第一章　非线性凸规划

Definition 1.0.1. 非线性规划的数学模型

$$\min f(X)$$

$$\text{s.t.} \quad h_i(X) = 0 \quad i = 1, 2, \ldots, m$$

$$g_j(X) \geq 0 \quad j = 1, 2, \ldots, l$$

$$X \in R^n$$

何老师的问题:非线性规划找的是局部最小点还是全局最小点?

实际上对于非线性凸规划，我们找到的局部最小点就是全局最小点。对于复杂的非线性规划，实际上我们不确定找到的是局部最小点还是全局最小点。

Definition 1.0.2. 凸函数与凹函数

凸函数为满足下列条件的函数:

$$f[\lambda X_1 + (1-\lambda)X_2] \leq \lambda f(X_1) + (1-\lambda) f(X_2) \quad 0 < \lambda < 1 \quad X_1, X_2 \in R^n$$

Proof. 证明凸函数的局部最小点就是全局最小点

对于任意 $\theta \in [0,1]$，验证：

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

展开左边并整理后，比较两边的差值为：

$$\frac{\theta(1-\theta)}{2}(x-y)^T A(x-y) \leq 0$$

由于 A 半正定且 $\theta(1-\theta) \geq 0$，不等式成立，故 f 是凸函数。 □

凸函数的性质

1. 凸函数非负线性组合仍为凸函数

$$\gamma_1 f_1(X) + \gamma_2 f_2(X) \quad \gamma_i \geq 0$$

2. 若 $f(X)$ 是定义在凸集 R_C 上的凸函数，则其 β 水平集 S_β 为凸集。

$$S_\beta = \{X \mid f(X) \leq \beta, X \in R_C\} \quad \beta\text{水平集}$$

3. 对于凸函数 $f(X)$，若存在 $X^* \in R^n$ 满足

$$\nabla f(X^*)^T (X - X^*) \geq 0 \quad \forall X \in R^n$$

则 X^* 为 $f(X)$ 的全局最小点。充要条件!

(1)(3)显然,(3)先证明是局部最小点，然后证明的全局最小点。

对于(2)，直接使用定义是可以证明的。

Definition 1.0.3. 凸规划问题

$$\min f(X)$$

s.t. $h_i(X) = 0 \quad i = 1, 2, ..., m$ 　线性函数

$$g_j(X) \geq 0 \quad j = 1, 2, ..., l \quad \text{凹函数}$$

$$X \in R^n$$

实际上第二个凹函数的要求，如果是左右同乘一个负号，实际上是一个凸函数。所以这个规划问题的可行域也是个凸集(线性函数是凸函数也是凹函数)。

Theorem 1.0.1. 若目标函数为严格凸函数，则如果全局最优解存在必为唯一全局最优解。（反证法）

如果不是严格，我们用定义可以说明凸函数的最优解解集一定是各个最优解的凸包

Definition 1.0.4. 海森矩阵

$$\nabla^2 f(X) = \begin{bmatrix} \frac{\partial^2 f(X)}{\partial x_1^2} & \frac{\partial^2 f(X)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(X)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(X)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(X)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(X)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(X)}{\partial x_n \partial x_1} & \frac{\partial^2 f(X)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(X)}{\partial x_n^2} \end{bmatrix}$$

$$f(X+\Delta X)=f(X)+\nabla f(X)^T\Delta X+\frac{1}{2}\Delta X^T\nabla^2 f(X)\Delta X+O(||\Delta X||^2)$$

如果某处梯度为0，海森矩阵为半正定，实际上这只是局部极小值的必要条件。除非加强为领域梯度均大于等于0。

Definition 1.0.5. 拉格朗日乘数法(针对等式约束):

定义Lagrange函数

$$\min L(X,\lambda)=f(X)+\lambda^T h(X)=f(X)+\sum_{i=1}^{m}\lambda_i h_i(X)$$

$$\lambda=\begin{bmatrix}\lambda_1 & \lambda_2 & \cdots & \lambda_m\end{bmatrix}^T$$

求无约束极值问题：

$$\left.\frac{\partial L(X,\lambda)}{\partial X}\right|_{X^*}=0 \quad\Rightarrow\quad \nabla f(X^*)+\sum_{i=1}^{m}\lambda_i\nabla h_i(X^*)=0$$

$$\left.\frac{\partial L(X,\lambda)}{\partial \lambda}\right|_{\lambda^*}=0 \quad\Rightarrow\quad h_i(X^*)=0 \quad i=1,2,\ldots,m$$

不等式约束极值问题:

$$\min f(X)$$

$$\text{s.t.}\quad g_j(X)\geq 0 \quad j=1,2,\ldots,l$$

$$X\in R^n$$

Definition 1.0.6. 定义：令目标函数值 $f(X^{(0)})$ 下降的方向，满足

$$\exists\lambda_0>0,\ \text{对于}\ \forall\lambda\in(0,\lambda_0],\ \text{有}\ f(X^{(0)}+\lambda P)<f(X^{(0)})$$

判别条件: $\nabla f(X^{(0)})^T P<0$

Theorem 1.0.2. Gordan 引理

A_j 为一组已知向量，不存在向量 P，使

$$A_j^T P<0 \quad j=1,2,...,l$$

同时成立的充要条件：

存在不全为零的非负实数 $\mu_j \geq 0$，使

$$\sum_{j=1}^{l} \mu_j A_j = 0$$

几何含义：A_j 不可能分布在任何超平面的同一侧。

Proof. 用Farkas引理证明Gordan引理

Farkas引理：对于矩阵$A \in \mathbb{R}^{m\times n}$和向量$b \in \mathbb{R}^m$，以下两个系统恰有一个有解：

1. 存在$x \geq 0$使得$Ax = b$;
2. 存在$y \in \mathbb{R}^m$使得$y^\top A \geq 0^\top$且$y^\top b < 0$。

证明Gordon引理：

1. **构造辅助系统：**

   将系统A加强为存在满足归一化条件的解。令 $e = (1, 1, \ldots, 1)^\top$，定义新系统：

$$\exists x \geq 0, \quad \begin{cases} Ax = 0 \\ e^\top x = 1 \end{cases}$$

   若此系统有解，则原系统A有非零解；反之则无。

2. **应用Farkas引理：**

   将上述系统写作矩阵形式 $\begin{bmatrix} A \\ e^\top \end{bmatrix} x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$。根据Farkas引理，恰有以下之一成立：

   (a) 存在 $x \geq 0$ 满足方程

   (b) 存在 $\begin{bmatrix} y \\ \lambda \end{bmatrix} \in \mathbb{R}^{m+1}$ 使得：

$$y^\top A + \lambda e^\top \geq 0^\top \quad \text{且} \quad y^\top 0 + \lambda \cdot 1 < 0$$

   第二个不等式简化为 $\lambda < 0$。

3. **分析对偶条件：**

   若原系统无解，则存在 y 和 $\lambda < 0$ 使得：

$$y^\top A_j + \lambda \geq 0 \quad (\forall j = 1, \ldots, n)$$

   令 $\mu = -\lambda > 0$，则有：

$$y^\top A_j \geq \mu > 0 \quad (\forall j)$$

即 $A^\top y > 0$，故系统B成立。

4. **反方向证明：**

   若系统B有解（存在 y 使 $A^\top y > 0$），假设系统A也有解（存在非零 $x \geq 0$ 使 $Ax = 0$），则：

$$y^\top Ax = (A^\top y)^\top x > 0 \quad (\text{因 } A^\top y > 0 \text{ 且 } x \geq 0, \neq 0)$$

   但左边等于 $y^\top 0 = 0$，矛盾。故系统B存在时系统A无解。

综上，系统A与系统B恰有一个成立，Gordon引理得证。 □

Theorem 1.0.3. Fritz John条件

若X*是局部极小点，存在不全为零的非负实数 $\mu_j \geq 0, j \in J(X*) \cup \{0\}$，使

$$\mu_0 \nabla f(X*) - \sum_{j \in J} \mu_j \nabla g_j(X*) = 0$$

Proof. 证明：Gordan引理的直接推论(不存在一个可行的下降方向)

$$\nabla f(X*)^T P < 0$$

$$-\nabla g_j(X*)^T P < 0$$

□

Remark. 等价Fritz John条件(将J显式化)

若X^*是局部极小点，则存在不全为零的非负实数 $\mu_j \geq 0$，$j = 0, 1, \ldots, l$，使下列条件成立：

$$\begin{cases} \mu_0 \nabla f(X^*) - \sum_{j=1}^{l} \mu_j \nabla g_j(X^*) = 0 & \text{Lagrange驻点条件} \\ \mu_j g_j(X^*) = 0 & j = 1, 2, \ldots, l \quad \text{互补松弛条件} \\ \mu_j \geq 0 & \sum_{j=0}^{m} \mu_j \neq 0 \quad j = 0, 1, \ldots, l \quad \text{强非负条件} \end{cases}$$

上述条件隐含了如下事实：$\mu_j = 0 \quad j \notin J$

Definition 1.0.7. Kuhn-Tucker条件

若Fritz John条件的$\mu_0 \geq 0$，则可推出KKT条件的数学形式：

$$\begin{cases} \nabla f(X^*) - \sum_{j=1}^{l} \mu_j^* \nabla g_j(X^*) = 0 & \text{Lagrange驻点条件} \\ \mu_j^* g_j(X^*) = 0 & j = 1, 2, \ldots, l \quad \text{互补松弛条件} \\ \mu_j^* \geq 0 & j = 1, 2, \ldots, l \quad \text{非负条件} \end{cases}$$

Karush(1939), Kuhn-Tucker (1951)

何老师的问题:什么条件下$\mu_0 \geq 0$?

在加上约束规格限制时$\mu_0 \geq 0$，比如说正则条件

- **Fritz John点**：满足Fritz John条件（允许目标函数乘子为零）
- **KKT点**：满足KKT条件（需约束规格，乘子非负）
- **正则KKT点**：满足LICQ或严格互补的KKT点
- **极值点**：局部最优解
  - **包含关系**：

    $$\text{Fritz John点} \supsetneq \text{KKT点} \supsetneq \text{正则KKT点}$$

  - **极值点归属**：
    * 若满足约束规格：极值点 $\subset$ KKT点
    * 若不满足约束规格：极值点 $\subset$ Fritz John点
  - **非极值点存在性**：
    * 存在非极值的KKT点（如鞍点）
    * 存在非极值的Fritz John点（目标乘子为零）

等价不等式约束极值问题（等式约束改写成不等式）

$$\min f(X)$$

$$\text{s.t.} \quad h_i(X) \geq 0 \quad i = 1, 2, \ldots, m$$

$$-h_i(X) \geq 0 \quad i = 1, 2, \ldots, m$$

$$g_j(X) \geq 0 \quad j = 1, 2, \ldots, l$$

$$X \in R^n$$

问题：此时约束是否满足正则条件？

显然不满足，等式约束的梯度函数线性相关

一般约束的KKT条件的推导

$$\mu_0^* \nabla f(X^*) - \sum_{j=1}^{l} \mu_j^* \nabla g_j(X^*) - \sum_{i=1}^{m} \mu_i'^* \nabla h_i(X^*) + \sum_{i=1}^{m} \mu_i''^* \nabla h_i(X^*) = 0$$

$$\mu_0^* \nabla f(X^*) - \sum_{j=1}^{l} \mu_j^* \nabla g_j(X^*) - \sum_{i=1}^{m} (\mu_i'^* - \mu_i''^*) \nabla h_i(X^*) = 0$$

$$\mu_0^* \nabla f(X^*) - \sum_{j=1}^{l} \mu_j^* \nabla g_j(X^*) - \sum_{i=1}^{m} \gamma_i^* \nabla h_i(X^*) = 0$$

$$\mu_i'^* \geq 0 \quad \mu_i''^* \geq 0 \quad \Rightarrow \quad \gamma_i^* = \mu_i'^* - \mu_i''^* \text{无符号约束} \quad i = 1, 2, \ldots, p$$

注意 μ_0, μ_j, γ_i 不可同时为0!

Definition 1.0.8. 一般约束下的Fritz John条件

若X*是局部极小点，存在不全为零的μj (j=0,1,2,...,m)和γi (i=0,1,2,...,p)，满足

$$\mu_0^* \nabla f(X^*) - \sum_{j=1}^{l} \mu_j^* \nabla g_j(X^*) - \sum_{i=1}^{m} \gamma_i^* \nabla h_i(X^*) = 0 \quad \text{Lagrange驻点条件}$$

$$\mu_j^* g_j(X^*) = 0 \quad j = 1, 2, ..., m \quad \text{互补松弛条件}$$

$$\mu_j^* \geq 0 \quad j = 0, 1, ..., m$$

$$\sum_{j=0}^{l} \mu_j^* + \sum_{i=1}^{m} |\gamma_i^*| \neq 0 \quad \text{强非负条件}$$

$$\gamma_i^* h_i(X^*) = 0 \quad i = 1, 2, ..., p \quad \text{等式互补松弛条件}$$

Definition 1.0.9. KKT条件的矩阵形式

$$\nabla f(X^*) - \nabla h(X^*)\gamma^* - \nabla g(X^*)\mu^* = 0$$

$$\mu^* \odot g(X^*) = 0 \quad \Leftrightarrow \quad \mu_j^* g_j(X^*) = 0 \quad j = 1, 2, \ldots, l$$

其中

$$g(X^*) \ge 0$$

$$h(X^*) = 0$$

$$\mu^* \ge 0$$

Theorem 1.0.4. 凸规划下的KKT条件为最优解的充要条件

Proof. KKT条件的充分性证明

$$\nabla f(X^*) - \sum_{i=1}^{m} \gamma_i^* \nabla h_i(X^*) - \sum_{j=1}^{l} \mu_j^* \nabla g_j(X^*) = 0 \quad \mu_j^* g_j(X^*) = 0 \quad \mu_j^* \ge 0$$

$f(X)$是凸函数 $\Rightarrow f(X) \ge f(X^*) + \nabla f(X^*)^T (X - X^*)$

$$= f(X^*) + \sum_{i=1}^{m} \gamma_i^* \nabla h_i(X^*)^T (X - X^*) + \sum_{j=1}^{l} \mu_j^* \nabla g_j(X^*)^T (X - X^*)$$

$h_i(X)$线性函数且 $h_i(X) = 0 \Rightarrow \nabla h_i(X^*)^T (X - X^*) = h_i(X) - h_i(X^*) = 0$ $g_j(X)$是凹函数 $\Rightarrow \nabla g_j(X^*)^T (X - X^*) \ge g_j(X) - g_j(X^*)$ 根据互补松弛性条件 $\mu_j^* g_j(X^*) = 0$ 且 $g_j(X) \ge 0$，有

$$f(X) \ge f(X^*) + \sum_{j=1}^{l} \mu_j^* \left[g_j(X) - g_j(X^*) \right] = f(X^*) + \sum_{j=1}^{l} \mu_j^* g_j(X) \ge f(X^*)$$

□