

BBCPOP: A Sparse Doubly Nonnegative Relaxation of Polynomial Optimization Problems with Binary, Box and Complementarity Constraints

N. Ito^{*}, S. Kim[†], M. Kojima[‡], A. Takeda[§], and K.-C. Toh[¶]

March, 2018

Abstract. The software package BBCPOP is a MATLAB implementation of a hierarchy of sparse doubly nonnegative (DNN) relaxations of a class of polynomial optimization (minimization) problems (POPs) with binary, box and complementarity (BBC) constraints. Given a POP in the class and a relaxation order, BBCPOP constructs a simple conic optimization problem (COP), which serves as a DNN relaxation of the POP, and then solves the COP by applying the bisection and projection (BP) method. The COP is expressed with a linear objective function and constraints described as a single hyperplane and two cones, which are the Cartesian product of positive semidefinite cones and a polyhedral cone induced from the BBC constraints. BBCPOP aims to compute a tight lower bound for the optimal value of a large-scale POP in the class that is beyond the comfort zone of existing software packages. The robustness, reliability and efficiency of BBCPOP are demonstrated in comparison to the state-of-the-art software SDP package SDPNAL+ on randomly generated sparse POPs of degree 2 and 3 with up to a few thousands variables, and ones of degree 4, 5, 6. and 8 with up to a few hundred variables. Comparison with other BBC POPs that arise from combinatorial optimization problems such as quadratic assignment problems are also reported. The software package **BBCPOP** is available at <https://sites.google.com/site/bbcpop1/>.

Key words. MATLAB software package, High-degree polynomial optimization problems with binary, box and complementarity constraints, Hierarchy of doubly nonnegative relaxations, Sparsity, Bisection and projection methods, Tight lower bounds, Efficiency.

AMS Classification. 90C20, 90C22, 90C25, 90C26.

^{*} Department of Mathematical Informatics, The University of Tokyo, Tokyo 113-8656, Japan. This work was supported by Grant-in-Aid for JSPS Research Fellowship JP17J07365. (naoki_ito@mist.i.u-tokyo.ac.jp).

[†] Department of Mathematics, Ewha W. University, 52 Ewhayeodae-gil, Sudaemoon-gu, Seoul 03760 Korea. The research was supported by NRF 2017-R1A2B2005119. (skim@ewha.ac.kr).

[‡] Department of Industrial and Systems Engineering, Chuo University, Tokyo 112-8551 Japan. This research was supported by Grant-in-Aid for Scientific Research (A) 26242027. (kojimamasakzu@mac.com).

[§] Department of Mathematical Analysis and Statistical Inference, The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan The work of this author was supported by Grant-in-Aid for Scientific Research (C), 15K00031. (atakeda@ism.ac.jp).

[¶] Department of Mathematics and Institute of Operations Research and Analytics, 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-146-000-256-114. (mattohkc@nus.edu.sg).

1 Introduction

We introduce a Matlab package BBCPOP for computing a tight lower bound of the optimal value of large-scale sparse polynomial optimization problems (POPs) with binary, box and complementarity (BBC) constraints. Let f_0 be a real valued polynomial function defined on the n -dimensional Euclidean space \mathbb{R}^n , I_{box} and I_{bin} a partition of $N \equiv \{1, 2, \dots, n\}$, *i.e.*, $I_{\text{box}} \cup I_{\text{bin}} = N$ and $I_{\text{box}} \cap I_{\text{bin}} = \emptyset$, and \mathcal{C} a family of subsets of N . BBCPOP finds a lower bound for the optimal value ζ^* of POPs described as

$$\zeta^* = \min_{\mathbf{x}} \left\{ f_0(\mathbf{x}) \left| \begin{array}{l} x_i \in [0, 1] \ (i \in I_{\text{box}}) \text{ (box constraint),} \\ x_j \in \{0, 1\} \ (j \in I_{\text{bin}}) \text{ (binary constraint),} \\ \prod_{j \in C} x_j = 0 \ (C \in \mathcal{C}) \text{ (complementarity constraint)} \end{array} \right. \right\}. \quad (1)$$

The above BBC constrained POP (1) has been widely studied as they have many applications in combinatorial optimization, signal processing [11, 25], transportation engineering [1], and optimal power flow [12, 26].

BBCPOP provides a MATLAB implementation to automatically generate a hierarchy of sparse doubly nonnegative (DNN) relaxations of POP (1) together with the BP (bisection and projection) method as a solver for the resulting DNN relaxations problems. This software is developed to find approximate optimal values of larger-scale POPs which are beyond the range that can be comfortably handled by existing software packages. More precisely, an approximate optimal value provided by BBCPOP for a POP is a valid lower bound for the actual optimal value of the POP that is generally NP-hard to compute. The hierarchy of sparse DNN relaxations implemented can be regarded as a variant of the hierarchies of sparse SDP relaxations considered in [38]. The BP method was first introduced by Kim, Kojima and Toh in [20] for the dense doubly nonnegative (DNN) relaxation of a class of QOPs such as binary quadratic problems, maximum clique problems, quadratic multi-knapsack problems, and quadratic assignment problems, and improved later in [4]. In their subsequent work [19], the BP method was generalized to handle the hierarchy of sparse DNN relaxations of a class of binary and box constrained POPs. Some numerical results on large-scale binary and box constrained POPs and the comparison of the BP method to SparsePOP [38] combined with SDPNAL+ [42] were reported in [19]. In this paper, we present an extended version of the BP method for a hierarchy of sparse DNN relaxations of a class of BBC constrained POPs.

Existing software packages available for solving general POPs include GloptiPoly [13], SOSTOOLS [32] and SparsePOP [38]. Notable numerical methods for POPs that have not been announced as software packages include (i) the application of the package SDPNAL [43] for solving SDP relaxations of POPs in [30]; (ii) the application of the solver SDPT3 to solve the bounded degree sums of squares (BSOS) SDP relaxations in [41] and the sparse BSOS relaxations in [41]. GloptiPoly [13], which is designed for the hierarchy of the dense semidefinite (SDP) relaxations by Lasserre in [22], and SOSTOOLS [32], which implements the SDP relaxation by Parrilo [33], can handle small-scale dense POPs with at most 20-40 variables. By exploiting the structured sparsity in POPs, the hierarchy of sparse SDP relaxations was proposed in [37] and implemented as SparsePOP [38]. It was shown that SparsePOP could solve medium-scale general POPs of degree up to 4, and unconstrained POPs with 5000 variables [39] in less than a minute if the sparsity in POPs can be characterized as a banded sparsity pattern such as in the minimization of the chained wood and

chained singular functions. More recently, BSOS [41] and its sparse version of BSOS [41] based on a bounded sums of squares of polynomial have been introduced, and it was shown that sparse BSOS [41] could solve POPs with up to 1000 variables for the same examples. We should note that solving large-scale unconstrained minimization of such functions is much easier than solving constrained POPs. In [30], it was demonstrated that SDPNAL [43] can solve POPs of degree up to 6 and 100 variables could be solved by assuming block diagonal structure sparsity.

Despite efforts to solve large-scale POPs by proposing new theoretical frameworks and various numerical techniques, large-scale POPs still remain very challenging to solve. This difficulty arises from solving large-scale SDP relaxations by SDP solvers, for instance, SeDuMi [35], SDPA [9], SDPT3 [36], and SDPNAL+ [42]. SDP solvers based on primal-dual interior-point algorithms such as SeDuMi, SDPA, SDPT3 have limitation in solving dense SDP relaxations where the size of the variable matrix is at most several thousands. As the size of SDP relaxations in the hierarchy of SDP relaxations for POPs grows exponentially with a parameter called the relaxation order determined by the degree of POPs, it is impossible to solve POPs with 20-30 variables using the SDP solvers based on primal-dual interior-point algorithms unless some special features of POPs such as sparsity or symmetry are utilized. Recently announced SDPNAL+ [42] employs a majorized semismooth Newton-CG augmented Lagrangian method and has illustrated its superior performance of solving large-scale SDPs.

SDPNAL+, however, tends to exhibit some numerical difficulties when handling degenerate SDP problems, in particular problems with many equality constraints. In these cases, it often cannot solve the degenerate SDP problems accurately and the lower bounds computed are even invalid. It is frequently observed in numerical computation that the SDP problems in the hierarchy of dense or sparse SDP relaxations generally become more degenerate as the relaxation order is increased to obtain tight lower bounds for the optimal value of POPs. As shown in [19], the degeneracy also increases as the number of variables and the degree of constrained POPs become larger. Moreover, the SDP relaxation of a high-degree POP can be degenerate even with the first relaxation order in many cases. When a POP from applications can be represented in several different formulations, the degeneracy of each formulation may differ. In particular, it was discussed in [16] that different formulations of equivalent conic relaxations of a combinatorial quadratic optimization problem (QOP) can significantly affect the degeneracy. It was also shown through the numerical results that SDPNAL+ worked efficiently on some conic relaxation formulations of a QOP but its performance on some other formulations was not satisfactory because of the degeneracy. Because of the limitation of SDPNAL+, the BP method in [20, 4] was specifically designed to handle potentially degenerate SDP relaxations problems. The BP method was demonstrated to be robust against the degeneracy in [19], while applying SDPNAL+ to such degenerate cases often leads to invalid bounds and slow convergence. Thus, it is essential to have a solver that can deal with the degeneracy of SDP relaxations for computing valid lower bounds of large-scale or high-degree POPs.

The robustness of BBCPOP is guaranteed by the results in [4] where the lower bounds obtained by the BP method is shown to be always valid. In addition, BBCPOP can effectively handle degenerate DNN relaxations of large-scale and/or high-degree POPs. We show through numerical experiment that BBCPOP can efficiently and robustly compute valid lower bounds for the optimal values of large-scale sparse POPs with BBC constraints

in comparison to SDPNAL+. The test instances whose valid lower bounds could be obtained successfully by BBCPOP in 2000 seconds include, for example, a degree 3 binary POP with complementarity constraints in 1444 variables and a degree 8 box constrained POP in 120 variables. For these instances, SDPNAL+ did not provide a comparable bound within 20000 seconds.

A distinctive feature of our package BBCPOP is that it not only automatically generate DNN relaxations for a BBC constrained POP but also integrate their computations with the robust BP algorithm that is designed specifically for solving them. Other available software packages and numerical methods for POPs such as GlotiPoly [13], SOSTOOLS [32], SparsePOP [38], BSOS [24], and SBSOS [41] in fact first generate the SDP relaxations for the underlying POPs, and then rely on existing SDP solvers such as SDPT3 or SeDuMi to solve the resulting SDP problems. As a result, their performance are heavily dependent on the SDP solver chosen.

This paper is organized as follows: In Section 2, we describe a simple COP (2) to which our DNN relaxation of POP (1) is reduced. We also briefly explain the accelerated proximal gradient method and the BP method for solving the COP. In Section 3, we first describe how to exploit the sparsity in the POP (1) and then derive a simple COP of the form (2) to serve as the sparse DNN relaxations of (1). In Section 4, we present computational aspects of BBCPOP and issues related to its efficient implementation. Section 5 contains numerical results on various BBC constrained POPs. Finally, we conclude in Section 6.

2 Preliminaries

We describe a simple COP (2) in Section 2.1, the accelerated proximal gradient method [5] in Section 2.2 and the bisection and projection (BP) method [20, 4] in Section 2.3. These two methods are designed to solve COP (2) and implemented in BBCPOP. In Section 3, we will reduce a DNN relaxation of POP (1) to the simple COP (2).

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 $\|\mathbf{X}\| = (\langle \mathbf{X}, \mathbf{X} \rangle)^{1/2}$ for every $\mathbf{X} \in \mathbb{V}$. Let \mathbb{K}_1 and \mathbb{K}_2 be closed convex cones in \mathbb{V} satisfying $(\mathbb{K}_1 \cap \mathbb{K}_2)^* = (\mathbb{K}_1)^* + (\mathbb{K}_2)^*$, where $\mathbb{K}^* = \{\mathbf{Y} \in \mathbb{V} : \langle \mathbf{X}, \mathbf{Y} \rangle \geq 0 \text{ for all } \mathbf{X} \in \mathbb{K}\}$ denotes the dual cone of a cone $\mathbb{K} \subset \mathbb{V}$. Let \mathbb{R}^n be the space of n -dimensional column vectors, \mathbb{R}_+^n the nonnegative orthant of \mathbb{R}^n , \mathbb{S}^n the space of $n \times n$ symmetric matrices, \mathbb{S}_+^n the cone of $n \times n$ symmetric positive semidefinite matrices, and \mathbb{N}^n the cone of $n \times n$ symmetric nonnegative matrices.

2.1 A simple conic optimization problem

Let $\mathbf{Q}^0 \in \mathbb{V}$ and $\mathbf{O} \neq \mathbf{H}^0 \in \mathbb{K}_1^* + \mathbb{K}_2^*$. We introduce the following conic optimization problem (COP):

$$\eta^* = \min_{\mathbf{Z}} \{ \langle \mathbf{Q}_0, \mathbf{Z} \rangle \mid \langle \mathbf{H}_0, \mathbf{Z} \rangle = 1, \mathbf{Z} \in \mathbb{K}_1 \cap \mathbb{K}_2 \}. \quad (2)$$

If we take $\mathbb{V} = \mathbb{S}^m$, $\mathbb{K}_1 = \mathbb{S}_+^m$, and \mathbb{K}_2 a polyhedral cone in \mathbb{S}^m for some m , respectively, then the problem (2) represents a general SDP. If in addition $\mathbb{K}_2 \subset \mathbb{N}^m$, then it forms a DNN

optimization problem. Let $\mathbf{G}(y_0) = \mathbf{Q}_0 - y_0\mathbf{H}_0$. The dual of (2) can be described as

$$y_0^* = \max_{y_0, \mathbf{Y}_2} \{y_0 \mid \mathbf{Q}_0 - y_0\mathbf{H}_0 - \mathbf{Y}_2 \in \mathbb{K}_1^*, \mathbf{Y}_2 \in \mathbb{K}_2^*\} \quad (3)$$

$$= \max_{y_0} \{y_0 \mid \mathbf{G}(y_0) \in \mathbb{K}_1^* + \mathbb{K}_2^*\}. \quad (4)$$

As shown in [2, Lemma 2.3], strong duality holds for (2) and (4), *i.e.*, $\eta^* = y_0^*$.

Since $\mathbf{H}_0 \in \mathbb{K}_1^* + \mathbb{K}_2^*$ in (4), we have the following inequality from [20]:

$$y_0 \leq y_0^* \text{ if and only if } \mathbf{G}(y_0) \in \mathbb{K}_1^* + \mathbb{K}_2^*. \quad (5)$$

Therefore, the approximate value of y_0^* can be computed by using the bisection method if the feasibility of any given y_0 , *i.e.*, whether $\mathbf{G}(y_0) \in \mathbb{K}_1^* + \mathbb{K}_2^*$, can be determined. The recently proposed bisection and projection (BP) method [20] (BP Algorithm described in Section 2.3) provides precisely the feasibility test for any given y_0 through a numerical algorithm (APG Algorithm described in Section 2.2) that is based on the accelerated proximal gradient method, where we employed the version in [5] that is modified from [29].

2.2 The accelerated proximal gradient algorithm for feasibility test

For an arbitrary fixed y_0 , let $\mathbf{G} = \mathbf{G}(y_0)$ for simplicity of notation. We also use the notation $\Pi_{\mathbb{K}}(\mathbf{Z})$ to denote the metric projection of $\mathbf{Z} \in V$ onto a closed convex cone $\mathbb{K} \subset V$. Then the problem of testing whether $\mathbf{G} \in \mathbb{K}_1^* + \mathbb{K}_2^*$ leads to the following problem:

$$f^* = \min_{\mathbf{Y}} \left\{ \frac{1}{2} \|\mathbf{G} - \mathbf{Y}\|^2 \mid \mathbf{Y} \in \mathbb{K}_1^* + \mathbb{K}_2^* \right\} \quad (6)$$

$$= \min_{\mathbf{Y}_1, \mathbf{Y}_2} \left\{ \frac{1}{2} \|\mathbf{G} - (\mathbf{Y}_1 + \mathbf{Y}_2)\|^2 \mid \mathbf{Y}_1 \in \mathbb{K}_1^*, \mathbf{Y}_2 \in \mathbb{K}_2^* \right\}$$

$$= \min_{\mathbf{Y}_1} \left\{ f(\mathbf{Y}_1) := \frac{1}{2} \|\Pi_{\mathbb{K}_2}(\mathbf{Y}_1 - \mathbf{G})\|^2 \mid \mathbf{Y}_1 \in \mathbb{K}_1^* \right\}. \quad (7)$$

Here we note that

$$\begin{aligned} \min_{\mathbf{Y}_2} \left\{ \frac{1}{2} \|\mathbf{G} - (\mathbf{Y}_1 + \mathbf{Y}_2)\|^2 \mid \mathbf{Y}_2 \in \mathbb{K}_2^* \right\} &= \frac{1}{2} \|\mathbf{G} - \mathbf{Y}_1 - \Pi_{\mathbb{K}_2^*}(\mathbf{G} - \mathbf{Y}_1)\|^2 \\ &= \frac{1}{2} \|\Pi_{\mathbb{K}_2}(\mathbf{Y}_1 - \mathbf{G})\|^2, \end{aligned}$$

where the last equality above follows from Moreau's decomposition theorem [28, 8]. Obviously, $f^* \geq 0$, and $f^* = 0$ if and only if $\mathbf{G} \in \mathbb{K}_1^* + \mathbb{K}_2^*$.

The gradient of the objective function $f(\mathbf{Y}_1)$ of (7) is given by $\nabla f(\mathbf{Y}_1) = \Pi_{\mathbb{K}_2}(\mathbf{Y}_1 - \mathbf{G})$. As the projection operator $\Pi_{\mathbb{K}}$ onto a convex set \mathbb{K} is nonexpansive [6, Proposition 2.2.1], we have that

$$\|\nabla f(\mathbf{Y}_1) - \nabla f(\mathbf{Y}_1')\| \leq \|(\mathbf{Y}_1 - \mathbf{G}) - (\mathbf{Y}_1' - \mathbf{G})\| = \|\mathbf{Y}_1 - \mathbf{Y}_1'\| \quad (\mathbf{Y}_1, \mathbf{Y}_1' \in \mathbb{V}).$$

Therefore, the gradient $\nabla f(\mathbf{Y}_1)$ is Lipschitz continuous with the Lipschitz constant $L_f = 1$. The KKT condition for $(\mathbf{Y}_1, \mathbf{Y}_2)$ to be the optimal solution of (6) is given by

$$\begin{aligned} \mathbf{X} &= \mathbf{G} - \mathbf{Y}_1 - \mathbf{Y}_2, \quad \langle \mathbf{X}, \mathbf{Y}_1 \rangle = 0, \quad \langle \mathbf{X}, \mathbf{Y}_2 \rangle = 0, \\ \mathbf{X} &\in \mathbb{K}_1 \cap \mathbb{K}_2, \quad \mathbf{Y}_1 \in \mathbb{K}_1^*, \quad \mathbf{Y}_2 \in \mathbb{K}_2^*. \end{aligned}$$

When the KKT condition above holds, we have that $\mathbf{G} \in \mathbb{K}_1^* + \mathbb{K}_2^*$ if and only if $\|\mathbf{X}\| = 0$.

Assume that the metric projections $\Pi_{\mathbb{K}_1}$ and $\Pi_{\mathbb{K}_2}$ onto the cones \mathbb{K}_1 and \mathbb{K}_2 can be computed without difficulty. In [20], APG Algorithm [5] below is applied to (7) to determine whether $f^* = 0$ numerically.

Our APG based algorithm employs the following error criterion

$$g(\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2) = \max \left\{ \frac{\langle \mathbf{X}, \mathbf{Y}_1 \rangle}{1 + \|\mathbf{X}\| + \|\mathbf{Y}_1\|}, \frac{\langle \mathbf{X}, \mathbf{Y}_2 \rangle}{1 + \|\mathbf{X}\| + \|\mathbf{Y}_2\|}, \frac{\Pi_{\mathbb{K}_1^*}(-\mathbf{X})}{1 + \|\mathbf{X}\|}, \frac{\Pi_{\mathbb{K}_2^*}(-\mathbf{X})}{1 + \|\mathbf{X}\|} \right\}$$

to check whether $(\mathbf{Y}_1, \mathbf{Y}_2) = (\mathbf{Y}_1^k, \mathbf{Y}_2^k) \in \mathbb{K}_1^* \times \mathbb{K}_2^*$ satisfies the KKT conditions approximately. It terminates if $\|\mathbf{X}^k\| < \epsilon$ or if $\|\mathbf{X}^k\| \geq \epsilon$ and $g(\mathbf{X}^k, \mathbf{Y}_1^k, \mathbf{Y}_2^k) < \delta$ for sufficiently small positive ϵ and δ , say $\epsilon = 10^{-12}$ and $\delta = 10^{-12}$. Note that $f(\mathbf{Y}_1^k)$ corresponds to $\frac{1}{2}\|\mathbf{X}^k\|^2$. Hence if $\|\mathbf{X}^k\|$ becomes smaller than $\epsilon > 0$, we may regard $(\mathbf{Y}_1^k, \mathbf{Y}_2^k) \in \mathbb{K}_1^* \times \mathbb{K}_2^*$ is an approximate optimal solution of (6) and $\mathbf{G} = \mathbf{G}(y_0) \in (\mathbb{K}_1^* + \mathbb{K}_2^*)$, which implies that y_0 is a feasible solution of the problem (4) and $y_0 \leq y_0^*$. On the other hand, if $\|\mathbf{X}^k\| \geq \epsilon$ and $g(\mathbf{X}^k, \mathbf{Y}_1^k, \mathbf{Y}_2^k) < \delta$, then the KKT optimality condition is almost satisfied, and we classify that $\mathbf{G} = \mathbf{G}(y_0)$ does not lie in $\mathbb{K}_1^* + \mathbb{K}_2^*$ according to $\|\mathbf{X}^k\| \geq \epsilon$. In the latter case, we determine that y_0 is not a feasible solution of (4) and $y_0 > y_0^*$.

APG Algorithm (the accelerated proximal gradient algorithm [5] for feasibility test)

Input: $\mathbf{G} \in \mathbb{V}$, $\mathbf{Y}_1^0 \in \mathbb{V}$, $\Pi_{\mathbb{K}_1}$, $\Pi_{\mathbb{K}_2}$, $\epsilon > 0$, $\delta > 0$, $k_{max} > 0$,

Output: $(\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2) = (\mathbf{X}^k, \mathbf{Y}_1^k, \mathbf{Y}_2^k)$

Initialize: $t_1 \leftarrow 1$, $L \leftarrow L_f (= 1)$, $\bar{\mathbf{Y}}_1^1 \leftarrow \mathbf{Y}_1^0$

for $k = 1, \dots, k_{max}$ **do**

$\mathbf{Y}_1^k \leftarrow \Pi_{\mathbb{K}_1^*} \left(\bar{\mathbf{Y}}_1^k - \frac{1}{L} \Pi_{\mathbb{K}_2}(\bar{\mathbf{Y}}_1^k - \mathbf{G}) \right)$

$\mathbf{Y}_2^{k+1} \leftarrow \Pi_{\mathbb{K}_2^*}(\mathbf{G} - \mathbf{Y}_1^k)$, $\mathbf{X}^k \leftarrow \mathbf{G} - \mathbf{Y}_1^k - \mathbf{Y}_2^k$

if $\|\mathbf{X}^k\| < \epsilon$ (implying $\mathbf{G} \in \mathbb{K}_1^* + \mathbb{K}_2^*$) **or** **if** $\|\mathbf{X}^k\| \geq \epsilon$ and $g(\mathbf{X}^k, \mathbf{Y}_1^k, \mathbf{Y}_2^k) < \delta$ (implying $\mathbf{G} \notin \mathbb{K}_1^* + \mathbb{K}_2^*$) **then**

break

end if

$t_{k+1} \leftarrow \frac{1 + \sqrt{1 + 4t_k^2}}{2}$

$\bar{\mathbf{Y}}_1^{k+1} \leftarrow \mathbf{Y}_1^k + \frac{(t_k) - 1}{t_{k+1}} (\mathbf{Y}_1^k - \mathbf{Y}_1^{k-1})$

end for

We note that the sublinear convergence of APG Algorithm, in the sense that $f(\mathbf{Y}_1^k) - f^* \leq O(1/k^2)$, is ensured for any optimal solution \mathbf{Y}_1^* of (7) in [5, Theorem 4.4].

2.3 The bisection and projection algorithm for COP

As numerically small numbers $\epsilon > 0$ and $\delta > 0$ must be used in APG Algorithm to decide whether $\|\mathbf{X}^k\|$ is equal to 0 on a finite precision floating-point arithmetic machine, an infeasible y_0 can sometimes be erroneously determined as a feasible solution of the problem (4) by APG Algorithm. Likewise, there is also a small possibility for a feasible solution to

be wrongly declared as infeasible due to numerical error. As a result, the feasibility test based on APG Algorithm may not be always correct.

To address the validity issue of the result obtained from APG Algorithm, an improved BP method was introduced in [4]. Here, a valid lower bound $y_0^{v\ell}$ for the optimal value y_0^* is always generated by the improved BP method, which assumes the following two conditions for a given interior point \mathbf{I} of \mathbb{K}_1^* :

(A1) There exists a known positive number $\rho > 0$ such that $\langle \mathbf{I}, \mathbf{Z} \rangle \leq \rho$ for every feasible solution \mathbf{Z} of (2).

(A2) For each $\mathbf{Z} \in \mathbb{V}$, $\lambda_{\min}(\mathbf{Z}) = \sup\{\lambda \mid \mathbf{Z} - \lambda\mathbf{I} \in \mathbb{K}_1^*\}$ can be computed accurately at a moderate cost.

If $\mathbb{K}_1 = \mathbb{K}_1^* = \mathbb{S}_+^n$ and \mathbf{I} is the identity matrix, then $\langle \mathbf{I}, \mathbf{Z} \rangle$ is the trace of \mathbf{Z} and $\lambda_{\min}(\mathbf{Z})$ is the minimum eigenvalue of \mathbf{Z} . Under the assumption (A1), the problem (2) is equivalent to

$$\min_{\mathbf{Z}} \left\{ \langle \mathbf{Q}_0, \mathbf{Z} \rangle \mid \langle \mathbf{H}_0, \mathbf{Z} \rangle = 1, \langle \mathbf{I}, \mathbf{Z} \rangle \leq \rho, \mathbf{Z} \in \mathbb{K}_1 \cap \mathbb{K}_2 \right\}. \quad (8)$$

Its dual

$$\sup_{y_0, \mathbf{Y}_2, \mu} \{y_0 + \rho\mu \mid \mathbf{G}(y_0) - \mathbf{Y}_2 - \mu\mathbf{I} \in \mathbb{K}_1^*, \mathbf{Y}_2 \in \mathbb{K}_2^*, \mu \leq 0\} \quad (9)$$

is equivalent to (4) with the optimal value y_0^* . Suppose that $\bar{y}_0 \in \mathbb{R}$ and $\bar{\mathbf{Y}}_2 \in \mathbb{K}_2^*$ are given. Let $\bar{\mu} = \min\{0, \lambda_{\min}(\mathbf{G}(\bar{y}_0) - \bar{\mathbf{Y}}_2)\}$. Then $(y_0, \mathbf{Y}_2, \mu) = (\bar{y}_0, \bar{\mathbf{Y}}_2, \bar{\mu})$ is a feasible solution of (9), and $y_0^{v\ell} = \bar{y}_0 + \rho\bar{\mu}$ provides a valid lower bound for y_0^* . If $(\bar{y}_0, \bar{\mathbf{Y}}_2)$ is a feasible solution of (3), then $y_0^{v\ell} = \bar{y}_0$ from $\bar{\mu} = 0$. The improved BP method in [4] is described in BP Algorithm below.

BP Algorithm (The improved bisection-projection algorithm [4])

Input: $y_0^\ell \leq y_0^u$, $tol > 0$, $\rho > 0$, $\epsilon > 0$, $\delta > 0$, η_r , $\Pi_{\mathbb{K}_1}$, $\Pi_{\mathbb{K}_2}$, k_{\max} .

Output: $y_0^{v\ell}$

Initialize: $y_0^m \leftarrow \frac{y_0^\ell + y_0^u}{2}$ if $-\infty < y_0^\ell$; otherwise, $y_0^m \leftarrow y_0^u$. $\hat{\mathbf{Y}}_1 \leftarrow \Pi_{\mathbb{K}_1^*}(\mathbf{G}(y_0^m))$. $y_0^{v\ell} \leftarrow -\infty$.

while $y_0^u - y_0^\ell > tol$ **do**

$\hat{\mathbf{Y}}_1^{\text{init}} \leftarrow \hat{\mathbf{Y}}_1$

$(\hat{\mathbf{X}}, \hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2) \leftarrow$ The output of APGR Algorithm with inputs $(\mathbf{G}(y_0^m), \hat{\mathbf{Y}}_1^{\text{init}}, \Pi_{\mathbb{K}_1}, \Pi_{\mathbb{K}_2}, \epsilon, \delta, k_{\max}, \eta_r)$.

$y_0^{v\ell} \leftarrow \max\{y_0^{v\ell}, y_0^m + \rho \min\{0, \lambda_{\min}(\mathbf{G}(y_0^m) - \hat{\mathbf{Y}}_2)\}\}$

if $\|\hat{\mathbf{X}}\| < \epsilon$ **then**

$y_0^\ell \leftarrow y_0^m$

else

$y_0^u \leftarrow y_0^m$, $y_0^\ell \leftarrow \max\{y_0^\ell, y_0^{v\ell}\}$

end if

$y_0^m \leftarrow (y_0^\ell + y_0^u)/2$

end while

Here APGR Algorithm is an enhanced version of APG Algorithm which is described in Section 4.2. One of the advantages of BP Algorithm above is that it does not require

an initial finite estimation of the lower bound y_0^ℓ . In fact, BP Algorithm implemented in the current version of BBCPOP sets $y_0^m \leftarrow y_0^u$ and $y_0^\ell \leftarrow -\infty$ at the initialization step. For a better upper bound y_0^u for y_0^* , a heuristic method applied to the original POP can be employed. For the value of ρ , the exact theoretical value of ρ can be computed from COP (2) in some cases; see [3] for example. We also present a method to estimate ρ in Section 4.1.

3 Sparse DNN relaxation of POP (1)

Exploiting sparsity is a key technique for solving large-scale SDPs and SDP relaxations of large-scale POPs. For example, see [10, 37, 38]. Throughout this section, we assume that POP (1) satisfies a certain structured sparsity described in Section 3.2, and derive its sparse DNN relaxation (17) of the same form as COP (2). The method to derive the DNN relaxation (17) from POP (1) can be divided into two main steps: lift POP (1) to an equivalent problem (14) with moment matrices in the vector variable \mathbf{x} , and replace the moment matrices by symmetric matrix variables after adding valid inequalities. In the software BBCPOP, the function BBCPOPtoDNN implements the two steps just mentioned. After introducing notation and symbols in Section 3.1, we present how the sparsity in POP (1) is exploited in Section 3.2, and the details of the above two steps in Section 3.3 and Section 3.4, respectively.

3.1 Notation and symbols

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

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 $\mathbf{X} \in \mathbb{S}^{\mathcal{A}}$ is written as $X_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ ($(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{A} \times \mathcal{A}$). The *inner product* of $\mathbf{X}, \mathbf{Y} \in \mathbb{S}^{\mathcal{A}}$ is defined by $\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \sum_{\boldsymbol{\beta} \in \mathcal{A}} X_{\boldsymbol{\alpha}\boldsymbol{\beta}} Y_{\boldsymbol{\alpha}\boldsymbol{\beta}}$, and the *norm* of $\mathbf{X} \in \mathbb{S}^{\mathcal{A}}$ is defined by $\|\mathbf{X}\| = (\langle \mathbf{X}, \mathbf{X} \rangle)^{1/2}$. Assuming that the elements of \mathcal{A} are enumerated in an appropriate order, we denote a $|\mathcal{A}|$ -dimensional column vector of monomials \mathbf{x}^α ($\boldsymbol{\alpha} \in \mathcal{A}$) by $\mathbf{x}^{\mathcal{A}}$, and a $|\mathcal{A}| \times |\mathcal{A}|$ symmetric matrix $(\mathbf{x}^{\mathcal{A}})(\mathbf{x}^{\mathcal{A}})^T$ of monomials $\mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}$ ($(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{A} \times \mathcal{A}$) by $\mathbf{x}^{\mathcal{A} \times \mathcal{A}} \in \mathbb{S}^{\mathcal{A}}$. We call $\mathbf{x}^{\mathcal{A} \times \mathcal{A}}$ a *moment matrix*.

For a 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, $\mathbf{x}^{\mathcal{A} \times \mathcal{A}} \in \mathbb{S}_+^{\mathcal{A}}$ for every $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{x}^{\mathcal{A} \times \mathcal{A}} \in \mathbb{S}_+^{\mathcal{A}} \cap \mathbb{N}^{\mathcal{A}}$ for every $\mathbf{x} \in \mathbb{R}_+^n$.

We denote the feasible region of POP (1) as

$$H = \{\mathbf{x} \in \mathbb{R}^n \mid x_i \in [0, 1] \ (i \in I_{\text{box}}), \quad x_j \in \{0, 1\} \ (j \in I_{\text{bin}}), \quad \mathbf{x}^\gamma = 0 \ (\gamma \in \Gamma)\},$$

where $\Gamma = \bigcup_{C \in \mathcal{C}} \{\boldsymbol{\gamma} \in \{0, 1\}^n \mid \gamma_i = 1 \text{ if } i \in C \text{ and } \gamma_j = 0 \text{ otherwise}\}$. Then, POP (1) is written as follows:

$$\zeta^* = \min_{\mathbf{x} \in \mathbb{R}^n} \{f_0(\mathbf{x}) \mid \mathbf{x} \in H\}. \quad (10)$$

We note that the POPs dealt with in [19] are special cases of (10) where $\Gamma = \emptyset$.

Let $\mathbf{r} : \mathbb{Z}_+^n \rightarrow \mathbb{Z}_+^n$ be defined by

$$(\mathbf{r}(\boldsymbol{\alpha}))_i = \begin{cases} \min\{\alpha_i, 1\} & \text{if } i \in I_{\text{bin}} \\ \alpha_i & \text{otherwise (i.e., } i \in I_{\text{box}}). \end{cases} \quad (11)$$

If $\mathbf{x} \in H$, then $\mathbf{x}^\alpha = \mathbf{x}^{\mathbf{r}(\boldsymbol{\alpha})}$ holds for all $\boldsymbol{\alpha} \in \mathbb{Z}_+^n$. Hence we may replace each monomial \mathbf{x}^α in $f_0(\mathbf{x})$ by $\mathbf{x}^{\mathbf{r}(\boldsymbol{\alpha})}$. Therefore $\text{supp} f_0 = \mathbf{r}(\text{supp} f_0)$ is assumed without loss of generality in the subsequent discussion.

3.2 Exploiting sparsity

Let $\nabla^2 f_0(\mathbf{x})$ denote the Hessian matrix of $f_0(\mathbf{x})$. For POP (10), we introduce *the* $n \times n$ *sparsity pattern matrix* \mathbf{R} whose (i, j) th element is defined by

$$R_{ij} = \begin{cases} 1 & \text{if } i = j \text{ or the } (i, j)\text{th element of } \nabla^2 f_0(\mathbf{x}) \text{ is not identically zero,} \\ 1 & \text{if } i, j \in C \text{ for some } C \in \mathcal{C}, \\ 0 & \text{otherwise.} \end{cases}$$

If $\mathcal{C} = \emptyset$, then \mathbf{R} represents the sparsity pattern of the Hessian matrix $\nabla^2 f_0(\mathbf{x})$.

Next, we choose a family of subsets V^k of $N = \{1, 2, \dots, n\}$ ($k = 1, \dots, \ell$) such that the union of $V^k \times V^k$ ($k = 1, \dots, \ell$) covers the set of indices (i, j) associated with the nonzero elements of \mathbf{R} , i.e.,

$$\{(i, j) \in N \times N \mid R_{ij} = 1\} \subseteq \bigcup_{k=1}^{\ell} V^k \times V^k. \quad (12)$$

Obviously, when the only set $V^1 = N$ is chosen for such a family, we get a dense DNN relaxation of POP (10). When \mathbf{R} is sparse (see Figure 1), *the sparsity pattern graph* $G(N, \mathcal{E})$ with the node set N and the edge set $\mathcal{E} = \{(i, j) \in N \times N : i < j \text{ and } R_{ij} = 1\}$ is utilized to create such a family V^k ($k = 1, \dots, \ell$). More precisely, let $G(N, \overline{\mathcal{E}})$ be a chordal extension of $G(N, \mathcal{E})$, and take the maximal cliques V^k ($k = 1, \dots, \ell$) for the family. The chordal extension and its maximal cliques can be found by using the technique [7] based on the symbolic Cholesky decomposition of the adjacency matrix of $G(N, \mathcal{E})$. We implicitly assume that POP (10) satisfies the structured sparsity which induces small size V^k with $|V^k| = O(1)$ ($k = 1, \dots, \ell$). Figure 1 shows such examples. The family V^k ($k = 1, \dots, \ell$) chosen this way satisfies nice properties; see [7, 10, 37] for more details. Here we only mention that the number ℓ of maximal cliques does not exceed n .

As representative sparsity patterns, we illustrate the following two types:

Arrow type: For given $\ell \geq 2$, $a \geq 2$, $b \in \{0, \dots, a-1\}$, and $c \geq 1$, we set

$$V^k = (\{(k-1)(a-b)\} + \{1, 2, \dots, a\}) \cup (\{(\ell-1)(a-b)+a\} + \{1, 2, \dots, c\}) \quad (k = 1, \dots, \ell)$$

(the left figure of Figure 1).

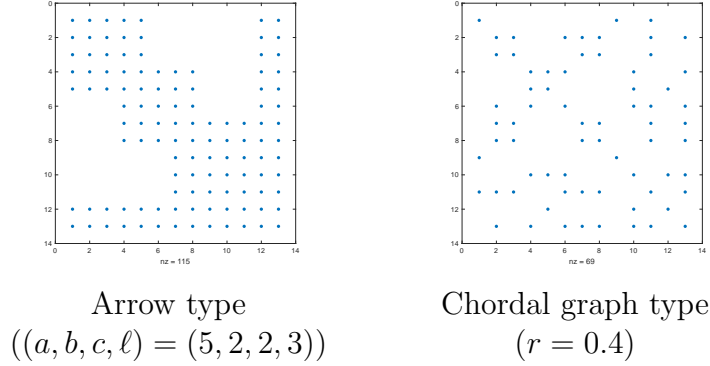


Figure 1: Examples of the sparsity pattern matrix \mathbf{R} with $n = 13$, where the dots correspond to 1 and the blank parts to 0's.

Chordal graph type: Let the number n of variables and the *radio range* $\rho > 0$ be given. For n points $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ drawn from a uniform distribution over the unit square $[0, 1]^2$, we construct the sparsity pattern graph $G(N, \mathcal{E})$ such that $\mathcal{E} = \{(i, j) \in N \times N \mid i < j, \|\mathbf{v}_i - \mathbf{v}_j\| \leq \rho\}$, where $N = \{1, \dots, n\}$. Let V^k ($k = 1, 2, \dots, \ell$) be the maximal cliques in a chordal extension of $G = (N, \mathcal{E})$. (the right figure of Figure 1).

In Section 5, we report numerical results on randomly generated instances of binary and box constrained POPs with these two types of sparsity patterns.

3.3 Lifting POP (10) with moment matrices in \mathbf{x}

Let

$$d = \max \{ \deg(f_0), \deg(\mathbf{x}^\gamma) \ (\gamma \in \Gamma) \}, \quad [d/2] \leq \omega \in \mathbb{Z}_+$$

$$\mathcal{A}_\omega^k = \{ \boldsymbol{\alpha} \in \mathbb{Z}_+^n \mid \alpha_i = 0 \ (i \notin V^k), \sum_{i \in V^k} \alpha_i \leq \omega \} \quad (k = 1, \dots, \ell).$$

Here the parameter ω is named as the *relaxation order*. We then see that

$$\Gamma \subseteq \bigcup_{k=1}^{\ell} (\mathcal{A}_\omega^k + \mathcal{A}_\omega^k) \text{ and } \text{supp}(f_0) \subseteq \bigcup_{k=1}^{\ell} (\mathcal{A}_\omega^k + \mathcal{A}_\omega^k). \quad (13)$$

By the first inclusion relation in (13), each monomial \mathbf{x}^γ ($\gamma \in \Gamma$) is involved in the moment matrix $\mathbf{x}^{\mathcal{A}_\omega^k \times \mathcal{A}_\omega^k}$ for some $k \in \{1, \dots, \ell\}$. By the second inclusion relation in (13), each monomial \mathbf{x}^α of the polynomial $f_0(\mathbf{x})$ is involved in the moment matrix $\mathbf{x}^{\mathcal{A}_\omega^k \times \mathcal{A}_\omega^k}$ for some $k \in \{1, \dots, \ell\}$. Hence the polynomial objective function $f_0(\mathbf{x})$ can be lifted to the space $\mathbb{V} = \mathbb{S}^{\mathcal{A}_\omega^1} \times \dots \times \mathbb{S}^{\mathcal{A}_\omega^\ell}$ such that

$$f_0(\mathbf{x}) = \langle \mathbf{F}_0, (\mathbf{x}^{\mathcal{A}_\omega^1 \times \mathcal{A}_\omega^1}, \dots, \mathbf{x}^{\mathcal{A}_\omega^\ell \times \mathcal{A}_\omega^\ell}) \rangle = \sum_{k=1}^{\ell} \langle \mathbf{F}_0^k, \mathbf{x}^{\mathcal{A}_\omega^k \times \mathcal{A}_\omega^k} \rangle$$

for some $\mathbf{F}_0 = (\mathbf{F}_0^1, \dots, \mathbf{F}_0^\ell) \in \mathbb{V}$, where the inner product $\langle \mathbf{A}, \mathbf{B} \rangle$ for each pair of $\mathbf{A} = (\mathbf{A}^1, \dots, \mathbf{A}^\ell)$ and $\mathbf{B} = (\mathbf{B}^1, \dots, \mathbf{B}^\ell)$ in $\mathbb{V} = \mathbb{S}^{\mathcal{A}_\omega^1} \times \dots \times \mathbb{S}^{\mathcal{A}_\omega^\ell}$ is defined by $\langle \mathbf{A}, \mathbf{B} \rangle =$

$\sum_{k=1}^{\ell} \langle \mathbf{A}^k, \mathbf{B}^k \rangle$. To lift the constraint set $H \subseteq \mathbb{R}^n$ to the space \mathbb{V} , we define

$$\begin{aligned} M &= \{(\mathbf{x}^{\mathcal{A}_\omega^1 \times \mathcal{A}_\omega^1}, \dots, \mathbf{x}^{\mathcal{A}_\omega^\ell \times \mathcal{A}_\omega^\ell}) \in \mathbb{V} \mid \mathbf{x} \in H\} \\ &= \left\{ (\mathbf{x}^{\mathcal{A}_\omega^1 \times \mathcal{A}_\omega^1}, \dots, \mathbf{x}^{\mathcal{A}_\omega^\ell \times \mathcal{A}_\omega^\ell}) \in \mathbb{V} \mid \begin{array}{l} x_i \in [0, 1] \ (i \in I_{\text{box}}), \ x_i \in \{0, 1\} \ (i \in I_{\text{bin}}), \\ \mathbf{x}^\gamma = 0 \ (\gamma \in \Gamma) \end{array} \right\}. \end{aligned}$$

By definition, $\mathbf{x} \in H$ if and only if $(\mathbf{x}^{\mathcal{A}_\omega^1 \times \mathcal{A}_\omega^1}, \dots, \mathbf{x}^{\mathcal{A}_\omega^\ell \times \mathcal{A}_\omega^\ell}) \in M$. Thus we can lift POP (10) to the space \mathbb{V} as follows:

$$\min_{\mathbf{Z}} \{ \langle \mathbf{F}_0, \mathbf{Z} \rangle \mid \mathbf{Z} = (\mathbf{Z}^1, \mathbf{Z}^2, \dots, \mathbf{Z}^\ell) \in M \}. \quad (14)$$

To illustrate the lifting procedure from POP (10) to the space \mathbb{V} , we consider the following example.

Example 3.1. Let us consider the following POP with $n = 3$, $\mathcal{C} = \{\{1, 2\}\}$, $I_{\text{box}} = \{1\}$, and $I_{\text{bin}} = \{2, 3\}$:

$$\min_{\mathbf{x} \in \mathbb{R}^3} \{ f_0(\mathbf{x}) = -x_1x_2 - x_2x_3 \mid x_1x_2 = 0, \ x_1 \in [0, 1], \ x_2, x_3 \in \{0, 1\}. \}. \quad (15)$$

Since $d = \max\{\deg(f_0), \deg(\mathbf{x}^\gamma) \ (\gamma \in \Gamma)\} = 2$, we can take $\omega = 1 \geq \lceil d/2 \rceil$. The sparsity pattern matrix \mathbf{R} turns out to be $\mathbf{R} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$. We show two different choices of V^k ($k = 1, \dots, \ell$).

Dense case: Let $\ell = 1$, $V^k = \{1, 2, 3\}$, and $\omega = 1$. Then $\mathcal{A}_\omega^1 = \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$. We have

$$\mathbf{x}^{\mathcal{A}_\omega^1} = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \text{and} \quad \mathbf{x}^{\mathcal{A}_\omega^1 \times \mathcal{A}_\omega^1} = \mathbf{x}^{\mathcal{A}_\omega^1} (\mathbf{x}^{\mathcal{A}_\omega^1})^T = \begin{pmatrix} 1 & x_1 & x_2 & x_3 \\ x_1 & x_1^2 & x_1x_2 & x_1x_3 \\ x_2 & x_1x_2 & x_2^2 & x_2x_3 \\ x_3 & x_1x_3 & x_2x_3 & x_3^2 \end{pmatrix}.$$

If we define

$$\mathbf{F}_0^1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & -0.5 \\ 0 & -0.5 & 0 & 0 \\ 0 & -0.5 & 0 & 0 \end{pmatrix},$$

then $f_0(\mathbf{x}) = \langle \mathbf{F}_0^1, \mathbf{x}^{\mathcal{A}_\omega^1 \times \mathcal{A}_\omega^1} \rangle$ holds.

Sparse case: Let $\ell = 2$, $V^1 = \{1, 2\}$, $V^2 = \{2, 3\}$, and $\omega = 1$. Then $\mathcal{A}_\omega^1 = \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ and $\mathcal{A}_\omega^2 = \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$. We have

$$\begin{aligned} \mathbf{x}^{\mathcal{A}_\omega^1} &= \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix}, & \mathbf{x}^{\mathcal{A}_\omega^2} &= \begin{pmatrix} 1 \\ x_2 \\ x_3 \end{pmatrix}, \\ \mathbf{x}^{\mathcal{A}_\omega^1 \times \mathcal{A}_\omega^1} &= \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & x_1^2 & x_1x_2 \\ x_2 & x_1x_2 & x_2^2 \end{pmatrix}, & \mathbf{x}^{\mathcal{A}_\omega^2 \times \mathcal{A}_\omega^2} &= \begin{pmatrix} 1 & x_2 & x_3 \\ x_2 & x_2^2 & x_2x_3 \\ x_3 & x_2x_3 & x_3^2 \end{pmatrix}. \end{aligned}$$

If we define

$$\mathbf{F}_0 = (\mathbf{F}_0^1, \mathbf{F}_0^2) = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -0.5 \\ 0 & -0.5 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -0.5 \\ 0 & -0.5 & 0 \end{pmatrix} \right),$$

then $f_0(\mathbf{x}) = \sum_{k=1}^{\ell} \langle \mathbf{F}_0^k, \mathbf{x}^{\mathcal{A}_\omega^k \times \mathcal{A}_\omega^k} \rangle$ holds.

3.4 Valid constraints and conic relaxations of POPs

Note that the objective function of the lifted minimization problem (14) is linear so that it is equivalent to the minimization of the same objective function over the convex hull of the feasible region M of (14). However, the resulting convex minimization problem as well as the original problem (14) are numerically intractable. The second step for deriving a DNN relaxation from the POP (10) is to relax the nonconvex feasible region M to a numerically tractable convex set, which is represented as the intersection of the hyperplane of the form $\{\mathbf{Z} \in \mathbb{V} \mid \langle \mathbf{H}_0, \mathbf{Z} \rangle = 1\}$ and two convex cones \mathbb{K}_1 and \mathbb{K}_2 in \mathbb{V} . Hence we obtain a COP of the form (2), which serves a DNN relaxation of POP (10).

By definition, $(\mathbf{0}, \mathbf{0}) \in \mathcal{A}_\omega^k \times \mathcal{A}_\omega^k$ ($k = 1, \dots, \ell$), which implies that $(\mathbf{x}^{\mathcal{A}_\omega^k \times \mathcal{A}_\omega^k})_{\mathbf{0}\mathbf{0}} = 1$. Thus, if $\mathbf{H}_0 = (\mathbf{H}_0^1, \dots, \mathbf{H}_0^\ell) \in \mathbb{V}$ is defined such that

$$(\mathbf{H}_0^k)_{\alpha\beta} = \begin{cases} 1/\ell & \text{if } \alpha = \beta = \mathbf{0} \\ 0 & \text{otherwise,} \end{cases}$$

then the hyperplane $\{\mathbf{Z} \in \mathbb{V} \mid \langle \mathbf{H}_0, \mathbf{Z} \rangle = 1\}$ will contain M . We also know $M \subseteq \mathbb{S}_+^{\mathcal{A}_\omega^1} \times \dots \times \mathbb{S}_+^{\mathcal{A}_\omega^\ell}$. As a result, we can take $\mathbb{K}_1 = \mathbb{S}_+^{\mathcal{A}_\omega^1} \times \dots \times \mathbb{S}_+^{\mathcal{A}_\omega^\ell}$.

To construct the polyhedral cone \mathbb{K}_2 , we consider the following valid equalities and inequalities for M :

$$\begin{aligned} (\mathbf{x}^{\mathcal{A}_\omega^k \times \mathcal{A}_\omega^k})_{\alpha\beta} &\geq 0, \\ (\mathbf{x}^{\mathcal{A}_\omega^k \times \mathcal{A}_\omega^k})_{\alpha\beta} &= (\mathbf{x}^{\mathcal{A}_\omega^{k'} \times \mathcal{A}_\omega^{k'}})_{\alpha'\beta'} && \text{if } \mathbf{r}(\alpha + \beta) = \mathbf{r}(\alpha' + \beta'), \\ (\mathbf{x}^{\mathcal{A}_\omega^k \times \mathcal{A}_\omega^k})_{\alpha\beta} &\geq (\mathbf{x}^{\mathcal{A}_\omega^{k'} \times \mathcal{A}_\omega^{k'}})_{\alpha'\beta'} && \text{if } \mathbf{r}(c(\alpha + \beta)) = \mathbf{r}(\alpha' + \beta') \text{ for some } c \geq 1, \\ (\mathbf{x}^{\mathcal{A}_\omega^k \times \mathcal{A}_\omega^k})_{\alpha\beta} &= 0 && \text{if } \mathbf{r}(\alpha + \beta) \geq \gamma \text{ for some } \gamma \in \Gamma, \end{aligned} \quad (16)$$

($k, k' \in \{1, \dots, \ell\}$, $\alpha, \beta \in \mathcal{A}_\omega^k$, $\alpha', \beta' \in \mathcal{A}_\omega^{k'}$). In the above, the first inequality follows from the fact that $\mathbf{x} \geq \mathbf{0}$, the second and third inequalities follow from the definition of $\mathbf{r} : \mathbb{Z}_+^n \rightarrow \mathbb{Z}_+^n$ and $\mathbf{x} \in [0, 1]^n$, and the last equality from the complementarity condition $\mathbf{x}^\gamma = 0$ ($\gamma \in \Gamma$). Now, by linearizing the equalities and inequalities above, i.e., replacing $(\mathbf{x}^{\mathcal{A}_\omega^1 \times \mathcal{A}_\omega^1}, \dots, \mathbf{x}^{\mathcal{A}_\omega^\ell \times \mathcal{A}_\omega^\ell})$ by an independent variable $\mathbf{Z} = (\mathbf{Z}^1, \dots, \mathbf{Z}^\ell) \in \mathbb{V}$, we obtain the linear equalities and inequalities to describe the cone \mathbb{K}_2 such that

$$\mathbb{K}_2 = \left\{ \mathbf{Z} = (\mathbf{Z}^1, \dots, \mathbf{Z}^\ell) \in \mathbb{V} \left| \begin{array}{l} \mathbf{Z}_{\alpha\beta}^k \geq 0 \quad (\text{nonnegativity}) \\ \mathbf{Z}_{\alpha\beta}^k = \mathbf{Z}_{\alpha'\beta'}^{k'} \quad \text{if } \mathbf{r}(\alpha + \beta) = \mathbf{r}(\alpha' + \beta'), \\ \mathbf{Z}_{\alpha\beta}^k \geq \mathbf{Z}_{\alpha'\beta'}^{k'} \quad \text{if } \mathbf{r}(c(\alpha + \beta)) = \mathbf{r}(\alpha' + \beta') \\ \quad \text{for some } c \geq 1, \\ \mathbf{Z}_{\alpha\beta}^k = 0 \quad \text{if } \mathbf{r}(\alpha + \beta) \geq \gamma \text{ for some } \gamma \in \Gamma \\ (k, k' \in \{1, \dots, \ell\}, \alpha, \beta \in \mathcal{A}_\omega^k, \alpha', \beta' \in \mathcal{A}_\omega^{k'}) \end{array} \right. \right\}.$$

Consequently, we obtain the following COP:

$$\zeta = \min_{\mathbf{Z}} \{ \langle \mathbf{F}_0, \mathbf{Z} \rangle \mid \langle \mathbf{H}_0, \mathbf{Z} \rangle = 1, \mathbf{Z} \in \mathbb{K}_1 \cap \mathbb{K}_2 \}, \quad (17)$$

which serves as a DNN relaxation of POP (10).

Note that COP (17) is exactly in the form of (2), to which BP Algorithm can be applied. In APG Algorithm, which is called within BP Algorithm, the metric projection $\Pi_{\mathbb{K}_1}(\mathbf{Z})$ of $\mathbf{Z} = (\mathbf{Z}^1, \dots, \mathbf{Z}^\ell) \in \mathbb{V}$ onto $\mathbb{K}_1 = \mathbb{S}_+^{\mathcal{A}_\omega^1} \times \dots \times \mathbb{S}_+^{\mathcal{A}_\omega^\ell}$ can be computed by the eigenvalue decomposition of $\mathbf{Z}^k \in \mathbb{S}^{\mathcal{A}_\omega^k}$ ($k = 1, \dots, \ell$), and the metric projection $\Pi_{\mathbb{K}_2}(\mathbf{Z})$ of $\mathbf{Z} = (\mathbf{Z}^1, \dots, \mathbf{Z}^\ell) \in \mathbb{V}$ onto \mathbb{K}_2 can be computed efficiently by Algorithm 3.3 of [19]. As a result, COP (17) can be solved efficiently by BP Algorithm. Note that the eigenvalue decomposition of $\mathbf{Z}^k \in \mathbb{S}^{\mathcal{A}_\omega^k}$, which requires $O(|\mathcal{A}_\omega^k|^3)$ arithmetic operations at each iteration of APG Algorithm, is the most time consuming part in BP Algorithm applied to COP (17). Therefore, it is crucial for the computational efficiency of BP Algorithm to choose smaller V^k ($k = 1, \dots, \ell$), which determines the size $|\mathcal{A}_\omega^k|$ of \mathcal{A}_ω^k ($k = 1, \dots, \ell$), by exploiting the sparsity of POP (10) as presented in Section 3.2. We also note that the primal-dual interior-point method for COP (17) would remain computationally expensive since \mathbb{K}_2 consists of a large number of inequality such as nonnegative constraints.

Given a POP of the form (10), the sparsity pattern matrix \mathbf{R} and the graph $G(N, \mathcal{E})$ are uniquely determined. However, the choice of the family V^k ($k = 1, \dots, \ell$) satisfying (12) is not unique. Recall that $\mathbf{F}_0, \mathbf{H}_0, \mathbb{K}_1$ and \mathbb{K}_2 in COP (17) depends on the relaxation order ω . Theoretically, the optimal value ζ of COP (17) is monotonically nondecreasing with respect to $\omega \geq \lceil d/2 \rceil$. Thus, a tighter lower bound for the optimal value ζ^* of POP (10) can be expected when a larger ω is used. However, the numerical cost of solving COP (17) by BP Algorithm increases very rapidly as ω increases. We should mention that the family of COP (17) with increasing $\omega \geq \lceil d/2 \rceil$ forms a hierarchy of DNN relaxations for POP (10). Specifically, if $I_{\text{box}} = \emptyset$, our hierarchy may be regarded as a variant of the sparse SDP relaxation proposed in [37, 39] if the SDP is replaced by a stronger DNN relaxation at each hierarchy level, although their formulations may look quite different. Therefore, if $I_{\text{box}} = \emptyset$, the convergence of the optimal value ζ of COP (17) to the optimal value ζ^* of POP (10) is guaranteed [21, 23] as the relaxation order ω increases to infinity.

We also mention that COP (17) can be strengthened by replacing the condition “if $\mathbf{r}(c(\boldsymbol{\alpha} + \boldsymbol{\beta})) = \mathbf{r}(\boldsymbol{\alpha}' + \boldsymbol{\beta}')$ for some $c \geq 1$ ” with the condition “if $\mathbf{r}(\boldsymbol{\alpha} + \boldsymbol{\beta}) \leq \mathbf{r}(\boldsymbol{\alpha}' + \boldsymbol{\beta}')$ ” in (16) and in the description of \mathbb{K}_2 . But we should take note that this replacement considerably increases the number of inequalities in \mathbb{K}_2 and makes the computation of the metric projection $\Pi_{\mathbb{K}_2}(\cdot)$ to be very complicated and expensive. As a result, APG Algorithm (hence BP Algorithm too) is not expected to be efficient for the COP with the strengthened \mathbb{K}_2 ; see [14] for the details.

4 Improving the solution quality

4.1 An upper bound ρ of the trace of the moment matrix

Under the assumptions (A1) and (A2), a valid lower bound $y_0^{v\ell}$ for the optimal value y_0^* of COP (2) is obtained from BP Algorithm. The quality of the valid lower bound $y_0^{v\ell}$ depends

noticeably on the choice of $\rho > 0$ satisfying (A1) where a smaller ρ will lead to a larger $y_0^{v\ell}$. In this section, we discuss the problem of minimizing ρ among the ones that satisfy (A1).

Throughout this section, we assume that COP (2) is constructed for a conic relaxation of POP (10) as described in Section 3. Thus, $y_0^{v\ell}$ serves as a valid lower bound of the POP. In this case, \mathbb{K}_1 turns out to be $\mathbb{S}_+^{A_\omega^1} \times \mathbb{S}_+^{A_\omega^2} \times \dots \times \mathbb{S}_+^{A_\omega^\ell}$. As a result, $\mathbf{I} = (\mathbf{I}_1, \dots, \mathbf{I}_\ell)$ can be taken as an interior point of \mathbb{K}_1 , where \mathbf{I}_k denotes the identity matrix in $\mathbb{S}^{A_\omega^k}$. The problem under consideration is written as

$$\begin{aligned} \hat{\rho} &= \max \{ \langle \mathbf{I}, \mathbf{Z} \rangle \mid \mathbf{Z} \text{ is a feasible solution of COP (2)} \} \\ &= \max \left\{ \sum_{k=1}^{\ell} \langle \mathbf{I}_k, \mathbf{Z}_k \rangle \mid \langle \mathbf{H}, \mathbf{Z} \rangle = 1, \mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_\ell) \in \mathbb{K}_1 \cap \mathbb{K}_2 \right\}. \end{aligned} \quad (18)$$

We may regard the problem (18) as a DNN relaxation of the following POP:

$$\begin{aligned} \rho^* &= \max \left\{ \sum_{k=1}^{\ell} \langle \mathbf{I}_k, \mathbf{x}^{A_\omega^k \times A_\omega^k} \rangle \mid \mathbf{x} \in H \text{ (i.e., a feasible solution of POP (10))} \right\} \\ &= \max \left\{ \sum_{k=1}^{\ell} \sum_{\alpha \in \mathcal{A}_\omega^k} \mathbf{x}^{\alpha+\alpha} \mid \begin{array}{l} x_i \in [0, 1] \text{ (} i \in I_{\text{box}} \text{)}, x_j \in \{0, 1\} \text{ (} j \in I_{\text{bin}} \text{)}, \\ \mathbf{x}^\gamma = 0 \text{ (} \gamma \in \Gamma \text{)} \end{array} \right\}. \end{aligned} \quad (19)$$

This implies that $\hat{\rho} \geq \rho^*$, and if $\rho \geq \rho^*$ and (y_0, \mathbf{Y}_2, μ) is a feasible solution of COP (9), then $y_0 + \rho\mu$ provides a valid lower bound for the optimal value of POP (10). Thus, it is more reasonable to consider (19) directly than its relaxation (18). It is easy to verify that the problem (19) has an optimal solution \mathbf{x} with $x_i \in \{0, 1\}$ for all $i = 1, \dots, n$. In addition, if $\alpha \geq \gamma$ for some $\gamma \in \Gamma$ then $\mathbf{x}^{\alpha+\alpha} = 0$. Hence, (19) is reduced to a combinatorial optimization problem given by

$$\rho^* = \max \left\{ \sum_{k=1}^{\ell} \sum_{\alpha \in \mathcal{B}_\omega^k} \mathbf{x}^{\alpha+\alpha} \mid \mathbf{x} \in \{0, 1\}^n, \mathbf{x}^\gamma = 0 \text{ (} \gamma \in \Gamma \text{)} \right\}, \quad (20)$$

where $\mathcal{B}_\omega^k = \{\alpha \in \mathcal{A}_\omega^k \mid \alpha \not\geq \gamma \text{ for any } \gamma \in \Gamma\}$ ($k = 1, \dots, \ell$). It would be ideal to use $\rho = \rho^*$ in BP Algorithm for a tight lower bound $y_0^{v\ell}$. As the problem (20) is numerically intractable in general, any upper bound ρ for ρ^* can be used in practice. In particular, the trivial upper bound $\rho = \sum_{k=1}^{\ell} |\mathcal{B}_\omega^k|$ may be used for an upper bound of ρ^* , but it may not be tight except for simple cases.

We can further reduce the problem (20) to a submodular function minimization under a set cover constraint for which efficient approximation algorithms [18, 40] exist for a tight lower bound of the its minimum value. For this purpose, the vector variable $\mathbf{x} \in \{0, 1\}^n$ is replaced by a set variable $S_0 \subseteq N = \{1, \dots, n\}$, which determines x_i to be 0 if $i \in S_0$ and 1 otherwise. For every $i \in N$, define

$$E_i = \bigcup_{k=1}^{\ell} \{(k, \alpha) \mid \alpha \in \mathcal{B}_\omega^k, \alpha_i \geq 1\}, \quad F_i = \{\gamma \in \Gamma \mid \gamma_i \geq 1\}.$$

The next Lemma states two properties of the above sets.

Lemma 4.1. Choose $\mathbf{x} \in \{0, 1\}^n$ arbitrarily. Let $S_0 = \{i \in N \mid x_i = 0\}$. Then we have that

$$(i) \quad \bigcup_{i \in S_0} E_i = \bigcup_{k=1}^{\ell} \{(k, \boldsymbol{\alpha}) \mid \boldsymbol{\alpha} \in \mathcal{B}_{\omega}^k, \mathbf{x}^{\boldsymbol{\alpha}} = 0\};$$

$$(ii) \quad \mathbf{x}^{\boldsymbol{\gamma}} = 0 \ (\boldsymbol{\gamma} \in \Gamma) \text{ if and only if } \bigcup_{i \in S_0} F_i = \Gamma \text{ or } \left| \bigcup_{i \in S_0} F_i \right| = |\Gamma|.$$

Proof. (i) Assume that $(k, \boldsymbol{\alpha}) \in E_i$ for some $i \in S_0$. By definition, $(k, \boldsymbol{\alpha}) \in \mathcal{B}_{\omega}^k$ and $\alpha_i \geq 1$. Hence $0 = x_i = x_i^{\alpha_i} = \mathbf{x}^{\boldsymbol{\alpha}}$. Thus we have shown the inclusion $\bigcup_{i \in S_0} E_i \subseteq \bigcup_{k=1}^{\ell} \{(k, \boldsymbol{\alpha}) \mid \boldsymbol{\alpha} \in \mathcal{B}_{\omega}^k, \mathbf{x}^{\boldsymbol{\alpha}} = 0\}$. Now assume that $\boldsymbol{\alpha} \in \mathcal{B}_{\omega}^k$ and $\mathbf{x}^{\boldsymbol{\alpha}} = 0$. It follows from $\mathbf{x}^{\boldsymbol{\alpha}} = 0$ and $\mathbf{x} \in \{0, 1\}^n$ that $\alpha_i \geq 1$ and $x_i = 0$ for some $i \in N$. Hence $i \in S_0$ and $(k, \boldsymbol{\alpha}) \in \bigcup_{i \in S_0} E_i$. Thus we have shown the converse inclusion.

(ii) Assume that $\mathbf{x}^{\boldsymbol{\gamma}} = 0$ ($\boldsymbol{\gamma} \in \Gamma$). The inclusion $\bigcup_{i \in S_0} F_i \subseteq \Gamma$ is straightforward by definition. If $\boldsymbol{\gamma} \in \Gamma$, then it follows from $\mathbf{x}^{\boldsymbol{\gamma}} = 0$ that $\gamma_i \geq 1$ and $x_i = 0$ for some $i \in N$; hence $i \in S_0$ and $\boldsymbol{\gamma} \in F_i$. Thus we have shown the converse inclusion, and the ‘‘only if’’ part of (ii). Now assume that $\bigcup_{i \in S_0} F_i = \Gamma$. Let $\boldsymbol{\gamma} \in \Gamma$. Then there is an $i \in S_0$ such that $\boldsymbol{\gamma} \in F_i$: Hence $x_i = 0$ and $\gamma_i \geq 1$, which implies that $0 = x_i = x_i^{\gamma_i} = \mathbf{x}^{\boldsymbol{\gamma}}$. Thus we have shown the ‘‘if’’ part of (ii). \square

By (ii) of Lemma 4.1, we can rewrite the constraint of the problem (20) as $S_0 \subseteq N$ and $\left| \bigcup_{i \in S_0} F_i \right| = |\Gamma|$, and the objective function as

$$\begin{aligned} \sum_{k=1}^{\ell} \sum_{\boldsymbol{\alpha} \in \mathcal{B}_{\omega}^k} \mathbf{x}^{\boldsymbol{\alpha} + \boldsymbol{\alpha}} &= \sum_{k=1}^{\ell} \sum_{\boldsymbol{\alpha} \in \mathcal{B}_{\omega}^k} \sum_{\mathbf{x}^{\boldsymbol{\alpha}} = 1} 1 \text{ (since } \mathbf{x}^{\boldsymbol{\alpha}} = 0 \text{ or } 1 \text{ for every } \boldsymbol{\alpha} \in \mathcal{B}_{\omega}^k) \\ &= \sum_{k=1}^{\ell} \left(|\mathcal{B}_{\omega}^k| - \sum_{\boldsymbol{\alpha} \in \mathcal{B}_{\omega}^k} \sum_{\mathbf{x}^{\boldsymbol{\alpha}} = 0} 1 \right) \\ &= \sum_{k=1}^{\ell} |\mathcal{B}_{\omega}^k| - \sum_{k=1}^{\ell} \sum_{\boldsymbol{\alpha} \in \mathcal{B}_{\omega}^k} \sum_{\mathbf{x}^{\boldsymbol{\alpha}} = 0} 1 \\ &= \sum_{k=1}^{\ell} |\mathcal{B}_{\omega}^k| - \left| \bigcup_{k=1}^{\ell} \{(k, \boldsymbol{\alpha}) \mid \boldsymbol{\alpha} \in \mathcal{B}_{\omega}^k, \mathbf{x}^{\boldsymbol{\alpha}} = 0\} \right| \\ &= \sum_{k=1}^{\ell} |\mathcal{B}_{\omega}^k| - \left| \bigcup_{i \in S_0} E_i \right| \text{ (by (i) of Lemma 4.1).} \end{aligned}$$

Therefore the problem (20) is equivalent to the problem

$$\min\{c(S_0) \mid S_0 \subseteq N, f(S_0) = |\Gamma|\}, \quad (21)$$

where c and f are *submodular* functions defined by

$$c(S_0) = \left| \bigcup_{i \in S_0} E_i \right| \quad \text{and} \quad f(S_0) = \left| \bigcup_{i \in S_0} F_i \right| \text{ for every } S_0 \subseteq N.$$

This problem is known as a submodular minimization problem under a submodular cover constraint. Approximation algorithms [18, 40] can be used to obtain a lower bound $\bar{c} \geq 0$ for the optimal value of (21). By construction, $\rho = \sum_{k=1}^{\ell} |\mathcal{B}_{\omega}^k| - \bar{c}$ provides an upper bound of (20), which is tighter than or equals to $\sum_{k=1}^{\ell} |\mathcal{B}_{\omega}^k|$. See [14] for the details.

4.2 Enhancing APG Algorithm

Although APG Algorithm has the strong theoretical complexity result such that $f(\mathbf{Y}_2^k) - f^* \leq O(1/k^2)$, in this subsection, we will propose some enhancements to improve its practical performance. We begin by noting that the term $\frac{(t_k)-1}{t_{k+1}}(\mathbf{Y}_1^k - \mathbf{Y}_1^{k-1})$ in the substitution $\bar{\mathbf{Y}}_1^{k+1} \leftarrow \mathbf{Y}_1^k + \frac{(t_k)-1}{t_{k+1}}(\mathbf{Y}_1^k - \mathbf{Y}_1^{k-1})$ can be seen as the momentum of the sequence, and the monotonically increasing sequence $\{\frac{(t_k)-1}{t_{k+1}}\}_{k=0}^\infty \subseteq [0, 1)$ determines the amount of the momentum. When the momentum is high, the sequence $\{\mathbf{Y}_1^k\}_{k=0}^\infty$ would overshoot and oscillate around the optimal solution. In order to avoid such an oscillation to further speed up the convergence, we incorporate the adaptive restarting technique [31] that resets the momentum back to zero ($t_k \leftarrow 1$) and takes a step back to the previous point \mathbf{Y}_1^{k-1} when the objective value increases, i.e., $\|\mathbf{X}^k\| - \|\mathbf{X}^{k-1}\| > 0$. In order to avoid frequent restarts, we also employ the technique of [27, 17] that prohibits the next restart for K_i iterations after the i th restart has occurred, where $K_0 \geq 2$ and $K_i = 2K_{i-1}$ ($i = 1, 2, \dots$). More precisely, we modify APG Algorithm to the following APGR Algorithm.

APGR Algorithm (Accelerated Proximal Gradient algorithm with restarting for feasibility test)

Input: $\mathbf{G} \in \mathcal{X}$, $\mathbf{Y}_1^0 \in \mathcal{X}$, $\Pi_{\mathbb{K}_1}$, $\Pi_{\mathbb{K}_2}$, $\epsilon > 0$, $\delta > 0$, $k_{max} > 0$, $\eta_r > 1$

Output: $(\mathbf{X}^k, \mathbf{Y}_1^k, \mathbf{Y}_2^k)$

Initialize: $t_1 \leftarrow 1$, $L_1 \leftarrow 0.8$, $\bar{\mathbf{Y}}_1^0 \leftarrow \mathbf{Y}_1^0$, $K_0 = 2$, $i = 1$, $k_{re} = 0$

for $k = 1, \dots, k_{max}$ **do**

$\mathbf{Y}_1^k \leftarrow \Pi_{\mathbb{K}_1^*}(\bar{\mathbf{Y}}_1^k - \frac{1}{L_k}\Pi_{\mathbb{K}_2}(\bar{\mathbf{Y}}_1^k - \mathbf{G}))$

$\mathbf{Y}_2^k \leftarrow \Pi_{\mathbb{K}_2^*}(\mathbf{G} - \mathbf{Y}_1^k)$, $\mathbf{X}^k \leftarrow \mathbf{G} - \mathbf{Y}_1^k - \mathbf{Y}_2^k$

if $\|\mathbf{X}^k\| < \epsilon$ or $g(\mathbf{X}^k, \mathbf{Y}_1^k, \mathbf{Y}_2^k) < \delta$ **then**

break

end if

$t_{k+1} \leftarrow \frac{1 + \sqrt{1 + 4t_k^2}}{2}$

$\bar{\mathbf{Y}}_1^{k+1} \leftarrow \mathbf{Y}_1^k + \frac{(t_k)-1}{t_{k+1}}(\mathbf{Y}_1^k - \mathbf{Y}_1^{k-1})$

if $\|\mathbf{X}^k\| - \|\mathbf{X}^{k-1}\| > 0$ and $k > K_i + k_{re}$ **then**

$t^{k+1} \leftarrow 1$, $\bar{\mathbf{Y}}_1^{k+1} \leftarrow \mathbf{Y}_1^k$, $k_{re} \leftarrow k$

$K_{i+1} \leftarrow 2K_i$, $i \leftarrow i + 1$

$L_{k+1} \leftarrow \eta_r L_k$

else

$L_{k+1} \leftarrow L_k$

end if

end for

Recall that $L = 1$ is a Lipschitz constant for the gradient of the function f defined in (7). For APGR Algorithm, the convergence complexity result of $f(\mathbf{Y}_1^k) - f^* \leq O((\log k/k)^2)$ is ensured by [17, 14]. Although it is slightly worse than the original theoretical convergence guarantee of $O(1/k^2)$, it often converges much faster in practice. To improve the practical

performance of APGR Algorithm, the initial estimate L_1 of L in the algorithm is set to a value less than 1. For many instances, $L_1 = 0.8$ provided good results.

Even with the aforementioned practical improvements, APGR Algorithm can still take a long time to compute a very accurate solution of the problem (6) for the purpose of deciding whether $f^* = 0$. If there exists sufficient evidence to show that the optimum value f^* is not likely to be 0, then APGR Algorithm can be terminated earlier to save computation time. To be precise, next we discuss the stopping criteria of APGR Algorithm. Let g^k denote the violation $g(\mathbf{X}^k, \mathbf{Y}_1^k, \mathbf{Y}_2^k)$ of the KKT condition. Assume that g^k is sufficiently small, i.e., the solution is nearly optimal. Then the ratio $\|\mathbf{X}^k\|/g^k$ of the optimal value and the KKT violation is a reasonable measure to indicate how far away f^* is from 0. To determine that the value $\|\mathbf{X}^k\|/g^k$ will not be improved much, the following values are computed and tested at every k th iteration ($k > 30$):

- $M_g^k =$ Geometric mean of $\{g^{k-i}/g^{k-i-30} \mid i = 0, 1, \dots, 9\}$.
- $M_X^k =$ Geometric mean of $\{\|\mathbf{X}^{k-i}\|/\|\mathbf{X}^{k-i-30}\| \mid i = 0, 1, \dots, 9\}$.

If the above values are close to 1, much improvement in the KKT residual and objective values (and $\|\mathbf{X}^k\|/g^k$) cannot be expected. Based on these observations, we implemented several heuristic stopping criteria in our practical implementation of APGR Algorithm in the MATLAB function called `projK1K2.fista.dualK.m`. For example, we have found the following stopping criterion

$$\|\mathbf{X}^k\|/g^k \geq 10^4, \quad g^k \leq \sqrt{\delta}, \quad M_g^k \geq 0.95, \quad M_X^k \geq 0.995$$

is useful to reduce the computational time substantially without losing the quality of our computed valid lower bound y_0^{vl} from BP Algorithm.

5 Numerical experiments

We present numerical results to compare the performance of BBCPOP with SDPNAL+ [42], a state-of-the-art solver for large scale semidefinite optimization problem with nonnegative constraints. Recall that BBCPOP is a MATLAB implementation of the DNN relaxation (17) of POP (10) based on BP and APGR Algorithms, which have been presented in Sections 2.3 and 4.2, respectively. SDPNAL+ is an augmented Lagrangian based method for which the main subproblem in each iteration is solved by a semismooth Newton method that employs the generalized Hessian of the underlying function. As already mentioned in the Introduction, the solver is able to handle nondegenerate problems efficiently but it is usually not efficient in solving degenerate problems and generally unable to solve such problems to high accuracy. The solver BBCPOP on the other hand, uses only first-order derivative information and is specifically designed to handle degenerate problems arising from the DNN relaxations of BBC constrained POPs. As a result, SDPNAL+ is expected to provide a more accurate solution than BBCPOP when the former is successful in solving the problem to the required accuracy. In Section 2.3, we have presented a method to compute a valid lower bound in BP Algorithm by introducing a primal-dual pair of COPs (8) and (9). Similarly, we also can generate a valid lower bound from a solution of SDPNAL+. For

fair comparison, we also computed valid lower bounds based on the approximate solutions generated by SDPNAL+ for all experiments.

For BP Algorithm incorporated in BBCPOP, the parameters $(tol, \gamma, \epsilon, \delta, k_{\max}, \eta_r)$ were set to $(10^{-5}, 10, 10^{-13}, 10^{-6}, 20000, 1.1)$. For SDPNAL+, the parameter “tol” was set to 10^{-6} and “stopoptions” to 2 so that the solver continues to run even if it encounters some stagnations. SDPNAL+ was terminated in 20000 iterations even if the stopping criteria were not satisfied. All the computations were performed in MATLAB on a Mac Pro with Intel Xeon E5 CPU (2.7 GHZ) and 64 GB memory.

In Tables 1 through 4, the meaning of the notation are as follows: “opt” (the optimal value), “LBv” (a valid lower bound), “sec” (the computation time in second), “apgit” (the total number of iterations in APG Algorithm), “bpit” (the number of iterations in BP Algorithm), “iter” (the number of iterations in SDPNAL+), and “term” (the termination code). The termination code of SDPNAL+ has the following meaning: 0 (problem is solved to required tolerance), -1,-2,-3 (problem is partially solved with the primal feasibility, the dual feasibility, both feasibility slightly violating the required tolerance, respectively), 1 (problem is not solved successfully due to stagnation), 2 (the maximum number of iterations reached). For BBCPOP, its termination code has the following meaning: 1,2 (problem is solved to required tolerance), 3 (the iteration is stopped due to minor improvements in valid lower bounds).

5.1 Randomly generated sparse polynomial optimization problems with binary, box and complementarity constraints

In this section, we present experimental results on randomly generated sparse POPs of the form (10). The objective function $f_0(\mathbf{x}) = \sum_{\alpha \in \mathcal{F}} c_{\alpha} \mathbf{x}^{\alpha}$ was generated as follows. Its degree was fixed to 2, 3, 4, 5, 6, and 8. For the support $\text{supp} f = \mathcal{F}$, we took

$$\mathcal{F} = \bigcup_{k=1}^{\ell} \left\{ \alpha \in \mathbb{Z}_+^n \mid \sum_{i=1}^n \alpha_i \leq d, \quad \alpha_i = 0 \text{ if } i \notin V^k \right\},$$

where V^k ($k = 1, \dots, \ell$) were chosen from the two types of graphs, the arrow and chordal graphs, given in Section 3.2; see Figure 1. Each c_{α} was chosen from the uniform distribution over $[-1, 1]$. I_{bin} was set to $N = \{1, 2, \dots, n\}$ (hence $I_{\text{box}} = \emptyset$) in Table 1, and \emptyset (hence $I_{\text{box}} = N$) in Table 2. In each table, \mathcal{C} was set to \emptyset and a set of $2n$ elements randomly chosen from $\bigcup_{k=1}^{\ell} (V^k \times V^k)$, respectively. For all computation, the relaxation order ω was set to $\lceil \frac{d}{2} \rceil$.

Table 1: The computation results of SDPNAL+ and BBCPOP for $I_{\text{bin}} = \{1, 2, \dots, n\}$ and $I_{\text{box}} = \emptyset$.

obj. \ constr.		$C = \emptyset$			C : randomly chosen		
type	d	BBCPOP		SDPNAL+		SDPNAL+	
		LBv (sec,apgit:bpit,term)	LBv (sec,apgit:bpit,term)	LBv (sec,apgit:bpit,term)	LBv (sec,apgit:bpit,term)	LBv (sec,apgit:bpit,term)	LBv (sec,apgit:bpit,term)
Arrow ($a = 10$, $b = 2$, $c = 2$)	2	1284(160)	-9.115078e2 (2.31e2, 10517:16, 2)	-9.114681e2 (3.65e3, 5597, 0)	-3.845385e2 (2.11e2, 9583:16, 2)	-3.845080e2 (7.65e3, 20000, -2)	-
	2	1924(240)	-1.365696e3 (6.71e2, 20249:16, 2)	-1.365674e3 (6.67e3, 7740, 0)	-5.620671e2 (3.69e2, 12016:17, 2)	-5.620345e2 (1.39e4, 20000, -2)	-
	2	2564(320)	-1.848387e3 (3.86e2, 9188:17, 2)	-	-7.870155e2 (9.57e2, 23251:18, 2)	-	-
	3	484(60)	-6.612489e2 (1.12e3, 12533:20, 2)	-6.612368e2 (5.62e3, 4269, 0)	-1.843149e2 (6.03e2, 6663:18, 2)	-1.843102e2 (2.29e3, 953, 0)	-
	3	964(120)	-1.387246e3 (2.28e3, 11989:18, 2)	-1.387269e3 (3.28e4, 5230, 0)	-3.781069e2 (1.50e3, 7500:18, 2)	-3.781066e2 (1.38e4, 2333, 0)	-
Arrow ($a = 5$, $b = 2$, $c = 2$)	3	1444(180)	-1.996709e3 (3.56e3, 13913:20, 2)	-	-5.413793e2 (1.87e3, 7410:18, 2)	-	-
	5	52(6)	-1.628047e2 (2.48e3, 1486:21, 2)	-1.730651e2 (7.62e3, 20000, 1)	-1.893068e1 (1.09e3, 6869:18, 2)	-1.892955e1 (2.92e2, 320, 0)	-
	5	100(12)	-3.153501e2 (4.15e3, 12200:19, 2)	-3.224262e2 (1.39e4, 2402, 0)	-4.600732e1 (2.47e3, 7543:18, 2)	-4.600496e1 (2.23e3, 789, 0)	-
	5	148(18)	-4.836598e2 (6.75e3, 14284:20, 2)	-	-6.020487e1 (2.46e3, 5495:19, 2)	-	-
	6	70(22)	-9.487728e1 (1.57e2, 7730:19, 2)	-9.487333e1 (1.70e2, 335, 0)	-1.583319e1 (7.99e1, 4433:17, 2)	-1.583321e1 (2.50e1, 155, 0)	-
Arrow ($a = 5$, $b = 2$, $c = 2$)	6	76(24)	-1.035035e2 (2.25e2, 10062:18, 2)	-1.035003e2 (2.89e2, 967, 0)	-1.639041e1 (9.84e1, 4872:18, 2)	-1.638897e1 (4.46e1, 317, 0)	-
	6	82(26)	-1.112050e2 (1.53e2, 6407:19, 2)	-1.112020e2 (2.99e2, 856, 0)	-1.797953e1 (1.29e2, 5948:17, 2)	-1.797902e1 (2.75e1, 154, 0)	-
	8	64(20)	-8.471209e1 (4.59e2, 10985:18, 2)	-8.471209e1 (7.77e2, 3878, 0)	-1.446632e1 (1.57e2, 4503:17, 2)	-1.446599e1 (3.96e1, 154, 0)	-
	8	70(22)	-9.238185e1 (4.01e2, 8686:19, 2)	-9.238109e1 (6.81e2, 1558, 0)	-1.767580e1 (1.63e2, 4272:17, 2)	-1.767573e1 (1.19e2, 631, 0)	-
	8	76(24)	-9.720811e1 (5.03e2, 10009:19, 2)	-9.720601e1 (7.95e2, 2019, 0)	-1.717366e1 (1.70e2, 3996:17, 2)	-1.717314e1 (4.20e1, 153, 0)	-
Chordal ($r = 0.1$)	2	400	-3.474744e2 (3.60e2, 12180:17, 2)	-3.474432e2 (7.94e3, 2178, 0)	-8.638335e1 (7.34e2, 24016:18, 2)	-8.637568e1 (2.21e4, 20000, -2)	-
	2	800	-1.205713e3 (1.26e3, 13868:18, 2)	-1.205601e3 (2.80e4, 2334, 0)	-1.972467e2 (1.05e3, 11445:19, 2)	-1.972365e2 (3.37e4, 2116, 0)	-
	2	1600	-3.750599e3 (5.89e3, 12502:18, 2)	-	-5.124576e2 (8.89e3, 20542:20, 2)	(, ,)	-
	3	200	-2.065760e2 (3.52e2, 13101:18, 2)	-2.065634e2 (8.55e2, 656, 0)	-7.533718e1 (1.54e2, 5459:19, 2)	-7.533653e1 (1.72e2, 158, 0)	-
	3	300	-4.769478e2 (1.17e3, 11293:18, 2)	-4.769380e2 (6.51e3, 983, 0)	-8.210410e1 (3.81e2, 3617:18, 2)	-8.210191e1 (3.93e2, 154, 0)	-
Chordal ($r = 0.1$)	3	400	-9.067157e2 (1.95e4, 27303:19, 2)	-	-1.082068e2 (4.44e3, 5756:19, 2)	(, ,)	-
	5	100	-1.097916e2 (4.67e1, 6843:17, 2)	-1.097897e2 (2.09e2, 663, 0)	-8.239900e1 (4.48e1, 6314:18, 2)	-8.239926e1 (6.08e1, 486, 0)	-
	5	200	-6.115672e2 (1.00e3, 8164:18, 2)	-6.115882e2 (1.55e4, 1386, 0)	-1.091617e2 (4.84e2, 4269:18, 2)	-1.091602e2 (4.07e2, 156, 0)	-
	5	300	-1.637701e3 (4.56e4, 28839:19, 2)	-	-1.381003e2 (5.80e3, 3855:18, 2)	(, ,)	-
	6	100	-1.229437e2 (3.17e1, 4790:16, 2)	-1.229360e2 (6.72e1, 171, 0)	-1.000038e2 (2.79e1, 4242:17, 2)	-1.000038e2 (2.21e1, 156, 0)	-
Chordal ($r = 0.1$)	6	150	-3.343328e2 (1.41e2, 7499:18, 2)	-3.343190e2 (6.56e2, 983, 0)	-1.224431e2 (7.60e1, 4225:19, 2)	-1.224399e2 (1.46e2, 166, 0)	-
	6	200	-8.586312e2 (2.14e3, 16750:18, 2)	-8.586086e2 (1.26e4, 1151, 0)	-1.151008e2 (4.89e2, 3994:18, 2)	-1.150950e2 (8.88e2, 159, 0)	-
	8	100	-2.384588e2 (2.76e1, 4051:15, 2)	-2.384465e2 (1.85e2, 496, 0)	-1.824317e2 (4.23e1, 6057:17, 2)	-1.824326e2 (3.59e1, 165, 0)	-
	8	150	-5.620558e2 (1.92e2, 8539:18, 2)	-5.620407e2 (8.13e2, 1152, 0)	-1.161660e2 (9.62e1, 4438:17, 2)	-1.161638e2 (8.16e1, 158, 0)	-
	8	170	-9.313252e2 (1.06e3, 13272:19, 2)	-9.313161e2 (9.53e3, 1839, 0)	-1.475300e2 (4.09e2, 5602:17, 2)	-1.475289e2 (3.96e2, 318, 0)	-

Table 2: The computation results of SDPNAL+ and BBCPOP for $I_{\text{bin}} = \emptyset$ and $I_{\text{box}} = \{1, 2, \dots, n\}$.

obj. \ constr.	$C = \emptyset$			$C : \text{randomly chosen}$		
	d	$n(\ell)$	SDPNAL+ LBv (sec,apgit:bpbit,term)	BBCPOP LBv (sec,apgit:bpbit,term)	SDPNAL+ LBv (sec,iter,term)	SDPNAL+ LBv (sec,iter,term)
Arrow ($a = 10$, $b = 2$, $c = 2$)	2	1284(160)	-9.629408e2 (2.37e2, 11521:19, 2)	-9.629396e2 (1.46e3, 2151, 0)	-4.440439e2 (1.82e2, 8714:18, 2)	-4.440437e2 (5.60e3, 6475, 0)
	2	1924(240)	-1.432735e3 (4.83e2, 15027:17, 2)	-1.432743e3 (2.85e3, 2283, 0)	-6.704890e2 (3.08e2, 9502:18, 2)	-6.704828e2 (8.74e3, 5829, 0)
	2	2564(320)	-1.948288e3 (4.51e2, 10379:18, 2)	-1.948220e3 (4.27e3, 3622, 0)	-8.826387e2 (4.66e2, 11886:18, 2)	-8.826803e2 (1.62e4, 20000, 0)
	3	164(20)	-2.815738e2 (9.97e2, 28142:19, 2)	-2.814370e2 (3.57e3, 20000, -1)	-9.442008e1 (4.35e2, 11791:18, 2)	-9.441908e1 (5.52e3, 20000, -3)
	3	324(40)	-5.878139e2 (2.90e3, 40618:19, 2)	-5.879896e2 (1.08e4, 20000, -3)	-1.817788e2 (9.68e2, 13033:18, 2)	-1.817602e2 (1.62e4, 18383, 0)
	3	484(60)	-9.342330e2 (3.95e3, 39757:21, 2)	-	-3.024086e2 (1.71e3, 16224:18, 2)	-
	5	20(2)	-1.541292e2 (4.11e3, 30618:19, 2)	-1.784141e2 (8.18e3, 20000, 1)	-2.023543e1 (1.32e3, 9907:18, 2)	-2.024220e1 (4.36e3, 20000, 1)
	5	36(4)	-2.773921e2 (1.13e4, 43006:18, 2)	-2.841172e2 (2.29e4, 20000, 1)	-4.620292e1 (4.71e3, 17307:19, 2)	-4.622614e1 (7.70e3, 20000, -3)
	5	52(6)	-3.736479e2 (2.08e4, 55131:20, 2)	-	-6.892798e1 (5.02e3, 12720:20, 2)	-
	6	28(8)	-1.543165e2 (5.85e2, 23268:19, 2)	-1.615659e2 (2.48e3, 20000, 1)	-2.575588e1 (6.16e2, 24285:18, 2)	-2.574200e1 (3.02e3, 20000, -3)
	6	34(10)	-1.926331e2 (4.70e2, 15913:18, 2)	-1.931423e2 (5.12e3, 20000, 1)	-2.901087e1 (3.48e2, 11568:17, 2)	-2.902573e1 (3.50e3, 20000, -3)
	6	40(12)	-2.229557e2 (1.14e3, 32413:20, 2)	-2.229243e2 (6.94e3, 20000, 1)	-3.887661e1 (3.39e2, 9359:18, 2)	-3.887465e1 (5.12e3, 20000, -3)
8	16(4)	-2.062974e2 (4.74e3, 42025:22, 2)	-2.064534e2 (1.18e4, 20000, 1)	-7.338345e0 (1.85e3, 16555:19, 2)	-7.335928e0 (3.99e3, 20000, -2)	
8	22(6)	-2.513190e2 (5.41e3, 52243:21, 2)	-2.517759e2 (3.44e4, 20000, 1)	-1.919867e1 (2.57e3, 14396:17, 2)	-1.919410e1 (1.85e4, 20000, -3)	
8	28(8)	-3.020144e2 (7.92e3, 35596:18, 2)	-	-2.725916e1 (1.95e3, 8482:17, 2)	-	
Chordal ($r = 0.1$)	2	400	-3.486683e2 (5.04e2, 17758:19, 2)	-3.486410e2 (1.02e4, 2681, 0)	-8.688482e1 (3.34e2, 11287:19, 2)	-8.687822e1 (1.30e4, 10887, 0)
	2	800	-1.206955e3 (2.66e3, 30865:18, 2)	-1.2066830e3 (2.76e4, 4012, 0)	-1.975751e2 (1.19e3, 13331:18, 2)	-1.975661e2 (1.56e4, 2561, 0)
	2	1200	-2.385826e3 (7.53e3, 35674:18, 2)	-	-3.455533e2 (3.04e3, 14767:19, 2)	-
	3	100	-5.668570e1 (8.91e1, 13436:19, 2)	-5.668556e1 (1.87e3, 20000, -3)	-5.304373e1 (1.04e2, 15051:20, 2)	-5.304241e1 (2.22e3, 17843, 0)
	3	200	-2.124543e2 (1.26e3, 37656:18, 2)	-2.123967e2 (1.46e4, 20000, -3)	-7.678323e1 (4.25e2, 12432:18, 2)	-7.678226e1 (1.66e4, 13853, 0)
	3	400	-9.481387e2 (2.91e4, 30987:20, 2)	-	-1.100166e2 (2.72e4, 29200:19, 2)	-
	5	80	-6.679212e1 (1.58e2, 21489:17, 2)	-6.678636e1 (2.19e3, 20000, -3)	-6.470095e1 (1.54e2, 20670:18, 2)	-6.470377e1 (2.56e3, 20000, -3)
	5	160	-3.162245e2 (1.49e3, 24909:19, 2)	-3.209856e2 (2.29e4, 20000, 1)	-1.200387e2 (1.26e3, 21131:20, 2)	-1.201353e2 (2.46e4, 20000, -3)
	5	240	-1.056989e3 (5.91e4, 64866:19, 2)	-	-1.376493e2 (1.25e4, 14180:18, 2)	-
	6	50	-5.433668e1 (3.36e1, 9673:17, 2)	-5.433165e1 (5.40e2, 12140, 0)	-5.237042e1 (3.71e1, 9883:19, 2)	-5.236817e1 (3.97e2, 11692, 0)
	6	100	-1.314831e2 (2.76e2, 24792:17, 2)	-1.314177e2 (4.17e3, 20000, -3)	-1.065003e2 (2.63e2, 23767:20, 2)	-1.064412e2 (4.73e3, 20000, -3)
	6	120	-2.130796e2 (1.01e3, 52465:19, 2)	-2.133697e2 (1.17e4, 20000, 1)	-1.122865e2 (3.94e2, 19744:19, 2)	-1.122791e2 (7.57e3, 20000, -3)
8	50	-3.542262e1 (1.19e2, 22885:18, 2)	-3.540774e1 (7.59e2, 17063, 0)	-3.305533e1 (1.39e2, 25186:18, 2)	-3.303788e1 (5.14e2, 8524, 0)	
8	100	-2.567746e2 (9.45e2, 31985:18, 2)	-2.565093e2 (8.29e3, 20000, 1)	-1.962108e2 (4.56e2, 15764:18, 2)	-1.961334e2 (8.47e3, 20000, -3)	
8	120	-3.106839e2 (1.88e3, 31566:21, 2)	-5.129922e2 (1.98e4, 20000, 1)	-1.392161e2 (1.20e3, 20163:18, 2)	-1.397772e2 (1.89e4, 20000, 1)	

Tables 1 and 2 show the results on POPs with binary constraints and those with box constraints, respectively. We see that the lower bounds of the POP obtained by BBCPOP are comparable to those by SDPNAL+. In particular, when the arrow type sparsity and $I_{\text{bin}} = \emptyset$ were used as in Table 1, BBCPOP could compute tighter lower bounds than SDPNAL+ in some instances. The computation time of BBCPOP is much smaller than SDPNAL+ in most instances. BBCPOP converged in $10^4 \sim 10^5$ seconds for very large scale POPs for which SDPNAL+ could not in 10^5 seconds. Although BBCPOP uses only the first-order derivative information contrary to SDPNAL+ that utilizes the second-order information, it could compute lower bounds comparable to SDPNAL+ in shorter computation time.

As the number of variables increases for the problems with $d = 2, 3, 5$ in Table 1, SDPNAL+ frequently failed to obtain a lower bound while BBCPOP succeeded to compute a valid lower bound. As mentioned in Section 1, SDPNAL+ could not deal with degenerate POPs well. BBCPOP, on the other hand, provided valid lower bounds for those POPs. Similar observation can be made for Table 2 where no lower bounds and computational time were reported for SDPNAL+ in some problems. The numerical results in Tables 1 and 2 demonstrate the robustness of BBCPOP against degeneracy. We see from the results that the performance of BBCPOP is very efficient and robust for solving the DNN relaxations of large scale POPs with BBC constraints.

For the ease of comparison, in Figures 2 and 3, we plot the relative gap of the valid lower bounds computed by SDPNAL+ and BBCPOP i.e., $\frac{\text{LBv}(\text{BBCPOP}) - \text{LBv}(\text{SDPNAL+})}{|\text{LBv}(\text{BBCPOP})|}$, and the ratios of the execution times taken by SDPNAL+ to those by BBCPOP for the instances tested in Tables 1 and 2, respectively. Note that in the plots, a positive value for the relative gap of the lower bounds means that BBCPOP has computed a tighter bound than SDPNAL+. We can observe that for the instances corresponding to $\mathcal{C} = \emptyset$ in Table 1 (Figure 2), BBCPOP is a few times faster than SDPNAL+ in solving most of the instances, and it can be 10-20 times faster on a few of the large instances. On the other hand, for the instances corresponding to a randomly chosen \mathcal{C} in Table 1 (Figure 2), BBCPOP and SDPNAL+ have comparable efficiency in solving most of the instances. But for a few smaller instances, SDPNAL+ is 2–5 times faster while for some other larger instances, BBCPOP is at least 10-20 times faster.

For the instances in Table 2 (Figure 3), the efficiency of BBCPOP completely dominates that of SDPNAL+. Here, except for a few instances, BBCPOP is at least 5–20 times faster than SDPNAL+ while the valid lower bounds generated by BBCPOP are comparable or much tighter than those generated by SDPNAL+.

5.2 Quadratic Assignment Problem

Various important combinatorial optimization problems such as the max-cut problem, the maximum stable set problem, and the quadratic assignment problem (QAP) are often formulated as quadratic optimization problems. In this section, we show the numerical performances of BBCPOP and SDPNAL+ in solving DNN relaxation problems of QAPs. We refer the readers to [20, 3] for results on the max-cut and the maximum stable set problems.

Let \mathbf{A} and \mathbf{B} be given $r \times r$ matrices. Then the QAP is described as

$$\zeta_{\text{QAP}}^* = \min \{ \langle \mathbf{X}, (\mathbf{A}\mathbf{X}\mathbf{B}^T) \rangle \mid \mathbf{X} \text{ is a permutation matrix} \}. \quad (22)$$

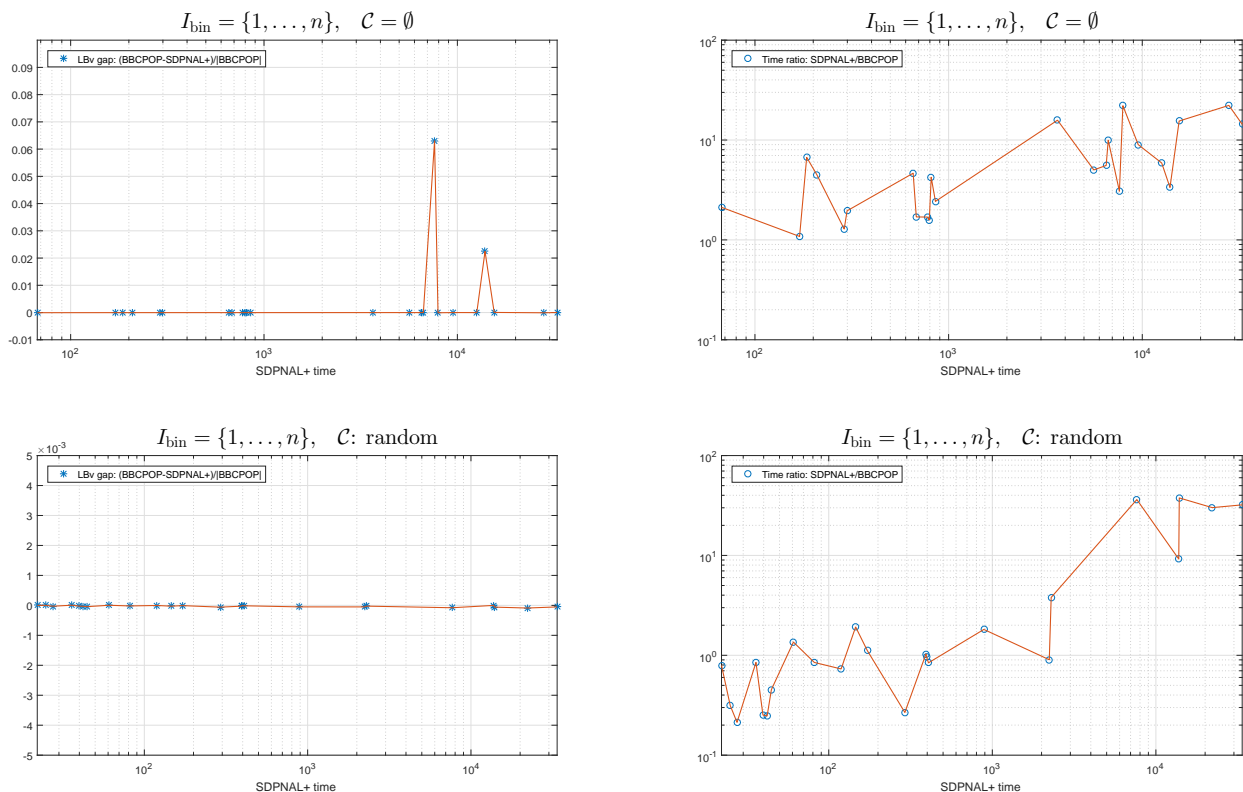


Figure 2: Comparison of the computational efficiency between BBCPOPOP and SDPNAL+ for the instances (with $I_{\text{bin}} = \{1, \dots, n\}$ and $I_{\text{box}} = \emptyset$) in Table 1.

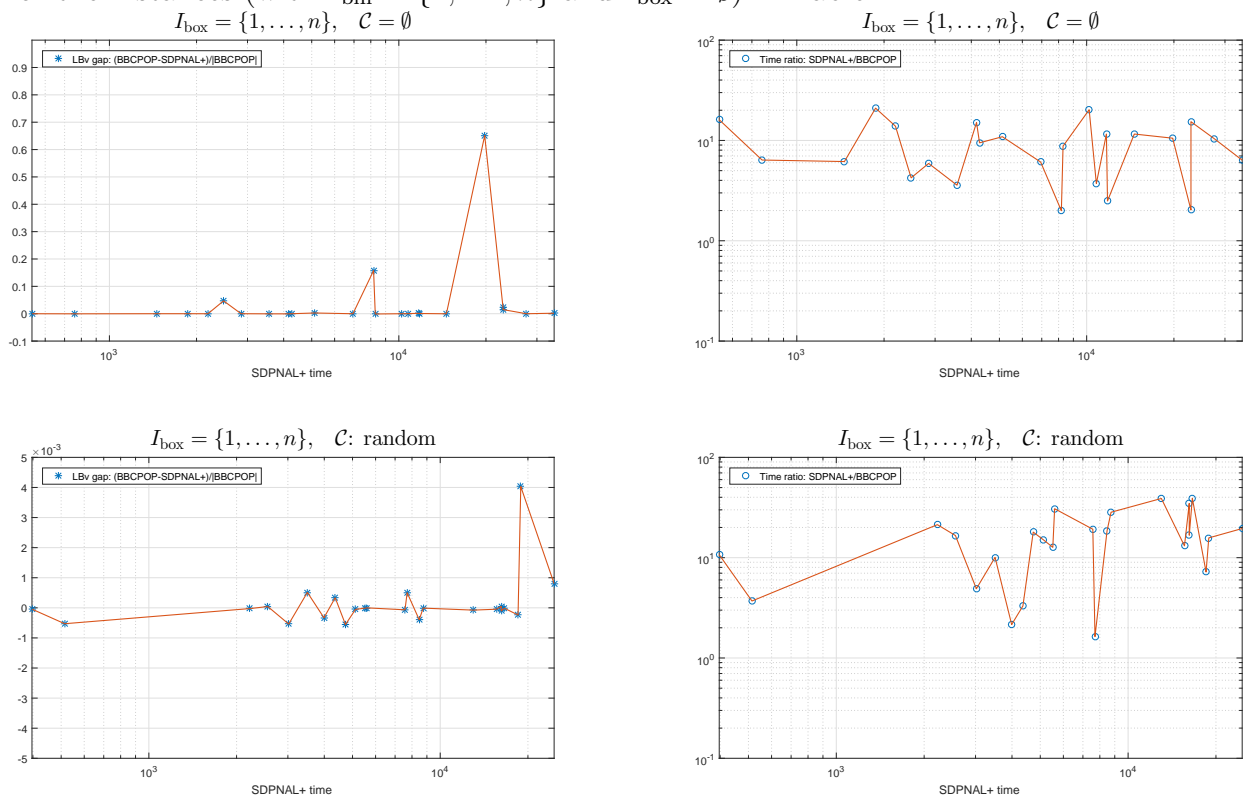


Figure 3: Comparison of the computational efficiency between BBCPOPOP and SDPNAL+ for the instances (with $I_{\text{bin}} = \emptyset$ and $I_{\text{box}} = \{1, \dots, n\}$) in Table 2.

Table 3: Computational results on small QAP instances. Both BBCPOP and SDPNAL+ were applied to the Lagrangian-DNN relaxation.

instance	BBCPOP	SDPNAL+	SDPNAL+ (AW+)
	LBv (sec, apgit:bpit)	LBv (sec, iter)	LBv (time, iter)
chr12a	9551.9 (8.98e0, 1984: 16)	9304.2 (1.06e2, 20000)	9551.9 (1.28e1, 1553)
chr12b	9741.8 (1.05e1, 2056: 16)	9669.5 (1.17e2, 20000)	9741.9 (1.10e1, 1553)
chr12c	11155.9 (8.63e0, 2042: 17)	10911.0 (1.28e2, 20000)	11156.0 (3.45e1, 3591)
had12	1651.9 (8.38e0, 1488: 16)	1616.0 (1.60e2, 20000)	1652.0 (1.72e1, 1522)
nug12	567.9 (2.04e1, 3701: 15)	565.6 (1.78e2, 20000)	567.8 (1.60e1, 1054)
rou12	235521.1 (4.62e1, 6681: 23)	229651.9 (1.44e2, 20000)	235518.8 (6.47e1, 5464)
scr12	31407.6 (2.33e1, 3843: 28)	30827.0 (1.62e2, 20000)	31410.0 (7.07e0, 495)
tai12a	224411.0 (1.68e1, 3154: 22)	219806.4 (1.42e2, 20000)	224416.0 (6.57e0, 426)
tai12b	39464040.0 (1.47e1, 2716: 19)	38647340.0 (1.41e2, 20000)	39464910.0 (1.95e1, 2355)

Here we characterize a permutation matrix \mathbf{X} as follows:

$$\begin{aligned} \mathbf{X} &\in [0, 1]^{r \times r}, \quad \mathbf{X}\mathbf{e} = \mathbf{X}^t\mathbf{e} = \mathbf{e}, \\ X_{ik}X_{jk} &= X_{ki}X_{kj} = 0 \quad (i, j, k = 1, 2, \dots, r; \quad i \neq j). \end{aligned}$$

Let $\mathbf{X} = (\mathbf{x}^1, \dots, \mathbf{x}^r) \in \mathbb{R}^{r \times r}$, where \mathbf{x}^p denotes the p -th column vector of \mathbf{X} . Let $n = r^2$. Arranging the columns \mathbf{x}^p ($1 \leq p \leq r$) of \mathbf{X} vertically into a long vector $\mathbf{x} = [\mathbf{x}^1; \dots; \mathbf{x}^r]$, we obtain the following reformulation of (22):

$$\min \left\{ \mathbf{x}^T (\mathbf{B} \otimes \mathbf{A}) \mathbf{x} \mid \begin{array}{l} \mathbf{x} \in [0, 1]^n, \quad (\mathbf{I} \otimes \mathbf{e}^T) \mathbf{x} = (\mathbf{e}^T \otimes \mathbf{I}) \mathbf{x} = \mathbf{e}, \\ x_i x_j = 0 \quad (i, j \in J_p; i \neq j; 1 \leq p \leq 2r), \end{array} \right\} \quad (23)$$

where

$$\begin{aligned} J_i &= \{(i-1)r + 1, (i-1)r + 2, \dots, (i-1)r + r\} \quad (1 \leq i \leq r), \\ J_{r+j} &= \{j, j+r, \dots, j+(r-1)r\} \quad (1 \leq j \leq r). \end{aligned}$$

Since it contains $2r$ equality constraints, the Lagrangian relaxation is applied to (23). More precisely, let $\mathbf{C} = [\mathbf{I} \otimes \mathbf{e}^T; \mathbf{e}^T \otimes \mathbf{I}]$ and $\mathbf{d} = [\mathbf{e}; \mathbf{e}]$, and consider the following Lagrangian relaxation problem of (23) with a penalty parameter $\lambda > 0$:

$$\min \left\{ \mathbf{x}^T (\mathbf{B} \otimes \mathbf{A}) \mathbf{x} + \lambda \frac{\|(\mathbf{B} \otimes \mathbf{A})\|}{\left\| \begin{pmatrix} \mathbf{d}^T \mathbf{d} & \mathbf{d}^T \mathbf{C} \\ \mathbf{C} \mathbf{d} & \mathbf{C}^T \mathbf{C} \end{pmatrix} \right\|} \|\mathbf{C}\mathbf{x} - \mathbf{d}\|^2 \mid \begin{array}{l} \mathbf{x} \in [0, 1]^n, \\ x_i x_j = 0 \quad (i, j \in J_p, i \neq j, \\ 1 \leq p \leq 2r) \end{array} \right\}, \quad (24)$$

which is a special case of POP (1) or (10). Thus, we can apply the DNN relaxation discussed in Section 3 to (24) and obtain a COP of the form (17), which serves as the Lagrangian-DNN relaxation [20, 4] of QAP (23). We set $\lambda = 10^5$ for all instances, and solved it by BBCPOP and SDPNAL+. For comparison, SDPNAL+ was also applied to another DNN relaxation, AW+ formulation [34], which is a benchmark formulation for SDPNAL+ in the paper [42]. We mention that the AW+ formulation does not use the Lagrangian relaxation.

In Table 3, we see that SDPNAL+ applied to the Lagrangian-DNN relaxation of QAP (23) shows inferior results compared to those obtained from BBCPOP. This is mainly because the Lagrangian-DNN relaxation is highly degenerated. BBCPOP, on the other hand, could solve such ill-conditioned problems successfully, which demonstrates the robustness of BP algorithm for ill-conditioned COPs. As the numerical results in [42] show that the AW+

Table 4: Computational results on large-scale QAP instances. BBCPOP was applied to the Lagrangian-DNN relaxation, and SDPNAL+ to the AW+ formulation.

instances	opt	BBCPOP			SDPNAL+ (AW+)				
		LBv	(sec,	apgit:bpit,	term)	LBv	(sec,	iter,	term)
chr15a	9896.0	9895.9	(3.78e1,	2572:18	, 3)	9892.2	(2.21e2,	6714,	0)
chr15b	7990.0	7989.9	(3.25e1,	2335:16	, 3)	7989.8	(5.25e1,	2066,	0)
chr15c	9504.0	9503.9	(3.11e1,	2346:19	, 1)	9504.0	(4.26e1,	1432,	0)
chr18a	11098.0	11097.9	(1.27e2,	3608:19	, 2)	11088.3	(2.77e2,	6586,	0)
chr18b	1534.0	1532.5	(2.76e2,	6117:17	, 1)	1533.9	(5.57e1,	917,	0)
chr20a	2192.0	2191.9	(1.97e2,	3264:17	, 1)	2191.7	(4.55e2,	6402,	0)
chr20b	2298.0	2298.0	(1.25e2,	2194:17	, 1)	2297.9	(2.91e2,	3412,	0)
chr20c	14142.0	14141.7	(1.86e2,	3352:20	, 2)	14139.4	(3.29e2,	5689,	0)
chr22a	6156.0	6156.0	(2.90e2,	2926:19	, 1)	6153.4	(7.65e2,	5791,	0)
chr25a	3796.0	3795.9	(5.62e2,	3139:18	, 1)	3795.8	(1.06e3,	3866,	0)
nug20	2570.0	2506.0	(2.24e2,	3024:18	, 1)	2505.9	(2.21e2,	2056,	0)
nug25	3744.0	3625.4	(8.68e2,	3911:19	, 1)	3625.3	(6.65e2,	2053,	0)
nug30	6124.0	5948.9	(2.30e3,	3634:21	, 1)	5948.7	(2.10e3,	2300,	0)
bur26a	5426670.0	5426095.0	(2.17e3,	7280:20	, 2)	5425904.2	(3.33e3,	9350,	0)
bur26b	3817852.0	3817277.4	(2.14e3,	7265:20	, 2)	3817148.6	(2.37e3,	6608,	0)
bur26c	5426795.0	5426203.7	(2.70e3,	9049:21	, 2)	5426457.7	(6.19e3,	18182,	0)
bur26d	3821225.0	3820014.6	(1.90e3,	6470:20	, 2)	3820601.2	(3.45e3,	10800,	0)
bur26e	5386879.0	5386572.0	(7.77e2,	3110:21	, 2)	5386585.4	(3.52e3,	11011,	0)
bur26f	3782044.0	3781834.8	(6.45e2,	2545:21	, 2)	3781760.8	(2.67e3,	8568,	0)
bur26g	10117172.0	10116571.1	(5.01e2,	2117:20	, 2)	10116504.8	(2.23e3,	8341,	0)
bur26h	7098658.0	7098236.7	(6.10e2,	2560:22	, 2)	7098381.4	(2.15e3,	8728,	0)
tai30a	1818146.0	1706789.7	(8.44e2,	1357:21	, 2)	1706814.3	(2.18e3,	2501,	0)
tai30b	637117113.0	598629428.0	(4.90e3,	7388:21	, 2)	598979464.0	(8.21e3,	12701,	0)
tai35a	2422002.0	2216540.3	(2.38e3,	1482:22	, 2)	2216573.2	(3.33e3,	1751,	0)
tai35b	283315445.0	269532369.0	(8.65e3,	6272:21	, 2)	269624118.0	(1.32e4,	8153,	0)
tai40a	3139370.0	2843198.6	(4.52e3,	1223:21	, 2)	2843245.1	(1.90e4,	2801,	0)
tai40b	637250948.0	608808415.0	(3.04e4,	8167:21	, 2)	608955916.0	(3.40e4,	6794,	0)
tai50a	4938796.0	4390743.4	(4.27e4,	3537:21	, 2)	4390862.9	(4.82e4,	2651,	0)
tai50b	458821517.0	431090745.0	(4.86e4,	5072:21	, 2)	431074160.0	(9.20e4,	7300,	0)
sko42	15812.0	15332.6	(1.89e4,	3898:21	, 2)	15332.5	(2.18e4,	3593,	0)
sko49	23386.0	22650.2	(3.62e4,	3383:21	, 2)	22650.4	(5.85e4,	3517,	0)
lipa40a	31538.0	31536.5	(9.24e3,	3062:22	, 2)	31538.0	(1.44e4,	3497,	0)
lipa40b	476581.0	476563.3	(1.18e4,	4405:21	, 2)	476581.0	(4.61e3,	935,	0)
lipa50a	62093.0	62089.6	(5.35e4,	2061:21	, 2)	62093.0	(6.69e4,	3099,	0)
lipa50b	1210244.0	1210195.2	(6.84e4,	2795:21	, 2)	1210244.0	(3.12e4,	1554,	0)
tho40	240516.0	226490.1	(1.82e4,	4826:21	, 2)	226482.4	(1.38e4,	2500,	0)
wil50	48816.0	48121.0	(6.66e4,	5453:21	, 2)	48120.1	(6.97e4,	4178,	0)

formulation works well for SDPNAL+, it is used in the subsequent experiments on large scale problems.

The results for large-scale QAPs are shown in Table 4. For most problems, BBCPOP produced comparable lower bounds with SDPNAL+ while it terminated slightly faster than SDPNAL+ in many cases. We note that the original QAP is not in the form of (1) to which BBCPOP can be applied. The results in Table 4, however, indicate that BBCPOP can solve the Lagrangian-DNN relaxations of very large-scale QAPs with high efficiency.

Finally, we note that various equivalent DNN relaxation formulations for QAPs have been proposed. The performance of the BP method and SDPNAL+ are expected to differ from one formulation to another; see Section 7 of [15] for more numerical results and investigation on the differences in the formulations and the performance of the BP method and SDPNAL+.

6 Concluding remarks

We have introduced a Matlab software package, BBCPOP, to compute valid lower bounds for the optimal values of large-scale sparse POPs with binary, box and complementarity (BBC) constraints. The performance of BBCPOP has been illustrated with the numerical results in Section 5 on various large-scale sparse POP instances with BBC constraints in comparison to those of SDPNAL+.

In terms of the number of variables, BBCPOP can handle efficiently larger POPs than the other available software packages and numerical methods for POPs such as GlotiPoly [13], SOSTOOLS [32], SparsePOP [38], BSOS [24], and SBSOS [41]. BBCPOP not only automatically generate sparse DNN relaxations of a given POP with BBC constraints, it also provides a robust numerical method, BP Algorithm, specially designed for solving the DNN relaxation problems. This is in contrast to the other software packages such as GloptiPoly, SparsePOP, SOSTOOLS, BSOS, and SBSOS that need to rely on an available SDP solver. As a result, their performance depends on the SDP solver chosen. One important feature of the lower bounds computed by BBCPOP is that it is theoretically guaranteed to be valid, whereas the lower bounds obtained by the other software packages, however tight they may be, is not guaranteed to be valid unless the relaxation problem is solved to very high accuracy.

For general POPs with polynomial equality and inequality constraints, Lagrangian relaxations can be applied for BBCPOP. The Lagrangian relaxation approach was successfully used to solve combinatorial QOPs as shown in [20] and in Section 5.2, and can be extended to general POPs. General POPs are difficult problems to solve with a computational method. It has been our experience that only the combination of SparsePOP [38] and the implementation of the primal-dual interior-point method, SeDuMi [35], could successfully deal with general POPs of moderate size. Other softwares including SDPNAL+ [42] have been unsuccessful to provide valid lower bounds for general POPs because the relaxation problems are generally highly degenerate. As a future work, we plan to investigate the Lagrangian relaxation methods together with BBCPOP for solving DNN relaxations of general POPs.

References

- [1] A. A. Ahmadi and A. Majumdar. Some applications of polynomial optimization in operations research and real-time decision making some applications of polynomial optimization in operations research and real-time decision making. *Optim. Lett.*, 10(4):709–729, 2016.
- [2] 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.
- [3] N. Arima, S. Kim, M. Kojima, and K. C. Toh. Lagrangian-conic relaxations, part II: Applications to polynomial optimization problems. Research report B-476, 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. *Comput. Optim. Appl.*, 66(3):453–479, 2017.
- [5] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2:183–202, 2009.
- [6] D. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Optimization and Computation Series. Athena Scientific, 2003.
- [7] J. R. S. Blair and B. Peyton. 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.
- [8] P. L. Combettes and N. N. Reyes. Moreau’s decomposition in banach spaces. *Math. Program.*, 139(1):103–114, June 2013.
- [9] K. Fujisawa, M. Fukuda, K. Kobayashi, M. Kojima, K. Nakata, M. Nakata, and M. Yamashita. SDPA (SemiDefinite Programming Algorithm) User’s Manual — Version 7.0.5. Research Report B-448, Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology, Oh-Okayama, Meguro, Tokyo 252-8552, February 2008.
- [10] 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.
- [11] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten. Convex optimization-based beamforming. *IEEE Signal Process. Mag.*, 27(3):62–75, 2010.
- [12] B. Ghaddar, J. Marecek, and M. Mevissen. Optimal power flow as a polynomial optimization problem. *IEEE Trans. Power Syst.*, 31(1):539–546, 2016.
- [13] D. Henrion and J. B. Lasserre. Gloptipoly: Global optimization over polynomials with matlab and sedumi. *ACM Trans. Math. Soft.*, 29(2):165–194, 2003.
- [14] N. Ito. *Applications of Accelerated Proximal Gradient Methods to Binary Classification and Polynomial Optimization*. PhD thesis, The university of Tokyo, 2018.

- [15] N. Ito, S. Kim, M. Kojima, A. Takeda, and K. C. Toh. Equivalences and differences in conic relaxations of combinatorial optimization problems. Research Rport B-483, Tokyo Institute of Technology, Department of Mathematical and Computing Sciences, Oh-Okayama, Meguro-ku, Tokyo 152-8552, July 2017.
- [16] N. Ito, S. Kim, M. Kojima, A. Takeda, and K. C. Toh. Numerical results on QAPs. http://www.opt.mist.i.u-tokyo.ac.jp/~n_ito/papers/IKKTT2017a_full_table.pdf, 2017.
- [17] N. Ito, A. Takeda, and K.-C. Toh. A unified formulation and fast accelerated proximal gradient method for classification. *J. Mach. Learn. Res.*, 18(16):1–49, 2017.
- [18] S. Iwata and K. Nagano. Submodular function minimization under covering constraints. In *50th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 671–680, 2009.
- [19] S. Kim, M. Kojima, and K. C. Toh. Doubly nonnegative relaxations for quadratic and polynomial optimization problems with binary and box constraints. Research Rport B-483, Tokyo Institute of Technology, Department of Mathematical and Computing Sciences, Oh-Okayama, Meguro-ku, Tokyo 152-8552, July 2016.
- [20] 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.
- [21] J. B. Lasserre. An explicit exact SDP relaxation for nonlinear 0-1 programs. In *Integer Programming and Combinatorial Optimization*, pages 293–303. Springer, 2001.
- [22] J. B. Lasserre. Global optimization with polynomials and the problems of moments. *SIAM J. Optim.*, 11:796–817, 2001.
- [23] J. B. Lasserre. Convergent SDP-Relaxations in polynomial optimization with sparsity. *SIAM J. Optimization*, 17(3):822–843, 2006.
- [24] J. B. Lasserre, K.-C. Toh, and S. Yang. A bounded degree SOS hierarchy for polynomial optimization. *EURO J. Comput. Optim.*, 5(1-2):87–117, 2017.
- [25] Z.-Q. Luo, Wing-Kin Ma, A. M.-C. So, Y. Ye, and S. Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Process. Mag.*, 27(3):20–34, 2010.
- [26] D. K. Molzahn and I. A. Hiskens. Moment-based relaxation of the optimal power flow problem. IEEE, 2014.
- [27] R. D. C. Monteiro, C. Ortiz, and B. F. Svaiter. An adaptive accelerated first-order method for convex optimization. *Comput. Optim. Appl.*, 64(31-73), 2016.
- [28] J. J. Moreau. Décomposition orthogonale d’un espace hilbertien selon deux cônes mutuellement polaires. *C. R. Acad. Sci.*, 255:238–240, 1962.
- [29] Y. Nesterov. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady*, 27(2):372–376, 1983.
- [30] J. Nie and L. Wang. Regularization methods for SDP relaxations in large scale polynomial optimization. *SIAM J. Optim.*, 22(2):408–428, 2012.

- [31] B. O’Donoghue and E. Candès. Adaptive restart for accelerated gradient schemes. *foundations of computational mathematics*. *Found. Comput. Math.*, 15(715-732), 2015.
- [32] A. Papachristodoulou, J. Anderson, G. Valmorbida, S. Prajna, P. Seiler, and P. Parrilo. *Sums of squares optimization toolbox for Matlab – User’s guide*. Department of Engineering Science, University of Oxford, Oxford, U.K., 2016.
- [33] P. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program.*, 96:293–320, 2003.
- [34] J. Povh and F. Rendl. A copositive programming approach to graph partitioning. *SIAM J. Optim.*, 18:223–241, 2007.
- [35] J. F. Sturm. SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods and Softw.*, 11&12:625–653, 1999.
- [36] R. H. Tütüncü, K. C. Toh, and M. J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Math. Program.*, 95:189–217, 2003.
- [37] 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.
- [38] H. Waki, S. Kim, M. Kojima, M. Muramatsu, and H. Sugimoto. Algorithm 883: Sparsepop: A sparse semidefinite programming relaxation of polynomial optimization problems. *ACM Trans. Math. Softw.*, 35(15), 2008.
- [39] H. Waki, S. Kim, M. Kojima, M. Muramatsu, H. Sugimoto, and M. Yamashita. *User Manual for SparsePOP: a Sparse Semidefinite Programming Relaxation of Polynomial Optimization Problems*. Tokyo Institute of Technology, B-414 edition, August 2009.
- [40] P. Wan, D.-Z. Du, P. Pardalos, and W. Wu. Greedy approximations for minimum submodular cover with submodular cost. *Comput. Optim. Appl.*, 45(2):463–474, 2010.
- [41] T. Weisser, J. B. Lasserre, and Kim-Chuan Toh. Sparse-BSOS: a bounded degree SOS hierarchy for large scale polynomial optimization with sparsity. *Math. Prog. Comp.*, DOI 10.1007/s12532-017-0121-6, 2017.
- [42] L. Q. Yang, D. F. Sun, and K. C. Toh. SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. *Math. Prog. Comp.*, 7:331–366, 2015.
- [43] X. Y. Zhao, D.F. Sun, and K.-C. Toh. A Newton-CG augmented Lagrangian method for semidefinite programming. *SIAM J. Optimization*, 20:1737–1765, 2010.