1 定义

1. 梯度

   若 $f:\mathbb{R}^n\to\mathbb{R}$ 在点 $\vec{x_0}=(x_1,x_2,\ldots,x_n)$ 处可微，即 $f(\vec{x})$ 在该点关于各变量的一阶偏导数存在，则 $f(\vec{x})$ 的 n 个偏导数构成的列向量

$$\left(\left.\frac{\partial f}{\partial x_1}\right|_{\vec{x}=\vec{x_0}},\left.\frac{\partial f}{\partial x_2}\right|_{\vec{x}=\vec{x_0}},\ldots,\left.\frac{\partial f}{\partial x_n}\right|_{\vec{x}=\vec{x_0}}\right)^{\mathrm{T}}$$

   称为 $f(\vec{x})$ 在点 $\vec{x_0}$ 处的梯度，记为 $\nabla f(\vec{x})|_{\vec{x}=\vec{x_0}}$.

2. Hesse 矩阵

   对于 $f:\mathbb{R}^n\to\mathbb{R}$，矩阵

$$H(f(\vec{x}))=\begin{pmatrix}\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2\partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n\partial x_1}\\ \frac{\partial^2 f}{\partial x_1\partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n\partial x_2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial^2 f}{\partial x_1\partial x_n} & \frac{\partial^2 f}{\partial x_2\partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2}\end{pmatrix}$$

   称为 $f(\vec{x})$ 的 Hesse 矩阵

3. 凸集

   设 S 是 $\mathbb{R}^n$ 中的一个集合。若对于任意两点 $\vec{x_1},\vec{x_2}\in S$ 及每个实数 $\lambda\in[0,1]$, 有

$$\lambda\vec{x_1}+(1-\lambda)\vec{x_2}\in S$$

   则称 S 为凸集。$\lambda\vec{x_1}+(1-\lambda)\vec{x_2}$ 称为 $\vec{x_1}$ 和 $\vec{x_2}$ 的凸组合。

   Remark

   线段与直线均为 $\mathbb{R}$ 上的凸集。

   对于由 $D=\{\vec{x}\mid f_i(\vec{x})=0,g_j(\vec{x})\geq 0\}$ 构成的集合，若 $f_i(\vec{x})$ 均为线性函数，$g_i(\vec{x})$ 均为凸函数，则 D 为凸集。

4. 凸包

   集合 $T\subset\mathbb{R}^n$ 的凸包是指所有包含 T 的凸集的交集，记为

$$\mathrm{conv}T=\bigcap_{T\subset C}C$$

   其中，C 为凸集。

5. 凸函数

   设 $S\subset\mathbb{R}^n$ 为非空凸集，函数 $f:S\to\mathbb{R}$. 若 $\forall\vec{x_1},\vec{x_2}\in S$, 和每一个 $\lambda\in(0,1)$, 都有

$$f(\lambda\vec{x_1}+(1-\lambda)\vec{x_2})\leq\lambda f(\vec{x_1})+(1-\lambda)f(\vec{x_2})$$

   则称 f 是 S 上的凸函数。若上面的不等式对于 $\vec{x_1}\neq\vec{x_2}$ 严格成立，则称 f 是 S 上的严格凸函数。

6. 上镜图
   对于函数 $f: S \to \mathbb{R}$, 称集合
   $$\left\{\begin{array}{c}\text{epi}f = \\ (x,y) \mid x \in S, y \in \mathbb{R}, y \geq f(x)\end{array}\right\}$$
   为 f 的上镜图。

7. 驻点与鞍点
   若 $f(\vec{\boldsymbol{x}})$ 在点 $\vec{\boldsymbol{x}}^*$ 处可微，且 $\nabla f(\vec{\boldsymbol{x}}^*) = 0$, 则称 $\vec{\boldsymbol{x}}^*$ 为 f 的一个驻点。既不是极大点也不是极小点的驻点称为鞍点。

8. 共轭方向
   设 A 是 $n \times n$ 对称正定矩阵，若 $\mathbb{R}^n$ 中的两个方向 $\vec{\boldsymbol{d}}^{(1)}$ 和 $\vec{\boldsymbol{d}}^{(2)}$ 满足
   $$\vec{\boldsymbol{d}}^{(1)^{\mathrm{T}}} A \vec{\boldsymbol{d}}^{(2)} = 0$$
   则称这两个方向关于 A 共轭。

9. 矩阵的内积
   定义 S^n 为一组 $n \times n$ 对称矩阵的集合。对于任意矩阵 $A, B \in S^n$, 定义 A 与 B 的内积为
   $$A \cdot B = \text{trace}(AB) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij}$$

2 定理

1. 凸集的性质
   设 S_1 和 S_2 是 $\mathbb{R}^n$ 中两个凸集，β 是实数，则
   - $\beta S_1 = \{\beta \vec{\boldsymbol{x}} \mid \vec{\boldsymbol{x}} \in S_1\}$ 是凸集
   - $S_1 \cap S_2$ 为凸集
   - $S_1 + S_2 = \{\vec{\boldsymbol{x}_1} + \vec{\boldsymbol{x}_2} \mid \vec{\boldsymbol{x}_1} \in S_1, \vec{\boldsymbol{x}_2} \in S_2\}$ 为凸集
   - $S_1 - S_2 = \{\vec{\boldsymbol{x}_1} - \vec{\boldsymbol{x}_2} \mid \vec{\boldsymbol{x}_1} \in S_1, \vec{\boldsymbol{x}_2} \in S_2\}$ 为凸集

2. 凸函数的性质
   设 S 为 $\mathbb{R}^n$ 的非空凸子集，则 f 是凸函数当且仅当其上镜图是凸集。

3. 凸函数的判定定理
   - 设 S 是 $\mathbb{R}^n$ 中的非空开凸集，$f: S \to \mathbb{R}$ 是可微的函数，则 f 是凸函数当且仅当 $\forall \vec{\boldsymbol{x}}^* \in S$, 有
     $$f(\vec{\boldsymbol{x}}) - f(\vec{\boldsymbol{x}}^*) \geq \nabla f(\vec{\boldsymbol{x}}^*)\left(\vec{\boldsymbol{x}} - \vec{\boldsymbol{x}}^*\right), \forall \vec{\boldsymbol{x}} \in S$$
     f 严格凸当且仅当 $\forall \vec{\boldsymbol{x}}^* \in S$, 有
     $$f(\vec{\boldsymbol{x}}) - f(\vec{\boldsymbol{x}}^*) \nabla f(\vec{\boldsymbol{x}}^*)\left(\vec{\boldsymbol{x}} - \vec{\boldsymbol{x}}^*\right), \forall \vec{\boldsymbol{x}} \in S, \vec{\boldsymbol{x}} \neq \vec{\boldsymbol{x}}^*$$
   - 设 S 是 $\mathbb{R}^n$ 中的非空开凸集，$f: S \to \mathbb{R}$ 是二次可微的函数，则

- f 是凸函数当且仅当 S 上每一点的 Hesse 矩阵是半正定的
  Remark
  实对称矩阵 A 半正定当且仅当其特征值均非负。
- f 是严格凸函数当且仅当 S 上每一点的 Hesse 矩阵是正定的
  Remark
  实对称矩阵 A 正定当且仅当其特征值均正。

3 方法

3.1 无约束极值问题 (UNLP)

3.1.1 定义

$$\min \quad f(\vec{\boldsymbol{x}}), \vec{\boldsymbol{x}} \in \mathbb{R}^n$$

其中 f 是 $\mathbb{R}^n$ 上的实值函数。

3.1.2 极值条件

- 极小点的必要条件
  设 $\vec{\boldsymbol{x}}^*$ 是问题 UNLP 的局部极小点。
  - 若 f 在 $\vec{\boldsymbol{x}}^*$ 处可微，则 $\nabla f(\vec{\boldsymbol{x}}^*) = 0$.
  - 若 f 在 $\vec{\boldsymbol{x}}^*$ 处二次可微，则 $\nabla f(\vec{\boldsymbol{x}}^*) = 0$ 且 Hesse 矩阵 $H(\vec{\boldsymbol{x}}^*)$ 半正定。
- 二阶充分条件
  假设 f 在 $\vec{\boldsymbol{x}}^*$ 点二次可微，若 $\nabla f(\vec{\boldsymbol{x}}^*) = 0$ 且 Hesse 矩阵 $H(\vec{\boldsymbol{x}}^*)$ 是正定的，则 $\vec{\boldsymbol{x}}^*$ 是 UNLP 的一个严格局部极小点。
- 充要条件
  假设 $f : \mathbb{R}^n \to \mathbb{R}$ 是可微的凸函数，则 $\vec{\boldsymbol{x}}^*$ 是 UNLP 的全局最小点当且仅当 $\nabla f(\vec{\boldsymbol{x}}^*) = 0$

3.1.3 一维搜索

得到点 $\vec{\boldsymbol{x}}^{(k)}$ 后，需要按某种规则确定一个方向 $\vec{\boldsymbol{d}}^{(k)}$,再从 $\vec{\boldsymbol{x}}^{(k)}$ 出发，沿方向 $\vec{\boldsymbol{d}}^{(k)}$ 在直线上求目标函数的极小点，从而得到 $\vec{\boldsymbol{x}}^{(k)}$ 的后继点 $\vec{\boldsymbol{x}}^{(k+1)}$. 重复以上做法，直至求得问题的解。这种方法称为一维搜索。

- 试探法
  - 0.618 法
    设 f 在 $[a, b]$ 上单峰。
    step 1 置初始区间 $[a_1, b_1]$ 及精度要求 $L > 0$. 计算试探点 λ_1 和 μ_1, 计算函数值 $f(\lambda_1)$ 和 $f(\mu_1)$. 计算公式是
    $$\lambda_1 = a_1 + 0.382(b_1 - a_1)$$
    $$\mu_1 = a_1 + 0.618(b_1 - a_1)$$
    令 $k = 1$

step 2 若 $b_k - a_k < L$, 则停止计算。否则，当 $f(\lambda_k) > f(\mu_k)$ 时，转**step 3**; 当 $f(\lambda_k) \leq f(\mu_k)$ 时，转**step 4**.

step 3 置 $a_{k+1} = \lambda_k, b_{k+1} = b_k, \lambda_{k+1} = \mu_k$

$$\mu_{k+1} = a_{k+1} + 0.618(b_{k+1} - a_{k+1})$$

计算函数值 $f(\mu_{k+1})$, 转**step 5**.

step 4 置 $a_{k+1} = a_k, b_{k+1} = \mu_k, \mu_{k+1} = \lambda_k$

$$\lambda_{k+1} = a_{k+1} + 0.382(b_{k+1} - a_{k+1})$$

计算函数值 $f(\lambda_{k+1})$, 转**step 5**.

step 5 置 $k = k + 1$, 返回**step 2**.

○ Fibonacci 法

设有数列 $\{F_k\}$, 满足

* $F_0 = F_1 = 1$
* $F_{k+1} = F_k + F_{k-1}$

则称 $\{F\}$ 为 Fibonacci 数列。

设 f 在 $[a, b]$ 上单峰。

step 1 置初始区间 $[a_1, b_1]$ 及精度要求 $L > 0$. 求计算函数值的次数 n, 使

$$F_n \geq \frac{b_1 - a_1}{L}$$

置辨别常数 $\delta > 0$. 计算试探点 λ_1 和 μ_1, 计算函数值 $f(\lambda_1)$ 和 $f(\mu_1)$. 计算公式是

$$\lambda_1 = a_1 + \frac{F_{n-2}}{F_n}(b_1 - a_1)$$
$$\mu_1 = a_1 + \frac{F_{n-1}}{F_n}(b_1 - a_1)$$

令 $k = 1$

step 2 若 $f(\lambda_k) > f(\mu_k)$ 时，转**step 3**; 当 $f(\lambda_k) \leq f(\mu_k)$ 时，转**step 4**.

step 3 置 $a_{k+1} = \lambda_k, b_{k+1} = b_k, \lambda_{k+1} = \mu_k$

$$\mu_{k+1} = a_{k+1} + \frac{F_{n-k-1}}{F_{n-k}}(b_{k+1} - a_{k+1})$$

若 $k = n - 2$, 则转**step 6**; 否则，计算函数值 $f(\mu_{k+1})$, 转**step 5**.

step 4 置 $a_{k+1} = a_k, b_{k+1} = \mu_k, \mu_{k+1} = \lambda_k$

$$\lambda_{k+1} = a_{k+1} + \frac{F_{n-k-2}}{F_{n-k}}(b_{k+1} - a_{k+1})$$

若 $k = n - 2$, 则转**step 6**; 否则，计算函数值 $f(\lambda_{k+1})$, 转**step 5**.

step 5 置 $k = k + 1$, 转**step 2**.

step 6 令 $\lambda_n = \lambda_{n-1}, \mu_n = \lambda_{n-1} + \delta$, 计算 $f(\lambda_n)$ 和 $f(\mu_n)$

若 $f(\lambda_n) > f(\mu_n)$, 则令

$$a_n = \lambda_n, b_n = b_{n-1}$$

若 $f(\lambda_n) \leq f(\mu_n)$, 则令

$$a_n = a_{n-1}, b_n = \lambda_n$$

停止计算，最小值包含于 $[a_n, b_n]$.

- 最速下降法
  最速下降法的迭代公式为
  $$\vec{x}^{(k+1)} = \vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}$$
  其中 $\vec{d}^{(k)}$ 是从 $\vec{x}^{(k)}$ 出发的搜索方向，这里取在点 $\vec{x}^{(k)}$ 处的最速下降方向，即
  $$\vec{d}^{(k)} = -\nabla f(\vec{x}^{(k)})$$
  λ_k 是从 $\vec{x}^{(k)}$ 出发沿方向 $\vec{d}^{(k)}$ 进行一维搜索的步长，即 λ_k 满足
  $$f(\vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}) = \min_{\lambda \geq 0} f(\vec{x}^{(k)} + \lambda \vec{d}^{(k)})$$
  step 1 给定初点 $\vec{x}^{(1)} \in \mathbb{R}^n$, 允许误差 $\varepsilon > 0$, 置 $k = 1$.
  step 2 计算搜索方向 $\vec{d}^{(k)} = -\nabla f(\vec{x}^{(k)})$
  step 3 若 $\left\|\vec{d}^{(k)}\right\| \leq \varepsilon$, 则停止计算；否则，从 $\vec{x}^{(k)}$ 出发，沿 $\vec{d}^{(k)}$ 进行一维搜索，求 λ_k, 使
  $$f(\vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}) = \min_{\lambda \geq 0} f(\vec{x}^{(k)} + \lambda \vec{d}^{(k)})$$
  step 4 令 $\vec{x}^{(k+1)} = \vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}$, 置 $k = k + 1$, 转**step 2**.

- Newton 法
  对于函数 $f(\vec{x})$, 记 $\nabla f(\vec{x})$ 为 f 在 $\vec{x}$ 处的梯度, $\nabla^2 f(\vec{x})$ 为在 $\vec{x}$ 处的 Hesse 矩阵, $\nabla^2 f(\vec{x})^{-1}$ 为在 $\vec{x}$ 处的 Hesse 矩阵的逆矩阵。则 Newton 法的迭代公式为
  $$\vec{x}^{(k+1)} = \vec{x}^{(k)} - \nabla^2 f(\vec{x}^{(k)})^{-1} \nabla f(\vec{x}^{(k)})$$

- 阻尼 Newton 法
  阻尼 Newton 法的迭代公式为
  $$\vec{x}^{(k+1)} = \vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}$$
  其中 $\vec{d}^{(k)} = -\nabla^2 f\left(\vec{x}^{(k)}\right)^{-1} \nabla\left(\vec{x}^{(k)}\right)$ 为 Newton 方向，λ_k 是由一维搜索得到的步长，即满足
  $$f(\vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}) = \min_{\lambda \geq 0} f(\vec{x}^{(k)} + \lambda \vec{d}^{(k)})$$
  step 1 给定初始点 $\vec{x}^{(1)}$, 允许误差 $\varepsilon > 0$, 置 $k = 1$.
  step 2 计算 $\nabla f\left(\vec{x}^{(k)}\right), \nabla^2 f\left(\vec{x}^{(k)}\right)^{-1}$
  step 3 若 $\left\|\nabla f\left(\vec{x}^{(k)}\right)\right\| < \varepsilon$, 则停止迭代；否则，令
  $$\vec{d}^{(k)} = -\nabla^2 f\left(\vec{x}^{(k)}\right)^{-1} \nabla\left(\vec{x}^{(k)}\right)$$
  step 4 从 $\vec{x}^{(k)}$ 出发，沿方向 $\vec{d}^{(k)}$ 作一维搜索，
  $$f(\vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}) = \min_{\lambda \geq 0} f(\vec{x}^{(k)} + \lambda \vec{d}^{(k)})$$
  令 $\vec{x}^{(k+1)} = \vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}$
  step 5 置 $k = k + 1$, 转**step 2**

- 修正 Newton 法

  step 1 给定初始点 $\vec{x}^{(1)}$, 允许误差 $\varepsilon > 0$, 置 $k = 1$.

step 2 计算 $\nabla f\left(\vec{x}^{(k)}\right), \nabla^2 f\left(\vec{x}^{(k)}\right)^{-1}$

step 3 若 $\left\|\nabla f\left(\vec{x}^{(k)}\right)\right\| < \varepsilon$, 则停止迭代；否则，置矩阵 $B_k = \nabla^2 f\left(\vec{x}^{(k)}\right) + E_k$, 其中 E_k 为修正矩阵（当 $\nabla^2 f\left(\vec{x}^{(k)}\right)$ 正定时，它取 0），令

$$\vec{d}^{(k)} = -B_k^{-1}\nabla\left(\vec{x}^{(k)}\right)$$

step 4 从 $\vec{x}^{(k)}$ 出发，沿方向 $\vec{d}^{(k)}$ 作一维搜索，

$$f(\vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}) = \min_{\lambda \geq 0} f(\vec{x}^{(k)} + \lambda \vec{d}^{(k)})$$

令 $\vec{x}^{(k+1)} = \vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}$

step 5 置 $k = k+1$, 转**step 2**

- 共轭梯度法

对于二次凸函数

$$f\left(\vec{x}\right) = \frac{1}{2}\vec{x}^{\mathrm{T}} A \vec{x} + b^{\mathrm{T}}\vec{x} + c$$

step 1 给定初始点 $\vec{x}^{(1)}$, 置 $k = 1$

step 2 计算 $g_k = \nabla f\left(\vec{x}^{(k)}\right)$. 若 $\|g_k\| = 0$, 停止迭代；否则，进行下一步

step 3 令

$$\vec{d}^{(k)} = -g_k + \beta_{k-1}\vec{d}^{(k-1)}$$

其中，

$$\beta_{k-1} = \begin{cases} 0 & k = 1 \\ \dfrac{\|g_k\|^2}{\|g_{k-1}\|^2} & k \geq 2 \end{cases}$$

step 4

$$\lambda_k = -\frac{g_k^{\mathrm{T}}\vec{d}^{(k)}}{\vec{d}^{(k)\mathrm{T}} A \vec{d}^{(k)}}$$

令 $\vec{x}^{(k+1)} = \vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}$

step 5 若 $k = n$, 则停止迭代；否则，置 $k = k+1$, 转**step 2**

- DFP 算法

step 1 给定初始点 $\vec{x}^{(1)} \in \mathbb{R}^n$, 允许误差 $\varepsilon > 0$

step 2 置 $H_1 = I_n$, 计算出在 $\vec{x}^{(1)}$ 处的梯度

$$g_1 = \nabla f\left(\vec{x}^{(1)}\right)$$

置 $k = 1$

step 3 令 $\vec{d}^{(k)} = -H_k g_k$

step 4 从 $\vec{x}^{(k)}$ 出发，沿方向 $\vec{d}^{(k)}$ 作一维搜索，

$$f(\vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}) = \min_{\lambda \geq 0} f(\vec{x}^{(k)} + \lambda \vec{d}^{(k)})$$

令 $\vec{x}^{(k+1)} = \vec{x}^{(k)} + \lambda_k \vec{d}^{(k)}$

step 5 检验是否满足收敛准则，若

$$\left\|\nabla f\left(\vec{x}^{(k+1)}\right)\right\| \leq \varepsilon$$

则停止迭代, 得到点 $\vec{x}^{(k+1)}$；否则，进行**step 6**

step 6 若 $k=n$, 则令 $\vec{\boldsymbol{x}}^{(1)}=\vec{\boldsymbol{x}}^{(k+1)}$, 返回**step 2**; 否则，进行**step 7**

step 7 令 $g_{k+1}=\nabla f\left(\vec{\boldsymbol{x}}^{(k+1)}\right)$, $\vec{\boldsymbol{p}}^{(k)}=\vec{\boldsymbol{x}}^{(k+1)}-\vec{\boldsymbol{x}}^{(k)}$, $\vec{\boldsymbol{q}}^{(k)}=\overrightarrow{\boldsymbol{g}_{k+1}}-\vec{\boldsymbol{g}_k}$

$$H_{k+1}=H_k+\frac{\vec{\boldsymbol{p}}^{(k)}\vec{\boldsymbol{p}}^{(k)\mathrm{T}}}{\vec{\boldsymbol{p}}^{(k)\mathrm{T}}\vec{\boldsymbol{q}}^{(k)}}-\frac{H_k\vec{\boldsymbol{q}}^{(k)}\vec{\boldsymbol{q}}^{(k)\mathrm{T}}H_k}{\vec{\boldsymbol{q}}^{(k)\mathrm{T}}H_k\vec{\boldsymbol{q}}^{(k)}}$$

置 $k=k+1$, 返回**step 3**

3.2 有约束极值问题

3.2.1 定义

$$\begin{aligned} \min \quad & f(\vec{\boldsymbol{x}}) \\ \text{s.t.} \quad & a_i(\vec{\boldsymbol{x}})=0, i=1,2,\ldots,p \\ & c_j(\vec{\boldsymbol{x}})\geq 0, j=1,2,\ldots,q \end{aligned}$$

其中 $a_i(\vec{\boldsymbol{x}})=0$ 称为等式约束，$c_j(\vec{\boldsymbol{x}})\geq 0$ 称为不等式约束。
其对应的 Lagrange 函数为

$$L\left(\vec{\boldsymbol{x}},\vec{\boldsymbol{\lambda}},\vec{\boldsymbol{\mu}}\right)=f(\vec{\boldsymbol{x}})-\sum_{i=1}^{p}\lambda_i a_i(\vec{\boldsymbol{x}})-\sum_{j=1}^{q}\mu_j c_j(\vec{\boldsymbol{x}})$$

3.2.2 极值条件

- KKT 条件
  若 $\overrightarrow{\boldsymbol{x}^*}$ 是极小点，则
  - $a_i(\overrightarrow{\boldsymbol{x}^*})=0, i=1,2,\ldots,p$
  - $c_j(\overrightarrow{\boldsymbol{x}^*})\geq 0, j=1,2,\ldots,q$
  - $\nabla f(\overrightarrow{\boldsymbol{x}^*})=\sum_{i=1}^{p}\lambda_i^*\nabla a_i(\overrightarrow{\boldsymbol{x}^*})+\sum_{j=1}^{q}\mu_j^*\nabla c_j(\overrightarrow{\boldsymbol{x}^*})$
  - $\mu_j^* c_j(\overrightarrow{\boldsymbol{x}^*})^{\mathrm{T}}=0, j=1,2,\ldots,q$
  - $\mu_j^*\geq 0, j=1,2,\ldots,q$

3.3 凸问题

3.3.1 定义

$$\begin{aligned} \min \quad & f(\vec{\boldsymbol{x}}) \\ \text{s.t.} \quad & a_i(\vec{\boldsymbol{x}})=0, i=1,2,\ldots,p \\ & c_j(\vec{\boldsymbol{x}})\geq 0, j=1,2,\ldots,q \end{aligned}$$

其中，f 为凸函数，$D=\{\vec{\boldsymbol{x}}\mid a_i(\vec{\boldsymbol{x}})=0, c_j(\vec{\boldsymbol{x}})\geq 0\}$ 是凸集。

3.3.2 极值条件

- 满足 KKT 条件的点一定为其极值点。
- Wolfe 对偶定理
  对于一个凸问题，定义其 Wolfe 对偶问题为

$$\begin{aligned} \max \quad & L(\vec{\boldsymbol{x}}) \\ \text{s.t.} \quad & \nabla_{\vec{\boldsymbol{x}}} L = 0 \\ & \mu_i \geq 0 \end{aligned}$$

  其中，L 为 f 的 Lagrange 函数。
  则此两个问题的极值点相同。

3.4 线性规划问题

3.4.1 定义

$$\begin{aligned} \min \quad & \vec{\boldsymbol{c}}^{\mathrm{T}} \vec{\boldsymbol{x}} \\ \text{s.t.} \quad & A\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}} \\ & \vec{\boldsymbol{x}} \geq \vec{\boldsymbol{0}} \end{aligned}$$

3.4.2 极值条件

- 原始对偶内点法
  设原问题的 Wolfe 对偶问题为

$$\begin{aligned} \max \quad & L(\vec{\boldsymbol{x}}) \\ \text{s.t.} \quad & \nabla_{\vec{\boldsymbol{x}}} L = 0 \\ & \mu_i > 0 \\ & \vec{\boldsymbol{x}} > 0 \end{aligned}$$

  其 KKT 条件为

$$\begin{aligned} A^{\mathrm{T}}\lambda + \mu = c \\ A\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}} \\ X\mu = \tau e \end{aligned}$$

  其中 $\tau > 0, e = [1, 1, \ldots, 1]^{\mathrm{T}}$

  step 1 在可行域里面找到一个初始点 $\vec{\boldsymbol{w}_0} = \{\vec{\boldsymbol{x}_0}, \vec{\boldsymbol{\lambda}_0}, \vec{\boldsymbol{\mu}_0}\}$

  step 2 在第 k 次迭代中，对 KKT 条件增加一阶扰动 $\tau = \tau_k$, 获得增量 δ_w 来更新下一个点 $\vec{\boldsymbol{w}_k}$. 解方程：

$$\begin{cases} A\delta_x = 0 \\ A\delta_\lambda + \delta_\mu = 0 \\ M\delta_x + X\delta_\mu = \tau_{k+1} - X\delta_\mu\delta_k \end{cases}$$

  其中 $M = \mathrm{diag}\{(\mu_k)_1, (\mu_k)_2, \ldots, (\mu_k)_n\}$

step 3 在下一轮迭代中的初始点更新为 $\overrightarrow{\boldsymbol{w}_{k+1}} = \{\overrightarrow{\boldsymbol{x}_k} + \overrightarrow{\boldsymbol{\delta}_x}, \overrightarrow{\boldsymbol{\lambda}_k} + \overrightarrow{\boldsymbol{\delta}_\lambda}, \overrightarrow{\boldsymbol{\mu}_k} + \overrightarrow{\boldsymbol{\delta}_\mu}\}$

step 4 在 KKT 条件中 τ 更新为更加靠近中心路径的 τ_{k+1}, 方法为

$$\tau_{k+1} = \frac{\mu_k^{\mathrm{T}} \overrightarrow{\boldsymbol{x}_k}}{n+\rho}$$

其中 $\rho > \sqrt{n}$

step 5 为了获得更好的 w_{k+1}, 可以采用一维搜索的方法

$$\overrightarrow{\boldsymbol{w}_{k+1}} = \overrightarrow{\boldsymbol{w}_k} + \alpha_k \overrightarrow{\boldsymbol{\delta}_w}$$

- 对偶形式

$$\begin{aligned} \max \quad & h(\overrightarrow{\boldsymbol{y}}) = b^{\mathrm{T}} \overrightarrow{\boldsymbol{y}} \\ \text{s.t.} \quad & c - A^{\mathrm{T}} \overrightarrow{\boldsymbol{y}} \geq 0 \end{aligned}$$

3.5 二次规划问题

3.5.1 定义

$$\begin{aligned} \min \quad & \overrightarrow{\boldsymbol{x}}^{\mathrm{T}} H \overrightarrow{\boldsymbol{x}} + \overrightarrow{\boldsymbol{p}}^{\mathrm{T}\overrightarrow{\boldsymbol{x}}} \\ \text{s.t.} \quad & \overrightarrow{\boldsymbol{a}_i}^{\mathrm{T}} \overrightarrow{\boldsymbol{x}} = \overrightarrow{\boldsymbol{b}_i}, i = 1, 2, \ldots, p \\ & \overrightarrow{\boldsymbol{c}_i}^{\mathrm{T}} \overrightarrow{\boldsymbol{x}} = \overrightarrow{\boldsymbol{d}_i}, i = 1, 2, \ldots, q \end{aligned}$$

考虑等式约束的二次规划问题

$$\begin{aligned} \min \quad & f(\overrightarrow{\boldsymbol{x}}) = \overrightarrow{\boldsymbol{x}}^{\mathrm{T}} H \overrightarrow{\boldsymbol{x}} + \overrightarrow{\boldsymbol{p}}^{\mathrm{T}\overrightarrow{\boldsymbol{x}}} \\ \text{s.t.} \quad & A\overrightarrow{\boldsymbol{x}} = \overrightarrow{\boldsymbol{b}} \end{aligned}$$

称为等式约束二次规划问题

3.5.2 极值条件

- Lagrange 法

  Lagrange 函数

$$L = f(\overrightarrow{\boldsymbol{x}}) - \overrightarrow{\boldsymbol{\lambda}}^{\mathrm{T}} \left(A\overrightarrow{\boldsymbol{x}} - \overrightarrow{\boldsymbol{b}} \right)$$

  求其 Langrage 矩阵

$$\begin{pmatrix} H & -A \\ -A & 0 \end{pmatrix} \begin{pmatrix} \overrightarrow{\boldsymbol{x}} \\ \overrightarrow{\boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} -\overrightarrow{\boldsymbol{p}} \\ -\overrightarrow{\boldsymbol{q}} \end{pmatrix}$$

3.6 半定问题

3.6.1 定义

原始 SDP 问题为

$$\begin{aligned} \min \quad & C \cdot X \\ \text{s.t.} \quad & A_i \cdot X = b_i, i = 1, 2, \ldots, p \\ & X \geq 0 \end{aligned}$$

3.6.2 对偶形式

$$
\begin{aligned}
\max \quad & b^{\mathrm{T}}\vec{\boldsymbol{y}} \\
\text{s.t.} \quad & C-\sum_{i=1}^{p} y_i A_i \geq 0
\end{aligned}
$$