

切锥(Tangent Cone)和表示规范的本质

我们已经注意到，解 x^* 的一阶局部性质是由邻近的可行域 Ω 决定的。同时，也是由在 x^* 出发的全部一阶可行方向集合决定的。这两者是否是等价的？应该不是，因为后者显然还和约束的具体代数表示形式有关，而前者完全是一个几何概念。这一节我们严格定义这两者，并分析它们之间的关系，为接下去的证明奠定基础。

首先，对于一个可行点 x ，若有序列 $\{z_k\} \rightarrow x$ ，且对充分大的 k ，均有 z_k 可行，则称序列 $\{z_k\}$ 为趋向(approaching) x 点的渐进可行序列(feasible sequence)。

定义 12.2 称向量 d 是 Ω 在 x 点的切向量(tangent)，若存在一个趋向 x 的可行序列 $\{z_k\}$ ，以及一个收敛到0的正数序列 $\{t_k\}$ ，使得

$$\lim_{k \rightarrow \infty} \frac{z_k - x}{t_k} = d. \quad (12.29)$$

Ω 在 x^* 点的全部切向量称为 Ω 在 x^* 点的切锥(tangent cone)，记作 $T_\Omega(x^*)$ 。

容易证明（但仍需证明）， $T_\Omega(x^*)$ 是一个锥。（锥的定义见参考书附录A.36。）

显然，切锥是一个完全几何的定义。从代数的角度，如果只考虑一阶局部信息，我们可以类似地给出一个定义：

定义 12.3 对可行点 x ，以及相应的活跃集 $\mathcal{A}(x)$ ，集合

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

称为线性化可行方向集。

显然（呵呵）， $\mathcal{F}(x)$ 也是一个锥。但是， $\mathcal{F}(x)$ 却不是纯粹几何的，它依赖 c_i ， $i \in \mathcal{E} \cup \mathcal{I}$ 的具体解析表达形式。

例 12.4 问题和例 12.1一样。对可行点 $x = (-\sqrt{2}, 0)^T$ ，我们构建趋于 x 的可行序列

$$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ -1/k \end{bmatrix}. \quad (12.30)$$

并选择 $t_k = \|z_k - x\|$ ，则 $d = (0, -1)^T$ 是切向量。而对另一个从相反方向趋于 x 的序列

$$z_k = \begin{bmatrix} -\sqrt{2 - 1/k^2} \\ 1/k \end{bmatrix},$$

则 $d = (0, 1)^T$ 是切向量。综合起来，在 $x = (-\sqrt{2}, 0)^T$ 的切锥是

$$\{(0, d_2)^T \mid d_2 \in \mathbb{R}\}.$$

而对 $d = (d_1, d_2)^T \in \mathcal{F}(x)$ ，有

$$0 = \nabla c_1(x)^T d = [2x_1 \quad 2x_2] \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -2\sqrt{2}d_1.$$

因此， $\mathcal{F}(x) = \{(0, d_2)^T \mid d_2 \in \mathbb{R}\}$ 。在这个例子中，我们看到， $T_\Omega(x) = \mathcal{F}(x)$ 。

但是如果我们把可行域改为

$$\Omega = \{x \mid c_1(x) = (x_1^2 + x_2^2 - 2)^2 = 0\}. \quad (12.31)$$

则

$$0 = \nabla c_1(x)^T d = \begin{bmatrix} 4(x_1^2 + x_2^2 - 2)x_1 & 4(x_1^2 + x_2^2 - 2)x_2 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix},$$

对任意的 $(d_1, d_2)^T$ 都是成立的，也即 $\mathcal{F}(x) = \mathbb{R}^2$ 。而 $T_\Omega(x)$ 没有任何变化（可行域的几何性质不变），因此在这种情况下，切锥和线性化可行方向集不一致。

例 12.5 再次回顾例12.2。现在 $x = (-\sqrt{2}, 0)^T$ 处的情况变得更加复杂。首先，趋于 x 的可行序列不但可以从边界趋近，也可以从内部，比如

$$z_k = (-\sqrt{2}, 0)^T + (1/k)w,$$

其中 w 是任意位于右半平面内的向量，即 $w = (w_1, w_2)^T, w_1 > 0$ 。当 $\|z_k\| \leq \sqrt{2}$ 时，序列值 z_k 是可行的，也即

$$(-\sqrt{2} + w_1/k)^2 + (w_2/k)^2 \leq 2,$$

化简得：

$$k \geq (w_1^2 + w_2^2)/(2\sqrt{2}w_1).$$

由于 $1/k$ 本身就是一个退化的正项序列，因此在 $x = (-\sqrt{2}, 0)^T$ 处的切锥就是

$$T_\Omega(x) = \{(w_1, w_2)^T | w_1 \geq 0\}.$$

而由

$$0 \leq \nabla c_1(x)^T d = \begin{bmatrix} -2x_1 & -2x_2 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = 2\sqrt{2}d_1,$$

知 $\mathcal{F}(x) = \{d = (d_1, d_2)^T | d_1 \geq 0\}$ ，因此在此表示下， $\mathcal{F}(x) = T_\Omega(x)$ 。

有上面的基础，我们现在可以讨论表示规范的实质。切锥代表了问题的客观几何性质，而线性化可行方向集则给我们了一个可以具体计算的解析表达形式。当这两者不一致时，一切算法和推导，都是基于有解析表达的线性化可行方向集，因此会给出错误的、和几何不一致的结果。所以，一个最原始、最基础的表示规范，就是

$$\mathcal{F}(x) = T_\Omega(x).$$

但是这个规范没有任何意义。因为问题的根源就在于我们无法直接操作几何属性的 $T_\Omega(x)$ 。因此，我们需要建立一些强条件下的特例，使得在满足这些条件时，事实上有 $\mathcal{F}(x) = T_\Omega(x)$ ，从而只需通过可计算和推导的 $\mathcal{F}(x)$ ，获得真正有几何意义的解。比如，LICQ（我们以后会严格证明这一点）。

最后我们再讨论一个例子：

$$c_1(x) = 1 - x_1^2 - (x_2 - 1)^2 \geq 0, c_2(x) = -x_2 \geq 0, \quad (12.32)$$

现在在这里实际的可行域只有单点 $\Omega = \{(0, 0)^T\}$ 。对该点 $x = (0, 0)^T$ ，显然切锥就是 $T_\Omega(x) = \{(0, 0)^T\}$ ，因为所有趋于 x 的可行序列必须满足当 k 充分大时， $z_k = x = (0, 0)^T$ 。进一步，我们也容易验证

$$\mathcal{F}(x) = \{(d_1, 0)^T | d_1 \in \mathbb{R}\}.$$

因此在这个例子，线性化可行方向集并没有在 x 点抓住问题的几何特征，也即它不满足约束规范。

一阶最优条件的证明

现在我们来考虑定理12.1的证明。这个证明尽管漫长，但确实是值得的。因为体现了整个优化领域的一些基本思想。

我们下面用 $A(x^*)$ 来表示

$$A(x^*)^T = [\nabla c_i(x^*)]_{i \in \mathcal{A}(x^*)}, \quad (12.37)$$

注意这里 $A(x^*)$ 是一个矩阵，它的行向量是活跃约束的梯度。

引理 12.2 令 x^* 是一个可行点。则以下两个命题成立：

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

证明：不失一般性(without loss of generality)，假设约束 $c_i(x)$ 在 x^* 活跃， $i = 1, 2, \dots, m$ 。对命题1，任取 $d \in T_\Omega(x^*)$ ，令序列 $\{z_k\}$ 和 $\{t_k\}$ 满足

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

且对所有 k ， $t_k > 0$ 。于是有

$$z_k = x^* + t_k d + o(t_k). \quad (12.38)$$

对 $i \in \mathcal{E}$ ，使用Taylor展开，我们有

$$\begin{aligned} 0 &= \frac{1}{t_k} c_i(z_k) \quad (c_i(z_k) = 0, \text{ 对 } k \text{ 充分大}) \\ &= \frac{1}{t_k} [c_i(x^*) + t_k \nabla c_i(x^*)^T d + o(t_k)] \quad (\text{注意首项为零}) \\ &= \nabla c_i(x^*)^T d + \frac{o(t_k)}{t_k} \end{aligned}$$

上式两端对 $k \rightarrow \infty$ ，得

$$\nabla c_i(x^*)^T d = 0, \forall i \in \mathcal{E}.$$

再对 $i \in \mathcal{A}(x^*) \cap \mathcal{I}$ ，类似地有

$$\begin{aligned} 0 &\leq \frac{1}{t_k} c_i(z_k) \\ &= \frac{1}{t_k} [c_i(x^*) + t_k \nabla c_i(x^*)^T d + o(t_k)] \quad (\text{首项也为零，因为 } i \in \mathcal{A}(x^*)) \\ &= \nabla c_i(x^*)^T d + \frac{o(t_k)}{t_k}. \end{aligned}$$

即

$$\nabla c_i(x^*)^T d \geq 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{I}.$$

综上，必有 $d \in \mathcal{F}(x^*)$ ，也即有命题1成立。

对命题2，由LICQ成立，矩阵 $A(x^*)$ 必然是行满秩的。注意 $A(x^*)$ 是 $m \times n$ 的，且 $n \geq m$ （为啥？）。故可令矩阵 Z 的列向量是 $A(x^*)$ 的零空间的基，也即

$$A(x^*)Z = 0, Z \in \mathbb{R}^{n \times (n-m)}, \quad (12.39)$$

其中 Z 列满秩。（即 $A(x^*)$ 作为一个线性方程组的系数矩阵，则该线性方程组是欠定的，它对应的齐次方程组的解是一个空间， Z 的列向量就是它的全部基。更进一步的理论参见《高等代数》课本。）任取 $d \in \mathcal{F}(x^*)$ ，令 $\{t_k\}_{k=0}^\infty$ 是任意的正项序列，且满足 $t_k \rightarrow 0, k \rightarrow \infty$ 。定义映射 $R: \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$ 以及参数方程组如下：

$$R(z, t) = \begin{bmatrix} c(z) - tA(x^*)d \\ Z^T(z - x^* - td) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (12.40)$$

(注意上面系统的第一块是 m 行, 而第二块是 $n - m$ 行, 所以总系统是 n 行的。)

这里断言(claim), 对充分大的 k , $t = t_k$ 充分小, 而 $z_k = z$ 为上述方程组的解, 则如此构建的 $\{z_k\}$ 和 $\{t_k\}$ 满足趋于 x^* 的渐近可行序列的定义, 也即

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

(从这个断言我们也可以想象一下为何几何条件很难直接应用...)

事实上, 对 $t = 0, z = x^*$, R 的Jacobi矩阵为

$$\nabla_z R(x^*, 0) = \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix}, \quad (12.41)$$

注意其第一块由LICQ是行满秩的, 而第二块由定义也是行满秩的。因此整体上该矩阵非奇异。由隐函数定理, 对充分小的 t_k , 方程组存在唯一解 z_k 。并且, 对 $i \in \mathcal{A}(x^*)$, 由(12.40)的前 m 行, 有

$$c(z_k) = t_k A(x^*)d,$$

因此有

$$i \in \mathcal{E} \Rightarrow c_i(z_k) = t_k \nabla c_i(x^*)^T d = 0, \quad (12.42a)$$

$$i \in \mathcal{A}(x^*) \cap \mathcal{I} \Rightarrow c_i(z_k) = t_k \nabla c_i(x^*)^T d \geq 0, \quad (12.42b)$$

所以, 对充分大的 k 以及相应充分小的 t_k , z_k 是可行的。(还没有结束!)

接下去我们还需要证明, 对此 $\{z_k\}$, 有

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

成立。这里我们对 $R(z_k, t_k)$ 应用Taylor定理, 有

$$\begin{aligned} 0 = R(z_k, t_k) &= \begin{bmatrix} c(z_k) - t_k A(x^*)d \\ Z^T(z_k - x^* - t_k d) \end{bmatrix} \\ &= [c(x^*) + \nabla c(x^*)^T(z_k - x^*) + o(\|z_k - x^*\|) - t_k A(x^*)d] \\ &\quad \text{(这里将第一项 } c(z_k) \text{ 在 } x^* \text{ 展开, 注意第一项是零, 第二项中 } \nabla c(x^*) \text{ 就是 } A(x^*)) \\ &= \begin{bmatrix} A(x^*)(z_k - x^*) + o(\|z_k - x^*\|) - t_k A(x^*)d \\ Z^T(z_k - x^* - t_k d) \end{bmatrix} \\ &= \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix} (z_k - x^* - t_k d) + o(\|z_k - x^*\|), \end{aligned}$$

两边同除以 t_k , 并注意到最后一行的系数矩阵是非奇异的, 有

$$\frac{z_k - x^*}{t_k} = d + o\left(\frac{\|z_k - x^*\|}{t_k}\right),$$

因此 $d \in T_\Omega(x^*)$, 证毕。□

定理 12.3. 若 x^* 是问题(12.1)的解, 则

$$\nabla f(x^*)^T d \geq 0, \forall d \in T_\Omega(x^*). \quad (12.43)$$

(这个定理说的是, 从最优解出发的全部几何方向, 都是不下降方向。这个显然是对的。这是一个非常基础的必要条件。)

证明: 反证, 假设存在 $d \in T_\Omega(x^*)$, 且

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

则令 $\{z_k\}$ 和 $\{t_k\}$ 是对应的可行序列和正序列，于是

$$\begin{aligned} f(z_k) &= f(x^*) + (z_k - x^*)^T \nabla f(x^*) + o(\|z_k - x^*\|) \\ &= f(x^*) + t_k d^T \nabla f(x^*) + o(t_k). \end{aligned}$$

这是因为由(12.38),

$$z_k = x^* + t_k d + o(t_k).$$

由 $d^T \nabla f(x^*) < 0$ ，当 k 充分大时，有

$$f(z_k) - f(x^*) = t_k d^T \nabla f(x^*) + o(t_k) < 0,$$

这与 x^* 是局部最小值点矛盾。□

显然，这个条件是不充分的，即便对 $\forall d \in T_\Omega(x^*)$ ，都有 $\nabla f(x^*)^T d \geq 0$ （从 x^* 点出发的切锥方向都不下降），也未必有 x^* 是局部极小值。

反例：

$$\min x_2, \quad \text{s.t. } x_2 \geq -x_1^2, \quad (12.44)$$

该问题的可行域是无界的，从草图不难看出它既没有全局最优解，也没有局部最优解。然而，对 $x^* = (0, 0)^T$ ，从它出发的切线均在上半平面，于是有 $\forall d \in T_\Omega(x^*)$ ， $\nabla f(x^*)^T d \geq 0$ 。这件事情的本质是即便 $T_\Omega(x^*)$ 也不是全部信息，它只是几何上的一阶信息，从我们在无约束优化中的经验，要确立充分条件，需要二阶局部信息参与（正定）。这个问题我们之后再讨论，目前先设法证明真正的一阶必要条件。

我们先考虑一个一般意义上的几何锥

$$K = \{By + Cw \mid y \geq 0\}, \quad (12.45)$$

其中， B 和 C 分别是 $n \times m$ 和 $n \times p$ 阶矩阵， y 和 w 是对应维数的向量。

（这个锥的定义比我们直观中的锥稍微更一般一点，它允许在一些方向上是全部而另一些方向上是正向。比如三维空间的一个半平面，按照这个定义也是一个锥。）

现在对于一个 $g \in \mathbb{R}^n$ ，显然要么有 $g \in K$ ，要么 $g \notin K$ 。但一个叫Farkas的几何学家指出，后者可以改为存在 $d \in \mathbb{R}^n$ ，使得

$$g^T d < 0, B^T d \geq 0, C^T d = 0. \quad (12.46)$$

这件事情的真正意义在于，将一件什么是锥，以及它如何划分空间这样一件纯几何的事情，和(12.46)这样的代数形式建立了联系。它本质上涉及到几何信息可以如何正确表达这样一件非常基础的事实，同时，也能解决我们现在遇到的问题。这个事实总结为下述引理：

引理12.4 (Farkas引理) 令锥 K 如(12.45)所定义。则对任意 $g \in \mathbb{R}^n$ ，要么 $g \in K$ ，否则必有 $d \in \mathbb{R}^n$ 满足(12.46)。

证明：我们首先证明，这两种情况不会同时成立。否则，若 $g \in K$ ，即有 $y \geq 0$ 和 w ，使

$$g = By + Cw,$$

同时又有 $d \in \mathbb{R}^n$ ，使 $g^T d < 0$ ， $B^T d \geq 0$ ， $C^T d = 0$ ，则

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

矛盾说明两种情况不会同时成立。

现证对 $g \notin K$, 必存在 d 满足(12.46)。但这里最意外的一点是首先要证明 K 是闭集。这一点似乎是显然的, 但是由于我们现在的证明是深入几何基础的, 所以仍然需要有一个证明。具体过程参见后面的引理 12.15。你会发现它的证明还是相当复杂的。我们下面直接接受这个结论。

令 \hat{s} 是 g 在 K 中的最佳逼近元, 即

$$\hat{s} = \operatorname{argmin} \|s - g\|_2^2, \quad s \in K. \quad (12.47)$$

由 $\hat{s} \in K$, 且 K 是锥, 故 $\alpha \hat{s} \in K, \forall \alpha \in \mathbb{R}$ 。由 \hat{s} 定义, 当 $\alpha = 1$ 时, 有

$$\|\alpha \hat{s} - g\|_2^2 = (\alpha \hat{s} - g)^T (\alpha \hat{s} - g)$$

取到最小 (因为 \hat{s} 就是最小)。故

$$\begin{aligned} \frac{d}{d\alpha} \|\alpha \hat{s} - g\|_2^2 \Big|_{\alpha=1} &= 0 \\ \Rightarrow \frac{d}{d\alpha} (\alpha \hat{s} - g)^T (\alpha \hat{s} - g) \Big|_{\alpha=1} &= \frac{d}{d\alpha} (\alpha^2 \hat{s}^T \hat{s} - 2\alpha g^T \hat{s} + g^T g) \Big|_{\alpha=1} \\ &= 2\alpha \hat{s}^T \hat{s} - 2g^T \hat{s} \Big|_{\alpha=1} \\ &= 2\hat{s}^T \hat{s} - 2g^T \hat{s} = 2\hat{s}^T (\hat{s} - g) = 0. \end{aligned} \quad (12.48)$$

(这是我们熟知的一个结论, 最佳逼近元事实上是正交投影。)

令 s 是 K 中任意向量, 由 K 凸,

$$\|\hat{s} + \theta(s - \hat{s}) - g\|_2^2 \geq \|\hat{s} - g\|_2^2, \forall \theta \in [0, 1].$$

(这里 $\hat{s} + \theta(s - \hat{s}) = (1 - \theta)\hat{s} + \theta s$ 是 s 和 \hat{s} 的凸组合, 由 $s, \hat{s} \in K$, 其凸组合必属于 K , 因此到 g 的距离必大于最佳逼近距离。)

将上式转成向量内积形式:

$$\begin{aligned} (\hat{s} - g)^T (\hat{s} - g) + 2\theta(s - \hat{s})^T (\hat{s} - g) + \theta^2(s - \hat{s})^T (s - \hat{s}) - \|\hat{s} - g\|_2^2 &\geq 0, \\ \Rightarrow 2\theta(s - \hat{s})^T (s - g) + \theta^2\|s - \hat{s}\|_2^2 &\geq 0, \end{aligned}$$

两边同除以 θ (对 $\theta \neq 0$, $\theta = 0$ 时结论是显然的), 并令 $\theta \rightarrow 0$, 有

$$(s - \hat{s})^T (\hat{s} - g) \geq 0.$$

由(12.48),

$$\hat{s}^T (\hat{s} - g) \geq 0,$$

于是

$$(s - \hat{s})^T (\hat{s} - g) = s^T (\hat{s} - g) - \hat{s}^T (\hat{s} - g) = s^T (\hat{s} - g) \geq 0, \forall s \in K. \quad (12.49)$$

到目前为止, 我们还没有涉及到底那个 d 在哪里这样的实质问题。现在, 令

$$d = \hat{s} - g \Leftrightarrow g = \hat{s} - d.$$

(这里 d 的几何意义也出现了。)

由 $g \notin K$, $d \neq 0$, 且

$$d^T g = d^T (\hat{s} - d) = d^T \hat{s} - d^T d = (\hat{s} - g)^T \hat{s} - d^T d = -\|d\|_2^2 < 0,$$

(这里又用了(12.48))

故 d 满足(12.46)的第一个条件。再由(12.49),

$$s^T(\hat{s} - g) = s^T d \geq 0, \forall s \in K,$$

于是

$$d^T(By + Cw) \geq 0, \forall y \geq 0, w \in \mathbb{R}^n.$$

令 $y = 0$, 则有

$$(C^T d)^T w \geq 0,$$

由 w 任意性, 唯有

$$C^T d = 0.$$

也即(12.46)第三项成立。

再令 $w = 0$, 有

$$(B^T d)^T y \geq 0, \forall y \geq 0 \Rightarrow B^T d \geq 0.$$

这是(12.46)第二项, 于是 d 满足(12.46), 证毕。□

推论 考虑Farkas引理的一个特例: 锥

$$N = \left\{ \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*), \lambda_i \geq 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \right\}, \quad (12.50)$$

并令

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

则要么

$$\nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*) = A(x^*)^T \lambda, \quad (12.51)$$

其中 $\lambda_i \geq 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{I}$; 要么存在 d , 使得

$$d^T \nabla f(x^*) < 0, d \in \mathcal{F}(x^*).$$

(这里实际上完整的 K 是

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

因此结论是要么 $g = \nabla f(x^*)$ 在 K 锥内, 也即存在满足要求的 λ 可以被如此表出, 要么存在 d 满足

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

也就是 $d \in \mathcal{F}(x^*)$, 且 d 上升。

)

定理12.1的证明: 设 $x^* \in \mathbb{R}^n$ 是问题(12.1)的一个可行点, 且满足LICQ。进一步, x^* 还是(12.1)的局部最优。由定理12.3, 任取 $d \in T_\Omega(x^*)$, 有

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

再由引理12.2, 对LICQ, 有

$$T_\Omega(x^*) = \mathcal{F}(x^*).$$

结合上述两个结论，我们有任取 $d \in \mathcal{F}(x^*)$,

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

于是由Fark's引理（引理12.4），在两个结论中只能前一个成立，也即(12.51)成立。即存在 λ 使得

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

其中 $\lambda_i \geq 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{I}$ 。现在对 λ^* 如下设置：

$$\lambda_i^* = \begin{cases} \lambda_i, & i \in \mathcal{A}(x^*), \\ 0, & i \in \mathcal{I} \setminus \mathcal{A}(x^*), \end{cases} \quad (12.52)$$

则不难验证，上述 λ^* 满足(12.34)。证毕。□