INFORMS OS Today

The Newsletter of the INFORMS Optimization Society

Volume 4 Number 1 May 2014

Contents

Send your comments and feedback to the Editor:

Shabbir Ahmed
School of Industrial & Systems Engineering
Georgia Tech, Atlanta, GA 30332
sahmed@gatech.edu

Chair's Column

Sanjay Mehrotra
Industrial Engineering and Management Sciences
Department, Northwestern University
mehrotra@iems.northwestern.edu

Optimization is now becoming a household word in communities across engineering, management, health and other areas of scientific research. I recall an advice from my department chairman when I came up for tenure over twenty years ago: 'You need to explain in layman terms the nature of optimization research, and demonstrate its use in practice. The promotion and tenure committee is full of bench scientists, and they don't understand what we do'. Hard work and persistence from our community members have lead in establishing optimization as a "core technology." The goal in front of us is to establish our leadership as drivers of research that changes the life of the common man. This is hard, since it requires simultaneous thinking in deep mathematical terms, as we all like to do, and in broad applied terms. However, I hope that the time to come will bring optimizers and optimization technology to new heights. Let us all work together to make it happen.

The present issue of the INFORMS Optimization Society newsletter, "INFORMS OS Today," features articles by the 2013 OS prize winners: Donald Goldfarb and Alexander Shapiro (Khachiyan Prize for Lifetime Accomplishments in Optimization), Pablo Parrilo (Farkas Prize for Mid-career Researchers), James Luedtke (Prize for Young Researchers), and Afonso Bandeira

(Student Paper Prize). These articles summarize their motivation of working on optimization problems, and the prize winning works. The articles by Don Goldfarb and Alex Shapiro describe their very interesting life journeys that brought them to optimization research that is beautifully intertwined with events in their personal lives. In a similar spirit Pablo's article "Adding squares: from control theory to optimization" takes us through the motivation of his work in polynomial optimization. Jim Luedtke summarizes his very interesting work in the paper "A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support" for which he won the Young Researcher award. Afonso Bandeira summarizes his work on "Sparse recovery in derivative-free optimization" that he completed with Katya Scheinberg and Luís N. Vicente with motivation from compressed sensing.

In this issue we also have announcements of key OS activities: Calls for nominations for the 2014 OS prizes, a call for nominations of candidates for OS officers, and a call for the OS 2016 Conference. The previous five conferences have had a diverse set of themes: "Optimization and HealthCare" (San Antonio, 2006); "Theory, Computation and Emerging Applications" (Atlanta, 2008); "Energy, Sustainability and Climate Change" (Gainesville, 2012); "Optimization and Analytics" (Coral Gables, 2012); and "Optimization and Big Data" (Houston, 2014). The 2014 edition of the INFORMS Optimization Society Conference was held March 6-8 in Houston, Texas, hosted by the Department of Computational and Applied Mathematics at Rice University. Many thanks go to Ilya Hicks who served as the general conference chair, and to the other members of the organizing committee: Matthias Heinkenschloss and Wotao Yin. The conference was attended by more than 125 participants from eight countries. It featured 104 talks in 29 sessions, 4 outstanding plenary talks, amazing food, and – in keeping with OS Conference tradition – at least one confirmed human pyramid. Please consider being active in the nomination process, as well as hosting the 2016 OS conference.

From root to leaves: Ilya Hicks, James Luedtke, Dan Steffy, Gustavo Angulo, Jeff Linderoth, Akshay Gupte

Optimization Society has traditionally had a very strong presence at the annual INFORMS meetings, the next one being in San Francisco, California on November 9-12 at the Hilton and Union Square and Parc 55 Wyndham. Our participation is organized via the OS sponsored clusters, which are organized by our Vice Chairs:

- Imre Polik, Computational Optimization and Software (imre@polik.net)
- Leo Liberti, Global Optimization (leoliberti@gmail.com)
- Juan Pablo Vielma, Integer and Discrete Optimization (jvielma@mit.edu)
- John Mitchell, Linear and Conic Optimization (mitchj@rpi.edu)
- Vladimir Boginski, Network Optimization (boginski@reef.uf.edu)
- Andreas Wächter, Nonlinear Optimization (andreas.waechter@northwestern.edu)
- Andrew Schaefer, Optimization under Uncertainty (schaefer@engr.pitt.edu)

You will note that the area names of several of the vice-chairs have changed. Nearly 225 membership votes were received for the name change ballot that was sent out earlier this year. Integer Programming is now called Integer and Discrete Optimization (92% in favor of change); Linear Programming and Complementarity is now called Linear and Conic Optimization (81% in favor of change); Networks is now called Network Optimization (90% in favor of change); Nonlinear Programming is now called Nonlinear Optimization (85% in favor of change); and Stochastic Programming is now called Optimization under Uncertainty (71% in favor of change).

Our impressive presence within INFORMS is due to the hard work of our vice-chairs, and it reflects the very large membership of the OS. Please contact appropriate Vice Chairs to get involved. I want to remind you that Pietro Belotti continues to be the OS webmaster, and he is always pleased to get your feedback on our website: www.informs.org/Community/Optimization-Society.

I look forward to seeing you at the INFORMS national meeting in November, in particular, at the OS Prize session, and the OS Business Meeting. The latter is always one of the highlights of an INFORMS meeting to have some refreshments, meet with old friends and make new ones.

My Nonlinear (Non-Straight) Optimization Path

Donald Goldfarb
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027
goldfarb@columbia.edu

It is truly a great honor to have been selected as a recipient of the 2013 INFORMS Khachiyan Prize for Lifetime Achievements in Optimization. I would like to take this opportunity to relate how I arrived at this point in my career. It was certainly not a direct path that I followed, but one that has brought me great satisfaction and deep and lasting friendships with other optimizers. However, first I wish to humbly thank the members of the prize committee who selected me, Jorge Nocedal (chair), Michael Todd, Jean-Philippe Vial, and Lawrence Wolsey and to my friends and colleagues, Sanjay Mehrotra, Michael Overton and Katya Scheinberg who nominated me.

I did not start out as a mathematician; in fact all my degrees are in chemical engineering. But from an early age I was fascinated by mathematics. I was admitted to the elite math-oriented Stuyvesant High School in New York City. However, due to my family's move to the outer reaches of Queens which made the daily commute to Stuyvesant prohibitively long, I went to an ordinary high school with little exposure to interesting mathematics. While in high school, I could not imagine how becoming a mathematician could lead to a career. I naively thought that to be a mathematician one had to be like Euclid and invent geometry. So when I applied to college, I chose chemical engineering as it involved science and mathematics, my strongest academic subjects. During my first week as a student at Cornell, I discovered that there was a program in Engineering Physics that required ten semesters of mathematics (engineering was a five-year program in those days). This appealed to me, but when I asked the Dean of Engineering to switch to it from the Chemical Engineering program that I had selected before arriving on campus, he counseled me to not change my selection as there was a common core program during the first two years in engineering, and I could transfer later if I did well. I did well, especially in math, but I had many Chem. E. friends and stayed on as a Chem. E., principally out of inertia. When I applied to graduate school, I again considered going on in math, but since I seemed to be good at chemical engineering and believed that doctoral studies in chemical engineering would be more challenging than my undergraduate studies, again inertia won out.

I arrived at Princeton in the fall of 1963 and chose to work on stochastic optimal control problems in chemical engineering. My initial approach was based on Pontryagin's maximum principle. This involved solving two-point boundary value problems, which required guessing the values of the adjoint (dual) variables at the initial time and then minimizing their deviation from their known values at the final time. Unfortunately, this process was extremely unstable and unless one's guess was very close to the correct initial values, the adjoint system of differential equations blew up. Fortuitously, around this time (fall 1965) I learned about gradient and second-order methods that had been recently proposed by Arthur Bryson and others to solve optimal control problems, and about the Davidon-Fletcher-Powell (DFP) quasi-Newton (aka variable metric) method for unconstrained nonlinear optimization. This method, originally proposed by William Davidon, was shown by Roger Fletcher and Michael Powell to minimize a strictly convex quadratic function of n variables in n steps. I was totally enthralled by this method - it managed to behave like a second-order method without having to compute second derivatives - and I wanted to apply it to my control problem. However, the latter, when discretized, gave a nonlinear program with linear equality (the ordinary differential equations in my models were linear) and inequality (the control variables had lower and upper bounds) constraints. Parenthetically, I do not believe that I was yet aware of the term nonlinear programming (NLP), and I know that I had no knowledge of the simplex method.

At this time, I was living in Manhattan having found the monastic life in Princeton not to my liking and having finished all of my course requirements. I had obtained permission from Peter Lax to use the Math Library at the Courant Institute and given access to their CDC 6600 computer. I naively asked Princeton, who received double tuition for me from my NSF doctoral fellowship, to pay for me to take courses at Courant, but was turned down; I did not know that universities do not like to give money to other universities. However, to satisfy my interest in mathematics, I attended lectures on linear programming given by Ralph Gomory who was visiting Courant (an amazing first exposure to the subject) and I traveled to Princeton one day a week to attend a course on Lyapunov stability theory given by one of the great mathematicians of the twentieth century, Solomon Lefschetz, who at that time was a professor emeritus. It was while I was taking the bus from NYC to Princeton to attend Lefschetz's class that I discovered how to extend the DFP method to my linearly constrained optimization problem. I vividly recall staying up most of the night working out my extension. Once I realized that the finite termination properties of the DFP method worked for my *projected* version of the DFP method, I knew that I had the basis for a thesis. My approach involved rank reducing the DFP inverse Hessian approximation when a bounding constraint became active and rank increasing it when it became necessary to leave a boundary constraint due to Lagrange multipliers having the wrong sign. I immediately dropped my studies on stochastic optimal control and focused only on my new NLP algorithm [1]. In May 1966 I gave my first conference talk on my method at the SIAM National Meeting in Iowa City, Iowa, where I met Michael Powell and John Dennis, beginning a long and gratifying friendship with both.

Early in 1966 I also started to look for a job. I consulted Professor Lefschetz, and he suggested that due to my interests that I apply for a post-doc in applied mathematics. This had never occurred to me, and I am forever grateful to him for his confidence in me and his advice. Moreover, he indicated that he could secure a postdoctoral position for me in the Lefschetz Center for Dynamical Systems in the Applied Mathematics Department at Brown. I did not take him up on that offer, but did ask him to write me a letter of recommendation for a postdoc at Harvard (where Bryson led a small applied mathematics group) and at Courant, since at that time I had met my future wife who was applying to graduate schools in Boston and New York City. I believe that Lefschetz's letter was crucial to my being offered positions at both Harvard and Courant, since I had no background in applied mathematics. When my future wife decided that she would continue her studies in New York, I accepted the Courant offer.

My two years at Courant were transformative. Although, there was no one else there at the time interested in optimization, my perspective was expanded by my interaction with some of the world's greatest applied mathematicians. During my first summer at Courant, I was asked to teach a masters level numerical analysis course. This gave me a strong foundation in numerical linear algebra that has served me well. During my second year at Courant, I was offered an assistant professorship in the Computer Science Department that was being created at the City College of New York (CCNY) and in September 1968, I and a graph theorist, Michael Plummer, became the founding faculty members of this Department. In 1968 there were only a smattering of CS departments in the U.S., or in fact, in the world, and very few people trained in this new discipline. After I moved to CCNY, I maintained contact with Courant, usually spending at least one day there each week. This was necessary for me to get any research done as the teaching load at CCNY was four courses per semester. Since my research was funded by grants, my teaching load was reduced to *ONLY* three courses per semester.

During the fourteen years that I was on the faculty at CCNY, I wrote twenty-three papers, only eight of which were co-authored, since there were few students in my Department and no other faculty at CCNY interested in optimization. My research initially focused on nonlinear optimization algorithms for both unconstrained and linearly constrained problems and the numerical linear algebra (matrix factorizations) [2] involved in implementing these algorithms. The first paper that I wrote after joining CCNY and after completing two papers on my PhD thesis research was the short paper [3], in which I developed the so-called Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. My work was based on a variational approach for deriving quasi-Newton updating formulas that John Greenstadt at IBM had proposed. I should digress that I have spent several summers and sabbatical leaves at the IBM Watson Research Center and one summer (1970) at an offshoot, the IBM New York Scientific Center. These visits played a very important role in my career and I am grateful for the interactions and collaborations that these visits made possible. In particular, I would like to single out Philip Wolfe, Andrew Conn, Michael Grigoriadis, John Forrest and my former student Katya Scheinberg, the latter three with whom I have written major papers.

I spent the 1974-75 academic year on sabbatical leave at the Atomic Energy Research Establishment in Harwell, England. Although I was invited to Harwell by Michael Powell, I spent a good deal of my time collaborating with John Reid. John was working on a code for the simplex method and asked me one day at lunch if I knew anything about a version of it, proposed by Paula Harris, called DEVEX. I did not, but when he described how DEVEX selected pivots, I remarked that it seemed like it was taking approximate steepest-edge steps. John then stated "doesn't the simplex method with the most negative cost pivot rule take steepest edge steps," pointing to the chapter "The Simplex Method Viewed as the Steepest Descent Along Edges" in George Dantzig's classic book Linear Programming and Extensions. I replied that it does, but only if one restricts the space of variables to only those that are nonbasic, and that such steps were not truly steepest because that depended how the active constraints intersected with one another. John did not fully understand my geometric explanation, so I went back to my office and wrote down some algebra to explain what I had told him. In doing so I wrote down recursions for implementing a steepest-edge simplex method and showed them to John. I also realized that I had discovered a way to implement such an algorithm efficiently. Philip Wolfe and L. Cutler had shown a dozen years earlier that the steepest edge rule resulted in far fewer pivots than other pivot rules, but concluded that it was not practicable due to the heavy computational cost it entailed at each iteration. I worked with John on implementing the steepest-edge simplex method resulting in the paper [4]. I also wrote some follow-up papers and later analyzed the method's worst-case behavior in [5]. Sixteen years later, while on another sabbatical leave which I spent at the IBM Watson Research Center, I was again asked about the DEVEX and steepest-edge methods, this time by John Forrest. He was working on improving IBM's OSL simplex code and how one could devise a dual steepest-edge implementation, since on many real-world problems the dual version of the simplex method out-performed the primal version. I provided the recurrence formulas that John needed and he implemented them in his $25,000$ line code in less than an hour and ran the new dual simplex variant on a difficult fleet scheduling problem that the standard primal and dual pivot rules took $3\frac{1}{4}$ and 54 hours, respectively, to solve. A short time later, John appeared at my office door and announced that my dual pivot rule *worked*. In fact it solved his problem in roughly a half hour. On a larger and more difficult fleet scheduling problem the dual steepest-edge rule took just 52 minutes to obtain an optimal solution while the standard dual pivot rule required 123 hours to do so. Needless to say, primal and dual steepest-edge pivot rules were quickly incorporated into the OSL code and other codes such as CPLEX . Our paper [6] on steepest-edge rules was awarded the INFORMS 1995 Prize for Excellence in the Interface between Operations Research and Computer Science.

I spent the 1979-80 academic year, as a visiting professor in both the Department of Computer Science and the School of Operations Research and Industrial Engineering (SORIE) at Cornell. This was an exciting time for the field of optimization. Soon after I arrived in Ithaca the news that Leonid Khachiyan had proved that linear programming problems could be solved in polynomial time using the ellipsoid method had reached Europe and the U.S. A major article about this appeared on November 11, 1979 in the New York Times. Interest at SORIE in the ellipsoid method was intense. A special weekly seminar on this topic was scheduled and Michael Todd and I began to work on trying to develop a practical version of the method. Unfortunately, the speed of the method in practice closely adhered to its theoretical speed even if deep and surrogate cuts were employed and hence it proved not to be a practical algorithm for linear programming [7]. Also during this time Michael Todd and Robert Bland and I were asked by Thomas Magnanti, the editor-in-Chief of Operations Research, to write a featured article for the journal presenting a survey on the ellipsoid method [8].

After returning to CCNY, I worked with Ashok Idnani, my first and next-to-last doctoral student at CCNY on dual and primal-dual methods for quadratic programming . Our dual method is heavily used and has been implemented in numerous codes [9]. At this time, I started to receive several inquiries about moving to another university. When I was contacted by the Industrial Engineering and Operations Research Department (IEOR) at Columbia, my interest was aroused, since joining Columbia would not require leaving New York City. So in July 1982, I joined the Columbia IEOR Department as a full professor. Shortly thereafter, it was again another very exciting time for the field of optimization. On November 19, 1984 an article with the heading "Breakthrough in Problem Solving" appeared on the front page of the New York Times announcing the development by Narendra Karmarkar of AT&T Bell Laboratories of an interior-point method for solving large-scale linear programs. Even before this, I had received a call from Karmarkar, telling me about his method, and a draft of his forthcoming paper. At the time I hosted a seminar series at Columbia called *Friends of Optimization* (FO) that had been started by Philip Wolfe. I was quite excited about Karmarkar's algorithm; although it had roughly the same complexity as the ellipsoid method, it appeared to be a very new approach to solving linear programs. Consequently, I invited Karmarkar to present his work to a specially called FO meeting. This was the very first talk that Karmarkar gave on his new method; it even preceded his first talk at Bell Labs on it. Subsequently, Karmarkar called me and asked me to work with him on developing a code for his method. Having valiantly tried with Michael Todd to make the ellipsoid method into a practical method, I turned down Karmarkar's request, believing that his algorithm, like the ellipsoid method, would prove to be good in theory but not in practice. Was I WRONG! However, once I realized the power of interior-point methods, I began working with my students on them. From the mid-1980's to the mid-1990's, working with my students Sanjay Mehrotra, Shucheng Liu, Siyun Wang, In-Chan Choi and Dong Xiao we developed relaxed and self-correcting variants of Karmarkar's method, and interior-point methods for convex quadratic programming (QP), quadratically constrained convex QP, multicommodity and other specially structured problems, and projective variants; e.g., see [10, 11]. Also during this period Michael Todd and I wrote a chapter on Linear Programming for the Optimization Volume of the Handbook in Operations Research and Management Science [12] based on the PhD-level course that I teach, and I worked with Michael Grigoriadis and Jim Orlin, my students Jianxiu Hao, Sheng-Roan Kai, Wei Chen and Yiqing Lin, and a postdoc Zhiying Jin on simplex and combinatorial algorithms for pure (e.g., [13, 14, 15]) and generalized network flow (e.g., [16]) problems that have respectively, strongly polynomial and polynomial complexity bounds .

In the early 1990's, interest in conic programming, and in particular interior-point methods for semidefinite (SDP) and second-order cone programming (SOCP) burgeoned after the publication of the seminal papers on this subject by Farid Alizadeh and Yuri Nesterov and Arkadii Nemirovskii. It was very natural to move from working on interior-point algorithms for linear and quadratic programming to the study of such methods for SOCP and SDP and related topics. Fortunately for me, Katya Scheinberg who had already worked with Nemrovskii in Russia was admitted to our doctoral program in 1992 and chose to work with me on SDP [17]. After she graduated, we continued to work in that area combining it with my long-held interest in factorization methods (see [18]). Also, Alizadeh spent the 2000 calendar year as a Visiting Associate Professor in our Department, which resulted in our collaborating on our highly cited SOCP survey paper [19]. Around this time (1998 to be exact), two other events occurred which led me into a new, but related, area of research: Garud Iyengar joined our Department as Assistant Professor and Aharon Ben-Tal and Arkadii Nemirovskii introduced the paradigm of robust optimization. Seeking robust solutions to variants of the well-studied Markowitz portfolio optimization problem seemed like a natural topic to pursue, and Garud and I developed an approach that combined SDP, SOCP, statistics and some nice linear algebra [20]. We followed this with other papers on robust quadratically constrained quadratic programming and robust active portfolio management.

In May of 2003, during a trip to Los Angeles to visit my daughter, I was invited by Stanley Osher (a friend from my days at Courant) to give a talk in the Mathematics Department at UCLA. As I had a strong interest at the time in SOCP, I spoke on that topic. Stanley, is one of the leading applied mathematicians in the world, and was widely known at that time for his work on partial differential equations, level-set methods and imaging. Clearly, he and I worked in very different domains. However, after my talk, Stanley told me that the imaging problems on which he worked involved minimizing the total variation (TV) of an image, but that he knew very little about optimization. I realized immediately that his problems could be formulated as SOCPs and as soon as I returned to New York, I started to work on them with Stanley, Martin Burger, who was visiting Stanley, and my student, Wotao Yin. This led to our Bregman iterative approach (which is equivalent to the augmented Lagrangian method) for solving image denoising and deblurring problems and to my currently second most cited paper [21] as well as many others on related topics (e.g., see [22, 23].

In 2005, intense interest in compressed sensing (CS) and the possibility of obtaining solutions to hard combinatorial optimization problems by solving tight convex relaxations of them was generated by the work of Emmanuel Candes, Terence Tao and David Donoho and their co-workers. This work showed that one could find the sparse solution to a linear system of equations by solving an l_1 minimization problem, and it ushered in an exciting new focus on optimization methods for problems whose solutions were sparse or had some other special structure. Since l_1 and TV minimization are very closely related, I was quickly drawn into developing algorithms for solving CS problems and their matrix analogs, rank minimization problems, like those that arise in the famous NETFLIX problem. My research on l_1 minimization was done in collaboration with Wotao Yin and several of my other TV-minimization collaborators, Yin Zhang and a new student

Zaiwen Wen [24, 25], while my research on rank minimization was done with two other new students Shiqian Ma and Lifeng Chen [26, 27]. Lifeng's main interest, however, was primarily in interior-point penalty function algorithms [28], which brought me back to topics that I had considered earlier in my career. Since, many of the extremely large-scale problems in machine learning (ML) are sparse or rank minimization problems, my work on algorithms for their convex relations, l_1 and nuclear norm minimization, led me quite naturally to the study of optimization methods for ML. These problems are so huge, having millions to hundreds of millions of variables, that first-order, FISTA, conditional gradient, alternating direction and linearization type methods become appropriate. So during approximately the last five years this has been the principal focus of my research which has involved my former students Katya Scheinberg, Wotao Yin, Shiqian Ma, Zaiwen Wen and Bo Huang (see [29]-[34]). Currently, I am working with my former and current students Bo Huang and Cun Mu, Zhiwei Qin, and a new colleague John Wright in the EE Department at Columbia on even larger tensor completion and robust and stable PCA problems. From my point of view, the field of continuous optimization has never been more exciting.

Some Final Remarks.

From the above description of the path that my research has taken, it may appear that I switched from one topic to another based on what was hot at the moment. This was not the case. I certainly prefer to work on optimization problems that are important and these tend to be in areas that receive a lot of attention, but in every case, my current or past research positioned me to venture into a new domain. I have often attributed this to luck, but I have come to realize that when one works on important topics, even if those topics become less relevant, research on them often leads in a logical way to new important topics. I would also like to stress how important it has been to me to have had fantastic doctoral students with whom to work. My students have often motivated me to get involved in new areas and I believe that I have learned more from them than they may have learned from me. They deserve much of the credit for the contributions to the field of continuous optimization for which I am being honored. While I am extremely proud of these contributions, I am also very proud for my service to the field as editor-in-chief, senior editor and associate editor of Mathematical Programming, editor of the SIAM Journal on Optimization and the SIAM Journal on Numerical Analysis, and associate editor of Operations Research and Mathematics of Computation. Finally, I also take pride in having served as Chairman of my Department for eighteen years and as Interim Dean of Engineering on two different occasions at Columbia. I have had and continue to have an exceptionally rewarding career in academia and am extremely grateful to those who helped make it so.

REFERENCES

[1] D. Goldfarb, Extension of Davidon's variable metric method to maximization under linear inequality and equality constraints, *SIAM Journal on Applied Mathematics*, 17 (4), 739-764, 1969.

[2] D. Goldfarb, Factorized variable metric methods for unconstrained optimization, *Mathematics of Computation* 30 (136), 796-811, 1976.

[3] D. Goldfarb, A family of variable-metric methods derived by variational means, *Mathematics of Computation*, 24 (109), 23-26, 1970.

[4] D. Goldfarb and J.K. Reid, A practicable steepest-edge simplex algorithm, *Mathematical Programming* 12 (1), 361-371, 1977.

[5] D. Goldfarb and W.Y. Sit, Worst case behavior of the steepest edge simplex method, *Discrete Applied Mathematics* 1 (4), 277-285, 1979.

[6] J.J. Forrest and D. Goldfarb, Steepest-edge simplex algorithms for linear programming, *Mathematical programming* 57 (1-3), 341-374, 1992.

[7] D. Goldfarb and M.J. Todd, Modifications and implementation of the ellipsoid algorithm for linear programming, *Mathematical Programming* 23 (1), 1-19, 1982.

[8] R.G. Bland, D. Goldfarb and M.J. Todd, The ellipsoid method: A survey, *Operations Research* 29 (6), 1039-1091, 1981.

[9] D. Goldfarb and A. Idnani, A numerically stable dual method for solving strictly convex quadratic programs, *Mathematical programming* 27 (1), 1-33, 1983.

[10] D. Goldfarb and S. Mehrotra, A relaxed version of Karmarkar's method," *Mathematical Programming* 40 (1-3), 289-315, 1988.

[11] D. Goldfarb and S. Liu, "An O (n 3 L) primal interior point algorithm for convex quadratic programming," *Mathematical Programming*, 49 (1-3), 325-340, 1990.

[12] D. Goldfarb and M.J. Todd, "Chapter II Linear programming, *Handbooks in Operations Research and Management Science* 1, 73-170, 1989.

[13] D. Goldfarb and J. Hao, A primal simplex algorithm that solves the maximum flow problem in at most nm pivots and $O(n^2m)$ time, *Mathematical Programming* 47 (1-3), 353-365, 1990.

[14] D. Goldfarb and M.D. Grigoriadis, A computational comparison of the Dinic and network simplex methods for maximum flow, *Annals of Operations Research* 13 (1), 81-123, 1988.

[15] D. Goldfarb, Efficient dual simplex algorithms for the assignment problem, *Mathematical Programming* 33 (2), 187-203, 1985.

[16] D. Goldfarb, Z. Jin and J.B. Orlin, Polynomial-time highest-gain augmenting path algorithms for the generalized circulation problem, *Mathematics of Operations Research* 22 (4), 793-802, 1997.

[17] D. Goldfarb and K. Scheinberg, Interior point trajectories in semidefinite programming, *SIAM Journal on Optimization* 8 (4), 871-886, 1988.

[18] D. Goldfarb and K. Scheinberg A product-form Cholesky factorization method for handling dense columns in interior point methods for linear programming, *Mathematical Programming* 99 (1), 1-34, 2004.

[19] F. Alizadeh and D. Goldfarb, Second-order cone programming, *Mathematical programming*, 95 (1), 3-51, 2003.

[20] D. Goldfarb and G. Iyengar, Robust portfolio selection problems, *Mathematics of Operations Research* 28 (1), 1-38, 2003.

[21] S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation-based image restoration, *Modeling and Simulation* 4 (2), 460-489, 2005.

[22] D. Goldfarb and W. Yin, Second-order cone programming methods for total variation-based image restoration, *SIAM Journal on Scientific Computing* 27 (2), 622-645, 2005.

[23] D. Goldfarb, W. Yin, Parametric maximum flow algorithms for fast total variation minimization, *SIAM Journal on Scientific Computing* 31 (5), 3712-3743, 2009.

[24] W. Yin, S. Osher, D. Goldfarb and J. Darbon, Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing, *SIAM Journal on Imaging Sciences*, 1 (1), 143-168, 2008.

[25] Z. Wen, W. Yin, D. Goldfarb and Y. Zhang, A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization, and continuation, *SIAM Journal on Scientific Computing* 32 (4), 1832-1857, 2010.

[26] S. Ma, D. Goldfarb, L. Chen, Fixed point and Bregman iterative methods for matrix rank minimization, *Mathematical Programming* 128 (1-2), 321-353, 2011.

[27] D. Goldfarb and S. Ma, Convergence of fixed point continuation algorithms for matrix rank minimization, *Foundations of Computational Mathematics* 11, 183-210, 2011.

[28] L. Chen and D. Goldfarb, Interior-point l_2-penalty methods for nonlinear programming with strong global convergence properties, *Mathematical Programming* 108, 1-36, 2006.

[29] Z. Wen, D. Goldfarb and W. Yin, Alternating direction augmented Lagrangian methods for semidefinite programming, *Mathematical Programming Computation* 2 (3-4), 203-230, 2010.

[30] K. Scheinberg, S. Ma and D. Goldfarb, Sparse Inverse Covariance Selection via Alternating Linearization Methods, *NIPS*, 2101-2109, 2010.

[31] D. Goldfarb, S. Ma and K. Scheinberg, Fast alternating linearization methods for minimizing the sum of two convex functions, *Mathematical Programming* 141, 1-34, 2012.

[32] D. Goldfarb and S .Ma, Fast multiple-splitting algorithms for convex optimization, *SIAM Journal on Optimization* 22 (2), 533-556, 2012.

[33] Z. Wen, D. Goldfarb and W. Yin, Alternating direction augmented Lagrangian methods for semidefinite programming, *Mathematical Programming Computation* 2, 203-230, 2010.

[34] B. Huang, S. Ma and D. Goldfarb, Accelerated linearized Bregman methods, *J. of Scientific Computing*, 54 (2-3), 428-453, 2013.

Looking Back at My life

Alexander Shapiro
School of Industrial and Systems Engineering
Georgia Institute of Technology, Atlanta, Georgia
30332-0205, USA.
ashapiro@isye.gatech.edu

1. Early years

This prize came as a big surprise for me. Since the prize is given for life-time accomplishments, I suppose I should write about my life. This is a daunting task. Anyway I will try.

I was born in Moscow the capital of Soviet Union (USSR) - the country that does not exist anymore. I started my school years at age 7 as did millions of other children in Soviet Union. My first years in school were unremarkable. I certainly was not the best pupil in my class. At that time the educational system in Soviet Union was very uniform, and looking back I realize how unique it was.

In sixth grade we started to learn algebra and plane (Euclidean) geometry. Every pupil in class had to learn how to derive and prove some simple geometrical constructions from basic axioms in a reasonably rigourous way. This is when I became somewhat different from the other children in my class. The study pace was very slow and I often was bored, so I was reading the textbook ahead of what we were supposed to know. At the same time I started to read books trying to solve mathematical puzzles.

OS Chair Sanjay Mehrotra, with Khachiyan Prize winners Alexander Shapiro and Don Goldfarb

In seventh grade I discovered the existence of mathematical workshops for children. These workshops were supervised by students from Moscow University, and were taught on a voluntary basis. Just coming to the old university building in the center of Moscow, to listen to and discuss new ideas was electrifying. Some of the children there were considerably more advanced and in the beginning I was quite intimidated. That summer I read parts of the excellent book by Rademacher and Toeplitz [5] (in Russian Translation). The idea that there are many different levels of infinity was amazing. Cantor's famous proof that the set of real numbers is uncountable is rather elementary and could be understood with a high school mathematical background.

In eighth grade I started to attend an evening school organized by Evgeny Borisovich Dynkin. Apart from being a great mathematician, Evgeny Borisovich was also an outstanding teacher and a most efficient organizer. We were listening to lectures given by first class mathematicians and took part in competitions for solving mathematical problems. One evening A.N. Kolmogorov gave a talk, I was 14 and didn't understand a thing he was talking about.

Next year, when I was supposed to start grade 9, Evgeny Borisovich started what became the famous School Number 2 for gifted children. About a hundred students were admitted and organized into 3 groups. The two years in this school were the most formative in my life. Evgeny Borisovich brought his students to teach us. Some of them became famous in their own right. I was taught how to think by great people. Our group was supervised by A.D. Wentzell. He had a somewhat unusual sense of humor. One day two students approached him arguing about the grades they received. Even though both of them solved the first 3 out of 4 questions the same way, one got a B and another C. His reply to the C student was: "But in the fourth question you showed negative knowledge". I wish I could say this to some of my students.

After graduating from high school I was admitted to the Mathematical Department of the Moscow University, the famous *Mech-Mat*. I was lucky, just a few years later Jews would not be accepted to Moscow University and this antisemitic policy continued until the late eighties. Mech-Mat was a unique place. Famous professors were giving classes, Kolmogorov was not teaching regular classes at that time, but I could see him sometimes walking along the corridors.

2. In Israel

By the end of my studies at Moscow University I already knew that I wanted to immigrate to Israel. In 1972 my wife and I left the "paradise" of the Soviet Union on our way to Israel. We were young, idealistic, naive and unprepared for the harsh reality. I started to study towards a Ph.D. at Hebrew University in Jerusalem. It didn't go well. To support my family I worked as a substitute teacher in high school. Then I got a tenured position in the Israeli Ministry of Communication. The decision to leave a secure government job, after about one year, to start a Ph.D. at Ben-Gurion University of the Negev in Beer Sheva was not an easy decision.

My new advisor was Giacomo Della Riccia. For my Ph.D. I was supposed to work on an applied project which involved minimization of the trace of a covariance matrix subject to keeping the reduced matrix positive semidefinite, the so-called Minimum Trace Factor Analysis. In today's terminology this can be classified as a semidefinite programming problem. Eventually this required development of an optimization algorithm and a related statistical inference. At that time I knew very little about Mathematical Programming and Statistics.

This optimization problem can be reformulated as a problem of minimization of the largest eigenvalue of a symmetric matrix. Since the corresponding objective function is convex and hence differentiable almost everywhere, I naively thought that I could apply a steepest decent type optimization algorithm. To my surprise it became clear that the algorithm didn't converge to the optimum, no matter how many iterations I used. This motivated me to start a systematic study of optimization techniques and convex and nonsmooth analysis.

Another question which I faced was to develop statistical inference of the optimal value and optimal solutions of that problem. For this I had to understand how small (random) perturbations of the sample covariance matrix effect changes of the corresponding optimal value and optimal solutions. This was the beginning of my interest in sensitivity analysis of optimization problems. Eventually, after many years and several countries, these studies were summarized in our book [2].

I defended my Ph.D. thesis in 1980. Several publications resulted from my thesis, from which I would like to point to [9]. It was shown there that for almost every covariance matrix its rank cannot be reduced below a certain bound by changing its diagonal elements. At that time the motivation was to show that for the sample covariance matrix, with probability one its rank cannot be significantly reduced by changing its diagonal elements.

At Ben-Gurion University I met Yosef Yomdin. He was a fresh Ph.D. immigrant from the USSR. We worked together on the theory of what is now known as the Difference Convex (DC) optimization. Unfortunately a couple of reports which we wrote were not published. One of interesting questions of that theory is to give an intrinsic characterization of functions $f : \mathbb{R}^n \to \mathbb{R}$ representable as a difference $f = g_1 - g_2$ of two convex functions $g_1, g_2 : \mathbb{R}^n \to \mathbb{R}$. For $n = 1$ there is a simple answer for this question. Convex functions $g_1, g_2 : \mathbb{R} \to \mathbb{R}$ are differentiable almost everywhere and their derivatives are monotonically nondecreasing. Therefore $f : \mathbb{R} \to \mathbb{R}$ is representable as a difference of two convex functions iff it is differentiable almost everywhere and its derivative has bounded variation. As far as

I know such an intrinsic characterization is not known for $n > 1$.

3. In South Africa

After receiving my Ph.D. diploma in 1980 I started to look for a (postdoc) position. At that time the USA was in a recession, I was unknown and did not have any connections. After one year and hundreds of applications I received a positive answer only from South Africa. I signed a contract for two years with the University of South Africa (UNISA) and eventually stayed in Pretoria for ten years (with the exception of one year, 1988-1989, which I spent on sabbatical in Technion - Israel Institute of Technology in Haifa).

At UNISA I started my lifelong friendship with an outstanding statistician, M.W. Browne. We worked together on statistical inference of covariance structures. From a technical point of view this can be considered as a branch of multivariate statistical analysis. On a practical side covariance structural models are very popular in Psychometric and Econometric applications. From that period I would single out three papers [3],[10] and [11].

In [3] we showed that large samples statistical inference, of a certain class of covariance structural models, based on the assumption of multivariate normal distribution of the population, holds for a much larger class of distributions. The result was surprising and unexpected. This started a new direction of research known as asymptotic robustness. In [11] we were first to suggest a *nonparametric* approach to estimation of variograms. Eventually it became known as Shapiro - Botha nonparametric variogram fitting method.

While at the Technion, Israel, I became acquainted with the late Reuven Rubinstein who introduced me to the art and science of Monte Carlo simulation. Our cooperation resulted in my first book [6].

During those years I also worked on nonsmooth and sensitivity analysis of optimization problems. UNISA had a very good library, but otherwise I worked quite in isolation. This was before the electronic age. I remember every morning checking my mail box hoping to receive news about submitted papers.

4. At Georgia Tech

In 1991 we immigrated to the United States where I got a position at Georgia Institute of Technology. The transition was not easy. I had to adjust to the American system of teaching, writing proposals and self promotion. Even today after 22 years I am not sure I adjusted well.

The agreement with Georgia Tech was that I would teach Statistics classes. And indeed I taught almost every graduate and undergraduate Statistics class. On the research side, I mainly worked on optimization theory and applications. I wrote a book with Frederic Bonnans on perturbation analysis of optimization problems [2]. It took about 4 years to write this book which was published in 2000. The book was translated into Chinese. It contains some original ideas on second order analysis of possibly nonsmooth (nondifferentiable) optimization problems. In particular, this theory could be applied to semi-definite and semi-infinite programming.

I also continued to work on stochastic programming. The idea of using Monte Carlo sampling techniques to approximate stochastic programming problems is not new of course. A statistical inference of what is now known as the Sample Average Approximation (SAA) method was already outlined in our book [6]. By the way the term *Sample Average Approximation* (SAA) was coined in our paper [4] in order to distinguish it from the *Stochastic Approximation* (SA) method. A natural question is how large should be the sample size in order to solve the "true" stochastic program with a given precision $\varepsilon > 0$ (cf., [4]). This can be viewed as a question of complexity of solving stochastic programs by randomization techniques. It turns out that from this point of view there is a principle difference between two and multi-stage stochastic programming problems. While a large class of two stage stochastic programs could be solved to a reasonable accuracy, generic multi-stage stochastic programs seem to be computationally intractable. This is discussed in details in publications [13] and [14]. This does not mean, of course, that some specific classes of multi-stage stochastic programs cannot be solved, say by approximate dynamic programming methods.

Together with Andrzej Ruszczynski I worked on theory and applications of risk measures. This topic of research started with the seminal paper [1]. The main findings of our contribution were published in [7] and [8], and later summarized in our monograph [15]. Eventually this theory was applied in a project for developing risk averse methodology for Brazilian operation planning of hydro plants (cf., [16]).

Slightly paraphrasing Thomas Edison: "Research is one percent inspiration, ninety-nine percent perspiration". But this "one percent" is important. In that area of research to formulate a problem which on one hand is interesting and important and on the other hand is doable, often is far more important than its technical solution. To find such a problem is not easy and it does not happen too often. After many years of constantly looking for new ideas I am still not sure that I was fortunate enough to initiate such a problem.

REFERENCES

[1] Artzner, P., Delbaen, F., Eber J.-M. and Heath D., Coherent measures of risk, *Mathematical Finance*, 9, 203-228, 1999.

[2] Bonnans, J.F. and Shapiro, A., *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

[3] Browne, M.W. and Shapiro, A., Robustness of normal theory methods in the analysis of linear latent variate models, *British Journal of Mathematical and Statistical Psychology*, 41, 193-208, 1988.

[4] Kleywegt, A.J., Shapiro, A. and Homem-de-Mello, T., The sample average approximation method for stochastic discrete optimization, *SIAM J. Optimization*, 12, 479-502, 2001.

[5] Rademacher, H. and Toeplitz, O. *The Enjoyment of Math*, Princeton University Press, Princeton, NJ, 1957.

[6] Rubinstein, R.Y. and Shapiro, A., *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, John Wiley and Sons, New York, 1993.

[7] Ruszczynski, A. and Shapiro, A., Optimization of convex risk functions, *Mathematics of Operations Research*, 31, 433-452, 2006.

[8] Ruszczynski, A. and Shapiro, A., Conditional risk mappings, *Mathematics of Operations Research*, 31, 544-561, 2006.

[9] Shapiro, A., Rank-reducibility of a symmetric matrix and sampling theory of Minimum Trace Factor Analysis, *Psychometrika*, 47, 187-199, 1982.

[10] Shapiro, A., Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints, *Biometrika*, 72, 133-144, 1985.

[11] Shapiro, A. and Botha, J.D., Variogram fitting with a general class of conditionally nonnegative definite functions", *Computational Statistics & Data Analysis*, 11, 87-96, 1991.

[12] Shapiro, A. and Homem-de-Mello, T., A simulation-based approach to stochastic programming with recourse, *Mathematical Programming*, 81, 301-325, 1998.

[13] Shapiro, A. and Nemirovski, A., On complexity of stochastic programming problems, in: *Continuous Optimization: Current Trends and Applications*, pp. 111-144, V. Jeyakumar and A.M. Rubinov (Eds.), Springer, 2005.

[14] Shapiro, A, On complexity of multistage stochastic programs, *Operations Research Letters*, 34, 1 - 8, 2006.

[15] Shapiro, A., Dentcheva, D. and Ruszczynski, A., *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.

[16] A. Shapiro, W. Tekaya, J. Paulo da Costa, M. Pereira Soares, Risk neutral and risk averse Stochastic Dual Dynamic Programming method, *European Journal of Operations Research*, 224, 375-391, 2013.

Adding Squares: From Control Theory to Optimization

Pablo A. Parrilo
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
parrilo@mit.edu

First of all, I thank the Farkas Prize committee: Dimitris Bertsimas (chair), George Nemhauser, Yurii Nesterov, and Yinyu Ye, for this deeply appreciated and unexpected recognition. It is a great honor, particularly given the broad influence of their work across the optimization field and in my own research.

It is a pleasure to be able to describe my work in optimization over the last few years. At the same time, it is a good opportunity to outline my somewhat indirect (or "non-central") path towards optimization, and to explain why algebraic considerations became of increasing importance in my work.

Beginnings in control My undergraduate education was in Electrical Engineering, in my native Buenos Aires, Argentina. Shortly afterwards, I began my PhD in Control and Dynamical Systems at the California Institute of Technology (Caltech), working on systems and control theory. Control theory is a very interdisciplinary area, where several branches of mathematics (e.g., dynamical systems, linear algebra, functional analysis, optimization, stochastics, etc.) provide distinct but complementary viewpoints. In the early 90's a strong unifying trend towards convex optimization (and particularly, the nascent area of semidefinite programming) became dominant, a viewpoint nicely crystallized in the book [6].


Sanjay Mehrotra, Pablo Parrilo and Dimitris Bertsimas

During my PhD I was generally interested in *robustness analysis*, and in particular, in extending well-understood techniques used in the linear case (the "structured singular value" [27]) to general nonlinear systems. The main mathematical challenge in this area is to understand how the solution set of a system of equations can change as a function of some unknown parameters. It quickly became clear that the case of uncertain polynomial equations required a solid understanding of the language and tools of real algebraic geometry, and I slowly began learning about this area.

Pieces of the puzzle, and further connections with optimization, were provided by the works of Shor and Reznick (particularly, [34] and the beautiful survey [31]). The unifying notion was *sum of squares decompositions*, which require a multivariate polynomial p to admit a representation

$$p(x) = \sum_i q_i^2(x),$$

where the q_i are also polynomials. This condition was shown to be tractable via semidefinite programming [34, 21, 23, 24]. The unifying role of sum of squares decompositions of polynomials as witnesses of nonnegativity and set emptiness, and their effective computation via semidefinite programming, proved to be incredible fertile notions; several of the key ideas are explained below.

Farkas, Positivstellensatz, and hierarchies of SDP relaxations The enormous power of this algebraic machinery quickly became evident, along with the realization that many earlier results of the field (e.g., the S-lemma) could be understood as particular cases of Positivstellensatz constructions.

To explain this, a good starting point is (quite appropriately!) the celebrated "lemma of the alternative" of Farkas. For *linear* inequalities, LP duality provides the following well-known characterization:

Theorem 1 (Farkas lemma)

$$\{x \in \mathbb{R}^n : Ax \geq b\} \ \textit{is empty}$$
$$\Updownarrow$$
$$\exists\, \lambda \geq 0 \ : A^T\lambda = 0, \quad b^T\lambda = 1.$$

Equivalently, the emptiness of a set defined by linear inequalities can be certified by any feasible solution of an auxiliary (dual) LP. A key insight is to realize that the second expression can be simply interpreted as a *polynomial identity* of the form

$$\lambda^T(Ax - b) \equiv -1.$$

This viewpoint is the one that generalizes (regardless of convexity!), to yield infeasibility certificates for *arbitrary* systems of polynomial equations and inequalities over the reals. The corresponding result, known as the *Positivstellensatz*, is a centerpiece of real algebraic geometry (e.g., [5]):

Theorem 2 (Positivstellensatz)

$$\left\{ x \in \mathbb{R}^n : \begin{array}{ll} f_i(x) = 0, & i = 1, \ldots, m \\ g_i(x) \geq 0, & i = 1, \ldots, p \end{array} \right\} \ \textit{is empty}$$
$$\Updownarrow$$
$$\exists\ F(x), G(x)\ \textit{s.t.} \left\{ \begin{array}{r} F(x) + G(x) \equiv -1 \\ F(x) \in \mathbf{ideal}(f_1, \ldots, f_m) \\ G(x) \in \mathbf{preorder}(g_1, \ldots, g_p). \end{array} \right.$$

Here $\mathbf{ideal}(f_1, \ldots, f_m)$ denotes the set of polynomials of the form $\sum_{i=1}^m q_i f_i$, where $q_i \in \mathbb{R}[x]$, and $\mathbf{preorder}(g_1, \ldots, g_p)$ is the *preorder* generated by the inequalities g_i, i.e., the set of polynomials of the form $\sum_\alpha s_\alpha h_\alpha$, where the $s_\alpha \in \mathbb{R}[x]$ are sum of squares and the h_α are squarefree products of the g_i.

The Positivstellensatz states that for every infeasible system of polynomial equations and inequalities, there is a simple algebraic identity that directly certifies the nonexistence of real solutions. Indeed, by construction, the evaluation of the polynomial $F(x) + G(x)$ at any feasible point yields a nonnegative number. However, since this expression is identically equal to the polynomial -1, we arrive at a contradiction. Remarkably, the Positivstellensatz holds under *no assumptions* whatsoever on the polynomials.

This characterization is extremely useful from the optimization viewpoint. Since it only involves linear equations and sum of squares constraints, it can easily be checked with semidefinite programming. This allows the formulation of natural hierarchies of semidefinite relaxations (for either feasibility or optimization) by constraining the degree of the possible certificates [24, 25, 30]. These hierarchies provide, in a fully algorithmic way, a general mechanism to approximate arbitrary semialgebraic problems.

Sums of squares and Lyapunov functions

One of the central problems in control theory is *asymptotic stability* of a dynamical system. Given a system of ordinary differential equations $dx(t)/dt = f(x(t))$, we are interested on whether the solutions $x(t)$ are "long-term stable", i.e., they satisfy $\lim_{t\to\infty} x(t) = 0$, for all initial conditions $x(0)$. The classical "second method" of Lyapunov provides a nice characterization in terms of *Lyapunov functions*, a generalization of the notion of "energy" or "potential function." These are functions $V : \mathbb{R}^n \to \mathbb{R}$ that satisfy

$$V(x) > 0, \qquad \frac{d}{dt}V(x(t)) = \left(\frac{dV}{dx}\right)^T f(x) < 0$$

for all $x \in \mathbb{R}^n/\{0\}$; intuitively, "energy" is nonnegative, and always decreases as the system evolves. Although this characterization is very appealing, it can be difficult to use since finding a suitable function V is often non-obvious.

From the optimization viewpoint, after parametrizing a given class of candidate functions V (say, polynomials) this looks like a simple feasibility problem for the unknown V. The expressions are clearly linear in the unknown function V. However, how should one deal with the positivity constraints? The solution, of course, was to impose a *sum of squares* condition, requiring both the function V and its derivative $-\frac{d}{dt}V$ to have SOS decompositions. This procedure, and other suitable modifications, enabled a significant extension of many results in control theory from the linear to the nonlinear case.

The techniques for parsing/solving these optimization problems, which we called *sum of squares programs*, where later implemented in the software package SOSTOOLS, written in collaboration with Antonis Papachristodoulou and Stephen Prajna [29].

Quantum interlude Long casual conversations at the Red Door (Caltech's coffeehouse and the source of many scientific breakthroughs over the years) with my "quantum" friends Andrew Doherty and Fecho Spedalieri quickly led us to the realization that sum of squares techniques could be extremely useful in the characterization of quantum entanglement.

Indeed, there is a very natural and fruitful isomorphism between three natural objects:

- "Entanglement witnesses:" observables that certify that a given quantum state (described by a density matrix ρ) cannot be explained purely in terms of classical probability,
- Matrix positive maps: linear maps $\Lambda : \mathcal{S}^n \to \mathcal{S}^m$, for which $X \succeq 0$ implies $\Lambda(X) \succeq 0$,
- Nonnegative biquadratic polynomials $p(\mathbf{x}, \mathbf{y})$: these are quadratic in $\mathbf{x}$ for any fixed $\mathbf{y}$, and vice versa.

In particular, the isomorphism between the last two is given by the relation $p(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \Lambda(\mathbf{y}\mathbf{y}^T)\mathbf{x}$. Since the last condition involved nonnegative polynomials, we started thinking that perhaps SOS methods could say something interesting about this...

Luckily for us, this was the case, and the outcome was the series of papers [12, 13, 14] where a complete hierarchy of quantum separability tests was developed. This work quickly resonated with the quantum information community. In a breakthrough paper in 2010, Brandão, Christandl and Yard [3] have shown that this hierarchy in fact provides a quasi-polynomial time algorithm for separability detection, and sparked very interesting developments.

Berkeley and Zurich After finishing my PhD, I went to Berkeley for a few months, hosted by Bernd Sturmfels and Laurent El Ghaoui, before starting a tenure-track position at ETH Zurich. That relatively short time proved to be incredibly exciting, and quite influential in my future research.

With Karin Gatermann, who was also visiting Berkeley at the time (and who sadly, would pass away a few years later at a young age) we began developing nice connections between SOS decompositions of group-invariant polynomials and representation theory. The key idea was to realize that the joint presence of symmetry and convexity allowed us to impose certain strong conditions on a sum of squares decomposition; without loss of generality, one could choose a representing Gram matrix from the *fixed point subspace*. This, in combination with Schur's lemma, made possible an equivalent reformulation, using smaller SDPs that were much easier to solve. These results would eventually become [15], and a number of applications of these methods, such as [28] followed.

While at Zurich, I became interested in "hyperbolic programming," an elegant formalization and abstraction of semidefinite programming developed by Güler, Tunçel and Renegar, among others. In joint work with Motakuri Ramana and Adrian Lewis [20], we realized the intimate connections between some recent work of Helton and Vinnikov on semidefinite representability of planar convex sets [19] and a classical conjecture of Peter Lax on determinantal representations of hyperbolic polynomials. Indeed, these two questions were essentially the same, modulo homogenization, and the Helton/Vinnikov theorem was used to settle this conjecture.

Despite working extremely well in practice (for those problems where we could solve them numerically!), there were relatively few theoretical results to assess the quality of SOS relaxations. Around 2003, in joint work with Etienne de Klerk and Monique Laurent [10], we showed that for the case of polynomial optimization over the simplex, SOS relaxations provided a fully polynomial-time approximation scheme (PTAS). In fact, we showed that the same result could be achieved by simpler algorithms such as a straightforward discretization scheme, and extended this to more general feasible sets such as polytopes with polynomially many vertices.

Incidentally, over the last few years there has been a surge of interest in the theoretical computer science community on the power and limitations of sum of squares hierarchies. In particular, recent works such as [4, 22] have provided novel insights and rounding algorithms, and in fact, suggest the possibility that the SOS approach may perhaps be used to settle Khot's "unique games" conjecture.

Playing games at MIT I moved to MIT in the Fall of 2004, and this proved to be (and still is!) an extremely stimulating intellectual experience. In both the Laboratory for Information and Decision Systems (LIDS) and the Operations Research Center (ORC) I have found amazing colleagues and students, and the source of many collaborations.

One of my first papers after arriving to MIT was a short note on polynomial games, a nice class of two-person zero-sum games originally studied by Dresher, Karlin and Shapley [11]. As it turned out, sum of squares and SDP methods were the perfect tool for the computation of minimax equilibria for these games. In particular, this gave a completely satisfactory generalization of the classical LP solution of bimatrix games, to the case where the players must choose actions on a continuous interval and the payoff is a polynomial function of the players' decisions [26]. In subsequent works with Noah Stein and Asu Ozdaglar we significantly extended these findings, by developing structural results and computational methods for other infinite games and related notions of equilibria, e.g., [35, 36].

Rank, sparsity, and beyond In 2006-07, the Institute for Mathematics and its Applications (IMA) hosted a year-long program on Applications of Algebraic Geometry. One frigid evening in January 2007, Ben Recht and I went out for dinner at a Chinese restaurant (that no longer exists) near the IMA, and started discussing some of the recent developments in "compressed sensing," a fairly new area at the time. It quickly become clear to us that it should be possible to extend many of the results about recovery of sparse vectors from linear measurements using ℓ_1 minimization, to the matrix situation where one minimizes the sum of singular values (i.e., the "nuclear norm" heuristic, that Maryam Fazel had developed for her PhD thesis). Maryam (who was also visiting the IMA) immediately joined us in the project, that quickly led to [32]. This starting point provided the spark for much follow-up work in the area.

The interaction between convex-algebraic ideas and probabilistic models proved to be very fruitful. In joint work with Venkat Chandrasekaran, Sujay Sanghavi, James Saunderson and Alan Willsky we introduced novel convex relaxations for several kinds of matrix decompositions (e.g., low-rank/sparse, or low-rank/diagonal), and the associated probabilistic descriptions such as Gaussian graphical models or correlation matrices [9, 33]. The geometry of the regularizing norms, given by convex hulls of certain algebraic varieties, played a crucial role. In particular, this led to the "atomic norm" framework [8] (with Venkat Chandrasekaran, Ben Recht, and Alan Willsky), where a very appealing geometric description of many "sensing" results was developed. This made possible a much simpler derivation of many of the known results, including a quantification of sample complexity in terms of natural geometric and probabilistic invariants such as the *Gaussian width*, as well as many extensions.

SOS convexity and SDP representability Despite the large body of theory, and the enormous number of successful applications, semidefinite programming still remains somewhat mysterious (at least, to me!). In particular, the question "which sets are SDP-representable?" still lacks a fully satisfactory answer, although great advances have been made in recent years.

A good contrast is the case of linear programming (LP). Consider for simplicity the case of convex sets that are full-dimensional and compact (i.e., convex bodies). In this case, the Minkowski-Weyl theorem exactly characterizes the feasible sets of LPs (polyhedra) as those sets with finitely many extreme points. A similarly complete characterization for SDP representability is not (yet?) available.

The notion of *sum of squares convexity* (or SOS-convexity for short), originally introduced by Helton and Nie [18], provides a nice algebraic analogue of classical "geometric" convexity for polynomial functions. Indeed, instead of mere positive definiteness, SOS-convexity requires the Hessian of a polynomial to factor in terms of polynomial matrices. In joint work with Amir Ali Ahmadi [1, 2] we provided a complete characterization and established when it is equivalent to "standard" convexity.

In recent work with João Gouveia and Rekha Thomas [17, 16] we began developing a systematic approach to SDP representability. In particular, this led to a very appealing characterization in terms of a quantity we call *positive semidefinite rank* (psd-rank). For simplicity, we describe only the case of polytopes; see [17] for general convex bodies. Given a nonnegative matrix $M \in \mathbb{R}_+^{n\times m}$, its psd-rank is the smallest integer k for which there exist positive semidefinite matrices $A_1, \ldots, A_n$ and $B_1, \ldots, B_m$ of size $k \times k$ such that

$$M_{ij} = \langle A_i, B_j \rangle.$$

It turns out that the psd-rank determines the size of the smallest SDP representation of a polytope with slack matrix M. This generalizes earlier work of Yannakakis [37] on the complexity of extended LP formulations of polytopes.

Convex algebraic geometry As is clear from the previous paragraphs, much of my work so far has been an attempt to understand the beautiful interactions between optimization, convex geometry, and algebraic geometry. The recent SIAM book [7] presents many of the developments in this evolving subject (known as "convex algebraic geometry"), and is the result of an NSF Focused Research Group (FRG) established a few years ago along with several of my colleagues (Helton, Nie, Sturmfels, Thomas). This is a truly exciting area, that has already shown to have a large number of potential applications, rich connections across the mathematical sciences, and provided novel tools for applied mathematics and engineering. Certainly much more work is needed in this area, which hopefully will keep me (and many others!) busy for the next few years...

Finally... I am enormously grateful to my students, friends, and colleagues for the many things they have taught me. John Doyle, Stephen Boyd, Bernd Sturmfels, Bill Helton, Dimitris Bertsimas and Rekha Thomas (in strict chronological order!) have been particularly influential figures throughout my career, as it is surely obvious from a careful analysis of the records. I want to use this opportunity to publicly thank them for their continuing support, and always helpful advice.

REFERENCES

[1] A. A. Ahmadi and P. A. Parrilo. A convex polynomial that is not sos-convex. *Mathematical Programming*, 135(1-2):275–292, 2012.

[2] A. A. Ahmadi and P. A. Parrilo. A complete characterization of the gap between convexity and sos-convexity. *SIAM Journal on Optimization*, 23(2):811–833, 2013.

[3] F. G. S. L. Brandão, M. Christandl, and J. Yard. A quasipolynomial-time algorithm for the quantum separability problem. In *Proceedings of the 43rd annual ACM Symposium on Theory of Computing*, STOC '11, pages 343–352, 2011.

[4] B. Barak, F. G. S. L. Brandão, A. W. Harrow, J. Kelner, D. Steurer, and Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their

applications. In *Proceedings of the 44th Symposium on Theory of Computing*, pages 307–326. ACM, 2012.

[5] J. Bochnak, M. Coste, and M-F. Roy. *Real Algebraic Geometry*. Springer, 1998.

[6] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, 1994.

[7] G. Blekherman, P. A. Parrilo, and R. Thomas, editors. *Semidefinite optimization and convex algebraic geometry*, volume 13 of *MOS-SIAM Series on Optimization*. SIAM, 2012.

[8] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.

[9] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

[10] E. de Klerk, M. Laurent, and P. A. Parrilo. A PTAS for the minimization of polynomials of fixed degree over the simplex. *Theoretical Computer Science*, 361(2-3):210–225, 2006.

[11] M. Dresher, S. Karlin, and L. S. Shapley. Polynomial games. In *Contributions to the Theory of Games*, Annals of Mathematics Studies, no. 24, pages 161–180. Princeton University Press, Princeton, N. J., 1950.

[12] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. Distinguishing separable and entangled states. *Physical Review Letters*, 88(18), 2002.

[13] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. Complete family of separability criteria. *Physical Review A*, 69:022308, 2004.

[14] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. Detecting multipartite entanglement. *Physical Review A*, 71(3):032333, 2005.

[15] K. Gatermann and P. A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *Journal of Pure and Applied Algebra*, 192(1-3):95–128, 2004.

[16] J. Gouveia, P. A. Parrilo, and R. R. Thomas. Approximate cone factorizations and lifts of polytopes. *arXiv:1308.2162*, 2013.

[17] J. Gouveia, P. A. Parrilo, and R. R. Thomas. Lifts of convex sets and cone factorizations. *Mathematics of Operations Research*, 38(2):248–264, 2013.

[18] J. W. Helton and J. Nie. Semidefinite representation of convex sets. *Mathematical Programming*, 122(1):21–64, 2010.

[19] J. W. Helton and V. Vinnikov. Linear matrix inequality representation of sets. *Comm. Pure Appl. Math.*, 60(5):654–674, 2007.

[20] A. S. Lewis, P. A. Parrilo, and M. V. Ramana. The Lax conjecture is true. *Proc. Amer. Math. Soc.*, 133(9):2495–2499, 2005.

[21] Y. Nesterov. Squared functional systems and optimization problems. In J.B.G. Frenk, C. Roos, T. Terlaky, and S. Zhang, editors, *High Performance Optimization*, pages 405–440. Kluwer Academic Publishers, 2000.

[22] R. O'Donnell and Y. Zhou. Approximability and proof complexity. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 1537–1556. SIAM, 2013.

[23] P. A. Parrilo. On a decomposition for multivariable forms via LMI methods. In *Proceedings of the American Control Conference*, 2000.

[24] P. A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, May 2000.

[25] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Prog.*, 96(2, Ser. B):293–320, 2003.

[26] P.A. Parrilo. Polynomial games and sum of squares optimization. In *Proceedings of the 45^{th} IEEE Conference on Decision and Control*, 2006.

[27] A. Packard and J. C. Doyle. The complex structured singular value. *Automatica*, 29(1):71–109, 1993.

[28] P. A. Parrilo and R. Peretz. An inequality for circle packings proved by semidefinite programming. *Discrete and Computational Geometry*, 31(3):357–367, 2004.

[29] S. Prajna, A. Papachristodoulou, and P. A. Parrilo. *SOSTOOLS: Sum of squares optimization toolbox for MATLAB*. Available from `www.cds.caltech.edu/sostools` and `www.mit.edu/~parrilo/sostools`.

[30] P.A. Parrilo and B. Sturmfels. Minimizing polynomial functions. In S. Basu and L. González-Vega, editors, *Algorithmic and Quantitative Real Algebraic Geometry*, volume 60 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. AMS, 2003.

[31] B. Reznick. Some concrete aspects of Hilbert's 17th problem. In *Contemporary Mathematics*, volume 253, pages 251–272. American Mathematical Society, 2000.

[32] B. Recht, M. Fazel, and P.A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[33] J. Saunderson, V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky. Diagonal and low-rank matrix decompositions, correlation matrices, and ellipsoid fitting. *SIAM Journal on Matrix Analysis and Applications*, 33(4):1395–1416, 2012.

[34] N. Z. Shor. Class of global minimum bounds of polynomial functions. *Cybernetics*, 23(6):731–734, 1987. (Russian orig.: Kibernetika, No. 6, (1987), 9–11).

[35] N. D. Stein, A. Ozdaglar, and P. A. Parrilo. Separable and low-rank continuous games. *International Journal of Game Theory*, 37(4):475–504, 2008.

[36] N. D. Stein, P. A. Parrilo, and A. Ozdaglar. Correlated equilibria in continuous games: Characterization and computation. *Games and Economic Behavior*, 71(2):436–455, 2011.

[37] M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. System Sci.*, 43(3):441–466, 1991.

Sanjay Mehrotra, James Luedtke and Alper Atamtürk

A Branch-and-Cut Decomposition Algorithm for Solving Chance-Constrained Mathematical Programs with Finite Support

James R. Luedtke
Department of Industrial and Systems Engineering
University of Wisconsin-Madison, Madison, WI, USA
jrluedt1@wisc.edu

This article summarizes the paper [4], which is concerned with solving a chance-constrained mathematical program (CCMP) of the form:

$$\min\Big\{f(x) : \mathbb{P}\{x \in P(\omega)\} \geq 1-\epsilon,\ x \in X\Big\}\ , \quad (1)$$

where $x \in \mathbb{R}^n$ is the vector of decision variables to be chosen to minimize $f : \mathbb{R}^n \to \mathbb{R}$, ω is a random vector assumed to have finite support, $P(\omega) \subseteq \mathbb{R}^n$ is a polyhedron for each ω, and $X \subseteq \mathbb{R}^n$ represents a set of deterministic constraints on x. $\epsilon \in (0,1)$ is a tolerance typically chosen close to zero, representing the desire to have the constraint $x \in P(\omega)$ hold with high probability. We let $\omega^k, k \in \mathcal{N} := \{1,\dots,N\}$ be the *scenarios* defining the support of ω, and assume for simplicity of exposition that $\mathbb{P}(\omega^k) = 1/N$ for $k \in \mathcal{N}$. For CCMPs in which ω does not have finite support, sample average approximation [5, 7] can be used to obtain approximations having finite support that can be used to construct feasible solutions and statistical estimates of solution quality of the original CCMP.

A generic example is when $P(\omega^k)$ is the set of points for which there exists a recourse decision y that satisfies a set of random linear constraints, i.e.,:

$$P(\omega^k) = \{x \in \mathbb{R}^n_+ : \exists y \in \mathbb{R}^d_+ \text{ with } T^k x + W^k y \geq b^k\}\ . \quad (2)$$

In this case the CCMP (1) can be formulated as a mixed-integer program (MIP) by introducing binary decision variables $z_k, k \in \mathcal{N}$, such that $z_k = 0$ implies $x \in P(\omega^k)$, and copies of the recourse variables $y^k,\ k \in \mathcal{N}$ to represent the recourse actions taken in each scenario,

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t. } T^k x + W^k y^k + z_k M_k \geq b^k, \quad k \in \mathcal{N} && \text{(3a)} \\ &\qquad \sum_{k=1}^{N} z_k \leq p\ , && \text{(3b)} \\ &\qquad x \in X, \quad z \in \{0,1\}^N, \quad y^k \in \mathbb{R}^d_+, \quad k \in \mathcal{N} && \text{(3c)} \end{aligned}$$

where $p = \lfloor \epsilon N \rfloor$. Here $M_k \in \mathbb{R}^m_+, k \in \mathcal{N}$ are sufficiently large to ensure that when $z_k = 1$, constraints (3a) are not active. On the other hand, when $z_k = 0$, constraints (3a) enforce $x \in P(\omega^k)$. This MIP formulation has two drawbacks: (i) it is large when the sample size N is large, due to having N copies of the recourse variables y^k and constraints (3a), and (ii) the "big-M" constraints (3a) are likely to lead to weak linear programming relaxations. The algorithm proposed in [4] uses decomposition and strong valid inequalities to overcome these drawbacks.

Algorithm

The decomposition approach is similar to that first proposed by Shen, Smith, and Ahmed [9]. A branch-and-cut algorithm is used to solve a master problem of the form

$$\begin{aligned} \min\ f(x) \text{ s.t. } &\sum_{k=1}^{N} z_k \leq p, && \text{(4a)} \\ &z_k = 0\ \Rightarrow x \in P(\omega^k), && \text{(4b)} \\ &x \in X, \quad z \in \{0,1\}^N\ , && \text{(4c)} \end{aligned}$$

Initially, the constraints (4b) and the integrality constraints are relaxed and a branch-and-bound search tree is constructed, where branching is done to enforce the integrality restrictions on z and any integer-constrained x variables.

When a solution $(\hat{x}, \hat{z})$ that satisfies the integrality constraints is obtained at a node relaxation solution (or via a heuristic), the logical conditions (4b) must be checked to determine if the solution is really feasible to (4). Thus, for each $k \in \mathcal{N}$ with $\hat{z}_k = 0$, a *separation* problem is solved which determines if $\hat{x} \in P(\omega^k)$ and if not, returns an inequality of the form

$$\alpha x \geq \beta \qquad (5)$$

valid for $P(\omega^k)$ and for which $\alpha \hat{x} < \beta$. It is assumed that the separation problem returns an inequality from a finite set. For example, if $P(\omega^k)$ is described as in (2), then the separation problem could be implemented with[1]

$$\begin{aligned} &\max_{\pi}\ \pi^T (b^k - T^k \hat{x}) \\ &\text{s.t. } \pi^T W^k \leq 0, \pi^T e \leq 1, \pi \in \mathbb{R}^m_+. \end{aligned}$$

If the optimal value of this linear program is positive with optimal solution $\hat{\pi}$, then $\alpha = \hat{\pi}^T T^k$ and $\beta = \hat{\pi}^T b^k$ yields the required inequality (5). If $\hat{x} \in P(\omega^k)$ for all k with $\hat{z}_k = 0$, then the solution is feasible and so the incumbent solution is updated. Otherwise, an inequality (a lazy constraint) must be added to the formulation to cut off solution $(\hat{x}, \hat{z})$. In particular, a basic decomposition algorithm can be obtained by adding an inequality of the form

$$\alpha x + M(\alpha) z_k \geq \beta,$$

where $M(\alpha)$ is large enough so that the inequality $\alpha x \geq \beta - M(\alpha)$ is satisfied by any solution feasible to (4). Such an algorithm accomplishes the goal of decomposition – separation problems are solved a single scenario at a time – but still suffers from the use of potentially weak big-M coefficients.

To obtain a set of stronger valid inequalities, when an inequality of the form (5) is obtained, the algorithm solves a set of single-scenario *optimization* problems:

$$h_k(\alpha) := \min\{\alpha x : x \in P(\omega^k) \cap \hat{X}\} \qquad (6)$$

where $\hat{X} \supseteq X$ contains a subset (possibly all) of the constraints on the x variables. Choosing $\hat{X} = X$ may yield the strongest valid inequalities, but at the expense of potentially making the subproblems (6) more difficult. At the other extreme, one could choose $\hat{X} = \mathbb{R}^n$.

[1] In [4], this problem has a regrettable typo: the constraint $\pi^T e \leq 1$ is incorrectly stated as $\pi^T e = 1$ in [4].

Having obtained the values $h_k(\alpha)$ for $k \in \mathcal{N}$, they are then sorted to obtain a permutation σ of $\mathcal{N}$ such that:

$$h_{\sigma_1}(\alpha) \geq h_{\sigma_2}(\alpha) \geq \cdots \geq h_{\sigma_N}(\alpha) .$$

These values can then be used to derive a set of inequalities valid for (4):

$$\alpha x + (h_{\sigma_i}(\alpha) - h_{\sigma_{p+1}}(\alpha)) z_{\sigma_i} \geq h_{\sigma_i}(\alpha), \quad i = 1, \ldots, p . \tag{7}$$

In particular, when $z_{\sigma_i} = 0$, the constraint (7) enforces $\alpha x \geq h_{\sigma_i}(\alpha)$ which is valid by definition of $h_{\sigma_i}(\alpha)$. When $z_{\sigma_i} = 1$, the constraint reduces to $\alpha x \geq h_{\sigma_{p+1}}(\alpha)$ which is valid due to (3b). One can interpret the coefficient $(h_{\sigma_i}(\alpha) - h_{\sigma_{p+1}}(\alpha))$ as a *strengthened* big-M coefficient.

The inequalities (7) can then be "mixed" using results of [1] or [2] to obtain an exponential class of additional strong valid inequalities:

$$\alpha x + \sum_{i=1}^{\ell} (h_{t_i}(\alpha) - h_{t_{i+1}}(\alpha)) z_{t_i} \geq h_{t_1}(\alpha) \tag{8}$$

for any $T = \{t_1, t_2, \ldots, t_\ell\} \subseteq \{\sigma_1, \ldots, \sigma_p\}$ with $h_{t_i}(\alpha) \geq h_{t_{i+1}}(\alpha)$ for $i = 1, \ldots, \ell$, and with $h_{t_{\ell+1}}(\alpha) := h_{\sigma_{p+1}}(\alpha)$. A most violated inequality in this class can be found efficiently using results from [1] or [2]. The use of mixing sets for solving CCMPs was first proposed in [6] for the case of single-stage problems having only right-hand side uncertainty, and was further studied in [3].

To summarize, when an integer feasible solution $(\hat{x}, \hat{z})$ is found in the branch-and-cut algorithm, a *single-scenario separation* problem is solved for each scenario with $\hat{z}_k = 0$ to determine if $\hat{x} \in P(\omega^k)$. If a scenario with $\hat{x} \notin P(\omega^k)$ is found, then a separating inequality (5) is found and the coefficient α is used as the objective in the *single-scenario optimization* problems (6). The optimal values of these problems are then used to construct the base valid inequalities (7) (with strengthened big-M coefficients), and finally a separation algorithm is used to search for one or more violated mixing inequalities of the form (8) which are then added to the formulation. Correctness of this algorithm is assured by demonstrating that this procedure is guaranteed to cut off any points $(\hat{x}, \hat{z})$ that do not satisfy the logical constraints (4b).

The computational performance of the algorithm can be improved in several ways, described in detail in [4]. First, to improve the LP relaxation bounds, the cut generation procedure described above can be called also at solutions $(\hat{x}, \hat{z})$ that are not integer feasible. Second, information about the success of past calls to the separation problems for all the scenarios is used to choose the sequence in which to solve the separation problems, with the hope of quickly finding a scenario k with $\hat{x} \notin P(\omega^k)$ if any such scenario exists. Finally, any time the problems (6) are solved, the values $\{h_k(\alpha) : k \in \mathcal{N}\}$ are saved and when the cut generation routine is called, it first checks for a violated mixing inequality (8) using any of these stored sets of values, and if a violated inequality is found it is immediately added and the LP relaxation is re-solved. This is advantageous because the separation of the inequalities (8) is very efficient compared to solving the all the single-scenario separation and optimization problems.

Summary of Computational Results

The algorithm was tested on a two-stage resource planning problem. In this problem, the amounts of a set of resources to have available must be decided before observing the random amounts of customer demands, and after the demands are known the resources are allocated to meet the customer demands. The goal is to minimize the cost of the resources while ensuring that, with high probability, all customer demands can be satisfied by some allocation of the resources. In the first version of the test problem, the set $P(\omega^k)$ has the form (2), but with the special structure that the coefficients in the constraints are all deterministic; only the right-hand side (the customer demands) is random. The algorithm was also tested on instances with random resource yields (some proportion of the planned resources are not available), and in which the

Table 1: Results for instances with random demands only.

(n, m)	ϵ	Basic Dec. Gap (%)	Strong Dec. Time (sec)
$(20, 30)$	0.05	6.8%	16
	0.1	23.0%	22
$(40, 50)$	0.05	16.6%	25
	0.1	25.4%	26

Table 2: Results for instances with random demands, yields, and service rates.

(n, m)	ϵ	Basic Dec. Gap (%)	Strong Dec. Gap (%)
$(10, 20)$	0.05	9.0%	0.3%
	0.1	15.7%	0.7%
$(20, 30)$	0.05	14.5%	1.4%
	0.1	20.9%	1.9%

customer service rates are random, leading to instances with random constraint coefficients.

The algorithm is compared to solving the large formulation (3) directly, and to a decomposition algorithm that does not use the mixing inequalities to obtain stronger valid inequalities. For the instance sizes tested, formulation (3) left very large optimality gaps after the one hour time limit, so in this summary we present only the comparison between the two decomposition algorithms. Table 1 shows a sample of the results for a set of instances with only random demands, having $N = 3000$ scenarios and having varying size in terms of number of resource types n and customer types m. Table 2 shows a sample of the results for a set of instances with random demands, yields, and service rates, having $N = 1500$ scenarios and varying sizes. The entries in both of these tables are averages over five instances.

The results indicate that the proposed algorithm can solve large instances of this problem, especially when only the demands (right-hand side of the constraints) are random, whereas a simple decomposition algorithm fails. The instances with random constraint coefficients are more difficult, and so the algorithms are tested on smaller instances. Even on these smaller instances the proposed algorithm is no longer able to solve all the instances in the time limit. However, the ending optimality gaps are very small, and much smaller than what is obtained using the basic decomposition algorithm.

We close this section by commenting on some experience with this algorithm that has been reported in more recent work on CCMPs having a different structure: no recourse variables and a small number of constraints, all having random coefficients. These problems have little benefit from a decomposition algorithm because there are no recourse variables and the number of cuts that would be added might easily exceed the size of the original formulation. Thus, the application of the proposed algorithm to these problems amounts to using the mixing inequalities to obtain stronger relaxations. Qiu et al. [8] and [10] have found that while using the proposed algorithm significantly outperforms the use of the standard MIP formulation with naively chosen big-M coefficients, just performing big-M coefficient strengthening, e.g., using the logic that leads to the constraints (7), already leads to a substantial improvement in the ability to solve the MIP formulation. For such instances it appears the incremental benefit in improving the bound using the mixing inequalities does not outweigh the computational burden of adding these additional cuts to the formulation.

Solving for the Efficient Frontier

A common question when solving a CCMP is how to choose the risk tolerance ϵ. The answer is to consider risk and cost as two competing objectives, and construct an (approximate) efficient frontier that shows for varying levels of risk level ϵ what the corresponding minimum cost is. A simple way to approximate this efficient frontier is to solve the CCMP for a variety of risk levels ϵ, and plot the resulting pairs of risk-cost values. In [4] we show how the proposed algorithm can be adapted to re-use information from solving a CCMP at one risk level when solving the next. In particular, by solving in increasing order or risk level, feasible solutions at one risk level remain feasible at the next risk level, and so can be used as initial incumbent solutions. More significantly, the information used to generate the valid inequalities in the algorithm, the coefficient vectors α and the corresponding values $h_k(\alpha), k \in \mathcal{N}$ can be saved and used to generate cuts when solving for different risk levels. This can save a significant amount of time, as finding these values represents the most time-consuming component of the cut-generating procedure. Computational results indicate that using these warm-start ideas to construct an approximate efficient frontier by solving at 16 different risk levels can lead to a 50-75% reduction in computational time compared to solving the problems independently.

Acknowledgements

I would like to thank the prize committee members Alper Atamtürk, Sam Burer, Andrzej Ruszczynski, and Nikolaos Sahinidis for this honor, and all Optimization Society prize committee members for their service to the Society and the community.

REFERENCES

[1] A. Atamtürk, G. L. Nemhauser, and M. W. P. Savelsbergh. The mixed vertex packing problem. *Math. Program.*, 89:35–53, 2000.

[2] O. Günlük and Y. Pochet. Mixing mixed-integer inequalities. *Math. Program.*, 90:429–457, 2001.

[3] S. Küçükuavuz. On mixing sets arising in chance-constrained programming. *Math. Program.*, 132:31–56, 2012.

[4] J. Luedtke. A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Math. Program.*, 2013. DOI 10.1007/s10107-013-0684-6.

[5] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM J. Optim.*, 19:674–699, 2008.

[6] J. Luedtke, S. Ahmed, and G. L. Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Math. Program.*, Oct 2008.

[7] B. Pagnoncelli, S. Ahmed, and A. Shapiro. Sample average approximation method for chance constrained programming: Theory and applications. *J. Optim. Theory Appl.*, 142(2):399–416, 2009.

[8] F. Qiu, S. Ahmed, S. Dey, and L. Wolsey. Covering linear programs with violations. *INFORMS J. Comput.*, 2013. Accepted for publication.

[9] S. Shen, J. Smith, and S. Ahmed. Expectation and chance-constrained models and algorithms for insuring critical paths. *Manage. Sci.*, 56:1794–1814, 2010.

[10] Y. Song, S. Küçükyavuz, and J. Luedtke. Chance-constrained binary packing problems. *INFORMS J. Comput.*, 2013. Accepted for publication.

Sparse Recovery in Derivative-Free Optimization

Afonso S. Bandeira
Program in Applied and Computational Mathematics
Princeton University, Princeton, NJ, USA
ajsb@math.princeton.edu

First of all, it is an honor to receive the INFORMS Optimization Society best student paper award. Before going into the content of the paper, I want to give a bit of context in which I did this research. This paper was essentially my master thesis work in the University of Coimbra, in Portugal, back in 2009-2010 (In fact, my master thesis has the same title [1]). Katya Scheinbery and Luís Nunes Vicente proposed me to work on the connection between the, then recent, developments in sparse recovery and Compressed Sensing (subject I was, and still am, quite interested in) and Derivative-Free Optimization, a subject I was not familiar with at the time but which I very quickly learned to enjoy. During this period I was extremely fortunate to have had the opportunity to work with both Luis and Katya and, together, we wrote the awarded paper, "Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization" [5]. This work has spurred both parallel [4] and future research [6] that I will also briefly describe below.

The framework is unconstrained optimization: one wants to minimize a (sufficiently smooth) function $f\colon \mathbb{R}^n \to \mathbb{R}$ over $\mathbb{R}^n$. In many applications function evaluations are particularly expensive and one has no access to function derivatives (an important example is when the goal is to do parameter optimization and each evaluation requires an expensive simulation). These applications motivate the interest in optimizing f without using its derivatives and using as few function evaluations as possible, this is known as Derivative-Free Optimization. An excellent introduction to the topic is given in the book [10].

One type of algorithms used in DFO are the, so called, model-based trust-region methods. Essentially, they operate by iteratively picking a small region, known as the trust-region, $B \subset \mathbb{R}^n$ (say a ball) and build (via approximation on samples of f) a model $m\colon B \to \mathbb{R}$ of f in B. The idea is for $m(x)$ to be easier to optimize in B while being a reliable model of f, and for minimizer of $m(x)$ to be an estimate for the minimizer of f in B. Depending on the location of the minimizer and its value on f the trust-region is updated and the procedure repeated.

Sanjay Mehrotra, Afonso Bandeira and Simge Küçükyavuz

A very popular class of models used is the quadratic polynomials, as these are reasonably simple while being able capturing curvature of the function (unlike linear models). For the sake of simplicity, let us suppose for the moment that f is itself a quadratic and we want to find the model $m = f$. Constructing $m(x)$ from p function evaluations of f corresponds to solving a linear system. Let $\phi_1, \ldots, \phi_N$ be a basis for the quadratic polynomials of n variables (meaning $N = \frac{(n+3)n+2}{2}$). We can write $m(x) = \sum_{i=1}^N \alpha_i \phi_i(x)$ and focus on estimating the $\{\alpha_i\}_{i=1}^N$. In fact, a function evaluation of f at y_j gives a linear constraint on α,

$$\sum_{i=1}^{N} \alpha_i \phi_i(y_j) = f(y_j).$$

A sample set Y of p points corresponds to p linear constraints on α which we represent by an interpolation matrix $M(\phi, Y)$

$$M(\phi, Y)\alpha = f(Y). \tag{1}$$

In order for (1) to be a determined system one needs $p \geq N = \frac{(n+3)n+2}{2}$ function evaluations in each iterations, which is very often too expensive.

Indeed, without any extra information on the structure of f, this is the best one can do. Fortunately, most functions that arise from applications have special structures. For example, in the parameter estimation problem, it is rather unlikely that every pair of parameters is interacting (in a relevant way). Pairs of parameters not interacting should correspond to zero joint derivatives which suggests sparsity of the function's Hessian. This motivated us to pursue techniques that exploited the Hessian sparsity in order to construct reliable models with far fewer samples, which is precisely the subject of the paper [5].

Provided we choose a basis $\{\phi\}_{i=1}^N$ such that Hessian sparsity translates into sparsity in α, the Hessian sparsity of f corresponds to sparsity of the solution of the linear system (1). Around a decade ago, the seminal work of Candes, Donoho, and others [8, 9, 11, 12, 13], spurred a vast body of exciting research in the area of sparse recovery (also known as Compressed Sensing) which provides very good understanding of when one is able to recover a sparse vector α from a underdetermined linear system. The main contribution of this paper is leveraging, and adapting, these results to estimate α (which gives a model $m(x)$), via (1), using significantly less than $N = \mathcal{O}(n^2)$ function evaluations.

In a nutshell, the theory of Compressed Sensing tells us that, if $M(\phi, Y)$ satisfies a certain property (known as the Restricted Isometry Property (RIP) [7]), the sparse vector α can be recovered by ℓ_1 minimization (essentially, it is the vector with minimum ℓ_1 norm which still satisfies the linear constraints). Matrices satisfying the Restricted Isometry Property are notably difficult to build and computationally hard to certify [2] (I have spent some time thinking about this myself [3]) but random constructions are known to yield RIP matrices for a number of rows (corresponding to samples) p on the order of $p = \mathcal{O}(k \log N)$, where k is the sparsity of the vector, and N the ambient dimension (in our case $N = \mathcal{O}(n^2)$). This means that, as long as $M(\phi, Y)$ is RIP, the number of samples needed is no longer on the order of vector dimension but instead, on the order of the sparsity of α (with a small logarithmic loss). Moreover, ℓ_1 minimization can be formulated as a linear program thus enjoying many efficient algorithms.

Classically, the results in Compressed Sensing guaranteeing the RIP property for random matrices mostly concern matrices with independent entries. In our setting, however, we are constrained to a very structured interpolation matrix $M(\phi, Y)$. Knowing how difficult constructing good deterministic RIP matrices seems to be, we opted to "inject randomness" in the matrix by taking the sample set Y to be random (while the basis $\{\phi\}_{i=1}^N$ is fixed and deterministic). In fact, provided that the basis $\{\phi\}_{i=1}^N$ satisfies certain properties, a sufficiently large random sample set Y gives an interpolation matrix $M(\phi, Y)$ which is known [17] to be RIP with high probability. In our paper [5], we are able to build a basis $\{\phi\}_{i=1}^N$ both inheriting sparsity from Hessian sparsity and satisfying the properties needed to yield RIP interpolation matrices. This, together with a particular choice of trust-region and sampling measure for Y allowed us to show that, with high probability, ℓ_1 minimization succeeds in recovering α from the linear measurements (1) with as few as

$$p = \mathcal{O}(n \log^4(n))$$

samples, provided that the Hessian of f has $\mathcal{O}(n)$ non-zero entries. Note that this number of samples (corresponding to function evaluations) is considerably less than the $\mathcal{O}(n^2)$ samples that would be needed in the classical case.

In general, f is not a quadratic polynomial. However, as long as f is sufficiently smooth, one can show that the procedure sketched above gives, with the same number of samples, a model m that approximates f in B essentially as well as its second-order Taylor approximation (these are known as fully-quadratic models). The idea is to replace f with its second-order Taylor approximation in the arguments above. In that case, each sample of f can be regarded as a noisy sample of the quadratic approximation. Fortunately, the guarantees given in the theory of sparse recovery often come with robustness to noise and, in this case, we can leverage such results to ensure the recovery of a fully-quadratic model of f, with high probability.

The sparsity assumption used is on the Hessian of the function, however the coefficients α also describe the gradient and constant term which may not be sparse. This means that there are some entries of the vector α that are not believed to be sparse. Motivated by this fact, we investigated the problem of sparse recovery for partially sparse vectors [4]. We showed that, not very surprisingly, one should do ℓ_1 minimization only on the entries that are believed to be sparse.

Using the machinery described above, we developed a model-based trust-region method based on minimum ℓ_1 model construction. In our experiments, this method was able to compete with state of the art Derivative-Free methods, such as NEWUOA [16, 15] on the standard problem data base CUTEr [14].

A natural question raised by this work regard the convergence of methods based on this type of random models. Recall that recovery is only ensured with high probability at each iteration and so it will likely fail on some iterations. This question was the target of further research [6] where we showed that, under somewhat general conditions, the convergence guarantees for model-based trust-region methods can be adapted to handle this uncertainty in the model construction step. Essentially, we showed that, as long as the probability of constructing a good model on each iteration is over one half, these methods still converge.

REFERENCES

[1] A. S. Bandeira, *Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization*, Master's thesis, Dept. of Mathematics, Univ. of Coimbra, 2010.

[2] A.S. Bandeira, E. Dobriban, D.G. Mixon, and W.F. Sawin, *Certifying the restricted isometry property is hard*, IEEE Trans. Inform. Theory **59** (2013), no. 6, 3448–3450.

[3] A. S. Bandeira, M. Fickus, D. G. Mixon, and P. Wong, *The road to deterministic matrices with the restricted isometry property*, Journal of Fourier Analysis and Applications **19** (2013), no. 6, 1123–1149.

[4] A. S. Bandeira, K. Scheinberg, and L. N. Vicente, *On partial sparse recovery*, Tech. report, CMUC, Department of Mathematics, University of Coimbra, Portugal, 2011.

[5] A. S. Bandeira, K. Scheinberg, and L. N. Vicente, *Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization*, Mathematical Programming **134** (2012), 223–257.

[6] A. S. Bandeira, K. Scheinberg, and L. N. Vicente, *Convergence of a trust region method based on probabilistic models*, Available online at arXiv:1304.2808 [math.OC] (2013).

[7] E. J. Candès, *The restricted isometry property and its implications for compressed sensing*, C. R. Acad. Sci. Paris, Ser. I **346** (2008), 589–592.

[8] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory **52** (2006), 489–509.

[9] E. J. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59** (2006), 1207–1223.

[10] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to derivative-free optimization*, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2009.

[11] E. J. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory **51** (2005), 4203–4215.

[12] E. J. Candès and T. Tao, *Near optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory **52** (2006), 5406–5425.

[13] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory **52** (2006), 1289–1306.

[14] N. I. M. Gould, D. Orban, and P. L. Toint, *Cuter and sifdec: A constrained and unconstrained testing environment, revisited*, ACM Trans. Math. Softw. **29** (2003), no. 4, 373–394.

[15] M. J. D. Powell, *On trust region methods for unconstrained minimization without derivatives*, Mathematical Programming **97** (2003), no. 3, 605–623.

[16] M. J. D. Powell, *The newuoa software for unconstrained optimization without derivatives*, Large-Scale Nonlinear Optimization (G. Pillo and M. Roma, eds.), Nonconvex Optimization and Its Applications, vol. 83, Springer US, 2006, pp. 255–297.

[17] H. Rauhut, *Compressive sensing and structured random matrices*, Theoretical Foundations and Numerical Methods for Sparse Recovery (M. Fornasier, ed.), Radon Series Comp. Appl. Math., vol. 9, deGruyter, 2010, pp. 1–92.

Nominations for Society Prizes Sought

The Society awards four prizes annually at the INFORMS annual meeting. We seek nominations (including self-nominations) for each of them, due by July 15, 2014. Details for each of the prizes, including eligibility rules and past winners, can be found by following the links from http://www.informs.org/Community/Optimization-Society/Prizes.

Each of the four awards includes a cash amount of US$1,000 and a citation plaque. The award winners will be invited to give a presentation in a special session sponsored by the Optimization Society during the INFORMS annual meeting in San Franciso, CA in November 2014 (the winners will be responsible for their own travel expenses to the meeting). Award winners are also asked to contribute an article about their award-winning work to the annual Optimization Society newsletter.

Nominations, applications, and inquiries for each of the prizes should be made via email to the corresponding prize committee chair.

The **Khachiyan Prize** is awarded for outstanding lifetime contributions to the field of optimization by an individual or team. The topic of the contribution must belong to the field of optimization in its broadest sense. Recipients of the INFORMS John von Neumann Theory Prize or the MPS/SIAM Dantzig Prize in prior years are not eligible for the Khachiyan Prize. This year's Khachiyan Prize committee is:

- Tamás Terlaky (Chair) tat208@lehigh.edu
- Daniel Bienstock
- Immanuel Bomze
- John Birge

The **Farkas Prize** is awarded for outstanding contributions by a mid-career researcher to the field of optimization, over the course of their career. Such contributions could include papers (published or submitted and accepted), books, monographs, and software. The awardee will be within 25 years of their terminal degree as of January 1 of the year of the award. The prize may be awarded at most once in their lifetime to any person. This year's Farkas Prize committee is:

- Yinyu Ye (Chair) yyye@stanford.edu
- Garud N. Iyengar
- Jean B. Lasserre
- Zhi-Quan (Tom) Lou

The **Prize for Young Researchers** is awarded to one or more young researcher(s) for an outstanding paper in optimization. The paper must be published in, or submitted to and accepted by, a refereed professional journal within the four calendar years preceding the year of the award. All authors must have been awarded their terminal degree within eight calendar years preceding the year of award. The prize committee for this year's Prize for Young Researchers is as follows:

- Andrzej Ruszczyński (Chair) rusz@rutcor.rutgers.edu
- Katya Scheinberg
- Javier Peña
- Jean-Philippe Richard

The **Student Paper Prize** is awarded to one or more student(s) for an outstanding paper in optimization that is submitted to and received or published in a refereed professional journal within three calendar years preceding the year of the award. Every nominee/applicant must be a student on the first of January of the year of the award. All coauthor(s) not nominated for the award must send a letter indicating that the majority of the nominated work was performed by the nominee(s). The prize committee for this year's Student Paper Prize is as follows:

- Santanu Dey (Chair) santanu.dey@isye.gatech.edu
- Serhat Aybat
- Guzin Bayraksan
- Francois Margot

Nominations of Candidates for Society Officers Sought

We would like to thank three Society Vice-Chairs who will be completing their two-year terms at the conclusion of the INFORMS meeting: Leo Liberti, Andreas Wächter, and Andrew Schaefer. We are currently seeking nominations of candidates for the following positions:

- Vice-Chair for Global Optimization
- Vice-Chair for Nonlinear Optimization
- Vice-Chair for Optimization Under Uncertainty

Self nominations for all of these positions are encouraged.

According to Society Bylaws, "The main responsibility of the Vice Chairs will be to help INFORMS Local Organizing committees identify cluster chairs and/or session chairs for the annual meetings. In general, the Vice Chairs shall serve as the point of contact with their sub-disciplines." Vice Chairs shall serve two-year terms.

Please send your nominations or self-nominations to Jim Luedtke (jrluedt1@wisc.edu), including contact information for the nominee, by Saturday, June 30, 2014. Online elections will begin in mid- August, with new officers taking up their duties at the conclusion of the 2014 INFORMS annual meeting.

Seeking a Host for the 2016 INFORMS Optimization Society Conference

The INFORMS Optimization Society Conference is held in the early part of the even years, typically in a warm location. The most recent OS conference, held in 2014 at Rice University, was a great success, offering an opportunity for researchers studying optimization-related topics to share their work in a focused venue. The Optimization Society is currently seeking candidate locations to host the 2016 conference. If you are interested in helping to host the conference, please contact the current Optimization Society chair, Sanjay Mehrotra (mehrotra@northwestern.edu), or the chair-elect Suvrajeet Sen (sen@datadrivendecisions.org).