Xinyi Li

http://notebook.xyli.me/Introducti
project-yangzhou/

主页
所有文章

这么早就说Hessian矩阵是半正定的，会不会给人一种凹函数的感觉？

2016-03-14

凸优化 微积分 线性代数

我大湾里技校网管系扛把子的顾凌峰同学说他已经决定了，让我来整理这篇回答，和大家传授一些关于Hessian矩阵和函数凸性关系的人生经验。原回答地址：怎么理解二阶偏导与凸函数的Hessian矩阵是半正定的？ - 顾凌峰的回答

本文将着力于分析多元函数Hessian矩阵的性质和函数凸性之间的充要关系，但由于我知识水平低下，也实在不是谦虚，严谨的数学证明还是另请高明，只试图给出一些直观的理解。

接下来我们将介绍凸函数的定义、一阶条件、二阶条件，以及这三者之间的**充要**关系。*表示内容为扩展的支线剧情，对主线没有影响，各位读者量力而行，觉得实在难以理解或者浪费时间可以直接跳过。

为了方便起见，我们忽略一些无关紧要的部分，假设函数具有良好的性质

- 定义域$\mathrm{dom}\, f$为凸集
- 在定义域上连续
- 在定义域上二次可导

这些条件当然不是必要的，我们将在后记部分进行详细的说明。

凸函数（*Convex Function*）

> 我一个看上去像“凹”字的函数，怎么就把我定义成凸函数了呢？

定义（*Definition*）

对于一元函数$y=f(\cdot)$来说，函数为凸的定义是:

对于定义域内两个任意的自变量x_1和x_2，且有任意$0\le\theta\le 1$，如果函数$f(\cdot)$满足

$$f(\theta x_1+(1-\theta)x_2)\le\theta f(x_1)+(1-\theta)f(x_2)$$

那么函数$f(\cdot)$是凸函数。

我们可以理解为连接函数图像两点构成的线段始终在函数曲线上方。

把这个定义推广到多元函数：多变量函数可以把自变量写成一个向量$\mathbf{x}=[x_1,x_2,\dots,x_n]^T$，同理对于定义域的任意两个自变量$\mathbf{x}_1$和$\mathbf{x}_2$，以及任意$0\le\theta\le 1$，如果函数$f(\cdot)$满足

$$f(\theta\mathbf{x_1}+(1-\theta)\mathbf{x_2})\le\theta f(\mathbf{x_1})+(1-\theta)f(\mathbf{x_2})$$

那么函数$f(\cdot)$是凸函数。

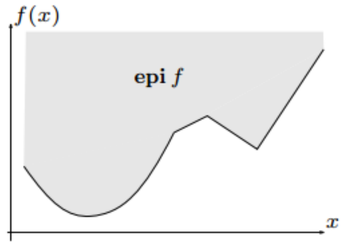
性质*（*Properties*）

一说到凸函数我们第一个想到的例子就是二次函数$y=x^2$，有人可能会疑惑这不是凹函数吗？是的，国内有很长一段时间对于函数的凹/凸的判定是混乱的，我初高中的时候受到的教育也是把$y=x^2$当作凹函数进行讲解的。$y=x^2$的图像直观看上去更像我们日常语境中说的“凹”，或者也有翻译者为了讨巧把convex说成“下凸”。你们这么望文生义是不行的，这么简单粗暴想当然就把它说成凹函数，就像在长安街看见坦克就大喊“占”或者“点”一样。

Xinyi Li

当然有数学工作者说后来国内出了个基本法明确的规定了凹凸的定义从此这些混乱才结束。

所以有人会不禁困惑，凸函数的定义和定义代表的几何意义到底和凸有什么关系？为什么要把满足这种性质的函数称为凸函数？

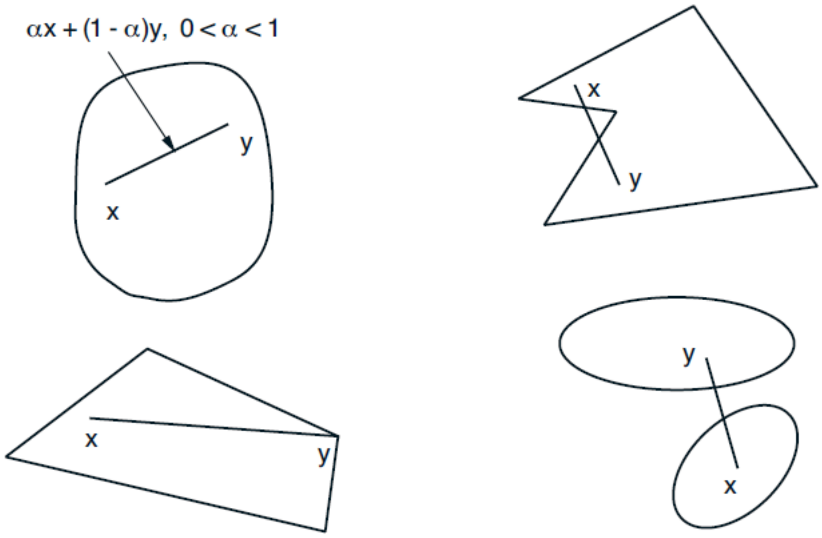


首先引入**上境图**（*Epigraph/Supergraph*）的概念：在函数$y=f(\cdot)$**曲线及其上方**的所有的点构成的集合。对于n元函数$y=f(\cdot)$这个$R^n\to R$的映射，其上境图为

$$epi=\{(\mathbf{x},y):\mathbf{x}\in R^n,y\in R,y\geq f(x)\}$$

当这个点集是凸集（*Convex Set*）的时候，对应的函数是凸函数，反之亦然。

凸集的“凸”是可以从几何上直观理解的，连接点集中的任意两点，线段上的每一点都在该点集内，则为凸集



Convex Sets　　　　Nonconvex Sets

这时我们对凸函数的定义有了更深刻的了解，它的定义不过是要求*上境图中的两点连线仍然包含于上境图*的一个特例，特别之处在于这两点都在集合的边界。考虑到连续函数的上境图必然是闭集，而验证闭集的凸性时，一般来说，这样的性质只要对于边界上的点成立，那么对于其他点就都成立。

你们给我弄的这个凸性啊，excited!这样去理解函数的“凸”就不会再出现凹凸定义混乱的情况了，不会听得风就是雨说$y=x^2$是凹函数了。

一阶条件（*First-order Condition*）

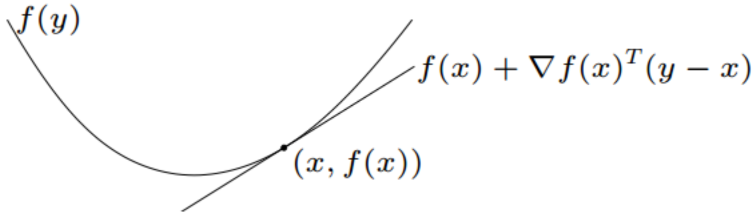
当且仅当对于定义域上任意自变量$\mathbf{x}_1$和$\mathbf{x}_2$，$y=f(\cdot)$都满足

$$f(\mathbf{x_1})\geq f(\mathbf{x}_2)+\nabla f(\mathbf{x}_2)\cdot(\mathbf{x}_1-\mathbf{x}_1)$$

时$y=f(\cdot)$为凸函数。

其中梯度向量$\nabla f(\mathbf{x})=\left(\frac{\partial f(\mathbf{x})}{\partial x_1},\frac{\partial f(\mathbf{x})}{\partial x_2},\ldots,\frac{\partial f(\mathbf{x})}{\partial x_n}\right)$也就是在该点对各个变量求偏导构成的向量。

应用到简单的一元函数可以看到



直观的理解是：凸函数的函数图像必然位于曲线上任意一点的切线的上方

必要性（*Necessity*）

本小节将证明凸函数必然具有上述性质。即对于在定义域内可导的凸函数$y=f(\cdot)$必然有不等式$f(\mathbf{x}_1)\geq f(\mathbf{x}_2)+\nabla f(\mathbf{x}_2)\cdot(\mathbf{x}_1-\mathbf{x}_1)$对于任意的$\mathbf{x}_1$和$\mathbf{x}_2$均成立。

假设函数$f(\cdot)$在定义域上是凸函数，那么根据定义有（假设$0<\theta\leq 1$）：

$$
\begin{aligned}
f(\theta\mathbf{x}_1+(1-\theta)\mathbf{x}_2) &\leq \theta f(\mathbf{x}_1)+(1-\theta)f(\mathbf{x}_2)\\
f(\mathbf{x}_2+\theta(\mathbf{x}_1-\mathbf{x}_2)) &\leq f(\mathbf{x}_2)+\theta(f(\mathbf{x}_1)-f(\mathbf{x}_2))\\
f(\mathbf{x}_2+\theta(\mathbf{x}_1-\mathbf{x}_2))-f(\mathbf{x}_2) &\leq \theta(f(\mathbf{x}_1)-f(\mathbf{x}_2))\\
\frac{f(\mathbf{x}_2+\theta(\mathbf{x}_1-\mathbf{x}_2))-f(\mathbf{x}_2)}{\theta} &\leq f(\mathbf{x}_1)-f(\mathbf{x}_2)\\
f(\mathbf{x}_1) &\geq f(\mathbf{x}_2)+\frac{f(\mathbf{x}_2+\theta(\mathbf{x}_1-\mathbf{x}_2))-f(\mathbf{x}_2)}{\theta}
\end{aligned}
$$

令$g(\theta)=f(\mathbf{x}_2+\theta(\mathbf{x}_1-\mathbf{x}_2))$，则$g(0)=f(\mathbf{x}_2)$，原不等式可以代换为$f(\mathbf{x}_1)\geq f(\mathbf{x}_2)+\frac{g(\theta)-g(0)}{\theta}$ 当θ趋近于0时，有$\lim\limits_{\theta\to 0}\frac{g(\theta)-g(0)}{\theta}=g'(0)$这一项也就是函数$g(\theta)$在$\theta=0$处的导数值，$g(\theta)$ 实际是$f(\mathbf{x})$与$\mathbf{x}=\mathbf{x}_2+\theta(\mathbf{x}_1-\mathbf{x}_2)$的复合函数，借助链式法则容易求导得

$$
\frac{dg}{d\theta}=(\mathbf{x}_1-\mathbf{x}_2)\frac{df(\mathbf{x}_2+\theta(\mathbf{x}_1-\mathbf{x}_2))}{d\mathbf{x}}
$$

由于只要求在$\theta=0$处的导数值所以容易得

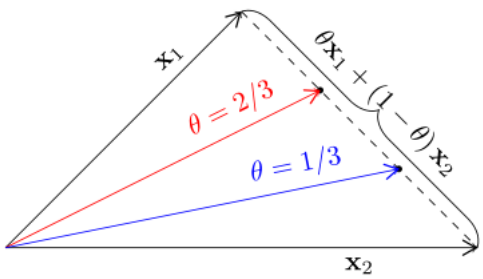
$$
\left.\frac{dg}{d\theta}\right|_{\theta=0}=(\mathbf{x}_1-\mathbf{x}_2)\frac{df(\mathbf{x}_2)}{d\mathbf{x}}=\nabla f(\mathbf{x}_2)\cdot(\mathbf{x}_1-\mathbf{x}_2)
$$

代入回不等式即可得到

$$
f(\mathbf{x}_1)\geq f(\mathbf{x}_2)+\nabla f(\mathbf{x}_2)\cdot(\mathbf{x}_1-\mathbf{x}_2)
$$

证毕。

这个推导过程中关键性的一步就是取了极限$\theta\to 0$，怎样理解这个操作？
$\theta\mathbf{x}_1+(1-\theta)\mathbf{x}_2$表示一个由$\mathbf{x}_1$和$\mathbf{x}_2$合成的向量



当我们规定$0\leq\theta\leq 1$时，得到的新向量必然落在向量$\mathbf{x}_1$和$\mathbf{x}_2$终点之间连接成的线段上，也就是图中的虚线位置，而θ的大小决定了新向量的终点更接近$\mathbf{x}_1$的终点还是$\mathbf{x}_2$的终点，可以把它理解为一种“**权重**”。当$\theta\to 0$时，我们取的这个新的点无限接近于$\mathbf{x}_2$所表示的点。

从凸函数的性质出发，函数曲线上任意两点连接而成的线段都在函数图像的上方，而考虑到定义域内可导的性质，所在直线的其他部分都在函数图像的下方（否则直接无法与函数相

交，而凸性对于整个定义域都成立则不可能存在两个以上的交点）。当所取的图像上的两点无限接近时，也就是$\theta \rightarrow 0$时，直线在函数上方的部分也趋近于0，则直线始终位于函数图像的下方，而这条直线可以用朴素（*naive*）的 *“切线逼近割线”* 去理解为函数在$\mathbf{x} = \mathbf{x}_2$处的切线，由于取点是任意的，一阶条件不等式表达的必要性 *“凸函数上任意一点的切线永远不高于函数图像”* 也是与图相合的，函数曲线比任意一点的切线不知道高到哪里去了。

充分性（*Sufficiency*）

本小节将证明一阶条件的充分性，即通过上述不等式的成立推导出具有这种性质的函数是凸函数。

对于定义域上任意自变量$\mathbf{x}_1$和$\mathbf{x}_2$，以及$0 \leq \theta \leq 1$，构造一个新变量$\mathbf{x} = \theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2$，由于定义域是凸集，必然有$\mathbf{x} \in \mathrm{dom} f$，即合成的新变量$\mathbf{x}$也被硬点在定义域内。所以该不等式对于$\mathbf{x}$，$\mathbf{x}_1$和$\mathbf{x}_2$这三个代表都适用，那么有：

$$\begin{cases} f(\mathbf{x}_1) & \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{x} - \mathbf{x}_1) \\ f(\mathbf{x}_2) & \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{x} - \mathbf{x}_2) \end{cases}$$

由于$0 \leq \theta \leq 1$，则有$0 \leq 1-\theta \leq 1$，分别以θ和$1-\theta$乘以上面两个不等式的两边，不等式方向不变：

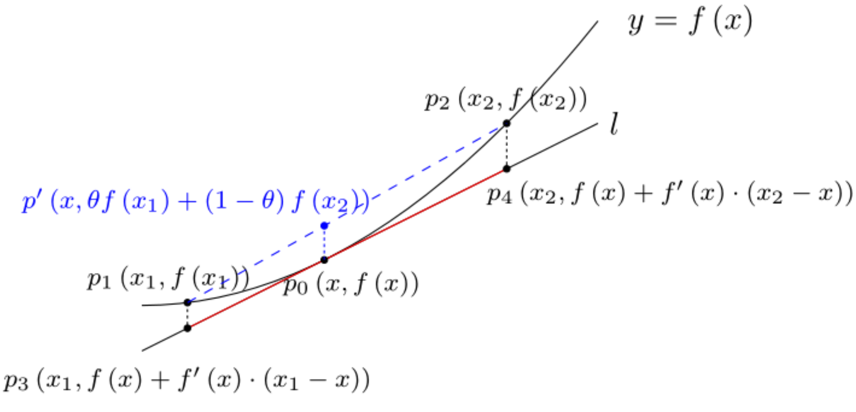
$$\begin{cases} \theta f(\mathbf{x}_1) & \geq \theta f(\mathbf{x}) + \theta \nabla f(\mathbf{x})(\mathbf{x} - \mathbf{x}_1) \\ (1-\theta) f(\mathbf{x}_2) & \geq (1-\theta) f(\mathbf{x}) + (1-\theta) \nabla f(\mathbf{x})(\mathbf{x} - \mathbf{x}_2) \end{cases}$$

将这两个不等式两边分别相加得

$$\begin{aligned} &\theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{x} - (\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2)) \\ &\theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) \geq f(\mathbf{x}) \\ &\theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) \geq f(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \end{aligned}$$

这个不等式恰好是凸函数的定义，至此我们可以说满足一阶条件不等式的可导函数符合凸函数的基本法，是凸函数。证毕。

这个证明的过程我们可以简化为一个一元函数$y = f(x)$来说明。根据已知条件，对于一组已知的x_1和x_2，找到之间任意的一点x（这个过程通过调节变量θ的大小），过函数图像上对于的点p_0作切线l，一阶条件不等式表达的含义即p_1和p_2必然不在这条切线l的下方，更具体来说就是p_1比p_3，p_2比p_4不知道高到哪里去了。



而我们在证明中把两个不等式分别乘以不同的系数再相加，这个操作定位了p'，这个点不低于p_0。两个不等式按一定的 “**权重**” 相加，由于(p_1, p_2)(Wallace组)对应的同横坐标的两个端点的纵坐标不小于(p_3, p_4)(Sharon组)，那么对于分别以两组点连接而成的线段，必然有(p_1, p_2)始终不低于(p_3, p_4)，即图中的蓝色虚线一定不在红色实线的下方。那么过点p垂直于y轴作直线与蓝色虚线的交点即p'一定也不在p的下方。

支持平面*（*Supporting Hyperplane*）

Xinyi Li

http://notebook.xyli.me/Introducti project-yangzhou/

> 它是凸函数，切平面怎么能不支持它？

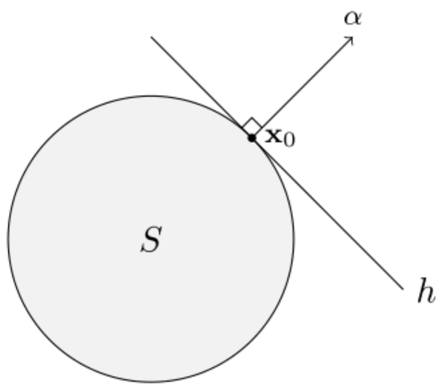
相比于定义对于割线的关注，一阶条件把注意力转移到了切线上。一阶条件不等式表达的几何意义在上述证明中我们简化为*"函数曲线比切线不知道高到哪里去了"*，这表达有时太肤浅了。

几何界有句老话，叫支持平面定理（*Supporting hyperplane theorem*）：（非空）凸集边界上的每个点，至少存在一个包含这个点的**支持平面。**

具体来说，对于一个非空凸集$S \subseteq R^n$，对于任意一点$\mathbf{x}_0 \in \mathrm{bd}S$或$\mathbf{x}_0 \notin S$，一定存在至少一个$\alpha \in R^n$且$\alpha \neq \mathbf{0}$使得：

$$\sup_{\mathbf{x} \in \mathbf{S}} \alpha^T \mathbf{x} \leq \alpha^T \mathbf{x}_0$$

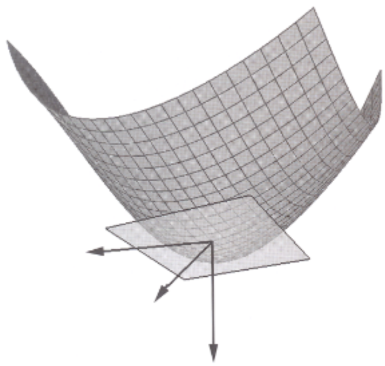
这个α是支持平面的法向量，这个不等式的意思就是整个集合的点在α上的投影都不高于$\mathbf{x}_0$在α上的投影



当然为了更加贴合前面的证明，我们可以取这里的α向量的相反向量代替α，那么有

$$\inf_{\mathbf{x} \in \mathbf{S}} \alpha^T \mathbf{x} \geq \alpha^T \mathbf{x}_0$$

整个集合的点在α上的投影都不低于$\mathbf{x}_0$在α上的投影，也就是法向量为α且过$\mathbf{x}_0$的平面h即点集S的支持平面，S始终在h的"上"方。



支持平面是怎样的平面？这是一个超平面，所以前面的证明中提到的切线也是一个支持平面。对于一元函数来说超平面就是切线，而n元函数的支持平面则是n维超平面，是函数在该点的**切平面**，梯度向量$\nabla f(\mathbf{x})$也就是这个切平面的法向量，集合S对应函数的上境图。支持平面定理中存在支持平面与凸集之间也是充要关系，这种关系有兴趣的读者可以自己尝试证明。

二阶条件（*Second-Order Condition*）

当且仅当对于定义域上任意自变量$\mathbf{x_0}$的Hessian矩阵$\mathbf{H}(\mathbf{x}_0)$均为半正定矩阵时，函数$y = f(\cdot)$是凸函数。

由于之前的证明已经说明了一阶条件和定义间的充要关系，接下来我们的证明将仅仅提供二阶条件与一阶条件间的等价推导，二阶矩阵与定义的关系我完全可以回答一句无可奉告。

Hessian矩阵（*Hessian Matrix*）

Xinyi Li

http://notebook.xyli.me/Introducti
project-yangzhou/

*Hessian矩阵*是一个函数的二阶偏导组成的方阵，用以描述一个点附近的曲率。有时我们也把$\mathbf{x_0}$的Hessian矩阵写为$\nabla^2 f\left(\mathbf{x}_0\right)$，可以把它理解为梯度向量在二阶的推广。毕竟梯度向量本身也是*Jacobian矩阵*的一种特例。

$f(\mathbf{x})$的Hessian矩阵第i行第j个元素为$f(\mathbf{x})$对第i个变量先求导，对第j个变量后求导的二阶导数，也就是$\mathbf{H}_{ij}=\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$，写成矩阵形式就是：

$$\mathbf{H}=\begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} \end{bmatrix}$$

必要性（*Necessity*）

本小节将从一阶条件出发，证明凸函数上任意一点对应的Hessian矩阵都是半正定的。

$f(\mathbf{x})$在$\mathbf{x_0}$的二阶*Taylor展开式*(带*Peano余项*)为：

$$f\left(\mathbf{x}\right)=f\left(\mathbf{x}_0\right)+\nabla f\left(\mathbf{x}_0\right)\left(\mathbf{x}-\mathbf{x}_0\right)+\frac{1}{2}\left(\mathbf{x}-\mathbf{x}_0\right)^T\mathbf{H}\left(\mathbf{x}_0\right)\left(\mathbf{x}-\mathbf{x}_0\right)+o\Big(\left(\mathbf{x}-\mathbf{x}_0\right)^T\left(\mathbf{x}-\right.$$

我们对整个等式稍作变换，规定$\Delta\mathbf{x}=t\mathbf{d}$表示增量，其中$\mathbf{d}$是一个微小的向量，仅仅表示变化的方向，而标量$t$则用以描述变化的步长，则上式可以改写为：

$$f\left(\mathbf{x}_0+t\mathbf{d}\right)=f\left(\mathbf{x}_0\right)+t\nabla f\left(\mathbf{x}_0\right)\mathbf{d}+\frac{t^2}{2}\mathbf{d}^T\mathbf{H}\left(\mathbf{x}_0\right)\mathbf{d}+o\left(t^2\right)$$

根据一阶条件有

$$f\left(\mathbf{x}_0+t\mathbf{d}\right)\geq f\left(\mathbf{x}_0\right)+t\nabla f\left(\mathbf{x}_0\right)\mathbf{d}$$

那么我们可以得到

$$\frac{t^2}{2}\mathbf{d}^T\mathbf{H}\left(\mathbf{x}_0\right)\mathbf{d}+o\left(t^2\right)\geq 0$$

恒成立，对于任意一个不为0的t，不等式两边同除以t^2，不等式方向不变：

$$\frac{1}{2}\mathbf{d}^T\mathbf{H}\left(\mathbf{x}_0\right)\mathbf{d}+\frac{o\left(t^2\right)}{t^2}\geq 0$$

注意到一阶条件对于两个变量x_1和x_2的限定词是“任意”，那么这个关系在取极限$t\rightarrow 0$时仍然成立，这是上面的不等式的第二项的极限$\lim\limits_{t\rightarrow 0}\frac{o(t^2)}{t^2}=0$，所以一阶条件蕴含的关系就是$\mathbf{d}^T\mathbf{H}\left(\mathbf{x}_0\right)\mathbf{d}\geq 0$对于任意的$\mathbf{d}$都成立，而这正好是$\mathbf{H}\left(\mathbf{x}_0\right)$为半正定矩阵的充要条件。

证毕。

充分性（*Sufficiency*）

本小节将从二阶条件出发逆推一阶条件，即证明任意一点的Hessian矩阵半正定时，一阶条件也满足，此时函数为凸函数。

现在已知在定义域内Hessian矩阵都是半正定的。那么在任意的$\mathbf{x}_0$使用带有*Lagrange余项*的一阶*Taylor展开式*为

$$f\left(\mathbf{x}\right)=f\left(\mathbf{x}_0\right)+\nabla f\left(\mathbf{x}_0\right)\left(\mathbf{x}-\mathbf{x}_0\right)+\frac{1}{2}\left(\mathbf{x}-\mathbf{x}_0\right)^T\mathbf{H}\left(\xi\right)\left(\mathbf{x}-\mathbf{x}_0\right)$$

其中$\mathbf{H}\left(\xi\right)$为在$\mathbf{x}=\xi$处的Hessian矩阵，其中$\xi$为$\mathbf{x}$与$\mathbf{x}_0$之间的一个值（闭区间）,由于$\mathbf{H}\left(\xi\right)$是半正定的，所以必然有$\frac{1}{2}\left(\mathbf{x}-\mathbf{x}_0\right)^T\mathbf{H}\left(\xi\right)\left(\mathbf{x}-\mathbf{x}_0\right)\geq 0$，因此对于任意的在定义域内的$\mathbf{x}$和$\mathbf{x}_0$都有

Xinyi Li

http://notebook.xyli.me/Introducti
project-yangzhou/

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$$

证毕。

Talyor定理*（*Taylor's Theorem*）

二阶条件的证明用到了非常重要的数学工具——Talyor展开式，我们知道Talyor展开是用以描述某点附近的近似值，使用一个多项式去逼近原函数。一元函数$y = f(x)$在点$(x_0, f(x_0))$处的n阶Talyor展开式为

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \ldots + \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n$$

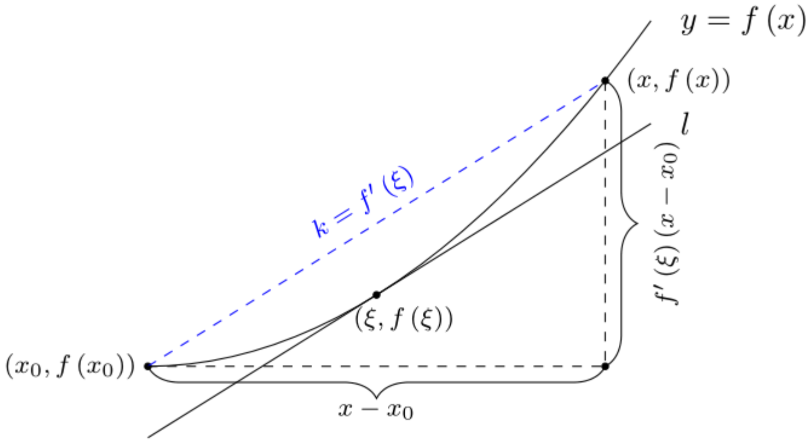
其中$R_n(x - x_0)$为余项，表示的是这个估计值与真实值的**偏差**，是一个$(x - x_0)^{n+1}$级的数，我们前面的证明用到了其中的两种形式。

我们可以看到“必要性”路线（定义⇒一阶条件⇒二阶条件）的证明我们一直都在增量趋近于无穷小的极限，所以展开式中使用的是*Peano余项*以方便我们忽略高阶无穷小的影响。我们知道随着阶数的增加，多项式估计值的准确度也在增加，取$\Delta x \to 0$时必须保证一阶条件的成立，那么微小的二阶增量也必须非负。

而“充分性”的证明路线（二阶条件⇒一阶条件⇒定义）则使用了*Lagrange余项*，去找到这个使等式成立的具体的某点附近的二阶导数。我们展开带有*Lagrange余项*的0阶Taylor展开式：

$$f(x) = f(x_0) + f'(\xi)(x - x_0)$$

这个式子明确的告诉我们，



高阶增量的系数就是区间内某个值的高阶导数，图中仅给出了0阶的情况，实际上这也是*Lagrange中值定理*的应用，可以推广到高阶。不同于“必要性”的宏观推导到微观的思想，“充分性”则是从每个定理成立的局部反推到全局成立，此时*Lagrange中值ξ*就扮演了非常重要的角色。

后记*（*Apply for Author*）

符号标记（*Notations*）

全文所用的符号和规定都出于我的个人习惯，比如多元函数的参数我使用$\mathbf{x}$是一个高高的列向量，可能其他的资料里面会使用宽宽的行向量，这时需要注意矩阵计算时一些变量的转置。毕竟你们非常熟悉行向量那一套，那么可以用你们自己的符号习惯再自己整理一下。

凹函数 （*Concave Function*）

是一个与凸函数对偶的概念，定义中的不等号方向改变，把一阶条件的$\geq$换成$\leq$，二阶条件中的半正定换成半负定，就是*凹函数*那一套。

严格凸函数（*Strictly Convex Function*）

类比严格单调函数的定义。在定义中去掉条件的不等号$0 < \theta < 1$，得到的几个结论也是没有等号成立的，对应的Hessian矩阵是**正定**的。

广义凸函数（*Generalized Convex Function*）

我们这么早就假设了函数是可导的，那么不可导的函数也可以是凸函数吗？当然啦。

我们规定可导只是方便我们推导一阶条件和二阶条件，如果导数不存在那么一阶条件和二阶条件也是不适用，但函数仍然是凸函数，它的上境图依然是凸集，在每个点上依然有支持平面。

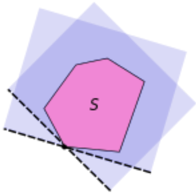
举个简单的例子就是分段函数

$$f(x)=\begin{cases}-x & x\le 0\\ 2x & x>0\end{cases}$$

Xinyi Li

http://notebook.xyli.me/Introducti project-yangzhou/

当然对于这种函数来说有些不可导的点上的支持平面就不止一个了。



为了**完善定义**，我们也可以扩展定义域（要求原定义域为凸集），原定义域上的函数值不变，空间内其他变量对应的函数值为无穷大，此时 $f\left(\theta x_1+(1-\theta)\,x_2\right)\le\theta f\left(x_1\right)+(1-\theta)\,f\left(x_2\right)$对于整个集合都适用，这个新函数也就是*广义凸函数*。

很惭愧，只做了一些微小的工作。

参考资料和延伸阅读（*References&Extensions*）

[1] Boyd S, Vandenberghe L. Convex optimization[M]. Cambridge university press, 2004.
[2] Wikipedia: https://en.wikipedia.org/wiki/Supporting_hyperplane
[3] Wikipedia: https://en.wikipedia.org/wiki/Taylor%27s_theorem
[4] *MATH2640 Introduction to Optimisation*: https://www1.maths.leeds.ac.uk/~cajones/math2640/notes2.pdf
[5] Lecture of Convex functions: https://www.seas.ucla.edu/~vandenbe/ee236b/lectures/functions.pdf
[6] Kolter Z. Convex Optimization Overview[J]. 2008. https://cs229.stanford.edu/section/cs229-cvxopt.pdf

< SMO算法：跑的比香港记者还快

我念了两句诗，然后就主动离开了知乎 >

Styling with Markdown is supported

Comment

Powered by Gitment

Xinyi Li

http://notebook.xyli.me/Introducti
project-yangzhou/