运筹学

9. 非线性规划基础

李 力
清华大学

Email: li-li@tsinghua.edu.cn

2023.4.

非线性规划部分主要内容

基础知识（凸优化问题，
凸函数，可行下降方法）

一维搜索（直线搜索）

无约束优化（下降方向）

约束优化（KT条件）

算法（简约梯度，罚函数，障碍函数）

基础知识

非线性规划的一般形式

$$
\begin{aligned}
\min\ & f(X) \\
\text{s.t.}\ & h_i(X)=0,\ i=1,2,\cdots,m \\
& g_j(X)\le 0,\ j=1,2,\cdots,l
\end{aligned}
$$

其中 $X=\left(x_1,x_2,\cdots,x_n\right)^T\in R^n$

定义可行集

$$
\Omega=\left\{X\in R^n \,\middle|\, h_i(X)=0,\ 1\le i\le m, g_j(X)\le 0, 1\le j\le l\right\}
$$

上述一般形式可简写成 $\min\limits_{X\in\Omega} f(X)$

$\hat{X}$ 的 ε 邻域: $B\left(\hat{X},\varepsilon\right)=\left\{X\in R^n \,\middle|\, \left\|X-\hat{X}\right\|<\varepsilon\right\}$

局部最优解 $\hat{X}$： $\hat{X}\in\Omega$，且存在 $\varepsilon>0$ 满足

$$f\left(\hat{X}\right)\le f\left(X\right),\ \forall X\in B\left(\hat{X},\varepsilon\right)\cap\Omega$$

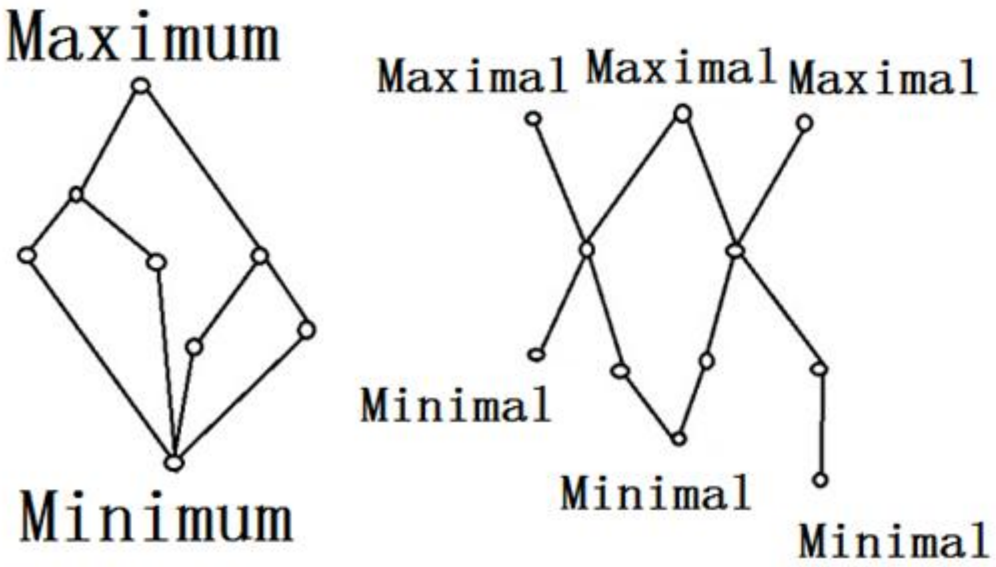
全局最优解 $\hat{X}$： $\hat{X}\in\Omega,\ f\left(\hat{X}\right)\le f\left(X\right),\ \forall X\in\Omega$

严格局部最优条件 $f\left(\hat{X}\right)<f\left(X\right),\ \forall X\in B\left(\hat{X},\varepsilon\right)\cap\Omega-\left\{\hat{X}\right\}$

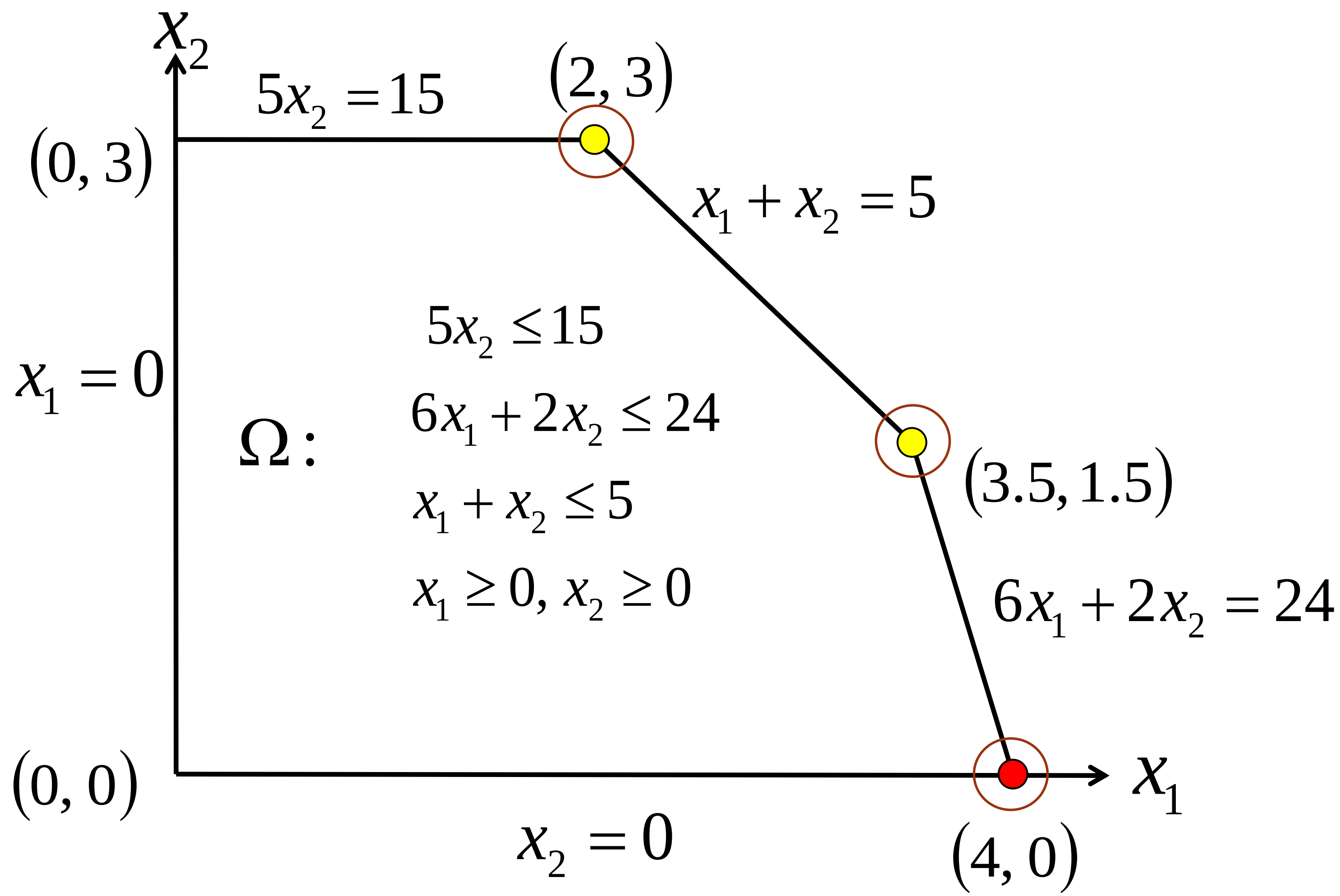
严格全局最优条件 $f\left(\hat{X}\right)<f\left(X\right),\ \forall X\in\Omega-\left\{\hat{X}\right\}$

函数的局部最小/大值（Minimal/Maximal）

和全局最小/大值（Minimum/Maximum）



$$\min_{X \in \Omega} \; -x_1^{\,2} - x_2^{\,2}$$



邻域　局部最优解　全局最优解

如何判断找到了局部或全局最优解？

什么样的非线性目标函数可以方便的判定是否找到了局部或全局最优解？

为此，人们开始研究凸优化问题

为什么研究凸优化？

1. 现实生活中，很多问题都属于凸优化问题。例如线性规划就是凸优化问题的特例。

2. 凸优化问题性质好，可以进行深入的理论分析。很大一部分凸优化问题是多项式时间可解问题问题，我们已经建立了十分有效的求解算法，可以快速求得全局最优解。

3. 凸优化是研究连续变量优化的起点和基础。目前很多非凸优化问题中非凸性的刻画都脱胎于凸优化，相关问题的求解也和凸优化联系在一起，常常有赖于找到这些非凸优化问题中“凸”的结构 。

非线性优化问题的定义和最优解要涉及以下知识：

一元函数的一阶/二阶导数、多元函数的梯度和Hesse阵

凸函数的定义及其性质

- 多元凸函数的基本定义和性质
- 凸函数的局部最优解和全局最优解
- 多元凸函数的一阶充要条件
- 多元凸函数的二阶充分条件和必要条件
- 多元凸函数和一元凸函数的关系
- 典型的凸函数

标量函数求偏导数（梯度）

$$\nabla f(X) = \frac{\partial f(X)}{\partial X} = \begin{pmatrix} \dfrac{\partial f(X)}{\partial x_1} \\ \dfrac{\partial f(X)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(X)}{\partial x_n} \end{pmatrix}$$

$$\nabla^T f(X) = \frac{\partial f(X)}{\partial X^T} = \left(\frac{\partial f(X)}{\partial x_1}, \frac{\partial f(X)}{\partial x_2}, \cdots, \frac{\partial f(X)}{\partial x_n} \right)$$

典型函数 $f(x1,x2) = x1^2 + 2\ x1x2 + x2^2$

向量函数求偏导数

$$F(X)=\left(f_1(X),f_2(X),\cdots,f_m(X)\right)^T$$

$$\frac{\partial F^T(X)}{\partial X}=\left(\frac{\partial f_1(X)}{\partial X},\frac{\partial f_2(X)}{\partial X},\cdots,\frac{\partial f_m(X)}{\partial X}\right)_{n\times m}$$
$$=\left(\nabla f_1(X),\nabla f_2(X),\cdots,\nabla f_m(X)\right)_{n\times m}$$

$$\frac{\partial F(X)}{\partial X^T}=\begin{pmatrix}\dfrac{\partial f_1(X)}{\partial X^T}\\ \dfrac{\partial f_2(X)}{\partial X^T}\\ \vdots\\ \dfrac{\partial f_m(X)}{\partial X^T}\end{pmatrix}_{m\times n}=\begin{pmatrix}\nabla^T f_1(X)\\ \nabla^T f_2(X)\\ \vdots\\ \nabla^T f_m(X)\end{pmatrix}_{m\times n}$$

海赛（Hesse，Hessian）矩阵

$$\nabla^2 f(X)=\frac{\partial \nabla f(X)}{\partial X^T}=\begin{bmatrix} \frac{\partial^2 f(X)}{\partial x_1 x_1} & \frac{\partial^2 f(X)}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(X)}{\partial x_1 x_n} \\ \frac{\partial^2 f(X)}{\partial x_2 x_1} & \frac{\partial^2 f(X)}{\partial x_2 x_2} & \cdots & \frac{\partial^2 f(X)}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(X)}{\partial x_n x_1} & \frac{\partial^2 f(X)}{\partial x_n x_2} & \cdots & \frac{\partial^2 f(X)}{\partial x_n x_n} \end{bmatrix}$$

黑塞矩阵（Hessian Matrix），又译作海森矩阵、海瑟矩阵、海塞矩阵等，是一个多元函数的二阶偏导数构成的方阵，描述了函数的局部曲率。黑塞矩阵最早于19世纪由德国数学家Ludwig Otto Hesse提出。

对向量函数的点积求偏导数

$$F(X)=\left(f_1(X), f_2(X),\cdots, f_m(X)\right)^T$$

$$G(X)=\left(g_1(X), g_2(X),\cdots, g_m(X)\right)^T$$

$$\frac{\partial\left(F^T(X)G(X)\right)}{\partial X}=\frac{\partial F^T(X)}{\partial X}G(X)+\frac{\partial G^T(X)}{\partial X}F(X)$$

$$\frac{\partial\left(F^T(X)G(X)\right)}{\partial X^T}=F^T(X)\frac{\partial G(X)}{\partial X^T}+G^T(X)\frac{\partial F(X)}{\partial X^T}$$

对常数矩阵和向量函数的乘积求偏导数

$$A \in R^{m\times m} \qquad F(X) = \left(f_1(X), f_2(X), \cdots, f_m(X)\right)^T$$

$$\frac{\partial \left(AF(X)\right)^T}{\partial X} = \frac{\partial \left(F^T(X)A^T\right)}{\partial X} = \frac{\partial F^T(X)}{\partial X}A^T$$

$$\frac{\partial AF(X)}{\partial X^T} = A\frac{\partial F(X)}{\partial X^T}$$

对二次函数求偏导数（$A^T = A \in R^{m\times m}$）

$$\frac{\partial \left(X^T AX\right)}{\partial X} = \frac{\partial X^T}{\partial X}AX + \frac{\partial (AX)^T}{\partial X}X = \left(A + A^T\right)X = 2AX$$

$$\frac{\partial \left(X^T AX\right)}{\partial X^T} = X^T\frac{\partial (AX)}{\partial X^T} + (AX)^T\frac{\partial X}{\partial X^T} = X^T\left(A + A^T\right) = 2X^T A$$

一元函数在原点的二阶泰勒（Taylor）展开

$$g(t)=g(0)+c_1 t+\frac{1}{2}c_2(\xi)t^2$$

需要假设存在连续二阶导数

其中 $0\le\xi\le t$

$$c_1=\left.\frac{dg(t)}{dt}\right|_{t=0}$$

$$c_2(\xi)=\left.\frac{d^2g(t)}{dt^2}\right|_{t=\xi}$$

我们经常通过类比一元函数和多元函数来增强感性理解

多元函数在给定点沿给定方向的二阶泰勒展开

$$f(\hat{X}+tD)=f(\hat{X})+c_1 t+\frac{1}{2}c_2(\xi)t^2$$

需要假设存在连续二阶导数

其中 $0\le\xi\le t$

$$c_1=\left.\frac{df(\hat{X}+tD)}{dt}\right|_{t=0}=\nabla^T f(\hat{X})D$$

$$c_2(\xi)=\left.\frac{d^2 f(\hat{X}+tD)}{dt^2}\right|_{t=\xi}=D^T\nabla^2 f\left(\hat{X}+\xi D\right)D$$

$$\Rightarrow\quad f(\hat{X}+tD)=f(\hat{X})+\nabla^T f(\hat{X})Dt+\frac{1}{2}D^T\nabla^2 f(\hat{X}+\xi D)Dt^2$$

（凸集上的）凸函数和凹函数

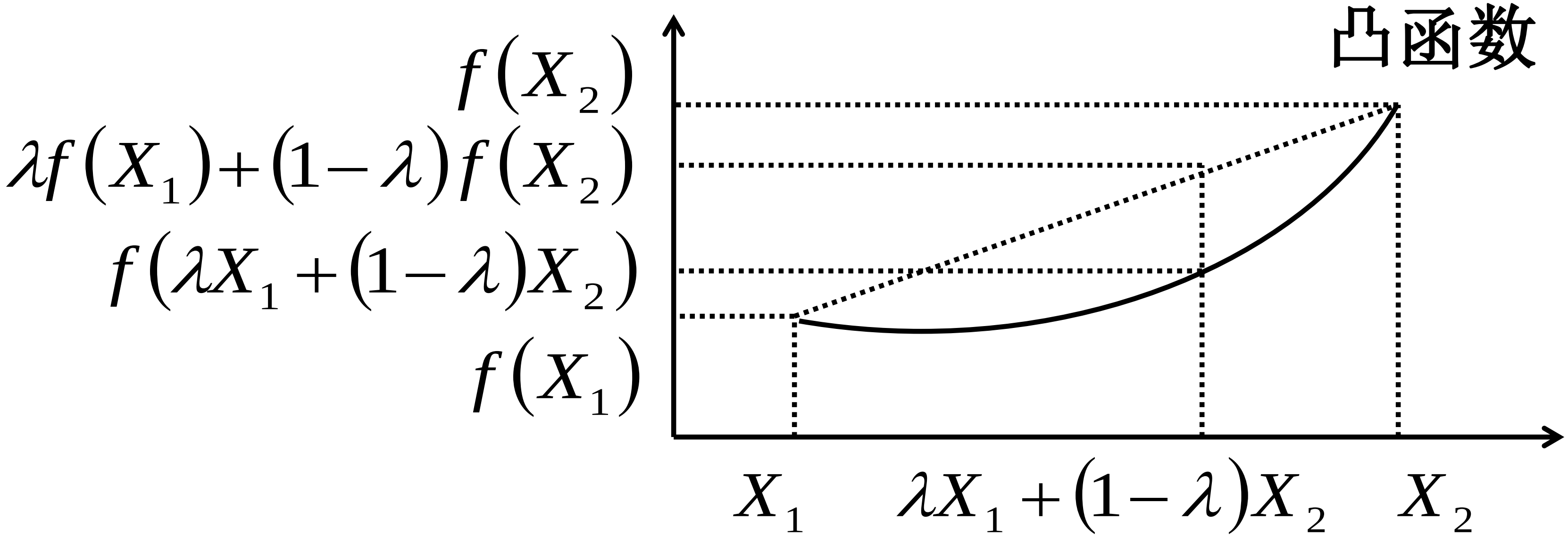
设 $f(X)$ 是定义在集合 $\Omega \subset R^n$ 上的函数，如果 Ω 是凸集，并且对 Ω 中任意两点 X_1, X_2 以及闭区间 $[0,1]$ 中任意一点 λ 都满足

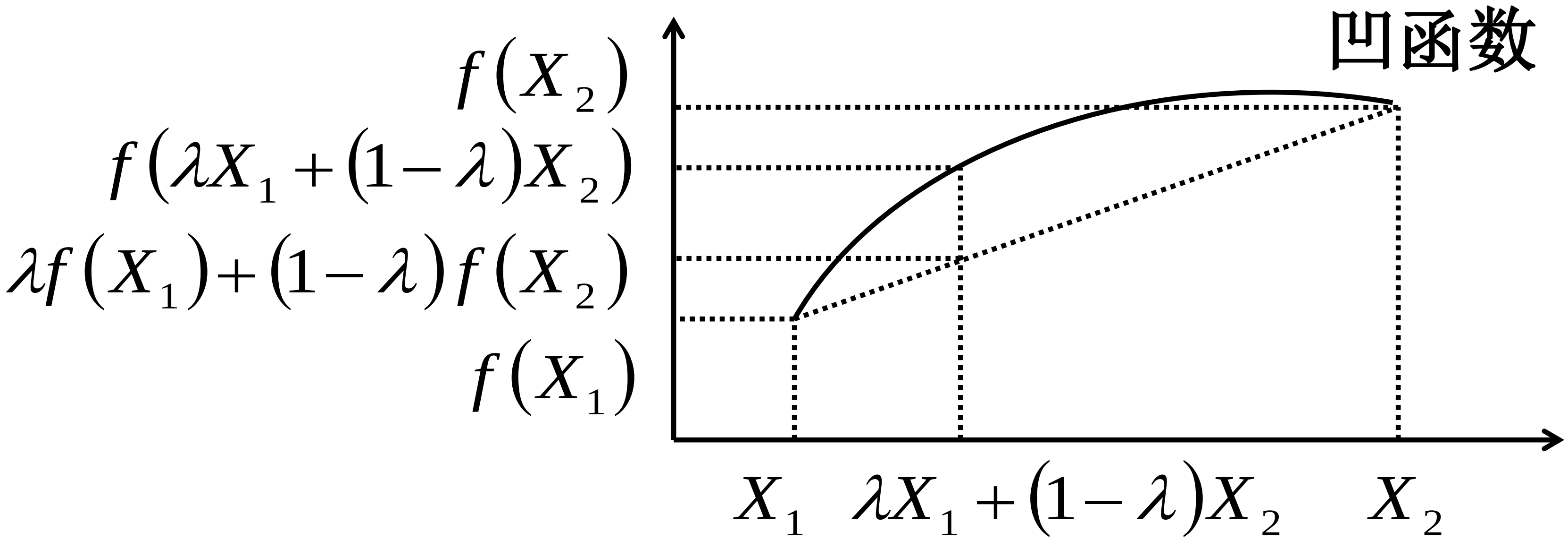
$$f\left(\lambda X_1 + (1-\lambda)X_2\right) \le \lambda f\left(X_1\right) + (1-\lambda) f\left(X_2\right)$$

则称 $f(X)$ 是（凸集 Ω上的）凸函数，如果 $-f(X)$ 是（凸集 Ω 上的）凸函数，则称 $f(X)$ 是（凸集 Ω 上的）凹函数，此时在上面的条件下应满足

$$f\left(\lambda X_1 + (1-\lambda)X_2\right) \ge \lambda f\left(X_1\right) + (1-\lambda) f\left(X_2\right)$$

一元凸（凹）函数的图象


凸函数
$f(X_2)$
$\lambda f(X_1)+(1-\lambda)f(X_2)$
$f(\lambda X_1+(1-\lambda)X_2)$
$f(X_1)$
X_1
$\lambda X_1+(1-\lambda)X_2$
X_2


凹函数
$f(X_2)$
$f(\lambda X_1+(1-\lambda)X_2)$
$\lambda f(X_1)+(1-\lambda)f(X_2)$
$f(X_1)$
X_1
$\lambda X_1+(1-\lambda)X_2$
X_2

多元凸函数和一元凸函数的关系

对定义在凸集 $\Omega \subset R^n$ 上的多元函数 $f(\cdot)$，任取 $X \in \Omega$ 和 $D \in R^n$ 定义一元函数 $\varphi(\cdot \mid X, D)$ 如下

$$\varphi(t \mid X, D) = f(X + tD), \ \forall X + tD \in \Omega$$

则存在以下关系

当且仅当

$f(\cdot)$ 是凸函数 $\Leftrightarrow$ $\varphi(\cdot \mid X, D), \forall X \in \Omega, D \in R^n$ 是凸函数

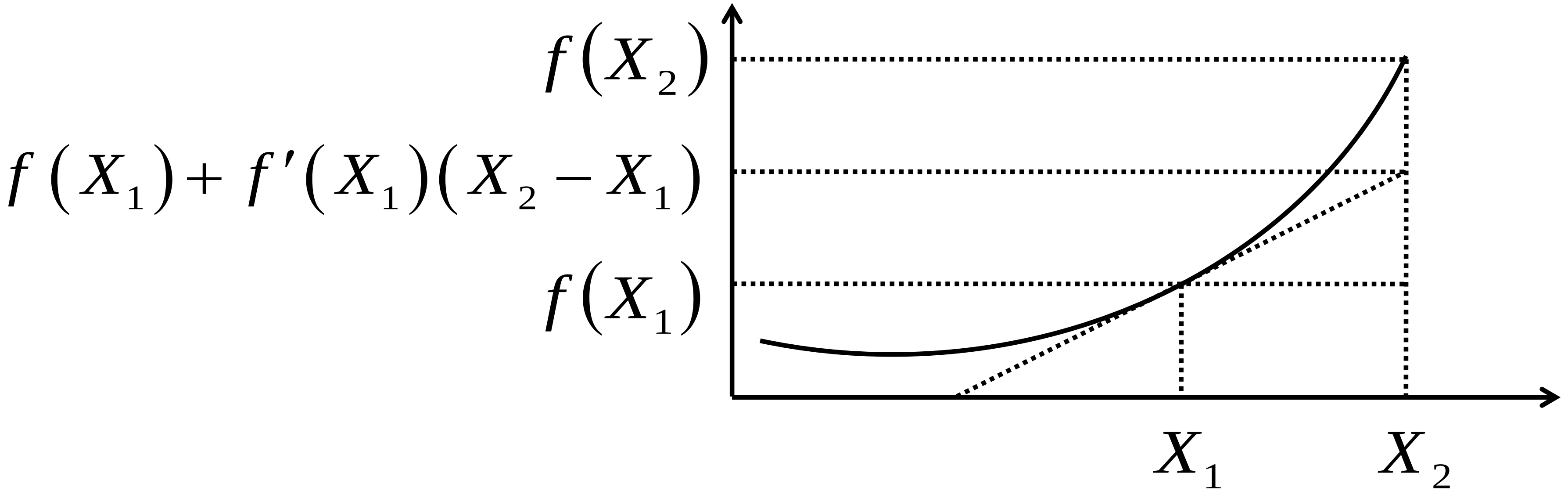
证明：正向利用

$$\varphi\left((1-\lambda)t_1 + \lambda t_2 \mid X, D\right) = f\left((1-\lambda)(X + t_1 D) + \lambda(X + t_2 D)\right)$$

反向利用 $f\left((1-t)X_1 + tX_2\right) = \varphi\left((1-t)\times 0 + t \times 1 \mid X_1, X_2 - X_1\right)$

一元可导凸函数的充要条件

$$f'(X_1)(X_2 - X_1) \le f(X_2) - f(X_1)$$



充分性： $X(t)=(1-t)X_1+tX_2$

$$f(X_1)-f(X(t))\geq f'(X(t))(X_1-X(t))$$

$$f(X_2)-f(X(t))\geq f'(X(t))(X_2-X(t))$$

$$\Rightarrow\quad (1-t)\big(f(X_1)-f(X(t))\big)+t\big(f(X_2)-f(X(t))\big)\geq 0$$

$$\Leftrightarrow\quad (1-t)f(X_1)+tf(X_2)\geq f\big((1-t)X_1+tX_2\big)$$

必要性： $f\big((1-t)X_1+tX_2\big)\leq(1-t)f(X_1)+tf(X_2)$

$$\Leftrightarrow\quad \frac{f\big(X_1+t(X_2-X_1)\big)-f(X_1)}{t}\leq f(X_2)-f(X_1)$$

（只需要一阶导数存在）

多元可导凸函数的一阶充要条件

$$\nabla^T f(X_1)(X_2 - X_1) \le f(X_2) - f(X_1),\ \forall X_1, X_2 \in \Omega$$

充分性： $g(t_i) = f(X + t_i V) = f(X_i),\ i = 1, 2$

$$\Rightarrow \quad g'(t_1) = \nabla f(X + t_1 V)^T V$$

$$\Rightarrow \quad g(t_2) - g(t_1) = f(X_2) - f(X_1) \ge \nabla f(X_1)^T (X_2 - X_1)$$

$$= \nabla f(X + t_1 V)^T V(t_2 - t_1) = g'(t_1)(t_2 - t_1)$$

必要性： $g'(t) = \nabla f(X_1 + t(X_2 - X_1))^T (X_2 - X_1)$

$$g'(0) = \nabla f(X_1)^T (X_2 - X_1)$$

$$g'(0)(1-0) \le g(1) - g(0),\ \forall X_1, X_2 \in \Omega$$

一元二阶可导凸函数的二阶充分条件

$$f''(X) \geq 0, \ \forall X \in \Omega$$

证明

已知条件 $\Rightarrow$ $f'(X_2) \geq f'(X_1), \ \forall X_2 \geq X_1$

情况1 $X_2 \geq X_1 \, (\xi \geq X_1)$

$$f(X_2) - f(X_1) = f'(\xi)(X_2 - X_1) \geq f'(X_1)(X_2 - X_1)$$

情况2 $X_2 \leq X_1 \, (\xi \leq X_1)$

$$f(X_2) - f(X_1) = f'(\xi)(X_2 - X_1) \geq f'(X_1)(X_2 - X_1)$$

由一阶充要条件知结论成立

（只需要二阶导数存在）

一元二阶可导凸函数的二阶必要条件

$$f''(X)\geq 0,\ \forall X\in\Omega$$

其中 Ω 不是单点集

证明　必存在 $|\Delta X|\in B(0,\varepsilon)$ $\Rightarrow$ $X+\Delta X\in\Omega$

$$\Rightarrow\quad f'(X)(\Delta X)\leq f(X+\Delta X)-f(X)$$

$$f'(X+\Delta X)(-\Delta X)\leq f\left(X+\Delta X+(-\Delta X)\right)-f(X+\Delta X)$$

$$\Rightarrow\quad \left(f'(X+\Delta X)-f'(X)\right)\Delta X\geq 0$$

$$\Rightarrow\quad \frac{\left(f'(X+\Delta X)-f'(X)\right)(\Delta X)^2}{\Delta X}\geq 0\quad\Rightarrow\quad f''(X)\geq 0$$

多元二阶可导凸函数的二阶充分条件

$$\nabla^2 f(X) \geq 0, \ \forall X \in \Omega$$

证明途径

$$g(t) = f\left(X_1 + t(X_2 - X_1)\right)$$

$$g'(t) = \nabla f\left(X_1 + t(X_2 - X_1)\right)^T (X_2 - X_1)$$

$$g''(t) = (X_2 - X_1)^T \nabla^2 f\left(X_1 + t(X_2 - X_1)\right)(X_2 - X_1)$$

$$\nabla^2 f(X) \geq 0, \ \forall X \in \Omega \quad \Rightarrow \quad g''(t) \geq 0$$

$\Rightarrow$ $g(t)$ 是一元凸函数

多元二阶可导凸函数的二阶必要条件

$$\nabla^2 f(X) \geq 0,\ \forall X \in \Omega$$ 其中 Ω是开集

证明途径

对任意的 $X \in \Omega$ 和模充分小的 $\Delta X \in R^n$

$g(t) = f(X + t\Delta X)$ 是含零开区间上的凸函数

$$g'(t) = \nabla f(X_1 + t\Delta X)^T \Delta X$$

$$g''(t) = \Delta X^T \nabla^2 f(X + t\Delta X) \Delta X$$

$$\Delta X^T \nabla^2 f(X) \Delta X = g''(0) \geq 0 \quad \Rightarrow \quad \nabla^2 f(X) \geq 0$$

<u>凸性对优化问题的基本作用</u> （回答**Page 8**的问题）

如果 Ω 是凸集，$f\left(X\right)$ 是其上的连续凸函数，称

$$\min_{X\in\Omega} f\left(X\right)$$

是凸规划问题

如果 $X^{*}\in\Omega$ 是凸规划问题的任意一个局部最优

解，那么它也是该问题的全局最优解

证明：如果存在 $\hat{X} \in \Omega$ 满足 $f(\hat{X}) < f(X^*)$

$$\Rightarrow \qquad \lambda f(\hat{X}) < \lambda f(X^*), \quad \forall \lambda > 0$$

又因为

$$f\left(\lambda \hat{X} + (1-\lambda) X^*\right) \le \lambda f(\hat{X}) + (1-\lambda) f(X^*), \ \forall 0 \le \lambda \le 1$$

$$\Rightarrow \qquad f\left(\lambda \hat{X} + (1-\lambda) X^*\right) < f(X^*), \ \forall 0 < \lambda \le 1$$

因为对充分小的 $\lambda > 0$，$\lambda \hat{X} + (1-\lambda) X^*$ 能够充分接近 X^*，说明 X^* 不是局部最优解，矛盾

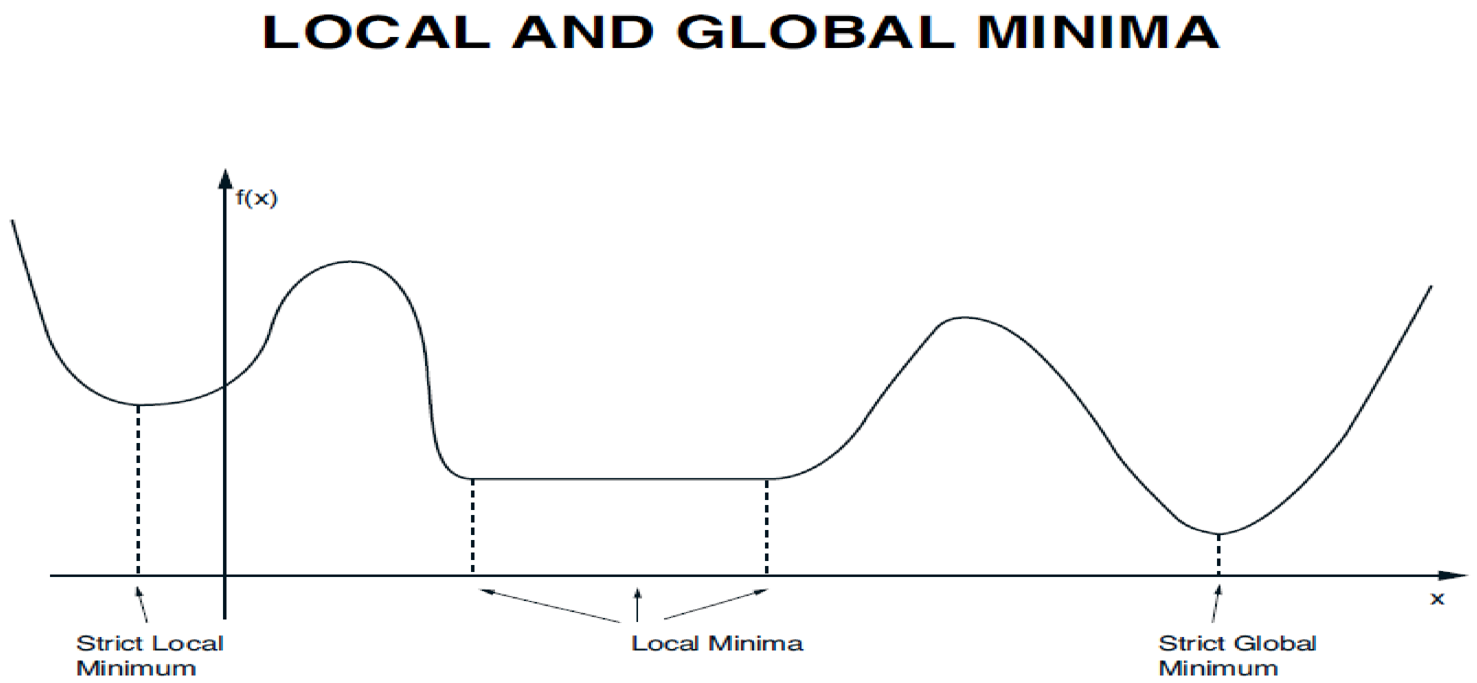
典型的凸函数包括：

- 指数函数 e^{ax}， 负对数函数
- 仿射函数（同时是凸函数和凹函数）
- 正定或者半正定二次函数 $x^T Ax + b^T x + c$
- 范数（无法利用凸性的一阶条件以及二阶条件进行证明，因为范数本身可能并不是处处可微的）
- 幂函数 x^p，绝对值幂函数 $|x|^p$（ $p \geq 1$ 为凸函数， $p \in (0,1)$ 为凹函数）

典型的凹函数包括：

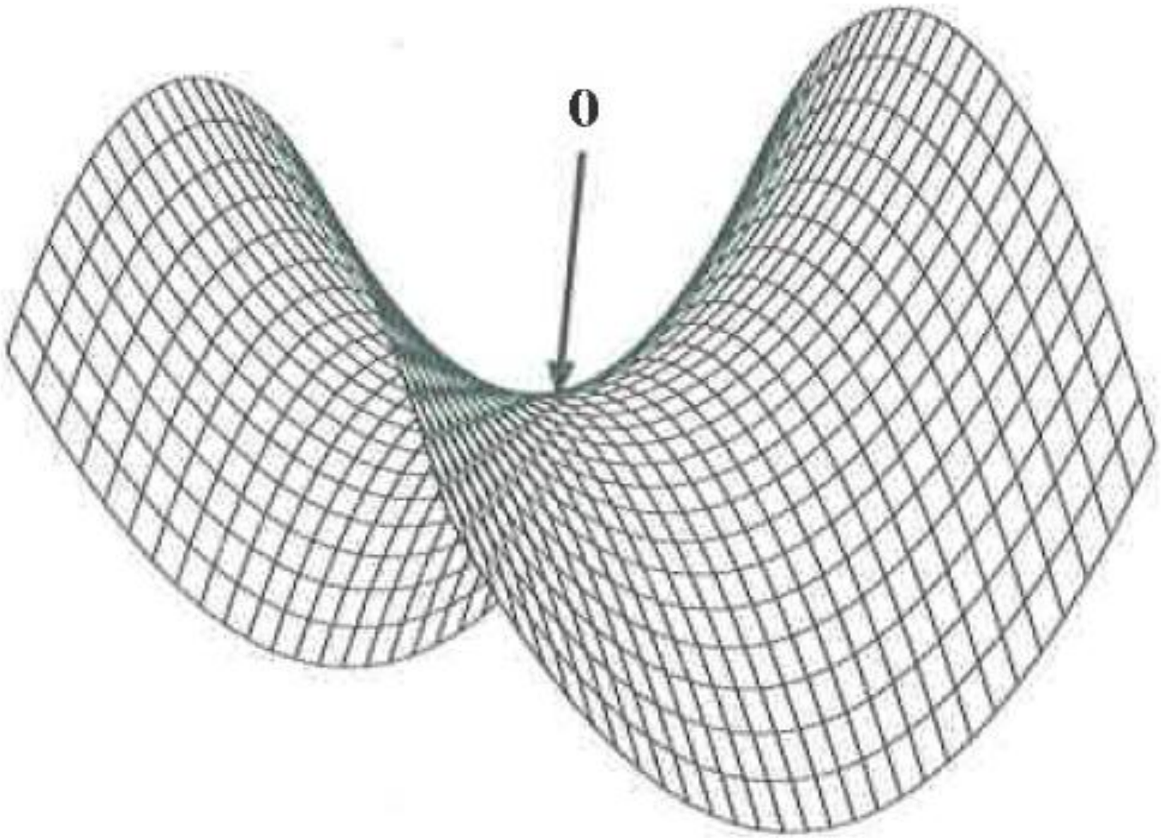
- 对数函数 $\log(x)$

非凸函数



Unconstrained local and global minima in one dimension.

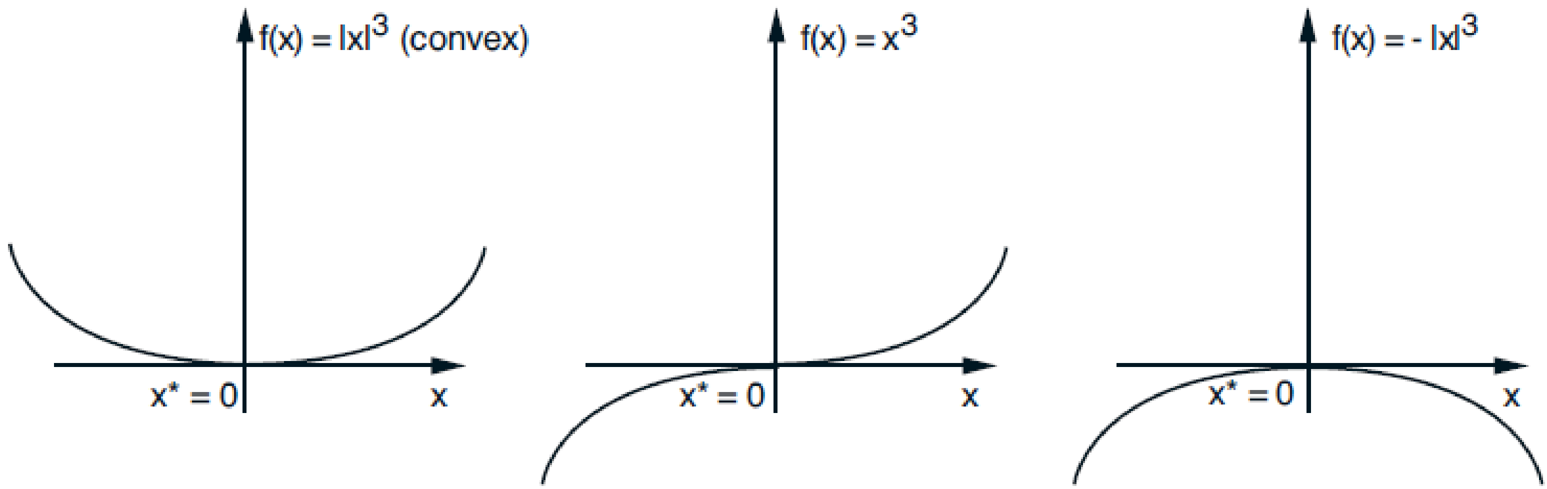
非凸函数



$$x_1^2 - x_2^2$$

非凸函数

- There may exist points that satisfy the 1st and 2nd order conditions but are not local minima



First and second order necessary optimality conditions for functions of one variable.