

Large-Scale Semidefinite Programming via Saddle Point Mirror-Prox Algorithm

Zhaosong Lu* Arkadi Nemirovski† Renato. D. C. Monteiro‡

November 11, 2004

Abstract

In this paper, we first develop “economical” representations for positive semidefiniteness of well-structured sparse symmetric 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 with efficiency $\mathcal{O}(\epsilon^{-1})$ developed in [6]. Some numerical implementations for large-scale Lovasz capacity and MAXCUT problems are finally present.

AMS 2000 subject classification: 90C22, 90C25, 90C51, 65K10

1 Introduction

Consider a semidefinite program

$$\min_x \{ \text{Tr}(cx) : x \in \mathcal{N} \cup \mathbf{S}_+ \}, \quad (1)$$

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 favourable sparsity pattern of a large-scale problem (1) (that is, 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 to produce high-accuracy solutions. Note, however, that IPMs are Newton-type methods, with an iteration which requires assembling and solving a

*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA (Email: zhaosong@isye.gatech.edu). This author was supported in part by NSF Grant CCR-0203113 and CCF-0430644.

†School of Industrial Engineering and Management, Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel (E-mail: nemirovs@ie.technion.ac.il).

‡School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA (Email: monteiro@isye.gatech.edu). This author was supported in part by NSF Grant CCR-0203113 and CCF-0430644.

Newton system of n linear equations with n unknowns, where $n = \min[\dim \mathcal{N}, \text{codim } \mathcal{N}]$ 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 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 μ of diagonal blocks in matrices from \mathbf{S} is not too large, say, $\mu \leq 1000$. 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 μ is much larger than the above limit, provided that diagonal blocks in the matrices $A \in \mathcal{N}$ possess favourable 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 of 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 [7, 8, 6]. 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 matrices from \mathbf{S} rather than to be able to compute few largest eigenvalues of matrices from \mathcal{N} . As a result, the algorithms from [7, 8, 6]) as applied to (1) become impractical, provided that the largest size μ of diagonal blocks in matrices from \mathbf{S} exceeds about 1000.

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 matrices from \mathcal{N} possess favourable 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 section 2, we explain what a “favourable sparsity pattern” is and introduce some notation and definitions which will be used throughout the paper. In section 3, we develop our main tool, specifically, demonstrate that positive definiteness of a large symmetric matrix A possessing a favourable 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 matrix¹⁾. In section 4 we utilize the aforementioned representations to derive saddle point

¹⁾This result, which we get “for free”, can be also obtained, with a moderate effort, from general results of [4]

formulations of some large-scale SDP problems, specifically, those of computing Lovasz capacity of a graph and the MAXCUT problem, with emphasis on the case when the incidence matrix of the underlying graph possesses a favourable 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 favourable sparsity pattern in the incidence matrix. In concluding section 5, we illustrate our constructions by numerical results for the MAXCUT and Lovasz capacity problems on well-structured sparse graphs.

2 Well-structured sparse symmetric matrices

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

Motivation. To get an idea what a “favourable sparsity pattern” might be, consider 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 matrix $A = \text{Diag}\{A^1, \dots, A^L\}$ from \mathcal{N} is or is not positive semidefinite. Assuming that we are checking positive semidefiniteness of sparse symmetric matrices A^ℓ by applying Cholesky factorization algorithm, that is, by trying to represent A^ℓ as $D_\ell D_\ell^T$ with lower triangular D_ℓ , the nonzeros in D_ℓ will, generically, be the entries i, j with $i - v_i \leq j \leq i$, where $i - v_i = \min\{j : A_{ij}^\ell \neq 0\}$. In other words, when adding to the original pattern of nonzero entries all entries i, j with $i - v_i \leq j \leq i$ (and all symmetric entries), we do not alter the fill in of the Cholesky factor. In other words, we do not lose much by assuming that the original pattern of nonzeros already was comprised of all sub-diagonal entries (i, j) with $i - v_i \leq j \leq i$, with added symmetric entries. For notational convenience, we prefer to work with “symmetric” situation, where the nonzero entries are super-diagonal entries i, j with $i \leq j \leq i + v_i$, and the symmetric sub-diagonal entries; note that matrices of the former type can be obtained from those of the latter one by reversing the order of rows and columns. 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 $i + v_i \leq n$ for all i ; the “hard zero” super-diagonal entries i, j ($i \leq j$) 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 $A = UU^T$, 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 $A = UU^T$ with upper triangular U).

We are about to introduce terminology and notation allowing to operate with “well-structured”, in the sense we have just motivated, sparsity patterns.

Notation. Let $J \subset \{1, \dots, n\}$ be an index set with $\ell > 0$ elements. We denote by $[B_{ij}]_{i,j \in J}$ $\ell \times \ell$ matrix with rows and columns indexed by elements from 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 .

on existence of positive semidefinite completions.

Simple sparsity structures and associated entities. Let $v \in \mathbf{R}^n$ be a *simple sparsity structure* – a nonnegative integral vector such that $i + v_i \leq n$ for all $i \leq n$. 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 $I = \{i_1 < i_2 < \dots < i_m\}$ of all integers representable as $i + v_i$ with $i \leq n$. Note that $i_m = n$, since $n + v_n = n$ (recall that $i + v_i \leq n$ and $v_i \geq 0$). We refer to m as to *number of blocks* in v .
3. The sets

$$J_k = \{i \leq i_k : i + v_i \geq i_k\}, \quad J'_k = \{i \in J_k : i \leq i_{k-1}\}, \quad k = 1, \dots, m, \quad (2)$$

where $i_0 = 0$ (that is, $J'_1 = \emptyset$). Note that $J_k \setminus J_{k-1} = \{i_{k-1} + 1, \dots, i_k\}$ and that $J'_k \subset J_{k-1}$, 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 $I = \{i_1, \dots, i_m\}$; 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 $D_{ij} = \{k_-, k_- + 1, \dots, k_+\}$ 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

$$\mathcal{L} = \text{Diag}\{\ell(1, 1)^{-1/2}, \dots, \ell(n, n)^{-1/2}\}, \quad \mathcal{K} = \text{Diag}\{\ell(1, 1), \dots, \ell(n, n)\}. \quad (3)$$

6. The Euclidean space \mathbf{B} comprised of collections $B = \{B_k = [B_{ij}^k = B_{ji}^k]_{i,j \in J_k}\}_{k=1}^m$ of symmetric matrices 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 .

7. For $B = \{B_k = [B_{ij}^k = B_{ji}^k]_{i,j \in J_k}\}_{k=1}^m \in \mathbf{B}$, we set

$$B^k =]B_{ij}^k[_{i,j \in J_k} \in \mathbf{S}^{(v)}, \quad k = 1, \dots, m.$$

and define linear mapping $\mathcal{S}(B) : \mathbf{B} \rightarrow \mathbf{S}^{(v)}$ as

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

Finally, in the sequel $\lambda_{\min}(A)$ ($\lambda_{\max}(A)$) denotes the minimal, resp., the maximal eigenvalue of a symmetric matrix A .

where $*$ mark nonzero blocks. Proposition 3.1.(i) says that such a matrix is positive semidefinite if and only if it is sum of positive semidefinite matrices of the form

$$\left[\begin{array}{c} \boxed{\begin{matrix} * & * & * \\ * & * & * \\ * & * & * \end{matrix}} \\ \vdots \end{array} \right], \quad \left[\begin{array}{c} \boxed{\begin{matrix} * & * & * \\ * & * & * \\ * & * & * \end{matrix}} \\ \vdots \end{array} \right], \quad \left[\begin{array}{c} \boxed{\begin{matrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{matrix}} \\ \vdots \end{array} \right],$$

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

Proof of Proposition 3.1. (i): Induction in m . For $m = 1$ the statement is evident. Assuming that the statement is valid for $m = s$, let us prove it for $m = s + 1$. 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 . For $\epsilon \geq 0$, let $A_\epsilon = \left[\begin{array}{c|c} P & Q \\ \hline Q^T & R + \epsilon I \end{array} \right]$ with $i_{m-1} \times i_{m-1}$ block P . For $\epsilon > 0$, let $B^\epsilon = \left[\begin{array}{c|c} Q(R + \epsilon I)^{-1}Q^T & Q \\ \hline Q^T & R + \epsilon I \end{array} \right]$, so that $B^\epsilon \succeq 0 \succeq 0$. By the Schur Complement Lemma, we have $A_\epsilon - B^\epsilon \succeq 0$, thus, B^ϵ remains bounded as $\epsilon \rightarrow +0$. Thus, we can find a sequence $\epsilon_t \rightarrow +0$, $t \rightarrow \infty$, such that $\exists B^m = \lim_{t \rightarrow \infty} B^{\epsilon_t}$; observe that both B^m and $A - B^m$ are $\succeq 0$. By construction, $B^m \succeq 0$ and $B^m =]B_{ij}^m[_{i,j \in J_m}$; 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 a $i_{m-1} \times i_{m-1}$ matrix $\tilde{C} \in \mathbf{S}^{(v')}$, 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 \tilde{C} the inductive hypothesis, we can find $m - 1$ matrices $B^k =]B_{ij}^k[_{i,j \in J_k} \succeq 0$, $k = 1, \dots, m - 1$, 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 $\text{Tr}(BC) \geq 0$. 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 $A_{ij} = \sum_{k:i,j \in J_k} B_{ij}^k$. We have

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

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

$$\text{Card}(\{k : i, j \in J_k\}) \leq \sqrt{\ell(i,i)\ell(j,j)} \quad (7)$$

for every i, j . This is evident due to

$$\text{Card}(\{k : i, j \in J_k\}) \leq \min[\text{Card}(\{k : i \in J_k\}), \text{Card}(\{k : j \in J_k\})] = \min[\ell(i, i), \ell(j, j)]. \blacksquare$$

Proposition 3.1(i) establishes a characterization for positive semidefiniteness of matrices from $\mathbf{S}^{(v)}$, but it does not give the explicit formulas for the matrices $B_k = [B_{ij}^k = B_{ji}^k]_{i,j \in J_k}$. We next develop an equivalent reformulation of positive semidefiniteness of matrices from $\mathbf{S}^{(v)}$ by introducing some additional variables.

Lemma 3.1 *Let $m > 1$. A matrix $A \in \mathbf{S}^{(v)}$ is $\succeq 0$ if and only if there exists matrix $\Delta^{m-1} = (\Delta^{m-1})^T = [\Delta_{ij}^{m-1}]_{i,j \in J_m}$ such that the matrices*

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

and

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

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 . Vice versa, 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 $d = i_m - i_{m-1} = n - i_{m-1}$, let Q be the South-East $d \times d$ angular block in A , and P be the matrix with $|J'_m|$ rows, indexed by $i \in J'_m$, and d columns indexed by $j \in J_m \setminus J_{m-1} = \{i_{m-1} + 1, i_{m-1} + 2, \dots, i_m = n\}$ given by $P_{ij} = A_{ij}$. Let $\epsilon > 0$. Setting $\Delta^\epsilon = P(Q + \epsilon I_d)^{-1} P^T$ and applying the Schur Complement Lemma, we conclude that the matrices

$$\mathcal{B}^\epsilon = [B_{ij}]_{i,j \in J_m} : B_{ij} = \begin{cases} A_{ij} + \delta_j^i \epsilon, & i \notin J'_m \text{ or } j \notin J'_m \\ \Delta_{ij}^\epsilon, & i, j \in J'_m \end{cases}$$

and

$$\mathcal{C}^\epsilon = [C_{ij}^\epsilon]_{i,j=1}^{i_{m-1}} : C_{ij}^\epsilon = \begin{cases} A_{ij}, & i \notin J'_m \text{ or } j \notin J'_m \\ A_{ij} - \Delta_{ij}^\epsilon, & i, j \in J'_m \end{cases}$$

are $\succeq 0$. As $\epsilon \rightarrow +0$, the matrices Δ^ϵ remain bounded; indeed, by construction $\Delta^\epsilon \succeq 0$ and $\Delta^\epsilon \preceq [A_{ij}]_{i,j \in J'_m}$ due to $\mathcal{C}^\epsilon \succeq 0$. Thus, we can find a sequence $\epsilon_t \rightarrow +0$ in such a way that $\Delta^{\epsilon_t} \rightarrow \Delta \equiv \Delta^{m-1}$ as $t \rightarrow \infty$, whence $0 \preceq \mathcal{B}^{\epsilon_t} \rightarrow \mathcal{B}_m(A, \Delta^{m-1})$ and $0 \preceq \mathcal{C}^{\epsilon_t} \rightarrow \mathcal{C}_m(A, \Delta^{m-1})$ as $t \rightarrow \infty$, that is, $\mathcal{B}_m(A, \Delta^{m-1}) \succeq 0, \mathcal{C}_m(A, \Delta^{m-1}) \succeq 0$, as required. \square

Observing that matrix $\mathcal{C} = \mathcal{C}_m(A, \Delta^{m-1})$ belongs to $\mathbf{S}^{(v')}$, where $v' = (v'_1, \dots, v'_{i_{m-1}})^T$, where $v'_i = \min[v_i, i_{m-1} - i]$, $1 \leq i \leq i_{m-1}$, and applying Lemma 3.1 recursively, we arrive at the following result.

Theorem 3.1 *Let $v \in \mathbf{R}^n$ be an integral nonnegative vector such that $i + v_i \leq n$ for all i , let $I = \{i_1 < i_2 < \dots < i_m\}$ be the image of $\{1, 2, \dots, n\}$ 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 $\Delta^k = [\Delta_k]^T = [\Delta_{ij}^k]_{i,j} \in J'_{k+1}$, $k = 1, 2, \dots, m-1$, to a solution of the explicit system of m LMIs

$$\mathcal{B}_k(A, \Delta) \succeq 0, \quad k = 1, \dots, m \quad (10)$$

given by the following recurrence:

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

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

If $k = 1$, terminate, otherwise set

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

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 k from the segment $k_-(i, j) \leq k \leq k_+(i, j)$, and for those k the corresponding entry ij in \mathcal{B}^k is

$$B_{ij}^k = \begin{cases} A_{ij}, & k_-(i, j) = k = k_+(i, j) \\ A_{ij} - \sum_{\nu=k_-(i,j)}^{k_+(i,j)-1} \Delta_{ij}^\nu, & k_-(i, j) = k < k_+(i, j) \\ \Delta_{ij}^{k-1}, & k_-(i, j) < k \leq k_+(i, j) \end{cases} \quad (11)$$

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 $\mathcal{B}_1, \dots, \mathcal{B}_m$ by adding zero rows and columns. We arrive at the following result.

Theorem 3.2 *A matrix $A \in \mathbf{S}^{(v)}$ is $\succeq 0$ if and only if there exist matrices $\Delta^k = [\Delta^k]^T = [\Delta_{ij}^k]_{i,j \in J'_{k+1}}$, $1 \leq k \leq m-1$, such that the matrices $\mathcal{B}_k = \mathcal{B}_k(A, \{\Delta^k\}_{k=1}^{m-1}) = [B_{ij}^k]_{i,j \in J_k}$ given by (11) are $\succeq 0$. Whenever this is the case, one has*

$$\begin{aligned} \Delta^k &\succeq 0, \quad k = 1, \dots, m-1 \\ \sum_{k=1}^{m-1} \text{Tr}(\Delta^k) &\leq \text{Tr}(A). \end{aligned} \quad (12)$$

Let $v \in \mathbf{R}^n$ be a fixed sparsity structure, and J_k, J'_k , $k = 1, \dots, m$, be the corresponding index sets. We set

$$\begin{aligned} \Delta &= \left\{ \Delta = \{ \Delta^k = [\Delta^k]^T = [\Delta_{ij}^k]_{i,j \in J'_{k+1}} \}_{k=1}^{m-1} \right\}, \\ \Delta_\rho &= \left\{ \Delta \in \Delta : \Delta^k \succeq 0, \quad k = 1, \dots, m-1, \quad \sum_{k=1}^{m-1} \text{Tr}(\Delta^k) \leq \rho \right\} \end{aligned} \quad (13)$$

and denote by $\mathcal{B}_k(A, \Delta) = [B_{ij}^k(A, \Delta)]_{i,j \in J_k}$ the linear matrix-valued functions of $A \in \mathbf{S}^{(v)}$, $\Delta \in \Delta$ defined by (11). Finally, let

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

The following proposition will be used in section 4.

Proposition 3.2 Let $A \in \mathbf{S}^{(v)}$, $\Delta \in \mathbf{\Delta}$ be such that $\lambda_{\min}(A, \Delta) = -\lambda < 0$. Then $A \succeq -\lambda\mathcal{K}$, where \mathcal{K} is given by (3).

Proof. Let $\widehat{\Delta}_{ij}^k = \begin{cases} \Delta_{ij}^k, & i \neq j \\ \Delta_{ij}^k + \lambda, & i = j \end{cases}$, and let $\widehat{A} = A + \lambda\mathcal{K}$. By (11), we have

$$i, j \in J_k \Rightarrow B_{ij}^k(\widehat{A}, \widehat{\Delta}) - B_{ij}^k(A, \Delta) = \lambda\delta_j^i,$$

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 $A \in \mathbf{S}^{(v)}$ as solvability of certain system \mathcal{S} of LMIs in variables A and additional matrix variables Δ^k , $k = 1, \dots, m-1$. 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 $S_k = |J_k|$, $k = 1, \dots, m$.
2. *Number of additional variables.* Let $d_k = i_k - i_{k-1}$, $k = 1, \dots, m$. 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 3.2.

Let $d = (d_0, d_1, \dots, d_\mu)$ be a *staircase structure* - collection of integers with $d_0 \geq 0$ and $d_1, \dots, d_\mu > 0$, and let $|d| = d_0 + \dots + d_\mu$. Collection d defines the subspace $\mathbf{S}^{[d]}$ 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 $A = A^T$ and $A_{ij} = 0$ for $0 < i < j - 1$:

$$A \in \mathbf{S}^{[d]} \Leftrightarrow A = \begin{bmatrix} A_{0,0} & A_{0,1} & A_{0,2} & \dots & A_{0,\mu-1} & A_{0,\mu} \\ A_{0,1}^T & A_{1,1} & A_{1,2} & & & \\ A_{0,2}^T & A_{1,2}^T & A_{2,2} & A_{2,3} & & \\ \vdots & & \ddots & \ddots & \ddots & \\ A_{0,\mu-1}^T & & & A_{\mu-2,\mu-1}^T & A_{\mu-1,\mu-1} & A_{\mu-1,\mu} \\ A_{0,\mu}^T & & & & A_{\mu-1,\mu}^T & A_{\mu,\mu} \end{bmatrix}.$$

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

$$v_i = \begin{cases} |d| - i, & i \leq d_0 \\ \sum_{j=0}^{k+1} d_j - i, & \sum_{j=0}^{k-1} d_j < i \leq \sum_{j=0}^k d_j \text{ for } k = 1, \dots, \mu - 1 \\ |d| - i, & \sum_{j=0}^{\mu-1} d_j < i \leq |d| \end{cases}$$

We also see that there are $m = \mu - 1$ elements $i_1 < \dots < i_m$ in I given by $i_k = \sum_{j=0}^{k+1} d_j$ for $k = 1, \dots, m$. Using Theorem 3.2, we immediately have the following result.

Proposition 3.3 *A d -staircase matrix $A = [A_{ij}]_{i,j=0}^{\mu}$ is positive semidefnite if and only if there exists*

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

such that

$$\begin{aligned} & \left[\begin{array}{c|c|c} A_{0,0} - \sum_{j=1}^{\mu-2} \Delta_{0,0}^j & A_{0,1} & A_{0,2} - \Delta_{0,1}^1 \\ \hline A_{0,1}^T & A_{1,1} & A_{1,2} \\ \hline A_{0,2}^T - [\Delta_{0,1}^1]^T & A_{1,2}^T & A_{2,2} - \Delta_{2,2}^1 \end{array} \right] \succeq 0, \\ & \left[\begin{array}{c|c|c} \Delta_{0,0}^{j-1} & \Delta_{0,1}^{j-1} & A_{0,j+1} - \Delta_{0,1}^j \\ \hline [\Delta_{0,1}^{j-1}]^T & \Delta_{1,1}^{j-1} & A_{j,j+1} \\ \hline A_{0,j+1}^T - [\Delta_{0,1}^j]^T & A_{j,j+1}^T & A_{j+1,j+1} - \Delta_{1,1}^j \end{array} \right] \succeq 0, \quad j = 2, \dots, \mu - 2 \\ & \left[\begin{array}{c|c|c} \Delta_{0,0}^{\mu-2} & \Delta_{0,1}^{\mu-2} & A_{0,\mu} \\ \hline [\Delta_{0,1}^{\mu-2}]^T & \Delta_{1,1}^{\mu-2} & A_{\mu-1,\mu} \\ \hline A_{0,\mu}^T & A_{\mu-1,\mu}^T & A_{\mu,\mu} \end{array} \right] \succeq 0. \end{aligned}$$

3.2 Positive semidefnite completion of matrices from $\mathbf{S}^{(v)}$

The cone $\mathbf{S}_+^{(v)}$ of positive semidefnite matrices from $\mathbf{S}^{(v)}$ is the intersection of the positive semidefnite cone \mathbf{S}_+^n and the linear subspace $\mathbf{S}^{(v)} \subset \mathbf{S}^n$. Since this subspace clearly intersects the interior of \mathbf{S}_+^n , the cone $\mathbf{C}^{(v)}$ of matrices from $\mathbf{S}^{(v)}$ which is dual to $\mathbf{S}_+^{(v)}$ w.r.t. the Frobenius inner product is exactly the cone of matrices Z from $\mathbf{S}^{(v)}$ admitting positive semidefnite completion, that is, those Z which can be made positive semidefnite by replacing ‘‘hard zero’’ entries ij (those with $j > i + v_i$ or $i > j + v_j$) with appropriately chosen nonzero entries. Proposition 3.1 implies the following result.

Proposition 3.4 *A matrix $Z = [Z_{ij}]_{i,j=1}^m \in \mathbf{S}^{(v)}$ belongs to $\mathbf{C}^{(v)}$ if and only if all matrices $[Z_{ij}]_{i,j \in J_k}$, $k = 1, 2, \dots, m$, are $\succeq 0$.*

Proof. By Proposition 3.1, $Z \in \mathbf{C}^{(v)}$ if and only if the optimal value in the optimization problem

$$\min_{\{X_{ij}^k\}_{i,j \in J_k} \succeq 0\}_{k=1}^m} \left\{ \text{Tr} \left(Z \sum_{k=1}^m X_{ij}^k \right) \right\} \quad (P)$$

is ≥ 0 . Since the problem clearly is strictly feasible and homogeneous, this is so if and only if the semidefnite dual of (P) is feasible, that is, if and only if

$$\exists \{ [Z_{ij}^k = Z_{ji}^k]_{i,j \in J_k} \succeq 0 \}_{k=1}^m : \sum \text{Tr}([Z_{ij}^k][B_{ij}^k]) \equiv \text{Tr}(Z \sum_{k=1}^m B_{ij}^k),$$

where \equiv holds true identically is $\{[B_{ij}^k = B_{ji}^k]_{i,j \in J_k}\}_{k=1}^m$, which is equivalent to $Z_{ij} = Z_{ij}^k$, $i, j \in J_k$, $k = 1, \dots, m$. In other words, $Z \in \mathbf{C}^{(v)}$ if and only if $[Z_{ij}]_{i,j \in J_k} \succeq 0$, $k = 1, \dots, m$. ■

Remark 3.1 *The result stated in Proposition 3.4 can be obtained, with a moderate effort, from the results of [4] on necessary and sufficient conditions for a partially defined symmetric matrix to admit positive semidefinite completion.*

Corollary 3.1 *For $A \in \mathbf{S}^{(v)}$ one has*

$$\lambda_{\max}(A) = \max_Y \left\{ \text{Tr}(AY) : Y \in \mathbf{S}^{(v)}, \text{Tr}(Y) = 1, [Y_{ij}]_{i,j \in J_k} \succeq 0, k = 1, 2, \dots, m \right\}. \quad (14)$$

Indeed, for $A \in \mathbf{S}^n$ we have $\lambda_{\max}(A) = \max_Y \{ \text{Tr}(AY) : Y \in \mathbf{S}_+^n, \text{Tr}(Y) = 1 \}$; when $A \in \mathbf{S}^{(v)}$, the latter formula clearly can be rewritten as $\lambda_{\max}(A) = \max_Y \{ \text{Tr}(AY) : Y \in \mathbf{C}^{(v)}, \text{Tr}(Y) = 1 \}$. Invoking Proposition 3.4, we arrive at (14).

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 3.4.

Proposition 3.5 *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*

$$\left[\begin{array}{c|c|c} B_{0,0} & B_{0,j} & B_{0,j+1} \\ \hline B_{0,j}^T & B_{j,j} & B_{j,j+1} \\ \hline B_{0,j+1}^T & B_{j,j+1}^T & B_{j+1,j+1} \end{array} \right] \succeq 0, \quad j = 1, \dots, \mu - 1. \quad (15)$$

4 Using the representations

In this section, we will use the representations presented in Subsections 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 Lovasz capacity and MAXCUT problems are given in Subsections 4.1, 4.2 and 4.3, respectively.

4.1 Semidefinite programs with well-structured sparse constraint matrices

Let v be a simple sparsity pattern. Consider semidefinite program

$$\text{Opt} = \max_x \left\{ c^T x : x \in X, A[x] \succeq 0 \right\}, \quad (16)$$

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 $\mathbf{S}^{(v)}$.

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 (16);
- A.3. We are given a finite upper bound, ρ , on the quantity

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

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

$$\nu = \max \{t : A[\bar{x}] \succeq t\mathcal{K}\}. \quad (17)$$

We start with the following simple fact (a kind of “exact penalty” statement):

Lemma 4.1 *Let $\mathcal{Y} = \{Y = \{Y^k = [Y_{ij}^k]_{i,j \in J_k}\}_{k=1}^m : Y^k \succeq 0, \sum_k \text{Tr}(Y^k) \leq 1\}$. Given $T \geq 0$, let us associate with (16) the saddle point problem*

$$F_T(x, \Delta) = \min_{Y \in \mathcal{Y}} \left[c^T x + T \sum_{k=1}^m \text{Tr}(Y^k \mathcal{B}_k(A[x], \Delta)) \right] \quad (18)$$

(for the definition of Δ_ρ , see (13)). Assume that

$$T \geq \frac{1}{\nu}(\text{Opt} - c^T \bar{x}). \quad (19)$$

If $(x_\epsilon \in X, \Delta_\epsilon)$ is an ϵ -solution to (18), then the point

$$x^\epsilon = \frac{1}{1+\gamma}x_\epsilon + \frac{\gamma}{1+\gamma}\bar{x}, \quad \gamma = \frac{\max[0, -\lambda_{\min}(A[x_\epsilon], \Delta_\epsilon)]}{\nu}, \quad (20)$$

is a feasible ϵ -solution to (16).

Proof. We clearly have

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

Further, by Theorem 3.2, $A[x]$ with $x \in X$ is $\succeq 0$ if and only if $\max_{\Delta \in \Delta_\rho} \lambda_{\min}(A[x], \Delta) \geq 0$; thus, when x is feasible for (16), we have $\sup_{\Delta \in \Delta_\rho} F_T(x, \Delta) \geq c^T x$, so that the optimal value of (18) is $\geq \text{Opt}$. Consequently, ϵ -optimality of x_ϵ for (18) implies that

$$F_T(x_\epsilon, \Delta_\epsilon) \equiv c^T x_\epsilon + T \min[\lambda_{\min}(A[x_\epsilon], \Delta_\epsilon), 0] \geq \text{Opt} - \epsilon. \quad (21)$$

It is possible that $\lambda_{\min}(A[x_\epsilon], \Delta_\epsilon) \geq 0$; then x_ϵ is feasible for (16) by Theorem 3.2, $x^\epsilon = x_\epsilon$, and (21) says that x^ϵ is ϵ -optimal solution to (16). Now let $\lambda_{\min}(A[x_\epsilon], \Delta_\epsilon) = -\lambda < 0$, so that $\gamma = \frac{\lambda}{\nu}$. Then (21) implies that

$$\begin{aligned} c^T x_\epsilon + \gamma c^T \bar{x} &\geq \text{Opt} - \epsilon + T\lambda + \gamma c^T \bar{x} = \text{Opt}(1 + \gamma) - \epsilon + \lambda T - \gamma[\text{Opt} - c^T \bar{x}] \\ &= \text{Opt}(1 + \gamma) - \epsilon + \lambda[T - \nu^{-1}[\text{Opt} - c^T \bar{x}]] \\ &\geq \text{Opt}(1 + \gamma) - \epsilon \end{aligned}$$

(we have used (19)), whence $c^T x^\epsilon \geq \text{Opt} - \epsilon$. It remains to note that $A[x_\epsilon] \succeq -\lambda\mathcal{K}$ by Proposition 3.2, while $A[\bar{x}] \succeq \nu\mathcal{K}$; it follows that

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

Lemma 4.1 combines with the results of [6] to yield the following

Theorem 4.1 Consider problem (16) 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 simplex $\{x \in \mathbf{R}^N : 0 \leq x, \sum_i x_i \leq R\}$, or the 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 $(\mathbf{S}^{(v)}, \|\cdot\|)$, 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).

Under the outlined assumptions, for every $\epsilon > 0$ one can find a feasible ϵ -solution x_ϵ to (16) (so that $A[x_\epsilon] \succeq 0$ and $c^T x_\epsilon \leq \text{Opt} + \epsilon$) in no more than

$$N(\epsilon) = O(1) \frac{[\text{Opt}^{\text{up}} - c^T \bar{x}] \sqrt{\ln n}}{\nu \epsilon} \times \begin{cases} [\chi R + \rho \sqrt{\ln n}], & \text{case of (a)} \\ [\chi R \sqrt{N} + \rho \sqrt{\ln n}], & \text{case of (b)} \\ [\chi R \sqrt{\ln(N)} + \rho \sqrt{\ln n}], & \text{case of (c)} \end{cases}, \quad (22)$$

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 $|J_k|$, $k = 1, \dots, m$, 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 (17).

Proof. Let $T = \frac{\text{Opt}^{\text{up}} - c^T \bar{x}}{\nu}$. By Lemma 4.1, an ϵ -solution to (16) is readily given by ϵ -solution to the saddle point problem (18) with T we have just defined. Now, problem (18) is of the form

$$\max_{u=(x,\Delta) \in X \times \Delta_\rho} \min_{Y \in \mathcal{Y} \subset \mathbf{S}} [\text{lin}(u, Y) + T \langle \mathcal{A}(x) + \mathcal{D}(\Delta), Y \rangle], \quad (23)$$

where

- $\text{lin}(u, Y)$ is an appropriate affine function of u, Y ,
- $Y = \text{Diag}\{Y^1, \dots, Y^m\}$, $Y^k = [Y_{ij}^k]_{i,j \in J_k}$, $k = 1, \dots, m$, \mathbf{S} is the linear space of all block-diagonal matrices Y of the indicated block-diagonal structure, and $\mathcal{Y} = \{Y \in \mathbf{S} : 0 \preceq Y, \text{Tr}(Y) \leq 1\}$;
- $\mathcal{A}(\cdot)$ 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 . k -th diagonal block $\mathcal{A}^k(x)$ in $A(x)$, $k = 1, \dots, m$, is the contribution of A to $\mathcal{B}_k(A[x], \Delta)$, see (11);
- $\mathcal{D}(\cdot)$ is the linear mapping from the space $\widehat{\mathbf{S}}$ of block-diagonal matrices $\Delta = \text{Diag}\{\Delta^1, \dots, \Delta^{m-1}\}$, $\Delta^\ell = [\Delta_{ij}^\ell]_{i,j \in J'_{\ell+1}}$, into \mathbf{S} defined as follows: 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 (11);
- Δ_ρ is the set of all positive semidefinite matrices from $\widehat{\mathbf{S}}$ with trace $\leq \rho$;
- 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 (23) within any given accuracy $\epsilon > 0$ in no more than

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

steps of the complexity indicated in Theorem 4.1, 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 $\|x\|_X \leq 1$, and let $A = A(x)$, so that $\|A\| \leq \chi$. Invoking (11), 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 $\leq \chi$, so that the “border” is of norm $\leq 2\chi$, whence $\|\mathcal{A}(x)\| \leq 2\chi$. Thus, $L_{XY} \leq 2\chi$. 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, which is the same, is a rank 1 matrix of the Frobenius norm equal to 1. For such a matrix Δ , it follows immediately from (11) 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 $L_{\Delta Y} \leq 2$. Combining our observations, we arrive at (22). ■

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 Lovasz capacity of a graph and to MAXCUT, with emphasis on utilizing favourable sparsity patterns of the underlying graphs.

4.2 Computing Lovasz capacity for a graph with a favourable sparsity pattern

Let $v = (v_1, v_1, \dots, v_{n+1})^T \in \mathbf{R}^{n+1}$ be a simple sparsity structure with $v_1 = n$, and let G be an undirected graph with n nodes, indexed $2, 3, \dots, n+1$, and the set of arcs E such that if $(i, j) \in E$ and $i \leq j$, then $j \leq i + v_i$. Let

$$\mu = \max_{2 \leq i \leq j, j-i \leq v_i} [k_+(i, j) - k_-(i, j) + 1]. \quad (24)$$

Note that μ is exactly the maximum, over nonzero entries ij , $j \geq i \geq 2$, of matrices from $\mathbf{S}^{(v)}$, number of those $k = 1, \dots, m$ for which $i, j \in J_k$.

Consider the Lovasz capacity problem

$$\begin{aligned} \vartheta(G) &= \min_{X, \lambda} \left\{ \lambda : \lambda I_n - ee^T - X \succeq 0, (i, j) \notin E \Rightarrow X_{i-1, j-1} = 0 \right\} \\ &= \min_{X, \lambda} \left\{ \lambda : \left[\begin{array}{c|c} \nu & \sqrt{\nu} e^T \\ \hline \sqrt{\nu} e & \lambda I_n - X \end{array} \right] \succeq 0, (i, j) \notin E \Rightarrow X_{i-1, j-1} = 0 \right\} \end{aligned} \quad (25)$$

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 (25) is given by the Schur Complement Lemma. Let \mathcal{M} be the

affine subspace in $\mathbf{S}^{(v)}$ comprised of all matrices of the form $\left[\begin{array}{c|c} \nu & \sqrt{\nu}e^T \\ \hline \sqrt{\nu}e & Z \end{array} \right]$ with Z constrained by the requirements

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

We equip $\mathbf{S}^{(v)}$ (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 $\mathcal{L} = \text{Diag} \{ \{ \ell^{-1/2}(i, i) \}_{i=1}^{n+1} \}$ (cf. (3)). The norm on $\mathbf{S}^{(v)}$ 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 $\mathbf{S}^{(v)}$ onto \mathcal{M} , so that for any $A \in \mathbf{S}^{(v)}$ one has

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

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 \hat{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 $\hat{\theta} \leq n$ on the Lovasz capacity, consider the following optimization problem:

$$\begin{aligned} \text{Opt} &= \min_{B = \{B_k = B_k^T = [B_{ij}^k]_{i,j \in J_k}\}_{k=1}^m} \left\{ \lambda(B) + T \| \mathcal{S}(B) - \mathcal{P}(\mathcal{S}(B)) \|_{\mathcal{L}} : \begin{array}{l} B_k \succeq 0, k = 1, \dots, m \\ \sum_{k=1}^m \text{Tr}(B_k^2) \leq R^2 \end{array} \right\} \\ \mathcal{S}(B) &= \sum_{k=1}^m B^k, \quad B^k = [B_{ij}^k]_{i,j \in J_k} \\ \lambda(B) &= \frac{\left(\sum_{i=2}^{n+1} \ell^{-1}(i, i) (\mathcal{S}(B))_{ii} \right) \left(\sum_{i=2}^{n+1} \ell^{-1}(i, i) \right)^{-1}}{\left(\sum_{i=2}^{n+1} \ell^{-1}(i, i) \right)^{-1}} = (\mathcal{P}(\mathcal{S}(B)))_{jj}, \quad j = 2, 3, \dots, n+1, \\ R &= \sqrt{\hat{\theta}^2(n + 2\text{Card}(E)) + \nu^2 + 2\nu n}, \end{aligned} \tag{26}$$

where $T \geq 1$.

Observe that

$$\text{Opt} \leq \vartheta(G). \tag{27}$$

Indeed, let X_* be the X -component of the optimal solution to (25). Then the matrix $Y_* = \left[\begin{array}{c|c} \nu & \sqrt{\nu}e^T \\ \hline \sqrt{\nu}e & \vartheta(G)I_n - X_* \end{array} \right]$ is $\succeq 0$ and belongs to $\mathbf{S}^{(v)}$; by Proposition 3.1, 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 (26); at this feasible solution, the objective of (26) clearly is equal to $\lambda(B^*) = \vartheta(G)$, and (27) follows.

Observe also that (26) is nothing but the saddle point problem

$$\min_{B \in \mathbf{B}} \max_{Y \in \mathcal{Y}} F(B, Y), \tag{28}$$

where

$$\begin{aligned}\mathcal{B} &= \{B \in \mathbf{B} : B_k \succeq 0, k = 1, \dots, m, \sum_k \|B_k\|_F^2 \leq R^2\} \\ \mathcal{Y} &= \{Y \in \mathbf{S}^{(v)} : \|Y\|_{\mathcal{L}} \leq 1\} \\ F(B, Y) &= \lambda(B) + T\langle Y, \mathcal{S}(B) - \mathcal{P}(\mathcal{S}(B)) \rangle_{\mathcal{L}}\end{aligned}\tag{29}$$

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 $\mathbf{S}^{(v)}$ equipped with the norm $\|\cdot\|_{\mathcal{L}} \leq 1$.

Since the mapping \mathcal{Q} is of norm ≤ 1 , from the results of [6] the saddle point problem (28) can be solved within accuracy $\epsilon > 0$ in no more than

$$N(\epsilon) = O(1) \frac{TR}{\epsilon}\tag{30}$$

steps, with the computational effort per step dominated by the necessity to find eigenvalue decompositions of m symmetric matrices of the sizes $\text{Card}(J_1), \dots, \text{Card}(J_m)$. Thus, computational effort per step does not exceed

$$\mathcal{C} = O(1) \sum_{k=1}^m \text{Card}^3(J_k).\tag{31}$$

Assume that we have found an ϵ -solution $\tilde{B} = \{\tilde{B}_k\}_{k=1}^m$ to (28), so that

$$\lambda(\tilde{B}) + T \underbrace{\|\mathcal{S}(\tilde{B}) - \mathcal{P}(\mathcal{S}(\tilde{B}))\|_{\mathcal{L}}}_{\delta} \leq \text{Opt} + \epsilon.\tag{32}$$

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

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

where X is of the structure required in (25). Since $\|\Delta\|_{\mathcal{L}} \equiv \|\mathcal{L}^{1/2}\Delta\mathcal{L}^{1/2}\|_F = \delta$, 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 $\mathcal{S}(\tilde{B}) \succeq 0$ results in $C \succeq -\delta\mathcal{L}^{-1}$.

We see that if $\mu = \max_{2 \leq i \leq n+1} \ell(i, i)$, then

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

whence

$$\left[\begin{array}{c|c} \nu & \sqrt{\nu}e^T \\ \hline \sqrt{\nu}e & \frac{\nu + m^{1/2}\delta}{\nu}[\lambda(\tilde{B}) + \mu^{1/2}\delta]I_n - \frac{\nu + m^{1/2}\delta}{\nu}X \end{array} \right] \succeq 0.\tag{33}$$

Thus, a feasible ϵ -solution \tilde{B} to (28) can be easily converted to a feasible solution $(\tilde{\lambda}, \tilde{X} = \frac{\nu + m^{1/2}\delta}{\nu}X)$ to (25) with the value of the objective

$$\begin{aligned}\tilde{\lambda} &= \frac{\nu + m^{1/2}\delta}{\nu}[\lambda(\tilde{B}) + \mu^{1/2}\delta] \\ &\leq \frac{\nu + m^{1/2}\delta}{\nu} \left[\text{Opt} + \epsilon - (T - \mu^{1/2})\delta \right] && \text{[see (32)]} \\ &\leq \frac{\nu + m^{1/2}\delta}{\nu} \left[\vartheta(G) + \epsilon - (T - \mu^{1/2})\delta \right] && \text{[see (27)]} \\ &= \vartheta(G) + \epsilon + \delta \left[\frac{m^{1/2}(\vartheta(G) + \epsilon)}{\nu} - (T - \mu^{1/2}) \frac{\nu + m^{1/2}\delta}{\nu} \right]\end{aligned}\tag{34}$$

We arrive at the following result:

Proposition 4.1 Let $\mu = \max_{2 \leq i \leq n+1} \ell(i, i)$, and let

$$T \geq \mu^{1/2} + \frac{m^{1/2}(\vartheta(G) + \epsilon)}{\nu}. \quad (35)$$

Then ϵ -solution to (29) induces a feasible ϵ -solution to (25). The number of steps required to get such a solution can be bounded by (30), (26), while the computational effort per step can be bounded by (31).

Corollary 4.1 Given an upper bound $\hat{\theta}$ on $\vartheta(G)$, let us set

$$\phi(\nu) = \left(\mu^{1/2} + \frac{m^{1/2}\hat{\theta}}{\nu} \right) \sqrt{\hat{\theta}^2(n + 2\text{Card}(E)) + \nu^2 + 2\nu n}$$

and

$$\bar{\nu} = \underset{\nu > 0}{\text{argmin}} \phi(\nu), \quad \hat{T} = \mu^{1/2} + \frac{m^{1/2}\hat{\theta}}{\bar{\nu}} \quad (36)$$

With $T = \hat{T}$, the outlined procedure allows, for every ϵ , $0 < \epsilon \leq \hat{\theta} - \vartheta(G)$, to find a feasible ϵ -solution to (25) in no more than

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

steps, with the complexity of a step given by (31).

Corollary 4.2 Let $\text{Card}(E) \geq n$. Then, setting

$$\nu = \min \left[\hat{\theta} \sqrt{\text{Card}(E)}, \hat{\theta}^2 \text{Card}(E) n^{-1} \right],$$

one gets

$$N(\epsilon) \leq O(1) \frac{\hat{\theta} \sqrt{m \text{Card}(E)}}{\epsilon}.$$

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

$$\phi(\nu) \leq O(1)\hat{\theta}\sqrt{\text{Card}(E)} \left(\underbrace{\mu^{1/2}}_{\leq m^{1/2}} + \max \left[\underbrace{\frac{m^{1/2}}{\sqrt{\text{Card}(E)}}}_{\leq 1}, \underbrace{\frac{m^{1/2}n}{\hat{\theta}\text{Card}(E)}}_{\leq m^{1/2}} \right] \right) \leq O(1)\hat{\theta}\sqrt{m\text{Card}(E)}. \quad \blacksquare$$

Example: staircase structure. Let m, p be positive integers, and $n = p(m + 1)$. Assume that the incidence matrix of the graph is from $\mathbf{S}^{[d]}$, where $d \in \mathbf{R}^{m+1}$ with $d_i = p$ for $i = 0, \dots, m$. Then, from (25), we see that

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

In the preceding notation, we have $i_k = 1 + (k+2)p$, $k = 1, \dots, m-1$, $\mu = m-1$, $\text{Card}(J_k) = 3p+1$, $\text{Card}(E) \leq O(1)mp^2$. Thus,

$$\mathcal{C} = O(1)mp^3, \quad \phi(\nu) \leq O(1) \left(m^{1/2} + \frac{m^{1/2}\hat{\theta}}{\nu} \right) (\hat{\theta}^2 p^2 m + \nu^2 + \nu pm)^{1/2};$$

Setting $\hat{\nu} = \hat{\theta}pm^{1/2}$, we get

$$\phi(\hat{\nu}) \leq O(1)m^{1/2} (\hat{\theta}^2 p^2 m + p^2 m^{3/2} \hat{\theta})^{1/2} \leq O(1)\hat{\theta}pm (1 + m^{1/2}\hat{\theta}^{-1})^{1/2}.$$

Since the stability number of the corresponding graph clearly is at least $O(m)$, we have $\phi(\hat{\nu}) \leq O(1)\hat{\theta}pm$. Consequently, computing Lovasz capacity within accuracy ϵ costs at most

$$O(1) \frac{\hat{\theta}pm}{\epsilon} \times mp^3 = O(1) \frac{\hat{\theta}p^4 m^2}{\epsilon}$$

operations. For comparison:

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

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

Consider a MAXCUT-type problem

$$\text{Opt} = \max_{X \in \mathbf{S}^n} \{ \text{Tr}(VX) : X \succeq 0, \text{diag}(X) = \mathbf{e} \} \quad (37)$$

where $\text{diag}(A)$ is the diagonal of a square matrix A and \mathbf{e} is the vector of ones. Assume that $V \in \mathbf{S}^{(v)}$ for a given simple sparsity structure v . By Proposition 3.4 problem (37) is equivalent to

$$\text{Opt} = \max_{X \in \mathbf{S}^{(v)}} \left\{ \text{Tr}(VX) : \text{diag}(X) = \mathbf{e}, X^k(X) \equiv [X_{ij}]_{i,j \in J_k} \succeq 0, k = 1, \dots, m \right\}. \quad (38)$$

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

$$\text{Opt}^+ = \max_{X \in \mathcal{X}} \Phi(X) \equiv \min_{Y \in \mathcal{Y}} \left[\text{Tr}(VX) + T \sum_{k=1}^m \text{Tr}(X^k(X)Y^k) \right], \quad (39)$$

where $T > 0$ is a parameter. Observe that the optimal value in (39) is $\geq \text{Opt}$. Indeed, if X_* is an optimal solution to (38), then clearly $X_* \in \mathcal{X}$, and $\Phi(X_*) = \text{Tr}(VX_*)$. Now let X be an ϵ -solution to (39), so that $X \in \mathcal{X}$ and

$$\begin{aligned} \text{Tr}(VX) - T\lambda &\geq \text{Opt}^+ - \epsilon \geq \text{Opt} - \epsilon, \\ \lambda &= \max[0, -\lambda_{\min}(X^1(X)), \dots, -\lambda_{\min}(X^m(X))]. \end{aligned} \quad (40)$$

It is possible that $\lambda = 0$, that is, X is feasible for (38); 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 (38). Setting $\hat{X} = X + \lambda I$, we have

$$\begin{aligned} \text{Tr}(V\hat{X}) &= \text{Tr}(VX) + \lambda\text{Tr}(V) \geq \text{Opt} - \epsilon + \lambda[\text{Tr}(V) + T] \\ \Rightarrow \text{Tr}(V\tilde{X}) &\geq (1 + \lambda)^{-1}[\text{Opt} - \epsilon] + \lambda[\text{Tr}(V) + T] \geq \text{Opt} - \epsilon + (1 + \lambda)^{-1}\lambda[T + \text{Tr}(V) - \text{Opt}] \end{aligned}$$

We see that if

$$T \geq \text{Opt} - \text{Tr}(V), \quad (41)$$

then \tilde{X} is a feasible ϵ -solution to (38). This observation suggests the following scheme for solving (38): given an upper bound Opt^{up} on Opt , we set $T = \text{Opt}^{\text{up}} - \text{Tr}(V)$ and solve saddle point problem (39) within accuracy ϵ , and then convert, in the just presented fashion, the resulting X into a feasible ϵ -solution to (38).

By [6], generating ϵ -solution to (39) costs $O(1)\frac{T\sqrt{\dim \mathbf{S}^{(v)}\sqrt{\ln n}}}{\epsilon}$ steps, with the computational effort per step dominated by the necessity to find eigenvalue decompositions of m matrices $X^k(X)$, $k = 1, \dots, m$, where $X \in \mathcal{X}$. We arrive at the following result:

Proposition 4.2 *Let an upper bound Opt^{up} on the optimal value in (37) be given. For every $\epsilon > 0$, and ϵ -solution to problem (37) with $V \in \mathbf{S}^{(v)}$ can be found in no more than*

$$N(\epsilon) = O(1)\frac{[\text{Opt}^{\text{up}} - \text{Tr}(V)]\sqrt{\ln n}\sqrt{\sum_{i=1}^n(1 + v_i)}}{\epsilon} \quad (42)$$

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

Remark 4.1 *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 $\text{Tr}(V) \leq \text{Opt} \leq 2\text{Tr}(V)$. In this case, one can set $\text{Opt}^{\text{up}} = 2\text{Tr}(V)$, thus converting (42) into the bound $N(\epsilon) \leq O(1)\frac{\text{Opt}}{\epsilon}\sqrt{\ln n}\sqrt{\sum_{i=1}^n(1 + v_i)}$.*

Example: staircase structure. Let m, p be positive integers, and $n = p(m + 1)$. Consider the staircase structure $d = (p, \dots, p) \in \mathbf{R}^{m+1}$, 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 (37) becomes the standard MAXCUT problem associated with (A, G) . By Remark 4.1, the outlined scheme allows to solve the latter problem within any accuracy $\epsilon > 0$ at the arithmetic cost of $O(1)\frac{\text{Opt}}{\epsilon}p^4m^{3/2}\sqrt{\ln(pm)}$ operations. Note that the arithmetic cost of a *single* interior point iteration as applied to the “most economical” dual reformulation of (37), is $O(1)p^3m^3$. It follows that when a “moderate” relative accuracy ϵ/Opt , say, $\epsilon/\text{Opt} = 0.01$ is sought and $m^{3/2} \gg p\sqrt{\ln(pm)}$, 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 $O(1)\sqrt{pm} \ln\left(\frac{\text{Opt}}{\epsilon}\right)p^3m^3$ operations for IPMs (the factor $O(1)\sqrt{pm} \ln\left(\frac{\text{Opt}}{\epsilon}\right)$ is the theoretical bound on the number of IPM iterations required to get ϵ -solution).

5 Numerical implementation

In this section, we present the results of numerical experiments with the Lovasz capacity problem (25) and the (semidefinite relaxation of the) MAXCUT problem (37). These problems were solved by the first order Mirror-Prox algorithm from [6] as applied to the saddle point reformulations (28), respectively, (39), of the problems.

In our experiments, the incidence matrix has staircase structure with $d = (p, \dots, p) \in \mathbf{R}^{m+1}$, with dense $p \times p$ blocks allowed by the structure. Note that the number of nodes in such a graph is $n = (m + 1)p$, while the number of arcs is $\frac{p^2(5m-1)-p(m+1)}{2}$. For our computations, we generated graphs with $p = 2, 3, \dots, 6$ 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. All computations are performed on Supermicro dual-2.66GHz Intel Xeon server with 2GB RAM.

Lovasz capacity problem. When solving this problem according to the scheme developed in section 4.2, one needs an a priori upper bound $\hat{\theta}$ on $\vartheta(G)$. Using the well-know result that Lovasz 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 $\hat{\theta} = m$, and these were the upper bounds used in our computations. The results are presented in Table 1. In the table, 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 Lovasz 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 this Table is identical to the one of Table 1.

References

- [1] Burer, S., Monteiro, R.D.C., Zhang, Y., “Solving a class of semidefinite programs via nonlinear programming”, *Mathematical Programming* 93 (2002), 97–122.
- [2] Burer, S., Monteiro, R.D.C., “A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization”, *Mathematical Programming, Series B* 95 (2003), 329–357.
- [3] Fukuda, M., Kojima, M., Murota, K., Nakata, K., “Exploiting Sparsity in Semidefinite Programming via Matrix Completion I: General Framework”, *SIAM J. on Optimization* 11 (2000), 647–674.
- [4] Grone, B., Johnson, C.R. Marques de Sa, E., Wolkowicz, H., “Positive definite completions of partial Hermitian matrices”, *Linear Algebra Appl.* 58 (1984), 109–124.
- [5] Helmberg, C., Rendl, F., “A Spectral Bundle Method for Semidefinite Programming”, *SIAM J. on Optimization* 10 (2000), 673–696.

| (m, 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 ^{''} |

Table 1: Computational result for the Lovasz capacity problem

| (m, 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 ^h 20'40" |
| (14999, 2) | 30000 | 134988 | 5.775e5 | 5.833e5 | 3963 | 3 ^h 50'19" |
| (19999, 2) | 40000 | 179988 | 7.725e5 | 7.802e5 | 4137 | 5 ^h 10'46" |
| (39999, 2) | 80000 | 359988 | 1.545e6 | 1.560e6 | 5622 | 13 ^h 53'19" |
| (3333, 3) | 10002 | 69987 | 2.862e5 | 2.891e5 | 3447 | 1 ^h 29'56" |
| (6666, 3) | 20001 | 139980 | 5.717e5 | 5.775e5 | 3765 | 3 ^h 14'50" |
| (9999, 3) | 30000 | 209973 | 8.574e5 | 8.660e5 | 4536 | 5 ^h 58'23" |
| (13333, 3) | 40002 | 279987 | 1.146e6 | 1.157e6 | 5355 | 9 ^h 4'41" |
| (26666, 3) | 80001 | 559980 | 2.291e6 | 2.314e6 | 7260 | 24 ^h 24'51" |
| (2499, 4) | 10000 | 94952 | 3.783e5 | 3.821e5 | 2673 | 1 ^h 21'31" |
| (4999, 4) | 20000 | 189952 | 7.585e5 | 7.660e5 | 3531 | 3 ^h 36'46" |
| (7499, 4) | 30000 | 284952 | 1.137e6 | 1.148e6 | 4317 | 6 ^h 49'26" |
| (9999, 4) | 40000 | 379952 | 1.515e6 | 1.530e6 | 4773 | 9 ^h 42'54" |
| (19999, 4) | 80000 | 759952 | 3.028e6 | 3.058e6 | 6393 | 25 ^h 53'39" |
| (1999, 5) | 10000 | 119925 | 4.703e5 | 4.750e5 | 3012 | 1 ^h 53'49" |
| (3999, 5) | 20000 | 239925 | 9.423e5 | 9.517e5 | 3177 | 3 ^h 53'26" |
| (5999, 5) | 30000 | 359925 | 1.417e6 | 1.431e6 | 3741 | 9 ^h 6'1" |
| (7999, 5) | 40000 | 479925 | 1.885e6 | 1.904e6 | 4338 | 10 ^h 32'40" |
| (15999, 5) | 80000 | 959925 | 3.771e6 | 3.809e6 | 5508 | 26 ^h 32'50" |
| (1666, 6) | 10002 | 144921 | 5.645e5 | 5.701e5 | 2487 | 1 ^h 44'37" |
| (3333, 6) | 20004 | 289950 | 1.127e6 | 1.138e6 | 3153 | 4 ^h 23'17" |
| (4999, 6) | 30000 | 434892 | 1.694e6 | 1.711e6 | 3558 | 9 ^h 42'34" |
| (6666, 6) | 40002 | 579921 | 2.257e6 | 2.279e6 | 4263 | 11 ^h 53'27" |
| (13333, 6) | 80004 | 1159950 | 4.514e6 | 4.559e6 | 5619 | 31 ^h 17'50" |

Table 2: Computational results for the MAXCUT problem

- [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” – to appear in *SIAM J. on Optimization*.
- [7] Nesterov, Yu. (2003), “Smooth minimization of nonsmooth functions” – CORE Discussion Paper 2003/12, February 2003.
- [8] Nesterov, Yu. (2003), “Dual extrapolation and its applications for solving variational inequalities and related problems” – CORE Discussion Paper 2003/68, September 2003. <http://www.core.ucl.ac.be/services/COREdp03.html>