Linear Algebra and its Applications 435 (2011) 1446–1463

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

A class of semidefinite programs with rank-one solutions

Guillaume Sagnol [1]

ZIB (Zuse Institut Berlin), Takustr. 7, 14195 Berlin, Germany

ARTICLE INFO

Article history:
Received 9 October 2009
Accepted 13 March 2011
Available online 29 April 2011

Submitted by H. Schneider

AMS classification:
90C22
62K05

Keywords:
SDP
Semidefinite packing problem
Rank-one solution
Low-rank solutions
SOCP
Optimal experimental design
Multiresponse experiments

ABSTRACT

We show that a class of semidefinite programs (SDP) admits a solution that is a positive semidefinite matrix of rank at most r, where r is the rank of the matrix involved in the objective function of the SDP. The optimization problems of this class are semidefinite packing problems, which are the SDP analogs to vector packing problems. Of particular interest is the case in which our result guarantees the existence of a solution of rank one: we show that the computation of this solution actually reduces to a Second Order Cone Program (SOCP). We point out an application in statistics, in the optimal design of experiments.

1. Introduction

In this paper, we study *semidefinite packing problems*. The latter, which are the semidefinite programming (SDP) analogs to the packing problems in linear programming, can be written as:

$$\begin{aligned} \max \quad & \langle C, X\rangle, \\ \text{s.t.} \quad & \langle M_i, X\rangle \leqslant b_i, \quad i \in [l], \\ & X \succeq 0, \end{aligned} \tag{P}$$

where $C \succeq 0$, and $M_i \succeq 0,\ i \in [l]$. The notation $X \succeq 0$ indicates that X belongs to the set $\mathbb{S}_n^+$ of $n \times n$ symmetric positive semidefinite matrices. Similarly, $X \succ 0$ stands for $X \in \mathbb{S}_n^{++}$, the set of $n \times n$ symmetric positive definite matrices. The space of $n \times n$ symmetric matrices $\mathbb{S}_n$ is equipped with the

E-mail address: sagnol@zib.de

[1] This work was done when the author was with INRIA Saclay and CMAP, École Polytechnique, being supported by the research contract CRE EB 257676 with Orange Labs.

0024-3795/$ - see front matter
doi:10.1016/j.laa.2011.03.027

inner product $\langle A, B\rangle = \mathrm{trace}(A^T B)$. We also make use of the standard notation $[l] := \{1, \dots, l\}$, and we use boldface letters to denote vectors. We denote the nullspace (resp. the range) of a matrix A by $\mathrm{Ker}\ A$ (resp. $\mathrm{Im}\ A$).

Semidefinite packing problems were introduced by Iyengar et al. [9]. They showed that these arise in many applications such as relaxations of combinatorial optimization problems or maximum variance unfolding, and gave an algorithm to compute approximate solutions, which is faster than the commonly used interior point methods.

Our main result is that when the matrix C is of rank r, Problem (P) has a solution that is of rank at most r (Theorem 2). In particular, when $r = 1$, the optimal SDP variable X can be factorized as $\boldsymbol{x}\boldsymbol{x}^{\boldsymbol{T}}$, and we show that finding $\boldsymbol{x}$ reduces to a Second Order Cone Program (SOCP) which is computationally more tractable than the initial SDP. We present this result and some applications in Section 2. Then, we extend our result to a wider class of semidefinite programs (Theorems 5 and 6), in which not all the constraints are of packing type. The proofs of the results of Section 2.1 are given in Section 4. Theorems 5 and 6 are proved in Appendix.

1.1. Related work

Solutions of small rank of semidefinite programs have been extensively studied over the past years. Barvinok [2] and Pataki [13] discovered independently that any SDP with l constraints has a solution X^* whose rank is at most

$$r^* = \left\lfloor \frac{\sqrt{8l+1}-1}{2} \right\rfloor,$$

where $\lfloor\cdot\rfloor$ denotes the integer part. This was one of the motivations of Burer and Monteiro for developing the SDPLR solver [5], which searches a solution of the SDP in the form $X = RR^T$, where R is a $n \times r^*$ matrix. The resulting problem is non-convex, and so the augmented Lagrangian algorithm proposed in [5] is not guaranteed to converge to a global optimum. However, it performs remarkably well in practice, and some conditions which ensure that the returned solution is an optimum of the SDP are provided in [6]. Our result shows that for a semidefinite packing problem in which the matrix C has rank r, one can force the matrix R to be of size $n \times r$ (rather than $n \times r^*$), which can lead to considerable gains in computation time when r is small.

We point out that the ratio between the optimal value of Problem (P) and the value of its best solution of rank one has been studied by Nemirovski et al. [12]. They show that the value v^* of the SDP and the value v_1^* of its best rank-one solution satisfy:

$$v^* \geqslant v_1^* \geqslant \frac{1}{2\ln(2l\mu)} v^*, \quad \text{where } \mu = \min\left(l, \max_{i\in[l]}\ \mathrm{rank}\, M_i\right). \tag{1}$$

This ratio can be considerably reduced in particular configurations, but to the best of our knowledge, the fact that the gap in (1) vanishes when the matrix C in the objective function is of rank 1 is new, except in the particular case in which every M_i is of rank 1, too [16].

2. Main result and consequences

In this section, we state the main result of this article and point out an application to statistics. We also discuss the significance of our result for combinatorial optimization problems (the hypothesis on the rank of the matrix C appears to be very restrictive). The results of this section are proved in Section 4.

2.1. The main result

We start with an algebraic characterization of the semidefinite packing problems that are feasible and bounded.

Theorem 1. *Problem* (P) *is feasible if and only if every b_i is nonnegative. Moreover if Problem* (P) *is feasible, then this problem is bounded if and only if the range of C is included in the range of $\sum_i M_i$.*

The reader should note that the range inclusion condition in Theorem 1 is in fact equivalent to the feasibility of the Lagrangian dual of Problem (P):

$$\begin{aligned} \min_{\boldsymbol{\mu}\geqslant 0} \quad & \boldsymbol{\mu}^{\boldsymbol{T}}\boldsymbol{b}, \\ \text{s.t.} \quad & \sum_i \mu_i M_i \succeq C. \end{aligned} \tag{D}$$

The main result of this article follows:

Theorem 2. *We assume that the conditions of Theorem 1 are fulfilled, so that Problem* (P) *is feasible and bounded. If* rank $C = r$, *then the semidefinite packing problem* (P) *has a solution which is a matrix of rank at most r.*

A consequence of Theorem 2 is that when the matrix in the objective function is of rank 1 ($C = \boldsymbol{c}\boldsymbol{c}^T$), the computation of a solution X of Problem (P) reduces to the computation of a vector $\boldsymbol{x}$ such that $X = \boldsymbol{x}\boldsymbol{x}^{\boldsymbol{T}}$. The next result shows that this can be done very efficiently by a Second Order Cone Program (SOCP).

Corollary 3. *We assume that the conditions of Theorem 1 are fulfilled, and that $C = \boldsymbol{c}\boldsymbol{c}^{\boldsymbol{T}}$ for a vector $\boldsymbol{c} \in \mathbb{R}^n$ (i.e.* rank $C = 1$*). Then, Problem* (P) *reduces to the SOCP:*

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{x}, \\ s.t. \quad & \|A_i\boldsymbol{x}\|_2 \leqslant \sqrt{b_i}, \quad i = 1 \in [l], \end{aligned} \tag{2}$$

where the matrices A_i are such that $M_i = A_i^T A_i$. Moreover, if $\boldsymbol{x}$ is any optimal solution of Problem (2)*, then $X = \boldsymbol{x}\boldsymbol{x}^{\boldsymbol{T}}$ is an optimal solution of Problem* (P)*, and the optimal value of* (P) *is $(\boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{x})^2$.*

Proof. The SOCP (2) is simply obtained from (P) by substituting $\boldsymbol{x}\boldsymbol{x}^{\boldsymbol{T}}$ from X and $A_i^T A_i$ from M_i. The objective function $\langle C, X\rangle$ becomes $(\boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{x})^2$, and we can remove the square by noticing that $\boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{x} \geqslant 0$ without loss of generality, since if $\boldsymbol{x}$ is optimal, so is $-\boldsymbol{x}$. □

In fact, the proof of Theorem 2 relies on the projection of Problem (P) on an appropriate subspace, which lets the reduced semidefinite packing problem be strictly feasible, as well as its dual. This reduction is not only of theoretical interest, since in some cases it may yield some important computational savings. Therefore, we next state this result as a proposition.

Let $\mathcal{I}_0 := \{i \in [l] : b_i = 0\}$ and $\mathcal{I} := [l]\backslash\mathcal{I}_0$. Let the columns of the $n \times n_0$ matrix U form an orthonormal basis of $\operatorname{Im}(\sum_{i\in[l]} M_i)$, and the columns of the $n_0 \times n'$ matrix V form an orthonormal basis of $\operatorname{Ker}(U^T \sum_{i\in\mathcal{I}_0} M_i U)$. We further define $C' := (UV)^T C(UV) \in \mathbb{S}_{n'}^+$ and $M_i' := (UV)^T M_i(UV) \in \mathbb{S}_{n'}^+$ (for $i \in \mathcal{I}$), and we consider the reduced problem

$$\begin{aligned} \max_{Z\in\mathbb{S}_{n'}^+} \quad & \langle C', Z\rangle, \\ \text{s.t.} \quad & \langle M_i', Z\rangle \leqslant b_i, \quad i \in \mathcal{I}. \end{aligned} \tag{P$'$}$$

Proposition 4. *We assume that the conditions of Theorem 1 are fulfilled, so that Problem* (P) *is feasible and bounded. Then, the following properties hold:*

(i) Problem (P′) *is strictly feasible, i.e.* $\exists \overline{Z} \succ 0 : \forall\, i \in \mathcal{I}, \langle M_i', \overline{Z}\rangle < b_i$.
(ii) The Lagrangian dual of (P′) *is strictly feasible, i.e.* $\exists \overline{\boldsymbol{\mu}} > \boldsymbol{0} : \sum_{i\in\mathcal{I}} \overline{\mu}_i M_i' \succ C'$.

(iii) If Z is a solution of Problem (P′), then $X := (UV)Z(UV)^T$ is an optimal solution of Problem (P) (which of course satisfies $\operatorname{rank} X \leqslant \operatorname{rank} Z$ *and* $\langle C, X\rangle = \langle C', Z\rangle$*).*

The present work grew out from an application to networks [4], in which the traffic between any two pairs of nodes must be inferred from a set of measurements. This can be modeled by the theory of optimal experimental design, which leads to a large SDP. Standard solvers relying on interior points methods, like SeDuMi [20], cannot handle problems of this size. However, in a followup work relying on the present reduction to an SOCP [19], we solve within seconds the same instances in SeDuMi. We next present this application.

2.2. Application to the optimal design of experiments

An interesting application arises in statistics, in the design of optimal experiments (for more details on the subject, the reader is referred to Pukelsheim [15]). An experimenter wishes to estimate the quantity $\boldsymbol{c^T\theta}$, where $\boldsymbol{\theta}$ is an unknown n-dimensional parameter, and $\boldsymbol{c}$ is a vector of n coefficients. To this end, she disposes of l available experiments, each one giving a linear measurement of the parameter $\boldsymbol{y_i} = A_i\boldsymbol{\theta}$, up to a (centered) measurement noise. If the amount of experimental effort spent on the ith experiment is w_i, it is known that the variance of the best linear unbiased estimator for $\boldsymbol{c^T\theta}$ is $\boldsymbol{c^T}(\sum_i w_iM_i)^\dagger\boldsymbol{c}$, where $M_i = A_i^TA_i$, and $M^\dagger$ denotes the Moore–Penrose inverse of M. The problem of distributing the experimental effort so as to minimize this variance is called the "$\boldsymbol{c}$-optimal problem", and can be formulated as:

$$\begin{aligned} \min_{\boldsymbol{w}\geqslant\boldsymbol{0}} \quad & \boldsymbol{c^T}\left(\sum_i w_iM_i\right)^\dagger\boldsymbol{c}, \\ \text{s.t.} \quad & \sum_{i=1}^{l} w_i = 1. \end{aligned} \tag{3}$$

It is classical to reformulate this problem as a semidefinite program, by using the Schur complement lemma and duality theory (see e.g. [16,18]). The c-optimal SDP already appeared in Pukelsheim and Titterington [14], hidden under a more general form:

$$\begin{aligned} \max \quad & \boldsymbol{c^T}X\boldsymbol{c}, \\ \text{s.t.} \quad & \langle M_i, X\rangle \leqslant 1, \quad i \in [l], \\ & X \succeq 0. \end{aligned} \tag{4}$$

In this problem, the design variable $\boldsymbol{w}$ is proportional to the dual variable associated to the constraints $\langle M_i, X\rangle \leqslant 1$. Note that this is a semidefinite packing problem, in which the matrix defining the objective function has rank 1 ($C = \boldsymbol{cc^T}$). More generally, if we want to estimate simultaneously r linear functions of the parameter $\boldsymbol{\zeta} = \left(\boldsymbol{c_1^T\theta}, \dots, \boldsymbol{c_r^T\theta}\right)$, the best unbiased estimator $\hat{\zeta}$ is now an r-dimensional vector with covariance matrix

$$\mathrm{Cov}_{\boldsymbol{w}}(\hat{\boldsymbol{\zeta}}) := K^T\left(\sum_{k=1}^{l} w_kM_k\right)^\dagger K,$$

where $K = [\boldsymbol{c_1}, \dots, \boldsymbol{c_r}]$. Several criteria can be used for this experimental design problem. Popular ones are the A-criterion and the E-criterion, which aim at minimizing respectively the trace and the largest eigenvalue of $\mathrm{Cov}_{\boldsymbol{w}}(\hat{\boldsymbol{\zeta}})$. These optimization problems can also be formulated as semidefinite packing problems. For A-optimality, this *packing* formulation is given in [18]:

$$\begin{aligned} \max \quad & \boldsymbol{\tilde{c}^T}X\boldsymbol{\tilde{c}}, \\ \text{s.t.} \quad & \langle \tilde{M}_i, X\rangle \leqslant 1, \quad i \in [l], \\ & X \succeq 0, \end{aligned} \tag{5}$$

where $\tilde{\boldsymbol{c}} = \left[\boldsymbol{c_1^T}, \ldots, \boldsymbol{c_r^T}\right]^T$, and $\tilde{M}_i$ is a block-diagonal matrix which contains r times the block M_i on its main diagonal. The matrix in the objective function is of rank 1 ($C = \tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^{\boldsymbol{T}}$), and so Problem (5) reduces to a SOCP by Corollary 3. This reduction is of great interest for the computation of optimal experimental designs, because SOCP solvers are much more efficient than SDP solvers, and take advantage of the sparsity of the matrices A_i (whereas the matrices $M_i = A_i^T A_i$ used in the original SDP formulation (5) are *not very sparse* in general).

The E-optimal design SDP is presented in [22] (for the special case in which $C = I$), and takes exactly the form of the semidefinite packing problem (P), with $b_i = 1$ for all $i \in [l]$ and $C = KK^T = \sum_{i=1}^r \boldsymbol{c_i}\boldsymbol{c_i}^T$. Here, the matrix C has rank r, and so Theorem 2 indicates that the E-optimal design SDP has a solution which is a matrix of rank at most r. This suggests the use of specialized low rank solvers for this SDP when r is small (cf. Section 1.1 at the end of the introduction), which can lead to a considerable improvement in terms of computation time.

2.3. Relation with combinatorial optimization

SDP relaxations of combinatorial optimization problems have motivated the authors of [9] to study semidefinite packing problems. Hence, we discuss the significance of our result for this class of problems in this section.

Semidefinite programs have been used extensively to formulate relaxations of NP-hard combinatorial optimization problems after the work of Goemans and Williamson on the approximability of MAXCUT [8]. These SDP relaxations often lead to optimal solutions of the related combinatorial optimization problems whenever the solution of the SDP is of small rank. As shown by Iyengar et al. [9], SDP relaxations of many combinatorial optimization problems can be cast as semidefinite packing programs. Our result therefore identifies a subclass of combinatorial optimization problems which are solvable in polynomial time. Unfortunately, this promising statement only helped us to identify trivial instances so far. For example, the MAXCUT semidefinite packing problem [9] yields an exact solution of the combinatorial problem whenever it has a rank 1 solution. The matrix C in the objective function of this SDP is the Laplacian of the graph, and so it is known that

$$\operatorname{rank} C = N - \kappa,$$

where N is the number of vertices and κ is the number of connected components in the graph. Our result therefore states that if a graph of N vertices has $N - 1$ connected components, then it defines a MAXCUT instance that is solvable in polynomial time. Such graphs actually consist in a pair of connected vertices, plus $N - 2$ isolated vertices, and the related MAXCUT instance is trivial.

Another limitation for the application of our theorem in this field is that most semidefinite packing problems arising in combinatorial optimization (including but not limited to the Lovász ϑ function SDP [11] and the related Szegedy number SDP [21], the vector coloring SDP [10], the sparsest cut SDP [1] and the sparse principal components analysis SDP [7]) can be written in the form of (P), with an additional trace equality constraint trace $(X) = 1$. In fact, we can show that if such an "equality constrained" problem is strictly feasible, then it is equivalent to the following "classical" semidefinite packing problem:

$$\begin{aligned} \max \quad & \langle C + \lambda I, X\rangle - \lambda \\ \text{s.t.} \quad & \langle M_i, X\rangle \leqslant b_i, \quad i \in [l], \\ & \operatorname{trace} X \leqslant 1, \\ & X \succeq 0, \end{aligned} \tag{6}$$

where λ is any scalar larger than $|\lambda^*|$, where λ^* is the optimal Lagrange multiplier associated to the constraint trace $(X) = 1$ (we omit the proof of this statement which is of secondary importance in this article). Since $C + \lambda I$ is a full rank matrix, our result does not seem to yield any valuable information for this class of problems.

3. Extension to “combined” problems

The proof of our main result also applies to a wider class of semidefinite programs, which can be written as:

$$\begin{aligned} \sup_{X,Y,\boldsymbol{\lambda}} \quad & \langle C, X\rangle + \langle R_0, Y\rangle + \boldsymbol{h_0}^T\boldsymbol{\lambda}, \\ \text{s.t.} \quad & \langle M_i, X\rangle \leqslant b_i + \langle R_i, Y\rangle + \boldsymbol{h_i}^T\boldsymbol{\lambda}, \quad i \in [l], \\ & X \in \mathbb{S}_n^+,\ Y \in \mathbb{S}_p^+,\ \boldsymbol{\lambda} \in \mathbb{R}^q, \end{aligned} \tag{P_{CMB}}$$

where **every matrix** M_i **and** C **are positive semidefinite, while the** R_i **are *arbitrary symmetric* matrices**. The vectors $\boldsymbol{h_i}$ are in $\mathbb{R}^q$. We denote by H the $q \times l$ matrix formed by the columns $\boldsymbol{h_1}, \ldots, \boldsymbol{h_l}$. The Lagrangian dual of Problem (P_{CMB}) is:

$$\begin{aligned} \inf_{\boldsymbol{\mu}\geqslant 0} \quad & \boldsymbol{b}^T\boldsymbol{\mu}, \\ \text{s.t.} \quad & \sum_{i=1}^{l} \mu_i M_i \succeq C, \\ & R_0 + \sum_{i=1}^{l} \mu_i R_i \preceq 0. \\ & \boldsymbol{h_0} + H\boldsymbol{\mu} = \boldsymbol{0}. \end{aligned} \tag{D_{CMB}}$$

We have seen in Section 2.1 that the feasibility of both the primal (P) and the dual (D) is sufficient to guarantee that Problem (P) has a solution of rank at most $r := \operatorname{rank} C$. For *combined* problems however, the feasibility of the couple of programs (P_{CMB})–(D_{CMB}) is not sufficient to guarantee the existence of a solution $(X, Y, \boldsymbol{\lambda})$ of Problem (P_{CMB}) in which $\operatorname{rank} X \leqslant r$. We give indeed an example (Example 1) where the optimum in Problem (P_{CMB}) is not even attained. However, we show in the next theorem that an asymptotic result subsists. Moreover, we shall see in Theorem 6 that a solution in which X is of rank at most r exists as soon as an additional condition holds (strict dual feasibility). The proof of Theorem 6 essentially mimics that of Theorem 2 and is therefore presented in Appendix A. Theorem 5 turns out to be a consequence of Theorem 6 and is proved in Appendix B.

Theorem 5. *We assume that Problems* (P_{CMB}) *and* (D_{CMB}) *are feasible. If* $\operatorname{rank} C = r$, *then there exists a sequence of feasible primal variables* $(X_k, Y_k, \boldsymbol{\lambda}_k)_{k\in\mathbb{N}}$ *such that* $\operatorname{rank} X_k \leqslant r$ *for all* $k \in \mathbb{N}$ *and* $\langle C, X_k\rangle + \langle R_0, Y_k\rangle + \boldsymbol{h_0}^T\boldsymbol{\lambda}_k$ *converges to the optimum of Problem* (P_{CMB}) *as* $k \to \infty$.

Theorem 6. *We assume that Problem* (P_{CMB}) *is feasible, and a refined Slater condition holds for Problem* (D_{CMB}), *i.e. there is a feasible dual variable which strictly satisfies the non-affine constraints:*

$$\exists \overline{\boldsymbol{\mu}} \geqslant \boldsymbol{0} : \sum_i \overline{\mu}_i M_i \succ C, \quad R_0 + \sum_i \overline{\mu}_i R_i \prec 0, \quad \boldsymbol{h_0} + H\overline{\boldsymbol{\mu}} = \boldsymbol{0}.$$

If $\operatorname{rank} C = r$, *then Problem* ($P_{\text{CMB}}$) *has a solution* $(X, Y, \boldsymbol{\lambda})$ *in which* $\operatorname{rank} X \leqslant r$. *Moreover, if* $C \neq 0$, *then every solution* $(X, Y, \boldsymbol{\lambda})$ *of Problem* (P_{CMB}) *is such that* $\operatorname{rank} X \leqslant n - \overline{r} + r$, *where* $\overline{r} := \min_{i\in[l]} \operatorname{rank} M_i$.

Example 1. Consider the following *combined* semidefinite packing problem:

$$\begin{aligned} \sup_{X\in\mathbb{S}_2^+,\ \boldsymbol{\lambda}\in\mathbb{R}^2} \quad & \frac{3}{100}\left\langle \begin{pmatrix} 81 & 9 \\ 9 & 1 \end{pmatrix}, X \right\rangle - \lambda_1 - 3\lambda_2, \\ \text{s.t.} \quad & 0 \leqslant 1 + \lambda_1, \\ & X_{1,1} \leqslant 1 + \lambda_2, \\ & X_{2,2} \leqslant 1 + 3\lambda_1 + \lambda_2. \end{aligned} \tag{7}$$

This problem is in the form of (P_{CMB}) indeed, with $C = \boldsymbol{c}\boldsymbol{c}^T$, $\boldsymbol{c} = \frac{\sqrt{3}}{10}[9 \quad 1]^T$, $\boldsymbol{h_0} = [-1 \quad -3]^T$,

$$M_1 = 0,\ M_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 1 \end{pmatrix}.$$

Problem (7) is clearly feasible (e.g. for $X = 0, \boldsymbol{\lambda} = \boldsymbol{0}$), and the reader can verify that $\boldsymbol{\mu} = \frac{1}{10}[1 \quad 27 \quad 3]^T$ is dual feasible (in fact, this is the only dual feasible vector, and hence the dual problem does not satisfy the Slater constraints qualification). The value of the optimum is $\frac{31}{10}$, and can be approached arbitrarily closely for the sequence of feasible variables $(\boldsymbol{x}_k\boldsymbol{x}_k^T, \boldsymbol{\lambda_k})_{k\in\mathbb{N}}$, where for all $k \geqslant 0$, $\boldsymbol{x}_k = [\sqrt{3+k} \quad \sqrt{k}]^T$, $\boldsymbol{\lambda}_k = [-1 \quad k+2]^T$, while this optimum is not attained by any couple $(X, \boldsymbol{\lambda})$ of (bounded) feasible variables.

As in the previous section, we have a result of reduction to a SOCP, which holds when C is of rank 1, every $R_i = 0$ and $\boldsymbol{h_0} = \boldsymbol{0}$. Recall that H denotes the matrix formed by the columns $\boldsymbol{h_1}, \ldots, \boldsymbol{h_l}$.

Corollary 7. *Consider the following "combined" semidefinite packing problem:*

$$\begin{aligned} \sup_{X\in\mathbb{S}_n,\ \boldsymbol{\lambda}\in\mathbb{R}^q} \quad & \langle C, X\rangle, \\ \text{s.t.} \quad & \langle M_i, X\rangle \leqslant \boldsymbol{h_i}^T\boldsymbol{\lambda} + b_i, \quad i \in [l], \\ & X \succeq 0. \end{aligned} \tag{8}$$

Assume that $C = \boldsymbol{c}\boldsymbol{c}^{\boldsymbol{T}}$ has rank 1. If Problem (8) and its Lagrangian dual are feasible, i.e.

(i) $\exists \overline{\boldsymbol{\lambda}} \in \mathbb{R}^q :\ H^T\overline{\boldsymbol{\lambda}} + \boldsymbol{b} \geqslant 0$;
(ii) $\exists \overline{\boldsymbol{\mu}} \geqslant \boldsymbol{0} :\ \sum_i \overline{\mu}_i M_i \succeq C,\ \boldsymbol{h_0} + H\overline{\boldsymbol{\mu}} = \boldsymbol{0}$,

then, Problem (8) is bounded, and its optimal value is the square of the optimal value of the following SOCP:

$$\begin{aligned} \sup_{\boldsymbol{x}\in\mathbb{R}^n,\ \boldsymbol{\lambda}\in\mathbb{R}^q} \quad & \boldsymbol{c}^T\boldsymbol{x}, \\ \text{s.t.} \quad & \left\| \begin{bmatrix} 2A_i\boldsymbol{x} \\ \boldsymbol{h_i}^T\boldsymbol{\lambda} + b_i - 1 \end{bmatrix} \right\|_2 \leqslant \boldsymbol{h_i}^T\boldsymbol{\lambda} + b_i + 1, \quad i \in [l], \end{aligned} \tag{9}$$

where the matrices A_i are such that $M_i = A_i^T A_i$. Moreover, if $(\boldsymbol{x}, \boldsymbol{\lambda})$ is a solution of Problem (9), then $(\boldsymbol{x}\boldsymbol{x}^{\boldsymbol{T}}, \boldsymbol{\lambda})$ is a solution of Problem (8), and the optimal value of (8) is $(\boldsymbol{c}^T\boldsymbol{x})^2$.

Proof. Theorem 5 guarantees the existence of a sequence of feasible variables $(X_k, \boldsymbol{\lambda}_k)_{k\in\mathbb{N}}$ in which X_k has rank 1, i.e. $X_k = \boldsymbol{x_k}\boldsymbol{x_k}^T$, and $\langle C, X_k\rangle = (\boldsymbol{c}^T\boldsymbol{x}_k)^2$ converges to the optimum of Problem (8). This optimal value is therefore equal to the supremum of $(\boldsymbol{c}^T\boldsymbol{x})^2$, over all the pairs of vectors $(\boldsymbol{x}, \boldsymbol{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^q$ such that $(\boldsymbol{x}\boldsymbol{x}^T, \boldsymbol{\lambda})$ is feasible for Problem (8). As in the proof of Corollary 3, we notice that if $(\boldsymbol{x}\boldsymbol{x}^T, \boldsymbol{\lambda})$ is feasible for Problem (8), so is $((-\boldsymbol{x})(-\boldsymbol{x})^T, \boldsymbol{\lambda})$, hence we can remove the square in the objective function.

The SOCP (9) is simply obtained from (8) by substituting $\boldsymbol{x}\boldsymbol{x}^{\boldsymbol{T}}$ from X and $A_i^T A_i$ from M_i. We also used the fact that for any vector $\boldsymbol{z}$ and for any scalar α, the hyperbolic constraint

$$\|\boldsymbol{z}\|_2^2 \leqslant \alpha$$

is equivalent to the second order cone constraint

$$\left\| \begin{bmatrix} 2\boldsymbol{z} \\ \alpha - 1 \end{bmatrix} \right\|_2 \leqslant \alpha + 1. \quad \square$$

3.1. Application: c-optimal design of experiments with multiple resource constraints

In a more general setting than the classical $\boldsymbol{c}$-optimal design problem (3) presented in the previous section, $\boldsymbol{w}$ no longer represents the percentage of experimental effort to spend on each experiment, but describes some resource allocation to the available experiments, that is subject to multiple linear constraints $P\boldsymbol{w} \leqslant \boldsymbol{d}$, where P is a $q \times l$ matrix with nonnegative entries and $\boldsymbol{d}$ is a $q \times 1$ vector. This problem arises for example in a network-wide optimal sampling problem [19], where $\boldsymbol{w}$ is the vector of the sampling rates of the monitoring devices on all links of the network, and is subject to linear constraints that limit the overhead of the routers. We next show that this problem is a "combined" semidefinite packing problem which reduces to an SOCP. The resource constrained $\boldsymbol{c}$-optimal design problem reads as follows:

$$\begin{aligned} \inf_{\boldsymbol{w}\geqslant\boldsymbol{0}} \quad & \boldsymbol{c}^{\boldsymbol{T}} \left(\sum_i w_i M_i\right)^{\dagger} \boldsymbol{c}, \\ \text{s.t.} \quad & P\boldsymbol{w} \leqslant \boldsymbol{d}. \end{aligned} \tag{10}$$

We assume that the optimal design problem is feasible, i.e. there exists a vector $\hat{\boldsymbol{w}} \geqslant \boldsymbol{0}$ such that $P\hat{\boldsymbol{w}} \leqslant \boldsymbol{d}$ and $\boldsymbol{c}$ is in the range of $\sum_i \hat{w}_i M_i$. Note that we can assume without loss of generality that $\hat{\boldsymbol{w}} > \boldsymbol{0}$. Otherwise, this would mean that the constraints $P\boldsymbol{w} \leqslant \boldsymbol{d},\ \boldsymbol{w} \geqslant \boldsymbol{0}$ force the equality $w_i = 0$ to hold for some coordinate $i \in [l]$, and in this case we could simply remove the experiment i from the set of available experiments.

We can now express the latter problem as an SDP thanks to the Schur complement lemma:

$$\begin{aligned} \inf_{t\in\mathbb{R},\ \boldsymbol{w}\geqslant\boldsymbol{0}} \quad & t \\ \text{s.t.} \quad & \left(\begin{array}{c|c} \sum_i w_i M_i & \boldsymbol{c} \\ \hline \boldsymbol{c}^T & t \end{array}\right) \succeq 0. \\ & P\boldsymbol{w} \leqslant \boldsymbol{d}. \end{aligned} \tag{11}$$

Since the optimal t is positive (we exclude the trivial case $\boldsymbol{c} = \boldsymbol{0}$), the latter matrix inequality may be rewritten as

$$\sum_i w_i M_i \succeq \frac{\boldsymbol{c}\boldsymbol{c}^{\boldsymbol{T}}}{t},$$

by using the Schur complement lemma again. Finally, we make the change of variables $\boldsymbol{\mu} = t\boldsymbol{w}$ and Problem (11) is equivalent to

$$\begin{aligned} \inf_{\boldsymbol{\mu}\geqslant\boldsymbol{0},t\geqslant 0} \quad & t \\ \text{s.t.} \quad & \sum_{i=1}^{l} \mu_i M_i \succeq \boldsymbol{c}\boldsymbol{c}^{\boldsymbol{T}} \\ & P\boldsymbol{\mu} \leqslant t\boldsymbol{d}. \end{aligned} \tag{12}$$

This problem is exactly in the form of Problem (D_{CMB}), for $C = \boldsymbol{c}\boldsymbol{c}^{\boldsymbol{T}}, \mu_{l+1} = t,\ \boldsymbol{b} = [0, \dots, 0, 1]^T \in \mathbb{R}^{l+1},\ M_{l+1} = 0,\ \boldsymbol{h_0} = \boldsymbol{0},\ H = [P, -\boldsymbol{d}]$, and for all $i \in 0, \dots, l+1,\ R_i = 0$ (we also need to introduce a nonnegative slack variable to handle the inequalities as equalities).

Let $\lambda := \boldsymbol{c}^T (\sum_i M_i)^{\dagger} \boldsymbol{c}^T$, so that $\lambda \sum_i M_i \succeq \boldsymbol{c}\boldsymbol{c}^T$. We set $\bar{t} = \max_{i\in[l]}(\lambda/\hat{w}_i)$ ($\bar{t}$ is well defined because $\hat{\boldsymbol{w}} > \boldsymbol{0}$). The vector $\overline{\boldsymbol{\mu}} := \bar{t}\hat{\boldsymbol{w}}$ is dual feasible, because $P\overline{\boldsymbol{\mu}} \leqslant \bar{t}\boldsymbol{d}$, and $\sum_{i=1}^{l} \overline{\mu}_i M_i \succeq \lambda \sum_{i=1}^{l} M_i \succeq \boldsymbol{c}\boldsymbol{c}^{\boldsymbol{T}}$. In addition, the corresponding primal problem is clearly feasible (for $\boldsymbol{\lambda} = \boldsymbol{0}$, since $\boldsymbol{b} \geqslant \boldsymbol{0}$), and thus we can use Corollary 7: the $\boldsymbol{c}$-optimal design problem with resource constraints (10) reduces to the SOCP (9). We give below this SOCP (with the parameters $\boldsymbol{b}$, M_i, H and the slacks defined as above), as well as its dual:

$$\begin{aligned}\sup_{\substack{\boldsymbol{x}\in\mathbb{R}^n\\ \boldsymbol{\lambda}\in\mathbb{R}^q}} \quad & \boldsymbol{c}^T\boldsymbol{x}\\ & \left\|\begin{bmatrix} 2A_i\boldsymbol{x}\\ \boldsymbol{p_i}^T\boldsymbol{\lambda}-1\end{bmatrix}\right\|_2 \leqslant \boldsymbol{p_i}^T\boldsymbol{\lambda}+1 \quad (\forall\, i\in[l]),\\ & \boldsymbol{d}^T\boldsymbol{\lambda}\leqslant 1,\\ & \boldsymbol{\lambda}\geqslant \boldsymbol{0}.\end{aligned} \qquad \begin{aligned}\inf_{\substack{\boldsymbol{\mu}\geqslant\boldsymbol{0},t\geqslant 0\\ \boldsymbol{\alpha}\geqslant\boldsymbol{0},(\boldsymbol{z_i})_{i\in[l]}}} \quad & \sum_{i=1}^l \alpha_i + t,\\ & \sum_{i=1}^l A_i^T\boldsymbol{z_i}=\boldsymbol{c},\\ & P\boldsymbol{\mu}\leqslant t\boldsymbol{d},\\ & \left\|\begin{bmatrix}\boldsymbol{z_i}\\ \alpha_i-\mu_i\end{bmatrix}\right\|_2 \leqslant \alpha_i+\mu_i\\ & \qquad (\forall\, i\in[l]),\end{aligned}$$

where the vectors $\boldsymbol{p_1},\ldots,\boldsymbol{p_l}\in\mathbb{R}^q$ are the columns of the matrix P, and for all $i\in[l]$, A_i is such that $A_i^TA_i=M_i$. The dual problem satisfies the (refined) Slater condition, because $\boldsymbol{c}\in\operatorname{Im}(\sum_i M_i)=\sum_i \operatorname{Im}(A_i^T)$, so that $\exists\bar{\boldsymbol{z}}_{\boldsymbol{1}},\ldots,\bar{\boldsymbol{z}}_{\boldsymbol{l}}:\sum_{i=1}^l A_i^T\bar{\boldsymbol{z}}_{\boldsymbol{i}}=\boldsymbol{c}, P\overline{\boldsymbol{\mu}}\leqslant\bar{t}\boldsymbol{d}$ and for $\overline{\boldsymbol{\alpha}}>\boldsymbol{0}$ large enough, the non-affine cone constraints are satisfied with a strict inequality. Hence, strong duality holds and the values of these two problems are equal. By construction, the optimal design variable $\boldsymbol{w}$ is related to the dual optimal variables $\boldsymbol{\mu}$ and t by the relation $\boldsymbol{w}=t^{-1}\boldsymbol{\mu}$. Moreover, Corollary 7 shows that the optimal value of Problem (10) is the square of the optimal value of these SOCPs.

4. Proofs of the theorems

Proof of Theorem 1. The fact that Problem (P) is feasible if and only if every b_i is nonnegative is clear, since $X=0$ is always feasible in this case and $M_i\succeq 0, X\succeq 0$, implies $\langle M_i,X\rangle\geqslant 0$.

Now, we assume that each b_i is nonnegative, and we show that Problem (P) is bounded if and only if $\operatorname{Im} C\subset\operatorname{Im}\sum_i M_i$. The positive semidefiniteness of the matrices M_i implies that there exists matrices A_i $(i\in[l])$ such that $A_i^TA_i=M_i$, and $[A_1^T,\ldots,A_l^T][A_1^T,\ldots,A_l^T]^T=\sum_i M_i$. We also consider a decomposition $C=\sum_{k=1}^r \boldsymbol{c_k}\boldsymbol{c_k}^T$. For any factorization $M=A^TA$ of a positive semidefinite matrix M, it is known that $\operatorname{Im} M=\operatorname{Im} A$, and so the following equivalence relations hold:

$$\begin{aligned}\operatorname{Im} C\subset\operatorname{Im}\sum_i M_i &\iff \forall\, k\in[r],\ \boldsymbol{c_k}\in\operatorname{Im}\left(\sum_i M_i\right)=\operatorname{Im}\left(\left[A_1^T,\ldots,A_l^T\right]\right)\\ &\iff \forall\, k\in[r],\ \boldsymbol{c_k}\in\left(\bigcap_{i=1}^l \operatorname{Ker}(A_i)\right)^{\perp}.\end{aligned} \tag{13}$$

We first assume that the range inclusion condition does not hold. Relation (13) shows that

$$\exists k\in[r],\quad \exists\boldsymbol{h}\in\mathbb{R}^n:\forall\, i\in[l],\quad A_i\boldsymbol{h}=0,\quad \boldsymbol{c_k}^T\boldsymbol{h}\neq 0.$$

Now, notice that $X=\alpha\boldsymbol{h}\boldsymbol{h}^T$ is feasible for all $\alpha>0$, since $\alpha\langle A_i^TA_i,\boldsymbol{h}\boldsymbol{h}^T\rangle=0\leqslant b_i$. This contradicts the fact that Problem (P) is bounded, because $\langle C,X\rangle\geqslant\alpha(\boldsymbol{c_k}^T\boldsymbol{h})^2$, and α can be chosen arbitrarily large.

Conversely, if the range inclusion holds, we consider the Lagrangian dual (D) of Problem (P): the range inclusion condition indicates that this problem is feasible, because it implies the existence of a scalar $\lambda>0$ such that $\lambda\sum_i M_i\succeq C$ (we point out that a convenient value for λ is $\sum_{k=1}^r \boldsymbol{c_k}^T(\sum_i M_i)^\dagger\boldsymbol{c_k}$; this can be seen with the help of the Schur complement lemma). This means that Problem (D) has a finite optimal value $OPT\leqslant\lambda\sum_i b_i$, and by weak duality, Problem (P) is bounded (its optimal value cannot exceed OPT). $\square$

Before proving Theorem 2, we need to show that we can project Problem (P) on a subspace such that the projected problem (P′) and its Lagrangian dual are strictly feasible (Proposition 4).

Proof of Proposition 4. Let $\mathcal{I}_0$, $\mathcal{I}$, U and V be defined as in the paragraph preceding the statement of the proposition. Note that every matrix M_i can be decomposed as $M_i = U\tilde{M}_iU^T$ for a given matrix $\tilde{M}_i$, because its range is included in the range of $\sum_i M_i$ (we have $\tilde{M}_i = U^T M_i U$). The same observation holds for C, which can be decomposed as $C = U\tilde{C}U^T$ (we have assumed the range inclusion $\operatorname{Im} C \subset \operatorname{Im} \sum_i M_i$). Hence, Problem (P) is equivalent to:

$$\begin{aligned} \max_{X\succeq 0} \quad & \langle \tilde{C}, U^T X U\rangle, \\ \text{s.t.} \quad & \langle \tilde{M}_i, U^T X U\rangle \leqslant b_i, \quad i \in [l]. \end{aligned}$$

After the change of variable $Z_0 = U^T X U$ (Z_0 is a positive semidefinite matrix if X is), we obtain a reduced semidefinite packing problem

$$\begin{aligned} \max_{Z_0\succeq 0} \quad & \langle \tilde{C}, Z_0\rangle, \\ \text{s.t.} \quad & \langle \tilde{M}_i, Z_0\rangle \leqslant b_i, \quad i \in [l]. \end{aligned} \tag{14}$$

By construction, if Z_0 is a solution of (14), then $X := UZ_0U^T$ is a solution of (P). Note that the projected matrices in the constraints now satisfy $\sum_i \tilde{M}_i = U^T(\sum_i M_i)U \succ 0$.

We shall now consider a second projection, in order to get rid of the constraints in which $b_i = 0$. Note that each constraint indexed by $i \in \mathcal{I}_0$ is equivalent to imposing that Z_0 belongs to the nullspace of the matrix $\tilde{M}_i$. Since the columns of V form a basis of $\bigcap_{i\in\mathcal{I}_0} \operatorname{Ker} \tilde{M}_i$, any semidefinite matrix Z_0 which is feasible for Problem (14) must be of the form VZV^T for some positive semidefinite matrix Z. Hence, Problem (14) reduces to:

$$\begin{aligned} \max_{Z\succeq 0} \quad & \langle V^T\tilde{C}V, Z\rangle, \\ \text{s.t.} \quad & \langle V^T\tilde{M}_iV, Z\rangle \leqslant b_i, \quad i \in \mathcal{I}. \end{aligned} \tag{15}$$

which is nothing but Problem (P′), because $V^T\tilde{M}_iV = V^TU^TM_iUV = M_i'$ and $V^T\tilde{C}V = C'$. By construction, If Z is a solution of (15)≡(P′), then VZV^T is a solution of (14), and $(UV)Z(UV)^T$ is a solution of the original problem (P). This proves the point (*iii*) of the proposition.

We have pointed out above that $\sum_i \tilde{M}_i \succ 0$. Therefore, there exists a real $\lambda > 0$ such that $\lambda \sum_i \tilde{M}_i \succ \tilde{C}$, and $\lambda \sum_i M_i' = V^T(\lambda \sum_i \tilde{M}_i)V \succ V^T\tilde{C}V = C'$. This proves the strict dual feasibility of Problem (P′) (point (*ii*) of the proposition). Finally, since every b_i is positive for $i \in \mathcal{I}$, it is clear that the matrix $\bar{Z} = \varepsilon I \succ 0$ is strictly feasible for Problem (P′) as soon as $\varepsilon > 0$ is sufficiently small. This establishes the point (*i*), and the proposition is proved. □

We can now prove the main result of this article. We will first show that the result holds when every M_i is positive definite, thanks to the complementary slackness relation. Then, the general result is obtained by continuity. We point out at the end of this section the sketch of an alternative proof of Theorem 2 for the case in which $r = 1$, based on the bidual of Problem (P) and Schur complements, that shows directly that Problem (P) reduces to the SOCP (2).

Proof of Theorem 2. We will show that the result of the theorem holds for any semidefinite packing problem which is strictly feasible, and whose dual is strictly feasible. Then, by Proposition 4, we can say that Problem (P′) has a solution Z of rank at most $r' := \operatorname{rank} C'$, and $X := (UV)^T Z(UV)$ is a solution of the original problem which is of rank at most $r' \leqslant r$.

So let us assume without loss of generality that (P) and (D) are strictly feasible:

$$\forall\, i \in [l],\, b_i > 0 \quad \text{and} \quad \exists \lambda > 0 : \lambda \sum_i M_i \succ C.$$

The Slater condition is fulfilled for this pair of programs, and so strong duality holds (the optimal value of (P) equals the optimal value of (D), and the dual problem attains its optimum). In addition, the strict dual feasibility implies that (P) also attains its optimum. The pairs of primal and dual solutions

$(X^*, \boldsymbol{\mu}^*)$ are characterized by the Karush–Kuhn–Tucker (KKT) conditions:

$$\begin{aligned}
&\text{Primal feasibility:} && \forall\, i \in [l], \quad \langle M_i, X^* \rangle \leqslant b_i;\\
& && X^* \succeq 0;\\
&\text{Dual feasibility:} && \boldsymbol{\mu^*} \geqslant 0, \quad \sum_{i=1}^{l} \mu_i^* M_i \succeq C;\\
&\text{Complementary slackness:} && \left(\sum_{i=1}^{l} \mu_i^* M_i - C\right) X^* = 0,\\
& && \forall\, i \in [l],\ \mu_i^* \left(b_i - \langle M_i, X^* \rangle\right) = 0.
\end{aligned}$$

Now, we consider the case in which $M_i \succ 0$ for all i, and we choose an arbitrary pair of primal and dual optimal solutions $(X^*, \boldsymbol{\mu}^*)$. The dual feasibility relation implies $\boldsymbol{\mu}^* \neq \mathbf{0}$, and so $\sum_i \mu_i^* M_i$ is a positive definite matrix (we exclude the trivial case $C = 0$). Since C is of rank r, we deduce that

$$\operatorname{rank}\left(\sum_i \mu_i^* M_i - C\right) \geqslant n - r.$$

Finally, the complementary slackness relation indicates that the columns of X^* belong to the nullspace of $(\sum_i \mu_i^* M_i - C)$, which is a vector space of dimension at most $n - (n - r) = r$, and so we conclude that $\operatorname{rank} X^* \leqslant r$.

We now turn to the study of the general case in which $M_i \succeq 0$. To this end, we consider the perturbed problems

$$\begin{aligned}
\max \quad & \langle C, X \rangle,\\
\text{s.t.} \quad & \langle M_i + \varepsilon I, X \rangle \leqslant b_i,\\
& X \succeq 0
\end{aligned} \tag{P_ε}$$

and

$$\begin{aligned}
\min_{\boldsymbol{\mu} \geqslant 0} \quad & \sum_{i=1}^{l} \mu_i b_i,\\
\text{s.t.} \quad & \sum_{i=1}^{l} \mu_i (M_i + \varepsilon I) \succeq C,
\end{aligned} \tag{D_ε}$$

where $\varepsilon \geqslant 0$. Note that the strict feasibility of the unperturbed problems (P) and (D) implies that of (P_ε) and (D_ε) on a neighborhood $\varepsilon \in [0, \varepsilon_0]$, $\varepsilon_0 > 0$. We denote by $(X^\varepsilon, \boldsymbol{\mu^\varepsilon})$ a pair of primal and dual solutions of (P_ε)–(D_ε).

If $\varepsilon > 0$, $M_i + \varepsilon I \succ 0$ and it follows from the previous discussion that X^ε is of rank at most r. We show below that we can choose the optimal variables $(X^\varepsilon, \boldsymbol{\mu^\varepsilon})_{\varepsilon \in]0, \varepsilon_0]}$ within a bounded region, so that we can construct a converging subsequence $(X^{\varepsilon_k}, \boldsymbol{\mu^{\varepsilon_k}})_{k \in \mathbb{N}},\ \varepsilon_k \to 0$ from these variables. To conclude, we will see that the limit $(X^0, \boldsymbol{\mu^0})$ satisfies the KKT conditions for Problems (P)–(D), and that X^0 is of rank at most r.

Let us denote the optimal value of Problems (P_ε)–(D_ε) by $OPT(\varepsilon)$. Since the constraints of the primal problem becomes tighter when ε grows, it is clear that $OPT(\varepsilon)$ is nonincreasing with respect to ε, so that

$$\forall\, \varepsilon \in [0, \varepsilon_0], \quad OPT(\varepsilon_0) \leqslant OPT(\varepsilon) \leqslant OPT(0).$$

We have:

$$\lambda\left(\sum_i M_i + \varepsilon I\right) - C \succ \lambda\left(\sum_i M_i\right) - C,$$

and so we can write

$$\begin{aligned}\left\langle \lambda \sum_i M_i - C, X^{\varepsilon} \right\rangle &\leqslant \left\langle \lambda \sum_i (M_i + \varepsilon I) - C, X^{\varepsilon} \right\rangle \\ &= \lambda \left\langle \sum_i (M_i + \varepsilon I)\,, X^{\varepsilon} \right\rangle - OPT(\varepsilon) \\ &\leqslant \lambda \sum_i b_i - OPT(\varepsilon_0),\end{aligned}$$

where the equality comes from the expression of $OPT(\varepsilon)$ and the latter inequality follows from the constraints of the Problem (P_ε). The matrix $\lambda \sum_i M_i - C$ is positive definite by assumption and its smallest eigenvalue λ' is therefore positive. Hence,

$$\lambda' \text{ trace } X^{\varepsilon} \leqslant \left\langle \lambda \sum_i M_i - C, X^{\varepsilon} \right\rangle \leqslant \overline{\boldsymbol{\mu}}^T \boldsymbol{b} - OPT(\varepsilon) \leqslant \lambda \sum_i b_i - OPT(\varepsilon_0).$$

This shows that the positive semidefinite matrix X^{ε} has its trace bounded, and therefore all its entries are bounded.

It remains to show that the dual optimal variable $\boldsymbol{\mu}^{\boldsymbol{\varepsilon}} \geqslant \mathbf{0}$ is bounded. This is simply done by writing:

$$\forall\, i \in [l], \quad b_i \mu_i^{\varepsilon} \leqslant \boldsymbol{b}^T \boldsymbol{\mu}^{\boldsymbol{\varepsilon}} = OPT(\varepsilon) \leqslant OPT(0).$$

By assumption, $b_i > 0$, and the entries of the vector $\boldsymbol{\mu}^{\boldsymbol{\varepsilon}} \geqslant \mathbf{0}$ are bounded.

We can therefore construct a sequence of pairs of primal and dual optimal solutions $(X^{\varepsilon}, \boldsymbol{\mu}^{\boldsymbol{\varepsilon_k}})_{k \in \mathbb{N}}$ that converges, with $\varepsilon_k \underset{k\to\infty}{\longrightarrow} 0$, $\varepsilon_k > 0$. The limit X^0 of this sequence is of rank at most r, because the rank is a lower semicontinuous function and rank $X^{\varepsilon_k} \leqslant r$ for all $k \in \mathbb{N}$. It remains to show that X^0 is a solution of Problem (P). The ε-perturbed KKT conditions must hold for all $k \in \mathbb{N}$, and so they hold for the pair $(X_0, \boldsymbol{\mu}^{\mathbf{0}})$ by taking the limit (the limit of any sequence of positive semidefinite matrices is a positive semidefinite matrix because $\mathbb{S}_n^+$ is closed). This concludes the proof. □

4.1. *Sketch of an alternative proof of Theorem 2 when $r = 1$*

By Proposition 4, we only need to show that the result holds for the reduced problem (P′), and so we assume without loss of generality that strong duality holds for all the optimization problems considered below.

When $r = 1$, there is a vector $\boldsymbol{c}$ such that $C = \boldsymbol{c}\boldsymbol{c}^{\boldsymbol{T}}$ and the dual problem of (P) takes the form:

$$\begin{aligned} \min_{\boldsymbol{\mu} \geqslant 0} \quad & \boldsymbol{\mu}^{\boldsymbol{T}} \boldsymbol{b}, \\ \text{s.t.} \quad & \boldsymbol{c}\boldsymbol{c}^{\boldsymbol{T}} \preceq \sum_i \mu_i M_i. \end{aligned} \tag{16}$$

Now, setting $t = \boldsymbol{\mu}^{\boldsymbol{T}} \boldsymbol{b}$, and $\boldsymbol{w} = \frac{\boldsymbol{\mu}}{t}$, so that the new variable $\boldsymbol{w}$ satisfies $\boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{b} = 1$, the constraint of the previous problem becomes $\frac{\boldsymbol{c}\boldsymbol{c}^{\boldsymbol{t}}}{t} \preceq \sum_i w_i M_i$. This matrix inequality, together with the fact that the optimal t is positive, can be reformulated thanks to the Schur complement lemma, and (16) is equivalent to:

$$\begin{aligned} \min_{t \in \mathbb{R}, \boldsymbol{w} \geqslant \mathbf{0}} \quad & t, \\ \text{s.t.} \quad & \left(\begin{array}{c|c} \sum_i w_i M_i & \boldsymbol{c} \\ \hline \boldsymbol{c}^{\boldsymbol{T}} & t \end{array} \right) \succeq 0, \\ & \boldsymbol{w}^{\boldsymbol{T}} \boldsymbol{b} = 1. \end{aligned} \tag{17}$$

We dualize this SDP once again to obtain the bidual of Program (P) (strong duality holds):

$$\max_{\beta\in\mathbb{R}, Z\in\mathbb{S}_{n+1}^+} \quad -\beta - 2\boldsymbol{v}^{\boldsymbol{T}}\boldsymbol{c}, \tag{18}$$

$$\text{s.t.} \quad \langle W, M_i\rangle \leqslant \beta b_i, \quad i \in [l]$$

$$Z = \left(\begin{array}{c|c} W & \boldsymbol{v} \\ \hline \boldsymbol{v}^{\boldsymbol{T}} & 1 \end{array}\right) \succeq 0.$$

We notice that the last matrix inequality is equivalent to $W \succeq \boldsymbol{v}\boldsymbol{v}^{\boldsymbol{T}}$, using a Schur complement. Since $M_i \succeq 0$, we can assume that $W = \boldsymbol{v}\boldsymbol{v}^{\boldsymbol{T}}$ without loss of generality, and (18) becomes:

$$\max_{\beta\in\mathbb{R}, \boldsymbol{v}\in\mathbb{R}^n} \quad -\beta - 2\boldsymbol{v}^{\boldsymbol{T}}\boldsymbol{c}, \tag{19}$$

$$\text{s.t.} \quad \|A_i\boldsymbol{v}\|^2 \leqslant \beta b_i, \quad i = 1 \in [l],$$

where A_i is a matrix such that $A_i^T A_i = M_i$.

We now define the new variables $\alpha = \sqrt{\beta}$ and $\boldsymbol{x} = \frac{\boldsymbol{v}}{\alpha}$, so that (19) becomes:

$$\max_{\boldsymbol{x}\in\mathbb{R}^n} \quad \left(\max_{\alpha} -\alpha^2 - 2\alpha\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{c}\right), \tag{20}$$

$$\text{s.t.} \quad \|A_i\boldsymbol{x}\| \leqslant \sqrt{b_i}, \quad i = 1 \in [l].$$

The reader can finally verify that the value of the max within parenthesis is $(\boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{x})^2$, and we have proved that the SDP (P) reduces to the SOCP (2). By the way, this guarantees that the SDP (P) has a rank-one solution. $\square$

Acknowledgments

The author thanks Stéphane Gaubert for his useful comments and enlightening discussions, as well as for his warm support. He also expresses his gratitude to two anonymous referees. In a previous version, the main result was restricted to the case in which $r = 1$. One referee suggested an alternative proof with an elegant complementary slackness argument, which led to the more general statement of Theorem 2. The author also thanks a second referee for his useful remarks, which helped to clarify the consequences and the presentation of these results.

A. Proof of Theorem 6

Before we give the proof of Theorem 6, we need one additional technical lemma, which shows that one can assume without loss of generality that the primal problem is strictly feasible, and that the vector space spanned by the vectors $\boldsymbol{h_0}, \boldsymbol{h_1}, \ldots, \boldsymbol{h_l}$ coincides with the cone generated by the same vectors. One can consider this lemma as the analog of Proposition 4 for combined problems.

Lemma 8. *We assume that the conditions of Theorem 6 are fulfilled. Then, there exists a subset $\mathcal{I} \subset [l]$, as well as matrices $C' \succeq 0$ and $M_i' \succeq 0$ ($i \in \mathcal{I}$), so that the reduced "combined" semidefinite packing problem*

$$\max_{Z\succeq 0,\ Y\succeq 0,\ \boldsymbol{\lambda}} \langle C', Z\rangle + \langle R_0, Y\rangle + \boldsymbol{h_0}^T\boldsymbol{\lambda} \quad \text{s.t.} \quad \forall\, i \in \mathcal{I},\ \langle M_i', Z\rangle \leqslant b_i + \langle R_i, Y\rangle + \boldsymbol{h_i}^T\boldsymbol{\lambda}$$

has the same optimal value as (P_{CMB}) and satisfies the following properties:

(*i*) $\exists (Z' \succ 0, Y' \succ 0, \boldsymbol{\lambda'}) : \forall\, i \in \mathcal{I},\ \langle M_i, Z'\rangle < b_i + \langle R_i, Y'\rangle + \boldsymbol{h_i}^T\boldsymbol{\lambda'}$.

(*ii*) *The cone K generated by the vectors $(\boldsymbol{h_i})_{i\in\{0\}\cup\mathcal{I}}$ is a vector space.*

(*iii*) $\operatorname{rank} C' \leqslant \operatorname{rank} C$.

(*iv*) *There is a matrix U with orthonormal columns such that if $(Z, Y, \boldsymbol{\lambda})$ is a solution of the reduced problem, then $(X := UZU^T, Y, \boldsymbol{\lambda})$ is a solution of Problem (P_{CMB}) (which of course satisfies* $\operatorname{rank} X \leqslant \operatorname{rank} Z$*).*

Proof. In this lemma, (*i*) and (*ii*) are the properties that we will need to prove Theorem 6. Properties (*iii*) and (*iv*) ensure that if the theorem holds for the reduced problem, then the result also holds for the initial problem (P_{CMB}). We handle separately the cases in which the initial problem does not satisfy the property (*i*) or (*ii*). If both cases arise simultaneously, we obtain the result of this lemma by applying successively the following two reductions.

Let $(X^*, Y^*, \boldsymbol{\lambda}^*)$ be an optimal solution of Problem (P_{CMB}); the existence of a solution is guaranteed by the (refined) Slater condition satisfied by the dual problem indeed (see e.g. [17,3]). We denote by $\mathcal{I}_0 \subset [l]$ the subset of indices for which $b_i + \langle R_i, Y^* \rangle + \boldsymbol{h_i}^T \boldsymbol{\lambda^*} = 0$ (note that we have $b_i + \langle R_i, Y^* \rangle + \boldsymbol{h_i}^T \boldsymbol{\lambda^*} \geqslant 0$ for all i because $M_i \succeq 0$ implies $\langle M_i, X^* \rangle \geqslant 0$). We define $\mathcal{I} := [l] \backslash \mathcal{I}_0$. In Problem ($P_{\text{CMB}}$), we can replace the constraint $\langle M_i, X \rangle \leqslant b_i + \langle R_i, Y \rangle + \boldsymbol{h_i}^T \boldsymbol{\lambda}$ by $\langle M_i, X \rangle = 0$ for all $i \in \mathcal{I}_0$, since $(X^*, Y^*, \boldsymbol{\lambda^*})$ satisfies this stronger set of constraints. For a feasible positive semidefinite matrix X, this implies $\langle \sum_{i \in \mathcal{I}_0} M_i, X \rangle = 0$, and even $\sum_{i \in \mathcal{I}_0} M_i X = 0$. Therefore, X is of the form UZU^T for some positive semidefinite matrix Z, where the columns of U form an orthonormal basis of the nullspace of $M_0 := \sum_{i \in \mathcal{I}_0} M_i$ (U is obtained by taking the eigenvectors corresponding to the vanishing eigenvalues of M_0). Hence, Problem (P_{CMB}) is equivalent to:

$$\begin{aligned} \max \quad & \langle U^T C U, Z \rangle + \langle R_0, Y \rangle + \boldsymbol{h_0}^T \boldsymbol{\lambda}, \\ \text{s.t.} \quad & \langle U^T M_i U, Z \rangle \leqslant b_i + \langle R_i, Y \rangle + \boldsymbol{h_i}^T \boldsymbol{\lambda}, \quad i \in \mathcal{I}, \\ & Z \succeq 0, \; Y \succeq 0. \end{aligned} \tag{A.1}$$

We have thus reduced the problem to one for which $b_i + \langle R_i, Y^* \rangle + \boldsymbol{h_i}^T \boldsymbol{\lambda^*} > 0$ for all i, and strict feasibility follows (i.e. property (*i*) holds, consider $\boldsymbol{\lambda'} = \boldsymbol{\lambda^*}$, $Y' = Y^* + \eta_1 I$, and $Z' = \eta_2 I$ for sufficiently small reals $\eta_1 > 0$ and $\eta_2 > 0$). Moreover, the projected matrix $C' := U^T C U$ in the objective function has a smaller rank than C (i.e. (*iii*) holds). Finally, (*iv*) holds for the reduced problem by construction: if $(Z, Y, \boldsymbol{\lambda})$ is a solution of Problem (A.1), then $(X := UZU^T, Y, \boldsymbol{\lambda})$ is a solution of Problem (P_{CMB}), both problems have the same optimal value, and of course $\operatorname{rank} X \leqslant \operatorname{rank} Z$.

We now handle the second case, in which Property (*ii*) does not hold for Problem (P_{CMB}). The set $K = \{ [\boldsymbol{h_0}, H]\boldsymbol{v}, \boldsymbol{v} \in \mathbb{R}^{l+1}, \boldsymbol{v} \geqslant \boldsymbol{0}\}$ is a closed convex cone. Hence, it is known that it can be decomposed as $K = L + Q$, where L is a vector space and $Q \subset L^{\perp}$ is a closed convex pointed cone ($L = K \cap (-K)$ is the *lineality space* of K). The interior of the dual cone Q^* is therefore nonempty, i.e. $\exists \boldsymbol{\lambda} : \forall \boldsymbol{q} \in Q \backslash \{\boldsymbol{0}\}, \boldsymbol{\lambda}^T \boldsymbol{q} > 0$. Let $\boldsymbol{\lambda_0}$ be the orthogonal projection of $\boldsymbol{\lambda}$ on $L^{\perp}$, so that $\boldsymbol{\lambda_0}^T \boldsymbol{q} = \boldsymbol{\lambda}^T \boldsymbol{q} > 0$ for all $\boldsymbol{q} \in Q \backslash \{\boldsymbol{0}\}$, and $\boldsymbol{\lambda_0}^T \boldsymbol{x} = 0$ for all $\boldsymbol{x} \in L$. Now, we define the set of indices $\mathcal{I} = \{i \in [l] : \boldsymbol{h_i} \in L\}$, and its complement $\mathcal{I}_0 = [l] \backslash \mathcal{I}$. For all $i \in \mathcal{I}_0$, $\boldsymbol{h_i} = \boldsymbol{x_i} + \boldsymbol{q_i}$ for a vector $\boldsymbol{x_i} \in L$ and a vector $\boldsymbol{q_i} \in Q \backslash \{\boldsymbol{0}\}$, so that $\boldsymbol{\lambda_0}^T \boldsymbol{h_i} = \boldsymbol{\lambda_0}^T \boldsymbol{x_i} + \boldsymbol{\lambda_0}^T \boldsymbol{q_i} = \boldsymbol{\lambda_0}^T \boldsymbol{q_i} > 0$. For the indices $i \in \mathcal{I}$, it is clear that $\boldsymbol{\lambda_0}^T \boldsymbol{h_i} = 0$. Finally, since $\boldsymbol{h_0} + H\overline{\boldsymbol{\mu}} = 0$, we have $-\boldsymbol{h_0} \in K$, so that $\boldsymbol{h_0} \in L$ and $\boldsymbol{h_0}^T \boldsymbol{\lambda} = 0$. To sum up, we have proved the existence of a vector $\boldsymbol{\lambda_0}$ for which

$$\forall\, i \in \{0\} \cup \mathcal{I}, \quad \boldsymbol{\lambda_0}^T \boldsymbol{h_i} = 0 \quad \text{and} \quad \forall\, i \in \mathcal{I}_0, \boldsymbol{\lambda_0}^T \boldsymbol{h_i} > 0.$$

Let $(X^*, Y^*, \boldsymbol{\lambda}^*)$ be an optimal solution of Problem (P_{CMB}). For all positive real t, $(X^*, Y^*, \boldsymbol{\lambda}^* + t\boldsymbol{\lambda_0})$ is also a solution, because it is feasible and has the same objective value. Letting $t \to \infty$, we see that the constraints of the problem that are indexed by $i \in \mathcal{I}_0$ may be removed without changing the optimum. We have thus reduced the problem to one for which (*ii*) holds. $\square$

We can now prove Theorem 6. The proof mimics that of Theorem 2, i.e. we first show that the result holds when each M_i is positive definite, and the general result is obtained by continuity. The only difference is how we show that we can choose optimal variables $(X^{\varepsilon}, Y^{\varepsilon}, \boldsymbol{\lambda^{\varepsilon}}, \boldsymbol{\mu^{\varepsilon}})_{\varepsilon \in]0, \varepsilon_0]}$ for a perturbed problem within a bounded region.

Proof of Theorem 6. By Lemma 8, we may assume without loss of generality that $K = \operatorname{cone}\{\boldsymbol{h_0}, \ldots, \boldsymbol{h_l}\} \supset -K$ and that the primal problem is strictly feasible. The strict feasibility of the primal problem ensures that strong duality holds, i.e. the optimal value of (P_{CMB}) equals the optimal value of (D_{CMB}), and the optimum is attained in the dual problem. Moreover, the (refined) Slater constraints qualification for the dual problem guarantees the existence of primal optimal variables as well (see e.g. Theorem 28.2 in [17]). The pairs of primal and dual solutions

$\big((X^*, Y^*, \boldsymbol{\lambda^*}), \boldsymbol{\mu}^*\big)$ are characterized by the Karush–Kuhn–Tucker (KKT) conditions:

$$\text{Primal feasibility:}\quad \forall\, i \in [l],\quad \langle M_i, X^*\rangle \leqslant b_i + \langle R_i, Y^*\rangle + \boldsymbol{h_i}^T\boldsymbol{\lambda^*},$$

$$X^* \succeq 0,\ Y^* \succeq 0;$$

$$\text{Dual feasibility:}\quad \boldsymbol{\mu^*} \geqslant 0,\quad \sum_{i=1}^{l} \mu_i^* M_i \succeq C,\quad R_0 + \sum_{i=1}^{l} \mu_i^* R_i \preceq 0,\quad \boldsymbol{h_0} + H\boldsymbol{\mu^*} = 0;$$

$$\text{Complementary slackness:}\quad \left(\sum_{i=1}^{l} \mu_i^* M_i - C\right) X^* = 0,\quad \left(R_0 + \sum_{i=1}^{l} \mu_i^* R_i\right) Y^* = 0,$$

$$\forall\, i \in [l],\ \mu_i^*(b_i + \langle R_i, Y^*\rangle + \boldsymbol{h_i}^T\boldsymbol{\lambda^*} - \langle M_i, X^*\rangle) = 0.$$

Now, we consider the case in which $M_i \succ 0$ for all i, and we choose an arbitrary pair of primal and dual optimal solutions $\big((X^*, Y^*, \boldsymbol{\lambda^*}), \boldsymbol{\mu}^*\big)$. The dual feasibility relation implies $\boldsymbol{\mu}^* \neq \mathbf{0}$, and so $\sum_i \mu_i^* M_i$ is a positive definite matrix (we exclude the trivial case $C = 0$). Since C is of rank r, we deduce that

$$\operatorname{rank}\left(\sum_i \mu_i^* M_i - C\right) \geqslant n - r.$$

Finally, the complementary slackness relation indicates that the columns of X^* belong to the nullspace of $(\sum_i \mu_i^* M_i - C)$, which is a vector space of dimension at most $n - (n - r) = r$, and so we conclude that $\operatorname{rank} X^* \leqslant r$.

We now turn to the study of the general case in which $M_i \succeq 0$. To this end, we consider the perturbed problems

$$\begin{aligned} \max\quad & \langle C, X\rangle + \langle R_0, Y\rangle + \boldsymbol{h_0}^T\boldsymbol{\lambda}, \\ \text{s.t.}\quad & \langle M_i + \varepsilon I, X\rangle \leqslant b_i + \langle R_i, Y\rangle + \boldsymbol{h_i}^T\boldsymbol{\lambda} \quad i \in [l], \\ & X \succeq 0,\ Y \succeq 0 \end{aligned} \tag{$P_{\text{CMB}}^{\varepsilon}$}$$

and

$$\begin{aligned} \min_{\boldsymbol{\mu} \geqslant 0}\quad & \sum_{i=1}^{l} \mu_i b_i, \\ \text{s.t.}\quad & \sum_{i=1}^{l} \mu_i (M_i + \varepsilon I) \succeq C, \\ & R_0 + \sum_{i=1}^{l} \mu_i R_i \preceq 0, \\ & \boldsymbol{h_0} + H\boldsymbol{\mu} = \mathbf{0}, \end{aligned} \tag{$D_{\text{CMB}}^{\varepsilon}$}$$

where $\varepsilon \geqslant 0$. Note that the refined Slater constraints qualification for the unperturbed problems (P_{CMB}) and (D_{CMB}) (i.e. simultaneous feasibility (resp. strict feasibility) of all the affine constraints (resp. non-affine constraints)) implies the qualification of the constraints for ($P_{\text{CMB}}^{\varepsilon}$) and ($D_{\text{CMB}}^{\varepsilon}$) on a neighborhood $\varepsilon \in [0, \varepsilon_0]$, $\varepsilon_0 > 0$. We denote by $\big((X^{\varepsilon}, Y^{\varepsilon}, \boldsymbol{\lambda^{\varepsilon}}), \boldsymbol{\mu^{\varepsilon}}\big)$ a pair of primal and dual solutions of ($P_{\text{CMB}}^{\varepsilon}$)–($D_{\text{CMB}}^{\varepsilon}$). If $\varepsilon > 0$, $M_i + \varepsilon I \succ 0$ and it follows from the previous discussion that X^{ε} is of rank at most r. We show below that we can choose the optimal variables $(X^{\varepsilon}, Y^{\varepsilon}, \boldsymbol{\lambda^{\varepsilon}}, \boldsymbol{\mu^{\varepsilon}})_{\varepsilon \in]0,\varepsilon_0]}$ within a bounded region, so that we can construct a converging subsequence $(X^{\varepsilon_k}, Y^{\varepsilon_k}, \boldsymbol{\lambda^{\varepsilon_k}}, \boldsymbol{\mu^{\varepsilon_k}})_{k \in \mathbb{N}},\ \varepsilon_k \to 0$ from these variables. To conclude, we will see that the limit $(X^0, Y^0, \boldsymbol{\lambda^0}, \boldsymbol{\mu^0})$ satisfies the KKT conditions for Problems (P_{CMB})–(D_{CMB}), and that X^0 is of rank at most r.

Let us denote the optimal value of Problems ($P_{\text{CMB}}^{\varepsilon}$)–($D_{\text{CMB}}^{\varepsilon}$) by $OPT(\varepsilon)$. Since the constraints of the primal problem becomes tighter when ε grows, it is clear that $OPT(\varepsilon)$ is nonincreasing with respect to ε, so that

$$\forall\, \varepsilon \in [0, \varepsilon_0],\quad OPT(\varepsilon_0) \leqslant OPT(\varepsilon) \leqslant OPT(0).$$

Now let $\varepsilon \in]0, \varepsilon_0]$. By assumption, there exists a vector $\overline{\boldsymbol{\mu}} \geqslant \mathbf{0}$ such that

$$\sum_i \overline{\mu}_i (M_i + \varepsilon I) \succeq \sum_i \overline{\mu}_i M_i \succ C, \quad \text{and} \quad R_0 + \sum_i \overline{\mu}_i R_0 \prec 0. \tag{A.2}$$

Therefore, we have

$$\begin{aligned} OPT(\varepsilon) = \langle C, X^\varepsilon \rangle + \langle R_0, Y^\varepsilon \rangle + \boldsymbol{h_0}^T \boldsymbol{\lambda^\varepsilon} &\leqslant \langle \sum_i \overline{\mu}_i \left(M_i + \varepsilon I\right), X^\varepsilon \rangle + \langle R_0, Y^\varepsilon \rangle + \boldsymbol{h_0}^T \boldsymbol{\lambda^\varepsilon} \\ &\leqslant \sum_i \overline{\mu}_i \left(b_i + \langle R_i, Y^\varepsilon \rangle + \boldsymbol{h_i}^T \boldsymbol{\lambda^\varepsilon} \right) + \langle R_0, Y^\varepsilon \rangle + \boldsymbol{h_0}^T \boldsymbol{\lambda^\varepsilon} \\ &= \overline{\boldsymbol{\mu}}^T \boldsymbol{b} + \langle \sum_i \overline{\mu}_i R_i + R_0, Y^\varepsilon \rangle + \underbrace{(\boldsymbol{h_0} + H\overline{\boldsymbol{\mu}})^T}_{=\mathbf{0}} \boldsymbol{\lambda^\varepsilon}, \end{aligned}$$

where the first inequality follows from (A.2), and the second one from the feasibility condition $\langle M_i + \varepsilon I, X^\varepsilon \rangle \leqslant b_i + \langle R_i, Y^\varepsilon \rangle + \boldsymbol{h_i}^T \boldsymbol{\lambda^\varepsilon}$. The assumption (A.2) moreover implies that $-(\sum_i \overline{\mu}_i R_i + R_0)$ is positive definite, so that its smallest eigenvalue λ' is positive, and

$$\lambda' \text{ trace } Y^\varepsilon \leqslant \left\langle - \left(\sum_i \overline{\mu}_i R_i + R_0 \right), Y^\varepsilon \right\rangle \leqslant \overline{\boldsymbol{\mu}}^T \boldsymbol{b} - OPT(\varepsilon) \leqslant \overline{\boldsymbol{\mu}}^T \boldsymbol{b} - OPT(\varepsilon_0).$$

This shows that the trace of Y^ε is bounded, and so $Y^\varepsilon \succeq 0$ is bounded.

Similarly, to bound X^ε, we write:

$$\begin{aligned} \left\langle \sum_i \overline{\mu}_i M_i - C, X^\varepsilon \right\rangle &\leqslant \left\langle \sum_i \overline{\mu}_i \left(M_i + \varepsilon I\right) - C, X^\varepsilon \right\rangle \\ &= \left\langle \sum_i \overline{\mu}_i \left(M_i + \varepsilon I\right), X^\varepsilon \right\rangle - OPT(\varepsilon) + \langle R_0, Y^\varepsilon \rangle + \boldsymbol{h_0}^T \boldsymbol{\lambda^\varepsilon} \\ &\leqslant \sum_i \overline{\mu}_i (b_i + \langle R_i, Y^\varepsilon \rangle + \boldsymbol{h_i}^T \boldsymbol{\lambda^\varepsilon}) - OPT(\varepsilon) + \langle R_0, Y^\varepsilon \rangle + \boldsymbol{h_0}^T \boldsymbol{\lambda^\varepsilon} \\ &= \overline{\boldsymbol{\mu}}^T \boldsymbol{b} - OPT(\varepsilon) + \underbrace{\left\langle \sum_i \overline{\mu}_i R_i + R_0, Y^\varepsilon \right\rangle}_{\leqslant 0} + \underbrace{(\boldsymbol{h_0} + H\overline{\boldsymbol{\mu}})^T}_{=\mathbf{0}} \boldsymbol{\lambda^\varepsilon}, \end{aligned}$$

where the first equality comes from the expression of $OPT(\varepsilon)$. The matrix $\sum_i \overline{\mu}_i M_i - C$ is positive definite and its smallest eigenvalue λ'' is therefore positive. Hence,

$$\lambda'' \text{ trace } X^\varepsilon \leqslant \overline{\boldsymbol{\mu}}^T \boldsymbol{b} - OPT(\varepsilon) \leqslant \overline{\boldsymbol{\mu}}^T \boldsymbol{b} - OPT(\varepsilon_0),$$

and this shows that the matrix $X^\varepsilon \succeq 0$ is bounded.

Now, note that the feasibility of $\boldsymbol{\lambda^\varepsilon}$ implies that the quantity $b_i + \langle R_i, Y^\varepsilon \rangle + \boldsymbol{h_i}^T \boldsymbol{\lambda^\varepsilon}$ is nonnegative for all $i \in [l]$. Since Y^ε is bounded, we deduce the existence of a lower bound $m_i \in \mathbb{R}$ such that $\boldsymbol{h_i}^T \boldsymbol{\lambda^\varepsilon} \geqslant m_i$ $(\forall\, i \in [l])$. Similarly, since $\boldsymbol{h_0}^T \boldsymbol{\lambda^\varepsilon} \geqslant OPT(\varepsilon_0) - \langle C, X^\varepsilon \rangle - \langle R_0, Y^\varepsilon \rangle$, there is a scalar m_0 such that $\boldsymbol{h_0}^T \boldsymbol{\lambda^\varepsilon} \geqslant m_0$. We now use the fact that every vector $(-\boldsymbol{h_i})$ may be written as a positive combination of the $\boldsymbol{h_k}$, $(k \in \{0\} \cup [l])$, and we obtain that the quantities $\boldsymbol{h_i}^T \boldsymbol{\lambda^\varepsilon}$ are also bounded from above. Let us denote by H_0 the matrix $[\boldsymbol{h_0}, H]$; we have just proved that the vector $H_0^T \boldsymbol{\lambda^\varepsilon}$ is bounded:

$$\exists \overline{m} \in \mathbb{R} : \quad \| H_0^T \boldsymbol{\lambda^\varepsilon} \|_2 \leqslant \overline{m}$$

(the latter bound does not depend on ε). Note that one may assume without loss of generality that $\boldsymbol{\lambda^\varepsilon} \in \operatorname{Im} H_0$ (otherwise we consider the projection $\boldsymbol{\lambda_P^\varepsilon}$ of $\boldsymbol{\lambda^\varepsilon}$ on $\operatorname{Im} H_0$ which is also a solution since $H_0^T \boldsymbol{\lambda^\varepsilon} = H_0^T \boldsymbol{\lambda_P^\varepsilon}$). We know from the Courant-Fisher theorem that the smallest positive eigenvalue of $H_0 H_0^T$ satisfies:

$$\lambda_{\min}^{>} \left(H_0 H_0^T \right) = \min_{\boldsymbol{v} \in \operatorname{Im} H_0 \setminus \{\mathbf{0}\}} \frac{\boldsymbol{v}^T H_0 H_0^T \boldsymbol{v}}{\boldsymbol{v}^T \boldsymbol{v}}.$$

Therefore, since we have assumed $\boldsymbol{\lambda}^{\boldsymbol{\varepsilon}} \in \operatorname{Im} H_0$:

$$\|\boldsymbol{\lambda}^{\boldsymbol{\varepsilon}}\|^2 \leqslant \frac{\|H_0^T \boldsymbol{\lambda}^{\boldsymbol{\varepsilon}}\|^2}{\lambda_{\min}^{>}\left(H_0 H_0^T\right)} \leqslant \frac{\overline{m}^2}{\lambda_{\min}^{>}\left(H_0 H_0^T\right)}.$$

It remains to show that the dual optimal variable $\boldsymbol{\mu}^{\boldsymbol{\varepsilon}}$ is bounded. Our strict primal feasibility assumption (which does not entail generality thanks to Lemma 8) ensures the existence of a matrix $\overline{Y} \succ 0$ and a vector $\overline{\boldsymbol{\lambda}}$ such that

$$\forall\, i \in [l], \quad \langle R_i, \overline{Y} \rangle + b_i + \boldsymbol{h_i}^T \overline{\boldsymbol{\lambda}} = \eta_i > 0.$$

By dual feasibility, $R_0 + \sum_i \mu_i^{\varepsilon} R_i$ is a negative semidefinite matrix, and we have:

$$0 \geqslant \langle R_0, \overline{Y} \rangle + \sum_{i=1}^{l} \mu_i^{\varepsilon} \langle R_i, \overline{Y} \rangle = \langle R_0, \overline{Y} \rangle + \sum_{i=1}^{l} \mu_i^{\varepsilon} (\eta_i - b_i - \boldsymbol{h_i}^T \overline{\boldsymbol{\lambda}}).$$

Hence, we have the following inequalities:

$$\begin{aligned} \forall\, k \in [l],\ \eta_k \mu_k^{\varepsilon} \leqslant \sum_{i=1}^{l} \eta_i \mu_i^{\varepsilon} &\leqslant \boldsymbol{b}^T \boldsymbol{\mu}^{\boldsymbol{\varepsilon}} + \overline{\boldsymbol{\lambda}}^T H \boldsymbol{\mu}^{\boldsymbol{\varepsilon}} - \langle R_0, \overline{Y} \rangle \\ &= OPT(\varepsilon) - \overline{\boldsymbol{\lambda}}^T \boldsymbol{h_0} - \langle R_0, \overline{Y} \rangle \\ &\leqslant OPT(0) - \overline{\boldsymbol{\lambda}}^T \boldsymbol{h_0} - \langle R_0, \overline{Y} \rangle, \end{aligned}$$

and we have shown that $\boldsymbol{\mu}^{\boldsymbol{\varepsilon}} \geqslant \boldsymbol{0}$ is bounded.

We can therefore construct a sequence of pairs of primal and dual optimal solutions $(X^{\varepsilon_k}, Y^{\varepsilon_k}, \boldsymbol{\lambda}^{\boldsymbol{\varepsilon_k}}, \boldsymbol{\mu}^{\boldsymbol{\varepsilon_k}})_{k \in \mathbb{N}}$ that converges, with $\varepsilon_k \underset{k\to\infty}{\longrightarrow} 0$, $\varepsilon_k > 0$. In this sequence, the limit X^0 of X^{ε_k} is of rank at most r, because the rank is a lower semicontinuous function and rank $X^{\varepsilon_k} \leqslant r$ for all $k \in \mathbb{N}$. It remains to show that $(X^0, Y^0, \boldsymbol{\lambda}^{\boldsymbol{0}})$ is a solution of Problem (P_{CMB}). The ε-perturbed KKT conditions must hold for all $k \in \mathbb{N}$, and so they hold for the pair $((X^0, Y^0, \boldsymbol{\lambda}^{\boldsymbol{0}}), \boldsymbol{\mu}^{\boldsymbol{0}})$ by taking the limit (this works because $\mathbb{S}_n^+$ is closed). This concludes the proof of the existence of a solution in which rank $X \leqslant r$.

It remains to show the second statement of this theorem, namely that if $C \neq 0$ and $\bar{r} := \min_{i \in [l]} \operatorname{rank} M_i$, then the rank of X is bounded by $n - \bar{r} + r$ for any solution $(X, Y, \boldsymbol{\lambda})$ of (P_{CMB}).

Let $(X^*, Y^*, \boldsymbol{\lambda}^{\boldsymbol{*}})$ be a solution of Problem (P_{CMB}). If the primal problem is strictly feasible, then there exists a Lagrange multiplier $\boldsymbol{\mu}^{\boldsymbol{*}} \geqslant \boldsymbol{0}$ such that the KKT conditions described at the beginning of this proof are satisfied. Since $C \neq 0$, we have $\boldsymbol{\mu}^{\boldsymbol{*}} \neq \boldsymbol{0}$, and we can write:

$$\operatorname{rank} \left(\sum_{i \in [l]} \mu_i^* M_i - C \right) \geqslant \bar{r} - r.$$

Hence, since by complementary slackness, X^* belongs to the nullspace of $(\sum_{i \in [l]} \mu_i^* M_i - C)$, we find rank $X^* \leqslant n - \bar{r} + r$.

If the primal problem is not strictly feasible, there must be an index $i \in [l]$ such that $\langle M_i, X^* \rangle = 0$ (otherwise, $(\eta_1 I, Y^* + \eta_2 I, \boldsymbol{\lambda}^{\boldsymbol{*}})$ would be strictly feasible for sufficiently small positive reals η_1 and η_2). Therefore, X^* is in the nullspace of a matrix of rank larger than $\bar{r}$, and rank $X^* \leqslant n - \bar{r} \leqslant n - \bar{r} + r$. $\square$

B. Proof of Theorem 5

We assume that Problems (P_{CMB}) and (D_{CMB}) are feasible, and for $\eta \geqslant 0$ we consider the following pair of primal and dual perturbed problems.

$$\begin{aligned} \sup\ \ & \langle C, X \rangle + \langle R_0, Y \rangle + \boldsymbol{h_0}^T \boldsymbol{\lambda}, \\ \text{s.t.}\ \ & \langle M_i, X \rangle \leqslant b_i + \langle R_i, Y \rangle + \boldsymbol{h_i}^T \boldsymbol{\lambda} \quad i \in [l], \\ & \eta\ (\operatorname{trace} X + \operatorname{trace} Y) \leqslant 1, \\ & X \succeq 0,\ Y \succeq 0 \end{aligned} \tag{P_η}$$

and

$$\begin{aligned} \inf_{\boldsymbol{\mu}\geqslant\mathbf{0},\ \sigma\geqslant 0} \quad & \sum_{i=1}^{l}\mu_i b_i+\sigma, \\ \text{s.t.} \quad & \sum_{i=1}^{l}\mu_i M_i+\sigma\eta I\succeq C, \\ & R_0+\sum_{i=1}^{l}\mu_i R_i-\sigma\eta I\preceq 0, \\ & \boldsymbol{h_0}+H\boldsymbol{\mu}=\mathbf{0}. \end{aligned} \tag{D_η}$$

It is clear that the feasibility of Problem (P_{CMB}) implies that of (P_η) if $\eta>0$ is sufficiently small. Let $\overline{\boldsymbol{\mu}}$ be a dual feasible variable for Problem (D_{CMB}), and $\overline{\sigma}>0$ be sufficiently large so that $\sum_{i=1}^{l}\overline{\mu}_i M_i+\overline{\sigma}\eta I\succ C$ and $R_0+\sum_{i=1}^{l}\overline{\mu}_i R_i-\overline{\sigma}\eta I\prec 0$: the refined Slater condition holds for the perturbed problem (D_η). Hence, by Theorem 6, there exists a solution $(X^\eta, Y^\eta, \boldsymbol{\lambda^\eta})$ of Problem (P_η) in which rank $X^\eta\leqslant r$. We next show that $\langle C, X^\eta\rangle+\langle R_0, Y^\eta\rangle+\boldsymbol{h_0}^T\boldsymbol{\lambda^\eta}$ converges to the value of the supremum in Problem (P_{CMB}) as $\eta\to 0^+$, which will complete this proof.

Let η_k be a positive sequence decreasing to 0, and define $\gamma_k:=\langle C, X^{\eta_k}\rangle+\langle R_0, Y^{\eta_k}\rangle+\boldsymbol{h_0}^T\boldsymbol{\lambda^{\eta_k}}$. It is clear that γ_k is a nondecreasing sequence, because the constraints in Problem (P_η) become looser as η gets smaller, and γ_k is bounded from above by the value of the supremum γ^* in Problem (P_{CMB}). Therefore, $(\gamma_k)_{k\in\mathbb{N}}$ converges. Assume (*ad absurdum*) that the limit of this sequence is $\gamma_\infty<\gamma*$. Then, there are some variables $(X_0, Y_0, \boldsymbol{\lambda_0})$ that are feasible for (P_{CMB}), and such that $\langle C, X_0\rangle+\langle R_0, Y_0\rangle+\boldsymbol{h_0}^T\boldsymbol{\lambda_0}>\gamma_\infty$. But then, $(X_0, Y_0, \boldsymbol{\lambda_0})$ is also feasible for Problem (P_η), when $\eta\leqslant\eta_0:=(\text{trace } X_0+\text{trace } Y_0)^{-1}$. For any $k\in\mathbb{N}$ such that $\eta_k\leqslant\eta_0$, this contradicts the optimality of $(X^{\eta_k}, Y^{\eta_k}, \boldsymbol{\lambda^{\eta_k}})$ for Problem (P_{η_k}). Hence, $\gamma_\infty=\gamma*$ and the proof is complete.

References

[1] S. Arora, S. Rao, U. Vazirani, Expander flows, geometric embeddings and graph partitioning, J. ACM 56 (2) (2009) 1–37.
[2] A.I. Barvinok, Problems of distance geometry and convex properties of quadratic maps, Discrete Comput. Geom. 13 (1995) 189–202.
[3] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, MA, 1995.
[4] M. Bouhtou, S. Gaubert, G. Sagnol, Optimization of network traffic measurement: a semidefinite programming approach, in: Proceedings of the International Conference on Engineering Optimization (ENGOPT), Rio De Janeiro, Brazil, 2008, ISBN 978-85-7650-152-7.
[5] S. Burer, R.D.C. Monteiro, A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization, Math. Program. (Series B) 95 (2) (2003) 329–357.
[6] S. Burer, R.D.C. Monteiro, Local minima and convergence in low-rank semidefinite programming, Math. Program. (Series A) 103 (3) (2005) 427–444.
[7] A. d' Aspremont, L. El Ghaoui, M.I. Jordan, G.R.G. Lanckriet, A direct formulation for sparse PCA using semidefinite programming, SIAM Rev. 49 (3) (2007) 434–448 (electronic).
[8] M.X. Goemans, S.P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problem using semidefinite programming, J. ACM 42 (6) (1995) 1115–1145.
[9] G. Iyengar, D.J. Phillips, C. Stein, Approximating semidefinite packing programs, SIAM J. Optim. 21 (1) (2011) 231–268.
[10] D. Karger, R. Motwani, M. Sudan, Approximate graph coloring by semidefinite programming, J. ACM 45 (2) (1998) 246–265.
[11] L. Lovász, On the Shannon capacity of a graph, IEEE Trans. Inform. Theory 25 (1979) 1–7.
[12] A. Nemirovski, C. Roos, T. Terlaky, On maximization of quadratic form over intersection of ellipsoids with common center, Math. Program. 86 (1999) 463–473.
[13] G. Pataki, On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues, Math. Oper. Res. 23 (2) (1998) 339–358.
[14] F. Pukelsheim, On linear regression designs which maximize information, J. Statist. Plann. Inference 4 (1980) 339–364.
[15] F. Pukelsheim, Optimal Design of Experiments, Wiley, 1993.
[16] P. Richtarik, Simultaneously solving seven optimization problems in relative scale, Optimization online, Preprint number 2185, 2008.
[17] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[18] G. Sagnol, Computing optimal designs of multiresponse experiments reduces to second-order cone programming, J. Statist. Plann. Inference 141 (5) (2011) 1684–1708.
[19] G. Sagnol, S. Gaubert, M. Bouhtou, Optimal monitoring on large networks by successive c-optimal designs, in: 22nd International Teletraffic Congress (ITC22), Amsterdam, The Netherlands, September 2010.
[20] J.F. Sturm, Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones, Optim. Methods Softw. 11–12 (1999) 625–653.
[21] M. Szegedy, A note on the ϑ number of Lovász and the generalized Delsarte bound, SFCS94: Proceedings of the 35th Annual Symposium on Foundations of Computer Science, IEEE Computer Society, Washington, DC, USA, 1994, pp. 36–39.
[22] L. Vandenberghe, S. Boyd, S. Wu, Determinant maximization with linear matrix inequality constraints, SIAM J. Matrix Anal. Appl. 19 (1998) 499–533.