# Mathematisches Forschungsinstitut Oberwolfach

# Mixed-integer Nonlinear Optimization: a hatchery for modern mathematics

Organized by
Leo Liberti, Palaiseau
Sebastian Sager, Magdeburg
Angelika Wiegele, Klagenfurt

2 June – 8 June 2019

ABSTRACT. The second MFO Oberwolfach Workshop on Mixed-Integer Nonlinear Programming (MINLP) took place between 2nd and 8th June 2019. MINLP refers to one of the hardest Mathematical Programming (MP) problem classes, involving both nonlinear functions as well as continuous and integer decision variables. MP is a formal language for describing optimization problems, and is traditionally part of Operations Research (OR), which is itself at the intersection of mathematics, computer science, engineering and econometrics. The scientific program has covered the three announced areas (hierarchies of approximation, mixed-integer nonlinear optimal control, and dealing with uncertainties) with a variety of tutorials, talks, short research announcements, and a special "open problems" session.

## Introduction by the Organizers

This report refers to the second workshop on MINLP organized at Oberwolfach. We refer to the report following the first workshop [4] for a somewhat longer definition of MINLP. In summary, MINLP is one of the most general classes of MP, which is itself a formal language used to describe optimization problems in terms of parameters (the input of the problem), decision variables (which will contain the output after the solution procedure), an objective function to be optimized, and some constraints to be satisfied.

The workshop was organized in 5 tutorial talks (one per day, one hour long, including 15 minutes for questions), 21 "normal" talks (45 minutes long, including 15 minutes for questions), 11 short research announcements (SRA — 15 minutes

long, including 5 minutes for questions), and one open problems session proposing 9 new open problems in the field of MINLP, and attended by everyone at the workshop. The discussion after practically all talks was lively and filled with questions from many attendees. As Oberwolfach tradition warrants, we spent Wednesday afternoon hiking towards a scrumptious Schwarzwälderkirschtorte in St. Roman, a little more than 8km away from the Institute.

## 1. Scientific areas

The scientific organization of this workshop was divided into three main areas.

1.1. **Hierarchies of approximation.** In general MINLP is undecidable, since Universal Diophantine Equations can be easily encoded within MINLP; if all decision variables are bounded, however, MINLP becomes decidable [2]. Most of the well-known decidable restrictions are **NP**-hard, including cases such as Quadratic Programming (QP) with all continuous variables, a quadratic form $x^\top Q x$ with a single negative eigenvalue in the objective function to be minimized, and linear constraints to be satisfied [5]. Among the few cases in **P** we find the (decision version of) minimization QPs with $Q$ positive semidefinite (psd), and the minimization of indefinite QPs over the unit ball [8]. As soon as integer variables appear, most of the instances arising in practice can reasonably be argued to belong to **NP**-hard classes. Although the nonlinear functions arising in MINLP may be transcendental (e.g. log, exp and trigonometric functions), most of the developed theory is limited to polynomials. This discussion on complexity motivates the choice of the first research area that our workshop hosted: *hierarchies of relaxations*. Given a set $S$ of polynomials with a given number of variables $n$ and a given maximum degree $d$, one can represent a sequence of polynomials $p_1(x), \ldots, p_m(x)$ by means of a matrix-by-vector product $Ay$, where $y$ is the vector of all of the $k$ possible monomials on $n$ variables of degree $d$, and $A$ is $m \times k$ matrix. Relaxations can be obtained over the restricted class of Sum-Of-Squares (SOS) polynomials using quadratic forms $y^\top Q y$, and then linearizing all of the monomials of degree $2d$ by means of a matrix $Y$ (representing $yy^\top$) and the corresponding linear form $Q \bullet Y$. Such relaxations can be used to solve difficult MINLP instances.

1.2. **Mixed-Integer Optimal Control.** MINLP applications arise in many real-life problems. A particularly challenging class is in parameter optimization of mixed-discrete dynamical systems, also known as *Mixed-Integer Nonlinear Optimal Control* (MINOC), the second research area of this workshop. In the smallest cases, and in absence of integer controls, the polynomial modelling referred to above can be used to solve MINOC problems in terms of SOS hierarchies. This avoids the need for discretizing time [3]. Applications of this methodology can be found in robotics, or control of unmanned aerial or underwater vehicles. In most other cases, a discretize-then-optimize approach is taken, which unfortunately yields large-scale instances [6]. It then becomes necessary to exploit problem structure in order to derive results allowing a simplification: e.g. the construction of relaxations yielding approximated solutions after rounding [7].

1.3. **Uncertainties.** The third research area this workshop focused on was *uncertainties*. In optimization, dealing with uncertainties in the data is often necessary, as input data from real problems are often noisy and even downright wrong. Classical topics in this area include: (a) robust optimization, which entails seeking for optimal solutions which remain optimal even when the input data differs from the actual real data up to an uncertainty set (which is often an interval or a polyhedron), (b) stochastic optimization, where the uncertainties are modelled by random variables with a given probability distribution. In the latter case, the distribution is often used in the framework of chance constraints, where, instead of imposing a constraint such as $g(x) \leq 0$, one asks that $\mathsf{Prob}(g(x) \leq 0) \geq 1 - \delta$ for some given $\delta$. The inverse distribution function is then used to turn the probabilistic constraint into a standard constraint that solution algorithms can deal with.

1.4. **Other areas.** Other areas of research on MINLP optimization were present at the workshop, including development on MINLP solution algorithms and their implementations, various types of non-hierarchical reformulations and relaxations, as well as many applications of MINLP to other branches of mathematics (geometry, statistics), computer science (graph theory, quantum computing, data science), physics (material science), engineering (logistics, safety), econometrics (game theory), medicine (automated cardiac arryhtmia classification).

## 2. Participants

The workshop was attended by 51 participants from 9 countries, distributed as follows:

| DE | US | AT | IT | FR | NL | UK | BR | CA |
|----|----|----|----|----|----|----|----|----|
| 22 | 10 | 5 | 5 | 3 | 2 | 2 | 1 | 1 |

Insofar as OR researchers are affiliated to a multitude of different departments, MINLP also appears to attract a very diverse variety of cultures, and this workshop was no exception: our participants are affiliated to mathematics departments, but also to computer science and engineering departments.

2.1. **Gender balance.** Genderwise, this workshop attracted 16 women (31% of the total number of participants), 12 of which gave talks (in particular, 2 tutorials and 6 normal talks were given by women). The organizers were congratulated on this account by many of the participants, as well as the vice-director of the Institute.

## 3. Open problems

The open problems session took place between 17:15 and 18:30 on thursday. Here follows the list of open problems and an account of the session proceedings.

(1) **S. Leyffer**, **V. Piccialli** and **L. Palagi** discussed the poor abilities of most local Nonlinear Programming (NLP) solvers with respect to taking into account convex cones in input formulations. Specifically, without taking such cones into account, these formulations become problematic, since

they incur into various forms of degeneracy (e.g. local optima may fail to satisfy constraint qualification conditions). On the other hand, it is easy to prove that such conditions are not really problematic if the cone constraints are taken explicitly into account. The proposers encourage the community to devise implementable solution algorithms capable of detecting and recognizing cone structures even in the presence of nonconvexity in the formulation. A discussion ensued:

- **Leyffer**: in Speakman's talk about perspective reformulation in NLP, the reformulated problem violates MFCQ conditions, and is problematic w.r.t. solvers; the true perspective, albeit nondifferentiable at zero, should not pose a problem in general; but when solving thousands or millions of time within sBB, then failures count.
- **Piccialli**: tried those problematic instances using the `filter`, `ipopt`, `knitro`, `snopt` solvers; best were `ipopt`, `knitro`, which are interior point methods (IPM); guessing failure reason for `filter`, `snopt`, both sequential quadratic programming (SQP) methods, is that that they move along the boundary where the MCFQ are not satisfied.
- **Leyffer** invokes a robust solution for perspective reformulations, and recalls that IPMs and SQPs converge on Mathematical Programs with Equilibrium Constraints (MPEC), where MFCQ also fail to be satisfied.
- **Frangioni** and **Linderoth**: success can be achieved using separation from cones, thus the need to integrate convex cone management in local NLP solvers.

(2) **Linderoth** asked for a method to perform sparse Cholesky factorization. Given a psd matrix $K$, find a permutation matrix $\Pi$ s.t. the Cholesky factorization of $\Pi^\top K \Pi$ is as sparse as possible. While it would be nice to find a formulation for obtaining the globally optimal sparsity, **Ahmadi** suggests that when using the Cholesky factorization in iterative Diagonally Dominant Programming (DDP), a fast and sub-optimal method would be sufficient.

(3) **Ahmadi** asks for the decidability (in the Turing model) of determining stability of a dynamical system in a half-space. Given an $n \times n$ rational matrix $G$, a rational vector $a \in \mathbb{Q}^n$, a scalar $b \in \mathbb{Q}$, and a rational vector $x$ such that $a^\top G x \leq b$, is there a finite algorithm to test whether

$$\forall k \in \mathbb{Z}_+ \quad a^\top G^k x \leq b \ ?$$

If the spectral radius of $G$ is $\rho(G) < 1$, then this problem is known to be **NP**-hard. **Poschka** proposes to use the Jordan decomposition of $G$.

(4) **Ahmadi** asks for the complexity class (whether in **P** or **NP**-hard in the Turing model) of the Hurwitz matrix completion problem. A real $n \times n$ matrix is Hurwitz (or stable) if all eigenvalues have negative real parts. Given a partially defined matrix $A$, is it possible to complete it to a Hurwitz matrix? The symmetric case is tractable. Another characterization is that

$A$ is Hurwitz iff

$\exists$ psd $n \times n$ matrix $P$ ($P - I_n$ is psd $\leftrightarrow I_n - A^T P - PA$ is psd).

The motivating application is that, in the differential equation $\dot{x} = Ax$, the origin is asymptotically stable iff $A$ is Hurwitz. Suggestion for an algorithm: alternating between $P$ and $A$ to prove local convergence.

(5) **Averkov** asks for a proof of second-order cone (SOC) or exponential cone representability for the set $C = \mathsf{conv}\{(x, y) \mid y = x^2 - x^p \wedge -\varepsilon \le x \le \varepsilon\}$ with, e.g., $p = 6$. Taking second derivatives, the form is positive around zero, so $y$ is convex around zero.

(6) **Averkov** asks for general SOC or exponential cone representability of convex sets in $\mathbb{R}^2$ or $\mathbb{R}^3$: are there any two or three-dimensional convex semialgebraic sets that are not SOC-representable? A discussion ensued about possible equivalence of exponential and second-order cones: they are not equivalent, since the exponential cone is not semi-algebraic.

(7) **Averkov** asks for the SDP extension complexity of correlation matrices. Consider the cone $\mathsf{CR}_k = \{A \in S_+^\ell \mid \mathsf{diag}(A) = \mathbf{1}\}$. The SDP extension complexity $\mathsf{sxdeg}(C)$ of a cone $C$ is defined as the minimum $n \in \mathbb{N}$ such that $C$ has an extended formulation with finitely many Linear Matrix Inequalities (LMI) of size $n$. The problem is to compute $s_k = \mathsf{sxdeg}(\mathsf{CR}_k)$. We only know that $s_2 = 1$ and $s_3 = 2$.

(8) **Gleixner** discusses the "nasty" QCQP instance presented by Bienstock during his tutorial talk, based on the feasible set

$$
\begin{aligned}
(x_1 - 1)^2 + x_2^2 &\ge 3 + s^2 \\
\forall i \quad a_i - a_{i-1}^2 &= 0 \\
a_0 &= s^2,
\end{aligned}
$$

where the error on $s$ is "amplified" through repeated square roots. Solvers provide a wrong solution. How do we recognize such problems automatically? Possible answer by Renegar is valid but computationally too expensive.

(9) **Liberti** asks for a Mixed-Integer Linear Programming (MILP) formulation of the Maximum Entropy Sampling Problem (MESP) presented by Anstreicher in his talk.

## 4. Short research announcements

Two SRA sessions were organized (on Tuesday and Thursday).

- A. Potschka, *Penalty alternating direction methods for mixed-integer optimal control with combinatorial constraints,* with one direction being towards satisfying integrality of binary variables, e.g. by means of a feasibility pump algorithm.
- S. Zhao, *Extension of Alternating Direction Method of Multiplier and Semidefinite Programming*, leading to natively taking into account inequality constraints, as well as fine-tuning of certain parameters.

- R. Schulz, *Algebraic gas polynomials* arising in gas network management. The flow and pressure equations lead to quadratic polynomial equations.
- N. Gusmeroli, *An Exact Penalty Method over Discrete Sets* is going to be implemented to a high-performance solver for binary quadratic problems called BiqBin (following BiqMac).
- A. Khajavirad, *Convexification of block-permutation invariant sets* is interrupted for lack of time, leaving half the audience in intolerable suspense for want of closure (the other half went off to get coffee).
- A. Frangioni, `SMS++`*: a modelling system with focus on "Large-Scale" structure* allows the C++ modelling of nested and permuted block-structure constraint matrices in such a way as to make it easier to handle reformulations, restrictions, relaxations and parallel computation.
- A. Fügenschuh, *Call for papers for a special issue on Optimization and Engineering* on mixed-integer optimization with differential equation constraints.
- M. Cerulli, *Flying safely on two levels* involves the solution of bilevel programming problems in order to keep aircraft from bumping into each other midflight and causing a flaming ball of fire.
- M. Hahn, *Binary Optimal Control Using NLP Methods In Measure Spaces* can be achieved in Banach spaces by optimizing over $\sigma$-algebras using sublevel sets of the Radon-Nikodym derivative.
- V. Piccialli, *A new branch and bound for finite Nash games with switching costs* based on a standard QP (stQP) formulation.
- S. Sager, *MINLP-Enhanced Machine Learning for Cardiac Arrhythmia Classification* presents the research at this workshop most likely to have a positive societal impact.

## 5. The future

This is the second edition of MINLP workshops at Oberwolfach. The attendees were enthusiastic, and many asked about future editions. Although there are a variety of MINLP workshops organized by different members of our community, the Oberwolfach ones are those which are more pertinent to mathematics. We shall accordingly endeavour to continue this trend, as long as the MFO Scientific Committee agrees with this idea. A special issue of Mathematical Programming Series B (co-edited by the organizers) is going to be dedicated to the topics of this workshop.

## References

[1] M. Bussieck, A. Drud, and A. Meeraus. MINLPLib – A collection of test models for mixed-integer nonlinear programming. *INFORMS Journal on Computing*, 15(1), 2003.

[2] L. Liberti. Undecidability and hardness in mixed-integer nonlinear programming. *RAIRO-Operations Research*, 53:81–109, 2019.

[3] A. Majumdar, A. Ahmadi, and R. Tedrake. Control and verification of high-dimensional systems with dsos and sdsos programming. In *Conference on Decision and Control*, volume 53, pages 394–401, Piscataway, 2014. IEEE.

[4] MFO-OWR mixed-integer nonlinear programming workshop report. In L. Liberti, S. Sager, and A. Wiegele, editors, *MFO-OWR MINLP Workshop Report*, volume 46 of *Oberwolfach Reports*. European Mathematical Society, Zürich, 2015.

[5] P. Pardalos and S. Vavasis. Quadratic Programming with one negative eigenvalue is NP-hard. *Journal of Global Optimization*, 1:15–22, 1991.

[6] K. Pruitt, S. Leyffer, A. Newman, and R. Braun. A mixed-integer nonlinear program for the optimal design and dispatch of distributed generation systems. *Optimization and Engineering*, 2013.

[7] S. Sager. *Numerical methods for mixed-integer optimal control problems.* PhD thesis, Universität Heidelberg, 2005.

[8] S.A. Vavasis. *Nonlinear Optimization: Complexity Issues.* Oxford University Press, Oxford, 1991.

## Workshop: Mixed-integer Nonlinear Optimization: a hatchery for modern mathematics

## Table of Contents

# Abstracts

### Learning dynamical systems with side information
AMIR ALI AHMADI

(joint work with Bachir El Khadir)

We study the problem of learning dynamical systems from very limited data but in presence of "side information", such as physical laws or contextual knowledge. This is motivated by safety-critical applications where an unknown dynamical system needs to be controlled after a very short learning phase where a few of its trajectories are observed. (Imagine, e.g., the task of autonomously landing a passenger airplane that has gone through sudden wing damage.) We show that sum of squares optimization is particularly suited for exploiting side information in order to assist the task of learning when data is limited.

### Efficient solution of maximum-entropy sampling problems
KURT ANSTREICHER

The maximum-entropy sampling problem (MESP) arises in spatial statistics. In a typical application, $C$ is a sample covariance matrix obtained from time-series observations of an environmental variable at $n$ locations, and it is desired to choose $s$ locations from which to collect subsequent data so as to maximize the information obtained. We assume throughout that $C$ is positive definite. The resulting problem is then

$$\text{MESP}: \qquad z(C, s) := \max\left\{\text{ldet } C[S, S] \ : \ S \subset N, \ |S| = s\right\},$$

where ldet denotes the natural logarithm of the determinant, $N = \{1, \ldots, n\}$, and for a subset $S \subset N$, $C[S, S]$ denotes the submatrix of $C$ having rows and columns indexed by $S$. The use of entropy as a metric for information, together with the assumption that values at the $n$ locations are drawn from a multivariate normal distribution, leads naturally to the problem MESP because ldet $C[S, S]$ is, up to constants, the entropy of the Gaussian random variables having covariance matrix $C[S, S]$. For survey articles describing the MESP see [5, 6].

The study of exact algorithms for MESP was initiated in [4]. Exact algorithms to compute a maximum-entropy design use the "branch-and-bound" (B&B) framework, for which a key ingredient is the methodology for producing an upper bound on $z(C, s)$. Subsequent nodes in the B&B tree, corresponding to indices being fixed into or out of $S$, result in problems of the same form as MESP but with modified data $(C', s')$. A fast method that can provide a good upper bound on $z(C, s)$ is critical to the success of this approach. The exact algorithm in [4] used a bound based on the eigenvalues of $C$. A variety of different bounding methods have subsequently been developed and investigated, and several of these have been incorporated into complete B&B algorithms. Recent results using the optimized "masked spectral" [1] and Boolean quadric polytope (BQP) bounds [2]

are the most promising so far, although both of these bounds involve challenging computational problems posed over $n \times n$ positive semidefinite matrices.

In this paper we consider a new bound for the MESP that is based on maximizing a function of the form $\operatorname{ldet} M(x)$ subject to the constraints $0 \leq x_i \leq 1$, $i = 1, \ldots, n$, $\sum_{i=1}^{n} x_i = s$, where solutions of MESP correspond to binary $x$. This bound has similarities with both the nonlinear programming (NLP) bound of [3] and the BQP bound [2]. The NLP bound is based on maximizing a function of the form $\operatorname{ldet} M(x)$ over the same constraints, where $M(x)$ is a nonlinear function of $x$. For appropriate parameter choices this function $\operatorname{ldet} M(x)$ is provably concave, although the form of $M(x)$ is too complex for $\operatorname{ldet} M(x)$ to be recognized as concave by a "disciplined" convex-programming system such as cvx. The BQP bound is based on maximizing a function of the form $\operatorname{ldet} M(X)$, where $M(X)$ is linear in the $n \times n$ semidefinite matrix variable $X$. We refer to the new bound as the "linx" bound because $M(x)$ is linear in $x$. Validity of the linx bound is based on the following simple but previously unexploited determinant identity.

**Lemma 1.** *For a subset $S \subset N = \{1, \ldots, n\}$ let $x_i = 1$, $i \in S$ and $x_i = 0$, $i \in N \setminus S$. Then $\operatorname{ldet}(C \operatorname{Diag}(x)C + I - \operatorname{Diag}(x)) = 2 \operatorname{ldet} C_{SS}$.*

Motivated by Lemma 1, we define the linx bound via the optimization problem

$$
\begin{aligned}
z_{\operatorname{linx}}(C, s) = \quad & \max \quad \frac{1}{2} \operatorname{ldet}\left(C \operatorname{Diag}(x)C + I - \operatorname{Diag}(x)\right) \\
& \text{s.t.} \quad e^T x = s \\
& \qquad\; 0 \leq x \leq e,
\end{aligned}
$$

(1)

where $e$ is a vector of ones. Validity of the bound, $z_{\operatorname{linx}}(C, s) \geq z(C, s)$ then follows immediately from Lemma 1, and solving MESP corresponds to finding the optimal binary solution of (1). The convex optimization problem (1) has a dual which also corresponds to a determinant maximization problem with linear constraints,

$$
\begin{aligned}
& \min \quad -\frac{1}{2} \operatorname{ldet} S \\
& \text{s.t.} \quad \operatorname{tr}(S) + su + e^T v = n \\
& \qquad\; \operatorname{diag}(CSC) - \operatorname{diag}(S) \leq ue + v \\
& \qquad\; v \geq 0, S \succeq 0.
\end{aligned}
$$

(2)

The objective gap between primal and dual solutions in (1) and (2), for convenience ignoring the factor $1/2$ in both, is then

$$
\begin{aligned}
-\operatorname{ldet} S - \operatorname{ldet} M(x) \;=\; & -\operatorname{ldet} S - \operatorname{ldet}\left(C \operatorname{Diag}(x)C + I - \operatorname{Diag}(x)\right) \\
=\; & -\operatorname{ldet} S^{1/2}(C \operatorname{Diag}(x)C + I - \operatorname{Diag}(x))S^{1/2} \\
\geq\; & -\operatorname{diag}(CSC)^T x - \operatorname{tr}(S) + \operatorname{diag}(S)^T x + n \\
=\; & w^T x - ue^T x - v^T x - \operatorname{tr}(S) + n \\
=\; & u(s - e^T x) + v^T(e - x) + w^T x + n - us - v^T e - \operatorname{tr}(S) \\
=\; & v^T(e - x) + w^T x,
\end{aligned}
$$

(3)

where the inequality uses the fact that for any $X \succ 0$, $\operatorname{ldet} X \le \operatorname{ldet} I + \operatorname{tr}(X - I) = \operatorname{tr}(X) - n$. It is straightforward to prove that strong duality always holds between problems (1) and (2). The weak duality condition (3) can be used to derive variable-fixing logic for MESP when the linx bound is applied. Let $\hat{z}$ be the objective value for a known solution of MESP. Assume that $\bar{x}$ solves (1), and $(\bar{u}, \bar{v}, \bar{w}, \overline{S})$ solves (2), where $\bar{w} = \bar{u} + \bar{v} + \operatorname{diag}(\overline{S}) - \operatorname{diag}(C\overline{S}C)$. Let $z_{\text{linx}} = \frac{1}{2} \operatorname{ldet} M(\bar{x}) = -\frac{1}{2} \operatorname{ldet} \overline{S}$. From (3) we know that for any other $x$ feasible in (1),

$$\operatorname{ldet} M(x) \le 2 z_{\text{linx}} - \bar{v}^T(e - x) - \bar{w}^T x.$$

It follows that if $x^*$ is a binary solution in (1) with objective value greater than $\hat{z}$,

$$\bar{w}_i \ge 2(z_{\text{linx}} - \hat{z}) \implies x_i^* = 0, \qquad \bar{v}_i \ge 2(z_{\text{linx}} - \hat{z}) \implies x_i^* = 1.$$

The dual variables $(\bar{v}, \bar{w})$ can also be used to devise an effective branching strategy in the context of a B&B algorithm for MESP.

Computation of the linx bound is facilitated by the fact that version 4.0 of the Matlab-based SDPT3 solver [7] allows for the objective $\operatorname{ldet} M(x)$ when $M(x)$ is linear in $x$. Comparisons on benchmark instances demonstrate outstanding results for the linx bound. A complete B&B algorithm using the linx bound obtains the first solutions of benchmark instances for a matrix $C$ with $n = 124$, summarized in Table 1. A notable feature of the B&B algorithm is that the dual-based branching strategy typically obtains the optimal solution very quickly, as shown in the last column of the table.

TABLE 1. B&B statistics for $n = 124$ instances

| $s$ | Heuristic Value | Optimal Value | Root Gap | Number Nodes | Time (hours) | Time to Opt (sec) |
|---|---|---|---|---|---|---|
| 10 | 43.918 | 43.918 | 0.363 | 207 | 0.04 | 155 |
| 20 | 76.999 | 77.827 | 2.306 | 1,239 | 0.21 | 181 |
| 30 | 106.674 | 106.700 | 2.010 | 3,413 | 0.47 | 147 |
| 40 | 130.162 | 131.055 | 3.304 | 11,202 | 2.42 | 167 |
| 50 | 148.661 | 149.498 | 4.197 | 159,302 | 51.81 | 1,515 |
| 60 | 163.371 | 164.012 | 3.992 | 171,383 | 67.01 | 171 |
| 70 | 172.243 | 172.528 | 3.680 | 140,240 | 37.93 | 215 |
| 80 | 174.813 | 175.091 | 3.298 | 42,989 | 12.71 | 230 |
| 90 | 171.262 | 171.262 | 2.918 | 21,495 | 4.82 | 249 |
| 100 | 162.616 | 162.865 | 2.393 | 5,334 | 1.15 | 298 |
| 110 | 147.730 | 147.933 | 1.919 | 1,589 | 0.40 | 343 |

## REFERENCES

[1] K. Anstreicher and J. Lee. A masked spectral bound for maximum-entropy sampling. In *mODa 7—Advances in model-oriented design and analysis*, Contrib. Statist., pages 1–10. Physica, Heidelberg, 2004.

[2] K. M. Anstreicher. Maximum-entropy sampling and the Boolean quadric polytope. *J. Global Optim.*, 72:603-618. 2018.

[3] K. M. Anstreicher, M. Fampa, J. Lee, and J. Williams. Using continuous nonlinear relaxations to solve constrained maximum-entropy sampling problems. *Mathematical Programming*, 85(2):221–240, 1999.

[4] C.-W. Ko, J. Lee, and M. Queyranne. An exact algorithm for maximum entropy sampling. *Operations Research*, 43(4):684–691, 1995.

[5] J. Lee. Semidefinite programming in experimental design. In R. Saigal, H. Wolkowicz, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming*, volume 27 of *International Series in Operations Research and Management Science*, pages 528–532. Kluwer Acad. Publ., Boston, MA, 2000.

[6] J. Lee. Maximum-entropy sampling. In A. H. El-Shaarawi and W. W. Piegorsch, editors, *Encyclopedia of Environmetrics*, volume 3, pages 1229–1234. John Wiley, 2001.

[7] K.-C. Toh, M. J. Todd, and R. H. Tutuncu. On the implementation and usage of SDPT3 - a MATLAB software package for semidefinite-quadratic-linear programming, version 4.0. In M. Anjos and J. Lasserre, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*, volume 166 of *International Series in Operations Research and Management Science*, pages 715–754. Springer, 2012.

# Semidefinite approaches to polynomial optimization: Power and limitations of the SOS cones

## Gennadiy Averkov

We give an overview on how semidefinite programming is used in polynomial optimization and also discuss the limitations of the current approach based on the SOS cones. Let's start by introducing conic and semidefinite programming. General conic programming with respect to a closed convex cone $K$ is the problem

$$\inf \left\{ c^\top x \,:\, Ax = b,\ x \in K \right\}$$

of optimization of a linear objective function subject to a system $Ax = b$ of linear and inequalities and the condition that the vector $x$ of the optimization variables is in the cone $K$. The case $\mathbb{R}^n_+$ gives linear programming, while the case

$$K = \mathcal{S}^k_+ := \{ k \times k \text{ symmetric psd matrices over } \mathbb{R} \}$$

gives *semidefinite programming (SDP)*. One can express SDP in terms of the so-called linear matrix inequalities (LMIs). Consider a $k \times k$ symmetric matrix

$$A(x) := \Big( a_{ij}(x) \Big)_{i,j=1,\ldots,k}$$

with entries $a_{ij}(x)$ being affine functions in $x \in \mathbb{R}^n$. The condition

$$A(x) \in \mathcal{S}^k_+$$

is called a *linear matrix inequality (LMI)* of size $k$ on $n$ real-valued variables $x \in \mathbb{R}^n$, while the respective set

$$\left\{ x \in \mathbb{R}^n \,:\, A(x) \in \mathcal{S}^k_+ \right\}$$

is called a *spectrahedron*. *Semidefinite programming* is optimization of a linear function subject to finitely many LMIs [WSV00, AL12]. SDP is efficiently solvable using interior-point methods under mild assumptions. *But* if you can avoid LMIs

of large size, you should really do that because of the running-time and numerical-stability issues.

A nice thing about SDP is that some very basic classes of algorithmic and optimization problems can naturally be phrased as a special case of SDP. Linear programming is a subset of SDP, since linear constraints are LMIs of size 1. Determination of the maximum eigenvalue of a symmetric matrix is a semidefinite problem with an LMI on one variable:

$$\min \left\{ \lambda \, : \, \lambda I - A \in \mathcal{S}_+^n \right\}.$$

LMIs frequently allow to convexify non-convex problems of algebraic nature so that afterwards SDP can be used to solve underlying optimization problems. Numerous application areas of SDP include problems in probability and statistics, coding theory, systems and control theory and combinatorial optimization [WSV00].

We say that a set $C$ has an *extended formulation* with $m$ LMIs of size $k$ if $C$ is a linear image of a spectrahedron described by $m$ LMIs of size $k$. A standard way to reduce optimization of a linear function over a given semi-algebraic convex set $C$ is by providing a semidefinite extended formulation of $C$ and lifting the underlying optimization problem over $C$ to an optimization problem over the respective spectrahedron.

In what follows, we deal with polynomial with real coefficients. Polynomial optimization is optimization of a polynomial objective function subject to finitely many polynomial inequalities. An approach of Lasserre to solving polynomial optimization problems is based on Positivstellensätze (which describe positivity of polynomials in terms of sum-of-squares certificates) and semidefinite formulations of the so-called sum-of-squares cones [Mar08, Lau09, Las15]. A polynomial is called a *sum of squares* if it can be represented as a sum of squares of finitely many polynomials. For given positive integers $n$ and $d$, we introduce the *sum-of-squares cone* $\Sigma_{n,2d}$ to be the cone of $n$-variate sum-of-squares polynomials of degree at most $2d$. The cone $\Sigma_{n,2d}$ is known to have a semidefinite extended formulation with one LMI of size $\binom{n+d}{n}$ [Las15, §2.1]. To approximate the optimal value of a polynomial optimization problem following Lasserre's approach, one establishes a hierarchy of SDPs which based on the mentioned semidefinite extended formulations of $\Sigma_{n,2d}$ with growing values of $d$. The approach allows to obtain strong approximations of polynomial optimization problems at a very high computational cost due to the LMIs of a very large size that are used in the hierarchy of the SDPs.

So far, it has not been clear if the known semidefinite formulation of $\Sigma_{n,2d}$ is optimal in terms of the size of the LMIs. We present a theorem that allows to confirm that the known semidefinite extended formulation of $\Sigma_{n,2d}$ is best possible.

We consider the *semidefinite extension complexity* of a set $C$ (denoted as $\mathrm{sxc}(C)$), which is the smallest $k$ such that $C$ has a semidefinite extended formulation with one LMI of size $k$, and introduce the *semidefinite extension degree* of $C$ (which we denote as $\mathrm{sxdeg}(C)$) to be the smallest $k$ such that $C$ has a semidefinite extended formulation with finitely many LMIs of size $k$.

**Theorem 1** (Main theorem)**.** *Let $X \subseteq \mathbb{R}^n$ be a set with non-empty interior. Let $C$ be a closed convex cone in the space of n-variate polynomials of degree at most $2d$ such that every polynomial in $C$ is non-negative on $X$ and there exist finite subsets $S$ of $X$ of arbitrarily large cardinality with the following property:*

(∗) *For every k-element subset $T$ of $S$, some polynomial $f$ in the cone $C$ is equal to zero on $T$ and is strictly positive on $S \setminus T$.*

*Then $\operatorname{sxdeg}(C) > k$.*

Using Theorem 1, we obtain

**Corollary 2.** $\operatorname{sxdeg}(\Sigma_{n,2d}) = \operatorname{sxc}(\Sigma_{n,2d}) = \binom{n+d}{n}$.

Corollary 2 shows that the known semidefinite formulation of $\Sigma_{n,2d}$ is best possible in terms of both the size and the number of the LMIs.

The case $d = 1$ of Corollary 2 yields the semidefinite extension degree of $\mathcal{S}_+^k$:

**Corollary 3.** $\operatorname{sxdeg}(\mathcal{S}_+^k) = k$.

Corollary 3 implies that the expressive power of the semidefinite optimization grows strictly with the growth of the size $k$ of the underlying LMIs. In other words, the family of all convex semialgebraic sets that have a semidefinite extended formulation (we call such sets *semidefinitely representable*) can be decomposed into the hierarchy of the families

$$\operatorname{SDR}(k) := \{S \subseteq \mathbb{R}^n \,:\, n \in \mathbb{N},\ \operatorname{sxdeg}(S) \le k\}$$

with each level of the hierarchy being strictly larger than the previous one. The lowest level $\operatorname{SDR}(1)$ of the hierarchy is just the family of all polyhedra. The family $\operatorname{SDR}(1)$ corresponds to linear optimization. The next level $\operatorname{SDR}(2)$ corresponds to the second-order cone programming.

Corollary 3 covers the result $\operatorname{sxdeg}(\mathcal{S}_+^3) = 3$ of Fawzi [Faw19] as a special case.

References

[AL12]   Miguel F. Anjos and Jean B. Lasserre (eds.), *Handbook on semidefinite, conic and polynomial optimization*, International Series in Operations Research & Management Science, vol. 166, Springer, New York, 2012.

[Faw19]  Hamza Fawzi, *On representing the positive semidefinite cone using the second-order cone*, Math. Program. **175** (2019), no. 1-2, Ser. A, 109–118.

[Las15]  Jean Bernard Lasserre, *An Introduction to Polynomial and Semi-algebraic Optimization*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2015.

[Lau09]  Monique Laurent, *Sums of squares, moment matrices and optimization over polynomials*, Emerging applications of algebraic geometry, IMA Vol. Math. Appl., vol. 149, Springer, New York, 2009, pp. 157–270.

[Mar08]  Murray Marshall, *Positive Polynomials and Sums of Squares*, Mathematical Surveys and Monographs, vol. 146, American Mathematical Society, Providence, RI, 2008.

[WSV00]  Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe (eds.), *Handbook of semidefinite programming*, International Series in Operations Research & Management Science, vol. 27, Kluwer Academic Publishers, Boston, MA, 2000, Theory, algorithms, and applications.

## A survey of QCQP results that I like

DANIEL BIENSTOCK

QCQPs, or quadratically constrained quadratic programs, present a number of complex challenges both theoretical and practical. They are very expressive – QCQPs can be used to efficiently represent any polynomial optimization problem. As a result, QCQPs can attain very badly behave feasible sets. For example, already in low dimension a QCQP can have a unique optimal solution, which is an isolated point and irrational. Hence any solution obtained via a standard computational algorithm will either be infeasible or strictly suboptimal, and (again) examples can be constructed where this subobtimality is substantial. Perhaps more important, one can construct examples of QCQPs that are *near* badly-behaved examples, with the result that solutions obtained by floating point procedures can be substantially *super*-optimal. From a practical perspective, even simple instances of the well-known ACOPF problem in electrical power transmission can exhibit substantial complexity, see [2].

In this talk we surveyed work by Renegar on solutions to polynomially constrained problems. One key result is that given a system of polynomial inequalities, and $\epsilon > 0$ there is a finite bit-model procedure that constructs a family $F$ of points such that each for each connected component of the solution set for the system, some point in $F$ is within distance $\epsilon$ of that component. The particular result we surveyed is a key result in [3], which provides a finite procedure, again in the bit-model of computing, for approximating all zeros of a system of polynomial equations. This procedure is based on a simplification of the well-known concept of *resultant* of a system of algebraic equations.

Next we considered the "2-QCQP" problem, which is a QCQP with just two constraints. This is a special case of QCQP with a fixed number of constraints, which is known to be polynomially solvable [1]. In the special case of 2-QCQP where one of the constraints is strictly convex, [4] presents an interesting procedure that solves the problem (in the real number model of computing) in polynomial time, relying on a construction that makes use of Bézoutians and Kronecker product of matrices.

### REFERENCES

[1] D. Bienstock, *A note on polynomial solvability of the CDT problem*, SIAM J. Optimization **26** (2016) 486–496.

[2] I.A. Hiskens and R.J.Davy, *Exploring the Power Flow Solution Space Boundary*, IEEE Trans. Pow. Sys **16** (2001), 389–395.

[3] J. Renegar, *On the Computational Complexity and Geometry of the First-order Theory of the Reals, Part I: Introduction. Preliminaries. The Geometry of Semi-algebraic Sets. The Decision Problem for the Existential Theory of the Reals*, J. Symbolic Computation **13** (1192), 255–299.

[4] S. Sakaue, Y. Nakatsusaka, A. Takeda and S. Iwata, *Solving Generalized CDT Problems via Two-Parameter Eigenvalues*, SIAM J. Optimization **26** (2016), 1669–1694.

# An active set algorithm for robust combinatorial optimization
Marianna De Santis
(joint work with Christoph Buchheim)

We address combinatorial optimization problems given in the general form

$$\text{(CP)} \qquad \begin{array}{ll} \min & c^\top x \\ \text{s.t.} & x \in P \cap \mathbb{Z}^n, \end{array}$$

where $P \subseteq \mathbb{R}^n$ is a compact convex set, say $P \subseteq [l, u]$ with $l, u \in \mathbb{R}^n$, and the objective function vector $c \in \mathbb{R}^n$ is assumed to be uncertain. This setting appears in many applications where the feasible set is certain, but the objective function coefficients may have to be estimated or result from imprecise measurements. As an example, when searching for a shortest path in a road network, the topology of the network is usually considered fixed, but the travel times may vary depending on the traffic conditions. We assume that the uncertain coefficients of the objective function are varying over ellipsoidal uncertainty sets. The robust counterpart of such a problem can be rewritten as the following second-order cone program (SOCP) with integrality constraints

$$\text{(P)} \qquad \begin{array}{ll} \min & f(x) = c^\top x + \sqrt{x^\top Q x} \\ \text{s.t.} & x \in P \cap \mathbb{Z}^n, \end{array}$$

being $Q$ a positive definite matrix. In order to solve Problem (P), we propose a branch-and-bound algorithm where dual bounds are computed by means of an active set algorithm. The latter is applied to the Lagrangian dual of the continuous relaxation, where the feasible set of the combinatorial problem is supposed to be given by a separation oracle. The method benefits from the closed form solution of the active set subproblems and from a smart update of pseudo-inverse matrices.

We tested our approach on randomly generated instances and on instances from different combinatorial problems, including the shortest path and the traveling salesman problem. Here we report the comparison of our branch-and-bound method `BB-ELLAS` with the state-of-the art mixed-integer SOCP solver of Gurobi on instances from the minimum spanning tree problem. Given an undirected weighted graph $G = (V, E)$, a minimum spanning tree is a subset of edges that connects all vertices, without any cycles and with the minimum total edge weight. Our approach uses the following formulation of the Robust Spanning Tree problem:

$$\begin{array}{rrcl} \min & c^\top x + \sqrt{x^\top Q x} & & \\ \text{s.t.} & \sum_{e \in E} x_e & = & |V| - 1 \\ & \sum_{e \subseteq X} x_e & \leq & |X| - 1 \quad \forall \emptyset \neq X \subseteq V \\ & x & \in & \{0, 1\}^E \end{array}$$

In the above model, the number of inequalities, taking into account also the non-negativity constraints, is $m = 2^{|V|} - 2 + |E|$. Since this number is exponential in the input size, we also have to use a separation algorithm for Gurobi. For

both `BB-EllAS` and Gurobi, we use a simple implementation based on the Ford-Fulkerson algorithm. For our experiments, we considered complete graphs with expected edge weights randomly generated in $[0.5, 1.5]$. The positive definite matrix $Q$ has been built as follows. We chose $n$ eigenvalues $\lambda_i$ uniformly at random from $[0, 1]$ and orthonormalized $n$ random vectors $v_i$, each entry of which was chosen uniformly at random from $[-1, 1]$. Setting $\bar{Q} = \sum_{i=1}^{n} \lambda_i v_i v_i^\top$, the entries of $Q$ are given as $Q_{ij} = c_i c_j \bar{Q}_{ij}$, where $c$ is the vector defining the linear term.

| | | | BB-EllAS | | | Gurobi | | |
|---|---|---|---|---|---|---|---|---|
| $|V|$ | $n$ | $m$ | #sol | time | nodes | #sol | time | nodes |
| 12 | 66 | 4,160 | 10 | 53.51 | 9.5e+4 | 10 | 110.61 | 1.1e+4 |
| 13 | 78 | 8,268 | 10 | 231.11 | 1.9e+5 | 10 | 546.84 | 2.9e+4 |
| 14 | 91 | 16,473 | 10 | 312.34 | 5.9e+5 | 7 | 1802.43 | 7.2e+4 |
| 15 | 105 | 32,871 | 5 | 2388.39 | 2.8e+6 | 2 | 3271.12 | 1.0e+5 |
| 16 | 120 | 65,654 | 1 | 1490.94 | 1.4e+6 | 0 | — | — |

In the above table we show the results obtained on 50 different problem instances: for each $|V| \in \{12, \dots, 16\}$ we generated 10 different complete instances. For both algorithms we report the number of instances solved within the time limit of one hour ($\#$ sol), the average running time, and the average number of branch-and-bound nodes. The branch-and-bound method `BB-EllAS` clearly outperforms the MISOCP solver of Gurobi on the instances considered, being able to solve significantly more instances than Gurobi within the time limit and with faster running time.

## Convexification of the Quadratic Knapsack Problem with Integrated Cut Strengthening

Marcia Fampa

(joint work with Daniela Lubke, Fei Wang, Henry Wolkowicz)

We study a convex quadratic programming (CQP) relaxation of the quadratic knapsack problem (QKP),

$$
\begin{aligned}
p_{\mathrm{QKP}}^* := \quad &\max \quad x^T Q x \\
(\textbf{QKP}) \quad &\text{s.t.} \quad w^T x \le c \\
&\quad\quad x \in \{0, 1\}^n,
\end{aligned}
$$

where $Q \in \mathbb{S}^n$ is a symmetric $n \times n$ nonnegative integer profit matrix, $w \in \mathbb{Z}_{++}^n$ is a vector of positive integer weights for the items, and $c \in \mathbb{Z}_{++}$ is the knapsack capacity with $c \ge w_i$, for all $i \in N := \{1, \dots, n\}$. The binary (vector) variable $x$ indicates which items are chosen for the knapsack, and the inequality in the model, known as a knapsack inequality, ensures that the selection of items does not exceed the knapsack capacity.

The QKP was proved to be NP-Hard in the strong sense by reduction from the clique problem. It is a generalization of the knapsack problem, which has the same feasible set of the QKP, and a linear objective function in $x$. Several papers

have proposed branch-and-bound algorithms for the QKP, and the main difference between them is the method used to obtain upper bounds for the subproblems. The well known trade-off between the strength of the bounds and the computational effort required to obtain them is intensively discussed in the literature.

A common approach to construct relaxations to **QKP** is to lift the problem to the symmetric matrix space determined by the equation $X = xx^T$, and then replace the quadratic objective function with a linear function in $X$, namely, $\mathsf{trace}(()QX)$. As the constraint $X = xx^T$ is nonconvex, it is relaxed by convex constraints in the relaxations. The well known McCormick inequalities and the semidefinite constraint, $X - xx^T \succeq 0$, for example, have been used to relax the nonconvex constraint $X = xx^T$ in relaxations of the QKP.

We investigate the following relaxation **CQP**$(Q_p)$, for **QKP**, where instead of linearizing all the objective function, we perturb the objective function Hessian $Q$, and maintain the (concave) perturbed version of the quadratic function in the objective, linearizing only the remaining part derived from the perturbation.

$$(\mathbf{CQP}\,(Q_p)) \quad p_{\mathrm{CQP}}^*(Q_p) := \max \quad x^T(Q - Q_p)x + \mathsf{trace}(()Q_p X)$$
$$\text{s.t.} \quad (x, X) \in \mathcal{P}.$$

where $Q - Q_p \preceq 0$ and $\mathcal{P}$ is a compact convex subset of $[0,1]^n \times \mathbb{S}^n$, such that

$$\{(x, X) \ : \ w^T x \leq c, \ X = xx^T, \ x \in \{0,1\}^n\} \subset \mathcal{P}.$$

The idea behind **CQP**$(Q_p)$ is to keep quadratic information of the original problem in the relaxation, and has been extensively applied to nonconvex quadratic problems (see, for example, [1] and references therein).

**CQP**$(Q_p)$ is a parametric convex quadratic problem, defined as a function of a matrix parameter $Q_p$, such that $Q - Q_p \preceq 0$. Our approach searches for the matrix parameter that minimizes the upper bound for **QKP**. For that, we consider the parametric problem

$$param_{\mathrm{QKP}}^* := \min_{Q - Q_p \preceq 0} p_{\mathrm{CQP}}^*(Q_p),$$

and develop a primal-dual Interior Point Method (IPM) to solve it.

During the iterations of the IPM, separation problems are solved to generate cuts that are added to **CQP**$(Q_p)$, and the search for the best matrix parameter $Q_p$ is adapted accordingly. The cuts added to the relaxation are derived from new valid inequalities on the lifted matrix variable, which are based on cover inequalities, and referred to as CILS (Cover Inequalities in the Lifted Space) and SCILS (Set of Cover Inequalities in the Lifted Space). The idea is also extended to generate valid inequalities based on knapsack inequalities. Finally, at each iteration of the IPM, lower bounds for the problem are also generated from feasible solutions constructed from a rank-one approximation of the solution of **CQP**$(Q_p)$.

We call our algorithm CWICS (Convexification with Integrated Cut Strengthening). CWICS alternates between optimizing the matrix parameter $Q_p$ with an IPM, and applying cutting planes generated by valid inequalities, to the convex quadratic programming relaxation **CQP**$(Q_p)$, of **QKP** (see [2] for more details). The main characteristics of CWICS are:

- Consists of an IPM with a BFGS step to improve the matrix parameter $Q_p$ on the convex quadratic relaxation $\mathbf{CQP}(Q_p)$, of $\mathbf{QKP}$.
- At every $m$ iterations of the IPM, separation problems are solved to generate cuts CILS and SCILS that are added to $\mathbf{CQP}(Q_p)$.
- When cuts are added to the relaxation, the computation of the search direction of the IPM changes accordingly, as it depends on the optimal solution of $\mathbf{CQP}(Q_p)$.
- At every iteration of the IPM, an upper bound is computed by solving $\mathbf{CQP}(Q_p)$, for the current matrix $Q_p$ and the current set of valid inequalities.
- At every iteration of the IPM, a lower bound is computed by a heuristic based on a rank-one approximation of the solution of $\mathbf{CQP}(Q_p)$.
- CWICS stops when the gap between upper and lower bounds is small enough, or at a maximum number of iterations.

### References

[1] M. Fampa, J. Lee, and W. Melo, *On global optimization with indefinite quadratics*, EURO Journal on Computational Optimization **5(3)** (2017), 309–337.

[2] M. Fampa, D. Lubke, F. Wang, H. Wolkowicz, *Parametric Convex Quadratic Relaxation of the Quadratic Knapsack Problem*, arXiv:1901.06714 [math.OC] (2019).

## Perspective Reformulations Beyond the Separable Case

### Antonio Frangioni

(joint work with Claudio Gentile, James Hungerford)

Perspective Reformulation techniques have shown to be a useful tool, in no small part due to their simplicity, to strengthen formulations of Mixed-Integer Nonlinear problems with disjunctive constraints like semi-continuous variables. One typical application is to problems with convex-cost semicontinuous (vector) variables, i.e.,

$$
\left\{
\begin{array}{ll}
\min & g(z) + \sum_{i \in N} f_i(x_i) + c_i y_i \\
& A_i x_i \le b_i y_i & i \in N \\
& y \in \{0, 1\}^n, \; x \in \Re^m, \; (x, y, z) \in \mathcal{O}
\end{array}
\right.
$$

where each $f_i$ is a closed convex function with (w.l.o.g.) $f_i(0) = 0$, each set $\mathcal{P}_i = \{x_i \in \Re^{m_i} : A_i x_i \le b_i\}$ is a polytope, and the "other" variables / constraints $z / \mathcal{O} \subset \Re^{m+n+q}$ ($m = \sum_{i \in N} m_i$) may in principle have any structure; however, for simplicity of discussion they are better assumed convex. The constraints impose the classical logical relationship $y_i = 0 \implies x_i = 0$, $y_i = 1 \implies x_i \in \mathcal{P}_i$, which is useful to model many real-life situations in applications like energy, transportation, finance and many others. By restricting to a single block

$$
(P) \qquad \min\{f(x) + cy \; : \; Ax \le by, \; y \in \{0, 1\}\}
$$

we have a structure that can be exploited: in fact, it is well-known [1] that the *Perspective Reformulation* of $(P)$

$$(PR) \qquad \min \big\{\, yf(\,x\,/\,y\,) + cy \ : \ Ax \le by \,, \ y \in \{\, 0\,,\, 1\,\} \,\big\}$$

corresponds to the *convex envelope* of the original objective on the non-convex feasible set, and therefore has a (much) better lower bound. The PR can be efficiently solved in different ways, comprised outer-approximation [1], reformulation as a conic program [3] or (possibly, approximate) projection [5]. Using this technique, however, requires separability of the disjunctive constraints/objective in question. There are some ways to sidestep this issue, namely the extraction of a separable part of nonseparable (quadratic) functions via either eigenvalue computation [1] or SemiDefinite Programming (SDP) techniques [2, 4].

We present an improvement to the standard approach doing the latter where, rather then extracting the diagonal of the Hessian of the function, we (approximately) decompose it as the sum of $2 \times 2$ matrices. This is based on the fact that, for the variant of $(P)$ with a nonseparable $2 \times 2$ objective

$$(P_2) \quad \begin{cases} \min & q_{11}x_1^2 + 2q_{12}x_1x_2 + q_{22}x_2^2 \\ & l_iy_i \le x_i \le u_iy_i \,, \ y_i \in \{\, 0\,,\, 1\,\} \quad i = 1, 2 \end{cases}$$

we can provide the apparently awkward reformulation

(1) $\qquad \min q_{11}(x_1^1)^2 + q_{22}(x_2^2)^2 + q_{11}(x_1^{12})^2 + 2q_{12}x_1^{12}x_2^{12} + q_{22}(x_2^{12})^2$

(2) $\qquad\quad x_i = x_i^i + x_i^{12} \quad , \qquad y_i = y^i + y^{12} \qquad\qquad i = 1, 2$

(3) $\qquad\quad l_iy^i \le x_i^i \le u_iy^i \ , \ l_iy^{12} \le x_i^{12} \le u_iy^{12} \qquad i = 1, 2$

(4) $\qquad\quad y^1 + y^2 + y^{12} \le 1 \ , \ y^1 \,, \ y^2 \,, \ y^{12} \in \{\, 0\,,\, 1\,\}$

which, by standard results, immediately provides a (PR) in a lifted space

$$(PR_2) \quad \begin{cases} \min & q_{11}(x_1^1)^2/y^1 + q_{22}(x_2^2)^2/y^2 + \\ & \quad \big[\, q_{11}(x_1^{12})^2 + 2q_{12}x_1^{12}x_2^{12} + q_{22}(x_2^{12})^2 \,\big]/y^{12} \\ & (2)\text{--}(4) \end{cases}$$

This can also be reduced by project away more than half of the extra variables via equalities, although the projection on the original space is known to be difficult to fully characterize [6]. For a general problem with nonseparable convex $f(x) = x^T Q x$, $Q \in S_+^{n \times n}$, one can consider an *approximate* $2 \times 2$-Decomposition (2×2D) of $Q$ as

$$Q = \Phi + \sum_{p \in P} E^p \Pi^p (E^p)^T$$

where $P$ is the set of all $O(n^2)$ unordered pairs $\{\, i\,,\, j\,\}$, $E^p = [\, e_i\,,\, e_j\,] \in \Re^{n \times 2}$ (with $e_h$ as usual the $h$-th vector of the canonical base of $\Re^n$), $\Phi \in S_+^{n \times n}$ and all $\Pi^p \in S_+^{2 \times 2}$. This allows to write an (approximate) *2×2 Perspective Reformulation*

(2×2PR)

$$\min g(z) + x^T \Phi x + q^T x + c^T y+$$

$$\sum_{p \in P} \left[ \Pi_{11}^p (x_i^{p,i})^2 / y^{p,i} + \Pi_{22}^p (x_j^{p,j})^2 / y^{p,j} + (x^{p,p})^T \Pi^p x^{p,p} / y^{p,p} \right]$$

$$(x, y, z) \in \mathcal{O}$$

$$x_i = x_i^{p,i} + x_i^{p,p} \quad , \quad y_i = y^{p,i} + y^{p,p} \qquad p \in P \ , \ i \in p$$

$$l_i y^{p,i} \le x_i^{p,i} \le u_i y^{p,i} \qquad p \in P \ , \ i \in p$$

$$l_i y^{p,p} \le x_i^{p,p} \le u_i y^{p,p} \qquad p \in P \ , \ i \in p$$

$$y^{p,p} + y^{p,i} + y^{p,j} \le 1 \qquad p \in P$$

$$y^{p,i} \in \{\, 0 \, , \, 1 \,\} \ , \ y^{p,p} \in \{\, 0 \, , \, 1 \,\} \qquad p \in P \ , \ i \in p$$

which, with proper choices of $\Phi$ and $\Pi^p$, should have an even improved continuous relaxation abound. We fully characterize the condition under which $\Phi = 0$ is possible, i.e., $Q$ has an *exact* 2×2D. This turns out to be equivalent to the fact that $Q$ is *Weakly Scaled Diagonally Dominant* (WSDD): that is, $\exists\ d > 0$ s.t. $M = \mathsf{diag}(d)\, Q\, \mathsf{diag}(d)$ is *Weakly Diagonally Dominant* (WDD), that is, $|M_{ii}| \ge \sum_{j \ne i} |M_{ij}|$ for all $i$. This can be detected by a simple eigenvalue computation, in that it is also equivalent to

$$\rho(\, |\, I - \mathsf{diag}(Q)^{-\frac{1}{2}}\, Q\, \mathsf{diag}(Q)^{-\frac{1}{2}}\, |\,) \le 1 \ ;$$

being $\lambda < 1$ the maximum eigenvalue, there exists a corresponding eigenvector $x > 0$, and the 2×2D is then provided by the explicit formula

$$Q_{ii}^{\{\, i\, ,\, j\,\}} \ = \ \frac{\lambda |Q_{ij}| \sqrt{Q_{ii}}}{\lambda \sqrt{Q_{jj}}} x_i^{-1} x_j + \frac{Q_{ii}(1 - \lambda)}{n - 1} \ .$$

Actually, the formula can be made parametric, and conditions for the unicity of the decomposition are also given [7]. We then describe six possible ways in which we can compute different approximate 2×2D of a given problem, combining the previous result and SDP computations. Indeed, it is obvious that one can, for instance, solve SDP problems of the kind

$$\min \left\{ \, \| \, \Phi \, \| \ : \ Q = \Phi + \sum_{p \in P} E^p \Pi^p (E^p)^T \ , \ \Pi^p \succeq 0 \ p \in P \ , \ \Phi \succeq 0 \, \right\}$$

to minimize some proper norm of the "residual" $\Phi$, but more sophisticated (and costly) versions are possible where the constraints of the problem are also taken into account to provide the maximum possible lower bound, extending the results in [4] for the extraction of the diagonal. We experimentally analyse the impact of the different choices on the quality of the obtained bounds. The computational results show promise of the approach, but also highlight a number of hurdles that need be resolved before it can be widely applied. Also, extensions of the approach to the case of $k \times k$-Decomposition with $k > 2$ are discussed, highlighting the interesting open problem of developing conditions for exact $k \times k$-Decomposability of a given matrix $Q$.

REFERENCES

[1] A. Frangioni and C. Gentile, *Perspective Cuts for a Class of Convex 0–1 Mixed Integer Programs*, Mathematical Programming **106**(2) (2006), 225–236.
[2] A. Frangioni and C. Gentile, *SDP Diagonalizations and Perspective Cuts for a Class of Nonseparable MIQP*, Operations Research Letters **35**(2) (2007), 181–185.
[3] A. Frangioni and C. Gentile, *A Computational Comparison of Reformulations of the Perspective Relaxation: SOCP vs. Cutting Planes*, Operations Research Letters **37**(3) (2009), 206–210.
[4] X. Zheng, X. Sun and D. Li, *Improving the Performance of MIQP Solvers for Quadratic Pro- grams with Cardinality and Minimum Threshold Constraints: A Semidefinite Program Approach* INFORMS Journal on Computing **26**(4) (2014), 690–703.
[5] A. Frangioni, F. Furini and C. Gentile, *Approximated Perspective Relaxations: a Project&Lift Approach*, Computational Optimization and Applications **63**(3) (2016), 705–735.
[6] H. Jeon, J. Linderoth, A. Miller *Quadratic cone cutting surfaces for quadratic programs with on–off constraints*, Discrete Optimization **24** (2017), 32–50.
[7] A. Frangioni, C. Gentile and J. Hungerford, *Decompositions of Semidefinite Matrices and the Perspective Reformulation of Nonseparable Quadratic Programs*, Mathematics of Operations Research, to appear (2019).

# Mixed-Integer Partial Differential Equation Constrained Optimization

ARMIN FÜGENSCHUH

(joint work with F. Gnegel, M. Hagel, S. Leyffer, M. Stiemer)

An area of research that recently gained momentum is the inclusion of continuous physical phenomena described by ODEs or PDEs into discrete decision processes as an additional type of constraints. For instance, Fügenschuh et al. [4, 3] describe network flow problems in which the transport equations are included, Koch et al. [7] and Hahn et al. [6] combine gas dynamics with discrete decisions related to the control of natural gas networks, and Frank et al. [2] consider the shortest path problem together with the heat equation. We will refer to this class of problems as *Mixed-Integer PDE-Constrained Optimization (MIPDECO)* problems, a term introduce in [8]; this notion includes ODEs as a one-dimensional special case of PDEs.

Classical mixed-integer linear programs have a finite number of constraints, for instance, one for each node or each arc of a finite graph structure. In contrast, differential equations are defined on continuous domains, they have to be valid at infinitely many points. In the numerical treatment of PDEs as well as in the field of Optimal Control, it is common practice to find a finite system that approximates the solution with high accuracy. A possible way for solving MIPDECO problems is thus to adapt known approximation techniques for PDEs in order to obtain an approximation with only finitely many constraints and variables. A common approach is to use finite-difference methods, which replaces the differential operators with suitable difference equations on a finite and regular grid. The approximated function values on the grid are then included in the MIPDECO formulation as additional variables, and the difference equations are additional constraints. The

downside of this approach is that even moderately fine grids – by PDE-numerics standards – lead to a huge number of variables and constraints – by mixed-integer programming standards. Thus the solution of the linear programming relaxation (the root node of the evolving branch-and-bound-tree) of such an instance can take multiple hours or may not even finish on a state-of-the-art numerical solver (such as Cplex). Even in the case of linear systems (for instance, resulting from approximations of the heat equation, the wave equation, or the transport equation) numerical issues uncommon for MILPs derived for combinatorial problems have to be taken into account. For example, the standard bounds strengthening routines in the presolve phase of the aforementioned solvers may result in a false certificate of infeasibility [1].

In the case of linear PDEs, it is sometimes possible to transfer the solution of the PDE into a preprocessing step. That is, it can be efficient to solve the PDE for a basis of the control space and apply the principle of superposition to obtain solutions by linear combinations of the basis [2, 5]. In addition to that, splitting the solution process of the PDE and the MILP has the additional benefit that other solution methods, such as Galerkin methods (in particular, finite element methods) can be applied more readily.

We illustrate these ideas for the convection-diffusion equation, where an additional vector of variables $w = (w_1, w_2, \ldots, w_n) \in \mathbb{R}^n$ representing a control is included in the righthand side. In full, we consider the following PDE system:

$$
\text{(1a)} \quad u_t(x,t) - \vec{c} \cdot \nabla u(x,t) - d\Delta u(x,t) = y(x,t,w) \qquad \forall (x,t) \in \Omega \times (0,T),
$$

$$
\text{(1b)} \quad \frac{\partial}{\partial \vec{n}} u(x,t) = h_R(u_R - u(x,t)) \qquad \forall (x,t) \in \partial\Omega \times (0,T),
$$

$$
\text{(1c)} \quad u(x,0) = u_0(x) \qquad \forall x \in \Omega.
$$

Here $\Omega$ is a sufficiently regular domain in which the time depending temperature distribution $u$ is studied. Further given are: a constant outer temperature $u_R$ on the boundary $\partial\Omega$, a coefficient $h_R$ for the heat transfer over the boundary, a temperature distribution $u_0(x)$ at time 0, and further constants $\vec{c}, d$. The function $y$ on the righthand side expresses the influence of $w$ on the continuous dynamics. We assume that $y$ depends linearly on $w$. Then there exist functions $y_i$ for all $i \in \{1, \ldots, n\}$, such that

$$
y(x,t,w) = \sum_{i=1}^{n} w_i y_i(x,t).
$$

By the principle of superposition the solution of the PDE can then be expressed as

$$
\text{(2)} \qquad u(x,t) = u_{\text{inh}}(x,t) + \sum_{i=1}^{n} w_i \hat{u}_i(x,t),
$$

where $u_{\text{inh}}$ (inh for inhomogeneous) is the solution of (1a) for $w = (0,0,\ldots,0)$ (which implies $y = 0$), and each $\hat{u}_i$ is the solution of (1a) for $y = y_i, u_R = 0$, and $u_0(x) = 0$.

Now the approximations of the functions on the righthand side of (2) are calculated in a preprocessing step. Hereto, a finite element method can be applied, for instance, which has the additional benefit that the basis functions can be evaluated at arbitrary points in the domain without further approximation techniques (i.e., not only at grid points), since finite element solutions are discrete functions that are naturally defined throughout $\Omega$.

We then replace the system (1) by its approximative solution (2), and thus decouple the discretization of the PDE from the further solution process. Constraints defined for the whole domain, such as a state constraint

$$(3) \qquad\qquad u(x,t) \geq 0, \qquad\qquad \forall (x,t) \in \Omega \times (0,T)$$

can be evaluated during the branch-and-bound solution process (technically, as lazy constraints) at a finite number of positions $(x,t)$, and are added on demand in case one is violated. Since these constraints are dense in $w_i$ (i.e., having many nonzero entries), it is not clear at first if this approach is computationally advantageous, and thus comparative numerical studies were carried out. The numerical results we obtained for two case studies – a wildfire hazard controlled by firefighters operating on a road network, and a contaminated subsurface water flow controlled by filtration stations – show a large reduction in computation time. It turned out that only very few lazy constraints were in fact generated, so that their density did no great harm to the solution process.

REFERENCES

[1] A. Dittel, A. Fügenschuh, S. Göttlich, and M. Herty. MIP Presolve Techniques for a PDE-based Supply Chain Model. *Optimization Methods & Software*, 24(3):427–445, 2009.

[2] M. Frank, A. Fügenschuh, M. Herty, and L. Schewe. The Coolest Path Problem. *Networks and Heterogeneous Media*, 5(1):143–162, 2010.

[3] A. Fügenschuh, S. Göttlich, C. Kirchner, M. Herty, and A. Martin. Efficient Reformulation and Solution of a Nonlinear PDE-Controlled Flow Network Model. *Computing*, 85(3):245–265, 2009.

[4] A. Fügenschuh, M. Herty, A. Klar, and A. Martin. Combinatorial and Continuous Models and Optimization for Traffic Flow on Networks. *SIAM Journal on Optimization*, 16(4):1155–1176, 2006.

[5] F. Gnegel, M. Dudzinski, A. Fügenschuh, and M. Stiemer. Mixed Integer PDE Constrained Optimization for the Control of a Wildfire Hazard. In *Operations Research Proceedings 2017*, pages 483–488. Springer Verlag, Heidelberg, 2018.

[6] M. Hahn, S. Leyffer, and V. M. Zavala. Mixed-Integer PDE-Constrained Optimal Control of Gas Networks. Technical report, Argonne National Laboratory ANL/MCS-P7095-0817, 2017. Available at URL: https://wiki.mcs.anl.gov/leyffer/images/2/27/GasNetMIP.pdf (visited on 06/25/2018).

[7] T. Koch, B. Hiller, M. E. Pfetsch, and L. Schewe. *Evaluating gas network capacities*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2015.

[8] S. Leyffer. Mixed-Integer PDE-Constrained Optimization. In A. Wiegele L. Liberti, S. Sager, editor, *Mixed-integer Nonlinear Optimization: A Hatchery for Modern Mathematics*, volume 46/2015, pages 2738–2740. Mathematisches Forschungsinstitut Oberwolfach, 2015. DOI: 10.4171/OWR/2015/46.

## Graph Theory + Algebraic Model + SOS-Hierarchy + SDP = Fun

ELISABETH GAAR

(joint work with Daniel Krenn, Susan Margulies, Angelika Wiegele)

In 1968 Vizing conjectured that the product of the domination numbers of two graphs is always less or equal to the domination number of their Cartesian product graph. A lot of research has been done on Vizing's conjecture since then, including results that confirm the conjecture for some graph classes and results on properties of possible minimal counterexamples; see for example the survey [2]. Today, we still don't know whether Vizing's conjecture is true or not.

We discuss a completely new approach of tackling Vizing's conjecture, which was introduced in [1]. First we fix 4 parameters representing the number of vertices and the domination number of each of the two graphs. Then we define an ideal depending on these 4 parameters, such that the variety of this ideal represents dominating sets in the product graphs. Next we translate Vizing's conjecture for these parameters into the question of whether a specific polynomial $f^*$ is nonnegative on the variety of this specific ideal. To tackle nonnegativity, we do another reformulation to the question of whether $f^*$ is a sum-of-squares (SOS) polynomial on a certain level of the SOS-hierarchy and discuss how we can use semidefinite programming (SDP) to answer these kind of questions. If we have found an SOS-representation of $f^*$, we refer to it as certificate. Along the way we have to transform the numeric SDP solution into an exact (potentially irrational) certificate.

We obtained several sparse, low-degree certificates for different values of the parameters. So even though we still do not have reached parameter values covering previously unknown results about Vizing's conjecture, we provided a proof of concept of our new method to tackle Vizing's conjecture.

In [1] the limiting factor is the computation of the Gröbner basis, which is needed in order to construct the SDP. In the course of the Oberwolfach workshop Mixed-integer Nonlinear Optimization 2019 the following questions arose. Can we use symmetries in order to produce certificates easier? How can we omit (or partially omit) the computation of the Gröbner basis in order to obtain certificates for larger values of parameters?

References

[1] E. Gaar, D. Krenn, S. Margulies, A. Wiegele, *An Optimization-Based Sum-of-Squares Approach to Vizing's Conjecture*, Proceedings of the International Symposium on Symbolic and Algebraic Computation (ISSAC '19), July 15–18, 2019, Beijing, China.
[2] B. Brešar, P. Dorbec, W. Goddard, B. L. Hartnell, M. A. Henning, S. Klavžar, and D. F. Rall. 2012. Vizing's conjecture: a survey and recent results. J. Graph Theory 69, 1-2 (2012), 46–76.

# Reformulation Techniques for Mixed Integer Quadratic Programs

LAURA GALLI

(joint work with Adam N. Letchford)

Consider an arbitrary optimisation problem in which some or all of the variables are integer-constrained. If those variables are explicitly bounded, then one can reformulate the problem by replacing each such variable with a collection of binary variables. For example, if the variable $x_i$ is integer-constrained and we know that $0 \leq x_i \leq u_i$, where $u_i$ is a positive integer, we can replace each occurrence of $x_i$ with

$$\sum_{s=0}^{\lfloor \log_2 u_i \rfloor} 2^s \, \tilde{x}_{is},$$

where the $\tilde{x}_{is}$ are new binary variables [8], although several other ways are also possible.

It has been shown that *bit representation* can be useful for deriving strong cutting planes for mixed-integer linear programs. A few people have also applied bit representation to *nonlinear* problems. Gupte *et al.* [5] show that bit representation can be useful for solving certain mixed-integer bilinear programs. Billionnet *et al.* [2] show that, under certain conditions, the bit representation can be used to convert non-convex mixed-integer quadratic programs (MIQPs) into convex mixed 0-1 quadratic programs. In this paper, we continue this line of research and show that, in the context of non-convex MIQPs, bit representation can tighten linear programming (LP) relaxations of the problem. First, we consider three LP relaxations of MIQPs, based on reformulations due to McCormick [7] (LPM), Harjunkoski *et al.* [6] (LPH) and Glover & Woolsey [4] (LPG). A natural question at this point is: which of the three strategies is to be preferred? We provide a *partial* answer, by analysing the quality of the corresponding LP relaxations in a very simple "special" case. Let us say that an MIQP is *simple* if $n = |I| = 1$ and there are no linear constraints apart from the trivial bounds $0 \leq x_1 \leq u_1$. To compare the three LP relaxations, we project their feasible regions into the McCormick $(x, X)$-space. This allows us to prove that LPM is stronger than LPH, but there are no other dominance relations. Then, we show ways to strengthen each of the three relaxations using *split* inequalities [3], *cover* inequalities [5] and RLT [1]. We analyse three strengthened versions of the original LP relaxations (called LPM$^+$, LPH$^+$ and LPG$^+$, respectively) and again compare them via polyhedral projection. We can prove that LPH$^+$ dominates LPM$^+$, and conjecture that LPG$^+$ is at least

as strong as $\mathtt{LPH}^+$. This suggests that *disaggregating* integer variables induces a "strengthening hierarchy" of the LP relaxations. We also investigate the interplay between *semidefinite* programming relaxations (SDP) and bit representation. We start proving that bit representation can never make the SDP bound worse, in fact we provide an example where it is strictly better. Next, we explore this phenomenon in more detail by comparing the projections of the feasible regions into the $(x, X)$-space. The reason for the improvement in the bound appears to be the constraint $\text{diag}(X) = x$, which is valid for binary variables, but has no analog in the general-integer case. In other words, the SDP relaxation of the original MIQP does not exploit the integrality of $x$ in any way, while bit representation does. A key outcome of this study is that the best relaxations based on bit representation are provably stronger than the others. We also present some preliminary experiments to compare our formulations from a computational point of view.

## References

[1] W.P. Adams, H.D. Sherali, *A tight linearization and an algorithm for zero-one quadratic programming problems*, Management Science **32** (1986), 1274–1290.

[2] A. Billionnet, S. Elloumi, A. Lambert, *Extending the QCR method to general mixed-integer programs*, Mathematical Programming **131** (2012), 381–401.

[3] S. Burer, A.N. Letchford, *Unbounded convex sets for non-convex mixed-integer quadratic programming*, Mathematical Programming **143** (2014), 231–256.

[4] F. Glover, E. Woolsey, *Converting the 0-1 polynomial program to a 0-1 linear program*, Operations Research **22** (1974), 180–182.

[5] A. Gupte, S. Ahmed, M.S. Cheon, S. Dey, *Solving mixed integer bilinear problems using MILP formulations*, SIAM Journal on Optimization **23** (2013), 721–744.

[6] I. Harjunkoski, R. Pörn, T. Westerlund, H. Skrivars, *Different strategies for solving bilinear integer non-linear programming problems with convex transformations*, Computers & Chemical Engineering **21** (1997), 5487–5492.

[7] G.P. McCormick, *Computability of global solutions to factorable nonconvex programs. Part I: convex underestimating problems*, Mathematical Programming **10** (1976), 147–175.

[8] L.J. Watters, *Reduction of integer polynomial programming problems to zero-one linear programming problems*, Operations Research **15** (1967), 1171–1174.

## A fresh look at surrogate duality for mixed-integer nonlinear programming

Ambros Gleixner

(joint work with Benjamin Müller, Gonzalo Muñoz, Maxime Gasse, Andrea Lodi)

The concept of surrogate duality dates back to a paper by Glover from 1965 [3], who described how to form integer programming relaxations by aggregating constraints. In contrast to Lagrangian duality, the violation of a surrogate constraint is not penalized as part of the objective function, but controlled directly by enforcing the surrogate constraint as a hard inequality. As a result, surrogate relaxations can yield tighter dual bounds if suitable dual multipliers are identified.

While previous works mostly focus on the mixed-integer linear case, our motivation is the solution or the computation of strong dual bounds for challenging

nonconvex mixed-integer nonlinear programs. Formally, we consider an MINLP

$$(1) \qquad \min \left\{ c^t x : g_j(x) \leq 0 \ \forall j = 1, \ldots, m, x \in X \right\}$$

with factorable continuous functions $g_j : \mathbb{R}^n \to \mathbb{R}$ and a mixed-integer ground set $X \subseteq \mathbb{R}^n$ defined by linear constraints and integrality restrictions on some of the variables. The classical *surrogate relaxation* with multiplier vector $\lambda \in \mathbb{R}^m$, $\lambda \geq 0$, reads

$$(2) \qquad S(\lambda) := \min \left\{ c^t x : \sum_j \lambda_j g_j(x) \leq 0, x \in X \right\}.$$

The task of computing the $\lambda$ giving the tightest dual bound leads to the *surrogate dual* $\max_{\lambda \geq 0} S(\lambda)$. While our computational results indicate that bounds from surrogate relaxations can be strong, we also observed that they become weaker as the number of nonline constraints increase. Hence, we propose to study the generalization from one to multiple constraint aggregations,

$$(3) \qquad S^K(\lambda) := \min \left\{ c^t x : \sum_j \lambda_j^k g_j(x) \leq 0 \ \forall k = 1, \ldots, K, x \in X \right\}$$

for $K \geq 2$. This yields a natural hierarchy of increasingly difficult and tight relaxations.

Although the quasi-concavity of $S^1$ fails to hold for larger $K$, we show that the classical separation algorithms developed in [1, 2, 4] can be adapted in practice. Using generalized surrogate relaxations implemented on top of the spatial branch-and-bound solver SCIP [6], we obtain significantly improved dual bounds for several hard MINLPs contained in MINLPLib2 [5]. These results, however, require a variety of sophisticated enhancements to the base algorithm including the use of refined relaxations, early termination criteria for master and subproblems, support stabilization, trust region techniques, and symmetry handling in the space of dual multipliers.

## References

[1] Dyer, M.E.: Calculating surrogate constraints. Mathematical Programming **19**(1), 255–278 (1980)

[2] Banerjee, K.: Generalized lagrange multipliers in dynamic programming. Ph.D. thesis, University of California, Berkeley (1971)

[3] Glover, F.: A multiphase-dual algorithm for the zero-one integer programming problem. Operations Research **13**(6), 879–919 (1965)

[4] Karwan, M.H., Rardin, R.L.: Some relationships between lagrangian and surrogate duality in integer programming. Mathematical Programming **17**(1), 320–334 (1979)

[5] Vigerske, S.: MINLPLib. A library of Mixed-Integer and Continuous Nonlinear Programming Instances. `http://www.minlplib.org/`

[6] Vigerske, S., Gleixner, A.: SCIP: Global optimization of mixed-integer nonlinear programs in a branch-and-cut framework. Optimization Methods and Software pp. 1–31 (2017)

# Improved convergence analysis of Lasserre's measure–based upper bounds for polynomial minimization on compact sets

Monique Laurent

(joint work with Etienne de Klerk and Lucas Slot)

We consider the problem of computing the minimum value $f_{\min,K}$ of an $n$-variate polynomial $f$ over a compact set $K \subseteq \mathbb{R}^n$. This is a computationally hard problem, which permits to model a wide range of applications from combinatorial optimization, control, global optimization and other areas. This problem can be reformulated as finding a probability measure $\nu$ on $K$ minimizing $\int_K f d\nu$. Lasserre [9] showed that it suffices to consider such measures of the form $\nu = q\mu$, where $q$ is a sum-of-squares polynomial (abbreviated as sos) and $\mu$ is a given Borel measure supported on $K$. By bounding the degree of $q$ by $2r$ one gets a converging hierarchy of upper bounds $f^{(r)}$ for $f_{\min,K}$. For any fixed integer $r$, the parameter $f^{(r)}$ can be expressed as a semidefinite program of size polynomial in $n$ or, as recalled below, as an eigenvalue optimization problem. The objective is to analyze the rate of convergence of the sequence $\mathbf{f^{(r)}} = f^{(r)} - f_{\min}$ in terms of the degree bound $r$. Depending on the class of compact sets $K$ and the reference measure $\mu$, different strategies can be employed for the analysis.

The first strategy is based on the fact that we are searching for a sos polynomial $q$ that approximates well the Dirac measure at a global minimizer of $f$ in $K$. Such sos polynomials were constructed using truncations of the Guassian measure in [7] and of the Boltzman distribution in [3]; they led to a convergence rate in $O(1/\sqrt{r})$ for general compac sets satisfying a mild geometric condition (see [7]) and in $O(1/r)$ for general convex bodies (see [3]). Using sos polynomials constructed from the so-called needle polynomials from approximation theory (see [8]), improved convergence rates in $O(\log r/r)$ and $O(\log^2 r/r^2)$ have been shown in [11] for these two classes, respectively.

The second strategy is based on an eigenvalue reformulation for the parameter $f^{(r)}$, which can be verified to be equal to the smallest value of the matrix

$$A_f(r) = \left( \int_K f(x) p_\alpha(x) p_\beta(x) d\mu(x) \right)_{|\alpha|,|\beta| \leq r}.$$

Here, $\{p_\alpha(x) : \alpha \in \mathbb{N}^n, |\alpha| \leq r\}$ is an orthonormal basis of the set of polynomials with degree at most $r$, with respect to the scalar product induced by the measure $\mu$ on $K$. Using this strategy a convergence rate in $O(1/r^2)$ can be shown for several special sets $K$ and measures $\mu$. A basic tool used here for the analysis is that one may restrict to the case when $f$ is quadratic (or even linear in some cases), since one may always replace $f$ by a quadratic upper estimator obtained using, e.g., Taylor theorem.

Consider first the interval $K = [-1, 1]$ equipped with the Jacobi-type measure $d\mu(x) = (1 - x^2)^\lambda$ where $\lambda > -1$. The $\Theta(1/r^2)$ convergence rate is shown in [4] for the case when $f$ is linear and any $\lambda > -1$, and for arbitrary $f$ when $\lambda = -1/2$ (the

Chebyshev case). These results extend easily to the hypercube $[-1, 1]^n$ equipped with the corresponding product measures.

In addition, these results can be 'lifted' to other sets $K$, such as the hypercube with more general measures, the sphere, the ball and the simplex. Another useful basic tool to do this relies on using the notion of 'local similarity' introduced in [11]. Roughly speaking, assume that $K \subseteq \widehat{K}$ are two compact sets that 'look similar' in a neighborhood of a global minimizer $x^*$ of $f$ in $K$, and that $K$ (resp., $\widehat{K}$) is equipped with a measure $d\mu(x) = w(x)dx$ (resp., $d\widehat{\mu}(x) = \widehat{w}(x)dx$), where the two weight functions $w, \widehat{w}$ also 'look similar' in the neighbourhood of $x^*$. Then, as shown in [11], information about the convergence rate for $(\widehat{K}, \widehat{w})$ implies information about the convergence rate for $(K, w)$.

The above mentioned convergence rates for the interval $[-1, 1]$ can be used to show the following further results. First, the bounds $\mathbf{f}^{(\mathbf{r})}$ also have a convergence rate in $O(1/r^2)$ for the case when $K$ is the unit sphere [6] (where an additional integration trick is used); this improves an earlier result of [1] who proved a rate in $O(1/r)$ for homogeneous polynomials. Second, the rate $O(1/r^2)$ extends for the hypercube $[-1, 1]^n$ equipped with the measure $\prod_{i=1}^{n}(1-x_i^2)^\lambda dx_i$ for any $\lambda \geq -1/2$, and for the unit ball equipped with the measure $(1 - \|x\|^2)^\lambda dx$ with $\lambda \geq 0$ (see [11]). Finally, the convergence rate $O(1/r^2)$ also holds for the simplex and for a large class of convex bodies, equipped with the Lebesgue measure (see [11]). These convex bodies are those that admit at each boundary point an inner tangent ball and an outer tangent ball; equivalently, they can be characterized as the smooth convex bodies which have a $C^2$ boundary and are 2-strictly convex.

By exploiting links to cubature rules as shown in [10] it is shown in [6] that the convergence rate $O(1/r^2)$ is tight for the optimization of linear polynomials over the unit sphere. Determining what is the exact rate for arbitrary convex bodies remains open. As shown in [2], the known results about the convergence rate of the sequences $\mathbf{f}^{(\mathbf{r})}$ can be applied to the general problem of moments, however at the price of 'loosing a square root' (see also [5] for details). Minimizing a rational function can also be cast as an instance of the general problem of moments, but in this case the same convergence rate as for polynomials is preserved (see [6]).

## References

[1] A.C. Doherty and S. Wehner. *Convergence of SDP hierarchies for polynomial optimization on the hypersphere*, arXiv:1210.5048.

[2] E. de Klerk, D. Kuhn and K. Postek, *Distributionally robust optimization with polynomial densities: theory, models and algorithms*, arXiv:1805.03588.

[3] E. de Klerk and M. Laurent, *Comparison of Lasserre's measure-based bounds for polynomial optimization to bounds obtained by simulated annealing*, Mathematics of Operations Research **43** (2017), 1317–1325.

[4] E. de Klerk and M. Laurent, *Worst-case examples for Lasserre's measure–based hierarchy for polynomial optimization on the hypercube*, arXiv:1804.05524 (to appear in Mathematics of Operations Research).

[5] E. de Klerk and M. Laurent, *A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis*, arXiv:1811.05439.

[6] E. de Klerk and M. Laurent, *Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere*, arXiv:1904.08828.

[7] E. de Klerk, M. Laurent and Z. Sun. *Convergence analysis for Lasserre's measure-based hierarchy of upper bounds for polynomial optimization*, Mathematical Programming **162(1)** (2017), 363–392.

[8] A. Kroó and J.J. Swetits, *On density of interpolation points, a Kadec-type theorem, and Saff's principle of contamination in Lp-approximation*, Constructive Approximation **8(1)** (1992), 87–103.

[9] J.B. Lasserre, *A New Look at Nonnegativity on Closed Sets and Polynomial Optimization*, SIAM Journal on Optimization **21(3)** (2010), 864–885.

[10] A. Martinez, F. PIazzon, A. Sommirava, and M. Vianello, *Quadrature-based polynomial optimization*, Optimization Letters (2019), https://doi.org/10.1007/s11590-019-01416-x

[11] L. Slot and M. Laurent. *Improved convergence analysis of Lasserre's measure–based upper bounds for polynomial minimization on compact sets*, arXiv:1905.08142.

## Local search for sparse reflexive generalized inverses

Jon Lee

(joint work with Marcia Fampa, Luze Xu)

For a rectangular diagonal matrix $\Sigma$, we get the Moore-Penrose pseudoinverse $\Sigma^+$ by taking the reciprocal of each non-zero element on the diagonal, leaving the zeros in place, and then transposing the matrix. For arbitrary $A$, if $A = U\Sigma V^\top$ is the singular value decomposition of the real matrix $A$, then the *Moore-Penrose pseudoinverse* $A^+$ is defined as $A^+ := V\Sigma^+ U^\top$. There are very efficient algorithms to calculate the singular value decomposition and hence the Moore-Penrose pseudoinverse, but unfortunately it can be quite dense, even when $A$ is sparse.

A classical results is that the Moore-Penrose pseudoinverse is the unique $n \times m$ matrix $H$ satisfying

(P1) $\qquad\qquad AHA = A \quad$ (*generalized inverse*)

(P2) $\qquad\qquad HAH = H \quad$ (*reflexive*)

(P3) $\qquad\qquad (AH)^\top = AH \quad$ (*"ah-symmetric"*)

(P4) $\qquad\qquad (HA)^\top = HA \quad$ (*"ha-symmetric"*)

With the terminology indicated above, any ah-symmetric generalized inverse $H$ of $A$ solves the least-squares problem $\min\{\|b - Ax\|_2 \ : \ x \in \mathbb{R}^n\}$ via $x := Hb$. We are particularly interested in the situation where $r := \text{rank}(A) \ll n \ll m$, $A^+$ is rather dense, and we have many right-hand sides $b$ to process (via $x := Hb$). What we seek is a *sparse* ah-symmetric generalized inverse $H$ of $A$. In what follows, we concentrate of seeking to minimize the vector 1-norm of $H$, rather than the intractable minimization of the "0-norm" of $H$.

We are also concerned about the rank of $H$. Low rank for $H$ can be interpreted as a kind of linear-algebraic "explainability". In fact, any generalized inverse $H$ of $A$ has rank at least $r$, and the rank of $H$ is equal to $r$ iff $H$ is reflexive. Conveniently, the quadratic equation P2 becomes linear in the presence of P1 and P3, because

P1+P3 implies that $AH = AA^+$. So, we could solve $\min\{\|H\|_1 : P1 + P2 + P3\}$ by LP. What is not clear is whether the LP-solution $H$ would actually have any sparsity and any structure.

Rather than using LP, we employ a natural *column block construction* for producing an ah-symmetric *reflexive* generalized inverse. Moreover, our construction has a nice "explainable strucure". For $T$, an ordered subset of $r$ elements from $\{1, \ldots, n\}$, let $\hat{A} := A[:, T]$ be the $m \times r$ submatrix of $A$ formed by columns $T$. If $\mathrm{rank}(\hat{A}) = r$, let

$$\hat{H} := \hat{A}^+ = (\hat{A}^\top \hat{A})^{-1} \hat{A}^\top.$$

The $n \times m$ matrix $H$ with all rows equal to zero, except rows $T$, which are given by $\hat{H}$, is an ah-symmetric reflexive generalized inverse of $A$ having at most $rm$ nonzeros.

In the least-squares setting, we are simply regressing onto the columns indexed by $T$. So our column-block solution always has a kind of explainability in the least-squares setting. Next, we seek to find such a good $T$ for the column block construction.

**Theorem:** Let $A$ be an $m \times n$, rank-$r$ matrix, and let $S$ be *any* ordered subset of $r$ elements from $\{1, \ldots, m\}$ indexing linearly-independent rows of $A$. Choose $\epsilon > 0$, and let $\tilde{A} := A[S, T]$ be a $(1 + \epsilon)$-local maximizer for the absolute determinant on the set of $r \times r$ nonsingular submatrices of $A[S, :]$, which can be calculated efficiently. Then the $H$ built via column block construction over $\hat{A} := A[:, T]$, is an ah-symmetric reflexive generalized inverse (having at most $rm$ nonzeros), satisfying $\|H\|_1 \leq r(1 + \epsilon)\|H_{opt}^{ah}\|_1$, where $H_{opt}^{ah}$ is a 1-norm minimizing ah-symmetric generalized inverse of $A$.

In fact, we have a family of examples to demonstrate that our local-search algorithm may in fact terminate with $\|H\|_1$ approximately $r$ times the 1-norm of a 1-norm minimizing ah-symmetric reflexive generalized inverse of $A$.

Because the 1-norm of $H$ is now controlled via our local-search algorithm, we can say that the entries of $H$ are under control. So we have essentially everything that we want for a good ah-symmetric generalized inverse: (i) we can hope that $H$ is sparse in the same way that everyone does who uses 1-norm as a tractable alternative to "0-norm", (ii) the entries of $H$ are under control because we approximately minimize an actual norm, (iii) $H$ has some guaranteed sparsity (no more that $rm$ nonzeros), (iv) $H$ has lowest possible rank $= r$, (v) $H$ has explainable structure, having all nonzeros in just $r$ rows — and so leads to a least-squares solution $x := Hb$ with guaranteed sparsity (no more than $r$ nonzeros).

Finally, we wish to point out that a more-obvious local-search algorithm based on minimizing $\|\tilde{A}^{-1}\|$ cannot give *any* approximation guarantee at all. We are currently carrying out computational experiments to see how practical our local search is. Our preliminary findings are that we can essentially even take $\epsilon = 0$ and get very rapid finite convergence to an $H$ with 1-norm much better than the guarantee.

Most of the results above and many more are in [2], which builds on [1]. Computational methodology and experiments will appear in [3].

## REFERENCES

[1] Marcia Fampa, Jon Lee. On sparse reflexive generalized inverses. *Operations Research Letters*, 46:605-610, 2018.

[2] Luze Xu, Marcia Fampa, Jon Lee. Aspects of symmetry for sparse reflexive generalized inverses. arXiv:1903.05744, 2019.

[3] Marcia Fampa, Jon Lee, Gabriel Ponte, Luze Xu. Experimental analysis of local search for sparse reflexive generalized inverses, forthcoming.

## Mixed-Integer PDE-Constrained Optimization

SVEN LEYFFER

(joint work with Bart van Bloemen Waanders, Mirko Hahn, Todd Munson, Lars Ruthotto, Meenarli Sharma, Ryan Vogt)

We consider the solution of mixed-integer partial differential equation (PDE) constrained optimization (MIPDECO) problems. This is a difficult class of problems that combines the combinatorial complexity of integer variables with the computational challenges of PDE constraints. We introduce a trust-region algorithm for MIPDECO and show its effectiveness on two classes of problems motivated by practical applications:

(1) *Determination and location of a set of discrete sources from noisy measurements.* This model is loosely motivated by applications in groundwater flow, where we want to find the location of pollutants in the subsurface; see, for example, [7, 3].

(2) *Design of an electromagnetic cloak.* This model is a mixed-integer formulation of the topology optimization formulation for an electromagnetic cloak design; see, for example, [5].

Both models include a PDE that is defined over a two- or three-dimensional domain and discretized by using quadrilateral finite elements. The source inversion model involves a linear advection-diffusion PDE, while the cloak-design is modeled by using a 2D Helmholtz equation. In both cases, the integer variables are binary indicator variables that model the presence of the source and the presence of cloaking material. Both models can be expressed abstractly as

$$(1) \qquad \left\{ \begin{array}{ll} \min\limits_{u,w} & \mathcal{J}(u,w) \\ \text{s.t.} & \mathcal{C}(u,w) = 0 \\ & w \in \{0,1\}^p \end{array} \right\} \quad \Leftrightarrow \quad \left\{ \begin{array}{ll} \min\limits_{w} & \mathcal{J}(u(w),w) \\ \text{s.t.} & w \in \{0,1\}^p \end{array} \right\},$$

where $\mathcal{F}$ is the objective function, $\mathcal{C}$ represents the PDE and boundary conditions, $u$ are the state variables, and $w$ are the binary control variables. We assume that given $w$, we can uniquely solve the PDE to obtain $u(w)$, resulting in an equivalent reduced-space formulation on the right. The presence of mesh-dependent integer variables in these models makes the use of commercial branch-and-cut (see, e.g.,

[1]) prohibitively expensive (and impractical for 3D extensions of our models). Consequently, we develop a trust-region heuristic that is described next.

## 1. Trust-Region Method for MIPDECO

We present a new improvement heuristic for MIPDECO that is motivated by trust-region methods for nonlinear optimization; see, for example, [2]. Our method is also related to local-branching heuristics for MINLP [4, 6].

The key idea is to work with the reduced-space formulation in (1) and to define a trust-region subproblem around a current iterate, $w^{(k)}$, as

$$
(2) \quad \begin{cases} \min_{w} & \mathcal{J}^{(k)} + g^{(k)^T}(w - w^{(k)}) \\ \text{s.t.} & \|w - w^{(k)}\|_1 \le \Delta_k, \quad \text{and } w \in \{0,1\}^p \end{cases},
$$

where $g^{(k)} := \nabla_w \mathcal{J}(u(w^{(k)}), w^{(k)})$ is the reduced gradient, and $\Delta_k \in \mathbb{Z}_+$ is the $\ell_1$ trust-region radius. We note that because $w \in \{0,1\}^p$, the trust-region constraint can be written equivalently as a single affine constraint. Given this subproblem, we define a trust-region algorithm as follows.

---

Given $w^{(0)} \in \{0,1\}^p$, gradient, $g^{(0)}$, set $\Delta_0 = \bar{\Delta}$ and $k \leftarrow 0$;

**while** $\Delta_k > 0$ **do**

    Solve trust-region subproblem (2) for $\hat{w}$;

    Evaluate $\mathcal{J}(\hat{w}, u(\hat{w}))$ (PDE) & $\rho_k = \frac{\mathcal{J}(w^{(k)}) - \mathcal{J}(\hat{w})}{-g^{k^T}(\hat{w} - w^{(k)})} = \frac{\text{ActRed}}{\text{PredRed}}$;

    **if** $\rho_k > \bar{\rho}$ **then**

        accept step: $w^{(k+1)} = \hat{w}$, possibly increase $\Delta_k$;

    **else**

        reject step: $w^{(k+1)} = w^{(k)}$, reduce $\Delta_{k+1} = \lfloor \frac{\Delta_k}{2} \rfloor$;

    Set $k \leftarrow k + 1$;

---

The algorithm requires two PDE solves (forward and adjoint) per successful iteration, and we can solve the subproblem (2) efficiently by recasting it as a simple knapsack problem.

## 2. Computational Results

The results of the trust-region approach to solving the source inversion problem are shown in Figure 1, which shows the location of the original sources, the observations, $u$, with the measurement locations in red, and the final solution from the trust-region approach, which has an intersection-over-union (IoU) score of 82.3%.

We have also applied this method to the design of an electromagnetic scatterer. Figure 2 shows the solution of the continuous relaxation final trust-region problem and the corresponding wave difference, or objective functional, demonstrating the effectiveness of our approach.
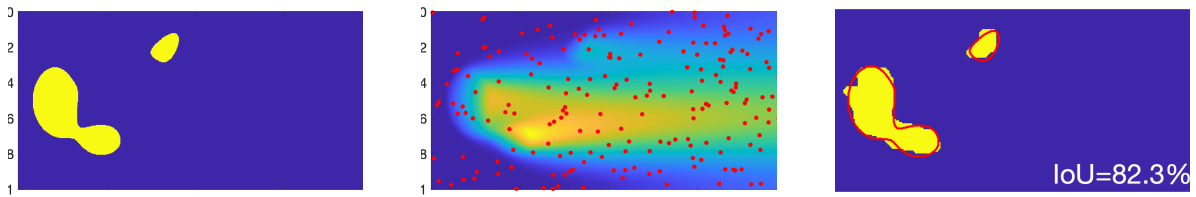
FIGURE 1. Original sources, the observations with measurement locations in red, and the final trust-region iterate.
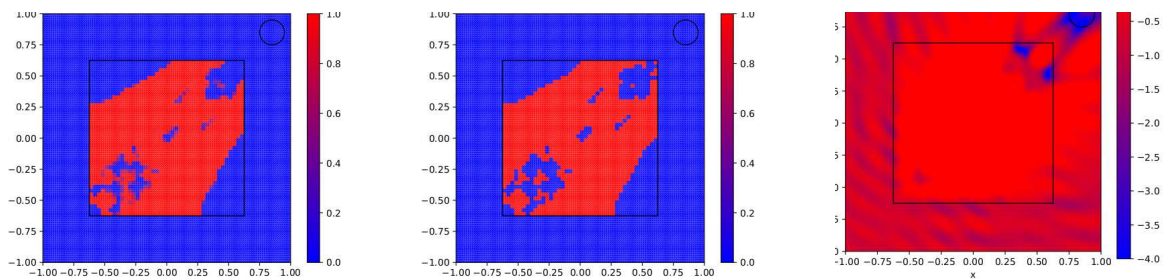


FIGURE 2. Solution of continuous relaxation final trust-region problem, and the logarithm of the wave difference.

## 3. CONCLUSIONS, OUTLOOK, AND OPEN PROBLEMS

We have presented a trust-region heuristic for solving MIPDECOs and have shown its effectiveness in solving realistic applications. Our approach leaves open a number of important questions and opportunities for future research.

(1) Currently, the algorithm stops when (2) cannot make any more progress. It would be interesting to see whether this stopping criterion can be replaced by a formal criterion based on the convergence of lower and upper bounds.

(2) Multigrid methods may provide an interesting refinement strategy for obtaining even better solutions.

(3) A formal convergence analysis of the algorithm based on topology optimization is an open problem, as well as the characterization of solutions under mesh refinement.

Despite these gaps in its theoretical justification, the proposed trust-region scheme performs well in practice, and we are working on other applications that could benefit from this approach.

## REFERENCES

[1] Belotti, P., Kirches, C., Leyffer, S., Linderoth, J., Luedtke, J., Mahajan, A.: Mixed-integer nonlinear optimization. Acta Numerica **22**, 1–131 (2013). DOI 10.1017/S0962492913000032. URL http://journals.cambridge.org/article\_S0962492913000032

[2] Conn, A.R., Gould, N.I., Toint, P.L.: Trust Region Methods. SIAM (2000)

[3] Fipki, S., Celi, A.: The use of multilateral well designs for improved recovery in heavy oil reservoirs. In: IADV/SPE Conference and Exhibition. SPE, Orlanda, Florida (2008)

[4] Fischetti, M., Lodi, A.: Local branching. Mathematical Programming **98**, 23–47 (2002)

[5] Haslinger, J., Mäkinen, R.A.: On a topology optimization problem governed by two-dimensional helmholtz equation. Computational Optimization and Applications **62**(2), 517–544 (2015)

[6] Nannicini, G., Belotti, P., Liberti, L.: A local branching heuristic for MINLPs. arXiv:0812.2188v1 [math.CO] (2008). `http://arxiv.org/abs/0812.2188`

[7] Ozdogan, U.: Optimization of well placement under time-dependent uncertainty. Master's thesis, Stanford University (2004)

# Robust Optimization: (Some) Theory, (Some) Algorithms, and (Some) Applications

### Frauke Liers

In the last two decades, protecting a mathematical optimization problem against uncertainties using robust optimization techniques has become a powerful tool. The research area of robust optimization has whitnessed fast algorithmic advances, novel theoretical insights, and the development of efficient solution approaches. However, still many challenges and settings exist where not yet efficient solution approaches are known.

Usually, the uncertainties are given to an optimization model via so-called uncertainty sets. A solution is robust feasible if it is feasible regardless how the uncertainty manifests itself within the set. Among the robust feasible solutions, a robust optimum one needs to be determined that yields the best guaranteed objective value. Due to the necessity that a robust solution needs to be feasible for all realizations of the uncertainty, semi-infinite models result for which no general solution algorithm exists.

In this overview talk, we have reviewed established as well as some recent developments in the field of (mixed-integer) linear, combinatorial, single and two-stage, convex, nonlinear, and PDE-constrained robust mathematical optimization. Although these areas are very wide, covering these topics is possible due to the fact that in particular three main approaches are often used in robust optimization, namely:

(1) Reformulation of the semi-infinite robust optimization problem so that the robust counterpart is a finite-dimensional and algorithmically tractable problem that can be solved to global optimality. Usually, reformulations use some form of duality theory such as linear, conic, or Fenchel duality.

(2) Decomposition of the robust problem into a master- and subproblem that are solved iteratively, similar as in an outer approximation or in a Benders decomposition approach.

(3) If the above is not easily possible, as for example in optimization with partial differential equations, often the robust problems are approximated, for example using Taylor expansion. Solution algorithms are then developed for the approximate robust counterparts.

In the talk, we have reviewed where and how these methodologies can be used successfully. We mentioned pointers to the literature. We closed by mentioning some theoretical and algorithmical challenges in the field, such as the development

of efficient global algorithms for multi-stage robust problems, in particular with non-convex stages that might involve integral decisions. Applications for such tasks appear in applications such as energy and logistics.

The field of distributional robustness can be understood as a combination of robust and stochastic optimization, with robust and stochastic optimization as special cases. This area currently attracts many researchers and has many relevant applications. The development of efficient solution approaches would be highly desirable.

## An exact algorithm for a class of mixed-integer programs with equilibrium constraints

ANDREA LODI

(joint work with Teodora Dan and Patrice Marcotte)

In this study [1], we consider a rich class of mathematical programs with equilibrium constraints (MPECs) involving both integer and continuous variables. Such a class, which subsumes mathematical programs with complementarity constraints, as well as bilevel programs involving lower level convex programs is, in general, extremely hard to solve due to complementarity constraints and integrality requirements. For its solution, we design an (exact) branch-and-bound (B&B) algorithm that treats each node of the B&B tree as a separate optimization problem and potentially changes its formulation and solution approach by designing, for example, a separate B&B tree. We refer to this algorithm as tree-of-trees B&B. The algorithm is implemented and computationally evaluated on a specific instance of MPEC, namely a competitive facility location problem that explicitly takes into account the queueing process that determines the equilibrium assignment of users to open facilities, and for which to date, no exact method has been proposed.

### REFERENCES

[1] T. Dan, A. Lodi, P. Marcotte, *An Exact Algorithm for a Class of Mixed-Integer Programs with Equilibrium Constraints*, Technical Report DS4DM-2018-010, CERC Data Science for Real-time Decision Making, 2018.

# Approximation properties of Sum-Up Rounding and consequences for Mixed-Integer PDE-Constrained Optimization

Paul Manns

(joint work with F. Bestehorn, C. Hansknecht, C. Kirches, F. Lenders)

For $n \in \mathbb{N}$, we abbreviate $[n] = \{1, \ldots, n\}$. We consider convexified mixed-integer optimal control problems, see [5], of the form

$$
\text{(BC)} \quad
\begin{aligned}
\inf_{y,u,\omega} \quad & J(y,u) \\
\text{s.t.} \quad & Ay = \sum_{i=1}^{M} \omega_i f_i(y,u), \\
& 0 \le \omega_i(s) c_i(y(s), u(s)) && \text{for a.a. } s \in \Omega_T \text{ and } i \in [M], \\
& \omega(s) \in \{0,1\}^M \text{ and } \sum_{i=1}^{M} \omega_i(s) = 1 \text{ for a.a. } s \in \Omega_T,
\end{aligned}
$$

in which $Ay = \sum_{i=1}^{M} \omega_i f_i(y,u)$ is the state equation of the underlying process, which is defined on a bounded domain or space-time cylinder $\Omega_T$, $A$ is a suitable differential operator and the $f_i$ are good-natured non-linearities. The $c_i$ are point-wise a.e. defined constraint functions. The function $\omega : \Omega_T \to \{0,1\}^M$ acts as an activation of the functions $f_1, \ldots, f_M$, i.e. $\omega_i(s) = 1$ for exactly one $i \in [M]$ and $\omega_j(s) = 0$ for $j \neq i$ for a.a. $s \in \Omega_T$. The fractional relaxation of (BC) is given by

$$
\text{(RC)} \quad
\begin{aligned}
\min_{y,u,\alpha} \quad & J(y,u) \\
\text{s.t.} \quad & Ay = \sum_{i=1}^{M} \alpha_i f_i(y,u), \\
& 0 \le \alpha_i(s) c_i(y(s), u(s)) && \text{for a.a. } s \in \Omega_T \text{ and } i \in [M], \\
& \alpha(s) \in [0,1]^M \text{ and } \sum_{i=1}^{M} \alpha_i(s) = 1 \text{ for a.a. } s \in \Omega_T.
\end{aligned}
$$

## Approximation arguments

Let (RC) be well-posed with solution $y, u, \alpha$. Let $\omega^h : \Omega_T \to \{0,1\}^M$ be a piecewise constant function satisfying $\sum_{i=1}^{M} \omega_i(s) = 1$ for a.a. $s \in \Omega_T$ that is computed from $\alpha$. The cell volume of the pieces on which $\omega^h$ is defined is denoted by $h$ and the algorithms that produce these $\omega^h$ are called Sum-Up Rounding (SUR) algorithms. Let $y^h$ solve the state equation for $u$ and $\omega^h$. We analyze the chain of arguments

$$
\text{(1)} \qquad h \to 0 \underset{(1)}{\implies} d(\omega^h, \alpha) \to 0 \underset{(2)}{\implies} \omega^h \rightharpoonup^* \alpha \underset{(3)}{\implies} y^h \to y,
$$

in which $d$ denotes a suitable pseudo-metric and $\rightharpoonup^*$ convergence in the weak-$^*$ topology of $L^\infty(\Omega_T, \mathbb{R}^M)$. Let the functions $J$ and $c_i$ be continuous in the first argument. Then, the approximation arguments (1) to (3) yield the optimality principle

$$
\min_{\text{(RC)}} J(y,u) = \lim_{h \to 0} J(y^h, u).
$$

Furthermore, if $c_i \equiv 0$ holds for all $i$, we have

$$
\min_{\text{(RC)}} J(y,u) = \inf_{\text{(BC)}} J(y,u).
$$

This gives well-definedness and finite termination for successive refinements of the grid, on which $\omega^h$ is computed by means of the SUR algorithm, up to a desired

tolerance with respect to the objective. We detail the approximation properties (1), (2), (3) below.

### Roundings are of vanishing integrality gap

Let $(\mathcal{T}_k)_{k\in\{1,\dots,N\}}$ be a finite partition of $\Omega_T$. For a given $\alpha$, which is feasible for (RC), we define the SUR algorithm in the presence of mixed constraints iteratively for $k = 1, \dots, N$ by

$$\omega(s) := \sum_{k=1}^{N} \chi_{\mathcal{T}_k}(s)W_k,$$

(SUR-VC) $\quad W_k(i) := \begin{cases} 1 & \text{if } i = \underset{j\in F_k}{\operatorname{argmax}} \int_{\bigcup_{\ell=1}^{k}\mathcal{T}_\ell} \alpha_j - \int_{\bigcup_{\ell=1}^{k-1}\mathcal{T}_\ell} \omega_j \\ 0 & \text{else,} \end{cases} \quad \text{for } i \in [M],$

$$F_k := \left\{ i \in [M] : \int_{\mathcal{T}_k} \alpha_i > 0 \right\}.$$

The resulting binary control function is also called a *rounding* of $\alpha$. The set $F_k$ contains the entries of $\omega$ that are admissible to be set to one in grid cell $k$. These are exactly the ones with a strictly positive mean of $\alpha$ over grid cell $k$. This restriction guarantees $0 \leq \liminf \omega_i^h c_i(y^h, u)$ for $0 \leq \alpha_i c(y, u)$. In [4], we have shown the approximation property below, which implies (1).

**Theorem 1.** *There exists $C > 0$ such that for all feasible $\alpha$ of* (RC) *and $\omega^h$ being computed with* (SUR-VC)*, the function $\omega^h$ satisfies the pointwise SOS1 constraint in* (BC) *and the estimate*

$$d(\omega^h, \alpha) := \max_{k\in[M]} \left\| \int_{\bigcup_{k'=1}^{k}\mathcal{T}_{k'}} \alpha(s) - \omega^h(s)\,\mathrm{d}\,s \right\|_\infty \leq Ch.$$

In the literature, the function $d$ is called *integrality gap* and the convergence $d(\omega^h, \alpha) \to 0$ is known as *vanishing integrality gap*.

### A vanishing integrality gap induces weak-* convergence

The following condition from [2] on rounding grid refinements induces (2).

**Definition 2.** *The sequence $\left(\{\mathcal{T}_1^n, \dots, \mathcal{T}_{N^{(n)}}^n\}\right)_n \subset 2^{\mathcal{B}(\Omega)}$ is called an **admissible sequence of refined rounding grids of** $\Omega_T$ if*

    (1) $\{\mathcal{T}_1^n, \dots, \mathcal{T}_{N^{(n)}}^n\}$ *is a finite partition of $\Omega_T$ for all $n \in \mathbb{N}$,*

    (2) $\max_{i\in\{1,\dots,N^{(n)}\}} \lambda(\mathcal{T}_i^n) \to 0$,

    (3) *For all $i \in \{1, \dots, N^{(n+1)}\}$, there exists $j \in \{1, \dots, N^{(n)}\}$ such that $\mathcal{T}_i^{n+1} \subset \mathcal{T}_j^n$,*

    (4) *the cells **shrink regularly**, i.e. there exists a constant $C > 0$ such that for each $\mathcal{T}_j^n$ there exists a ball $B_j^n$ such that $\mathcal{T}_j^n \subset B_j^n$ and $\lambda(\mathcal{T}_j^n) \geq C\lambda(B_j^n)$.*

**Theorem 3.** *Let $\alpha$ be feasible for* (RC) *and $(\omega^n)_n$ be computed from $\alpha$ with* (SUR-VC) *on an admissible sequence of rounding grids. Then, $\omega^n \rightharpoonup^* \alpha$ in $L^\infty(\Omega_T, \mathbb{R}^M)$.*

## State vector convergence

The property of the solution operator of the state equation to establish (3) is complete continuity, which is stated below.

**Definition 4.** *Let $X$, $Y$ be Banach spaces. Then, a mapping $T : X \to Y$ is called **completely continuous** if, for every weakly convergent sequence $x^n \rightharpoonup x$, we have $Tx^n \to Tx$ in norm.*

We refer to [3] for elliptic equations and to [2] for a class of semilinear evolution equations constraining (BC) and (RC). We note that other completely continuous operators exist, which enable us to employ the derived approximation properties outside the context of differential equations. An example are convolutions with a fixed kernels that can be used to model filtering in signal processing problems.

## Switching-cost aware rounding

(SUR-VC) often produces controls that exhibit high-frequency switching, which may be undesirable. Thus, one can penalize the number of switches in the objective and obtain a trade-off between approximation quality and switching costs. However, these costs cannot be bounded for $h \to 0$ as the weak-$*$ approximation of fractional-valued functions may require arbitrarily high switching in the binary-valued approximants for $h \to 0$, see [1].

## References

[1] F. Bestehorn, C. Hansknecht, C. Kirches, and P. Manns. A switching cost aware rounding method for relaxations of mixed-integer optimal control problems. *submitted*, 2019.
[2] P. Manns and C. Kirches. Multi-dimensional sum-up rounding for elliptic control systems. *submitted*, 2018.
[3] P. Manns and C. Kirches. Improved regularity assumptions for partial outer convexification of mixed-integer PDE-constrained optimization problems. *ESAIM: Control, Optimisation and Calculus of Variations*, 2019. to appear.
[4] P. Manns, C. Kirches, and F. Lenders. A linear bound on the integrality gap for Sum-Up Rounding in the presence of vanishing constraints. *submitted*, 2017.
[5] S. Sager, H.G. Bock, and M. Diehl. The Integer Approximation Error in Mixed-Integer Optimal Control. *Mathematical Programming, Series A*, 133(1–2):1–23, 2012.

## Modeling and optimization of traffic at traffic-light controlled intersections

Maximilian Merkert

(joint work with Gennadiy Averkov, Do Duc Le, Sebastian Sager, Stephan Sorgatz)

Assisted and autonomous driving is a growing field of interest. The coordination of vehicular traffic at traffic-light controlled intersections offers great potential for optimization: Intersections in urban areas represent intrinsic bottlenecks for the movement of cars, potentially causing long waiting times, traffic jams, air pollution and increased energy consumption. Imperfect human driving behavior

and inefficient traffic light switches further amplify congestion. Consequently, controlling traffic light signals plays a vital role in effective traffic management.

**Traffic Model.** We use a MINLP model to describe the traffic flow on a simple urban road network within a fixed time interval at a microscopic level. It is based on the model presented in [1]. The movement of vehicles is modeled by a time-discretized ODE system while the states of the traffic lights are represented via binary variables. Our goal is to find a globally optimal traffic flow, i.e. to determine controls for the movement of each car as well as for each traffic light that minimize a global objective—such as the total transit time or energy consumption of all cars—while avoiding collisions. Collision prevention on lanes is ensured by constraints that require each car to keep a predefined safety distance to its predecessor. However, successor-predecessor relations may change after turning, which necessitates the introduction of further binary indicator variables. Collisions on the intersection are prevented due to traffic lights. A trigger mechanism models the logic of traffic lights, which basically prohibits cars to be in the intersection area at certain time steps.

**Traffic light regulations.** Traffic light controls have to fulfill certain requirements in order to be reasonable or even legal such as minimum and maximum green and red phases or minimum and maximum cycle times of individual traffic lights. Other regulations may affect several traffic lights at a given intersection simultaneously. Incorporating these rules substantially increases the computational difficulty of the overall problem formulation. We demonstrate that many natural requirements can be implemented by finite automata—enabling flow-based extended formulations of the corresponding 0-1 polytope [2]. As an important example, the requirement of minimum green and red phases leads to min-up/min-down polytopes, which have exponentially many facets in the original space [3]. Flow-based extended formulations allow us to recover the *turn on/off* inequalities that lead to a complete description of min-up/min-down polytopes in extended space known from the context of unit commitment problems [4] by projecting out certain flow variables. However, given the size of the extended formulations for certain traffic light regulations, it still has to be investigated how to exploit them most efficiently in practice.

**Effects on traffic flow.** The resulting models are very challenging to solve to global optimality, in particular when involving realistic traffic light regulations. Our approach yields a solution from a global point of view where everything is coordinated centrally. Though unlikely to become reality in the very near future, solutions obtained from our model serve as a benchmark and basis for decentralized concepts. Using our model, we computationally investigate different scenarios with respect to traffic density, traffic light regulations, and equipment rate (percentage of coordinated vehicles). We compare the resulting traffic with the aid of the microscopic traffic simulation tool SUMO (*Simulation of Urban MObility*) [5]. Experiments suggest that waiting times can be significantly reduced. Moreover, optimized solutions show interesting driving behavior that might improve

efficiency of real-world urban traffic already in the case where communication is just possible from infrastructure to cars. For instance, cars slow down or stop at some distance from a traffic light (instead of right in front of it) and plan ahead to be at the maximum allowed speed when reaching a green light. This leads to cars spending less time in the crossing area.

## References

[1] S. Sorgatz, *Optimization of Vehicular Traffic at Traffic-Light Controlled Intersections*, Ph. D. Dissertation, Otto-von-Guericke-Universität Magdeburg (2016).
[2] S. Fiorini, K. Pashkovich, *Uncapacitated flow-based extended formulations*, Mathematical Programming **153** (2015), 117–131.
[3] J. Lee, J. Leung, F. Margot, *Min-up/min-down polytopes*, Discrete Optimization **117** (2004), 77–85.
[4] D. Rajan, S. Takriti, *Minimum Up/Down Polytopes of the Unit Commitment Problem With Start-Up Costs*, IBM Research (2005).
[5] P. Alvarez Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, E. Wießner, *Microscopic Traffic Simulation using SUMO*, IEEE Intelligent Transportation Systems Conference (2018).

## Developing spatial branch & bound solvers

Ruth Misener

(joint work with Radu Baltean-Lugojan, Francesco Ceccon, Miten Mistry)

**Summary**

This survey presentation, which was meant as a broad overview, discussed the ingredients common to branch and bound solvers. We also presented recent advances in developing different solver components. Finally, the presentation offered a to-do list for developing next generation solvers. We specifically focused on the theoretical and algorithmic contributions needed to effectively solve mixed-integer nonlinear optimization problems (MINLP).

Because this survey talk was meant for a fairly broad audience, the slides are available on Twitter[1]. Twitter interaction with the slides includes a discussion of whether the convexity structure of a special function in energy efficiency [7] can be represented using cones[2].

**Definitions & solvers**

The first section of the presentation defined MINLP and the tools for representing and solving MINLP. An example showcased several challenges arising in modern MINLP solution methods:

- The difficulty of developing MINLP relaxations, i.e. that most state-of-the-art relaxations are based on box variable bounds.
- The impact of special structure detection in building a relaxation, i.e. how succeeding or failing at structure detection influences relaxation strength.

---

[1] https://twitter.com/RuthMisener/status/1135421089683836928
[2] https://twitter.com/JeffLinderoth/status/1135517902688440320

- The challenges associated with branching, i.e. a branch may not necessarily reject the current feasible solution.

This discussion motivated some challenges where we returned at the conclusion.

## Data structures

This section discussed the data structures that are common to all exact MINLP solvers. The purpose was to motivate discussion with respect to (i) what data structures we should be incorporating into modern solvers and (ii) how we can make modern mathematical advances fit into the constraints of the best-known data structures.

In particular, the presentation detailed: (i) directed acyclic graphs for representing optimization model equations, (ii) hard-coded expression types for storing special structure knowledge related to each type of important expression, and (iii) undirected [weighted] graphs for specifically addressing quadratic objectives and constraints. There are two paradigms for managing special structure: automatic detection and disciplined convex programming [5]. The presentation considered the trade-offs between these two methods, especially with respect to the extensible convexity detector SUSPECT [3] and an array of applications common in MINLP.

## Branch & bound components

This section presented components common to every branch and bound solver: (i) relaxations, (ii) branching, (iii) bounds tightening, (iv) primal heuristics, (v) and cutting planes. We particularly focussed on the challenges and opportunities associated with relaxations and branching:

- **Relaxations** for special functions is a long-time research area in MINLP. Within the bounds of practical development, the best MINLP strategy is heterogenous, i.e. uses multiple types of relaxations simultaneously [1, 2]. A major challenge for the future is relaxations that are *not* based on boxes. There are a few references developing relaxations beyond boxes[3], but we highlighted the need for more development along this line.
- **Branching** is an especially difficult area in solving MINLP because a branch may not necessarily eliminate the current feasible point [6]. There is some interesting new work with respect to branching to get disjunctions [4], but a lot more to do in this area.

## Challenges

We closed the presentation by offering a number of open challenges and invited the audience to add to the list.

### References

[1] K. M. Anstreicher. Semidefinite programming versus the reformulation-linearization technique for nonconvex quadratically constrained quadratic programming. *J Glob Optim*, 43 (2):471–484, 2009.

---

[3]Slide 34, `https://www.doc.ic.ac.uk/~rmisener/MINLPSolver_Tutorial.pdf`

[2] A. Bompadre and A. Mitsos. Convergence rate of McCormick relaxations. *J Glob Optim*, 52 (1):1–28, 2012.

[3] F. Ceccon, J. D. Siirola, and R. Misener. SUSPECT: MINLP special structure detector for Pyomo. *Optim Letters*, 2019. DOI 10.1007/s11590-019-01396-y.

[4] S. S. Dey, A. Santana, and Y. Wang. New SOCP relaxation and branching rule for bipartite bilinear programs. *Optim Eng*, 20(2):307–336, 2019.

[5] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

[6] S. Lee and I. E. Grossmann. New algorithms for nonlinear generalized disjunctive programming. *Comput Chem Eng*, 24(9-10):2125–2141, 2000.

[7] M. Mistry and R. Misener. Optimising heat exchanger network synthesis using convexity properties of the logarithmic mean temperature difference. *Comput Chem Eng*, 94:1 – 17, 2016.

# A snapshot of quantum computing algorithms for optimization

GIACOMO NANNICINI

A quantum computer (QC) with $n$ qubits is a device that computes a vector $x \in \mathbb{C}^{2^n}, \|x\| = 1, x = \sum_{z \in \{0,1\}^n} \alpha_z e_z$. Here and in the following, we use $e$ to denote standard basis vectors and we index them with binary strings in $\{0,1\}^n$, since there are $2^n$ such vectors. The QC is initialized with the state $e_{\vec{0}}$, and we obtain the final state $x$ applying a unitary $U = \mathbb{C}^{2^n \times 2^n}$. The unitary must be obtained combining a polynomial number (in $n$) of basic operations; the basic operations are certain $2 \times 2$ and $4 \times 4$ matrices, combined via tensor products and matrix multiplication.

The only operation to extract information out of a QC is a measurement. The outcome of a measurement on the quantum state $x$ is a sample from a random variable $X$, with sample space $\Omega = \{0,1\}^n$ and $\Pr(X = z) = |\alpha_z|^2$.

Given this paradigm of computation, we are interested in optimization algorithms and their interplay with mixed-integer nonlinear programming techniques. We will divide our discussion into algorithms for combinatorial optimization, and for continuous optimization.

## 1. COMBINATORIAL OPTIMIZATION

In combinatorial optimization we want to solve:

$$\text{(OPT)} \qquad \min_{z \in \{0,1\}^n} f(z) = c^\top z + z^\top Q z.$$

This can be formulated as an eigenvalue problem. Define $H := \text{diag}(f(\vec{0}), \ldots, f(\vec{1}))$, where the matrix $H$ is typically called a Hamiltonian and encodes the energy of a quantum system; in this case, we can simply think of the Hamiltonian as encoding

the possible objective function values. Then:

$$\min_{x:\|x\|=1} x^* H x = \min_{x:\|x\|=1} \sum_{z\in\{0,1\}^n} f(z)|\alpha_z|^2 = \min_{x:\|x\|=1} \sum_{z\in\{0,1\}^n} f(z)\Pr(X=z)$$

$$= \min_{x:\|x\|=1} \mathbb{E}[f(X)] = (\text{OPT}).$$

Stated in this form, this is equivalent to finding the minimum eigenvalue of $H$. However, this problem has dimension $2^n$, therefore it may be difficult to optimize.

1.1. **VQE.** To reduce the size of the eigenvalue formulation, [Peruzzo et al., 2014] have proposed the Variational Quantum Eigensolver (VQE). It works as follows. Define a parametrized unitary $U(\theta)$, where $\theta$ is a small vector of parameters and $x = U(\theta)e_{\vec{0}}$. Then solve:

$$\min_{\theta} e_{\vec{0}}^{\top} U(\theta)^* H U(\theta) e_{\vec{0}}.$$

This is now a small-dimensional problem, but it is nonconvex and we do not have any guarantee on the approximation unless we study $U(\theta)$. If we can at least compute $e_{\vec{0}}^{\top} U(\theta)^* H U(\theta) e_{\vec{0}}$, and hopefully its derivatives, then one could apply a classical optimization algorithm. It is easy to compute this quantity with a QC. After we have chosen $\theta$, we prepare the state $x = U(\theta)e_{\vec{0}}$, then estimate the expectation $\mathbb{E}[f(X)]$. One can in principle even get partial derivatives of $\mathbb{E}[f(X)]$ with respect to $\theta$, knowing the form of $U(\theta)$. In fact, since we know that the optimum of (OPT) is a basis vector, it is not difficult to construct a simple $U(\theta)$ that is guaranteed to span all basis vectors. However, if we do so we are essentially solving a discrete problem by parametrizing it as a continuous problem, which is generally a terrible approach in practice.

1.2. **QAOA.** The Quantum Approximate Optimization Algorithm (QAOA) is a form of VQE with much more rigorous theoretical guarantees. QAOA prescribes the form of $U(\theta)$. More specifically, we choose a number of iterations $p$, and we use the unitary:

$$U(\beta,\gamma,p) = \prod_{i=1}^{p} \left[U_B(\beta_i)U_C(\gamma_i)\right] W,$$

where $W$ is the Walsh-Hadamard transform (an operation that can be implemented with just $n$ basic quantum gates). Each block $U_B(\beta_i), U_C(\gamma_i)$ is defined as a matrix exponential, parametrized by a single parameter $\beta_i, \gamma_i \in [0, 2\pi]$, and easy to implement with basic quantum operations.

**Theorem 1.1** ([Farhi et al., 2014]). *We have* $\lim_{p\to\infty} \min_{\beta,\gamma} e_{\vec{0}}^{\top} W U(\beta,\gamma,p)^* H U(\beta,\gamma,p)We_{\vec{0}} = (OPT)$. *Furthermore, for MAXCUT on 3-regular graphs,*

$$\max_{\beta,\gamma} e_{\vec{0}}^{\top} W U(\beta,\gamma,1)^* H U(\beta,\gamma,1)We_{\vec{0}} \geq 0.6924(OPT).$$

The advantages of QAOA over VQE are that here are theoretical guarantees that with a large enough $p$, the circuit will span the optimum. However, practical performance is typically poor in numerical tests.

**Proposition 1.2** (N. *et al.*). *Let $x = \sum_z \alpha_{p,z} e_z$ be the quantum state constructed by $p$ iterations of QAOA. Then $|\alpha_{p,z}| \leq (2^{n+1}(2 - \Delta_{p-1} - \delta) + 1)^p \frac{1}{\sqrt{2^n}}$ for all $z \in \{0,1\}^n$, where $\delta$ is the largest fraction of $z \in \{0,1\}^n$ that have the same value, and $\Delta_p$ is a parameter that measures how different the amplitudes are after iteration $p$ (with $\Delta_0 = 1$).*

This implies that for certain problems, i.e., feasibility problems with few feasible solutions, the state vector is "flat" and therefore sampling from it is not much better than random guessing.

## 2. Continuous optimization

There has been significant progress in continuous optimization in recent years. One reason for that is the improved solution of linear systems. This is what we give an overview of next.

2.1. **Solution of linear systems.** Suppose we want to solve $Ay = b$, with $A \in \mathbb{R}^{N \times N}$ and $N = 2^n$. Then there exists a quantum algorithm that solves it in time $\text{poly}(n, \log 1/\epsilon)$, where $\epsilon$ is a precision parameter [Childs et al., 2017]. We give a description of the first algorithm of this type [Harrow et al., 2009], which is easier to understand although it achieves worse running time. The description here is quite inaccurate and just meant to convey intuition.

First, we note that there is an algorithm (Quantum Phase Estimation) that given unitary $U \in \mathbb{C}^{2^n \times 2^n}$ and an eigenvector $v$ such that $Uv = \exp(2\pi i\theta)v$, outputs $\theta$ in an additional register with precision $\epsilon$, and requires running time $\text{poly}(n, \log 1/\epsilon)$. This dates back to [Shor, 1997].

Assume $A$ Hermitian and let $v^1, \ldots, v^N$ be an eigenbasis with eigenvectors $2\pi\lambda_1, \ldots, 2\pi\lambda_N$. Then $b = \sum_{j \in [N]} w_j v_j$. Apply the transformation $\exp(iA)$ to obtain $b = \sum_{j \in [N]} w_j \exp(iA)v_j = \sum_{j \in [N]} w_j \exp(2\pi i\lambda_j)v_j$, and then QPE yielding all the $\lambda_j$. Notice that we stated QPE as applying to a single vector, but because all quantum algorithms are linear operations, then they can also be applied to linear combinations of vectors. Via another quantum operation, we compute $\sum_{j \in [N]} \frac{w_j}{2\pi\lambda_j} v_j = A^{-1}b$. This algorithm has a certain (bounded) probability of failing. It provides an exponential speedup in the size of the linear system, but that is under the assumption that input and output are available in "quantum form": naive input or output would require time proportional to $2^n$.

2.2. **Interior point.** [Kerenidis and Prakash, 2018] give an interior point algorithm for SDP and LP with runtime $\tilde{O}(\frac{n^{3.5}}{\xi^2}\kappa^2 \log \frac{1}{\epsilon})$, where $\epsilon$ is the optimality gap and $\xi$ the constraint satisfaction tolerance. For comparison, the best classical algorithm is $O(n^6)$ in the best case. We remark that this is a polynomial improvement, not exponential, even though it uses an exponentially faster linear systems algorithm: this is because there are bottlenecks in preparing the data for the linear systems, and in extracting the solution of the linear system. The algorithm uses QRAM, a form of storage that is assumed to allow access in superposition. QRAM has not been realized experimentally and it is unclear if it will ever be.

We aim to solve a primal/dual pair:

$$\text{Opt}(P) := \min_{x \in \mathbb{R}^m} \{c^\top x | \sum_{k \in [m]} x_k A^{(k)} \succeq B\}$$

$$\text{Opt}(D) := \max_{Y \succeq 0} \{\text{Tr}(BY) | Y \succeq 0, \text{Tr}(YA^{(j)}) = c_j\}.$$

Let $L = \text{span}_{k \in [m]}(A^{(k)})$, $L^\perp$ its orthogonal complement, and $C$ an arbitrary dual feasible solution. Then the primal and problem can be written as:

$$\text{Opt}(P') := \min_{S \succeq 0} \{\text{Tr}(CS) + \text{Tr}(BC) | S \in (L - B)\}$$

$$\text{Opt}(D) := \max_{Y \succeq 0} \{\text{Tr}(BY) | Y \in (L^\perp + C)\}.$$

The algorithm can be described as follows.

- Input: Matrices $A^{(k)}, k \in [m], B \in \mathbb{R}^{n \times n}, c \in \mathbb{R}^m$ in QRAM, parameters $\epsilon, \delta > 0$.
- Find feasible initial point $(S_0, Y_0, \nu_0)$ and store in QRAM. Set $S \leftarrow S_0, Y \leftarrow Y_0, \nu \leftarrow \nu_0$.
- Repeat $O(\sqrt{n} \log(n/\epsilon))$ times:
    (1) Compute matrices $Y^{-1}$ and $\nu I - SY$ classically, and store in QRAM.
    (2) Solve the Newton linear system:

    $$dS \in L, dY \in L^\perp, SdY + dSY = (1 - \frac{1}{10\sqrt{n}})\nu I - SY.$$

    using block encodings to find estimate of the norm of $\|(dS, dY)\|$.
    (3) Solve the Newton linear system to obtain a quantum state proportional to $(dS, dY)$ to accuracy $\delta^2/n^3$. Recover the vector classically (an operation known as state tomography) and use the norm estimate to obtain classical estimate of $(dS, dY)$.
    (4) Update $S \leftarrow S + dS, Y \leftarrow Y + dY$ in QRAM. Update $\nu \leftarrow \text{Tr}(SY)/n$.
- Output $S, Y$.

This algorithm has polynomial speedup as compared to classical, and has classical input and classical output. However, it relies heavily on QRAM to prepare data structure efficiently (in particular, to update the Newton system).

Another algorithm for SDP is discussed in [Brandao and Svore, 2017] and (indepedently) by [Van Apeldoorn et al., 2017]. It is based on the Arora/Kale multiplicative weights update algorithm. It has bad dependency on some numerical parameters and is not designed to be a practical algorithm, but it does not require QRAM.

## References

[Brandao and Svore, 2017] Brandao, F. G. and Svore, K. M. (2017). Quantum speed-ups for solving semidefinite programs. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 415–426. IEEE.

[Childs et al., 2017] Childs, A. M., Kothari, R., and Somma, R. D. (2017). Quantum algorithm for systems of linear equations with exponentially improved dependence on precision. *SIAM Journal on Computing*, 46(6):1920–1950.

[Farhi et al., 2014] Farhi, E., Goldstone, J., and Gutmann, S. (2014). A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*.

[Harrow et al., 2009] Harrow, A. W., Hassidim, A., and Lloyd, S. (2009). Quantum Algorithm for Linear Systems of Equations. *Physical Review Letters*, 103(15):150502.

[Kerenidis and Prakash, 2018] Kerenidis, I. and Prakash, A. (2018). A quantum interior point method for lps and sdps. *arXiv preprint arXiv:1808.09266*.

[Peruzzo et al., 2014] Peruzzo, A., McClean, J., Shadbolt, P., Yung, M.-H., Zhou, X.-Q., Love, P. J., Aspuru-Guzik, A., and O'brien, J. L. (2014). A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5:4213.

[Shor, 1997] Shor, P. W. (1997). Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509.

[Van Apeldoorn et al., 2017] Van Apeldoorn, J., Gilyén, A., Gribling, S., and de Wolf, R. (2017). Quantum sdp-solvers: Better upper and lower bounds. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 403–414. IEEE.

# Recent developments in the BARON project

## Nikolaos V. Sahinidis

The BARON project began in the early 1990s with the aim to develop an efficient computational tool for the solution of nonconvex mixed-integer nonlinear optimization problems. The BARON code has more than doubled in the past decade as a result of several developments. The preset talk serves as an overview of some of these recent developments, including recently published and unpublished work:

- portfolios or linear and nonlinear relaxations that exploit convex nonlinear relaxations and convexity of intermediates in the functional decomposition of factorable functions [2]
- MIP relaxations, including the parallel solution of MIPs at every node of BARON's branch-and-bound tree [3, 4]
- cutting planes from relaxations of convex-transformable functions [1, 7]
- cutting planes from running intersection inequalities [5]
- bound propagation from optimality conditions [6, 8]
- software engineering for the solution of large-scale problems.

Several examples are used to illustrate the basic principles and extensive computational results are presented on problem collections that include MINLPlib, GlobalLib, PrincetonLib, and QAPlib.

## References

[1] A. Khajavirad, J. J. Michalek and N. V. Sahinidis, *Relaxations of factorable functions with convex-transformable intermediates*, Mathematical Programming **144** (2014), 107–140.

[2] A. Khajavirad and N. V. Sahinidis, *A hybrid LP/NLP paradigm for global optimization relaxations*, Mathematical Programming Computation **10** (2018), 383–421.

[3] M. Kılınç and N. V. Sahinidis, *Exploiting integrality in the global optimization of mixed-integer nonlinear programming problems in BARON*, Optimization Methods and Software **33** (2019), 540–562.

[4] K. Zhou, M. Kılınç, X. Chen and N. V. Sahinidis, *An efficient strategy for the activation of MIP relaxations in a multicore global MINLP solver*, Journal of Global Optimization **32** (1990), 120–140.

[5] A. Del Pia, A. Khajavirad and N. V. Sahinidis, *On the impact of running intersection inequalities for globally solving polynomial optimization problems*, Mathematical Programming Computation (2019), submitted.

[6] Y. Puranik and N. V. Sahinidis, *Bounds tightening based on optimality conditions for non-convex box-constrained optimization*, Journal of Global Optimization **67** (2017), 59–77.

[7] C. J. Nohra and N. V. Sahinidis, *Global optimization of nonconvex problems with convex-transformable intermediates*, Journal of Global Optimization **72** (2018), 255–276.

[8] Y. Zhang, N. V. Sahinidis, C. J. Nohra and G. Rong, *Optimality-based domain reduction for inequality-constrained NLP and MINLP problems*, Journal of Global Optimization (2018), submitted.

## On the exactness of Lasserre relaxations of SPIs and POPs

Markus Schweighofer

(joint work with Tom-Lukas Kriel)

We consider a finite system of non-strict real polynomial inequalities (SPI). Its Lasserre relaxation of degree $d$ is a certain natural linear matrix inequality (LMI) in the original variables and one additional variable for each nonlinear monomial of degree at most $d$. This LMI defines a spectrahedron that projects down to a convex semialgebraic set containing the solution set of the SPI. In the best case, the projection equals the convex hull of the solution set of the SPI. We say that the Lasserre hierarchy *eventually becomes exact* if this is the case for all sufficiently large $d$.

We suppose that the SPI satisfies the *Archimedean* condition which is nearly equivalent to its solution set being compact in the following sense: An Archimedean SPI has compact solution set. Conversely, an SPI with compact solution set can be made Archimedean by adding certain appropriate redundant inequalities.

In [1], Kriel and myself showed that if the solution set of the SLI "bulges outwards" with positive curvature, then the Lasserre hierarchy very often eventually becomes exact. The proof combines ingredients from several areas:

- Real closed fields and real quantifier elimination from real algebraic geometry,
- pure states and separation theorems from functional analysis (applied to vector spaces over real closed fields which are considered as real vector spaces),
- Lagrange multipliers and the Karush-Kuhn-Tucker theorem from convexity (after being transferred to real closed fields),
- the finiteness theorem from first order logic (which follows for example from Gödel's completeness theorem).

The major drawback of this theorem is that it does in general not allow for linear constraints in the SLI . In the talk, we give an example of an SLI with one quartic inequality in two variables that defines the disjoint union of two disks in

the plane of different radii and one linear inequality defining an affine half space that cuts out part of the bigger disk but precisely preserves the smaller disk. In this example, the Lasserre hierarchy does *not* eventually become exact.

With completely different and more traditional techniques, Kriel and myself showed in [2] a similar second theorem whose advantage is that it allows for linear constraints and more generally constraints satisfying a certain "relative sums-of-squares concavity condition". However, this alternative theorem supposes the solution set of the SLI to be convex.

Neither of our two theorems on SLIs seems to be accessible to the techniques we used to prove the respective other theorem.

Now let additionally a polynomial objective function be given, i.e., consider a polynomial optimization problem (POP). Its Lasserre relaxation of degree $d$ is now a semidefinite program (SDP). In the best case, the optimal values of the POP and the SDP agree. In [1], Kriel and myself proved that this often happens if the relaxation degree exceeds some bound that depends on the constraints of the POP and certain characteristicae of the objective like the mutual distance of its global minimizers on the feasible region.

## References

[1] T.-L. Kriel, M. Schweighofer, *On the Exactness of Lasserre Relaxations and Pure States Over Real Closed Fields*, Foundations of Computational Mathematics, Online First (2019), 100–120.

[2] T.-L. Kriel, M. Schweighofer, *On the exactness of Lasserre relaxations for compact convex basic closed semialgebraic sets*, SIAM J. Optim. **28** (2018), 1796–1816.

## Gaining or losing perspective

EMILY SPEAKMAN

(joint work with Jon Lee, Daphne Skipper)

Our interest is in studying "perspective reformulations". This technique has been used in the presence of indicator variables: when an indicator is "off", a vector of decision variables is forced to a specific point, and when it is "on", the vector of decision variables must belong to a specific convex set. [3] studied such a situation where binary variables manage terms in a separable-quadratic objective function, with each continuous variable $x$ being either 0 or in a positive interval (also see [2]). The perspective-reformulation approach (see [3] and the references therein) leads to very strong conic-programming relaxations, but not all MINLO (mixed-integer nonlinear optimization) solvers are equipped to handle these. So one of our interests is in determining when a natural and simpler non conic-programming relaxation may be adequate.

Generally, our view is that MINLO modelers and algorithm/software developers can usefully factor in analytic comparisons of relaxations in their work. $d$-dimensional volume is a natural analytic measure for comparing the size of a pair

of convex bodies in $\mathbb{R}^d$ and [4] introduced the idea of using volume as a measure for comparing relaxations in the context of an optimization problem.

Our view of the current relevant convex-MINLO software environment is that it is very unsettled with a lot to come. One of the best algorithmic options for convex-MINLO is "outer approximation", but this is not usually appropriate when constraint functions are not convex (even when the feasible region of the continuous relaxation is a convex set). Even "NLP-based B&B" for convex-MINLO may not be appropriate when the underlying NLP solver is presented with a formulation where a constraint qualification does not hold at likely optima. In some situations (like ours), the relevant convex sets can be represented as convex cones, thus handling the constraint-qualification issue — but introducing non-differentiability at likely optima. In this way of thinking, conic constraints are not well handled by general convex-MINLO software (like Knitro, Ipopt, Bonmin, etc.). The only conic solver that handles integer variables (via B&B) is MOSEK, and then only quadratic cones, and "as long as they do not contain both quadratic objective or constraints and conic constraints at the same time". So not all of our work can be applied today, within the current convex-MINLO software environment, and so we see our work as forward looking.

We study MINLO formulations of the disjunction $x \in \{0\} \cup [l, u]$, where $z$ is a binary indicator of $x \in [l, u]$, and $y$ "captures" $x^p$, for $p > 1$ (see [1], for example). We investigate a family of relaxations for this model, employing the inequality $yz^q \geq x^p$, parameterized by the "lifting exponent" $q \in [0, p-1]$; we make the convention that $0^0 = 1$ (relevant when $z = 0$ and $q = 0$). These models are higher-dimensional-power-cone representable, and hence tractable in theory. We bound our formulations using the linear inequality $u^p z \geq y$ which is always satisfied at optimality (for the typical application where $y$ replaces $x^p$ in a minimization objective).

For $q = 0$, we have the most *most naive relaxation* using $y \geq x^p$. For $q = 1$, we have the *naive perspective relaxation* using $yz \geq x^p$. For $q = p - 1$, we get the *true perspective relaxation* using $yz^{p-1} \geq x^p$, which gives the convex hull. Interestingly, this last fact seems to be only very-well known when $p = 2$, in which case $p - 1 = 1$ and the naive perspective relaxation is the true perspective relaxation. So some might think, even for $p > 2$, that $q = 1$ would give the convex hull — but this naive perspective relaxation is not the strongest; we need to use $q = p - 1$ to get the convex hull.

We present a formula for the volumes of all of these relaxations as a means of comparing them. In doing so, we quantify, in terms of $l, u, p$, and $q$, how much stronger the convex hull is compared to the weaker relaxations, and when, in terms of $l$ and $u$, there is much to be gained at all by considering more than the weakest relaxation. Using our formula, and thinking of the baseline of $q = 1$, namely the naive perspective relaxation, we quantify the impact of "losing perspective" (e.g., going to $q = 0$, namely the most naive relaxation) and of "gaining perspective" (e.g., going to $q = p - 1$, namely the true perspective relaxation). Depending on $l$ and $u$ for a particular $x$ (of which there may be a great many in a real model),

we may adopt different relaxations based on the differences of the volumes of the various relaxation choices and on the solver environment.

Compared to earlier work on volume formulae and related branching-point selection relevant to comparing convex relaxations, our present results are the first involving convex sets that are not polytopes. Thus we demonstrate that we can get meaningful results that do not rely on triangulation of polytopes.

Finally, we present some computational experiments (for $p = 2$) which bear out our theory, as we verify that volume can be used to determine which variables are more important to handle by perspective relaxation.

## References

[1] M.S. Akturk, A. Atamturk, S. Gurel *A strong conic quadratic reformulation for machine-job assignment with controllable processing times*, Operations Research Letters **37:3** (2009), 187–191.

[2] A. Frangioni, C. Gentile *Perspective cuts for a class of convex 0–1 mixed integer programs*, Mathematical Programming **106:2** (2006), 225–236.

[3] O. Gunluk, J. Linderoth *Perspective reformulations of mixed integer nonlinear programs with indicator variables*, Mathematical Programming, Series B **124** (2010), 183–205.

[4] J. Lee, W. Morris *Geometric comparison of combinatorial polytopes*, Discrete Applied Mathematics **55:2** (1994), 163–182.

[5] J. Lee, D. Skipper, E. Speakman *Gaining or losing perspective*, In: Le Thi H., Le H., Pham Dinh T. (eds) Optimization of Complex Systems: Theory, Models, Algorithms and Applications. WCGO 2019. Advances in Intelligent Systems and Computing, **991**. Springer, Cham (2019), 387–397.

# Recent developments in mixed-integer PDE- and ODE-constrained optimal control

Stefan Ulbrich

(joint work with Oliver Habeck, Kristina Janzen, Christian Kirches, Paul Manns, Marc E. Pfetsch)

In this survey talk we discuss three timely directions of research for mixed-integer optimal control problems (MIOCP) with ODEs and PDEs. In the first part we consider MIOCPs with tight continuous relaxations [9, 10, 11, 12]. Here, on the right hand side of the ODE or the PDE besides a continuous control also an integer control appears that can attain only finitely many different values. From an equivalent partially convex reformulation one obtains a partially convex relaxation that is a standard optimal control problem. From the solution of this problem by a sum-up rounding rule on a grid of size $h$ an $h$-optimal solution of the original MIOCP can be obtained [1, 11, 12]. We state a corresponding relaxation algorithm [9, 11]. Moreover, we discuss in which cases the results can be extended to PDEs and multidimensional problems [10].

In the second part we consider MIOCPs for ODE-/PDE-networks which are inspired by optimization problems for gas, heat, water, traffic and similar networks [5, 6, 7, 8, 13]. Here, we require that only evaluations of the solutions of the ODEs at the node locations enter the problem. We focus on the ODE-constrained case

motivated by nomination problems for stationary gas networks [7, 13]. Then the problem can be reduced by using the solution operator of the ODE on each edge mapping the initial data and controls to the end value of the solution. Assuming that there are computable and convergent upper and lower bounds for the solution operator obtained by numerical schemes, a nonconvex relaxation can be obtained [7, 14]. Convex relaxations of the nonconvex constraints yield a convex relaxation. They can be used in a spatial Branch-and-Bound method that adaptively refines the bounding numerical discretizations of the ODE-solution operator while exploring the Branch-and-Bound tree. For the case of gas networks we show how appropriate discretization schemes for the isothermal Euler equations can be constructed and convex relaxations can be obtained [7]. We present numerical results for GasLib-40 and GasLib-582 instances by using a SCIP-based implementation of the adaptive spatial Branch-and-Bound method.

In the third part we discuss recent results showing that MIOCPs for semilinear elliptic PDEs enjoy under some convexity assumptions on the semilinear terms pointwise concavity and submodularity properties [2, 4]. Hence, for a variety of constraints and objective functions outer approximation algorithms can be applied efficiently [2, 3].

We end the talk by pointing out several challenges and topics for future research.

## References

[1] P. Belotti, C. Kirches, S. Leyffer, J. Linderoth, J. Luedtke, and A. Mahajan. *Mixed-integer nonlinear optimization.* Acta Numer. **22** (2013), 1–131.

[2] C. Buchheim, R. Kuhlmann, and C. Meyer. *Combinatorial optimal control of semilinear elliptic PDEs.* Comput. Optim. Appl. **70** (2018), 641–675.

[3] M. A. Duran and I. E. Grossmann. *An outer-approximation algorithm for a class of mixed-integer nonlinear programs.* Math. Programming **36** (1986), 307–339.

[4] J. Edmonds. Submodular functions, matroids, and certain polyhedra [mr0270945]. In *Combinatorial optimization—Eureka, you shrink!*, volume 2570 of *Lecture Notes in Comput. Sci.*, 11–26. Springer, Berlin, 2003.

[5] B. Geiß ler, O. Kolb, J. Lang, G. Leugering, A. Martin, and A. Morsi. *Mixed integer linear models for the optimization of dynamical transport networks.* Math. Methods Oper. Res. **73** (2011), 339–362.

[6] M. Gugat, G. Leugering, A. Martin, M. Schmidt, M. Sirvent, and D. Wintergerst. *Towards simulation based mixed-integer optimization with differential equations.* Networks **72** (2018), 60–83.

[7] O. Habeck, M. E. Pfetsch, and S. Ulbrich. Global optimization of mixed-integer ode constrained network problems using the example of stationary gas transport. Preprint TRR154-186, Technische Universität Darmstadt, 2018.

[8] M. Hahn, S. Leyffer, and V. Zavala. Mixed-integer PDE-constrained optimal control of gas networks. Preprint ANL/MCS-P7095-0817, Mathematics and Computer Science Division, 2017.

[9] F. M. Hante and S. Sager. *Relaxation methods for mixed-integer optimal control of partial differential equations.* Comput. Optim. Appl. **55** (2013), 197–225.

[10] P. Manns and C. Kirches. Multi-dimensional sum-up rounding for elliptic control systems. Preprint DFG SPP 1962, Technische Universität Braunschweig, 2018.

[11] S. Sager. *Numerical methods for mixed-integer optimal control problems.* Der Andere Verlag, Tönning, Lübeck, Marburg, 2005.

[12] S. Sager, H. G. Bock, and M. Diehl. *The integer approximation error in mixed-integer optimal control.* Math. Program. **133** (2012), 1–23.
[13] L. Schewe, T. Koch, A. Martin, and M. E. Pfetsch. Mathematical optimization for evaluating gas network capacities. In *Evaluating gas network capacities*, volume 21 of *MOS-SIAM Ser. Optim.*, 87–102. SIAM, Philadelphia, PA, 2015.
[14] J. K. Scott and P. I. Barton. *Improved relaxations for the parametric solutions of ODEs using differential inequalities.* J. Global Optim. **57** (2013), 143–176.

# New Relaxations for Composite Functions

Mohit Tawarmalani

(joint work with Taotao He)

We introduce a new relaxation framework for MINLPs. These new relaxations, called composite relaxations, are tighter than the prevalent factorable programming relaxations [6], implemented in many state-of-the-art solvers.

While most relaxation strategies consider only bounds on the inner-functions, or, consider at most one estimator for each inner function at a time, we develop ways to exploit multiple estimators for each inner functions while generating valid inequalities. Our relaxation procedure proceeds in two steps. First, we encode the inner function structure into a polytope in a generic fashion. Second, the graph of the outer-function over this polytope is relaxed into a convex set [3].

Although the separation problem for the convex hull of the graph of the outer-function over the constructed polytope is NP-Hard in general, we show that separation is tractable in several cases. In particular, consider an outer-function, which is supermodular and concave-extendable and let $d$ and $n$ denote respectively the number of inner functions and the number of estimators for each inner-function. For this case, we develop an $O(dn \log d)$ algorithm [4] for separating the hypograph of the outer-function. We remark that this result generalizes various specially structured concave envelope results over the hypercube, because our domain includes the hypercube as a special case. We show that tractable separation is also possible when the outer-functions are not supermodular, but $d$ is fixed. When the outer-function is convex in each argument, the limiting relaxation obtained with infinitely many estimators for each inner-function [4] is shown to be related to the solution of an optimal transport problem.

We argue that the composite relaxations are particularly well-suited for constructing MIP relaxations via discretization strategies and for relaxations of functions with discrete domains [5]. By exploiting our results for discrete domains, we show that, for certain compositions of univariate functions, we can construct a sequence of polyhedral relaxations that converge, in the limit, to the concave envelope. Our proposed discretization is obtained by reinterpreting the incremental formulation [2] and combining it with the composite relaxation. Moreover, we show that composite relaxations can also be used to improve most relaxation hierarchies [1] without introducing additional variables. Our results extend to

simultaneous convexification of functions. We conclude with preliminary computational experience with these relaxations that demonstrates that these techniques help reduce the gap significantly on some specially structured problem instances.

## References

[1] H. D. Sherali and W. Adams, *A reformulation-linearization technique for solving discrete and continuous nonconvex problems*, Springer Science & Business Media, 2013.
[2] G. B. Dantzig, *On the significance of solving linear programming problems with some integer variables*, Econometrica (1960), 30–44.
[3] T. He and M. Tawarmalani, *A new framework to relax composite functions in nonlinear programs*, Mathematical Programming (2018), submitted.
[4] T. He and M. Tawarmalani, *Tractable relaxations of composite functions*, Mathematics of Operations Research (2018), submitted.
[5] T. He and M. Tawarmalani, *Extracting structure from expression trees for discrete and continuous relaxations of MINLPs*, (2019), working paper.
[6] G. P. McCormick, *Computability of global solutions to factorable nonconvex programs: Part I – Convex underestimating problems*, Mathematical Programming **10** (1976), 147–175.

## Sums-of-squares for extremal discrete geometry on the unit sphere

Frank Vallentin

(joint work with D. de Laat, F.C. Machado. F.M. de Oliveira Filho)

The goal of this talk is to present a hierarchy of $k$-point semidefinite programming upper bounds for the maximum number of equiangular lines in $n$-dimensional Euclidean space. In [4] we apply symmetry reduction techniques for invariant semidefinite programs to compute the bound for $k = 4$, 5, and 6, finding many improved bounds for the maximum number of equiangular lines with fixed common angle.

A set of *equiangular lines* is a set of lines having a common intersection point such that every pair of lines defines the same angle. It is an interesting and in general open question to determine the maximum size of a set of equiangular lines. Next to being fundamental objects in discrete geometry, equiangular lines have applications, for example in the field of signal processing (compressive sensing, finite Hilbert space frames).

A general framework for hierarchies of semidefinite programming bounds for geometric packing problems was developed by de Laat and Vallentin [5]. This framework is an infinite-dimensional generalization of Lasserre's hierarchy for computing the independence number of a finite graph. Let $G = (V, E)$ be a graph. A subset of $V$ is *independent* if it does not contain pairs of adjacent vertices. The *independence number* of $G$, denoted by $\alpha(G)$, is the maximum cardinality of an independent set. For an integer $k \geq 0$, let $I_k$ be the set of independent sets in $G$ of size at most $k$ and $I_{=k}$ be the set of independent sets in $G$ of size exactly $k$. The maximum number of equiangular lines in $n$ dimensions with common angle $\arccos a$ is equal to $M_a(n) = \alpha(G_{n,a})$ where the infinite graph $G_{n,a}$

has vertex set $V = S^{n-1}$, the unit sphere, and $\{x, y\}$ forms an edge if and only if $x \cdot y \in [-1, 1) \setminus \{\pm a\}$.

The aim is to compute $k$-point SDP bounds for larger values of $k$, like $k = 4, 5, 6$. For achieving this we apply the framework of de Laat and Vallentin and derive a new block-diagonal hierarchy of semidefinite $k$-point bounds which is especially suited for computing bounds for equiangular lines.

Our block-diagonal hierarchy (presented below in its dual form) is similar to the one of Gvozdenović, Laurent, Vallentin [6] for finite graphs. Define the convex cone $\mathcal{C}(V^2 \times I_{k-2})_{\succeq 0}$ by $T \in \mathcal{C}(V^2 \times I_{k-2})_{\succeq 0}$ if and only if for every $Q \in I_{k-2}$ the kernel $(x, y) \mapsto T(x, y, Q)$ is psd. Define the operator $B_k \colon \mathcal{C}(V^2 \times I_{k-2})_{sym} \to \mathcal{C}(I_k \setminus \{\emptyset\})$ by

$$B_k T(S) = \sum_{\substack{Q \subseteq S \\ |Q| \leq k-2}} \sum_{\substack{x,y \in S \\ Q \cup \{x,y\} = S}} T(x, y, Q).$$

Our new $k$-point semidefinite programming bound is defined as

$$\Delta_k(G)^* = \inf\{\, 1 + \lambda : \lambda \in \mathbb{R}, \ T \in \mathcal{C}(V^2 \times I_{k-2})_{\succeq 0}, B_k T \leq \lambda \chi_{I_{=1}} - 2\chi_{I_{=2}} \,\},$$

where $\chi_{I_{=1}}$, $\chi_{I_{=2}}$ are characteristic functions. Note that the blocks are indexed by the set $I_{k-2}$.

**Theorem.** $\alpha(G) \leq \Delta_k(G)^*$ for every compact topological packing graph $G$.

The poof is easy: Let $C \subseteq V$ be independent, let $(\lambda, T)$ be feasible for $\Delta_k(G)^*$. On the one hand,

$$\sum_{\substack{S \subseteq C \\ |S| \leq k, \ S \neq \emptyset}} B_k T(S) \leq \binom{|C|}{1} \lambda + \binom{|C|}{2}(-2) = |C|(1 + \lambda - |C|).$$

On the other hand,

$$\sum_{\substack{S \subseteq C \\ |S| \leq k, \ S \neq \emptyset}} B_k T(S) = \sum_{\substack{S \subseteq C \\ |S| \leq k, \ S \neq \emptyset}} \sum_{\substack{Q \subseteq S \\ |Q| \leq k-2}} \sum_{\substack{x,y \in S \\ Q \cup \{x,y\} = S}} T(x, y, Q)$$

$$= \sum_{\substack{Q \subseteq C \\ |Q| \leq k-2}} \sum_{x,y \in C} T(x, y, Q) \geq 0.$$

Symmetry reduction plays a key role in the computation of $\Delta_k(G)^*$ via semidefinite programming. In a sense, *we block-diagonalize the block-diagonal hierarchy twice*. Let $\Gamma$ be a subgroup of the automorphism group $\mathrm{Aut}(G)$. The group $\Gamma$ acts on $\mathcal{C}(V^2 \times I_{k-2})_{sym}$ by $(\gamma T)(x, y, S) = T(\gamma^{-1}x, \gamma^{-1}y, \gamma^{-1}S)$. Given feasible $(\lambda, T)$ of $\Delta_k(G)^*$, the pair $(\lambda, \overline{T})$ with

$$\overline{T}(x, y, S) = \int_\Gamma T(\gamma^{-1}x, \gamma^{-1}y, \gamma^{-1}S) \, d\gamma,$$

is also feasible, with the same objective function. So we can restrict to $\mathcal{C}(V^2 \times I_{k-2})_{\succeq 0}^\Gamma$ of $\Gamma$-invariant kernels.

To parametrize $\mathcal{C}(V^2 \times I_{k-2})_{\succeq 0}^\Gamma$, we define $\mathcal{R}_{k-2}$ as the complete set of representatives of the orbits of $I_{k-2}/\Gamma$. The stabilizer is defined as $\mathrm{Stab}_\Gamma(R) = \{\, \gamma \in \Gamma : \gamma R = R \,\}$ of $R \in \mathcal{R}_{k-2}$.

**Theorem.** If $I_{k-2}/\Gamma$ is finite, then

$$\mathcal{C}(V^2 \times I_{k-2})^\Gamma \cong \bigoplus_{R \in \mathcal{R}_{k-2}} \mathcal{C}(V^2)^{\mathrm{Stab}_\Gamma(R)}.$$

The isomorphism is explicit and preserves positivity.

Thus, to apply it for equiangular lines we need to parametrize $\mathcal{C}((S^{n-1})^2)^{\mathrm{Stab}_{O(n)}(R)}_{\succeq 0}$. Here, the following theorem of Musin [8] is useful. Musin's theorem generalizes results by Schoenberg [9] ($m = 0$) and by Bachoc, Vallentin [1] ($m = 1$).

**Theorem.** Let $R \subseteq S^{n-1}$ with $m = \dim(\mathrm{span}(R)) \leq n - 2$ and let $E$ be an $m \times n$ matrix whose rows form an orthonormal basis for $\mathrm{span}(R)$. Fix $d \geq 0$ and, for each $0 \leq l \leq d$, let $F_l$ be a positive semidefinite matrix of size $\binom{d-l+m}{m} \times \binom{d-l+m}{m}$. Then $K \colon S^{n-1} \times S^{n-1} \to \mathbb{R}$ given by

$$K(x,y) = \sum_{l=0}^d \left\langle F_l, Y_l^{n,m}(x \cdot y, Ex, Ey) \right\rangle$$

is a positive, continuous, and $\mathrm{Stab}_{O(n)}(\mathrm{span}(R))$-invariant kernel. Every such kernel can be uniformly approximated by this formula.

Here, matrix $Y_l^{n,m}$ is matrix of polynomials $Y_l^{n,m}(t,u,v) = P_l^{n,m}(t,u,v)z_{d-l}(u)z_{d-l}(v)^\mathsf{T}$, where the multivariate Gegenbauer polynomials are used

$$P_l^{n,m}(t,u,v) = ((1 - \|u\|^2)(1 - \|v\|^2))^{l/2} P_l^{n-m}\left( \frac{t - u \cdot v}{\sqrt{(1 - \|u\|^2)(1 - \|v\|^2)}} \right),$$

and $z_l(u)$ column vector containing basis of polynomials of degree at most $l$.

Back to the problem of bounding equiangular lines. The literature on this subject is vast and we point out only some major theorems and conjectures in this area.

**Theorem.** (Lemmens, Seidel [7]) If $n \geq 15$, then $M_{1/3}(n) = 2n - 2$.

**Theorem.** (Balla, Dräxler, Keevash, Sudakov [2]) $M_a(n) \leq 1.93n$ for $a \in (0,1) \setminus \{1/3\}$ for large enough $n = n(a)$.

**Conjecture.** (Lemmens, Seidel [7]) We have

$$M_{1/5}(n) = \begin{cases} 276 & \text{for } 23 \leq n \leq 185; \\ \lfloor \tfrac{3}{2}(n-1) \rfloor & \text{for } n \geq 185. \end{cases}$$

**Conjecture.** (Bukh [3] If $1/a \geq 3$ is an odd integer, then $M_a(n) = \frac{1+a}{1-a}n + O(1)$.

These conjectures spur the following question: What is the smallest $n$ such that $M_a(n) = (1/a^2 - 2)(1/a^2 - 1)/2$? To adress it we computed $\Delta_k(G)^*$ for $k = 3, 4, 5, 6$. Computing and verifying $\Delta_6(G)^*$ rigorously takes several days. We got the following table where the values in the columns $\Delta_k^*$ give the largest dimension for which $\Delta_k^*$ is equal to $(1/a^2 - 2)(1/a^2 - 1)/2$.

| $a$ | $(1/a^2 - 2)(1/a^2 - 1)/2$ | $\Delta_3^*$ | $\Delta_4^*$ | $\Delta_5^*$ | $\Delta_6^*$ |
|---|---|---|---|---|---|
| 1/5 | 276 | 60 | 65 | 69 | 70 |
| 1/7 | 1128 | 131 | 145 | 158 | 169 |
| 1/9 | 3160 | 227 | 251 | 273 | 300 |
| 1/11 | 7140 | 347 | 381 | 413 | 448 |

On the other hand Bukh's construction shows that $(1/a^2 - 2)(1/a^2 - 1)/2$ is achieved in the dimensions 185, 847, 2529, 5951 for $a = 1/5, 1/7, 1/9, 1/11$.

We end with a few open questions:

(1) Why does $\Delta_k(G_{n,a})^*$ stabilize at $(1/a^2 - 2)(1/a^2 - 1)/2$?
(2) Is $\Delta_k(G_{n,a})^* = \Theta(n^2)$ for fixed $k$ and $a$?
(3) Is $\Delta_{\alpha(G)}(G) = \alpha(G)$ similarly to the Lasserre hierarchy?

## References

[1] Christine Bachoc and Frank Vallentin. New upper bounds for kissing numbers from semi-definite programming. *J. Amer. Math. Soc.*, 21(3):909–924, 2008.
[2] Igor Balla, Felix Dräxler, Peter Keevash, and Benny Sudakov. Equiangular lines and spherical codes in Euclidean space. *Invent. Math.*, 211(1):179–212, 2018.
[3] Boris Bukh. Bounds on equiangular lines and on related spherical codes. *SIAM J. Discrete Math.*, 30(1):549–554, 2016.
[4] David de Laat, Fabrício Caluza Machado, Fernando Mário de Oliveira Filho and Frank Vallentin. $k$-point semidefinite programming bounds for equiangular lines. arXiv:1812.06045 [math.OC], 2018.
[5] David de Laat and Frank Vallentin. A semidefinite programming hierarchy for packing problems in discrete geometry. *Math. Program.*, 151(2, Ser. B):529–553, 2015.
[6] Nebojsa Gvozdenović, Monique Laurent, and Frank Vallentin. Block-diagonal semidefinite programming hierarchies for 0/1 programming. *Oper. Res. Lett.*, 37(1):27–31, 2009.
[7] P.W.H. Lemmens and Johan J. Seidel. Equiangular lines. *J. Algebra*, 24:494–512, 1973.
[8] Oleg R. Musin. Multivariate positive definite functions on spheres. In *Discrete geometry and algebraic combinatorics*, volume 625 of *Contemp. Math.*, pages 177–190. Amer. Math. Soc., Providence, RI, 2014.
[9] Isaac J. Schoenberg. Positive definite functions on spheres. *Duke Math. J.*, 9:96–108, 1942.

*Reporter: Leo Liberti*

# Participants

**Dr. Amir Ali Ahmadi**
Department of Operations Research
and Financial Engineering
Princeton University
Princeton, NJ 08540
UNITED STATES

**Prof. Dr. Kurt M. Anstreicher**
Department of Business Analytics
The University of Iowa
Iowa City IA 52242-1994
UNITED STATES

**Prof. Dr. Gennadiy Averkov**
Fakultät für Mathematik
Otto-von-Guericke-Universität
Magdeburg
Postfach 4120
39016 Magdeburg
GERMANY

**Prof. Dr. Pietro Belotti**
FICO - Xpress Optimization Team
International Square
Starley Way, Marston Green
Birmingham B37 7GN
UNITED KINGDOM

**Prof. Dr. Daniel Bienstock**
Industrial Engineering and Operations
Research
Columbia University
500 West 120th Street
New York, NY 10027
UNITED STATES

**Prof. Dr. Christoph Buchheim**
Fakultät für Mathematik
Technische Universität Dortmund
Vogelpothsweg 87
44227 Dortmund
GERMANY

**Martina Cerulli**
Laboratoire d'Informatique (LIX)
École Polytechnique
91128 Palaiseau Cedex
FRANCE

**Prof. Dr. Claudia D'Ambrosio**
Laboratoire d'Informatique (LIX)
École Polytechnique
91128 Palaiseau Cedex
FRANCE

**Prof. Dr. Etienne de Klerk**
Faculty of Economic Sciences
Tilburg University
Warandelaan 2
P.O. Box 90153
5000 LE Tilburg
NETHERLANDS

**Prof. Dr. Marianna De Santis**
Dipartimento di Ingegneria Informatica
Automatica e Gestionale Antonio
Ruberti
Sapienza Università di Roma
Via Ariosto 25
00185 Roma
ITALY

**Prof. Dr. Mirjam Dür**
Institut für Mathematik
Universität Augsburg
86135 Augsburg
GERMANY

**Prof. Dr. Marcia Fampa**
Instituto de Matematica
Universidade Federal do Rio de Janeiro
Av. Pedro Calmon, N° 550
Rio de Janeiro 21945-970
BRAZIL

**Prof. Dr. Antonio Frangioni**
Dipartimento di Informatica
Università di Pisa
Largo B. Pontecorvo, 3
56127 Pisa
ITALY

**Prof. Dr. Armin Fügenschuh**
Brandenburgische Technische Universität
Cottbus - Senftenberg
Fachgebiet Ingenieurmathematik und
Numerik der Optimierung
Platz der Deutschen Einheit 1
03046 Cottbus
GERMANY

**Dipl.-Ing. Dr. Elisabeth Gaar**
Institut für Mathematik
Alpen-Adria-Universität Klagenfurt
Universitätsstrasse 65-67
9020 Klagenfurt
AUSTRIA

**Dr. Laura Galli**
Dipartimento di Informatica
- Operations Research -
Università di Pisa
Largo B. Pontecorvo, 3
56127 Pisa
ITALY

**Dr. Ambros Gleixner**
Research Group for Mathematical
Optimization Methods
Konrad-Zuse-Zentrum für
Informationstechnik (ZIB)
Takustrasse 7
14195 Berlin
GERMANY

**Nicolò Gusmeroli**
Institut für Mathematik
Alpen-Adria-Universität Klagenfurt
Universitätsstrasse 65-67
9020 Klagenfurt
AUSTRIA

**Mirko Hahn**
Fakultät für Mathematik
Otto-von-Guericke-Universität
Magdeburg
Postfach 4120
39016 Magdeburg
GERMANY

**Prof. Dr. Christoph Helmberg**
Fakultät für Mathematik
Technische Universität Chemnitz
Reichenhainer Strasse 39
09126 Chemnitz
GERMANY

**Prof. Dr. Volker Kaibel**
Fakultät für Mathematik
Otto-von-Guericke-Universität
Magdeburg
Universitätsplatz 2
39106 Magdeburg
GERMANY

**Dr. Aida Khajavirad**
Courant Institute of Mathematical
Sciences
New York University
251, Mercer Street
New York, NY 10012-1110
UNITED STATES

**Prof. Dr. Christian Kirches**
Institute for Mathematical Optimization
Technische Universität Braunschweig
Universitätsplatz 2
38106 Braunschweig
GERMANY

**Prof. Dr. Monique Laurent**
Centrum Wiskunde and Informatica
(CWI)
Postbus 94079
1090 GB Amsterdam
NETHERLANDS

**Prof. Dr. Jon X. Lee**
Department of Industrial and
Operations Engineering
University of Michigan
1205 Beal Avenue
Ann Arbor MI 48109-2117
UNITED STATES

**Dr. Sven Leyffer**
Mathematics and Computer Science
Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439-4844
UNITED STATES

**Prof. Dr. Leo Liberti**
Laboratoire d'Informatique (LIX)
École Polytechnique
91128 Palaiseau Cedex
FRANCE

**Prof. Dr. Frauke Liers**
Department Mathematik
Universität Erlangen-Nürnberg
Cauerstrasse 11
91058 Erlangen
GERMANY

**Prof. Dr. Jeffrey T. Linderoth**
Department of Industrial Engineering
The University of Wisconsin-Madison
1513 University Avenue
Madison, WI 53706-1572
UNITED STATES

**Prof. Dr. Andrea Lodi**
Department of Mathematics and
Industrial Engineering - École
Polytechnique de Montréal
GERAD - HEC Montréal, Office 5507
3000, Côte-Sainte-Catherine Road
Montreal QC H3T 2A7
CANADA

**Paul Manns**
Institute for Mathematical Optimization
Technische Universität Braunschweig
Universitätsplatz 2
38106 Braunschweig
GERMANY

**Prof. Dr. Alexander Martin**
Lehrstuhl für Wirtschaftsmathematik
Universität Erlangen-Nürnberg
Cauerstrasse 11
91058 Erlangen
GERMANY

**Dr. Maximilian Merkert**
Fakultät für Mathematik
Otto-von-Guericke-Universität
Magdeburg
Postfach 4120
39016 Magdeburg
GERMANY

**Dr. Ruth Misener**
Department of Computing
Imperial College London
Huxley Building
London SW7 2AZ
UNITED KINGDOM

**Prof. Dr. Giacomo Nannicini**
IBM Corporation
Thomas J. Watson Research Center
1101 Kitchawan Road
Yorktown Heights, NY 10598
UNITED STATES

**Prof. Dr. Laura Palagi**
Dipartimento di Ingegneria Informatica
Automatica e Gestionale Antonio
Ruberti
Sapienza Università di Roma
Via Ariosto, 25
00185 Roma
ITALY

**Prof. Dr. Veronica Piccialli**
Dipartimento di Ingegneria Civile
e Ingegneria Informatica
Università degli Studi di Roma "Tor
Vergata"
Via del Politecnico, 1
00133 Roma
ITALY


**PD Dr. Andreas Potschka**
Interdisziplinäres Zentrum für
Wissenschaftliches Rechnen
Universität Heidelberg
Im Neuenheimer Feld 205
69120 Heidelberg
GERMANY


**Prof. Dr. Franz Rendl**
Institut für Mathematik
Alpen-Adria-Universität Klagenfurt
Universitätsstrasse 65-67
9020 Klagenfurt
AUSTRIA


**Prof. Dr. Sebastian Sager**
Fakultät für Mathematik
Otto-von-Guericke-Universität
Magdeburg
Postfach 4120
39016 Magdeburg
GERMANY


**Prof. Dr. Nick Sahinidis**
Chemical Engineering Department
Carnegie Mellon University
5000, Forbes Avenue
Pittsburgh PA 15213
UNITED STATES


**Prof. Dr. Claus Scheiderer**
Fachbereich für Mathematik und
Statistik
Universität Konstanz
Fach D 203
78457 Konstanz
GERMANY

**Ihno Schrot**
Bogenstrasse 3
69124 Heidelberg
GERMANY


**Prof. Dr. Rüdiger Schultz**
Fachbereich Mathematik
Universität Duisburg-Essen
Thea-Leymann-Strasse 9
45127 Essen
GERMANY


**Prof. Dr. Markus Schweighofer**
Fachbereich für Mathematik und
Statistik
Universität Konstanz
78457 Konstanz
GERMANY


**Dr. Emily Speakman**
Institut für Mathematische Optimierung
Universität Magdeburg
Postfach 4120
39016 Magdeburg
GERMANY


**Prof. Dr. Mohit Tawarmalani**
Krannert School of Management
Purdue University
4070 Rawls Hall
100, S. Grant Street
West Lafayette IN 47907-2076
UNITED STATES


**Prof. Dr. Stefan Ulbrich**
Fachbereich Mathematik
Fachgebiet: Nichtlineare Optimierung
Technische Universität Darmstadt
Dolivostraße 15
64293 Darmstadt
GERMANY

**Prof. Dr. Frank Vallentin**
Mathematisches Institut
Universität zu Köln
Weyertal 86-90
50931 Köln
GERMANY

**Shudian Zhao**
Institut für Mathematik
Alpen-Adria-Universität Klagenfurt
Universitätsstrasse 65-67
9020 Klagenfurt
AUSTRIA

**Prof. Dr. Angelika Wiegele**
Institut für Mathematik
Alpen-Adria-Universität Klagenfurt
Universitätsstrasse 65-67
9020 Klagenfurt
AUSTRIA