Solving Nonlinear Problems via Disjunctions

Daniel Bienstock

Columbia University

October 2019

Some of my papers that were influenced by Egon's work and which relate to today's talk

- *Subset Algebra Lift Operators for 0-1 Integer Programming*, with M. Zuckerberg (2004).
- *Tree-width and the Sherali-Adams operator*, with N. Özbay (2004)
- *Strong formulations for convex functions over nonconvex sets*, with A. Michalka (2014).
- *LP formulations for polynomial optimization problems*, with G. Muñoz (2015).
- *Outer-Product-Free Sets for Polynomial Optimization and Oracle-Based Cuts*, with C. Chen and G. Muñoz (2018).
- *Principled Deep Neural Network Training through Linear Programming*, with G. Muñoz and S. Pokutta (2019).

Some of my papers that were influenced by Egon's work and which relate to today's talk

- *Subset Algebra Lift Operators for 0-1 Integer Programming*, with M. Zuckerberg (2004).
- *Tree-width and the Sherali-Adams operator*, with N. Özbay (2004)
- *Strong formulations for convex functions over nonconvex sets*, with A. Michalka (2014).
- **LP formulations for polynomial optimization problems**, with G. Muñoz (2015).
- **Outer-Product-Free Sets for Polynomial Optimization and Oracle-Based Cuts**, with C. Chen and G. Muñoz (2018).
- **Principled Deep Neural Network Training through Linear Programming**, with G. Muñoz and S. Pokutta (2019).

We wanteed to talk about ongoing computational work

But it is not ready

So we will settle for some theory

Approximate optimization of well-behaved functions

Prototype problem:

$$\begin{aligned} \min \ & c^T x \\ \text{s.t. } & f_i(x) \leq 0, \qquad i = 1, \ldots, m \\ & x \in [0,1]^n, \ x_j \in \{0,1\}, \ j \in J \end{aligned}$$

Each $\boldsymbol{f_i}$ is "well-behaved": Lipschitz constant $\boldsymbol{\mathcal{L}_i}$
Note: it appears redundant to say that some variables are binary

Toolset:

- **Intersection graph**
  A vertex for each variable and an edge whenever two variables appear in the same $\boldsymbol{f_i}$
- **Tree-width** Min clique number (minus one) over all chordal supergraphs of G

Prototype problem:

$$\begin{aligned} \min \ & c^T x \\ \text{s.t. } & f_i(x) \le 0, \qquad i = 1, \ldots, m \\ & x \in [0,1]^n, \ x_j \in \{0,1\}, \, j \in J \end{aligned}$$

An extension of work in B. and Muñoz 2015, SIOPT 2018.

Suppose:

the intersection graph has tree-width ω and $\boldsymbol{f_i}$ has Lipschitz constant $\boldsymbol{\mathcal{L}_i} \le \boldsymbol{\mathcal{L}}$.

If problem is feasible, for every $\mathbf{0} < \epsilon < \mathbf{1}$ there is an **LP** relaxation with

$\boldsymbol{O\left((\mathcal{L}/\epsilon)^{\omega+1}\,(n+m)\log(\mathcal{L}/\epsilon)\right)}$ variables and constraints, and

optimality and feasibility errors $\boldsymbol{O(\epsilon)}$

Main technique: approximation through pure-binary problems

Glover, 1975 (extended)
Let x be a variable, with bounds $0 \le x \le 1$. Let $0 < \gamma < 1$. Then we can approximate

$$x \approx \sum_{h=1}^{K} 2^{-h} y_h$$

where each y_h is a **binary variable**. In fact, choosing $K = \lceil \log_2 \gamma^{-1} \rceil$, we have

$$x \le \sum_{h=1}^{K} 2^{-h} y_h \le x + \gamma.$$

→ Given a mixed-integer well-behaved problem
apply this technique to each continuous variable x_j

The main result

$$\begin{aligned} c^* \doteq \min \ & c^T x \\ \text{s.t. } & f_i(x) \le 0, \qquad i = 1, \ldots, m \\ & x \in [0,1]^n, \quad x_j \in \{0,1\}, j \in J \end{aligned}$$

- Intersection graph with tree-width ω,
- Each f_i has Lipschitz constant $\mathcal{L}_i \le \mathcal{L}$.

For $0 < \epsilon < 1$, an **LP** relaxation of size $O\left((1/\epsilon)^{\omega+1} (n+m) \log(1/\epsilon)\right)$

yields $\hat{x} \in [0,1]^n, \ \hat{x}_J \in \{0,1\}^J$ with

- $c^T \hat{x} \ \le \ c^* + O(\|c\|_1 \epsilon)$
- $f_i(\hat{x}) \le O(\mathcal{L}_i \epsilon), \qquad i = 1, \ldots, m$

How is it done?

- **Lifting hierarchies** in 0-1 linear integer programming
- Balas, Lovász-Schrijver, Sherali-Adams, Balas-Ceria-Cornuéjols
- Specific version: zeta-function idea of Lovász-Schrijver

$$x \in \{0,1\}^n \ \rightarrow \ X \in \{0,1\}^{2^{[n]}}$$

Applications

- Fixed-charge network flow problems, on networks with small treewidth

- ACOPF problem, on networks with small treewidth

- In both cases, LP of size $\boldsymbol{O}\left((\mathbf{1}/\epsilon)^{\omega+1}\, \boldsymbol{n} \log(\mathbf{1}/\epsilon)\right)$

$$\begin{aligned} \omega = &\ \text{treewidth of network} \\ \leq &\ \text{treewidth of intersection graph of formulation} \end{aligned}$$

- A famous scientist: "lifting hierarchies do not work"

Another “application”

Positivestellensätze for semi-algebraic systems

- Given a system of **polynomial** inequalities $\boldsymbol{p_i(x) \leq 0, \quad i \in I}$
- A positivstellensatz is a **proof of infeasibility** of the system
- Under assumptions (e.g. compactness) such statements exist (rich literature)

(2017) Amir-Ali Ahmadi, Georgina Hall $\Rightarrow$ a new positivstellensatz, under compactness (containment in a known ball)

The lifted LP relaxation hierarchy gives a similar result

We are given a system

$$f_i(x) \le 0, \qquad i \in I \tag{1a}$$
$$x \in [0,1]^n, \quad x_j \in \{0,1\}, j \in J. \tag{1b}$$

where each $\boldsymbol{f_i}$ is a function with Lipschitz constant $\boldsymbol{\mathcal{L}}$.

$\rightarrow$ If the system is **infeasible**, is there a short proof thereof? Yes:

1. Given ϵ, our **LP** $\mathbf{= LP(\epsilon)}$, **if feasible**, yields $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}(\boldsymbol{\epsilon})$ with

$$\boldsymbol{f_i(\hat{x}) \le O(\mathcal{L}\epsilon) \quad \forall i \in I, \;\; \hat{x} \in [0,1]^n, \;\; \hat{x}_j \in \{0,1\}, j \in J}$$

2. Let $\epsilon \rightarrow \mathbf{0}$: $\hat{\boldsymbol{x}}$ has accumulation point $\boldsymbol{x^*}$
3. But system (1a), (1b) is infeasible, so $\boldsymbol{x^*}$ cannot be feasible!
4. **Conclusion:** **LP(ϵ) is not feasible for** some ϵ small enough
5. But **LP(ϵ)** is a relaxation for (1a), (1b)

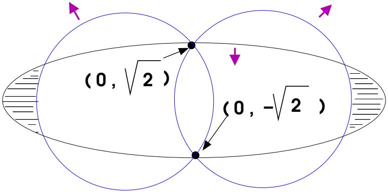
A bad QCQP

```
Maximize x2

s.t.
       x3 - x1 = -1
       x4 - x1 =  1
   o1: [ x3^2 + x2^2 - sneaky^2 ] >= 3
   o2: [ x4^2 + x2^2 ] >= 3
   e1: [ .1 x1^2 + x2^2 ] <= 2

  bad:  distraction + [ sneaky^2 ]         >= 0.1

joke1: - a + [ distraction^2 + sneaky^2 ]  <= 0.0

cruel:  - sneaky + [ a^2 + sneaky^2 ]      <= 0.0

Bounds
x1 free
x2 free
x3 free
x4 free
End
```

→ Gurobi, SCIP, other codes: value ≈ **1.4142**
Wrong, actual value ≈ **1.22**

What's going on?

$$\begin{array}{rrcl} & \max & x_2 & \\ \text{s.t.} & (x_1-1)^2 + x_2^2 & \geq & 3 \\ & (x_1+1)^2 + x_2^2 & \geq & 3 \\ & \frac{x_1^2}{10} + x_2^2 & \leq & 2 \end{array}$$



But this is NOT the problem being solved ..

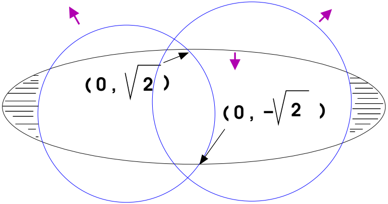
The bad QCQP

```
Maximize x2

s.t.
       x3 - x1 = -1
       x4 - x1 =  1
   o1: [ x3^2 + x2^2 - sneaky^2 ] >= 3
   o2: [ x4^2 + x2^2 ] >= 3
   e1: [ .1 x1^2 + x2^2 ] <= 2

  bad:  distraction + [ sneaky^2 ]          >= 0.1

joke1: - a + [ distraction^2 + sneaky^2 ]  <= 0.0

cruel:  - sneaky + [ a^2 + sneaky^2 ]      <= 0.0

Bounds
x1 free
x2 free
x3 free
x4 free
End
```

→ Gurobi, SCIP, other codes: value ≈ **1.4142**
Wrong, actual value ≈ **1.22**

Actually THIS is the problem being solved:

$$
\begin{array}{rrcl}
 & \max \quad x_2 & & \\
\text{s.t.} & (x_1-1)^2 + x_2^2 & \geq & 3 + \phi \qquad\qquad (\phi > 0) \\
 & (x_1+1)^2 + x_2^2 & \geq & 3 \\
 & \dfrac{x_1^2}{10} + x_2^2 & \leq & 2
\end{array}
$$



Fri.Oct.25.162243.2019@littleboy