

最优化理论

Optimality Theory



- 01 课程简介(Introduction)**
- 02 线性规划(Linear Programming)**
- 03 非线性规划(Non-Linear Programming)**
- 04 整数规划(Integer Programming)**
- 05 动态规划(Dynamic Programming)**

■ 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！



最优化条件

Optimality Condition

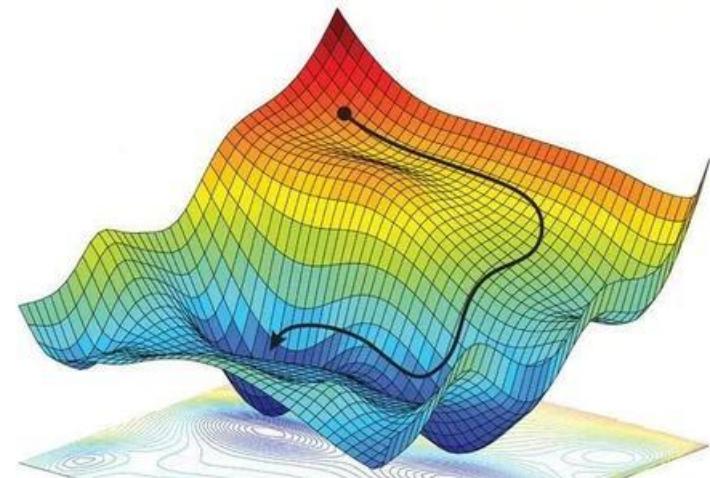
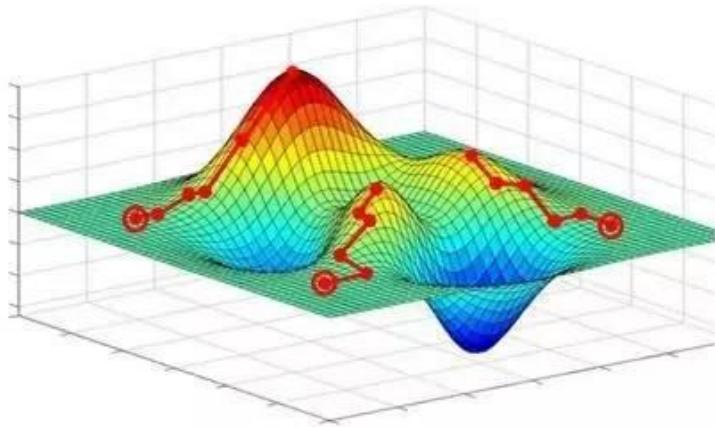
■ 基本概念

最优化问题可归结成如下数学形式：

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

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

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





■ 基本概念

- 对于极大化问题，可类似定义全局极大点和局部极大点；
- 全局极小点也是局部极小点；
- 局部极小点不一定是全局极小点；
- 某些条件下(如凸规划)，局部极小点也是全局极小点

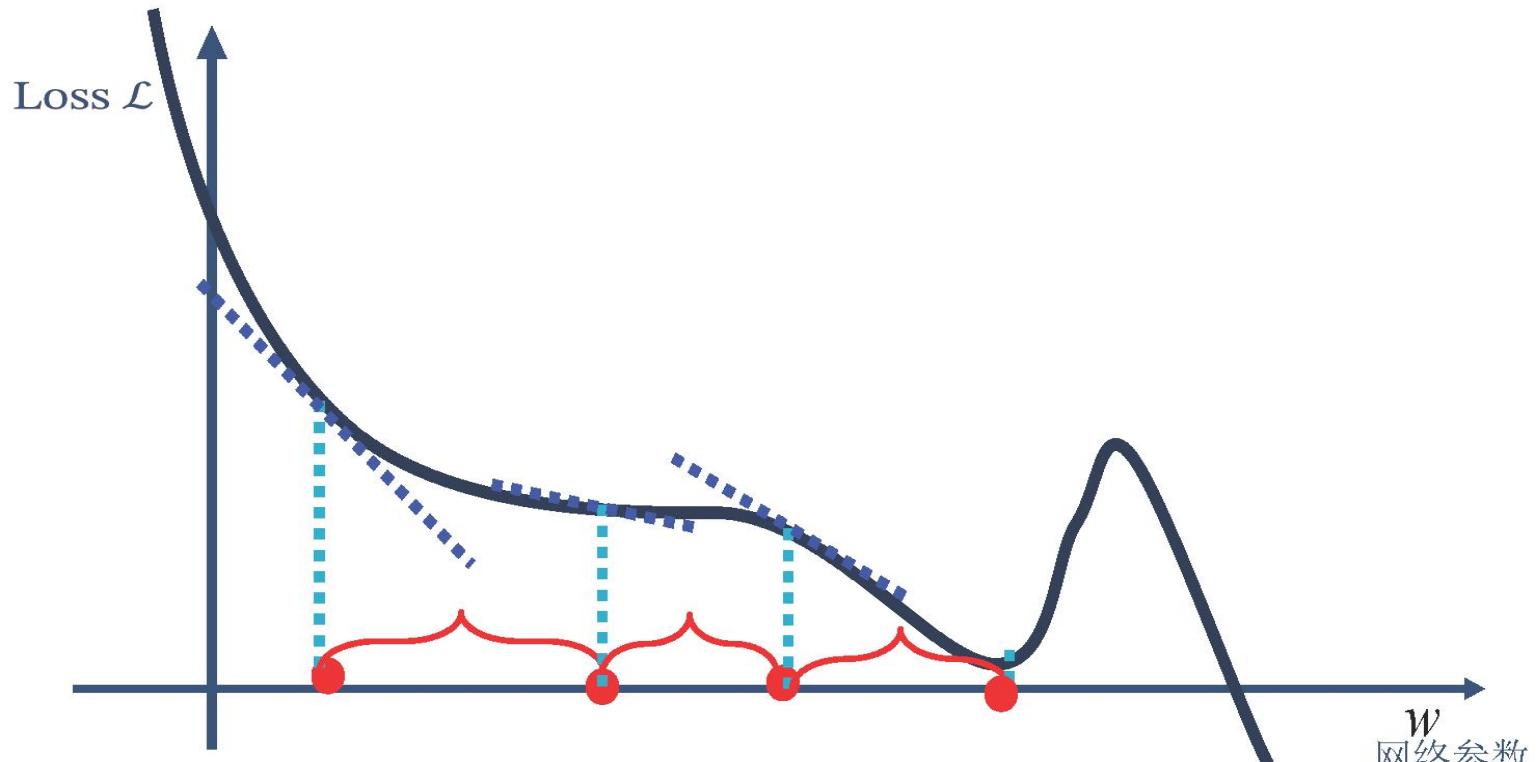
■ 求函数的极小值

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

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

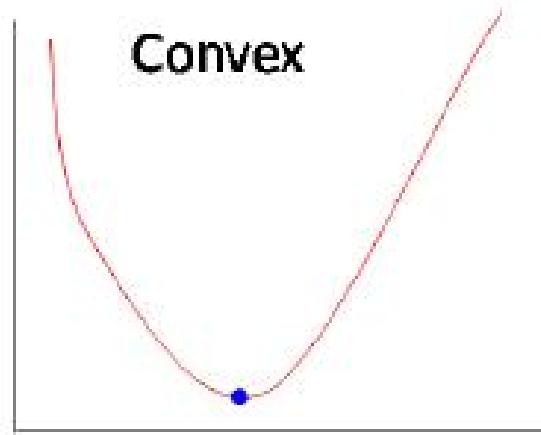
■ 梯度下降 (Gradient Descent)

➤ 梯度下降类似盲人下山

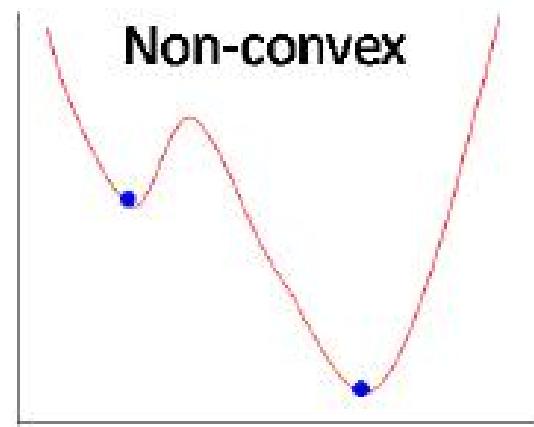


$$\text{梯度: } \frac{\partial f(w)}{\partial w} = \lim_{\Delta w \rightarrow 0} \frac{f(w + \Delta w) - f(w)}{\Delta w}$$

■ 学习准则 (Learning Rule)



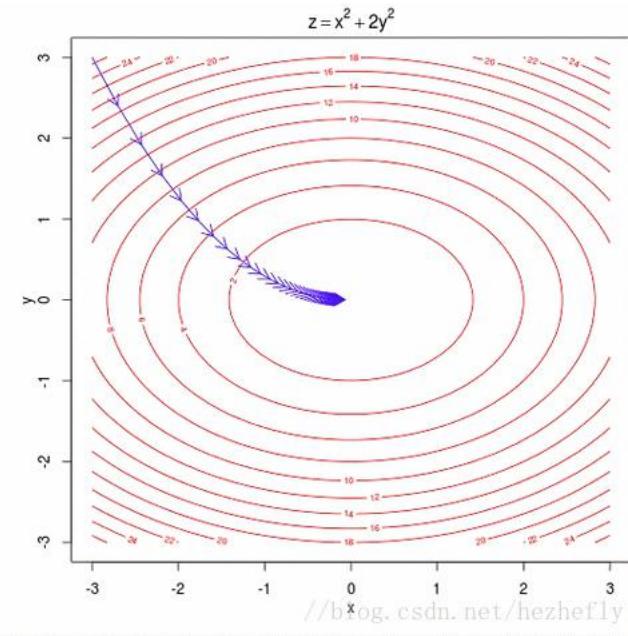
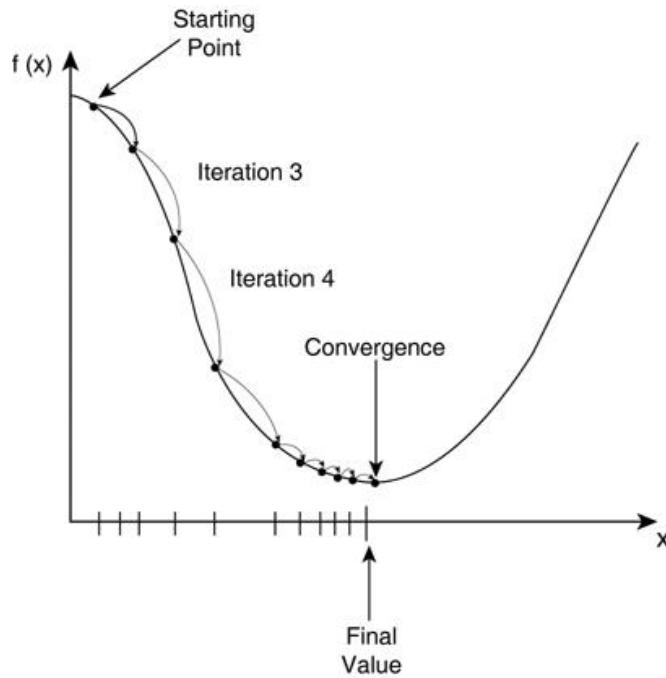
Convex



Non-convex

$$\min_{\mathbf{x}} f(\mathbf{x})$$

■ 梯度下降法(Gradient Descent)

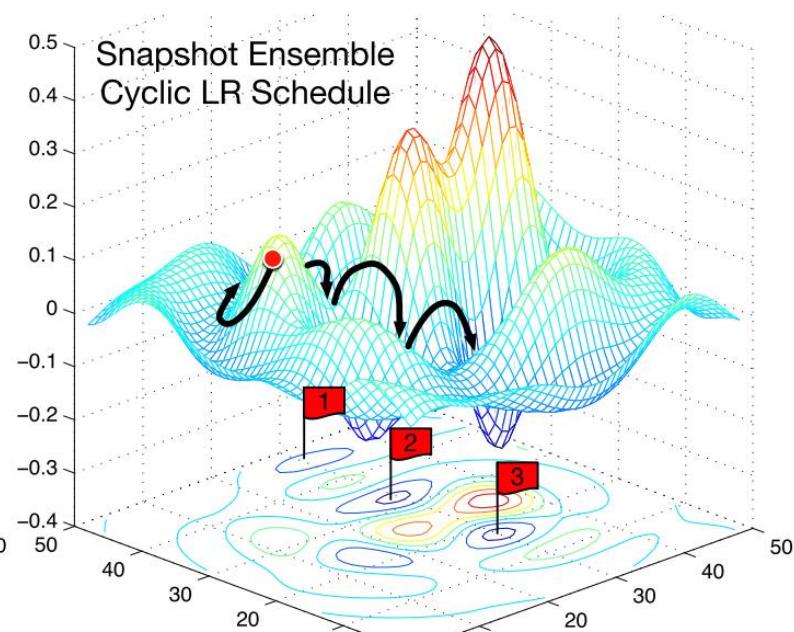
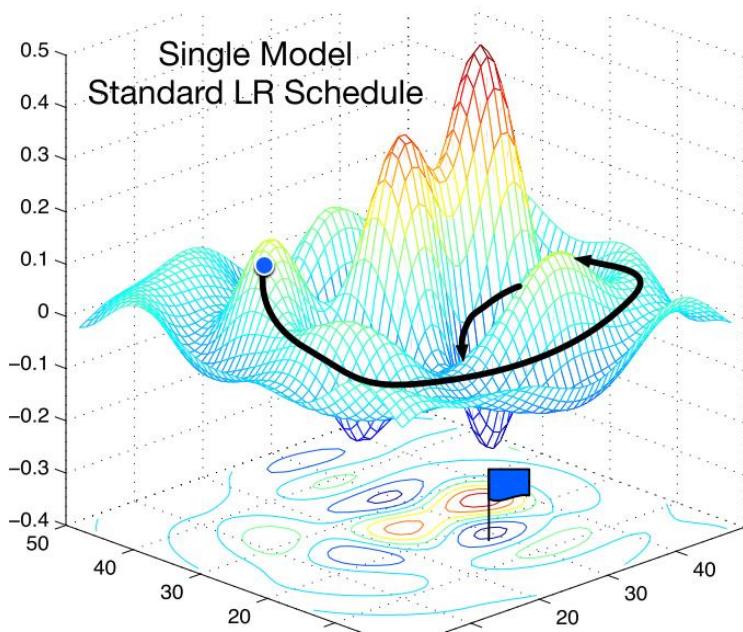


$$\begin{aligned}
 \theta_{t+1} &= \theta_t - \alpha \frac{\partial \mathcal{R}(\theta)}{\partial \theta_t} \\
 &= \theta_t - \alpha \frac{1}{N} \sum_{i=1}^N \frac{\partial \mathcal{L}(\theta_t; x^{(i)}, y^{(i)})}{\partial \theta}.
 \end{aligned}$$

搜索步长 α 中也叫作**学习率**(Learning Rate)

■ 学习率(Learning Rate)

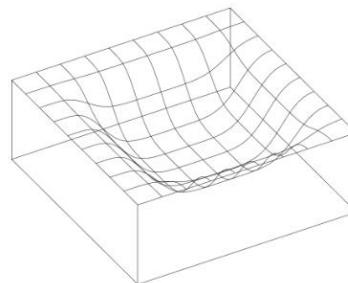
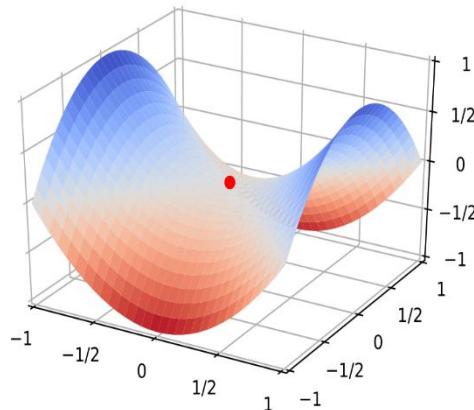
➤ 周期性学习率调整



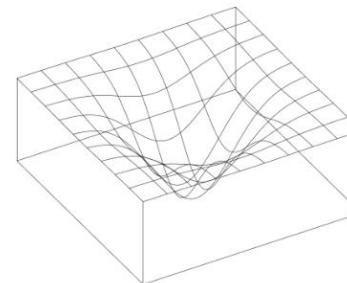
■ 非凸优化问题(Non-convex Optimal Problems)

➤ 平坦最小值问题

鞍点(Saddle Point)、驻点(Stationary Point)



(a) 平坦最小值



(b) 尖锐最小值

1. 梯度值为0，在平坦最小值的邻域内，所有点对应的训练损失都比较接近
2. 大部分的局部最小解是等价的
3. 局部最小解对应的训练损失都可能非常接近于全局最小解对应的训练损失

■ 优化算法(Optimization Methods)

➤ 可视化

[鞍点](#)

[优化曲面](#)

- 上面两种情况都可以看出，有的算法几乎很快就找到了正确的方向并前进，收敛速度也相当快，而其它方法要么很慢，要么走了很多弯路才找到
- 那么是什么时候才能有最优解(局部)？如何找出呢？

■ 主要内容(非常重要)

本章研究非线性规划的最优解所满足的必要条件和充分条件，为各种算法(如SVM、牛顿迭代、割线法)的推导和分析提供必不可少的理论基础。

- 无约束问题的极值条件
- 约束极值问题的最优化条件
- 对偶及鞍点*

■ 无约束问题的极值条件

无约束极值问题

考虑非线性规划问题

$$\min f(x), \quad x \in E^n$$

其中 $f(x)$ 是定义在 E^n 上的实值函数

——称为无约束极值问题 (UNLP)

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

■ 无约束问题的极值条件

定理 7.1.1(非极小点的充分条件)

设 $f(x)$ 在点 x^* 处可微, 若存在方向 $d(\neq 0) \in \mathbb{R}^n$, 使得 $\nabla f(x^*)'d < 0$, 则存在 $\delta > 0$, 使得对任意 $\lambda \in (0, \delta)$, 有 $f(x^* + \lambda d) < f(x^*)$.

x^* 一定不是极小点

此时, 我们称 d 为 $f(x)$ 在 x^* 的一个下降方向.

证明. 由 $f(x)$ 在 x^* 可微, 则

$$f(x^* + \lambda d) = f(x^*) + \lambda \nabla f(x^*)' d + \lambda \|d\| \alpha(x^*; \lambda d),$$

其中 $\alpha(x^*; \lambda d) \rightarrow 0$ (当 $\lambda \rightarrow 0$).

■ 无约束问题的极值条件

移项且两边同除以 λ ($\lambda \neq 0$),得

$$(f(x^* + \lambda d) - f(x^*))/\lambda = \nabla f(x^*)d + \|d\|\alpha(x^*; \lambda d)$$

由于 $\nabla f(x^*)d < 0$ 且 $\alpha(x^*; \lambda d) \rightarrow 0$ (当 $\lambda \rightarrow 0$).

从而存在 $\delta > 0$ 使得对任意 $\lambda \in (0, \delta)$,

$$\nabla f(x^*)d + \|d\|\alpha(x^*; \lambda d) < 0 .$$

定理立明.

■ 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 \mathbf{x}^* 处是问题(UNLP)的局部极小点.

- (1) 当 $f(\mathbf{x})$ 在 \mathbf{x}^* 可微时, 则梯度 $\nabla f(\mathbf{x}^*) = \mathbf{0}$.
- (2) 当 $f(\mathbf{x})$ 在 \mathbf{x}^* 二次可微时, 则 $\nabla f(\mathbf{x}^*) = \mathbf{0}$ 且 Hessian 矩阵 $\mathbf{H}(\mathbf{x}^*)$ 是半正定的

证明(1). 若 $\nabla f(\mathbf{x}^*) \neq \mathbf{0}$, 作 $\mathbf{d} = -\nabla f(\mathbf{x}^*)$.

则有 $\nabla f(\mathbf{x}^*) \mathbf{d} < 0$

由定理 7.1.1, 存在 $\delta > 0$ 使得 $f(\mathbf{x}^* + \lambda \mathbf{d}) < f(\mathbf{x}^*), \forall \lambda \in (0, \delta)$, 此与 \mathbf{x}^* 为局部极小相矛盾, 故 $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

■ 1.5 函数的可微性与展开

设 $f: R^n \rightarrow R$ 是连续函数, 对于 $x^0 \in R^n$ 和单位向量

$e^i = (\delta_1^i, \delta_2^i, \dots, \delta_n^i)^T$, $\delta_i^i = 1$, $\delta_j^i = 0$ ($i \neq j$), 一元函数 $f(x^0 + te^i)$

在 $t = 0$ 的导数(若存在的话)称为 f 在点 x^0 关于 x_i 的 **一阶偏导数**.
($i = 1, 2, \dots, n$)

对于任意 $i = 1, 2, \dots, n$, 若 f 在点 x 关于 x_i 的一阶偏导数存在, 则称 $f(x)$ 在 x 点存在一阶偏导数. 此时, $f(x)$ 在 x 点的 **梯度** 定义为

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^T.$$

对于 $x^0 \in R^n$, $p \in R^n$, $p \neq 0$, 函数 f 在点 x^0 关于方向 p 的 **方向导数** 定义为:

$$\frac{\partial f(x^0)}{\partial p} = \lim_{t \rightarrow 0^+} \frac{f(x^0 + tp) - f(x^0)}{t}.$$

■ 1.5 函数的可微性与展开

我们也用 $Df(x^0; p)$ 表示 f 在点 x^0 关于方向 p 的方向导数 当的一阶偏导连续时有 $Df(x^0; p) = \nabla f(x^0)^T p$

当 $f(x)$ 在 x 点存在二阶偏导时, 函数 f 在点 x 的 Hesse 矩阵 定义为

$$\nabla^2 f(x) \triangleq \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \dots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix} \quad \left(\text{其中 } \frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_j} \left(\frac{\partial f(x)}{\partial x_i} \right) \right)$$

■ 1.5 函数的可微性与展开

Th1.9(Taylor) 设 $f: R^n \rightarrow R$ 连续可微, 向量 $p \in R^n$, 则

$$\begin{aligned} f(x + p) &= f(x) + \int_0^1 \nabla f(x + tp)^T p dt = f(x) + \nabla f(x + \xi p)^T p, \xi \in (0,1) \\ &= f(x) + \nabla f(x)^T p + o(\|p\|) \end{aligned}$$

进而, 若 f 二阶连续可微, 则

$$\begin{aligned} \nabla f(x + p) &= \nabla f(x) + \int_0^1 \nabla^2 f(x + tp)^T p dt = \nabla f(x) + \nabla^2 f(x + \xi p) p, \xi \in (0,1) \\ &= \nabla f(x) + \nabla^2 f(x) p + o(\|p\|) \end{aligned}$$

$$\begin{aligned} f(x + p) &= f(x) + \nabla f(x)^T p + \int_0^1 (1-t) p^T \nabla^2 f(x + tp)^T p dt \\ &= f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x + \xi p) p, \xi \in (0,1) \\ &= f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x) p + o(\|p\|^2) \end{aligned}$$

■ 1.5 函数的可微性与展开

对向量值函数 $F = (f_1, f_2, \dots, f_n) : R^n \rightarrow R^m$, 若每个分量函数 f_i 是(连续)可微的, 则称函数 F 是(连续)可微的。向量函数 F 在 x 的导数 $F' \in R^{m \times n}$ 是指它在 x 点的 *Jacobi* 矩阵, 记为 $F'(x)$ 或 $J_F(x)$. 为与标量函数梯度对应, 定义 *Jacobi* 矩阵的转置为 F 在 x 点的梯度, 记为

$$\nabla F(x) = J_F(x)^T = (\nabla f_1, \nabla f_2, \dots, \nabla f_n)$$

类似地, 设 $F : R^n \rightarrow R^m$ 连续可微, $\forall x, p \in R^n$, 则

$$F(x + p) = F(x) + \int_0^1 J_F(x + tp)^T p dt$$

■ 1.5 函数的可微性与展开

Df1.12 给定映射 $G: R^n \rightarrow R^m$, 点 $x \in R^n$, 若存在常数 $L > 0$, 使对任意 $y \in R^n$, 下式成立: $\|G(x) - G(y)\| \leq L \|x - y\|$

则称 G 在 x 是 Lipschitz 连续的, L 称为 Lipschitz 常数。若上式对 $\forall x, y \in R^n$ 成立, 则称 G 在 R^n 内 Lipschitz 连续。

Th1.10 设向量值函数 $F: R^n \rightarrow R^m$ 连续可微, 则对 $\forall u, v, x \in R^n$, 有 $\|F(u) - F(v) - J_F(x)(u - v)\| \leq \sup_{t \in [0,1]} \|J_F(v + t(u - v)) - J_F(x)\| \|u - v\|$

进而, 若 Jacobi 矩阵映射 J_F 在 R^n 内是 Lipschitz 连续的, 记 L 为 Lipschitz 常数, 则

$$\|F(u) - F(v) - J_F(x)(u - v)\| \leq L \frac{\|u - x\| + \|v - x\|}{2} \|u - v\|$$

■ 无约束问题的极值条件

证明(2): 给定任意向量 \mathbf{d} . 由 $f(\mathbf{x})$ 在 \mathbf{x}^* 的二次可微性, 有

$$f(\mathbf{x}^* + \lambda \mathbf{d}) = f(\mathbf{x}^*) + \lambda \nabla f(\mathbf{x}^*) \mathbf{d} + \lambda \mathbf{d}' \mathbf{H}(\mathbf{x}^*) \mathbf{d} / 2 + \lambda \|\mathbf{d}\|^2 \alpha(\mathbf{x}^*; \lambda \mathbf{d}) \quad (\text{I}),$$

其中 $\alpha(\mathbf{x}^*; \lambda \mathbf{d}) \rightarrow 0$ ($\lambda \rightarrow 0$). 由 (1) 的证明有 $\nabla f(\mathbf{x}^*) = 0$.

移项整理并两端除以 λ , 得

$$(f(\mathbf{x}^* + \lambda \mathbf{d}) - f(\mathbf{x}^*)) / \lambda = \mathbf{d}' \mathbf{H}(\mathbf{x}^*) \mathbf{d} / 2 + \|\mathbf{d}\|^2 \alpha(\mathbf{x}^*; \lambda \mathbf{d}) \quad (\text{II}).$$

因 \mathbf{x}^* 局部极小, 对充分小 λ 有 $f(\mathbf{x}^* + \lambda \mathbf{d}) \geq f(\mathbf{x}^*)$

由 (II), 显见

$$\mathbf{d}' \mathbf{H}(\mathbf{x}^*) \mathbf{d} / 2 + \|\mathbf{d}\|^2 \alpha(\mathbf{x}^*; \lambda \mathbf{d}) \geq 0$$

对充分小的 λ 成立, 对 $\lambda \rightarrow 0$ 取极限, 则有 $\mathbf{d}' \mathbf{H}(\mathbf{x}^*) \mathbf{d} \geq 0$,
从而, $\mathbf{H}(\mathbf{x}^*)$ 半正定



■ 无约束问题的极值条件 (二阶充分条件)

定义1 若 $f(x)$ 在点 x^* 处可微, 且 $\nabla f(x^*) = 0$, 则称 x^* 为 $f(x)$ 的一个驻点或平稳点.

既不是极大点也不是极小点的驻点称为鞍点.

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

■ 无约束问题的极值条件

证明. 因 f 在 \mathbf{x}^* 二次可微, 故对任意 \mathbf{x} , 有

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + (\mathbf{x} - \mathbf{x}^*)^T \mathbf{H}(\mathbf{x}^*) (\mathbf{x} - \mathbf{x}^*)/2 + \|\mathbf{x} - \mathbf{x}^*\|^2 \alpha(\mathbf{x}^*; \mathbf{x} - \mathbf{x}^*),$$

这里 $\alpha(\mathbf{x}^*; \mathbf{x} - \mathbf{x}^*) \rightarrow 0$, 当 $\mathbf{x} \rightarrow \mathbf{x}^*$.

假设命题不真, \mathbf{x}^* 不是局部极小, 则存在序列 $\{\mathbf{x}_k\}$ 收敛到 \mathbf{x}^* , 并使得 $f(\mathbf{x}_k) < f(\mathbf{x}^*)$ 对每一 k 成立. 定义序列

$$(\mathbf{x}_k - \mathbf{x}^*) / \|\mathbf{x}_k - \mathbf{x}^*\| = \mathbf{d}_k.$$

则上述方程蕴含 $\mathbf{d}_k^T \mathbf{H}(\mathbf{x}^*) \mathbf{d}_k / 2 + \alpha(\mathbf{x}^*; \mathbf{x}_k - \mathbf{x}^*) \leq 0 \quad (\forall k)$

但对每一 k , $\|\mathbf{d}_k\| = 1$, 从而 $\{\mathbf{d}_k\} \rightarrow \mathbf{d}$, 当 $k \rightarrow \infty$, 这里 $\|\mathbf{d}\| = 1$.

由 $\nabla f(\mathbf{x}^*) = \mathbf{0}$, 我们有 $\mathbf{d}^T \mathbf{H}(\mathbf{x}^*) \mathbf{d} \leq 0$, 此与其正定性矛盾.

■ 无约束问题的极值条件(充要条件)

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

证明. (必要性) 若 x^* 全局最优, 由 Th7.1.2, $\nabla f(x^*) = 0$.

(充分性) 设 $\nabla f(x^*) = 0$, 则 $\nabla f(x^*)(x - x^*) = 0$, $\forall x^* \in E^n$

由 $f(x)$ 可微凸, 有 (Th1.4.14)

$$f(x) \geq f(x^*) + \nabla f(x^*)(x - x^*) = \nabla f(x^*)$$

定理得证.

■ 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：

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

解：先求驻点， 由于

$$\frac{\partial f}{\partial x_1} = 4x_1^3 - 2x_1 - 2, \frac{\partial f}{\partial x_2} = 2x_2$$

$$\text{令 } \nabla f(x) = 0, \text{ 即 } 4x_1^3 - 2x_1 - 2 = 0, 2x_2 = 0$$

解得驻点 $x^* = (1, 0)'$

$$\text{又 Hessian 阵 } \nabla^2 f(x) = \begin{bmatrix} 12x_1^2 - 2 & 0 \\ 0 & 2 \end{bmatrix} \Rightarrow \nabla^2 f(x^*) = \begin{bmatrix} 10 & 0 \\ 0 & 2 \end{bmatrix}$$

正定， 故 $x^* = (1, 0)'$ 是局部极小点。

■ 无约束问题的极值条件

例7.1.2 利用极值条件解下列问题：

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

解：先求驻点，由于

$$\frac{\partial f}{\partial x_1} = x_1^2 - 1, \frac{\partial f}{\partial x_2} = x_2^2 - 2x_2$$

$$\text{令 } \nabla f(x) = 0, \text{ 即 } x_1^2 - 1 = 0, x_2^2 - 2x_2 = 0$$

$$\text{解得驻点 } x^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, x^{(2)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, x^{(3)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, x^{(4)} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

■ 无约束问题的极值条件

又Hessian阵 $\nabla^2 f(x) = \begin{bmatrix} 2x_1 & 0 \\ 0 & 2x_2 - 2 \end{bmatrix} \Rightarrow$

$$\nabla^2 f(x^{(1)}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}, \nabla^2 f(x^{(2)}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\nabla^2 f(x^{(3)}) = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}, \nabla^2 f(x^{(4)}) = \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix}$$

由于 $\nabla^2 f(x^{(1)})$, $\nabla^2 f(x^{(3)})$, $\nabla^2 f(x^{(4)})$ 不定或负定, 仅 $\nabla^2 f(x^{(2)})$ 正定,

故 $x^{(2)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 是局部极小点。

■ 约束极值问题的最优化条件

约束极值问题

有约束的极值问题一般表示为

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

其中 $g_i(x) \geq 0$ 称为不等式约束, $h_j(x) = 0$ 称为等式约束,

集合 $S = \left\{ x \left| \begin{array}{l} g_i(x) \geq 0, i = 1, 2, \dots, m \\ h_j(x) = 0, \quad j = 1, 2, \dots, l \end{array} \right. \right\}$

称为可行集合或可行域

■ 约束极值问题的最优化条件

可行方向和下降方向

定义 7.2.1. 设 $f(\mathbf{x})$ 定义在 \mathbf{R}^n 上的实值函数. $\mathbf{x}^* \in \mathbf{R}^n$, $\mathbf{d} \neq 0$, 若存在 $\delta > 0$ 使得 $f(\mathbf{x}^* + \lambda \mathbf{d}) < f(\mathbf{x}^*) (\forall \lambda \in (0, \delta))$, 则 \mathbf{d} 称为 $f(\mathbf{x})$ 在 \mathbf{x}^* 的下降方向(Decedent Direction)

设 $f(\mathbf{x})$ 在 \mathbf{x}^* 可微. 若存在向量 \mathbf{d} 满足 $\nabla f(\mathbf{x}^*) \mathbf{d} < 0$, 则 由 Th7.1.1, \mathbf{d} 是 $f(\mathbf{x})$ 在 \mathbf{x}^* 的下降方向。记所有这样的向量集合为

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

定义 7.2.2 设 $\mathbf{x} \in \text{cl}S$, $\mathbf{d} \in \mathbf{R}^n$. 若 $\exists \delta > 0$, 使得 $\forall \lambda \in [0, \delta]$, $\mathbf{x} + \lambda \mathbf{d} \in S$, 则称 \mathbf{d} 为集合 S 在点 \mathbf{x} 的一个可行方向. 集合 S 在 \mathbf{x} 点的所有可行方向集合称为 S 在 \mathbf{x} 点的可行方向锥, 记为 D (或 $FD(\mathbf{x}, S)$)

$$D = \{d \mid d \neq 0, \mathbf{x} \in \text{cl}S, \exists \delta > 0, \text{使得对 } \forall \lambda \in (0, \delta), \text{ 有 } \mathbf{x} + \lambda \mathbf{d} \in S\}$$

■ 约束极值问题的最优化条件

由可行方向定义和下降方向知，从点 \mathbf{x}^* 沿可行方向 $\mathbf{d} \in \mathbf{D}(\mathbf{x}^*)$ 作一个很小的移动还是可行点。

进一步，由定理 7.1.1, 若 $\nabla f(\mathbf{x}^*)\mathbf{d} < 0$ ，则 \mathbf{d} 是 f 在 \mathbf{x}^* 的下降方向。下面定理将说明若 \mathbf{x}^* 是局部最优且 $\nabla f(\mathbf{x}^*)\mathbf{d} < 0$, 则 $\mathbf{d} \notin \mathbf{D}(\mathbf{x}^*)$. 即不是可行方向。

定理 7.2.1. (必要条件) 考虑极小化问题：

$$\min f(\mathbf{x}), \text{ s. t } \mathbf{x} \in \mathbf{S},$$

其中 \mathbf{S} 是 \mathbf{R}^n 中非空集合。设 $f(\mathbf{x})$ 在 \mathbf{x}^* 可微，若 \mathbf{x}^* 是局部极小点，则 $\mathbf{F}_0(\mathbf{x}^*) \cap \mathbf{D} = \emptyset$, 其中 $\mathbf{F}_0(\mathbf{x}^*) = \{\mathbf{d} \mid \nabla f(\mathbf{x}^*)\mathbf{d} < 0\}$ ， \mathbf{D} 是 \mathbf{S} 在 \mathbf{x}^* 的可行方向锥。

■ 约束极值问题的最优化条件

证明：反证法，设存在向量 $d \in F_0(x^*) \cap D$. 由 Th7.1.1 知

(1) : 存在 $\delta_1 > 0$, $f(x^* + \lambda d) < f(x^*)$, $\forall \lambda \in (0, \delta_1)$,

又由 D 的定义,

(2) : 存在 $\delta_2 > 0$, $x^* + \lambda d \in S$, $\forall \lambda \in (0, \delta_2)$

此与局部极小矛盾。

■ 不等式约束的一阶最优化条件

考察非线性规划

$$\min f(x)$$

$$s, t \quad g_i(x) \geq 0, \quad i = 1, 2, \dots, m$$

$$\text{可行域 } S = \{x \mid g_i(x) \geq 0, i = 1, 2, \dots, m\}$$

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 $x^* \in S$ 处有两种情形

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

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

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

■ 不等式约束的一阶最优化条件

定理 7.2.2. (必要条件) 考虑极小化问题

$$\min f(x)$$

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

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 $F_0(x^*) \cap G_0(x^*) = \emptyset$,

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

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

■ 不等式约束的一阶最优化条件

证明概要.

设 $d \in G_0(x^*)$. 因 S 为开集, 则存在 $\delta_1 > 0$ 使得对 $\forall \lambda \in (0, \delta_1)$, $x^* + \lambda d \in S$ 。

另外存在 $\delta_2 > 0$ 使得对 $\forall \lambda \in (0, \delta_2)$, $i \notin I$, 有 $g_i(x^* + \lambda d) > 0$,
最后存在 $\delta_3 > 0$ 使得对任意 $\lambda \in (0, \delta_3)$, $i \in I$, 有 $g_i(x^* + \lambda d) > g_i(x^*)$,
从而 $d \in D$, D 是 x^* 处的可行方向锥。

于是 $G(x^*) \subset D$. 由定理 7.2.1, 立明。

■ 不等式约束的一阶最优化条件

$$\text{Min } (x-3)^2 + (y-2)^2$$

$$\text{s.t. } x^2 + y^2 \leq 5$$

$$x + y \leq 3$$

$$x, y \geq 0.$$

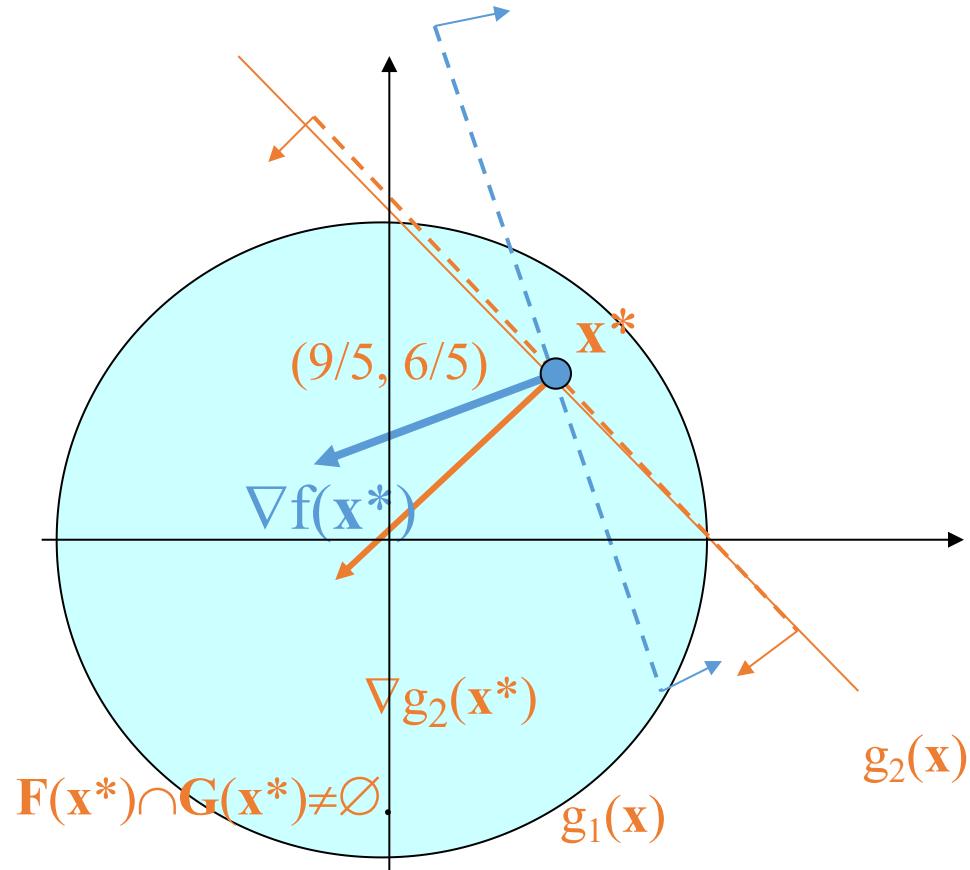
$$g_1(x) = -x^2 - y^2 + 5,$$

$$g_2(x) = -x - y + 3,$$

$$g_3(x) = x,$$

$$g_4(x) = y; \quad I = \{2\}$$

$$\nabla f(x) = [2(x-3), 2(y-2)]$$



■ 不等式约束的一阶最优化条件

$$\text{Min } (x-3)^2 + (y-2)^2$$

$$\text{s.t. } x^2 + y^2 \leq 5$$

$$x + y \leq 3$$

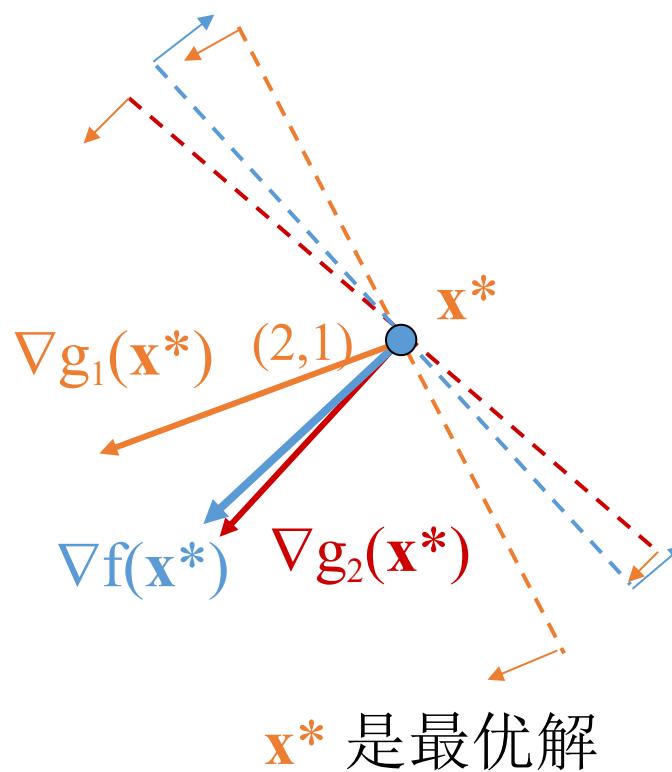
$$x, y \geq 0.$$

$$x^* = (2, 1)$$

$$\nabla g_1(x^*) = (-4, -2),$$

$$\nabla g_2(x^*) = (-1, -1),$$

$$\nabla f(x^*) = (-2, -2)$$



■ 不等式约束的一阶最优化条件

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

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

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

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

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

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

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

■ 不等式约束的一阶最优化条件

证明. 由Th7.2.2, 不存在向量 \mathbf{d} 同时满足 $\nabla f(\mathbf{x}^*)\mathbf{d} < 0$ 和 $-\nabla g_i(\mathbf{x}^*)\mathbf{d} < 0$, ($\forall i \in I$). 设 \mathbf{A} 是其行由 $\nabla f(\mathbf{x}^*)$ 和 $-\nabla g_i(\mathbf{x}^*)(i \in I)$ 组成的矩阵. 则 $\mathbf{A}\mathbf{d} < 0$ 无解.

由Godan定理, 必存在非零向量 $\mathbf{p} \geq 0$ 使得 $\mathbf{A}\mathbf{p} = 0$. 记 \mathbf{p} 的分量为 u_0 和 $u_i (i \in I)$, 则定理第一部分得证. 第二部分让 $u_i = 0 (i \notin I)$ 立明.

在 Fritz John 条件中, u_0, u_i 称为 Lagrangian 乘子.

而条件 $u_i g_i(\mathbf{x}^*) = 0$ 称为互补松弛条件(Complementary Slackness Condition).

■ 不等式约束的一阶最优化条件

例1 已知 $x^* = (3, 1)'$ 是下列问题的最优解：

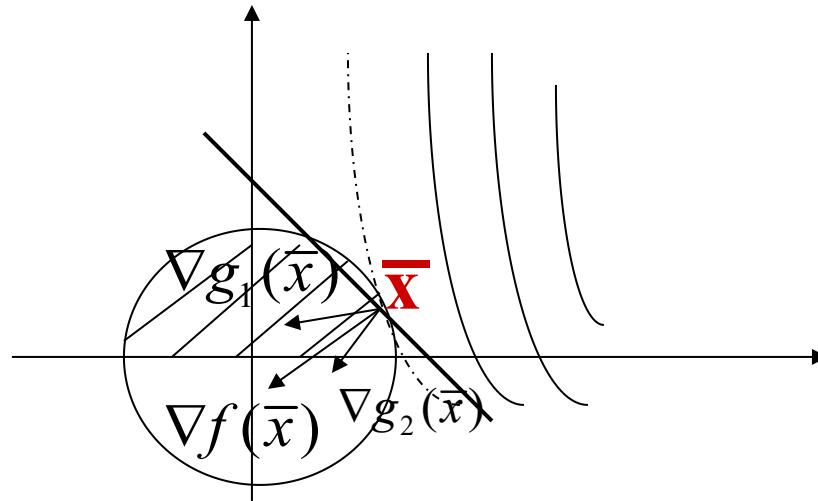
$$\min f(x) = (x_1 - 7)^2 + (x_2 - 3)^2$$

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

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

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

验证满足FJ条件.



■ 不等式约束的一阶最优化条件

在点 $\bar{x} = (3,1)'$ 处, 前两个约束是起作用约束, 即 $I = \{1,2\}$, 计算在 \bar{x} 处目标函数及起作用约束函数的梯度, 有

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

$$\text{设 } w_0 \begin{bmatrix} -8 \\ -4 \end{bmatrix} - w_1 \begin{bmatrix} -6 \\ -2 \end{bmatrix} - w_2 \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

此方程有分量不全为0的非负解, 如 $w=(1,1,2)$
于是在 \bar{x} 处满足FJ条件。

■ 不等式约束的一阶最优化条件

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。

在点 \bar{x} 处，两约束都起作用。目标函数及约束函数的梯度为：

■ 不等式约束的一阶最优化条件

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

$$\text{设 } w_0 \begin{bmatrix} 0 \\ -1 \end{bmatrix} - w_1 \begin{bmatrix} -2 \\ 0 \end{bmatrix} - w_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

此方程有解 $w = (0, k, 2k)$ ， k 可取任何正数
于是在 \bar{x} 处满足 FJ 条件。

■ 不等式约束的一阶最优化条件

若 **Lagrangian 乘子** $u_0 = 0$, 则 **Fritz John 条件** 不包含 $f(\mathbf{x})$ 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] 独立给出的最优化必要条件恰是 **Fritz John 条件** 加上 $u_0 > 0$ 。

■ 不等式约束的一阶最优化条件

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

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

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. ∇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

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

若还有 $g_i (i \notin I)$ 在 x^* 可微, 则

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

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

■ 不等式约束的一阶最优化条件

Karush-Kuhn-Tucker

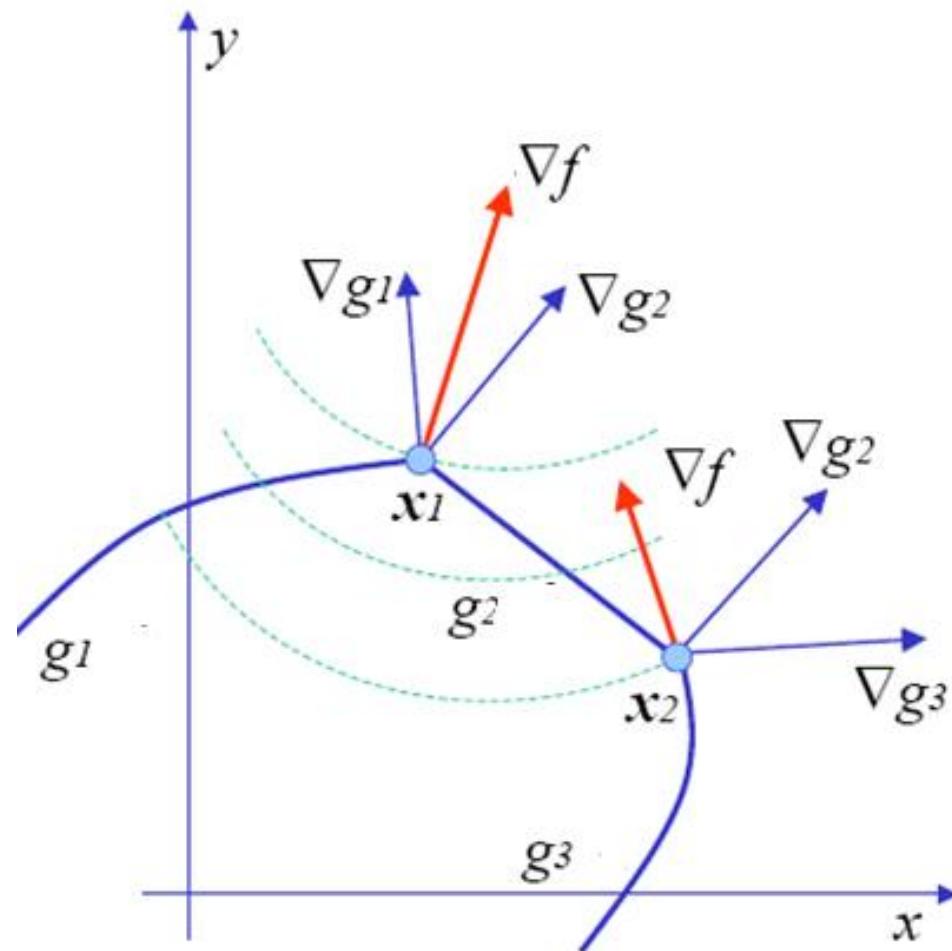
条件可写成向量形式

$$\nabla f(\mathbf{x}^*) - \mathbf{u} \nabla g(\mathbf{x}^*) = \mathbf{0},$$

$$\mathbf{u}g(\mathbf{x}^*) = \mathbf{0},$$

$$\mathbf{u} \geq \mathbf{0}.$$

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化条件

例3给定非线性规划

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

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

$$-x_1 + x_2 \geq 0$$

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化条件

记 $f(x) = (x_1 - 2)^2 + x_2^2$

$$g_1(x) = x_1 - x_2^2$$

$$g_2(x) = -x_1 + x_2$$

目标函数和约束函数的梯度是

$$\nabla f(x) = \begin{bmatrix} 2(x_1 - 2) \\ 2x_2 \end{bmatrix}, \quad \nabla g_1(x) = \begin{bmatrix} 1 \\ -2x_2 \end{bmatrix}, \quad \nabla g_2(x) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

先验证 $x^{(1)}$ 在这一点, $g_1(x) \geq 0$, $g_2(x) \geq 0$ 都是起作用约束
目标函数和约束函数的梯度分别是

$$\nabla f(x^{(1)}) = \begin{bmatrix} -4 \\ 0 \end{bmatrix}, \quad \nabla g_1(x^{(1)}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \nabla g_2(x^{(1)}) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

■ 不等式约束的一阶最优化条件

设 $\begin{bmatrix} -4 \\ 0 \end{bmatrix} - w_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - w_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

解此方程组, 得 $w_1 = -4, w_2 = 0$ 。

由于 $w_1 < 0$, 故不是KKT点。

再验证 $x^{(2)}$, 在这一点, $g_1(x) \geq 0, g_2(x) \geq 0$ 都是起作用约束
目标函数和约束函数的梯度分别是:

$$\nabla f(x^{(2)}) = \begin{bmatrix} -2 \\ 2 \end{bmatrix}, \quad \nabla g_1(x^{(2)}) = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad \nabla g_2(x^{(2)}) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

■ 不等式约束的一阶最优化条件

设 $\begin{bmatrix} -2 \\ 2 \end{bmatrix} - w_1 \begin{bmatrix} 1 \\ -2 \end{bmatrix} - w_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

解此方程组，得 $w_1=0, w_2=2$ 。

故它是KKT点。

■ 不等式约束的一阶最优化条件

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

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

其中 S 是 E^n 中非空开集. 设 x^* 为可行点, $I = \{i \mid g_i(x^*) = 0\}$.

设 $f(x)$ 和 贡 $-g_i$ 是凸的, 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

■ 小结

- 无约束问题的极值条件
- 约束极值问题的最优化条件
- 对偶及鞍点*