A Spectral Bundle Method for Sparse Semidefinite Programs

Hesam Mojtahedi, Feng-Yi Liao, and Yang Zheng

***Abstract*— Semidefinite programs (SDPs) have found a wide range of applications in the field of control. To improve scalability, it is important to exploit the inherent sparsity for solving SDPs. In this paper, we develop a new spectral bundle algorithm that solves sparse SDPs without introducing additional variables. We first apply the standard chordal decomposition to replace a large positive semidefinite (PSD) constraint with a set of smaller coupled constraints. Then, we use the exact penalty method to move the PSD constraints into the cost function. This leads to an equivalent non-smooth penalized problem, which can be solved by bundle methods. We present a new efficient spectral bundle algorithm, where subgradient information is incorporated to update a lower approximation function at each iteration. We further establish sublinear convergences in terms of objective value, primal feasibility, dual feasibility, and duality gap. Under Slater's condition, the algorithm converges with the rate of $\mathcal{O}\left(1/\epsilon^3\right)$, and the rate improves to $\mathcal{O}\left(1/\epsilon\right)$ when strict complementarity holds. Our numerical experiments support the theoretical analysis.**

I. Introduction

Semidefinite programs (SDPs) are an important sub-field of optimization that involves the minimization of a linear objective function over the cone of positive semidefinite (PSD) matrices with linear constraints. The standard primal and dual forms of SDPs are in the form of

$$\begin{aligned} \min_{X} \quad & \langle C, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i, \quad i=1,\dots,m \\ & X \in \mathbb{S}^n_+, \end{aligned} \tag{1}$$

$$\begin{aligned} \max_{y,Z} \quad & b^\top y \\ \text{subject to} \quad & Z + \sum_{i=1}^m A_i y_i = C, \\ & Z \in \mathbb{S}^n_+, \end{aligned} \tag{2}$$

where $A_1, A_2, \dots, A_m, C \in \mathbb{S}^n$ and $b \in \mathbb{R}^m$ are problem data, $\mathbb{S}^n_+$ stands for the cone of PSD matrices, and $\langle \cdot, \cdot \rangle$ denotes the standard matrix inner product. SDPs have important applications in numerous fields such as combinatorial optimization [1], control theory [2], machine learning [3]. Furthermore, many graph-theoretic problems can be addressed using SDPs, including maximum cut, graph partitioning, and matrix completion [4].

In theory, SDPs can be solved to arbitrary accuracy in polynomial time using interior-point methods [5]. However, due to its computational complexity, it is often impractical to solve large-scale SDPs considering memory and time constraints [3], [6]. The state-of-the-art solvers for SDPs, such as MOSEK [7], can only solve medium-sized problems reliably (e.g., $n, m \leq 1000$ in (1)) on regular laptops. Improving the scalability of solving SDPs has received extensive research interest [3], [6], [8]. First-order algorithms are one promising direction for computational scalability when solutions of moderate accuracy are required. For example, a general conic solver based on the alternating direction method of multipliers (ADMM) was developed in [9]. This approach has been extended in [10] to exploit the underlying sparsity in SDPs based on chordal decomposition. We refer interested readers to [6, Section 3] for a recent survey. Despite the efficiency of first-order methods per iteration, obtaining high-accuracy solutions remains challenging and may require an unacceptable number of iterations due to the slow convergence.

Another approach is to apply structured decomposition to decompose a large PSD matrix $X \in \mathbb{S}^n_+$ into structured ones that are easier to impose positivity [6], [11], [12]. For a sparse matrix X, we can associate it with a graph, and the principal submatrices can be identified by maximal cliques of the sparsity graph (see Section II). If the sparsity graph is *chordal*, which means that all cycles of length greater than three have an edge between nonconsecutive vertices in a cycle, a clique-based decomposition is guaranteed to exist for sparse PSD matrices [13]. In this case, it is possible to equivalently replace a large matrix constraint $X \in \mathbb{S}^n_+$ with a set of smaller and coupled matrix constraints. This chordal decomposition strategy, combined with a dual result on the existence of PSD matrix completions, are promising to significantly reduce the computational complexity of SDPs that involve sparse PSD matrices [6], [14]–[16].

In this paper, we focus on the spectral bundle method proposed in [17], which shows fast practical convergence and enjoys low computation complexity per iteration. In particular, the dual SDP (2) is transformed into an equivalent eigenvalue optimization by exploring the constant trace property in [17]. Very recently, [18] generalized the spectral bundle method to any SDPs and showed convergence in terms of primal feasibility, dual feasibility, and primal-dual duality gap. Furthermore, a linear convergence rate of the spectral bundle method is discovered under certain assumptions [18].

In this work, inspired by [17], [18], we propose a first-order spectral bundle method to solve a primal sparse SDP that is characterized by a chordal graph or chordal extension. Specifically, instead of solving (1) directly, benefiting from chordal sparsity property, we decompose the large semidefinite constraint in (1) into several smaller ones. We emphasize that the smaller PSD constraints are interdependent in gen-

*This work is supported by NSF ECCS-2154650.

The authors are with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093, USA. Emails: hmojtahedi@ucsd.edu; fliao@ucsd.edu; zhengy@eng.ucsd.edu.

eral. In many existing methods, such as those outlined in [6, Section 3], different additional consensus constraints have been introduced to handle the coupled constraints. Instead, we solve an equivalent penalized problem without introducing extra variables, which is in the form of constrained non-smooth eigenvalue optimization. Similar to [17], [18], this problem is well-suited to be solved via bundle methods [19]. In particular, we adapt and tailor the techniques in [17], [18] to solve the resulting non-smooth problem, leading to a new spectral bundle algorithm for sparse SDPs. Assuming Slater's condition, we prove that the algorithm converges as $\mathcal{O}\left(1/\epsilon^3\right)$. If the problem satisfies strict complementarity, the convergence rate is enhanced to $\mathcal{O}\left(1/\epsilon\right)$. Our numerical results show that the algorithm converges to a highly accurate solution within a moderate number of iterations.

The rest of this paper is structured as follows. We cover some preliminaries on chordal graphs and bundle methods in Section II. In Section III, we introduce an exact penalization for sparse SDPs. This allows us to develop a new spectral bundle algorithm in Section IV. We present numerical results in Section V, and conclude the paper in Section VI. Some technical proofs are provided in our report [20].

Notation. For simplicity, in (1) and (2), we define the affine mapping $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ as $\mathcal{A}(\cdot) = \left[\langle A_1, \cdot\rangle, \ldots, \langle A_m, \cdot\rangle\right]^\mathsf{T}$ and the adjoint operator $\mathcal{A}^* : \mathbb{R}^m \to \mathbb{S}^n$ as $\mathcal{A}^*(y) = \sum_{i=1}^m A_i y_i$. Given a matrix $X \in \mathbb{R}^{n\times m}$, we denote $\|X\|_2 = \sqrt{\lambda_{\max}\left(X^\mathsf{T}X\right)}$ and $\|X\|_\mathrm{F} = \sqrt{\langle X, X\rangle}$ as the spectral norm and Frobenius norm, respectively. We also use $\|x\|_2 = \sqrt{x^\mathsf{T}x}$ to denote the l_2 norm of a vector $x \in \mathbb{R}^n$.

II. Preliminaries

In this section, we review some background on graph theory, matrix decomposition, and bundle methods for non-smooth optimization.

A. Chordal graphs and matrix decomposition

A graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is defined by a set of vertices $\mathcal{V}$ and a set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. A graph is called *undirected* if the edges do not have orientations, i.e., $(i, j) \in \mathcal{E} \Leftrightarrow (j, i) \in \mathcal{E}$. A subset of vertices $\mathcal{C} \subseteq \mathcal{V}$ is called a clique if every pair of vertices in $\mathcal{C}$ is connected by an edge. A clique is maximal if it is not a subset of any other clique. We use $|\mathcal{C}|$ to denote the number of vertices in the clique. A cycle in a graph is defined as a sequence of vertices and edges that begins and ends at the same vertex. A chord is an edge between two non-consecutive vertices in a cycle. An undirected graph $\mathcal{G}$ is called *chordal* if it contains at least one chord in every cycle of length greater than three. There are efficient algorithms, such as the maximal cardinality search [21] to identify chordal graphs and their maximal cliques.

Given a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, a matrix $X \in \mathbb{S}^n$ has sparsity pattern $\mathcal{E}$ if $X_{i,j} = X_{j,i} = 0$, $\forall (i, j) \notin \mathcal{E}, i \neq j$. We denote the space of sparse symmetric matrices by

$$\mathbb{S}^n(\mathcal{E}, 0) := \{X \in \mathbb{S}^n \mid X_{ij} = 0, \text{ if } (i, j) \notin \mathcal{E}, i \neq j\}.$$

Given a matrix $X \in \mathbb{S}^n$, let $\mathbb{P}_{\mathbb{S}^n(\mathcal{E},0)}(X)$ be the projection onto $\mathbb{S}^n(\mathcal{E}, 0)$ with respect to the Frobenius norm, i.e., $M = \mathbb{P}_{\mathbb{S}^n(\mathcal{E},0)}(X)$ with $M_{ij} = 0$, if $(i, j) \notin \mathcal{E}, i \neq j$ and $M_{ij} = X_{ij}$, otherwise. Then, we define the cone of positive-semidefinite completable matrices as

$$\mathbb{S}^n_+(\mathcal{E}, ?) := \mathbb{P}_{\mathbb{S}^n(\mathcal{E},0)}\left(\mathbb{S}^n_+\right).$$

In other words, $X \in \mathbb{S}^n_+(\mathcal{E}, ?)$ if some (or all) of the zero entries X_{ij} with $(i, j) \notin \mathcal{E}, i \neq j$ can be replaced with nonzeros to obtain a PSD matrix $\bar{X} \in \mathbb{S}^n_+$. We call $\bar{X}$ the PSD completion of $X \in \mathbb{S}^n_+(\mathcal{E}, ?)$.

Given a clique $\mathcal{C}_k$ of graph $\mathcal{G}$, we define an index matrix $E_{\mathcal{C}_k} \in \mathbb{R}^{|\mathcal{C}_k|\times n}$ as follow

$$(E_{\mathcal{C}_k})_{ij} = \begin{cases} 1, & \text{if } \mathcal{C}_k(i) = j \\ 0, & \text{otherwise}. \end{cases}$$

Given a matrix $X \in \mathbb{S}^n$, the operation $E_{\mathcal{C}_k} X E_{\mathcal{C}_k}^\mathsf{T} \in \mathbb{S}^{|\mathcal{C}_k|}$ selects the submatrix indexed by $\mathcal{C}_k$. Alternatively, given $Y \in \mathbb{S}^{|\mathcal{C}_k|}$, the operation $E_{\mathcal{C}_k}^\mathsf{T} Y E_{\mathcal{C}_k} \in \mathbb{S}^n$ expands Y into a sparse $n \times n$ matrix that contains Y as its principal submatrix indexed $\mathcal{C}_k$, and zero otherwise.

Theorem 1 ([6]): Given a chordal graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with maximal cliques $\mathcal{C}_1, \ldots, \mathcal{C}_p$, we have $X \in \mathbb{S}^n_+(\mathcal{E}, ?)$ if and only if $E_{\mathcal{C}_k} X E_{\mathcal{C}_k}^\mathsf{T} \in \mathbb{S}^{|\mathcal{C}_k|}_+$, $\forall k = 1, \ldots, p$.

This result replaces a large constraint $X \in \mathbb{S}^n_+(\mathcal{E}, ?)$ with a set of smaller PSD constraints, indexing by the cliques. If the chordal graph has small cliques, we can expect computational improvements, which have been widely used in the literature (see [6] for a survey). In this paper, we will exploit Theorem 1 to develop a new spectral bundle method for sparse SDPs.

B. Bundle methods

The bundle method [19] is a standard technique to solve a non-smooth convex optimization problem of the form

$$\begin{aligned} \min_{x\in\mathcal{X}_0} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{X}_0, \end{aligned} \tag{3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex (possibly non-differentiable) function and $\mathcal{X}_0$ is a simple convex set. We refer the interested reader to [19] for a detailed discussion on bundle methods. We only introduce a few key ingredients below.

One key step in the bundle method is to construct a lower approximation $\hat{f}_t(x)$ of the objective function $f(x)$ at each iteration t, i.e., $\hat{f}_t(x) \leq f(x), \forall x \in \mathcal{X}_0$. One standard way is to use a subgradient to form an under-estimator $\hat{f}(x) = f(\hat{x}) + \langle g, x - \hat{x}\rangle$ where g is a subgradient of f at point $\hat{x}$, but other methods also exist [17]. At each iteration of the bundle method, we perform the following proximal step

$$y_{t+1} \in \operatorname*{argmin}_x \quad \hat{f}_t(x) + \frac{\alpha}{2}\left\|x - x_t\right\|^2, \tag{4}$$

where x_t is the current reference point and $\alpha > 0$ regulates the deviation from the reference point. If the candidate point y_{t+1} gives a sufficient descent in the true cost function, i.e. let $\beta \in (0, 1)$, we have

$$f\left(y_{t+1}\right) \leq f\left(x_t\right) - \beta\left(f\left(x_t\right) - \hat{f}_t\left(y_{t+1}\right)\right),$$

then we update the current (reference) iterate $x_{t+1} = y_{t+1}$ (*descent step*); otherwise, the reference point does not change, $x_{t+1} = x_t$ (*null step*). In either case, y_{t+1} will be used to update the lower approximation function $\hat{f}_{t+1}(x)$.

The bundle method is guaranteed to return a converging sequence x_t to a minimizer of (3) (if exists), when the lower approximation $\hat{f}_t$ satisfies three properties [22]

$$\hat{f}_{t+1}(x) \le f(x), \ \forall x \in \mathcal{X}_0, \quad (5a)$$

$$\hat{f}_{t+1}(x) \ge f\left(y_{t+1}\right) + \left\langle g_{t+1}, x - y_{t+1}\right\rangle, \forall x \in \mathcal{X}_0, \quad (5b)$$

$$\hat{f}_{t+1}(x) \ge \hat{f}\left(y_{t+1}\right) + \left\langle s_{t+1}, x - y_{t+1}\right\rangle, \forall x \in \mathcal{X}_0, \quad (5c)$$

where g_{t+1} is a subgradient of $f(x)$ at y_{t+1}, and $s_{t+1} = \alpha(x_t - y_{t+1})$. We refer the readers to [22] for more details.

III. Exact penalization for sparse SDPs

In this section, we introduce an exact penalization of sparse SDPs (1) into the form of (3). This allows us to develop the spectral bundle method in the next section.

A. Exact penalization for constrained convex optimization

Consider a constrained convex optimization problem:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \ i = 1, \dots, m, \\ & x \in \mathcal{X}_0, \end{aligned} \quad (6)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \dots, m$ are (possibly non-differentiable) convex functions, $\mathcal{X}_0 \subseteq \mathbb{R}^n$ is a convex closed set. The idea of exact penalty methods is to reformulate (6) by introducing an exact penalty function

$$P(x) = \sum_{i=1}^{m} \max\{0, g_i(x)\}.$$

We then consider a penalized problem

$$\begin{aligned} \min_{x} \quad & \Phi_\rho(x) := f(x) + \rho P(x) \\ \text{subject to} \quad & x \in \mathcal{X}_0, \end{aligned} \quad (7)$$

where $\rho > 0$ is a penalty parameter. It is known that when choosing ρ large enough and assuming Slater's condition, problems (6) and (7) are equivalent to each other in the sense that they have the same optimal value and solution set [23, Theorem 7.21]. Therefore, we can transform some nonsmooth constraints that are hard to handle in (6) into the nonsmooth cost in (7). Then, we can apply the bundle method (cf. Section II-B) to solve the nonsmooth problem (7).

B. Non-smooth penalization of sparse SDPs

We consider the standard primal SDP (1). In many practical applications, the matrices $A_1, \dots, A_m, C$ in problem data are often sparse [6]. If they share a common sparsity pattern $\mathcal{G}(\mathcal{V}, \mathcal{E})$, i.e., $C \in \mathbb{S}^n(\mathcal{E}, 0), A_i \in \mathbb{S}^n(\mathcal{E}, 0), i = 1, \dots, m$, we refer to this graph as *aggregate sparsity* pattern. It is not difficult to verify that (1) is equivalent to

$$\begin{aligned} \min_{X} \quad & \langle C, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i, \quad i = 1, \dots, m, \\ & X \in \mathbb{S}^n_+(\mathcal{E}, ?). \end{aligned} \quad (8)$$

Without loss of generality, we assume the aggregate sparsity pattern $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is a chordal graph with maximal cliques $\mathcal{C}_1, \dots, \mathcal{C}_p$ (otherwise a suitable chordal extension can be performed). Then, Theorem 1 allows us to reformulate problem (8) into

$$\begin{aligned} \min_{X} \quad & \langle C, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i, \quad i = 1, \dots, m \\ & E_{\mathcal{C}_k} X E_{\mathcal{C}_k}^{\mathsf{T}} \in \mathbb{S}_+^{|\mathcal{C}_k|}, \quad k = 1, \dots, p. \end{aligned} \quad (9)$$

The possibly large semidefinite constraint in (1) is replaced by several smaller constraints in (9). This decomposition (9) underpins many scalable algorithms for solving sparse SDPs [6], [10], [24]. Note that the submatrices $E_{\mathcal{C}_k} X E_{\mathcal{C}_k}^{\mathsf{T}}$ induced by maximal cliques may overlap, therefore the semidefinite constraints in (9) are coupled. Many previous techniques in [6], [10], [24] introduce a large number of extra consensus constraints such as $X_k = E_{\mathcal{C}_k} X E_{\mathcal{C}_k}^{\mathsf{T}}$.

In this work, we introduce further reformulations which allow us to solve (9) directly without adding extra variables. It is clear that (9) is equivalent to

$$\begin{aligned} \min_{X} \quad & \langle C, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i, \quad i = 1, \dots, m \\ & \lambda_{\max}\left(E_{\mathcal{C}_k}(-X)E_{\mathcal{C}_k}^{\mathsf{T}}\right) \le 0, \ k = 1, \dots, p. \end{aligned} \quad (10)$$

The eigenvalue constraints in (10) are non-smooth. Similar to (7), we then apply the exact penalty method to get the following formulation

$$\begin{aligned} \min_{X} \quad & \langle C, X\rangle + \rho \sum_{k=1}^{p} \max\left\{0, \lambda_{\max}\left(E_{\mathcal{C}_k}(-X)E_{\mathcal{C}_k}^{\mathsf{T}}\right)\right\} \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i, \quad i = 1, \dots, m. \end{aligned} \quad (11)$$

Unlike previous results in [6], [10], [24], this formulation (11) only has a single matrix variable X (with no extra variables). It is clear (11) is in the form of (3). We will develop a spectral bundle method to solve (11) in Section IV.

C. Properties and assumptions

Before developing the spectral bundle method, we expect that problems (11) and (1) are equivalent when the penalty parameter ρ is large enough. In particular, let us consider the Lagrange dual problem of (9), which is

$$\begin{aligned} \max_{y, Y_k} \quad & b^{\mathsf{T}} y \\ \text{subject to} \quad & C - \sum_{i=1}^{m} y_i A_i = \sum_{k=1}^{p} E_{\mathcal{C}_k}^{\mathsf{T}} Y_k E_{\mathcal{C}_k}, \\ & Y_k \in \mathbb{S}_+^{|\mathcal{C}_k|}, \quad k = 1, \dots, p. \end{aligned} \quad (12)$$

It can be verified that (12) is also equivalent to (2). We denote the optimal solution set of the decomposed primal SDP (9) and dual SDP (12) by $\mathcal{P}^\star$ and $\mathcal{D}^\star$ respectively. Throughout the paper, we make the following assumptions.

Assumption 1: The primal and dual SDPs (1) and (2) satisfy Slater's condition, i.e., they are strictly feasible.

Assumption 2: The optimal solution sets for (9) and (12), i.e., $\mathcal{P}^\star$ and $\mathcal{D}^\star$, are compact.

It can be verified that the problem (9) and (12) also satisfy Slater's condition by Assumption 1. We have the following technical result, and its proof is provided in our report [20].

Proposition 1: Under Assumptions 1 and 2, the nonsmooth penalized formulation (10) is equivalent to the original primal SDP (1) if we choose

$$\rho > \mathcal{D}_{y^\star} := \max_{(y^\star,\{Y_k^\star\})\in\mathcal{D}^\star} \left\{\mathbf{tr}(Y_1^\star), \mathbf{tr}(Y_2^\star), \ldots, \mathbf{tr}(Y_p^\star)\right\}.$$

We conclude this section with a notion of strict complementarity.

Definition 1 (strict complementarity): A pair of optimal solutions $(X^\star, \{y^\star, Y_k^\star\}) \in \mathcal{P}^\star \times \mathcal{D}^\star$ in (9) and (12) satisfies *strict complementarity* if

$$\operatorname{rank}\left(E_{\mathcal{C}_k} X^\star E_{\mathcal{C}_k}^\mathsf{T}\right) + \operatorname{rank}\left(Y_k^\star\right) = |\mathcal{C}_k|,\ k = 1, \ldots, p.$$

If such a pair exists, we say the decomposed SDPs (9) and (12) satisfy strict complementarity.

IV. A Spectral Bundle Method

In this section, we introduce a spectral bundle algorithm to solve the penalized nonsmooth problem (11).

A. Constructions of lower-approximation model

For simplicity, we denote the cost function in (11) as

$$G(X) := \langle C, X\rangle + \rho \sum_{k=1}^{p} \max\left\{0, \lambda_{\max}\left(E_{\mathcal{C}_k}(-X)E_{\mathcal{C}_k}^\mathsf{T}\right)\right\}.$$

As discussed in Section II-B, one key step in the bundle method is to construct an appropriate lower approximation model for $G(X)$ that satisfies (5a) to (5c). Our strategy for constructing a lower-approximation model is motivated by [17] and [18, Section 2.2] that focus on the dual SDP (2).

In particular, for each clique k, we use an orthonormal matrix $P_k \in \mathbb{R}^{|\mathcal{C}_k|\times r}$, where $r \le \max_k |\mathcal{C}_k|$, such that $P_k^\mathsf{T} P_k = I_{|\mathcal{C}_k|\times|\mathcal{C}_k|}$, and construct a lower approximation:

$$\begin{aligned}\hat{G}_{\{P_k\}}(X) = \langle C, X\rangle + \\ \rho \sum_{k=1}^{p} \max_{\substack{S_k\in\mathbb{S}_+^r,\\ \mathbf{tr}(S_k)\le 1}} \left\langle P_k S_k P_k^\mathsf{T}, E_{\mathcal{C}_k}(-X)E_{\mathcal{C}_k}^\mathsf{T}\right\rangle.\end{aligned} \tag{13}$$

By definition, we observe

$$\hat{G}_{\{P_k\}}(X) \le G(X), \quad \forall X \in \mathbb{S}^n$$

due to the fact

$$\max\{\lambda_{\max}(-X), 0\} = \max_{S\in\mathbb{S}_+^n, \mathbf{tr}(S)\le 1} \langle S, -X\rangle, \quad \forall X \in \mathbb{S}^n,$$

and $\{P_k S P_k^\mathsf{T} \in \mathbb{S}_+^n \mid S \in \mathbb{S}_+^r, \mathbf{tr}(S) \le 1\} \subseteq \{S \in \mathbb{S}_+^n \mid \mathbf{tr}(S) \le 1\}$. Therefore, $\hat{G}_{\{P_k\}}(X)$ in (13) serves as an underestimator that meets (5a).

We can also verify the subgradient lower bound condition (5b) when we choose P_k spanning the top eigenvector associated with $E_{\mathcal{C}_k}(-X)E_{\mathcal{C}_k}$. Another modification is required to ensure that the condition (5c) is fulfilled. Along with selecting past and current eigenvectors to generate P_k, the spectral bundle approach in [17] retains a thoughtfully chosen weight to incorporate past information. Notably, we introduce a constant matrix $\bar{W}_k \in \mathbb{S}_+^{|\mathcal{C}_k|}$ for each clique k with $\mathbf{tr}\left(\bar{W}_k\right) = 1$, and define the set, $k = 1, \ldots, p$

$$\begin{aligned}\hat{\mathcal{W}}_k := \{\gamma_k \bar{W}_k + P_k S_k P_k^\mathsf{T} \mid S_k \in \mathbb{S}_+^r, \\ \gamma_k \ge 0, \gamma_k + \mathbf{tr}\left(S_k\right) \le 1\},\end{aligned} \tag{14}$$

and refine the lower approximation function below

$$\begin{aligned}&\hat{G}_{\{\bar{W}_k, P_k\}}(X) \\ &= \langle C, X\rangle + \rho \sum_{k=1}^{p} \max_{W_k\in\hat{\mathcal{W}}_k} \left\langle W_k, E_{\mathcal{C}_k}(-X)E_{\mathcal{C}_k}^\mathsf{T}\right\rangle.\end{aligned} \tag{15}$$

It is evident that the lower approximation model (15) provides a better estimate than (13). Letting $\gamma_k = 0$ reduce (15) to (13); hence (15) satisfies (5a), (5b) , and as well as (5c) by meticulously constructing $\bar{W}_k$ and P_k at each iteration.

B. A spectral bundle algorithm

Following Section II-B, we present a spectral bundle algorithm to solve (11) based on the lower approximation model (15). In this algorithm, we will construct a lower approximation model $\hat{G}_{\{\bar{W}_{t,k}, P_{t,k}\}}(X_t)$ and update the model parameters $\{\bar{W}_{t,k}\}$, $\{P_{t,k}\}$, and the set $\hat{\mathcal{W}}_{t,k} := \{\gamma_{t,k}\bar{W}_{t,k} + P_{t,k}S_{t,k}P_{t,k}^\mathsf{T} \mid \gamma_{t,k} \ge 0, S_{t,k} \in \mathbb{S}_+^r, \gamma_{t,k} + \mathbf{tr}(S_{t,k}) \le 1\}$ at each iteration t. The overall algorithm is listed in Algorithm 1, which has the following steps:

Pre-processing: The algorithm starts by extracting aggregate sparsity pattern of the problem data and computing the maximal cliques $\mathcal{C}_1, \ldots, \mathcal{C}_p$. This step can be performed very efficiently, e.g., using algorithms in [25].

Initialization: The algorithm is initiated with a random reference point $\Omega_0 \in \mathbb{S}^n$ and $P_{0,k} \in \mathbb{R}^{|\mathcal{C}_k|\times r}$ by setting the top r eigenvectors of $E_{\mathcal{C}_k}(-\Omega_0)E_{\mathcal{C}_k}^T$ as their columns, We then choose any symmetric matrix $\bar{W}_{0,k} \in \mathbb{S}_+^n$ with $\mathbf{tr}\left(\bar{W}_{0,k}\right) = 1$, and construct the initial under-estimator $\hat{G}_{\{\bar{W}_{0,k}, P_{0,k}\}}(X)$ as in (15).

Solving the master problem: Similar to (4), our algorithm solves the following problem at iteration $t \ge 0$ to get the next iteration parameters and the candidate reference point

$$\begin{aligned}&\left(X_{t+1}^\star, S_{t,k}^\star, \gamma_{t,k}^\star\right) \\ &= \operatorname*{argmin}_{X\in\mathcal{X}_0} \hat{G}_{\{\bar{W}_{t,k}, P_{t,k}\}}(X) + \frac{\alpha}{2}\left\|X - \Omega_t\right\|_F^2,\end{aligned} \tag{16}$$

where $\mathcal{X}_0 = \{X \in \mathbb{S}^n \mid \langle A_i, X\rangle = b_i, i = 1, 2, \ldots, m\}$, and Ω_t is the reference point at iteration t and α is the penalization parameter which penalizes the deviation from the reference point at each iteration.

Update reference point: The algorithm updates the reference point if the following condition holds

$$\begin{aligned}&\beta\left(G\left(\Omega_t\right) - \hat{G}_{\{\bar{W}_{t,k}, P_{t,k}\}}\left(X_{t+1}^\star\right)\right) \\ &\le G\left(\Omega_t\right) - G\left(X_{t+1}^\star\right),\end{aligned} \tag{17}$$

where $\beta \in (0, 1)$. This condition indicates that if the actual cost reduction $G\left(\Omega_t\right) - G\left(X_{t+1}^\star\right)$ is greater or equal

Algorithm 1 Spectral bundle method for sparse SDPs

Require: Problem data $A_1, \ldots, A_m, C \in \mathbb{S}^n$, $b \in \mathbb{R}^n$.
Require: Parameters $r_{\mathrm{p}} \geq 0, r_{\mathrm{c}} \geq 1$, $\alpha > 0$, $\beta \in (0,1)$, $\epsilon \geq 0$. An initial point $\Omega_0 \in \mathbb{S}^n$.
Pre-processing: Extract aggregate sparsity pattern of the problem data and compute maximal cliques.
Initialization: Let $r = r_{\mathrm{p}} + r_{\mathrm{c}}$. Initialize $\bar{W}_{0,k} \in \mathbb{S}_+^{|\mathcal{C}_k|}$, with $\mathbf{tr}\left(\bar{W}_{0,k}\right) = 1$, and construct $P_{0,k} \in \mathbb{R}^{|\mathcal{C}_k| \times r}$ with its columns set to the top r orthonormal eigenvectors of $E_{\mathcal{C}_k}(-\Omega_0)E_{\mathcal{C}_k}^T$.
for $t = 0, \ldots, t_{\max}$ **do**
  Solve (16) to obtain $X_{t+1}^\star, \gamma_{k,t}^\star$, and $S_{k,t}^\star$. *master problem*
  If $G(\Omega_t) - \hat{G}_{\{\bar{W}_{t,k}, P_{t,k}\}}(X_{t+1}^\star) \leq \epsilon$, then stop.
  Set $\Omega_{t+1} = \begin{cases} X_{t+1}^\star, & \text{if (17) holds.} \quad \text{\\*descent step*} \\ \Omega_t, & \text{otherwise.} \quad \text{\\*null step*} \end{cases}$
  Compute $P_{k,t+1}$ as (18) and $\bar{W}_{k,t+1}$ as (19). *update model*
end for

than β portion of the approximate reduction $G(\Omega_t) - \hat{G}_{\{\bar{W}_{t,k}, P_{t,k}\}}\left(X_{t+1}^\star\right)$, a *decent step* happens and the algorithm updates the reference point, i.e., $\Omega_{t+1} = X_{t+1}^\star$. Otherwise, a *null step* happens and the reference point does not change, i.e., $\Omega_{t+1} = \Omega_t$.

Update the under-estimator model: The algorithm updates the approximator at each iteration to improve the approximation accuracy at the candidate reference point. Similar to [18, Section 2.2], to compute $\bar{W}_{t+1,k}, P_{t+1,k}$, we apply eigenvalue decomposition to small matrices $S_{k,t}^\star$ as below

$$S_{t,k}^\star = \begin{bmatrix} Q_{k,1} & Q_{k,2} \end{bmatrix} \begin{bmatrix} \Sigma_{k,1} & 0 \\ 0 & \Sigma_{k,2} \end{bmatrix} \begin{bmatrix} Q_{k,1}^{\mathsf{T}} \\ Q_{k,2}^{\mathsf{T}} \end{bmatrix},$$

where $Q_{k,1} \in \mathbb{R}^{r \times r_{\mathrm{p}}}$ and $Q_{k,2} \in \mathbb{R}^{r \times r_{\mathrm{c}}}$ contain the orthonormal eigenvectors associated with eigenvalues $\Sigma_{k,1}$ and $\Sigma_{k,2}$ respectively, $\Sigma_{k,1}$ consists of the largest r_{p} eigenvalues, and $\Sigma_{k,2}$ consists of the rest of the eigenvalues. We compute the $V_{t,k} \in \mathbb{R}^{|\mathcal{C}_k| \times r_{\mathrm{c}}}$ with its columns being the top $r_{\mathrm{c}} \geq 1$ orthonormal eigenvectors of $E_{\mathcal{C}_k}(-X_{t+1}^\star)E_{\mathcal{C}_k}^{\mathsf{T}}$ which captures the current sub-gradient information of the spectral bundle. Then, the next parameter $P_{t+1,k}$ is updated as

$$P_{t+1,k} = \mathrm{orth}\left(\begin{bmatrix} V_{t,k}, & P_{t,k}Q_{1,k} \end{bmatrix}\right). \tag{18}$$

The update of the weight matrices $\bar{W}_{t,k}$ captures the remaining past information

$$\bar{W}_{t+1,k} = \frac{\left(\gamma_{t,k}^\star \bar{W}_{t,k} + P_{t,k}Q_{k,2}\Sigma_{k,2}Q_{k,2}^{\mathsf{T}}P_{t,k}^{\mathsf{T}}\right)}{\gamma_{t,k}^\star + \mathbf{tr}\left(\Sigma_{k,2}\right)}, \tag{19}$$

where $\bar{W}_{t+1,k}$ has been normalized with $\mathbf{tr}(\bar{W}_{t+1,k}) = 1$. In the case that $r_{\mathrm{p}} = 0$, the parameter updates in (18) and (19) become $P_{t+1,k} = V_{t,k} \in \mathbb{R}^{|\mathcal{C}_k| \times r}$ and $\bar{W}_{t+1,k} = \frac{W_{t,k}^\star}{\mathbf{tr}(W_{t,k}^\star)}$ respectively, where $W_{t,k}^\star$ is the optimal solution of $\gamma_k \bar{W}_{t,k} + P_{t,k}S_kP_{t,k}^{\mathsf{T}}$ in (16).

C. Computational details

Solving the regularized master problem in (16) is the main computation in Algorithm 1. Therefore, it is crucial to solve the master problem efficiently. We summarize the computation details in Proposition 2. For notational simplicity, we define the linear mapping $\mathcal{E}_{\mathcal{C}_k} : \mathbb{S}^n \to \mathbb{S}^{|\mathcal{C}_k|}$ as $\mathcal{E}_{\mathcal{C}_k}(X) = E_{\mathcal{C}_k}XE_{\mathcal{C}_k}^{\mathsf{T}}$ and $\hat{\mathcal{E}}_{\mathcal{C}_k} : \mathbb{S}^{|\mathcal{C}_k|} \to \mathbb{S}^n$ as $\hat{\mathcal{E}}_{\mathcal{C}_k}(X) = E_{\mathcal{C}_k}^{\mathsf{T}}XE_{\mathcal{C}_k}$.

Proposition 2: The master problem (16) is equivalent to the following problem

$$\begin{aligned} \max_{\substack{W_k \in \hat{\mathcal{W}}_{t,k} \\ y \in \mathbb{R}^m}} & \left\langle C - \rho \sum_{k=1}^p \hat{\mathcal{E}}_{\mathcal{C}_k}(W_k), \Omega_t \right\rangle + \langle b - \mathcal{A}(\Omega_t), y \rangle \\ & - \frac{1}{2\alpha} \left\| \rho \sum_{i=k}^p \hat{\mathcal{E}}_{\mathcal{C}_k}(W_k) + \mathcal{A}^*(y) - C \right\|_{\mathrm{F}}^2. \end{aligned} \tag{20}$$

The optimal solution of X in (16) is recovered by

$$X_{t+1}^\star = \Omega_t + \frac{1}{\alpha}\left(\rho \sum_{k=1}^p \hat{\mathcal{E}}_{\mathcal{C}_k}(W_k) + \mathcal{A}^*(y) - C\right). \tag{21}$$

The proof is based on strong duality and the details are provided in the report [20]. When the iterate Ω_t is feasible, i.e., $\mathcal{A}(\Omega_t) = b$, (20) can be simplified as

$$\begin{aligned} \max_{\substack{W_k \in \hat{\mathcal{W}}_{t,k} \\ y \in \mathbb{R}^m}} & \left\langle C - \rho \sum_{k=1}^p \hat{\mathcal{E}}_{\mathcal{C}_k}(W_k), \Omega_t \right\rangle \\ & - \frac{1}{2\alpha} \left\| \rho \sum_{k=1}^p \hat{\mathcal{E}}_{\mathcal{C}_k}(W_k) + \mathcal{A}^*(y) - C \right\|_{\mathrm{F}}^2. \end{aligned}$$

D. Convergence guarantees

We present the convergence guarantee for Algorithm 1 when strong duality holds for (9) and (12). Our analysis is motivated by [18, Section 3] that solves dual SDP (2).

Theorem 2: Suppose strong duality holds for (9) and (12). Given any $\beta \in (0,1)$, $r_{\mathrm{c}} \geq 1$, $r_{\mathrm{p}} \geq 0$, $\alpha > 0$, $r = r_{\mathrm{c}} + r_{\mathrm{p}}$, $\rho > (2\mathcal{D}_{\mathcal{Y}^\star} + 1)$, $P_{0,k} \in \mathbb{R}^{n \times r}, \forall 1 \leq k \leq p$, $\Omega_0 \in \mathbb{S}^n$, and target accuracy $\epsilon > 0$, then Algorithm 1 outputs iterates $\left(\Omega_t, \{W_{t,k}^\star\}, y_t^\star\right)$ with

$$G(\Omega_t) - G(X^\star) \leq \epsilon, \tag{22a}$$

$$\left\| \rho \sum_{k=1}^p \hat{\mathcal{E}}_{\mathcal{C}_k}\left(W_{t,k}^\star\right) - C + \mathcal{A}^*(y_t^\star) \right\|_{\mathrm{F}}^2 \leq \epsilon, \; W_{t,k}^\star \succeq 0, \tag{22b}$$

$$\lambda_{\min}\left(E_{\mathcal{C}_k}\Omega_t E_{\mathcal{C}_k}^{\mathsf{T}}\right) \geq -\epsilon, 1 \leq k \leq p, \tag{22c}$$

$$|\langle C, \Omega_t \rangle - \langle b, y_t^\star \rangle| \leq \sqrt{\epsilon}, \tag{22d}$$

by taking some iteration $t \leq \mathcal{O}\left(1/\epsilon^3\right)$. If, additionally, the strict complementarity (see Definition 1) holds, then the conditions above are reached by some iteration $t \leq \mathcal{O}(1/\epsilon)$.

One main step for the proof of Theorem 2 is to connect the primal feasibility, dual feasibility, and primal-dual optimality to the cost value gap. We summarize those connections in the following lemma.

Lemma 1: Under the parameters in Theorem 2, at each descent step $t > 0$, the following results hold.

- The approximate dual feasibility satisfies

$$\left\| \rho \sum_{k=1}^{p} \hat{\mathcal{E}}_{\mathcal{C}_k}(W_{t,k}^\star) - C + \mathcal{A}^*(y_t^\star) \right\|_{\mathrm{F}}^2 \leq \frac{2\alpha}{\beta}(G(\Omega_t) - G(X^\star)).$$

- The approximate primal feasibility satisfies

$$\lambda_{\min}(\mathcal{E}_{\mathcal{C}_k}(\Omega_{t+1})) \geq \frac{-(G(\Omega_t) - G(X^\star))}{\mathcal{D}_{\mathcal{Y}^\star} + 1}, 1 \leq k \leq p, \quad \text{and} \quad \mathcal{A}(\Omega_{t+1}) = b.$$

- The approximate primal-dual optimality satisfies

$$\langle C, \Omega_{t+1}\rangle - \langle b, y_t^\star\rangle \geq \left(-(G(\Omega_t) - G(X^\star))\frac{p\rho}{\mathcal{D}_{\mathcal{Y}^\star}+1}\right) - \mathcal{D}_{\Omega_0}\sqrt{\frac{2\alpha}{\beta}(F(\Omega_t) - F(X^\star))},$$

$$\langle C, \Omega_{t+1}\rangle - \langle b, y_t^\star\rangle \leq \left(\frac{1-\beta}{\beta}(G(\Omega_t) - G(X^\star))\right) + \mathcal{D}_{\Omega_0}\sqrt{\frac{2\alpha}{\beta}(G(\Omega_t) - G(X^\star))},$$

where $\mathcal{D}_{\Omega_0} = \sup_{G(\Omega_t)\leq G(\Omega_0)} \|\Omega_t\|_{\mathrm{F}}$ is maximum norm value over the sublevel set which is bounded.

Lemma 2 (quadratic growth): Under Assumption 2 and the selection of parameters in Theorem 2, for any fixed $\epsilon > 0$ and X in the sub-level set $\mathcal{S}_\epsilon := \{X \in \mathbb{S}^n \mid G(X) \leq G(X^\star) + \epsilon, \mathcal{A}(X) = b, \|X\|_{\mathrm{F}} < \infty\}$, there exist some constants $\zeta \geq 1$ and $\mu > 0$ such that

$$G(X) - G(X^\star) \geq \mu \cdot \mathrm{dist}^\zeta(X, \mathcal{P}^\star).$$

Furthermore, if SDPs (9) and (12) satisfy strict complementarity, the exponent term $\zeta = 2$.

The proofs for Lemma 1 and 2 are provided the in the report [20]. With Lemmas 1 and 2, we are ready to prove Theorem 2 by utilizing convergence results in [22], [26].

Proof of Theorem 2. It remains to verify that Algorithm 1 satisfy (5a) to (5c). For notational convenience, we use $\hat{G}_t(X)$ to denote the approximation model $\hat{G}_{\{\bar{W}_{t,k}, P_{t,k}\}}(X)$ at iteration t. First, by the construction of $P_{t,k}$ and $\bar{W}_{t,k}$ in (18) and (19), we have $(P_{t,k})^{\mathsf{T}} P_{t,k} = I$, $\bar{W}_{t,k} \succeq 0$, and $\mathbf{tr}(\bar{W}_{t,k}) = 1$. It follows that

$$\hat{\mathcal{W}}_{t,k} \subset \{W \in \mathbb{S}_+^{|\mathcal{C}_k|} \mid \mathbf{tr}(W) \leq 1\},$$

which implies

$$\begin{aligned}
&\max_{W_k \in \hat{\mathcal{W}}_{t,k}} \langle W_k, \mathcal{E}_{\mathcal{C}_k}(-X)\rangle \\
\leq &\max_{\mathbf{tr}(W)\leq 1, W \in \mathbb{S}_+^r} \langle W, \mathcal{E}_{\mathcal{C}_k}(-X)\rangle \\
= &\max\{\lambda_{\max}\left(\mathcal{E}_{\mathcal{C}_k}(X)\right), 0\}, \forall X \in \mathbb{S}^n, 1 \leq k \leq p.
\end{aligned}$$

Therefore, the global lower-bound property (5a) is satisfied.

Second, since the update in (18) includes a subgradient information of $\lambda_{\max}(\mathcal{E}_{\mathcal{C}_k}(-X_{t+1}^\star))$, there exists a normalized vector $e_k \in \mathbb{R}^r$ such that $P_{t+1,k} e_k = v_k$, where v_k is a subgradient of $\lambda_{\max}(\mathcal{E}_{\mathcal{C}_k}(-X_{t+1}^\star))$. If $\lambda_{\max}(\mathcal{E}_{\mathcal{C}_k}(-X_{t+1}^\star)) > 0$, we let $\gamma_{t+1,k} = 0$ and $S_{t+1,k} = e_k(e_k)^{\mathsf{T}}$, otherwise, we let $\gamma_{t+1,k} = 0$ and $S_{t+1,k} = 0$. It follows that

$$\begin{aligned}
&\hat{G}_{t+1}(X) \\
\geq &\langle C, X\rangle + \rho \sum_{k\in\mathcal{I}} \left\langle P_{t+1,k}(e_k(e_k)^{\mathsf{T}})P_{t+1,k}^{\mathsf{T}}, \mathcal{E}_{\mathcal{C}_k}(-X)\right\rangle \\
= &\langle C, X\rangle + \rho \sum_{k\in\mathcal{I}} \left\langle v_k v_k^{\mathsf{T}}, \mathcal{E}_{\mathcal{C}_k}(-X)\right\rangle \\
= &\langle C, X_{t+1}^\star\rangle + \rho \sum_{k\in\mathcal{I}} \lambda_{\max}(\mathcal{E}_{\mathcal{C}_k}(-X_{t+1}^\star)) \\
&+ \langle C - \rho \sum_{k\in\mathcal{I}} \hat{\mathcal{E}}_{\mathcal{C}_k}(v_k v_k^{\mathsf{T}}), X - X_{t+1}^\star\rangle \\
= &G(X_{t+1}^\star) + \langle g_{t+1}, X - X_{t+1}^\star\rangle, \ \ \forall X \in \mathcal{X}_0,
\end{aligned}$$

where $g_{t+1} = C - \rho\sum_{k\in\mathcal{I}} \hat{\mathcal{E}}_{\mathcal{C}_k}(v_k v_k^{\mathsf{T}}) \in \partial G(X_{t+1}^\star)$ and $\mathcal{I} = \{k \mid \lambda_{\max}\left(\mathcal{E}_{\mathcal{C}_k}(-X_{t+1}^\star)\right) > 0, 1 \leq k \leq p\}$, which verified the second property (5b).

Third, we first argue that the optimal solution at step t is within the feasible region of the next iteration. Recalling the update procedure of the matrix $P_{t+1,k}$ in (18), there exists matrices $\bar{Q}_k \in \mathbb{R}^{|\mathcal{C}_k|\times r}$ with normalized columns such that $P_{t+1,k}\bar{Q}_k = P_{t,k}Q_{1,k}$. By letting $\gamma_{t+1} = \gamma_{t,k}^\star + \mathbf{tr}(\Sigma_{k,2})$ and $S_{t+1,k} = \bar{Q}_k \Sigma_{1,k} \bar{Q}_k^{\mathsf{T}}$, we recover the solution

$$\begin{aligned}
W_{t,k}^\star &= \gamma_{t+1}\bar{W}_{t+1,k} + P_{t+1,k}S_{t+1,k}P_{t+1,k}^{\mathsf{T}} \\
&= (\gamma_{t,k}^\star + \mathbf{tr}(\Sigma_{k,2}))\bar{W}_{t,k} + P_{t+1,k}\bar{Q}\Sigma_{1,k}\bar{Q}^{\mathsf{T}}P_{t+1,k}^{\mathsf{T}} \\
&= (\gamma_{t,k}^\star + \mathbf{tr}(\Sigma_{k,2}))\bar{W}_{t,k} + P_{t,k}Q_{1,k}\Sigma_{1,k}Q_{1,k}^{\mathsf{T}}P_{t,k}^{\mathsf{T}}.
\end{aligned}$$

Hence, it follows that

$$\begin{aligned}
&\hat{G}_{t+1}(X) \\
\geq &\langle C, X\rangle + \rho \sum_{k=1}^{p} \langle W_{t,k}^\star, \mathcal{E}_{\mathcal{C}_k}(-X)\rangle \\
= &\left\langle C - \rho \sum_{k=1}^{p} \hat{\mathcal{E}}_{\mathcal{C}_k}(W_{t,k}^\star), X_{t+1}^\star \right\rangle \\
&+ \left\langle C - \rho \sum_{k=1}^{p} \hat{\mathcal{E}}_{\mathcal{C}_k}(W_{t,k}^\star) + \mathcal{A}^*(y_t^\star), X - X_{t+1}^\star \right\rangle \\
= &\hat{G}_{t+1}(X_{t+1}^\star) + \langle \alpha(X_{t+1}^\star - \Omega_t), X - X_{t+1}^\star\rangle, \forall X \in \mathcal{X}_0,
\end{aligned}$$

where the first equality uses the fact that $\langle \mathcal{A}^*(y_t^\star), X - X_{t+1}^\star\rangle = 0$, the second equality uses the optimality condition (21) and the normal cone of an affine set $\mathcal{X}_0$ is $\mathcal{N}_{\mathcal{X}_0}(X_{t+1}^\star) = \{\mathcal{A}^*(y) \mid y \in \mathbb{R}^m\}$. Hence, we have proved $\alpha(X_{t+1}^\star - \Omega_t) \in \partial\hat{G}(X_{t+1}^\star) + \mathcal{N}_{\mathcal{X}_0}(X_{t+1}^\star)$ and the third property (5c) is verified. Finally, we show that $G(X)$ is $(\|C\|_{\mathrm{F}} + \rho p)$-Lipschitz continuous. In particular,

$$\begin{aligned}
&|G(Y) - G(X)| \\
\leq &|\langle C, Y - X\rangle| + \rho \sum_{i=1}^{p} \big| \max\{0, \lambda_{\max}\left(\mathcal{E}_{\mathcal{C}_k}(-Y)\right)\} \\
&- \max\{0, \lambda_{\max}\left(\mathcal{E}_{\mathcal{C}_k}(-X)\right)\} \big|, \forall Y, X \in \mathbb{S}^n.
\end{aligned}$$

The first term can be bounded by Cauchy Schwarz inequality

$$|\langle C, Y - X\rangle| \leq \|C\|_{\mathrm{F}}\|Y - X\|_{\mathrm{F}}.$$

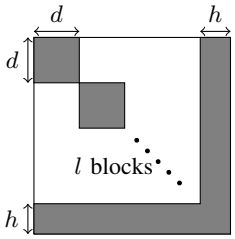

Fig. 1: Block-arrow sparsity pattern [10]: the number of blocks: l; block size: d; the width of the arrow head: h.

The second term can be bounded using the fact that the function $\max\{0, x\}$ is 1-Lipschitz and $\lambda_{\max}(Y - X) \leq \|Y - X\|_2, \forall X, Y \in \mathbb{R}^{n\times n}$. It follows that $\forall 1 \leq k \leq p$,

$$\begin{aligned}
&\max\{0, \lambda_{\max}\left(\mathcal{E}_{\mathcal{C}_k}(-Y)\right)\} - \max\{0, \lambda_{\max}\left(\mathcal{E}_{\mathcal{C}_k}(-X)\right)\} \\
\leq&|\lambda_{\max}\left(\mathcal{E}_{\mathcal{C}_k}(-Y + X)\right)| \\
\leq&\|\mathcal{E}_{\mathcal{C}_k}(-Y + X)\|_2 \\
\leq&\| - Y + X\|_2 \\
\leq&\|Y - X\|_{\mathrm{F}}.
\end{aligned}$$

Combining the upper bounds leads to

$$|G(Y) - G(X)| \leq (\|C\|_{\mathrm{F}} + \rho p)\, \|Y - X\|_{\mathrm{F}}.$$

By the results in [22], the convergence rate is $\mathcal{O}(1/\epsilon^3)$ when the objective function is Lipschitz continuous and improved to $\mathcal{O}(1/\epsilon)$ when quadratic growth Lemma 2 is satisfied.

V. Implementation and Numerical Results

In this section, we present the numerical results of Algorithm 1 to show the efficiency and convergence of the algorithm. All the experiments were executed in MATLAB R2022b on Windows PC with 3.61 GHz speed and 32.0 GB RAM. Similar to [6], [10], we consider SDPs with a block-arrow sparsity pattern and overlapping maximal cliques. The block-arrow sparsity pattern is shown in Figure 1 which has l maximal cliques of size $d+h$. The problem data are randomly generated which satisfy the strict complementarity property and have at least one high-rank primal optimal solution[1].

For the implementation, we reformulate the master problem (20) in Algorithm 1 into a quadratic SDP of the form,

$$\begin{aligned}
\min_{v\in\mathbb{R}^{p+pr^2}} \quad & v^{\mathsf{T}}Qv + q^{\mathsf{T}}v + c \\
\text{subject to} \quad & \gamma_k \geq 0, S_k \in \mathbb{S}^r_+, \\
& \gamma_k + \mathbf{tr}(S_k) \leq \rho, k = 1, \ldots, p,
\end{aligned}$$

where $v = \begin{bmatrix}\gamma_1 & \cdots & \gamma_p & \mathrm{vec}(S_1)^{\mathsf{T}} & \cdots & \mathrm{vec}(S_p)^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$,and $\mathrm{vec}(\cdot)$ denotes the vectorization operation, then solve it using MOSEK [7] (See our technical report [20] for the detailed construction of matrix Q and q). We note that the above problem only involves p nonnegative variable and p small PSD variables, which can be efficiently solved.

We run Algorithm 1 for two settings:

[1]See our code and experiments at https://github.com/hsmmoj/A-Spectral-Bundle-Method-for-Sparse-SDPs.

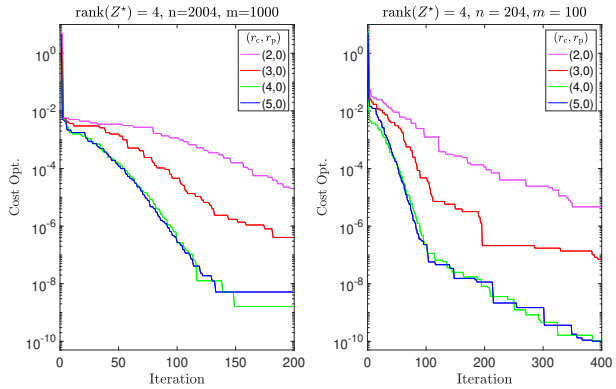

Fig. 2: The relative optimality gap of different choices of $(r_{\mathrm{c}}, r_{\mathrm{p}})$ for two random SDPs with $\mathrm{rank}(X^\star) = 2000$ and $\mathrm{rank}(X^\star) = 200$ on the left and right respectively.

TABLE I: Computational results on solving two random SDPs with block-arrow sparsity pattern and primal solutions being high-rank. Dimension of the data is calculated by $n = d\times l+h$. The number of iterations for the small problem and the large problem is 400 and 200 iterations respectively.

Dimension	$(r_{\mathrm{p}}, r_{\mathrm{c}})$	Semi Opt.	Affine Opt.	Dual Gap	Cost Opt.
d=20	$(0, 2)$	$-0.63e{-2}$	$1.06e{-7}$	$2.08e{-6}$	$4.64e{-6}$
l=10	$(0, 3)$	$-3.67e{-4}$	$1.14e{-9}$	$2.36e{-7}$	$7.05e{-8}$
h=4	$(0, 4)$	$-3.47e{-8}$	$2.45e{-11}$	$4.54e{-9}$	$1.04e{-10}$
m=100	$(0, 5)$	$-1.01e{-10}$	$4.74e{-12}$	$1.61e{-8}$	$1.01e{-10}$
d=50	$(0, 2)$	$-0.292e{-1}$	$2.05e{-8}$	$9.07e{-8}$	$2.02e{-5}$
l=40	$(0, 3)$	$-0.11e{-2}$	$3.83e{-10}$	$1.76e{-8}$	$4.03e{-7}$
h=4	$(0, 4)$	$-2.25e{-5}$	$6.30e{-11}$	$9.16e{-11}$	$1.61e{-9}$
m=1000	$(0, 5)$	$-1.03e{-5}$	$4.17e{-11}$	$9.67e{-10}$	$5.15e{-9}$

1) A small-scale problem with dimensions $d = 20, l = 10, h = 4,$ and $m = 100$.
2) A large-scale problem with dimensions $d = 50, l = 40, h = 4,$ and $m = 1000$.

Inspired by [18, Section 5], we choose different configurations of the parameters r_{p} and r_{c}. The parameter r_{p} is fixed to be 0, while different r_{c} is considered since r_{p} does not have much influence on the convergence rate both theoretically and numerically as shown in [18, Section 5]. The numerical results are presented in Figure 2 and Table I. In Table I, "Semi Opt.", "Affine Opt.", "Dual Gap", and "Cost Opt." denote the following optimality criteria

$$\begin{aligned}
&\lambda_{\min}\left(\Omega_{t+1}\right), \quad \frac{\|C - \mathcal{A}^*(y) - W_t^\star\|_{\mathrm{F}}}{1 + \|C\|}, \\
&\frac{|\langle C, \Omega_{t+1}\rangle - \langle b, \omega_t\rangle|}{1 + |\langle C, \Omega_{t+1}\rangle| + \langle b, \omega_t\rangle\ |}, \frac{G\left(\Omega_{t+1}\right) - G^\star}{G^\star},
\end{aligned}$$

where $G^\star$ denotes the true optimal value. In both small-scale and large-scale problems, Algorithm 1 converges to the optimal solution set in terms of objective value, primal feasibility, dual feasibility, and duality gap within a moderate number of iterations. In addition, we empirically observe that there is a major improvement in the accuracy when

the parameter r_c is chosen as large as $\operatorname{rank}(Z^\star)$ in (2) for two different problems, which coincides with the analogous theoretical guarantees in [18].

VI. Conclusions

In this paper, we develop a new variant of the spectral bundle method for sparse SDPs. This approach breaks down a large PSD constraint into several smaller ones by performing chordal decomposition. We introduce an equivalent non-smooth convex optimization problem by taking the constraints into the objective function. Instead of introducing additional consensus variables as the previous work [6], [10], this framework solves the non-smooth problem using a new spectral bundle method in Algorithm 1. Under a mild condition, the algorithm converges as $\mathcal{O}\left(1/\epsilon^3\right)$. If the problem satisfies strict complementarity, the convergence rate is improved to $\mathcal{O}\left(1/\epsilon\right)$. This novel algorithm opens up a new way to handle scalability issues in solving large-scale SDPs. Ongoing research directions include performing more numerical comparisons, examining additional problems with the aforementioned assumptions, and investigating convergence behaviors when the master problem is solved inexactly.

References

[1] F. Alizadeh, "Interior point methods in semidefinite programming with applications to combinatorial optimization," *SIAM Journal on Optimization*, vol. 5, no. 1, pp. 13–51, 1995.

[2] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. SIAM, 1994.

[3] A. Majumdar, G. Hall, and A. Ahmadi, "Recent scalability improvements for semidefinite programming with applications in machine learning, control, and robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 331–360, May 2020.

[4] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *Journal of the ACM (JACM)*, vol. 42, no. 6, pp. 1115–1145, 1995.

[5] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, 1996.

[6] Y. Zheng, G. Fantuzzi, and A. Papachristodoulou, "Chordal and factor-width decompositions for scalable semidefinite and polynomial optimization," *Annual Reviews in Control*, vol. 52, pp. 243–279, 2021.

[7] M. ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*, 2019.

[8] Y. Zheng, *Chordal sparsity in control and optimization of large-scale systems*. PhD thesis, University of Oxford, 2019.

[9] B. O'donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," *J. Optim. Theory Appl.*, vol. 169, p. 1042–1068, jun 2016.

[10] Y. Zheng, G. Fantuzzi, A. Papachristodoulou, P. Goulart, and A. Wynn, "Chordal decomposition in operator-splitting methods for sparse semidefinite programs," *Mathematical Programming*, vol. 180, no. 1-2, pp. 489–532, 2020.

[11] A. A. Ahmadi and A. Majumdar, "Dsos and sdsos optimization: more tractable alternatives to sum of squares and semidefinite optimization," *SIAM Journal on Applied Algebra and Geometry*, vol. 3, no. 2, pp. 193–230, 2019.

[12] J. Miller, Y. Zheng, M. Sznaier, and A. Papachristodoulou, "Decomposed structured subsets for semidefinite and sum-of-squares optimization," *Automatica*, vol. 137, p. 110125, 2022.

[13] J. Agler, W. Helton, S. McCullough, and L. Rodman, "Positive semidefinite matrices with a given sparsity pattern," *Linear Algebra and its Applications*, vol. 107, pp. 101–149, 1988.

[14] M. Fukuda, M. Kojima, K. Murota, and K. Nakata, "Exploiting sparsity in semidefinite programming via matrix completion i: General framework," *SIAM Journal on Optimization*, vol. 11, no. 3, pp. 647–674, 2001.

[15] L. Vandenberghe and M. S. Andersen, "Chordal graphs and semidefinite optimization," *Foundations and Trends® in Optimization*, vol. 1, no. 4, pp. 241–433, 2015.

[16] Y. Zheng and G. Fantuzzi, "Sum-of-squares chordal decomposition of polynomial matrix inequalities," *Mathematical Programming*, vol. 197, no. 1, pp. 71–108, 2023.

[17] C. Helmberg and F. Rendl, "A spectral bundle method for semidefinite programming," *SIAM Journal on Optimization*, vol. 10, no. 3, pp. 673–696, 2000.

[18] L. Ding and B. Grimmer, "Revisit of spectral bundle methods: Primal-dual (sub) linear convergence rates," *arXiv preprint arXiv:2008.07067*, 2020.

[19] C. Lemaréchal and J. Zowe, *A Condensed Introduction to Bundle Methods in Nonsmooth Optimization*, pp. 357–382. Dordrecht: Springer Netherlands, 1994.

[20] M. Hesam, F.-Y. Liao, and Y. Zheng, "A spectral bundle method for sparse semidefinite programs." Technical report https://github.com/hsmmoj/A-Spectral-Bundle-Method-for-Sparse-SDPs/raw/main/Spectral_Bundle_Method_for_Decomposed_SDPs.pdf, 2023.

[21] R. E. Tarjan and M. Yannakakis, "Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs," *SIAM Journal on computing*, vol. 13, no. 3, pp. 566–579, 1984.

[22] M. Díaz and B. Grimmer, "Optimal convergence rates for the proximal bundle method," *arXiv preprint arXiv:2105.07874*, 2021.

[23] A. Ruszczynski, *Nonlinear Optimization*. Princeton University Press, 2011.

[24] Y. Sun, M. S. Andersen, and L. Vandenberghe, "Decomposition in conic optimization with partially separable structure," *SIAM Journal on Optimization*, vol. 24, no. 2, pp. 873–897, 2014.

[25] R. E. Tarjan and M. Yannakakis, "Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs," *SIAM Journal on Computing*, vol. 13, no. 3, pp. 566–579, 1984.

[26] K. C. Kiwiel, "Efficiency of proximal bundle methods," *Journal of Optimization Theory and Applications*, vol. 104, no. 3, pp. 589–603, 2000.