Improved Approximation Algorithms for Low-Rank Problems Using Semidefinite Optimization

Ryan Cory-Wright
Department of Analytics, Marketing and Operations, Imperial College 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

Inspired by the impact of the Goemans-Williamson algorithm on combinatorial optimization, we construct an analogous relax-then-sample strategy for low-rank optimization problems. First, for orthogonally constrained quadratic optimization problems, we derive a semidefinite relaxation and a randomized rounding scheme, which obtains provably near-optimal solutions, mimicking the blueprint from Goemans and Williamson for the Max-Cut problem. We then extend our approach to generic low-rank optimization problems by developing new semidefinite relaxations that are both tighter and more broadly applicable than those in prior works. Although our original proposal introduces large semidefinite matrices as decision variables, we show that most of the blocks in these matrices can be safely omitted without altering the optimal value, hence improving the scalability of our approach. Using several examples (including matrix completion, basis pursuit, and reduced-rank regression), we show how to reduce the size of our relaxation even further. Finally, we numerically illustrate the effectiveness and scalability of our relaxation and our sampling scheme on orthogonally constrained quadratic optimization and matrix completion problems.

Key words: Low-rank; semidefinite relaxation; randomized rounding; approximation algorithm

1. Introduction

Many important optimization problems feature semi-orthogonal matrices, i.e., matrices $\boldsymbol{U} \in \mathbb{R}^{n\times m}$ such that $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_m$. Orthogonality constraints force the columns of $\boldsymbol{U}$ to be orthogonal and unit length, and are central to quadratic assignment (Gilman et al. 2022), quantum locality (Briët et al. 2011), control theory (Ben-Tal and Nemirovski 2002), and sparse PCA (Cory-Wright and Pauphilet 2022) problems. The set of semi-orthogonal matrices is often called the Stiefel manifold (Burer and Park 2023, Gilman et al. 2022). Orthogonality constraints are also related to the rank of a matrix, which models a matrix's complexity in data imputation (Bell and Koren 2007), factor analysis (Bertsimas et al. 2017), and multi-task regression (Negahban and Wainwright 2011) settings.

For any semi-orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{n\times m} : \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_m$, the matrix $\boldsymbol{Y} := \boldsymbol{U}\boldsymbol{U}^\top$ is an orthogonal projection matrix of rank m, i.e., it satisfies $\boldsymbol{Y}^2 = \boldsymbol{Y}$. Moreover, for any semi-orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{n\times n}$, the matrix $\boldsymbol{Y} = \frac{1}{2}\boldsymbol{U} + \frac{1}{2}\boldsymbol{I}_n$ is a projection matrix. Building on the algebraic similarities between binary variables and projection matrices (which solve the polynomial equations

$z^2 = z$ and $\boldsymbol{Y}^2 = \boldsymbol{Y}$), efficient approaches for mixed-integer optimization have been extended to rank-constrained optimization problems, including outer approximation (Bertsimas et al. 2022), perspective relaxations (Bertsimas et al. 2023c), and branch-and-bound (Bertsimas et al. 2023b).

In mixed-integer optimization, a major advance in the design of approximation algorithms occurred with the relax-and-round algorithm of Goemans and Williamson (1995). The objective of this paper is to construct an analogous approximation algorithm that runs in polynomial time and gives high-quality approximation guarantees for semi-orthogonal quadratic and low-rank optimization problems. As low-rank optimization is strongly NP-hard (Gillis and Glineur 2011), indeed $\exists\mathbb{R}$-hard, our theoretical approximation guarantees will only hold for the semi-orthogonal case, but our algorithms are applicable in both cases.

The algorithm of Goemans and Williamson (1995) provides a constant factor approximation guarantee of 0.87856 for Max-Cut problems. Their algorithm is significant because, before 1995, no polynomial-time algorithm achieved an approximation ratio better than 1/2, and since 1995, no polynomial-time algorithm with a better worst-case approximation guarantee has been found. The theoretical and computational success of Goemans and Williamson (1995)'s algorithm has implications far beyond Max-Cut. Their algorithm provides a $2/\pi$-approximation for general binary quadratic optimization (BQO) problems (Nesterov 1998), and can be extended to linearly-constrained BQO problems (Bertsimas and Ye 1998). More recently, Dong et al. (2015) developed a sampling scheme *à la* Goemans and Williamson for a broad class of mixed-integer optimization problems with logical constraints. Conceptually, the Goemans-Williamson algorithm propelled semidefinite optimization and correlated rounding at the core of approximation algorithms for combinatorial optimization (see Wolkowicz et al. 1998, Williamson and Shmoys 2011).

In this paper, we extend the Goemans-Williamson algorithm to quadratic semi-orthogonal and rank-constrained optimization problems by leveraging the deep connection between binary and low-rank optimization (Bertsimas et al. 2022), thus enriching the toolbox of semidefinite relaxations and approximation algorithms for low-rank optimization. Our main theoretical contributions are a guarantee on the expected performance of our Goemans-Williamson type algorithm for a class of quadratic semi-orthogonal optimization problems (Theorem 1 in Section 2.4), as well as new semidefinite relaxations for generic low-rank optimization problems (Proposition 4 and Theorem 2 in Section 4.1). Before presenting our contributions, we briefly describe the Goemans-Williamson algorithm for BQO and review the literature on approximation algorithms for some orthogonally constrained quadratic optimization problems.

1.1. Binary Quadratic Optimization and the Goemans-Williamson Algorithm

Binary quadratic optimization (BQO) is a canonical optimization problem with numerous applications throughout machine learning, statistics, and quantum computing (see Luo et al. 2010, for a review). As we discuss in detail in Section 3.1, it also drives logically constrained optimization problems with quadratic objectives.

Formally, given a matrix $\boldsymbol{Q} \succeq \boldsymbol{0}$, BQO selects a vector $\boldsymbol{z}$ in $\{-1,1\}^n$ that solves

$$\max_{\boldsymbol{z}\in\{-1,1\}^n} \quad \langle \boldsymbol{Q}, \boldsymbol{z}\boldsymbol{z}^\top \rangle. \tag{1}$$

Problem (1) is NP-hard and often challenging to solve to certifiable optimality when $n \geq 100$s (Rehfeldt et al. 2023). Accordingly, a popular approach for obtaining near-optimal solutions is to sample from a distribution parameterized by the solution of (1)'s convex relaxation. Namely, introduce a rank-one matrix $\boldsymbol{Z}$ to model the product $\boldsymbol{z}\boldsymbol{z}^\top$. Then, (1) is equivalent to

$$\max_{\boldsymbol{Z}\in\mathcal{S}^n_+} \quad \langle \boldsymbol{Q}, \boldsymbol{Z} \rangle \text{ s.t. } \operatorname{diag}(\boldsymbol{Z}) = \boldsymbol{e}, \operatorname{rank}(\boldsymbol{Z}) = 1.$$

We obtain a valid semidefinite relaxation of (1) by relaxing the rank constraint, as in Shor (1987):

$$\max_{\boldsymbol{Z}\in\mathcal{S}^n_+} \quad \langle \boldsymbol{Q}, \boldsymbol{Z} \rangle \text{ s.t. } \operatorname{diag}(\boldsymbol{Z}) = \boldsymbol{e}. \tag{2}$$

Probabilistically speaking, (2) is a device for constructing a pseudodistribution over $\boldsymbol{z} \in \{-1,1\}^n$, which aims to match the first two moments of the distribution of optimal solutions to the original binary quadratic problem (d'Aspremont and Boyd 2003, Barak et al. 2014). This suggests a procedure for converting an optimal solution to the relaxation to a near-optimal feasible solution: sample from a distribution parameterized by the relaxed solution and round to restore feasibility, as proposed by Goemans and Williamson (1995) for Max-Cut and described in Algorithm 1.

Algorithm 1 The Goemans-Williamson rounding algorithm for Problem (1)

Require: Positive semidefinite matrix $\boldsymbol{Q} \in \mathcal{S}^n_+$

Compute $\boldsymbol{Z}^\star$ solution of (2)

Sample $\hat{\boldsymbol{z}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Z}^\star)$

Construct $\bar{\boldsymbol{z}} \in \{-1,1\}^n : \bar{z}_i := \operatorname{sign}(\hat{z}_i)$

return $\bar{\boldsymbol{z}}$ solution to Problem (1)

The overall idea of Algorithm 1 is that the random solution $\hat{\boldsymbol{z}}$ is feasible in expectation, i.e., $\mathbb{E}[\operatorname{diag}(\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}^\top)] = \boldsymbol{e}$, and the random solution obtains an expected objective value equal to the optimal value of the semidefinite relaxation, i.e., $\mathbb{E}\langle \boldsymbol{Q}, \hat{\boldsymbol{z}}\hat{\boldsymbol{z}}^\top \rangle = \langle \boldsymbol{Q}, \boldsymbol{Z}^\star \rangle$. Thus, we obtain a high-quality feasible solution in expectation by rounding $\hat{\boldsymbol{z}}$ to restore feasibility. We have the following guarantee for the performance of Algorithm 1 (see Nesterov 1998, Bertsimas and Ye 1998)

LEMMA 1. *: Let $\boldsymbol{Q} \succeq \boldsymbol{0}$. Then, taking expectations over solutions to Algorithm 1 we have:*

$$\langle \boldsymbol{Q}, \boldsymbol{Z}^\star \rangle \geq \left\langle \boldsymbol{Q}, \mathbb{E}[\bar{\boldsymbol{z}}\bar{\boldsymbol{z}}^\top] \right\rangle \geq \frac{2}{\pi} \langle \boldsymbol{Q}, \boldsymbol{Z}^\star \rangle .$$

Lemma 1 implies that Goemans-Williamson rounding provides a constant factor guarantee for BQO in expectation. Moreover, if we further assume that $\boldsymbol{Q}$ is the Laplacian matrix of a graph, i.e., $\boldsymbol{Q}$ is a diagonally dominant matrix, then the very same approach yields a $\frac{2}{\pi}\min_{0\leq\theta\leq\pi}\left(\frac{\theta}{1-\cos\theta}\right) = 0.8786$-approximation (Goemans and Williamson 1995).

We remark that Problem (1) can easily be rewritten as a binary quadratic problems over $z_i \in \{0,1\}$ by setting $z_{i,\text{new}} = (z_i+1)/2$. In this case, the randomized rounding step in Algorithm 1 involves sampling from $\mathcal{N}(\boldsymbol{z}^\star, \boldsymbol{Z}^\star - \boldsymbol{z}^\star\boldsymbol{z}^{\star\top})$, where we impose the constraint $\boldsymbol{Z} \succeq \boldsymbol{z}\boldsymbol{z}^\top$ in our relaxation for a vector $\boldsymbol{z} \in [0,1]^n$, and the rounding step involves rounding to $\{0,1\}^n$ (cf. Bertsimas and Ye 1998, d'Aspremont and Boyd 2003). However, it is not possible to obtain purely multiplicative approximation guarantees when optimizing quadratic matrices over $\{0,1\}^n$, because the process of converting from $\{0,1\}^n$ to $\{\pm1\}^n$ involves adding/removing an additive term from the objective.

1.2. Orthogonally Constrained Quadratic Optimization

In this work, we generalize Algorithm 1 to address orthogonally and rank-constrained optimization problems. To achieve this, we consider a general family of orthogonally constrained quadratic problems that subsumes binary quadratic optimization. Formally, given a positive semidefinite matrix $\boldsymbol{A} \in \mathcal{S}^{nm}_+$, we optimize over semi-orthogonal matrices $\boldsymbol{U} \in \mathbb{R}^{n\times m}$ the problem:

$$\max_{\boldsymbol{U}\in\mathbb{R}^{n\times m}} \quad \langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{U})\mathrm{vec}(\boldsymbol{U})^\top\rangle \quad \text{s.t.} \quad \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_m, \tag{3}$$

where the vec(·) operator stacks the column of $\boldsymbol{U}$ together into a single vector, and we require that $n \geq m$ so that the problem is feasible. The similarities between Problems (3) and (1) are striking: by letting $n = m$ and $\boldsymbol{U}$ be a diagonal matrix in Problem (3), we recover Problem (1). Actually, Problem (3) is a strict generalization of Problem (1), because the set of semi-orthogonal matrices is not mixed-integer convex representable (Lubin et al. 2022, Corollary 4.1). From a worst-case complexity perspective, Problem (3) is NP-hard by reduction from Max-Cut, indeed $\exists\mathbb{R}$-hard by reduction from the problem of representing a graph with orthogonal vectors (Arends et al. 2011).

As discussed in the introduction, Problem (3) arises in a wide variety of problem settings, including quantum non-locality and generalized trust regions, and has been extensively studied in the literature. For conciseness, we refer the reader to Gilman et al. (2022, Appendix D) for a more extensive review. In addition, as will become clear in Section 4, Problem (3) appears as a relevant substructure for rank-constrained quadratic optimization problems of the form

$$\min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}} \quad \lambda\cdot\mathrm{rank}(\boldsymbol{X}) + \left\langle \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top, \boldsymbol{H}\right\rangle + \langle \boldsymbol{D}, \boldsymbol{X}\rangle \quad \text{s.t.} \quad \mathrm{rank}(\boldsymbol{X}) \leq k, \tag{4}$$

such as matrix completion (Candès and Recht 2009) or reduced rank regression (Negahban and Wainwright 2011), in the same way as (1) is a natural substructure for logically constrained quadratic optimization (as we recall in Section 3).

To the best of our knowledge, there are currently no rounding mechanisms for solving (3) approximately at this level of generality; the closest is a greedy rounding mechanism proposed by Bertsimas et al. (2022), which applies only when $\boldsymbol{H}$ is decomposable into a non-negative diagonal matrix plus a positive semidefinite matrix. Accordingly, most of the literature focuses on exact approaches for special cases of Problem (3) or approximation algorithms for other classes of quadratically constrained optimization problems, which we review below.

Exact Approaches for Special Cases: Some instances of (3) can be solved in polynomial time. For instance, if $\boldsymbol{A}$ is a block diagonal matrix with identical on-diagonal blocks $\boldsymbol{\Sigma}$ (a condition trivially met if $m=1$), then Problem (3) is equivalent to the principal component analysis (PCA) problem, which is solvable in polynomial time. Indeed, an optimal solution corresponds to m leading eigenvectors of $\boldsymbol{\Sigma}$. PCA with $m=1$ is also sometimes called a trust-region problem. More generally, instances of (3) with permutation-invariant objectives can often be solved in polynomial time via the exactness of their semidefinite relaxations (Anstreicher and Wolkowicz 2000, Kim et al. 2022). Problem (3) can also be solved to provable optimality via global solvers such as `Gurobi` or `BARON`, or custom branch-and-bound schemes (Bertsimas et al. 2023b), although the scalability of these global solvers is limited by current technology.

Approximation Algorithms: To our knowledge, known approximation algorithms do not apply to (3) in its full generality, although they do apply to some variants. Briët et al. (2010) propose an approximation algorithm for the problem

$$\max_{\boldsymbol{U}\in\mathbb{R}^{n\times m}} \operatorname{tr}(\boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{U}) \text{ s.t. } \boldsymbol{u}_i^\top \boldsymbol{u}_i = 1 \ \forall i \in [n],$$

ultimately obtaining an approximation ratio of $2/\pi+\Theta(1/m)$. However, this work does not maintain orthogonality between the columns of $\boldsymbol{U}$ and thus is less general than the problem studied here. Building on this idea, a second line of prior work (Nemirovski 2007, Bandeira et al. 2016) proposes approximation algorithms for the problem

$$\max_{\boldsymbol{U}_i\in\mathbb{R}^{n\times m}, i=1,\dots,k} \sum_{i,i'\in[k]} \langle \boldsymbol{A}^{(i,i')}, \boldsymbol{U}_i^\top \boldsymbol{U}_{i'}\rangle \text{ s.t. } \boldsymbol{U}_i^\top \boldsymbol{U}_i = \boldsymbol{I}_m \ \forall i \in [k], \tag{5}$$

where $\boldsymbol{A}^{(i,i')} \in \mathbb{R}^{m\times m}$ and the large matrix $\boldsymbol{A} := [\boldsymbol{A}^{(i,i')}]_{i,i'} \in \mathcal{S}_+^{mk}$ is positive semidefinite. In particular, Nemirovski (2007) develops non-trivial approximation guarantees via Talagrand's inequality, while Bandeira et al. (2016) connect their projected solution with the singular value decomposition of a Gaussian random matrix with i.i.d real entries.

Unfortunately, Problem (5) is not equivalent to (3) and the proof techniques in the aforementioned works do not extend to our case. There are two key differences in the objective function of (5). Compared to (3), Problem (5) involves the *inner*-products between columns of *different* semi-orthogonal matrices $\boldsymbol{U}_i$, $\boldsymbol{U}_{i'}$ for $i \neq i'$, where the terms $\langle \boldsymbol{A}^{(i,i)}, \boldsymbol{U}_i^\top \boldsymbol{U}_i\rangle$ are constant because of the orthogonality constraints. On the other hand, the objective in (3) depends on *outer*-products between columns of the *same* matrix $\boldsymbol{U}$. In particular, we can restore feasibility for each matrix $\boldsymbol{U}_i$, $i=1,\ldots,k$ in (5) separately, while the columns of $\boldsymbol{U}$ in (3) need to be considered together.

1.3. Contributions and Structure

Our main contribution is the development of a Goemans-Williamson sampling algorithm for the class of semi-orthogonal problems (3) and its extension to rank-constrained optimization.

We begin by studying approximation algorithms for Problem (3) in Section 2. We derive a semidefinite relaxation and propose a sampling procedure, which strictly generalizes Algorithm 1 to semi-orthogonal quadratic optimization. We provide theoretical performance guarantees for our method in our first main result (Theorem 1).

To prepare the extension of our approach to a broader class of low-rank optimization problems, we review the extension of the traditional Goemans-Williamson algorithm to mixed-binary quadratic optimization with logical constraints originally proposed by Dong et al. (2015) in Section 3. This extension relies on the existence of compact semidefinite relaxations for these problems, known as Shor relaxations (named after Shor 1987). While most of the results in this section are not new, we present a new proof for the compact Shor relaxation (Proposition 3), which brings the advantage of being constructive and thus more easily generalizable to rank-constrained problems.

Following the steps in the mixed-binary optimization case and leveraging the connection between binary variables and projection matrices, we extend our approach to low-rank optimization problems in Section 4. To facilitate this extension, we first derive new Shor relaxations for low-rank optimization problems. Unlike prior works (Recht et al. 2010, Bertsimas et al. 2023c, Kim et al. 2022, Li and Xie 2024), our relaxations do not require a spectral or permutation-invariant term in the objective or constraints. Conscious that these relaxations involve a number of additional semidefinite variables that may be prohibitively large in practice, we show how to eliminate many of these variables in the relaxation without altering its optimal value (Theorem 2). Finally, we describe a sampling algorithm to generate high-quality solutions from this relaxation.

To illustrate our approach, we apply our Shor relaxation to three prominent low-rank optimization problems in Section 5. In particular, we show how to exploit further problem structure and eliminate more variables from our relaxations, making our new relaxation more scalable.

Finally, in Section 6, we numerically benchmark our convex relaxations and randomized rounding schemes on quadratic semi-orthogonal and low-rank matrix completion problems.

1.4. Notations

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

For a matrix $\boldsymbol{X} \in \mathbb{R}^{n\times m}$, we let $\boldsymbol{x}_i$ denote its ith column and $\boldsymbol{X}_{i,\cdot}$ denote a column vector containing its ith row. We let $\mathrm{vec}(\boldsymbol{X}) : \mathbb{R}^{n\times m} \to \mathbb{R}^{nm}$ denote the vectorization operator which maps matrices to vectors by stacking columns. For a square matrix $\boldsymbol{X}$, $\mathrm{diag}(\boldsymbol{X})$ compiles the diagonal entries of $\boldsymbol{X}$ into a vector, while $\mathrm{Diag}(\boldsymbol{x})$ is a square matrix with diagonal equal to $\boldsymbol{x}$. For a matrix $\boldsymbol{W}$, we may find it convenient to describe it as a block matrix composed of equally sized blocks and denote the (i,i') block by $\boldsymbol{W}^{(i,i')}$. The dimension of each block will be clear from the context, given the size of the matrix $\boldsymbol{W}$ and the number of blocks. For example, the objective in (3) can be written

$$\langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{U})\mathrm{vec}(\boldsymbol{U})^\top \rangle = \sum_{i,i' \in [m]} \langle \boldsymbol{A}^{(i,i')}, \boldsymbol{u}_i \boldsymbol{u}_{i'}^\top \rangle,$$

with $\boldsymbol{A}^{(i,i')} \in \mathbb{R}^{n\times n}$. In particular, $\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}$ with $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ denotes an $nm \times nm$ block-diagonal matrix whose m diagonal blocks are equal to $\boldsymbol{\Sigma}$ (see Gupta and Nagar 2018, Chapter 1.2, for an introduction to the Kronecker product $\otimes$). With this notation, $\mathrm{vec}(\boldsymbol{\Sigma}\boldsymbol{X}) = (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma})\,\mathrm{vec}(\boldsymbol{X})$.

We let $\boldsymbol{X}^\dagger$ be the pseudoinverse of $\boldsymbol{X}$, which occurs in the Schur complement lemma (Boyd et al. 1994, Eqn. 2.41). We let $\mathcal{Y}^k_n := \{\boldsymbol{Y} \in \mathcal{S}^n_+ : \boldsymbol{Y}^2 = \boldsymbol{Y}, \mathrm{tr}(\boldsymbol{Y}) \leq k\}$ denote the set of orthogonal projection matrices with rank at most k, whose convex hull is $\{\boldsymbol{P} : \boldsymbol{0} \preceq \boldsymbol{P} \preceq \boldsymbol{I}_n, \mathrm{tr}(\boldsymbol{P}) \leq k\}$ (Overton and Womersley 1992, Theorem 3). We have $\mathrm{rank}(\boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{Y})$ for any projection matrix $\boldsymbol{Y}$.

Finally, our sampling procedure invokes the multivariate Gaussian probability measure: we let $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ denote a centered multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}$; see Grimmett and Stirzaker (2020) for an overview of the Gaussian distribution and Gupta and Nagar (2018) for an overview of its matrix extensions.

2. A Goemans-Williamson Approach for Orthogonality Constraints

In this section, we propose a new Goemans-Williamson approach for semi-orthogonal quadratic optimization problems, mirroring the development of the Goemans-Williamson algorithm for BQO in Section 1.1. First, in Section 2.1, we review a semidefinite relaxation for semi-orthogonal quadratic optimization originally developed by Burer and Park (2023). Then, we propose a randomized rounding scheme to generate high-quality solutions in Section 2.2, which is a strict generalization of Goemans-Williamson rounding as explained in Section 2.3. We derive performance guarantees for the rounding mechanism in Section 2.4, and provide technical results in Sections 2.5–2.6.

The rest of the paper extends the Goemans-Williamson rounding scheme developed in this section from low-rank orthogonal to low-rank quadratic optimization.

2.1. A Shor Relaxation

We study quadratic optimization over orthogonality constraints as described in Problem (3). As reviewed in §1.2 and derived by Burer and Park (2023), Problem (3) admits the relaxation:

$$\max_{\boldsymbol{W}\in\mathcal{S}_+^{mn}} \quad \langle \boldsymbol{A}, \boldsymbol{W}\rangle \quad \text{s.t.} \quad \operatorname{tr}\left(\boldsymbol{W}^{(j,j')}\right) = \delta_{j,j'} \ \forall j,j'\in[m], \ \sum_{i\in[m]} \boldsymbol{W}^{(i,i)} \preceq \boldsymbol{I}_n, \tag{6}$$

where the matrix $\boldsymbol{W}$ encodes for the outer-product of $\operatorname{vec}(\boldsymbol{U})$ with itself, and the trace constraints on the blocks of $\boldsymbol{W}$ stem from the columns of $\boldsymbol{U}$ having unit norm and being pairwise orthogonal.

Similarly to semidefinite relaxation of (1), imposing the constraint that $\boldsymbol{W}$ is rank-one in (6) would result in an exact reformulation of (3). Accordingly, Problem (6)'s relaxation is tight whenever some optimal solution is rank-one. However, the optimal solutions to (6) are often all high-rank. Indeed, ignoring the semidefinite constraint $\sum_{i\in[m]} \boldsymbol{W}^{(i,i)} \preceq \boldsymbol{I}_n$, the Pataki-Barvinok bound only guarantees the existence of a rank-m solution (cf. Pataki 1998, Barvinok 2001), which suggests that (6) is unlikely to yield a rank-one matrix $\boldsymbol{W}$.

Therefore, an interesting question is how to generate a high-quality feasible solution to (3) given a matrix $\boldsymbol{W}^\star$ which solves (6) but is not rank-one. In the next section, we address this question by proposing a sampling scheme followed by a projection step.

2.2. A Sample-Then-Project Procedure

We propose a randomized rounding scheme to generate high-quality feasible solutions to (3) from an optimal solution to (6). First, we solve (6) and obtain a semidefinite matrix $\boldsymbol{W}^\star$. Second, using $\boldsymbol{W}^\star$, we sample an $n\times m$ matrix $\boldsymbol{G}$ such that $\operatorname{vec}(\boldsymbol{G})$ follows a normal distribution with mean $\boldsymbol{0}_{nm}$ and covariance matrix $\boldsymbol{W}^\star$. Third, from the matrix $\boldsymbol{G}$, we generate a feasible solution to (3) by projecting $\boldsymbol{G}$ onto the feasible space of (3) using the singular value decomposition. Specifically, we compute a singular value decomposition of $\boldsymbol{G}$, $\boldsymbol{G}=\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$, and define $\boldsymbol{Q}:=\boldsymbol{U}\boldsymbol{V}^\top$. We summarize our procedure in Algorithm 2. In practice, we run the sampling and rounding steps multiple times and return the best solution, since rounding is significantly cheaper than solving the relaxation.

We remark that the normal distribution in our second step differs from the most widely used 'matrix normal distribution' (see, e.g., Gupta and Nagar 2018, Chapter 2) and, to the best of our knowledge, has only been studied by Barratt (2018). In contrast with other definitions of matrix Gaussian distributions, the entries of $\boldsymbol{G}$ in our sampling are neither independent nor identically distributed. In our implementation of Algorithm 2, we can sample $\operatorname{vec}(\boldsymbol{G})\sim\mathcal{N}(\boldsymbol{0}_{nm},\boldsymbol{W}^\star)$ even when $\boldsymbol{W}^\star$ is rank-deficient via the following construction—which will also be relevant for the theoretical analysis in Sections 2.4–2.5. Denoting $r=\operatorname{rank}(\boldsymbol{W}^\star)$, we first construct a Cholesky decomposition of $\boldsymbol{W}$: $\boldsymbol{W}=\sum_{k\in[r]}\operatorname{vec}(\boldsymbol{B}_k)\operatorname{vec}(\boldsymbol{B}_k)^\top$ with $\boldsymbol{B}_k\in\mathbb{R}^{n\times m}$. Then, we sample $\operatorname{vec}(\boldsymbol{G})=\sum_{k\in[r]}\operatorname{vec}(\boldsymbol{B}_k)z_k$ with $\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{0}_r,\boldsymbol{I}_r)$. This procedure ensures that $\operatorname{vec}(\boldsymbol{G})\in\operatorname{span}(\boldsymbol{W}^\star)$ almost surely, and that if the semidefinite relaxation is tight then $\boldsymbol{G}$ is optimal almost surely.

Algorithm 2 A Goemans-Williamson Algorithm for Orthogonality Constrained Optimization

Require: Positive semidefinite matrix $\boldsymbol{A} \in \mathcal{S}_+^{nm}$

Compute $\boldsymbol{W}^\star$ solution of (6)

Sample $\boldsymbol{G}$ according to $\mathrm{vec}(\boldsymbol{G}) \sim \mathcal{N}(\boldsymbol{0}_{nm}, \boldsymbol{W}^\star)$

Construct $\boldsymbol{Q} \in \arg\min_{\boldsymbol{U} \in \mathbb{R}^{n\times m}:\ \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_m} \ \|\boldsymbol{U} - \boldsymbol{G}\|_F^2$

return Semi-orthogonal matrix $\boldsymbol{Q}$

Similar to the original algorithm of Goemans and Williamson (1995), the intuition behind Algorithm 2 is that the sampled matrix $\boldsymbol{G}$ achieves an average performance equal to the relaxation value ($\mathbb{E}[\langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{G})\mathrm{vec}(\boldsymbol{G})^\top\rangle] = \langle \boldsymbol{A}, \boldsymbol{W}^\star\rangle$) and is feasible on average ($\mathbb{E}[\boldsymbol{G}^\top \boldsymbol{G}] = \boldsymbol{I}_m$). Therefore, the objective value of the projected solution $\boldsymbol{Q}$ should not be too different from that of $\boldsymbol{G}$.

2.3. Connection with Goemans-Williamson Algorithm for Binary Quadratic Optimization

We now connect Algorithms 1 and 2.

Consider a fixed but arbitrary instance of Problem (1). We embed each variable $z_i \in \{\pm 1\}$ into an n-dimensional vector $\boldsymbol{u}_i = z_i \boldsymbol{e}_i$ where $\boldsymbol{e}_i$ is the i-th standard basis vector. By construction, the matrix $\boldsymbol{U} = [\boldsymbol{u}_1 \cdots \boldsymbol{u}_n]$ satisfies $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_n$. With these notations, (1) is equivalent to an orthogonally constrained optimization problem (3) with the matrix $\boldsymbol{A}$ defined by blocks as $\boldsymbol{A}^{(j,j')} = Q_{j,j'} \boldsymbol{e}_j \boldsymbol{e}_{j'}^\top$.

The sparsity pattern of $\boldsymbol{A}$ implies that we can impose a similar sparsity pattern on $\boldsymbol{W}$ in the relaxation (6), i.e., we can restrict our attention to matrices $\boldsymbol{W}$ of the form $\boldsymbol{W}^{(j,j')} = \omega_{j,j'} \boldsymbol{e}_j \boldsymbol{e}_{j'}^\top$ without loss of optimality. Hence, because $\mathrm{vec}(\boldsymbol{G}) \in \mathrm{span}(\boldsymbol{W}^\star)$, Algorithm 2 generates matrices $\boldsymbol{G}$ that are diagonal. In other words, we have $\boldsymbol{G} = \mathrm{Diag}(\boldsymbol{\gamma})$ with $\boldsymbol{\gamma} \sim (\boldsymbol{0}_n, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = (\omega^\star_{j,j'})_{j,j'}$.

By construction, the columns of $\boldsymbol{G}$ are orthogonal, so the projection step boils down to normalizing each diagonal entry: $\boldsymbol{Q} = \mathrm{Diag}(\mathrm{sign}(\boldsymbol{\gamma}))$. This is precisely Algorithm 1. Thus, Algorithm 2 inherits the performance guarantees of Algorithm 1 in the special case where Problem (3) reduces to binary quadratic optimization. We now derive a more general performance guarantee.

2.4. Performance Analysis

In this section, we derive an approximation guarantee for Algorithm 2 and thus for the semi-orthogonal quadratic optimization problem (3).

Our result is a α-multiplicative and m-additive high probability guarantee for Algorithm 2:

THEOREM 1. *Let $\boldsymbol{A} \in \mathcal{S}_+^{nm}$ be a semidefinite matrix. For any $\delta > 0$, the semi-orthogonal matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times m}$ generated by Algorithm 2 satisfies with probability $1 - \delta$*

$$\alpha \langle \boldsymbol{A}, \boldsymbol{W}^\star\rangle - m\|\boldsymbol{A}\|_F \le \left\langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top \right\rangle,$$

with $\alpha = 1/(\log(n+m) + \log(1/\delta))$.

Our guarantee is a probabilistic statement that directly connects the practical implementation of the algorithm with its performance. Indeed, in practice, we sample N solutions using Algorithm 2 and retain the best one. For any $\delta > 0$, the best solution found from this procedure satisfies the guarantee in Theorem 1 with probability $1-\delta^N$, which vanishes to 0 as $N \to \infty$.

The striking feature of our bound is that the multiplicative factor — $\mathcal{O}(1/\log(n+m))$— depends on the problem dimension logarithmically, as opposed to the linear baseline for uniform rounding we lay out at the end of this section. On the other hand, we should acknowledge the presence of an $O(m)$-additive error term. We believe this additive term could be a limitation of our proof technique, and discuss this point in more detail in the next section. We now provide some intuition for the proof, before formally proving our main result.

Intuition for the Proof of Theorem 1: By construction, we have $\langle \boldsymbol{A}, \boldsymbol{W}^\star\rangle = \langle \boldsymbol{A}, \operatorname{vec}(\boldsymbol{G})\operatorname{vec}(\boldsymbol{G})^\top\rangle$ in expectation. Therefore, to derive a meaningful performance guarantee, we need to control the distance between $\operatorname{vec}(\boldsymbol{Q})\operatorname{vec}(\boldsymbol{Q})^\top$ and $\operatorname{vec}(\boldsymbol{G})\operatorname{vec}(\boldsymbol{G})^\top$. Recall that we can compute $\boldsymbol{Q}$ as $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{V}^\top$, where $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$ is a singular value decomposition of $\boldsymbol{G}$.

Unfortunately, given that the entries of $\boldsymbol{G}$ are neither independent nor identically distributed, the distribution of $(\boldsymbol{U}, \boldsymbol{S}, \boldsymbol{V})$ is highly non-trivial. Indeed, in the i.i.d. case (i.e., $\boldsymbol{W}^\star = \frac{1}{n}\boldsymbol{I}_{nm}$), $\boldsymbol{U}$, $\boldsymbol{S}$, and $\boldsymbol{V}$ are independent, $\boldsymbol{U}$ and $\boldsymbol{V}$ are uniformly distributed over the set of semi-orthogonal matrices, and the distribution of $\boldsymbol{S}$ is well characterized (Bandeira et al. 2016, Section 3).

Our setting diverges from this special case. Our relaxation is tight when $\boldsymbol{W}^\star$ is rank-one, so we should expect the distribution of $\operatorname{vec}(\boldsymbol{G})$ to be far from isotropic. Therefore, we first bound the projection error $\|\operatorname{vec}(\boldsymbol{G})\operatorname{vec}(\boldsymbol{G})^\top - \operatorname{vec}(\boldsymbol{Q})\operatorname{vec}(\boldsymbol{Q})^\top\|_F$ as a function of the largest singular value and the Frobenius norm of $\boldsymbol{G}$ —as stated in the following lemma— and then resort to concentration inequalities on $\sigma_{\max}(\boldsymbol{G})$ —which we formally derive in Section 2.6.

Lemma 2. *Consider a matrix $\boldsymbol{G} \in \mathbb{R}^{n\times m}$ and its singular value decomposition $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$. Define $\boldsymbol{Q} := \boldsymbol{U}\boldsymbol{V}^\top$ a projection of $\boldsymbol{G}$ onto the set of semi-orthogonal matrices. For any $\alpha > 0$, we have*

$$\|\alpha \operatorname{vec}(\boldsymbol{G})\operatorname{vec}(\boldsymbol{G})^\top - \operatorname{vec}(\boldsymbol{Q})\operatorname{vec}(\boldsymbol{Q})^\top\|_F^2 \le \alpha^2\left(1 - \frac{2}{\alpha\,\sigma_{\max}(\boldsymbol{G})^2}\right)\|\boldsymbol{G}\|_F^4 + m^2. \tag{7}$$

Let us observe that, $\mathbb{E}[\|\boldsymbol{G}\|_F^4] \ge (\mathbb{E}[\|\boldsymbol{G}\|_F^2])^2 = m^2$, so the right-hand side of (7) comprises two terms that scale like m^2. The constant in front of the first term, however, is unsigned. The essence of the proof for our guarantee is scaling the parameter α to control the error term in a non-trivial way.

We postpone the proof of Lemma 2 to Section 2.5 and formally prove Theorem 1.

Proof of Theorem 1 First, by interpreting $\boldsymbol{G}$ as a Gaussian matrix series, we can derive sub-Gaussian tail bounds for $\sigma_{\max}(\boldsymbol{G})$ (proof deferred to Lemma 4 in Section 2.6):

$$\mathbb{P}\left(\sigma_{\max}(\boldsymbol{G}) \ge t\right) \le (n+m)e^{-t^2/2}, \quad \forall t > 0.$$

In other words, with probability $1-\delta$, we have

$$\sigma_{\max}(\boldsymbol{G})^2 \leq 2\log(n+m) + 2\log(1/\delta),$$

which implies, for any $\alpha > 0$,

$$1 - \frac{2}{\alpha\,\sigma_{\max}(\boldsymbol{G})^2} \leq 1 - \frac{1}{\alpha\,(\log(n+m)+\log(1/\delta))}.$$

Picking $\alpha < (\log(n+m)+\log(1/\delta))^{-1}$ ensures that the right-hand side is negative so that, from Lemma 2, we have

$$\|\alpha\,\mathrm{vec}(\boldsymbol{G})\,\mathrm{vec}(\boldsymbol{G})^\top - \mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top\|_F^2 < m^2.$$

By combining Cauchy-Schwarz, $\forall \boldsymbol{W} : \langle \boldsymbol{A}, \boldsymbol{W}\rangle \leq \|\boldsymbol{A}\|_F\,\|\boldsymbol{W}\|_F$, with the bound above, we have

$$\left\langle \boldsymbol{A}, \alpha\,\mathrm{vec}(\boldsymbol{G})\,\mathrm{vec}(\boldsymbol{G})^\top\right\rangle - \left\langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top\right\rangle < \|\boldsymbol{A}\|_F\, m. \qquad \square$$

To evaluate the performance of Algorithm 2, it is interesting to compare its performance to a naive baseline where we draw $\boldsymbol{Q}$ uniformly from the set of semi-orthogonal matrices. Note that this is analogous to generating i.i.d. Bernoulli vectors in binary quadratic optimization, which successfully achieves a 1/2 approximation ratio in the Max-Cut case:

PROPOSITION 1. *Let $\boldsymbol{Q} \in \mathbb{R}^{n\times m}$ be distributed uniformly over $\{\boldsymbol{U} \in \mathbb{R}^{n\times m} : \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_m\}$. We have*

$$\mathbb{E}[\langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{Q})\mathrm{vec}(\boldsymbol{Q})^\top\rangle] \leq \max_{\boldsymbol{U}\in\mathbb{R}^{n\times m}:\boldsymbol{U}^\top\boldsymbol{U}=\boldsymbol{I}_m} \langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{U})\mathrm{vec}(\boldsymbol{U})^\top\rangle \leq nm\,\mathbb{E}[\langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{Q})\mathrm{vec}(\boldsymbol{Q})^\top\rangle].$$

Proposition 1 implies that taking $\boldsymbol{Q}$ to be uniformly distributed gives a nm-factor approximation algorithm for Problem (3). This is a worse multiplicative term than the logarithmic approximation guarantee in our main result, but does not contain any additive term, which suggests that, in future work, it may also be possible to omit the additive term from our main result.

Proof of Proposition 1 By optimality, $\boldsymbol{Q}$ being feasible for (3),

$$\langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{Q})\mathrm{vec}(\boldsymbol{Q})^\top\rangle \leq \max_{\boldsymbol{U}\in\mathbb{R}^{n\times m}:\boldsymbol{U}^\top\boldsymbol{U}=\boldsymbol{I}_m} \langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{U})\mathrm{vec}(\boldsymbol{U})^\top\rangle,$$

which leads to the first inequality.

Furthermore,

$$\max_{\boldsymbol{U}\in\mathbb{R}^{n\times m}:\boldsymbol{U}^\top\boldsymbol{U}=\boldsymbol{I}_m} \langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{U})\mathrm{vec}(\boldsymbol{U})^\top\rangle \leq \max_{\boldsymbol{u}\in\mathbb{R}^{nm}:\|\boldsymbol{u}\|=m} \langle \boldsymbol{A}, \boldsymbol{u}\boldsymbol{u}^\top\rangle = m\lambda_{\max}(\boldsymbol{A}) \leq m\,\mathrm{tr}(\boldsymbol{A}).$$

To conclude, observe that since $\boldsymbol{Q}$ is distributed according to the Haar measure, we have $\mathbb{E}[\boldsymbol{q}_i\boldsymbol{q}_i^\top] = \frac{1}{n}\boldsymbol{I}_n$ and $\mathbb{E}[\boldsymbol{q}_i\boldsymbol{q}_j^\top] = \boldsymbol{0}$ for $i \neq j$ (cf. Meckes 2019). Therefore, we have $\mathbb{E}[\mathrm{vec}(\boldsymbol{Q})\mathrm{vec}(\boldsymbol{Q})^\top] = \frac{1}{n}\boldsymbol{I}_{nm}$ and $\mathbb{E}[\langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{Q})\mathrm{vec}(\boldsymbol{Q})^\top\rangle] = \frac{1}{n}\,\mathrm{tr}(\boldsymbol{A})$. $\square$

2.5. Proof of Lemma 2

We start this section by providing the proof of Lemma 2, to support a discussion on the limitations of our proof techniques and opportunities for future research.

Proof of Lemma 2 Using Kronecker products, we can write $\mathrm{vec}(\boldsymbol{G}) = \mathrm{vec}(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top) = (\boldsymbol{V} \otimes \boldsymbol{U})\,\mathrm{vec}(\boldsymbol{S})$ and $\mathrm{vec}(\boldsymbol{Q}) = (\boldsymbol{V} \otimes \boldsymbol{U})\,\mathrm{vec}(\boldsymbol{I}_m)$. The matrix $(\boldsymbol{V} \otimes \boldsymbol{U})$ being unitary, we have

$$\|\alpha\,\mathrm{vec}(\boldsymbol{G})\,\mathrm{vec}(\boldsymbol{G})^\top - \mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top\|_F^2 = \|\alpha\,\mathrm{vec}(\boldsymbol{S})\,\mathrm{vec}(\boldsymbol{S})^\top - \mathrm{vec}(\boldsymbol{I}_m)\,\mathrm{vec}(\boldsymbol{I}_m)^\top\|_F^2 = \sum_{i,j\in[m]} (\alpha S_{i,i}S_{j,j} - 1)^2.$$

Expanding the square yields

$$\sum_{i,j\in[m]} (\alpha S_{i,i}S_{j,j} - 1)^2 = \alpha^2 (\sum_{i\in[m]} S_{i,i}^2)^2 - 2\alpha(\sum_{i\in[m]} S_{i,i})^2 + m^2.$$

From Hölder's inequality, we have $\sum_{i\in[m]} S_{i,i}^2 \le \left(\max_i S_{i,i}\right) \sum_{i\in[m]} S_{i,i}$, so that

$$\sum_{i,j\in[m]} (\alpha S_{i,i}S_{j,j} - 1)^2 \le \alpha^2 (\sum_{i\in[m]} S_{i,i}^2)^2 - \frac{2\alpha}{\max_i S_{i,i}^2}(\sum_{i\in[m]} S_{i,i}^2)^2 + m^2.$$

Substituting

$$\sum_{i\in[m]} S_{i,i}^2 = \|\boldsymbol{G}\|_F^2, \quad \text{and} \quad \max_i S_{i,i}^2 = \lambda_{\max}(\boldsymbol{G}^\top\boldsymbol{G}) = \sigma_{\max}(\boldsymbol{G})^2,$$

concludes the proof. □

Remark 1. From the proof, we see that our bound is tight whenever Hölder's inequality is tight, i.e., when the singular values $S_{i,i}$ are equal.

A key limitation in our performance guarantee (Theorem 1) is the presence of an additive error term scaling as m. In our proof, this factor comes from (7) and the fact that we bound the first term of the right-hand side of (7) by 0. A tighter upper bound, especially one that scales like m^2, would yield a smaller additive error term. To do so, we need a lower bound on $\|\boldsymbol{G}\|_F^4$ (with high probability) that scales like m^2. Although $\|\boldsymbol{G}\|_F$ concentrates around its mean, we can only say that $\mathbb{E}[\|\boldsymbol{G}\|_F^2] \ge m - 4\lambda_{\max}(\boldsymbol{W})$ (see Lemma EC.2) and $\lambda_{\max}(\boldsymbol{W})$ is not negligible compared to m (in the fortunate case where $\mathrm{rank}(\boldsymbol{W}^\star) = 1$, we even have $\lambda_{\max}(\boldsymbol{W}) = m$). So, this direction cannot lead to significant improvement in reducing the additive error term in Theorem 1, at least without modifying the relax-and-sampling strategy to control fourth moments (Barak et al. 2014, Deshpande and Montanari 2015), which would impact the tractability of our procedure.

Alternatively, before applying Hölder's inequality in the proof of Lemma 2, we have

$$\|\alpha\,\mathrm{vec}(\boldsymbol{G})\,\mathrm{vec}(\boldsymbol{G})^\top - \mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top\|_F^2 = \alpha^2\|\boldsymbol{G}\|^4 - 2\alpha\|\boldsymbol{G}\|_*^2 + m^2,$$

where $\|\boldsymbol{G}\|_*$ denotes the nuclear norm of $\boldsymbol{G}$. To our knowledge, no concentration bound exists that bounds $\|\boldsymbol{G}\|_*^2$ from below. For this reason, we used the fact that $\|\boldsymbol{G}\|_* \ge \|\boldsymbol{G}\|_F^2/\sigma_{\max}(\boldsymbol{G})$ and apply results on the Frobenius norm and the largest singular values, with the complication that our bound involves their ratio. Deriving concentration inequalities for the nuclear norm of the matrix Gaussian series constitutes an interesting future research direction that could benefit our analysis.

Finally, our proof of Theorem 1 applies Lemma 2 for a fixed value of α. Instead, one could consider parameter values that explicitly depend on $\boldsymbol{G}$, with the additional complication that $\mathbb{E}[\alpha \operatorname{vec}(\boldsymbol{G}) \operatorname{vec}(\boldsymbol{G})^\top] \neq \alpha \mathbb{E}[\operatorname{vec}(\boldsymbol{G}) \operatorname{vec}(\boldsymbol{G})^\top]$ in this case. To the best of our knowledge, $\alpha = \|\boldsymbol{G}\|_F^{-2}$ is the only choice amenable to analysis (see Lemma EC.3).

2.6. Bounding the Singular Values of Stochastic Matrices and Further Technical Discussion

As described in Section 2.2, in our implementation of Algorithm 2, we sample $\operatorname{vec}(\boldsymbol{G}) \sim \mathcal{N}(\boldsymbol{0}_{nm}, \boldsymbol{W}^\star)$ as $\operatorname{vec}(\boldsymbol{G}) = \sum_{k\in[r]} \operatorname{vec}(\boldsymbol{B}_k) z_k$ with $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}_r, \boldsymbol{I}_r)$ and $\boldsymbol{W}^\star = \sum_{k\in[r]} \operatorname{vec}(\boldsymbol{B}_k)\operatorname{vec}(\boldsymbol{B}_k)^\top$ a Cholesky decomposition of $\boldsymbol{W}^\star$ This construction interprets $\boldsymbol{G}$ as a matrix series, $\boldsymbol{G} = \sum_{k\in[r]} \boldsymbol{B}_k z_k$, as studied in the statistics literature (see, e.g., Tropp 2015). A key parameter in concentration inequalities for such matrix series is the so-called variance parameter, defined as

$$\sigma^2 = \max\left\{\lambda_{\max}\left(\sum_k \boldsymbol{B}_k^\top \boldsymbol{B}_k\right), \lambda_{\max}\left(\sum_k \boldsymbol{B}_k \boldsymbol{B}_k^\top\right)\right\}.$$

Because the matrix $\boldsymbol{W}^\star$ satisfies the constraints in (6), we have $\sigma^2 \leq 1$, as a direct consequence of the following lemma (proof in Appendix EC.3.2):

Lemma 3. *Let $\boldsymbol{W}$ be a feasible solution of (6) and consider a Cholesky decomposition of $\boldsymbol{W}$, $\boldsymbol{W} = \sum_{k\in[r]} \operatorname{vec}(\boldsymbol{B}_k)\operatorname{vec}(\boldsymbol{B}_k)^\top$ with $r = \operatorname{rank}(\boldsymbol{W})$ and $\boldsymbol{B}_k \in \mathbb{R}^{n\times m}$. Then,*

$$\sum_k \boldsymbol{B}_k^\top \boldsymbol{B}_k = \boldsymbol{I}_m, \quad \text{and} \quad \sum_k \boldsymbol{B}_k \boldsymbol{B}_k^\top \preceq \boldsymbol{I}_n.$$

As a result, Tropp (2012, Corollary 4.2) applies:

Lemma 4. *Consider a random matrix $\boldsymbol{G}$ sampled according to $\operatorname{vec}(\boldsymbol{G}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{W})$, where the matrix $\boldsymbol{W}$ is a feasible solution to (6). Then, the largest singular value of $\boldsymbol{G}$ satisfies:*

$$\mathbb{P}\left(\sigma_{\max}(\boldsymbol{G}) \geq t\right) \leq (n+m)e^{-t^2/2}, \quad \forall t > 0.$$

Among others, such tail bounds are useful in providing asymptotically tight control on the expectation of random variables. Formally, for $\boldsymbol{G}$, we have:

Corollary 1. *Consider a random matrix $\boldsymbol{G}$ sampled according to $\operatorname{vec}(\boldsymbol{G}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{W})$, where the matrix $\boldsymbol{W}$ is a feasible solution to (6). Then, $\boldsymbol{G}$ satisfies the following*

$$\mathbb{E}\left[\|\boldsymbol{G}\|_F^2\right] = m, \quad \text{and} \quad \mathbb{E}[\sigma_{\max}(\boldsymbol{G})^2] \leq 2\log(n+m) + 2.$$

Proof of Corollary 1 For the Frobenius norm, we have $\mathbb{E}\left[\|\boldsymbol{G}\|_F^2\right] = \operatorname{tr}(\mathbb{E}\left[\boldsymbol{G}^\top \boldsymbol{G}\right]) = \operatorname{tr}\left(\sum_k \boldsymbol{B}_k^\top \boldsymbol{B}_k\right) = m$, driven by the m constraints $\operatorname{tr}(\boldsymbol{W}^{(i,i)}) = 1$ in our semidefinite relaxation.

For the largest singular value, this is a consequence of sub-Gaussian tail bounds (Lemma 4), which we derive for the sake of consistency in Lemma EC.1. Note that this result is the rectangular analog of the result obtained in Tropp (2012, section 4.4) for random symmetric matrices. □

Corollary 1 shows that the largest singular value of $\boldsymbol{G}$ does not explode with n and m, but rather remains bounded (up to logarithmic terms), while $\|\boldsymbol{G}\|_F^2$ scales as m in expectation. Together, these results suggest that the singular values of $\boldsymbol{G}$ are of comparable size, since $\|\boldsymbol{G}\|_F^2 = \sum_{i\in[m]} \sigma_i(\boldsymbol{G})^2$ and thus all singular values must contribute to $\|\boldsymbol{G}\|_F^2$ to achieve the correct scaling. This observation is important for quantifying the distance between $\mathrm{vec}(\boldsymbol{G})\,\mathrm{vec}(\boldsymbol{G})^\top$ and $\mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top$.

3. Goemans-Williamson and Logically Constrained Optimization

Given our objective to generalize our approximation algorithm for semi-orthogonal quadratic optimization to low-rank optimization, we review, in this section, how the classical Goemans-Williamson algorithm for BQO has been generalized in mixed-integer optimization problems. Precisely, we review a semidefinite relaxation and randomized rounding scheme for logically constrained problems, which prepares the ground for the extension of the semidefinite relaxation and randomized rounding scheme from Section 2 to rank-constrained optimization in Section 4.

3.1. A Shor Relaxation and Its Compact Version

We consider a quadratic optimization problem that unfolds over two stages, as occurs in sparse regression, portfolio selection, and network design problems; see Bertsimas et al. (2021) for a review. In the first stage, a decision-maker activates binary variables subject to resource budget constraints and activation costs. Subsequently, in the second stage, the decision-maker optimizes over the continuous variables. Formally, we consider the problem

$$\min_{\boldsymbol{z}\in\mathcal{Z}_n^k} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad \boldsymbol{c}^\top \boldsymbol{z} + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{x} \text{ s.t. } \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b},\ x_i = z_i x_i \ \forall i \in [n], \tag{8}$$

where $\mathcal{Z}_n^k := \{\boldsymbol{z} \in \{0,1\}^n : \boldsymbol{e}^\top \boldsymbol{z} \le k\}$, $\boldsymbol{Q} \succeq \boldsymbol{0}$ is a positive definite matrix, and $\boldsymbol{c} \in \mathbb{R}^n_+$. Note that the bilinear constraints $x_i = z_i x_i$ for $z_i \in \{0,1\}$ enforces the logical relationships '$x_i = 0$ if $z_i = 0$'.

Problem (8) has a convex quadratic objective function. Therefore, a viable technique for obtaining a strong convex relaxation is introducing semidefinite matrices to model products of variables. This technique was first proposed by Shor (1987) in the context of non-convex quadratic optimization and has since been studied by many other authors; see Han et al. (2022) for a review. In particular, we introduce the block matrix $\boldsymbol{W} \in \mathcal{S}^{2n}_+$ to represent the outer product of $\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{z}\end{pmatrix}$ with itself. Specifically, we partition $\boldsymbol{W}$ into four blocks: $\boldsymbol{W}_{x,x}$, $\boldsymbol{W}_{z,z}$, $\boldsymbol{W}_{x,z}$, and $\boldsymbol{W}_{x,z}^\top$, which model $\boldsymbol{x}\boldsymbol{x}^\top$, $\boldsymbol{z}\boldsymbol{z}^\top$, $\boldsymbol{x}\boldsymbol{z}^\top$, and $\boldsymbol{z}\boldsymbol{x}^\top$, respectively. With these additional variables, we have the following semidefinite relaxation for Problem (8):

Proposition 2. *The optimization problem*

$$\min_{\boldsymbol{z}\in[0,1]^n:\,\boldsymbol{e}^\top\boldsymbol{z}\le k}\ \min_{\substack{\boldsymbol{x}\in\mathbb{R}^n:\,\boldsymbol{A}\boldsymbol{x}\le\boldsymbol{b}\\ \boldsymbol{W}\in\mathcal{S}^{2n}_+}}\ \boldsymbol{c}^\top\boldsymbol{z}+\frac{1}{2}\langle\boldsymbol{Q},\boldsymbol{W}_{x,x}\rangle+\boldsymbol{d}^\top\boldsymbol{x}$$
$$\text{s.t.}\quad \boldsymbol{W}\succeq\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{z}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{z}\end{pmatrix}^\top,\ \mathrm{diag}(\boldsymbol{W}_{z,z})=\boldsymbol{z},\ \mathrm{diag}(\boldsymbol{W}_{x,z})=\boldsymbol{x}, \tag{9}$$

is a valid convex relaxation of Problem (8)*.*

Proof of Proposition 2 It suffices to show that any feasible solution to (8) corresponds to a feasible solution in (9) with the same objective value. To see this, fix $\boldsymbol{z},\boldsymbol{x}$ in (8), and set

$$\boldsymbol{W}:=\begin{pmatrix}\boldsymbol{W}_{x,x} & \boldsymbol{W}_{x,z}\\ \boldsymbol{W}_{x,z}^\top & \boldsymbol{W}_{z,z}\end{pmatrix}=\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{z}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{z}\end{pmatrix}^\top.$$

Furthermore, $(\boldsymbol{W}_{z,z})_{i,i}=z_i^2=z_i$ because z_i is binary, and $(\boldsymbol{W}_{x,z})_{i,i}=x_iz_i=x_i$. Hence, the solution $(\boldsymbol{z},\boldsymbol{x},\boldsymbol{W})$ is feasible in (9) and attains the same objective value. □

Remark 2. Problem (9) is a relaxation of Problem (8) by allowing $\boldsymbol{z}\in[0,1]^n$ and by omitting the rank-1 constraint on $\boldsymbol{W}$. Reimposing the rank-one constraint obtains an equivalent reformulation of Problem (8).

Remark 3. We can strengthen Problem (9) by applying the Reformulation-Linearization Technique (RLT; e.g., Bao et al. 2011) to the linear constraints on $\boldsymbol{x}$, $\boldsymbol{A}\boldsymbol{x}\le\boldsymbol{b}$, leading to $\boldsymbol{A}\boldsymbol{W}_{x,x}\boldsymbol{A}^\top+\boldsymbol{b}\boldsymbol{b}^\top\ge\boldsymbol{b}\boldsymbol{x}^\top\boldsymbol{A}+\boldsymbol{A}^\top\boldsymbol{b}\boldsymbol{x}^\top$. All results follow identically with RLT constraints on $(\boldsymbol{x},\boldsymbol{W}_{x,x})$.

While Problem (9) is a valid convex relaxation, it may be expensive to solve, because it involves large semidefinite matrices. Surprisingly, Han et al. (2022) demonstrated that Problem (9) is equivalent to the so-called "optimal perspective relaxation" originally proposed by Zheng et al. (2014), Dong et al. (2015), which is much more compact. We now recall this compact relaxation and prove its equivalence. We acknowledge that this result has been proven previously in Han et al. (2022, Theorem 6), in a non-constructive way. Here, we develop a new, constructive proof for it, which we will be able to extend to rank-constrained optimization.

Proposition 3. *Problem* (9) *is equivalent to*

$$\min_{\boldsymbol{z}\in[0,1]^n:\,\boldsymbol{e}^\top\boldsymbol{z}\le k}\ \min_{\substack{\boldsymbol{x}\in\mathbb{R}^n:\,\boldsymbol{A}\boldsymbol{x}\le\boldsymbol{b}\\ \boldsymbol{X}\in\mathcal{S}^{n}}}\ \boldsymbol{c}^\top\boldsymbol{z}+\frac{1}{2}\langle\boldsymbol{Q},\boldsymbol{X}\rangle+\boldsymbol{d}^\top\boldsymbol{x}$$
$$\text{s.t.}\quad \boldsymbol{X}\succeq\boldsymbol{x}\boldsymbol{x}^\top,\ x_i^2\le X_{i,i}z_i,\forall i\in[n], \tag{10}$$

Proposition 3 shows that we can solve the semidefinite relaxation (9) by solving the much smaller semidefinite optimization problem (10), which only involves one semidefinite variable $\boldsymbol{X}\in\mathcal{S}^n_+$, and reconstruct an optimal solution involving $\boldsymbol{W}\in\mathcal{S}^{2n}_+$ to (9). Our proof of Proposition 3 makes this reconstruction step explicit.

Proof of Proposition 3 We show that any feasible solution to Problem (9) generates a feasible solution to (10) with an equal or lower objective and vice versa.

First, we consider a feasible solution to (9), $(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{W})$, and show that the solution $(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{X}) = (\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{W}_{x,x})$ is a feasible solution to (10), with the same objective value. To establish feasibility, we only need to verify that $x_i^2 \leq X_{i,i} z_i$, since the remaining constraints in (10) are present in (9). From the non-negativity of the 2×2 minors of the semidefinite matrix $\boldsymbol{W}$, we have $(\boldsymbol{W}_{x,x})_{i,i}(\boldsymbol{W}_{z,z})_{i,i} \geq (\boldsymbol{W}_{x,z})_{i,i}^2$. Substituting the identities $(\boldsymbol{W}_{x,z})_{i,i} = x_i$ and $(\boldsymbol{W}_{z,z})_{i,i} = z_i$ yields the result.

Next, consider a feasible solution $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{X})$ to (10). Observe that the constraint $x_i^2 \leq X_{i,i} z_i$ imposes $x_i = 0$ if $X_{i,i} = 0$. Since $\boldsymbol{c} \geq \boldsymbol{0}$, it follows that, if the constraint $z_i X_{i,i} \geq x_i^2$ is not binding for some index i, we can decrease z_i without impacting feasibility or worsening the objective value. Accordingly, we can assume $z_i = x_i^2 / X_{i,i}$ without loss of generality (with the convention $0/0 = 0$ so that $z_i = 0$ if $X_{i,i} = 0$). We now define a matrix $\boldsymbol{W}$ such that $(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{W})$ is feasible for (9) and achieves the same objective value. Observe that the matrix $\boldsymbol{M}$

$$\underbrace{\begin{pmatrix} 1 & \boldsymbol{x}^\top & \boldsymbol{z}^\top \\ \boldsymbol{x} & \boldsymbol{X} & \boldsymbol{W}_{x,z} \\ \boldsymbol{z} & \boldsymbol{W}_{x,z}^\top & \boldsymbol{W}_{z,z} \end{pmatrix}}_{\boldsymbol{M}} := \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_n \\ \boldsymbol{0} & \mathrm{Diag}(\boldsymbol{u}) \end{pmatrix} \begin{pmatrix} 1 & \boldsymbol{x}^\top \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_n \\ \boldsymbol{0} & \mathrm{Diag}(\boldsymbol{u}) \end{pmatrix}^\top$$

with $u_i = \frac{x_i}{X_{i,i}}$ if $X_{i,i} > 0$ and 0 otherwise, is positive semidefinite as a positive semidefinite matrix left and right multiplied by a matrix and the same matrix transposed. Hence, we consider the matrices $\boldsymbol{W}_{z,z}, \boldsymbol{W}_{x,z}$ as defined above. Moreover, we note that the vector $\boldsymbol{z}$ defined as a block of the matrix $\boldsymbol{M}$ is equal to our original $\boldsymbol{z}$. Indeed, $(\mathrm{Diag}(\boldsymbol{u})\boldsymbol{x})_i = (\boldsymbol{x} \circ \boldsymbol{u})_i = \frac{x_i^2}{X_{i,i}} = z_i$.

To complete the proof, we verify that $\boldsymbol{M}$ gives a feasible solution to (9). First, by the Schur complement lemma, we have

$$\boldsymbol{M} \succeq \boldsymbol{0} \text{ if and only if } \boldsymbol{W} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{W}_{x,z} \\ \boldsymbol{W}_{x,z}^\top & \boldsymbol{W}_{z,z} \end{pmatrix} \succeq \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix}^\top.$$

Second, by the definition of $\boldsymbol{W}_{z,z}$, we have

$$(\boldsymbol{W}_{z,z})_{ii} = X_{i,i} u_i^2 = \begin{cases} \frac{x_i^2}{X_{i,i}} & \text{if } X_{i,i} > 0 \\ 0 & \text{if } X_{i,i} = 0 \end{cases} = z_i,$$

because $x_i^2 / X_{i,i} = z_i$. Finally, by the definition of $\boldsymbol{W}_{x,z}$, we have

$$(\boldsymbol{W}_{x,z})_{ii} = X_{i,i} u_i = \begin{cases} x_i & \text{if } X_{i,i} > 0 \\ 0 & \text{if } X_{i,i} = 0 \end{cases} = x_i.$$

Therefore, $(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{W})$ is feasible in (10) and attains an equal objective value. □

We close this section by pointing out that Proposition 3 does *not* imply that Problem (9) cannot be useful in practice. Indeed, the variables $\boldsymbol{W}_{z,z}$ and $\boldsymbol{W}_{x,z}$ enable to express constraints that further tighten the relaxation. For example, one can tighten Problem (9)'s relaxation by imposing the so-called triangle inequalities on $(\boldsymbol{z}, \boldsymbol{W}_{z,z})$, as derived by Padberg (1989). As we demonstrate via a simple sparse linear regression example in Appendix EC.1, the equivalence demonstrated in Proposition 3 does not hold in the presence of these triangle inequalities.

3.2. Goemans-Williamson Rounding for Logically Constrained Optimization

The equivalence result in Proposition 3 reveals that it is possible to reconstruct an optimal $\boldsymbol{W}_{z,z}$ given an optimal solution to the semidefinite relaxation (10) that involves $\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{X} = \boldsymbol{W}_{x,x}$ only. This raises the following research question: how to use the reconstructed solution $\boldsymbol{W}_{z,z}$ as part of a rounding scheme for constructing a high-quality solution to (8). To answer this question, Dong et al. (2015) observe, in the context of sparse regression, that the variable $\boldsymbol{x}$ being fixed, the objective function in Problem (8) is quadratic in $\boldsymbol{z}$, given that $x_i = z_i x_i$. This observation suggests that the rounding mechanism of Goemans and Williamson (1995) is a good candidate for generating high-quality feasible solutions $\boldsymbol{z}$ to (8). In particular, rounding for a binary $\boldsymbol{z}$ using a Goemans-Williamson scheme, then solving for $\boldsymbol{x}$ with $\boldsymbol{z}$ being fixed to $\bar{\boldsymbol{z}}$. Accordingly, we now describe a Goemans-Williamson rounding to logically constrained quadratic optimization problems, in Algorithm 3 (see also Dong et al. (2015)).

Algorithm 3 A Goemans-Williamson Rounding Method for Logically Constrained Optimization

Compute solution $\boldsymbol{z}^\star, \boldsymbol{W}^\star_{z,z}$ either by solving (9), or solving(10) and reconstructing $\boldsymbol{W}^\star_{z,z}$.

Sample $\hat{\boldsymbol{z}} \sim \mathcal{M}(\boldsymbol{z}^\star, \boldsymbol{W}^\star_{z,z} - \boldsymbol{z}^\star \boldsymbol{z}^{\star\top})$

Construct $\bar{\boldsymbol{z}} \in \{0,1\}^n : \bar{z}_i = \text{Round}(\hat{z}_i)$.

Compute $\bar{\boldsymbol{x}}(\bar{\boldsymbol{z}})$, an optimal $\boldsymbol{x}$ given $\bar{\boldsymbol{z}}$ by solving

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \quad \frac{1}{2} \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{x} \text{ s.t. } \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \; x_i = 0 \text{ if } \bar{z}_i = 0, \forall i \in [n]$$

return $\bar{\boldsymbol{z}}$, $\bar{\boldsymbol{x}}(\bar{\boldsymbol{z}})$

We remark that $\hat{\boldsymbol{z}}$ is sampled according to a normal distribution with covariance matrix $\boldsymbol{W}^\star_{z,z} - \boldsymbol{z}^\star \boldsymbol{z}^{\star\top}$ in Algorithm 3 to ensure that $\mathbb{E}[\hat{\boldsymbol{z}} \hat{\boldsymbol{z}}^\top] = \boldsymbol{W}^\star_{z,z}$, and thus the random solution $\hat{\boldsymbol{z}}$ is feasible and has an objective value equal to the optimal value of the semidefinite relaxation in expectation.

Unfortunately, it is challenging to produce a constant-factor approximation guarantee for Algorithm 3, as discussed for the case of sparse linear regression by Dong et al. (2015). This is perhaps unsurprising, indeed, solving logically constrained quadratic optimization problems is strongly NP-hard (Chen et al. 2019). Nonetheless, the Goemans-Williamson algorithm is useful in practice even in settings where we cannot obtain constant-factor theoretical guarantees.

In the rest of the paper, we mirror this extension of Goemans-Williamson to logically constrained quadratic optimization to further generalize our Shor relaxation and Goemans Williamson sampling scheme for semi-orthogonal quadratic optimization to low-rank quadratic optimization problems.

4. New Relaxations and Sampling for Low-Rank Optimization Problems

In this section, we generalize our Goemans-Williamson algorithm for semi-orthogonal quadratic optimization (Algorithm 2 in Section 2) to generic rank-constrained optimization. Our overall approach is to mirror the generalization of the traditional Goemans-Williams algorithm to logically constrained quadratic optimization from Section 3.

We proceed in three steps: First, we derive new Shor relaxations for rank-constrained optimization problems (§4.1). Unlike prior work (Recht et al. 2010, Bertsimas et al. 2023c, Kim et al. 2022, Li and Xie 2024), our relaxations do not require the presence of a spectral or permutation-invariant term in the objective or constraints. Interestingly, we show that many of the variables in our Shor relaxations can be omitted without altering the objective value, leading to a more compact and tractable formulation. Compared with Bertsimas et al. (2023c), we show that our new relaxations are stronger and more broadly applicable. Second, we discuss how our common ideas in logically constrained optimization, such as RLT, can be generalized to our context and further strengthen our relaxation (§4.2). Finally, we describe our sampling algorithm for these problems in §4.3.

4.1. A New Shor Relaxation and Its Compact Formulation

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

$$\begin{aligned} \min_{\boldsymbol{Y}\in\mathcal{Y}_n^k}\ \min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}} \quad & \langle \boldsymbol{C},\boldsymbol{Y}\rangle + \left\langle \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top, \boldsymbol{H}\right\rangle + \langle \boldsymbol{D},\boldsymbol{X}\rangle \\ \text{s.t.} \quad & \langle \boldsymbol{A}_i,\boldsymbol{X}\rangle \le b_i \ \forall i\in[m],\ \boldsymbol{X}=\boldsymbol{Y}\boldsymbol{X}, \end{aligned} \tag{11}$$

where $\boldsymbol{H}\in\mathcal{S}_+^{nm}, \boldsymbol{C}\in\mathcal{S}_+^n$ are positive semidefinite matrices, $\boldsymbol{D}\in\mathbb{R}^{n\times m}$ is a rectangular matrix. As demonstrated in Bertsimas et al. (2022), any rank-constrained optimization problem of the form (4) can be formulated as an optimization over $(\boldsymbol{X},\boldsymbol{Y})$ of the form (11), where the additional decision variable $\boldsymbol{Y}$ is a projection matrix which encodes the span of $\boldsymbol{X}$ and whose trace bounds $\mathrm{rank}(\boldsymbol{X})$. Here, we write $\mathrm{vec}(\boldsymbol{X}^\top)$ rather than the mathematically equivalent $\mathrm{vec}(\boldsymbol{X})$ to simplify the notation in our relaxations. Since there exists a permutation matrix $\boldsymbol{K}_{n,m}\in\mathbb{R}^{nm\times nm}$ such that $\mathrm{vec}(\boldsymbol{A}^\top)=\boldsymbol{K}_{n,m}\,\mathrm{vec}(\boldsymbol{A})$ for any $\boldsymbol{A}\in\mathbb{R}^{n\times m}$ ($\boldsymbol{K}_{n,m}$ is also called a commutation matrix, see Magnus and Neudecker 1979), both formulations are equivalent.

Problem (11) is quite a general formulation. It models matrix completion objectives like $\sum_{(i,j)\in\Omega}(X_{i,j}-A_{i,j})^2$ (as we detail in Section 5.1) and optimal power flow terms like $X_{i,j}X_{k,l}$. As a result of this generality, it is also challenging to solve.

We now develop a convex relaxation of (11). We remark that previous works on developing low-rank relaxations like Bertsimas et al. (2023c), Kim et al. (2022) require a spectral or permutation invariant term in the objective to develop a valid convex relaxation, hence do not apply to (11). Thus, designing a computationally tractable convex relaxation for (11) is arguably an open problem. Following the Shor relaxation blueprint outlined for logically constrained MIO in Section 3.1, we introduce matrices $\boldsymbol{W}_{x,x}$, $\boldsymbol{W}_{x,y}$, $\boldsymbol{W}_{y,y}$ to model the outer products $\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$, $\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{Y})^\top$, and $\mathrm{vec}(\boldsymbol{Y})\mathrm{vec}(\boldsymbol{Y})^\top$ respectively.

Proposition 4. *The convex semidefinite optimization problem*

$$\min_{\substack{\boldsymbol{Y}\in\mathcal{S}^n_+:\boldsymbol{Y}\preceq\boldsymbol{I},\mathrm{tr}(\boldsymbol{Y})\le k\\ \boldsymbol{W}_{y,y}\in\mathcal{S}^{n^2}_+}} \min_{\substack{\boldsymbol{X}\in\mathbb{R}^{m\times n}:\langle\boldsymbol{A}_i,\boldsymbol{X}\rangle\le b_i, i\in[m]\\ \boldsymbol{W}_{x,x}\in\mathcal{S}^{nm}_+,\boldsymbol{W}_{x,y}\in\mathbb{R}^{nm\times n^2}}} \langle\boldsymbol{C},\boldsymbol{Y}\rangle+\langle\boldsymbol{W}_{x,x},\boldsymbol{H}\rangle+\langle\boldsymbol{D},\boldsymbol{X}\rangle$$
$$\text{s.t.} \quad \begin{pmatrix} 1 & \mathrm{vec}(\boldsymbol{X}^\top)^\top & \mathrm{vec}(\boldsymbol{Y})^\top\\ \mathrm{vec}(\boldsymbol{X}^\top) & \boldsymbol{W}_{x,x} & \boldsymbol{W}_{x,y}\\ \mathrm{vec}(\boldsymbol{Y}) & \boldsymbol{W}_{x,y}^\top & \boldsymbol{W}_{y,y}\end{pmatrix}\succeq\boldsymbol{0}, \tag{12}$$
$$\sum_{i=1}^n \boldsymbol{W}_{y,y}^{(i,i)}=\boldsymbol{Y},\ \sum_{i=1}^n \boldsymbol{W}_{x,y}^{(i,i)}=\boldsymbol{X}^\top$$

is a valid convex relaxation of Problem (11).

Remark 4. If an optimal solution to (12) is such that $\boldsymbol{W}_{x,x}$ is a rank-one matrix then $\boldsymbol{W}_{x,x}=\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$ and the optimal values of (12) and (11) coincide.

Proof of Proposition 4 Fix $(\boldsymbol{X},\boldsymbol{Y})$ in (11) and set

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

It is sufficient to verify that $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W}_{x,x},\boldsymbol{W}_{x,y},\boldsymbol{W}_{y,y})$ is feasible for (12)—it obviously attains the same objective value. First, by construction, the semidefinite constraint is satisfied (at equality). Moreover, we have

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

□

Unfortunately, (12) is not compact and involves $n^2\times n^2$ and $nm\times nm$ matrices. Therefore, motivated by Proposition 3, a natural research question is whether it is possible to eliminate any variables from (12) without altering its optimal objective value. We answer this question affirmatively, using a proof technique reminiscent of our constructive approach to Proposition 3:

THEOREM 2. *Problem* (12) *is equivalent to*

$$\min_{\boldsymbol{Y}\in\mathcal{S}^n_+:\,\boldsymbol{Y}\preceq\boldsymbol{I},\ \mathrm{tr}(\boldsymbol{Y})\leq k}\quad \min_{\substack{\boldsymbol{X}\in\mathbb{R}^{n\times m}:\langle\boldsymbol{A}_i,\boldsymbol{X}\rangle\leq b_i, i\in[m]\\ \boldsymbol{W}_{x,x}\in\mathcal{S}^{nm}_+}}\quad \langle\boldsymbol{C},\boldsymbol{Y}\rangle+\langle\boldsymbol{W}_{x,x},\boldsymbol{H}\rangle+\langle\boldsymbol{D},\boldsymbol{X}\rangle$$

$$\text{s.t.}\quad \boldsymbol{W}_{x,x}\succeq \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top,\qquad (13)$$
$$\begin{pmatrix}\sum_{i\in[n]}\boldsymbol{W}^{(i,i)}_{x,x} & \boldsymbol{X}^\top\\ \boldsymbol{X} & \boldsymbol{Y}\end{pmatrix}\succeq\boldsymbol{0}.$$

Proof of Theorem 2 We show that given a feasible solution to either problem we can generate an optimal solution to the other problem with an equal or lower objective value.

Suppose that $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W}_{x,x},\boldsymbol{W}_{x,y},\boldsymbol{W}_{y,y})$ is feasible in (12). Then, by summing appropriate semidefinite submatrices of the overall PSD matrix, we have that

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

Moreover, from (12) we have that $\sum_{i\in[n]}\boldsymbol{W}^{(i,i)}_{y,x}=\boldsymbol{X}^\top$ and $\sum_{i\in[n]}\boldsymbol{W}^{(i,i)}_{y,y}=\boldsymbol{Y}$. Thus, $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W}_{x,x})$ is feasible in (13) and attains the same objective value.

Next, suppose that $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W}_{x,x})$ is feasible in (13). By the Schur complement lemma, we must have $\boldsymbol{Y}\succeq\boldsymbol{X}(\sum_i\boldsymbol{W}^{(i,i)}_{x,x})^\dagger\boldsymbol{X}^\top$. Since $\boldsymbol{C}\succeq\boldsymbol{0}$, we can set $\boldsymbol{Y}:=\boldsymbol{X}(\sum_i\boldsymbol{W}^{(i,i)}_{x,x})^\dagger\boldsymbol{X}^\top$ without loss of optimality—doing so cannot increase the objective value. To construct admissible matrices $\boldsymbol{W}_{x,y}$ and $\boldsymbol{W}_{y,y}$, let us first define the auxiliary matrix

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

and observe that $\boldsymbol{Y}=\boldsymbol{U}^\top\boldsymbol{X}^\top=\boldsymbol{X}\boldsymbol{U}$. Then, we define $\boldsymbol{W}_{x,y}$, $\boldsymbol{W}_{y,y}$ as the blocks of the matrix

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

defined as

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

Since $\boldsymbol{Y}=\boldsymbol{U}^\top\boldsymbol{X}^\top$, we have $\mathrm{vec}(\boldsymbol{Y})=\mathrm{vec}(\boldsymbol{U}^\top\boldsymbol{X}^\top)=(\boldsymbol{I}_n\otimes\boldsymbol{U}^\top)\mathrm{vec}(\boldsymbol{X}^\top)$ and thus our construction is consistent with the existing value of $\boldsymbol{Y}$. We now verify that $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W}_{x,x},\boldsymbol{W}_{x,y},\boldsymbol{W}_{y,y})$ is feasible for (12). By construction, $\boldsymbol{M}\succeq\boldsymbol{0}$. Thus, $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W}_{x,x},\boldsymbol{W}_{x,y},\boldsymbol{W}_{y,y})$ satisfies the semidefinite constraint in (12). Next, by construction, $\boldsymbol{W}_{x,y}$ and $\boldsymbol{W}_{y,y}$ can be decomposed into $n\times n$ blocks satisfying:

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

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

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

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

Therefore, we conclude that $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}_{x,x}, \boldsymbol{W}_{x,y}, \boldsymbol{W}_{y,y})$ is feasible in (12) and attains an equal or lower objective value. Thus, both relaxations are equivalent. □

Problem (13) is much more compact that (12), as it does not require to introduce of the variables $\boldsymbol{W}_{y,y} \in \mathcal{S}_+^{n^2}$ nor $\boldsymbol{W}_{x,y} \in \mathbb{R}^{nm\times n}$. The proof of Theorem 2 provides a recipe for reconstructing an optimal $\boldsymbol{W}_{y,y}$ given an optimal solution $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{W}_{x,x})$ to (13). Namely, compute the auxiliary matrix $\boldsymbol{U} := \left(\sum_{i\in[n]} \boldsymbol{W}_{x,x}^{(i,i)}\right)^\dagger \boldsymbol{X}^\top$ and set $\boldsymbol{W}_{y,y} := (\boldsymbol{I}_n \otimes \boldsymbol{U})^\top \boldsymbol{W}_{x,x} (\boldsymbol{I}_n \otimes \boldsymbol{U})$. With this observation, one can implement the Goemans-Williamson sampling scheme for $\boldsymbol{Y}$ we propose in Section 4.3, even without solving a relaxation that explicitly involves $\boldsymbol{W}_{y,y}$.

Finally, it is interesting to consider whether the relaxation developed here is at least as strong as the matrix perspective relaxation developed by Bertsimas et al. (2023c). We now prove this is indeed the case. Bertsimas et al. (2023c) only applies to partially separable objectives. Hence, we first need to impose more structure on the objective of (11) to compare relaxations.

Proposition 5. *Assume that the term* $\langle \boldsymbol{H}, \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top\rangle + \langle \boldsymbol{D}, \boldsymbol{X}\rangle$ *in Problem* (11) *can be rewritten as the partially separable term* $\frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2 + h(\boldsymbol{X})$, *where* h *is convex in* $\boldsymbol{X}$. *Then, the optimal value of Problem* (12) *is at least as large as the relaxation of Bertsimas et al. (2023c)*

$$\min_{\boldsymbol{Y}\in \mathrm{Conv}\left(\mathcal{Y}_n^k\right)} \min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}, \boldsymbol{\theta}\in\mathcal{S}_+^m} \quad \langle \boldsymbol{C}, \boldsymbol{Y}\rangle + \frac{1}{2\gamma}\mathrm{tr}(\boldsymbol{\theta}) + h(\boldsymbol{X}) \tag{14}$$
$$s.t. \quad \langle \boldsymbol{A}_i, \boldsymbol{X}\rangle \le b_i \ \forall i\in[m], \ \begin{pmatrix} \boldsymbol{\theta} & \boldsymbol{X}^\top \\ \boldsymbol{X} & \boldsymbol{Y}\end{pmatrix} \succeq \boldsymbol{0},$$

Proof of Proposition 5 Given the equivalence between Problems (12)–(13) proven in Theorem 2, it suffices to show that the constraints in (13) imply the constraints in (14). Letting $\boldsymbol{\theta} := \sum_{i\in[n]} \boldsymbol{W}_{x,x}^{(i,i)}$, we observe that $\boldsymbol{\theta}$ is feasible for (14). In addition, given the additional assumption that the objective involves $\|\boldsymbol{X}\|_F^2 = \mathrm{tr}(\boldsymbol{X}^\top\boldsymbol{X})$, the objective in the relaxation is

$$\langle \boldsymbol{H}, \boldsymbol{W}_{x,x}\rangle = \mathrm{tr}\left(\sum_{i\in[n]} \boldsymbol{W}_{x,x}^{(i,i)}\right) = \mathrm{tr}\left(\boldsymbol{\theta}\right),$$

which completes the proof. □

The proof of Proposition 5 reveals that our Shor relaxation (13) can be perceived as decomposing the variable $\boldsymbol{\theta}$ in (14), and strengthening the relaxation by imposing additional constraints on the elements of this decomposition.

4.2. Strategies for Strengthening the Shor Relaxation

Theorem 2 might give the unfair impression that Problem (12) is not a useful relaxation, because it is equivalent to the much more compact optimization problem (13). However, explicit decision variables $\boldsymbol{W}_{y,y}\boldsymbol{W}_{x,y}$ allow us to express additional valid inequalities to strengthen the relaxation:

- The matrix $\boldsymbol{Y}$ being symmetric, $\mathrm{vec}(\boldsymbol{Y}) = \mathrm{vec}(\boldsymbol{Y}^\top) = \boldsymbol{K}_{n,n}\,\mathrm{vec}(\boldsymbol{Y})$, which leads to the constraints

$$\begin{aligned}\mathrm{vec}(\boldsymbol{Y})\,\mathrm{vec}(\boldsymbol{Y})^\top = \boldsymbol{K}_{n,n}\,\mathrm{vec}(\boldsymbol{Y})\,\mathrm{vec}(\boldsymbol{Y})^\top \boldsymbol{K}_{n,n}^\top &\implies \boldsymbol{W}_{y,y} = \boldsymbol{K}_{n,n}\boldsymbol{W}_{y,y}\boldsymbol{K}_{n,n}^\top,\\ \mathrm{vec}(\boldsymbol{X}^\top)\,\mathrm{vec}(\boldsymbol{Y})^\top = \mathrm{vec}(\boldsymbol{X}^\top)\,\mathrm{vec}(\boldsymbol{Y})^\top \boldsymbol{K}_{n,n}^\top &\implies \boldsymbol{W}_{x,y} = \boldsymbol{W}_{x,y}\boldsymbol{K}_{n,n}^\top.\end{aligned} \tag{15}$$

- If we further require the matrix $\boldsymbol{X}$ to be symmetric (implying $n = m$), then we can impose the additional linear equalities $\boldsymbol{W}_{x,x} = \boldsymbol{K}_{n,n}\boldsymbol{W}_{x,x}\boldsymbol{K}_{n,n}^\top$ and $\boldsymbol{W}_{x,y} = \boldsymbol{K}_{n,n}\boldsymbol{W}_{x,y}$.
- As in binary optimization, we can impose triangle inequalities on $\boldsymbol{Y}$ and $\boldsymbol{W}_{yy}$. Indeed, from the fact that $0 \le Y_{i,i} \le 1$, we have that any triplet (i, j, ℓ) satisfies

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

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

Finally, similarly to the previous section, one can tighten Problem (12) and Problem (13) by applying RLT. Any constraint of the form $\boldsymbol{A}\mathrm{vec}(\boldsymbol{X}) \le \boldsymbol{b}$ leads to the valid inequalities $\boldsymbol{A}\boldsymbol{W}_{x,x}\boldsymbol{A}^\top + \boldsymbol{b}\boldsymbol{b}^\top \ge \boldsymbol{b}\mathrm{vec}(\boldsymbol{X})^\top\boldsymbol{A} + \boldsymbol{A}^\top\boldsymbol{b}\mathrm{vec}(\boldsymbol{X})^\top$, as reviewed by Bao et al. (2011).[1]

4.3. Generalization of Goemans-Williamson Rounding to Low-Rank Optimization

Mirroring Section 3.2, we apply the Goemans-Williamson rounding scheme proposed in Algorithm 2 to low-rank optimization. First, we observe that under the constraint $\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}$, the term

$$\langle \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top, \boldsymbol{H}\rangle + \langle \boldsymbol{D}, \boldsymbol{X}\rangle$$

in the objective function of (11) is, through the identity $\mathrm{vec}(\boldsymbol{X}^\top) = (\boldsymbol{X} \otimes \boldsymbol{I}_n)\mathrm{vec}(\boldsymbol{Y}^\top)$, equal to

$$\langle \mathrm{vec}(\boldsymbol{Y})\mathrm{vec}(\boldsymbol{Y})^\top, (\boldsymbol{X} \otimes \boldsymbol{I}_n)\boldsymbol{H}(\boldsymbol{X} \otimes \boldsymbol{I}_n)\rangle + \langle \boldsymbol{Y}, \boldsymbol{X}^\top\boldsymbol{D}\rangle.$$

Thus, Problem (11) can be rewritten as the following optimization problem

$$\begin{aligned}\min_{\boldsymbol{X} \in \mathbb{R}^{n\times m}} \min_{\boldsymbol{Y} \in \mathcal{Y}_n^k} \quad & \left\langle \boldsymbol{C} + \boldsymbol{X}^\top\boldsymbol{D}, \boldsymbol{Y}\right\rangle + \langle \mathrm{vec}(\boldsymbol{Y})\mathrm{vec}(\boldsymbol{Y})^\top, (\boldsymbol{X} \otimes \boldsymbol{I}_n)\boldsymbol{H}(\boldsymbol{X} \otimes \boldsymbol{I}_n)\rangle\\ \text{s.t.} \quad & \langle \boldsymbol{A}_i, \boldsymbol{X}\rangle \le b_i \ \forall i \in [m],\ \boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X},\end{aligned} \tag{16}$$

where the lower-level optimization problem is quadratic in $\boldsymbol{Y}$ and very much reminiscent of the orthogonally constrained problem studied in Section 2. This suggests the Goemans-Williamson mechanism proposed in Algorithm 2 is a good candidate for generating feasible solutions to (16).

By analogy from our Goemans-Williamson rounding for logical constraints in §3.2, we formalize our Goemans-Williamson rounding scheme for rank-constrained optimization in Algorithm 4.

We make the following remarks on our implementation of Algorithm 4

Algorithm 4 A Goemans-Williamson Rounding Method for Logically Constrained Optimization

Generate solution to the semidefinite relaxation $\boldsymbol{Y}^\star, \boldsymbol{W}^\star_{y,y}$

Compute $\hat{\boldsymbol{Y}} : \mathrm{vec}(\hat{\boldsymbol{Y}}) \sim \mathcal{N}(\mathrm{vec}(\boldsymbol{Y}^\star), \boldsymbol{W}^\star_{y,y} - \mathrm{vec}(\boldsymbol{Y}^\star)\mathrm{vec}(\boldsymbol{Y}^\star)^\top)$

Construct $\bar{\boldsymbol{Y}} \in \mathcal{Y}^k_n$ which solves $\min_{\boldsymbol{Y} \in \mathcal{Y}^k_n} \|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\|_F$ (by performing an SVD)

Compute $\bar{\boldsymbol{X}}(\bar{\boldsymbol{Y}})$, an optimal $\boldsymbol{X}$ given $\bar{\boldsymbol{Y}}$ by solving

$$\begin{aligned} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}} \quad & \langle \boldsymbol{C}, \boldsymbol{Y} \rangle + \left\langle \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top, \boldsymbol{H} \right\rangle + \langle \boldsymbol{D}, \boldsymbol{X} \rangle \\ \text{s.t.} \quad & \langle \boldsymbol{A}_i, \boldsymbol{X} \rangle \le b_i \ \forall i \in [m], \ \boldsymbol{X} = \bar{\boldsymbol{Y}} \boldsymbol{X} \end{aligned}$$

return $\bar{\boldsymbol{Y}}$, $\bar{\boldsymbol{X}}(\bar{\boldsymbol{Y}})$ feasible solution to (16)

- To obtain a solution to our Shor relaxation, $\boldsymbol{Y}^\star$, $\boldsymbol{W}^\star_{y,y}$, we can either solve (12), or solve the equivalent compact relaxation (13) and reconstruct $\boldsymbol{W}^\star_{y,y}$ as $\boldsymbol{W}^\star_{y,y} := (\boldsymbol{I}_n \otimes \boldsymbol{U})^\top \boldsymbol{W}_{x,x} (\boldsymbol{I}_n \otimes \boldsymbol{U})$ where $\boldsymbol{U} := \left(\sum_{i \in [n]} \boldsymbol{W}^{(i,i)}_{x,x}\right)^\dagger \boldsymbol{X}^\top$.
- We take the second moment of our Gaussian distribution to be $\boldsymbol{W}^\star_{y,y} - \mathrm{vec}(\boldsymbol{Y}^\star)\mathrm{vec}(\boldsymbol{Y}^\star)^\top$ so that $\mathbb{E}[\mathrm{vec}(\hat{\boldsymbol{Y}})\mathrm{vec}(\hat{\boldsymbol{Y}})^\top] = \boldsymbol{W}^\star_{y,y}$.
- The sampled matrices $\hat{\boldsymbol{Y}}$ are not necessarily symmetric in general. However, if $\boldsymbol{W}^\star_{y,y}$ satisfies (15), $\hat{\boldsymbol{Y}}$ is symmetric almost surely. Consequently, it could be beneficial to sample using a moment matrix that satisfies the permutation-invariance constraints (15). In numerical experiments, we investigate the benefits of projecting $\boldsymbol{W}^\star_{y,y}$ onto the set of matrices satisfying (15) before sampling.
- In practice, we randomly round multiple times from the solution to the Shor relaxation and return the best solution $\bar{\boldsymbol{Y}}$ found, rather than only rounding once. This repetition improves the quality of the returned solution significantly, and comes at a low increase in computational cost because solving the Shor relaxation is more expensive than sampling $\bar{\boldsymbol{Y}}$ and computing $\bar{\boldsymbol{X}}(\bar{\boldsymbol{Y}})$.

5. Examples of Low-Rank Relaxations

This section applies the Shor relaxation technique proposed in §4 to several important problems from the low-rank literature. By exploiting problem structure, we demonstrate that it is often possible to reduce our Shor relaxation to a relaxation that does not involve any $n^2 \times n^2$ matrices.

5.1. Matrix Completion

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

$$\min_{\boldsymbol{Y} \in \mathcal{Y}_n} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}} \quad \|\mathcal{P}(\boldsymbol{A}) - \mathcal{P}(\boldsymbol{X})\|_F^2 + \lambda \cdot \mathrm{tr}(\boldsymbol{Y}) \text{ s.t.} \quad \boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}, \tag{17}$$

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

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

is a linear map which masks the hidden entries of $\boldsymbol{A}$. By expanding the quadratic $\|\mathcal{P}(\boldsymbol{A}) - \mathcal{P}(\boldsymbol{X})\|_F^2$, and invoking Theorem 2, we obtain the following relaxation of (17)

$$\begin{aligned} \min_{\boldsymbol{Y} \in \mathrm{Conv}(\mathcal{Y}_n)} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}, \boldsymbol{W} \in \mathcal{S}_+^{nm}} \quad & \sum_{i \in [n]} \langle \boldsymbol{W}^{(i,i)}, \boldsymbol{H}^i \rangle - 2 \langle \mathcal{P}(\boldsymbol{X}), \mathcal{P}(\boldsymbol{A}) \rangle + \langle \mathcal{P}(\boldsymbol{A}), \mathcal{P}(\boldsymbol{A}) \rangle + \lambda \cdot \mathrm{tr}(\boldsymbol{Y}) \\ \text{s.t.} \quad & \boldsymbol{W} \succeq \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top, \ \begin{pmatrix} \sum_{i \in [n]} \boldsymbol{W}_{x,x}^{(i,i)} & \boldsymbol{X}^\top \\ \boldsymbol{X} & \boldsymbol{Y} \end{pmatrix} \succeq \boldsymbol{0}, \end{aligned} \tag{18}$$

where $\boldsymbol{H}^i$ is a diagonal matrix which takes entries $\boldsymbol{H}_j^i = 1$ if $(i,j) \in \Omega$ and $\boldsymbol{H}_j^i = 0$ otherwise.

Compared with the matrix perspective relaxation of Bertsimas et al. (2023c), our relaxation is directly applicable to (17), while Bertsimas et al. (2023c) requires the presence of an additional Frobenius regularization term $+\frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2$ in the objective. With this additional term, our approach leads to relaxations of the form (18) after redefining $\boldsymbol{H}_i \leftarrow \boldsymbol{H}_i + \frac{1}{2\gamma}\boldsymbol{I}_m$, which are at least as strong as the relaxation of Bertsimas et al. (2023c) per Proposition 5.

We observe that the off-diagonal blocks of $\boldsymbol{W}$ do not appear in either the objective of (18) or any constraints other than $\boldsymbol{W} \succeq \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$. For this reason, we can omit them entirely:

Proposition 6. *Problems* (18) *attains the same optimal objective value as*

$$\begin{aligned} \min_{\boldsymbol{Y} \in \mathrm{Conv}(\mathcal{Y}_n)} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}, \boldsymbol{S}^i \in \mathcal{S}_+^{m}} \quad & \sum_{i \in [n]} \langle \boldsymbol{S}^i, \boldsymbol{H}^i \rangle - 2 \langle \mathcal{P}(\boldsymbol{X}), \mathcal{P}(\boldsymbol{A}) \rangle + \langle \mathcal{P}(\boldsymbol{A}), \mathcal{P}(\boldsymbol{A}) \rangle + \lambda \cdot \mathrm{tr}(\boldsymbol{Y}) \\ s.t. \quad & \boldsymbol{S}^i \succeq \boldsymbol{X}_{i,.}\boldsymbol{X}_{i,.}^\top, \ \begin{pmatrix} \sum_{i \in [n]} \boldsymbol{S}^i & \boldsymbol{X}^\top \\ \boldsymbol{X} & \boldsymbol{Y} \end{pmatrix} \succeq \boldsymbol{0}. \end{aligned} \tag{19}$$

Proof of Proposition 6 It suffices to show that given any feasible solution to (19) we can construct a feasible solution to (18) with the same objective value; the converse is immediate. Let $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{S}^i)$ be feasible in (19). Define the block matrix $\boldsymbol{W}$ by setting $\boldsymbol{W}^{(i,i)} = \boldsymbol{S}^i$ and $\boldsymbol{W}^{(i,j)} = (\boldsymbol{X}^\top)_i(\boldsymbol{X}^\top)_j^\top$. Then, it is not hard to see that $\boldsymbol{W} - \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$ is a block matrix with zero off-diagonal blocks and on-diagonal blocks $\boldsymbol{S}^i - \boldsymbol{X}_{i,.}\boldsymbol{X}_{i,.}^\top \succeq \boldsymbol{0}$. Thus, $\boldsymbol{W} - \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$ is a positive semidefinite matrix, and $\boldsymbol{W} \succeq \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$. Moreover, $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W})$ is feasible in (18) and attains the same objective value. □

Remark 5. Suppose that two columns of $\boldsymbol{A}$ have an identical sparsity pattern with respect to the known entries Ω, i.e., $\boldsymbol{H}^i = \boldsymbol{H}^j$. Then, we can replace the matrices $\boldsymbol{S}^i, \boldsymbol{S}^j$ with their sum $\tilde{\boldsymbol{S}}^{i,j} := \boldsymbol{S}^i + \boldsymbol{S}^j$ and rewrite (19) even more compactly, by omitting the matrices $\boldsymbol{S}^i, \boldsymbol{S}^j$, substituting $\tilde{\boldsymbol{S}}^{i,j}$ for $\boldsymbol{S}^i + \boldsymbol{S}^j$ in the objective/constraints, and requiring that $\tilde{\boldsymbol{S}}^{i,j} \succeq (\boldsymbol{X}^\top)_i(\boldsymbol{X}^\top)_i^\top + (\boldsymbol{X}^\top)_j(\boldsymbol{X}^\top)_j^\top$. This observation also applies if $k \geq 2$ columns share the same sparsity pattern.

In Section EC.4 we support our discussion on low-rank matrix completion by demonstrating that analogous reductions hold for low-rank basis pursuit.

5.2. Reduced Rank Regression

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

$$\min_{\boldsymbol{X}\in\mathbb{R}^{p\times n}} \quad \|\boldsymbol{B} - \boldsymbol{A}\boldsymbol{X}\|_F^2 + \mu \cdot \mathrm{rank}(\boldsymbol{X}), \tag{20}$$

where $\mu > 0$ controls the complexity of the estimator. For this problem, our Shor relaxation (13) is equivalent to the (improved) matrix perspective relaxation of Bertsimas et al. (2023c), which generalizes the optimal perspective relaxation for sparse regression (Dong et al. 2015).

Indeed, by invoking Theorem 2, we obtain (20)'s Shor relaxation

$$\begin{aligned} \min_{\boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n)} \min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}, \boldsymbol{W}\in\mathcal{S}_+^{nm}} \quad & \left\langle \boldsymbol{A}^\top \boldsymbol{A}, \sum_{i\in[n]} \boldsymbol{W}^{(i,i)} \right\rangle + \langle \boldsymbol{B}, \boldsymbol{B}\rangle - 2\langle \boldsymbol{A}\boldsymbol{X}, \boldsymbol{B}\rangle + \mu\cdot \mathrm{tr}(\boldsymbol{Y}) \\ \text{s.t.} \quad & \boldsymbol{W} \succeq \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top, \ \begin{pmatrix} \sum_{i\in[n]} \boldsymbol{W}^{(i,i)} & \boldsymbol{X}^\top \\ \boldsymbol{X} & \boldsymbol{Y} \end{pmatrix} \succeq \boldsymbol{0}, \end{aligned} \tag{21}$$

for which we show the following equivalence result:

Proposition 7. *Problem* (21) *attains the same objective value as*

$$\begin{aligned} \min_{\boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n)} \min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}, \boldsymbol{\theta}\in\mathcal{S}_+^{m}} \quad & \langle \boldsymbol{A}^\top \boldsymbol{A}, \boldsymbol{\theta}\rangle + \langle \boldsymbol{B}, \boldsymbol{B}\rangle - 2\langle \boldsymbol{A}\boldsymbol{X}, \boldsymbol{B}\rangle + \mu\cdot \mathrm{tr}(\boldsymbol{Y}) \\ s.t. \quad & \begin{pmatrix} \boldsymbol{\theta} & \boldsymbol{X}^\top \\ \boldsymbol{X} & \boldsymbol{Y} \end{pmatrix} \succeq \boldsymbol{0}, \end{aligned} \tag{22}$$

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

Proof of Proposition 7 We show that for any solution to (22) one can construct a solution to (21) with the same objective value or vice versa. Indeed, for any feasible solution $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{W})$ to (21), $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\theta} = \sum_{i\in[n]} \boldsymbol{W}^{(i,i)})$ is feasible for (22) with the same objective value. Conversely, let us consider $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\theta})$ a feasible solution to (22). Then,

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

because both matrices are PSD given that $\boldsymbol{Y} \preceq \boldsymbol{I}$. Therefore, it follows from the Schur complement lemma that $\boldsymbol{\theta} \succeq \boldsymbol{X}^\top \boldsymbol{X} = \sum_{i\in[n]} \boldsymbol{X}_{i,.}\boldsymbol{X}_{i,.}^\top$. Thus, there exists a decomposition $\boldsymbol{\theta} = \sum_{i\in[n]} \boldsymbol{S}^i$ with $\boldsymbol{S}_i \succeq \boldsymbol{X}_{i,.}\boldsymbol{X}_{i,.}^\top$ for each i. Finally, let us define the matrix $\boldsymbol{W}$ such that $\boldsymbol{W}^{(i,i)} = \boldsymbol{S}^i$ and $\boldsymbol{W}^{(i,j)} = \boldsymbol{X}_{i,.}\boldsymbol{X}_{j,.}^\top$ for $i \neq j$. Then, $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W})$ is feasible for (21) and attains the same objective value. The relaxation (22) is precisely the relaxation developed in Bertsimas et al. (2023c). □

Proposition 7's proof technique uses the fact that $\boldsymbol{X}$ enters the objective quadratically via $\boldsymbol{X}^\top\boldsymbol{X}$, rather than properties specific to reduced rank regression. This suggests other low-rank problems which are quadratic through $\boldsymbol{X}^\top \boldsymbol{X}$ (or $\boldsymbol{X}\boldsymbol{X}^\top$), 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 Shor relaxations.

We have shown in this section that for a wide variety of quadratic low-rank problems, it is possible to eliminate enough variables in the Shor relaxation that no matrices of size $n^2 \times n^2$ remain. This suggests that while Shor relaxations involving $n^2 \times n^2$ matrices may appear to be too large to be useful in practice, they can often be reduced to forms that are useful in practice.

6. Numerical Results

In this section, we benchmark our relax-then-sample schemes on synthetic semi-orthogonal quadratic and low-rank matrix completion problems. We also compare the performance of our schemes with the matrix perspective relaxation proposed by Bertsimas et al. (2022, 2023c).

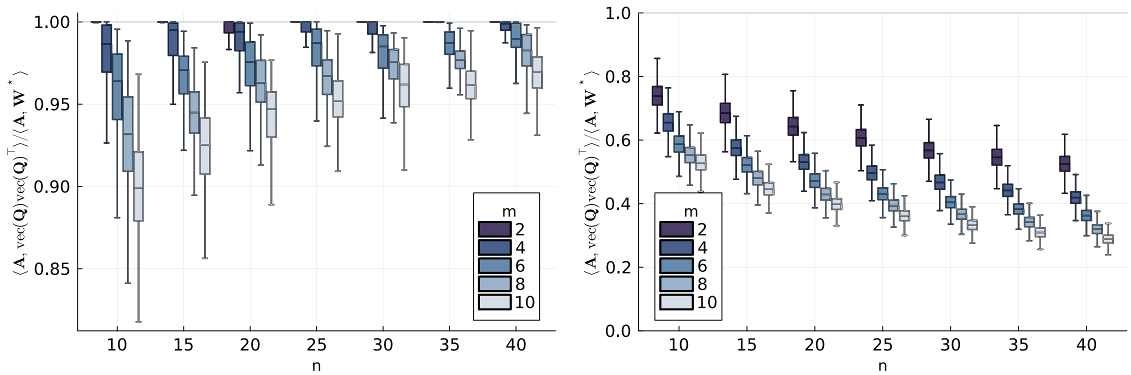
All experiments are conducted on a MacBook Pro laptop with a 36 GB Apple M3 CPU and 36 GB main memory, using MOSEK version 10.1, Julia version 1.9, and JuMP.jl version 1.13.0. All solver parameters are set to their default values. We divide our discussion into two parts. First, in §6.1, we study the quality of our relax-and-round scheme for semi-orthogonal quadratic problems. Second, in §6.2 we investigate the quality of our relax-and-round scheme for low-rank matrix completion problems and compare with prior literature.

6.1. Semi-Orthogonal Quadratic Optimization

We evaluate the performance of our Shor relaxation and Algorithm 2 for semi-orthogonal quadratic optimization problems (3). For fixed (n,m), we generate a random semidefinite matrix $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}^\top \in \mathcal{S}^{nm}_+$ where the entries of $\boldsymbol{B}$ are standard independent random variables. We solve the Shor relaxation (6) and sample $N = 2{,}000$ feasible solutions from Algorithm 2. For comparison, we also sample N solutions uniformly at random. We consider $n \in \{10, 15, \ldots, 40\}$ and $m \in \{2, 4, \ldots, 10\}$. We generate five instances for each size (n,m).

Figure 1 reports the distribution of performance ratios, $\langle \boldsymbol{A}, \operatorname{vec}(\boldsymbol{Q}) \operatorname{vec}(\boldsymbol{Q})^\top \rangle / \langle \boldsymbol{A}, \boldsymbol{W}^\star \rangle$, for solutions sampled according to Algorithm 2 (left panel) compared with uniformly sampled ones (right panel). We observe that Algorithm 2's average performance ratio is above 90% for all instances—for $m = 2$, the performance ratio is so close to 1 that the box plot appears as a thick line. Overall, performance improves with n and worsens with m. In practice, we may retain only the best solution out of N samples, so the upper quartile of the distribution is more representative of the quality of the returned solution. As a point of comparison, solutions generated uniformly at random achieve performance ratios around 60%, and their performance degrades as both n and m increase.

We should note that the empirical performance of Algorithm 2 is not entirely consistent with our theoretical bound: Theorem 1 guarantees a performance ratio that decreases with n, and



(a) Algorithm 2 (b) Uniformly sampled $\boldsymbol{Q}$

Figure 1 Distribution (boxplots) of performance ratios, $\langle \boldsymbol{A}, \mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top \rangle / \langle \boldsymbol{A}, \boldsymbol{W}^\star \rangle$, for solutions generated by Algorithm 2 (left panel) vs. uniformly sampled ones (right panel)

around 0.3–0.4 for these instances (see Figure EC.1). This discrepancy could be an artifact of the instance generation process. Our proof technique for Theorem 1 relies on controlling the distance $\|\alpha\,\mathrm{vec}(\boldsymbol{G})\,\mathrm{vec}(\boldsymbol{G})^\top - \mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top\|_F$ for some appropriately scaled α. Figure EC.2 displays the empirical distribution of $\|\alpha\,\mathrm{vec}(\boldsymbol{G})\,\mathrm{vec}(\boldsymbol{G})^\top - \mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top\|_F$ for several values of m. The solid vertical line indicates the value of m. Figure EC.2 shows that our value of α is appropriately scaled to ensure $\|\alpha\,\mathrm{vec}(\boldsymbol{G})\,\mathrm{vec}(\boldsymbol{G})^\top - \mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top\|_F < m$ with high probability, suggesting that tighter theoretical guarantees could only be achieved by using a different proof strategy.

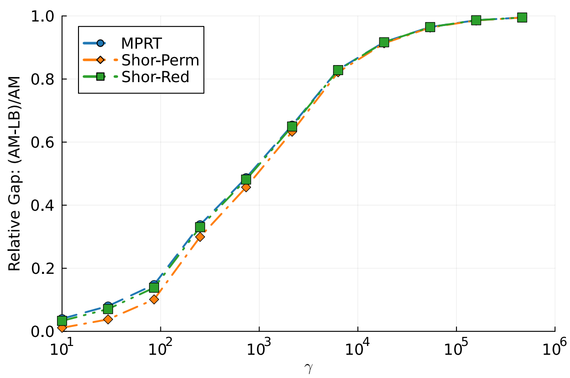
6.2. Low-Rank Matrix Completion

In this section, we evaluate the performance of Algorithm 4 on synthetic low-rank matrix completion instances. We use the data generation process of Candès and Recht (2009): We construct a matrix of observations, $\boldsymbol{A}_{\mathrm{full}} \in \mathbb{R}^{n\times m}$, from a rank-$r$ model: $\boldsymbol{A}_{\mathrm{full}} = \boldsymbol{U}\boldsymbol{V} + \epsilon\,\boldsymbol{Z}$, where the entries of $\boldsymbol{U} \in \mathbb{R}^{n\times r}, \boldsymbol{V} \in \mathbb{R}^{r\times m}$, and $\boldsymbol{Z} \in \mathbb{R}^{n\times m}$ are drawn independently from a standard normal distribution, and $\epsilon \geq 0$ models the degree of noise. We fix $\epsilon = 0.1, m = n$ and $r = 2$ for all experiments. We sample a random subset $\Omega \subseteq [n] \times [m]$, of predefined size (see also Candès and Recht 2009, section 1.1.2). Each result reported in this section is averaged over 10 random seeds.

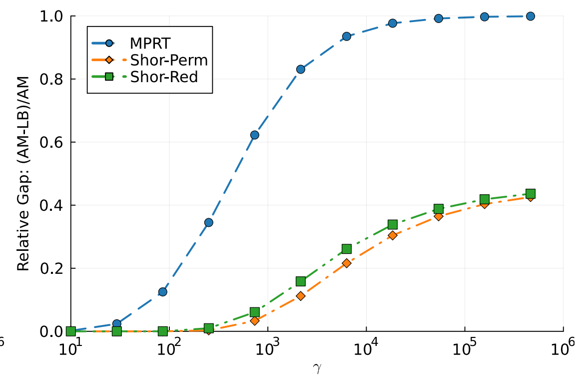
We first evaluate the quality of our new relaxations, compared with the matrix perspective relaxation of Bertsimas et al. (2023c, MPRT). Unfortunately, MPRT does not apply to (17) as it requires a Frobenius regularization term in the objective. Hence, instead of (17), we consider

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

for some regularization parameter $\gamma > 0$. As $\gamma \to \infty$, we recover the solution of (17). We compare the (lower) bounds obtained by three different approaches: MPRT, our full Shor relaxation (12) with



(a) $p = 0.5$

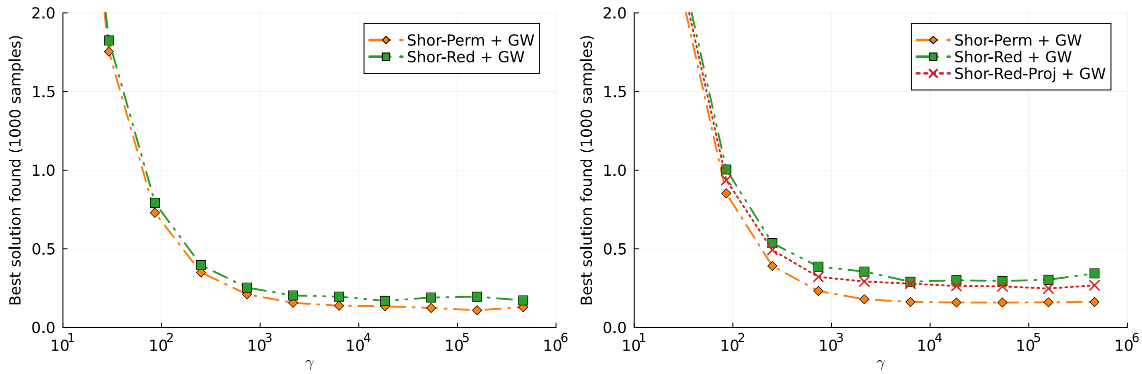


(b) $p = 0.95$

Figure 2 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 permutation equalities (15), hereafter denoted "Shor-Perm", and our compact Shor relaxation (19) ("Shor-Red"). Figure 2 reports the lower bounds achieved by each approach —in relative terms compared with an upper bounds achieved by the alternating minimization method of Burer and Monteiro (2003) initialized with a truncated SVD of $\mathcal{P}(\boldsymbol{A})$ (absolute values are reported in Figures EC.3–EC.4)— as γ increases, for different proportion of entries sampled $p = |\Omega|/mn$ ($n = 8$ being fixed). Supporting Proposition 5, we observe that Shor-Perm and Shor-Red obtains smaller optimality gaps than MPRT, for all values of γ, and that the benefit increases as the fraction of sampled entries p increases. In particular, when $p = 0.95$, there is a regime of values of γ (around 10^2) where both Shor relaxations are tight (as evidenced by a gap of 0%), while MPRT is not. In addition, as γ increases, MPRT achieves an uninformative gap of 100% (by returning a trivial lower bound of 0, see Figure EC.3), while our Shor relaxations provide non-trivial bounds (and gaps). From this experiment, it seems that imposing the permutation equalities (15) on $\boldsymbol{W}_{y,y}$ in our Shor relaxation (Shor-Perm vs. Shor-Red) does not lead so significantly tighter bounds, while being computationally much more expensive (see Figure EC.5 for computational times).

Our second experiment investigates the performance of our rounding strategy for the Shor relaxations, on the same instances. The relaxation Shor-Perm provides a matrix $\boldsymbol{W}_{y,y}$ directly. From a solution to the compact relaxation Shor-Red, we can reconstruct a matrix $\boldsymbol{W}_{y,y}$ using the reconstruction strategy discussed in Section 4.3. Figure 3 reports the best upper bound found from 1,000 sampled solutions. Interestingly, we observe that while the lower bounds from both relaxations in Figure 2 are rather similar, Shor-Perm provides a substantial improvement in the quality of the upper bound obtained, especially for higher values of p. Intuitively, this can be explained by the fact that the constraints (15) ensure that the sampled solution $\hat{\boldsymbol{Y}}$ is symmetric almost surely, hence is closer to being feasible. However, the matrix $\boldsymbol{W}_{y,y}$ recovered from Shor-Red does not satisfy



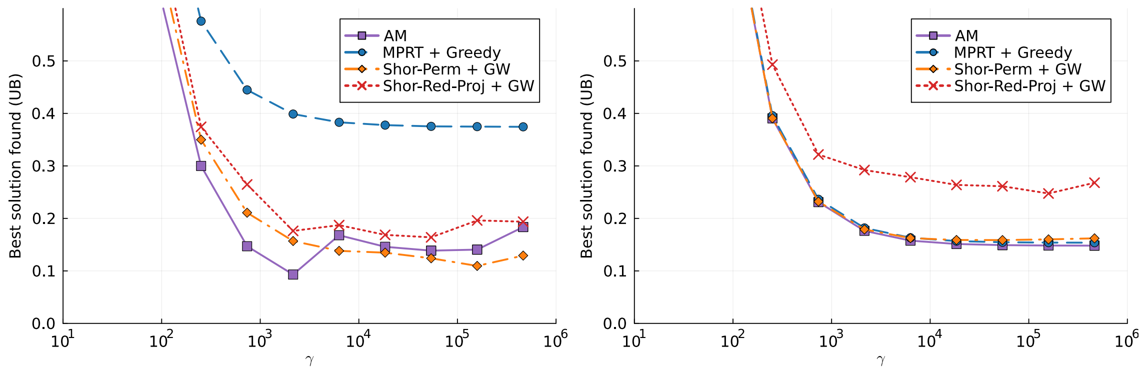
(a) $p = 0.5$ (b) $p = 0.95$

Figure 3 Average quality of GW rounding as we vary γ, for rounding the full Shor relaxation ("GW-Full") and the reduced relaxation with and without projecting $\boldsymbol{W}_{y,y}$ ("GW-Red-Proj", "GW-Red-NoProj").

these constraints. To support this intuition, we consider a third approach where we project the matrix $\boldsymbol{W}_{y,y}$ recovered from (15) onto the set $\{\boldsymbol{W} \in \mathcal{S}_+^{nm} : \boldsymbol{W} = \boldsymbol{K}_{n,m}\boldsymbol{W}\boldsymbol{K}_{n,m}^\top\}$ before sampling ("Shor-Red-Proj"). As displayed on the right panel of Figure 3, this additional projection step improves the quality of the solutions sampled from Shor-Red further, without significant additional computational cost, thus we use this projection technique for the rest of our numerics.

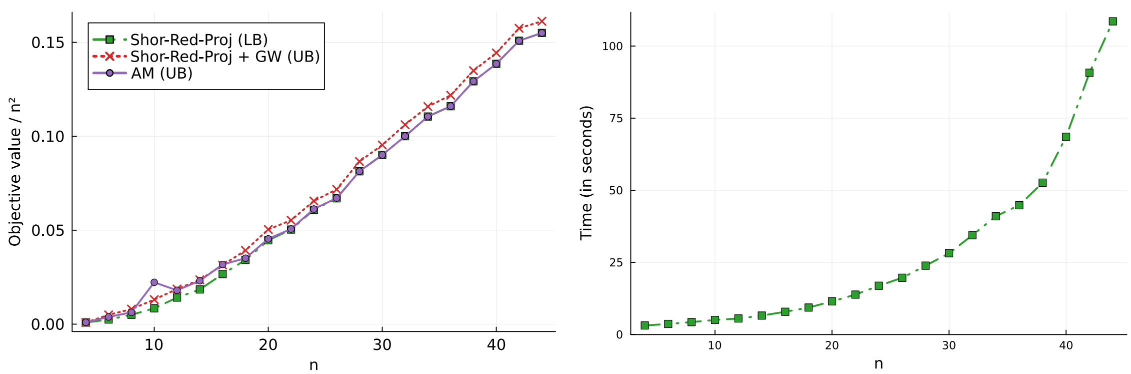
On the same instances, our third experiment compares Goemans-Williamson rounding with two other methods for generating feasible solutions: taking a truncated SVD of the MPRT relaxation (as advocated in Bertsimas et al. 2023c, "MPRT + Greedy") and alternating minimization initialized with a truncated SVD of $\mathcal{P}(\boldsymbol{A})$ ("AM"). Figure 4 depicts the upper bounds achieved by each method. Among the rounding-based schemes, we observe that Goemans-Williamson rounding on Shor-Perm performs significantly better than MPRT + Greedy when $p = 0.5$ and comparably when $p = 0.95$. The alternating minimization method of Burer and Monteiro (2003) is generally the best-performing method, except for instances with $p = 0.5$ and $\gamma \geq 10^4$. For these particularly challenging instances, which have many local optima, our Goemans-Williamson rounding could serve as an alternative or the initialization of the AM algorithm.

Our final experiment benchmarks the scalability of our reduced Shor relaxation and Goemans-Williamson rounding as we vary $n = m$ with the proportion of entries fixed at $p = 0.5$. We set $\gamma = 10^4/n^2$. We report the average upper and lower bound (divided by n^2 so that quantities have the same meaning as we vary n; left) and the average computational time (right) in Figure 5. We also report the average objective value obtained by alternating minimization as a baseline. Note that we do not consider the full Shor relaxation in this experiment, as it requires more RAM than is available for these experiments when $n = 10$. For any $n \in \{4, \ldots, 42\}$, the Shor relaxation can be solved in seconds, while when $n > 44$, Mosek runs out of RAM. Moreover, the lower bound from the Shor relaxation is tight for $n \geq 18$, although only alternating minimization matches the bound.



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

Figure 4 Average quality of feasible methods we vary γ, for GW rounding on the full Shor relaxation ("GW-Full"), on the reduced relaxation with projecting $\boldsymbol{W}_{y,y}$ ("GW-Red-Proj"), greedily rounding the matrix perspective relaxation ("MPRT-GD"), and alternating minimization ("AM").



(a) Objective values (b) Runtimes

Figure 5 Objective value (left panel) and runtime (right panel) as we vary $n = m$ with $p = 0.5$ for our reduced Shor relaxation followed by Goemans-Williamson a rounding. Results are averaged over 10 replications.

7. Conclusion

This paper proposes a new technique for relaxing and rounding quadratic optimization problems over semi-orthogonal matrices, and generalizes it to a broader class of low-rank optimization problems. We obtain new semidefinite relaxation by vectorizing the matrices and modeling the outer product of this vectorization with itself. By exploiting problem structure to eliminate most of the variables in our semidefinite relaxations, we show how to solve our relaxation efficiently. By interpreting the new decision variables in these relaxations as the second moment of a multivariate Gaussian distribution, we propose a sampling procedure, reminiscent of the Goemans-Williamson algorithm for BQO, which obtains high-quality solutions to low-rank problems in polynomial time.

References

Anstreicher K, Wolkowicz H (2000) On Lagrangian relaxation of quadratic matrix constraints. *SIAM Journal on Matrix Analysis and Applications* 22(1):41–55.

Arends F, Ouaknine J, Wampler CW (2011) On searching for small Kochen-Specker vector systems (extended version). *Graph-Theoretic Concepts in Computer Science*, volume 6986, 23—-34 (Springer, Heidelberg).

Bandeira AS, Kennedy C, Singer A (2016) Approximating the little Grothendieck problem over the orthogonal and unitary groups. *Mathematical Programming* 160:433–475.

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

Barak B, Kelner JA, Steurer D (2014) Rounding sum-of-squares relaxations. *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, 31–40.

Barratt S (2018) A matrix Gaussian distribution. *arXiv preprint arXiv:1804.11010* .

Barvinok A (2001) A remark on the rank of positive semidefinite matrices subject to affine constraints. *Discrete & Computational Geometry* 25(1):23–31.

Bell RM, Koren Y (2007) Lessons from the Netflix prize challenge. *ACM SIGKDD Explorations Newsletter* 9(2):75–79.

Ben-Tal A, Nemirovski A (2002) On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty. *SIAM Journal on Optimization* 12(3):811–833.

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(1):12478–12528.

Bertsimas D, Cory-Wright R, Lo S, Pauphilet J (2023b) Optimal low-rank matrix completion: Semidefinite relaxations and eigenvector disjunctions. *arXiv preprint arXiv:2305.12292* .

Bertsimas D, Cory-Wright R, Pauphilet J (2021) A unified approach to mixed-integer optimization problems with logical constraints. *SIAM Journal on Optimization* 31(3):2340–2367.

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 (2023c) A new perspective on low-rank optimization. *Mathematical Programming* 202:47–92.

Bertsimas D, Ye Y (1998) Semidefinite relaxations, multivariate normal distributions, and order statistics. *Handbook of Combinatorial Optimization*, 1473–1491 (Springer).

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

Briët J, Buhrman H, Toner B (2011) A generalized Grothendieck inequality and nonlocal correlations that require high entanglement. *Communications in Mathematical Physics* 305(3):827–843.

Briët J, de Oliveira Filho FM, Vallentin F (2010) The positive semidefinite Grothendieck problem with rank constraint. *International Colloquium on Automata, Languages, and Programming*, 31–42 (Springer).

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

Burer S, Park K (2023) A strengthened SDP relaxation for quadratic optimization over the Stiefel manifold. *Journal of Optimization Theory and Applications* 1–20.

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

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

Chen Y, Ye Y, Wang M (2019) Approximation hardness for a class of sparse optimization problems. *Journal of Machine Learning Research* 20(38):1–27.

Cory-Wright R, Pauphilet J (2022) Sparse PCA with multiple components. *arXiv preprint arXiv:2209.14790* .

Deshpande Y, Montanari A (2015) Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. *Conference on Learning Theory*, 523–562 (PMLR).

Dong H, Chen K, Linderoth J (2015) Regularization vs. relaxation: A conic optimization perspective of statistical variable selection. *arXiv preprint arXiv:1510.06083* .

d'Aspremont A, Boyd S (2003) Relaxations and randomized methods for nonconvex QCQPs. *EE392o Class Notes, Stanford University* 1:1–16.

Gillis N, Glineur F (2011) Low-rank matrix approximation with weights or missing data is NP-hard. *SIAM Journal on Matrix Analysis and Applications* 32(4):1149–1165.

Gilman K, Burer S, Balzano L (2022) A semidefinite relaxation for sums of heterogeneous quadratics on the Stiefel manifold. *arXiv preprint arXiv:2205.13653* .

Goemans MX, Williamson DP (1995) Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)* 42(6):1115–1145.

Grimmett G, Stirzaker D (2020) *Probability and Random Processes* (Oxford University Press).

Gupta AK, Nagar DK (2018) *Matrix Variate Distributions* (Chapman and Hall/CRC).

Han S, Gómez A, Atamtürk A (2022) The equivalence of optimal perspective formulation and Shor's SDP for quadratic programs with indicator variables. *Operations Research Letters* 50(2):195–198.

Kim J, Tawarmalani M, Richard JPP (2022) Convexification of permutation-invariant sets and an application to sparse principal component analysis. *Mathematics of Operations Research* 74(4):2547–2584.

Li Y, Xie W (2024) On the partial convexification of the low-rank spectral optimization: Rank bounds and algorithms. *International Conference on Integer Programming and Combinatorial Optimization*, 265–279 (Springer).

Lubin M, Vielma JP, Zadik I (2022) Mixed-integer convex representability. *Mathematics of Operations Research* 47(1):720–749.

Luo ZQ, Ma WK, So AMC, Ye Y, Zhang S (2010) Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine* 27(3):20–34.

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

Meckes ES (2019) *The Random Matrix Theory of the Classical Compact Groups*, volume 218 (Cambridge University Press).

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

Nemirovski A (2007) Sums of random symmetric matrices and quadratic optimization under orthogonality constraints. *Mathematical Programming* 109(2-3):283–317.

Nesterov Y (1998) Semidefinite relaxation and nonconvex quadratic optimization. *Optimization Methods and Software* 9(1-3):141–160.

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.

Padberg M (1989) The Boolean quadric polytope: some characteristics, facets and relatives. *Mathematical Programming* 45:139–172.

Pataki G (1998) On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Mathematics of Operations Research* 23(2):339–358.

Recht B, Fazel M, Parrilo PA (2010) Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review* 52(3):471–501.

Rehfeldt D, Koch T, Shinano Y (2023) Faster exact solution of sparse MaxCut and QUBO problems. *Mathematical Programming Computation* 15(3):445–470.

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

Tropp JA (2012) User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics* 12:389–434.

Tropp JA (2015) An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning* 8(1-2):1–230.

Wainwright MJ (2019) *High-dimensional Statistics: A Non-Asymptotic Viewpoint*, volume 48 (Cambridge University Press).

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

Williamson DP, Shmoys DB (2011) *The Design of Approximation Algorithms* (Cambridge university press).

Wolkowicz H, Saigal R, Vandenberghe L (1998) *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, volume 27 (Springer Science & Business Media).

Zheng X, Sun X, Li D (2014) Improving the performance of MIQP solvers for quadratic programs with cardinality and minimum threshold constraints: A semidefinite program approach. *INFORMS Journal on Computing* 26(4):690–703.

Supplementary Material

EC.1. Non-Equivalence of Shor and Optimal Perspective Relaxations

Consider a sparse linear regression problem setting of the form

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \quad \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2 \text{ s.t. } \|\boldsymbol{\beta}\|_0 \leq k,$$

and its semidefinite relaxations (a) Problem (9) reinforced with the triangle inequalities

$$z_i + z_j + z_l \leq Z_{i,j} + Z_{i,k} + Z_{j,k} + 1 \qquad \forall i, j, k \in [n],$$
$$Z_{i,j} + Z_{i,k} \leq z_i + Z_{j,k} \qquad \forall i, j, k \in [n],$$

and (b) the more compact semidefinite relaxation (10), which as proven in Proposition 3 is equivalent to Problem (9) (without the triangle inequalities). Let the problem data be $p = 6, n = 8, k = 3$ and

$$\boldsymbol{X} = \begin{pmatrix} 1.04 & 0.97 & 0.35 & 0.34 & 0.04 & 0.62 \\ 1.13 & 1.08 & 0.66 & 0.78 & 0.85 & 0.45 \\ 1.50 & 2.54 & 1.73 & 0.11 & -1.06 & -0.41 \\ 0.65 & -1.42 & -1.52 & -1.03 & -0.11 & 0.81 \\ 0.49 & -1.17 & -1.58 & 0.60 & 0.70 & 1.53 \\ 0.51 & -1.34 & -1.53 & 0.07 & -0.10 & 0.17 \\ 0.81 & 2.63 & -0.90 & 1.73 & 1.36 & 1.73 \\ 0.76 & 0.71 & 0.08 & -0.20 & -0.57 & -0.13 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} 0.43 \\ 0.84 \\ 1.15 \\ -2.22 \\ -1.44 \\ -1.94 \\ -3.18 \\ -2.44 \end{pmatrix}$$

Then, using Mosek version 10.2 to solve all relaxations and Gurobi version 10.0.2 to solve the mixed-integer problem:

- Problem (9) reinforced with the triangle inequalities described in Remark EC.1 has an optimal objective value of 1.45886.
- Problem (8) has an optimal objective value of 1.4118.
- The original MINLO has an optimal objective value of 1.5336.

Thus, the Shor relaxation with triangle inequalities and the more compact semidefinite relaxation are not equivalent.

EC.2. Non-Equivalence of Reduced Shor Relaxation and Shor Relaxation in Presence of Permutation Equalities

Consider a low-rank matrix completion problem of the form

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

and its semidefinite relaxations (a) the matrix perspective relaxation as introduced in the paper Bertsimas et al. (2023c), (b) the semidefinite relaxation (12) with the inequalities on $\boldsymbol{W}_{x,y}$ and $\boldsymbol{W}_{y,y}$, (c) the semidefinite relaxation (13). Let the problem data be $\gamma=100, k=2, n=7, m=5$, and suppose we are trying to impute the following matrix, where $*$ denotes a missing entry:

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

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

- The matrix perspective relaxation as introduced in the paper Bertsimas et al. (2023c) has an optimal objective value of 3.9275.
- The semidefinite relaxation (12) has an optimal objective value of 5.1387.
- The more compact semidefinite relaxation (13) has an objective value of 4.314.
- The method of Burer and Monteiro (2003) finds a feasible solution with objective value 9.495.

Thus, we conclude that the permutation inequalities in (12) are not redundant, and the reduction in Theorem 2 does not hold in the presence of these inequalities. Nonetheless, the reduction is useful because it produces a non-trivial lower bound after solving a smaller semidefinite problem.

EC.3. Additional Technical Results

EC.3.1. From Sub-Gaussian Tails to Bound on Expectation

LEMMA EC.1. *Consider a non-negative random variable Z that satisfies the following tail bound*

$$\forall t>0, \quad \mathbb{P}\left(Z>t\right)\le Ce^{-t^2/2\sigma^2},$$

for $C,\sigma>0$. Then, we have

$$\mathbb{E}[Z^2]\le 2\sigma^2\log(C)+2\sigma^2.$$

Proof of Lemma EC.1 Using the characterization of the expected value for non-negative random variables, we have

$$\mathbb{E}[Z^2]=\int_0^\infty \mathbb{P}\left(Z^2\ge t\right)dt=\int_0^\infty \mathbb{P}\left(Z\ge\sqrt{t}\right)dt$$

$$= \int_0^{2\sigma^2 \log(C)} \mathbb{P}\left(Z \geq \sqrt{t}\right) dt + \int_{2\sigma^2 \log(C)}^{\infty} \mathbb{P}\left(Z \geq \sqrt{t}\right) dt.$$

We bound $\mathbb{P}\left(Z \geq \sqrt{t}\right)$ by 1 in the first integral and $Ce^{-t/2\sigma^2}$ in the second one and get

$$\mathbb{E}[Z^2] \leq 2\sigma^2 \log(C) + C\left[-\frac{\sigma^2}{2} e^{-t/2\sigma^2}\right]_{2\sigma^2 \log(C)}^{\infty} = 2\sigma^2 \log(C) + 2\sigma^2.$$

□

EC.3.2. Proof of Lemma 3

Proof of Lemma 3 Noting that $\boldsymbol{W}^{(i,j)} = \sum_{k\in[r]} \boldsymbol{B}_k \boldsymbol{e}_i \boldsymbol{e}_j^\top \boldsymbol{B}_k^\top$, we have

$$\left(\sum_k \boldsymbol{B}_k^\top \boldsymbol{B}_k\right)_{i,j} = \sum_{k\in[r]} \boldsymbol{e}_i^\top \boldsymbol{B}_k^\top \boldsymbol{B}_k \boldsymbol{e}_j = \operatorname{tr}(\boldsymbol{W}^{(i,j)}),$$
$$\sum_{k\in[r]} \boldsymbol{B}_k \boldsymbol{B}_k^\top = \sum_{k\in[r]} \sum_{i\in[m]} \boldsymbol{B}_k \boldsymbol{e}_i \boldsymbol{e}_i^\top \boldsymbol{B}_k^\top = \sum_{i\in[m]} \boldsymbol{W}^{(i,i)}.$$

We conclude by using the fact that $\boldsymbol{W}$ satisfies the constraints in (6). □

EC.3.3. Concentration of the Frobenius Norm

LEMMA EC.2. *Consider a random matrix $\boldsymbol{G}$ sampled according to $\operatorname{vec}(\boldsymbol{G}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{W})$, where the matrix $\boldsymbol{W}$ is a feasible solution to (6). Then, the deviation of $\|\boldsymbol{G}\|_F$ from its mean satisfies the following tail bound: For any $t \geq 0$,*

$$\mathbb{P}\left(|\|\boldsymbol{G}\|_F - \mathbb{E}[\|\boldsymbol{G}\|_F]| \geq t\right) \leq 2e^{-t^2/2\lambda_{\max}(\boldsymbol{W})}.$$

Proof of Lemma EC.2 Recall that $\boldsymbol{G}$ is described as a Gaussian matrix series: $\boldsymbol{G} = \sum_{k\in[nm]} \boldsymbol{B}_k z_k$ where z_k are i.i.d. standard random variables. Denoting $f(\boldsymbol{z}) := \|\sum_{k\in[nm]} \boldsymbol{B}_k z_k\|_F$, we have

$$|f(\boldsymbol{z}) - f(\boldsymbol{z}')| \leq \left\|\sum_k \boldsymbol{B}_k (z_k - z_k')\right\|_F = \sup_{\boldsymbol{U}:\|\boldsymbol{U}\|_F=1} \sum_k \langle \boldsymbol{U}, \boldsymbol{B}_k\rangle (z_k - z_k') \leq \sup_{\boldsymbol{U}:\|\boldsymbol{U}\|_F=1} \sqrt{\sum_k \langle \boldsymbol{U}, \boldsymbol{B}_k\rangle^2}\, \|\boldsymbol{z} - \boldsymbol{z}'\|,$$

so the function f is L-Lipschitz with

$$L^2 = \sup_{\boldsymbol{U}:\|\boldsymbol{U}\|_F=1} \sum_k \langle \boldsymbol{U}, \boldsymbol{B}_k\rangle^2 = \sup_{\boldsymbol{u}:\|\boldsymbol{u}\|=1} \sum_k \boldsymbol{u}^\top \operatorname{vec}(\boldsymbol{B}_k) \operatorname{vec}(\boldsymbol{B}_k)^\top \boldsymbol{u} = \sup_{\boldsymbol{u}:\|\boldsymbol{u}\|=1} \boldsymbol{u}^\top \boldsymbol{W} \boldsymbol{u} = \lambda_{\max}(\boldsymbol{W}).$$

The result follows from concentration inequalities for Lipschitz function of i.i.d. standard random variables (Wainwright 2019, Theorem 2.26). □

COROLLARY EC.1. *Under the assumptions of Lemma EC.2, we have*

$$\mathbb{E}\left[\|\boldsymbol{G}\|_F\right]^2 \geq m - 4\lambda_{\max}(\boldsymbol{W}).$$

Proof of Corollary EC.1 Denote $Z = |\|\boldsymbol{G}\|_F - \mathbb{E}[\|\boldsymbol{G}\|_F]|$. On one side,

$$\mathbb{E}\left[Z^2\right] = \mathbb{E}\left[\|\boldsymbol{G}\|_F^2\right] - \mathbb{E}\left[\|\boldsymbol{G}\|_F\right]^2 = m - \mathbb{E}\left[\|\boldsymbol{G}\|_F\right]^2.$$

On the other side, the tail bound in Lemma EC.2 combined with Lemma EC.1 leads to

$$\mathbb{E}\left[Z^2\right] \leq 2\log(2)\lambda_{\max}(\boldsymbol{W}) + 2\lambda_{\max}(\boldsymbol{W}) \leq 4\lambda_{\max}(\boldsymbol{W}).$$

Eventually, we get

$$m - \mathbb{E}\left[\|\boldsymbol{G}\|_F\right]^2 \leq 4\lambda_{\max}(\boldsymbol{W}).$$

□

EC.3.4. Gaussian Vector and its Normalized Version

In this section, we report results comparing the expectation of a Gaussian random vector and that of its normalized version.

LEMMA EC.3. *Consider one vector* $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}_k, \boldsymbol{\Sigma})$ *with* $\boldsymbol{\Sigma} \succ \boldsymbol{0}$. *Define* $\boldsymbol{u} := \frac{1}{\|\boldsymbol{z}\|}\boldsymbol{z}$.

$$\frac{1}{k\lambda_{\max}(\boldsymbol{\Sigma})}\mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top] \preceq \mathbb{E}[\boldsymbol{u}\boldsymbol{u}^\top] \preceq \frac{1}{k\lambda_{\min}(\boldsymbol{\Sigma})}\mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top].$$

REMARK EC.1. Here, we assume that $\boldsymbol{\Sigma}$ is full-rank and take $\boldsymbol{u} \in \mathbb{S}^k$, the unit sphere in dimension k. If $\boldsymbol{\Sigma}$ is rank-deficient, however, $\boldsymbol{z} \in \operatorname{span}(\boldsymbol{\Sigma})$ a.s.–and so $\boldsymbol{u} \in \operatorname{span}(\boldsymbol{\Sigma})$ as well. As a result, we can apply the same reasoning to $\operatorname{span}(\boldsymbol{\Sigma})$, which is homeomorphic to $\mathbb{R}^{\operatorname{rank}(\boldsymbol{\Sigma})}$, instead of the entire space $\mathbb{R}^k$.

Proof of Lemma EC.3 Denote $r := \|\boldsymbol{z}\|$. Then, the joint density of $(\boldsymbol{u}, r) \in \mathbb{S}^k \times \mathbb{R}_+$ is

$$p(r,\boldsymbol{u}) := \frac{r^{l-1}}{\sqrt{|\boldsymbol{\Sigma}|}(2\pi)^{\frac{k}{2}}} e^{-\frac{1}{2}r^2\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{u}}.$$

With these notations, and implicitly using Fubini's theorem to write the integration over $(r,\boldsymbol{u})$ as an integration over r, followed by an integration over $\boldsymbol{u}$, we have

$$\begin{aligned}\mathbb{E}[\boldsymbol{u}\boldsymbol{u}^\top] &= \int_{\boldsymbol{u},r} \boldsymbol{u}\boldsymbol{u}^\top \frac{r^{k-1}}{\sqrt{|\boldsymbol{\Sigma}|}(2\pi)^{\frac{k}{2}}} e^{-\frac{1}{2}r^2\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{u}} d\boldsymbol{u}dr \\ &= \frac{1}{\sqrt{|\boldsymbol{\Sigma}|}(2\pi)^{\frac{k}{2}}} \int_{\boldsymbol{u}} \boldsymbol{u}\boldsymbol{u}^\top I_{k-1}\left(\tfrac{1}{2}\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\right) d\boldsymbol{u},\end{aligned}$$

where we define the Gaussian integral

$$I_n(b) := \int_0^{+\infty} x^n e^{-bx^2} dx,$$

for any integer n and scalar $b > 0$.

Similarly,

$$\begin{aligned}\mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top] &= \mathbb{E}[r^2\boldsymbol{u}\boldsymbol{u}^\top] \\ &= \int_{\boldsymbol{u},r} \boldsymbol{u}\boldsymbol{u}^\top \frac{r^{k+1}}{\sqrt{|\boldsymbol{\Sigma}|}(2\pi)^{\frac{k}{2}}} e^{-\frac{1}{2}r^2\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{u}} d\boldsymbol{u}dr \\ &= \frac{1}{\sqrt{|\boldsymbol{\Sigma}|}(2\pi)^{\frac{k}{2}}} \int_{\boldsymbol{u}} \boldsymbol{u}\boldsymbol{u}^\top I_{k+1}\left(\tfrac{1}{2}\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\right) d\boldsymbol{u}.\end{aligned}$$

By integration by part, we have the recursive formula

$$I_{k+1}(b) = \frac{k}{2b} I_{k-1}(b),$$

so that

$$\mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top] = \frac{k}{\sqrt{|\boldsymbol{\Sigma}|}(2\pi)^{\frac{k}{2}}} \int \boldsymbol{u}\boldsymbol{u}^\top \left(\boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{u}\right)^{-1} I_{n+1}\left(\tfrac{1}{2}\boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{u}\right) d\boldsymbol{u}.$$

Since $\lambda_{\min}(\boldsymbol{\Sigma}^{-1}) \leq \boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{u} \leq \lambda_{\max}(\boldsymbol{\Sigma}^{-1})$ and $\lambda_{\min}(\boldsymbol{\Sigma}) = \lambda_{\max}(\boldsymbol{\Sigma}^{-1})^{-1}$, $\lambda_{\max}(\boldsymbol{\Sigma}) = \lambda_{\min}(\boldsymbol{\Sigma}^{-1})^{-1}$, we have

$$k\lambda_{\min}(\boldsymbol{\Sigma})\, \mathbb{E}[\boldsymbol{u}\boldsymbol{u}^\top] \quad \preceq \mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top] \preceq \quad k\lambda_{\max}(\boldsymbol{\Sigma})\, \mathbb{E}[\boldsymbol{u}\boldsymbol{u}^\top],$$

which concludes the proof. □

EC.4. Basis Pursuit Discussion

In this section, we support our discussion of compact relaxations for low-rank matrix completion problems (Section 5.1) by demonstrating analogous results hold in the low-rank basis pursuit case.

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

$$\min_{\boldsymbol{Y} \in \mathcal{Y}_n} \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}} \quad \mathrm{tr}(\boldsymbol{Y}) \text{ s.t. } \quad \mathcal{P}(\boldsymbol{A}) = \mathcal{P}(\boldsymbol{X}), \boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}, \tag{EC.1}$$

where $\mathcal{P}(\boldsymbol{A})$ denotes a linear map that masks the hidden entries of $\boldsymbol{A}, \boldsymbol{X}$ such that $\mathcal{P}(\boldsymbol{A})_{i,j} = A_{i,j}$ if $(i,j) \in \Omega$ and 0 otherwise. Following Theorem 2 and applying RLT to the constraints $A_{i,j} - X_{i,j} = 0, \forall (i,j) \in \Omega$ leads to the following relaxation

$$\begin{aligned} \min_{\boldsymbol{Y} \in \mathrm{Conv}(\mathcal{Y}_n)} \ \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}, \boldsymbol{W} \in \mathcal{S}^{nm}_+} & \mathrm{tr}(\boldsymbol{Y}) \\ \text{s.t.} \quad & A_{i,j}A_{k,\ell} - A_{k,\ell}X_{i,j} - A_{i,j}X_{k,\ell} + (\boldsymbol{W}^{(i,k)})_{j,\ell} = 0, \forall (i,j),(k,\ell) \in \Omega \times \Omega \\ & A_{i,j} = X_{i,j}, \forall (i,j) \in \Omega \\ & \boldsymbol{W} \succeq \mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top, \ \begin{pmatrix} \sum_{i \in [n]} \boldsymbol{W}^{(i,i)} & \boldsymbol{X}^\top \\ \boldsymbol{X} & \boldsymbol{Y} \end{pmatrix} \succeq \boldsymbol{0}, \end{aligned} \tag{EC.2}$$

Similarly to the low-rank matrix completion case, the structure of the compact Shor relaxation means that the off-diagonal blocks of $\boldsymbol{W}$ do not appear in either the objective nor any constraint involving $\boldsymbol{Y}$. As we prove below, the off-diagonal blocks can, therefore, be eliminated from the relaxation without impacting its optimal value:

Proposition EC.1. *Problem* (EC.2) *attains the same objective value as*

$$\begin{aligned}\min_{\boldsymbol{Y}\in\mathrm{Conv}(\mathcal{Y}_n)}\ \min_{\boldsymbol{X}\in\mathbb{R}^{n\times m},\ \boldsymbol{S}^i\in\mathcal{S}^m_+, i\in[n]}\quad & \mathrm{tr}(\boldsymbol{Y})\\ s.t.\quad & A_{i,j}A_{i,\ell}-A_{i,\ell}X_{i,j}-A_{i,j}X_{i,\ell}+(\boldsymbol{S}^i)_{j,\ell}=0, \forall (i,j),(i,\ell)\in\Omega\times\Omega\\ & A_{i,j}=X_{i,j}, \forall (i,j)\in\Omega\\ & \boldsymbol{S}^i\succeq \boldsymbol{X}_{i,\cdot}\boldsymbol{X}_{i,\cdot}^\top,\ \begin{pmatrix}\sum_{i\in[n]}\boldsymbol{S}^i & \boldsymbol{X}^\top\\ \boldsymbol{X} & \boldsymbol{Y}\end{pmatrix}\succeq \boldsymbol{0},\end{aligned} \tag{EC.3}$$

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

Proof of Proposition EC.1 From a solution to (EC.2), defining $\boldsymbol{S}^i := \boldsymbol{W}^{(i,i)}$ yields a feasible solution to (EC.3) with same objective value. In turn, let us consider a feasible solution to (EC.2), $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{S}^i)$. Define the block matrix $\boldsymbol{W}\in\mathcal{S}^{nm}$ by setting $\boldsymbol{W}^{(i,i)}=\boldsymbol{S}^i$ and $\boldsymbol{W}^{(i,k)}=\boldsymbol{X}_{i,\cdot}\boldsymbol{X}_{k,\cdot}^\top$. Then, it is not hard to see that $\boldsymbol{W}-\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top$ is a block diagonal matrix with on-diagonal blocks $\boldsymbol{S}^i-\boldsymbol{X}_{i,\cdot}\boldsymbol{X}_{i,\cdot}^\top\succeq\boldsymbol{0}$. Thus, $\boldsymbol{W}-\mathrm{vec}(\boldsymbol{X}^\top)\mathrm{vec}(\boldsymbol{X}^\top)^\top\succeq\boldsymbol{0}$. Moreover,

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

So the linear constraints indexed by $(i,j),(i,\ell)\in\Omega\times\Omega$ are all satisfied. Thus, $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{W})$ is feasible in (EC.2) and attains the same objective value. □

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

EC.5. Additional Numerical Results

This section complements Section 6.

EC.5.1. Theoretical Performance Guarantees of Algorithm 2

We should note that the empirical performance of Algorithm 2 is not entirely consistent with our theoretical bound: Figure EC.1 displays the theoretical performance ratio $\alpha = 1/\log(n+m)$ for the n and m values considered in the numerical experiments in Section 6.1.

Moreover, Figure EC.2 depicts the empirical distribution of $\|\alpha\,\mathrm{vec}(\boldsymbol{G})\,\mathrm{vec}(\boldsymbol{G})^\top - \mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top\|_F$ for different values of m:

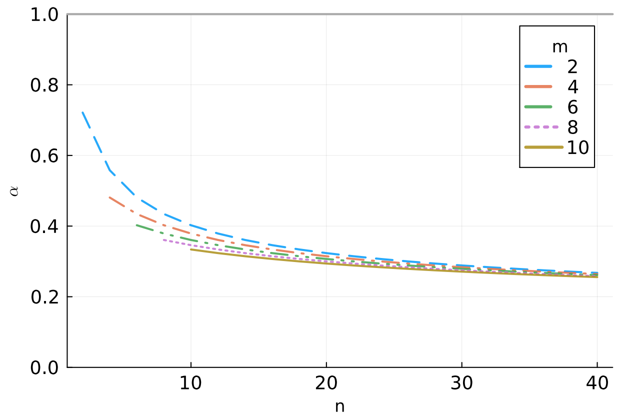


Figure EC.1 Theoretical performance (multiplicative) ratio α of Algorithm 2 as n increases, for different values of m. Note that we start the curves at $n = m$ because $n < m$ is not feasible.

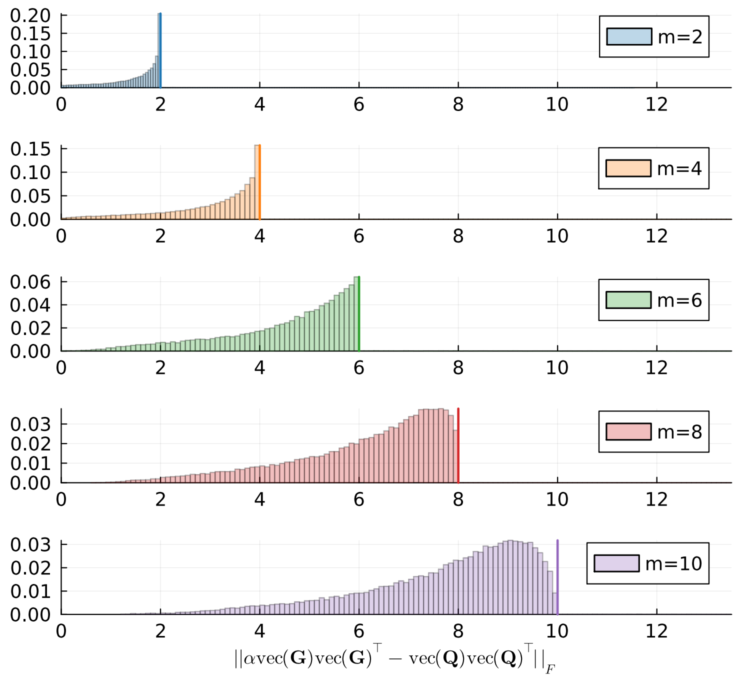
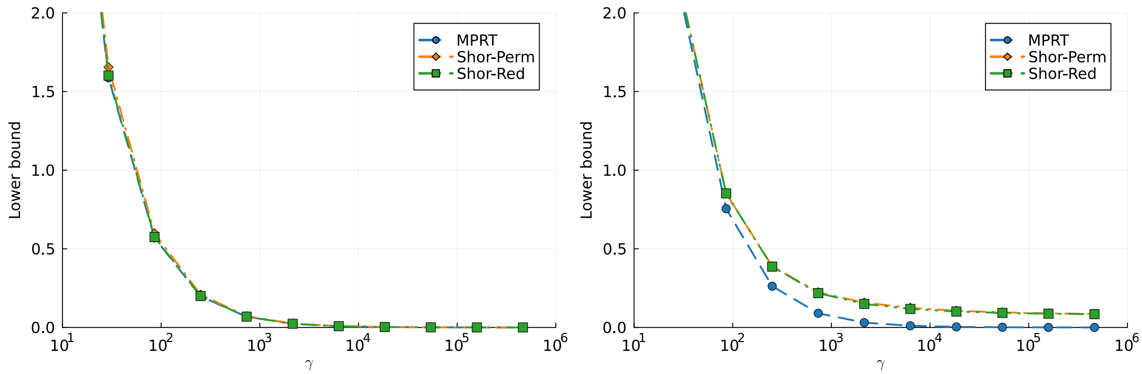


Figure EC.2 Empirical distribution of $\|\alpha\,\mathrm{vec}(\boldsymbol{G})\,\mathrm{vec}(\boldsymbol{G})^\top - \mathrm{vec}(\boldsymbol{Q})\,\mathrm{vec}(\boldsymbol{Q})^\top\|_F$ for different values of m. The solid vertical line indicates the value of m.

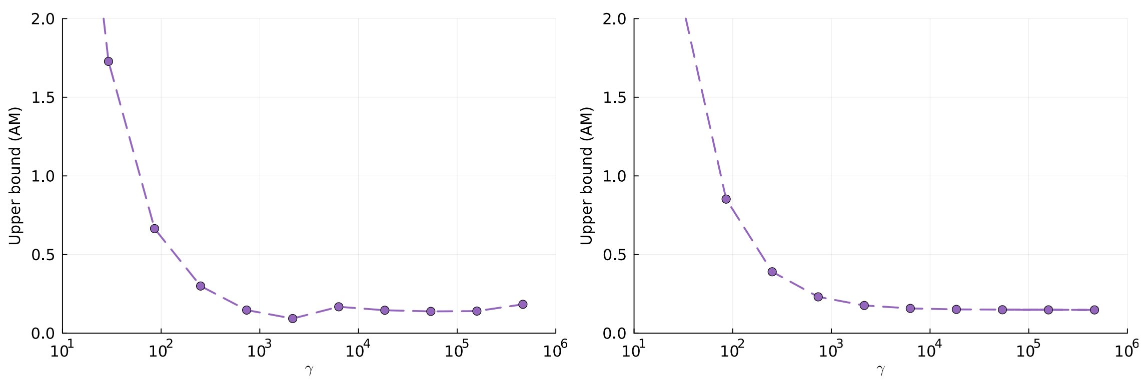
EC.5.2. Additional Results for Low-Rank Matrix Completion

Figure 2 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 one upper bound (obtained by alternating minimization, AM). Figure EC.3 and EC.4 report the lower and upper bounds separately.



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

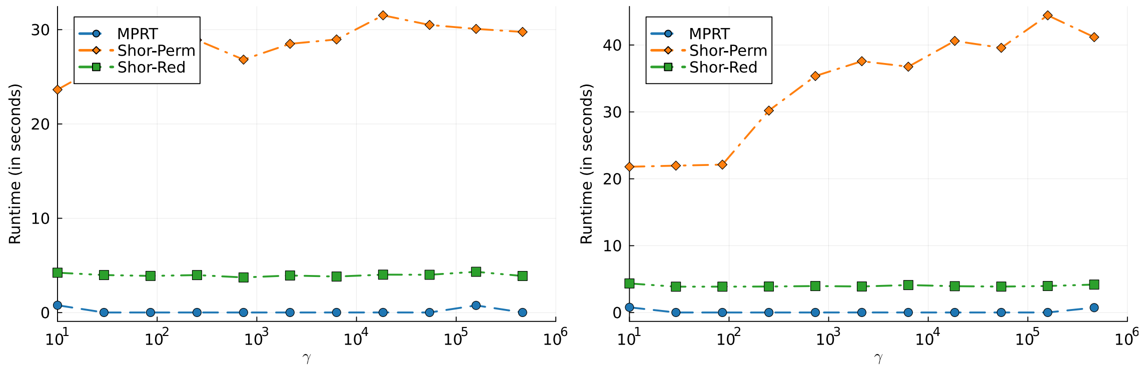
Figure EC.3 Absolute lower bounds as we vary γ for (a) a matrix perspective relaxation ("MPRT"), (b) our Shor relaxation with permutation equalities ("Shor-Perm"), (c) our compact Shor relaxation with no permutation equalities ("GW-Red"), for $p \in \{0.5, 0.95\}$ and $n = 8$.



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

Figure EC.4 Absolute upper bounds as we vary γ for the alternating minimization method of Burer and Monteiro (2003) initialized at a rank-r SVD of $\mathcal{P}(\boldsymbol{A})$ for $p \in \{0.5, 0.95\}$ and $n = 8$.

Figure EC.5 compares the same three relaxations in terms of computational time.



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

Figure EC.5 Runtimes for (a) a matrix perspective relaxation ("MPRT"), (b) our Shor relaxation with permutation equalities ("Shor-Perm"), (c) our Shor relaxation with no permutation equalities ("GW-Red"), for $p \in \{0.5, 0.95\}$, $n = 8$, and increasing γ.