最优化理论
Optimality Theory

目 录 (CONTENT)

大连理工大学 Dalian University Of Technology

课程简介

■ 课程考核(32学时)

➢期末考试：

●考试构成： 70.0%, 百分制;

课程过程中和课后习题(期末考试)

●放假时间： 2021年1月16日-2021年2月28日

●结课时间： 2020年1月1日(周五，最后一课)

●考试时间：计划2020年1月13日(101人)

具体时间待定，看学校安排！

07

PART SEVEN

最优性条件

Optimality Condition

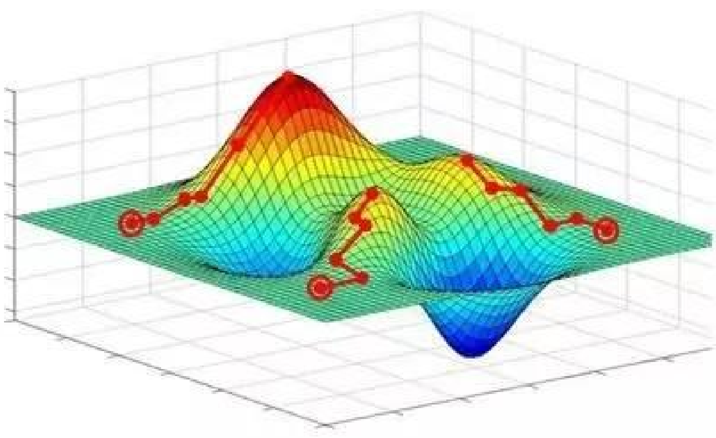
■ 基本概念

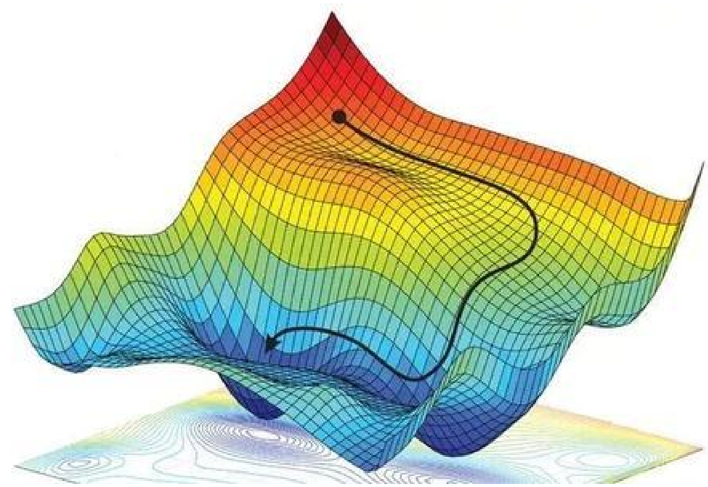
最优化问题可归结成如下数学形式：

$$\min_{x \in R^n} f(x) \qquad \text{---目标函数}$$

$$s.t. \quad g_i(x) \geq 0, i \in I$$

$$h_j(x) = 0, j \in E$$





■ 求函数的极小值

$$\phi(x)=\frac{x_1+x_2}{3+x_1^2+x_2^2+x_1x_2}$$

$$\begin{aligned}&\min \quad (x_1-3)^2+(x_2-2)^2\\&s.t. x_1^2+x_2^2\le 5,\\&x_1+2x_2=4,\\&x_1,x_2\ge 0.\end{aligned}$$

无约束问题的极值条件

定理 7.1.2-3(极小点的必要条件)设 x^* 处是问题(UNLP)的局部极小点.

(1)当 $f(x)$ 在 x^* 可微时,则梯度 $\nabla f(x^*)=0$.

(2)当 $f(x)$ 在 x^* 二次可微时,则 $\nabla f(x^*)=0$ 且 **Hessian** 矩阵 $H(x^*)$ 是半正定的

定理 7.1.4 (二阶充分条件). 假设 $f(x)$ 在 x^* 点二次可微,$\nabla f(x^*)=0$ 且 **Hessian** 矩阵 $H(x^*)$ 是正定的,则 x^* 是(UNLP)的一个(严格)局部极小点

定理 7.1.5 (充要条件). 假设 $f(x):R^n \to R$ 是可微的凸函数,则 x^* 是(UNLP)的全局最小点当且仅当 $\nabla f(x^*)=0$.

■ 无约束问题的极值条件

例7.1.1　利用极值条件解下列问题：

$$\min f(x) = (x_1^2 - 1)^2 + x_1^2 + x_2^2 - 2x_1$$

例7.1.2 利用极值条件解下列问题：

$$\min f(x) = \frac{1}{3}x_1^3 + \frac{1}{3}x_2^3 - x_2^2 - x_1$$

不等式约束的一阶最优性条件

考察非线性规划

$$\min \quad f(x)$$

$$s,t \quad g_i(x) \ge 0, \qquad i = 1,2,..,m$$

可行域S＝$\left\{x \,\middle|\, g_i(x) \ge 0, i = 1,2,..,m\right\}$

为把最优性的几何条件用代数来表示，引入起作用约束的概念。问题的约束条件在点x*∈S处有两种情形

1，$\mathbf{I}=\{i|\ g_i(\mathbf{x}^*)=0\}$——在x*处起作用约束

2，$g_i(\mathbf{x}^*)>0.\ i\notin \mathbf{I}$——在x*处不起作用约束

$\mathbf{G}_0(\mathbf{x}^*)=\{\mathbf{d}\ |\nabla g_i(\mathbf{x}^*)\mathbf{d}>0\ ,\ i\in \mathbf{I}\ \}.$

■ 不等式约束的一阶最优性条件

定理 7.2.2. (必要条件) 考虑极小化问题

$$\min \ f(x)$$

$$\text{s.t } g_i(x)\geq 0,\quad i=1,\ldots,m,\quad x\in S,$$

其中S是R^n中的非空开集。

设 x^* 为可行点, $I=\{i|g_i(x^*)=0\}$. 进一步假设，$f(x)$和 $g_i(x)$ $(i\in I)$ 在 x^* 可微，g_i $(i\notin I)$在 x^*连续.

若 x^* 是局部最优解，则 $F_0(x^*)\cap G_0(x^*)=\varnothing$,

其中$F_0(x^*)=\{d\ |\nabla f(x^*)d<0\}$,

$G_0(x^*)=\{d\ |\nabla g_i(x^*)d>0,\ i\in I\}$.

大连理工大学 Dalian University Of Technology

不等式约束的一阶最优性条件

定理 **7.2.3**. (**Fritz John Condition**, 1948)考虑极小化问题

$$\min\ f(x),\quad \text{s.t}\ g_i(x)\geq 0,\quad i=1,\ldots,m,\quad x\in S,$$

其中 S 是 E^n中非空开集. 设 x* 为可行点, $I=\{i \mid g_i(x^*)=0\}$. 进一步假设 f(x) 和 $g_i(x)(i\in I)$在 x* 可微, $g_i\ (i\notin I)$ 在 x*连续. 若 x* 是局部最优解,则存在一组非负数 u_0, $u_i\ (i\in I)$使得

$$u_0\nabla f(\mathbf{x}^*)-\sum_{i\in I}u_i\nabla g_i(\mathbf{x}^*)=0,\ u_0,\ u_i\geq 0 \text{ for } i\in I \text{ and } (u_0, u_I)\neq 0.$$

进一步, 若 $g_i(\mathbf{x})\ (i\notin I)$ 在 x*也可微，则

$$u_0\nabla f(\mathbf{x}^*)-\sum_{i=1}^{i=m}u_i\nabla g_i(\mathbf{x}^*)=0,$$

$u_i g_i(\mathbf{x}^*)=0$, u_0, u_i (所有 i)， 且 $(u_0, \mathbf{u})\neq 0$.

■ 不等式约束的一阶最优性条件

例1 已知x*=(3,1)′是下列问题的最优解：

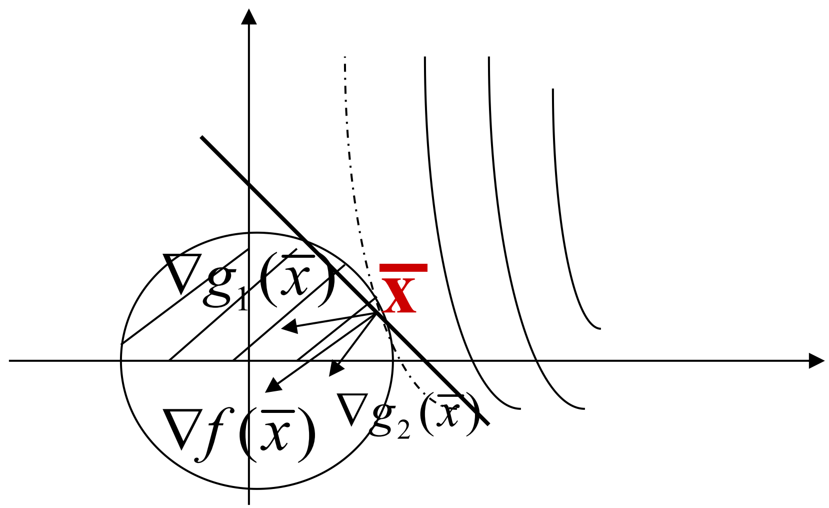
$$\min\ f(x)=(x_1-7)^2+(x_2-3)^2$$

$$s.t.\quad g_1(x)=10-x_1^{\ 2}-x_2^{\ 2}\ge 0$$

$$g_2(x)=4-x_1-x_2\ge 0$$

$$g_3(x)=x_2\ge 0$$

验证满足FJ条件.



■ 不等式约束的一阶最优性条件

例2 给定非线性规划问题

$$\min \ f(x) = -x_2$$

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

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

验证$\bar{x} = (0,2)'$处FJ条件满足。

■ 不等式约束的一阶最优性条件

若 **Lagrangian 乘子** $u_0=0$, 则 **Fritz John 条件** 不包含f(x)的任何信息，它仅仅是表明可以把起作用约束的梯度作一个非负的非平凡的线性组合而成为零向量。从而对我们的最优解没有多少实用价值。

为保证$u_0>0$,可以对约束强加某种限制，这种限制条件叫做约束规格或约束品性(**Constraint Qualifications**).已有很多的约束规格，特别的, **Karush** [1939, MS Thesis, Dept of Math, Univ of Chicago] , **Kuhn** 和 **Tucker** [1951] 独立给出的最优性必要条件恰是 **Fritz John 条件**加上 $u_0>0$。

最优性条件

■ 不等式约束的一阶最优性条件

定理 7.2.4. (Karush-Kuhn-Tucker 必要条件)考虑极小化问题

$$\min\ f(x)\ \text{s.t}\ g_i(x)\geq 0,\quad i=1,\ldots,m,\quad x\in S,$$

其中 S 是 E^n中非空开集. 设 x^* 为可行点, $I=\{i|\ g_i(x^*)=0\}$.
进一步假设 $f(x)$ 和 $g_i(x)(i\in I)$在 x^* 可微, g_i $(i\notin I)$ 在 x^*连续.
∇g_i对$i\in I$ 线性独立.
若 x^* 是局部最优解,则存在一组非负数$u_i(i\in I)$使得

$$\nabla f(x^*)-\sum_{i\in I} u_i\nabla g_i(x^*)=0,\ u_i\geq 0\ (i\in I).$$

若还有 g_i $(i\notin I)$在 x^*可微, 则

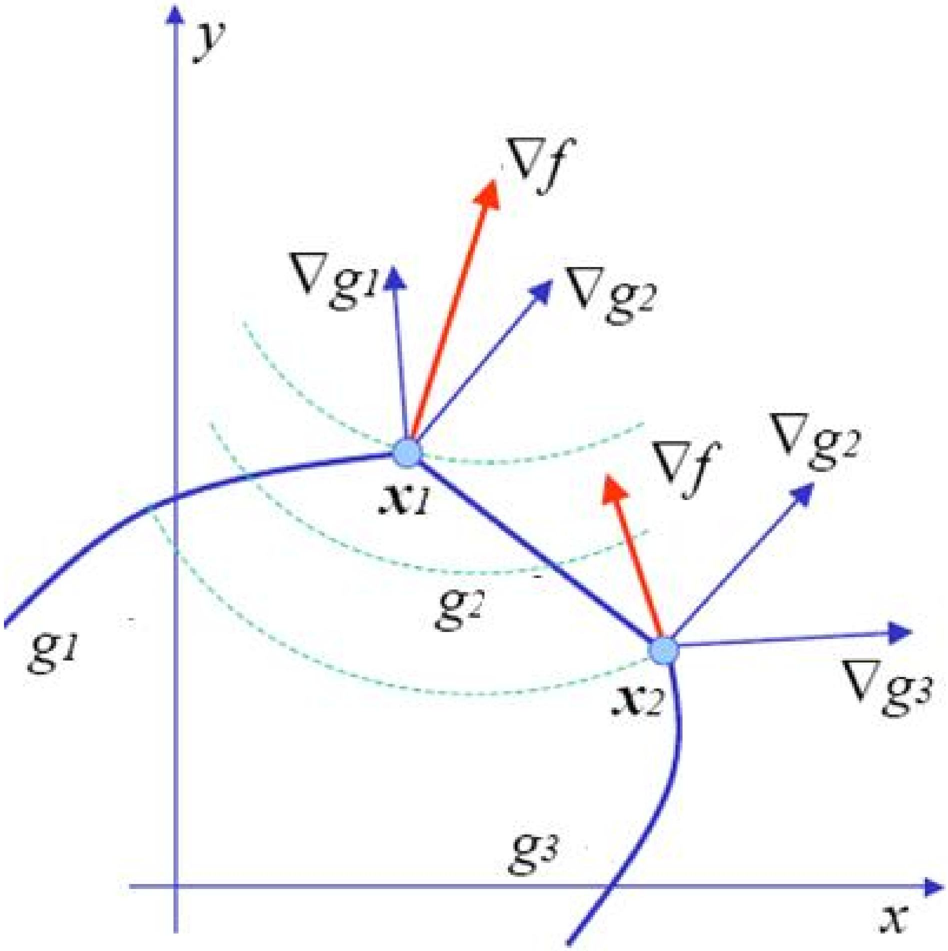
$$\nabla f(x^*)-\sum_{i=1}^{i=m} u_i\nabla g_i(x^*)=0,$$

$$u_i g_i(x^*)=0,\ u_i\geq 0,\quad i=1,\ldots,m.$$

■ 不等式约束的一阶最优性条件

Karush-Kuhn-Tucker
条件可写成向量形式
$\nabla f(\mathbf{x}^*) - \mathbf{u}\nabla \mathbf{g}(\mathbf{x}^*) = \mathbf{0}$,
$\mathbf{u}\mathbf{g}(\mathbf{x}^*) = 0$,
$\mathbf{u} \geq \mathbf{0}$.

这表明 $\nabla f(\mathbf{x}^*)$ 属于起作用约束的这些约束的梯度所形成的锥中。



■ 不等式约束的一阶最优性条件

例3给定非线性规划

$$\min \quad (x_1 - 2)^2 + x_2^2$$

$$s.t. \quad x_1 - x_2^2 \geq 0$$

$$-x_1 + x_2 \geq 0$$

验证下列两点

$x^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$和$x^{(2)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$是否为K-T点

■ 不等式约束的一阶最优性条件

定理 7.2.5. (**Karush-Kuhn-Tucker 充分条件**)考虑极小化问题

$$\min\ f(x)\ \ s.\,t\ \ g_i(x)\geq 0,\quad i=1,\ldots,m\ ,\quad x\in S,$$

其中S是E^n中非空开集. 设 x^* 为可行点, $I=\{i|\ g_i(x^*)=0\}$.
设$f(x)$和g_i是凹的，进一步假设 $f(x)$ 和 $g_i(x)(i\in I)$在 x^* 可微, $g_i\ (i\notin I)$ 在x^*连续. 若 K-K-T条件在 x^*成立，则x^*是全局最优解.

证明略

■ 一般约束问题的一阶最优性条件

考虑约束优化(COP)

$$\begin{aligned} &\min \quad f(x) \quad x \in E^n \\ s,t \quad & g_i(x) \geq 0, \quad i = 1,2,..,m \qquad (7.2.1) \\ & h_j(x) = 0, \quad j = 1,2,...,l \end{aligned}$$

记

$$g(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \\ \cdots \\ g_m(x) \end{pmatrix}, h(x) = \begin{pmatrix} h_1(x) \\ h_2(x) \\ \cdots \\ h_l(x) \end{pmatrix}$$

(COP) 矩阵形式

$$\begin{aligned} &\min \quad f(x) \quad x \in E^n \\ s,t \quad & g(x) \geq 0, \qquad (7.2.33) \\ & h(x) = 0, \end{aligned}$$

■ 一般约束问题的一阶最优性条件

定义7.2.3 设$\bar{x}$为可行点，不等式约束中在$\bar{x}$点起作用约束下标集记作I,若向量组$\{\nabla g_i(\bar{x}),\nabla h_j(\bar{x}) \mid i=1,2,\ldots,m, j=1,2,\ldots,l\}$线性无关，则称$\bar{x}$为约束$g(x)$和$h(x)$的正则点。

定义7.2.4 点集$\{x=x(t) \mid t_0 \le t \le t_1\}$称为曲面$S=\{x \mid h(x)=0\}$上的一条曲线,如果对所有$t\in[t_0,t_1]$均有$h(x(t))=0$.

显然,曲线上点是参数t的函数,若导数$x'(t)=\left. dx(t) \middle/ dt \right.$存在则称曲线是可微的.曲线$x(t)$的一阶导数$x'(t)$是曲线在点$x(t)$处的切向量.曲面S上在点$x$处所有可微曲线的切向量组成的集合,称为曲面S在点$x$的切平面,记做$T(x)$.

一般约束问题的一阶最优性条件

定义集合

$$H=\left\{d \,\middle|\, \nabla h_j(x)'d=0, \quad j=1,2,\ldots,l\right\}$$

可证明，当$\nabla h_1(x)$，$\nabla h_2(x),\ldots,\nabla h_l(x)$线性无关时，H等于等式约束

$$\begin{cases} h_1(x)=0 \\ h_2(x)=0 \\ \quad\cdots \\ h_l(x)=0 \end{cases}$$

所定义的超曲面在$\bar{x}$处的切平面$T(\bar{x})$。

一般约束问题的一阶最优性条件

$Th7.2.6$设$\bar{x}$是曲面$S=\{x\mid h(x)=0\}$上一个正则点(即$\nabla h_1(x)$，$\nabla h_2(x),\ldots,\nabla h_l(x)$线性无关), 则在点$\bar{x}$切平面$T(\bar{x})$等于子空间$H=\{d\mid\nabla h(\bar{x})^T d=0\}$.

证明略

■ 一般约束问题的一阶最优性条件

$Th7.2.7$设在约束极值问题(7.2.1)中,$\bar{x}$为可行点,

$I=\{i\mid g_i(\bar{x})=0\}$, f和$g_i(i\in I)$在点$\bar{x}$可微, $g_i(i\notin I)$

在点$\bar{x}$连续, $h_j(j=1,2,\ldots,l)$在点$\bar{x}$连续可微,且

$\nabla h_1(\bar{x})$, $\nabla h_2(\bar{x}),\ldots,\nabla h_l(\bar{x})$线性无关. 若$\bar{x}$是局部

最优解,则在点$\bar{x}$处,有 $F_0\cap G_0\cap H_0=\Phi$

其中

$$F_0=\{d\mid d^T\nabla f(x)<0\}$$

$$G_0=\{d\mid d^T\nabla g_i(x^*)>0\ (i\in I)\}$$

$$H_0=\{d\mid d^T\nabla h_j(x^*)=0\ (j=1,2,\ldots,l)\}$$

■ 一般约束问题的一阶最优性条件

下面给出一阶必要条件的代数表示. 我们有如下定理.

*Th*7.2.8(Fritz John条件)设在约束极值问题(7.2.1)中, $\bar{x}$为可行点,$I=\{i\,|\,g_i(\bar{x})=0\}$, f和$g_i(i\in I)$在点$\bar{x}$可微, $g_i(i\notin I)$在点$\bar{x}$连续,$h_j(j=1,2,\ldots,l)$在点$\bar{x}$连续可微. 若$\bar{x}$是局部最优解, 则∃不全为零的数$w_0, w_i\ge 0(i\in I), v_j\in R(j=1,2,\ldots,l)$, 使得

$$w_0\nabla f(\bar{x})-\sum_{i\in I}w_i\nabla g_i(\bar{x})-\sum_{j=1}^{l}v_j\nabla h_j(\bar{x})=0$$

■ 一般约束问题的一阶最优性条件

例:

$$\min \quad x_1^2 + x_2^2$$

$$\begin{aligned} s.t \quad & x_1^2 + x_2^2 \le 5, \\ & -x_1 \le 0, \\ & -x_2 \le 0, \\ & x_1 + 2x_2 = 4. \end{aligned}$$

验证在$x^* = (4/5, 8/5)$处

Fritz John条件成立.

此处只有一个等式约束.

注意到在x^*处I=ϕ, 因此

与不等式相应的乘子为零.

$\nabla f(x^*) = (8/5, 16/5)^T$,

$\nabla h_1(x^*) = (1, 2)^T$

于是$w\nabla f(x^*) + v\nabla h_1(x^*) = 0$即

$$w\begin{pmatrix} 8/5 \\ 16/5 \end{pmatrix} + v\begin{pmatrix} 1 \\ 2 \end{pmatrix} = 0$$

有合适解. 如取$w = 5, v = -8$

■ 一般约束问题的一阶最优性条件

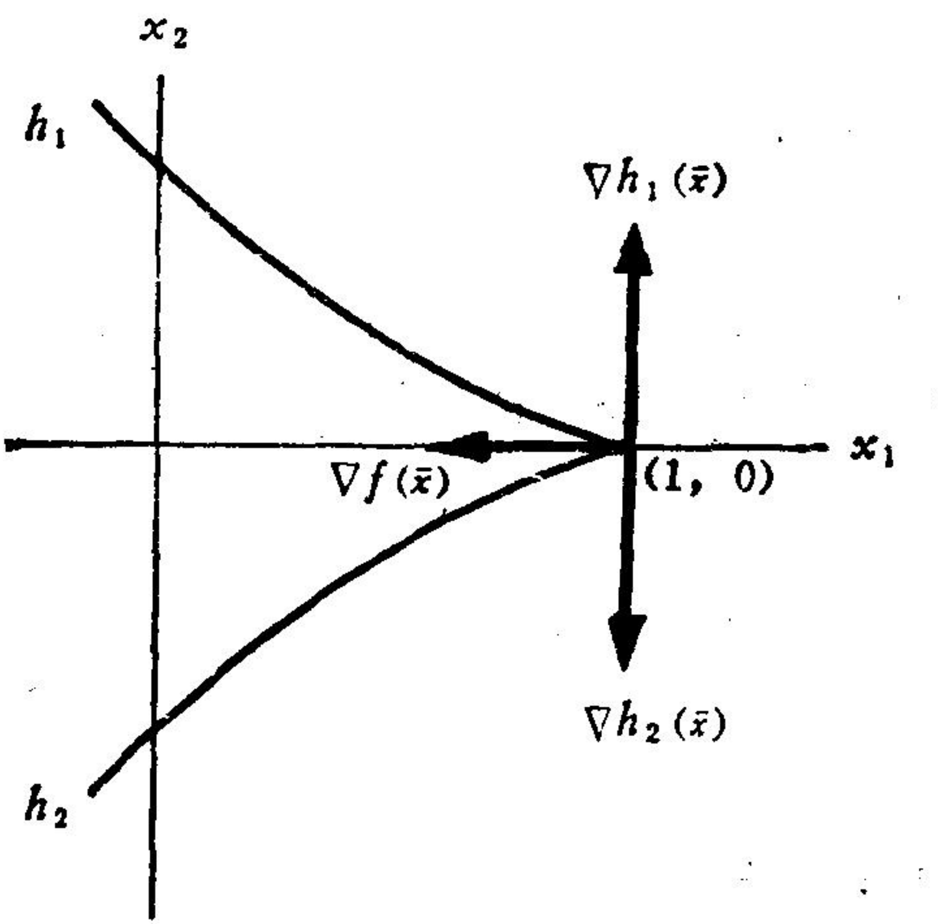
例：

$$\min -x_1$$

$$x_2-(1-x_1)^3=0,$$

$$-x_2-(1-x_1)^3=0,$$

验证可行点(1,0)是否满足FJ条件。



■ 一般约束问题的一阶最优性条件

$$\min -x_1$$
$$x_2-(1-x_1)^3=0,$$
$$-x_2-(1-x_1)^3=0,$$

此问题只有一个可行解$x^*=(1,0)^T$.

在此点处, 有

$\nabla f(x^*)=(-1,0)^T$,

$\nabla h_1(x^*)=(0,1)^T, \nabla h_2(x^*)=(0,-1)^T$,

FJ条件 $\quad w_0\begin{pmatrix}-1\\0\end{pmatrix}+v_1\begin{pmatrix}0\\1\end{pmatrix}+v_2\begin{pmatrix}0\\-1\end{pmatrix}$

当$w_0=0, v_1=-v_2=a$(a为任意数) 成立.

故在x^*满足FJ条件.

一般约束问题的一阶最优性条件

前面FJ条件中w_0不一定为正, 在下面定理中。我们将前面提到的KKT最优性必要条件（Th2.4）加以推广。这是通过增加约束规格来实现的.

定理 7.2 .9(一阶必要条件(Karush-Kuhn-Tucker 定理) :代数特征)

设$x^* \in S$是问题(7. 2. 1)的局部极小点, 函数$f, g_i, (i \in I)$在点x^*处可微, $g_i, (i \notin I)$在点x^*处连续, h_j在点x^*处连续可微. 若

$$\{\nabla g_i(x^*), \nabla h_j(x^*) \mid i \in I, j = 1, 2, \ldots, l\}$$

线性无关，则$\exists w_i \geq 0 (i \in I), v_j \in R (j = 1, 2, \ldots, l)$, 使得

$$\nabla f(x^*) = \sum_{i \in I} w_i \nabla g_i(x^*) + \sum_{j \in E} v_j \nabla h_j(x^*)$$

一般约束问题的一阶最优性条件

进一步假设,g_i ($i \notin \mathbf{I}$)在 $\mathbf{x}^*$, 连续可微,则

$$\nabla f(\mathbf{x}^*) - \sum_{i=1}^{i=m} u_i \nabla g_i(\mathbf{x}^*) - \sum v_j \nabla h_j(\mathbf{x}^*) = \mathbf{0},$$

$$u_i g_i(\mathbf{x}^*) = 0,\ u_i \geq 0 (i=1,\ldots,m.)$$

□若采用矩阵和向量记号,则KKT可如下简洁表示

定义Lagrange函数为

$$L(\boldsymbol{x}, \lambda, \mu) = \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{g}^T \lambda - \boldsymbol{h}^T \mu \quad (7.2.48)$$

其中$\boldsymbol{g}$和$\boldsymbol{h}$分别是由$\boldsymbol{g}_i$和$\boldsymbol{h}_j$组成的向量函数.

■ 一般约束问题的一阶最优性条件

记向量值函数$\boldsymbol{g}$和$\boldsymbol{h}$在$\boldsymbol{x}$点的梯度为

$$\nabla \boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{J}_g(\boldsymbol{x})^T = (\nabla \boldsymbol{g}_i, \boldsymbol{i} \in \boldsymbol{I}),$$

$$\nabla \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{J}_h(\boldsymbol{x})^T = (\nabla \boldsymbol{h}_j, \boldsymbol{j} = 1, 2, \ldots, \boldsymbol{l}),$$

其中$\boldsymbol{J}_g(\boldsymbol{x}), \boldsymbol{J}_h(\boldsymbol{x})$分别是$\boldsymbol{g}, \boldsymbol{h}$在$\boldsymbol{x}$点的Jacobi矩阵.

则KKT条件可表为

$$\begin{aligned}
&\nabla \boldsymbol{f}(\boldsymbol{x}^*) = \nabla \boldsymbol{g}(\boldsymbol{x}^*)\lambda + \nabla \boldsymbol{h}(\boldsymbol{x}^*)\mu \\
&\boldsymbol{g}(\boldsymbol{x}^*)^T \lambda = 0, \qquad \lambda \geq 0 \qquad\qquad (7.2.49) \\
&\boldsymbol{g}(\boldsymbol{x}^*) \geq 0, \quad \boldsymbol{h}(\boldsymbol{x}^*) = 0
\end{aligned}$$

■ 一般约束问题的一阶最优性条件

定义若$\boldsymbol{x}^* \in \boldsymbol{R}^n, \lambda \in \boldsymbol{R}^{|I|}, \mu \in \boldsymbol{R}^l$满足KKT条件(7.2.49),
则$\boldsymbol{x}^*$称为约束优化问题(COP)的一个KKT点, λ和μ称为$\boldsymbol{x}^*$处的Lagrange乘子.特别的, $\boldsymbol{g}(\boldsymbol{x}^*)^T\lambda=0$称为互补松弛条件.

■ 一般约束问题的一阶最优性条件

定义：设$x \in S, d \in R^n$，若∃向量序列$\{d^k\}$和正实数列$\{\delta_k\}$使得对$\forall k = 1,2,\ldots$有$x + \delta_k d^k \in S$，且$d^k \to d, \delta_k \to 0$则称$d$为$S$在点$x$的一个序列化可行方向. S在点x的所有序列化可行方向的集合，称为序列化可行方向锥，记为$SFD(x, S)$

定理 (充分条件)设$x^* \in S$，函数f, g_i, h_j在点x^*处可微，且

$$d^T \nabla f(x^*) > 0, \forall d \in SFD(x^*, S) \setminus \{0\}$$

则x^*为(COP)的严格局部极小点.

■ 一般约束问题的一阶最优性条件

定理7.2.10 (充分条件:凸规划情形)

设f为凸函数，g_i为凹函数, h_j为线性函数。对于$x^* \in S$, 若f, g_i, h_j在点x^*处可微, 并且KKT条件(7. 2. 49)成立，则x^*为优化问题（COP）的全局极小点。

■ 第五次作业

第243页第七章2.3.4.5.6.8.9

感兴趣的同学学习7.3节--对偶及鞍点问题
(和线性规划的对偶问题类似)

11月23日--------习题课!
(相同时间答疑)

■ 小结

- ➢ 无约束问题的极值条件
- ➢ 约束极值问题的最优性条件
- ➢ 对偶及鞍点*