AFRL-AFOSR-VA-TR-2019-0127

Exploiting Geometry and Degeneracy in Large Scale Structured Optimization

Dmitriy Drusvyatskiy
UNIVERSITY OF WASHINGTON

03/07/2019
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
02/27/2019

2. REPORT TYPE
final

3. DATES COVERED *(From - To)*
6/01/2015-10/31/2018

4. TITLE AND SUBTITLE
Exploiting Geometry and Degeneracy in Large Scale Structured Optimization

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-15-1-0237

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Drusvyatskiy, Dmitriy

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Washington, Office of Sponsored Programs, 4333 Brooklyn Ave NE, Seattle, WA 98195-0001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/PK3, USAF, AFRL DUNS 143574726, AF OFFICE of SCIENTIFIC RESEARCH, 875 NORTH RANDOLPH STREET, RM 3112, ARLINGTON, VA 22203-1954

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION A

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Funding from the Air Force has played an instrumental role in my research. The broad goals of the proposed work were two-fold: (1) ``to develop `facial reduction algorithms' for large-scale highly structured problems'' and (2) to ``develop new algorithms for nonsmooth and nonconvex problems, which converge rapidly under favorable conditiones.'' In parallel this work aimed to apply the techniques to pervasive large scale problems in computational mathematics and the applied sciences. All the objectives were successfully met. Some highlights of the research include a new robust algorithm for graph

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Funding from the Air	Unclassifie d	Unclassified		19b. TELEPHONE NUMBER *(Include area code)*

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATE COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33315-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report. e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

PROJECT SUMMARY PI: DMITRIY DRUSVYATSKIY

Exploiting Geometry and Degeneracy in Large Scale Structured Optimization

Grant: FA9550-15-1-0237

Final Report: 6/01/2015 - 10/31/2018

Funding from the Air Force has played an instrumental role in my research. The broad goals of the proposed work were two-fold: (1) "to develop 'facial reduction algorithms' for large-scale highly structured problems" and (2) to "develop new algorithms for nonsmooth and nonconvex problems, which converge rapidly under favorable conditions." In parallel, this project aimed to apply the techniques to pervasive large scale problems in computational mathematics and the applied sciences. All objectives were successfully met, as I explain below. The results of the research were summarized in over 20 publications in high calibre journals, including Math Prog., SIAM J. Optim., Math of Oper. Res. and Found. Comput. Math. Support from the AFOSR has funded three of my students, Kellie MacPhee, Scott Roy, and Courtney Paquette. The three students participated in all aspects of the project, coauthoring papers and presenting at conferences. Sections 1-5 describe the research highlights of the project, organized by topic; the names of the three students appear in blue in the bibliographic citations. The final Section 6 lists the conferences and colloquia that the PI attended as an invited speaker during the award period.

1 Facial Reduction: exploiting degeneracy in large-scale optimization

The central thrust of this project was the realization that numerous large-scale computational tasks lead to optimization problems that are inherently degenerate. Degeneracy is meant here in a precise mathematical sense; intuitively, it means that the computational task has much fewer degrees of freedom than the standard problem formulation suggests, thereby leading to ill-conditioning. Typical examples include convex relaxations of various NP-hard problems, such as sensor network localization, graph partitioning, and quadratic assignment. Yet another rich class of example comes from sum-of-squares relaxation techniques. Rather than a nuisance, such degeneracy can be exploited leading to smaller and more stable problem formulations. The formal procedure for exploiting such degeneracies is called *facial reduction.* A large part of the project focused on exploring the relationship between facial reduction and the convex geometry of the optimization problem. For example, the paper 3 showed that facial reduction can complete in a single step if and only if the data of the problem lies on a facially exposed face of a certain convex cone. This shows that nonexposed faces form an obstruction to efficient use of facial reduction. In parallel, we developed efficient facial reduction procedures for concrete problems. In particular, we developed a new robust algorithm for graph realization and extensively tested it on sensor network localization problems. This line of work led to the articles 2-4 below, as well as the expository survey article 1. The work 5 is another example of exploiting degeneracy, although in a different sense. There, we considered a wide class of optimization problems with PDE constraints. For this class of problems, it appears very natural to apply the classical quadratic penalization technique of nonlinear programming. The main difficulty is that quadratic penalization leads to problem formulations with highly ill-conditioned Hessians. Remarkably, we showed that this degeneracy can be eliminated entirely by implicitly and iteratively minimizing out the objective in one set of the variables. We numerically validated the proposed algorithms on boundary control and optimal transport problems.

1. D. Drusvyatskiy, H. Wolkowicz, 'The many faces of degeneracy in conic optimization', *Foundations and Trends in Optimization*, Vol. 3, No. 2, pp 77-170, 2017.

2. D. Drusvyatskiy, N. Krislock, Y.-L. Voronin, and H. Wolkowicz, 'Noisy Euclidean distance realization: robust facial reduction and the Pareto frontier', *SIAM J. Optim.*, 27-4 (2017), 2301-2331.

3. D. Drusvyatskiy, G. Pataki, H. Wolkowicz, 'Coordinate shadows of semidefinite and Euclidean distance matrices', *SIAM J. Optim.* 25 (2015), no. 2, 1160-1178.

4. D. Drusvyatskiy, G. Li, H. Wolkowicz, 'A note on alternating projections for ill-posed semidefinite feasibility problems'. *Math. Program.*, 162 (2017), no. 1-2, Ser. A, 537-548.

5. A.Y. Aravkin, D. Drusvyatskiy, T. van Leeuwen, 'Efficient quadratic penalization through the partial minimization technique'. *IEEE Trans. Automat. Contr.*, 63 (2018), no. 7, 2131-2138.

2 Regularity and rapid convergence of algorithms

The second thrust of the proposal was to develop new algorithms for nonsmooth and nonconvex problems, which converge rapidly under favorable conditions. The PI pursued this agenda in a series of publications listed below. Roughly speaking, this work applied to two different types of optimization problems: nonconvex feasibility and convex composite minimization. The nonconvex feasibility problem is to find a point in the intersection of two nonconvex sets—a ubiquitous task across computational mathematics. A typical assumption one makes is that finding the nearest-point to each individual set is computationally tractable. Under this assumption, the method of alternating projections proceeds by iteratively finding the nearest point to the first set, then the second, then the first, and so on. In the publication 5, the PI together with coauthors showed that a simple geometric condition, called transversality, guarantees the method converges linearly to a point in the intersection. This is the sharpest currently available guarantee for the method. In parallel, the PI investigated algorithms for minimizing composition of convex functions with smooth maps. Such problems appear often in applied mathematics and data science. The publications 1-2 developed subgradient and Gauss-Newton methods that provably converge linearly (or faster) under the mild regularity assumption that the objective grows sharply away from the solution set. Moreover, the publications 2-3 theoretically and numerically validated the approach on phase retrieval, quadratic sensing, and blind deconvolution problems. Finally, the publication 4 investigated the nature of regularity conditions more abstractly. This work showed that under mild assumptions on the functional components of the problem (semi-algebraicity), regularity conditions that enable rapid convergence of numerical methods hold generically in a precise mathematical sense.

1. D. Drusvyatskiy, A.S. Lewis, 'Error bounds, quadratic growth, and linear convergence of proximal methods'. *Math. Oper. Res.*, 43 (2018), no. 3, 919-948.

2. D. Davis, D. Drusvyatskiy, K.J. MacPhee, C. Paquette, 'Subgradient methods for sharp weakly convex functions'. *J. Optim. Theory. Appl.*, 179 (2018), no. 3, 962-982.

3. D. Davis, D. Drusvyatskiy, C. Paquette, 'The nonsmooth landscape of phase retrieval', To appear in *IMA J. Numer. Anal.*, 2018.

4. D. Drusvyatskiy, A.D. Ioffe, A.S. Lewis, 'Generic minimizing behavior in semialgebraic optimization'. *SIAM J. Optim.* 26 (2016), no. 1, 513-534.

5. D. Drusvyatskiy, A.D. Ioffe, A.S. Lewis, 'Transversality and alternating projections for nonconvex sets'. *Found. Comput. Math.* 15 (2015), no. 6, 1637-1651.

3 Complexity of first-order algorithms

The contemporary need to extract meaningful conclusions from exascale data provides formidable challenges for optimization specialists. In such circumstances, the high per-iteration cost of algorithms using second-order information can render them prohibitively expansive. Instead, first-order methods have dominated much of large-scale optimization as of late. A mature complexity theory, beginning with Nemirovsky-Yudin '82, is an attractive feature of the subject. For specific problem classes, there are known lower complexity bounds, which express limitations on the "efficiency" that any "algorithm" can be guaranteed to have. Methods achieving such best possible convergence rates are called optimal for the class. A large part the PI's work aimed to develop new first-order algorithms, with an emphasis on their complexity guarantees. In particular, the paper 1 developed a new optimal first-order method for smooth strongly convex minimization. In contrast to existing work, the method is completely transparent geometrically and allows to incorporate limited memory for improved performance. The two papers 2-3 developed efficient first-order algorithms for minimizing convex functions over regions with complex geometries, such as those defined by a constraint on the size of the measured residuals. The papers evaluated the performance of the proposed algorithms on statistical and signal processing tasks. The paper 4 established the first complexity guarantees for first-order methods on convex composite optimization problems. Moreover, this work proposed a novel algorithm that automatically accelerates when the problem is nearly convex.

1. D. Drusvyatskiy, M. Fazel, S. Roy, 'An optimal first order method based on optimal quadratic averaging'. *SIAM J. Optim.*, 28 (2018), no. 1, 251-271.

2. A.Y. Aravkin, J.V. Burke, D. Drusvyatskiy, M.P. Friedlander, and S. Roy, 'Level-set methods for convex optimization', To appear in *Math. Program. Ser. B*, 2018.

3. A.Y. Aravkin, J.V. Burke, D. Drusvyatskiy, M.P. Friedlander, K. MacPhee, 'Foundations of gauge and perspective duality'. *SIAM J. Optim.*, 28 (2018), no. 3, 2406-2434.

4. D. Drusvyatskiy, C. Paquette, 'Efficiency of minimizing compositions of convex functions and smooth maps', To appear in *Math. Program.*, 2018.

4 Stochastic algorithms beyond smoothness and convexity

The stochastic subgradient method plays a central role in stochastic optimization and its numerous applications in data science and engineering. Its popularity is in large part due to its simplicity, ease of implementation, and well-documented success in large scale applications. Indeed, the stochastic subgradient method forms a core numerical subroutine for several widely used solvers, including Google's TensorFlow and the open source PyTorch library. Though variants of the method date back to Robbins-Monro's pioneering 1951 work, convergence guarantees in the nonsmooth and nonconvex setting have remained elusive. In the two papers below, the PI with coauthors, developed the first convergence guarantees for the stochastic subgradient, proximal point, and Gauss-Newton methods on nonsmooth and nonconvex problems. Indeed, most of the results are new even in deterministic settings.

1. D. Davis, D. Drusvyatskiy, S. Kakade, J.D. Lee, 'Stochastic subgradient method converges on tame functions', To appear in *Found. Comput. Math.*, 2018.

2. D. Davis, D. Drusvyatskiy, 'Stochastic model-based minimization of weakly convex functions', *SIAM J. Optim.*, 29 (2019), no. 1, 207-239.

5 Spectral variational analysis

Another recent line of work is geared towards eigenvalue optimization problems. Central objects of interest are functions of symmetric matrices that depend on the matrix only through its eigenvalues. A rudimentary example is the nuclear norm of a symmetric matrix, which is simply the ℓ_1-norm of its eigenvalues. Though such spectral functions are typically nonsmooth, the way in which the nonsmoothness arises is highly structured. The general philosophy (called the transfer principle) is that many "variational properties" of the spectral function and that of its restriction to the subspace of diagonal matrices (ℓ_1-norm in the example above) are in one-to-one correspondence. Convexity and smoothness are early examples of such properties.

A central result in this line of research is a formula established by Lewis relating "generalized derivatives" of the spectral function to those of its diagonal restriction. In the paper 1, the PI with his student C. Paquette derived a new, short, and entirely transparent proof of this foundational result, in contrast to the largely opaque original proof.

In a parallel work 2, the PI with coauthors showed that the so-called Euclidean distance degree satisfies the transfer principle. The Euclidean distance degree, introduced by Draisma-Horobet-Ottaviani- Sturmfels-Thomas, is an important invariant measuring the complexity of distance minimization problems to algebraic varieties.

1. D. Drusvyatskiy, C. Paquette, 'Variational analysis of spectral functions simplified'. *J. Conv. Anal.* 25 (2018), No. 1, 119–134.

2. D. Drusvyatskiy, H.-L. Lee, G. Ottaviani, R.R. Thomas, 'The Euclidean distance degree of orthogonally invariant matrix varieties'. *Israel J. Math.* 221 (2017), no. 1, 291-316.

6 Invited conference and seminar presentations

Support from the AFOSR has enabled the PI to travel to the following conferences and colloquia as an invited speaker. The travel resulted in a number of new collaborations that developed during the award period.

1. Oct. 2018: 'Convergence Rates of Stochastic Algorithms in Nonsmooth Nonconvex Optimization', Computational and Applied Mathematics Colloquium, U. Chicago, IL.
2. Aug. 2018: 'Stochastic methods for nonsmooth nonconvex optimization', Variational Analysis and Applications, Erice, Sicily.
3. Jul. 2018: 'Convergence rates of stochastic algorithms for nonsmooth nonconvex problems', Modern Trends in Nonconvex Optimization for Machine Learning, ICML 2018 Workshop, Stockholm, Sweden.
4. Jul. 2018: 'Stochastic methods for nonsmooth nonconvex optimization', Nonconvex Formulations and Algorithms in Data Sciences, U. Wisconsin Madison, WI.
5. Jul. 2018: 'Stochastic subgradient method converges on tame functions', International Symposium on Mathematical Programming (ISMP), Bordeaux, France.
6. Jan. 2018: 'Slope and geometry in variational mathematics', CNA Seminar, Carnegie Mellon University, Pittsburgh, PA.
7. Dec. 2017: 'Algorithms for minimizing compositions of convex functions and smooth maps', CS Theory Seminar, University of Washington, Seattle, WA.
8. Nov. 2017: 'Structure, complexity, and conditioning in nonsmooth optimization', Mathematics colloquium, UCSD, San Diego, CA.
9. Nov. 2017: 'Composite nonlinear models at scale', ORIE colloquium, Cornell, Ithaca.
10. Jul. 2017: 'Efficiency of minimizing compositions of convex functions and smooth maps', Foundations of Computational Mathematics (FoCM), Barcelona, Spain.

11. May 2017: ‘Accelerated first-order methods beyond convexity’, Workshop on nonsmooth optimization and its applications, Hausdorff center for mathematics, Bonn, Germany.

12. May 2017: ‘Taylor-like models in nonsmooth optimization’, SIAM Conference on Optimization, University of British Columbia (UBC), Vancouver, USA.

13. Apr. 2017: ‘Accelerated first-order methods beyond convexity’, Workshop on optimization and statistical learning, Les Houches, France.

14. Apr. 2017: ‘Accelerated first-order methods beyond convexity’, AMS Spring Western Sectional Meeting, Washington State University, Pullman, WA.

15. Jul. 2016: ‘Expanding the reach of optimal methods’, SIAM Annual meeting, Boston, Massachusetts, USA.

16. May 2016: ‘Error bounds, quadratic growth, and linear convergence of proximal methods’, CORS Annual conference (session organizer), Banff, Alberta, Canada.

17. May 2016: ‘Expanding the reach of optimal methods’, West Coast Optimization Meeting (WCOM 2016) (conference organizer), University of Washington, Seattle, WA.

18. Apr. 2016: ‘ Geometry of orthogonally invariant matrix varieties’, Algebra & Discrete Math. seminar, UC Davis, CA, USA.

19. Aug. 2015: ‘Tame variational analysis’, Workshop on Variational Analysis and Applications, Erice, Sicily.

20. Jul. 2015: ‘Slope and variational geometry in optimization’, A.W. Tucker prize session, International Symposium on Mathematical Programming (ISMP 2015), Pittsburgh, USA.

21. Jul. 2015: ‘Singularity degree in semi-definite programming’, International Symposium on Mathematical Programming (ISMP 2015), Pittsburgh, USA.