Title: Follow-the-Regularized-Leader and Mirror Descent: Equivalence Theorems and L1 Regularization

Author: H.Brendan McMahan

Source: Google

Abstact:  
证明大部分的凸优化的真实下降算法都能够用FTRL算法进行解释。FOBOS, FTRL, RDA算法的不同在于如何处理L1 的累积惩罚。文章介绍的FTRL-proximal算法是FOBOS和RDA算法的结合，并且更适用于现实世界。

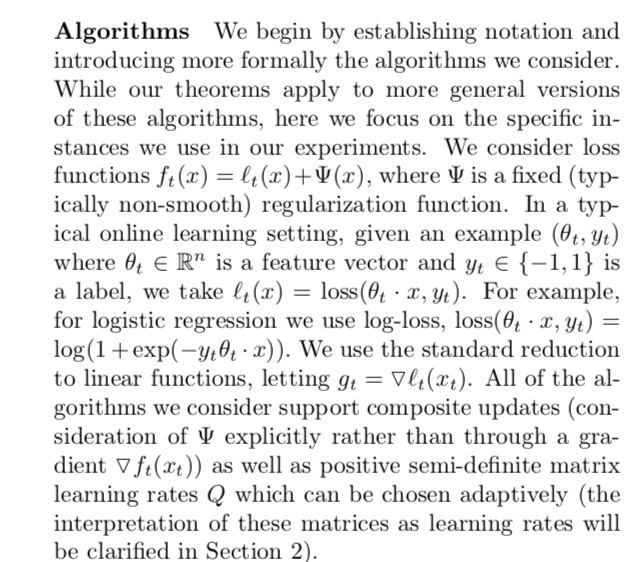
Introduction:  
从表面上看呢FTRL算法相比于使用梯度下降的FOBOS算法和RDA算法更相似。但是实际上通过正则化之后，这三者的不同主要在两个方面：  
1. How they choose to center the additional strong convexity used to guarantee low regret: RDA 在开始的时候就会预先定义；而FOBOS在训练的过程中确定

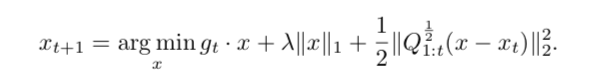
2. How they handle an arbitrary non-smooth regularization function. This include the mechanism of projection onto a feasible set and how L1 regularization is handled

为了比较以上三者呢，提出了FTRL-proximal算法。这个算法如果去掉了non-smooth term就和FOBOS一样。它的更新过程和dual averaging 一样，而它的additional strong convexity Is centered at the current feasible point.

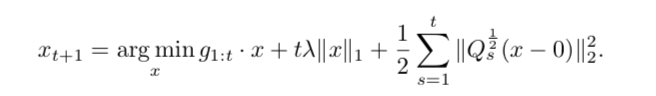
Algorithms:

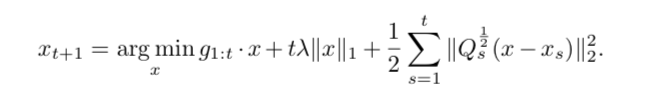
各参数的形式话定义：

  
FOBOS算法的形式：



RDA算法的形式：

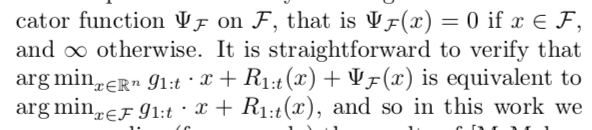
  
FTRL-Proximal算法的形式：



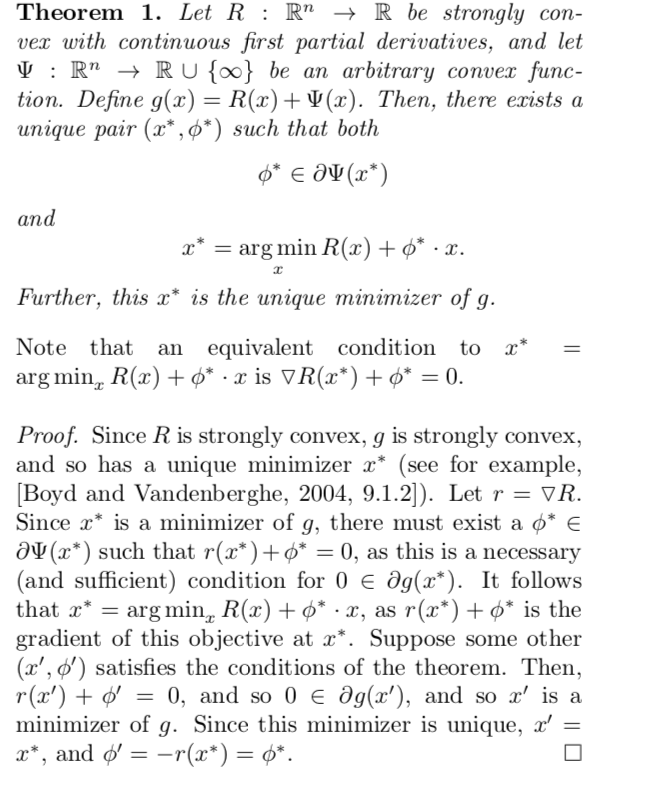
Efficient Implementations:  
对于有k个非零分量的gt可以在O(k)的时间内完成更新。

Feasible sets:

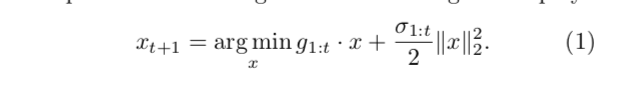
我们只需要讨论函数fi，而不需要讨论解空间



Notation and technical background

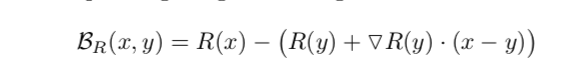


Mirror descent follows the leader

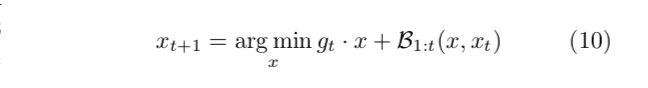
FTRL的形式：  


Mirror descent:  
the version captures the notion that we don’t want to change our hypothesis xt too much, but we do want to move in a direction that decreases the loss of our hypothesis on the most recently seen example.

通过定义Bregman divergence之后



它将形式变化为了：



同样的Composite-objective mirror descent(COMID)可以变成下面的形式：  


2.1 an equivalence theorem for proximal regularization

Mirror descent算法可以被看作FTRL算法：