

约束优化问题

我们先快速回顾一下约束优化的基本问题形式。其一般形式为：

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & \begin{cases} c_i(x) = 0, & i \in \mathcal{E}, \\ c_i(x) \geq 0, & i \in \mathcal{I}. \end{cases} \end{aligned} \quad (12.1)$$

这里 f 和 c_i 都是定义在 \mathbb{R}^n 上的实值、光滑函数。而 \mathcal{I} 和 \mathcal{E} 是下标的有限集（默认为自然数的子集）。这里 f 仍然称为目标函数， $c_i \in \mathcal{E}$ 称为等值约束(equality constraints)，而 $c_i \in \mathcal{I}$ 称为不等值约束(inequality constraints)。满足所有约束条件的点集全体称为可行域(feasible set)，用

$$\Omega = \{x \mid c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\} \quad (12.2)$$

表示。这里指出一下不等值约束都是大于等于零只是一种约定，局限于本书。任何具体的文献或软件，均可以有自己的约定，在阅读或使用前应刚先搞清楚。

于是，问题(12.1)也可以写成紧凑形式：

$$\min_{x \in \Omega} f(x). \quad (12.3)$$

接下去我们重新讨论一下问题(12.3)的解的必要条件和充分条件，特别是约束在其中起的作用。

局部和全局解

我们直接引入例子使得讨论更加直观：

$$\min_{x \in \mathbb{R}^n} \|x\|_2^2, \quad \text{s.t. } \|x\|_2^2 \geq 1.$$

若不考虑约束，显然问题有全局最优解 $x = \vec{0}$ 。而当加了约束之后，所有 $\|x\|_2^2 = 1$ 的点都是问题的全局最优解。当 $n \geq 2$ 时，这样的解有无穷多个。

(草图)

再一个例子，

$$\min (x_2 + 100)^2 + 0.01x_1^2, \quad \text{s.t. } x_2 - \cos x_1 \geq 0. \quad (12.4)$$

如果不考虑约束，显然 (12.4) 有唯一的全局最小值点 $(0, -100)^T$ ，考虑约束之后，在

$$x^{(k)} = (k\pi, -1)^T, k = \pm 1, \pm 3, \dots$$

附近有无穷多个局部最小值点。

(草图)

因此我们有必要重新严格讨论在带约束条件下，局部解和全局解的定义。

定义 向量 x^* 称为(12.3)的一个局部最优解，若 $x^* \in \Omega$ ，且存在 x^* 的邻域 \mathcal{N} ，满足当 $x \in \mathcal{N} \cap \Omega$ 时，有 $f(x) \geq f(x^*)$ 。

定义 向量 x^* 称为(12.3)的一个严格局部最优解，若 $x^* \in \Omega$ ，且存在 x^* 的邻域 \mathcal{N} ，满足当 $x \in \mathcal{N} \cap \Omega, x \neq x^*$ 时，有 $f(x) > f(x^*)$ 。

定义 向量 x^* 称为(12.3)的一个孤立局部最优解，若 $x^* \in \Omega$ ，且存在 x^* 的邻域 \mathcal{N} ，使 x^* 是 $\mathcal{N} \cap \Omega$ 中 f 唯一的局部最优解。

(草图：解的类型判定。)

光滑性

现在光滑性不但指目标函数，也包括约束函数 $c_i(x)$ 。直观上，即可行域 Ω 的边界是光滑的，还是包含了尖点和跳跃。这里指的注意的是，即便全部约束函数 $c_i(x)$ 都是光滑的，也不表示可行域 Ω 一定有光滑的边界。例如：

区域

$$\|x\|_1 = |x_1| + |x_2| \leq 1 \quad (12.5)$$

等价于下述约束函数构成的可行域：

$$\begin{cases} x_1 + x_2 \leq 1, \\ x_1 - x_2 \leq 1, \\ -x_1 + x_2 \leq 1, \\ -x_1 - x_2 \leq 1. \end{cases} \quad (12.6)$$

(草图)

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。另一方面，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。例如：

$$\min f(x) = \max(x^2, x), \quad (12.7)$$

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：

$$\begin{aligned} \min \quad & t, \\ \text{s. t.} \quad & t \geq x, t \geq x^2. \end{aligned} \quad (12.8)$$

例子

接下去我们通过三个例子来建立对约束优化中的一系列基本概念的直观解释。

例 12.1： 含一个等值约束的二维问题。

$$\begin{cases} \min & x_1 + x_2, \\ \text{s. t.} & x_1^2 + x_2^2 - 2 = 0. \end{cases} \quad (12.9)$$

按之前约定，这里 $f(x) = x_1 + x_2$ ， $\mathcal{I} = \phi$ ， $\mathcal{E} = \{1\}$ ， $c_1(x) = x_1^2 + x_2^2 - 2$ 。因此它的可行域是一个圆（就是圆，没有内部），半径是 $\sqrt{2}$ ，圆心是原点。从等高线 $f(x) = x_1 + x_2 = c$ 可以看出，

(草图)

全局最优解为 $(-1, -1)^T$ 。在草图中观察，我们不难发现， x^* 的一个性质是 ∇c_1 和 ∇f 平行，但方向相反，这一点是特例么？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变化方向都有两个，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。直至到达 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，也即 f 不会下降。所以总结起来， x^* 真正满足的条件是： x^* 的任何保持在可行域内部的变化方向，都不是 f 的下降方向。这个直观是很容易接受的，现在我们通过分析的手段将它严格化。

从 $\forall x \in \Omega = \{x : c_1(x) = 0\}$ 出发的方向，保持 x 停留在 Ω 内部的变化方向 d （简称为“可行方向”或“ x 点的可行方向”）必须满足

$$c_1(x + d) = 0,$$

由Taylor展开（并带入 $c_1(x) = 0$ ），有

$$0 = c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d = \nabla c_1(x)^T d. \quad (12.11)$$

或者说，如果限制在一阶近似条件下，有

$$\nabla c_1(x)^T d = 0. \quad (12.12)$$

另一方面，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

$$0 > f(x + d) - f(x) \approx \nabla f(x)^T d \Rightarrow \nabla f(x)^T d < 0. \quad (12.13)$$

于是，只要(12.12)和(12.13)能同时成立，那么我们就可以找到一个从 x 出发的方向 d ，沿此方向：

- $x + d$ 停留在 Ω 内部；
- $f(x)$ 继续下降。

换言之，若 x 是问题的解，则上述两点一定不能**同时**成立。这里我们事实上已经给出了问题(12.9)的一阶必要条件：若 x^* 是(12.9)的解，则从 x^* 出发的任何方向 d ，不会同时满足(12.12)和(12.13)。由于在 x 点， f 的下降方向覆盖一个半平面（不含边界），而可行方向总是相对出现（ d 可行，则 $-d$ 必可行），

（草图）

因此惟有 $\nabla c_1(x)$ 和 $\nabla f(x)$ 平行才不会出现“既可行，又下降”的方向。此时，根据平行定义，有

$$\nabla f(x) = \lambda \nabla c_1(x), \lambda \in \mathbb{R}. \quad (12.10)$$

为方便计算，引入Lagrange乘子函数

$$L(x, \lambda) = f(x) - \lambda_1 c_1(x), \quad (12.16)$$

则(12.10)也可等价地表示为

$$\nabla_x L(x, \lambda_1) = \nabla f(x) - \lambda_1 \cdot \nabla c_1(x) = 0. \quad (12.17)$$

于是(12.9)的解的必要条件现在可以表示为：在解 x^* ，存在 $\lambda_1^* \in \mathbb{R}$ ，使

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

成立。这里 λ_1 称为约束 $c_1(x) = 0$ 对应的Lagrange乘子(multiplier)。

显然，再次强调这个条件是非充分的，比如(12.9)还有一个可行点 $(1, 1)^T$ 满足这个条件，但它是极大值点。此外， λ^* 的符号并不能说明对应 x^* 点的极值性质，因为这个符号实际上反映了我们记录 $c_1(x)$ 的一种人为约定，比如

$$c_1(x) = x_1^2 + x_2^2 - 2 = 0$$

可以等价写为

$$c_1(x) = 2 - x_1^2 - x_2^2 = 0,$$

从而将 $x^* = (-1, -1)^T$ 处的 λ_1^* 值从 $-\frac{1}{2}$ 变为 $\frac{1}{2}$ 。

例 12.2：只有一个不等值约束的优化。

$$\begin{cases} \min & x_1 + x_2 \\ \text{s. t.} & 2 - x_1^2 - x_2^2 \geq 0, \end{cases} \quad (12.18)$$

和上一个例子不同，现在可行域变为圆的内部和边界。对应Lagrange函数， λ_1^* 还是 $\frac{1}{2}$ ，最优解还是 $(-1, -1)^T$ ，但此时， λ_1^* 的符号是重要的。（注意此时 $c_1(x)$ 也不可随意变号，因为符号现在对应的是圆的内部和外部。）

回到分析手段, $\forall x \in \Omega$, 现在确保 $x + d$ 停留在可行域的前进方向在一阶Taylor展开中的表现为

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

现在无法消去第一项, 完整的一阶可行方向条件是

$$c_1(x) + \nabla c_1(x)^T d \geq 0. \quad (12.19)$$

而使 f 的下降方向的条件仍然是(12.13)。现在我们参照上一个例子的思路, 来考虑同时满足(12.13)和(12.19)的 d 都有哪些情况。

情况 I: x 严格在 Ω 内部, 也即有 $c_1(x) > 0$ 。此时, 从 x 出发的任何方向 d 都满足(12.19)。因此, 如果有 x^* 使得从 x^* 出发的任何方向都不同时满足(12.13)和(12.19), 那么剩下的唯一可能就是

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

也就是 x^* 事实上就是对应无约束问题下 f 的稳定点。这也是自然的, 因为这种情况下, (12.19)在一阶方向的选择上, 根本没有起到任何约束作用。

(草图)

情况 II: 当 x 落在 Ω 边界上, 也就是 $c_1(x) = 0$ 时。此时, 我们发现(12.19)的第一项又可以消去了, 因此从 x 出发的方向 d , 存在继续优化的前提变为

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

这里的第一个不等式限定了一个半平面 (不含边界), 第二个不等式限定了另一个半平面 (含边界)。要使这两个半平面相交为空, 惟有 $\nabla f(x)$ 和 $\nabla c_1(x)$ 是同一个方向 (比平行更严格), 也即

$$\nabla f(x) = \lambda_1 \nabla c_1(x), \lambda_1 \geq 0. \quad (12.21)$$

(草图)

而情况 I 和 II 可以用 Lagrange 函数的形式统一地表示为: 当 x^* 可行时, 存在 λ_1^* , 满足

$$\nabla_x L(x^*, \lambda_1^*) = 0, \lambda_1^* \geq 0, \quad (12.22)$$

并且, 必须有

$$\lambda_1^* c_1(x^*) = 0 \quad (12.23)$$

成立。这个条件被称为互补性条件, 它要求 λ_1^* 和 $c_1(x^*)$ 至少有一个为零。前者实际对应情况 I, 后者对应情况 II。注意情况 I 和情况 II 也有可能同时成立, 也即 λ_1^* 和 $c_1(x^*)$ 同时为零。这种情况要个别再讨论。

(草图)

这里我们注意到, 如果对于一个可行点 x , 如果对应的不等值约束 $c(x) \neq 0$, 那么实际上这个约束对 x 是无效的。也就是只有正好处于其边界的约束, 对一个点是否是最优值点的判定才会有意义。基于这一事实, 我们提出了

定义 12.1. 下标集合 $\mathcal{A}(x)$ 称为点 x 的活跃集 (active set, 也有翻作“积极集”), 它由全部等值约束的下标和在 x 点满足 $c_i(x) = 0$ 的不等值约束的下标组成。即

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

对于任何一个可行点 x , 一个不等值约束 c_i , $i \in \mathcal{I}$ 称为活跃的, 如果它的下标 $i \in \mathcal{A}(x)$, 也就是在 x 点, 约束 $c_i(x) = 0$; 称它为不活跃的, 如果 $c_i(x) > 0$ 。

例 12.3: 两个不等值约束。

$$\begin{cases} \min & x_1 + x_2 \\ \text{s. t.} & 2 - x_1^2 - x_2^2 \geq 0 \\ & x_2 \geq 0. \end{cases} \quad (12.24)$$

(草图)

显然最优解现在是 $(-\sqrt{2}, 0)^T$ 。我们注意到在这一点，两个约束都是活跃的。对于这样两个约束都活跃的点（我们这里先无视由于维数的原因导致这样的点只有两个） x ，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 d ，这个逻辑我们重复使用到现在了），那么必然有方向 d 满足

$$\nabla c_i(x)^T d \geq 0, i \in \{1, 2\}, \nabla f(x)^T d < 0. \quad (12.25)$$

比如从最优解 $x^* = (-\sqrt{2}, 0)^T$ 出发就不存在这样的方向。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， $\nabla f(x^*)^T d > 0$ ，也即只有严格上升方向。

(草图)

我们仍然将这一结论统一到Lagrange函数的框架下。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。于是 $\nabla f(x)$ 和 $c_i(x)$, $i \in \mathcal{I}$ 的关系可以表达为：存在 $\lambda^* \geq 0$ （表示每一个分量都大于等于零），使得：

$$\nabla_x L(x^*, \lambda^*) = 0, \quad (12.26)$$

并且满足互补性条件

$$\lambda_1^* c_1(x^*) = 0, \quad \lambda_2^* c_2(x^*) = 0. \quad (12.27)$$

这里我们看到互补性条件其实表达了这么一种意思：一个约束要么是活跃的（ $c_i(x) = 0$ ），因此会在(12.26)中参与形成和 $\nabla f(x)$ 的线性关系；要么是不活跃的，此时它在(12.26)中实际上是不存在的（ $\lambda_i = 0$ ）。

我们下面具体地形式检查一下 $x^* = (-\sqrt{2}, 0)^T$ ，有

$$\nabla f(x^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} -2x_1^* \\ -2x_2^* \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

代入

$$\begin{aligned} \nabla_x L(x^*, \lambda^*) &= \nabla f(x^*) - \lambda_1^* \nabla c_1(x^*) - \lambda_2^* \nabla c_2(x^*) \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2\sqrt{2}\lambda_1^* \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ \lambda_2^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \end{aligned}$$

解得：

$$\begin{cases} \lambda_1^* &= \frac{1}{2\sqrt{2}}, \\ \lambda_2^* &= 1, \end{cases}$$

故 $\lambda^* \geq 0$ ，且由于 $c_1(x)$ 和 $c_2(x)$ 在 x^* 点均活跃，故

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

即互补性条件成立。

再观察一下另一个两个约束都活跃的点 $x = (\sqrt{2}, 0)^T$ 。此时

$$\nabla f(x^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} -2\sqrt{2} \\ 0 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

于是

$$\nabla_x L(x, \lambda) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} \lambda_1 \cdot (-2\sqrt{2}) \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} \lambda_1 = -\frac{1}{2\sqrt{2}}, \\ \lambda_2 = 0. \end{cases}$$

不满足条件(12.26)。

而点 $x = (1, 0)^T$ ，则只有 $c_2(x)$ 在该点活跃，也即 $c_1(x) > 0$ ，因此由互补性条件，必有 $\lambda_1 = 0$ 。于是我们只需验证

$$\nabla f(x) - \lambda_2 \nabla c_2(x) = 0 \Rightarrow \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ \lambda_2 \end{bmatrix} = 0.$$

这个等式显然不可能被满足。

一阶优化条件

现在我们将上述例子带来的结论推广到一般形式。我们注意到，对于 x^* 是问题最优解的一阶必要性条件，需要包含两个部分：

1. 存在 λ^* 使得 $\nabla_x L(x^*, \lambda^*) = 0$ ，其中 $\lambda_i^* \geq 0$ ，对 $i \in \mathcal{I}$ ；
2. 互补性条件： $\lambda_i^* c_i(x^*) = 0$ ， $i \in \mathcal{I}$ 。

为得到更加严格的结果，我们先给出Lagrange函数的一般形式：

$$L(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x). \quad (12.33)$$

可行点 x 的活跃集 $\mathcal{A}(x)$ 的定义为全部等值约束的下标和全部在 x 点等号成立的不等值约束的下标：

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

而全部活跃的 $\nabla c_i(x^*)$ 的实际起的作用是构成 $\nabla_x f(x^*)$ 的一个线性表出关系：

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

其中 λ_i^* 实际就是表出系数，那么问题就来了，这样的 $\nabla c_i(x^*)$ 都是线性无关的么？这里值得注意的一个关键问题是， $\nabla c_i(x)$ 其实有很多等价的解析表达形式，而 $\nabla c_i(x^*)$ 的线性无关性和具体的表达形式都是相关的。例如，在(12.9)中，做如下替换

$$c_1(x) = (x_1^2 + x_2^2 - 2)^2 = 0 \Leftrightarrow c_1(x) = x_1^2 + x_2^2 - 2 = 0,$$

从几何上看，完全一致，但是注意，对左边的约束，总有 $\nabla c_1(x) = 0$ ，直接导致形式上

$$\nabla f(x) = \lambda_1 \nabla c_1(x)$$

结构从表达约束的角度失去作用。这里由于特殊的代数表达形式， $\nabla c_1(x)$ 发生了退化（从一个向量角度说，退化就是失去线性无关性）。这提示我们，事实上，对 $c_i(x)$ 的解析表达式的写法是有要求的。必须提出一种“正确的写法”，或者称为“规范”(qualification)。具体的规范有很多，下面是一种简单而常用的规范：

定义 12.4 (LICQ) 对 x^* 以及由上述定义12.1定义的 $\mathcal{A}(x^*)$ ，称满足线性无关约束条件(linear independence constraint qualification, LICQ)，若其活跃的约束的梯度集合

$$\{\nabla c_i(x^*), i \in \mathcal{A}(x^*)\}$$

是线性无关组。

在此基础上，我们终于能完整地提出约束优化问题的一阶必要条件。

定理 12.1 假设 x^* 是(12.1)的局部最优解，且在 x^* 满足LICQ，则存在Lagrange乘子 λ^* ，其中 $\lambda_i^*, i \in \mathcal{E} \cup \mathcal{I}$ 在 (x^*, λ^*) 满足：

$$\nabla_x L(x^*, \lambda^*) = 0, \quad (12.34a)$$

$$c_i(x^*) = 0, \quad \forall i \in \mathcal{E}, \quad (12.34b)$$

$$c_i(x^*) \geq 0, \quad \forall i \in \mathcal{I}, \quad (12.34c)$$

$$\lambda_i^* \geq 0, \quad \forall i \in \mathcal{I}, \quad (12.34d)$$

$$\lambda_i^* \cdot c_i(x^*) = 0, \quad \forall i \in \mathcal{E} \cup \mathcal{I}. \quad (12.34e)$$

条件(12.34)称为Karush-Kuhn-Tucker条件，即KKT条件。由互补性条件(12.34e)知，当 $i \notin \mathcal{A}(x^*)$ 时，相应的 λ_i^* 必为零，因此我们可以把(12.34a)改写为

$$0 = \nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*). \quad (12.35)$$

上式直观地表达了在可行点 x^* ，要么 $\nabla f(x^*) = 0$ ，因此是无约束问题的最优解；要么全部从 x^* 出发的可行方向（在 $\nabla c_i(x^*)$ 交集构成的“锥体”内），没有 f 的下降方向。（从 x^* 出发，不存在既可行，又下降的方向。）

这里(12.34e)其实留下了一个漏洞，即若 λ_i 和 $c_i(x^*)$ 都等于零，会发生上述哪一种情况？这种情况其实需要进一步分析。不过我们这里先用一种简单粗暴的方式排斥这种例外。

定义 12.5 严格互补(strict complementarity) 对(12.1)的局部解 x^* 和满足(12.34)的向量 λ^* ，我们称严格互补条件成立，若对每一个 $i \in \mathcal{I}$ ， λ_i^* 和 $c_i(x^*)$ 始终有且仅有一个为零。或者等价地， $\lambda_i^* > 0$ 对全部 $i \in \mathcal{I} \cap \mathcal{A}(x^*)$ 成立。

一般地，对给定的问题(12.1)，和解 x^* ，可能有多个向量 λ^* 满足(12.34)，但当LICQ成立时， λ^* 是唯一的。这一点由

$$\nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*)$$

是显然的。同时，上述表示也看出， λ_i^* 的实际意义是对应 $\nabla c_i(x^*)$ 在 $\nabla f(x^*)$ 中的权重， λ_i^* 越大的约束，在约束 f 无法下降中起的作用也越大，从算法角度看，越敏感。

比如，如果对 $c_i(x)$ 的函数值施加一个小扰动，导致 $c_i(x) \geq 0$ 变为 $c_i(x) \geq -\varepsilon \|\nabla c_i(x^*)\|$ ，则对 x^* 的活跃约束有

$$\frac{c_i(x^*(\varepsilon)) - c_i(x^*)}{x^*(\varepsilon) - x^*} = \nabla c_i(x^*) \Rightarrow c_i(x^* + \varepsilon) - c_i(x^*) \approx (x^*(\varepsilon) - x^*)^T \nabla c_i(x^*).$$

这里 $x^*(\varepsilon)$ 对应扰动后的新解。而对 $j \neq i$ ，因为扰动未改变 $c_j(x)$ ，故 $x^*(\varepsilon)$ 和 x^* 对 $c_j(x)$ 而言有相同的函数值，即

$$0 = c_j(x^*(\varepsilon)) - c_j(x^*) \approx (x^*(\varepsilon) - x^*)^T \nabla c_j(x^*).$$

于是对 $f(x^*(\varepsilon))$ ，有

$$\begin{aligned} f(x^*(\varepsilon)) - f(x^*) &\approx (x^*(\varepsilon) - x^*)^T \nabla f(x^*) \\ &= \sum_{j \in \mathcal{A}(x^*)} \lambda_j^* (x^*(\varepsilon) - x^*)^T \nabla c_j(x^*) \\ &\approx -\varepsilon \|\nabla c_i(x^*)\| \lambda_i^*. \end{aligned}$$

最后一步是因为对 $j \neq i$ ，求和项均为零。进而，

$$\frac{df(x^*(\varepsilon))}{d\varepsilon} = \lambda_i^* \|\nabla c_i(x^*)\|. \quad (12.80)$$

因此, 严格地说, $\lambda_i^* \|\nabla c_i(x^*)\|$ 才是第 i 个约束对解目标值的敏感影响因子。

定义 12.8 令 x^* 是问题(12.1)的解, 且满足KKT条件(12.34)。称不等值约束 $c_i(x)$ 是强约束的(strongly active)或绑定的(binding), 若有满足(12.34)的 λ^* 使 $i \in \mathcal{A}(x^*)$ 且 $\lambda_i^* > 0$ 。称 $c_i(x)$ 是弱活跃的(weakly active), 若对所有满足(12.34)的 λ^* , 有 $i \in \mathcal{A}(x^*)$ 且 $\lambda_i^* = 0$ 。

这个定义实际上和定义12.5讲述的是同样的事。

为了证明定理12.1, 我们需要再做一些技术准备。