Project Proposal: Exemplar-based Solvers in Intel MKL Libraries

Xin Lin
TACC_ID: jimmylin
jimmylin@utexas.edu

Yi-Chu Chen
TACC_ID: ycchen
ycchen886@utexas.edu

Juan
TACC_ID: jtrejo13
jtrejo13@utexas.edu

1. Introduction

The Intel MKL library [1] have implemented a wide range of optimizations for linear algebra computation. To the best of our knowledge, its functional completeness and robustness are unprecedented among all linear-algebra libaries. Thus, the MKL library should serve as the standardized infrastructure for most contexts of academic research and scientific computing.

On the other hand, exemplar-based methods gain its popularities in the domain of Bayesian Nonparametrics. Despite of the guaranteed optimality of solutions, the solvers of exemplar-based methods are not practically applicable for its non-linear polynomial computational complexity. To make this method more scalable, one considerable option is to implement all of associated linear algebra operations with advanced scientific solvers (i.e. Intel MKL Library).

2. Technical Plan

In this research-oriented final project, we plan to re-implement two ADMM-based iterative solvers by means of Intel MKL library calls. These two solvers are originally developed for demonstrating the theoretical optimality of associated statistical models and computational algorithms in solving the medoid clustering problem and the multiple sequence alignment problem. Both problem solvers are based on the ADMM computational framework and exemplar-based approach in the domain of Bayesian Nonparametrics.

The first problem solver is to resolve the medoid clustering problem. The statistical modelling behind is recorded in [3]. The iterative ADMM-based program will be decomposed into two subproblems: (a) frank-wolfe algorithm and (b) proximal algoirthm. Operations in both subproblems will be refined for better computational efficiency.

The second problem solver is much more complicated. It targets at solving Multiple Sequence Alignment with the guaranteed optimality [2]. Similarly, the ADMM framework requires us to iteratively resolve two subproblems: (a) tensor-based frank-wolfe algorithms and (b) tensor-based proximal algorithm.

We expect our new implementation to improve the solvers' performance by a large margin for the computational optimizations already achieved in the MKL library. And hopefully, the final version of implementation could make the applications of exemplar-based methods more practical for their scientific usages.

3. Partner Plan

Workload Allocation Yi-Chu and Juan will be in charge of re-implementing the examplar-based medoid clustering solver, where all system variables are restored in two-dimensional arrays. Xin will be working on re-implementing the multiple sequence alignment solver, where high-dimensional tensors are involved and the performance improvement could be more complicated.

Timing Schedule After deliberate group discussions, the timing schedule we all agree to is as follows: we leave two weeks for most of codebase-related work (developing the MKL-based version of the solver and running performance experiments), and two weeks for composing posters and the final project report.

Team Coordination Our team work will be coordinated through the github repository `https://github.com/JimmyLin192/SDS394_ExemplarPerfOpt` for project source codes and the sharelatex: `https://www.sharelatex.com/project/580facdb4f7d6d58762b9a0e` for paper works.

References

[1] M. Intel. Intel math kernel library, 2007.

[2] I. E. Yen, X. Lin, E. J. Zhang, E. P. Ravikumar, and I. S. Dhillon. A convex atomic-norm approach to multiple sequence alignment and motif discovery.

[3] I. E. Yen, X. Lin, K. Zhong, P. Ravikumar, and I. Dhillon. A convex exemplar-based approach to mad-bayes dirichlet process mixture models. In *Proceedings of The 32nd International Conference on Machine Learning*, pages 2418–2426, 2015.