# DOUBLY NONNEGATIVE RELAXATIONS FOR QUADRATIC AND POLYNOMIAL OPTIMIZATION PROBLEMS WITH BINARY AND BOX CONSTRAINTS

SUNYOUNG KIM*, MASAKAZU KOJIMA†, AND KIM-CHUAN TOH‡

**Abstract.** We propose doubly nonnegative (DNN) relaxations for polynomial optimization problems (POPs) with binary and box constraints to find tight lower bounds for their optimal values using a bisection and projection (BP) method. This work is an extension of the work by Kim, Kojima and Toh in 2016 from quadratic optimization problems (QOPs) to POPs. We show how the dense and sparse DNN relaxations are reduced to a simple conic optimization problem (COP) to which the BP algorithm can be applied. To compute the metric projection efficiently in the BP method, we introduce a class of polyhedral cones as a basic framework for various DNN relaxations. Moreover, we prove using the basic framework why the tight lower bounds of QOPs were obtained numerically by the Lagrangian-DNN relaxation of QOPs in the work by the authors in 2016. Preliminary numerical results on randomly generated POPs with binary and boxed constraints and the maximum complete satisfiability problems are provided to demonstrate the numerical efficiency of the proposed method for attaining tight bounds.

**Key words.** Polynomial optimization problems with nonnegative variables, doubly nonnegative relaxations, a class of polyhedral cones, the bisection and projection algorithm.

**AMS subject classifications.** 90C22, 90C25, 90C26

**1. Introduction.** Consider a polynomial optimization problem (POP): minimize a real-valued polynomial $f(\boldsymbol{x})$ in $\boldsymbol{x} \in \mathbb{R}^n$ over a basic semi-algebraic subset of $\mathbb{R}^n$. The problem serves as a fundamental nonconvex model in global optimization, notably, quadratic optimization problems (QOPs) with continuous and binary variables are its special cases. As many problems from applications, including combinatorial optimization, can be formulated as POPs with nonnegative variables, our focus is on designing efficient convex relaxation methods and algorithms for solving such POPs.

As a relaxation method for QOPs with nonnegative variables, the semidefinite programming (SDP) relaxation with the additional nonnegative constraint on the variable matrix, known as the doubly nonnegative (DNN) relaxation, were used in [10, 25]. This is a natural approach in the sense that it employs SDP relaxations [8, 21], which have proved to be very successful in solving various QOPs. The DNN relaxation approach for QOPs, however, suffers from computational inefficiency if the SDP relaxation is solved by the primal-dual interior-point method [15, 20]. The inefficiency arises mainly from the rapidly increasing sizes of the positive semidefinite matrix variables and polyhedral constraints in the DNN relaxations. More precisely, when the DNN relaxation of a QOP is converted to an SDP, the single nonnegative constraint imposed on the DNN matrix variable becomes nonnegative constraints on the elements of the matrix in the SDP, which makes the size of the SDP grow

*Department of Mathematics, Ewha W. University, 52 Ewhayeodae-gil, Seoul 120-750 Korea. The research of this author was supported by NRF 2014-R1A2A1A11049618. (skim@ewha.ac.kr)

†Department of Industrial and Systems Engineering, Chuo University, Tokyo 192-0393, Japan. This research of this author was supported by Grant-in-Aid for Scientific Research (A) 26242027 and the Japan Science and Technology Agency (JST), the Core Research of Evolutionary Science and Technology (CREST) research project. (kojimamasakazu@icloud.com)

‡Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076. Research supported in part by the Ministry of Education, Singapore, Academic Research Fund under Grant R-146-000-194-112. (mattohkc@nus.edu.sg)

quadratically with the size of the DNN matrix variable.

The SDP relaxations for QOPs can be regarded as a special case of Lasserre's SDP relaxation [17] with the lowest hierarchy applied to QOPs. The hierarchy of SDP relaxations for POPs by Lasserre [17] is a powerful method supported by the convergence result to the optimal values of POPs under moderate assumptions. However, the size of Lasserre's SDP relaxation increases exponentially with the size of a given POP and/or the degree of polynomials. An approach to mitigate this difficulty is by exploiting the sparsity in SDPs and SDP relaxations, as proposed in [9, 14, 23]. However, solving large-sized DNN relaxations of QOPs and SDP relaxations of POPs still remains a very challenging problem.

Recently, a bisection and projection (BP) algorithm was proposed by Kim, Kojima and Toh [13] (see also [4]) for a Lagrangian and doubly nonnegative (Lagrangian-DNN) relaxation, which was developed as a numerically tractable relaxation of the completely positive programming (CPP) reformulation [1, 7] of a class of QOPs with linear equality, binary and complementarity constraints in nonnegative variables. In their Lagrangian-DNN relaxation, a QOP in the class is relaxed into a simple DNN problem which minimizes a linear function in matrix $\boldsymbol{X} \in \mathbb{S}^n$ subject to the DNN constraint $\boldsymbol{X} \in \mathbb{S}^n_+ \cap \mathbb{N}^n$ and a single linear equality constraint $X_{11} = 1$, where $\mathbb{S}^n$ denotes the linear space of $n \times n$ symmetric matrices, $\mathbb{S}^n_+$ the cone of positive semidefinite matrices in $\mathbb{S}^n$ and $\mathbb{N}^n$ the cone of matrices in $\mathbb{S}^n$ with nonnegative elements. It was demonstrated through numerical results that the BP algorithm could efficiently solve large-scale Lagrangian-DNN relaxation problems. Here we extend the application of the BP algorithm [13] to DNN relaxations of POPs in this paper.

We are mainly concerned with a POP in the following form:

$$\min \{ f(\boldsymbol{x}) : \boldsymbol{x} \in S \}, \tag{1.1}$$

where $S = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \ x_i \in \{0,1\} \ (i \in I_{\rm bin}), \ x_i \in [0,1] \ (i \in I_{\rm box}) \ \right\}$, with $I_{\rm bin}$ and $I_{\rm box}$ being a partition of $\{1.\ldots,n\}$ such that $I_{\rm bin} \cup I_{\rm box} = \{1.\ldots,n\}$ and $I_{\rm bin} \cap I_{\rm box} = \emptyset$. *Binary* variables are denoted by $x_i$ $(i \in I_{\rm bin})$, and *box constrained* variables by $x_j$ $(j \in I_{\rm box})$. Obviously, a QOP with binary and box constraints is a special case of (1.1), where the degree of the objective polynomial function $f(\boldsymbol{x})$ is 2.

The purpose of this paper is to develop a basic framework where DNN relaxations of POP (1.1) are derived and the application of the BP algorithm can be performed with the efficiency and effectiveness for computing tight bounds of POP (1.1). The DNN relaxations of POP (1.1) that can be derived include the standard dense and sparse DNN relaxations of QOPs with binary and box constraints, the Lagrangian-DNN relaxation of a class of QOPs with binary and box constraints, and hierarchies of dense and sparse DNN relaxations of POPs with binary and box constraints. The last two relaxations may be regarded as variants of the hierarchy of SDP relaxation of POPs [17] and its sparse version [23], respectively. All the DNN relaxations are indeed reduced to a simple conic optimization problem (COP) of minimizing a linear function in $\boldsymbol{X} \in \mathbb{V}$ subject to $\boldsymbol{X} \in \mathbb{K}^1 \cap \mathbb{K}^2$ and a single linear equality constraint $\langle \boldsymbol{H}^0, \boldsymbol{X} \rangle = 1$. Here $\mathbb{V}$ is a finite dimensional linear space endowed with the inner product $\langle \cdot, \cdot \rangle$, $\mathbb{K}^1$ and $\mathbb{K}^2$ are closed convex cones in $\mathbb{V}$, and $\boldsymbol{H}^0 \in \mathbb{V}$. The resulting simple COP satisfies a few additional assumptions: $\mathbb{V}$ is the Cartesian product of some linear spaces of symmetric matrices, $\mathbb{K}^1$ the the Cartesian product of cones of some positive semidefinite matrices in $\mathbb{V}$, $\mathbb{K}^2$ a polyhedral cone in $\mathbb{V}$, and $\boldsymbol{H}^0 \neq \boldsymbol{O}$ lies in the dual of $\mathbb{K}^1 \cap \mathbb{K}^2$. Under these assumptions, the BP algorithm can be applied to the simple COP.

Two main issues, which are closely related, should be dealt with for our purpose. The first one is the construction of the polyhedral cone $\mathbb{K}^2$ for the simple COP. The second is the efficient computation of the metric projection from $\mathbb{K}^2$ onto $\mathbb{V}$, which is a very important step of the BP algorithm for the overall computational efficiency. To handle these two issues for the various DNN relaxations in a unified manner, we introduce a class of polyhedral cones in $\mathbb{V}$ onto which the metric projection can efficiently be computed. Then, each cone is discussed as a special case of the class of polyhedral cones, as a result, the various DNN relaxations can be derived depending on the cones. The BP algorithm applied to the DNN relaxations of POP (1.1) is shown to perform more efficiently than the primal-dual interior-point method in solving the standard DNN relaxation when it is tested on randomly generated POPs with binary and box constraints and the maximum complete satisfiability problem [11]. Another aspect of using the basic framework is that it reveals why the Lagrangian-DNN relaxation developed in [2, 13] provided a tighter bound than the standard DNN relaxation. This is the second contribution of the paper.

In Section 2, we define the simple COP in a precise form, and describe the BP and accelerated BP algorithm [4] for the COP. We also review additional concepts necessary for the subsequent discussions. Section 3 includes the main results of this paper. We define a class of polyhedral cones in $\mathbb{V}$ onto which the metric projection can be computed, and study its properties. In Sections 4 and 5, the dense and sparse DNN relaxations of POP (1.1) are derived, respectively, from the results established in Section 3. Section 6 introduces strong DNN relaxations of binary and box constrained QOPs, a full-DNN relaxation and a twin DNN relaxation, and shows that the tight lower bounds attained by the Lagrangian-DNN relaxation developed in [13] for QOPs are at least as strong as the ones provided by the two strong DNN relaxations. In Section 7, we present preliminary numerical results. We first show numerical results on the dense and sparse DNN relaxations of randomly generated binary POPs with degree 3 and 5 objective polynomials and the maximum complete satisfiability problem [11]. The performance of the accelerated BP algorithm is compared to SDPT3 [22], one of the popular primal-dual interior point methods. Then the Lagrangian-DNN relaxation of binary and box constrained QOPs mentioned in Section 7 is compared with their standard DNN relaxation mentioned in Section 4 to demonstrate the theoretical advantage of the Lagrangian-DNN relaxation. Finally, we conclude in Section 8.

## 2. Preliminaries.

**2.1. A simple conic optimization problem.** Let $\mathbb{V}$ be a finite dimensional vector space endowed with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$ such that $\|\boldsymbol{X}\| = (\langle \boldsymbol{X}, \boldsymbol{X} \rangle)^{1/2}$ for every $\boldsymbol{X} \in \mathbb{V}$. Let $\mathbb{K}^1$ and $\mathbb{K}^2$ be closed convex cones in $\mathbb{V}$ satisfying $(\mathbb{K}^1 \cap \mathbb{K}^2)^* = (\mathbb{K}^1)^* + (\mathbb{K}^2)^*$, $\boldsymbol{Q}^0 \in \mathbb{V}$ and $\boldsymbol{O} \neq \boldsymbol{H}^0 \in (\mathbb{K}^1)^* + (\mathbb{K}^2)^*$, where $\mathbb{K}^* = \{\boldsymbol{Y} \in \mathbb{V} : \langle \boldsymbol{X}, \boldsymbol{Y} \rangle \geq 0 \text{ for all } \boldsymbol{X} \in \mathbb{K}\}$ denotes the dual cone of a cone $\mathbb{K} \subset \mathbb{V}$.

We introduce a simple conic optimization problem (COP):

$$\varphi^* = \min \left\{ \langle \boldsymbol{Q}^0, \boldsymbol{X} \rangle : \langle \boldsymbol{H}^0, \boldsymbol{X} \rangle = 1, \ \boldsymbol{X} \in \mathbb{K}^1 \cap \mathbb{K}^2 \right\}, \tag{2.1}$$

and its dual

$$y_0^* = \max \left\{ y_0 : \boldsymbol{Q}^0 - \boldsymbol{H}^0 y_0 - \boldsymbol{W} = \boldsymbol{Y}, \ \boldsymbol{Y} \in (\mathbb{K}^1)^*, \ \boldsymbol{W} \in (\mathbb{K}^2)^* \right\}. \tag{2.2}$$

The primal-dual pair of COPs (2.1) and (2.2) plays a crucial role throughout the paper. All the dense and sparse DNN relaxations for QOPs and POPs with binary and box constraints discussed in Sections 4, 5 and 6 are eventually reduced to COP (2.1). Then, the accelerated bisection and projection algorithm [4], which is described as Algorithm 2.1, is applied to the primal-dual pair of COPs to solve the relaxations.

**2.2. The accelerated bisection and projection algorithm.** The bisection and projection (BP) algorithm was originally proposed in [13] as a numerical method for solving the Lagrangian-DNN relaxation of a class of QOPs with linear equality, binary and complementarity constraints in nonnegative variables. While the special case where $\mathbb{V} = \mathbb{S}^n$ (the linear space of $n \times n$ symmetric matrices), $\mathbb{K}^1 = \mathbb{S}^n_+$ (the cone of positive semidefinite matrices in $\mathbb{S}^n$) and $\mathbb{K}^1 = \mathbb{N}^n$ (the cone of nonnegative matrices in $\mathbb{S}^n$) was dealt with in [13], the algorithm in [13] could be extended to the more general COP (2.1) as in [3].

Define $\boldsymbol{G}(y_0) = \boldsymbol{Q}^0 - \boldsymbol{H}^0 y_0$ and $g(y_0) = \min\left\{\|\boldsymbol{G}(y_0) - \boldsymbol{Z}\| : \boldsymbol{Z} \in (\mathbb{K}^1)^* + (\mathbb{K}^2)^*\right\}$ for every $y_0 \in \mathbb{R}$. Then, $y_0$ is a feasible solution of (2.2) if and only if $g(y_0) = 0$. As a result, (2.2) is equivalent to $\max\{y_0 : g(y_0) = 0\}$. Recall that $\boldsymbol{H}^0 \in (\mathbb{K}^1)^* + (\mathbb{K}^2)^*$. Hence, if $g(\bar{y}_0) = 0$, or equivalently, $\boldsymbol{G}(\bar{y}_0) \in (\mathbb{K}^1)^* + (\mathbb{K}^2)^*$ for some $\bar{y}_0 \in \mathbb{R}$, then for every $y_0 \leq \bar{y}_0$, $\boldsymbol{G}(y_0) = \boldsymbol{G}(\bar{y}_0) + (\bar{y}_0 - y_0)\boldsymbol{H}^0 \in (\mathbb{K}^1)^* + (\mathbb{K}^2)^*$ or equivalently, $g(y_0) = 0$. Thus, $g(y_0) = 0$ for every $y_0 \in (-\infty, y_0^*]$. Furthermore, by Lemma 4.1 of [3], $g$ is continuously differentiable, convex and monotonically increasing on $[y_0^*, +\infty)$.

For every $\boldsymbol{G} \in \mathbb{V}$, let $\Pi(\boldsymbol{G})$ and $\Pi^*(\boldsymbol{G})$ denote the metric projection of $\boldsymbol{G}$ onto the cone $\mathbb{K}^1 \cap \mathbb{K}^2$ and its dual cone $(\mathbb{K}^1)^* + (\mathbb{K}^2)^*$, respectively, i.e.,

$$\Pi(\boldsymbol{G}) = \mathrm{argmin}\left\{\|\boldsymbol{G} - \boldsymbol{X}\| : \boldsymbol{X} \in \mathbb{K}^1 \cap \mathbb{K}^2\right\},$$
$$\Pi^*(\boldsymbol{G}) = \mathrm{argmin}\left\{\|\boldsymbol{G} - \boldsymbol{Z}\| : \boldsymbol{Z} \in (\mathbb{K}^1)^* + (\mathbb{K}^2)^*\right\}.$$

By the decomposition theorem of Moreau [19], we know that $\boldsymbol{G}(y_0) = \Pi^*(\boldsymbol{G}(y_0)) - \Pi(-\boldsymbol{G}(y_0))$. Let $\widehat{\boldsymbol{X}}(y_0)$ denote $\Pi(-\boldsymbol{G}(y_0)) \in \mathbb{K}^1 \cap \mathbb{K}^2$. Then

$$g(y_0) = \|\boldsymbol{G}(y_0) - \Pi^*(\boldsymbol{G}(y_0))\| = \|-\Pi(-\boldsymbol{G}(y_0))\| = \left\|\widehat{\boldsymbol{X}}(y_0)\right\|.$$

Next, we describe the basic idea of the BP algorithm [13]. Given an initial interval $[\ell^0, u^0]$ of lower and upper bounds for $y_0^*$, it starts with $p = 0$. First, set $y_0^{p+1} = (\ell^p + u^p)/2$. If $g(y_0^{p+1}) = \left\|\widehat{\boldsymbol{X}}(y_0^{p+1})\right\| > 0$, update the bounds such that $\ell^{p+1} = \ell^p$ and $u^{p+1} = y_0^{p+1}$. Otherwise, $g(y_0^{p+1}) = \left\|\widehat{\boldsymbol{X}}(y_0^{p+1})\right\| = 0$. Then, update the bounds such that $\ell^{p+1} = y_0^{p+1}$ and $u^{p+1} = u^p$. In either case, $y_0^* \in [\ell^{p+1}, u^{p+1}]$ holds. After replacing $p + 1$ by $p$, the iteration continues until the length of the interval becomes sufficiently small.

For the BP algorithm in [13] to be used as a reliable solution method, a few numerical issues should carefully be addressed. The first issue is on the computation of the projection $\widehat{\boldsymbol{X}}(y_0) = \Pi(-\boldsymbol{G}(y_0))$. For this computation, the accelerated proximal gradient method [5] was applied in [13]; See Section 4.2 and Algorithm C of [13]. Nevertheless, the projection $\widehat{\boldsymbol{X}}(y_0) = \Pi(-\boldsymbol{G}(y_0))$ cannot be obtained exactly for a given $y_0 \in \mathbb{R}$ with double-precision floating point arithmetic, which frequently results in a positive numerical value of $g(y_0) = \left\|\widehat{\boldsymbol{X}}(y_0)\right\|$ even when $y_0 \leq y_0^*$. The second numerical difficulty of the BP algorithm is that an initial interval $[\ell^0, u^0]$ containing $y_0^*$ is required. In many practical applications, $u^0$ is available but a reasonable $\ell^0$ is not.

For the aforementioned numerical difficulties, Arima, Kim, Kojima and Toh [4] proposed an improvement of the BP algorithm under the additional assumptions:

- We know an $\boldsymbol{I} \in \mathbb{V}$ which lies in the interior of $(\mathbb{K}^1)^*$ and a positive number $\rho$ such that $\langle \boldsymbol{I},\, \boldsymbol{X} \rangle \leq \rho$ holds for every feasible solution $\boldsymbol{X}$ of COP (2.1).
- For any $\boldsymbol{A} \in \mathbb{V}$, it is easy to compute $\mu(\boldsymbol{A}) = \sup\{\mu : \boldsymbol{A} - \mu \boldsymbol{I} \in (\mathbb{K}^1)^*\}$. (Since $\boldsymbol{I}$ lies in the interior of $(\mathbb{K}^1)^*$, $\mu(\boldsymbol{A})$ is finite for any $\boldsymbol{A} \in \mathbb{V}$).

The problem with $\mathbb{V} = \mathbb{S}^n$, $\mathbb{K}^1 = \mathbb{S}^n_+$, $\mathbb{K}^2 = \mathbb{N}^n$ and $\boldsymbol{I} =$ the $n \times n$ identity matrix was considered in [4], where $\mu(\boldsymbol{A})$ coincides with the minimum eigenvalue of $\boldsymbol{A} \in \mathbb{S}^n$. Under the assumptions previously mentioned, COP (2.1) is equivalent to

$$\varphi^* = \min\left\{\langle \boldsymbol{Q}^0,\, \boldsymbol{X}\rangle : \langle \boldsymbol{H}^0,\, \boldsymbol{X}\rangle = 1,\ \langle \boldsymbol{I},\, \boldsymbol{X}\rangle \leq \rho,\ \boldsymbol{X} \in \mathbb{K}^1 \cap \mathbb{K}^2\right\}.$$

For the computation of a valid lower bound for $\varphi^*$, we consider the dual:

$$\max\left\{y_0 + \rho t : \boldsymbol{Q}^0 - \boldsymbol{H}^0 y_0 - \boldsymbol{I}t - \boldsymbol{W} = \boldsymbol{Y},\ \boldsymbol{Y} \in (\mathbb{K}^1)^*,\ \boldsymbol{W} \in (\mathbb{K}^2)^*,\ t \leq 0\right\} \tag{2.3}$$

Suppose that a decomposition $\widehat{\boldsymbol{W}}(y_0^p) + \widehat{\boldsymbol{Y}}(y_0^p) - \widehat{\boldsymbol{X}}(y_0^p)$ of $\boldsymbol{G}(y_0^p)$ is computed approximately at the $p$th iteration of the BP algorithm, specifically, that $\boldsymbol{G}(y_0^p)$ is nearly equal to $\widehat{\boldsymbol{W}}(y_0^p) + \widehat{\boldsymbol{Y}}(y_0^p) - \widehat{\boldsymbol{X}}(y_0^p)$ with $\widehat{\boldsymbol{W}}(y_0^p) \in (\mathbb{K}^2)^*$. We note that either $\widehat{\boldsymbol{Y}}(y_0^p) \in (\mathbb{K}^1)^*$ or $\widehat{\boldsymbol{X}}(y_0^p) \in \mathbb{K}^1 \cap \mathbb{K}^2$ is not required at this point. Let

$$\begin{array}{rcl} t^p & = & \min\{\mu(\boldsymbol{Q}^0 - \boldsymbol{H}^0 y_0^p - \widehat{\boldsymbol{W}}(y_0^p)), 0\}, \\ \widetilde{\boldsymbol{Y}}^p & = & \boldsymbol{Q}^0 - \boldsymbol{H}^0 y_0^p - \widehat{\boldsymbol{W}}(y_0^p) - \boldsymbol{I}t^p. \end{array} \tag{2.4}$$

Then $t^p \leq 0$ and $\widetilde{\boldsymbol{Y}}^p \in (\mathbb{K}^1)^*$. As a result, $y_0^p + \rho t^p$ provides a valid lower bound for $\varphi^*$ by the weak duality.

The BP algorithm incorporating this technique can generate a valid lower bound $\underline{\ell}^{p+1} = y_0^p + \rho t^p$ for $\varphi^*$ at each iteration, even if the algorithm starts with $\ell^0 = -\infty$. The numerical results in [4] showed that this technique also served to accelerate the BP algorithm. In the numerical results in Section 7, the accelerated BP algorithm has been used.

To describe the accelerated BP algorithm, we denote the numerical value of $\left\|\widehat{\boldsymbol{X}}(y_0)\right\| / \max\{1, \|\boldsymbol{G}(y_0)\|\}$by $h(y_0)$.

ALGORITHM 2.1. (Accelerated BP Algorithm)

Step 0. Choose positive numbers $\epsilon$ and $\delta$ sufficiently small (e.g., $\epsilon = 1.0$ e-11 and $\delta = 1.0$ e-4). Here $\delta$ determines the target length of an interval $[\ell^p,\ u^p] \subset \mathbb{R}$ which is expected to contain $y_0^*$. Let $p = 0$.

Step 1. Find a $u^0 \in \mathbb{R}$ such that $y_0^* \leq u^0$. Let $\ell^0 = \underline{\ell}^0 = -\infty$. Choose $y_0^0 \leq u^0$.

Step 2. If $u^p - \ell^p < \delta \max\{1, |\ell^p|, |u^p|\}$, output $\underline{\ell}^p$ as a lower bound for $y_0^*$. Otherwise, compute a decomposition $\boldsymbol{G}(y_0^p) = \widehat{\boldsymbol{W}}(y_0^p) + \widehat{\boldsymbol{Y}}(y_0^p) - \widehat{\boldsymbol{X}}(y_0^p)$.

Step 3. Take $t^p \in \mathbb{R}$ and $\widetilde{\boldsymbol{Y}}^p \in \mathbb{V}$ as in (2.4). Let $\underline{\ell}^{p+1} = \max\{\underline{\ell}^p, y_0^p + \rho t^p\}$. If $h(y_0^p) \leq \epsilon$, then let $\ell^{p+1} = y_0^p$ and $u^{p+1} = u^p$. Otherwise, let $\ell^{p+1} = \max\{\underline{\ell}^{p+1},\ \ell^p\}$ and $u^{p+1} = y_0^p$.

Step 4. Let $y_0^{p+1} = (\ell^{p+1} + u^{p+1})/2$. Replace $p + 1$ by $p$ and go to Step 2.

See Section 3 of [4] and Section 4 of [13] for more details.

In the simple COP (2.1), $\mathbb{K}^1$ is usually the cone (or the Cartesian product of cones) of positive semidefinite matrices as in Section 4, 5 and 6, and the metric projection $\Pi_{\mathbb{K}^1}(\boldsymbol{Z})$ of each $\boldsymbol{Z} \in \mathbb{V}$ onto $\mathbb{K}^1$ can be computed by the eigenvalue decomposition. Thus the only remaining important step is the metric projection onto the cone $\mathbb{K}^2$, which should be carried out efficiently for the overall computational efficiency. In Section 3, we present a class of polyhedral cones onto which the metric projection can efficiently be computed.

**2.3. QOPs with binary and box constraints and their standard DNN relaxations.** Any quadratic function in $\boldsymbol{x} \in \mathbb{R}^n$ can be written as $\langle \boldsymbol{Q}^0,\, (1,\boldsymbol{x})^T(1,\boldsymbol{x})\rangle$, where $\boldsymbol{x}$ is an $n$-dimensional row vector and $\boldsymbol{Q}^0 \in \mathbb{S}^{1+n}$. As a special case of (1.1), we consider the QOP with binary and box constraints

$$\min\left\{\langle \boldsymbol{Q}^0,\, (1,\boldsymbol{x})^T(1,\boldsymbol{x})\rangle : \boldsymbol{x} \in S\right\}. \tag{2.5}$$

Introducing a symmetric matrix variable $\begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \in \mathbb{S}^{1+n}$, and a subset

$$T = \left\{ \begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \in \mathbb{S}^{1+n} : \boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^T,\ \boldsymbol{x} \in S,\ X_{00} = 1 \right\},$$

we rewrite (2.5) as

$$\min\left\{ \left\langle \boldsymbol{Q}^0, \begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \right\rangle : \begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \in T \right\}. \tag{2.6}$$

Thus QOP (2.5) in $\mathbb{R}^n$ is *lifted* to the equivalent problem (2.6) in $\mathbb{S}^{1+n}$. Between a feasible solution $\boldsymbol{x} \in \mathbb{R}^n$ of (2.5) and a feasible solution $\begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \in \mathbb{S}^{1+n}$ of (2.6), the following correspondence holds:

$$1 \leftrightarrow X_{00},\ x_i \leftrightarrow x_i\ (i = 1,\ldots,n),\ x_ix_j \leftrightarrow X_{ij}\ (i = 1,\ldots,n, j = 1,\ldots,n). \tag{2.7}$$

To derive a DNN relaxation of QOP (2.5), we relax the feasible region $T$ of the lifted problem (2.6) to a convex subset of $\mathbb{S}^{1+n}$, which is described as the intersection of a hyperplane and two closed convex cones $\mathbb{K}^1$, $\mathbb{K}^2$ as in (2.1). By construction, if $\begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \in T$ then $\begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \in \mathbb{S}^{1+n}_+$, $\begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \geq \boldsymbol{O}$, $\left\langle \boldsymbol{H}^0, \begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \right\rangle = X_{00} = 1$. Here $\boldsymbol{H}^0$ denotes the $(1+n)\times(1+n)$ matrix in $\mathbb{S}^{1+n}$ with 1 as the $(0,0)$th element and 0 elsewhere. It follows from $\boldsymbol{x} \in S$ that the identity $x_i = x_i^2$ $(i \in I_{\rm bin})$ and the inequality $x_j \geq x_j^2$ $(j \in I_{\rm box})$ hold for every $\boldsymbol{x} \in S$. Since $x_i^2$ corresponds to $X_{ii}$ $(i = 1,\ldots,n)$, the identity and inequality induce $x_i = X_{ii}$ $(i \in I_{\rm bin})$ and $x_j \geq X_{jj}$ $(j \in I_{\rm box})$, respectively. By defining

$$\mathbb{L} = \left\{ \begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \in \mathbb{N}^{1+n} : x_i = X_{ii}\ (i \in I_{\rm bin}),\ x_j \geq X_{jj}\ (j \in I_{\rm box}) \right\}, \tag{2.8}$$

which forms a polyhedral cone in $\mathbb{S}^{1+n}$, we see that every $\begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \in T$ lies in $\mathbb{L}$. Here $\mathbb{N}^{1+n}$ denotes the cone of $(1+n)\times(1+n)$ symmetric matrices with all elements nonnegative. Therefore, COP (2.1) with $\mathbb{K}^1 = \mathbb{S}^{1+n}_+$ and $\mathbb{K}^2 = \mathbb{L}$ serves as the *standard DNN relaxation* of QOP (2.5).

**2.4. Notation and symbols for POPs and their DNN relaxations.** For the extension of the discussion in Section 2.3 to a general POP (1.1) with binary and box constraints and its dense and sparse DNN relaxations in the subsequent sections, we introduce the following notation and symbols.

Let $\mathbb{Z}^n_+$ denote the set of $n$-dimensional nonnegative integer vectors. For each $\boldsymbol{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n$ and $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_n) \in \mathbb{Z}^n_+$, let $\boldsymbol{x}^{\boldsymbol{\alpha}} = x^{\boldsymbol{\alpha}_1}\cdots x^{\boldsymbol{\alpha}_n}$ denote a *monomial.* We call $\deg(\boldsymbol{x}^{\boldsymbol{\alpha}}) = \max\{\alpha_i : i = 1, \dots, n\}$ the *degree* of a monomial $\boldsymbol{x}^{\boldsymbol{\alpha}}$. Each polynomial $f(\boldsymbol{x})$ is represented as $f(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{F}} c(\boldsymbol{\alpha}) \boldsymbol{x}^{\boldsymbol{\alpha}}$ for some nonempty finite subset $\mathcal{F}$ of $\mathbb{Z}^n_+$ and $c(\boldsymbol{\alpha}) \in \mathbb{R}$ ($\boldsymbol{\alpha} \in \mathcal{F}$). We call $\text{supp} f = \{\boldsymbol{\alpha} \in \mathcal{F} : c(\boldsymbol{\alpha}) \neq 0\}$ the *support* of $f(\boldsymbol{x})$; hence $f(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \text{supp} f} c(\boldsymbol{\alpha}) \boldsymbol{x}^{\boldsymbol{\alpha}}$ is the minimal representation of $f(\boldsymbol{x})$. We call $\deg f = \max\{\deg(\boldsymbol{x}^{\boldsymbol{\alpha}}) : \boldsymbol{\alpha} \in \text{supp} f\}$ the *degree* of $f(\boldsymbol{x})$.

Let $\mathcal{A}$ be a nonempty finite subset of $\mathbb{Z}^n_+$ with cardinality $|\mathcal{A}|$, and let $\mathbb{S}^{\mathcal{A}}$ denote the linear space of $|\mathcal{A}| \times |\mathcal{A}|$ symmetric matrices whose rows and columns are indexed by $\mathcal{A}$. The $(\boldsymbol{\alpha}, \boldsymbol{\beta})$th component of each $\boldsymbol{X} \in \mathbb{S}^{\mathcal{A}}$ is written as $X_{\boldsymbol{\alpha\beta}}$ ($(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{A} \times \mathcal{A}$). The *inner product* of $\boldsymbol{X},\ \boldsymbol{Y} \in \mathbb{S}^{\mathcal{A}}$ is defined by $\langle \boldsymbol{X},\, \boldsymbol{Y} \rangle = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \sum_{\boldsymbol{\beta} \in \mathcal{A}} X_{\boldsymbol{\alpha\beta}} Y_{\boldsymbol{\alpha\beta}}$, and the *norm* of $\boldsymbol{X} \in \mathbb{S}^{\mathcal{A}}$ by $\|\boldsymbol{X}\| = (\langle \boldsymbol{X},\, \boldsymbol{X} \rangle)^{1/2}$. We denote a $|\mathcal{A}|$-dimensional row vector of monomials $\boldsymbol{x}^{\boldsymbol{\alpha}}$ ($\boldsymbol{\alpha} \in \mathcal{A}$) by $\boldsymbol{x}^{\mathcal{A}}$, and a $|\mathcal{A}| \times |\mathcal{A}|$ symmetric matrix $(\boldsymbol{x}^{\mathcal{A}})^T (\boldsymbol{x}^{\mathcal{A}})$ of monomials $\boldsymbol{x}^{\boldsymbol{\alpha} + \boldsymbol{\beta}}$ ($(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{A} \times \mathcal{A}$) by $\boldsymbol{x}^{\square \mathcal{A}} \in \mathbb{S}^{\mathcal{A}}$. We call $\boldsymbol{x}^{\square \mathcal{A}}$ a *moment matrix.*

For pair of subsets $\mathcal{A}$ and $\mathcal{B}$ of $\mathbb{Z}^n_+$, let $\mathcal{A} + \mathcal{B} = \{\boldsymbol{\alpha} + \boldsymbol{\beta} : \boldsymbol{\alpha} \in \mathcal{A},\ \boldsymbol{\beta} \in \mathcal{B}\}$ denote their Minkowski sum. Let $\mathbb{S}^{\mathcal{A}}_+$ denote the cone of positive semidefinite matrices in $\mathbb{S}^{\mathcal{A}}$, and $\mathbb{N}^{\mathcal{A}}$ the cone of nonnegative matrices in $\mathbb{S}^{\mathcal{A}}$. By construction, $\boldsymbol{x}^{\square \mathcal{A}} \in \mathbb{S}^{\mathcal{A}}_+$ for every $\boldsymbol{x} \in \mathbb{R}^n$, and $\boldsymbol{x}^{\square \mathcal{A}} \in \mathbb{N}^{\mathcal{A}}$ for every $\boldsymbol{x} \in \mathbb{R}^n_+$.

**2.5. Lifting POP (1.1) in $\mathbb{R}^n$ to the space of higher dimensional symmetric matrices.** Define $\boldsymbol{r} : \mathbb{Z}^n_+ \to \mathbb{Z}^n_+$ by

$$r_i(\boldsymbol{\alpha}) = \begin{cases} 0 & \text{if } i \in I_{\text{bin}} \ \text{ and } \alpha_i = 0, \\ 1 & \text{if } i \in I_{\text{bin}} \ \text{ and } \alpha_i > 0, \\ \alpha_i & \text{otherwise } (i.e.,\ i \in I_{\text{box}}). \end{cases} \tag{2.9}$$

Then, for each monomial $\boldsymbol{x}^{\boldsymbol{\alpha}}$ with $\boldsymbol{\alpha} \in \mathbb{Z}^n_+$ and $\boldsymbol{x} \in S$, we observe that

$$\boldsymbol{x}^{\boldsymbol{\alpha}} = \Pi_{i \in I_{\text{bin}}} x_i^{\alpha_i} \Pi_{j \in I_{\text{box}}} x_j^{\alpha_j} \ = \ \Pi_{i \in I_{\text{bin}}} x_i^{r_i(\boldsymbol{\alpha})} \Pi_{j \in I_{\text{box}}} x_j^{\alpha_j} \ = \ \boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\alpha})}.$$

Thus, each monomial $\boldsymbol{x}^{\boldsymbol{\alpha}}$ in the objective function $f(\boldsymbol{x})$ of POP (1.1) with binary and box constraints may be replaced by $\boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\alpha})}$. We may assume without loss of generality that $\text{supp}(f) = \boldsymbol{r}(\text{supp}(f))$.

To construct a doubly nonnegative (DNN) relaxation of POP (1.1), we first decide a nonempty finite subset $\mathcal{A}$ of $\mathbb{Z}^n_+$ satisfying the condition

$$\boldsymbol{0} \in \mathcal{A} = \boldsymbol{r}(\mathcal{A}) \text{ and } \text{supp} f = \boldsymbol{r}(\text{supp}(f)) \subset \mathcal{A} + \mathcal{A}. \tag{2.10}$$

By choosing a $|\mathcal{A}| \times |\mathcal{A}|$ matrix $\boldsymbol{Q}^0 \in \mathbb{S}^{\mathcal{A}}$ such that $f(\boldsymbol{x}) = \langle \boldsymbol{Q}^0,\, \boldsymbol{x}^{\square \mathcal{A}} \rangle$, we rewrite POP (1.1) as

$$\min\{\langle \boldsymbol{Q}^0,\, \boldsymbol{X} \rangle : \boldsymbol{X} \in T\}, \tag{2.11}$$

where $T = \left\{ \boldsymbol{x}^{\square \mathcal{A}} \in \mathbb{S}^{\mathcal{A}} : \boldsymbol{x} \in S \right\}$. As mentioned in Section 2.4, we will relax $T$ to a convex subset of $\mathbb{S}^{\mathcal{A}}$, which is the intersection of a hyperplane and two polyhedral cones $\mathbb{K}^1$, $\mathbb{K}^2$ in $\mathbb{S}^{\mathcal{A}}$.

**3. A class of polyhedral cones and the metric projection onto them.** Before describing the class of polyhedral cones we are interested in, we first discuss why the cones become essential for the subsequent discussions. Recall that POP (1.1) in the $n$-dimensional space $\mathbb{R}^n$ has been lifted to the problem (2.11) in the symmetric matrix space $\mathbb{S}^{\mathcal{A}}$. The two problems are equivalent under the correspondence $\mathbb{S}^{\mathcal{A}} \ni \boldsymbol{x}^{\square \mathcal{A}} \leftrightarrow \boldsymbol{X} \in \mathbb{S}^{\mathcal{A}}$, or componentwisely $(\boldsymbol{x}^{\square \mathcal{A}})_{\boldsymbol{\alpha\beta}} = \boldsymbol{x}^{\boldsymbol{\alpha} + \boldsymbol{\beta}} \leftrightarrow X_{\boldsymbol{\alpha\beta}}$ $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{A} \times \mathcal{A}$. In addition to that, many identities and inequalities hold among elements $(\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\alpha\beta}} = \boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}$ of $\boldsymbol{x}^{\square\mathcal{A}}$ $((\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathcal{A}\times\mathcal{A})$ for all $\boldsymbol{x} \in S$, which are translated to equalities and inequalities in $X_{\boldsymbol{\alpha\beta}}$ $(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathcal{A}\times\mathcal{A}$.

In particular, if the relation $\boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\beta}) = \boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\delta})$ holds, then $(\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\alpha\beta}} = \boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}} = (\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\gamma\delta}} = \boldsymbol{x}^{\boldsymbol{\gamma}+\boldsymbol{\delta}}$. Thus, the relation $\boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\beta}) = \boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\delta})$ naturally induces an equivalence relation $\sim$ in $\mathcal{A}\times\mathcal{A}$ such that $(\boldsymbol{\alpha},\boldsymbol{\beta}) \sim (\boldsymbol{\gamma},\boldsymbol{\delta})$ if and only if $\boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\beta}) = \boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\gamma})$ holds. With this equivalence relation $\sim$, a common nonnegative value can be assigned to $X_{\boldsymbol{\alpha\beta}}$ for all $(\boldsymbol{\alpha},\boldsymbol{\beta})$ in each equivalence class. The use of the equivalence class in this way considerably simplifies the description of the polyhedral cone $\mathbb{L}$ used for $\mathbb{K}^2$ in the DNN relaxation (2.1).

Translating the inequalities in $(\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\alpha\beta}}$ $(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathcal{A}\times\mathcal{A}$ to the ones in $X_{\boldsymbol{\alpha\beta}}$ $(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathcal{A}\times\mathcal{A}$ for the construction of $\mathbb{L}$ is not as straightforward as in the case for equalities. In fact, only a subset of the inequalities in $(\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\alpha\beta}}$ is translated to the ones for $\mathbb{L}$. If all the inequalities from $(\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\alpha\beta}}$ $(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathcal{A}\times\mathcal{A}$ are included in the resulting cone $\mathbb{L}$, then the computation of the metric projection from $\mathbb{S}^{\mathcal{A}}$ onto $\mathbb{L}$ may be neither efficient nor accurate. Thus, the resulting cone $\mathbb{L}$ should be constructed with the simplest possible structure for efficient and accurate computation of the metric projection.

For this purpose, a chain of equivalence classes $[(\boldsymbol{\alpha}^1,\boldsymbol{\beta}^1)],\ldots,[(\boldsymbol{\alpha}^m,\boldsymbol{\beta}^m)]$ along which the chain of inequalities $(\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\alpha}_1\boldsymbol{\beta}_1} \geq \cdots \geq (\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\alpha}_m\boldsymbol{\beta}_m}$ is satisfied for every $\boldsymbol{x} \in S$ is defined by introducing a preorder $\succeq$ in $\mathcal{A}\times\mathcal{A}$. The inequalities generated from a family of disjoint chains are used to construct $\mathbb{L}$. The cone $\mathbb{L}$ constructed this way is essential for the algorithm in Section 3.4, as the most crucial step of the algorithm is computing the metric projection onto $\mathbb{L}$.

Suppose that $S = [0,1]^2$. Then $x_1 \geq x_1^2 \geq x_1^2x_2,\ x_2 \geq x_2^2 \geq x_1x_2^2,\ x_1x_2 \geq x_1^2x_2^2$ form a family of three disjoint chains of inequalities, which can be translated to the corresponding inequalities in $X_{\boldsymbol{\alpha\beta}}$ $((\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathcal{A}\times\mathcal{A})$ to construct $\mathbb{L}$. But an inequality $x_1x_2 \geq x_1x_2^2$ cannot be added since it is not disjoint with the last two chains. As for the case when $x_1 \in \{0,1\}$ and $x_2 \in [0,1]$, equivalence classes should be considered more carefully. For example, $x_1^2x_2$ and $x_1x_2$ are monomials in an equivalence class since $\boldsymbol{r}((2,1)) = \boldsymbol{r}((1,1))$. Thus, the family of chains above itself is not disjoint.

Throughout this section, for simplicity, the symmetric matrix space $\mathbb{S}^{\mathcal{A}}$ is identified with the $|\mathcal{A}|\times|\mathcal{A}|$-dimensional Euclidean space. Although the symmetry is lost by this generalization, it does not affect the construction of the class of polyhedral cones. Indeed, this generalization enables us to apply the result and the method in this section directly to the sparse DNN relaxation of POP (1.1) in Section 5, and to the discussion related to the Lagrangian-DNN relaxation derived from the CPP reformulation of a class of QOPs with equality, binary and complementarity constraints [13] in Section 6. This is another motivation for the generalization.

**3.1. Preorder, equivalence relations and a chain of equivalence classes.** Let $\boldsymbol{\Theta}$ be a nonempty finite set, and $\mathbb{R}^{\boldsymbol{\Theta}}$ be the $|\boldsymbol{\Theta}|$-dimensional Euclidean space of row vectors $\boldsymbol{X}$ whose elements are indexed by $\boldsymbol{\Theta}$; the $\theta$th element of $\boldsymbol{X}$ is denoted by $X_\theta$ $(\theta \in \boldsymbol{\Theta})$. Let $\mathbb{R}^{\boldsymbol{\Theta}}_+$ denote the nonnegative orthant $\{\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{\Theta}} : X_{\boldsymbol{\theta}} \geq 0 \text{ for every } \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ of $\mathbb{R}^{\boldsymbol{\Theta}}$. We call a binary relation $\succeq$ in $\boldsymbol{\Theta}$ a *preorder* if it satisfies

$$\boldsymbol{\theta} \succeq \boldsymbol{\theta} \text{ for every } \boldsymbol{\theta} \in \boldsymbol{\Theta} \ (\textit{reflexive}),$$
$$\boldsymbol{\theta}^1 \succeq \boldsymbol{\theta}^2 \text{ and } \boldsymbol{\theta}^2 \succeq \boldsymbol{\theta}^3 \Rightarrow \boldsymbol{\theta}^1 \succeq \boldsymbol{\theta}^3 \ (\textit{transitive}).$$

The preorder $\succeq$ induces a *strict preorder* $\succ$ and an *equivalence relation* $\sim$:

$$\begin{array}{lll} \boldsymbol{\theta} \succ \boldsymbol{\eta} & \Leftrightarrow & \boldsymbol{\theta} \succeq \boldsymbol{\eta} \text{ and } \boldsymbol{\eta} \not\succeq \boldsymbol{\theta}, \\ \boldsymbol{\theta} \sim \boldsymbol{\eta} & \Leftrightarrow & \boldsymbol{\theta} \succeq \boldsymbol{\eta} \text{ and } \boldsymbol{\eta} \succeq \boldsymbol{\theta}. \end{array} \tag{3.1}$$

Let $[\boldsymbol{\theta}]$ denote the *equivalence class* which contains $\boldsymbol{\theta} \in \boldsymbol{\Theta}$; $[\boldsymbol{\theta}] = \{\boldsymbol{\eta} \in \boldsymbol{\Theta} : \boldsymbol{\eta} \sim \boldsymbol{\theta}\}$. We can consistently use $\succeq$ and $\succ$ in the family of equivalence classes $\{[\boldsymbol{\theta}] : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ such that $[\boldsymbol{\theta}] \succeq [\boldsymbol{\eta}]$ (or $[\boldsymbol{\theta}] \succ [\boldsymbol{\eta}]$) if $\boldsymbol{\theta} \succeq \boldsymbol{\eta}$ (or $\boldsymbol{\theta} \succ \boldsymbol{\eta}$). In fact, $\succeq$ acts as a *partial order* in the family, which is reflexive, transitive, and *antisymmetry, i.e.*, $[\boldsymbol{\theta}] \succeq [\boldsymbol{\eta}]$ and $[\boldsymbol{\eta}] \succeq [\boldsymbol{\theta}] \Rightarrow [\boldsymbol{\theta}] = [\boldsymbol{\eta}]$. An equivalence class is frequently denoted by $E_{\boldsymbol{\sigma}}$, where $\boldsymbol{\sigma}$ means a representative element of the equivalence class or a symbol attached to the class.

A finite sequence of equivalence classes $E_{\boldsymbol{\sigma}^1}, \ldots, E_{\boldsymbol{\sigma}^\ell}$ for $\ell \geq 1$ is called a *chain* if $E_{\boldsymbol{\sigma}^1} \succ \cdots \succ E_{\boldsymbol{\sigma}^\ell}$. In particular, a single equivalence class itself is a chain. For simplicity of notation, each chain $E_{\boldsymbol{\sigma}^1}, \ldots, E_{\boldsymbol{\sigma}^\ell}$ is identified with the family of equivalence classes $\{E_{\boldsymbol{\sigma}^1}, \ldots, E_{\boldsymbol{\sigma}^\ell}\}$. A chain $\Gamma$ is *maximal* if it is not a proper subfamily of any other chain. Two chains $\Gamma^1$ and $\Gamma^2$ are *disjoint* if $\Gamma^1 \cap \Gamma^2 = \emptyset$.

**3.2. A class of polyhedral cones.** Let $\{\Gamma_1, \ldots, \Gamma_r\}$ be a family of chains of equivalence classes. Define

$$\mathbb{L} = \left\{ \boldsymbol{X} \in \mathbb{R}^{\boldsymbol{\Theta}}_+ : \begin{array}{l} X_{\boldsymbol{\theta}} = X_{\boldsymbol{\eta}} \text{ if } \boldsymbol{\theta} \sim \boldsymbol{\eta}, \\ X_{\boldsymbol{\theta}} \geq X_{\boldsymbol{\eta}} \text{ if } \Gamma_p \ni [\boldsymbol{\theta}] \succ [\boldsymbol{\eta}] \in \Gamma_p \ (p = 1, \ldots, r) \end{array} \right\}.$$

Obviously, $\mathbb{L}$ forms a polyhedral cone in $\mathbb{R}^{\boldsymbol{\Theta}}$.

Remark 3.1. As a special case of families $\{\Gamma_1, \ldots, \Gamma_r\}$ of chains of equivalence classes, we can consider "the finest family" $\{\{[\boldsymbol{\theta}], [\boldsymbol{\tau}]\} : \boldsymbol{\theta} \in \boldsymbol{\Theta},\ \boldsymbol{\tau} \in \boldsymbol{\Theta},\ \boldsymbol{\theta} \succ \boldsymbol{\tau}\}$ to impose all the inequalities induced from the preorder $\succeq$ on $\mathbb{L}$. In this case, the resulting "smallest" cone $\mathbb{L}$ is of the form

$$\mathbb{L} = \left\{ \boldsymbol{X} \in \mathbb{R}^{\boldsymbol{\Theta}}_+ : \begin{array}{l} X_{\boldsymbol{\theta}} = X_{\boldsymbol{\eta}} \text{ if } \boldsymbol{\theta} \sim \boldsymbol{\eta}, \\ X_{\boldsymbol{\theta}} \geq X_{\boldsymbol{\eta}} \text{ if } [\boldsymbol{\theta}] \succ [\boldsymbol{\eta}] \end{array} \right\}.$$

The finest family, however, is not necessarily disjoint. We impose the disjoint property on the family $\{\Gamma_1, \ldots, \Gamma_r\}$ in the next section for efficient computation of the metric projection from $\mathbb{R}^{\boldsymbol{\Theta}}$ onto $\mathbb{L}$.

Since $[\boldsymbol{\theta}]$ itself is a chain of equivalence classes ($\boldsymbol{\theta} \in \boldsymbol{\Theta}$), we may assume without loss of generality that the family $\{\Gamma_1, \ldots, \Gamma_r\}$ covers the family of equivalence classes $\{[\boldsymbol{\theta}] : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ such that

$$\bigcup_{p=1}^{r} \Gamma_p = \{[\boldsymbol{\theta}] : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}. \tag{3.2}$$

Then the cone $\mathbb{L}$ can be rewritten as

$$\mathbb{L} = \left\{ \boldsymbol{X} \in \mathbb{R}^{\boldsymbol{\Theta}} : \begin{array}{l} X_{\boldsymbol{\theta}} = \xi_{\boldsymbol{\sigma}} \in \mathbb{R}_+ \text{ if } \boldsymbol{\theta} \in E_{\boldsymbol{\sigma}} \in \Gamma_p \ (p = 1, \ldots, r), \\ \xi_{\boldsymbol{\sigma}} \geq \xi_{\boldsymbol{\tau}} \text{ if } \Gamma_p \ni E_{\boldsymbol{\sigma}} \succ E_{\boldsymbol{\tau}} \in \Gamma_p \ (p = 1, \ldots, r) \end{array} \right\}. \tag{3.3}$$

If the chain $\Gamma_p$ is represented as

$$\Gamma_p = \{E_{\boldsymbol{\sigma}^1}, \ldots, E_{\boldsymbol{\sigma}^\ell}\}, \text{ where } E_{\boldsymbol{\sigma}^1} \succ \cdots \succ E_{\boldsymbol{\sigma}^\ell} \tag{3.4}$$

for a fixed $p \in \{1, \ldots, r\}$, the inequalities $\xi_{\boldsymbol{\sigma}} \geq \xi_{\boldsymbol{\tau}}$ for $\Gamma_p \ni E_{\boldsymbol{\sigma}} \succ E_{\boldsymbol{\tau}} \in \Gamma_p$ are written as $\xi_{\boldsymbol{\sigma}^j} \geq \xi_{\boldsymbol{\sigma}^{j+1}}$ $(j = 1, \ldots, \ell - 1)$. Specifically, when $\Gamma_p$ consists of a single equivalence class $E_{\boldsymbol{\sigma}^1}$, these inequalities vanish.

**3.3. The metric projection from $\mathbb{R}^{\boldsymbol{\Theta}}$ onto $\mathbb{L}$.** Now we consider the metric projection $\Pi_{\mathbb{L}}$ from $\mathbb{R}^{\boldsymbol{\Theta}}$ onto $\mathbb{L}$. First, we show a fundamental property of $\Pi_{\mathbb{L}}$. Let $\boldsymbol{Z} \in \mathbb{R}^{\boldsymbol{\Theta}}$ and $\overline{\boldsymbol{X}} = \Pi_{\mathbb{L}}(\boldsymbol{Z})$. By definition, $\overline{\boldsymbol{X}}$ is the unique optimal solution of the strictly convex quadratic optimization problem:

$$\begin{aligned}
&\min\left\{\sum_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} (X_{\boldsymbol{\theta}} - Z_{\boldsymbol{\theta}})^2 : \boldsymbol{X} \in \mathbb{L}\right\} \\
&= \min\left\{\sum_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} (\xi_{\boldsymbol{\sigma}} - Z_{\boldsymbol{\theta}})^2 : \begin{array}{l} X_{\boldsymbol{\theta}} = \xi_{\boldsymbol{\sigma}} \in \mathbb{R}_+ \text{ if } \boldsymbol{\theta} \in E_{\boldsymbol{\sigma}} \in \Gamma_p \ (p = 1,\ldots,r), \\ \xi_{\boldsymbol{\sigma}} \geq \xi_{\boldsymbol{\tau}} \text{ if } \Gamma_p \ni E_{\boldsymbol{\sigma}} \succ E_{\boldsymbol{\tau}} \in \Gamma_p \ (p = 1,\ldots,r) \end{array}\right\}. \quad (3.5)
\end{aligned}$$

To characterize $\overline{\boldsymbol{X}} = \Pi_{\mathbb{L}}(\boldsymbol{Z})$ in a compact way, we use the notation $\mu(\boldsymbol{Z}, E) = \left(\sum_{\boldsymbol{\theta}\in E} Z_{\boldsymbol{\theta}}\right) / |E|$ to denote the average of $Z_{\boldsymbol{\theta}}$ $(\boldsymbol{\theta} \in E)$, and $\mu_+(\boldsymbol{Z}, E) = \max\{\mu(\boldsymbol{Z}, E), \boldsymbol{0}\}$ for every $\boldsymbol{Z} \in \mathbb{R}^{\boldsymbol{\Theta}}$ and $E \subset \boldsymbol{\Theta}$.

LEMMA 3.2. *Let $\boldsymbol{Z} \in \mathbb{R}^{\boldsymbol{\Theta}}$ and $\overline{\boldsymbol{X}} = \Pi_{\mathbb{L}}(\boldsymbol{Z})$. Assume that the family of chains $\Gamma_1, \ldots, \Gamma_r$ is a partition of the family of equivalence classes $\{[\boldsymbol{\theta}] : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, i.e., assume that $\Gamma_i \bigcap \Gamma_j = \emptyset$ $(i \neq j)$ in addition to* (3.2)*. For an arbitrary fixed $p \in \{1, \ldots, r\}$, denote the chain $\Gamma_p$ as in* (3.4)*.*

*(a) If $\ell = 1$ (i.e., $\Gamma_p$ consists of a single equivalence set $E_{\boldsymbol{\sigma}^1}$) or $\mu(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^1}) \geq \cdots \geq \mu(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^\ell})$, then $\overline{X}_{\boldsymbol{\theta}} = \mu_+(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^j})$ $(\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^j}, j = 1, \ldots, \ell)$.*

*(b) Suppose that $\mu(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^k}) < \mu(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^{k+1}})$ for some $k = 1, \ldots, \ell-1$. Then $\overline{X}_{\boldsymbol{\theta}} = \overline{X}_{\boldsymbol{\eta}}$ for every $\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^k}$ and $\boldsymbol{\eta} \in E_{\boldsymbol{\sigma}^{k+1}}$.*

*Proof.* By the assumption, we can transform the problem (3.5) further into

$$\begin{aligned}
&\min\left\{\sum_{q=1}^{r} \sum_{E_{\boldsymbol{\sigma}}\in\Gamma_q} \sum_{\boldsymbol{\theta}\in E_{\boldsymbol{\sigma}}} (\xi_{\boldsymbol{\sigma}} - Z_{\boldsymbol{\theta}})^2 : \begin{array}{l} X_{\boldsymbol{\theta}} = \xi_{\boldsymbol{\sigma}} \in \mathbb{R}_+ \text{ if } \boldsymbol{\theta} \in E_{\boldsymbol{\sigma}} \in \Gamma_q \ (q = 1,\ldots,r), \\ \xi_{\boldsymbol{\sigma}} \geq \xi_{\boldsymbol{\tau}} \text{ if } \Gamma_q \ni E_{\boldsymbol{\sigma}} \succ E_{\boldsymbol{\tau}} \in \Gamma_q \ (q = 1,\ldots,r) \end{array}\right\} \\
&= \sum_{q=1}^{r} \min\left\{\sum_{E_{\boldsymbol{\sigma}}\in\Gamma_q} \sum_{\boldsymbol{\theta}\in E_{\boldsymbol{\sigma}}} (\xi_{\boldsymbol{\sigma}} - Z_{\boldsymbol{\theta}})^2 : \begin{array}{l} X_{\boldsymbol{\theta}} = \xi_{\boldsymbol{\sigma}} \in \mathbb{R}_+ \text{ if } \boldsymbol{\theta} \in E_{\boldsymbol{\sigma}} \in \Gamma_q, \\ \xi_{\boldsymbol{\sigma}} \geq \xi_{\boldsymbol{\tau}} \text{ if } \Gamma_q \ni E_{\boldsymbol{\sigma}} \succ E_{\boldsymbol{\tau}} \in \Gamma_q \end{array}\right\}.
\end{aligned}$$

This implies that the computation of $\overline{\boldsymbol{X}} = \Pi_{\mathbb{L}}(\boldsymbol{Z})$ is divided into $r$ subproblems

$$\min\left\{\sum_{E_{\boldsymbol{\sigma}}\in\Gamma_q} \sum_{\boldsymbol{\theta}\in E_{\boldsymbol{\sigma}}} (\xi_{\boldsymbol{\sigma}} - Z_{\boldsymbol{\theta}})^2 : \begin{array}{l} X_{\boldsymbol{\theta}} = \xi_{\boldsymbol{\sigma}} \in \mathbb{R}_+ \text{ if } \boldsymbol{\theta} \in E_{\boldsymbol{\sigma}} \in \Gamma_q, \\ \xi_{\boldsymbol{\sigma}} \geq \xi_{\boldsymbol{\tau}} \text{ if } \Gamma_q \ni E_{\boldsymbol{\sigma}} \succ E_{\boldsymbol{\tau}} \in \Gamma_q \end{array}\right\} \qquad (3.6)$$

$(q = 1, \ldots, r)$. By assumption, the $p$th problem to compute $(X_{\boldsymbol{\theta}}$ $(\theta \in E_{\boldsymbol{\sigma}} \in \Gamma_p)$ is written as

$$\min\left\{\sum_{j=1}^{\ell} \sum_{\boldsymbol{\theta}\in E_{\boldsymbol{\sigma}^j}} (\xi_{\boldsymbol{\sigma}^j} - Z_{\boldsymbol{\theta}})^2 : \begin{array}{l} X_{\boldsymbol{\theta}} = \xi_{\boldsymbol{\sigma}^j} \in \mathbb{R}_+ \text{ if } \boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^j} \ (j = 1,\ldots,\ell), \\ \xi_{\boldsymbol{\sigma}^j} \geq \xi_{\boldsymbol{\sigma}^{j+1}} \quad (j = 1,\ldots,\ell-1) \end{array}\right\}. \quad (3.7)$$

We note that each term $\sum_{\boldsymbol{\theta}\in E_{\boldsymbol{\sigma}^j}} (\xi_{\boldsymbol{\sigma}^j} - Z_{\boldsymbol{\theta}})^2$ of the objective function of the $p$th subproblem (3.7) is a strictly convex function in a single variable $\xi_{\boldsymbol{\sigma}^j} \in \mathbb{R}$ with the unique minimizer $\mu(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^j})$. Hence, if $\bar{\xi}_{\boldsymbol{\sigma}^j} > \hat{\xi}_{\boldsymbol{\sigma}^j} \geq \mu(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^j})$ or $\mu(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^j}) \geq \hat{\xi}_{\boldsymbol{\sigma}^j} > \bar{\xi}_{\boldsymbol{\sigma}^j}$, then $\sum_{\boldsymbol{\theta}\in E_{\boldsymbol{\sigma}^j}} (\hat{\xi}_{\boldsymbol{\sigma}^j} - Z_{\boldsymbol{\theta}})^2 < \sum_{\boldsymbol{\theta}\in E_{\boldsymbol{\sigma}^j}} (\bar{\xi}_{\boldsymbol{\sigma}^j} - Z_{\boldsymbol{\theta}})^2$. This fact is used to prove both (a) and (b).

(a) For each $j = 1, \ldots, \ell$, $\bar{\xi}_{\boldsymbol{\sigma}^j} = \mu_+(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^j})$ is the minimizer of the problem

$$\min \left\{ \sum_{\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^j}} (\xi_{\boldsymbol{\sigma}^j} - Z_{\boldsymbol{\theta}})^2 : \xi_{\boldsymbol{\sigma}^j} \geq 0 \right\}.$$

Since they satisfy $\bar{\xi}_{\boldsymbol{\sigma}^1} \geq \cdots \geq \bar{\xi}_{\boldsymbol{\sigma}^\ell}$, $\overline{X}_{\boldsymbol{\theta}} = \bar{\xi}_{\boldsymbol{\sigma}^j}$ $(\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^j}, j = 1, \ldots, \ell)$ provide the minimizer of the subproblem.

(b) Since $\overline{\boldsymbol{X}}$ is feasible, there exist $\bar{\xi}_j$ such that $\overline{X}_{\boldsymbol{\theta}} = \bar{\xi}_j \in \mathbb{R}_+$ if $\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^j}$ $(j = 1, \ldots, \ell)$. Assume on the contrary that $\bar{\xi}_{\boldsymbol{\sigma}^k} > \bar{\xi}_{\boldsymbol{\sigma}^{k+1}}$. Consider $\hat{\xi}$ such that $\hat{\xi}_{\boldsymbol{\sigma}^j} = \bar{\xi}_{\boldsymbol{\sigma}^j}$ $(j \neq k,\ j \neq k+1)$. We consider two cases. First, suppose that $\mu_+(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^k}) < \bar{\xi}_{\boldsymbol{\sigma}^k}$. Set $\hat{\xi}_{\boldsymbol{\sigma}^k} = \hat{\xi}_{\boldsymbol{\sigma}^{k+1}} = \max\{\mu_+(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^k}), \bar{\xi}_{\boldsymbol{\sigma}^{k+1}}\}$. Then we observe that

$$\hat{\xi}_{\boldsymbol{\sigma}^1} \geq \cdots \geq \hat{\xi}_{\boldsymbol{\sigma}^{k-1}} \geq \bar{\xi}_{\boldsymbol{\sigma}^k} > \hat{\xi}_{\boldsymbol{\sigma}^k} = \hat{\xi}_{\boldsymbol{\sigma}^{k+1}} \geq \bar{\xi}_{\boldsymbol{\sigma}^{k+1}} \geq \hat{\xi}_{\boldsymbol{\sigma}^{k+2}} \geq \cdots \geq \hat{\xi}_{\boldsymbol{\sigma}^\ell}.$$

Hence

$$\sum_{\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^k}} (\hat{\xi}_{\boldsymbol{\sigma}^k} - Z_{\boldsymbol{\theta}})^2 < \sum_{\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^k}} (\bar{\xi}_{\boldsymbol{\sigma}^k} - Z_{\boldsymbol{\theta}})^2 \text{ (since } \mu_+(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^k}) \leq \hat{\xi}_{\boldsymbol{\sigma}^k} < \bar{\xi}_{\boldsymbol{\sigma}^k}),$$

$$\sum_{\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^{k+1}}} (\hat{\xi}_{\boldsymbol{\sigma}^{k+1}} - Z_{\boldsymbol{\theta}})^2 \begin{cases} < \displaystyle\sum_{\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^{k+1}}} (\bar{\xi}_{\boldsymbol{\sigma}^{k+1}} - Z_{\boldsymbol{\theta}})^2 & \text{if } \bar{\xi}_{\boldsymbol{\sigma}^{k+1}} < \mu_+(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^k}) \leq \mu_+(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^{k+1}}), \\ = \displaystyle\sum_{\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^{k+1}}} (\bar{\xi}_{\boldsymbol{\sigma}^{k+1}} - Z_{\boldsymbol{\theta}})^2 & \text{otherwise.} \end{cases}$$

It follows that $(\widehat{X}_{\boldsymbol{\theta}} = \hat{\xi}_{\boldsymbol{\sigma}^j}$ $(\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^j},\ j = 1, \ldots, \ell))$ is a feasible solution of the subproblem (3.7), and that the objective value $\sum_{i=1}^{\ell} \sum_{\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^j}} (\hat{\xi}_{\boldsymbol{\sigma}^j} - Z_{\boldsymbol{\theta}})^2$ is smaller than the optimal value $\sum_{j=1}^{\ell} \sum_{\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^j}} (\bar{\xi}_{\boldsymbol{\sigma}^j} - Z_{\boldsymbol{\theta}})^2$. This is a contradiction. Next suppose that $\bar{\xi}_{\boldsymbol{\sigma}^k} \leq \mu_+(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^k})$. In this case, set $\hat{\xi}_{\boldsymbol{\sigma}^k} = \hat{\xi}_{\boldsymbol{\sigma}^{k+1}} = \bar{\xi}_{\boldsymbol{\sigma}^k}$. Then

$$\hat{\xi}_{\boldsymbol{\sigma}^1} \geq \cdots \geq \hat{\xi}_{\boldsymbol{\sigma}^{k-1}} \geq \bar{\xi}_{\boldsymbol{\sigma}^k} = \hat{\xi}_{\boldsymbol{\sigma}^k} = \hat{\xi}_{\boldsymbol{\sigma}^{k+1}} > \bar{\xi}_{\boldsymbol{\sigma}^{k+1}} \geq \hat{\xi}_{\boldsymbol{\sigma}^{k+2}} \geq \cdots \geq \hat{\xi}_{\boldsymbol{\sigma}^\ell},$$

$$\sum_{\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^{k+1}}} (\hat{\xi}_{\boldsymbol{\sigma}^{k+1}} - Z_{\boldsymbol{\theta}})^2 < \sum_{\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^{k+1}}} (\bar{\xi}_{\boldsymbol{\sigma}^{k+1}} - Z_{\boldsymbol{\theta}})^2 \text{ (since } \bar{\xi}_{\boldsymbol{\sigma}^{k+1}} < \bar{\xi}_{\boldsymbol{\sigma}^k} = \hat{\xi}_{\boldsymbol{\sigma}^{k+1}} \leq \mu_+(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^k})).$$

As we can similarly derive a contradiction, assertion (b) follows. □

Based on Lemma 3.2, we present an algorithm for computing the metric projection $\overline{\boldsymbol{X}} = \Pi_{\mathbb{L}}(\boldsymbol{Z})$ of a given $\boldsymbol{Z} \in \mathbb{S}^{\mathcal{A}}$.

ALGORITHM 3.3. For $p = 1, \ldots, r$, execute the following steps.

**Step 0.** Represent $\Gamma_p$ as in (3.4). Compute $\mu_+(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^j})$ $(j = 1, \ldots, \ell)$.

**Step 1.** If $\ell = 1$ or $\mu(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^1}) \geq \cdots \geq \mu(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^\ell})$, output $\overline{X}_{\boldsymbol{\theta}} = \mu_+(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^j})$ $(\boldsymbol{\theta} \in E_{\boldsymbol{\sigma}^j}, j = 1, \ldots, \ell)$.

**Step 2.** Otherwise, find $k \in \{1, \ldots, \ell - 1\}$ such that $\mu(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^k}) < \mu(\boldsymbol{Z}, E_{\boldsymbol{\sigma}^{k+1}})$. Replace $\Gamma_p$ by $\{E_{\boldsymbol{\sigma}^1}, \ldots, E_{\boldsymbol{\sigma}^{k-1}}, E_{\boldsymbol{\sigma}^k} \cup E_{\boldsymbol{\sigma}^{k+1}}, E_{k+2}, \ldots, E_{\boldsymbol{\sigma}^\ell}\}$, where $E_{\boldsymbol{\sigma}^k} \cup E_{\boldsymbol{\sigma}^{k+1}}$ is regarded as a single equivalence class, and go to Step 0.

**4. Dense DNN relaxations of POP** (1.1)**.** Recall that POP (1.1) in $\mathbb{R}^n$ has been lifted in Section 2.5 to the problem (2.11) in $\mathbb{S}^{\mathcal{A}}$, where $\mathcal{A} \subset \mathbb{Z}^n_+$ satisfies (2.10). Then, the lifted problem (2.11) is relaxed to a COP of the form (2.1). The construction of the cone $\mathbb{L}_{\rm d}$ for $\mathbb{K}^2$ used in (2.1) is presented in Section 4.1 using the results in Section 3. For the computation of the metric projection $\Pi_{\mathbb{L}_{\mathrm{d}}}$, Lemma 3.2 and Algorithm 3.3 can be applied. We illustrate the computation with an example in Section 4.2. In Sections 4.3 and 4.4, we describe the standard choices of $\mathcal{A}$ for DNN relaxations of QOPs with binary and box constraints, and a hierarchy of DNN relaxations of POPs with binary and box constraints, respectively.

**4.1. Construction of a polyhedral cone $\mathbb{L}_{\mathbf{d}}$ for $\mathbb{K}^2$.** Assume that a nonempty subset $\mathcal{A}$ of $\mathbb{Z}^n_+$ satisfying (2.10) is given throughout this section. Let $\boldsymbol{\Theta} = \mathcal{A} \times \mathcal{A}$. Then, $\mathbb{S}^{\mathcal{A}}$ can be identified with $\mathbb{R}^{\boldsymbol{\Theta}}$ where the $(\boldsymbol{\alpha},\boldsymbol{\beta})$th element of $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{\Theta}}$ is written as $X_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ ($(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \boldsymbol{\Theta}$). We may regard $\boldsymbol{Q}^0 \in \mathbb{R}^{\boldsymbol{\Theta}}$ and $\boldsymbol{x}^{\square\mathcal{A}} \in \mathbb{R}^{\boldsymbol{\Theta}}$ in (2.11). To generalize the inequality $x_i \geq x_i^2$ for every $x_i \in [0,1]$ to monomials in $\boldsymbol{x}^{\square\mathcal{A}}$, we introduce a binary relation $\succeq$ in $\boldsymbol{\Theta}$ such that for every pair of $(\boldsymbol{\alpha},\boldsymbol{\beta}),(\boldsymbol{\gamma},\boldsymbol{\delta}) \in \boldsymbol{\Theta}$, $(\boldsymbol{\alpha},\boldsymbol{\beta}) \succeq (\boldsymbol{\gamma},\boldsymbol{\delta})$ if and only if there exists a positive number $c \geq 1$ such that $r_i(\boldsymbol{\alpha}+\boldsymbol{\beta}) = r_i(\boldsymbol{\gamma}+\boldsymbol{\delta})$ ($i \in I_{\mathrm{bin}}$) and $c(\alpha_j+\beta_j) = \gamma_j+\delta_j$ ($j \in I_{\mathrm{box}}$). Using the definition (2.9) of $\boldsymbol{r} : \mathbb{Z}^n_+ \to \mathbb{Z}^n_+$, it is easy to verify that $\succeq$ is a preorder. By definition, we see that

$$(\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\alpha}\boldsymbol{\beta}} = \boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}} = \boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\beta})} \geq \boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\gamma}+\boldsymbol{\delta})} = \boldsymbol{x}^{\boldsymbol{\gamma}+\boldsymbol{\delta}} = (\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\gamma}\boldsymbol{\delta}} \quad \text{if } \boldsymbol{x} \in S \text{ and } (\boldsymbol{\alpha},\boldsymbol{\beta}) \succeq (\boldsymbol{\gamma},\boldsymbol{\delta}), \tag{4.1}$$

where $(\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ denotes the $(\boldsymbol{\alpha},\boldsymbol{\beta})$th element of $\boldsymbol{x}^{\square\mathcal{A}} \in \mathbb{S}^{\mathcal{A}}$.

As shown in Section 3, the preorder $\succeq$ induces a strict preorder $\succ$ and an equivalence relation $\sim$ by (3.1). Let $(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \boldsymbol{\Theta}$. Then $[\boldsymbol{\theta}] = \{(\boldsymbol{\gamma},\boldsymbol{\delta}) \in \boldsymbol{\Theta} : \boldsymbol{r}(\boldsymbol{\gamma}+\boldsymbol{\delta}) = \boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\beta})\}$. Each equivalence class is characterized by $\boldsymbol{\sigma} \in \boldsymbol{r}(\mathcal{A}+\mathcal{A})$, and the family of equivalence classes is denoted by $\{E_{\boldsymbol{\sigma}}\ (\boldsymbol{\sigma} \in \boldsymbol{r}(\mathcal{A}+\mathcal{A})\}$ where $E_{\boldsymbol{\sigma}} = \{(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \boldsymbol{\Theta} : \boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\beta}) = \boldsymbol{\sigma}\}$. By definition, $(\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\alpha}\boldsymbol{\beta}} = (\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\gamma}\boldsymbol{\delta}}$ if and only if $(\boldsymbol{\alpha},\boldsymbol{\beta}) \succeq (\boldsymbol{\gamma},\boldsymbol{\delta})$ and $(\boldsymbol{\gamma},\boldsymbol{\delta}) \succeq (\boldsymbol{\alpha},\boldsymbol{\beta})$. Hence, by (4.1),

$$(\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\alpha}\boldsymbol{\beta}} = (\boldsymbol{x}^{\square\mathcal{A}})_{\boldsymbol{\gamma}\boldsymbol{\delta}} \text{ if } \boldsymbol{x} \in S \text{ and } (\boldsymbol{\alpha},\boldsymbol{\beta}) \sim (\boldsymbol{\gamma},\boldsymbol{\delta}). \tag{4.2}$$

LEMMA 4.1.

*(a) For each equivalence class $E_{\boldsymbol{\sigma}}$ ($\boldsymbol{\sigma} \in \boldsymbol{r}(\mathcal{A}+\mathcal{A})$), there exists a unique maximal chain $\Gamma$containing $E_{\boldsymbol{\sigma}}$, which is represented as*

$$\Gamma = \{E_{\boldsymbol{\tau}} : \tau_i = \sigma_i \ (i \in I_{bin}),\ c\tau_j = \sigma_j \ (j \in I_{box}) \text{ for some } c > 0\ \}.$$

*(b) The family of maximal chains partitions the family of equivalence classes. If we denote the family of maximal chains as $\Gamma_1,\ldots,\Gamma_r$ then $\bigcup_{p=1}^r \Gamma_p = \{E_{\boldsymbol{\sigma}} : \boldsymbol{\sigma} \in \boldsymbol{r}(\mathcal{A}+\mathcal{A})\}$ and $\Gamma_p \bigcap \Gamma_q = \emptyset$ ($p \neq q$).*

*(c) If $I_{box} = \emptyset$, then every maximal chain consists of a single equivalence class $E_{\boldsymbol{\sigma}}$ for some $\boldsymbol{\sigma} \in \boldsymbol{r}(\mathcal{A}+\mathcal{A})$.*

*Proof.* It suffices to show assertion (a) because assertions (b) and (c) follow from (a). Since each equivalence class itself is a chain, there is at least one maximal chain containing it. We show the uniqueness. Let $\boldsymbol{\sigma} \in \boldsymbol{r}(\mathcal{A}+\mathcal{A})$ be fixed. By definition,

$$\begin{aligned}
&E_{\boldsymbol{\sigma}} = \{(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \boldsymbol{\Theta} : \boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\beta}) = \boldsymbol{\sigma}\},\\
&E_{\boldsymbol{\tau}} \succ E_{\boldsymbol{\sigma}} \Leftrightarrow \tau_i = \sigma_i \ (i \in I_{\mathrm{bin}}) \text{ and } c\tau_j = \sigma_j \ (j \in I_{\mathrm{box}}) \text{ for some } c > 1,\\
&E_{\sigma} \succ E_{\boldsymbol{\tau}} \Leftrightarrow \tau_i = \sigma_i \ (i \in I_{\mathrm{bin}}) \text{ and } c\sigma_j = \tau_j \ (j \in I_{\mathrm{box}}) \text{ for some } c > 1.
\end{aligned}$$

Therefore, the maximal chain $\Gamma$ containing $E_{\boldsymbol{\sigma}}$ is uniquely determined as described in (a). ▯

Let $\Gamma_1, \ldots, \Gamma_r$ be the family of maximal chains of equivalence classes. Define

$$\begin{aligned}
\boldsymbol{H}^0 &= \text{the } |\mathcal{A}| \times |\mathcal{A}| \text{ matrix in } \mathbb{S}^{\mathcal{A}} \text{ with 1 at the } (\mathbf{0},\mathbf{0})\text{th element and 0 elsewhere,} \\
\mathbb{L}_{\rm d} &= \left\{ \boldsymbol{X} \in \mathbb{N}^{\mathcal{A}}_+ : \begin{array}{ll} X_{\boldsymbol{\alpha\beta}} = X_{\boldsymbol{\gamma\delta}} & \text{if } (\boldsymbol{\alpha},\boldsymbol{\beta}) \sim (\boldsymbol{\gamma},\boldsymbol{\delta}), \\ X_{\boldsymbol{\alpha\beta}} \geq X_{\boldsymbol{\alpha\beta}} & \text{if } \Gamma_p \ni [(\boldsymbol{\alpha},\boldsymbol{\beta})] \succ [(\boldsymbol{\gamma},\boldsymbol{\delta})] \in \Gamma_p \\ & (p = 1, \ldots, r) \end{array} \right\} \\
&= \left\{ \boldsymbol{X} \in \mathbb{S}^{\mathcal{A}} : \begin{array}{ll} X_{\boldsymbol{\alpha\beta}} = \xi_{\boldsymbol{\sigma}} \in \mathbb{R}_+ & \text{if } (\boldsymbol{\alpha},\boldsymbol{\beta}) \in E_{\boldsymbol{\sigma}} \in \Gamma_p \ (p = 1, \ldots, r), \\ \xi_{\boldsymbol{\sigma}} \geq \xi_{\boldsymbol{\tau}} & \text{if } \Gamma_p \ni E_{\boldsymbol{\sigma}} \succ E_{\boldsymbol{\tau}} \in \Gamma_p \ (p = 1, \ldots, r) \end{array} \right\}.
\end{aligned}$$

Here the last identity follows from (b) of Lemma 4.1. If $\boldsymbol{x}$ is a feasible solution of (2.11) with the objective value $\langle \boldsymbol{Q}^0, \boldsymbol{x}^{\square \mathcal{A}} \rangle = f(\boldsymbol{x})$, then $\boldsymbol{X} = \boldsymbol{x}^{\square \mathcal{A}}$ satisfies that

$$\begin{aligned}
&\langle \boldsymbol{Q}^0, \boldsymbol{x}^{\square \mathcal{A}} \rangle = \langle \boldsymbol{Q}^0, \boldsymbol{X} \rangle, \ \boldsymbol{X} \in \mathbb{S}^{\mathcal{A}}_+, \ \boldsymbol{X} \in \mathbb{N}^{\mathcal{A}}_+, \ \langle \boldsymbol{H}^0, \boldsymbol{X} \rangle = X_{\mathbf{00}} = 1, \\
&X_{\boldsymbol{\alpha\beta}} = X_{\boldsymbol{\gamma\delta}} \text{ if } (\boldsymbol{\alpha},\boldsymbol{\beta}) \sim (\boldsymbol{\gamma},\boldsymbol{\delta}), \\
&X_{\boldsymbol{\alpha\beta}} \geq X_{\boldsymbol{\gamma\delta}} \text{ if } \Gamma_p \ni [(\boldsymbol{\alpha},\boldsymbol{\beta})] \succ [(\boldsymbol{\gamma},\boldsymbol{\delta})] \in \Gamma_p \ (p = 1, \ldots, r).
\end{aligned}$$

Note that the last two relations are obtained by (4.1) and (4.2). This implies that $\boldsymbol{X}$ is a feasible solution of COP (2.1) with $\mathbb{K}^1 = \mathbb{S}^{\mathcal{A}}_+$ and $\mathbb{K}^2 = \mathbb{L}_{\rm d}$ and attains the same objective value $\langle \boldsymbol{Q}^0, \boldsymbol{x}^{\square \mathcal{A}} \rangle$. Therefore, COP (2.1) serves as a DNN relaxation of POP (1.1). Notice that the construction of $\mathbb{L}_{\rm d}$ depends on the choices of $\mathcal{A}$. The standard choices of $\mathcal{A}$ are discussed for QOPs in Section 4.3 and for POPs in Section 4.4.

LEMMA 4.2. *Let* $\boldsymbol{X}$ *be a feasible solution of COP* (2.1) *with* $\mathbb{K}^1 = \mathbb{S}^{\mathcal{A}}_+$ *and* $\mathbb{K}^2 = \mathbb{L}_d$.

*(a)* $0 \leq X_{\boldsymbol{\alpha\beta}} \leq 1$ *for every* $((\boldsymbol{\alpha},\boldsymbol{\beta}) \in \boldsymbol{\Theta})$.

*(b)* $\langle \boldsymbol{I}, \boldsymbol{X} \rangle \leq |\mathcal{A}|$, *where* $\boldsymbol{I}$ *denotes the identity matrix in* $\mathbb{S}^{\mathcal{A}}$.

*Proof.* As assertion (b) follows from assertion (a), we only prove (a). It follows from the definition of $\mathbb{L}_{\rm d}$ that each element of $\boldsymbol{X}$ is nonnegative. Since $\boldsymbol{X} \in \mathbb{S}^{\mathcal{A}}_+$, it suffices to prove that all the diagonal elements of $\boldsymbol{X}$ is not greater than 1. We know that $1 = \langle \boldsymbol{H}^0, \boldsymbol{X} \rangle = X_{\mathbf{00}}$. Let $\mathbf{0} \neq \boldsymbol{\alpha} \in \mathcal{A}$ be fixed. We will show $X_{\boldsymbol{\alpha\alpha}} \leq 1$. Consider the $2 \times 2$ principal submatrix $\begin{pmatrix} X_{\mathbf{00}} & X_{\mathbf{0}\boldsymbol{\alpha}} \\ X_{\boldsymbol{\alpha}\mathbf{0}} & X_{\boldsymbol{\alpha\alpha}} \end{pmatrix} = \begin{pmatrix} 1 & X_{\mathbf{0}\boldsymbol{\alpha}} \\ X_{\boldsymbol{\alpha}\mathbf{0}} & X_{\boldsymbol{\alpha\alpha}} \end{pmatrix}$ of $\boldsymbol{X} \in \mathbb{S}^{\mathcal{A}}_+$. The positive semdifiniteness of the submatrix implies that $0 \leq X_{\boldsymbol{\alpha\alpha}} - X^2_{\mathbf{0}\boldsymbol{\alpha}}$. On the other hand, we see that $r_i(\mathbf{0}+\boldsymbol{\alpha}) = r_i(\boldsymbol{\alpha}+\boldsymbol{\alpha})$ $(i \in I_{\rm bin})$ and $2r_j(\mathbf{0}+\boldsymbol{\alpha}) = 2\alpha_j = r_j(\boldsymbol{\alpha}+\boldsymbol{\alpha})$ $(j \in I_{\rm box})$. As a result, $(\mathbf{0},\boldsymbol{\alpha}) \succeq (\boldsymbol{\alpha},\boldsymbol{\alpha})$, and $X_{\mathbf{0}\boldsymbol{\alpha}} \geq X_{\boldsymbol{\alpha\alpha}}$. It follows that $0 \leq X_{\boldsymbol{\alpha\alpha}} - X^2_{\boldsymbol{\alpha\alpha}} = X_{\boldsymbol{\alpha\alpha}}(1 - X_{\boldsymbol{\alpha\alpha}})$, which implies $0 \leq X_{\boldsymbol{\alpha\alpha}} \leq 1$. □

By (b) of Lemma 4.2, we can apply Algorithm 2.1 (Accelerated Bisection Algorithm) to COP (2.1) with $\mathbb{K}^1 = \mathbb{S}^{\mathcal{A}}_+$ and $\mathbb{K}^2 = \mathbb{L}_{\rm d}$ generated previously.

**4.2. Computation of the metric projection $\Pi_{\mathbb{L}_{\rm d}}$ from $\mathbb{S}^{\mathcal{A}}$ onto $\mathbb{L}_{\rm d}$.** We have already shown in Lemma 4.1 that the family of maximal chains $\{\Gamma_1, \ldots, \Gamma_r\}$ partitions the family of equivalence classes $\{E_{\boldsymbol{\sigma}} : \boldsymbol{\sigma} \in \boldsymbol{r}(\mathcal{A}+\mathcal{A})\}$. Therefore, we can use Lemma 3.2 and Algorithm 3.3 for the computation of the metric projection $\Pi_{\mathbb{L}_{\rm d}}$ from $\mathbb{S}^{\mathcal{A}}$ onto $\mathbb{L}_{\rm d}$.

EXAMPLE 4.3. Let $I_{\rm bin} = \{1\}$, $I_{\rm box} = \{2\}$, $S = \{(x_1, x_2) \in \mathbb{R}^2 : x_i \in \{0,1\} \ (i \in I_{\rm bin}), \ x_j \in [0,1] \ (j \in I_{\rm box})\}$ and $\mathcal{A} = \{(0,0),(1,0),(0,1),(1,1)\}$. In this case, we see

$\boldsymbol{r}(\mathcal{A}) = \mathcal{A}$ and

$$\boldsymbol{x}^{\square\mathcal{A}} = \begin{pmatrix} 1 & x_1 & x_2 & x_1x_2 \\ x_1 & x_1^2 & x_1x_2 & x_1^2x_2 \\ x_2 & x_1x_2 & x_2^2 & x_1x_2^2 \\ x_1x_2 & x_1^2x_2 & x_1x_2^2 & x_1^2x_2^2 \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_2 & x_1x_2 \\ x_1 & x_1 & x_1x_2 & x_1x_2 \\ x_2 & x_1x_2 & x_2^2 & x_1x_2^2 \\ x_1x_2 & x_1x_2 & x_1x_2^2 & x_1x_2^2 \end{pmatrix}$$

for every $\boldsymbol{x} \in S$. We also see that $\boldsymbol{r}(\mathcal{A}+\mathcal{A}) = \{(0,0),(1,0),(0,1),(1,1),(0,2),(1,2)\}$. As a result, we have 6 equivalence classes corresponding to the different monomials $1,\ x_1,\ x_2,\ x_1x_2,\ x_2^2$ and $x_1x_2^2$ in the reduced moment matrix on the right, and the equivalence classes are $E_{(0,0)},\ E_{(1,0)},\ E_{(0,1)},\ E_{(1,1)},\ E_{(0,2)},\ E_{(1,2)}$. For example, $E_{(1,2)}$ consists of $(\boldsymbol{\alpha},\boldsymbol{\beta}) = ((0,1),(1,1)),\ ((1,1),(0,1)),\ ((1,1),(1,1))$ corresponding to 3 monomials $x_1x_2^2$ appeared in the reduced moment matrix. The maximal chains are $\Gamma_1 = \{E_{(0,0)}\},\ \Gamma_2 = \{E_{(1,0)}\},\ \Gamma_3 = \{E_{(0,1)},\ E_{(0,2)}\},\ \Gamma_4 = \{E_{(1,1)},\ E_{(1,2)}\}$. Thus, $\mathbb{L}_{\rm d}$ is represented as

$$\mathbb{L}_{\rm d} = \left\{ \boldsymbol{X} \in \mathbb{S}^{\mathcal{A}} : \begin{array}{l} X_{\boldsymbol{\alpha\beta}} = \xi_{\boldsymbol{\sigma}} \in \mathbb{R}_+ \text{ if } (\boldsymbol{\alpha},\boldsymbol{\beta}) \in E_{\boldsymbol{\sigma}} \in \Gamma_p\ (p = 1,\ldots,4), \\ \xi_{(0,1)} \geq \xi_{(0,2)},\ \xi_{(1,1)} \geq \xi_{(1,2)} \end{array} \right\}.$$

To illustrate how we compute $\overline{\boldsymbol{X}} = \Pi_{\mathbb{L}_{\rm d}}(\boldsymbol{Z})$ for a given $\boldsymbol{Z} \in \mathbb{S}^{\mathcal{A}}$, take $\Gamma_p = \Gamma_4 = \{E_{(1,1)},\ E_{(1,2)}\}$. By Lemma 3.2 (or applying Steps 0, 1 and 2 of Algorithm 3.3 with $p = 4$), we obtain that

$$\begin{array}{l} \overline{X}_{\boldsymbol{\alpha\beta}} = \begin{cases} \mu_+(\boldsymbol{Z}, E_{(1,1)}) & \text{if } \mu(\boldsymbol{Z}, E_{(1,1)}) \geq \mu(\boldsymbol{Z}, E_{(1,2)}) \\ \mu_+(\boldsymbol{Z}, E_{(1,1)} \cup E_{(1,2)}) & \text{otherwise,} \end{cases} \\ \qquad \text{for all } (\boldsymbol{\alpha},\boldsymbol{\beta}) \in E_{(1,1)}, \\ \overline{X}_{\boldsymbol{\alpha\beta}} = \begin{cases} \mu_+(\boldsymbol{Z}, E_{(1,2)}) & \text{if } \mu(\boldsymbol{Z}, E_{(1,1)}) \geq \mu(\boldsymbol{Z}, E_{(1,2)}) \\ \mu_+(\boldsymbol{Z}, E_{(1,1)} \cup E_{(1,2)}) & \text{otherwise,} \end{cases} \\ \qquad \text{for all } (\boldsymbol{\alpha},\boldsymbol{\beta}) \in E_{(1,2)}. \end{array}$$

**4.3. Choosing $\mathcal{A}$ for the standard DNN relaxation of a QOP (2.5) with binary and box constraints.** For the standard DNN relaxation QOP (2.5) with binary and box constraints in Section 2.3, we choose $\mathcal{A} = \left\{\boldsymbol{\alpha} \in \mathbb{Z}^n_+ : \sum_{i=1}^n \alpha_i \leq 1\right\} = \{\boldsymbol{0}, \boldsymbol{e}^1, \ldots, \boldsymbol{e}^n\}$, where $\boldsymbol{e}^i$ denotes the $i$th coordinate vector with 1 at the $i$th element and 0 elsewhere ($i = 1,\ldots,n$). Then the moment matrix $\boldsymbol{x}^{\square\mathcal{A}}$ coincides with $(1,\boldsymbol{x})(1,\boldsymbol{x})^T$. The equivalence classes are $E_{\boldsymbol{e}^j} = \{(\boldsymbol{0},\boldsymbol{e}^j),(\boldsymbol{e}^j,\boldsymbol{0}),(\boldsymbol{e}^j,\boldsymbol{e}^j)\}$ ($j \in I_{\rm bin}$) in addition to the trivial ones from the symmetry of the moment matrix $\boldsymbol{x}^{\square\mathcal{A}}$. They induce the identities $x_j = x_j^2$ ($j \in I_{\rm bin}$). The maximal chains which have more than one equivalence classes are $\{E_{\boldsymbol{e}^i},\ E_{2\boldsymbol{e}^i}\}$ ($i \in I_{\rm box}$), which induce the inequalities $x_i \geq x_i^2$ ($i \in I_{\rm box}$). Therefore, the cone $\mathbb{L}_{\rm d}$ used for $\mathbb{K}^2$ of (2.1) is represented as $\mathbb{L}_{\rm d} = \{\boldsymbol{X} \in \mathbb{N}^{\mathcal{A}} : X_{\boldsymbol{0}\boldsymbol{e}^j} = X_{\boldsymbol{e}^j\boldsymbol{e}^j}\ (j \in I_{\rm bin}),\ X_{\boldsymbol{0}\boldsymbol{e}^i} \geq X_{\boldsymbol{e}^i\boldsymbol{e}^i}\ (i \in I_{\rm box})\}$, which coincides with the cone $\mathbb{L}$ defined by (2.8) if we use the coordinate $0,1,\ldots,n$ instead of $\mathcal{A}$ and identify $\mathbb{S}^{\mathcal{A}}$ with $\mathbb{S}^{1+n}$.

**4.4. Choosing $\mathcal{A}$ for a hierarchy of DNN relaxations POPs with binary and box constraints.** In this subsection, we briefly discuss how the idea of the hierarchy of SDP relaxations proposed by Lasserre [17] for general POPs is incorporated in our DNN relaxation of POP (1.1) with binary and box constraints. Let $\omega_0$ be the positive smallest integer not less than $(\deg f)/2$. For each $\omega \geq \omega_0$, let $\mathcal{A}^\omega = \{\boldsymbol{r}(\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \mathbb{Z}^n_+,\ \sum_{i=1}^n \alpha_i \leq \omega\}$. We call $\omega$ a relaxation order.

First, we apply $\boldsymbol{r}$ to the exponents of all monomials of $f(\boldsymbol{x})$, *i.e.*, replace every monomial $\boldsymbol{x}^{\boldsymbol{\gamma}}$ of $f_0(\boldsymbol{x})$ by $\boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\gamma})}$. Then condition (2.10) is clearly satisfied with $\mathcal{A} = \mathcal{A}^\omega$ for every $\omega \geq \omega_0$. By increasing the relaxation order $\omega$ from $\omega_0$ and constructing a DNN relaxation with each $\mathcal{A} = \mathcal{A}^\omega$, a hierarchy of DNN relaxations of POP (1.1) can be obtained. If $I_{\rm bin} = \{1, \ldots, n\}$, *i.e.*, $S = \{0,1\}^n$, this construction is essentially the same as Lasserre's when it is applied to POP (1.1) with $S = \{0,1\}^n$, except that a stronger DNN relaxation than a SDP relaxation is employed at each level of the hierarchy; See [16]. When $I_{\rm bin} \neq \{1, \ldots, n\}$, our DNN relaxation and Lasserre's SDP relaxation at each level of the hierarchies are different. Although the theoretical convergence of the optimal values of our hierarchy of DNN relaxations to the optimal value of POP (1.1) is not provided, it is possible to show the effectiveness of our hierarchy of DNN relaxations with numerical experiments, which we will leave as future work.

**5. Sparse DNN relaxations of POP** (1.1)**.** The sparse DNN relaxation of POP (1.1) presented in this section is based on [23], where a sparse version of the hierarchy of SDP relaxations of general POPs with inequality constraints was proposed by exploiting certain structured sparsity in the objective polynomial and constraints. Notice that POP (1.1) involves only simple constraints $x_i - x_i^2 = 0$ for $x_i \in \{0,1\}$ and $x_i - x_i^2 \geq 0$ for $x_i \in [0,1]$, which are both separable in $x_i$ ($i = 1, \ldots, n$). As a result, we can focus on the objective polynomial $f(\boldsymbol{x})$ for sparsity exploitation. The structured sparsity of $f(\boldsymbol{x})$ is readily observed in the nonzero pattern of its Hessian matrix.

For each subset $C$ of $\{1, \ldots, n\}$, let $\mathbb{Z}^C_+ = \{\boldsymbol{\alpha} \in \mathbb{Z}^n_+ : \alpha_i = 0 \text{ if } i \notin C\}$. We first assume that the objective polynomial $f(\boldsymbol{x})$ with $\text{supp} f = \boldsymbol{r}(\text{supp} f)$ is represented as the sum of $m$ polynomials $f_k(\boldsymbol{x})$ $(k = 1, \ldots, m)$ with $\text{supp} f_k = \boldsymbol{r}(\text{supp} f_k) \subset \mathbb{Z}^{C_k}_+$ for some $C_k \subset \{1, \ldots, n\}$. We assume that $m$ and the size $|C_k|$ of each $C_k$ are both small, *e.g.*, $m \leq n$ and $|C_k| = O(1)$. Under this assumption, the construction of the cone $\mathbb{L}_{\rm s}$ used for $\mathbb{K}^2$ of the sparse DNN relaxation of the form (2.1) is described in Section 5.1. More details on choices of $C_1, \ldots, C_m$ from the Hessian matrix of the objective function $f(\boldsymbol{x})$ are presented in Section 5.2.

**5.1. Construction of a polyhedral cone $\mathbb{L}_{\rm s}$ for $\mathbb{K}^2$.** For each $k$, choose $\mathcal{A}_k$ to be a finite subset of $\mathbb{Z}^{C_k}_+$ such that $\boldsymbol{0} \in \mathcal{A}_k = \boldsymbol{r}(\mathcal{A}_k)$ and $\text{supp} f_k = \boldsymbol{r}(\text{supp} f_k) \subset \mathcal{A}_k + \mathcal{A}_k$, and take a matrix $\boldsymbol{Q}^0_k \in \mathbb{S}^{\mathcal{A}_k}$ such that $f_k(\boldsymbol{x}) = \langle \boldsymbol{Q}^0_k, \, \boldsymbol{x}^{\square \mathcal{A}_k} \rangle$. Let $\boldsymbol{\Theta} = \{(k, \boldsymbol{\alpha}, \boldsymbol{\beta}) : (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{A}_k \times \mathcal{A}_k, \ k = 1, \ldots, m\}$.

For each $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m) \in \mathbb{S}^{\mathcal{A}_1} \times \cdots \times \mathbb{S}^{\mathcal{A}_m}$, we denote the $(\boldsymbol{\alpha}, \boldsymbol{\beta})$th element of $\boldsymbol{X}_k$ by $X_{k\boldsymbol{\alpha}\boldsymbol{\beta}}$ $((k, \boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\Theta})$ $(k = 1, \ldots, m)$. Then we may identify $\mathbb{S}^{\mathcal{A}_1} \times \cdots \times \mathbb{S}^{\mathcal{A}_m}$ with the $|\boldsymbol{\Theta}|$-dimensional Euclidian space $\mathbb{R}^{\boldsymbol{\Theta}}$ with the inner product $\langle \boldsymbol{W}, \, \boldsymbol{X} \rangle = \sum_{(k,\boldsymbol{\alpha},\boldsymbol{\beta}) \in \boldsymbol{\Theta}} W_{k\boldsymbol{\alpha}\boldsymbol{\beta}} X_{k\boldsymbol{\alpha}\boldsymbol{\beta}}$. Let $\boldsymbol{Q}^0 = (\boldsymbol{Q}^0_1, \ldots, \boldsymbol{Q}^0_m) \in \mathbb{R}^{\boldsymbol{\Theta}}$. Then, POP (1.1) is equivalent to

$$\min \left\{ \langle \boldsymbol{Q}^0, \, \boldsymbol{X} \rangle : \boldsymbol{X} \in T \right\}, \tag{5.1}$$

where $T = \left\{ \boldsymbol{X} = (\boldsymbol{x}^{\square \mathcal{A}_1}, \ldots, \boldsymbol{x}^{\square \mathcal{A}_m}) \in \mathbb{R}^{\boldsymbol{\Theta}} : \boldsymbol{x} \in S \right\}$. Notice that neither the representation of $f(\boldsymbol{x})$ in terms of $f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})$ nor the representation of each $f_k(\boldsymbol{x})$ in terms of $\boldsymbol{Q}^0_k \in \mathbb{S}^{\mathcal{A}}$ is unique.

The sparse DNN relaxations of POP (1.1) is derived from (5.1) in the same way as the dense DNN relaxations of (1.1) from (2.11) in Section 4. Using $\boldsymbol{r} : \mathbb{Z}^n_+ \to \mathbb{Z}^n_+$ defined by (2.9), we introduce a preorder $\succeq$ in $\boldsymbol{\Theta}$: for every pair of $(k, \boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\Theta}$ and

$(\ell, \boldsymbol{\gamma}, \boldsymbol{\delta}) \in \boldsymbol{\Theta}$, $(k, \boldsymbol{\alpha}, \boldsymbol{\beta}) \succeq (\ell, \boldsymbol{\gamma}, \boldsymbol{\delta})$ if and only if there exists a positive number $c \geq 1$ such that $r_i(\boldsymbol{\alpha}+\boldsymbol{\beta}) = r_i(\boldsymbol{\gamma}+\boldsymbol{\delta})$ $(i \in I_{\rm bin})$ and $c(\alpha_j + \beta_j) = \gamma_j + \delta_j$ $(j \in I_{\rm box})$. By definition, we see that

$$(\boldsymbol{x}^{\Box \mathcal{A}_k})_{\boldsymbol{\alpha}\boldsymbol{\beta}} = \boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}} = \boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\beta})} \geq \boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\gamma}+\boldsymbol{\delta})} = \boldsymbol{x}^{\boldsymbol{\gamma}+\boldsymbol{\delta}} = (\boldsymbol{x}^{\Box \mathcal{A}_\ell})_{\boldsymbol{\gamma}\boldsymbol{\delta}}$$
$$\text{if } \boldsymbol{x} \in S \text{ and } \boldsymbol{\Theta} \ni (k, \boldsymbol{\alpha}, \boldsymbol{\beta}) \succeq (\ell, \boldsymbol{\gamma}, \boldsymbol{\delta}) \in \boldsymbol{\Theta}. \tag{5.2}$$

The preorder $\succeq$ induces a strictly preorder $\succ$ and an equivalence relation $\sim$ by (3.1). Let $(\ell, \boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\Theta}$. Then $[(\ell, \boldsymbol{\alpha}, \boldsymbol{\beta})] = \{(k, \boldsymbol{\gamma}, \boldsymbol{\delta}) \in \boldsymbol{\Theta} : \boldsymbol{r}(\boldsymbol{\gamma}+\boldsymbol{\delta}) = \boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\beta})\}$ (the equivalence class containing $(\ell, \boldsymbol{\alpha}, \boldsymbol{\beta})$). As a result, each equivalence class is characterized by $\boldsymbol{\sigma} \in \bigcup_{k=1}^m \boldsymbol{r}(\mathcal{A}_k + \mathcal{A}_k)$, and the family of equivalence classes is denoted by $\{E_{\boldsymbol{\sigma}}\ (\boldsymbol{\sigma} \in \bigcup_{k=1}^m\ \boldsymbol{r}(\mathcal{A}_k + \mathcal{A}_k))\}$ where $E_{\boldsymbol{\sigma}} = \{(k, \boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\Theta} : \boldsymbol{r}(\boldsymbol{\alpha}+\boldsymbol{\beta}) = \boldsymbol{\sigma}\}$. By definition and (5.2),

$$(\boldsymbol{x}^{\Box \mathcal{A}_k})_{\boldsymbol{\alpha}\boldsymbol{\beta}} = (\boldsymbol{x}^{\Box \mathcal{A}_\ell})_{\boldsymbol{\gamma}\boldsymbol{\delta}} \text{ if } \boldsymbol{x} \in S \text{ and } (k, \boldsymbol{\alpha}, \boldsymbol{\beta}) \sim (\ell, \boldsymbol{\gamma}, \boldsymbol{\delta}). \tag{5.3}$$

We can extend Lemma 4.1 established for the dense DNN relaxation to the sparse DNN relaxation. In fact, the extended lemma ensures that the family of maximal chains of equivalence classes, denoted by $\{\Gamma_1, \ldots, \Gamma_r\}$, partitions the family of equivalence classes $\{E_{\boldsymbol{\sigma}}\ (\boldsymbol{\sigma} \in \bigcup_{k=1}^m\ \boldsymbol{r}(\mathcal{A}_k + \mathcal{A}_k))\}$. Now define

$$\begin{aligned}
\boldsymbol{H}_1^0 &= \text{the } |\mathcal{A}_1| \times |\mathcal{A}_1| \text{ matrix in } \mathbb{S}^{\mathcal{A}_1} \text{ with 1 at the } (\boldsymbol{0}, \boldsymbol{0})\text{th element} \\
&\quad\ \text{and 0 elsewhere}, \\
\boldsymbol{H}^0 &= (\boldsymbol{H}_1^0, \boldsymbol{O}, \ldots, \boldsymbol{O}) \in \mathbb{R}^{\boldsymbol{\Theta}} = \mathbb{S}^{\mathcal{A}_1} \times \cdots \times \mathbb{S}^{\mathcal{A}_m}, \\
\mathbb{L}_{\rm s} &= \left\{ \boldsymbol{X} \in \mathbb{R}^{\boldsymbol{\Theta}} : \begin{array}{ll} X_{k\boldsymbol{\alpha}\boldsymbol{\beta}} = \xi_{\boldsymbol{\sigma}} \in \mathbb{R}_+ & \text{if } (k, \boldsymbol{\alpha}, \boldsymbol{\beta}) \in E_{\boldsymbol{\sigma}} \in \Gamma_p \\ \xi_{\boldsymbol{\sigma}} \geq \xi_{\boldsymbol{\tau}} & \text{if } \Gamma_p \ni E_{\boldsymbol{\sigma}} \succ E_{\boldsymbol{\tau}} \in \Gamma_p \\ & (p = 1, \ldots, r) \end{array} \right\}.
\end{aligned}$$

If $\boldsymbol{x}$ is a feasible solution, then $\boldsymbol{X} = (\boldsymbol{x}^{\Box \mathcal{A}_1}, \ldots, \boldsymbol{x}^{\Box \mathcal{A}_m}) \in \mathbb{R}^{\boldsymbol{\Theta}}$ satisfies $\boldsymbol{X} \in \mathbb{S}_+^{\mathcal{A}_1} \times \cdots \times \mathbb{S}_+^{\mathcal{A}_m}$, $\langle \boldsymbol{H}^0, \boldsymbol{X} \rangle = 1$, $\boldsymbol{X} \in \mathbb{L}_{\rm s}$ and $f(\boldsymbol{x}) = \langle \boldsymbol{Q}^0, \boldsymbol{X} \rangle$. Thus, we can see that it is a feasible solution of COP (2.1) with $\mathbb{K}^1 = \mathbb{S}_+^{\mathcal{A}_1} \times \cdots \times \mathbb{S}_+^{\mathcal{A}_m}$ and $\mathbb{K}^2 = \mathbb{L}_{\rm s}$, and that it also attains the same objective value $f(\boldsymbol{x})$. This implies that COP (2.1) serves as a sparse DNN relaxation of POP (1.1).

Lemma 4.2 can be extended to the sparse DNN relaxation of POP (1.1), so that Algorithm 2.1 can be used for solving COP (2.1) with $\mathbb{K}^1 = \mathbb{S}_+^{\mathcal{A}_1} \times \cdots \times \mathbb{S}_+^{\mathcal{A}_m}$ and $\mathbb{K}^2 = \mathbb{L}_{\rm s}$. Lemma 3.2 and Algorithm 3.3 can also be applied to the metric projection $\Pi_{\mathbb{L}_{\rm s}}$ from $\mathbb{R}^{\boldsymbol{\Theta}}$ onto $\mathbb{L}_{\rm s}$.

**5.2. The chordal graph sparsity of the Hessian matrix of the objective polynomial of POP** (1.1)**.** In this section, we describe how the objective polynomial $f(\boldsymbol{x})$ of POP (1.1) can be represented as the sum of $m$ polynomials $f_k(\boldsymbol{x})$ $(k = 1, \ldots, m)$ with ${\rm supp} f_k \subset \mathbb{Z}^{C_k}$ for some $C_k \subset \{1, \ldots, n\}$. We assume that $f(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in {\rm supp} f} c(\boldsymbol{\alpha}) \boldsymbol{x}^{\boldsymbol{\alpha}}$ with ${\rm supp} f = \boldsymbol{r}({\rm supp} f) \subset \mathbb{Z}_+^n$ .

Let $\boldsymbol{H}f(\boldsymbol{x})$ denotes the $n \times n$ Hessian matrix of $f(\boldsymbol{x})$. Each element of $\boldsymbol{H}f(\boldsymbol{x})$ is a polynomial. Throughout this section, we assume that many elements of $\boldsymbol{H}f(\boldsymbol{x})$ are identically zero. To find a structured sparsity of $\boldsymbol{H}f(\boldsymbol{x})$ which can be exploited, we introduce an undirected graph $G(N, \mathcal{E})$ with a node set $N = \{1, \ldots, n\}$ and an edge set $\mathcal{E} \subset N \times N$ consisting of all $(i,j)$s with $i \neq j$ such that the $(i,j)$th element of $\boldsymbol{H}f(\boldsymbol{x})$ is not identically zero. We identify each edge $(i,j) \in \mathcal{E}$ with $(j,i)$. Let $G(N, \overline{\mathcal{E}})$

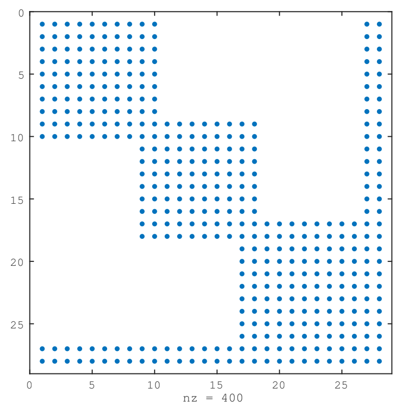


FIG. 5.1. *The arrow-type nonzero pattern of the Hessian matrix* $\boldsymbol{H}f(\boldsymbol{x})$ *of* $f(\boldsymbol{x})$. $m = 3,\ a = 2,\ b = 10,\ c = 2$.

be a chordal extension of $G(N, \mathcal{E})$, a chordal graph such that $\mathcal{E} \subset \overline{\mathcal{E}}$. Here a graph is called chordal when every cycle consisting of more than 3 edges has a chord. Let $C_1, \ldots, C_m \subset N$ denote the maximal cliques of $G(N, \overline{\mathcal{E}})$. See [6, 18] for chordal graphs and their fundamental properties.

EXAMPLE 5.1. (An arrow-type nonzero pattern) For nonnegative integers $m$, $a$, $b$ and $c$ with $m \geq 1$, $b \geq 1$ and $a \leq b/2$, let $n = b + (b-a)(m-1) + c,\ N = \{1, \ldots, n\}$,

$$B = \{1, \ldots, b\},\ F = \{1, \ldots, c\} + \{n - c\},$$
$$B_k = B + \{(b-a)(k-1)\}\ (k = 1, \ldots, m),\ C_k = B_k \cup F\ (k = 1, \ldots, m),$$

where + stands for the Minkowski sum of two sets. Every $C_k$ consists of two disjoint subsets $B_k$ and $F$ of $N$. We see that $|C_k| = b + c$, $|B_k \cap B_{k+1}| = a$ $(k = 1, \ldots, m-1)$ and $B_j \cap B_k = \emptyset$ if $|j - k| \geq 2$. We may regard $C_1, \ldots, C_m \subset N$ as the maximal cliques which characterize the nonzero pattern of the Hessian matrix $\boldsymbol{H}f(\boldsymbol{x})$ of a sparse polynomial $f(\boldsymbol{x})$. Figure 5.1 shows the nonzero pattern of the Hessian matrix $\boldsymbol{H}f(\boldsymbol{x})$ of such a polynomial with $m = 3$, $b = 10$, $a = 2$ and $c = 2$. We observe that each $B_k$ induces a $b \times b$ nonzero block along the diagonal and $F$ leads to $c$ nonzero rows and columns at the bottom and the right end of $\boldsymbol{H}f(\boldsymbol{x})$. Furthermore, the graph $G(N, \mathcal{E})$ with the node set $N = \{1, \ldots, n\}$ and the edge set $\mathcal{E} = \{(i, j) \in N \times N : (i, j) \in C_k \times C_k$ and $i < j$ for some $k\}$ forms a chordal graph having the maximal cliques $C_1, \ldots, C_m$. We see that this example with the described $\boldsymbol{H}f(\boldsymbol{x})$ of a sparse polynomial $f(\boldsymbol{x})$ illustrates the previous discussion.

To represent $f(\boldsymbol{x})$ as the sum of $m$ polynomials $f_k(\boldsymbol{x})$ $(k = 1, \ldots, m)$ with $\mathrm{supp} f_k = \boldsymbol{r}(\mathrm{supp} f_k) \subset \mathbb{Z}^{C_k}$ in an iterative way, we initially set $f_k(\boldsymbol{x}) \equiv 0$ $(k = 1, \ldots, m)$ and $\mathcal{F} = \mathrm{supp} f = \boldsymbol{r}(\mathrm{supp} f)$. Choose $\boldsymbol{\alpha} \in \mathcal{F}$. If $C = \{i \in N : \alpha_i \geq 1\} = \{i\}$ for some $i \in N$ or equivalently $\boldsymbol{\alpha} = c\boldsymbol{e}^i$ for some nonzero $c \in \mathbb{Z}_+$, then $C$ is contained in some maximal clique $C_k$ since $\bigcup_{k=1}^m C_k = N$. Thus, $\boldsymbol{\alpha} \in \mathbb{Z}_+^{C_k}$. Otherwise $C$ consists of more than one element, say, $1, \ldots, \ell$ or $\boldsymbol{\alpha} = \sum_{i=1}^{\ell} c_i \boldsymbol{e}^i$ for some nonzero $c_i \in \mathbb{Z}_+$. In this case, for every $(i, j) \in C \times C$, the $(i, j)$th element of the Hessian matrix $\boldsymbol{H}f(\boldsymbol{x})$ of $f(\boldsymbol{x})$ is not zero, so that $C$ forms a clique of $G(N, \mathcal{E})$. Hence, $C$ is contained in some maximal clique $C_k$ and $\boldsymbol{\alpha} \in \mathbb{Z}_+^{C_k}$. In both cases, update $f_k(\boldsymbol{x}) \leftarrow f_k(\boldsymbol{x}) + c(\boldsymbol{\alpha})\boldsymbol{x}^{\boldsymbol{\alpha}}$ and $\mathcal{F} \leftarrow \mathcal{F} \backslash \{\boldsymbol{\alpha}\}$. Repeating this procedure until $\mathcal{F}$ becomes empty, we obtain the desired $m$ polynomials to represent $f(\boldsymbol{x})$.

**5.3. Choosing $\mathcal{A}_k$ $(k = 1, \ldots, m)$ for sparse DNN relaxations of QOPs and POPs with binary and box constraints.** Suppose that $\deg f = 2$ in POP (1.1) to deal with a QOP with binary and box constraints. If we represent the quadratic objective function $f(\boldsymbol{x}) = \langle \boldsymbol{Q}^0, \boldsymbol{x}\boldsymbol{x}^T \rangle + \boldsymbol{a}\boldsymbol{x}^T$, then the Hessian matrix $\boldsymbol{H}f(\boldsymbol{x})$ coincides wtih $\boldsymbol{Q}^0$. Therefore, the nonzero pattern of $\boldsymbol{Q}^0$ determines the graph $G(N, \mathcal{E})$, which induces a chordal extension of $G(N, \overline{\mathcal{E}})$ and its maximal cliques $C_1, \ldots, C_m$. We choose $\left\{\boldsymbol{0},\ \boldsymbol{e}^i\ (i \in C_k)\right\}$ for $\mathcal{A}^\omega_k$ $(k = 1, \ldots, m)$ to construct the cone $\mathbb{L}_{\rm s}$ and the sparse DNN relaxation of the form (2.1).

We now extend the hierarchy of dense DNN relaxations of POP (1.1) with binary and box constraints presented in Section 4.4, to the sparse case. Let $\omega_0$ be the smallest positive integer not less than $(\deg f)/2$. First, construct nonempty subsets $C_k$ of $N$ and polynomial functions $f_k(\boldsymbol{x})$ with $\mathrm{supp} f_k = \boldsymbol{r}(\mathrm{supp} f_k \subset \mathbb{Z}^{C_k})$ $(k = 1, \ldots, m)$ from the Hessian matrix $\boldsymbol{H}f(\boldsymbol{x})$ of $f(\boldsymbol{x})$ as in Section 5.2. For a positive integer $\omega \geq \omega_0$, let $\mathcal{A}^\omega_k = \left\{ \boldsymbol{r}(\boldsymbol{\alpha}) \in \mathbb{Z}^{C_k}_+ : \sum_{i \in N} \alpha_i \leq \omega \right\}$ $(k = 1, \ldots, m)$. Then, $\mathrm{supp} f_k \subset \mathcal{A}^\omega_k + \mathcal{A}^\omega_k$ holds $(k = 1, \ldots, m)$. Thus, we can find a matrix $\boldsymbol{Q}^0_k \in \mathbb{S}^{\mathcal{A}^\omega_k}$ such that $f_k(\boldsymbol{x}) = \langle \boldsymbol{Q}^0_k, \boldsymbol{x}^{\mathcal{A}^\omega_k} \rangle$ $(k = 1, \ldots, m)$ to lift POP (1.1) to the problem (5.1) in $\mathbb{R}^{\boldsymbol{\Theta}}$ with $\boldsymbol{\Theta} = \mathbb{S}^{\mathcal{A}_1} \times \cdots \mathbb{S}^{\mathcal{A}_k}$. Note that $\mathcal{A}^\omega_k$ $(k = 1, \ldots, m)$ with $I_{\rm bin}$ and $I_{\rm box}$ determine the preorder $\succeq$, the equivalence relation $\sim$, the family of maximal chains of equivalence classes, $\mathbb{L}_{\rm s}$, and eventually the sparse DNN relaxation of the form (2.1) with the relaxation order $\omega$.

**6. A full preorder $\succeq_{\rm f}$ in $\mathcal{A} \times \mathcal{A}$ and a Lagrangian-DNN relaxation.** Let $\mathcal{A}$ be a nonempty subset of $\mathbb{Z}^n_+$ satisfying (2.10), and let $\succeq$ denote the preorder introduced in Section 4. A binary relation $\succeq_{\rm f}$ in $\boldsymbol{\Theta} = \mathcal{A} \times \mathcal{A}$ defined by $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \succeq_{\rm f} (\boldsymbol{\gamma}, \boldsymbol{\delta}) \Leftrightarrow \boldsymbol{r}(\boldsymbol{\alpha} + \boldsymbol{\beta}) \leq \boldsymbol{r}(\boldsymbol{\gamma} + \boldsymbol{\delta})$ also serves a preorder in $\boldsymbol{\Theta}$ satisfying (4.1). We call $\succeq_{\rm f}$ the *full preorder* in $\mathcal{A} \times \mathcal{A}$.

LEMMA 6.1.

*(a) The preorders $\succeq$ and $\succeq_f$ induce a common equivalence relation $\sim$.*

*(b) For every pair of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $(\boldsymbol{\gamma}, \boldsymbol{\delta})$ in $\boldsymbol{\Theta}$, $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \succeq_f (\boldsymbol{\gamma}, \boldsymbol{\delta})$ if and only if $\boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}} \geq \boldsymbol{x}^{\boldsymbol{\gamma}+\boldsymbol{\delta}}$ for every $\boldsymbol{x} \in S$.*

*Proof.* Assertion (a) follows directly from the definitions of $\succeq$ and $\succeq_{\rm f}$. For assertion (b), it suffices to show that $\boldsymbol{r}(\boldsymbol{\sigma}) \leq \boldsymbol{r}(\boldsymbol{\tau})$ if and only if $\boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\sigma})} \geq \boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\tau})}$ for every $\boldsymbol{x} \in S$. The 'only if' part is obvious by definition. To prove the 'if part', we assume that $\boldsymbol{r}(\boldsymbol{\sigma}) \not\leq \boldsymbol{r}(\boldsymbol{\tau})$ holds, and we will show that $\boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\sigma})} < \boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\tau})}$ for some $\boldsymbol{x} \in S$. By the assumption, there exists an $i$ such that $r_i(\boldsymbol{\sigma}) > r_i(\boldsymbol{\tau})$. Fix $x_j = 1$ for every $j \neq i$. Then $x_j^{r_i(\boldsymbol{\sigma})} = x_j^{r_i(\boldsymbol{\tau})} = 1$ for every $j \neq i$. If $i \in I_{\rm bin}$, then $r_i(\boldsymbol{\sigma}) > r_i(\boldsymbol{\tau})$ implies that $r_i(\boldsymbol{\sigma}) = 1 > r_i(\boldsymbol{\tau}) = 0$. In this case, taking $x_i = 0$, we obtain that $\boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\sigma})} = x_i^{r_i(\boldsymbol{\sigma})} = 0 < 1 = x_i^{r_i(\boldsymbol{\tau})} = \boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\tau})}$. Otherwise, $i \in I_{\rm box}$ and $r_i(\boldsymbol{\sigma}) = \sigma_i > r_i(\boldsymbol{\tau}) = \tau_i$. In this case, taking $x_i = 0.5$, we obtain that $\boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\sigma})} = x_i^{\boldsymbol{\sigma}_i} < x_i^{\boldsymbol{\tau}_i} = \boldsymbol{x}^{\boldsymbol{r}(\boldsymbol{\tau})}$. $\square$

Now, let us consider to use the preorder $\succeq_{\rm f}$ instead of $\succeq$ to derive a DNN relaxation of POP (1.1), which will be called as the *full DNN relaxation*. By (b) of Lemma 6.1, the preorder $\succeq_{\rm f}$ may be regarded as the strongest preorder to generate inequalities between two monomials $\boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}$ and $\boldsymbol{x}^{\boldsymbol{\gamma}+\boldsymbol{\delta}}$ for every $\boldsymbol{x} \in S$. In particlular, the preorder $\succeq_{\rm f}$ is stronger than $\succeq$ in the sense that if $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \succeq (\boldsymbol{\gamma}, \boldsymbol{\delta})$ then $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \succeq_{\rm f} (\boldsymbol{\gamma}, \boldsymbol{\delta})$. This ensures that every chain with respect to $\succeq$ is a chain with respect to $\succeq_{\rm f}$. Therefore, the resulting cone, which is denoted by $\mathbb{L}_{\rm f}$, is expected to be smaller than the original one, and the resulting DNN relaxation of POP (1.1) is exepcted to be stronger than the DNN relaxation given in Section 4. However, the family of maximal chains with respect to $\succeq_{\rm f}$ are not necessarily disjoint with each other. This makes the computation of the metric projection from $\mathbb{S}^{\mathcal{A}}$ onto $\mathbb{L}_{\rm f}$ highly challenging and expensive. To apply Lemma 3.2 and Algorithm 3.3, we need to choose a family of chains (with respect to $\succeq_{\rm f}$) which partitions the family of equivalence classes $\{E_{\boldsymbol{\sigma}}\ (\boldsymbol{\sigma} \in \boldsymbol{r}(\mathcal{A}+\mathcal{A})\}$. The family of maximal chains with respect to $\succeq$ may be regarded as such a family.

In this section, we focus on QOP (2.5) with binary and box constraints to describe its full DNN relaxation using $\succeq_{\rm f}$, and show its close relations with the Lagrangian-DNN relaxation proposed in [2] for a class of QOPs with linear, binary and complementarity constraints.

**6.1. A full DNN relaxation of a QOP with binary and box constraints.** First we observe that the preorder $\succeq_{\rm f}$ induces the identities $x_i = x_i^2\ (i \in I_{\rm bin})$ and inequalities $x_i \geq x_i^2\ (i \in I_{\rm box})$, $x_i \geq x_i x_j\ (1 \leq i < j \leq n)$ which hold for every $\boldsymbol{x} \in S$. Hence, the finest family of chains of equivalence classes mentioned in Remark 3.1 leads to the *full DNN relaxation*

$$\psi_1 = \min \left\{ \left\langle \boldsymbol{Q}^0, \begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \right\rangle :\ X_{00} = 1, \begin{pmatrix} X_{00} & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in \mathbb{S}^{1+n}_+ \cap \mathbb{L}_{\rm f} \right\} \tag{6.1}$$

where

$$\mathbb{L}_{\rm f} = \left\{ \begin{pmatrix} V_{00} & \boldsymbol{v}^T \\ \boldsymbol{v} & \boldsymbol{V} \end{pmatrix} \in \mathbb{N}^{1+n} :\ \begin{array}{l} v_i = V_{ii}\ (i \in I_{\rm bin}), \\ v_i \geq V_{ii}\ (i \in I_{\rm box}), \\ v_i \geq V_{ij}\ (1 \leq i < j \leq n) \end{array} \right\}.$$

**6.2. Strengthening the full DNN relaxation** (6.1) **with slack variables.** Introducing slack variables $u_i \geq 0\ (i = 1, \ldots, n)$ into QOP (2.5), we consider

$$\min\{\langle \boldsymbol{Q}^0,\ (1, \boldsymbol{x})(1, \boldsymbol{x})^T \rangle : \boldsymbol{x} \in S,\ \boldsymbol{x} + \boldsymbol{u} = \boldsymbol{e},\ \boldsymbol{u} \in S\},$$

where $\boldsymbol{e}$ denotes the $n$-dimensional row vector of 1's. Obviously this QOP is equivalent to (2.5). Applying the discussion in Section 6.1 twice in the $\boldsymbol{x}$ space and $\boldsymbol{u}$ space, we have a DNN relaxation of this problem as

$$\psi_2 = \min \left\{ \left\langle \boldsymbol{Q}^0, \begin{pmatrix} X_{00} & \boldsymbol{x} \\ \boldsymbol{x}^T & \boldsymbol{X} \end{pmatrix} \right\rangle :\ \begin{array}{l} X_{00} = U_{00} = 1,\ \boldsymbol{x} + \boldsymbol{u} = \boldsymbol{e}, \\ \begin{pmatrix} X_{00} & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in \mathbb{S}^{1+n}_+ \cap \mathbb{L}_{\rm f}, \\ \begin{pmatrix} U_{00} & \boldsymbol{u}^T \\ \boldsymbol{u} & \boldsymbol{U} \end{pmatrix} \in \mathbb{S}^{1+n}_+ \cap \mathbb{L}_{\rm f} \end{array} \right\}. \tag{6.2}$$

We call (6.2) the *twin DNN relaxation* of QOP (2.5). It is easy to see that $\psi_1 \leq \psi_2$.

Another DNN relaxation will be derived with cones $\mathbb{K}^1 = \mathbb{S}^{1+2n}_+$ and $\mathbb{K}^2 = \mathbb{L}_{\rm d} \in \mathbb{S}^{1+2n}$, which is at least as strong as (6.2) and can be approximately solved by the accelerated BP algorithm (Algorithm 2.1) combined with Algorithm 3.3 for the computation the metric computation of $\mathbb{S}^{1+2n}$ onto $\mathbb{L}_{\rm d}$.

**6.3. Representing binary constraints with complementarity constraints.** Since each binary variable $x_i$ is characterized by $x_i + u_i = 1$, $x_i \geq 0$, $u_i \geq 0$ and $x_i u_i = 0$ with a slack box variable $u_i$, QOP (2.5) is equivalent to

$$\min \left\{ \langle \boldsymbol{Q}^0,\ (1, \boldsymbol{x})(1, \boldsymbol{x})^T \rangle :\ \begin{array}{l} x_0 = 1,\ \boldsymbol{x} \in [0,1]^n,\ \boldsymbol{u} \in [0,1]^n, \\ \left((-1,\ \boldsymbol{e}_j,\ \boldsymbol{e}_j)(1, \boldsymbol{x}, \boldsymbol{u})^T\right)^2 = 0\ (j = 1, \ldots, n), \\ x_i u_i = 0\ (i \in I_{\rm bin}) \end{array} \right\} \tag{6.3}$$

which becomes a box constrained QOP with equality constraints. Here $\boldsymbol{e}^j$ denotes the $j$th coordinate row vector in $\mathbb{R}^n$. Note that the equality constraints $\left((-1,\ \boldsymbol{e}_j,\ \boldsymbol{e}_j)(1,\boldsymbol{x},\boldsymbol{u})^T\right)^2 = 0$ $(j=1,\ldots,n)$ are equivalent to a linear equality $\boldsymbol{x}+\boldsymbol{u}=\boldsymbol{e}$. Define

$$
\begin{aligned}
&\boldsymbol{Q}^1_j = (-1,\ \boldsymbol{e}_j,\ \boldsymbol{e}_j)^T(-1,\ \boldsymbol{e}_j,\ \boldsymbol{e}_j) \in \mathbb{S}^{1+2n}_+ \ (j=1,\ldots,n),\\
&\boldsymbol{Q}^2_i = \begin{pmatrix} 0 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0}^T & \boldsymbol{O} & \boldsymbol{E}_{ii} \\ \boldsymbol{0}^T & \boldsymbol{E}_{ii} & \boldsymbol{O} \end{pmatrix} \in \mathbb{N}^{1+2n} \ (i\in I_{\rm bin}),\\
&\boldsymbol{E}_{ii} = \text{the } n\times n \text{ matrix with 1 at the } (i,i)\text{th element and 0 elsewhere } (i\in I_{\rm bin}).
\end{aligned}
$$

Then we can rewrite problem (6.3) as

$$
\min\left\{ \langle \boldsymbol{Q}^0,\ (1,\boldsymbol{x})(1,\boldsymbol{x})^T\rangle :\ \begin{array}{l} x_0 = 1, \boldsymbol{x}\in[0,1]^n,\ \boldsymbol{u}\in[0,1]^n, \\ \langle \boldsymbol{Q}^1_j,\ (1,\boldsymbol{x},\boldsymbol{u})(1,\boldsymbol{x},\boldsymbol{u})^T\rangle = 0 \ (j=1,\ldots,n), \\ \langle \boldsymbol{Q}^2_i,\ (1,\boldsymbol{x},\boldsymbol{u})(1,\boldsymbol{x},\boldsymbol{u})^T\rangle = 0 \ (i\in I_{\rm bin}). \end{array} \right\}
$$

Replacing $(1,\boldsymbol{x},\boldsymbol{u})(1,\boldsymbol{x},\boldsymbol{u})^T$ by $\boldsymbol{Z} = \begin{pmatrix} X_{00} & \boldsymbol{x} & \boldsymbol{u} \\ \boldsymbol{x}^T & \boldsymbol{X} & \boldsymbol{W} \\ \boldsymbol{u}^T & \boldsymbol{W}^T & \boldsymbol{U} \end{pmatrix}$ and taking account of $\boldsymbol{x}\in[0,1]^n,\ \boldsymbol{u}\in[0,1]^n$, we obtain a DNN relaxation of QOP (6.3):

$$
\psi_3 = \left\{ \left\langle \boldsymbol{Q}^0,\ \begin{pmatrix} X_{00} & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \right\rangle :\ \begin{array}{l} \boldsymbol{Z}\in\mathbb{S}^{1+2n}_+\cap\mathbb{L}_{\rm d},\ \langle \boldsymbol{H}^0,\ \boldsymbol{Z}\rangle = 0, \\ \langle \boldsymbol{Q}^1_j,\ \boldsymbol{Z}\rangle = 0 \ (j=1,\ldots,n), \\ \langle \boldsymbol{Q}^2_i,\ \boldsymbol{Z}\rangle = 0 \ (i\in I_{\rm bin}) \end{array} \right\}, \tag{6.4}
$$

where

$$
\begin{aligned}
&\boldsymbol{H}^0 = \text{the } (1+2n)\times(1+2n) \text{ matrix with 1 at the } (0,0)\text{th element},\\
&\mathbb{L}_{\rm d} = \left\{\boldsymbol{Z}\in\mathbb{N}^{1+2n} : x_i \ge X_{ii},\ u_i \ge U_{ii} \ (i=1,\ldots,n)\right\}.
\end{aligned}
$$

Since $\mathbb{L}_{\rm d}$ is constructed using the preorder $\succeq$ in the $(\boldsymbol{x},\boldsymbol{u})$ space, $\mathbb{R}^{2n}$ (see Section 4.3), Algorithm 3.3 can be applied for computation of the metric projection $\Pi_{\mathbb{L}_{\rm d}}$ from $\mathbb{S}^{1+2n}$ onto $\mathbb{L}_{\rm d}$.

Now, we prove that the DNN relaxation (6.4) is at least as strong as the twin DNN relaxation (6.2) by showing that $\psi_3 \ge \psi_2$. To see this, suppose that $\boldsymbol{Z}$ is a feasible solution of (6.4). Obviously,

$$
X_{00} = 1,\ \begin{pmatrix} X_{00} & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \in \mathbb{S}^{1+n}_+\cap\mathbb{N}^{1+n},\ \begin{pmatrix} X_{00} & \boldsymbol{u}^T \\ \boldsymbol{u} & \boldsymbol{U} \end{pmatrix} \in \mathbb{S}^{1+n}_+\cap\mathbb{N}^{1+n}.
$$

Since $\boldsymbol{Z}\in\mathbb{S}^{1+2n}_+$, we see from the identity $\langle \boldsymbol{Q}^1_j,\ \boldsymbol{Z}\rangle = 0$ $(j=1,\ldots,n)$ and the definition of $\boldsymbol{Q}^1_j$ that $\boldsymbol{Z}(-1,\ \boldsymbol{e}_j,\ \boldsymbol{e}_j)^T = \boldsymbol{0}$ $(j=1,\ldots,n)$, or equivalently,

$$
\begin{aligned}
&X_{00} - x_j - u_j = 0 \ (j=1,\ldots,n),\\
&x_i - X_{ij} - W_{ij} = 0 \ (1\le j\le n),\ u_i - U_{ij} - W_{ji} = 0 \ (1\le j\le n).
\end{aligned}
$$

We also know from $\langle \boldsymbol{Q}^2_i,\ \boldsymbol{Z}\rangle = 0$ $(i\in I_{\rm bin})$ and $\boldsymbol{W}\ge\boldsymbol{O}$ that $W_{ii}=0$ $(i\in I_{\rm bin})$, $W_{ii}\ge 0$ $(i\in I_{\rm box})$ and $W_{ij}\ge 0$ $(1\le j\le n)$. Thus, the previous relations imply

$$
\begin{aligned}
&\boldsymbol{x}+\boldsymbol{u} = \boldsymbol{e},\\
&x_i = X_{ii}\ (i\in I_{\rm bin}),\ x_i\ge X_{ii}\ (i\in I_{\rm box}),\ x_i \ge X_{ij}\ (1\le i<j\le n),\\
&u_i = U_{ii}\ (i\in I_{\rm bin}),\ u_i\ge U_{ii}\ (i\in I_{\rm box}),\ u_i \ge U_{ij}\ (1\le i<j\le n),\\
&x_i - W_{ij}\ge 0\ (1\le i\le j\le n),\ u_i - W_{ji}\ge 0\ (1\le i\le j\le n).
\end{aligned}
$$

As a result, both $\begin{pmatrix} X_{00} & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix}$ and $\begin{pmatrix} X_{00} & \boldsymbol{u}^T \\ \boldsymbol{u} & \boldsymbol{U} \end{pmatrix}$ lie in $\mathbb{L}_{\rm f}$, and they provide a feasible solution of (6.2). Consequently, $\psi_3 \geq \psi_2$ follows. The inequalities in the last line above are irrelevant for the conclusion, but it shows that the DNN relaxation (6.4) also incorporates the inequalities induced from $x_i - x_i u_j \geq 0$ $(1 \leq i \leq j \leq n)$ and $u_i - x_j u_i \geq 0$ $(1 \leq i \leq j \leq n)$.

**6.4. A Lagrangian-DNN relaxation.** Since $\langle \boldsymbol{Q}^1_j, \boldsymbol{Z}\rangle \geq 0$ for every $\boldsymbol{Z} \in \mathbb{S}^{1+n}_+$ $(j = 1,\ldots,n)$ and $\langle \boldsymbol{Q}^2_i, \boldsymbol{Z}\rangle \geq 0$ for every $\boldsymbol{Z} \in \mathbb{N}^{1+n}_+$ $(i \in I_{\rm bin})$, we can rewrite (6.4) as

$$\psi_3 = \min\left\{ \left\langle \boldsymbol{Q}^0, \begin{pmatrix} X_{00} & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \right\rangle : \begin{array}{l} \boldsymbol{Z} \in \mathbb{S}^{1+2n}_+ \cap \mathbb{L}_{\rm d},\ \langle \boldsymbol{H}^0, \boldsymbol{Z}\rangle = 1, \\ \langle \boldsymbol{H}^1, \boldsymbol{Z}\rangle = 0, \end{array} \right\}, \tag{6.5}$$

where $\boldsymbol{H}^1 = \sum_{j=1}^n \boldsymbol{Q}^1_j + \sum_{i\in I_{\rm bin}} \boldsymbol{Q}^2_i$. We notice that (6.5) is not in the form COP (2.1) yet. The Lagrangian relaxation is further applied to (6.5) with a $\lambda > 0$ to obtain the Lagrangian-DNN relaxation of (6.3)

$$\psi_4(\lambda) = \min\left\{ \left\langle \boldsymbol{Q}^0, \begin{pmatrix} X_{00} & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \right\rangle + \lambda\langle \boldsymbol{H}^1, \boldsymbol{Z}\rangle : \begin{array}{l} \boldsymbol{Z} \in \mathbb{S}^{1+2n}_+ \cap \mathbb{L}_{\rm d}, \\ \langle \boldsymbol{H}^0, \boldsymbol{Z}\rangle = 1, \end{array} \right\}, \tag{6.6}$$

which is in the same form as COP (2.1). Since the term $\langle \boldsymbol{H}^1, \boldsymbol{Z}\rangle$ is nonnegative for every $\boldsymbol{Z} \in \mathbb{S}^{1+2n}_+ \cap \mathbb{L}_{\rm d}$, it serves as a penalty term such that if $\boldsymbol{Z} \in \mathbb{S}^{1+2n}_+ \cap \mathbb{L}_{\rm d}$ is not feasible for (6.5), then the term diverges to $\infty$ as $\lambda > 0$ tends to $\infty$. Using this fact, it is easy to prove that the optimal value $\psi_4(\lambda)$ of (6.6) converges to the optimal value $\psi_3$ of (6.5) as $\lambda > 0$ tends to $\infty$.

Applying the Lagrangian-DNN relaxation (6.6) to QOP (6.3) with box constraints serves as a a relaxation of QOP (2.5) with binary and box constraitns. As a result, we can say that the Lagrangian-DNN relaxation of (6.3) provides another way to strengthen the standard relaxation of QOP (2.5) mentioned in Sections 2.3 and 4.3, in addition to applying the hierarchy of DNN relaxations to (2.5). This Lagrangian-DNN relaxation was originally proposed in [2] for the CPP reformulation of a class of QOPs with linear, binary and complementarity constraints. See also [13, 4]. We compare the standard DNN relaxation applied to the QOP (2.5) and the Lagraingian-DNN relaxation applied to (6.3) through numerical results in Section 7.3, which shows that the Lagrangian-DNN relaxation is more effective in obtaining a tight lower bound for the optimal value of QOP (2.5), and that the standard DNN relaxation requires less computational time.

**7. Preliminary numerical results.** We tested Algorithm 2.1 (the accelerated BP algorithm) on randomly generated binary QOPs and POPs and the maximum complete satisfiability problem [11]. The purpose of the numerical experiments is to demonstrate the performance of Algorithm 2.1 in comparison to the primal-dual interior-point method in solving the hierarchy of DNN relaxations.

Each dense test problem solved by Algorithm 2.1 is of the form (2.1) with $\mathbb{K}^1 = \mathbb{S}^{\mathcal{A}}_+$ and $\mathbb{K}^2 = \mathbb{L}_{\rm d} \subset \mathbb{S}^{\mathcal{A}}$ for some $\mathcal{A} \subset \mathbb{Z}^n_+$ (see Section 4), while each sparse test problem is of the form (2.1) with $\mathbb{K}^1 = \mathbb{S}^{\mathcal{A}_1}_+ \times \cdots \times \mathbb{S}^{\mathcal{A}_m}_+$ and $\mathbb{K}^2 = \mathbb{L}_{\rm s} \subset \mathbb{S}^{\mathcal{A}_1} \times \cdots \times \mathbb{S}^{\mathcal{A}_1}$ for some $\mathcal{A}_1,\ldots,\mathcal{A}_m \subset \mathbb{Z}^n_+$ (see Section 5). In both cases, $\mathbb{K}^2$ is a polyhedral cone, so that (2.1) is equivalent to an SDP of minimizing the same objective function $\langle \boldsymbol{Q}^0, \boldsymbol{X}\rangle$ subject to $\langle \boldsymbol{H}^0, \boldsymbol{X}\rangle = 1$, $\boldsymbol{X} \in \mathbb{K}^1$ and the additional equalities and inequalities describing the cone $\mathbb{K}^2$. Such an SDP problem was constructed directly from QOP (2.5) and

POP (1.1) by SparsePOP [24], and then SDPT3 [22], an implementation of the primal-dual interior-point method, was applied to the SDP. All the numerical experiments were run in Matlab on a Mac Pro with Intel Xeon E5 8-core CPU (3.0 GHZ) and 64 GB memory.

**7.1. Dense binary POPs.** We randomly generated binary POPs with the degrees of the objective polynomials being 3 and 5. Algorithm 2.1 was applied to COPs of the form (2.1), which is induced from the dense DNN relaxation of the binary POPs. For the hierarchy of DNN relaxations, the relaxation order 2 and 3 were taken, respectively. See Section 4.4.

We observe from Table 7.1 that

(i) Both Algorithm 2.1 and SDPT3 attain tight lower bounds for small-sized problems.

(ii) For large-sized problems, Algorithm 2.1 computes lower bounds more reliably in less time than SDPT3.

(iii) The lower bounds obtained by SDPT3 deteriorate as the size of the problem increases.

(iv) Algorithm 2.1 requires much less memory than SDPT3 for large-sized problems.

The reason for (iii) is that the dual of the SDP problem converted from a DNN problem has no interior feasible solution and such an ill-posed problem is often difficult to solve by primal-dual interior-point methods. However, the ill-posedness of the problem with no interior feasible solution does not affect the performance of Algorithm 2.1.

Table 7.1

*Randomly generated dense binary POPs, POP* (1.1) *with* $S = \{0,1\}^n$. *Dense DNN means the dense DNN relaxation of POP* (1.1) *in Section 4.4.* † : *SDPT3 failed to attain an accurate optimal solution, and stopped with relative duality gap* $\geq$ 1.0e-2.

| | | | Lower bounds (seconds) | |
|---|---|---|---|---|
| | | | SDPT3 | Algorithm 2.1 |
| Deg | $n$ | Opt.Val | Dense DNN | Dense DNN |
| 3 | 10 | -1.0746e1 | -1.0746e1 (8.36e-1) | -1.0746e1 (2.89e0) |
| | 15 | -2.7910e1 | -2.7910e1 (7.92e0) | -2.7911e1 (5.68e0) |
| | 20 | -3.2664e1 | -4.1077e1 (2.83e2)$^\dagger$ | -3.2667e1 (2.82e1) |
| | 30 | | -1.0620e2 (6.29e3)$^\dagger$ | -7.9430e1 (1.54e2) |
| | 35 | | Out of Memory | -1.0047e2 (2.19e2) |
| | 70 | | | -1.0268e4 (1.63e4) |
| 5 | 5 | -3.3130e0 | -3.3130e0 (3.10e-1) | -3.3130e0 (2.50e0) |
| | 10 | -1.5062e1 | -1.5062e1 (2.29e1) | -1.5063e1 (9.76e0) |
| | 15 | -3.4952e1 | -3.9496e1 (3.18e3)$^\dagger$ | -3.4953e1 (2.35e2) |
| | 20 | -1.3405e2 | Out of Memory | -1.3408e2 (3.91e3) |

**7.2. Sparse binary POPs.** The test problems in Table 7.2 are randomly generated binary POPs of the form POP (1.1) with $S = \{0,1\}^n$ and $\deg f = 3,\ 5$. The optimal value was computed by generating all feasible solutions for the problems with $n \leq 20$. The nonzero pattern of the $n \times n$ Hessian matrix $\boldsymbol{H}f(\boldsymbol{x})$ of $f(\boldsymbol{x})$ in each problem is of the arrow-type as shown in Example 5.1, which is characterized by the maximal cliques $C_1, \ldots, C_m$ of a chordal graph. See also Figure 5.1. We fixed $a = 2$, $b = 10$ and $c = 2$, where the size of each clique is $b + c = 12$, and increased $m$ from 2 to 120 for the problems of degree 3, and 2 to 12 for the problems of degree 5. The relaxation order $\omega = 2$ for the problems of degree 3, and $\omega = 3$ for the problems of degree 5 were used.

The observations (i), (ii), (iii) and (iv) for the dense DNN relaxation in Section 7.1

remain valid if Algorithm 2.1 is replaced by Algorithm 2.1 applied to Sparse DNN. In addition, Table 7.2 demonstrates the efficiency of the sparse DNN relaxation of POP (1.1) in Section 5.3, compared to the dense DNN relaxation in Section 4.4.

TABLE 7.2
*Randomly generated sparse binary POPs, POP* (1.1) *with* $S = \{0,1\}^n$. *Dense DNN: the dense DNN relaxation of POP* (1.1) *in Section 4.4. Sparse DNN: the sparse DNN relaxation of POP* (1.1) *in Section 5.3.* † : *SDPT3 failed to attain an accurate optimal solution, and stopped with relative duality gap* $\geq$ 1.0e-2.

| | | | | Lower bounds (seconds) | | |
|---|---|---|---|---|---|---|
| | | | | SDPT3 | Algorithm 2.1 | |
| deg | $m$ | $n$ | Opt.Val | Sparse DNN | Dense DNN | Sparse DNN |
| 3 | 2 | 20 | -2.2059e1 | -2.2059e1 (4.57e0) | -2.2084e1 (3.43e1) | -2.2059e1 (1.77e1) |
| 3 | 4 | 36 | | -4.0217e1 (2.66e1) | -4.0557e1 (5.11e2) | -4.0218e1 (2.26e1) |
| 3 | 6 | 52 | | -6.8233e1 (6.14e1) | -6.8243e1 (4.92e3) | -6.8235e1 (3.91e1) |
| 3 | 20 | 164 | | -2.0396e2 (3.46e2) | | -2.0380e2 (2.01e2) |
| 3 | 40 | 324 | | -4.2420e2 (2.70e3)† | | -4.2275e2 (2.99e2) |
| 3 | 60 | 484 | | Out of Memory | | -6.5594e2 (5.00e2) |
| 3 | 120 | 964 | | | | -1.2603e3 (8.64e2) |
| 5 | 2 | 20 | -5.0021e1 | -5.0423e1 (3.93e2) | | -5.0023e1 (2.69e2) |
| 5 | 4 | 36 | | -1.1732e2 (5.49e3)† | | -9.1637e1 (2.51e2) |
| 5 | 6 | 52 | | -1.3377e2 (1.22e4)† | | -1.1799e2 (6.85e2) |
| 5 | 12 | 100 | | | | -2.7068e2 (1.91e3) |

In Table 7.3, we show the numerical results on the maximum complete satisfiability problem [11]. Let $x_i$ denote a boolean variable which takes 1 for true and 0 for false ($i = 1, \ldots, n$). For each $k = 1, \ldots, m$, let $I_k$ and $J_k$ be disjoint subsets of $\{1, \ldots, n\}$, and define a conjunctive clause $(\wedge_{i \in I_k} x_i) \wedge (\wedge_{j \in J_k}(1 - x_j))$ . The maximum complete satisfiability problem [11] is to find an $\boldsymbol{x} = (x_1, \ldots, x_n) \in \{0,1\}^n$ that maximizes the total weighted sum of the satisfied clauses. This problem is different from the usual maximum satisfiability problem (see *e.g*, [12]) and it is NP-hard in general. Let $\boldsymbol{w} = (w_1, \ldots, w_m)$ denote a weight vector. Then the problem is formulated as a binary POP with the objective polynomial $f(\boldsymbol{x}) = \sum_{k=1}^m w_k \left(\Pi_{i \in I_k} x_i\right) \left(\Pi_{j \in J_k}(1 - x_j)\right)$. For the numerical experiment, each problem is constructed as follows: Let $10 \leq m = n \leq 120$, $d = 4$, and choose $w_k$ randomly from $\{1, \ldots, n\}$ ($k = 1, \ldots, m$). Choose $d$ elements for $H_k$ ($k = 1, \ldots, m$) with the first element $k$ and the other $d - 1$ elements randomly from $\{1, \ldots, n\}$; eliminate any duplicated element from $H_k$. Finally, distribute the elements of $H_k$ randomly to $I_k$ and $J_k$ ($k = 1, \ldots, m$) so that $H_k = I_k \cup J_k$ and $I_k \cap J_k = \emptyset$.

In Table 7.3, we can make the same observations as (i), (ii), (iii) and (iv) in Section 7.1 for Algorithm 2.1 applied to Sparse DNN, instead of Algorithm 2.1 applied to Dense DNN. The structured sparsity of the test problems does not exist to the extent to be exploited in comparison to the sparse binary POP instances presented in Table 7.2 since each conjunctive clause is generated randomly. Thus, the sparse DNN relaxation provides no great advantage over the dense DNN relaxation. Nevertheless, as the number of conjunctive clauses with the fixed size $d = 4$ increases, the structured sparsity characterized by a chordal graph gradually increase so that Sparse DNN becomes faster than Dense DNN in Table 7.3.

**7.3. Binary and box constrained QOPs.** For the experiments on fully dense QOPs, QOPs with binary and box constraints were randomly generated. The results are displayed in Table 7.4 and 7.5, respectively. The number of variables $n$ ranges from 10 to 800. When $n \leq 20$ in Table 7.4, the optimal value was computed by

TABLE 7.3
*The maximum complete satisfiability problem [11]. Dense DNN: the dense DNN relaxation of POP* (1.1) *in Section 4.4. Sparse DNN: the sparse DNN relaxation of POP* (1.1) *in Section 5.3.* † *: SDPT3 failed to attain an accurate optimal solution, and stopped with relative duality gap* $\geq$ 1.0e-2.

| | | Lower bounds (seconds) | | |
|---|---|---|---|---|
| | | SDPT3 | Algorithm 2.1 | |
| $n$ | Opt.Val | Sparse DNN | Dense DNN | Sparse DNN |
| 10 | -2.8000e1 | -2.8000e1 (5.07e-1) | -2.8000e1 (4.79e0) | -2.8001e1 (5.43e0) |
| 15 | -4.7000e1 | -4.7000e1 (1.78e0) | -4.7025e1 (1.42e1) | -4.7008e1 (3.42e1) |
| 20 | -7.7000e1 | -7.7000e1 (6.09e0) | -7.7043e1 (8.69e1) | -7.7034e1 (8.03e1) |
| 30 | | -2.3129e2 (3.68e1) | -2.3128e2 (2.32e2) | -2.3104e2 (2.12e2) |
| 40 | | -5.0975e2 (5.13e2)$^{\dagger}$ | -3.3940e2 (1.08e3) | -3.3904e2 (4.69e2) |
| 50 | | -5.3206e2 (2.93e3)$^{\dagger}$ | -5.2237e2 (3.41e3) | -5.2205e2 (1.07e3) |
| 60 | | Out of Memory | | -8.1284e2 (1.31e3) |
| 90 | | | | -1.7122e3 (1.90e4) |
| 120 | | | | -5.7807e3 (7.27e4) |

generating all feasible solutions.

In Tables 7.4 and 7.5, we observe that SDPT3 solved the SDP, which is equivalent to COP (2.1), slightly more accurately than Algorithm 2.1, but spent much longer time and required more memory as $n$ increased. Notice that SDPT3 could not solve the problem with $n = 400$ due to "out of memory" error in Matlab, while Algorithm 2.1 could solve successfully the problems of $n = 800$.

In Section 6, we have shown that the Lagrangian-DNN relaxation of (2.5), which can be also solved by Algorithm 2.1, provides at least as tight lower bounds as the standard DNN relaxation derived from QOP (2.5) in Sections 2.3 and 4.3. The last column of Tables 7.4 and 7.5 verifies this theoretical assertion. Algorithm 2.1 consumed longer time to solve the Lagrangian-DNN relaxation than to solve the standard DNN relaxation, but is still much faster than SDPT3 in solving the standard DNN relaxation for large-sized QOPs.

TABLE 7.4
*Randomly generated dense binary QOPs, QOP* (2.5) *with* $S = \{0,1\}^n$. *Dense DNN: the standard DNN relaxation derived from QOP* (2.5) *in Section 4.3. Dense Lag-Dual: the Lagrangian-DNN relaxation of* (2.5) *presented in Section 6.*

| | | Lower bounds (seconds) | | |
|---|---|---|---|---|
| | | SDPT3 | Algorithm 2.1 | |
| $n$ | Opt.Val | Dense DNN | Dense DNN | Dense Lag-DNN |
| 10 | -5.8817e0 | -5.9802e0 (2.11e-1) | -5.9804e0 (1.86e0) | -5.8851e0 (1.96e0) |
| 20 | -1.7669e1 | -1.7833e1 (3.27e-1) | -1.7835e1 (1.03e0) | -1.7699e1 (2.42e0) |
| 100 | | -1.5645e2 (5.21e1) | -1.5647e2 (9.85e0) | -1.5109e2 (3.10e1) |
| 200 | | -5.1778e2 (1.85e3) | -5.1788e2 (2.80e1) | -5.0562e2 (1.13e2) |
| 300 | | -9.2689e2 (2.09e4) | -9.2710e2 (9.62e1) | -9.0634e2 (3.21e2) |
| 400 | | Out of Memory | -1.4219e3 (1.46e2) | -1.3912e3 (4.92e2) |
| 800 | | | -4.3072e3 (9.22e2) | -4.2293e3 (2.58e3) |

**8. Concluding remarks.** For POPs with binary and box constraints, we have provided a theoretical framework under which many DNN relaxations can be obtained and uniformly formulated as a simple COP. Moreover, the computation of the metric projection onto the polyhedral cone in the COP can be carried out efficiently and accurately. The framework has also been used to prove why the (accelerated) BP algorithm was successful in obtaining tight bounds when applied to the COP from the QOPs in [4, 13]. As the most important step of the BP algorithm is in computing

Table 7.5

*Randomly generated dense box constrained QOPs, QOP (2.5) with $S = [0,1]^n$. Dense DNN: the standard DNN relaxation derived from QOP (2.5) in Section 4.3. Dense Lag-Dual: the Lagrangian-DNN relaxation of (2.5) in Section 6.*

| | Lower bounds (sec) | | |
|---|---|---|---|
| | SDPT3 | Algorithm 2.1 | |
| $n$ | Dense DNN | Dense DNN | Dense Lag-DNN |
| 100 | -2.1397e2 (4.50e1) | -2.1399e2 (8.38e0) | -2.0544e2 (1.17e1) |
| 200 | -6.9528e2 (1.87e3) | -6.9535e2 (3.78e1) | -6.7238e2 (6.50e1) |
| 300 | -1.3139e3 (2.05e4) | -1.3140e3 (1.15e2) | -1.2753e3 (1.97e2) |
| 400 | Out of Memory | -2.1484e3 (2.00e2) | -2.1000e3 (3.99e2) |
| 800 | | -5.7459e3 (9.70e2) | -5.6527e3 (2.32e3) |

the metric projection, the framework presented in this paper expands the applicability of the BP algorithm to the POPs from the QOPs. In fact, a wide range of general POPs can be solved with the proposed method as described in the following.

We see that any variable $y_i$ bounded by an interval $[\ell_i, u_i]$ with $-\infty < \ell_i < u_i < \infty$ can be scaled to $x_i = (y_i - \ell_i)/(u_i - \ell_i) \in [0,1]$. Thus, it is possible to assume that all the lower and upper bounds for variables are 0 and 1 in POPs with bounded variables. Moreover, if an additional inequality constraint $h(\boldsymbol{x}) \geq 0$ is included in POP (1.1), it can be written as an quality constraint $h(\boldsymbol{x}) - ay = 0$ with a slack variable $y \geq 0$, where $a$ is a positive number. Since $h(\boldsymbol{x})$ is bounded in $S$, $y \in [0,1]$ can be added by taking a sufficiently large $a > 0$.

Now, suppose that equality constraints $g_i(\boldsymbol{x}) = 0$ $(i = 1,\ldots,m)$ are added to (1.1):

$$\text{minimize } f_0(\boldsymbol{x}) \text{ subject to } \boldsymbol{x} \in S, \; g_i(\boldsymbol{x}) = 0 \; (i = 1,\ldots,m). \tag{8.1}$$

Assume that the resulting problem (8.1) is feasible. Obviously, each equality constraint $g_i(\boldsymbol{x}) = 0$ can be rewritten as $g_i(\boldsymbol{x})^2 = 0$. Let $\lambda > 0$ be a sufficiently large number. Consider a Lagrangian relaxation of (8.1):

$$\text{minimize } f_0(\boldsymbol{x}) + \lambda \textstyle\sum_{i=1}^m g_i(\boldsymbol{x})^2 \text{ subject to } \boldsymbol{x} \in S. \tag{8.2}$$

Notice that $g_i(\boldsymbol{x})^2 \geq 0$ for every $\boldsymbol{x} \in \mathbb{R}^n$. Hence the term $\lambda \sum_{i=1}^m g_i(\boldsymbol{x})^2$ added to the objective function $f_0(\boldsymbol{x})$ of (8.1) serves as a penalty in the sense that if $\boldsymbol{x} \in S$ is not a feasible solution of (8.1), then the objective value $f_0(\boldsymbol{x}) + \lambda \sum_{i=1}^m g_i(\boldsymbol{x})^2$ of (8.2) diverges to $+\infty$ as $\lambda$ tends to $+\infty$. Using the compactness of the feasible region $S$ of (8.2), it is easy to prove that the optimal value of (8.2) converges to the optimal value of (8.1) as $\lambda \to \infty$. Consequently, a lower bound of the optimal value of POP (8.1) can be computed by applying Algorithm 2.1 to the DNN relaxation of POP (8.2) with binary and box constraints for a sufficiently large $\lambda > 0$.

**Acknowledgements.** The authors would like to thank Professor Akiko Takeda for suggesting the maximum complete satisfiability problem as a test example.

## REFERENCES

[1] N. Arima, S. Kim, and M. Kojima. A quadratically constrained quadratic optimization model for completely positive cone programming. *SIAM J. Optim.*, 23(4):2320–2340, 2013.

[2] N. Arima, S. Kim, and M. Kojima. Simplified copositive and lagrangian relaxations for linearly constrained quadratic optimization problems in continuous and binary variables. *Pacific J. of Optim.*, 10:437–451, 2014.

[3] N. Arima, S. Kim, M. Kojima, and K. C. Toh. Lagrangian-conic relaxations, part I: A unified framework and its applications to quadratic optimization problems. Research report B-475, Tokyo Institute of Technology, Department of Mathematical and Computing Sciences, Oh-Okayama, Meguro-ku, Tokyo 152-8552, January 2014.

[4] N. Arima, S. Kim, M. Kojima, and K.C. Toh. A robust Lagrangian-DNN method for a class of quadratic optimization problems. Research Report B-482, Tokyo Institute of Technology, Department of Mathematical and Computing Sciences, Oh-Okayama, Meguro-ku, Tokyo 152-8552, February 2016.

[5] A. Beck and Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2:183–202, 2009.

[6] J. R. S. Blair and Peyton B. An introduction to chordal graphs and clique trees. In Liu J.W.H. George A., Gilbert J. R., editor, *Graph Theory and Sparse Matrix Computation*. Springer-Verlag, New York, 1993.

[7] S. Burer. Optimizatin a polyhedral-semidefinite relaxation of completely positive programs. *Math. Prog. Comp.*, 2:1–19, 2010.

[8] T. Fujie and M. Kojima. Semidefinite programming relaxation for nonconvex quadratic programs. *J. of Global Optim.*, 10:367–380, 1997.

[9] M. Fukuda, M. Kojima, K. Murota, and K. Nakata. Exploiting sparsity in semidefinite programming via matrix completion. I: General framework. *SIAM J. Optim.*, 11:647–674, 2000.

[10] D. Ge and Y. Ye. On doubly positive semidefinite programming relaxations. http://www.optimization-online.org/DB_HTML/2010/08/2709.html, August 2010.

[11] S. He, Z. Li, and S. Zhang. Approximation algortithms for discrete polynomial optimization. *J. of Oper. Res. Soc. China*, 1(1):3–36, 2013.

[12] D. S. Johnson. Approximation algorithms for combinatorial prolblems. *J. Comput. Syst. Sci.*, 9:256–278, 1974.

[13] S. Kim, M. Kojima, and K. C. Toh. A lagrangian-dnn relaxation: a fast method for computing tight lower bounds for a class of quadratic optimization problems. *Math. Program.*, 156:161–187, 2016.

[14] M. Kobayashi, S. Kim, and M. Kojima. Correlative sparsity in primal-dual interior-point methods for lp, sdp, and socp. *Appl. Math. Optim.*, 58:69–88, 2008.

[15] M. Kojima, S. Shindoh, and S. Hara. Interior-point methods for the monotone linear complementarity problem in symmetric matrices. *SIAM J. Optim.*, 7:86–125, 1997.

[16] J. B. Lasserre. An explicit exact SDP relaxation for nonlinear 0-1 programs. In *Integer Programming and Combinatorial Optimization*, pages 293–303. Springer, 2001.

[17] J. B. Lasserre. Global optimization with polynomials and the problems of moments. *SIAM J. Optim.*, 11:796–817, 2001.

[18] Columbic M.C. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, New York, 1980.

[19] J. J. Moreau. Décomposition orthogonale d'un espace hilbertien selon deux cones mutuellement polaires. *C. R. Acad. Sci.*, 255:238–240, 1962.

[20] Y. E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. Optim.*, 8:324–363, 1998.

[21] N. Z. Shor. Quadratic optimization problems. *Soviet journal of Computer and Systems Sciences*, 25:1–11, 1987.

[22] R. H. Tütüncü, K. C. Toh, and M. J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Math. Program.*, 95:189–217, 2003.

[23] H. Waki, S. Kim, M. Kojima, and M. Muramatsu. Sums of squares and semidefinite programming relaxations for polynomial optimization problems with structured sparsity. *SIAM J. Optim.*, 17:218–242, 2006.

[24] H. Waki, S. Kim, M. Kojima, M. Muramatsu, and H. Sugimoto. Algorithm 883: Sparsepop: A sparse semidefinite programming relaxation of polynomial pptimization problems. *ACM Trans. Math. Soft.*, 35(15), 2008.

[25] A. Yoshise and Y. Matsukawa. On optimization over the doubly nonnegative cone. IEEE Multi-conference on Systems and Control, 2010.