

5.8 拉格朗日乘子法

SMaLL

¹ 中国石油大学（华东）

SMaLL 课题组

small.sem.upc.edu.cn

liangxijunsd@163.com

2023

拉格朗日乘子法

1. 惩罚函数和增广拉格朗日函数

1.1 二次惩罚函数法

1.2 乘子方法的主要思想

惩罚函数

惩罚函数的基本思想:

代价函数 \rightarrow 添加一个惩罚项，从而增加不可行点的代价。

与这些方法相关:

- (1) 惩罚项的权重 c 称为惩罚参数
- (2) 它决定了惩罚的程度以及无约束问题与原始问题的近似程度
- (3) 当 c 取值越大时，近似效果越好。

约束问题

首先考虑等式约束问题

$$\begin{aligned} & \text{minimize } f(x) \\ & \text{subject to } h(x) = 0, \quad x \in X \end{aligned} \tag{1}$$

其中 $f: \mathbf{R}^n \rightarrow \mathbf{R}$, $h: \mathbf{R}^n \rightarrow \mathbf{R}^m$ 是给定的函数, X 是给定的 \mathbf{R}^n 的子集.

二阶充分性条件

假设 x^* 、 λ^* 是满足定理 1 的最优解和拉格朗日乘子向量。

定理 1: (二阶充分性条件) 设 f 和 h 二阶连续可微, 令 $x^* \in \mathbf{R}^n$ 且 $\lambda^* \in \mathbf{R}^m$ 满足

$$\nabla_x L(x^*, \lambda^*) = 0, \quad \nabla_\lambda L(x^*, \lambda^*) = 0$$

$$y' \nabla_{xx}^2 L(x^*, \lambda^*) y > 0$$

对于任意 $y \neq 0$ 且 $\nabla h(x^*)' y = 0$. 则 x^* 是 f 满足 $h(x) = 0$ 的局部最小值. 实际上存在 $\gamma > 0$ $\epsilon > 0$ 使得

$$f(x) \geq f(x^*) + \frac{\gamma}{2} \|x - x^*\|^2, \quad \forall x \text{ with } h(x) = 0 \text{ and } \|x - x^*\| < \epsilon$$

增广拉格朗日函数

增广拉格朗日函数 $L_c : \Re^n \times \Re^m \mapsto \Re$:

$$L_c(x, \lambda) = f(x) + \lambda' h(x) + \frac{c}{2} \|h(x)\|^2$$

其中 c 是惩罚参数

两种方法

通过最小化 $L_c(\cdot, \lambda)$ 来获得接近 x^* 的解有两种方法:

(a) 令 λ 接近 λ^* .

如果 c 大于某一确定的数值, 那么对于 $\gamma > 0$ 且 $\epsilon > 0$:

$$L_c(x, \lambda^*) \geq L_c(x^*, \lambda^*) + \frac{\gamma}{2} \|x - x^*\|^2, \quad \forall x \text{ with } \|x - x^*\| < \epsilon$$

对应于 λ^* 的增广拉格朗日函 $L_c(\cdot, \lambda^*)$ 在 x^* 处取到严格局部极小值。这表明, 如果 λ 接近 λ^* , 可以通过无约束地最小 $L_c(\cdot, \lambda^*)$ 来找到 x^* 的一个很好的近似。

两种方法

(b) 设置充分大的惩罚系数 c .

- ▶ 对于很大的 c , 因为取不可行点的代价很高, 所以 $L_c(\cdot, \lambda)$ 的无约束最小值将是几乎可行的.
- ▶ 对于可行解 x 有 $L_c(x, \lambda) = f(x)$, 所以对于近似可行解 x 有 $L_c(x, \lambda) \approx f(x)$.
- ▶ 因此, 当 c 较大时, 我们希望通过无约束最小化问题 $\min L_c(\cdot, \lambda)$ 来获得 x^* 的良好近似。

示例

考虑以下二元函数极小化问题

$$\begin{array}{ll}\text{minimize} & f(x) = \frac{1}{2} (x_1^2 + x_2^2) \\ \text{subject to} & x_1 = 1\end{array}\quad (2)$$

最优解 $x^* = (1, 0)$ 对应拉格朗日乘子 $\lambda^* = -1$. 因此, 增广拉格朗日函数为

$$L_c(x, \lambda) = \frac{1}{2} (x_1^2 + x_2^2) + \lambda (x_1 - 1) + \frac{c}{2} (x_1 - 1)^2$$

令该函数梯度为零, 我们可以证明它有唯一的最小值点 $x(\lambda, c)$ 其坐标为

$$x_1(\lambda, c) = \frac{c - \lambda}{c + 1}, \quad x_2(\lambda, c) = 0 \quad (3)$$

示例

因此，对于所有的 $c > 0$,

$$\lim_{\lambda \rightarrow \lambda^*} x_1(\lambda, c) = x_1(-1, c) = 1 = x_1^*, \quad \lim_{\lambda \rightarrow \lambda^*} x_2(\lambda^*, c) = 0 = x_2^*$$

即当 λ 接近 λ^* , $L_c(x, \lambda)$ 的最小值点接近有约束情形下的最优解, (如图 1).

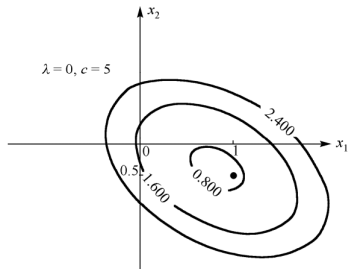
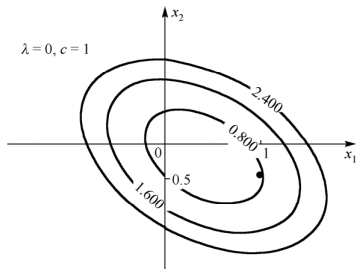


图: 1 增广拉格朗日函数的等值线

$$L_c(x, \lambda) = \frac{1}{2} (x_1^2 + x_2^2) + \lambda (x_1 - 1) + \frac{c}{2} (x_1 - 1)^2$$

对于 $c = 1$ 和两个不同的 λ . $L_c(x, \lambda)$ 的无约束最小值接近约束最小值 $x^* = (1, 0)$ 当 $\lambda \rightarrow \lambda^* = -1$

$$x_1(\lambda, c) = \frac{c - \lambda}{c + 1}, \quad x_2(\lambda, c) = 0 \quad (2)$$

利用 Eq. (2), 我们得到对于所有 λ ,

$$\lim_{c \rightarrow \infty} x_1(\lambda, c) = 1 = x_1^*, \quad \lim_{c \rightarrow \infty} x_2(\lambda, c) = 0 = x_2^*$$

表明随着 c 的增加, 无约束最小值 $L_c(x, \lambda)$ 接近约束最小值。(如图 2).

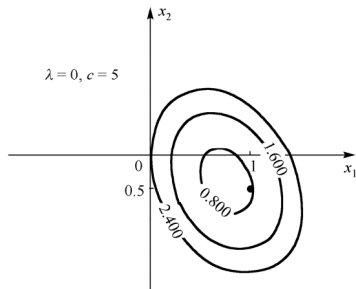
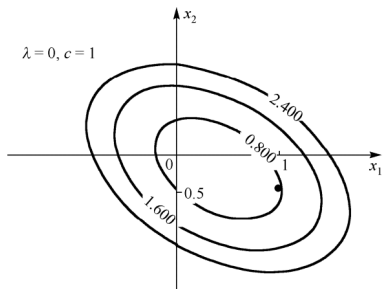


图: 2 增广拉格朗日函数的等值线

$$L_c(x, \lambda) = \frac{1}{2} (x_1^2 + x_2^2) + \lambda (x_1 - 1) + \frac{c}{2} (x_1 - 1)^2$$

对于示例, 当 $\lambda = 0$ 且两个不同的值 c . $L_c(x, \lambda)$ 的无约束最小值接近约束最小值 $x^* = (1, 0)$ 当 $c \rightarrow \infty$.

二次惩罚函数法

二次惩罚函数法

它求解一系列如下形式的问题

$$\text{minimize } L_{c^k}(x, \lambda^k), \text{ subject to } x \in X$$

$\{\lambda^k\}$ 是 \mathbf{R}^m 中的一个序列, $\{c^k\}$ 是一个正惩罚参数序列, $c^k \rightarrow \infty$

定理

定理 2 [收敛] 假设 f 和 h 是连续函数, X 是闭集, 约束集 $\{x \in X \mid h(x) = 0\}$ 非空. 对于 $k = 0, 1, \dots$, 设 x^k 为如下问题的全局最优解

$$\begin{aligned} & \text{minimize} && L_{c^k}(x, \lambda^k) \\ & \text{subject to} && x \in X \end{aligned} \tag{4}$$

其中 $\{\lambda^k\}$ 是有界的, $0 < c^k < c^{k+1}$, 且 $c^k \rightarrow \infty$.

\Rightarrow 序列 $\{x^k\}$ 的每个极限点都是原问题的全局最优解。

证明

证明:

令 \bar{x} 是序列 $\{x^k\}$ 的极限. 根据 x^k 的定义, 有

$$L_{c^k}(x^k, \lambda^k) \leq L_{c^k}(x, \lambda^k), \quad \forall x \in X \quad (5)$$

记 f^* 为原问题 (1) 式的最优值, 则有

$$\begin{aligned} f^* &= \inf_{h(x)=0, x \in X} f(x) \\ &= \inf_{h(x)=0, x \in X} \left\{ f(x) + \lambda^{k'} h(x) + \frac{c^k}{2} \|h(x)\|^2 \right\} \\ &= \inf_{h(x)=0, x \in X} L_{c^k}(x, \lambda^k) \end{aligned}$$

因此，通过取 Eq. (5) 式右边在条件 $x \in X, h(x) = 0$ 的下界，得：

$$L_{c^k}(x^k, \lambda^k) = f(x^k) + \lambda^{k'} h(x^k) + \frac{c^k}{2} \|h(x^k)\|^2 \leq f^*.$$

因为序列 $\{\lambda^k\}$ 是有界的 \Rightarrow 所以该序列有极限 $\bar{\lambda}$.

不失一般性，假设 $\lambda^k \rightarrow \bar{\lambda}$.

在上述不等式两边取上极限, 利用 f 和 h 的连续性:

$$f(\bar{x}) + \bar{\lambda}' h(\bar{x}) + \limsup_{k \rightarrow \infty} \frac{c^k}{2} \|h(x^k)\|^2 \leq f^* \quad (6)$$

因为 $\|h(x^k)\|^2 \geq 0$ 且 $c^k \rightarrow \infty$, 由此得出 $h(x^k) \rightarrow 0$ 且

$$h(\bar{x}) = 0, \quad (7)$$

Eq. (6) 式的左边将等于正无穷, 而 $f^* < \infty$ (因为约束集是非空的), 两者相矛盾.

由于 X 是闭集 $\Rightarrow \bar{x} \in X \Rightarrow$ 故 \bar{x} 是可行的.

根据 Eq (6) 和 (7) $\Rightarrow f(\bar{x}) \leq f^* \Rightarrow \bar{x}$ 是最优的.

拉格朗日乘子估计-不精确最小化

命题 1

假定增广拉格朗日函数的最小值能被精确地计算出。

特别地，当 $X = \mathbf{R}^n$ ， f 和 h 都可微时，对于如下无约束优化问题

$$\begin{array}{ll} \text{minimize} & L_{c^k}(x, \lambda^k) \\ \text{subject to} & x \in \mathbf{R}^n \end{array} \quad (8)$$

算法通常会在 x^k 满足

$$\|\nabla_x L_{c^k}(x^k, \lambda^k)\| \leq \epsilon^k$$

时终止，其中 ϵ^k 是一个非常小的量。

定理

定理 3:

- ▶ 假定 $X = \mathbb{R}^n$, f, h : 都是连续可微的.
- ▶ 令 $k = 0, 1, \dots$, x^k 满足 $\|\nabla_x L_{c^k}(x^k, \lambda^k)\| \leq \epsilon^k$
其中 $\{\lambda^k\}$ 是有界的;
- ▶ $\{\epsilon^k\}$ 和 $\{c^k\}$ 满足

$$0 < c^k < c^{k+1}, \quad \forall k, \quad c^k \rightarrow \infty$$

$$0 \leq \epsilon^k, \quad \forall k, \quad \epsilon^k \rightarrow 0$$

- ▶ 假设子序列 $\{x^k\}_K$ 收敛到向量 x^* : 且 $\nabla h(x^*)$ 的秩为 m .

\Rightarrow

- ▶ $\{\lambda^k + c^k h(x^k)\}_K \rightarrow \lambda^*$
- ▶ 其中向量 λ^* 和 x^* 满足一阶必要条件

$$\nabla f(x^*) + \nabla h(x^*) \lambda^* = 0, \quad h(x^*) = 0$$

证明

证明:

不失一般性, 我们假设整个序 $\{x^k\}$ 收敛到 x^* , 对于所有 k , 定义

$$\tilde{\lambda}^k = \lambda^k + c^k h(x^k)$$

那么有

$$\nabla_x L_{c^k}(x^k, \lambda^k) = \nabla f(x^k) + \nabla h(x^k) (\lambda^k + c^k h(x^k)) = \nabla f(x^k) + \nabla h(x^k) \tilde{\lambda}^k \quad (9)$$

因为 $\nabla h(x^*)$ 的秩为 m , 所以对充分大的 k , $\nabla h(x^k)$ 的秩也为 m . 不失一般性, 假设对于所有 k , $\nabla h(x^k)$ 的秩都为 m . 然后对 Eq. (6) 式两边同时左乘

$$\left(\nabla h(x^k)' \nabla h(x^k) \right)^{-1} \nabla h(x^k)'$$

得到

$$\tilde{\lambda}^k = \left(\nabla h(x^k)' \nabla h(x^k) \right)^{-1} \nabla h(x^k)' (\nabla_x L_{c^k}(x^k, \lambda^k) - \nabla f(x^k)) \quad (10)$$

因为假设 $\nabla_x L_{c^k}(x^k, \lambda^k) \rightarrow 0$, 所以根据 Eq. (7) 式, 有

$$\tilde{\lambda}^k \rightarrow \lambda^*$$

其中

$$\lambda^* = - \left(\nabla h(x^*)' \nabla h(x^*) \right)^{-1} \nabla h(x^*)' \nabla f(x^*)$$

再次利用 $\nabla_x L_{c^k}(x^k, \lambda^k) \rightarrow 0$ 和 Eq. (6), 得到

$$\nabla f(x^*) + \nabla h(x^*) \lambda^* = 0$$

因为序列 $\{\lambda^k\}$ 是有界的, 且 $\lambda^k + c^k h(x^k) \rightarrow \lambda^*$, 所以 $\{c^k h(x^k)\}$ 也是有界的. 因为 $c^k \rightarrow \infty$, 所以 $h(x^k) \rightarrow 0$ 故 $h(x^*) = 0$.

关于条件数较大的问题

现在考虑二次惩罚法的实际应用 p 假设第 k 次迭代中 L_{c^k} (对应的无约束最小化问题在

$$\|\nabla_x L_{c^k}(x^k, \lambda^k)\| \leq \epsilon^k$$

时结束其中 $\epsilon^k \rightarrow 0$. 那么有三种可能:

- (a) 没有满足 $\|\nabla_x L_{c^k}(x^k, \lambda^k)\| \leq \epsilon^k$ 的 x^k 存在, 该方法失效.
- 情况 (a) 经常在 $L_{c^k}(\cdot, \lambda^k)$ 无界时发生.

(b) 满足 $\|\nabla_{\mathbf{x}} L_{c_k}(\mathbf{x}_k, \lambda_k)\| \leq \epsilon_k$ 的序列 $\{\mathbf{x}_k\}$ 存在, 但该序列要么没有极限, 要么它的极限 \mathbf{x}_* 使矩阵 $\nabla h(\mathbf{x}_*)$ 具有线性相关的列 (不满足列满秩条件).

情况 (b) 通常在 $L_{c^k}(\cdot, \lambda^k)$ 有界时发生, 但原问题没有可行解。

(c) 满足 $\|\nabla_{\mathbf{x}} L_{c_k}(\mathbf{x}_k, \lambda_k)\| \leq \epsilon_k$ 的序列 $\{\mathbf{x}_k\}$ 存在, 且极限 \mathbf{x}_* 使矩阵 $\nabla h(\mathbf{x}_*)$ 的秩为 m 。根据命题, \mathbf{x}_* 和 λ_* 满足最优化问题的一阶必要条件.

情况 (c) 是最常见的, 无约束最小化算法对于每个 k 都成功找到最优解, 通常情况下 $\{\mathbf{x}_k\}$ 收敛到一个可行解; 当然, $\{\mathbf{x}_k\}$ 收敛到局部最优解 x^* 是有可能的。那么, 如果 x^* 不存在对应的拉格朗日乘子向量, 则序列 $\{\lambda^k + c^k h(x^k)\}$ 发散且没有极限。

实践表明

惩罚函数法:

- (1) 在总体上是相当可靠的
- (2) 通常收敛于原问题的局部最优解.

当它失效时，通常是由于随着 $c^k \rightarrow \infty$, 最小化如下子问题的难度越来越大

$$\begin{aligned} & \text{minimize } L_{c^k}(x, \lambda^k) \\ & \text{subject to } x \in X \end{aligned}$$

假设 $\mathbf{X} = \mathbf{R}^n$, f 和 h 都是二次可微的, 那么根据梯度方法的收敛速度, 最小化的难度取决于 Hessian 矩阵 $\nabla_{xx}^2 L_{c^k}(x^k, \lambda^k)$ 的最大特征值与最小特征值之比 (条件数), 且该比值随 c_k 增大而增大。

示例

考虑二元函数极小化问题

$$\begin{array}{ll}\text{minimize} & f(x) = \frac{1}{2} (x_1^2 + x_2^2) \\ \text{subject to} & x_1 = 1\end{array}\quad (11)$$

增广拉格朗日函数为

$$L_c(x, \lambda) = \frac{1}{2} (x_1^2 + x_2^2) + \lambda (x_1 - 1) + \frac{c}{2} (x_1 - 1)^2 \quad (12)$$

其 Hessian 矩阵为

$$\nabla_{xx}^2 L_c(x, \lambda) = I + c \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1+c & 0 \\ 0 & 1 \end{pmatrix}$$

该矩阵的特征值分别为 $1 + c$ 和 1 ，因此，Hessian 矩阵的最大特征值与最小特征值之比为 $1 + c$ 。当 c 趋于正无穷时，该比值也趋于正无穷。对于较大的 c ，其增广拉格朗日函数具有较窄水平集，可知函数是病态的。

乘子方法的主要思想

考虑问题

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && h(x) = 0 \end{aligned} \tag{13}$$

前面提到，在两种情况下可以通过极小化增广拉格朗日函数 $L_c(\cdot, \lambda)$ ，很好地近似这个问题的最优解：

- (a) 向量 λ 接近于拉格朗日乘子.
- (b) 惩罚参数 c 非常大.

除了有界性之外，没有对序列 $\{\lambda^k\}$ 进行任何假设，在对 f 和 h 最小化的假设下，即使 c_k 没有趋近无穷，通过使用更好的方案更新 λ_k ，从而缓解了增广 Lagrange 函数条件过大的困难，也能显著提高收敛速度。

在一些合理的假设下，即使 c_k 没有趋近无穷，这种方法也是可行的。

乘子法

乘子法

在二次惩罚方法中 λ^k 的更新公式为

$$\lambda^{k+1} = \lambda^k + c^k h(x^k). \quad (14)$$

如果生成的序列 $\{x^k\}$ 收敛到局部最小解 x^* , 那么 $\{\lambda^k + c^k h(x^k)\}$ 收敛到相应的拉格朗日乘子向量 λ^* .

使用上述 λ^k 更新公式的二次惩罚法称为乘子法.

示例

再次考虑问题

$$\begin{array}{ll}\text{minimize} & f(x) = \frac{1}{2} (x_1^2 + x_2^2) \\ \text{subject to} & x_1 = 1\end{array}\quad (15)$$

最优解 $x^* = (1, 0)$ 对应的拉格朗日乘子 $\lambda^* = -1$. 增广拉格朗日函数为

$$L_c(x, \lambda) = \frac{1}{2} (x_1^2 + x_2^2) + \lambda (x_1 - 1) + \frac{c}{2} (x_1 - 1)^2$$

向量 x^k 通过乘子法来最小 $L_c k(\cdot, \lambda^k)$, 并且可以由下式给出

$$x^k = \left(\frac{c^k - \lambda^k}{c^k + 1}, 0 \right)$$

使用该表达式, 乘子更新公式 (8) 可以写成

$$\lambda^{k+1} = \lambda^k + c^k \left(\frac{c^k - \lambda^k}{c^k + 1} - 1 \right) = \frac{\lambda^k}{c^k + 1} - \frac{c^k}{c^k + 1}$$

或者通过引入拉格朗日乘数 $\lambda^* = -1$,

$$\lambda^{k+1} - \lambda^* = \frac{\lambda^k - \lambda^*}{c^k + 1}$$

评述

- (1) 对于每一个非递减序列 $\{c^k\}$, $\lambda^k \rightarrow \lambda^* = -1$ 且 $x^k \rightarrow x^* = (1, 0)$ 因为 $1/(c^k + 1)$ 乘 $\lambda^k - \lambda^*$ 总是小于 1。
- (2) c^k 越大, 收敛速度越快: 事实上, 如果 $c^k \rightarrow \infty$, $\{|\lambda^k - \lambda^*|\}$ 是超线性收敛的。
- (3) 当然, 我们没有必要让 c^k 趋近无穷, 尽管这样做能提高收敛速度。

作业

利用乘子法 minimize $f(x) = x_1^2 + x_1x_2 + x_2^2$, subject to $x_1 = 1$