Dipartimento di Informatica e Sistemistica "Antonio Ruberti"
Università degli Studi di Roma "La Sapienza"

An unconstrained minimization method for solving low rank SDP relaxations of the max cut problem

Luigi Grippo
Laura Palagi
Veronica Piccialli

Technical Report n. 7
2007

Aracne

I *Technical Reports* del Dipartimento di Informatica e Sistemistica "Antonio Ruberti" svolgono la funzione di divulgare tempestivamente, in forma definitiva o provvisoria, i risultati di ricerche scientifiche originali. Per ciascuna pubblicazione vengono soddisfatti gli obblighi previsti dall'art. 1 del D.L.L. 31.8.1945, n. 660 e successive modifiche.
Copie della presente pubblicazione possono essere richieste alla Redazione.

Dipartimento di Informatica e Sistemistica "Antonio Ruberti"
Università degli Studi di Roma "La Sapienza"

Via Eudossiana, 18 - 00184 Roma
Via Buonarroti, 12 - 00185 Roma
Via Salaria, 113 - 00198 Roma

www.dis.uniroma1.it

www.aracneeditrice.it
info@aracneeditrice.it

Redazione:
00173 Roma
via Raffaele Garofalo, 133 A/B
06 93781065
telefax 72678427

ISBN 978–88–548–1276–5

I edizione: luglio 2007

Finito di stampare nel mese di luglio del 2007
dalla tipografia « Braille Gamma S.r.l. » di Santa Rufina di Cittaducale (Ri)
per conto della « Aracne editrice S.r.l. » di Roma
Printed in Italy

An unconstrained minimization method for solving low rank SDP relaxations of the max cut problem

L. Grippo*, L. Palagi*, V Piccialli*

*Università degli Studi di Roma "La Sapienza"
Dipartimento di Informatica e Sistemistica "A. Ruberti"
Via Ariosto 25 - 00185 Roma - Italy

e-mail (Grippo): grippo@dis.uniroma1.it
e-mail (Palagi): palagi@dis.uniroma1.it
e-mail (Piccialli): piccialli@dis.uniroma1.it

Abstract

In this paper we consider low-rank semidefinite programming (LRSDP) relaxations of the max cut problem. Using the Gramian representation of a positive semidefinite matrix, the LRSDP problem is transformed into the nonconvex nonlinear programming problem of minimizing a quadratic function with quadratic equality constraints. First, we establish some new relationships among these two formulations and we give necessary and sufficient conditions of global optimality. Then we propose a continuously differentiable exact merit function that exploits the special structure of the constraints and we use this function to define an efficient and globally convergent algorithm for the solution of the LRSDP problem. Finally, we test our code on an extended set of instances of the max cut problem and we report comparisons with other existing codes.

Keywords: semidefinite programming - low-rank factorization - max cut problem - nonlinear programming - exact penalty functions

1 Introduction

This paper concerns the solution of large scale Semidefinite Programming (SDP) problems arising as relaxations of the max cut problem in a graph. Given a simple undirected graph $G = (V, E)$ weighted on the edges, the max cut problem consists in finding a partition of vertices such that the sum of the weights on the edges between the two parts of the partition is maximum. The max cut problem is a well known NP-hard problem, and good bounds can be obtained by using convex SDP relaxations [17]. The simplest SDP relaxation of the max cut problem is of the form:

$$\begin{aligned} \min \quad & \mathrm{trace}\,(QX) \\ & \mathrm{diag}(X) = e \\ & X \succeq 0, \quad X \in \mathcal{S}^n, \end{aligned} \tag{1}$$

where the data matrix Q is an $n \times n$ real symmetric matrix, $\mathrm{trace}\,(QX)$ denotes the trace-inner product of matrices, $\mathrm{diag}(X)$ is the vector of dimension n containing all the diagonal elements of X and the $n \times n$ matrix variable X is required to be symmetric and positive semidefinite, as indicated with the notation $X \succeq 0,\ X \in \mathcal{S}^n$.
Several algorithms have been proposed in the literature for solving SDP problems, many of them belonging to the interior point class (see for example the survey [27] and references therein). In alternative to interior point methods, a recent trend has been developing algorithms based on nonlinear programming reformulations of the SDP problem. The first idea goes back to Homer and Peinado [22] who use the change of variables $X_{ij} = v_i^T v_j / \|v_i\| \|v_j\|$ for the elements of X, to transform problem (1) into an unconstrained optimization problem in the new variables $v_i \in R^n$ for $i = 1, \ldots, n$. In particular, they define a parallel computational scheme, in order to cope with the large dimensionality of the new problem. Burer and Monteiro in [6] propose a variant of Homer and Peinado's approach where they use the change of variables $X = LL^T$ where L is a lower triangular matrix.
More recently, Burer and Monteiro in [7, 8] recast a general linear SDP problem as a low rank semidefinite programming problem (LRSDP) by applying the change of variables $X = VV^T$, where V is a $n \times r$, $r < n$, rectangular matrix. The value of r is chosen by exploiting the result proved in Barvinok [1] and Pataki [31], that states that there exist an optimal solution of a linearly constrained SDP problem with rank r satisfying $r(r + 1) \leq m$, where m is the number of linear constraints. For the solution of the LRSDP problem, Burer and Monteiro propose an augmented Lagrangian method, which requires the solution of a sequence

of unconstrained problems for different values of a penalty parameter and of the Lagrange multipliers estimates.

In this paper we focus on the max cut SDP relaxation (1), and on the corresponding LRSDP problem. In fact, we also consider the reduced problem where we replace the variable X with a rectangular matrix V of dimension $n \times r$. First, we study optimality conditions and we establish necessary and sufficient conditions expressed in terms of the Lagrange multipliers for guaranteeing that a stationary point of the Lagrangian function of LRSDP problem yields a global minimizer that solves the original SDP problem. In particular, we show that known sufficient optimality conditions [7] can be proven to be also necessary.

Then we define a new unconstrained differentiable exact merit function for the computation of stationary points of the LRSDP problem. The augmented Lagrangian approach introduced in [7], although quite effective in practice, has two intrinsic drawbacks, since a sequence of unconstrained minimizations need to be performed, and some *a posteriori* assumptions on the behavior of the sequence are needed in order to prove global convergence. The exact penalty method defined in this paper overcomes both these drawbacks. Indeed, we need only a single unconstrained minimization of the merit function for a fixed sufficiently small value of the penalty parameter, and we can prove the global convergence of the algorithm without imposing any assumption on the behavior of the generated sequence. Therefore, we feel that, at least in the particular case of the problem arising from the relaxation of max cut, our approach fills the gap left by the seminal work of Burer and Monteiro [7, 8].

The paper is organized as follows. In Section 2, we report some known results on the max cut problem and we state the non linear relaxation that will be addressed in the paper. In particular, we use a Kronecker product notation to reformulate the problem in a standard NLP form. In Section 3, we review the main optimality conditions for the non linear programming problem and we state necessary and sufficient conditions for global optimality. In Section 4 we recast the original equality constrained problem as an unconstrained one, using a penalty function approach that exploits the special structure of the problem. In Section 5 we define a globally convergent algorithm for the computation of stationary points of the LRSDP problem and finally in Section 6 we report extensive numerical results on standard instances of the max cut problem.

1.1 Notation and terminology

We denote by $\mathbb{R}^n$ the space of real $n-$dimensional columns vector and by $\mathbb{R}^{n\times m}$ the space of real $n \times m$ matrices. By $\mathcal{S}^n$ we indicate the space of real $n \times n$ symmetric matrices.
Given two $n\times n$ square matrices Q and A we define the usual trace-inner product by letting

$$\text{trace}\,(QA) = \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}a_{ij},$$

and we indicate by $\|A\|$ the induced Frobenius norm:

$$\|A\|^2 = \text{trace}\,(A^TA).$$

If $v \in \mathbb{R}^n$, $\|v\|$ is intended as the Euclidean norm of v. Given a square matrix $A \in \mathbb{R}^{n\times n}$, we denote by $\text{diag}(A)$ the vector of dimension n containing all the diagonal elements of A. Given a vector $a \in \mathbb{R}^n$, we denote by $\text{Diag}(a)$ the diagonal square matrix of dimension n, with the elements of a on the diagonal. Moreover, we indicate by e_i the vector of zeroes elements except for the i-th equal to one, by e the vector of all ones, and we set $E_{ii} = \text{Diag}(e_i)$, $I_p = \text{Diag}(e)$ with $e \in \mathbb{R}^p$.
In the paper, we make use of the Kronecker product $\otimes$ (see, for instance, [24]). We recall that given two matrices A $m \times n$ and B $p \times q$, the Kronecker product $A \otimes B$ is the $mp \times nq$ matrix given by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}B & a_{n2}B & \dots & a_{mn}B \end{bmatrix}.$$

The basic properties of the Kronecker product are the following identities

$$A \otimes B \otimes C = (A \otimes B) \otimes C = A \otimes (B \otimes C),$$

$$(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D,$$

$$(A \otimes B)(C \otimes D) = (AC \otimes BD),$$

where we assume that all the matrix operations appearing in each identity can be performed. Note that here and in the sequel, in order to simplify notation, we indicate by $(AC \otimes BD)$ the Kronecker product $((AC) \otimes (BD))$.
The transpose of a Kronecker product is: $(A \otimes B)^T = A^T \otimes B^T$. Given a matrix $A \in \mathcal{S}^n$, with spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$ and a matrix

$B \in \mathcal{S}^m$, with spectrum $\sigma(B) = \{\mu_1, \dots, \mu_m\}$, the spectrum of $A \otimes B$ is given by:

$$\sigma(A \otimes B) = \{\lambda_i \mu_j \ : i = 1, \dots, n; \ j = 1, \dots, m\}.$$

Furthermore, in the Frobenius norm, we have that

$$\|A \otimes B\| = \|A\| \|B\|.$$

2 SDP formulation and relaxations of max cut

Let $G(V, E)$ be a weighted undirected graph, with $n = |V|$ nodes and weights w_{ij} for $(i, j) \in E$. Let $A \in \mathcal{S}^n$ be the weighted adjacency matrix

$$a_{ij} = \begin{cases} w_{ij} & (i,j) \in E \\ 0 & \text{otherwise} \end{cases}.$$

The max cut problem consists in finding a partition of the set of nodes V of the weighted undirected graph G so as to maximize the sum of the weights on the edges that have one end in each side of the partition. Let the vector $x \in \{-1, 1\}^n$ represent any cut in the graph, i.e. the sets $\{i \in 1, \dots, n : x_i = +1\}$ and $\{i \in 1, \dots, n : x_i = -1\}$ constitute a partition of the sets of nodes. Then the weight of the cut induced by the partition is given by

$$\sum_{i<j} a_{ij} \frac{(1 - x_i x_j)}{2}.$$

Let $L := \mathrm{Diag}(Ae) - A$ denote the Laplacian matrix associated with the graph. Then it is straightforward to check that

$$\frac{1}{4} x^T L x = \sum_{i<j} a_{ij} \frac{(1 - x_i x_j)}{2},$$

and hence the max cut problem can be formulated as

$$\begin{array}{ll} \max & \frac{1}{4} x^T L x \\ & x_i^2 = 1, \quad i = 1, \dots, n. \end{array} \qquad (2)$$

By letting $Q = \frac{1}{4} A$, we can solve equivalently the problem

$$\begin{array}{ll} \min & x^T Q x \\ & x_i^2 = 1, \quad i = 1, \dots, n. \end{array} \qquad (3)$$

Indeed, if $x^*_{\rm MC}$ is an optimal solution of problem (3), and we denote by

$$z^*_{MC} = {x^*_{\rm MC}}^T Q x^*_{\rm MC},$$

the value of the optimal cut is given by

$$m_C = -z^*_{MC} + \frac{1}{2}\sum_{i<j} a_{ij}.$$

It is possible to reformulate the max cut problem as an SDP problem with a rank constraint (see for example [33], [16]). Indeed, let us introduce the following rank one matrix:

$$X = xx^T,$$

whose generic elements are $X_{ij} = x_i x_j$. It is easy to see that

$$x^T Q x = \sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j = \sum_{i=1}^n \sum_{j=1}^n q_{ij} X_{ij} = \text{trace}\,(QX).$$

Moreover, as a rank one matrix $X \succeq 0$ can always be written in the form xx^T for some $x \in \mathbb{R}^n$, we can rewrite problem (3) as

$$\begin{array}{ll} \min & \text{trace}\,(QX) \\ & \text{diag}(X) = e \\ & \text{rank}\,(X) = 1 \\ & X \succeq 0, X \in \mathcal{S}^n, \end{array} \tag{4}$$

where $\text{diag}(X) = e$ replaces the original constraints $x_i^2 = 1$ for $i = 1, \dots, n$.

A well known SDP relaxation of max cut (see, for instance, [9], [17], [26], [30]) is obtained by removing from formulation (4) the rank one constraint, so that we get the SDP problem:

$$\begin{array}{ll} \min & \text{trace}\,(QX) \\ & \text{diag}(X) = e \\ & X \succeq 0, X \in \mathcal{S}^n. \end{array} \tag{SDP$_{\rm MC}$}$$

We recall (see for example [23]) that a matrix $X \in \mathcal{S}^n$ is positive semidefinite if and only if there exists a set of vectors $S = \{v_1, \dots, v_n\}$, $v_i \in \mathbb{R}^n$ such that X is the Gram matrix of S with respect to the Euclidean inner product, i.e. $X_{ij} = v_i^T v_j$. Thus it is possible to make the change of variables $X = VV^T$, where $V = [v_1 \dots v_n]^T \in \mathbb{R}^{n\times n}$ so that the semidefiniteness constraint can be removed and Problem ($\mathrm{SDP_{MC}}$) is equivalent to the following problem:

$$\begin{aligned} \min \quad & \mathrm{trace}\,(QVV^T) \\ & \mathrm{diag}(VV^T) = e, \quad V \in \mathbb{R}^{n\times n}. \end{aligned} \tag{5}$$

The equivalence between problem (5) and problem ($\mathrm{SDP_{MC}}$) means that there is one-to-one correspondence among global solutions.
We observe that the relaxation (5) can be derived directly from Problem (3) by enlarging the space of variables. In fact, if we replace the variable x_i with a vector $v_i \in \mathbb{R}^n$ we get exactly Problem (5). However, deriving this relaxation by SDP, gives the important additional information that this problem can be solved in polynomial time (see [19], [29], [30]).
Problem (5) has n^2 variables, but it is possible to show that the actual number of variables needed to obtain an optimal solution of Problem ($\mathrm{SDP_{MC}}$) is much smaller.
Let $\mathcal{X}^*_{\mathrm{SDP}}$ be the set of all the optimal solutions of Problem ($\mathrm{SDP_{MC}}$) and define the integer $r_{\min}$ as

$$r_{\min} = \min_{X\in\mathcal{X}^*_{\mathrm{SDP}}} \mathrm{rank}\,(X).$$

It is well known (see for example Theorem 4.5.8 in [23]) that a $n \times n$ matrix $X \succeq 0$ of rank r is congruent to a diagonal matrix consisting of all zeros and r ones, namely

$$X = \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} I_r & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = V_1 V_1^T,$$

where V_1 is a $n \times r$ matrix. Thus, X is the Gram matrix in $\mathbb{R}^r$ of the set of rows of V_1. Hence if the value $r_{\min}$ were known, the dimension of the matrix V in Problem (5) could be reduced by using the above result. Indeed, let $v \in \mathbb{R}^{nr}$ be the vector $v = (v_1^T, \dots, v_n^T)^T$. Then, we can consider the following quadratic problem with quadratic constraints

$$\begin{aligned} \min \quad & q_r(v) := \sum_{i=1}^n \sum_{j=1}^n q_{ij} v_i^T v_j \\ & \|v_i\|^2 = 1, \quad i = 1,\dots,n, \qquad v_i \in \mathbb{R}^r. \end{aligned} \tag{6}$$

For any $r \ge r_{\min}$ a global solution of Problem (6) gives a global solution of Problem (5) and hence of Problem ($\mathrm{SDP_{MC}}$). The value of $r_{\min}$ is not known, but an upper bound on $r_{\min}$ can be easily computed. In fact, in [1] and [31] a result for SDP problems has been proved that, specialized to the max cut relaxation ($\mathrm{SDP_{MC}}$), can be stated as follows.

Proposition 1 *There exists an $X \in \mathcal{S}^n$ optimal solution of* (SDP$_{\rm MC}$) *with rank r satisfying the inequality*

$$r(r+1)/2 \le n. \qquad \square$$

This result implies that $r_{\min} \le \widehat{r}$ where

$$\widehat{r} = \max\{k \in N : k(k+1)/2 \le n\} = \left\lfloor \frac{\sqrt{1+8n}-1}{2} \right\rfloor. \tag{7}$$

Thus, Problem (6) gives a global solution of Problem (SDP$_{\rm MC}$) for all $r \ge \widehat{r}$, and $\widehat{r}$ can be evaluated. This result has already been exploited in [17] and [7].
To summarize, let $X_r^* \in \mathcal{X}_{\rm SDP}^*$ be an optimal solution of Problem (SDP$_{\rm MC}$) with rank $r \ge r_{\min}$, and denote by $z_{\rm SDP}^*$ the optimal objective function value , i.e. $z_{\rm SDP}^* = \mathrm{trace}\,(QX_r^*)$. Let q_r^* denote the optimal objective function value of Problem (6) where $r \ge r_{\min}$.
Problem (6) is equivalent to Problem (SDP$_{\rm MC}$), in the sense that every global solution $v^* \in R^{nr}$ of Problem (6) gives a global solution X^* of Problem (SDP$_{\rm MC}$) where $X_{ij}^* = {v_i^*}^T v_j^*$ and we have $z_{\rm SDP}^* = q_r^*$. We can conclude that for all $r \ge r_{\min}$, it holds

$$z_{\rm SDP}^* = q_n^* = \ldots = q_r^* = \ldots = q_{\widehat{r}}^* = \ldots = q_{r_{\min}}^*. \tag{8}$$

Moreover, when $r = 1$ we get a problem equivalent to Problem (4), so that $q_1^* \equiv z_{\rm MC}^*$, and for increasing values of r, we get non increasing values of the optimal value (the feasible region is enlarging). Hence we can write

$$z_{\rm SDP}^* \equiv q_n^* \equiv q_{\widehat{r}}^* \equiv q_{r_{\min}}^* \le q_r^* \le q_1^* \equiv z_{\rm MC}^*. \tag{9}$$

Problem (6) is a non convex problem with quadratic objective function and quadratic constraints that can be recast in compact NLP form using Kronecker products. In fact, letting $e_i \in \mathbb{R}^n$, we can write the vector $v \in \mathbb{R}^{nr}$ as

$$v = \sum_{i=1}^n (e_i \otimes v_i) = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

so that we have

$$v_i = (e_i \otimes I_r)^T v.$$

Therefore we can write the objective function of Problem (6) as:

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n q_{ij} v_i^T v_j = & \sum_{i=1}^n \sum_{j=1}^n q_{ij} \left((e_i \otimes I_r)^T v\right)^T (e_j \otimes I_r)^T v \\ = & \sum_{i=1}^n \sum_{j=1}^n v^T (q_{ij} e_i e_j^T \otimes I_r) v = v^T (Q \otimes I_r) v, \end{aligned}$$

while the constraints can be written as

$$\|v_i\|^2 = v^T\,(E_{ii} \otimes I_r)\,v.$$

Thus, we obtain the nonlinear programming problem

$$\begin{array}{ll} \min & v^T(Q \otimes I_r)v \;=\; q_r(v) \\ & v^T\,(E_{ii} \otimes I_r)\,v = 1 \qquad i = 1, \dots, n, \end{array} \qquad (\mathrm{NLP_r})$$

that is the problem we will focus on in the rest of the paper.

3 Optimality conditions

In this section we study both local and global optimality conditions for problem (NLP_r).
The Lagrangian function for Problem (NLP_r) is, for an arbitrary fixed value $r \geq 1$,

$$\begin{aligned} L(v, \lambda) = &\quad v^T\,(Q \otimes I_r)\,v + \sum_{i=1}^{n} \lambda_i \left(v^T\,(E_{ii} \otimes I_r)\,v - 1\right) \\ = &\quad v^T[(Q + \Lambda) \otimes I_r]v - \lambda^T e \end{aligned} \qquad (10)$$

where $\lambda = (\lambda_1, \dots, \lambda_n)^T$ and we set $\Lambda = \mathrm{Diag}\{\lambda\}$. We recall the definition of a stationary point of Problem (NLP_r).

Definition 1 (Stationary point) *A point $\hat{v} \in \mathbb{R}^{nr}$ is a stationary point of Problem* (NLP_r) *if there exists a Lagrange multiplier $\hat{\lambda} \in \mathbb{R}^n$ such that $(\hat{v}, \hat{\lambda}) \in \mathbb{R}^{nr} \times \mathbb{R}^n$ satisfies*

$$\begin{array}{l} \hat{v}^T\,(E_{ii} \otimes I_r)\,\hat{v} = 1, \quad i = 1, \dots, n \\ \nabla_v L(\hat{v}, \hat{\lambda}) = 0, \end{array}$$

where $\nabla_v L(\hat{v}, \hat{\lambda}) \equiv 2[(Q + \hat{\Lambda}) \otimes I_r]\hat{v}$. □

In the following proposition we state the well known necessary optimality conditions for Problem (NLP_r).

Proposition 2 (First order necessary conditions) *Let $\hat{v} \in \mathbb{R}^{nr}$ be a local minimizer of Problem* (NLP_r)*. Then there exists unique $\hat{\lambda} \in \mathbb{R}^n$ such that $(\hat{v}, \hat{\lambda}) \in \mathbb{R}^{nr} \times \mathbb{R}^n$ satisfies:*

$$\begin{array}{lll} [(\,Q + \hat{\Lambda}\,) \otimes I_r]\,\hat{v} & = & 0 \\ \hat{v}^T\,(E_{ii} \otimes I_r)\,\hat{v} & = & 1, \quad i = 1, \dots, n. \end{array} \qquad (11)$$

Proof. As the constraints of Problem (NLP$_r$) satisfy the linear independence constraints qualification, then $\hat{v}$ is a stationary point and the Lagrange multiplier is unique. By Definition 1 of stationary point we have (11). □

The second order necessary conditions for Problem (NLP$_r$) are given in the next proposition.

Proposition 3 (Second order necessary conditions) *Let $\hat{v} \in \mathbb{R}^{nr}$ be a local minimizer of Problem* (NLP$_r$)*. Then there exists $\hat{\lambda} \in \mathbb{R}^n$ such that $(\hat{v}, \hat{\lambda}) \in \mathbb{R}^{nr} \times \mathbb{R}^n$ satisfies* (11) *and*

$$z^T \left[\left(Q + \hat{\Lambda} \right) \otimes I_r \right] z \geq 0$$

for every $z \in \mathbb{R}^{nr}$ such that $\hat{v}^T (E_{ii} \otimes I_r) z = 0$ for $i = 1, \ldots, n$.

Now, we state some useful properties of the Lagrange multipliers at a stationary point.

Proposition 4 *Let $\hat{v} \in \mathbb{R}^{nr}$ be a stationary point of Problem* (NLP$_r$)*. Then we have:*

$$\hat{\lambda}_i = -\hat{v}^T (E_{ii} Q \otimes I_r) \hat{v}, \quad i = 1, \ldots, n, \tag{12}$$

and

$$\sum_{i=1}^{n} \hat{\lambda}_i = -q_r(\hat{v}). \tag{13}$$

Proof. Let $\hat{v}$ be a stationary pair of Problem (NLP$_r$) so that

$$\frac{1}{2} \nabla_v L(\hat{v}, \hat{\lambda}) \equiv (Q \otimes I_r) \hat{v} + \sum_{j=1}^{n} \hat{\lambda}_j (E_{jj} \otimes I_r) \hat{v} = 0. \tag{14}$$

Premultiplying both sides of (14) by $\hat{v}^T (E_{ii} \otimes I_r)$ we can write:

$$\hat{v}^T (E_{ii} \otimes I_r) (Q \otimes I_r) \hat{v} + \sum_{j=1}^{n} \hat{\lambda}_j \hat{v}^T (E_{ii} \otimes I_r) (E_{jj} \otimes I_r) \hat{v} = 0,$$

whence, using the properties of Kronecker products and noting that $E_{ii} E_{jj} = 0_{n \times n}$ for $i \neq j$ and $E_{ii} E_{ii} = E_{ii}$, we get

$$\hat{v}^T (E_{ii} Q \otimes I_r) \hat{v} + \hat{\lambda}_i \hat{v}^T (E_{ii} \otimes I_r) \hat{v} = 0.$$

Therefore, as $\hat{v}^T(E_{ii} \otimes I_r)\hat{v} = 1$, we obtain expression (12) for the multipliers. Finally, summing up all the multipliers, and noting that $\sum_{i=1}^{n} E_{ii} = I_n$ we obtain

$$\begin{aligned}\sum_{i=1}^n \hat{\lambda}_i &= -\hat{v}^T \sum_{i=1}^n (E_{ii} Q \otimes I_r)\hat{v} \\ &= -\hat{v}^T [(\sum_{i=1}^n E_{ii}) Q \otimes I_r]\hat{v} = -\hat{v}^T (Q \otimes I_r)\hat{v},\end{aligned}$$

which yields (13). □

We pointed out in Section 2 that for suitable values of r any global solution of the non convex Problem (NLP$_r$) gives a global solution of the convex Problem (SDP$_{\text{MC}}$). Duality theory applied to Problem (SDP$_{\text{MC}}$) and the connections between Problem (SDP$_{\text{MC}}$) and Problem (NLP$_r$) give us a straightforward global necessary and sufficient optimality condition. Letting $u \in \mathbb{R}^n$, consider the standard Lagrangian dual of Problem (SDP$_{\text{MC}}$):

$$\begin{array}{ll}\max & e^T u \\ & Q - \text{Diag}(u) \succeq 0.\end{array} \tag{15}$$

Denote by $u^* \in \mathbb{R}^n$ an optimal solution of Problem (15), and let

$$e^T u^* = z^*_{\text{DUAL}}.$$

Since Slater's constraint qualification holds both for the primal and the dual problem, it is well known that there is no duality gap. Therefore, X^* and u^* are optimal solutions of the primal problem (SDP$_{\text{MC}}$) and of its dual (15) respectively if and only if:

$$\begin{aligned} z^*_{\text{SDP}} = \text{trace}(QX^*) &= e^T u^* = z^*_{\text{DUAL}} \\ \text{diag}(X^*) &= e \\ X^* &\succeq 0 \\ Q - \text{Diag}(u^*) &\succeq 0. \end{aligned} \tag{16}$$

By posing $u_i = -y_i$ for $i = 1, \ldots, n$ we can write problem (15) as

$$\begin{array}{ll}\min & e^T y \\ & Q + Y \succeq 0,\end{array}$$

where $Y = \text{Diag}(y)$.
The next proposition gives a necessary and sufficient global optimality condition for Problem (NLP$_r$) that simply derives from the optimality conditions (16) and from the relation (8).

We remark that, from now on, for sake of simplicity, we adopt the following terminology: whenever we say that a point $v^* \in \mathbb{R}^{nr}$ *solves Problem* (SDP$_{\text{MC}}$) we mean that $X^* = V^*V^{*T}$, where $V^* = (v_1^* \dots v_n^*)^T$, is an optimal solution of Problem (SDP$_{\text{MC}}$), namely

$$V^*V^{*T} \in \mathcal{X}^*_{\text{SDP}}, \text{ and } q_r(v^*) = z^*_{\text{SDP}}.$$

This implies, by definition, that $r \geq r_{\min}$.

Proposition 5 *A point $v^* \in \mathbb{R}^{nr}$ is a global minimizer of Problem* (NLP$_r$) *that solves Problem* (SDP$_{\text{MC}}$) *if and only if there exists a $y^* \in \mathbb{R}^n$ such that (v^*, y^*) satisfies*

$$\begin{array}{l} -e^T y^* = q_r(v^*) \\ Q + Y^* \succeq 0 \\ v^{*T}(E_{ii} \otimes I_r)\; v^* = 1, \qquad i = 1, \dots, n. \end{array} \tag{17}$$

Proof The proof easily follows from the primal dual SDP optimality conditions (16), where we set $u^* = -y^*$, and from relation (8) that together give for all $r \geq r_{\min}$

$$q_r^* = z^*_{\text{SDP}} = z^*_{\text{DUAL}} = -e^T y^* = q_r(v^*).$$

□

We note that Proposition 5 gives a necessary and sufficient condition of global optimality in terms of primal-dual relationships, so that checking this condition requires the solution of the dual problem (15), which is an SDP problem as well. However, we show that the solution u^* of the dual problem actually is obtained by the Lagrange multiplier λ^* associated with the solution v^* of Problem (NLP$_r$). This yields a necessary and sufficient global optimality condition that is similar in structure to the one valid for the trust region problem [28], in that it requires the Hessian of the Lagrangian to be positive semidefinite at a stationary point.

Proposition 6 *A point $v^* \in \mathbb{R}^{nr}$ is a global minimizer of Problem* (NLP$_r$) *that solves Problem* (SDP$_{\text{MC}}$) *if and only if there exists a $\lambda^* \in \mathbb{R}^n$ such that*

$$\begin{array}{l} [(Q + \Lambda^*) \otimes I_r]\, v^* = 0 \\ Q + \Lambda^* \succeq 0 \\ v^{*T}(E_{ii} \otimes I_r)\; v^* = 1, \qquad i = 1, \dots, n. \end{array} \tag{18}$$

Proof First assume that (18) are satisfied. By (13), we have $q_r(v^*) = -e^T\lambda^*$. The vector $u^* = -\lambda^*$ is feasible for the dual problem (15), and hence u^* is optimal for the dual. Therefore, the primal dual optimality conditions (16) and (8) together give

$$q_r^* = z_{\text{SDP}}^* = z_{\text{DUAL}}^* = -e^T\lambda^* = q_r(v^*).$$

As for the necessity part, the first order necessary conditions state that a unique $\lambda^* \in \mathbb{R}^n$ exists such that

$$[(Q + \Lambda^*) \otimes I_r]\, v^* = 0. \tag{19}$$

Moreover, since by assumption $v^* \in \mathbb{R}^{nr}$ solves Problem (SDP_{MC}), we get by Proposition 5 that there exist $y^* \in \mathbb{R}^n$ such that

$$\begin{aligned} &-e^T y^* = q_r(v^*) \\ &Q + Y^* \succeq 0, \end{aligned}$$

and hence:

$$-e^T y^* = v^{*T}\,(Q \otimes I_r)\, v^*. \tag{20}$$

Since $v^{*T}(E_{ii} \otimes I_r)v^* = 1, \quad i = 1, \ldots, n$, we can write

$$e^T y^* = \sum_{i=1}^{n} v^{*T}(y_i^* E_{ii} \otimes I_r)v^* = v^{*T}(Y^* \otimes I_r)v^*$$

that summed up with (20) gives

$$v^{*T}\,[(Q + Y^*) \otimes I_r]\, v^* = 0.$$

Since $Q + Y^*$ is positive semidefinite, it follows that

$$[(Q + Y^*) \otimes I_r]\, v^* = 0.$$

The unicity of the multiplier λ^* satisfying (19) implies that $\lambda^* = y^*$, and hence

$$Q + \Lambda^* = Q + Y^* \succeq 0,$$

so that (18) holds. □

The advantage of the condition stated in Proposition 5 is that this condition can be computationally checked without solving the dual problem (15), since it requires only the knowledge of the Lagrange multiplier associated to the point v^*.
We point out once more that problem (NLP_r) is a non convex optimization problem so that necessary and sufficient global optimality conditions are usually not available. However, we were able to derive the above condition, thanks to the strict connection between problem (NLP_r) for

$r \geq r_{\min}$ and the convex problem (SDP_{MC}), which indicates some sort of hidden convexity. Hidden convexity of other quadratic problems with quadratic constraints has been exploited to derive global optimality conditions for non convex problems [3, 28, 35]. We remark that, for $r < r_{\min}$, conditions (18) do not apply and global optimality conditions are not available.
A different sufficient global optimality condition has been proved in [7] by Burer and Monteiro for a more general class of SDP problems. In particular, for $r < n$ they prove the following result (here specialized to Problem (SDP_{MC})) that gives a sufficient condition of global optimality.

Proposition 7 (Proposition 4 in [7]) *Let $v^* \in \mathbb{R}^{nr}$, with $r < n$, be a local minimum point of Problem (NLP_r). Let $\hat{v} \in \mathbb{R}^{n(r+1)}$ be a point with components $\hat{v}_i \in \mathbb{R}^{r+1}$ such that*

$$\hat{v}_i = \begin{pmatrix} v_i^* \\ 0 \end{pmatrix}.$$

If $\hat{v}$ is a local minimum of Problem (NLP_{r+1}), then v^* is a global minimum point of Problem (NLP_r) that solves Problem (SDP_{MC}).

□

Actually, by looking at the details of the proof of the above result in [7], it emerges that the only assumption needed is that v^* and $\hat{v}$ satisfy only necessary conditions. We show that this condition is also necessary, and by exploiting this result we can establish another necessary and sufficient condition that can be computationally checked. Indeed, we can state the following result.

Proposition 8 *A point $v^* \in \mathbb{R}^{nr}$, with $r < n$, is a global minimum point of Problem (NLP_r) that solves Problem (SDP_{MC}) if and only if the following conditions hold:*

(i) *v^* is a stationary point for Problem (NLP_r),*

(ii) *the point $\hat{v} \in \mathbb{R}^{n(r+1)}$ with components $\hat{v}_i \in \mathbb{R}^{r+1}$ defined as*

$$\hat{v}_i = \begin{pmatrix} v_i^* \\ 0 \end{pmatrix} \tag{21}$$

is a stationary point for Problem (NLP_{r+1}) satisfying the second order necessary optimality conditions.

Proof First of all, we prove sufficiency. Let $\lambda^* \in \mathbb{R}^n$ be the unique Lagrange multiplier associated to $v^* \in \mathbb{R}^{nr}$ for Problem (NLP_r) and let

$\hat{\lambda} \in \mathbb{R}^n$ be the unique Lagrange multiplier associated to $\hat{v} \in \mathbb{R}^{n(r+1)}$, such that $(\hat{v}, \hat{\lambda})$ satisfies the first order necessary optimality conditions (11) for Problem (NLP$_{r+1}$). As a first step, we show that $\hat{\lambda} = \lambda^*$. In fact, it follows from the expression of $\hat{v}$, setting $q^*_{ij} = (Q + \Lambda^*)_{ij}$:

$$
\begin{aligned}
\left[(Q+\Lambda^*)\otimes I_{r+1}\right]\hat{v} &= \begin{pmatrix} q^*_{11} I_r & 0_r & \cdots & q^*_{1n} I_r & 0_r \\ 0_r^T & q^*_{11} & & 0_r^T & q^*_{1n} \\ \vdots & & \ddots & \vdots & \\ q^*_{n1} I_r & 0_r & \cdots & q^*_{nn} I_r & 0_r \\ 0_r^T & q^*_{n1} & & 0_r^T & q^*_{nn} \end{pmatrix} \begin{pmatrix} v^*_1 \\ 0 \\ \vdots \\ v^*_n \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^n q^*_{1j} v^*_j \\ 0 \\ \vdots \\ \sum_{j=1}^n q^*_{nj} v^*_j \\ 0 \end{pmatrix}.
\end{aligned}
\tag{22}
$$

Since (v^*, λ^*) is a stationary point of Problem (NLP$_r$), it follows from (22) that

$$\left[(Q+\Lambda^*)\otimes I_{r+1}\right]\hat{v} = 0. \tag{23}$$

As the Lagrange multiplier $\hat{\lambda}$ is unique, (23) implies $\hat{\lambda} = \lambda^*$. Hence $\hat{v}$ is a stationary point of Problem (NLP$_r$) with Lagrange multiplier λ^*.

Now, for any $w = (w_1, \ldots, w_n)^T \in \mathbb{R}^n$, let us define the vector $z \in \mathbb{R}^{nr}$

$$z^T = (\, 0_r^T \;\; w_1 \;\; 0_r^T \;\; w_2 \;\; \cdots \;\; 0_r^T \;\; w_n \,)$$

which satisfies

$$\hat{v}^T \left[E_{ii} \otimes I_{r+1}\right] z = 0, \qquad \text{for all } i = 1, \ldots, n. \tag{24}$$

Hence, by the second order necessary conditions for Problem (NLP$_{r+1}$), we must have $z^T \left[(Q+\Lambda^*)\otimes I_{r+1}\right] z \geq 0$ and therefore, by the expression of z we get

$$0 \le z^T \left[(Q+\Lambda^*)\otimes I_{r+1}\right] z = \begin{pmatrix} 0_r^T & w_1 & \dots & 0_r^T & w_n \end{pmatrix} \begin{pmatrix} 0_r \\ \sum_{j=1}^n q_{1j}^* w_j \\ \vdots \\ 0_r \\ \sum_{j=1}^n q_{nj}^* w_j \end{pmatrix}$$

$$= \sum_{j=1}^n \sum_{i=1}^n q_{ij}^* w_i w_j = w^T(Q+\Lambda^*)w, \qquad (25)$$

where w is any vector in $\mathbb{R}^n$, which implies $(Q+\Lambda^*) \succeq 0$. Then the global optimality of v^* follows from relation (13), that says $q_r(v^*) = -e^T\lambda^*$, and from Proposition 5.

Now we prove the necessity part. Let $v^* \in \mathbb{R}^{nr}$ be a global minimum point of Problem (NLP$_r$) that solves Problem (SDP$_{\mathrm{MC}}$). Then, there exists $\lambda^* \in \mathbb{R}^n$ such that (v^*, λ^*) satisfies first order necessary optimality conditions for Problem (NLP$_r$). Let us define the vector $\hat{v} \in \mathbb{R}^{n(r+1)}$ with vector components defined by (21), which is obviously feasible for Problem (NLP$_{r+1}$). We have that

$$q_{r+1}(\hat{v}) = q_r(v^*) = z^*_{\mathrm{SDP}},$$

so that, by (8), $\hat{v}$ is a global minimum point of Problem (NLP$_{r+1}$). Moreover, by using again (23), it follows from v^* being a stationary point of Problem (NLP$_r$) with Lagrange multiplier λ^* that $(\hat{v}, \lambda^*)$ satisfies

$$\left[(Q+\Lambda^*)\otimes I_{r+1}\right]\hat{v} = 0,$$

namely $\hat{v}$ is a stationary point for Problem (NLP$_{r+1}$) with Lagrange multiplier λ^*. Again, by the uniqueness of the multipliers for Problem (NLP$_{r+1}$) and by the global optimality of $\hat{v}$, we have that $(\hat{v}, \lambda^*)$ satisfies also the second order optimality conditions of Problem (NLP$_{r+1}$), and this completes the proof. □

The optimality conditions derived so far can be related to the eigenvalue bounds given in [32] specialized to Problem (SDP$_{\mathrm{MC}}$). Indeed, we observe that Problem (NLP$_r$) is equivalent to the following problem, where we add the redundant constraint $\|v\|^2 = n$:

$$\begin{array}{ll} \min & v^T (Q \otimes I_r)\, v \\ & v^T (E_{ii} \otimes I_r)\, v = 1, \quad i = 1, \dots, n \\ & \|v\|^2 = n. \end{array} \tag{26}$$

In connection with problem (26) we can define the function

$$\Psi(\lambda) \quad = \quad \min_{\|v\|^2 = n} L(v, \lambda),$$

where $L(v, \lambda)$ is the Lagrangian function of Problem (NLP$_r$) given by (10). We can state the following result.

Proposition 9 *For every $\lambda \in \mathbb{R}^n$, and for every v such that $\|v_i\|^2 = 1$ for $i = 1, \dots, n$, the following inequalities hold*

$$n\lambda_{\min}\,[Q + \Lambda] - e^T\lambda \leq z^*_{\mathrm{SDP}} \leq q^*_r \leq q_r(v). \tag{27}$$

Proof Recalling the properties of Rayleigh quotient, and the properties of the spectrum of the Kronecker product of two matrices, we get

$$\Psi(\lambda) = n\lambda_{\min}\,[Q + \Lambda] - e^T\lambda.$$

It is easy to see that for every $\lambda \in \mathbb{R}^n$ $\Psi(\lambda) \leq q^*_r$, and, since $\Psi(\lambda)$ does not depend on r, we have also $\Psi(\lambda) \leq z^*_{\mathrm{SDP}}$. The right side of the inequality follows from the definition of q^*_r. □

Now, let $(\hat{v}, \hat{\lambda})$ be a stationary point of Problem (NLP$_r$) and assume that

$$\lambda_{\min}(Q + \hat{\Lambda}) = 0.$$

Then, keeping into account again (13), we have

$$q(\hat{v}) = -e^T\hat{\lambda} \leq q^*_r,$$

so that we must have $q(\hat{v}) = q^*_r = z^*_{\mathrm{SDP}}$, which is exactly the sufficient condition stated in Proposition 6.

4 An exact penalty function

In this section, for an arbitrary $r \geq 1$ we consider Problem (NLP$_r$) that is:

$$\begin{array}{ll} \min & q_r(v) \\ & h_i(v) = 1, \quad i = 1, \dots, n, \end{array}$$