数学优化

范数、可微性、展开

- **向量范数**

  若实值函数 $\|\cdot\|: \mathbb{R}^n \to \mathbb{R}$ 满足

  - $\forall \boldsymbol{x} \in \mathbb{R}^n, \|\boldsymbol{x}\| \geq 0$ 且 $\|\boldsymbol{x}\| = 0 \Longleftrightarrow \boldsymbol{x} = \mathbf{0}$
  - $\forall \alpha \in \mathbb{R}, \forall \boldsymbol{x} \in \mathbb{R}^n, \|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|$
  - $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, \|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

  则称其为向量范数，其中 $\mathbb{R}^n$ 表示 n 维向量空间

  设 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n$

  - L_1 范数：$\|\boldsymbol{x}\|_1 = \sum\limits_{i=1}^{m} |x_i|$
  - L_2 范数：$\|\boldsymbol{x}\|_2 = \left(\sum\limits_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$
  - L_∞ 范数：$\|\boldsymbol{x}\|_\infty = \max\limits_{i} |x_i|$
  - L_p 范数：$\|\boldsymbol{x}\|_p = \left(\sum\limits_{j=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$

- **矩阵范数**

  Frobenius 范数：$\|\boldsymbol{A}\|_F = \left(\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} |a_{ij}|^2\right)^{\frac{1}{2}} = \left(\operatorname{tr}\left(\boldsymbol{A}^T \boldsymbol{A}\right)\right)^{\frac{1}{2}}$

  $\boldsymbol{A} \in \mathbb{R}^{m \times n}$，诱导的矩阵范数为 $\|\boldsymbol{A}\| = \max\limits_{x \neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$

  设 $\boldsymbol{A} = (a_{ij})_{m \times n}$，$\lambda_{\boldsymbol{A}^T \boldsymbol{A}}$ 表示 $\boldsymbol{A}^T \boldsymbol{A}$ 的最大特征值

  - $\|\boldsymbol{A}\|_1 = \max\limits_{j} \sum\limits_{i=1}^{m} |a_{ij}|$
  - $\|\boldsymbol{A}\|_2 = \sqrt{\lambda_{\boldsymbol{A}^T \boldsymbol{A}}}$（谱范数）
  - $\|\boldsymbol{A}\|_\infty = \max\limits_{i} \sum\limits_{j=1}^{n} |a_{ij}|$

  若矩阵 $\boldsymbol{A}$ 正定，则条件数 $\operatorname{cond}(\boldsymbol{A}) = \|A\| \|A^{-1}\|$

  - 在 L_2 范数定义下，$\operatorname{cond}(\boldsymbol{A}) = \frac{\lambda_{\max}(\boldsymbol{A})}{\lambda_{\min}(\boldsymbol{A})}$
  - 条件数是判断矩阵病态与否的一种度量，条件数越大矩阵越病态

- **梯度**

  $\nabla f(\boldsymbol{x}) = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_n}\right]^T$

  f 在点 x^0 关于 p 的方向导数 $\frac{\partial f(x^0)}{\partial p} = \lim\limits_{t \to 0^+} \frac{f(x^0 + tp) - f(x^0)}{t}$

  - $\nabla \boldsymbol{b}^T \boldsymbol{x} = \boldsymbol{b}$
  - $\nabla \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} = \left(\boldsymbol{A} + \boldsymbol{A}^T\right) \boldsymbol{x} \xlongequal{\text{对称矩阵}} 2\boldsymbol{A}\boldsymbol{x}$
  - $\nabla \left(\frac{1}{2} \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{x} + \boldsymbol{c}\right) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$

- **Jacobi** 矩阵

$$\boldsymbol{h}'(\boldsymbol{x}) = \nabla \boldsymbol{h}(\boldsymbol{x})^T = \begin{pmatrix} \frac{\partial h_1(\boldsymbol{x})}{\partial x_1} & \frac{\partial h_1(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial h_1(\boldsymbol{x})}{\partial x_n} \\ \frac{\partial h_2(\boldsymbol{x})}{\partial x_1} & \frac{\partial h_2(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial h_2(\boldsymbol{x})}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial h_m(\boldsymbol{x})}{\partial x_1} & \frac{\partial h_m(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial h_m(\boldsymbol{x})}{\partial x_n} \end{pmatrix}$$

- **Hessian 矩阵**

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

- **Taylor 展开**
  - 一阶展开

    $f: \mathbb{R}^n \to \mathbb{R}$ 连续可微，向量 $\boldsymbol{p} \in \mathbb{R}^n$

    $f(\boldsymbol{x}+\boldsymbol{p}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{p} + o\left(\|\boldsymbol{p}\|\right)$
  - 二阶展开

    $f: \mathbb{R}^n \to \mathbb{R}$ 二阶连续可微

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

凸集

- **凸集的定义**
  - 仿射集

    设 S 为 n 维欧氏空间 $\mathbb{R}^n$ 中的一个集合。若对任意两点 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S$ 及每个实数 $\lambda \in \mathbb{R}$，有 $\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)} \in S$，则称 S 为仿射集。$\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}$ 称为 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}$ 的仿射组合

    - 泛化定义：$\forall x_1, x_2, \cdots, x_k \in S, \lambda_1, \lambda_2, \cdots, \lambda_k \in \mathbb{R}, \lambda_1 + \lambda_2 + \cdots + \lambda_k = 1$，有 $\lambda_1 x_1 + \lambda_2 x_2 + \cdots \lambda_k x_k \in S$
  - 仿射包

    $$\text{aff} S = \left\{ \lambda_1 x_1 + \lambda_2 x_2 + \cdots \lambda_k x_k \,\middle|\, \begin{array}{c} \forall x_1, x_2, \cdots, x_k \in S \\ \lambda_1 + \lambda_2 + \cdots + \lambda_k = 1 \end{array} \right\}$$
  - 凸集

    设 S 为 n 维欧氏空间 $\mathbb{R}^n$ 中的一个集合。若对任意两点 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S$ 及每个实数 $\lambda \in [0,1]$，有 $\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)} \in S$，则称 S 为凸集。$\lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}$ 称为 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}$ 的凸组合
  - 凸包

    $$\text{conv} C = \left\{ \lambda_1 x_1 + \lambda_2 x_2 + \cdots \lambda_k x_k \,\middle|\, \begin{array}{c} \forall x_1, x_2, \cdots, x_k \in S \\ \lambda_1 + \lambda_2 + \cdots + \lambda_k = 1 \\ \lambda_1, \lambda_2, \cdots, \lambda_k \in [0,1] \end{array} \right\}$$
  - 锥

设有集合 $C \subset \mathbb{R}^n$，若对没一点 $x \in C$，当 $\lambda > 0$ 时，都有 $\lambda x \in C$，称 C 为锥

- 凸锥

  任意两点 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S$ 及实数 $\lambda_1, \lambda_2 \geq 0$ 都有 $\lambda_1 \boldsymbol{x}^{(1)} + \lambda_2 \boldsymbol{x}^{(2)} \in C$，称 C 为凸锥

- **典型凸集**
  - 超平面

    $H = \left\{\boldsymbol{x} \mid \boldsymbol{p}^T \boldsymbol{x} = \alpha\right\}$，其中 $\boldsymbol{p}$ 为 n 维列向量，α 为实数

    - 半空间

      正的闭半空间 $H^+ = \left\{\boldsymbol{x} \mid \boldsymbol{p}^T \boldsymbol{x} \geq \alpha\right\}$

      负的闭半空间 $H^- = \left\{\boldsymbol{x} \mid \boldsymbol{p}^T \boldsymbol{x} \leq \alpha\right\}$

      正的开半空间 $\dot{H}^+ = \left\{\boldsymbol{x} \mid \boldsymbol{p}^T \boldsymbol{x} > \alpha\right\}$

      负的开半空间 $\dot{H}^- = \left\{\boldsymbol{x} \mid \boldsymbol{p}^T \boldsymbol{x} < \alpha\right\}$

  - 多面体

    有限个线性等式和不等式的解集

    有限个半空间与超平面的交集

  - 多胞形

    有限点集 $\left\{\boldsymbol{x}^0, \boldsymbol{x}^1, \cdots, \boldsymbol{x}^m\right\} \subset \mathbb{R}^n$ 的凸包称为多胞形

  - 单纯形

    $\left\{\boldsymbol{x}^0, \boldsymbol{x}^1, \cdots, \boldsymbol{x}^m\right\}$ 仿射无关，$\boldsymbol{x}^1 - \boldsymbol{x}^0, \cdots, \boldsymbol{x}^m - \boldsymbol{x}^0$ 线性无关，对应的凸包称为 m 维单纯形，向量 $\boldsymbol{x}^i$ 称为该单纯形的顶点

  - 极点

    非空凸集中的点 $\boldsymbol{x}$ 称为极点，若 $\boldsymbol{x} = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2, \lambda \in (0,1), \boldsymbol{x}_1, \boldsymbol{x}_2 \in S$，则 $\boldsymbol{x} = \boldsymbol{x}_1 = \boldsymbol{x}_2$

- **集合的保凸运算**
  - 保凸运算

    设 S_1, S_2 为 $\mathbb{R}^n$ 中的两个凸集，β 为实数，则

    - $\beta S_1 = \{\beta \boldsymbol{x} \mid \boldsymbol{x} \in S_1\}$ 为凸集
    - $S_1 + \beta = \{\boldsymbol{x} + \beta \mid \boldsymbol{x} \in S_1\}$ 为凸集
    - $S_1 \cap S_2$ 为凸集
    - $S_1 + S_2 = \left\{\boldsymbol{x}^{(1)} + \boldsymbol{x}^{(2)} \mid \boldsymbol{x}^{(1)} \in S_1, \boldsymbol{x}^{(2)} \in S_2\right\}$ 为凸集
    - $S_1 - S_2 = \left\{\boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)} \mid \boldsymbol{x}^{(1)} \in S_1, \boldsymbol{x}^{(2)} \in S_2\right\}$ 为凸集

  - 集合的保凸运算

    函数 $f: \mathbb{R}^n \to \mathbb{R}^m$ 是仿射的，当 $f(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}, A \in \mathbb{R}^{m \times n}, \boldsymbol{b} \in \mathbb{R}^m$，若 $S \in \mathbb{R}^n$ 为凸集，则 $f(S) = \{f(\boldsymbol{x}) \mid x \in S\}$ 为凸

凸函数

- **凸函数定义**

  设 $S \subset \mathbb{R}^n$ 是非空凸集，函数 $f: S \to \mathbb{R}$，若对任意 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$，每一 $\lambda \in (0,1)$ 都有 $f\left(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2\right) \leq \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2)$ 则称 f 是 S 上的凸函数

若对于 $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ 严格成立，则称 f 是 S 上的严格凸函数

  - 一阶条件定理

    设 $S \subset \mathbb{R}^n$ 是非空开凸集，函数 $f: S \rightarrow \mathbb{R}$ 是可微函数，则 $f(\boldsymbol{x})$ 是凸函数当且仅当对任意的 $\boldsymbol{x}^* \in S$，有 $\forall \boldsymbol{x} \in S, f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^T (\boldsymbol{x} - \boldsymbol{x}^*)$

    $f(\boldsymbol{x})$ 严格凸当且仅当 $\forall \boldsymbol{x} \in S, f(\boldsymbol{x}) > f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^T (\boldsymbol{x} - \boldsymbol{x}^*)$

    - 推广

      设 $S \subset \mathbb{R}^n$ 是非空开凸集，函数 $f: S \rightarrow \mathbb{R}$ 是可微函数，则 $f(\boldsymbol{x})$ 是凸函数当且仅当对任意的 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$，有 $(\nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2))^T (\boldsymbol{x}_2 - \boldsymbol{x}_1) \geq 0$

      $f(\boldsymbol{x})$ 严格凸当且仅当对任意的 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$，有 $(\nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2))^T (\boldsymbol{x}_2 - \boldsymbol{x}_1) > 0$

  - 二阶条件定理

    设 $S \subset \mathbb{R}^n$ 是非空凸集，函数 $f: S \rightarrow \mathbb{R}$ 的二次可微函数，则 $f(\boldsymbol{x})$ 是凸函数当且仅当 S 上每一点的 Hessian 矩阵是半正定的

    $f(\boldsymbol{x})$ 严格凸当且仅当 S 上每一点的 Hessian 矩阵是正定的

  - 切西瓜定理

    设 $S \subset \mathbb{R}^n$ 是非空凸集，函数 $f: S \rightarrow \mathbb{R}$ 的可微函数，则 $f(\boldsymbol{x})$ 是凸函数当且仅当对任意 $\boldsymbol{x} \in S$，任意的方向 $\boldsymbol{v} \in \mathbb{R}^n$，关于 t 的一元函数 $g(t) = f(\boldsymbol{x} + t\boldsymbol{v})$ 为凸函数，其中 $\mathrm{dom} g = \{t \mid \boldsymbol{x} + t\boldsymbol{v} \in \mathrm{dom} f\}$

- **常见函数的凹凸性**
- **函数的保凸运算**
  - 非负加权和：$f = w_1 f_1 + \cdots + w_m f_m$
  - 仿射映射：$g(x) = f(Ax + b)$
  - 逐点最大：$f(x) = \max\{f_1(x), \cdots, f_m(x)\}$
  - 函数透视：$g(x, t) = t f(x/t)$
  - 函数共轭：$f^*(y) = \sup\limits_{x \in \mathrm{dom} f} (y^T x - f(x))$
  - 复合函数：$f(x) = h(g(x))$，若 h 非减，则凹凸性一致；若 h 非增，则凹凸性相反
- **凸函数与凸集的关系**
  - 凸函数 f 的水平集 $L(f, \alpha) = \{\boldsymbol{x} \mid \boldsymbol{x} \in S, f(\boldsymbol{x}) \leq \alpha\}, \alpha \in \mathbb{R}$ 和上镜图 $\mathrm{epi}(f) = \{(x, y) \mid x \in S, y \in \mathbb{R}, y \geq f(x)\}$ 都是凸集
  - 拟凸函数：若函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 所有的 α-sublevel sets $S_\alpha \{\boldsymbol{x} \mid f(\boldsymbol{x}) \leq \alpha, \boldsymbol{x} \in \mathrm{dom} f\}$ 都为凸集合，则 f 为拟凸函数

    拟凹函数：若函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 所有的 α-superlevel sets $S_\alpha \{\boldsymbol{x} \mid f(\boldsymbol{x}) \geq \alpha, \boldsymbol{x} \in \mathrm{dom} f\}$ 都为凸集合，则 f 为拟凹函数

    若 $\mathrm{dom} f$ 为凸集且 $f(\theta \boldsymbol{x} + (1 - \theta) \boldsymbol{y}) \leq \max\{f(\boldsymbol{x}), f(\boldsymbol{y})\}$，则 f 为拟凸函数

凸优化问题

- **凸规划**
  - 目标函数：凸函数

    可行域：凸集

- 对凸问题，任何局部最优解都是全局最优解

- **等价形式**
  - 变量替换

    $\boldsymbol{x}=\phi(\boldsymbol{z}),\tilde{f}_i(\boldsymbol{z})=f_i(\phi(\boldsymbol{z}))$

    等价于

    $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}}\quad & \tilde{f}_0(\boldsymbol{z})\\ \text{subject to}\quad & \tilde{f}_i(\boldsymbol{z})\le 0\quad i=1,\cdots,m\\ & \tilde{h}_i(\boldsymbol{z})=0\quad i=1,\cdots,p\end{aligned}$$

  - 消除/引入等式约束

    $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}}\quad & f_0(\boldsymbol{x})\\ \text{subject to}\quad & f_i(\boldsymbol{x})\le 0\quad i=1,\cdots,m\\ & \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\end{aligned}$$

    $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\leftrightarrow\boldsymbol{x}=\boldsymbol{F}\boldsymbol{z}+\boldsymbol{x}_0$

    等价于

    $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}}\quad & f_0(\boldsymbol{F}\boldsymbol{z}+\boldsymbol{x}_0)\\ \text{subject to}\quad & f_i(\boldsymbol{F}\boldsymbol{z}+\boldsymbol{x}_0)\le 0\quad i=1,\cdots,m\end{aligned}$$

  - 引入松弛变量

    $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}}\quad & f_0(\boldsymbol{x})\\ \text{subject to}\quad & \boldsymbol{a}_i^T\boldsymbol{x}\le b_i\quad i=1,\cdots,m\end{aligned}$$

    等价于

    $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}}\quad & f_0(\boldsymbol{x})\\ \text{subject to}\quad & \boldsymbol{a}_i^T\boldsymbol{x}+s_i\le b_i\quad i=1,\cdots,m\\ & s_i\ge 0\qquad\qquad\quad i=1,\cdots,m\end{aligned}$$

  - 目标函数放到约束条件

    $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}}\quad & f_0(\boldsymbol{x})\\ \text{subject to}\quad & f_i(\boldsymbol{x})\le 0\quad i=1,\cdots,m\\ & \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\end{aligned}$$

    等价于

    $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}}\quad & t\\ \text{subject to}\quad & f_0(\boldsymbol{x})-t\le 0\\ & f_i(\boldsymbol{x})\le 0\qquad i=1,\cdots,m\\ & \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\end{aligned}$$

  - 先优化某些变量

    $$\begin{aligned}\underset{\boldsymbol{x},\boldsymbol{y}}{\text{minimize}}\quad & f_0(\boldsymbol{x},\boldsymbol{y})\\ \text{subject to}\quad & f_i(\boldsymbol{x})\le 0\quad i=1,\cdots,m\end{aligned}$$

    $\inf_{\boldsymbol{x},\boldsymbol{y}}f(\boldsymbol{x},\boldsymbol{y})=\inf_{\boldsymbol{x}}\tilde{f}(\boldsymbol{x})$

    $\tilde{f}(\boldsymbol{x})=\inf_{\boldsymbol{y}}f(\boldsymbol{x},\boldsymbol{y})$

    等价于

$$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}} \quad & \tilde{f}_0(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \le 0 \quad i=1,\cdots,m\end{aligned}$$

- **线性规划（LP）**
  - 标准形式：

    $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}} \quad & \boldsymbol{c}^T\boldsymbol{x}+d \\ \text{subject to} \quad & \boldsymbol{G}\boldsymbol{x} \le \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\end{aligned}$$

  - 可行域是一个多面体

二次规划

- **二次规划（QP）**

  $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}} \quad & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}+\boldsymbol{q}^T\boldsymbol{x}+r \\ \text{subject to} \quad & \boldsymbol{G}\boldsymbol{x} \le \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\end{aligned}$$

- **二次约束二次规划（QCQP）**

  $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}} \quad & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{P}_0\boldsymbol{x}+\boldsymbol{q}_0^T\boldsymbol{x}+r_0 \\ \text{subject to} \quad & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{P}_i\boldsymbol{x}+\boldsymbol{q}_i^T\boldsymbol{x}+r_i \le 0 \quad i=1,\cdots,m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\end{aligned}$$

- **半正定规划（SDP）**

  $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}} \quad & \operatorname{tr}(\boldsymbol{C}\boldsymbol{X}) \\ \text{subject to} \quad & \operatorname{tr}(\boldsymbol{A}_i\boldsymbol{X}) = b_i \\ & \boldsymbol{X} \succeq 0\end{aligned}$$

  $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}} \quad & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} \quad & x_1\boldsymbol{F}_1+\cdots x_n\boldsymbol{F}_n \preceq \boldsymbol{G} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\end{aligned}$$

对偶问题

- **拉格朗日**

  标准形式最优化问题

  $$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0 \quad i=1,\cdots,m \\ & h_i(x) = 0 \quad i=1,\cdots,p\end{aligned}$$

  $D=\bigcap_{i=0}^m \operatorname{dom} f_i \cap \bigcap_{i=0}^p \operatorname{dom} h_i$

  拉格朗日：目标函数和约束函数的加权和

  $$L(x,\lambda,v)=f_0(x)+\sum_{i=1}^m \lambda_i f_i(x)+\sum_{i=1}^p v_i h_i(x)$$

- **对偶函数**

  $g: x \in R^m \times R^p \to R$

$$
\begin{aligned}
g(\lambda, v) &= \inf_{x \in D} L(x, \lambda, v) \\
&= \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x) \right)
\end{aligned}
$$

对偶函数总是凹函数

- **对偶问题**

  $$
  \begin{aligned}
  &\underset{\boldsymbol{x}}{\text{minimize}} \quad && g(\lambda, v) \\
  &\text{subject to} \quad && \lambda \geq 0
  \end{aligned}
  $$

  设原问题最优化目标函数值为 p^*，对偶问题为 d^*，则 $d^* \leq p^*$

  对偶问题总是凸问题

- **弱对偶和强对偶**
  - 弱对偶

    $d^* \leq p^*$

    对偶间隙 $d^* - p^*$

  - 强对偶

    $d^* = p^*$

    Slater 约束规范：对凸优化问题，只要有一个严格可行点，强对偶即成立

最优性条件

- **KKT 条件**
  - 原问题可行

    $f_i(\boldsymbol{x}^*) \leq 0, i = 1, \cdots, m, h_i(\boldsymbol{x}^*) = 0, i = 1, \cdots, p$

  - 对偶问题可行

    $\boldsymbol{\lambda}^* \geq 0$

  - 互补松弛条件

    $\forall i, \lambda_i^* f_i(\boldsymbol{x}^*) = 0$

  - 拉格朗日对 $\boldsymbol{x}$ 求导为零

    $$\nabla f_0(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\boldsymbol{x}^*) + \sum_{i=1}^{p} v_i \nabla h_i(\boldsymbol{x}^*) = 0$$

算法

0.618 法

- **单峰函数**：若 $\exists \alpha^* \in [a, b]$，使得 $f(x)$ 在 $[a, \alpha^*]$ 上严格递减，在 $[\alpha^*, b]$ 上严格递增，则 $f(x)$ 是 $[a, b]$ 上的单峰函数
- 设 $\varphi(\alpha)$ 是搜索区间 $[a_1, b_1]$ 上的单峰函数，设在第 k 次迭代时搜索区间为 $[a_k, b_k]$。取两个试探点 $\lambda_k, \mu_k \in [a_k, b_k], \lambda_k < \mu_k$，计算 $\varphi(\lambda_k), \varphi(\mu_k)$

- 若 $\varphi(\lambda_k) \leq \varphi(\mu_k)$，令 $a_{k+1} = a_k, b_{k+1} = \mu_k$
- 若 $\varphi(\lambda_k) > \varphi(\mu_k)$，令 $a_{k+1} = \lambda_k, b_{k+1} = b_k$

牛顿法

$x^{(k+1)} = x^{(k)} - \frac{f'\left(x^{(k)}\right)}{f''\left(x^{(k)}\right)}$

割线法

$x^{(k+1)} = x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f'(x^{(k)}) - f'\left(x^{(k-1)}\right)} f'\left(x^{(k)}\right)$

梯度下降法

- 给定初始点 $x^{(1)}$
- 计算负梯度方向 $d^{(k)} = -\nabla f\left(x^{(k)}\right)$
- 计算步长 λ，满足 $\varphi(\lambda_k) = f\left(x^{(1)} + \lambda_k d^{(k)}\right)$，$\varphi'(\lambda_k) = 0$
- 更新 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$
- 中止准则：$\|\nabla f(x)\|_2 \leq \epsilon$

牛顿法

$x^{(k+1)} = x^{(k)} - \left(\nabla^2 f\left(x^{(k)}\right)\right)^{-1} \nabla f\left(x^{(k)}\right)$