子午线

学习笔记

作者：leekarry
组织：果壳
时间：October 25, 2019
版本：0.1

ElegantLaTeX Program

我很快，快到时间都会变慢。而我一慢，时间就会过得飞快。

目　　录

第 1 章 学科简述与基本概念

1.1 学科简述

最优化理论与算法是一个重要的数学分支, 它所研究的问题是讨论在众多的方案中什么样的方案最优以及怎样找出最优方案。

1. 对于一个优化问题,通常有一个优化目标函数 $f(x)$, x 为参数变量,$c(x)$ 为约束。
2. 最优化问题的标注形式为

$$\begin{aligned} min f(x) \quad & x \in R^n \\ s.t. C_i(x) = 0 \quad & i \in \epsilon \\ C_i(x) \geq 0 \quad & i \in I \end{aligned} \tag{1.1}$$

3. 其中 ϵ 表示等式集合 I 表示不等式集合
4. 其中满足约束的解称之为 ** 可行解 **

1.2 线性与非线性规划问题

目标函数和约束函数都是线性的,称之为 ** 线性规划问题 **。数学模型中含有非线性函数, 因此称为 ** 非线性规划问题 **。在线性规划与非线性规划中, 满足约束条件的点称为 ** 可行点 **, 全体可行点组成的集合称为 ** 可行集 ** 或 ** 可行域 **。如果一个问题的可行集是整个空间。那么此问题就称为 ** 无约束问题 **。

设 $f(x)$ 为目标函数,S 为可行域,若存在 $\bar{x} \in S$,若对每个 $x \in S$,有 $f(x) \geq f(\bar{x})$,则称 $\bar{x}$ 为 $f(x)$ 在 S 上的 ** 全局极小点 **。

设 $f(x)$ 为目标函数, S 为可行域, 若存在 $\bar{x} \in S$, 若存在 $x \in S, \epsilon > 0$ 邻域 $N(\bar{x}, \epsilon) = \{x \| x - \bar{x} \| < \epsilon\}$,使得对每个 $x \in S \cap N(\bar{x}, \epsilon)$, 有 $f(x) \geq f(\bar{x})$, 则称 $\bar{x}$ 为 $f(x)$ 在 S 上的 ** 局部极小点 **。

1.3 几个数学概念

向量范数和矩阵范数

1. 向量范数

$$\|x\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{\frac{1}{p}} \tag{1.2}$$

2. 矩阵范数设 A 为 $m \times n$ 矩阵, $\| \cdot \|_\alpha$ 是 R^m 上向量范数, $\| \cdot \|_\beta$ 是 R^n 上向量范数, 定义矩阵范数

$$\|A\| = \max_{\|x\|_\beta - 1} \|Ax\|_\alpha \tag{1.3}$$

序列的极限

设 $\{x^{(k)}\}$ 是 R^n 中一个向量序列, $\bar{x} \in R^n$, 如果对每个任给的 $\epsilon > 0$ 存在正整数 K, 使得当 $k > K_\epsilon$ 时就有 $\|x^{(k)} - \bar{x}\| < \epsilon$, 则称序列收敛到 $\bar{x}$, 或称序列以 $\bar{x}$ 为**极限**, 记作

$$\lim_{k\to\infty} x^{(k)} = \bar{x} \tag{1.4}$$

$\bar{x}$ 称为序列的一个**聚点**。

梯度、Hesse 矩阵、Taylor 展开式

函数 f 在 x 处的梯度为 n 维列向量:

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n}\right]^T \tag{1.5}$$

f 在 x 处的 Hesse 矩阵为 $n \times n$ 矩阵 $\nabla^2 f(x)$, 第 i 行第 j 列元素为

$$[\nabla^2 f(x)]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad 1 \le i, j \le n \tag{1.6}$$

Jacobi 矩阵、链式法则和隐函数存在定理

1.4 凸集和凸函数

凸集

凸集和凸函数是线性规划和非线性规划都要涉及的基本概念。关于凸集和凸函数的一些定理在最优化问题的理论证明及算法研究中具有重要作用。设 S 为 n 维欧氏空间 R^n 中一个集合。若对 S 中任意两点, 联结它们的线段仍属于 S; 换言之, 对 S 中任意两点 $x^{(1)}, x^{(2)}$ 及每个实数 $\lambda \in [0, 1]$, 都有

$$\lambda x^{(1)} + (1-\lambda)x^{(2)} \in S \tag{1.7}$$

则称 S 为**凸集**。

设有集合 $C \subset R^n$, 若对 C 中每一点 x, 当 λ 取任何非负数时, 都有 $\lambda x \in C$, 则称 C 为**锥**, 又若 C 为凸集, 则称 C 为**凸锥**。

有限个半空间的交

$$\{x | Ax \le b\}$$

称为**多面集**, 其中 A 为 $m \times n$ 矩阵, b 为 m 维向量。

设 S 为非空凸集, $x \in S$, 若 x 不能表示成 S 中两个不同点的凸组合; 换言之, 若假设 $x = \lambda x^{(1)} + (1-\lambda)x^{(2)} (\lambda \in (0,1)), x^{(1)}, x^{(2)} \in S$, 必推得 $x = x^{(1)} = x^{(2)}$, 则称 x 是凸集 S 的**极点**。

设 S 为 R^n 中的闭凸集,d 为非零向量,如果对 S 中的每一个 x,都有射线

$$\{x + \lambda d | \lambda \leq 0\} \subset S \tag{1.8}$$

则称向量 d 为 S 的方向,又设 $d^{(1)}$和$d^{(2)}$ 是 S 的两个方向,若对任何正数 λ 有 $d^{(1) \neq \lambda d^{(2)}}$,则称 $d^{(1)}$ 和 $d^{(2)}$ 是两个不同的方向。若 S 的方向 d 不能表示成该集合的两个不同方向的正的线性组合,则称 d 为 S 的**极方向**。

凸集的另一个重要性质是**分离定理**.

设 S_1 和 S_2 是 R^n 中两个非空集合,$H = \{x | P^T x = \alpha\}$ 为超平面。如果对每个 $x \in S_1$,都有 $P^T x \geq \alpha$,对于每个 $x \in S_2$,都有 $P^T x \leq \alpha$,则称超平面 H 分离集合 S_1 和 S_2。

凸函数

设 S 为 R^n 中的非空凸集,f 是定义在 S 上的实函数。如果对任意的 $x^{(1)}, x^{(2)} \in S$ 及每个数 $\lambda \in (0, 1)$,都有

$$f(\lambda x^{(1)} + (1 + \lambda) x^{(2)}) \leq \lambda f(x^{(1)}) + (1 + \lambda) f(x^{(2)}) \tag{1.9}$$

则称 f 为 S 上的 ** 凸函数 **。

凸函数的判别

利用凸函数的定义及有关性质可以判别一个函数是否为凸函数, 但有时计算比较复杂,使用很不方便,因此需要进一步研究凸函数的判别问题。

1. 设 S 是 R^n 中非空开凸集,$f(x)$ 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 $x^{(1)}, x^{(2)} \in S$,都有

$$f(x^{(2)}) \geq f(x^{(1)}) + \nabla f(x^{(1)})^T (x^{(2)} - x^{(1)}) \tag{1.10}$$

2. 设 S 是 R^n 中非空开凸集, $f(x)$ 是定义在 S 上的二次可微函数, 如果在每一点 $x \in S$, $Hesse$ 矩阵正定,则 $f(x)$ 为严格凸函数。

凸规划

对于标准形式目标函数为凸函数,等式约束为线性约束;不等式约束为凹函数。

第 2 章 对偶原理及灵敏度分析

第 3 章 最优性条件

第 4 章 线性规划

线性规划是数学规划的一个重要分支，它在理论和算法上都比较成熟，在实践上有着广泛的应用，不仅许多实际课题属于线性规划问题，而且运筹学其他分支中的一些问题也可以转化为线性规划来计算，因此线性规划在最优化学科中占有重要地位。

4.1 标准形式及图解法

一般线性规划问题总可以写成下列标准形式：

$$\begin{aligned} min \quad & cx \\ s.t. \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{4.1}$$

其中 A 是 $m \times n$ 矩阵，c 是 n 维行向量，b 是 m 维列向量。

基本性质

1. 可行域。线性规划的可行域是凸集。
2. 最优极点。设线性规划的可行域非空，则有下列结论：
   (a). 线性规划存在有限最优解的条件是所有 $cd^{(j)}$ 为非负数。其中 $d^{(j)}$ 是可行域的极方向。
   (b). 若线性规划存在有限最优解，则目标函数的最优值可在某个极点上达到。
3. 最优基本可行解。极点是个几何概念，有直观性强的优点，但不便于演算，因此需要研究极点的代数含义。

$$x = \begin{bmatrix} x_B \\ x_N \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} \tag{4.2}$$

称为方程组 $Ax = b$ 的一个**基本解**。B 称为基矩阵。简称为基。x_B 的各分量称为基变量，基变量的全体 $x_{B_1}, x_{B_2}, \ldots, x_{B_m}$ 称为一组基。x_N 的各分量称为非基变量。又若 $B^{-1}b \geq 0$，则称 x 为约束条件 $Ax = b, x \geq 0$ 的基本可行解。相应地，称 B 为可行基矩阵，$x_{B_1}, x_{B_2}, \ldots, x_{B_m}$ 为一组可行基。若 $B^{-1}b > 0$，即基变量的取值均为正数，则称基本可行解是非退化的。如果满足 $B^{-1}b \geq 0$ 且至少有一个分量是零，则称基本可行解是退化的基本可行解。

令 $K = \{x|Ax = b, x \geq 0\}$，$A$ 是 $m \times n$ 矩阵，A 的秩为 m，则 K 的极点集与 $Ax = b, x \geq 0$ 的基本可行解集等价。

当线性规划存在最优解时，则一定存在一个基本可行解，它是最优解。这样，线性规划问题的求解，可归结为求最优基本可行解。

4. 基本可行解的存在问题。如果 $Ax=b, x\geq 0$ 有可行解，则一定存在基本可行解。其中 A 是 $m\times n$ 矩阵，A 的秩为 m。

4.2 单纯形方法

单纯形方法原理

若线性规划（标准形式）有最优解，则必存在最优基本可行解。因此求解线性规划问题归结为找最优基本可行解。单纯形方法的基本思想，就是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；通过不断改进基本可行解，力图达到最优基本可行解。

考虑问题：

$$\begin{aligned} min \quad & f \overset{def}{=} cx \\ s.t. \quad & Ax=b, \\ & x\geq 0, \end{aligned} \tag{4.3}$$

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

$$A=(p_1,p_2,\ldots,p_n) \tag{4.4}$$

现将 A 分解成 (B,N)，使得其中 B 是基矩阵，N 是非基矩阵，设

$$x^{(0)}=\begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} \tag{4.5}$$

是基本可行解，在 $x^{(0)}$ 处的目标函数值

$$\begin{aligned} f_0 = cx^{(0)} &= (c_B, c_N)\begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} \\ &= c_B B^{-1}b, \end{aligned} \tag{4.6}$$

现在分析怎样从基本可行解 $x^{(0)}$ 出发，求一个改进的基本可行解，设

$$x=\begin{bmatrix} x_B \\ x_N \end{bmatrix} \tag{4.7}$$

是任一个可行解，则由 $Ax=b$ 得到

$$\begin{aligned} AX=(B,N)\begin{pmatrix} x_B \\ x_N \end{pmatrix} &= Bx_B+Nx_N=b \\ \Longrightarrow\ x_B &= B^{-1}b-B^{-1}Nx_N \end{aligned} \tag{4.8}$$

在点 x 处的目标函数值

$$
\begin{aligned}
f = cx &= (c_B, c_N)\begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} \\
&= c_B x_B + c_N x_N \\
&= c_B(B^{-1}b - B^{-1}N x_N) + c_N x_N \\
&= c_B B^{-1} b - (c_B B^{-1} N - c_N) x_N \\
&= f_0 - \sum_{j\in R} (c_B B^{-1} p_j - c_j) x_j \\
&= f_0 - \sum_{j\in R} (z_j - c_j) x_j
\end{aligned}
\tag{4.9}
$$

适当选取自由未知量 $x_j(j \in R)$ 的数值就有可能使得

$$\sum_{j\in R} (z_j - c_j) x_j > 0 \tag{4.10}$$

这里假设 $z_k - c_k > 0, x_k$ 由零变为正数后，得到方程组 $Ax = b$ 的解

$$x_B = B^{-1}b - B^{-1}p_k x_k = \bar{b} - y_k x_k \tag{4.11}$$

其中 $\bar{b}$ 和 y_k 是 m 维列向量，$\bar{b} = B^{-1}b, y_k = B^{-1}p_k$。从而得到使目标函数值减少的新的基本可行解。因此选择 x_k，使

$$z_k - c_k = max_{j\in R}\{z_j - c_j\} \tag{4.12}$$

在新得到的点，目标函数值是

$$f = f_0 - (z_k - c_k)x_k \tag{4.13}$$

再来分析怎样确定 x_k 的取值。一方面，x_k 取值越大函数值下降越多；另一方面，x_k 的取值受到可行性的限制，它不能无限增大。

$$x_k \le \frac{\bar{b}_i}{y_{ik}} \tag{4.14}$$

因此，为使 $x_B \ge 0$ 应令

$$x_k = min\left\{ \frac{\bar{b}_i}{y_{ik}} \middle| y_{ik} > 0 \right\} = \frac{\bar{b}_r}{y_{rk}} \tag{4.15}$$

x_k 取值 $\frac{\bar{b}_r}{y_{rk}}$ 后，原来的基变量 $x_B = 0$，得到新的可行解

$$x = (\overbrace{x_{B_1}, \ldots, x_{B_{r-1}}, 0, x_{b_{r+1}}, \ldots}^{x_B}, \underbrace{0, \ldots, x_k, 0, \ldots, 0}_{x_N})^T \tag{4.16}$$

这个解一定是基本可行解。经上述转换，x_k 由原来的非基变量变成基变量，而原来的基变

量 x_{B_r} 变成非基变量。在新的基本可行解处，目标函数值比原来减少了 $(z_k - c_k)x_k$。重复以上过程，可以进一步改进基本可行解，直到所有 $z_j - c_j$ 均非正数，以致任何一个非基变量取正值都不能使目标函数值减少时为止。

在线性规划中，通常称 $z_j - c_j$ 为**判别数**或**检验数**。

单纯性方法计算步骤

1. 解 $Bx_B = b$，求得 $x_B = B^{-1}b = \bar{b}$，令 $x_N = 0$，计算目标函数值 $f = c_B x_B$；
2. 求单纯性乘子 w，解 $wB = c_B$，得到 $w = c_B B^{-1}$。对于所有非基变量，计算判别数 $z_j - c_j = wp_j - c_j$ 令
$$z_k - c_k = max_{j \in R}\{z_j - c_j\}$$
若 $z_k - c_k \le 0$，则对于所有非基变量 $z_j - c_j \le 0$，对应基变量的判别数总是零，因此停止计算，现行基本可行解是最优解。否则，进行下一步；
3. 解 $By_k = p_k$，得到 $y_k = B^{-1}p_k$，若 $y_k \le 0$，即 y_k 的每个分量均非正数，则停止计算，问题不存在有限最优解。否则，进行步骤(4)；
4. 确定下标 r，使
$$\frac{\bar{b}_r}{y_{rk}} = min\left\{\frac{\bar{b}_i}{y_{ik}}\middle| y_{ik} > 0\right\} \tag{4.17}$$
x_{B_r} 为离基变量，x_k 为进基变量，用 p_k 替换 p_{B_r}，得到新的基矩阵 B，返回步骤(1)。

收敛性

对于非退化问题，单纯性方法经有限次迭代或达到最优基本可行解，或得出无界的结论。

第 5 章　非线性规划