数值线性代数

Вычислительные методы линейной алгебры

作者：Galois 爱求五次根

组织：深北莫数学学社分析小组

时间：2023/2/14

宗旨：执象而求，咫尺千里

时间是个常数，但对勤奋者来说，是个‘变数’。用‘分’来计算时间的人比用‘小时’来计算时间的人时间多 59 倍——雷巴柯夫

目录

第 1 章 部分内容回顾: 线性代数和数学分析 **1**

1.1 度量空间 . . . 1
1.2 度量空间上的极限过程 . . . 1
1.3 度量空间中的紧性 . . . 2
1.4 赋范空间 . . . 2
1.5 向量的 $p-$ 范数 . . . 3
1.5.1 Hölder 不等式 . . . 3
1.5.2 Minkovsky 不等式 . . . 4
1.6 矩阵范数 . . . 4
1.7 矩阵的算子范数 . . . 5

第 2 章 线性代数前置内容 **6**

2.1 欧几里得（酉）空间 . . . 6
2.2 等距矩阵 . . . 6
2.3 $p = 2$ 时的保距矩阵 . . . 7
2.4 正规矩阵 . . . 8
2.5 定号矩阵 . . . 8
2.6 矩阵的奇异值分解 . . . 9
2.7 酉不变范数 . . . 10
2.8 构造奇异值分解的替代方法 (альтернативный метод) . . . 10
2.9 秩亏近似 . . . 11

第 3 章 线性方程组求解的扰动理论 **12**

3.1 扰动理论 . . . 12
3.2 矩阵级数 . . . 12
3.3 逆矩阵 . . . 13
3.4 线性方程组的条件数 (обусловленность) . . . 14
3.5 解的相对误差估计 . . . 14
3.6 条件数 (Число обусловленности) . . . 14
3.7 矩阵和右侧部分的相容性 (согласованность) . . . 15
3.8 对角线优超 (диагональное преобладание) . . . 15

第 4 章 矩阵特征值的扰动与特征值的定位 (локализация) **17**

4.1 特征值的扰动 . . . 17
4.2 特征值的扰动与一般情况 . . . 17
4.3 代数多项式的根的连续性 . . . 18
4.4 特征值的定位（局部化） . . . 19

第 5 章 矩阵的特征值与向量的小扰动、Hadamard 不等式和矩阵范数的性质 **21**

5.1 特征值和向量的小扰动 . . . 21
5.2 简单特征值条件化 . . . 22
5.3 扰动分析 . . . 23
5.4 Hadamard 不等式 (Неравенство Адамара) . . . 23

5.5 辅助命题 . . . 24
5.6 矩阵 Euclid 范数的酉不变性 . . . 24

第 6 章 矩阵特征值的扰动（续）：谱距离和 Hermite 矩阵的谱 25
6.1 谱距离 . . . 25
6.2 谱距离的辅助推论 . . . 26
6.3 谱距离（续） . . . 27
6.4 Hermite 矩阵的谱 . . . 28
6.5 Hermite 矩阵的谱：辅助推论 . . . 28
6.6 Hermite 矩阵的谱（续） . . . 29

第 7 章 机器算数的特点 (особенности машинной арифметики) 30
7.1 机器数 . . . 30
7.2 机器算数公理 . . . 30
7.3 数字加减误差 . . . 31
7.4 数字乘除误差 . . . 32
7.5 例子：标量积的舍入误差 . . . 32
7.6 “坏”操作 . . . 32
7.7 例子 . . . 33
7.8 运算先后顺序 . . . 33
7.9 求解三角形方程组 . . . 33

第 8 章 线性方程组的直接解法和矩阵的 LU 分解 35
8.1 解决问题的直接方法 . . . 35
8.2 LU 分解 . . . 35
8.3 LU 分解的存在性 . . . 36
8.4 LU 分解的唯一性 . . . 36
8.5 LU 分解和 Gauss 消元法之间的联系 . . . 36
8.6 矩阵 LU 分解的舍入错误 . . . 37
8.7 主元的选择 . . . 37
8.8 矩阵的 LDL 分解和 Cholesky 分解 . . . 38
8.9 Cholesky 方法 . . . 38
8.10 Cholesky 方法中的误差 . . . 39

第 9 章 方阵的 QR 分解与反射、旋转矩阵 40
9.1 QR 分解 . . . 40
9.2 反射矩阵 . . . 40
9.3 排除具有反射的元素 (Исключение элементов с помощью отражений) . . . 41
9.4 旋转矩阵 . . . 41
9.5 反射和旋转方法的机器实现 . . . 42
9.6 正交化方法 . . . 42
9.7 正交性损失 . . . 43
9.8 正交性损失的处理 . . . 43
9.9 修正的 Gram-Schmidt 算法 . . . 44
9.10 双对角线化 . . . 44

第 10 章伪逆矩阵及其在求解线性方程组中的应用 45

10.1 Moore-Penrose 伪逆矩阵 . . . 45
10.2 矩阵的骨架（标架）分解 . . . 45
10.3 Moore-Penrose 伪逆矩阵 (псевдообратная матрица Мура-Пенроуза) . . . 46
10.4 伪逆矩阵的性质 . . . 47
10.5 应用伪逆矩阵求解线性方程组 . . . 47
10.6 线性方程组的正规伪解 . . . 48

第 11 章求特征值和特征向量的问题 50
11.1 特征值问题 . . . 50
11.2 特征值问题的稳定性 (устойчивость) . . . 50
11.3 特征多项式插值法 . . . 51
11.4 三对角线矩阵的特征多项式 . . . 52
11.5 求特征向量的逆迭代法 . . . 52

第 12 章求 Hermite 矩阵的特征值特征向量 53
12.1 Hermite 矩阵的反射法 . . . 53
12.2 直接旋转法 (прямой метод вращений) . . . 54
12.3 迭代旋转法 (итерационный метод вращений) . . . 55

第 13 章寻找非 Hermite 矩阵的特征值和特征向量与部分特征值问题 57
13.1 初等变换法 . . . 57
13.2 初等变换法：第一步 . . . 57
13.3 初等变换法：第二步 . . . 58
13.4 QR 算法 . . . 59
13.5 部分特征值问题 . . . 59
13.6 移位取逆迭代 (обратные итерации со сдвигом) . . . 60

第 14 章求解线性方程组的迭代算法（例子和迭代方法的收敛性） 61
14.1 迭代法求解线性方程组 . . . 61
14.2 矩阵的表述形式 . . . 61
14.3 迭代法的加速 . . . 62
14.4 显式和隐式法 . . . 62
14.5 迭代法的收敛性 . . . 63
14.6 迭代法的收敛性：Jacobi 法 . . . 63
14.7 迭代法的收敛性：上松弛法 (метод верхней релаксации) . . . 64
14.8 迭代法的收敛性: 简单迭代法 . . . 64
14.9 稳定迭代法的收敛准则 . . . 65

第 15 章求解线性方程组的迭代算法（迭代法的收敛速度估计） 66
15.1 迭代法的收敛速度 . . . 66
15.2 收敛速度估计 . . . 66
15.3 最优迭代参数 . . . 67
15.4 简单迭代法的收敛速度 . . . 68
15.5 非对称矩阵 B 的情况下的误差估计 . . . 68

第 16 章求解线性方程组的迭代算法 (带有 Chebyshev 参数集的迭代方法) 70
16.1 Chebyshev 多项式 . . . 70

16.2 Chebyshev 多项式在任意区间上的情况 . . . 71
16.3 Chebyshev 多项式与交错归一化 . . . 71
16.4 Chebyshev 多项式的应用 . . . 72
16.5 带有 Chebyshev 参数集的迭代法 . . . 72
16.6 隐式 Chebyshev 迭代法 . . . 74

第 17 章求解线性方程组的迭代算法（变分类型的迭代方法） 75
17.1 变分迭代算法 (итерационные методы вариационного типа) . . . 75
17.2 最小残差法 . . . 75
17.3 最小校正法 . . . 76
17.4 最速下降法 . . . 77
17.5 共轭梯度法 (метод сопряжённых градиентов) . . . 77
17.6 共轭梯度法中的误差最小化 . . . 78

第 1 章 部分内容回顾: 线性代数和数学分析

1.1 度量空间

在解决线性代数的相关计算问题时，需要对各种不同的研究对象引入“接近程度”的概念。接近程度通常采用度量（或距离）来刻画

定义 1.1 (度量与度量空间)

设 M 是一个非空集合，而 $\rho(x,y)$ 是一个非负（定）函数，对一切的 $x,y\in M$ 都有定义，且满足下述性质:

- $\rho(x,y)\geqslant 0,\forall x,y\in M$, 其中 $\rho(x,y)=0\Leftrightarrow x=y$ （正定性）
- $\rho(x,y)=\rho(y,x),\forall x,y\in M$ （对称性）
- $\rho(x,z)\leqslant\rho(x,y)+\rho(y,z),\forall x,y,z\in M$ （三角不等式）

那么函数 $\rho(x,y)$ 称为度量 (метрика)（或 x 与 y 之间的距离），集合 M 称为度量空间 (метрическое пространство)

♣

例题 1.1

- 集合 $\mathbb{R}$ 上的度量可以引入为绝对值函数：$\rho(x,y)=|x-y|$
- 空间 $\mathbb{R}^n$ 上的度量可以引入为 $p-$ 范数：$\rho_\alpha(x,y)=(\sum_{i=1}^n|x_i-y_i|^\alpha)^{\frac{1}{\alpha}},\alpha\geqslant 1$, 其中当 $\alpha=2$ 时称为欧几里得范数
- 闭区间连续函数空间 $C[a,b]$ 上的度量可以定义成：$\rho(x,y)=\max_{t\in[a,b]}|x(t)-y(t)|$

1.2 度量空间上的极限过程

定义 1.2 (序列的收敛)

序列 $\{x_n\},x_n\in M$ 如果满足：$\exists x\in M:\lim_{n\to\infty}\rho(x_n,x)=0$ 则称为是收敛的。在这种情况下，点 x 称为序列 $\{x_n\}$ 的极限。记为：$x=\lim_{n\to\infty}x_n$

♣

定义 1.3 (Cauchy 列)

序列 $\{x_n\},x_n\in M$ 称为柯西列（或基本列），如果满足:

$$(\forall\varepsilon>0)(\exists N>0):\forall n,m\geqslant N\Rightarrow\rho(x_n,x_m)\leqslant\varepsilon$$

♣

定义 1.4 (度量空间的完备)

如果度量空间 M 中的任意 Cauchy 列都收敛（满足序列收敛的 Cauchy 准则），则称该度量空间为完备的 (полный)

♣

定义 1.5 (闭集)

集合 $X\subset M$ 称为闭集 (замкнутое множество)，如果它包含自身所有极限点，即 $\forall\{x_n\},x_n\in X,\exists\lim_{n\to\infty}x_n=x\Rightarrow x\in X$（这里 x 是集合 X 的极限点）

♣

集合 X 的闭包 $\bar{X}$ 是集合 X 及其所有极限点的闭集

定义 1.6 (函数单点集的连续性)

数值函数 $f(x), x \in M$ 如果满足：$\forall \{x_n\}, x_n \neq x_0, \exists \lim_{n\to\infty} x_n = x_0 \Rightarrow \exists \lim_{n\to\infty} f(x_n) = f(x_0)$ (Heine 极限判定准则)，称函数 $f(x)$ 在点 x_0 处连续 ♣

1.3 度量空间中的紧性

定义 1.7 (紧集)

闭集 $X \subset M$ 是紧的 (компактное)，对 $\forall \{x_n\}, x_n \in X, \Rightarrow$，那么根据 Bolzano-Weierstrass 定理都可以找到一个子列 $\{x_{n_k}\} : \exists \lim_{k\to\infty} x_{n_k} = x \in X$ ♣

注 紧集的任何开覆盖中必有有限子覆盖

集合 $B_r(a) = \{x \in M : \rho(x, a) < r\}$ 是圆心为 a 半径为 r 的开球 (открытый шар)，集合 $B_r^c(a) = \{x \in M : \rho(x, a) \leqslant r\}$ 是闭球 (замкнутый шар)

定义 1.8 (开集)

集合 $X \subset M$ 称为开集，如果满足：

$$(\forall x \in X)(\exists \varepsilon > 0) : B_\varepsilon(x) \subset X$$

♣

定义 1.9 (有界集)

集合 $X \subset M$ 称为有界集，如果满足：

$$(\exists R > 0)(\exists a \in X) : X \subseteq B_R(a)$$

♣

定理 1.1 (球的嵌套定理/о вложенных шарах)

度量空间 M 是完备的 $\Leftrightarrow$ 对于任意的闭球序列 $B_{r_n}^c(a_n)$ 有：

$$(\exists \lim_{n\to\infty} r_n = 0)\left(B_{r_{n+1}}^c(a_{n+1}) \subset B_{r_n}^c(a_n)\right)(\forall n) \Rightarrow \exists x^* \in B_{r_n}^c(a_n), \forall n$$

♡

1.4 赋范空间

考虑一个实数或复数线性 (向量) 空间 V，在其上定义函数 $f(x) = \|x\|$ 满足：

1. $\|x\| \geqslant 0, \forall x \in V; \|x\| = 0 \Leftrightarrow x = 0$（正定性）
2. $\|\alpha x\| = |\alpha| \cdot \|x\|, \forall \alpha \in \mathbb{R}(\mathbb{C}), x \in V$（绝对齐次性）
3. $\|x + y\| \leqslant \|x\| + \|y\|$（三角不等式）

满足上面性质定义的函数 $\|x\|$ 是向量 x 的**范数**，空间 V 是**赋范空间 (нормированное пространство)**

在任何赋范空间中，可以借助如下范数引入度量

$$\rho(x, y) = \|x - y\|$$

定义 1.10 (Banach 空间/Банаховое пространство)

相对于与范数对应的度量完备的赋范空间称为 Banach 空间 ♣

定义 1.11

空间 V 上的两个范数 $\|\cdot\|_*$ 和 $\|\cdot\|_{**}$ 是等价的，如果满足：$\exists c_1, c_2 > 0$，有

$$c_1\|x\|_* \leqslant \|x\|_{**} \leqslant c_2\|x\|_*, \forall x \in V \quad \text{(你中有我，我中有你)}$$

♣

定理 1.2 (范数的等价性)

在任何有限维空间中，任何两个范数都是等价的

♡

1.5 向量的 $p-$ 范数

定义 1.12 (向量的 $p-$ 范数)

设 $V=\mathbb{C}^n$ 或 $V=\mathbb{R}^n$，且 $p\geqslant 1(p\in\mathbb{R})$。向量 $x=(x_1,\ldots,x_n)^T\in V$，称

$$\|x\|_p=\left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p}$$

为向量 x 的 $p-$ 范数

♣

引理 1.1 (范数比较引理)

函数 $f(p)=\|x\|_p$ 具有下述性质：

$$(\forall x\in V)(p_1,p_2>0):\|x\|_{p_1}=1\Rightarrow\begin{cases}\|x\|_{p_2}\leqslant 1, & p_2>p_1\\ \|x\|_{p_2}\geqslant 1, & p_2<p_1\end{cases}$$

♡

引理 1.2 (Young 不等式/неравенство Юнга)

设 $p,q>1:\frac{1}{p}+\frac{1}{q}=1$（称 p 与 q 共轭)。那么对任意的 $a,b\geqslant 0$，有

$$ab\leqslant\frac{a^p}{p}+\frac{b^q}{q}$$

♡

证明 当 $a=0$ 或 $b=0$ 时，不等式显然成立。此外，假设 $a,b>0$。辅助函数 $f(x)=\ln(x)$ 当 $x>0$ 时是凹（上凸）函数（因为 $f''(x)<0$)，因此：

$$(\forall\lambda\in[0,1]):\lambda\ln(x)+(1-\lambda)\ln(y)\leqslant\ln(\lambda x+(1-\lambda)y)$$

现在假设 $\lambda=\frac{1}{p},1-\lambda=\frac{1}{q},x=a^p,y=b^q$，那么有

$$\lambda\ln(x)+(1-\lambda)\ln(y)=\ln(ab)\leqslant\ln(\lambda x+(1-\lambda)y)=\ln\left(\frac{a^p}{p}+\frac{b^q}{q}\right)$$

1.5.1 Hölder 不等式

定理 1.3 (Hölder 不等式/неравенство Гёльдера)

对任意的 $p,q>1:\frac{1}{p}+\frac{1}{q}=1$，任意的 $x,y\in V=\mathbb{C}^n\,(\mathbb{R}^n)$

$$\left|\sum_{i=1}^{n}x_iy_i\right|\leqslant\|x\|_p\cdot\|y\|_q$$

♡

证明 当 $x=0$ 或者 $y=0$ 时，不等式显然成立。此外，假设 $x,y\neq 0$。

设 $\tilde{x}=\frac{x}{\|x\|_p},\tilde{y}=\frac{y}{\|y\|_q}$，那么 $\|\tilde{x}\|_p=\|\tilde{y}\|_q=1$。当 $a=\tilde{x}_i,b=\tilde{y}_i,i=1,\ldots,n$ 时，由引理（Young 不等式1.2）得：

$$|\tilde{x}_i|\cdot|\tilde{y}_i|\leqslant\frac{|\tilde{x}_i|^p}{p}+\frac{|\tilde{y}_i|^q}{q}.$$

对一切不同的 $i=1,\dots,n$，不等式两边求和：

$$\frac{1}{\|x\|_p\|y\|_q}\left|\sum_{i=1}^n x_iy_i\right|\leqslant\sum_{i=1}^n|\tilde{x}_i|\cdot|\tilde{y}_i|\leqslant\sum_{i=1}^n\left(\frac{|\tilde{x}_i|^p}{p}+\frac{|\tilde{y}_i|^q}{q}\right)=\frac{\|\tilde{x}_i\|_p^p}{p}+\frac{\|\tilde{y}_i\|_q^q}{q}=1$$

注 在空间 $\mathbb{R}^n$ 中，Hölder 不等式将变为等式，若 $\exists\alpha>0$：

$$|x_i|^p=\alpha\,|y_i|^q\,,x_iy_i\geqslant 0,\forall i=1,\dots,n$$

1.5.2 Minkovsky 不等式

定理 1.4 (Minkovsky 不等式/неравенство Минковского)

对任意的 $p>1$，任意的 $x,y\in V=\mathbb{C}^n\,(\mathbb{R}^n)$：

$$\|x+y\|_p\leqslant\|x\|_p+\|y\|_p$$

♡

证明 使用两次 Hölder 不等式 (1.3)（这里 $(p-1)q=p$）：

$$\sum_{i=1}^n|x_i|\,|x_i+y_i|^{p-1}\leqslant\|x\|_p\cdot\left(\sum_{i=1}^n\left(|x_i+y_i|^{p-1}\right)^q\right)^{\frac{1}{q}}=\|x\|_p\cdot(\|x+y\|_p)^{p/q}$$

$$\sum_{i=1}^n|y_i|\,|x_i+y_i|^{p-1}\leqslant\|y\|_p\cdot\left(\sum_{i=1}^n\left(|x_i+y_i|^{p-1}\right)^q\right)^{\frac{1}{q}}=\|y\|_p\cdot(\|x+y\|_p)^{p/q}$$

然后

$$\|x+y\|_p^p=\sum_{i=1}^n|x_i+y_i|^p\leqslant\sum_{i=1}^n(|x_i|+|y_i|)\,|x_i+y_i|^{p-1}\leqslant(\|x+y\|_p)^{p/q}\,(\|x\|_p+\|y\|_p)$$

剩下的就是不等式两边同时除以 $(\|x+y\|_p)^{p/q}$

注 Minkovsky 不等式是向量 $p-$ 范数定义下的三角不等式

注 向量 $p-$ 范数的**特殊情况**：

- $p=2\Rightarrow$ 向量的 Euclid 范数 $\|x\|_2$。在这种情况下，Hölder 不等式等价于 Cauchy-Буняковский 不等式
- $p=1\Rightarrow$

$$\|x\|_1=\sum_{i=1}^n|x_i|$$

Manhattan 距离诱导的范数

- 极限情况：

$$\|x\|_\infty=\max_{i=1,\dots,n}|x_i|=\lim_{p\to\infty}\|x\|_p$$

Chebyshev 距离诱导的范数

1.6 矩阵范数

在固定矩阵尺寸的矩阵空间中，矩阵的范数可以作为向量范数引入，通过将矩阵看作元素排列成表格形式的向量

例如：（矩阵 $p-$ 范数）

$$A=(a_{ij})\in\mathbb{R}^{n\times m}\Rightarrow\|A\|_{vec,p}=\left(\sum_{i=1}^n\sum_{j=1}^m|a_{ij}|^p\right)^{1/p},p\geqslant 1$$

然而，除了向量范数的通常性质外，矩阵范数还有一个额外的**乘法性质 (свойство мультипликативности)**：

$$\|AB\|\leqslant\|A\|\cdot\|B\|$$

例题 1.2 Euclid 范数（Frobenius 范数）即当 $p=2$ 时，矩阵的 $p-$ 范数：

$$\|A\|_E = \left(\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}|^2\right)^{1/2}, A\in\mathbb{C}^{n\times m}$$

$$若A=[a_1,\ldots,a_m]\in\mathbb{C}^{n\times m}, B=[b_1,\ldots,b_m]^T\in\mathbb{C}^{m\times k}, 则有AB=a_1b_1^T+\ldots+a_mb_m^T$$

$$\|AB\|_E\leqslant\left\|a_1b_1^T\right\|_E+\ldots+\left\|a_mb_m^T\right\|_E=\|a_1\|_2\cdot\|b_1\|_2+\ldots+\|a_m\|_2\cdot\|b_m\|_2\leqslant\|A\|_E\cdot\|B\|_E$$

其中第二个不等号借助 Cauchy-Буняковский 不等式完成证明

1.7 矩阵的算子范数

考虑矩阵空间 $A\in\mathbb{C}^{n\times m}$，设在空间 $\mathbb{C}^n$ 中给定范数 $\|\cdot\|_*$，而在另一个空间 $\mathbb{C}^m$ 中给定范数 $\|\cdot\|_{**}$。由此定义矩阵 A 的范数：

$$\|A\|_{***}=\max_{x\neq0}\frac{\|Ax\|_*}{\|x\|_{**}}=\max_{x:\|x\|_{**}=1}\|Ax\|_*$$

上述范数也是算子的诱导范数，满足向量范数与算子诱导范数的**相容性 (свойство согласованности)**：

$$\|Ax\|_*\leqslant\|A\|_{***}\cdot\|x\|_{**}$$

例题 1.3

$$\|A\|_p=\max_{x\neq0}\frac{\|Ax\|_p}{\|x\|_p}$$

是矩阵范数，因为当 $AB\neq0$：

$$\|AB\|_p=\max_{x\neq0,Bx\neq0}\frac{\|ABx\|_p}{\|Bx\|_p}\frac{\|Bx\|_p}{\|x\|_p}\leqslant\max_{Bx\neq0}\frac{\|ABx\|_p}{\|Bx\|_p}\cdot\max_{x\neq0}\frac{\|Bx\|_p}{\|x\|_p}\leqslant\|A\|_p\cdot\|B\|_p$$

在线性代数课程中已证明 ($A\in\mathbb{C}^{n\times m}$) 的诱导范数 (подчиненный норм) 的性质：

- $\|A\|_1=\max_{j=\overline{1,m}}\sum_{i=1}^{n}|a_{ij}|$（每列元素和的最大值）
- $\|A\|_\infty=\max_{i=\overline{1,n}}\sum_{j=1}^{m}|a_{ij}|$（每行元素和的最大值）
- $\|A\|_2$ 等于矩阵 A 的最大奇异值，即 $\sqrt{|\lambda|}$ 中的最大值，其中 λ 是矩阵 AA^* 和 A^*A 的特征值。$\|A\|_2$ 称为矩阵的**谱范数 (спектральная норма)**

第 2 章 线性代数前置内容

2.1 欧几里得（酉）空间

定义 2.1 (标量积)

设 V 是实或复向量空间，在空间 V 上对每一个二元有序向量对 x, y 定义一个数 $\langle x, y\rangle$，满足：

1. $\langle x, x\rangle \geqslant 0, \forall x \in V; \langle x, x\rangle = 0 \Leftrightarrow x = \theta$ （正定性）
2. $\langle x, y\rangle = \overline{\langle y, x\rangle}$ （共轭对称性）
3. $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle, \forall \alpha \in \mathbb{C}(\mathbb{R})$ （半双齐次性）
4. $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ （双可加性）

数 $\langle x, y\rangle$ 称为向量 x 和 y 的标量积 (скалярное произведение) 或内积

♣

定义 2.2 (Euclid 空间与酉空间)

具有标量积的实向量空间称为 Euclid 空间 (евклидовое)；具有标量积的复向量空间称为酉空间 (унитарное)

♣

如果 $\langle x, y\rangle = 0$，那么向量 x 和 y **正交 (ортогональны)**

设 $e = \{e_1, \ldots, e_n\}$ 是空间 V 的一个基，数组 $x_1, \ldots, x_n$ 和 $y_1, \ldots, y_n$ 分别是向量 x、y 在该基下的坐标。若 e 是标准正交基 (ортонормированный)，当且仅当：

$$\langle x, y\rangle = x_1\overline{y_1} + \ldots + x_n\overline{y_n}, \forall x, y \in V$$

由标量积诱导的向量的范数（长度）：$\|x\| = \sqrt{\langle x, x\rangle}$

定理 2.1

线性空间 V 中的某一范数能够由一个标量积诱导生成，充分必要条件是该范数满足平行四边形恒等式 (тождество параллелограмма)

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2, \forall x, y \in V$$

♡

2.2 等距矩阵

定义 2.3 (等距同构/изометричной)

矩阵 $Q \in \mathbb{C}^{n\times n}$ 如果满足：

$$(\forall x \in \mathbb{C}^n) : \|Qx\| = \|x\|$$

则称该矩阵保持空间 $\mathbb{C}^n$ 中的范数 $\|\cdot\|$（或关于范数等距同构）

♣

例题 2.1 $Q-$ 置换矩阵 (матрица перестановки)，即经过行或列置换的单位矩阵

研究保持向量 $p-$ 范数的矩阵 Q 的结构：设范数为 $p-$ 范数 $\|\cdot\| = \|\cdot\|_p$，$x = e_i$ 是恒等矩阵的第 i 列 $\Rightarrow \|Qe_i\| = 1$。即矩阵 Q 的每一个列向量的 $p-$ 范数都是 1，此外由 Hölder 不等式 (1.3) 可知：

$$|y^TQx| \leq \|Q^Ty\|_q \cdot \|x\|_p$$

$$\|Q^Ty\|_q = \max_{x\neq\theta}\frac{|y^TQx|}{\|x\|_p} = \max_{x\neq\theta}\frac{|y^TQx|}{\|Qx\|_p} = \max_{z\neq\theta}\frac{|y^Tz|}{\|z\|_p} = \|y\|_q$$

其中第一个等式和最后一个等式由 Hölder 不等式完成证明，因此矩阵 Q 的每一个行向量的范数也等于 1

设 $p<2, q>2, \frac{1}{p}+\frac{1}{q}=1$ 对矩阵 Q 的任意列 q_i 和任意行 $\tilde{q}_j^T$，有：

$$\|q_i\|_2 \leqslant \|q_i\|_p = 1, \left\|\tilde{q}_j^T\right\|_2 \geqslant \left\|\tilde{q}_j^T\right\|_q = 1, \quad \sum_{i=1}^{n}\|q_i\|_2^2 = \sum_{j=1}^{n}\|\tilde{q}_j\|_2^2$$

这只有在满足 $(\forall i,j):\|q_i\|_2=\|\tilde{q}_j\|_2=1 \Rightarrow (\forall i):\|q_i\|_p=\|q_i\|_2=1 \Rightarrow$ 在第 q_i 列中，恰好只有一个元素的模等于 1，其余的元素都是零。同理可得矩阵 Q 行的相关结构。当 $p>2$ 时结果也一样成立

因此，当 $p\neq 2$ 时，$Q=P\operatorname{diag}(d_1,\ldots,d_n)$，其中 P 为置换矩阵，$(\forall i):|d_i|=1$

2.3 $p=2$ 时的保距矩阵

设 $p=2$，然后空间 V 为 Euclid（酉）空间

设 Q 是一个等距矩阵

标量积可以借助相应空间中的范数诱导表示（极化恒等式）：

- 在空间$V=\mathbb{R}^n$中：$\langle x,y\rangle=\frac{1}{2}\left(\|x+y\|^2-\|x\|^2-\|y\|^2\right)$
- 在空间$V=\mathbb{C}^n$中：$\langle x,y\rangle=\frac{1}{2}\left((\|x+y\|^2-\|x\|^2-\|y\|^2)+i(\|x+iy\|^2-\|x\|^2-\|iy\|^2)\right)$

由此可知，等距矩阵保持标量积：

$$(\forall x,y\in V):\langle Qx,Qy\rangle=\langle x,y\rangle$$

若 $x=e_i, y=e_j$，有 $\langle q_i,q_j\rangle=\delta_{ij}$（Kronecker 符号）$\Rightarrow$ 矩阵 Q 的列构成标准正交向量组，即有：$Q^*Q=I$，(Q^* 是矩阵 Q 的共轭转置)。因此矩阵 Q 是酉矩阵：$Q^*=Q^{-1}$

在 2− 范数的情况下，酉矩阵 Q 是等距矩阵

注 如果矩阵 Q 是酉矩阵，那么矩阵 Q^{-1} 也是酉矩阵

定理 2.2 (Schur 分解定理/теорема Шура)

设矩阵 $A\in\mathbb{C}^{n\times n}, \lambda_1,\lambda_2,\ldots,\lambda_n\in\operatorname{Spec}(A)$（即 $\lambda_1,\ldots,\lambda_n$ 是矩阵 A 的全部特征值），则存在 n 阶酉矩阵 Q 满足：

- Q^*AQ 为上三角阵
- $\operatorname{diag}(Q^*AQ)=\operatorname{diag}(\lambda_1,\ldots,\lambda_n)$

♡

证明 设 (λ_1,x_1) 是矩阵 A 的一个特征对，即：

$$Ax_1=\lambda_1x_1$$

由 x_1 是一个非零向量及 Householder 变换性质得：存在一个 Householder 变换 H_1 满足

$$H_1x_1=\delta e_1,\quad \delta\neq 0$$

注意到方程 (2.3) 等价于 $H_1AH_1^*H_1x_1=\lambda_1H_1x_1$，则由式 $H_1x_1=\delta e_1,\ \delta\neq 0$ 有 $H_1AH_1^*e_1=\lambda_1e_1$，则矩阵 $H_1AH_1^*$ 必有如下形式：

$$H_1AH_1^*=\begin{bmatrix}\lambda_1 & *\\ 0 & A_1\end{bmatrix},\quad A_1\in\mathbb{C}^{(n-1)\times(n-1)}$$

由上述讨论定理归纳即证

2.4 正规矩阵

定义 2.4 (正规矩阵)

矩阵 A，如果满足 $A^*A = AA^*$，则称为正规矩阵

♣

注 特别的，正规矩阵可能是 Hermite（对称）矩阵，满足 $(A^* = A)$（共轭对称）；也有可能是酉（正交）矩阵，满足 $\left(A^* = A^{-1}\right)$

定理 2.3 (正规矩阵的充要条件)

矩阵 $A \in \mathbb{C}^{n\times n}$ 是正规矩阵，当且仅当，空间 $\mathbb{C}^n$ 中存在一个由矩阵 A 的特征向量构成的标准正交基

♡

定理 2.4 (Hermite 矩阵的充要条件)

正规矩阵 A 称为 Hermite 矩阵，当且仅当，其所有特征值都是实数

♡

定理 2.5 (酉矩阵的充要条件)

正规矩阵称为酉矩阵，当且仅当，其所有特征值模等于 1

♡

定义 2.5 (Hermite 分解)

对任意的矩阵 $A \in \mathbb{C}^{n\times n}$ 都有有效的 Hermite 分解式：

$$\exists H, K : A = H + iK, H^* = H, K^* = K$$

♣

2.5 定号矩阵

定义 2.6 (正定矩阵/положительно определенная)

矩阵 $A \in \mathbb{C}^{n\times n}$ 是正定的，如果满足 $(\forall x \in \mathbb{C}^n, x \neq \theta) : \langle x, Ax\rangle > 0$

♣

定义 2.7 (负定矩阵/отрицательно определенная)

矩阵 A 是负定的，如果满足 $(\forall x \in \mathbb{C}^n, x \neq \theta) : \langle x, Ax\rangle < 0$

♣

同理可以定义非正定和非负定矩阵

注 如果矩阵 A 是定号的，那么在 Hermite 分解式 $A = H + iK$ 中，$K = 0$

注 定号矩阵 A 不一定都是对称矩阵

定理 2.6

矩阵 $A \in \mathbb{C}^{n\times n}$ 是非负定（正定）的 $\Leftrightarrow$ 矩阵的所有特征值都是非负的（正的）

♡

定理 2.7 (Sylvester 方法)

若矩阵 A 是非负定（正定）的，则其任意顺序主子矩阵 B 也都是非负定（正定）的

♡

证明 设 $y = [\tilde{y}, \theta]^T \in \mathbb{C}^n$

$$(\forall \tilde{y}) : \langle y, Ay\rangle = y^*Ay = y^* \begin{pmatrix} B & * \\ * & * \end{pmatrix} y = (\tilde{y})^* B\tilde{y} \geqslant 0 (> 0)$$

对任意的矩阵 A 来说，矩阵 A^*A 总是非负定的：$A^*A \geqslant 0$（既是正矩阵，又是自伴矩阵）

2.6 矩阵的奇异值分解

定理 2.8 (奇异值分解/сингулярное разложение матрицы)

设 $A \in \mathbb{C}^{n\times m}, r = \operatorname{rg}(A)$，则存在奇异值 $\sigma_1 \geqslant \ldots \geqslant \sigma_r > 0$ 与酉矩阵 $U \in \mathbb{C}^{m\times m}, V \in \mathbb{C}^{n\times n}$，满足：

$$A = V\Sigma U^*, \text{其中} \Sigma = \begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \vdots \\ & & \sigma_r & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix} \in \mathbb{C}^{n\times m}$$

♡

证明 矩阵 A^*A 是正规的 Hermite 矩阵 $\Rightarrow \exists U : U = [u_1, \ldots, u_m] \in \mathbb{C}^{m\times m}$ 是酉矩阵，且满足

$$U^*A^*AU = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_m^2\right), \sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_m$$

设 $\sigma_r > 0, (\forall i > r) : \sigma_i = 0$。然后：

$$U_r = [u_1, \ldots, u_r], \Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r), U_r^*A^*AU_r = \Sigma_r^2 \Rightarrow \left(\Sigma_r^{-1}U_r^*A^*\right)\left(AU_r\Sigma_r^{-1}\right) = I$$

下记 $V_r = AU_r\Sigma_r^{-1} = [v_1, \ldots, v_r]$，则 $V_r^*V_r = I \Rightarrow$ 矩阵 V_r 是酉矩阵

这时 $(\forall i = 1, \ldots, r) : Au_i = \sigma_i v_i$，而 $(\forall i \geqslant r+1) : u_i^*A^*Au_i = \|Au_i\|_2^2 = 0 \Rightarrow Au_i = 0$

任意地扩充 V_r 直到为酉矩阵 $V \in \mathbb{C}^{n\times n}$，则有：

$$AU = V\begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow V^*AU = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}$$

则由乘上非退化矩阵秩不变性质 $\Rightarrow r = \operatorname{rg}(A)$

性质 奇异值分解的性质，左右奇异向量分别是 v_i 和 u_i：

1) 奇异值 σ_i 唯一定义
2) 若 $\sigma_1 > \ldots > \sigma_r > 0$，则奇异向量 $u_1, \ldots, u_r$ 和 $v_1, \ldots, v_r$ 是唯一定义的，精确到因子 $c : |c| = 1$
3)

$$Au_i = \begin{cases} \sigma_i v_i, & i = \overline{1, r} \\ 0, & i = \overline{r+1, m} \end{cases}$$

$$A^*v_i = \begin{cases} \sigma_i u_i, & i = \overline{1, r} \\ 0, & i = \overline{r+1, n} \end{cases}$$

4) $A = \sum_{i=1}^{r} \sigma_i v_i u_i^*$
5) $\ker(A) = \mathcal{L}(u_{r+1}, \ldots, u_m)$
6) $\operatorname{im}(A) = \mathcal{L}(v_1, \ldots, v_r)$
7) $\ker(A^*) = \mathcal{L}(v_{r+1}, \ldots, v_n)$
8) $\operatorname{im}(A^*) = \mathcal{L}(u_1, \ldots, u_r)$

注 $\operatorname{im}(A) = \ker(A^*)^T, \quad \operatorname{im}(A^*) = \ker(A)^T$

2.7 酉不变范数

定义 2.8 (酉不变范数/унитарно инвариантный норм)

设 $\forall A \in \mathbb{C}^{n\times m}, \forall U \in \mathbb{C}^{k\times n}, V \in \mathbb{C}^{m\times l}$，且 $\|\cdot\|$ 为矩阵范数。若

$$U, V \in \mathcal{U} \Rightarrow \|A\| = \|UAV\|$$

则称矩阵范数 $\|\cdot\|$ 为酉不变范数

♣

例题 2.2 酉不变范数的例子：

- 范数 $\|A\|_2$：

$$\|UAV\|_2 = \max_{x\neq\theta} \frac{\|UAVx\|_2}{\|x\|_2} = \max_{z\neq\theta} \frac{\|UAVV^*z\|_2}{\|V^*z\|_2} = \max_{z\neq\theta} \frac{\|UAz\|_2}{\|z\|_2} = \max_{z\neq\theta} \frac{\|Az\|_2}{\|z\|_2} = \|A\|_2.$$

这里替换 $z = Vx$，同时 $\|V^*z\|_2 = \sqrt{\langle V^*z, V^*z\rangle} = \sqrt{\langle z, VV^*z\rangle} = \|z\|_2$。

- 范数 $\|A\|_E$：

$$\|UAV\|_E^2 = \|A\|_E^2$$

范数若为酉不变范数，那么借助矩阵的奇异值分解 $A = V\Sigma U^*$，可以得到 $\|A\| = \|\Sigma\|$，即酉不变范数由矩阵的奇异值唯一决定。特别地：

- $\|A\|_2 = \sigma_1$ （最大奇异值）
- $\|A\|_E = \left(\sigma_1^2 + \ldots + \sigma_r^2\right)^{1/2}$

2.8 构造奇异值分解的替代方法 (альтернативный метод)

考虑算子 2 范数 $\|A\|_2$ 的定义，有：

$$\|A\|_2 = \max_{x:\|x\|_2=1} \|Ax\|_2 = \sigma \geqslant 0$$

根据 Weierstrass 定理可知，连续函数 $y = \|Ax\|_2$ 在紧集 $\{x : \|x\|_2 = 1\}$ 中的点 z 上取到最大值。那么

$$\|z\|_2 = 1, \quad Az = \sigma y, \quad \|y\|_2 = 1$$

构造酉矩阵 $U = [z, U_1]$ 和 $V = [y, V_1]$，然后有：

$$V^*AU = \begin{bmatrix} y^* \\ V_1^* \end{bmatrix} [\sigma y, AU_1] = \begin{pmatrix} \sigma & b^* \\ 0 & B \end{pmatrix}$$

$$\left\| V^*AU \begin{bmatrix} \sigma \\ b \end{bmatrix} \right\|_2^2 \geqslant \left(\sigma^2 + b^*b\right)^2, \quad \|(\sigma, b^*)\|_2^2 = \sigma^2 + b^*b \Rightarrow \|V^*AU\|_2^2 \geqslant \sigma^2 + b^*b$$

谱范数 $\|\cdot\|_2$ 酉不变 $\Rightarrow \|A\|_2 = \|V^*AU\|_2 = \sigma$。因此

$$b = 0, \quad V^*AU = \begin{pmatrix} \sigma & 0 \\ 0 & B \end{pmatrix}$$

接下来，借助数学归纳法，可以对子矩阵 B 构造奇异值分解

2.9 秩亏近似

定理 2.9

设 $k < \mathrm{rg}(A), A_k = \sum_{i=1}^{k} \sigma_i v_i u_i^*$。那么有:

$$\|A - A_k\|_2 = \min_{B:\mathrm{rg}(B)=k} \|A - B\|_2 = \sigma_{k+1}$$

♡

证明 对矩阵 B 应用关于秩与亏的定理，有:

$\mathrm{rg}(B) = k \Rightarrow \dim \ker(B) = n - k \Rightarrow \exists z \neq \theta, z \in \ker B \cap \mathcal{L}(u_1, \ldots, u_{k+1})$.

(最后一步是因为: 若 $\ker B \cap \mathcal{L}(u_1, \ldots, u_{k+1}) = \{\theta\}$，则由 $\ker B$ 中的向量与向量 $u_1, \ldots, u_{k+1}$ 可以构成 $\mathbb{R}^n$ 的基，与维数矛盾)

再假设 $\|z\|_2 = 1$, $z = a_1 u_1 + \ldots + a_{k+1} u_{k+1}$, $a_1^2 + \ldots + a_{k+1}^2 = 1$，那么有

$$Az = a_1 \sigma_1 v_1 + \ldots + a_{k+1} \sigma_{k+1} v_{k+1}$$

$$\|A - B\|_2^2 \geqslant \|(A - B)z\|_2^2 = \|Az\|_2^2 = \sum_{i=1}^{k+1} a_i^2 \sigma_i^2 \geqslant \sigma_{k+1}^2$$

推论 2.1

(在谱范数的意义下) 非退化方阵的最小奇异值等于到最近退化矩阵的距离

♡

第 3 章 线性方程组求解的扰动理论

3.1 扰动理论

线性代数中出现的主要任务之一是通过已知的 x 找到（计算）某些映射 $f(x)$ 下的值。通常算法计算会出现一些误差（错误）。进而出现了下述问题：随着 x 的微小扰动，映射值 $f(x)$ 的计算结果会有多大的差距?

例题 3.1 设 $f(x)$ 是关于变量 x 的可微函数，Δx 是自变量 x 的微小扰动。然后根据 Taylor 公式有 $f(x+\Delta x)\approx f(x)+f'(x)\Delta x$。假设 $x\neq 0, f'(x)\neq 0$。然后可以引入下述对扰动敏感度的度量：

$$\frac{f(x+\Delta x)-f(x)}{\|f(x)\|}\approx\left(\frac{f'(x)}{\|f(x)\|}\|x\|\right)\frac{\Delta x}{\|x\|}$$

定义 3.1 (条件数的相对度量)

称

$$\operatorname{cond} f(x)=\frac{\|f'(x)\|}{\|f(x)\|}\|x\|$$

为目标问题敏感度（条件数）的相对度量 (Относительной мерой чувствительности задачи (числом обусловленности))

♣

关于线性方程组 $Ax=f$ 解的扰动估计，当误差可能出现在右侧部分和矩阵中时：

$$f\to f+\Delta f, A\to A+\Delta A,\quad \Delta f,\Delta A\text{-微小数值}$$

3.2 矩阵级数

假设 $A_k\in\mathbb{C}^{n\times n}, k=1,2,\ldots$。考虑矩阵级数

$$\sum_{k=0}^{\infty}A_k \tag{3.1}$$

定义 3.2 (矩阵级数的收敛)

若存在矩阵 $A\in\mathbb{C}^{n\times n}$ 满足:

$$\exists \lim_{N\to\infty}\|S_N-A\|=0, \text{部分和序列} S_N=\sum_{k=0}^{N}A_k$$

则称矩阵级数 (3.1) 收敛

♣

定理 3.1 (矩阵级数收敛的充分条件)

设矩阵级数 $\sum_{k=0}^{\infty}\|A_k\|$ 收敛，则矩阵级数 $\sum_{k=0}^{\infty}A_k$ 也收敛

♡

证明 证明借助 Cauchy 收敛准测和估计

$$\|S_{n+p}-S_n\|=\left\|\sum_{k=n+1}^{n+p}A_k\right\|\leqslant\sum_{k=n+1}^{n+p}\|A_k\|$$

定义 3.3 (Neumann 级数/ряд Неймана)

矩阵级数 $\sum_{k=0}^{\infty}F^k$ 称为 Neumann 级数

♣

若 $\|F\|<1$，那么矩阵级数 $\sum_{k=0}^{\infty}\|F^k\|$ 收敛 $\Rightarrow$Neumann 级数收敛。（借助不等式 $\|F^k\|\leqslant\|F\|^k$）

定义 3.4 (谱半径/спектральный радиус)

矩阵 F 的最大模特征值称为谱半径，记为 $\rho(F)$

♣

定义 3.5 (矩阵收敛的必要条件)

若 $\rho(F)<1$，则称矩阵 F 收敛

♣

引理 3.1

矩阵 $F\in\mathbb{C}^{n\times n}$ 的 Neumann 级数收敛 $\Leftrightarrow$ F 收敛

♡

证明

$\Leftarrow$：由 Schur 分解定理2.2可知，存在酉矩阵 $Q\left(Q^{-1}=Q^*\right)$ 满足：

$$T=(t_{ij})=Q^{-1}FQ-\text{ 上三角矩阵}$$

$T^k=Q^{-1}F^kQ\Rightarrow$ 矩阵级数 $\sum_{k=0}^{\infty}F^k$ 收敛 $\Leftrightarrow$ 级数 $\sum_{k=0}^{\infty}T^k$ 收敛

假设 $D_\varepsilon=\operatorname{diag}\left(1,\varepsilon,\ldots,\varepsilon^{n-1}\right)$，那么当 $i\leqslant j$ 时，有 $\left\{D_\varepsilon^{-1}TD_\varepsilon\right\}_{ij}=\varepsilon^{j-i}t_{ij}$

矩阵 $D_\varepsilon^{-1}TD_\varepsilon$ 的对角线元绝对值都小于 $1\Rightarrow$ 当 $\varepsilon>0$ 且充分小时，有 $\left\|D_\varepsilon^{-1}TD_\varepsilon\right\|_2<1\Rightarrow$ 矩阵 $D_\varepsilon^{-1}TD_\varepsilon$ 的 Neumann 级数收敛

$\left(D_\varepsilon^{-1}TD_\varepsilon\right)^k=D_\varepsilon^{-1}T^kD_\varepsilon\Rightarrow$ 矩阵级数 $\sum_{k=0}^{\infty}T^k$ 收敛 $\Leftrightarrow$ 级数 $\sum_{k=0}^{\infty}\left(D_\varepsilon^{-1}TD_\varepsilon\right)^k$ 收敛

证明了矩阵 F 的 Neumann 级数的收敛性

$\Rightarrow$：反证法。假设 $(\exists\lambda:|\lambda|\geqslant 1)(\exists\tilde{x}\neq\theta):F\tilde{x}=\lambda\tilde{x}$，则有：

$$\sum_{k=0}^{N}F^k\tilde{x}=\left(1+\lambda+\ldots+\lambda^N\right)\tilde{x}$$

但 $\nexists\lim_{N\to+\infty}\left(1+\lambda+\ldots+\lambda^N\right)\Rightarrow$Neumann 级数发散。矛盾！

3.3 逆矩阵

引理 3.2 (矩阵可逆的充分条件)

如果矩阵 F 满足 $\|F\|<1$，那么矩阵 $A=I-F$ 可逆，并且有：

$$(I-F)^{-1}=\sum_{k=0}^{\infty}F^k;\quad\left\|(I-F)^{-1}\right\|\leqslant\frac{\|I\|}{1-\|F\|}$$

♡

证明 将矩阵相乘并取极限 $N\to\infty$：

$$(I-F)\left(\sum_{k=0}^{N}F^k\right)=I-F+F-F^2+F^2-F^3+\ldots+F^N-F^{N+1}=I-F^{N+1}\to I$$

因此，矩阵 $I-F$ 可逆，且第一个等式得证

接下来证明关于范数的不等式：

$$\left\|F^k\right\|\leqslant\|I\|\cdot\|F\|^k,\quad\left\|\sum_{k=0}^{N}F^k\right\|\leqslant\|I\|\cdot\sum_{k=0}^{N}\|F\|^k\leqslant\frac{\|I\|}{1-\|F\|}$$

推论 3.1

设矩阵 A 时非退化方阵，ΔA 为扰动矩阵，若 $\left\|A^{-1}\Delta A\right\|<1$，则矩阵 $(A+\Delta A)$ 可逆，且有

$$(A+\Delta A)^{-1}=\left(I+A^{-1}\Delta A\right)^{-1}A^{-1}=\sum_{k=0}^{\infty}\left(-A^{-1}\Delta A\right)^kA^{-1}=A^{-1}\sum_{k=0}^{\infty}\left(-\Delta AA^{-1}\right)^k$$

♡

3.4 线性方程组的条件数 (обусловленность)

考虑一个具有非退化矩阵 A 的方程组 $Ax = f, f \neq \theta$

再考虑扰动方程组 $(A + \Delta A)\tilde{x} = f + \Delta f$

问：$\tilde{x}$ 与 x 有多大的差别?

若 $\left\|A^{-1}\Delta A\right\| < 1$，那么有

$$\tilde{x} - x = (A + \Delta A)^{-1}(f + \Delta f) - A^{-1}f = \left((A + \Delta A)^{-1} - A^{-1}\right) f + (A + \Delta A)^{-1}\Delta f =$$
$$= \left(\left(I + A^{-1}\Delta A\right)^{-1} - I\right)\left(A^{-1}f\right) + \left(I + A^{-1}\Delta A\right)^{-1}\left(A^{-1}\Delta f\right) =$$
$$= \left(\sum_{k=1}^{\infty}\left(-A^{-1}\Delta A\right)^{k}\right)\left(A^{-1}f\right) + \left(\sum_{k=0}^{\infty}\left(-A^{-1}\Delta A\right)^{k}\right)\left(A^{-1}\Delta f\right)$$

用 $\|\tilde{x} - x\|$ 除以 $\|x\| = \left\|A^{-1}f\right\|$ 并估计解的相对误差 (относительная погрешность решения):

又根据不等式 $\|f\| \leqslant \|A\| \cdot \|x\|$，有

$$\frac{\|\tilde{x} - x\|}{\|x\|} \leqslant \left\|A^{-1}\Delta A\right\| \|I\| \sum_{k=0}^{\infty}\left\|A^{-1}\Delta A\right\|^{k} + \frac{\left\|A^{-1}\right\| \|\Delta f\| \|A\|}{\|f\|}\|I\| \sum_{k=0}^{\infty}\left\|A^{-1}\Delta A\right\|^{k} =$$
$$= \|I\| \frac{\left\|A^{-1}\right\| \cdot \|A\|}{1 - \|A^{-1}\Delta A\|}\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta f\|}{\|f\|}\right)$$

定义 3.6 (条件数/число обусловленности)

$\operatorname{cond} A = \left\|A^{-1}\right\| \cdot \|A\|$ 称为矩阵 A 的条件数 ♣

条件数表示出解 x 对矩阵和右侧部分的微小扰动的敏感性，对于退化矩阵一般有 $A = \infty$。

3.5 解的相对误差估计

取定的非退化矩阵 A 的可获得的相对误差估计是通过选择的右侧部分及其扰动来实现的

$$A = \sum_{k=1}^{n} \sigma_k v_k u_k^* - \text{奇异值分解}$$
$$f = v_1, \quad x = \sigma_1^{-1} u_1, \quad \Delta f = \varepsilon v_n, \quad \Delta A = 0$$

然后有：

$$\operatorname{cond}_2 A = \left\|A^{-1}\right\|_2 \cdot \|A\|_2 = \frac{\sigma_1}{\sigma_n}, \quad \|f\| = 1, \quad \|\Delta f\|_2 = \varepsilon$$
$$\|\tilde{x} - x\|_2 = \left\|A^{-1}\Delta f\right\|_2 = \left\|\varepsilon\sigma_n^{-1}u_n\right\|_2 = \varepsilon\sigma_n^{-1}$$
$$\Rightarrow \frac{\|\tilde{x} - x\|_2}{\|x\|_2} = \frac{\varepsilon\sigma_n^{-1}}{\sigma_1^{-1}} = \left\|A^{-1}\right\|_2 \cdot \|A\|_2 \cdot \frac{\|\Delta f\|_2}{\|f\|_2} = \operatorname{cond}_2 A \cdot \frac{\|\Delta f\|_2}{\|f\|_2}$$

3.6 条件数 (Число обусловленности)

矩阵的条件数显著取决于当前所使用的矩阵范数

例题 3.2 考虑范数 $\|A\|_2 = \sigma_1(A) = \sigma_{\max}(A)$。然后 $\left\|A^{-1}\right\|_2 = \sigma_{\max}\left(A^{-1}\right) = \frac{1}{\sigma_{\min}(A)}$，则：

$$\operatorname{cond}_2 A = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

注 几何解释：$\operatorname{cond}_2 A$ 表明变换 A 沿其主方向上不均匀地拉伸空间

注 在范数 $\|\cdot\|_1$ 和 $\|\cdot\|_\infty$ 下计算矩阵的条件数时，考虑利用

$$\forall A : \|A\|_\infty = \left\|A^T\right\|_1 \Rightarrow \operatorname{cond}_\infty A = \operatorname{cond}_1 A^T$$

很有用

例题 3.3

$$A=\begin{pmatrix}5&4\\4&3\end{pmatrix}, A^{-1}=\begin{pmatrix}-3&4\\4&-5\end{pmatrix}, \sigma_{\max}(A)\approx 8.12, \sigma_{\min}(A)\approx 0.12$$

$$\|A\|_1=\|A\|_\infty=9, \left\|A^{-1}\right\|_1=\left\|A^{-1}\right\|_\infty=9$$

$$\operatorname{cond}_2 A\approx 66, \quad \operatorname{cond}_1 A=\operatorname{cond}_\infty A=81$$

如果矩阵 A 没有发生扰动，并且使用范数 $\|\cdot\|_2$，则右侧部分 1% 的误差将导致解 66% 的误差

3.7 矩阵和右侧部分的相容性 (согласованность)

如果可以人为地限制可容许扰动的类别，则可以改进对右侧矩阵和向量的小扰动造成的的相对误差的一般估计。

定义 3.7

由矩阵 A 的奇异值分解，有

$$A=\sum_{k=1}^{n}\sigma_k v_k u_k^* \Rightarrow A^{-1}=\sum_{k=1}^{n}\sigma_k^{-1}u_k v_k^* \Rightarrow x=\sum_{k=1}^{n}\frac{v_k^* f}{\sigma_k}u_k$$

假定矩阵 A 是固定的 (没有扰动)，即 $\Delta A=0$

在一般情形下，借助估计

$$\sqrt{\sum_{k=1}^{n}\left(\frac{v_k^*\Delta f}{\sigma_k}\right)^2}\leqslant\frac{\|\Delta f\|_2}{\sigma_n}$$

有

$$\frac{\|\tilde{x}-x\|_2}{\|x\|_2}=\sqrt{\frac{\sum_{k=1}^{n}\left(\frac{v_k^*\Delta f}{\sigma_k}\right)^2}{\sum_{k=1}^{n}\left(\frac{v_k^* f}{\sigma_k}\right)^2}}\leqslant\frac{\sigma_1}{\sigma_n}\cdot\frac{\|\Delta f\|_2}{\|f\|_2}=\operatorname{cond}_2 A\cdot\frac{\|\Delta f\|_2}{\|f\|_2}$$

若存在某一个 r，使得

$$v_{r+1}^*\Delta f=\ldots=v_n^*\Delta f=0$$

成立，则

$$\frac{\|\tilde{x}-x\|_2}{\|x\|_2}\leqslant\frac{\sigma_1}{\sigma_r}\cdot\frac{\|\Delta f\|_2}{\|f\|_2}, \quad \frac{\sigma_1}{\sigma_r}<\frac{\sigma_1}{\sigma_n}$$

此时，称矩阵和右侧部分是相容的 (согласованный)

3.8 对角线优超 (диагональное преобладание)

定义 3.8 (行 (列) 对角线优超/строчное(столбцовое) диагональное преобладание)

矩阵 $A=\{a_{ij}\}\in\mathbb{C}^{n\times n}$ 若满足：

$$|a_{ii}|>r_i=\sum_{j=1,j\neq i}^{n}|a_{ij}|, i=1,\ldots,n$$

则称矩阵 A 行对角线优超

对称地，若满足：

$$|a_{jj}| > c_j = \sum_{i=1, i\neq j}^{n} |a_{ij}|, j = 1, \ldots, n$$

则称矩阵 A 列对角线优超 ♣

注 对角线优超又名**对角占优、对角优势**

定理 3.2

具有行或列对角线优超的矩阵是非退化的矩阵 ♡

证明 设 $\operatorname{diag}(A) = \operatorname{diag}(a_{11}, \ldots, a_{nn}), \operatorname{off}(A) = A - \operatorname{diag}(A)$

由对角线优超的定义可得：所有的对角元素都非 0

不妨设矩阵 A 有严格的行对角线优超 $\Leftrightarrow \left\|(\operatorname{diag}(A))^{-1}\operatorname{off}(A)\right\|_\infty < 1$。这意味着：$A = \operatorname{diag}(A)\left(I + (\operatorname{diag}(A))^{-1}\operatorname{off}(A)\right)$ 是非退化的，由引理3.2即得

列对角线优超证明同理，仅需将 A 变成 A^T 即可

对于具有对角线优超的矩阵，也可以推导出 $\left\|A^{-1}\right\|_\infty$ 的估计值：设 $\alpha = \min_i(|a_{ii}| - r_i) > 0$，并且对于向量 x，有 $|x_k| = \|x\|_\infty$。那么

$$\alpha\|x\|_\infty = \alpha|x_k| \leqslant |a_{kk}x_k| - \sum_{j\neq k}|a_{kj}||x_j| \leqslant |a_{kk}x_k| - \left|\sum_{j\neq k} a_{kj}x_j\right| \leqslant \left|\sum_j a_{kj}x_j\right| \leqslant$$

$$\leqslant \max_k \left|\sum_j a_{kj}x_j\right| = \|Ax\|_\infty$$

$$\left\|A^{-1}\right\|_\infty = \max_{z\neq 0}\frac{\left\|A^{-1}z\right\|_\infty}{\|z\|_\infty} = \max_{x=A^{-1}z\neq 0}\frac{\|x\|_\infty}{\|Ax\|_\infty} \leqslant \frac{1}{\alpha} = \frac{1}{\min_i(|a_{ii}| - r_i)}$$

又已知 $\|A\|_\infty \leqslant \max_i(|a_{ii}| + r_i)$

因此：

$$\operatorname{cond}_\infty A \leqslant \frac{\max_i(|a_{ii}| + r_i)}{\min_i(|a_{ii}| - r_i)}$$

通过过渡到矩阵的转置，也可以得到估计值

$$\|A\|_1 \leqslant \max_j(|a_{jj}| + c_j), \quad \left\|A^{-1}\right\|_1 \leqslant \frac{1}{\beta}, \beta = \min_j(|a_{jj}| - c_j)$$

$$\operatorname{cond}_1 A \leqslant \frac{\max_j(|a_{jj}| + c_j)}{\min_j(|a_{jj}| - c_j)}$$

第 4 章 矩阵特征值的扰动与特征值的定位 (локализация)

4.1 特征值的扰动

设 $\lambda(A)$ 是矩阵 A 的谱 (спектр)，即其所有特征值构成的集合

引理 4.1

设 $\mu \in \lambda(A+\Delta A)$，但 $\mu \notin \lambda(A)$。那么：$\frac{1}{\|(A-\mu I)^{-1}\|_2} \leqslant \|\Delta A\|_2$。

♡

证明 $\mu \notin \lambda(A) \Rightarrow \exists (A-\mu I)^{-1}$

$$(A+\Delta A)-\mu I = (A-\mu I)+\Delta A - \text{退化} \Rightarrow$$
$$\Rightarrow I + (A-\mu I)^{-1}\Delta A \text{是退化的} \Rightarrow \left\|(A-\mu I)^{-1}\Delta A\right\|_2 \geqslant 1 \Rightarrow$$
$$\Rightarrow \left\|(A-\mu I)^{-1}\right\|_2 \cdot \|\Delta A\|_2 \geqslant 1$$

定理 4.1

假设矩阵 A 可对角化 (диагонализуема)，即 $\exists P : P^{-1}AP = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) = \Lambda$。若 $\mu \in \lambda(A+\Delta A)$，则

$$\min_{i=1,\ldots,n} |\mu - \lambda_i| \leqslant \left\|P^{-1}\right\|_2 \cdot \|P\|_2 \cdot \|\Delta A\|_2$$

♡

证明

若 $\mu \in \lambda(A)$，则 $\min_{i=1,\ldots,n} |\mu-\lambda_i| = 0$。不等式显然成立

若 $\mu \notin \lambda(A)$，则 $\mu \notin \lambda(\Lambda)$，但是 $\mu \in \lambda\left(\Lambda + P^{-1}\Delta A P\right)$。根据引理4.1得：

$$\min_{i=1,\ldots,n} |\mu-\lambda_i| = \frac{1}{\|(\Lambda-\mu I)^{-1}\|_2} \leqslant \left\|P^{-1}\Delta A P\right\|_2 \leqslant \left\|P^{-1}\right\|_2 \cdot \|P\|_2 \cdot \|\Delta A\|_2$$

谱对小扰动的敏感性由特征向量矩阵 P 的条件数表出

4.2 特征值的扰动与一般情况

定理 4.2

设 $P^{-1}AP = J$，其中 J 是矩阵 A 的 Jordan 型，且 $\mu \in \lambda(A+\Delta A)$。那么存在 $\lambda \in \lambda(A)$，满足：

$$\frac{|\mu-\lambda|^m}{1+|\mu-\lambda|+\ldots+|\mu-\lambda|^{m-1}} \leqslant \left\|P^{-1}\right\|_2 \cdot \|P\|_2 \cdot \|\Delta A\|_2$$

其中，m 是对应 λ 的 Jordan 块 (жорданова клетка) 的最大阶数

♡

证明

若 $\mu \in \lambda(A)$，不等式显然成立；

若 $\mu \notin \lambda(A)$，则由引理4.1可知：$\frac{1}{\|(J-\mu I)^{-1}\|_2} \leqslant \left\|P^{-1}\Delta A P\right\|_2$。设 Jordan 型 J 由 Jordan 块 $J_1, \ldots, J_k$ 组成。那么

$$\left\|(J-\mu I)^{-1}\right\|_2 \leqslant \max_{i=1,\ldots,k} \left\|(J_i-\mu I)^{-1}\right\|_2$$

再假设 $J_i = \lambda I + N_i$，且该 Jordan 块阶数为 m。那么 $N_i^m = 0$，$\|N_i\|_2 = 1$

$$\left\|(J_i - \mu I)^{-1}\right\|_2 = \left\|((\lambda-\mu)I + N_i)^{-1}\right\|_2 \leqslant \frac{1}{|\lambda-\mu|}\left\|\left(I + (\lambda-\mu)^{-1}N_i\right)^{-1}\right\|_2 \leqslant$$

$$\leqslant \frac{1}{|\lambda-\mu|}\left(\|I\|_2 + \left\|(\lambda-\mu)^{-1}N_i\right\|_2 + \ldots + \left\|(\lambda-\mu)^{-1}N_i\right\|_2^{m_i-1}\right) = \frac{1 + |\mu-\lambda| + \ldots + |\mu-\lambda|^{m_i-1}}{|\lambda-\mu|^{m_i}}$$

如果具有最大 Jordan 块阶数 m 的矩阵有 ε 阶的扰动，那么扰动矩阵的任意特征值都可能与原始矩阵的某个特征值相差 $|\varepsilon|^{\frac{1}{m}}$ 个数量级

4.3 代数多项式的根的连续性

考虑具有复系数的 n 次多项式 $P(z)$:

$$P(z) = z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0$$

令带有复系数 $a_{i,s}$ 的多项式序列

$$P_s(z) = z^n + a_{n-1,s}z^{n-1} + \ldots + a_{1,s}z + a_{0,s}$$

收敛于多项式 $P(z) : \lim_{s\to\infty} a_{i,s} = a_i, \forall i$

问题是：当 s 取很大的数值时，多项式 $P_s(z)$ 与 $P(z)$ 的根有什么样的关系?

引理 4.2 (关于根的定位引理/о локализации корня)

对任意的 n 次多项式 $P(z)$ 和任意的复数 z_0，至少有多项式 $P(z)$ 的一个根位于圆

$$|z - z_0| \leqslant \sqrt[n]{|P(z_0)|}$$

中

♡

证明 由 Vieta 定理可以推出：若 $z_1, \ldots, z_n$ 是多项式 $P(z)$ 所有的根，那么有 $|z_1 \ldots z_n| = |a_0|$。这意味着至少有一个根位于圆

$$|z| \leqslant \sqrt[n]{|a_0|}$$

上。用以点 z_0 为中心的 Taylor 展开式分解多项式 $P(z)$:

$$P(z) = (z - z_0)^n + \ldots + P(z_0)$$

再一次使用 Vieta 定理 $\Rightarrow \exists z-$ 根：$|z - z_0| \leqslant \sqrt[n]{|P(z_0)|}$

$z_1, \ldots, z_r$ 表示多项式 $P(z)$ 的两两不同的根。根据引理4.2，在每一个圆

$$|z - z_i| \leqslant \sqrt[n]{|P_s(z_i)|}, i = 1, \ldots, r$$

内都存在多项式 $P_s(z)$ 的至少 (по крайней мере) 一个根 $z_{i,s}$

此外，$(\forall i) : \lim_{s\to\infty} P_s(z_i) = 0$。这意味着当 s 充分大时，圆无公共点。若 $r = n$，则根 $z_{1,s}, \ldots, z_{r,s}$ 是两两不同的。也就意味着:

$$\lim_{s\to\infty} z_{i,s} = z_i, i = 1, \ldots, r$$

如果多项式 $P(z)$ 只有重数为 1 的简单根 (простый корень)，则证明其根对系数的连续依赖性

现在假设 $z_1, \ldots, z_n$ 和 $z_{1,s}, \ldots, z_{n,s}$ 分别是多项式 $P(z)$ 和 $P_s(z)$ 的全部根组，并且根之间可以相等（成重数关系）

定理 4.3

多项式 $P_s(z)$ 的根可以以下面的方式重新进行编号 (перенумеровать)，即满足关系:

$$\lim_{s\to\infty} z_{i,s} = z_i, i = 1, \ldots, n$$

♡

证明 （借助第二类数学归纳法完成证明）对于 $n=1$ 次多项式，定理显然成立

设定理对任意次数 $\leqslant n-1$ 的多项式都成立。证明定理对于 n 次多项式也成立

在多项式 $P_s(z)$ 的根中可以找到一个根 $z_{1,s}$，满足 $\lim_{s\to\infty} z_{1,s}=z_1$。分别用 $z-z_1$ 和 $z-z_{1,s}$ 因式分解多项式 $P(z)$ 与 $P_s(z)$：

$$P(z)=R(z)\,(z-z_1)\,,R(z)=z^{n-1}+b_{n-2}z^{n-2}+\ldots+b_1z+b_0$$

$$P_s(z)=R_s(z)\,(z-z_{1,s})\,,R_s(z)=z^{n-1}+b_{n-2,s}z^{n-2}+\ldots+b_{1,s}z+b_{0,s}$$

$$b_{n-2}=a_{n-1}+z_1,b_{n-3}=a_{n-2}+z_1b_{n-2},\ldots,b_0=a_1+z_1b_1$$

$$b_{n-2,s}=a_{n-1,s}+z_{1,s},b_{n-3,s}=a_{n-2,s}+z_{1,s}b_{n-2,s},\ldots,b_{0,s}=a_{1,s}+z_{1,s}b_{1,s}$$

因此，$\lim_{s\to\infty} R_s(z)=R(z)$

$z_2,\ldots,z_n$ 是多项式 $R(x)$ 的根，而 $z_{2,s},\ldots,z_{n,s}$ 是多项式 $R_s(z)$ 的根。根据归纳假设，这些根可以按照定理条件的方式进行排序

已证，代数多项式的根是多项式系数在其任意变化域（基域）上的连续函数

该命题可用于特征值-矩阵 A 特征多项式的根理论

4.4 特征值的定位（局部化）

矩阵 A 特征值的定位问题：需要确定（特征值）所在的复平面上的那些大致区域。区域的计算应该比数值地求特征值要简单得多

最简单估计：矩阵 A 至少有一个特征值

$$\lambda:|\lambda|\leqslant|\det A|^{1/n},\quad \text{引理4.2的推论},\quad P(z)=\det(A-zI),z_0=0$$

另一种最简单估计：

$$\forall\lambda-A\text{的特征值}\Rightarrow|\lambda|\leqslant\|A\|,\quad \forall\|\cdot\|-\text{诱导范数}$$

$$\|A\|=\max_{x\neq0}\frac{\|Ax\|}{\|x\|}\geqslant\frac{\|Ax_\lambda\|}{\|x_\lambda\|}=|\lambda|,\quad Ax_\lambda=\lambda\cdot x_\lambda$$

特别的，$|\lambda|\leqslant\|A\|_2=\sigma_1=\sigma_{\max}(A)$

如果同样地考虑矩阵 A^{-1}，则可以得到 $\sigma_n\leqslant|\lambda|$。因此对任意的矩阵 A 及其任意的特征值 λ，有以下结论：

$$\sigma_n\leqslant|\lambda|\leqslant\sigma_1$$

定理 4.4

令 $A\in\mathbb{C}^{n\times n}$，考虑圆：

$$R_i=\{z\in\mathbb{C}:|a_{ii}-z|\leqslant r_i\},\quad C_i=\{z\in\mathbb{C}:|a_{ii}-z|\leqslant c_i\}$$

其中

$$r_i=\sum_{j=1,j\neq i}^{n}|a_{ij}|,\quad c_j=\sum_{i=1,i\neq j}^{n}|a_{ij}|$$

然后，对 $\forall\lambda\in\lambda(A)$

$$\lambda\in\bigcup_{i=1}^{n}R_i,\quad \lambda\in\bigcup_{i=1}^{n}C_i$$

♡

证明 若 $\lambda\notin\cup_iR_i$，则矩阵 $A-\lambda I$ 具有行对角优超 $\Rightarrow$ 由定理3.2知，矩阵 $A-\lambda I$ 非退化 $\Rightarrow\lambda$ 不是特征值

对于矩阵 C_i 具有列对角线优超的情况，证明同理

定义 4.1 (Gerschgorin 圆盘/круг Гершгорина)

圆 R_i, C_i 称为 Gerschgorin 圆盘

$$R_i = \{z \in \mathbb{C} : |a_{ii} - z| \leqslant r_i\}, \quad C_i = \{z \in \mathbb{C} : |a_{ii} - z| \leqslant c_i\}$$

♣

注 详情建议参考视频 Proving From the Graph – The Gershgorin Theorem： address

定理 4.5 (Gerschgorin 圆盘定理)

若 m 个 Gerschgorin 圆构成了一个与其他圆孤立（不相交）的区域 G，那么区域 G 恰好包含矩阵 A 的 m 个特征值

♡

证明 令 $A(t) = \operatorname{diag}(A) + t \cdot \operatorname{off}(A), t \in [0,1], \operatorname{off}(A) = A - \operatorname{diag}(A)$。用 $G(t)$ 表示与 G 中的圆具有相同中心的 Gerschgorin 圆盘的并集，用 $G'(t)$ 表示其余圆盘的并集。那么 $A(1) \equiv A$，

$$G(t) \subset G(1) = G, \quad G'(t) \subset G'(1) = G'$$

根据条件 $G \cap G' = \emptyset$，因此 $G(t) \cap G'(t) = \emptyset, \forall t \in [0,1]$

由定理4.3可知，存在连续的特征值函数 $\lambda_1(t), \ldots, \lambda_m(t)$，满足

$$\{\lambda_1(0), \ldots, \lambda_m(0)\} = G(0)$$

$$(\forall t \geqslant 0) : \lambda_1(t), \ldots, \lambda_m(t) \in \lambda(A(t))$$

再设 $t_i = \sup\{t \geqslant 0 : \lambda_i(t) \in G(t)\}$

若 $t_i < 1$，那么 $(\forall t > t_i) : \lambda_i(t) \in G'(t)$。但（由 $\lambda_i(t)$ 连续性），有：$\lambda_i(t_i) \in G(t_i) \cap G'(t_i) \neq \emptyset$，这是不可能的。矛盾!$\Rightarrow (\forall i = 1, \ldots, m) : t_i \geqslant 1$，

$$\Rightarrow \lambda_1(1), \ldots, \lambda_m(1) \in G$$

推论 4.1

若 Gerschgorin 圆盘两两互不相交，那么 Gerschgorin 圆盘族中的每一个都恰好仅包含一个特征值

♡

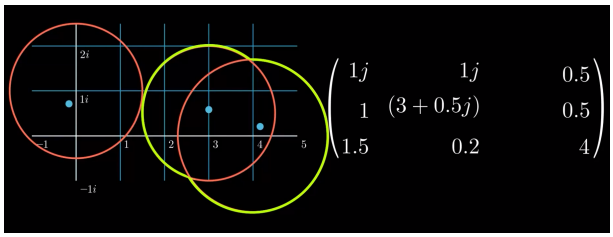


图 4.1: Gerschgorin 圆盘

第 5 章 矩阵的特征值与向量的小扰动、Hadamard 不等式和矩阵范数的性质

5.1 特征值和向量的小扰动

假设矩阵 A 只有简单的（两两不同的）特征值，并且扰动矩阵形如

$$A(\varepsilon) = A + A_1\varepsilon + \mathcal{O}\left(\varepsilon^2\right)$$

令 P 是 A 的特征向量矩阵，$\Lambda = P^{-1}AP$ 是矩阵 A 的特征值构成的对角矩阵。假定：

$$\Omega(\varepsilon) = P^{-1}A(\varepsilon)P = \Lambda + \Omega_1\varepsilon + \mathcal{O}\left(\varepsilon^2\right), \quad \Omega_1 = P^{-1}A_1P$$

矩阵 $\Omega(\varepsilon)$ 与 $A(\varepsilon)$ 具有相同的特征值。对于很小的 ε，矩阵 $\Omega(\varepsilon)$ 的 Gerschgorin 圆不会相交，表明矩阵 $\Omega(\varepsilon)$ 具有简单的（两两不同的）特征值

令 $\Lambda(\varepsilon)$ 和 $Z(\varepsilon)$ 分别为矩阵 $\Omega(\varepsilon)$ 的特征值对角矩阵和特征向量矩阵：

$$\Lambda(\varepsilon) = Z^{-1}(\varepsilon)\Omega(\varepsilon)Z(\varepsilon), \quad \Lambda(\varepsilon) - \text{对角阵}$$

$$\Lambda(\varepsilon) = \Lambda + \Lambda_1\varepsilon + \tilde{\Lambda}(\varepsilon), \quad \Lambda(0) = \Lambda, \quad \tilde{\Lambda}(\varepsilon) = \bar{o}(\varepsilon)$$

$$Z(\varepsilon) = I + Z_1\varepsilon + \tilde{Z}(\varepsilon), \quad Z(0) = I, \quad \tilde{Z}(\varepsilon) = \bar{o}(\varepsilon)$$

对于足够小的 $\varepsilon > 0, \operatorname{diag} Z(\varepsilon) \neq 0$。矩阵 $Z(\varepsilon)$ 在具有到列归一化的精度下被定义。因此，认为：

$$\operatorname{diag} Z(\varepsilon) = I, \quad \operatorname{diag} Z_1 = 0, \quad \operatorname{diag} \tilde{Z}(\varepsilon) = 0$$

将上述矩阵的分解式代入 $\Omega(\varepsilon)Z(\varepsilon) = Z(\varepsilon)\Lambda(\varepsilon)$：

$$\left(\Lambda + \Omega_1\varepsilon + \mathcal{O}\left(\varepsilon^2\right)\right)\left(I + Z_1\varepsilon + \tilde{Z}(\varepsilon)\right) = \left(I + Z_1\varepsilon + \tilde{Z}(\varepsilon)\right)\left(\Lambda + \Lambda_1\varepsilon + \tilde{\Lambda}(\varepsilon)\right) \tag{5.1}$$

拆开括号，合并同类项，其中关于 ε 的线性项为：

$$\left(\Lambda Z_1 - Z_1\Lambda + \Omega_1 - \Lambda_1\right)\varepsilon + \ldots = 0$$

因此，对于很小的 ε，有：

$$\Lambda Z_1 - Z_1\Lambda = \Lambda_1 - \Omega_1 \Rightarrow \Lambda_1 = \operatorname{diag}\Omega_1, \quad \Lambda Z_1 - Z_1\Lambda = -\operatorname{off}\Omega_1$$

回到关系式 (5.1)：

$$\left(\Lambda + \Omega_1\varepsilon + \mathcal{O}\left(\varepsilon^2\right)\right)\left(I + Z_1\varepsilon\right) - \left(I + Z_1\varepsilon\right)\left(\Lambda + \Lambda_1\varepsilon\right) = \mathcal{O}\left(\varepsilon^2\right)$$

当 $\varepsilon : \varepsilon > 0$ 很小时

$$\left\|\left(I + Z_1\varepsilon\right)^{-1}\right\| = \mathcal{O}(1) \Rightarrow \left(I + Z_1\varepsilon\right)^{-1}\left(\Lambda + \Omega_1\varepsilon + \mathcal{O}\left(\varepsilon^2\right)\right)\left(I + Z_1\varepsilon\right) - \left(\Lambda + \Lambda_1\varepsilon\right) = \mathcal{O}\left(\varepsilon^2\right)$$

因此矩阵 $\Lambda + \Omega_1\varepsilon + \mathcal{O}\left(\varepsilon^2\right)$ 的特征值具有到 $\mathcal{O}\left(\varepsilon^2\right)$ 的精度，且与矩阵 $\Lambda + \Lambda_1\varepsilon$ 的对角元一致。所以 $\tilde{\Lambda} = \mathcal{O}\left(\varepsilon^2\right)$

矩阵 $\tilde{Z}$ 满足关系

$$\left(\Lambda + \Omega_1\varepsilon + \mathcal{O}\left(\varepsilon^2\right)\right)\tilde{Z} - \tilde{Z}\left(\Lambda + \Lambda_1\varepsilon + \mathcal{O}\left(\varepsilon^2\right)\right) = \mathcal{O}\left(\varepsilon^2\right), \quad \operatorname{diag}\tilde{Z} = 0$$

这是一个关于矩阵 $\tilde{Z}$ 的元素的线性方程组，具有非退化的（当 ε 很小时）方阵。因此 $\tilde{Z} = \mathcal{O}\left(\varepsilon^2\right)$

定理 5.1

设 $P^{-1}AP = \Lambda$ 是一个对角矩阵，对角元为矩阵 A 的两两不同的特征值。那么对于很小的 ε，矩阵 $A(\varepsilon) = A + A_1\varepsilon + \mathcal{O}\left(\varepsilon^2\right)$ 可对角化：

$$P^{-1}(\varepsilon)A(\varepsilon)P(\varepsilon) = \Lambda(\varepsilon)$$

并且有

$$\Lambda(\varepsilon)=\Lambda+\Lambda_1\varepsilon+\mathcal{O}\left(\varepsilon^2\right),\quad P(\varepsilon)=P\left(I+Z_1\varepsilon+\mathcal{O}\left(\varepsilon^2\right)\right)$$
$$\Lambda_1=\operatorname{diag}\left(P^{-1}A_1P\right)$$

而矩阵 Z_1 由方程

$$\operatorname{diag}Z_1=0,\quad \Lambda Z_1-Z_1\Lambda=-\operatorname{off}\left(P^{-1}A_1P\right)$$

唯一确定

♡

推论 5.1

矩阵 $A(\varepsilon)$ 的特征值 $\lambda_i(\varepsilon)$ 具有形式

$$\lambda_i(\varepsilon)=\lambda_i+q_i^TA_1p_i\varepsilon+\mathcal{O}\left(\varepsilon^2\right)$$

其中 q_i^T 是矩阵 P^{-1} 的每一行。除此之外

$$\left|q_i^TA_1p_i\right|\leqslant\|A_1\|_2\cdot\|q_i\|_2\cdot\|p_i\|_2=\|A_1\|_2\cdot s\left(\lambda_i\right)$$

数 $s\left(\lambda_i\right)$ 称为特征值 λ_i 的条件数 (обусловленность)。该数表示出到重特征值 λ_i 的矩阵的距离

♡

5.2 简单特征值条件化

定理 5.2

设矩阵 A 带有条件数为 $s\left(\lambda_i\right)$ 的简单特征值（代数重数等于 1 的特征值）λ_i。那么存在矩阵 $A+\Delta A$，对于该矩阵而言：λ_i 是重特征值，并且成立

$$\|\Delta A\|_2\leqslant\frac{\|A\|_2}{\sqrt{s^2\left(\lambda_i\right)-1}}$$

♡

证明 假设矩阵 A 已化简为上三角型（根据 Schur 定理2.2）：

$$A=\begin{bmatrix}\lambda_i & z^T\\ 0 & B\end{bmatrix}$$

然后

$$p_1=\begin{bmatrix}1 & 0 & \dots & 0\end{bmatrix}^T,\quad q_1^T=\begin{bmatrix}1 & v^T\end{bmatrix}\Rightarrow s\left(\lambda_i\right)=\left(\|v\|_2^2+1\right)^{1/2}$$
$$v^TB+z^T=\lambda_iv^T\left(\text{因为：}q_1^T-\text{矩阵}A\text{的特征向量，即}q_1^TA=\lambda_iq_1^T\right)$$

令 $\tilde{B}=B+\frac{vz^T}{\|v\|_2^2}\Rightarrow\lambda_i$ 是矩阵 $\tilde{B}$ 的特征值。现假设

$$\Delta A=\begin{bmatrix}0 & 0\\ 0 & \frac{vz^T}{\|v\|_2^2}\end{bmatrix}\Rightarrow\|\Delta A\|_2\leqslant\frac{\left\|z^T\right\|_2}{\|v\|_2}\leqslant\frac{\|A\|_2}{\|v\|_2}=\frac{\|A\|_2}{\sqrt{s^2\left(\lambda_i\right)-1}}$$

数 λ_i 是矩阵 $A+\Delta A$ 的重数 $\geqslant2$ 的特征值

5.3 扰动分析

设对一切的 $\varepsilon : |\varepsilon| < \varepsilon_0$，矩阵级数 $A(\varepsilon) = \sum_{k=0}^{\infty} A_k \varepsilon^k$ 收敛。假设矩阵 A_0 仅有简单特征值（代数重数为 1 的特征值）。那么当 ε 足够小时，矩阵 $A(\varepsilon)$ 通过 $P(\varepsilon)$ 对角化：

$$P^{-1}(\varepsilon) A(\varepsilon) P(\varepsilon) = \Lambda(\varepsilon)$$

$$\Lambda(\varepsilon) = \sum_{k=0}^{\infty} \Lambda_k \varepsilon^k, \quad P(\varepsilon) = \sum_{k=0}^{\infty} P_k \varepsilon^k$$

求矩阵 Λ_k，P_k。令 $Z_k = P_0^{-1} P_k$，$\Omega_k = P_0^{-1} A_k P_0$，然后有

$$(\Lambda_0 + \Omega_1 \varepsilon + \ldots)(I + Z_1 \varepsilon + \ldots) = (I + Z_1 \varepsilon + \ldots)(\Lambda_0 + \Lambda_1 \varepsilon + \ldots)$$

令 $\varepsilon^k, k = 1, 2, 3, \ldots$ 项的系数对应相等：

$$\Lambda_0 Z_k - Z_k \Lambda_0 = \Lambda_k - \Phi_k, \quad \Phi_k = \sum_{i=1}^{k-1} (\Omega_i Z_{k-i} - Z_{k-i} \Omega_i) + \Omega_k$$

因此

$$\Lambda_k = \operatorname{diag} \Phi_k, \quad \Lambda_0 Z_k - Z_k \Lambda_0 = -\operatorname{off} \Phi_k$$

已知当 $i \leqslant k-1$ 时 Λ_i，Z_i 的值，（由归纳）可以得到 Λ_k，Z_k

5.4 Hadamard 不等式 (Неравенство Адамара)

定理 5.3

假设矩阵 $A = [a_1 \ldots a_n] \in \mathbb{C}^{n \times n}$，$\det A \neq 0$。那么

$$|\det(A)| \leqslant \|a_1\|_2 \cdot \ldots \cdot \|a_n\|_2$$

♡

证明 对向量 $a_1, \ldots, a_n$ 进行 Gram-Schmidt 正交化 (процесс ортогонализации Грама-Шмидта) 过程，得到标准正交向量组 $b_1, \ldots, b_n$。令 $B = [b_1 \ldots b_n]$ 是一个标准正交矩阵。由 Gram-Schmidt 过程的流程可知：

$$\exists C = (c_{ij}) \text{-上三角阵}: \ A = BC$$

$$(1 \leqslant i \leqslant j) : c_{ij} = \langle a_j, b_i \rangle ; \quad (j \leqslant i \leqslant n) : c_{ij} = 0$$

$$(\forall i = 1, \ldots, n) : a_i = \sum_{j=1}^{i} \langle a_i, b_j \rangle b_j, \quad \|a_i\|_2^2 = \sum_{j=1}^{i} |\langle a_i, b_j \rangle|^2$$

$$|\det A|^2 = \det(A^* A) = \det(C^* B^* B C) = \det C^* C = |\det C|^2 = \prod_{i=1}^{n} |\langle a_i, b_i \rangle|^2 \leqslant \prod_{i=1}^{n} \sum_{j=1}^{i} |\langle a_i, b_j \rangle|^2 = \prod_{i=1}^{n} \|a_i\|_2^2$$

推论 5.2 (Hadamard 等式成立的充要条件)

Hadamard 不等式变成等式 $\Leftrightarrow$

$$(\forall i) : |\langle a_i, b_i \rangle|^2 = \sum_{j=1}^{i} |\langle a_i, b_j \rangle|^2 \Leftrightarrow (\forall i) : a_i = \langle a_i, b_i \rangle b_i$$

♡

5.5 辅助命题

引理 5.1

对任意的矩阵 $A \in \mathbb{C}^{n\times m}$，成立

$$\|A\|_E^2 = \operatorname{tr}(AA^*)$$

♡

证明 令 $A = (a_{ij}), /AA^* = (b_{ij})$，那么

$$(\forall i,j): b_{ij} = \sum_{k=1}^{m} a_{ik}\bar{a}_{jk} \Rightarrow b_{ii} = \sum_{k=1}^{m} a_{ik}\bar{a}_{ik} = \sum_{k=1}^{m} |a_{ik}|^2 \Rightarrow \operatorname{tr}(AA^*) = b_{11} + \ldots + b_{nn} = \sum_{i=1}^{n}\sum_{k=1}^{m} |a_{ik}|^2 = \|A\|_E^2$$

引理 5.2

对任意的矩阵 $A, B \in \mathbb{C}^{n\times m}$，成立

$$\|A-B\|_E^2 = \|A\|_E^2 + \|B\|_E^2 - 2\operatorname{Re}\operatorname{tr}(BA^*)$$

♡

证明 Euclid 矩阵范数 $\|\cdot\|_E$ 是由标量积诱导范数。因此

$$\|A-B\|_E^2 = \langle A-B, A-B\rangle = \langle A,A\rangle + \langle B,B\rangle - \langle B,A\rangle - \langle A,B\rangle = \|A\|_E^2 + \|B\|_E^2 - \langle B,A\rangle - \overline{\langle B,A\rangle}$$

以下为矩阵 $A = (a_{ij})$ 和 $B = (b_{ij})$ 标量积的定义：

$$\langle B,A\rangle = \sum_{i=1}^{n}\sum_{k=1}^{m} b_{ik}\bar{a}_{ik}$$

同时，若 $BA^* = (c_{ij})$，则

$$c_{ij} = \sum_{k=1}^{m} b_{ik}\bar{a}_{jk} \Rightarrow \operatorname{tr}(BA^*) = \sum_{i=1}^{n}\sum_{k=1}^{m} b_{ik}\bar{a}_{ik} = \langle B,A\rangle \quad \Rightarrow \langle B,A\rangle + \overline{\langle B,A\rangle} = 2\operatorname{Re}\operatorname{tr}(BA^*)$$

5.6 矩阵 Euclid 范数的酉不变性

推论 5.3 (引理5.1的推论：矩阵范数 $\|\cdot\|_E$ 的酉不变性)

矩阵范数 $\|\cdot\|_E$ 具有酉不变性

♡

证明 设 $A \in \mathbb{R}^{n\times m}$。对任意的酉矩阵 $U \in \mathbb{R}^{n\times n}, V \in \mathbb{R}^{m\times m}$，成立：

$$\|UAV\|_E = \|A\|_E$$

接下来证明：

$$\|UAV\|_E^2 = \operatorname{tr}((UAV)^*(UAV)) = \operatorname{tr}(V^*(A^*A)V)$$

设 $V = (v_{ij})$，$(A^*A) = (a_{ij})$，$V^*(A^*A)V = (c_{ij})$。那么

$$c_{ii} = \sum_{j=1}^{m} \bar{v}_{ji}\sum_{k=1}^{m} a_{jk}v_{ki} = \sum_{j=1}^{m}\sum_{k=1}^{m} a_{jk}\bar{v}_{ji}v_{ki}$$

$$\sum_{i=1}^{m} c_{ii} = \sum_{j=1}^{m}\sum_{k=1}^{m} a_{jk}\sum_{i=1}^{m}\bar{v}_{ji}v_{ki} = \sum_{j=1}^{m}\sum_{k=1}^{m} a_{jk}\langle v_k^T, v_j^T\rangle = \sum_{j=1}^{m} a_{jj}$$

因此

$$\operatorname{tr}(V^*(A^*A)V) = \operatorname{tr}(A^*A) = \|A\|_E^2$$

第 6 章 矩阵特征值的扰动（续）：谱距离和 Hermite 矩阵的谱

6.1 谱距离

理论课的主要目的是估计矩阵的谱之间的距离：扰动和未扰动情况下

定义 6.1 (矩阵 Hausdorff 谱距离/Хаусдорфово спектральное расстояние)

设 $A,B\in\mathbb{C}^{n\times n}$，且 $\{\lambda_i\},\{\mu_j\}$ 分别是矩阵 A,B 的谱，即其所有特征值构成的集合

矩阵 A 和 B 之间的 Hausdorff 谱距离：

$$\operatorname{hd}(A,B)=\max\left\{\max_i\min_j|\lambda_i-\mu_j|,\max_j\min_i|\lambda_i-\mu_j|\right\}$$

♣

定义 6.2 (矩阵谱的 $p-$ 距离/спектральное p-расстояние)

设 $\lambda(A)=[\lambda_1,\ldots,\lambda_n]^T,\lambda(B)=[\mu_1,\ldots,\mu_n]^T$

矩阵 A 和 B 之间的谱 $p-$ 距离：

$$d_p(A,B)=\min_P\{\|\lambda(A)-P\lambda(B)\|_p:P-\text{置换矩阵}\}$$

♣

定理 6.1

满足估计:$\operatorname{hd}(A,B)\leqslant(\|A\|_2+\|B\|_2)^{1-\frac{1}{n}}\cdot\|A-B\|_2^{\frac{1}{n}}$

♡

证明 令酉矩阵 X 的列由向量 $x_1,\ldots,x_n$ 构成，并且 $Bx_1=\mu x_1$。若 $\lambda_1,\ldots,\lambda_n$ 是矩阵 A 的特征值，那么

$$\left(\min_i|\lambda_i-\mu|\right)^n\leqslant\prod_{i=1}^n|\lambda_i-\mu|=|\det((A-\mu I)X)|\leqslant\prod_{i=1}^n\|(A-\mu I)x_i\|_2\leqslant$$

$$\leqslant\|(A-B)x_1\|_2\cdot\prod_{i=2}^n\|(A-\mu I)x_i\|_2\leqslant\|A-B\|_2\cdot(\|A\|_2+\|B\|_2)^{n-1}$$

定理 6.2

谱 $\infty-$ 距离满足估计

$$d_\infty(A,B)\leqslant(2n-1)\operatorname{hd}(A,B)$$

♡

证明 令 λ_i 是矩阵 A 的特征值。考虑圆

$$D_i=\{z:|z-\lambda_i|\leqslant\operatorname{hd}(A,B)\},\quad D_i(\tau)=\{z:|z-\lambda_i|\leqslant\varepsilon(\tau)\}$$

其中

$$\varepsilon(\tau)=\left(\|A\|_2+\max_{t\in[0,1]}\|A+t(B-A)\|_2\right)^{1-\frac{1}{n}}\cdot\|\tau(B-A)\|_2^{\frac{1}{n}},\quad\tau\in[0,1]$$

根据定理 (6.1)，矩阵 $A+\tau(B-A)$ 的所有特征值都包含在圆 $D_i(\tau),i=1,\ldots,n$ 的并集中。类比 Gerschgorin 圆，若 m 个圆 $D_i(\tau)$ 与其余的圆孤立，那么其并集正好包含矩阵 $A+\tau(B-A)$ 的 m 个特征值

若 m 个圆 $D_i(1)$ 与其余的圆均孤立，那么对应的圆 D_i 的并集正好包含 m 个特征值 μ_i（因为 $hd(A,B)\leqslant\varepsilon(1)$）

假设矩阵 B 的圆和特征值 μ_i 的编号满足:

$$\mu_i \in \bigcup_{1\leqslant k\leqslant m} D_k,\ i=1,\ldots,m$$

那么 $(\forall i,j=1,\ldots,m):|\mu_i-\lambda_j|\leqslant(1+2(m-1))\,\mathrm{hd}(A,B)$。最坏的情况为 $m=n$

引理 6.1 (Birkhoff 定理)

任意的双随机 (двоякостохастическая) 矩阵 S 都可以表示为有限个置换矩阵 P_k 的凸组合 (выпуклая комбинация):

$$S=\sum_{k=1}^{m}\nu_k P_k,\quad \nu_1+\ldots+\nu_m=1,\quad \nu_1,\ldots,\nu_m\geqslant 0$$

♡

定理 6.3

正规矩阵 A 和 B 满足估计

$$d_2(A,B)\leqslant\|A-B\|_E$$

♡

证明 矩阵 A,B 是正规矩阵 $\Rightarrow$ 存在酉矩阵 P 和 Q，使得

$$D_A=\mathrm{diag}\,(\lambda_i)=P^*AP,\quad D_B=\mathrm{diag}\,(\mu_i)=Q^*BQ$$

成立。令 $Z=P^*Q$, $Z^*Z=I$，根据 Euclid 范数的酉不变性质和引理5.2:

$$\|A-B\|_E^2=\|D_A-ZD_BZ^*\|_E^2=\|D_A\|_E^2+\|D_B\|_E^2-2\,\mathrm{Retr}\,(ZD_BZ^*D_A^*)$$

为了得到估计的下界，记

$$\gamma=2\,\mathrm{Retr}\,(ZD_BZ^*D_A^*)=\sum_{i=1}^{n}\sum_{j=1}^{n}s_{ij}\alpha_{ij},\quad s_{ij}=|z_{ij}|^2,\alpha_{ij}=2\,\mathrm{Re}\,\lambda_i^*\mu_j$$

对于固定的 α_{ij}，泛函 $\gamma=\gamma(S)$ 在矩阵 $S=(s_{ij})$ 的集合上是线性的。重点关心的是其在有非负元素并且所有的行和、列和都等于 1（根据矩阵 P 和 Q 的酉性的推论）的矩阵 S 集合上的最大值。这样的矩阵称为双随机矩阵 **(двоякостохастическая матрица)**

根据 Birkhoff 定理6.1，$\gamma(S)$ 在其中一个置换矩阵上达到最大值:

$$\gamma(S)\leqslant\max_{k=1,\ldots,m}\gamma\,(P_k)=\gamma(\Pi),\quad \Pi\text{是一个置换矩阵}$$

然后

$$\|A-B\|_E^2=\|D_A\|_E^2+\|D_B\|_E^2-2\,\mathrm{Retr}\,(ZD_BZ^*D_A^*)\geqslant$$
$$\geqslant\|D_A\|_E^2+\|\Pi D_B\Pi^*\|_E^2-2\,\mathrm{Retr}\,(\Pi D_B\Pi^*D_A^*)=\|D_A-\Pi D_B\Pi^*\|_E^2\geqslant d_2^2(A,B)$$

此处考虑到矩阵 Π 是酉矩阵，借助引理5.2完成证明

6.2 谱距离的辅助推论

引理 6.2

假设矩阵 M 和 N 是正规矩阵，Ω 为对角矩阵，那么

$$\mathrm{tr}\,((M\Omega-\Omega N)(M-N)^*+(M-N)(M\Omega-\Omega N)^*)=2\,\mathrm{tr}\,(\Omega\,((M-N)(M-N)^*))$$

♡

证明 令 $M=(m_{ij}),\Omega=\mathrm{diag}\,(w_1,\ldots,w_n),M\Omega M^*=(a_{ij}),\Omega MM^*=(b_{ij})$。由矩阵 M 的正规性，可以得到: $MM^*=M^*M\Rightarrow$

$$\sum_{i=1}^{n}m_{ki}\bar{m}_{ji}=\sum_{i=1}^{n}\bar{m}_{ik}m_{ij}\Rightarrow\sum_{i=1}^{n}|m_{ki}|^2=\sum_{i=1}^{n}|m_{ik}|^2$$

$$a_{ij}=\sum_{k=1}^{n} m_{ik}w_k\bar{m}_{jk},\Rightarrow \operatorname{tr} M\Omega M^*=\sum_{i=1}^{n}\sum_{k=1}^{n} w_k\left|m_{ik}\right|^2$$

$$b_{ij}=\sum_{k=1}^{n} w_i m_{ik}\bar{m}_{jk}\Rightarrow \operatorname{tr}\Omega MM^*=\sum_{i=1}^{n}\sum_{k=1}^{n} w_i\left|m_{ik}\right|^2=\sum_{i=1}^{n}\sum_{k=1}^{n} w_k\left|m_{ki}\right|^2=\sum_{i=1}^{n}\sum_{k=1}^{n} w_k\left|m_{ik}\right|^2$$

表明 $\operatorname{tr} M\Omega M^*=\operatorname{tr}\Omega MM^*$。同理 $\operatorname{tr} MM^*\Omega=\operatorname{tr}\Omega MM^*=\operatorname{tr} M\Omega M^*$。

对于正规矩阵 N 和正规矩阵 $M-N$，可以证明类似的关系。注意 $(\forall A):\operatorname{tr}(A\Omega)=\operatorname{tr}(\Omega A)$（由于矩阵 $Omega$ 是对角矩阵）。现考虑下述基本关系：

$$\begin{aligned}
&\operatorname{tr}\left((M\Omega-\Omega N)(M-N)^*+(M-N)(M\Omega-\Omega N)^*\right)=\\
&=\operatorname{tr}\left(M\Omega M^*-\Omega NM^*-M\Omega N^*+\Omega NN^*+M\Omega M^*-MN^*\Omega-N\Omega M^*+NN^*\Omega\right)=\\
&=\operatorname{tr}\left((M-N)\Omega(M-N)^*-\Omega NM^*+M\Omega M^*-MN^*\Omega+NN^*\Omega\right)=\\
&=\operatorname{tr}\left((M-N)\Omega(M-N)^*+\Omega\left(-NM^*+MM^*-MN^*+NN^*\right)\right)=\\
&=\operatorname{tr}\left((M-N)\Omega(M-N)^*+\Omega(M-N)(M-N)^*\right)=2\operatorname{tr}\left(\Omega(M-N)(M-N)^*\right)
\end{aligned}$$

6.3 谱距离（续）

现在考虑将定理6.3推广到任意可对角化但不一定是正规矩阵的情况

定理 6.4

假设矩阵 A,B 可对角化，矩阵 P,Q 为对应的特征向量矩阵，即

$$A=PD_AP^{-1},\quad B=QD_BQ^{-1},\quad D_A,D_B-\text{对角矩阵}$$

然后有

$$d_2(A,B)\leqslant \operatorname{cond}_2(P)\cdot\operatorname{cond}_2(Q)\cdot\|A-B\|_E$$

$$\operatorname{cond}_2(P)=\left\|P^{-1}\right\|_2\cdot\|P\|_2,\quad \operatorname{cond}_2(Q)=\left\|Q^{-1}\right\|_2\cdot\|Q\|_2$$

♡

证明

$$\|A-B\|_E=\left\|PD_AP^{-1}-QD_BQ^{-1}\right\|_E=\left\|P\left(D_A\left(P^{-1}Q\right)-\left(P^{-1}Q\right)D_B\right)Q^{-1}\right\|_E\geqslant$$

$$\geqslant\frac{1}{\|P^{-1}\|_2\|Q\|_2}\|D_AZ-ZD_B\|_E,\quad Z=P^{-1}Q,1=\|I\|_2=\left\|PP^{-1}\right\|_2\leqslant\|P\|_2\cdot\left\|P^{-1}\right\|_2$$

考虑矩阵的奇异值分解 $Z=V\Sigma U^*$，那么有

$$\|D_AZ-ZD_B\|_E=\|V\left(\left(V^*D_AV\right)\Sigma-\Sigma\left(U^*D_BU\right)\right)U^*\|_E=\|M\Sigma-\Sigma N\|_E$$

其中矩阵 $M=V^*D_AV$ 和 $N=U^*D_BU$ 是正规矩阵。根据定理6.3，可以推出：

$$d_2(A,B)=d_2\left(D_A,D_B\right)=d_2(M,N)\leqslant\|M-N\|_E$$

设 $\sigma_{\min}-$ 矩阵Z 的最小特征值，那么

$$\sigma_{\min}=\frac{1}{\|Z^{-1}\|_2}\geqslant\frac{1}{\|Q^{-1}\|_2\|P\|_2}$$

接下来证明

$$\|M\Sigma-\Sigma N\|_E\geqslant\sigma_{\min}\|M-N\|_E$$

令 $\Omega = \Sigma - \sigma_{\min} I \geqslant 0$，然后有

$$\|M\Sigma - \Sigma N\|_E^2 = \|(M\Omega - \Omega N) + \sigma_{\min}(M - N)\|_E^2 = \{\text{引理 (5.2)}\} =$$
$$= \|M\Omega - \Omega N\|_E^2 + \sigma_{\min}^2 \|M - N\|_E^2 + \sigma_{\min} \operatorname{tr}\left((M\Omega - \Omega N)(M - N)^* + (M - N)(M\Omega - \Omega N)^*\right)$$
$$\sigma_{\min} \operatorname{tr}\left((M\Omega - \Omega N)(M - N)^* + (M - N)(M\Omega - \Omega N)^*\right) =$$
$$= \{\text{引理6.2}\} = 2\sigma_{\min} \operatorname{tr}\Omega\left((M - N)(M - N)^*\right) \geqslant 0$$

结合已得到的不等式，可得

$$d_2(A, B) \leqslant \|M - N\|_E \leqslant \tfrac{1}{\sigma_{\min}} \|M\Sigma - \Sigma N\|_E \leqslant \left\|P^{-1}\right\|_2 \cdot \|P\|_2 \left\|Q^{-1}\right\|_2 \cdot \|Q\|_2 \|A - B\|_E =$$
$$= \operatorname{cond}_2(P) \cdot \operatorname{cond}_2(Q) \cdot \|A - B\|_E$$

6.4 Hermite 矩阵的谱

定理 6.5 (Courant-Fischer 定理)

令 $\lambda_1 \geqslant \ldots \geqslant \lambda_n$ 是 Hermite 矩阵 $A \in \mathbb{C}^{n\times n}$ $(A = A^*)$ 的特征值。那么成立

$$\lambda_k = \max_{L:\dim L = k} \min_{x\in L, x\neq\theta} \frac{x^*Ax}{x^*x}, \quad \lambda_k = \min_{L:\dim L = n-k+1} \max_{x\in L, x\neq\theta} \frac{x^*Ax}{x^*x}$$

♡

定理 6.6

设 A 是一个 n 级 Hermite 矩阵，而 B 是 A 的 $n-1$ 级顺序（主）子矩阵。那么矩阵 A 的特征值 $\lambda_1 \geqslant \ldots \geqslant \lambda_n$ 和矩阵 B 的特征值 $\mu_1 \geqslant \ldots \geqslant \mu_{n-1}$ 满足划分关系 (соотношение разделения)：

$$\lambda_k \geqslant \mu_k \geqslant \lambda_{k+1}, \quad k = 1, \ldots, n-1$$

♡

证明 令 M 是由向量 $x \in \mathbb{C}^n$ 构成的线性子空间，x 形如

$$x = \begin{bmatrix} \tilde{x} \\ 0 \end{bmatrix}, \tilde{x} \in \mathbb{C}^{n-1} \Rightarrow \frac{x^*Ax}{x^*x} = \frac{\tilde{x}^*B\tilde{x}}{\tilde{x}^*\tilde{x}}$$

由 Courant-Fischer 定理6.5，可得

$$l\mu_k = \max_{\tilde{L}:\dim\tilde{L}=k} \min_{\tilde{x}\in\tilde{L},\tilde{x}\neq\theta} \frac{\tilde{x}^*B\tilde{x}}{\tilde{x}^*\tilde{x}} = \max_{L:\dim L=k, L\subseteq M} \min_{x\in L, x\neq\theta} \frac{x^*Ax}{x^*x} \leqslant \max_{L:\dim L=k} \min_{x\in L, x\neq\theta} \frac{x^*Ax}{x^*x} = \lambda_k$$

$$\mu_k = \max_{\tilde{L}:\dim\tilde{L}=(n-1)-k+1} \min_{\tilde{x}\in\tilde{L},\tilde{x}\neq\theta} \frac{\tilde{x}^*B\tilde{x}}{\tilde{x}^*\tilde{x}} = \min_{L:\dim L=(n-1)-k+1, L\subseteq M} \max_{x\in L, x\neq\theta} \frac{x^*Ax}{x^*x} \geqslant$$

$$\geqslant \min_{L:\dim L=(n-1)-k+1} \max_{x\in L, x\neq\theta} \frac{x^*Ax}{x^*x} = \lambda_{k+1}$$

6.5 Hermite 矩阵的谱：辅助推论

引理 6.3

假设 $A \in \mathbb{C}^{n\times n}$ 是一个 Hermite 矩阵，$p \in \mathbb{C}^n$。那么有

$$\max_{\dim L = k} \min_{x\in L, x\perp p, x\neq\theta} \frac{x^*Ax}{x^*x} \leqslant \max_{\dim L = k-1} \min_{x\in L, x\neq\theta} \frac{x^*Ax}{x^*x}$$

♡

证明 空间 $L : \dim L = k$ 的所有子空间都可以分成两部分

1) $p \perp L$

2) p 不正交于 L

在第一个情况下，对 $\forall x \in L \Rightarrow x \perp p \Rightarrow$

$$\min_{x\in L, x\perp p, x\neq\theta} \frac{x^*Ax}{x^*x} = \min_{x\in L, x\neq\theta} \frac{x^*Ax}{x^*x} \leqslant \min_{x\in \tilde{L}, x\neq\theta} \frac{x^*Ax}{x^*x}, \forall \tilde{L} \subset L, \dim \tilde{L} = k-1$$

在第二个情况下，考虑 $\tilde{L} = \left\{x \in L : x^T p = 0\right\}, \dim \tilde{L} = k-1 \Rightarrow$

$$\min_{x\in L, x\perp p, x\neq\theta} \frac{x^*Ax}{x^*x} = \min_{x\in \tilde{L}, x\neq\theta} \frac{x^*Ax}{x^*x}$$

现在结合已获得的不等式并在 L 上取最大值：

$$\max_{\dim L=k} \min_{x\in L, x\perp p, x\neq\theta} \frac{x^*Ax}{x^*x} \leqslant \max_{\dim L=k} \min_{x\in \tilde{L}, x\neq\theta} \frac{x^*Ax}{x^*x} \leqslant \max_{\dim L^*=k-1} \min_{x\in L^*, x\neq\theta} \frac{x^*Ax}{x^*x}$$

该公式表明 $\tilde{L}$ 依赖于 L，并且 $\dim \tilde{L} = k-1$，但在最后一个最大值中，已考虑所有可能的 $k-1$ 维的空间 L^*，不一定和 L 有联系

6.6 Hermite 矩阵的谱（续）

定理 6.7

令矩阵 A 和 B 都是 n 级 Hermite 矩阵，并且其特征值分别为 $\lambda_1 \geqslant \ldots \geqslant \lambda_n$ 和 $\mu_1 \geqslant \ldots \geqslant \mu_n$。若

$$B = A + \varepsilon pp^*, \varepsilon > 0, p \in \mathbb{C}^n, \|p\|_2 = 1$$

那么满足如下的划分关系：

$$\mu_1 \geqslant \lambda_1 \geqslant \mu_2 \geqslant \ldots \geqslant \lambda_{n-1} \geqslant \mu_n \geqslant \lambda_n$$

在这种情况下，有

$$\mu_k = \lambda_k + t_k\varepsilon,\ t_k \geqslant 0,\ k = 1, \ldots, n; \quad t_1 + \ldots + t_n = 1$$

♡

证明 由 Courant-Fischer 定理6.5，可得：

$$\lambda_k = \max_{\dim L=k} \min_{x\in L, x\neq\theta} \frac{x^*Ax}{x^*x} \leqslant \max_{\dim L=k} \min_{x\in L, x\neq\theta} \frac{x^*Bx}{x^*x} = \mu_k$$

当 $k \geqslant 2$ 时，根据引理6.3，有

$$\mu_k \leqslant \max_{\dim L=k} \min_{x\in L, x\perp p, x\neq\theta} \frac{x^*Bx}{x^*x} = \max_{\dim L=k} \min_{x\in L, x\perp p, x\neq\theta} \frac{x^*Ax}{x^*x} \leqslant \max_{\dim L=k-1} \min_{x\in L, x\neq\theta} \frac{x^*Ax}{x^*x} = \lambda_{k-1}$$

由 Courant-Fischer 定理6.5，可以推出 $(\forall k) : \mu_k \leqslant \lambda_k + \varepsilon$。因此可以找到满足 $\mu_k = \lambda_k + t_k\varepsilon$ 成立的 $t_k \in [0, 1]$。将这些等式对所有的 k 求和：

$$\operatorname{tr} B = \operatorname{tr} A + \varepsilon\left(t_1 + \ldots + t_n\right)$$

但又有 $\operatorname{tr} B = \operatorname{tr} A + \varepsilon \Rightarrow t_1 + \ldots + t_n = 1$

第 7 章 机器算数的特点 (особенности машинной арифметики)

7.1 机器数

用计算机解决数学问题时，只能使用一组有限的数

定义 7.1 (机器数/машинное число)

机器数形如：

$$a = \pm\left(\frac{d_1}{p} + \frac{d_2}{p^2} + \ldots + \frac{d_t}{p^t}\right)\cdot p^{\alpha}$$

其中 $p, \alpha, d_1, \ldots, d_t$ 是整数

♣

注 • 数 $p > 0$ 称为**算数基 (основание арифметики)**

- 数 $\left(\frac{d_1}{p} + \frac{d_2}{p^2} + \ldots + \frac{d_t}{p^t}\right)$ 称为**尾数 (мантисса)**
- 数 α 称为机器数 a **阶数 (порядок)**
- 数 $d_i \in \{0, 1, \ldots, p-1\}$ 称为数值的**数位 (разряд)**
- 数 t 称为**尾数长度 (длинная мантисса)**

通常假设 $d_1 \neq 0$ （所谓的**规范系 (нормализованная система)**）。除此之外，整数 L 和 U 给出 α 的上下界：$L \leqslant \alpha \leqslant U$。$a = 0$ 是一个特殊的机器数

允许的机器数集由参数 p, t, L, U 决定

7.2 机器算数公理

当将数字输入计算机并使用机器数执行操作时，通常会发生数字的舍入操作。舍入 (округление) 是某个由实数集到机器数集的映射

如果 x 是一个实数，$\mathfrak{fl}(x)$ 是该实数舍入映射的值，则有公理：$\mathfrak{fl}(x) = x(1+\varepsilon)$，其中当 $\mathfrak{fl}(x) \neq 0 : |\varepsilon| \leqslant \eta$。再假设 η 为 $|\varepsilon|$ 取值的上确界。$\varepsilon-$ **相对舍入误差 (относительная погрешность округления)**

如果在对实数进行四舍五入时，选择最接近其的机器数，那么 $\eta = \frac{1}{2}p^{1-t}$。例如：

设 a 数字的真值，$\tilde{a} = \mathfrak{fl}(a)$ 为其对应得机器数，那么有：

$$a = \pm\left(\frac{d_1}{p} + \frac{d_2}{p^2} + \ldots + \frac{d_t}{p^t} + \frac{d_{t+1}}{p^{t+1}} + \ldots\right)\cdot p^{\alpha}$$

若考察规范系，那么尾数的模长总是 $\geqslant \frac{1}{p}$，因此 $|a| \geqslant p^{\alpha-1}$。对于通过寻找最近元素的方法对尾数进行四舍五入后的绝对误差，满足估计：

$$|a-\tilde{a}| \leqslant \frac{1}{2}p^{\alpha}\frac{1}{p^{t+1}}\left(1 + \frac{1}{p} + \frac{1}{p^2} + \ldots\right) \leqslant \frac{1}{2}p^{\alpha-t} \text{即} |a-\tilde{a}| \leqslant \frac{1}{2}p^{\alpha}\left(\frac{p-1}{p^{t+1}} + \frac{p-1}{p^{t+2}} + \ldots\right) = \frac{1}{2}p^{\alpha}\frac{1}{p^t} = \frac{1}{2}p^{\alpha-t} \Rightarrow$$

$$\Rightarrow \frac{|a-\tilde{a}|}{|a|} \leqslant \frac{1}{2}p^{1-t}$$

同理，通过删除数位 $\eta = p^{1-t}$ 舍入

机器数 a 和 b 的运算 $(*)$ 的结果用 $\mathfrak{fl}(a*b)$ 表示。因此，假设如果有 $\mathfrak{f}(a*b) \neq 0$，则 $\mathfrak{f}(a*b) = (a*b)(1+\varepsilon), |\varepsilon| \leqslant \eta$。该关系式是研究算法中舍入误差的影响的主要公理。

推论 7.1

只有当运算结果不是机器数零时，相对误差 ε 才会很小

♡

有时，舍入运算是通过舍弃“不必要的、额外的”数位实现的。在这种情况下，等式 $x = a * b$ 一般不会推出 $\mathrm{fl}(a) * \mathrm{fl}(b) = \mathrm{fl}(x)$。例如，设 $p = 2, t = 2, a = 0.11, b = 0.0001$ 且 $x = a - b = 0.1011$。在给定的情况下 $\mathrm{fl}(x) = 0.10$, 那么 $\mathrm{fl}(a) - \mathrm{fl}(b) = 0.11$ （数字运算操作在 2 位的“加法器”中执行）

使用浮点乘法，当两个非零数的零乘积小于最小非零元素 ε 的绝对值时，可能会出现机器数 0。其中 ε 称为**机器数 ε(машинное эпсилон)**。

机器数 ε 是最小浮点数 ε，满足：

$$1 + \varepsilon > 1$$

(是 1 与可表示为机器编号的下一个最接近的数字之间的差)

例题 7.1 В стандарте IEEE для чисел с основанием арифметики $p = 2$, с одинарной точностью (single) $\varepsilon = 2^{-23}$; для чисел с двойной точностью (double) $\varepsilon = 2^{-52}$.

7.3 数字加减误差

如果准确数的和等于：

$$S = a_1 + a_2 + \ldots + a_n$$

近似数的和为

$$\tilde{S} = a_1 + \Delta a_1 + a_2 + \Delta a_2 + \ldots + a_n + \Delta a_n$$

其中 $(i = 1, \ldots, n) : \Delta a_i$ 是使用机器浮点数表示的数字的绝对误差。那么绝对误差的和等于

$$\Delta S = \Delta a_1 + \ldots + \Delta a_n$$

和的相对误差为

$$\delta S = \frac{\Delta S}{S} = \frac{a_1}{S}\frac{\Delta a_1}{a_1} + \frac{a_2}{S}\frac{\Delta a_2}{a_2} + \ldots + \frac{a_n}{S}\frac{\Delta a_n}{a_n} = \frac{a_1\delta a_1 + a_2\delta a_2 + \ldots + a_n\delta a_n}{S}$$

其中 $\delta a_i = \frac{\Delta a_i}{a_i} -$ 表示数字的相对误差

相同符号的几个数和的相对误差位于被加数的相对误差最大值、最小值之间：

$$\min_i \delta a_i \leqslant \delta S \leqslant \max_i \delta a_i$$

当加上异号的数字或减去同号的数字时，如果数字彼此接近，相对误差可能会非常大。当给定的 δa_i 很小时，数值 S 可能也会很小。因此，计算算法必须以避免减去接近数的方式构建

值得注意的是，计算误差也取决于计算的顺序

例题 7.2

$$S = x_1 + x_2 + x_3$$

从左到右求和

$$\tilde{S}_1 = (x_1 + x_2)(1 + \varepsilon) \Rightarrow \tilde{S} = \left(\tilde{S}_1 + x_3\right)(1 + \varepsilon) = (x_1 + x_2)(1 + \varepsilon)^2 + x_3(1 + \varepsilon)$$

在另一种操作顺序下，误差会有所不同：

$$\tilde{S}_1 = (x_3 + x_2)(1 + \varepsilon) \Rightarrow \tilde{S} = (x_3 + x_2)(1 + \varepsilon)^2 + x_1(1 + \varepsilon)$$

执行某些因舍入误差而失真的算法的结果与进行相同算法（无舍入）的结果一致，但其初始数据不准确。也就是说，可以借助反向分析：减少舍入误差对原始数据的扰动的影响。那么在求和的第一种情况中：

$$\tilde{S} = \tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3, \quad \tilde{x}_1 = x_1(1 + \varepsilon)^2, \tilde{x}_2 = x_2(1 + \varepsilon)^2, \tilde{x}_3 = x_3(1 + \varepsilon)$$

7.4 数字乘除误差

当乘或除近似数时，相对误差相加减

$$S = a_1 \cdot a_2$$

$$\tilde{S} = a_1 \cdot a_2 \left(1+\varepsilon_1\right)\left(1+\varepsilon_2\right) \approx a_1 a_2 \left(1+\varepsilon_1+\varepsilon_2\right)$$

(精确到对于 ε 的二阶小量)

若 $S = \frac{a_1}{a_2}$，则

$$\Delta S = \frac{a_1\left(1+\varepsilon_1\right)}{a_2\left(1+\varepsilon_2\right)} - \frac{a_1}{a_2} = \frac{a_1\left(\varepsilon_1-\varepsilon_2\right)}{a_2\left(1+\varepsilon_2\right)} \approx \frac{a_1}{a_2}\left(\varepsilon_1-\varepsilon_2\right) = \frac{a_1}{a_2}\left(1+\left(\varepsilon_1-\varepsilon_2-1\right)\right) \Rightarrow$$

$$\Rightarrow \tilde{S} = S + \Delta S \approx \frac{a_1}{a_2}\left(1+\left(\varepsilon_1-\varepsilon_2\right)\right)$$

7.5 例子：标量积的舍入误差

假设计算中的所有相对误差都是相同的，并记为 ε。令 $\tilde{\alpha}$ 为机器上得到的的标量积运算：

$$\left(x=[x_1,\ldots,x_n]^T, y=[y_1,\ldots,y_n]^T, x_i,y_i\in\mathbb{R}\right): \alpha = x^T y$$

假设使用以下算法：

1) $\alpha = 0$
2) $\alpha = \alpha + x_1 y_1$
3) $\alpha = \alpha + x_2 y_2$

$\ldots$

$n+1)$ $\alpha = \alpha + x_n y_n$

那么

$$\tilde{\alpha} = \sum_{i=1}^{n} x_i y_i (1+\varepsilon)^{n+1-i}$$

由此获得公式可以从两个不同的方面来解释。使用直接分析的方法可以估计准确值和实际计算值之间的接近程度（这里 $|x|, |y|$ 表示分别由向量 x, y 的分量的模长（绝对值）组成的向量）：

$$|\tilde{\alpha} - \alpha| \leqslant n\eta \left|x^T\right| |y| + \mathcal{O}\left(\eta^2\right)$$

反向分析建议将实际计算结果表示为带有扰动数据的精确计算结果，并对这种“等价的”扰动进行估计

$$\tilde{a} = \tilde{x}^T \tilde{y}, \quad |\tilde{x} - x| \leqslant \frac{1}{2} n\eta |x| + \mathcal{O}\left(\eta^2\right), \quad |\tilde{y} - y| \leqslant \frac{1}{2} n\eta |y| + \mathcal{O}\left(\eta^2\right)$$

(在 x 和 y 之间可能还存在不同的扰动分布方案)

7.6 “坏”操作

通常，求解问题时产生的大的误差并不是大量运算操作的结果（因此产生大量错误），而是某一个“坏”操作造成的结果。

例题 7.3 令 $a \approx \tilde{a}, b \approx \tilde{b}$，那么

$$\mathrm{f}(\tilde{a} - \tilde{b}) = (\tilde{a} - \tilde{b})(1+\varepsilon) = (a-b)(1+\varepsilon+\delta), \quad \delta = \left(\frac{(\tilde{a}-a)-(\tilde{b}-b)}{a-b}\right)(1+\varepsilon)$$

如果数 a 和数 b 彼此十分接近，那么 δ 的数值可能会非常大

7.7 例子

在使用某种程序计算有 2×2 子矩阵块的分块三角阵的特征向量时，发现恰好一个特征向量的误差超过其他向量的误差 3 个数量级。在这种情况下，主要问题归结为具有 2 个方程、2 个未知元的其次方程组

$$a_1x_1+a_2x_2=0,\quad b_1x_1+b_2x_2=0,\quad \|x\|_\infty=1$$

的解。设向量 $a=[a_1,a_2]^T$ 和 $b=[b_1,b_2]^T$ 非零且近似共线:$|a_1b_2-a_2b_1|=\delta$，其中 δ 取很小的数值，可以取

$$x_1=-\frac{a_2}{\|a\|_\infty},x_2=\frac{a_1}{\|a\|_\infty}$$

然后得到误差

$$r_1=|a_1\tilde{x}_1+a_2\tilde{x}_2|\leqslant 2\eta\|a\|_\infty,\quad r_2=|b_1\tilde{x}_1+b_2\tilde{x}_2|\leqslant\frac{\delta}{\|a\|_\infty}+2\eta\|b\|_\infty$$

也可以用另一种方式构建:

$$x_1=-\frac{b_2}{\|b\|_\infty},\quad x_2=\frac{b_1}{\|b\|_\infty}$$

然后有

$$r_1\leqslant\frac{\delta}{\|b\|_\infty}+2\eta\|a\|_\infty,\quad r_2\leqslant 2\eta\|b\|_\infty$$

由此产生的残差估计在 $\frac{\delta}{\|a\|_\infty}$ 和 $\frac{\delta}{\|b\|_\infty}$ 两方面有所不同。如果 $\|b\|_\infty$ 的值超过 $\|a\|_\infty$ 三个数量级，那么一般来说，第一种方法中的误差会大于三个数量级

因此，若 $\|a\|_\infty\geqslant\|b\|_\infty$，那么需要使用第一种方法；否则需要使用第二种方法

7.8 运算先后顺序

考察一个与积分计算相关的不稳定算法的例子

$$a_n=\int_0^1 x^ne^{x-1}dx,\quad n=1,2,\dots$$

分部积分:

$$(\forall n\geqslant 1):\int_0^1 x^ne^{x-1}dx=1-n\int_0^1 x^{n-1}e^{x-1}dx\Rightarrow a_n=1-na_{n-1}$$

此外 $a_0=1-e^{-1},a_1=e^{-1}$

如果“自下而上”求解得到的递推关系，那么在 a_n 的计算中即使 a_1 的一个不大的误差，乘以 $n!$ 也会变得非常大

为了解决误差累积的问题，可以“自上而下”解决递归关系

$$a_{k-1}=k^{-1}(1-a_k),\quad k=N,N-1,\dots,3,n+1$$

为了找到初始条件，考虑:

$$(k\to\infty):0<a_k<\int_0^1 x^kdx=\frac{1}{k+1}\to 0\Rightarrow x_N\approx 0$$

在新的方案中，误差按比值 $\frac{1}{N(N-1)\cdots(n+1)}=\frac{n!}{N!}$ 减小

7.9 求解三角形方程组

考虑具有下三角矩阵 $L=(l_{ij})$ 的线性方程组 (СЛАУ): $Lx=b$。可以使用**直接替换法 (метод прямой подстановки)** 求解

$$(i=1,\dots,n):x_i=\left(b_i-\sum_{j=1}^{i-1}l_{ij}x_j\right)/l_{ii}$$

若 $\tilde{x}_i$ 是一个实际的计算值，那么

$$\tilde{x}_i = \left(\left(b_i - \sum_{j=1}^{i-1} l_{ij} \tilde{x}_j (1+\varepsilon)^{i-j} \right) / l_{ii} \right) (1+\varepsilon)^2$$

令

$$\tilde{l}_{ij} = \begin{cases} l_{ij}(1+\varepsilon)^{i-j} & ,i > j \\ l_{ij}/(1+\varepsilon)^2 & ,i = j \\ 0 & ,i < j \end{cases} \Rightarrow \tilde{x}_i = \left(b_i - \sum_{j=1}^{i-1} \tilde{l}_{ij} \tilde{x}_j \right) / \tilde{l}_{ii},$$

即 $\tilde{x}_i$ 是具有相同的右侧（非齐次）部分和扰动矩阵 $\tilde{L} = \left(\tilde{l}_{ij} \right)$ 的方程组准确解的一个分量，$\tilde{L}$ 和 L 的接近程度可以估计为

$$\left| \tilde{l}_{ij} - l_{ij} \right| \leqslant \begin{cases} |l_{ij}| \, (i-j)\eta + \mathcal{O}\left(\eta^2\right) & ,i > j \\ |l_{ij}| \, 2\eta + \mathcal{O}\left(\eta^2\right) & ,i = j \\ 0 & ,i < j \end{cases}$$

由此证明下述定理

定理 7.1

对于直接替换法，方程组 $Lx = b$ 的实际计算解 $\tilde{x}$ 也满足扰动方程组 $\tilde{L}\tilde{x} = b$，其中 $\tilde{L}$ 是使 $|\tilde{L} - L| \leqslant n\eta|L| + \mathcal{O}\left(\eta^2\right)$ 成立的三角阵

♡

对有上三角阵的线性方程组的逆代换法求解可以得到相同的结果

第 8 章 线性方程组的直接解法和矩阵的 LU 分解

8.1 解决问题的直接方法

数学问题的求解可以表示为一个基于某些初等运算集制定的算法的形式。如果仅需要有限次数量的初等运算来解决问题，那么这种方法称为**解决问题的直接方法 (прямой метод решения задач)**。

例题 8.1 如果将算术运算 $(+,-,\cdot,\div)$ 视为初等运算，那么方程 $x^2=2$ 无法使用有限次初等运算求解。若添加“数的平方根”运算，那么就会有解决问题的直接方法

注 使用算术运算和平方根运算，（由 Galois 理论：五次方程不存在根式解）不可能构造直接方法来找出任意一个次数 $\geqslant 5$ 的多项式的根 $\Rightarrow$ 不存在计算任意阶数 $\geqslant 5$ 的矩阵的特征值的直接方法。

线性方程组 $Ax=b$（在一般情况下，对没有任何特征 (специфика) 的密集矩阵 (плотная матрица)）求解存在直接方法。接下来使用算术运算，有时候还有平方根运算作为初等运算。线性方程组求解最简单的直接方法是 **Gauss 消元法 (метод Гаусса)** 和**旋转和反射变换法 (метод, использующий преобразования вращения или отражения)**。这些方法，与系数矩阵的 LU **分解**或 QR **分解**的获得有关

8.2 LU 分解

定义 8.1 (严格正则矩阵/строго регулярная матрица)

如果矩阵 A 的所有顺序主子矩阵（包括 A 本身）都是非退化的（即顺序主子式非 0），那么矩阵 A 称为严格正则矩阵

$$A=\begin{pmatrix} a_{11} & a_{12} & \dots \\ a_{21} & a_{22} & \dots \\ \dots & \dots & \dots \end{pmatrix} \Rightarrow a_{11}\neq 0,\ \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0, \dots,\ \det(A)\neq 0$$

♣

定义 8.2 (LU 分解/LU-разложение)

矩阵 A 的 LU 分解是形如 $A=LU$ 的等式，其中

- 矩阵 L 是下三角阵，且对角线元素都是 1
- 矩阵 u 是非退化的上三角阵

♣

推论 8.1

有时，矩阵的 LU 分解以不同的方式定义：矩阵 L 是非退化的下三角阵，而矩阵 U 是对角线元素都为 1 的上三角阵

$$1)\quad L=\begin{pmatrix} 1 & 0 & \dots & 0 \\ l_{21} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ l_{n1} & l_{n2} & \dots & 1 \end{pmatrix},\ U=\begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & u_{nn} \end{pmatrix}$$

$$2)\quad L=\begin{pmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{pmatrix},\ U=\begin{pmatrix} 1 & u_{12} & \dots & u_{1n} \\ 0 & 1 & \dots & u_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

♡

8.3 LU 分解的存在性

定理 8.1 (LU 分解存在的充要条件)

矩阵 A 具有 LU 分解的充分必要条件是：矩阵 A 是严格正则矩阵

♡

证明

必要性：若 $A=LU$，则矩阵 A 及其任何顺序主子矩阵 B 都可表示为两个非退化（下和上）三角阵乘积的形式 $\Rightarrow$ 矩阵 A,B 都是非退化矩阵 $\Rightarrow$ A 是严格正则矩阵

充分性：通过对矩阵 A 规模大小进行归纳完成证明。假设

$$A=\begin{bmatrix} a & c^T \\ b & D \end{bmatrix},\ a\neq 0,\ z=\frac{1}{a}b,\ A_1=D-\frac{1}{a}bc^T$$

$$\Rightarrow \begin{bmatrix} 1 & 0 \\ -z & I \end{bmatrix}\begin{bmatrix} a & c^T \\ b & D \end{bmatrix}=\begin{bmatrix} a & c^T \\ 0 & A_1 \end{bmatrix}$$

矩阵 A_1 是严格正则矩阵。那么根据归纳假设，A_1 有 LU 分解:$A_1=L_1U_1$。令

$$L=\begin{bmatrix} 1 & 0 \\ z & L_1 \end{bmatrix},\quad U=\begin{bmatrix} a & c^T \\ 0 & U_1 \end{bmatrix}$$

$$\Rightarrow LU=\begin{bmatrix} a & c^T \\ b & L_1U_1+\frac{1}{a}bc^T \end{bmatrix}=\begin{bmatrix} a & c^T \\ b & D \end{bmatrix}=A$$

8.4 LU 分解的唯一性

推论 8.2

矩阵 A 的所有顺序主子式都是正的 $\Leftrightarrow$ 在矩阵 A 的 LU 分解中，矩阵 U 有正的对角元

♡

定理 8.2 (LU 分解的唯一性)

LU 分解是唯一定义的。

♡

证明 若 $L_1U_1=L_2U_2$，那么假设

$$D=L_2^{-1}L_1=U_2U_1^{-1}$$

矩阵 L_2^{-1} 和 $L_2^{-1}L_1$ 都是下三角阵，矩阵 U_1^{-1} 和 $U_2U_1^{-1}$ 都是上三角阵。因此 D 同时为上三角阵和下三角阵 $\Rightarrow$ D 是对角阵。但是矩阵 $L_2^{-1}L_1$ 对角线元素都是 1$\Rightarrow$ $D=I\Rightarrow L_1=L_2,U_1=U_2$

8.5 LU 分解和 Gauss 消元法之间的联系

在代数课程中已经学习求解线性方程组 $Ax=b$ 的 Gauss 消元法，它主要包含两个主要步骤：

1. 正向过程 (прямой ход)：借助初等变换将线性方程组简化为具有上阶梯（三角）矩阵的线性方程组
2. 反向过程 (обратный ход)：求解阶梯型（三角）矩阵的线性方程组

令矩阵 A 是非退化的方阵。那么在矩阵术语中，Gauss 消元法可以如下进行描述

1. 正向过程：寻找对角元都是 1 的下三角阵，满足 $Ux=L^{-1}Ax=L^{-1}b$ 具有上三角阵的线性方程组
2. 反向过程：求解线性方程组 $Ux=L^{-1}b$，即找 $x=U^{-1}L^{-1}b$

那么有 $A=LU$

若已知矩阵 A 的 LU 分解，那么很容易计算得到矩阵 L^{-1} 和 U^{-1}。因此，可以通过

$$Ax = b, A = LU \Rightarrow x = U^{-1}L^{-1}b$$

求解具有不同右侧部分 b 的线性方程组

8.6 矩阵 LU 分解的舍入错误

定理 8.3

令矩阵 A 是 n 级可用机器算数表示的实严格正则矩阵，那么对于实际计算的矩阵 $\tilde{L}, \tilde{U}$ 成立不等式

$$|\tilde{L}\tilde{U} - A| \leqslant 3n\eta(|A| + |\tilde{L}| \cdot |\tilde{U}|) + \mathcal{O}\left(\eta^2\right) \tag{8.1}$$

♡

证明 假设具有如下形式的矩阵完全相等：

$$A = LU,\ A = \begin{bmatrix} a & c^T \\ b & D \end{bmatrix},\ L = \begin{bmatrix} 1 & 0 \\ z & L_1 \end{bmatrix},\ U = \begin{bmatrix} a & c^T \\ 0 & U_1 \end{bmatrix}$$

对于实际计算矩阵，有：

$$|\tilde{L}\tilde{U} - A| = \begin{bmatrix} 1 & 0 \\ \tilde{z} & \tilde{L}_1 \end{bmatrix} \begin{bmatrix} a & \tilde{c}^T \\ 0 & \tilde{U}_1 \end{bmatrix} - \begin{bmatrix} a & c^T \\ b & D \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ (\tilde{z} - z)a & \tilde{L}_1\tilde{U}_1 - H \end{bmatrix},\ H = D - \tilde{z}c^T$$

令 $\tilde{H} = \mathrm{fl}\left(D - \tilde{z}c^T\right)$。矩阵 $\tilde{L}_1$ 和 $\tilde{U}_1$ 构成矩阵 $\tilde{H}$ 实际计算 LU 分解。

接下来，通过对维数 n 归纳完成证明。当 $n = 1$ 不等式显然成立（等式左侧为 0）。由归纳假设，有

$$\left|\tilde{L}_1\tilde{U}_1 - \tilde{H}\right| \leqslant 3(n-1)\eta\left(|\tilde{H}| + \left|\tilde{L}_1\right| \cdot \left|\tilde{U}_1\right|\right) + \mathcal{O}\left(\eta^2\right)$$

由机器算数公理7.2，有

$$|\tilde{H} - H| \leqslant \left(|D| + |\tilde{z}| \cdot \left|c^T\right|\right)\eta \Rightarrow |\tilde{H}| \leqslant |H| + \mathcal{O}(\eta) \leqslant |D| + |\tilde{z}| \cdot \left|c^T\right| + \mathcal{O}(\eta)$$

因此

$$\left|\tilde{L}_1\tilde{U}_1 - H\right| \leqslant \left|\tilde{L}_1\tilde{U}_1 - \tilde{H}\right| + |\tilde{H} - H| \leqslant (3n-2)\eta|D| + (3n-1)\eta|\tilde{z}| \cdot \left|c^T\right| +$$

$$+(3n-3)\eta\left|\tilde{L}_1\right| \cdot \left|\tilde{U}_1\right| + \mathcal{O}\left(\eta^2\right) \leqslant 3n\eta\left(|D| + |\tilde{z}| \cdot \left|c^T\right| + \left|\tilde{L}_1\right| \cdot \left|\tilde{U}_1\right|\right) + \mathcal{O}\left(\eta^2\right)$$

因为 $az = b$，则 $|(\tilde{z} - z)a| \leqslant \eta|az| = \eta|b|$，因此不等式 (8.1) 得证

8.7 主元的选择

从所得的舍入误差估计可以看出，Gauss 消元法的“瓶颈”在右侧部分的项 $|\tilde{L}| \cdot |\tilde{U}|$。也就是说，由于三角因子中元素的增长，可能会导致较大的误差

例题 8.2 令 $n = 2, p$ 是算数基数 , t 是尾数长度。假设

$$A = \begin{bmatrix} p^{-t} & 1 \\ 1 & 1 \end{bmatrix} \Rightarrow \tilde{L} = \begin{bmatrix} 1 & 0 \\ p^t & \tilde{L}_1 \end{bmatrix}, \tilde{U} = \begin{bmatrix} p^{-t} & 1 \\ 0 & \tilde{U}_1 \end{bmatrix}$$

令 $\mathrm{fl}\left(1 - p^t\right) = -p^t$ （p^t 是一个非常大的数）。那么 $\tilde{L}_1 = 1,\ \tilde{U}_1 = -p^t$:

$$\tilde{L}\tilde{U} - A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

得到非常大的误差。原因是**主元 (ведущий член)**$a = p^{-t}$ 的值非常小

由于 $\tilde{L}_1\tilde{U}_1 = \mathrm{fl}\left(D - \frac{1}{a}bc^T\right)$，为了减小矩阵 $\tilde{L}_1$ 和 $\tilde{U}_1$ 并使误差更小，应选择尽可能大的主元 a。例如通过矩阵 A 的行变换交换主元

若对矩阵 A 进行额外的行变换(以扩大主元),结果得到**矩阵的 LUP 分解 (LUP-разложение матрицы)**:

$$PA = LU, \quad \text{其中} P \text{是置换矩阵}$$

8.8 矩阵的 LDL 分解和 Cholesky 分解

定理 8.4

若矩阵 $A \in \mathbb{C}^{n\times n}$ 严格正则且对称 $\left(A = A^T\right)$，则有 LDL^T 分解：$A = LDL^T$，其中 L 是对角元都是 1 的下三角阵；D 是非退化对角矩阵 ♡

证明 设 $A = LU, D = \operatorname{diag}(U)$。然后有 $A = LDD^{-1}U = A^T = \left(U^TD^{-1}\right)\left(DL^T\right)$。矩阵 U^TD^{-1} 是对角元为 1 的下三角矩阵。由 LU 分解的唯一性可以推出 $DL^T = U \Rightarrow A = LDL^T$

下面的定理可以类似地证明:

定理 8.5

若矩阵 $A \in \mathbb{C}^{n\times n}$ 严格正则且为 Hermite 矩阵 $(A = A^*)$，则有 LDL^* 分解：$A = LDL^*$，其中 L 是对角元都是 1 的下三角阵；D 是非退化对角矩阵 ♡

定义 8.3 (Cholesky 分解/разложение Холецкий)

假设矩阵 C 是一个主对角元都是正数的下三角矩阵。$A = CC^*$ 分解称为 Cholesky 分解 (разложение Холецкий) ♣

定理 8.6 (Cholesky 分解的充要条件)

矩阵 $A \in \mathbb{C}^{n\times n}$ 有 Cholesky 分解的充分必要条件是：矩阵 A 是顺序主子式都是正数的 Hermite 矩阵 ♡

证明 充分性的证明:

令矩阵 $A = LDL^T$(或 $A = LDL^*$),$D = \operatorname{diag}(d_1, \ldots, d_n)$。假定 $C = L\operatorname{diag}\left(\sqrt{d_1}, \ldots, \sqrt{d_n}\right)$，那么有 $A = CC^T$ (或 $A = CC^*$)

8.9 Cholesky 方法

设矩阵 A 是一个顺序主子式都是正数的实对称矩阵。构造其 Cholesky 分解

例题 8.3 令 $n = 3$

$$\begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} c_{11} & 0 & 0 \\ c_{21} & c_{22} & 0 \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \cdot \begin{bmatrix} c_{11} & c_{21} & c_{31} \\ 0 & c_{22} & c_{32} \\ 0 & 0 & c_{33} \end{bmatrix}$$

依次将第一个矩阵的元素乘以第二个矩阵的第 $1, 2, 3$ 列，那么有:

$$c_{11} = \sqrt{a_{11}}, \quad c_{21} = \frac{a_{21}}{c_{11}}, \quad c_{31} = \frac{a_{31}}{c_{11}}$$

$$c_{22} = \sqrt{a_{22} - c_{21}^2}, \quad c_{32} = \frac{a_{32} - c_{31}c_{21}}{c_{22}}$$

$$c_{33} = \sqrt{a_{33} - c_{31}^2 - c_{32}^2}$$

在一般情况下，**Cholesky 算法 (алгоритм Холецкого)** 如下所示对每个 $k = 1, \ldots, n$:

- 计算对角元:

$$c_{kk} = \left(a_{kk} - \sum_{j=1}^{k-1} c_{kj}^2 \right)^{1/2}$$

- 对每个 $i = k+1, \ldots, n$:

$$c_{ik} = \left(a_{ik} - \sum_{j=1}^{k-1} c_{ij} c_{kj} \right) / c_{kk}$$

假定上述指定表达式的求值方式使得对任何算数运算和求根运算的相对误差 ε，估计 $|\varepsilon| \leqslant \eta$ 成立

8.10 Cholesky 方法中的误差

对于实际计算的值 $\tilde{c}_{ij}$，有关系式

$$\tilde{c}_{kk}^2 = \left(a_{kk} - \sum_{j=1}^{k-1} \tilde{c}_{kj}^2 (1+\varepsilon)^{k-j} \right) (1+\varepsilon)^3$$

$$\tilde{c}_{ik}\tilde{c}_{kk} = \left(a_{ik} - \sum_{j=1}^{k-1} \tilde{c}_{ij}\tilde{c}_{kj} (1+\varepsilon)^{k-j} \right) (1+\varepsilon)^3$$

从第一个关系式可以得到，对于小的 ε，有

$$\sqrt{\sum_{j=1}^{k} \tilde{c}_{kj}^2} \leqslant \sqrt{a_{kk} + \mathcal{O}(\eta)} \leqslant \sqrt{a_{kk}} + \mathcal{O}(\eta)$$

从第二个关系式可以得到，对于小的 ε，成立

$$\tilde{c}_{ik}\tilde{c}_{kk} \left(1 - 3\varepsilon + \mathcal{O}\left(\varepsilon^2\right)\right) = a_{ik} - \sum_{j=1}^{k-1} \tilde{c}_{ij}\tilde{c}_{kj} \left(1 + (k-j)\varepsilon + \mathcal{O}\left(\varepsilon^2\right)\right),$$

$$\sum_{j=1}^{k} \tilde{c}_{ij}\tilde{c}_{kj} - a_{ik} = 3\varepsilon \tilde{c}_{ik}\tilde{c}_{kk} - \sum_{j=1}^{k-1} \tilde{c}_{ij}\tilde{c}_{kj}(k-j)\varepsilon + \mathcal{O}\left(\varepsilon^2\right)$$

$$\Rightarrow \left| \sum_{j=1}^{k} \tilde{c}_{ij}\tilde{c}_{kj} - a_{ik} \right| \leqslant \eta(k+1) \sum_{j=1}^{k} |\tilde{c}_{ij}| \cdot |\tilde{c}_{kj}| + \mathcal{O}\left(\eta^2\right) \leqslant$$

$$\leqslant \eta(k+1) \sqrt{\sum_{j=1}^{k} |\tilde{c}_{ij}|^2} \sqrt{\sum_{j=1}^{k} |\tilde{c}_{kj}|^2} + \mathcal{O}\left(\eta^2\right) \leqslant$$

$$\leqslant \eta(n+1) \sqrt{\sum_{j=1}^{k} |\tilde{c}_{ij}|^2} \sqrt{\sum_{j=1}^{i} |\tilde{c}_{kj}|^2} + \mathcal{O}\left(\eta^2\right) \leqslant \eta(n+1)\sqrt{a_{ii}}\sqrt{a_{kk}} + \mathcal{O}\left(\eta^2\right)$$

那么

$$\left| \tilde{C}\tilde{C}^T - A \right| \leqslant \eta(n+1) \begin{bmatrix} \sqrt{a_{11}} \\ \ldots \\ \sqrt{a_{nn}} \end{bmatrix} \left[\sqrt{a_{11}}, \ldots, \sqrt{a_{nn}} \right] + \mathcal{O}\left(\eta^2\right)$$

因此在 Cholesky 方法中没有由于矩阵元素的增长而增加误差的问题，和 LU 分解一样

第 9 章 方阵的 QR 分解与反射、旋转矩阵

9.1 QR 分解

定义 9.1 (QR 分解/QR-разложение)

设 $A\in\mathbb{C}^{n\times n}$。矩阵 A 的 QR 分解 **(QR-разложение)**：$A=QR$，其中 Q 是酉矩阵，R 是上三角阵 ♣

推论 9.1

因为 $R=Q^*A$，则

$$\max_{i,j}|r_{ij}|\leqslant\max_{i,j}|q_i^*a_j|\leqslant\max_j\|a_j\|_2\leqslant\sqrt{n}\max_{i,j}|a_{ij}|$$

♡

定理 9.1

任何方阵 A 都存在 QR 分解 ♡

证明 假设矩阵 A 是非退化矩阵，那么矩阵 A^*A 是正定的，并且 A^*A 中的所有顺序主子矩阵也都是正定的（根据正定矩阵的 Sylvester 判据）。因此存在 Cholesky 分解：$A^*A=R^*R$，其中 R 是对角元都是正数的上三角矩阵。现证矩阵 $Q=AR^{-1}$ 是酉矩阵：

$$Q^*Q=\left(AR^{-1}\right)^*\left(AR^{-1}\right)=\left(R^{-1}\right)^*\left(A^*A\right)R^{-1}=\left(R^{-1}\right)^*R^*RR^{-1}=I$$

若矩阵 A 是退化的，则对任意充分大的 $n\in\mathbb{N}$，扰动矩阵 $A_n=A+\frac{1}{n}I$ 非退化。令 $\lambda_1,\ldots,\lambda_k$ 是矩阵 A 的全部（不重复的）特征值，且 $\lambda_1=0$。那么矩阵 A_n 的特征值形如 $\frac{1}{n},\lambda_2+\frac{1}{n},\ldots,\lambda_k+\frac{1}{n}$ 且对于充分小的 n 都不等于 0

矩阵 A_n 存在 QR 分解：$A_n:A_n=Q_nR_n$。酉矩阵 Q_n 集合是紧集（关于范数有界 ($\|Q\|_2=1$)，构成一个闭集，且位于有限维空间 $\mathbb{C}^{n\times n}$ 上）。根据 Bolzano-Weierstrass 定理，存在子列 $Q_{n_k}\to Q$，当 $k\to\infty$ 时，极限矩阵 Q 是酉矩阵。那么 $Q_{n_k}^*A_{n_k}\to Q^*A=R$。矩阵 R 为上三角阵，因为矩阵 $Q_{n_k}^*A_{n_k}$ 中的每一个都是上三角阵

9.2 反射矩阵

定义 9.2 (Householder 矩阵/матрица Хаусхолдера)

矩阵 $H=H(u)=I-2uu^*$，其中 $\|u\|_2=1$，称为反射矩阵 **(матрица отражения)** 或 Householder 矩阵 ♣

性质 （反射矩阵最简单的性质）

- 矩阵 H 是酉矩阵：$HH^*=(I-2uu^*)(I-2uu^*)^*=I-4uu^*+4u(u^*u)u^*=I$
- 矩阵 H 是 Hermite 矩阵：$H^*=H$
- $Hu=-u$，$Hv=v$，$\forall v\perp u$

定理 9.2

对任意相同长度的向量 $a,b\in\mathbb{C}^n$，存在数 γ 和反射矩阵 H，满足：

$$Ha=\gamma b,\quad|\gamma|=1$$

♡

证明 从方程 $Ha=a-2u(u^*a)=\gamma b$ 中找到合适的向量 u。如果 a 和 b 是两个非零共线向量，那么可以取

$u = \frac{a}{\|a\|_2}, \gamma = \pm 1$

否则假定

$$u = \frac{a - \gamma b}{\|a - \gamma b\|_2},\ a - \gamma b \neq 0$$

带入后可得方程

$$2\left(u^* a\right) = \|a - \gamma b\|_2 \Leftrightarrow 2\left(a^* a - \gamma^* b^* a\right) = \|a - \gamma b\|_2^2 = \|a\|_2^2 + \|b\|_2^2 - 2\operatorname{Re}\left(\gamma^* b^* a\right)$$

因为 $\|a\|_2 = \|b\|_2$，则有等式 $\gamma^* b^* a = \operatorname{Re}\left(\gamma^* b^* a\right) \Leftrightarrow \gamma^* b^* a \in \mathbb{R}$。所以当 $b^* a = 0$ 时，可以任意取定 $\gamma : |\gamma| = 1$。否则 $\gamma = \pm \frac{b^* a}{|b^* a|}$（两种可能）

9.3 排除具有反射的元素 (Исключение элементов с помощью отражений)

对任意的 $a \in \mathbb{C}^n$ 都存在反射矩阵 H，满足

$$Ha = \gamma \left[\|a\|_2, 0, \ldots, 0\right]^T,\quad |\gamma| = 1$$

在这种情况下，矩阵 H 由向量 $u = \frac{v}{\|v\|_2}$ 定义，其中 $v = \left[a_1 - \gamma\|a\|_2, a_2, \ldots, a_n\right]^T$。若 $a_1 \neq 0$，则取 $\gamma = -\frac{a_1}{|a_1|}$

定理 9.3

对任意矩阵 $A \in \mathbb{C}^{n\times n}$ 都存在反射矩阵 $H_1, \ldots, H_{n-1}$，满足：矩阵 $R = H_{n-1} \ldots H_1 A$ 是上三角阵 ♡

这里矩阵 H_i 由向量 u_i 定义，并且向量 u_i 的前 $i-1$ 个分量为 0：

$$u_i^T = [0, \ldots, 0, *, \ldots, *]$$

这样的向量指定了一个不会改变自变量前 $i-1$ 个坐标的反射矩阵

$$H_i \left(x_1, \ldots, x_{i-1}, x_i, \ldots, x_n\right)^T = \left(x_1, \ldots, x_{i-1}, \tilde{x}_i, \ldots, \tilde{x}_n\right)^T$$

此外，假设取定向量 u_i 可以消除（清零）矩阵 $H_{i-1} \ldots H_1 A$ 第 i 列的第 $i+1, i+2, \ldots, n$ 个元素。结果是得到一个上三角阵

酉矩阵的乘积还是酉矩阵。因此 $Z = H_{n-1} \ldots H_1$ 是酉矩阵，$A = QR$，$Q = Z^*$。也就得到了 QR 分解的等价推导

9.4 旋转矩阵

定义 9.3 (旋转矩阵或 Givens 矩阵)

如果矩阵 $G_{kl} \in \mathbb{R}^{n\times n}$ 和单位矩阵仅有位于第 k, l 行、第 k, l 列交叉处元素构成的 2×2 子矩阵不同，且形如

$$M(\varphi) = \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix}$$

则该矩阵称为旋转矩阵 (**матрица вращения**) 或 **Givens** 矩阵 (**матрица Гивенса**)。 ♣

矩阵 G_{kl} 是正交矩阵，满足:$G_{kl} G_{kl}^T = I$。

若向量 $[a_1, a_2]^T \in \mathbb{R}^2$ 非零，且

$$\cos(\varphi) = \frac{a_1}{\sqrt{a_1^2 + a_2^2}},\quad \sin(\varphi) = -\frac{a_2}{\sqrt{a_1^2 + a_2^2}}$$

那么对应的旋转矩阵能删除向量的第二个分量:$M(\varphi)\left[a_1, a_2\right]^T = [\alpha, 0]^T, \alpha = \sqrt{a_1^2 + a_2^2}$。这个想法是基于**借助旋**

转消除元素法得到矩阵 A 的 QR 分解：矩阵 A 左乘旋转矩阵 G_{kl} 可以在位于第 k 或第 l 行，第 k 或第 l 列任何位置得到 0。也就是说矩阵 A 可以通过左乘旋转矩阵序列的方式化简成上三角阵 R：

$$G_{n,n-1}\dots G_{42}G_{32}G_{n1}\dots G_{31}G_{21}A=R$$

其中，矩阵 G_{ij} 将 (i,j) 位置的元素清零

9.5 反射和旋转方法的机器实现

使用反射或旋转矩阵 $\tilde{Q}$、$\tilde{R}$ 的实际计算，满足

$$\|A-\tilde{Q}\tilde{R}\|\leqslant c_1(n)\eta\|A\|+\mathcal{O}\left(\eta^2\right)$$
$$\left\|\tilde{Q}^*\tilde{Q}-I\right\|\leqslant c_2(n)\eta+\mathcal{O}\left(\eta^2\right)$$

其中 $c_1(n)$ 和 $c_2(n)$ 是关于 n 的某函数，取决于所使用的范数和算法实现的特征

使用反射矩阵得到矩阵的 QR 分解，需要

$$\frac{4}{3}n^3+\mathcal{O}\left(n^2\right)$$

次算术运算

使用旋转矩阵得到矩阵的 QR 分解，需要

$$2n^3+\mathcal{O}\left(n^2\right)$$

次算数运算

9.6 正交化方法

矩阵 QR 分解可以不借助反射或旋转得到

例题 9.1 （Gram-Schmidt 正交化过程）

令 $n=3$，目的是满足等式 $A=QR$，有：

$$\begin{bmatrix}a_1 & a_2 & a_3\end{bmatrix}=\begin{bmatrix}q_1 & q_2 & q_3\end{bmatrix}\begin{bmatrix}r_{11} & r_{12} & r_{13}\\0 & r_{22} & r_{23}\\0 & 0 & r_{33}\end{bmatrix}$$

当构建矩阵 QR 分解时，引入辅助向量 p_1,p_2,p_3，分别与 q_1,q_2,q_3 共线

矩阵第一列相等的条件：

$$p_1=a_1,\quad r_{11}=\|p_1\|_2,\quad q_1=\frac{p_1}{r_{11}}$$

使第二列相等，有

$$a_2=q_1r_{12}+q_2r_{22}$$

向量 q_2 应该与向量 q_1 正交。等式两侧同时乘以 q_1^*：

$$r_{12}=q_1^*a_2,\quad p_2=a_2-q_1r_{12},$$
$$r_{22}=\|p_2\|_2,\quad q_2=\frac{p_2}{r_{22}}.$$

使第三列相等，有

$$a_3=q_1r_{13}+q_2r_{23}+q_3r_{33}$$

等式两侧依次乘以 q_1^* 和 q_2^*，有：

$$r_{13}=q_1^*a_3,\quad r_{23}=q_2^*a_3,$$
$$p_3=a_3-q_1r_{13}-q_2r_{23},\quad r_{33}=\|p_3\|_2,\quad q_3=\frac{p_3}{r_{33}}$$

上述得到的算法是经典的 **Gram-Schmidt 正交化过程 (процесс ортогонализации Грама-Шмидта)**。一般形式如下:

$$p_j = a_j - \sum_{i=1}^{j-1} q_i \left(q_i^* a_j\right), \quad q_j = \frac{p_j}{\|p_j\|_2}, \quad j = 1, \ldots, k, \quad r_{ij} = q_i^* a_j$$

标准正交 n 个 n 维向量，需要 $2n^3 + \mathcal{O}\left(n^2\right)$ 次算术运算。

9.7 正交性损失

在机器算数的条件下，通过正交化方法得到的实际计算出的向量 $\tilde{q}_1, \ldots, \tilde{q}_k$ 的线性包络与由带有小扰动 $f_1, \ldots, f_k$ 的扰动向量 $a_1 + f_1, \ldots, a_k + f_k$ 张成的线性包络一致

通常在计算的结果中，向量 $\tilde{q}_1, \ldots, \tilde{q}_k$ 完全不正交。令 $\tilde{Q}_i = [\tilde{q}_1, \ldots, \tilde{q}_i], i = 1, \ldots, k$。作为向量 $\tilde{q}_1, \ldots, \tilde{q}_i$ 正交性的度量，可以取 $\delta_i = \left\|\tilde{Q}_i^* \tilde{Q}_i - I\right\|_2$

假设计算准确地在第 $(i+1)$ 步执行，$\|\tilde{q}_{i+1}\|_2 = 1$。那么:

$$\beta_{i+1} = \tilde{Q}_i^* \tilde{q}_{i+1} = \frac{1}{\|\tilde{p}_{i+1}\|_2} \tilde{Q}_i^* \left(a_{i+1} - \tilde{Q}_i \tilde{Q}_i^* a_{i+1}\right) = \frac{1}{\|\tilde{p}_{i+1}\|_2} \left(I - \tilde{Q}_i^* \tilde{Q}_i\right) \tilde{Q}_i^* a_{i+1}$$

因此

$$\|\beta_{i+1}\|_2 \leqslant \delta_i \sqrt{1+\delta_i} \frac{\|a_{i+1}\|_2}{\|\tilde{p}_{i+1}\|_2}$$

$$\tilde{Q}_{i+1} = \left[\tilde{Q}_i \tilde{q}_{i+1}\right] \Rightarrow \delta_{i+1} = \left\| \begin{bmatrix} \tilde{Q}_i^* \tilde{Q}_i - I & \beta_{i+1} \\ \beta_{i+1}^* & 0 \end{bmatrix} \right\|_2 \leqslant \delta_i + 2\|\beta_{i+1}\|_2$$

那么

$$\delta_{i+1} \leqslant \delta_i \left(1 + 2\sqrt{1+\delta_i} \frac{\|a_{i+1}\|_2}{\|\tilde{p}_{i+1}\|_2}\right)$$

这表明即使 δ_i 很小，那么对于小的 $\|\tilde{p}_{i+1}\|_2$，数值 δ_{i+1} 可以很大。这就是算法的 “瓶颈”

9.8 正交性损失的处理

假设计算出的向量 $\tilde{q}_1, \ldots, \tilde{q}_i$ “几乎正交”: $\delta_i < 1$

为了提高第 $(i+1)$ 步的正交性，考察下面的重新正交化过程:

$$p^{(0)} = a_{i+1}$$

$$p^{(j)} = \left(I - \tilde{Q}_i \tilde{Q}_i^*\right) p^{(j-1)}, j = 1, 2, \ldots$$

当 $j = 1$ 时的迭代过程对应 Gram-Schmidt 正交化过程一般步骤。固定某一 $j > 1$，且在第 j 次迭代时停止: $q_{i+1} = \frac{p^{(j)}}{\|p^{(j)}\|_2}$。然后

$$\beta^{(j)} = \tilde{Q}_i^* p^{(j)} = \left(I - \tilde{Q}_i^* \tilde{Q}_i\right)^j \tilde{Q}_i^* p^{(0)}$$

如果 $\delta_i < 1$，那么当 $j \to \infty$ 时，$\left(I - \tilde{Q}_i^* \tilde{Q}_i\right)^j \to 0$。因此，向量 $\tilde{q}_{i+1}$ 相比先前的所有向量有更高的正交性，即使它们是正交的，准确度也低得多。

9.9 修正的 Gram-Schmidt 算法

常规算法	修正算法
对每一个$j=1,\dots,k$：	对每一个$j=1,\dots,k$：
$p_j = a_j$	$p_j^{(0)} = a_j$
对每一个$i=1,\dots,j-1$：	对每一个$m=1,2,\dots,M$：
$p_j = p_j - q_i\,(q_i^* a_j)$	对每一个$i=1,\dots,j-1$：
$q_j = \frac{p_j}{\|p_j\|_2}$	$p_j^{(m)} = p_j^{(m-1)} - q_i\left(q_i^* p_j^{(m-1)}\right)$
	$q_j = \frac{p_j^{(M)}}{\left\|p_j^{(M)}\right\|_2}$

9.10 双对角线化

定义 9.4

矩阵 $B=(b_{ij})$ 如果满足 $(\forall i,j:(i>j)$或$(i+1<j)):b_{ij}=0$，则称为（上）双对角线型（**верхняя двухдиагональная** 或 **бидиагональная**）。

♣

任意矩阵 $A\in\mathbb{C}^{n\times n}$ 都可化为双对角线型 $B=PAQ$，其中 P 和 Q 是有限个反射或旋转矩阵的乘积

假设使用反射矩阵。首先在矩阵 A 左乘反射矩阵，将第一列对角线以下的所有元素清零。然后将所得的结果再右乘反射矩阵，将第一行从第 3 到第 n 个位置的元素清零。在这种情况下，前面从第一列得到的 0 不会改变

此外，通过左乘，将第二列所有对角线以下的元素归零。然后，通过右乘，在第二行的第 4 到第 n 个位置得到零。以此类推。每次乘以反射矩阵后，所有先前得到的 0 都保留在原处

借助酉双对角线矩阵，矩阵 A 的一些问题简化成关于双对角线矩阵 B 的更简单的问题。例如：

- 如果能得到双对角线矩阵 B 的奇异值分解，那么任意矩阵 A 的奇异值分解也就可以得到
- 矩阵 A 的最小二乘问题 $(\|Ax-b\|_2\to\min_x)$ 可以通过变量替换 $x=Qu, b=Pf$ 简化为双对角线矩阵 B 的最小化问题 $\|Bu-f\|_2\to\min_u$。这种情况下，保持范数

第 10 章 伪逆矩阵及其在求解线性方程组中的应用

10.1 Moore-Penrose 伪逆矩阵

Moore-Penrose 伪逆矩阵 (псевдообратная матрица Мура-Пенроуза)。

之前已经考虑过含有非退化方阵 A 的线性方程组 $Ax=b$。但是，在实践中经常会出现矩阵 A 是退化的或 A 是长方形矩阵（不是方阵）的情况。在这种情况下, 可以使用 Moore-Penrose 伪逆矩阵 A^+ 代替 A^{-1} 求解线性方程组

要确定伪逆矩阵，可以对矩阵 A 进行奇异值分解：

矩阵 A 的奇异值分解 $A=D\Lambda C^T$，这里 $\Lambda=\operatorname{diag}(\rho_1,\ldots,\rho_r,0,\ldots,0)$，$\rho_i$ 是矩阵 A 的奇异值

伪逆矩阵 $A^+=C\tilde{\Lambda}D^T$，其中 $\tilde{\Lambda}=\operatorname{diag}\left(\rho_1^{-1},\ldots,\rho_r^{-1},0,\ldots,0\right)$。

矩阵 C,D 是正交矩阵。这些矩阵的列包含两个奇异基的向量的坐标

另一种伪逆矩阵的构造方法是借助**骨架（标架）分解 (скелетное разложение)**。

10.2 矩阵的骨架（标架）分解

定义 10.1 (矩阵的骨架（标架）分解/skeleton Decomposition 或 скелетное разложение)

假设 $A\in\mathbb{C}^{m\times n},\operatorname{rg}A=r>0$

$$A=BC,\quad B\in\mathbb{C}^{m\times r},\quad C\in\mathbb{C}^{r\times n}$$

称为矩阵 A 的骨架分解 ♣

引理 10.1

在矩阵骨架分解中 $\operatorname{rg}B=\operatorname{rg}C=\operatorname{rg}A=r$ ♡

证明 矩阵乘积的秩不会超过因子矩阵的秩。那么有

$$r=\operatorname{rg}A=\operatorname{rg}BC\leqslant\min\{\operatorname{rg}B,\operatorname{rg}C\}$$

但 $\operatorname{rg}B\leqslant r,\operatorname{rg}C\leqslant r$。因此 $\operatorname{rg}B=\operatorname{rg}C=r$

引理 10.2

对任意的矩阵 $A\in\mathbb{C}^{m\times n}$ 都存在骨架（标架）分解 ♡

证明 为了得到矩阵的骨架（标架）分解，需取矩阵 A 的任意 r 个线性无关的列作为矩阵 $B=[b_1,\ldots,b_r]$ 的列，或者任意 r 个线性无关的并且可以线性表出矩阵 A 的列向量的列就足够了。那么矩阵 A 的任意第 j 列 $a_j,j=\overline{1,n}$ 都是矩阵 B 的列向量的线性组合，组合系数为 $c_{1j},\ldots,c_{rj}$。这些系数构成矩阵 C 的第 j 列：$C:a_j=Bc_j,j=\overline{1,n}\Rightarrow A=BC$

注 若矩阵 $A\in\mathbb{C}^{m\times n}$ 行满秩 $(\operatorname{rg}A=n)$，则可以方便地将矩阵 A 本身作为矩阵 B，将单位矩阵 I_n 作为矩阵 C

引理 10.3

矩阵 A 的骨架（标架）分解不是唯一的 ♡

证明 如果取 $B_1=BS$ 和 $C_1=S^{-1}C$ 代替矩阵 B 和 C，其中 S 是任意的非退化 $r\times r$ 矩阵，则 $A=B_1C_1$ 就是骨架分解另一种表示

引理 10.4

若矩阵 B 和 C 是矩阵 A 的骨架（标架）分解的组成成分，则矩阵 B^*B 和 CC^* 是非退化的 ♡

证明 假设 x 是方程 $B^*Bx=0$ 的任意解，证明其只能是 0。将方程两侧左乘 x^*：$x^*B^*Bx=(Bx)^*Bx=0$。等价于 $Bx=0$。但是矩阵 B 列满秩，因此 $x=0$ 是唯一解。由此可得，$\det B^*B\neq 0$。与 CC^* 的非退化性证明类似

10.3 Moore-Penrose 伪逆矩阵 (псевдообратная матрица Мура-Пенроуза)

考察矩阵方程

$$AXA=A$$

如果矩阵 A 是方阵且非退化，则该方程有唯一解 $X=A^{-1}$。如果矩阵 A 是任意 $m\times n$ 形长方形矩阵，则所求的解是 $n\times m$ 形矩阵，且不是唯一确定的。一般情况下，方程有无穷解集

定义 10.2 (Moore-Penrose 伪逆或广义逆矩阵)

矩阵 $A^+\in\mathbb{C}^{n\times m}$ 称为矩阵 $A\in\mathbb{C}^{m\times n}$ 的 **Moore-Penrose 伪逆或广义逆矩阵 (псевдообратная или обобщённая обратная матрица Мура-Пенроуза)**，如果满足条件

$$AA^+A=A,\quad A^+=UA^*=A^*V$$

其中 $U\in\mathbb{C}^{n\times n},V\in\mathbb{C}^{m\times m}$ 是某复矩阵。即有矩阵 A^+ 的行（列）是矩阵 A^* 的行（列）的线性组合 ♣

定理 10.1

任意的矩阵 A 的伪逆矩阵 A^+ 存在且唯一，并由公式

$$A^+=C^+B^+=C^*\left(CC^*\right)^{-1}\left(B^*B\right)^{-1}B^*$$

表示，其中 B 和 C 是矩阵 A 的骨架（标架）分解的组成成分 ♡

证明

存在性：若 $A=0$，则假设 $A^+=0$。现在令 $A\neq 0$。那么就有骨架分解 $A=BC$。求出矩阵 B^+ 和 C^+。根据伪逆矩阵的定义，有：

$$BB^+B=B,\ B^+=\tilde{U}B^*\Rightarrow B\tilde{U}B^*B=B$$

将等式两侧同时左乘矩阵 B^*：

$$B^*B\tilde{U}B^*B=B^*B\Rightarrow\tilde{U}=(B^*B)^{-1}$$

（这里关于 B^*B 的非退化性，使用引理10.4）因此

$$B^+=(B^*B)^{-1}B^*$$

类似地，可以得到公式 $C^+=C^*\left(CC^*\right)^{-1}$

证明矩阵 C^+B^+ 是矩阵 A 的伪逆矩阵，令 $K=\left(CC^*\right)^{-1}\left(B^*B\right)^{-1}$，那么

$$AA^+A=BCC^*\left(CC^*\right)^{-1}\left(B^*B\right)^{-1}B^*BC=BC=A$$

$$A^+=C^*KB^*=C^*K\left(CC^*\right)^{-1}CC^*B^*=UA^*$$

$$A^+=C^*KB^*=C^*B^*B\left(B^*B\right)^{-1}KB^*=A^*V$$

其中，$U=C^*K\left(CC^*\right)^{-1}C,V=B\left(B^*B\right)^{-1}KB^*$

唯一性：假设存在两个伪逆矩阵 A_1^+ 和 A_2^+，然后

$$AA_1^+A = AA_2^+A = A,\ A_1^+ = U_1A^* = A^*V_1,\ A_2^+ = U_2A^* = A^*V_2$$

$$\Rightarrow A\left(A_1^+ - A_2^+\right)A = 0,\ A_1^+ - A_2^+ = (U_1 - U_2)A^* = A^*(V_1 - V_2)$$

令 $D = A_1^+ - A_2^+$，$U = U_1 - U_2$，$V = V_1 - V_2$，那么

$$ADA = 0,\ D = UA^* = A^*V$$

$$\Rightarrow (DA)^*DA = A^*D^*DA = A^*V^*ADA = 0 \Rightarrow DA = 0$$

因此得到

$$DD^* = DAU^* = 0 \Rightarrow D = 0 \Rightarrow A_1^+ = A_2^+$$

10.4 伪逆矩阵的性质

性质

1) $(A^+)^* = (A^*)^+$
2) $(A^+)^+ = A$
3) $(AA^+)^* = AA^+$
4) $(A^+A)^* = A^+A$
5) $A^+AA^+ = A^+$
6) 矩阵 A，A^+，AA^+ 和 A^+A 具有相同的秩
7) $A^+ = A^*(AA^*)^{-1}$，若矩阵 A 行满秩
8) $A^+ = (A^*A)^{-1}A^*$，若矩阵 A 列满秩
9) $A = 0 \Leftrightarrow A^+ = 0$

10.5 应用伪逆矩阵求解线性方程组

定理 10.2

齐次方程组 $Ax = 0$ 的通解由等式

$$x = \left(I - A^+A\right)q$$

给出，其中 q 是长度（规模）合适的任意向量 ♡

证明 对任意的向量 q：

$$A\left(\left(I - A^+A\right)q\right) = Aq - AA^+Aq = Aq - Aq = 0$$

除此之外，对方程组的 $Ax = 0$ 的任意解 x，可以假设 $x = q$，那么有

$$\left(I - A^+A\right)x = x - A^+Ax = x$$

结论 （线性方程组解的唯一性的充要条件）

假设矩阵 A 列满秩。那么 $A^+A = I$。也就是说，线性方程组 $Ax = 0$ 的解唯一当且仅当矩阵 A 列满秩。如果方程组的解不唯一，那么存在解的无穷集（解集的势无穷大）

定理 10.3

设 $A \in \mathbb{C}^{m\times n}, b \in \mathbb{C}^m$。以下命题等价：

1. 方程 $Ax = b$ 有解
2. $\operatorname{rg}[A, b] = \operatorname{rg} A$

3. $AA^{+}b=b$ ♡

证明 条件 1 和条件 2 的等价性由 Kronecker-Capelli 定理 (теорема Кронекера-Капелли) 推出

证明条件 1 和条件 3 的等价性。假设方程 $Ax=b$ 是相容的。即存在 $\tilde{x}$，满足 $A\tilde{x}=b$。那么

$$b=A\tilde{x}=AA^{+}A\tilde{x}=AA^{+}b$$

即满足条件 3

现在假设 $AA^{+}b=b$，取 $\tilde{x}=A^{+}b$，那么有

$$A\tilde{x}=AA^{+}b=b$$

定理 10.4 (线性方程组解存在的充要条件)

使方程 $Ax=b$ 有解的充分必要条件是

$$AA^{+}b=b$$

在这种情况下，通解可以表示成 $x=A^{+}b+\left(I-A^{+}A\right)q$ 的形式，其中 q 是任意向量 ♡

证明 证明通解由公式 $x=A^{+}b+(I-A^{+}A)q$ 给出。假设 $AA^{+}b=b$，并令 $x^{0}=x-A^{+}b$。那么

$$Ax=b\Leftrightarrow Ax=AA^{+}b\Leftrightarrow A\left(x-A^{+}b\right)=0\Leftrightarrow Ax^{0}=0$$

由定理10.2可得 $x^{0}=\left(I-AA^{+}\right)q$，即 $x=A^{+}b+\left(I-A^{+}A\right)q$

10.6 线性方程组的正规伪解

对于方程组 $Ax=b$，考虑误差向量 $r=b-Ax$

向量 x^{0} 称为方程组 $Ax=b$ 的**伪解 (псевдорешение)**，如果对应该向量的数 $\|r\|=\left\|b-Ax^{0}\right\|$ 最小。具有最小长度的伪解称为线性方程组的**正规伪解 (нормальное псевдорешение)**

定理 10.5 (正规伪解的存在性与唯一性)

方程组 $Ax=b$ 的正规伪解总是存在且唯一，并由公式

$$x^{0}=A^{+}b$$

定义 ♡

证明 考虑任意向量 x，然后有

$$b-Ax=\left(b-Ax^{0}\right)+A\left(x^{0}-x\right)=u+v,\ u=b-Ax^{0}=b-AA^{+}b,\ v=A\left(x^{0}-x\right)$$

那么

$$\|b-Ax\|^{2}=(b-Ax)^{*}(b-Ax)=(u+v)^{*}(u+v)=u^{*}u+u^{*}v+v^{*}u+v^{*}v$$

$$v^{*}u=\left(x^{0}-x\right)^{*}A^{*}\left(b-AA^{+}b\right)=\left(x^{0}-x\right)^{*}\left(A^{*}-A^{*}AA^{+}\right)b$$

使用矩阵 A 的骨架（标架）分解：

$$A^{*}AA^{+}=C^{*}B^{*}BCC^{*}\left(CC^{*}\right)^{-1}\left(BB^{*}\right)^{-1}B^{*}=C^{*}B^{*}=A^{*}$$

由此推出 $v^{*}u=0$。同时又有 $u^{*}v=\left(v^{*}u\right)^{*}=0$，因此:

$$\|b-Ax\|^{2}=u^{*}u+v^{*}v=\|u\|^{2}+\|v\|^{2}=\left\|b-Ax^{0}\right\|^{2}+\left\|A\left(x^{0}-x\right)\right\|^{2}$$

即对任意的 x，有

$$\|b-Ax\|^{2}\geqslant\left\|b-Ax^{0}\right\|^{2}$$

因此 x^{0} 是方程组 $Ax=b$ 的伪解

接下来证明伪解 x^0 是正规的。令 x 是使得

$$\|b-Ax\|^2=\left\|b-Ax^0\right\|^2$$

成立的向量。那么（因为 $\|b-Ax\|^2=\left\|b-Ax^0\right\|^2+\left\|A\left(x^0-x\right)\right\|^2$）：

$$Az=0,\quad z=x^0-x$$

另一方面

$$\|x\|^2=\left(x^0-z\right)^*\left(x^0-z\right)=\left\|x^0\right\|^2+\|z\|^2-\left(x^0\right)^*z-z^*x^0$$

由于 $A^+=A^*V$，那么

$$\left(x^0\right)^*z=\left(A^+b\right)^*z=\left(A^*Vb\right)^*z=b^*V^*Az=0$$

类似地 $z^*x^0=\left(\left(x^0\right)^*z\right)^*=0$。因此 $\|x\|^2=\left\|x^0\right\|^2+\|z\|^2$。由此可得

$$\|x\|^2\geq\left\|x^0\right\|^2$$

并且等号仅仅存在于 $z=0$ 时，即当 $x=x^0$ 的时候，其中 $x^0=A^+b$

第 11 章 求特征值和特征向量的问题

11.1 特征值问题

特征值问题是与特征值（谱）和特征向量有关问题的总和

需要求出矩阵 $A \in \mathbb{C}^{n\times n}$ 的所有特征值。最简单的想法是：求出特征多项式的系数并将问题简化成计算该多项式的根的问题。为了求得特征多项式的系数可以构造的需要 $\mathcal{O}\left(n^3\right)$ 算术运算的直接方法。

借助特征多项式的想法并不总是好的。问题是特征值可能对矩阵元素的微小扰动弱敏感，但对其特征多项式的系数的微小扰动强敏感。也就是说一个好的问题变“不好”了

特别困难的是特征值对矩阵元素小扰动强敏感的情况

解决谱问题的普遍方法之一是研究**谱图 (спектральный портрет)**。对于给定的矩阵 A 和参数 $\varepsilon > 0$，谱图是集合：

$$S(\varepsilon) = \{\lambda \in \mathbb{C} : f(\lambda) = \sigma_{\min}(A - \lambda I) \leqslant \varepsilon\}$$

若 $\tilde{\lambda}-$ 矩阵 A 的特征值，则奇异值 $\sigma_{\min}(A - \tilde{\lambda}I) = 0$，因此 $\tilde{\lambda} \in S(\varepsilon)$

对于一些简单形式的矩阵，其特征值问题很容易解决：**对角阵 (диагональная матрица)**、**三对角线阵 (трёхдиагональная матрица)**、**三角阵 (треугольная матрица)**、**(分块)准三角阵 (почти треугольная матрица)**。三角阵或对角阵的特征值等于其对角元

很多求解特征值问题的数值方法都是基于使用相似变换 $B = P^{-1}AP$ 将矩阵简化为上述指定的简单形式之一。令 η 和 y 分别是矩阵 B 的特征值和特征向量。那么 $\eta y = By = P^{-1}APy \Rightarrow \eta(Py) = A(Py)$。因此 η 和 Py 是矩阵 A 的特征值和特征向量。那么有以下结论

推论 11.1

相似变换 (преобразование подобия) 不改变矩阵的特征值！ ♡

11.2 特征值问题的稳定性 (устойчивость)

假设矩阵 A 的特征向量能构成一个基。设 λ_i 是矩阵 A 的简单特征值，x_i 是对应的特征向量

如果稍微改动矩阵 A 的元素，那么精确到二阶小量 (с точностью до величин второго порядка малости)

$$(A + \Delta A)(x_i + \Delta x_i) = (\lambda_i + \Delta\lambda_i)(x_i + \Delta x_i)$$

后，特征值及其对应的特征向量的扰动满足线性化方程：

$$A \cdot \Delta x_i + \Delta A \cdot x_i \approx \Delta\lambda_i \cdot x_i + \lambda_i \cdot \Delta x_i$$

用未扰动的特征向量分解 Δx_i，有

$$\Delta x_i = \sum_{j=1}^{n} \xi_{ij} x_j$$

扰动的特征向量被确定，在一个因子的精度下，并取该因子使得 $\xi_{ii} = 0$

令 y_j 是矩阵 A^* 的对应的特征向量：$A^* y_j = \overline{\lambda_j} y_j$，那么

$$0 = \langle y_j, Ax_i\rangle - \langle A^* y_j, x_i\rangle = \langle y_j, \lambda_i x_i\rangle - \left\langle \overline{\lambda_j} y_j, x_i\right\rangle \Rightarrow \left(\overline{\lambda_i} - \overline{\lambda_j}\right)\langle y_j, x_i\rangle = 0$$

$$\Rightarrow \langle y_j, x_i\rangle = 0, \quad \text{若} \lambda_i \neq \lambda_j$$

$$\langle y_j, A \cdot \Delta x_i\rangle = \langle A^* y_j, \Delta x_i\rangle = \overline{\lambda_j}\langle y_j, \Delta x_i\rangle = \overline{\lambda_j}\xi_{ij}\langle y_j, x_j\rangle, \ \forall j$$

$$\langle y_j, \lambda_i \Delta x_i\rangle = \overline{\lambda_i}\xi_{ij}\langle y_j, x_j\rangle$$

现在将约等式先标量乘以 y_i，然后当 $j \neq i$ 时乘以 y_j，并且代入 Δx_i 在基向量 $x_1, \ldots, x_n$ 下的分解式，有

$$\langle y_i, x_i \rangle \Delta\lambda_i \approx \langle y_i, \Delta A \cdot x_i \rangle, \quad \xi_{ij} \left(\overline{\lambda_i} - \overline{\lambda_j} \right) \langle y_j, x_j \rangle \approx \langle y_j, \Delta A \cdot x_i \rangle$$

矩阵 ΔA 的扰动可以是任意的

特征值和特征向量的分量的最大误差不超过（精确到被舍弃的较高阶小量的无穷小项）以下值：

$$|\Delta\lambda_i| \lesssim |\chi_i| \cdot \max_{i,j} |(\Delta A)_{ij}|, \quad |\xi_{ij}| \lesssim \frac{|\chi_j| \cdot \|x_i\|_2}{|\lambda_i - \lambda_j| \cdot \|x_j\|_2} \cdot \max_{i,j} |(\Delta A)_{ij}|, \ j \neq i$$

其中 χ_i 是矩阵的**偏斜因子或扭曲因子 (коэффициент перекоса)**：

$$\chi_i = \frac{\|x_i\|_2 \cdot \|y_i\|_2}{\langle x_i, y_i \rangle} = \frac{1}{\cos(\varphi_i)}, \quad |\chi_i| \geqslant 1$$

其中 φ 是矩阵 A 和 A^* 对应特征向量之间的夹角

注 Hermite 矩阵 $A = A^*$ 的所有偏斜因子都等于 1，因为对应的向量都是一致的 (совпадают)，也就是说 Hermite 矩阵的特征值对矩阵元素的扰动不是很敏感

例题 11.1 假设矩阵 $A \in \mathbb{C}^{2\times 2}$ 有 Jordan 块的形式：

$$A = \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}, x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, A^* = \begin{bmatrix} a^* & 0 \\ 1 & a^* \end{bmatrix}, y_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

则 $\langle x_1, y_1 \rangle = 0$，且偏斜因子等于无穷大 ∞

例题 11.2

$$A = \begin{bmatrix} 20 & 20 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 19 & 20 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 18 & 20 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & 2 & 20 \\ \varepsilon & 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix}, \quad f(\lambda) = \det(A - \lambda I) = \prod_{k=1}^{20} (k - \lambda) - 20^{19}\varepsilon = 0$$

当 $\varepsilon = 0$ 时，多项式最低次项的系数 $a_0 = 20!$ 最小模特征值 $\lambda_{\min} = 1$。当 $\varepsilon = 20^{-19} \cdot 20! \approx 5 \cdot 10^{-7}$ 时 $a_0 = 0, \lambda_{\min} = 0$。可以看出，该矩阵的稳定性较差

结论 如果矩阵对应的偏斜因子较小，则矩阵特征值相对于矩阵元素的扰动是稳定的。若偏斜系数较大，那么稳定性可能会很差。如果所有的偏斜系数都较小，那么特征向量关于矩阵的所有元素都是稳定的

11.3 特征多项式插值法

如果已知特征多项式 $f(\lambda)$，那么其根可以通过近似计算得到（例如，求方程近似解的切线法或弦法）

为了快速求出矩阵的特征多项式，通常使用**插值法 (метод интерполяции)**。n 次多项式 $f(\lambda)$ 由其在 $n+1$ 个不同的点 λ_k 处的值唯一确定。也就是说，计算 $n+1$ 级矩阵 $A - \lambda_k I$ 的行列式就足够了，然后用 Newton 插值公式唯一地恢复补齐多项式

$$\lambda_0 < \lambda_1 < \lambda_2 < \ldots < \lambda_n - \text{ 插值节点}$$

$$f(\lambda) = f(\lambda_0) + \sum_{k=1}^{n} (\lambda - \lambda_0)(\lambda - \lambda_1) \ldots (\lambda - \lambda_{k-1}) \Delta f(\lambda_0, \lambda_1, \ldots, \lambda_k)$$

其中 $\Delta f(\lambda_0, \lambda_1, \ldots, \lambda_k)$ 是插值节点划分的差 (разделённые разности)

$$\Delta f(\lambda_0, \lambda_1) = \frac{f(\lambda_1) - f(\lambda_0)}{\lambda_1 - \lambda_0}$$

$$\Delta f(\lambda_0, \lambda_1, \lambda_2) = \frac{\Delta f(\lambda_1, \lambda_0) - \Delta f(\lambda_1, \lambda_2)}{\lambda_2 - \lambda_0}$$

等等。插值法仅对较小的 n 有意义

11.4 三对角线矩阵的特征多项式

令矩阵 $A = (a_{ij}) \in \mathbb{C}^{n\times n}$ 是三对角线矩阵。那么其特征多项式可以在不使用插值公式的情况下高效计算出，并且不用多次计算矩阵的行列式。令 $D_k(\lambda)$ 是矩阵 $A - \lambda I$ 的 k 阶主子式，则 $D_n(\lambda) = f(\lambda)$

对于 $k = 3, \dots, n$，通过最后一行分解行列式 $D_k(\lambda)$：

$$D_k(\lambda) = (a_{kk} - \lambda)\, D_{k-1}(\lambda) - a_{k,k-1}B_{k,k-1}(\lambda)$$

通过最后一列分解行列式 $B_{k,k-1}(\lambda)$：

$$B_{k,k-1}(\lambda) = a_{k-1,k}D_{k-2}(\lambda)$$

因此

$$D_k(\lambda) = (a_{kk} - \lambda)\, D_{k-1}(\lambda) - a_{k,k-1}a_{k-1,k}D_{k-2}(\lambda)$$

因此计算 $D_1(\lambda), D_2(\lambda)$ 并使用递推关系来计算 $D_n(\lambda)$ 足够了

注 对准三角矩阵 A 来说，也可以写出递推关系以便更有效地计算特征多项式。

11.5 求特征向量的逆迭代法

假设 λ_i 是矩阵 A 的特征值，$\tilde{\lambda}_i$ 是其近似值。那么可以得到 $\det\left(A - \tilde{\lambda}_i I\right) \neq 0$，但这个行列式的值很小。所以无法从线性方程组 $\left(A - \tilde{\lambda}_i I\right) x = \theta$ 中求出对应的特征向量，因为该方程组有唯一解 $x = \theta$。

使用**逆迭代法 (метод обратной итерации)** 求特征向量：取定任意向量 b 并考察线性方程组

$$\left(A - \tilde{\lambda}_i I\right) x = b$$

假设 n 级矩阵 A 有 n 个线性无关的特征向量 x_j（例如正规矩阵）：$(\forall j = 1, \dots, n) : Ax_j = \lambda_j x_j$。通过基 x_j 分解 x 和 b：

$$x = \sum_{j=1}^{n} \xi_j x_j,\ b = \sum_{j=1}^{n} \beta_j x_j \Rightarrow \sum_{j=1}^{n} \left(\xi_j \left(\lambda_j - \tilde{\lambda}_i\right) - \beta_j\right) x_j = 0$$

因此

$$\xi_j = \frac{\beta_j}{\lambda_j - \tilde{\lambda}_i}, \forall j$$

接下来考虑两种可能的情况

- **第一种情况**：λ_i- 简单特征值。那么在所有 ξ_j，$j = 1, \dots, n$ 中，只有一个是非常大的 $\Rightarrow$ 向量 $\frac{x}{\|x\|_2}$ 与 $\frac{x_i}{\|x_i\|_2}$ 几乎相同
- **第二种情况**：λ_i- 重特征值，即存在 $(1 < p \leqslant n) : \lambda_1 = \lambda_2 = \ldots = \lambda_p$。在这种情况下特征向量 $x_1, \dots, x_p$ 并不是唯一确定的。它们任意的线性组合也是 λ_1 对应的特征向量

系数 $\xi_1, \xi_2, \dots, \xi_p$ 将很大，而其余的系数小。所以找到是 $x_1, \dots, x_p$ 线性组合的向量 x 也就意味着 x 是所求的特征向量

注 逆迭代法也可以用于没有一组特征向量构成基的矩阵（没有 n 个线性无关特征向量的矩阵）

例题 11.3 令 $A = J_4(0)$ 是有特征值 $\lambda = 0$ 的 4 级 Jordan 块。假设 $\tilde{\lambda} = \varepsilon$，$b = [1, 1, 1, 1]^T$。那么在该方法的一次迭代后，可得

$$\begin{bmatrix} \varepsilon & 1 & 0 & 0 \\ 0 & \varepsilon & 1 & 0 \\ 0 & 0 & \varepsilon & 1 \\ 0 & 0 & 0 & \varepsilon \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \varepsilon x_1 + x_2 \\ \varepsilon x_2 + x_3 \\ \varepsilon x_3 + x_4 \\ \varepsilon x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \Rightarrow \begin{cases} x_4 = \varepsilon^{-1} \\ x_3 = -\varepsilon^{-2} + \varepsilon^{-1} \\ x_2 = \varepsilon^{-3} - \varepsilon^{-2} + \varepsilon^{-1} \\ x_1 = -\varepsilon^{-4} + \varepsilon^{-3} - \varepsilon^{-2} + \varepsilon^{-1} \end{cases}$$

将解乘以 $-\varepsilon^4$，得到 $x_1 = 1 + \mathcal{O}(\varepsilon)$，$x_2 = \mathcal{O}(\varepsilon)$，$x_3 = \mathcal{O}\left(\varepsilon^2\right)$，$x_4 = \mathcal{O}\left(\varepsilon^3\right)$，即 $x = e_1 + \mathcal{O}(\varepsilon), e_1 = [1, 0, 0, 0]^T$

第 12 章 求 Hermite 矩阵的特征值特征向量

12.1 Hermite 矩阵的反射法

设矩阵 $A \in \mathbb{C}^{n\times n}$ 是 Hermite 矩阵：$A = A^*$。这样的矩阵，即使对于较大的级数 n，也存在高效且稳定的求出所有特征值的方法，其基于通过相似变换将矩阵化成三对角线型或其他简单的形式，特征值问题很容易得到解决

详细研究**反射法 (метод отражений)**。证明对任意矩阵 A 都可以选出一个有限的反射序列，作用后将矩阵化成上准三角型（分块上三角矩阵）。为此，每次轮流反射都将矩阵 A 的下半部分中最长的非零列清零

假设矩阵第 $q-1$ 列已清零，将矩阵 A 分块：

$$\left[\begin{array}{ccc|cccc} * & * & * & * & * & * & * \\ * & * & * & * & * & * & * \\ 0 & * & * & * & * & * & * \\ \hline 0 & 0 & * & * & * & * & * \\ 0 & 0 & * & * & * & * & * \\ 0 & 0 & * & * & * & * & * \\ 0 & 0 & * & * & * & * & * \end{array}\right] = \left[\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array}\right]$$

方块 $A_1 \in \mathbb{C}^{q\times q}$ 是上准三角型，而在矩形块 $A_3 \in \mathbb{C}^{(n-q)\times q}$ 只有最后一列是非零的

使用前 q 个分量为零（$u_i = 0, \forall i \leqslant q$）的向量 $u = (u_i)$ 进行反射。那么反射矩阵 $P_q = I - 2uu^*$ 可以划分出和矩阵 A 大小相同的块：

$$P_q = P_q^{-1} = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & W \end{array}\right],\ W = (w_{ij}),\ w_{ij} = \delta_{ij} - 2u_i\bar{u}_j,\ q+1 \leqslant i,\ j \leqslant n$$

所需的相似变换具有以下形式：

$$B = P_q^{-1}AP_q = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & W \end{array}\right] \cdot \left[\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array}\right] \cdot \left[\begin{array}{c|c} I & 0 \\ \hline 0 & W \end{array}\right] = \left[\begin{array}{c|c} A_1 & A_2W \\ \hline WA_3 & WA_4W \end{array}\right] = \left[\begin{array}{c|c} B_1 & B_2 \\ \hline B_3 & B_4 \end{array}\right]$$

矩阵 B 的左上角块已经具有所需的准三角型的形式。由于矩阵 A_3 只有最后一列非零，那么矩阵 $B_3 = WA_3$ 也可以只有最后一列非零并且其元素等于

$$b_{iq} = a_{iq} - \alpha u_i,\ q+1 \leqslant i \leqslant n,\ \alpha = 2\sum_{j=q+1}^{n} \bar{u}_j a_{jq}$$

为了将矩阵 B_3 的最后一列的所有元素都归零，除了最上面的。还需要假定

$$(q+2 \leqslant i \leqslant n) : u_i = \frac{a_{iq}}{\alpha}$$

注意，如果将向量 u 乘以 $e^{i\varphi}, \forall\varphi \in \mathbb{R}$，那么反射矩阵不会发生改变。因此可以取 φ，使得 $\alpha \in \mathbb{R}$

当 $i = q+1$，从 b_{iq} 的公式中，可以得到

$$u_{q+1} = \frac{a_{q+1,q} - b_{q+1,q}}{\alpha}$$

要求 $u^*u = 1$：

$$\alpha^2 = \sum_{i=q+1}^{n} |a_{iq}|^2 + |b_{q+1,q}|^2 - \left(b_{q+1,q}^* a_{q+1,q} + b_{q+1,q} a_{q+1,q}^*\right)$$

现将 u_i 的表达式带入到 α 的定义中：

$$\alpha^2 = 2\sum_{i=q+1}^{n} |a_{iq}|^2 - 2b_{q+1,q}^* a_{q+1,q}$$

因此 $b_{q+1,q}^* a_{q+1,q} \in \mathbb{R} \Rightarrow \arg b_{q+1,q}^* = \pi k - \arg a_{q+1,q}$。可以取数字 k：例如 $k = \pm 1$。然后有 $b_{q+1,q}^* a_{q+1,q} \leqslant 0$ 和

$\alpha \neq 0$。当 $k=-1$ 时，$\arg b_{q+1,q} = \pi + \arg a_{q+1,q}$。即可将 α^2 的两个不同的表达式等同起来：

$$|b_{q+1,q}|^2 = \sum_{i=q+1}^{n} |a_{iq}|^2 \Rightarrow \alpha = \left(2\,|b_{q+1,q}|\,(|b_{q+1,q}| + |a_{q+1,q}|)\right)^{1/2}$$

因此，得到计算反射矩阵的显式公式，这些公式不使用复数值来计算实矩阵 A

依次取 $q = 1, 2, \ldots, n-2$ 并确定其对应的向量 $u = u(q)$，将矩阵 A 简化成上准三角型。若矩阵 A 是 Hermite 矩阵，则结果得到三对角线矩阵 $B = P^{-1}AP$，其中 P 是酉矩阵

那么如何提高每一步的计算效率：

- 矩阵块 A_1 不变。在块 $B_3 = WA_3$ 中，只有一个非零元 $b_{q+1,q}$，是在求反射矩阵时计算出来的
- 在求两个另外的块时，可以简化矩阵乘反射矩阵的运算：(矩阵乘法) 代替 A_4W 可以考虑关系式 $A_4W = A_4\,(I - 2ww^*) = A_4 - 2\,(A_4w)\,w^*$，这里已经两次需要计算矩阵向量乘法！大约快了 n 倍！对于左乘矩阵 W 也是类似的
- 若矩阵 A 为 Hermite 矩阵，那么 $B_2 = B_3^* \Rightarrow$ 矩阵块 B_2 可以单独不被计算（省略）。同样在块 B_4 中可以只计算下半部分

若矩阵 A 已简化成三对角线型 B，则矩阵 A 和 B 的特征值相同。若 $B = P^{-1}AP$，且 y 是矩阵 B 的特征向量，则 Py 是 A 的对应的特征向量

12.2 直接旋转法 (прямой метод вращений)

旋转法在速度上略逊于反射法，但其计算公式较为简单。这种方式允许借助相似变换将非 Hermite 矩阵变为准三角型矩阵，将 Hermite 矩阵变为三对角线矩阵

对复向量来说，初等旋转矩阵具有以下形式

$$G_{kl} = \begin{bmatrix} 1 & 0 & \ldots & 0 & 0 & 0 & 0 \\ & \ddots & & & & & \\ & & & \alpha & \ldots & -\beta^* & \\ & & & & \ddots & & \\ & & & \beta & \ldots & \alpha & \\ & & & & & \ddots & \\ 0 & \ldots & \ldots & \ldots & \ldots & \ldots & 1 \end{bmatrix},\ \alpha = \cos(\varphi),\ \beta = e^{-i\psi}\sin(\varphi),\ \alpha^2 + |\beta|^2 = 1$$

对于实向量 $\psi = 0$

矩阵 $B = AG_{kl}$ 与矩阵 A 的区别仅在于第 k 列和第 l 列的元素有不同：

$$b_{ik} = a_{ik}\alpha + a_{il}\beta,\ b_{il} = -a_{ik}\beta^* + a_{il}\alpha,\ 1 \leqslant i \leqslant n$$

$$b_{ij} = a_{ij},\ \forall j \neq k, l,\ 1 \leqslant i \leqslant n$$

矩阵 $C = G_{kl}^* B$ 与矩阵 B 的区别仅在于第 k 行和第 l 行的元素有不同：

$$c_{ki} = b_{ki}\alpha + b_{li}\beta^*,\ c_{li} = -b_{ki}\beta + b_{li}\alpha,\ 1 \leqslant i \leqslant n$$

$$c_{ji} = b_{ji},\ \forall j \neq k, l,\ 1 \leqslant i \leqslant n$$

矩阵 $C = G_{kl}^* A G_{kl}$ 与矩阵 A 的区别仅在两行和两列。若 A 是 Hermite 矩阵，那么 C 也是 Hermite 矩阵。

可以选取旋转角度，以消除位于平面旋转子矩阵左下角前面的元素 $c_{l,k-1}$

$$-b_{k,k-1}\beta + b_{l,k-1}\alpha = 0 \Rightarrow \alpha a_{l,k-1} = \beta a_{k,k-1}$$

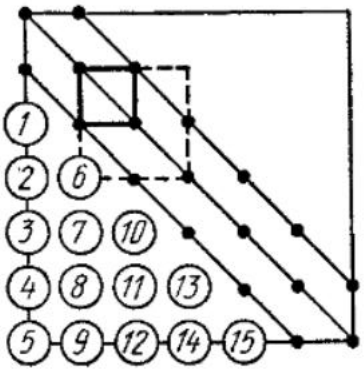
只需取

$$\alpha = \frac{|a_{k,k-1}|}{\sqrt{|a_{k,k-1}|^2 + |a_{l,k-1}|^2}},\quad \beta = \frac{\alpha a_{l,k-1}}{a_{k,k-1}}$$

在实数情况下，将公式改写成对称的形式更合适

$$\alpha = \frac{a_{k,k-1}}{\sqrt{a_{k,k-1}^2 + a_{l,k-1}^2}}, \quad \beta = \frac{a_{l,k-1}}{\sqrt{a_{k,k-1}^2 + a_{l,k-1}^2}}$$

现按照图中所示的顺序将矩阵的这些元素归零



矩阵第一个元素通过矩阵 G_{23} 消除（变为零），第二个元素被矩阵 G_{24} 清除。第二次旋转后，矩阵 A 中改变了第二行和第四行、第二列和第四列的元素。也就是说之前归零的第一个元素将保持等于 0

重复推理，可以确保一旦一个元素在这样的作用序列中被清除，将保持为零。因此，所有步骤完成后，矩阵将变成准上三角型矩阵（分块上三角阵）$(a_{ij} = 0, \forall i > j + 1)$。若矩阵 A 是 Hermite 矩阵，那么在这些步骤完成后将得到三对角线型矩阵

- 如何提高每一步中计算的效率？对于 Hermite 矩阵，只需要计算矩阵下半部分的变化元素
- 求得的三对角线型矩阵的特征值和原矩阵 A 的特征值相同
- 若 y 是三对角线矩阵的特征向量，则 $x = G_{23}G_{24}\dots G_{n-1,n}y$ 是矩阵 A 的特征向量。在不反复乘矩阵但向量 y 依次左乘旋转矩阵的情况下计算该向量更加有效。乘以矩阵 G_{kl} 只改变向量的第 k 个和第 l 个分量 $\Rightarrow$ 计算非常简单

12.3 迭代旋转法 (итерационный метод вращений)

所考虑的寻找特征值和特征向量的直接方法有一个共同的方案：

1) 原始矩阵的转换

2) 算出变换矩阵特征多项式的根

3) 通过反向迭代法求解特征向量

可是，对某些类矩阵，这种通用方法可能不是最有效的。因此考虑迭代法是有意义的，这些方法通常情况下速度较慢，但对部分特定类型的矩阵具有优势

对于 Hermite 矩阵，最著名的是**迭代旋转法 (итерационный метод вращений)**。它基于对满足在极限情况下将 Hermite 矩阵 A 转换成对角阵的初等旋转矩阵序列的选择。在这种情况下，旋转变换用于与直接旋转法作用相同类型的矩阵，但旋转的顺序先后及其角度是以完全不同的方式选择的

研究初等旋转如何作用于矩阵范数的平方：

$$S = \|A\|_E^2 = \sum_{i,j=1}^{n} |a_{ij}|^2$$

考虑右乘旋转矩阵：$B = AG_{kl}$。通过此变换，第 k 和第 l 列的元素以保持模长平方的成对和大小的方式发生改变

$$|b_{ik}|^2 + |b_{il}|^2 = |a_{ik}|^2 + |a_{il}|^2, \quad 1 \leqslant i \leqslant n \Rightarrow \|B\|_E^2 = \|A\|_E^2$$

将 Euclid 范数中包含的和分成对角线和、非对角线和：

$$S_1 = \sum_{i=1}^{n} |a_{ii}|^2, \quad S_2 = \sum_{i,j=1, i\neq j}^{n} |a_{ij}|^2$$

经过初等变换 $C = G_{kl}^* A G_{kl}$，矩阵非对角线元素 a_{ik}, a_{il}, a_{ki}，$a_{li}, i \neq k, l$ 以其模长（绝对值）的平方和保持不变的方式改变。除了这些元素以外，矩阵对角线之外还有另一个变化的元素 a_{kl}。这意味着 S_2 数值变化量和 $|a_{kl}|^2$ 的变化量一样多。进一步选取旋转使 S_2 减小

为了在一次旋转中最大程度减小 S_2，选取旋转角度以使 a_{kl} 归零。进一步假设 $A \in \mathbb{R}^{n\times n}, A = A^T$。然后

$$c_{kl} = b_{kl}\alpha + b_{ll}\beta = a_{kl}\left(\alpha^2 - \beta^2\right) + (a_{ll} - a_{kk})\,\alpha\beta$$

令 $c_{kl} = 0$ 且满足归一化条件 $\alpha^2 + \beta^2 = 1$。那么有

$$\beta^2 = 1 - \alpha^2, \quad \alpha^4 - \alpha^2 + a_{kl}^2\left(4a_{kl}^2 + (a_{kk} - a_{ll})^2\right)^{-1} = 0$$

可以选择生成的双二次方程 (биквадратное уравнение) 的四个根中的任何一个。例如

$$\alpha = \sqrt{\frac{1}{2}\left(1 + \frac{1}{\sqrt{1+\mu^2}}\right)}, \ \mu = \frac{2a_{kl}}{a_{kk} - a_{ll}}, \ \beta = (\operatorname{sgn}\mu)\sqrt{\frac{1}{2}\left(1 - \frac{1}{\sqrt{1+\mu^2}}\right)}$$

每次旋转时，S_2 减少而 S_1 增加，因为 $S_1 + S_2$ 保持不变。如果选择一个旋转序列使得 $S_2 \to 0$，那么在足够次数的旋转之后，所有对角线以外的元素变得任意小，且矩阵 A 转换成对角阵

得到的对角阵的对角元素就是所求的特征值

在有限次旋转中消除所有对角线以外的元素是不可能的，因为，与直接旋转法不同，这里，每次轮流的旋转后，之前为零的元素可能再次变成非零元

在（依次）当前旋转时，可以选择将对角线外最大模元素清零。这将导致 S_2 以最快速度降低（最高下降率）。但是寻找最大元素需要很多的时间

另一种方法是：选择最佳元素。假设

$$r_i = \sum_{j=1, j\neq i}^{n} |a_{ij}|^2$$

选取最大的 r_i 的总和，并在其中选取最大模元素。这是最理想的

选择最优元素的算法比寻找矩阵最大模元素更经济有效。r_i 的总和不需要在每一步完全重新计算，因为每次旋转只有 r_k 和 r_l 发生变化：

$$r_k' = r_k + (a_{kk}')^2 - a_{kk}^2 - a_{kl}^2, \quad r_l' = r_l + r_k - r_k'$$

最优元素是 r_i 的总和的不少于 $\frac{1}{n-1}$ 的部分，而整个 r_i 总和至少是 S_2 的 $\frac{1}{n}$。表明在一个旋转的过程中，S_2 至少减少 $\frac{2S_2}{n(n-1)}$。N 次旋转后 S_2 将至少减少 $\left(1 - \frac{2}{n(n-1)}\right)^N$ 倍。但

$$\left(1 - \frac{2}{n(n-1)}\right)^N \to 0, N \to \infty$$

因此，选择最优元素的过程收敛

第 13 章 寻找非 Hermite 矩阵的特征值和特征向量与部分特征值问题

13.1 初等变换法

经过相似变换（反射或旋转），任何非 Hermite 矩阵都能化成准三角型。此外，存在一种更有效的算法，可以将任何矩阵化成三对角线型，称为**初等变换法 (метод элементарных преобразований)**

该算法由两步骤组成。第一步将矩阵化成准三角型，第二步再化成三对角线型矩阵。在每一步都需要构建类似反射的初等相似变换的矩阵序列。第一步的变换依次将矩阵的列的下半部分变为 0，而第二步使矩阵的行的上半部分变为 0

第一步：在第 q 次迭代时，使用下述矩阵进行变换

$$N = \left[\begin{array}{c|c} E_q & 0 \\ \hline 0 & N_q \end{array}\right], E_q \in \mathbb{C}^{q\times q}, N_q \in \mathbb{C}^{(n-q)\times(n-q)}, N_q = N_q(v) = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ v_{q+2} & 1 & 0 & \dots & 0 \\ v_{q+3} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ v_n & 0 & 0 & \dots & 1 \end{bmatrix}$$

矩阵 N 不是酉矩阵：$N^*N \neq I$。但是 $N(v)N(-v) = I$，因此 $N^{-1}(v) = N(-v)$。矩阵 N 用于归零第 q 列的下半部分，如果前面的列已被处理

13.2 初等变换法：第一步

将矩阵 A 划分称大小规模相同的子矩阵块。使用和矩阵 N 相似的变换

$$B = N^{-1}AN = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & N_q(-v) \end{array}\right] \cdot \left[\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array}\right] \cdot \left[\begin{array}{c|c} I & 0 \\ \hline 0 & N_q(v) \end{array}\right] =$$

$$= \left[\begin{array}{c|c} A_1 & A_2N_q(v) \\ \hline N_q(-v)A_3 & N_q(-v)A_4N_q(v) \end{array}\right] = \left[\begin{array}{c|c} B_1 & B_2 \\ \hline B_3 & B_4 \end{array}\right]$$

矩阵块 A_3 和 $B_3 = N_q(-v)A_3$ 只有最后一列非零。该列的元素形如

$$b_{q+1,q} = a_{q+1,q}, \quad b_{iq} = a_{iq} - v_i a_{q+1,q}, \ q+2 \leqslant i \leqslant n$$

为了将矩阵块 B_3 中除了 $b_{q+1,q}$ 以外的所有元素都归零，需要规定：

$$v_i = \frac{a_{iq}}{a_{q+1,q}}, q+2 \leqslant i \leqslant n$$

该公式确定了矩阵所需的初等变换，该方法比找反射矩阵更加简单

可能的问题：$a_{q+1,q} = 0$ 或 $a_{q+1,q}$ 的数值非常小但异于 0

由于矩阵 N 的结构，乘以其自身和乘以向量的速度一样快。例如，找矩阵块 $B_2 = A_2N_q(v)$：

$$b_{ij} = a_{ij}, \quad q+2 \leqslant j \leqslant n, 1 \leqslant i \leqslant q$$

$$b_{i,q+1} = a_{i,q+1} + \sum_{j=q+2}^{n} a_{ij}v_j, 1 \leqslant i \leqslant q$$

也就是说，当右乘 N_q 时，只有块 A_2 的第一列发生变化，而块 B_2 的其余列分别等于块 A_2 中相对应的列。乘积 $C_4 = A_4N_q(v)$ 由类似的算法计算，但唯一的区别是：$i \in [q+1, n]$（而不是 $i \in [1, q]$）。左乘矩阵 N_q 会得到

另一个表达式。例如，对第四个块 $B_4 = N_q(-v)C_4$ 按元素（逐一元素）相乘，得到

$$b_{q+1,j} = c_{q+1,j}, \quad q+1 \leqslant j \leqslant n$$

$$b_{ij} = c_{ij} - v_i c_{q+1,j}, \quad q+1 \leqslant j \leqslant n, q+2 \leqslant i \leqslant n$$

也就是说几乎矩阵块的所有元素都发生变化

注 对 Hermite 矩阵来说，初等变换法没什么好处，因为在这种情况下，Hermite 矩阵不会得到保留。那么最后的矩阵将会是准三角型而不是三对角线型，就像反射法一样

基于所得公式的计算可能不够稳定。如果在当前迭代的计算过程中出现小的**主元** $a_{q+1,q}$，则分量 v_i 将会很大。当计算矩阵剩余块时，元素乘以这些分量，误差也会大大增加。为使方法稳定，应选择已处理列的主要（模最大）元素，那么有：

$$|a_{rq}| = \max_i |a_{iq}|, \quad q+2 \leqslant r, i \leqslant n$$

并且置换第 $(q+1)$ 和第 r 行，使其成为前导行（主行）。这时 $|v_i| \leqslant 1$，且误差几乎没有增加

两行的置换相当于左乘**置换矩阵 (матрица перестановки)** $P_{q+1,r}$：

$$A' = P_{q+1,r}A, \quad P_{q+1,r} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & & \ddots & & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}, \quad P = P^{-1} = P^*$$

行置换改变矩阵的特征值。为了使特征值不发生改变，需要进行相似变换 $A'' = P^{-1}AP = PAP = A'P$。右乘矩阵 P，相当于对矩阵 A' 进行列变换

13.3 初等变换法：第二步

原则上第一步将矩阵化简成（上）准三角形后，已经可以确定特征值。但是求解准三角型矩阵的完整特征值问题是困难的。为了利用初等变换法的高速度，需要将矩阵化成更合适的形式。将准三角型矩阵化成三对角线型矩阵

注 若矩阵 A 是一个下准三角型矩阵（当 $i+1 < j$ 时，$a_{ij} = 0$），那么在与矩阵 N 相似的变换作用后，矩阵 A 依旧是下准三角型矩阵。块 A_1 在此变换下不改变。在块 A_2 中，只有左下角元素 $a_{q,q+1}$ 非零 $\Rightarrow$ 在块 B_2 中，也只有元素 $b_{q,q+1} = a_{q,q+1}$ 非零。块 B_4 也有所需的形式

第二步：（将变换与矩阵对应）转置第一步所有变换。当转置 $N^{-1}(v)$ 时，得到矩阵

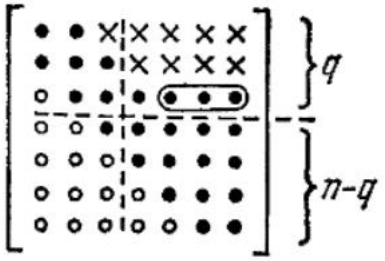
$$M = \left[\begin{array}{c|c} E_q & 0 \\ \hline 0 & M_q \end{array}\right], M_q = M_q(v) = \begin{bmatrix} 1 & -v_{q+2} & -v_{q+3} & \dots & -v_n \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}, M^{-1}(v) = M(-v)$$

若使用矩阵 M 对上三角阵进行相似变换，则生成的矩阵保持上准三角型

将相似变换矩阵链 $M^{-1}AM$ 应用在第一步结果得到的矩阵上。在每次迭代时，选择初等矩阵 $M(v)$，以便将当前行右半部分元素清零

对于前 $q-1$ 行所得到的矩阵应该是下准三角型矩阵（但上面已经指出其同时还保持上准三角型不变。这意味着该矩阵是一个三对角线型矩阵），接下来应该将第 q 行的部分元素清零

第二步的公式可以通过对第一步的公式应用转置运算很容易得到



例如，设前 $q-1$ 行已经被处理过，且需要将第 q 行的部分归零。那么

$$(q+2 \leqslant j \leqslant n): v_j = \frac{a_{q,j}}{a_{q,q+1}}$$

由于在第二步中，矩阵块 A_3 只有一个非零元素，但在块 A_4 中近一半元素为零。因此准三角型矩阵到三对角线型矩阵的相似变换是非常快速的。

如果在第二步的某一迭代中，主元 $a_{q,q+1}$ 变得非常小，那么计算结果变得很不稳定。这里通过置换行与列**不能提高稳定性**，因为这样的置换违背了矩阵的结构（不再是上准三角型矩阵）

13.4 QR 算法

假设矩阵 A 是任意非退化方阵。那么该矩阵可以被 QR 分解表示成：$A = Q_1 R_1$，其中 Q_1 是酉矩阵，R_1 是上三角阵。由此推出 $R_1 = Q_1^{-1} A$，并且 $A_1 = R_1 Q_1 = Q_1^{-1} A Q_1$ 与矩阵 A 相似

根据下述准则构造矩阵序列 $A_n, n = 1, 2, \ldots$：

- 将矩阵 A_n 分解成酉矩阵和上三角阵的乘积形式，形如 $A_n = Q_{n+1} R_{n+1}$。
- 令 $A_{n+1} = R_{n+1} Q_{n+1}$。

由于 $A_{n+1} = Q_{n+1}^{-1} A_n Q_{n+1}$，则所有矩阵 A_n 都彼此相似且相似于矩阵 A。

用形如 $A = Q \Lambda Q^{-1}$ 的方式表示矩阵 A，其中 Λ 是正规 Jordan 型矩阵，即当 $j < i$ 或 $j > i+1$ 时 $\lambda_{ij} = 0$。$\lambda_{ii} = \lambda_i$ 为矩阵 A 的特征值，$\lambda_{i,i+1}$ 等于 0 或 1。总可以取矩阵 Q 使矩阵 Λ 的对角元按照模长非增的顺序

$$|\lambda_1| = \ldots = |\lambda_{l_1}| > |\lambda_{l_1+1}| = \ldots = |\lambda_{l_2}| > \ldots > \left|\lambda_{l_{s-1}+1}\right| = \ldots = |\lambda_{l_s}|$$

进行排列

定理 13.1

假设在矩阵 A 的 QR 分解中，矩阵 Q 的所有对角线子式都非退化。那么矩阵 A_n 序列当 $n \to \infty$ 时，形式上收敛到分块上三角型 $\hat{A}$ 的形式，并且每一个块都对应特征值模的确定值 ♡

这里的意思是，对矩阵 A 同时进行一些相同序号的行置换和列置换后，将满足关系：若 $l_k < i \leqslant l_{k+1}, j < i$ 或 $l_{k+1} < j, k = 1, \ldots, s$，则 $a_{ij}^{(n)} \to 0 \left(A_n = \left\{a_{ij}^{(n)}\right\}\right)$。

如果 n 取很大的数值（足够大），将矩阵 A_n 在极限中应该变为 0 的所有元素归零，那么的得到一个分块上三角型的矩阵。其特征多项式等于其对角线各块的特征多项式的乘积

如果不仅需要找到矩阵 A 的特征值，还要找到其特征向量和伴随向量（присоединённый вектор）那么在构造矩阵 A_n 序列的过程中，记住正交矩阵 $P_n = Q_1 \ldots Q_n : P_{n+1} = P_n Q_{n+1}$

注 若矩阵 A 有 n 个不同的特征值 $\lambda_1, \ldots, \lambda_n$，则矩阵 $\hat{A}$ 是对角阵，特征值 $\lambda_1, \ldots, \lambda_n$ 位于其主对角线上

13.5 部分特征值问题

在很多问题中，并不是对矩阵的所有特征值都感兴趣，而仅仅是其中的一小部分。例如，最大模特征值或最小模特征值。如果矩阵 A 的维数非常大，那么在这种情况下求解矩阵的所有特征值是毫无益处的。搜索单个特

征值的大多数方法都是迭代法，也就是说需要构造收敛于单个特征值和特征向量的数值序列

幂方法 (Степенной метод) 求矩阵的最大模特征值。令 $|\lambda_1| > |\lambda_2| \geqslant |\lambda_3| \geqslant \ldots$ 构造迭代过程，假设矩阵 A 是正规矩阵：

$$(s = 0, 1, 2, \ldots) : x^{(s+1)} = Ax^{(s)}$$

按矩阵的特征向量分解第 0 次近似值：$x^{(0)} = \sum_i \xi_i x_i$。然后有 $x^{(s)} = \sum_i \lambda_i^s \xi_i x_i$，并且当 s 足够大时，$x^{(s)} \approx \lambda_1^s \xi_1 x_1$，若 $\xi_1 \neq 0$，那么向量 $x^{(s)}$ 按方向收敛到特征向量：

$$\exists \lim_{s\to\infty} \left\| \frac{x^{(s)}}{\left\|x^{(s)}\right\|} - \operatorname{sgn}(\xi_1) \frac{x_1}{\|x_1\|} \right\| = 0$$

对于足够大的 s，有 $x^{(s+1)} \approx \lambda_1 x^{(s)}$。特征值的模可以通过

$$|\lambda_1| \approx \frac{\left|x^{(s+1)}\right|}{\left|x^{(s)}\right|}$$

求得

若 $\{x_i\}$ 是由特征向量构成的标准正交基，且 $\lambda_1^s > 0$，在该情况下证明极限过程，那么有

$$\left\| \frac{x^{(s)}}{\left\|x^{(s)}\right\|} - \operatorname{sgn}(\xi_1) \frac{x_1}{\|x_1\|} \right\|^2 = \left\| \frac{\sum_i \lambda_i^s \xi_i x_i}{\sqrt{\sum_i (\lambda_i^s \xi_i)^2}} - \operatorname{sgn}(\xi_1) x_1 \right\| =$$

$$= \left\| \frac{\xi_1 x_1 + \sum_{i>1} \lambda_i^s (\lambda_1)^{-s} \xi_i x_i}{\sqrt{\xi_1^2 + \sum_{i>1} \left(\lambda_i^s (\lambda_1)^{-s} \xi_i\right)^2}} - \operatorname{sgn}(\xi_1) x_1 \right\| \to \|\operatorname{sgn}(\xi_1) x_1 - \operatorname{sgn}(\xi_1) x_1\| = 0$$

这里借助极限

$$\lambda_i^s (\lambda_1)^{-s} = \left(\frac{\lambda_i}{\lambda_1}\right)^s \to 0, \quad s \to \infty$$

和范数连续性的性质

13.6 移位取逆迭代 (обратные итерации со сдвигом)

假设矩阵 A 是非退化矩阵。考虑与幂方法相反的迭代过程：

$$x^{(s+1)} = A^{-1} x^{(s)}$$

收敛到矩阵 A^{-1} 的最大模特征值，也就是收敛到矩阵 A 的最小模特征值。

通过移位法可以加快收敛速度。该方法也可以用于搜索中间的特征值，不一定是最大模或最小模。

假定大致知道一些特征值 $\tilde{\lambda}_i$（的近似值）。则位移矩阵 $\left(A - \tilde{\lambda}_i I\right)$ 有特征值 $\left(\lambda - \tilde{\lambda}_i\right)$，其中 λ 是矩阵 A 的特征值。在该矩阵中，感兴趣的特征值 $\lambda_i - \tilde{\lambda}_i$ 的模比其他矩阵的特征值的模小得多。因此用移位矩阵反向迭代法

$$\left(A - \tilde{\lambda}_i I\right) x^{(s+1)} = x^{(s)}$$

可以非常快速地收敛且确定需要的特征值 $\lambda_i - \tilde{\lambda}_i$。最终的公式（需要标准化以免溢出）：

$$\left(A - \tilde{\lambda}_i I\right) y^{(s)} = x^{(s)}, \; x^{(s+1)} = \frac{y^{(s)}}{\left\|y^{(s)}\right\|}, \; \lambda_i^{(s)} - \tilde{\lambda}_i \approx \frac{\left\|x^{(s)}\right\|}{\left\|y^{(s)}\right\|}$$

第 14 章 求解线性方程组的迭代算法（例子和迭代方法的收敛性）

14.1 迭代法求解线性方程组

考察线性方程组

$$(x \in \mathbb{R}^n, b \in \mathbb{R}^n) : Ax = b, \quad A = (a_{ij}) \in \mathbb{R}^{n\times n}, \quad \det A \neq 0$$

求解线性方程组的任意迭代法，其主要目的都是构建一个序列 $\left\{x^{(k)}\right\}, k = 1, 2, \ldots$：当 $k \to \infty$ 时，$x^{(k)} \to x^*$，且满足 $Ax^* = b$

假设 $(\forall i = 1, \ldots, n) : a_{ii} \neq 0$。将方程组转换为以下形式

$$(i = 1, 2, \ldots, n) : x_i = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j - \sum_{j=i+1}^{n} \frac{a_{ij}}{a_{ii}} x_j + \frac{b_i}{a_{ii}}$$

迭代法的两个例子：

定义 14.1

- **Jacobi 迭代法 (Метод Якоби)**:

$$x_i^{(k+1)} = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k)} - \sum_{j=i+1}^{n} \frac{a_{ij}}{a_{ii}} x_j^{(k)} + \frac{b_i}{a_{ii}}, i = 1, 2, \ldots, n, k = 0, 1, \ldots, k^*, \forall x^{(0)}$$

- **Gauss-Seidel(G-S) 迭代法，即 Seidel 迭代法 (Метод Зейделя)**:

$$x_i^{(k+1)} = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k+1)} - \sum_{j=i+1}^{n} \frac{a_{ij}}{a_{ii}} x_j^{(k)} + \frac{b_i}{a_{ii}}, i = 1, 2, \ldots, n, k = 0, 1, \ldots, k^*, \forall x^{(0)}$$

♣

14.2 矩阵的表述形式

将矩阵 A 想象成三个矩阵的和：

$$A = A_1 + D + A_2, \quad D = \operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})$$

其中，矩阵 A_1 是对角线元素全为 0 的下三角阵，A_2 是对角线元素全为 0 的上三角阵。则线性方程组可以重新写成下述形式：

$$x = -D^{-1}A_1 x - D^{-1}A_2 x + D^{-1}b$$

Jacobi 迭代法在向量表示中如下所示：

$$x^{(k+1)} = -D^{-1}A_1 x^{(k)} - D^{-1}A_2 x^{(k)} + D^{-1}b$$

或

$$Dx^{(k+1)} + (A_1 + A_2)\, x^{(k)} = b \Leftrightarrow D\left(x^{(k+1)} - x^{(k)}\right) + Ax^{(k)} = b$$

Seidel 迭代法具有如下的矩阵形式：

$$x^{(k+1)} = -D^{-1}A_1 x^{(k+1)} - D^{-1}A_2 x^{(k)} + D^{-1}b$$

或

$$(D + A_1)\, x^{(k+1)} + A_2 x^{(k)} = b \Leftrightarrow (D + A_1)\left(x^{(k+1)} - x^{(k)}\right) + Ax^{(k)} = b$$

若 Jacobi 法或 Seidel 法收敛，则收敛到原始方程组的解

14.3 迭代法的加速

通常，为了加速收敛，将数值参数引入迭代法，这些参数可能取决于迭代的次数。例如在 Jacobi 迭代法和 Seidel 迭代法中分别引入迭代参数 τ_{k+1}，有

$$D\frac{x^{(k+1)}-x^{(k)}}{\tau_{k+1}}+Ax^{(k)}=b$$

$$(D+A_1)\frac{x^{(k+1)}-x^{(k)}}{\tau_{k+1}}+Ax^{(k)}=b$$

上面导出的 Jacobi 迭代法和 Seidel 迭代法属于**单步迭代法 (одношаговый итерационный метод)**，为了算出 $x^{(k+1)}$，仅需要知道前一次迭代的 $x^{(k)}$。有时也使用**多步迭代法 (многошаговые итерационные методы)**，其中 $x^{(k+1)}$ 通过 $x^k,\dots,x^{(k-l)}$ 的数值确定

单步迭代法的标准形式：

$$B^{(k+1)}\frac{x^{(k+1)}-x^{(k)}}{\tau_{k+1}}+Ax^{(k)}=b, k=0,1,\dots,k^*$$

这里，矩阵 $B^{(k+1)}$ 指定一个具体的方法，τ_{k+1} 表示迭代参数。假设存在矩阵 $\left(B^{(k+1)}\right)^{-1}$，并且 $x^{(k+1)}$ 可以如下定义:

$$B^{(k+1)}x^{(k+1)}=F^{(k)}, F^{(k)}=\left(B^{(k+1)}-\tau_{k+1}A\right)x^{(k)}+\tau_{k+1}b$$

14.4 显式和隐式法

定义 14.2

迭代方法如果满足 $B^{(k)}=I$，则称为显式方法 **(явный метод)**。其余的方法都是隐式方法 **(неявный метод)**

♣

隐式迭代法只有在每一个矩阵 $B^{(k)}$ 都比矩阵 A 本质上更容易求逆的情况下才有意义。例如，在 Seidel 法中矩阵 $B^{(k)}$ 是三角阵，因此更容易求逆。隐式方法的优点是收敛速度更快

定义 14.3

迭代法如果满足 $B^{(k+1)}\equiv B$，$\tau_{k+1}\equiv\tau$，则称稳定的 **(стационарный)**，反之是不稳定的 **(нестационарный)**

♣

例题 14.1

- 简单迭代法 (метод простой итерации)（显式、稳定的方法)：
$$\frac{x^{(k+1)}-x^{(k)}}{\tau}+Ax^{(k)}=b$$
- Richardson 迭代法 (итерационный метод Ричардсона)（显式、不稳定的方法)：
$$\frac{x^{(k+1)}-x^{(k)}}{\tau_{k+1}}+Ax^{(k)}=b$$
- 超松弛 (SOR) 迭代法 (метод верхней релаксации)：
$$(D+\omega A_1)\frac{x^{(k+1)}-x^{(k)}}{\omega}+Ax^{(k)}=b$$

14.5 迭代法的收敛性

考虑规范形式的单步、稳定迭代法：

$$B\frac{x^{(k+1)}-x^{(k)}}{\tau}+Ax^{(k)}=b, k=0,1,2,\dots$$

其中向量 $x^{(0)}$ 给定。迭代法收敛，如果满足：当 $k\to\infty$ 时，$\left\|x^{(k)}-x^*\right\|\to 0$。假定在以下任何地方都使用 Euclid 向量范数 $\|\cdot\|_2$。令 $z^{(k)}=x^{(k)}-x^*$ 为方法在第 k 次迭代的误差，且满足方程

$$B\frac{z^{(k+1)}-z^{(k)}}{\tau}+Az^{(k)}=0,\quad k=0,1,2,\dots,z^{(0)}=x^{(0)}-x^*$$

定理 14.1 (Самарский theorem)

假设 A 是对称且正定矩阵，$\tau>0$，并且满足不等式 $B-0.5\tau A>0$。则迭代法收敛

证明 证明当 $k\to\infty$ 时，有 $\left\|z^{(k)}\right\|\to 0$。从迭代法的方程中可得

$$z^{(k+1)}=\left(I-\tau B^{-1}A\right)z^{(k)}\Rightarrow Az^{(k+1)}=\left(A-\tau AB^{-1}A\right)z^{(k)}$$

因此

$$\left\langle Az^{(k+1)},z^{(k+1)}\right\rangle=\left\langle Az^{(k)},z^{(k)}\right\rangle-\tau\left\langle AB^{-1}Az^{(k)},z^{(k)}\right\rangle-$$
$$-\tau\left\langle Az^{(k)},B^{-1}Az^{(k)}\right\rangle+\tau^2\left\langle AB^{-1}Az^{(k)},AB^{-1}Az^{(k)}\right\rangle$$

由矩阵 A 的对称性，有：

$$\left\langle AB^{-1}Az^{(k)},z^{(k)}\right\rangle=\left\langle Az^{(k)},B^{-1}Az^{(k)}\right\rangle$$

因此

$$\left\langle Az^{(k+1)},z^{(k+1)}\right\rangle=\left\langle Az^{(k)},z^{(k)}\right\rangle-2\tau\left\langle (B-0.5\tau A)B^{-1}Az^{(k)},B^{-1}Az^{(k)}\right\rangle$$
$$B-0.5\tau A>0\Rightarrow\left\langle Az^{(k+1)},z^{(k+1)}\right\rangle<\left\langle Az^{(k)},z^{(k)}\right\rangle$$

这意味着，序列 $J_k=\left\langle Az^{(k)},z^{(k)}\right\rangle$ 单调递减且有下界 0（因为 $A>0$）。由此可知，$\exists\lim_{k\to\infty}J_k=J^*$

除此之外

$$B-0.5\tau A>0\Rightarrow\exists\delta>0:\left\langle (B-0.5\tau A)B^{-1}Az^{(k)},B^{-1}Az^{(k)}\right\rangle\geqslant\delta\left\|B^{-1}Az^{(k)}\right\|^2$$
$$\Rightarrow J_{k+1}-J_k+2\delta\tau\left\|B^{-1}Az^{(k)}\right\|^2\leqslant 0$$

当 $k\to\infty$ 时，趋向极限 $\Rightarrow\exists\lim_{k\to\infty}\left\|w^{(k)}\right\|=0,\quad w^{(k)}=B^{-1}Az^{(k)}$。矩阵 A,B 可逆，所以当 $k\to\infty$ 时，有 $z^{(k)}=A^{-1}Bw^{(k)}\to\theta$

注 矩阵 A 和矩阵 B 都是对称（自伴）实矩阵

14.6 迭代法的收敛性：Jacobi 法

在 Jacobi 法中

$$D\left(x^{(k+1)}-x^{(k)}\right)+Ax^{(k)}=b,\quad B=D,\quad \tau=1$$

定理 14.2

假设矩阵 A 是对称、正定的，且有行对角线优超：

$$a_{ii}>\sum_{j\neq i}|a_{ij}|,\forall i=1,2,\dots,n$$

那么 Jacobi 迭代法收敛

证明 收敛的条件 $B-0.5\tau A>0$，可以表示成 $A<2D$。接下来证明其由定理的条件推出。令 $\langle Ax,x\rangle=\sum_{i,j=1}^{n}a_{ij}x_ix_j$

$$\Rightarrow\langle Ax,x\rangle\leqslant\frac{1}{2}\sum_{i,j=1}^{n}|a_{ij}|\,x_i^2+\frac{1}{2}\sum_{i,j=1}^{n}|a_{ij}|\,x_j^2=\frac{1}{2}\sum_{i,j=1}^{n}|a_{ij}|\,x_i^2+\frac{1}{2}\sum_{i,j=1}^{n}|a_{ji}|\,x_i^2$$

$$\Rightarrow\langle Ax,x\rangle\leqslant\sum_{i,j=1}^{n}|a_{ij}|\,x_i^2=\sum_{i=1}^{n}x_i^2\left(\sum_{j\neq i}|a_{ij}|+a_{ii}\right)<2\sum_{i=1}^{n}a_{ii}x_i^2=2\langle Dx,x\rangle,\forall x\neq\theta$$

因此 $A<2D$

14.7 迭代法的收敛性：上松弛法 (метод верхней релаксации)

在上松弛法中

$$(D+\omega A_1)\,\frac{x^{(k+1)}-x^{(k)}}{\omega}+Ax^{(k)}=b,\quad B=D+\omega A_1,\quad \tau=\omega$$

定理 14.3

假设矩阵 A 是对称、正定矩阵。那么当 $\omega\in(0,2)$ 时，上松弛迭代法收敛。特别地，(当 $\omega=1$ 时) Seidel 法收敛

♡

证明 因为 $A_1=A_2^T$，那么有：

$$\langle Ax,x\rangle=\langle Dx,x\rangle+\langle A_1x,x\rangle+\langle A_2x,x\rangle=\langle Dx,x\rangle+2\langle A_1x,x\rangle$$

收敛的条件有下述形式：

若 $\omega\in(0,2)$

$$\langle Bx,x\rangle-0.5\omega\langle Ax,x\rangle=\langle(D+\omega A_1)\,x,x\rangle-0.5\omega(\langle Dx,x\rangle+2\langle A_1x,x\rangle)=$$

$$=(1-0.5\omega)\langle Dx,x\rangle>0$$

这里，使用性质：若 $A>0$，则 $D>0$

14.8 迭代法的收敛性: 简单迭代法

在简单迭代法中

$$\frac{x^{(k+1)}-x^{(k)}}{\tau}+Ax^{(k)}=b,\quad B=I$$

定理 14.4

假设 A 是对称、正定矩阵。那么简单迭代法收敛的条件具有以下形式：

$$I-0.5\tau A>0$$

令 $\lambda_i,i=1,2,\ldots,n$ 是矩阵 A 的特征值。若矩阵 $I-0.5\tau A$ 的所有特征值都是正的，则 (简单迭代法) 方法收敛。这等价于不等式 $1-0.5\tau\lambda_{\max}>0$。也就是说收敛性条件为：

$$\tau<\frac{2}{\lambda_{\max}(A)}$$

♡

证明 证明这个条件不仅仅是充分的，也是必要条件

令 $x^{(0)}=x^*+\mu$ 其中 μ 是矩阵 A 的对应于特征值 $\lambda_{\max}$ 的特征向量：$A\mu=\lambda_{\max}\mu$。那么 $z^{(0)}=\mu$

$$z^{(k)}=(I-\tau A)^kz^{(0)}=(I-\tau A)^k\mu\Rightarrow z^{(k)}=(1-\tau\lambda_{\max})^k\,\mu,\left\|z^{(k)}\right\|=|1-\tau\lambda_{\max}|^k\,\|\mu\|$$

若 $\tau=2\,(\lambda_{\max})^{-1}$，则 $\left\|z^{(k)}\right\|=\|\mu\|\nrightarrow 0$

若 $\tau > 2\left(\lambda_{\max}\right)^{-1}$，那么 $\left\|z^{(k)}\right\| \to \infty$

14.9 稳定迭代法的收敛准则

考虑稳定迭代过程：

$$B\frac{x^{(k+1)}-x^{(k)}}{\tau}+Ax^{(k)}=b,\quad B\frac{z^{(k+1)}-z^{(k)}}{\tau}+Az^{(k)}=\theta \Leftrightarrow z^{(k+1)}=Sz^{(k)}$$

迭代过程的收敛性取决于矩阵 $S=I-\tau B^{-1}A$ 的性质

定理 14.5

迭代法在任意初始近似 $x^{(0)}$ 下收敛，当且仅当 $\forall s$ 是矩阵 $S=I-\tau B^{-1}A$ 的特征值 $\Rightarrow |s|<1$ ♡

证明

必要性：假设矩阵 S 有特征值 $s:|s|>1$。令 μ 是特征值 s 对应的特征向量，$x^{(0)}=x^*+\mu$。那么

$$z^{(0)}=\mu, z^{(k)}=S^kz^{(0)}=s^kz^{(0)}=s^k\mu \Rightarrow \left\|z^{(k)}\right\|=|s|^k\cdot\|\mu\|\to\infty$$

充分性：首先考虑矩阵 S 有 n 个线性无关特征向量 $\mu_k, k=1,\ldots,n$ 的情况。令 $s_k, k=1,2,\ldots,n$ 是对应的特征值。用向量 μ_k 分解初始误差 $z^{(0)}=x^{(0)}-x^*$：

$$z^{(0)}=\sum_{j=1}^{n}c_j\mu_j \Rightarrow z^{(k)}=S^kz^{(0)}=\sum_{j=1}^{n}c_js_j^k\mu_j \Rightarrow \left\|z^{(k)}\right\|\leqslant\rho^k\sum_{j=1}^{n}|c_j|\cdot\|\mu_j\|\to 0$$

其中 $\rho=\max_{j=1,\ldots,n}|s_j|$ 是矩阵 S 的谱半径，$|\rho|<1$

在一般情况下，当矩阵 S 没有由特征向量构成的基（矩阵无法对角化），则使用矩阵的 Jordan 型 $\tilde{S}=P^{-1}SP, \tilde{S}=\operatorname{diag}\left(\tilde{S}_1,\ldots,\tilde{S}_l\right)$ 是分块对角矩阵，$\tilde{S}_k$ 是 Jordan 块。令 Jordan 块 $\tilde{S}_k$ 对应于特征值 s_k

构建改良 Jordan 型 $\hat{S}=D^{-1}\tilde{S}D$，其中 $D=\operatorname{diag}\left(1,\varepsilon,\ldots,\varepsilon^{n-1}\right), \varepsilon>0$。矩阵 $\hat{S}$ 具有和 $\tilde{S}$ 相同的分块对角型，但 Jordan 块此时具有下述形式：

$$\hat{S}_k=\begin{bmatrix} s_k & \varepsilon & 0 & 0 & \ldots & 0\\ 0 & s_k & \varepsilon & 0 & \ldots & 0\\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & 0 & 0 & \ldots & \varepsilon\\ 0 & 0 & 0 & 0 & \ldots & s_k \end{bmatrix}$$

矩阵 S 和 $\hat{S}$ 由等式 $\hat{S}=Q^{-1}SQ, Q=PD$ 联系起来。此外

$$\|\hat{S}\|_\infty\leqslant\rho(S)+\varepsilon,\quad \rho(S)=\max_k|s_k|$$

令 $\|y\|_*=\left\|Q^{-1}y\right\|_\infty$，然后有

$$\|S\|_*=\max_{y\neq\theta}\frac{\|Sy\|_*}{\|y\|_*}=\max_{y\neq\theta}\frac{\left\|Q^{-1}Sy\right\|_\infty}{\left\|Q^{-1}y\right\|_\infty}=\max_{x\neq\theta}\frac{\left\|Q^{-1}SQx\right\|_\infty}{\|x\|_\infty}=\|\hat{S}\|_\infty\leqslant\rho(S)+\varepsilon$$

根据条件 $\rho(S)<1$，那么 $\exists\varepsilon>0:\|S\|_*\leqslant\rho(S)+\varepsilon\leqslant q<1$。然后 $\left\|z^{(k)}\right\|_*\leqslant\|S\|_*^k\cdot\left\|z^{(0)}\right\|_*\leqslant q^k\left\|z^{(0)}\right\|_*$，则有：当 $k\to\infty$ 时，$\left\|z^{(k)}\right\|_*\to 0$

第 15 章 求解线性方程组的迭代算法（迭代法的收敛速度估计）

15.1 迭代法的收敛速度

考虑线性方程组 $Ax=b$ 的迭代法，从其获得近似值序列 $\left\{x^{(k)}\right\}, k=0,1,2,\ldots$

如果迭代法的误差满足以下条件：

$$(k=0,1,2,\ldots,\quad 0<q<1):\left\|x^{(k)}-x^{*}\right\| \leqslant q^{k}\left\|x^{(0)}-x^{*}\right\|$$

则称该方法以分母为 q 的几何级数的速度收敛。然后可以确定足以使初始误差减少所需次数倍的迭代次数：$\forall \varepsilon>0$，找到（合适的）$k \in \mathbb{N}$ 使得 $q^{k}<\varepsilon$，即 $k \geqslant K(\varepsilon)=\frac{\ln(1/\varepsilon)}{\ln(1/q)}$。那么

$$\left\|x^{(k)}-x^{*}\right\| \leqslant \varepsilon\left\|x^{(0)}-x^{*}\right\|$$

数 $K(\varepsilon)$ 的整数部分称为**获得给定 ε 精度所需的最小迭代次数 (минимальный число итераций, необходимый для получения заданной точности ε)**。表达式 $\ln(1/q)$ 称为**迭代法的收敛速度 (скорость сходимости итерационного метода)**。收敛速度由过度矩阵 S 的性质确定，且不依赖于迭代序号 k、不依赖于初始近似 $x^{(0)}$ 的选取、也不依赖于给定 ε 的准确度。不同迭代法的性能通常通过其收敛速度来比较：$\ln(1/q)$ 值越大，方法越好

15.2 收敛速度估计

考虑定常迭代法

$$B\frac{x^{(k+1)}-x^{(k)}}{\tau}+Ax^{(k)}=b$$

直接研究矩阵 $S=I-\tau B^{-1}A$ 的谱是非常困难的。需要更简单的收敛条件用来估计收敛速度。

定理 15.1

假设 A,B 是对称且正定的矩阵，有

$$\gamma_1 B \leqslant A \leqslant \gamma_2 B,\quad \gamma_1,\gamma_2>0,\gamma_2>\gamma_1$$

对于 $\tau=\frac{2}{\gamma_1+\gamma_2}$，迭代过程收敛且对于误差成立估计式

$$\left\|x^{(k)}-x^{*}\right\|_A \leqslant \rho^{k}\left\|x^{(0)}-x^{*}\right\|_A, k=0,1,2,\ldots$$

$$\left\|x^{(k)}-x^{*}\right\|_B \leqslant \rho^{k}\left\|x^{(0)}-x^{*}\right\|_B, k=0,1,2,\ldots$$

$$\rho=\frac{1-\xi}{1+\xi},\quad \xi=\frac{\gamma_1}{\gamma_2}$$

在证明定理之前，先考虑两个辅助引理。

♡

误差 $z^{(k)}=x^{(k)}-x^{*}$ 的方程形如：

$$B\frac{z^{(k+1)}-z^{(k)}}{\tau}+Az^{(k)}=0,\quad k=0,1,2,\ldots \Leftrightarrow z^{(k+1)}=Sz^{(k)}, S=I-\tau B^{-1}A$$

引理 15.1

假设 A,B 是对称且正定的矩阵，数 $\rho>0$。矩阵不等式

$$\frac{1-\rho}{\tau}B\leqslant A\leqslant\frac{1+\rho}{\tau}B$$

成立的充分必要条件是对任意的 $z^{(0)}$，迭代法误差满足估计 $\left\|z^{(k+1)}\right\|_A\leqslant\rho\left\|z^{(k)}\right\|_A,k=0,1,\ldots$

♡

证明 设 $w^{(k)}=A^{1/2}z^{(k)},\left\|w^{(k)}\right\|=\sqrt{\left\langle w^{(k)},w^{(k)}\right\rangle}$。然后可以将误差估计改写成 $\left\|w^{(k+1)}\right\|\leqslant\rho\left\|w^{(k)}\right\|$ 的形式。另外 $w^{(k+1)}=\tilde{S}w^{(k)}$，其中 $\tilde{S}=A^{1/2}SA^{-1/2}=I-\tau C,C=A^{1/2}B^{-1}A^{1/2}$。矩阵 $\tilde{S}$ 是对称矩阵，所以

$$\left\|w^{(k+1)}\right\|^2=\left\langle\tilde{S}w^{(k)},\tilde{S}w^{(k)}\right\rangle=\left\langle\tilde{S}^2w^{(k)},w^{(k)}\right\rangle$$

因此，所研究的估计等价于不等式 $\tilde{S}^2\leqslant\rho^2I\Leftrightarrow-\rho I\leqslant\tilde{S}\leqslant\rho I\Leftrightarrow\frac{1-\rho}{\tau}I\leqslant C\leqslant\frac{1+\rho}{\tau}I\Leftrightarrow\frac{1-\rho}{\tau}C^{-1}\leqslant I\leqslant\frac{1+\rho}{\tau}C^{-1}\Leftrightarrow\frac{1-\rho}{\tau}A^{-1/2}BA^{-1/2}\leqslant I\leqslant\frac{1+\rho}{\tau}A^{-1/2}BA^{-1/2}\Leftrightarrow\frac{1-\rho}{\tau}B\leqslant A\leqslant\frac{1+\rho}{\tau}B$

引理 15.2

假设 A,B 是对称且正定的矩阵，数 $\rho>0$。矩阵不等式

$$\frac{1-\rho}{\tau}B\leqslant A\leqslant\frac{1+\rho}{\tau}B$$

成立的充分必要条件是对任意的 $z^{(0)}$，迭代法误差满足估计 $\left\|z^{(k+1)}\right\|_B\leqslant\rho\left\|z^{(k)}\right\|_B,k=0,1,\ldots$

♡

证明 假设 $w^{(k)}=B^{1/2}z^{(k)},C=B^{-1/2}AB^{-1/2}$，那么

$$\left\|z^{(k+1)}\right\|_B\leqslant\rho\left\|z^{(k)}\right\|_B\Leftrightarrow\left\|w^{(k+1)}\right\|\leqslant\rho\left\|w^{(k)}\right\|$$

另外，$w^{(k+1)}=B^{1/2}SB^{-1/2}w^{(k)}=(I-\tau C)w^{(k)}$。

$$\left\|w^{(k+1)}\right\|\leqslant\rho\left\|w^{(k)}\right\|\Leftrightarrow\frac{1-\rho}{\tau}I\leqslant B^{-1/2}AB^{-1/2}\leqslant\frac{1+\rho}{\tau}I\Leftrightarrow$$
$$\Leftrightarrow\frac{1-\rho}{\tau}B^{1/2}A^{-1}B^{1/2}\leqslant I\leqslant\frac{1+\rho}{\tau}B^{1/2}A^{-1}B^{1/2}\Leftrightarrow$$
$$\frac{1-\rho}{\tau}A^{-1}\leqslant B^{-1}\leqslant\frac{1+\rho}{\tau}A^{-1}\Leftrightarrow\frac{1-\rho}{\tau}B\leqslant A\leqslant\frac{1+\rho}{\tau}B$$

接下来借助引理来证明定理：

证明 定理 (15.1) 的证明

假设

$$\rho=\frac{1-\xi}{1+\xi},\xi=\frac{\gamma_1}{\gamma_2},\tau=\frac{2}{\gamma_1+\gamma_2}$$

然后有

$$\frac{1-\rho}{\tau}=\frac{2\xi}{(1+\xi)\tau}=\frac{\gamma_1+\gamma_2}{1+\frac{\gamma_2}{\gamma_1}}=\gamma_1$$
$$\frac{1+\rho}{\tau}=\frac{2}{(1+\xi)\tau}=\frac{\gamma_1+\gamma_2}{1+\frac{\gamma_1}{\gamma_2}}=\gamma_2$$

并且定理的证明来自上述引理

15.3 最优迭代参数

考虑已证明定理的一些结论。

考察广义的特征值问题

$$A\mu=\lambda B\mu$$

若 $\gamma_1B\leqslant A\leqslant\gamma_2B$，则对任意的特征向量 μ，有

$$\gamma_1\langle B\mu,\mu\rangle\leqslant\langle A\mu,\mu\rangle=\lambda\langle B\mu,\mu\rangle\leqslant\gamma_2\langle B\mu,\mu\rangle$$

那么

$$\gamma_1 \leqslant \lambda_{\min}(A,B), \quad \gamma_2 \geqslant \lambda_{\max}(A,B) \tag{15.1}$$

其中 $\lambda_{\min}(A,B), \lambda_{\max}(A,B)$ 分别是最小和最大的特征值。因此，满足不等式 $\gamma_1 B \leqslant A \leqslant \gamma_2 B$ 的最精确常量是 $\gamma_1 = \lambda_{\min}(A,B), \gamma_2 = \lambda_{\max}(A,B)$

在满足条件 (15.1) 的参数 γ_1, γ_2 的集合上，两个变量

$$\rho = \frac{1-\xi}{1+\xi}, \quad \xi = \frac{\gamma_1}{\gamma_2}$$

的函数在点

$$\tau_0 = \frac{2}{\lambda_{\min}(A,B) + \lambda_{\max}(A,B)}, \quad \xi = \frac{\lambda_{\min}(A,B)}{\lambda_{\max}(A,B)}$$

处达到最小值。τ_0 是**最优迭代参数 (оптимальный итерационный параметр)**

$$\rho(\gamma_1,\gamma_2) = \tfrac{\gamma_2-\gamma_1}{\gamma_2+\gamma_1}$$
$$\rho'_{\gamma_1} = \tfrac{-2\gamma_2}{(\gamma_2+\gamma_1)^2} < 0, \quad \rho'_{\gamma_2} = \tfrac{2\gamma_1}{(\gamma_2+\gamma_1)^2} > 0$$

因此，函数 $\rho(\gamma_1,\gamma_2)$ 当

$$\gamma_1 = \gamma_{1,\max} = \lambda_{\min}(A,B), \gamma_2 = \gamma_{2,\min} = \lambda_{\max}(A,B)$$

时，达到最小值

15.4 简单迭代法的收敛速度

在应用中，当比值 $\frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$ 很大时，经常出现不良条件矩阵 (плохо обусловленная матрица)A 有关的问题。在这种情况下，ξ 近似于 0，数 ρ 近似于 1，且简单迭代法（对于 $B=I, \tau=\tau_0$）收敛缓慢。估计迭代次数 $k_0(\varepsilon)$ 是在 ξ 很小的情况下，达到给定的 ε 的精确度所需要的，即得到估计值

$$\left\|x^{(k)} - x^*\right\| \leqslant \varepsilon \left\|x^{(0)} - x^*\right\|$$

当 ξ 较小时，有 $\frac{1}{\rho} = \frac{1+\xi}{1-\xi} = (1+\xi)(1+\xi+\bar{o}(\xi)) = 1+2\xi+\bar{o}(\xi)$

$$\rho^k < \varepsilon \Leftrightarrow k \geqslant K(\varepsilon) = \frac{\ln(1/\varepsilon)}{\ln(1/\rho)} \approx \frac{\ln(1/\varepsilon)}{2\xi} = \mathcal{O}\left(\frac{1}{\xi}\right)$$

在 ξ 较小的情况下简单迭代法是缓慢收敛的

有两种方法可以加快迭代方法的收敛速度：

1) 通过使用隐式迭代法 ($B \neq I$) 计算

2) 保留在显式迭代法类中，可以根据迭代次数选择 $\tau = \tau_k$

也可以使用这两种迭代法的组合，即使用带有可变迭代参数的隐式迭代法。

在使用隐式方法时，通过矩阵 B 的选择，可以借助和 $\frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}$ 的对比，增加 $\xi = \frac{\lambda_{\min}\left(B^{-1}A\right)}{\lambda_{\max}(B^{-1}A)}$ 的值（换句话说，增加 $\xi = \frac{\lambda_{\min}\left(B^{-1}A\right)}{\lambda_{\max}(B^{-1}A)}$ 与 $\frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}$ 的比值）

15.5 非对称矩阵 B 的情况下的误差估计

现在放弃对矩阵 B 对称性的要求

引理 15.3

假设矩阵 D 是任意一个对称正定矩阵。若 $\exists B^{-1}$，则满足估计

$$(k=0,1,\ldots): \left\|z^{(k+1)}\right\|_D \leqslant \rho \left\|z^{(k)}\right\|_D$$

的充分必要条件是矩阵不等式

$$\rho^2 D \geqslant S^T D S$$

成立，其中 $S = I - \tau B^{-1} A, z^{(k+1)} = S z^{(k)}$

♡

证明 不等式 $\left\|z^{(k+1)}\right\|_D \leqslant \rho \left\|z^{(k)}\right\|_D$ 等价于

$$\left\langle DSz^{(k)}, Sz^{(k)} \right\rangle \leqslant \rho^2 \left\langle Dz^{(k)}, z^{(k)} \right\rangle \Leftrightarrow \rho^2 \left\langle Dz^{(k)}, z^{(k)} \right\rangle \geqslant \left\langle S^T DSz^{(k)}, z^{(k)} \right\rangle, k = 0, 1, \ldots$$

由于 $z^{(0)}$ 的任意性，则指定的不等式只有在满足 $\rho^2 D \geqslant S^T D S$ 的条件下成立

反过来：

$$\left\|z^{(k+1)}\right\|_D^2 = \left\langle Dz^{(k+1)}, z^{(k+1)} \right\rangle = \left\langle S^T DSz^{(k)}, z^{(k)} \right\rangle \leqslant \rho^2 \left\langle Dz^{(k)}, z^{(k)} \right\rangle = \rho^2 \left\|z^{(k)}\right\|_D^2$$

定理 15.2

假设矩阵 A 是一个对称正定矩阵，B 是一个非退化矩阵。如果以下含有某一与 k 无关的常数 $\rho \in (0,1)$ 的矩阵不等式成立

$$\frac{B^T + B}{2} - \frac{\tau}{2} A \geqslant \frac{1 - \rho^2}{2\tau} B^T A^{-1} B$$

则迭代法收敛且误差满足下述估计

$$\left\|x^{(k)} - x^*\right\|_A \leqslant \rho^k \left\|x^{(0)} - x^*\right\|_A$$

♡

证明 证明当 $D = A$ 时满足引理 (15.3) 的条件：

$$\rho^2 D \geqslant S^T D S \Leftrightarrow \rho^2 A \geqslant \left(I - \tau A \left(B^{-1}\right)^T\right) A \left(I - \tau B^{-1} A\right) \Leftrightarrow$$

$$\Leftrightarrow \tau A \left(\left(B^{-1}\right)^T + B^{-1}\right) A \geqslant \left(1 - \rho^2\right) A + \tau^2 A \left(B^{-1}\right)^T A B^{-1} A$$

将得到的不等式的右乘矩阵 $A^{-1}B$，左乘矩阵 $B^T A^{-1}$。然后得到等价的不等式

$$\tau \left(B + B^T\right) \geqslant \left(1 - \rho^2\right) B^T A^{-1} B + \tau^2 A$$

除 2τ 并从定理中得到所需的不等式。由引理可以推出 $\left\|z^{(k+1)}\right\|_A \leqslant \rho \left\|z^{(k)}\right\|_A$，并且 $\rho \in (0,1)$。因此当 $k \to \infty$ 时，$\left\|z^{(k)}\right\|_A \to 0$，该方法收敛，定理得证

第 16 章 求解线性方程组的迭代算法（带有 Chebyshev 参数集的迭代方法）

16.1 Chebyshev 多项式

考虑以下辅助问题：在所有首项系数为 1 的 n 次多项式中，需要找到一个多项式 $T_n(x)$，其值 $\max_{x\in[-1,1]}|T_n(x)|$ 是最小的。具有此性质的多项式称为 **Chebyshev 多项式 (многочлен Чебышёва)** 或**在闭区间 $[-1,1]$ 上偏离 0 最小的多项式 (многочлен,наименее уклоняющимся от нуля на отрезке $[-1,1]$)**。证明

$$T_n(x)=2^{1-n}\cos(n\arccos(x))$$

首先考察函数

$$P_n(x)=\cos(n\arccos(x)),x\in[-1,1]$$

借助三角公式展开后有公式

$$\cos((n+1)\arccos(x))+\cos((n-1)\arccos(x))=2\cos(n\arccos(x))\cdot\cos(\arccos(x))=2xP_n(x)$$

因此

$$P_{n+1}(x)-2xP_n(x)+P_{n-1}(x)=0,\quad P_0(x)=1,P_1(x)=x$$

通过归纳法可以证明，多形式 $P_n(x)$ 是首相系数为 2^{n-1} 的 n 次多项式 $\Rightarrow T_n(x)$ 是首相系数为 1 的 n 次多项式

接下来考虑该多项式的性质：多项式 $T_n(x)$ 的根位于点

$$x_k=\cos\frac{(2k+1)\pi}{2n},k=0,1,\ldots,n-1$$

极值在点

$$x'_k=\cos\frac{k\pi}{n},k=0,1,\ldots,n$$

处，并且

$$T_n\left(x'_k\right)=(-1)^k2^{1-n},k=0,1,\ldots,n$$

因此

$$\max_{x\in[-1,1]}|T_n(x)|=2^{1-n}$$

证明，在所有首相系数为 1 的 n 次多项式中，多项式 $T_n(x)$ 在区间 $[-1,1]$ 上与 0 的偏差较小，即

$$\max_{x\in[-1,1]}|T_n(x)|=\min\left\{\max_{x\in[-1,1]}|Q_n(x)|:Q_n(x)=x^n+\ldots\right\}$$

令 $Q_n(x)$ 是任意首相系数为 1 的 n 次多项式，设

$$\|Q_n\|=\max_{x\in[-1,1]}|Q_n(x)|$$

引理 16.1

假设存在一个点集

$$-1\leqslant x'_n<x'_{n-1}<\ldots<x'_1<x'_0\leqslant 1$$

满足

$$(k=0,1,\ldots,n):\left|Q_n\left(x'_k\right)\right|=\|Q_n\|$$

此外，数 $Q_n\left(x'_l\right)$ 有交替的符号。那么在所有首项系数为 1 的 n 次多项式中，多项式 $Q_n(x)$ 与 0 的偏离

量最小 ♡

证明 假设存在首项系数为 1 的 n 次多项式 $\hat{Q}_n(x)$，使得 $\left\|\hat{Q}_n\right\| < \|Q_n\|$，则有

$$\max_{x\in[-1,1]}\left|\hat{Q}_n(x)\right| < \|Q_n\|$$

设 $R(x) = Q_n(x) - \hat{Q}_n(x)$。该多项式次数不会超过 $n-1$，也不恒等于零

设 $Q_n\left(x'_k\right) = (-1)^k \|Q_n\|, k = 0, 1, \ldots, n$，那么

$$R\left(x'_k\right) = (-1)^k \|Q_n\| - \hat{Q}_n\left(x'_k\right), k = 0, 1, \ldots, n$$

区间 $[-1, 1]$ 上的多项式 $R(x)$ 改变符号 n 次，即其至少有 n 个根。但这是不可能的，因为多项式 $R(x)$ 是 $n-1$ 次多项式。得到矛盾！

推论 16.1

Chebyshev 多项式 $T_n(x)$ 满足该引理的条件 ♡

16.2 Chebyshev 多项式在任意区间上的情况

现在考虑在给定区间 $[a, b]$ 上所有的首项系数为 1 的 n 次多项式中寻找和零偏差最小的多项式的问题。通过将区间 $x \in [a, b]$ 变成区间 $t \in [-1, 1]$ 的变换 $t = \frac{2}{b-a}x - \frac{b+a}{b-a}$，将该问题简化为已考虑过的问题。通过该替换，Chebyshev 多项式 $T_n(t) = 2^{1-n}\cos(n\arccos(t))$ 转换成形式

$$F_n(x) = 2^{1-n}\cos\left(n\arccos\frac{2x-(b+a)}{b-a}\right)$$

并且项 x^n 的系数等于 $\frac{2^n}{(b-a)^n}$。因此，在所有首项系数为 1 的 n 次多项式中，在区间 $[a, b]$ 上和 0 偏离量最小的多项式如下：

$$T_n(x) = \frac{(b-a)^n}{2^{2n-1}}\cos\left(n\arccos\frac{2x-(b+a)}{b-a}\right)$$

该多项式的根位于点

$$(k = 0, 1, \ldots, n-1) : x_k = \frac{a+b}{2} + \frac{b-a}{2}\cos\frac{(2k+1)\pi}{2n}$$

与零的最大偏差是

$$\max_{x\in[a,b]}|T_n(x)| = \frac{(b-a)^n}{2^{2n-1}}$$

16.3 Chebyshev 多项式与交错归一化

考虑这个问题：在所有 $x = 0$ 时取值为 1 的 n 次多项式中，找到一个 n 次多项式 $P_n(x)$，使得在区间 $[a, b]$ 上其与 0 的偏差最小。所需多项式与 Chebyshev 多项式的不同之处仅在于归一化：

$$P_n(x) = \frac{T_n(x)}{T_n(0)}, \text{ если } T_n(0) \neq 0$$

当 $T_n(0) = 0$ 时，问题无解

替换 $T_n(x)$ 的表达式，可得

$$P_n(x) = p_n\cos\left(n\arccos\frac{2x-(b+a)}{b-a}\right), p_n = \left(\cos\left(n\arccos\frac{b+a}{a-b}\right)\right)^{-1}$$

令

$$\xi = \frac{a}{b}, \rho_0 = \frac{1-\xi}{1+\xi}. \Rightarrow p_n = \left(\cos\left(n\arccos\left(-\frac{1}{\rho_0}\right)\right)\right)^{-1}$$

下述比值正确：

$$\cos(n\arccos(-z))=(-1)^n\cos(n\arccos z)=\frac{(-1)^n}{2}\left(\left(z+\sqrt{z^2-1}\right)^n+\left(z-\sqrt{z^2-1}\right)^n\right)$$

当 $z=\frac{1}{\rho_0}$ 时，有

$$z-\sqrt{z^2-1}=\frac{1}{\rho_0}-\sqrt{\frac{1}{\rho_0^2}-1}=\frac{1-\sqrt{1-\rho_0^2}}{\rho_0}=\frac{1-\sqrt{\xi}}{1+\sqrt{\xi}},\quad z+\sqrt{z^2-1}=\frac{1+\sqrt{\xi}}{1-\sqrt{\xi}}$$

因此

$$p_n=(-1)^n\left(\frac{1}{2}\left(\rho_1^n+\frac{1}{\rho_1^n}\right)\right)^{-1}=(-1)^n\frac{2\rho_1^n}{1+\rho_1^{2n}},\rho_1=\frac{1-\sqrt{\xi}}{1+\sqrt{\xi}}$$

最后

$$P_n(x)=(-1)^nq_n\cos\left(n\arccos\frac{2x-(b+a)}{b-a}\right)$$

$$q_n=\frac{2\rho_1^n}{1+\rho_1^{2n}},\rho_1=\frac{1-\sqrt{\xi}}{1+\sqrt{\xi}},\xi=\frac{a}{b},b>a$$

多项式 $P_n(x)$ 的根位于点

$$(k=1,2,\ldots,n):x_k=\frac{a+b}{2}+\frac{b-a}{2}\cos\frac{2k-1}{2n}\pi$$

处

16.4 Chebyshev 多项式的应用

考虑多项式

$$f_n(\lambda)=(1-\tau_1\lambda)(1-\tau_2\lambda)\ldots(1-\tau_n\lambda)$$

需要选择合适的参数 $\tau_k>0,k=1,\ldots,n$ 代入，使

$$\max_{\gamma_1\leqslant\lambda\leqslant\gamma_2}|f_n(\lambda)|$$

的值最小。因为 $f_n(0)=1$，所以问题的解就是上面所考察的 Chebyshev 多项式 $P_n(\lambda)$。多项式 $f_n(\lambda)$ 的根等于 $\lambda_k=\tau_k^{-1},k=1,\ldots,n$，应该与多项式

$$P_n(\lambda)=(-1)^nq_n\cos\left(n\arccos\frac{2\lambda-(\gamma_1+\gamma_2)}{\gamma_2-\gamma_1}\right),\quad q_n=\frac{2\rho_1^n}{1+\rho_1^{2n}},\rho_1=\frac{1-\sqrt{\xi}}{1+\sqrt{\xi}},\xi=\frac{\gamma_1}{\gamma_2}$$

的根一致。这样的多项式 $P_n(\lambda)$ 的根位于点

$$\tilde{\lambda}_k=\frac{\gamma_1+\gamma_2}{2}+\frac{\gamma_2-\gamma_1}{2}\cos\frac{(2k-1)\pi}{2n},k=1,2,\ldots,n$$

上。因此，最佳参数为

$$\tau_k=\left(\frac{\gamma_1+\gamma_2}{2}+\frac{\gamma_2-\gamma_1}{2}\cos\frac{(2k-1)\pi}{2n}\right)^{-1},k=1,2,\ldots,n,\min_{\tau_k}\max_{\lambda}|f_n(\lambda)|=q_n$$

16.5 带有 Chebyshev 参数集的迭代法

考虑具有对称正定矩阵 A 的线性方程组 $Ax=b$。将使用非定常迭代法 (нестационарный итерационный метод)

$$\frac{x^{(k+1)}-x^{(k)}}{\tau_{k+1}}+Ax^{(k)}=b,k=0,1,\ldots,n-1$$

求解该问题，其中向量 $x^{(0)}$ 给定

考虑关于迭代参数的最优选择问题，即关于寻找正数 $\tau_1,\tau_2,\ldots,\tau_n$，使得第 n 次迭代的误差 $x^{(n)}-x^*$ 的范数最小

定理 16.1

假设矩阵 A 是对称正定矩阵，且 $\lambda_{\min}(A)>0,\lambda_{\max}(A)>0$ 分别为其最小和最大特征值。假设给定迭代次数为 n。在类型 (16.5) 的迭代法中使误差 $\left\|x^{(n)}-x^*\right\|$ 最小的方法是

$$\tau_k=\frac{\tau_0}{1+\rho_0 t_k},k=1,2,\ldots,n$$

其中

$$\tau_0=\frac{2}{\lambda_{\min}(A)+\lambda_{\max}(A)},\rho_0=\frac{1-\xi}{1+\xi},\xi=\frac{\lambda_{\min}(A)}{\lambda_{\max}(A)},t_k=\cos\frac{(2k-1)\pi}{2n},k=1,2,\ldots,n$$

此外，对误差的合理估计有

$$\left\|x^{(n)}-x^*\right\|\leqslant q_n\left\|x^{(0)}-x^*\right\|$$

$$q_n=\frac{2\rho_1^n}{1+\rho_1^{2n}},\rho_1=\frac{1-\sqrt{\xi}}{1+\sqrt{\xi}}$$

♡

证明 对于误差 $z^{(k)}=x^{(k)}-x^*$，方程

$$\frac{z^{(k+1)}-z^{(k)}}{\tau_{k+1}}+Az^{(k)}=0,k=0,1,\ldots,n-1,z^{(0)}=x^{(0)}-x^*$$

为真。因此

$$z^{(k)}=(I-\tau_kA)(I-\tau_{k-1}A)\ldots(I-\tau_1A)z^{(0)},k=1,2,\ldots,n$$

或

$$z^{(n)}=T_nz^{(0)},T_n=(I-\tau_nA)(I-\tau_{n-1}A)\ldots(I-\tau_1A)$$

矩阵 T_n 是对称矩阵，$\|T_n\|=\nu$，其中 ν 是矩阵 T_n 的最大模特征值。那么 $\left\|z^{(n)}\right\|\leqslant\nu\left\|z^{(0)}\right\|$，并且这种估算没有得到改善。现选取参数 $\tau_1,\tau_2,\ldots,\tau_n$，以便将 ν 最小化

令 $\lambda_k,k=1,\ldots,m$ 是矩阵 A 的特征值，可考虑

$$0<\lambda_{\min}(A)=\lambda_1\leqslant\lambda_2\leqslant\ldots\leqslant\lambda_m=\lambda_{\max}(A)$$

推出

$$\nu=\max_{k=1,\ldots,m}\left|(1-\tau_1\lambda_k)(1-\tau_2\lambda_k)\ldots(1-\tau_n\lambda_k)\right|$$

$$\nu\leqslant\max_{\lambda_{\min}(A)\leqslant\lambda\leqslant\lambda_{\max}(A)}\left|f_n(\lambda)\right|,f_n(\lambda)=(1-\tau_1\lambda)\ldots(1-\tau_n\lambda)$$

最优参数 $\tau_1,\ldots,\tau_n$ 由之前考虑的问题的解确定

$$\min_{\tau_1,\ldots,\tau_n}\lambda_{\min}\,(A)\leqslant\lambda\leqslant\lambda_{\max}(A)\left|f_n(\lambda)\right|$$

在定理16.1中找到的带有参数 τ_k 的迭代法，称为**带有 Chebyshev 参数集的显式迭代法 (явный итерационный метод с чебышёвским набором параметров)**

推论 16.2

当 $n=1$ 时，带有 Chebyshev 参数集的迭代法与简单迭代法一致

$$\frac{x^{(k+1)}-x^{(k)}}{\tau}+Ax^{(k)}=b,\quad\tau=\tau_0=\frac{2}{\lambda_{\min}(A)+\lambda_{\max}(A)}$$

♡

令 $\varepsilon>0-$ 所需的精度：$\left\|x^{(n)}-x^*\right\|\leqslant\varepsilon\left\|x^{(0)}-x^*\right\|$。若 $q_n<\varepsilon$，那么该条件将成立，即

$$\frac{1+\rho_1^{2n}}{\rho_1^n}>\frac{2}{\varepsilon}\Leftrightarrow\frac{1}{\rho_1^n}>\frac{1+\sqrt{1-\varepsilon^2}}{\varepsilon}$$

该条件将得以满足，若

$$\frac{1}{\rho_1^n}\geqslant\frac{2}{\varepsilon}\Leftrightarrow n\geqslant n_0(\varepsilon)=\frac{\ln(2/\varepsilon)}{\ln(1/\rho_1)}$$

在最坏的情况下，当 $\xi=\frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}$ 很小时

$$\ln\left(1/\rho_1\right)=\ln\left(\frac{1+\sqrt{\xi}}{1-\sqrt{\xi}}\right)\approx 2\sqrt{\xi},\quad n_0(\varepsilon)\approx\frac{\ln(2/\varepsilon)}{2\sqrt{\xi}}=\mathcal{O}\left(\frac{1}{\sqrt{\xi}}\right)$$

收敛速度比简单迭代法 $\left(n_0(\varepsilon)=\mathcal{O}\left(\frac{1}{\xi}\right)\right)$ 更好

16.6 隐式 Chebyshev 迭代法

现在考虑隐式迭代法

$$B\frac{x^{(k+1)}-x^{(k)}}{\tau_{k+1}}+Ax^{(k)}=b,k=0,1,\dots$$

向量 $x^{(0)}$ 已给定。矩阵 B 是对称正定的。该方法的收敛速度将不再由表达式 $\frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}$ 确定，而是表达式 $\frac{\lambda_{\min}\left(B^{-1}A\right)}{\lambda_{\max}(B^{-1}A)}$。需要选择合适的矩阵 B，以便该分数尽可能大。

定理 16.2

假设矩阵 A 和 B 是对称正定矩阵，而 $\lambda_{\min}\left(B^{-1}A\right),\lambda_{\max}\left(B^{-1}A\right)$ 分别是广义问题 $A\mu=\lambda B\mu$ 的最小、最大特征值。设给定的迭代次数为 n。方法有最小误差 $\left\|x^{(n)}-x^*\right\|_B$，若参数 τ_k 从以下公式确定

$$\tau_k=\frac{\tau_0}{1+\rho_0 t_k},k=1,2,\dots,n$$

其中

$$\tau_0=\frac{2}{\lambda_{\min}\left(B^{-1}A\right)+\lambda_{\max}\left(B^{-1}A\right)},\rho_0=\frac{1-\xi}{1+\xi},\xi=\frac{\lambda_{\min}\left(B^{-1}A\right)}{\lambda_{\max}\left(B^{-1}A\right)},t_k=\cos\frac{(2k-1)\pi}{2n},k=1,2,\dots,n$$

因此合理估计

$$\left\|x^{(n)}-x^*\right\|_B\leqslant q_n\left\|x^{(0)}-x^*\right\|_B$$

$$q_n=\frac{2\rho_1^n}{1+\rho_1^{2n}},\rho_1=\frac{1-\sqrt{\xi}}{1+\sqrt{\xi}}$$

♡

证明 误差 $z^{(k)}=x^{(k)}-x^*$ 满足齐次方程

$$B\frac{z^{(k+1)}-z^{(k)}}{\tau_{k+1}}+Az^{(k)}=0,k=0,1,\dots,z^{(0)}=x^{(0)}-x^*$$

将方程乘以矩阵 $B^{-1/2}$，并且有表达式 $v^{(k)}=B^{1/2}z^{(k)},C=B^{-1/2}AB^{-1/2}$。则得到方程：

$$\frac{v^{(k+1)}-v^{(k)}}{\tau_{k+1}}+Cv^{(k)}=0,k=0,1,\dots,v^{(0)}=B^{1/2}\left(x^{(0)}-x^*\right)$$

矩阵 C 是正定对称的。其谱是广义的特征值问题 $A\mu=\lambda B\mu$ 的解。对于得到的带有矩阵 C 的方程，可以使用定理16.1（将 A 换成 C）。因此 $\xi=\frac{\lambda_{\min}(C)}{\lambda_{\max}(C)}=\frac{\lambda_{\min}\left(B^{-1}A\right)}{\lambda_{\max}(B^{-1}A)}$。满足误差估计

$$\left\|v^{(n)}\right\|\leqslant q_n\left\|v^{(0)}\right\|\Leftrightarrow\left\|z^{(n)}\right\|_B\leqslant q_n\left\|z^{(0)}\right\|_B$$

推论 16.3

在定理 (16.2) 的条件中，可以得到估计：

$$\left\|x^{(n)}-x^*\right\|_A\leqslant q_n\left\|x^{(0)}-x^*\right\|_A$$

♡

为此，在证明中只需要进行替换 $v^{(k)}=A^{1/2}z^{(k)},C=A^{1/2}B^{-1}A^{1/2}$ 就足够了

第 17 章 求解线性方程组的迭代算法（变分类型的迭代方法）

17.1 变分迭代算法 (итерационные методы вариационного типа)

考虑形如

$$B\frac{x^{(k+1)}-x^{(k)}}{\tau_{k+1}}+Ax^{(k)}=b$$

的迭代法。在前面考虑的方法中，参数 τ_k 是借助边界 $\gamma_1 \leqslant \lambda_{\min}(A)$ 和 $\gamma_2 \geqslant \lambda_{\max}(A)$ 选取的矩阵 A 的特征值

现在将从对给定误差

$$\left\|x^{(k)}-x^*\right\|_D$$

来说，误差最小的条件

$$\left\|x^{(k+1)}-x^*\right\|_D$$

中选择这些参数。其中 D 是一个固定的对称正定矩阵

$$\|v\|_D=\sqrt{\langle Dv,v\rangle}$$

对于不同的矩阵 B 和 D 得到不同的迭代法

这些方法的收敛速度不高于 Chebyshev 多项式，但其优点是不需要知道矩阵 A 谱的边界

17.2 最小残差法

假设矩阵 A 是一个对称正定矩阵。令 $r^{(k)}=\left(Ax^{(k)}-b\right)$ 是第 k 次迭代的**残差 (невязка)**。误差 $z^{(k)}$ 和残差 $r^{(k)}$ 通过等式 $Az^{(k)}=r^{(k)}$ 联系起来。此外

将显式 $(B=I)$ 迭代法重新改写成 $x^{(k+1)}=x^{(k)}-\tau_{k+1}r^{(k)}$ 的形式

最小残差法 (метод минимальных невязок) 是一种迭代法，其中参数 τ_{k+1} 从给定范数 $\left\|r^{(k)}\right\|$ 下，$\left\|r^{(k+1)}\right\|$ 取最小值的条件中选择。得到关于迭代参数 τ_{k+1} 的显式表达式

$$Ax^{(k+1)}=Ax^{(k)}-\tau_{k+1}Ar^{(k)} \Rightarrow r^{(k+1)}=r^{(k)}-\tau_{k+1}Ar^{(k)}$$

即残差满足和误差相同的方程 $z^{(k)}=x^{(k)}-x^*$。将方程的两侧同时变成标量平方：

$$\left\|r^{(k+1)}\right\|^2=\left\|r^{(k)}\right\|^2-2\tau_{k+1}\left\langle r^{(k)},Ar^{(k)}\right\rangle+\tau_{k+1}^2\left\|Ar^{(k)}\right\|^2$$

当 $\tau_{k+1}=\tau_{k+1}^*=\frac{\left\langle Ar^{(k)},r^{(k)}\right\rangle}{\left\|Ar^{(k)}\right\|^2}$，达到最小值 $\left\|r^{(k+1)}\right\|$

在最小残差法中，从第 k 次迭代到第 $k+1$ 次迭代由以下方式实现：

1) 根据找到的 $x^{(k)}$ 的值计算残差向量 $r^{(k)}=Ax^{(k)}-b$

2) 按指定公式，求 $\tau_{k+1}=\tau_{k+1}^*$

3) 根据迭代法基本方程计算 $x^{(k+1)}$。

最小残差法以不低于具有最优参数 τ 的简单迭代法的速度收敛

定理 17.1

假设矩阵 A 是对称且正定的矩阵，最小残差法的误差满足估计：

$$(n=0,1,\ldots):\left\|A\left(x^{(n)}-x^*\right)\right\| \leqslant \rho_0^n\left\|A\left(x^{(0)}-x^*\right)\right\|$$

其中

$$\rho_0=\frac{1-\xi}{1+\xi},\quad \xi=\frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}$$

♡

证明 如果在代替最小化值 τ_{k+1}^* 的等式中

$$\left\|r^{(k+1)}\right\|^2 = \left\|r^{(k)}\right\|^2 - 2\tau_{k+1}\left\langle r^{(k)}, Ar^{(k)}\right\rangle + \tau_{k+1}^2\left\|Ar^{(k)}\right\|^2$$

取 $\tau_{k+1} = \tau_0 = \frac{2}{\lambda_{\min}(A)+\lambda_{\max}(A)}$，则有以下不等式

$$\left\|r^{(k+1)}\right\|^2 \leqslant \left\|(I-\tau_0 A)\, r^{(k)}\right\|^2 \Rightarrow \left\|r^{(k+1)}\right\| \leqslant \|I-\tau_0 A\| \cdot \left\|r^{(k)}\right\|, \quad \left\|r^{(k)}\right\| = \left\|z^{(k)}\right\|_A$$

在第十四课理论课中 (15.1) 已经得到估计

$$\|I-\tau_0 A\| = \sigma_{\max}(I-\tau_0 A) = \rho_0 = \frac{1-\xi}{1+\xi}, \xi = \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}$$

所以

$$\left\|r^{(k+1)}\right\| \leqslant \rho_0\left\|r^{(k)}\right\| \Rightarrow \left\|A\left(x^{(k+1)}-x^*\right)\right\| \leqslant \rho_0\left\|A\left(x^{(k)}-x^*\right)\right\|$$

17.3 最小校正法

考虑隐式迭代法

$$x^{(k+1)} = x^{(k)} - \tau B^{-1}r^{(k)}, r^{(k)} = Ax^{(k)} - b$$

向量 $w^{(k)} = B^{-1}r^{(k)}$ 称为第 $(k+1)$ 次迭代的**校正 (поправка)**。校正满足与隐式方法下误差 $z^{(k)} = x^{(k)} - x^*$ 相同的方程：

$$B\frac{w^{(k+1)} - w^{(k)}}{\tau_{k+1}} + Aw^{(k)} = 0$$

令矩阵 B 是对称正定矩阵。**最小校正法 (метод минимальных поправок)** 是一种隐式迭代法，其中参数 τ_{k+1} 从对给定的 $w^{(k)}$ 使 $\left\|w^{(k+1)}\right\|_B = \left\langle Bw^{(k+1)}, w^{(k+1)}\right\rangle^{1/2}$ 最小的条件中选择。在 $B = I$ 的情况下，最小校正法和最小残差法相吻合。重写修正方程

$$w^{(k+1)} = w^{(k)} - \tau_{k+1}B^{-1}Aw^{(k)}$$

$$\left\|w^{(k+1)}\right\|_B^2 = \left\|w^{(k)}\right\|_B^2 - 2\tau_{k+1}\left\langle Aw^{(k)}, w^{(k)}\right\rangle + \tau_{k+1}^2\left\langle B^{-1}Aw^{(k)}, Aw^{(k)}\right\rangle$$

如果 $\tau_{k+1} = \tau_{k+1}^* = \frac{\left\langle Aw^{(k)}, w^{(k)}\right\rangle}{\left\langle B^{-1}Aw^{(k)}, Aw^{(k)}\right\rangle}$，校正 $\left\|w^{(k+1)}\right\|_B$ 将是最小的

在每次迭代的最小校正法中，需要：

1) 求解线性方程组 $Bw^{(k)} = r^{(k)}$ 并求校正 $w^{(k)}$

2) 求解线性方程组 $Bv^{(k)} = Aw^{(k)}$ 并求向量 $v^{(k)} = B^{-1}Aw^{(k)}$

3) 计算 τ_{k+1}

最小校正法的收敛速度由广义的特征值问题 $Ax = \lambda Bx$ 的谱的边界决定

定理 17.2

假设 A, B 是对称正定矩阵，$\lambda_{\min}\left(B^{-1}A\right), \lambda_{\max}\left(B^{-1}A\right)$ 分别是广义问题下的最小和最大特征值。对于最小校正法的误差满足估计：

$$\left\|A\left(x^{(n)}-x^*\right)\right\|_{B^{-1}} \leqslant \rho_0^n \cdot \left\|A\left(x^{(0)}-x^*\right)\right\|_{B^{-1}}, n = 0, 1, \ldots$$

$$\rho_0 = \frac{1-\xi}{1+\xi}, \xi = \frac{\lambda_{\min}\left(B^{-1}A\right)}{\lambda_{\max}\left(B^{-1}A\right)}$$

♡

证明 设 $v^{(k)} = B^{1/2}w^{(k)}, C = B^{-1/2}AB^{-1/2}$，那么

$$\frac{v^{(k+1)} - v^{(k)}}{\tau_{k+1}} + Cv^{(k)} = 0, \quad \tau_{k+1}^* = \frac{\left\langle Cv^{(k)}, v^{(k)}\right\rangle}{\left\|Cv^{(k)}\right\|^2}$$

此外，根据定理17.1可以类似地得到：

$$\left\|v^{(k+1)}\right\| \leqslant \rho_0 \left\|v^{(k)}\right\|, \rho_0 = \frac{1-\xi}{1+\xi}, \xi = \frac{\lambda_{\min}(C)}{\lambda_{\max}(C)}$$

而 $\lambda_{\min}(C) = \lambda_{\min}\left(B^{-1}A\right), \lambda_{\max}(C) = \lambda_{\max}\left(B^{-1}A\right)$，此外

$$\left\|v^{(k)}\right\| = \left\|B^{1/2}w^{(k)}\right\| = \left\|B^{-1/2}r^{(k)}\right\| = \left\|r^{(k)}\right\|_{B^{-1}} = \left\|A\left(x^{(k)} - x^*\right)\right\|_{B^{-1}}$$

17.4 最速下降法

再次，考虑显式迭代法，并从对给定的 $z^{(k)}$ 使 $\left\|z^{(k+1)}\right\|_A$ 取最小值得条件中选择迭代参数 τ_{k+1}，其中 $z^{(k)} = x^{(k)} - x^*$。误差 $z^{(k)}$ 满足方程：

$$z^{(k+1)} = z^{(k)} - \tau_{k+1}Az^{(k)}$$

因此

$$\left\|z^{(k+1)}\right\|_A^2 = \left\|z^{(k)}\right\|_A^2 - 2\tau_{k+1}\left\langle Az^{(k)}, Az^{(k)}\right\rangle + \tau_{k+1}\left\langle A^2z^{(k)}, Az^{(k)}\right\rangle$$

如果 $\tau_{k+1} = \frac{\left\langle Az^{(k)}, Az^{(k)}\right\rangle}{\left\langle A^2z^{(k)}, Az^{(k)}\right\rangle}$，那么误差 $\left\|z^{(k+1)}\right\|_A$ 将是最小的

$z^{(k)}$ 的值未知，但 $Az^{(k)} = r^{(k)} = Ax^{(k)} - b$ 已知。因此 $\tau_{k+1} = \frac{\left\langle r^{(k)}, r^{(k)}\right\rangle}{\left\langle Ar^{(k)}, r^{(k)}\right\rangle}$

在结果中得到了**显式最速下降法 (явный метод скорейшего спуска)**。

与定理17.1一样，可以证明最速下降法收敛的速度和具有最优参数 $\tau = \tau_0$ 的简单迭代法收敛的速度相同。最速下降法的误差，满足估计

$$\left\|x^{(n)} - x^*\right\|_A \leqslant \rho_0^n \left\|x^{(0)} - x^*\right\|_A, n = 0, 1, \ldots, \rho_0 = \frac{1-\xi}{1+\xi}, \xi = \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}$$

以下方法称为**隐式最速下降法 (неявный метод скорейшего спуска)**：

$$B\frac{x^{(k+1)} - x^{(k)}}{\tau_{k+1}} + Ax^{(k)} = b$$

其中参数 τ_{k+1} 在使 $\left\|z^{(k+1)}\right\|_A$ 最小的条件中选择

误差 $z^{(k)} = x^{(k)} - x^*$ 满足方程

$$z^{(k+1)} = z^{(k)} - \tau_{k+1}B^{-1}Az^{(k)}$$

即有：

$$\left\|z^{(k+1)}\right\|_A^2 = \left\|z^{(k)}\right\|_A^2 - 2\tau_{k+1}\left\langle Az^{(k)}, B^{-1}Az^{(k)}\right\rangle + \tau_{k+1}^2\left\langle AB^{-1}Az^{(k)}, B^{-1}Az^{(k)}\right\rangle \Rightarrow$$

$$\Rightarrow \left\|z^{(k+1)}\right\|_A^2 = \left\|z^{(k)}\right\|_A^2 - 2\tau_{k+1}\left\langle r^{(k)}, w^{(k)}\right\rangle + \tau_{k+1}^2\left\langle Aw^{(k)}, w^{(k)}\right\rangle$$

如果 $\tau_{k+1} = \frac{\left\langle r^{(k)}, w^{(k)}\right\rangle}{\left\langle Aw^{(k)}, w^{(k)}\right\rangle}$，那么误差 $\left\|z^{(k+1)}\right\|_A$ 将是最小的

因此对于隐式最速下降法，估计

$$\left\|x^{(n)} - x^*\right\|_A \leqslant \rho_0^n \left\|x^{(0)} - x^*\right\|_A, n = 0, 1, \ldots, \rho_0 = \frac{1-\xi}{1+\xi}, \xi = \frac{\lambda_{\min}\left(B^{-1}A\right)}{\lambda_{\max}\left(B^{-1}A\right)}$$

成立

17.5 共轭梯度法 (метод сопряжённых градиентов)

共轭梯度法是一个两步迭代法，即为了求出新的迭代 $x^{(k+1)}$，要使用前面两次迭代 $x^{(k)}$ 和 $x^{(k-1)}$，这也就增大了所需的内存体积。但与此同时，在正确选择参数的情况下，收敛速度将高于一步迭代法。特别是所考虑的共轭梯度法在任何初始近似下，都在有限次迭代中收敛

假设 A 是一个系统矩阵 (матрица системы)，B 是一个对称正定矩阵。考虑下面一类隐式两步迭代法：

$$B\frac{\left(x^{(k+1)}-x^{(k)}\right)+\left(1-\alpha_{k+1}\right)\left(x^{(k)}-x^{(k-1)}\right)}{\tau_{k+1}\alpha_{k+1}}+Ax^{(k)}=b,k=1,2,\ldots$$

其中 α_{k+1},τ_{k+1} 是迭代参数。为了开始计算需要指定两个初始近似值 $x^{(0)}$ 和 $x^{(1)}$，初始近似值 $x^{(0)}$ 可任意取定，而向量 $x^{(1)}$ 要通过一步公式计算（当 $k=0,a_1=1$）：

$$B\frac{x^{(1)}-x^{(0)}}{\tau_1}+Ax^{(0)}=b$$

若求出参数 α_{k+1},τ_{k+1}，那么新的近似值 $x^{(k+1)}$ 通过前面的两个值 $x^{(k)}$ 和 $x^{(k-1)}$ 表出，公式如下：

$$x^{(k+1)}=\alpha_{k+1}x^{(k)}+\left(1-\alpha_{k+1}\right)x^{(k-1)}-\tau_{k+1}\alpha_{k+1}w^{(k)}$$

其中 $w^{(k)}=B^{-1}r^{(k)},r^{(k)}=Ax^{(k)}-b$

17.6 共轭梯度法中的误差最小化

误差 $z^{(k)}=x^{(k)}-x^*$ 满足方程

$$z^{(k+1)}=\alpha_{k+1}\left(I-\tau_{k+1}B^{-1}A\right)z^{(k)}+\left(1-\alpha_{k+1}\right)z^{(k-1)},k=1,2,\ldots,\quad z^{(1)}=\left(I-\tau_1B^{-1}A\right)z^{(0)}$$

令 $v^{(k)}=A^{1/2}z^{(k)},\left\|v^{(k)}\right\|=\left\|z^{(k)}\right\|_A,C=A^{1/2}B^{-1}A^{1/2}$。函数 $v^{(k)}$ 满足方程

$$v^{(k+1)}=\alpha_{k+1}\left(I-\tau_{k+1}C\right)v^{(k)}+\left(1-\alpha_{k+1}\right)v^{(k-1)},k=1,2,\ldots,\quad v^{(1)}=\left(I-\tau_1C\right)v^{(0)}$$

假设矩阵 A 和 B 是对称且正定矩阵，满足不等式

$$\gamma_1B\leqslant A\leqslant\gamma_2B,\gamma_2>\gamma_1>0$$

那么 $C=C^T>0$，并且 $\gamma_1I\leqslant C\leqslant\gamma_2I$。所以如果依次求得向量 $v^{(1)},v^{(2)},\ldots,v^{(k-1)}$，必可以得到 $v^{(k)}=P_k(C)v^{(0)}$，其中 $P_k(C)$ 是算子 C 对应的满足条件 $P_k(0)=I$ 的 k 次矩阵多项式

需要选择合适的迭代参数 τ_k,α_k，以便于对任意 $n=1,2,\ldots,\ \left\|v^{(n)}\right\|=\left\|z^{(n)}\right\|_A$ 最小

参数 τ_1 是从对给定的向量 $v^{(0)}$ 使 $\left\|v^{(1)}\right\|$ 取最小值的条件中求得的，类似于最速下降法 $\tau_1=\frac{\left\langle Cv^{(0)},v^{(0)}\right\rangle}{\left\|Cv^{(0)}\right\|^2}$。注意，对于以上选取的参数 τ_1，满足等式 $\left\langle Cv^{(1)},v^{(0)}\right\rangle=0$，令 $\alpha_1=1$

考虑在第 k 次迭代的误差 $v^{(k)}=P_k(C)v^{(0)}$。将多项式 $P_k(C)$ 写成以下形式：

$$P_k(C)=I+\sum_{i=1}^{k}a_i^{(k)}C^i,a_i^{(k)}=a_i^{(k)}\left(\alpha_i,\tau_i\right),i=1,2,\ldots,k$$

然后有

$$v^{(k)}=v^{(0)}+\sum_{i=1}^{k}a_i^{(k)}C^iv^{(0)},k=1,2,\ldots$$

$$\left\|v^{(n)}\right\|^2=\sum_{i,j=1}^{n}a_i^{(n)}a_j^{(n)}\left\langle C^iv^{(0)},C^jv^{(0)}\right\rangle+2\sum_{j=1}^{n}a_j^{(n)}\left\langle v^{(0)},C^jv^{(0)}\right\rangle+\left\|v^{(0)}\right\|^2$$

因此，$\left\|v^{(n)}\right\|^2$ 是关于变量 $a_1^{(n)},\ldots,a_n^{(n)}$ 的二次多项式。让导数 $\frac{\partial\left\|v^{(n)}\right\|^2}{\partial a_j^{(n)}},j=1,2,\ldots,n$ 等于 0，得到关于 $a_i^{(n)}$ 的方程组

$$\sum_{i=1}^{n}a_i^{(n)}\left\langle C^jv^{(0)},C^iv^{(0)}\right\rangle+\left\langle C^jv^{(0)},v^{(0)}\right\rangle=0$$

其解与最小值 $\left\|v^{(n)}\right\|^2$ 对应

使用已获得的方程，现在需要找到参数 $\alpha_k,\tau_k,k=1,2,\ldots,n$。注意，搜索 $a_i^{(n)}$ 的方程可以写成以下形式：

$$(j=1,2,\ldots,n):\left\langle C^jv^{(0)},v^{(n)}\right\rangle=0\tag{17.1}$$

引理 17.1

条件 (17.1) 等价于条件 $\left\langle Cv^{(j)}, v^{(n)} \right\rangle = 0, j = 0, 1, \ldots, n-1$ ♡

现在借助引理，要从条件

$$\left\langle Cv^{(j)}, v^{(n)} \right\rangle = 0, n = 1, 2, \ldots, j = 0, 1, \ldots, n-1$$

中求出迭代参数。也就是说，向量 $v^{(0)}, v^{(1)}, \ldots, v^{(k)}, \ldots$ 在标量积 $\langle u, v \rangle_C = \langle Cu, v \rangle$ 的意义上是两两正交的

假设参数 $\tau_1, \tau_2, \ldots, \tau_k, \alpha_1, \alpha_2, \ldots, \alpha_k$ 已经是最优选择的。使用关系式

$$\left\langle Cv^{(j)}, v^{(k+1)} \right\rangle = 0, j = 0, 1, \ldots, k$$

构造最优的参数 τ_{k+1}, α_{k+1}

可以说明，如果取

$$\tau_{k+1} = \frac{\left\langle Cv^{(k)}, v^{(k)} \right\rangle}{\left\| Cv^{(k)} \right\|^2}, \quad \alpha_{k+1} = \left(1 - \frac{\tau_{k+1}}{\tau_k} \cdot \frac{1}{\alpha_k} \cdot \frac{\left\langle Cv^{(k)}, v^{(k)} \right\rangle}{\left\langle Cv^{(k-1)}, v^{(k-1)} \right\rangle} \right)^{-1}$$

那么指定的等式将会满足

注意

$$v^{(k)} = A^{1/2} z^{(k)}, C = A^{1/2} B^{-1} A^{1/2}, z^{(k)} = x^{(k)} - x^*, Az^{(k)} = Ax^{(k)} - b = r^{(k)}, B^{-1} r^{(k)} = w^{(k)}$$

$$Cv^{(k)} = A^{1/2} w^{(k)}, \left\langle Cv^{(k)}, v^{(k)} \right\rangle = \left\langle w^{(k)}, r^{(k)} \right\rangle, \left\langle Cv^{(k)}, Cv^{(k)} \right\rangle = \left\langle Aw^{(k)}, w^{(k)} \right\rangle$$