Math. Program., Ser. B (2007) 109:211–237
DOI 10.1007/s10107-006-0031-2

FULL LENGTH PAPER

Large-scale semidefinite programming via a saddle point Mirror-Prox algorithm

Zhaosong Lu · Arkadi Nemirovski · Renato D. C. Monteiro

Received: 11 November 2004 / Accepted: 25 January 2006 /
Published online: 24 November 2006

Abstract In this paper, we first demonstrate that positive semidefiniteness of a large well-structured sparse symmetric matrix can be represented via positive semidefiniteness of a bunch of smaller matrices linked, in a linear fashion, to the matrix. We derive also the "dual counterpart" of the outlined representation, which expresses the possibility of positive semidefinite completion of a well-structured partially defined symmetric matrix in terms of positive semidefiniteness of a specific bunch of fully defined submatrices of the matrix. Using the representations, we then reformulate well-structured large-scale semidefinite problems into smooth convex–concave saddle point problems, which can be solved by a Prox-method developed in [6] with efficiency $\mathcal{O}(\epsilon^{-1})$. Implementations and some numerical results for large-scale Lovász capacity and MAXCUT problems are finally presented.

Keywords Semidefinite programming · Saddle point problem · Mirror-Prox method

Mathematics Subject Classification (2000) 90C22 · 90C25 · 90C51 · 65K10

Z. Lu (✉)
Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 156, Canada
e-mail: zhaosong@sfu.ca

A. Nemirovski · R. D. C. Monteiro
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
e-mail: nemirovs@isye.gatech.edu

R. D. C. Monteiro
e-mail: monteiro@isye.gatech.edu

1 Introduction

Consider a semidefinite program

$$\min_x \{\mathrm{Tr}(cx) : x \in \mathcal{N} \cap \mathbf{S}_+\}, \tag{1}$$

where $\mathbf{S}_+$ is the cone of positive semidefinite matrices in the space $\mathbf{S}$ of symmetric block-diagonal matrices with a given block-diagonal structure, $\mathcal{N}$ is an affine subspace in $\mathbf{S}$ and $c \in \mathbf{S}$. The goal of this paper is to investigate the possibility of utilizing favorable sparsity patterns of a large-scale problem (1) (that is, the sparsity pattern of diagonal blocks in matrices from $\mathcal{N}$) when solving the problem by a simple first-order method. To motivate our goal, let us start with discussing whether it makes sense to solve (1) by first-order methods, given the breakthrough developments in the theory and implementation of interior point methods (IPMs) for semidefinite programming (SDP) we have witnessed during the last decade. Indeed, IPMs are polynomial-time methods and as such allow to solve SDPs within accuracy ϵ at a low iteration count (proportional to $\ln(1/\epsilon)$) and thus capable of producing high-accuracy solutions. Note, however, that IPMs are Newton-type methods, with an iteration which requires assembling and solving a Newton system of n linear equations with n unknowns, where $n = \min\left[\dim\mathcal{N}, \mathrm{codim}\,\mathcal{N}\right]$ is the minimum of the design dimensions of the problem and its dual. Typically, the Newton system is dense, so that the cost of solving it by standard linear algebra techniques is $O(n^3)$ arithmetic operations. It follows that in reality the scope of IPMs in SDP is restricted to problems with n at most few thousands – otherwise a single iteration will "last forever". At the present level of our knowledge, the only way to process numerically SDPs with n of order of 10^4 or more seems to use simple first-order optimization techniques with computationally cheap iterations. Although all known first-order methods in the large-scale case exhibit slow – sublinear – convergence and thus are unable to produce high-accuracy solutions in realistic time, medium-accuracy solutions are still achievable. Historically, the first SDP algorithm of the latter type was the *spectral bundle* method [5] – a version of the well-known bundle method for nonsmooth convex minimization "tailored" to semidefinite problems. A strong point of the present method is in its modest requirements on our abilities to handle matrices from $\mathcal{N}$ – all we need is to compute few largest eigenvalues and associated eigenvectors of such matrices. This task can be carried out routinely when the largest size ζ of diagonal blocks in matrices from $\mathbf{S}$ is not too large, say, $\zeta \leq 1,000$. Note that under this limitation, n still can be of order of 10^5, meaning that (1) is far beyond the scope of IPMs. Moreover, the task in question still can be carried out when ζ is much larger than the above limit, provided that diagonal blocks in the matrices $A \in \mathcal{N}$ possess favorable sparsity patterns. A weak point of the spectral bundle method, at least from the theoretical viewpoint, is the convergence rate: the inaccuracy in terms of the objective can decrease with the iteration count t as slowly as $O(t^{-1/2})$ (this is the best possible, in the large-scale case, rate of convergence of first-order methods on nonsmooth convex programs). Also, theoretical convergence rate results are not established for the first-order SDP algorithms proposed recently in [1,2]. Recently, novel $O(t^{-1})$-converging first-order algorithms, based on smooth saddle-point reformulation of nonsmooth convex programs were developed [6–8]. Numerical results presented in these papers (including those on genuine SDP with n as large as 100,000–190,000 [6]) demonstrate high computational potential of the proposed methods. However, theoretical and computational advantages exhibited by the $O(t^{-1})$-converging methods as compared to algorithms like spectral bundle have their price, specifically, the necessity to operate with eigenvalue decompositions of the matrices from $\mathbf{S}$ rather than computing a few largest eigenvalues of matrices from $\mathcal{N}$. As a result, the algorithms from [6–8] as applied to (1) become impractical, when the largest size ζ of diagonal blocks in the matrices from $\mathbf{S}$ exceeds about 1,000.

The goal of this paper is to demonstrate that one can extend the scope of $O(t^{-1})$-converging first-order methods as applied to semidefinite program (1) beyond the just outlined limits by assuming that diagonal blocks in the matrices from $\mathcal{N}$ possess favorable sparsity patterns. This type of semidefinite program (1) has also been studied in [3] via matrix completion in the context of IPM. The outline of the paper is as follows. In Sect. 2, we explain what a "favorable sparsity pattern" is and introduce some notation and definitions which will be used throughout the paper. In Sect. 3, we develop our main tool, specifically, demonstrate that positive semidefiniteness of a large symmetric matrix A possessing a favorable sparsity pattern can be represented via positive semidefiniteness of a bunch of smaller matrices linked, in a linear fashion, to A. We derive also the "dual counterpart" of the outlined representation, which expresses the possibility of positive semidefinite completion of a "well-structured" partially defined symmetric matrix in terms of positive semidefiniteness of a specific bunch of fully defined submatrices of the matrix.[1] In Sect. 4 we utilize the aforementioned representations to derive saddle point formulations of some large-scale SDP problems, specifically, those of computing Lovász capacity of a graph and the MAXCUT problem, with emphasis on the case when the incidence matrix of the underlying graph possesses a favorable sparsity pattern. We demonstrate that the complexity of solving these problems within a fixed relative accuracy by an appropriate $O(t^{-1})$-converging first-order method (namely, the Mirror-Prox algorithm from [6]) is by orders of magnitude less than complexity associated with IPMs, and show that with our approach, we indeed can utilize a favorable sparsity pattern in the incidence matrix. In concluding Sect. 5, we illustrate our constructions by numerical results for the MAXCUT and Lovász capacity problems on well-structured sparse graphs.

[1] This result, which we get "for free", can be also obtained from general results of [4] on existence of positive semidefinite completions.

2 Well-structured sparse symmetric matrices

In this section, we motivate and define the notion of a symmetric matrix with "favorable sparsity pattern" and introduce notation to be used throughout the paper.

Motivation. To get an idea what a "favorable sparsity pattern" might be, consider the semidefinite program (1), and let A^ℓ, $\ell = 1, \dots, L$, be the diagonal blocks of a generic matrix from $\mathcal{N}$. Assume that these blocks possess certain sparsity patterns. How could we utilize this sparsity? Our first observation is that even high sparsity by itself can be of no use. Indeed, consider the simplest SDP-related computational issue, that is, checking whether a symmetric $n \times n$ matrix A is positive semidefinite. Assuming that we are checking positive semidefiniteness of a sparse symmetric matrix A by applying Cholesky factorization algorithm, that is, by trying to represent A as UU^{T} with upper triangular U, the nonzeros in U will, generically, be the entries i, j with $i \le j \le i + v_i$, where $v_i = 0$ when $A_{ii} = 0$ and $i + v_i = \max\{j : A_{ij} \neq 0\}$ otherwise. In other words, when adding to the original pattern of nonzero entries of A all entries i, j with $i \le j \le i + v_i$ (and all symmetric entries), we do not alter the fill in of the Cholesky factor. Therefore, we do not lose much by assuming that the original pattern of nonzeros already was comprised of all super-diagonal entries (i, j) with $i \le j \le i + v_i$, with added symmetric entries (by performing reordering some rows/columns of the matrix if necessary).

We arrive at the notion of a *well-structured* sparse $n \times n$ symmetric matrix with sparsity pattern given by a nonnegative integral vector v_i such that $i + v_i \le n$ for all i; the "hard zero" super-diagonal entries i, j $(i \le j)$ in such a matrix are those with $j > i + v_i$. Note that for such a matrix A, the "hard zeros" in the *upper triangular* factor U of the Cholesky factorization $A = UU^{\mathrm{T}}$ are exactly the same as hard zeros in the upper triangular part of A. In particular, if A is a well-structured sparse symmetric matrix with $\sum_i v_i \ll n^2$, then it is relatively easy to check whether or not $A \succeq 0$; to this end, it suffices to apply to A the Cholesky factorization algorithm (where the factorization being sought is $A = UU^{\mathrm{T}}$ with upper triangular U).

Next, we introduce terminology and notation for dealing with "well-structured", in the sense we have just motivated, sparsity patterns.

Simple sparsity structures and associated entities. Let $v \in \mathbf{R}^n$ be a *simple sparsity structure* – a nonnegative integral vector such that $i + v_i \le n$ for all $i \le n$. We associate with structure v the following entities:

1. A subspace $\mathbf{S}^{(v)}$ in the space $\mathbf{S}^n$ of symmetric $n \times n$ matrices; $\mathbf{S}^{(v)}$ is comprised of all matrices $[A_{ij}]_{i,j=1}^n$ from $\mathbf{S}^n$ such that $A_{ij} = 0$ for $j > i + v_i$.
2. The set $I = \{i_1 < i_2 < \cdots < i_m\}$ of all integers representable as $i + v_i$ with $i \le n$. Note that $i_m = n$, since $n + v_n = n$ (recall that $i + v_i \le n$ and $v_i \ge 0$). We refer to m as the *number of blocks* in v.
3. The sets

$$J_k = \{i \le i_k : i + v_i \ge i_k\}, \quad J'_k = \{i \in J_k : i \le i_{k-1}\}, \;\; k = 1, \dots, m, \tag{2}$$

where $i_0 = 0$ (that is, $J_1' = \emptyset$). Note that $J_k \backslash J_{k-1} = \{i_{k-1}+1, \ldots, i_k\}$ and that $J_k' = J_{k-1} \cap J_k$, where $J_0 = \emptyset$.

4. The set of *occupied* cells ij – those with $i \leq j \leq i + v_i$. For an occupied cell ij, both integers $i + v_i$ and $j + v_j$ are elements of the set $I = \{i_1, \ldots, i_m\}$; thus, $\min[i + v_i, j + v_j] = i_{k_+}$ for certain $k_+ = k_+(i,j) \leq m$. Since $j \leq i + v_i$, we have $j \leq \min[i + v_i, j + v_j] = i_{k_+}$. Therefore, the smallest k, let it be called $k_- = k_-(i,j)$, such that $j \leq i_k$, satisfies $k_- \leq k_+$. Since $j + v_j$ is one of i_s, we conclude that $j + v_j \geq i_{k_-}$. Note that the segment $D_{ij} = \{k_-, k_- + 1, \ldots, k_+\}$ is exactly the segment of those k for which i and j belong to J_k; we denote by $\ell(i,j)$ the cardinality of D_{ij}.
5. Two diagonal matrices $\mathcal{L}$ and $\mathcal{K}$ defined as

$$\mathcal{L} = \mathrm{Diag}\{\ell(1,1)^{-1/2}, \ldots, \ell(n,n)^{-1/2}\}, \qquad \mathcal{K} = \mathrm{Diag}\{\ell(1,1), \ldots, \ell(n,n)\}. \tag{3}$$

We now provide an example to illustrate the definitions given above. Consider a subspace of $\mathbf{S}^7$, consisting of all symmetric matrices with nonzero entries specified as below

$$\begin{bmatrix} * & * & * & * & & & \\ * & * & * & & & & \\ * & * & * & * & * & * & \\ * & & * & * & * & & \\ & & * & * & * & * & * \\ & & * & & * & * & * \\ & & & & * & * & * \end{bmatrix} \in \mathbf{S}^7.$$

We observe that the subspace defined above is $\mathbf{S}^{(v)}$ with $v = (3,1,3,1,2,1,0)^{\mathrm{T}}$. We easily see that $m = 5$, and $i_1 = 3, i_2 = 4, i_3 = 5, i_4 = 6$ and $i_5 = 7$, and hence $I = \{3,4,5,6,7\}$. Using (2), we have

$$\begin{gathered} J_1 = \{1,2,3\},\ J_2 = \{1,3,4\},\ J_3 = \{3,4,5\},\ J_4 = \{3,5,6\},\ J_5 = \{5,6,7\}, \\ J_1' = \emptyset,\ J_2' = \{1,3\},\ J_3' = \{3,4\},\ J_4' = \{3,5\},\ J_5' = \{5,6\}. \end{gathered}$$

Using the definition of D_{ij}, we have

$$\begin{gathered} D_{11} = \{1,2\},\ D_{22} = \{1\},\ D_{33} = \{1,2,3,4\},\ D_{44} = \{2,3\}, \\ D_{55} = \{3,4,5\},\ D_{66} = \{4,5\},\ D_{77} = \{5\}, \end{gathered}$$

and hence,

$$\begin{aligned} &\ell(1,1) = 2,\ \ell(2,2) = 1,\ \ell(3,3) = 4,\ \ell(4,4) = 2,\ \ell(5,5) = 3, \\ &\ell(6,6) = 2,\ \ell(7,7) = 1. \end{aligned}$$

Therefore, we obtain that

$$\mathcal{L} = \mathrm{Diag}\left\{\frac{1}{\sqrt{2}},\ 1,\ \frac{1}{2},\ \frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{2}},\ 1\right\},\ \ \mathcal{K} = \mathrm{Diag}\{2,\ 1,\ 4,\ 2,\ 3,\ 2,\ 1\}.$$

Finally, in the sequel $\lambda_{\min}(A)$ (resp., $\lambda_{\max}(A)$) denotes the minimal (resp., maximal) eigenvalue of a symmetric matrix A. δ^i_j denotes the Kronecker delta. For a finite set $\mathcal{J}$, we denote its cardinality by $|\mathcal{J}|$.

3 Representation results for well-structured sparse symmetric matrices

Consider again the semidefinite program (1). Assuming that the diagonal blocks A^ℓ in a generic matrix $A \in \mathcal{N}$ to be sparse, with well-structured sparsity pattern as defined in Sect. 2, it is relatively easy to verify whether the Linear Matrix Inequalities (LMIs) are satisfied at a given point (since the Cholesky factorization $A^\ell = U_\ell U_\ell^{\mathrm{T}}$ with upper triangular U_ℓ does not increase fill in). This possibility, however, in many respects is not sufficient. When solving SDPs by numerous advanced methods, including interior point ones, we would prefer to deal with many small dense LMIs rather than with few large sparse ones, at least in the case when the total row size of the former system of LMIs is of the same order of magnitude as the total row size of the latter system. In this respect, the following question is of definite interest:

> Given a well-structured sparse matrix A, is it possible to express the fact that $A \succeq 0$ by a system of relatively small LMIs in variables A_{ij} and perhaps additional variables?

We are about to give an affirmative answer to this question.

3.1 Positive semidefiniteness of well-structured sparse matrices

In this subsection, we will provide some necessary and sufficient conditions for a matrix from $\mathbf{S}^{(v)}$ to be positive semidefinite. The following notations will be used throughout the remaining paper.

Notation. Let $J \subset \{1, \ldots, n\}$ be an index set with $\ell > 0$ elements. We denote by $[B_{ij}]_{i,j\in J}$ the $\ell \times \ell$ matrix obtained from B by extracting the rows and columns with indices in J, and by $]B_{ij}[_{i,j\in J}$ – the $n \times n$ matrix with entries B_{ij} for all $i, j \in J$ and zero entries for the remaining pairs i, j.

The following notations will be used in this subsection and Subsect. 4.2 of this paper.

Let $v \in \mathbf{R}^n$ be a simple sparsity structure, and J_k, $k = 1, \ldots, m$, be the corresponding index sets (see Sect. 2). We define $\mathbf{B}$ as an Euclidean space comprised of collections $B = \left\{B_k = \left[B^k_{ij} = B^k_{ji}\right]_{i,j\in J_k}\right\}_{k=1}^{m}$ of symmetric matrices, i.e.,

$$\mathbf{B} \equiv \left\{ B = (B_1, \ldots, B_m) : B_k = \left[B^k_{ij} = B^k_{ji} \right]_{i,j \in J_k}, \; k = 1, \ldots, m \right\},$$

and equipped with natural linear operations and the norm

$$\|B\|_F = \sqrt{\sum_{k=1}^{m} \|B_k\|_F^2},$$

where $\|B_k\|_F$ is the Frobenius norm of B_k. For $B = \{B_k = [B^k_{ij} = B^k_{ji}]_{i,j \in J_k}\}_{k=1}^m \in$ $\mathbf{B}$, we set

$$B^k =]B^k_{ij}[_{i,j \in J_k} \in \mathbf{S}^{(\nu)}, \; k = 1, \ldots, m.$$

and define the linear mapping $\mathcal{S}(B) : \mathbf{B} \to \mathbf{S}^{(\nu)}$ as

$$\mathcal{S}(B) = \sum_{k=1}^{m} B^k.$$

Proposition 1 (i) *A matrix* $A \in \mathbf{S}^{(\nu)}$ *is* $\succeq 0$ *if and only if there exists* $B = \{B_k = [B^k_{ij} = B^k_{ji}]_{i,j \in J_k} \succeq 0\}_{k=1}^m \in \mathbf{B}$ *such that*

$$A = \mathcal{S}(B) \equiv \sum_{k=1}^{m} B^k. \tag{4}$$

(ii) *Whenever* $B = \{B_k = [B^k_{ij}]_{i,j \in J_k} \succeq 0\}_{k=1}^m$ *satisfies* (4)*, one has, for any* $n \times n$ *real matrix* W*,*

$$\sum_{k=1}^{m} \|W^{\mathrm{T}} B^k W\|_F^2 \le \|W^{\mathrm{T}} A W\|_F^2. \tag{5}$$

(iii) *We have*

$$\forall B \in \mathbf{B} : \|\mathcal{L}^{1/2} \mathcal{S}(B) \mathcal{L}^{1/2}\|_F \le \|B\|_F, \tag{6}$$

where $\mathcal{L}$ *is given by* (3).

Illustration: Overlapping block-diagonal structure. Before proving Proposition 1, it makes sense to "visualize" its simplest "overlapping block-diagonal" version. Consider a symmetric block-matrix of the form

$$A = \begin{bmatrix} * & * & * & & & & \\ * & * & * & & & & \\ * & * & * & * & * & & \\ & & * & * & * & * & * \\ & & * & * & * & * & * \\ & & & * & * & * & * \\ & & & * & * & * & * \end{bmatrix}$$

where $*$ marks nonzero blocks. Proposition 1.(i) says that such a matrix is positive semidefinite if and only if it is the sum of positive semidefinite matrices of the form

$$\begin{bmatrix} \boxed{\begin{matrix}*&*&*\\ *&*&*\\ *&*&*\end{matrix}} & & \\ & & \\ & & \end{bmatrix}, \begin{bmatrix} & & \\ & \boxed{\begin{matrix}*&*&*\\ *&*&*\\ *&*&*\end{matrix}} & \\ & & \end{bmatrix}, \begin{bmatrix} & & \\ & & \\ & & \boxed{\begin{matrix}*&*&*&*\\ *&*&*&*\\ *&*&*&*\\ *&*&*&*\end{matrix}} \end{bmatrix},$$

and similarly when the number of overlapping diagonal blocks is > 3.

Proof of Proposition 1. (i) Induction in m. For $m = 1$ the statement is evident. Assuming that the statement is valid for $m - 1$, let us prove it for m. The "if" part is evident; thus, assume that $A \in \mathbf{S}^{(v)}$ is $\succeq 0$, and let us prove the existence of the required B_k. Let $A = \left[\begin{array}{c|c} P & Q \\ \hline Q^{\mathrm{T}} & R \end{array}\right]$ with $i_{m-1} \times i_{m-1}$ block P. For $\epsilon \geq 0$, let $A_\epsilon = \left[\begin{array}{c|c} P & Q \\ \hline Q^{\mathrm{T}} & R + \epsilon I \end{array}\right]$ and $B^\epsilon = \left[\begin{array}{c|c} Q(R+\epsilon I)^{-1}Q^{\mathrm{T}} & Q \\ \hline Q^{\mathrm{T}} & R + \epsilon I \end{array}\right]$. We easily see that $B^\epsilon \succeq 0$. By the Schur Complement Lemma, we have $A_\epsilon - B^\epsilon \succeq 0$, thus, B^ϵ remains bounded as $\epsilon \to +0$. Thus, we can find a sequence $\epsilon_t \to +0$, $t \to \infty$, and a matrix B^m such that

$$B^m = \lim_{t\to\infty} B^{\epsilon_t}. \tag{7}$$

Observe that both B^m and $A - B^m$ are $\succeq 0$. By construction, $B^m =]B^m_{ij}[_{i,j\in J_m}$; besides this, the rows i and the columns j in $C = A - B^m$ with $i, j > i_{m-1}$ are zero. Removing these rows and columns, we get an $i_{m-1} \times i_{m-1}$ matrix $\bar{C} \in \mathbf{S}^{(v')}$, where $v'_i = \min[i_{m-1} - i, v_i]$, $1 \leq i \leq i_{m-1} = \dim v'$. Clearly, the number of blocks in v' is $m - 1$, and the corresponding index sets J_k, $1 \leq k \leq m - 1$, are the same as for v. Applying to $\bar{C}$ the inductive hypothesis, we can find $m - 1$ matrices $B^k =]B^k_{ij}[_{i,j\in J_k} \succeq 0$, $k = 1, \ldots, m-1$, such that $C = \sum_{k=1}^{m-1} B^k$, whence $A = C + B^m = \sum_{k=1}^{m} B^k$ with $B^k \succeq 0$ of the required structure. The induction is over.

(ii) For matrices $B, C \succeq 0$, one has $\mathrm{Tr}(BC) \geq 0$. It follows that under the premise of (ii) one has $\| \sum_k W^T B^k W\|_F^2 \geq \sum_k \|W^T B^k W\|_F^2$.

(iii) Let $A = \mathcal{S}(B)$, so that $A_{ij} = \sum_{k:i,j\in J_k} B^k_{ij}$. Recall from Sect. 2 that $\ell(i,j) = |\{k : i,j \in J_k\}|$ for every i,j. We have

$$\begin{aligned}
\|\mathcal{L}^{1/2} A \mathcal{L}^{1/2}\|_F^2 &= \sum_{i,j} A_{ij}^2 \ell^{-1/2}(i,i)\ell^{-1/2}(j,j) \\
&= \sum_{i,j} \left(\sum_{k:i,j\in J_k} B^k_{ij} \right)^2 \ell^{-1/2}(i,i)\ell^{-1/2}(j,j) \\
&\leq \sum_{i,j} \sum_{k:i,j\in J_k} (B^k_{ij})^2 \frac{\ell(i,j)}{\sqrt{\ell(i,i)\ell(j,j)}} \\
&\leq \left(\max_{i,j} \frac{\ell(i,j)}{\sqrt{\ell(i,i)\ell(j,j)}} \right) \sum_{i,j} \sum_{k:i,j\in J_k} (B^k_{ij})^2 \\
&= \left(\max_{i,j} \frac{\ell(i,j)}{\sqrt{\ell(i,i)\ell(j,j)}} \right) \|B\|_F^2;
\end{aligned}$$

thus, in order to prove (iii) it suffices to verify that

$$\ell(i,j) \leq \sqrt{\ell(i,i)\ell(j,j)}$$

for every i,j. This is evident due to

$$\ell(i,j) = |\{k : i,j \in J_k\}| \leq \min \left[|\{k : i \in J_k\}|, |\{k : j \in J_k\}| \right] = \min[\ell(i,i), \ell(j,j)].$$

$\square$

Proposition 1(i) establishes a characterization for positive semidefiniteness of matrices from $\mathbf{S}^{(\nu)}$, but it does not give the explicit formulas for the matrices $B_k = [B^k_{ij} = B^k_{ji}]_{i,j\in J_k}$. We next develop an equivalent reformulation of positive semidefiniteness of matrices from $\mathbf{S}^{(\nu)}$ by introducing some additional variables.

Lemma 1 *Let $m > 1$. A matrix $A \in \mathbf{S}^{(\nu)}$ is $\succeq 0$ if and only if there exists a matrix $\Delta^{m-1} = (\Delta^{m-1})^T = [\Delta^{m-1}_{ij}]_{i,j\in J'_m}$ such that the matrices*

$$\mathcal{B} \equiv \mathcal{B}_m(A, \Delta^{m-1}) = \left[B_{ij}\right]_{i,j\in J_m} : B_{ij} = \begin{cases} A_{ij}, & i \notin J'_m \text{ or } j \notin J'_m \\ \Delta^{m-1}_{ij}, & i,j \in J'_m \end{cases} \tag{8}$$

and

$$\mathcal{C} \equiv \mathcal{C}_m(A, \Delta^{m-1}) = \left[C_{ij}\right]_{i,j=1}^{i_{m-1}} : C_{ij} = \begin{cases} A_{ij}, & i \notin J'_m \text{ or } j \notin J'_m \\ A_{ij} - \Delta^{m-1}_{ij}, & i,j \in J'_m \end{cases} \tag{9}$$

are positive semidefinite.

Proof A is the sum of matrices obtained from $\mathcal{B}$ and $\mathcal{C}$ by adding a number of zero rows and columns; thus, if $\mathcal{B}$ and $\mathcal{C}$ are $\succeq 0$, so is A. Conversely, assuming $A \succeq 0$, let us prove that there exists Δ^{m-1} such that the corresponding matrices $\mathcal{B}, \mathcal{C}$ are $\succeq 0$. Let B^m be defined in (7). Recall from the proof of Proposition 1(i) that $B^m \succeq 0$ and $A - B^m \succeq 0$. Now, let $\Delta^{m-1} = \left[B^m_{ij}\right]_{i,j \in J'_m}$. From the construction of B^m, we see that $\mathcal{B}$ defined as in (8) satisfies $\mathcal{B} = \left[B^m_{ij}\right]_{i,j \in J_m}$, and hence $\mathcal{B} \succeq 0$. Similarly, $\mathcal{C}$ defined as in (9) is actually the North-Western $i_{m-1} \times i_{m-1}$ block in $A - B^m$, and hence $\mathcal{C} \succeq 0$. □

Observing that matrix $\mathcal{C} = \mathcal{C}_m(A, \Delta^{m-1})$ belongs to $\mathbf{S}^{(v')}$, where $v' = (v'_1, \ldots, v'_{i_{m-1}})^T$, where $v'_i = \min[v_i, i_{m-1} - i]$, $1 \leq i \leq i_{m-1}$, and applying Lemma 1 recursively, we arrive at the following result.

Theorem 1 *Let $v \in \mathbf{R}^n$ be an integral nonnegative vector such that $i + v_i \leq n$ for all i, let $I = \{i_1 < i_2 < \cdots < i_m\}$ be the image of $\{1, 2, \ldots, n\}$ under the mapping $i \mapsto i + v_i$, and let the sets J_k, J'_k be defined by* (2)*. A matrix $A \in \mathbf{S}^{(v)}$ is $\succeq 0$ if and only if this matrix can be extended, by properly chosen matrices $\Delta^k = [\Delta^k]^T = \left[\Delta^k_{ij}\right]_{i,j \in J'_{k+1}}$, $k = 1, 2, \ldots, m-1$, to a solution of an explicit system $\mathcal{S}$ of m LMIs*

$$\mathcal{B}_k(A, \Delta) \succeq 0, \; k = 1, \ldots, m$$

given by the following recurrence:

<u>Initialization:</u> *Set $k = m$, $\mathcal{C}^m = A$.* <u>Step k, $m \geq k \geq 1$:</u> *Given matrix $\mathcal{C}^k \in \mathbf{S}^{(v^k)}$, with $v^k_i = \min[i_k - i, v_i]$, $i = 1, 2, \ldots, i_k$, set*

$$\mathcal{B}_k(A, \Delta) = [B^k_{ij}]_{i,j \in J_k} : B^k_{ij} = \begin{cases} \mathcal{C}^k_{ij}, & i \in J_k \backslash J'_k \quad or\, j \in J_k \backslash J'_k \\ \Delta^{k-1}_{ij}, & i, j \in J'_k \end{cases}$$

If $k = 1$, terminate, otherwise set

$$\mathcal{C}^{k-1} = \left[\mathcal{C}^{k-1}_{ij}\right]_{i,j=1}^{i_{k-1}} : \mathcal{C}^{k-1}_{ij} = \begin{cases} \mathcal{C}^k_{ij}, & i \notin J'_k \; or\; j \notin J'_k \\ \mathcal{C}^k_{ij} - \Delta^{k-1}_{ij}, & i, j \in J'_k \end{cases}$$

replace k with $k - 1$ and loop.

From the construction of $\mathcal{B}_k \equiv \mathcal{B}_k(A, \Delta)$ above, we see that each cell ij with $i \leq j$ belongs to $\mathcal{B}_k$ exactly for all $k \in D_{ij}$ and for those k the corresponding entry ij in B^k is

$$B^k_{ij} = \begin{cases} A_{ij}, & k_-(i,j) = k = k_+(i,j) \\ A_{ij} - \sum\limits_{\nu = k_-(i,j)}^{k_+(i,j)-1} \Delta^\nu_{ij}, & k_-(i,j) = k < k_+(i,j) \\ \Delta^{k-1}_{ij}, & k_-(i,j) < k \leq k_+(i,j) \end{cases} \quad (10)$$

Note that Δ^k is the principal sub-matrix in $\mathcal{B}_{k+1}$ corresponding to $i,j \in J'_{k+1}$, and that A is the sum of matrices obtained from $\mathcal{B}_1, \ldots, \mathcal{B}_m$ by adding zero rows and columns. We arrive at the following result.

Theorem 2 *A matrix $A \in \mathbf{S}^{(\nu)}$ is $\succeq 0$ if and only if there exist matrices $\Delta^k = [\Delta^k]^T = [\Delta^k_{ij}]_{i,j \in J'_{k+1}}$, $1 \le k \le m-1$, such that the matrices $\mathcal{B}_k = \mathcal{B}_k(A, \Delta) = [B^k_{ij}]_{i,j \in J_k}$ given by* (10) *are $\succeq 0$. Whenever this is the case, one has*

$$\begin{aligned} \Delta^k &\succeq 0, \quad k = 1, \ldots, m-1 \\ \sum_{k=1}^{m-1} \mathrm{Tr}(\Delta^k) &\le \mathrm{Tr}(A). \end{aligned}$$

Let $\nu \in \mathbf{R}^n$ be a simple sparsity structure, and J_k, J'_k, $k = 1, \ldots, m$, be the corresponding index sets (see Sec. 2). We define $\boldsymbol{\Delta}$ as an Euclidean space comprised of collections $\boldsymbol{\Delta} = \{\boldsymbol{\Delta}^k = [\boldsymbol{\Delta}^k]^{\mathrm{T}} = [\boldsymbol{\Delta}^k_{ij}]_{i,j \in J'_{k+1}}\}_{k=1}^{m-1}$, i.e.,

$$\boldsymbol{\Delta} \equiv \left\{ \boldsymbol{\Delta} = \{\boldsymbol{\Delta}^k = [\boldsymbol{\Delta}^k]^T = [\boldsymbol{\Delta}^k_{ij}]_{i,j \in J'_{k+1}}\}_{k=1}^{m-1} \right\}.$$

Let $\boldsymbol{\Delta}_\rho$ be a subset of $\boldsymbol{\Delta}$ defined as

$$\boldsymbol{\Delta}_\rho = \left\{ \boldsymbol{\Delta} \in \boldsymbol{\Delta} : \boldsymbol{\Delta}^k \succeq 0,\, k = 1, \ldots, m-1, \sum_{k=1}^{m-1} \mathrm{Tr}(\boldsymbol{\Delta}^k) \le \rho \right\}. \tag{11}$$

We denote by $\mathcal{B}_k(A, \Delta) = [B^k_{ij}(A, \Delta)]_{i,j \in J_k}$ the linear matrix-valued functions of $A \in \mathbf{S}^{(\nu)}$, $\boldsymbol{\Delta} \in \Delta$ defined by (10). Finally, let

$$\lambda_{\min}(A, \Delta) = \min_{1 \le k \le m} \lambda_{\min}(\mathcal{B}_k(A, \Delta)).$$

The following proposition will be used in Sect. 4.

Proposition 2 *Let $A \in \mathbf{S}^{(\nu)}$, $\Delta \in \Delta$ be such that $\lambda_{\min}(A, \Delta) = -\lambda < 0$. Then $A \succeq -\lambda \mathcal{K}$, where $\mathcal{K}$ is given by* (3).

Proof Let $\widehat{\Delta}^k_{ij} = \begin{cases} \Delta^k_{ij}, & i \neq j \\ \Delta^k_{ij} + \lambda, & i = j \end{cases}$, and let $\widehat{A} = A + \lambda \mathcal{K}$. By (10), we have

$$i,j \in J_k \Rightarrow B^k_{ij}(\widehat{A}, \widehat{\Delta}) - B^k_{ij}(A, \Delta) = \lambda \delta^i_j,$$

whence $\mathcal{B}_k(\widehat{A}, \widehat{\Delta}) \succeq 0$, and $\widehat{A} = A + \lambda \mathcal{K} \succeq 0$. □

Sizes of $\mathcal{S}$. We have expressed positive semidefiniteness of $A \in \mathbf{S}^{(\nu)}$ as solvability of an explicit system $\mathcal{S}$ (see Theorem 1) of LMIs in matrix A and additional matrix variables Δ^k, $k = 1, \ldots, m-1$. The sizes of $\mathcal{S}$ are as follows:

1. *Number and sizes of LMIs.* $\mathcal{S}$ contains m LMIs $\mathcal{B}_k(A, \Delta) \succeq 0$ of row sizes $S_k = |J_k|$, $k = 1, \ldots, m$.
2. *Number of additional variables.* Let $d_k = i_k - i_{k-1}$, $k = 1, \ldots, m$. Clearly, step $k \geq 2$ of our construction adds $V_k = \frac{(|J_k|-d_k)(|J_k|-d_k+1)}{2}$ additional variables, and step $k = 1$ does not add new variables. Thus, the total number of additional variables is

$$V = \sum_{k=2}^{m} \frac{(|J_k| - d_k)(|J_k| - d_k + 1)}{2}.$$

Example: staircase structure. Before ending this subsection, we present an example for positive semidefinite *staircase* matrices to illustrate the result established in Theorem 2.

Let $d = (d_0, d_1, \ldots, d_\mu)$ be a *staircase structure* – collection of integers with $d_0 \geq 0$ and $d_1, \ldots, d_\mu > 0$, and let $|d| = d_0 + \cdots + d_\mu$. Let $\mathbf{S}^{[d]}$ be the subspace of *d-staircase* symmetric matrices in $\mathbf{S}^{|d|}$, which is comprised of $(\mu+1) \times (\mu+1)$ block matrices $[A_{ij}]_{i,j=0}^{\mu}$ with $d_i \times d_j$ blocks A_{ij} such that $A = A^{\mathrm{T}}$ and $A_{ij} = 0$ for $0 < i < j-1$:

$$A \in \mathbf{S}^{[d]} \Leftrightarrow A = \left[\begin{array}{c|c|c|c|c|c} A_{0,0} & A_{0,1} & A_{0,2} & \cdots & A_{0,\mu-1} & A_{0,\mu} \\ \hline A_{0,1}^{\mathrm{T}} & A_{1,1} & A_{1,2} & 0 & 0 & 0 \\ \hline A_{0,2}^{\mathrm{T}} & A_{1,2}^{\mathrm{T}} & A_{2,2} & A_{2,3} & 0 & 0 \\ \hline \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \hline A_{0,\mu-1}^{\mathrm{T}} & 0 & 0 & A_{\mu-2,\mu-1}^{\mathrm{T}} & A_{\mu-1,\mu-1} & A_{\mu-1,\mu} \\ \hline A_{0,\mu}^{\mathrm{T}} & 0 & 0 & 0 & A_{\mu-1,\mu}^{\mathrm{T}} & A_{\mu,\mu} \end{array}\right].$$

In view of the definition of simple sparsity structure, we easily see that $A \in \mathbf{S}^{[d]}$ iff $A \in \mathbf{S}^{(\nu)}$, where ν is a simple sparsity structure defined as

$$\nu_i = \begin{cases} |d| - i, & i \leq d_0 \\ \sum\limits_{j=0}^{k+1} d_j - i, & \sum\limits_{j=0}^{k-1} d_j < i \leq \sum\limits_{j=0}^{k} d_j \quad \text{for } k = 1, \ldots, \mu - 1 \\ |d| - i, & \sum\limits_{j=0}^{\mu-1} d_j < i \leq |d| \end{cases}$$

We also see that there are $m = \mu - 1$ elements $i_1 < \cdots < i_m$ in I given by $i_k = \sum\limits_{j=0}^{k+1} d_j$ for $k = 1, \cdots, m$. Using Theorem 2, we immediately have the following result.

Proposition 3 *A d-staircase matrix $A = [A_{ij}]_{i,j=0}^{\mu}$ is positive semidefinite if and only if there exists*

$$\Delta = \left\{ \Delta^j = \left[\begin{array}{c|c} \Delta^j_{0,0} & \Delta^j_{0,1} \\ \hline [\Delta^j_{0,1}]^{\mathrm{T}} & \Delta^j_{1,1} \end{array} \right] : \Delta^j_{0,0} \in \mathbf{S}^{d_0}, \Delta^{j+1}_{1,1} \in \mathbf{S}^{d_j} \right\}_{j=1}^{\mu-2}$$

such that

$$\left[\begin{array}{c|c|c} A_{0,0} - \sum\limits_{j=1}^{\mu-2} \Delta^j_{0,0} & A_{0,1} & A_{0,2} - \Delta^1_{0,1} \\ \hline A^{\mathrm{T}}_{0,1} & A_{1,1} & A_{1,2} \\ \hline A^{\mathrm{T}}_{0,2} - [\Delta^1_{0,1}]^{\mathrm{T}} & A^{\mathrm{T}}_{1,2} & A_{2,2} - \Delta^1_{2,2} \end{array} \right] \succeq 0,$$

$$\left[\begin{array}{c|c|c} \Delta^{j-1}_{0,0} & \Delta^{j-1}_{0,1} & A_{0,j+1} - \Delta^j_{0,1} \\ \hline [\Delta^{j-1}_{0,1}]^{\mathrm{T}} & \Delta^{j-1}_{1,1} & A_{j,j+1} \\ \hline A^{\mathrm{T}}_{0,j+1} - [\Delta^j_{0,1}]^{\mathrm{T}} & A^{\mathrm{T}}_{j,j+1} & A_{j+1,j+1} - \Delta^j_{1,1} \end{array} \right] \succeq 0, \quad j = 2, \ldots, \mu - 2$$

$$\left[\begin{array}{c|c|c} \Delta^{\mu-2}_{0,0} & \Delta^{\mu-2}_{0,1} & A_{0,\mu} \\ \hline [\Delta^{\mu-2}_{0,1}]^{\mathrm{T}} & \Delta^{\mu-2}_{1,1} & A_{\mu-1,\mu} \\ \hline A^{\mathrm{T}}_{0,\mu} & A^{\mathrm{T}}_{\mu-1,\mu} & A_{\mu,\mu} \end{array} \right] \succeq 0.$$

3.2 Positive semidefinite completion of matrices from $\mathbf{S}^{(\nu)}$

Let $\mathbf{C}^{(\nu)}$ be the cone of matrices Z from $\mathbf{S}^{(\nu)}$ admitting positive semidefinite completion, that is, those Z which can be made positive semidefinite by replacing "hard zero" entries ij (those with $j > i + \nu_i$ or $i > j + \nu_j$) with appropriate, perhaps nonzero, entries. Also, let $\mathbf{S}^{(\nu)}_+ = \mathbf{S}^{(\nu)} \cap \mathbf{S}^n_+$, $\mathbf{S}^{(\nu)}_\perp = \left\{Z \in \mathbf{S}^n : \mathrm{Tr}(XZ) = 0 \; \forall X \in \mathbf{S}^{(\nu)}\right\}$, and $\mathbf{C} = \left\{Z \in \mathbf{S}^{(\nu)} : \mathrm{Tr}(XZ) \geq 0 \forall X \in \mathbf{S}^{(\nu)}_+\right\}$. We claim that $\mathbf{C}^{(\nu)} = \mathbf{C}$. Indeed, we easily observe that for any $Z \in \mathbf{C}$,

$$\min\left\{\mathrm{Tr}(XZ) : X \in \mathbf{S}^{(\nu)}_+\right\} = 0. \tag{12}$$

Since the subspace $\mathbf{S}^{(\nu)}$ clearly intersects the interior of $\mathbf{S}^n_+$, the dual of (12) is solvable and hence the dual feasible region $\left\{Y \in \mathbf{S}^{(\nu)}_\perp : Z - Y \succeq 0\right\} \neq \emptyset$, which immediately implies that $Z \in \mathbf{C}^{(\nu)}$, and hence $\mathbf{C} \subseteq \mathbf{C}^{(\nu)}$. The converse $\mathbf{C}^{(\nu)} \subseteq \mathbf{C}$ is evident. Thus, $\mathbf{C}^{(\nu)} = \mathbf{C} = \left\{Z \in \mathbf{S}^{(\nu)} : \mathrm{Tr}(XZ) \geq 0 \; \forall X \in \mathbf{S}^{(\nu)}_+\right\}$, and $\mathbf{C}^{(\nu)}$ can be viewed as the dual cone of the cone $\mathbf{S}^{(\nu)}_+$ in the space $\mathbf{S}^{(\nu)}$. This together with Proposition 1 implies the following result.

Proposition 4 *A matrix* $Z = \left[Z_{ij}\right]_{i,j=1}^{n} \in \mathbf{S}^{(\nu)}$ *belongs to* $\mathbf{C}^{(\nu)}$ *if and only if all matrices* $\left[Z_{ij}\right]_{i,j\in J_k}$*,* $k = 1,2,\ldots,m$*, are* $\succeq 0$.

Proof Since $\mathbf{C} = \mathbf{C}^{(\nu)}$, we see that $Z \in \mathbf{C}^{(\nu)}$ iff $\mathrm{Tr}(XZ) \geq 0$ for any $X \in \mathbf{S}_{+}^{(\nu)}$. Invoking Proposition 1, we conclude that $Z \in \mathbf{C}^{(\nu)}$ if and only if the optimal value in the optimization problem

$$\min_{\{]X^k_{ij}[_{i,j\in J_k}\succeq 0\}_{k=1}^{m}} \left\{ \mathrm{Tr}\left(Z \sum_{k=1}^{m}]X^k_{ij}[_{i,j\in J_k} \right) \right\}$$

is ≥ 0. Clearly, this is so if and only if

$$\min_{]X^k_{ij}[_{i,j\in J_k}\succeq 0} \left\{ \mathrm{Tr}\left( ]Z^k_{ij}[_{i,j\in J_k}]X^k_{ij}[_{i,j\in J_k} \right) \right\} \geq 0, \quad k = 1,\ldots,m,$$

which implies that $]Z^k_{ij}[_{i,j\in J_k}\succeq 0$ for $k = 1,\ldots,m$ due to a well-known result that, for a real symmetric matrix A, $\min_{X\succeq 0} \mathrm{Tr}(AX) \geq 0$ if and only if $A \succeq 0$. In other words, $Z \in \mathbf{C}^{(\nu)}$ if and only if $[Z_{ij}]_{i,j\in J_k} \succeq 0$, $k = 1,\ldots,m$. □

Remark 1 Alternatively to our exposition, the result stated in Proposition 4 can be obtained directly from the necessary and sufficient conditions, found in [4], for a partially defined symmetric matrix to admit positive semidefinite completion.

Corollary 1 *For* $A \in \mathbf{S}^{(\nu)}$ *one has*

$$\lambda_{\max}(A) = \max_{Y} \left\{ \mathrm{Tr}(AY) : Y \in \mathbf{S}^{(\nu)}, \mathrm{Tr}(Y) = 1, \left[Y_{ij}\right]_{i,j\in J_k} \succeq 0, k = 1,2,\ldots,m \right\}. \tag{13}$$

Indeed, for $A \in \mathbf{S}^n$ we have $\lambda_{\max}(A) = \max_{Y} \left\{ \mathrm{Tr}(AY) : Y \in \mathbf{S}_{+}^{n}, \mathrm{Tr}(Y) = 1 \right\}$; when $A \in \mathbf{S}^{(\nu)}$, the latter formula clearly can be rewritten as $\lambda_{\max}(A) = \max_{Y} \left\{ \mathrm{Tr}(AY) : Y \in \mathbf{C}^{(\nu)}, \mathrm{Tr}(Y) = 1 \right\}$. Invoking Proposition 4, we arrive at (13).

Before ending this subsection, we give an example on positive semidefinite completion of staircase matrices from $\mathbf{S}^{[d]}$ to illustrate the result established in Proposition 4.

Proposition 5 *Let d be a staircase structure with* $\mu > 1$*, and* $\mathbf{C}^{[d]}$ *be the cone of d-staircase matrices B admitting positive semidefinite completion. Then, a matrix* $B \in \mathbf{S}^{[d]}$ *belongs to* $\mathbf{C}^{[d]}$ *if and only if*

$$\left[\begin{array}{c|c|c} B_{0,0} & B_{0,j} & B_{0,j+1} \\ \hline B_{0,j}^T & B_{j,j} & B_{j,j+1} \\ \hline B_{0,j+1}^T & B_{j,j+1}^T & B_{j+1,j+1} \end{array} \right] \succeq 0, \quad j = 1,\ldots,\mu - 1.$$

4 Using the representations

In this section, we will use the representations presented in Subsects. 3.1 and 3.2 to reformulate some large-scale SDP problems into saddle point problems. The saddle point problem reformulations for a class of SDPs, and SDP relaxations of Lovász capacity and MAXCUT problems are given in Subsects. 4.1, 4.2 and 4.3, respectively.

4.1 Semidefinite programs with well-structured sparse constraint matrices

Let ν be a simple sparsity pattern. Consider the semidefinite program

$$\mathrm{Opt} = \max_x \left\{ c^{\mathrm{T}} x : x \in X, A[x] \succeq 0 \right\}, \tag{14}$$

where X is a "simple" (see below) convex compact set in $\mathbf{R}^N$ and $A[x]$ is affine matrix-valued function on X taking values in $\mathbf{S}^{(\nu)}$.

Throughout this subsection, we make the following assumptions:

A.1. *We know a point $\bar{x} \in X$ such that $A[\bar{x}] \succ 0$;*
A.2. *We are given a finite upper bound,* Opt^{up}*, on the optimal value* Opt *in* (14)*;*
A.3. *We are given a finite upper bound, ρ, on the quantity*

$$\max_x \{\mathrm{Tr}(A[x]) : x \in X, A[x] \succeq 0\}.$$

Given a point $\bar{x}$ mentioned in A.1, let

$$\nu = \max\{t : A[\bar{x}] \succeq t\mathcal{K}\}, \tag{15}$$

where $\mathcal{K}$ is defined in (3). We start with the following simple fact (a kind of "exact penalty" statement):

Lemma 2 *Let $\mathcal{Y} = \{Y = \{Y^k = [Y^k_{ij}]_{i,j \in J_k}\}_{k=1}^m : Y^k \succeq 0, \sum_k \mathrm{Tr}(Y^k) \le 1\}$. Given $T \ge 0$, let us associate with* (14) *the saddle point problem*

$$\begin{gathered} \max_{x \in X, \Delta \in \Delta_\rho} F_T(x, \Delta) \\ F_T(x,\Delta) = \min_{Y \in \mathcal{Y}} \left[c^T x + T \sum_{k=1}^m \mathrm{Tr}(Y^k \mathcal{B}_k(A[x], \Delta)) \right] \end{gathered} \tag{16}$$

(*for the definition of $\mathbf{\Delta}_\rho$, see* (11)). *Assume that*

$$T \ge \frac{1}{\nu}(\mathrm{Opt} - c^T \bar{x}). \tag{17}$$

Let $(x_\epsilon, \Delta_\epsilon)$ *be an* ϵ*-solution to* (16)*, that is,* $x_\epsilon \in X$*,* $\Delta_\epsilon \in \Delta_\rho$ *and* $F_T(x_\epsilon, \Delta_\epsilon) \geq \max\limits_{x \in X, \Delta \in \Delta_\rho} F_T(x, \Delta) - \epsilon$*. Then, the point*

$$x^\epsilon = \frac{1}{1+\gamma} x_\epsilon + \frac{\gamma}{1+\gamma} \bar{x}, \quad \gamma = \frac{\max[0, -\lambda_{\min}(A[x_\epsilon], \Delta_\epsilon)]}{\nu},$$

is a feasible ϵ*-solution to* (14)*, i.e.,* $x^\epsilon \in X$*,* $A[x^\epsilon] \succeq 0$*, and* $c^T x^\epsilon \geq \mathrm{Opt} - \epsilon$*.*

Proof We clearly have

$$F_T(x, \Delta) = c^{\mathrm{T}} x + T \min[\lambda_{\min}(A[x], \Delta), 0].$$

Further, by Theorem 2, $A[x]$ with $x \in X$ is $\succeq 0$ if and only if $\max\limits_{\Delta \in \mathbf{\Delta}_\rho} \lambda_{\min}(A[x], \Delta) \geq 0$; thus, when x is feasible for (14), we have $\sup\limits_{\Delta \in \mathbf{\Delta}_\rho} F_T(x, \Delta) \geq c^T x$, so that the optimal value of (16) is $\geq$ Opt. Consequently, ϵ-optimality of x_ϵ for (16) implies that

$$F_T(x_\epsilon, \Delta_\epsilon) \equiv c^{\mathrm{T}} x_\epsilon + T \min[\lambda_{\min}(A[x_\epsilon], \Delta_\epsilon), 0] \geq \mathrm{Opt} - \epsilon. \tag{18}$$

It is possible that $\lambda_{\min}(A[x_\epsilon], \Delta_\epsilon) \geq 0$; then x_ϵ is feasible for (14) by Theorem 2, $x^\epsilon = x_\epsilon$, and (18) says that x^ϵ is a feasible ϵ-solution to (14). Now let $\lambda_{\min}(A[x_\epsilon], \Delta_\epsilon) = -\lambda < 0$, so that $\gamma = \lambda/\nu$. Then (18) implies that

$$c^{\mathrm{T}} x_\epsilon + \gamma c^{\mathrm{T}} \bar{x} \geq \mathrm{Opt} - \epsilon + T\lambda + \gamma c^{\mathrm{T}} \bar{x} \geq \mathrm{Opt}(1+\gamma) - \epsilon$$

where we have used (17) to get $T\lambda \geq \gamma(\mathrm{Opt} - c^{\mathrm{T}} \bar{x})$, whence $c^{\mathrm{T}} x^\epsilon \geq \mathrm{Opt} - \epsilon$. It remains to note that $A[x_\epsilon] \succeq -\lambda \mathcal{K}$ by Proposition 2, while $A[\bar{x}] \succeq \nu \mathcal{K}$; it follows that

$$A[x^\epsilon] = (1+\gamma)^{-1}(A[x_\epsilon] + \gamma A[\bar{x}]) \succeq (1+\gamma)^{-1}\left[-\lambda \mathcal{K} + \frac{\lambda}{\nu} \nu \mathcal{K}\right] = 0.$$

□

Lemma 2 combines with the results of [6] to yield the following

Theorem 3 *Consider problem* (14) *satisfying Assumptions A.1 – A.3, and let* X *be either*

(a) *the Euclidean ball* $\{x \in \mathbf{R}^N : \|x\|_2 \leq R\}$*, or the intersection of this ball with nonnegative orthant,*

or

(b) *the box* $\{x \in \mathbf{R}^N : \|x\|_\infty \leq R\}$*,*

or

(c) *the* $\|\cdot\|_1$*-ball* $\{x \in \mathbf{R}^N : \|x\|_1 \leq R\}$*, or the full-dimensional simplex* $\{x \in \mathbf{R}^N : 0 \leq x, \sum_i x_i \leq R\}$*, or the "flat" simplex* $\{x \in \mathbf{R}^N : 0 \leq x, \sum_i x_i = R\}$*.*

Assume that we are given an upper bound χ *on the norm of the homogeneous part of* $A[\cdot]$ *considered as a linear mapping from* $(\mathbf{R}^N, \|\cdot\|_X)$ *to* $(\mathbf{S}^{(\nu)}, \|\cdot\|)$*, where*

$\|\cdot\|_X$ *is* $\|\cdot\|_2$ *in the cases of* (a), (b), *and is* $\|\cdot\|_1$ *in the case of* (c), *while* $\|\cdot\|$ *is the standard matrix norm (the largest singular value) throughout the remaining part of this subsection.*

Under the outlined assumptions, for every $\epsilon > 0$ *one can find a feasible* ϵ*-solution* x_ϵ *to* (14) *(so that* $x^\epsilon \in X$, $A[x^\epsilon] \succeq 0$ *and* $c^T x^\epsilon \le \mathrm{Opt} + \epsilon$*) in no more than*

$$N(\epsilon) = \mathrm{O}(1)\frac{[\mathrm{Opt}^{\mathrm{up}} - c^T\bar{x}]\sqrt{\ln n}}{\nu\epsilon} \times \begin{cases} [\chi R + \rho\sqrt{\ln n}], & \text{case of (a)} \\ [\chi R\sqrt{N} + \rho\sqrt{\ln n}], & \text{case of (b)} \\ [\chi R\sqrt{\ln(N)} + \rho\sqrt{\ln n}], & \text{case of (c)} \end{cases}, \tag{19}$$

steps, with computational effort per step dominated by the necessity

- *to compute* $A[x]$*, for a given* x*;*
- *to compute, given* m *symmetric matrices of the row sizes* $|J_k|$, $k = 1, \ldots, m$*, the eigenvalue decompositions of the matrices.*

Above, O(1) *is an absolute constant,* $N = \dim x$*,* n *is the row dimension of* $A[\cdot]$*, and* ν *is given by* (15).

Proof Let $T = \frac{\mathrm{Opt}^{\mathrm{up}} - c^{\mathrm{T}}\bar{x}}{\nu}$. By Lemma 2, a feasible ϵ-solution to (14) is readily given by an ϵ-solution to the saddle point problem (16) with T we have just defined. Now, problem (16) is of the form

$$\max_{u=(x,\Delta)\in X\times\Delta_\rho} \min_{Y\in\mathcal{Y}\subset\mathbf{S}} \left[\mathrm{lin}(u,Y) + T\langle\mathcal{A}(x) + \mathcal{D}(\Delta), Y\rangle\right], \tag{20}$$

where

- $\mathrm{lin}(u, Y)$ is an appropriate affine function of u, Y,
- $Y = \mathrm{Diag}\{Y^1, \ldots, Y^m\}$, $Y^k = [Y^k_{ij}]_{i,j\in J_k}$, $k = 1, \ldots, m$, $\mathbf{S}$ is the linear space of all block-diagonal matrices Y of the indicated block-diagonal structure, and $\mathcal{Y} = \{Y \in \mathbf{S} : 0 \preceq Y, \mathrm{Tr}(Y) \le 1\}$;
- $\mathcal{A}(\cdot)$ is the linear mapping from $\mathbf{R}^N$ into $\mathbf{S}$ defined as follows. Given $x \in \mathbf{R}^N$, we compute the homogeneous part $A = A(x) = A[x] - A[0]$ of the mapping $A[\cdot]$ at x. The k-th diagonal block $\mathcal{A}^k(x)$ in $\mathcal{A}(x)$, $k = 1, \ldots, m$, is the contribution of A to $\mathcal{B}_k(A[x], \Delta)$, see (10);
- $\mathcal{D}(\cdot)$ is the linear mapping from the space $\widehat{\mathbf{S}}$ of block-diagonal matrices $\Delta = \mathrm{Diag}\{\Delta^1, \ldots, \Delta^{m-1}\}$, $\Delta^\ell = [\Delta^\ell_{ij}]_{i,j\in J'_{\ell+1}}$, $\ell = 1, \ldots, m-1$, into $\mathbf{S}$ defined as follows: The k-th diagonal block $\mathcal{D}^k(\Delta)$ in $\mathcal{D}(\Delta)$ is the contribution of Δ to $\mathcal{B}_k(A[x], \Delta)$, see (10);
- $\boldsymbol{\Delta}_\rho$ is the set of all positive semidefinite matrices from $\widehat{\mathbf{S}}$ with trace $\le \rho$;
- finally, $\langle\cdot,\cdot\rangle$ is the Frobenius inner product on $\mathbf{S}$.

Now, as shown in [6], the Mirror-Prox algorithm from [6] solves problem (20) within any given accuracy $\epsilon > 0$ in no more than

$$N(\epsilon) = \mathrm{O}(1)T\frac{L_{XY}\sqrt{\Theta_X\Theta_Y} + L_{\Delta Y}\sqrt{\Theta_\Delta\Theta_Y}}{\epsilon}$$

steps of the complexity indicated in Theorem 3, where

$$\Theta_X = \begin{cases} R^2, & \text{case (a)} \\ R^2 N, & \text{case (b)}, \\ R^2 \ln N, & \text{case (c)} \end{cases} \quad \Theta_Y = \ln n, \quad \Theta_\Delta = \rho^2 \ln n,$$

L_{XY} is the norm of the linear mapping $\mathcal{A}$ considered as a mapping from $(\mathbf{R}^N, \|\cdot\|_X)$ to $(\mathbf{S}, \|\cdot\|)$, and $L_{\Delta Y}$ is the norm of the linear mapping $\mathcal{D}$ considered as the mapping from $(\widehat{\mathbf{S}}, |\cdot|_1)$ to $(\mathbf{S}, \|\cdot\|)$, where $|\Delta|_1$ is the sum of modulae of eigenvalues of $\Delta \in \widehat{\mathbf{S}}$. It remains to evaluate L_{XY} and $L_{\Delta Y}$. Let $x \in \mathbf{R}^N$ satisfy $\|x\|_X \le 1$, and let $A = A(x)$, so that $\|A\| \le \chi$. Invoking (10), it is immediately seen that $\mathcal{A}^k(x)$, for every k, is a "border" in A: there exist two principal submatrices in A embedded one into another such that $\mathcal{A}^k(x)$ is obtained from the larger submatrix by replacing the entries belonging to the smaller one by zeros. By eigenvalue interlacement Theorem, both submatrices are of norm $\le \chi$, so that the "border" is of norm $\le 2\chi$, whence $\|\mathcal{A}(x)\| \le 2\chi$. Thus, $L_{XY} \le 2\chi$. Now let us bound $L_{\Delta Y}$. The extreme points of the unit $|\cdot|_1$-ball D in $\widehat{\mathbf{S}}$ are block-diagonal matrices with just one nonzero diagonal block, which is a symmetric rank 1 matrix of the corresponding size with the only nonzero singular value equal to 1, or equivalently, is a rank 1 matrix of the Frobenius norm equal to 1. For such a matrix Δ, it follows immediately from (10) that the Frobenius (and then – the matrix) norm of every block in $\mathcal{D}(\Delta)$ is at most 2. Since $L_{\Delta Y}$ is the maximum of the quantities $\|\mathcal{D}(\Delta)\|$ over the extreme points Δ of D, we conclude that $L_{\Delta Y} \le 2$. Combining our observations, we arrive at (19). □

We have presented a rather general approach to solving SDPs by reducing them to saddle point problems which are further solved by the $O(t^{-1})$-converging Mirror-Prox algorithm from [6]. In the sequel, we apply this scheme to the problems of computing Lovász capacity of a graph and to MAXCUT, with emphasis on utilizing favorable sparsity patterns of the underlying graphs.

4.2 Computing Lovász capacity for a graph with a favorable sparsity pattern

Let $v = (v_1, v_2, \ldots, v_{n+1})^T \in \mathbf{R}^{n+1}$ be a simple sparsity structure with $v_1 = n$, and let G be an undirected graph with n nodes, indexed by $2, 3, \ldots, n+1$, and the set of arcs E such that if $(i,j) \in E$ and $i \le j$, then $j \le i + v_i$. Recall from Section 2 that, for each entry ij with $1 \le i \le j \le i + v_i$, $\ell(i,j)$ is the number of sets J_k $(k = 1, \ldots, m)$ such that $i, j \in J_k$. Let

$$\eta = \max_{2 \le i \le n+1} \ell(i,i). \tag{21}$$

Consider the Lovász capacity problem

$$
\begin{aligned}
\vartheta(G) &= \min_{X,\lambda}\left\{\lambda : \lambda I_n - ee^{\mathrm{T}} - X \succeq 0, (i,j)\notin E \Rightarrow X_{i-1,j-1}=0\right\}\\
&= \min_{X,\lambda}\left\{\lambda : \left[\begin{array}{c|c}\nu & \sqrt{\nu}e^{\mathrm{T}}\\ \hline \sqrt{\nu}e & \lambda I_n - X\end{array}\right] \succeq 0, (i,j)\notin E \Rightarrow X_{i-1,j-1}=0\right\}
\end{aligned}
\tag{22}
$$

where $e \in \mathbf{R}^n$ is the vector of ones and $\nu > 0$ is a parameter. Note that the equivalence of the two optimization problems in (22) is given by the Schur Complement Lemma. Let $\mathcal{M}$ be the affine subspace in $\mathbf{S}^{(\nu)}$ comprised of all matrices of the form $\left[\begin{array}{c|c}\nu & \sqrt{\nu}e^{\mathrm{T}}\\ \hline \sqrt{\nu}e & Z\end{array}\right]$ with Z constrained by the requirements

$$
Z_{11} = Z_{22} = \cdots = Z_{nn};\ (i<j \ \&\ (i,j)\notin E) \Rightarrow Z_{i-1,j-1}=0.
$$

We equip $\mathbf{S}^{(\nu)}$ (and thus $\mathcal{M}$) with the Euclidean structure given by the inner product

$$
\langle A, B\rangle_{\mathcal{L}} = \langle \mathcal{L}^{1/2} A \mathcal{L}^{1/2}, \mathcal{L}^{1/2} B \mathcal{L}^{1/2}\rangle,
$$

where $\langle P, Q\rangle$ is the Frobenius inner product and $\mathcal{L} = \mathrm{Diag}\left\{\{\ell^{-1/2}(i,i)\}_{i=1}^{n+1}\right\}$ (cf. (3)). The norm on $\mathbf{S}^{(\nu)}$ corresponding to the inner product $\langle\cdot,\cdot\rangle_{\mathcal{L}}$ will be denoted $\|\cdot\|_{\mathcal{L}}$. We denote by $\mathcal{P}$ the orthogonal projector of $\mathbf{S}^{(\nu)}$ onto $\mathcal{M}$, so that for any $A \in \mathbf{S}^{(\nu)}$ one has

$$
\mathcal{P}(A) = \left[\begin{array}{c|c}\nu & \sqrt{\nu}e^{\mathrm{T}}\\ \hline \sqrt{\nu}e & \gamma(A) I_n + \widehat{A}\end{array}\right],
$$

where $\gamma(A) = \left(\sum_{i=2}^{n+1}\ell^{-1}(i,i)A_{ii}\right)\left(\sum_{i=2}^{n+1}\ell^{-1}(i,i)\right)^{-1}$ and the matrix $\widehat{A}$ is obtained from the South-Eastern $n\times n$ angular block of A by replacing all diagonal entries and all entries ij with $(i,j)\notin E$ with zeros.

Given an upper bound $\widehat{\theta} \le n$ on the Lovász capacity, consider the following optimization problem:

$$
\begin{aligned}
\mathrm{Opt} = &\min_{B=\{B_k = B_k^T = [B_{ij}^k]_{i,j\in J_k}\}_{k=1}^m}\\
&\times\left\{\lambda(B) + T\|\mathcal{S}(B) - \mathcal{P}(\mathcal{S}(B))\|_{\mathcal{L}} : \begin{array}{c} B_k \succeq 0, k=1,\dots,m\\ \sum_{k=1}^m \mathrm{Tr}(B_k^2) \le R^2\end{array}\right\}\\
\mathcal{S}(B) &= \sum_{k=1}^m B^k,\ B^k =]B_{ij}^k[_{i,j\in J_k}\\
\lambda(B) &= \left(\sum_{i=2}^{n+1}\ell^{-1}(i,i)(\mathcal{S}(B))_{ii}\right)\left(\sum_{i=2}^{n+1}\ell^{-1}(i,i)\right)^{-1} = (\mathcal{P}(\mathcal{S}(B)))_{jj},\ j=2,3,\dots,n+1,\\
R &= \sqrt{\widehat{\theta}^{\,2}(n+2|E|) + \nu^2 + 2\nu n},
\end{aligned}
\tag{23}
$$

where $T \geq 1$.

Observe that

$$\text{Opt} \leq \vartheta(G). \tag{24}$$

Indeed, let X_* be the X-component of the optimal solution to (22). Then the matrix $Y_* = \left[\begin{array}{c|c} \nu & \sqrt{\nu}e^{\mathrm{T}} \\ \hline \sqrt{\nu}e & \vartheta(G)I_n - X_* \end{array}\right]$ is $\succeq 0$ and belongs to $\mathbf{S}^{(\nu)}$; by Proposition 1, this matrix is $\mathcal{S}(B^*)$ for certain $B^* \in \mathbf{B}$ with components $B^*_k \succeq 0$. From the latter fact and (5) it follows $\sum_k \|B^*_k\|_F^2 \leq \|Y_*\|_F^2 \leq R^2$, with the latter inequality readily given by the fact that $|(X_*)_{ij}| \leq \vartheta(G)$ due to $Y_* \succeq 0$. Thus, B^* is feasible for (23); at this feasible solution, the objective of (23) clearly is equal to $\lambda(B^*) = \vartheta(G)$, and (24) follows.

Observe also that (23) is nothing but the saddle point problem

$$\min_{B \in \mathcal{B}} \max_{Y \in \mathcal{Y}} F(B, Y), \tag{25}$$

where

$$\begin{aligned} \mathcal{B} &= \{B \in \mathbf{B} : B_k \succeq 0,\ k = 1, \ldots, m, \sum_k \|B_k\|_F^2 \leq R^2\} \\ \mathcal{Y} &= \{Y \in \mathbf{S}^{(\nu)} : \|Y\|_{\mathcal{L}} \leq 1\} \\ F(B, Y) &= \lambda(B) + T\langle Y, \mathcal{S}(B) - \mathcal{P}(\mathcal{S}(B))\rangle_{\mathcal{L}} \end{aligned} \tag{26}$$

Note that by (6) the norm of the linear part of the affine mapping

$$B \mapsto \mathcal{Q}(B) = \mathcal{S}(B) - \mathcal{P}(\mathcal{S}(B)),$$

treated as the mapping from the space $\mathbf{B}$ equipped with the norm $\|B\|_F = \sqrt{\sum_{k=1}^m \|B_k\|_F^2}$ to the space $\mathbf{S}^{(\nu)}$ equipped with the norm $\|\cdot\|_{\mathcal{L}}$ is ≤ 1.

Since the mapping $\mathcal{Q}$ is of norm ≤ 1, from the results of [6] the saddle point problem (25) can be solved within accuracy $\epsilon > 0$ in no more than

$$N(\epsilon) = \mathrm{O}(1)\frac{TR}{\epsilon} \tag{27}$$

steps, with the computational effort per step dominated by the necessity to find eigenvalue decompositions of m symmetric matrices of the sizes $|J_1|, \ldots, |J_m|$. Thus, for all practical purposes computational effort per step does not exceed

$$\mathcal{C} = \mathrm{O}(1)\sum_{k=1}^m |J_k|^3. \tag{28}$$

Assume that we have found an ϵ-solution $\widetilde{B} = \{\widetilde{B}_k\}_{k=1}^m \in \mathcal{B}$ to (25), so that

$$\lambda(\widetilde{B}) + T \underbrace{\| \overbrace{\mathcal{S}(\widetilde{B}) - \mathcal{P}(\mathcal{S}(\widetilde{B}))}^{\Delta} \|_{\mathcal{L}}}_{\delta} \leq \mathrm{Opt} + \epsilon. \tag{29}$$

and $\widetilde{B}_k \succeq 0$ for all k, whence $\mathcal{S}(\widetilde{B}) \succeq 0$. Observe that

$$\mathcal{P}(\mathcal{S}(\widetilde{B})) = \left[\begin{array}{c|c} \nu & \sqrt{\nu} e^{\mathrm{T}} \\ \hline \sqrt{\nu} e & \lambda(\widetilde{B}) I_n - X \end{array}\right], \tag{30}$$

where X is of the structure required in (22). Since $\|\Delta\|_{\mathcal{L}} \equiv \|\mathcal{L}^{1/2} \Delta \mathcal{L}^{1/2}\|_F = \delta$ (where Δ, δ are defined as in (29)), we have $\mathcal{L}^{1/2} \Delta \mathcal{L}^{1/2} \preceq \delta I_{n+1}$, whence $\Delta \preceq \delta \mathcal{L}^{-1}$. This combined with $\mathcal{S}(\widetilde{B}) \succeq 0$ results in $\mathcal{P}(\mathcal{S}(\widetilde{B})) \succeq -\delta \mathcal{L}^{-1}$. This together with (3), (21) and (30) implies that

$$\left[\begin{array}{c|c} \nu + m^{1/2}\delta & \sqrt{\nu} e^{\mathrm{T}} \\ \hline \sqrt{\nu} e & [\lambda(\widetilde{B}) + \eta^{1/2}\delta] I_n - X \end{array}\right] \succeq 0,$$

whence

$$\left[\begin{array}{c|c} \nu & \sqrt{\nu} e^{\mathrm{T}} \\ \hline \sqrt{\nu} e & \frac{\nu + m^{1/2}\delta}{\nu} [\lambda(\widetilde{B}) + \eta^{1/2}\delta] I_n - \frac{\nu + m^{1/2}\delta}{\nu} X \end{array}\right] \succeq 0.$$

Thus, an ϵ-solution $\widetilde{B}$ to (25) can be easily converted to a feasible solution $(\widetilde{\lambda}, \widetilde{X} = \frac{\nu + m^{1/2}\delta}{\nu} X)$ to (22) with the value of the objective

$$\begin{aligned} \widetilde{\lambda} &= \tfrac{\nu + m^{1/2}\delta}{\nu} [\lambda(\widetilde{B}) + \eta^{1/2}\delta] \\ &\leq \tfrac{\nu + m^{1/2}\delta}{\nu} \left[\mathrm{Opt} + \epsilon - (T - \eta^{1/2})\delta\right] && \text{[see (29)]} \\ &\leq \tfrac{\nu + m^{1/2}\delta}{\nu} \left[\vartheta(G) + \epsilon - (T - \eta^{1/2})\delta\right] && \text{[see (24)]} \\ &= \vartheta(G) + \epsilon + \delta \left[\tfrac{m^{1/2}(\vartheta(G) + \epsilon)}{\nu} - (T - \eta^{1/2}) \tfrac{\nu + m^{1/2}\delta}{\nu}\right]. \end{aligned}$$

We arrive at the following result:

Proposition 6 *Let η be defined as in* (21)*, and let*

$$T \geq \eta^{1/2} + \frac{m^{1/2}(\vartheta(G) + \epsilon)}{\nu}.$$

Then an ϵ-solution to (25) *induces a feasible ϵ-solution to* (22)*. The number of steps required to get such a solution can be bounded by* (27)*, while the computational effort per step can be bounded by* (28).

Corollary 2 *Given an upper bound $\widehat{\theta}$ on $\vartheta(G)$, let us set*

$$\phi(\nu) = \left(\eta^{1/2} + \frac{m^{1/2}\widehat{\theta}}{\nu}\right)\sqrt{\widehat{\theta}^2(n+2|E|) + \nu^2 + 2\nu n}$$

and

$$\bar{\nu} = \operatorname*{argmin}_{\nu>0} \phi(\nu), \quad \widehat{T} = \eta^{1/2} + \frac{m^{1/2}\widehat{\theta}}{\bar{\nu}}$$

With $T = \widehat{T}$, the outlined procedure allows, for every ϵ, $0 < \epsilon \leq \widehat{\theta} - \vartheta(G)$, to find a feasible ϵ-solution to (22) *in no more than*

$$N(\epsilon) = \mathrm{O}(1)\frac{\phi(\bar{\nu})}{\epsilon}$$

steps, with the complexity of a step given by (28).

Corollary 3 *Let $|E| \geq n$. Then, setting*

$$\nu = \min\left[\widehat{\theta}\sqrt{|E|}, \widehat{\theta}^2|E|n^{-1}\right],$$

one gets

$$N(\epsilon) \leq \mathrm{O}(1)\frac{\widehat{\theta}\sqrt{m|E|}}{\epsilon}.$$

Proof Indeed, with ν in question, we clearly have $\sqrt{\widehat{\theta}^{\,2}(n+2|E|) + \nu^2 + 2\nu n} \leq \mathrm{O}(1)\widehat{\theta}\sqrt{|E|}$. Consequently,

$$\phi(\nu) \leq \mathrm{O}(1)\widehat{\theta}\sqrt{|E|}\left(\underbrace{\eta^{1/2}}_{\leq m^{1/2}} + \max\left[\underbrace{\frac{m^{1/2}}{\sqrt{|E|}}}_{\leq 1}, \underbrace{\frac{m^{1/2}n}{\widehat{\theta}|E|}}_{\leq m^{1/2}}\right]\right) \leq \mathrm{O}(1)\widehat{\theta}\sqrt{m|E|}.$$

This together with Corollary 2 implies that the conclusion holds. □

Example: staircase structure Let p, q be positive integers, and $n = p(q+1)$. Assume that the incidence matrix of the graph is from $\mathbf{S}^{[d]}$, where $d \in \mathbf{R}^{q+1}$ with $d_i = p$ for $i = 0, \ldots, q$. Then, from (22), we see that

$$i + v_i = \begin{cases} n+1, & 1 \leq i < 2+p \\ 1+(k+1)p, & 2+(k-1)p \leq i < 2+kp, \quad k = 2, \ldots, q \\ 1+(q+1)p, & 2+pq \leq i \leq n+1 \end{cases}$$

In the preceding notation, we have $i_k = 1+(k+2)p, k = 1,\ldots,q-1, \eta = q-1$, $|J_k| = 3p+1$, $|E| \le \mathrm{O}(1)p^2q$. Thus,

$$\mathcal{C} = \mathrm{O}(1)p^3q,\ \phi(\nu) \le \mathrm{O}(1)\left(q^{1/2} + \frac{q^{1/2}\widehat{\theta}}{\nu}\right)\left(\widehat{\theta}^2p^2q + \nu^2 + \nu pq\right)^{1/2}.$$

Setting $\widehat{\nu} = \widehat{\theta}pq^{1/2}$, we get

$$\phi(\widehat{\nu}) \le \mathrm{O}(1)q^{1/2}\left(\widehat{\theta}^{\,2}p^2q + p^2q^{3/2}\widehat{\theta}\right)^{1/2} \le \mathrm{O}(1)\widehat{\theta}pq\left(1 + q^{1/2}\widehat{\theta}^{-1}\right)^{1/2}.$$

Since the stability number of the corresponding graph clearly is at least $O(q)$, we have $\phi(\widehat{\nu}) \le \mathrm{O}(1)\widehat{\theta}pq$. Consequently, computing Lovász capacity within accuracy ϵ costs at most

$$\mathrm{O}(1)\frac{\widehat{\theta}pq}{\epsilon} \times p^3q = \mathrm{O}(1)\frac{\widehat{\theta}p^4q^2}{\epsilon}$$

operations. For comparison:

1. Saddle point approach, similar to the above one, as applied to computing Lovász capacity for a *general* pq-node graph G with $O(p^2q)$ arcs and $\vartheta(G) \le \widehat{\theta}$, results in the bound $O(1)\frac{\widehat{\theta}p^4q^{7/2}\sqrt{\ln(pq)}}{\epsilon}$, see [6];
2. The arithmetic cost of a single interior point iteration in the problem of computing Lovász capacity of a *general* pq-node graph is as large as $O(1)p^6q^6$, and is at least p^6q^3 even in the case of graph possessing the structure in question.

4.3 The MAXCUT problem on a graph with a favorable sparsity pattern

Consider a MAXCUT-type problem

$$\mathrm{Opt} = \max_{X\in\mathbf{S}^n}\left\{\mathrm{Tr}(VX) : X \succeq 0, \mathrm{diag}(X) = \mathbf{e}\right\} \tag{31}$$

where $\mathrm{diag}(A)$ is the diagonal of a square matrix A and $\mathbf{e}$ is the vector of ones. Assume that $V \in \mathbf{S}^{(\nu)}$ for a given simple sparsity structure ν. By Proposition 4 problem (31) is equivalent to

$$\mathrm{Opt} = \max_{X\in\mathbf{S}^{(\nu)}}\left\{\mathrm{Tr}(VX) : \mathrm{diag}(X) = \mathbf{e}, X^k \equiv [X_{ij}]_{i,j\in J_k} \succeq 0,\ k = 1,\ldots,m\right\}. \tag{32}$$

Let $\mathcal{X} = \{X \in \mathbf{S}^{(\nu)} : |X_{ij}| \le 1 \forall i, j, X_{ii} = 1 \forall i\}$, $\mathcal{Y} = \{Y = \{Y^k = [Y^k_{ij}]_{i,j \in J_k}\}_{k=1}^m : Y^k \succeq 0, \sum_k \mathrm{Tr}(Y^k) \le 1\}$. Consider the saddle point problem

$$\mathrm{Opt}^+ = \max_{X \in \mathcal{X}} \Phi(X) \equiv \min_{Y \in \mathcal{Y}} \left[\mathrm{Tr}(VX) + T \sum_{k=1}^m \mathrm{Tr}(X^k Y^k) \right], \tag{33}$$

where $T > 0$ is a parameter. Observe that the optimal value in (33) is $\ge$ Opt. Indeed, if X_* is an optimal solution to (32), then clearly $X_* \in \mathcal{X}$, and $\Phi(X_*) = \mathrm{Tr}(VX_*)$. Now let X be an ϵ-solution to (33), so that $X \in \mathcal{X}$ and

$$\begin{aligned} \mathrm{Tr}(VX) - T\lambda \ge \mathrm{Opt}^+ - \epsilon \ge \mathrm{Opt} - \epsilon, \\ \lambda = \max[0, -\lambda_{\min}(X^1), \ldots, -\lambda_{\min}(X^m)]. \end{aligned}$$

It is possible that $\lambda = 0$, that is, X is feasible for (32); in this case, X is a feasible ϵ-solution to the latter problem. Now consider the case when $\lambda > 0$, and let $\tilde{X} = (1 + \lambda)^{-1}(X + \lambda I)$. Clearly, $\tilde{X}$ is feasible for (32). Setting $\widehat{X} = X + \lambda I$, we have

$$\begin{aligned} \mathrm{Tr}(V\widehat{X}) &= \mathrm{Tr}(VX) + \lambda \mathrm{Tr}(V) \ge \mathrm{Opt} - \epsilon + \lambda[\mathrm{Tr}(V) + T] \\ \Rightarrow \mathrm{Tr}(V\tilde{X}) &\ge (1 + \lambda)^{-1}[\mathrm{Opt} - \epsilon] + \lambda[\mathrm{Tr}(V) + T] \\ &\ge \mathrm{Opt} - \epsilon + (1 + \lambda)^{-1}\lambda \left[T + \mathrm{Tr}(V) - \mathrm{Opt}\right]. \end{aligned}$$

We see that *if*

$$T \ge \mathrm{Opt} - \mathrm{Tr}(V),$$

then $\tilde{X}$ *is a feasible* ϵ*-solution to* (32). This observation suggests the following scheme for solving (32): given an upper bound $\mathrm{Opt}^{\mathrm{up}}$ on Opt, we set $T = \mathrm{Opt}^{\mathrm{up}} - \mathrm{Tr}(V)$ and solve saddle point problem (33) within accuracy ϵ, and then convert, in the just presented fashion, the resulting X into a feasible ϵ-solution to (32).

By [6], generating an ϵ-solution to (33) costs $\mathrm{O}(1)\frac{T\sqrt{\dim \mathbf{S}^{(\nu)}}\sqrt{\ln n}}{\epsilon}$ steps, with the computational effort per step dominated by the necessity to find eigenvalue decompositions of m matrices X^k, $k = 1, \ldots, m$, where X^k is defined as in (32) for $X \in \mathcal{X}$. We arrive at the following result:

Proposition 7 *Let an upper bound* $\mathrm{Opt}^{\mathrm{up}}$ *on the optimal value in* (31) *be given. For every* $\epsilon > 0$*, a feasible* ϵ*-solution to problem* (31) *with* $V \in \mathbf{S}^{(\nu)}$ *can be found in no more than*

$$N(\epsilon) = \mathrm{O}(1)\frac{[\mathrm{Opt}^{\mathrm{up}} - \mathrm{Tr}(V)]\sqrt{\ln n}\sqrt{\sum_{i=1}^n (1 + \nu_i)}}{\epsilon} \tag{34}$$

steps of Mirror-Prox algorithm [6]*, with* $\mathrm{O}(1)\sum_{k=1}^m |J_k|^3$ *operations per step.*

Table 1 Computational result for the Lovász capacity problem

(q,p)	Nodes	Edges	LwBnd	UpBnd	Iter	CPU
(4999,2)	10000	44988	2.497e3	2.516e3	11757	$3^h 55' 16''$
(9999,2)	20000	89988	4.996e3	5.045e3	17238	$11^h 50' 58''$
(14999,2)	30000	134988	7.484e3	7.545e3	30162	$32^h 58' 22''$
(19999,2)	40000	179988	9.952e3	1.004e4	34833	$51^h 49' 3''$
(3333,3)	10002	69987	1.660e3	1.676e3	10770	$4^h 22' 41''$
(6666,3)	20001	139980	3.329e3	3.358e3	20097	$16^h 28' 4''$
(9999,3)	30000	209973	4.998e3	5.046e3	24615	$33^h 5' 22''$
(13333,3)	40002	279987	6.643e3	6.708e3	29154	$51^h 2' 55''$
(2499,4)	10000	94952	1.249e3	1.259e3	8313	$4^h 11' 37''$
(4999,4)	20000	189952	2.491e3	2.514e3	17412	$17^h 52' 14''$
(7499,4)	30000	284952	3.747e3	3.784e3	21315	$34^h 10' 7''$
(9999,4)	40000	379952	4.972e3	5.022e3	28737	$61^h 11' 53''$
(1999,5)	10000	119925	9.970e2	1.001e3	9792	$5^h 43' 27''$
(3999,5)	20000	239925	1.995e3	2.013e3	13041	$15^h 29' 2''$
(5999,5)	30000	359925	2.989e3	3.016e3	23625	$42^h 46' 4''$
(7999,5)	40000	479925	3.990e3	4.022e3	27381	$75^h 41' 56''$
(1666,6)	10002	144921	8.301e2	8.382e2	9999	$6^h 56' 21''$
(3333,6)	20004	289950	1.659e3	1.676e3	14205	$20^h 27' 42''$
(4999,6)	30000	434892	2.496e3	2.517e3	17403	$46^h 12' 42''$
(6666,6)	40002	579921	3.331e3	3.364e3	21621	$62^h 33' 58''$

Remark 2 When V is a diagonal-dominated matrix: $V_{ii} \geq \sum_{j \neq i} |V_{ij}|$ (as it is the case in the true MAXCUT problem), one clearly has $\mathrm{Tr}(V) \leq \mathrm{Opt} \leq 2\mathrm{Tr}(V)$. In this case, one can set $\mathrm{Opt}^{\mathrm{up}} = 2\mathrm{Tr}(V)$, thus converting (34) into the bound $N(\epsilon) \leq O(1)\frac{\mathrm{Tr}(V)}{\epsilon}\sqrt{\ln n}\sqrt{\sum_{i=1}^{n}(1+v_i)}$.

Example: staircase structure Let p, q be positive integers, and $n = p(q+1)$. Consider the staircase structure $d = (p, \ldots, p) \in \mathbf{R}^{q+1}$, and assume that we are given an n-node graph G with incidence matrix belonging to $\mathbf{S}^{[d]}$. Given a matrix A of nonnegative weights of arcs in G, let $V_{ij} = \frac{1}{4}\begin{cases} -A_{ij}, & i \neq j \\ \sum_j A_{ij}, & i = j \end{cases}$, so that (31) becomes the standard MAXCUT problem associated with (A, G). By Remark 2, the outlined scheme allows to solve the latter problem within any accuracy $\epsilon > 0$ at the arithmetic cost of $\mathrm{O}(1)\frac{\mathrm{Opt}}{\epsilon}p^4 q^{3/2}\sqrt{\ln(pq)}$ operations. Note that the arithmetic cost of a *single* interior point iteration as applied to the "most economical", in terms of the design dimension, dual reformulation of (31), is $\mathrm{O}(1)p^3q^3$. It follows that when a "moderate" relative accuracy ϵ/Opt, say, $\epsilon/\mathrm{Opt} = 0.01$ is sought and $q^{3/2} \gg p\sqrt{\ln(pq)}$, the Mirror-Prox algorithm as applied to the MAXCUT problem by far outperforms Interior Point techniques. The difference becomes even more significant when we compare the complexity bound for Mirror-Prox with the theoretical complexity bound of $\mathrm{O}(1)\sqrt{pq}\ln\left(\frac{\mathrm{Opt}}{\epsilon}\right)p^3q^3$ operations for IPMs (the factor $\mathrm{O}(1)\sqrt{pq}\ln\left(\frac{\mathrm{Opt}}{\epsilon}\right)$

Table 2 Computational results for the MAXCUT problem

(q,p)	Nodes	Edges	LwBnd	UpBnd	Iter	CPU
(4999,2)	10000	44988	1.921e5	1.940e5	2604	$51'22''$
(9999,2)	20000	89988	3.859e5	3.898e5	3711	$2^h20'40''$
(14999,2)	30000	134988	5.775e5	5.833e5	3963	$3^h50'19''$
(19999,2)	40000	179988	7.725e5	7.802e5	4137	$5^h10'46''$
(39999,2)	80000	359988	1.545e6	1.560e6	5622	$13^h53'19''$
(3333,3)	10002	69987	2.862e5	2.891e5	3447	$1^h29'56''$
(6666,3)	20001	139980	5.717e5	5.775e5	3765	$3^h14'50''$
(9999,3)	30000	209973	8.574e5	8.660e5	4536	$5^h58'23''$
(13333,3)	40002	279987	1.146e6	1.157e6	5355	$9^h4'41''$
(26666,3)	80001	559980	2.291e6	2.314e6	7260	$24^h24'51''$
(2499,4)	10000	94952	3.783e5	3.821e5	2673	$1^h21'31''$
(4999,4)	20000	189952	7.585e5	7.660e5	3531	$3^h36'46''$
(7499,4)	30000	284952	1.137e6	1.148e6	4317	$6^h49'26''$
(9999,4)	40000	379952	1.515e6	1.530e6	4773	$9^h42'54''$
(19999,4)	80000	759952	3.028e6	3.058e6	6393	$25^h53'39''$
(1999,5)	10000	119925	4.703e5	4.750e5	3012	$1^h53'49''$
(3999,5)	20000	239925	9.423e5	9.517e5	3177	$3^h53'26''$
(5999,5)	30000	359925	1.417e6	1.431e6	3741	$9^h6'1''$
(7999,5)	40000	479925	1.885e6	1.904e6	4338	$10^h32'40''$
(15999,5)	80000	959925	3.771e6	3.809e6	5508	$26^h32'50''$
(1666,6)	10002	144921	5.645e5	5.701e5	2487	$1^h44'37''$
(3333,6)	20004	289950	1.127e6	1.138e6	3153	$4^h23'17''$
(4999,6)	30000	434892	1.694e6	1.711e6	3558	$9^h42'34''$
(6666,6)	40002	579921	2.257e6	2.279e6	4263	$11^h53'27''$
(13333,6)	80004	1159950	4.514e6	4.559e6	5619	$31^h17'50''$

is the theoretical bound on the number of IPM iterations required to get an ϵ-solution).

5 Numerical implementation

In this section, we present the results of numerical experiments with the Lovász capacity problem (22) and the (semidefinite relaxation of the) MAXCUT problem (31). These problems were solved by the first-order Mirror-Prox algorithm from [6] as applied to the saddle point reformulations (25), respectively, (33), of the problems.

In our experiments, the incidence matrix has staircase structure with $d = (p, \ldots, p) \in \mathbf{R}^{q+1}$ for some $q > 0$, with dense $p \times p$ blocks allowed by the structure. Note that the number of nodes in such a graph is $n = (q+1)p$, while the number of arcs is $\frac{p^2(5q-1)-p(q+1)}{2}$. For our computations, we generated graphs with $p = 2, 3, \ldots, 6$ and n ranging from about 10,000 to about 80,000 (so that the number of arcs varied from about 50,000 to about 1,100,000). We terminate the computations when the relative error, as given by valid on-line inaccuracy bounds generated by the Mirror-Prox algorithm, became less than 1% for both problems. Our code is written in ANSI C. All computations are performed on Supermicro dual−2.66 GHz Intel Xeon server with 2 GB RAM.

Lovász capacity problem When solving this problem according to the scheme developed in Sect. 4.2, one needs an a priori upper bound $\widehat{\theta}$ on $\vartheta(G)$. Using the well-known result that Lovász capacity number of a graph G is bounded above by the chromatic number of the complement graph, it easy to see that for the graphs we are generating one can take $\widehat{\theta} = q$, and these were the upper bounds used in our computations. The results are presented in Table 1. In Table 1, the first three columns report the sizes of our generated graphs. The fourth and the fifth columns present the valid upper, respectively, lower bounds on $\vartheta(G)$ as reported by the Mirror-Prox algorithm. The last two columns report the number of steps and the CPU time.

Semidefinite relaxation of MAXCUT The graphs used in our experiments have the same structure as in the case of Lovász capacity problems. The weights of the arcs were picked at random from the uniform distribution in $[1, 11]$. The results are presented in Table 2; the structure of Table 2 is identical to the one of Table 1.

Acknowledgements This work was partially supported by NSF Grants CCR-0203113 and CCF-0430644 and ONR Grant N00014-05-1-0183.

References

1. Burer, S., Monteiro, R.D.C., Zhang, Y.: Solving a class of semidefinite programs via nonlinear programming. Math. Program. **93**, 97–122 (2002)
2. Burer, S., Monteiro, R.D.C.: A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Math. Program. Ser. B **95**, 329–357 (2003)
3. Fukuda, M., Kojima, M., Murota, K., Nakata, K.: Exploiting sparsity in semidefinite programming via matrix completion I: general framework. SIAM J Optim. **11**, 647–674 (2000)
4. Grone, B., Johnson, C.R. Marques de Sa, E., Wolkowicz, H.: Positive definite completions of partial Hermitian matrices. Linear Algebr. Appl. **58**, 109–124 (1984)
5. Helmberg, C., Rendl, F.: A spectral bundle method for semidefinite programming. SIAM J. Optim. **10**, 673–696 (2000)
6. Nemirovski, A.: Prox-method with rate of convergence O$(1/t)$ for variational inequalities with lipschitz continuous monotone operators and smooth convex–concave saddle point problems. SIAM J. Optim. **15**, 229–251 (2004)
7. Nesterov, Yu.: Smooth minimization of non-smooth functions. Math. Program. **103**, 127–152 (2005)
8. Nesterov, Yu.: Dual extrapolation and its applications for solving variational inequalities and related problems. – CORE Discussion Paper 2003/68, http://www.core.ucl.ac.be/services/COREdp03.html (2003)