总结报告25

Primal-dual Problem

（2020.4.5）

一、理论基础

|  |  |  |  |
| --- | --- | --- | --- |
| Source | Details | Remark | |
| 张贤达：矩阵分析与应用P216-P222 |  | 对于正交约束问题，ΘTΘ-I=0，由于它不是仿射函数（大概是说形如Ax+b，线性函数）的形式，所以，它**不能够**借助Lagrangian松弛方法换成一个凹函数的对偶无约束问题。  对于Loss（f0（x））它可以不是凸函数。对偶之后仍然是凹函数优化问题，但是这个时候就会有duality gap，不是原问题的最优解，而是次优解。 | |
| Yuxin Chen，Princeton University：Dual and primal-dual methods PPT  （接上） | Dual的推导：  Primal-dual的推导：    Can we update both primal and dual variables simultaneously and take advantage of both proxf and proxh?也就是这种情况利用primal-dual problem | 从推导可以看出，是否把f（x）+<λ，Ax>结合起来写成共轭的形式，就决定了最后到底产生的是dual形式还是primal dual形式。  Dual形式，只有λ一个变量，很多文献也给出了λ转化成x的公式；  而primal-dual形式，有x和λ两个变量，也就是一个鞍点问题。  这个PPT当中也给出了dual和primal-dual的prox算法，原始-对偶在每一次迭代中，原始变量和对偶变量都要更新，而dual可以只更新对偶变量，最后根据等式转换成原始变量解，（当然也可以x和λ都更新）。  这个PPT中dual和primal-dual methods的本质区别在于，是否原始变量和对偶变量的迫近算子都用到了。  目标函数是： | |
| TinTin博客：Primal-dual problem | Primal： | 其中，是regularizer，f(x)是分布式的loss | |
| 百度文库：对偶和鞍点问题PPT |  | | 鞍点定理其实和第一行张贤达说的是一个道理，要转换成鞍点问题，约束要满足凸函数、线性函数的条件；  鞍点与KKT条件之间的关系表明，满足KKT条件的点就是鞍点。 |

二、论文

|  |  |  |
| --- | --- | --- |
| Title | Objective functions | Note  (The property of regularizer) |
| Stochastic Dual Coordinate Ascent Methods for Regularized Loss Minimization | Primal:  Dual: | 正则项：L2-norm |
| Trading Computation for Communication: Distributed Stochastic Dual Coordinate Ascent | Primal:    Dual: | 正则项：g(w)要求是凸函数，但不要求光滑。比如，g（w）中含有L1-norm |
| A General Distributed Dual Coordinate Optimization Framework for Regularized Loss Minimization | Primal：  分布式原始：  分布式对偶： | g(w) is a strongly convex regularizer, h(w) is another convex regularizer  比如： |
| Stochastic Primal-Dual Coordinate Method for Regularized Empirical Risk Minimization | **Stochastic Primal-Dual Coordinate (SPDC)method：**basic idea: to approach the saddle point of f(x, y) defined in (4), we alternatively maximize f with respect to y, and minimize f with respect to x. Since the dual vector y has n coordinates and each coordinate is associated with a feature vector ai ∈ Rd, maximizing f with respect to y takes O(nd) computation, which can be very expensive if n is large. We reduce the computational cost by **randomly picking a single coordinate of y at a time, and maximizing f only with respect to this coordinate**. Consequently, the computation of each iteration is O(d). | g（x）凸 |
| Chambolle and Pock（2011） |  | 正则项要求凸。 |
| 鞍点问题和约束优化的几个一阶算法 尤燕飞 |  | 凸，非光滑。 |
| Stochastic Primal-Dual Method for Empirical Risk Minimization with O（1）Per-Iteration Complexity |  | 有一类算法的目标函数为  f（w）+g（Aw），这里的A可以data matrix. |

三、summary

总的来说，对于应用primal-dual，因为Loss应该就是ERM，那么对于regularizer来说，都是要求是凸的，但是可以不光滑，比如L2-norm，L1-norm,至于非凸的目前还没有看到。根据张贤达的矩阵分析与应用，应该这种正交约束难以用对偶的方法解决，但是，一些论文给出了正交约束的算法的迭代步骤。也就是当单独优化Θ：，这种算法是有一些的。