# Thesis Summary Ryan Cory-Wright

## *Integer and Matrix Optimization: A Nonlinear Approach*

Many important problems from the operations research, machine learning, and statistics literature exhibit either (a) logical relations between continuous variables $x$ and binary variables $z$ of the form "$x = 0$ if $z = 0$", or (b) rank constraints. Indeed, start-up costs in machine scheduling, financial transaction costs, and cardinality constraints exhibit logical relations. Moreover, important problems such as factor analysis and matrix completion, which model notions of minimal complexity, low dimensionality, or orthogonality in a system, contain rank constraints. These constraints are viewed as separate entities and studied by separate subfields—integer and global optimization—that propose different strategies for optimizing over them.

Since the work of Glover [12], logical relations have been well studied by the integer optimization community. They are typically enforced through a "big-$M$" constraint of the form $-Mz \leq x \leq Mz$ for a large constant $M > 0$, and optimized over via branch-and-bound or branch-and-cut. Glover's work has been so influential that big-$M$ constraints are now considered as intrinsic components of the problem formulations themselves, to the extent that textbooks in the field introduce facility location, network design or sparse portfolio problems with big-$M$ constraints *by default* [see, e.g., 2], although they are actually *reformulations* of logical constraints.

On the other hand, rank constraints are commonly regarded as intractable by the global optimization and machine learning communities, since they cannot be represented using mixed-integer convex optimization [15], and there did not exist any generic codes that solve low-rank optimization problems to certifiable optimality at even moderate problem sizes at the time this thesis was written (although we have since designed one using the techniques described in this thesis [3]). This state of affairs has led influential works on low-rank optimization such as [5] to characterize low-rank optimization as intractable and advocate convex relaxations or heuristics that do not enjoy assumption-free optimality guarantees.

In this thesis, we adopt a different perspective on logical and rank constraints. We interpret both constraints as purely algebraic ones: logical constraints are nonlinear constraints on a continuous scalar variable $x$ of the form $x = z \circ x$ for $x$ continuous and $z$ binary (meaning $z^2 = z$), while rank constraints, $\mathrm{Rank}(\boldsymbol{X}) \leq k$, are a nonlinear constraint on a matrix $\boldsymbol{X} \in \mathbb{R}^{n \times m}$ of the form $\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}$ intersected with a linear constraint $\mathrm{tr}(\boldsymbol{Y}) \leq k$ for an orthogonal projection matrix $\boldsymbol{Y}$ (meaning $\boldsymbol{Y} \in \mathcal{S}^n, \boldsymbol{Y}^2 = \boldsymbol{Y}$). Under this lens, we show that both constraints can be modeled in the same way, and thus optimized over using the same techniques (i.e., greedy rounding for provably near-optimal solutions, and branch-and-cut for certifiably optimal solutions). This reveals that while sparsity constraints and rank constraints arise in different applications, and are addressed by different research communities using different algorithms, they are two different facets of the same unified story. In particular, algorithms that have been known to the mixed-integer community for almost $50$ years can, when adapted appropriately, solve both sparsity and rank-constrained problems more accurately and more efficiently than algorithms that are currently considered to be state-of-the-art; see Table $1$ for a summary of the connections between logical and rank constraints made throughout the thesis.

The thesis comes in two parts: in the first part—consisting of Chapters $2$–$4$—we explore how our framework applies to logically constrained problems, and undertake detailed studies of its scalability for sparse portfolio selection and sparse principal component analysis problems. In the second part—consisting of Chapters $5$–$6$—we explore its implications for low-rank problems, by proposing a technique for solving low-rank problems exactly and undertaking a detailed study of the convex relaxations of a class of low-rank problems.

In the rest of this document, we provide a summary of the publications that have arisen as a result of this thesis, and provide a chapter-by-chapter summary of the main contributions of the thesis.

**Table** 1: Analogy between mixed-integer and mixed-projection optimization derived throughout thesis.

| Framework | Chapters 2–4 | Chapters 5–6 |
|---|---|---|
| Parsimony concept<br>Non-convex set | Cardinality ($\|\boldsymbol{x}\|_0 \leq k$ for $\boldsymbol{x} \in \mathbb{R}^n$)<br>Binaries ($\exists z \in \mathbb{R} : z^2 = z$) | Rank ($\mathrm{Rank}(\boldsymbol{Y}) \leq k$ for $\boldsymbol{X} \in \mathbb{R}^{m \times n}$)<br>Projection matrices ($\exists \boldsymbol{Y} \in \mathcal{S}^n : \boldsymbol{Y}^2 = \boldsymbol{Y}$) |
| Non-linear formulation | $\|\boldsymbol{x}\|_0 \leq k \iff \exists \boldsymbol{z} : \boldsymbol{z} \circ \boldsymbol{z} = \boldsymbol{z}$,<br>$\boldsymbol{x} = \boldsymbol{z} \circ \boldsymbol{x}, \boldsymbol{e}^\top \boldsymbol{z} \leq k$ | $\mathrm{Rank}(\boldsymbol{X}) \leq k \iff \exists \boldsymbol{Y} : \boldsymbol{Y}^2 = \boldsymbol{Y}$,<br>$\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}, \mathrm{tr}(\boldsymbol{Y}) \leq k$ |
| Big-M reformulation<br>(for some large $M > 0$) | $\|\boldsymbol{x}\|_0 \leq k, \|\boldsymbol{x}\|_\infty \leq M \iff$<br>$\exists \boldsymbol{z} : \boldsymbol{z} \circ \boldsymbol{z} = \boldsymbol{z}$,<br>$-M\boldsymbol{z} \leq \boldsymbol{x} \leq M\boldsymbol{z}, \boldsymbol{e}^\top \boldsymbol{z} \leq k$ | $\mathrm{Rank}(\boldsymbol{X}) \leq k, \|\boldsymbol{X}\|_\sigma \leq M \iff$<br>$\exists \boldsymbol{Y} : \boldsymbol{Y}^2 = \boldsymbol{Y}$,<br>$\begin{pmatrix} M\boldsymbol{Y} & \boldsymbol{X} \\ \boldsymbol{X}^\top & M\mathbb{I} \end{pmatrix} \succeq \boldsymbol{0}, \mathrm{tr}(\boldsymbol{Y}) \leq k$ |
| Perspective reformulation<br>(for some $\gamma > 0$) | $\|\boldsymbol{x}\|_0 \leq k, \frac{1}{2\gamma}\|\boldsymbol{x}\|_2^2 \leq \rho \iff$<br>$\exists \boldsymbol{z}, \boldsymbol{\theta} : \boldsymbol{z} \circ \boldsymbol{z} = \boldsymbol{z}, \boldsymbol{e}^\top \boldsymbol{\theta} \leq \rho$,<br>$\begin{pmatrix} \theta_i & x_i \\ x_i & z_i \end{pmatrix} \succeq \boldsymbol{0}\ \forall i, \boldsymbol{e}^\top \boldsymbol{z} \leq k$ | $\mathrm{Rank}(\boldsymbol{X}) \leq k, \frac{1}{2\gamma}\|\boldsymbol{X}\|_F^2 \leq \rho \iff$<br>$\exists \boldsymbol{Y}, \boldsymbol{\theta} : \boldsymbol{Y}^2 = \boldsymbol{Y}, \mathrm{tr}(\boldsymbol{\theta}) \leq \rho$,<br>$\begin{pmatrix} \boldsymbol{\theta} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{Y} \end{pmatrix} \succeq \boldsymbol{0}, \mathrm{tr}(\boldsymbol{Y}) \leq k$ |
| Applications | Sparse regression, portfolio selection, | Rank regression, matrix completion |

**A More General Story:** The framework proposed in this thesis exhibits an interesting connection to the theory of Euclidean Jordan algebras. To model both sparsity and rank constraints, we work with idempotent elements $z_i$ embedded in a Jordan algebra such that $z_i^2 = z_i$, $z_i z_j = 0$ if $i \neq j$ and $\sum_{i=1}^n z_i = e$, where $e$ denotes an identity of appropriate dimension; see [7] for an introduction to analysis over symmetric cones. We can interpret both the cardinality of a vector $\boldsymbol{x}$ and the rank of a matrix $\boldsymbol{X}$ as a special case of the Jordan-algebraic rank, i.e., the minimum number of idempotents required to provide a spectral decomposition $\boldsymbol{x} = \sum_{i=1}^n \lambda_i z_i$ for a binary unit vector $z_i$ or $\boldsymbol{X} = \sum_{i=1}^n \lambda_i \boldsymbol{Y}_i$ for projection matrices $\boldsymbol{Y}_i$. This suggests this thesis could be extended to model other notions of Jordan-algebraic rank, e.g., the CP-rank of a tensor.

**Related Publications:**

The contents of this thesis are based on the following published papers (authors ordered alphabetically):

- **Chapter 2:** D. Bertsimas, R. Cory-Wright, and J. Pauphilet. A unified approach to mixed-integer optimization problems with logical constraints. *SIAM Journal on Optimization* 31(3), 2340-2367, 2021.
- **Chapter 3:** D. Bertsimas and R. Cory-Wright. A scalable algorithm for sparse portfolio selection. *INFORMS Journal on Computing* 34(3), 1489-1511, 2022.
- **Chapter 4:** D. Bertsimas, R. Cory-Wright, and J. Pauphilet. Solving Large-Scale Sparse PCA to Certifiable (Near) Optimality. *Journal of Machine Learning Research* 23, 13:1-13:35, 2022.
- **Chapter 5:** D. Bertsimas, R. Cory-Wright, and J. Pauphilet. Mixed-projection conic optimization: A new paradigm for modeling rank constraints. *Operations Research* 70(6), 3321-3344, 2022.
- **Chapter 6:** D. Bertsimas, R. Cory-Wright, and J. Pauphilet. A new perspective on low-rank optimization. *Mathematical Programming*, 202:47–92, 2023.

In addition, the following preprints and publications have arisen partly as follow-up work to this thesis:

- D. Bertsimas and R. Cory-Wright. On polyhedral and second-order cone decompositions of semidefinite optimization problems. *Operations Research Letters*, 48(1), 78-85, 2023.
- D. Bertsimas, R. Cory-Wright, and N. A. G. Johnson. Sparse Plus Low Rank Matrix Decomposition: A Discrete Optimization Approach. *Journal of Machine Learning Research*, 24: 1–51, 2023.
- D. Bertsimas, R. Cory-Wright, S. Lo, and J. Pauphilet. Optimal Low-Rank Matrix Completion: Semidefinite Relaxations and Eigenvector Disjunctions. arXiv:2305.12292, 2023. (under review at *Operations Research*)
- D. Bertsimas, R. Cory-Wright, J. Pauphilet, and P. Petridis. A Stochastic Benders Decomposition Scheme for Large-Scale Data-Driven Network Design. arXiv:2303.07695, 2023. (major revision at *IJOC*)
- R. Cory-Wright and J. Pauphilet. Sparse PCA With Multiple Components. arXiv:2209.14790.v2, 2023. (under review at *Operations Research*).

We are currently synthesizing this thesis with the aforementioned follow-up publications and some other papers authored by Dimitris Bertsimas and Jean Pauphilet into a book with the same name as this thesis.

## Main Contributions of Thesis: Chapter-by-Chapter

We now provide an overview of each chapter and its main contributions.

**Chapter 2: A Unified Approach to Mixed-Integer Optimization Problems With Logical Constraints** We consider optimization problems that unfold over two stages. In the first stage, a decision-maker activates binary variables, while satisfying resource budget constraints and incurring activation costs. In the second stage, the decision-maker optimizes over the continuous variables. Formally, we consider:

$$\min_{\boldsymbol{z}\in\mathcal{Z}} f(\boldsymbol{z}) \text{ where } f(\boldsymbol{z}) := \min_{\boldsymbol{x}\in\mathbb{R}^n} \; \boldsymbol{c}^\top \boldsymbol{z} + g(\boldsymbol{x}) + \Omega(\boldsymbol{x}) \text{ s.t. } x_i = 0 \text{ if } z_i = 0,\ \forall i \in [n], \tag{1}$$

where $\mathcal{Z} \subseteq \{0,1\}^n$, $\boldsymbol{c} \in \mathbb{R}^n$ is a cost vector, $g(\cdot)$ is a generic convex function, and $\Omega(\cdot)$ is either: (a) a big-$M$ penalty, $\Omega(\boldsymbol{x}) = 0$ if $\|\boldsymbol{x}\|_\infty \leq M$ and $\infty$ otherwise or (b) a ridge penalty, $\Omega(\boldsymbol{x}) = \frac{1}{2\gamma}\|\boldsymbol{x}\|_2^2$ (one could also consider other separable convex penalties, as discussed in Chapter $6$). This problem arises in many important applications, including facility location [8] and network design [6]. In contrast to previous approaches, we reformulate the constraint "$x_i = 0$ if $z_i = 0$" in a non-linear way, by substituting $z_i x_i$ for $x_i$ in (1), and invoking strong duality. This supplies a tractable reformulation of (1):

$$\min_{\boldsymbol{z}\in\mathcal{Z}} f(\boldsymbol{z}) \text{ where } f(\boldsymbol{z}) := \max_{\boldsymbol{\alpha}\in\mathbb{R}^n} \quad \boldsymbol{c}^\top \boldsymbol{z} + h(\boldsymbol{\alpha}) - \sum_{i=1}^n z_i\, \Omega^\star(\alpha_i), \tag{2}$$

where $h(\boldsymbol{\alpha}) := \inf_{\boldsymbol{v}} g(\boldsymbol{v}) - \boldsymbol{v}^\top \boldsymbol{\alpha}$ is, up to a sign, the Fenchel conjugate of $g$, and $\Omega^\star$ is either (a) $\Omega^\star(\beta) = M|\beta|$ under the big-$M$ penalty or (b) $\Omega^\star(\beta) = \frac{\gamma}{2}\beta^2$ under the ridge penalty.

This non-linear reformulation allows us to make significant progress in solving (1):

- Reformulating $f(\boldsymbol{z})$ as a maximization problem proves that $f(\boldsymbol{z})$ is convex in $\boldsymbol{z}$. Consequently, we can minimize $f(\boldsymbol{z})$ to provable optimality by iteratively constructing a piecewise linear outer approximation.
- From the Boolean relaxation of the problem, we obtain valid lower approximations of $f(\boldsymbol{z})$ to initialize the cutting-plane algorithm, and an optimal continuous solution. We provide theoretical guarantees on the integrality gap, the performance of a randomized rounding scheme, and propose numerically efficient methods to solve the Boolean relaxation.

**Chapter 3: A Scalable Algorithm for Sparse Portfolio Selection** We provide two main contributions. First, we propose augmenting sparse portfolio selection problems with a ridge regularization term. This yields a more practically tractable problem for two reasons. First, the duality gap between a sparse porfolio selection problem and its second-order cone relaxation decreases as we increase the amount of regularization and becomes $0$ with a sufficiently large but finite amount of regularization. Second, as we numerically establish in computational experiments, the algorithms developed in this chapter converge more rapidly with more regularization. Our second main contribution is specializing the outer-approximation method developed in the previous chapter to sparse portfolio selection problems, and demonstrating that this allows us to solve sparse portfolio selection problems with up to $3,200$ securities to certifiable optimality in hundreds or thousands of seconds, one order of magnitude larger than previous attempts as summarized in the below table:

Table $2$: Largest sparse portfolio instances reliably solved by each approach

| Reference | Solution method | Largest instance size solved |
|---:|---|---:|
| Frangioni and Gentile [10] | Perspective cut+SDP | $400$ |
| Bonami and Lejeune [4] | Nonlinear B&B | $200$ |
| Gao and Li [11] | Lagrangian relaxation B&B | $300$ |
| Zheng et al. [16] | Branch-and-Cut+SDP | $400$ |
| Chapter $3$ | Outer approximation with ridge regularization | $3,200$ |

We now compare big-$M$ and ridge regularization for a sparse portfolio selection problem on the securities contained in the Russell $1000$, with a sparsity budget of $k = 5$ in Figure 1. We depict the relationship between an optimal allocation of funds $\boldsymbol{x}^\star$ and the regularization parameter $M$ (left) and $\gamma$ (right). Qualitatively, both parameters impact the investment profiles comparably and eventually select the same stocks. Yet, we observe two main differences: First, setting $M < \frac{1}{k}$ renders the entire problem infeasible, while the problem remains feasible for any $\gamma > 0$. This is a serious concern when a lower bound on $M$ is not known. Second, the profile for ridge regularization seems smoother than its equivalent with big-$M$.

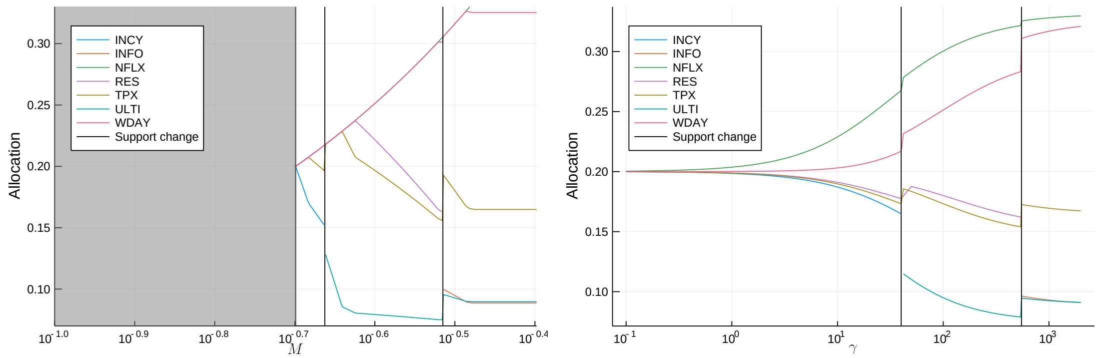


**Figure** $1$**:** Optimal allocation of funds between securities as the regularization parameter ($M$ or $\gamma$) increases.

**Chapter** $4$**: Solving Large-Scale Sparse PCA to Certifiable (Near) Optimality** We provide two main contributions. First, we reformulate sparse PCA exactly as a mixed-integer semidefinite optimization problem; a reformulation which is, to the best of our knowledge, novel (as of the time this chapter was first submitted for publication). This reformulation is significant because sparse PCA is traditionally treated as a low-rank optimization problem, which as we show in the next chapter is a harder class of problems to address. Second, we leverage this MISDO formulation to design efficient algorithms for solving non-convex quadratic integer problems, such as sparse PCA, to certifiable optimality or within $1-2\%$ of optimality in practice at a larger scale than existing state-of-the-art methods. A key feature of our approach is that we sidestep the computational difficulties in solving SDOs by solving second-order cone relaxations as in [14, 1].

**Chapter** $5$**: Mixed-Projection Conic Optimization: A New Paradigm for Modeling Rank Constraints** We consider a rank-penalized rank-constrained conic optimization problem:

$$\min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}} \lambda \cdot \mathrm{Rank}(\boldsymbol{X}) + \langle \boldsymbol{C}, \boldsymbol{X} \rangle \ : \boldsymbol{A}\boldsymbol{X} = \boldsymbol{B}, \mathrm{Rank}(\boldsymbol{X}) \leq k, \boldsymbol{X} \in \mathcal{K}, \tag{3}$$

where $\mathcal{K}$ is a proper cone, and $\lambda, k \geq 0$ are parameters. First, we observe that the rank constraint can be formulated in terms of a trace constraint on a projection matrix $\boldsymbol{Y}$ that spans the column space of $\boldsymbol{X}$, i.e.,

$$\mathrm{Rank}(\boldsymbol{X}) \leq k \quad \iff \exists \boldsymbol{Y} \in \mathcal{S}^n : \ \boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}, \boldsymbol{Y} = \boldsymbol{Y}^2, \mathrm{tr}(\boldsymbol{Y}) \leq k. \tag{4}$$

We then observe that the projection matrix $\boldsymbol{Y}$—which satisfies $\boldsymbol{Y}^2 = \boldsymbol{Y}$—can be viewed as the matrix analog of a binary variable which satisfies $z^2 = z$. Letting $\mathcal{Y}^k_n = \{\boldsymbol{Y} \in S^n : \boldsymbol{Y} = \boldsymbol{Y}^2, \mathrm{tr}(\boldsymbol{Y}) \leq k\}$ denote the set of symmetric rank-$k$ projections on $\mathbb{R}^{n \times n}$, we obtain a mixed-projection reformulation of rank optimization:

$$\min_{\boldsymbol{Y} \in \mathcal{Y}^k_n} f(\boldsymbol{Y}) \text{ where } f(\boldsymbol{Y}) = \min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}} \ \lambda \cdot \mathrm{tr}(\boldsymbol{Y}) + \langle \boldsymbol{C}, \boldsymbol{X} \rangle + \Omega(\boldsymbol{X}) : \boldsymbol{A}\boldsymbol{X} = \boldsymbol{B}, \boldsymbol{Y}\boldsymbol{X} = \boldsymbol{X}, \boldsymbol{X} \in \mathcal{K}. \tag{5}$$

As in MIO, we impose a regularizer $\Omega$ to make the problem tractable. It's either: (a) a spectral penalty, $\Omega(\boldsymbol{X}) = 0$ if $\|\boldsymbol{X}\|_\sigma \leq M$ and $\infty$ otherwise or (b) a Frobenius penalty, $\Omega(\boldsymbol{X}) = (\frac{1}{2\gamma})\|\boldsymbol{X}\|_F^2$. Taking the dual of the inner maximization problem $(3)$ then supplies a tractable reformulation of the constraint $\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}$:

$$\min_{\boldsymbol{Y} \in \mathcal{Y}^k_n} f(\boldsymbol{Y}) \text{ where } f(\boldsymbol{Y}) = \max_{\boldsymbol{\alpha}, \boldsymbol{\Pi}} \ \lambda \cdot \mathrm{tr}(\boldsymbol{Y}) + \langle \boldsymbol{B}, \boldsymbol{\Pi} \rangle - \Omega^\star(\boldsymbol{\alpha}, \boldsymbol{Y}) \text{ s.t. } \boldsymbol{C} - \boldsymbol{\alpha} - \boldsymbol{A}^\top \boldsymbol{\Pi} \in \mathcal{K}^\star, \tag{6}$$

where $\Omega^\star$—the conjugate of $\Omega$—is linear in $\boldsymbol{Y}$. Problem $(4)$ allows us to make significant progress on $(1)$:

- Reformulating $f(\boldsymbol{Y})$ as a maximization problem proves $f(\boldsymbol{Y})$ is convex in $\boldsymbol{Y}$. Consequently, we can minimize $(4)$ to provable optimality by iteratively constructing a piecewise linear lower approximation and minimizing the approximation over $\mathcal{Y}_n^k$ using spatial branch-and-bound.
- We solve matrix completion problems with $10$s of features to certifiable optimality, and with $100$s of features to near optimality. To our knowledge, this is the first work to solve low-rank problems to optimality.

**Chapter $6$: A New Perspective on Low-Rank Optimization** The main contributions of this chapter are twofold. First, we propose a general reformulation technique for obtaining high-quality relaxations of low-rank optimization problems: introducing an orthogonal projection matrix to model a low-rank constraint, and strengthening the formulation by taking the matrix perspective of an appropriate substructure of the problem. Specifically, given an optimization problem of the form:

$$\min_{\boldsymbol{Y}\in\mathcal{Y}_n^k}\min_{\boldsymbol{X}\in\mathcal{S}^n} \quad \langle \boldsymbol{C},\boldsymbol{X}\rangle + \mu\cdot\mathrm{tr}(\boldsymbol{Y}) + \mathrm{tr}(f(\boldsymbol{X})) \text{ s.t. } \langle \boldsymbol{A}_i,\boldsymbol{X}\rangle = b_i \quad \forall i\in[m],\ \boldsymbol{X}=\boldsymbol{Y}\boldsymbol{X},\ \boldsymbol{X}\in\mathcal{K}, \tag{7}$$

where $\mathcal{K}$ is a proper cone, $\boldsymbol{Y}\in\mathcal{Y}_n^k$ is the set of $n\times n$ orthogonal projection matrices with trace at most $k$:

$$\mathcal{Y}_n^k := \left\{\boldsymbol{Y}\in\mathcal{S}_+^n : \boldsymbol{Y}^2=\boldsymbol{Y},\ \mathrm{tr}(\boldsymbol{Y})\le k\right\},$$

and we let the orthogonal projection matrix $\boldsymbol{Y}$ model the rank of $\boldsymbol{X}$ as discussed in Chapter $5$. Then, since $\boldsymbol{Y}$ is an orthogonal projection matrix, imposing the nonlinear constraint $\boldsymbol{X}=\boldsymbol{Y}\boldsymbol{X}$ and introducing the term $\Omega(\boldsymbol{X})=\mathrm{tr}(f(\boldsymbol{X}))$ in the objective is equivalent to introducing the term $\mathrm{tr}(g_f(\boldsymbol{X},\boldsymbol{Y}))+(n-\mathrm{tr}(\boldsymbol{Y}))\omega(0)$ in the objective, where $g_f$ is the matrix perspective of $f$, and thus Problem (7) is equivalent to:

$$\begin{aligned}\min_{\boldsymbol{Y}\in\mathcal{Y}_n^k}\min_{\boldsymbol{X}\in\mathcal{S}^n} \quad & \langle \boldsymbol{C},\boldsymbol{X}\rangle + \mu\cdot\mathrm{tr}(\boldsymbol{Y}) + \mathrm{tr}(g_f(\boldsymbol{X},\boldsymbol{Y})) + (n-\mathrm{tr}(\boldsymbol{Y}))\omega(0)\\ \text{s.t.} \quad & \langle \boldsymbol{A}_i,\boldsymbol{X}\rangle = b_i \quad \forall i\in[m],\ \boldsymbol{X}\in\mathcal{K}.\end{aligned}$$

Moreover, relaxing $\mathcal{Y}_n^k$ to its convex hull $\mathrm{Conv}(\mathcal{Y}_n^k)=\{\boldsymbol{Y}\in\mathcal{S}_+^n : \boldsymbol{Y}\preceq\mathbb{I}, \mathrm{tr}(\boldsymbol{Y})\le k\}$ immediately yields a valid semidefinite relaxation of a low-rank problem. We term this technique the "Matrix Perspective Reformulation Technique" in recognition of the fact that we impose the rank constraint implicitly via the domain of the matrix convex function $f$'s matrix perspective function, which is defined as

$$g_{f_\omega}(\boldsymbol{X},\boldsymbol{Y}) = \begin{cases}\boldsymbol{Y}^{\frac{1}{2}} f_\omega(\boldsymbol{Y}^{-\frac{1}{2}}\boldsymbol{X}\boldsymbol{Y}^{-\frac{1}{2}})\boldsymbol{Y}^{\frac{1}{2}} & \text{if } \mathrm{Span}(\boldsymbol{X})\subseteq\mathrm{Span}(\boldsymbol{Y}), \boldsymbol{Y}\succeq\boldsymbol{0},\\ \infty & \text{otherwise,}\end{cases}$$

We remark that this technique is a generalization of the perspective reformulation technique—a popular preprocessing technique for obtaining tight relaxations of mixed-integer problems [9, 13].

Second, by applying this technique, we obtain explicit characterizations of convex hulls of low-rank sets which frequently arise in low-rank problems. In particular, we are able to establish the following result:

**Main Theoretical Contribution:** *Consider a matrix convex operator function $f=f_\omega$, which satisfies some mild regularity conditions (see Chapter $6$), and let $\omega$ be the scalar function associated with $f_\omega$. Let*

$$\mathcal{T} = \{\boldsymbol{X}\in\mathcal{S}^n : \mathrm{tr}(f(\boldsymbol{X})) + \mu\cdot\mathrm{Rank}(\boldsymbol{X}) \le t, \mathrm{Rank}(\boldsymbol{X})\le k\} \tag{8}$$

*be a set where $t\in\mathbb{R}, k\in\mathbb{N}$ are fixed. Then, an extended formulation of the convex hull of $\mathcal{T}$ is given by:*

$$\mathcal{T}^c = \left\{(\boldsymbol{X},\boldsymbol{Y})\in\mathcal{S}^n\times\mathrm{Conv}(\mathcal{Y}_n^k) : \mathrm{tr}(g_f(\boldsymbol{X},\boldsymbol{Y})) + \mu\cdot\mathrm{tr}(\boldsymbol{Y}) + (n-\mathrm{tr}(\boldsymbol{Y}))\omega(0)\le t\right\},$$

*where $\mathrm{Conv}(\mathcal{Y}_n^k)=\{\boldsymbol{Y}\in\mathcal{S}_+^n : \boldsymbol{Y}\preceq\mathbb{I}, \mathrm{tr}(\boldsymbol{Y})\le k\}$ is the convex hull of trace-$k$ projection matrices, and $g_f$ is the matrix perspective function of $f$.*

As the interplay between convex hulls of indicator sets and perspective functions has engineered algorithms that outperform state-of-the-art heuristics in sparse linear regression and sparse portfolio selection (see Chapters 2–3), we hope that this work will empower similar developments for low-rank problems.

## Publications


[1] Dimitris Bertsimas and Ryan Cory-Wright. On polyhedral and second-order cone decompositions of semidefinite optimization problems. *Operations Research Letters*, 48(1):78–85, 2020.

[2] Dimitris Bertsimas and Robert Weismantel. *Optimization over integers*. Dynamic Ideas Belmont, MA, 2005.

[3] Dimitris Bertsimas, Ryan Cory-Wright, Sean Lo, and Jean Pauphilet. Optimal low-rank matrix completion: Semidefinite relaxations and eigenvector disjunctions. *arXiv preprint arXiv:2305.12292*, 2023.

[4] Pierre Bonami and Miguel A Lejeune. An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Operations Research*, 57(3):650–670, 2009.

[5] Emmanuel J Candes and Yaniv Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6): 925–936, 2010.

[6] Teodor Gabriel Crainic, Mike Hewitt, Francesca Maggioni, and Walter Rei. Partial Benders decomposition: general methodology and application to stochastic network design. *Transportation Science*, 55(2):414–435, 2021.

[7] Jacques Faraut and A Koranyi. Analysis on symmetric cones. *Oxford, UK*, 1994.

[8] Matteo Fischetti, Ivana Ljubić, and Markus Sinnl. Redesigning Benders decomposition for large-scale facility location. *Management Science*, 63(7):2146–2162, 2017.

[9] Antonio Frangioni and Claudio Gentile. Perspective cuts for a class of convex 0–1 mixed integer programs. *Mathematical Programming*, 106(2):225–236, 2006.

[10] Antonio Frangioni and Claudio Gentile. A computational comparison of reformulations of the perspective relaxation: SOCP vs. cutting planes. *Operations Research Letters*, 37(3):206–210, 2009.

[11] Jianjun Gao and Duan Li. Optimal cardinality constrained portfolio selection. *Operations Research*, 61 (3):745–761, 2013.

[12] Fred Glover. Improved linear integer programming formulations of nonlinear integer problems. *Management Science*, 22(4):455–460, 1975.

[13] Oktay Günlük and Jeff Linderoth. Perspective reformulations of mixed integer nonlinear programs with indicator variables. *Mathematical Programming*, 124:183–205, 2010.

[14] Georgina Hall. *Optimization over nonnegative and convex polynomials with and without semidefinite programming*. PhD thesis, Princeton University, 2018.

[15] Miles Lubin, Ilias Zadik, and Juan Pablo Vielma. Mixed-integer convex representability. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 392–404. Springer, 2017.

[16] Xiaojin Zheng, Xiaoling Sun, and Duan Li. 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, 2014.