第一部分 矩阵论

1 λ 矩阵与 Jordan 标准型

1.1 λ 矩阵

定义 1.1 (多项式矩阵 λ-矩阵)**.** 若 $a_{ij}(\lambda), i=1,2,\ldots,m, j=1,2,\ldots,n$ 为 λ 的多项式，构成的矩阵 $A(\lambda)$ 或记为 A 称为**多项式矩阵**或 **λ-矩阵**。

定义 1.2 (秩)**.** λ-矩阵的秩定义为矩阵中最大的主子式并非全零的阶数。注意行列式为零指为多项式 0，而不是能取到 0。

定义 1.3 (可逆 逆矩阵)**.** 对给定 n 阶方形 λ-阵 $A(\lambda)$，定义 $A(\lambda)$ 可逆为 $\exists B(\lambda), A(\lambda)B(\lambda)=B(\lambda)A(\lambda)=I$，称 $B(\lambda)$ 为 $A(\lambda)$ 的逆矩阵，记为 $A^{-1}(\lambda)$。

定理 1.4.

$$A^{-1}=\frac{1}{\det A(\lambda)}A^*(\lambda)$$

其中右侧为行列式与伴随矩阵。

推论 1.5. 对方阵 $A(\lambda)$，$A(\lambda)$ 可逆当且仅当 $\det A(\lambda)\equiv c(c\neq 0)$。

定义 1.6 (初等变换 初等矩阵)**.** λ-矩阵的初等变换：

- 交换两行
- 行乘 c（$c\neq 0$）
- 行乘以 $\phi(\lambda)$ 加到另一行

单位矩阵施加以上操作得到的矩阵称为初等矩阵。

定理 1.7. 初等矩阵均可逆。

推论 1.8. 初等矩阵的累积可逆，可逆矩阵可写作初等矩阵的积。

定理 1.9. 初等行变换等价于左侧乘以初等矩阵，初等列变换等价于右侧乘以初等矩阵。

定义 1.10. 若 $A(\lambda)$ 可经过有限次初等变换得到 $B(\lambda)$，定义两 λ-矩阵等价，记为 $A(\lambda)\cong B(\lambda)$。

推论 1.11. $A\cong B\Leftrightarrow \exists P(\lambda),Q(\lambda),B(\lambda)=P(\lambda)A(\lambda)Q(\lambda)$

1.2 Smith 标准型与不变因子、行列式因子

定义 1.12 (Smith 标准型 不变因子)**.** 对任意 n 阶 λ-矩阵 $A(\lambda)$，总是有对角阵 $\mathrm{diag}(d_1(\lambda),d_2(\lambda),\ldots,d_r(\lambda),0,\ldots,0)$ 与之等价。其中 $d_i(\lambda)$ 都是首项系数为 1 的多项式，且 $d_i|d_{i+1}$。称这个对角阵为 $A(\lambda)$ 的 **Smith 标准型**，其中 $d_i(\lambda)$ 称为 $A(\lambda)$ 的**不变因子**。

定义 1.13 (行列式因子)**.** 对 $A(\lambda)$，定义所有 k 阶子式的最大公因式 $D_k(\lambda)$ 为 $A(\lambda)$ 的 k 阶**行列式因子**。约定其首项系数为 1。

性质 1.14. 行列式因子具有以下性质：

- $D_i(\lambda)|D_{i+1}(\lambda)$
- $D_i=1\Rightarrow D_{i-1}=D_{i-2}=\cdots=D_1=1$
- 存在 k 阶子式非零常数，则 $D_k=1$
- 存在两个 k 阶子式无公因式，则 $D_k=1$
- $D_k(\lambda)=a\det A(\lambda)$

定理 1.15. 初等变换不改变行列式因子和秩。

定理 1.16. $D_k(\lambda)=\prod_{i=1}^{k}d_i(\lambda)$

推论 1.17.

$$A\cong B\Leftrightarrow \forall k, D_{Ak}(\lambda)=D_{Bk}(\lambda)\Leftrightarrow \forall i, d_{Ai}(\lambda)=d_{Bi}(\lambda)$$

定理 1.18. *Smith* 标准型唯一。

定理 1.19.

$$A(\lambda)\text{可逆}\Leftrightarrow A(\lambda)\cong I$$

1.3 初等因子

定义 1.20. 按 $d_r(\lambda)$ 的 s 个互异复根 λ_j 分解多项式为一次多项式之积，则由整除关系，$d_i=\prod_j(\lambda-\lambda_j)^{e_{ij}}, i=1,2,\ldots,r, j=1,2,\ldots,s$，且 $0\le e_{1j}\le e_{2j}\le\cdots\le e_{rj}$。其中全部指数大于零的因子 $(\lambda-\lambda_j)^{e_{ij}}$，允许重复，称为其**初等因子**。

举例：

$$\begin{aligned}
&d_1(\lambda)=1\\
&d_2(\lambda)=\lambda(\lambda-1)\\
&d_3(\lambda)=\lambda(\lambda-1)^2(\lambda+1)^2\\
&d_4(\lambda)=\lambda^2(\lambda-1)^3(\lambda+1)^3(\lambda-2)
\end{aligned}$$

则 $A(\lambda)$ 初等因子为 $\lambda,\lambda,\lambda^2,\lambda-1,(\lambda-1)^2,(\lambda-1)^3,(\lambda+1)^2,(\lambda+1)^3,(\lambda-2)$

性质 1.21. 若不变因子相同，初等因子必然相同。若初等因子相同且秩相同，不变因子必然相同。

前者显然，后者是因为同一个根构成的 k 个因子一定分别是来自 $d_r,d_{r-1},\ldots,d_{r-k+1}$ 的因子。

推论 1.22. 矩阵等价当且仅当有相同的秩和初等因子。

定理 1.23. 分块对角矩阵的初等因子的全体为所有子矩阵的初等因子的全体。

1.4 数字矩阵相似与 λ 矩阵等价

定理 1.24. 数字矩阵 A,B 相似当且仅当 $\lambda I-A,\lambda I-B$ 等价。

定义 1.25. 对数字矩阵 A，称 $\lambda I - A$ 的不变因子、初等因子为 A 的不变因子、初等因子。

考虑对数字矩阵 A，$\lambda I - A$ 必然满秩，有

定义 1.26. 对数字矩阵 A, B，以下条件等价：

- 相似。
- 有相同初等因子。
- 有相同不变因子。
- 有相同行列式因子。

1.5 Jordan 标准型

1.5.1 求解等价的标准型

定理 1.27. 对数字矩阵 A，记初等因子为 $(\lambda - a_1)^{n_1}$，$(\lambda - a_2)^{n_2}, \cdots, (\lambda - a_s)^{n_s}$，则 $A \sim J$，其中 J 是 *Jordan* 矩阵 $J = \operatorname{diag}(J_1, J_2, \ldots, J_s)$，且 J_i 是对角元为 a_i 的 n_i 阶 *Jordan* 块。

推论 1.28. 数字矩阵 A 可以对角化当且仅当初等因子都是一次式。

通过特征值和矩阵秩也可用于求解 Jordan 标准型。每个特征值的代数重数（特征多项式根的重数）是特征值在对角线上出现的次数，即对应 Jordan 块阶数和，几何重数（对应不变子空间维数）是 Jordan 块数。但是这个方法没办法区分 Jordan 块的划分，如果存在同一特征值的 2 个 2 阶 Jordan 块和 1 个 1 阶 1 个 3 阶都会表现为代数重数 4，几何重数 2。因此这一方法仅能用于低阶矩阵 Jordan 标准型的求解。

定理 1.29.

$$\begin{bmatrix} a & b \\ & a \end{bmatrix} \sim \begin{bmatrix} a & 1 \\ & a \end{bmatrix}$$

1.5.2 求解相似变换矩阵

关于

$$P^{-1}AP = J$$

求解 P。

$$AP = PJ$$

按若尔当矩阵分块书写这一矩阵可得关于向量的方程组，

$$[AP_1, AP_2, \cdots, AP_t] = [P_1J_1, P_2J_2, \cdots, P_tJ_t]$$

在每一个 $AP_i = P_iJ_i$ 中，按列分块 $P = [X_1, X_2, \ldots, X_{n_i}]$ 有

$$\begin{aligned} AX_1 &= \lambda_i X_1 \\ AX_2 &= X_1 + \lambda_i X_2 \\ AX_3 &= X_2 + \lambda_i X_3 \\ &\cdots \\ AX_n &= X_{n_i - 1} + \lambda_i X_{n_i} \end{aligned}$$

X_1 的选取需要保证 X_2 有解，依此类推，且需要解出线性无关的特征向量。

2 内积空间、正规矩阵与 Hermite 阵

2.1 内积空间

2.1.1 内积空间

定义 2.1 (内积空间)**.** 对给定 $F = \mathbb{C}$ 上的向量空间 V，定义一个 $V \times V \to F, \alpha, \beta \mapsto (\alpha, \beta)$ 的运算 (α, β)，若这个运算满足：

- 共轭对称性 $(\alpha, \beta) = \overline{(\beta, \alpha)}$
- 对 α 线性 $(\alpha_1 + \alpha_2, \beta) = (\alpha_1, \beta) + (\alpha_2, \beta), (k\alpha, \beta) = k(\alpha, \beta)$
- 正定性 $(\alpha, \alpha) \geq 0$，且 $(\alpha, \alpha) = 0 \Leftrightarrow \alpha = \mathbf{0}$

称这个运算是一个**内积**运算。增添了内积结构的复向量空间称为**酉空间**，实数域上为**欧几里得空间**，统称内积空间。

性质 2.2.

$$\left(\sum_{i=1}^{t} k_i \alpha_i, \beta\right) = \sum_{i=1}^{t} k_i (\alpha_i, \beta)$$
$$\left(\alpha, \sum_{i=1}^{t} k_i \beta_i\right) = \sum_{i=1}^{t} \bar{k}_i (\alpha, \beta_i)$$

定义 2.3 (度量矩阵)**.** 对 n 维酉空间 V 上的一组基 $\{\alpha_i\}$，定义

$$G = \Big((\alpha_i, \alpha_j)\Big)_{n \times n}$$

为基底 $\{\alpha_i\}$ 的度量矩阵。

性质 2.4. 度量矩阵是 *Hermite* 阵。

已省略其他定义。

- 内积
- 酉空间/欧氏空间
- 长度
- 夹角
- 单位化
- 正交向量组/正交基
- 标准正交向量组/标准正交基
- Schmidt 正交化

2.1.2 酉矩阵/正交矩阵

定义 2.5. *Hermite* 算子 $^{\mathrm{H}}$ 定义为 $A^{\mathrm{H}} = \bar{A}^{\mathrm{T}} = \overline{A^{\mathrm{T}}}$，称为 A 的共轭转置矩阵。对方阵 A，若 $A = A^{\mathrm{H}}$ 定义为 *Hermite* 阵，简称 *H*-阵；若 $A = -A^{\mathrm{H}}$ 定义为反 *Hermite* 阵，简称反 *H*-阵。对应实数域上的对称矩阵和反对称矩阵。

定义 2.6. 定义满足 $UU^{\mathrm{H}} = U^{\mathrm{H}}U = I$ 的 n 阶方阵 U 为酉矩阵，记集合为 $U \in U^{n\times n}$。实数域上为正交矩阵。

定理 2.7. *Householder* 阵 $H = I - 2\alpha\alpha^{\mathrm{H}}, |\alpha| = 1$ 为酉矩阵。

性质 2.8. 若 A 为酉矩阵：

- $A^H = A^{-1} \in U^{n\times n}$
- $|\det A| = 1$
- $AB, BA \in U^{n\times n}$

定理 2.9. 给定矩阵，是酉矩阵当且仅当列向量标准正交。

定义 2.10. 定义酉空间上保持内积不变的线性变换为酉变换。欧氏空间上保持内积不变的称为正交变换。

性质 2.11. 酉变换保持向量的长度、夹角不变。

定理 2.12. 给定酉空间 V 和 V 上线性变换 σ，以下等价：

- σ 是酉变换
- $\forall\alpha \in V, \|\sigma(\alpha)\| = \|\alpha\|$
- 将 V 中的一组标准正交基变换为标准正交基
- σ 在标准正交基下的表示为酉矩阵

定义 2.13 (次酉矩阵)**.** 给定列满秩矩阵 $U \in \mathbb{C}_r^{m\times r}$，若列向量标准正交，称为**次酉矩阵**，记 $U \in U_r^{m\times r}$。

定理 2.14. 给定 $U \in \mathbb{C}^{m\times n}$，$U$ 是次酉矩阵当且仅当 $U^{\mathrm{H}}U = I_r$。

2.1.3 幂等矩阵

定义 2.15. 对方阵 A，若 $A^2 = A$，称 A 幂等。

性质 2.16. 给定幂等矩阵 A，有：

- $A^{\mathrm{T}}, A^{\mathrm{H}}, I - A$ 幂等
- $A(I - A) = (I - A)A = 0$
- $\mathrm{N}(A) = \mathrm{R}(I - A)$
- $Ax = x \Leftrightarrow x \in R(A)$
- $\mathbb{C}^n = \mathrm{R}(A) \oplus \mathrm{N}(A)$

定理 2.17. $A \in \mathbb{C}_r^{n\times n}$，$A$ 幂等 $\Leftrightarrow A \sim \mathrm{diag}\left(I_r, \mathbf{0}_{n-r}\right)$

推论 2.18. A 幂等 $\Rightarrow \mathrm{r}(A) = \mathrm{tr}(A)$

定理 2.19. 方阵 A，A 幂等，当且仅当，存在次酉矩阵 $U \in U_r^{n\times r}$，使得 $A = UU^{\mathrm{H}}$

2.2 正规矩阵

2.2.1 Schur 引理与 Schur 不等式

定理 2.20 (Schur 引理)**.** 任意一个复方阵 $A \in \mathbb{C}^{n\times n}$ 酉相似于一个上三角阵。

可使用数学归纳法，根据基的扩充定理和施密特正交化证明。

定义 2.21. 给定 n 阶方阵 A，有

$$U^{\mathrm{H}}AU = T = \begin{bmatrix} \lambda_1 & \cdots & * \\ & \ddots & \vdots \\ & & \lambda_n \end{bmatrix}$$

则有

$$A = UTU^{\mathrm{H}}$$

称为 A 的 *Schur* 分解。

定理 2.22 (Schur 不等式)**.** 对给定 n 阶方阵 $A = \left(a_{ij}\right)_{n\times n}$，特征值为 $\lambda_1, \lambda_2, \ldots, \lambda_n$，有

$$\sum_i |\lambda_i|^2 \le \sum_i \sum_j |a_{ij}|^2$$

2.2.2 正规矩阵

定义 2.23 (正规矩阵)**.** 对方阵 A，若 $AA^{\mathrm{H}} = A^{\mathrm{H}}A$，称为**正规矩阵**。实数域上称实正规矩阵。

引理 2.24. 与正规矩阵酉相似的矩阵一定是正规矩阵。

引理 2.25. 同时是正规矩阵、三角形矩阵，则必然是对角矩阵。

定理 2.26. 正规矩阵的充要条件是酉相似于对角矩阵。

推论 2.27. 正规矩阵有 n 个线性无关特征向量，且对应不同特征值的特征向量互相正交。

定理 2.28. 对正规矩阵：

- 是 *Hermite* 矩阵，当且仅当，特征值全部落在实轴上。
- 是反 *Hermite* 矩阵，当且仅当，特征值全部落在虚轴上。
- 是酉矩阵，当且仅当，特征值全部落在单位圆上。

2.3 Hermite 二次型

2.3.1 Hermite 矩阵

定理 2.29. 对 *Hermite* 矩阵 A，有：

- $A^{\mathrm{H}}A, AA^{\mathrm{H}}$ 都是 *Hermite* 矩阵。
- $A + A^{\mathrm{H}}$ 是 *Hermite* 矩阵，但 $A - A^{\mathrm{H}}$ 是反 *Hermite* 矩阵。

- A^k 是 *Hermite* 矩阵，若可逆，A^{-1} 也是 *Hermite* 矩阵。
- $\lambda \in \mathbb{R}$，则 λA 是 *Hermite* 矩阵，$\lambda i A$ 是反 *Hermite* 矩阵。

对 *Hermite* 矩阵 A, B，有：

- 实系数线性组合也是 *Hermite* 矩阵。
- 乘积 AB 是 *Hermite* 矩阵当且仅当可交换。

定理 2.30. 对方阵 $A \in \mathbb{C}^n$，以下三项等价：

- A 是 *Hermite* 矩阵
- $\forall X \in \mathbb{C}^n$，$X^H AX$ 是实数
- $\forall B \in \mathbb{C}^{n\times n}$，$B^H AB$ 是 *Hermite* 矩阵

2.3.2 Hermite 二次型

定义 2.31 (Hermite 二次型)**.** n 个复变量 $x_1, x_2, \ldots, x_n$，二次齐次多项式

$$f(x_1, x_2, \ldots, x_n) = \sum_{i,j} a_{ij}\overline{x_i}x_j$$

其中 $a_{ij} = \overline{a_{ji}}$，则称为 Hermite 二次型。

类似实数域上二次型的记号，记向量 $X = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$，矩阵 $A = \left(a_{ij}\right)_{m\times n}$，则

$$f(x_1, x_2, \ldots, x_n) = X^{\mathrm{H}}AX$$

其中矩阵 A 称为 Hermite 二次型的的矩阵，是一个 *Hermite* 阵，A 的秩为 Hermite 二次型的秩。

可以对 Hermite 二次型进行可逆线性替换，即使用 $X = CY$，其中 C 为可逆方阵，代替，得到

$$X^{\mathrm{H}}AX = Y^{\mathrm{H}}C^{\mathrm{H}}ACY = Y^{\mathrm{H}}BY$$

定义 2.32. 矩阵为对角矩阵的 *Hermite* 二次型

$$f() = \lambda_1\overline{x_1}x_1 + \lambda_2\overline{x_2}x_2 + \cdots + \lambda_n\overline{x_n}x_n$$

称为二次型的标准型。

定理 2.33. 对给定 *Hermite* 二次型，总是存在酉线性替换将其转化为标准型。存在举例：标准型的对角矩阵的对角线元素为原矩阵的特征值。这一线性替换的分解就是原矩阵的特征值分解。但线性替换是合同不是相似，不要求酉线性替换的情况下，变换结果对角矩阵不唯一。

定义 2.34. 矩阵为对角矩阵，且对角线上元素具有前 s 个为 1，然后 $r-s$ 个元素为 -1，其余 $n-r$ 个为 0 的形式

$$f() = \overline{x_1}x_1 + \cdots + \overline{x_s}x_s - \overline{x_{s+1}}x_{s+1} - \cdots - \overline{x_r}x_r$$

的二次型称为二次型的规范型。

定理 2.35. 对给定 *Hermite* 二次型，总是存在唯一对应的规范型。矩阵的标准型总是有固定的正系数、负系数、零系数项（正负惯性指数），将其按这个顺序重排并将其绝对值平方根转换到两侧线性替换中，则得到规范型矩阵。

2.3.3 正定 Hermite 二次型

由 Hermite 矩阵的性质，对任意 X 都有 $X^{\mathrm{H}}AX$ 为实数，因此 Hermite 二次型结果总是实数。

定义 2.36 (正定 半正定)**.** 对 *Hermite* 矩阵 A，若 $\forall X \in \mathbb{C}^n \backslash \{\mathbf{0}\}, X^{\mathrm{H}}AX > 0$，称 A 正定；若 $\forall X \in \mathbb{C}^n, X^{\mathrm{H}}AX \geq 0$，称 A 半正定。

定理 2.37. 对给定 *Hermite* 二次型 $f(X) = X^{\mathrm{H}}AX$，以下各项等价：

- $f(X)$ 正定
- 对任意可逆 P，$P^{\mathrm{H}}AP$ 正定
- A 的 n 个特征值均为正数（正惯性指数为 n）
- 存在可逆 P，$P^{\mathrm{H}}AP = I$（规范型）
- 存在可逆 Q，$A = Q^{\mathrm{H}}Q$（上一条的推论）
- 存在正线上三角 R，$A = R^{\mathrm{H}}R$

定理 2.38. 对给定 *Hermite* 二次型 $f(X) = X^{\mathrm{H}}AX$，以下各项等价：

- $f(X)$ 半正定
- 对任意可逆 P，$P^{\mathrm{H}}AP$ 半正定
- A 的 n 个特征值均为非负实数（负惯性指数为 0）
- 存在可逆 P，$P^{\mathrm{H}}AP = \mathrm{diag}(I_r, \mathbf{0})$（规范型）
- 存在秩为 r 的 n 阶方阵 Q，$A = Q^{\mathrm{H}}Q$（上一条推论）

定理 2.39. A 为正定（半正定）*Hermite* 矩阵，则存在唯一的正定（半正定）Q，满足 $A = Q^2$。记为 $Q = A^{\frac{1}{2}}$。

2.3.4 Hermite 矩阵偶合同标准型

定理 2.40. 对给定二次型 A, B，若 B 正定，存在非退化线性替换 $X = PY$ 同时满足：

$$P^{\mathrm{H}}AP = \sum_{i=1}^{n} \lambda_i \overline{y_i} y_i$$

$$P^{\mathrm{H}}BP = \sum_{i=1}^{n} \overline{y_i} y_i$$

其中 λ_i 是 $|\lambda B - A| = 0$ 的根，且均为实数。

先对角化 B 并化为规范型，即 $P_1^{\mathrm{H}}BP_1 = I$，其中 P_1 可逆，因此 $P_1^{\mathrm{H}}AP_1$ 也是正定矩阵，然后酉相似对角化 $P_1^{\mathrm{H}}AP_1$，得到 $P_2^{\mathrm{H}}P_1^{\mathrm{H}}AP_1P_2 = \Lambda$，其中 P_2 为酉矩阵。此时有 $P_2^{\mathrm{H}}P_1^{\mathrm{H}}BP_1P_2 = P_2^{\mathrm{H}}IP_2 = I$，因此可逆矩阵 $P = P_1P_2$ 即为符合条件的方阵。

定义 2.41. 对给定 n 阶 *Hermite* 矩阵 A, B，其中 B 正定，则使方程

$$AX = \lambda BX$$

有非零解的 λ 的充要条件是

$$|\lambda B - A| = 0$$

类比与特征方程、特征值与特征向量，称这个关于 λ 的方程为矩阵 A 相对于矩阵 B 的特征方程，根 $\lambda_1, \lambda_2, \ldots, \lambda_n$ 为 A 相对于 B 的广义特征值，关于 $\lambda = \lambda_i$ 时 X 的方程的解向量为对应 λ_i 的广义特征向量。

性质 2.42. 给定 *Hermite* 矩阵 A, B，其中 B 正定，则 A 相对于 B，总是有 n 个广义特征值，且分别对应 n 个线性无关的广义特征向量。适当选取广义特征向量 $X_1, X_2, \ldots, X_n$，可满足 $X_i^{\mathrm{H}} B X_j = \delta_{ij}, X_i^{\mathrm{H}} A X_j = \lambda_j \delta_{ij}$

3 矩阵的分解

3.1 满秩分解

定义 3.1 (满秩分解)**.** 对矩阵 $A \in \mathbb{C}_r^{m\times n}, \exists B \in \mathbb{C}_r^{m\times r}, C \in \mathbb{C}_r^{r\times n}$，使得 $A = BC$，称为矩阵的满秩分解。

求解方法：

1. 将矩阵 A 化为行最简形/rref 得到 R，获得主元位置。
2. A 中主元所在列构成 B。
3. R 中主元所在行构成 C。

定理 3.2. 矩阵的满秩分解不唯一，若 $A = BC = B_1C_1$ 均为 A 的满秩分解，有：$\exists \theta \in \mathbb{C}_n^{n\times n}, B = B_1\theta, C = \theta^{-1}C_1$，且具有不变量 $C^{\mathrm{H}}(CC^{\mathrm{H}})^{-1}(B^{\mathrm{H}}B)^{-1}B^{\mathrm{H}}$。

3.2 正交三角分解

定义 3.3 (正交三角分解)**.** 给定可逆方阵 $A \in \mathbb{C}_n^{n\times n}$，存在唯一的酉矩阵 U 和正线上三角阵 R 使得 $A = UR$；对于转置的情况，存在唯一的酉矩阵 U_1 和正线下三角阵 R_1 使得 $A = R_1U_1$

定理 3.4. 给定矩阵 $A \in \mathbb{C}_r^{m\times r}$，存在唯一的次酉矩阵 $U \in U^{m\times r}$ 和 r 阶正线上三角矩阵 R 使得 $A = UR$。

求法：

1. 将 A 的列向量组施密特正交化得到 U。
2. $A = UR \Rightarrow R = U^{\mathrm{H}}A$

推论 3.5. 对给定矩阵 $A \in \mathbb{C}_r^{m\times n}$，存在一个分解 $A = U_1R_1L_2U_2$，其中 $U_1 \in U^{m\times r}, U_2 \in U^{r\times n}$，且 R_1, L_2 为 r 阶上三角阵和下三角阵。

3.3 奇异值分解

引理 3.6. 对任意矩阵 A，有 $r(AA^{\mathrm{H}}) = r(A^{\mathrm{H}}A) = r(A)$

引理 3.7. 对任意矩阵 A，有 AA^{H} 和 $A^{\mathrm{H}}A$ 都是半正定 *Hermite* 阵。

定理 3.8. 对任意矩阵 $A \in \mathbb{C}_r^{m\times n}$，记 AA^{H} 的特征值从大到小排列为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_m = 0$，$A^{\mathrm{H}}A$ 的特征值从大到小排列为 $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_r > \mu_{r+1} = \cdots = \mu_n = 0$，则有 $\lambda_i = \mu_i > 0, i = 1, 2, \ldots, r$。

定理 3.9. 对正规矩阵 A，AA^{H} 和 $A^{\mathrm{H}}A$ 的特征值是 A 的特征值的模长平方。

定义 3.10 (奇异值)**.** 对矩阵 $A \in \mathbb{C}_r^{m\times n}$，记 AA^{H} 的非零特征值从大到小排列为 $\lambda_1, \lambda_2, \ldots, \lambda_r$，称 $\sigma_i = \sqrt{\lambda_i} = \sqrt{\mu_i} > 0$ 为矩阵的正奇异值，简称奇异值。

定理 3.11. 对矩阵 $A \in \mathbb{C}_r^{m\times n}$，记奇异值为 $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r$，则存在 $U \in U^{m\times m}, V \in U^{n\times n}$，使得 $U^{\mathrm{H}}AV = \begin{bmatrix} \Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$，其中 $\Delta = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$ 为奇异值由大到小排列构成的对角阵，填充 0 后的矩阵为 $m \times n$ 矩阵。

定义 3.12 (奇异值分解)**.** 对矩阵 A，定义 $A = U \begin{bmatrix} \Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} V^{\mathrm{H}}$ 为矩阵的奇异值分解式。

考虑将 U, V 的前 r 列与后 $m-r$ 和 $n-r$ 列分块，则得到 $A = U_1\Delta V_1$，即：

推论 3.13. 对矩阵 $A \in \mathbb{C}_r^{m\times n}$，记奇异值为 $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r$，存在次酉矩阵 $U \in U^{m\times r}, V \in U^{r\times n}$，使得 $A = U\Sigma V^{\mathrm{H}}$

求解方法：

1. 计算 AA^{H}（或 $A^{\mathrm{H}}A$）。
2. 特征值分解 AA^{H}（或 $A^{\mathrm{H}}A$），得到特征值 $\lambda_1, \lambda_2, \ldots, \lambda_r$，以及对应标准正交特征向量组，分别构成对角矩阵 Δ 和酉矩阵 U（或 V）。
3. 将 U 分块，记对应非零特征值的前 r 列为 U_1，对应特征值 0 的后 $m-r$ 列为 U_2。同样记 V 的前 r 列与后 $n-r$ 列为 V_1, V_2。
4. $A = U_1\Delta V_1^{\mathrm{H}} \Rightarrow U_1 = AV_1\Delta^{-1}V_1, = A^{\mathrm{H}}U_1\Delta^{-1}$。求出另一个矩阵前 r 列。
5. 选取使得 V（或 U）是酉矩阵的 V_2（或 U_2）。可通过选取线性无关向量进行施密特正交化完成。

3.4 极分解

定义 3.14 (极分解)**.** 对可逆矩阵 $A \in \mathbb{C}_n^{n\times n}$，存在 $U \in U^{n\times n}$ 与正定 H_1, H_2 满足 $A = H_1U = UH_2$。分解唯一且 $AA^{\mathrm{H}} = H_1^2, A^{\mathrm{H}}A = H_2^2$，称为矩阵的极分解表达式。对一般矩阵 $A \in \mathbb{C}_r^{n\times n}$，存在 $U \in U^{n\times n}$ 与半正定 H_1, H_2 满足以上各式。

定理 3.15 (使用奇异值分解表达极分解)**.**

$$A = U\Delta V = \underbrace{U\Delta U^{\mathrm{H}}}_{H_1}\underbrace{UV}_{U} = \underbrace{UV}_{U}\underbrace{V^{\mathrm{H}}\Delta V}_{H_2}$$

定理 3.16. 正规矩阵当且仅当极分解为 $A = HU = UH$。且此时 $A^{\mathrm{H}}A = H^2$

3.5 谱分解

定义 3.17 (谱分解). 对正规矩阵 $A \in \mathbb{C}^{n\times n}$，将其特征值分解后按特征值矩阵的对角元素分块，得到

$$A = \lambda_1\alpha_1\alpha_1^{\mathrm{H}} + \lambda_2\alpha_2\alpha_2^{\mathrm{H}} + \cdots + \lambda_n\alpha_n\alpha_n^{\mathrm{H}}$$

称为矩阵 A 的谱分解表达式。

将各项按特征值合并，得到 $A = \sum\limits_i \lambda_i \sum\limits_j \alpha_{ij}\alpha_{ij}^{\mathrm{H}} = \sum\limits_i \lambda_i G_i$，其中 $G_i = \sigma_j\alpha_j\alpha_j^{\mathrm{H}}$，是正交的投影矩阵的和，也有投影矩阵的性质（Hermite、幂等），同时 G_i 间两两正交，故称为正交投影矩阵。

定理 3.18. 对矩阵 A，有 n 个互异特征值 $\lambda_1, \lambda_2, \ldots, \lambda_s$，且其中第 i 个特征值的重数为 n_i，则有：A 正交当且仅当存在 s 个 n 阶矩阵 $G_1, G_2, \ldots, G_s$ 满足：

- $A = \sum\limits_i \lambda_i G_i$
- $G = G^{\mathrm{H}}, G = G^2$
- $G_iG_k = 0 (i \neq k)$
- $\sum\limits_i G_i = I$
- $\mathrm{r}(G_i) = n_i$

此时后者仅存在唯一一组。

求解步骤：对给定可对角化矩阵：

- 求出全部互异特征值 $\lambda_1, \lambda_2, \ldots, \lambda_s$。
- 求解每个特征值 λ_i 的线性无关特征向量 $\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in_i}$，得到变换矩阵 P。
- 对 $P = [\alpha_1, \alpha_2, \ldots, \alpha_n]$，将 $P^{-\mathrm{T}}$ 按行分块记为 $[\beta_1, \beta_2, \ldots, \beta_n]$。
- 令 $G_i = \sum\limits_j \alpha_{ij}\beta_{ij}^{\mathrm{T}}$
- 得到 $A = \sum\limits_i \lambda_i G_i$

4 向量与矩阵范数

4.1 向量范数

4.1.1 定义

定义 4.1 (范数). 给定 F 上的线性空间 V，对函数 $\|\cdot\| : V \to \mathbb{R}, \alpha \mapsto \|\alpha\|$，若满足：

- 非负：$\forall\alpha \in V, \|\alpha\| \geq 0$，且 $\|\alpha\| = 0 \Leftrightarrow \alpha = \mathbf{0}$
- 齐次：$\|k\alpha\| = |k|\|\alpha\|, \forall k \in F, \alpha \in V$
- 三角不等式：$\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|, \forall\alpha, \beta \in V$

则称 $\|\alpha\|$ 是 α 的一个范数。

4.1.2 常见向量范数

引理 4.2 (Hölder 不等式).

$$\sum_{i=1}^n |a_ib_i| \leq \left(\sum_{i=1}^n |a_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n |b_i|^q\right)^{\frac{1}{q}}$$

其中 $p > 1, q > 1, 1/p + 1/q = 1$

先证明 $uv \leq \frac{u^p}{p} + \frac{v^q}{q}$，然后令 $u = \frac{a_k}{a}, v = \frac{b_k}{b}$，求和后令 $a = \left(\sum\limits_k a_k^p\right)^{\frac{1}{p}}, b = \left(\sum\limits_k b_k^p\right)^{\frac{1}{q}}$ 消去 a 和 b，可得上式。

引理 4.3 (Minkowski 不等式).

$$\left(\sum_{i=1}^n |a_i + b_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^n |a_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |b_i|^p\right)^{\frac{1}{p}}$$

由于 $|\cdot|^p = |\cdot||\cdot|^{p-1} = |\cdot||\cdot|^{p\cdot\frac{p-1}{p}} = |\cdot|(|\cdot|^p)^{\frac{p-1}{p}}$，累和后按上面的展开为

$$\begin{aligned}
&\sum_i |a_i + b_i|^p \\
\leq&\sum_i |a_i + b_i||a_i + b_i|^{p-1} \\
\leq&\sum_i (|a_i| + |b_i|)|a_i + b_i|^{p-1} (\text{三角不等式}) \\
=&\sum_i |a_i||a_i + b_i|^{p-1} + \sum_i |b_i||a_i + b_i|^{p-1} \\
\leq&\left(\sum_j |a_j|^p\right)^{\frac{1}{p}} \left(\sum_i |a_i + b_i|^{(p-1)q}\right)^{\frac{1}{q}} \\
&+ \left(\sum_j |b_j|^p\right)^{\frac{1}{p}} \left(\sum_i |a_i + b_i|^{(p-1)q}\right)^{\frac{1}{q}} \\
&(\text{分别 Hölder 不等式}) \\
=&\left[\left(\sum_j |a_j|^p\right)^{\frac{1}{p}} + \left(\sum_j |b_j|^p\right)^{\frac{1}{p}}\right] \left(\sum_i |a_i + b_i|^p\right)^{1-\frac{1}{p}}
\end{aligned}$$

约去右侧第二个因式，得到

$$\left(\sum_i |a_i + b_i|^p\right)^{\frac{1}{p}} = \left(\sum_j |a_j|^p\right)^{\frac{1}{p}} + \left(\sum_j |b_j|^p\right)^{\frac{1}{p}}$$

定义 4.4 (p-范数). 定义向量 α 的 p-范数为：

$$\|\alpha\|_p = (\sum_{i=1}^n |\alpha_i|^p)^{\frac{1}{p}}$$

非负、齐次显然成立，三角不等式由 Minkowski 不等式保证。

定义 4.5 (1-范数 2-范数 ∞-范数). 以下是 p-范数的常见形式：

- 向量 1-范数：令 $p=1$，得

$$\|\alpha\|_1=\sum_{i=1}^{n}|a_i|$$

- 向量 2-范数：令 $p=2$，得

$$\|\alpha\|_2=\sqrt{\sum_{i=1}^{n}|a_i|^2}$$

，与向量长度相同，也称欧氏范数

- 向量 ∞-范数：令 $p\to\infty$，得

$$\|\alpha\|_\infty=\max_{i=1}^{n}|a_i|$$

4.1.3 向量范数的等价

定义 4.6 (向量范数的等价). 对两个范数 $\|\cdot\|_a,\|\cdot\|_b$，若 $\exists d_1,d_2\in\mathbb{R}^+$，使得 $\forall\alpha,d_1\|\alpha\|_b\le\|\alpha\|_a\le d_2\|\alpha\|_b$，则称两向量范数等价，此时也有 $\frac{1}{d_2}\|\alpha\|_a\le\|\alpha\|_b\le\frac{1}{d_1}\|\alpha\|_a$ 成立。

定理 4.7. 有限维线性空间 V 上任意两向量范数等价。

4.1.4 向量范数的构造

定理 4.8. 对给定线性空间 $\mathbb{C}^m$ 上的范数 $\|\cdot\|_b$，若有列满秩矩阵 $A\in\mathbb{C}_n^{m\times n}$，则 $\|\alpha\|_a=\|A\alpha\|_b$ 是线性空间 $\mathbb{C}^n$ 上的范数。

这是将一个线性空间变换到一个更大的线性空间里使用后者的范数作为自己的范数。

4.2 矩阵范数

4.2.1 定义与 Frobenious 范数

定义 4.9 (矩阵范数). 对矩阵函数 $\|\cdot\|:\mathbb{C}^{m\times n}\to\mathbb{R},A\mapsto\|A\|$，若满足：

- 非负：$\forall A\in\mathbb{C}^{m\times n},\|A\|\ge 0$，且 $\|A\|=0\Leftrightarrow A=\mathbf{0}$
- 齐次：$\|kA\|=|k|\|A\|,\forall k\in F,A\in\mathbb{C}^{m\times n}$
- 三角不等式：$\|A+B\|\le\|A\|+\|B\|,\forall A,B\in\mathbb{C}^{m\times n}$
- 与乘法相容：$\|AB\|\le\|A\|\|B\|,\forall A,B\in\mathbb{C}^{m\times n}$

则称 $\|A\|$ 是矩阵 A 的一个范数。

简单地想法是把矩阵看作线性空间，然后考虑满足乘法相容的性质，由此我们得到了 1-范数，2-范数，∞-范数到矩阵范数的推广：

- 对矩阵 $A=\left(a_{ij}\right)_{m\times n}$，$\|A\|=\sum\limits_{i,j}|a_{ij}|$ 是矩阵的范数。
- 对矩阵 $A=\left(a_{ij}\right)_{m\times n}$，$\|A\|=\sqrt{\sum\limits_{i,j}|a_{ij}|^2}$ 是矩阵的范数。一般称为 Frobenious 范数，记作 $\|A\|_F$。
- 对方阵 $A=\left(a_{ij}\right)_{n\times n}$，$\|A\|=n\max\limits_{i,j}|a_{ij}|$ 是矩阵的范数。

定义 4.10. 与向量范数类似地, 对两个矩阵范数 $\|\cdot\|_a,\|\cdot\|_b$，若 $\exists d_1,d_2\in\mathbb{R}^+$，使得 $\forall\alpha,d_1\|\alpha\|_b\le\|\alpha\|_a\le d_2\|\alpha\|_b$，则称两矩阵范数等价，此时也有 $\frac{1}{d_2}\|\alpha\|_a\le\|\alpha\|_b\le\frac{1}{d_1}\|\alpha\|_a$ 成立。

定理 4.11. 对两种不同的矩阵范数 $\|A\|_\alpha,\|B\|_\beta$，总是存在 $d_1,d_2\in\mathbb{R}^+$ 使其等价。

定义 4.12 (Frobenious 范数).

$$\|A\|_F=\sqrt{\sum_{i,j}|a_{ij}|^2}=\sqrt{\mathrm{tr}(AA^{\mathrm{H}})}=\sqrt{\mathrm{tr}(A^{\mathrm{H}}A)}$$

性质 4.13. F 范数 $\|\cdot\|_F$ 具有以下性质：

- 将矩阵 A 的列向量组记为 $\alpha_1,\alpha_2,\ldots,\alpha_n$，则 $\|A\|_F^2=\sum\limits_{i}\|\alpha_i\|_2^2$
- $\|A\|_F^2=\mathrm{tr}(AA^{\mathrm{H}})=\sum_i\lambda_i(AA^{\mathrm{H}})$，其中 $\lambda_i(A)$ 指 A 的第 i 个特征值
- 共轭转置、乘以酉矩阵都不改变矩阵的 F 范数。

4.2.2 诱导范数

定义 4.14 (矩阵范数和向量范数相容). 对向量范数 $\|X\|_\alpha$ 与矩阵范数 $\|A\|_\beta$，若对任意 X,A 都有：$\|AX\|_\alpha\le\|A\|_\beta\|X\|_\alpha$ 则称矩阵范数 $\|A\|_\beta$ 与向量范数 $\|X\|_\alpha$ 是相容的。

定理 4.15. 矩阵 *Frobenious* 范数与向量 *2*-范数相容。

定义 4.16 (诱导范数). 对向量范数 $\|X\|_\alpha$，定义

$$\|A\|_i=\max_{X\neq\mathbf{0}}\frac{\|AX\|_\alpha}{\|X\|_\alpha}$$

则 $\|A\|_i$ 是一个矩阵范数，且与 $\|X\|_\alpha$ 相容。称这个矩阵范数为由向量范数 $\|X\|_\alpha$ 所诱导的**诱导范数**或**算子范数**。

对任意一个诱导范数，总有 $\|I\|=1$。但不是所有的范数都是诱导范数。

定义 4.17 (矩阵 p-范数). 由向量的 p-范数诱导的矩阵范数称为矩阵的 p-范数。

定义 4.18 (矩阵 1-范数 矩阵 2-范数 矩阵 ∞-范数). 以下是矩阵 p-范数的常见形式：

- 矩阵 1-范数：由 $\|A\|_1=\max\limits_{X\neq\mathbf{0}}\frac{\|AX\|_1}{\|X\|_1}$，考虑 $\|AX\|_1\le\sum\limits_{i}|x_i|\|\alpha_i\|_1\le\sum\limits_{i}|x_i|\cdot\max\limits_{i}\|\alpha_i\|_1$，等号取在 $\|x\|$ 是标准基向量时，得

$$\|A\|_1=\max_{j}\sum_{i}|a_{ij}|$$

为列绝对值之和的最大值，称为**列和范数**。最大值取在 $\|X\|$ 是标准基向量时。

- 矩阵 2-范数：由 $\|A\|_2=\max\limits_{X\neq\mathbf{0}}\frac{\|AX\|_2}{\|X\|_2}$，考虑 $\frac{\|AX\|_2}{\|X\|_2}=\frac{X^{\mathrm{H}}A^{\mathrm{H}}AX}{X^{\mathrm{H}}X}$，然后对角化 $A^{\mathrm{H}}A$，则上式 $=\frac{Y^{\mathrm{H}}\Lambda Y}{Y^{\mathrm{H}}Y}=\frac{\sum\limits_i\lambda_i y_i^{\mathrm{H}}y_i}{\sum\limits_i y_i^{\mathrm{H}}y_i}$，得
$$\|A\|_2=\max_i\sqrt{\lambda_i(A^{\mathrm{H}}A)}$$
为 $A^{\mathrm{H}}A$ 最大特征值平方根，即 A 的最大奇异值，称为**谱范数**。最大值取在 $\|X\|$ 是 $A^{\mathrm{H}}A$ 对应这一特征值的特征向量时。
- 矩阵 ∞-范数：由 $\|A\|_\infty=\max\limits_{X\neq\mathbf{0}}\frac{\|AX\|_\infty}{\|X\|_\infty}$，考虑 $\|AX\|_\infty=\max\limits_i\left|\sum\limits_j a_{ij}x_j\right|\leq\max\limits_i\sum\limits_j|a_{ij}||x_j|$，等号取在 x_j 和 a_{ij} 成比例时，然后上式 $\leq\max\limits_i\sum\limits_j|\alpha_{ij}|\cdot\max\limits_k|x_k|=\|X\|_\infty\max\limits_i\sum\limits_j|a_{ij}|$，得
$$\|A\|_\infty=\max_i\sum_j|a_{ij}|$$
为行绝对值之和的最大值，称为**行和范数**。最大值取在 $\|X\|$ 中所有项按 A 的最大行的符号选择 ±1 时。

定理 4.19. $\|AA^{\mathrm{H}}\|_2=\|A^{\mathrm{H}}A\|_2=\|A\|_2^2$

原因是 AA^{H} 和 $A^{\mathrm{H}}A$ 是 Hermite 阵，$(A^{\mathrm{H}}A)^{\mathrm{H}}A^{\mathrm{H}}A=(A^{\mathrm{H}}A)^2$，而同时 A^2 的特征值就是 A 的特征值平方。

与由向量范数诱导矩阵范数相反的方向：

定理 4.20. 对给定矩阵范数 $\|A\|_*$，存在向量范数 $\|X\|$ 使得 $\|AX\|\leq\|A\|_*\|X\|$。

可验证对任意 $\alpha\neq\mathbf{0}$，$\|X\|=\|X\alpha^{\mathrm{H}}\|_*$ 总符合向量范数的定义，且总满足上式。

4.2.3 矩阵谱半径

定义 4.21 (谱半径)**.** 对方阵 A，记其特征值为 $\lambda_1,\lambda_2,\ldots,\lambda_n$，称
$$\rho(A)=\max_i|\lambda_i|$$
为矩阵 A 的**谱半径**。

定理 4.22. 对任意矩阵范数 $\|\cdot\|$，对任意方阵 A 总有
$$\rho(A)\leq\|A\|$$

由于特征值定义，$AX=\lambda X$，取范数则有 $|\lambda|\|X\|=\|AX\|\leq\|A\|\|X\|$，因此 $|\lambda|\leq\|A\|$，则特征值最大值也比右侧小。

定理 4.23. 对正规矩阵 A，有 $\rho(A)=\|A\|_2$

这是因为 $A=U\Lambda U^{\mathrm{H}}$ 时，有 $A^{\mathrm{H}}A=U\Lambda^H U^{\mathrm{H}}U\Lambda U^{\mathrm{H}}=U\overline{\Lambda}\Lambda U^H$，因此 $\lambda_i(A^{\mathrm{H}}A)=\overline{\lambda_i(A)}\lambda_i(A)=|\lambda_i(A)|^2$，两侧加上 max 得到 $\|A\|_2^2=\rho^2(A)$。

定理 4.24.
$$\|A\|_2^2\leq\|A\|_1\|A\|_\infty$$

考虑 $A^{\mathrm{H}}A$ 取无穷范数，
$$\rho(A^{\mathrm{H}}A)\leq\|A^{\mathrm{H}}A\|_\infty=\|A^{\mathrm{H}}\|_\infty\|A\|_\infty=\|A\|_1\|A\|_\infty$$

4.3 矩阵序列、极限、级数 *

5 矩阵函数

5.1 矩阵的多项式表示与极小多项式

定义 5.1. 对多项式 $f(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$ 和方阵 A，称 $f(A)=a_nA^n+a_{n-1}A^{n-1}+\cdots+a_1A+a_0I$ 为**矩阵多项式**。

表示成 Jordan 标准型，则

定义 5.2. 若 $A=PJP^{-1}$，则 $f(A)=Pf(J)P^{-1}=P\,\mathrm{diag}(f(J_1),f(J_2),\ldots,f(J_s))P^{-1}$，称为矩阵多项式的 *Jordan* 表示。

由 Jordan 块的特点，
$$J_i=\begin{bmatrix}\lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{bmatrix}_{d_i\times d_i}$$
$$J_i^k=\begin{bmatrix}\lambda_i^k & \binom{k}{1}\lambda_i^{k-1} & \cdots & \binom{k}{d_i-1}\lambda_i^{k-d_i+1}\\ & \lambda_i^k & \ddots & \vdots\\ & & \ddots & \binom{k}{1}\lambda_i^{k-1}\\ & & & \lambda_i^k\end{bmatrix}$$
$$f(J_i)=\begin{bmatrix}f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \cdots & \frac{1}{(d_i-1)!}f^{(d_i-1)}(\lambda_i)\\ & f(\lambda_i) & \ddots & \vdots\\ & & \ddots & \frac{1}{1!}f'(\lambda_i)\\ & & & f(\lambda_i)\end{bmatrix}$$

因此 $f(A)$ 可以被进一步简化用于运算。

定义 5.3 (零化多项式)**.** 给定方阵 A 和多项式 $f(x)$，若 $f(A)=\mathbf{0}$，称 $f(x)$ 是 A 的**零化多项式**。

定理 5.4 (Hamilton-Cayley 定理)**.** 对方阵 A，特征多项式 $f(\lambda)=\det(\lambda I-A)$ 是其零化多项式。

定义 5.5 (最小多项式)**.** 对方阵 A，其全部零化多项式中次数最低且首项系数为 1 的称为其**最小多项式**，通常记为 $m(\lambda)$。

性质 5.6. 已知 $A\in\mathbb{C}^{n\times n}$，关于 A 的最小多项式：

- 最小多项式唯一

- 最小多项式整除任一零化多项式
- 相似矩阵有相同最小多项式

求解方法：

1. 将矩阵化为 Jordan 标准型。
2. 对 Jordan 块：$m(\lambda) = (\lambda - \lambda_i)^{d_i}$
3. 对分块对角矩阵：最小多项式为各分块最小多项式的最小公倍式。

5.2 矩阵函数 *

定义 5.7. 类似多项式逼近的方式定义：

- $\exp A = \sum\limits_{n=0}^{\infty} \frac{1}{n!} A^n = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \dots$
- $\sin A = \sum\limits_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} A^(2n+1) = A - \frac{1}{3!}A^3 + \frac{1}{5!}A^5 - \dots$
- $\cos A = \sum\limits_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} A^(2n) = I - \frac{1}{2!}A^2 + \frac{1}{4!}A^4 - \dots$

6 函数矩阵和矩阵微分方程 *

7 广义逆矩阵

7.1 广义逆矩阵

定义 7.1 (广义逆矩阵). 对数域 F 上的矩阵 $A \in F_r^{m\times n}$，矩阵方程 $AXA = A$ 总有解。考虑将 A 变为标准型的分解 $A = P\operatorname{diag}(I_r, \mathbf{0})Q$，方程的通解可以表示为

$$X = Q^{-1}\begin{bmatrix} I_r & B_{(m-r)\times r} \\ C_{r\times(n-r)} & D_{(m-r)\times(n-r)} \end{bmatrix} P^{-1}$$

其中 B, C, D 是满足标注大小的任意矩阵。这样的矩阵称为 A 的**广义逆矩阵**，简称**广义逆**，记为 A^-。

由定义中 X 的表达形式可知，若 $m = n = r$，则 $X = A^{-1}$。

定理 7.2 (非齐次方程组相容性定理). 非齐次线性方程组 $AX = \beta$ 有解当且仅当 $\beta = AA^-\beta$。

定理 7.3 (非齐次线性方程组解的结构定理). 非齐次线性方程组 $AX = \beta$ 的通解为 $X = A^-\beta$，其中 A^- 取遍 A 的所有广义逆。

定理 7.4 (齐次线性方程组解的结构定理). 齐次线性方程组 $AX = \mathbf{0}$ 的通解为 $X = (I_n - A^-A)Z$，其中 A^- 是 A 任意给定的广义逆，Z 取遍任意一个 n 维列向量。

推论 7.5. 非齐次线性方程组 $AX = \beta$ 有解，则通解为 $X = A^-\beta + (I_n - A^-A)Z$，其中 A^- 是 A 任意给定的广义逆，Z 取遍任意一个 n 维列向量。

7.2 伪逆矩阵

定义 7.6 (伪逆矩阵). 对矩阵 $A \in \mathbb{C}^{m\times n}$，若存在 $M \in \mathbb{C}^{n\times m}$，满足：

- $AMA = A$
- $MAM = M$
- $(AM)^{\mathrm{H}} = AM$
- $(MA)^{\mathrm{H}} = MA$

则称 M 为 A 的**伪逆矩阵**，记作 A^+，或 $A^\dagger$。以上条件称为 *Moore-Penrose* 方程。

由定义，伪逆矩阵是一个广义逆矩阵，且当 A 可逆时取 A^{-1}。

定理 7.7 (伪逆矩阵的求解). 对矩阵 $A \in \mathbb{C}^{m\times n}$，将矩阵进行满秩分解 $A = BC$，则：

$$A^+ = C^{\mathrm{H}}(CC^{\mathrm{H}})^{-1}(B^{\mathrm{H}}B)^{-1}B^{\mathrm{H}}$$

是 A 的伪逆矩阵。

定理 7.8. 伪逆矩阵唯一。

性质 7.9.

$$\begin{gathered}
\left(A^+\right)^+ = A \\
\left(AA^H\right)^+ = \left(A^H\right)^+ A^+ = \left(A^+\right)^H A^+ \\
\left(A^HA\right)^+ = A^+\left(A^H\right)^+ = A^+\left(A^+\right)^H \\
A^+ = A^H\left(AA^H\right)^+ = \left(A^HA\right)^+ A^H
\end{gathered}$$

定理 7.10 (伪逆矩阵的求解 2). 对给定 $A \in \mathbb{C}_r^{m\times n}$，对角化 $A^{\mathrm{H}}A$ 得 $U^{\mathrm{H}}A^{\mathrm{H}}AU = \Lambda$，则 $A^+ = U\Lambda^+U^{\mathrm{H}}A^{\mathrm{H}}$，其中 $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_r, 0, \dots, 0)$，$\Lambda^+ = \operatorname{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_r^{-1}, 0, \dots, 0)$。

定理 7.11 (伪逆矩阵的求解 3). 对给定 $A \in \mathbb{C}_r^{m\times n}$，奇异值分解得 $A = U\Sigma V^{\mathrm{H}}$，则 $A^+ = V\Sigma^+U^{\mathrm{H}}$，其中 $\Sigma = \begin{bmatrix} \Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{m\times n}$，$\Sigma^+ = \begin{bmatrix} \Delta^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{n\times m}$。

7.3 最小二乘解

定义 7.12 (最小二乘解). 对方程组 $Ax = b$，给定 $x_0 \in \mathbb{C}^n$，若 $\forall x \in \mathbb{C}^n, \|Ax_0 - b\|^2 \le \|Ax - b\|^2$，则称 x_0 为 $Ax = b$ 的**最小二乘解**。

定义 7.13 (最佳最小二乘解). 对方程组 $Ax = b$，给定最小二乘解 x_0，若对任意最小二乘解 x，都有 $\|x_0\|^2 \le \|x\|^2$，则称 x_0 为 $Ax = b$ 的**最佳最小二乘解**。

定理 7.14. 方程组 $Ax = b$ 的最佳最小二乘解是 $x = A^+b$。

第二部分 最优化理论

8 预备知识

8.1 概念与形式化

定义 8.1 (极值问题及不同约束及记号). 形如

$$\min f(x_1, x_2, \ldots, x_n)$$

的问题称为**无约束问题**。形如

$$\begin{aligned} \min \quad & f(x_1, x_2, \ldots, x_n) \\ \text{s.t.} \quad & h_i(x_1, x_2, \ldots, x_n) = 0, i = 1, 2, \ldots, l \\ & s_j(x_1, x_2, \ldots, x_n) \geq 0, j = 1, 2, \ldots, m \end{aligned}$$

或记作

$$\begin{aligned} \min \quad & f(\vec{x}) \\ \text{s.t.} \quad & \vec{h}(\vec{x}) = \vec{0} \\ & \vec{s}(\vec{x}) \geq \vec{0} \end{aligned}$$

的问题是**有约束问题**。其中 $f(\vec{x})$ 为**目标函数**，$\vec{h}(\vec{x}) = \vec{0}$ 为**等式约束**，$\vec{s}(\vec{x}) \geq \vec{0}$ 为**不等式约束**。

对于 $\max g(x_1, x_2, \ldots, x_n)$，通常转化为 $\min f(x_1, x_2, \ldots, x_n), f = -g$ 处理；对于 $s(x) \leq 0$ 的约束也按 $-s(x) \geq 0$ 处理。

定义 8.2 (容许集 容许解). 满足约束的 $\vec{x}$ 的范围记为 D，称为**容许集/可行集**，其中的 $\vec{x}$ 的取值称为**容许解/可行解**。

定义 8.3 (极小点及其分类). 对 $\vec{x}^* \in D$，若 $\forall \vec{x} \in D, f(\vec{x}^*) \leq f(\vec{x})$，则 $\vec{x}^*$ 称为**最优点/极小点**，$f(\vec{x}^*)$ 称为**最优值/极小值**，$(\vec{x}^*, f(\vec{x}^*))$ 称为**最优解**。根据局部或全局分为**局部极小点**和**全局极小点**。在这个范围内，取值达到最小，称为**严格极小点**，取值没有比它更小的，称为**非严格极小点**。

8.2 简单的情况

- 令一阶导数或梯度为 0
- Lagrange 乘子法
- 图解法

8.3 梯度与 Hesse 矩阵

8.3.1 梯度

定义 8.4 (可微). 对 $f : \mathbb{R}^n \to \mathbb{R}, x_0 \in D$，若 $\exists l \in \mathbb{R}^n, \forall p \in \mathbb{R}^n$, 有：

$$\lim_{\|\vec{p}\| \to 0} \frac{f(\vec{x}_0 + \vec{p}) - f(\vec{x}_0) - \vec{l}^T \vec{p}}{\|\vec{p}\|} = 0$$

则称 f 在 x_0 处可微。此时一阶导数总存在，且为 $\vec{l}$ 的各个分量：

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)^{\mathrm{T}}$$

定义为在 $\vec{x}$ 处的梯度。

性质 8.5. 对多元函数的梯度，以下性质成立：

- 梯度处处垂直于等值面。
- 梯度是 x_0 处变化最快方向。
- 对 $\exists \delta > 0, \forall t \in (0, \delta), f(\vec{x}_0 + t\vec{p}) > f(\vec{x}_0)$ 则称为上升方向，否则为下降方向。

方向导数的定义略。上升方向、下降方向取决于方向导数的正负。方向导数大小为梯度与方向向量的内积。

标量值函数对向量求梯度的常见公式：

$$\nabla C = 0$$
$$\nabla (\vec{b}^{\mathrm{T}} \vec{x}) = \vec{b}$$
$$\nabla (\vec{x}^{\mathrm{T}} Q \vec{x}) = 2Q\vec{x}$$

8.3.2 向量值函数梯度

定义 8.6 (向量值函数可微).

$$\lim_{\|\vec{p}\| \to 0} \frac{\vec{g}\,(\vec{x}_0 + \vec{p}) - \vec{g}\,(\vec{x}_0) - \nabla \vec{g}\,(\vec{x}_0)^T \vec{p}}{\|\vec{p}\|} = \overrightarrow{0}$$

称为向量值函数 $\vec{g}$ 在 $\vec{x}_0$ 处可微，其中

$$\nabla \vec{g}(\vec{x}_0) = \left(\frac{\partial g_j(\vec{x}_0)}{\partial x_i} \right)_{n \times m}$$

称为向量值函数在 $\vec{x}_0$ 处的梯度。$\nabla \vec{g}^{\mathrm{T}}$ 称为 *Jacobi* 矩阵。

向量值函数求梯度的常见公式：

$$\nabla \vec{c} = \mathbf{0}$$
$$\nabla \vec{x} = I$$
$$\nabla (A\vec{x}) = A^{\mathrm{T}}$$

定义 8.7. 记 $\nabla^2 f(x) = \nabla [\nabla f(x)] = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{n \times n}$ 为 $f(\vec{x})$ 的二阶导数，也称为 $f(\vec{x})$ 的 **Hesse 矩阵**。

定理 8.8. 对 $\varphi(t) = f(\vec{x} + t\vec{p})$，有 $\varphi'(t) = \nabla f(\vec{x} + t\vec{p})^{\mathrm{T}} \vec{p}$，$\varphi''(t) = \vec{p}^{\mathrm{T}} \nabla f(\vec{x} + t\vec{p}) \vec{p}$，因此可得三项泰勒展开式：

$$f(\vec{x} + \vec{p}) = f(\vec{x}) + \nabla f(\vec{x})^T \vec{p} + \frac{1}{2} \vec{p}^T \nabla^2 f(\vec{x}) \vec{p} + o(\|\vec{p}\|)$$

8.4 凸函数与凸规划

8.4.1 凸集

定义 8.9 (凸组合 严格凸组合). $x = \sum_i a_i x_i$，其中 $\sum_i a_i = 1$。若 $a_i \geq 0$，称为凸组合，若 $a_i > 0$，称为严格凸组合。

定义 8.10 (凸集). 集合中任意两点的任意凸组合属于集合，称为凸集。

定理 8.11 (凸集判定). 凸集，当且仅当，集合中任意两点的中点总是在集合内。

超平面、椭球等都是凸集，平面、半平面也是。特别规定空集也是。

定理 8.12. 凸集的交集是凸集。

8.4.2 凸函数

定义 8.13. 设 C 为凸集，对 $f: C \subseteq \mathbb{R}^n \to \mathbb{R}$，对任意互异两点 $\vec{x}_1, \vec{x}_2$，对任意和为 1 且非负的 α_1, α_2，总有 $f(\alpha_1\vec{x}_1 + \alpha_2\vec{x}_2) \le \alpha_1 f(\vec{x}_1) + \alpha_2 f(\vec{x}_2)$，则称 f 为定义在凸集 C 上的**凸函数**。若对任意和为 1 且正的 α_1, α_2 有上式且取不到等号，则称 f 为**严格凸函数**。对凸集 C 上的函数 g，若 $f = -g$ 为凸函数，称 g 为**凹函数**。

定理 8.14. 有限个凸函数的非负组合仍然是凸函数。

定理 8.15. 对凸集 C 上的凸函数 f，水平集 $\{\vec{x} \mid f(\vec{x}) \le \beta, \vec{x} \in C\}$ 是凸集。

定理 8.16 (凸函数判定). 设 C 为凸集，$f: C \subseteq \mathbb{R}^n \to \mathbb{R}$ 为可微函数，则：

- f 为凸函数 $\Leftrightarrow \forall \vec{x}_1, \vec{x}_2, f(\vec{x}_2) > f(\vec{x}_1) + \nabla f(\vec{x}_1)^{\mathrm{T}}(\vec{x}_2 - \vec{x}_1)$
- f 为严格凸函数 $\Leftrightarrow \forall \vec{x}_1 \ne \vec{x}_2, f(\vec{x}_2) > f(\vec{x}_1) + \nabla f(\vec{x}_1)^{\mathrm{T}}(\vec{x}_2 - \vec{x}_1)$

定理 8.17. C 非空开凸集，f 二阶可微，则 f 为凸函数 $\Leftrightarrow \nabla^2 f$ 在 C 上半正定。

定理 8.18. C 非空凸集，f 二阶可微，则 f 为严格凸函数 $\Rightarrow \nabla^2 f$ 在 C 上正定。

8.4.3 凸规划

定义 8.19 (凸规划). 设 C 是凸集，f 是凸集上的凸函数。则问题 $\min f(\vec{x}) \text{s.t.} \vec{x} \in C$ 称为**凸规划问题**。

对有约束情况，若等式约束 $\vec{h}(\vec{x}) = 0$ 是线性函数，$s_i(\vec{x})$ 都是凹函数，则容许集是凸集，在容许集上的凸函数也是一种凸规划问题。

性质 8.20. 凸规划具有以下性质：

- 极小点 $\vec{x}^*$ 是全局极小点。
- 严格凸函数的极小点是唯一全局极小点。
- 极小点集合是凸集。

定理 8.21. C 非空凸集，f 是 C 上可微凸函数，则：

$$\vec{x}^* \text{是极小点} \Leftrightarrow \forall \vec{x} \in C, \nabla f(\vec{x}^*)^{\mathrm{T}}(\vec{x} - \vec{x}^*)$$

，即极小点的任意方向都不是下降方向。

推论 8.22. C 非空凸集，f 是 C 上可微凸函数，则：

$$\nabla f(\vec{x}^*) = \vec{0} \Rightarrow \vec{x}^* \text{是极小点}$$

8.4.4 二次规划

定义 8.23 (二次函数 正定二次函数). 多元二次函数表示为 $f(\vec{x}) = \frac{1}{2}\vec{x}^{\mathrm{T}}Q\vec{x} + \vec{b}^{\mathrm{T}}\vec{x} + c$，若 $Q > 0$，称为正定二次函数。

对二次函数 f，有 $\nabla f(\vec{x}) = Q\vec{x} + \vec{b}$，$\nabla^2 f(\vec{x}) = Q$。因此有：

- Q 正定的二次函数是严格凸函数。
- Q 半正定的二次函数的凸函数。

定义 8.24 (二次规划). 最优化问题 $\min f(\vec{x}) = \frac{1}{2}\vec{x}^{\mathrm{T}}Q\vec{x} + \vec{b}^{\mathrm{T}}\vec{x} + c \text{ s.t.} A\vec{x} \ge \vec{p}, C\vec{x} = \vec{d}$ 称为二次规划问题，简称二次规划。若 Q 半正定（或正定），称为二次凸规划问题。

8.5 极小点判定

定理 8.25. f 是 D 上的函数，有连续一阶偏导，$\vec{x}^* \in \operatorname{int} D$，则 $\vec{x}^*$ 是局部极小点 $\Rightarrow \nabla f(\vec{x}^*) = \vec{0}$

定理 8.26. C 非空开凸集，f 是 C 上可微凸函数，$\vec{x}^* \in C$，则 $\vec{x}^*$ 是全局极小点 $\Leftrightarrow \nabla f(\vec{x}^*) = \vec{0}$

定义 8.27 (驻点). f 是 D 上的可微函数，$\vec{x}^* \in D$，若 $\nabla f(\vec{x}^*) = \vec{0}$，则称 $\vec{x}^*$ 为 $f(\vec{x})$ 的**驻点**。

定理 8.28. f 是 D 上函数，有连续二阶偏导，$\vec{x}^* \in D$，则 $\nabla f(\vec{x}^*) = \vec{0}, \nabla^2 f(\vec{x}^*) > 0 \Rightarrow \vec{x}^*$ 是严格局部极小点

定理 8.29. f 是 $\mathbb{R}^n$ 上函数，有连续二阶偏导，$\vec{x}^* \in \mathbb{R}^n$，则 $\nabla f(\vec{x}^*) = \vec{0}, \forall \vec{x} \in N(\vec{x}^*, \delta), \nabla^2 f(\vec{x}) \ge 0 \Rightarrow \vec{x}^*$ 是局部极小点

因此正定二次函数有唯一极小点 $\vec{x}^* = -Q^{-1}\vec{b}$。

若极小点附近 Hesse 矩阵正定，等值面近似同心椭球面族。

8.6 下降迭代算法

8.6.1 下降迭代算法

定义 8.30. 在集合 X 上选取一个初始值 x_0，然后依据某种规则 A 从第 k 次迭代点 x_k 得到第 $k+1$ 次迭代点 x_{k+1}。则称为迭代算法。对最优化问题，X 为容许集时，若迭代过程中使得 $f(x_{k+1}) < f(x_k)$，则称为下降迭代算法。

算法中的迭代步骤：

1. 选定初始点 $\vec{x}_0$，设置 $k = 0$

2. 确定迭代方向 $\vec{p}_k$
3. 确定迭代步长 t_k
4. 迭代 $\vec{x}_{k+1} = \vec{x}_k + t_k\vec{p}_k$
5. 判定是否符合终值准则。若未结束，返回第 2 步。

8.6.2 直线搜索

已确定 $\vec{p}_k$ 后选择 t_k。

定义 8.31. 选择步长 t_k 时选择使得 $\min_t f(\vec{x}_k + t\vec{p}_k)$ 的 t 作为 t_k，这个方法称为直线搜索。其中取值 t_k 称为最优步长因子。

优点是下降量最多，但同时有计算量大的缺点。

性质 8.32. 对直线搜索的结果 $\vec{z} = \vec{x}_k + t_k\vec{p}_k$ 有 $\nabla f(\vec{z})^{\mathrm{T}} \cdot \vec{p}_k = 0$

8.6.3 计算终止准则

理想：$\|\vec{x}^* - \vec{x}\| < \varepsilon$。但 $\vec{x}^*$ 未知因此需要使用其他方法逼近。

简单用变化量刻画终止准则：

$$\begin{cases} \|\vec{x}_{k+1} - \vec{x}_k\| < \varepsilon_1 \\ \|f(\vec{x}_{k+1}) - f(\vec{x}_k)\| < \varepsilon_2 \end{cases}$$

无量纲处理：

$$\begin{cases} \frac{\|\vec{x}_{k+1} - \vec{x}_k\|}{\|\vec{x}_k\|+1} < \varepsilon_1 \\ \frac{\|f(\vec{x}_{k+1}) - f(\vec{x}_k)\|}{\|f(\vec{x}_k)\|+1} < \varepsilon_2 \end{cases}$$

也可以结合 $\|\nabla f\| < \varepsilon_3$ 进行判断。

9 线性规划 *

定义 9.1. 目标函数和等式约束都是线性，且不等式约束各 x 非负的问题，$\min \sum c_i x_i \text{s.t.} \sum a_{ij} x_j = b_i, x_i \geq 0$，称为线性规划问题。一般也记作 $\min C\vec{x} \text{s.t.} A\vec{x} = \vec{b}$。

线性规划的最优解总是出现在顶点。

10 无约束优化方法

无约束优化问题分为根据梯度或 Hesse 矩阵的方法和直接方法。前一类因为需要计算梯度和 Hesse 矩阵，对函数有一定要求，后一类适应性强但收敛较慢。前者按梯度进行搜索，首先我们需要再次考虑直线搜索问题。也就是求解一元函数极小化问题 $\min \varphi(t)$ 的方法。

10.1 直线搜索

对 $\varphi'(t) = 0$ 可求解的问题，显然二分法查找零点是最朴素的方法。此外的分为精确的和不精确的两类。精确的直线搜索方法包括下面的区间收缩和函数逼近两种。

10.1.1 区间收缩法

定义 10.1 (单谷函数)**.** 对函数 $\varphi(t)$，存在全局极小点 t^*，对任意 $t_1 < t_2$，有：若 $t_1 < t_2 \leq t^*$，则 $f(t_1) > f(t_2)$；若 $t^* \leq t_1 < t_2$，则 $f(t_1) < f(t_2)$。即 $\varphi(t)$ 在定义域上，t^* 的左侧严格单调递减，右侧严格单调递增。称这样的 $\varphi(t)$ 为单谷函数。

定义 10.2 (搜索区间)**.** 对函数 $\varphi(t)$，记其全局极小点为 t^*，对区间 $[t_1, t_2]$，若 $t^* \in [t_1, t_2]$，则称为 $\varphi(t)$ 极小点的一个搜索区间，通常记作 $\{t_1, t_2\}$。有时也选择一个 $t_3 \in (t_1, t_2)$ 记作 $\{t_1, t_3, t_2\}$。

单谷函数的定义域就是一个搜索区间。

定理 10.3. 对单谷函数 $\varphi(t)$ 极小点的一个搜索区间 $\{a, b\}$，任取 $t_1, t_2 \in (a, b) t_1 < t_2$，则：若 $\varphi(t_1) \leq \varphi(t_2)$，则 $\{a, t_2\}$ 是一个搜索区间；若 $\varphi(t_1) \geq \varphi(t_2)$，则 $\{t_1, b\}$ 是一个搜索区间。

因此，得到确定搜索区间的算法。

1. 确定初始点 t_0 与步长 h
2. 计算 $\varphi_0 = \varphi(t_0), t_2 = t_0 + h, \varphi_2 = \varphi(t_2)$
3. 若 $\varphi_0 \leq \varphi_2$，令 $t_1 = t_0, \varphi_1 = \varphi_0, h = -h$（翻转搜索方向），转 5；否则转 4
4. 令 $t_1 = t_0, \varphi_1 = \varphi_0, t_0 = t_2, \varphi_0 = \varphi_2, h = 2h$（前进并倍增区间长度）
5. $t_2 = t_0 + h, \varphi_2 = \varphi(t_2)$。若 $\varphi_2 > \varphi_0$ 转 6；否则转 4（前进，若完成则结束，否则倍增）
6. 令 $a = \min\{t_1, t_2\}, b = \max\{t_1, t_2\}$，得到搜索区间 $\{a, b\}$

得到 $t_0 < t_1 < t_2$ 且公差为 h，且 $\varphi_0 > \varphi_1 < \varphi_2$ 的搜索区间。t_0 与 h 需要事先指定，经验上，初始步长 h 选择 1，然后在第 k 次更新中，选择 $h = \beta\frac{\|\vec{x}_k - \vec{x}_{k-1}\|}{\|\vec{p}_k\|}$。

黄金分割法 若考虑复用上次的函数值，使用其中仍然处在中间的点，则问题转化为保证子区间 $\{a, t_1, t_2\}$ 或 $\{t_1, t_2, b\}$ 中间的点仍然处在类似比例的位置，然后就会得到黄金分割比。

对 a, b 取 $\beta = \frac{\sqrt{5}-1}{2}$，然后 $t_2 = \beta(b-a) + a, t_1 = a + b - t_2$，则判断保留 $\{t_1, t_2, b\}$ 时，t_2 在区间的比例成为新的 t_1，保留 $\{a, t_1, t_2\}$ 时，t_1 在区间中的比例成为新的 t_2。

区间每次缩小到原来的 β 倍，较二分法求 $\varphi'(t) = 0$ 缩小速度慢，但是求值的时间会更快。

10.1.2 函数逼近法——二次插值法

二次插值法，也称抛物线插值法，是一种函数逼近法。适用于连续单谷函数求极小值。

算法：

1. 给定搜索区间 $\{t_1,t_2,t_3\}$，过 $(t_1,\varphi_1),(t_2,\varphi_2),(t_3,\varphi_3)$ 拟合抛物线 $Q(t)=pt^2+qt+r$，则得到极小值点 $t_4=-\frac{q}{2p}$。
2. 检查终止条件 $\frac{|\varphi_4-\varphi_2|}{|\varphi_2|+1}<\epsilon$。
3. 比较 t_4,t_2 以及 φ_4,φ_2。
4. 以比较结果选择新的搜索区间。若 $t_1<t_2<t_4<t_3$：若 $\varphi_2<\varphi_4$，取左侧 $\{t_1,t_2,t_4\}$；若 $\varphi_2<\varphi_4$，取右侧 $\{t_2,t_4,t_3\}$。若 $t_1<t_4<t_2<t_3$：若 $\varphi_4>\varphi_2$，取右侧 $\{t_4,t_2,t_3\}$；若 $\varphi_4<\varphi_2$，取左侧 $\{t_1,t_4,t_2\}$。
5. 迭代。

10.1.3 不精确一维搜索

Goldstein 和 Powell 指出了关于设计不精确一维搜索的迭代方法的两个条件。

$$
\begin{gathered}
f(\vec{x}_k)-f(\vec{x}_{k+1})\geq-\rho t_k\nabla f(\vec{x}_k)^T\vec{p}_k\\
\nabla f(\vec{x}_{k+1})^T\vec{p}_k\geq\sigma\nabla f(\vec{x}_k)^T\vec{p}_k
\end{gathered}
$$

其中 ρ,σ 为满足 $0<\rho<\sigma<1$ 的常数。

10.2 最速下降法

回顾前面梯度部分的定义，最速下降方向指梯度的反方向。方便起见，记 $f_k=f(\vec{x}_k),\vec{g}_k=\nabla f(\vec{x}_k)$。

基本思想：第 k 次迭代时，当前已迭代得到 $\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_k$，取搜索方向 $\vec{p}_k=-\nabla\vec{x}_k$，然后选择使得 $\min\limits_t f(\vec{x}_k+t\vec{p}_k)$ 的步长 t_k 进行迭代。

性质 10.4. *每次迭代有 $\nabla f(\vec{x}_k)\perp\nabla f(\vec{x}_{k+1})$，即*

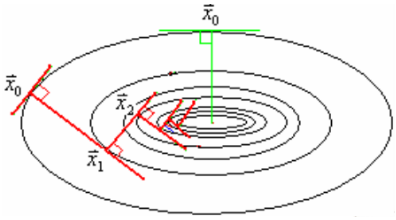
$$\vec{g}_k^{\mathrm{T}}\vec{g}_{k+1}=0$$

以正定二次函数为例，$\vec{g}_k=Q\vec{x}_k+\vec{b}$，代入 $\vec{x}_{k+1}=\vec{x}_k-t_k\vec{g}_k$，因此有 $\vec{g}_{k+1}=Q\vec{x}_{k+1}+\vec{b}=Q\vec{x}_k-t_kQ\vec{g}_k=\vec{g}_k-t_kQ\vec{g}_k$，从 $\vec{g}_k{}^{\mathrm{T}}(\vec{g}_k-t_kQ\vec{g}_k)=0$，可解出对正定二次函数：

$$t_k=\frac{\vec{g}_k^{\mathrm{T}}\vec{g}_k}{\vec{g}_k^{\mathrm{T}}Q\vec{g}_k}$$

性质 10.5 (收敛性)**.** *对任意点收敛，且总是线性收敛的。*

缺点：具有“锯齿现象”。在最速下降法接近最优点时，每次一定垂直上次的方向不断接近目标，造成出现锯齿，不能直接接近目标，这会导致大量的额外计算量。



10.3 Newton 法

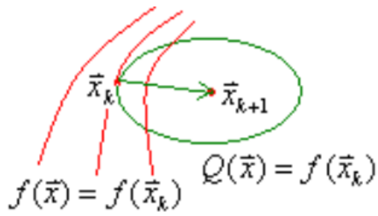
10.3.1 Newton 法

要求有二阶连续偏导数，且 Hesse 矩阵必须正定且显式。记 $f_k=f(\vec{x}_k),\vec{g}_k=\nabla f(\vec{x}_k),G_k=\nabla^2f(\vec{x}_k)$。

对 Taylor 展开式的 $f(\vec{x}^*)\approx f(\vec{x}_k)+g(\vec{x}_k)(\vec{x}^*-\vec{x}_k)+\frac{1}{2}(\vec{x}^*-\vec{x}_k)^{\mathrm{T}}G(\vec{x}_k)(\vec{x}^*-\vec{x}_k)$ 观察右侧关于 $\vec{x}^*$ 的二次函数，最小值取在 $\vec{x}^*-\vec{x}_k=G(\vec{x}_k)^{-1}g(\vec{x}_k)$。因此选择迭代选择 $\vec{x}_{k+1}=\vec{x}_k-G(\vec{x}_k)^{-1}g(\vec{x}_k)$。

对正定二次函数，$G(\vec{x}_k)^{-1}g(\vec{x}_k)=Q^{-1}(Q\vec{x}_k+\vec{b})=\vec{x}_k+Q^{-1}\vec{b}$，因此下一步直接收敛得到 $\vec{x}_{k+1}=-Q^{-1}\vec{b}$。

Newton 法的几何解释是使用二次曲面逼近局部，然后取二次曲面上的极小值点为下次的迭代点。



10.3.2 修正 Newton 法 *

10.4 共轭向量法与共轭梯度法

10.4.1 共轭向量

关于从上次迭代点按二阶导数项拟合一个直接指向最小值的方向，得到 $\vec{p}_0Q\vec{p}_1=0$，因此做出以下定义：

定义 10.6. *若 Q 是 n 阶正定矩阵，非零向量 $\vec{p}_1,\vec{p}_2,\ldots,\vec{p}_m$ 满足 $\vec{p}_i^{\mathrm{T}}Q\vec{p}_j=0$，称这组向量是 Q 共轭向量，或者这组向量是 Q 共轭的，这组向量对应的方向称为 Q 共轭方向。*

定理 10.7. *Q 共轭的向量组线性无关。*

定义 10.8. *对线性无关向量组 $\vec{p}_1,\vec{p}_2,\ldots,\vec{p}_n$，定义全体 $\vec{z}=\vec{x}_0+\sum_i a_i\vec{p}_i$ 构成的集合为点 $\vec{x}_0$ 与向量组 $\vec{p}_1,\vec{p}_2,\ldots,\vec{p}_n$ 生成的线性流形，记为 $L[\vec{x}_0;\vec{p}_1,\vec{p}_2,\ldots,\vec{p}_n]$。*

10.4.2 共轭方向法

定理 10.9. *给定 n 阶正定矩阵 Q，Q 共轭非零向量组 $\vec{p}_0,\vec{p}_1,\ldots,\vec{p}_{m-1}$，任意选取 $\vec{x}_0$ 作为初始点进行 m 次直*

线搜索得到 $\vec{x}_1,\dots,\vec{x}_m$，则有：$\vec{p}_j^{\mathrm{T}}\nabla f(\vec{x}_m)=0$，且 $\vec{x}_m$ 是二次函数 $f(\vec{x})=\frac{1}{2}\vec{x}^{\mathrm{T}}Q\vec{x}+\vec{b}^{\mathrm{T}}\vec{x}+\vec{c}$ 极小点。

因为提供共轭方向的方法不同，共轭方向法有多种。

10.4.3 共轭梯度法

共轭梯度法是一种共轭向量法。

1. 第一次迭代：
   (a) 搜索方向取 $\vec{p}_0=-\vec{g}_0$；
   (b) 步长 $t_0=-\frac{\vec{p}_0^{\mathrm{T}}\vec{g}_0}{\vec{p}_0^{\mathrm{T}}Q\vec{p}_0}$；
   (c) 迭代点 $\vec{x}_1=\vec{x}_0+t_0\vec{p}_0$。
2. 第二次迭代：
   (a) 搜索方向：考虑和 $\vec{p}_0$ 线性无关，设 $\vec{p}_1=\vec{g}_1+\alpha_0\vec{p}_0$，结合共轭条件 $\vec{p}_0^{\mathrm{T}}Q\vec{p}_1=0$，可得 $\alpha_0=-\frac{\vec{g}_1^{\mathrm{T}}Q\vec{p}_0}{\vec{p}_0^{\mathrm{T}}Q\vec{p}_0}$，由此得到 $\vec{p}_1$；
   (b) 此时步长 $t_1=-\frac{\vec{p}_1^{\mathrm{T}}\vec{g}_1}{\vec{p}_1^{\mathrm{T}}Q\vec{p}_1}$；
   (c) 第二个迭代点 $\vec{x}_2=\vec{x}_1+t_1\vec{p}_0$。
3. 第 k 次（$k>3$）：
   (a) 搜索方向：考虑和上次的 $\vec{p}_{k-1}$ 线性无关，设 $\vec{p}_k=\vec{g}_k+\alpha_{k-1}\vec{p}_{k-1}$，结合共轭条件 $\vec{p}_k^{\mathrm{T}}Q\vec{p}_{k-1}=0$，可得 $\alpha_{k-1}=-\frac{\vec{g}_k^{\mathrm{T}}Q\vec{p}_0}{\vec{p}_0^{\mathrm{T}}Q\vec{p}_0}$，由此得到 $\vec{p}_k$；
   (b) 此时步长 $t_k=-\frac{\vec{p}_k^{\mathrm{T}}\vec{g}_k}{\vec{p}_k^{\mathrm{T}}Q\vec{p}_k}$；
   (c) 第二个迭代点 $\vec{x}_k=\vec{x}_{k-1}+t_1\vec{p}_k$。

不使用 Hesse 矩阵的共轭梯度法 *

10.5 拟 Newton 法

考虑使用其他矩阵 H 近似 Newton 法中的 G^{-1}。每次使用 $\vec{x}_{k+1}=\vec{x}_k-t_kH_k\vec{g}_k$ 迭代。然后考查使 H 确实近似 G^{-1} 的条件：

- H_k 正定，以保证每次的搜索方向都是下降方向。
- 易于计算的迭代关系，迭代 H_k 的公式称为校正公式。
- 逆 Newton 条件：$\vec{x}_{k+1}-\vec{x}_k\approx H_{k+1}(\vec{g}_{k+1}-\vec{g}_k)$
- DFP：选择 $H_{k+1}=H_k+\alpha_k\vec{u}_k\vec{u}_k^{\mathrm{T}}+\beta_k\vec{v}_k\vec{v}_k^{\mathrm{T}}$ 作为校正公式，其中 $\alpha_k,\beta_k,\vec{u}_k,\vec{v}_k$ 都是待定值。
- BFGS(*)
- Broyden 算法族 (*)

10.6 步长加速法 *

11 约束最优化方法

最常见的是容许方向法和罚函数法。

11.1 最优性条件

11.1.1 等式约束

对含有等式约束的最优化问题，

$$\begin{array}{ll}\min & f(\vec{x})\\ \text{s.t.} & \vec{h}(\vec{x})=0\end{array}$$

令 $L=f(\vec{x})-\sum_i\lambda_ih_i(\vec{x})$，极值点问题转化为求 Lagrange 函数的极值点，即

$$\nabla L=\begin{bmatrix}\nabla_{\vec{x}}L\\ \nabla_{\vec{\lambda}}L\end{bmatrix}=\begin{bmatrix}\nabla f(\vec{x})+\sum\limits_i\lambda_i\nabla h_i(\vec{x})\\ h_1(\vec{x})\\ \vdots\\ h_m(\vec{x})\end{bmatrix}=\vec{0}$$

这一方法称为 Lagrange 乘子法。

11.1.2 不等式约束的几何最优性条件

定义 11.1. 对容许集中一点，去到等号的不等式称为**起作用约束**，否则称为**不起作用约束**。对内点，每个约束都是不起作用约束，对边界点，有至少一个起作用约束。

定义 11.2. 对从一点 $\vec{x}$ 和方向向量 $\vec{p}$，若 $\exists\delta>0$ 使得 $\forall t\in(0,\delta),\vec{x}+t\vec{p}\in D$，称为**容许方向**。

定义 11.3 (锥。凸锥。)**.**

定义 11.4. 全体容许方向向量构成**容许方向锥**。全体下降方向向量构成**下降方向锥**。

定理 11.5 (几何最优性条件)**.** 下降方向锥与容许方向锥交集为空。

这是一个必要条件。

11.1.3 F-J 条件

引理 11.6 (Farkas 引理)**.** 对 n 维向量 $a_1,a_2,\dots,a_m,b$，有：$a_i^{\mathrm{T}}p>0,i=1,2,\dots,m\Rightarrow b^{\mathrm{T}}p>0\Leftrightarrow\exists\gamma_1,\gamma_2,\dots,\gamma_m>0,b=\sum\limits_i\gamma_ia_i$。

引理 11.7 (Gordan 引理)**.** 对 n 维向量 $a_1,a_2,\dots,a_m$，有：$\nexists p,\vec{a}^{\mathrm{T}}p<0,i=1,2,\dots,m\Rightarrow\exists\gamma_1,\gamma_2,\dots,\gamma_m\geq0,b=\sum\limits_i\gamma_ia_i=\vec{0}$。

定理 11.8 (Fritz-John 条件)**.** $\vec{x}^*$ 是极小点 $\Rightarrow\exists\mu_0,\mu_1,\dots,\mu_m$不全为零，使得：

$$\begin{cases}\mu_0\nabla f(\vec{x}^*)-\sum\limits_i\mu_i\nabla s_i(\vec{x}^*)=0 & \textit{(Lagrange 函数)}\\ \mu_is_i(\vec{x}^*)=0,i=1,2,\dots,m & \textit{(互补松弛条件)}\\ \mu_i\geq0,i=1,2,\dots,m\end{cases}$$

这个必要条件称为 *F-J* 条件，满足 *F-J* 条件的所有点称为 *F-J* 点。式中的 $\mu_0,\mu_1,\dots,\mu_m$ 称为 *Lagrange* 乘子。

11.1.4　K-T 条件

约束 F-J 条件的 $\mu_0 = 1$ 得到 K-T 条件。

定理 11.9 (Kuhn-Tucker 条件)**.** *$\vec{x}^*$ 是极小点 $\Rightarrow \exists \mu_1, \ldots, \mu_m$，使得：*

$$\begin{cases} \mu_0 \nabla f(\vec{x}^*) = \sum_i \mu_i \nabla s_i(\vec{x}^*) & \textit{(Lagrange 函数)} \\ \mu_i s_i(\vec{x}^*) = 0, i = 1, 2, \ldots, m & \textit{(互补松弛条件)} \\ \mu_i \geq 0, i = 1, 2, \ldots, m \end{cases}$$

这个必要条件称为 K-T 条件或 KKT 条件 (KKT 指 Karush-Kuhn-Tucker)，满足 K-T 条件的所有点称为 K-T 点。

定理 11.10. *起作用的 ∇s 线性无关，且 $\mu_0 \neq 0$ 的 F-J 点必为 K-T 点。*

定理 11.11. *对凸规划，其中 f 可微凸函数，s 可微凹函数，h 线性函数，则 K-T 点都是全局最优点。*

11.2　外部罚函数法

对最优化问题，定义增广目标函数

$$F(\vec{x}, \mu) = f(\vec{x}) + \mu \alpha(\vec{x})$$
$$\alpha(\vec{x}) = \sum_j [h_j(\vec{x})]^2 + \sum_i [s_i(\vec{x})]^2 u(s_i(\vec{x}))$$

F 称为增广目标函数，μ 为罚因子，第二项 $\mu\alpha(\vec{x})$ 称为惩罚项，其中的 $u(t)$ 是一个阶跃函数，定义为

$$u(t) = \begin{cases} 0, t \geq 0 \\ 1, t < 0 \end{cases}$$

定理 11.12. *给定 μ，若 $\vec{x}_\mu$ 是无约束问题 $\min F(\vec{x}, \mu)$ 的极小点，则 $\vec{x}_\mu$ 是有约束问题 $\min f(\vec{x})$s.t.$h_j(\vec{x}) = 0, s_i(\vec{x}) \geq 0$ 的极小点的充要条件是 $\vec{x}_\mu$ 是有约束问题的容许点。*

对于一般的 $\vec{x} \notin C$，需要增大 μ 计算，一般认为 $\mu = 1, 2, \ldots$ 得到序列 $\{\vec{x}_k\}$，则若 $\{\vec{x}_k\}$ 收敛，一定能收敛在极小点。这种方法称为**外部罚函数法**或**外点法**。这种通过解无约束问题进而解决有约束问题的方法，也称为“序列无约束最小化技术”(SUMT)。

11.3　内部罚函数法

外点法取在外部并逐渐向容许集边界移动，因此其可能难以取到容许集内的点，或者可能函数在容许集外根本无法定义，导致不能用外点法计算出极小点。为保证迭代点总是内点，则产生了内部罚函数法。

初始点 $\vec{x}_0$ 选择内点，然后设计障碍函数

$$F(\vec{x}, \mu) = f(\vec{x}) + \mu \beta(\vec{x})$$
$$\beta(\vec{x}) = \sum_i \frac{1}{s_i(\vec{x})}$$

F 为障碍函数，μ 为罚因子，第二项 $\mu\beta(\vec{x})$ 为惩罚项。β 也可以取如 $\sum \ln \frac{1}{s(\vec{x})}, \sum \frac{1}{s^2(\vec{x})}, \sum \frac{1}{s(\vec{x})u[-s(\vec{x})]}$ 等的函数。

选择一个内点作为初始点，通过 $\mu \to 0^+$，得到极小点。

注意内部罚函数法有无法处理等式约束的严重缺陷，产生的结合两种罚函数法的优势得到了混合罚函数法。

11.4　乘子法

外部罚函数法中，随着罚因子增大，Hesse 矩阵条件数增大，数值计算的稳定性变差导致计算结果不精确，于是产生了乘子法。

H 乘子法 *

R 乘子法 *

11.5　Zoutendijk 容许方向法 *

12　多目标规划

同时对多个目标进行优化即多目标规划问题。

12.1　数学模型

不等式约束转换为大于等于后，可得数学模型：

$$\begin{aligned} \min \quad & f_1(x_1, x_2, \ldots, x_n) \\ \min \quad & f_2(x_1, x_2, \ldots, x_n) \\ & \ldots \\ \min \quad & f_p(x_1, x_2, \ldots, x_n) \\ \max \quad & g_1(x_1, x_2, \ldots, x_n) \\ \max \quad & g_2(x_1, x_2, \ldots, x_n) \\ & \ldots \\ \max \quad & g_q(x_1, x_2, \ldots, x_n) \\ \text{s.t.} \quad & s_i(x_1, x_2, \ldots, x_n) \geq 0 \\ & h_j(x_1, x_2, \ldots, x_n) = 0 \end{aligned}$$

将最大转换为最小，写成向量形式，记

$$\begin{aligned} \vec{x} &= (x_1, x_2, \ldots, x_n)^{\mathrm{T}} \\ \vec{f}(\vec{x}) &= (f_1(\vec{x}), f_2(\vec{x}), \ldots, f_p(\vec{x}), f_{p+1}(\vec{x}), \ldots, f_r(\vec{x}))^{\mathrm{T}} \\ f_{p+k}(\vec{x}) &= -g_k(\vec{x}) \\ D &= \{\vec{x} \mid s_i(\vec{x}) \geq 0, h_i(\vec{x}) = \vec{0}\} \end{aligned}$$

则原问题可记为

$$\mathbf{v}\text{-} \min_{\vec{x} \in D} \vec{f}(\vec{x})$$

称为向量极小化模型，$\vec{f}(\vec{x})$ 称为向量目标函数，其分量称为分量目标函数。

对各分量目标函数为线性的情况，可通过线性替换使得 $\vec{s}(\vec{x})$ 转化为

$$\begin{aligned} \mathbf{v}-\min_{\vec{x}} \quad & \vec{f}(\vec{x}) = C\vec{x} \\ \text{s.t.} \quad & A\vec{x} = \vec{b} \\ & \vec{x} \geq \vec{0} \end{aligned}$$

12.2 解的概念与性质

定义 12.1 (向量的序关系)**.**

$$\vec{a} < \vec{b} \Leftrightarrow \vec{b} > \vec{a} \Leftrightarrow \forall i, a_i < b_i$$
$$\vec{a} \leq \vec{b} \Leftrightarrow \vec{b} \geq \vec{a} \Leftrightarrow \forall i, a_i \leq b_i$$
$$\vec{a} \prec \vec{b} \Leftrightarrow \vec{b} \succ \vec{a} \Leftrightarrow \forall i, a_i \leq b_i \land \exists i, a_i \neq b_i$$

定义 12.2 (不同程度的最优解)**.** 为确定“好”解，定义：

- 对 $\vec{f}(\vec{x})$，给定 $\vec{x}^*$，若 $\forall \vec{x} \in D, \vec{f}(\vec{x}) \geq \vec{f}(\vec{x}^*)$，则称 $\vec{x}^*$ 为**绝对最优解**。全体绝对最优解的集合为**绝对最优解集**，记作 $X^*(\vec{f}, D)$。
- 对 $\vec{f}(\vec{x})$，给定 $\vec{x}^*$，若 $\nexists \vec{x} \in D, \vec{f}(\vec{x}) \prec \vec{f}(\vec{x}^*)$，则称 $\vec{x}^*$ 为**有效解**或 **Pareto 解**。全体有效解的集合为**有效解集**，记作 $P(\vec{f}, D)$。
- 对 $\vec{f}(\vec{x})$，给定 $\vec{x}^*$，若 $\nexists \vec{x} \in D, \vec{f}(\vec{x}) < \vec{f}(\vec{x}^*)$，则称 $\vec{x}^*$ 为**弱有效解**。全体弱有效解的集合为**弱有效解集**，记作 $P_w(\vec{f}, D)$。

定理 12.3. $X^* \subseteq P \subseteq P_w \subseteq D$，且 $X^* = \bigcap\limits_i X_i^*, \bigcup\limits_i X_i^* \subseteq P_w$。其中 X^*, P, P_w 分别为绝对最优、有效、弱有效解，X_i^* 为第 i 个分量目标函数的最优解集。

定理 12.4. 若 $X^* \neq \varnothing$，则 $P = X^*$。若 D 是凸集，每个分量目标函数 f_i 是凸函数，则 $P = P_w$。

12.3 评价函数法

定义 12.5 (多元函数单调性)**.** 多元函数 $\varphi : \mathbb{R}^r \to \mathbb{R}$：若 $\forall \vec{z}_1 < \vec{z}_2, \varphi(\vec{z}_1) < \varphi(\vec{z}_2)$，称 $\varphi(\vec{z})$ 单调增函数；若 $\forall \vec{z}_1 \prec \vec{z}_2, \varphi(\vec{z}_1) < \varphi(\vec{z}_2)$，称 $\varphi(\vec{z})$ 严格单调增函数。

定理 12.6. 对多目标规划 $\vec{f}(\vec{x})$，考虑 $\min\limits_{x \in D} \varphi(\vec{f}(\vec{x}))$ 有极小点 $\vec{x}^*$，则：

- 若 φ 是严格增函数，$x^* \in P(\vec{f}, D)$
- 若 φ 是增函数，$x^* \in P_w(\vec{f}, D)$

称这样的 φ 为评价函数。

12.3.1 常见评价函数

线性加权和 取评价函数 $\varphi(\vec{z}) = \vec{u}\vec{z}, u_i \geq 0, \sum u_i = 1$ 为**线性加权和函数**，其中 $\vec{u}$ 为**权向量**，且：$\vec{u} > 0$ 时，严格单调增；$\vec{u} \succ \vec{0}$ 时，单调增。这样的方法称为**线性加权和法**。

理想点函数 取评价函数 $\varphi(\vec{z}) = \|\vec{z} - \vec{z}^*\|$ 为**理想点函数**，其中 $\vec{z}^*$ 是每个分量函数能取到的最小值，称为**理想点**，这样的方法称为**理想点法**。

平方加权函数 考虑理想点函数中范数进行平方，取加权的二次方和的情况，取评价函数 $\varphi(\vec{z}) = \sum\limits_i (\vec{z}_i - \vec{z}_i^*)$。这样的方法称为**平方加权和法**。

极大函数 取评价函数 $\varphi(\vec{z}) = \max\{z_i\}$ 或加权的 $\max\{u_i z_i\}$，这样的方法称为**极大极小法**。由于直接求这个极大值的极小值较难，通常转化为 $\min v \text{s.t.} u_i f_i(\vec{x}) \leq v, x \in D$ 来求解，其中 v 称为松弛变量。

12.3.2 权重的确定

去量纲处理 首先需要将各不同目标函数的量纲去除。

老手法 选择多位相关专家和有实践经验的工作者，统计其对不同项目的权重的估计值，然后求取平均。求平均后需要检查每个人与平均结果的最大偏差，若偏差足够小则认为大家的经验无差别，可以使用，否则需要再次进行协商调整并重复此流程。

α 法 借助各分量目标函数在容许集上的极小值点确定权重的方法。通过最小化不同的分量目标函数得到极小值点 $\vec{x}_1^*, \vec{x}_2^*, \ldots, \vec{x}_r^*$，在自变量取这些值时得到向量目标函数的取值 $\vec{z}_i^* = \vec{f}(\vec{x}_i^*)$。若这是 r 个不同的点，撑起一个 r 维空间中的超平面，设 $\vec{u}^{\mathrm{T}}\vec{z} = \alpha$，改变合适的 α 值使得 $\vec{u}$ 每个分量非负且和为 1，则得到一个可用的权重向量 $\vec{u}$。

13 标量的矩阵求导术

定义 13.1. 对矩阵函数 $f : F^{m \times n} \to F$，定义对矩阵的导数 $\frac{\partial f}{\partial A} = \left(\frac{\partial f}{\partial a_{ij}}\right)_{m \times n}$

- $f(A) = X^{\mathrm{T}}AY, \frac{\partial f}{\partial A} = XY^{\mathrm{T}}$
- $f(A) = \mathrm{tr}(AB) = \mathrm{tr}(BA), \frac{\partial f}{\partial A} = B^{\mathrm{T}}$
- $f(A) = \mathrm{tr}(A^{\mathrm{T}}C) = \mathrm{tr}(CA^{\mathrm{T}}) = A \bullet C, \frac{\partial f}{\partial A} = C$
- $f(A) = \mathrm{tr}(A^{\mathrm{T}}A) = \mathrm{tr}(AA^{\mathrm{T}}) = \|A\|_F^2, \frac{\partial f}{\partial A} = 2A$
- $f(A) = \mathrm{tr}(A^{\mathrm{T}}AB), \frac{\partial f}{\partial A} = AB + AB^{\mathrm{T}}$
- $f(A) = \mathrm{tr}(A^{\mathrm{T}}BA), \frac{\partial f}{\partial A} = BA + B^{\mathrm{T}}A$
- $f(A) = \mathrm{tr}(PAQ) = \mathrm{tr}(AQP), \frac{\partial f}{\partial A} = P^{\mathrm{T}}Q^{\mathrm{T}}$