

Scalable Online Optimization algorithms and the Structural SMV model

William St-Arnaud, Elyes Lamouchi, Frédéric Boileau

Introduction

Certain structured SVM's with multilabel outputs have an exponentially large number of constraints, which makes the problem inefficient or intractable in practice. There has been a lot of research focused on providing a solution to that issue. The Frank-Wolfe algorithm [1] is a popular method for constrained convex optimization consisting in taking a first-order approximation of the objective function and doing a linear search over the set of constraints to update the current point. Over the years, many improvements and alternatives were developed to address the particularities of various frameworks. Jaggi [2] presents a variety of stronger convergence results for Frank-Wolfe-type algorithms found in the literature (e.g. F-W with approximate linear subproblems, away steps, etc.). Lacoste-Julien et al. [4] propose a randomized block-coordinate variant of the F-W algorithm. The paper goes on to demonstrate its use in the case of SVM's and the advantages it yields over stochastic subgradient methods and other available SVM solvers. S. Lacoste-Julien and M.Jaggi [3] also go on to show that the "Away-steps", "Pairwise FW", "fully-corrective FW", and "Minimum norm point" variants of the F-W algorithm all achieve linear convergence under a weaker condition than strong convexity.

Project

The type of project we wish to present is at the crossing of two lines of work: analysis and practical evaluation. Our project will, therefore, be composed of two main parts:

- The first part is a literature review on the subject. We intend to go deeper into some of the papers cited above, studying the assumptions they make and identifying the critical points in the proofs that forbid using softer assumptions. We will also compare the papers, showing how they relate to each other and how they overcome previous difficulties related to structured predictions. Finally, this part would serve as an overview of all previous results that will allow us to identify the insights and limitations of this class of results and potentially some holes in the results which could be used as a starting point for future research.
- The second part consists of a practical evaluation of the method proposed in and of other methods depending on the time constraints. The idea is to see first-handedly the effect of the various variants of FW on real data. This way, we can better compare/justify the usage of these methods in real-world scenarios. We intend to run some evaluation of the model on both toy/synthetic datasets and real datasets with a few numbers of network configurations. datasets, we aim to construct a practical example in which this method helps the model to converge to an interesting global minimum. On real datasets, the objective would be to determine if the gain from using this method is significant or not in real case uses when compared to more conventional methods.

Our project will take the form of an article detailing the analysis of the papers and our experiments on real data, as well as a repository including the code used for the experiments.

References

- [1] Marguerite Frank and Philip Wolfe. “An algorithm for quadratic programming”. In: *Naval Research Logistics Quarterly* 3.1-2 (1956), pp. 95–110. URL: <https://EconPapers.repec.org/RePEc:wly:navlog:v:3:y:1956:i:1-2:p:95-110>.
- [2] Martin Jaggi. “Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization.” In: *Proceedings of the 30th International Conference on Machine Learning* (2013), pp. 427–435. URL: <http://infoscience.epfl.ch/record/229246>.
- [3] Simon Lacoste-Julien and Martin Jaggi. “On the Global Linear Convergence of Frank-Wolfe Optimization Variants”. In: *arXiv e-prints*, arXiv:1511.05932 (2015), arXiv:1511.05932. arXiv: 1511.05932 [math.OC].
- [4] Simon Lacoste-Julien et al. “Stochastic Block-Coordinate Frank-Wolfe Optimization for Structural SVMs”. In: *CoRR* abs/1207.4747 (2012). arXiv: 1207.4747. URL: <http://arxiv.org/abs/1207.4747>.
- [5] Benjamin Taskar, Simon L. Julien, and Michael I. Jordan. “Structured Prediction, Dual Extragradient and Bregman Projections”. In: *Journal of Machine Learning Research* 7 (2006), pp. 1627–1653. URL: <http://dblp.uni-trier.de/rec/bibtex/journals/jmlr/TaskarLJ06>.