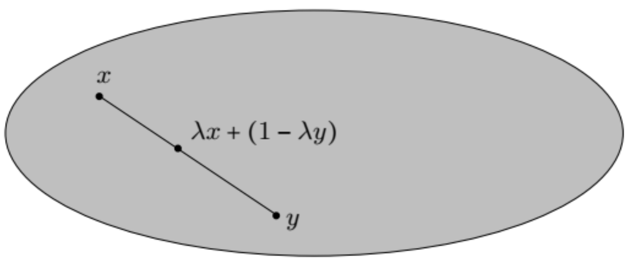
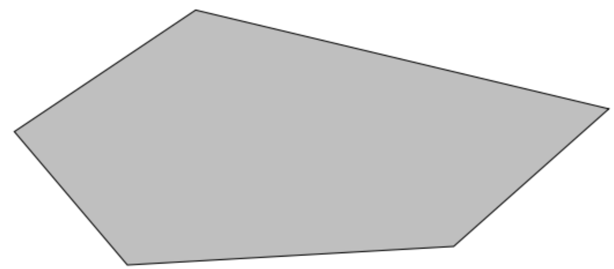
凸集与凸函数

凸集 对任意的 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{S}$ 和任意的 $\lambda \in [0,1]$, 都有

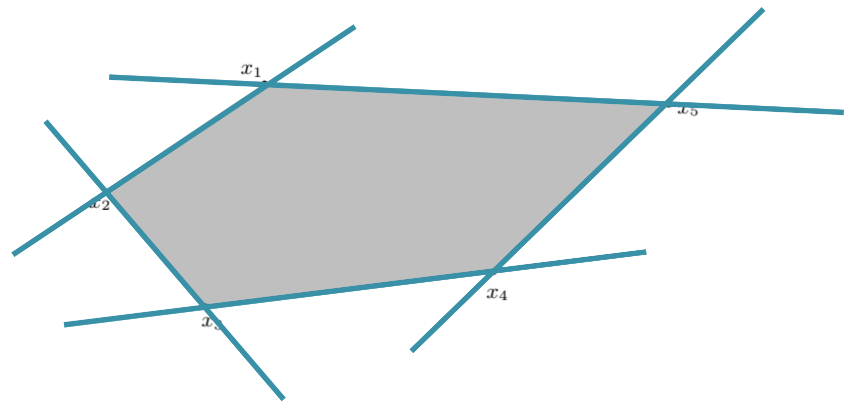
$$\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in \mathcal{S}$$





凸组合 设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m \in \mathbb{R}^n$, $\lambda_i \geqslant 0, i = 1, 2, \cdots, m$ 满足 $\lambda_1 + \lambda_2 + \cdots + \lambda_m = 1$, 称 $\lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2 + \cdots + \lambda_m \boldsymbol{x}_m$ 为 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$ 的一个凸组合.

凸包 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$ 的所有凸组合构成的集合称为由其生成的凸包.



凸包： 含有限个顶点的闭凸集, 又称多面胞.

$$S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{b}\}$$

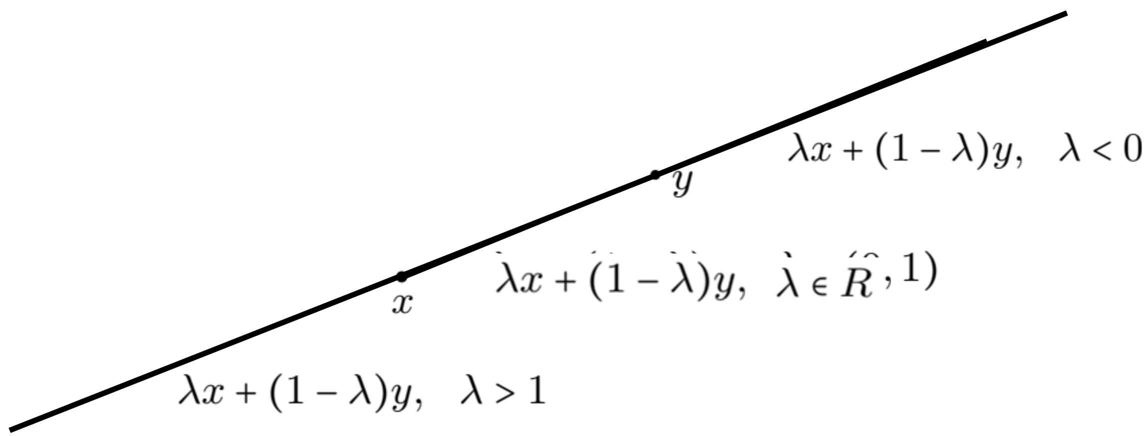
凸多面体

有界的凸多面体就是多面胞

仿射集 设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m \in \mathbb{R}^n$, $\lambda_i \in \mathbb{R}, i = 1, 2, \cdots, m$ 满足

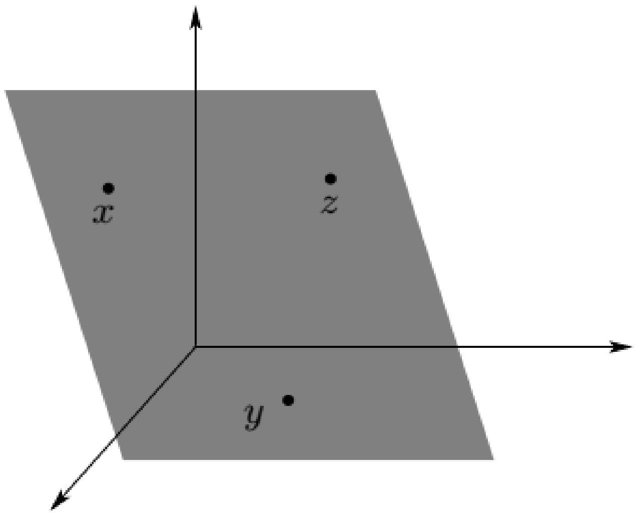
$\lambda_1 + \lambda_2 + \cdots + \lambda_m = 1$,则称 $\lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2 + \cdots + \lambda_m \boldsymbol{x}_m$

为 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$ 的仿射组合。



两个点的仿射组合是一条直线！

三个点的仿射组合是一个平面



若仿射集包含原点，则是一子空间。否则称为仿射子空间。

仿射组合 $\lambda_1\boldsymbol{x}_1+\lambda_2\boldsymbol{x}_2+\cdots+\lambda_m\boldsymbol{x}_m$

$$\boldsymbol{x}_1+\lambda_2(\boldsymbol{x}_2-\boldsymbol{x}_1)+\cdots+\lambda_m(\boldsymbol{x}_m-\boldsymbol{x}_1)\ ,\ \lambda_2,\cdots,\lambda_m\in\mathbb{R}$$

$\lambda_2(\boldsymbol{x}_2-\boldsymbol{x}_1)+\cdots+\lambda_m(\boldsymbol{x}_m-\boldsymbol{x}_1),\quad \lambda_2,\cdots,\lambda_m\in\mathbb{R}$

线性组合

仿射集是子空间的一个平移.

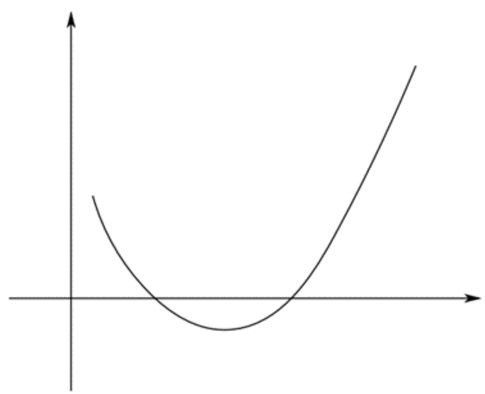
对仿射集 $\mathcal{S}$ 中任一点 $\boldsymbol{x}$, 集合 $\mathcal{S}-\{\boldsymbol{x}\}$ 为 $\mathbb{R}^n$ 的子空间.

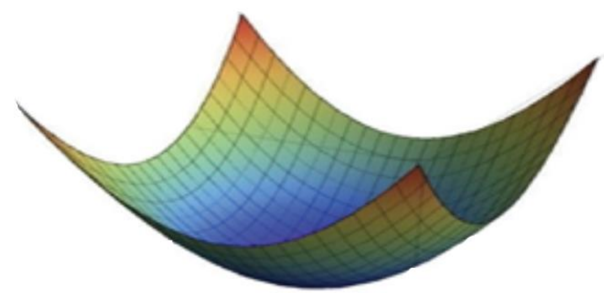
从外形看，子空间和仿射集是一样的，只是一个通过原点，另一个不通过原点.

子空间 $\mathcal{S}-\{\boldsymbol{x}\}$ 的维数称为仿射集 $\mathcal{S}$ 的维数.

凸函数

$$f\left(\lambda \boldsymbol{x}+(1-\lambda)\boldsymbol{y}\right)\leqslant \lambda f(\boldsymbol{x})+(1-\lambda)f(\boldsymbol{y}),\quad \forall\ \boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n,\ \lambda\in[0,1]$$





凸函数的等价定义

$$f\left(\lambda \boldsymbol{x}+(1-\lambda)\boldsymbol{y}\right)\leqslant \lambda f(\boldsymbol{x})+(1-\lambda)f(\boldsymbol{y}),\quad \forall\ \boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n,\ \lambda\in[0,1]$$

$$f\left(\sum_{i=1}^{m}\lambda_i\boldsymbol{x}_i\right)\leqslant\sum_{i=1}^{m}\lambda_i f(\boldsymbol{x}_i),\ \forall\boldsymbol{x}_i\in\mathbb{R}^n,\ \lambda_i\geqslant 0,\ i=1,2,\cdots,m,\ \sum_{i=1}^{m}\lambda_i=1$$

凸组合 设 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m\in\mathbb{R}^n$,

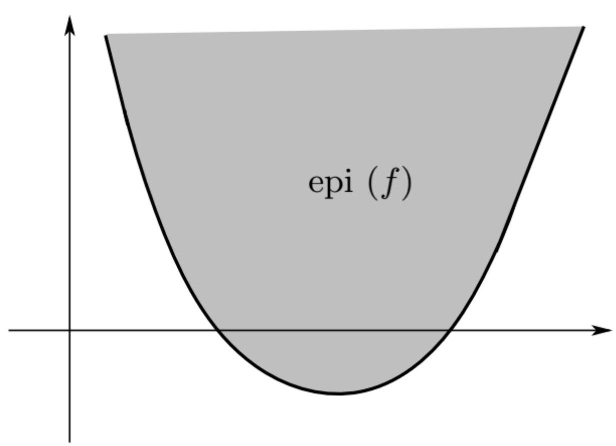
$$\lambda_1+\lambda_2+\cdots+\lambda_m=1,\quad \lambda_1,\lambda_2,\cdots,\lambda_m\in(0,1)$$

则称 $\lambda_1\boldsymbol{x}_1+\lambda_2\boldsymbol{x}_2+\cdots+\lambda_m\boldsymbol{x}_m$ 为 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m$ 的一个凸组合.

凸函数的等价定义

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leqslant \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}), \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,\ \lambda \in [0,1]$$

- $\mathrm{epi}(f) = \{(\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid \alpha \geqslant f(\boldsymbol{x})\}$ 是凸集。



凸函数的等价定义

凸函数 $f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leqslant \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}), \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,\ \lambda \in [0,1]$

连续可微

理论分析

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geqslant \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}), \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

或 $(\nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}))^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) \geqslant 0, \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

二阶连续可微

$$\boldsymbol{h}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}) \boldsymbol{h} \geqslant 0, \quad \forall\ \boldsymbol{x}, \boldsymbol{h} \in \mathbb{R}^n \longrightarrow f''(x) \geqslant 0$$

Hesse阵半正定

判定准则

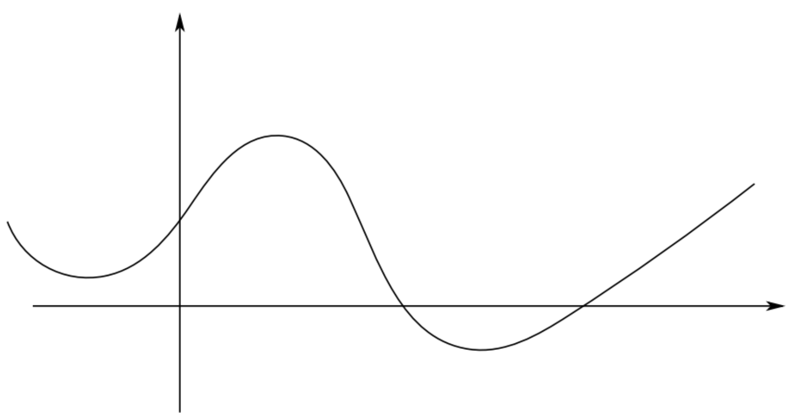
严格凸函数： 取严格不等号

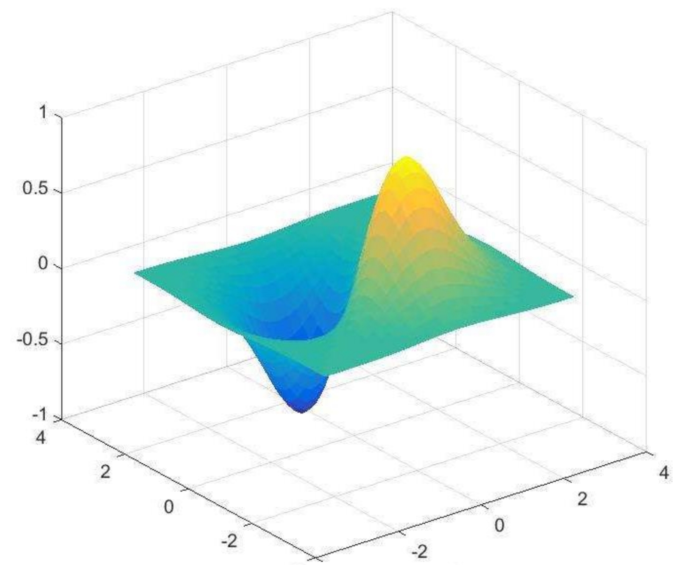
无约束优化最优性条件

光滑优化最优性条件

数学模型 $$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x})$$

$f:\mathbb{R}^n\to\mathbb{R}$ 连续可微(光滑函数).





(b)

最优解判定：直观地，验证最优解，需将其和附近点的目标函数值逐一比较，工作量太大.

利用目标函数的连续可微性质，可建立一个简易的判断方法，这就是光滑优化问题的最优性条件.

连续可微

无约束优化问题 $\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x})$

定理 $\boldsymbol{x}^*$是(局部)最优解，则 $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.

证明： 反证法。取 $\boldsymbol{d} = -\nabla f(\boldsymbol{x}^*), \alpha > 0$ 充分小。则

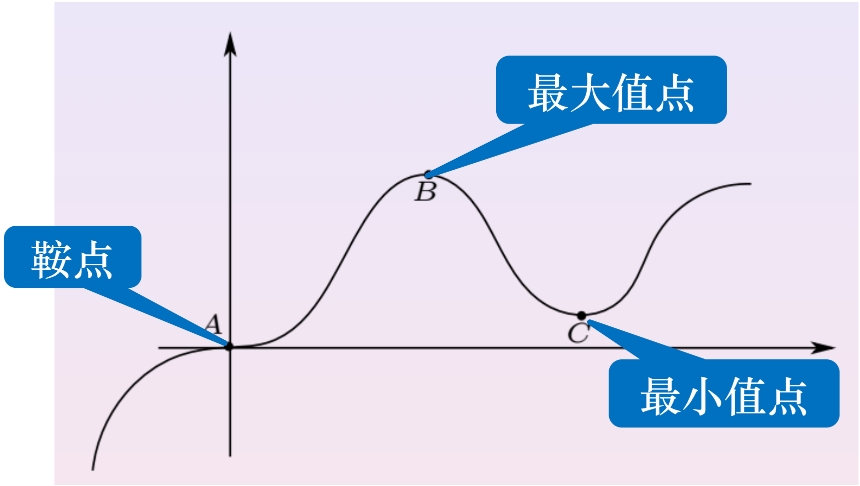
$$f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) = f(\boldsymbol{x}^*) + \alpha \nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{d} + o(\alpha)$$

$$= f(\boldsymbol{x}^*) - \alpha \|\nabla f(\boldsymbol{x}^*)\|^2 + o(\alpha)$$

$$< f(\boldsymbol{x}^*).$$

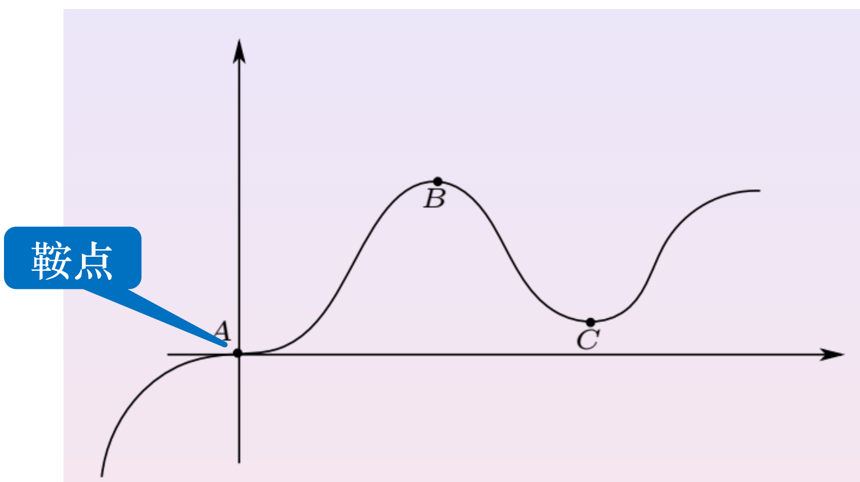
■

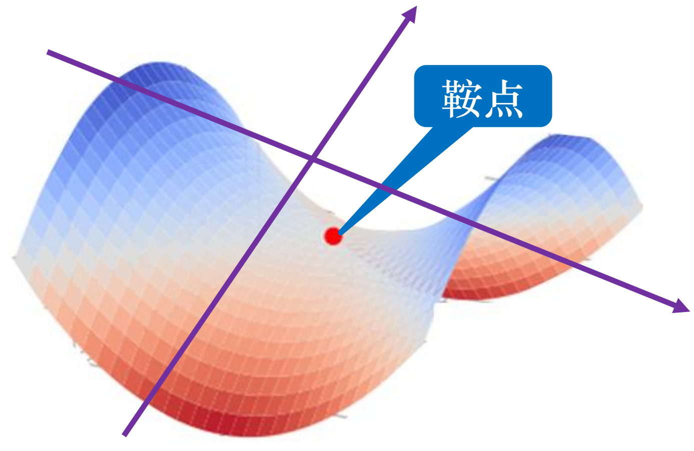
梯度零点称为优化问题的**稳定点**. 它可能是最大值点，可能是最小值点, 也可能二者都不是。



鞍点：梯度为零，且从该点出发的一个方向上是最大值点，另一个方向上是最小值点.

鞍点：比上不足比下有余！





马鞍面

问题：什么条件下，稳定点是最优值点？

光滑优化最优性条件（续）

凸优化 $\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x})$

$f:\mathbb{R}^n \to \mathbb{R}$ 为连续可微的凸函数。

定理 凸优化问题的稳定点为其全局最优解。

即若 $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$，则 $\boldsymbol{x}^*$为优化问题的全局最小值点。

证明： 利用凸函数的性质

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \geqslant \langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle = 0.$$

无约束优化二阶最优性条件

定理 设 $\boldsymbol{x}^*$ 为优化问题 $\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$ 的最优解,

则 $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$, $\nabla^2 f(\boldsymbol{x}^*)$ 半正定。

证明： 结论 $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ 已证。

反设 $\nabla^2 f(\boldsymbol{x}^*)$ 非半正定，则存在向量 $\boldsymbol{d}$ 使得

$$\boldsymbol{d}^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} < 0.$$

取 $\alpha > 0$ 充分小。由Taylor展式

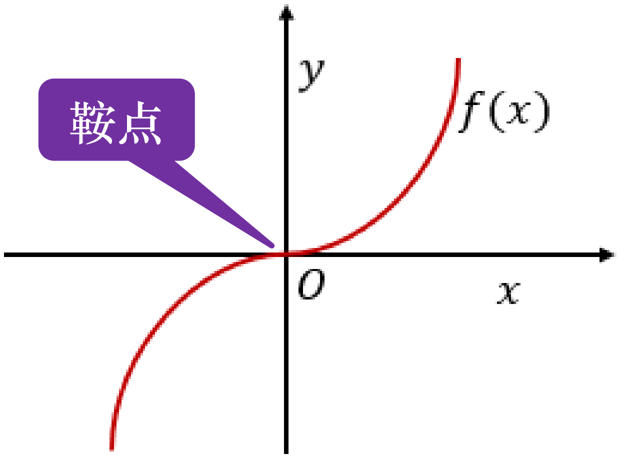
$$f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) = f(\boldsymbol{x}^*) + \alpha\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{d} + \frac{1}{2}\alpha^2\boldsymbol{d}^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} + o(\alpha^2)$$

$$< f(\boldsymbol{x}^*).$$

得矛盾。

二阶必要条件的非充分性

函数 $f(x)=x^3$ 在零点满足 $\nabla f(\boldsymbol{x}^*)=\mathbf{0}$, $\nabla^2 f(\boldsymbol{x}^*)$ 半正定，但零点不是其最优值点。



二阶充分条件

定理 对优化问题 $\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$ ，设$\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ ，$\nabla^2 f(\boldsymbol{x}^*)$ 正定，则 $\boldsymbol{x}^*$ 是该优化问题的严格最优解.

证明： 对任意充分靠近 $\boldsymbol{x}^*$ 的 $\boldsymbol{x}$，存在单位向量 $\boldsymbol{d}$ ，使得

$$\boldsymbol{x} = \boldsymbol{x}^* + \alpha \boldsymbol{d}$$

其中 $\alpha > 0$ 充分小。 则

$$f(\boldsymbol{x}) = f(\boldsymbol{x}^*) + \frac{1}{2}\alpha^2 \boldsymbol{d}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} + o(\alpha^2) > f(\boldsymbol{x}^*).$$

二阶充分条件的非必要性

上述结论给出的二阶充分性条件不是必要的. 如零点是单元函数 $f(x)=x^4$ 的严格最优解, 但上述二阶充分性条件在该点并不成立.

无约束优化问题不存在充分必要的最优性条件!

最优性条件的应用

例 $$\min_{x>0,y\geqslant 0} f(x,y)=\frac{10}{x}+\frac{(x-y)^2}{2x}+\frac{3y^2}{2x}$$

分析与求解：先忽略约束。

利用 $$\min_{x,y} f(x,y)=\min_{x}\min_{y} f(x,y)$$

先固定 x，关于 y 的内层优化，再求解关于 x 的外层优化

$$\min_{y} f(x,y)=\frac{10}{x}+\frac{(x-y)^2}{2x}+\frac{3y^2}{2x}$$

目标函数关于 y 为凸函数，利用最优性条件得最优解

$$y=\frac{1}{4}x$$

将上述最优解代入目标函数得外层优化问题

$$\min \ f(x) = \frac{10}{x} + \frac{3}{8}x$$

目标函数关于x 为凸函数。再利用最优性条件得

$$x = \tfrac{4}{3}\sqrt{15}$$

这样 $y = \tfrac{1}{4}x = \tfrac{1}{3}\sqrt{15}$

它们满足约束条件，自然为原问题的最优解。 ■