

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 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 the moment matrix are redundant and derive an equivalent compact relaxation that only involves two semidefinite constraints of dimension $nm + 1$ and $n + m$, respectively. We also derive a new class of valid inequalities for low-rank problems, which we call projection cuts, that exploit the fact that rank constraints are inherited by linear images of a low-rank matrix, to strengthen our low-rank relaxations substantially. For matrix completion and reduced-rank regression problems, among others, we exploit additional structure to obtain even more compact formulations involving semidefinite matrices of dimension at most the sum of the two dimensions of the low-rank decision matrix (i.e., of size at most $n + m$). 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

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

where $\mathbf{H}, \mathbf{Q}_i \in \mathcal{S}^{nm}$ are $nm \times nm$ symmetric matrices, \mathbf{D}, \mathbf{E}_i are $n \times m$ matrices, $\lambda \in \mathbb{R}_+, k \in \mathbb{N}$ are parameters which control the complexity of $\mathbf{X} \in \mathbb{R}^{n \times m}$ by respectively penalizing and constraining its rank, \mathcal{I} denotes the index set of constraints, and $\text{vec}(\cdot)$ denotes the vectorization of the matrix \mathbf{X} . Note that we write $\text{vec}(\mathbf{X}^\top)$ rather than $\text{vec}(\mathbf{X})$ to simplify the notation in our subsequent 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 $\{\mathbf{X} : \text{rank}(\mathbf{X}) \leq k, f(\mathbf{X}) \leq 0\}$, where f is a quasiconvex function in the singular values of \mathbf{X} , (ii) Bertsimas et al. (2023b), who, building upon the work of Bertsimas et al. (2022), leverage partial separability to obtain compact semidefinite relaxations for low-rank problems with a spectral term in the objective of the form $\text{tr}(f(\mathbf{X}))$ 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 $\text{vec}(\mathbf{X}^\top)$. 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 representation of low-rank matrices with projection matrices as proposed in Bertsimas et al. (2022), thus enriching the toolbox of semidefinite relaxations for low-rank optimization. Our main theoretical contributions are a new class of compact semidefinite relaxations for generic low-rank optimization problems (Theorem 1 in Section 2.1), and a new family of valid inequalities, which we term projection cuts (Theorem 2 in Section 2.2), that often close most of the gap between this convex relaxation and feasible solutions. To the best of our knowledge, this is the first work that obtains non-trivial and computationally tractable lower bounds for generic 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. To illustrate our approach, we apply our lifted relaxation to low-rank matrix completion and reduced rank regression in Section 3. 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 \mathbf{x} denote vectors, uppercase boldface characters such as \mathbf{X} denote matrices, and calligraphic uppercase characters such as \mathcal{Z} denote sets. We let $[n]$ denote the set of running indices $\{1, \dots, n\}$. The cone of $n \times n$ symmetric (resp. positive semidefinite) matrices is denoted by \mathcal{S}^n (resp. \mathcal{S}_+^n).

For a matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$, we let \mathbf{X}_i denote its i th column and $\mathbf{X}_{i,\cdot}$ denote a vector containing its i th row. We let $\text{vec}(\mathbf{X}) : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{nm}$ denote the vectorization operator which maps matrices to vectors by stacking columns. For a matrix \mathbf{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 $\mathbf{W}^{(i, i')}$. The dimension of each block will be clear from the context, given the size of the matrix \mathbf{W} and the number of blocks. For any matrices \mathbf{A}, \mathbf{B} , we use $\mathbf{A} \otimes \mathbf{B}$ to denote their Kronecker product (see Gupta and Nagar 2018, Chapter 1.2).

We let \mathbf{X}^\dagger be the pseudoinverse of \mathbf{X} , which is used in the Schur complement lemma (Boyd et al. 1994, Eqn. 2.41). We let $\mathcal{Y}_n^k := \{\mathbf{Y} \in \mathcal{S}_+^n : \mathbf{Y}^2 = \mathbf{Y}, \text{tr}(\mathbf{Y}) \leq k\}$ denote the set of orthogonal projection matrices with rank at most k , whose convex hull is $\{\mathbf{P} \in \mathcal{S}_+^n : \mathbf{P} \preceq \mathbf{I}_n, \text{tr}(\mathbf{P}) \leq k\}$ (Overton and Womersley 1992, Theorem 3).

2. Lifted Relaxations for Low-Rank Optimization Problems

In this section, we demonstrate how to apply a lifted relaxation technique to low-rank quadratic optimization problems in a manner that yields tractable convex relaxations. Throughout the section, we study the following reformulation of Problem (1):

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

As demonstrated¹ in Bertsimas et al. (2022), any rank-constrained optimization problem of the form (1) can be formulated as an optimization over (\mathbf{X}, \mathbf{Y}) of the form (2), where the additional decision matrix \mathbf{Y} is a projection matrix which encodes the span of \mathbf{X} and whose trace bounds $\text{rank}(\mathbf{X})$. However, unlike Problem (1), Problem (2) concentrates all rank-induced nonconvexity within the constraints $\mathbf{Y}^2 = \mathbf{Y}$ and $\mathbf{X} = \mathbf{Y} \mathbf{X}$, making it amenable to lifted relaxation techniques.

The rest of the section is organized as follows: First, in Section 2.1, we design a lifted semidefinite relaxation for Problem (2) (Proposition 1) and show that many of the additional semidefinite variables can be safely omitted (Theorem 1). Second, in Section 2.2, we develop a family of additional valid inequalities which we call projection cuts. Third, in Section 2.3, we specialize our results to the case where \mathbf{H} and \mathbf{Q}_i are block diagonal matrices, and show that our semidefinite relaxations and projection cuts can be written in terms of optimization problems that do not involve any $n^2 \times n^2$ matrices in this case. Finally, in Section 2.4, we revisit the nonlinear formulation of Burer and Monteiro (2003), and discuss why a naive lifted relaxation designed around this nonlinear formulation cannot yield strong lower bounds, while a more sophisticated version of the relaxation is equivalent to the relaxation designed in Section 2.1.

2.1. A New Lifted Relaxation and Its Compact Formulation

We introduce matrices $\mathbf{W}_{x,x}$, $\mathbf{W}_{x,y}$, $\mathbf{W}_{y,y}$ to model the outer products $\text{vec}(\mathbf{X}^\top)\text{vec}(\mathbf{X}^\top)^\top$, $\text{vec}(\mathbf{X}^\top)\text{vec}(\mathbf{Y})^\top$, and $\text{vec}(\mathbf{Y})\text{vec}(\mathbf{Y})^\top$. By including these decision matrices, we have:

PROPOSITION 1. *The convex semidefinite optimization problem*

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

is a valid convex relaxation of Problem (2).

REMARK 1. If an optimal solution to (3) is such that $\mathbf{W}_{x,x} = \text{vec}(\mathbf{X}^\top)\text{vec}(\mathbf{X}^\top)^\top$ then the optimal values of (3) and (2) coincide.

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

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

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

Moreover, we have

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

□

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

THEOREM 1. *Problem (3) is equivalent to*

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

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

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

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

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

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

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

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

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

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

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

Since $\mathbf{Y} = \mathbf{U}^\top \mathbf{X}^\top$, we have $\text{vec}(\mathbf{Y}) = \text{vec}(\mathbf{U}^\top \mathbf{X}^\top) = (\mathbf{I}_n \otimes \mathbf{U}^\top) \text{vec}(\mathbf{X}^\top)$ and thus our construction is consistent with the existing value of \mathbf{Y} . We now verify that $(\mathbf{X}, \mathbf{Y}, \mathbf{W}_{x,x}, \mathbf{W}_{x,y}, \mathbf{W}_{y,y})$ is feasible for (3). By construction, $\mathbf{M} \succeq \mathbf{0}$. Thus, $(\mathbf{X}, \mathbf{Y}, \mathbf{W}_{x,x}, \mathbf{W}_{x,y}, \mathbf{W}_{y,y})$ satisfies the semidefinite constraint in (3). Next, by construction, $\mathbf{W}_{x,y}$ and $\mathbf{W}_{y,y}$ can be decomposed into $n \times n$ blocks satisfying:

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

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

$$\begin{aligned} \sum_{i \in [n]} \mathbf{W}_{x,y}^{(i,i)} &= \sum_{i \in [n]} \mathbf{W}_{x,x}^{(i,i)} \mathbf{U} = \left(\sum_{i \in [n]} \mathbf{W}_{x,x}^{(i,i)} \right) \left(\sum_{j \in [n]} \mathbf{W}_{x,x}^{(j,j)} \right)^\dagger \mathbf{X}^\top = \mathbf{X}^\top, \\ \sum_{i \in [n]} \mathbf{W}_{y,y}^{(i,i)} &= \sum_{i \in [n]} \mathbf{U}^\top \mathbf{W}_{x,x}^{(i,i)} \mathbf{U} = \mathbf{U}^\top \left(\sum_{i \in [n]} \mathbf{W}_{x,x}^{(i,i)} \right) \mathbf{U} = \mathbf{U}^\top \mathbf{X}^\top = \mathbf{Y}. \end{aligned}$$

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

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

Finally, it is interesting to consider whether the relaxation developed here is at least as strong as the matrix perspective relaxation of Bertsimas et al. (2023b). We now prove this is indeed the case. The relaxation of 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 \mathbf{H} is positive definite, so that the term $\langle \mathbf{H}, \text{vec}(\mathbf{X}^\top) \text{vec}(\mathbf{X}^\top)^\top \rangle + \langle \mathbf{D}, \mathbf{X} \rangle$ in Problem (2) can be rewritten as the partially separable term $\frac{1}{2\gamma} \|\mathbf{X}\|_F^2 + h(\mathbf{X})$, where h is convex in \mathbf{X} . Further, assume the matrices \mathbf{Q}_i are positive definite, so that the terms $\langle \mathbf{Q}_i, \text{vec}(\mathbf{X}^\top) \text{vec}(\mathbf{X}^\top)^\top \rangle + \langle \mathbf{E}_i, \mathbf{X} \rangle$ in Problem (2) can be rewritten as the partially separable terms $\frac{1}{2\gamma} \|\mathbf{X}\|_F^2 + h_i(\mathbf{X})$ for a common γ , where h_i are convex in \mathbf{X} . Then, the optimal value of Problem (3) is at least as large as the relaxation of Bertsimas et al. (2023b)*

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

Proof of Proposition 2 Given the equivalence between Problems (3)–(4) proven in Theorem 1 and the fact that the quadratic forms are all positive semidefinite, it suffices to show that the constraints in (4) imply the constraints in (5). Letting $\boldsymbol{\theta} := \sum_{i \in [n]} \mathbf{W}_{x,x}^{(i,i)}$, we observe that $\boldsymbol{\theta}$ is feasible for (5). Similarly, convexity of h, h_i implies $\mathbf{H} - \frac{1}{2\gamma} \mathbf{I} \succeq \mathbf{0}$, $\mathbf{Q}_i - \frac{1}{2\gamma} \mathbf{I} \succeq \mathbf{0}$. Therefore, the constraint $\mathbf{W}_{x,x} \succeq \text{vec}(\mathbf{X}^\top) \text{vec}(\mathbf{X}^\top)^\top$ yields the inequalities

$$\begin{aligned} \frac{1}{2\gamma} \text{tr}(\boldsymbol{\theta}) + h(\mathbf{X}) &\leq \langle \mathbf{H}, \mathbf{W}_{x,x} \rangle + \langle \mathbf{D}, \mathbf{X} \rangle, \\ \frac{1}{2\gamma} \text{tr}(\boldsymbol{\theta}) + h_i(\mathbf{X}) &\leq \langle \mathbf{Q}_i, \mathbf{W}_{x,x} \rangle + \langle \mathbf{E}_i, \mathbf{X} \rangle \leq b_i, \quad \forall i \in \mathcal{I}, \end{aligned}$$

which completes the proof. \square

The proof of Proposition 2 reveals that in the boundary case of reduced rank regression (see §3.2), our lifted relaxation (4) reduces to a perspective-type relaxation which can be perceived as decomposing the variable $\boldsymbol{\theta}$ in (5), and strengthening the relaxation by imposing additional constraints on elements of this decomposition.

2.2. Projection Cuts for Strengthening the Lifted Relaxation

In this section, we propose a strategy for improving our compact lifted relaxations, sometimes significantly, by exploiting the fact that the rank constraint $\text{rank}(\mathbf{X}) \leq k$ is inherited by many linear images of \mathbf{X} . In particular, if \mathbf{A}, \mathbf{B} are matrices of arbitrary dimension, then the rank constraint implies $\text{rank}(\mathbf{AXB}) \leq k$. Formally, we have the following result:

THEOREM 2. *Let $\mathbf{A} \in \mathbb{R}^{p \times n}, \mathbf{B} \in \mathbb{R}^{m \times q}$ be real matrices, and let us denote $\mathbf{D} := \mathbf{A} \otimes \mathbf{B}^\top \in \mathbb{R}^{pq \times nm}$. In particular, for any $\mathbf{X} \in \mathbb{R}^{n \times m}$, we have $\text{vec}((\mathbf{AXB})^\top) = \mathbf{D} \text{vec}(\mathbf{X}^\top)$. Further, for the lifted matrix $\mathbf{W}_{x,x} \in \mathcal{S}^{nm}$, let the auxiliary matrix $\mathbf{D}\mathbf{W}_{x,x}\mathbf{D}^\top \in \mathcal{S}^{pq}$ be partitioned into $p \times p$ blocks of size $q \times q$. Then, the following constraint is valid for Problem (2), and may be imposed in (4):*

$$\exists \mathbf{Z}_D \in \text{Conv}(\mathcal{Y}_p^k) \quad \text{s.t.} \quad \begin{pmatrix} \sum_{a=1}^p (\mathbf{D}\mathbf{W}_{x,x}\mathbf{D}^\top)^{(a,a)} & (\mathbf{AXB})^\top \\ \mathbf{AXB} & \mathbf{Z}_D \end{pmatrix} \succeq \mathbf{0}, \mathbf{A}\mathbf{Y}\mathbf{A}^\top \preceq \rho_A \mathbf{Z}_D, \quad (6)$$

where $\rho_A := \lambda_{\max}(\mathbf{A}\mathbf{A}^\top)$, and \mathbf{Y} is the semidefinite matrix in (4).

REMARK 2. A natural choice for \mathbf{B} in Theorem 2 is taking $q = m$ and $\mathbf{B} = \mathbf{I}_m$. Indeed, any variable \mathbf{Z}_D satisfying (6) for \mathbf{I}_m also satisfies (6) for any $\mathbf{B} \in \mathbb{R}^{m \times q}$ (simply left/right multiply the first semidefinite inequality by $\text{Diag}(\mathbf{B}^\top, \mathbf{I}_p)/\text{Diag}(\mathbf{B}, \mathbf{I}_p)$). The benefit of using a generic \mathbf{B} (especially one with $q < m$) is computational tractability.

Proof of Theorem 2 Let (\mathbf{X}, \mathbf{Y}) be feasible for Problem (2), and take the exact lifted matrix $\mathbf{W}_{x,x} = \text{vec}(\mathbf{X}^\top) \text{vec}(\mathbf{X}^\top)^\top$. Further, for ease of notation, define $\mathbf{W}_D := \mathbf{D}\mathbf{W}_{x,x}\mathbf{D}^\top \in \mathcal{S}^{pq}$, and $\Theta_D(\mathbf{W}_{x,x}) := \sum_{a=1}^p \mathbf{W}_D^{(a,a)} \in \mathcal{S}^q$. By definition of \mathbf{D} , we have $\mathbf{W}_D = \text{vec}((\mathbf{AXB})^\top) \text{vec}((\mathbf{AXB})^\top)^\top$ and $\Theta_D(\mathbf{W}_{x,x}) = \sum_{a=1}^p ((\mathbf{AXB})^\top)_{a,\cdot}^\top ((\mathbf{AXB})^\top)_{a,\cdot} = (\mathbf{AXB})^\top \mathbf{AXB}$.

Let \mathbf{Z}_D be the orthogonal projector onto the column space of $\mathbf{A}\mathbf{Y}$. Since $\text{rank}(\mathbf{A}\mathbf{Y}) \leq \text{rank}(\mathbf{Y}) \leq k$, the matrix \mathbf{Z}_D has rank at most k , so $\mathbf{Z}_D \in \mathcal{Y}_p^k$. In addition, from $\mathbf{Z}_D \mathbf{A}\mathbf{Y} = \mathbf{A}\mathbf{Y}$ and $\mathbf{X} = \mathbf{Y}\mathbf{X}$, we have the relationship $\mathbf{Z}_D \mathbf{AXB} = \mathbf{AXB}$. Consequently,

$$\begin{pmatrix} \Theta_D(\mathbf{W}_{x,x}) & (\mathbf{AXB})^\top \\ (\mathbf{AXB}) & \mathbf{Z}_D \end{pmatrix} = \begin{pmatrix} (\mathbf{AXB})^\top \mathbf{AXB} & (\mathbf{AXB})^\top \\ \mathbf{AXB} & \mathbf{Z}_D \end{pmatrix} = \begin{pmatrix} (\mathbf{AXB})^\top \\ \mathbf{Z}_D \end{pmatrix} \begin{pmatrix} (\mathbf{AXB})^\top \\ \mathbf{Z}_D \end{pmatrix}^\top \succeq \mathbf{0}.$$

Thus (6) is satisfied by every feasible point of Problem (2) under its exact lift. For the remaining inequality, observe that $\mathbf{A}\mathbf{Y}\mathbf{A}^\top = \mathbf{A}\mathbf{Y}(\mathbf{A}\mathbf{Y})^\top$, because \mathbf{Y} is an orthogonal projection matrix. By construction, $\text{span}(\mathbf{Z}_D) = \text{span}(\mathbf{A}\mathbf{Y}) = \text{span}(\mathbf{A}\mathbf{Y}\mathbf{A}^\top)$. Since $\mathbf{0} \preceq \mathbf{A}\mathbf{Y}\mathbf{A}^\top \preceq \mathbf{A}\mathbf{A}^\top \preceq \rho_A \mathbf{I}_p$, it follows that $\mathbf{A}\mathbf{Y}\mathbf{A}^\top \preceq \rho_A \mathbf{Z}_D$. Finally, replacing $\mathbf{Z}_D \in \mathcal{Y}_p^k$ by $\mathbf{Z}_D \in \text{Conv}(\mathcal{Y}_p^k)$ preserves validity. \square

We support our discussion via Example EC.1 in Section EC.1.1, which shows that for a specific low-rank matrix completion problem, (3) gives a relaxation with an optimality gap of around 50%, but imposing a small number of projection cuts of the above form closes the gap to 0.

2.3. More Compact Relaxations for Block Diagonal Optimization Problems

In this section, we show that if the quadratic objective matrix \mathbf{H} and the constraint matrices \mathbf{Q}_i are block diagonal, then we can omit the off-diagonal blocks of $\mathbf{W}_{x,x}$. We demonstrate the utility of

this observation throughout Section 3, by leveraging it repeatedly to eliminate most of the decision variables from Problem (4) without altering the objective value, for several low-rank problems:

THEOREM 3. *In Problem (4), let the quadratic objective matrix \mathbf{H} be block diagonal, i.e.:*

$$\mathbf{H} = \text{Diag}(\mathbf{H}_1, \dots, \mathbf{H}_n) : \mathbf{H}_i \in \mathcal{S}^m,$$

and let the constraint matrices \mathbf{Q}_i similarly be block decomposable into n matrices $\mathbf{Q}_{i,j}$, i.e.,

$$\mathbf{Q}_i = \text{Diag}(\mathbf{Q}_{i,1}, \dots, \mathbf{Q}_{i,n}) : \mathbf{Q}_{i,j} \in \mathcal{S}^m.$$

Then, letting $\mathbf{X}_{i,\cdot}$ denote a column vector containing the i th row of \mathbf{X} , Problem (4) attains the same optimal value as the following optimization problem:

$$\begin{aligned} \min_{\mathbf{Y} \in \text{Conv}(\mathcal{Y}_n^k)} \min_{\mathbf{X} \in \mathbb{R}^{n \times m}, \mathbf{S}^1, \dots, \mathbf{S}^n \in \mathcal{S}_+^m} \quad & \langle \mathbf{D}, \mathbf{X} \rangle + \sum_{i=1}^n \langle \mathbf{H}_i, \mathbf{S}^i \rangle + \lambda \cdot \text{tr}(\mathbf{Y}) \\ \text{s.t.} \quad & \mathbf{S}^i \succeq \mathbf{X}_{i,\cdot} \mathbf{X}_{i,\cdot}^\top, \quad i \in [n], \quad \begin{pmatrix} \sum_{i=1}^n \mathbf{S}^i & \mathbf{X}^\top \\ \mathbf{X} & \mathbf{Y} \end{pmatrix} \succeq \mathbf{0}, \\ & \sum_{j=1}^n \langle \mathbf{Q}_{i,j}, \mathbf{S}^j \rangle + \langle \mathbf{E}_i, \mathbf{X} \rangle \leq b_i \quad \forall i \in \mathcal{I}. \end{aligned} \quad (7)$$

Proof of Theorem 3 It suffices to show that given any feasible solution to (7), we can construct a feasible solution to (4) with the same objective value; the converse is immediate. Let $(\mathbf{X}, \mathbf{Y}, \mathbf{S}^i)$ be feasible in (7). Define the block matrix $\mathbf{W} \in \mathcal{S}_+^{nm \times nm}$ by setting $\mathbf{W}^{(i,i)} = \mathbf{S}^i$ and $\mathbf{W}^{(i,j)} = \mathbf{X}_{i,\cdot} \mathbf{X}_{j,\cdot}^\top$. Then, it is not hard to see that $\mathbf{W} - \text{vec}(\mathbf{X}^\top) \text{vec}(\mathbf{X}^\top)^\top$ is a block matrix with zero off-diagonal blocks and on-diagonal blocks $\mathbf{S}^i - \mathbf{X}_{i,\cdot} \mathbf{X}_{i,\cdot}^\top \succeq \mathbf{0}$. Thus, $\mathbf{W} - \text{vec}(\mathbf{X}^\top) \text{vec}(\mathbf{X}^\top)^\top$ is a positive semidefinite matrix, and $\mathbf{W} \succeq \text{vec}(\mathbf{X}^\top) \text{vec}(\mathbf{X}^\top)^\top$. Moreover, due to the block diagonal nature of the lifted semidefinite matrices \mathbf{H} and \mathbf{Q}_i , $(\mathbf{X}, \mathbf{Y}, \mathbf{W})$ is feasible in (4) and attains the same objective value. \square

Compared with (3) or (4), the semidefinite relaxation (7) involves n positive semidefinite variables of dimension m (vs. one PSD matrix of dimension nm in (3)) and $n + 1$ semidefinite constraints of dimension $m + 1$ (vs. one of dimension $nm + 1$)—both relaxations involve one semidefinite constraint of dimension $n + m$.

Finally, we now present a restriction of Theorem 2 that provides valid inequalities involving only the matrices $\mathbf{S}^i = \mathbf{X}_{i,\cdot} \mathbf{X}_{i,\cdot}^\top$, which can be used for the more compact relaxation (7).

COROLLARY 1. *Consider a matrix $\mathbf{A} \in \mathbb{R}^{p \times n}$ where each row and column has at most one non-zero coordinate. Namely, row j of \mathbf{A} is of the form $\alpha_j \mathbf{e}_{i(j)}^\top$ for some $\alpha_j \in \mathbb{R}$ and $i : j \mapsto i(j) \in [n]$ is injective. Then, denoting $\mathbf{S}^i = \mathbf{X}_{i,\cdot} \mathbf{X}_{i,\cdot}^\top, i \in [n]$, the following constraint is valid for (2), and may be imposed in (7) :*

$$\exists \mathbf{Z}_A \in \text{Conv}(\mathcal{Y}_p^k) \quad \text{s.t.} \quad \begin{pmatrix} \sum_{j \in [p]} \alpha_j^2 \mathbf{S}^{i(j)} & (\mathbf{A}\mathbf{X})^\top \\ \mathbf{A}\mathbf{X} & \mathbf{Z}_A \end{pmatrix} \succeq \mathbf{0}, \quad \mathbf{A}\mathbf{Y}\mathbf{A}^\top \preceq \left(\max_{j \in [p]} \alpha_j^2 \right) \mathbf{Z}_A. \quad (8)$$

Proof of Corollary 1 The result follows from applying Theorem 2 to $(\mathbf{A}, \mathbf{I}_m)$. Indeed, in this case, $\mathbf{A}^\top \mathbf{A} = \sum_{j \in [p]} \alpha_j^2 \mathbf{e}_{i(j)} \mathbf{e}_{i(j)}^\top$ and the $i(j)$'s are distinct, so

$$\rho_A = \lambda_{\max}(\mathbf{A}\mathbf{A}^\top) = \lambda_{\max}(\mathbf{A}^\top \mathbf{A}) = \max_{j \in [p]} \alpha_j^2,$$

$$\text{and } \sum_{a \in [p]} (\mathbf{D}\mathbf{W}_{x,x} \mathbf{D}^\top)^{(a,a)} = (\mathbf{A}\mathbf{X}\mathbf{B})^\top \mathbf{A}\mathbf{X}\mathbf{B} = \mathbf{X}^\top \mathbf{A}^\top \mathbf{A}\mathbf{X} = \sum_{j \in [p]} \alpha_j^2 \mathbf{X} \mathbf{e}_{i(j)} \mathbf{e}_{i(j)}^\top \mathbf{X}^\top = \sum_{j \in [p]} \alpha_j^2 \mathbf{S}^{i(j)}. \quad \square$$

2.4. Partial Failure of Lifted Shor Relaxation Without Projection Matrices

This subsection² justifies the use of a mixed-projection reformulation of Problem (1) to design our convex relaxations in Section 2. In the presence of a hard rank constraint only ($\lambda = 0$), it is common to decompose \mathbf{X} into $\mathbf{X} = \mathbf{U}\mathbf{V}^\top$ with $\mathbf{U} \in \mathbb{R}^{n \times k}$, $\mathbf{V} \in \mathbb{R}^{m \times k}$ to enforce $\text{rank}(\mathbf{X}) \leq k$ by design, as proposed in the literature (Burer and Monteiro 2003). However, we now show that a naive first-order lift of this reformulation cannot provide a better bound than the trivial convex relaxation obtained by disregarding the rank constraint.

To obtain a lifted relaxation, we vectorize $\mathbf{X}, \mathbf{U}, \mathbf{V}$ and let the matrix \mathbf{W} collect the appropriate lifted blocks. Analogously to Proposition 1, the corresponding lifted relaxation is then

$$\begin{aligned} & \min_{\substack{\mathbf{X} \in \mathbb{R}^{n \times m}, \mathbf{U} \in \mathbb{R}^{n \times k}, \\ \mathbf{V} \in \mathbb{R}^{m \times k}, \mathbf{W} \in \mathcal{S}_+^{n(m+k)+mk}}} \langle \mathbf{H}, \mathbf{W}_{x,x} \rangle + \langle \mathbf{D}, \mathbf{X} \rangle \\ & \text{s.t. } \begin{pmatrix} 1 & \text{vec}(\mathbf{X}^\top)^\top & \text{vec}(\mathbf{U})^\top & \text{vec}(\mathbf{V})^\top \\ \text{vec}(\mathbf{X}^\top) & \mathbf{W}_{x,x} & \mathbf{W}_{x,u} & \mathbf{W}_{x,v} \\ \text{vec}(\mathbf{U}) & \mathbf{W}_{x,u}^\top & \mathbf{W}_{u,u} & \mathbf{W}_{u,v} \\ \text{vec}(\mathbf{V}) & \mathbf{W}_{x,v}^\top & \mathbf{W}_{u,v}^\top & \mathbf{W}_{v,v} \end{pmatrix} \succeq 0, \\ & \sum_{j=1}^k \mathbf{W}_{u,v}^{(j,j)} = \mathbf{X}, \\ & \langle \mathbf{Q}_i, \mathbf{W}_{x,x} \rangle + \langle \mathbf{E}_i, \mathbf{X} \rangle \leq b_i, \quad \forall i \in I, \end{aligned} \quad (9)$$

where \mathbf{W} collects the lifted moment blocks and the constraint $\mathbf{X} = \sum_{j=1}^k \mathbf{W}_{u,v}^{(j,j)}$ encodes $\mathbf{X} = \mathbf{U}\mathbf{V}^\top$.

We now show that (9) is no stronger than simply dropping the rank constraint. Formally, we have the following result (proof deferred to Section EC.1.2):

PROPOSITION 3. *For any $k \geq 1$, Problem (9) attains the same optimal value as*

$$\begin{aligned} & \min_{\mathbf{X}, \mathbf{W}_{x,x}} \langle \mathbf{H}, \mathbf{W}_{x,x} \rangle + \langle \mathbf{D}, \mathbf{X} \rangle \\ & \text{s.t. } \begin{pmatrix} 1 & \text{vec}(\mathbf{X}^\top)^\top \\ \text{vec}(\mathbf{X}^\top) & \mathbf{W}_{x,x} \end{pmatrix} \succeq 0, \\ & \langle \mathbf{Q}_i, \mathbf{W}_{x,x} \rangle + \langle \mathbf{E}_i, \mathbf{X} \rangle \leq b_i, \quad \forall i \in I. \end{aligned} \quad (10)$$

Proposition 3 shows that the first-order lifting of the factorized formulation $\mathbf{X} = \mathbf{U}\mathbf{V}^\top$ does not encode any nontrivial rank information: once the rank-one condition on the moment matrix is dropped, the blocks involving \mathbf{U} and \mathbf{V} can always be completed so as to represent an arbitrary matrix \mathbf{X} . In other words, Proposition 3 holds because there are no additional constraints on \mathbf{U}, \mathbf{V} .

In contrast, our relaxation starts from the mixed-projection reformulation (2), which relies on a structured decomposition of \mathbf{X} .

In particular, we can recover a non-trivial compact semidefinite relaxation from a decomposition $\mathbf{X} = \mathbf{U}\mathbf{V}^\top$ if we further require that $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_k$. In this case, we must have $\mathbf{X}^\top\mathbf{X} = \mathbf{V}\mathbf{V}^\top$. These two relationships translate into additional constraints in the lifted semidefinite relaxation: $\text{tr}(\mathbf{W}_{u,u}^{(j,l)}) = \delta_{j,l} \ \forall j, l \in [k]$ and $\sum_{i \in [n]} \mathbf{W}_{x,x}^{(i,i)} = \sum_{j \in [k]} \mathbf{W}_{v,v}^{(j,j)}$, respectively. With these additional constraints, we recover our semidefinite relaxation (4) (proof deferred to Section EC.1.3):

PROPOSITION 4. *Let $\lambda = 0$. The following problem attains the same optimal value as (4):*

$$\begin{aligned}
& \min_{\substack{\mathbf{X} \in \mathbb{R}^{n \times m}, \mathbf{U} \in \mathbb{R}^{n \times k}, \\ \mathbf{V} \in \mathbb{R}^{m \times k}, \mathbf{W} \in \mathcal{S}_+^{n(m+k)+mk}}} \langle \mathbf{H}, \mathbf{W}_{x,x} \rangle + \langle \mathbf{D}, \mathbf{X} \rangle \\
& \text{s.t.} \quad \begin{pmatrix} 1 & \text{vec}(\mathbf{X}^\top)^\top & \text{vec}(\mathbf{U})^\top & \text{vec}(\mathbf{V})^\top \\ \text{vec}(\mathbf{X}^\top) & \mathbf{W}_{x,x} & \mathbf{W}_{x,u} & \mathbf{W}_{x,v} \\ \text{vec}(\mathbf{U}) & \mathbf{W}_{x,u}^\top & \mathbf{W}_{u,u} & \mathbf{W}_{u,v} \\ \text{vec}(\mathbf{V}) & \mathbf{W}_{x,v}^\top & \mathbf{W}_{u,v}^\top & \mathbf{W}_{v,v} \end{pmatrix} \succeq 0, \\
& \sum_{j=1}^k \mathbf{W}_{u,v}^{(j,j)} = \mathbf{X}, \text{tr}(\mathbf{W}_{u,u}^{(j,l)}) = \delta_{j,l} \ \forall j, l \in [k], \\
& \sum_{i \in [n]} \mathbf{W}_{x,x}^{(i,i)} = \sum_{j \in [k]} \mathbf{W}_{v,v}^{(j,j)}, \\
& \langle \mathbf{Q}_i, \mathbf{W}_{x,x} \rangle + \langle \mathbf{E}_i, \mathbf{X} \rangle \leq b_i, \ \forall i \in I.
\end{aligned} \tag{11}$$

The proof of Proposition 4 reveals that, upon imposing $\mathbf{X} = \mathbf{U}\mathbf{V}^\top, \mathbf{U}^\top\mathbf{U} = \mathbf{I}_k, \mathbf{X}^\top\mathbf{X} = \mathbf{V}\mathbf{V}^\top$, the block matrix $\sum_{j \in [k]} \mathbf{W}_{u,u}^{(j,j)}$ dominates the relaxed projection variable $\mathbf{Y} = \mathbf{X}\mathbf{\Theta}^\dagger\mathbf{X}^\top : \mathbf{\Theta} = \sum_{j \in [k]} \mathbf{W}_{v,v}^{(j,j)}$. All in all, a lifted relaxation of a Burer and Monteiro (2003) factorization recovers relaxation (4) in extended form, provided additional structure on the factors \mathbf{U}, \mathbf{V} is imposed.

REMARK 3. One can show that Proposition 3 holds even when adding the constraints $\text{tr}(\mathbf{W}_{u,u}^{(j,l)}) = \delta_{j,l} \ \forall j, l \in [k]$. To obtain a non-trivial relaxation in Proposition 4, it is crucial to observe that $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_k$ implies $\mathbf{X}^\top\mathbf{X} = \mathbf{V}\mathbf{V}^\top$ and to use the latter constraints to relate the diagonal blocks of $\mathbf{W}_{x,x}$ (which appears in the objective) with those of $\mathbf{W}_{v,v}$.

3. Examples of Low-Rank Relaxations

This section applies the lifted relaxation proposed in Section 2 to two important problems from the low-rank literature: (i) matrix completion (Section 3.1), which possesses a row-separable quadratic loss, and (ii) reduced rank regression (Section 3.2), which possesses an objective dependent explicitly on $\mathbf{X}\mathbf{X}^\top$. By exploiting their respective problem structure, we show that our lifted relaxation can be further reduced to an equivalent relaxation with no semidefinite matrices or constraints of dimension larger than the sum of the two dimensions of the low-rank matrix, as opposed to the $O((n+m)^2)$ dimension blocks in our lifted relaxations from the previous section.

Section EC.2 supports this discussion by applying our lifted relaxation technique to three additional families of low-rank problems: basis pursuit, sparse plus low-rank matrix decomposition, and low-rank factor analysis.

3.1. Matrix Completion

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

$$\min_{\mathbf{Y} \in \mathcal{Y}_n^k} \min_{\mathbf{X} \in \mathbb{R}^{n \times m}} \|\mathcal{P}(\mathbf{A}) - \mathcal{P}(\mathbf{X})\|_F^2 + \lambda \cdot \text{tr}(\mathbf{Y}) \quad \text{s.t.} \quad \mathbf{X} = \mathbf{Y}\mathbf{X}, \quad (12)$$

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

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

is a linear map which masks the hidden entries of \mathbf{A} by outputting a matrix equal to \mathbf{A} on Ω and 0 otherwise. By expanding the quadratic $\|\mathcal{P}(\mathbf{A}) - \mathcal{P}(\mathbf{X})\|_F^2 = \sum_{i \in [n]} \left(\sum_{j \in [m]: (i,j) \in \Omega} (X_{i,j} - A_{i,j})^2 \right)$, we can invoke Theorem 3 to obtain the following relaxation of (12)

$$\begin{aligned} \min_{\mathbf{Y} \in \text{Conv}(\mathcal{Y}_n^k)} \min_{\mathbf{X} \in \mathbb{R}^{n \times m}, \mathbf{S}^i \in \mathcal{S}_+^m} & \sum_{i \in [n]} \langle \mathbf{S}^i, \mathbf{H}^i \rangle - 2 \langle \mathcal{P}(\mathbf{X}), \mathcal{P}(\mathbf{A}) \rangle + \langle \mathcal{P}(\mathbf{A}), \mathcal{P}(\mathbf{A}) \rangle + \lambda \cdot \text{tr}(\mathbf{Y}) \\ \text{s.t.} & \mathbf{S}^i \succeq \mathbf{X}_{i,\cdot} \mathbf{X}_{i,\cdot}^\top, \left(\sum_{i \in [n]} \mathbf{S}^i \begin{array}{c} \mathbf{X}^\top \\ \mathbf{Y} \end{array} \right) \succeq \mathbf{0}. \end{aligned} \quad (13)$$

where \mathbf{H}^i is a diagonal matrix which takes entries $\mathbf{H}_{j,j}^i = 1$ if $(i,j) \in \Omega$ and $\mathbf{H}_{j,j}^i = 0$ otherwise, and we retain the constant term $\langle \mathcal{P}(\mathbf{A}), \mathcal{P}(\mathbf{A}) \rangle$ 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 (12), while Bertsimas et al. (2023b) requires the presence of an additional Frobenius regularization term $+\frac{1}{2\gamma} \|\mathbf{X}\|_F^2$ in the objective.

Because each matrix \mathbf{H}^i in Problem (13) is binary, we can eliminate the individual matrices \mathbf{S}^i and keep only one additional semidefinite variable $\Theta \in \mathcal{S}_+^m$ to encode their sum. Formally, we have the following result (proof deferred to Section EC.2.1):

PROPOSITION 5. *Define $\bar{\mathbf{H}} := \text{Diag}(\bar{\mathbf{h}})$ where $\bar{h}_j := \min_{i \in [n]} H_{j,j}^i$. Then, Problem (13) attains the same optimal value as the following optimization problem:*

$$\begin{aligned} \min_{\mathbf{Y} \in \text{Conv}(\mathcal{Y}_n^k)} \min_{\mathbf{X} \in \mathbb{R}^{n \times m}, \Theta \in \mathcal{S}_+^m} & \langle \bar{\mathbf{H}}, \Theta \rangle + \sum_{i \in [n]} \langle \mathbf{H}^i - \bar{\mathbf{H}}, \mathbf{X}_{i,\cdot} \mathbf{X}_{i,\cdot}^\top \rangle - 2 \langle \mathcal{P}(\mathbf{X}), \mathcal{P}(\mathbf{A}) \rangle + \langle \mathcal{P}(\mathbf{A}), \mathcal{P}(\mathbf{A}) \rangle + \lambda \text{tr}(\mathbf{Y}) \\ \text{s.t.} & \begin{pmatrix} \Theta & \mathbf{X}^\top \\ \mathbf{X} & \mathbf{Y} \end{pmatrix} \succeq \mathbf{0}. \end{aligned} \quad (14)$$

We point out that if $\bar{\mathbf{H}} = \mathbf{0}$, relaxation (14) is weak. Indeed, in this case, any completion that satisfies $\mathcal{P}(\mathbf{X} - \mathbf{A}) = \mathbf{0}$, $\mathbf{Y} = \epsilon \mathbf{I}$, $\Theta = \epsilon^{-1} \mathbf{X}^\top \mathbf{X}$ is feasible and attains objective value 0 as $\epsilon \rightarrow 0$. This situation ($\bar{\mathbf{H}} = \mathbf{0}$) occurs if each column is missing for at least one row (meaning $\min_{i \in [n]} H_{j,j}^i = 0$). In other words, to obtain nontrivial relaxations via the compact unregularized relaxation (14), there must exist at least one column j which is observed in every row. When \mathbf{X} represents a dataset of n p -dimensional observations, this means that there must exist at least one variable which is consistently available for all observations. For example, Goldberg et al. (2001) consider a collaborative filtering setting where all users are provided with universal queries before they begin rating other items and Athey et al. (2021) consider a panel data setting possibly involving complete rows for control units where the goal is to infer missing elements of a control outcome matrix.

On the other hand, to obtain non-trivial relaxations when $\bar{\mathbf{H}} = \mathbf{0}$, one must either add a Frobenius regularization term to the objective or strengthen the relaxation via projection cuts (or both).

With the additional regularization term $+\frac{1}{2\gamma} \|\mathbf{X}\|_F^2$ in the objective, our approach leads to relaxations of the form (13) after redefining $\mathbf{H}^i \leftarrow \mathbf{H}^i + \frac{1}{2\gamma} \mathbf{I}_m$. Per Proposition 2, our resulting relaxations are at least as strong as the relaxation of Bertsimas et al. (2023b). However, the matrices \mathbf{H}^i are no longer binary and the compact reformulation in Proposition 5 no longer holds.

Applying Corollary 1, Problem (13) can be strengthened by imposing projection cuts. Let $\mathcal{R} = \{r_1, \dots, r_{|\mathcal{R}|}\}$ be a nonempty subset of the rows of \mathbf{X} of cardinality $|\mathcal{R}| \geq k + 1$, let $\mathbf{E}_{\mathcal{R}} \in \{0, 1\}^{|\mathcal{R}| \times n}$ denote the corresponding row-selection matrix, and define $\mathbf{X}_{\mathcal{R}} := \mathbf{E}_{\mathcal{R}} \mathbf{X} \in \mathbb{R}^{|\mathcal{R}| \times m}$. Then, since $\text{rank}(\mathbf{X}_{\mathcal{R}}) \leq \text{rank}(\mathbf{X})$ and $\mathbf{E}_{\mathcal{R}} \mathbf{E}_{\mathcal{R}}^\top = \mathbf{I}_{|\mathcal{R}|}$, we have the following auxiliary valid inequalities for Problem (13):

$$\exists \mathbf{Z}_{\mathcal{R}} \in \text{Conv}(\mathcal{Y}_{|\mathcal{R}|}^k) \text{ s.t. } \begin{pmatrix} \sum_{i \in \mathcal{R}} \mathbf{S}^i & \mathbf{X}_{\mathcal{R}}^\top \\ \mathbf{X}_{\mathcal{R}} & \mathbf{Z}_{\mathcal{R}} \end{pmatrix} \succeq \mathbf{0}, \mathbf{E}_{\mathcal{R}} \mathbf{Y} \mathbf{E}_{\mathcal{R}}^\top \preceq \mathbf{Z}_{\mathcal{R}}. \quad (15)$$

We require that $|\mathcal{R}| > k$ in order that one cannot take $\mathbf{Z}_{\mathcal{R}} = \mathbf{I}$, rendering the cut trivial.

3.2. Reduced Rank Regression

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

$$\min_{\mathbf{X} \in \mathbb{R}^{p \times m}} \|\mathbf{B} - \mathbf{A}\mathbf{X}\|_F^2 + \mu \cdot \text{rank}(\mathbf{X}), \quad (16)$$

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 $\text{vec}(\mathbf{X})\text{vec}(\mathbf{X})^\top$ rather than the mathematically equivalent $\text{vec}(\mathbf{X}^\top)\text{vec}(\mathbf{X}^\top)^\top$ for notational convenience—we obtain (16)’s lifted relaxation

$$\begin{aligned} \min_{\mathbf{Y} \in \text{Conv}(\mathcal{Y}_m)} \min_{\mathbf{X} \in \mathbb{R}^{p \times m}, \mathbf{W} \in \mathcal{S}_+^{pm}} \left\langle \mathbf{A}^\top \mathbf{A}, \sum_{i \in [m]} \mathbf{W}^{(i,i)} \right\rangle + \langle \mathbf{B}, \mathbf{B} \rangle - 2\langle \mathbf{A}\mathbf{X}, \mathbf{B} \rangle + \mu \cdot \text{tr}(\mathbf{Y}) \\ \text{s.t. } \mathbf{W} \succeq \text{vec}(\mathbf{X})\text{vec}(\mathbf{X})^\top, \begin{pmatrix} \sum_{i \in [m]} \mathbf{W}^{(i,i)} & \mathbf{X} \\ \mathbf{X}^\top & \mathbf{Y} \end{pmatrix} \succeq \mathbf{0}, \end{aligned} \quad (17)$$

for which we have the following corollary to Theorem 3:

COROLLARY 2. *Problem (17) attains the same objective value as*

$$\begin{aligned} \min_{\mathbf{Y} \in \text{Conv}(\mathcal{Y}_m)} \min_{\mathbf{X} \in \mathbb{R}^{p \times m}, \boldsymbol{\theta} \in \mathcal{S}_+^p} \langle \mathbf{A}^\top \mathbf{A}, \boldsymbol{\theta} \rangle + \langle \mathbf{B}, \mathbf{B} \rangle - 2\langle \mathbf{A}\mathbf{X}, \mathbf{B} \rangle + \mu \cdot \text{tr}(\mathbf{Y}) \\ \text{s.t. } \begin{pmatrix} \boldsymbol{\theta} & \mathbf{X} \\ \mathbf{X}^\top & \mathbf{Y} \end{pmatrix} \succeq \mathbf{0}, \end{aligned} \quad (18)$$

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

Corollary 2 suggests other low-rank problems which are quadratic through $\mathbf{X}\mathbf{X}^\top$ (or $\mathbf{X}^\top\mathbf{X}$), 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.

In this section, we have leveraged Theorem 3 to show that for two quadratic low-rank problems (plus three additional examples in Section EC.2), it is possible to eliminate enough variables in the lifted relaxation that no matrices of size $n^2 \times n^2$ remain. That is, while naive lifted relaxations may appear to be too large to be useful, they can often be reduced to practical sizes.

4. Numerical Results

In this section, we report four numerical experiments on low-rank matrix completion problems. Sections 4.1–4.3 use synthetic instances to isolate the quality of our bounds compared to the matrix perspective relaxation of Bertsimas et al. (2023b), the impact of the projection cuts developed in this paper, and the scalability of our approach, respectively. Section 4.4 then evaluates the same modeling ideas on real datasets with native missingness patterns.

Our first experiment was conducted on a MacBook Pro laptop (Apple M3, 36 GB), using MOSEK version 10.2, Julia version 1.9, and JuMP.jl version 1.13.0. All other experiments were conducted on the Imperial High Performance Computing environment (AMD EPYC 7742, 1 TB RAM) using MOSEK version 11.0.30, Julia version 1.11, and JuMP.jl version 1.29.3. All solver parameters are set to their default values for all experiments except where explicitly stated otherwise.

All synthetic experiments in this section use the data generation process of Candès and Recht (2009): We construct a matrix of observations, $\mathbf{A}_{\text{full}} \in \mathbb{R}^{n \times m}$, from a rank- k model: $\mathbf{A}_{\text{full}} = \mathbf{U}\mathbf{V} + \epsilon\mathbf{Z}$, where the entries of $\mathbf{U} \in \mathbb{R}^{n \times k}$, $\mathbf{V} \in \mathbb{R}^{k \times m}$, and $\mathbf{Z} \in \mathbb{R}^{n \times m}$ are drawn independently from a standard

normal distribution, and $\epsilon \geq 0$ models the degree of noise. We fix $\lambda = 0, \epsilon = 0.1$ and set n, m, k as indicated in each experiment. We then sample a random subset $\Omega \subseteq [n] \times [m]$ of predefined size $|\Omega|$ uniformly at random (see also Candès and Recht 2009, section 1.1.2). Each reported result is averaged over 10 random seeds.

4.1. Comparison of Lifted Relaxations With Matrix Perspective Relaxation

We first evaluate the quality of our new relaxations, compared with the matrix perspective relaxation of Bertsimas et al. (2023b, MPRT). We consider the regularized problem

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

for some regularization parameter $\gamma > 0$, with $m = n = 8$ and $k = 2$. As $\gamma \rightarrow \infty$, we recover the solution of (12). We compare the lower bounds obtained by three different approaches: MPRT, our lifted relaxation (13), hereafter denoted ‘‘Lifted’’, and our lifted relaxation augmented with the projection cuts (15) for all three-row subsets $\mathcal{R} \subseteq [n]$, i.e., $|\mathcal{R}| = 3$ (‘‘Lifted-proj-cuts’’). Figure 1 reports the relative optimality gap achieved by each approach—using the alternating minimization method of Burer and Monteiro (2003), Jain et al. (2013) initialized with a truncated SVD of $\mathcal{P}(\mathbf{A})$ for the upper bound (absolute values are reported in Figures EC.1–EC.2)—as γ increases, for different proportions of sampled entries $p = |\Omega|/mn$. We compute the relative gap as $\text{Gap} = \frac{UB-LB}{UB}$ with $UB > 0$.

Consistent with Proposition 2, we observe that Lifted and Lifted-proj-cuts 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 $p = 0.95$, MPRT achieves an uninformative gap of 100% as γ increases (by returning a trivial lower bound of 0, see Figure EC.1) while Lifted achieves a gap lower than 50% for the largest values of γ . Lifted-proj-cuts has an average optimality gap of 0 for all values of γ (average absolute gap $< 10^{-5}$ for all γ). From this experiment, it seems that imposing the projection cuts leads to significantly tighter bounds in some circumstances ($p = 0.95$), and only moderately tighter bounds in others ($p = 0.5$). The computational cost of imposing projection cuts is also rather modest at this scale; an average solve time of around 0.3 seconds per instance with projection cuts compared with around 0.05 seconds without projection cuts at this scale (see Figure EC.3 for computational times).

4.2. Performance of Projection Cuts on Synthetic Data

Our second experiment isolates the effect of projection cuts on the strength and computational cost of our lifted relaxation (13). We use the same experimental setup as Figure 1, with varying n, k, γ and fixing $m = 10, \epsilon = 0.1, \frac{|\Omega|}{nm} = 0.5$. For each triple (n, k, γ) we compare three lower bounds:

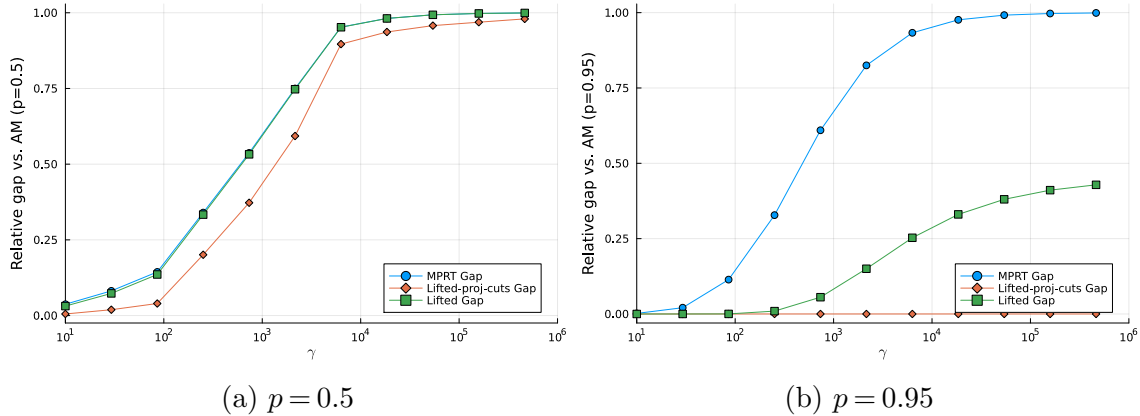


Figure 1 Relative gap obtained with different relaxations of the regularized matrix completion problem as we vary γ . We fix $n = 8$. Results are averaged over 10 replications.

the lifted relaxation (13) without projection cuts, the same relaxation with 100 randomly selected projection cuts (15) of cardinality $k + 1$, and the same relaxation with all projection cuts of cardinality $k + 1$. If fewer than 100 cuts are available, the random cut variant imposes all of them. We report average runtimes and average relative optimality gaps over 10 instances per triple (n, k, γ) , using the same alternating minimization setup as in Figure 1 to compute upper bounds. All SDP relaxations are solved with a relative optimality tolerance of 10^{-8} and a time limit of 16 hours. For SDP instances that return the status `SLOW_PROGRESS` (approximately 25% of instances), we use the objective bound when `Mosek` terminates as a surrogate for the semidefinite bound.

Table 1 shows that the base relaxation (13) is very strong when n is large or when γ is relatively small (i.e., the amount of regularization is large), but its gap increases as γ increases or as n decreases. Furthermore, we observe that projection cuts can substantially strengthen the lifted relaxation (13), especially in the weakly regularized regime where γ is very large. In this case, imposing all cardinality- $k + 1$ projection cuts often closes most of the remaining gap, at the cost of a substantially larger SDP. Moreover, the random cut variant captures a meaningful portion of this performance improvement while remaining much cheaper than imposing all cuts. This supports a simple strategy: solve the lifted relaxation without cuts, and perform alternating minimization, then impose projection cuts only when the base relaxation is not sufficiently strong.

4.3. Scalability of Lifted Relaxation on Synthetic Data

Next, motivated by the observation that relaxations without projection cuts may sometimes be preferable as n increases, our third experiment benchmarks the scalability of our reduced lifted relaxation (14) without projection cuts, as we increase n with $m = 50, k = 2$. We do not impose any Frobenius regularization for this experiment, and thus the method of Bertsimas et al. (2023b)

Table 1 Average performance of lifted relaxation (13) with and without the projection cuts (15) on synthetic $n \times 10$ matrix completion instances. “No. cuts” denotes the relaxation (13), “Rand.” denotes (13) augmented with 100 uniformly sampled projection cuts (15) of cardinality $k + 1$, while “All” denotes (13) augmented with all such cuts. Gaps are relative gaps $(UB - LB)/UB$, averaged over 10 replications; runtimes are in seconds. The reported regularization parameter is γ/nm . A dash denotes that an instance could not be solved within the given time/memory budget.

n	γ/nm	$k = 2$						$k = 3$					
		No cuts		Rand. cuts		All cuts		No cuts		Rand. cuts		All cuts	
		Gap (%)	T (s)	Gap (%)	T (s)	Gap (%)	T (s)	Gap (%)	T (s)	Gap (%)	T (s)	Gap (%)	T (s)
10	10^4	16.60%	3.34	4.98%	1.47	1.29 %	1.86	9.79%	3.37	7.91%	2.25	7.06 %	6.24
10	10^5	53.01%	3.38	7.99%	1.63	6.74 %	2.16	19.80%	3.37	17.47%	1.97	15.73 %	5.13
10	10^6	89.56%	3.38	27.97%	1.89	25.83 %	2.11	41.74%	3.32	39.48%	1.90	37.40 %	4.96
20	10^4	0.93%	4.28	0.04%	16.61	0.00 %	113.75	2.74%	4.19	1.27%	34.76	0.02 %	4234.1
20	10^5	25.17%	4.20	2.45%	23.50	0.00 %	120.29	17.32%	4.45	12.55%	38.26	0.49 %	4779.1
20	10^6	79.53%	4.82	21.95%	23.46	0.00 %	131.48	64.27%	4.20	56.13%	34.69	12.77 %	5439.7
40	10^4	0.00%	11.64	0.00%	1562.3	–	–	0.02%	23.98	0.00%	1891.7	–	–
40	10^5	0.63%	13.98	0.00%	1641.4	–	–	1.36%	13.80	0.32%	2751.7	–	–
40	10^6	43.23%	13.21	12.49%	2439.5	–	–	35.10%	13.02	26.95%	2948.2	–	–
60	10^4	0.00%	105.65	0.00%	20375.3	–	–	0.00%	108.16	–	–	–	–
60	10^5	0.00%	118.27	0.00%	19914.8	–	–	0.00%	131.56	–	–	–	–
60	10^6	20.46%	117.87	7.62%	42081.5	–	–	15.41%	119.30	–	–	–	–
80	10^4	0.00%	467.88	–	–	–	–	0.00%	496.72	–	–	–	–
80	10^5	0.00%	808.08	–	–	–	–	0.00%	467.07	–	–	–	–
80	10^6	5.09%	603.24	–	–	–	–	4.28%	568.44	–	–	–	–
100	10^4	0.00%	1720.5	–	–	–	–	0.00%	1577.2	–	–	–	–
100	10^5	0.00%	1995.8	–	–	–	–	0.00%	1803.1	–	–	–	–
100	10^6	0.46%	2521.3	–	–	–	–	0.51%	2747.1	–	–	–	–

is not applicable. We use the same data generation process as the previous experiment to generate \mathbf{A} . To ensure that our relaxation (14) is non-trivial, we include a universal-query component in the observation design. Specifically, we set $q = \lfloor \rho m \rfloor$, observe the first q columns in every row, and then observe each entry in the remaining $m - q$ columns independently with probability $\tau = \frac{p - q/m}{1 - q/m}$ in order that the expected number of observed entries is pnm . This combines the universal query motivation of Goldberg et al. (2001) with the missing completely at random sampling model of Rubin (1976).

We report the average lower bound over 10 random seeds (divided by nm so that quantities have the same scale as we vary n ; left) and the average computational time (right) in Figure 2 for $p = 0.4, \rho = 0.3$, with Mosek’s relative optimality tolerance set to 10^{-4} . We also report the average objective value obtained by running alternating minimization from 10 random initialization points and taking the best solution (Burer and Monteiro 2003, Jain et al. 2013) as a baseline. We observe that the lifted compact relaxation (14) scales up to $n = 350$. Further improvements in SDP algorithms and hardware may extend this range further. We observe similar scalability performance, albeit with a larger optimality gap, for $p = 0.4, \rho = 0.2$ in Figure EC.4.

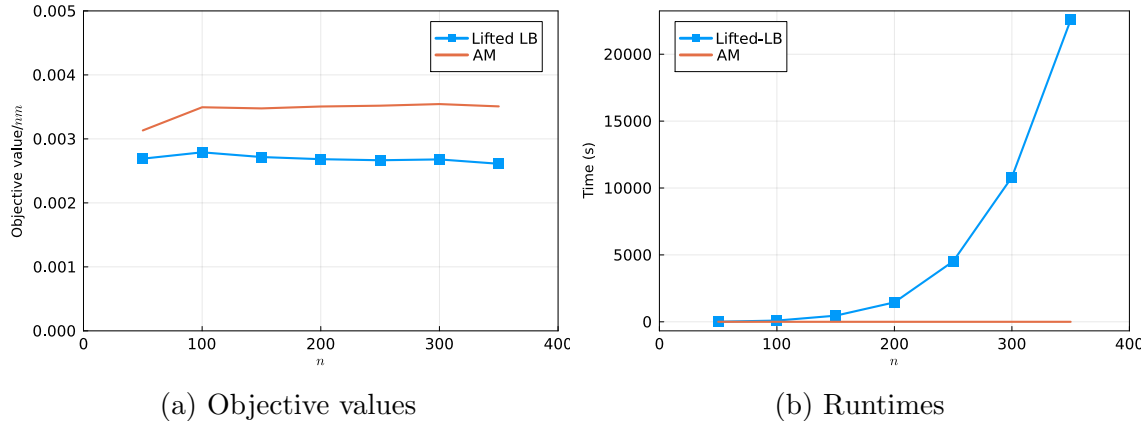


Figure 2 Normalized objective values (left panel) and runtime for the compact lifted relaxation (right panel) as we vary n with $m = 50$ for our compact lifted relaxation (14) with $p = 0.4, \rho = 0.3$.

4.4. Performance on Real Data

Our final experiment demonstrates the real-world applicability of our methods by evaluating our lifted relaxations on real missing-data matrices. We select five datasets from the UCI and RDatasets repos as indicated in Table 2. For each dataset, we drop the response variable and categorical variables, standardize numerical columns using observed entries, and preserve the native missingness pattern. We then construct ten 30×10 submatrices by selecting missing-affected numerical columns and rows enriched for missing entries, without masking any additional values. For $k \in \{2, 3\}$ and $\gamma/(nm) = 10^5$, we solve the lifted relaxation (13) with and without 1000 randomly sampled projection cuts (15) of cardinality $k + 1$. Table 2 reports average runtimes and relative gaps against the same alternating-minimization upper bound used above. All SDP solver parameters are the same as in the previous experiment, except that we impose a relative optimality tolerance of 10^{-8} , a memory limit of 128 GB of RAM, and a time limit of 8 hours.

We observe that the strength of our base relaxation (13) depends strongly on the dataset used, in the sense that it is tight (0% gap when $k = 3$) with no projection cuts on the movies dataset, and 100% with no projection cuts on the ambientNOxCH dataset. However, for datasets where the gap between our base relaxation (13) and alternating minimization is not 0%, imposing 1000 projection cuts is highly effective, reducing the optimality gap by more than half in most cases.

5. Conclusion

This paper develops scalable semidefinite relaxations for rank-constrained quadratic optimization problems. Our starting point is a mixed-projection reformulation of the rank-constrained problems and a lifting, obtained by introducing moment variables for the quadratic terms in \mathbf{X} and the projection matrix \mathbf{Y} . While a direct lifted relaxation involves large moment matrices with $n^2 \times n^2$

Table 2 Real-data results for the lifted relaxation (13) with and without 1000 randomly selected projection cuts (15). Each row reports averages over ten 30×10 submatrices constructed from the native missingness pattern of the indicated dataset. We fix $\gamma/(nm) = 10^5$. “Obs. frac.” denotes the average fraction of observed entries in the selected 30×10 submatrices, while “Miss. cols.” denotes the average number of the 10 selected columns that contain at least one missing entry.

Dataset	Source	Obs. frac.	Miss. cols.	$k = 2$				$k = 3$			
				No cuts		Rand.		No cuts		Rand.	
				Gap (%)	T (s)	Gap (%)	T (s)	Gap (%)	T (s)	Gap (%)	T (s)
Arrhythmia	UCI	0.891	2.7	26.00%	9.58	10.21 %	58.75	28.05%	8.97	14.84 %	104.43
Cars93	RDatasets	0.953	3	53.89%	9.24	8.32 %	54.74	64.93%	9.14	19.90 %	115.71
MCAS	RDatasets	0.882	3	35.04%	9.18	11.7 %	58.64	46.41%	9.85	23.66 %	124.89
ambientNOxCH	RDatasets	0.858	9.8	100%	9.33	24.48 %	67.66	100%	8.91	42.67 %	143
movies	RDatasets	0.806	2	0.06%	9.26	0.06 %	54.67	0.00%	9.26	0.00 %	100.4

semidefinite blocks, we show that most of these blocks are redundant and can be removed without altering the relaxation’s optimal value. Our key insight is that, in a lifted relaxation, the effective size of a low-rank relaxation is governed by the subset of moments of the matrix queried by the objective, side constraints, and coupled with a projection matrix, rather than by the size of the matrix squared.

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

Endnotes

1. The constraints $\mathbf{X} = \mathbf{Y}\mathbf{X}$, $\mathbf{Y}^2 = \mathbf{Y}$ imply that $\text{rank}(\mathbf{X}) \leq \text{tr}(\mathbf{Y})$, which can be made to hold with equality by letting $\mathbf{Y} = \mathbf{U}\mathbf{U}^\top$ for $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ a singular value decomposition of \mathbf{X} .
2. We are grateful to Amir Ali Ahmadi for an insightful question which prompted this subsection.

References

- Athey S, Bayati M, Doudchenko N, Imbens G, Khosravi K (2021) Matrix completion methods for causal panel data models. *Journal of the American Statistical Association* 116(536):1716–1730.
- 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, Recht B (2009) Exact matrix completion via convex optimization. *Foundations of Computational Mathematics* 9(6):717–772.
- Goldberg K, Roeder T, Gupta D, Perkins C (2001) Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval* 4(2):133–151.
- Gupta AK, Nagar DK (2018) *Matrix Variate Distributions* (Chapman and Hall/CRC).
- 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.
- 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.
- Rubin DB (1976) Inference and missing data. *Biometrika* 63(3):581–592.
- Shor NZ (1987) Quadratic optimization problems. *Soviet Journal of Computer and Systems Sciences* 25:1–11.
- Wang AL, Kilinç-Karzan F (2022) On the tightness of SDP relaxations of QCQPs. *Mathematical Programming* 193(1):33–73.

Supplementary Material

EC.1. Supplemental Theoretical Results and Examples for Section 2

EC.1.1. Example Showing Valid Inequalities May Strengthen Lifted Relaxation

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

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

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:

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

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

- The relaxation of Bertsimas et al. (2023b) has an optimal objective value of 4.637.
- The semidefinite relaxation (13) has an objective value of 5.0875.
- The semidefinite relaxation (13) augmented with all projection cuts (15) for $|\mathcal{R}| = 3$ has an objective value of 10.142.
- The method of Burer and Monteiro (2003) finds a feasible solution with objective value 10.142. That is, the lifted relaxation itself has an optimality gap of around 50%, but projection cuts are sufficient to close this gap to 0.

EC.1.2. Proof of Proposition 3

Proof of Proposition 3 Let $(\mathbf{X}, \mathbf{U}, \mathbf{V}, \mathbf{W})$ be feasible for (9). Then, by considering the appropriate principal submatrix it is immediate that

$$\begin{pmatrix} 1 & \text{vec}(\mathbf{X}^\top)^\top \\ \text{vec}(\mathbf{X}^\top) & \mathbf{W}_{x,x} \end{pmatrix} \succeq 0.$$

Moreover, the pair $(\mathbf{X}, \mathbf{W}_{x,x})$ is feasible for (10) and attains the same objective value as in (9). Therefore, the optimal value of (9) is at least that of (10).

Conversely, fix any feasible solution $(\mathbf{X}, \mathbf{W}_{x,x})$ of (10). Set $\mathbf{U} = \mathbf{0}$, $\mathbf{V} = \mathbf{0}$, $\mathbf{W}_{x,u} = \mathbf{0}$, $\mathbf{W}_{x,v} = \mathbf{0}$. Next, partition $\mathbf{W}_{u,v}$ into $k \times k$ blocks $\mathbf{W}_{u,v}^{(j,\ell)} \in \mathbb{R}^{n \times m}$ and define $\mathbf{W}_{u,v}^{(1,1)} := \mathbf{X}$, $\mathbf{W}_{u,v}^{(j,\ell)} := \mathbf{0}$ for $(j,\ell) \neq (1,1)$. Next, set $\mathbf{W}_{u,u}^{(j,j)} := \frac{1}{n} \mathbf{I}_n \ \forall j \in [k]$, $\mathbf{W}_{u,u}^{(j,l)} := \mathbf{0} \ \forall j, l \in [k] : j \neq l$, $\mathbf{W}_{v,v}^{(1,1)} := n \mathbf{X}^\top \mathbf{X}$ with every other block of $\mathbf{W}_{v,v}$ equal to zero. By construction, $\sum_{j=1}^k \mathbf{W}_{u,v}^{(j,j)} = \mathbf{X}$ Moreover, the matrix

$$\begin{pmatrix} \mathbf{W}_{u,u} & \mathbf{W}_{u,v} \\ \mathbf{W}_{u,v}^\top & \mathbf{W}_{v,v} \end{pmatrix}$$

is block diagonal, with the only nontrivial coupled block being

$$\begin{pmatrix} \frac{1}{n} \mathbf{I}_n & \mathbf{X} \\ \mathbf{X}^\top & n \mathbf{X}^\top \mathbf{X} \end{pmatrix},$$

which is clearly positive semidefinite. Therefore, the full moment matrix in (9) is block diagonal:

$$\begin{pmatrix} 1 & \text{vec}(\mathbf{X}^\top)^\top & 0 & 0 \\ \text{vec}(\mathbf{X}^\top) & \mathbf{W}_{x,x} & 0 & 0 \\ 0 & 0 & \mathbf{W}_{u,u} & \mathbf{W}_{u,v} \\ 0 & 0 & \mathbf{W}_{u,v}^\top & \mathbf{W}_{v,v} \end{pmatrix} \succeq 0,$$

because its two diagonal blocks are positive semidefinite. Thus, the constructed point is feasible for (9) and attains the same objective value as $(\mathbf{X}, \mathbf{W}_{x,x})$ in (10). It follows that the optimal value of (9) is at most that of (10). Combining the two inequalities yields the result. \square

REMARK EC.1. The feasible solution constructed for the reverse inequality also satisfies the constraint $\text{tr}(\mathbf{W}_{u,u}^{(j,l)}) = \delta_{j,l}, \forall j, l \in [k]$.

EC.1.3. Proof of Proposition 4

Proof of Proposition 4 Let z_{BM} denote the optimal value of (11) and z_{proj} denote the optimal value of (4). We show that $z_{\text{BM}} = z_{\text{proj}}$.

First, let $(\mathbf{X}, \mathbf{Y}, \mathbf{W}_{x,x})$ be feasible for (4). We show that $(\mathbf{X}, \mathbf{Y}, \mathbf{W}_{x,x})$ generates a solution in (11) with an equal or lower objective, and hence $z_{\text{BM}} \leq z_{\text{proj}}$.

To show this, define $\Theta := \sum_{i=1}^n \mathbf{W}_{x,x}^{(i,i)}$, $\tau := \frac{k - \text{tr}(\mathbf{Y})}{kn} \geq 0$. Moreover, set $\mathbf{U} = \mathbf{V} = \mathbf{0}$, $\mathbf{W}_{x,u} = \mathbf{W}_{x,v} = \mathbf{0}$, all off-diagonal factor blocks to zero, and for each $j \in [k]$ define

$$\mathbf{W}_{u,u}^{(j,j)} := \frac{1}{k} \mathbf{Y} + \tau \mathbf{I}_n, \quad \mathbf{W}_{u,v}^{(j,j)} := \frac{1}{k} \mathbf{X}, \quad \mathbf{W}_{v,v}^{(j,j)} := \frac{1}{k} \Theta.$$

Then, we have

$$\text{tr}(\mathbf{W}_{u,u}^{(j,j)}) = 1, \quad \sum_{j=1}^k \mathbf{W}_{u,v}^{(j,j)} = \mathbf{X}, \quad \sum_{j=1}^k \mathbf{W}_{v,v}^{(j,j)} = \Theta.$$

Moreover, for each j ,

$$\begin{pmatrix} \mathbf{W}_{u,u}^{(j,j)} & \mathbf{W}_{u,v}^{(j,j)} \\ (\mathbf{W}_{u,v}^{(j,j)})^\top & \mathbf{W}_{v,v}^{(j,j)} \end{pmatrix} = \frac{1}{k} \begin{pmatrix} \mathbf{Y} & \mathbf{X} \\ \mathbf{X}^\top & \Theta \end{pmatrix} + \begin{pmatrix} \tau \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \succeq \mathbf{0}.$$

Hence, the factor block is PSD, and together with

$$\begin{pmatrix} 1 & \text{vec}(\mathbf{X}^\top)^\top \\ \text{vec}(\mathbf{X}^\top) & \mathbf{W}_{x,x} \end{pmatrix} \succeq 0$$

this gives a feasible point of (11) with the same objective value and the same side constraints.

Conversely, let $(\mathbf{X}, \mathbf{U}, \mathbf{V}, \mathbf{W})$ be feasible for (11), and define

$$\tilde{\mathbf{Y}} := \sum_{j=1}^k \mathbf{W}_{u,u}^{(j,j)}, \quad \Theta := \sum_{j=1}^k \mathbf{W}_{v,v}^{(j,j)} = \sum_{i=1}^n \mathbf{W}_{x,x}^{(i,i)}.$$

Summing the principal (u_j, v_j) -blocks yields

$$\begin{pmatrix} \tilde{\mathbf{Y}} & \mathbf{X} \\ \mathbf{X}^\top & \Theta \end{pmatrix} \succeq \mathbf{0}, \quad \text{tr}(\tilde{\mathbf{Y}}) = k,$$

while the principal $(1, \mathbf{X})$ -block gives

$$\mathbf{W}_{x,x} \succeq \text{vec}(\mathbf{X}^\top) \text{vec}(\mathbf{X}^\top)^\top.$$

Summing the diagonal $m \times m$ blocks of the latter inequality gives

$$\Theta \succeq \mathbf{X}^\top \mathbf{X}.$$

Now set

$$\mathbf{Y} := \mathbf{X} \Theta^\dagger \mathbf{X}^\top.$$

Since $\Theta \succeq \mathbf{X}^\top \mathbf{X}$, we have $\ker(\Theta) \subseteq \ker(\mathbf{X})$, and therefore

$$\begin{pmatrix} \Theta & \mathbf{X}^\top \\ \mathbf{X} & \mathbf{Y} \end{pmatrix} \succeq \mathbf{0}.$$

Also, with $\mathbf{R} := \mathbf{X} \Theta^{\dagger/2}$,

$$\mathbf{R}^\top \mathbf{R} = \Theta^{\dagger/2} \mathbf{X}^\top \mathbf{X} \Theta^{\dagger/2} \preceq \Theta^{\dagger/2} \Theta \Theta^{\dagger/2} = P_{\text{range}(\Theta)} \preceq \mathbf{I},$$

so $\mathbf{0} \preceq \mathbf{Y} = \mathbf{R} \mathbf{R}^\top \preceq \mathbf{I}$. Finally, applying the generalized Schur complement to

$$\begin{pmatrix} \tilde{\mathbf{Y}} & \mathbf{X} \\ \mathbf{X}^\top & \Theta \end{pmatrix} \succeq \mathbf{0}$$

gives

$$\mathbf{Y} = \mathbf{X} \Theta^\dagger \mathbf{X}^\top \preceq \tilde{\mathbf{Y}},$$

hence

$$\text{tr}(\mathbf{Y}) \leq \text{tr}(\tilde{\mathbf{Y}}) = k.$$

Thus $(\mathbf{X}, \mathbf{Y}, \mathbf{W}_{x,x})$ is feasible for (4) with the same objective value and the same side constraints.

Therefore, $z_{\text{proj}} \leq z_{\text{BM}}$. \square

EC.2. Additional Examples Supporting Section 3

In this section, we first provide a proof of Proposition 5 and then support our discussion in Section 3 by providing three additional families of low-rank problems that admit compact semidefinite relaxations using the lifting outlined in Section 2.1.

EC.2.1. Proof of Proposition 5

Proof of Proposition 5 We fix (\mathbf{X}, \mathbf{Y}) in (13) and consider the partial minimization problem with respect to \mathbf{S}^i . We write

$$\mathbf{S}^i = \mathbf{X}_{i,\cdot} \mathbf{X}_{i,\cdot}^\top + \mathbf{U}^i, \quad \mathbf{U}^i \succeq \mathbf{0}.$$

Then, by the Schur complement lemma, the slack matrices \mathbf{U}^i must satisfy

$$\sum_{i \in [n]} \mathbf{U}^i \succeq \mathbf{R} := \mathbf{X}^\top \mathbf{Y}^\dagger \mathbf{X} - \mathbf{X}^\top \mathbf{X},$$

whenever $\text{range}(\mathbf{X}) \subseteq \text{range}(\mathbf{Y})$. Since $\mathbf{0} \preceq \mathbf{Y} \preceq \mathbf{I}$, we have $\mathbf{R} \succeq \mathbf{0}$. Minimizing for \mathbf{S}^i is then equivalent to solving

$$\phi_H(\mathbf{R}) := \min_{\substack{\mathbf{U}^i \succeq \mathbf{0} \\ \sum_i \mathbf{U}^i \succeq \mathbf{R}}} \sum_{i \in [n]} \langle \mathbf{H}^i, \mathbf{U}^i \rangle = \max_{\substack{\mathbf{M} \succeq \mathbf{0} \\ \mathbf{M} \preceq \mathbf{H}^i \quad \forall i \in [n]}} \langle \mathbf{M}, \mathbf{R} \rangle,$$

where strong duality holds because the minimization problem satisfies Slater's constraint qualification.

When the matrices \mathbf{H}^i are diagonal and binary, we have

$$\{\mathbf{M} \succeq \mathbf{0} : \mathbf{M} \preceq \mathbf{H}^i \quad \forall i \in [n]\} = \{\mathbf{M} : \mathbf{0} \preceq \mathbf{M} \preceq \bar{\mathbf{H}}\}.$$

Indeed, let S^i denote the support of the matrix \mathbf{H}^i . By construction, the support of $\bar{\mathbf{H}}$ is $\bar{S} := \bigcap_{i \in [n]} S^i$. If $\mathbf{M} \preceq \mathbf{H}^i$ for some $i \in [n]$, then the matrix \mathbf{M} is supported only on the coordinates in S^i . Applying this reasoning for all $i \in [n]$, we can say that matrices in $\{\mathbf{M} \succeq \mathbf{0} : \mathbf{M} \preceq \mathbf{H}^i \quad \forall i \in [n]\}$ are only supported on coordinates in \bar{S} . Restricted to \bar{S} , the diagonal matrices $\mathbf{H}^i, i \in [n]$ and $\bar{\mathbf{H}}$ are all equal to the identity matrix, showing that both sets are equal. Hence, because $\mathbf{R} \succeq \mathbf{0}$, we must have $\phi_H(\mathbf{R}) = \langle \bar{\mathbf{H}}, \mathbf{R} \rangle$. Therefore, for fixed (\mathbf{X}, \mathbf{Y}) , the optimized \mathbf{S}^i -dependent contribution in (13) equals

$$\sum_{i \in [n]} \langle \mathbf{H}^i, \mathbf{X}_{i,\cdot} \mathbf{X}_{i,\cdot}^\top \rangle + \langle \bar{\mathbf{H}}, \mathbf{X}^\top \mathbf{Y}^\dagger \mathbf{X} - \mathbf{X}^\top \mathbf{X} \rangle,$$

which can be rewritten as

$$\sum_{i \in [n]} \langle \mathbf{H}^i - \bar{\mathbf{H}}, \mathbf{X}_{i,\cdot} \mathbf{X}_{i,\cdot}^\top \rangle + \langle \bar{\mathbf{H}}, \mathbf{X}^\top \mathbf{Y}^\dagger \mathbf{X} \rangle.$$

Finally, minimizing $\langle \bar{\mathbf{H}}, \Theta \rangle$ subject to the Schur complement constraint leads to the objective $\langle \bar{\mathbf{H}}, \mathbf{X}^\top \mathbf{Y}^\dagger \mathbf{X} \rangle$ since $\bar{\mathbf{H}} \succeq \mathbf{0}$. \square

EC.2.2. Sparse Plus Low-Rank Matrix Decomposition

In this section, we leverage the techniques proposed in this paper to obtain a strong convex relaxation for the sparse plus low-rank matrix decomposition problem:

$$\min_{\mathbf{X} \in \mathbb{R}^{n \times m}, \mathbf{F} \in \mathbb{R}^{n \times m}} \|\mathbf{A} - \mathbf{X} - \mathbf{F}\|_F^2 \quad \text{s.t.} \quad \text{rank}(\mathbf{X}) \leq k, \|\mathbf{F}\|_0 \leq s, \quad (\text{EC.1})$$

where \mathbf{A} is a data matrix which is to be approximately decomposed into a low-rank matrix \mathbf{X} plus a sparse matrix \mathbf{F} . We remark that convex relaxations for this problem have been proposed for the special case where the objective is augmented with regularization by Bertsimas et al. (2023a). However, their matrix perspective relaxation gives a trivial lower bound of 0 for (EC.1).

We now demonstrate that a non-trivial convex relaxation can be obtained by leveraging the ideas described in this paper. Specifically, we follow Section 2.1 in introducing a projection matrix $\mathbf{Y} \in \mathcal{Y}_n^k$ to model the rank of \mathbf{X} , a binary matrix \mathbf{Z} to model the sparsity of \mathbf{F} , and imposing the bilinear constraints $\mathbf{X} = \mathbf{Y}\mathbf{X}$, $\mathbf{F} = \mathbf{Z} \circ \mathbf{F}$. This gives the following reformulation of (EC.1):

$$\min_{\mathbf{Z} \in \{0,1\}^{n \times m}, \mathbf{Y} \in \mathcal{Y}_n^k} \min_{\mathbf{X} \in \mathbb{R}^{n \times m}, \mathbf{F} \in \mathbb{R}^{n \times m}} \|\mathbf{A} - \mathbf{X} - \mathbf{F}\|_F^2 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{Y}\mathbf{X}, \mathbf{F} = \mathbf{Z} \circ \mathbf{F}, \text{tr}(\mathbf{Y}) \leq k, \langle \mathbf{E}, \mathbf{Z} \rangle \leq s, \quad (\text{EC.2})$$

which is nonconvex through quadratic constraints and thus is amenable to a lifted relaxation. Following Proposition 1, we vectorize all matrices, introduce a lifted matrix, and relax the constraints $\mathbf{Z} \in \{0,1\}^{n \times m}$ and $\mathbf{Y} \in \mathcal{Y}_n^k$ to obtain the following semidefinite relaxation of (EC.1).

$$\begin{aligned} & \min_{\substack{\mathbf{X}, \mathbf{F} \in \mathbb{R}^{n \times m}, \mathbf{Y} \in \text{Conv}(\mathcal{Y}_n^k), \\ \mathbf{Z} \in [0,1]^{n \times m}: \langle \mathbf{E}, \mathbf{Z} \rangle \leq s, \\ \mathbf{W}_{x,x}, \mathbf{W}_{y,y}, \mathbf{W}_{f,f}, \mathbf{W}_{z,z} \succeq 0, \\ \mathbf{W}_{x,y}, \mathbf{W}_{x,f}, \mathbf{W}_{f,z}, \mathbf{W}_{x,z}, \mathbf{W}_{y,f}, \mathbf{W}_{y,z}}} & \langle \mathbf{A}, \mathbf{A} \rangle - 2\langle \mathbf{A}, \mathbf{X} + \mathbf{F} \rangle + \text{tr}(\mathbf{W}_{x,x}) + \text{tr}(\mathbf{W}_{f,f}) + 2\text{tr}(\mathbf{W}_{x,f}) \quad (\text{EC.3}) \\ & \text{s.t.} \quad \begin{pmatrix} 1 & \text{vec}(\mathbf{X}^\top)^\top & \text{vec}(\mathbf{Y}^\top)^\top & \text{vec}(\mathbf{F}^\top)^\top & \text{vec}(\mathbf{Z}^\top)^\top \\ \text{vec}(\mathbf{X}^\top) & \mathbf{W}_{x,x} & \mathbf{W}_{x,y} & \mathbf{W}_{x,f} & \mathbf{W}_{x,z} \\ \text{vec}(\mathbf{Y}^\top) & \mathbf{W}_{x,y}^\top & \mathbf{W}_{y,y} & \mathbf{W}_{y,f} & \mathbf{W}_{y,z} \\ \text{vec}(\mathbf{F}^\top) & \mathbf{W}_{x,f}^\top & \mathbf{W}_{y,f}^\top & \mathbf{W}_{f,f} & \mathbf{W}_{f,z} \\ \text{vec}(\mathbf{Z}^\top) & \mathbf{W}_{x,z}^\top & \mathbf{W}_{y,z}^\top & \mathbf{W}_{f,z}^\top & \mathbf{W}_{z,z} \end{pmatrix} \succeq \mathbf{0}, \\ & \sum_{i \in [n]} \mathbf{W}_{y,y}^{(i,i)} = \mathbf{Y}, \quad \sum_{i \in [n]} \mathbf{W}_{x,y}^{(i,i)} = \mathbf{X}^\top, \\ & \text{diag}(\mathbf{W}_{z,z}) = \text{vec}(\mathbf{Z}^\top), \quad \text{diag}(\mathbf{W}_{f,z}) = \text{vec}(\mathbf{F}^\top). \end{aligned}$$

This relaxation is *not* practically tractable. However, we observe that, for the support variables \mathbf{Z} , only the *diagonal* product information is needed, and that the Frobenius objective depends only on row-wise quadratic terms, so we can fix all off-diagonal moment blocks without loss of generality, using the same off-diagonal elimination arguments as in Theorem 1 and Theorem 3 (note that Theorem 3 is not directly applicable due to the sparse matrix). This gives rise to the following application of Theorem 1 to Problem (EC.3), where \mathbf{E} denotes a matrix of all ones:

PROPOSITION EC.1. *Problem (EC.3) attains the same value as the following compact relaxation:*

$$\begin{aligned}
& \min_{\substack{\mathbf{X}, \mathbf{F}, \mathbf{T}, \mathbf{V} \in \mathbb{R}^{n \times m}, \mathbf{S}^i \in \mathcal{S}_+^m, \\ \mathbf{Y} \in \text{Conv}(\mathcal{Y}_n^k), \mathbf{Z} \in [0,1]^{n \times m}; \langle \mathbf{E}, \mathbf{Z} \rangle \leq s}} \langle \mathbf{A}, \mathbf{A} \rangle - 2\langle \mathbf{A}, \mathbf{X} + \mathbf{F} \rangle + \sum_{i \in [n]} \text{tr}(\mathbf{S}^i) + \langle \mathbf{E}, \mathbf{T} \rangle + 2\langle \mathbf{E}, \mathbf{V} \rangle \quad (\text{EC.4}) \\
& \text{s.t.} \quad \begin{pmatrix} \sum_{i \in [n]} \mathbf{S}^i & \mathbf{X}^\top \\ \mathbf{X} & \mathbf{Y} \end{pmatrix} \succeq \mathbf{0}, \quad \mathbf{S}^i \succeq \mathbf{X}_{i\cdot}, \mathbf{X}_{i\cdot}^\top, \quad \forall i \in [n], \\
& \quad \begin{pmatrix} T_{ij} & F_{ij} \\ F_{ij} & Z_{ij} \end{pmatrix} \succeq 0, \quad \begin{pmatrix} 1 & X_{ij} & F_{ij} \\ X_{ij} & S_{jj}^i & V_{ij} \\ F_{ij} & V_{ij} & T_{ij} \end{pmatrix} \succeq 0, \quad \forall (i, j) \in [n] \times [m].
\end{aligned}$$

REMARK EC.2. If only entries $\Omega \subseteq [n] \times [m]$ are observed in \mathbf{A} , we can replace the original objective function by $\|\mathcal{P}_\Omega(\mathbf{A} - \mathbf{X} - \mathbf{F})\|_F^2$ as in Section 3.1. The convex relaxation (EC.4) can be applied to this modified problem after modifying (EC.4)'s objective function with row masks \mathbf{H}^i as in Section 3.1. Further, these row masks allow us to make the relaxation more tractable by omitting the variables $V_{i,j}, T_{i,j}, Z_{i,j}$ for $(i, j) \notin \Omega$ and corresponding 2×2 and 3×3 minors.

Proof of Proposition EC.1 We show that for any feasible solution to (EC.3), there exists a feasible solution to (EC.4) with the same objective value and vice versa.

To prove the result, we require additional notation. Namely, since $\text{vec}(\mathbf{X}^\top)$ stacks the rows of \mathbf{X} , we index its entries by $\ell(i, j) := (i-1)m + j$, $(i, j) \in [n] \times [m]$, so that $\text{vec}(\mathbf{X}^\top)_{\ell(i, j)} = X_{ij}$.

Problem (EC.4) is a relaxation of Problem (EC.3): Fix a solution to (EC.3) and define

$$\mathbf{S}^i := \mathbf{W}_{x,x}^{(i,i)} \in \mathcal{S}_+^m, \quad i \in [n], \quad T_{ij} := (\mathbf{W}_{f,f})_{\ell(i,j), \ell(i,j)}, \quad V_{ij} := (\mathbf{W}_{x,f})_{\ell(i,j), \ell(i,j)}.$$

We claim that $(\mathbf{X}, \mathbf{F}, \mathbf{Y}, \mathbf{Z}, \{\mathbf{S}^i\}_{i \in [n]}, \mathbf{T}, \mathbf{V})$ is feasible for (EC.4) and attains the same objective value. First, the fact that the constraints

$$\begin{pmatrix} \sum_{i \in [n]} \mathbf{W}_{x,x}^{(i,i)} & \sum_{i \in [n]} \mathbf{W}_{x,y}^{(i,i)} \\ \left(\sum_{i \in [n]} \mathbf{W}_{x,y}^{(i,i)} \right)^\top & \sum_{i \in [n]} \mathbf{W}_{y,y}^{(i,i)} \end{pmatrix} \succeq 0, \quad \begin{pmatrix} \sum_{i \in [n]} \mathbf{S}^i & \mathbf{X}^\top \\ \mathbf{X} & \mathbf{Y} \end{pmatrix} \succeq 0, \quad \mathbf{S}^i = \mathbf{W}_{x,x}^{(i,i)} \succeq \mathbf{X}_{i\cdot}, \mathbf{X}_{i\cdot}^\top$$

hold follows identically to the proof of Theorem 1.

Second, for any (i, j) , consider the principal submatrix of the lifted PSD constraint indexed by (F_{ij}, Z_{ij}) , i.e., the entries $\ell(i, j)$ in $\text{vec}(\mathbf{F}^\top)$ and $\text{vec}(\mathbf{Z}^\top)$. This gives

$$\begin{pmatrix} (\mathbf{W}_{f,f})_{\ell(i,j), \ell(i,j)} & (\mathbf{W}_{f,z})_{\ell(i,j), \ell(i,j)} \\ (\mathbf{W}_{f,z})_{\ell(i,j), \ell(i,j)} & (\mathbf{W}_{z,z})_{\ell(i,j), \ell(i,j)} \end{pmatrix} \succeq 0.$$

By the diagonal constraints in (EC.3), $(\mathbf{W}_{f,z})_{\ell(i,j), \ell(i,j)} = F_{ij}$ and $(\mathbf{W}_{z,z})_{\ell(i,j), \ell(i,j)} = Z_{ij}$. Hence

$$\begin{pmatrix} T_{ij} & F_{ij} \\ F_{ij} & Z_{ij} \end{pmatrix} \succeq 0,$$

which is exactly the second set of LMIs in (EC.4).

Third, for any (i, j) , the principal submatrix of the lifted PSD constraint indexed by $(1, X_{ij}, F_{ij})$, i.e., $(1, \text{vec}(\mathbf{X}^\top)_{\ell(i,j)}, \text{vec}(\mathbf{F}^\top)_{\ell(i,j)})$, yields

$$\begin{pmatrix} 1 & X_{ij} & F_{ij} \\ X_{ij} & (\mathbf{W}_{x,x})_{\ell(i,j),\ell(i,j)} & (\mathbf{W}_{x,f})_{\ell(i,j),\ell(i,j)} \\ F_{ij} & (\mathbf{W}_{x,f})_{\ell(i,j),\ell(i,j)} & (\mathbf{W}_{f,f})_{\ell(i,j),\ell(i,j)} \end{pmatrix} \succeq 0.$$

Using $(\mathbf{W}_{x,x})_{\ell(i,j),\ell(i,j)} = (\mathbf{S}^i)_{jj}$, the definitions of T_{ij} and V_{ij} , we obtain

$$\begin{pmatrix} 1 & X_{ij} & F_{ij} \\ X_{ij} & (\mathbf{S}^i)_{jj} & V_{ij} \\ F_{ij} & V_{ij} & T_{ij} \end{pmatrix} \succeq 0,$$

which proves $(\mathbf{X}, \mathbf{F}, \mathbf{Y}, \mathbf{Z}, \{\mathbf{S}^i\}_{i \in [n]}, \mathbf{T}, \mathbf{V})$ is indeed feasible for (EC.4) with the same objective.

Problem (EC.3) is a relaxation of (EC.4): Fix a feasible solution $(\mathbf{X}, \mathbf{F}, \mathbf{Y}, \mathbf{Z}, \{\mathbf{S}^i\}, \mathbf{T}, \mathbf{V})$ to (EC.4), with a view to construct a solution to (EC.3) with the same objective value.

First, define $\mathbf{W}_{x,x} \in \mathcal{S}_+^{nm}$ blockwise via $\mathbf{W}_{x,x}^{(i,i)} := \mathbf{S}^i$, $\mathbf{W}_{x,x}^{(i,k)} := \mathbf{X}_{i,\cdot} \mathbf{X}_{k,\cdot}^\top$ ($i \neq k$). Then, $\mathbf{W}_{x,x}$ is positive semidefinite by construction and $\text{tr}(\mathbf{W}_{x,x}) = \sum_i \text{tr}(\mathbf{S}^i)$.

Second, let $\Theta := \sum_{i \in [n]} \mathbf{S}^i$. From (EC.4) we have $\mathbf{Y} \succeq \mathbf{X} \Theta^\dagger \mathbf{X}^\top$. Since \mathbf{Y} does not appear in the objective, we may replace \mathbf{Y} by $\tilde{\mathbf{Y}} := \mathbf{X} \Theta^\dagger \mathbf{X}^\top \preceq \mathbf{Y}$, which preserves feasibility in $\text{Conv}(\mathcal{Y}_n^k)$. Therefore, define $\mathbf{U} := \Theta^\dagger \mathbf{X}^\top \in \mathbb{R}^{m \times n}$ and set the diagonal blocks $\mathbf{W}_{x,y}^{(i,i)} := \mathbf{S}^i \mathbf{U}$, $\mathbf{W}_{y,y}^{(i,i)} := \mathbf{U}^\top \mathbf{S}^i \mathbf{U}$. Then $\sum_i \mathbf{W}_{x,y}^{(i,i)} = \mathbf{X}^\top$ and $\sum_i \mathbf{W}_{y,y}^{(i,i)} = \tilde{\mathbf{Y}}$. By Theorem 1's construction, we can complete the remaining (off-diagonal) blocks of $\mathbf{W}_{x,y}$ and $\mathbf{W}_{y,y}$ so the lifted PSD constraint on $(1, \text{vec}(\mathbf{X}^\top), \text{vec}(\mathbf{Y}^\top))$ holds.

Third, let

$$\mathbf{M}_{x,y} := \begin{pmatrix} 1 & \text{vec}(\mathbf{X}^\top)^\top & \text{vec}(\tilde{\mathbf{Y}}^\top)^\top \\ \text{vec}(\mathbf{X}^\top) & \mathbf{W}_{x,x} & \mathbf{W}_{x,y} \\ \text{vec}(\tilde{\mathbf{Y}}^\top) & \mathbf{W}_{x,y}^\top & \mathbf{W}_{y,y} \end{pmatrix} \succeq \mathbf{0}$$

denote the positive semidefinite matrix constructed above, and let

$$g_0, \{g_{\ell(i,j)}\}_{(i,j) \in [n] \times [m]}, \{h_q\}_{q=1}^{n^2}$$

be vectors whose Gram matrix equals $\mathbf{M}_{x,y}$. In particular,

$$\langle g_0, g_{\ell(i,j)} \rangle = X_{ij}, \quad \|g_{\ell(i,j)}\|^2 = (\mathbf{S}^i)_{jj}.$$

For each (i, j) , write

$$r_{ij} := g_{\ell(i,j)} - X_{ij} g_0,$$

so that $r_{ij} \perp g_0$ and $\|r_{ij}\|^2 = (\mathbf{S}^i)_{jj} - X_{ij}^2$.

Next, fix $(i, j) \in [n] \times [m]$. Since

$$\begin{pmatrix} 1 & X_{ij} & F_{ij} \\ X_{ij} & (\mathbf{S}^i)_{jj} & V_{ij} \\ F_{ij} & V_{ij} & T_{ij} \end{pmatrix} \succeq 0,$$

there exists a vector q_{ij} such that

$$\langle g_0, q_{ij} \rangle = F_{ij}, \quad \langle g_{\ell(i,j)}, q_{ij} \rangle = V_{ij}, \quad \|q_{ij}\|^2 = T_{ij}.$$

For example, if $\|r_{ij}\|^2 > 0$, one may take

$$q_{ij} = F_{ij}g_0 + \alpha_{ij}r_{ij} + \beta_{ij}e_{ij}, \quad \alpha_{ij} := \frac{V_{ij} - X_{ij}F_{ij}}{\|r_{ij}\|^2},$$

where e_{ij} is a unit vector orthogonal to all previously defined vectors, and

$$\beta_{ij} := \sqrt{T_{ij} - F_{ij}^2 - \alpha_{ij}^2\|r_{ij}\|^2},$$

which is well-defined by the above 3×3 LMI. If $\|r_{ij}\|^2 = 0$, then the same LMI implies $V_{ij} = X_{ij}F_{ij}$, and we may take

$$q_{ij} = F_{ij}g_0 + \sqrt{T_{ij} - F_{ij}^2} e_{ij}.$$

Now define

$$u_{ij} := q_{ij} - F_{ij}g_0,$$

so that $u_{ij} \perp g_0$ and $\|u_{ij}\|^2 = T_{ij} - F_{ij}^2$. Since

$$\begin{pmatrix} T_{ij} & F_{ij} \\ F_{ij} & Z_{ij} \end{pmatrix} \succeq 0 \quad \text{and} \quad 0 \leq Z_{ij} \leq 1,$$

there exists a vector z_{ij} such that

$$\langle g_0, z_{ij} \rangle = Z_{ij}, \quad \langle q_{ij}, z_{ij} \rangle = F_{ij}, \quad \|z_{ij}\|^2 = Z_{ij}.$$

Indeed, if $\|u_{ij}\|^2 > 0$, we may take

$$z_{ij} = Z_{ij}g_0 + a_{ij}u_{ij} + b_{ij}\tilde{e}_{ij}, \quad a_{ij} := \frac{F_{ij}(1 - Z_{ij})}{\|u_{ij}\|^2},$$

where \tilde{e}_{ij} is a unit vector orthogonal to all previously defined vectors, and

$$b_{ij} := \sqrt{Z_{ij} - Z_{ij}^2 - a_{ij}^2\|u_{ij}\|^2},$$

which is well-defined because

$$Z_{ij} - Z_{ij}^2 - a_{ij}^2\|u_{ij}\|^2 = \frac{(1 - Z_{ij})(Z_{ij}T_{ij} - F_{ij}^2)}{T_{ij} - F_{ij}^2} \geq 0.$$

If $\|u_{ij}\|^2 = 0$, then $q_{ij} = F_{ij}g_0$, and we may take

$$z_{ij} = Z_{ij}g_0 + \sqrt{Z_{ij} - Z_{ij}^2} \tilde{e}_{ij},$$

noting that in this case $\langle q_{ij}, z_{ij} \rangle = F_{ij}$ follows from $F_{ij}^2 = T_{ij} \leq T_{ij}Z_{ij} = F_{ij}^2Z_{ij}$.

Finally, define the remaining lifted blocks so that the full lifted matrix in (EC.3) is the Gram matrix of

$$\{g_0\} \cup \{g_{\ell(i,j)}\}_{(i,j)} \cup \{h_q\}_{q=1}^{n^2} \cup \{q_{ij}\}_{(i,j)} \cup \{z_{ij}\}_{(i,j)}.$$

This matrix is positive semidefinite by construction, and satisfies

$$\text{diag}(W_{f,f}) = \vec{T}^\top, \quad \text{diag}(W_{f,z}) = \vec{F}^\top, \quad \text{diag}(W_{z,z}) = \vec{Z}^\top,$$

and also satisfies $\text{tr}(W_{x,f}) = \langle E, V \rangle$. Since the blocks $W_{x,z}$, $W_{y,f}$, $W_{y,z}$ are unconstrained in (EC.3), the Gram values assigned to them are admissible. Therefore, the constructed solution is feasible for (EC.3) and has the same objective value as in (EC.4). \square

EC.2.3. Example: Basis Pursuit

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

$$\min_{\mathbf{Y} \in \mathcal{Y}_n} \min_{\mathbf{X} \in \mathbb{R}^{n \times m}} \text{tr}(\mathbf{Y}) \text{ s.t. } \mathcal{P}(\mathbf{A}) = \mathcal{P}(\mathbf{X}), \mathbf{X} = \mathbf{Y}\mathbf{X}, \quad (\text{EC.5})$$

where $\mathcal{Y}_n := \mathcal{Y}_n^n$ denote the set of $n \times n$ orthogonal projection matrices of any rank, and $\mathcal{P}(\mathbf{A})$ denotes a linear map that masks the hidden entries of \mathbf{A} , \mathbf{X} such that $\mathcal{P}(\mathbf{A})_{i,j} = A_{i,j}$ if $(i,j) \in \Omega$ and 0 otherwise. Following Theorem 1 and pairwise multiplying the constraints $A_{i,j} - X_{i,j} = 0, \forall (i,j) \in \Omega$ leads to the following relaxation

$$\begin{aligned} \min_{\mathbf{Y} \in \text{Conv}(\mathcal{Y}_n)} \min_{\mathbf{X} \in \mathbb{R}^{n \times m}, \mathbf{W} \in \mathcal{S}_+^{nm}} \text{tr}(\mathbf{Y}) \\ \text{s.t. } & A_{i,j}A_{k,\ell} - A_{k,\ell}X_{i,j} - A_{i,j}X_{k,\ell} + (\mathbf{W}^{(i,k)})_{j,\ell} = 0, \forall (i,j), (k,\ell) \in \Omega \times \Omega \\ & A_{i,j} = X_{i,j}, \forall (i,j) \in \Omega \\ & \mathbf{W} \succeq \text{vec}(\mathbf{X}^\top)\text{vec}(\mathbf{X}^\top)^\top, \left(\begin{array}{c} \sum_{i \in [n]} \mathbf{W}^{(i,i)} \mathbf{X}^\top \\ \mathbf{X} \quad \mathbf{Y} \end{array} \right) \succeq \mathbf{0}, \end{aligned} \quad (\text{EC.6})$$

Similarly to the low-rank matrix completion case, the structure of the compact lifted relaxation means that if we set $\mathbf{W}_{x,x}^{(i,j)} = \mathbf{X}_{i,\cdot}\mathbf{X}_{j,\cdot}^\top$ for $i \neq j$ then $X_{i,j} = A_{i,j}$ and $X_{k,l} = A_{k,l}$ for observed entries $(i,j), (k,l) \in \Omega$. Hence, any constraints involving the off-diagonal blocks of $\mathbf{W}_{x,x}$ are satisfied by the construction in the proof of Theorem 3, and we have the following corollary to it:

COROLLARY EC.1. *Problem (EC.6) attains the same objective value as*

$$\begin{aligned} \min_{\mathbf{Y} \in \text{Conv}(\mathcal{Y}_n)} \min_{\mathbf{X} \in \mathbb{R}^{n \times m}, \mathbf{S}^i \in \mathcal{S}_+^m, i \in [n]} \text{tr}(\mathbf{Y}) \\ \text{s.t. } & A_{i,j}A_{i,\ell} - A_{i,\ell}X_{i,j} - A_{i,j}X_{i,\ell} + (\mathbf{S}^i)_{j,\ell} = 0, \forall (i,j), (i,\ell) \in \Omega \times \Omega \\ & A_{i,j} = X_{i,j}, \forall (i,j) \in \Omega \\ & \mathbf{S}^i \succeq \mathbf{X}_{i,\cdot}\mathbf{X}_{i,\cdot}^\top, \left(\begin{array}{c} \sum_{i \in [n]} \mathbf{S}^i \mathbf{X}^\top \\ \mathbf{X} \quad \mathbf{Y} \end{array} \right) \succeq \mathbf{0}, \end{aligned} \quad (\text{EC.7})$$

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

REMARK EC.3. The number of linear equality constraints in (EC.7) grows quadratically in $|\Omega|$, thus, in practice, one may subsample or aggregate these constraints to improve the tractability of the semidefinite relaxation.

EC.2.4. Example: Low-Rank Factor Analysis

In this section, we apply the lifted relaxation technique developed in Section 2.1 to derive a strong convex relaxation for low-rank factor analysis as explored in Bertsimas et al. (2017):

$$\min_{\mathbf{X} \in \mathcal{S}_+^n, \mathbf{D} \in \mathcal{D}} \|\mathbf{A} - \mathbf{X} - \mathbf{D}\|_F^2 \quad \text{s.t.} \quad \text{rank}(\mathbf{X}) \leq k, \quad (\text{EC.8})$$

where $\mathbf{A} \in \mathcal{S}_{++}^n$ is a covariance matrix which is to be approximately decomposed into a positive semidefinite low-rank matrix \mathbf{X} plus a diagonal matrix \mathbf{D} , which is contained in a convex set \mathcal{D} (e.g., $\mathcal{D} = \{\text{Diag}(\mathbf{d}) : \mathbf{d} \geq \mathbf{0}\}$). Note that (EC.8) is nonconvex *only* through the rank constraint.

As in Section 2.1, we introduce a projection matrix $\mathbf{Y} \in \mathcal{Y}_n^k$ and impose $\mathbf{X} = \mathbf{Y}\mathbf{X}$ to model the rank constraint. This yields the equivalent mixed-projection formulation

$$\min_{\mathbf{Y} \in \mathcal{Y}_n^k} \min_{\mathbf{X} \in \mathcal{S}_+^n, \mathbf{D} \in \mathcal{D}} \|\mathbf{A} - \mathbf{X} - \mathbf{D}\|_F^2 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{Y}\mathbf{X}. \quad (\text{EC.9})$$

As in previous examples, we now develop a semidefinite relaxation by vectorizing and lifting. This yields the following relaxation:

$$\begin{aligned} \min_{\substack{\mathbf{X} \in \mathcal{S}_+^n, \mathbf{d} \in \mathbb{R}^n, \mathbf{Y} \in \text{Conv}(\mathcal{Y}_n^k), \\ \text{Diag}(\mathbf{d}) \in \mathcal{D}, \mathbf{W}_{x,x}, \mathbf{W}_{y,y} \in \mathcal{S}_+^{n^2}, \\ \mathbf{W}_{d,d} \in \mathcal{S}_+^n, \mathbf{W}_{x,y} \in \mathbb{R}^{n^2 \times n^2}, \\ \mathbf{W}_{x,d} \in \mathbb{R}^{n^2 \times n}, \mathbf{W}_{y,d} \in \mathbb{R}^{n^2 \times n}}} \quad & \langle \mathbf{A}, \mathbf{A} \rangle - 2\langle \mathbf{A}, \mathbf{X} \rangle - 2\text{diag}(\mathbf{A})^\top \mathbf{d} + \text{tr}(\mathbf{W}_{x,x}) + \text{tr}(\mathbf{W}_{d,d}) + 2\langle \mathbf{P}, \mathbf{W}_{x,d} \rangle \end{aligned} \quad (\text{EC.10})$$

$$\begin{aligned} \text{s.t.} \quad & \begin{pmatrix} 1 & \text{vec}(\mathbf{X}^\top)^\top & \text{vec}(\mathbf{Y}^\top)^\top & \mathbf{d}^\top \\ \text{vec}(\mathbf{X}^\top) & \mathbf{W}_{x,x} & \mathbf{W}_{x,y} & \mathbf{W}_{x,d} \\ \text{vec}(\mathbf{Y}^\top) & \mathbf{W}_{x,y}^\top & \mathbf{W}_{y,y} & \mathbf{W}_{y,d} \\ \mathbf{d} & \mathbf{W}_{x,d}^\top & \mathbf{W}_{y,d}^\top & \mathbf{W}_{d,d} \end{pmatrix} \succeq \mathbf{0}, \\ & \sum_{i \in [n]} \mathbf{W}_{y,y}^{(i,i)} = \mathbf{Y}, \quad \sum_{i \in [n]} \mathbf{W}_{x,y}^{(i,i)} = \mathbf{X}^\top, \end{aligned}$$

where, to simplify notation, we define the diagonal selector $\mathbf{P} \in \mathbb{R}^{n^2 \times n}$ by $\text{vec}(\text{Diag}(\mathbf{d})) = \mathbf{P}\mathbf{d}$.

We now eliminate most lifted blocks without loss of generality, analogously to Theorem 1 and Theorem 3 (note that Theorem 3 is not directly applicable due to the diagonal matrix). Formally, we have the following result:

PROPOSITION EC.2. *Problem (EC.10) attains the same value as the following compact relaxation:*

$$\begin{aligned} \min_{\substack{\mathbf{X} \in \mathcal{S}_+^n, \mathbf{d} \in \mathbb{R}^n, \mathbf{t}, \mathbf{v} \in \mathbb{R}^n, \\ \mathbf{S}^i \in \mathcal{S}_+^n, \mathbf{Y} \in \text{Conv}(\mathcal{Y}_n^k), \text{Diag}(\mathbf{d}) \in \mathcal{D}}} \quad & \langle \mathbf{A}, \mathbf{A} \rangle - 2\langle \mathbf{A}, \mathbf{X} \rangle - 2\text{diag}(\mathbf{A})^\top \mathbf{d} + \sum_{i \in [n]} \text{tr}(\mathbf{S}^i) + \mathbf{e}^\top \mathbf{t} + 2\mathbf{e}^\top \mathbf{v} \quad (\text{EC.11}) \\ \text{s.t.} \quad & \begin{pmatrix} \sum_{i \in [n]} \mathbf{S}^i & \mathbf{X}^\top \\ \mathbf{X} & \mathbf{Y} \end{pmatrix} \succeq \mathbf{0}, \quad \mathbf{S}^i \succeq \mathbf{X}_{i,\cdot} \mathbf{X}_{i,\cdot}^\top \quad \forall i \in [n], \\ & \begin{pmatrix} 1 & X_{ii} & d_i \\ X_{ii} & S_{ii}^i & v_i \\ d_i & v_i & t_i \end{pmatrix} \succeq \mathbf{0}, \quad \forall i \in [n]. \end{aligned}$$

REMARK EC.4. Comparison with Bertsimas et al. (2017): Relaxation (EC.11) is a strong *one-shot* semidefinite lower bound for the least-squares objective in (EC.8), without requiring additional regularization or bounding-box refinements, and which also applies when \mathbf{A} is indefinite (although we lose the statistical interpretation in this case). By contrast, Bertsimas et al. (2017) develop semidefinite bounds tailored to classical factor analysis structure through a spectral reformulation that strongly leverages an additional constraint $\mathbf{A} - \mathbf{D} \succeq \mathbf{0}$, and their approach does not apply without this constraint. Thus, the approach derived here is more general, although it may be weaker after the McCormick branch-and-bound procedure in Bertsimas et al. (2017) is applied.

Proof of Proposition EC.2 We show that for any feasible solution to (EC.10), there exists a feasible solution to (EC.11) with the same objective value and vice versa.

To prove the result, we require additional notation. Since $\text{vec}(\mathbf{X}^\top)$ stacks the rows of \mathbf{X} , we index its entries by $\ell(i, j) := (i-1)n + j$, $(i, j) \in [n] \times [n]$, so that $\text{vec}(\mathbf{X}^\top)_{\ell(i, j)} = X_{ij}$.

Problem (EC.11) is a relaxation of Problem (EC.10): Fix a solution to (EC.10) and define

$$\mathbf{S}^i := \mathbf{W}_{x,x}^{(i,i)} \in \mathcal{S}_+^n, \quad i \in [n], \quad t_i := (\mathbf{W}_{d,d})_{ii}, \quad v_i := (\mathbf{W}_{x,d})_{\ell(i,i),i}, \quad i \in [n].$$

We claim that $(\mathbf{X}, \mathbf{d}, \mathbf{Y}, \{\mathbf{S}^i\}_{i \in [n]}, \mathbf{t}, \mathbf{v})$ is feasible for (EC.11) and attains the same objective value. First, the feasibility with respect to the constraints

$$\begin{pmatrix} \sum_{i \in [n]} \mathbf{S}^i & \mathbf{X}^\top \\ \mathbf{X} & \mathbf{Y} \end{pmatrix} \succeq \mathbf{0}, \quad \mathbf{S}^i = \mathbf{W}_{x,x}^{(i,i)} \succeq \mathbf{X}_i \mathbf{X}_i^\top,$$

follows identically to the proof of Theorem 1.

Second, for any $i \in [n]$, consider the principal submatrix of the lifted PSD constraint in (EC.10) indexed by $(1, \text{vec}(\mathbf{X}^\top)_{\ell(i,i)}, d_i)$, i.e., $(1, X_{ii}, d_i)$. This yields

$$\begin{pmatrix} 1 & X_{ii} & d_i \\ X_{ii} & (\mathbf{W}_{x,x})_{\ell(i,i),\ell(i,i)} & (\mathbf{W}_{x,d})_{\ell(i,i),i} \\ d_i & (\mathbf{W}_{x,d})_{\ell(i,i),i} & (\mathbf{W}_{d,d})_{ii} \end{pmatrix} \succeq \mathbf{0}.$$

Using $(\mathbf{W}_{x,x})_{\ell(i,i),\ell(i,i)} = (\mathbf{S}^i)_{ii}$ and the definitions of t_i and v_i gives

$$\begin{pmatrix} 1 & X_{ii} & d_i \\ X_{ii} & (\mathbf{S}^i)_{ii} & v_i \\ d_i & v_i & t_i \end{pmatrix} \succeq \mathbf{0},$$

which is exactly the second set of LMIs in (EC.11).

Finally, the objective values agree because $\text{tr}(\mathbf{W}_{x,x}) = \sum_{i \in [n]} \text{tr}(\mathbf{W}_{x,x}^{(i,i)}) = \sum_{i \in [n]} \text{tr}(\mathbf{S}^i)$, $\text{tr}(\mathbf{W}_{d,d}) = \sum_{i \in [n]} t_i$, and $\langle \mathbf{P}, \mathbf{W}_{x,d} \rangle = \sum_{i \in [n]} (\mathbf{W}_{x,d})_{\ell(i,i),i} = \mathbf{e}^\top \mathbf{v}$ by the definition of \mathbf{P} .

Problem (EC.10) is a relaxation of Problem (EC.11): Fix a feasible solution $(\mathbf{X}, \mathbf{d}, \mathbf{Y}, \{\mathbf{S}^i\}_{i \in [n]}, \mathbf{t}, \mathbf{v})$ to (EC.11), with a view to construct a feasible solution to (EC.10).

First, define $\mathbf{W}_{x,x} \in \mathcal{S}_+^{n^2}$ blockwise via $\mathbf{W}_{x,x}^{(i,i)} := \mathbf{S}^i$, $\mathbf{W}_{x,x}^{(i,k)} := \mathbf{X}_i \mathbf{X}_k^\top$ ($i \neq k$). Then $\mathbf{W}_{x,x} \succeq \mathbf{0}$ by construction and $\text{tr}(\mathbf{W}_{x,x}) = \sum_{i \in [n]} \text{tr}(\mathbf{S}^i)$.

Second, let $\Theta := \sum_{i \in [n]} \mathbf{S}^i$, $\mathbf{U} := \Theta^\dagger \mathbf{X}^\top \in \mathbb{R}^{n \times n}$ and set the diagonal blocks $\mathbf{W}_{x,y}^{(i,i)} := \mathbf{S}^i \mathbf{U}$ and $\mathbf{W}_{y,y}^{(i,i)} := \mathbf{U}^\top \mathbf{S}^i \mathbf{U}$. Then, identically to the proof of Theorem 1, we can complete the remaining (off-diagonal) blocks of $\mathbf{W}_{x,y}$ and $\mathbf{W}_{y,y}$ so that

$$\begin{pmatrix} 1 & \text{vec}(\mathbf{X}^\top)^\top & \text{vec}(\tilde{\mathbf{Y}}^\top)^\top \\ \text{vec}(\mathbf{X}^\top) & \mathbf{W}_{x,x} & \mathbf{W}_{x,y} \\ \text{vec}(\tilde{\mathbf{Y}}^\top) & \mathbf{W}_{x,y}^\top & \mathbf{W}_{y,y} \end{pmatrix} \succeq \mathbf{0}.$$

Third, let $\mathbf{M}_{x,y}$ denote the PSD matrix in the display above, and let $\mathbf{g}_0, \{\mathbf{g}_p\}_{p=1}^{n^2}, \{\mathbf{h}_q\}_{q=1}^{n^2}$ be vectors whose Gram matrix equals $\mathbf{M}_{x,y}$ (e.g., take a Cholesky factor of $\mathbf{M}_{x,y}$). In particular, $\|\mathbf{g}_0\|^2 = 1$, $\mathbf{g}_0^\top \mathbf{g}_{\ell(i,i)} = X_{ii}$, and $\|\mathbf{g}_{\ell(i,i)}\|^2 = (\mathbf{W}_{x,x})_{\ell(i,i),\ell(i,i)} = S_{ii}^i$. Next, fix $i \in [n]$ and write

$$\mathbf{g}_{\ell(i,i)} = X_{ii} \mathbf{g}_0 + \mathbf{r}_i, \quad \text{where } \mathbf{r}_i \perp \mathbf{g}_0, \quad \|\mathbf{r}_i\|^2 = S_{ii}^i - X_{ii}^2.$$

Introduce a new unit vector \mathbf{e}_i orthogonal to $\text{span}\{\mathbf{g}_0, \mathbf{g}_p, \mathbf{h}_q : p, q \in [n^2]\}$, and define

$$\alpha_i := \begin{cases} \frac{v_i - X_{ii} d_i}{\|\mathbf{r}_i\|^2}, & \text{if } \|\mathbf{r}_i\|^2 > 0, \\ 0, & \text{if } \|\mathbf{r}_i\|^2 = 0, \end{cases} \quad \beta_i := \sqrt{t_i - d_i^2 - \alpha_i^2 \|\mathbf{r}_i\|^2}.$$

The quantity under the square root is nonnegative because the constraint $\begin{pmatrix} 1 & X_{ii} & d_i \\ X_{ii} & S_{ii}^i & v_i \\ d_i & v_i & t_i \end{pmatrix} \succeq \mathbf{0}$ implies,

by Schur complement, that $\begin{pmatrix} \|\mathbf{r}_i\|^2 & v_i - X_{ii} d_i \\ v_i - X_{ii} d_i & t_i - d_i^2 \end{pmatrix} \succeq \mathbf{0}$. Now, set $\mathbf{q}_i := d_i \mathbf{g}_0 + \alpha_i \mathbf{r}_i + \beta_i \mathbf{e}_i$. Then, $\mathbf{g}_0^\top \mathbf{q}_i = d_i$, $\|\mathbf{q}_i\|^2 = t_i$, and

$$\mathbf{g}_{\ell(i,i)}^\top \mathbf{q}_i = X_{ii} d_i + \mathbf{r}_i^\top (\alpha_i \mathbf{r}_i) = X_{ii} d_i + \alpha_i \|\mathbf{r}_i\|^2 = v_i.$$

Define the remaining lifted blocks by

$$(\mathbf{W}_{d,d})_{ij} := \mathbf{q}_i^\top \mathbf{q}_j, \quad (\mathbf{W}_{x,d})_{p,i} := \mathbf{g}_p^\top \mathbf{q}_i, \quad (\mathbf{W}_{y,d})_{q,i} := \mathbf{h}_q^\top \mathbf{q}_i.$$

By construction, the full lifted matrix in (EC.10) is the Gram matrix of $\{\mathbf{g}_0, \mathbf{g}_p, \mathbf{h}_q, \mathbf{q}_i\}$, hence is PSD. Moreover, $\text{tr}(\mathbf{W}_{d,d}) = \sum_{i \in [n]} \|\mathbf{q}_i\|^2 = \mathbf{e}^\top \mathbf{t}$ and $\langle \mathbf{P}, \mathbf{W}_{x,d} \rangle = \sum_{i \in [n]} (\mathbf{W}_{x,d})_{\ell(i,i),i} = \sum_{i \in [n]} \mathbf{g}_{\ell(i,i)}^\top \mathbf{q}_i = \mathbf{e}^\top \mathbf{v}$. Therefore, the constructed solution is feasible for (EC.10) and has the same objective value as (EC.11). \square

EC.3. Additional Low-Rank Matrix Completion Results Supporting Section 4

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). Figures 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.

Figure EC.4 reports similar results to Figure 2, except with $\rho = 0.2$ rather than $\rho = 0.3$.

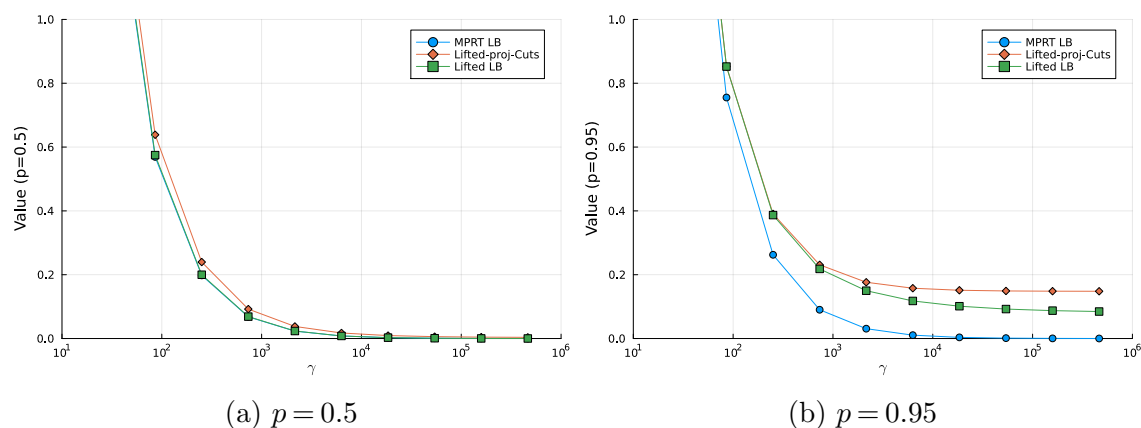


Figure EC.1 Absolute lower bounds as we vary γ for (i) a matrix perspective relaxation (“MPRT”), (ii) our lifted relaxation (“Lifted”), (iii) our lifted relaxation with projection cuts (“Lifted-proj-cuts”), for $p = 0.5$ in panel (a) and $p = 0.95$ in panel (b) for $n = 8$.

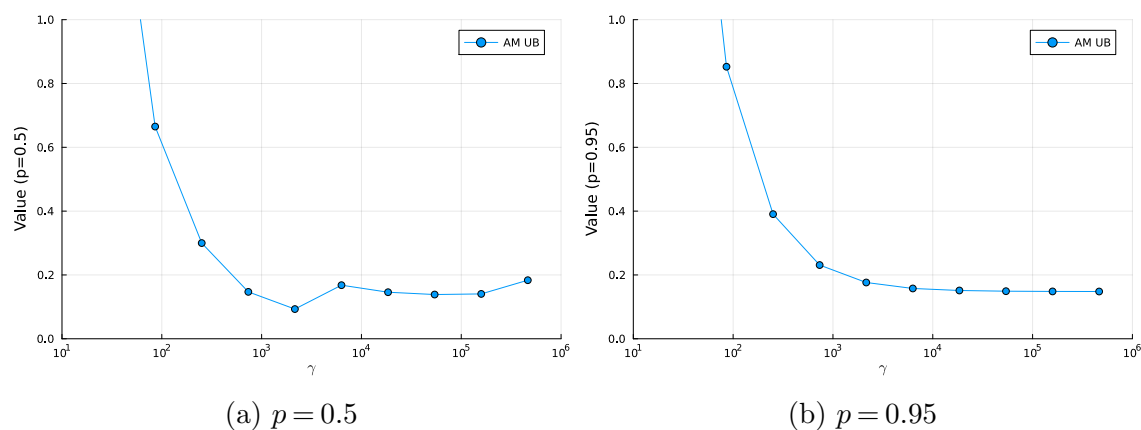


Figure EC.2 Absolute upper bounds as we vary γ for alternating minimization initialized at a rank- k SVD of $\mathcal{P}(\mathbf{A})$ for $p \in \{0.5, 0.95\}$ and $n = 8$.

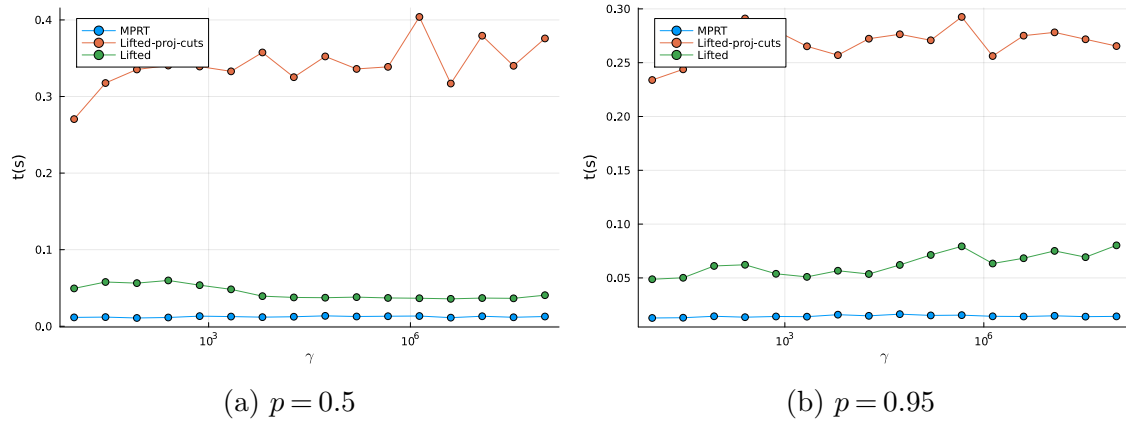


Figure EC.3 Runtimes as we vary γ for (i) a matrix perspective relaxation (“MPRT”), (ii) our lifted relaxation (“Lifted”), (iii) our lifted relaxation with projection cuts (“Lifted-proj-cuts”), for $p = 0.5$ in panel (a) and $p = 0.95$ in panel (b) for $n = 8$.

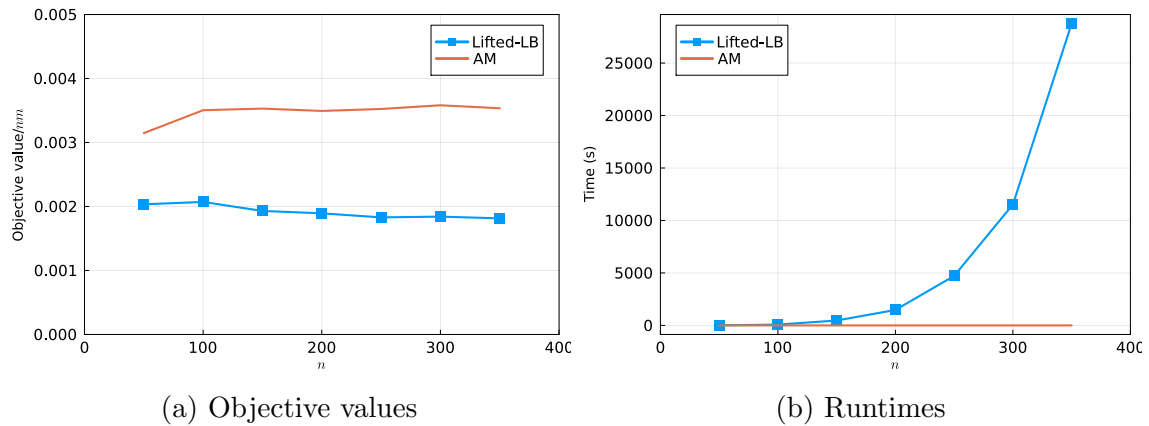


Figure EC.4 Objective value (left panel) and runtime for the compact lifted relaxation (right panel) as we vary n with $m = 50$ for our compact lifted relaxation (14) with $p = 0.4$, $\rho = 0.2$. Results are averaged over 10 replications.