CIG Final Report Summary

PI: Elad Hazan

1 FINAL PUBLISHABLE SUMMARY REPORT

This final report pertains to project ”Sublinear Optimization for Machine Learning” carried out at the Technion, PI: Elad Hazan, and generously funded through the Marie Curie program.

Can we solve optimization problems in using computational resources proportional to the information necessary to represent and verify the solution?

The interplay between information and computation is at the core of the hardest problems of computer science and mathematics. However, statistical problems arising in machine learning exhibit a much more attainable version of this question, which has been a major focus of this research project. In the past few years we have been able to successfully design algorithms that run in time proportional to the information theoretic limit of verifying a solution. These algorithms run in time which is lesser than time to perform even a linear scan of the input, thus are called sublinear optimization algorithms. Surprisingly, this phenomenon is not limited to isoteric or contrived computational models, but rather hold for perhaps the most widely used optimization algorithms in use today: linear classification and kernel machines.

The main investigation topic of our project is to develop provably correct algorithms which run in sublinear time and/or space - i.e. they do not observe all the data even once, hence applicable to super-scale problems. Specific objectives listed in the project proposal are:

1. Develop a sublinear-time solver for the SVM (support vector machine) optimization problem. Our solver will depend sublinearly on the two main complexity parameters of the corresponding optimization problem: the number of examples and the dimension. Current solvers have at least linear dependence on both of these parameters. Two variants of this tool were considered, soft margin SVM, and ”Lasso”.

2. Enable efficient use of Kernels. We will develop sublinear time algorithms for classification using kernels, i.e. mappings to high dimensional space, whose structure enables efficient inner-product computation. The kernel method allows one to compute non-linear classifiers, and in particular Gaussian classifiers, which are immensely useful in practice.

3. Develop sublinear algorithms for PCA (principle component analysis) and matrix factorization. These matrix operations are the basic building blocks of collaborative filtering and other methods for recommendation systems.

Indeed, all three components of these investigation areas were researched during our project. In particular, we have attained the following results (which were all published in broad audience top-tier machine learning venues):

1. The first sublinear soft-margin SVM algorithm was designed, implemented and benchmarked. A publication describing this development was accepted and published in the proceedings of NIPS (a major annual conference of machine learning) with the title of ”Beating SGD: Learning SVMs in Sublinear Time”.

2. The above paper discusses applications to kernels, along the lines of the original sublinear algorithms for machine learning, which were enhanced with full detail to application to various kernels including the Gaussian and polynomial kernels in the paper titled ”Sublinear optimization for machine learning” (see bibliography below).

3. Application of sublinear time technics to the operator and matrix world were explored with respect to application to semi-definite programming. In particular, we have developed the first sublinear time algorithm for semi-definite programming. A publication describing this result was accepted and published in the proceedings of NIPS 2011, titled ”Approximating Semidefinite Programs in Sublinear Time”.
4. We have extended the sublinear time framework to consider optimization and learning with partial access to information. A publication detailing near-optimal algorithms for learning with partial attributes was presented in ICML 2012, titled ”Linear Regression with Limited Observation”.

After researching these aspects of our project, we have continued to explore further two main directions. The first is projection free optimization and learning, a subject on which two further publication have been published in top venues: The 29th International Conference on Machine Learning (ICML 2012) and the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2013). These are detailed below.

The other direction is applications of our sublinear time learning and optimization algoirthms to the field of robust optimization. This has resulted in the first oracle-based robust optimization method, and a paper was accepted for publication in ”Operations Research” journal.

2 USE AND DISSEMINATION OF FOREGROUND

The main product of our research are scientific publications that have passed the scrutiny of the scientific community through the peer-review process, as well as the education of researchers that join the EU scientific community. Our findings were published as papers in broad-audience scientific conferences and journals as detailed below.

Four Ph.D students were trained and are near completion of their thesis. One student is expected to graduate this year. Three others next year. Two masters student was trained and graduated as part of our project. One more is expected to graduate this year.

The PI is well underway to write a graduate text book, which is freely available of his website.

A project website is in place at: http://sublrn.net.technion.ac.il/

Publications that have directly been produced as a product of this research are listed below. Co-authors are listed alphabetically. with asterix* appended to to their name

Published journal articles and peer-reviewed conference papers

1. Oracle-Based Robust Optimization via Online Learning
Aharon Ben-Tal, Elad Hazan, Tomer Koren* and Shie Mannor
accepted to Operations Research (OR)
2. Playing Non-linear Games with Linear Oracles
Dan Garber* and Elad Hazan
54th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2013)
3. Projection-free Online Learning.
E. Hazan and S. Kale
The 29th International Conference on Machine Learning (ICML 2012)
4. Linear Regression with Limited Observation.
E. Hazan and T. Koren*
Proceedings of the 29th International Conference on Machine Learning (ICML), 2012, pages 807–814
ICML 2012 Best Student Paper Runner Up
5. Beating SGD: Learning SVMs in Sublinear Time
E. Hazan, T. Koren* and N. Srebro
Twenty-Fifth Annual Conference on Neural Information Processing Systems (NIPS 2011)

6. Approximating Semidefinite Programs in Sublinear Time
   D. Garber* and E. Hazan
   Twenty-Fifth Annual Conference on Neural Information Processing Systems (NIPS 2011)

References

[CHW12] Kenneth L. Clarkson, Elad Hazan, and David P. Woodruff. Sublinear optimization for machine learning. *J. ACM*, 59(5):23:1–23:49, November 2012.

[FGKP06] Vitaly Feldman, Parikshit Gopalan, Subhash Khot, and Ashok Kumar Ponnuswami. New results for learning noisy parities and halfspaces. *Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 563–574, 2006.

[GH11] Dan Garber and Elad Hazan. Approximating semidefinite programs in sublinear time. In John Shawe-Taylor, Richard S. Zemel, Peter L. Bartlett, Fernando C. N. Pereira, and Kilian Q. Weinberger, editors, *NIPS*, pages 1080–1088, 2011.

[HKS11] Elad Hazan, Tomer Koren, and Nati Srebro. Beating sgd: Learning svms in sublinear time. In *25th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 1233–1241, 2011.

[MP88] Marvin Minsky and Seymour Papert. *Perceptrons: An introduction to computational geometry*. MIT press Cambridge, Mass, 1988.

[Nov63] Albert B. Novikoff. On convergence proofs for perceptrons. In *Proceedings of the Symposium on the Mathematical Theory of Automata*, volume 12, pages 615–622, 1963.

[SS02] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press, 2002.

[SS03] Bernhard Schölkopf and Alexander J. Smola. A short introduction to learning with kernels. *Advanced lectures on machine learning*, pages 41–64, 2003.