

INFORMS OS Today

The Newsletter of the INFORMS Optimization Society

Volume 6 Number 1

May 2016

Contents

<i>Chair's Column</i>	1
<i>The IOS 2016 Optimization Conference</i>	36
<i>Nominations for OS Prizes</i>	38
<i>Nominations for OS Officers</i>	39
<i>Host Sought for the 2018 OS Conference</i>	39

Featured Articles

Thanks!

Jean-Bernard Lasserre	3
-----------------------------	---

From Linear to Nonlinear Integer Optimization

Robert Weismantel	5
-------------------------	---

Disjunctive Conic Sets, Conic Minimal Inequalities, and Cut-Generating Functions

Fatma Kilinç-Karzan	13
---------------------------	----

Semidefinite Relaxations for Nonlinear Optimization Problems

Somayeh Sojoudi and Javad Lavaei	20
--	----

A New Perspective on Boosting in Linear Regression via Subgradient Optimization and Relatives

Paul Grigas	29
-------------------	----

Chair's Column

Suvrajeet Sen

Industrial and Systems Engineering
University of Southern California
USA
s.sen@usc.edu

Dear Fellow IOS Members:

I had the pleasure to meet many of you at the INFORMS Optimization Society (IOS) conference in Princeton a couple of months back. First off, I have to thank Warren Powell (Princeton) for putting together a contemporary, vibrant, and well-organized conference. For those of you who could not make it to the conference, I want you to know that our field continues to attract top-notch scholars from so many fields that there is something for every optimization sub-specialty and taste. I feel fortunate to have made this my area of study as far back as the late 70's. The staying power of our discipline derives

Send your comments and feedback to the Editor:

Marina A Epelman
Industrial and Operations Engineering
University of Michigan, Ann Arbor, USA
mepelman@umich.edu



2015 IOS Prizes winners (left to right): Javad Lavaei, Somayeh Sojoudi, Jean-Bernard Lasserre, Robert Weismantel, Fatma Kilinç-Karzan, and Paul Grigas

from the dedication and creativity of the Optimization community which continues to challenge itself via new connections with other academic disciplines (e.g., statistics, signal processing, robotics) while at the same time, addressing pressing real-world challenges of today (e.g. energy, health care, sustainability, security). The breadth of this field, spanning theory and applications, is not only visible in the IOS as a whole, but even in the research portfolio of its members. This is the essence of a truly vibrant research area, and I am honored to have the opportunity to serve as the Chair of the IOS.

As many of you are aware, the IOS is preparing a proposal to launch a new INFORMS journal devoted to Optimization. Over the past year, there have been several occasions to provide inputs for this new venture. A committee, consisting of John Birge (University of Chicago, chair), Dimitris Bertsimas (MIT), David Morton (Northwestern), Warren Powell (Princeton) and David Shmoys (Cornell) put together an initial vision for the journal. I am very grateful for their service, and thank them on behalf of our community. The recommendation, which was presented at the IOS Business Meeting, was further refined and circulated among the IOS membership earlier this year. The report presented the pros and cons of using both Optimization and Analytics as focal points for the journal. In response, the community spoke clearly in favor of a journal dedicated to Optimization alone, and there appeared to be a fairly significant group of members who were not supportive of including Analytics as part of the journal name. Instead, many suggested an editorial area for Analytics so that authors working at the interface between Optimization and Analytics would find a high-quality outlet for their work. This would give the IOS an opportunity to play a significant role in the direction of Analytics within INFORMS, while avoiding the risk associated with using a new area in the name of the journal. We will propose such a structure, with the remaining areas being aligned with the Special Interest Groups (SIGS) or Subdivisions within IOS. Over the next couple of months, we will be moving forward with a proposal to the INFORMS Publications Committee for a Journal of the INFORMS Optimization Society. Incidentally, the flagship INFORMS journal, *Operations Research*, is now inviting submissions in a new category of papers

described as “data-based principles of operational science.” I suspect that this area will be similar to the Analytics area of our proposed journal.

As part of this newsletter, I would like to acknowledge various awards committees which worked to identify outstanding contributions by individuals who are at different points in their careers. One of the most prestigious awards in the field of Optimization is the society’s Lifetime Award, named after Leonid Khachiyan. The 2015 Khachiyan award was given to Jean Bernard Lasserre (Laboratory for Analysis and Architecture of Systems, France) who has been a pioneer in polynomial and semi-algebraic optimization. The committee for the Khachiyan Prize was composed of Ilan Adler (Chair, UC Berkeley), Michael Ball (University of Maryland), Donald Goldfarb (Columbia University) and Werner Römisch (Humboldt-University Berlin). The Farkas Prize for mid-career researchers for 2015 went to Robert Weismantel (ETH-Zurich, Switzerland) for his contributions to discrete mathematics and optimization. The committee for this award was Ariela Sofer (Chair, George Mason University), Warren Adams (Clemson University), Sanjay Mehrotra (Northwestern University) and Zelda Zabinsky (University of Washington). The third award, for Young Researchers, was shared by Fatma Kılınç-Karzan (Carnegie Mellon University), and Javad Lavaei and Somayeh Sojoudi (UC Berkeley). Fatma’s award was for her work on Mixed-Integer Conic Programs (which appeared recently in *Math. of OR*), and Javad and Somayeh were awarded for their joint work on Semidefinite Relaxations with underlying Graph Structures (which appeared in *SIAM J. on Optimization*). The committee for the prize for young researchers was chaired by Nick Sahinidis (CMU), and other committee members were Daniel Bienstock (Columbia University), Sam Burer (University of Iowa), and Andrew Schaefer (Rice University). And finally, the winner of the student paper competition was Paul Grigas (MIT) for his paper on connections between boosting and subgradient optimization (which was coauthored with Robert Freund and Rahul Mazumder). This committee was chaired by Mohit Tawarmalani (Purdue University), and others who served were Fatma Kılınç-Karzan (CMU), Warren Powell (Princeton University) and Uday Shanbhag (Penn State).

I want to take this opportunity to thank the work of the officers who ended their terms in 2015. They included Jim Luedtke (Secretary/Treasurer), Imre Pólik (Computational Optimization and Software), Juan Pablo Vielma (Integer and Discrete Optimization), John Mitchell (Linear and Conic Optimization), Vladimir Boginski (Network Optimization), and Shabbir Ahmed (IOS Newsletter Editor). These highly motivated individuals have been replaced by an equally committed group in Burcu Keskin (Secretary/Treasurer), Hande Benson (Computational Optimization and Software), Amitabh Basu (Integer and Discrete Optimization), Saukeh Siddiqui (Linear and Conic Optimization), Austin Buchanan (Network Optimization) and Marina Epelman (IOS Newsletter Editor). Last, and certainly not the least, I would also like to thank our Chair-Elect David Morton for his willingness to serve as the Chair of IOS starting his two-year term in 2017.

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 hope to see you all in Nashville. (You can find me by looking for the Indian guy wearing a cowboy hat.)

Thanks!

Jean-Bernard Lasserre

LAAS-CNRS and
Institute of Mathematics
University of Toulouse
France
lasserre@laas.fr

I am very honored and grateful for being awarded the Khachiyan prize of the Optimization Society of INFORMS and of course I first want to thank the highly-respected members of the Khachiyan prize committee. Recognition (official or not) by peers is what any researcher would be proud of. However such recognition is rarely due to the sole merit of the awardee. Many (of which I am one) will agree that “research” is usually a mix between solitary intellectual efforts and collaborative work with PhD students, Post-Docs, and colleagues. In fact, and some philosophers would say it better than I, the notion of “merit” in any of our actions in a lifetime is itself questionable.

I cannot forget how much I owe to several persons at different stages of my career, for scientific and non-scientific reasons: After my PhD in Toulouse (France) on large-scale linear programs for production planning, and thanks to G. Giralt (from CNRS), I spent a wonderful one-year Post-Doc at the EECS department of the University of California at Berkeley (1978–1979). Being in such a stimulating environment was extraordinary (in the plain sense of the



Suvrajeet Sen and Jean-Bernard Lasserre

word) for a young student! In particular, I met P. Varaiya, an extraordinary man (in every aspect) who became my supervisor, and there, I realized that “research” was really what I wanted to do in my life! In addition, in Berkeley more than anywhere else in the world one could still enjoy some remainders of the sixties... I liked the place so much that I came back seven years later as an NSF Research Fellow (again visiting P. Varaiya).

Back to Toulouse, and after one year of military service (detached in a research lab) I was recruited at CNRS (Le Centre National de la Recherche Scientifique) and for me “C. N. R. S.” sounded like the four magic letters of a prestigious research institution! I worked again on *production planning and scheduling* but now with emphasis on *planning horizons*. I was then contacted by O. Hernandez-Lerma (IPN, Mexico) to start a collaboration on *Markov Control (or Decision) Processes* (mainly on Borel spaces) within a CNRS-CONACYT program (CNRS has many such bilateral cooperation programs with countries all over the world). This friendly collaboration lasted about 15 years during which I learnt a lot! In particular, seminal works of the sixties (notably by de Ghellinck, d’Epenoux, Manne) expressed discounted and average-cost MDPs with *finite* states and actions as *linear programs*. It turns out that such LP formulations can be extended to MDPs with (infinite-dimensional) Borel state and action spaces. But then I had to learn about real and functional analysis so as to become familiar with infinite-dimensional linear programs on appropriate spaces of measures. This was indeed possible because a nice thing about CNRS is that you can enjoy almost total freedom and so you may explore any research direction that you want. This collaboration with O. Hernandez-Lerma (and some other nice people in this small but friendly community) resulted in two Springer books on Markov Control Processes and one Birkhauser book on invariant measures for Markov Chains (MCs).

Then I started to be motivated by practical numerical evaluation of ergodic criteria for MCs on Borel spaces, directly via an invariant measure rather than via an estimator through “simulation”. It turns out that if the stochastic kernel P associated to the MC (the infinite-dimensional analogue of the transition matrix in the finite-state case) maps poly-

nomials into polynomials then the invariance property $\mu P = \mu$ translates into *countably many linear equations* between *moments* of the invariant measure μ . Hence if f is a polynomial, the ergodic functional $\int f d\mu$ is a linear combination of finitely many moments y of μ which satisfy a system of countably many linear equations (involving higher-order moments). So it remains to express conditions under which the vector y is coming from moments of some measure μ . This is where I had to understand the old K -moment problem in functional analysis (initiated by famous mathematicians at the end of the nineteenth century). After some time and efforts I ended up reading Schmüdgen and Putinar’s *Positivstellensätze* which in my mind are one of the few examples of a very powerful mathematical theorem whose statement (not proof) can be understood by freshmen at a university and can be used in so many applications (namely, every time where one has to state that a polynomial is positive on a compact semi-algebraic set). In fact, Putinar’s (and Schmüdgen’s) Positivstellensatz has two dual facets (one in real algebraic geometry about positivity on K , and one in functional analysis on the K -moment problem).

Last but not least, this beautiful and fascinating duality is nicely captured by standard duality of convex optimization in appropriate convex cones and can be implemented via a hierarchy of semidefinite programs. It was the beginning of an exciting period during which I could meet and interact with people from various areas at workshops in some very nice mathematics institutes around the world. Many became “colleagues” but in trying to cite them I would probably do unfortunate omissions. Since then I have tried to popularize the field of *Polynomial Optimization* (i) which is at the crossroad of several disciplines, (ii) whose list of important (practical and theoretical) applications is almost endless, and (iii) which meets convex optimization for practical implementation (via hierarchies of convex relaxations). It provides a lot of interesting research issues as well as challenges for practical implementation that should attract PhD students and researchers in optimization (in a broad sense) motivated by the multi- or inter-disciplinary aspects of the field!

Finally, in parallel with polynomial optimization, I was (and still am) also interested in the fascinating (at least in my mind) connections between four seem-

ingly unrelated problems defined on the same convex polytope K , namely, (i) maximizing $c'x$ on K (LP), (ii) maximizing $c'x$ on $K \cap \mathbb{Z}^n$ (Integer Programming (IP)), (iii) integrating $\exp c'x$ on K (linear integration), and (iv) summing up $\exp c'x$ on $K \cap \mathbb{Z}^n$ (linear counting). It turns out that again a nice “magic” formula from algebraic geometers (Brion & Vergne, Khovanskii & Pukhlikov) provides a result in closed-form for (iii) and (iv) in which all basic fundamental ingredients of (i) (LP) (basis, reduced cost, and dual vector) appear! Moreover, the same asymptotic result links (i) with (iii) and links (ii) with (iv). The formula for (iv) also encodes Gomory’s *corner polyhedron* associated with (ii). So exactly as the Legendre-Fenchel transform provides a duality for convex optimization, the Laplace transform (respectively, Z-transform) provides a duality for integration (respectively, for counting) with striking analogies and parallels in the linear case. In my taste, among many excitements and motivations in doing research, an important one is trying to reveal and understand connections between seemingly different fields. I enjoyed a lot writing a book on this topic even though I am somehow disappointed to have got almost no feedback so far! May be in some near future?

To conclude, in retrospect I have been very lucky and fortunate to have been able to change research topics and enter new fields at my convenience. As nicely expressed by D. Bertsekas in his last year’s Khachiyan speech, “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.” I will never be grateful enough to the CNRS institution that provides its researchers with this almost total freedom!

From Linear to Nonlinear Integer Optimization

Robert Weismantel

Department of Mathematics and
Institute for Operations Research
ETH Zurich
Switzerland

robert.weismantel@ifor.math.ethz.ch

1. Introduction

First of all I am very grateful to the Farkas Prize Committee 2015 consisting of Ariela Sofer (Chair), Warren Adams, Sanjay Mehrotra and Zeld Zabinsky. I feel greatly honored that I am the recipient of the prize in 2015.

I would like to take this opportunity to reflect about the field and present my motivation for some selected research in integer programming that I have conducted in the past years. The main intention of this article, though, is to outline three research directions that I personally regard as very interesting for the future.

When my friend and colleague Alexander Martin and I started in 1989 as PhD students in the group of Martin Grötschel in Augsburg and later in Berlin, the common belief was that general-purpose methods for integer programming do not work. Instead one would try to study a combinatorial / integer optimization problem from a polyhedral point of view,



Robert Weismantel and Ariela Sofer

i.e., detect classes of valid inequalities and derive necessary and sufficient conditions on the input instance under which such inequalities become facet-defining. Back then computers were very slow compared to current standards. The software that was available for solving integer programs often could not solve an integer optimization problem with 100 binary variables without exploiting special structure of the underlying problem.

This situation has changed drastically within twenty five years. State-of-the-art software for mixed integer linear programs can typically handle instances of moderate to large size without a priori knowledge about the combinatorics of the problem. This is a fascinating development for the field in general. Most important to me is the fact it allows us to move forward to more complicated mixed integer optimization models that involve nonlinear functions, for instance.

In the same vein as linear integer programming heavily relies on our ability to solve linear optimization problems, we can now go one step further. We assume that we have at our disposal an oracle for solving linear integer problems and exploit this oracle to tackle more involved problems.

The past fifty years have also seen an enormous development of the theory of combinatorial and integer programming. Besides cutting plane theory, also integer programming in fixed dimension, matroid theory, graph theory, totally unimodular matrices, Hilbert bases, theory of approximation algorithms have become subfields within mathematical optimization with various links to other mathematical disciplines such as probability theory, combinatorics and algebraic geometry. To the best of my knowledge, apart from basics of cutting plane theory, other theories are still not used by solvers. The question emerges whether we can bridge this gap? For my work, this question has been a major source of inspiration and motivation.

My attention was and still is caught by questions related to a systematic study of feasible solutions to systems of linear equations and inequalities over integers. The basic mathematical objects here are bases of lattices and their refinements to cones. Let me introduce these notions rigorously.

Definition 1. (*Bases of lattices and cones*)

- (a) Let B be an n -by- n invertible matrix. B is a basis of the lattice $\{B\lambda \mid \lambda \in \mathbb{Z}^n\}$.
- (b) Let $C \subseteq \mathbb{R}^n$ be a rational polyhedral cone. A set $H = \{h^1, \dots, h^k\} \subseteq C \cap \mathbb{Z}^n$ is a Hilbert basis of C if for every $x \in C \cap \mathbb{Z}^n$ there exist multipliers $\lambda_1, \dots, \lambda_k \in \mathbb{Z}_+$ such that

$$x = \sum_{i=1}^k \lambda_i h_i.$$

On the first glance, the definition (b) sounds simple. In fact, it is known for centuries that such a Hilbert basis of a rational polyhedral cone exists and is finite, see for instance Gordan [7]. Van der Corput showed that if the cone is pointed, then a Hilbert basis of minimal cardinality is uniquely determined [8]. Despite the fact that Hilbert bases occur in many areas of mathematics, surprisingly little is known about geometric and algebraic properties of such bases. Inspired by this fact I began to work with my colleague and friend Martin Henk on this topic. We were studying several parameters: the height of Hilbert bases [14], the integer Caratheodory number [15] and connections between Hilbert bases and lattice bases for simultaneous diophantine approximation [16, 17]. From the point of view of integer optimization, there are several important connections with Hilbert bases. Let me mention three below. Edmonds and Giles [11] introduced the notion of totally dual integral systems. Those systems have been characterized by Giles and Pulleyblank [12] by means of Hilbert bases of cones associated with the tight constraints at faces. Cook, Gerards, Schrijver and Tardos showed that Hilbert bases allow us to bound the distances between vertices of the linear programming optimum and optimal integral solutions [9]. Graver, in his work [10], noticed that Hilbert bases can be used to verify optimality of a given integer point that is feasible for an associated integer programming problem. This optimality certificate is usually too big to be computed efficiently. There are, however, special cases even in variable dimension in which this can be accomplished. As one such nice example let me refer to the family of N -fold integer programming problems that were introduced and studied with my colleagues Raymond Hemmecke, Jesús de Loera and Shmuel Onn in [19]. There are several further examples where one can

shed light on integer programming problems using Hilbert bases. Together with my colleagues Raymond, Jesús, Matthias, Shmuel and also Jon Lee we have been investigating this topic intensively. I will refrain from listing all our joint papers here. Let me at least mention that it was a great pleasure having had this very nice collaboration over a period of about ten years. As a team we could push this theory to a level that I am satisfied with today. Jointly with my former students Utz Haus and Matthias Köppe we turned lattice- and Hilbert-basis theory into a primal algorithm, the integral basis method [18]. One starts with a solution and turns it into a basic feasible one of an appropriate simplex tableau. Then columns in this tableau are removed and replaced by integer combinations of the removed columns with non-removed ones in a way that no primal feasible integral solution is lost. This replacement step may be viewed as an affine version of a Hilbert basis computation and can be accomplished very fast for special integer programs. Proceeding this way, one either arrives at a simplex tableau proof that the current solution is optimal or detects an improving direction. This is why the method is a primal integer programming algorithm. It would have been interesting to combine this method with cutting plane schemes to develop a primal/dual simplex tableau based algorithm. To the best of my knowledge nobody experimented with this idea yet. Hopefully, we will at some point arrive at a primal / dual algorithm for integer programming. This was always a dream of mine.

2. A link between fixed and variable dimension

Despite the fact that there are numerous beautiful and deep results about integer optimization problems with a constant number of variables such as Lenstra's algorithm [5], it seems difficult to apply these results to practically relevant instances. One obvious explanation is that such instances typically live in very high dimensions. Moreover, these high-dimensional problems are typically mixed integer problems with *binary* variables. Given that fixed-dimension theory applies to few variables with a large range of potential values for the individual

variables, it is not straightforward to apply the algorithm of Lenstra, for instance, in a direct manner. One attempt to use fixed-dimension integer programming theory for the study of high-dimensional optimization problems is presented in my joint work with Rico Zenklusen, Robert Hildebrand and Jörg Bader [1]. Consider the “high-dimensional” integer optimization problem

$$\max \{c^T x \mid Ax \leq b, x \in \mathbb{Z}^n\}.$$

Let us try to find a matrix $W \in \mathbb{Z}^{k \times n}$ with k small and relate the two polyhedra

$$\text{conv}\{x \in \mathbb{Z}^n \mid Ax \leq b\}$$

and

$$\text{proj}_x \left(\text{conv}\{(x, z) \in \mathbb{R}^n \times \mathbb{Z}^k \mid Ax \leq b, Wx = z\} \right).$$

Since W is integral, the following inclusion property is obvious: the polyhedron on the left is always contained in the one on the right. Furthermore, if W is the n -by- n identity matrix, then the two polyhedra coincide. The objectives here are two-fold:

- If we request equality between the two polyhedra, then the goal is to find an integer matrix W with as few rows k as possible so that one can model the integer hull of the original feasible region.
- If we do not require equality between the two polyhedra, then the question emerges whether we can find good approximations of the integer hull of all solutions to a specific combinatorial / integer problem by means of solutions of a mixed integer problem with few integrality constraints.

In other words, in the first setting we consider reformulations based only on the constraint matrix A , and hence these reformulations apply to all integral right-hand sides b . For this, we study decompositions of the matrix A that access underlying totally unimodular (TU) properties of the matrix. We say a matrix factorization $A = UW$ is a *TU decomposition* of A if U is an integral matrix and W is TU. From a mathematical point of view, this matrix property is interesting on its own. For algorithmic applications,

we introduce a more general notion: A decomposition $A = \tilde{A} + UW$ is called an *affine TU decomposition* of A if U is an integral matrix and the joint matrix $[\tilde{A}; W]$ is TU.

Our motivation for studying affine TU decompositions of matrices comes from the following simple fact.

Theorem 2. *Let $A = \tilde{A} + UW \in \mathbb{Z}^{m \times n}$ with $W \in \{0, \pm 1\}^{k \times n}$ be an affine TU decomposition, and $b \in \mathbb{Z}^m$. Then $\text{conv}(\{x \in \mathbb{R}^n \mid Ax \leq b, Wx \in \mathbb{Z}^k\})$ is an integral polyhedron.*

This is one answer to the first question we were posing before: Polyhedra

$$P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$$

with an affine TU decomposition $A = \tilde{A} + UW$ have the property

$$\begin{aligned} & \text{conv}\{x \in \mathbb{R}^n \mid Ax \leq b\} \\ &= \text{proj}_x(\text{conv}\{(x, z) \in \mathbb{R}^n \times \mathbb{Z}^k \mid Ax \leq b, Wx = z\}). \end{aligned} \quad (1)$$

Among all affine TU decompositions of A , we call the minimal number of rows needed for W the *affine TU-dimension* of A . This is the reformulation that we are really interested in, since the number of integer variables is a measure of complexity for the underlying problem.

At this point a reader might wonder whether an affine TU decomposition of a matrix is too much to ask for if we only request equality between the two polyhedra in Eq. (1). In a somewhat restricted setting a converse of the above theorem holds.

Theorem 3. *Let $A \in \mathbb{Z}^{m \times n}$, and let $W \in \mathbb{Z}^{k \times n}$ have rank k such that the polyhedron*

$$\{x \in \mathbb{R}^n \mid Ax \leq b, Wx = d\}$$

is integral for all $b \in \mathbb{Z}^m$ and for all $d \in \mathbb{Z}^k$. Then there exist matrices $U \in \mathbb{Z}^{m \times k}$ and $W' \in \mathbb{Z}^{k \times n}$ such that $A = \tilde{A} + UW'$ is an affine TU decomposition. Moreover, for every $b \in \mathbb{Z}^m$,

$$\begin{aligned} & \text{conv}\left(\{x \in \mathbb{R}^n \mid Ax \leq b, Wx \in \mathbb{Z}^k\}\right) \\ &= \text{conv}\left(\{x \in \mathbb{R}^n \mid Ax \leq b, W'x \in \mathbb{Z}^k\}\right). \end{aligned}$$

One can now study various specific combinatorial optimization problems and exhibit constructions that lead to an affine TU-dimension of a given matrix. In particular, there are various knapsack problems for which the affine TU-dimension of the given vector of numbers is not too big. This then automatically leads to a corresponding mixed integer model with few integer variables. Let me also mention in this context that when k and the number of rows of A are fixed, then one can find a polynomial time algorithm to determine if the affine TU-dimension of A is equal to k . On the other hand it is NP-hard to decide if for a given matrix $A \in \mathbb{Z}^{m \times n}$ its affine TU-dimension is equal to n . It remains yet open whether determining the affine TU dimension for a general matrix A is polynomial-time solvable when only k is a constant.

Let me now briefly turn to the second question. If we do not require equality between the two polyhedra, then the question emerges whether we can find good approximations of the integer hull of all solutions to a specific combinatorial / integer problem by means of solutions of a mixed integer problem with few integrality constraints. This question is still widely open. Indeed, besides some basic examples we know very little about this topic. For instance, the parity polytope has a simple inequality description if we add one additional integrality constraint. There also exist classes of knapsack polytopes that (i) have an exponential-sized polyhedral description of its convex hull of integer solutions and (ii) the convex hull can be described by linearly many linear inequality constraints together with one integrality constraint. These are two examples that illustrate that it is well conceivable to find interesting approximations of a convex hull of integer solutions with a small number of integrality constraints.

3. Lattice free sets and mixed integer convex optimization

In my view one of the most important objects to study in order to better understand integer programming are lattice-free sets. Before explaining this in some more detail, let me make precise what lattice-free sets are.

Definition 4. (*Lattice-free sets*)

Let $S \subset \mathbb{R}^{n+d}$ be a convex set. S is called *lattice free* if

$$\text{int}(S) \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset.$$

Furthermore, we call S *maximally lattice free* if S is maximal with respect to inclusion.

Basic examples of lattice-free sets are strips: for an integer normal vector $c \in \mathbb{Z}^n$ whose greatest common divisor of all entries is one and an integer number γ , the set $\{x \in \mathbb{R}^{n+d} \mid \gamma \leq [c, 0]^T x \leq \gamma + 1\}$ is maximally lattice free. However, the geometry of maximally lattice-free sets is complicated in general. The following properties established by Lovász and Doignon are well-known and fundamentally important.

Lemma 5. [4] [6] *Let $S \subset \mathbb{R}^{n+d}$ be a maximal, full-dimensional, lattice-free, convex set. Then, the following properties hold:*

- a) *Let K be the orthogonal projection of S onto \mathbb{R}^n . Then, K is maximally lattice free (with respect to \mathbb{R}^n) and $S = K \times \mathbb{R}^d$,*
- b) *S is a polyhedron,*
- c) *each facet of S contains an integer point in its relative interior and*
- d) *S has at most 2^n facets.*

A substantial amount of research in mixed integer programming is dedicated to the question of how to derive inequalities from a description of P that are satisfied by all the points in $P \cap (\mathbb{Z}^n \times \mathbb{R}^d)$. Such inequalities naturally define relaxations of $P_I = \text{conv}(P \cap (\mathbb{Z}^n \times \mathbb{R}^d))$ in form of polyhedra that are contained in P and that contain P_I . In order to obtain such relaxations, one can use an operator introduced in Andersen, Louveaux and Weismantel [2] that may be viewed as a special disjunctive programming approach, a general and widely applicable framework invented by Egon Balas in the nineteen-seventies. More formally, if L is a lattice-free polyhedron, then

$$P_I \subset P \setminus L := \overline{\text{conv}}(P \setminus \text{int } L) \subseteq P.$$

Here, int denotes the topological interior and $\overline{\text{conv}}$ — the closed convex hull. Such an operation can be used iteratively and for all lattice-free sets. Some questions arising in this context are: which lattice-free polyhedra L should one use in order to

- approximate P_I sufficiently well, or
- be able to obtain a polyhedral closure, or
- prove finite convergence to the mixed integer hull, or
- develop a cutting plane proof?

These questions have been addressed in joint work with my former postdocs, Kent Andersen, Quentin Louveaux and Alberto del Pia, [20, 21]

As a second example of why lattice-free polyhedra are intrinsically related with an understanding of mixed integer problems, I would now like to move away from linear mixed integer optimization and consider a more general convex optimization problem in mixed integer variables. I refer here to my joint work with Michel Baes and Timm Oertel [2]. My intention is to indicate informally that optimality certificates and duality in convex optimization have a very natural mixed integer analogue. Recall first that a duality theory in Euclidean space follows from a precise interplay between points — which are viewed as primal objects — and hyperplanes interpreted as dual objects. It turns out that there is a similar interplay in the mixed integer setting. Here, the primal objects are sets of points, whereas the dual objects are lattice-free polyhedra.

Let $f : \text{dom}(f) \mapsto \mathbb{R}$ be a continuous convex function. In order to simplify our exposition we may assume w.l.o.g. here that $\text{dom}(f) = \mathbb{R}^n$. Assume that f has a, not necessarily unique, minimizer x^* . Then a necessary and sufficient certificate for x^* being a minimizer of f is that $0 \in \partial f(x^*)$, i.e., the zero-function is in the subdifferential of f at x^* . Hence,

$$x^* = \text{argmin}_{x \in \mathbb{R}^n} f(x) \iff 0 \in \partial f(x^*).$$

A question emerges: how do we obtain a certificate that a point $x^* \in \mathbb{Z}^n \times \mathbb{R}^d$ solves the corresponding mixed integer convex problem

$$x^* = \text{argmin}_{x \in \mathbb{Z}^n \times \mathbb{R}^d} f(x)? \quad (2)$$

Let us explain the idea of our approach. By definition, $x^* = \text{argmin}_{x \in \mathbb{Z}^n \times \mathbb{R}^d} f(x)$ if and only if

$$\{x \in \mathbb{Z}^n \times \mathbb{R}^d \mid f(x) < f(x^*)\} = \emptyset. \quad (3)$$

The level set $\{x \in \mathbb{R}^{n+d} \mid f(x) \leq f(x^*)\}$ is convex. If it is nonempty, then its projection to its first n

components, that is, to the subspace spanned by the integer variables is again a convex set. Clearly, $x^* = \operatorname{argmin}_{x \in \mathbb{Z}^n \times \mathbb{R}^d} f(x)$ if and only if

$$Q := \{z \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^d, x = (z, y) \text{ and } f(x) < f(x^*)\} \cap \mathbb{Z}^n = \emptyset.$$

From the theorem of Lovász, inclusion-wise maximally lattice-free convex sets are polyhedra [4]: we can restrict our attention to such polyhedra P that contain the (convex) projection Q . From the theorem of Doignon [6], it follows that a subset of at most 2^n inequalities in the description of P is enough to prove that $\operatorname{int}(P) \cap \mathbb{Z}^n = \emptyset$. It remains to show how to relate these 2^n inequalities to the function f . The following theorem clarifies this relationship, providing a necessary and sufficient optimality condition for our original mixed integer convex problem. Each of these 2^n inequalities is related to a mixed integer point, the set of which constitutes our *optimality certificate*. Condition (a) ensures that x^* is one of the points of the optimality certificate and is the best of them. Also, in view of condition (c), every point x in the certificate minimizes f on its own fiber, that is, on the set $\{(x_1, \dots, x_n)^T\} \times \mathbb{R}^d$. Finally, the subgradient of f at each point of the certificate defines a half-space. The interior of their intersection defines a polyhedron whose projection on the first n components is lattice-free by condition (b).

One direct implication of the result below is that it provides us with a “good” certificate for the optimality of a given mixed-integer point. The verification can be performed in polynomial time, provided that the number of integer variables is a constant and regardless of how many continuous variables are involved.

Theorem 6. $x^* = \operatorname{argmin}_{x \in \mathbb{Z}^n \times \mathbb{R}^d} f(x)$ if and only if there exist $k \leq 2^n$ points

$x_1 = x^*, x_2, \dots, x_k \in \mathbb{Z}^n \times \mathbb{R}^d$ and vectors $h_i \in \partial f(x_i)$

such that the following conditions hold:

- (a) $f(x_1) \leq \dots \leq f(x_k)$.
- (b) $\{x \in \mathbb{R}^{n+d} \mid h_i^T(x - x_i) < 0 \text{ for all } i\} \cap (\mathbb{Z}^n \times \mathbb{R}^d) = \emptyset$.
- (c) $h_i \in \mathbb{R}^n \times \{0\}^d$ for $i = 1, \dots, k$.

In the same vein as one can generalize the unconstrained optimality conditions to the KKT theorem, the result above can be extended to convex constraints. The Lagrangean relaxation method then leads to a formalism of duality in convex optimization and allows us to develop a duality result for mixed integer convex programs. The only difference is that a single point is replaced by a lattice free polyhedron whose inequality description is polynomial whenever the number of integer variables is a constant. I view this as a first step towards the ultimate goal of designing new mixed-integer convex algorithms. Any lattice free polyhedron that we construct in the course of such an algorithm provides us with a lower bound on the optimal objective function value. This is why lattice free polyhedra can be viewed as dual objects for integer and mixed integer programs.

4. An axiomatic approach to nonlinear integer optimization

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a nonlinear function. We consider the discrete optimization problem

$$\min\{f(x) : x \in P \cap \mathbb{Z}^n\}. \quad (4)$$

Let us assume that P is a polytope presented by means of an inequality description, i.e., $P = \{x \in \mathbb{R}^n : Ax \leq b\}$, $A \in \mathbb{Z}^{m \times n}$, and $b \in \mathbb{Z}^m$. We say that Problem (4) can be solved in polynomial time if in time bounded by a polynomial in the size of its input, we can either determine that the problem is infeasible, or we can find a feasible minimizer.

When $f(x)$ is a convex function, Problem (4) with fixed n and bounded P can be solved in polynomial time by a Lenstra-type algorithm, see [5, 3] and [26]. On the other hand if f is concave with fixed n , then by computing the integer hull of P using [25] the problem is also polynomial time solvable. There are a few other polynomially solvable special cases when n is fixed. In particular, when f is a polynomial of fixed degree, then in joint work with my colleagues Jesús de Loera, Raymond Hemmecke and Matthias Köppe [24] we developed an FPTAS (fully polynomial time approximation scheme) for maximizing non-negative polynomials over integer points in a polytope. For $n \leq 2$, and f being a poly-

mial of degree at most three, the problem is also polynomial-time solvable, see [27] and [22].

This raises the question of how we can combine all these techniques to obtain complexity results for larger classes of functions.

In order to develop an FPTAS for classes of nonlinear functions to be minimized over integer points in polyhedra, we recently proposed a framework. This is joint work with Robert Hildebrand and Kevin Zemmer [13]. The idea is to combine the techniques of Papadimitriou and Yannakakis [23] with ideas similar to those commonly used to derive certificates of positivity for polynomials over semialgebraic sets. Generally speaking, in the latter context one is given a finite number of “basic polynomials” f_1, \dots, f_m which are known to be positive over the integers in a polyhedron P . A sufficient condition to prove that another polynomial f is positive over $P \cap \mathbb{Z}^n$ is to find a decomposition of f as a sum of products of a *sum of squares* (SOS) polynomial and a basic function f_i . A polynomial $p(x)$ is SOS if there exist polynomials $q_1(x), \dots, q_m(x)$ such that $p(x) = \sum_{i=1}^m q_i^2(x)$.

We would like to use a similar approach to arrive at an FPTAS. Again we work with classes of “basic functions.” Then, for a given f , we try to detect a decomposition of f as a finite sum of products of a so-called “sliceable function” and a basic function f_i . Roughly speaking, sliceable functions — thanks to the result of [23] — can be approximated by subdividing the given polyhedron.

For instance, the set of all convex functions presented by a first-order oracle that are nonnegative over $P \cap \mathbb{Z}^n$ could serve as a class of basic functions, because we can solve Problem (4) for any member in the class in polynomial time when n is fixed. The nonnegativity assumption implies sign-compatibility, which is a necessary property of the set of basic functions. Another example is the set of all concave functions presented by an evaluation oracle that are nonnegative over $P \cap \mathbb{Z}^n$. The same property holds true in this case. These two examples demonstrate that we consider not only polynomials f_i , but also more general classes of basic functions. For example, we can decompose the polynomial $x^2 + y^2 - z^2$ as the product of two non-polynomial functions: a basic function $\sqrt{x^2 + y^2} - |z|$ and a sliceable function $\sqrt{x^2 + y^2} + |z|$. Our technique also ap-

plies, for instance, to the Motzkin polynomial, and even to functions f that are not polynomials.

In the aforementioned paper we explain axiomatically what we mean by basic and sliceable functions that for the sake of brevity I refrain from explaining here in detail. As a consequence of our technique we easily derive the following result

Theorem. *Let $Q \in \mathbb{Z}^{n \times n}$ be a symmetric matrix and let n be fixed. Then there is an FPTAS for Problem (4) with $f(x) = x^T Q x$ in the following cases:*

1. Q has at most one negative eigenvalue;
2. Q has at most one positive eigenvalue.

This decomposition technique might be viewed as a first step towards a systematic approach of which classes f of functions admit an FPTAS for the corresponding optimization problem (4) when n is a constant.

REFERENCES

- [1] J. Bader, R. Hildebrand, R. Weismantel, R. Zenklusen, *Mixed Integer Reformulations of Integer Programs and the Affine TU-dimension of a Matrix*, Manuscript 2015.
- [2] M. Baes, T. Oertel, R. Weismantel, *Duality for Mixed-Integer Convex Minimization*, Manuscript 2014, Mathematical Programming to appear.
- [3] M. Grötschel, L. Lovász, A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*, Springer Berlin, 1988.
- [4] L. Lovász. Geometry of numbers and integer programming. In M. Iri and K. Tanabe, editors, *Mathematical Programming: State of the Art*, pages 177–201, Mathematical Programming Society, 1989.
- [5] H.-W. Lenstra. *Integer programming with a fixed number of variables*. Mathematics of Operations Research 8, 538–548, 1983.
- [6] J.-P. Doignon. *Convexity in crystallographical lattices*. J. Geometry, 3, 71–85, 1973.
- [7] P. Gordan. *Über die Auflösung linearer Gleichungen mit reellen Coefficienten*. Math. Ann., 6, 23–28, 1873.
- [8] J.-G. Van der Corput. *Über Systeme von linear-homogenen Gleichungen und Ungleichungen*. Proceedings Koninklijke Akademie van Wetenschappen te Amsterdam 34, 368–371, 1931.

- [9] W. Cook, B. Gerards, A. Schrijver, E. Tardos. *Sensitivity theorems in integer linear programming*. Mathematical Programming 34, 251 – 264, 1986.
- [10] J. Graver. *On the foundation of linear and integer programming I*. Mathematical Programming 9, 207 – 226, 1975.
- [11] J. Edmonds, F. R. Giles. *A min-max relation for submodular functions on graphs*. Annals of Discrete Mathematics 1, 185 – 204, 1977.
- [12] F. R. Giles, W. R. Pulleyblank. *Total dual integrality and integer polyhedra*. Linear Algebra and Its Applications 25, 191 – 196, 1979.
- [13] R. Hildebrand, R. Weismantel, K. Zemmer. *An FPTAS for Minimizing Indefinite Quadratic Forms over Integers in Polyhedra*. Manuscript 2015, SODA 2016 to appear.
- [14] M. Henk, R. Weismantel, *The height of minimal Hilbert bases*, Results in Mathematics 32, 298 – 303, 1997.
- [15] W. Bruns, J. Gubaladze, M. Henk, A. Martin, R. Weismantel, *A counterexample to an integer analogue of Carathéodory's Theorem*, Journal of Pure and Applied Mathematics 510, 179 – 185, 1999.
- [16] M. Henk, R. Weismantel, *On minimal solutions of Diophantine equations*, Contributions to Algebra and Geometry 41, 49 – 55, 2000.
- [17] M. Henk, R. Weismantel, *Diophantine approximations and integer points of cones*, Combinatorica 22, 401 – 408, 2002.
- [18] U. Haus, M. Köppe, R. Weismantel, *A primal all-integer algorithm based on irreducible solutions*, Mathematical Programming 96, 205 – 246, 2003.
- [19] R. Hemmecke, J. De Loera, S. Onn, R. Weismantel, *N-fold Integer Programming*, Discrete Optimization 5 (2), 231 – 241, 2008.
- [20] K. Andersen, Q. Louveaux, R. Weismantel, *An analysis of mixed integer linear sets based on lattice point free convex sets*, Mathematics of Operations Research 35, 233 – 256, 2010.
- [21] A. Del Pia, R. Weismantel, *On convergence in mixed integer programming*, Mathematical Programming 135, 397 – 412, 2012.
- [22] A. Del Pia, R. Hildebrand, R. Weismantel, R. Zemmer, *Minimizing Cubic and Homogeneous Polynomials over Integers in the Plane*, Manuscript 2014, Mathematics of Operations Research to appear.
- [23] C.H. Papadimitriou, M. Yannakakis, *On the Approximability of Trade-offs and Optimal Access of Web Sources*, Proceedings of the 41st Annual Symposium on Foundations of Computer Science, 86 – 92, 2000.
- [24] J. De Loera, M. Köppe, R. Hemmecke, R. Weismantel, *Integer polynomial optimization in fixed dimension*, Mathematics of Operations Research 31, 147 – 153, 2006.
- [25] B. Cook, M. Hartmann, R. Kannan, C. McDiarmid, *On integer points in polyhedra*, Combinatorica 12, 27 – 37, 1992.
- [26] L. Khachiyan, L. Porkolab, *Integer Optimization on Convex Semialgebraic Sets*, Discrete and Computational Geometry 23, 207 – 224, 2002.
- [27] A. Del Pia, R. Weismantel, *Integer quadratic programming in the plane*, Proceedings of SODA 2014, 840 – 846, 2014.

Disjunctive Conic Sets, Conic Minimal Inequalities, and Cut-Generating Functions

Fatma Kılınç-Karzan

Tepper School of Business
Carnegie Mellon University,
USA

fkilinc@andrew.cmu.edu

This article summarizes the paper [15] and some recent developments which are concerned with the *disjunctive conic sets* of form

$$\mathcal{S}(A, \mathcal{K}, \mathcal{B}) := \{x \in \mathbb{E} : Ax \in \mathcal{B}, x \in \mathcal{K}\},$$

where \mathbb{E} is a finite dimensional Euclidean space with inner product $\langle \cdot, \cdot \rangle$, $A : \mathbb{E} \rightarrow \mathbb{R}^m$ is a linear map, $\emptyset \neq \mathcal{B} \subset \mathbb{R}^m$ is a set of right-hand side vectors, and $\mathcal{K} \subset \mathbb{E}$ is a *regular* (closed, convex, full-dimensional, and pointed) cone. We restrict our attention to the interesting cases where $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$ is nonempty and nonconvex. Thus, we assume $\mathcal{B} \neq \emptyset$ but make no other assumptions on \mathcal{B} ; in particular, \mathcal{B} may be either finite or infinite. Examples of regular cones include the nonnegative orthant \mathbb{R}_+^n , the second-order (Lorentz) cone \mathbb{L}^n , and the positive semidefinite cone \mathbb{S}_+^n .

Disjunctive conic sets arise naturally in the solution set representations of Mixed Integer Conic Programs (MICPs) where nonlinear convex relations

among variables are captured in the conic constraint $x \in \mathcal{K}$ and integrality restrictions are encoded in A and \mathcal{B} by an appropriate selection. These sets also form the basis of fundamental structured relaxations used in generating cutting planes/surfaces for MICPs. For example, a disjunctive conic sets $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$ can represent multi-term (or split) disjunctions on regular cones and their cross-sections. Besides, the separation of a fractional solution from the feasible set of a Mixed Integer Linear Program (MILP) can be encoded as a set $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ with a closed set \mathcal{B} satisfying $0 \notin \mathcal{B}$ [5, 12, 14]. Moreover, the flexibility in the choice of \mathcal{B} makes these sets a relevant model for conic optimization problems with complementarity constraints. See [15, Sec 1.2] for illustrative examples.

The set $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ has compelled significant attention. When \mathcal{B} is a finite set, $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ is nothing but a disjunctive set such as those introduced and studied by Balas [2]. Johnson [14] characterized minimal valid linear inequalities for $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ through support functions of certain sets. Jeroslow [12] and Blair [5] presented similar characterizations via the value functions of MILPs with bounded feasible sets (in the former) and with rational data (in the latter). This body of work has strong connections to the strong duality theory for MILPs [1, 11].

In this paper, we generalize earlier results on classification and characterization of strong valid linear inequalities for the convex hull description of $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$ to the case where \mathcal{K} is a general regular cone without relying on the prior assumptions such as the finiteness of \mathcal{B} , etc. In order to capture dominance relations among valid linear inequalities, we introduce the notion of *conic minimality* of an inequality. This definition exposes a shortcoming in the usual minimality definition and offers a potential remedy via using \mathcal{K} to encode structural information on the problem. We perform a systemic study of conic minimal inequalities in terms of their existence, sufficiency, strength, necessary conditions and sufficient conditions for their characterization, and establish connections with functions that generate these inequalities.



Fatma Kılınç-Karzan and Nick Sahinidis

Introducing some notation

For a set $Q \subset \mathbb{R}^n$, we denote its topological interior by $\text{int}(Q)$ and its closed convex hull by $\overline{\text{conv}}(S)$. The support function of a set $Q \subset \mathbb{R}^n$ is defined as $\sigma_Q(z) := \sup_{q \in \mathbb{R}^n} \{z^\top q : q \in Q\}$. Support functions are sublinear (positively homogeneous, subadditive, and thus convex); and when $Q \neq \emptyset$, we also have $\sigma_Q(0) = 0$.

Given two Euclidean spaces \mathbb{E}, \mathbb{F} , we define the kernel of a linear map $A : \mathbb{E} \rightarrow \mathbb{F}$ as $\text{Ker}(A) := \{u \in \mathbb{E} : Au = 0\}$ and its image as $\text{Im}(A) := \{Au : u \in \mathbb{E}\}$. We use A^* to denote the conjugate linear map given by the identity $\langle y, Ax \rangle_{\mathbb{F}} = \langle A^*y, x \rangle_{\mathbb{E}} \quad \forall (x \in \mathbb{E}, y \in \mathbb{F})$. When the Euclidean space \mathbb{E} is just \mathbb{R}^n , we use the dot product as the corresponding inner product.

For a given cone $\mathcal{K} \subset \mathbb{E}$, we let \mathcal{K}^* denote its dual cone given by $\mathcal{K}^* := \{y \in \mathbb{E} : \langle x, y \rangle \geq 0 \quad \forall x \in \mathcal{K}\}$ and $\text{Ext}(\mathcal{K})$ denote the set of the extreme rays of \mathcal{K} . We let $[n] := \{1, \dots, n\}$ for any positive integer n .

A hierarchy on valid linear inequalities

We pursue a principled study of the structure of valid linear inequalities defining the closed convex hull of $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$. Given any vector $\mu \in \mathbb{E}$ and a number $\mu_0 \leq \vartheta(\mu)$ where $\vartheta(\mu)$ is defined as

$$\vartheta(\mu) := \inf_{x \in \mathbb{E}} \{\langle \mu, x \rangle : x \in \mathcal{S}(A, \mathcal{K}, \mathcal{B})\},$$

the linear inequality of the form $\langle \mu, x \rangle \geq \mu_0$ is a *valid inequality* for $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$. We refer to a valid inequality $\langle \mu, x \rangle \geq \mu_0$ as *trivial* if $\mu_0 = -\infty$, and as *tight* if $\mu_0 = \vartheta(\mu)$. We say that a valid linear inequality for $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$ is *extreme* if it is a valid equation or if it cannot be written as the sum of two distinct valid linear inequalities (sums of valid equations are excluded here). While extreme inequalities are necessary and sufficient for a complete description of $\overline{\text{conv}}(\mathcal{S}(A, \mathcal{K}, \mathcal{B}))$, their identification or algebraic characterization is often quite complicated. We compromise on this by examining the structure of slightly larger classes of inequalities — *minimal* and *sublinear* inequalities defined with respect to the cone \mathcal{K} .

Let us start by pointing out a simple class of valid inequalities. From the definition of \mathcal{K}^* , any inequality $\langle \delta, x \rangle \geq 0$ with $\delta \in \mathcal{K}^*$ is valid for $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$ since $\mathcal{S}(A, \mathcal{K}, \mathcal{B}) \subseteq \mathcal{K}$. We refer to these as *cone-implied inequalities*. Cone-implied inequalities may be extreme in certain cases; even so, they are not interesting because the constraint $x \in \mathcal{K}$ captures all of them.

Cone \mathcal{K} in the description of $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$ plays a critical role in identifying dominance relations among valid linear inequalities. Consider two valid inequalities for $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$ given by $\langle \mu, x \rangle \geq \mu_0$ and $\langle \rho, x \rangle \geq \rho_0$. We say that $\langle \rho, x \rangle \geq \rho_0$ *dominates* $\langle \mu, x \rangle \geq \mu_0$ with respect to the cone \mathcal{K} whenever $\mu - \rho \in \mathcal{K}^* \setminus \{0\}$ and $\rho_0 \geq \mu_0$. In fact, when $\langle \rho, x \rangle \geq \rho_0$ dominates $\langle \mu, x \rangle \geq \mu_0$, we have

$$\langle \mu, x \rangle = \underbrace{\langle \rho, x \rangle}_{\geq \rho_0} + \underbrace{\langle \mu - \rho, x \rangle}_{\geq 0} \geq \rho_0 \geq \mu_0,$$

where the first inequality follows from $x \in \mathcal{K}$ and $\mu - \rho \in \mathcal{K}^*$. Then in such a case, $\langle \mu, x \rangle \geq \mu_0$ is a consequence of the inequality $\langle \rho, x \rangle \geq \rho_0$ and the conic constraint $x \in \mathcal{K}$. This motivates our definition of *conic minimal* inequalities:

Definition 1. A valid inequality $\langle \mu, x \rangle \geq \mu_0$ for $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$ is called \mathcal{K} -minimal if for all inequalities $\langle \rho, x \rangle \geq \rho_0$ valid for $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$ with $\mu - \rho \in \mathcal{K}^* \setminus \{0\}$, we have $\rho_0 < \mu_0$.

Conic minimality definition specifically restricts our attention to the class of valid inequalities that cannot be written as the sum of another valid inequality and a cone-implied inequality. Thus, none of the cone-implied inequalities is \mathcal{K} -minimal. However, some \mathcal{K} -minimal inequalities can be expressed as the sum of two other non-cone-implied valid inequalities. Hence, not all \mathcal{K} -minimal inequalities are extreme.

In finite and infinite relaxations associated with MILPs, minimality of a valid inequality is traditionally defined with respect to the nonnegative orthant, i.e., $\mathcal{K} = \mathbb{R}_+^n$. That is, a valid inequality $\langle \mu, x \rangle \geq \mu_0$ is \mathbb{R}_+^n -minimal if reducing any coefficient μ_i for $i \in [n]$ leads to a strict reduction in the right-hand side value μ_0 (see [14]). Therefore, our conic minimality concept for disjunctive conic sets is a natural generalization of \mathbb{R}_+^n -minimality.

Extending earlier results from [11, 14] given in the case of $\mathcal{K} = \mathbb{R}_+^n$ to general regular cones \mathcal{K} , we can easily see that \mathcal{K} -minimal inequalities exist only if the following assumption holds (see [15, Prop 1]):

Assumption 2. *For each $\delta \in \mathcal{K}^* \setminus \{0\}$, there exists some $x_\delta \in \mathcal{S}(A, \mathcal{K}, \mathcal{B})$ such that $\langle \delta, x_\delta \rangle > 0$.*

When, for example, $\overline{\text{conv}}(\mathcal{S}(A, \mathcal{K}, \mathcal{B}))$ is full dimensional, Assumption 2 is satisfied and hence \mathcal{K} -minimal inequalities exist. Also, under Assumption 2, all non-cone-implied, extreme inequalities are \mathcal{K} -minimal.

Proposition 3 ([15, Prop 2 and Cor 2]). *Under Assumption 2, \mathcal{K} -minimal inequalities together with the conic constraint $x \in \mathcal{K}$ are sufficient to describe $\overline{\text{conv}}(\mathcal{S}(A, \mathcal{K}, \mathcal{B}))$.*

This prompts an interest in \mathcal{K} -minimal inequalities and suggests that in an efficient cutting plane procedure we should at the least aim at separating inequalities from this class.

On the selection of cone \mathcal{K} in disjunctive conic representations

In all of the previous literature, minimality of an inequality is defined with respect to the nonnegative orthant. We next expose a shortcoming of this and illustrate how encoding structural information in the cone \mathcal{K} is rather pivotal in providing a more refined characterization of extreme inequalities. This point is important even in the case of a disjunctive set associated with an MILP; yet it has been completely overlooked in the literature.

First note that we are essentially interested in the closed convex hull characterizations of disjunctive conic sets and because of our flexibility in selecting \mathcal{B} and \mathcal{K} , we may have a choice among several different representations $\mathcal{S}(A_1, \mathcal{K}_1, \mathcal{B}_1)$, $\mathcal{S}(A_2, \mathcal{K}_2, \mathcal{B}_2)$, etc. Moreover, whether a valid inequality is necessary for the convex hull description, i.e., extreme, depends on only the closed convex hull and is independent of the choice of A , \mathcal{B} , and \mathcal{K} used in the representation. Besides, as long as the closed convex hull remains the same, \mathcal{K} -minimality definition is independent of A and \mathcal{B} used in the representation but depends on only \mathcal{K} . That said, when $\mathcal{K}_1 \neq \mathcal{K}_2$, \mathcal{K}_1 -minimal inequalities might differ significantly from \mathcal{K}_2 -minimal inequalities even

when $\overline{\text{conv}}(\mathcal{S}(A_1, \mathcal{K}_1, \mathcal{B}_1)) = \overline{\text{conv}}(\mathcal{S}(A_2, \mathcal{K}_2, \mathcal{B}_2))$. For example, suppose $\mathcal{K}_1 \subset \mathcal{K}_2$ as well as $\overline{\text{conv}}(\mathcal{S}(A_1, \mathcal{K}_1, \mathcal{B}_1)) = \overline{\text{conv}}(\mathcal{S}(A_2, \mathcal{K}_2, \mathcal{B}_2))$; then all \mathcal{K}_1 -minimal inequalities are also \mathcal{K}_2 -minimal but not vice versa. This, in the light of Proposition 3, demonstrates how the selection of cone \mathcal{K} in disjunctive conic representations is critical in identifying more refined dominance relations among valid inequalities. We consequently deduce that minimality should be defined with respect to the smallest cone \mathcal{K} as it encodes the largest amount of structural information. See [15, Rem 1, 5, and 7 and Sec 2.2] as well.

Usually, additional structural information of a problem is available in the form of a convex or polyhedral relaxation; and such information can be encoded in a cone \mathcal{K} in a lifted space by a single additional variable through homogenization as described in [15, Ex 4].

\mathcal{K} -minimality and tightness

A first and foremost desirable feature of a strong valid inequality $\langle \mu, x \rangle \geq \mu_0$ is its tightness, i.e., $\mu_0 = \vartheta(\mu)$. The concepts of tightness and \mathcal{K} -minimality are intrinsically different. Still, for certain vectors $\mu \in \mathbb{E}$, \mathcal{K} -minimality not only immediately implies tightness of the inequality but also determines the sign of $\vartheta(\mu)$.

Proposition 4 ([15, Prop 3]). *Let $\langle \mu, x \rangle \geq \mu_0$ with $\mu \in \pm\mathcal{K}^*$ be a \mathcal{K} -minimal inequality. Then $\mu_0 = \vartheta(\mu)$; and furthermore, $\mu \in \mathcal{K}^*$ (resp., $\mu \in -\mathcal{K}^*$) implies $\vartheta(\mu) > 0$ (resp., $\vartheta(\mu) < 0$).*

However, there are \mathcal{K} -minimal inequalities with $\mu \notin \pm\mathcal{K}^*$ that are not tight. In fact, a pathology occurs when $\text{Ker}(A) \cap \text{int}(\mathcal{K}) \neq \emptyset$ and $\mu \in \text{Im}(A^*)$.

Proposition 5 ([15, Prop 4]). *Suppose $\text{Ker}(A) \cap \text{int}(\mathcal{K}) \neq \emptyset$. Then, for any $\mu \in \text{Im}(A^*)$, the inequality $\langle \mu, x \rangle \geq \mu_0$ with any $\mu_0 \in (-\infty, \vartheta(\mu)]$ is \mathcal{K} -minimal; yet only one of these is tight.*

Because tightness has a direct characterization through $\vartheta(\mu)$, we keep it as a separate consideration.

Algebraic necessary conditions

\mathcal{K} -minimality concept has a number of algebraic implications.

Any nontrivial valid inequality $\langle \mu, x \rangle \geq \mu_0$ with $\mu_0 \in \mathbb{R}$ has to satisfy $\mu \in \mathcal{K}^* + \text{Im}(A^*)$ (see [15, Prop 6]). Based on this, we can then associate with such an inequality the following nonempty set

$$D_\mu := \{\lambda \in \mathbb{R}^m : \mu - A^*\lambda \in \mathcal{K}^*\}.$$

Because of their structure and relation to cut-generating functions, we refer to these sets D_μ as *cut-generating sets*. Given a nontrivial valid inequality, there is a unique set D_μ associated with it. Yet, it is possible to have two distinct vectors μ' and μ yielding the same set $D_\mu = D_{\mu'}$ (see [15, Ex 8]).

The support function σ_{D_μ} plays an important role in our analysis. First of all, given $\mu \in \mathcal{K}^* + \text{Im}(A^*)$, σ_{D_μ} is helpful in determining a lower bound on $\vartheta(\mu)$, i.e., $\vartheta(\mu) \geq \inf_{b \in \mathcal{B}} \sigma_{D_\mu}(b)$ and thus ensuring the validity of $\langle \mu, x \rangle \geq \mu_0$ for $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$. For $\mathcal{K} = \mathbb{R}_+^n$, this result was first proven in [14, Thm 9]. Below, we provide its refinement and generalization for arbitrary regular cones \mathcal{K} .

Proposition 6 ([15, Prop 7 and 8]). *For any $\mu \in \mathcal{K}^* + \text{Im}(A^*)$, $\vartheta(\mu) \geq \inf_{b \in \mathcal{B}} \sigma_{D_\mu}(b)$. Moreover, when at least one of the following conditions holds: (i) \mathcal{K} is polyhedral, (ii) $\text{Ker}(A) \cap \text{int}(\mathcal{K}) \neq \emptyset$, (iii) $\mu \in \text{int}(\mathcal{K}^*) + \text{Im}(A^*)$, we have $\vartheta(\mu) = \inf_{b \in \mathcal{B}} \sigma_{D_\mu}(b)$.*

For any nontrivial valid inequality $\langle \mu, x \rangle \geq \mu_0$, there exists at least one $z \in \text{Ext}(\mathcal{K})$ such that $\sigma_{D_\mu}(Az) = \langle \mu, z \rangle$ (see [15, Lem 2, Cor 3, and Prop 9]). Further, there is a much more elegant connection between \mathbb{R}_+^n -sublinear inequalities and the support functions of cut-generating sets D_μ . This has striking consequences that we will comment more on later.

A key necessary condition for \mathcal{K} -minimality is based on a certain non-expansiveness property. For this, we introduce the cone of $\mathcal{K}^* - \mathcal{K}^*$ positive linear maps given by $\mathcal{F}_\mathcal{K} := \{(Z : \mathbb{E} \rightarrow \mathbb{E}) : Z \text{ is a linear map, and } Z^*v \in \mathcal{K} \ \forall v \in \mathcal{K}\}$, where Z^* denotes the conjugate linear map of Z .

Proposition 7 ([15, Prop 5]). *A valid inequality $\langle \mu, x \rangle \geq \mu_0$ is \mathcal{K} -minimal only if $\mu - Z\mu \notin \mathcal{K}^* \setminus \{0\}$ for all $Z \in \mathcal{F}_\mathcal{K}$ such that $AZ^* = A$.*

Description of $\mathcal{F}_\mathcal{K}$, unfortunately, can be rather nontrivial. For example, deciding whether a given linear map takes \mathbb{S}_+^n to itself is an NP-Hard optimization problem [4]. Because of the general difficulty of working with $\mathcal{F}_\mathcal{K}$ and thereby verifying the

necessary condition for \mathcal{K} -minimality stated above, we next consider an appropriate relaxation of this condition and introduce the class of \mathcal{K} -sublinear inequalities.

Definition 8. *Given $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$, a valid inequality $\langle \mu, x \rangle \geq \mu_0$ is \mathcal{K} -sublinear if for all $\alpha \in \text{Ext}(\mathcal{K}^*)$ it satisfies $0 \leq \langle \mu, u \rangle$ for all u such that $Au = 0$ and $\langle \alpha, v \rangle u + v \in \mathcal{K} \ \forall v \in \text{Ext}(\mathcal{K})$.*

Every \mathcal{K} -minimal inequality is also \mathcal{K} -sublinear [15, Thm 1]. Without any technical assumptions such as Assumption 2, the existence, sufficiency, properties of \mathcal{K} -sublinear inequalities, and their connection with CGFs are pursued further in [16].

Sufficient conditions

The following sufficient conditions complement our necessary conditions and also suggest practical ways of verifying \mathcal{K} -sublinearity and/or \mathcal{K} -minimality of inequalities.

Proposition 9 ([15, Prop 10]). *Let $\langle \mu, x \rangle \geq \mu_0$ be a nontrivial valid inequality. If there exists a collection I of vectors $x^i \in \text{Ext}(\mathcal{K})$ such that $\sigma_{D_\mu}(Ax^i) = \langle \mu, x^i \rangle$ for all $i \in I$ and $\sum_{i \in I} x^i \in \text{int}(\mathcal{K})$, then $\langle \mu, x \rangle \geq \mu_0$ is \mathcal{K} -sublinear.*

Proposition 10 ([15, Prop 11]). *Suppose Assumption 2 holds. Consider a valid inequality $\langle \mu, x \rangle \geq \mu_0$. If there exists a collection I of vectors $x^i \in \mathcal{K}$ such that $\sum_{i \in I} x^i \in \text{int}(\mathcal{K})$, $Ax^i \in \mathcal{B}$ and $\langle \mu, x^i \rangle = \mu_0$, then $\langle \mu, x \rangle \geq \mu_0$ is \mathcal{K} -minimal.*

Proposition 10 in particular states that a valid inequality is \mathcal{K} -minimal whenever the inequality is satisfied as equality at a point at the intersection of $\text{int}(\mathcal{K})$ and $\text{conv}(\mathcal{S}(A, \mathcal{K}, \mathcal{B}))$. For MILP problems, this resembles a sufficient condition for an inequality to be facet defining. Nonetheless, conic minimality notion is much weaker than extremality.

Cut generating functions

Given a nonconvex set $\mathcal{B} \subset \mathbb{R}^m$, an important class of problems is defined by the infinite family of sets of form $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ given by any realization of $n \in \mathbb{N}$ and $A \in \mathbb{R}^{m \times n}$. This family of sets is characterized by solely \mathcal{B} which, in its most general form, is assumed to be a closed set satisfying $0 \notin \mathcal{B}$. Then

$0 \notin \overline{\text{conv}}(\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B}))$ follows easily [7, Lem 2.1]. This motivates the definition of *cut-generating functions* (CGFs) — a priori formulas to generate cuts that separate the origin from the convex hull of any instance of $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ determined by n and A :

Definition 11. *Given a nonempty and closed set $\mathcal{B} \subset \mathbb{R}^m$ satisfying $0 \notin \mathcal{B}$, a cut-generating function for \mathcal{B} is a function $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$ such that the inequality given by $\sum_{i=1}^n \psi(A_i)x_i \geq 1$ is valid for $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ where A_i is the i -th column of the matrix A , for any natural number $n \in \mathbb{N}$ and any matrix $A \in \mathbb{R}^{m \times n}$.*

This framework has its roots in Gomory functions [9] and Gomory and Johnson's infinite group relaxations studied in the MILP context [10, 13, 1]. Recent work has focused on a variety of structural assumptions on \mathcal{B} such as \mathcal{B} is a general lattice [6], \mathcal{B} is composed of lattice points contained in a rational polyhedron [8, 3], and \mathcal{B} is a closed set [7], and demonstrated strong connections between \mathbb{R}_+^n -minimal inequalities and CGFs obtained from the gauge functions of maximal lattice-free sets.

This framework and CGFs are intimately connected to our results on \mathbb{R}_+^n -sublinear inequalities and their relation with support functions of cut-generating sets. We discuss this next; see [15, Sec 4.3] for a detailed account.

Decades ago, Johnson [14] considered $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ with $\mathcal{K} = \mathbb{R}_+^n$ and introduced *subadditive inequalities*. These inequalities are equivalent to the \mathbb{R}_+^n -sublinear inequalities (see e.g., [15, Rem 9]). We restate their definition below:

Definition 12. *Given $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$, a valid inequality $\langle \mu, x \rangle \geq \mu_0$ is \mathbb{R}_+^n -sublinear if for all $i \in [n]$, $\langle \mu, u \rangle \geq 0$ holds for all u such that $Au = 0$ and $u + e_i \in \mathbb{R}_+^n$ where e_i denotes the i^{th} unit vector in \mathbb{R}^n .*

A fundamental result of Johnson [14, Thm 10] asserts that the cut coefficient vector of any \mathbb{R}_+^n -sublinear inequality is generated by its support function σ_{D_μ} , which is also piecewise linear. Our Proposition 9 complements this result and proves that the conditions of Proposition 9 are necessary and sufficient for \mathbb{R}_+^n -sublinearity. That is, a valid inequality $\langle \mu, x \rangle \geq \mu_0$ is \mathbb{R}_+^n -sublinear and tight if and only if

its support function σ_{D_μ} generates its coefficient vector μ and its right-hand side value μ_0 . The following theorem summarizes these results [14, Thm 10] and [15, Props 6, 8, and 10, and Thm 4] for $\mathcal{K} = \mathbb{R}_+^n$; see also [15, Rem 10 and 11].

Theorem 13. *Consider $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$. Then any nontrivial valid inequality $\langle \mu, x \rangle \geq \mu_0$ satisfies $\mu \in \mathbb{R}_+^n + \text{Im}(A^\top)$ and $\vartheta(\mu) = \inf_{b \in \mathcal{B}} \sigma_{D_\mu}(b) \geq \mu_0 > -\infty$. Moreover, $\langle \mu, x \rangle \geq \mu_0$ is \mathbb{R}_+^n -sublinear if and only if it is valid ($\vartheta(\mu) \geq \mu_0$) and $\mu_i = \sigma_{D_\mu}(A_i)$ for all $i \in [n]$ where A_i denotes the i -th column of the matrix A .*

More recently, Kılınç-Karzan and Steffy [16] noted that support function σ_{D_μ} associated with any nontrivial valid inequality $\langle \mu, x \rangle \geq \mu_0$ can be utilized in obtaining a stronger and \mathbb{R}_+^n -sublinear inequality.

Proposition 14 ([16, Prop 3]). *Any nontrivial valid inequality $\langle \mu, x \rangle \geq \mu_0$ for $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ is equivalent to or dominated by an \mathbb{R}_+^n -sublinear inequality given by $\sum_{i=1}^n \sigma_{D_\mu}(A_i)x_i \geq \inf_{b \in \mathcal{B}} \sigma_{D_\mu}(b) \geq \mu_0$ where the domination is with respect to the cone $\mathcal{K} = \mathbb{R}_+^n$.*

Thus, \mathbb{R}_+^n -sublinear inequalities are *always sufficient* to describe $\overline{\text{conv}}(\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B}))$. Proposition 14 also inspired the following definition of *relaxed* CGFs as the support functions of nonempty sets D in [16]:

Definition 15. *Given $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ and a set $\emptyset \neq D \subset \mathbb{R}^m$, we say that the support function $\sigma_D : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ of D is a relaxed cut-generating function for $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$.*

Clearly, the support functions associated with \mathbb{R}_+^n -sublinear inequalities are relaxed CGFs. Although the relaxed CGFs such as σ_{D_μ} are seemingly tied to a particular set $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ defined by fixed n, A , and \mathcal{B} , the subadditivity of these support functions permits us at once to generate valid inequalities for any instance $\mathcal{S}(A', \mathbb{R}_+^{n'}, \mathcal{B})$ with data $A' \in \mathbb{R}^{m \times n'}$, i.e., varying n and A , as long as the set \mathcal{B} is kept the same.

Proposition 16 ([16, Prop 4]). *Suppose $\mathcal{B} \subset \mathbb{R}^m$ is given. Let $\sigma_D(\cdot)$ be a relaxed CGF for $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ associated with a nonempty set $D \subset \mathbb{R}^m$. Then, the inequality $\sum_{i=1}^{n'} \sigma_D(A'_i)x_i \geq \inf_{b \in \mathcal{B}} \sigma_D(b)$ is valid for any $\mathcal{S}(A', \mathbb{R}_+^{n'}, \mathcal{B})$ where the dimension n' and the matrix $A' \in \mathbb{R}^{m \times n'}$ are arbitrary, and A'_i denotes the i -th column of the matrix A' for all $i \in [n']$.*

When \mathcal{B} is a closed set satisfying $0 \notin \mathcal{B}$, Proposition 16 essentially binds together relaxed CGFs and regular CGFs. For a relaxed CGF σ_D to be a regular CGF, we need to ensure: (i) $\inf_{b \in \mathcal{B}} \sigma_D(b) \geq 1$ and (ii) σ_D is finite valued. All \mathbb{R}_+^n -sublinear inequalities of form $\langle \mu, x \rangle \geq 1$ immediately have $\inf_{b \in \mathcal{B}} \sigma_{D_\mu}(b) \geq 1$. Then we infer from Theorem 13 and Propositions 14 and 16 that without any structural or technical assumptions, the relaxed CGFs, specifically the ones associated with the sets D_μ of \mathbb{R}_+^n -sublinear inequalities, are sufficient to generate all necessary inequalities for the description of $\overline{\text{conv}}(\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B}))$ for all choices of n and A . When the set \mathcal{B} is composed of lattice points, a classical result [6, Thm 1.2] states that all \mathbb{R}_+^n -minimal inequalities are generated by sublinear functions which are also piecewise linear. Johnson's [14] analysis along with ours easily recovers this. Sufficiency of regular CGFs for generating all cuts separating the origin in the case of general \mathcal{B} relies on additional structural assumptions [7, Ex 6.1 and Thm 6.3]. This is in contrast to the sufficiency of relaxed CGFs for any \mathcal{B} . In this respect, the main challenge in transforming a relaxed CGF σ_D into a regular CGF resides in ensuring finite valuedness of σ_D while maintaining $\inf_{b \in \mathcal{B}} \sigma_D(b) \geq 1$. Whenever σ_D is not finite valued, i.e., D is unbounded, under certain assumptions, the relaxed CGFs obtained from bounded sets $\hat{D} \subset D$ offer a solution for this challenge.

The sufficiency of CGFs for describing the convex hulls of disjunctive conic sets is intrinsically related to the strong duality theory for integer programs. Morán et al. [19, Thm 2.4] has extended the strong duality theory for MILPs to MICPs of a specific form. Under technical assumptions, these theorems assert that for every integer programming instance, there is a dual problem achieving zero duality gap where the 'dual variables' are finite-valued subadditive functions that are nondecreasing with respect to the underlying cone. These functions indeed act locally on each variable x_i and produce cut coefficient μ_i by considering only the data A_i associated with x_i ; therefore, they are simply CGFs. Then the sufficiency of CGFs for generating all cuts of the form $\langle \mu, x \rangle \geq 1$ follows from strong MICP duality theorem. Nevertheless, not only the strong duality results for MILPs and MICPs rely on some technical assumptions but also the sets $\mathcal{S}(A, \mathbb{R}_+^n, \mathcal{B})$ represent-

ing MILPs and the specific form of MICPs from [19] impose a specific structure on \mathcal{B} (see [15, Ex 3]). Additional discussion relating [19] to CGFs is given in [15, Rem 12] and [16, Rem 2].

Our results naturally capture some of the earlier results from the MILP setup and generalize them to the cases with arbitrary nonconvex sets \mathcal{B} . That said, our study also reveals some problems associated with such a CGF-based view that treats the data associated with each individual variable independently in the case of general regular cones other than the nonnegative orthant. Namely, [15, Ex 8 and Rem 12] features an extreme inequality for a set $\mathcal{S}(A, \mathcal{K}, \mathcal{B})$ with $\mathcal{K} = \mathbb{L}^3$ that cannot be generated by any CGF or relaxed CGF.

Final remarks

In the context of disjunctive conic sets, characterization of \mathcal{K} -minimal and tight inequalities has underlied the development of structured convex (or conic representable) cuts for two-term linear disjunctions applied to a second-order cone (see [18]). The flexible representation structure offered by disjunctive conic sets can easily allow us to pursue a similar principled study of other simple, yet fundamental, nonconvex sets defined by multi-term disjunctions or quadratics on regular cones. In this regard, characterizations of extreme inequalities beyond \mathcal{K} -minimality are very appealing as well.

We also hope that the understanding and connections we built on CGFs and relaxed CGFs will be instrumental in understanding when minimal or extreme CGFs will produce strong linear inequalities such as facets for given problem instances. On a related note, the sufficiency of CGFs to generate all valid inequalities for the convex hull description of disjunctive sets or all cuts that separate the origin from the convex hull of disjunctive sets is an indispensable question for the justification of this research focus on CGFs. Along these lines, our results have recently contributed to the foundation of the most general conditions guaranteeing the sufficiency of CGFs for general \mathcal{B} [17].

Acknowledgments

I would like to thank the Young Researcher Prize Committee consisting of Nick Sahinidis (chair), Daniel Bienstock, Sam Burer, and Andrew Schaefer; it is truly a great honor to have been selected as a recipient of this prize.

This research was supported in part by NSF CAREER grant 1454548.

REFERENCES

- [1] A. Bachem, E. Johnson, and R. Schrader. A characterization of minimal valid inequalities for mixed integer programs. *Operations Research Letters*, 1:63–66, 1982.
- [2] E. Balas. Disjunctive programming. *Annals of Discrete Mathematics*, 5:3–51, 1979.
- [3] A. Basu, M. Conforti, G. Cornuéjols, and G. Zambelli. Minimal inequalities for an infinite relaxation of integer programs. *SIAM Journal on Discrete Mathematics*, 24(1):158–168, 2010.
- [4] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.
- [5] C. Blair. Minimal inequalities for mixed integer programs. *Discrete Mathematics*, 24:147–151, 1978.
- [6] V. Borozan and G. Cornuéjols. Minimal valid inequalities for integer constraints. *Mathematics of Operations Research*, 34(3):538–546, 2009.
- [7] M. Conforti, G. Cornuéjols, A. Daniilidis, C. Lemaréchal, and J. Malick. Cut-generating functions and s -free sets. *Mathematics of Operations Research*, 40(2):276–301, 2015.
- [8] S. Dey and L. Wolsey. Constrained infinite group relaxation of MIPs. *SIAM Journal on Optimization*, 20:2890–2912, 2010.
- [9] R. Gomory. An algorithm for integer solutions to linear programs. In R. Graves and P. Wolfe, editors, *Recent Advances in Mathematical Programming*, pages 269–302. McGraw-Hill, New York, NY, 1963.
- [10] R. Gomory and E. Johnson. Some continuous functions related to corner polyhedra. *Mathematical Programming*, 3:23–85, 1972.
- [11] R. Jeroslow. Cutting plane theory: Algebraic methods. *Discrete Mathematics*, 23:121–150, 1978.
- [12] R. Jeroslow. Minimal inequalities. *Mathematical Programming*, 17:1–15, 1979.
- [13] E. Johnson. On the group problem for mixed integer programming. *Mathematical Programming*, 2:137–179, 1974.
- [14] E. Johnson. Characterization of facets for multiple right-hand side choice linear programs. *Mathematical Programming Study*, 14:137–179, 1981.
- [15] F. Kılınç-Karzan. On minimal inequalities for mixed integer conic programs. *Mathematics of Operations Research*, 41(2):477–510, 2015.
- [16] F. Kılınç-Karzan and D. E. Steffy. On sublinear inequalities for mixed integer conic programs. *Mathematical Programming*, 2016. <http://dx.doi.org/10.1007/s10107-015-0968-0>.
- [17] F. Kılınç-Karzan and B. Yang. Sufficient conditions and necessary conditions for the sufficiency of cut-generating functions. Technical report, December 2015. <http://www.andrew.cmu.edu/user/fkilinc/files/draft-sufficiency-web.pdf>.
- [18] F. Kılınç-Karzan and S. Yıldız. Two term disjunctions on the second-order cone. *Mathematical Programming*, 154:463–491, 2015.
- [19] D. Moran R, S. Dey, and J. Vielma. A strong dual for conic mixed-integer programs. *SIAM Journal on Optimization*, 22(3):1136–1150, 2012.

Semidefinite Relaxations for Nonlinear Optimization Problems

Somayeh Sojoudi

Department of EECS

and Javad Lavaei

Department of IEOR

University of California, Berkeley

USA

sojoudi@berkeley.edu and lavaei@berkeley.edu

Convex optimization has found a wide range of applications across engineering and economics [1]. In the past several years, great effort has been devoted to casting many real-world problems as convex optimization problems. Nevertheless, several classes



Nick Sahinidis and Somayeh Sojoudi



Nick Sahinidis and Javad Lavaei

of optimization problems, including polynomial optimization and quadratically constrained quadratic program (QCQP) as a special case, are nonlinear, non-convex, and NP-hard in the worst case. In particular, there is no known effective optimization technique for integer and combinatorial optimization as a small subclass of QCQP. Given a non-convex optimization problem, there are several techniques to find a solution that is locally optimal. However, seeking a global or near-global solution in polynomial time is a daunting challenge. There is a large body of literature on nonlinear optimization witnessing the complexity of this problem.

To reduce the computational complexity of a non-convex optimization problem, several convex relaxation methods based on linear matrix inequality (LMI), semidefinite programming (SDP) and second-order cone programming (SOCP) have gained popularity [2, 3]. These techniques enlarge the possibly non-convex feasible set into a convex set that is characterizable via convex functions, and then provide the exact or a lower bound on the optimal objective value associated with a global solution. The effectiveness of this technique has been substantiated in different contexts [4, 5, 6, 7, 8, 9, 10]. The SDP relaxation converts an optimization problem with a vector variable to a convex program with a matrix variable, via a lifting technique. The exactness of the relaxation can then be interpreted as the existence of a low-rank (e.g., rank-1) matrix solution for the SDP relaxation.

To explain the SDP relaxation technique, consider the polynomial optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \dots, m \end{aligned} \quad (1)$$

where f_0, \dots, f_m are arbitrary polynomial functions. This problem can be reformulated as

$$\begin{aligned} \min_{\tilde{x} \in \mathbb{R}^{\tilde{n}}} \quad & \tilde{x}^* M_0 \tilde{x} \\ \text{s.t.} \quad & \tilde{x}^* M_i \tilde{x} \leq 0, \quad i = 1, 2, \dots, \tilde{m} \end{aligned} \quad (2)$$

for some non-unique numbers \tilde{n} and \tilde{m} . In this problem, the variable \tilde{x} consists of multiple copies of the entries of x as well as some auxiliary parameters, and the matrices $M_0, M_1, \dots, M_{\tilde{m}}$ are all sparse [11]. Given a convex function $\mu(\cdot)$, consider the convex

program

$$\begin{aligned} \min_{W \in \mathbb{S}^{\tilde{n}}} \quad & \text{trace}\{M_0 W\} + \mu(W) \\ \text{s.t.} \quad & \text{trace}\{M_i W\} \leq 0, \quad i = 1, \dots, \tilde{m} \\ & W \succeq 0 \end{aligned} \quad (3)$$

where W is a matrix variable. We refer to this problem as *penalized SDP*. This problem is equivalent to the QCQP problem (2) under the additional rank constraint $\text{rank}\{W\} = 1$, provided that $\mu(W) = 0$ (note that W plays the role of $\tilde{x}\tilde{x}^*$). Hence, penalized SDP is a convex relaxation of the original problem (1) whenever $\mu(W) = 0$, but it is advantageous to design a nonzero function $\mu(W)$ to compensate for the dropped rank constraint $\text{rank}\{W\} = 1$. It is shown in [11] that the non-unique conversion from (1) to (2) can be carried out in such a way that the penalized SDP will have a low-rank solution W^{opt} for a wide class of penalty functions $\mu(\cdot)$ (including $\mu(W) = 0$). This result has two implications. First, the NP-hardness of various subclasses of polynomial optimization, e.g., combinatorial optimization, is only related to the existence of a low-rank SDP solution that is not rank 1. Second, by approximating the low-rank solution of the SDP relaxation with a rank-1 matrix, an approximate solution of the original problem may be obtained whose closeness to the global solution can also be upper bounded. Some of our recent results on low-rank SDP relaxations together with their applications in electrical power networks will be explained next.

Notations Before providing a survey of our recent results, the notations used in this paper will be defined below. \mathbb{R} , \mathbb{C} , \mathbb{S}^n and \mathbb{H}^n denote the sets of real numbers, complex numbers, $n \times n$ symmetric matrices and $n \times n$ Hermitian matrices, respectively. $\text{Re}\{W\}$, $\text{Im}\{W\}$, $\text{rank}\{W\}$ and $\text{trace}\{W\}$ denote the real part, imaginary part, rank and trace of a matrix W , respectively. The notation $W \succeq 0$ means that W is Hermitian and positive semidefinite. Given a matrix W , its (l, m) entry is denoted as W_{lm} . The superscript $(\cdot)^{\text{opt}}$ is used to show the globally optimal value of an optimization parameter. The symbol $(\cdot)^*$ represents the conjugate transpose operator. Given an undirected graph \mathcal{G} , the notation $i \in \mathcal{G}$ means that i is a vertex of \mathcal{G} , and $(i, j) \in \mathcal{G}$ means that (i, j) is an edge of \mathcal{G} and besides $i < j$.

1. Motivation: Power Optimization Problems

The real-time operation of an electric power network depends heavily on several large-scale optimization problems solved from every few minutes to every year. State estimation, optimal power flow (OPF), security constrained OPF, unit commitment, transmission planning, sizing of capacitor banks, and network reconfiguration are some fundamental optimization problems solved for transmission and distribution networks. Since these different problems have all been built upon the power flow equations, they are referred to as *OPF-based optimization* in this paper. Regardless of their large-scale nature, it is a daunting challenge to solve these problems efficiently. This is a consequence of the nonlinearity/non-convexity created by two different sources: (i) discrete variables such as the ratio of a tap-changing transformer, the on/off status of a line switch, or the commitment parameter of a generator, and (ii) the laws of physics. Issue (i) is more or less universal and researchers in many fields of study have proposed various sophisticated methods to handle integer variables. In contrast, Issue (ii) is pertinent to power systems, and it demands new specialized techniques and approaches. More precisely, complex power being a quadratic function of complex bus voltages imposes quadratic constraints on OPF-based optimization problems. Issue (ii) makes these problems NP-hard and has the following implications [12]:

- The well-established numerical algorithms, e.g., Gradient descent, Newton's method and primal-dual algorithm, may only find non-global local minima.
- Given a local solution, it is hard to verify how close to a global minimum the solution is.
- These algorithms may not converge if they start from a bad initial point.
- These algorithms may suffer from the lack of numerical robustness.
- In the case of non-convergence, it is hard to determine whether the problem was infeasible or the initial guess for the solution was not good enough.

OPF is at the heart of Independent System Operator (ISO) power markets and vertically integrated utility dispatch [13]. This problem needs to be solved annually for system planning, daily for day-ahead commitment markets, and every 5-15 minutes for real-time market balancing. Due to the issues outlined above, the existing solvers for OPF-based optimization either make potentially very conservative approximations or deploy general-purpose local-search algorithms. For example, a linearized version of OPF, named DC OPF, is normally solved in practice, whose solution may not be physically meaningful due to approximating the laws of physics. Although OPF has been studied for 50 years, the algorithms deployed by ISOs suffer from several issues, which may incur tens of billions of dollars annually [13]. More sophisticated OPF-based problems are even harder to solve and may need much coarser approximations. As the power industry moves towards the upgrade of legacy grids into *smart grids*, new optimization problems emerge for both distribution and transmission systems, which are large-scale (with hundreds of thousands of variables) and may need to be solved on a short time scale (to cope with the intermittency and variability of renewable energy). The non-convexity issue worsens the situation greatly.

The power flow equations for a power network are quadratic in the complex voltage vector. Using this fact, it can be shown that all constraints of the above-mentioned OPF-based problems can often be cast as quadratic constraints after introducing certain auxiliary parameters. We have shown in a series of papers that an SDP relaxation is exact and finds global solutions for benchmarks examples of OPF-based optimization [14, 15, 16, 17, 18, 9, 19, 20, 21, 22, 23]. By leveraging the physics of power grids, it is also theoretically proven that the SDP relaxation is always exact for every distribution network and every transmission network containing a sufficient number of transformers (under some technical assumptions) [17, 24]. The papers [22] and [19] show that if the SDP relaxation is not exact due to the violation of certain assumptions, a penalized SDP relaxation would work for a carefully chosen penalty term, which leads to recovering a near-global solution. This technique is tested on several real-world grids and the outcome is partially reported in Table 1

Test cases	Near-optimal cost	Global optimality guarantee	Run time
Polish 2383wp	1874322.65	99.316%	529
Polish 2736sp	1308270.20	99.970%	701
Polish 2737sop	777664.02	99.995%	675
Polish 2746wop	1208453.93	99.985%	801
Polish 2746wp	1632384.87	99.962%	699
Polish 3012wp	2608918.45	99.188%	814
Polish 3120sp	2160800.42	99.073%	910

Table 1: Performance of penalized SDP for OPF (run time is in seconds).

(see our solver [25] for more details). It can be observed that the SDP relaxation has found operating points for the nationwide grid of Poland in different times of the year, where the global optimality guarantee of each solution is at least 99%, implying that the unknown global minima are at most 1% away from the obtained feasible solutions. The above observations show the significant potential of conic relaxation for structured optimization problems, such as those appearing in electrical grids. Some of our theoretical results on structured optimization problems will be outlined below.

2. Highly Structured Optimization

In this part, we offer some theoretical results on how structure helps. Consider the problem

$$\begin{aligned} \min_{x \in \mathbb{D}^n} \quad & x^* M_0 x \\ \text{s.t.} \quad & x^* M_i x \leq 0, \quad i = 1, 2, \dots, m \end{aligned} \quad (4)$$

where \mathbb{D} is either \mathbb{R} or \mathbb{C} . Note that every polynomial optimization problem can be reformulated as above, and therefore (4) includes a broad class of problems. Assume that M_0, \dots, M_m are symmetric matrices in the real case and Hermitian matrices in the complex case. An SDP relaxation of this non-convex optimization problem can be naturally found based on the discussion provided in the preceding section. This relaxation is exact if it has a rank-1 solution W^{opt} , in which case an optimal solution of problem (4) can be recovered from W^{opt} . To study

the exactness of the SDP relaxation, consider a *generalized weighted graph* \mathcal{G} with n nodes constructed as follows:

- Given every two nodes $i, j \in \{1, \dots, n\}$ such that $i \neq j$, there exists an edge between the nodes i and j if and only if the (i, j) off-diagonal entry of at least one of the matrices M_0, \dots, M_m is nonzero.
- For every $(i, j) \in \mathcal{G}$, the union of the nonzero $(i, j)^{\text{th}}$ entries of M_0, \dots, M_m will be assigned as a *weight set* to the edge (i, j) .

This graph captures the sparsity of the problem and the connection among its coefficients. It is desirable to relate the exactness of the SDP relaxation to certain properties of the generalized weighted graph \mathcal{G} . To this end, we will introduce the notion of “sign-definite set.” A finite set $\mathcal{T} \subset \mathbb{R}$ is said to be *sign definite with respect to* \mathbb{R} if its elements are either all negative or all nonnegative. \mathcal{T} is called *negative* if its elements are negative and is called *positive* if its elements are nonnegative. A finite set $\mathcal{T} \subset \mathbb{C}$ is said to be *sign definite with respect to* \mathbb{C} if when the sets \mathcal{T} and $-\mathcal{T}$ are mapped into two collections of points in \mathbb{R}^2 , then there exists a line separating the two sets (the elements of the sets are allowed to lie on the line). The following results hold in both real and complex cases (see [24] for more details):

Real case $\mathbb{D} = \mathbb{R}$: The SDP relaxation is exact if two groups of conditions are satisfied for \mathcal{G} :

- *Edge conditions:* The weight set for every edge of \mathcal{G} is sign definite with respect to \mathbb{R} .
- *Cycle conditions:* Every cycle of \mathcal{G} has an even number of edges with positive weight sets.

The above conditions are naturally satisfied in three special cases: (1) \mathcal{G} is acyclic with sign-definite edge weight sets, (2) \mathcal{G} is bipartite with positive weight sets, (3) \mathcal{G} is arbitrary with negative weight sets. If the SDP relaxation is not exact, it still has a low-rank solution for a broad class of graphs. In particular, if the edge conditions are satisfied but some of the cycle conditions are violated, then the SDP relaxation has a solution whose rank is upper bounded by $n - R_{\min}(\mathcal{G})$, where $R_{\min}(\mathcal{G})$ denotes the minimum positive-semidefinite rank of the graph. The number

$n - R_{\min}(\mathcal{G})$ turns out to be small for a large class of graphs.

Complex case $\mathbb{D} = \mathbb{C}$: The SDP relaxation is exact if all weight sets are sign definite with respect to \mathbb{C} and one of the following conditions is satisfied:

- \mathcal{G} is acyclic.
- \mathcal{G} is bipartite and weakly cyclic with positive and negative real weight sets (a graph is called *weakly cyclic* if its cycles are edge disjoint).
- \mathcal{G} is a weakly cyclic graph with only imaginary weight sets such that all elements in each set have the same sign for their imaginary parts.
- \mathcal{G} can be decomposed as a union of edge-disjoint subgraphs in an acyclic way such that each subgraph satisfies one of the structural properties (i)-(iii) stated above.

As shown in [24], the above results are valid for a large class of optimization problems, which goes far beyond (4) (it includes polynomial, exponential and logarithmic problems).

2.1 Applications in Power Systems

A majority of real-world optimization problems can be regarded as “optimization problems with graph structures”, meaning that each of those problems has an underlying graph structure characterizing a physical system. For example, optimization problems in circuits, antenna systems and communication networks fall within this category. Then, the question of interest is: how is the computational complexity of an optimization problem related to the structure of the system over which the optimization problem is performed? This question will be explored here in the context of electrical power grids. Consider an arbitrary AC power network with n nodes (known as buses). For every edge (i, j) of the network, the nodes i and j are connected to each other via a transmission line with the admittance $g_{ij} + b_{ij}i$ (note that the symbol “ i ” denotes the imaginary unit). Assume that each node of the network is connected to an external device, which exchanges electrical power with the power network.

Figure 1(a) exemplifies a sample power network in which two external devices generate power while the

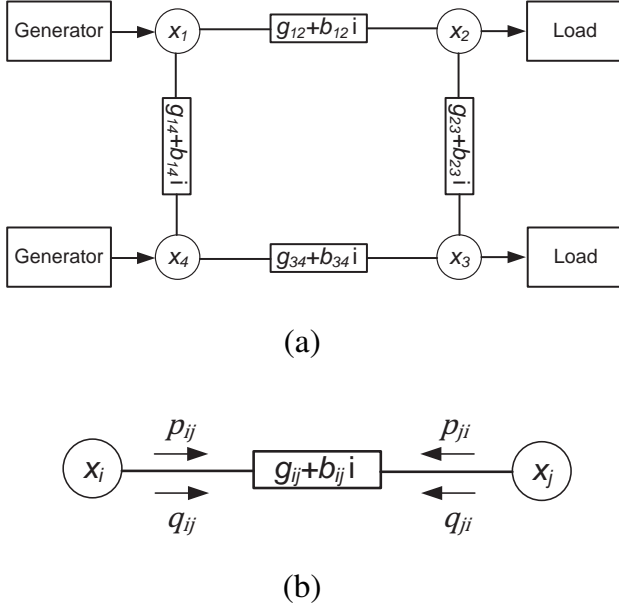


Figure 1: (a) An example of a power network; (b) this figure illustrates that each transmission line has four flows.

remaining ones consume power. As shown in Figure 1(b), each line (i, j) of the network is associated with four power flows:

- p_{ij} : Active power entering the line from node i
- p_{ji} : Active power entering the line from node j
- q_{ij} : Reactive power entering the line from node i
- q_{ji} : Reactive power entering the line from node j

Note that $p_{ij} + p_{ji}$ and $q_{ij} + q_{ji}$ represent the active and reactive losses incurred in the line. Let x_i denote the complex voltage (phasor) for node i of the network. One can write:

$$\begin{aligned} p_{ij}(x) &= \text{Re} \{x_i(x_i - x_j)^*(g_{ij} - b_{ij}i)\}, \\ p_{ji}(x) &= \text{Re} \{x_j(x_j - x_i)^*(g_{ij} - b_{ij}i)\}, \\ q_{ij}(x) &= \text{Im} \{x_i(x_i - x_j)^*(g_{ij} - b_{ij}i)\}, \\ q_{ji}(x) &= \text{Im} \{x_j(x_j - x_i)^*(g_{ij} - b_{ij}i)\}. \end{aligned}$$

Note that since the flows all depend on x , the argument x has been added to the above equations (e.g., $p_{ij}(x)$ instead of p_{ij}). The flows $p_{ij}(x)$, $p_{ji}(x)$, $q_{ij}(x)$

and $q_{ji}(x)$ can all be expressed in terms of $|x_i|^2$, $|x_j|^2$ and $\text{Re} \{c_{ij}^{(k)} x_i x_j^*\}$ for $k = 1, 2, 3, 4$, where

$$\begin{aligned} c_{ij}^{(1)} &= -g_{ij} + b_{ij}i, & c_{ij}^{(2)} &= -g_{ij} - b_{ij}i, \\ c_{ij}^{(3)} &= b_{ij} + g_{ij}i, & c_{ij}^{(4)} &= b_{ij} - g_{ij}i. \end{aligned}$$

Define $p(x)$ as the vector of all active flows $p_{ij}(x)$ and $p_{ji}(x)$ for every line (i, j) of the network. Likewise, define $q(x)$ as the vector of all line reactive flows. Consider the optimization problem

$$\begin{aligned} \min_{x \in \mathbb{C}^n} \quad & h_0(p(x), q(x), y(x)) \\ \text{s.t.} \quad & h_j(p(x), q(x), y(x)) \leq 0, \quad j = 1, 2, \dots, m \end{aligned} \quad (5)$$

for given functions h_0, \dots, h_m , where $y(x)$ is the vector of squared voltage magnitudes $|x_i|^2$ for $i = 1, 2, \dots, n$. Assume that the function $h_j(\cdot, \cdot, \cdot)$ accepting three arguments (inputs) is monotonic with respect to its first and second vector arguments, for $j = 0, \dots, m$. The above optimization problem aims to optimize the flows in a power grid. The constraints of this optimization problem account for network, technological and physical constraints. For example, they limit line flows, voltage magnitudes, power delivered to each load, and power supplied by each generator. Observe that $p(x)$ and $q(x)$ are both quadratic in x . The SDP relaxation method introduced before can be used to eliminate the effect of quadratic terms, which replaces $p(x)$, $q(x)$ and $y(x)$ with linear functions of a matrix W (playing the role of xx^*). To study under what conditions the relaxed matrix optimization problem is exact (or has a rank-1 solution W^{opt}), we can map the structure of the problem into a generalized weighted graph \mathcal{G} . This graph has the same topology as the physical power network, where the weight set for each edge (i, j) is equal to $\{c_{ij}^{(1)}, c_{ij}^{(2)}, c_{ij}^{(3)}, c_{ij}^{(4)}\}$. A customary transmission line is a passive device with nonnegative resistance and inductance, implying the inequalities $g_{ij} \geq 0$ and $b_{ij} \leq 0$. As a result of this property, the set $\{c_{ij}^{(1)}, c_{ij}^{(2)}, c_{ij}^{(3)}, c_{ij}^{(4)}\}$ turns out to be sign definite with respect to \mathbb{C} . Using a general version of the result outlined in the preceding section, it can be concluded that the proposed matrix relaxation is exact as long as \mathcal{G} is acyclic (note that most distribution networks are acyclic). This result also holds for cyclic (transmission) networks having a sufficient

number of phase shifters [17]. This implies that the physics of power networks reduce the computational complexity.

3. Low Rank Solutions

In this section, we study the existence of low-rank solutions for SDP relaxations. This is helpful for retrieving a near-global solution of the original non-convex problem in the case where the relaxation is not exact.

3.1 Positive Semidefinite Matrix Completion

The low-rank positive semidefinite matrix completion problem aims to design the unknown entries of a partially filled matrix so that the resulting matrix becomes positive semidefinite with a minimum rank. This fundamental problem serves as a basis for studying the SDP relaxation for polynomial optimization problems. To introduce the problem, consider a simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with n vertices together with a known positive-definite matrix $\widehat{W} \in \mathbb{S}^n$ (the symbols \mathcal{V} and \mathcal{E} denote the vertex set and edge set of the graph). The goal is to solve the following optimization problem:

$$\min_{W \in \mathbb{S}^n} \quad \text{rank}\{W\} \quad (6a)$$

$$\text{s.t.} \quad W_{ij} = \widehat{W}_{ij}, \quad \forall (i, j) \in \mathcal{E} \quad (6b)$$

$$W_{kk} = \widehat{W}_{kk}, \quad \forall k \in \mathcal{V} \quad (6c)$$

$$W \succeq 0 \quad (6d)$$

Note that the matrix W inherits the values of its diagonal and those off-diagonal entries corresponding to the edges of \mathcal{G} from the given matrix \widehat{W} . This problem is difficult to tackle due to its non-convex objective function. To reduce the complexity of the problem, we will propose two convex relaxations based on the graph notions of OS and treewidth.

Definition 1. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $\mathcal{O} = \{o_k\}_{k=1}^s$ be a sequence of vertices of \mathcal{G} with s elements. Denote as \mathcal{G}_k the subgraph induced by $\{o_1, \dots, o_k\}$, for $k = 1, \dots, s$. Let \mathcal{G}'_k be the connected component of \mathcal{G}_k containing o_k . \mathcal{O} is called an OS-vertex sequence of \mathcal{G} if for every $k \in \{1, \dots, s\}$ there

exists a vertex $w_k \in \mathcal{V}$ with the following three properties:

1. w_k is a neighbor of o_k , i.e., $(o_k, w_k) \in \mathcal{E}$
2. w_k does not belong to the set $\{o_1, o_2, \dots, o_k\}$
3. w_k is not connected to any vertex in \mathcal{G}'_k other than o_k

Denote the maximum cardinality among all OS-vertex sequences of \mathcal{G} as $\text{OS}(\mathcal{G})$ [26].

To develop a convex relaxation for the matrix completion problem (6), let $\mathcal{G}^c = (\mathcal{V}^c, \mathcal{E}^c)$ be an arbitrary graph such that $\mathcal{V}^c = \mathcal{V}$ and $\mathcal{E}^c \cap \mathcal{E} = \emptyset$.

Convex Relaxation I: This problem is defined as

$$\min_{W \in \mathbb{S}^n} \quad \sum_{(i,j) \in \mathcal{E}^c} t_{ij} W_{ij} \quad (7a)$$

$$\text{s.t.} \quad W_{ij} = \widehat{W}_{ij}, \quad \forall (i, j) \in \mathcal{E} \quad (7b)$$

$$W_{kk} = \widehat{W}_{kk}, \quad \forall k \in \mathcal{V} \quad (7c)$$

$$W \succeq 0 \quad (7d)$$

where t_{ij} 's are arbitrary nonzero scalars.

As shown in [27], every solution of Convex Relaxation I, denoted as \widehat{W}^{opt} , satisfies the inequality

$$\text{rank}\{\widehat{W}^{\text{opt}}\} \leq n - \min_{\mathcal{G}^s} \left\{ \text{OS}(\mathcal{G}^s \cup \mathcal{G}^c) \mid \mathcal{G}^s \subseteq \mathcal{G} \right\} \quad (8)$$

where

- The notation $\mathcal{G}^s \subseteq \mathcal{G}$ means that \mathcal{G}^s is a graph with n vertices whose edge set is a subset of the edge set of \mathcal{G} .
- $\mathcal{G}^s \cup \mathcal{G}^c$ denotes the edge-wise union of the graphs \mathcal{G}^s and \mathcal{G}^c .

Note that the inequality (8) holds for all possible nonzero values of the coefficients t_{ij} 's. Hence, the convex program (7) provides a suboptimal solution for the non-convex problem (6) together with an upper bound on its optimal objective value. Roughly speaking, a suitable choice of \mathcal{G}^c makes the upper bound $n - \min_{\mathcal{G}^s \subseteq \mathcal{G}} \text{OS}(\mathcal{G}^s \cup \mathcal{G}^c)$ very small for a large class of sparse graphs \mathcal{G} . In particular, given an arbitrary tree decomposition of \mathcal{G} with width t , the

graph \mathcal{G}^c can be designed based on the tree in a way that

$$n - \min_{\mathcal{G}^s} \left\{ \text{OS}(\mathcal{G}^s \cup \mathcal{G}^c) \mid \mathcal{G}^s \subseteq \mathcal{G} \right\} \leq t + 1, \quad (9)$$

provided that all supernodes of the tree decomposition have the same size (see [27] for the general case where the supernodes have different sizes). Hence, the convex problem (7) is able to provide a suboptimal solution for problem (6) with the property that $\text{rank}\{W^{\text{opt}}\} \leq t + 1$. In particular, if an optimal tree decomposition is deployed for the construction of \mathcal{G}^c , then the relation $\text{rank}\{W^{\text{opt}}\} \leq \text{tw}(\mathcal{G}) + 1$ holds for all nonzero values of the coefficients t_{ij} 's, where $\text{tw}(\mathcal{G})$ denotes the treewidth of \mathcal{G} . Note that the existence of a solution for problem (6) of rank at most $\text{tw}(\mathcal{G}) + 1$ has already been proved in [28] for real-valued problems, but the technique stated above works for both real and complex problems. In addition, the above technique designs infinitely many optimization problems, each of which returns such a solution. The importance of this result will become clear later in this paper.

Assume that \mathcal{G} is a large-scale graph with no clear sparsity pattern. In this case, it could be difficult to find a good tree decomposition or directly design a subgraph \mathcal{G}^c minimizing the upper bound $n - \min_{\mathcal{G}^s \subseteq \mathcal{G}} \text{OS}(\mathcal{G}^s \cup \mathcal{G}^c)$. Under this circumstance, we use another convex relaxation for (6).

Convex Relaxation II: This problem is defined as

$$\min_{W \in \mathbb{H}^n} \sum_{(i,j) \in \mathcal{E} \cup \mathcal{E}^c} t_{ij} \text{Im}\{W_{ij}\} \quad (10a)$$

$$\text{s.t.} \quad \text{Re}\{W_{ij}\} = \widehat{W}_{ij}, \quad \forall (i,j) \in \mathcal{E} \quad (10b)$$

$$W_{kk} = \widehat{W}_{kk}, \quad \forall k \in \mathcal{V} \quad (10c)$$

$$W \succeq 0 \quad (10d)$$

with nonzero coefficients t_{ij} 's, where the variable of the optimization is the complex-valued matrix W .

Let W^{opt} denote an arbitrary solution of the above optimization problem. The matrix $\text{Re}\{W^{\text{opt}}\}$ turns out to be a suboptimal real-valued solution of the matrix completion problem (6) satisfying the inequality

$$\text{rank}\{\text{Re}\{W^{\text{opt}}\}\} \leq 2(n - \text{OS}(\mathcal{G} \cup \mathcal{G}^c)). \quad (11)$$

At the cost of adding the factor 2, the bound provided in (11) is simpler than the one given in (8)

due to obviating the need for taking the minimum of $\text{OS}(\cdot)$ over a set of subgraphs \mathcal{G}^s . The above bound is very useful since it is small for a large class of sparse graphs, even in the case where \mathcal{G}^c is considered as a trivial graph with no edges.

3.2 Sparse Quadratic Optimization

Consider the standard non-convex QCQP:

$$\min_{x \in \mathbb{R}^{n-1}} x^* A_0 x + 2b_0^* x + c_0 \quad (12a)$$

$$\text{s.t.} \quad x^* A_k x + 2b_k^* x + c_k \leq 0, \quad k = 1, \dots, m \quad (12b)$$

where $A_k \in \mathbb{S}^{n-1}$, $b_k \in \mathbb{R}^{n-1}$ and $c_k \in \mathbb{R}$, for $k = 0, \dots, m$. Define

$$M_k = \begin{bmatrix} c_k & b_k^* \\ b_k & A_k \end{bmatrix}. \quad (13)$$

The problem (12) can be reformulated as

$$\begin{aligned} \min_{W \in \mathbb{S}^n} \quad & \text{trace}\{M_0 W\} \\ \text{s.t.} \quad & \text{trace}\{M_k W\} \leq 0, \quad k = 1, \dots, m \\ & W_{11} = 1, \\ & W \succeq 0, \\ & \text{rank}\{W\} = 1, \end{aligned} \quad (14)$$

where W plays the role of

$$\begin{bmatrix} 1 \\ x \end{bmatrix} [1 \quad x^*]. \quad (15)$$

In the above reformulation of QCQP (12), the constraint $\text{rank}\{W\} = 1$ carries all the non-convexity. Neglecting this constraint yields an SDP relaxation. Let \widehat{W} denote an arbitrary solution of the SDP relaxation of the non-convex problem (12). There are cases where \widehat{W} has full rank and yet there exists a low-rank solution simultaneously. Indeed, the SDP relaxation could naturally have infinitely many solutions, and therefore a solution with the lowest rank should be sought.

Low-Rank Solution: In an effort to find a low-rank SDP solution, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with n vertices such that $(i, j) \in \mathcal{G}$ if the (i, j) entry of at least one of the matrices M_0, M_1, \dots, M_m is nonzero. The graph \mathcal{G} captures the sparsity of the optimization problem (12). Observe that those off-diagonal entries of \widehat{W} that correspond to non-existent edges

of \mathcal{G} play no direct role in the SDP relaxation. As a result, it can be inferred that every solution W^{opt} to the matrix completion problem (6) or its convex relaxations (7) and (10) is also a solution to the SDP relaxation of the QCQP problem (12). Depending on the choice of \mathcal{G}^c in (7) and (10), different low-rank solutions of the SDP relaxation can be generated for a sparse graph \mathcal{G} . In particular, there are infinitely many optimization problems with linear objectives such that each one generates a solution W^{opt} of the SDP relaxation with rank at most $\text{tw}\{\mathcal{G}\} + 1$, provided that the optimal tree decomposition of \mathcal{G} is known. Without taking advantage of a tree decomposition, we can generate a solution with rank at most $2(n - \text{OS}(\mathcal{G}))$ in polynomial time (note that this solution can be found efficiently, even though computing the theoretical upper bound on its rank would be an NP-hard problem). Note that it is shown in [11] that every polynomial optimization problem can be reformulated in a higher dimensional space (still polynomial size description) such that its corresponding treewidth becomes 1. For such a formulation, the SDP relaxation has rank at most 2. Although this reformulation improves the rank for approximation purposes, it offers a looser (rather than tightened) lower bound on the globally optimal objective value, and this introduces a trade-off.

Penalized SDP Relaxation: The strategy delineated above consists of two steps: (i) finding an arbitrary (potentially high-rank) solution \widehat{W} of the SDP relaxation for QCQP, and (ii) turning the solution into a lower rank solution W^{opt} by solving a second convex optimization based on the matrix completion approach. It is advantageous to integrate these two steps. This will be carried out in the sequel. Consider the convex optimization problem

$$\begin{aligned} \min_{W \in \mathbb{S}^n} \quad & \text{trace}\{M_0 W\} + \varepsilon_1 \text{trace}\{W\} + \varepsilon_2 \sum_{(i,j) \in \mathcal{E}^c} t_{ij} W_{ij} \\ \text{s.t.} \quad & \text{trace}\{M_k W\} \leq 0, \quad k = 1, \dots, m \\ & W_{11} = 1, \\ & W \succeq 0 \end{aligned} \tag{16}$$

for a given graph \mathcal{G}^c , a scalar ε_1 , and nonzero numbers ε_2 and t_{ij} 's. Notice that the objective of this optimization has two penalty terms: (i) a trace term motivated by the nuclear norm technique for

rank compensation, and (ii) a weighted sum of some off-diagonal entries of W motivated by the matrix completion approach described earlier. As before and under some technical assumptions, every solution W^{opt} of the above penalized SDP problem satisfies the inequality

$$\text{rank}\{W^{\text{opt}}\} \leq n - \min_{\mathcal{G}^s \subseteq \mathcal{G}} \text{OS}(\mathcal{G}^s \cup \mathcal{G}^c) \tag{17}$$

where the right side of the inequality can be replaced by $t + 1$ if \mathcal{G}^c is constructed from a tree decomposition of \mathcal{G} with width t such that its supernodes are of identical size. Note that the penalized SDP may become arbitrarily close to the SDP problem by making ε_1 sufficiently small or equal to zero. This means that an ε -approximation of a low-rank solution of the SDP relaxation of QCQP can be obtained through the penalized SDP problem. In other words, the proposed penalization eliminates high-rank solutions of the SDP relaxation. A similar penalization technique can be derived based on problem (10), leading to the upper bound $2(n - \text{OS}(\mathcal{G} \cup \mathcal{G}^c))$ on the rank of all solutions of the corresponding penalized SDP.

Consider a QCQP problem whose underlying sparsity graph \mathcal{G} has a relatively small treewidth. The above penalized convex relaxation generates only low-rank solutions for an infinite choice of the coefficients ε_1 , ε_2 and t_{ij} 's. Our simulations on thousands of power optimization and optimal distributed control problems suggest that it is possible to generate a near-global rank-1 solution by meticulously devising t_{ij} 's and tuning the regularization parameters ε_1 and ε_2 [10, 29, 22, 30, 31].

Acknowledgments

This work was supported by DARPA YFA, ONR YIP Award, NSF CAREER Award 1351279 and NSF EECS Award 1406865.

REFERENCES

- [1] A. Ben-Tal and A. S. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.
- [2] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, "Linear matrix inequalities in system and control theory," *Studies in Applied Mathematics, SIAM*, 1994.

- [3] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, 2004.
- [4] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *Journal of the ACM*, vol. 42, pp. 1115–1145, 1995.
- [5] Y. Nesterov, "Semidefinite relaxation and nonconvex quadratic optimization," *Optimization Methods and Software*, vol. 9, pp. 141–160, 1998.
- [6] Y. Ye, "Approximating quadratic programming with bound and quadratic constraints," *Mathematical Programming*, vol. 84, pp. 219–226, 1999.
- [7] Z. Luo, N. Sidiropoulos, P. Tseng, and S. Zhang, "Approximation bounds for quadratic optimization with homogeneous quadratic constraints," *SIAM Journal on Optimization*, vol. 18, pp. 1–28, 2007.
- [8] S. He, Z. Luo, J. Nie, and S. Zhang, "Semidefinite relaxation bounds for indefinite homogeneous quadratic optimization," *SIAM Journal on Optimization*, vol. 19, pp. 503–523, 2008.
- [9] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 92–107, 2012.
- [10] S. Sojoudi, R. Madani, and J. Lavaei, "Low-rank solution of convex relaxation for optimal power flow problem," *IEEE SmartGridComm*, 2013.
- [11] S. Sojoudi, R. Madani, G. Fazelnia, and J. Lavaei, "Graph-theoretic algorithms for polynomial optimization problems," *53rd IEEE Conference on Decision and Control*, 2014.
- [12] D. Bienstock and A. Verma, "Strong NP-hardness of AC power flows feasibility," Dec. 2015, [Online]. Available: <http://arxiv.org/pdf/1512.07315v1.pdf>.
- [13] "Report by Federal Energy Regulatory Commission." [Online]. Available: <http://www.ferc.gov/industries/electric/indus-act/market-planning/opf-papers.asp>
- [14] J. Lavaei and S. H. Low, "Convexification of optimal power flow problem," *48th Annual Allerton Conference*, 2010.
- [15] J. Lavaei, "Zero duality gap for classical OPF problem convexifies fundamental nonlinear power problems," *American Control Conference*, 2011.
- [16] J. Lavaei, A. Rantzer, and S. H. Low, "Power flow optimization using positive programming," *18th IFAC World Congress*, 2011.
- [17] S. Sojoudi and J. Lavaei, "Physics of power networks makes hard optimization problems easy to solve," *IEEE Power & Energy Society General Meeting*, 2012.
- [18] J. Lavaei, B. Zhang, and D. Tse, "Geometry of power flows and optimization in distribution networks," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 572–583, 2014.
- [19] R. Madani, S. Sojoudi, and J. Lavaei, "Convex relaxation for optimal power flow problem: Mesh networks," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 199–211, 2015.
- [20] S. Sojoudi and J. Lavaei, "Convexification of optimal power flow problem by means of phase shifters," *IEEE SmartGridComm*, 2013.
- [21] S. Sojoudi and S. H. Low, "Optimal charging of plug-in hybrid electric vehicles in smart grids," in *IEEE Power and Energy Society General Meeting*, 2011.
- [22] R. Madani, M. Ashraphijuo, and J. Lavaei, "Promises of conic relaxation for contingency-constrained optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1297–1307, 2016.
- [23] R. Madani, J. Lavaei, and R. Baldick, "Convexification of power flow problem over arbitrary networks," *54th IEEE Conference on Decision and Control*, 2015.
- [24] S. Sojoudi and J. Lavaei, "Exactness of semidefinite relaxations for nonlinear optimization problems with underlying graph structure," *SIAM Journal on Optimization*, vol. 4, pp. 1746 – 1778, 2014.
- [25] R. Madani, M. Ashraphijuo, and J. Lavaei, "OPF Solver," <http://www.ieor.berkeley.edu/~lavaei/Software.html>, 2014.
- [26] P. Hackney, B. Harris, M. Lay, L. H. Mitchell, S. K. Narayan, and A. Pascoe, "Linearly independent vertices and minimum semidefinite rank," *Linear Algebra and its Applications*, vol. 431, no. 8, pp. 1105–1115, 2009.
- [27] R. Madani, G. Fazelnia, S. Sojoudi, and J. Lavaei, "Low-rank solutions of matrix inequalities with applications to polynomial optimization and matrix completion problems," *53rd IEEE Conference on Decision and Control*, 2014, extended version available at http://ieor.berkeley.edu/~lavaei/LMI_Low_Rank.pdf.
- [28] M. Laurent and A. Varvitsiotis, "A new graph parameter related to bounded rank positive semidefinite matrix completions," *Mathematical Programming*, pp. 1–35, 2012.
- [29] J. Lavaei, "Optimal decentralized control problem as a rank-constrained optimization," *51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2013.

- [30] G. Fazelnia, R. Madani, A. Kalbat, and J. Lavaei, “Convex relaxation for optimal distributed control problems,” *to appear in IEEE Transactions on Automatic Control*, 2016.
- [31] A. Kalbat, R. Madani, G. Fazelnia, and J. Lavaei, “Efficient convex relaxation for stochastic optimal distributed control problem,” *52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2014.

A New Perspective on Boosting in Linear Regression via Subgradient Optimization and Relatives

Paul Grigas

Operations Research Center
Massachusetts Institute of Technology
USA
pgrigas@mit.edu

1. Boosting Algorithms in Linear Regression

Boosting [15, 6, 9, 16, 12] is an extremely successful and popular supervised learning technique that combines multiple “weak” learners into a more powerful “committee.” AdaBoost [7, 16, 12], developed in the context of classification, is one of the earliest and most influential boosting algorithms. In our paper [5], we analyze boosting algorithms in linear regression [8, 9, 3] from the perspective of modern first-order methods in convex optimization. This perspective has two primary upshots: *(i)* it leads to first-ever computational guarantees for existing boosting algorithms, and *(ii)* it leads to new boosting algorithms with novel connections to the LASSO [18].



Paul Grigas and Mohit Tawarmalani

Notation We use the usual linear regression notation with model matrix $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_p] \in \mathbb{R}^{n \times p}$, response vector $\mathbf{y} \in \mathbb{R}^{n \times 1}$, and regression coefficients $\beta \in \mathbb{R}^p$. Each column of \mathbf{X} corresponds to a particular feature or predictor variable, and each row corresponds to a particular observed sample. We assume herein that the features \mathbf{X}_i have been centered to have zero mean and unit ℓ_2 norm, i.e., $\|\mathbf{X}_i\|_2 = 1$ for $i = 1, \dots, p$, and \mathbf{y} is also centered to have zero mean. For a regression coefficient vector β , the predicted value of the response is given by $\mathbf{X}\beta$ and $r = \mathbf{y} - \mathbf{X}\beta$ denotes the residuals. Let e_j denote the j^{th} unit vector in \mathbb{R}^p and let $\|v\|_0$ denote the number of nonzero coefficients in the vector v . Denote the empirical least squares loss function by $L_n(\beta) := \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2$, let $L_n^* := \min_{\beta \in \mathbb{R}^p} L_n(\beta)$, and let $\hat{\beta}_{\text{LS}}$ denote an arbitrary minimizer of $L_n(\beta)$, i.e., $\hat{\beta}_{\text{LS}} \in \arg \min_{\beta \in \mathbb{R}^p} L_n(\beta)$. Finally, let $\lambda_{\text{pmin}}(\mathbf{X}^T \mathbf{X})$ denote the smallest nonzero (and hence positive) eigenvalue of $\mathbf{X}^T \mathbf{X}$.

Boosting and Implicit Regularization The first boosting algorithm we consider is the Incremental Forward Stagewise algorithm [12, 3] presented below, which we refer to as FS_ε .

Algorithm: Incremental Forward Stagewise Regression – FS_ε

- Fix the learning rate $\varepsilon > 0$ and number of iterations M .
- Initialize at $\hat{r}^0 = \mathbf{y}$, $\hat{\beta}^0 = 0$, $k = 0$.
- For $0 \leq k \leq M$ do the following:
 1. Compute: $j_k \in \arg \max_{j \in \{1, \dots, p\}} |(\hat{r}^k)^T \mathbf{X}_j|$
 2. $\hat{\beta}_{j_k}^{k+1} \leftarrow \hat{\beta}_{j_k}^k + \varepsilon \operatorname{sgn}((\hat{r}^k)^T \mathbf{X}_{j_k})$ and $\hat{\beta}_j^{k+1} \leftarrow \hat{\beta}_j^k, j \neq j_k$
 - $\hat{r}^{k+1} \leftarrow \hat{r}^k - \varepsilon \operatorname{sgn}((\hat{r}^k)^T \mathbf{X}_{j_k}) \mathbf{X}_{j_k}$.

At the k^{th} iteration, FS_ε chooses a column \mathbf{X}_{j_k} , corresponding to a particular feature that is the most correlated (in absolute value) with the current residuals and then updates the corresponding regression coefficient by an amount $\varepsilon > 0$, called the learning rate (or shrinkage factor).

A close cousin of FS_ε is the least squares boosting algorithm, or $\text{LS-BOOST}(\varepsilon)$, proposed in [8]. The $\text{LS-BOOST}(\varepsilon)$ algorithm is identical to FS_ε except that $\text{LS-BOOST}(\varepsilon)$ changes the amount by which the selected coefficient is updated at each iteration – at the k^{th} iteration, $\text{LS-BOOST}(\varepsilon)$ updates:

$$\begin{aligned} \hat{\beta}_{j_k}^{k+1} &\leftarrow \hat{\beta}_{j_k}^k + \varepsilon ((\hat{r}^k)^T \mathbf{X}_{j_k}) \text{ and } \hat{\beta}_j^{k+1} \leftarrow \hat{\beta}_j^k, j \neq j_k \\ \hat{r}^{k+1} &\leftarrow \hat{r}^k - \varepsilon ((\hat{r}^k)^T \mathbf{X}_{j_k}) \mathbf{X}_{j_k}, \end{aligned}$$

where now $\varepsilon \in (0, 1]$.

Note that both algorithms often lead to models with attractive statistical properties [8, 12, 1, 2]. In this linear regression setting, while there may be several important concerns, it is often of primary importance to produce a parsimonious model with good out of sample predictive performance. When p is small relative to n , minimizing the empirical least squares loss function $L_n(\beta)$ usually achieves this goal. On the other hand, when $n, p \gg 0$ (and particularly when $p > n$), $\hat{\beta}_{\text{LS}}$ often has poor predictive performance; in other words, $\hat{\beta}_{\text{LS}}$ *overfits* the training data. Additionally $\hat{\beta}_{\text{LS}}$ is almost always fully dense. Regularization techniques enable one to find a model with better predictive performance by balancing two competing objectives: (i) data fidelity, or how well the model fits the training data, and (ii) “shrinkage,” or a measure of model simplicity. Shrinkage is often measured using $\|\beta\|$ for some appropriate norm $\|\cdot\|$, whereby a coefficient vector with a relatively small value of $\|\beta\|$ exhibits more shrinkage. The FS_ε and $\text{LS-BOOST}(\varepsilon)$ algorithms are effective, even in settings where $n, p \gg 0$ and/or $p > n$, because they each impart a type of *implicit regularization* by tracing out a path of models with varying levels of data fidelity and shrinkage.

For both FS_ε and $\text{LS-BOOST}(\varepsilon)$, the choices of ε and M play crucial roles in the statistical behavior of the algorithm. Let us consider $\text{LS-BOOST}(\varepsilon)$ alone for now. Setting $\varepsilon = 1$ corresponds to minimizing the empirical least squares loss function $L_n(\beta)$ along the direction of the selected feature, i.e., it holds that $(\hat{r}^k)^T \mathbf{X}_{j_k} = \arg \min_{u \in \mathbb{R}} L_n(\beta^k + u e_{j_k})$. Qualitatively speaking, $\text{LS-BOOST}(\varepsilon)$ does eventually minimize the empirical least squares loss function as long as $\varepsilon > 0$, but a small value of ε (for example, $\varepsilon = 0.001$) slows down the rate of convergence as compared to the choice $\varepsilon = 1$. Thus it may seem counterintuitive to set $\varepsilon < 1$; however with a small value of ε it

is possible to explore a larger class of models, with varying degrees of shrinkage. It has been observed empirically that small values of ε often lead to models with better predictive power [8]. In practice, one might set ε relatively small and use a holdout dataset to select the best performing model found throughout the course of the algorithm; in many instances the selected model is found long before convergence to the empirical least squares solution. The role of M and ε in FS_ε is very similar. In short, both M and ε together control the training error (data fidelity) and the amount of shrinkage (regularization) for both $\text{LS-BOOST}(\varepsilon)$ and FS_ε . We refer the reader to Figure 1, depicting the evolution of the algorithmic properties of FS_ε and $\text{LS-BOOST}(\varepsilon)$ as a function of M and ε .

2. Computational Guarantees for FS_ε and $\text{LS-Boost}(\varepsilon)$ Through the Lens of Subgradient Descent

Up until the present work, and as pointed out by [12], the understanding of how the algorithmic parameters ε and M control the tradeoffs between data fidelity and shrinkage in FS_ε and $\text{LS-BOOST}(\varepsilon)$ has been rather qualitative. One of the contributions of the full paper is a precise quantification of this tradeoff, for both FS_ε and $\text{LS-BOOST}(\varepsilon)$. Indeed, the paper presents, for the first time, precise descriptions of how the quantities ε and M control the amount of training error and regularization in FS_ε and $\text{LS-BOOST}(\varepsilon)$. These precise computational guarantees are enabled by new connections to first-order methods in convex optimization. In particular, the paper presents a new unifying framework for interpreting FS_ε and $\text{LS-BOOST}(\varepsilon)$ as instances of the subgradient descent method of convex optimization, applied to the problem of minimizing the largest correlation between residuals and predictors.

Boosting as Subgradient Descent Let $P_{\text{res}} := \{r \in \mathbb{R}^n : r = \mathbf{y} - \mathbf{X}\beta \text{ for some } \beta \in \mathbb{R}^p\}$ denote the affine space of residuals and consider the following

convex optimization problem:

$$\begin{aligned} \text{CM:} \quad & \min_r f(r) := \|\mathbf{X}^T r\|_\infty \\ & \text{s.t. } r \in P_{\text{res}}, \end{aligned} \quad (1)$$

which we dub the “Correlation Minimization” problem, or CM for short, since $f(r)$ is the largest absolute correlation between the residual vector r and the predictors. Note an important subtlety in the CM problem, namely that the optimization variable in CM is the *residual* r and *not* the regression coefficient vector β .

The subgradient descent method (see [17], for example) is a simple generalization of the method of gradient descent to the case when $f(\cdot)$ is not differentiable. As applied to the CM problem (1), the subgradient descent method has the following update scheme:

Compute a subgradient of $f(\cdot)$ at r^k :

$$g^k \in \partial f(r^k) \quad (2)$$

Perform update at r^k :

$$r^{k+1} \leftarrow \Pi_{P_{\text{res}}}(r^k - \alpha_k g^k), \quad (3)$$

where $\partial f(r)$ denotes the set of subgradients of $f(\cdot)$ at r and $\Pi_{P_{\text{res}}}$ denotes the (Euclidean) projection operator onto P_{res} , namely $\Pi_{P_{\text{res}}}(\bar{r}) := \arg \min_{r \in P_{\text{res}}} \|r - \bar{r}\|_2$.

The following proposition states that the boosting algorithms FS_ε and $\text{LS-BOOST}(\varepsilon)$ can be viewed as instantiations of the subgradient descent method to solve the CM problem (1).

Proposition 1. *Consider the subgradient descent method (2)–(3) with step-size sequence $\{\alpha_k\}$ to solve the correlation minimization (CM) problem (1), initialized at $\hat{r}^0 = \mathbf{y}$. Then:*

- (i) *the FS_ε algorithm is an instance of subgradient descent, with a constant step-size $\alpha_k := \varepsilon$ at each iteration,*
- (ii) *the $\text{LS-BOOST}(\varepsilon)$ algorithm is an instance of subgradient descent, with non-uniform step-sizes $\alpha_k := \varepsilon |\tilde{u}_{j_k}|$ at iteration k , where $\tilde{u}_{j_k} := (\hat{r}^k)^T \mathbf{X}_{j_k} = \arg \min_u \|\hat{r}^k - \mathbf{X}_{j_k} u\|_2^2$.*

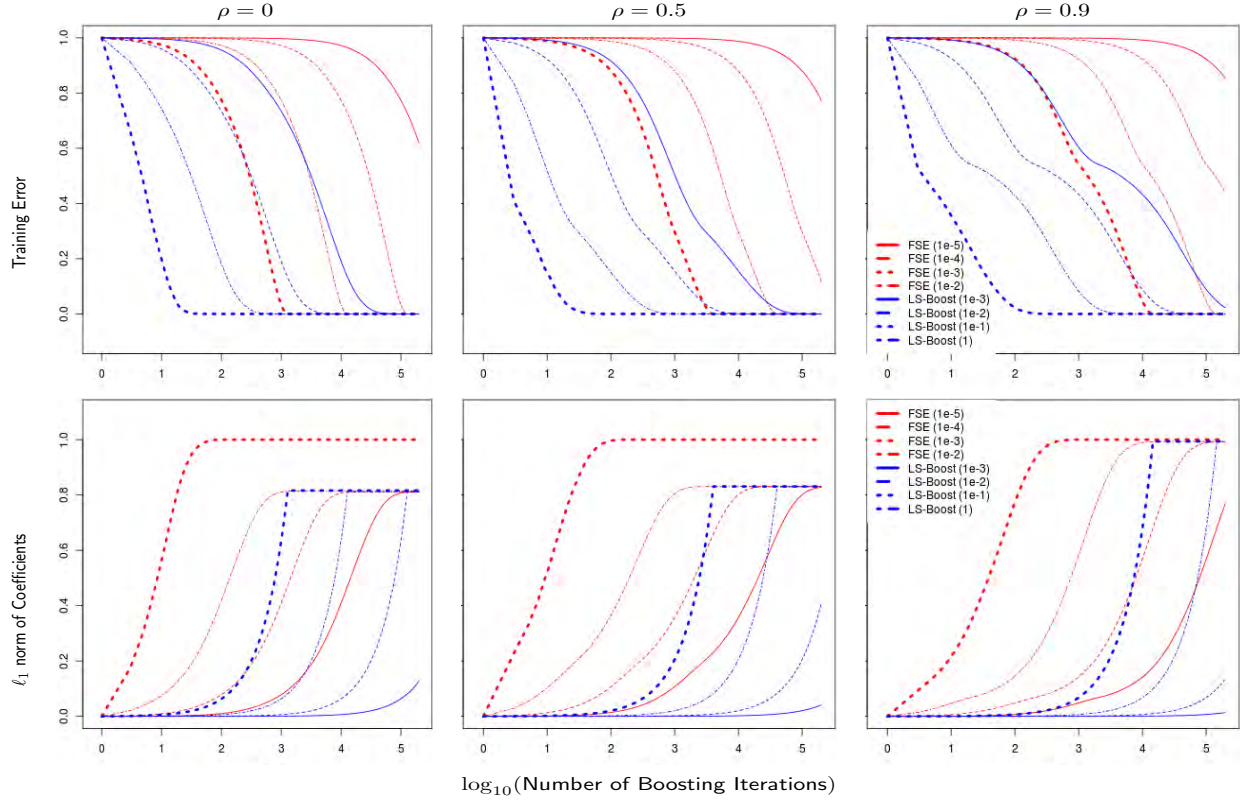


Figure 1: Evolution of LS-BOOST(ε) and FS_ε versus iterations (in the log-scale), run on a synthetic dataset with $n = 50$, $p = 500$; the features are drawn from a Gaussian distribution with pairwise correlations ρ . The true β has ten nonzeros with $\beta_i = 1, i \leq 10$ and SNR = 1. Three different values of ρ have been considered ($\rho = 0, 0.5$ and 0.9) and ε varies from $\varepsilon = 10^{-5}$ to $\varepsilon = 1$. The top row shows the training errors for different learning rates, and the bottom row shows the ℓ_1 norm of the coefficients produced by the different algorithms for different learning rates. (Here the values have all been re-scaled so that the y-axis lies in $[0, 1]$).

Some Computational Guarantees for FS_ε

Proposition 1 is interesting especially since FS_ε and LS-BOOST(ε) have been traditionally interpreted as greedy coordinate descent or steepest descent type procedures [12, 10]. Furthermore the following theorem presents relevant convergence properties of FS_ε , some of which are direct consequences of Proposition 1 based on well-known computational guarantees associated with the subgradient descent method [14, 13].

Theorem 2. (Some Convergence Properties of FS_ε) Consider the FS_ε algorithm with learning rate ε . Let $M \geq 0$ be the total number of iterations. Then there exists an index $i \in \{0, \dots, M\}$ for which the following bounds hold:

(i) (training error):

$$L_n(\hat{\beta}^i) - L_n^* \leq \frac{p}{2n\lambda_{\min}(\mathbf{X}^T\mathbf{X})} \left[\frac{\|\mathbf{X}\hat{\beta}_{LS}\|_2^2}{\varepsilon(M+1)} + \varepsilon \right]^2$$

(ii) (ℓ_1 -shrinkage of coefficients): $\|\hat{\beta}^i\|_1 \leq M\varepsilon$

(iii) (sparsity of coefficients): $\|\hat{\beta}^i\|_0 \leq M$.

Theorem 2 gives a flavor of some of the computational guarantees included in the full paper; the paper includes additional results regarding convergence of regression coefficients, prediction distances, and correlation values. Furthermore, the paper also includes an analogous theorem for LS-BOOST(ε), which highlights the differences in convergence patterns between the two algorithms. Theorem 2 (and related results included in the paper) provides, for

the first time, a precise theoretical description of the amount of data fidelity and shrinkage/regularization imparted by running FS_ε for a fixed but arbitrary number of iterations, for *any* dataset. Moreover, this result sheds light on the data fidelity *vis-à-vis* shrinkage characteristics of FS_ε . In particular, Theorem 2 demonstrates explicitly how (bounds on) the training error and ℓ_1 -shrinkage depend on the algorithmic parameters ε and M , which implies an explicit tradeoff between data fidelity and shrinkage that is controlled by these parameters. Indeed, let TBOUND and SBOUND denote the training error bound and shrinkage bound in parts (i) and (ii) of Theorem 2, respectively. Then simple manipulation of the arithmetic in these two bounds yields the following tradeoff equation:

$$\text{TBOUND} = \frac{p}{2n\lambda_{\min}(\mathbf{X}^T\mathbf{X})} \left[\frac{\|\mathbf{X}\hat{\beta}_{\text{LS}}\|_2^2}{\text{SBOUND} + \varepsilon} + \varepsilon \right]^2.$$

In the full paper, we extensively discuss the consequences of Theorem 2 and related results in terms of improved understanding of the behavior of FS_ε and $\text{LS-BOOST}(\varepsilon)$.

3. Boosting and Lasso

As mentioned previously, FS_ε and $\text{LS-BOOST}(\varepsilon)$ are effective even in high-dimensional settings where $p > n$ since they implicitly deliver regularized models. An alternative and very popular approach in such settings is based on an explicit regularization scheme, namely ℓ_1 -regularized regression, i.e., LASSO [18]. The constraint version of LASSO with regularization parameter $\delta \geq 0$ is given by the following convex quadratic optimization problem:

$$\text{LASSO} : L_{n,\delta}^* := \min_{\beta} \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 \quad \text{s.t.} \quad \|\beta\|_1 \leq \delta. \quad (4)$$

Although LASSO and the previously discussed boosting methods originate from different perspectives, there are interesting similarities between the two, as is nicely explored in [12, 3, 11]. Figure 2 (top panel) shows an example where the LASSO profile/path (the set of solutions of (4) as δ varies) is similar to the trajectories of FS_ε and $\text{LS-BOOST}(\varepsilon)$ (for small values

of ε). Although they are different in general (Figure 2, bottom panel), [3, 11] explores the connection more deeply.

One of the aims of our work is to contribute some substantial further understanding of the relationship between LASSO, FS_ε , and $\text{LS-BOOST}(\varepsilon)$, particularly for *arbitrary* datasets where such understanding is still fairly limited. Motivated thusly, we introduce a new boosting algorithm, called $\text{R-FS}_{\varepsilon,\delta}$ (regularized FS_ε), that includes an additional shrinkage step as compared to FS_ε . That is $\text{R-FS}_{\varepsilon,\delta}$ first shrinks all of the coefficients, then adds ε to the selected coefficient; $\text{R-FS}_{\varepsilon,\delta}$ replaces Step 2 of FS_ε by:

$$\begin{aligned} \hat{\beta}_{j_k}^{k+1} &\leftarrow (1 - \frac{\varepsilon}{\delta}) \hat{\beta}_{j_k}^k + \varepsilon \text{sgn}((\hat{r}^k)^T \mathbf{X}_{j_k}) \text{ and } \hat{\beta}_j^{k+1} \leftarrow \\ &\quad (1 - \frac{\varepsilon}{\delta}) \hat{\beta}_j^k, j \neq j_k \\ \hat{r}^{k+1} &\leftarrow \hat{r}^k - \varepsilon [\text{sgn}((\hat{r}^k)^T \mathbf{X}_{j_k}) \mathbf{X}_{j_k} + \frac{1}{\delta}(\hat{r}^k - \mathbf{y})] \end{aligned}$$

where $\delta > 0$ is an additional algorithmic parameter. Note that one can easily verify the formula for updating the residuals based on the coefficient update. Furthermore, $\text{R-FS}_{\varepsilon,\delta}$ with $\delta = +\infty$ is exactly FS_ε .

It turns out that $\text{R-FS}_{\varepsilon,\delta}$ is precisely related to the LASSO problem through duality. Consider the following parametric family of optimization problems indexed by $\delta \in (0, \infty]$:

$$\begin{aligned} \text{RCM}_\delta : \quad \min_r \quad f_\delta(r) &:= \|\mathbf{X}^T r\|_\infty + \frac{1}{2\delta} \|r - \mathbf{y}\|_2^2 \\ \text{s.t.} \quad r &\in P_{\text{res}}, \end{aligned} \quad (5)$$

where $P_{\text{res}} = \{r \in \mathbb{R}^n : r = \mathbf{y} - \mathbf{X}\beta \text{ for some } \beta \in \mathbb{R}^p\}$ and “RCM” connotes Regularized Correlation Minimization.

In the full paper, we establish the following connections between $\text{R-FS}_{\varepsilon,\delta}$, the RCM problem, and the LASSO problem:

1. The RCM problem (5) is equivalent to the dual problem of the LASSO (4).
2. $\text{R-FS}_{\varepsilon,\delta}$ is an instance of subgradient descent applied to the RCM problem (5).

The $\text{R-FS}_{\varepsilon,\delta}$ algorithm is also related to a variant of the Frank-Wolfe method in convex optimization [4], applied directly to LASSO.

Furthermore, we show the following properties of the new algorithm $\text{R-FS}_{\varepsilon,\delta}$:

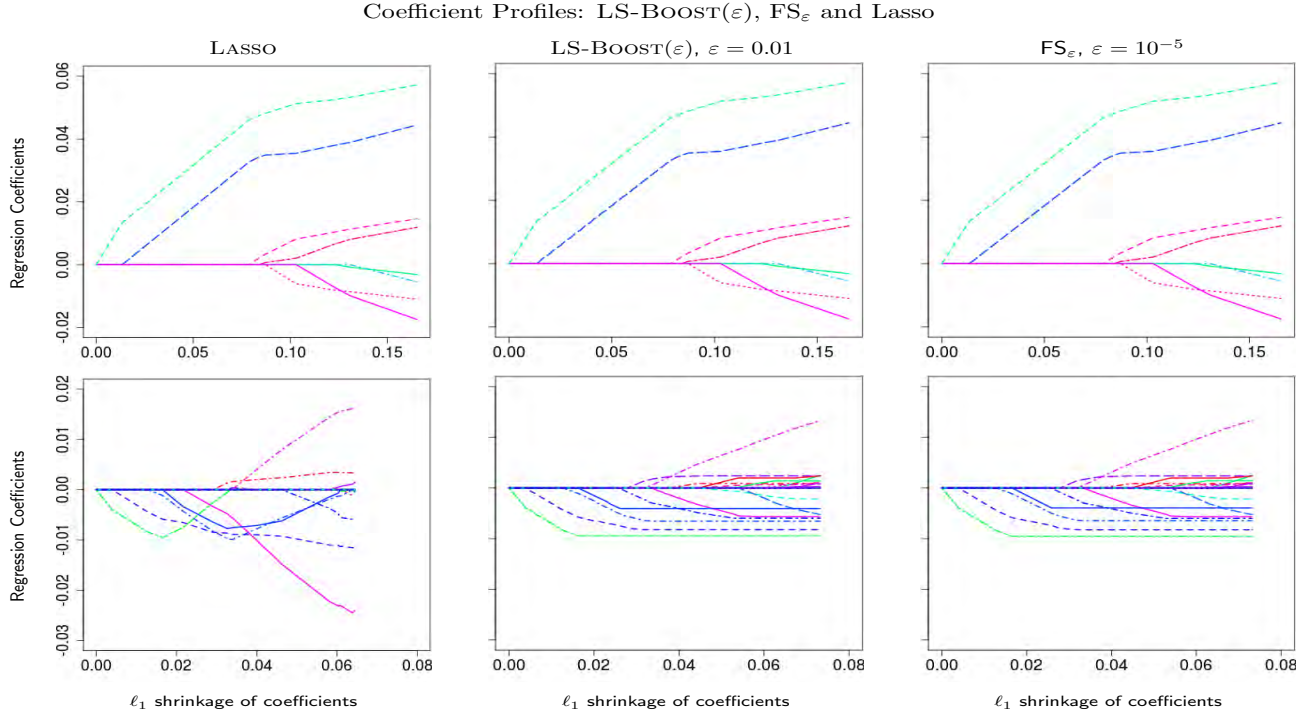


Figure 2: Coefficient Profiles for different algorithms as a function of the ℓ_1 norm of the regression coefficients on two different datasets. The top row corresponds to a dataset where the coefficient profiles look very similar, and the bottom row corresponds to a dataset where the coefficient profile of LASSO is seen to be different from FS $_{\varepsilon}$ and LS-BOOST(ε).

- As the number of iterations become large, “grand” algorithm: the subgradient descent method applied to the RCM problem (5). R-FS $_{\varepsilon,\delta}$ delivers an approximate LASSO solution.
- R-FS $_{\varepsilon,\delta}$ has computational guarantees analogous to Theorem 2 that provide a precise description of data-fidelity *vis-à-vis* ℓ_1 shrinkage.
- R-FS $_{\varepsilon,\delta}$ specializes to FS $_{\varepsilon}$, LS-BOOST(ε) and the LASSO depending on the parameter value δ and the learning rates (step-sizes) used therein.
- An adaptive version of R-FS $_{\varepsilon,\delta}$, which we call PATH-R-FS $_{\varepsilon}$, is shown to approximate the path of LASSO solutions with precise bounds that quantify the approximation error over the path.
- In our computational experiments, we observe that R-FS $_{\varepsilon,\delta}$ leads to models with statistical properties that compare favorably with the LASSO and FS $_{\varepsilon}$. R-FS $_{\varepsilon,\delta}$ also leads to models that are sparser than FS $_{\varepsilon}$.

In total, we establish that FS $_{\varepsilon}$, LS-BOOST(ε) and LASSO can be viewed as special instances of one

4. Summary

We analyze boosting algorithms in linear regression from the perspective modern first-order methods in convex optimization. We show that classic boosting algorithms in linear regression, FS $_{\varepsilon}$ and LS-BOOST(ε), can be viewed as subgradient descent to minimize the maximum absolute correlation between features and residuals. We also propose a modification of FS $_{\varepsilon}$ that yields an algorithm for the LASSO, and that computes the LASSO path. Our perspective leads to first-ever comprehensive computational guarantees for all of these boosting algorithms, which provide a precise theoretical description of the amount of data-fidelity and regularization imparted by running a boosting algorithm with a pre-specified learning rate for a fixed but arbitrary number of iterations, for any dataset.

5. Acknowledgments

The results described here and in the full paper is joint work with Robert Freund and Rahul Mazumder, and I am grateful for their collaboration and their general support. I would also like to thank the prize committee — Mohit Tawarmalani, Fatma Kılınç-Karzan, Warren Powell, and Uday Shanbhag — for choosing to honor this work, and more importantly for their service to the optimization community in general.

REFERENCES

- [1] P. Bühlmann. Boosting for high-dimensional linear models. *Annals of Statistics*, pages 559–583, 2006.
- [2] P. Bühlmann and T. Hothorn. Boosting algorithms: regularization, prediction and model fitting (with discussion). *Statistical Science*, 22(4):477–505, 2008.
- [3] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression (with discussion). *Annals of Statistics*, 32(2):407–499, 2004.
- [4] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3:95–110, 1956.
- [5] R. M. Freund, P. Grigas, and R. Mazumder. A new perspective on boosting in linear regression via sub-gradient optimization and relatives. *arXiv preprint arXiv:1505.04243*, 2015.
- [6] Y. Freund. Boosting a weak learning algorithm by majority. *Information and computation*, 121(2):256–285, 1995.
- [7] Y. Freund and R. Schapire. Experiments with a new boosting algorithm. In *Machine Learning: Proceedings of the Thirteenth International Conference*, pages 148–156. Morgan Kaufman, San Francisco, 1996.
- [8] J. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232, 2001.
- [9] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting (with discussion). *Annals of Statistics*, 28:337–307, 2000.
- [10] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2000.
- [11] T. Hastie, J. Taylor, R. Tibshirani, and G. Walther. Forward stagewise regression and the monotone lasso. *Electronic Journal of Statistics*, 1:1–29, 2007.
- [12] T. Hastie, R. Tibshirani, and J. Friedman. *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Verlag, New York, 2009.
- [13] Y. E. Nesterov. *Introductory lectures on convex optimization: a basic course*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, 2003.
- [14] B. Polyak. *Introduction to Optimization*. Optimization Software, Inc., New York, 1987.
- [15] R. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):197–227, 1990.
- [16] R. Schapire and Y. Freund. *Boosting: Foundations and Algorithms*. Adaptive computation and machine learning. Mit Press, 2012.
- [17] N. Z. Shor. *Minimization Methods for Non-Differentiable Functions*, volume 3 of *Springer Series in Computational Mathematics*. Springer, Berlin, 1985.
- [18] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288, 1996.

The IOS 2016 Optimization Conference: A Whole Lot of Learning Going on!

Warren B. Powell (conference general chair)
 Princeton University
 powell@princeton.edu

The 2016 edition of the INFORMS Optimization Society conference held March 17–19, 2016 at Princeton University was a tremendous success, attracting over 200 participants attending 52 sessions (which included 8 tutorial sessions) along with four outstanding plenaries.

The theme of the conference was “Optimization and Learning,” and while this attracted considerable interest from the interface of optimization and machine learning, the real goal was to focus on learning new applications and new research challenges, regardless of the problem domain or research focus.

Arguably the highlight of the conference were the four outstanding plenary speakers. Richard O’Neill from the Federal Energy Regulatory Commission (FERC), with his extensive experience investigating the interface of optimization and power systems, gave a lecture that highlighted the many optimization challenges that remain in getting our arms around a robust grid. These span integer programming problems for deterministic unit commitment that continue to challenge the limits of the most powerful commercial packages, stochastic optimization problems to handle the growing presence of renewables, and complex nonlinear models of optimal power flow problems at increasingly fine-grained time scales.

Professor Han Liu from Princeton University then gave an exceptionally clear overview of optimization challenges in machine learning, listing four major areas: matrix optimization, distributed optimization, nonconvex optimization, and big data optimization. These were elegantly illustrated in the context of non-paranormal and exponential family graphical models, motivated by applications in genomics, neuroscience, climate science, and finance. The growing interest in this area was reflected by 11 separate sessions on machine learning-related topics.

Russ Tedrake from MIT then followed with an entertaining and informative summary of the optimal control problems that arise in robotics. He introduced the audience to some cutting-edge research in robotics, showcasing a trend towards optimization-based algorithms for some fundamental problems of the field. These include motion planning, collision avoidance, sensing, safety verification, walking over rough terrains, and autonomous flying, just to name a few. The problems introduce challenges in linear complementarity problems, automatic construction of Lyapunov functions, and sum of squares and semidefinite optimization. The growing presence of robotics is sure to create a source of exciting problems for optimization.

Peter Frazier from Cornell University closed the conference with an entertaining lecture on stochastic optimization problems (in particular, learning problems) that he has encountered working with Uber (where he is spending a sabbatical) and Yelp. The companies offer a range of stochastic optimization spanning learning how a market will respond to new websites, to the combined response of drivers and passengers to surge pricing. Stochastic optimization introduces issues such as modeling the different types of uncertainty, in addition to finding effective policies that reflect the different types of metrics that arise. 15 separate sessions addressed some form of optimization under uncertainty.

These talks alone contained enough research topics to keep an entire team of researchers busy for a lifetime! But these were just the highlights. The conference featured a series of nine well-attended tutorials:

- Machine Learning in Policy Search, by Haitham Bou Ammar and Jose Marcio Luna
- Optimization over Nonnegative Polynomials: Algorithms and Applications, by Amir Ali Ahmadi
- The Lasserre hierarchy for polynomial optimization, by Etienne de Klerk
- Applications of machine learning and optimization in online revenue management, by David Simchi-Levi
- Multilevel/multistage discrete optimization problems, by Ted Ralphs

- Nonlinear programming challenges in the optimal power flow problem, by Javad Lavaei
- Robust Optimization methods in Stochastic Analysis, by Chaitanya Bandi
- Online Convex Optimization, by Elad Hazan
- Optimization with equilibrium constraints, by Uday Shanbhag

Each tutorial was given a double time-slot, and then accompanied with closely related talks to build on the area.

While the plenaries and tutorials helped to highlight exciting research challenges and problem domains, the real strength of the conference came from the consistently high quality of talks spread over the 44 remaining sessions. There were complete tracks on optimization problems in machine learning, and stochastic and robust optimization (including a session on stochastic optimization in learning), with two additional sessions in parallel tracks on optimal learning and low-rank approximations in approximate dynamic programming. Sessions dedicated to both convex and nonconvex optimization (including convex relaxations of nonconvex problems) were evidence of the continued interest in these areas. Integer programming and multilevel optimization were also represented, along with application sessions in logistics, finance, and energy.

If there was a consistent theme throughout the talks, it was the interdisciplinary nature of the work: optimization in machine learning, machine learning in stochastic optimization, convex optimization for nonconvex problems, integer programming for stochastic optimization, stochastic programming in statistics (I am sure I have missed a few). Further, while most of the work in “optimization” would be properly classified as a form of traditional mathematical programming, topics such as optimal learning and dynamic programming (which often address problems with discrete actions) address optimization problems that are more familiar in probability and simulation.

The traditional barriers between fields have fallen, which emphasizes the need for students to get broad, interdisciplinary educations that span statistics/machine learning, probability/stochastic modeling, and optimization.

A word cloud image formed from the titles of all the talks provides an indication of the range of topics.



While most of the credit for the success of the conference is due to the exceptional content provided by the attendees, special thanks are due to the local program committee (Amir Ali Ahmadi, Mengdi Wang, Jonathan Eckstein, and Andrzej Ruszczyński), the INFORMS office (Paulette Bronis and Ellen Tralongo), Princeton's office of Conference and Event Services (Tara Zarillo), and a team of 15 graduate students from Princeton who manned the registration desk and provided a variety of support services.

The conference was not all work. Tours of the historical university campus were conducted after sessions on Thursday. This led to a wine and cheese party Thursday evening in the elegant Chancellor Green Hall that featured jazz music from our own Jonathan Eckstein and Carlos Oliveira. A nice banquet dinner on Friday provided a setting for more conversation. And as far as I could tell, we met the most important goal of the conference: flow in = flow out!

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, 2016. 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 Nashville, TN in November 2016 (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:

- Tamás Terlaky (Chair)
terlaky@lehig.edu
- Philip Gill
- Dorit Hochbaum
- 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
- Ignacio Grossmann
- Ted Ralphs
- Alex Shapiro

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:

- Sam Burer (Chair)
samuel-burer@uiowa.edu
- Jon Lee
- Andrew Schaefer
- Cole Smith

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:

- James Luedtke (Chair)
jrluedt1@wisc.edu
- Güzin Bayraksan
- Shiqian Ma
- Mohit Tawarmalani

Nominations of Candidates for Society Officers Sought

We would like to thank three Society Vice-Chairs who will be completing their two-year terms at the conclusion of the INFORMS 2016 annual meeting: Aida Khajavirad, Daniel Robinson, and Warren Powell.

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

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

Self-nominations for all of these positions are encouraged.

Vice-Chairs 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.”

Please send your nominations or self-nominations to Burcu Keskin (bkeskin@cba.ua.edu), including

contact information for the nominee, by June 30, 2016. Online elections will begin in mid-August, with new officers taking up their duties at the conclusion of the 2016 INFORMS annual meeting.

Seeking a Host for the 2018 INFORMS Optimization Society Conference

The INFORMS Optimization Society Conference is held in the early part of the even years, often in a warm, or otherwise attractive, location. The most recent OS conference, held in 2016 at Princeton University, was a great success, offering an opportunity for researchers studying optimization-related topics to share their work in a focused venue. (See the short report on the conference by its General Chair, Warren Powell, earlier in this issue.) The Optimization Society is currently seeking candidate locations to host the 2018 conference. If you are interested in helping to host the conference, please contact the current Optimization Society chair, Suvrajeet Sen (sen@datadrivendecisions.org), or the chair-elect David Morton (david.morton@northwestern.edu).