Compact Lifted Relaxations for Low-Rank Optimization

Ryan Cory-Wright
Department of Analytics, Marketing and Operations, Imperial Business School, London, UK
ORCID: 0000-0002-4485-0619
r.cory-wright@imperial.ac.uk

Jean Pauphilet
Management Science and Operations, London Business School, London, UK
ORCID: 0000-0001-6352-0984
jpauphilet@london.edu

We develop tractable convex relaxations for rank-constrained quadratic optimization problems over $n \times m$ matrices, a setting for which tractable relaxations are typically only available when the objective or constraints admit spectral (permutation-invariant) structure. We derive lifted semidefinite relaxations that do not require such spectral terms. Although a direct lifting introduces a large semidefinite constraint in dimension $n^2 + nm + 1$, we prove that many blocks of moment matrix are redundant and derive an equivalent compact relaxation that only involves two semidefinite constraints of dimension $nm + 1$ and $n + m$ respectively. For matrix completion, basis pursuit, and reduced-rank regression problems, we exploit additional structure to obtain even more compact formulations involving semidefinite matrices of dimension at most $2\max(n, m)$. Overall, we obtain scalable semidefinite bounds for a broad class of low-rank quadratic problems.

Key words: Low-rank optimization; semidefinite programming; matrix completion

---

1. Introduction

This work develops tractable relaxations for rank-constrained optimization problems of the form

$$\begin{aligned} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}} \quad & \lambda \cdot \operatorname{rank}(\boldsymbol{X}) + \left\langle \boldsymbol{H}, \operatorname{vec}(\boldsymbol{X}^\top)\operatorname{vec}(\boldsymbol{X}^\top)^\top \right\rangle + \langle \boldsymbol{D}, \boldsymbol{X} \rangle \\ \text{s.t.} \quad & \operatorname{rank}(\boldsymbol{X}) \leq k, \ \left\langle \boldsymbol{Q}_i, \operatorname{vec}(\boldsymbol{X}^\top)\operatorname{vec}(\boldsymbol{X}^\top)^\top \right\rangle + \langle \boldsymbol{E}_i, \boldsymbol{X} \rangle \leq b_i \ \forall i \in \mathcal{I} \end{aligned} \tag{1}$$

where $\boldsymbol{H}, \boldsymbol{Q}_i \in \mathcal{S}^{nm}$ are $nm \times nm$ symmetric matrices, $\boldsymbol{D}, \boldsymbol{E}_i$ are $n \times m$ matrices, $\lambda \in \mathbb{R}_+, k \in \mathbb{N}$ are parameters which control the complexity of $\boldsymbol{X} \in \mathbb{R}^{n \times m}$ by respectively penalizing and constraining its rank, $\mathcal{I}$ denotes the index set of constraints, and $\operatorname{vec}(\cdot)$ denotes the vectorization of the matrix $\boldsymbol{X}$. Note that we write $\operatorname{vec}(\boldsymbol{X}^\top)$ rather than $\operatorname{vec}(\boldsymbol{X})$ to simplify the notation in our relaxations; both formulations are equivalent up to a fixed permutation of the coordinates.

Problem (1) is a very general class of problems: it models unregularized and Frobenius-regularized matrix completion (Candès and Recht 2009) and reduced rank regression (Negahban and Wainwright 2011) as special cases, as we discuss in detail in Section 3. However, to our knowledge, there are no tractable convex relaxations for low-rank quadratic optimization problems like (1) except those that exploit a spectral term in the objective or constraints that may not persist in practice. Accordingly, this paper proposes computationally tractable lifted relaxations of Problem (1).

The closest works to this paper are (i) Kim et al. (2022), who generalize the work of Hiriart-Urruty and Le (2012) to propose an extended formulation for sets of the form $\{\boldsymbol{X} : \mathrm{rank}(\boldsymbol{X}) \leq k,\ f(\boldsymbol{X}) \leq 0\}$, where f is a quasiconvex function in the singular values of $\boldsymbol{X}$, (ii) Bertsimas et al. (2023b), who leverage partial separability to obtain compact semidefinite relaxations for low-rank problems with a spectral term in the objective of the form $\mathrm{tr}(f(\boldsymbol{X}))$ for f matrix convex, via a matrix perspective relaxation, and (iii) Li and Xie (2025), who propose a column generation algorithm for solving the convex relaxation of a low-rank spectrally constrained problem. All three works yield powerful relaxations when their structural assumptions hold, but do not yield generic, scalable convex relaxations for general low-rank quadratic optimization in which the objective and constraints are arbitrary quadratic forms in $\mathrm{vec}(\boldsymbol{X}^\top)$. We complement this literature by providing a lifted SDP relaxation for general quadratic objectives and constraints, and by showing how to systematically eliminate variables to obtain implementable relaxations for low-rank problems.

Main contributions: In this work, we extend the lifted relaxation idea of Shor (1987) to Problem (1) by leveraging the connection between binary and low-rank optimization as explored in (Bertsimas et al. 2022), thus enriching the toolbox of semidefinite relaxations for low-rank optimization. Our main theoretical contribution is a class of new semidefinite relaxations for generic low-rank optimization problems (Proposition 1 and Theorem 1 in Section 2.1). To the best of our knowledge, this is the first work that obtains non-trivial and computationally tractable lower bounds for quadratic low-rank optimization problems like (1), without depending on spectral terms.

Structure: We propose lifted relaxations of low-rank quadratic optimization problems in Section 2. Conscious that these relaxations involve a number of additional semidefinite variables that may be prohibitively large in practice, we show how to eliminate many of these variables in the relaxation without altering its optimal value (Theorem 1). To illustrate our approach, we apply our lifted relaxation to three prominent low-rank optimization problems in Section 3. In particular, on these special cases, we show how to exploit further problem structure and eliminate more variables from our relaxations, making our new relaxation more scalable. Finally, in Section 4, we numerically benchmark our convex relaxations on low-rank matrix completion problems.

1.1. Notation

We let non-boldface characters such as b denote scalars, lowercase boldface characters such as $\boldsymbol{x}$ denote vectors, uppercase boldface characters such as $\boldsymbol{X}$ denote matrices, and calligraphic uppercase characters such as $\mathcal{Z}$ denote sets. We let $[n]$ denote the set of running indices $\{1, ..., n\}$. The cone of $n \times n$ symmetric (resp. positive semidefinite) matrices is denoted by $\mathcal{S}^n$ (resp. $\mathcal{S}^n_+$). Inner products are denoted $\langle \cdot, \cdot \rangle$, and are associated with the Euclidean norm $\|\boldsymbol{x}\|$ for vectors and the Frobenius norm $\|\boldsymbol{X}\|_F$ for matrices.

For a matrix $\boldsymbol{X} \in \mathbb{R}^{n\times m}$, we let $\boldsymbol{X}_i$ denote its ith column and $\boldsymbol{X}_{i,\cdot}$ denote a vector containing its ith row. We let $\mathrm{vec}(\boldsymbol{X}) : \mathbb{R}^{n\times m} \to \mathbb{R}^{nm}$ denote the vectorization operator which maps matrices to vectors by stacking columns. For a matrix $\boldsymbol{W}$, we may find it convenient to describe it as a block matrix composed of equally sized blocks and denote the (i,i') block by $\boldsymbol{W}^{(i,i')}$. The dimension of each block will be clear from the context, given the size of the matrix $\boldsymbol{W}$ and the number of blocks. In particular, $\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}$ with $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ denotes an $nm \times nm$ block-diagonal matrix whose m diagonal blocks are equal to $\boldsymbol{\Sigma}$. With this notation, $\mathrm{vec}(\boldsymbol{\Sigma X}) = (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma})\,\mathrm{vec}(\boldsymbol{X})$.

We let $\boldsymbol{X}^\dagger$ be the pseudoinverse of $\boldsymbol{X}$, which is used in the Schur complement lemma (Boyd et al. 1994, Eqn. 2.41). We let $\mathcal{Y}_n^k := \{\boldsymbol{Y} \in \mathcal{S}_+^n : \boldsymbol{Y}^2 = \boldsymbol{Y}, \mathrm{tr}(\boldsymbol{Y}) \le k\}$ denote the set of orthogonal projection matrices with rank at most k, whose convex hull is $\{\boldsymbol{P} \in \mathcal{S}_+^n : \boldsymbol{P} \preceq \boldsymbol{I}_n, \mathrm{tr}(\boldsymbol{P}) \le k\}$ (Overton and Womersley 1992, Theorem 3). Analogously, we let $\mathcal{Y}_n$ denote the set of $n \times n$ orthogonal projection matrices of any rank, with convex hull $\mathrm{Conv}(\mathcal{Y}_n) = \{\boldsymbol{P} \in \mathcal{S}_+^n : \boldsymbol{P} \preceq \boldsymbol{I}_n\}$. In particular, we have $\mathrm{rank}(\boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{Y})$ for any projection matrix $\boldsymbol{Y}$.

2. Lifted Relaxations for Low-Rank Optimization Problems

In this section, we demonstrate how to apply the lifted relaxation technique to low-rank quadratic optimization problems in a manner that yields tractable convex relaxations.

First, we derive new lifted relaxations for rank-constrained optimization problems (§2.1). Interestingly, we show that many variables in our lifted relaxations can be omitted without altering the objective value, yielding a more compact and tractable formulation. Compared with Bertsimas et al. (2023b), we show that our new relaxations are stronger and more broadly applicable. Second, we discuss how common ideas in logically constrained optimization, such as the reformulation-linearization technique (RLT Adams and Sherali 1986), can be generalized to low-rank optimization to further strengthen our relaxations (§2.2).

2.1. A New Lifted Relaxation and Its Compact Formulation

We study a quadratic low-rank optimization problem with linear constraints, which encompasses low-rank matrix completion (Candès and Recht 2009), and reduced rank regression (Negahban and Wainwright 2011) problems, among others; see Bertsimas et al. (2022) for a review of low-rank optimization. Formally, we study the following reformulation of Problem (1):

$$\begin{aligned}
\min_{\boldsymbol{Y} \in \mathcal{Y}_n^k} \min_{\boldsymbol{X} \in \mathbb{R}^{n\times m}} \quad & \lambda \cdot \mathrm{tr}(\boldsymbol{Y}) + \left\langle \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top, \boldsymbol{H} \right\rangle + \langle \boldsymbol{D}, \boldsymbol{X} \rangle \\
\text{s.t.} \quad & \left\langle \boldsymbol{Q}_i, \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top \right\rangle + \langle \boldsymbol{E}_i, \boldsymbol{X} \rangle \le b_i, \ \forall i \in \mathcal{I}, \ \boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}.
\end{aligned} \tag{2}$$

As demonstrated[1] in Bertsimas et al. (2022), any rank-constrained optimization problem of the form (1) can be formulated as an optimization over $(\boldsymbol{X}, \boldsymbol{Y})$ of the form (2), where the additional decision variable $\boldsymbol{Y}$ is a projection matrix which encodes the span of $\boldsymbol{X}$ and whose trace bounds $\mathrm{rank}(\boldsymbol{X})$. However, unlike Problem (1), Problem (2) concentrates all nonconvexity within the quadratic constraint $\boldsymbol{Y}^2 = \boldsymbol{Y}$, thus making it especially amenable to the lifted relaxation technique.

We now develop a convex relaxation of (2). Specifically, we introduce matrices $\boldsymbol{W}_{x,x}$, $\boldsymbol{W}_{x,y}$, $\boldsymbol{W}_{y,y}$ to model the outer products $\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$, $\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{Y})^\top$, and $\mathrm{vec}(\boldsymbol{Y})\mathrm{vec}(\boldsymbol{Y})^\top$:

Proposition 1 **(Lifted relaxation)**. *The convex semidefinite optimization problem*

$$
\begin{aligned}
\min_{\substack{\boldsymbol{Y}\in\mathcal{S}^n_+ : \boldsymbol{Y}\preceq \boldsymbol{I}, \mathrm{tr}(\boldsymbol{Y})\le k \\ \boldsymbol{W}_{y,y}\in\mathcal{S}^{n^2}_+}} \quad \min_{\substack{\boldsymbol{X}\in\mathbb{R}^{n\times m}, \\ \boldsymbol{W}_{x,x}\in\mathcal{S}^{nm}_+, \boldsymbol{W}_{x,y}\in\mathbb{R}^{nm\times n^2}}} \quad & \lambda\cdot\mathrm{tr}(\boldsymbol{Y}) + \langle \boldsymbol{W}_{x,x}, \boldsymbol{H}\rangle + \langle \boldsymbol{D}, \boldsymbol{X}\rangle \\
\text{s.t.} \quad & \begin{pmatrix} 1 & \mathrm{vec}(\boldsymbol{X}^\top)^\top & \mathrm{vec}(\boldsymbol{Y})^\top \\ \mathrm{vec}(\boldsymbol{X}^\top) & \boldsymbol{W}_{x,x} & \boldsymbol{W}_{x,y} \\ \mathrm{vec}(\boldsymbol{Y}) & \boldsymbol{W}_{x,y}^\top & \boldsymbol{W}_{y,y} \end{pmatrix} \succeq \boldsymbol{0}, \\
& \sum_{i=1}^n \boldsymbol{W}_{y,y}^{(i,i)} = \boldsymbol{Y}, \ \sum_{i=1}^n \boldsymbol{W}_{x,y}^{(i,i)} = \boldsymbol{X}^\top, \\
& \langle \boldsymbol{Q}_i, \boldsymbol{W}_{x,x}\rangle + \langle \boldsymbol{E}_i, \boldsymbol{X}\rangle \le b_i \ \forall i \in \mathcal{I},
\end{aligned} \tag{3}
$$

is a valid convex relaxation of Problem (2).

Remark 1. If an optimal solution to (3) is such that $\boldsymbol{W}_{x,x}$ is a rank-one matrix then $\boldsymbol{W}_{x,x} = \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$ and the optimal values of (3) and (2) coincide.

Proof of Proposition 1 Fix $(\boldsymbol{X}, \boldsymbol{Y})$ in (2) and set

$$(\boldsymbol{W}_{x,x}, \boldsymbol{W}_{x,y}, \boldsymbol{W}_{y,y}) := (\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top, \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{Y})^\top, \mathrm{vec}(\boldsymbol{Y})\mathrm{vec}(\boldsymbol{Y})^\top).$$

It is sufficient to verify that $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}_{x,x}, \boldsymbol{W}_{x,y}, \boldsymbol{W}_{y,y})$ is feasible for (3)—it obviously attains the same objective value. First, by construction, the semidefinite constraint is satisfied (at equality). Moreover, we have

$$
\begin{aligned}
\boldsymbol{Y}\boldsymbol{Y}^\top = \boldsymbol{Y} &\implies \sum_{i=1}^n \boldsymbol{W}_{y,y}^{(i,i)} = \boldsymbol{Y}, \\
\boldsymbol{X}^\top\boldsymbol{Y}^\top = \boldsymbol{X}^\top &\implies \sum_{i\in[n]} \boldsymbol{W}_{x,y}^{(i,i)} = \boldsymbol{X}^\top. \quad \square
\end{aligned}
$$

Unfortunately, (3) is not compact as it involves one semidefinite constraint of dimension $n^2 + nm + 1$. a natural research question is whether it is possible to eliminate any variables from (3) without altering its optimal objective value. We answer this question affirmatively.

Theorem 1 **(Elimination theorem).** *Problem* (3) *is equivalent to*

$$\begin{aligned}\min_{\boldsymbol{Y}\in\mathcal{S}^n_+:\boldsymbol{Y}\preceq\boldsymbol{I},\ \mathrm{tr}(\boldsymbol{Y})\le k}\quad \min_{\substack{\boldsymbol{X}\in\mathbb{R}^{n\times m}\\ \boldsymbol{W}_{x,x}\in\mathcal{S}^{nm}_+}}\quad & \lambda\cdot\mathrm{tr}(\boldsymbol{Y})+\langle \boldsymbol{W}_{x,x},\boldsymbol{H}\rangle+\langle\boldsymbol{D},\boldsymbol{X}\rangle\\ s.t.\quad & \boldsymbol{W}_{x,x}\succeq \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top,\\ & \langle\boldsymbol{Q}_i,\boldsymbol{W}_{x,x}\rangle+\langle\boldsymbol{E}_i,\boldsymbol{X}\rangle\le b_i,\ \forall i\in\mathcal{I},\\ & \begin{pmatrix}\sum_{i\in[n]}\boldsymbol{W}^{(i,i)}_{x,x} & \boldsymbol{X}^\top\\ \boldsymbol{X} & \boldsymbol{Y}\end{pmatrix}\succeq\boldsymbol{0}.\end{aligned}\tag{4}$$

Remark 2. As we demonstrate later in this section (Example 1), Theorem 1's equivalence result relies on the off-diagonal blocks $\boldsymbol{W}_{x,y}$ and $\boldsymbol{W}_{y,y}$ appearing only inside the block PSD matrix in (3). Constraints on the off-diagonal blocks $\boldsymbol{W}_{x,y}$ and $\boldsymbol{W}_{y,y}$ may render them non-eliminable.

Proof of Theorem 1 We show that given a feasible solution to either problem, we can generate an optimal solution to the other problem with an equal or lower objective value.

Suppose that $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W}_{x,x},\boldsymbol{W}_{x,y},\boldsymbol{W}_{y,y})$ is feasible in (3). Then, by summing the PSD principal submatrices containing $\boldsymbol{W}^{(i,i)}_{x,x},\boldsymbol{W}^{(i,i)}_{x,y},\boldsymbol{W}^{(i,i)}_{y,y}$ for each $i\in[n]$, we have that

$$\begin{pmatrix}\sum_{i\in[n]}\boldsymbol{W}^{(i,i)}_{x,x} & \sum_{i\in[n]}\boldsymbol{W}^{(i,i)}_{x,y}\\ \sum_{i\in[n]}\boldsymbol{W}^{(i,i)\top}_{x,y} & \sum_{i\in[n]}\boldsymbol{W}^{(i,i)}_{y,y}\end{pmatrix}\succeq\boldsymbol{0}.$$

Moreover, from (3) we have that $\sum_{i\in[n]}\boldsymbol{W}_{x,y}{}^{(i,i)}=\boldsymbol{X}^\top$ and $\sum_{i\in[n]}\boldsymbol{W}^{(i,i)}_{y,y}=\boldsymbol{Y}$. Thus, $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W}_{x,x})$ is feasible in (4) and attains the same objective value.

Next, suppose that $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W}_{x,x})$ is feasible in (4). By the Schur complement lemma, we must have $\boldsymbol{Y}\succeq\boldsymbol{X}(\sum_i\boldsymbol{W}^{(i,i)}_{x,x})^\dagger\boldsymbol{X}^\top$. Since the cost matrix associated with $\boldsymbol{Y}$ in the objective, $\lambda\boldsymbol{I}$, is positive semidefinite, we can set $\boldsymbol{Y}:=\boldsymbol{X}(\sum_i\boldsymbol{W}^{(i,i)}_{x,x})^\dagger\boldsymbol{X}^\top$ without loss of optimality—doing so cannot increase the objective value nor make the constraints $\boldsymbol{0}\preceq\boldsymbol{Y}\preceq\boldsymbol{I}_n$ and $\mathrm{tr}(\boldsymbol{Y})\le k$ violated. To construct admissible matrices $\boldsymbol{W}_{x,y}$ and $\boldsymbol{W}_{y,y}$, let us first define the auxiliary matrix

$$\boldsymbol{U}:=\left(\sum_{i\in[n]}\boldsymbol{W}^{(i,i)}_{x,x}\right)^\dagger\boldsymbol{X}^\top\in\mathbb{R}^{m\times n},$$

and observe that $\boldsymbol{Y}=\boldsymbol{U}^\top\boldsymbol{X}^\top=\boldsymbol{X}\boldsymbol{U}$. Then, we define the matrix

$$\boldsymbol{M}:=\begin{pmatrix}1&\boldsymbol{0}&\boldsymbol{0}\\ \boldsymbol{0}&\boldsymbol{I}_{nm}&\boldsymbol{0}\\ \boldsymbol{0}&\boldsymbol{0}&\boldsymbol{I}_n\otimes\boldsymbol{U}\end{pmatrix}^\top\begin{pmatrix}1&\mathrm{vec}(\boldsymbol{X}^\top)^\top&\mathrm{vec}(\boldsymbol{X}^\top)^\top\\ \mathrm{vec}(\boldsymbol{X}^\top)&\boldsymbol{W}_{x,x}&\boldsymbol{W}_{x,x}\\ \mathrm{vec}(\boldsymbol{X}^\top)&\boldsymbol{W}_{x,x}&\boldsymbol{W}_{x,x}\end{pmatrix}\begin{pmatrix}1&\boldsymbol{0}&\boldsymbol{0}\\ \boldsymbol{0}&\boldsymbol{I}_{nm}&\boldsymbol{0}\\ \boldsymbol{0}&\boldsymbol{0}&\boldsymbol{I}_n\otimes\boldsymbol{U}\end{pmatrix}.$$

We let $\boldsymbol{W}_{x,y}$, $\boldsymbol{W}_{y,y}$ denote the relevant off-diagonal blocks, i.e.,

$$\boldsymbol{M}=:\begin{pmatrix}1&\mathrm{vec}(\boldsymbol{X}^\top)^\top&\mathrm{vec}(\boldsymbol{Y})^\top\\ \mathrm{vec}(\boldsymbol{X}^\top)&\boldsymbol{W}_{x,x}&\boldsymbol{W}_{x,y}\\ \mathrm{vec}(\boldsymbol{Y})&\boldsymbol{W}_{x,y}^\top&\boldsymbol{W}_{y,y}\end{pmatrix}.$$

Since $\boldsymbol{Y}=\boldsymbol{U}^\top\boldsymbol{X}^\top$, we have $\mathrm{vec}(\boldsymbol{Y})=\mathrm{vec}(\boldsymbol{U}^\top\boldsymbol{X}^\top)=(\boldsymbol{I}_n\otimes\boldsymbol{U}^\top)\mathrm{vec}(\boldsymbol{X}^\top)$ and thus our construction is consistent with the existing value of $\boldsymbol{Y}$. We now verify that $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W}_{x,x},\boldsymbol{W}_{x,y},\boldsymbol{W}_{y,y})$ is feasible for

(3). By construction, $\boldsymbol{M} \succeq \boldsymbol{0}$. Thus, $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}_{x,x}, \boldsymbol{W}_{x,y}, \boldsymbol{W}_{y,y})$ satisfies the semidefinite constraint in (3). Next, by construction, $\boldsymbol{W}_{x,y}$ and $\boldsymbol{W}_{y,y}$ can be decomposed into $n \times n$ blocks satisfying:

$$\boldsymbol{W}_{x,y}^{(i,j)} = \boldsymbol{W}_{x,x}^{(i,j)} \boldsymbol{U}, \ \boldsymbol{W}_{y,y}^{(i,j)} = \boldsymbol{U}^\top \boldsymbol{W}_{x,x}^{(i,j)} \boldsymbol{U}.$$

Summing the on-diagonal blocks of these matrices then reveals that

$$\sum_{i \in [n]} \boldsymbol{W}_{x,y}^{(i,i)} = \sum_{i \in [n]} \boldsymbol{W}_{x,x}^{(i,i)} \boldsymbol{U} = \left(\sum_{i \in [n]} \boldsymbol{W}_{x,x}^{(i,i)}\right) \left(\sum_{j \in [n]} \boldsymbol{W}_{x,x}^{(j,j)}\right)^\dagger \boldsymbol{X}^\top = \boldsymbol{X}^\top,$$

$$\sum_{i \in [n]} \boldsymbol{W}_{y,y}^{(i,i)} = \sum_{i \in [n]} \boldsymbol{U}^\top \boldsymbol{W}_{x,x}^{(i,i)} \boldsymbol{U} = \boldsymbol{U}^\top \left(\sum_{i \in [n]} \boldsymbol{W}_{x,x}^{(i,i)} \boldsymbol{U}\right) = \boldsymbol{U}^\top \boldsymbol{X}^\top = \boldsymbol{Y}.$$

Therefore, we conclude that $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}_{x,x}, \boldsymbol{W}_{x,y}, \boldsymbol{W}_{y,y})$ is feasible in (3) and attains an equal or lower objective value. Thus, both relaxations are equivalent. □

Problem (4) is much more compact than (3), as it does not require introducing the variables $\boldsymbol{W}_{y,y} \in \mathcal{S}_+^{n^2}$ or $\boldsymbol{W}_{x,y} \in \mathbb{R}^{nm \times n^2}$. Instead, (4) only involves two semidefinite constraints of dimension $nm+1$ and $n+m$ respectively. The proof of Theorem 1 provides a recipe for reconstructing an optimal $\boldsymbol{W}_{y,y}$ given an optimal solution $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{W}_{x,x})$ to (4). Namely, compute the auxiliary matrix $\boldsymbol{U} := \left(\sum_{i \in [n]} \boldsymbol{W}_{x,x}^{(i,i)}\right)^\dagger \boldsymbol{X}^\top$ and set $\boldsymbol{W}_{y,y} := (\boldsymbol{I}_n \otimes \boldsymbol{U})^\top \boldsymbol{W}_{x,x} (\boldsymbol{I}_n \otimes \boldsymbol{U})$.

Finally, it is interesting to consider whether the relaxation developed here is at least as strong as the matrix perspective relaxation developed by Bertsimas et al. (2023b). We now prove this is indeed the case. Bertsimas et al. (2023b) only applies to partially separable objectives. Hence, we first need to impose more structure on the objective of (2) to compare relaxations.

Proposition 2. *Assume that the term $\langle \boldsymbol{H}, \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top \rangle + \langle \boldsymbol{D}, \boldsymbol{X} \rangle$ in Problem (2) can be rewritten as the partially separable term $\frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2 + h(\boldsymbol{X})$, where h is convex in $\boldsymbol{X}$. Further, assume the terms $\langle \boldsymbol{Q}_i, \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top \rangle + \langle \boldsymbol{E}_i, \boldsymbol{X} \rangle$ in Problem (2) can be rewritten as the partially separable terms $\frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2 + h_i(\boldsymbol{X})$, where h_i are convex in $\boldsymbol{X}$. Then, the optimal value of Problem (3) is at least as large as the relaxation of Bertsimas et al. (2023b)*

$$\min_{\boldsymbol{Y} \in \mathrm{Conv}\left(\mathcal{Y}_n^k\right)} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}, \boldsymbol{\theta} \in \mathcal{S}_+^m} \quad \lambda \cdot \mathrm{tr}(\boldsymbol{Y}) + \frac{1}{2\gamma}\mathrm{tr}(\boldsymbol{\theta}) + h(\boldsymbol{X}) \tag{5}$$
$$\text{s.t.} \quad \frac{1}{2\gamma}\mathrm{tr}(\boldsymbol{\theta}) + h_i(\boldsymbol{X}) \leq b_i, \ \forall i \in \mathcal{I}, \quad \begin{pmatrix} \boldsymbol{\theta} & \boldsymbol{X}^\top \\ \boldsymbol{X} & \boldsymbol{Y} \end{pmatrix} \succeq \boldsymbol{0},$$

Proof of Proposition 2 Given the equivalence between Problems (3)–(4) proven in Theorem 1, it suffices to show that the constraints in (4) imply the constraints in (5). Letting $\boldsymbol{\theta} := \sum_{i \in [n]} \boldsymbol{W}_{x,x}^{(i,i)}$, we observe that $\boldsymbol{\theta}$ is feasible for (5), which completes the proof. □

The proof of Proposition 2 reveals that our lifted relaxation (4) can be perceived as decomposing the variable $\boldsymbol{\theta}$ in (5), and strengthening the relaxation by imposing additional constraints on the elements of this decomposition.

2.2. Strategies for Strengthening the Lifted Relaxation

Theorem 1 might give the impression that Problem (3) is not a useful relaxation, because it is equivalent to the more compact optimization problem (4). However, this is arguably unfair, because explicit decision variables $\boldsymbol{W}_{y,y}, \boldsymbol{W}_{x,y}$ allow us to express additional valid inequalities to strengthen the relaxation. Indeed, as we demonstrate by example in this section, these inequalities sometimes allow (3) to be strictly tighter than (4), when strengthened with additional inequalities:

Symmetry constraints: The matrix $\boldsymbol{Y}$ being symmetric, $\mathrm{vec}(\boldsymbol{Y}) = \mathrm{vec}(\boldsymbol{Y}^\top) = \boldsymbol{K}_{n,n}\,\mathrm{vec}(\boldsymbol{Y})$ for the commutation matrix $\boldsymbol{K}_{n,n} \in \mathbb{R}^{n^2 \times n^2}$ (Magnus and Neudecker 1979), which leads to the constraints

$$\begin{aligned}
\mathrm{vec}(\boldsymbol{Y})\,\mathrm{vec}(\boldsymbol{Y})^\top = \boldsymbol{K}_{n,n}\,\mathrm{vec}(\boldsymbol{Y})\,\mathrm{vec}(\boldsymbol{Y})^\top \boldsymbol{K}_{n,n}^\top &\implies \boldsymbol{W}_{y,y} = \boldsymbol{K}_{n,n}\boldsymbol{W}_{y,y}\boldsymbol{K}_{n,n}^\top,\\
\mathrm{vec}(\boldsymbol{X}^\top)\,\mathrm{vec}(\boldsymbol{Y})^\top = \mathrm{vec}(\boldsymbol{X}^\top)\,\mathrm{vec}(\boldsymbol{Y})^\top \boldsymbol{K}_{n,n}^\top &\implies \boldsymbol{W}_{x,y} = \boldsymbol{W}_{x,y}\boldsymbol{K}_{n,n}^\top.
\end{aligned} \tag{6}$$

Moreover, if we further require the matrix $\boldsymbol{X}$ to be symmetric (implying $n = m$), then we can impose the additional linear equalities $\boldsymbol{W}_{x,x} = \boldsymbol{K}_{n,n}\boldsymbol{W}_{x,x}\boldsymbol{K}_{n,n}^\top$ and $\boldsymbol{W}_{x,y} = \boldsymbol{K}_{n,n}\boldsymbol{W}_{x,y}$.

Triangle inequalities: As in binary optimization, we can impose triangle inequalities on $\boldsymbol{Y}$ and $\boldsymbol{W}_{yy}$. Indeed, from the fact that $0 \le Y_{i,i} \le 1$ (which is a consequence of $\boldsymbol{0} \preceq \boldsymbol{Y} \preceq \boldsymbol{I}_n$), we have that any triplet (i, j, ℓ) satisfies

$$\begin{aligned}
&(1 - Y_{i,i})(1 - Y_{j,j})(1 - Y_{\ell,\ell}) \ge 0\\
\iff &1 - Y_{i,i} - Y_{j,j} - Y_{\ell,\ell} + Y_{i,i}Y_{j,j} + Y_{i,i}Y_{\ell,\ell} + Y_{j,j}Y_{\ell,\ell} - Y_{i,i}Y_{j,j}Y_{\ell,\ell} \ge 0\\
\implies &1 - Y_{i,i} - Y_{j,j} - Y_{\ell,\ell} + Y_{i,i}Y_{j,j} + Y_{i,i}Y_{\ell,\ell} + Y_{j,j}Y_{\ell,\ell} \ge 0,
\end{aligned}$$

which can be expressed as a linear constraint in $(\boldsymbol{Y}, \boldsymbol{W}_{yy})$ after replacing each bilinear term with the appropriate entry of $\boldsymbol{W}_{yy}$. We can derive additional triangle inequalities by starting from the fact that $Y_{i,i}(1 - Y_{j,j})(1 - Y_{\ell,\ell}) \ge 0$ or $Y_{i,i}Y_{j,j}(1 - Y_{\ell,\ell}) \ge 0$. Triangle inequalities involving $Y_{i,j} \in [-1, 1]$ rather than $Y_{i,i}$ follow similarly.

RLT inequalities: Finally, as binary quadratic optimization (Sherali and Alameddine 1992), one can tighten Problem (3) and Problem (4) by applying RLT. Let $\boldsymbol{x} = \mathrm{vec}(\boldsymbol{X}^\top)$. Then, any constraint of the form $\boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}$ leads to the valid inequalities $\boldsymbol{A}\boldsymbol{W}_{x,x}\boldsymbol{A}^\top + \boldsymbol{b}\boldsymbol{b}^\top \ge \boldsymbol{b}\boldsymbol{x}^\top\boldsymbol{A}^\top + \boldsymbol{A}\boldsymbol{x}\boldsymbol{b}^\top$, as reviewed by Bao et al. (2011).[2]

We now support our discussion by providing an example from low-rank matrix completion, which demonstrates that Problem (3) with additional symmetry constraints (6) is strictly stronger than the more compact optimization problem (4), and that both are stronger than the matrix perspective relaxation of Bertsimas et al. (2023b).

Example 1. Consider a low-rank matrix completion problem of the form

$$\min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}} \quad \frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2 + \frac{1}{2}\sum_{(i,j)\in\Omega}(X_{i,j} - A_{i,j})^2 \text{ s.t. } \mathrm{rank}(\boldsymbol{X}) \le k.$$

Let the problem data be $\gamma = 100, k = 2, n = 7, m = 5$, and suppose we are trying to impute the following matrix, where $*$ denotes a missing entry:

$$\boldsymbol{A} = \begin{pmatrix} -2 & * & -1 & 1 & -1 \\ * & 4 & -4 & -5 & -4 \\ * & -3 & 1 & 4 & 3 \\ 3 & 5 & -5 & -5 & -1 \\ 7 & 8 & -10 & -8 & 1 \\ 3 & 1 & -2 & * & 5 \\ 7 & 7 & -13 & -8 & * \end{pmatrix}.$$

Then (using Mosek version 10.2 to solve all semidefinite relaxations):

- The relaxation of Bertsimas et al. (2023b) has an optimal objective value of 3.9275.
- The semidefinite relaxation (3) augmented with symmetry constraints (6) has an optimal objective value of 5.1387.
- The more compact semidefinite relaxation (4) has an objective value of 4.314.

In spite of Example 1, Theorem 1 is useful because it produces a non-trivial lower bound after solving a smaller semidefinite problem. Accordingly, we now explore problem-specific decompositions to make Problem (4) even more compact without altering its optimal value.

3. Examples of Low-Rank Relaxations

This section applies the lifted relaxation proposed in Section 2 to several important problems from the low-rank literature. By exploiting the problem structure, we show that it is often possible to further reduce our lifted relaxation. In particular, in all these examples, we show that our lifted relaxation can be reduced to an equivalent relaxation with no semidefinite matrices or constraints of dimension larger than $2\max(n,m)$, as opposed to the $O((n+m)^2)$ dimension blocks in our lifted relaxations in the previous section.

3.1. Matrix Completion

Given a random sample $\{A_{i,j} : (i,j) \in \Omega \subseteq [n] \times [m]\}$ of a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$, the goal of the low-rank matrix completion problem is to reconstruct the matrix $\boldsymbol{A}$, by assuming it is approximately low-rank (Candès and Recht 2009). This problem admits the formulation:

$$\min_{\boldsymbol{Y} \in \mathcal{Y}_n^k} \min_{\boldsymbol{X} \in \mathbb{R}^{n\times m}} \quad \|\mathcal{P}(\boldsymbol{A}) - \mathcal{P}(\boldsymbol{X})\|_F^2 + \lambda \cdot \mathrm{tr}(\boldsymbol{Y}) \quad \text{s.t.} \quad \boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}, \tag{7}$$

where $\lambda \geq 0$ is a penalty multiplier on the rank of $\boldsymbol{X}$ through the trace of $\boldsymbol{Y}$, k is a hard constraint on the rank of $\boldsymbol{X}$ through the trace of $\boldsymbol{Y}$, and

$$\mathcal{P}(\boldsymbol{A})_{i,j} = \begin{cases} A_{i,j} & \text{if } (i,j) \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

is a linear map which masks the hidden entries of $\boldsymbol{A}$ by outputting a matrix equal to $\boldsymbol{A}$ on Ω and 0 otherwise. By expanding the quadratic $\|\mathcal{P}(\boldsymbol{A}) - \mathcal{P}(\boldsymbol{X})\|_F^2 = \sum_{i\in[n]}\left(\sum_{j\in[m]:(i,j)\in\Omega}(X_{i,j}-A_{i,j})^2\right)$, we can invoke Theorem 1 to obtain the following relaxation of (7)

$$\begin{aligned}\min_{\boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n^k)}\ \min_{\boldsymbol{X}\in\mathbb{R}^{n\times m},\boldsymbol{W}\in\mathcal{S}_+^{nm}} \quad & \sum_{i\in[n]}\langle \boldsymbol{W}^{(i,i)},\boldsymbol{H}^i\rangle - 2\langle\mathcal{P}(\boldsymbol{X}),\mathcal{P}(\boldsymbol{A})\rangle + \langle\mathcal{P}(\boldsymbol{A}),\mathcal{P}(\boldsymbol{A})\rangle + \lambda\cdot\mathrm{tr}(\boldsymbol{Y})\\ \text{s.t.}\quad & \boldsymbol{W}\succeq \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top,\ \begin{pmatrix}\sum_{i\in[n]}\boldsymbol{W}^{(i,i)}_{x,x} & \boldsymbol{X}^\top\\ \boldsymbol{X} & \boldsymbol{Y}\end{pmatrix}\succeq\boldsymbol{0},\end{aligned} \tag{8}$$

where $\boldsymbol{H}^i$ is a diagonal matrix which takes entries $\boldsymbol{H}^i_{j,j}=1$ if $(i,j)\in\Omega$ and $\boldsymbol{H}^i_{j,j}=0$ otherwise, and we retain the constant term $\langle\mathcal{P}(\boldsymbol{A}),\mathcal{P}(\boldsymbol{A})\rangle$ to ensure that all relaxations are directly comparable.

Compared with the matrix perspective relaxation of Bertsimas et al. (2023b), our relaxation is directly applicable to (7), while Bertsimas et al. (2023b) requires the presence of an additional Frobenius regularization term $+\frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2$ in the objective. With this additional term, our approach leads to relaxations of the form (8) after redefining $\boldsymbol{H}_i \leftarrow \boldsymbol{H}_i + \frac{1}{2\gamma}\boldsymbol{I}_m$, which are at least as strong as the relaxation of Bertsimas et al. (2023b) per Proposition 2.

We observe that the off-diagonal blocks of $\boldsymbol{W}$ do not appear in either the objective of (8) or any constraints other than $\boldsymbol{W}\succeq \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$. For this reason, we can omit them entirely:

Proposition 3. *Problem* (8) *attains the same optimal objective value as*

$$\begin{aligned}\min_{\boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n^k)}\ \min_{\boldsymbol{X}\in\mathbb{R}^{n\times m},\boldsymbol{S}^i\in\mathcal{S}_+^{m}} \quad & \sum_{i\in[n]}\langle \boldsymbol{S}^i,\boldsymbol{H}^i\rangle - 2\,\langle\mathcal{P}(\boldsymbol{X}),\mathcal{P}(\boldsymbol{A})\rangle + \langle\mathcal{P}(\boldsymbol{A}),\mathcal{P}(\boldsymbol{A})\rangle + \lambda\cdot\mathrm{tr}(\boldsymbol{Y})\\ s.t.\quad & \boldsymbol{S}^i\succeq \boldsymbol{X}_{i,.}\boldsymbol{X}_{i,.}^\top,\ \begin{pmatrix}\sum_{i\in[n]}\boldsymbol{S}^i & \boldsymbol{X}^\top\\ \boldsymbol{X} & \boldsymbol{Y}\end{pmatrix}\succeq\boldsymbol{0}.\end{aligned} \tag{9}$$

Proof of Proposition 3 It suffices to show that given any feasible solution to (9) we can construct a feasible solution to (8) with the same objective value; the converse is immediate. Let $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{S}^i)$ be feasible in (9). Define the block matrix $\boldsymbol{W}\in\mathcal{S}_+^{nm\times nm}$ by setting $\boldsymbol{W}^{(i,i)}=\boldsymbol{S}^i$ and $\boldsymbol{W}^{(i,j)}=\boldsymbol{X}_{i,.}\boldsymbol{X}_{j,.}^\top$. Then, it is not hard to see that $\boldsymbol{W}-\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$ is a block matrix with zero off-diagonal blocks and on-diagonal blocks $\boldsymbol{S}^i-\boldsymbol{X}_{i,.}\boldsymbol{X}_{i,.}^\top\succeq\boldsymbol{0}$. Thus, $\boldsymbol{W}-\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$ is a positive semidefinite matrix, and $\boldsymbol{W}\succeq\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$. Moreover, $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W})$ is feasible in (8) and attains the same objective value. □

Compared with (8), the semidefinite relaxation (9) involves n positive semidefinite variables of dimension m (vs. one PSD matrix of dimension nm in (8)) and $n+1$ semidefinite constraints of dimension $m+1$ (vs. one of dimension $nm+1$)—both relaxations involve one semidefinite constraints of dimension $n+m$.

Remark 3. Let $\boldsymbol{H}^1,\ldots,\boldsymbol{H}^G$ be the *distinct* diagonal masks appearing among $\{\boldsymbol{H}^i\}_{i\in[n]}$, and let

$$\mathcal{I}_g := \{\, i\in[n] : \boldsymbol{H}^i = \boldsymbol{H}^g \,\},\qquad g=1,\ldots,G,$$

so that $\{\mathcal{I}_g\}_{g=1}^G$ forms a partition of $[n]$. Introduce one aggregated semidefinite variable $\boldsymbol{S}^g \in \mathbb{S}^m_+$ per group, which represents $\boldsymbol{S}^g = \sum_{i \in \mathcal{I}_g} \boldsymbol{S}^i$. Then the relaxation (9) can be rewritten as the more compact relaxation:

$$\begin{aligned} \min_{\boldsymbol{Y} \in \mathrm{Conv}(\mathcal{Y}_n^k)} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}, \boldsymbol{S}^g \in \mathcal{S}^m_+} \quad & \sum_{g=1}^G \langle \boldsymbol{S}^g, \boldsymbol{H}^g \rangle - 2\langle \mathcal{P}(\boldsymbol{X}), \mathcal{P}(\boldsymbol{A}) \rangle + \langle \mathcal{P}(\boldsymbol{A}), \mathcal{P}(\boldsymbol{A}) \rangle + \lambda \cdot \mathrm{tr}(Y) \\ \text{s.t.} \quad & \boldsymbol{S}^g \succeq \sum_{i \in \mathcal{I}_g} \boldsymbol{X}_{i,.} \boldsymbol{X}_{i,.}^\top, \; g = 1, \ldots, G, \; \begin{pmatrix} \sum_{g=1}^G \boldsymbol{S}^g & \boldsymbol{X}^\top \\ \boldsymbol{X} & \boldsymbol{Y} \end{pmatrix} \succeq 0. \end{aligned} \tag{10}$$

Moreover, if rows of $\boldsymbol{A}$ have a similar sparsity pattern but with minor differences, we can obtain a computationally cheaper yet looser relaxation by masking all entries that do not appear in both rows, i.e., setting $\boldsymbol{H}^i, \boldsymbol{H}^j$ to both be equal to the matrix product $\boldsymbol{H}^i \boldsymbol{H}^j$ and proceeding as above.

3.2. Reduced Rank Regression

Given a response matrix $\boldsymbol{B} \in \mathbb{R}^{n \times m}$ and a predictor matrix $\boldsymbol{A} \in \mathbb{R}^{n \times p}$, an important problem in high-dimensional statistics is to recover a low-complexity model which relates the matrices $\boldsymbol{B}$ and $\boldsymbol{A}$. A popular choice for doing so is to assume that $\boldsymbol{B}, \boldsymbol{A}$ are related via $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{E}$, where $\boldsymbol{X} \in \mathbb{R}^{p \times m}$ is a coefficient matrix, $\boldsymbol{E}$ is a matrix of noise, and we require that the rank of $\boldsymbol{X}$ is small so that the linear model is parsimonious (Negahban and Wainwright 2011). This gives:

$$\min_{\boldsymbol{X} \in \mathbb{R}^{p \times m}} \quad \|\boldsymbol{B} - \boldsymbol{A}\boldsymbol{X}\|_F^2 + \mu \cdot \mathrm{rank}(\boldsymbol{X}), \tag{11}$$

where $\mu > 0$ controls the complexity of the estimator. For this problem, our lifted relaxation (4) is equivalent to the (improved) matrix perspective relaxation of Bertsimas et al. (2023b).

Indeed, by invoking Theorem 1–on $\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^\top$ rather than the mathematically equivalent $\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$ for notational convenience–we obtain (11)'s lifted relaxation

$$\begin{aligned} \min_{\boldsymbol{Y} \in \mathrm{Conv}(\mathcal{Y}_m)} \min_{\boldsymbol{X} \in \mathbb{R}^{p \times m}, \boldsymbol{W} \in \mathcal{S}^{pm}_+} \quad & \left\langle \boldsymbol{A}^\top \boldsymbol{A}, \sum_{i \in [m]} \boldsymbol{W}^{(i,i)} \right\rangle + \langle \boldsymbol{B}, \boldsymbol{B} \rangle - 2\langle \boldsymbol{A}\boldsymbol{X}, \boldsymbol{B} \rangle + \mu \cdot \mathrm{tr}(\boldsymbol{Y}) \\ \text{s.t.} \quad & \boldsymbol{W} \succeq \mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^\top, \; \begin{pmatrix} \sum_{i \in [m]} \boldsymbol{W}^{(i,i)} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{Y} \end{pmatrix} \succeq \boldsymbol{0}, \end{aligned} \tag{12}$$

for which we show the following equivalence result:

Proposition 4. *Problem* (12) *attains the same objective value as*

$$\begin{aligned} \min_{\boldsymbol{Y} \in \mathrm{Conv}(\mathcal{Y}_m)} \min_{\boldsymbol{X} \in \mathbb{R}^{p \times m}, \boldsymbol{\theta} \in \mathcal{S}^p_+} \quad & \langle \boldsymbol{A}^\top \boldsymbol{A}, \boldsymbol{\theta} \rangle + \langle \boldsymbol{B}, \boldsymbol{B} \rangle - 2\langle \boldsymbol{A}\boldsymbol{X}, \boldsymbol{B} \rangle + \mu \cdot \mathrm{tr}(\boldsymbol{Y}) \\ s.t. \quad & \begin{pmatrix} \boldsymbol{\theta} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{Y} \end{pmatrix} \succeq \boldsymbol{0}, \end{aligned} \tag{13}$$

which corresponds to the improved relaxation of Bertsimas et al. (2023b, Equation 7)

Proof of Proposition 4 We show that for any solution to (13) one can construct a solution to (12) with the same objective value or vice versa. Indeed, for any feasible solution $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{W})$ to

(12), $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta} = \sum_{i\in[m]} \boldsymbol{W}^{(i,i)})$ is feasible for (13) with the same objective value. Conversely, let us consider $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\theta})$ a feasible solution to (13). Then,

$$\begin{pmatrix} \boldsymbol{\theta} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{I}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{Y} \end{pmatrix} + \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0}^\top & \boldsymbol{I}_m - \boldsymbol{Y} \end{pmatrix} \succeq \boldsymbol{0},$$

because both matrices are PSD given that $\boldsymbol{Y} \preceq \boldsymbol{I}_m$. Therefore, it follows from the Schur complement lemma that $\boldsymbol{\theta} \succeq \boldsymbol{X}\boldsymbol{X}^\top = \sum_{i\in[m]} \boldsymbol{X}_i \boldsymbol{X}_i^\top$. Thus, we can set $\boldsymbol{S}^i := \boldsymbol{X}_i\boldsymbol{X}_i^\top + \frac{1}{m}(\boldsymbol{\theta} - \boldsymbol{X}\boldsymbol{X}^\top) \succeq \boldsymbol{0}$ for each i. Finally, let us define the matrix $\boldsymbol{W}$ such that $\boldsymbol{W}^{(i,i)} = \boldsymbol{S}^i$ and $\boldsymbol{W}^{(i,j)} = \boldsymbol{X}_i \boldsymbol{X}_j^\top$ for $i \neq j$. Then, $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W})$ is feasible for (12) and attains the same objective value. The relaxation (13) is precisely the relaxation developed in Bertsimas et al. (2023b). $\square$

Proposition 4's proof technique uses the fact that $\boldsymbol{X}$ enters the objective quadratically via $\boldsymbol{X}\boldsymbol{X}^\top$, rather than properties specific to reduced rank regression. This suggests other low-rank problems which are quadratic through $\boldsymbol{X}\boldsymbol{X}^\top$ (or $\boldsymbol{X}^\top\boldsymbol{X}$), e.g., low-rank factor analysis (Bertsimas et al. 2017), sparse plus low-rank matrix decompositions (Bertsimas et al. 2023a) and quadratically constrained programming (Wang and Kılınç-Karzan 2022) admit similarly compact lifted relaxations.

3.3. Basis Pursuit

Given a sample $\{A_{i,j}, (i,j) \in \Omega \subseteq [n] \times [m]\}$ of an *exactly* low-rank matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$, the goal of the low-rank basis pursuit problem is to recover the lowest rank matrix $\boldsymbol{X}$ that exactly matches all observed entries of $\boldsymbol{A}$ (Candès and Recht 2009). This problem admits the formulation:

$$\min_{\boldsymbol{Y}\in\mathcal{Y}_n} \min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}} \quad \mathrm{tr}(\boldsymbol{Y}) \text{ s.t. } \quad \mathcal{P}(\boldsymbol{A}) = \mathcal{P}(\boldsymbol{X}), \boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}, \tag{14}$$

where $\mathcal{P}(\boldsymbol{A})$ denotes a linear map that masks the hidden entries of $\boldsymbol{A}, \boldsymbol{X}$ such that $\mathcal{P}(\boldsymbol{A})_{i,j} = A_{i,j}$ if $(i,j) \in \Omega$ and 0 otherwise. Following Theorem 1 and applying RLT to the constraints $A_{i,j} - X_{i,j} = 0, \forall (i,j) \in \Omega$ leads to the following relaxation

$$\begin{aligned}
\min_{\boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n)} \; \min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}, \boldsymbol{W}\in\mathcal{S}_+^{nm}} \; & \mathrm{tr}(\boldsymbol{Y}) \\
\text{s.t.} \quad & A_{i,j}A_{k,\ell} - A_{k,\ell}X_{i,j} - A_{i,j}X_{k,\ell} + (\boldsymbol{W}^{(i,k)})_{j,\ell} = 0, \forall (i,j),(k,\ell) \in \Omega\times\Omega \\
& A_{i,j} = X_{i,j}, \forall (i,j) \in \Omega \\
& \boldsymbol{W} \succeq \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top, \; \begin{pmatrix} \sum_{i\in[n]} \boldsymbol{W}^{(i,i)} & \boldsymbol{X}^\top \\ \boldsymbol{X} & \boldsymbol{Y} \end{pmatrix} \succeq \boldsymbol{0},
\end{aligned} \tag{15}$$

Similarly to the low-rank matrix completion case, the structure of the compact lifted relaxation means that the off-diagonal blocks of $\boldsymbol{W}$ do not appear in either the objective nor any constraint involving $\boldsymbol{Y}$. They only appear in some of the RLT constraints, but these constraints can always be satisfied by setting $\boldsymbol{W}^{(i,k)} = \boldsymbol{X}_{i,\cdot}\boldsymbol{X}_{j,\cdot}^\top$. As we prove below, the off-diagonal blocks can, therefore, be eliminated from the relaxation without impacting its optimal value:

Proposition 5. *Problem (15) attains the same objective value as*

$$
\begin{aligned}
\min_{\boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n)} \min_{\boldsymbol{X}\in\mathbb{R}^{n\times m},\ \boldsymbol{S}^i\in\mathcal{S}^m_+, i\in[n]} \quad & \mathrm{tr}(\boldsymbol{Y}) \\
\text{s.t.} \quad & A_{i,j}A_{i,\ell} - A_{i,\ell}X_{i,j} - A_{i,j}X_{i,\ell} + (\boldsymbol{S}^i)_{j,\ell} = 0, \forall (i,j),(i,\ell)\in\Omega\times\Omega \\
& A_{i,j} = X_{i,j}, \forall (i,j)\in\Omega \\
& \boldsymbol{S}^i \succeq \boldsymbol{X}_{i,\cdot}\boldsymbol{X}_{i,\cdot}^\top, \ \begin{pmatrix} \sum_{i\in[n]}\boldsymbol{S}^i & \boldsymbol{X}^\top \\ \boldsymbol{X} & \boldsymbol{Y}\end{pmatrix} \succeq \boldsymbol{0},
\end{aligned} \tag{16}
$$

where $\boldsymbol{X}_{i,\cdot}$ denotes a column vector containing the ith row of $\boldsymbol{X}$.

Remark 4. The number of linear equality constraints in (16) grows quadratically in $|\Omega|$, thus, in practice, one may subsample or aggregate these constraints to improve the tractability of the semidefinite relaxation.

Proof of Proposition 5 From a solution to (15), defining $\boldsymbol{S}^i := \boldsymbol{W}^{(i,i)}$ yields a feasible solution to (16) with same objective value. In turn, let us consider a feasible solution to (16), $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{S}^i)$. Define the block matrix $\boldsymbol{W}\in\mathcal{S}^{nm}$ by setting $\boldsymbol{W}^{(i,i)} = \boldsymbol{S}^i$ and $\boldsymbol{W}^{(i,k)} = \boldsymbol{X}_{i,\cdot}\boldsymbol{X}_{k,\cdot}^\top$. Then, it is not hard to see that $\boldsymbol{W} - \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$ is a block diagonal matrix with on-diagonal blocks $\boldsymbol{S}^i - \boldsymbol{X}_{i,\cdot}\boldsymbol{X}_{i,\cdot}^\top \succeq \boldsymbol{0}$. Thus, $\boldsymbol{W} - \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top \succeq \boldsymbol{0}$. Moreover,

$$
(\boldsymbol{W}^{(i,k)})_{j,\ell} = \begin{cases} (\boldsymbol{S}^i)_{j,\ell} & \text{if } i = k, \\ X_{i,j}X_{k,\ell} & \text{otherwise.}\end{cases}
$$

So the linear constraints indexed by $(i,j),(k,\ell)\in\Omega\times\Omega$ are all satisfied. Thus, $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W})$ is feasible in (15) and attains the same objective value. □

The preprocessing techniques proposed here also apply directly to phase retrieval problems (cf. Candès and Li 2014). Indeed, phase retrieval is essentially basis pursuit, except we replace the linear constraint $\mathcal{P}(\boldsymbol{A}-\boldsymbol{X}) = \boldsymbol{0}$ with other constraints $\langle \boldsymbol{g}_i\boldsymbol{g}_i^\top, \boldsymbol{X}\rangle = b_i \ \forall i\in\mathcal{I}$. However, the unstructured nature of the linear constraints implies that eliminating as many variables may not be possible.

We have shown in this section that for three quadratic low-rank problems, it is possible to eliminate enough variables in the lifted relaxation that no matrices of size $n^2\times n^2$ remain. Thus, our proof technique can likely be applied to other low-rank problems of practical interest (e.g., sensor location). This suggests that while lifted relaxations involving $n^2\times n^2$ matrices may appear to be too large to be useful in practice, they can often be reduced to forms that are useful.

4. Numerical Results

In this section, we benchmark our low-rank convex relaxations against the matrix perspective relaxation proposed by Bertsimas et al. (2023b). We emphasize that we introduce $\|\cdot\|_F^2$ regularization to perform this comparison, as the relaxation of Bertsimas et al. (2023b) is not applicable to generic low-rank quadratic optimization. All experiments are conducted on a MacBook Pro laptop (Apple M3, 36 GB), using MOSEK version 10.1, Julia version 1.9, and JuMP.jl version 1.13.0. All solver parameters are set to their default values.

4.1. Low-Rank Matrix Completion

In this section, we evaluate the performance of our new low-rank relaxations on synthetic low-rank matrix completion instances. We use the data generation process of Candès and Recht (2009): We construct a matrix of observations, $\boldsymbol{A}_{\text{full}} \in \mathbb{R}^{n\times m}$, from a rank-$r$ model: $\boldsymbol{A}_{\text{full}} = \boldsymbol{U}\boldsymbol{V} + \epsilon \boldsymbol{Z}$, where the entries of $\boldsymbol{U} \in \mathbb{R}^{n\times r}, \boldsymbol{V} \in \mathbb{R}^{r\times m}$, and $\boldsymbol{Z} \in \mathbb{R}^{n\times m}$ are drawn independently from a standard normal distribution, and $\epsilon \geq 0$ models the degree of noise. We fix $\epsilon = 0.1, m = n$ and $r = 2$ for all experiments. We sample a random subset $\Omega \subseteq [n] \times [m]$, of predefined size (see also Candès and Recht 2009, section 1.1.2). Each result reported in this section is averaged over 10 random seeds.

We first evaluate the quality of our new relaxations, compared with the matrix perspective relaxation of Bertsimas et al. (2023b, MPRT). Unfortunately, MPRT does not apply to (7) as it requires a Frobenius regularization term in the objective. Hence, instead of (7), we consider

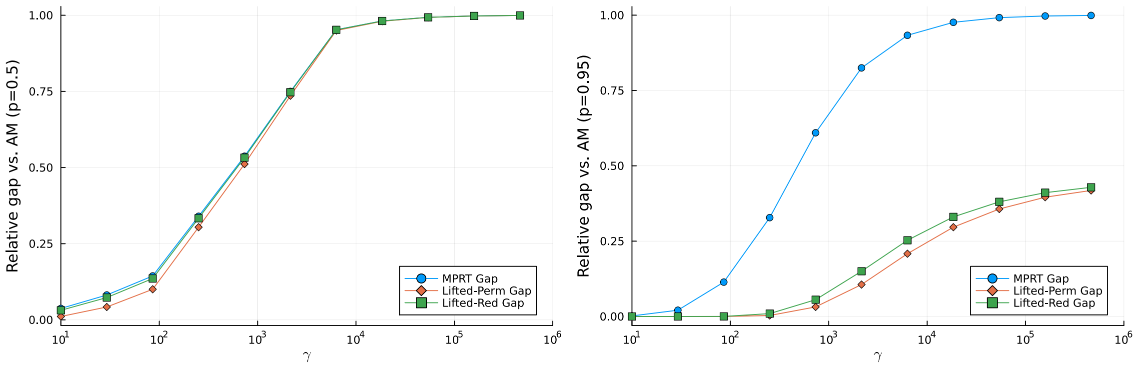
$$\min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}} \quad \frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2 + \frac{1}{2}\sum_{(i,j)\in\Omega}(A_{i,j} - X_{i,j})^2 \text{ s.t. } \operatorname{rank}(\boldsymbol{X}) \leq r.$$

for some regularization parameter $\gamma > 0$. As $\gamma \to \infty$, we recover the solution of (7). We compare the (lower) bounds obtained by three different approaches: MPRT, our full lifted relaxation (3) with the permutation equalities (6), hereafter denoted "Lifted-Perm", and our compact lifted relaxation (9) ("Lifted-Red"). Figure 1 reports the lower bounds achieved by each approach —in relative terms compared with an upper bound achieved by the alternating minimization method of Burer and Monteiro (2003), Jain et al. (2013) initialized with a truncated SVD of $\mathcal{P}(\boldsymbol{A})$ (absolute values are reported in Figures EC.1–EC.2)— as γ increases, for different proportion of entries sampled $p = |\Omega|/mn$ ($n = 8$ being fixed). Specifically, we compute the relative gap as $\text{Gap} = \frac{UB-LB}{UB}$ with $UB > 0$.

Supporting Proposition 2, we observe that Lifted-Perm and Lifted-Red obtain smaller optimality gaps than MPRT, for all values of γ, and that the benefit increases as the fraction of sampled entries p increases. In particular, when $p = 0.95$, there is a regime of values of γ (around 10^2) where both lifted relaxations are tight (as evidenced by a gap of 0%), while MPRT is not.

In addition, as γ increases, MPRT achieves an uninformative gap of 100% (by returning a trivial lower bound of 0, see Figure EC.1), while our lifted relaxations provide non-trivial bounds (and gaps). From this experiment, it seems that imposing the permutation equalities (6) on $\boldsymbol{W}_{y,y}$ in our lifted relaxation (Lifted-Perm vs. Lifted-Red) does not lead to significantly tighter bounds, while being computationally much more expensive (see Figure EC.3 for computational times).

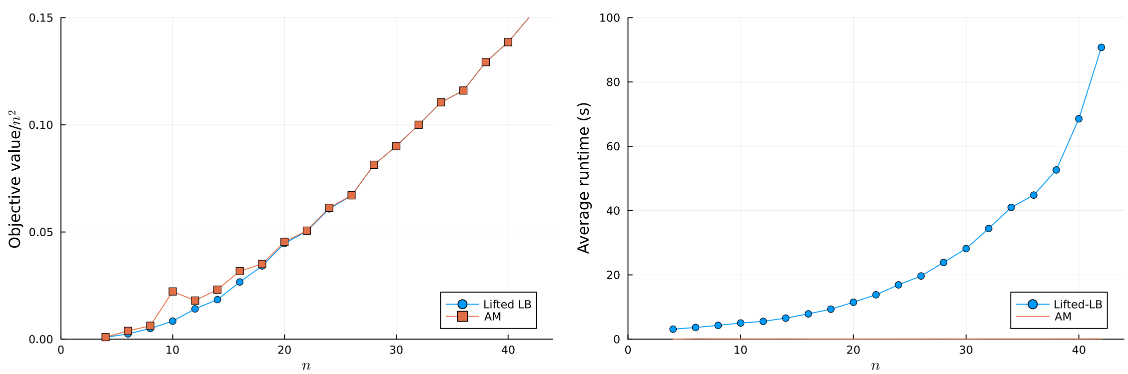
Our second experiment benchmarks the scalability of our reduced lifted relaxation as we vary $n = m$ with the proportion of entries fixed at $p = 0.5$. We set $\gamma = 10^4/n^2$. We report the average lower bound (divided by n^2 so that quantities have the same meaning as we vary n; left) and the average computational time (right) in Figure 2. We also report the average objective value obtained



(a) $p = 0.5$ (b) $p = 0.95$

Figure 1 Relative gap obtained with different relaxations of the regularized matrix completion problem as we vary γ. We fix $n = 8$. Results are averaged over 10 replications.

by alternating minimization as a baseline. Note that we do not consider the full lifted relaxation in this experiment, as it requires more RAM than is available for these experiments when $n = 10$. For any $n \in \{4, \dots, 42\}$, the lifted relaxation can be solved in seconds, while when $n > 44$, Mosek runs out of RAM. Moreover, the lower bound from the lifted relaxation is tight for $n \geq 18$, as identified by the fact that alternating minimization matches the bound.



(a) Objective values (b) Runtimes

Figure 2 Objective value (left panel) and runtime for Lifted-Red (right panel) as we vary $n = m$ with $p = 0.5$ for our reduced lifted relaxation. Results are averaged over 10 replications.

5. Conclusion

This paper develops scalable semidefinite relaxations for rank-constrained quadratic optimization problems, which are available in the literature only in the special case where the objective or constraints admit spectral substructures. Our starting point is a lifting for rank-constrained formulations with linear constraints, obtained by introducing moment variables for the quadratic terms in $\boldsymbol{X}$ and the projection matrix $\boldsymbol{Y}$. While a direct lifted relaxation involves large moment matrices with $n^2 \times n^2$ semidefinite blocks, we show that most of these blocks are redundant and can be removed without altering the relaxation's optimal value.

Beyond this generic reduction, we derive a practical modeling workflow for exploiting additional structure. For matrix completion and basis pursuit, the relaxation can be expressed using only row-wise $m \times m$ semidefinite variables and a single $(m+n) \times (m+n)$ coupling constraint, eliminating any need to manipulate any matrix whose dimension is $O((n+m)^2)$. Collectively, our examples suggest that large lifted relaxations can often be transformed into implementable SDPs.

Endnotes

1. Specifically, the constraints $\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}, \boldsymbol{Y}^2 = \boldsymbol{Y}$ imply that $\mathrm{Rank}(\boldsymbol{X}) \leq \mathrm{tr}(\boldsymbol{Y})$, which can be made to hold with equality by letting $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{U}^\top$ for $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ a singular value decomposition of $\boldsymbol{X}$.
2. One may also be tempted to impose the inequalities $\boldsymbol{A}\boldsymbol{W}_{x,x}\boldsymbol{A}^\top + \boldsymbol{b}\boldsymbol{b}^\top \succeq \boldsymbol{b}\boldsymbol{x}^\top\boldsymbol{A}^\top + \boldsymbol{A}\boldsymbol{x}\boldsymbol{b}^\top$ but they are actually implied by $\boldsymbol{W}_{x,x} \succeq \boldsymbol{x}\boldsymbol{x}^\top$.

References

Adams WP, Sherali HD (1986) A tight linearization and an algorithm for zero-one quadratic programming problems. *Management Science* 32(10):1274–1290.

Bao X, Sahinidis NV, Tawarmalani M (2011) Semidefinite relaxations for quadratically constrained quadratic programming: A review and comparisons. *Mathematical Programming* 129:129–157.

Bertsimas D, Copenhaver MS, Mazumder R (2017) Certifiably optimal low rank factor analysis. *Journal of Machine Learning Research* 18(29):1–53.

Bertsimas D, Cory-Wright R, Johnson NA (2023a) Sparse plus low rank matrix decomposition: A discrete optimization approach. *Journal of Machine Learning Research* 24(267):1–51.

Bertsimas D, Cory-Wright R, Pauphilet J (2022) Mixed-projection conic optimization: A new paradigm for modeling rank constraints. *Operations Research* 70(6):3321–3344.

Bertsimas D, Cory-Wright R, Pauphilet J (2023b) A new perspective on low-rank optimization. *Mathematical Programming* 202:47–92.

Boyd S, El Ghaoui L, Feron E, Balakrishnan V (1994) *Linear Matrix Inequalities in System and Control Theory* (SIAM).

Burer S, Monteiro RD (2003) A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming* 95(2):329–357.

Candès EJ, Li X (2014) Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. *Foundations of Computational Mathematics* 14:1017–1026.

Candès EJ, Recht B (2009) Exact matrix completion via convex optimization. *Foundations of Computational Mathematics* 9(6):717–772.

Hiriart-Urruty JB, Le HY (2012) Convexifying the set of matrices of bounded rank: applications to the quasiconvexification and convexification of the rank function. *Optimization Letters* 6(5):841–849.

Jain P, Netrapalli P, Sanghavi S (2013) Low-rank matrix completion using alternating minimization. *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, 665–674.

Kim J, Tawarmalani M, Richard JPP (2022) Convexification of permutation-invariant sets and an application to sparse principal component analysis. *Mathematics of Operations Research* 47(4):2547–2584.

Li Y, Xie W (2025) On the partial convexification for low-rank spectral optimization: rank bounds and algorithms. *Mathematical Programming* 1–58.

Magnus JR, Neudecker H (1979) The commutation matrix: some properties and applications. *The Annals of Statistics* 7(2):381–394.

Negahban S, Wainwright MJ (2011) Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Annals of Statistics* 39:1069–1097.

Overton ML, Womersley RS (1992) On the sum of the largest eigenvalues of a symmetric matrix. *SIAM Journal on Matrix Analysis and Applications* 13(1):41–45.

Sherali HD, Alameddine A (1992) A new reformulation-linearization technique for bilinear programming problems. *Journal of Global optimization* 2(4):379–410.

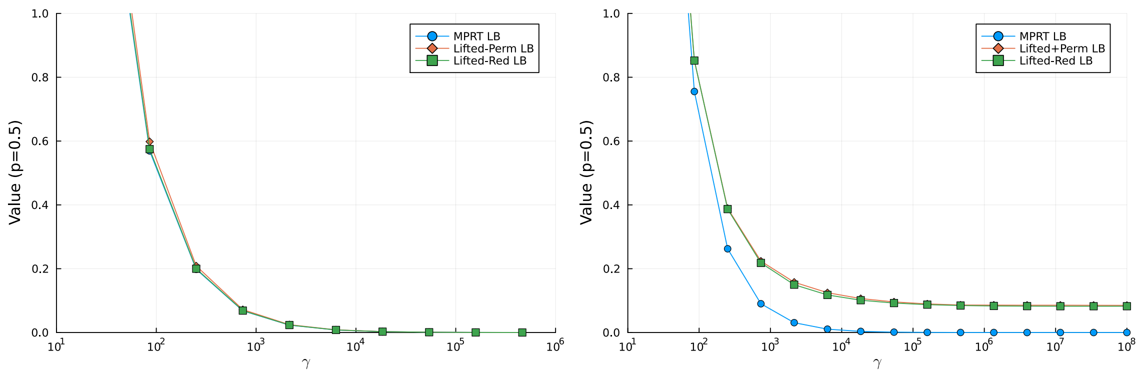
Shor NZ (1987) Quadratic optimization problems. *Soviet Journal of Computer and Systems Sciences* 25:1–11.

Wang AL, Kılınç-Karzan F (2022) On the tightness of SDP relaxations of QCQPs. *Mathematical Programming* 193(1):33–73.

Supplementary Material

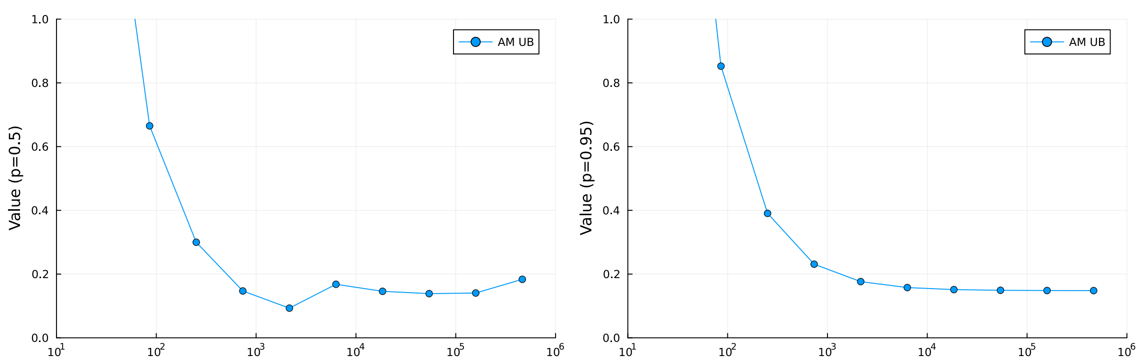
EC.1. Additional Results for Low-Rank Matrix Completion

Figure 1 compares the quality of different relaxations for low-rank matrix completion by returning the optimality gap achieved, defined as the relative difference between the lower bound (obtained by each relaxation) and our upper bound (obtained by alternating minimization, AM). Figure EC.1 and EC.2 report the lower and upper bounds separately. Moreover, Figure EC.3 compares the same three relaxations in terms of computational time.



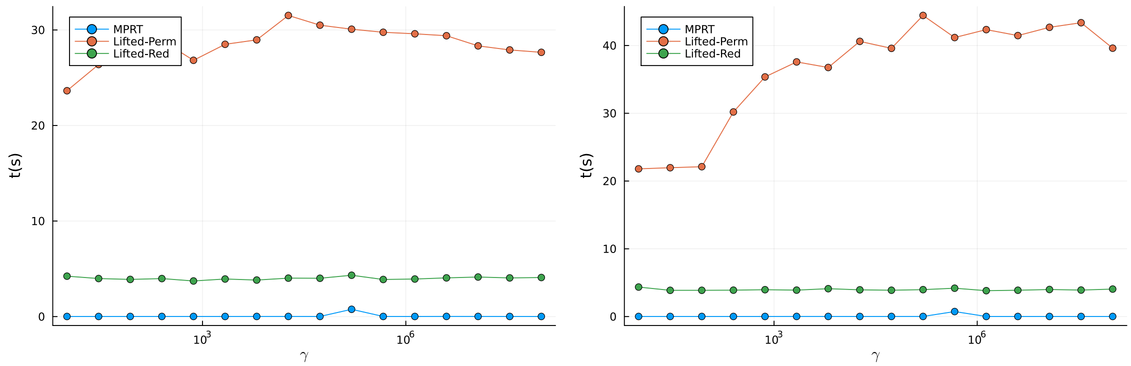
(a) $p = 0.5$ (b) $p = 0.95$

Figure EC.1 Absolute lower bounds as we vary γ for (a) a matrix perspective relaxation ("MPRT"), (b) our lifted relaxation with permutation equalities ("Lifted-Perm"), (c) our compact lifted relaxation with no permutation equalities ("Lifted-Red"), for $p \in \{0.5, 0.95\}$ and $n = 8$.



(a) $p = 0.5$ (b) $p = 0.95$

Figure EC.2 Absolute upper bounds as we vary γ for alternating minimization initialized at a rank-r SVD of $\mathcal{P}(\boldsymbol{A})$ for $p \in \{0.5, 0.95\}$ and $n = 8$.



(a) $p = 0.5$ (b) $p = 0.95$

Figure EC.3 Runtimes for (a) a matrix perspective relaxation ("MPRT"), (b) our lifted relaxation with permutation equalities ("Lifted-Perm"), (c) our lifted relaxation with no permutation equalities ("Lifted-Red"), for $p \in \{0.5, 0.95\}$, $n = 8$, and increasing γ.