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

1 Introduction

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

对于 n 阶实对称矩阵 A，下列命题等价：

- A^TAX 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 $A = P^TP$

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

对于 n 阶实对称矩阵 A，下列命题等价：

- A^TAX 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

Remark 1.3. 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。

- $x_1, \dots, x_k \in S, \lambda_1 + \cdots + \lambda_k = 1$，称 $\sum_{i=1}^k \lambda_i x_i$ 为 $x_1, \dots, x_k$ 的凸组合
- 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。
- S 是 E^n 中的闭凸集，$d \in E^n, d \neq 0$，如果对 $\forall x \in S$，有 $\{x + \lambda d | \lambda > 0\} \subset S$，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。

Remark 1.4. 设 $S = \{x | Ax = b, x \geq 0\}$ 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

- (多面集表示定理)$x \in S$ 的充要条件是

$$x = \sum_{j=1}^{k} \lambda_j x^{(j)} + \sum_{j=1}^{l} \mu_j d^{(j)}$$
$$\lambda_j \geq 0, \forall j = 1, \dots, k$$
$$\sum_{j=1}^{k} \lambda_j = 1$$
$$\mu_j \geq 0, \forall j = 1, \dots, l$$

Remark 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 $\overline{x} \in S$，使得

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

  $\overline{x}$ 是这一最小距离点 $\Leftrightarrow (y - \overline{x})^T(\overline{x} - x) \geq 0, \forall x \in S$。

- 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 $\varepsilon > 0$，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 $\forall x \in clS$（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

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

Remark 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 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$，下列两个系统只有一个有解

$$\boldsymbol{A}x < 0$$
$$\boldsymbol{y} \geq 0, \boldsymbol{y} \neq 0, \boldsymbol{A}^T \boldsymbol{y} = 0$$

Remark 1.7. 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。

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 上的局部极小点事整体极小点，且极小点的集合是凸集。

Remark 1.8. $f(x)$ 是凸集 S 上的凸函数，对每一个实数 c，则集合 $S_c = \{x|x \in S, f(x) \leq c\}$ 是凸集。

Remark 1.9. 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。

Remark 1.10. 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。

Remark 1.11. 凸规划：求凸函数在凸集上的极小点。

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

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

Remark 2.1. LP 基本定理:

- 可行域的极点对应 LP 问题的基本可行解
- LP 的最优解一定可以在基本可行解中找到
- 设 S 是 LP 的可行域，$\bar{x} \in S$，则 $\bar{x}$ 是 S 的极点 $\Longleftrightarrow$ $\bar{x}$ 是 LP 的基本可行解
- 如果 LP 有可行解，则一定存在基本可行解
- 如果 LP 有最优解，则存在一个基本可行解是最优解
- 若 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

Remark 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，求一个使目标丽数值有所改善的基本可行解:通过不断改进基本可行解，力图达到最优基本可行解。

Remark 3.2.

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

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。

  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

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

1. 如果 $\forall j \in R$，有 $z_j - c_j \leq 0$，则 $x^{(0)}$ 为最优解；
2. 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，$x_k = 0 \to x_k > 0$。

- 出基下标和进基变量的值

  $Ax = b$ 解的变化，原 $x_B = B^{-1}b - B^{-1}N x_N$，$x_k$ 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 $\forall x_k \Longrightarrow x_{B_i} > 0, x_k$ 可以取无限大，故 $f \to -\infty$，原问题无界。
  2. 要满足 $x_B = \bar{b} - y_k x_k \geq 0$，则取 $x_k = \min(\frac{\bar{b}}{y_{ik}}, y_{ik} > 0) = \frac{\bar{b}}{y_{rk}} > 0$，$r$ 为出基下标。

Remark 3.3. 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。

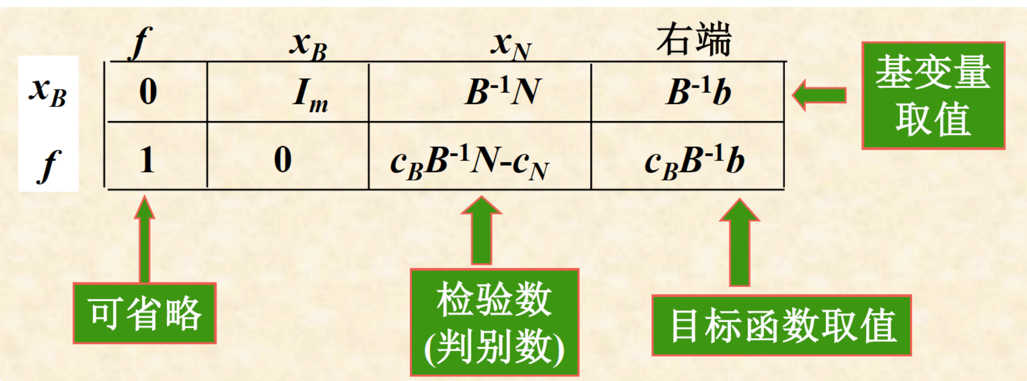


Figure 1: 单纯形表

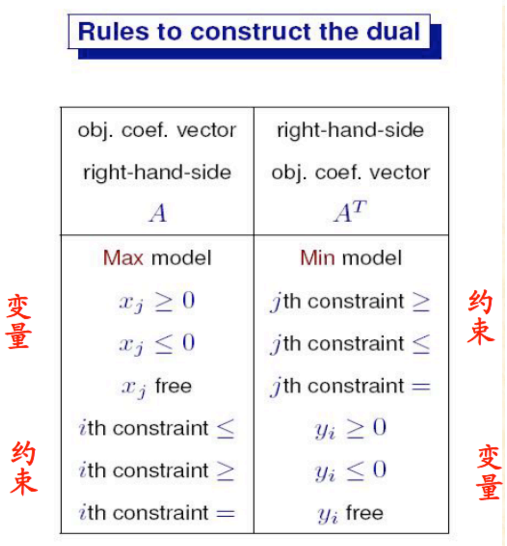
4 对偶理论

Remark 4.1. 对称形式的对偶规划的要点

- min 变成 max，价值系数 c 与右端向量 b 互换，系数矩阵 A 转置，$\geq$ 变为 $\leq$
- 原问题中约束条件的个数 = 对偶问题中变量的个数
- 原问题中变量的个数 = 对偶问题中约束条件的个数
- $\begin{cases} \min & c^T x \\ s.t. & Ax \geq b \\ & x \geq 0 \end{cases} \implies \begin{cases} \max & b^T w \\ s.t. & A^T w \leq c \\ & w \geq 0 \end{cases}$
- 对偶问题的对偶问题是原问题
- 一般情形 LP 问题的对偶问题

$$\begin{cases} \min & c^T x \\ s.t. & A_1 x \geq b_1 \\ & A_2 x = b_2 \\ & A_3 x \leq b_3 \\ & x \geq 0 \end{cases} \implies \begin{cases} \max & b_1^T w_1 + b_2^T w_2 + b_3^T w_3 \\ s.t. & A_1^T w_1 + A_2^T w_2 + A_3^T w_3 \leq c \\ & w_1 \geq 0, w_2 \text{ free}, w_3 \leq 0 \end{cases}$$

图2中，右边是原问题，左边是对偶问题。



Rules to construct the dual

	obj. coef. vector right-hand-side A	right-hand-side obj. coef. vector A^T	
	Max model	Min model	
变量	$x_j \geq 0$	jth constraint $\geq$	约束
	$x_j \leq 0$	jth constraint $\leq$	
	x_j free	jth constraint $=$	
约束	ith constraint $\leq$	$y_i \geq 0$	变量
	ith constraint $\geq$	$y_i \leq 0$	
	ith constraint $=$	y_i free	

Figure 2: LP 问题的对偶问题转换规则

$$\begin{cases} \min & 2x_1+x_2+2x_3 \\ s.t. & x_1+x_2+2x_3 \geq 1 \\ & x_1-x_2+x_3 \leq 2 \\ & -x_1+x_2+x_3=1 \\ & x_1 \geq 0, x_2 \text{ free}, x_3 \leq 0 \end{cases} \Longrightarrow \begin{cases} \max & w_1+2w_2+w_3 \\ s.t. & w_1+w_2-w_3 \leq 2 \\ & w_1-w_2+w_3=1 \\ & 2w_1+w_2+w_3 \geq 2 \\ & w_1 \geq 0, w_2 \leq 0, w_3 \text{ free} \end{cases}$$

原问题的变量对应对偶问题的约束，并且**符号改变**。

Remark 4.2. 对偶问题的基本性质

- 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。
  - Proof:
    * $Ax^{(0)} \geq b, x^{(0)} \geq 0$，同时 $A^T w^{(0)} \leq c, w^{(0)} \geq 0$，故 $c^T x^{(0)} \geq (A^T w^{(0)})^T x^{(0)} = (w^{(0)})^T A x^{(0)} \geq (w^{(0)})^T b = b^T w^{(0)}$
  - 推论 1：若原问题 (P) 或对偶问题 (D) 有无界解，则其对偶问题 (D) 或原问题 (P) 无可行解
  - 推论 2：极大化问题的任何一个可行解所对应的目标函数值都是其对偶问题的目标函数值的下界。
  - 推论 3：极小化问题的任何一个可行解所对应的目标函数值都是其对偶问题的目标函数值的上界。
- 最优性准则：若 $x^{(0)}, w^{(0)}$ 分别为原问题 (P) 对偶问题 (D) 的可行解，且 $cx^{(0)} = w^{(0)}b$，则 $x^{(0)}, w^{(0)}$ 分别为原问题 (P) 对偶问题 (D) 的最优解
- 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)(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) 的最优单纯形表中松弛变量的检验数的相反数为单纯形乘子 $w = c_B B^{-1}$，也就是其对偶问题 (D) 的最优解.

Remark 4.3. 总结：

- 原问题有最优解，对偶问题一定有最优解，且相同
- 原问题有无界解，则对偶问题无可行解

- 原问题无可行解，则对偶问题有无界解或无可行解

Remark 4.4. 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 $\forall i, j (1 \le i \le m, 1 \le j \le n)$ 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。

互补松弛定理（非对称形式）

设 $x^{(0)}$ 和 $w^{(0)}$ 分别是 $\begin{cases} \min & c^T x \\ s.t. & Ax = b \\ & x \ge 0 \end{cases}$ 和 $\begin{cases} \max & b^T w \\ s.t. & A^T w \le c \end{cases}$ 的可行解，则 $x^{(0)}$ 和 $w^{(0)}$ 是最优解的充要条件是 $\forall j$

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

Remark 4.5. 对偶单纯形法步骤：

1. 化标准型，建立初始单纯形表
2. 判断，若 $B^{-1} b \ge 0$，则已得到最优解
3. 换基迭代
   (a) 确定换出变量，$\bar{b_r} = \min_i \left\{ \bar{b_i} \right\} < 0, x_r$ 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 $y_{rj} \ge 0$，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

Remark 4.6. 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 $wP_j - c_j \le 0$，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。

5 算法概述

Remark 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。

Remark 5.2. 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：

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

Remark 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）

Remark 5.4. Q-收敛速率 (Quotient)

设序列 $\left\{\gamma^{(k)}\right\}$ 收敛于 γ^*，定义满足

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

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

Remark 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 $\|x_k - x^*\| \le \kappa q^k$，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 $\|x_k - x^*\| \le \kappa \prod_{i=1}^k q_i$，则称点列 $\{x_k\}$ R-超线性收敛到 x^*。

- Q-(超) 线性收敛 $\Longrightarrow$ R-(超) 线性收敛

Remark 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

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

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。

Remark 5.7. 单纯形算法的复杂度为指数时间复杂度。

Remark 5.8. 对于线性规划问题

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

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。

Remark 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

Remark 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。

$$\min \quad f(x)$$

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

定理. 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 $\left\{x^{(k)}\right\}$ 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。

Remark 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。