INFORMS OS Today

The Newsletter of the INFORMS Optimization Society

Volume 5 Number 1 May 2015

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

Suvrajeet Sen
University of Southern California
s.sen@usc.edu

Dear Fellow IOS Members:

It is my pleasure to assume the role of Chair of the IOS, one of the largest, and most scholarly groups within INFORMS. This year marks the Twentieth Anniversary of the formation of the INFORMS Optimization Section, which started with 50 signatures petitioning the Board to form a section. As it stands today, the INFORMS Optimization Society has matured to become one of the larger Societies within INFORMS. With seven active special interest groups, four highly coveted awards, and of course, its strong presence in every INFORMS Annual Meeting, this Society is on a roll. In large part, this vibrant community owes a lot to many of the past officer bearers, without whose hard work, this Society would not be as strong as it is today. For the current health of our Society, I wish to thank my predecessors, and in particular, my immediate predecessor Sanjay Mehrotra under whose leadership, we gathered the momentum for our conferences, a possible new journal, and several other initiatives. I also take this opportunity to express my gratitude to Jim Luedtke, Secretary/Treasurer, whose soft-spoken assurance, and attention to detail keeps the society humming. Another person whose support is instrumental to many communications is our Web Editor: Pietro Belotti; thanks Pietro.

The major agenda item in the upcoming years is the possibility of a new journal, sponsored by the Society. Because of the growth in a variety of applications in Communications and Control, Data Science, and Machine Learning, there has been an explosive growth in Optimization. This field has always attracted advanced applications such as Energy, Finance, Health Care and others. What is remarkable today is that as the area matures, many advanced applications are motivating new optimization concepts, including fundamental mathematical breakthroughs (e.g., the mathematics of risk), algorithmic speed-ups and robust software (e.g., convex, and mixed-integer optimization), as well as new modeling paradigms (e.g., multi-agent and data driven optimization). These are but a few of the promising directions today, and there are too many more to mention in this column.

Given the above backdrop, it is no surprise that our Society has shown overwhelming support for a new INFORMS journal devoted to Optimization. Exactly what the focus of such a journal should be is not clear at this point. There are other open questions, and we hope to resolve them as the year progresses. Our plan is to start a dialogue using INFORMS Connect as a platform for discussion, and we may even hold some meetings at affiliated conferences, such as at the International Symposium on Mathematical Programming in Pittsburgh this summer. Please be on the lookout for our calls at conferences of these affiliated groups.

Our Society is at the heart of INFORMS, and it draws tremendous strength from this sense of leadership. There is another source of strength: the caliber of scholars that the IOS attracts. In order to celebrate the accomplishments of our researchers, the IOS awards four highly coveted prizes: the Khachiyan Prize, the Farkas Prize, the Young-Researcher Prize, and the Student-paper Prize. This newsletter provides a glimpse of the personal stories of the award winners, and their unique perspectives on optimization. The awards committees for these prizes showed impeccable taste in their choices, and I wish to thank all members of these committees for their diligence with the process. The winners for the year 2014 Khachiyan Prize were Dimitris Bertsekas (MIT) and Clovis Gonzaga (Federal University of Santa Catarina, Florianópolis, Brazil). These two titans of optimization have made multi-faceted contributions.

- Bertsekas was awarded for his pioneering role in dynamic programming (with uncountable state spaces, approximate, neuro-dynamic and approximate dynamic programming), Lagrangean methods, dual-based bounds for nonconvex problems, network optimization, and applications in communications, power, transportation and others. Moreover his books on many of these topics have been adopted at many universities, world-wide.
- Gonzaga for his leadership in interior point methods (IPM), augmented Lagrangean methods, on filter methods for constrained non-linear optimization, and on fast steepest descent methods. He has recently developed steepest descent algorithms for the minimization of convex quadratic functions with the same performance bound as Krylov space methods. He is also the author of several influential reviews which have introduced IPMs to scores of young researchers.

The 2014 Farkas' Prize was awarded to Farid Alizadeh (Rutgers) who laid the groundwork for the field of semidefinite programming (SDP), which has grown quickly and steadily over the past 25 years. The citation continues the description of his contributions via connections between SDP and combinatorial optimization, complementarity, non-degeneracy and algorithmic methodology.

The 2014 Prize for Young Researchers was awarded to Jiawang Nie (UC - San Diego) for this paper "Optimality conditions and finite convergence of Lasserre's hierarchy," Mathematical Programming, Ser. A (2014) 146:97-121, DOI: 10.1007/s10107-013-0680-x. Prior to this proof of finite convergence, the existing theory could only establish the convergence of this family of relaxations in the limit. No proof of finite convergence had been previously found in spite of the significant numerical evidence for finite convergence that had been collected over a decade or so.

The winner of the 2014 Student Paper Prize was Damek Davis (UCLA) for his paper "Convergence rate analysis of several splitting schemes," arXiv preprint arXiv:1406.4834 (2014), with Wotao Yin.

This paper "makes a very significant contribution to the theory of optimization by using a rigorous and unified convergence rate analysis of important splitting schemes."

The awards committees for this year have been appointed, and they are listed elsewhere in this newsletter. I encourage the optimization community to identify strong candidates for these awards by sending forth nominations in each of the categories mentioned above.

Finally, I wish to welcome our newly elected Vice Chairs: Aida Khajavirad (Global Optimization), Warren Powell (Optimization Under Uncertainty), and Daniel Robinson (Nonlinear Optimization). They are joining the continuing team of Vice Chairs: Vladimir Boginski (Network Optimization),John Mitchell (Linear and Conic Optimization), Imre Polik (Computational Optimization and Software), and Juan Pablo Vielma (Integer and Discrete Optimization). These individuals are truly the reason for the strong presence of IOS at the INFORMS Annual Conferences, and I thank them for their leadership.

Before I close, I wish to also express my sincere appreciation for the work and leadership of Shabbir Ahmed. He has taken our newsletter to a whole new level, and provided it continuity for the past four years. Such dedication to the Society is what makes IOS tick. Thank you Shabbir!

I hope my enthusiasm for our society is palpable in this column, and hope that you will all get involved in the most vital society within INFORMS. Have a great summer, and see you all in either Pittsburgh or Philadelphia.

p.s: For those interested in reading about some new areas for INFORMS researchers, follow this link which needs you to login at INFORMS Connect:
`http://connect.informs.org/communities/community-home/librarydocuments/viewdocument/?DocumentKey=d7e454e3-1872-4826-900b-7871063a5980`

A Journey through Optimization

Dimitri P. Bertsekas
Lab. for Information and Decision Sciences
Massachusetts Institute of Tech., Cambridge, MA 02139
dimitrib@mit.edu

I feel honored and grateful for being awarded the Khachiyan Prize. It is customary in such circumstances to thank one's institutions, mentors, and collaborators, and I have many to thank. I was fortunate to be surrounded by first class students and colleagues, at high quality institutions, which gave me space and freedom to work in any direction I wished to go. It is also customary to chart one's intellectual roots and journey, and I will not depart from this tradition.

We commonly advise scholarly Ph.D. students in optimization to take the time to get a broad many-course education, with substantial mathematical content, and special depth in their research area. Then upon graduation, to use their Ph.D. research area as the basis and focus for further research, while gradually branching out into neighboring fields, and networking within the profession. This is good advice, which I often give, but this is not how it worked for me!

I came from Greece with an undergraduate degree in mechanical engineering, got my MS in control theory at George Washington University in three semesters, while holding a full-time job in an unrelated field, and finished two years later my Ph.D. thesis on control under set membership uncertainty at MIT in 1971. I benefited from the stimulating intellectual atmosphere of MIT's Electronic Systems Laboratory (later LIDS), nurtured by Mike Athans and Sanjoy Mitter, but because of my short stay there, I graduated with little knowledge beyond Kalman filtering and linear-quadratic control theory. Then I went to teach at Stanford in a department that combined mathematical engineering and operations research (in which my background was rather limited) with economics (in which I had no exposure at all). In my department there was little interest in control theory, and none at all in my thesis work. Never having completed a first course in analysis, my first assignment was to teach to unsuspecting students optimization by functional analytic methods from David Luenberger's wonderful book. The optimism and energy of youth carried me through, and I found inspiration in what I saw as an exquisite connection between elegant mathematics and interesting practical problems. Studying David Luenberger's other works (including his Nonlinear Programming book) and working next door to him had a lasting effect on me.

Two more formative experiences at Stanford were studying Terry Rockafellar's Convex Analysis book (and teaching a seminar course from it), and most importantly teaching a new course on dynamic programming, for which I studied Bellman's books in great detail. My department valued rigorous mathematical analysis that could be broadly applied, and provided a stimulating environment where both could thrive. Accordingly, my course aimed to combine Bellman's vision of wide practical applicability with the emerging mathematical theory of Markov Decision Processes. The course was an encouraging success at Stanford, and set me on a good track. It has survived to the present day at MIT, enriched by subsequent developments in theoretical and approximation methodologies.

After three years at Stanford, I taught for five years in the quiet and scholarly environment of the University of Illinois. There I finally had a chance to consolidate my mathematics and optimization background, through research to a great extent. In particular, it helped a lot that with the spirit of youth, I took the plunge into the world of the measure-theoretic foundations of stochastic optimal control, aiming to expand the pioneering Borel space framework of David Blackwell, in the company of my then Ph.D. student Steven Shreve.

I changed again direction by moving back to MIT in 1979, to work in the then emerging field of data networks and the related field of distributed computation. There I had the good fortune to meet two colleagues with whom I interacted closely over many years: Bob Gallager, who coauthored with me a book on data networks in the mid-80s, and John Tsitsiklis, who worked with me first while a doctoral student and then as a colleague, and over time coauthored with me two research monographs on distributed algorithms and neuro-dynamic programming, and a probability textbook. Working with Bob and John, and writing books with them was exciting and rewarding, and made MIT a special place for me.

Nonetheless, at the same time I was getting distracted by many side activities, such as books in nonlinear programming, dynamic programming, and network optimization, getting involved in applications of queueing theory and power systems, and personally writing several network optimization codes. By that time, however, I realized that simultaneous engagement in multiple, diverse, and frequently changing intellectual activities (while not recommended broadly) was a natural and exciting mode of operation that worked well for me, and also had some considerable benefits. It stimulated the cross-fertilization of ideas, and provided perspective for more broadly integrated courses and books. Over time I settled in a pattern of closely connected activities that reinforced each other: focus on an interesting subject to generate content and fill in missing pieces, teaching to provide a synthesis and to write class notes, and writing a book to understand the subject in depth and put together a story that would make sense to others. Writing books and looking for missing pieces was an effective way to generate new research, and working on a broad range of subjects was a good way to discover interconnections, broaden my perspective, and fight stagnation.

In retrospect I was very fortunate to get into methodologies that eventually prospered. Dynamic programming developed perhaps beyond Bellman's own expectation. He correctly emphasized the curse of dimensionality as a formidable impediment in its use, but probably could not have foreseen the transformational impact of the advances brought about by reinforcement learning, neuro-dynamic programming, and other approximation methodologies. When I got into convex analysis and optimization, it was an emerging theoretical subject, overshadowed by linear, nonlinear, and integer programming. Now, however, it has taken center stage thanks to the explosive growth of machine learning and large scale computation, and it has become the lynchpin that holds together most of the popular optimization methodologies. Data networks and distributed computation were thought promising when I got involved, but it was hard to imagine the profound impact they had on engineering, as well as the world around us today. Even set membership description of uncertainty, my Ph.D. thesis subject, which was totally overlooked for nearly fifteen years, eventually came to the mainstream, and has connected with the popular areas of robust optimization, robust control, and model predictive control. Was it good judgement or fortunate accident that steered me towards these fields? I honestly cannot say. Albert Einstein wisely told us that "Luck is when opportunity meets preparation." In my case, I also think it helped that I resisted overly lengthy distractions in practical directions that were too specialized, as well as in mathematical directions that had little visible connection to the practical world.

An academic journey must have companions to learn from and share with, and for me these were my students and collaborators. In fact it is hard to draw a distinction, because I always viewed my Ph.D. students as my collaborators. On more than one occasion, a Ph.D. thesis evolved into a book, as in the cases of Angelia Nedic and Asuman Ozdaglar, or into a long multi-year series of research papers after graduation, as in the cases of Paul Tseng and Janey Yu. My collaborators were many and I cannot mention them all, but they were special to me and I was fortunate to have met them. Thank you all for sharing this exciting journey with me.

Tamás Terlaky, Dimitri Bertsekas, Clóvis Gonzaga and Sanjay Mehrotra

Optimizing in the Third World

Clóvis C. Gonzaga
Department of Mathematics, Federal University of Santa Catarina, Florianópolis, SC, Brazil
ccgonzaga1@gmail.com

Introduction. In the year of my seventieth birthday I was amazed by being awarded the Khachiyan Prize, for which I thank the committee composed by Tamás Terlaky (chair), Daniel Bienstock, Immanuel Bomze and John Birge.

I was born in South Brazil, and raised in Joinville, Santa Catarina, a small industrial town, formerly a German settlement. I decided very soon to be an engineer (I had the knack). A local company instituted a prize: the student with best grades in Mathematics, Physics, Chemistry and Drawing during the three years of high school accepted in an Engineering school got a five year scholarship. I won this prize, and passed the entry exam in the best Engineering school in the Country, the Technological Institute of Aeronautics. ITA is a civilian school belonging to the Air Force, with courses in Electronic, Mechanical, Aeronautical and (nowadays) Computational Engineering. I studied Electronics, with emphasis in control and servomechanisms. ITA was founded in the fifties, organized by a team from MIT. We lived in the campus supported by the Air Force, in a very demanding system: with one grade D you are expelled, a grade C had to be replaced by B during the vacations (a maximum of five times). Plenty of jobs were offered to graduates.

I graduated in 1967, and went immediately to the graduate school of Engineering (COPPE) of the Federal University of Rio de Janeiro, for an MSc in Electrical Engineering. COPPE and the Catholic University of Rio de Janeiro were the first Brazilian graduate schools with MSc and DSc titles in the American system. It was founded in the sixties with a substantial support from the Brazilian government. We were pioneers, mostly isolated from the world, in a time when slide rules were still used, internet had not been imagined even by science fiction writers. The school had an IBM 1130 computer with

32 Kbytes of memory and no floating point processor, which allowed you to process your fortran program in punched cards once a day.

I was an MSc student with a full time job, taking care of a TR48 analog computer (young people do not know what that is), doing maintenance, operating and teaching analog computation. I started a research on adaptive control and read a text on identification of a control system, based on a method called Levenberg-Marquardt (!). I got hooked. By this time an IBM researcher and Berkeley part time professor, Jean-Paul Jacob, came to Brazil for a year. Jean-Paul had a PhD from UC Berkeley, advised by Lucien Polak, and gave us a short course on non-linear programming in 1969, the first to be taught in Brazil. I wrote a dissertation with him, on "reduction of optimal control problems to non-linear programming". We had by then a group of nine quite talented students. Jean-Paul proposed to the University to start a PhD program on Systems Engineering in Brazil, and brought Lotfi Zadeh from Berkeley to organize it. We made an agreement with UC, and American professors came to Brazil to teach us advanced courses. Most of us finished the PhD in Berkeley, but I refused to go abroad.

My doctoral research was done mostly alone. I worked on two unrelated topics.

Graph search methods. Stephen Coles from Stanford visited us and taught a course on artificial intelligence, with emphasis on graph search methods, especially the A* algorithm for computing minimum cost paths using look-ahead heuristics. I used search methods for the solution of sequential decision problems, for which at this time electrical engineers knew only dynamical programming. I devised an algorithm for planning the long term expansion of power transmission systems, using a partial ordering of nodes (a preference relation) for pruning less promising paths through a graph of system configurations. This was done in a contract with the Brazilian governmental power systems agency, using a half million dollars IBM 370 computer with 256K of memory. The program took two hours of CPU, and for several years it was used in the actual planning of the Brazilian interconnected system.

Convex optimization. My main interest was in continuous optimization. I was the first Brazilian to teach a full course on non-linear programming, based on the fresh from press book by Olvi Mangasarian. I read several texts on convex optimization and point to set mappings, by authors like Rockafellar, Berge, Bertsekas, Geoffrion. I especially liked the idea of epsilon-subgradients, and had the idea of using this concept to assemble subradients of a function and devise a general method for minimizing convex functions. I travelled to the US, where Jean-Paul scheduled interviews with two of my "idols", Polak in Berkeley and Geoffrion in UCLA. I told them about my ideas, and they asked the obvious question, about my adviser. I had none, and they agreed that the idea was great, but unattainable by a young man isolated in Brazil. Then Phillip Wolfe came to Brazil, and I was in charge of showing him some of Rio de Janeiro. I took him to the Sugarloaf, to Corcovado, to my home, and eventually told him about what I was doing. He said that he was doing precisely the same thing. I was delighted – Wolfe and I had the same ideas! In the following day I abandoned the topic. The end of this tale was also frustrating for Wolfe: Claude Lemarechal finished the bundle method before him, and certainly much before I would have done it, if ever.

I wrote a thesis on graph search methods for solving sequential decision problems, with application to the Brazilian transmission system. The results were nice, but I never tried to publish any journal paper. No tradition, no advisor, no need to publish.

Berkeley. In 1975 I went to Berkeley for a post-doc with my "grandfather" Lucien Polak, working on semi-infinite problems, using cutting sets for solving control systems design problems. We wrote a nice paper [10], and I had my first contact with a real research environment. In my first technical interview with him, he was in his office with another professor, unknown to me. Extremely nervous, stuttered "I read your paper with Prof. Maine, and... and...". "Did you find an error?" I said "y-y-yes." And he said "Show us. By the way, meet David Maine, from Imperial College". Resisting the temptation to faint, I went to the blackboard and showed the mistake. He asked if I knew how to do it right, I did, and this was the beginning of a life-time friendship.

Back to Brazil, I worked on the long term planning of the operation of hydro-thermal power systems, using stochastic dynamic programming, and on the beautiful theory of water value. I went to many congresses, met many researchers in discrete systems, combinatorics and in power planning, but again published no journal papers. The political situation in Brazil deteriorated, under a violent military dictatorship. Our salaries went down and research activities became irrelevant.

Interior points. In 1985 Lucien Polak invited me to replace him in UC during a sabbatical year. I got a visiting professor position, and stayed in Berkeley for two and a half years, teaching courses on automatic control. Karmarkar's paper had just appeared, and I got very interested. He had studied with Richard Karp, and was an expert in combinatorics, complexity theory and linear programming, with no background in NLP, I believe.

After much effort in rephrasing his quite obscure method, I saw that it could be reduced to a scaled steepest descent method applied to his potential function, and wrote a paper on this. Then I realized that he had not used the full potential of Newton's method, and that it should be possible to improve his result. I worked day and night, found dozens of ways of not improving his complexity results, got very depressed, went to a Chinese restaurant and the fortune cookie said "You will solve a problem that will be very important in your life." I had already decided for the logarithmic barrier function (instead of the potential function). I went home and started dealing with the third derivatives, leading to the large quadratic convergence region needed to build the path following method represented by the inchworm in the figure. Properties of path following methods had already been obtained by Megiddo and by Renegar. I presented my result in a workshop organized by Nimrod Megiddo in Asilomar, California, in the beginning of 1987, and the paper was published in the workshop proceedings. All references about these early results are in [4].

My life changed instantly. I met all the main brains in the field, and saw myself in the middle of the interior point revolution. A bunch of extremely clever guys like Anstreicher, Adler, Goldfarb, Kojima, McCormick, Megiddo, Mizuno, Monteiro, Nemirovsky, Nesterov, Renegar, Roos, Todd, Vial, Ye (forgive me if I forgot your name) travelled around the world like rodeo riders, mounting a new horse in each congress, with results that became obsolete in a matter of months. In 1992 I published the state of the art paper [4] in SIAM Review, later endowed with the "citation classics award" for the most cited paper in Mathematics and Computer Sciences written by a Brazilian in the 90's decade.

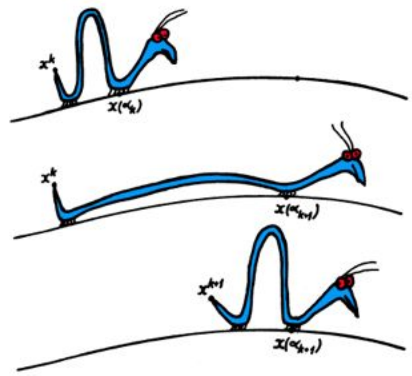


Figure 1: The inchworm

Back in Rio de Janeiro in 1987 I kept working on interior point methods, alone in Brazil, sending and receiving technical reports by regular mail and travelling frequently abroad. Brazil was again learning how to be a democracy. When computer communications arrived, science in the third world became feasible.

I spent 1993 at INRIA in Paris, working with Frédéric Bonnans [2, 5] and Jean-Charles Gilbert [3], worked for a semester in TU Delft, in the company of Roos and Terlaky.

In 1994 I moved back to South Brazil, transferred (with a group of DSc students) to the Math. Department of the Federal University of Santa Catarina in Florianópolis, a city on a beautiful island surrounded by 40 beaches, where we have organized two international workshops (clovisfest 60 and 70, in 2004 and 2014, with a plan for another one in 2024, maybe...). I worked on linear complementarity problems, augmented Lagrangian methods, Nesterov optimal descent methods, filter methods, and lately on the complexity of descent methods for quadratic minimization. I wrote papers with my former students Hugo Lara [13], Luis C. Matioli [14], Romulo Castillo [6], Ademir Ribeiro [12, 16], Diane Rossetto [8], Marcia Vanti [9], Roger Behling [1], Fernanda

Raupp [15], Marli Cardia, and have a long lasting collaboration with Elizabeth Karas [7, 8, 9] from the Federal University of Paraná in Brazil.

I am a member of the Brazilian Academy of Sciences and of TWAS. I am a SIAM fellow, I received the Great Cross of the Brazilian Order of Scientific Merit, and now this very prestigious prize in honour of Leonid Khachiyan.

Now I am retired, leisurely working in Florianópolis as a voluntary researcher. I thank my students, old and new [11], for the joy they brought to my life in these many years.

REFERENCES

[1] Roger Behling, C. C. Gonzaga, and G Haeser. Primal-dual relationship between levenberg-marquardt and central trajectories for linearly constrained convex optimization. *Journal of Optimization Theory and Applications*, 162:705–717, 2013.

[2] F. Bonnans and C. C. Gonzaga. Convergence of interior point algorithms for monotone linear complementarity problems. *Mathematics of Operations Research*, 21:1–25, 1996.

[3] J. C. Gilbert, C. C. Gonzaga, and E. Karas. Examples of ill-behaved central paths in convex optimization. *Mathematical Programming*, 103(1):63–94, 2005.

[4] C. C. Gonzaga. Path following methods for linear programming. *SIAM Review*, 34(2):167–227, 1992.

[5] C. C. Gonzaga and F Bonnans. Fast convergence of the simplified largest step path following algorithm. *Mathematical Programming*, 76:95–116, 1997.

[6] C. C. Gonzaga and R Castillo. A nonlinear programming algorithm based on non-coercive penalty functions. *Mathematical Programming*, 96:87–101, 2003.

[7] C. C. Gonzaga and E. W. Karas. Fine tuning Nesterov's steepest descent algorithm for differentiable convex programming. *Mathematical Programming*, 138(1–2):141–166, 2013.

[8] C. C. Gonzaga, E. W. Karas, and D. R. Rossetto. An optimal algorithm for constrained differentiable convex optimization. *SIAM Journal on Optimization*, 2013. To appear.

[9] C. C. Gonzaga, E. W. Karas, and M. Vanti. A globally convergent filter method for nonlinear programming. *SIAM J. Optimization*, 14(3):646–669, 2003.

[10] C. C. Gonzaga and E Polak. On constraint dropping schemes and optimality functions for a class of outer approximation algorithms. *SIAM Journal on Control and Optimization*, 17, 1979.

[11] C. C. Gonzaga and R. Schneider. Improving the efficiency of steepest descent algorithms for minimizing convex quadratics. Technical report, Federal University of Santa Catarina, Brazil, February 2015.

[12] E. W. Karas, C. C. Gonzaga, and A. A. Ribeiro. Local convergence of filter methods for nonlinear programming. *Optimization*, 59:1153–1171, 2010.

[13] H Lara and C. C. Gonzaga. A note on properties of condition numbers. *Linear Algebra and its Applications*, 261:269–273, 1997.

[14] L. C. Matioli and C. C. Gonzaga. A new family of penalties for augmented lagrangian methods. *Numerical Linear Algebra with Applications*, 15:925–944, 2008.

[15] F. M. Raupp and C. C. Gonzaga. A center cutting plane algorithm for a likelihood estimate problem. *Computational Optimization and Applications*, 21:277–300, 2000.

[16] A. A. Ribeiro, E. W. Karas, and C. C. Gonzaga. Global convergence of filter methods for nonlinear programming. *SIAM Journal on Optimization*, 19:1231 – 1249, 2008.

From Lovász θ Function and Karmarkar's Algorithm to Semidefinite Programming

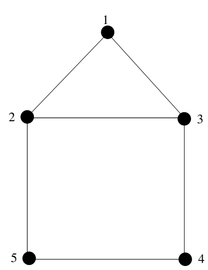
Farid Alizadeh
School of Business, Rutgers University,
Piscataway, NJ 08854
farid.alizadeh@rutgers.edu

It is truly an honor, and quite an unexpected one, to be awarded the Farkas Prize, especially considering the remarkable people who have received this award in the past.

Below I will give an account of how I was steered into the study of semidefinite programming from a brilliant construction by Lovász.

1. The marvelous θ function

While studying as a graduate masters student at the University of Nebraska–Lincoln in 1987, I came across the little book of Lovász [27] (see also the earlier [26]). In it there was a derivation of a graph invariant which was so simple and yet astonishingly beautiful. Here is the idea: Suppose a graph has a clique (as subset of vertices all connected to each other) of size k. Construct a symmetric matrix $A(\mathbf{x})$, where the i, j entry is 1 if vertices i and j are connected in the graph or $i = j$; otherwise set $[A(\mathbf{x})]_{ij} = [A(\mathbf{x})]_{ji} = x_{ij}$ a real-valued free variable. Consider this example:



The matrix of this graph as described above, and with $\mathbf{x} = \big(x_{14}, x_{15}, x_{24}, x_{35}\big)$ is given by:

$$A(\mathbf{x}) \overset{\text{def}}{=} \begin{pmatrix} 1 & 1 & 1 & x_{14} & x_{15} \\ 1 & 1 & 1 & x_{24} & 1 \\ 1 & 1 & 1 & 1 & x_{35} \\ x_{14} & x_{24} & 1 & 1 & 1 \\ x_{15} & 1 & x_{35} & 1 & 1 \end{pmatrix}.$$

From elementary linear algebra, we know that the largest eigenvalue of a principal submatrix is bounded by the largest eigenvalue of the matrix itself. So, since the submatrix corresponding to the vertices of the clique is a $k \times k$ matrix of all ones, and since the sole nonzero eigenvalue of such matrix is equal to k, it follows that $k \leq \lambda_{[1]}\big(A(\mathbf{x})\big)$. In particular, the size of the largest clique $\omega(G) \leq \min_{\mathbf{x}} \lambda_{[1]}\big(A(\mathbf{x})\big)$.

What is amazing is that the right hand side of the inequality, known as the *Lovász θ function*, is the solution of a *convex optimization* problem, even though there are no "neat" formulas for eigenvalues of a matrix. As a result, the θ invariant of a graph can be estimated in polynomial time by the *ellipsoid* method.

2. Karmarkar's algorithms

Completely independent of the θ function I also got interested in the (then) new Karmarkar's algorithm for linear programming (LP). While the ellipsoid method was known to solve LP's in polynomial time, it had not proved to be useful in practice. Karmarkar's algorithm was both polynomial-time, and also practical. I read Karmarkar's paper [24], and subsequently attended the SIAM Optimization Conference in Houston in 1987. Many major players in interior point methods, including Ilan Adler, Don Goldfarb, Clovis Gonzaga, Narendra Karmarkar, Michael J. Todd and Yinyu Ye were presenting talks.


Yinyu Ye, Farid Alizadeh and Sanjay Mehrotra

Right brain meets left brain: Can we extend Karmarkar's algorithm to compute the θ function?

Driving back from Houston to Lincoln, and juggling in my mind what I had learned in the SIAM conference, I thought of a crazy idea: Can we apply interior point methods to compute the θ function?

After transferring to Minnesota in 1988 to study with late Ben Rosen, I worked on this idea and realized that it was not so crazy after all. Applying the logarithmic barrier to the optimization problem characterizing the θ function we get:

$$\begin{aligned}
&\min_{z,\mathbf{x}} z - \mu \sum_i \ln\big(z - \lambda_i\big(A(\mathbf{x})\big)\big) \\
=&\min_{z,\mathbf{x}} z - \mu \ln \prod_i \Big(z - \lambda_i\big(A(\mathbf{x})\big)\Big) \\
=&\min_{z,\mathbf{x}} z - \mu \ln P_{A(\mathbf{x})}(z),
\end{aligned}$$

where $P_{A(\mathbf{x})}(z)$ is the characteristic polynomial of $A(\mathbf{x})$. Now, the last line can be easily computed using unconstrained optimization techniques based on Newton's, or quasi/truncated Newton methods. Using the array of supercomputers available at Minnesota, especially the Cray-2, and the NAG library optimization software, I could use the above formulation and compute θ of graphs with hundreds of vertices and a few thousand edges.

3. Onward to semidefinite programming

In the Spring of 1990 I got hold of the wonderful book by Grötschel, Lovász and Schrijver [21]. To my delight an entire chapter was devoted to the θ function. A particular "dual" formulation caught my eye:

$$\theta(G) = \begin{array}{ll} \max & J \bullet Y \\ \text{s.t.} & Y_{ij} = 0 \text{ for all } ij \text{ edges in } G \\ & Y \succcurlyeq 0 \end{array},$$

where $A \bullet B \stackrel{\text{def}}{=} \sum_{ij} A_{ij} B_{ij}$, and $Y \succcurlyeq 0$ means Y is a symmetric positive semidefinite matrix.

Looking at this characterization, I realized that it resembled the so-called "standard-form" LP: $\min\{\mathbf{c}^\top \mathbf{x} \mid A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$; only the nonnegative orthant was replaced by the positive semidefinite cone. Based on this observation I came up with a strategy to extend design and analysis of interior point methods:

1. First, I would focus my attention not on the θ function per se, but on a more general optimization problem and its dual:

   $$\begin{array}{llll} \min & C \bullet X & \max & \mathbf{b}^\top \mathbf{y} \\ \text{s.t.} & A_i \bullet X = b_i & \text{s.t.} & \sum_i y_i A_i + S = C \\ & X \succcurlyeq 0 & & S \succcurlyeq 0 \end{array}$$

   (This was the time I thought of the term *semidefinite programming (SDP)* for this problem[1]).

2. Next, I would take one of literally hundreds of papers on the complexity of various interior point methods for LP, and try to extend it, *word for word*, to SDP.

My intuition was based on the belief that Karmarkar's, and other interior point methods did not rely in significant ways on the combinatorial nature of linear programming (the way the simplex methods did, for instance). So techniques for their analysis should be extensible to more general setting above.

In the meantime, around May of 1990, Yinyu Ye, who was at the University of Iowa at the time, visited Minnesota. We had an hour long conversation and I discussed my project with him. Yinyu was very interested and encouraged me to pursue the project. He also suggested that I use one of the algorithms he had developed, which were based on the Todd-Ye-Mizuno potential function.

I spent the next few weeks studying Ye's two papers [48, 47] which presented two potential reduction algorithms for linear programming with iteration complexity $\mathcal{O}\big(\sqrt{n}L\big)$. While reading these papers, I realized that in almost every place that the vector $\mathbf{x}_k$ (the current estimate for the optimal solution) occurs as a parameter of a nonlinear function, that function is *symmetric* in $\mathbf{x}_k$. For instance, in

[1] The term "positive (semi)-definite programming" had been used much earlier by Dantzig and Cottle to describe a different problem: The linear complementarity problem with a semidefinite matrix[14]. I was not aware of this at the time, and in any case the term had not caught on.

the potential function given by $\mathbf{c}^\top\mathbf{x} - q\sum_i \ln x_i$, the term $\sum_i \ln x_i$ is symmetric. So are expressions like $\|\mathbf{x} - \mathbf{1}\|$, where $\mathbf{1}$ is the vector of all ones, and $\|\cdot\|$ corresponds to either the Euclidean norm, the maximum (∞) norm or the 1-norm. So it was straightforward to translate statements involving such functions to SDP. All I had to do was to replace $\mathbf{x}_k$ (in LP) with the *eigenvalues* of X_k (in SDP). For instance, an inequality used by Ye, and earlier by Karmarkar, was the following: For $\mathbf{0} < \mathbf{x} < \mathbf{1}$:

$$\sum_i \ln x_i \geq \sum_i x_i - n - \frac{\|\mathbf{x} - \mathbf{1}\|_2}{2\big(1 - \|\mathbf{x} - \mathbf{1}\|_\infty\big)}.$$

This inequality is easily proved using the Taylor expansion of $\ln(1-x)$. Now let X be a symmetric matrix and let $\lambda_i(X)$ be its eigenvalues. Then $0 < \lambda_i(X) < 1$ for all i is equivalent to $0 \prec X \prec I$. And the inequality above, upon replacing $\lambda_i(X)$ for x_i, becomes

$$\ln \operatorname{Det}(X) \geq \operatorname{trace}(X) - n - \frac{\|X - I\|_F}{2\big(1 - \rho(X - I)\big)},$$

where $\rho(A)$ is the spectral radius of A (the largest eigenvalue norm $\max_i |\lambda_i(X)|$ for symmetric matrices).

Another key ingredient of LP interior point methods was some form of linear or projective transformation that mapped the nonnegative orthant back to itself and brought the current interior feasible point $\mathbf{x}_k$ to $\mathbf{1}$. For instance, a linear transformation achieving this would be $\mathbf{x} \to \operatorname{Diag}(\mathbf{x}_k)^{-1}\mathbf{x}$. Typically a number of identities and inequalities are proved for this transformed vector, ultimately resulting in a fixed reduction in a potential function. The analogous transformation for SDP seemed to be $X \to X_k^{-1}X$ which maps the current interior feasible point X_k to I, the identity matrix. As in the inequality above, I could extend all the needed identities and inequalities involving $\operatorname{Diag}(\mathbf{x}_k)^{-1}\mathbf{x}$ in Ye's papers to the spectrum of $X_k^{-1}X$. As a result I could prove a fixed reduction to an analogous matrix potential function. So my strategy of "word-for-word" turning LP interior point algorithms and their analyses to SDP seemed to work.

Of course there was a major flaw in all this which made the whole thing somewhat absurd: While in LP the transformation $\mathbf{x} \to \operatorname{Diag}(\mathbf{x}_k)^{-1}\mathbf{x}$ mapped the nonnegative orthant back to itself, the corresponding $X \to X_k^{-1}X$ did not map the semidefinite cone to itself; the product of two symmetric matrices is not even a symmetric matrix in general.

I was stuck on this issue for a few weeks, until I decided to pay a visit to Yinyu and seek his advice. In summer of 1990, I drove five hours from Minneapolis to Iowa City. Yinyu was gracious to treat me to pizza, and even more gracious afterwards to listen to my construction and the dilemma about non-symmetry of $X_k^{-1}X$. As soon as I described the problem, and without hesitation, Yinyu proposed that I use instead the transformation: $X \to X_k^{-1/2}XX_k^{-1/2}$. The moment he uttered these words I knew that the problem was solved. $X \to X_k^{-1/2}XX_k^{-1/2}$ is an automorphism of the semidefinite cone. At the same time it has the same spectrum as $X_k^{-1}X$, so all the identities and inequalities I had already proved for its spectrum remained valid for this transformation as well.

I should mention here that Yinyu in his earlier visit to Minnesota also brought to my attention a preliminary version of Nesterov and Nemirovski's monumental work [32] (as an e-book on a floppy, in 1990!) I did not quite understand the book, but fortunately there was a workshop in early summer of 1990 in Madison, Wisconsin, in which both Yuri and Arkadi did also attend. I talked to them about my work on the θ function, and they re-assured me that their *self-concordance theory* implied that θ can be estimated in $\mathcal{O}\big(\sqrt{n}\big)$ iterations. Based on this discussion I quickly wrote a paper [2] on randomized parallel computation of θ, and became confident that my "word-for-word" approach had a strong chance of success.

4. Further research on applications of SDP

After writing my dissertation [1] and a few papers based on the research above [3, 4], I spent two years (1992-1994) as an NSF postdoctoral associate at the International Computer Science Institute at the University of California–Berkeley. I was extremely privileged to work with Dick Karp on applications of combinatorial optimization to molecular biology. I should also mention my joyful collaboration with

Andrew Goldberg in implementing the push-relabel method for the maximum-flow problem on the CM-2 architecture [6].

At this time I also got to know Ilan Adler and through interactions with him got interested both in primal-dual algorithms for SDP and also in second order cone programming. My interest in SOCP arose after reading a preliminary version of a paper of Nemirovski and Scheinberg [31] which essentially made a "word-for-word" extension of Karmarkar's algorithm from LP to SOCP.

In 1994 I moved to RUTCOR at Rutgers University. Below I briefly outline my activities in the past twenty years:

1994-1998: Primal-Dual algorithms I started my collaboration with Jean-Pierre Haeberly of (then) Fordham, and Michael Overton of New York University. After studying the primal-dual method of Helmberg et al. [23] we developed our own method, which computed Newton's direction for primal and dual feasibility and the complementarity relation $\frac{XS+SX}{2} = \mu I$ [11]; this direction was termed by the community as the AHO direction. We also developed *SDPpack*, the first publicly available software package that could solve optimization problems with any combination of linear, second order and semidefinite constraints [8]. Later, in a seminal work, Monteiro and Zhang extended the AHO method by adding an automorphic *scaling* of the semidefinite cone [30]. Renato Monteiro showed that under all such scalings, a "short-step" approach yields $\mathcal{O}\big(\sqrt{n}|\ln\epsilon|\big)$ iteration complexity [29]. The Monteiro-Zhang family includes the fundamental Nesterov-Todd methods [33, 34] and the Helmberg et al. method mentioned above.

1995-1997: Degeneracy in SDP With Jean-Pierre and Michael, we also studied the notion of (non)degeneracy and strict complementarity in SDP [10]. This work was further expanded by Pataki [43], and Faybusovich [17].

1998-2003: Jordan algebras Through the work of Güler [22], Faybusovich [17, 16, 18], and brilliant texts of Faraut and Korány [15] and Koecher [28] Euclidean Jordan algebras became accessible to the optimization community. With Stefan Schmieta we showed that just about anything said about the Monteiro-Zhang family can be extended verbatim to optimization over symmetric cones [44, 45]. We also showed that the word-for-word extension of Ye's algorithm in LP can be extended all the way to symmetric cones [12]. With Yu Xia we extended the Q method—developed earlier with Haeberly and Overton [9]—to symmetric cones as well [13].

2008-2014: nonnegative and SOS polynomials As a result of the work of N. Z. Shor [46], and later Lassere [25] Parrilo [42, 41] and Nesterov [35], it became evident that the set of polynomials which are expressible as sum-of-squares (SOS) of other polynomials, is SD-representable. With David Papp, we studied the notion of sum-of-squares in *abstract algebras* and, building on the work of Nesterov [35] and Faybusovich [19] showed SDP-representability of SOS cones in the most general setting possible [40]. With RUTCOR students Gabor Rudolf, Nilay Noyan, David Papp, we showed that the complementarity conditions for the cone of univariate nonnegative polynomials and its dual, the moment cone, cannot be expressed by more than four *independent* bilinear relations (unlike nonnegative orthant, and other symmetric cones where complementarity could be expressed by n bilinear relations). In the process we introduced the notion of *bilinearity rank* of cones. Later, Gowda and Tao showed that bilinearity rank of a cone is exactly the dimension of the Lie algebra of its automorphism group [20].

2003-2011: Shape-constrained statistical estimation With my colleague at Rutgers, Jonathan Eckstein, and former students Gabor Rudolf and Nilay Noyan, we applied SDP to estimate the (nonnegative) arrival rate of nonhomogeneous Poisson processes, and used our approach to estimate arrival rate of e-mails [5]. And with David Papp we estimated the two dimensional arrival rate of accidents in New Jersey Turnpike as a function of both time and millage post [39]. Also with David we used similar techniques for shape constrained regression and probability density estimation [40].

5. Final word

I am fortunate to have been associated with great mentors, colleagues, and students throughout the years. Early in my career, my advisor at Nebraska, David Klarner, sparked the love of algorithms in me. And Ben Rosen at Minnesota had the confidence to agree to support me after just a ten minute conversation.

Michael Overton, who had already extensively studied eigenvalue optimization [36, 38, 37], did substantial hand-holding in the early stages of my investigation. I also had stimulating discussions both in person and by e-mail with Arkadi Nemirovski and Yuri Nesterov (who read the first drafts of my dissertation and pointed out several sloppy errors). Laci Lovász was gracious and encouraged me to pursue my ideas about the θ function and its efficient computation. Gene Golub was kind in allowing me to use Stanford student offices, and also having many discussions about numerical stability of eigenvalue optimization algorithms.

Over the years discussions with Stephen Boyd, Gabor Pataki, Mutakuri Ramanna, Michel Goemans, David Williamson and Garud Iyengar were very fruitful. In addition, I had a great sabbatical leave at Columbia IEOR department during calendar year of 2001. A result of this visit was a very satisfying collaboration with Don Goldfarb which resulted in our paper on SOCP [7].

Finally, I should acknowledge the great pleasure I have had to work with (former) graduate students Reuben Sattergren, Stefan Schmieta, Yu Xia, Gabor Rudolf, Nilay Noyan, David Papp, and RUTCOR and Rutgers MSIS (current) graduate students Marta Cavaleiro, Mohammad Ranjbar and Deniz Seyed Eskandani.

REFERENCES

[1] F. Alizadeh. *Combinatorial Optimization with Interior Point Methods and Semi-Definite Matrices.* PhD thesis, Computer Science Department, University of Minnesota, Minneapolis, Minnesota, 1991.

[2] F. Alizadeh. A sublinear-time randomized parallel algorithm for the maximum clique problem in perfect graphs. In *Proc. 2nd ACM-SIAM Symposium on Discrete Algorithms*, 1991.

[3] F. Alizadeh. Optimization Over Positive Semi-Definite Cone; Interior-Point Methods and Combinatorial Applications. In P. Pardalos, editor, *Advances in Optimization and Parallel Computing*. North-Holland, Amsterdam, 1992.

[4] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM J. Optim.*, 5(1):13–51, 1995.

[5] F. Alizadeh, J. Eckstein, N. Noyan, and G. Rudolf. Arrival Rate Approximation by Nonnegative Cubic Splines . *Operations Research*, 56(1):140–156, 2008.

[6] F. Alizadeh and A. V. Goldberg. Implementing the Push-Relabel Method for the Maximum Flow Problem on a Connection Machine. In David S. Johnson and Catherine C. McGeoch, editors, *Network Flows and Matching, Fisrt DIMACS Implementation Challenge*, volume 12 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, 1993.

[7] F. Alizadeh and D. Goldfarb. Second-Order Cone Programming. *Math. Programming Series B*, 95:3–51, 2003.

[8] F. Alizadeh, J.-P.A. Haeberly, V. Nayakkankuppam, M.L. Overton, and S.A. Schmieta. SDPpack User Guide (Version 0.9 Beta). Technical Report 737, Courant Institute of Mathematical Sciences, New York University, New York, NY, 1997. (Available at `http://www.cs.nyu.edu/faculty/overton/sdppack`).

[9] F. Alizadeh, J.P. Haeberly, and M.L. Overton. A new primal-dual interior point method for semidefinite programming. In *Proc. 5th SIAM Conference on Appl. Linear Algebra*, Snowbird, Utah, 1994.

[10] F. Alizadeh, J.P. Haeberly, and M.L. Overton. Complementarity and nondegeneracy in semidefinite programming. *Math. Programming*, 77(2), 1997.

[11] F. Alizadeh, J.P. Haeberly, and M.L. Overton. Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results. *SIAM J. Optim.*, 8(3):746–768, 1998.

[12] F. Alizadeh and S.H. Schmieta. Symmetric Cones, Potential Reduction Methods and Word-By-Word Extensions. In R. Saigal, L. Vandenberghe, and H. Wolkowicz, editors, *Handbook of Semidefinite Programming, Theory, Algorithms and Applications*, pages 195–233. Kluwer Academic Publishers, 2000.

[13] F. Alizadeh and Y. Xia. The Q Method for Symmetric Cone Programming. *Journal of Optimzation Theorey and Applications*, 149(1):102–137, 2010.

[14] G. B. Danztig and R. W. Cottle. Positive (semi-)definite programming. In J. Abadie, editor, *Nonlinear Programming*, pages 55–73. North Holland, Amsterdam, 1967.

[15] J. Faraut and A. Korányi. *Analysis on Symmetric Cones*. Oxford University Press, Oxford, UK, 1994.

[16] L. Faybusovich. Euclidean Jordan algebras and interior-point algorithms. *Positivity*, 1(4):331–357, 1997.

[17] L. Faybusovich. Linear systems in Jordan algebras and primal-dual interior point algorithms. *Journal of Computational and Applied Mathematics*, 86:149–175, 1997.

[18] L. Faybusovich. A Jordan-algebraic approach to potential-reduction algorithms. *Math. Z.*, 239:117–129, 2002.

[19] L. Faybusovich. On nesterov's approach to semi-infinite programming. *Acta. Appl. Math.*, 74(2):195–215, 2002.

[20] S. Gowda and J. Tao. On the bilinearity rank of a proper cone and lyapunov-like transformations. *Math. Programming Series A*, 147:155–170, 2014.

[21] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer Verlag, 1988.

[22] O. Güler. Barrier Functions in Interior Point Methods. *Math. of Oper. Res.*, 21:860–885, 1996.

[23] C. Helmberg, F. Rendl, R.J. Vanderbei, and H. Wolkowicz. An interior-point method for semidefinite programming. *SIAM J. Optim.*, 6:342–361, 1996.

[24] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[25] Lasserre, J. B. Global Optimization with Polynomials and The Problem of Moments. *SIAM J. Optim.*, 11(8):708–817, 2001.

[26] L. Lovász. On the shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(1):1–7, January 1979.

[27] L. Lovász. *An Algorithmic Theory of Numbers, Graphs and Convexity*. CBMS-NSF50. SIAM, 1986.

[28] M. Koecher. *The Minnesota Notes on Jordan Algebras and Their Applications*. Springer-Verlag, 1999. Edited by Kreig, Aloys and Walcher, Sebastian based on Lectures given in The University of Minnesota, 1960.

[29] R.D.C. Monteiro. Polynomial convergence of primal-dual algorithms for semidefinite programming based on monteiro and zhang family of directions. *SIAM J. Optim.*, 8:797–812, 1998.

[30] R.D.C. Monteiro and Y. Zhang. A unified analysis for a class of path-following primal-dual interior-point algorithms for semidefinite programming. *Mathematical Programming*, 81:281–299, 1998.

[31] A. Nemirovski and K. Scheinberg. Extension of Karmarkar's algorithm onto convex quadratically constrained quadratic programming. *Math. Programming*, 72:273–289, 1996.

[32] Y. Nesterov and A. Nemirovski. Self-concordant functions and polynomial time methods in convex programming. Central Economical and Mathematical Institute, U.S.S.R Academy of Science, Moscow, 1990.

[33] Y.E. Nesterov and M.J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. of Oper. Res.*, 22:1–42, 1997.

[34] Y.E. Nesterov and M.J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. Optim.*, 8:324–364, 1998.

[35] Yu. Nesterov. Squared functional systems and optimization problems. In H. Frenk, K. Roos, T. Terlaky, and S. Zhang, editors, *High Performance Optimization*, Appl. Optim., pages 405–440. Kluwer, Dordrecht, Netherlands, 2000.

[36] M. L. Overton. On minimizing the maximum eigenvalue of a symmetric matrix. *SIAM J. Matrix Anal. Appl.*, 9(2):256–268, 1988.

[37] M. L. Overton and R. S. Womersley. On the sum of the largest eigenvalues of a symmetric matrix. *SIAM J. Matrix Anal. Appl.*, 13:41–45, 1992.

[38] M. L. Overton and R. S. Womersley. Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. *Math. Programming*, 62(2):321–357, 1993.

[39] D. Papp and F. Alizadeh. Multivariate Arrival Rate Estimation by Sum-Of-Squares Polynomial Splines and Semidefinite Programming. In *Proc. 2011 Winter Simulation Conf.*, 2011.

[40] D. Papp and F. Alizadeh. Shape-Constrained Estimation Using Nonnegative Splines. *Comput. Graph. Statist.*, 23(1):211–231, 2014.

[41] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96(2):293–320, 2003.

[42] Parrilo, P. A. *Structured Semidefinite Programs and semialgebraic Methods*. PhD thesis, The Californai Inst. of Technology, 2000.

[43] G. Pataki. On the rank of extreme matrices in semidefinite programming and the multiplicity of optimal eigenvalues. *Math. of Oper. Res.*, 23(2):339–358, 1998.

[44] S.H. Schmieta and F. Alizadeh. Associative and Jordan Algebras, and Polynomial Time Interior-Point Algorithms for Symmetric Cones. *Math. of Oper. Res.*, 26(3):543–564, 2001.

[45] S.H. Schmieta and F. Alizadeh. Extension of Commutative Class of Primal-Dual Interior Point Algorithms to Symmetric Cones. *Math. Programming*, 96:409–438, 2003.

[46] N. Z. Shor. A class of global minimum bounds of polynomial functions. *Cybernatics*, 23(6):731–734, 1987.

[47] Y. Ye. A class of projective transformations for linear programming. *SIAM J. Comput.*, 19(3):457–466, 1990.

[48] Y. Ye. An $O(n^3L)$ Potential Reduction Algorithm for Linear Programming. *Math. Programming*, 50(2):239–258, 1991.

Local Versus Global Conditions in Polynomial Optimization

Jiawang Nie
Department of Mathematics
University of California San Diego,
La Jolla, CA 92093, USA
njw@math.ucsd.edu

This paper briefly reviews the relationship between local and global optimality conditions in [15]. Consider the polynomial optimization problem

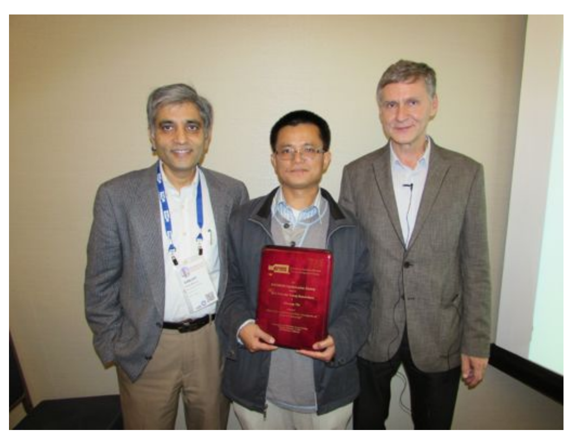
$$\left\{\begin{array}{ll} \min & f(x) \\ s.t. & h_i(x) = 0 \, (i = 1, \ldots, m_1), \\ & g_j(x) \geq 0 \, (j = 1, \ldots, m_2), \end{array}\right. \tag{1}$$

where $f, h_1, \ldots, h_{m_1}, g_1, \ldots, g_{m_2}$ are real polynomials in $x := (x_1, \ldots, x_n)$. For convenience, denote

$$h := (h_1, \ldots, h_{m_1}), \quad g := (g_1, \ldots, g_{m_2})$$

and $g_0 := 1$. Let K be the feasible set of (1). When there are no equality (resp., inequality) constraints, the tuple $h = \emptyset$ and $m_1 = 0$ (resp., $g = \emptyset$ and $m_2 = 0$).

The problem (1) can be treated as a general nonlinear program. By classical nonlinear optimization methods, we can typically get a Karush-Kuhn-Tucker (KKT) point of (1). Theoretically, it is NP-hard to check whether a KKT point is a local minimizer or not. However, it is often not too hard to



Sanjay Mehrotra, Jiawang Nie and Andrzej Ruszczyński do that in practice. This is because there exist standard conditions ensuring local optimality. On the other hand, it is often much harder to get a global minimizer. In practice, sometimes we may be able to get a global optimizer, but it is typically hard to verify the global optimality. A major reason for this is lack of easily checkable global optimality conditions in nonlinear programming theory.

Local and global optimality conditions are presumably very different, except special cases like convex optimization. For general nonconvex optimization, little is known about global conditions. However, for polynomial optimization, this is possible by using representations of nonnegative polynomials. Interestingly, global optimality conditions are closely related to the local ones, which was discovered in the paper [15].

1. Local Optimality Conditions

Let u be a local minimizer of (1) and

$$J(u) \;:=\; \{j_1,\ldots,j_r\}$$

be the index set of active inequality constraints. If the *constraint qualification condition (CQC)* holds at u, i.e., the gradient vectors

$$\nabla h_1(u),\ldots,\nabla h_{m_1}(u),\nabla g_{m_1}(u),\ldots,\nabla g_{j_r}(u)$$

are linearly independent, then there exist Lagrange multipliers $\lambda_1,\ldots,\lambda_{m_1}$ and $\mu_1,\ldots,\mu_{m_2}$ satisfying

$$\nabla f(u)=\sum_{i=1}^{m_1}\lambda_i\nabla h_i(u)+\sum_{j=1}^{m_2}\mu_j\nabla g_j(u), \qquad (2)$$

$$\left.\begin{array}{c}\mu_1 g_1(u)=\cdots=\mu_{m_2}g_{m_2}(u)=0,\\ \mu_1\geq 0,\ldots,\mu_{m_2}\geq 0.\end{array}\right\} \qquad (3)$$

The equation (2) is called the *first order optimality condition (FOOC)*, and (3) is called the *complementarity condition*. If it further holds that

$$\mu_1+g_1(u)>0,\ldots,\mu_{m_2}+g_{m_2}(u)>0, \qquad (4)$$

then the *strict complementarity condition (SCC)* holds at u. The strict complementarity is equivalent to $\mu_j>0$ for every $j\in J(u)$. Let $L(x)$ be the Lagrange function

$$L(x):=f(x)-\sum_{i=1}^{m_1}\lambda_i h_i(x)-\sum_{j\in J(u)}\mu_j g_j(x).$$

Clearly, (2) implies the gradient $\nabla_x L(u)=0$. The polynomials f,h_i,g_j are smooth functions. Thus, under the constraint qualification condition, the *second order necessity condition (SONC)* holds:

$$v^T\nabla_x^2 L(u)v\geq 0 \qquad \forall\, v\in G(u)^{\perp}. \qquad (5)$$

In the above, $G(u)$ denotes the Jacobian of the active constraining polynomials

$$G(u)=\text{Jacobian}\Big(h_1,\ldots,h_{m_1},g_{j_1},\ldots,g_{j_r}\Big)\Big|_{x=u}$$

and $G(u)^{\perp}$ denotes the null space of $G(u)$. If it holds that

$$v^T\nabla_x^2 L(u)v>0 \quad \text{for all } \ 0\neq v\in G(u)^{\perp}, \qquad (6)$$

then the *second order sufficiency condition (SOSC)* holds at u. The relations among the above conditions can be summarized as follows: if CQC holds at u, then (2), (3) and (5) are necessary conditions for u to be a local minimizer, but they may not be sufficient; if (2), (3), (4) and (6) hold at $u\in K$, then u is a strict local minimizer of (1). We refer to [1, Section 3.3] for such classical results.

Mathematically, CQC, SCC and SOSC are sufficient for local optimality, but may not be necessary. However, for *generic* cases, they are sufficient and necessary conditions. This is a major conclusion of [15]. Denote by $\mathbb{R}[x]_d$ the set of real polynomials in x and with degrees at most d. Let $[m]:=\{1,\ldots,m\}$. The following theorem is from [15].

Theorem 1. *Let $d_0,d_1,\ldots,d_{m_1},d_1',\ldots,d_{m_2}'$ be positive integers. Then there exist a finite set of nonzero polynomials $\varphi_1,\ldots,\varphi_L$, which are in the coefficients of polynomials $f\in\mathbb{R}[x]_{d_0}$, $h_i\in\mathbb{R}[x]_{d_i}$ $(i\in[m_1])$, $g_j\in\mathbb{R}[x]_{d_j'}$ $(j\in[m_2])$ such that if*

$$\begin{array}{c}\varphi_1(f,h_1,\ldots,h_{m_1},g_1,\ldots,g_{m_2})\neq 0,\\ \vdots\\ \varphi_L(f,h_1,\ldots,h_{m_1},g_1,\ldots,g_{m_2})\neq 0,\end{array}$$

then CQC, SCC and SOSC hold at every local minimizer of (1).

Theorem 1 implies that the local conditions CQC, SCC and SOSC hold at every local minimizer in the space of input polynomials with given degrees, except a union of finitely many hypersurfaces. So, they hold in an open dense set in the space of input polynomials. Therefore, CQC, SCC and SOSC can be used as sufficient and necessary conditions in checking local optimality, for *almost all* polynomial optimization problems. This fact was observed in nonlinear programming.

2. A global optimality condition

Let u be a feasible point for (1). By the definition, u is a global minimizer if and only if

$$f(x) - f(u) \geq 0 \quad \forall\, x \in K. \tag{7}$$

Typically, it is quite difficult to check (7) directly. In practice, people are interested in easily checkable conditions ensuring (7). For polynomial optimization, this is possible by using sum-of-squares type representations.

Let $\mathbb{R}[x]$ be the ring of real polynomials in $x := (x_1, \ldots, x_n)$. A polynomial $p \in \mathbb{R}[x]$ is said to be *sum-of-squares* (SOS) if $p = p_1^2 + \cdots + p_k^2$ for $p_1, \ldots, p_k \in \mathbb{R}[x]$. A sufficient condition for (7) is that there exist polynomials $\phi_1, \ldots, \phi_{m_1} \in \mathbb{R}[x]$ and SOS polynomials $\sigma_0, \sigma_1, \ldots, \sigma_{m_2} \in \mathbb{R}[x]$ such that

$$f(x) - f(u) = \sum_{i=1}^{m_1} \phi_i(x) h_i(x) + \sum_{j=0}^{m_2} \sigma_j(x) g_j(x). \tag{8}$$

The equality in (8) is a polynomial identity in the variables of x. Note that for every feasible point x in (1), the right hand side in (8) is always nonnegative. This is why (8) ensures that u is a global minimizer. The condition (8) was investigated by Lasserre [6]. It was a major tool for solving the optimization problem (1) globally. We call (8) a global optimality condition for (1).

People wonder when the global optimality condition holds. The representation of $f(x) - f(u)$ in (8) was motivated by Putinar's Positivstellensatz [16], which gives SOS type certificates for positive or nonnegative polynomials on the set K. Denote

$$\langle h \rangle := h_1 \mathbb{R}[x] + \cdots + h_{m_1} \mathbb{R}[x],$$

which is the ideal generated by the polynomial tuple h. Let $\Sigma[x]$ be the set of all SOS polynomials in $\mathbb{R}[x]$. The polynomial tuple g generates the quadratic module:

$$Q(g) := \Sigma[x] + g_1 \Sigma[x] + \cdots + g_{m_2} \Sigma[x].$$

If there exists a polynomial $p \in \langle h \rangle + Q(g)$ such that the set $\{x \in \mathbb{R}^n : p(x) \geq 0\}$ is compact, then $\langle h \rangle + Q(g)$ is said to be *archimedean*. The archimedeanness of $\langle h \rangle + Q(g)$ implies the compactness of K, while the reverse is not necessary. However, when K is compact, we can always add a redundant condition like $R - \|x\|_2^2 \geq 0$ to the tuple g so that $\langle h \rangle + Q(g)$ is archimedean. Hence, archimedeanness of $\langle h \rangle + Q(g)$ is almost equivalent to the compactness of K. Putinar's Positivstellensatz [16] says that if $\langle h \rangle + Q(g)$ is archimedean, then every polynomial which is strictly positive on K belongs to $\langle h \rangle + Q(g)$ (cf. [16]).

The global optimality condition (8) is equivalent to the membership

$$f(x) - f(u) \in \langle h \rangle + Q(g).$$

When u is a global minimizer of (1), the polynomial

$$\tilde{f}(x) := f(x) - f(u)$$

is nonnegative on K, but not strictly positive on K. This is because u is always a zero point of $\tilde{f}$ on K. So, Putinar's Positivstellensatz itself does not imply the global optimality condition (8). Indeed, there are counterexamples that (8) may not hold. For instance, when f is the Motzkin polynomial $x_1^2 x_2^2 (x_1^2 + x_2^2 - 3x_3^2) + x_3^6$ and K is the unit ball, then (8) fails to hold.

However, the global optimality condition (8) holds for *almost all* polynomials f, h_i, g_j, i.e., it holds in an open dense set in the space of input polynomials. This is a major conclusion of [15]. The ideal $\langle h \rangle$ is said to be *real* if every polynomial in $\mathbb{R}[x]$ vanishing on the set $\{x \in \mathbb{R}^n : h(x) = 0\}$ belongs to $\langle h \rangle$ (cf. [2]). This is a general condition. For instance, if $\langle h \rangle$ is a prime ideal and h has a nonsingular real zero, then $\langle h \rangle$ is real (cf. [2]). As pointed out earlier, when the feasible set K is compact, we can generally assume that $\langle h \rangle + Q(g)$ is archimedean. Interestingly, the local conditions CQC, SCC and SOSC imply the global optimality condition (8), under the archimedeanness of $\langle h \rangle + Q(g)$. The following theorem is a consequence of the results in [15].

Theorem 2. *Assume that the ideal $\langle h \rangle$ is real and the set $\langle h \rangle + Q(g)$ is archimedean. If the constraint qualification condition, strict complementarity condition, and second order sufficiency condition hold* *at every global minimizer of (1), then the global optimality condition (8) holds.*

Proof. At every global minimizer u of f on K, the CQC, SCC and SOSC conditions implies that the boundary hessian condition holds at u, by Theroem 3.1 of [15]. The boundary hessian condition was introduced by Marshall [11] (see Condition 2.3 of [15]). Let f_{min} be the global minimum value of (1). Denote $V = \{x \in \mathbb{R}^n : h(x) = 0\}$. Let $I(V)$ be the set of all polynomials vanishing on V. By Theorem 9.5.3 of [10] (also see Theorem 2.4 of [15]), we have

$$f(x) - f_{min} \in I(V) + Q(g).$$

Because $\langle h \rangle$ is real, $\langle h \rangle = I(V)$ and

$$f(x) - f(u) \in \langle h \rangle + Q(g).$$

So, the global optimality condition (8) holds. □

By Theorem 1, the local conditions CQC, SCC and SOSC hold generically, i.e., in an open dense set in the space of input polynomials. Therefore, the global optimality condition (8) also holds generically, when $\langle h \rangle$ is real and $\langle h \rangle + Q(g)$ is archimedean.

3. Lasserre's hierarchy

Lasserre [6] introduced a sequence of semidefinite relaxations for solving (1) globally, which is now called *Lasserre's hierarchy* in the literature. It can be desribed in two equivalent versions. One version uses SOS type representations, while the other one uses moment and localizing matrices. They are dual to each other, as shown in [6]. For convenience of description, we present the SOS version here. For each $k \in \mathbb{N}$ (the set of nonnegative integers), denote the sets of polynomials (note $g_0 = 1$)

$$\langle h \rangle_{2k} := \left\{ \sum_{i=1}^{m_1} \phi_i h_i \;\middle|\; \begin{array}{c} \text{each } \phi_i \in \mathbb{R}[x] \\ \text{and } \deg(\phi_i h_i) \le 2k \end{array} \right\},$$

$$Q_k(g) := \left\{ \sum_{j=0}^{m_2} \sigma_j g_j \;\middle|\; \begin{array}{c} \text{each } \sigma_j \in \Sigma[x] \\ \text{and } \deg(\sigma_j g_j) \le 2k \end{array} \right\}.$$

Note that $\langle h \rangle_{2k}$ is a truncation of $\langle h \rangle$ and $Q_k(g)$ is a truncation of $Q(g)$. The SOS version of Lasserre's hierarchy is the sequence of relaxations

$$\max \quad \gamma \quad \text{s.t.} \quad f - \gamma \in \langle h \rangle_{2k} + Q_k(g) \tag{9}$$

for $k = 1, 2, \ldots,$. The problem (9) is equivalent to a *semidefinite program* (SDP). So it can be solved as an SDP by numerical methods. For instance, the software `GloptiPoly 3` [3] and `SeDuMi` [18] can be used to solve it. We refer to [7, 9, 10] for recent work in polynomial optimization.

Let f_{min} be the minimum value of (1) and f_k denote the optimal value of (9). It was shown that (cf. [6])

$$\cdots \le f_k \le f_{k+1} \le \cdots \le f_{min}.$$

When $\langle h \rangle + Q(g)$ is archimedean, Lasserre [6] proved the asymptotic convergence

$$f_k \to f_{min} \quad \text{as} \quad k \to \infty.$$

If $f_k = f_{min}$ for some k, Lasserre's hierarchy is said to have *finite convergence*. It is possible that the sequence $\{f_k\}$ has only asymptotic, but not finite, convergence. For instance, this is the case when f is the Motzkin polynomial $x_1^2 x_2^2(x_1^2 + x_2^2 - 3x_3^2) + x_3^6$ and K is the unit ball [12, Example 5.3]. Indeed, such f always exists whenever $\dim(K) \ge 3$, which can be implied by [17, Prop. 6.1]. However, such cases do not happen very much. Lasserre's hierarchy often has finite convergence in practice, which was demonstrated by extensive numerical experiments in polynomial optimization (cf. [4, 5]).

A major conclusion of [15] is that Lasserre's hierarchy almost always has finite convergence. Specifically, it was shown that Lasserre's hierarchy has finite convergence when the local conditions CQC, SCC and SOSC are satisfied, under the archimedeanness. The following theorem is shown in [15].

Theorem 3. *Assume that $\langle h \rangle + Q(g)$ is archimedean. If the constraint qualification, strict complementarity and second order sufficiency conditions hold at every global minimizer of (1), then Lasserre's hierarchy of (9) has finite convergence.*

By Theorem 1, the local conditions CQC, SCC and SOSC at every local minimizer, in an open dense set in the space of input polynomials. This implies that, under the archimedeanness of $\langle h \rangle + Q(g)$, Lasserre's hierarchy has finite convergence, in an open dense set in the space of input polynomials. That is, Lasserre's hierarchies almost always (i.e., generically) have finite convergence. This is a major conclusion of [15].

If one of the assumptions in Theorem 3 does not hold, then $\{f_k\}$ may fail to have finite convergence. The counterexamples were shown in §3 of [15]. On the other hand, there exists other non-generic conditions than ensures finite convergence of $\{f_k\}$. For instance, if h has finitely many real or complex zeros, then $\{f_k\}$ has finite convergence (cf. [8, 14]).

Since the minimum value f_{min} is typically not known, a practical concern is how to check $f_k = f_{min}$ in computation. This issue was addressed in [13]. Flat truncation is generally a sufficient and necessary condition for checking finite convergence.

For non-generic polynomial optimization problems, it is possible that the sequence $\{f_k\}$ does not have finite convergence to f_{min}. People are interested in methods that have finite convergence for minimizing *all* polynomials over a given set K. The Jacobian SDP relaxation proposed in [12] can be applied for this purpose. It gives a sequence of lower bounds that have finite converge to f_{min}, for every polynomial f that has a global minimizer over a general set K.

Acknowledgement The research was partially supported by the NSF grants DMS-0844775 and DMS-1417985.

REFERENCES

[1] D. Bertsekas. *Nonlinear Programming*, second edition. Athena Scientific, 1995.

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

[3] D. Henrion, J. Lasserre and J. Loefberg. *GloptiPoly 3: moments, optimization and semidefinite programming.* `http://homepages.laas.fr/henrion/software/gloptipoly3/`

[4] D. Henrion and J.B. Lasserre. GloptiPoly : Global Optimization over Polynomials with Matlab and SeDuMi. *ACM Trans. Math. Soft.* 29, pp. 165–194, 2003.

[5] D. Henrion and J.B. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. *Positive polynomials in control*, 293310, Lecture Notes in Control and Inform. Sci., 312, Springer, Berlin, 2005.

[6] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3): 796-817, 2001.

[7] J.B. Lasserre. *Moments, Positive Polynomials and Their Applications*, Imperial College Press, 2009.

[8] M. Laurent. Semidefinite representations for finite varieties. *Mathematical Programming*, Vol. 109, pp. 1–26, 2007.

[9] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. *Emerging Applications of Algebraic Geometry, Vol. 149 of IMA Volumes in Mathematics and its Applications* (Eds. M. Putinar and S. Sullivant), Springer, pages 157-270, 2009.

[10] M. Marshall. *Positive Polynomials and Sums of Squares.* Mathematical Surveys and Monographs, 146. American Mathematical Society, Providence, RI, 2008.

[11] M. Marshall. Representation of non-negative polynomials, degree bounds and applications to optimization. *Canad. J. Math.*, 61 (2009), pp. 205–221.

[12] J. Nie. An exact Jacobian SDP relaxation for polynomial optimization. *Mathematical Programming*, Ser. A, Vol. 137, pp. 225–255, 2013.

[13] J. Nie. Certifying convergence of Lasserre's hierarchy via flat truncation. *Mathematical Programming*, Ser. A, Vol 142, No. 1-2, pp. 485-510, 2013.

[14] J. Nie. Polynomial optimization with real varieties. *SIAM Journal On Optimization*, Vol 23, No.3, pp. 1634-1646, 2013.

[15] J. Nie. Optimality Conditions and Finite Convergence of Lasserre's Hierarchy Mathematical Programming, Ser. A, Vol 146, No. 1-2, pp. 97-121, 2014.

[16] M. Putinar. Positive polynomials on compact semialgebraic sets, *Ind. Univ. Math. J.* 42 (1993), 969–984.

[17] C. Scheiderer. Sums of squares of regular functions on real algebraic varieties. *Trans. Am. Math. Soc.*, 352, 1039-1069 (1999).

[18] J.F. Sturm. SeDuMi 1.02: a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11&12 (1999), 625-653. `http://sedumi.ie.lehigh.edu`

Convergence Rate Analysis of Several Splitting Schemes

Damek Davis
Department of Mathematics
University of California, Los Angeles, CA
damek@math.ucla.edu

Our paper [12] is concerned with the following unconstrained minimization problem:

$$\underset{x\in\mathcal{H}}{\text{minimize}}\, f(x)+g(x) \qquad (1)$$

where $\mathcal{H}$ is a Hilbert space and $f, g : \mathcal{H} \to (-\infty, \infty]$ are closed, proper, and convex functions. Through duality, the results generalize to the problem:

$$\underset{x\in\mathcal{H}_1, y\in\mathcal{H}_2}{\text{minimize}}\, f(x)+g(y) \quad \text{subject to } Ax+By=0, \qquad (2)$$

where $A : \mathcal{H}_1 \to \mathcal{G}$ and $B : \mathcal{H}_2 \to \mathcal{G}$ are linear mappings and $\mathcal{H}_1, \mathcal{H}_2$, and $\mathcal{G}$ are Hilbert spaces. At first this decomposition of the objective into two functions may seem superfluous, but it captures the motivation of operator-splitting methods: the whole convex problem may be as complicated as you wish, but it can be decomposed into two or more (possibly) simpler terms.

The above decomposition suggests that we try to minimize the complicated objective $f+g$ by alternatively solving a series of subproblems related to minimizing f and g separately. The following strongly convex optimization problem is one of the main subproblems we use:

$$\mathbf{prox}_f(x) := \text{argmin}_{y\in\mathcal{H}}\, f(y) + \frac{1}{2}\|y-x\|^2,$$

where the input x will encode some information about $f+g$, so the subproblem will not just minimize f alone. It turns out that many functions arising in machine learning, signal processing, and imaging have simple or closed form proximal operators [7, 20, 4, 16], which is why operator-splitting algorithms are so effective for large-scale problems.

Since von Neumann devised the method of alternating projections [24], several operator splitting methods have been proposed [18, 21, 23] to solve a variety of decomposable problems [6, 14, 8, 9, 25, 5]. Although splitting methods are observed to perform well in practice, their convergence rates were unknown for many years; this is a big limitation for practitioners. Without convergence rates, practitioners cannot predict practical performance, compare competing methods without implementing both, or understand why methods perform well on nice problems but are slow on other problems of the same size. The award winning manuscript [12], and follow up work [13, 10, 11] resolves these issues for a variety of splitting algorithms.

Analyzing convergence rates of splitting methods is challenging because the classical objective error analysis breaks down. Here, the objective no longer monotonically decreases; in some cases, they can be completely useless. For example, finding a point in the intersection of sets C_1 and C_2 can be formulated as Problem (1) with $f = \iota_{C_1}$ and $g = \iota_{C_2}$. Notice that the objective function $f+g$ is finite (and equal to 0) only at a solution $x \in C_1 \cap C_2$. The Douglas-Rachford splitting method (which is a special case of the Peaceman-Rachford splitting method (see Equation 3)) produces a sequence of points converging to a solution. In general, all points in the sequence have an infinite objective value, but the limit point has a zero objective value! Thus, we must take a different approach to analyze operator-splitting methods: we first prove convergence rates for the fixed-point residual associated to a nonexpansive mapping, which implies bounds for certain subgradients that we use to bound the objective function.

Sanjay Mehrotra, Damek Davis and Serhat Aybat

A sketch of the upper complexity proof.

A key property used in our analysis is the *firm nonexpansiveness* of $\mathbf{prox}_f$, which means that $2\mathbf{prox}_f - I_{\mathcal{H}}$ is nonexpansive, i.e., 1-Lipschitz continuous [1, Proposition 4.2]. Because the composition of two nonexpansive operators remains nonexpansive, this property implies that the *Peaceman-Rachford operator* $T_{\mathrm{PRS}} := (2\mathbf{prox}_f - I) \circ (2\mathbf{prox}_g - I)$ is nonexpansive [18]. Thus, the Krasnosel'skiĭ-Mann (KM) [17, 19] iteration with damping by $\lambda \in (0,1)$

$$z^{k+1} = (1-\lambda)z^k + \lambda T_{\mathrm{PRS}}(z^k) \quad k = 0, 1, \ldots \tag{3}$$

will weakly converge to a fixed point of T_{PRS} whenever one exists [1, Theorem 5.14]. We are interested in fixed-points z^* of T_{PRS} because they immediately yield minimizers through the relation $\mathbf{prox}_g(z^*) \in \operatorname{argmin}(f+g)$ [1, Theorem 25.6]. Furthermore, the sequence $(\mathbf{prox}_g(z^j))_{j\geq 0}$ weakly converges to a minimizer of $f+g$ when one exists [22].

For any nonexpansive mapping $T : \mathcal{H} \to \mathcal{H}$, the fixed-point residual (FPR) $\|Tz - z\|$ measures how "close" z is to being a fixed point. A key result in our paper establishes the convergence rate of this quantity:

Theorem 4. *Let $\lambda \in (0,1)$. For any nonexpansive map $T : \mathcal{H} \to \mathcal{H}$ with nonempty fixed-point set, we have*

$$\|Tz^k - z^k\| = o\left(\frac{1}{\sqrt{k+1}}\right)$$

where $z^0 \in \mathcal{H}$ and for all $k \geq 0$, $z^{k+1} = (1-\lambda)z^k + \lambda T z^k$.

Roughly, this theorem follows by showing $(\|z^{j+1} - z^j\|^2)_{j\geq 0}$ is monotonic and summable. The convergence then follows from the following simple lemma:

Lemma 5. *Let $(a_j)_{j\geq 0} \subseteq \mathbf{R}_{++}$ be a monotonic down and summable sequence. Then $a_k = o(1/(k+1))$.*

Proof. $(k+1)a_{2k} \leq a_{2k} + \cdots + a_k \overset{k\to\infty}{\to} 0$. □

We already mentioned that the convergence rate of the fixed-point residual would play a role in bounding subgradients associated to the relaxed PRS algorithm. Using some basic convex analysis, Figure 1 "expands" a single iteration of Algorithm 3 into a series of "subgradient steps."

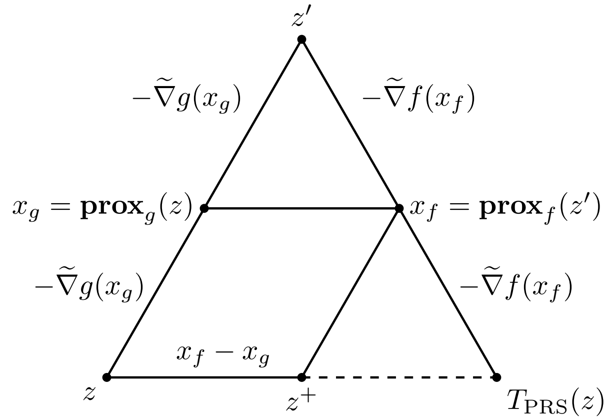


Figure 1: A single iteration of the relaxed PRS algorithm. Note that $z^+ = (1/2)z + (1/2)T_{\mathrm{PRS}}(z)$ and $\widetilde{\nabla} g(x_g) \in \partial g(x_g)$ and $\widetilde{\nabla} f(x_f) \in \partial f(x_f)$.

A few identities immediately follow from Figure 1:

$$z^+ - z = x_f - x_g = -(\widetilde{\nabla} g(x_g) + \widetilde{\nabla} f(x_f)) \tag{4}$$

Note that whenever $T_{\mathrm{PRS}}(z) = z$, we have $x^* := x_f = x_g$ and $\widetilde{\nabla} g(x^*) + \widetilde{\nabla} f(x^*) = 0$, so $x^* = \mathbf{prox}_g(z)$ is optimal for Problem 1. With this identity and Theorem 4, we get the following result:

Theorem 6. *Suppose that $(z^j)_{j\geq 0}$ is generated by Algorithm 3. For all $k \geq 0$, with $x_g^k, \widetilde{\nabla} g(x_g^k), x_f^k$, and $\widetilde{\nabla} f(x_f^k)$ defined as in Figure 3, we have*

$$\|x_f^k - x_g^k\|^2 = \|\widetilde{\nabla} g(x_g^k) + \widetilde{\nabla} f(x_f^k)\|^2 = o\left(\frac{1}{k+1}\right).$$

By applying the basic inequality $h(y) \geq h(x) + \langle \widetilde{\nabla} h(x), y - x\rangle$ on f and g, we obtain the next theorem.

Theorem 7. *Suppose that $(z^j)_{j\geq 0}$ is generated by Algorithm (3). Let x^* be a minimizer of $f+g$. For all $k \geq 0$, let x_g^k and x_f^k be defined as in Figure 3. Then*

$$|f(x_f^k) + g(x_g^k) - f(x^*) - g(x^*)| = o\left(\frac{1}{\sqrt{k+1}}\right).$$

Furthermore, when f is Lipschitz continuous, we have

$$0 \leq (f+g)(x_g^k) - (f+g)(x^*) = o\left(\frac{1}{\sqrt{k+1}}\right).$$

Theorem 7 shows that the relaxed PRS algorithm is strictly faster than the slowest performing first-order (subgradient) methods. Splitting methods are known to perform well in practice, so it is natural to suspect that this is not the best we can do. The following section shows that this intuition is partially true.

Acceleration through averaging.

Can we improve upon the worst-case $o(1/\sqrt{k+1})$ complexity in Theorem 7? The following example gives us some insight into how we might improve.

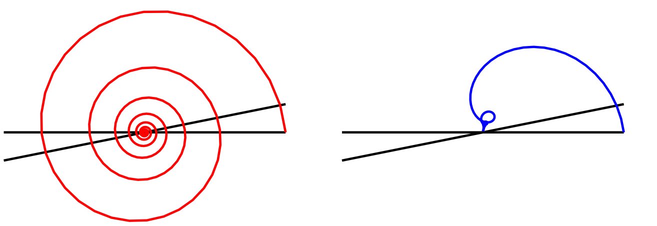


Figure 2: The relaxed PRS algorithm ($\lambda = 1/2$) applied to the intersection of two lines L_1 and L_2 (the horizontal axis and the line through the origin with slope .2). Formally, $g = \iota_{L_1}$ and $f = \iota_{L_2}$ are the indicator functions of L_1 and L_2. The left side (red) is the actual trajectory of the DRS algorithm (i.e., the sequence $(z^j)_{j\geq 0}$). The right side (blue) is obtained through averaging $\overline{z}^k := (1/(k+1))\sum_{i=0}^k z^i$.

Figure 2 replaces the k-th iterate z^k generated by Algorithm (3) with the averaged, or *ergodic*, iterate $\overline{z}^k := (1/(k+1))\sum_{i=0}^k z^i$. It is clear that, at least initially, the ergodic iterate outperforms the standard, or *nonergodic*, iterate because of cancellation induced from averaging the spiraling sequence. In actuality, the nonergodic iterate eventually surpasses the nonergodic iterate and converges linearly with rate ≈ 0.9806 [2], but this example gives us an idea of why averaging can help.

Inspired by Figure 2, we introduced the following averaging scheme for relaxed-PRS to obtain a faster $O(1/(k+1))$ worst-case convergence rate.

Theorem 8. *Suppose that $(z^j)_{j\geq 0}$ is generated by Algorithm (3). Let x^* be a minimizer of $f+g$. For all $k \geq 0$, let x_g^k and x_f^k be defined as in Figure 3. Furthermore, let $\overline{x}_g^k := (1/(k+1))\sum_{i=0}^k x_g^i$ and $\overline{x}_f^k := (1/(k+1))\sum_{i=0}^k x_f^i$. Then*

$$|f(\overline{x}_f^k) + g(\overline{x}_g^k) - f(x^*) - g(x^*)| = O\left(\frac{1}{k+1}\right).$$

Furthermore, when f is Lipschitz continuous, we have

$$0 \leq (f+g)(\overline{x}_g^k) - (f+g)(x^*) = O\left(\frac{1}{k+1}\right).$$

The rate $O(1/(k+1))$ is a big improvement over the $o(1/\sqrt{k+1})$ rate from Theorem 7. However, it is not clear that averaging will always result in faster practical performance. Indeed, consider the following basis pursuit problem:

$$\underset{x\in\mathbf{R}^d}{\text{minimize}}\ \|x\|_1 \quad \text{subject to: } Ax = b \tag{5}$$

where $A \in \mathbf{R}^{m\times d}$ is a matrix and $b \in \mathbf{R}^m$ is a vector. We can apply the relaxed PRS algorithm to this problem by setting $g(x) = \|x\|_1$ and $f(x) = \iota_{\{y|Ay=b\}}(x)$ where ι_C denotes the $\{0,\infty\}$-valued indicator function of a set C.

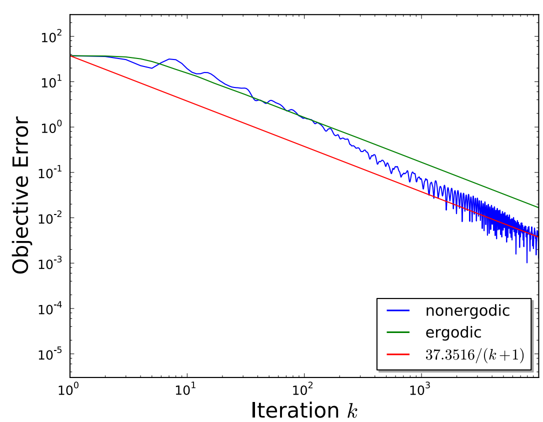


Figure 3: The objective error of the basis pursuit problem (5)

Figure 3 shows a numerical experiment on the basis-pursuit problem. Three things are clear: (i) the nonergodic iterate is highly oscillatory as suggested early; (ii) the performance of the nonergodic iterate eventually surpasses the ergodic iterate; (iii) the ergodic iterate eventually converges with rate exactly $O(1/(k+1))$ and no faster. This situation is typical, and suggests that averaging is not necessarily a good idea in all scenarios. Nevertheless obtaining $\overline{x}_g^k$ is a nearly costless procedure, so it never hurts to average.

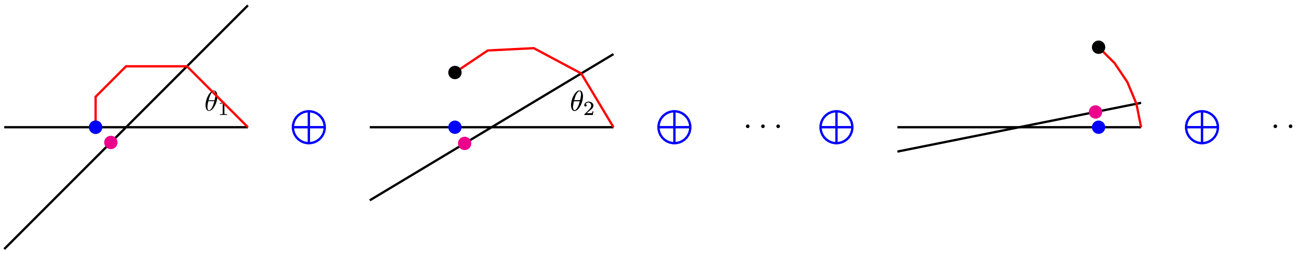


Figure 4: This graph shows 5 iterations of relaxed PRS applied to an infinite direct sum of examples similar to the one shown in Figure 2. The angles $(\theta_j)_{j\geq 0}$ are converging to 0. The black and blue dots coincide in the first component. The infinite vectors of black (on tip of spiral), blue (on horizontal axis), and magenta (on non horizontal line) dots are the vectors z^5, x_g^5, and the projection of x_g^5 onto the non horizontal line, respectively. The objective functions in Theorem 9 are as follows: g is the indicator of the horizontal axis and f is the distance function to the non horizontal line. Thus, $g(x_g^5) = 0$ and $f(x_g^5)$ is the distance between the blue and magenta dots.

The worst-case nonergodic rate is sharp.

Although averaging can improve the rate of convergence of relaxed PRS, we observed in Figure 3 that the nonergodic iterate can perform better than the ergodic iterate. Does this mean that our convergence rate in Theorem 7 can be improved? It turns out that the answer is no, at least in infinite dimensional problems:

Theorem 9. *There is a Hilbert space $\mathcal{H}$ and two closed, proper, and convex functions $f, g : \mathcal{H} \to (-\infty, \infty]$ such that f is Lipschitz continuous and the following holds: for all $\varepsilon > 0$, there is an initial point $z^0 \in \mathcal{H}$ and a stepsize $\gamma \in \mathbf{R}_{++}$ with*

$$(f+g)(x_g^k) - (f+g)(x^*) = \Omega\left(\frac{1}{(k+1)^{1/2+\varepsilon}}\right)$$

where $(z^j)_{j\geq 0}$ is generated by Algorithm (3) with $\lambda = 1/2$ and x^* is the unique minimizer of $f+g$.

The example that proves the sharpness of our $o(1/\sqrt{k+1})$ rate is shown in Figure 4. The proof of Theorem 9 is technical and involves estimating an oscillatory integral, so we cannot get into the details in this brief article.

Follow up work.

In a later paper [13], we show that relaxed PRS and ADMM automatically adapt to the regularity of Problems 1 and 2. For example, we show that for all minimizers x^* of $f+g$, the "best" objective error satisfies

$$\min_{i=0,\cdots,k}\left((f+g)(x_g^i) - (f+g)(x^*)\right) = o\left(\frac{1}{k+1}\right)$$

whenever f is differentiable and ∇f is Lipschitz continuous. Thus, relaxed PRS obtains the same rate of convergence as the forward-backward splitting algorithm (also called ISTA) [3], which requires knowledge of the Lipschitz constant of ∇f to ensure convergence (or a suitable line search procedure). The paper also includes general conditions for linear convergence, linear convergence of feasibility problems with nice intersection, and convergence rates under strong convexity assumptions. Finally, the papers [10, 11] prove convergence rates for a wide class of algorithms that are significantly more powerful than relaxed PRS and ADMM.

Conclusion.

The results of our paper show that the relaxed PRS method is nearly as slow as the subgradient method (Theorem 9). This is unexpected because splitting methods tend to perform well in practice. We can get around this issue by computing the ergodic sequence (Theorem 8) and testing whether it has a smaller objective value than the nonergodic sequence. This procedure is essentially costless, but it tends to perform worse in practice (Figure 3). In the paper, we include several other complexity results including the first example of arbitrarily slow convergence for the Douglas-Rachford splitting method. We also analyze extensions of Problem (1) and get similar results for the alternating direction method of multipliers (ADMM) [15].

Acknowledgements.

I would like to say thank you to the prize committee for this honor. I am also grateful to Professor Wotao Yin for taking me on as a student and collaborating with me on this project.

REFERENCES

[1] Heinz H. Bauschke and Patrick L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011.

[2] Heinz H. Bauschke, J.Y. Bello Cruz, Tran T.A. Nghia, Hung M. Phan, and Xianfu Wang. The rate of linear convergence of the Douglas-Rachford algorithm for subspaces is the cosine of the Friedrichs angle. *Journal of Approximation Theory*, 185(0):63 – 79, 2014.

[3] Amir Beck and Marc Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[4] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[5] Luis M. Briceño-Arias and Patrick L. Combettes. A Monotone+Skew Splitting Model for Composite Monotone Inclusions in Duality. *SIAM Journal on Optimization*, 21(4):1230–1250, 2011.

[6] Antonin Chambolle and Thomas Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[7] Patrick L. Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer Optimization and Its Applications, pages 185–212. Springer New York, 2011.

[8] Patrick L. Combettes and Jean-Christophe Pesquet. Primal-Dual Splitting Algorithm for Solving Inclusions with Mixtures of Composite, Lipschitzian, and Parallel-Sum Type Monotone Operators. *Set-Valued and Variational Analysis*, 20(2):307–330, 2012.

[9] Laurent Condat. A Primal-Dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms. *Journal of Optimization Theory and Applications*, 158(2):460–479, 2013.

[10] Damek Davis. Convergence rate analysis of primal-dual splitting schemes. *arXiv preprint arXiv:1408.4419v2*, 2014.

[11] Damek Davis. Convergence rate analysis of the forward-Douglas-Rachford splitting scheme. *arXiv preprint arXiv:1410.2654v3*, 2014.

[12] Damek Davis and Wotao Yin. Convergence rate analysis of several splitting schemes. *arXiv preprint arXiv:1406.4834v2*, 2014.

[13] Damek Davis and Wotao Yin. Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions. *arXiv preprint arXiv:1407.5210v2*, 2014.

[14] Ernie Esser, Xiaoqun Zhang, and Tony Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.

[15] Roland Glowinski and Américo Marrocco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet nonlinéaires. *Rev. Francaise dAut. Inf. Rech. Oper*, R-2:41–76, 1975.

[16] Thomas Goldstein and Stanley Osher. The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[17] Mark Aleksandrovich Krasnosel'skii. Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk*, 10(1):123–127, 1955.

[18] Pierre-Louis Lions and Bertrand Mercier. Splitting Algorithms for the Sum of Two Nonlinear Operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[19] W. Robert Mann. Mean Value Methods in Iteration. *Proceedings of the American Mathematical Society*, 4(3):pp. 506–510, 1953.

[20] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, 2014.

[21] Gregory B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *Journal of Mathematical Analysis and Applications*, 72(2):383 – 390, 1979.

[22] Benar F. Svaiter. On Weak Convergence of the Douglas-Rachford Method. *SIAM Journal on Control and Optimization*, 49(1):280–287, 2011.

[23] Paul Tseng. A Modified Forward-Backward Splitting Method for Maximal Monotone Mappings. *SIAM Journal on Control and Optimization*, 38(2):431–446, 2000.

[24] John von Neumann. *Functional operators: The geometry of orthogonal spaces*. Princeton University Press, 1951. (Reprint of mimeographed lecture notes first distributed in 1933.).

[25] Băng Công Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, 38(3):667–681, 2013.

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, 2015. 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 Philadelphia, PA in November 2015 (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. The prize committee for this year's Khachiyan Prize is as follows:

- Ilan Adler (Chair)
  adler@ieor.berkeley.edu
- Michael Ball
- Don Goldfarb
- Werner Römisch

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. The prize committee for this year's Farkas Prize is as follows:

- Ariela Sofer (Chair)
  asofer@gmu.edu
- Jon Lee
- Sanjay Mehrotra
- Zelda Zabinski

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:

- Nick Sahinidis (Chair)
  sahinidis@cmu.edu
- Dan Bienstock
- Sam Burer
- Andrew Schaefer

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:

- Mohit Tawarmalani (Chair)
  mtawarma@purdue.edu
- Fatma Kilinç-Karzan
- Warren Powell
- Uday Shanbhag

Nominations of Candidates for Society Officers Sought

Sanjay Mehtotra will complete his term as Most-Recent Past-Chair of the Society at the conclusion of the 2015 INFORMS annual meeting. Suvrajeet Sen is continuing as Chair through 2016. Jim Luedtke will also complete his term as Secretary/Treasurer at the conclusion of the INFORMS meeting. The Society is indebted to Sanjay and Jim for their work.

We would also like to thank four Society Vice-Chairs who will be completing their two-year terms at the conclusion of the INFORMS meeting: Imre Pólik, Juan Pablo Vielma, John Mitchell, and Vladimir Boginski.

Finally, we thank the Newsletter Editor, Shabbir Ahmed, who is also stepping down this year.

We are currently seeking nominations of candidates for the following positions:

- Chair-Elect
- Secretary/Treasurer
- Vice-Chair for Computational Optimization and Software
- Vice-Chair for Integer and Discrete Optimization
- Vice-Chair for Linear and Conic Optimization
- Vice-Chair for Network Optimization
- Newsletter Editor

Self nominations for all of these positions are encouraged.

To ensure a smooth transition of the chairmanship of the Society, the Chair-Elect serves a one-year term before assuming a two-year position as Chair; thus this is a three-year commitment. As stated in the Society Bylaws, "The Chair shall be the chief administrative officer of the OS and shall be responsible for the development and execution of the Society's program. He/she shall (a) call and organize meetings of the OS, (b) appoint ad hoc committees as required, (c) appoint chairs and members of standing committees, (d) manage the affairs of the OS between meetings, and (e) preside at OS Council meetings and Society membership meetings."

The Secretary/Treasurer serves a two-year term. According to Society Bylaws, "The Secretary/Treasurer shall conduct the correspondence of the OS, keep the minutes and records of the Society, maintain contact with INFORMS, receive reports of activities from those Society Committees that may be established, conduct the election of officers and Members of Council for the OS, make arrangements for the regular meetings of the Council and the membership meetings of the OS. As treasurer, he/she shall also be responsible for disbursement of the Society funds as directed by the OS Council, prepare and distribute reports of the financial condition of the OS, help prepare the annual budget of the Society for submission to INFORMS. It will be the responsibility of the outgoing Secretary/Treasurer to make arrangements for the orderly transfer of all the Society's records to the person succeeding him/her." The Secretary/Treasurer is allowed to serve at most two consecutive terms; as Jim Luedtke is now completing his second term, the OS must elect a new Secretary/Treasurer this cycle.

Vice-Chairs also serve a two-year term. 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."

The Newsletter Editor is appointed by the Society Chair. The Newsletter Editor is responsible for editing this newsletter, which is produced in the spring each year. Although this is not an elected position, nominations for this position are being sought at this time for consideration by the chair.

Please send your nominations or self-nominations to Jim Luedtke (jrluedt1@wisc.edu), including contact information for the nominee, by June 1, 2015. Online elections will begin in mid- June, with new officers taking up their duties at the conclusion of the 2015 INFORMS annual meeting.

Optimization groups at Princeton and Rutgers will be hosting the 2016 IOS Conference, and Warren Powell will serve as the General Chair. Stay tuned for more information.