Perspective Reformulations Beyond the Separable Case

Antonio Frangioni

Dipartimento di Informatica, Università di Pisa

with C. Gentile, J. Hungerford

Oberwolfach Workshop ID 1923
Mixed-integer Nonlinear Optimization:
a hatchery for modern mathematics
June 7, 2019

Outline

The Target Structure

- MINLP with convex-cost semicontinuous (vector) variables

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

  - f_i closed convex, $f_i(0) = 0$, $\mathcal{P}_i = \{\, x_i \in \mathbb{R}^{m_i} \;:\; A_i x_i \leq b_i \,\}$ compact
  - $y_i = 0 \Longrightarrow x_i = 0$, $y_i = 1 \Longrightarrow x_i \in \mathcal{P}_i$
  - $z \,/\, \mathcal{O} \subset \mathbb{R}^{m+n+q}$ ($m = \sum_{i \in N} m_i$) "other" variables / constraints

- Basic tool: $y \in \{\, 0 \,,\, 1 \,\}^n \rightarrow y \in [\, 0 \,,\, 1 \,]^n \equiv$ continuous relaxation $(\underline{P})$

- If g, $\mathcal{O}$ convex, $(\underline{P})$ "easy", $\nu(\underline{P}) \leq \nu(P)$, continuous solution useful for many things (heuristics, branching, ...)

- All hinges on $\nu(\underline{P})$ being a "good bound", but it may not be

Exploiting the Structure

- To improve the bound we must exploit the structure
- Single block problem

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

$$f(x, y) = \begin{cases} 0 & \text{if } y = 0 \text{ and } x = 0 \\ f(x) + cy & \text{if } y = 1 \text{ and } Ax \leq b \\ +\infty & \text{otherwise} \end{cases}$$

- $co\, f$ = convex envelope = "best" lower convex approximation (actually closed convex envelope . . .)
- Computing convex envelopes is hard in general, but in this case

$$co\, f(x, y) = \begin{cases} 0 & \text{if } y = 0 \text{ and } x = 0 \\ yf(x/y) + cy & \text{if } y \in (0, 1] \text{ and } Ax \leq by \\ +\infty & \text{otherwise} \end{cases}$$

The Perspective Reformulation / Relaxation

- Fundamental facts: $co\, f \leq f$, $co\, f(\, x\, ,\, y\,) = f(\, x\, ,\, y\,)$ if $y \in \{\, 0\, ,\, 1\, \}$
- This immediately leads to the Perspective Reformulation of (P)

$$(PR) \qquad \min \left\{\, y f(\, x \,/\, y\,) + cy \;:\; Ax \leq by\, ,\; y \in \{\, 0\, ,\, 1\, \} \,\right\}$$

- Why? Of course because then the Perspective Relaxation of (P)

$$(\underline{PR}) \qquad \min \left\{\, y f(\, x \,/\, y\,) + cy \;:\; Ax \leq by\, ,\; y \in [\, 0\, ,\, 1\,] \,\right\}$$

$\equiv$ continuous relaxation of (PR) has

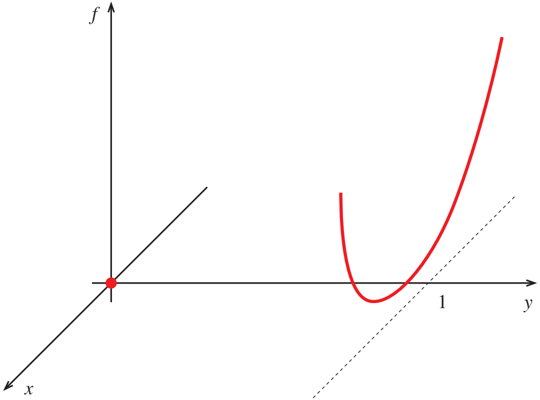
$$\nu(\underline{P}) \leq \nu(\underline{PR}) \leq \nu(PR) = \nu(P)$$

$\equiv$ better bound

- How much better? Plenty enough to matter

Where Does Such a Cool Name Comes From?

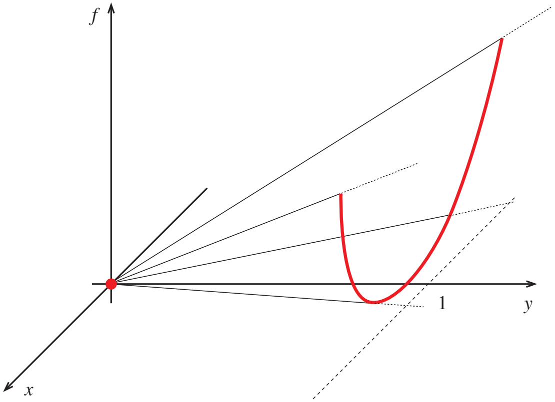
- *co f* is a section of the perspective function $\phi(x, y) = yf(x/y)$



$f(x) = ax^2 + bx$

Where Does Such a Cool Name Comes From?

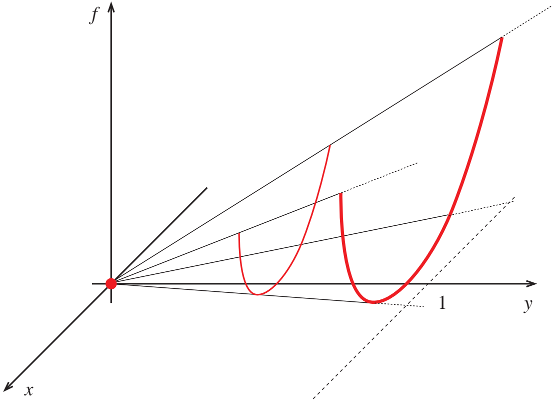
- *co f* is a section of the perspective function $\phi(x, y) = yf(x/y)$



$$f(x) = ax^2 + bx$$
$$\Downarrow$$

Where Does Such a Cool Name Comes From?

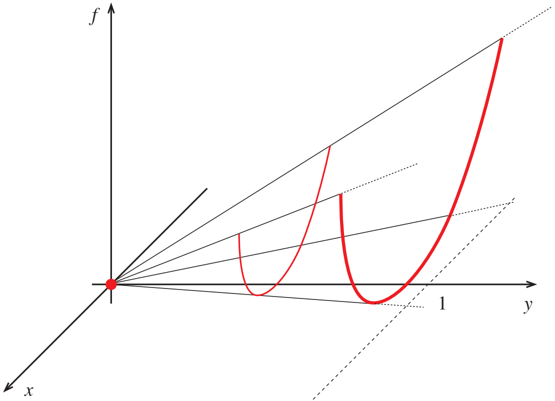
- *co f* is a section of the perspective function $\phi(x, y) = yf(x/y)$



$$f(x) = ax^2 + bx$$
$$\Downarrow$$
$$co\, f(x, y) = (a/y)x^2 + bx + cy$$
$$\geq$$
$$ax^2 + bx + cy$$

Where Does Such a Cool Name Comes From?

- $co\,f$ is a section of the perspective function $\phi(x, y) = yf(x/y)$



$$f(x) = ax^2 + bx$$
$$\Downarrow$$
$$co\,f(x, y) =$$
$$(a/y)x^2 + bx + cy$$
$$\geq$$
$$ax^2 + bx + cy$$

- Well-known tool in convex analysis, convex if f is and $y \geq 0$
- Apparently ill-defined for $y = 0$, but $co\,f(0, 0) = 0$ by continuity
- Yet, "much more nonlinear" than $f(x) + cy$:

  if (P) was a MIQP, (PR) no longer is so

Solution Methods

- How do you solve ($\underline{PR}$) (within a B&C)?
- Actually a few different ways:
  - Perspective Cuts[1,2,3,4]
  - Second Order Cone Programming reformulation[3,5,6,7]

    $v \geq x^2/y \mapsto vy \geq x^2$ (the perspective of a SOCP f is SOCP)
  - Projected Perspective Reformulation (P^2R)[8,9]
  - Approximated Projected Perspective Reformulation (AP^2R)[9,10,11]
- It does matter: orders-of magnitude differences

---

[1] F., Gentile "Perspective Cuts for a Class of Convex 0-1 Mixed Integer Programs", *Math. Prog.*, 2006

[2] F., Gentile "SDP Diagonalizations and Perspective Cuts for a Class of Nonseparable MIQP" *ORL* 2007

[3] F., Gentile "A Computational Comparison of [. . .]: SOCP vs. Cutting Planes" *ORL* 2009

[4] D'Ambrosio, F., Gentile "Strengthening the Sequential Convex MINLP Technique by Perspective Reformulations" *OL* 2019

[5] Günlük, Linderoth "Perspective reformulations of MINLPs with indicator variables", *Math. Prog.* 2010

[6] F., Galli, Scutellà "Delay-Constrained Shortest Paths: Approximation Algorithms and SOCP Models" *JOTA* 2015

[7] F., Galli, Stea "Delay-constrained Routing Problems: Accurate Scheduling Models and Admission Control" *C&OR* 2017

[8] F., Gentile, Grande, Pacifici "Projected Perspective Reformulations with Applications in Design Problems" *Op. Res.* 2010

[9] Castro, F., Gentile "Perspective Reformulations of the CTA Problem with L_2 Distances" *OR* 2014

[10] F., Furini, Gentile "Approximated Perspective Relaxations: a Project&Lift Approach" *COAP* 2016

[11] F., Furini, Gentile "Improving the Approximated Projected Perspective Reformulation by Dual Information" *ORL* 2018

Is it Useful? You Bet!

- Of course works in constraints as well
- Only to mention the applications we personally dabbled with:
  - Electrical Power Production Management (Unit Commitment)
  - Network Design with Nonlinear Delay function
  - Nonlinear (Uncapacited) Facility Location
  - Routing with Delay Constraints
  - Sensor Placement
  - Data Protection in Statistical Tables
  - Univariate Nonlinear Optimization
  - Portfolio Optimization
- Admittedly, the last one we made up,
  but nonetheless there was a lot of interest in it (this talk)

Outline

Why Nonseparable Functions?

- Motivation: ($\approx$ fake) application in Portfolio Optimization
- Mean-Variance (MV) problem with min buy-in thresholds

$$\min \left\{ x^T Q x \;\middle|\; \begin{array}{l} \sum_{i \in N} x_i = 1 \;\;,\;\; \mu^T x \geq \rho \;, \\ l_i y_i \leq x_i \leq u_i y_i \;,\; y_i \in \{\, 0 \,,\, 1 \,\} \;\; i \in N \end{array} \right\}$$

$\mu =$ expected returns, $Q =$ covariance matrix, $\rho =$ desired return

- (MV) with cardinality constraint: just add $\sum_{i=1}^{n} y_i \leq K$
- But $f(x) = x^T Q x$ nonseparable (Q not diagonal)
  $\Longrightarrow$ PR not (directly) applicable
- Yet, a Referee wanted another application, that the only one we had
- If you have an hammer but not a nail, make one

Extension to the nonseparable functions (2)

- Exceedingly dirty trick: choose $D \succeq 0$ diagonal s.t. $R = Q - D \succeq 0$

$$\min \left\{ x^T D x + x^T R x \;\middle|\; \begin{array}{l} \sum_{i \in N} x_i = 1 \;\; , \;\; \mu^T x \geq \rho \\ l_i y_i \leq x_i \leq u_i y_i \;,\; y_i \in \{\, 0 \,,\, 1 \,\} \;\; i \in N \end{array} \right\}$$

  split objective into "separable + nonseparable", use separable part

- (Partial) Perspective Reformulation

$$\min \left\{ \sum_{i \in N} d_i \frac{x_i^2}{y_i} + x^T R x \;\middle|\; \begin{array}{l} \sum_{i \in N} x_i = 1 \;\; , \;\; \mu^T x \geq \rho \\ l_i y_i \leq x_i \leq u_i y_i \;,\; y_i \in \{\, 0 \,,\, 1 \,\} \; i \in N \end{array} \right\}$$

- How to choose D?

- Simple choice $D = \lambda_{min}(Q) I$, clearly not the best ($\lambda_{min}(Q) = 0$?)

- D should be "as large as possible" $\Longrightarrow$ an optimization problem

Choosing D via SDP

- Assuming $tr(D)$ the relevant metric, the "largest" $D = diag(d)$ solves

$$\max \{ \sum_{i=1}^{n} d_i \ : \ Q - \sum_{i=1}^{n} [\, e_i e_i^T \,] d_i \succeq 0 \ , \ d \geq 0 \}$$

$$\min \{ tr(QX) \ : \ diag(X) \geq e \ , \ X \succeq 0 \}$$

  dual pair of SemiDefinite (= convex = easy) Problems

- Several efficient, open-source SDP codes

- Interesting relaxation: removing $d \geq 0$ in the primal gives

$$\min \{ tr(QX) \ : \ diag(X) = e \ , \ X \succeq 0 \}$$

  - $d^* > 0$ anyway in all our tests
  - most often faster to solve in practice by all codes
  - *constant trace* = max eigenvalue problem, specialized approaches

- Is SDP solver fast enough? Does the bound improve enough?

The instances

- 30 randomly-generated instances for each $n \in \{200, 300, 400\}$
- $\mu_i \in [0.002, 0.01]$, $l_i \in [0.075, 0.125]$, $u_i \in [0.375, 0.425]$ (uniformly)
- Q = well-known random generator [Pardalos, Rodgers '90]
- Effectiveness of Perspective Relaxation heavily impacted by dominance index

$$S = \text{average} \left\{ \frac{Q_{ii} - \sum_{j \neq i} |Q_{ij}|}{Q_{ii}} \;:\; i = 1, \ldots, n \right\}$$

- For each n, three classes of instances (10 each):
  - "$+$" instances, $S \approx 0.6$ (diagonally dominant)
  - "0" instances, $S \approx 0$ (diagonally quasi-dominant)
  - "$-$" instances, $S \approx -0.5$ (not diagonally dominant)
- Available at `http://www.di.unipi.it/optimize/Data`

Is it Worth?

	SDP			ME			no-PR		
	time	gap	r.gap	time	gap	r.gap	p.gap	gap	r.gap
200^+	164		1.14	904		6.48	0.14	45.33	85.63
200^0	161		2.14	320		6.10	0.38	51.27	84.47
200^-	1902		3.65	3306	0.02	6.69	0.24	42.09	78.88
300^+	818		4.54	2061		5.62	0.41	64.68	92.01
300^0	856		1.97	1715		6.28	0.43	59.91	87.87
300^-	1709		2.68	2797	0.05	7.04	0.53	45.11	78.77
400^+	2264		4.79	4756	0.10	6.15	1.03	61.47	89.06
400^0	4378	0.10	2.29	7421	0.16	6.53	1.18	68.68	90.03
400^-	6311	0.23	3.06	6901	0.36	6.49	1.60	65.88	88.47

- Root node gap divided by 10 with ME, by 2 more with SDP
- Effectiveness worsens as Q less dominant (not all the 400 instances solved), but much better than no-PR: none solved in 10000s
- Just a `cutcallback` and a single SDP to reformulate

Extract a Better Diagonal

- Generic MIQP case

$$\min x^T Qx + qx + cy$$
$$Ax + By \leq b \;,\; l_i y_i \leq x_i \leq u_i y_i \;,\; y_i \in \{0,1\} \quad i \in N \qquad (1)$$

- Find the diagonal giving the best (root node) bound

$$\max_d \; \min_{x,y} \; qx + cy + \textstyle\sum_{i \in N} d_i x_i^2 / y_i + x^T (Q - diag(d)) x$$
$$(1) \;,\; d \geq 0 \;,\; Q - diag(d) \succeq 0 \;,\; y \in [0,1]^n$$

- Using SDP duality can be written as a single (larger) SDP[12]

$$\min qx + cy + \textstyle\sum_{i \in N} Q_{ii} w_i + \langle Q , F \rangle$$
$$(1) \;,\; y \in [0,1]^n \;,\; diag(F) \geq 0$$
$$\begin{bmatrix} 1 & x^T \\ x & F + diag(w) \end{bmatrix} \succeq 0 \;,\; \begin{bmatrix} w_i & x_i \\ x_i & y_i \end{bmatrix} \succeq 0 \qquad i \in N$$

[12] Zheng, Sun, Li "Improving the Performance of MIQP [...]: A Semidefinite Program Approach" *IJOC* 2014

Is it Worth?

- "Large SDP" few seconds vs. split second, has to be worth it
- D^l "best" at root node only, at times $D^c = D^l/2 + D^s/2$ better (??)

	D^s time	D^c time	D^l time
200^+	1.17	0.93	1.41
200^0	1256.99	47.54	4.46
200^-	21941.53	622.25	14.99
300^+	4.60	2.36	3.18
300^0	2299.75	71.40	8.81
300^-	68043.87	1649.26	26.86
400^+	4.98	3.17	5.33
400^0	37836.53	613.01	50.60
400^-	175843.65	8095.26	52.91

- Yet, exploiting "more of Q" most definitely works

Outline

Take the Nonseparable (2×2) Bull by the Horns

- Obvious next step: 2×2 MIQP with semi-continuous variables

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

$Q \succeq 0 \equiv q_{11} > 0, \ q_{22} > 0, \ q_{11} q_{22} \geq (q_{12})^2$

- Might seem a very minor step, but actually not so
- The convex envelope of this is not easy to write
- Culprit: the damn mixed term $2q_{12}x_1x_2$
- It is was not there, we could use The PR of Alternatives

The PR of Alternatives

- Multi-piece generalization of (P)
  - $x = [\, x^k \,]_{k \in K}$, $f(x) = \sum_{k \in K} f^k(x^k)$
  - each x^k either 0 or $x^k \in \mathcal{P}^k = \{\, x^k \;:\; A^k x^k \leq b^k \,\}$
  - at most one of the $x^k \neq 0$

$$f(x) = \begin{cases} f^k(x^k) + c^k & \text{if } x^k \in \mathcal{P}^k \text{ and } x^h = 0 \;\forall\, h \in K \setminus \{\, k \,\} \\ 0 & \text{if } x = 0 \\ +\infty & \text{otherwise} \end{cases}$$

- Obvious application of the convex hull of disjunctions:

$$co\, f(x) \;=\; \begin{cases} \min & \sum_{k \in K} \theta^k f^k(\, x^k \,/\, \theta^k \,) \\ & \sum_{k \in K} \theta^k \leq 1 \\ & A^k x^k \leq b^k \theta^k \;,\; \theta_k \geq 0 \quad k \in K \end{cases}$$

basically the sum of individual convex envelopes

The PR of Alternatives to the Rescue

- Somewhat awkward reformulation:

$$\begin{aligned} \min\ & q_{11}(x_1^1)^2 + q_{22}(x_2^2)^2 + q_{11}(x_1^{12})^2 + 2q_{12}x_1^{12}x_2^{12} + q_{22}(x_2^{12})^2 \\ & x_i = x_i^i + x_i^{12} \quad , \quad y_i = y^i + y^{12} \qquad i = 1, 2 \\ & l_i y^i \leq x_i^i \leq u_i y^i \ , \ l_i y^{12} \leq x_i^{12} \leq u_i y^{12} \qquad i = 1, 2 \\ & y^1 + y^2 + y^{12} \leq 1 \ , \ y^1 \ , \ y^2 \ , \ y^{12} \in \{ 0 , 1 \} \end{aligned}$$

"get rid by the mixed term" by creating its "private copy" of the x

- Why? Because PR of alternatives then gives

$$(PR_2) \quad \begin{cases} \min & q_{11}(x_1^1)^2/y^1 + q_{22}(x_2^2)^2/y^2 + \\ & \quad \left[q_{11}(x_1^{12})^2 + 2q_{12}x_1^{12}x_2^{12} + q_{22}(x_2^{12})^2\right]/y^{12} \\ & \dots \end{cases}$$

and can project away half of the extra variables via equalities

- But projecting on the original variable space nontrivial[13]

[13] Jeon, Linderoth, Miller "Quadratic cone cutting surfaces for quadratic programs with on–off constraints" *Disc. Opt.* 2017

$2 \times 2 \mapsto n \times n$ in One Fell Swoop

- Nice idea: decompose Q as the sum of 2×2 PSD matrices
  (could never work with $k = 1$, but can work as soon as $k > 1$)
- P = set of all $O(n^2)$ undirected pairs $\{\, i \,,\, j \,\} \subset N \times N$ $(i \neq j)$
  $n \times n$ matrices $Q^{ij} \succeq 0$ with 2×2 support such that
  $$Q = \sum_{\{\, i \,,\, j \}\in P} Q^{ij}$$

Example

$$Q = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

- Q^{ij} such that $Q^{ij}_{hk} = 0$ if $\{\, h \,,\, k \,\} \not\subseteq \{\, i \,,\, j \,\} \equiv$
  $Q^{ij} = E^p \Pi^p (E^p)^T$ where $E^p = [\, e_i \,,\, e_j \,] \in \mathbb{R}^{n\times 2}$, $\Pi^p \succeq 0 \in \mathbb{R}^{2\times 2}$
  (rank-two matrix with 2×2 semidefinite core)

When is Q 2×2-Decomposable?

- Q Weakly Diagonally Dominant (WDD) $\equiv |Q_{ii}| \geq \sum_{j \neq i} |Q_{ij}| \quad i \in N$

Lemma

Q is (WDD) $\Longrightarrow$ Q has 2×2D

- Q Weakly Scaled Diagonally Dominant (WSDD) $\equiv$ $\exists\ d > 0$ s.t. $diag(d)\, Q\, diag(d)$ is WDD

Theorem

Q has a 2×2D $\iff$ Q is (WSDD)

- A 2×2D is available by closed formula by computing an eigenpair (λ, x) s.t. $\lambda = \rho(\,|\, I - diag(Q)^{-\frac{1}{2}}\, Q\, diag(Q)^{-\frac{1}{2}}\,|\,) \leq 1$ and $x > 0$

$$Q^{ij}_{ii} \;=\; \frac{\lambda |Q_{ij}| \sqrt{Q_{ii}}}{\lambda \sqrt{Q_{jj}}} x_i^{-1} x_j + \frac{Q_{ii}(1-\lambda)}{n-1}$$

- Actually a more general parametric formula, condition for unicity

What if Q is not 2×2-Decomposable?

- Extract "as much as possible" by a "simple" SDP

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

- Approximate 2×2 Perspective Reformulation (2×2PR)

$$\min x^T \Phi x + q^T x + c^T y +$$

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

$$Ax + By \leq b$$

$$x_i = x_i^{p,i} + x_i^{p,p} \;\;,\;\; y_i = y^{p,i} + y^{p,p} \qquad p \in P \;,\; i \in p \qquad (2)$$

$$l_i y^{p,i} \leq x_i^{p,i} \leq u_i y^{p,i} \qquad p \in P \;,\; i \in p \qquad (3)$$

$$l_i y^{p,p} \leq x_i^{p,p} \leq u_i y^{p,p} \qquad p \in P \;,\; i \in p \qquad (4)$$

$$y^{p,p} + y^{p,i} + y^{p,j} \leq 1 \qquad p \in P \qquad (5)$$

$$y^{p,i} \in \{0, 1\} \;,\; y^{p,p} \in \{0, 1\} \qquad p \in P \;,\; i \in p \qquad (6)$$

- (2×2PR) rather more expensive to solve (by standard SOCP)

Extract a Better 2×2D

- Extract the best possible 2×2D: usual max / min
- SDP duality leads to a large SDP

$$\min\; q^T x + c^T y + \langle\, Q \,,\, F \,\rangle$$

$$Ax + By \leq b \;\;,\;\; (2)\text{–}(5) \;\;,\;\; \begin{bmatrix} 1 & x^T \\ x & F \end{bmatrix} \succeq 0$$

$$D^p F D^p \succeq E^p \left(\begin{bmatrix} w_i^p & 0 \\ 0 & w_j^p \end{bmatrix} + W^p \right) (E^p)^T \qquad p \in P$$

$$\begin{bmatrix} W^p & x^{p,p} \\ (x^{p,p})^T & y^{p,p} \end{bmatrix} \succeq 0 \qquad p \in P$$

$$\begin{bmatrix} w_i^p & x_i^{p,i} \\ x_i^{p,i} & y^{p,i} \end{bmatrix} \succeq 0 \qquad p \in P \;\;,\;\; i \in p$$

(last ones actually SOCP constraints)

- Positively huge! 1e+5 seconds for $n = 50$ (by standard SDP)

Cheaper alternatives to find a "good" 2×2D

- We want to write $Q = \Phi + X$, where X is 2×2D and $\Phi \succeq 0$ "small"
- We know how to quickly test if X has a 2×2D
- We know (more than one) D s.t. $Q - D \succeq 0$
- We assume $\Phi(\varepsilon) = \varepsilon(Q - D) \equiv X(\varepsilon) = Q - \Phi(\varepsilon) = (1 - \varepsilon)Q + \varepsilon D$
- If $Q = X(0)$ has a 2×2D, stop: $\Phi(0) = 0$
- $X(1) = D$ surely has a 2×2D
- Find the smallest ε s.t. $X(\varepsilon)$ has a 2×2D
- Binary search with eigenvalue computation inside, very quick
- Depends on the D one starts from

Outline

The instances

- $n \in \{ 25 , 50 \}$ because we could go no further
- Suddenly realised $Q \geq 0$ in all our cases, but did not need to be
- "p" = "+", "z" = "0", "n" = "−" instances
- "o", "y", "m" instances produced starting "p", "z", "n": changing sign of all the off-diagonals, correct diagonal if $\not\succeq 0$
- This makes Q look like/be a M-matrix[14]
- CXV + SeDuMi for SDPs, Cplex for MI-SOCPs
- D_s, D_l: 1×1 with D from "small SDP" and "large SDP"
- $2\times2_s$, $2\times2_l$: 2×2 with 2×2D from "small SDP" and "large SDP"
- $2\times2_{h,s}$, $2\times2_{h,l}$: 2×2 with 2×2D from heuristic starting by D_s, D_l
- Report root node gaps of PR (forget about running times . . .)

[14] Atamtuřk, Goómez "Strong Formulations for Quadratic Opt. with M-Matrices and Semi-Continuous Variables" *RR* 2018

Some Results, $n = 25$

	D_s	D_l	$2{\times}2_s$	$2{\times}2_{h,s}$	$2{\times}2_{h,l}$	$2{\times}2_l$
25-p-a	2.28	2.21	2.21	2.21	2.21	2.21
25-p-i	0.71	0.56	2.33	2.33	2.33	0.55
25-o-c	8.23	5.09	2.48	2.48	2.48	2.47
25-o-i	14.93	13.65	6.58	6.58	6.58	6.58
25-z-c	3.06	2.44	15.27	15.83	15.72	1.64
25-z-e	0.02	0.00	0.07	0.07	0.08	0.00
25-y-e	2.14	2.10	0.32	0.32	0.32	0.21
25-y-g	24.57	24.57	9.14	9.14	9.14	8.71
25-n-a	2.79	1.78	11.97	12.10	11.24	0.27
25-n-c	3.04	2.04	12.12	13.04	12.66	1.42
25-m-c	99.55	99.55	98.59	98.59	98.59	98.59
25-m-g	5.32	5.32	4.79	4.79	4.79	4.66
25-m-i	100	100	100	100	100	100

Some Results, $n = 50$

	D_s	D_l	$2{\times}2_s$	$2{\times}2_{h,s}$	$2{\times}2_{h,l}$	$2{\times}2_l$
50-p-a	1.47	1.23	1.20	1.20	1.20	1.20
50-p-e	0.89	0.33	1.40	1.40	1.40	0.26
50-o-c	15.71	9.32	2.44	2.44	2.44	2.44
50-o-g	19.84	15.90	6.32	6.32	6.32	6.31
50-z-e	2.86	1.16	10.97	11.45	11.30	0.51
50-z-i	2.92	1.79	3.65	4.72	4.83	1.36
50-y-a	8.80	8.80	0.16	0.16	0.16	0.14
50-y-i	94.62	94.62	73.27	73.27	73.27	73.26
50-n-e	3.12	1.54	9.23	11.90	10.85	0.36
50-n-i	3.92	1.98	6.75	10.10	9.07	0.57
50-m-a	59.97	59.97	15.47	15.47	15.47	15.38
50-m-c	100	100	100	100	100	100
50-m-i	98.51	98.51	90.22	90.22	90.22	90.22

So, is it Worth?

- Not using standard means
- A bad 2×2D is worse than a good D, even if obtained by it
- Both SDPs and SOCP for 2×2D PR (extremely) more costly
- Cheap heuristics for 2×2D sometimes good, sometimes horrible
- Improvement w.r.t. diagonal ("standard" or "large")
  hugely dependent on Q: sometimes huge ($Q_{ij} < 0$), sometimes null
- Surely something more to say theoretically
- Yet the approach looks interesting to us (sparse case ...)

Outline

$k \times k$ Convex Envelope

- Example: 3×3 case, three indices $t = \{ 1 , 2 , 3 \}$
- $2^3 - 1 = 7$ *configurations* in $C(t) = 2^t \setminus \emptyset$
  $= \{ \{ 1 \} , \{ 2 \} , \{ 3 \} , \{ 1 , 2 \} , \{ 1 , 3 \} , \{ 2 , 3 \} , \{ 1 , 2 , 3 \} \}$
- For $c \in C(t)$, $Q^c = Q$ restricted to these indices, $x^c = [x_i^c]_{i \in c}$

$$
(PR_3) \quad \left\{ \begin{array}{lll}
\min & \sum_{c \in C(t)} \left[(x^c)^T Q^c x^c \right] / y^c & \\
\text{s.t.} & x_i = \sum_{c \in C(t) \,:\, i \in c} x_i^c & i \in t \\
& y_i = \sum_{c \in C(t) \,:\, i \in c} y^c & i \in t \\
& l_i y^c \leq x_i^c \leq u_i y^c & c \in C(t) \;,\; i \in c \\
& \sum_{c \in C(t)} y^c \leq 1 & \\
& y^c \in \{0, 1\} & c \in C(t)
\end{array} \right.
$$

- Can clearly go to any $k \times k$ at the only cost of 2^k configurations for each of the $O(n^k)$ subsets of indices
- Does it look promising? You tell me ...

$k \times k$ PR Requires a $k \times k$D

- Set $T = \{\, (\, i\, ,\, j\, ,\, k\,) \in N \times N \times N \;:\; i < j < k \,\}$ of all possible triples, $E^t = [\, e_i\, ,\, e_j\, ,\, e_k\,]$, $\Gamma^t \succeq 0 \in \mathbb{R}^{3\times 3}$ for all $t \in T$
- Approximate 3×3D $Q = \Phi + \sum_{t\in T} E^t \Gamma^t (E^t)^T$ gives the 3×3PR

$$
\begin{aligned}
\min\ & x^T \Phi x + q^T x + c^T y + \\
& \textstyle\sum_{t\in T} \sum_{c\in C(t)} (x^{t,c})^T (\Gamma^t)^c x^{t,c} / y^{t,c} \\
\text{s.t. } & Ax + By \le b \\
& x_i = \sum_{c\in C(t)\,:\, i\in c} x_i^{t,c} \;,\; y_i = \sum_{c\in C(t)\,:\, i\in c} y^{t,c} && t\in T \;,\; i \in N \quad (7) \\
& l_i y^{t,c} \le x_i^{t,c} \le u_i y^{t,c} && t\in T\,,\, c\in C(t)\,,\, i\in c \quad (8) \\
& \textstyle\sum_{c\in C(t)} y^{t,c} \le 1 && t\in T \quad (9) \\
& y^{t,c} \in \{0,1\} && t\in T \;,\; c \in C(t) \quad (10)
\end{aligned}
$$

- Algebraic characterization of $k \times k$D? Heuristics??

Finding a $k \times k$D of Q

- "Small" SDP: $E^t = [\, e_i \,,\, e_j \,,\, e_k \,]$, $\Gamma^t \succeq 0 \in \mathbb{R}^{3\times 3}$ for all $t \in T$

$$\min \Big\{ \|\, \Phi \,\|^2 \;:\; Q = \Phi + \textstyle\sum_{t \in T} E^t \Gamma^t (E^t)^T \;\;,\;\; \Gamma^t \succeq 0 \quad t \in T \;\;,\;\; \Phi \succeq 0 \Big\}$$

- You can play the same "max / min + SDP duality" game to get

$$
\begin{array}{ll}
\min q^T x + c^T y + \sum_{p \in P} 2 Q_p f_p + \sum_{i \in N} Q_{ii} f_i & \\
\text{s.t. } Ax + By \leq b \;\;,\;\; (7)\text{–}(10) \;\;,\;\; y \in [0,1]^n & \\
\left[\begin{array}{cc} 1 & x^T \\ x & \sum_{(i,j) \in P} O^{ij} f_{ij} + \sum_{i \in N} D^i f_i \end{array} \right] \succeq 0 & \\
\sum_{(i,j) \subset t} O^{t,ij} f_{ij} + \sum_{i \in t} D^{t,i} f_i \succeq \sum_{c \in C(t)} \bar{W}^{t,c} & t \in T \\
\left[\begin{array}{cc} W^{t,c} & x^{t,c} \\ (x^{t,c})^T & y^{t,c} \end{array} \right] \succeq 0 & t \in T \;,\; c \in C(t)
\end{array}
$$

($\bar{W}^{t,c} = W^{t,c}$ extended in $\mathbb{R}^{n \times n}$, a few of the latter SOCP)

- Perhaps better say you theoretically could ...

Conclusions and Several Open Questions

- Perspective Reformulation is very simple, yet surprisingly effective in the separable case
- Extensions to the nonseparable case by diagonal extraction reasonably simple (one SDP) and surprisingly effective as well
- Nice general idea: extraction of 2×2 SDP matrices (may have further applications?)
- Full theoretical characterization:
  - matrices with exact 2×2 decomposition
  - SDP formulation for approximated 2×2 decomposition
- Still a lot of work to be competitive w.r.t. diagonal case (will we ever get there?)
- More theory for $k > 2$? Possible practical value?