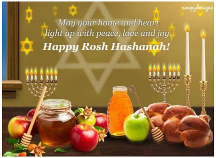
Hard Combinatorial Problems, Doubly Nonnegative Relaxations, Facial and Symmetry Reduction, and Alternating Direction Method of Multipliers

Henry Wolkowicz
Dept. Comb. and Opt., University of Waterloo, Canada



Happy 5781

Fri. Sept. 18/20, 3:30-4:30PM; Tutte Seminar U. of W.

Collaborators

Hao Hu (University of Waterloo)
Renata Sotirov (Tilburg University)
in HSW:
Facial Reduction for Symmetry Reduced Semidefinite Programs
arXiv 1912.10245 [13].

Xinxin Li (Jilin University)
Ting Kei Pong (The Hong Kong Polytechnic University)
Naomi Graham, Haesol Im, Hao Sun (University of Waterloo)
Danilo Oliveira (UFdU), Yangyang Xu (RPI)

NP-Hard problems and SDP

- Solving hard combinatorial/discrete optimization problems requires: efficient upper/lower bounding techniques.
- These problems are often modelled using quadratic objectives and/or quadratic constraints, i.e., QQPs.
- Lagrangian relaxations of QQPs lead to Semidefinite Programming, SDP, and SDP relaxations, e.g., Handbooks on SDP and Cone Optimization; [25, 1].

Solving Large Scale Problems; Reductions

- SDP s (relaxations) are expensive to solve using the (early methods of choice) interior-point approaches. This becomes *doubly* expensive when cutting planes are added, e.g., using Doubly Nonnegative, DNN, relaxations; i.e., these methods do not scale well and generally do NOT provide high accuracy solutions.

- There are currently few techniques that: exploit structure; reduce size of data; and handle large scale problems:
  - chordality reduction
  - facial reduction and regularization, FR
  - symmetry reduction, SR
  - first order methods (splittings, e.g., ADMM)

Outline/Background/Motivation III

Facial Reduction, FR ; and Symmetry Reduction, SR

- Strict feasibility (regularity) fails for many of the SDP relaxations of many hard combinatorial problems. (Compare Rademacher Theorem: Loc. Lip. functions are differentiable a.e.)
  FR , e.g., [2, 3, 4, 9, 19] provides a means of regularizing the SDP relaxations, while simultaneously reducing the size.
- SR e.g., Schrijver [20]; [19, 23, 6, 10, 11], is used to obtain a (simplified) block diagonal form, for problems that are invariant under the action of a symmetry group. Essentially, the problem can be restricted to a matrix *-algebra that contains the data matrices. Then a rotation results in the block diagonal simplified, smaller, reduced, structure.

FR and SR together into ADMM to Solve Humongous Problems

- FR, SR appear to provide a regularization and natural splitting of variables for the application of Alternating Direction Method of Multipliers, ADMM , type methods for large scale problems;
  and for exploiting structure.

- Classes of Problems:
  Min-Cut; Maxcut; Graph Partitioning; Vertex Separator;
  and here: Quadratic Assignment Problem, QAP

What is the QAP?

$A, B \in \mathbb{S}^n$ real symmetric $n \times n$ matrices, C real $n \times n$,
$\langle \cdot , \cdot \rangle$ denotes trace inner product, $\langle Y, X \rangle = \operatorname{trace} YX^\top$,
and Π_n set of $n \times n$ permutation matrices (permutations ϕ)

assign n facilities to n locations; minimize total cost

flow is A_{ij} between facilities i, j and it multiplies
distance $B_{\phi(i)\phi(j)}$ to get the total cost of assigning facilities i, j
to locations $\phi(i), \phi(j)$, respectively;
then add location costs in $-\frac{1}{2}\left(C_{i\phi(i)} + C_{j\phi(j)}\right)$

Discrete Optimization Model

The quadratic assignment problem, QAP, in the trace formulation

$$\text{(QAP)} \qquad p^* := \min_{X \in \Pi_n} \langle AXB - 2C, X \rangle \quad \left(= \operatorname{trace}(AXB - 2C)X^T \right)$$

Applications Include:

Koopmans-Beckmann ’57 [14]; Nyberg et al ’12 [17]

- facility location planning: Universities, hospital layout, airport gate assignment, wiring problems/circuit boards/VLSI, typewriter keyboards (though max?)
- Bandwith minimization of a graph
- Image processing
- Scheduling
- Supply Chains
- Economics
- Molecular conformations in chemistry
- Manufacturing lines
- Includes as special case: Traveling salesman problem and Maximum cut problem

QQP : Quadratic-Quadratic Model for $X \in \Pi$

$Xe = e, X^T e = e, X \geq 0$, doubly stochastic; ($e$ – ones vector)
turn linear constraints into quadratic

Start with Quadratic-Quadratic Model for $X \in \Pi$, a QQP

$$
\begin{array}{lll}
\min_X & \langle AXB - 2C, X \rangle & \\
\text{s.t.} & \|Xe - e\|^2 + \|X^T e - e\|^2 = 0 & \text{(r-c sums)} \\
 & XX^T = X^T X = I_n & \text{(orthogonality)} \\
 & X_{ij} X_{ik} = 0, \; X_{ji} X_{ki} = 0, \; \forall i, \; \forall j \neq k, & \text{(gangster)} \\
 & X_{ij}^2 - X_{ij} = 0, \; \forall i, j, & (0-1) \\
 & X \geq 0 & \text{(nonnegativity)}
\end{array}
$$

Dual of Dual is SDP Relaxation

The Lagrangian dual is an SDP.
The (Lagrangian) dual of this SDP is equivalent to the SDP relaxation of the QQP. BUT, strict feasibility (Slater) fails!

Derivation of FR, SDP Relax. in ZKRW [26], '98;

Start new derivation; QQP with fewer constraints; OWX [18] '18

$$
\begin{array}{lll}
\min_X & \langle AXB - 2C, X\rangle & \\
\text{s.t.} & X_{ij}X_{ik} = 0,\ X_{ji}X_{ki} = 0,\ \forall i,\ \forall j \neq k, & \text{(gangster)} \\
 & X_{ij}^2 - X_{ij} = 0,\ \forall i,j, & (0-1) \\
 & \sum_{i=1}^n X_{ij}^2 - 1 = 0,\ \forall j,\ \sum_{j=1}^n X_{ij}^2 - 1 = 0,\ \forall i. & \text{(r-c sums)}
\end{array}
$$

linearization/lifting to $Y \in \mathbb{S}^{n^2+1}$: $Y_{(ij)(st)} \cong X_{ij}X_{st}$

Gangster constraints

- The first set of constraints, the elementwise orthogonality of the row and columns of X, are the gangster constraints. They are particularly strong constraints and enable many of the other constraints (such as orthogonality $XX^T = I, X^TX = I$, row and columns sums are 1) to be redundant.
- In fact, after the facial reduction, FR, many of these constraints also become redundant.

Facial reduction, FR

Lifting; blocked appropriately; $x = \mathrm{vec}\,(X)$ columnwise

$$Y = \begin{pmatrix} x_0 \\ x \end{pmatrix} \begin{pmatrix} x_0 \\ x \end{pmatrix}^T =: \begin{bmatrix} Y_{00} & Y_{0\,1:n^2} \\ Y_{1:n^2\,0} & \overline{Y} \end{bmatrix} \in \mathbb{S}^{n^2+1},$$

$$Y_{1:n^2\,0} := \begin{bmatrix} Y_{(10)} \\ Y_{(20)} \\ \vdots \\ Y_{(n^2,0)} \end{bmatrix}; \quad \overline{Y} := \begin{bmatrix} \overline{Y}_{(11)} & \overline{Y}_{(12)} & \cdots & \overline{Y}_{(1n)} \\ \overline{Y}_{(21)} & \overline{Y}_{(22)} & \cdots & \overline{Y}_{(2n)} \\ \vdots & \ddots & \ddots & \vdots \\ \overline{Y}_{(n1)} & \ddots & \ddots & \overline{Y}_{(nn)} \end{bmatrix}$$

Objective

$$\operatorname{trace} AXBX^T = \operatorname{trace} L_A Y, \text{ where } L_A := \begin{bmatrix} 0 & 0 \\ 0 & B \otimes A \end{bmatrix}.$$

where $\otimes$ is Kronecker product

SDP Constraints (after the lifting/linearization)

E.g., the arrow constraint (linearization from the 0, 1 constraint)

$$\operatorname{arrow}(Y) := \operatorname{diag}(Y) - \begin{bmatrix} 0 \\ Y_{1:n^2\,0} \end{bmatrix} = e_0,$$

e_0 first (0-th) unit vector
(redundant in the final SDP relaxation)

DNN, doubly nonnegative

$$Y \in \mathrm{DNN} = \{Y \in \mathbb{S}_+^{n^2+1} : 0 \le Y \ (\le 1)\}$$

DNN is doubly nonnegative cone, i.e., intersection of positive semidefinite cone and nonnegative orthant.

SDP constraints and FR cont. . .

Trace constraints (from linear equality constraints

$$\operatorname{trace} D_1 Y = 0, \qquad D_1 := \begin{bmatrix} n & -e_n^T \otimes e_n^T \\ -e_n \otimes e_n & (e_n e_n^T) \otimes I_n \end{bmatrix},$$

$$\operatorname{trace} D_2 Y = 0, \qquad D_2 := \begin{bmatrix} e^T e & -e^T \otimes e_n^T \\ -e \otimes e_n & I_n \otimes (e_n e_n^T) \end{bmatrix},$$

e_j vector of ones of dimension j; $D_i \succeq 0, i = 1, 2$; nullspaces of these matrices yield the facial reduction $Y = VRV^T$.

Block: trace, diagonal and off-diagonal

$$\begin{array}{rcl} \mathcal{D}_t(Y) & := & \left(\operatorname{trace} \overline{Y}_{(ij)}\right) = I \in \mathbb{S}^n; \\ \mathcal{D}_d(Y) & := & \sum_{i=1}^n \operatorname{diag} Y_{(ii)} = e_n \in \mathbb{R}^n; \\ \mathcal{D}_o(Y) & := & \left(\sum_{s \neq t} \left(\overline{Y}_{(ij)}\right)_{st}\right) = \hat{I} \in \mathbb{S}^n, \end{array}$$

where $\hat{I} := ee^T - I$.

SDP constraints cont. . .

trace $Y = n + 1$; and Gangster constraints on Y

The Hadamard product and orthogonal type constraints lead to

gangster constraints

i.e., simple constraints that restrict elements to be zero (shoot holes in the matrix) and/or restrict entire blocks.

gangster and restricted gangster constraint on Y:

$$\mathcal{G}_H(Y) = 0,$$

for specific index sets H, e.g., Hadamard orthogonal rows of $X \in \Pi$ yields

$$i \neq j : \Longrightarrow \; X_{ik}X_{jk} = 0, \forall k \; \Longrightarrow \; Y_{(ik),(jk)} = 0, \forall k.$$

SDP relaxation

SDP Relaxation with Many (some redundant) Constraints

$$
\begin{array}{rll}
\text{qap}(n,A,B) \geq p^*_{\text{SDP}} := & \min & \operatorname{trace} L_A Y \\
& \text{s.t.} & \operatorname{arrow}(Y) = e_0 \\
& & \operatorname{trace} D_1 Y = 0, \ \operatorname{trace} D_2 Y = 0 \\
& & \mathcal{G}_{J_0}(Y) = 0, \ \ Y_{00} = 1 \\
& & \mathcal{D}_t(Y) = I, \mathcal{D}_d(Y) = e, \mathcal{D}_o(Y) = \widehat{I} \\
& & Y \in \mathbb{S}_+^{n^2+1}
\end{array}
$$

Equivalent FR greatly simplified SDP; with $Y = \widetilde{V} R \widetilde{V}^T$

$$
\begin{array}{rcll}
\text{qap}(n,A,B) \geq p^*_{\text{SDP}} & = & \min & \operatorname{trace}\left(\widetilde{V}^T L_A \widetilde{V}\right) R \\
& & \text{s.t.} & \mathcal{G}_{\widehat{J}_{\mathcal{I}}}(\widetilde{V} R \widetilde{V}^T) = \mathcal{G}_{\widehat{J}_{\mathcal{I}}}(e_0 e_0^T) \\
& & & R \in \mathbb{S}_+^{(n-1)^2+1}
\end{array}
$$

Splitting methods and facial reduction, FR

Natural Splitting? $Y \in \mathcal{P}$, $R \in \mathbb{S}^r_+$

$Y = VRV^T$

$$Y \in \mathcal{P} \subset \mathbb{S}^{N+1}_+, \qquad R \in \mathbb{S}^r_+, \quad r < N+1$$

Facial reduction provides a guarantee that strict feasibility holds for the primal and that the dual of the dual is the primal. (In our instance of QAP, strict feasibility holds for primal and dual.)
AND: it provides a reduction in dimension AND so rank.

Natural separation/splitting

There is a natural separation of constraints where

$$Y \in \mathcal{P} \text{ polyhedral} \qquad R \in \mathbb{S}^r_+ \text{ sdp cone}$$

Group invariance and symmetry reduction, SR

General primal-dual SDP

$$p^*_{\mathrm{SDP}} = \min\{\langle C, X\rangle \mid \mathcal{A}(X) = b \in \mathbb{R}^m, \quad X \in \mathbb{S}^n_+\},$$

where $A_i \in \mathbb{S}^n$, $\mathcal{A}(X) = (\operatorname{trace} A_i X)$

$$d^*_{\mathrm{SDP}} = \max\{\langle b, y\rangle \mid \mathcal{A}^*(y) \preceq C, \quad y \in \mathbb{R}^m\}$$

where $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$; $\mathcal{A}^*(y) = \sum_i y_i A_i$.

SR: substitute using $\tilde{B}^*$; obtain SR block diagonal form

- use procedure for simplifying an SDP that is invariant under the action of a symmetry group, Schrijver [20];
- the appropriate algebra isomorphism follows from the Artin-Wedderburn theory [24].

SR continued

Framework

- $\mathcal{G}$ - nontrivial group of permutation matrices of size n.
- commutant, $A_{\mathcal{G}}$ (or centralizer ring) of $\mathcal{G}$:

$$\begin{aligned} A_{\mathcal{G}} &= \{X \in \mathbb{R}^{n\times n} \mid PX = XP, \ \forall P \in \mathcal{G}\} \\ &= \{X \in \mathbb{R}^{n\times n} \mid \mathcal{R}_{\mathcal{G}}(X) = X\}, \end{aligned}$$

  where $\mathcal{R}_{\mathcal{G}}(X) := \frac{1}{|\mathcal{G}|}\sum_{P\in\mathcal{G}} PXP^T$, is the Reynolds operator, or group average, and is orthogonal projection onto the commutant;
- the commutant $A_{\mathcal{G}}$ is a matrix $*$-algebra, i.e., closed under addition, scalar multiplication, matrix multiplication, and taking transposition.

Group invariance and symmetry reduction, SR

Basis for $A_{\mathcal{G}}$: $\{B_1, \ldots, B_d\}$, $B_i \in \{0,1\}^{n\times n}$

- basis for $A_{\mathcal{G}}$ from the orbits of the action of $\mathcal{G}$ on ordered pairs of vertices, where the orbit of $(u_i, u_j) \in \{0,1\}^n \times \{0,1\}^n$ under the action of $\mathcal{G}$ is the set $\{(Pu_i, Pu_j) \mid P \in \mathcal{G}\}$, and $u_i \in \mathbb{R}^n$ is the i-th unit vector.

Definition (coherent configuration (J ones matrix))

A set of zero-one $n \times n$ matrices $\{B_1, \ldots, B_d\}$ is called a coherent configuration of rank d if

1. $\sum_{i\in\mathcal{I}} B_i = I$ for some $\mathcal{I} \subset \{1, \ldots, d\}$, and $\sum_{i=1}^{d} B_i = J$;
2. $B_i^{\mathrm{T}} \in \{B_1, \ldots, B_d\}$ for $i = 1, \ldots, d$;
3. $B_iB_j \in \operatorname{span}\{B_1, \ldots, B_d\}$, $\forall i, j \in \{1, \ldots, d\}$.

Restrict SDP to feasible points in a matrix $*$-Algebra

Theorem (de Klerk et al, [7])

Let $A_{\mathcal{G}}$ denote a matrix $*$-algebra that contains the data matrices of an SDP problem as well as the identity matrix. If the SDP problem has an optimal solution, then it has an optimal solution in $A_{\mathcal{G}}$, the centralizer ring.

Corollary (can reduce size of feasible set to consider)

We can restrict the feasible set of the optimization problem to its intersection with $A_{\mathcal{G}}$. In particular, we can use the basis matrices and assume that

$$X \in \mathcal{F}_X \cap A_{\mathcal{G}} \Leftrightarrow \left[X = \sum_{i=1}^{d} x_i B_i =: \mathcal{B}^*(x) \in \mathcal{F}_X, \text{ for some } x \in \mathbb{R}^d \right].$$

First SR using substitution $X = \mathcal{B}^*(x)$

We assume that the group of permutation matrices $\mathcal{G}$ is such (small enough) that the centralizer/commutant $A_{\mathcal{G}}$ contains our data matrices, (A_i, C).

$$p^*_{\mathrm{SDP}} = \min\{\langle C, X\rangle \mid \mathcal{A}(X) = b, \ \ X \succeq 0\}$$

Feasible set reduced; optimal value unchanged

$$p^*_{\mathrm{SDP}} = \min\{\langle \mathcal{B}(C), x\rangle \mid (\mathcal{A} \circ \mathcal{B}^*)(x) = b, \ \ \mathcal{B}^*(x) \succeq 0\}$$

Here, $\mathcal{B} = \mathcal{B}^{**}$ is the adjoint of $\mathcal{B}^*$.
In the case of a doubly nonnegative relaxation, the structure of our basis allows us to set/constrain $x \geq 0$.

Second $\mathbb{SR}$, block diagonal form

Basic $*$-algebra

$\mathcal{M}$ is called basic if $\mathcal{M} = \{\oplus_{i=1}^{t} M \mid M \in \mathbb{C}^{m\times m}\}$, where $\oplus$ denotes the direct sum of matrices.

Theorem (Wedderburn [24])

Let $\mathcal{M}$ be a matrix $*$-algebra containing the identity matrix. Then there exists a unitary matrix Q such that $Q^*\mathcal{M}Q$ is a direct sum of basic matrix $*$-algebras.

Second SR

Mutual block diagonalization with orthogonal Q, t blocks

$$\tilde{B}_j := Q^T B_j Q =: \text{Blkdiag}((\tilde{B}_j^k)_{k=1}^t), \forall j = 1, \ldots, d.$$

Linear transformation for $Q^T X Q = \sum_{j=1}^d x_j \tilde{B}_j =: \tilde{\mathcal{B}}^*(x)$

$$\sum_{j=1}^d x_j \tilde{B}_j = \begin{bmatrix} \tilde{\mathcal{B}}_1^*(x) & & \\ & \ddots & \\ & & \tilde{\mathcal{B}}_t^*(x) \end{bmatrix} =: \text{Blkdiag}((\tilde{\mathcal{B}}_k^*(x))_{k=1}^t)$$

where $\tilde{\mathcal{B}}_k^*(x) =: \sum_{j=1}^d x_j \tilde{B}_j^k \in \mathcal{S}_+^{n_i}$ is k-th diagonal block of $\tilde{\mathcal{B}}^*(x)$, and sum of t block sizes $n_1 + \ldots + n_t = n$.

For any feasible X

$$X = \mathcal{B}^*(x) = Q\tilde{\mathcal{B}}^*(x)Q^T \in \mathcal{F}_X$$

Second $\mathcal{SR}$ block diagonal form using $X = Q\tilde{\mathcal{B}}^*(x)Q^T$

Block diagonal problem

$$p^*_{\rm SDP} = \min\{\langle\tilde{\mathcal{B}}(\tilde{C}), x\rangle \mid (\tilde{\mathcal{A}} \circ \tilde{\mathcal{B}}^*)(x) = b, \;\; \tilde{\mathcal{B}}^*(x) \succeq 0\},$$

After appropriate simplifications

$$p^*_{\rm SDP} = \min\{c^T x \mid Ax = b, \;\; \tilde{\mathcal{B}}^*_k(x) \succeq 0, \; k = 1, \ldots, t\}.$$

feasible set and feasible slacks are

$$\mathcal{F}_x := \{x \mid \tilde{\mathcal{B}}^*(x) \succeq 0, \; Ax = b, \; x \in \mathbb{R}^d\}$$

$$\mathcal{S}_x := \{\tilde{\mathcal{B}}^*(x) \succeq 0 \mid Ax = b, \; x \in \mathbb{R}^d\}.$$

$\tilde{\mathcal{B}}^*(x)$ is a block-diagonal matrix

get smaller problem typically: $x \in \mathbb{R}^d, \quad d \ll \sum_{i=1}^d t(n_i) \ll t(n)$,
where $t(k) = k(k+1)/2$ is the triangular number.

FR for symmetric reduced program; exposing vectors

Maximum rank preserving properties of SR

$$\begin{aligned}\max\{\operatorname{rank}(X) : X \in \mathcal{F}_X\} &= \operatorname{rank}(X),\ \forall X \in \operatorname{ri}(\mathcal{F}_X)\\ &= \operatorname{rank}(X),\ \forall X \in \operatorname{ri}(\operatorname{face}(\mathcal{F}_X)),\end{aligned}$$

$\operatorname{face}(\mathcal{F}_X)$ is minimal face of $\mathbb{S}^n_+$ containing feasible set.

Theorem

Let $r = \max\{\operatorname{rank}(X) \ : \ X \in \mathcal{F}_X\}$. *Then*

$$\begin{aligned} r &= \max\left\{\operatorname{rank}\left(\tfrac{1}{|\mathcal{G}|}\textstyle\sum_{P\in\mathcal{G}} P^T X P\right) \ : \ X \in \mathcal{F}_X\right\}\\ &= \max\{\operatorname{rank}(X) \ : \ X \in \mathcal{F}_X \cap A_{\mathcal{G}}\} \ \textit{centralizer}\\ &= \max\{\operatorname{rank}(\tilde{\mathcal{B}}^*(x)) \ : \ \tilde{\mathcal{B}}^*(x) \in \mathcal{S}_x\} \ \textit{slacks}\end{aligned}$$

FR; exposing vectors; symmetry reduced programs

For many combinatorial problems, the semidefinite relaxation is not strictly feasible. Therefore it is degenerate and ill-posed. Therefore, the symmetry reduced problem is degenerate as well.

We want to implement both SR and FR together and do it efficiently and robustly.

Key is exposing vectors

The exposing vectors of symmetry reduced program can be obtained from the exposing vectors from original program. (Therefore, we can exploit structure of original problem.)

Exposing vectors for FR

Let $0 \neq W = UU^T$ be an exposing vector of the minimal face of $\mathbb{S}^n_+$ containing the feasible region $\mathcal{F}_X$:
$X \in \mathcal{F}_X \implies \operatorname{trace} WX = 0$;
let $U \in \mathbb{R}^{n \times (n-r)}$ full column rank;
let $V \in \mathbb{R}^{n \times r}$ with $\operatorname{Range}(V) = \operatorname{Null}(U^T)$.

FR: use substitution $X = \mathcal{V}^*(R) = VRV^T$

obtain equivalent, smaller,

$$\min\{\langle V^T C V, R\rangle \mid \langle V^T A_i V, R\rangle = b_i, \quad i = 1, \ldots, m, \quad R \in \mathbb{S}^r_+\}.$$

In fact, with appropriate V, $\hat{R}$ strictly feasible corresponds to $\hat{X} = \mathcal{V}^*(\hat{R}) \in \operatorname{ri}(\mathcal{F}_X)$. Moreover, at least one constraint becomes redundant at each FR step.
(So at most $\min\{m, n-1\}$ FR steps.)

Exposing vectors for $\mathbb{SR}$ in commutant $A_{\mathcal{G}}$

Lemma

Let W be an exposing vector of rank d of a face of $\mathbb{S}^n_+$ containing $\mathcal{F}_X$. Then there exists an exposing vector $W_{\mathcal{G}} \in A_{\mathcal{G}}$ with $\operatorname{rank}(W_{\mathcal{G}}) \geq d$.

Proof.

Let W be the exposing vector of rank d, i.e.,$W \succeq 0$ and $X \in \mathcal{F}_X \implies \langle W, X \rangle = 0$.
Since the original problem is $\mathcal{G}$-invariant, $PXP^T \in \mathcal{F}_X$ for every $P \in \mathcal{G}$, we conclude that

$$\langle W, PXP^T \rangle = \langle P^T WP, X \rangle = 0.$$

Therefore, $P^T WP \succeq 0$ is an exposing vector of rank d. Thus $W_{\mathcal{G}} = \frac{1}{|\mathcal{G}|} \sum_{P \in \mathcal{G}} P^T WP$ is an exposing vector of $\mathcal{F}_X$.
That the rank is at least d follows from taking the sum of nonsingular congruences of $W \succeq 0$. □

Exposing vectors for $\mathbb{SR}$ in block diagonal form

Lemma

Let W be an exposing vector of face of $\mathbb{S}^n_+$ containing $\mathcal{F}_X$, and assume that $W \in A_{\mathcal{G}}$. Let Q be the orthogonal matrix given above in the block diagonalization. Then $\widetilde{W} = Q^T WQ$ exposes a face of $\mathbb{S}^n_+$ containing $\mathcal{S}_x$.

Theorem

Let $W \in A_{\mathcal{G}}$ be an exposing vector of $\operatorname{face}(\mathcal{F}_X)$, the minimal face of $\mathbb{S}^n_+$ containing $\mathcal{F}_X$. Then the block-diagonal matrix $\widetilde{W} = Q^T WQ$ exposes $\operatorname{face}(\mathcal{S}_x)$, the minimal face of $\mathbb{S}^n_+$ containing $\mathcal{S}_x$.

Facial and symmetry reduced program

$\widetilde{W} = Q^T W Q$ exposes the minimal face of $\mathbb{S}^n_+$ containing $\mathcal{S}_x$;

$\widetilde{W} = \text{Blkdiag}(\widetilde{W}_1, \ldots, \widetilde{W}_t)$, $\widetilde{W}_i = \tilde{U}_i \tilde{U}_i^T$, $\tilde{U}_i$ full rank, $i = 1, \ldots, t$

Let $\tilde{V}_i$ be a full rank matrix $\text{Range}(V_i) = \text{Null}(U_i^T)$
$\tilde{V} = \text{Blkdiag}(\tilde{V}_1, \ldots, \tilde{V}_t)$.
FR:

$$\begin{aligned} p^*_{FR} &= \min\{c^T x \mid Ax = b,\ \tilde{\mathcal{B}}^*(x) = \tilde{V}\tilde{R}\tilde{V}^T,\ \tilde{R} \succeq 0\} \\ &= \min\{c^T x | Ax = b, \tilde{\mathcal{B}}^*_k(x) = \tilde{V}_k \tilde{R}_k \tilde{V}_k^T, \tilde{R}_k \succeq 0, k = 1 : t\} \end{aligned}$$

where $\tilde{V}_k \tilde{R}_k \tilde{V}_k^T$ is the corresponding k-th block of $\tilde{\mathcal{B}}^*(x)$, and
$\tilde{R} = \text{Blkdiag}(\tilde{R}_1, \ldots, \tilde{R}_t)$.

Singularity degree for FR and SR

Definition

The singularity degree of a feasible region $\mathcal{F}$, denoted by sd($\mathcal{F}$), is the smallest number of steps required for the FR algorithm to terminate.

Holder error bound, Sturm '00 [21]

For a feasible set $\mathcal{F}_X = \mathcal{L} \cap \mathbb{S}^n_+$, for a linear manifold $\mathcal{L}$, Sturm showed that a Holder error bound always holds, i.e., the distance of any X to $\mathcal{F}_X$ can be bounded by a multiple of a certain power of the distance to $\mathcal{L}$ and to $\mathbb{S}^n_+$ separately. Sturm showed that the Holder exponent can be set to $2^{-sd(\mathcal{F}_X)}$. (It does NOT depend on the size or rank of the matrices, only the singularity degree.)

Theorem

$sd(\mathcal{F}_x) \leq sd(\mathcal{F}_X)$.

Motivation for first order methods and bounding

Difficulties for primal-dual interior-point methods for SDP

- solving large problems
- obtaining high accuracy solutions
- exploiting sparsity
- adding on nonnegativity and other cutting plane constraints

First order operator splitting methods for SDP

- FR/SR: regularization/dim. size reduction/natural splitting, $Y = VRV^T$
- Flexibility in dealing with additional constraints
- separable/split optimization steps are inexpensive

Alternating direction method of multipliers, ADMM

It is extremely successful for splittings with two cones. The ADMM is well suited for large-scaled DNN problems, where one can split between simple polyhedral and convex cone projections, e.g., survey Boyd et al ’11 [5]; applications to QAP, Mincut e.g., [18, 15, 12].

Augmented Lagrangian for: $\tilde{\mathcal{B}}^*(x) = \tilde{V}\tilde{R}\tilde{V}^T$

Let $\tilde{V} = \mathrm{Blkdiag}(\tilde{V}_1, \ldots, \tilde{V}_t)$ and $\tilde{R} = \mathrm{Blkdiag}(\tilde{R}_1, \ldots, \tilde{R}_t)$.

The augmented Lagrangian

$$\begin{aligned} \mathcal{L}(x, \tilde{R}, \tilde{Z}) &= \langle \tilde{C}, \tilde{\mathcal{B}}^*(x) \rangle + \langle \tilde{Z}, \tilde{\mathcal{B}}^*(x) - \tilde{V}\tilde{R}\tilde{V}^T \rangle \\ &\quad + \frac{\beta}{2} || \tilde{\mathcal{B}}^*(x) - \tilde{V}\tilde{R}\tilde{V}^T ||^2 \end{aligned}$$

where, $\tilde{C} = Q^T C Q$ is block-diagonal matrix as $C \in A_{\mathcal{G}}$;
Lagrange multiplier $\tilde{Z}$ is also in block-diagonal form;
$\beta > 0$ is the penalty parameter.

$$\max_{\tilde{Z}} \min_{x \in P, \tilde{R} \succeq 0} \mathcal{L}(x, \tilde{R}, \tilde{Z}),$$

P is a simple polyhedral set: $Ax = b, x \geq 0$

Simple subproblems

Splitting yields three subproblems

find following updates $(x_+, \tilde{R}_+, \tilde{Z}_+)$:

$$\begin{aligned} x_+ &= \arg\min_{x \in P} \mathcal{L}(x, \tilde{R}, \tilde{Z}), \\ \tilde{R}_+ &= \arg\min_{\tilde{R} \succeq 0} \mathcal{L}(x_+, \tilde{R}, \tilde{Z}), \\ \tilde{Z}_+ &= \tilde{Z} + \gamma\beta(\tilde{\mathcal{B}}^*(x_+) - \tilde{V}\tilde{R}_+\tilde{V}^T). \end{aligned}$$

$\gamma \in (0, \frac{1+\sqrt{5}}{2})$ - step size for updating dual variable $\tilde{Z}$.

On solving $\tilde{R}$-subproblem explicitly

Complete square

$$\begin{aligned}\tilde{R}_+ &= \min_{\tilde{R}\succeq 0} \|\tilde{\mathcal{B}}^*(x) - \tilde{V}\tilde{R}\tilde{V}^T + \tfrac{1}{\beta}\tilde{Z}\|^2 \\ &= \min_{\tilde{R}\succeq 0} \|\tilde{R} - \tilde{V}^T(\tilde{\mathcal{B}}^*(x) + \tfrac{1}{\beta}\tilde{Z})\tilde{V}\|^2 \\ &= \sum_{k=1}^{t} \min_{\tilde{R}_k\succeq 0} \|\tilde{R}_k - \left(\tilde{V}^T(\tilde{\mathcal{B}}^*(x) + \tfrac{1}{\beta}\tilde{Z})\tilde{V}\right)_k\|^2.\end{aligned}$$

Solve k small problems/psd projections

$$\tilde{R}_k = \mathcal{P}_{\mathbb{S}_+}\left(\tilde{V}^T(\tilde{\mathcal{B}}^*(x) + \frac{1}{\beta}\tilde{Z})\tilde{V}\right)_k, \quad k = 1, \ldots, t,$$

On solving the x-subproblem

$$x_+ = \arg\min_{x \in P} \left\| \tilde{\mathcal{B}}^*(x) - \tilde{V}\tilde{R}\tilde{V}^T + \frac{\tilde{C} + \tilde{Z}}{\beta} \right\|^2 .$$

• For many combinatorial optimization problems, some of the constraints such as in $Ax = b$ become redundant after FR of their semidefinite programming relaxations.
• Thus, the set P often collapses to a simple set. This often leads to an analytic solution for the x-subproblem.
• This happens for the quadratic assignment, graph partitioning, vertex separator, and shortest path problems.

Numerical results for the QAP

Tests using:

• computer: DellPowerEdge; two Intel Xeon E5-2637v3 4-core 3.5 GHz (Haswell) processors; 64GB of memory
• Mosek as the interior point solver
• We include *huge* problems of sizes up to $n = 512$, i.e. the SDP relaxation is of size $n^2 + 1 = 1 + 512^2$ and this therefore includes order $n^4 = 625 * 10^8$ nonnegativity constraints.

Stopping

We terminate when the primal and dual residuals are small or we are not making progress in decreasing the duality gap.

Results

Significant improvements for huge problems

• The following table shows that we significantly improve bounds for all eng1_n and eng9_n instances.

• Moreover, we are able to compute bounds for huge QAP instances with $n = 256$ and $n = 512$ in a reasonable amount of time.

• Note that for each instance from of size $n = 2^d$, the DNN relaxation boils down to $d + 1$ positive semidefinite blocks of order n. There are currently no interior point algorithms that are able to solve such huge problems.

Mittlemann and Peng problems '10 [16]

Table: Lower and upper bounds for different QAP instances.

		MandP '10 [16]		ADMM			
problem	UB	LB	time	OBJ	LB	time	res.
Harper_16	2752	2742	1	2743	2742	1.92	4.50e-05
Harper_32	27360	27328	3	27331	27327	9.70	1.67e-04
Harper_64	262260	262160	56	262196	261168	36.12	1.12e-05
Harper_128	2479944	2446944	1491	2446800	2437880	186.12	3.86e-05
Harper_256	22370940	-	-	22369996	22205236	432.10	9.58e-06
Harper_512	201329908	-	-	201327683	200198783	1903.66	9.49e-06
eng1_16	1.58049	1.5452	1	1.5741	1.5740	2.28	3.87e-05
eng1_32	1.58528	1.24196	4	1.5669	1.5637	14.63	5.32e-06
eng1_64	1.58297	0.926658	56	1.5444	1.5401	38.35	4.69e-06
eng1_128	1.56962	0.881738	1688	1.4983	1.4870	389.04	2.37e-06
eng1_256	1.57995	-	-	1.4820	1.3222	971.48	9.95e-06
eng1_512	1.53431	-	-	1.4553	1.3343	9220.13	9.66e-06
eng9_16	1.02017	0.930857	1	1.0014	1.0013	3.58	2.11e-06
eng9_32	1.40941	1.03724	3	1.3507	1.3490	12.67	3.80e-05
eng9_64	1.43201	0.887776	68	1.3534	1.3489	74.89	6.60e-05
eng9_128	1.43198	0.846574	2084	1.3331	1.3254	700.27	8.46e-06
eng9_256	1.45132	-	-	1.3152	1.2610	1752.72	9.74e-06
eng9_512	1.45914	-	-	1.3074	1.1168	23191.96	9.96e-06
VQ_32	297.29	294.49	3	296.3241	296.1351	11.82	1.27e-05
VQ_64	353.5	352.4	45	352.7621	351.4358	43.17	4.22e-04
VQ_128	399.09	393.29	2719	398.4269	396.2794	282.28	6.19e-04
rand_256	126630.6273	-	-	124589.4215	124469.2129	2054.61	3.78e-05
rand_512	577604.8759	-	-	570935.1468	569915.3034	9694.71	1.32e-04

Solving some to optimality using only DNN relaxation

		SDPNAL+: STYZ'20 [22]		ADMM: OWX'15 [18]		SDP: KS'10 [8]		ADMM			
inst.	opt	LB	time	LB	time	LB	time	OBJ	LB	time	res
esc16a	68	63.2750	16	64	20.14	63.2756	0.75	63.2856	63.2856	2.48	1.17e-11
esc16b	292	289.9730	24	290	3.10	289.8817	1.04	290.0000	290.0000	0.78	9.95e-13
esc16c	160	153.9619	65	154	8.44	153.8242	1.78	154.0000	153.9999	2.11	2.56e-09
esc16d	16	13.0000	2	13	17.39	13.0000	0.89	13.0000	13.0000	1.04	9.94e-13
esc16e	28	26.3367	2	27	24.04	26.3368	0.51	26.3368	26.3368	1.21	9.89e-13
esc16f	0	-	-	0	3.22e+02	0	0.14	0	0	0.01	2.53e-14
esc16g	26	24.7388	4	25	33.54	24.7403	0.51	24.7403	24.7403	1.40	9.95e-13
esc16h	996	976.1857	10	977	4.01	976.2244	0.79	976.2293	976.2293	2.51	7.73e-13
esc16i	14	11.3749	6	12	100.79	11.3749	0.73	11.3749	11.3660	6.15	2.53e-06
esc16j	8	7.7938	4	8	56.90	7.7942	0.42	7.7942	7.7942	0.21	9.73e-13
esc32a	130	103.3206	333	104	2.89e+03	103.3194	114.88	103.3211	103.0465	12.36	3.62e-06
esc32b	168	131.8532	464	132	2.52e+03	131.8718	5.58	131.8843	131.8843	4.64	9.59e-13
esc32c	642	615.1600	331	616	4.48e+02	615.1400	3.70	615.1813	615.1813	8.04	2.05e-10
esc32d	200	190.2273	67	191	8.68e+02	190.2266	2.09	190.2271	190.2263	5.86	7.45e-08
esc32e	2	1.9001	149	2	1.81e+03	-	-	1.9000	1.9000	0.70	4.49e-13
esc32f	2	-	-	2	1.80e+03	-	-	1.9000	1.9000	0.76	4.49e-13
esc32g	6	5.8336	65	6	6.04e+02	5.8330	1.80	5.8333	5.8333	3.50	9.97e-13
esc32h	438	424.3256	1076	425	3.02e+03	424.3382	7.16	424.4027	424.3184	5.89	1.03e-06
esc64a	116	-	-	98	1.64e+04	97.7499	12.99	97.7500	97.7500	5.33	8.95e-13
esc128	64	-	-	-	-	53.0844	140.36	51.7518	51.7518	137.71	1.18e-12

Table: Esc instances

Conclusion

- We discussed strategies for finding new, strengthened lower and upper bounds, for large discrete optimization problems from the resulting HUGE DNN relaxations.
- In particular, we combined FR with SR efficiently to obtain a regularized problem reduced in dimension and in size. We exploited the resulting natural splitting with a ADMM approach.
- Interesting theoretical results about singularity degree and rank preservation arose for the SR.

References I

- A.F. Anjos and J.B. Lasserre (eds.), *Handbook on semidefinite, conic and polynomial optimization*, International Series in Operations Research & Management Science, Springer-Verlag, 2011.
- J.M. Borwein and H. Wolkowicz, *Characterization of optimality for the abstract convex program with finite-dimensional range*, J. Austral. Math. Soc. Ser. A **30** (1980/81), no. 4, 390–411. MR 83i:90156
- ______, *Facial reduction for a cone-convex programming problem*, J. Austral. Math. Soc. Ser. A **30** (1980/81), no. 3, 369–380. MR 83b:90121
- ______, *Regularizing the abstract convex program*, J. Math. Anal. Appl. **83** (1981), no. 2, 495–530. MR 83d:90236

References II

- S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Machine Learning **3** (2011), no. 1, 1–122.
- E. de Klerk, *Exploiting special structure in semidefinite programming: A survey of theory and applications*, European J. Oper. Res. **201** (2010), no. 1, 1–10.
- E. de Klerk, C. Dobre, and D.V. Ṗasechnik, *Numerical block diagonalization of matrix ∗-algebras with application to semidefinite programming*, Math. Program. **129** (2011), no. 1, Ser. B, 91–111. MR 2831404
- E. de Klerk and R. Sotirov, *Exploiting group symmetry in semidefinite programming relaxations of the quadratic assignment problem*, Math. Program. **122** (2010), no. 2, Ser. A, 225–246. MR MR2546331

- D. Drusvyatskiy and H. Wolkowicz, *The many faces of degeneracy in conic optimization*, Foundations and Trends® in Optimization **3** (2017), no. 2, 77–170.
- K. Gatermann and P.A. Parrilo, *Symmetry groups, semidefinite programs, and sums of squares*, Journal of Pure and Applied Algebra **192** (2004), no. 1-3, 95–128.
- D. Gijswijt, *Matrix algebras and semidefinite programming techniques for codes*, PhD Thesis (2010).
- N. Graham, H. Hu, H. Im, X. Li, and H. Wolkowicz, *A restricted dual Peaceman-Rachford splitting method for QAP*, Tech. report, University of Waterloo, Waterloo, Ontario, 2020, 29 pages, research report.

- Hao Hu, Renata Sotirov, and Henry Wolkowicz, *Facial reduction for symmetry reduced semidefinite programs*, 2019.
- T.C. Koopmans and M.J. Beckmann, *Assignment problems and the location of economic activities*, Econometrica **25** (1957), 53–76.
- X. Li, T.K. Pong, H. Sun, and H. Wolkowicz, *A strictly contractive Peaceman-Rachford splitting method for the doubly nonnegative relaxation of the minimum cut problem*, Tech. report, University of Waterloo, Waterloo, Ontario, 2019, 40 pages, research report.

References V

- H. Mittelmann and J. Peng, *Estimating bounds for quadratic assignment problems associated with hamming and manhattan distance matrices based on semidefinite programming*, SIAM Journal on Optimization **20** (2010), no. 6, 3408–3426.
- Axel Nyberg and Tapio Westerlund, *A new exact discrete linear reformulation of the quadratic assignment problem*, European Journal of Operational Research **220** (2012), no. 2, 314–319.
- D.E. Oliveira, H. Wolkowicz, and Y. Xu, *ADMM for the SDP relaxation of the QAP*, Math. Program. Comput. **10** (2018), no. 4, 631–658.
- F.N. Permenter, *Reduction methods in semidefinite and conic optimization*, Ph.D. thesis, Massachusetts Institute of Technology, 2017.

References VI

A. Schrijver, *A comparison of the Delsarte and Lovász bounds*, IEEE Transactions on Information Theory **25** (1979), no. 4, 425–429.

J.F. Sturm, *Error bounds for linear matrix inequalities*, SIAM J. Optim. **10** (2000), no. 4, 1228–1248 (electronic). MR 1777090 (2001i:90057)

D. Sun, K.C. Toh, Y. Yuan, and X.Y. Zhao, *SDPNAL $+$: A Matlab software for semidefinite programming with bound constraints (version 1.0)*, Optimization Methods and Software **35** (2020), no. 1, 87–115.

F. Vallentin, *Symmetry in semidefinite programs*, Linear Algebra Appl. **430** (2009), no. 1, 360–369. MR 2460523

J.H.M. Wedderburn, *On Hypercomplex Numbers*, Proc. London Math. Soc. (2) **6** (1908), 77–118. MR 1575142

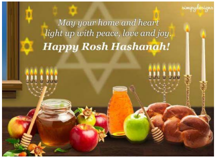
References VII

- H. Wolkowicz, R. Saigal, and L. Vandenberghe (eds.), *Handbook of semidefinite programming*, International Series in Operations Research & Management Science, 27, Kluwer Academic Publishers, Boston, MA, 2000, Theory, algorithms, and applications. MR MR1778223 (2001k:90001)
- Q. Zhao, S.E. Karisch, F. Rendl, and H. Wolkowicz, *Semidefinite programming relaxations for the quadratic assignment problem*, vol. 2, 1998, Semidefinite programming and interior-point approaches for combinatorial optimization problems (Toronto, ON, 1996), pp. 71–109. MR 1616871

Thanks for your attention!

Hard Combinatorial Problems, Doubly Nonnegative Relaxations, Facial and Symmetry Reduction, and Alternating Direction Method of Multipliers

Henry Wolkowicz
Dept. Comb. and Opt., University of Waterloo, Canada



Happy 5781

Fri. Sept. 18/20, 3:30-4:30PM; Tutte Seminar U. of W.