Probabilistically Exact Convexification of Mixed-Integer Quadratic Optimization with Indicators

Salar Fattahi[1] and Andrés Gómez[2]

[1] University of Michigan, Ann Arbor MI, USA fattahi@umich.edu
[2] University of Southern California, Los Angeles CA, USA gomezand@usc.edu

Abstract. We study the convexification of mixed-integer quadratic optimization with indicator variables. The problem remains NP-hard even when the quadratic matrix has bandwidth two, posing a fundamental obstacle to tractable exact convexification. To overcome this challenge in realistic settings, we introduce *probabilistically exact (inner) convexification*, a framework that relaxes the requirement of exactness for every objective while preserving exact optimality for *statistically relevant ones*. Our approach is motivated by inference applications, where the objective is determined by a data vector drawn from a distribution and a regularization vector chosen from a prescribed family. Our convexification preserves the exactness of the optimal value with high probability over the data distribution, simultaneously for every regularization vector in this family. We apply this notion to quadratic matrices of any fixed bandwidth. Despite their worst-case NP-hardness, we construct probabilistically exact convexifications that admit exact extended second-order cone formulations of size polynomial in the dimension n and the inverse failure probability $1/\delta$. The required assumptions for their high-probability exactness are surprisingly mild: they accommodate both light-tailed and heavy-tailed data, including distributions without finite mean, as well as arbitrary nonnegative regularization coefficients.

Keywords: Mixed-integer optimization · Convexification · Dynamic programming.

1 Introduction

Mixed-integer quadratic optimization with indicator variables provides a natural framework for statistical estimation with logical decisions. Given a positive definite matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, we consider optimization problems of the form

$$\min_{(t,\boldsymbol{x},\boldsymbol{z})\in\mathcal{P}} \left\{t + \boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\beta}^\top \boldsymbol{z}\right\}, \tag{1}$$

$$\text{where} \quad \mathcal{P} := \left\{(t,\boldsymbol{x},\boldsymbol{z}) \in \mathbb{R}_+ \times \mathbb{R}^n \times \{0,1\}^n : t \geq \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}, \quad \boldsymbol{x} \circ (\boldsymbol{1} - \boldsymbol{z}) = \boldsymbol{0}\right\}, \tag{2}$$

and $\circ$ denotes the entrywise product. The indicator constraints $\boldsymbol{x} \circ (\boldsymbol{1} - \boldsymbol{z}) = \boldsymbol{0}$ are equivalent to the logical condition $x_i \neq 0 \implies z_i \neq 0$ for all $i \in [n]$; such constraints are used to impose logical conditions on the support of the continuous variables $\boldsymbol{x}$.

An application particularly relevant to this paper is sparse inference of spatio-temporal processes [2]. Given noisy observations $\boldsymbol{\alpha} \in \mathbb{R}^n$ and a sparse positive definite matrix

$\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, a representative formulation is [1, 8, 9]

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{z}\in\{0,1\}^n} \quad & \frac{1}{2}\|\boldsymbol{\alpha}-\boldsymbol{x}\|_2^2 + \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \beta \mathbf{1}^\top \boldsymbol{z} \\ \text{s.t.} \quad & \boldsymbol{x} \circ (\mathbf{1}-\boldsymbol{z}) = \mathbf{0}. \end{aligned} \tag{3}$$

Variables $\boldsymbol{x}$ represent estimates of the underlying process, while the indicators z_i determine which signal components x_i may be nonzero. The parameter $\beta \geq 0$ penalizes the number of active components, with $\mathbf{1}^\top \boldsymbol{z}$ counting the nonzero signal components at optimality. Finally, $\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$ regularizes the signal according to its dependence structure, where $Q_{ij} \neq 0$ indicates that components i and j are spatially or temporally related. In particular, $\boldsymbol{Q}$ is typically banded in time-series models [13, 10].

Convexification-based approaches to solving (1) exploit convex formulations of the underlying mixed-integer set [18, 19]. An exact characterization of the convex hull $\text{conv}(\mathcal{P})$ yields an equivalent convex reformulation of (1). For a fixed quadratic matrix $\boldsymbol{Q}$, this formulation can be reused across different (data) vectors $\boldsymbol{\alpha}$ and (hyperparameter) vectors $\boldsymbol{\beta}$. When tractable, it therefore facilitates repeated optimization and cross-validation in statistical applications. However, obtaining a tractable exact convexification of $\mathcal{P}$ is challenging, and available results rely on additional structural assumptions. For example, tridiagonal quadratics admit exact formulations of quadratic size [13, 10], while decision diagrams provide a framework for constructing exact and approximate formulations for broader classes of structured quadratics [6]. Despite these advances, tractable exact convexification remains elusive in general: problem (1) remains NP-hard even when $\boldsymbol{Q}$ is banded with bandwidth two [14, Theorem 1 and Corollary 2].[3] Consequently, banded $\boldsymbol{Q}$ alone does not guarantee an exact convex reformulation that can be constructed and solved efficiently.

To alleviate the burdens associated with exact convexifications, which guarantee exactness for *all* possible data vectors and regularization parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, we ask whether it is possible to construct simpler formulations that are exact only for the *relevant statistical inputs*. Indeed, in statistical applications, the regularization parameters $\boldsymbol{\beta}$ are typically nonnegative. Moreover, the data vector $\boldsymbol{\alpha}$ is often modeled as a random draw from a well-behaved distribution. Motivated by these observations, we show that, when $\boldsymbol{Q}$ is banded, it is possible to construct compact (inner) convex approximations of the closure of $\text{conv}(\mathcal{P})$ that recover exact optimal solutions of (1) *with high probability* over the distribution of $\boldsymbol{\alpha}$, uniformly for all nonnegative choices of $\boldsymbol{\beta}$.

Due to space constraints, most proofs are deferred to the appendix.

2 Probabilistically Exact Approximation and Convexification

For any closed set $C \subseteq \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^n$, define its *restricted support function* $h_C : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ by

$$h_C(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \inf_{(t,\boldsymbol{x},\boldsymbol{z})\in C} \left\{ t + \boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\beta}^\top \boldsymbol{z} \right\}, \tag{4}$$

[3] The cited result appears in an AI-generated note released in the repository accompanying [12]. While the authors from the repository do not claim to have verified correctness of this note, we have independently checked the proof of the cited result and verified its correctness.

where $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$, and we adopt the convention that $h_C(\boldsymbol{\alpha}, \boldsymbol{\beta}) = +\infty$ when $C = \emptyset$. The term "restricted" refers to the fact that h_C is the restriction of the full support function to directions for which the coefficient of t is equal to one.

Definition 1 (Upward-closedness). *A set $C \subseteq \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^n$ is said to be* upward-closed (in its first element) *if C is closed and*

$$(t, \boldsymbol{x}, \boldsymbol{z}) \in C \quad \Longrightarrow \quad (t + s, \boldsymbol{x}, \boldsymbol{z}) \in C \qquad \text{for every } s \geq 0.$$

Lemma 1 (Convex hull via restricted support functions). *Suppose that $\mathcal{X} \subseteq \mathcal{P}$ is nonempty and upward-closed. Then, both* $\operatorname{conv}(\mathcal{X})$ *and* $\operatorname{conv}(\mathcal{P})$ *are closed and*

$$h_{\mathcal{X}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = h_{\mathcal{P}}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \quad \forall (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathbb{R}^n \times \mathbb{R}^n, \quad \Longleftrightarrow \quad \operatorname{conv}(\mathcal{X}) = \operatorname{conv}(\mathcal{P}). \quad (5)$$

The preceding lemma characterizes the exact convexification of $\mathcal{P}$ by equality of the restricted support functions in all directions $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Motivated by this characterization, we introduce a relaxed notion of convexification for $\mathcal{P}$: rather than requiring the restricted support function of $\mathcal{X} \subseteq \mathcal{P}$ to coincide with that of $\mathcal{P}$ for every pair $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, we require exactness only for $\boldsymbol{\beta}$ belonging to a prescribed set $\Gamma \subseteq \mathbb{R}^n$ and for a set of coefficients $\boldsymbol{\alpha}$ having sufficiently high probability under a probability measure μ.

Definition 2 (probabilistically exact approximation and convexification). *Let $\Gamma \subseteq \mathbb{R}^n$, let μ be a probability measure on $\mathbb{R}^n$, and let $\delta \in [0, 1]$. A set $\mathcal{X} \subseteq \mathcal{P}$ is called a* (μ, Γ, δ)-probabilistically exact (inner) approximation *of $\mathcal{P}$ if it is nonempty and upward-closed, and satisfies*

$$\mathbb{P}_{\boldsymbol{\alpha} \sim \mu} \left(h_{\mathcal{X}}(\alpha, \beta) = h_{\mathcal{P}}(\alpha, \beta) \quad \text{for all } \beta \in \Gamma \right) \geq 1 - \delta. \quad (6)$$

The corresponding convex set $\operatorname{conv}(\mathcal{X})$ *is called a* (μ, Γ, δ)-probabilistically exact (inner) convexification *of $\mathcal{P}$.*

The definition is tailored to realistic data-driven settings in which neither $\boldsymbol{\alpha}$ nor $\boldsymbol{\beta}$ is treated as arbitrary. In particular, inspired by the inference problem (3), the (data) vector $\boldsymbol{\alpha}$ defining the linear objective in the x-variables is modeled as a random vector with distribution μ, whereas the (regularization) vector $\boldsymbol{\beta}$ is selected from a prescribed set Γ (e.g., nonnegative vectors). The condition (6) requires that, with probability at least $1 - \delta$ over $\boldsymbol{\alpha} \sim \mu$, restricting the feasible region from $\mathcal{P}$ to $\mathcal{X}$ preserves the optimal value simultaneously for every $\boldsymbol{\beta} \in \Gamma$. Thus, $\mathcal{X}$ need only retain the portions of $\mathcal{P}$ relevant to objective coefficients that occur with sufficiently high probability under μ; exactness may fail on a set of coefficients $\boldsymbol{\alpha}$ having probability at most δ.

The key advantage of probabilistic approximation and convexification is that they can yield more tractable formulations by allowing a controlled probability of inexactness. We demonstrate this advantage for the important class of problems with banded $\boldsymbol{Q}$: although linear optimization over $\mathcal{P}$ remains NP-hard for general banded matrices, we construct a compact probabilistically exact convexification under mild assumptions.

3 Application to Banded Structures

In this section, we derive polynomial-size probabilistically exact approximation and convexification of $\mathcal{P}$ when $\boldsymbol{Q}$ is banded. For an integer $b \geq 1$, we call $\boldsymbol{Q}$ *b-banded* if $\boldsymbol{Q}_{ij} = 0$ whenever $|i - j| > b$.

Assumption 1 (Bounded spectrum and nonnegative regularization) *There are constants* $0 < \lambda_{\min} \leq \lambda_{\max} < \infty$ *and* $\beta_0 > 0$*, independent of n, such that*

$$\lambda_{\min} \boldsymbol{I} \preceq \boldsymbol{Q} \preceq \lambda_{\max} \boldsymbol{I}, \tag{7}$$

$$\Gamma \subseteq \{\mathbf{0}\} \cup [\beta_0, \infty)^n. \tag{8}$$

Assumption (8) requires positive regularization parameters to be bounded away from zero, consistent with positive penalties used to promote sparsity in $\boldsymbol{x}$. We also allow $\boldsymbol{\beta} = \mathbf{0}$ to accommodate the unregularized problem, which is commonly considered in cross-validation. We summarize our main result in the next informal theorem.

Theorem 1 (Probabilistic exactness for banded structures; informal). *Suppose that* $\boldsymbol{Q}$ *is b-banded and satisfies* (7)*,* Γ *satisfies* (8)*, and* μ *is a product measure whose coordinates satisfy:* (i) *uniformly positive probability near zero; and* (ii) *uniformly bounded absolute moments of some positive order. Then, for any* $\delta \in (0, 1]$*, there exists a* (μ, Γ, δ)*-probabilistically exact approximation* $\mathcal{X}$ *of* $\mathcal{P}$ *such that* $\mathrm{conv}(\mathcal{X})$ *is closed and admits an exact extended second-order cone programming (SOCP) formulation with*

$$\begin{cases} O\left(n\left[1 + \log\left(\frac{n}{\delta}\right)\right]^2\right), & b = 1, \\ O\left(\vartheta_b^b n\left[1 + \log\left(\frac{n}{\delta}\right)\right]\left(\frac{n}{\delta}\right)^{\gamma_b}\right), & b \geq 2, \end{cases} \tag{9}$$

variables and constraints. Here, for $b \geq 2$*,* $\vartheta_b \in (1, 2)$ *is the unique positive root of* $x^b = x^{b-1} + \cdots + x + 1$*, and* $\gamma_b = O(\log \vartheta_b)$*. In particular, for any fixed bandwidth b, the formulation has size polynomial in n and* $1/\delta$*.*

The assumptions on Γ and μ are satisfied in a broad range of practical settings. In particular, the restrictions on Γ only exclude negative regularization parameters, consistent with sparsity-inducing regularization in practice. Similarly, the conditions on μ accommodate a broad class of distributions, including Gaussian, Laplace, and bounded distributions with sufficient probability mass near zero, as well as heavy-tailed distributions such as Cauchy and Student-t. Notably, the latter class includes distributions with infinite mean.

Comparison with existing results. For tridiagonal matrices ($b = 1$), Liu et al. [13] establish an exact semidefinite programming formulation of $\mathrm{conv}(\mathcal{P})$ with $O(n^2)$ variables and linear constraints, together with a single $(n+1) \times (n+1)$ positive semidefinite constraint. Subsequently, Gomez et al. [10] derive an exact SOCP formulation with $O(n^2)$ variables and constraints. Under our distributional assumptions, we improve upon these size bounds by constructing a probabilistically exact convexification with only $O(n \log^2 n)$ variables and constraints for fixed δ and model parameters. For general banded matrices with $b \geq 2$, the situation is fundamentally different: problem (1) remains NP-hard even for $b = 2$ [14, Theorem 1 and Corollary 2], implying that bounded bandwidth alone cannot guarantee an efficiently constructible and solvable exact convex formulation unless P = NP. To address this challenge, Gomez et al. [10] and Bienstock and Chen [4] develop a deterministic approximation scheme that obtains the optimal value within an additive error ϵ, with complexity polynomial in n and $\|\boldsymbol{\alpha}\|_\infty^2/\epsilon$ for fixed bandwidth

and spectral parameters. Our approach offers a different tradeoff between tractability and accuracy: rather than allowing a bounded approximation error for every instance, we construct polynomial-size convexifications that preserve the optimal value *exactly* with probability at least $1 - \delta$ over $\boldsymbol{\alpha} \sim \mu$. Moreover, this exactness guarantee holds uniformly for all $\boldsymbol{\beta} \in \Gamma$, allowing the same formulation to be reused across different regularization choices without sacrificing optimality on the high-probability event.

3.1 Key Idea: Dynamic Programming

Our construction relies on dynamic programming (DP). Throughout the paper, we index the variables according to the row and column ordering of $\boldsymbol{Q}$, and write $[\ell : k] := \{\ell, \ell + 1, \ldots, k\}$ for integers $\ell \leq k$. A support vector $\boldsymbol{z} \in \{0,1\}^n$ is said to contain a *resetting sequence of length* b if there exists $s \in [1 : n - b + 1]$ such that $z_s = \cdots = z_{s+b-1} = 0$. Since $\boldsymbol{Q}$ is b-banded, these b consecutive zeros eliminate all quadratic interactions between variables indexed before and after the resetting sequence. This separation allows the problem to be decomposed into smaller subproblems, forming the basis of our DP construction.

To formalize this decomposition, we introduce two families of subproblems. For $s \in \{0, \ldots, n\}$, let $f_s^\star$ denote the optimal value of (1) with $z_{s+1} = \cdots = z_n = 0$. For $1 \leq \ell \leq k \leq n$, let $q^\star_{[\ell:k]}$ denote the optimal value over support vectors restricted to $[\ell : k]$ and containing no resetting sequence entirely within this interval. Formally,

$$q^\star_{[\ell:k]} = \min_{\substack{(t,\boldsymbol{x},\boldsymbol{z}) \in \mathcal{P} \\ \boldsymbol{z} \in \mathcal{N}_{[\ell:k],b}}} \left\{ t + \boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\beta}^\top \boldsymbol{z} \right\}, \tag{10}$$

where

$$\mathcal{N}_{[\ell:k],b} := \left\{ \boldsymbol{z} \in \{0,1\}^n : \begin{array}{ll} z_i = 0, & \forall i \notin [\ell : k], \\ \sum_{j=s}^{s+b-1} z_j \geq 1, & \forall s \in [\ell : k - b + 1] \end{array} \right\}. \tag{11}$$

Let $\boldsymbol{z}^\star \in \{0,1\}^n$ be an optimal support vector for $f_i^\star$, and let s denote the starting index of its last resetting sequence of length b, with $s \leq 0$ if no such sequence exists. Conditioned on the location of s, we obtain the DP recurrence

$$f_i^\star = \min_{-b+1 \leq s \leq i-b+1} \left\{ f^\star_{s-1} + q^\star_{[s+b:i]} \right\}, \qquad i \in [n], \tag{12}$$

where $f_j^\star = 0$ for $j \leq 0$ and $q^\star_{[i+1:i]} = 0$. In particular, $f_n^\star$ is the optimal value of (1); Figure 1 exemplifies this decomposition. Evidently, the efficiency of this DP hinges on the cost of computing the values $q^\star_{[\ell:k]}$. Once these are computed, the partial costs $f_1^\star, \ldots, f_n^\star$ can be obtained sequentially using (12). We distinguish two cases:

- **Tridiagonal case ($b = 1$).** For $b = 1$, $\mathcal{N}_{[\ell:k],1}$ is a singleton, with $z_i = 1$ for $i \in [\ell : k]$ and $z_i = 0$ otherwise. Thus, by (10), each $q^\star_{[\ell:k]}$ can be computed by solving a single QP. In fact, Liu et al. [13] show that all $\left\{ q^\star_{[\ell:k]} \right\}_{\ell \leq k}$ can be computed in $O(n^2)$ time. Consequently, (12) yields an $O(n^2)$-time algorithm for solving (1).

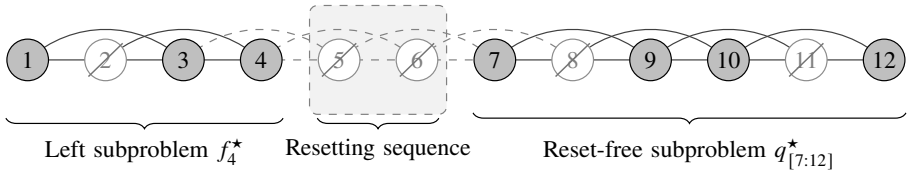


Fig. 1. Illustration of the DP recurrence for $b = 2$ and $n = 12$. The last resetting sequence, $z_5^\star = z_6^\star = 0$, separates the indices $[1:4]$ from the reset-free indices $[7:12]$.

– **General banded case ($b \geq 2$).** For $b \geq 2$, we show that $|\mathcal{N}_{[\ell:k],b}| = \Theta\left(\vartheta_b^{k-\ell+1}\right)$, where $\vartheta_b \in (1,2)$ is the unique positive root of

$$x^b = x^{b-1} + \cdots + x + 1.$$

Hence, directly evaluating each $q^\star_{[\ell:k]}$ by enumerating admissible supports in $\mathcal{N}_{[\ell:k],b}$ can require exponentially many QP solves, making this approach computationally prohibitive. This difficulty is expected in light of the NP-hardness of (1) for $b \geq 2$.

For $b \geq 2$, the computational bottleneck in (12) is evaluating $q^\star_{[s+b:i]}$ over long intervals. Our key idea is therefore to avoid these costly subproblems by restricting the feasible region to support vectors whose reset-free intervals have length at most a prescribed threshold.

Definition 3 (Reset support sets). *For integers $b \in [n]$ and $M \geq b$, define the* (M,b)-reset support set *by*

$$\mathcal{Z}_{M,b} := \left\{ \boldsymbol{z} \in \{0,1\}^n : \begin{array}{l} \textit{every block of } M \textit{ consecutive entries of } \boldsymbol{z} \\ \textit{contains a resetting sequence of length } b \end{array} \right\}. \tag{13}$$

When $M > n$, the condition is vacuous. Additionally, define the corresponding augmented (M,b)-reset support set *by*

$$\widehat{\mathcal{Z}}_{M,b} := \begin{cases} \mathcal{Z}_{M,b} \cup \{\mathbf{1}\}, & \mathbf{0} \in \Gamma, \\ \mathcal{Z}_{M,b}, & \mathbf{0} \notin \Gamma. \end{cases} \tag{14}$$

We define the associated restricted feasible set by

$$\mathcal{P}_{M,b} := \left\{ (t, \boldsymbol{x}, \boldsymbol{z}) \in \mathcal{P} : \boldsymbol{z} \in \widehat{\mathcal{Z}}_{M,b} \right\}. \tag{15}$$

Apart from the possible inclusion of $\boldsymbol{z} = \mathbf{1}$, the set $\mathcal{P}_{M,b}$ restricts $\mathcal{P}$ to support vectors for which every block of M consecutive coordinates contains a resetting sequence of length b. To formulate the corresponding DP, let $\widehat{f}_i$ and $\widehat{q}_{[\ell:k]}$ be defined as $f_i^\star$ and $q^\star_{[\ell:k]}$, respectively, with the additional constraint $\boldsymbol{z} \in \mathcal{Z}_{M,b}$. The resulting DP recurrence is

$$\widehat{f}_i = \min_{\substack{1-b \leq s \leq i-b+1 \\ |[s+b:i]| \leq M-1}} \left\{ \widehat{f}_{s-1} + \widehat{q}_{[s+b:i]} \right\}, \qquad i \in [n], \tag{16}$$

where $\widehat{f}_j = 0$ for $j \le 0$, $\widehat{q}_{[i+1:i]} = 0$, and infeasible subproblems have value $+\infty$. Unlike (12), this restricted recurrence considers only reset-free intervals of length at most $M-1$. Each interval cost $\widehat{q}_{[\ell:k]}$ appearing in (16) can therefore be computed by solving at most $O(\vartheta_b^M)$ QPs. Since (16) involves $O(nM)$ such costs, computing all of them requires $O(nM\vartheta_b^M)$ QP solves. Once these costs are available, the recurrence computes $\widehat{f}_n$ in $O(nM)$ time. If $\mathbf{0} \in \Gamma$, we also evaluate the full support $\boldsymbol{z} = \mathbf{1}$ by solving one additional QP and take the smaller of its optimal value and $\widehat{f}_n$. This yields the optimal value of (1) over the restricted feasible set $\mathcal{P}_{M,b}$. This approach raises two natural questions:

1. Does restricting $\mathcal{P}$ to $\mathcal{P}_{M,b}$ preserve optimality? In Section 3.2, we show that, if Γ satisfies (8) and μ satisfies a *barrier-free property* (Definition 5), there exists a choice of

$$M = O\left(\log\left(\frac{n}{\delta}\right) + b\right)$$

for which $\mathcal{P}_{M,b}$ and $\mathrm{conv}(\mathcal{P}_{M,b})$ are, respectively, a (μ, Γ, δ)-probabilistically exact approximation and convexification of $\mathcal{P}$. Thus, with probability at least $1-\delta$ over $\boldsymbol{\alpha} \sim \mu$, replacing $\mathcal{P}$ with either set preserves the optimal value of (1) simultaneously for every $\boldsymbol{\beta} \in \Gamma$. Moreover, in Section 3.3, we show that the barrier-free property required for μ holds under mild and verifiable conditions, including those delineated in Theorem 1.

2. Does restricting $\mathcal{P}$ to $\mathcal{P}_{M,b}$ guarantee tractability? As discussed above, for $b \ge 2$, solving (1) over $\mathcal{P}_{M,b}$ via the restricted DP recurrence (16) requires solving $O(nM\vartheta_b^M)$ QPs. More generally, in Section 3.4, we show that $\mathrm{conv}(\mathcal{P}_{M,b})$ admits an exact extended SOCP formulation with $O(nM\Phi_b(M))$ variables and constraints, where

$$\Phi_b(M) := \begin{cases} M, & b = 1, \\ \vartheta_b^M, & b \ge 2. \end{cases}$$

With $M = O(\log(n/\delta)+b)$, both the running time and the formulation size are polynomial in n and $1/\delta$ for fixed bandwidth b and model parameters.

3.2 Barriers and Probabilistic Reset Support Sets

Under Assumption 1, let

$$\kappa := \frac{\lambda_{\max}}{\lambda_{\min}}, \qquad \rho := \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{1/b} \in [0,1), \qquad \tau_b := \lambda_{\min}\rho^{b-1}\sqrt{\frac{\beta_0}{2\lambda_{\max}}}. \tag{17}$$

We focus on the nontrivial case $\rho \in (0,1)$. If $\rho = 0$, then $\kappa = 1$ and $\boldsymbol{Q}$ is a scalar multiple of the identity, in which case $\mathrm{conv}(\mathcal{P})$ admits the standard linear-size perspective formulation [7, 11]. We begin by introducing a local weight whose value triggers a reset.

Definition 4 (Local weight). *The* local weight *at $i \in [n]$ is defined as $W_i(\boldsymbol{\alpha}) := \sum_{j=1}^{n} \rho^{|i-j|}|\alpha_j|$. We call $i \in [n-b+1]$ a* barrier *when $W_i(\boldsymbol{\alpha}) \le \tau_b$.*

Lemma 2 (Barriers force resets). *Fix $\boldsymbol{\alpha} \in \mathbb{R}^n$ and $i \in [n-b+1]$. If $W_i(\boldsymbol{\alpha}) \le \tau_b$, then, for every $\boldsymbol{\beta} \in [\beta_0, \infty)^n$, every optimal support vector $\boldsymbol{z}^\star$ of (1) has a resetting sequence of length b beginning at i; that is, $z_i^\star = z_{i+1}^\star = \cdots = z_{i+b-1}^\star = 0$.*

By the above lemma, to ensure that every optimal solution belongs to $\mathcal{P}_{M,b}$, it suffices that, within every block of M consecutive coordinates, there exists at least one barrier i such that $W_i(\boldsymbol{\alpha}) \leq \tau_b$. We show next that this property holds with high probability whenever the probability measure μ governing $\boldsymbol{\alpha}$ satisfies a barrier-free property.

Definition 5 (Barrier-free property). *Let $C_{\mathrm{v}} \geq 1$ and $c_{\mathrm{v}} > 0$ be constants independent of n. A probability measure μ on $\mathbb{R}^n$ is said to satisfy the* $(C_{\mathrm{v}}, c_{\mathrm{v}})$-barrier-free property *if, for every interval $\mathcal{I} \subseteq [n - b + 1]$ of consecutive indices,*

$$\mathbb{P}_{\boldsymbol{\alpha}\sim\mu}\left(W_i(\boldsymbol{\alpha}) > \tau_b \text{ for all } i \in \mathcal{I}\right) \leq C_{\mathrm{v}} \exp(-c_{\mathrm{v}}|\mathcal{I}|)\,. \tag{18}$$

Theorem 2 (Probabilistic exactness under the barrier-free property). *Suppose that μ satisfies the $(C_{\mathrm{v}}, c_{\mathrm{v}})$-barrier-free property. For any $\delta \in (0, 1]$, define*

$$M := \left\lceil \frac{1}{c_{\mathrm{v}}} \log\left(\frac{C_{\mathrm{v}} n}{\delta}\right)\right\rceil + b - 1. \tag{19}$$

Then $\mathcal{P}_{M,b}$ is a (μ, Γ, δ)-probabilistically exact approximation of $\mathcal{P}$, and $\mathrm{conv}(\mathcal{P}_{M,b})$ *is a (μ, Γ, δ)-probabilistically exact convexification of $\mathcal{P}$.*

The preceding theorem establishes the probabilistic exactness of $\mathcal{P}_{M,b}$ and its convex hull under the barrier-free property. To demonstrate the practical relevance of this result, we next show that the barrier-free property holds for broad classes of distributions.

3.3 Measures Satisfying the Barrier-free Property

Our next result establishes the barrier-free property for product measures under two very mild conditions.

Proposition 1 (Product measures satisfying the barrier-free property). *Suppose that μ is a product measure on $\mathbb{R}^n$, that is, the coordinates of $\boldsymbol{\alpha} \sim \mu$ are independent. Fix any $g \in (1, \rho^{-1})$ and define*

$$a_g := \tau_b \frac{1 - \rho g}{1 + \rho g} > 0, \tag{20}$$

where τ_b is the barrier threshold defined in (17)*. Assume that there exist constants $p_0 \in (0, 1]$, $q > 0$, and $M_q < \infty$, independent of n, such that*

$$\min_{i\in[n]} \mathbb{P}(|\alpha_i| \leq a_g) \geq p_0, \qquad \max_{i\in[n]} \mathbb{E}|\alpha_i|^q \leq M_q. \tag{21}$$

Then μ satisfies the (C_v, c_v)-barrier-free property for some constants $C_v \geq 1$ and $c_v > 0$ that depend only on ρ, g, τ_b, p_0, q, and M_q, and are independent of n.

In Appendix B, we verify that these assumptions hold for independent coordinates with Gaussian, Laplace, Cauchy, or Student-t distributions, as well as uniformly bounded coordinates satisfying the first condition in (21). For example, for i.i.d. standard normal coordinates, one may choose $p_0 = 2\Phi(a_g) - 1$, $q = 1$, and $M_1 = \sqrt{2/\pi}$, where Φ denotes the standard normal cumulative distribution function.

We devote the remainder of this subsection to outlining the proof of Proposition 1. A key idea underlying our analysis is to interpret the random coefficients $\alpha_1, \ldots, \alpha_n$ as inducing a random covering of the indices. For notational convenience, extend $\alpha_1, \ldots, \alpha_n$ to a sequence including large or negative indexes $(\alpha_j)_{j\in\mathbb{Z}}$, by setting $\alpha_j = 0$ for every $j \notin [n]$. The added coordinates are deterministic and hence independent of the original coordinates, and they trivially satisfy the bounds in (21). This extension allows us to formulate our argument without separately treating boundary effects.

Definition 6 (Index coverage). *For each index $j \in \mathbb{Z}$, define its* coverage radius R_j *and* coverage neighborhood $\mathcal{B}_j$ *as*

$$\mathcal{B}_j := \{i \in \mathbb{Z} : |i - j| < R_j\}, \quad \textit{where} \quad R_j := \min\{k \in \mathbb{Z}_+ : |\alpha_j| \le a_g g^k\}. \tag{22}$$

We say that an index $i \in \mathbb{Z}$ is covered *by $j \in \mathbb{Z}$ if $i \in \mathcal{B}_j$. Accordingly, an index $i \in \mathbb{Z}$ is called* uncovered *if it is not covered by any $j \in \mathbb{Z}$.*

Lemma 3 (Uncovered indices are barriers). *Any uncovered index $i \in [n]$ is a barrier.*

By Lemma 3, to establish that an index $i \in [n]$ is a barrier, it suffices to show that it is uncovered. Consequently, verifying the barrier-free property reduces to the following probabilistic question: how likely is it that a long interval has at least one uncovered index? The main difficulty is that the event "index i is uncovered" depends on all neighborhoods $\mathcal{B}_j$. As a result, uncovered-index events are generally dependent, even though the coordinates α_j themselves are independent. We address this issue by separating coverage due to nearby indices from coverage due to remote indices.

Definition 7 (Local and remote coverage). *Given $D > 0$, an index $i \in \mathbb{Z}$ is said to be* D-locally uncovered *if $i \notin \mathcal{B}_j$ for every $j \in \mathbb{Z}$ such that $|i - j| \le D$. An index $j \in \mathbb{Z}$ is called a* D-remote coverer *of $i \in \mathbb{Z}$ if $|i - j| > D$ and $i \in \mathcal{B}_j$.*

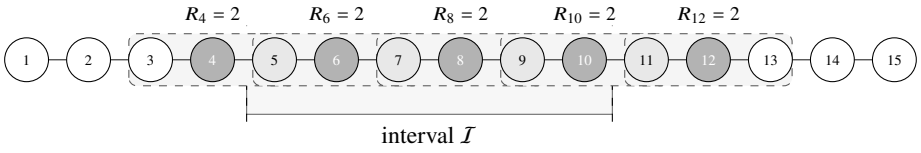
Figure 2 illustrates these two coverage mechanisms. The local–remote decomposition separates the dependence issue into two parts. First, D-local uncoveredness depends only on a block of coordinates around each candidate index. Thus, candidates separated by more than $2D$ have disjoint coordinate blocks, making their D-local-uncovered events independent under the product-measure assumption. Second, once sufficiently many locally uncovered candidates are identified, it remains to bound the probability that they are covered by remote indices. The next lemma formalizes the first observation.

Lemma 4 (Probability and independence of local-uncovered events). *Under the assumptions of Proposition 1, fix $D \in \mathbb{N}$ and, for each $i \in [n]$, let A_i denote the event that i is D-locally uncovered. Then, there exists a constant $p_D > 0$, depending only on $\tau_b, g, \rho, q, p_0, M_q$ and independent of n, such that, for every $i \in [n]$,*

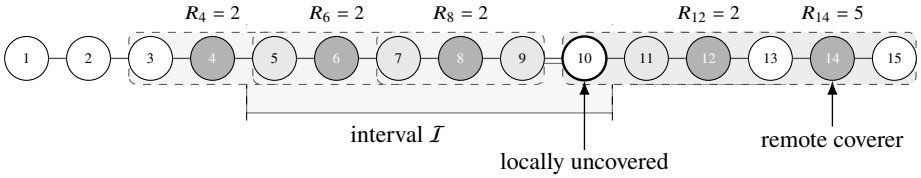
$$\mathbb{P}_{\boldsymbol{\alpha}\sim\mu}(A_i) \ge p_D.$$

Moreover, if $i, j \in [n]$ satisfy $|i - j| \ge 2D + 1$, then A_i and A_j are independent.

Lemma 5 (Many locally uncovered candidates). *Under the assumptions of Lemma 4, fix $D \in \mathbb{N}$, and let $c_1, \ldots, c_m \in [n]$ be deterministic candidate indices, where $m \ge 1$, whose*



(a) Every node in I is covered locally.



(b) A node in $\mathcal{I}$ is not covered locally, but a remote coverer reaches it.

Fig. 2. Two ways in which an interval can contain no uncovered nodes, illustrated for $D = 3$. In panel (a), every node in the interval is covered by a node within distance D. In panel (b), node 10 is D-locally uncovered, but it is covered by the remote coverer 14.

pairwise distances are at least $2D+1$. *Let* $X := |\{k \in [m] : c_k \text{ is } D\text{-locally uncovered}\}|$ *denote the number of locally uncovered candidates. Then*

$$\mathbb{P}_{\boldsymbol{\alpha}\sim\mu}\left(X > \frac{p_D m}{2}\right) \geq 1 - \exp\left(-\frac{p_D m}{8}\right).$$

Choose candidates $c_1, \ldots, c_m \in \mathcal{I} \subseteq [n]$ spaced $2D+1$ apart. By Lemma 5, a positive fraction are D-locally uncovered with high probability for sufficiently large m. However, these candidates may still have remote coverers (see Figure 2). Define

$$N_{\geq D}(j) := \left|\{k \in [m] : D < |j - c_k|,\ c_k \in \mathcal{B}_j\}\right|, \qquad j \in \mathbb{Z}$$

as the number of candidates for which index j is a D-remote coverer. The next lemma bounds $\sum_{j\in\mathbb{Z}} N_{\geq D}(j)$, the total number of remote coverings counted with multiplicity, by a small fraction of m with high probability.

Lemma 6 (Few remotely covered candidates). *Under the assumptions of Proposition 1, there exist an integer* $D \geq 1$ *and a constant* $0 < p_R \leq p_D$, *depending only on* $\tau_b, g, \rho, q, p_0, M_q$ *and independent of* n, *such that the following holds: for every collection of candidate indices* $c_1, \ldots, c_m$ *separated by at least* $2D+1$,

$$\mathbb{P}_{\boldsymbol{\alpha}\sim\mu}\left(\sum_{j\in\mathbb{Z}} N_{\geq D}(j) \leq \frac{p_R m}{2}\right) \geq 1 - \exp\left(-\frac{3p_R m}{8}\right).$$

By Lemmas 5 and 6, with high probability, locally uncovered candidates outnumber remote coverings. Thus, a barrier exists by Lemma 3, proving Proposition 1.

3.4 Compact SOCP Representation

We now present an SOCP-representable formulation of $\operatorname{conv}(\mathcal{P}_{M,b})$ that agrees with the restricted DP (16), including the case $\mathbf{0} \in \Gamma$. Let $M \geq b$ be an integer; as before, the reset condition is vacuous when $M > n$. Define the acyclic graph $G = (\mathcal{V}, \mathcal{A})$ by

$$\begin{aligned}\mathcal{V} &= \{0\} \cup [1 : n - b + 1] \cup \{n + 1\},\\ \mathcal{A} &= \{(0, j) : j \in \mathcal{V} \setminus \{0\},\ j \leq M\}\\ &\quad \cup \{(i, j) : i \in [1 : n - b + 1],\ j \in \mathcal{V},\ b \leq j - i \leq b + M - 1\}.\end{aligned}$$

In the graph, vertex 0 is a source vertex and $n + 1$ is a terminal vertex. Each internal vertex i marks the start of a chosen resetting sequence. Arc $(0, j)$ represents the interval $[1 : j - 1]$ before the first chosen resetting sequence. Arc (i, j) with $i \geq 1$ represents the block $[i : j - 1]$, consisting of b zeros followed by a reset-free interval of length at most $M - 1$. For integers $0 \leq \ell \leq k$ and a positive definite matrix $\mathbf{\Sigma} \in \mathbb{R}^{k\times k}$, define

$$\begin{aligned}\mathcal{X}_{\ell,k}(\mathbf{\Sigma}) = \Big\{(t, \boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}_+ \times \mathbb{R}^k \times \{0, 1\}^k : t \geq \boldsymbol{x}^\top \mathbf{\Sigma} \boldsymbol{x}, \quad \boldsymbol{x} \circ (\mathbf{1} - \boldsymbol{z}) = \mathbf{0},\\ z_i = 0 \quad \forall i \in [\ell], \qquad \sum_{i=j}^{j+b-1} z_i \geq 1 \quad \forall j \in [\ell + 1 : k - b + 1]\Big\}.\end{aligned}$$

Constraints indexed by an empty interval are vacuous. For $k = 0$, we interpret $\mathcal{X}_{0,0}$ as the set consisting of $t \geq 0$ and empty vectors $\boldsymbol{x}, \boldsymbol{z}$. For each arc (i, j), define $I_{ij} := [\max\{1, i\} : j - 1]$ and let $\boldsymbol{Q}_{I_{ij},I_{ij}}$ denote the submatrix of $\boldsymbol{Q}$ indexed by I_{ij}. Let $\mathcal{X}^{ij}$ consist of the points $(t, \boldsymbol{x}, \boldsymbol{z}) \in \mathcal{P}$ with $x_\ell = z_\ell = 0$ for $\ell \notin I_{ij}$, $\boldsymbol{z} \in \mathcal{Z}_{M,b}$, and

$$(t, \boldsymbol{x}_{I_{0j}}, \boldsymbol{z}_{I_{0j}}) \in \mathcal{X}_{0,j-1}(\boldsymbol{Q}_{I_{0j},I_{0j}}) \text{ if } i = 0, \qquad (t, \boldsymbol{x}_{I_{ij}}, \boldsymbol{z}_{I_{ij}}) \in \mathcal{X}_{b,j-i}(\boldsymbol{Q}_{I_{ij},I_{ij}}) \text{ if } i \geq 1.$$

We discard every arc (i, j) for which $\mathcal{X}^{ij} = \emptyset$. We now show how, inspired by disjunctive programming representations of unions of convex sets [5], a description of $\operatorname{conv}(\mathcal{X}^{ij})$ can be used to construct an extended representation of $\operatorname{conv}(\mathcal{P}_{M,b})$. Specifically, let f^{ij} be the convex indicator function of $\operatorname{conv}\left(\mathcal{X}^{ij}\right)$, that is,

$$f^{ij}(t, \boldsymbol{x}, \boldsymbol{z}) = \begin{cases} 0 & \text{if } (t, \boldsymbol{x}, \boldsymbol{z}) \in \operatorname{conv}\left(\mathcal{X}^{ij}\right)\\ \infty & \text{otherwise.}\end{cases}$$

We defer to Proposition 3 the discussion of a mechanism to compute functions f^{ij}. Introduce flow variables r^{ij} and local variables $(t^{ij}, \boldsymbol{x^{ij}}, \boldsymbol{z^{ij}}) \in \mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^n$ for $(i, j) \in \mathcal{A}$. Additional variables $r^\star, t^\star \in \mathbb{R}$ and $\boldsymbol{x}^\star \in \mathbb{R}^n$ represent the full-support $\boldsymbol{z} = \mathbf{1}$. Consider

$$\boldsymbol{x} = \sum_{(i,j)\in\mathcal{A}} \boldsymbol{x^{ij}} + \boldsymbol{x}^\star, \qquad \boldsymbol{z} = \sum_{(i,j)\in\mathcal{A}} \boldsymbol{z^{ij}} + \mathbf{1}r^\star, \tag{23a}$$

$$t = \sum_{(i,j)\in\mathcal{A}} t^{ij} + t^\star, \tag{23b}$$

$$r^{ij} f^{ij}\left(\frac{t^{ij}}{r^{ij}}, \frac{\boldsymbol{x^{ij}}}{r^{ij}}, \frac{\boldsymbol{z^{ij}}}{r^{ij}}\right) \leq 0 \qquad \forall (i, j) \in \mathcal{A}, \tag{23c}$$

$$(\boldsymbol{x}^\star)^\top \boldsymbol{Q}\boldsymbol{x}^\star \le t^\star r^\star, \qquad t^\star \ge 0, \quad r^\star \ge 0, \tag{23d}$$

$$\sum_{(0,j)\in\mathcal{A}} r^{0j} = 1 - r^\star, \qquad \sum_{(i,n+1)\in\mathcal{A}} r^{i,n+1} = 1 - r^\star, \tag{23e}$$

$$\sum_{(i,k)\in\mathcal{A}} r^{ik} - \sum_{(k,j)\in\mathcal{A}} r^{kj} = 0 \qquad \forall k \in \mathcal{V}\setminus\{0,n+1\}, \tag{23f}$$

$$r^\star = 0 \qquad \text{if } \mathbf{0} \notin \Gamma, \tag{23g}$$

$$r^\star \ge 0, \qquad r^{ij} \ge 0 \qquad \forall (i,j) \in \mathcal{A}, \tag{23h}$$

where the function in (23c) is the closure of the perspective function of f^{ij}.

Proposition 2. *The projection of* (23) *onto* $(t, \boldsymbol{x}, \boldsymbol{z})$ *is* $\mathrm{conv}(\mathcal{P}_{M,b})$.

Proposition 3. *Fix an arc* (i, j) *with* $\mathcal{X}^{ij} \neq \emptyset$*, and let* $\boldsymbol{v}^1, \dots, \boldsymbol{v}^N$ *list all its admissible support vectors, that is, all* $\boldsymbol{v} \in \{0,1\}^n$ *such that* $(0, \mathbf{0}, \boldsymbol{v}) \in \mathcal{X}^{ij}$*. Then* $f^{ij}(t, \boldsymbol{x}, \boldsymbol{z}) \le 0$ *if and only if there exist* $\lambda_h, t^h \ge 0$ *and* $\boldsymbol{x}^h \in \mathbb{R}^n$*,* $h \in [N]$*, satisfying*

$$\begin{aligned} &\sum_{h=1}^N \lambda_h = 1, \qquad t \ge \sum_{h=1}^N t^h, \qquad \boldsymbol{x} = \sum_{h=1}^N \boldsymbol{x}^h, \qquad \boldsymbol{z} = \sum_{h=1}^N \lambda_h \boldsymbol{v}^h, \\ &\boldsymbol{x}^h \circ (\mathbf{1} - \boldsymbol{v}^h) = \mathbf{0}, h \in [N], \qquad (\boldsymbol{x}^h)^\top \boldsymbol{Q}\boldsymbol{x}^h \le t^h \lambda_h, h \in [N]. \end{aligned} \tag{24}$$

Proof. The disjunctive formulation of Ceria and Soares [5], applied to the quadratic epigraph for each fixed support $\boldsymbol{v}^h$, gives (24). □

Let $\mathcal{Y}$ be obtained from (23) by replacing each constraint (23c) with (24), normalized by $\sum_h \lambda_h = r^{ij}$, and eliminating fixed-zero coordinates. The following theorem summarizes our main result: $\mathcal{Y}$ is an exact extended SOCP representation of $\mathrm{conv}(\mathcal{P}_{M,b})$ and, under Theorem 2, yields a polynomial-size probabilistically exact convexification of $\mathcal{P}$.

Theorem 3 (Extended SOCP representation). *The projection of* $\mathcal{Y}$ *onto* $(t, \boldsymbol{x}, \boldsymbol{z})$ *equals* $\mathrm{conv}(\mathcal{P}_{M,b})$*. Moreover,* $\mathcal{Y}$ *admits an SOCP description in which the number of scalar variables, the number of linear constraints, and the total second-order-cone dimension are each bounded by* $O(nM\Phi_b(M))$

$$O(nM\Phi_b(M)) = \begin{cases} O(nM^2), & b = 1, \\ O(nM\vartheta_b^M), & b \ge 2. \end{cases} \tag{25}$$

Consequently, under the assumptions of Theorem 2 and with M *chosen as in* (19)*,* $\mathrm{conv}(\mathcal{P}_{M,b})$ *is a* (μ, Γ, δ)*-probabilistically exact convexification of* $\mathcal{P}$ *with an exact extended SOCP representation* $\mathcal{Y}$ *satisfying the size bounds in* (9)*.*

Declaration of AI Use

The authors independently developed the concept of probabilistically exact convexification and the initial proof strategy for banded structures in 2025, without the use of AI. They subsequently used ChatGPT 6 Pro to assist with scrutinizing mathematical arguments, checking proofs, and language polishing. All AI-assisted material was carefully reviewed and revised by the authors, who assume full responsibility for the mathematical correctness, content, and conclusions of this paper.

Acknowledgements

Salar Fattahi is supported, in part, by grants CCF-2337776 and DMS-2152776 from the National Science Foundation and grant N000142612074 from the Office of Naval Research. Andrés Gómez is supported, in part, by grant N00014-26-1-2117 from the Office of Naval Research and grant 2346058 from the National Science Foundation.

Bibliography

[1] Alper Atamtürk, Andrés Gómez, and Shaoning Han. Sparse and smooth signal estimation: Convexification of ℓ_0-formulations. *Journal of Machine Learning Research*, 22(52):1–43, 2021. URL https://jmlr.org/papers/v22/18-745.html.

[2] Julian Besag, Jeremy York, and Annie Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, 1991. https://doi.org/10.1007/BF00116466.

[3] Aaresh Bhathena, Salar Fattahi, Andrés Gómez, and Simge Küçükyavuz. Solving convex quadratic optimization with indicators over structured graphs. *arXiv preprint arXiv:2603.02103*, 2026.

[4] Daniel Bienstock and Tongtong Chen. Solving convex QPs with structured sparsity under indicator conditions. *arXiv preprint arXiv:2411.11722*, 2024.

[5] Sebastián Ceria and João Soares. Convex programming for disjunctive convex optimization. *Mathematical Programming*, 86(3):595–614, 1999.

[6] Soobin Choi, Salar Fattahi, Andres Gomez, Shaoning Han, and Leonardo Lozano. Convexification of mixed-integer quadratic optimization via decision diagrams. *arXiv preprint arXiv:2608.22815*, 2026.

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

[8] Andrés Gómez. Outlier detection in time series via mixed-integer conic quadratic optimization. *SIAM Journal on Optimization*, 31(3):1897–1925, 2021.

[9] Andrés Gómez and Shaoning Han. Convex submodular minimization with indicator variables. *arXiv preprint arXiv:2209.13161v2*, 2025. URL https://arxiv.org/abs/2209.13161v2.

[10] Andres Gomez, Shaoning Han, and Leonardo Lozano. Real-time solution of quadratic optimization problems with banded matrices and indicator variables. *arXiv preprint arXiv:2405.03051*, 2024.

[11] Oktay Günlük and Jeff Linderoth. Perspective reformulations of mixed integer nonlinear programs with indicator variables. *Mathematical Programming*, 124(1): 183–205, 2010.

[12] Sergey Gusev and David E. Bernal Neira. AI agent swarms as researchers: Progress, challenges, and open questions, 2026. URL https://arxiv.org/abs/2609.35719.

[13] Peijing Liu, Salar Fattahi, Andrés Gómez, and Simge Küçükyavuz. A graph-based decomposition method for convex quadratic optimization with indicators. *Mathematical Programming*, 200(2):669–701, 2023.

[14] M18. Exact indicator quadratic optimization at bandwidth two. Research note, topic M18, SECQUOIA/agent-swarm-research repository, September 2026. URL https://github.com/SECQUOIA/agent-swarm-research/blob/84c6be7aad17d085e5343e789885c1cbcbbd4e07/minlp-notes/results/indicator-quadratic-treewidth-two-hardness.md. AI-generated re-

search note, dated September 22, 2026. Version: commit 84c6be7 (paper-v1). Accessed October 23, 2026.
[15] R. Tyrrell Rockafellar. *Convex Analysis*, volume 28 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1970. https://doi.org/10.1515/9781400873173.
[16] R Tyrrell Rockafellar and Roger JB Wets. *Variational analysis*. Springer, 1998.
[17] William F Trench. Conditional convergence of infinite products. *The American mathematical monthly*, 106(7):646–651, 1999.
[18] Linchuan Wei, Andrés Gómez, and Simge Küçükyavuz. Ideal formulations for constrained convex optimization problems with indicator variables. *Mathematical Programming*, 192(1):57–88, 2022.
[19] Linchuan Wei, Alper Atamtürk, Andrés Gómez, and Simge Küçükyavuz. On the convex hull of convex quadratic optimization problems with indicators. *Mathematical Programming*, 204(1):703–737, 2024.

A Proofs

A.1 Proof of Lemma 1

We first show that both $\operatorname{conv}(\mathcal{X})$ and $\operatorname{conv}(\mathcal{P})$ are closed. Let C denote either $\mathcal{X}$ or $\mathcal{P}$. To prove that $\operatorname{conv}(C)$ is closed, we show that it can be written as the epigraph of a lower-semicontinuous function. By [16, Theorem 1.6], this implies that C is closed. Define $\varphi_C(\boldsymbol{x}, \boldsymbol{z}) := \inf\{t : (t, \boldsymbol{x}, \boldsymbol{z}) \in C\}$, with the convention that $\varphi_C(\boldsymbol{x}, \boldsymbol{z}) = +\infty$ when $\{t : (t, \boldsymbol{x}, \boldsymbol{z}) \in C\}$ is empty. Since C is upward-closed, $C = \{(t, \boldsymbol{x}, \boldsymbol{z}) : t \geq \varphi_C(\boldsymbol{x}, \boldsymbol{z})\}$, which is the epigraph of $\varphi(\boldsymbol{x}, \boldsymbol{z})$. Since this epigraph is closed, we conclude that φ_C is lower semicontinuous [16, Theorem 1.6]. Moreover, since $C \subseteq \mathcal{P}$,

$$\varphi_C(\boldsymbol{x}, \boldsymbol{z}) \geq \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} \geq \lambda_{\min} \|\boldsymbol{x}\|_2^2. \tag{26}$$

Therefore, φ is closed and $\frac{\varphi_C(\boldsymbol{x}, \boldsymbol{z})}{\|(\boldsymbol{x}, \boldsymbol{z})\|} \to +\infty$ whenever $\|(\boldsymbol{x}, \boldsymbol{z})\| \to +\infty$ (recall that $\boldsymbol{z} \in \{0, 1\}^n$ is bounded). Therefore, the conditions of [16, Corollary 3.47] are satisfied, which implies that the convex envelope of φ_C, denoted by $\operatorname{conv} \varphi_C$, is proper, lower semicontinuous, and coercive, and its defining infimum over convex combinations is attained (see [16, Proposition 2.31]). Therefore,

$$\operatorname{conv}(C) = \{(t, \boldsymbol{x}, \boldsymbol{z}) : t \geq (\operatorname{conv} \varphi_C)(\boldsymbol{x}, \boldsymbol{z})\}. \tag{27}$$

Thus, $\operatorname{conv}(C)$ is the epigraph of a lower-semicontinuous function and is therefore closed.

Now, set $\widehat{\mathcal{X}} := \operatorname{conv}(\mathcal{X})$ and $\widehat{\mathcal{P}} := \operatorname{conv}(\mathcal{P})$. As established above, both sets are closed and convex. Moreover, they are upward-closed and $\widehat{\mathcal{X}} \subseteq \widehat{\mathcal{P}}$. Suppose, toward a contradiction, that this inclusion is strict. Choose $(\bar{t}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{z}}) \in \widehat{\mathcal{P}} \setminus \widehat{\mathcal{X}}$. Because $\widehat{\mathcal{X}}$ is nonempty, closed, and convex, the strong separating hyperplane theorem [15, Corollary 11.4.2] yields $(\gamma, \boldsymbol{\alpha}, \boldsymbol{\beta}) \neq (0, \mathbf{0}, \mathbf{0})$ and $\eta \in \mathbb{R}$ such that

$$\gamma \bar{t} + \boldsymbol{\alpha}^\top \bar{\boldsymbol{x}} + \boldsymbol{\beta}^\top \bar{\boldsymbol{z}} < \eta \leq \gamma t + \boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\beta}^\top \boldsymbol{z}, \qquad \forall (t, \boldsymbol{x}, \boldsymbol{z}) \in \widehat{\mathcal{X}}. \tag{28}$$

We must have $\gamma \geq 0$. Indeed, if $\gamma < 0$, then for any $(t, \boldsymbol{x}, \boldsymbol{z}) \in \widehat{\mathcal{X}}$, upward-closedness implies that $(t+s, \boldsymbol{x}, \boldsymbol{z}) \in \widehat{\mathcal{X}}$ for every $s \geq 0$. Consequently, $\gamma(t+s)+\boldsymbol{\alpha}^\top \boldsymbol{x}+\boldsymbol{\beta}^\top \boldsymbol{z} \to -\infty$ as $s \to +\infty$, contradicting (28). If $\gamma = 0$, define $\Delta := \eta - \boldsymbol{\alpha}^\top \bar{\boldsymbol{x}} - \boldsymbol{\beta}^\top \bar{\boldsymbol{z}} > 0$. Since $\bar{t} \geq 0$, choose $\gamma_+ > 0$ sufficiently small that $\gamma_+ \bar{t} < \Delta$. Then

$$\gamma_+ \bar{t} + \boldsymbol{\alpha}^\top \bar{\boldsymbol{x}} + \boldsymbol{\beta}^\top \bar{\boldsymbol{z}} < \eta \leq \gamma_+ t + \boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\beta}^\top \boldsymbol{z}, \qquad \forall (t, \boldsymbol{x}, \boldsymbol{z}) \in \widehat{\mathcal{X}}, \tag{29}$$

where the second inequality follows from $\gamma_+ > \gamma$ and $t \geq 0$. Thus, replacing γ by γ_+, we may assume without loss of generality that $\gamma > 0$. Dividing the strict separation inequality by γ gives

$$\bar{t} + \left(\frac{\boldsymbol{\alpha}}{\gamma}\right)^\top \bar{\boldsymbol{x}} + \left(\frac{\boldsymbol{\beta}}{\gamma}\right)^\top \bar{\boldsymbol{z}} < h_{\widehat{\mathcal{X}}}\left(\frac{\boldsymbol{\alpha}}{\gamma}, \frac{\boldsymbol{\beta}}{\gamma}\right). \tag{30}$$

On the other hand, since $(\bar{t}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{z}}) \in \widehat{\mathcal{P}}$,

$$h_{\widehat{\mathcal{P}}}\left(\frac{\boldsymbol{\alpha}}{\gamma}, \frac{\boldsymbol{\beta}}{\gamma}\right) \leq \bar{t} + \left(\frac{\boldsymbol{\alpha}}{\gamma}\right)^\top \bar{\boldsymbol{x}} + \left(\frac{\boldsymbol{\beta}}{\gamma}\right)^\top \bar{\boldsymbol{z}}. \tag{31}$$

Therefore,

$$h_{\widehat{\mathcal{P}}}\left(\frac{\boldsymbol{\alpha}}{\gamma}, \frac{\boldsymbol{\beta}}{\gamma}\right) < h_{\widehat{\mathcal{X}}}\left(\frac{\boldsymbol{\alpha}}{\gamma}, \frac{\boldsymbol{\beta}}{\gamma}\right). \tag{32}$$

Recalling that $\widehat{\mathcal{X}} := \mathrm{conv}(\mathcal{X})$ and $\widehat{\mathcal{P}} := \mathrm{conv}(\mathcal{P})$, we have $h_{\widehat{\mathcal{X}}} = h_{\mathcal{X}}$ and $h_{\widehat{\mathcal{P}}} = h_{\mathcal{P}}$, which contradicts the assumed equality of the restricted lower-support functions. Hence,

$$\mathrm{conv}(\mathcal{X}) = \mathrm{conv}(\mathcal{P}).$$

The converse follows immediately from $h_{\widehat{\mathcal{X}}} = h_{\mathcal{X}}$ and $h_{\widehat{\mathcal{P}}} = h_{\mathcal{P}}$. □

A.2 Proof of Lemma 2

The proof relies on the following auxiliary lemma, which is a special case of [3, Lemma 5].

Lemma 7 (inverse decay; Lemma 5 of [3]). *Let $\mathcal{T} \subseteq [n]$ be any subset of the rows/columns of $\boldsymbol{Q}$, and let $\pi : \mathcal{T} \to [n]$ be the canonical indexing map that assigns to each $i \in \mathcal{T}$ its corresponding row/column position within the submatrix $\boldsymbol{Q}_{\mathcal{T},\mathcal{T}}$. For all indices $i, j \in \mathcal{T}$,*

$$\left|\left(\boldsymbol{Q}_{\mathcal{T},\mathcal{T}}^{-1}\right)_{\pi(i),\pi(j)}\right| \leq \frac{2}{\lambda_{\min}} \rho^{|i-j|}, \tag{33}$$

Proof (Proof of Lemma 2). Fix $\boldsymbol{\beta} \in [\beta_0, \infty)^n$ and an optimal solution $(t^\star, \boldsymbol{x}^\star, \boldsymbol{z}^\star)$ of (1), and let $\mathcal{T} := \{j \in [n] : z_j^\star = 1\}$. The vector $\boldsymbol{x}^\star$ minimizes $\boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{\alpha}^\top \boldsymbol{x}$ subject to $\boldsymbol{x}_{\mathcal{T}^c} = 0$. Hence

$$\boldsymbol{x}_{\mathcal{T}}^\star = -\frac{1}{2} \boldsymbol{Q}_{\mathcal{T},\mathcal{T}}^{-1} \boldsymbol{\alpha}_{\mathcal{T}}, \qquad \boldsymbol{x}_{\mathcal{T}^c}^\star = 0.$$

Invoking Lemma 7, for every $j \in T$,

$$|x_j^\star| \le \frac{1}{2} \sum_{q \in \mathcal{T}} \left| \left(\boldsymbol{Q}_{\mathcal{T},\mathcal{T}}^{-1} \right)_{\pi(j),\pi(q)} \right| |\alpha_q| \le \frac{1}{\lambda_{\min}} \sum_{q \in \mathcal{T}} \rho^{|j-q|} |\alpha_q| \le \frac{1}{\lambda_{\min}} W_j(\boldsymbol{\alpha}).$$

Due to the definition of the W_i (Definition 4) for every $k \ge 0$ with $i + k \le n$,

$$W_{i+k}(\boldsymbol{\alpha}) \le \rho^{-k} W_i(\boldsymbol{\alpha}). \tag{34}$$

Set $\xi := \sqrt{\frac{\beta_0}{2\lambda_{\max}}}$. For every $k \in \{0, \dots, b-1\}$, using (34) and the definition of τ_b,

$$|x_{i+k}^\star| \le \frac{1}{\lambda_{\min}} W_{i+k}(\boldsymbol{\alpha}) \le \frac{\rho^{-k}}{\lambda_{\min}} W_i(\boldsymbol{\alpha}) \le \rho^{b-1-k} \xi \le \xi.$$

Let $\mathcal{A} := \mathcal{T} \cap [i : i + b - 1]$. We show that $\mathcal{A} = \varnothing$. Suppose, toward a contradiction, that $\mathcal{A} \neq \varnothing$. Define $\widetilde{\boldsymbol{x}}$ by setting the coordinates of $\boldsymbol{x}^\star$ indexed by $\mathcal{A}$ to zero and leaving all other coordinates unchanged. Then $\widetilde{\boldsymbol{x}}$ is feasible for the support $\mathcal{T} \setminus \mathcal{A}$. For any support $\mathcal{S} \subseteq [n]$, write

$$f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{x}, \mathcal{S}) := \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{\alpha}^\top \boldsymbol{x} + \beta(\mathcal{S}), \qquad \beta(\mathcal{S}) := \sum_{j \in \mathcal{S}} \beta_j.$$

Since $\boldsymbol{x}^\star$ minimizes the objective over the support $\mathcal{T}$, its first-order optimality conditions give $(2\boldsymbol{Q}\boldsymbol{x}^\star + \boldsymbol{\alpha})_{\mathcal{T}} = 0$. By expanding $f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{x}^\star, \mathcal{T})$ and $f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\widetilde{\boldsymbol{x}}, \mathcal{T} \setminus \mathcal{A})$ over $\mathcal{A}$ and $\mathcal{T} \setminus \mathcal{A}$, we obtain

$$\begin{aligned}
&f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{x}^\star, \mathcal{T}) - f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\widetilde{\boldsymbol{x}}, \mathcal{T} \setminus \mathcal{A}) \\
&\quad = (\boldsymbol{x}_{\mathcal{A}}^\star)^\top \boldsymbol{Q}_{\mathcal{A}\mathcal{A}} \boldsymbol{x}_{\mathcal{A}}^\star + 2(\boldsymbol{x}_{\mathcal{A}}^\star)^\top \boldsymbol{Q}_{\mathcal{A},\mathcal{T}\setminus\mathcal{A}} \boldsymbol{x}_{\mathcal{T}\setminus\mathcal{A}}^\star + \boldsymbol{\alpha}_{\mathcal{A}}^\top \boldsymbol{x}_{\mathcal{A}}^\star + \beta(\mathcal{A}) \\
&\quad = 2(\boldsymbol{x}_{\mathcal{A}}^\star)^\top (\boldsymbol{Q}\boldsymbol{x}^\star)_{\mathcal{A}} - (\boldsymbol{x}_{\mathcal{A}}^\star)^\top \boldsymbol{Q}_{\mathcal{A}\mathcal{A}} \boldsymbol{x}_{\mathcal{A}}^\star + \boldsymbol{\alpha}_{\mathcal{A}}^\top \boldsymbol{x}_{\mathcal{A}}^\star + \beta(\mathcal{A}) \\
&\quad = \beta(\mathcal{A}) - (\boldsymbol{x}_{\mathcal{A}}^\star)^\top \boldsymbol{Q}_{\mathcal{A}\mathcal{A}} \boldsymbol{x}_{\mathcal{A}}^\star.
\end{aligned} \tag{35}$$

On the other hand, $\boldsymbol{Q}_{\mathcal{A}\mathcal{A}} \preceq \lambda_{\max} \boldsymbol{I}$ and $|x_j^\star| \le \xi$ for every $j \in \mathcal{A}$, so

$$(\boldsymbol{x}_{\mathcal{A}}^\star)^\top \boldsymbol{Q}_{\mathcal{A}\mathcal{A}} \boldsymbol{x}_{\mathcal{A}}^\star \le \lambda_{\max} \|\boldsymbol{x}_{\mathcal{A}}^\star\|_2^2 \le \lambda_{\max} |\mathcal{A}| \xi^2 = \frac{|\mathcal{A}| \beta_0}{2} \le \frac{\beta(\mathcal{A})}{2}.$$

Substituting this bound into (35) gives

$$f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{x}^\star, \mathcal{T}) - f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\widetilde{\boldsymbol{x}}, \mathcal{T} \setminus \mathcal{A}) \ge \frac{\beta(\mathcal{A})}{2} > 0,$$

where the strict inequality follows from $\mathcal{A} \neq \varnothing$ and $\beta_j \ge \beta_0 > 0$. Thus, removing the indices in $\mathcal{A}$ strictly improves the objective, contradicting the optimality of $(\boldsymbol{x}^\star, \boldsymbol{z}^\star)$. Hence $\mathcal{A} = \varnothing$, and therefore $z_i^\star = \cdots = z_{i+b-1}^\star = 0$. □

A.3 Proof of Theorem 2

Define

$$L := M - b + 1 = \left\lceil \frac{1}{c_{\mathrm{v}}} \log\left(\frac{C_{\mathrm{v}} n}{\delta}\right) \right\rceil.$$

If $M > n$, then there is no block of M consecutive coordinates. Under the convention that the defining condition of $\mathcal{Z}_{M,b}$ is then vacuous, we have $\widehat{\mathcal{Z}}_{M,b} = \{0,1\}^n$ and hence $\mathcal{P}_{M,b} = \mathcal{P}$. The conclusion is therefore immediate. We may thus assume throughout the remainder of the proof that $M \le n$. For each $s \in [n - M + 1]$, let $\mathcal{B}_s := [s : s + M - 1]$ be the corresponding block of M consecutive coordinates. The indices at which a resetting sequence of length b can begin while remaining entirely within B_s form the interval

$$\mathcal{I}_s := [s : s + M - b] = [s : s + L - 1].$$

In particular, $|\mathcal{I}_s| = L$ and $\mathcal{I}_s \subseteq [n - b + 1]$. By the $(C_{\mathrm{v}}, c_{\mathrm{v}})$ barrier-free property,

$$\mathbb{P}_{\boldsymbol{\alpha}\sim\mu}\left(W_i(\boldsymbol{\alpha}) > \tau_b \text{ for every } i \in \mathcal{I}_s\right) \le C_{\mathrm{v}} e^{-c_{\mathrm{v}} L} \le \frac{\delta}{n},$$

where the second inequality follows from the definition of L. Define the event

$$G := \left\{\boldsymbol{\alpha} \in \mathbb{R}^n \;\middle|\; \text{for every } s \in [n - M + 1],\ \text{there exists } i \in \mathcal{I}_s \text{ such that } W_i(\boldsymbol{\alpha}) \le \tau_b\right\}.$$

Since there are $n - M + 1 \le n$ blocks of length M, a union bound gives

$$\mathbb{P}_{\boldsymbol{\alpha}\sim\mu}(G^c) \le (n - M + 1)\frac{\delta}{n} \le \delta.$$

where G^c is the complement of G. Thus, $\mathbb{P}_{\boldsymbol{\alpha}\sim\mu}(G) \ge 1 - \delta$. Now, fix $\boldsymbol{\alpha} \in G$. We show that

$$h_{\mathcal{P}_{M,b}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = h_{\mathcal{P}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \qquad \text{for every } \boldsymbol{\beta} \in \Gamma.$$

First, consider any $\boldsymbol{\beta} \in [\beta_0, \infty)^n$, and let $(t^\star, \boldsymbol{x}^\star, \boldsymbol{z}^\star)$ be an optimal solution over $\mathcal{P}$. For every $s \in [n - M + 1]$, the definition of G yields an index $i \in \mathcal{I}_s$ such that $W_i(\boldsymbol{\alpha}) \le \tau_b$. Lemma 2 then implies $z_i^\star = z_{i+1}^\star = \cdots = z_{i+b-1}^\star = 0$. Since $i \in \mathcal{I}_s$, this resetting sequence is contained entirely within $\mathcal{B}_s$. As this holds for every $s \in [n - M + 1]$, every block of M consecutive coordinates of $\boldsymbol{z}^\star$ contains a resetting sequence of length b. Hence $\boldsymbol{z}^\star \in \mathcal{Z}_{M,b} \subseteq \widehat{\mathcal{Z}}_{M,b}$, and therefore $(t^\star, \boldsymbol{x}^\star, \boldsymbol{z}^\star) \in \mathcal{P}_{M,b}$. This implies $h_{\mathcal{P}_{M,b}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \le h_{\mathcal{P}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Because $\mathcal{P}_{M,b} \subseteq \mathcal{P}$, we have $h_{\mathcal{P}_{M,b}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \ge h_{\mathcal{P}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Thus,

$$h_{\mathcal{P}_{M,b}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = h_{\mathcal{P}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \qquad \forall \boldsymbol{\beta} \in \Gamma \cap [\beta_0, \infty)^n.$$

It remains to consider the possible case $\boldsymbol{\beta} = \mathbf{0}$. Suppose that $\mathbf{0} \in \Gamma$. The all-one support imposes no restrictions on $\boldsymbol{x}$, and it belongs to $\widehat{\mathcal{Z}}_{M,b}$ by definition. In particular, setting

$$\bar{\boldsymbol{x}} := -\frac{1}{2}\boldsymbol{Q}^{-1}\boldsymbol{\alpha}, \qquad \bar{\boldsymbol{z}} := \mathbf{1}, \qquad \bar{t} := \bar{\boldsymbol{x}}^\top \boldsymbol{Q} \bar{\boldsymbol{x}}$$

gives a point $(\bar{t}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{z}}) \in \mathcal{P}_{M,b}$. Moreover, $\bar{\boldsymbol{x}}$ is the unique minimizer of $\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{\alpha}^\top \boldsymbol{x}$ over $\mathbb{R}^n$. Since the all-one support also belongs to $\mathcal{P}$, this point is optimal over $\mathcal{P}$ when $\boldsymbol{\beta} = \mathbf{0}$. Therefore,

$$h_{\mathcal{P}_{M,b}}(\boldsymbol{\alpha}, \mathbf{0}) = h_{\mathcal{P}}(\boldsymbol{\alpha}, \mathbf{0}).$$

Under the standing assumption $\Gamma \subseteq \{\mathbf{0}\} \cup [\beta_0, \infty)^n$, for every $\boldsymbol{\alpha} \in G$, we have $h_\chi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = h_\mathcal{P}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ for all $\boldsymbol{\beta} \in \Gamma$. Hence

$$\mathbb{P}_{\boldsymbol{\alpha} \sim \mu} \left(h_\chi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = h_\mathcal{P}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \quad \text{for all } \boldsymbol{\beta} \in \Gamma \right) \geq 1 - \delta.$$

Finally, $\mathcal{P}_{M,b}$ is nonempty and upward-closed in t. Therefore, $\mathcal{P}_{M,b}$ is a (μ, Γ, δ)-probabilistically exact approximation of $\mathcal{P}$ according to Definition 2. By the same definition, $\operatorname{conv}(\mathcal{P}_{M,b})$ is a (μ, Γ, δ)-probabilistically exact convexification. □

A.4 Proof of Lemma 3

By definition of R_j and $R_j \leq |i - j|$, we have $|\alpha_j| \leq a_g g^{|i-j|}$. Therefore

$$W_i(\boldsymbol{\alpha}) = \sum_{j=1}^{n} \rho^{|i-j|} |\alpha_j| \leq \sum_{j \in \mathbb{Z}} \rho^{|i-j|} |\alpha_j| \leq a_g \sum_{d \in \mathbb{Z}} (\rho g)^{|d|} = a_g \frac{1 + \rho g}{1 - \rho g} = \tau_b.$$

□

A.5 Proof of Lemma 4

We first derive a uniform tail bound for the coverage radii. By the definition of R_j, for every $j \in [n]$ and $d \in \mathbb{Z}_+$,

$$\{R_j > d\} = \{|\alpha_j| > a_g g^d\}.$$

Define $C_q := M_q a_g^{-q}$ and $\psi_q := g^{-q} \in (0, 1)$. Markov's inequality and (21) give

$$\mathbb{P}(R_j > d) \leq \frac{\mathbb{E}|\alpha_j|^q}{\left(a_g g^d\right)^q} \leq M_q a_g^{-q} g^{-qd} = C_q \psi_q^d. \quad \Longrightarrow \quad \mathbb{P}(R_j \leq d) \geq 1 - C_q \psi_q^d. \tag{36}$$

On the other hand, since $g^d \geq 1$,

$$\{|\alpha_j| \leq a_g\} \subseteq \{|\alpha_j| \leq a_g g^d\} = \{R_j \leq d\}.$$

This implies that $\mathbb{P}(R_j \leq d) \geq p_0$ since, by the first condition of (21), we have $\mathbb{P}(|\alpha_j| \leq a_g) \geq p_0$. Combining the two bounds yields

$$\mathbb{P}(R_j \leq d) \geq \max\{p_0, 1 - C_q \psi_q^d\}. \tag{37}$$

Now fix $i \in [n]$. The event A_i can be written as

$$\begin{aligned} A_i &= \{i \notin \mathcal{B}_i\} \cap \bigcap_{d=1}^{D} \left(\{i \notin \mathcal{B}_{i-d}\} \cap \{i \notin \mathcal{B}_{i+d}\} \right) \\ &= \{R_i = 0\} \cap \bigcap_{d=1}^{D} \left(\{R_{i-d} \leq d\} \cap \{R_{i+d} \leq d\} \right), \end{aligned}$$

Each R_j is a function only of α_j. Since μ is a product measure, the radii associated with distinct indices in $[n]$ are independent. Moreover,

$$\mathbb{P}(R_i = 0) = \mathbb{P}(|\alpha_i| \le a_g) \ge p_0.$$

Using independence together with (37), and noting that for each $d = 1, \dots, D$ there are at most two relevant indices at distance d from i, we obtain

$$\mathbb{P}(A_i) \ge p_D := p_0 \prod_{d=1}^{D} \max\left\{p_0,\, 1 - C_q \psi_q^d\right\}^2.$$

Finally, A_i depends only on the coordinates indexed by $[i - D : i + D] \cap [n]$, while A_j depends only on those indexed by $[j - D : j + D] \cap [n]$. If $|i - j| \ge 2D + 1$, these two index sets are disjoint. Since μ is a product measure, the corresponding coordinate blocks are independent. Therefore, A_i and A_j are independent. □

A.6 Proof of Lemma 5

For each $k \in [m]$, recall that A_{c_k} denotes the event that c_k is D-locally uncovered. The separation of the candidates together with Lemma 4 implies that the events $A_{c_1}, \dots, A_{c_m}$ are mutually independent. By Lemma 4, $\mathbb{P}(A_{c_k}) \ge p_D$ for every $k \in [m]$. Thus $X = \sum_{k=1}^{m} \mathbf{1}_{A_{c_k}}$ is a sum of independent Bernoulli variables with $\mathbb{E}X \ge p_D m$. The Chernoff bound gives

$$\mathbb{P}\left(X > \frac{p_D m}{2}\right) \ge 1 - \mathbb{P}\left(X \le \frac{\mathbb{E}X}{2}\right) \ge 1 - \exp\left(-\frac{\mathbb{E}X}{8}\right) \ge 1 - \exp\left(-\frac{p_D m}{8}\right).$$

□

A.7 Proof of Lemma 6

For brevity, write $S := \sum_{j \in \mathbb{Z}} N_{\ge D}(j)$. Recall that $\alpha_j = 0$ for $j \notin [n]$, and hence $R_j = N_{\ge D}(j) = 0$ for such indices. Thus, the sum defining S is effectively finite. Define

$$p_R := p_0 \prod_{d=1}^{\infty} \max\left\{p_0,\, 1 - C_q \psi_q^d\right\}^2 > 0,$$

where $C_q := M_q a_g^{-q}$ and $\psi_q := g^{-q} \in (0, 1)$, as defined before. We first show that $p_R > 0$. Let $p_d := \max\{p_0, 1 - C_q \psi_q^d\}$ for $d \ge 1$. Since $0 < p_0 \le p_d \le 1$ and

$$\sum_{d=1}^{\infty} |p_d - 1| \le C_q \sum_{d=1}^{\infty} \psi_q^d = \frac{C_q \psi_q}{1 - \psi_q} < \infty,$$

the standard infinite-product convergence criterion [17, Theorem 1] implies that $\prod_{d=1}^{\infty} p_d$ converges to a strictly positive limit. Consequently, $p_R = p_0 \prod_{d=1}^{\infty} p_d^2 > 0$.

Set $s_D := 2D + 1$. Since the candidate indices $c_1, \dots, c_m$ are separated by at least s_D, a radius R_j can cover at most $1 + 2R_j/s_D$ candidate indices. Consequently,

$N_{\geq D}(j) \leq (1+2R_j/s_D)\mathbf{1}_{\{R_j>D\}}$. For fixed candidate indices, $N_{\geq D}(j)$ is a deterministic function of R_j alone. Since R_j depends only on α_j and μ is a product measure, the variables $(N_{\geq D}(j))_{j\in\mathbb{Z}}$ are independent. Therefore,

$$\log \mathbb{E}e^S = \sum_{j\in\mathbb{Z}} \log \mathbb{E}e^{N_{\geq D}(j)} \leq \sum_{j\in\mathbb{Z}} \left(\mathbb{E}e^{N_{\geq D}(j)} - 1\right) \leq \sum_{j\in\mathbb{Z}} \mathbb{E}\left[N_{\geq D}(j)e^{N_{\geq D}(j)}\right],$$

where the equality follows from the independence of $(N_{\geq D}(j))_{j\in\mathbb{Z}}$; the first inequality follows from $\log u \leq u-1$ for every $u>0$; and the second inequality follows from $e^u - 1 \leq ue^u$ for $u \geq 0$. By the definition of $N_{\geq D}(j)$, we have $N_{\geq D}(j) = \sum_{k=1}^m \mathbf{1}_{\{D<|j-c_k|<R_j\}}$. Moreover, $N_{\geq D}(j) \leq 1 + 2R_j/s_D$. Therefore, for each $j \in \mathbb{Z}$, we have

$$N_{\geq D}(j)e^{N_{\geq D}(j)} = \sum_{k=1}^m \mathbf{1}_{\{D<|j-c_k|<R_j\}} e^{N_{\geq D}(j)} \leq e\sum_{k=1}^m \mathbf{1}_{\{D<|j-c_k|<R_j\}} e^{2R_j/s_D}.$$

Hence,

$$\begin{aligned}
\log \mathbb{E}e^S &\leq e\sum_{j\in\mathbb{Z}}\sum_{k=1}^m \mathbb{E}\left[\mathbf{1}_{\{D<|j-c_k|<R_j\}} e^{2R_j/s_D}\right] \\
&\leq e\sum_{k=1}^m \sum_{\substack{j\in\mathbb{Z} \\ |j-c_k|>D}} \mathbb{E}\left[\mathbf{1}_{\{R_j>|j-c_k|\}} e^{2R_j/s_D}\right] \\
&\leq 2em\sum_{d=D+1}^{\infty} \sup_{j\in[n]} \mathbb{E}\left[\mathbf{1}_{\{R_j>d\}} e^{2R_j/s_D}\right],
\end{aligned}$$

where the factor 2 accounts for the two possible indices at each distance d from a candidate index. Define $u_D := \psi_q e^{2/s_D}$. Recall that $\psi_q \in (0,1)$. Therefore, for D sufficiently large, $u_D < 1$. For every $d \geq D+1$,

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_{\{R_j>d\}} e^{2R_j/s_D}\right] &= \sum_{r=d+1}^{\infty} e^{2r/s_D}\mathbb{P}(R_j = r) \\
&\leq \sum_{r=d+1}^{\infty} e^{2r/s_D}\mathbb{P}(R_j > r-1) \\
&\leq \frac{C_q}{\psi_q}\sum_{r=d+1}^{\infty} u_D^r \\
&= \frac{C_q}{\psi_q}\frac{u_D^{d+1}}{1-u_D},
\end{aligned}$$

where the last inequality follows from the fact that, for every $j \in [n]$ and $r \in \mathbb{Z}_+$, $\mathbb{P}(R_j > r-1) \leq C_q\psi_q^{r-1}$. Substituting this estimate above yields

$$\log \mathbb{E}e^S \leq m\varepsilon_D, \quad \text{where} \quad \varepsilon_D := \frac{2eC_q}{\psi_q}\frac{u_D^{D+2}}{(1-u_D)^2}.$$

Since $u_D = \psi_q e^{2/(2D+1)} \to \psi_q < 1$ as $D \to \infty$, we have $\varepsilon_D \to 0$ as $D \to \infty$. We may therefore choose a fixed D, depending only on g, ρ, q, p_0, M_q, such that $\varepsilon_D \le p_R/8$. Finally, Markov's inequality gives

$$\mathbb{P}\left(S > \frac{p_R m}{2}\right) \le \exp\left(-\frac{p_R m}{2}\right) \mathbb{E}e^S \le \exp\left[-\left(\frac{p_R}{2} - \varepsilon_D\right) m\right] \le \exp\left(-\frac{3p_R m}{8}\right).$$

Equivalently,

$$\mathbb{P}\left(\sum_{j\in\mathbb{Z}} N_{\ge D}(j) \le \frac{p_R m}{2}\right) \ge 1 - \exp\left(-\frac{3p_R m}{8}\right),$$

which proves the claim. □

A.8 Proof of Proposition 1

Let $0 < p_R \le p_D$ and $D \ge 1$ be as in Lemma 6. Fix a nonempty interval $\mathcal{I} \subseteq [n-b+1]$ and select $m = \lceil |\mathcal{I}|/(2D+1) \rceil$ candidates in $\mathcal{I}$ separated by at least $2D+1$. By Lemmas 4 and 5 and noting $p_R \le p_D$, more than $p_R m/2$ candidates are D-locally uncovered, except on an event of probability at most $e^{-p_R m/8}$. Meanwhile, Lemma 6 ensures that the total number of remote coverings of these candidates, counted with multiplicity, is at most $p_R m/2$, except on an event of probability at most $e^{-3p_R m/8}$. When both conclusions hold, at least one locally uncovered candidate has no remote coverer. This candidate is globally uncovered and hence is a barrier by Lemma 3. A union bound therefore gives

$$\mathbb{P}_{\boldsymbol{\alpha}\sim\mu}\left(W_i(\boldsymbol{\alpha}) > \tau_b \text{ for every } i \in \mathcal{I}\right) \le e^{-p_R m/8} + e^{-3p_R m/8} \le 2\exp\left(-\frac{p_R|\mathcal{I}|}{8(2D+1)}\right).$$

Thus, μ satisfies the barrier-free property with $C_{\mathrm{v}} = 2$ and $c_{\mathrm{v}} = p_R/[8(2D+1)] > 0$. These constants are independent of n and $\mathcal{I}$. □

A.9 Proof of Proposition 2

Consider the minimization of an arbitrary linear function of the form

$$\min \boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\beta}^\top \boldsymbol{z} + t \text{ s.t. (23).} \tag{38}$$

We show that minimization (38) is equivalent to the restricted DP (16), and thus is equivalent to minimizing the same function over $\mathcal{P}_{M,b}$. First, using constraints (23a)-(23b), we can rewrite (38) as

$$\min \boldsymbol{\alpha}^\top \left(\sum_{(i,j)\in\mathcal{A}} \boldsymbol{x^{ij}} + \boldsymbol{x^\star}\right) + \boldsymbol{\beta}^\top \left(\sum_{(i,j)\in\mathcal{A}} \boldsymbol{z^{ij}} + \mathbf{1}r^\star\right) + \sum_{(i,j)\in\mathcal{A}} t^{ij} + t^\star \text{ s.t. (23c) – (23h).} \tag{39}$$

Now observe that if variables $\boldsymbol{r}$ are fixed, problem (39) is separable across variable $(t^{ij}, \boldsymbol{x^{ij}}, \boldsymbol{z^{ij}})$ and $(t^\star, \boldsymbol{x^\star}, \mathbf{1})$. If $r^{ij} = 0$, then constraint (23c) forces $\boldsymbol{z^{ij}} = \boldsymbol{x^{ij}} = \mathbf{0}$ and

$t^{ij} = 0$ in optimal solutions; otherwise, if $i \geq 1$, then the cost associated with variables $(t^{ij}, \boldsymbol{x^{ij}}, \boldsymbol{z^{ij}})$ is

$$\begin{aligned} &\min_{(t^{ij}, \boldsymbol{x^{ij}}, \boldsymbol{z^{ij}})} \boldsymbol{\alpha}^\top \boldsymbol{x^{ij}} + \boldsymbol{\beta}^\top \boldsymbol{z^{ij}} + t^{ij} \text{ s.t. } r^{ij} f^{ij}\left(\frac{t^{ij}}{r^{ij}}, \frac{\boldsymbol{x^{ij}}}{r^{ij}}, \frac{\boldsymbol{z^{ij}}}{r^{ij}}\right) \leq 0 \\ \Leftrightarrow \quad &\min_{(\bar{t}^{ij}, \bar{\boldsymbol{x}}^{\boldsymbol{ij}}, \bar{\boldsymbol{z}}^{\boldsymbol{ij}})} r^{ij}\left(\boldsymbol{\alpha}^\top \bar{\boldsymbol{x}}^{\boldsymbol{ij}} + \boldsymbol{\beta}^\top \bar{\boldsymbol{z}}^{\boldsymbol{ij}} + \bar{t}^{ij}\right) \text{ s.t. } f^{ij}(\bar{t}^{ij}, \bar{\boldsymbol{x}}^{\boldsymbol{ij}}, \bar{\boldsymbol{z}}^{\boldsymbol{ij}}) \leq 0, \end{aligned} \tag{40}$$

where the equivalence is obtained by doing the change of variables $(\bar{t}^{ij}, \bar{\boldsymbol{x}}^{\boldsymbol{ij}}, \bar{\boldsymbol{z}}^{\boldsymbol{ij}}) = \frac{1}{r^{ij}}(t^{ij}, \boldsymbol{x^{ij}}, \boldsymbol{z^{ij}})$. From the definition of f^{ij} as a convex indicator function, we find that

$$\begin{aligned} (40) \Leftrightarrow &\min_{(\bar{t}^{ij}, \bar{\boldsymbol{x}}^{\boldsymbol{ij}}, \bar{\boldsymbol{z}}^{\boldsymbol{ij}})} r^{ij}\left(\boldsymbol{\alpha}^\top \bar{\boldsymbol{x}}^{\boldsymbol{ij}} + \boldsymbol{\beta}^\top \bar{\boldsymbol{z}}^{\boldsymbol{ij}} + \bar{t}^{ij}\right) \text{ s.t. } (\bar{t}^{ij}, \bar{\boldsymbol{x}}^{\boldsymbol{ij}}, \bar{\boldsymbol{z}}^{\boldsymbol{ij}}) \in \operatorname{conv}(\mathcal{X}^{ij}) \\ = &\min_{(\bar{t}^{ij}, \bar{\boldsymbol{x}}^{\boldsymbol{ij}}, \bar{\boldsymbol{z}}^{\boldsymbol{ij}})} r^{ij}\left(\boldsymbol{\alpha}^\top \bar{\boldsymbol{x}}^{\boldsymbol{ij}} + \boldsymbol{\beta}^\top \bar{\boldsymbol{z}}^{\boldsymbol{ij}} + \bar{t}^{ij}\right) \text{ s.t. } (\bar{t}^{ij}, \bar{\boldsymbol{x}}^{\boldsymbol{ij}}, \bar{\boldsymbol{z}}^{\boldsymbol{ij}}) \in \mathcal{X}^{ij} \\ = &r^{ij} \widehat{q}_{[i+b:j-1]}, \end{aligned} \tag{41}$$

where $\widehat{q}_{[i+b:j-1]}$ is the restricted DP cost of the reset-free interval following the b zeros at $[i : i+b-1]$, as defined in Section 3.1. Observe that cost (41) also holds when $r^{ij} = 0$. Through identical arguments, the cost associated with a source arc $(0, j)$ is $r^{0j}\widehat{q}_{[1:j-1]}$, since its associated interval $[1 : j-1]$ must be reset-free. The cost associated with $(t^\star, \boldsymbol{x^\star}, \mathbf{1})$ is $r^\star\left(-\frac{1}{4}\boldsymbol{\alpha}^\top \boldsymbol{Q^{-1}} \boldsymbol{\alpha} + \boldsymbol{\beta}^\top \mathbf{1}\right)$, which is the cost of minimizing the quadratic program assuming a full support multiplied by $r^\star$. Thus, projecting out all auxiliary variables except $\boldsymbol{r}$ from (39), we find the equivalent form

$$\begin{aligned} \min \quad &\sum_{\substack{(i,j)\in\mathcal{A} \\ i\geq 1}} r^{ij}\widehat{q}_{[i+b:j-1]} + \sum_{(0,j)\in\mathcal{A}} r^{0j}\widehat{q}_{[1:j-1]} \\ &\quad + r^\star\left(-\frac{1}{4}\boldsymbol{\alpha}^\top \boldsymbol{Q^{-1}} \boldsymbol{\alpha} + \boldsymbol{\beta}^\top \mathbf{1}\right) \\ \text{s.t. } &\text{(23e)} - \text{(23h)}. \end{aligned}$$

This optimization problem corresponds precisely to a shortest path problem over the state transition graph of the restricted DP (16), and thus their optimal objective values coincide.

We next verify that both $\operatorname{conv}(\mathcal{P}_{M,b})$ and the projection of (23) onto $(t, \boldsymbol{x}, \boldsymbol{z})$ are closed. The set $\mathcal{P}_{M,b} \subseteq \mathcal{P}$ is upward-closed. Hence, by Lemma 1, $\operatorname{conv}(\mathcal{P}_{M,b})$ is closed. For the projection, consider any convergent sequence of projected points and choose a feasible lift of each point. The flow variables lie in $[0, 1]$, the nonnegative local epigraph variables are bounded by t, and the perspective constraints together with $\boldsymbol{Q} \succ 0$ bound all remaining local variables. The lifts therefore have a convergent subsequence, whose limit is feasible because the extended feasible set is closed. Thus, the limiting projected point also belongs to the projection, proving its closedness.

Both sets are therefore convex, upward-closed, and contained in $\{(t, \boldsymbol{x}, \boldsymbol{z}) : t \geq 0\}$. Since their optimal values agree for every objective $t + \boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\beta}^\top \boldsymbol{z}$, the separation argument in the proof of Lemma 1 yields the claimed convex-hull equality. □

A.10 Proof of Theorem 3

We first provide an upper and lower bounds on the number of reset-free patterns, which will be useful in establishing a bound on the size of $\mathcal{Y}$.

Lemma 8 (Number of reset-free patterns). *Let $N_b(k)$ be the number of binary support vectors of length k containing no b consecutive zeros, with $N_b(0) = 1$. Set $\vartheta_1 = 1$ and, for $b \geq 2$, let $\vartheta_b \in (1, 2)$ be the unique positive root of*

$$x^b = x^{b-1} + \cdots + x + 1.$$

There are constants $c_b, C_b > 0$, depending only on b, such that

$$c_b \vartheta_b^k \leq N_b(k) \leq C_b \vartheta_b^k \quad \forall k \geq 0, \qquad \sum_{k=0}^{M-1} N_b(k) \leq C_b \Phi_b(M), \tag{42}$$

where

$$\Phi_b(M) := \begin{cases} M, & b = 1, \\ \vartheta_b^M, & b \geq 2. \end{cases}$$

Proof. If $b = 1$, the only reset-free vector of each length is the all-ones vector, including the empty vector when $k = 0$. Thus $N_1(k) = 1$ and $\sum_{k=0}^{M-1} N_1(k) = M = \Phi_1(M)$, so both bounds hold with $c_1 = C_1 = 1$. Henceforth assume $b \geq 2$.

For $0 \leq k < b$, every binary vector of length k is reset-free, and hence $N_b(k) = 2^k$. We first derive a recurrence for $N_b(k)$. Fix $k \geq b$. For a binary vector $\boldsymbol{z}$ of length k, let $t(\boldsymbol{z})$ denote its number of trailing zeros, with $t(\boldsymbol{z}) = 0$ when $z_k = 1$. Partition the reset-free vectors according to $t(\boldsymbol{z})$. Since such a vector cannot contain b consecutive zeros, $t(\boldsymbol{z}) \in \{0, \ldots, b-1\}$. If $t(\boldsymbol{z}) = j$, its last $j+1$ entries are a one followed by j zeros. Deleting these entries leaves a reset-free vector of length $k - j - 1$. Conversely, appending a one followed by j zeros to any reset-free vector of length $k - j - 1$ produces a reset-free vector with exactly j trailing zeros. This bijection gives

$$N_b(k) = \sum_{j=0}^{b-1} N_b(k - j - 1) = \sum_{j=1}^{b} N_b(k - j), \qquad k \geq b. \tag{43}$$

The polynomial $p_b(x) := x^b - x^{b-1} - \cdots - x - 1$ has at most one positive root, since $p_b(x)/x^b = 1 - \sum_{j=1}^{b} x^{-j}$ is strictly increasing for $x > 0$. Moreover,

$$p_b(1) = 1 - b < 0, \qquad p_b(2) = 2^b - \sum_{j=0}^{b-1} 2^j = 1 > 0.$$

Thus its unique positive root ϑ_b lies in $(1, 2)$. Define the positive constants

$$c_b := \min_{0 \leq r < b} \frac{N_b(r)}{\vartheta_b^r}, \qquad D_b := \max_{0 \leq r < b} \frac{N_b(r)}{\vartheta_b^r}.$$

We claim that $c_b \vartheta_b^k \le N_b(k) \le D_b \vartheta_b^k$ for every $k \ge 0$. The claim holds for $0 \le k < b$ by construction. Suppose it holds for all indices smaller than some $k \ge b$. Using (43) and the defining equation for ϑ_b, we obtain

$$\begin{aligned} N_b(k) &= \sum_{j=1}^{b} N_b(k-j) \le D_b \sum_{j=1}^{b} \vartheta_b^{k-j} \\ &= D_b \vartheta_b^{k-b}(\vartheta_b^{b-1} + \cdots + \vartheta_b + 1) = D_b \vartheta_b^k. \end{aligned}$$

The same induction hypothesis gives $N_b(k) \ge c_b \sum_{j=1}^{b} \vartheta_b^{k-j} = c_b \vartheta_b^k$, proving the claim. Finally, since $\vartheta_b > 1$, the cumulative count satisfies

$$\begin{aligned} \sum_{k=0}^{M-1} N_b(k) &\le D_b \sum_{k=0}^{M-1} \vartheta_b^k = D_b \frac{\vartheta_b^M - 1}{\vartheta_b - 1} \\ &\le \frac{D_b}{\vartheta_b - 1} \vartheta_b^M. \end{aligned}$$

Taking $C_b = D_b \max\{1, (\vartheta_b - 1)^{-1}\}$ makes the same constant valid in both upper bounds in (42). □

Proof (Proof of Theorem 3). For $r^{ij} > 0$, dividing the local variables by r^{ij} recovers (24) with normalization $\sum_h \lambda_h = 1$. When $r^{ij} = 0$, nonnegativity forces $\lambda_h = 0$ for every h, and $\boldsymbol{Q} \succ 0$ then forces $\boldsymbol{x}^h = \mathbf{0}$. The resulting local points are exactly $\boldsymbol{x}^{ij} = \boldsymbol{z}^{ij} = \mathbf{0}$, $t^{ij} \ge 0$, as required by the closed perspective. Thus, (24) with the normalization $\sum_h \lambda_h = r^{ij}$ is an equivalent SOCP representation of (23c). The projection equality then follows from Proposition 2.

To bound the size, fix an arc $(i, j) \in A$ and let $k < M$ be the length of the reset-free interval in the definition of $\mathcal{X}^{ij}$; explicitly, $k = j - 1$ if $i = 0$ and $k = j - i - b$ otherwise. Since all support coordinates outside this interval are fixed to zero, each admissible support of $\mathcal{X}^{ij}$ is determined by a reset-free binary vector of length k. Thus, $\mathcal{X}^{ij}$ has at most $N_b(k)$ admissible support patterns. After omitting coordinates fixed to zero, each pattern requires at most k coordinates of $\boldsymbol{x}^h$, the two nonnegative scalars λ_h, t^h, and a second-order cone of dimension at most $k + 2$. The local vectors $\boldsymbol{x}^{ij}$ and $\boldsymbol{z}^{ij}$ each retain at most k coordinates. Together with t^{ij}, r^{ij}, and the aggregation and normalization constraints, these contribute an additional $O(k + 1)$ scalar variables and linear constraints. Hence the local SOCP formulation associated with (23c) has $O((k + 1)N_b(k))$ scalar variables, linear constraints, and total second-order-cone dimension.

For each k, there are at most $n + 1$ such arcs, since the starting vertex and k determine the endpoint. The global variables, flow balance equations, and full-support block contribute only $O(n)$ additional terms. Consequently, the size of $\mathcal{Y}$ is bounded by

$$O\left(n \sum_{k=0}^{M-1} (k + 1)N_b(k)\right) = O_b(nM\Phi_b(M)),$$

where we used $k + 1 \le M$ and $\sum_{k=0}^{M-1} N_b(k) = O_b(\Phi_b(M))$ from Lemma 8. This gives (25). Finally, Theorem 2 gives probabilistic exactness for the choice (19). Substituting $M \le b + c_v^{-1} \log(C_v n/\delta)$ into (25) yields (9), with $\gamma_b = \log(\vartheta_b)/c_v$. □

B Specific Distributions Satisfying the Barrier-Free Property

Corollary 1 (Explicit families satisfying the barrier-free property). *The conditions of Proposition 1 hold in each of the following cases, with the displayed choices of q, p_0, and M_q.*

1. ***Gaussian distributions.*** *Suppose that $\alpha_i \sim \mathcal{N}(0, \sigma_i^2)$ with $0 < \sigma_i \le \bar{\sigma} < \infty$ for every $i \in [n]$. One may take $q = 1$, $p_0 = 2\Phi(a_g/\bar{\sigma}) - 1$, $M_1 = \bar{\sigma}\sqrt{\frac{2}{\pi}}$, where Φ is the standard normal distribution function.*
2. ***Laplace distributions.*** *Suppose that α_i has density $f_i(x) = (2\sigma_i)^{-1}\exp(-|x|/\sigma_i)$, with $0 < \sigma_i \le \bar{\sigma} < \infty$ for every $i \in [n]$. One may take $q = 1, p_0 = 1 - \exp(-a_g/\bar{\sigma}), M_1 = \bar{\sigma}$.*
3. ***Bounded distributions with positive mass near zero.*** *Suppose that there exist constants $B > 0$ and $\pi_0 \in (0, 1]$ such that $|\alpha_i| \le B$ almost surely and $\mathbb{P}(|\alpha_i| \le a_g) \ge \pi_0$ for every $i \in [n]$. One may take $q = 1, p_0 = \pi_0, M_1 = B$.*
4. ***Cauchy distributions.*** *Suppose that $\alpha_i \sim \mathrm{Cauchy}(0, \sigma_i)$ with $0 < \sigma_i \le \bar{\sigma} < \infty$ for every $i \in [n]$. One may take $q = \frac{1}{2}, p_0 = \frac{2}{\pi}\arctan(a_g/\bar{\sigma}), M_{1/2} = \sqrt{2\bar{\sigma}}$.*
5. ***Student-t distributions.*** *Suppose that $\alpha_i = \sigma_i T_i$, where each T_i has a standard Student-t_ν distribution for a fixed $\nu > 0$, and $0 < \sigma_i \le \bar{\sigma} < \infty$ for every $i \in [n]$. For any fixed $q \in (0, \nu)$, one may take*

$$p_0 = 2F_\nu(a_g/\bar{\sigma}) - 1, \qquad M_q = \bar{\sigma}^q \nu^{q/2} \frac{\Gamma\big((q+1)/2\big)\Gamma\big((\nu-q)/2\big)}{\sqrt{\pi}\,\Gamma(\nu/2)},$$

where F_ν is the standard Student-t_ν distribution function and Γ is the gamma function.

Proof. For each distribution with chosen parameters, it suffices to verify the local-mass and moment bounds in (21).

We first start with the bounded family. The local-mass condition for this family is assumed, and the moment bound $\mathbb{E}|\alpha_i| \le B$ follows from $|\alpha_i| \le B$ almost surely.

We next consider the local-mass condition for the Gaussian, Laplace, Cauchy, and Student-t families. In each case, $\alpha_i = \sigma_i Z_i$, where the standardized variables Z_i have a common continuous distribution function F that is symmetric about zero. Since $\sigma_i \le \bar{\sigma}$,

$$\mathbb{P}(|\alpha_i| \le a_g) = 2F(a_g/\sigma_i) - 1 \ge 2F(a_g/\bar{\sigma}) - 1.$$

For the Gaussian and Student-t families, this gives the choices of p_0 directly. For the standard Laplace distribution, $2F(u) - 1 = 1 - e^{-u}$ for $u \ge 0$, while for the standard Cauchy distribution, $2F(u) - 1 = (2/\pi)\arctan u$. These identities give the remaining choices of p_0. Each is strictly positive because $a_g/\bar{\sigma} > 0$ and the corresponding density is positive in a neighborhood of zero.

We next verify the moment bounds. For a standard normal variable Z_G and a standard Laplace variable Z_L, direct integration gives

$$\mathbb{E}|Z_G| = \frac{2}{\sqrt{2\pi}} \int_0^\infty x e^{-x^2/2}\,dx = \sqrt{\frac{2}{\pi}}, \qquad \mathbb{E}|Z_L| = \int_0^\infty x e^{-x}\,dx = 1.$$

Since $\mathbb{E}|\alpha_i| = \sigma_i \mathbb{E}|Z_i|$ and $\sigma_i \le \bar{\sigma}$, the asserted Gaussian and Laplace moment bounds follow. For a standard Student-t_ν variable T and $q \in (0, \nu)$, its density gives

$$\begin{aligned}\mathbb{E}|T|^q &= \frac{2\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \int_0^\infty x^q \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} dx \\ &= \frac{\nu^{q/2}\Gamma((\nu+1)/2)}{\sqrt{\pi}\,\Gamma(\nu/2)} \int_0^\infty u^{(q-1)/2}(1+u)^{-(\nu+1)/2}\, du \\ &= \nu^{q/2} \frac{\Gamma((q+1)/2)\Gamma((\nu-q)/2)}{\sqrt{\pi}\,\Gamma(\nu/2)}.\end{aligned}$$

Here the second equality uses $u = x^2/\nu$, and the third uses Euler's beta integral. Multiplying by $\sigma_i^q \le \bar{\sigma}^q$ proves the stated Student-t moment bound.

Finally, the standard Cauchy distribution is the Student-t distribution with $\nu = 1$. Taking $q = 1/2$ in the preceding identity yields $\mathbb{E}|T|^{1/2} = \Gamma(3/4)\Gamma(1/4)/\pi = \sqrt{2}$, where the last equality follows from the gamma reflection formula. Thus $\mathbb{E}|\alpha_i|^{1/2} = \sqrt{2\sigma_i} \le \sqrt{2\bar{\sigma}}$, as claimed. □