ARC-Fellowship: Implementation of fast first-order methods for solving large-scale convex optimization problems

By: Camilo Ortiz*
Advisor: Renato Monteiro†

June 30, 2011

For some years now, there has been an increased interest in the use and development of first-order methods in convex optimization. Even though progress has been made in sophisticated interior point (second-order) methods (e.g., several variants and combinations of a quasi-Newton method with a technique to solve linear equations), when it comes to solving large-scale problems, the need for "cheaper" iterations continues to exist. Thus, we can argue that there is always going to be an instance of a problem or an application where the current methods fail to succeed strictly because of the size of the data.

The difficulty with the size of the data is not necessarily a storing issue. More frequently, failure arises because of the limitations that the size imposes on the set of operations that can be performed within reasonable time. Here, first-order methods play a key role as they only use cheap operations at every iteration. Moreover, various recent results in first-order methods like in [8, 9, 12, 13, 14, 19] efficiently solve important classes of convex optimization problems that even include non-smooth functions.

In our research, we focus on the particular case of solving large-scale (millions of variables/constraints) convex optimization problems formulated as follows:

$$z^* := \min \{ f(x) \ : \ x \in \Omega \}, \tag{1}$$

where $\Omega \subset \mathbb{R}^n$ is a closed convex nonempty set and $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex, lower-semicontinuous and C^1 in Ω. This is a very general optimization problem that includes conic semidefinite programming problems (SDPs) as a special case, namely when f is linear ($f(\cdot) = \langle c, \cdot \rangle$) and Ω is the intersection of an affine manifold and the positive semidefinite cone ($\Omega = \{x \in \mathbb{R}^{n\times n} : \mathcal{A}x = b, x \succeq 0\}$). It is well-known that within the many applications of solving SDPs we have the solution of good relaxations of NP-hard combinatorial problems including max-clique problems (see [6]), frequency assignment problems (or max k-cut problem, see [1]), quadratic assignment problems (see[15]), traveling salesman problems (see [2]) and general binary integer quadratic problems.

Our general approach for solving (1) is the application of a family of approximate proximal-point methods, called hybrid proximal extragradient (HPE) methods, first introduced in [17, 18] by Solodov and Svaiter, and whose complexity is derived in [8, 10]. These methods solve monotone inclusion problems and we specialize them to solve the particular case of convex optimization (as the optimality conditions can be viewed as an inclusion problem). So far, in [7], we obtained very encouraging results with an algorithm that applies a block-decomposition HPE (BD-HPE) method (introduced in [9]) to SDPs. This algorithm outperformed the two most competitive codes for large-scale conic semidefinite programs, namely: the boundary point method introduced in [14] and the Newton-CG augmented Lagrangian method in [20]. The benchmark reported in [7] compares our method against these two alternatives by testing various types of SDPs, including those arising from randomly generated sparse

*PhD student at ISyE: camiort@gatech.edu
†Professor at ISyE: renato.monteiro@isye.gatech.edu

matrices. In [7], several speed-up ingredients are introduced to the BD-HPE method in its pure form, drastically improving the efficiency of the resulting algorithm.

It is important to point out that interior point methods are still the leading approach when it comes to solving conic programming problems with high accuracy. Moreover, the method in [20] shows that a second-order Newton-CG augmented Lagrangian method solves even large-scale SDPs extremely efficiently by exploiting the structure of an estimator of the generalized Hessian. However, second-order methods like this one rely heavily on the starting point, and therefore, use first-order methods as a first phase to provide a more robust overall method (e.g., [20] uses the first-order method in [14] to find the starting point).

Now, the goal of our research is to obtain efficient implementations of first-order algorithms for solving (1), including instances that may not be SDPs, with the same approach of specializing a variant of the HPE methods as described above. In fact, recently it has been proved in [11] that an accelerated version (A-HPE) of the HPE method generalizes a variant of Nesterov's optimal method as in [13, 19]. Motivated by these results, we plan to implement a first-order application of the A-HPE method to solve 1, using various speed-up ingredients, including the ones previously used in [7].

Among the ingredients considered to speed-up the HPE methods in its pure form, we include: an aggressive choice of stepsize for performing the extragradient step; the use of scaled inner products in the primal and dual spaces; a dynamic update of the scaled inner product in the primal space for properly balancing the primal and dual relative residuals; and proper choices of the initial primal and dual iterates, as well as the initial parameter for the primal scaled inner product. In addition, we are interested in approximately solving some of the matrix multiplications and singular value decompositions (SVD) performed at each iteration of the algorithm. This implies that the matrix multiplication (including matrix-vector multiplication) would be approximately computed by random sampling approaches as in [3] or [5], and the SVD would be approximately computed by adaptive sampling (see [3]), volume sampling (see [4]) or random projection (see [16]) approaches. These last ingredients tackle two mayor bottlenecks of first-order methods when applied to the special case of large-scale SDPs: matrix multiplications; and projections of symmetric matrices onto the semidefinite cone.

Even though there is an explicit calculation of the iteration complexity of the HPE methods, it is not guaranteed that we will derive equal or better iteration complexity for the resulting algorithms (obtained after applying all of the speed-up ingredients mentioned). The approximate calculations and the structure of many of these techniques may overlay on basic assumptions required for obtaining the convergence theoretical bounds of the method implemented. We believe this is one of the main theoretical challenges when introducing randomness and heuristics into algorithms that solve deterministic problems. However, in practice, we will make numerical tests to give us some insight of how the theoretical complexity may behave, as well as to check the empirical value of our approach.

References

[1] Samuel Burer, Renato D.C. Monteiro, and Yin Zhang. A computational study of a gradient-based log-barrier algorithm for a class of large-scale SDPs. *Mathematical Programming*, 95:359–379, 2003. 10.1007/s10107-002-0353-7.

[2] Etienne de Klerk, Dmitrii V. Pasechnik, and Renata Sotirov. On semidefinite programming relaxations of the traveling salesman problem. *SIAM Journal on Optimization*, 19(4):1559–1573, 2008.

[3] Amit Deshpande, Luis Rademacher, Santosh Vempala, and Grant Wang. Matrix approximation and projective clustering via volume sampling. *Theory of Computing*, 2(1):225–247, 2006.

[4] Amit Deshpande and Santosh Vempala. Adaptive sampling and fast low-rank matrix approximation. In *APPROX-RANDOM'06*, pages 292–303, 2006.

[5] Petros Drineas and Ravi Kannan. Fast Monte-Carlo algorithms for approximate matrix multiplication. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 452–459, 2001.

[6] L. Lovasz. On the Shannon capacity of a graph. *Information Theory, IEEE Transactions on*, 25(1):1 – 7, jan 1979.

[7] Renato D. C. Monteiro, Camilo Ortiz, and Benar F. Svaiter. Implementation of a block-decomposition algorithm for solving large-scale conic semidefinite programming problems. *Optimization-online preprint 3032*, pages 1–32, 2011.

[8] Renato D. C. Monteiro and Benar F. Svaiter. Complexity of variants of Tseng's modified F-B splitting and Korpelevich's methods for generalized variational inequalities with applications to saddle point and convex optimization problems. *Optimization-online preprint 2675*, pages 1–30, 2010.

[9] Renato D. C. Monteiro and Benar F. Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating minimization augmented Lagrangian method. *Optimization-online preprint 2713*, pages 1–28, 2010.

[10] Renato D. C. Monteiro and Benar F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20(6):2755–2787, 2010.

[11] Renato D. C. Monteiro and Benar F. Svaiter. An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods. *Optimization-online preprint 3030*, pages 1–32, 2011.

[12] Arkadi Nemirovski. Prox-method with rate of convergence O(1/t) for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

[13] Yu. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103:127–152, 2005. 10.1007/s10107-004-0552-5.

[14] J. Povh, F. Rendl, and A. Wiegele. A boundary point method to solve semidefinite programs. *Computing*, 78:277–286, 2006. 10.1007/s00607-006-0182-2.

[15] Janez Povh and Franz Rendl. Copositive and semidefinite relaxations of the quadratic assignment problem. *Discrete Optimization*, 6(3):231 – 241, 2009.

[16] Tamas Sarlos. Improved approximation algorithms for large matrices via random projections. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 143–152, Washington, DC, USA, 2006. IEEE Computer Society.

[17] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient – proximal point algorithm using the enlargement of a maximal monotone operator. *SetValued Analysis*, 7(4):323–345, 1999.

[18] M. V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. *Journal of Convex Analysis*, 6(1):59–70, 1999.

[19] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. *submitted to SIAM Journal on Optimization*, 2008.

[20] Xin-Yuan Zhao, Defeng Sun, and Kim-Chuan Toh. A Newton-CG augmented lagrangian method for semidefinite programming. *SIAM Journal on Optimization*, 20(4):1737–1765, 2010.