

PRACTICAL OPTIMIZATION ALGORITHMS

徐 翔

数学科学学院
浙江大学

MAY 6, 2022

CHAPTER VII: 约束优化理论

简介(INTRODUCTION)

简介

- 一般形式：

$$\min_{x \in R^n} f(x), \quad (7.1)$$

$$s.t. \quad \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{J}. \end{cases} \quad (7.2)$$

- 可行集(feasible set)

$$\Omega = \left\{ x \mid c_i(x) = 0, x \in \mathcal{E}; c_i(x) \geq 0, x \in \mathcal{J} \right\} \quad (7.3)$$

- 写成紧凑形式

$$\min_{x \in \Omega} f(x) \quad (7.4)$$

局部解和全局解

● 例1

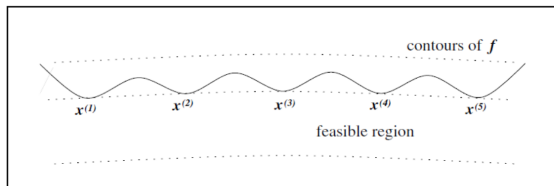
$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^n x_i^2 + 2, \quad s.t. \quad \|x\|_2^2 \geq 1$$

如果不考虑约束，显然该问题有全局最优解 $x = 0$ 。当考虑了约束之后，所有 $\|x\|_2^2 = 1$ 上的点都是全局最优解。当 $n \geq 2$ 时，这样的解有无穷多个。

● 例2

$$\min (x_2 + 100)^2 + 0.01x_1^2, \quad s.t. \quad x_2 - \cos x_1 \geq 0.$$

如果不考虑约束，显然该问题有全局最优解 $x = (0, -100)^T$ 。当考虑了约束之后，在 $x^{(k)} = (k\pi, -1)^T$, $k = \pm 1, \pm 3, \dots$ 附近都是局部最小值点。



局部解和全局解

定义

- ① 向量 x^* 被称为是(7.4)的一个局部最优解: 若 $x \in \Omega$, 且存在 x^* 的邻域 \mathcal{N} , 满足 $x \in \mathcal{N} \cap \Omega$ 时, 有 $f(x) \geq f(x^*)$ 。
- ② 向量 x^* 被称为是(7.4)的一个严格局部最优解: 若 $x \in \Omega$, 且存在 x^* 的邻域 \mathcal{N} , 满足 $x \in \mathcal{N} \cap \Omega$ 且 $x \neq x^*$ 时, 有 $f(x) > f(x^*)$ 。
- ③ 向量 x^* 被称为是(7.4)的一个孤立局部最优解: 若 $x \in \Omega$, 且存在 x^* 的邻域 \mathcal{N} , 满足 x^* 是 $\mathcal{N} \cap \Omega$ 中唯一的局部最优解。

光滑性

- 光滑性可以保证目标函数和约束条件具有合理的可预测性，可以在算法中选取好的搜索方向。
- 即便全部约束函数都是光滑的，也不表示可行域一定有光滑的边界。

$$\|x\|_1 = |x_1| + |x_2| \leq 1.$$

$$x_1 + x_2 \leq 1, x_1 - x_2 \leq 1, -x_1 + x_2 \leq 1, -x_1 - x_2 \leq 1.$$

显然，以上约束函数每一个都是光滑的，但是它们构成的可行域实际是一个菱形。

- 不光滑的无约束优化问题，有时可以转为等价的光滑约束问题。

$$\min f(x) = \max(x^2, x).$$

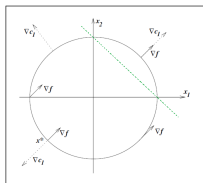
它的目标函数有两个尖点，位置在 $x = 0$ 和 $x = 1$ ，而 $x^* = 0$ 是全局最优解。等价地，可以将其转为光滑的约束优化问题：

$$\begin{aligned} \min \quad & t, \\ \text{s.t.} \quad & t \geq x, t \geq x^2. \end{aligned}$$

第一个例子

例1: 一个等值约束

$$\begin{array}{ll}\min & x_1 + x_2 \\ \text{s.t.} & x_1^2 + x_2^2 - 2 = 0.\end{array}$$



- 全局最优解 $(-1, -1)^T$
- x^* 的一个性质是 ∇f 和 ∇c_1 平行, 但方向相反, 这一点是必要条件吗?
- 我们在 c_1 上取任何非 x^* 的点, 让 x 保持在可行域内部的线性方向都有两个, 一个顺时针, 一个逆时针。这两个方向中, 总有一个是和 $-\nabla f$ 成锐角的, 也即是 f 的下降方向。直到 x^* 点, ∇c_1 和 f 平行, 这时沿 c_1 的任何移动方向, 都和 ∇f 成直角, 即 f 不会下降。
- x^* 真正满足的条件是:

$$\left\{x^* \text{ 处的可行方向} \right\} \cap \left\{f \text{ 在 } x^* \text{ 处的下降方向} \right\} = \emptyset$$

第一个例子

严格化分析

- $\forall x \in \Omega$, 保持 x 停留在 Ω 内部的可行方向 d , 满足 $c_1(x+d) = 0$.
- 由泰勒展开知道, 有 $0 = c_1(x+d) \approx c_1(x) + \nabla c_1(x)^T d = \nabla c_1(x)^T d$
- 若 x 不是 f 的局部最小值点, 则从 x 出发, 在可行域内部必有 f 的下降方向, 同样在Taylor展开下, 有

$$0 > f(x+d) - f(x) \approx \nabla f(x)^T d \rightarrow \nabla f(x)^T d < 0.$$

- 于是, 如果以上两个同时成立, 我们就可以找到一个从 x 出发的方向 d , 沿此方向: (1) $x+d$ 停留在 Ω 内部; (2) $f(x)$ 继续下降。
- 如果 x 是问题的解, 那么上述两点一定不能同时成立。这其实已经给了我们一阶必要性条件: 唯有 ∇c_1 和 ∇f 平行, 才不会出现既可行又下降的方向, 即

$$\nabla f(x) = \lambda \nabla c_1(x), \lambda \in R.$$

。

第一个例子

严格化分析

- 引入Lagrange乘子函数

$$L(x, \lambda_1) := f(x) - \lambda_1 c_1(x)$$

- 则平行条件等价于

$$\nabla_x L(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x) = 0$$

- 于是，非约束优化问题的解的**必要条件**可以表述为：

$$\text{存在 } \lambda_1^* \in R, \text{ 使得 } \nabla_x L(x^*, \lambda_1^*) = 0$$

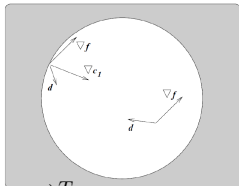
这里 λ_1 称为约束条件 $c_1(x) = 0$ 对应的**Lagrange乘子**(Lagrange Multiplier)。

- 显然，这个条件是**非充分的**。比如该例子还有一个可行点 $(1, 1)^T$ 也满足这个条件，但是它是极大值点。而且 λ_1^* 的符号并不能说明 x^* 的极值性质，因为 $c_1(x) = x_1^2 + x_2^2 - 2 = 0$ 或者 $c_1(x) = 2 - x_1^2 - x_2^2 = 0$ 。从而 $x^* = (-1, -1)^T$ 处的 λ_1^* 值从 $-\frac{1}{2}$ 变为 $\frac{1}{2}$ 。

第二个例子

例2：一个不等值约束的优化

$$\begin{array}{ll}\min & x_1 + x_2 \\ \text{s.t.} & 2 - x_1^2 - x_2^2 \geq 0\end{array}$$



- 可行域变为圆的内部和边界，最优解还是 $(-1, -1)^T$ 。
- $\forall x \in \Omega$ ，现在确保 $x + d$ 留在 Ω 内的搜索方向在泰勒展开中：

$$0 \leq c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d$$

- 现在无法消去第一项 $c_1(x)$ ，完整的一阶可行方向条件是

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

- 使得 f 下降的方向仍然是

$$\nabla f(x)^T d < 0.$$

第二个例子

例2：分两种情况讨论

- x 在 Ω 内部，即 $c_1(x) > 0$ 。此时任何 d 都是可行方向，如果从 x^* 出发的方向都**不能同时满足两个条件**（即交集是空集），那么只剩下唯一的可能

$$\nabla f(x^*) = 0$$

这是无约束问题下对应的必要条件。

- x 在 Ω 边界，也就是 $c_1(x) = 0$ 。此时又回到第一例子的情形，如果 x 出发还存在优化的可能方向 d ，必须满足

$$\nabla f(x)^T d < 0, \nabla c_1(x)^T d \geq 0.$$

此时第一不等式限定了半个平面，第二个不等式限定了另外半个平面（包含边界）。要使得两个半平面相交为**空集**，唯有 ∇f 和 $\nabla c_1(x)$ 是同一个方向(比平行更严格)，即

$$\nabla f(x) = \lambda_1 \nabla c_1(x), \lambda_1 \geq 0$$

第二个例子

总结以上两种情况，可以统一地表示为：

- 当 x^* 没有可行的下降方向时，存在 λ_1^* ，使得

$$\nabla_x L(x^*, \lambda_1^*) = 0, \lambda_1^* \geq 0$$

并且

$$\lambda_1^* c_1(x) = 0$$

- 这个条件被称为**互补性条件(complementarity condition)**。它要求 λ_1^* 和 $c_1(x^*)$ 至少有一个为零。前者对应情况 x^* 落在 Ω 内，后者对应情况 x^* 落在 Ω 边界上。
- 如果对于一个可行点 x ，其对应的不等值约束 $c_i(x) \neq 0$ ，那么实际上这个约束对 x 是无效的。只有正好处在边界的约束，对一个点是否是最优值点的判定才会有意义。

活跃集

定义: 活跃集(active set, 也称为有效集、积极集)

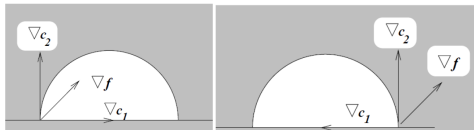
下标集合 $\mathcal{A}(x)$ 称为点 x 的活跃集(也被称为积极集), 是有全部等值约束的下标、在 x 点满足 $c_i(x) = 0$ 的不等值约束的下标组成。即

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} | c_i(x) = 0\}.$$

第三个例子

例3：两个不等式约束

$$\begin{aligned} \min \quad & x_1 + x_2 \\ \text{s.t.} \quad & 2 - x_1^2 - x_2^2 \geq 0 \\ & x_2 \geq 0. \end{aligned}$$



- 显然最优解 $x^* = (-\sqrt{2}, 0)^T$ 。
- 两个约束都是活跃的。
- 如果它不是一个最优值点（意味着从它出发还有进一步可以下降的方向），那么必然有 d 方向满足

$$\nabla c_i(x)^T d \geq 0, \quad \nabla f(x)^T d < 0.$$

- 显然最优解 $x^* = (-\sqrt{2}, 0)^T$ 是不存在这个方向的。因为满足条件的方向必须在 $\nabla c_1(x^*)$ 和 $\nabla c_2(x^*)$ 的夹角内，但是在此范围内的所有方向满足 $\nabla f(x^*)^T d > 0$ ，都是严格上升方向。

第三个例子

- 定义Lagrange乘子函数

$$L(x, \lambda) = f(x) - \lambda_1 c_1(x) - \lambda_2 c_2(x)$$

这里 $\lambda = (\lambda_1, \lambda_2)^T$ 是Lagrange乘子。

- $\nabla f(x)$ 和 $c_i(x)$ 的关系可以表示为：存在 $\lambda^* \geq 0$ ，使得

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

并且满足互补性条件

$$\lambda_1 c_1(x^*) = 0, \quad \lambda_2^* c_2(x^*) = 0.$$

- 互补性条件:

一个约束要么是活跃的($c_i(x) = 0$)，因此会在 $\nabla_x L(x, \lambda) = 0$ 中参与；要么是不活跃的，此时它在 $\nabla_x L(x, \lambda) = 0$ 中实际上是不存在的。

约束条件“资格”(Constration Qualification)

约束条件“资格”(Constration Qualification)

- 在前面的分析中，我们在考虑是否存在可行的下降方向时，是用一阶泰勒展开分析的。
- 这需要假设一阶线性展开能够捕捉可行域的几何性质。如果在 x 附近，一阶泰勒展开式与可行域相差巨大，那么我们不可能期望一阶线性逼近能够很好地近似原问题。
- 因此我们需要对活跃的约束条件 c_i 在 x 处的性质，保证线性逼近能够很好地近似真实的可行域。
- 约束条件“资格”：是一些假设条件，使得可行集 Ω 和一阶线性逼近在 x^* 的领域内很接近。

切锥(Tangent Cone)

定义：切锥

- 给定一个可行点 $x \in \Omega$ ，我们称 $\{z_k\}$ 是一列逼近 x 的可行序列，如果对任意的充分大的 k ，满足 $z_k \in \Omega$ 且 $z_k \rightarrow x$ 。
- “切向”是可行序列的极限方向。
- 严格的数学定义：设 $x^* \in \Omega$ ， $d \neq 0$ 且 $d \in R^n$ ，如果存在 $\delta > 0$ 使得

$$x^* + t d \in \Omega, \forall t \in [0, \delta]$$

则称 d 是 Ω 在 x^* 处的线性化方向。

- 在 x^* 处所有的切向组成的集合称为切锥(Tangent Cone)，记为 $T_{\Omega}(x^*)$ 。
- 从切锥的定义可以看出，这个定义跟 Ω 的代数形式无关，只能 Ω 的几何性质有关。

线性化可行方向

定义：线性化可行方向

给定一个可行点 $x \in \Omega$ 以及活跃集 \mathcal{A} ，线性可行方向集

$$\mathcal{F}(x) = \{d \mid d^T \nabla c_i(x) = 0, \forall i \in \mathcal{E}; d^T \nabla c_i(x) \geq 0, \forall i \in \mathcal{A}(x) \cap \mathcal{J}\}.$$

显然，线性化可行方向集仅依赖于约束条件 c_i , $i \in \mathcal{E} \cup \mathcal{J}$ 。

约束条件限制

- 约束条件限制是一些假设，能够保证“切锥” $T_{\Omega}(x)$ 和“线性可行方向集” $\mathcal{F}(x)$ 是相似的。
- 这些限制条件，可以保证 $\mathcal{F}(x)$ (由 $c_i(x)$ 一阶线性化构造)能够很好地捕捉到 Ω 在 x 处的本质几何性质(essential geometric features)。
- 绝大多数约束条件限制都可以保证 $\mathcal{F}(x) = T_{\Omega}(x)$ 。

约束条件限制

两种限制

- **线性无关约束限制(LICQ)**: 给定一个可行点 $x \in \Omega$ 以及活跃集 $\mathcal{A}(x)$, 如果活跃的约束条件的梯度 $\nabla c_i(x)$, $i \in \mathcal{A}(x)$ 是线性无关的。
- **Mangasarian-Fromovitz约束限制(MFCQ)**: 存在一个向量 $w \in \mathbb{R}^n$ 使得

$$\begin{aligned}\nabla c_i(x^*)^T w &= 0, \quad \forall i \in \mathcal{E}, \\ \nabla c_i(x^*)^T w &> 0, \quad \forall i \in \mathcal{A}(x^*) \cap \mathcal{I},\end{aligned}$$

并且所有的等值约束条件的梯度 $\{\nabla c_i(x^*), i \in \mathcal{E}\}$ 是线性无关的。

- **特殊情况**: 如果所有的活跃约束 $c_i(\cdot)$, $i \in \mathcal{A}(x^*)$ 都是线性函数, 那么显然有 $\mathcal{F}(x^*) = T_{\Omega}(x^*)$ 。

一阶最优性条件(First-Order Necessary Conditions)

拉格朗日函数

定义 Lagrangian Function

拉格朗日函数

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

这里 λ_i 称为 $c_i(x)$ 的拉格朗日乘子(Lagrange Multiplier)。

一阶最优性条件(First-Order Necessary Conditions)

定理 (KKT条件, Karush-Kuhn-Tucker)

假设 x^* 是一个局部最优解并且在 x^* 处满足LICQ条件。则存在一个拉格朗日乘子向量 λ^* ，每个分量是 λ_i ， $i \in \mathcal{E} \cup \mathcal{J}$ ，使得如下条件满足

$$\begin{aligned}\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, \\ c_i(x^*) &= 0, & \forall i \in \mathcal{E}, \\ c_i(x^*) &\geq 0, & \forall i \in \mathcal{J}, \\ \lambda_i^* &\geq 0, & \forall i \in \mathcal{J}, \\ \lambda_i^* c_i(x^*) &= 0, & \forall i \in \mathcal{E} \cup \mathcal{J}.\end{aligned}$$

Remark: 通常来说，对于给定的优化问题及局部最优解 x^* ，可能存在很多个 λ^* 满足上述KKT条件。但是当LICQ满足时，最优的 λ^* 是唯一的。

互补性条件(Complementary Condition)

- KKT条件中最后一个条件 $\lambda_i^* c_i(x^*) = 0$ 称为互补性条件。这表明要么约束条件是活跃的，要么 $\lambda_i^* = 0$ ，或者两者同时成立。
- 如果约束条件不是活跃的，那么对应的Lagrange乘子为0，也就是对应的Lagrangian函数可以简化为：

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*).$$

- **严格互补条件** 给定一个局部最优解 x^* 和相应的 λ^* 满足KKT条件，如果对所有的不等式约束里 λ_i^* 和 $c_i(x^*)$ 中只有一个为0，即对于 $i \in \mathcal{I} \cap \mathcal{A}(x^*)$ 有 $\lambda_i^* > 0$ ，这样的情况，我们称为满足严格互补条件。
- 显然，如果满足严格互补条件，可以使得算法比较容易确定活跃集 $\mathcal{A}(x^*)$ ，而且算法收敛比较快。

拉格朗日乘子 λ_i 的意义

KKT条件中第一个条件: $\nabla f(x^*) = \sum \lambda_i^* \nabla c_i(x^*)$, $i \in \mathcal{A}(x^*)$

可以看出 λ_i^* 的实际意义是 $\nabla c_i(x^*)$ 在 $\nabla f(x^*)$ 中的权重。

- λ_i^* 越大, 说明第 i 个约束在 f 无法下降中起的作用也越大, 对算法也越敏感。极端情况 $\lambda_i^* = 0$, 说明约束没起作用。
- 如果对 $c_i(x) \geq 0$ 施加小扰动, 变为 $c_i(x) \geq -\varepsilon \|\nabla c_i(x^*)\|$, 那么活跃的约束变为

$$\frac{c_i(x^*(\varepsilon) - x^*)}{x^*(\varepsilon) - x^*} = \nabla c_i(x^*), \text{ 即: } c_i(x^* + \varepsilon) - c_i(x^*) \approx (x^*(\varepsilon) - x^*)^T \nabla c_i(x^*)$$

这里 $x^*(\varepsilon)$ 对应扰动后的新解。

- 对于 $j \neq i$, 扰动后 c_j 不变, 有

$$0 = c_j(x^*(\varepsilon)) - c_j(x^*) \approx (x^*(\varepsilon) - x^*)^T \nabla c_j(x^*). \text{ 于是对 } f(x^*(\varepsilon)), \text{ 有}$$

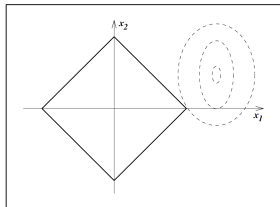
$$f(x^*(\varepsilon)) - f(x^*) \approx (x^*(\varepsilon) - x^*)^T \nabla f(x^*) = \sum \lambda_j^* (x^*(\varepsilon) - x^*)^T \nabla c_j(x^*) \approx -\varepsilon \lambda_i^* \|\nabla c_i(x^*)\|$$

进而 $\frac{df(x^*(\varepsilon))}{d\varepsilon} = -\lambda_i^* \|\nabla c_i(x^*)\|$ 。因此, $\lambda_i^* \|\nabla c_i(x^*)\|$ 是第 i 个约束对解目标值的敏感影响因子。

一个例子

例4：考虑如下的优化问题

$$\begin{aligned}
 \min \quad & (x_1 - \frac{3}{2})^2 + (x_2 - \frac{1}{2})^4 \\
 \text{s.t.} \quad & 1 - x_1 - x_2 \geq 0 \\
 & 1 - x_1 + x_2 \geq 0 \\
 & 1 + x_1 - x_2 \geq 0 \\
 & 1 + x_1 + x_2 \geq 0.
 \end{aligned}$$



- 最优解是 $(1, 0)^T$ 。第一个和第二个约束条件是活跃的，记为 c_1 , c_2 。
-

$$\begin{aligned}
 \nabla f(x^*) &= \begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\
 \lambda^* &= (\frac{3}{4}, \frac{1}{4}, 0, 0)^T
 \end{aligned}$$

证明一阶最优性条件(KKT)

证明主要分为三步。第一步是“切锥”和“一阶可行方向集”之间的关系，第二步是基本必要条件，第三步是Farkas引理，如何判断一个方向落在一个锥里。

引理1

假定 x^* 是一个可行点，则有

- ① $T_{\Omega}(x^*) \subset \mathcal{F}(x^*)$;
- ② 若在 x^* 点有LICQ成立，则 $T_{\Omega}(x^*) = \mathcal{F}(x^*)$

引理2

假定 x^* 是一个局部最优解，则有

$$\nabla f(x^*)^T d \geq 0, \text{ 对于任意的 } d \in T_{\Omega}(x^*)$$

即：切锥里所有的方向都不是下降方向。

引理1的证明

先证明第一个结论 $T_{\Omega}(x^*) \subset \mathcal{F}(x^*)$

- 假设 c_i 是活跃的约束条件, $i = 1, \dots, m$ 。对于任意的 $d \in T_{\Omega}(x^*)$, 令序列 $\{z_k\}$ 和 $\{t_k > 0\}$ 满足

$$\lim_{k \rightarrow \infty} \frac{z_k - x^*}{t_k} = d, \quad \text{即 } z_k = x^* + t_k d + o(t_k).$$

- 对于等式约束 $i \in \mathcal{E}$, 有

$$0 = \frac{c_i(z_k)}{t_k} = \frac{c_i(x^*) + t_k \nabla c_i(x^*)^T d + (t_k)}{t_k} = \nabla c_i(x^*)^T d + \frac{o(t_k)}{t_k}$$

两边同取极限可以得到 $\nabla c_i(x^*)^T d = 0$.

- 对于不等式约束 $i \in \mathcal{I} \cap \mathcal{A}(x^*)$, 有

$$0 \leq \frac{c_i(z_k)}{t_k} = \frac{c_i(x^*) + t_k \nabla c_i(x^*)^T d + (t_k)}{t_k} = \nabla c_i(x^*)^T d + \frac{o(t_k)}{t_k}$$

两边同取极限可以得到 $\nabla c_i(x^*)^T d \geq 0$.

综上所述, $d \in \mathcal{F}(x^*)$ 即: $T_{\Omega}(x^*) \subset \mathcal{F}(x^*)$ 。

引理1的证明

再证明引理1的第二个结论：如果满足LICQ，则 $T_{\Omega}(x^*) = \mathcal{F}(x^*)$

- 所有的活跃约束的梯度向量 $\nabla c_i(x^*)$ 可以组成一个矩阵，记为 $A(x^*) := [\nabla c_i(x^*)]^T, i \in \mathcal{A}(x^*)$ 。这最多是一个 $m \times n$ 的矩阵。
- 由于满足LICQ，则 $A(x^*)$ 行满秩。可以找到 $A(x^*)$ 的核空间的一组基函数 $Z \in \mathbb{R}^{n \times (n-m)}$ ，即 $A(x^*)Z = 0$ 。
- 对于 $\mathcal{F}(x^*)$ 中的任意方向 d ，定义如下的映射 $R: \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$

$$R(z, t) = \begin{bmatrix} c(z) - tA(x^*)d \\ Z^T(z - x^* - td) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- 这里断言，当 $t_k > 0$ 趋于0，上述方程解存在记为 z_k ，则 z_k 和 t_k 满足渐进可行序列的要求，即 $\lim_{k \rightarrow \infty} \frac{z_k - x^*}{t_k} = d$ 。(下面分两小步说明)

引理1的证明

- 第一小步, 即 z_k 是可行的。当 $t = 0$, $z = x^*$ 时, R 对 z 的Jacobi矩阵为 $[A(x^*)^T; Z]^T$ 是非奇异的, 所以根据隐函数存在定理, 对于充分小的 t_k , 方程组存在唯一的解 z_k 。这些解 z_k 满足 $c_i(z_k) = t_k A(x^*)d$ 。
 当 $i \in \mathcal{E}$ 时, 即 $c_i(z_k) = t_k \nabla c_i(x^*)^T d = 0$ 。
 当 $i \in \mathcal{A}(x^*) \cap \mathcal{J}$ 时, $c_i(z_k) = t_k \nabla c_i(x^*)^T d \geq 0$ 。
- 第二小步: 对于可行的 z_k , 满足 $\lim_{k \rightarrow \infty} \frac{z_k - x^*}{t_k} = d$ 。对于 $R(z_k, t_k)$ 应用泰勒展开可以得到:

$$\begin{aligned}
 0 = R(z_k, t_k) &= \begin{bmatrix} c(z_k) - t_k A(x^*)d \\ z^T(z_k - x^* - t_k d) \end{bmatrix} \\
 &= \begin{bmatrix} c(x^*) + \nabla c(x^*)^T(z_k - x^*) + o(\|z_k - x^*\|) - t_k A(x^*)d \\ z^T(z_k - x^* - t_k d) \end{bmatrix} \\
 &= \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix} (z_k - x^* - t_k d) + o(\|z_k - x^*\|)
 \end{aligned}$$

两边同除以 t_k , 并求极限可得:

$$\lim_{k \rightarrow \infty} \frac{z_k - x^*}{t_k} = d$$

引理2的证明

引理2

假定 x^* 是一个局部最优解，则有

$$\nabla f(x^*)^T d \geq 0, \text{ 对于任意的 } d \in T_{\Omega}(x^*)$$

证明：（使用反证法）。

- 假设存在 $d \in T_{\Omega}(x^*)$ 满足 $\nabla f(x^*)^T d < 0$ ，则可以构造序列 $t_k \rightarrow 0^+$ ， $z_k - x^* = t_k d + o(t_k)$ ，满足

$$f(z_k) - f(x^*) = \nabla f(x^*)^T (z_k - x^*) + o(\|z_k - x^*\|) = t_k \nabla f(x^*)^T d + o(t_k)$$

如果 $\nabla f(x^*)^T d < 0$ ，那么 $f(z_k) - f(x^*) < 0$ ，这与 x^* 是局部最小值点是矛盾的。

第三个引理：Farkas引理

我们先定义一般意义下的几何锥

$$K = \{By + Cw \mid y \geq 0\}$$

其中 B 和 C 分别是 $n \times m$ 和 $n \times p$ 阶矩阵， $y \in \mathbb{R}^m$ 和 $w \in \mathbb{R}^p$ 是对应维数的向量。**(如何判断一个向量不在锥 K 里?)**

Farkas引理

如上定义锥 K 。对于任意的一个向量 $g \in \mathbb{R}^n$ ，要么 $g \in K$ ，否则必定可以存在一个向量 $d \in \mathbb{R}^n$ 满足

$$g^T d < 0, \quad B^T d \geq 0, \quad C^T d = 0.$$

我们首先说明两种情况不能同时成立(反证法)。

如果能同时成立，那么 $g \in K$ ，即存在 $y \geq 0$ 和 w 满足 $g = By + Cw$ 。同时又存在 $d \in \mathbb{R}^n$ ，满足 $g^T d < 0$ ， $B^T d \geq 0$ ， $C^T d = 0$ ，则容易计算 $d^T g$ 得到如下的矛盾：

$$0 > d^T g = d^T (By + Cw) = (B^T d)^T y + (C^T d)^T w \geq 0$$

下面接着**构造性**地证明如果 $g \notin K$ ，必须要存在 d 满足 $g^T d < 0$ ， $B^T d \geq 0$ ， $C^T d = 0$ 。

第三个引理：Farkas引理

下面接着构造性地证明如果 $g \notin K$ ，必须要存在 d 满足 $g^T d < 0$, $B^T d \geq 0$, $C^T d = 0$ 。

- 令 $\hat{s} = \arg \min_{s \in K} \|s - g\|_2^2$ 是 g 在 K 中的最佳逼近元，则 \hat{s} 和 $\hat{s} - g$ 是垂直的，即 $\hat{s}^T(\hat{s} - g) = 0$ 。
- 构造 $d = \hat{s} - g$ ，验证引理中的结论。
 - $g^T d = (\hat{s} - d)^T d = \hat{s}^T d - d^T d = \hat{s}(\hat{s} - g) - d^T d = -\|d\|^2 < 0$
 - 由于 K 是一个凸集，假定 $s \in K$, $\theta \in [0, 1]$ ，则 $\hat{s} + \theta(s - \hat{s})$ 是凸组合也属于 K ，因此 g 与该向量的距离肯定比 g 与 \hat{s} 的距离要大。即

$$\|(\hat{s} + \theta(s - \hat{s})) - g\|_2^2 \geq \|\hat{s} - g\|_2^2$$

由此可以推出 $s^T(\hat{s} - g) \geq 0$ ，即 $s^T d \geq 0$ 于是 $d^T(B^T y + C^T w) \geq 0$, $\forall y \geq 0, w \in \mathbb{R}^n$. 取 $y = 0$ ，有 $(C^T d)^T w \geq 0$ ，由 w 的任意性，只有 $C^T d = 0$ 。

(3) 再令 $w = 0$ ，则可以得到 $(B^T d)^T y \geq 0, \forall y \geq 0$ ，可以推出 $B^T d \geq 0$ 。

证明KKT条件

我们定义如下的锥 N ,

$$K = \left\{ \sum_{i \in \mathcal{A}(x^*) \cap \mathcal{J}, \lambda_i \geq 0} \lambda_i \nabla c_i(x^*) + \sum_{i \in \mathcal{E}} \lambda_i \nabla c_i(x^*) \right\}$$

$$g = \nabla f(x^*)$$

显然, 对于 g , 要么 $g \in K$, 即

$$\nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*) = A(x^*)^T \lambda^*, \text{ 其中 } \lambda_i \geq 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{J}$$

要么 $g \notin K$, 即存在某个方向满足

$$\nabla f(x^*)^T d < 0, \quad \nabla c_i^T d \geq 0, i \in \mathcal{A}(x^*) \cap \mathcal{J}, \quad \nabla c_i^T d = 0, i \in \mathcal{E}.$$

也就是 $d \in \mathcal{F}(x^*)$, 而且 d 是 f 的上升方向。

最后, 根据引理2, 我们知道当 x^* 是局部最优解时, 第二种情况是不能出现的, 所以只能出现第一种情况, 即证明了KKT条件。

二阶最优性条件(Second Order Optimality Conditions)

二阶条件

- KKT条件中仅考虑了 ∇c_i 和 ∇f 的性质。
- 如果在 x^* 处的切锥里存在这样的方向 $w \in \mathcal{F}(x^*)$, 满足 $w^T \nabla f(x^*) = 0$, 如何判断 x^* 是不是一个局部最优解? 此时要判断 $f(x)$ 沿着 w 到底是不是可以继续下降, 要看 f 的二阶导数(tiebreaking)!
- 定义一个关键锥(critical cone)

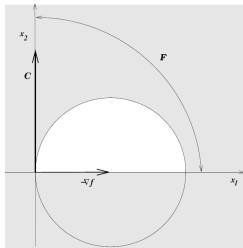
$$\mathcal{C}(x^*, \lambda^*) = \left\{ w \in \mathcal{F}(x^*) \mid \nabla c_i^T w = 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{J}, \lambda_i > 0 \right\}.$$

- 注: 为什么 $\mathcal{C}(x^*, \lambda^*)$ 中只需要考虑 $\lambda_i > 0$ 的部分? 这是因为当考虑 $w^T \nabla f(x^*) = w^T \sum_{i \in \mathcal{A}(x^*)} w^T \lambda_i \nabla c_i(x)$ 等于0时, 有很多部分已经自动为0, 比如
 - (1) 当 $\forall i \in \mathcal{E}$ 时, $\nabla c_i(x^*)^T w = 0$.
 - (2) 当 $\forall i \in \mathcal{A}(x^*) \cap \mathcal{J}$ 而且 $\lambda_i = 0$ 时, $\nabla c_i(x^*)^T w$ 可以 ≥ 0 .
 - (3) 当 $\forall i \in \mathcal{A}(x^*) \cap \mathcal{J}$ 而且 $\lambda_i > 0$ 时, $\nabla c_i(x^*)^T w$ 必须 $= 0$.
 也就是说 $w^T \nabla f(x^*) = 0$, 不确定的方向就是关键锥里的方向。

二阶条件

例5：考虑如下的优化问题

$$\begin{aligned} \min \quad & x_1 \\ \text{s.t.} \quad & x_2 \geq 0 \\ & 1 - (x_1 - 1)^2 - x_2^2 \geq 0 \end{aligned}$$



- 全局最优解是 $x^* = (0, 0)^T$, $\mathcal{A}(x^*) = \{1, 2\}$, $\lambda^* = (0, 0.5)^T$ 。
- 注意到，在 x^* 点处，活跃集中的梯度分别是 $\nabla c_1(x^*) = (0, 1)^T$, $\nabla c_2(x^*) = (2, 0)^T$ ，满足LICQ条件。
- $\mathcal{F}(x^*) = \{d | d \geq 0\}$ 。注意到在 \mathcal{F} 中，都是 f 非减的方向，这与 x^* 是局部最优解并不矛盾。但是要判定 x^* 就是最优解，还不够。
- 因为在关键锥内的方向上，也就是 $\mathcal{C}(x^*, \lambda^*) = \{(0, w_2) | w_2 \geq 0\}$ 上， $\nabla f(x^*) = 0$ ， f 到底是增还是减，不好判断。

二阶必要性条件

定理(二阶必要性条件)

假设 x^* 是局部最优解，且LICQ成立， λ^* 是对应的拉格朗日乘子，则

$$w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w \geq 0, \quad \forall w \in \mathcal{C}(x^*, \lambda^*).$$

证明：令 $z_k = x^* + t_k w + o(t_k^2)$, $w \in \mathcal{C}(x^*, \lambda^*)$ ，主要使用泰勒展开到第二阶

$$\begin{aligned} \mathcal{L}(z_k, \lambda^*) &= \mathcal{L}(x^*, \lambda^*) + (z_k - x^*)^T \nabla_x \mathcal{L}(x^*, \lambda^*) \\ &\quad + \frac{1}{2} (z_k - x^*)^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) (z_k - x^*) + o(\|z_k - x^*\|^2) \\ &= f(x^*) + \frac{1}{2} t_k^2 w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w + o(t_k^2) \end{aligned}$$

如果 $w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w < 0$ ，则 $f(z_k) < f(x^*)$ 产生矛盾。

二阶充分性条件

定理(二阶充分性条件)

假设对于可行点 x^* 处存在 λ^* 使得KKT条件成立，并且对于

$$w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w > 0, \quad \forall w \in \mathcal{C}(x^*, \lambda^*), w \neq 0.$$

则 x^* 是严格的局部最优解。证明：略。

二阶充分性条件

$$\begin{aligned} \min f(x) &= x_1 + x_2 \\ \text{s.t.} \quad c_1(x) &= 2 - x_1^2 - 2x_2^2 \geq 0, \end{aligned}$$

- Lagrange函数 $\mathcal{L}(x_1, x_2; \lambda_1) = (x_1 + x_2) - \lambda_1(2 - x_1^2 - x_2^2)$
- 由一阶最优性条件得到

$$\nabla_x \mathcal{L} = \begin{bmatrix} 1 + 2\lambda_1 x_1 \\ 1 + 2\lambda_1 x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, x_1^2 + x_2^2 = 2, \lambda_1 > 0$$

得到 $x^* = (-1, -1)^T$, $\lambda_1^* = \frac{1}{2}$ 。

- 验证是否满足充分性：

$$\nabla_{xx}^2 \mathcal{L} = \begin{bmatrix} 2\lambda_1^* & 0 \\ 0 & 2\lambda_1^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

于是对于任意的 $w \neq 0$ 都满足 $w^T \nabla_{xx}^2 \mathcal{L} w = w^T w > 0$ 。
即 x^* 一定是一个局部最优解。

二阶充分性条件

$$\begin{aligned} \min f(x) &= -0.1(x_1 - 4)^2 + x_2^2 \\ \text{s.t. } c_1(x) &= x_1^2 + x_2^2 - 1 \geq 0 \end{aligned}$$

- 首先可行域是单位圆的外部，目标函数值是可以趋于无穷大 $(\infty, 0)$ ，因此没有全局最优解，但是在可行域的边界上还是可以去到局部最小值。
- 定义 $\mathcal{L}(x, \lambda_1) = -0.1(x_1 - 4)^2 + x_2^2 - \lambda_1(x_1^2 + x_2^2 - 1)$ 使用KKT条件，

$$\nabla f - \lambda_1 \nabla c_1 = \begin{bmatrix} -0.2(x_1 - 4) - 2\lambda_1 x_1 \\ 2x_2 - 2\lambda_1 x_2 \end{bmatrix} = 0, \quad \lambda_1(x_1^2 + x_2^2 - 1) = 0$$

- 求解得到 $x^* = (1, 0)^T$, $\lambda_1^* = 0.3$ 。 $\mathcal{A}(x^*) = \{1\}$ 。
- 检验二阶最优性充分条件

$$\nabla c_1(x^*) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \mathcal{C}(x^*, \lambda^*) = \{(0, w_2)^T \mid w_2 \in \mathbb{R}\},$$

$$w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^2) w = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \begin{bmatrix} -0.8 & 0 \\ 0 & 1.4 \end{bmatrix} \begin{bmatrix} 0 \\ w_2 \end{bmatrix} = 1.4w_2^2 > 0.$$

二阶条件和投影Hessians

- 二阶条件(必要性、充分性)的验证通常比较复杂, 为此人们提出了一种稍微弱化但比较方便的形式。
- 当满足KKT条件的 λ^* 是唯一的(比如LICQ), 并且严格互补性条件成立, 此时关键锥的定义可以简化为:

$$\mathcal{C}(x^*, \lambda^*) = \text{Null}[\nabla c_i(x^*)^T]_{i \in \mathcal{A}(x^*)} = \text{Null}A(x^*) = \{Zu \mid u \in \mathbb{R}^{|\mathcal{A}(x^*)|}\}$$

这里 Z 表示 $A(x^*)$ 的零空间的基组成的矩阵, $|\mathcal{A}(x^*)|$ 表示活跃约束的个数。

- 可以判定 $Z^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) Z$ 的半正定性或正定性。
- 如何得到 Z ? 对 $A(x^*)$ 进行QR分解即可:

$$A(x^*)^T = [Q_1, Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} = Q_1 R$$

如果 R 非奇异, 则 $Z = Q_2$ 。如果 R 奇异(LICQ不成立)时, 可以对QR分解做适当的列交换来确定 Z 。

其他形式的约束限制

约束限制，**本质上**是对可行区域 Ω 而言，其线性化代数表示，能否准确地抓住 x^* 邻域内的几何形状。所以当活跃的约束都是线性函数即 $c_i(x) = a_i^T x + b_i$ ，显然是满足条件的。

引理：假设 $x^* \in \Omega$ 处的活跃约束条件 $c_i(x^*)$, $i \in \mathcal{A}(x^*)$ 都是线性函数，则有 $\mathcal{F}(x^*) = T_\Omega(x^*)$ 。

证明：首先 $T_\Omega \subset \mathcal{F}(x^*)$ 。我们只需要证明 $\mathcal{F}(x^*) \subset T_\Omega(x^*)$ 。即要证明 $\forall d \in \mathcal{F}(x^*)$, $d \in T_\Omega(x^*)$ ，也就是要证明对于 d ，存在一个序列 $z_k \in \Omega$ ，满足 $\lim_{t_k \rightarrow 0^+} \frac{z_k - x^*}{t_k} = d$ 。

$$\mathcal{F}(x^*) = \left\{ d \mid \begin{array}{l} a_i^T d = 0, \quad \forall i \in \mathcal{E}, \\ a_i^T d \geq 0, \quad \forall i \in \mathcal{A}(x^*) \cap \mathcal{J} \end{array} \right\}.$$

- 首先 $i \in \mathcal{J} \setminus \mathcal{A}(x^*)$ (即不活跃)，则存在 \bar{t} ，使得 $x^* + td$ 也不活跃，即 $c_i(x^* + tw) > 0$, $t \in [0, \bar{t}]$ 。
- 构造序列 $z_k = x^* + (\bar{t}/k)d$ ，**可以分三种约束**分别验证 $z_k \in \Omega$ 。满足 $d \in T_\Omega(x^*)$

其他形式的约束限制

定义Mangasarian-Fromovitz约束限制(MFCQ):

如果存在 w 使得

$$\begin{aligned}\nabla c_i(x^*)w &= 0, & \forall i \in \mathcal{E} \\ \nabla c_i(x^*)^T w &> 0, & \forall i \in \mathcal{A}(x^*) \cap \mathcal{J}.\end{aligned}$$

并且等式约束的梯度 $\nabla c_i(x^*)$, $i \in \mathcal{E}$ 都是线性无关的。

- 上面的不等式限制是严格的。
- MFCQ比LICQ条件弱。
- 不管是MFCQ还是LICQ, 都只是 $\mathcal{F}(x^*) = T_{\Omega}(x^*)$ 的充分条件, 而不是必要条件。反例

$$x_1^2 + x_2 \geq 0, \quad x_1^2 - x_2 \geq 0$$

在 $x^* = (0, 0)^T$ 处, 约束限制都不成立, 但是其线性化方向集

$$\mathcal{F}(x^*) = \{(w_1, 0) | w_1 \in \mathbb{R}\}$$

确实准确地反映了 x^* 附近可行域的几何性质。

其他形式的一阶必要性条件

定义法向锥(Normal Cone)

如果 $x \in \Omega$, 称 $N_{\Omega}(x) = \{v | v^T w \leq 0 \text{ 对于任意的 } w \in T_{\Omega}(x)\}$ 为法向锥

定理

假设 x^* 是 f 在 Ω 内的一个局部最优解, 则

$$-\nabla f(x^*) \in N_{\Omega}(x^*).$$

主问题和对偶问题(Primal and Dual Problem)

主问题

$$\min_{x \in \mathcal{R}^n} f(x), \quad s.t. \quad c(x) \geq 0$$

其中 $c(x) = (c_1(x), \dots, c_m(x))^T$, $f(x)$ 和 $-c_i(x)$ 都是凸函数。

定义其拉格朗日函数为

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

再定义对偶问题的目标函数：

$$q(\lambda) \equiv \inf_x \mathcal{L}(x, \lambda)$$

通常情况下，对于某些 λ 该最小值可能是 $-\infty$ 。我们定义 λ 的可行区域为

$$\mathcal{D} \equiv \{\lambda \mid q(\lambda) > -\infty\}$$

对偶问题

$$\max_{\lambda \in \mathcal{R}^m} q(\lambda), \quad s.t. \quad \lambda \geq 0$$

弱对偶性

定理：对偶问题的目标函数 $q(\lambda)$ 以及它的区域 \mathcal{D} 都是凸的。

证明：

$$\begin{aligned}\mathcal{L}(x, (1-\alpha)\lambda^0 + \alpha\lambda^1) &= (1-\alpha)\mathcal{L}(x, \lambda^0) + \alpha\mathcal{L}(x, \lambda^1), \quad \forall \lambda^0, \lambda^1 \in \mathcal{R}^m, \alpha \in [0, 1] \\ q(x, (1-\alpha)\lambda^0 + \alpha\lambda^1) &\geq (1-\alpha)q(x, \lambda^0) + \alpha q(x, \lambda^1).\end{aligned}$$

定理：对于任意的 \bar{x} ，如果是满足主问题的解，和任意的 $\bar{\lambda} \geq 0$ ，都有 $q(\bar{\lambda}) \leq f(\bar{x})$

证明：

$$q(\bar{\lambda}) = \inf_x f(x) - \bar{\lambda}^T c(x) \leq f(\bar{x}) - \bar{\lambda}^T c(\bar{x}) \leq f(\bar{x}).$$

主问题和对偶问题的关联性

定理：如果 \bar{x} 是主问题的一个解，而且 f 和 $-c_i(x)$ 都是凸可微函数，那么对于任意满足KKT条件的 $\bar{\lambda}$ 都是其对偶问题的解

证明：假设 $(\bar{x}, \bar{\lambda})$ 满足KKT条件，则有

$$\nabla_x \mathcal{L}(\bar{x}, \bar{\lambda}) = 0, c(\bar{x}) \geq 0, \bar{\lambda} \geq 0, \bar{\lambda}^T c(\bar{x}) = 0.$$

根据 $\mathcal{L}(\cdot, \bar{\lambda})$ 是凸可微函数，即对任意的 x ，

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

因此

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

由于 $q(\bar{\lambda}) \leq f(\bar{x})$, $\forall \lambda \geq 0$ ，我们可以立即得到 $q(\bar{\lambda}) = f(\bar{x})$ ，所以 $\bar{\lambda}$ 是对偶问题的一个解。

主问题和对偶问题的关联性

定理：如果 f 和 $-c_i$ 都是凸可微，而且 \bar{x} 是主问题的一个解，并且满足 LICQ。假设 $\hat{\lambda}$ 是对偶问题的一个解， $\mathcal{L}(\cdot, \hat{\lambda})$ 是严格凸函数，并且在 \hat{x} 处取到 $\inf_x \mathcal{L}(x, \hat{\lambda})$ 的极小值。那么 $\hat{x} = \bar{x}$

- 如果 $\bar{x} \neq \hat{x}$ ，那么根据 \bar{x} 处满足 LICQ，则一定至少存在一个 $\bar{\lambda}$ ，使得 $(\bar{x}, \bar{\lambda})$ 是原主问题的解。
- 再根据上面的定理，可以得到 $\bar{\lambda}$ 是对偶问题的解，即 $\mathcal{L}(\bar{x}, \bar{\lambda}) = q(\bar{\lambda})$ 。
- 由题意， $q(\hat{\lambda})$ 也取到最小值，所以 $q(\bar{\lambda}) = q(\hat{\lambda}) = \mathcal{L}(\hat{x}, \hat{\lambda})$ 。
- 由于 $\hat{x} = \arg \min_x \mathcal{L}(x, \hat{\lambda})$ ，根据一阶最优性条件可以得到 $\nabla_x \mathcal{L}(\hat{x}, \hat{\lambda}) = 0$ 。
- 再根据 $\mathcal{L}(\cdot, \hat{\lambda})$ 的严格凸性，可以得到

$$\mathcal{L}(\bar{x}, \hat{\lambda}) - \mathcal{L}(\hat{x}, \hat{\lambda}) > \nabla_x \mathcal{L}(\hat{x}, \hat{\lambda})^T (\bar{x} - \hat{x}) = 0$$

- 所以，我们得到 $\mathcal{L}(\bar{x}, \hat{\lambda}) > \mathcal{L}(\hat{x}, \hat{\lambda}) = \mathcal{L}(\bar{x}, \bar{\lambda})$ 。
- 特别地，我们有 $-\hat{\lambda}^T c(\bar{x}) > -\bar{x}c(\bar{x}) = 0$ ，这里最后的不等式显然是矛盾。

Wolfe对偶

为了计算简便，对偶问题有一个更简洁的形式：

Wolfe Dual

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

定理：如果 f 和 $-c_i$ 是凸可微的。 $(\bar{x}, \bar{\lambda})$ 是主问题的一组解，并且满足LICQ。那么 $(\bar{x}, \bar{\lambda})$ 也是Wolfe对偶问题的一组解。

证明：

- 由于 $(\bar{x}, \bar{\lambda})$ 满足KKT条件，显然也满足Wolfe对偶问题的约束条件，并且 $\mathcal{L}(\bar{x}, \bar{\lambda}) = f(\bar{x})$ 。
- 再验证Wolfe对偶问题的第一个条件。对于任意的 (x, λ) ,

$$\begin{aligned} \mathcal{L}(\bar{x}, \bar{\lambda}) = f(\bar{x}) &\geq f(\bar{x}) - \lambda^T c(\bar{x}) = \mathcal{L}(\bar{x}, \lambda) \\ &\geq \mathcal{L}(x, \lambda) + \nabla_x \mathcal{L}(x, \lambda)(\bar{x} - x) = \mathcal{L}(x, \lambda) \end{aligned}$$

- 这样 $(\bar{x}, \bar{\lambda})$ 就是Wolfe对偶问题的一个极大值解。

对偶问题的例子

线性规划例子

$$\min c^T x \quad \text{subject to } Ax - b \geq 0,$$

对应的对偶目标函数

$$q(\lambda) = \inf_x [c^T x - \lambda^T (Ax - b)] = \inf_x [(c - A^T \lambda)^T x + b^T \lambda]$$

如果 $(c - A^T \lambda) \neq 0$, 极小值可以取到 $-\infty$ 。如果 $c - A^T \lambda = 0$, 则目标函数就是 $b^T \lambda$ 。后面要对 $q(\lambda)$ 取极大值, 所以我们可以排除掉 $c - A^T \lambda \neq 0$ 的情形。因此, 我们可以直接写出对偶问题:

$$\max_{\lambda} b^T \lambda \quad \text{subject to } A^T \lambda = c, \lambda \geq 0.$$

Wolfe对偶问题可以写为:

$$\max_{\lambda} c^T x - \lambda^T (Ax - b) \quad \text{subject to } A^T \lambda = c, \lambda \geq 0$$

对偶问题的例子

二次凸规划例子

$$\min \frac{1}{2}x^T Gx + c^T x \quad \text{subject to } Ax - b \geq 0, \quad G \text{ 对称正定.}$$

$$\text{对偶目标函数: } q(\lambda) = \inf_x \mathcal{L}(x, \lambda) = \inf_x \left[\frac{1}{2}x^T Gx + c^T x - \lambda^T (Ax - b) \right].$$

因为 G 是正定的, 所以 $\mathcal{L}(\cdot, \lambda)$ 是严格凸函数, 取到最小值的条件是 $\nabla_x \mathcal{L}(x, \lambda) = 0$, 即 $Gx + c - A^T \lambda = 0$. 从上式可以求出 $x = G^{-1}(A^T \lambda - c)$, 然后代入 $q(\lambda)$ 中得到对偶问题的目标函数:

$$q(\lambda) = -0.5(A^T \lambda - c)^T G^{-1}(A^T \lambda - c)^T + b^T \lambda$$

或者, 我们可以写出相应的Wolfe对偶问题:

$$\max_{(\lambda, x)} 0.5x^T Gx + c^T x - \lambda^T (Ax - b), \quad \text{s.t. } Gx + c - A^T \lambda = 0, \lambda \geq 0.$$

当然如果我们利用约束条件 $(c - A^T \lambda)^T x = -x^T Gx$, 可以重新写成对偶问题

$$\max_{(\lambda, x)} -0.5x^T Gx + \lambda^T b, \quad \text{s.t. } Gx + c - A^T \lambda = 0, \lambda \geq 0.$$

THANKS FOR YOUR ATTENTION