

目录

| | | |
|----------|---------------------|-----------|
| 1 | Introduction | 3 |
| 1.1 | 基本概念 | 3 |
| 1.2 | 凸集 | 4 |
| 1.3 | 凸集分离定理 | 5 |
| 1.4 | 凸函数 | 7 |
| 2 | LP 基本性质 | 9 |
| 2.1 | LP 标准形 | 9 |
| 2.2 | LP 的图解法 | 11 |
| 2.3 | LP 的基本性质 | 11 |
| 3 | 单纯形法 | 14 |
| 3.1 | 单纯形法 | 14 |
| 3.2 | 两阶段法 | 17 |
| 3.3 | 线性规划的最优性条件 | 17 |
| 4 | 对偶理论 | 19 |
| 4.1 | LP 对偶问题的形式 | 19 |
| 4.2 | 对偶问题的基本性质 | 19 |
| 4.3 | 对偶问题的经济解释 | 22 |
| 4.4 | 对偶单纯形法 | 22 |
| 4.5 | 原 - 对偶算法 | 23 |
| 5 | 算法概述 | 24 |
| 5.1 | 算法分类 | 24 |
| 5.2 | 算法收敛性 | 24 |
| 5.3 | 算法复杂度 | 25 |
| 6 | 非线性规划的最优性条件 | 27 |
| 6.1 | 无约束优化的最优性条件 | 27 |
| 6.2 | 约束优化问题的最优性条件 | 28 |
| 6.3 | 一般约束问题的一阶最优性条件 | 29 |
| 6.4 | 约束优化问题的二阶最优性条件 | 29 |
| 7 | Lagrange 对偶 | 31 |
| 7.1 | Lagrange 对偶 | 31 |
| 7.2 | 对偶定理 | 31 |
| 7.3 | 鞍点问题 | 31 |
| 7.4 | Lagrange 乘子的经济学解释 | 31 |

| | | |
|-----------|-------------------------|-----------|
| 8 | 使用导数的最优化方法 | 32 |
| 8.1 | 最速下降法 | 32 |
| 8.2 | 牛顿法 | 33 |
| 8.3 | 共轭梯度法 | 34 |
| 9 | 罚函数法 | 36 |
| 9.1 | 外点罚函数法 | 36 |
| 9.2 | 内点罚函数法 | 38 |
| 9.3 | 乘子罚函数法 | 40 |
| 9.3.1 | 等式约束优化的乘子罚函数法 | 40 |
| 10 | 最小二乘问题 | 42 |

重点内容

总成绩 = 平时成绩 (30%) + 小论文 (20%) + 期末考试成绩 (50%)

| | |
|--|----|
| 最优化问题简介、发展史、分类 无约束优化理论 无约束优化的最优性条件 | |
| 算法概述和基础 使用导数的最优化方法, 最速下降法, 牛顿法 | 重点 |
| 牛顿法, 拟牛顿法 | 重点 |
| 共轭梯度法 | 重点 |
| 最小二乘法 | |
| 约束优化理论 约束优化的最优性条件, Lagrange 乘子 | 重点 |
| 约束优化的最优性条件: 二阶条件, 凸规划 | 重点 |
| 凸规划的性质、对偶理论、鞍点定理 | 重点 |
| 线性规划的基本性质, 极点和基解 | |
| 线性规划的单纯形方法 | |
| 约束优化问题的罚函数法 | 重点 |
| 增广 Lagrangian 方法/乘子罚函数方法 | |

1 Introduction

1.1 基本概念

定义 1.1 (二次型的正定性). 给定二次型 $f(X) = X^t A X$, 若对 $\forall X \neq 0$, 都有 $f(X) = X^t A X > 0$ 成立, 则称 $f(X)$ 为正定二次型, A 为正定矩阵。

定理 1.1. 对于 n 阶实对称矩阵 A , 下列命题等价:

- $X^t A X$ 是正定二次型 (或 A 是正定矩阵)
- A 的 n 个顺序主子式都大于 0
- A 的 n 个特征值都大于 0
- 存在可逆矩阵 P , 使得 $A = P^t P$

定义 1.2 (二次型的半正定性). 给定二次型 $f(X) = X^t A X$, 若对 $\forall X \neq 0$, 都有 $f(X) = X^t A X \geq 0$ 成立, 则称 $f(X)$ 为半正定二次型, A 为半正定矩阵。

定理 1.2. 对于 n 阶实对称矩阵 A , 下列命题等价:

- $A^t A X$ 是半正定二次型 (或 A 是半正定矩阵)
- A 的所有主子式都大于等于 0, 而且至少有一个等于 0
- A 的 n 个特征值都大于等于 0, 而且至少有一个等于 0

1.2 凸集

定义 1.3 (凸集). 设 $S \subseteq E^n$, 若对 $\forall x, y \in S, \forall \lambda \in [0, 1]$, 都有 $\lambda x + (1 - \lambda)y \in S$, 则称 S 为凸集。

• $x_1, \dots, x_k \in S, \lambda_1 + \dots + \lambda_k = 1$, 称 $\sum_{i=1}^k \lambda_i x_i$ 为 x_1, \dots, x_k 的凸组合。

定理 1.3. 设 S_1 和 S_2 为 E^n 中的两个凸集, β 是实数, 则

1. $\beta S_1 = \{\beta x \mid x \in S_1\}$
2. $S_1 \cap S_2$
3. $S_1 + S_2 = \{x^{(1)} + x^{(2)} \mid x^{(1)} \in S_1, x^{(2)} \in S_2\}$
4. $S_1 - S_2 = \{x^{(1)} - x^{(2)} \mid x^{(1)} \in S_1, x^{(2)} \in S_2\}$

都是凸集。

定义 1.4 (凸锥). 给定集合 $C \subset E^n$, 若对 C 中每一个 x , 及任意的 $\lambda \geq 0$, 都有 $\lambda x \in C$, 则称 C 为锥; 若 C 为凸集, 则称 C 为凸锥。

定义 1.5 (极点). S 是非空凸集, $x \in S$, 若由 $x = \lambda x_1 + (1 - \lambda)x_2$, 其中 $\lambda \in (0, 1), x_1, x_2 \in S$, 必推出 $x = x_1 = x_2$, 则称 x 是 S 的极点。

定义 1.6 (极方向). S 是 E^n 中的闭凸集, $d \in E^n, d \neq 0$, 如果对 $\forall x \in S$, 有

$$\{x + \lambda d \mid \lambda > 0\} \subset S$$

则称向量 d 为 S 的方向。

若 S 的方向 d 不能表示为集合的两个不同方向的正线性组合, 则称 d 为 S 的极方向。

例 1.1. 给定 $S = \{(x_1, x_2)^t \mid x_2 \geq |x_1|\}$, $d^{(1)} = (1, 1)^t, d^{(2)} = (-1, 1)^t$, 证明 $d^{(1)}, d^{(2)}$ 是 S 的极方向。

证明. $\forall x = (x_1, x_2) \in S, x_2 \geq |x_1|$,

$$x + \lambda d^{(1)} = (x_1 + \lambda, x_2 + \lambda).$$

由

$$x_2 + \lambda \geq |x_1| + \lambda \geq |x_1 + \lambda|$$

可得,

$$x + \lambda d^{(1)} \in S.$$

故 $d^{(1)}$ 是 S 的方向。

设

$$d^{(1)} = \lambda_1 \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \lambda_2 \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

其中 $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ 和 $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ 是 S 的方向, 可得

$$u_1 - u_2 = \frac{\lambda_2}{\lambda_1}(v_2 - v_1)$$

$$u_2 \geq |u_1|$$

$$v_2 \geq |v_1|$$

故 $u_1 = u_2, v_1 = v_2$, $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ 和 $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ 是两个相同的方向。所以 $d^{(1)}$ 是 S 的极方向。同理可证 $d^{(2)}$ 是 S 的极方向。 \square

定理 1.4 (多面集表示定理). 设 $S = \{x \mid Ax = b, x \geq 0\}$ 为非空多面集, 则有

- 极点集非空, 且存在有限个极点;
- 极方向集合为空集的充要条件是 S 有界; 若 S 无界, 则存在有限个极方向;
- $x \in S$ 的充要条件是

$$x = \sum_{j=1}^k \lambda_j x^{(j)} + \sum_{j=1}^l \mu_j d^{(j)}$$

$$\lambda_j \geq 0, \forall j = 1, \dots, k$$

$$\sum_{j=1}^k \lambda_j = 1$$

$$\mu_j \geq 0, \forall j = 1, \dots, l.$$

例 1.2. 用定义验证下列集合为凸集。

$$S = \{(x_1, x_2) \mid x_1 + 2x_2 \geq 1, x_1 - x_2 \geq 1\}$$

答案. $\forall x, y \in S$, 设 $\lambda \in [0, 1]$, $z = \lambda x + (1 - \lambda)y$.

则

$$z_1 = \lambda x_1 + (1 - \lambda)y_1$$

$$z_2 = \lambda x_2 + (1 - \lambda)y_2$$

$$z_1 + 2z_2 = \lambda(x_1 + 2x_2) + (1 - \lambda)(y_1 + 2y_2) \geq \lambda + (1 - \lambda) \geq 1$$

$$z_1 + 2z_2 = \lambda(x_1 - x_2) + (1 - \lambda)(y_1 - y_2) \geq \lambda + (1 - \lambda) \geq 1$$

故 $z \in S$, S 为凸集。

1.3 凸集分离定理

定理 1.5 (凸集分离定理).

- 设 S 为 E^n 的闭凸集, $y \notin S$, 则存在唯一的 $\bar{x} \in S$, 使得

$$\|y - \bar{x}\| = \inf_{x \in S} \|y - x\| > 0$$

\bar{x} 是这一最小距离点 $\iff (y - \bar{x})^t(\bar{x} - x) \geq 0, \forall x \in S$.

- 设 S 是 E^n 的非空闭凸集, $y \notin S$, 则存在非零向量 p 以及数 $\varepsilon > 0^1$, 使得对 $\forall x \in S$, 有 $p^t y \geq \varepsilon + p^t x$.
- 设 S 是 E^n 的非空凸集, $y \in \partial S$, 则存在非零向量 p , 使得对 $\forall x \in cl S$ (S 的闭包, 由 S 的内点和边界点组成), 有 $p^t y \geq p^t x$.
- 设 S_1 和 S_2 是 E^n 的两个非空凸集, $S_1 \cap S_2 = \emptyset$, 则存在非零向量, 使得

$$p^t y \geq p^t x \quad \text{其中 } \forall y \in S_1, \forall x \in S_2$$

定理 1.6. 两个系统恰有一个有解:

- *Farkas* 引理: 设 A 为 $m \times n$ 矩阵, c 为 n 维列向量, 则 $Ax \leq 0, c^t x > 0$ 有解的充分条件是 $A^t y = c, y \geq 0$ 无解。
- *Gordan* 定理: 给定矩阵 $A \in \mathbb{R}^{m \times n}$, 下列两个系统只有一个有解

$$Ax < 0$$

$$y \geq 0, y \neq 0, A^t y = 0$$

例 1.3. 证明 $Ax \leq 0, c^t x > 0$ 有解。其中

$$A = \begin{bmatrix} 1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix}, c = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

证明. 证明 $Ax \leq 0, c^t x > 0$ 有解, 即证 $A^t y = c, y \geq 0$ 无解。

$$A^t y = c$$

$$\begin{bmatrix} 1 & -1 \\ -2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

即

$$\begin{cases} y_1 - y_2 & = 2 \\ -2y_1 + y_2 & = 1 \\ y_1 + y_2 & = 0 \end{cases}$$

该方程组无解, 故原系统有解。 □

例 1.4. 设 A 是 $m \times n$ 矩阵, B 是 $l \times n$ 矩阵, $c \in \mathbb{R}^n$, 证明下列两个系统恰有一个有解:

- 系统 1: $Ax \leq 0, Bx = 0, c^t x > 0$, 对某些 $x \in \mathbb{R}^n$
- 系统 2: $A^t y + B^t z = c, y \geq 0$, 对某些 $y \in \mathbb{R}^m$ 和 $z \in \mathbb{R}^l$

证明. 证系统 1 有解, 即

$$\begin{bmatrix} A \\ B \\ -B \end{bmatrix} x \leq 0, c^t x > 0$$

¹ 令 $p = y - \bar{x}$, $\varepsilon = p^t(y - \bar{x})$, 有 $p^t(y - x) = p^t(y - \bar{x} + \bar{x} - x) = \varepsilon + (y - \bar{x})^t(\bar{x} - x) \geq \varepsilon$. 即可证明.

有解，则根据 Farkas 定理，有

$$\begin{pmatrix} A^t & B^t & -B^t \end{pmatrix} \begin{bmatrix} y \\ z_1 \\ z_2 \end{bmatrix} = c, \begin{bmatrix} y \\ z_1 \\ z_2 \end{bmatrix} \geq 0$$

无解，即 $A^t y + B^t z_1 - B^t z_2 = c, y \geq 0, z_1 \geq 0, z_2 \geq 0$ 无解，即 $A^t + B^t z = c, y \geq 0$ 无解。

反之，若 $A^t + B^t z = c, y \geq 0$ 有解，即 $A^t y + B^t z_1 - B^t z_2 = c, y \geq 0, z_1 \geq 0, z_2 \geq 0$ 有解，根据 Farkas 定理，有

$$\begin{bmatrix} A \\ B \\ -B \end{bmatrix} x \leq 0, c^t x > 0$$

无解，即 $Ax \leq 0, Bx = 0, c^t x > 0$ 无解。 □

1.4 凸函数

定义 1.7 (凸函数). 设 S 是 E^n 中的非空凸集， $f(x)$ 是定义在 S 上的实函数，如果对于每一对 $x_1, x_2 \in S$ 及每一个 $\lambda, 0 \leq \lambda \leq 1$ 都有

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

则称函数 $f(x)$ 为 S 上的凸函数。

定理 1.7. 凸函数的性质：

- 设 $f_1(x), f_2(x)$ 是凸集 S 上的凸函数，则函数 $f_1(x) + f_2(x)$ 在 S 上也是凸函数。
- 设 $f(x)$ 是凸集 S 上的凸函数，则对任意的 $a \geq 0$ ，函数 $af(x)$ 是凸的。
- 设 $f(x)$ 是凸集 S 上的凸函数，对每一个实数 c ，则集合

$$S_c = \{x \mid x \in S, f(x) \leq c\}$$

是凸集。

- 凸函数的根本重要性：设 S 是 E^n 中的非空凸集， f 是定义在 S 上的凸函数，则 f 在 S 上的局部极小点是整体极小点，且极小点的集合是凸集。

定理 1.8. 凸函数的判别：

- (一阶充要条件) 设 S 是 E^n 中的非空开凸集， $f(x)$ 是定义在 S 上的可微函数，则 $f(x)$ 为凸函数的充要条件是对任意两点 $x_1, x_2 \in S$ ，有

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^t (x_2 - x_1)$$

$f(x)$ 为严格凸函数的充要条件是对任意互不相同两点 $x_1, x_2 \in S$ ，有

$$f(x_2) > f(x_1) + \nabla f(x_1)^t (x_2 - x_1)$$

几何意义： $f(x)$ 是凸函数当且仅当任意点处的切线增量不超过函数的增量。

- (二阶充要条件) 设 S 是 E^n 中的非空开凸集， $f(x)$ 是定义在 S 上的二次可微函数，则 $f(x)$ 为凸函数的充要条件是对任意 $x \in S$ ， $f(x)$ 在 x 处的 Hessian 矩阵 $\nabla^2 f(x)$ 是半正定的。
 $f(x)$ 为严格凸函数的充要条件是对任意 $x \in S$ ， $f(x)$ 在 x 处的 Hessian 矩阵 $\nabla^2 f(x)$ 是正定的。

笔记 1.1. 设 $f(x)$ 是定义在凸集 S 上的可微凸函数, 若 $\exists x^* \in S$, 使对 $\forall x \in S$, 都有

$$\nabla f(x^*)^t(x - x^*) \geq 0$$

则 x^* 是 $f(x)$ 在凸集 S 上的全局极小点。

例 1.5. 判断下列函数是否为凸函数。

1. $f(x_1, x_2) = x_1^2 - 2x_1x_2 + x_2^2 + x_1 + x_2$
2. $f(x_1, x_2) = x_1e^{-(x_1+x_2)}$
3. $f(x_1, x_2, x_3) = x_1x_2 + 2x_1^2 + x_2^2 + 2x_3^2 - 6x_1x_3$

答案. 计算 $\nabla^2 f(x) \succeq 0$ 即可。

定义 1.8. 凸规划: 求凸函数在凸集上的极小点。

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i = 1, \dots, m \\ & h_j(x) = 0, j = 1, \dots, l \end{aligned}$$

若 $f(x)$ 是凸函数, $g_i(x)$ 是凸函数, $h_j(x)$ 是线性函数, 则原问题为凸规划。

- 凸规划的局部极小点就是整体极小点
- 且极小点的集合为凸集

2 LP 基本性质

LP: 线性规划.

2.1 LP 标准形

定义 2.1. LP 标准形:

1. 极小化型
2. 约束方程为等式
3. 所有的决策变量为非负值
4. 约束方程的右端项系数为非负值

$$\begin{aligned} \min \quad & z = c^t x \\ \text{s.t.} \quad & Ax = b \\ & x \succeq 0 \\ & c, x \in \mathbb{R}^n, b \in \mathbb{R}_+^m, A \in \mathbb{R}^{m \times n} \end{aligned}$$

笔记 2.1. 非标准形 LP 模型转化为标准形 LP 模型

- 目标函数是极大值的转化

$$\max c^t x \implies \min -c^t x$$

- 决策变量无约束转化为非负约束

$$x = x^+ - x^-, \quad x^+, x^- \geq 0$$

- 不等约束转化为等式约束 (松弛变量)

$$a^t x \leq b \implies a^t x + s = b, \quad s \geq 0$$

$$a^t x \geq b \implies a^t x - s = b, \quad s \geq 0$$

- 决策变量有上下界的转换

$$1 \leq x_1 \leq 6$$

$$x_1' = x_1 - 1 \geq 0, x_1' + x_2 = 5, x_2 \geq 0$$

- 带绝对值

$$\begin{cases} x_1 = \frac{x+|x|}{2} \\ x_2 = \frac{x-|x|}{2} \end{cases}$$

则 $x_1, x_2 \geq 0$, $x = x_1 - x_2$, $|x| = x_1 + x_2$.

例 2.1.

$$\begin{cases} \max & 3x_1 - 2x_2 + 3 \\ s.t. & x_1 + x_2 \leq 7 \\ & x_1 - x_2 + x_3 \geq 5 \\ & 1 \leq x_3 \leq 6 \\ & x_1 \geq 0, x_2 \text{ free} \end{cases}$$

转化为 LP 模型.

答案.

$$\begin{cases} \min & -3x_1 + 2(x_2^+ - x_2^-) - (x_3' + 1) \\ s.t. & x_1 + (x_2^+ - x_2^-) + x_4 = 7 \\ & x_1 - (x_2^+ - x_2^-) + x_3' - x_5 = 4 \\ & x_3' + x_6 = 5 \\ & x_1, x_2^+, x_2^-, x_3', x_4, x_5, x_6 \geq 0 \end{cases}$$

例 2.2.

$$\begin{cases} \max & -|x| - |y| \\ s.t. & x + y \geq 2 \\ & x \leq 3 \end{cases}$$

转换为 LP 模型.

答案.

$$\begin{cases} x_1 = \frac{x+|x|}{2} \\ x_2 = \frac{x-|x|}{2} \end{cases}, \begin{cases} y_3 = \frac{y+|y|}{2} \\ y_4 = \frac{y-|y|}{2} \end{cases}$$

则

$$x_1 \geq 0, x_2 \geq 0, x = x_1 - x_2, |x| = x_1 + x_2$$

$$x_3 \geq 0, x_4 \geq 0, y = x_3 - x_4, |y| = x_3 + x_4$$

$$\begin{cases} \min & (x_1 + x_2) + (x_3 + x_4) \\ s.t. & x_1 - x_2 + x_3 - x_4 - x_5 = 2 \\ & x_1 - x_2 + x_6 = 3 \\ & x_i \geq 0, i = 1, \dots, 6 \end{cases}$$

例 2.3. 将下列线性规划变成标准形式

$$\begin{cases} \min & 3x_1 - x_2 \\ s.t. & x_1 + x_2 \leq 9 \\ & 1 \leq x_1 \leq 5 \\ & 0 \leq x_2 \leq 6 \end{cases}$$

答案.

$$\begin{cases} \min & 3(x'_1 + 1) - x_2 \\ s.t. & x'_1 + x_2 + x_3 = 8 \\ & x'_1 + x_4 = 4 \\ & x_2 + x_5 = 6 \\ & x'_1, x_2, x_3, x_4, x_5 \geq 0 \end{cases}$$

2.2 LP 的图解法

笔记 2.2. 线性不等式的几何意义——半平面.

图解法步骤:

1. 作出 LP 问题的可行域
2. 作出目标函数的等值线
3. 移动等值线到可行域边界得到最优点

定理 2.1. 若 LP 问题存在最优解, 则必在可行域的某个极点上找到。

2.3 LP 的基本性质

定义 2.2 (可行解). 满足 LP 模型的约束条件且满足非负条件的解是可行解.

例 2.4.

$$\begin{cases} \max & z = 3x_1 + 2x_2 \\ s.t. & x_1 + 3x_2 \geq 6 \\ & x_1 - 2x_2 \leq 4 \\ & x_1, x_2 \geq 0 \end{cases}$$

判断 $X = (5, 1)^t$, $X = (-1, 3)^t$, $X = (2, 1)^t$ 是否为可行解。

答案. 只有 $X = (5, 1)^t$ 是可行解。

定理 2.2. 线性规划的可行域是凸集。

定理 2.3. 设线性规划的可行域非空, 则

1. LP 存在有限最优解的充要条件是对任意的 j , $cd^{(j)} \geq 0$, 其中 $d^{(j)}$ 为可行域的极方向, c 为目标函数的系数。
2. 若 LP 存在有限最优解, 则目标函数的最优值可在某个极点达到。

定义 2.3 (基矩阵和基变量). 对 LP 问题

$$\begin{cases} \min & z = cx \\ s.t. & Ax = b \\ & x \geq 0 \end{cases}$$

将 A 按列分块为 (P_1, \dots, P_n) , 则 $Ax = b$ 等价于

$$P_1x_1 + \dots + P_nx_n = b.$$

系数矩阵 A 中任意 m 列所组成的 m 阶可逆子方阵 B , 称为 LP 的一个基 (矩阵), 变量 x_j 对应的 P_j 包含在基 B 中, 则称 x_j 为基变量, 否则称为非基变量。

基变量的个数最多为 $\binom{n}{m}$ 。

定义 2.4 (基本解). 接定义 2.3, 设 $A = \begin{bmatrix} B & N \end{bmatrix}$, 其中 $r(B) = m$, 设 $x = \begin{bmatrix} x_B & x_N \end{bmatrix}$.

由 $Ax = b$ 得

$$Bx_B + Nx_N = b$$

解得

$$x_B = B^{-1}b - B^{-1}Nx_N$$

令 $x_N = 0$, 得

$$x = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$$

称 x 为 LP 的基本解。

定义 2.5 (基本可行解). 接定义 2.3, 若 $B^{-1}b \geq 0$, 则称 $x = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$ 为 LP 的基本可行解, B 称为可行基矩阵, x_{B_1}, \dots, x_{B_m} 为一组可行基。

若 $B^{-1}b > 0$, 则称基本可行解是非退化的, 否则称为退化的。

例 2.5. 优化问题的约束条件为

$$\begin{cases} x_1 + 3x_2 \geq 6 \\ x_1 - 2x_2 \leq 4 \\ x_1, x_2 \geq 0 \end{cases}$$

首先引入松弛变量, 转化为等式约束

$$\begin{cases} x_1 + 3x_2 - x_3 = 6 \\ x_1 - 2x_2 + x_4 = 4 \end{cases}$$

系数矩阵

$$A = \begin{pmatrix} 1 & 3 & -1 & 0 \\ 1 & -2 & 0 & 1 \end{pmatrix}$$

基矩阵为

$$B_1 = \begin{pmatrix} 1 & 3 \\ 1 & -2 \end{pmatrix}, B_2 = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}, B_3 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \\ B_4 = \begin{pmatrix} 3 & -1 \\ -2 & 0 \end{pmatrix}, B_5 = \begin{pmatrix} 3 & 0 \\ -2 & 1 \end{pmatrix}, B_6 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

对于 B_1 求基解

$$B_1^{-1}b = -\frac{1}{5} \begin{pmatrix} -2 & -3 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 6 \\ 4 \end{pmatrix} = \begin{pmatrix} \frac{24}{5} \\ \frac{2}{5} \end{pmatrix}$$

基本解为 $x^{(1)} = (\frac{24}{5}, \frac{2}{5}, 0, 0)^t$.

同理可得

$$x^{(1)} = (\frac{24}{5}, \frac{2}{5}, 0, 0)^t, x^{(2)} = (4, 0, -2, 0)^t, x^{(3)} = (6, 0, 0, -2)^t \\ x^{(4)} = (0, -2, -12, 0)^t, x^{(5)} = (0, 2, 0, 8)^t, x^{(6)} = (0, 0, -6, 4)^t$$

只有 $x^{(1)}$ 和 $x^{(5)}$ 为基本可行解。

定理 2.4. 基本可行解与极点之间的关系：

1. 可行解 \bar{x} 是基本可行解 $\iff \bar{x}$ 的非零分量所对应的 A 的列向量线性无关。
2. 设 S 是 LP 的可行域, $\bar{x} \in S$, 则 \bar{x} 是 S 的极点等价于 \bar{x} 是 LP 的基本可行解。

定理 2.5. 基本可行解的存在性：

1. 如果 LP 有可行解, 则一定存在基本可行解。
2. 如果 LP 有最优解, 则一定存在一个基本可行解是最优解。
3. 如果 LP 问题有最优解, 则要么最优解唯一, 要么有无穷多最优解。

3 单纯形法

3.1 单纯形法

笔记 3.1. 最优解一定在极点达到，而极点对应于基本可行解，求解线性规划问题归结为找最优基本可行解。可以从一个基本可行解出发，求一个使目标函数值有所改善的基本可行解：通过最优性条件判断是否达到最优解，并不断改进基本可行解，力图达到最优基本可行解。

笔记 3.2. 对于线性规划问题

$$(LP) \begin{cases} \min & f(x) = cx \\ s.t. & Ax = b \\ & x \geq 0 \end{cases}$$

其中 A 是 $m \times n$ 矩阵，秩为 m ， c 是 n 维列向量， $b \geq 0$ 是 m 维列向量。

- 初始基本可行解 $A = (P_1, \dots, P_m, P_{m+1}, \dots, P_n) = \begin{pmatrix} B & N \end{pmatrix}$

基本解 $x^{(0)} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$ ，若 $B^{-1}b \geq 0$ ，则 $x^{(0)}$ 是基本可行解。

$$\text{目标函数 } f_0 = cx^{(0)} = \begin{pmatrix} C_B & C_N \end{pmatrix} \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix} = C_B B^{-1}b$$

- 从初始基本可行解出发，求一个改进的基本可行解。
- 进基和终止条件

$$x = \begin{pmatrix} x_B \\ x_N \end{pmatrix}, Ax = b \Rightarrow \begin{pmatrix} B & N \end{pmatrix} \begin{pmatrix} x_B \\ x_N \end{pmatrix} = Bx_B + Nx_N = b \Rightarrow x_B = B^{-1}b - B^{-1}Nx_N$$

目标函数值：

$$\begin{aligned} f = cx &= \begin{pmatrix} c_B & c_N \end{pmatrix} \begin{pmatrix} x_B \\ x_N \end{pmatrix} = c_B x_B + c_N x_N \\ &= c_B (B^{-1}b - B^{-1}Nx_N) + c_N x_N \\ &= c_B B^{-1}b - (c_B B^{-1}N - c_N) x_N \\ &= f_0 - \sum_{j \in R} (c_B B^{-1}P_j - c_j) x_j \quad R: \text{非基变量下标集} \\ &= f_0 - \sum_{j \in R} (z_j - c_j) x_j \end{aligned}$$

1. $z_j - c_j$ 称为检验数或判别数，基变量的检验数为 0。
 2. 如果 $\forall j \in R$ ，有 $z_j - c_j \leq 0$ ，则 $x^{(0)}$ 为最优解；
 3. 否则 $z_k - c_k = \max_{j \in R} \{z_j - c_j\}$ ， P_k 为进基向量， x_k 为进基变量， $x_k = 0 \rightarrow x_k > 0$ 。
- 出基下标和进基变量的值
 $Ax = b$ 解的变化，原 $x_B = B^{-1}b - B^{-1}Nx_N$ ， x_k 由 0 变为正以后， $x_B = B^{-1}b - B^{-1}P_k x_k$ ， $x_B = \bar{b} - y_k x_k$ ，其中 $\bar{b} = B^{-1}b$ ， $y_k = B^{-1}P_k$ 。
 1. 若 $y_{ik} \leq 0$ ，则 $\forall x_k \Rightarrow x_{B_i} > 0$ ， x_k 可以取无限大，故 $f \rightarrow -\infty$ ，原问题无界。
 2. 要满足 $x_B = \bar{b} - y_k x_k \geq 0$ ，则取 $x_k = \min(\frac{\bar{b}}{y_{rk}}, y_{ik} > 0) = \frac{\bar{b}}{y_{rk}} > 0$ ， r 为出基下标。

笔记 3.3. 单纯形法计算步骤

1. 初始基为 B , 初始基本可行解为 $x^{(0)} = (B^{-1}b, 0)^t$, $\bar{b} = B^{-1}b$;
2. 判断是否所有的 $c_B B^{-1}P_j - c_j \leq 0$, 即检验数是否全部非正, 如果是, 那么 $x^{(0)}$ 就是最优解, 结束流程;
3. 否则取 $z_k - c_k = \max\{z_j - c_j\} > 0$, 判断是否 $y_k = B^{-1}P_k \leq 0$, 如果是, 则解无界;
4. 否则取 $\frac{\bar{b}_r}{y_{rk}} = \min\{\frac{\bar{b}_i}{y_{ik}} \mid y_{ik} > 0\}$, 以 p_k 代替 P_{b_r} 换基, 即 x_{b_r} 为离基变量, x_k 为进基变量, 跳转步骤 2。

笔记 3.4. 单纯形表

- 做初等行变换
- 若 $z_j - c_j > 0$, 对应的系数列向量 ≤ 0 , 则该 LP 存在无界解;
- 若某个非基变量的检验数为 0, 则该 LP 存在多个最优解.

| | f | x_B | x_N | 右端 |
|-------|-----|-------|---------------------|---------------|
| x_B | 0 | I_m | $B^{-1}N$ | $B^{-1}b$ |
| f | 1 | 0 | $c_B B^{-1}N - c_N$ | $c_B B^{-1}b$ |

可省略
检验数
(判别数)
目标函数取值
基变量
取值

图 1: 单纯形表

例 3.1. 用单纯形法求最优解

$$\begin{cases} \min & -x_2 + 2x_3 \\ s.t. & x_1 - 2x_2 + x_3 = 2 \\ & x_2 - 3x_3 \leq 1 \\ & x_2 - x_3 \leq 2 \\ & x_1, x_2, x_3 \geq 0 \end{cases}$$

答案. 引入松弛变量化为标准形

$$\begin{cases} \min & -x_2 + 2x_3 \\ s.t. & x_1 - 2x_2 + x_3 = 2 \\ & x_2 - 3x_3 + x_4 = 1 \\ & x_2 - x_3 + x_5 = 2 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{cases}$$

选择初始基 $B = (P_1, P_4, P_5) = I$ 。

| | x_1 | x_2 | x_3 | x_4 | x_5 | |
|-------|-------|-------|-------|-------|-------|---|
| x_1 | 1 | -2 | 1 | 0 | 0 | 2 |
| x_4 | 0 | 1 | -3 | 1 | 0 | 1 |
| x_5 | 0 | 1 | -1 | 0 | 1 | 2 |
| | 0 | 1 | -2 | 0 | 0 | 0 |

选取检验数最大的列，最大的列中 $\frac{1}{1} < \frac{2}{1}$

| | x_1 | x_2 | x_3 | x_4 | x_5 | |
|-------|-------|-------|-------|-------|-------|----|
| x_1 | 1 | 0 | -5 | 2 | 0 | 4 |
| x_2 | 0 | 1 | -3 | 1 | 0 | 1 |
| x_5 | 0 | 0 | 2 | -1 | 1 | 1 |
| | 0 | 0 | 1 | -1 | 0 | -1 |

选取检验数最大的列，最大的列中只有 $\frac{1}{2}$ 可选

| | x_1 | x_2 | x_3 | x_4 | x_5 | |
|-------|-------|-------|-------|-------|-------|------|
| x_1 | 1 | 0 | 0 | -1/2 | 5/2 | 13/2 |
| x_2 | 0 | 1 | 0 | -1/2 | 3/2 | 5/2 |
| x_3 | 0 | 0 | 1 | -1/2 | 1/2 | 1/2 |
| | 0 | 0 | 0 | -1/2 | -1/2 | -3/2 |

可得 $x^* = (\frac{13}{2}, \frac{5}{2}, \frac{1}{2})^t$, $f_{\min} = -\frac{3}{2}$ 。

例 3.2. 用单纯形法求最优解

$$\begin{cases} \max & 2x_1 + x_2 - x_3 \\ s.t. & x_1 + x_2 + 2x_3 \leq 6 \\ & x_1 + 4x_2 - x_3 \leq 4 \\ & x_1, x_2, x_3 \geq 0 \end{cases}$$

引入松弛变量化为标准形

$$\begin{cases} \max & 2x_1 + x_2 - x_3 \\ s.t. & x_1 + x_2 + 2x_3 + x_4 = 6 \\ & x_1 + 4x_2 - x_3 + x_5 = 4 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{cases}$$

选择初始基 $B = (P_4, P_5) = I$ 。

| | x_1 | x_2 | x_3 | x_4 | x_5 | |
|-------|-------|-------|-------|-------|-------|---|
| x_4 | 1 | 1 | 2 | 1 | 0 | 6 |
| x_5 | 1 | 4 | -1 | 0 | 1 | 4 |
| | -2 | -1 | 1 | 0 | 0 | 0 |

选取检验数最小的列，其中 $\frac{4}{1} < \frac{6}{1}$

| | x_1 | x_2 | x_3 | x_4 | x_5 | |
|-------|-------|-------|-------|-------|-------|---|
| x_4 | 0 | -3 | 3 | 1 | -1 | 2 |
| x_1 | 1 | 4 | -1 | 0 | 1 | 4 |
| | 0 | 7 | -1 | 0 | 2 | 8 |

选取检验数最小的列，只有 $\frac{3}{2}$

| | x_1 | x_2 | x_3 | x_4 | x_5 | |
|-------|-------|-------|-------|-------|-------|------|
| x_4 | 0 | -1 | 1 | 1/3 | -1/3 | 2/3 |
| x_1 | 1 | 3 | 0 | 1/3 | 2/3 | 14/3 |
| | 0 | 6 | 0 | 1/3 | 5/3 | 26/3 |

故 $x^* = (\frac{14}{3}, 0, \frac{2}{3})^t$, $f_{\max} = \frac{26}{3}$ 。

3.2 两阶段法

(感觉很 useless 啊)

对于如下线性规划问题

$$\begin{cases} \min & c^t x \\ s.t. & Ax = b \\ & x \geq 0 \end{cases}$$

两阶段法步骤:

1. 用单纯形法把人工变量变为非基变量, 求出原问题的一个基本可行解。

求解下列模型

$$\begin{cases} \min & e^t x_a \\ s.t. & Ax + x_a = b \\ & e = (1, \dots, 1)^t \\ & x, x_a \geq 0 \end{cases}$$

得到最优解为 $(\bar{x}^t, \bar{x}_a^t)^t$, 最优值为 $e^t \bar{x}_a$ 。

(a) 若 $\bar{x}_a \neq 0$, 则无可行解

(b) $\bar{x}_a = 0$ 而且所有的人工变量都是非基变量, 则 \bar{x} 是基本可行解

(c) $\bar{x}_a = 0$ 但 \bar{x}_a 的某个分量 \bar{x}_{a_j} 为基变量, 则设法将 \bar{x}_{a_j} 从基变量中去掉。

2. 从得到的基本可行解出发, 用单纯形法求最优解。

3.3 线性规划的最优性条件

即 KKT 条件。

对于如下凸优化问题

$$\begin{cases} \min & f(x) \\ s.t. & f_i(x) \leq 0, 1 \leq i \leq m \\ & h_i(x) = 0, 1 \leq i \leq p \end{cases}$$

有拉格朗日函数 $L(x, \lambda, \mu) = f(x) + \sum \lambda_i f_i(x) + \sum \mu_i h_i(x)$ 。

x 是最优解当且仅当

1. 满足约束条件: $f_i(x) \leq 0$, $h_i(x) = 0$
2. 非负约束: $\lambda \succeq 0$

3. 互补松弛: $\lambda_i f_i(x) = 0$

4. $\partial_x L(x, \lambda, \mu) = 0$

4 对偶理论

4.1 LP 对偶问题的形式

笔记 4.1. 图 2 中，左边是对偶问题 (D)，右边是原问题 (P)。

| Rules to construct the dual | | | |
|-----------------------------|--------------------------|--------------------------|----|
| | obj. coef. vector | right-hand-side | |
| | right-hand-side | obj. coef. vector | |
| | A | A^T | |
| | Max model | Min model | |
| 变量 | $x_j \geq 0$ | j th constraint \geq | 约束 |
| | $x_j \leq 0$ | j th constraint \leq | |
| | x_j free | j th constraint $=$ | |
| 约束 | i th constraint \leq | $y_i \geq 0$ | 变量 |
| | i th constraint \geq | $y_i \leq 0$ | |
| | i th constraint $=$ | y_i free | |

图 2: LP 问题的对偶问题转换规则

$$\left\{ \begin{array}{l} \min \quad 2x_1 + x_2 + 2x_3 \\ s.t. \quad x_1 + x_2 + 2x_3 \geq 1 \\ \quad \quad x_1 - x_2 + x_3 \leq 2 \\ \quad \quad -x_1 + x_2 + x_3 = 1 \\ \quad \quad x_1 \geq 0, x_2 \text{ free}, x_3 \leq 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \max \quad w_1 + 2w_2 + w_3 \\ s.t. \quad w_1 + w_2 - w_3 \leq 2 \\ \quad \quad w_1 - w_2 + w_3 = 1 \\ \quad \quad 2w_1 + w_2 + w_3 \geq 2 \\ \quad \quad w_1 \geq 0, w_2 \leq 0, w_3 \text{ free} \end{array} \right.$$

- 原问题的变量对应对偶问题的约束，并且符号改变。
- 原问题的约束对应对偶问题的变量，符号保持，等号对应 free。

4.2 对偶问题的基本性质

$$\text{原问题 (P)} \left\{ \begin{array}{l} \min \quad c^t x \\ s.t. \quad Ax \geq b \\ \quad \quad x \geq 0 \end{array} \right. \quad \text{对偶问题 (D)} \left\{ \begin{array}{l} \max \quad b^t w \\ s.t. \quad A^t w \leq c \\ \quad \quad w \geq 0 \end{array} \right.$$

定理 4.1 (弱对偶定理). 弱对偶定理: 若 $x^{(0)}, w^{(0)}$ 分别是原问题 (P) 和对偶问题 (D) 的可行解, 则 $c^T x^{(0)} \geq b^T w^{(0)}$, 即最小化目标的函数值大于等于最大化目标的函数值。

定理 4.2 (弱对偶推论). 若问题 (P) 或 (D) 有无界解, 则其对偶问题 (D) 或 (P) 无可行解。

定理 4.3 (最优性准则). 若 $x^{(0)}, w^{(0)}$ 分别为 (P), (D) 的可行解且 $c^T x^{(0)} = b^T w^{(0)}$, 则 $x^{(0)}, w^{(0)}$ 分别为 (P), (D) 问题的最优解。

定理 4.4 (强对偶定理). 若原问题 (P) 和对偶问题 (D) 均有可行解, 则原问题 (P) 和对偶问题 (D) 均有最优解, 且 (P) 和 (D) 的最优目标函数值相等。

- 推论: 若问题 (P) 或 (D) 无可行解, 则其对偶问题 (D) 或 (P) 或者无可行解, 或者目标函数值趋于无穷。
- 推论: 在用单纯形法求解 LP 问题 (P) 的松弛变量的检验数的相反数为对偶问题 (D) 的最优解。
- 推论: 在用单纯形法求解 LP 问题 (D) 的松弛变量的检验数的为原问题 (P) 的最优解。

例 4.1. 求下列问题对偶问题的最优解

$$\begin{cases} \max & 2x_1 + 3x_2 \\ s.t. & x_1 + 2x_2 \leq 8 \\ & 4x_1 \leq 16 \\ & 4x_2 \leq 12 \\ & x_1, x_2 \geq 0 \end{cases}$$

答案. 化为标准形

$$\begin{cases} \max & 2x_1 + 3x_2 \\ s.t. & x_1 + 2x_2 + x_3 = 8 \\ & 4x_1 + x_4 = 16 \\ & 4x_2 + x_5 = 12 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{cases}$$

| | x_1 | x_2 | x_3 | x_4 | x_5 | |
|-------|-------|-------|-------|-------|-------|----|
| x_3 | 1 | 2 | 1 | 0 | 0 | 8 |
| x_4 | 4 | 0 | 0 | 1 | 0 | 16 |
| x_5 | 0 | 4 | 0 | 0 | 1 | 12 |
| | -2 | -3 | 0 | 0 | 0 | 0 |
| | x_1 | x_2 | x_3 | x_4 | x_5 | |
| x_3 | 1 | 0 | 1 | 0 | -1/2 | 2 |
| x_4 | 4 | 0 | 0 | 1 | 0 | 16 |
| x_2 | 0 | 1 | 0 | 0 | 1/4 | 3 |
| | -2 | 0 | 0 | 0 | 3/4 | 9 |
| | x_1 | x_2 | x_3 | x_4 | x_5 | |
| x_1 | 1 | 0 | 1 | 0 | -1/2 | 2 |
| x_4 | 0 | 0 | -4 | 1 | 2 | 8 |
| x_2 | 0 | 1 | 0 | 0 | 1/4 | 3 |
| | 0 | 0 | 2 | 0 | -1/4 | 13 |

| | x_1 | x_2 | x_3 | x_4 | x_5 | |
|-------|-------|-------|-------|-------|-------|----|
| x_1 | 1 | 0 | 0 | 1/4 | 0 | 4 |
| x_5 | 0 | 0 | -2 | 1/2 | 1 | 4 |
| x_2 | 0 | 1 | 1/2 | -1/8 | 0 | 2 |
| | 0 | 0 | 3/2 | 1/8 | 0 | 14 |

此时获得最优解为 $x^* = (4, 2)^t$, $\max Z = 14$ 。

那么对偶问题的最优解为 $x^* = (\frac{3}{2}, \frac{1}{8}, 0)^t$, $\min Z = 14$ 。

定理 4.5 (互补松弛定理). 互补松弛定理: 设 $x^{(0)}, w^{(0)}$ 分别是 $(P), (D)$ 问题的可行解, 则 $x^{(0)}, w^{(0)}$ 分别为 $(P), (D)$ 的最优解的充要条件是 $\forall i, j (1 \leq i \leq m, 1 \leq j \leq n)$ 有

- 若 $x_j^{(0)} > 0$, 则 $w^{(0)} P_j = c_j$
- 若 $w^{(0)} P_j < c_j$, 则 $x_j^{(0)} = 0$
- 若 $w_i^{(0)} > 0$, 则 $A_i x^{(0)} = b_i$
- 若 $A_i x^{(0)} > b_i$, 则 $w_i^{(0)} = 0$

即 $\begin{cases} (c - w^{(0)} A) x^{(0)} = 0 \\ w^{(0)} (A x^{(0)} - b) = 0 \end{cases}$, 其中 P_j 是 A 的第 j 列, A_i 是 A 的第 i 行。

定理 4.6 (互补松弛定理非对称形式).

设 $x^{(0)}$ 和 $w^{(0)}$ 分别是 $\begin{cases} \min & c^t x \\ s.t. & Ax = b \\ & x \geq 0 \end{cases}$ 和 $\begin{cases} \max & b^t w \\ s.t. & A^t w \leq c \end{cases}$ 的可行解, 则 $x^{(0)}$ 和 $w^{(0)}$ 是最

优解的充要条件是 $\forall j$

- $x_j^{(0)} > 0 \implies w^{(0)} P_j = c_j$
- $w^{(0)} P_j < c_j \implies x_j^{(0)} = 0$

例 4.2. 考虑下面问题

$$(P) \begin{cases} \max & x_1 + 2x_2 + 3x_3 + 4x_4 \\ s.t. & x_1 + 2x_2 + 2x_3 + 3x_4 \leq 20 \\ & 2x_1 + x_2 + 3x_3 + 2x_4 \leq 20 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{cases} \quad (D) \begin{cases} \min & 20y_1 + 20y_2 \\ s.t. & y_1 + 2y_2 \geq 1 \\ & 2y_1 + y_2 \geq 2 \\ & 2y_1 + 3y_2 \geq 3 \\ & 3y_1 + 2y_2 \geq 4 \\ & y_1, y_2 \geq 0 \end{cases}$$

已知 (D) 的最优解为 $y^* = (\frac{6}{5}, \frac{1}{5})^t$, 用互补松弛定理求出 (P) 的最优解。

答案.

$$\begin{cases} x_1 + 2x_2 + 2x_3 + 3x_4 = 20 \\ 2x_1 + x_2 + 3x_3 + 2x_4 = 20 \\ x_1 = 0, x_2 = 0, x_3 > 0, x_4 > 0 \end{cases}$$

解得 $x^* = (0, 0, 4, 4)^t$, $\min Z = \max Z = 28$ 。

4.3 对偶问题的经济解释

4.4 对偶单纯形法

笔记 4.2. 对偶单纯形法步骤 (找的两次要小于 0, 和单纯形两次都要大于 0 相反):

1. 化标准型, 建立初始单纯形表
2. 判断, 若 $B^{-1}b \geq 0$, 则已得到最优解
3. 换基迭代
 - (a) 确定换出变量, $\bar{b}_r = \min_i \{\bar{b}_i\} < 0$, x_r 为换出变量
 - (b) 确定换入变量, $\min_j \left\{ \frac{z_j - c_j}{y_{rj}} \mid y_{rj} < 0 \right\} = \frac{z_k - c_k}{y_{rk}}$, x_k 为换入变量 (若所有 $y_{rj} \geq 0$, 则该 LP 问题无可行解)
 - (c) 换基迭代, y_{rk} 为主元
4. 回到第 2 步

笔记 4.3. 对偶单纯形法与原单纯形法的区别:

- 原单纯形法保持原问题的可行性, 对偶单纯形法保持所有检验数 $wP_j - c_j \leq 0$, 即保持对偶问题的可行性。
- 特点: 先选择出基变量, 再选择进基变量。

例 4.3. 用对偶单纯形法求解

$$\begin{cases} \min & x_1 + x_2 + x_3 \\ s.t. & 3x_1 + x_2 + x_3 \geq 1 \\ & -x_1 + 4x_2 + x_3 \geq 2 \\ & x_1, x_2, x_3 \geq 0 \end{cases}$$

答案. 引入松弛变量, 化为标准形

$$\begin{cases} \min & x_1 + x_2 + x_3 \\ s.t. & -3x_1 - x_2 - x_3 + x_4 = -1 \\ & x_1 - 4x_2 - x_3 + x_5 = -2 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{cases}$$

| | x_1 | x_2 | x_3 | x_4 | x_5 | |
|-------|-------|-------|-------|-------|-------|------|
| x_4 | -3 | -1 | -1 | 1 | 0 | -1 |
| x_5 | 1 | -4 | -1 | 0 | 1 | -2 |
| | -1 | -1 | -1 | 0 | 0 | 0 |
| | x_1 | x_2 | x_3 | x_4 | x_5 | |
| x_4 | -13/4 | 0 | -3/4 | 1 | -1/4 | -1/2 |
| x_2 | -1/4 | 1 | 1/4 | 0 | -1/4 | 1/2 |
| | -5/4 | 0 | -3/4 | 0 | -1/4 | 1/2 |

| | x_1 | x_2 | x_3 | x_4 | x_5 | |
|-------|-------|-------|-------|-------|-------|------|
| x_1 | 1 | 0 | 3/13 | -4/13 | 1/13 | 2/13 |
| x_2 | 0 | 1 | 4/13 | -1/13 | -3/13 | 7/13 |
| | 0 | 0 | -6/13 | -5/13 | -2/13 | 9/13 |

故 $x^* = (\frac{2}{13}, \frac{7}{13}, 0, 0, 0)^t$ 是最优解, $f_{\min} = \frac{9}{13}$, 对偶问题的最优解是 $(\frac{5}{13}, \frac{2}{13})^t$ 。

笔记 4.4. (对偶) 单纯形法步骤总结

- 单纯形法求解最小值问题, 第一步找最大的判别数, 第二步找最小的比值 (分母大于零)
- 单纯形法求解最大值问题, 第一步找最小的判别数, 第二步找最小的比值 (分母大于零)
- 对偶单纯形法求解最小值问题, 第一步找最小的 $B^{-1}b$, 第二步找最小的比值 (分母小于零)

4.5 原 - 对偶算法

原 - 对偶算法不考。

5 算法概述

5.1 算法分类

常见的解析法，图解法，智能算法，下降迭代算法。

应用广泛的智能算法有：遗传算法、蚁群算法、模拟退火算法、神经网络算法等。

笔记 5.1. 一类线搜索下降迭代算法的步骤：

1. 选定某一初始点 $x^{(0)}$ ，置 $k = 0$ ；
 2. 确定搜索方向 $d^{(k)}$ ；
 3. 从 $x^{(0)}$ 出发，沿方向 $d^{(k)}$ 求步长 λ_k ，以产生下一个迭代点 $x^{(k+1)}$ ；
 4. 检查 $x^{(k+1)}$ 是否为极小点或近似极小点，若是，则停止迭代；否则，令 $k := k + 1$ ，返回 2.
- 选取搜索方向是最关键的一步，各种算法的区别，主要在于确定搜索方向的方法不同。

笔记 5.2. 确定步长 λ_k 的主要方法

1. 令它等于某一常数。
2. 可接受点算法，即只要能使目标函数值下降，可任意选取步长。
3. 精确一维搜索：基于沿搜索方向使目标函数值下降最多，这样确定的步长为最佳步长。

$$\lambda_k = \arg \min f(x^{(k)} + \lambda d^{(k)})$$

5.2 算法收敛性

定义 5.1 (解集合). 把满足某些条件的点集定义为解集合。当迭代点属于该集合时，停止迭代。

常用的解集合：

- $\Omega = \{\bar{x} \mid \|\nabla f(\bar{x})\| = 0\}$
- $\Omega = \{\bar{x} \mid \bar{x} \text{ 为 KKT 点}\}$
- $\Omega = \{\bar{x} \mid \bar{x} \in S, f(\bar{x}) \leq b\}$ ，其中 b 是某个可接受的目标函数值。

笔记 5.3. 实用收敛准则

- $\|x^{(k+1)} - x^{(k)}\| < \varepsilon$ 或者 $\frac{\|x^{(k+1)} - x^{(k)}\|}{\|x^{(k)}\|} < \varepsilon$
- $f(x^{(k)}) - f(x^{(k+1)}) < \varepsilon$ 或者 $\frac{f(x^{(k)}) - f(x^{(k+1)})}{|f(x^{(k)})|} < \varepsilon$
- $\|\nabla f(x^{(k)})\| < \varepsilon$ (无约束最优化中)

笔记 5.4. Q-收敛速率 (Quotient)

设序列 $\{\gamma^{(k)}\}$ 收敛于 γ^* ，定义满足

$$\lim_{k \rightarrow +\infty} \frac{\|\gamma^{(k+1)} - \gamma^*\|}{\|\gamma^{(k)} - \gamma^*\|^p} = \beta < \infty$$

的非负数 p 的上确界为序列 $\{\gamma^{(k)}\}$ 的收敛级。

- 若序列的收敛级为 p ，则称序列是 p 级收敛的。
- 若 $p = 1$ 且 $0 < \beta < 1$ ，则称序列是以收敛比 β 线性收敛的。
- 若 $p > 1$ ，或者 $p = 1$ 且 $\beta = 0$ ，则称序列是超线性收敛的。
- 收敛级 p 越大，序列收敛得越快；当收敛级 p 相同时，收敛比 β 越小，序列收敛得越快。

笔记 5.5. R-收敛速率 (Root):

设点列 $\{x_k\}$ 收敛到 x^* . 若存在 $\kappa > 0, q \in (0, 1)$ 使 $\|x_k - x^*\| \leq \kappa q^k$, 则称点列 $\{x_k\}$ **R-线性收敛** 到 x^* ; 若存在 $\kappa > 0$ 和收敛到 0 的正数列 $\{q_k\}$ 使 $\|x_k - x^*\| \leq \kappa \prod_{i=1}^k q_i$, 则称点列 $\{x_k\}$ **R-超线性收敛** 到 x^* .

• **Q-(超)线性收敛 \implies R-(超)线性收敛**

笔记 5.6. 算法的二次终止性: 若某个算法对任意的**正定二次函数**, 从任意的初始点出发, 都能经有限步迭代达到其极小点, 则称该算法具有二次终止性.

用二次终止性作为判断算法优劣的原因:

1. 正定二次函数具有某些较好的性质, 因此一个好的算法应能够在有限步内达到其极小点.
2. 对于一般的目标函数, 若在其极小点处 **Hesse 矩阵** 正定

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

因此可以猜想, 对正定二次函数好的算法, 对于一般目标函数也应具有较好的性质.

5.3 算法复杂度

笔记 5.7. 单纯形算法的复杂度为指数时间复杂度。

笔记 5.8. 对于线性规划问题

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

如果求得的 λ_k , 使得

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

则称该一维搜索为**精确一维搜索**, λ_k 为最优步长。否则, 该一维搜索为**非精确一维搜索**。

笔记 5.9. 一维搜索

- **精确线搜索**
 - 试探法: 黄金分割法、Fibonacci 法、二分法
 - 函数逼近法: Newton 法、割线法、抛物线法、三次插值法
- **非精确线搜索:** Armijo 步长规则、Goldstein 步长规则、Wolfe 步长规则

笔记 5.10. 函数逼近法: 牛顿法

基本思想: 在极小点附近用二阶 Taylor 多项式近似.

$$\min f(x)$$

令 $\varphi(x) = f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) + \frac{1}{2}f''(x^{(k)})(x - x^{(k)})^2$, 又令 $\varphi'(x) = f'(x^{(k)}) + f''(x^{(k)})(x - x^{(k)}) = 0$, 得 $\varphi(x)$ 的驻点, 记 $x^{(k+1)}$, 则 $x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}$.

定理 5.1. 设 $f(x)$ 存在连续三阶导数, \bar{x} 满足 $f'(\bar{x}) = 0$, $f''(\bar{x}) \neq 0$, 初点 $x^{(1)}$ 充分接近 \bar{x} , 则牛顿法产生的序列 $\{x^{(k)}\}$ 至少以二级收敛速度收敛于 \bar{x} .

算法步骤:

1. 给定初始点 $x^{(0)}$, 允许误差 $\varepsilon > 0$, 置 $k = 0$.
2. 若 $|f'(x^{(k)})| < \varepsilon$, 则停止计算, 得点 $x^{(k)}$; 否则转 3

3. 计算点 $x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}$, 置 $k = k + 1$, 返回 2

缺点: 初始点选择十分重要. 如果初始点靠近极小点, 则可能很快收敛; 如果初始点远离极小点, 迭代产生的点列可能不收敛于极小点.

笔记 5.11. 非精确搜索:

- **Armijo 步长规则**

- 设 $\beta > 0, \gamma \in (0, 1), \sigma \in (0, 1)$. 取步长 $\lambda_k = \beta\gamma^{m_k}$, 其中 m_k 是满足下式的最小非负整数:

$$f\left(x^{(k)} + \beta\gamma^m d^{(k)}\right) \leq f\left(x^{(k)}\right) + \sigma\beta\gamma^m \nabla f\left(x^{(k)}\right)^T d^{(k)}$$

- 根据目标函数的 Taylor 展开式, 满足这种规则的步长一定存在.

- **Goldstein 步长规则**

- 设 $\sigma \in (0, \frac{1}{2})$. 取步长满足下式

$$f\left(x^{(k)} + \lambda d^{(k)}\right) \leq f\left(x^{(k)}\right) + \sigma\lambda \nabla f\left(x^{(k)}\right)^T d^{(k)}$$

$$f\left(x^{(k)} + \lambda d^{(k)}\right) > f\left(x^{(k)}\right) + (1 - \sigma)\lambda \nabla f\left(x^{(k)}\right)^T d^{(k)}$$

- 由于 $\lambda > 0$ 充分小时, 第二式必不成立, 故改规则在保证目标函数下降的前提下, 使下一迭代点尽可能远离当前迭代点.

- **Wolfe 步长规则**

- 设 $0 < \sigma_1 < \sigma_2 < 1$. 取步长 λ_k 满足下式

$$f\left(x^{(k)} + \lambda d^{(k)}\right) \leq f\left(x^{(k)}\right) + \sigma_1 \lambda \nabla f\left(x^{(k)}\right)^T d^{(k)}$$

$$\nabla f\left(x^{(k)} + \lambda d^{(k)}\right)^T d^{(k)} \geq \sigma_2 \nabla f\left(x^{(k)}\right)^T d^{(k)}$$

- 该规则使函数 $f(x^{(k)} + \lambda d^{(k)})$ 的陡度在 λ_k 点比在 $\lambda = 0$ 点有所减缓, 从而使下一迭代点尽可能远离当前迭代点.

6 非线性规划的最优性条件

6.1 无约束优化的最优性条件

无约束优化问题

$$\begin{cases} \min & f(x) \\ \text{s.t.} & x \in E^n \end{cases}$$

定义 6.1 (下降方向). 设 $\bar{x} \in E^n$ 是任给一点, $d \neq 0$, 若存在 $\delta > 0$, 使得对任意的 $\lambda \in (0, \delta)$, 有 $f(\bar{x} + \lambda d) < f(\bar{x})$, 则称 d 为 $f(x)$ 在点 \bar{x} 处的下降方向。

定理 6.1. 设函数 $f(x)$ 在点 \bar{x} 可微, 若存在 $d \neq 0$ 使 $\nabla f(\bar{x})^t d < 0$, 则存在 $\delta > 0$, 使对 $\forall \lambda \in (0, \delta)$, 有 $f(\bar{x} + \lambda d) < f(\bar{x})$, 即 d 是下降方向。

定理 6.2. 无约束优化问题的最优性条件

- 一阶必要条件: 设函数 $f(x)$ 在点 \bar{x} 处可微, 若 \bar{x} 是局部极小点, 则 $\nabla f(\bar{x}) = 0$ 。
- 二阶必要条件: 设函数 $f(x)$ 在 \bar{x} 处二阶可微, 若 \bar{x} 是局部极小点, 则 $\nabla f(\bar{x}) = 0$, 且 $\nabla^2 f(\bar{x})$ 是半正定的。
- 充分条件: 设函数 $f(x)$ 在点 \bar{x} 处二次可微, 若 $\nabla f(\bar{x}) = 0$, 且 $\nabla^2 f(\bar{x})$ 正定, 则 \bar{x} 是严格局部极小点。
- 设函数 $f(x)$ 在点 \bar{x} 的邻域内二次可微, 若 $\nabla f(\bar{x}) = 0$, 且 $\nabla^2 f(x)$ 在该邻域内半正定, 则 \bar{x} 是局部极小点。特别地, 对于邻域内的任意点 $x \neq \bar{x}$, 若 $\nabla^2 f(x)$ 正定, 则 \bar{x} 是一个严格的局部极小点。
- 设 $f(x)$ 是定义在 E^n 上的可微凸函数, $\bar{x} \in E^n$, 则 \bar{x} 为整体极小点的充要条件是 $\nabla f(\bar{x}) = 0$ 。

例 6.1. 求 $f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$ 的极小点。

答案.

$$\nabla f(x) = \begin{pmatrix} 4(x_1 - 2)^3 + 2(x_1 - 2x_2) \\ -4(x_1 - 2x_2) \end{pmatrix}$$

令 $\nabla f(x) = 0$, 解得驻点 $x^* = (2, 1)^t$ 。

$$\nabla^2 f(x) = \begin{pmatrix} 12(x_1 - 2)^2 + 2 & -4 \\ -4 & 8 \end{pmatrix}$$

$\nabla^2 f(x)$ 在 x^* 邻域内半正定, 故 $x^* = (2, 1)^t$ 是局部极小点。

笔记 6.1. 驻点

- 定义: 若 $f(x)$ 在点 x^* 可微, 并且 $\nabla f(x^*) = 0$. 则 x^* 称为 $f(x)$ 的一个驻点 (平稳点), 既不是极小点, 也不是极大点的驻点称为鞍点。

6.2 约束优化问题的最优性条件

定义 6.2 (下降方向). 对 $\min_{x \in E^n} f(x)$, 设 $\bar{x} \in E^n$ 是任给一点, $d \neq 0$, 若存在 $\delta > 0$, 使得对任意的 $\lambda \in (0, \delta)$ 有 $f(\bar{x} + \lambda d) < f(\bar{x})$, 则称 d 为 $f(x)$ 在点 \bar{x} 处的下降方向。

$$F_0 = \{d \mid \nabla f(\bar{x})^t d < 0\}$$

称为点 \bar{x} 处的下降方向集。

定义 6.3 (可行方向). 设集合 $S \subset E^n$, $\bar{x} \in clS$, d 为非零向量, 若存在数 $\delta > 0$, 使得对任意 $\lambda \in (0, \delta)$, 都有 $\bar{x} + \lambda d \in S$ 则称 d 为集合 S 在 \bar{x} 的可行方向。

$$D = \{d \mid d \neq 0, \bar{x} \in clS, \exists \delta > 0, \forall \lambda \in (0, \delta), \bar{x} + \lambda d \in S\}$$

是 \bar{x} 处的可行方向锥。

定理 6.3 (几何最优性条件). 考虑问题

$$\begin{cases} \min & f(x) \\ \text{s.t.} & x \in S \end{cases}$$

设 S 是 E^n 的非空集合, $\bar{x} \in S$, $f(x)$ 在 \bar{x} 处可微, 若 \bar{x} 是局部最优解, 则下降方向集和可行方向锥没有交集. 即 $F_0 \cap D = \emptyset$ 。

定理 6.4. 不等式约束优化问题的一阶最优性条件

$$\begin{cases} \min & f(x) \\ \text{s.t.} & g_i(x) \geq 0, \quad i = 1, \dots, m \end{cases}$$

可行域 $S = \{x \mid g_i(x) \geq 0, i = 1, \dots, m\}$

- 记 $I = \{i \mid g_i(\bar{x}) = 0, \bar{x} \in S\}$ 为起作用约束 (等式约束)。
- $G_0 = \{d \mid \nabla g_i(\bar{x})^t d > 0, i \in I\}$ 称为 S 在点 \bar{x} 处的局部约束方向锥 (内方向锥)。
- 几何最优性条件: 设 $\bar{x} \in S$, $f(x)$ 和 $g_i(x) (i \in I)$ 在 \bar{x} 处可微, $g_i(x) (i \notin I)$ 在 \bar{x} 处连续, 如果 \bar{x} 是局部最优解, 则下降方向集和内方向锥的交集为空, 即 $F_0 \cap G_0 = \emptyset$ 。

定理 6.5. Fritz John 条件: 设 $\bar{x} \in S$, $f(x), g_i(x) (i \in I)$ 在 \bar{x} 处可微, $g_i(x) (i \notin I)$ 在 \bar{x} 处连续, 若 \bar{x} 是局部最优解, 则存在不全为零的数 $w_0, w_i (i \in I)$, 使得

$$\begin{cases} w_0 \nabla f(\bar{x}) - \sum_{i \in I} w_i \nabla g_i(\bar{x}) = 0 \\ w_0, w_i \geq 0, \quad i \in I \end{cases}$$

\bar{x} 称为 Fritz John 点, 即满足 Fritz John 条件的点。

定理 6.6. KKT 条件: 设 $\bar{x} \in S$, $f(x), g_i(x) (i \in I)$ 在 \bar{x} 处可微, $g_i(x) (i \notin I)$ 在 \bar{x} 处连续, 若 \bar{x} 是局部最优解, 存在非负数 $w_i, i \in I$, 使得

$$\nabla f(\bar{x}) - \sum_{i \in I} w_i \nabla g_i(\bar{x}) = 0$$

设 $\bar{x} \in S, f, g_i$ 在 \bar{x} 可微, $\{\nabla g_i(\bar{x}) \mid i \in I\}$; 线性无关, 若 \bar{x} 是局部最优解, 则存在数 $w_i, i =$

$1, 2, \dots, m$, 使得

$$\begin{cases} \nabla f(\bar{x}) - \sum_{i=1}^m w_i \nabla g_i(\bar{x}) = 0 \\ w_i g_i(\bar{x}) = 0 \quad i = 1, 2, \dots, m \\ w_i \geq 0 \quad i = 1, 2, \dots, m \end{cases}$$

对于凸规划, 则若在 \bar{x} 点处 KKT 条件成立, 则 \bar{x} 为整体极小值点.

6.3 一般约束问题的一阶最优性条件

定理 6.7. 设 $\bar{x} \in S$, $f(x)$ 和 $g_i(x) (i \in I)$ 在 \bar{x} 处可微, $g_i(x) (i \notin I)$ 在 \bar{x} 处连续, $h_j (j = 1, \dots, l)$ 在 \bar{x} 处连续可微, 且 \bar{x} 是 $\mathcal{S} = \{x \mid h(x) = 0\}$ 的正则点. 如果 \bar{x} 是问题 (NP) 的局部最优解, 则在 \bar{x} 处, 有

$$F_0 \cap G_0 \cap H_0 = \emptyset$$

其中

$$\begin{aligned} F_0 &= \{d \mid \nabla f(\bar{x})^T d < 0\} \\ G_0 &= \{d \mid \nabla g_i(\bar{x})^T d > 0, i \in I\} \\ H_0 &= \{d \mid \nabla h_j(\bar{x})^T d = 0, j = 1, \dots, l\} \end{aligned}$$

定理 6.8. 若系统 $Ax < 0$, $Bx = 0$ 无解, 则系统 $A^t y + B^t z = 0, y \geq 0$, 且 $y \neq 0$ 或 $z \neq 0$ 有解.

笔记 6.2. 一阶充分条件 (凸优化问题):

$$\begin{cases} \min & f(x) \\ \text{s.t.} & g_i(x) \geq 0 \quad i = 1, \dots, m \\ & h_j(x) = 0, \quad j = 1, \dots, l \end{cases}$$

f 是凸函数, g_i 是凹函数, h_j 是线性函数, S 为可行域, $\bar{x} \in S$, $I = \{i \mid g_i(\bar{x}) = 0\}$. f 和 $g_i (i \in I)$ 在点 \bar{x} 可微, h_j 在点 \bar{x} 连续, $g_i (i \notin I)$ 在点 \bar{x} 连续, 且在 \bar{x} 处 KKT 条件成立, 则 \bar{x} 为整体极小点.

6.4 约束优化问题的二阶最优性条件

定理 6.9 (二阶充分条件). 对于优化问题

$$\begin{cases} \min & f(x) \\ \text{s.t.} & g_i(x) \geq 0 \\ & h_j(x) = 0 \end{cases}$$

设 f, g_i, h_j 是二次连续可微函数, \bar{x} 为可行解, 若存在 \bar{w}, \bar{v} , 使 $(\bar{x}, \bar{w}, \bar{v})$ 满足 KKT 条件且矩阵 $\nabla_x^2 L(\bar{x}, \bar{w}, \bar{v})$ 在子空间 G 上是正定的, 则 \bar{x} 是严格局部极小点。其中

$$G = \left\{ d \neq 0 \mid \begin{array}{l} \nabla g_i(\bar{x})^t d = 0, i \in I(\bar{x}) \quad \text{and} \quad \bar{w}_i > 0 \\ \nabla g_i(\bar{x})^t d \geq 0, i \in I(\bar{x}) \quad \text{and} \quad \bar{w}_i = 0 \\ \nabla h_j(\bar{x})^t d = 0 \end{array} \right\}$$

例 6.2. 求解如下问题的局部最优解

$$\begin{cases} \min & f(x) = x_1 x_2 \\ \text{s.t.} & c(x) = x_1^2 + x_2^2 - 1 = 0 \end{cases}$$

答案. 拉格朗日函数 $L(x, \mu) = x_1 x_2 + \mu(x_1^2 + x_2^2 - 1)$

有 KKT 条件

$$\begin{cases} x_1^2 + x_2^2 - 1 = 0 \\ x_2 + 2\mu x_1 = 0 \\ x_1 + 2\mu x_2 = 0 \end{cases}$$

解得

$$\mu = -\frac{1}{2}, x^* = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)^t \quad \text{or} \quad x^* = \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)^t$$

或

$$\mu = \frac{1}{2}, x^* = \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)^t \quad \text{or} \quad x^* = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)^t$$

根据二阶条件, 第一种情况

$$\nabla^2 L_x = \begin{pmatrix} 2\mu & 1 \\ 1 & 2\mu \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

$\mu = -\frac{1}{2}$, 对任意的 $d \neq 0$, 满足 $d_1 = -d_2$, 故

$$d^t \nabla_x^2 L d = -4d_1^2 < 0$$

故 $x^* = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)^t$ or $x^* = \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)^t$ 是局部极大值点。

第二种情况

$$\nabla^2 L_x = \begin{pmatrix} 2\mu & 1 \\ 1 & 2\mu \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$\mu = \frac{1}{2}$, 对任意的 $d \neq 0$, 满足 $d_1 = d_2$, 故

$$d^t \nabla_x^2 L d = 4d_1^2 > 0$$

故 $x^* = \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)^t$ or $x^* = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)^t$ 是局部极小值点。

7 Lagrange 对偶

7.1 Lagrange 对偶

$$\begin{cases} \min & f(x) \\ \text{s.t.} & g(x) \geq 0 \\ & h(x) = 0 \\ & x \in D \end{cases} \quad \begin{cases} \max & \theta(w, v) = \inf_{x \in D} L(x, w, v) \\ \text{s.t.} & g(x) \geq 0 \\ & h(x) = 0 \\ & x \in D \end{cases}$$

7.2 对偶定理

定理 7.1 (弱对偶定理). 设 x 和 (w, v) 分别是原问题和对偶问题的可行解, 则

$$f(x) \geq \theta(w, v).$$

记对偶间隙为 $\delta = f_{\min} - \theta_{\max} \geq 0$ 。

定理 7.2 (强对偶定理). 设 D 为非空开凸集, f 和 g_i 分别是 E^n 上的凸函数和凹函数, h_j 是 E^n 上的线性函数, 即 $h(x) = A(x) - b$, 存在 $\hat{x} \in D$, 使得

$$g(\hat{x}) > 0, h(\hat{x}) = 0, 0 \in \{h(x) \mid x \in D\}$$

则 $f_{\min} = \theta_{\max}$ 。

7.3 鞍点问题

定义 7.1 (鞍点). 设 $L(x, w, v)$ 为 Lagrange 函数, $\bar{x} \in E^n$, $\bar{w} \in E^m$, $\bar{w} \geq 0$, $\bar{v} \in E^l$, 如果对任意 x, w, v 都有

$$L(\bar{x}, w, v) \leq L(\bar{x}, \bar{w}, \bar{v}) \leq L(x, \bar{w}, \bar{v})$$

则称 $L(\bar{x}, \bar{w}, \bar{v})$ 为 $L(x, w, v)$ 的鞍点。

定理 7.3. Lagrange 函数的鞍点必是 Lagrange 函数关于 x 的极小点及关于 $(w, v)(w \geq 0)$ 的极大点。

定理 7.4 (鞍点定理). 设 $(\bar{x}, \bar{w}, \bar{v})$ 是原问题的 Lagrange 函数 $L(x, w, v)$ 的鞍点, 则 \bar{x} 和 (\bar{w}, \bar{v}) 分别是原问题和对偶问题的最优解。

定理 7.5 (鞍点定理). 假设 f 是凸函数, $g_i(x)$ 是凹函数, $h_j(x)$ 是线性函数, 且 A 是行满秩矩阵, 又设存在 \hat{x} , 使 $g(\hat{x}) > 0$, $h(\hat{x}) = 0$, 如果 \bar{x} 是原问题的最优解, 则存在 $(\bar{w}, \bar{v})(\bar{w} \geq 0)$, 使 $(\bar{x}, \bar{w}, \bar{v})$ 是 Lagrange 函数的鞍点。

定理 7.6 (鞍点和 KKT 条件的关系). 凸优化问题中, 鞍点一定满足 KKT 条件, KKT 条件成立的点一定是鞍点。

7.4 Lagrange 乘子的经济学解释

不会。

8 使用导数的最优化方法

最速下降法、牛顿法、共轭梯度法、拟牛顿法、信赖域法、最小二乘法。

8.1 最速下降法

笔记 8.1. 最速下降法, $f(x)$ 具有一阶连续偏导数

$$\begin{cases} \min & f(x) \\ \text{s.t.} & x \in E^n \end{cases}$$

带精确线搜索的最速下降法

1. 给定初始点 $x^{(1)} \in E^n$, 允许误差 $\varepsilon > 0$, 置 $k = 1$.
2. 取搜索方向: $d^{(k)} = -\nabla f(x^{(k)})$
3. 若 $\|d^{(k)}\| \leq \varepsilon$, 则停止计算; 否则, 从 $x^{(k)}$ 出发, 沿 $d^{(k)}$ 进行一维搜索, 求 λ_k , 使

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

4. 令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$, 置 $k := k + 1$, 返回 2

例 8.1. 求 $\min f(x) = (x_1 - 1)^2 + (x_2 - 1)^2$, 取 $x^{(1)} = (0, 0)^t, \varepsilon = 1e - 5$

答案. $\nabla f(x) = (2(x_1 - 1), 2(x_2 - 1))^t$

第一次迭代

- $d^{(1)} = -(-2, -2)^t = (2, 2)^t, \|d^{(1)}\| = 2\sqrt{2} > \varepsilon$
- $\min_{\lambda \geq 0} \varphi(\lambda) = \min_{\lambda \geq 0} f(x^{(1)} + \lambda d^{(1)})$, 得 $\lambda = \frac{1}{2}$.
- $x^{(2)} = x^{(1)} + \lambda d^{(1)} = (1, 1)^t$

第二次迭代

- $d^{(2)} = (0, 0)^t, \|d\| = 0 < \varepsilon$

故 $(1, 1)^t$ 为最优解。

笔记 8.2. 最速下降法 (二次情形): 对任意 $x^{(0)} \in E^n$, 最速下降法产生得序列收敛于 $f(x)$ 的唯一极小点 x^* , 而且, 对任意的 k , 有

$$E(x^{(k+1)}) \leq \left(\frac{A-a}{A+a} \right) E(x^{(k)})$$

其中 $E(x) = \frac{1}{2}(x - x^*)^t Q (x - x^*)$, A 为 $f(x)$ 的 Hesse 矩阵的最大特征值, $a > 0$ 为最小特征值。

最速下降法 (非二次情形): 设 $f(x)$ 存在连续二阶偏导数, \bar{x} 是局部极小点, Hesse 矩阵 $\nabla^2 f(\bar{x})$ 的最小特征值 $a > 0$, 最大特征值为 A , 算法产生的序列 $\{x^{(k)}\}$ 收敛于点 \bar{x} , 则目标函数值的序列 $\{f(x^{(k)})\}$ 以不大于 $\left(\frac{A-a}{A+a}\right)^2$ 的收敛比线性的收敛于 $f(\bar{x})$. 令条件数 $r = \frac{A}{a}$, 则 $\left(\frac{A-a}{A+a}\right)^2 = \left(\frac{r-1}{r+1}\right)^2 < 1$.

定理 8.1. 在相继两次迭代中, 梯度方向互相正交。

证明. 令 $\varphi(\lambda) = f(x^{(k)} + \lambda d^{(k)})$, $d^{(k)} = -\nabla f(x^{(k)})$, 为求出从 $x^{(k)}$ 出发沿方向 $d^{(k)}$ 的极小点, 令

$$\varphi'(\lambda_k) = \nabla f(x^{(k)} + \lambda_k d^{(k)})^t d^{(k)} = 0$$

得 $-\nabla f(x^{(k+1)})^t \nabla f(x^{(k)}) = 0$, 即方向 $d^{(k+1)} = -\nabla f(x^{(k+1)})$ 与 $d^{(k)} = -\nabla f(x^{(k)})$ 正交。 □

8.2 牛顿法

笔记 8.3. 牛顿法计算步骤:

1. 给定初始点 $x^{(0)} \in E^n$, 允许误差 $\varepsilon > 0$, 置 $k = 0$
2. 若 $\|\nabla f(x^{(k)})\| < \varepsilon$, 则停止计算;
3. 令

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

即迭代函数 (好记)

$$\text{next}(x) = x - (\nabla^2 f(x))^{-1} \nabla f(x)$$

置 $k := k + 1$, 返回 2.

例 8.2. 求 $\min f(x) = x_1^2 + 25x_2^2$ 。

答案.

$$\nabla f = \begin{pmatrix} 2x_1 \\ 50x_2 \end{pmatrix}, \nabla^2 f = \begin{pmatrix} 2 & 0 \\ 0 & 50 \end{pmatrix}$$

取 $x^{(0)} = (2, 2)^t$, $x^{(1)} = x^{(0)} - \nabla^2 f(x^{(0)})^{-1} \nabla f(x^{(0)}) = (0, 0)^t$.

$\|\nabla f(x^{(1)})\| = 0$, 则 $x^* = (0, 0)$ 。

笔记 8.4. 牛顿迭代法的缺点:

1. 可能会出现某步迭代时, 目标函数值上升
2. 当初始点远离极小点时, 牛顿法产生的点列可能不收敛, 或者收敛到鞍点, 或者 Hesse 矩阵不可逆, 无法计算
3. 需要计算 Hesse 矩阵, 计算量大

牛顿迭代法的优点:

1. 产生的点列 $\{x^{(k)}\}$ 若收敛, 则收敛速度快, 具有至少二阶收敛速率。
2. 牛顿法具有二次终止性 (正定二次函数可以一次迭代到达最优点)。

笔记 8.5. 阻尼牛顿法计算步骤:

1. 给定初始点 $x^{(1)} \in E^n$, 允许误差 $\varepsilon > 0$, 置 $k = 1$
2. 计算 $\nabla f(x^{(k)}), \nabla^2 f(x^{(k)})^{-1}$
3. 若 $\|\nabla f(x^{(k)})\| < \varepsilon$, 则停止迭代; 否则, 令 $d^{(k)} = \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$
4. 从 $x^{(k)}$ 出发, 沿方向 $d^{(k)}$ 作一维搜索:

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

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

5. 置 $k := k + 1$, 转步骤 2.

笔记 8.6. 修正牛顿法计算步骤:

1. 给定初始点 $x^{(1)} \in E^n$, 允许误差 $\varepsilon > 0$, 置 $k = 1$
2. 计算 $\nabla f(x^{(k)}), \nabla^2 f(x^{(k)})^{-1}$

3. 若 $\|\nabla f(x^{(k)})\| < \varepsilon$, 则停止迭代; 否则, 置 $B_k = G_k + \varepsilon_k I$, 其中 ε_k 是一个非负数, 选取 ε_k , 使得 B_k 是对称正定矩阵, 计算修正牛顿方向 $d^{(k)} = -B_k^{-1} \nabla f(x^{(k)})$
4. 从 $x^{(k)}$ 出发, 沿方向 $d^{(k)}$ 作一维搜索:

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

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

5. 置 $k := k + 1$, 转步骤 2.

笔记 8.7. 牛顿 - 最速下降法计算步骤:

1. 给定初始点 $x^{(1)} \in E^n$, 允许误差 $\varepsilon > 0$, 置 $k = 1$
2. 计算 $g_k = \nabla f(x^{(k)})$. 若 $\|g_k\| \leq \varepsilon$, 算法终止, 输出 $x^{(k)}$ 作为近似最优解; 否则, 转步骤 3
3. 计算 $G_k = \nabla^2 f(x^{(k)})$. 解线性方程组

$$G_k d^{(k)} + \nabla f(x_k) = 0$$

若有解 $d^{(k)}$ 且满足 $g_k^t d^{(k)} < 0$, 转步骤 4, 否则令 $d^{(k)} = -g_k$, 转步骤 4

4. 从 $x^{(k)}$ 出发, 沿方向 $d^{(k)}$ 作一维搜索:

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

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

5. 置 $k := k + 1$, 转步骤 2.

8.3 共轭梯度法

定义 8.1 (2-共轭方向). 设 A 是 n 阶对称正定矩阵, 若 E^n 中的两个方向 $d^{(1)}$ 和 $d^{(2)}$ 满足

$$\left(d^{(1)}\right)^t A d^{(2)} = 0$$

则称这两个方向关于 A 共轭, 或称它们关于 A 正交。

定义 8.2 (k -共轭方向). 若 $d^{(1)}, \dots, d^{(k)}$ 是 E^n 中 k 个方向, 它们两两关于 A 共轭, 即

$$\left(d^{(i)}\right)^t A d^{(j)} = 0, \quad \forall i \neq j$$

则称这组方向是 A 共轭的, 或称它们为 A 的 k 个共轭方向。

定理 8.2. 设 A 是 n 阶对称正定矩阵, $d^{(1)}, \dots, d^{(k)}$ 是 k 个 A 共轭的非零向量, 则这 k 个向量线性无关。

证明. 对于正定矩阵 A , 假设

$$\alpha_1 d^{(1)} + \alpha_2 d^{(2)} + \dots + \alpha_k d^{(k)} = 0$$

左乘 $d^{(i)t} A$ 可得

$$\alpha_i d^{(i)t} A d^{(i)} + \sum_{j \neq i} \alpha_j d^{(i)t} A d^{(j)} = 0$$

由于

$$\alpha_i d^{(i)t} A d^{(j)} = 0$$

故 $\alpha_i = 0$ 。根据假设可得这 k 个向量线性无关。 □

定理 8.3. 二次函数 $f(x) = \frac{1}{2}x^t Ax + b^t x + c$, 其中 $A_{n \times n}$ 是对称正定矩阵, $d^{(1)}, \dots, d^{(n-1)}$ 是 A 共轭的非零向量, 从任意一点 $x^{(0)} \in E^n$ 出发, 依次沿这组向量进行一维搜索

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

$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}, k = 0, \dots, n-1$, 则

$$\nabla f(x^{(k+1)})^t d^{(j)} = 0, j = 0, \dots, k$$

即搜索方向和之前所有的搜索方向都正交, 并且最多经过 n 步收敛。

笔记 8.8. 记 $g_i = \nabla f(x^{(i)})$

FR 共轭梯度法 (二次凸函数)

1. 给定初始点 $x^{(1)}$, 置 $k = 1$
2. 计算 $g_k = \nabla f(x^{(k)})$, 若 $\|g_k\| = 0$, 则停止计算, 否则进行下一步
3. 令 $d^{(k)} = -g_k + \beta_{k-1} d^{(k-1)}$, 其中, 当 $k = 1$ 时, $\beta_0 = 0$, 当 $k > 1$ 时, $\beta_{k-1} = \frac{d^{(k-1)t} A g_k}{d^{(k-1)t} A d^{(k-1)}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$
4. 令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$, 其中 $\lambda_k = \frac{g_k^t g_k}{d^{(k)t} A d^{(k)}}$
5. 若 $k = n$, 则停止计算, 否则置 $k = k + 1$, 返回步骤 2.

定理 8.4. 对正定二次函数, FR 法在 $m \leq n$ 次一维搜索后终止, 且对 $\forall i (1 \leq i \leq m)$, 下列关系成立

1. $d^{(i)t} A d^{(j)} = 0, j = 1, \dots, i-1$
2. $g_i^t g_j = 0, j = 1, \dots, i-1$
3. $g_i^t d^{(i)} = -g_i^t g_i$

笔记 8.9. 一般函数的共轭梯度法

1. 步长 λ_k 不能再公式 $\lambda_k = -\frac{g_k^t g_k}{d^{(k)t} A d^{(k)}}$ 计算, 必须用其他一维搜索方法来确定
2. 凡用到矩阵 A 之处, 需用现行点的 Hessian 矩阵 $\nabla^2 f(x^{(k)})$ 代替
3. 有限步迭代达不到极小点

| 无约束优化算法的对比 | | | | |
|------------|--------|--|-----------------------|--------------|
| 算法 | | 搜索方向 | 步长 | 特点 |
| 最速下降法 | | $-\nabla f(x^{(k)})$ | 最优步长、Armijo步长、Wolfe步长 | 线性收敛, 锯齿现象 |
| 牛顿法 | | $-\left[\nabla^2 f(x^{(k)})\right]^{-1} \nabla f(x^{(k)})$ | 1 | 二阶收敛, 二次终止性 |
| 阻尼牛顿法 | | $-\left[\nabla^2 f(x^{(k)})\right]^{-1} \nabla f(x^{(k)})$ | 最优步长、Armijo步长、Wolfe步长 | 二阶收敛, 二次终止性 |
| 修正牛顿法 | | $-\left[\nabla^2 f(x^{(k)}) + \mu_k I\right]^{-1} \nabla f(x^{(k)})$ | 最优步长、Armijo步长、Wolfe步长 | 二阶收敛, 二次终止性 |
| 拟牛顿法 | | $-H_k \nabla f(x^{(k)})$ | 最优步长、Wolfe步长 | 超线性收敛, 二次终止性 |
| 共轭梯度法 | 正定二次函数 | $-\nabla f(x^{(k)}) + \beta_{k-1} d^{(k-1)}$ | 最优步长、Wolfe步长 | 超线性收敛, 二次终止性 |
| | 一般函数 | $-\nabla f(x^{(k)}) + \beta_{k-1} d^{(k-1)}$ | 最优步长、Wolfe步长 | 超线性收敛 |

图 3: 无约束优化算法的对比

9 罚函数法

外点罚函数法、内点罚函数法、乘子罚函数法。

9.1 外点罚函数法

笔记 9.1. 对于有约束优化问题

$$\begin{cases} \min & f(x) \\ \text{s.t.} & g_i(x) \geq 0, i = 1, \dots, m \\ & h_j(x) = 0, j = 1, \dots, l \end{cases}$$

- 通过引入罚项转换为无约束优化问题

$$\begin{aligned} \min F(x, \sigma) &= f(x) + \sigma p(x) \\ &= f(x) + \sigma \left(\sum_{i=1}^m \varphi(g_i(x)) + \sum_{j=1}^l \psi(h_j(x)) \right) \end{aligned}$$

- 其中 σ 是很大的正数, $\varphi(y), \psi(y)$ 是连续函数。
- 一般取 $\varphi(g_i(x)) = (\max\{0, -g_i(x)\})^\alpha$, $\psi(h_j(x)) = |h_j(x)|^\beta$, 通常 $\alpha = \beta = 2$ 。

笔记 9.2. 外点罚函数法步骤:

1. 给定初始点 $x^{(0)}$, 初始罚因子 $\sigma_1 > 0 (\sigma_1 = 1)$, 放大系数 $c > 1$, 允许误差 $\varepsilon > 0$, 置 $k = 1$.
2. 以 $x^{(k-1)}$ 为初始点, 求解无约束问题 $\min f(x) + \sigma_k p(x)$ 设其极小点为 $x^{(k)}$.
3. 若 $\sigma_k p(x^{(k)}) < \varepsilon$, 则停止计算, 得到点 $x^{(k)}$; 否则令 $\sigma_{k+1} = c\sigma_k$, 置 $k := k + 1$, 返回步骤 2.

定理 9.1. 对于由外点法所产生的序列 $\{x^{(k)}\}$, 总有

1. $F(x^{(k+1)}, \sigma_{k+1}) \geq F(x^{(k)}, \sigma_k)$
2. $p(x^{(k+1)}) \leq p(x^{(k)})$
3. $f(x^{(k+1)}) \geq f(x^{(k)})$

定理 9.2. 设 x^* 是问题 (A) 的一个最优解, 则对 $\forall k$, 有 $f(x^*) = F(x^*, \sigma^*) \geq F(x^{(k)}, \sigma_k) \geq f(x^{(k)})$ 。

笔记 9.3. 外点罚函数法的重要特点: 函数 $F(x, \sigma)$ 是在整个空间 E^n 内进行优化, 初始点可任意选择, 且外点法也可用于非凸规划的最优化。

缺点

1. 惩罚项 $\sigma p(X)$ 的二阶偏导数一般不存在;
2. 外点法的中间结果不是可行解, 不能作为近似解;
3. 当点 $x^{(k)}$ 接近最优解时, 罚因子 σ_k 很大。可能使罚函数性质变坏, 使搜索产生极大困难。

例 9.1. 使用外点罚函数法求解优化问题

$$\begin{cases} \min & x_1 + x_2 \\ \text{s.t.} & x_1 - x_2^2 = 0 \end{cases}$$

答案. 定义罚函数

$$F(x, \sigma) = x_1 + x_2 + \sigma(x_1 - x_2^2)^2$$

求偏导

$$\partial_{x_1} F = 1 + 2\sigma(x_1 - x_2^2)$$

$$\partial_{x_2} F = 1 - 4\sigma x_2(x_1 - x_2^2)$$

令 $\partial_x F = 0$, 解得

$$\bar{x}_\sigma = \left(\frac{1}{4} - \frac{1}{\sigma}, -\frac{1}{2}\right)^t \rightarrow \left(\frac{1}{4}, -\frac{1}{2}\right) (\sigma \rightarrow \infty)$$

例 9.2. 使用外点罚函数法求解优化问题

$$\begin{cases} \min & f(x) = x_1 + x_2 \\ s.t. & g_1(x) = -x_1^2 + x_2 \geq 0 \\ & g_2(x) = x_1 \geq 0 \end{cases}$$

答案. 定义外罚函数

$$F(x, \sigma) = x_1 + x_2 + \sigma(\max(0, x_1^2 - x_2))^2 + \sigma(\max(0, -x_1))^2$$

求偏导

$$\partial_{x_1} F = 1 + 4\sigma x_1 \max(0, x_1^2 - x_2) - 2\sigma \max(0, -x_1)$$

$$\partial_{x_2} F = 1 - 2\sigma x_2 \max(0, x_1^2 - x_2)$$

令 $\partial_x F = 0$, 解得

$$\bar{x}_\sigma = \left(-\frac{1}{2+2\sigma}, \frac{1}{(2+2\sigma)^2} - \frac{1}{2\sigma}\right)^t \rightarrow (0, 0)^t (\sigma \rightarrow \infty)$$

例 9.3. 考虑下列问题:

$$\begin{cases} \min & x_1 x_2 \\ s.t. & g(x) = -2x_1 + x_2 + 3 \geq 0. \end{cases}$$

1. 用二阶最优性条件证明点 $\bar{x} = (\frac{3}{4}, -\frac{3}{2})$ 是局部最优解, 并说明它是否为全局最优解。
2. 定义障碍函数

$$G(x, r) = x_1 x_2 - r \ln g(x),$$

试用内点法求解, 并说明内点法产生的序列趋向点 \bar{x} .

答案.

1. $L(x, \lambda) = x_1 x_2 + \lambda(2x_1 - x_2 - 3)$, 由 KKT 条件

$$\begin{cases} \lambda(2x_1 - x_2 - 3) & = 0 \\ \lambda & \geq 0 \\ \partial_{x_1} L = x_2 + 2\lambda & = 0 \\ \partial_{x_2} L = x_1 - \lambda & = 0 \end{cases}$$

解得

$$x = \left(\frac{3}{4}, -\frac{3}{2}\right)^t, \quad \lambda = \frac{3}{4}.$$

由

$$\lambda > 0, \nabla g(x)^t d = 0$$

可得

$$d = (d_1, d_2)^t, \quad d_1 - 2d_2 = 0, \quad d \neq 0.$$

此时

$$\begin{aligned} d^t \nabla_x^2 L(x, \lambda) d &= \begin{pmatrix} 2d_2 & d_2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2d_2 \\ d_2 \end{pmatrix} \\ &= 4d_2^2 \\ &> 0 \end{aligned}$$

故矩阵 $\nabla^2 L(x, \lambda)$ 在子空间 G 上正定，故是极小值点。

只有唯一的 KKT 点，故也是全局最优解。

2. 令

$$\begin{cases} \partial_{x_1} G = x_2 + \frac{2r}{-2x_1 + x_2 + 3} = 0 \\ \partial_{x_2} G = x_1 - \frac{r}{-2x_1 + x_2 + 3} = 0 \end{cases}.$$

可得

$$x = \frac{1}{8}(3 \mp \sqrt{9 - 16r}, -6 \pm 2\sqrt{9 - 16r})^t.$$

需在 $r > 0$ 时满足约束条件，故

$$x = \frac{1}{8}(3 + \sqrt{9 - 16r}, -6 - 2\sqrt{9 - 16r})^t.$$

令 $r \rightarrow 0$, 可得 $x = \left(\frac{3}{4}, -\frac{3}{2}\right)^t$ 是全局最优解。

9.2 内点罚函数法

基本思想：迭代总是从内点出发，并保持可行域内部进行搜索。

笔记 9.4. 障碍函数：

$$G(x, r) = f(x) + rB(x)$$

其中 r 是很小的正数， $B(x)$ 定义在可行域内部，它满足两个条件：

1. $B(x)$ 是连续函数
2. 当点 x 趋向可行域边界时， $B(x) \rightarrow +\infty$

两种最重要的形式：

$$B(x) = \sum_{i=1}^m \frac{1}{g_i(x)} \quad B(x) = -\sum_{i=1}^m \ln g_i(x)$$

笔记 9.5. 算法步骤：

1. 给定初始点 $x^{(0)} \in \text{int } S$ ，允许误差 $\varepsilon > 0$ ，初始参数 r_1 ，缩小系数 $\beta \in (0, 1)$ ，置 $k = 1$ 。

2. 以 $x^{(k-1)}$ 为初始点, 求解下列问题

$$\begin{cases} \min & f(x) + r_k B(x) \\ \text{s.t.} & x \in \text{int } S \end{cases}$$

设其极小点为 $x^{(k)}$.

3. 若 $r_k B(x^{(k)}) < \varepsilon$, 则停止计算, 得到点 $x^{(k)}$; 否则, 令 $r_{k+1} = \beta r_k$, 置 $k := k + 1$, 返回 2.

定理 9.3. 对于由内点法所产生的序列 $\{x^{(k)}\}$, 总有

1. $G(x^{(k+1)}, r_{k+1}) \leq G(x^{(k)}, r_k)$
2. $B(x^{(k+1)}) \geq B(x^{(k)})$
3. $f(x^{(k+1)}) \leq f(x^{(k)})$

定理 9.4. 设 $\{x^{(l)}\}$ 是由内点法产生的一个序列, 则 $\{x^{(k)}\}$ 的任何收敛子序列的极限都是原问题的最优解。

笔记 9.6. 求初始内点的迭代步骤:

1. 任取 $x^{(0)} \in E^n, r_0 > 0$ (如取 $r_0 = 1$), 置 $k := 0$.
2. 令 $S_k = \{i \mid g_i(x^{(k)}) \leq 0, 1 \leq i \leq m\}, T_k = \{i \mid g_i(x^{(k)}) > 0, 1 \leq i \leq m\}$
3. 若 $S_k = \emptyset$, 停止计算; 否则, 转 4.
4. 构造函数

$$\tilde{P}(x, r_k) = - \sum_{i \in S_k} g_i(x) + r_k \sum_{i \in T_k} \frac{1}{g_i(x)}, (r_k > 0)$$

记 $\tilde{R}_k = \{x \mid g_i(x) > 0, i \in T_k\}$

5. 以 $x^{(k)}$ 为初始点, 在 \tilde{R}_k 域内, 求障碍函数 $\tilde{P}(x, r_k)$ 的极小点:

$$\begin{cases} \min & \tilde{P}(x, r_k) \\ \text{s.t.} & x \in \tilde{R}_k \end{cases}$$

得 $x^{(k+1)}$, 转 6

6. 令 $0 < r_{k+1} < r_k$ (如取 $r_{k+1} = \frac{1}{10} r_k$), 置 $k := k + 1$, 转 2.

例 9.4. 考虑约束优化问题:

$$\begin{cases} \min & x_1 x_2 \\ \text{s.t.} & 2x_1 - x_2 - 3 = 0 \end{cases}$$

1. 给定 $\bar{x} = (\frac{3}{4}, -\frac{3}{2})^t$, 利用约束优化问题局部解的一阶必要条件合二阶充分条件判断 \bar{x} 是否是局部最优解?
2. 定义外罚函数为

$$G(x, c) = x_1 x_2 + \frac{c}{2} (2x_1 - x_2 - 3)^2$$

试用外罚函数法求解, 并说明产生的序列趋向点。

答案.

1. Lagrange 乘子函数为 $L(x, \mu) = x_1x_2 + \mu(2x_1 - x_2 - 3)$.

由 KKT 条件可得

$$\begin{cases} 2x_1 - x_2 - 3 &= 0 \\ \partial_{x_1} L = x_2 + 2\mu &= 0 \\ \partial_{x_2} L = x_1 - \mu &= 0 \end{cases}$$

解得

$$x^* = \left(\frac{3}{4}, -\frac{3}{2} \right).$$

由二阶最优性条件可得

$$\nabla h(x)^t d = 0$$

故

$$d = (d_1, d_2) \neq 0, \quad 2d_1 - d_2 = 0$$

则 Lagrange 乘子函数的 Hessian 矩阵在空间 G 上

$$\begin{aligned} d^t \nabla^2 L_x(x, \mu) d &= \begin{pmatrix} d_1 & 2d_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} d_1 \\ 2d_1 \end{pmatrix} \\ &= 4d_1^2 \\ &> 0 \end{aligned}$$

故 $\nabla^2 L_x(x, \mu)$ 正定, $\bar{x} = \left(\frac{3}{4}, -\frac{3}{2} \right)$ 是局部最优解。

2. 对 x 求偏导可得

$$\partial_{x_1} G = x_2 + c(2x_1 - x_2 - 3) \cdot 2 = 0$$

$$\partial_{x_2} G = x_1 + c(2x_1 - x_2 - 3) \cdot (-1) = 0$$

解得

$$x^* = \left(\frac{3c}{4c-1}, \frac{-6c}{4c-1} \right).$$

令 $c \rightarrow +\infty$, 得

$$x^* = \left(\frac{3}{4}, -\frac{3}{2} \right).$$

笔记 9.7. 内点罚函数法优点: 迭代总在可行域内进行, 每一个中间结果都是可行解, 可以作为近似解。

内点罚函数法缺点: 选取初始可行点较困难, 且只适用于含不等式约束的非线性规划问题, 否则没有严格内点。

9.3 乘子罚函数法

9.3.1 等式约束优化的乘子罚函数法

笔记 9.8. 对于等式约束优化问题

$$\begin{cases} \min & f(x) \\ s.t. & c_i(x) = 0, i \in \varepsilon \end{cases}$$

$P(x, \lambda, \pi)$ 为乘子罚函数，也称增广 **Lagrange** 函数。

$$\min_x P(x, \lambda, \pi) := f(x) - \sum_{i \in \varepsilon} \lambda_i c_i(x) + \pi \sum_{i \in \varepsilon} c_i^2(x)$$

对于凸规划，任意 $\pi > 0$ 下，原问题的最优解一定是乘子罚函数的最优解。

并且要使 $c_i(x_k) \rightarrow 0$ ，可以使乘子 $\lambda_i^{(k)}$ 足够靠近最优 **Lagrange** 乘子 λ_i^* ，而不一定需要罚因子 $\pi_k \rightarrow +\infty$ 。

笔记 9.9. 乘子罚函数算法：

1. 给定 $\pi_0 > 0, \lambda^{(0)} = 0$ ，初始点 $x_{-1} \in \mathbb{R}^n$ ，增长因子 $\gamma > 1$ 和允许误差 $\varepsilon > 0$ 。令 $k = 0$ 。
2. 以 x_{k-1} 为初始点，用无约束优化方法计算函数 $P(x, \lambda^{(k)}, \pi_k)$ 的最小值点 x_k 。
3. 若 $\max\{|c_i(x_k)| \mid i \in \varepsilon\} \leq \varepsilon$ ，算法终止。否则，转下一步。
4. 若 $\|c(x_k)\|_\infty \geq \|c(x_{k-1})\|_\infty$ ，令 $\pi_{k+1} = \gamma\pi_k, \lambda^{(k+1)} = \lambda_k$ ，置 $\lambda^{(k+1)} = \lambda^{(k)}$ ，转步 2。否则，转步 5。
5. 若 $\pi_k > \pi_{k-1}$ 或 $\|c(x_k)\|_\infty \leq \frac{1}{4}\|c(x_{k-1})\|_\infty$ ，令 $\pi_{k+1} = \pi_k$ ，根据 $\lambda_i^{(k+1)} = \lambda_i^{(k)} - 2\pi_k c_i(x_k)$ 调整 $\lambda^{(k+1)}$ ，置 $k = k + 1$ ，转步 2。否则，令 $\pi_{k+1} = \gamma\pi_k, \lambda^{(k+1)} = \lambda^{(k)}$ ，置 $k = k + 1$ ，转步 2。

10 最小二乘问题