

# 数值优化 (Numerical Optimization) 学习系列-带约束最优化 (Constrained Optimization)

下一步 于 2015-12-27 18:53:46 发布



数值优化 专栏收录该内容

94 订阅 17 篇文章

已订阅

## 概述

带约束的最优化问题是区别于无约束最优化问题的，另外一大类最优化问题。在实际中也经常遇到，本节主要介绍了带约束最优化问题的相关理论以及最优解满足的必要条件。

1. 带约束最优化问题概述
2. 带约束最优化问题解的必要条件示例
3. 最优解满足的一阶和二阶必要条件
4. 对偶
5. 总结

## 带约束最优化问题

数学形式为

$$\begin{aligned} \min f(x) \\ \text{st. } c_i(x) = 0, i \in \mathcal{E} \\ c_i(x) \geq 0, i \in \mathcal{I} \end{aligned}$$

其中  $\mathcal{I}$  和  $\mathcal{E}$  都是指示集合分别表示不等式约束和等式约束。

0. 不同书籍上对于不等式约束可能会表示为小于等于号，其实原理都一致。

1. 可行解集合定义为  $\Omega = \{x | c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\}$

2. 几个概念，局部解、严格局部解和独立局部解。

3. 平滑的约束不代表解的可行域是平滑的；另外很多不平滑的无约束问题可以转换为带约束的问题。

4. 激活集合  $\mathcal{A}(x)$  的定义  $\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} | c_i(x) = 0\}$ ，激活集合是指满足等式约束和不等式约束中等号的解。

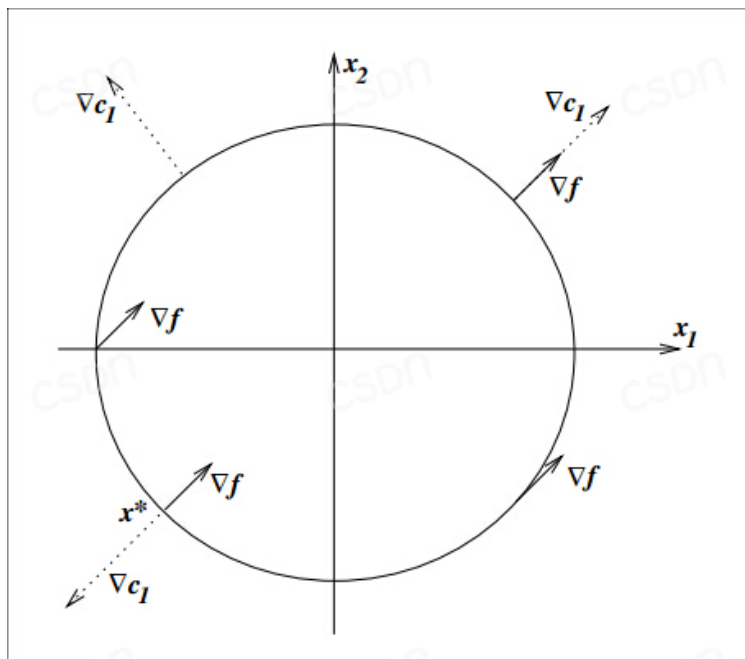
## 带约束问题解性质示例

## 等式约束

例子：

$$\min x_1 + x_2 \quad \text{st.} \quad x_1^2 + x_2^2 - 2 = 0$$

可行解集如下图所示，即在圆边上的点。



从图中可以看出，最优解的 $\nabla c_1(x^*)$ 和 $\nabla f(x^*)$ 是平行的，即 $\nabla f(x^*) = \lambda^* \nabla c_1(x^*)$ 。

下面简单证明该条件是必要的，主要思路采用反证法。

1. 分别对约束和目标函数在某可行点 $x$ 处进行一阶近似泰勒展开，结果如下
2. 约束展开形式， $c_1(x + s) = c_1(x) + \nabla c_1(x)^T s = 0$ ，由于需要满足等式约束，所以下一步搜索步长需要满足： $\nabla c_1(x)^T s = 0$
3. 目标函数展开形式， $f(x + s) = f(x) + \nabla f(x)^T s$ ，由于下一搜索方向需要满足函数值递减，所以下一步搜索步长需要满足： $\nabla f(x)^T s < 0$
4. 现在假设搜索方向归一化后表示为 $d = s/\|s\|$ ，则 $d$ 应该满足 $\nabla c_1(x)^T d = 0$ 和 $\nabla f(x)^T d < 0$
5. 采用反证法，如果 $\nabla c_1(x)$ 和 $\nabla f(x)$ 不平行，则总是可以找到某个搜索步长 $d = \left( \frac{\nabla c_1(x) \nabla c_1(x)^T}{\|\nabla c_1(x)\|^2} - \mathcal{I} \right) \nabla f(x)$ 满足需求，很容易验证 $d$ 满足该约束。因此结论得证。

下面看一下，通过引入拉格朗日函数如何表示该结论。

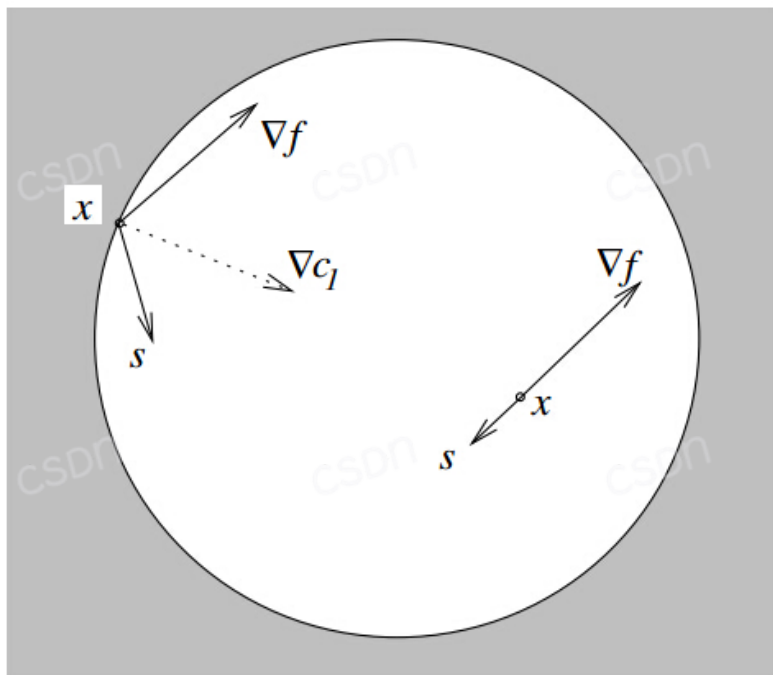
1. 该问题的拉格朗日函数为  $\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x)$
2. 对该问题求最优解有  $\nabla_x \mathcal{L} = \nabla f(x) - \lambda_1 \nabla c_1(x) = 0$
3. 拉格朗日因子  $\lambda$  不能保证该问题是最优解时的唯一性。

## 不等式约束

示例

$$\min x_1 + x_2; \quad \text{st. } 2 - x_1^2 - x_2^2 \geq 0$$

最优解图示如下，可行区域在圆内包括边缘



该问题的最优解性质和拉格朗日因子有很大关系，下面也是通过反正法指明需要满足的一些约束条件。  
根据泰勒展开公式可以知道，搜索方向需要满足一下条件：

内容来源：csdn.net

作者昵称：下一步

原文链接：[https://blog.csdn.net/fangqingan\\_java/article/details/49311679](https://blog.csdn.net/fangqingan_java/article/details/49311679)

作者主页：[https://blog.csdn.net/fangqingan\\_java](https://blog.csdn.net/fangqingan_java)

$$c_1(x+s) = c_1(x) + \nabla c_1(x)^T s \geq 0$$

$$\nabla f(x)^T s < 0$$

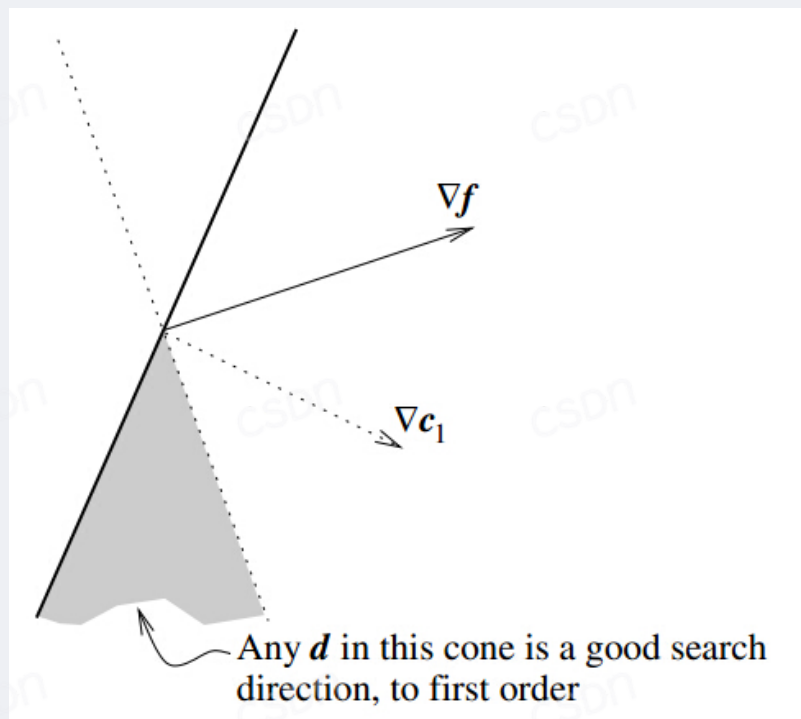
针对第一个不等式需要分情况讨论

第一种情况，当前可行解严格满足不等式约束，即  $c_1(x) > 0$ ，则有

1. 下一个搜索步长只要尽可能短就能满足  $c_1(x+s) \geq 0$ ，因此只要  $s$  满足第二个约束条件就行，即选择  $s = -\alpha \nabla f(x)$
2. 只要沿着负梯度方向总是可以找到某个下降方向，如果当前点是最优解除非  $\nabla f(x) = 0$

第二种情况，当前可行解满足等式约束，即  $c_1(x) = 0$ ，则有

1. 搜索步长应该满足  $\nabla f(x)^T s < 0$  和  $\nabla c_1(x)^T s \geq 0$
2. 根据不等式可以看到  $s$  应该和  $\nabla f(x)$  夹角大于  $180^\circ$ ，同时和  $\nabla c_1(x)$  夹角小于等  $180^\circ$ 。第一个定义了一个半开区间，第二个定义了半闭区间，可行步长是他们的交集，图示如下



3. 如果  $\nabla f(x) = \lambda \nabla c_1(x)$ ,  $\lambda \geq 0$ ，此时当前点是最优解。

内容来源：csdn.net

作者昵称：下一步

原文链接：[https://blog.csdn.net/fangqingan\\_java/article/details/49311679](https://blog.csdn.net/fangqingan_java/article/details/49311679)

作者主页：[https://blog.csdn.net/fangqingan\\_java](https://blog.csdn.net/fangqingan_java)

通过引入拉格朗日函数，综合表示上述两种情况

1. 拉格朗日函数表示为:  $\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x)$
2. 最优解形式  $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$ , and  $\lambda^* \geq 0$
3. 除此之外需要满足一个补充条件, 即  $\lambda^* c_1(x^*) = 0$
4. 补充条件解释如下, 由于要满足  $\nabla f(x) - \lambda \nabla c_1(x) = 0$ . 当  $\lambda = 0$  时, 此时有  $\nabla f(x) = 0$  对应于第一种情况. 当  $\lambda \neq 0$  时, 此时有  $\nabla f(x) = \lambda \nabla c_1(x)$  对应于第二种情况

## 其他

前面两个例子只是以一个约束进行了举例, 当有多个约束时上述必要条件同样满足。

## 一阶和二阶必要条件

### 概念

1. 在上面的例子中, 仅仅对目标函数和约束函数进行泰勒一阶展开, 即对进行了线性近似, 在此条件下
2. 约束规格 (Constraint qualification) 保证在最优解  $x^*$  附近约束集合和线性近似集合是类似的。
3. LICQ (Linear independence CQ) 表达如下  $\{\nabla c_i(x), i \in \mathcal{A}(x)\}$  是线性独立的。即激活集合的一阶梯度是线性独立的。

## 一阶条件

对于原始问题

$$\begin{aligned} \min f(x) \\ \text{st. } c_i(x) = 0, i \in \mathcal{E} \\ c_i(x) \geq 0, i \in \mathcal{I} \end{aligned}$$

对应的拉格朗日函数为

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{I} \cup \mathcal{E}} (\lambda_i c_i(x))$$

一阶条件表示如下

内容来源: csdn.net

作者昵称: 下一步

原文链接: [https://blog.csdn.net/fangqingan\\_java/article/details/49311679](https://blog.csdn.net/fangqingan_java/article/details/49311679)

作者主页: [https://blog.csdn.net/fangqingan\\_java](https://blog.csdn.net/fangqingan_java)

前提假设目标函数和约束函数都是连续可导的，并且LICQ在局部最优解 $x^*$ 满足，则 $x^*$ 需要满足如下条件，其中 $\lambda^*$ 是最优的拉格朗日因子。

1.  $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$
2.  $c_i(x^*) = 0, \quad i \in \mathcal{E}$
3.  $c_i(x^*) \geq 0, \quad i \in \mathcal{I}$
4.  $\lambda_i \geq 0, \quad i \in \mathcal{I}$
5.  $\lambda_i c_i(x^*) = 0, \quad i \in \mathcal{I} \cup \mathcal{E}$

以上条件也称之为**KKT条件**，其中最后一条也称之为补充条件。

## 二阶必要条件

前提假设目标函数和约束函数都是连续可导的，并且LICQ在局部最优解 $x^*$ 满足，则 $x^*$ 需要满足如下条件，其中 $\lambda^*$ 是最优的拉格朗日因子

1.  $w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w \geq 0, \quad w \in \mathcal{C}(x^*, \lambda^*)$
- 即需要在一定范围内保证正定，和无约束问题类似。

## 证明

略。

## 对偶 (Duality)

原始问题的对偶问题有很多很好的性质，很多问题都会转换为分析其对偶问题。同时对偶问题的解释原始解的下界。

## 拉格朗日函数

本节主要针对不等式约束的最优问题进行讨论，数学形式表达为

$$\min f(x) \quad s.t. \quad c_i(x) \geq 0, \quad i = 1 \dots m$$

对应的拉格朗日函数为

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x), \quad \lambda_i \geq 0$$

本节主要针对该拉格朗日函数和因子进行相关讨论。

## 原始问题和对偶问题

原始问题：

内容来源：csdn.net

作者昵称：下一步

原文链接：[https://blog.csdn.net/fangqingan\\_java/article/details/49311679](https://blog.csdn.net/fangqingan_java/article/details/49311679)

作者主页：[https://blog.csdn.net/fangqingan\\_java](https://blog.csdn.net/fangqingan_java)

1. 原始问题可以表达为  $\min_x \max_{\lambda} \mathcal{L}(x, \lambda)$ ，由于先对  $\lambda$  求最大值结果肯定为  $f(x)$ ，在求最小值就是目标问题。

## 对偶问题

1. 对偶函数  $q(\lambda) = \inf_x \mathcal{L}(x, \lambda)$ ，即对拉格朗日函数求下界，很多情况下对偶函数都是无下界的，因此定义对偶函数的定义域为  $\mathcal{D} = \{\lambda | q \geq -\infty\}$
2. 求解对偶函数的下界，相当于求解拉格朗日函数的最优解，是一个无约束的最优化问题，如果目标函数和约束都是凸函数则  $\mathcal{L}(x, \lambda)$  也是凸函数，此时局部最优也是全局最优解。
3. 对偶问题定义为  $\max_{\lambda} q(\lambda), s.t. \lambda \geq 0$ ，对偶问题也是一个约束最优化问题。
4. 对偶问题是求解拉格朗日因子的最优化问题，而原始问题的解  $x$  是关于  $\lambda$  的某个表达式。

## 性质

性质1：对偶函数是凹函数定义域是凸集。

证明思路就是定义。

性质2：弱对偶性，对于任意可行解  $\bar{x}$  和任意  $\bar{\lambda} \geq 0$ ，有  $q(\bar{\lambda}) \leq f(\bar{x})$

证明很简单， $q(\bar{\lambda}) = \inf_x f(x) - \bar{\lambda} c(x) \leq f(\bar{x}) - \bar{\lambda} c(\bar{x}) \geq f(\bar{x})$

即对偶问题是原始问题的下界，最大化对偶问题就会解决原始问题的最优解。其实这是一个普遍成立的条件

$$\min_x \max_y \theta(x) \geq \max_y \min_x \theta(x)$$

即最大值中的最小值也会大于等于最小值中的最大值。

性质3：强对偶性

1. 原始问题和对偶问题的差值也称之为对偶间隙。
2. 鞍点定义为  $\mathcal{L}(x^*, \lambda) \leq \mathcal{L}(x^*, \lambda^*) \leq \mathcal{L}(x, \lambda^*)$
3. 如果鞍点分别是原始问题的最优解并且对偶间隙问题0，则强对偶
4. 线性规划问题和线性约束的凸规划问题都满足强对偶，有些非凸优化问题也会满足强对偶。

性质4：根据KKT条件，该问题的KKT条件为

1.  $\nabla f(\bar{x}) - \nabla c(\bar{x}) \bar{\lambda} = 0$
2.  $c(\bar{x}) \geq 0$

内容来源：csdn.net

作者昵称：下一步

原文链接：[https://blog.csdn.net/fangqingan\\_java/article/details/49311679](https://blog.csdn.net/fangqingan_java/article/details/49311679)

作者主页：[https://blog.csdn.net/fangqingan\\_java](https://blog.csdn.net/fangqingan_java)

$$3. \bar{\lambda} \geq 0$$

$$4. \bar{\lambda} c(\bar{x}) = 0, i = 1 \dots m$$

性质5: 假设目标函数和约束函数都是可约的, 则有任意满足KKT条件的 $(\bar{x}, \bar{\lambda})$ 也是对偶问题的解。

证明

由于 $(\bar{x}, \bar{\lambda})$ 满足KKT条件, 则有

$$\mathcal{L}(x, \bar{\lambda}) \geq \mathcal{L}(\bar{x}, \bar{\lambda}) + \nabla_x \mathcal{L}(\bar{x}, \bar{\lambda})^T (x - \bar{x}) = \mathcal{L}(\bar{x}, \bar{\lambda})$$

由于 $\mathcal{L}$ 是凸函数, 根据泰勒公司展开会满足上述不等式, 根据KKT的第一个条件得到上述等式。

$$q(\bar{\lambda}) = \inf_x \mathcal{L}(x, \bar{\lambda}) = \mathcal{L}(\bar{x}, \bar{\lambda}) = f(\bar{x}) - \bar{\lambda}^T c(\bar{x}) = f(\bar{x})$$

$\inf_x \mathcal{L}(x, \bar{\lambda}) = \mathcal{L}(\bar{x}, \bar{\lambda})$ 即在已知 $\lambda = \bar{\lambda}$ 的情况下求解最优解

$f(\bar{x}) - \bar{\lambda}^T c(\bar{x}) = f(\bar{x})$ 根据KKT的补充条件得到。

由于 $q(\lambda) \leq f(\bar{x})$ 又 $q(\bar{\lambda}) = f(\bar{x})$ , 因此 $\bar{\lambda}$ 是对偶问题的解。

性质6: 如果目标函数和约束函数都是凸函数并且可导, 假设 $\bar{x}$ 是原始问题的最优解并且LICQ满足。假设 $\hat{\lambda}$ 是对偶问题的最优解并且最优 $\lambda$ 对应的 $x$ 为 $\hat{x}$ , 即 $\inf_x \mathcal{L}(x, \hat{\lambda})$ 的最优解为 $\hat{x}$ , 并且 $\mathcal{L}(x, \hat{\lambda})$ 是一个严格凸函数, 则有 $\bar{x} = \hat{x}$ 即原始问题和对偶问题最优解相同。对应于强对偶条件。

证明略。

性质7: Wolfe 对偶

$$\begin{aligned} \max_{x, \lambda} \mathcal{L}(x, \lambda) \\ s. t. \nabla_x \mathcal{L}(x, \lambda) = 0, \lambda \geq 0 \end{aligned}$$

又转换为了带有等式和不等式约束的最优化问题。

其实就是将求解 $\inf_x \mathcal{L}(x, \lambda)$ 的步骤放入约束中, 即 $\nabla_x L(x, \lambda) = 0$

## 总结

该小结需要了解一下知识点:

1. 带约束问题的一般形式
2. 带约束最优化问题解的一阶和二阶必要条件, 并且能够通过简单例子进行解释。
3. 对偶问题和原始问题关系以及相关性

内容来源: csdn.net

作者昵称: 下一步

原文链接: [https://blog.csdn.net/fangqingan\\_java/article/details/49311679](https://blog.csdn.net/fangqingan_java/article/details/49311679)

作者主页: [https://blog.csdn.net/fangqingan\\_java](https://blog.csdn.net/fangqingan_java)



