最优化理论和方法

约束问题的最优性条件

第九章　约束问题的最优性条件

第一节　可行方向

第二节　约束问题最优性条件

考虑一般约束问题：

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & g_i(x) \geq 0, i \in I = \{1, 2, \cdots, m_1\} \\ & h_j(x) = 0, j \in E = \{m_1 + 1, \cdots, \mathrm{m}\} \end{aligned} \qquad (9.1)$$

可行域： $D = \{x : g_i(x) \geq 0, i \in I; h_j(x) = 0, j \in E\}$

这里我们假设函数 f, g_i, h_j 连续可微

显然可行域 D 为闭集.

非光滑无约束优化有时可转化为光滑的约束优化问题

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

$$\Downarrow$$

$$\min t \qquad \text{s.t.} \qquad t \geq x, \qquad t \geq x^2.$$

第一节　可行方向

在第二章我们提到,约束问题问题的最优性条件有四个,为导出这些条件,我们需做一些准备工作

首先,我们需介绍与约束条件有关的可行方向.

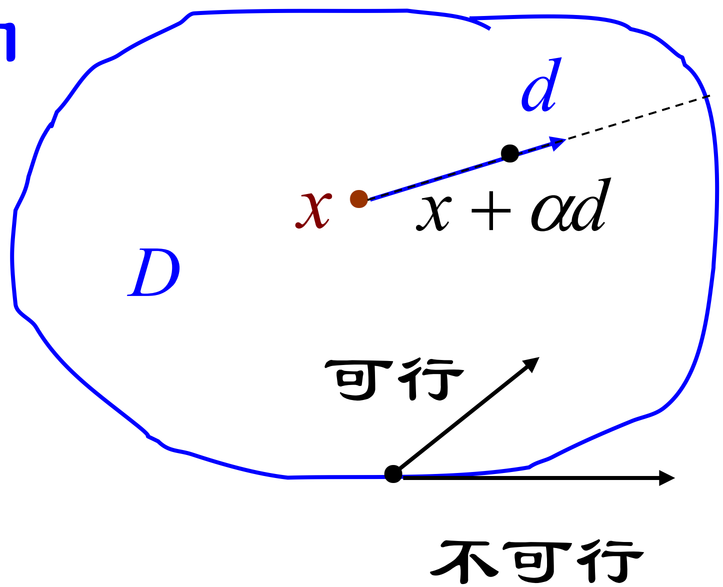
定义9.1.1 设$x \in D, d \in R^n$.若存在数$\delta > 0$,使得

$$x + \alpha d \in D, \forall \alpha \in (0, \delta],$$

则称d是D在x处的一个可行方向.

记x处所有可行方向的集合为$\mathrm{FD}(x, D)$

若记x处函数f 的所有下降方向集合为$\mathrm{GD}(x)$



容易看出,如果x^*是(9.1)的最优解, 则在该点不存在既下降又可行的方向, 即

$$\mathrm{GD}(x^*) \bigcap \mathrm{FD}(x^*, D) = \varnothing$$

该条件称为几何最优性条件

我们的目的是将几何最优性条件转化为便于计算的代数最优性条件. 这要求$\mathrm{GD}(x)$和$\mathrm{FD}(x,D)$的代数条件

$\forall x \in D$, 对于$\mathrm{GD}(x)$, 我们有

$$\mathrm{GD}(x) \supseteq \{d \in R^n \mid \nabla f(x)^T d < 0\}$$

但可行方向集$\mathrm{FD}(x,D)$的计算是困难的

事实上： 对于$d \in R^n$, 类似于$\mathrm{GD}(x)$的计算

等式 $h_j(x)=0$： $\nabla h_j(x)^T d > 0$ 不是可行方向

$\nabla h_j(x)^T d < 0$ 不是可行方向

$\nabla h_j(x)^T d = 0$ 包含可行方向

不等式 $g_i(x)=0$: $\nabla g_i(x)^T d \geq 0$ 是可行方向

不等式 $g_i(x)>0$： $\forall 0 \neq d \in R^n$ 是可行方向

由上面分析可知：$\forall d \in \mathrm{FD}(x, D)$, 则有

$$\begin{cases} \nabla h_j(x)^T d = 0, & \forall j \in E \\ \nabla g_i(x)^T d \geq 0, & \forall i \in I \text{ 且 } g_i(x) = 0 \end{cases}$$

但反之不一定成立.

为方便起见, 记

$$\mathrm{LFD}(x, D) = \left\{ d \in R^n \;\middle|\; \begin{array}{ll} \nabla h_j(x)^T d = 0, & \forall j \in E \\ \nabla g_i(x)^T d \geq 0, & \forall i \in I \text{ 且 } g_i(x) = 0 \end{array} \right\}$$

显然 $\mathrm{FD}(x, D) \subseteq \mathrm{LD}(x, D)$

为了更好地描述$\mathrm{FD}(x, D)$, 我们去掉$\mathrm{LD}(x, D)$中的某些"多余"的向量, 而得到一个新的方向集合. 具体如下：

定义9.1.2 设$x \in D, d \in R^n$. 若存在向量序列$\{d_k\}$和正数序列$\{\delta_k\}$, 使得

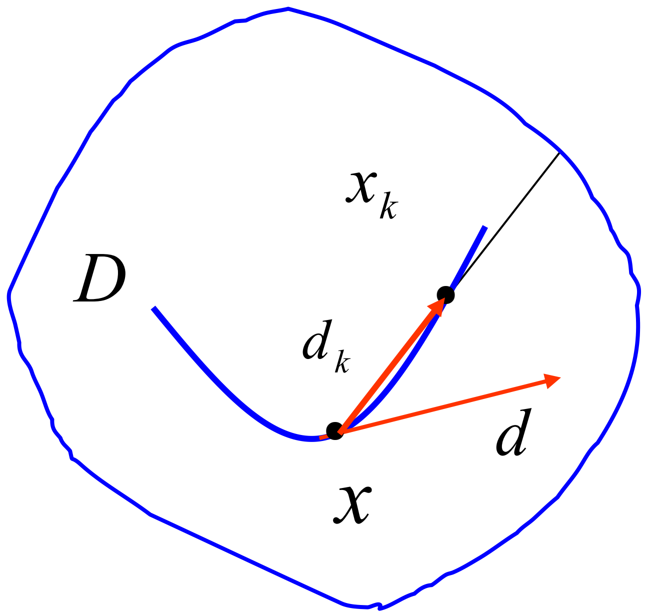
$$x + \delta_k d_k \in D, \quad \forall k$$

且 $\lim\limits_{k\to\infty} d_k = d$ 和 $\lim\limits_{k\to\infty} \delta_k = 0$

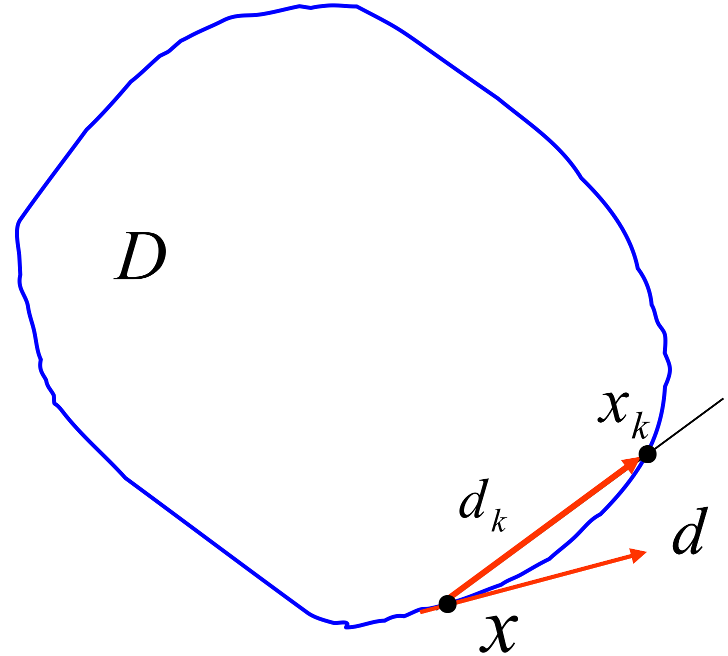
则称d是D在x处的一个序列可行方向. D在x处的所有序列可行方向的集合记作$\mathrm{SFD}(x, D)$.

令 $x_k = x + \delta_k d_k$, 由定义9.1.2知, $\{x_k\} \subset D$.

为理解序列可行方向, 我们来看看它的几何解释:



(a) 点x在D内部



(b) 点x在D的边界上

序列可行方向实际上就是可行方向

序列可行方向包含可行方向和边界的切线方向

显然，$\quad \mathrm{FD}(x, D) \subseteq \mathrm{SFD}(x, D) \quad$ (只需取$d_k = d$)

可行方向必是序列可行方向，但反之不然.

Tangent cone(切锥)

Definition 12.2.

The vector d *is said to be a* tangent (*or* tangent vector) *to* Ω *at a point* x *if there are a feasible sequence* $\{z_k\}$ *approaching* x *and a sequence of positive scalars* $\{t_k\}$ *with* $t_k \to 0$ *such that*

$$\lim_{k\to\infty} \frac{z_k - x}{t_k} = d. \tag{12.29}$$

The set of all tangents to Ω *at* x^* *is called the* tangent cone *and is denoted by* $T_\Omega(x^*)$.

定义9.1.2 设$x \in D, d \in R^n$. 若存在向量序列$\{d_k\}$和正数序列$\{\delta_k\}$，使得

$$x + \delta_k d_k \in D, \quad \forall k$$

且 $\lim_{k\to\infty} d_k = d$ 和 $\lim_{k\to\infty} \delta_k = 0$

则称d是D在x处的一个序列可行方向. D在x处的所有序列可行方向的集合记作$\mathrm{SFD}(x, D)$.

定理9.1.1 设$x^* \in D$是问题(9.1)的一个局部最优解, 则

$$\nabla f(x^*)^T d \ge 0, \quad \forall d \in \mathrm{SFD}(x^*, D)$$ 必要条件

证明： 因为x^*是局部最优解, 由定义, 必存在邻域$N(x^*)$, 使得

$$f(x) \ge f(x^*), \quad \forall x \in N(x^*)$$

另一方面, $\forall d \in \mathrm{SFD}(x^*, D)$, 存在可行点序列$\{x_k\}$满足

$$x_k = x^* + \delta_k d_k \to x^*$$

其中$\delta_k \to 0, d_k \to d.$ 所以, 当k充分大时, $x_k \in N(x^*)$.

故 $$f(x^*) \le f(x_k) = f(x^* + \delta_k d_k)$$
$$= f(x^*) + \delta_k \nabla f(x^*)^T d_k + o(\| \delta_k d_k \|)$$

在上式两端除以δ_k, 然后令$\delta_k \to 0$, 取极限即可得

$$\nabla f(x^*)^{\mathrm{T}} d \ge 0$$

定理表明最优解处的任何序列可行方向不可能是目标函数在该点处的下降方向.

类似于无约束问题的极值条件，进一步我们有关于SFD的最优解的充分条件：

定理9.1.2 设$x^* \in D$且满足

$$\nabla f(x^*)^T d > 0, \quad \forall 0 \neq d \in \mathrm{SFD}(x^*, D)$$

则x^*问题(9.1)的一个严格局部最优解．

证明：构造性的反证法

若x^*不是问题(9.1)的最优解，则必定存在序列$\{x_k\} \subset D$，使得

$$f(x^*) \geq f(x_k) \text{ 且 } x_k \to x^* \quad (x_k \neq x^*)$$

令 $d_k = \dfrac{x_k - x^*}{\| x_k - x^* \|}, \delta_k = \| x_k - x^* \|$. 则$x_k = x^* + \delta_k d_k$且$\delta_k \to 0$.

由于序列$\{d_k\}$有界，必存在收敛的子列. 不妨设$d_k \to d \neq 0$.

由$\mathrm{SFD}(x^*, D)$的定义可知，所构造的向量$d \in \mathrm{SFD}(x^*, D)$.

然而,由

$$f(x^*) \geq f(x_k) = f(x^* + \delta_k d_k)$$
$$= f(x^*) + \delta_k \nabla f(x^*)^T d_k + o(\| \delta_k d_k \|)$$

在上式两端除以δ_k,然后令$\delta_k \to 0$,取极限即可得

$$\nabla f(x^*)^{\mathrm{T}} d \leq 0$$

这与定理假设矛盾.因此x^*是问题(9.1)的一个严格局部最优解.　　证毕

定理表明:若可行点x^*处的所有序列可行方向都是目标函数 f 在该点处的上升方向,则x^*必定是严格局部最优解.

我们注意到：由于序列可行方向集$\mathrm{SFD}(x^*, D)$ 没有便于计算的公式,上面两个定理给出的最优解的判别条件仅具有理论意义,并没有实际的应用价值.

为将上面给出的最优解的判别条件代数化,我们需研究可行方向集的代数表示.

策略：将$\mathrm{SFD}(x^*, D)$放大到某一具有代数表示式的方向集－从前面介绍的$\mathrm{LD}(x, D)$（具有代数表示）得到启示：

将约束条件线性化而得到某种可行方向.

$$\Downarrow$$

线性化可行方向

在前面,我们已经看到：$\forall x \in D$,在x处的可行方向d的要求与约束有关,具体如下：

对等式 $h_j(x) = 0, \forall j \in E$：

要求 $\nabla h_j(x)^{\mathrm{T}} d = 0$

对不等式：$g_i(x) \geq 0, \forall j \in I$,分两种情形：

当$g_i(x) = 0$时，要求$\nabla g_i(x)^T d \geq 0$

当$g_i(x) > 0$时，则对$d \neq 0$无要求

基于上面的分析,我们引入如下的定义：

定义　对 $x \in D$，记索引集合

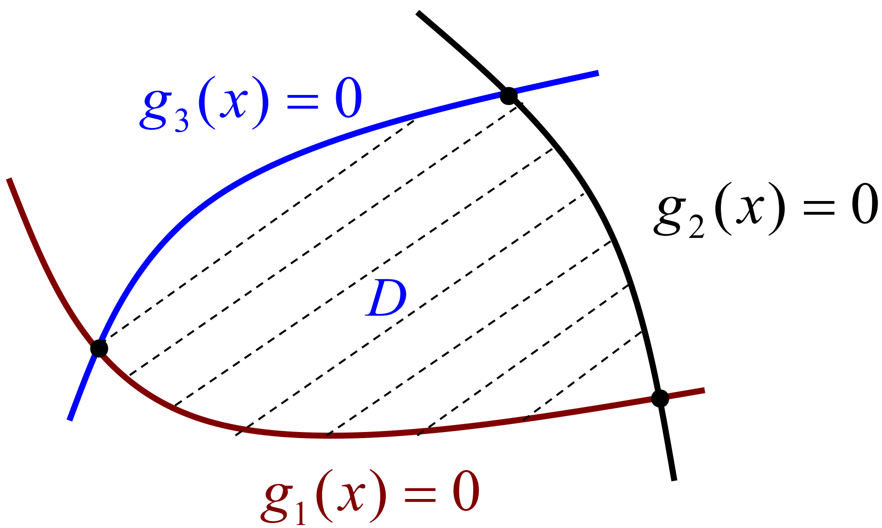
$$I(x)=\{i \in I \mid g_i(x)=0\},\ A(x)=E \bigcup I(x)$$

称集合$A(x)$为可行点x处的有效集或积极集.

若$i \in I(x)$ 或 $j \in E$，称相应的约束为x处的有效约束，即有

$$g_i(x)=0 \text{ 或 } h_i(x)=0$$

其它约束称为x处的非有效约束或无效约束.



如图：观察不同点处的有效约束

D内的点没有有效约束

仅边界上的点存在有效约束

总结：x处的有效约束是指该约束对该点处的可行方向起限制作用，反之无效约束即对该点处的可行方向没有任何影响.不同的可行点一般有不同的有效约束.

所有存在有效约束的点构成可行域的边界.

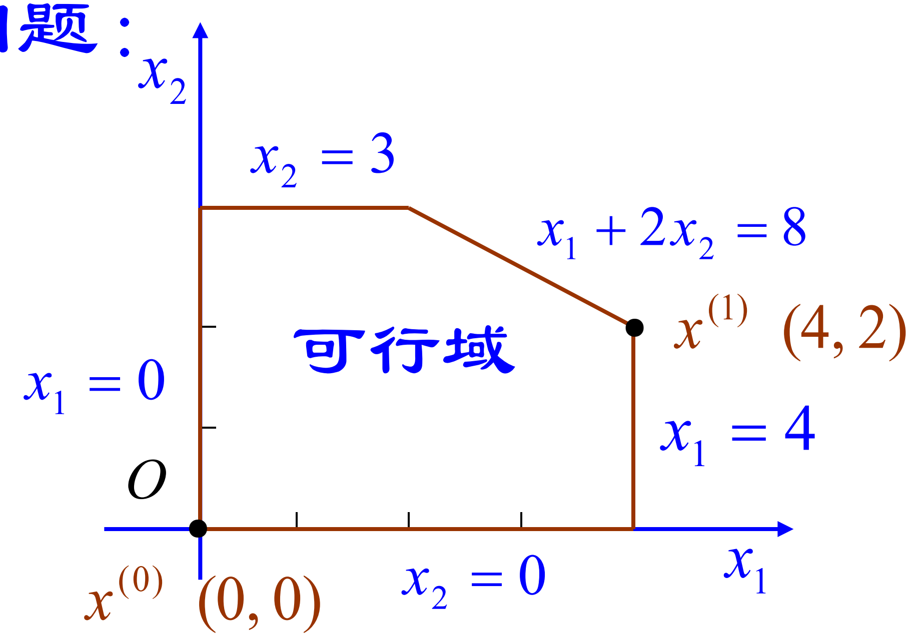
例9.1.1 考察如下约束问题：

$$\min \quad f(x) = x_1^2 + x_2^2$$

$$s.t. \quad x_1 + 2x_2 \le 8$$

$$0 \le x_1 \le 4$$

$$0 \le x_2 \le 3$$



标准形式

$$g_1(x) = 8 - x_1 - 2x_2 \ge 0$$

$$g_2(x) = x_1 \ge 0$$

$$g_3(x) = 4 - x_1 \ge 0$$

$$g_4(x) = x_2 \ge 0$$

$$g_5(x) = 3 - x_2 \ge 0$$

$$A(x^{(0)}) = \{2, 4\}$$

$$A(x^{(1)}) = \{1, 3\}$$

利用有效集,我们给出下面线性化可行方向的定义:

定义9.1.3 设$x \in D$, 集合

$$\mathrm{LFD}(x,D)=\left\{d \in R^n \;\middle|\; \begin{array}{l} d^T \nabla g_i(x) \geq 0, \forall i \in I(x) \\ d^T \nabla h_j(x) = 0, \forall j \in E \end{array}\right\}$$

中的向量d称为D在x处的线性化可行方向.

线性化可行方向只与有效集有关,且具有线性表达式,

便于计算.而且下面的命题成立:

命题 设 $x \in D$,则

$$\mathrm{FD}(x,D) \subseteq \mathrm{SFD}(x,D) \subseteq \mathrm{LFD}(x,D)$$

定理9.1.1 设$x^*\in D$是问题(9.1)的一个局部最优解, 则

$$\nabla f(x^*)^T d \ge 0, \quad \forall d \in \mathrm{SFD}(x^*, D)$$

证明：只需证第二个关系. $\forall d\in \mathrm{SFD}(x,D)$, 则存在

$d_k\to d, \delta_k\to 0$, 使得$x+\delta_k d_k\in D$. 故

怎么办？

$\forall i\in I(x)$, 由于$g_i(x)=0$, 因此

不用急啊

$$g_i(x+\delta_k d_k) = \delta_k \nabla g_i(x)^T d_k + o(\|\delta_k d_k\|) \ge 0$$

$\forall j\in E$, 由于$h_j(x)=0$, 因此

我们下面来研究在什么条件下：

$$h_j(x+\delta_k d_k) = \delta_k \nabla h_j(x)^T d_k + o(\|\delta_k d_k\|) = 0$$

$$\mathrm{SFD}(x,D) = \mathrm{LFD}(x,D)$$

在上面两式的两端同时处以δ_k, 然后令$k\to\infty$, 即有

$$\nabla g_i(x)^T d \ge 0 \text{ 和 } \nabla h_j(x)^T d = 0, \text{即} d\in \mathrm{LFD}(x,D)$$

注意：尽管 LFD(x,D)具有代数表示,但上面的命题表明 SFD(x,D)是LFD(x,D)的一个子集,因此还不能用 LFD(x,D)替换定理 9.1.1 中的SFD(x,D)

定理9.1.3 若$g_i(x), i \in I, h_j(x), j \in E$ 都是线性函数,则

$$\mathrm{FD}(x, D) = \mathrm{SFD}(x, D) = \mathrm{LFD}(x, D)$$

证明： $\forall x \in D$,由前面的命题知,现在我们只需证

$$\mathrm{LFD}(x, D) \subseteq \mathrm{FD}(x, D).$$

令 $$g_i(x) = a_i^T x + b_i, \quad i \in I$$

$$h_j(x) = a_j^T x + b_j, \quad j \in E$$

对$\forall d \in \mathrm{LFD}(x,D)$,则有$d$ 满足

$$\begin{cases} a_j^T d = 0, j \in E \\ a_i^T d \geq 0, i \in I(x) \end{cases}$$

分三种情形讨论：

(1) $\forall j \in E$, 由于$h_j(x) = a_j^T x + +b_j = 0$,则对任意的$\alpha \geq 0$,

$$h_j(x + \alpha d) = a_j^T (x + \alpha d) + b_j = a_j^T x + b_j + \alpha a_j^T d = 0,$$

(2) $\forall i \in I(x)$, 由于$g_i(x) = a_i^T x + b_i = 0$,则对任意的$\alpha \geq 0$,

$$g_i(x+\alpha d) = a_i^T(x+\alpha d) + b_i = a_i^T x + b_i + \alpha a_i^T d \geq 0,$$

(3) $\forall i \in I \setminus I(x)$, 由于$g_i(x) = a_i^T x + b_i > 0$,即$-a_i^T x - b_i < 0$

则要使

$$g_i(x+\alpha d) = a_i^T(x+\alpha d) + b_i = a_i^T x + b_i + \alpha a_i^T d \geq 0,$$

若$a_i^T d \geq 0$,显然上式对任意$\alpha \geq 0$成立,

若$a_i^T d < 0$,则要求$\alpha \leq \dfrac{-a_i^T x - b_i}{a_i^T d}$,此时我们令

$$\overline{\alpha} = \min\left\{ \frac{-a_i^T x - b_i}{a_i^T d} \,\middle|\, i \in I \setminus I(x)\ a_i^T d < 0 \right\}$$

则对任意的$i \in I \setminus I(x)$,我们有

$$a_i^T x + b_i + \alpha a_i^T d \geq 0, \forall \alpha \in (0, \ \overline{\alpha})$$

综上所述,我们证明了: $d \in \mathrm{FD}(x, D)$.　　　　证毕

除约束函数是线性函数外,人们更多研究的是对一般的非线性函数在什么情况下满足条件:

$$\mathrm{SFD}(x,D)=\mathrm{LFD}(x,D)$$

上面的条件首先由Abadie (1972年)提出,称之为Abadie约束品性(Constaint Qualificat ion),简称为ACQ约束品性对研究约束问题的最优性条件是非常重要的,ACQ是最基本的约束品性.除此之外,人们研究了一些较强的约束品性.

定义9.1.4 设$x\in D$,若向量组

$$\{\nabla g_i(x),\nabla h_j(x),i\in I(x),j\in E\}$$

线性无关,则称在x处线性无关约束品性成立,简称为在x处LICQ (Libear independence Constraint Qualification)成立.

引理9.1.1 设$x \in D$，若向量组

$$\{\nabla g_i(x), \nabla h_j(x), i \in I(x), j \in E\}$$

线性无关，则

$$\mathrm{SFD}(x, D) = \mathrm{LFD}(x, D)$$

证明：有点难，还是省略吧！感兴趣的同学自己琢磨.

由引理9.1.1知：$\mathrm{LICQ} \Rightarrow \mathrm{ACQ}$

现在我们再回头看看定理9.1.1，

定理9.1.1 设$x^* \in D$是问题(9.1)的一个局部最优解，则

$$\nabla f(x^*)^T d \geq 0, \quad \forall d \in \mathrm{SFD}(x^*, D)$$

ACQ成立

$\mathrm{LFD}(x, D)$

发现了什么？

定理9.1.1 设$x^* \in D$是问题(9.1)的一个局部最优解，则

$$\nabla f(x^*)^T d \geq 0, \quad \forall d \in \mathrm{SFD}(x^*, D)$$

如果我们将$\mathrm{SFD}(x, D)$换成$\mathrm{LFD}(x, D)$，那么定理9.1.1可以等价地描述为：

定理9.1.1 设$x^* \in D$是问题(9.1)的一个局部最优解，则关于向量$d \in R^n$的线性系统：

$$\begin{cases} \nabla f(x^*)^T d < 0 \\ \nabla g_i(x^*)^T d \geq 0, \ i \in I(x^*) \\ \nabla h_j(x^*)^T d = 0, \ j \in E \end{cases}$$

无解.

Farkas 定理还有谁记得？说的什么？

思考：上面的线性系统无解又等价于什么？

Farkas定理的推论：设A为$m\times n$矩阵,B为$p\times n$矩阵�$c \in R^n$，则

$$Ax \le 0, Bx = 0, c^T x > 0 \text{无解} \Leftrightarrow A^T y + B^T z = c, y \ge 0 \text{有解}$$

引理9.1.2 不等式 $\nabla f(x)^T d \ge 0$对所有 $d \in \mathrm{LFD}(x, D)$成立，则存在 $\lambda_i \ge 0, i \in I(x); \mu_j, j \in E$,使得

$$\nabla f(x) - \sum_{i \in I(x)} \lambda_i \nabla g_i(x) - \sum_{j \in E} \mu_{\mathrm{j}} \nabla h_j(x) = 0$$

该引理可以由Farkas定理的推论直接写出

第二节 约束问题的最优性条件

1、一阶必要条件

定义函数 $L: R_{n+m} \to R$:

$$L(x,\lambda,\mu) = f(x) - \lambda^T g_I(x) - \mu^T h_E(x)$$
$$= f(x) - \sum_{i\in I} \lambda_i g_i(x) - \sum_{j\in E} \mu_j h_j(x)$$

该函数称为问题(9.1)的Lagrange函数，其中 $\lambda \in R^{m_1}$，$\mu \in R^{m\text{-}m_1}$ 称为Lgrange乘子.

$$\nabla_x L(x,\lambda,\mu) = \nabla f(x) - \sum_{i\in I} \lambda_i \nabla g_i(x) - \sum_{j\in E} \mu_j \nabla h_j(x)$$

$$\nabla_x^2 L(x,\lambda,\mu) = \nabla^2 f(x) - \sum_{i\in I} \lambda_i \nabla^2 g_i(x) - \sum_{j\in E} \mu_j \nabla^2 h_j(x)$$

定理9.1.1的改进版：

定理 9.2.1 设$x^* \in D$是问题(9.1)的一个局部最优解，如果

$$\mathrm{SFD}(x^*, D) = \mathrm{LFD}(x^*, D) \tag{9.6}$$

则存在 Lagrange 乘子向量：$\lambda^* \in R^{m_1}, \mu^* \in R^{m-m_1}$

使得

$$\begin{cases} \nabla_x L(x^*, \lambda^*, \mu^*) = 0 \\ h_j(x^*) = 0, j \in E \\ g_i(x^*) \geq 0, \lambda_i^* \geq 0, \underline{\lambda_i^* g_i(x^*) = 0, i \in I} \end{cases} \tag{9.7}$$

互补松弛条件

这里，

$$\nabla_x L(x^*, \lambda^*, \mu^*) = \nabla f(x^*) - \sum_{i \in I} \lambda_i^* \nabla g_i(x^*) - \sum_{j \in E} \mu_j^* \nabla h_j(x^*)$$

(9.7)称为问题(9.1)的一阶必要条件－K-K-T条件

满足(9.7)的点x^*称为问题(9.1) 的一个KKT点.

由于(9.7)中无需已知$I(x^*)$, (9.7)除能用于最优解的判别,
而且能用来计算KKT点 — 可能的最优解.

约束问题的KKT点 类似于无约束问题的驻点.

思考

若函数可导,无约束问题的极值点一定是驻点,

请问约束问题的局部最优解一定是KKT点吗？？

不一定啦

？？？

$$\mathrm{SFD}(x^*, D) = \mathrm{LFD}(x^*, D)$$

最优解 x^* ⟶ KKT点 x^*

一般的KKT条件(KKT系统)

$$\begin{cases} \nabla_x L(x, \lambda, \mu) = 0 \\ h_j(x) = 0, j \in E \\ g_i(x) \geq 0, \lambda_i \geq 0, \lambda_i g_i(x) = 0, i \in I \end{cases} \quad (9.7b)$$

或向量形式：

$$H(x, \lambda, \mu) = \begin{pmatrix} \nabla_x L(x, \lambda, \mu) \\ h_E(x) \\ \min\{\lambda_I, g_I(x)\} \end{pmatrix} = 0 \quad (9.7c)$$

其中向量函数的极小值按分量进行比较.

利用KKT条件, 我们可以计算约束问题的KKT点

一般来说，KKT点的判别用定理9.1.1改进版的条件，即

$$\begin{cases}\nabla f(x^*)-\sum_{i\in I(x^*)}\lambda_i^*\nabla g_i(x^*)-\sum_{j\in E}\mu_j^*\nabla h_j(x^*)=0\\ \lambda_i^*\geq 0, i\in I(x^*)\end{cases}$$

而KKT点的计算则采用KKT系统(9.7b)

例 9.2.1 已知 $\bar{x}=(3,1)^{\mathrm{T}}$是下列问题的最优解

$$\begin{cases}\min & f(x)=(x_1-7)^2+(x_2-3)^2\\ \text{s.t.} & g_1(x)=10-x_1^2-x_2^2\geq 0\\ & g_2(x)=x_1-x_2-2\leq 0\\ & g_3(x)=x_2\geq 0\end{cases}$$

试问 $\bar{x}$ 是KKT点吗？

解：　经验证：$I(\bar{x}) = \{1,\ 2\}$，计算目标函数及有效约束在点$\bar{x}$处的梯度为

$$\nabla f(\bar{x}) = \begin{pmatrix} -8 \\ -4 \end{pmatrix}, \nabla g_1(x) = \begin{pmatrix} -6 \\ -2 \end{pmatrix}, \nabla g_2(x) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

显然，有效约束的梯度$\nabla g_1(\bar{x}), \nabla g_2(\bar{x})$线性无关，从而

$$\mathrm{SFD}(\bar{x}.D) = \mathrm{LFD}(\bar{x}, D)$$

因此，最优解一定是KKT点．　事实上，令

$$\begin{pmatrix} -8 \\ -4 \end{pmatrix} - \lambda_1 \begin{pmatrix} -6 \\ -2 \end{pmatrix} - \lambda_2 \begin{pmatrix} -1 \\ -1 \end{pmatrix} =$$

$$\begin{cases} \min & f(x) = (x_1 - 7)^2 + (x_2 - 3)^2 \\ \text{s.t.} & g_1(x) = 10 - x_1^2 - x_2^2 \geq 0 \\ & g_2(x) = x_1 - x_2 - 2 \leq 0 \\ & g_3(x) = x_2 \geq 0 \end{cases}$$

即得
$$\begin{cases} 6\lambda_1 + \lambda_2 - 8 = 0 \\ 2\lambda_1 + \lambda_2 - 4 = 0 \end{cases}$$

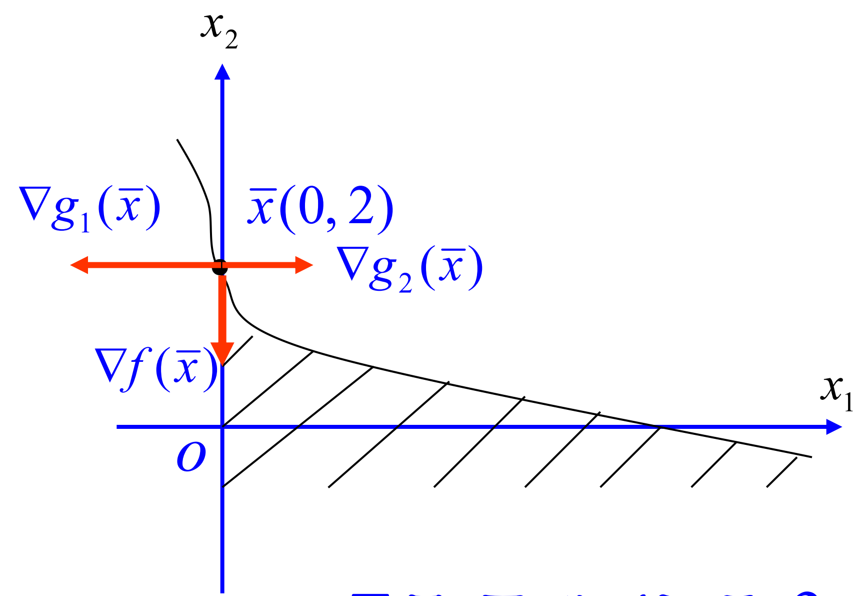
则方程组存在非负解，$\lambda_1 = 1, \lambda_2 = 2$，即$\bar{x}$是KKT点．

例9.2.2 已知约束问题

$$\min \quad f(x) = -x_2$$
$$s.t. \quad g_1(x) = -2x_1 + (2 - x_2)^3 \geq 0$$
$$g_2(x) = x_1 \geq 0$$

请问$\bar{x} = (0, 2)^T$是KKT点吗，

是最优解吗？



$\bar{x}$是最优解吗？

解：$I(\bar{x}) = \{1,\ 2\}$，即两个约束都是有效约束.

$$\nabla f(\bar{x}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \nabla g_1(\bar{x}) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \quad \nabla g_2(\bar{x}) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

是

显然，$\nabla g_1(\bar{x}), \nabla g_2(\bar{x})$线性相关. 令

$$\begin{pmatrix} 0 \\ -1 \end{pmatrix} - \lambda_1 \begin{pmatrix} -2 \\ 0 \end{pmatrix} - \lambda_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0$$

$\nabla f(\bar{x}) \perp \nabla g_1(\bar{x}), \nabla f(\bar{x}) \perp \nabla g_2(\bar{x})$，

$\nabla f(\bar{x})$不能用$\nabla g_1(\bar{x}), \nabla g_2(\bar{x})$的正线性组合表示

方程组无解，见图，故$\bar{x}$不是KKT点

例9.2.3 求下列约束问题的KKT点

$$
\begin{aligned}
\min \quad & f(x) = x_1 \\
s.t. \quad & g(x) = -(x_1 - 4)^2 - x_2^2 + 16 \geq 0 \\
& h(x) = x_1^2 + (x_2 - 2)^2 - 4 = 0
\end{aligned}
$$

解：该问题的KKT条件为：

$$
\begin{cases}
\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 2(x_1 - 4) \\ 2x_2 \end{pmatrix} - \mu \begin{pmatrix} 2x_1 \\ 2(x_2 - 2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
-(x_1 - 4)^2 - 2x_2^2 + 16 \geq 0 \\
x_1^2 + (x_2 - 2)^2 - 4 = 0 \\
\lambda \geq 0, \lambda[-(x_1 - 4)^2 - x_2^2 + 16] = 0
\end{cases}
$$

解上面的KKT系统，得KKT点为：

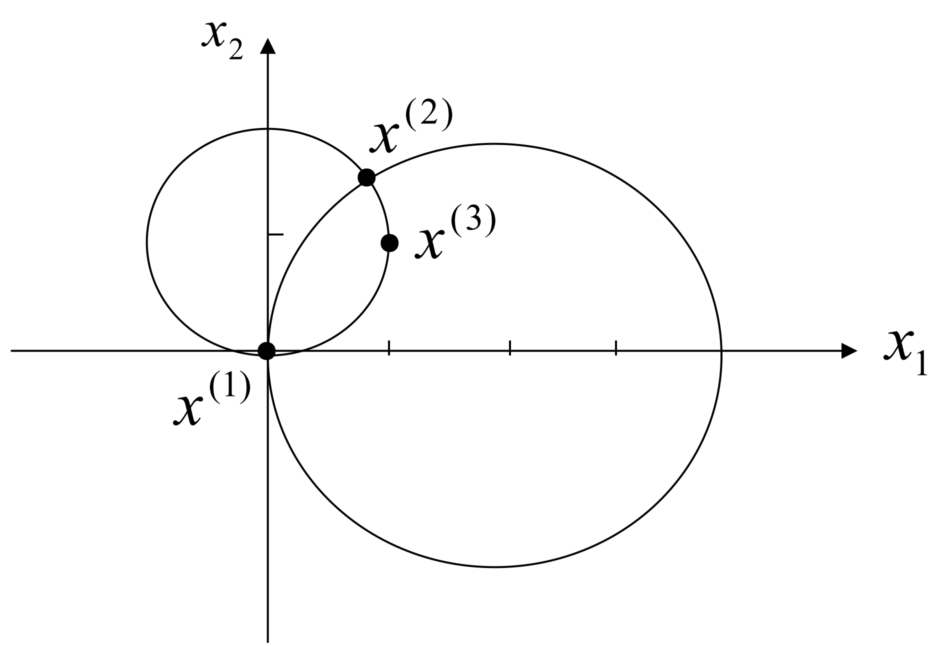
$$\begin{pmatrix} x^{(1)} \\ \lambda_1 \\ \mu_1 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ 1/8 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} x^{(2)} \\ \lambda_2 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 8/5 \\ 16/5 \end{pmatrix} \\ 3/40 \\ 1/5 \end{pmatrix}, \quad \begin{pmatrix} x^{(3)} \\ \lambda_3 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} \\ 0 \\ 1/4 \end{pmatrix}$$

由右图可知，

$x^{(1)}$是最优解，

$x^{(2)}$? 是最优解

$x^{(3)}$不是最优解



与无约束问题的一阶必要条件类似，对于凸规划，KKT条件也是充分条件

定理9.2.4 设 f 是凸函数，$g_i(x), i\in I$是凹函数，$h_j(x), j\in E$ 是线性函数. 若在x^*处满足KKT条件(9.7), 则x^*是问题(9.1)的全局最优解.

证明：$\forall x\in D$，由已知条件，我们有

$$f(x)-f(x^*)\ge \nabla f(x^*)^{\mathrm{T}}(x-x^*)$$

$$0\le g_i(x)=g_i(x)-g_i(x^*)\le \nabla g_i(x^*)^{\mathrm{T}}(x-x^*), i\in I(x^*)$$

$$0=h_j(x)-h_j(x^*)=\nabla h_j(x^*)^{\mathrm{T}}(x-x^*), j\in E$$

从而由KKT条件（9.7）得

$$f(x)-f(x^*)\ge \nabla f(x^*)^{\mathrm{T}}(x-x^*)$$

$$=\sum_{i\in I(x^*)}\lambda_i^*\nabla g_i(x^*)^{\mathrm{T}}(x-x^*)+\sum_{j\in E}\mu_j^*\nabla h_j(x^*)^{\mathrm{T}}(x-x^*)\ge 0$$

所以x^*是全局最优解.

2、二阶条件

设$x^* \in D$是(9.1)的一个局部最优解. 为介绍其二阶条件，

我们首先看看前面介绍的定理9.1.1和定理9.1.2

定理9.1.1 设$x^* \in D$是问题(9.1)的一个局部最优解，则

$$\nabla f(x^*)^T d \geq 0, \quad \forall d \in \mathrm{SFD}(x^*, D)$$

定理9.1.2 设$x^* \in D$且满足

$$\nabla f(x^*)^T d > 0, \quad \forall 0 \neq d \in \mathrm{SFD}(x^*, D)$$

则x^*问题(9.1)的一个严格局部最优解．

如果存在$0 \neq d \in \mathrm{SFD}(x^*, D)$，使得

有什么想法？

提示：

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

会怎么样？

定理9.1.2失效

研究二阶条件的目的：当

$$\nabla f(x^*)^T d = 0, \quad 0 \neq d \in \mathrm{SFD}(x^*, D) \qquad (9.**)$$

时,最优解的判别条件.

假定在x^*处$\mathrm{SFD}(x^*, D) = \mathrm{LFD}(x^*, D)$成立,则$x^*$是一个KKT点,即存在Lagrange 乘子$\lambda_i^* \geq 0, i \in I, \mu_j^*, j \in E$满足KKT条件(9.7). 则

$$(9.**) \iff 0 \neq d \in \mathrm{LFD}(x^*, D) \text{且}$$

$$0 = \nabla f(x^*)^T d = \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*)^{\mathrm{T}} d + \sum_{j \in E} \mu_j^* \nabla h_j(x^*)^T d,$$

$$= \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*)^{\mathrm{T}} d$$

$$\iff \lambda_i^* d^T \nabla g_i(x^*) = 0, i \in I(x^*), \forall d \in \mathrm{LFD}(x^*, D)$$

记 $z^* = (x^*, \lambda^*, \mu^*)$, 定义集合：

$$S(z^*) = \left\{ d \in \mathrm{LFD}(x^*, D) \,\middle|\, \lambda_i^* d^T \nabla g_i(x^*) = 0, i \in I(x^*) \right\}$$

称 $d \in S(z^*)$ 是 x^* 处的线性化零约束方向.

注意： 当(9.1)只有等式约束时：$S(z^*) = \mathrm{LFD}(x^*, D)$

定理9.2.2 设 $f(x), g_i(x), h_j(x) (i \in I, j \in E)$ 二次连续可微, x^* 是(9.1)的一个局部最优解, 且在该点处LICQ成立. 则存在Lgrange乘子 $\lambda^* \in R^{m_1}, \mu^* \in R^{m-m_1}$ 满足(9.7) 且

$$d^T \nabla_x^2 L(z^*) d \geq 0, \quad \forall d \in S(z^*) \qquad (9.8)$$

注意：

如果将定理中条件 LICQ 改成 ACQ,

定理结论不一定成立

定理9.2.3 设 $f(x), g_i(x), h_j(x)(i \in I, j \in E)$二次连续可微. 设$x^* \in D$且存在乘子$\lambda^* \in R^{m_1}, \mu^* \in R^{m-m_1}$满足(9.7). 如果

$$d^T \nabla_x^2 L(z^*) d > 0, \quad \forall 0 \neq d \in \mathrm{S}(z^*) \tag{9.9}$$

成立, 则x^*是问题(9.1)的一个严格局部最优解.

证明：采用反证法：假定x^*不是(9.1)的严格局部最优解,则存在序列$\{x_k\} \subset D$且$x_k \to x^*$,使得 $f(x_k) \le f(x^*)$.

不妨设

$$d_k = \frac{x_k - x^*}{\| x_k - x^* \|} \to d \neq 0$$

由于 $\delta_k = x_k - x^* \to 0$, 且$x_k = x^* + \delta_k d_k \in D$.

故 $d \in \mathrm{SFD}(x^*, D) \subseteq \mathrm{LFD}(x^*, D)$.

由 $f(x^*) \geq f(x_k) = f(x^*) + \delta_k \nabla f(x^*)^T d_k + o(\delta_k)$

推出： $d^T \nabla f(x^*) \leq 0$

另一方面，由KKT条件(9.7)知，

$$d^T \nabla f(x^*) = \sum_{i \in I(x^*)} \lambda_i^* d^T \nabla g_i(x^*) + \sum_{j \in E} \mu_j^* d^T \nabla h_j(x^*) \geq 0$$

所以

$$d^T \nabla f(x^*) = 0, \quad \lambda_i^* d^T \nabla g_i(x^*) = 0, \forall i \in I(x^*),$$

即 $d \in S(z^*)$.

又 $L(x_k, \lambda^*, \mu^*) = f(x_k) - \sum_{i \in I} \lambda_i^* g_i(x_k) - \sum_{j \in E} \mu_j^* h_j(x_k)$

$$= f(x_k) - \sum_{i \in I(x^*)} \lambda_i^* g_i(x_k) \leq f(x_k)$$

以及

$$L(x^*, \lambda^*, \mu^*) = f(x^*) - \sum_{i \in I} \lambda_i^* g_i(x^*) - \sum_{j \in E} \mu_j^* h_j(x^*) = f(x^*)$$

所以

$$L(x^*,\lambda^*,\mu^*) \geq L(x_k,\lambda^*,\mu^*)$$

$$= L(x^*,\lambda^*,\mu^*) + \delta_k \nabla_x L(x^*,\lambda^*,\mu^*)^{\mathrm{T}} d_{\mathrm{k}}$$

$$+ \frac{1}{2}\delta_k^2 d_k^T \nabla_{xx}^2 L(x^*,\lambda^*,\mu^*) d_{\mathrm{k}} + o(\|\delta_k\|^2)$$

$$= L(x^*,\lambda^*,\mu^*) + \frac{1}{2}\delta_k^2 d_k^T \nabla_{xx}^2 L(x^*,\lambda^*,\mu^*) d_{\mathrm{k}} + o(\|\delta_k\|^2)$$

上式两端除以δ_k^2并令$k \to \infty$, 得

$$d^T \nabla_{xx}^2 L(x^*,\lambda^*,\mu^*) d \leq 0.$$

这与定理假设矛盾.所以定理成立.

从约束问题的二阶最优性条件可以看出,我们不是从$\nabla_x^2 L(z^*)$的正定性来判断x^*是否是最优解.而是考察其在R^n的子集$S(z^*)$上正定.

例9.2.4 求下列约束问题的最优解

$$
\begin{aligned}
\min \quad & f(x) = x_1 \\
s.t. \quad & g(x) = -(x_1 - 4)^2 - x_2^2 + 16 \ge 0 \\
& h(x) = x_1^2 + (x_2 - 2)^2 - 4 = 0
\end{aligned}
$$

解：在例9.2.3中我们已计算出KKT点：

$$
\begin{pmatrix} x^{(1)} \\ \lambda_1 \\ \mu_1 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ 1/8 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} x^{(2)} \\ \lambda_2 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 8/5 \\ 16/5 \end{pmatrix} \\ 3/40 \\ 1/5 \end{pmatrix}, \quad \begin{pmatrix} x^{(3)} \\ \lambda_3 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} \\ 0 \\ 1/4 \end{pmatrix}
$$

$$\begin{pmatrix} x^{(1)} \\ \lambda_1 \\ \mu_1 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ 1/8 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} x^{(2)} \\ \lambda_2 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 8/5 \\ 16/5 \end{pmatrix} \\ 3/40 \\ 1/5 \end{pmatrix}, \quad \begin{pmatrix} x^{(3)} \\ \lambda_3 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} \\ 0 \\ 1/4 \end{pmatrix}$$

经计算得：

$$\nabla_x^2 L(z) = \begin{pmatrix} 2(\lambda - \mu) & 0 \\ 0 & 2(\lambda - \mu) \end{pmatrix}$$

在$x^{(1)}$处：

$$\nabla_x^2 L(z^{(1)}) = \begin{pmatrix} 1/4 & 0 \\ 0 & 1/4 \end{pmatrix}$$

即在$x^{(1)}$处$\nabla_x^2 L(z^{(1)})$正定，故$x^{(1)}$是最优解

在$x^{(2)}$处： $$\nabla_x^2 L(z^{(2)}) = \begin{pmatrix} -1/4 & 0 \\ 0 & -1/4 \end{pmatrix}$$

在$x^{(2)}$处$\nabla_x^2 L(z^{(2)})$负定，但$I(x^{(2)}) = \{1\}, \lambda_2 = 3/40 > 0$

$$\nabla g(x^{(2)}) = \begin{pmatrix} 24/5 \\ -32/5 \end{pmatrix}, \quad \nabla h(x^{(2)}) = \begin{pmatrix} 8/5 \\ 6/5 \end{pmatrix}$$ 线性无关

$$S(z^{(2)}) = \left\{ d \in R^2 \,\middle|\, \begin{array}{l} 24d_1 - 32d_2 = 0 \\ 8d_1 + 6d_2 = 0 \end{array} \right\} = \{(0, 0)^{\mathrm{T}}\}$$

由于 $S(z^{(2)})$中没有非零向量,因此无法判定$x^{(2)}$是最优解.

在$x^{(3)}$处 $$\nabla_x^2 L(z^{(3)}) = \begin{pmatrix} -1/2 & 0 \\ 0 & -1/2 \end{pmatrix}$$

在$x^{(3)}$处$\nabla_x^2 L(z^{(3)})$负定, $I(x^{(3)}) = \phi, \lambda_3 = 0,$

$$\nabla h(x^{(3)}) = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$$

$$S(z^{(3)}) = \left\{ d \in R^2 \middle| d_1 = 0, d_2 \in R \right\} = \{(0, d_2)^{\mathrm{T}}\} \mid d_2 \in R\}$$

对$\forall 0 \neq d \in S(z^{(3)})$, 有

$$d^T \nabla_x^2 L(z^{(3)}) d = -\frac{1}{2} d_2^2 < 0,$$

不满足二阶必要条件, 故$x^{(3)}$不是最优解.

例9.2.5 考虑下列约束问题

$$\min \quad f(x) = x_1^2 + (x_2 - 2)^2$$
$$\text{s.t.} \quad h(x) = \beta x_1^2 - x_2 = 0$$

其中$\beta \in R$,讨论取何值时,点$x^{(0)} = (0, 0)^{\mathrm{T}}$ 是最优解.

解:$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 2(x_2 - 2) \end{pmatrix}, \quad \nabla h(x) = \begin{pmatrix} 2\beta x_1 \\ -1 \end{pmatrix}$

$$\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \nabla^2 h(x) = \begin{pmatrix} 2\beta & 0 \\ 0 & 0 \end{pmatrix}$$

$$\nabla_x^2 L(x, \mu) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} - \mu \begin{pmatrix} 2\beta & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 - 2\beta\mu & 0 \\ 0 & 2 \end{pmatrix}$$

将$x^{(0)}$代入得:

$$\nabla f(x^{(0)}) = \begin{pmatrix} 0 \\ -4 \end{pmatrix}, \quad \nabla h(x^{(0)}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

令 $\nabla_x L(x^{(0)},\mu)=\nabla f(x^{(0)})-\mu\nabla h(x^{(0)})=0$, 即

$$\begin{pmatrix}0\\-4\end{pmatrix}-\mu\begin{pmatrix}0\\-1\end{pmatrix}=0,\quad \text{得}\ \mu=4$$

所以在$x^{(0)}$处：

$$\nabla_x^2 L(x^{(0)},\mu)=\begin{pmatrix}2-8\beta & 0\\0 & 2\end{pmatrix}$$

计算$S(x^{(0)},\mu)$：令

$$\nabla h(x^{(0)})^T d=0,\ (0\quad -1)\begin{pmatrix}d_1\\d_2\end{pmatrix}=-d_2=0$$

所以，　$S(x^{(0)},\mu)=\{(d_1,0)^{\mathrm{T}}\mid d_1\in R\}$

此时，$\forall d\in S(x^{(0)},\mu)$

$$d^T\nabla_x^2 L(x^{(0)},\mu)d=(d_1\quad 0)\begin{pmatrix}2-8\beta & 0\\0 & 2\end{pmatrix}\begin{pmatrix}d_1\\0\end{pmatrix}=2(1-4\beta)d_1^2$$

$\forall d \in S(x^{(0)}, \mu)$

$$d^T \nabla_x^2 L(x^{(0)}, \mu) d = (d_1 \ \ 0)\begin{pmatrix} 2-8\beta & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} d_1 \\ 0 \end{pmatrix} = 2(1-4\beta)d_1^2$$

(1) 当$\beta < \dfrac{1}{4}$时, $d^T \nabla_x^2 L(x^{(0)}, \mu) d > 0$, 故$x^{(0)}$是最优解

(2) 当$\beta > \dfrac{1}{4}$时, $d^T \nabla_x^2 L(x^{(0)}, \mu) d < 0$, 故$x^{(0)}$不是最优解

(3) 当$\beta = \dfrac{1}{4}$时, $d^T \nabla_x^2 L(x^{(0)}, \mu) d = 0$, 二阶条件失效, 消去$x_1$

原问题等价于:

$$\min \ f(x) = 4x_2 + (x_2 - 2)^2 = x_2^2 + 4$$

显然, $x^{(0)} = (0, 0)^{\mathrm{T}}$是最优解.

综上所述: 当$\beta \le \dfrac{1}{4}$时, $x^{(0)}$是最优解, $\beta > \dfrac{1}{4}$时则不是.

作 业

- 习题9
- 4，8，18