总结报告24

论文CoCoA : A General Framework for Communication-Efficient Distributed Optimization及相关论文

（2020.3.20）

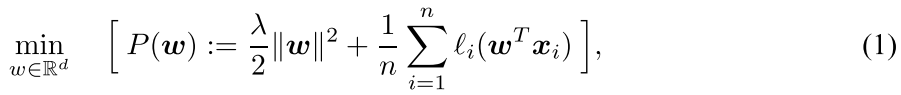
一、Contents

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **简称** | **题目** | **备注** |
| 1  (2014) | CoCoA | Jaggi, M., Smith, V., Takáč, M., Terhorst, J., Krishnan, S., Hofmann, T., & Jordan, M. I. (2014). **Communication-efficient distributed dual coordinate ascent**. Advances in Neural Information Processing Systems, 4(January), 3068–3076. | The core idea is to use the dual variables to efficiently merge the parallel updates from the different workers without much conflict, by exploiting the fact that they all work on disjoint sets of dual variables.  Only require communication of a single vector for each worker, that is ∆wk∈Rd. |
| 2  (2018) | Generalized  CoCoA | Smith, V., Takáč, M., & Jordan, M. I. (2018). **CoCoA : A General Framework for Communication-Efficient Distributed Optimization.** 18, 1–49. | A General Framework：  Leveraging a novel approach in handling non-strongly-convex regularizers and non-smooth loss functions. |
| 3  (2015) | CoCoA+ | Ma, C., Smith, V., Jaggi, M., Jordan, M. I., Richtárik, P., & Takáč, M. (2015). **Adding vs. averaging in distributed primal-dual optimization**. 32nd International Conference on Machine Learning, ICML 2015, 3, 1973–1982. | A novel generalization of the local COCOA **subproblems**:  The set of locally computed updates from the modified subproblems can be combined more efficiently.  COCOA+ updates can be aggressively added(Controlled by γ), which yields much faster convergence (extremely K becomes large). |
| 4  (2015) | PROX CoCoA+ | Smith, V., Forte, S., Jordan, M. I., & Jaggi, M. (2015). **L1-Regularized Distributed Optimization: A Communication-Efficient Primal-Dual Framework**. 1–23. | Cover non-strongly convex regularizers and losses.  PROX COCOA+ will result in distributing the data column-wise (by feature), and having the vector **v** be of length equal to the number of data points, which reduces communication costs. |

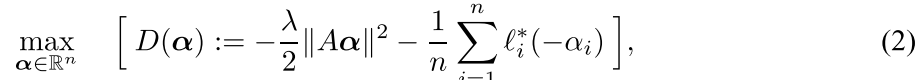
注：方法的不断演进是根据：一方面是对于子问题设计的改进；另一方面是模型更加general，即将更多的非凸非光滑losses或regularizers情况包含进去。

|  |  |  |
| --- | --- | --- |
| 1 | CoCoA | Jaggi, M., Smith, V., Takáč, M., Terhorst, J., Krishnan, S., Hofmann, T., & Jordan, M. I. (2014). **Communication-efficient distributed dual coordinate ascent**. Advances in Neural Information Processing Systems, 4(January), 3068–3076. |

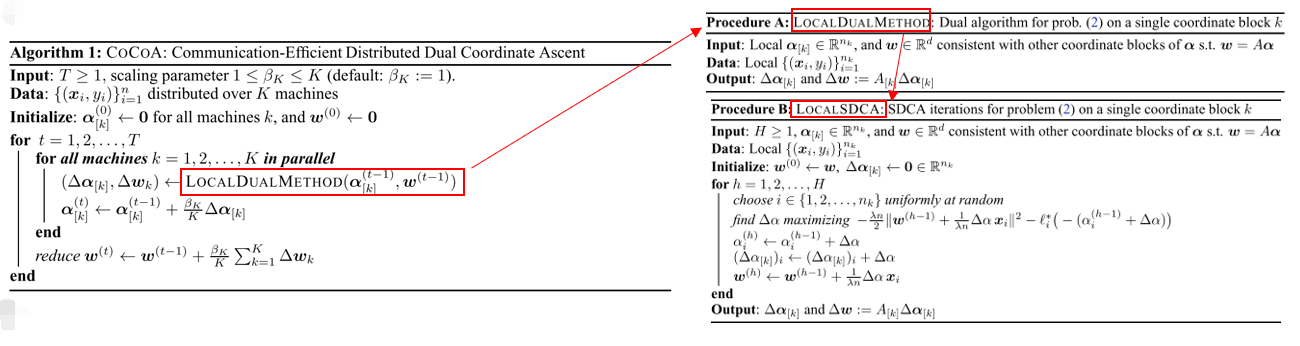
原始：



对偶：



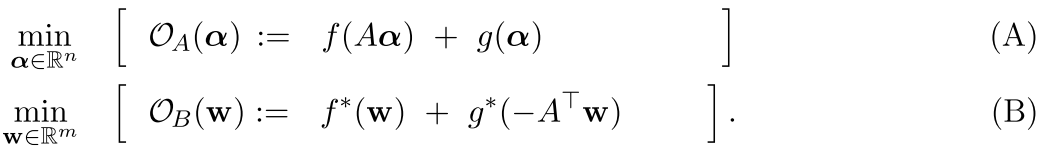
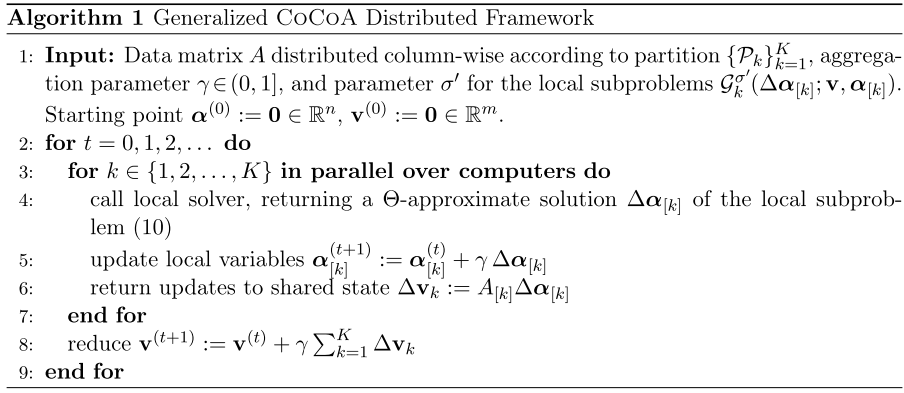
算法：

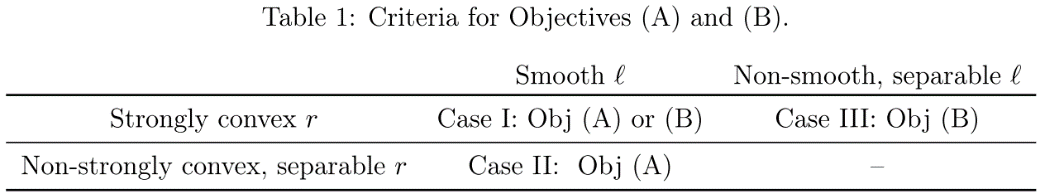


注：Local SDCA（Stochastic Dual Coordinate descent）是Local Dual Method的其中一种方法，当然也可以采用其他方法。

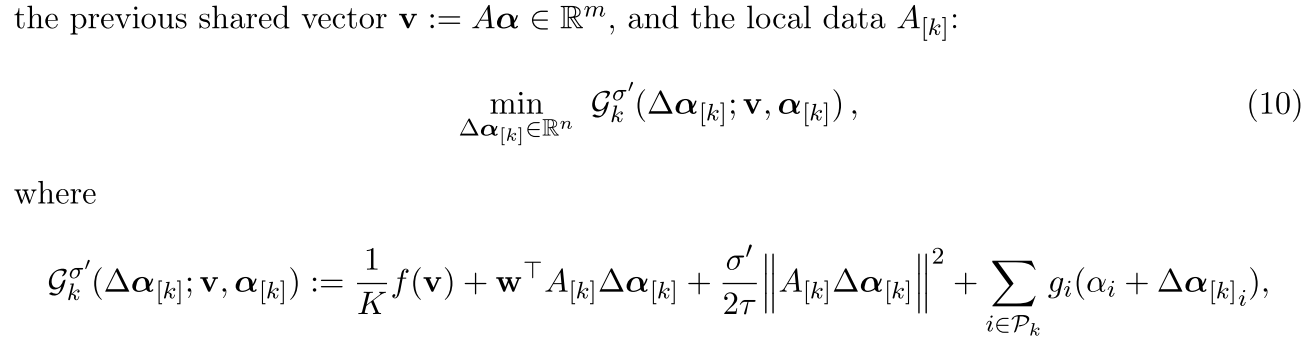
|  |  |  |
| --- | --- | --- |
| 2 | Generalized  CoCoA | Smith, V., Takáč, M., & Jordan, M. I. (2018). **CoCoA : A General Framework for Communication-Efficient Distributed Optimization.** 18, 1–49. |

原始/对偶： 算法：



几种非凸非光滑情况：

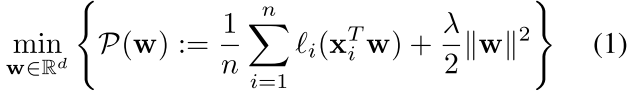
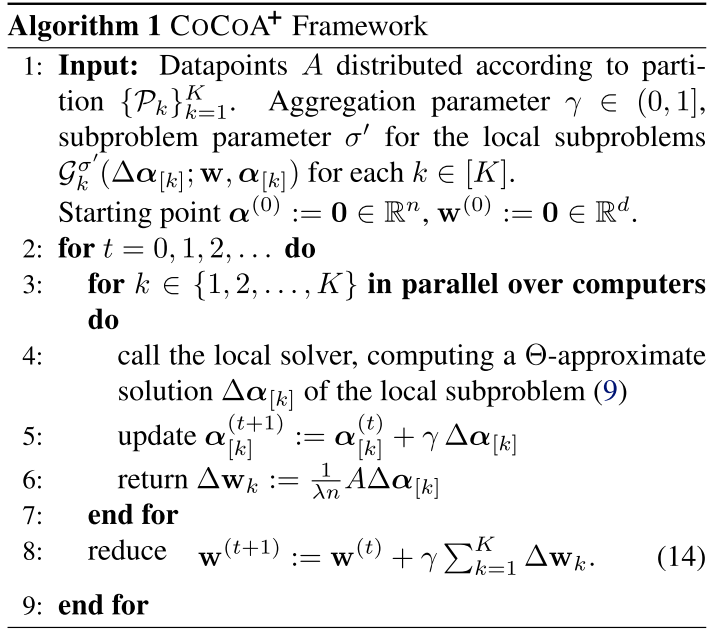
子问题（同论文4）：



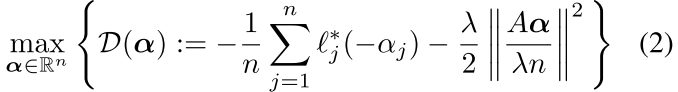
注：至于子问题subproblem为什么这么定义，见原论文Appendix D，Lemma 1

|  |  |  |
| --- | --- | --- |
| 3 | CoCoA+ | Ma, C., Smith, V., Jaggi, M., Jordan, M. I., Richtárik, P., & Takáč, M. (2015). **Adding vs. averaging in distributed primal-dual optimization**. 32nd International Conference on Machine Learning, ICML 2015, 3, 1973–1982. |

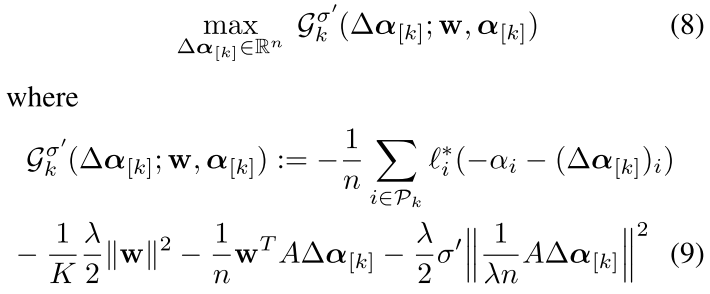
原始： 算法：



对偶：



子问题：

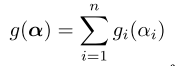


**核心**：1. 修改了subproblem

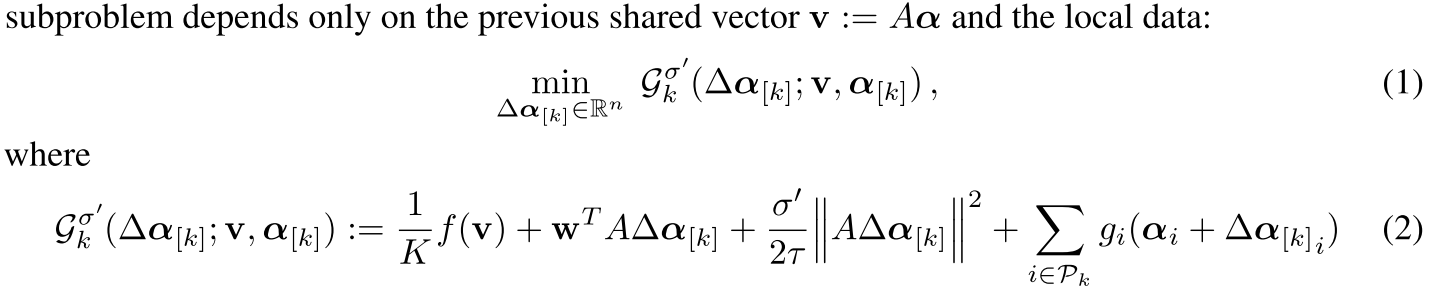
2. A aggregation parameter γ

|  |  |  |
| --- | --- | --- |
| 4 | PROX CoCoA+ | Smith, V., Forte, S., Jordan, M. I., & Jaggi, M. (2015). **L1-Regularized Distributed Optimization: A Communication-Efficient Primal-Dual Framework**. 1–23. |

目标函数：

 其中，

Subproblem（同论文2）：



Note:σ’——the subproblem parameter, which is a data-dependent term measuring the difficulty of the data partitioning

算法：

