# 从统计学到机器学习的算法基础2
# 凸优化

《Python数据科学：全栈技术详解》

讲师：Ben

# 自我介绍

- 天善商业智能和大数据社区　　讲师 –Ben
- 天善社区 ID - Ben_Chang
- https://www.hellobi.com – 学习过程中有任何相关的问题都可以提到技术社区数据挖掘版块。

# 主要内容

- 概述和基本数学知识
- 凸函数
- 问题求解
  - 最速下降法
  - 牛顿法
  - Lagrange对偶问题

# 概述和基本数学知识

# 凸优化介绍

1、其应用非常广泛，机器学习中很多优化问题都要通过凸优化来求解；

2、在非凸优化中，凸优化同样起到重要的作用，很多非凸优化问题，可以转化为凸优化问题来解决；

3、如上引用所述，凸优化问题可以看作是具有成熟求解方法的问题，而其他优化问题则未必。

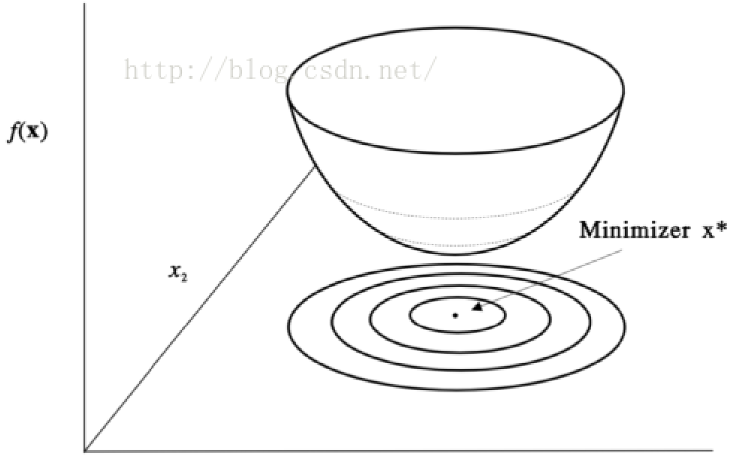
- 自变量为标量的函数$f:\mathbb{R}\rightarrow\mathbb{R}$

$$\min f(x) \quad x\in\mathbb{R}$$

- 自变量为矢量的函数$f:\mathbb{R}^n\rightarrow\mathbb{R}$

$$\min f(\mathbf{x}) \quad \mathbf{x}\in\mathbb{R}^n$$

- 熟悉Contour




在下方平面的部分叫做等高线。表示三维空间中所有的横切面（俯视图）构成的集合。

# 标准优化问题

$$
\begin{aligned}
\min \quad & f_0(x) & & \\
s.t. \quad & f_i(x) \le 0, & & i=1,\cdots,m \\
& h_i(x) = 0, & & i=1,\cdots,p
\end{aligned}
\qquad (1)
$$

表示在所有满足 $f_i(x) \le 0, i=1,\cdots,m$ 及 $h_i(x)=0, i=1,\cdots,p$ 的 $x$ 中找出使 $f_0(x)$ 最小的 $x$ 。

这里，$x \in \mathbf{R}^n$ 称为问题的优化变量，函数 $f_0 : \mathbf{R}^n \to \mathbf{R}$ 称为目标函数，

不等式 $f_i(x) \le 0$ 称为不等式约束，相应的 $f_i : \mathbf{R}^n \to \mathbf{R}, i=1,2,\cdots m$ 称为不等式约束函数，方程组 $h_i(x)=0$ 称为等式约束，相应的 $h_i : \mathbf{R}^n \to \mathbf{R}, i=1,2,\cdots p$ 称为等式约束。

如果没有约束，即 $m=p=0$，称问题(1)为无约束问题。

对目标和所有约束函数有定义的点的集合，

$$\mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{j=1}^{p} \operatorname{dom} h_j ,$$

称为优化问题(1)的定义域。

# 凸优化问题

$$
\begin{array}{lll}
\min & f_0(x) & \\
s.t. & f_i(x) \le 0, & i=1,\cdots,m \\
 & a_i^T x = b_i, & i=1,\cdots,p
\end{array}
\qquad (2)
$$

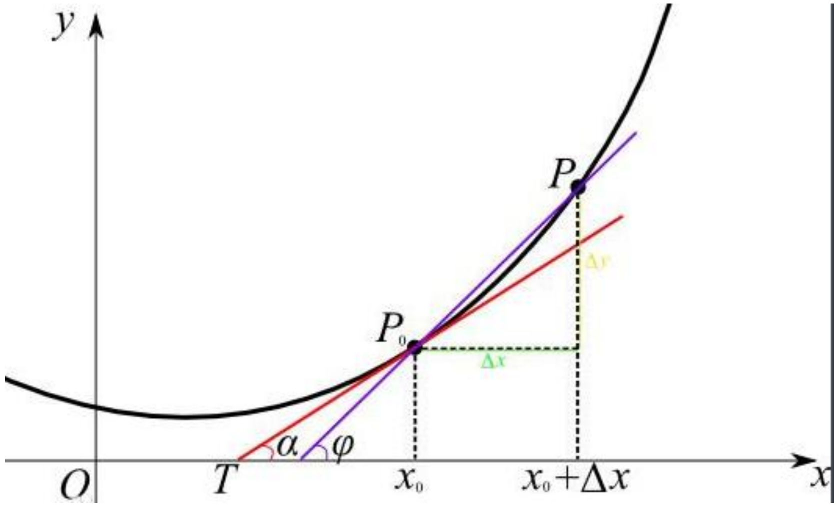
上述优化问题中，$f_i(x), i=0,\cdots,m$ 是凸函数。此类优化问题称为凸优化问题。

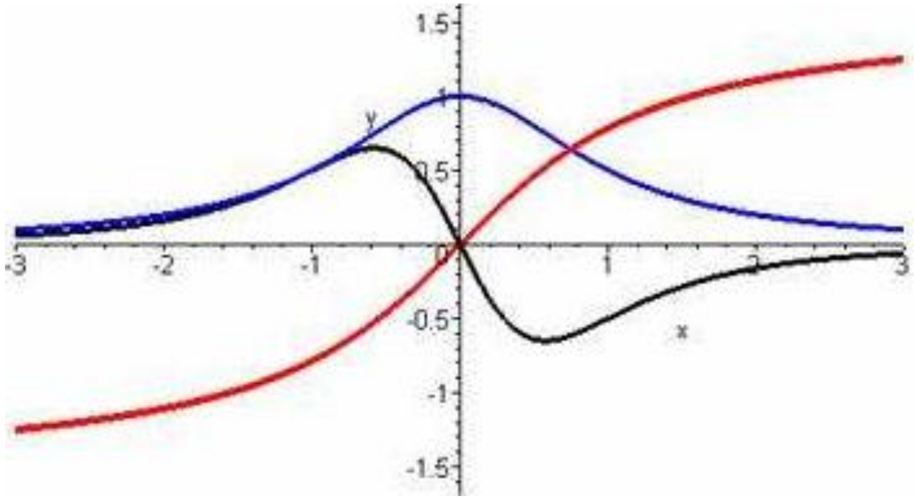
对比优化问题(1)，也就是说，目标函数和不等式约束为凸函数，等式约束是仿射函数的优化问题属于凸优化问题。

# 导数

简单的说，导数就是曲线的斜率，是曲线变化快慢的反应。

二阶导数是斜率变化快慢的反应，表征曲线的凸凹性。



# 常用函数的导数

（1） $C'=0$（$C$ 为常数）；（2） $(x^n)'=nx^{n-1}$（$n\in Q$）；

（3） $(\sin x)'=\cos x$；（4） $(\cos x)'=-\sin x$；

（5） $(a^x)'=a^x\ln a$；（6） $(e^x)'=e^x$；

（7） $(\log_a x)'=\frac{1}{x}\log_a e$；（8） $(\ln x)'=\frac{1}{x}$.

$$(u+v)'=u'+v'$$

$$(uv)'=u'v+uv'$$

# Taylor公式

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\Lambda+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+R_n(x)$$

$$f(x)=f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\Lambda+\frac{f^{(n)}(0)}{n!}x^n+o(x^n)$$

# Taylor公式的应用1

数值计算：初等函数值的计算(在原点展开)

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} + \Lambda + (-1)^{m-1} \frac{x^{2m-1}}{(2m-1)!} + R_{2m}$$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \Lambda + \frac{x^n}{n!} + R_n$$

在实践中，往往需要做一定程度的变换。

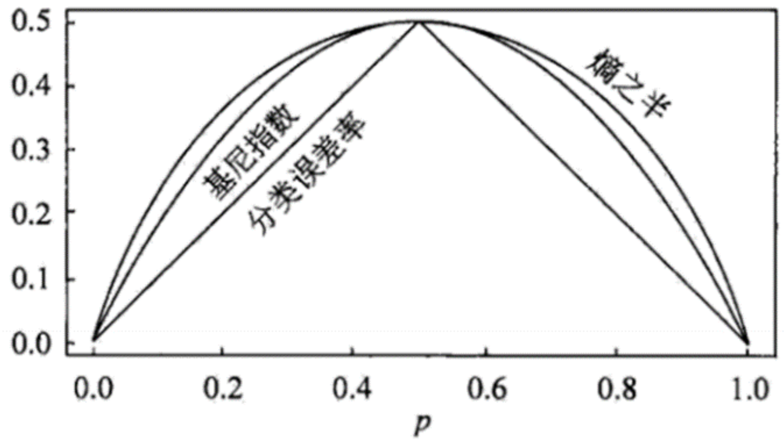
# Taylor公式的应用2

考察基尼系数、熵、分类误差率三者之间的关系

将f(x)=-lnx在x=1处一阶展开，忽略高阶无穷小，得到f(x)≈1-x

$$H(X) = -\sum_{k=1}^{K} p_k \ln p_k$$

$$\approx \sum_{k=1}^{K} p_k (1 - p_k)$$

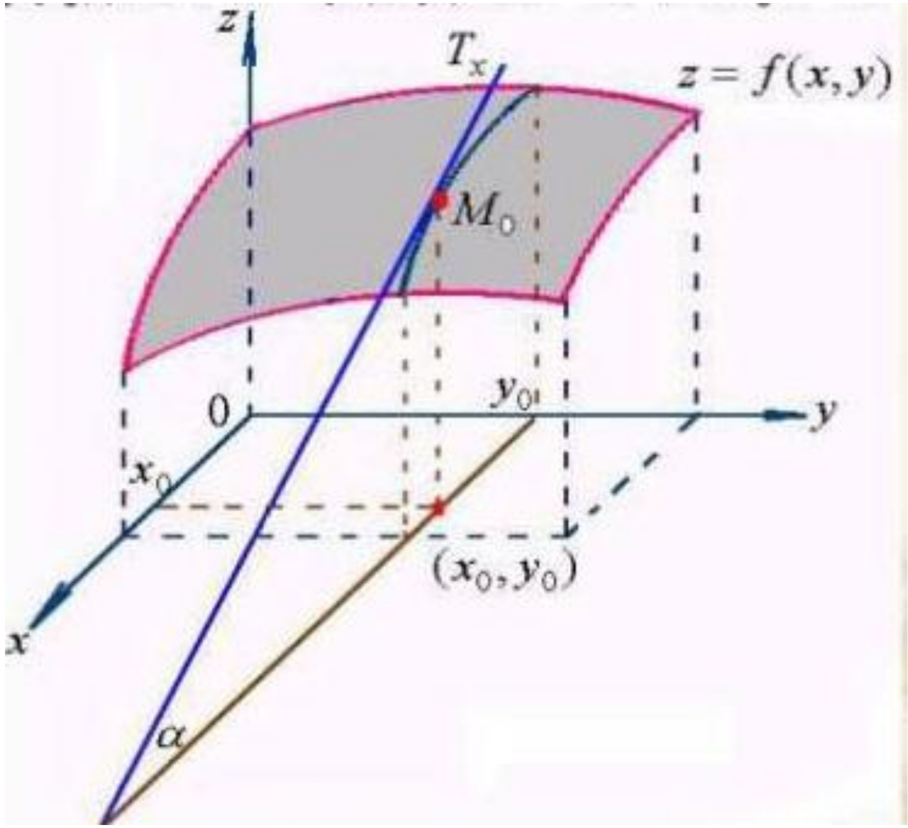



天善智能 TIANSHAN SOFT

# 偏导数

设函数z=f(x,y)在平面区域D内具有一阶连续偏导数。
符号记为：

$$\frac{\partial f}{\partial x}$$



# 方向导数

如果函数z=f(x,y)在点P(x,y)是可微分的，那么，函数在该点沿任一方向L的方向导数都存在，且有：

$$\frac{\partial f}{\partial l} = \frac{\partial f}{\partial x}\cos\varphi + \frac{\partial f}{\partial y}\sin\varphi$$

其中，ψ为x轴到方向L的转角。

# 梯度

设函数z=f(x,y)在平面区域D内具有一阶连续偏导数，则对于每一个点P(x,y)∈D，向量

$$\left(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}\right)$$

为函数z=f(x,y)在点P的梯度，记做**grad**f(x,y)

梯度的方向是函数在该点变化最快的方向。

考虑一座解析式为z=H(x,y)的山，在$(x_0,y_0)$的梯度是在该点坡度变化最快的方向。

梯度下降法

思考：若下山方向和梯度呈θ夹角，下降速度是多少？

其他概念请参考:

http://blog.csdn.net/thystar/article/details/51469469

# 凸集

# 仿射集(Affine set)

定义：通过集合C中任意两个不同点的直线仍然在集合C内，则称集合C为仿射集。

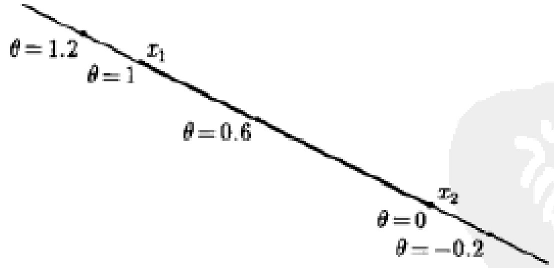
$$\forall x_1, x_2 \in C, \forall \theta \in R, \text{则} x = \theta \cdot x_1 + (1-\theta) \cdot x_2 \in C$$

仿射集的例子：直线、平面、超平面

# 直线上的点

$\forall x, y \in \mathbf{R}^n$ 且 $x \neq y$，那么

$$z = \theta x + (1-\theta) y, \theta \in$$




是一条穿越 $x$ 和 $y$ 的直线。

图 1 给出了二维情况下的直观解释。

A、$\theta=0$ 时，$z=y$；

B、$\theta=1$时，$z=x$；

C、$0 \le \theta \le 1$ 时，$z$ 是 $x$ 和 $y$ 之间的线段上的点;

D、$\theta<0$ 或 $\theta>1$ 时，线段上的点 $z$ 是 $x$ 和 $y$ 线段之外的点。

# 凸集

集合C内任意两点间的线段均在集合C内，则称集合C为凸集。

k个点的版本：$\forall x_1, x_2 \in C, \forall \theta \in [0,1]$, 则 $x = \theta \cdot x_1 + (1-\theta) \cdot x_2 \in C$

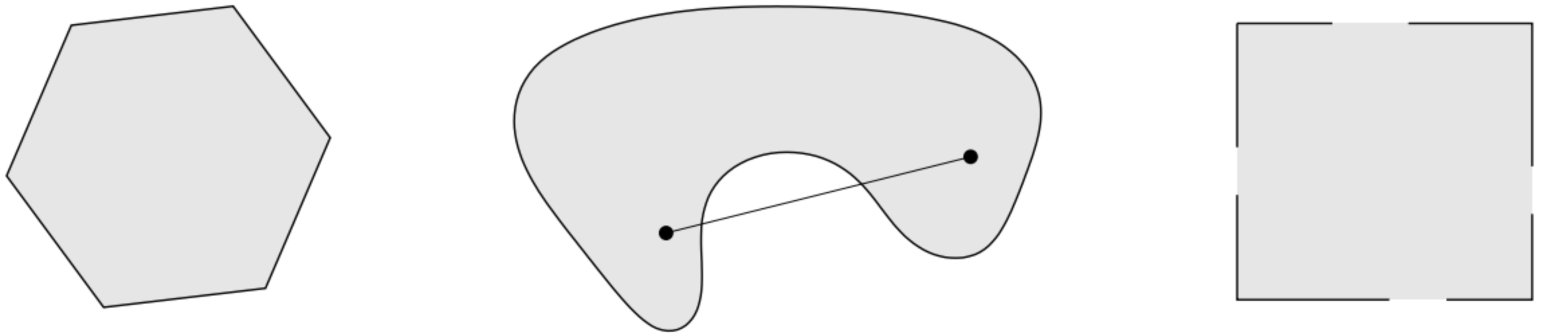
$$\forall x_1, x_2, \ldots x_k \in C, \theta_i \in [0,1] \text{且} \sum_{i=1}^{k} \theta_i = 1, \text{则} x = \sum_{i=1}^{k} \theta_i x_i \in C$$

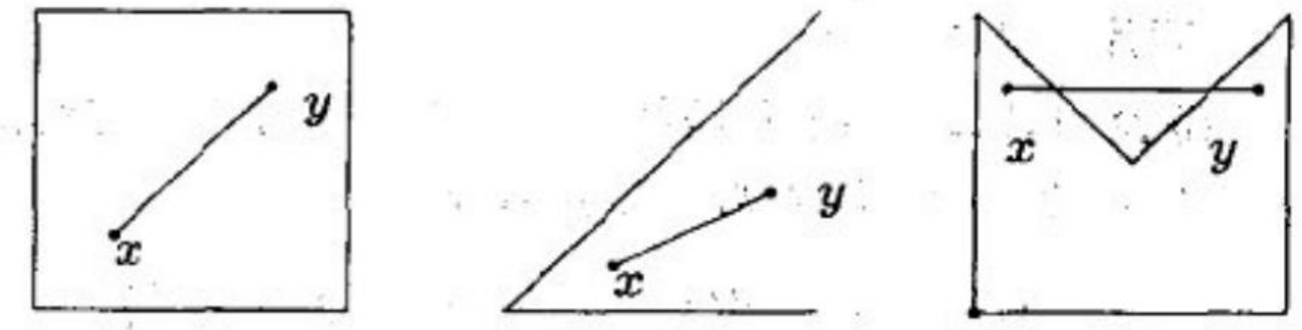
思考：两种表述的内涵一样吗？

# 仿射集和凸集的关系

因为仿射集的条件比凸集的条件强，所以，仿射集必然是凸集。

# 凸集






(a) 有界凸集　(b) 无界凸集　(c) 非凸集

# 典型的凸集

A、线段，射线，直线

B、超平面，半空间

C、仿射集

D、欧几里得球，范数球，椭球等

E、凸锥，范数锥等

# 其它相关知识

**保凸运算**

交集、仿射函数、线性分式函数及透视函数

**超平面分离定理**

两个不相交的凸集，存在一个超平面将其分离

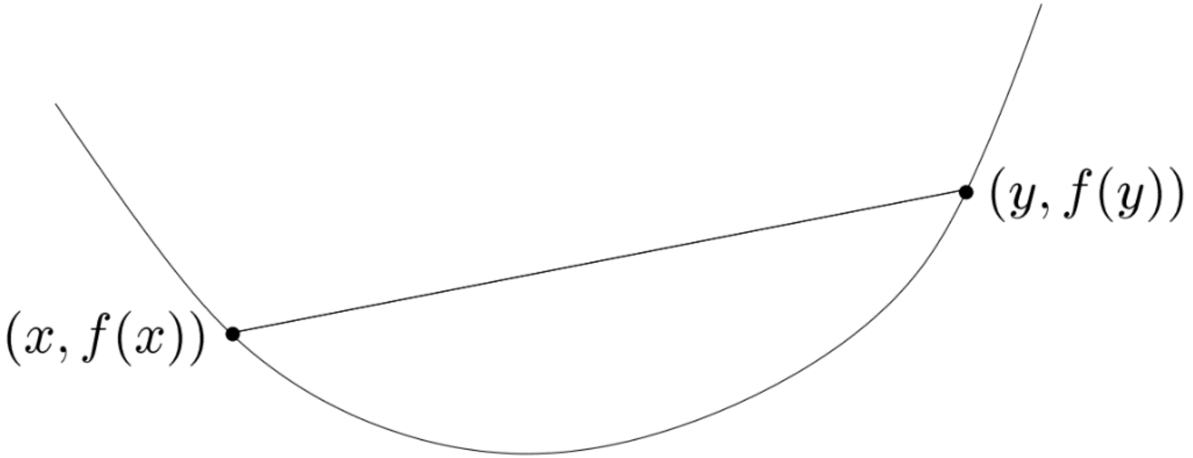
# 凸函数

# 凸函数

若函数f的定义域domf为凸集，且满足

$$\forall x, y \in \operatorname{dom} f, 0 \le \theta \le 1,\text{有}$$

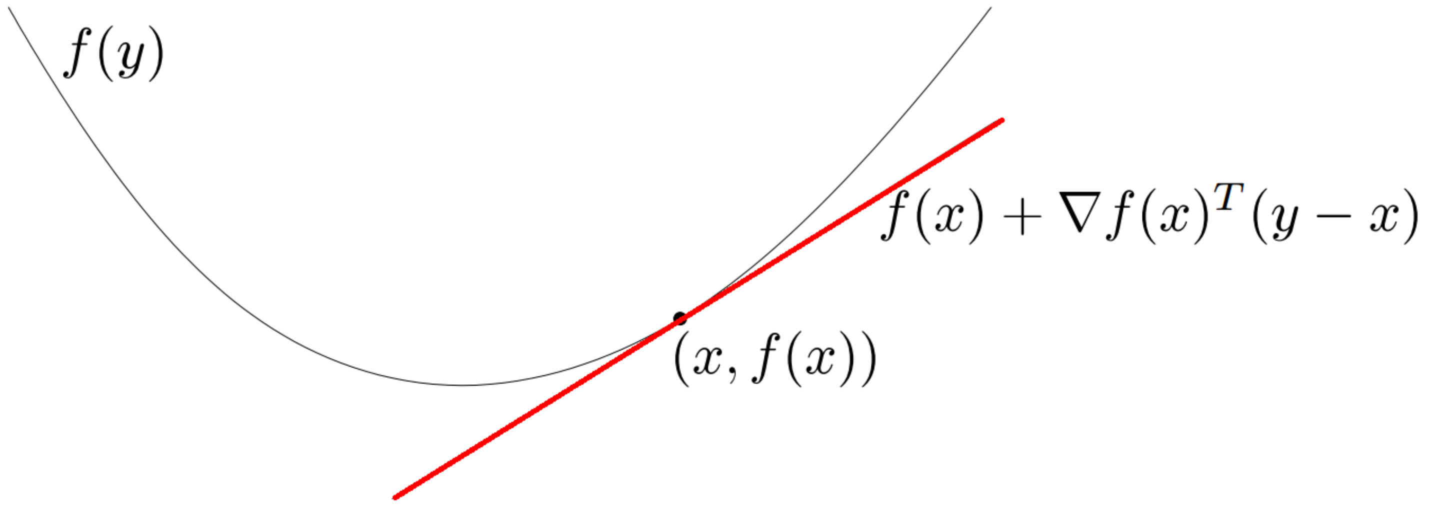
$$f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$$



# 一阶可微

若f一阶可微，则函数f为凸函数当前仅当f的定义域domf为凸集，且

$$\forall x, y \in \mathrm{dom} f,\ f(y) \geq f(x) + \nabla f(x)^T (y - x)$$



# 进一步思考

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

结合凸函数图像和支撑超平面理解该问题

对于凸函数，其一阶Taylor近似本质上是该函数的全局下估计。

反之，如果一个函数的一阶Taylor近似总是起全局下估计，则该函数是凸函数。

该不等式说明从一个函数的局部信息，可以得到一定程度的全局信息。

# 二阶可微

若函数f二阶可微，则函数f为凸函数当前仅当dom为凸集，且

$$\nabla^2 f(x) \succ= 0$$

若f是一元函数，上式表示二阶导大于等于0

若f是多元函数，上式表示二阶导Hessian矩阵半正定。

# 凸函数举例

- 指数函数 $e^{ax}$
- 幂函数 $x^a, x \in R_+, a \geq 1 \text{ or } a \leq 0$
- 负对数函数 $-\log x$
- 负熵函数 $x \log x$
- 范数函数 $\|x\|_p$

$$f(x) = \max(x_1, \ldots, x_n)$$

$$f(x) = x^2 / y, y > 0$$

$$f(x) = \log(e^{x_1} + \ldots + e^{x_n})$$

# 下水平集

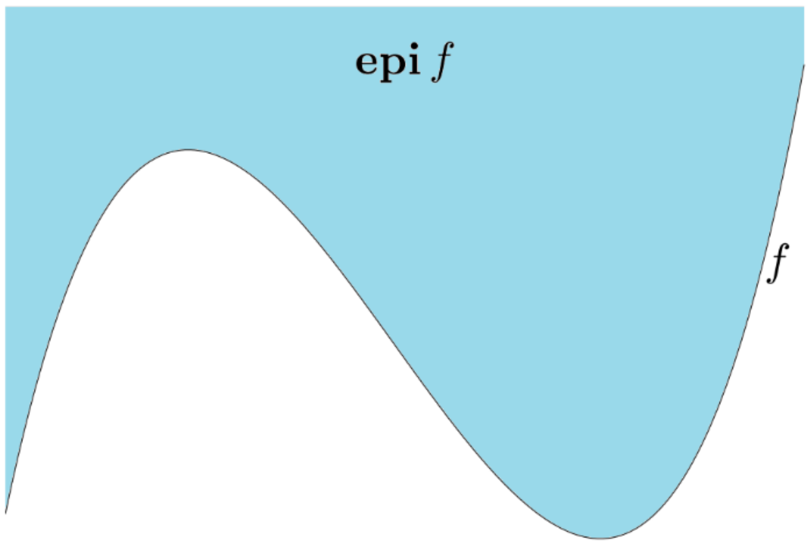
函数 $f$ 的下水平集 $C_\alpha = \{x \in \mathrm{dom}\, f \mid f(x) \le \alpha\}$ 是其定义域的子集。凸函数的下水平集是凸集。

# 上境图

函数f的图像定义为： $\{(x, f(x)) \mid x \in \mathbf{dom}\, f\}$

函数f的上境图(epigraph)定义为：

$$\mathbf{epi}\, f = \{(x, t) \mid x \in \mathbf{dom}\, f,\ f(x) \le t\}$$



# 凸函数与凸集

一个函数是凸函数，当且仅当其上境图是凸集。

思考：如何证明？(提示：定义)

进一步，一个函数是凹函数，当且仅当其亚图(hypograph)是凸集。

$$\mathbf{hypo}\, f = \{(x,t) \mid t \le f(x)\}$$

# 保持函数凸性的算子

凸函数的非负加权和

$$f(x) = \omega_1 f_1(x) + \ldots + \omega_n f_n(x)$$

凸函数与仿射函数的复合

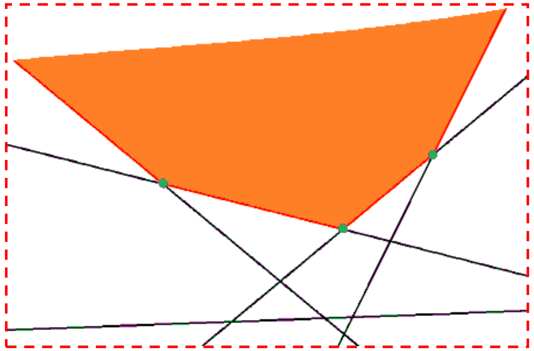
$$g(x) = f(Ax + b)$$

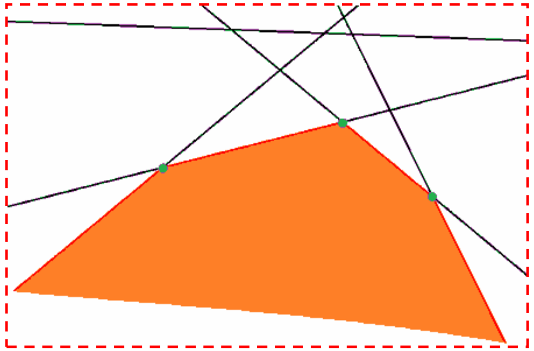
凸函数的逐点最大值、逐点上确界

$$f(x) = \max(f_1(x), \ldots, f_n(x))$$

$$f(x) = \sup_{y \in A} g(x, y)$$

# 逐点上确界和上境图的关系





逐点上确界和上境图的关系

一系列函数逐点上确界函数对应着这些函数上境图的交集。

直观例子

Oxy平面上随意画N条直线(直线是凸的——虽然直线也是凹的)，在每个x处取这些直线的最大的点，则构成的新函数是凸函数；

同时：N条直线逐点求下界，是凹函数；

在Lagrange对偶函数中会用到该结论。

# Jensen不等式：若f是凸函数

基本Jensen不等式

若 $$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$$

则 $$\theta_1, \ldots, \theta_k \ge 0,\ \theta_1 + \cdots + \theta_k = 1$$

若 $$f(\theta_1 x_1 + \cdots + \theta_k x_k) \le \theta_1 f(x_1) + \cdots + \theta_k f(x_k)$$

则 $$p(x) \ge 0 \text{ on } S \subseteq \mathbf{dom}\, f,\ \int_S p(x)\, dx = 1$$

$$f\left(\int_S p(x) x\, dx\right) \le \int_S f(x) p(x)\, dx$$

$$\boxed{f(\mathbf{E}\, x) \le \mathbf{E}\, f(x)}$$

# 问题求解

# 最优性准则

对于凸优化问题

$$\begin{aligned} \min \quad & f_0(x) \\ s.t. \quad & f_i(x) \le 0, \quad i=1,\cdots,m \\ & a_i^T x = b_i, \quad i=1,\cdots,p \end{aligned} \qquad (1)$$

其中，$f_i(x), i=0,\cdots,m$ 是凸函数。

$x^* \in X$ 是可微函数 $f_0$ 的最优解，当且仅当

$$\nabla f_0(x^*)^T(y-x^*) \ge 0, \forall y \in X \qquad (2)$$

# 无约束凸优化的最优性准则

无约束凸优化问题(式(1)中，$m = p = 0$)，$x^* \in X$ 是可微函数 $f$ 的最优解的重要条件是众所周知的

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

# 无约束凸优化问题求解

**解析解**

对于少数一些简单的凸优化问题，可以利用最优性准则通过解析来求解。但对于大多数凸优化问题来讲，是没有办法通过解析来求解的, 主要原因如下:

a. 函数不可导；
b.可导，但是变量太多无法求解；
c. 解可能是一个集合。

练习求解线性回归目标函数的解析解

$$\ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2$$

# 无约束凸优化问题求解

## 下降方法——迭代法

下降方法将产生一个优化点列 $x^{(k)}, k=1,\cdots$，其中

$$x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)},\quad t^{(k)} > 0,$$

使得

$$f(x^{(k+1)}) < f(x^{(k)}),$$

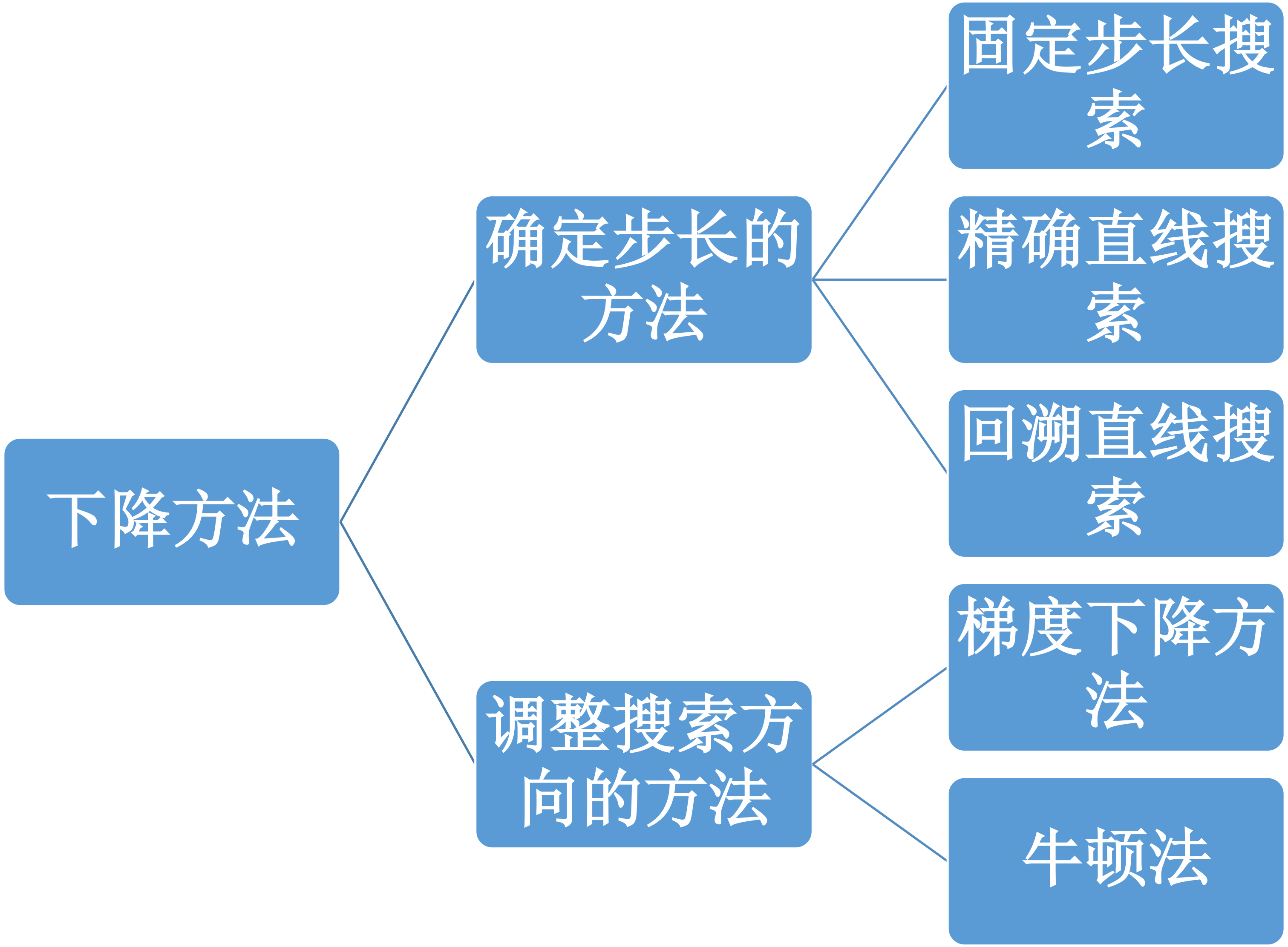
$\Delta x^{(k)}$ 被称为**搜索方向**，$t^{(k)}$ 被称为**步长**。

由凸函数的一阶条件可知，下降方法中的搜索方向必须满足

$$\nabla f(x^{(k)})^T \Delta x^{(k)} < 0,$$

即搜索方向和负梯度方向的夹角必须是锐角。

# 无约束凸优化问题求解



# 梯度下降方法

下降方法中必须满足，搜索方向和负梯度成锐角。用负梯度作为搜索方向就自然而然了，即

$$\Delta x = -\nabla f(x)$$

# 牛顿法

用目标函数的二阶泰勒展开近似该目标函数，通过求解这个二次函数的极小值来求解凸优化的搜索方向。

目标函数 $f(x)$ 在 $x$ 处的二阶 Taylor 展开为

$$f(x+v) \approx f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v,$$

因为在最小值附近, 因此 $\frac{\partial}{\partial v} f(x+v) = 0$

使上式两边对 $v$ 求偏导

$$\frac{\partial}{\partial v} f(x+v) = \nabla f(x) + \nabla^2 f(x) v,$$

当 $f(x+v)$ 最小时，$f(x+v)$ 对 $v$ 的偏导为零，即

$$\nabla f(x) + \nabla^2 f(x) v = 0$$

则有

$$v = -\nabla^2 f(x)^{-1} \nabla f(x)$$

# 牛顿法

将这个方向作为搜索方向，这个方向被称作 Newton 步径，即

$$\Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x) ,$$

$\nabla^2 f(x)$ 即为函数 $f(x)$ 的 Hessian 矩阵，由凸函数的 Hessian 矩阵的正定性可知，除 $f(x)$ 到达极值点 $\nabla f(x) = 0$ 外，有

$$\nabla f(x)^T \Delta x_{nt} = -\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) < 0$$

因此 Newton 步径是下降方向。

这种方法就是牛顿法(Newton 法)。

# 牛顿法

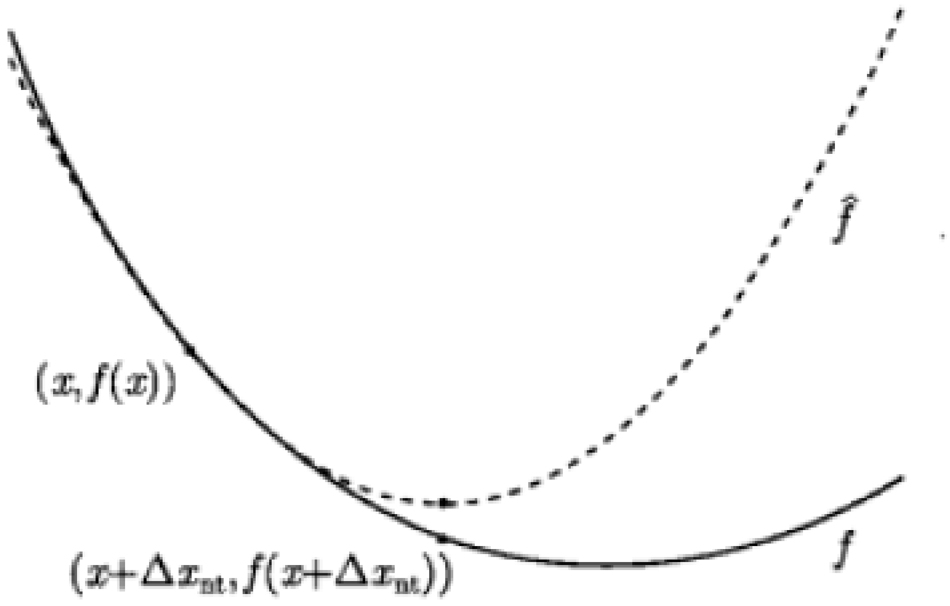
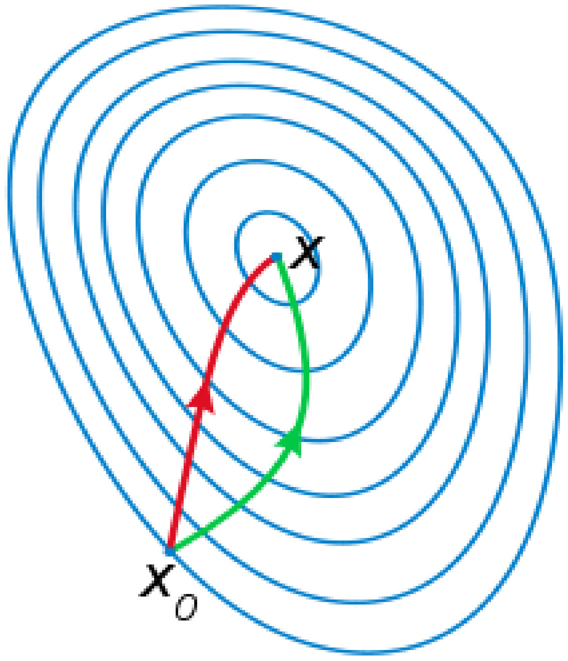



图 1 大概说明了牛顿法的原理，实线是目标函数 $f(x)$，虚线是 $f(x)$ 在 $x$ 处的 Taylor 二阶展开。

# 牛顿法

牛顿法就是用一个二次曲面去拟合当前所处位置的局部曲面，而梯度下降法是用一个平面去拟合当前的局部曲面，通常情况下，二次曲面的拟合会比平面更好，所以牛顿法选择的下降路径会更符合真实的最优下降路径。




红色为牛顿方法搜索路径，绿色为梯度下降法搜索路径。

牛顿法在GBDT中将有直接运用。

牛顿法的介绍:
http://blog.csdn.net/itplus/article/details/21896453

# 有约束凸优化的最优化准则

等式约束凸优化问题(式(1)中，$m=0$)，$x^* \in X$ 是可微函数 $f_0$ 的最优解的重要条件是，存在 $\mathbf{v}=[v_1,\cdots,v_p]^T \in \mathbf{R}^p$，使得

$$\nabla f_0(x^*)+\sum_{i=1}^{p} v_i a_i = 0$$

其中，这里 $a_i \in \mathbf{R}^n$。这个可以通过 Lagrange 对偶函数的 KKT 条件证明得到。

将等式约束凸优化问题描述为矩阵形式

$$\begin{array}{ll} \min & f(x) \\ s.t. & \mathbf{A}x=b \end{array}$$

对应的，$x^* \in X$ 是可微函数 $f$ 的最优解的重要条件为，存在 $\mathbf{v} \in \mathbf{R}^p$，使得

$$\nabla f(x^*)+\mathbf{A}^T\mathbf{v}=0$$

其中，$\mathbf{A} \in \mathbf{R}^{p\times n}$。

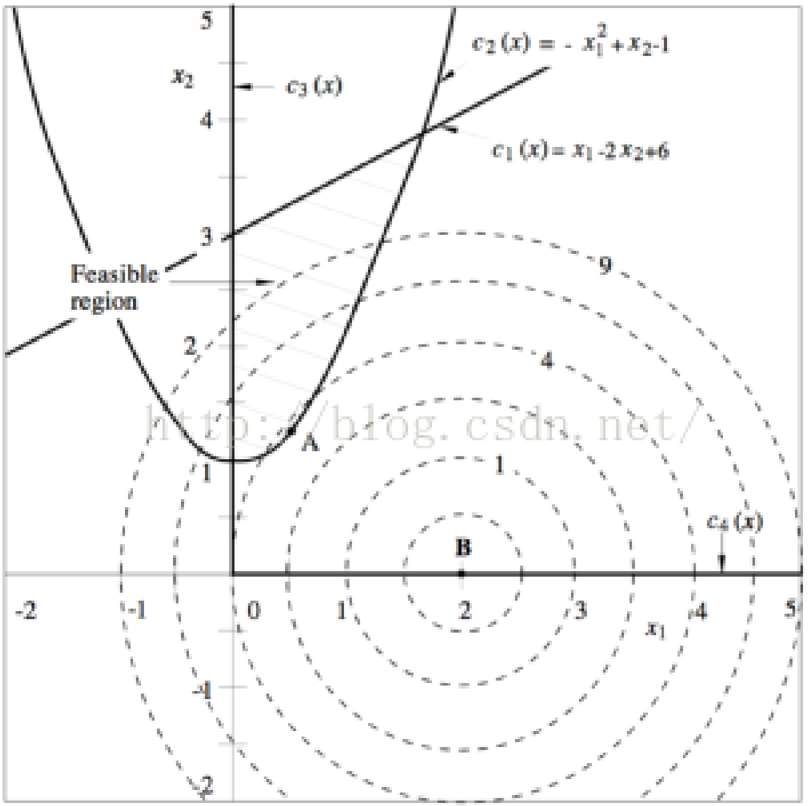
天善智能 TIANSHAN SOFT

# 示例

$$\text{minimize} \quad f(\mathbf{x}) = x_1^2 + x_2^2 - 4x_1 + 4 = (x_1-2)^2 + x_2^2$$

$$\text{subject to} \quad c_1(\mathbf{x}) = x_1 - 2x_2 + 6 \geq 0$$

$$c_2(\mathbf{x}) = -x_1^2 + x_2 - 1 \geq 0, c_3(\mathbf{x}) = x_1 \geq 0, c_4(\mathbf{x}) = x_2 \geq 0$$




可行域上的最优解一定会碰到一个边界。即至少有一个c (x) = 0;

在约束条件下，可行域围成的区间为抛物线以上，直线以下的那部分区域。最优解必须在可行域中。
最优解为A点。最优解会碰到c2边界。

# 有约束凸优化的优化问题求解

任何等式约束优化问题都可以通过消除等式约束转化为等价的无约束优化问题，然后利用无约束的方法求解。

$$\begin{aligned} \min \quad & f_0(x) & \\ s.t. \quad & f_i(x) \le 0, & i=1,\cdots,m \\ & h_i(x) = 0, & i=1,\cdots,p \end{aligned}$$

这里，约束条件有等式，也有不等式。即要满足$f_0(x)$，也要满足约束条件，即可行域：满足$f_0(x)$定义域和约束条件的x的集合。

可行域上的最优解一定会碰到一个边界。即至少有一个$f_i(x) = 0$
如果$x^*$是限制优化问题的局部最小解，那么存在拉格朗日乘子$\lambda_i^*$和$v_j^*$使得

$$\nabla f_0(x^*) = \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{j=1}^{p} v_j^* \nabla h_i(x^*)$$

# Lagrange函数

一般优化问题的拉格朗日乘子法

$$
\begin{aligned}
\text{minimize} \quad & f_0(x), \quad x \in \mathbf{R}^n \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\
& h_j(x) = 0, \quad j = 1, \ldots, p
\end{aligned}
$$

拉格朗日函数

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

则 $\nabla_x \mathrm{L}(x^*, \lambda^*, \nu^*) = 0$ 等价于

$$\nabla f_0(x^*) = \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{j=1}^{p} v_j^* \nabla h_i(x^*)$$

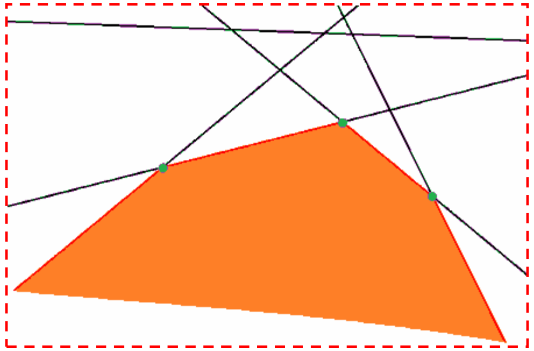
天善智能 TIANSHAN SOFT

# 拉格朗日（Lagrange）对偶函数

定义 Lagrange 对偶函数为 Lagrange 函数关于 $x$ 取最小值，即对于 $\lambda \in \mathbf{R}^m$，$\nu \in \mathbf{R}^p$，有

$$g(\lambda,\nu)=\inf_{x\in\mathcal{D}} L(x,\lambda,\nu)=\inf_{x\in\mathcal{D}}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\nu_i h_i(x)\right)。$$

Lagrange 对偶函数是一族关于 $(\lambda,\nu)$ 的仿射函数的逐点下确界，即使原问题不是凸的，对偶函数也是凹函数。



天善智能 TIANSHAN SOFT

# 拉格朗日对偶函数(dual function)

拉格朗日对偶函数

$$g(\lambda,\nu)=\inf_{x\in D} L(x,\lambda,\nu)=\inf_{x\in D}\left(f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)+\sum_{i=1}^{p}\nu_i h_i(x)\right)$$

若没有下确界，定义：

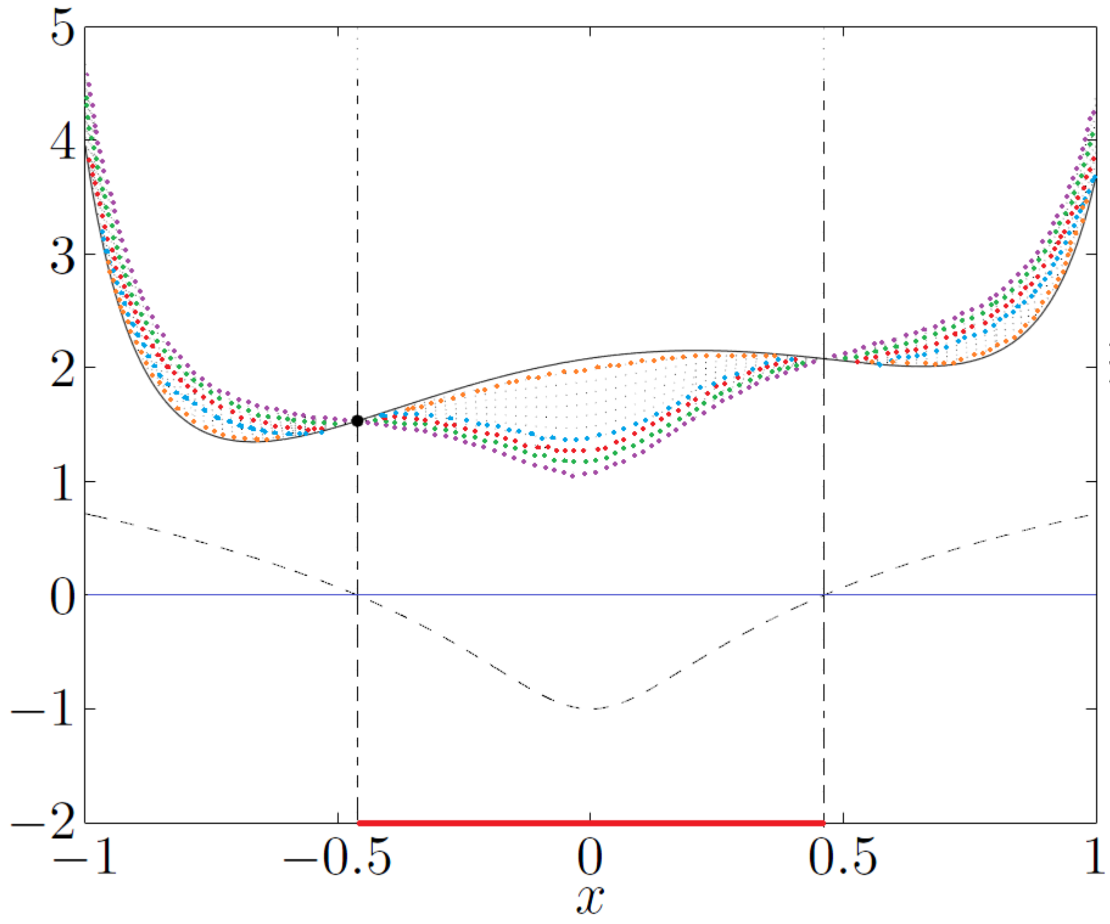
$$g(\lambda,\nu)=-\infty$$

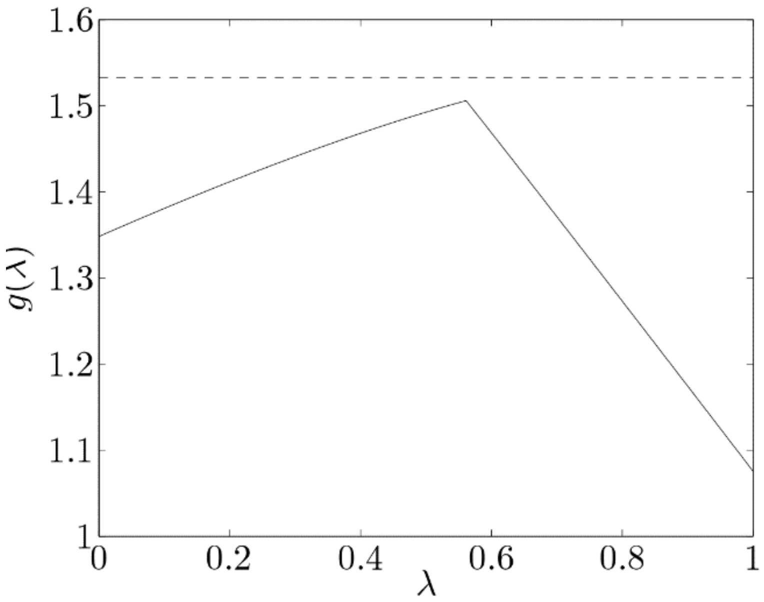
根据定义，显然有：对∀λ>0，∀v，若原优化问题有最优值p*，则

$$g(\lambda,\nu)\le p^*$$

进一步：拉格朗日对偶函数为凹函数。

# 左侧为原函数，右侧为对偶函数




天善智能
TIANSHAN SOFT

# 鞍点解释

为表述方便，假设没有等式约束，只考虑不等式约束，结论可方便地扩展到等式约束。

假设x0不可行，即存在某些i，使得$f_i(x)>0$。则选择　　　　，对于其他乘子，

$$\lambda_i \to \infty \qquad \lambda_j = 0,\ j \neq i$$

假设x0可行，则有$f_i(x)\le 0,(i=1,2,\ldots,m)$,选择

有：

$$\lambda_i = 0, i = 1,2,\Lambda\ ,m$$

$$\sup_{\lambda\ge 0} L(x,\lambda) = \sup_{\lambda\ge 0}\left( f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right) = \begin{cases} f_0(x) & f_i(x) \le 0, i = 1,2\Lambda\ ,m \\ \infty & otherwise \end{cases}$$

# 鞍点：最优点

而原问题是：
$$\inf_x f_0(x)$$

从而，原问题的本质为：
$$\inf_x \sup_{\lambda\geq 0} L(x,\lambda)$$

而对偶问题，是求对偶函数的最大值，即：

$$\sup_{\lambda\geq 0} \inf_x L(x,\lambda)$$

而：

$$\sup_{\lambda\geq 0} \inf_x L(x,\lambda) \leq \inf_x \sup_{\lambda\geq 0} L(x,\lambda)$$

# 证明： $\max_x \min_y f(x,y) \le \min_y \max_x f(x,y)$

对于任意的(x,y)∈domf

$$f(x,y) \le \max_x f(x,y)$$

$$\Rightarrow \min_y f(x,y) \le \min_y \max_x f(x,y)$$

$$\Rightarrow \max_x \min_y f(x,y) \le \min_y \max_x f(x,y)$$

# 强对偶条件

若要对偶函数的最大值即为原问题的最小值，考察需要满足的条件：

$$
\begin{aligned}
f_0(x^\star) &= g(\lambda^\star, \nu^\star) \\
&= \inf_x \left( f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{i=1}^{p} \nu_i^\star h_i(x) \right) \\
&\leq f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star h_i(x^\star) \\
&\leq f_0(x^\star).
\end{aligned}
$$

# Karush-Kuhn-Tucker (KKT)条件

$$
\begin{aligned}
f_i(x^\star) &\leq 0, \quad i = 1, \ldots, m \\
h_i(x^\star) &= 0, \quad i = 1, \ldots, p \\
\lambda_i^\star &\geq 0, \quad i = 1, \ldots, m \\
\lambda_i^\star f_i(x^\star) &= 0, \quad i = 1, \ldots, m
\end{aligned}
$$

$$
\nabla f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^p \nu_i^\star \nabla h_i(x^\star) = 0
$$

# 参考书


《凸优化》，Stephen Boyd等著，王书宁等译