[优化中的subgradient方法](http://blog.csdn.net/lansatiankongxxc/article/details/46386341)

分类： [【优化】](http://blog.csdn.net/lansatiankongxxc/article/category/2228721)2015-06-06 11:03 301人阅读 [评论](http://blog.csdn.net/lansatiankongxxc/article/details/46386341#comments)(0) [收藏](javascript:void(0);) [举报](http://blog.csdn.net/lansatiankongxxc/article/details/46386341#report)

[优化](http://www.csdn.net/tag/%e4%bc%98%e5%8c%96)[次梯度](http://www.csdn.net/tag/%e6%ac%a1%e6%a2%af%e5%ba%a6)[gradient](http://www.csdn.net/tag/gradient)

目录[(?)](http://blog.csdn.net/lansatiankongxxc/article/details/46386341" \o "系统根据文章中H1到H6标签自动生成文章目录)[[+]](http://blog.csdn.net/lansatiankongxxc/article/details/46386341)

哎，刚刚submit上paper比较心虚啊，无心学习，还是好好码码文字吧。

subgradient介绍

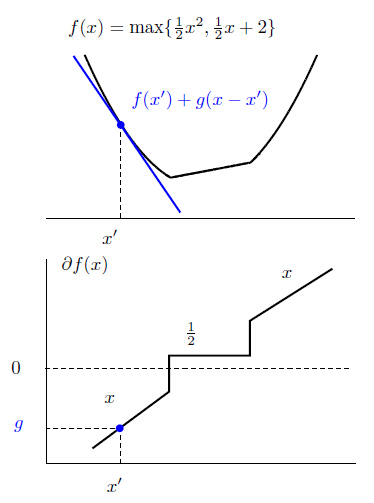
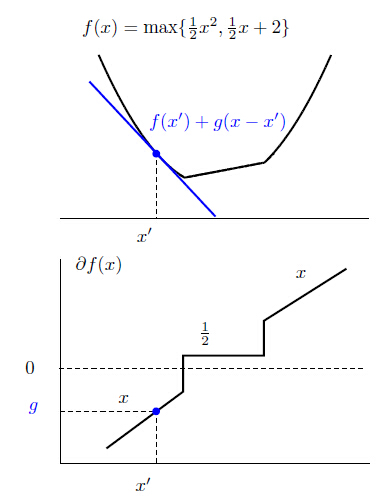
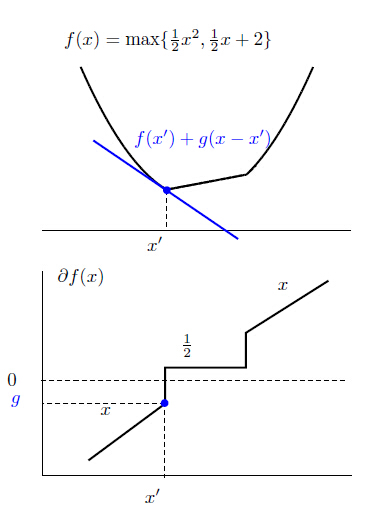
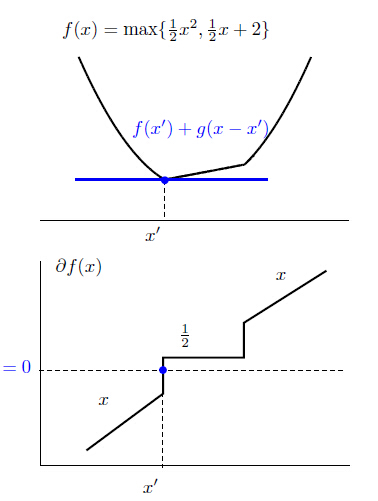
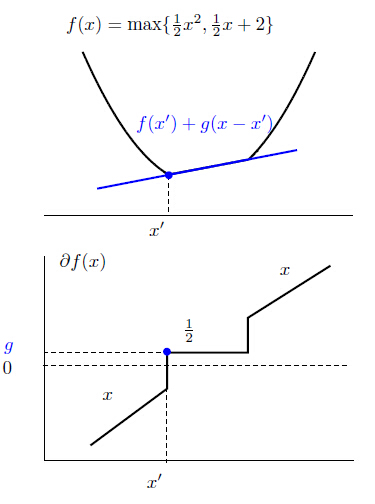
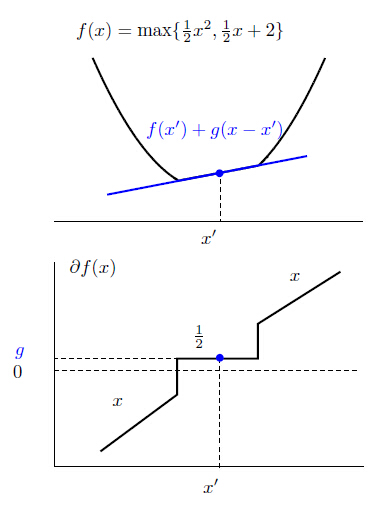
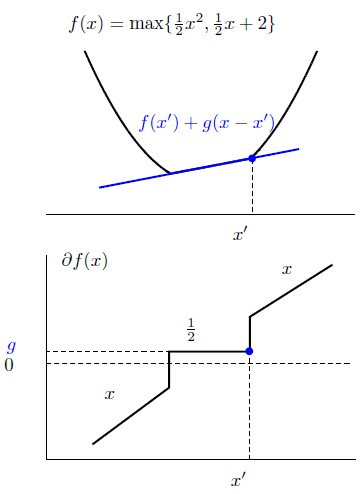
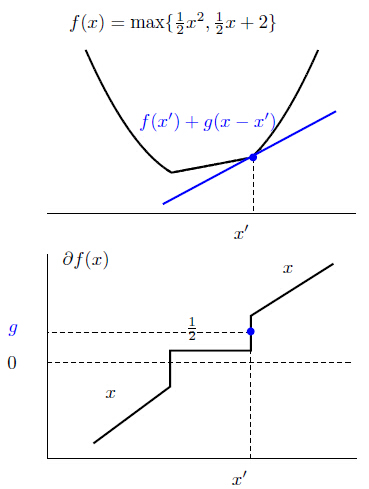
subgradient中文名叫次梯度，和梯度一样，完全可以多放梯度使用，至于为什么叫子梯度，是因为有一些凸函数是不可导的，没法用梯度，所以subgradient就在这里使用了。注意到，子梯度也是求解凸函数的，只是凸函数不是处处可导。

f:X→R是一个凸函数，X∈Rn是一个凸集。   
若是f在x′处∇f(x′)可导，考虑一阶泰勒展开式： 

f(x)≥f(x′)+∇(f(x′)T(x−x′),∀x∈X

能够得到f(x)的一个下届（f(x)是一个凸函数）   
若是f(x)在x′处不可导，仍然，可以得到一个f(x)的下届 

f(x)≥f(x′)+gT(x−x′),∀x∈X

这个g就叫做f(x)的子梯度，g∈Rn   
很明显，在一个点会有不止一个次梯度，在点x所有f(x)的次梯度集合叫做此微分∂f(x)   
   
   
   
   
   
   
   
   
我们可以看出，当f(x)是凸集并且在x附近有界时，∂f(x)是非空的，并且∂f(x)是一个闭凸集。

次梯度性质

∂f(x)={g}⇔f(x)可微并且g=∇f(x)

满足：   
1）scaling：

∂(αf(x))=α∂f(x),if α>0

2）addition：

∂(f1(x)+f2(x))=∂fz(x)+∂f2(x)

3）point-wise maximum:f(x)=maxi=1,...,mfi(x)并且fi(x)是可微的，那么： 

∂f(x)=Co{∇fi(x)∣fi(x)=f(x)}

即所有该点函数值等于最大值的函数的梯度的凸包。   
在非约束最优化问题中，要求解一个凸函数f:Rn→R的最小值 

x∗∈argminx∈Rnf(x)

很显然，若是f可导，那么我们只需要求解导数为0的点 

f(x∗=minx∈Rn⇔0=∇f(x∗)

当f不可导的时候，上述条件就可以一般化成 

f(x∗)=minx∈Rn⇔0∈∇f(x∗)

也即0满足次梯度的定义

f(x)≥f(x′)+0T(x−x′),∀x∈Rn

下面是次梯度法的一般方法：

1.t=1选择有限的正的迭代步长{αt}∞t=1   
2.计算一个次梯度g∈∂f(xt)   
3.更新xt+1=xt−αtgt   
4.若是算法没有收敛，则t=t+1返回第二步继续计算

次梯度方法性质：

1.简单通用性：就是说第二步中，∂f(xt)任何一个次梯度都是可以的.   
2.收敛性：只要选择的步长合适，总会收敛的   
3.收敛慢：需要大量的迭代才能收敛   
4.非单调收敛：−gt不需要是下降方向，在这种情况下，不能使用线性搜索选择合适的αt   
5.没有很好的停止准则

对于不同步长的序列的收敛结果

不妨设ftbest=min{f(x1),..,f(xt)}是t次迭代中的最优结果   
1.步长和不可消时（Non-summable diminishing step size）：   
limt→∞αt=0 并且∑∞t=1αt==∞   
这种情况能够收敛到最优解：limt→∞ftbest−f(x∗)=0   
2.Constant step size:   
αt=γ,where γ>0   
收敛到次优解：limt→∞ftbest−f(x∗)≤αG2/2   
3.Constant step length:   
αt=γ||gt||(i.e. ||xt+1−xt||=γ)，||g||≤G,∀g∈∂f   
能够收敛到次优解limt→∞ftbest−f(x∗)≤γG/2   
4.Polyak’s rule: αt=f(xt)−f(x∗)||gt||2   
若是最优值f(x∗)可知则可以用这种方法。

不等式约束的凸二次优化问题

问题formulate

一个不等式约束的凸二次优化问题可以表示为： 

(w∗,b∗,ξ∗)=argminw,b,ξ[12||w||2+C∑i=1mξi]

s.t.       yi(wTxi+b)ξi≥1−ξi,   ≥0              i=1,⋯,m,i=1,⋯,m.

注意到ξi≥max(0,1−yi(wTxi+b)),而且当目标函数取得最优的时候，这里的等号是成立的，所以可以进行代替:   
ξi=max(0,1−yi(wTxi+b))   
所以就可以将这个二次悠哈问题改写成一个非约束凸优化问题 

(w∗,b∗)=argminw,bf(w,b)=argminw,b[12||w||2f0(w,b)+C∑i=1mmax(0,1−yi(wTxi+b))fi(w,b)]

问题求解

因为

f0(w,b)=12||w||2

是可微的,并且   
∂wf0(w,b)=w,  ∂bf0(w,b)=0   
函数fi(w,b)=max0,1−yi(wTxi+b)是一个点最大值，所以其次微分可以写作，所有active function的梯度的convex combination

| i**-th function** | ∂wfi(w,b) | ∂bfi(w,b) |
| --- | --- | --- |
| I+={i|yi(wTxi+b)>1} | 0 | 0 |
| I0={i|yi(wTxi+b)=1} | Co{0,−yixi} | Co{0,−yi} |
| I−={i|yi(wTxi+b)<1} | −yixi | −yi |

所以次微分可以写作∂f(w,b)=∂f0(w,b)+C∑mi=1∂fi(w,b)可以使用参数话的表示方法，设0≤βi≤1,i∈I0,所以就有g=[w′b′]∈∂f(x)

w′(β)b′(β)=w−C∑i∈I0βiyixi−C∑i∈I−yixi=−C∑i∈I0βiyi−C∑i∈I−yi