Perspective Reformulations Beyond the Separable Case

Antonio Frangioni
Dipartimento di Informatica, Università di Pisa
with C. Gentile, J. Hungerford

One-day symposium on Integer Programming and Algorithms
Ecole des Ponts Paristech
November 9, 2019
1. The Perspective Relaxation
2. “Quick & Dirty” Extension to Nonseparable Functions
3. Firmly Beyond Separability
4. Computational Results
5. Going Even Further
6. Advertisement
7. Conclusions
The Target Structure

- MINLP with convex-cost semicontinuous (vector) variables

\[
\begin{align*}
\min_{\mathcal{O}} & \quad g(z) + \sum_{i \in \mathcal{N}} f_i(x_i) + c_i y_i \\
\text{subject to} & \quad A_i x_i \leq b_i y_i \\
& \quad y \in \{0, 1\}^n, \quad x \in \mathbb{R}^m, \quad (x, y, z) \in \mathcal{O}
\end{align*}
\]

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

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

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

- All hinges on \( \nu(P) \) being a “good bound”, but it may not be
Exploiting the Structure

- To improve the bound we must exploit the structure

- Single block problem

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

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

- \( co f = \text{convex envelope} = \text{“best” lower convex approximation} \)

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

\[
(PR) \quad \min \left\{ yf(x/y) + cy : Ax \leq by, \ y \in \{0, 1\} \right\}
\]

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

\[
(PR) \quad \min \left\{ yf(x/y) + cy : Ax \leq by, \ y \in [0, 1] \right\}
\]

\[\equiv \text{continuous relaxation of } (PR) \text{ has} \]

\[\nu(P) \leq \nu(PR) \leq \nu(PR) = \nu(P)\]

\[\equiv \text{better bound}\]

- How much better? Plenty enough to matter
Where Does Such a Cool Name Come From?

- \( \text{co } f \) 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 $\phi(x, y) = yf(x/y)$

$$f(x) = ax^2 + bx$$
Where Does Such a Cool Name Come From?

- \( \text{co } f \) is a section of the perspective function \( \phi(x, y) = y f(x/y) \)

\[
f(x) = ax^2 + bx
\]

\[
\downarrow
\]

\[
\text{co } f(x, y) = (a/y)x^2 + bx + cy
\]

\[
\lor
\]

\[
a x^2 + bx + cy
\]
Where Does Such a Cool Name Comes From?

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

Well-known tool in convex analysis, convex if \( f \) is and \( y \geq 0 \)

Apparently ill-defined for \( y = 0 \), but \( \text{co } f(0, 0) = 0 \) 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 $(PR)$ (within a B&C)?

- Actually a few different ways:
  - Perspective Cuts$^{1,2,3,4}$
  - Second Order Cone Programming reformulation$^{3,5,6,7}$
    \[ v \geq x^2/y \iff vy \geq x^2 \] (the perspective of a SOCP $f$ is SOCP)
  - Projected Perspective Reformulation (P$^2$R)$^{8,9}$
  - Approximated Projected Perspective Reformulation (AP$^2$R)$^{9,10,11}$

- It does matter: orders-of-magnitude differences

---

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
7. F., Galli, Stea “Delay-constrained Routing Problems: Accurate Scheduling Models and Admission Control” *C&OR* 2017
9. Castro, F., Gentile “Perspective Reformulations of the CTA Problem with L$_2$ Distances” *OR* 2014
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 (Uncapacitated) 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

1. The Perspective Relaxation
2. “Quick & Dirty” Extension to Nonseparable Functions
3. Firmly Beyond Separability
4. Computational Results
5. Going Even Further
6. Advertisement
7. Conclusions
Why Nonseparable Functions?

- Motivation: (≈ fake) application in Portfolio Optimization
- Mean-Varience (MV) problem with min buy-in thresholds

\[
\min \left\{ x^T Q x \mid \sum_{i \in N} x_i = 1 , \quad \mu^T x \geq \rho , \quad l_i y_i \leq x_i \leq u_i y_i , \quad y_i \in \{0, 1\} \quad i \in N \right\}
\]

\(\mu = \) 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)
  \[\implies \text{PR not (directly) applicable}\]

- 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

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

\[
\min \left\{ x^T Dx + x^T Rx \mid \sum_{i \in N} x_i = 1, \quad \mu^T x \geq \rho \\
\quad l_i y_i \leq x_i \leq u_i y_i, \quad y_i \in \{0, 1\} \quad i \in N \right\}
\]

split objective into “separable + nonseparable”, use separable part

- (Partial) Perspective Reformulation

\[
\min \left\{ \sum_{i \in N} d_i \frac{x_i^2}{y_i} + x^T Rx \mid \sum_{i \in N} x_i = 1, \quad \mu^T x \geq \rho \\
\quad l_i y_i \leq x_i \leq u_i y_i, \quad y_i \in \{0, 1\} \quad i \in N \right\}
\]

- How to choose $D$?

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

- $D$ should be “as large as possible” $\implies$ an optimization problem
Choosing \( D \) via SDP

- Assuming \( tr(D) \) the relevant metric, the “largest” \( D = \text{diag}(d) \) solves
  \[
  \max \left\{ \sum_{i=1}^{n} d_i : Q - \sum_{i=1}^{n} [e_i e_i^T]d_i \succeq 0 , \ d \geq 0 \right\}
  \]
  \[
  \min \left\{ \text{tr}(QX) : \text{diag}(X) \succeq e , \ X \succeq 0 \right\}
  \]
  dual pair of SemiDefinite (= convex = easy) Problems

- Several efficient, open-source SDP codes

- Interesting relaxation: removing \( d \geq 0 \) in the primal gives
  \[
  \min \left\{ \text{tr}(QX) : \text{diag}(X) = e , \ X \succeq 0 \right\}
  \]
  \( d^* > 0 \) 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?

<table>
<thead>
<tr>
<th></th>
<th>SDP</th>
<th></th>
<th>ME</th>
<th></th>
<th>no-PR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>time</td>
<td>gap</td>
<td>r.gap</td>
<td>time</td>
<td>gap</td>
<td>r.gap</td>
</tr>
<tr>
<td>200⁺</td>
<td>164</td>
<td>1.14</td>
<td></td>
<td>904</td>
<td>6.48</td>
<td></td>
</tr>
<tr>
<td>200₀</td>
<td>161</td>
<td>2.14</td>
<td></td>
<td>320</td>
<td>6.10</td>
<td></td>
</tr>
<tr>
<td>200⁻</td>
<td>1902</td>
<td>3.65</td>
<td></td>
<td>3306</td>
<td>0.02</td>
<td>6.69</td>
</tr>
<tr>
<td>300⁺</td>
<td>818</td>
<td>4.54</td>
<td></td>
<td>2061</td>
<td>5.62</td>
<td></td>
</tr>
<tr>
<td>300₀</td>
<td>856</td>
<td>1.97</td>
<td></td>
<td>1715</td>
<td>6.28</td>
<td></td>
</tr>
<tr>
<td>300⁻</td>
<td>1709</td>
<td>2.68</td>
<td></td>
<td>2797</td>
<td>0.05</td>
<td>7.04</td>
</tr>
<tr>
<td>400⁺</td>
<td>2264</td>
<td>4.79</td>
<td></td>
<td>4756</td>
<td>0.10</td>
<td>6.15</td>
</tr>
<tr>
<td>400₀</td>
<td>4378</td>
<td>0.10</td>
<td>2.29</td>
<td>7421</td>
<td>0.16</td>
<td>6.53</td>
</tr>
<tr>
<td>400⁻</td>
<td>6311</td>
<td>0.23</td>
<td>3.06</td>
<td>6901</td>
<td>0.36</td>
<td>6.49</td>
</tr>
</tbody>
</table>

- **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

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

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

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

- Using SDP duality can be written as a single (larger) SDP\(^{12}\)

\[
\min qx + cy + \sum_{i \in N} Q_{ii} w_i + \langle Q, F \rangle \\
(1), \ y \in [0, 1]^n, \ \text{diag}(F) \succeq 0
\]

\[
\begin{bmatrix}
1 & x^T \\
x & F + \text{diag}(w)
\end{bmatrix} \succeq 0, \quad \begin{bmatrix}
w_i & x_i \\
x_i & y_i
\end{bmatrix} \succeq 0 \quad i \in N
\]

\(^{12}\)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 (??)

<table>
<thead>
<tr>
<th></th>
<th>$D^s$ time</th>
<th>$D^c$ time</th>
<th>$D^l$ time</th>
</tr>
</thead>
<tbody>
<tr>
<td>200+</td>
<td>1.17</td>
<td>0.93</td>
<td>1.41</td>
</tr>
<tr>
<td>200^0</td>
<td>1256.99</td>
<td>47.54</td>
<td>4.46</td>
</tr>
<tr>
<td>200^-</td>
<td>21941.53</td>
<td>622.25</td>
<td>14.99</td>
</tr>
<tr>
<td>300+</td>
<td>4.60</td>
<td>2.36</td>
<td>3.18</td>
</tr>
<tr>
<td>300^0</td>
<td>2299.75</td>
<td>71.40</td>
<td>8.81</td>
</tr>
<tr>
<td>300^-</td>
<td>68043.87</td>
<td>1649.26</td>
<td>26.86</td>
</tr>
<tr>
<td>400+</td>
<td>4.98</td>
<td>3.17</td>
<td>5.33</td>
</tr>
<tr>
<td>400^0</td>
<td>37836.53</td>
<td>613.01</td>
<td>50.60</td>
</tr>
<tr>
<td>400^-</td>
<td>175843.65</td>
<td>8095.26</td>
<td>52.91</td>
</tr>
</tbody>
</table>

- Yet, exploiting “more of $Q$” most definitely works
1. The Perspective Relaxation
2. “Quick & Dirty” Extension to Nonseparable Functions
3. Firmly Beyond Separability
4. Computational Results
5. Going Even Further
6. Advertisement
7. Conclusions
Obvious next step: $2 \times 2$ MIQP with semi-continuous variables

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

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

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 \((P)\)
  
  \[ x = [x^k]_{k \in K}, \quad f(x) = \sum_{k \in K} f^k(x^k) \]
  
  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:

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

  basically the sum of individual convex envelopes

- Little nice result useful in more than one way
The PR of Alternatives to the Rescue

Somewhat awkward reformulation:

$$\begin{align*}
\min & \quad q_{11}(x_1^1)^2 + q_{22}(x_2^2)^2 + q_{11}(x_1^{12})^2 + 2q_{12}x_1^{12}x_2^{12} + q_{22}(x_2^{12})^2 \\
x_i &= x_i^i + x_i^{12}, \quad y_i = y_i^i + y_i^{12} \quad i = 1, 2 \\
l_iy_i^i \leq x_i^i \leq u_iy_i^i, \quad l_iy_i^{12} \leq x_i^{12} \leq u_iy_i^{12} \quad i = 1, 2 \\
y_1^1 + y_2^2 + y_1^{12} \leq 1, \quad y_1^1, y_2^2, y_1^{12} \in \{0, 1\}
\end{align*}$$

“get rid by the mixed term” by creating its “private copy” of the $x$

Why? Because PR of alternatives then gives

$$(PR_2) \begin{cases} 
\min & q_{11}(x_1^1)^2/y_1^1 + q_{22}(x_2^2)^2/y_2^2 + \\
& \left[ q_{11}(x_1^{12})^2 + 2q_{12}x_1^{12}x_2^{12} + q_{22}(x_2^{12})^2 \right] / y_1^{12} \\
\end{cases}$$

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
Nice idea: decompose $Q$ as the sum of $2 \times 2$ 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 $2 \times 2$ 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 $\{ h, k \} \not\subseteq \{ i, j \}$

$$Q^{ij} = E^p \Pi^p (E^p)^T \text{ where } E^p = [e_i, e_j] \in \mathbb{R}^{n \times 2}, \, \Pi^p \succeq 0 \in \mathbb{R}^{2 \times 2}$$

(rank-two matrix with $2 \times 2$ 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) $\implies Q$ has 2×2D

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

**Theorem**

$Q$ has a 2×2D $\iff Q$ is (WSDD)

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

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

- Actually a more general parametric formula, condition for unicity
Not really a new result

- Turned out it was known already\textsuperscript{14}

- Fundamental piece of the proof independently found in completely different application\textsuperscript{15}

- Independently found a completely application in optimization\textsuperscript{16}

- Our own result slightly more refined in some aspects:
  - explicit parametrized formulæ
  - characterization of unicity and existence low-rank decomposition

- Open question: algebraic characterization of $k \times kD$ for $k > 2$

- Only (weak) bounds are known\textsuperscript{14}

---

\textsuperscript{14} Boman, Chen, Parekh, Toledo “On factor width and symmetric H -matrices” \textit{Linear Algebra and its Applications}, 2005

\textsuperscript{15} Ruozzi, Tatikonda “Message-passing algorithms for quadratic minimization” \textit{Journal of Machine Learning}, 2013

\textsuperscript{16} Ahmadi, Majumdar “DSOS and SDSOS Optimization: More Tractable […]” \textit{SIAM J. Appl. Algebra Geometry}, 2019
What if $Q$ is not $2 \times 2$-Decomposable?

- Extract “as much as possible” by a “simple” SDP

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

- **Approximate $2 \times 2$ Perspective Reformulation ($2 \times 2$PR)**

\[
\min x^T \Phi x + q^T x + c^T y + \\
\quad \sum_{p \in P} \left[ \Pi_{11}^p (x_{i}^p, i)^2 / y_{p,i} + \Pi_{22}^p (x_{j}^p, j)^2 / y_{p,j} + (x_{p,p})^T \Pi_{p,p} x_{p,p} / y_{p,p} \right] \\
A x + B y \leq b \\
x_i = x_{i}^p, i + x_{i}^p, p \quad , \quad y_i = y_{p,i} + y_{p,p} \quad p \in P , \quad i \in p \quad (2) \\
l_i y_{p,i} \leq x_{i}^p, i \leq u_i y_{p,i} \quad p \in P , \quad i \in p \quad (3) \\
l_i y_{p,p} \leq x_{i}^p, p \leq u_i y_{p,p} \quad p \in P , \quad i \in p \quad (4) \\
y_{p,p} + y_{p,i} + y_{p,j} \leq 1 \quad p \in P \quad (5) \\
y_{p,i} \in \{0, 1\} \quad , \quad y_{p,p} \in \{0, 1\} \quad p \in P , \quad i \in p \quad (6)
\]

- ($2 \times 2$PR) rather more expensive to solve (by standard SOCP)
Extract a Better $2 \times 2D$

- Extract the best possible $2 \times 2D$: usual max / min

- SDP duality leads to a large SDP

\[
\min q^T x + c^T y + \langle Q, F \rangle
\]

\[
Ax + By \leq b, \quad (2)-(5), \quad \begin{bmatrix} 1 & x^T \\ x & F \end{bmatrix} \succeq 0
\]

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

\[
\begin{bmatrix} W^p & x^{p,p} \\ (x^{p,p})^T & y^{p,p} \end{bmatrix} \succeq 0
\]

\[
\begin{bmatrix} w_i^p & x_i^{p,i} \\ x_i^{p,i} & y^{p,i} \end{bmatrix} \succeq 0
\]

(last ones actually SOCP constraints)

- Positively huge! $1e+5$ seconds for $n = 50$ (by standard SDP)
Cheaper alternatives to find a “good” $2 \times 2D$

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

1. The Perspective Relaxation
2. “Quick & Dirty” Extension to Nonseparable Functions
3. Firmly Beyond Separability
4. Computational Results
5. Going Even Further
6. Advertisement
7. Conclusions
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 $\not\preceq 0$
- This makes $Q$ look like/be a M-matrix
- $\text{CXV} + \text{SeDuMi}$ for SDPs, Cplex for MI-SOCPs
- $D_s, D_l$: $1 \times 1$ with $D$ from “small SDP” and “large SDP”
- $2 \times 2_s, 2 \times 2_l$: $2 \times 2$ with $2 \times 2D$ from “small SDP” and “large SDP”
- $2 \times 2_{h,s}, 2 \times 2_{h,l}$: $2 \times 2$ with $2 \times 2D$ from heuristic starting by $D_s, D_l$
- Report root node gaps of PR (forget about running times . . . )

---

17 Atamtürk, Goómez “Strong Formulations for Quadratic Opt. with M-Matrices and Semi-Continuous Variables” RR 2018
Some Results, $n = 25$

<table>
<thead>
<tr>
<th></th>
<th>$D_s$</th>
<th>$D_l$</th>
<th>$2 \times 2_s$</th>
<th>$2 \times 2_{h,s}$</th>
<th>$2 \times 2_{h,l}$</th>
<th>$2 \times 2_l$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25-p-a</td>
<td>2.28</td>
<td>2.21</td>
<td>2.21</td>
<td>2.21</td>
<td>2.21</td>
<td>2.21</td>
</tr>
<tr>
<td>25-p-i</td>
<td>0.71</td>
<td>0.56</td>
<td>2.33</td>
<td>2.33</td>
<td>2.33</td>
<td>0.55</td>
</tr>
<tr>
<td>25-o-c</td>
<td>8.23</td>
<td>5.09</td>
<td>2.48</td>
<td>2.48</td>
<td>2.48</td>
<td>2.47</td>
</tr>
<tr>
<td>25-z-c</td>
<td>3.06</td>
<td>2.44</td>
<td>15.27</td>
<td>15.83</td>
<td>15.72</td>
<td>1.64</td>
</tr>
<tr>
<td>25-z-e</td>
<td>0.02</td>
<td>0.00</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
<td>0.00</td>
</tr>
<tr>
<td>25-y-e</td>
<td>2.14</td>
<td>2.10</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.21</td>
</tr>
<tr>
<td>25-n-a</td>
<td>2.79</td>
<td>1.78</td>
<td>11.97</td>
<td>12.10</td>
<td>11.24</td>
<td>0.27</td>
</tr>
<tr>
<td>25-n-c</td>
<td>3.04</td>
<td>2.04</td>
<td>12.12</td>
<td>13.04</td>
<td>12.66</td>
<td>1.42</td>
</tr>
<tr>
<td>25-m-c</td>
<td>99.55</td>
<td>99.55</td>
<td>98.59</td>
<td>98.59</td>
<td>98.59</td>
<td>98.59</td>
</tr>
<tr>
<td>25-m-g</td>
<td>5.32</td>
<td>5.32</td>
<td>4.79</td>
<td>4.79</td>
<td>4.79</td>
<td>4.66</td>
</tr>
<tr>
<td>25-m-i</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Some Results, $n = 50$

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

- **Not** using standard means

- A bad $2 \times 2\text{D}$ is worse than a good $D$, even if obtained by it

- Both SDPs and SOCP for $2 \times 2\text{D}$ PR (extremely) more costly

- Cheap heuristics for $2 \times 2\text{D}$ 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 . . . )
\(k \times k\) 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, \(x^c = [x^c_i]_{i \in c}\)

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

- 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 . . .
Set \( T = \{ (i, j, k) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N} : i < j < k \} \) 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 \( 3 \times 3D \) \( Q = \Phi + \sum_{t \in T} E^t \Gamma^t (E^t)^T \) gives the \( 3 \times 3 \) \( PR \)

\[
\begin{align*}
\min \ x^T \Phi x + q^T x + c^T y + \\
\sum_{t \in T} \sum_{c \in C(t)} \left( x^{t,c} \right)^T (\Gamma^t)^c x^{t,c} / y^{t,c}
\end{align*}
\]

s.t. \( Ax + By \leq b \)

\[
\begin{align*}
x_i &= \sum_{c \in C(t) : i \in c} x^{t,c} , \quad y_i &= \sum_{c \in C(t) : i \in c} y^{t,c} & t \in T , \ i \in \mathbb{N} & (7) \\
l_i y^{t,c} & \leq x^{t,c} \leq u_i y^{t,c} & t \in T , \ c \in C(t) , \ i \in c & (8) \\
\sum_{c \in C(t)} y^{t,c} & \leq 1 & t \in T & (9) \\
y^{t,c} & \in \{0, 1\} & t \in T , \ c \in C(t) & (10)
\end{align*}
\]

Algebraic characterization of \( k \times kD \)? Heuristics??
Finding a $k \times kD$ of $Q$

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

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

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

$$\min q^T x + c^T y + \sum_{p \in P} 2Q_p f_p + \sum_{i \in N} Q_{ii} f_i$$

s.t. $Ax + By \leq b \ , \ (7)-(10) \ , \ y \in [0,1]^n$

$$\left[ \begin{array}{c} 1 \\ x \end{array} \right] \begin{bmatrix} x^T \\ \sum_{(i,j) \in P} O^{ij} f_{ij} + \sum_{i \in N} D^i f_i \end{bmatrix} \succeq 0$$

$$\sum_{(i,j) \subset t} O^{t,ij} f_{ij} + \sum_{i \in t} D^{t,i} f_i \succeq \sum_{c \in C(t)} \bar{W}^{c, t}$$

$$\left[ \begin{array}{cc} W^{t, c} & x^{t, c} \\ (x^{t, c})^T & y^{t, c} \end{array} \right] \succeq 0$$

$t \in T$

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

- Perhaps better say you theoretically could . . .
Outline

1. The Perspective Relaxation
2. “Quick & Dirty” Extension to Nonseparable Functions
3. Firmly Beyond Separability
4. Computational Results
5. Going Even Further
6. Advertisement
7. Conclusions
Where will you publish the characterization of $k \times k$ decomposition?
Where will you publish the characterization of $k \times k$ decomposition?

With classical publishers, your own institution will have to pay dearly to access it, unless you pay ($\pm 2k$) for “gold” Open Access.
Where will you publish the characterization of $k \times k$ decomposition?

With classical publishers, your own institution will have to pay dearly to access it, unless you pay $(\pm 2k)$ for “gold” Open Access ... and your institution will pay the same to access all the others.

OJMO provides a free OA alternative (thanks to Mersenne) https://ojmo.centre-mersenne.org/

Papers will have DOI, indexation will follow.
Where will you publish the characterization of $k \times k$ decomposition?

With classical publishers, your own institution will have to pay dearly to access it, unless you pay $(\pm 2k)$ for “gold” Open Access ... and your institution will pay the same to access all the others.

OJMO provides a free OA alternative (thanks to Mersenne)

https://ojmo.centre-mersenne.org/

Papers will have DOI, indexation will follow.

We seek high-quality contributions.

Steering Commitee

- Dimitris Bertsimas
- Martine Labbé
- Eva K. Lee
- Marc Teboulle

Area Editors

- Continuous Optimization – David Russell Luke
- Discrete Optimization – Sebastian Pokutta
- Optimization under Uncertainty – Guzin Bayraksan
- Computational aspects and applications – Bernard Gendron
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.

- $2 \times 2$ PR perhaps still reasonable, but not any $k > 2$.

- Nice general idea: extraction of $2 \times 2$ SDP matrices (has at least another application, may have others).

- Full theoretical characterization:
  - matrices with exact $2 \times 2$ decomposition
  - SDP formulation for approximated $2 \times 2$ 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?