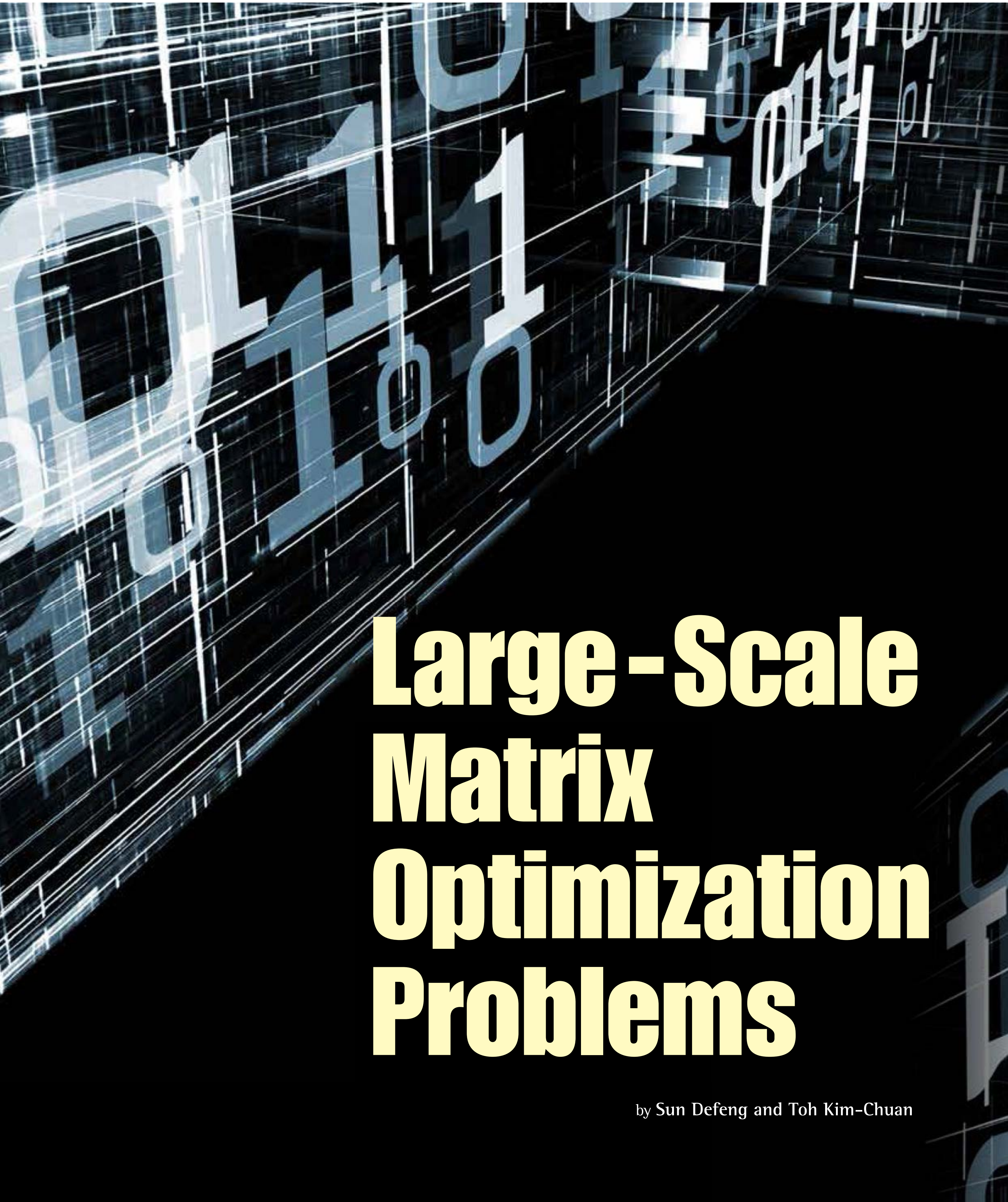
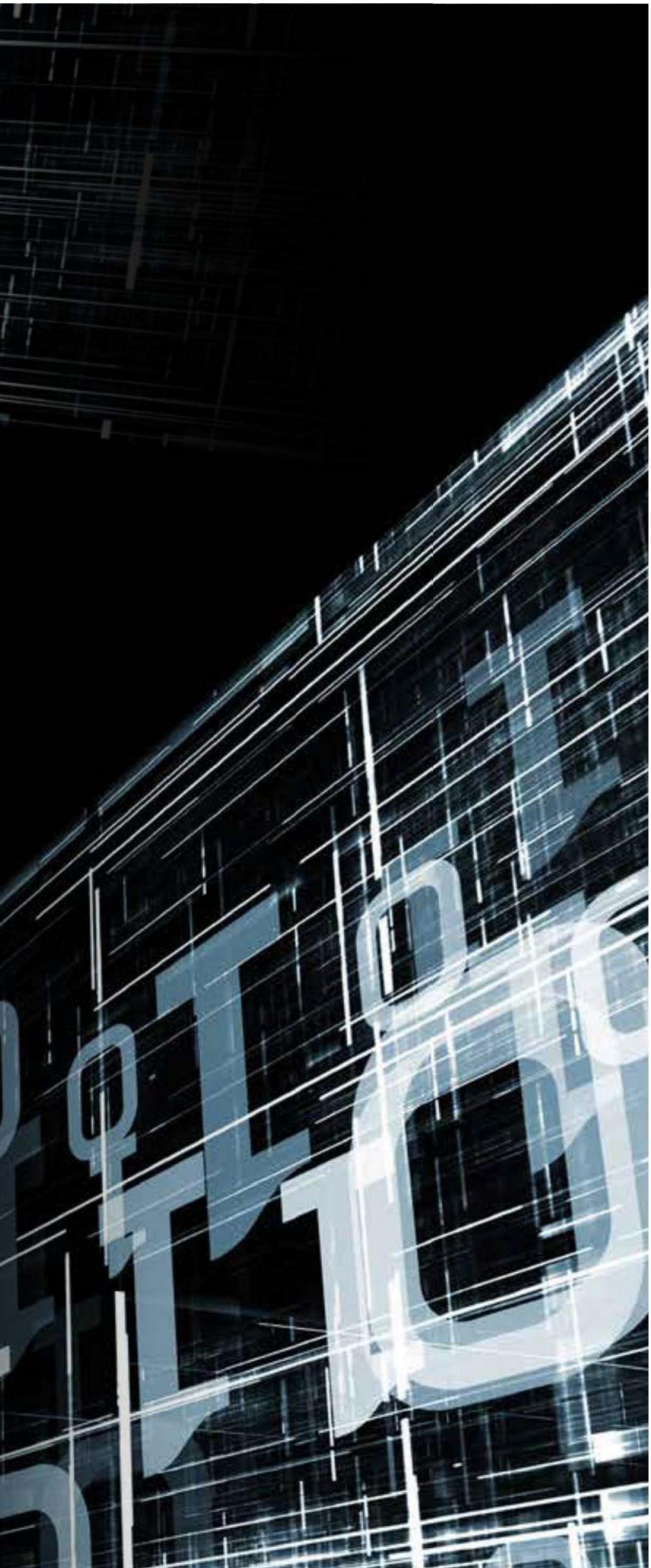


Large-Scale Matrix Optimization Problems

by Sun Defeng and Toh Kim-Chuan



Convex matrix cone programming (including the most notable class of semidefinite programming (SDP) [13]), is a major development in optimization in the last two decades, for which some thought to be comparable to the revolutionary development of linear programming in the 1950s. The optimization problems under consideration deal with convex cost functions and linear constraints, but the matrix variables are constrained to be in some specific convex cones. For the case of a linear SDP, the problem is to optimize a linear cost function subject to the constraint that the matrix decision variable must lie in the cone of symmetric positive semidefinite matrices while also satisfying some linear equality and/or inequality constraints.

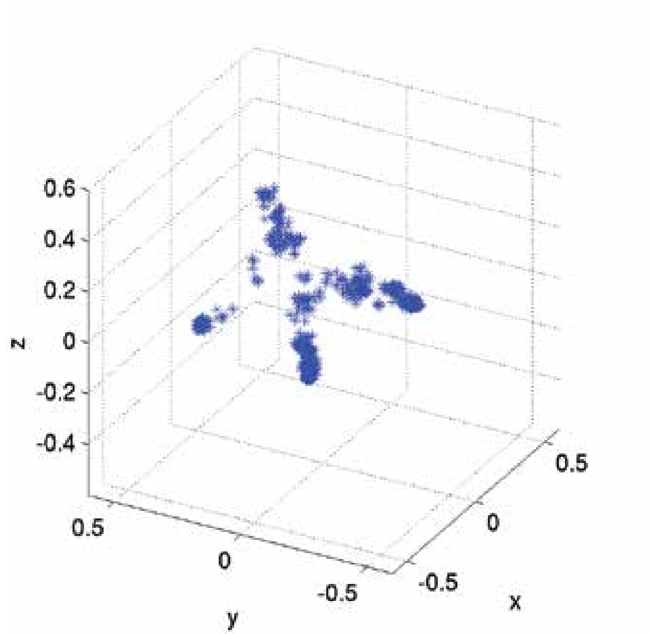
It is known that SDP includes linear programming as an important special class which on its own has been widely used as a quantitative modeling tool in engineering/economic/business. However the modeling power and applications of SDP and its generalizations are far wider. Today, the applications of SDP, and more generally matrix optimization (MatOpt) problems, span a wide variety of areas including linear matrix inequality in systems and control, combinatorial optimization, quantum information, machine learning, signal processing and communications, structural optimization, Euclidean metric embedding and sensor network localization, covariance matrix estimations in statistics, risk management and finance, robust optimization, and matrix approximation problems.

The applications of MatOpt appear to be unlimited, as one may extrapolate from exciting developments in the past two decades, and especially the recent explosive developments in diverse areas. For example, in matrix completion, SDP technique, via solving a nuclear norm minimization problem, is used to fill in missing entries in a partially specified matrix, such as movie ratings by viewers on movies they had watched. In many applications such as matrix completion, the success depends critically on our ability to solve large-scale MatOpt problems efficiently.

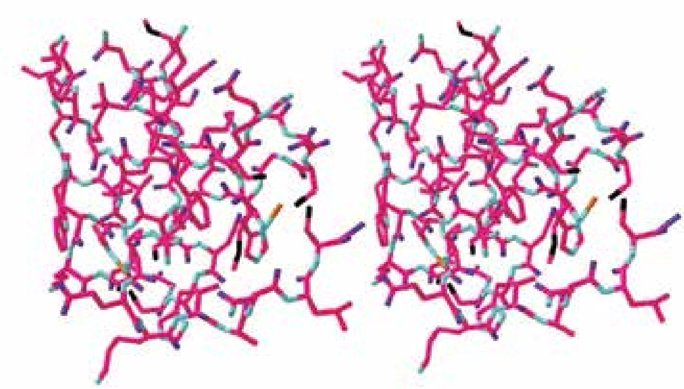
While many MatOpt problems can be reformulated as standard SDPs by adding auxiliary constraints and variables and/or increasing the dimension of the matrix variables, such a reformulation is however not practically viable since the reformulated SDPs would contain unnecessarily large number of additional constraints and/or have unnecessarily large matrix dimensions. This unfortunate situation is exacerbated by the fact that algorithms (as implemented in interior-point method based solvers such as SeDuMi by J. Sturm [9] or SDPT3 [14,16] by the second author) which are highly successful in handling medium-scale general SDP problems can no longer meet the demand of large-scale problems because the interior-point algorithmic framework would generally lead to excessive computational cost and core memory requirements. Thus a different algorithmic framework must be designed for large problems with structures which can be exploited effectively to reduce the computational demand.

In our work over the past five years, we have designed, analyzed and developed a variety of semismooth Newton-CG (conjugate gradient) proximal-point and augmented Lagrangian algorithms for solving several important classes of large scale MatOpt problems. In particular, we have designed a highly successful two-phase augmented Lagrangian method (called SDPNAL+) for solving large-scale linear SDP+ problems for which the matrix variables are constrained to be both positive semdefinite and nonnegative [19,20]. The class of SDP+ problems is important because such problems typically arise from convex relaxation of NP-hard combinatorial problems such the quadratic assignment problems, maximum stable set problems, as well as the k-clustering problems. As an illustration of the efficiency of our method, to solve an SDP+ problem whose matrix dimension is 200 and the number of linear equality constraints is 10000, the most advanced interior-point based SDP solvers such as SDPT3 or SeDuMi would take at least 20 hours to solve such a problem in a high-end desktop PC available today. But our SDPNAL+ solver can solve such problem in a matter of a few minutes.

Unlike interior-point algorithms, the design of efficient and robust proximal-point and augmented Lagrangian-based algorithms are highly dependent on the structures available for the class of problems under consideration. The key to successfully design efficient and robust proximal-point or augmented Lagrangian-based algorithms hinges critically on our ability to extract and exploit the specific features/structures in various problem classes so that the inner optimization sub-problems appearing in each iteration of the algorithms are computationally tractable. As a result, different classes of problems would require different algorithmic designs in order to successfully exploit the underlying structures to achieve computational efficiency. Besides having developed highly efficient solvers for linear SDP and SDP+ problems, in our work, we have also developed solvers for other important classes of MatOpt problems including (a) large scale convex quadratic SDP [6]; (b) log-determinant maximization problems with mixed-norm regularizations which are important for estimating covariance matrices in the high-dimension low-sample-size setting [17,18]; (c) covariance selection problems; (d) large scale Euclidean distance matrix estimation problems which arise in various Euclidean metric embedding problems such as embedding protein sequences in three-dimensional space for clustering purpose, sensor network localization and molecular conformation [8]; (e) large scale nuclear/spectral norm regularized matrix least squares problems with structural polyhedral constraints [5,7] which arise naturally in relaxations of low-rank matrix completion [1], and robust principal component analysis [2].



Clustering of 630 protein sequences via Euclidean metric embedding of sequence dissimilarities.


Simulated molecular conformation via semidefinite programming.

In our approach to design and develop new algorithms based on the inexact proximal or augmented Lagrangian frameworks, we analyzed and showed that certain large-scale dense linear systems of equations arising in the semismooth Newton proximal-point and augmented Lagrangian algorithms will be moderately well-conditioned when certain constraint nondegeneracy conditions hold whereas the corresponding counterpart in the commonly used interior-point methods would inherently be ill-conditioned. The implication of such a contrasting property is that algorithms based on proximal-point and augmented Lagrangian framework would have a great potential to efficiently solve large-scale MatOpt problems compared to the popular interior-point framework, because the Krylov subspace iterative solvers (such as CG) necessarily needed to solve the large linear systems of equations would be efficient for the former but not the latter. Provided that the inner optimization sub-problems have desirable theoretical properties such as satisfying constraint nondegeneracy conditions and strong semismoothness, we can show that the efficient variants of semismooth Newton-CG methods we have developed to solve them are guaranteed to have at least superlinear local convergence.

The success of our designed algorithms is built on the comprehensive theoretical studies we have conducted on the closed-form solutions and semismoothness properties of the metric projectors onto various classes of convex cones as well as the proximal-point mappings of various nonsmooth functions such as the nuclear norm and spectral norm (see [3,4,10]). These theoretical results not only provide the fundamental computational building blocks for our design of algorithms for solving large-scale problems involving those cones or nonsmooth functions, they are also important for the convergence analysis of the algorithms developed. In particular, the theoretical results are essential for analyzing the constraint nondegeneracy conditions needed for local quadratic convergence of semismooth/smoothing Newton methods [11] designed to solve the inner sub-problems.

In addition to semismooth Newton-CG proximal-point and augmented Lagrangian-based algorithms which made use of second-order information wisely, we have developed highly successful first-order algorithms based on accelerated proximal gradient methods and semi-proximal alternating direction methods of multipliers (SPADMM) for the various classes of MatOpt problems mentioned above [6,12,15]. In particular, we have recently developed a convergent semi-proximal alternating direction method of multipliers for MatOpt problems with four-block constraints. The importance of these first-order methods lies in the fact that (a) they are simpler to implement compared to methods based on second-order information (b) sometimes they are highly efficient for finding a solution of moderate accuracy (c) they can be used to warm-start the second-order based methods. Indeed, the highly efficient convergent SPADMM we have developed served a very important role in warm-starting our SDPNAL+ solver.

Currently, our research on algorithms for large-scale MatOpt problems are among the most general and advanced, and our algorithms are considered to be among the most successful for solving a variety of large-scale problems with millions of constraints. *i*

References

[1] E.J. Candes and B. Recht, Exact matrix completion via convex optimization, Foundation of Computational Mathematics, 2008, 9:717-772.

[2] E.J. Candes, X. Li, Y. Ma and J. Wright, Robust principal component analysis? Journal of the ACM, 2009, 58:1-37.

[3] C. Ding, D.F Sun and K.C. Toh, An introduction to a class of matrix cone programming, Mathematical Programming, 2014, 144:141-179.

[4] C. Ding, D.F. Sun, J. Sun, and K.C. Toh, Spectral operators of matrices, arXiv preprint, arXiv:1401.2269, 2014.

[5] K.F. Jiang, D.F. Sun, and K.C. Toh, A partial proximal point algorithm for nuclear norm regularized matrix least squares problems, Mathematical Programming Computation, 2014, 6:281-325.

[6] K.F. Jiang, D.F. Sun, and K.C. Toh, An inexact accelerated proximal gradient method for large scale linearly constrained convex SDP, SIAM Journal on Optimization, 2012, 22:1042-1064.

[7] Y.J. Liu, D.F. Sun, and K.C. Toh, An implementable proximal point algorithmic framework for nuclear norm minimization, Mathematical Programming, 2012, 133:399-436.

[8] N.-H. Z. Leung and K.C. Toh, An SDP-based divide-and-conquer algorithm for large scale noisy anchor-free graph realization, SIAM Journal on Scientific Computing, 2009, 31:4351-4372.

[9] J.F. Sturm, Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones, Optimization Methods and Software, 1999, 11-12:625-633.

[10] D.F. Sun and J. Sun, Semismooth Matrix Valued Functions, Mathematics of Operations Research, 2002, 27:150-169.

[11] D.F. Sun, The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications, Mathematics of Operations Research, 2006, 31:761-776.

[12] D.F. Sun, K.C. Toh and L.Q. Yang, A convergent proximal alternating direction method of multipliers for conic programming with 4-block constraints, arXiv preprint, arXiv:1404.5378, 2014.

[13] M.J. Todd, Semidefinite optimization, Acta Numerica, 2001, 10:515-560.

[14] K.C. Toh, M.J. Todd, and R.H. Tutuncu, SDPT3 --- a Matlab software package for semidefinite programming, Optimization Methods and Software, 1999, 11:545-581.

[15] K.C. Toh, and S.W. Yun An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems, Pacific Journal of Optimization, 2010, 6:615-640.

[16] R.H. Tutuncu, K.C. Toh, and M.J. Todd, Solving semidefinite-quadratic linear programs using SDPT3, Mathematical Programming, 2003, 95:189-217.

[17] C.J. Wang, D.F. Sun, and K.C. Toh, Solving log-determinant optimization problems by a Newton-CG primal proximal point algorithm, SIAM Journal on Optimization, 2010, 20:2994-3013.

[18] J.F. Yang, D.F. Sun, and K.C. Toh, A proximal point algorithm for log-determinant optimization with group Lasso regularization, SIAM Journal on Optimization, 2013, 23:857-893.

[19] L.Q. Yang, D.F. Sun, and K.C. Toh, SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints, arXiv preprint, arXiv:1406.0942, 2014.

[20] X.Y. Zhao, D.F. Sun, and K.C. Toh, A Newton-CG augmented Lagrangian method for semidefinite programming, SIAM Journal on Optimization, 2010, 20:1737-1765.

Sun Defeng and Toh Kim-Chuan Toh are professors in the Department of Mathematics at the National University of Singapore.

For more information about this article, please contact Professor Sun at matsundf@nus.edu.sg or Professor Toh at mattohkc@nus.edu.sg