
DIPARTIMENTO DI INFORMATICA
E SISTEMISTICA ANTONIO RUBERTI

SAPIENZA
UNIVERSITÀ DI ROMA


***An unconstrained approach for solving low rank SDP relaxations of {-1, 1} quadratic problems***


**Luigi Grippo**
**Laura Palagi**
**Mauro Piacentini**
**Veronica Piccialli**



**Technical Report n. 13, 2009**

# An unconstrained approach for solving low rank SDP relaxations of $\{-1, 1\}$ quadratic problems

Luigi Grippo*, Laura Palagi*, Mauro Piacentini*, Veronica Piccialli†

**Abstract**

We consider low-rank semidefinite programming (LRSDP) relaxations of $\pm 1$ quadratic problems that can be formulated as the nonconvex nonlinear programming problem of minimizing a quadratic function subject to separable quadratic equality constraints. We prove the equivalence of the LRSDP problem with the unconstrained minimization of a new merit function and we define an efficient and globally convergent algorithm for finding critical points 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 - boolean quadratic problem - nonlinear programming

*Sapienza Università di Roma - Dipartimento di Informatica e Sistemistica "A. Ruberti" - via Ariosto, 25 - 00185 Roma, Italy.

†Università di Tor Vergata - Dipartimento di Ingegneria dell'Impresa - via del Politecnico, 1-00133 Roma, Italy

## 1 Introduction

We consider a semidefinite programming problem in the form

$$\min_X \ \{Q \bullet X : \ \mathrm{diag}(X) = e, \ X \succeq 0\}, \tag{SDP}$$

where $Q \in \mathcal{S}^n$ is given, $X \in \mathcal{S}^n$ and $e \in R^n$ is the vector of all ones. Here $\mathcal{S}^n$ denotes the space of $n \times n$ symmetric matrices, and $X \succeq 0$ indicates that $X$ is positive semidefinite. Semidefinite Programming (SDP) problems of this form arise as relaxations of $\{-1, 1\}$ quadratic problems (see e.g. [6], [8], [14]) :

$$\begin{array}{ll} \min & x^T Q x \\ \text{s.t.} & x \in \{-1, 1\}^n. \end{array} \tag{1}$$

Efficient solution of problem (SDP) is of great interest because it can be exploited in a branch and bound scheme for solving the corresponding integer problem (1) (see e.g. [16, 17]).

Problem (SDP) may be solved in principle by any interior point method. However, this approach becomes impractical when the size of the combinatorial problem becomes larger than few thousand variables. For this reason, the special structure of the constraints of problem (SDP) has been exploited in the literature to define ad-hoc algorithms based on nonlinear programming reformulations. The first idea goes back to Homer and Peinado [12], where the change of variables $X_{ij} = v_i^T v_j / \|v_i\| \|v_j\|$ for the elements of $X$ enabled to formulate (SDP) as the unconstrained optimization problem

$$\min_{v \in \mathbb{R}^{n^2}} \ f_n(v) = \sum_{i=1}^n \sum_{j=1}^n q_{ij} \frac{v_i^T v_j}{\|v_i\| \|v_j\|} \tag{$\mathrm{Q_n}$}$$

in $n^2$ new variables $v_i \in \mathbb{R}^n$, $i = 1, \ldots, n$. In order to tackle the large dimension of the resulting problem a parallel gradient method was proposed. Later Burer and Monteiro in [4, 5] recast a general linear semidefinite programming problem as a low rank semidefinite programming problem (LRSDP) by applying the change of variables $X = VV^T$, where $V$ is an $n \times r$, $r < n$, rectangular matrix. This new formulation leads to a nonlinear optimization problem with dimension $nr$, which is solved in [4, 5] by means of an Augmented Lagrangian approach. In the computational result section of [4], in connection with the special problem (SDP), Burer and Monteiro resumed the unconstrained formulation proposed by by Homer and Peinado and mixed it with the low rank idea by introducing the change of variables $X_{ij} = v_i^T v_j / \|v_i\| \|v_j\|$ where $v_i \in \mathbb{R}^r$, $i = 1, \ldots, n$, with $r << n$. The resulting algorithm SDPLR-MC was computationally efficient, but the theory was not deeply investigated.

In [9] a specialized approach was proposed for solving problem (LRSDP): it was reformulated as the unconstrained minimization of an exact penalty function and a globally convergent algorithm was defined. Furthermore, the exactness of the merit function implied that a single minimization for a fixed positive value of a penalty parameter was enough to provide a stationary point of the LRSDP problem. Computational experiments in [9] showed that this unconstrained approach compares favorably with the best codes available in literature.

In this paper, we use the change of variables adopted in [4] for problem (SDP) to get a different unconstrained formulation, for which we prove equivalence with problem (SDP). The specific feature of this formulation is that we add to the function $f_r(v)$, where $v \in \mathbb{R}^{nr}$ as in [4], a shifted barrier penalty term that ensures compactness of the level sets of the new merit function. This allows us to use standard unconstrained optimization algorithms. In particular, we define a globally convergent algorithm based on the nonmonotone Barzilai-Borwein gradient method proposed in [11]. Numerical results show that the proposed approach outperforms the best existing methods for solving problem (SDP).

The paper is structured as follows: in Section 2, we report some useful results about the low rank reformulation of problem (SDP). In Section 3, we define the new unconstrained reformulation of problem (LRSDP), while in section 4 we define the solution algorithm employed for solving this formulation. In Section 5 we define the solution scheme for (SDP) and, finally, in Section 6 we report the numerical results.

Throughout the paper, given a matrix $M$ we denote by $\mathrm{diag}(M)$ the vector containing its diagonal and by $\mathrm{vec}(M)$ the vector obtained columnwise by the matrix $M$. Given a vector $v$, we denote by $\mathrm{Diag}(v)$ the diagonal matrix having as diagonal the vector $v$ and by $B_\rho(v)$ the closed ball centered in $v$ with radius $\rho > 0$, namely $B_\rho(v) = \{y \in \Re^m : \|y - v\| \le \rho\}$. For a given scalar $x$ we denote by $(x)_+$ the maximum between $x$ and zero, namely $(x)_+ \equiv \max(x, 0)$.

## 2 Some useful results about the low rank SDP formulation

Using the Gramian representation, any given matrix $X \succeq 0$ with rank $r$ can be written as $X = VV^T$, where $V$ is a $n \times r$ real matrix. Therefore the positive semidefiniteness constraint can be eliminated, and problem (SDP) reduces to

$$\min_V \left\{ Q \bullet VV^T : \ \mathrm{diag}(VV^T) = e \right\}. \tag{2}$$

A global minimum point of problem (2) is a solution of problem (SDP) provided that

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

where $\mathcal{X}^*_{\mathrm{SDP}}$ denotes the optimal solution set of problem (SDP). Although the value of $r_{\min}$ is not known, an upper bound can easily be computed by exploiting the result proved in [1, 15], that gives

$$r_{\min} \le \widehat{r} = \max\{k \in N : k(k+1)/2 \le n\}. \tag{3}$$

Thus, to get equivalence with (SDP), the dimension of the matrix $V$ in problem (2) can be fixed to $n \times r$ with $r \geq \widehat{r}$. For a fixed $r$, problem (2) can be written as

$$\min_{v} \left\{ q_r(v) = \sum_{i=1}^{n} \sum_{j=1}^{n} q_{ij} v_i^T v_j : \; \|v_i\|^2 = 1, \; i = 1, \ldots, n \right\} \qquad (\mathrm{NLP_r}),$$

where $v_i$, $i = 1, \ldots, n$, are the columns of the matrix $V^T$ and $v = \mathrm{vec}(V^T) \in \mathrm{I\!R}^{nr}$. We denote with $\mathcal{F}$ the feasible set of problem (NLP$_\mathrm{r}$), namely

$$\mathcal{F} = \{ v \in \Re^{nr} : \; \|v_i\|^2 = 1, \; i = 1, \ldots, n \}.$$

We say that a point $v^* \in \mathrm{I\!R}^{nr}$ *solves problem* (SDP) if $X^* = V^* V^{*T}$ is an optimal solution of problem (SDP). This implies, by definition, that $r \geq r_{\min}$.

Although reformulation (2) results in the non convex problem (NLP$_\mathrm{r}$), the primal-dual optimality condition for (SDP) combined with necessary optimality conditions for (NLP$_\mathrm{r}$) lead to some global optimality conditions [4, 9] that can be exploited from the computational point of view.

The standard first order necessary optimality condition for problem (NLP$_\mathrm{r}$) states that given a local minimizer $\hat{v} \in \mathrm{I\!R}^{nr}$ of problem (NLP$_\mathrm{r}$), there exists a unique $\hat{\lambda} \in \mathrm{I\!R}^n$ such that $(\hat{v}, \hat{\lambda})$ satisfies

$$\begin{array}{ll} \sum_{j=1}^{n} q_{ij} \hat{v}_j + \hat{\lambda}_i \hat{v}_i = 0, & \quad i = 1, \ldots, n \\ \|\hat{v}_i\|^2 = 1, & \quad i = 1, \ldots, n \end{array} \tag{4}$$

We define stationary point of problem (NLP$_\mathrm{r}$) a point $\hat{v} \in \mathrm{I\!R}^{nr}$ satisfying (4) with a suitable multiplier $\hat{\lambda} \in \mathrm{I\!R}^n$.

We note that, given a pair $(\hat{v}, \hat{\lambda})$ satisfying the conditions (4), the multiplier $\hat{\lambda}$ can be expressed uniquely as a function of $\hat{v}$ (see [9]), namely

$$\hat{\lambda}_i = \lambda_i(\hat{v}) = -\hat{v}_i^T \sum_{j=1}^{n} q_{ij} \hat{v}_j \,, \quad i = 1, \ldots, n. \tag{5}$$

By substituting the expression of $\hat{\lambda}$ in the first condition of (4), we get

$$\sum_{j=1}^{n} q_{ij} \left( I_r - \hat{v}_i \hat{v}_i^T \right) \hat{v}_j = 0 \quad i = 1, \ldots, n. \tag{6}$$

Next proposition, that extends the sufficient conditions given in [4], states the global optimality conditions obtained by exploiting the primal-dual properties for problem (SDP).

**Proposition 2.1 (Global optimality conditions)** *A point* $v^* \in \mathrm{I\!R}^{nr}$ *is a global minimizer of problem* (NLP$_\mathrm{r}$) *that solves problem* (SDP) *if and only if it is a stationary point of problem* (NLP$_\mathrm{r}$) *and satisfies*

$$Q + \mathrm{Diag}(\lambda(v^*)) \succeq 0,$$

*where* $\lambda(v^*)$ *is computed according to* (5)

The proof of this result can be found in [9] and for more general problems in [10, 13]. According to the above proposition, given a stationary point $\hat{v}$, we can prove its optimality just checking that a certain matrix is positive semidefinite. Furthermore in [12] it has been proved that, if $r = n$, there is no local minimum point of problem ($\mathrm{NLP_r}$), which is not global.

# 3 A new unconstrained formulation of problem (SDP)

We consider the unconstrained problem

$$\min_{v \in \mathbb{R}^{nr}} \ f_r(v) \equiv \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij} \frac{v_i^T v_j}{\|v_i\|\|v_j\|}, \tag{$\mathrm{Q_r}$}$$

that has been used to obtain a solution of Problem ($\mathrm{NLP_r}$) in [4]. We note that problem ($\mathrm{Q_r}$) presents some peculiarities that make standard convergence results not immediately applicable. Indeed, standard unconstrained algorithms can be proved to be globally convergent if the objective function is continuously differentiable and has compact level sets. Function $f_r(v)$ is not even defined at points where $\|v_i\| = 0$ for at least one index $i$. In principle, it is possible to modify standard algorithms by looking not at the sequence $\{(v_1, \dots, v_n)^k\}$ but at the normalized sequence $\{(v_1/\|v_1\|, \dots, v_n/\|v_n\|)^k\}$. However, this may cause difficulties in the use of many optimization algorithms.

In this paper, we propose to modify $f_r$ in such a way to get an unconstrained problem that can be solved by standard methods. In particular, we add the term

$$\sum_{i=1}^{n} \frac{(\|v_i\|^2 - 1)^2}{d(v_i)}, \tag{7}$$

where

$$d(v_i) \equiv \delta^2 - \Big(1 - \|v_i\|^2\Big)_+^2, \quad 0 < \delta < 1. \tag{8}$$

Therefore, the function we propose is

$$f_\varepsilon(v) \equiv f_r(v) + \frac{1}{\varepsilon}\sum_{i=1}^{n} \frac{(\|v_i\|^2 - 1)^2}{d(v_i)}, \tag{9}$$

where $\varepsilon > 0$. For a fixed $\varepsilon > 0$ we consider the unconstrained minimization problem

$$\min_{v \in S_\delta} \ f_\varepsilon(v), \tag{$\mathrm{RQ_r}$}$$

where the open set $S_\delta$ is defined as

$$S_\delta \equiv \{v \in \mathbb{R}^{nr} \ : \ \|v_i\|^2 > 1 - \delta, \quad i = 1, \dots, n\}.$$

The added term (7) ensures that the level sets of $f_\varepsilon$ are contained in the set $S_\delta$ and are compact. Hence, Problem ($\mathrm{RQ}_r$) allows us to overcome all the theoretical drawbacks of Problem ($\mathrm{Q}_r$). In particular, we will show that solving problem ($\mathrm{RQ}_r$) for a single value of $\varepsilon$ is equivalent to solve problem ($\mathrm{NLP_r}$).

We start by investigating the theoretical properties of function $f_\varepsilon(v)$. The gradient of function $f_\varepsilon(v)$ in the set $S_\delta$ is

$$\nabla_{v_i} f_\varepsilon(v) = \nabla_{v_i} f_r(v) + \frac{4}{\varepsilon}\frac{(\|v_i\|^2-1)}{d(v_i)}\left[1 - \frac{(\|v_i\|^2-1)(1-\|v_i\|^2)_+}{d(v_i)}\right] v_i$$

where

$$\nabla_{v_i} f_r(v) = \frac{2}{\|v_i\|}\left[\sum_{j=1}^n q_{ij}\left(I_r - \frac{v_i}{\|v_i\|}\frac{{v_i}^T}{\|v_i\|}\right)\frac{v_j}{\|v_j\|}\right]$$

The first important property is the compactness of the level sets of function $f_\varepsilon(v)$, that guarantees the existence of a solution of problem ($\mathrm{RQ}_r$).

**Proposition 3.1** *For every $v \in S_\delta$ and for every given $\varepsilon > 0$, the following condition holds*

$$f_\varepsilon(v) \geq -C + \frac{1}{\varepsilon}\frac{(\|v_i\|^2-1)^2}{\delta^2}, \quad \forall\ i = 1,\dots,n, \tag{10}$$

*where $C = \sum_{i=1}^n\sum_{j=1}^n |q_{ij}|$. Furthermore, for every given $\varepsilon > 0$ and for every given $v^0 \in S_\delta$, the level sets*

$$\mathcal{L}_\varepsilon(v^0) = \{v \in S_\delta :\ f_\varepsilon(v) \leq f_\varepsilon(v^0)\}$$

*of function $f_\varepsilon(v)$ are compact.*

**Proof** First, for every $v$, we have that

$$\begin{aligned} f_r(v) &= \sum_{i=1}^n\sum_{j=1}^n q_{ij}\frac{v_i^T v_j}{\|v_i\|\|v_j\|} \geq -\sum_{i=1}^n\sum_{j=1}^n |q_{ij}|\frac{|v_i^T v_j|}{\|v_i\|\|v_j\|} \\ &\geq -\sum_{i=1}^n\sum_{j=1}^n |q_{ij}|\frac{\|v_i\|\|v_j\|}{\|v_i\|\|v_j\|} = -C. \end{aligned}$$

Hence, (10) follows from simple majorizations. Now, we prove boundedness of $\mathcal{L}_\varepsilon(v^0)$. Let $\{v^k\} \in \mathcal{L}_\varepsilon(v^0)$ be a sequence of points such that $\|v^k\| \to \infty$. Assume without loss of generality that $\|v_1^k\| \to \infty$. By using (10), we can write:

$$f_\varepsilon(v^k) \geq -C + \frac{1}{\varepsilon}\frac{(\|v_1^k\|^2-1)^2}{\delta^2},$$

so that $f_\varepsilon(v)$ is coercive and the level set is bounded. On the other hand, any limit point of a sequence cannot belong to the boundary of $S_\delta$. Indeed, if $\|\hat{v}_i\|^2 = 1-\delta$ for some $i$, then (8) implies $d(\hat{v}_i) = 0$, and hence

$$\lim_{k\to\infty} f_\varepsilon(v^k) = \infty,$$

but this contradicts $v^k \in \mathcal{L}_\varepsilon(v^0)$ for $k$ sufficiently large. Therefore the level set $\mathcal{L}_\varepsilon(v^0)$ is also closed, and the thesis follows. $\square$

Next proposition gives a bound on the value of $\|v_i\|$ for all $i = 1, \ldots, n$ in the level set.

**Proposition 3.2** *Let $\varepsilon > 0$ and $v^0 \in \mathcal{F}$. Then, we have*

$$\mathcal{L}_\varepsilon(v^0) \subseteq \left\{ v \in \Re^{nr} : \|v_i\|^2 \leq (2C\varepsilon\delta^2)^{\frac{1}{2}} + 1, \quad i = 1, \ldots, n \right\}.$$

**Proof**
For any given $v \in \mathcal{L}_\varepsilon(v^0)$, because $v^0 \in \mathcal{F}$, we can write

$$f_\varepsilon(v) \leq f_\varepsilon(v^0) = f_r(v^0) \leq C,$$

where $C$ is defined in Proposition 3.1. Moreover, using (10), we have

$$f_\varepsilon(v) \geq -C + \frac{1}{\varepsilon} \frac{(\|v_j\|^2 - 1)^2}{\delta^2}, \quad j = 1, \ldots, n,$$

so that

$$\|v_j\|^2 \leq (2C\varepsilon\delta^2)^{\frac{1}{2}} + 1, \quad j = 1, \ldots, n.$$

$\square$

An interesting property of the objective function $f_r(v)$ of problem ($\mathrm{Q_r}$) is that, given a point $v$ in $S_\delta$, its gradient with respect to $v_i$ is orthogonal to the vector $v_i$, namely, for every $v \in S_\delta$ and for every $i = 1, \ldots, n$

$$v_i^T \nabla_{v_i} f_r(v) = 2 \left[ \sum_{j=1}^{n} q_{ij} \left( \frac{v_i^T}{\|v_i\|} - \frac{v_i^T v_i}{\|v_i\|^2} \frac{{v_i}^T}{\|v_i\|} \right) \frac{v_j}{\|v_j\|} \right] = 0. \tag{11}$$

The following theorem states the equivalence between stationary points, local/global minimizers of ($\mathrm{RQ}_r$) and the corresponding stationary points, local/global minimizers of ($\mathrm{NLP_r}$).

**Theorem 3.3 (Exactness properties of** ($\mathrm{RQ}_r$)**)** *For any $\varepsilon > 0$ the following correspondences hold:*

(i) *a point $\hat{v} \in \mathbb{R}^{nr}$ is a stationary point of Problem* ($\mathrm{RQ}_r$) *if and only if it is a stationary point of problem* ($\mathrm{NLP_r}$)*.*

(ii) *a point $\hat{v} \in \mathbb{R}^{nr}$ is a global minimizer of problem* ($\mathrm{RQ}_r$) *if and only if it is a global minimizer of problem* ($\mathrm{NLP_r}$)*.*

(iii) *a point $\hat{v} \in \mathbb{R}^{nr}$ is a local minimizer of problem* ($\mathrm{RQ_r}$) *if and only if it is a local minimizer of problem* ($\mathrm{NLP_r}$)*.*

**Proof** First, we recall that, for every $v \in S_\delta$, $v_i \neq 0$ for all $i = 1, \ldots, n$. Furthermore, by definition of $\nabla_{v_i} f_\varepsilon$ and by (11), we get for every $v_i$ and for $i = 1, \ldots, n$

$$v_i^T \nabla_{v_i} f_\varepsilon(v) = \frac{4}{\varepsilon} \frac{(\|v_i\|^2 - 1) v_i^T v_i}{d(v_i)} \left( 1 - \frac{(\|v_i\|^2 - 1)(1 - \|v_i\|^2)_+}{d(v_i)} \right).$$

Therefore we get, if $\|v_i\|^2 \geq 1$,

$$v_i^T \nabla_{v_i} f_\varepsilon(v) = \frac{4}{\varepsilon} \frac{(\|v_i\|^2 - 1)\|v_i\|^2}{\delta^2}, \tag{12}$$

otherwise

$$v_i^T \nabla_{v_i} f_\varepsilon(v) = \frac{4}{\varepsilon} \frac{(\|v_i\|^2 - 1)\|v_i\|^2}{d(v_i)} \left(1 + \frac{(\|v_i\|^2 - 1)^2}{d(v_i)}\right). \tag{13}$$

Further, if $v \in \mathcal{F}$

$$f_\varepsilon(v) = f_r(v) = q_r(v) \tag{14}$$

$$\nabla_{v_i} f_\varepsilon(v) = 2 \sum_{j=1}^{n} q_{ij}(I_r - v_i {v_i}^T)v_j, \quad i = 1, \ldots, n. \tag{15}$$

Now we can prove the three statements.
**(i)** *Sufficiency.* Let $\hat{v}$ be a stationary point for problem (NLP$_\mathrm{r}$). Therefore $\hat{v}$ satisfies (6) and $\hat{v} \in \mathcal{F}$. Then (15) implies

$$\nabla_{v_i} f_\varepsilon(\hat{v}) = 2 \sum_{j=1}^{n} q_{ij}(I_r - \hat{v}_i \hat{v}_i^T)\hat{v}_j = 0, \quad i = 1, \ldots, n.$$

*Necessity.* By (12) and (13), $\hat{v} \in S_\delta$ being a stationary point of $f_\varepsilon$ implies $\hat{v} \in \mathcal{F}$. Hence, as a result of (15), $\hat{v}$ is stationary point also for problem (NLP$_\mathrm{r}$).

**(ii)** *Necessity.* By Proposition 3.1, the function $f_\varepsilon$ admits a global minimizer $\hat{v}$ , which is obviously a stationary point of $f_\varepsilon$ and hence we have that $\hat{v} \in \mathcal{F}$, so that $f_\varepsilon(\hat{v}) = q_r(\hat{v})$. We proceed by contradiction. Assume that a global minimizer $\hat{v}$ of $f_\varepsilon$ is not a global minimizer of problem (NLP$_\mathrm{r}$). Then there exists a point $v^* \in \mathcal{F}$, global minimizer of problem (NLP$_\mathrm{r}$), such that

$$f_\varepsilon(\hat{v}) = q_r(\hat{v}) > q_r(v^*) = f_\varepsilon(v^*),$$

but this contradicts the assumption that $\hat{v}$ is a global minimizer of $f_\varepsilon$.
*Sufficiency.* True by similar arguments.

**(iii)** *Necessity.* Since $\hat{v}$ is a local minimizer of $f_\varepsilon$, it is a stationary point of $f_\varepsilon$, so that $\hat{v} \in \mathcal{F}$. Thus, $f_\varepsilon(\hat{v}) = q_r(\hat{v})$. Since $\hat{v}$ is a local minimizer of $f_\varepsilon$, there exists a $\rho > 0$ such that for all $v \in S_\delta \cap B_\rho(\hat{v})$ such that

$$q_r(\hat{v}) = f_\varepsilon(\hat{v}) \leq f_\varepsilon(v).$$

Therefore, by using (14), for all $v \in v \in \mathcal{F} \cap B_\rho(\hat{v})$ we have that

$$q_r(\hat{v}) \leq f_\varepsilon(v) = q_r(v).$$

and hence $\hat{v}$ is a local minimizer for problem (NLP$_\mathrm{r}$).

*Sufficiency.* Since $\hat{v} \in \mathcal{F}$ and is a local minimizer of $(\mathrm{NLP_r})$, there exists a $\rho > 0$ such that for all $v \in \mathcal{F} \cap B_\rho(\hat{v})$

$$q_r(\hat{v}) = f_\varepsilon(\hat{v}) \leq q_r(v) = f_\varepsilon(v).$$

We want to show that there exists $\gamma$ such that for all $v \in S_\delta \cap B_\gamma(\hat{v})$ we get

$$f_\varepsilon(\hat{v}) \leq f_\varepsilon(v).$$

It is sufficient to show that there is a $\gamma > 0$ such that for all $v \in S_\delta \cap B_\gamma(\hat{v})$, we have that $p(v) \in B_\gamma(\hat{v})$, where

$$p(v) \equiv \begin{pmatrix} \dfrac{v_1}{\|v_1\|} \\ \vdots \\ \dfrac{v_n}{\|v_n\|} \end{pmatrix}.$$

Actually in this case we have

$$q_r(\hat{v}) = f_\varepsilon(\hat{v}) \leq q_r(p(v)) = f_\varepsilon(p(v)) \leq f_\varepsilon(v).$$

It is well known that, given any point $x \neq 0 \in \mathbb{R}^n$, its projection over the unit norm set is simply $\dfrac{x}{\|x\|}$. Hence, for any $\gamma \leq \frac{\rho}{2}$ we can write

$$\begin{aligned}
\|p(v) - \hat{v}\|^2 &= \sum_{i=1}^n \|\hat{v}_i - \frac{v_i}{\|v_i\|}\|^2 = \sum_{i=1}^n \|\hat{v}_i - \frac{v_i}{\|v_i\|} + v_i - v_i\|^2 \\
&\leq \sum_{i=1}^n \left( \|\hat{v}_i - v_i\|^2 + \|v_i - \frac{v_i}{\|v_i\|}\|^2 + 2\|\hat{v}_i - v_i\|\|v_i - \frac{v_i}{\|v_i\|}\| \right) \\
&\leq \sum_{i=1}^n 4\|\hat{v}_i - v_i\|^2 = 4\|\hat{v} - v\|^2 \leq 4\gamma^2 \leq \rho^2
\end{aligned}$$

Therefore, for a proper $\gamma$, we have for all $v \in S_\delta \cap B_\gamma(\hat{v})$

$$f_\varepsilon(\hat{v}) \leq f_\varepsilon(v),$$

so that $\hat{v}$ is a local minimum also for $(\mathrm{RQ}_r)$. □

Theorem 3.3 states a tight relation between Problem $(\mathrm{RQ}_r)$ and $(\mathrm{NLP_r})$ and hence allows us to solve problem $(\mathrm{NLP_r})$ by minimizing $f_\varepsilon(v)$. We stress that all the properties of problem $(\mathrm{RQ}_r)$ hold for any given $\varepsilon > 0$.

Proposition 3.3 implies that we can solve problem $(\mathrm{NLP_r})$ by solving problem $(\mathrm{RQ_r})$, and Proposition 3.1 implies that any standard minimization method can be used for solving it. To be more precise, we assume that an unconstrained minimization procedure UNC satisfying the following property is available.

**Property A** *Given a continuously differentiable function with compact level sets, starting from any initial point, procedure* UNC *produces a sequence of points belonging to the initial level set such that:*

(i) *it admits at least an accumulation point,*

(ii) *every accumulation point is a stationary point of the objective function.*

In the literature there are many unconstrained minimization methods satisfying Property A, see for example [3]. We note that the function $f_\varepsilon(v)$ is continuously differentiable over the set $S_\delta$ and, by Proposition 3.1, it has compact level sets. Hence we can apply procedure UNC to find a stationary point of $f_\varepsilon$. We can easily state the following convergence result that does not require any further assumption.

**Proposition 3.4** *Let $r$ be given and $v^0 \in \mathcal{F}$. The Procedure* UNC *applied to the merit function $f_\varepsilon$ starting from $v^0$ produces a sequence $\{v^k\} \in \mathbb{R}^{nr}$ such that*

(i) *$\{v^k\}$ is bounded and it admits at least an accumulation point;*

(ii) *every accumulation point is a stationary point of problem* (NLP$_\mathrm{r}$)*;*

(iii) *if $\hat{v}$ is an accumulation point then $q_r(\hat{v}) \le q_r(v^0)$.*

**Proof** Function $f_\varepsilon(v)$ is continuously differentiable over the set $S_\delta$ and Proposition 3.1 implies that it has compact level sets. Therefore Property A implies that UNC produces a sequence that has at least an accumulation point and all the accumulation points are stationary points of problem (RQ$_\mathrm{r}$). Finally Theorem 3.3 implies that the stationary points of $f_\varepsilon$ are stationary points of problem (NLP$_\mathrm{r}$), and we have

$$q_r(\hat{v}) = f_\varepsilon(\hat{v}) \le f_\varepsilon(v^0) = q_r(v^0).$$

□

# 4 A Gradient based method for solving problem (RQ$_\mathrm{r}$)

The barrier term plays a key role to make standard optimization methods be globally convergent for problem (RQ$_r$). Nevertheless, generally, a barrier term affects negatively the performance behavior of any optimization method, especially when the produced sequence gets closer to the boundary of $S_\delta$.

Starting with $v^0 \in \mathcal{F}$, we define an iteration of the form

$$v_i^{k+1} = v_i^k - \alpha^k \nabla_{v_i} f_\varepsilon(v^k) \quad i = 1, \dots, n, \tag{16}$$

where $\alpha^k > 0$ is obtained by a suitable linesearch procedure satisfying

$$f_\varepsilon(v^{k+1}) \le f_\varepsilon(v^0), \tag{17}$$

that is a mild standard assumption. We prove that for $\varepsilon$ sufficiently large, the produced sequence stays in the set $\{v \in \mathbb{R}^{nr} : \|v_i\|^2 \ge 1,\, i = 1, \dots, n\}$. This result implies that the term (7) reduces simply to a penalty term on the feasibility of problem (NLP$_\mathrm{r}$). In particular, the following proposition holds.

**Proposition 4.1** *Let $v^0 \in \mathcal{F}$ and let $\{v^k\}$ be the sequence generated with the iterative scheme* (16)*, where each $\alpha^k$ satisfies* (17) *and $\alpha^k \le \alpha_M$. Then, there exists $\bar{\varepsilon} > 0$ such that, for any $\varepsilon \ge \bar{\varepsilon}$, we have for $k = 1, 2, \ldots$*

$$\|v_i^k\| \ge 1, \quad i = 1, \ldots, n.$$

**Proof** By (17), for a fixed value $\varepsilon > 0$ the sequence $\{v^k\}$ stays in the compact level set $\mathcal{L}_\varepsilon(v^0)$. The proof is by induction. Assume that there exists $\bar{\varepsilon} > 0$ such that, for any $\varepsilon \ge \bar{\varepsilon}$, it is true that $\|v_i^k\|^2 \ge 1$. We show that is true also for $k+1$. We can write

$$\begin{aligned}
\|v_i^{k+1}\|^2 &= \|v_i^k\|^2 + (\alpha^k)^2 \|\nabla_{v_i} f_\varepsilon(v^k)\|^2 - 2\alpha^k (v_i^k)^T \nabla_{v_i} f_\varepsilon(v^k) \\
&= \|v_i^k\|^2 + (\alpha^k)^2 \|\nabla_{v_i} f_\varepsilon(v^k)\|^2 - \frac{8\alpha^k}{\varepsilon} \frac{(\|v_i^k\|^2 - 1)\|v_i^k\|^2}{\delta^2} \\
&\ge \|v_i^k\|^2 - \frac{8\alpha_M}{\varepsilon\delta^2} (\|v_i^k\|^2 - 1)\|v_i^k\|^2,
\end{aligned}$$

where the second equality derives from (12), keeping in mind that $\|v_i^k\| \ge 1$. If $\|v_i^k\| = 1$, then $\|v_i^{k+1}\|^2 \ge 1$. Otherwise, if $\|v_i^k\| > 1$, we need to verify that a value of $\bar{\varepsilon}$ exists such that for all $\varepsilon \ge \bar{\varepsilon}$

$$(\|v_i^k\|^2 - 1) - \frac{8\alpha_M}{\varepsilon\delta^2} (\|v_i^k\|^2 - 1)\|v_i^k\|^2 \ge 0,$$

namely

$$1 - \frac{8\alpha_M}{\varepsilon\delta^2} \|v_i^k\|^2 \ge 0. \tag{18}$$

By Proposition 3.2, we have that for all $k$

$$\|v_i^k\|^2 \le (2C\varepsilon\delta^2)^{\frac{1}{2}} + 1 \quad i = 1, \ldots, n. \tag{19}$$

Therefore (19) combined with (18) implies

$$\varepsilon - 8\frac{\alpha_M}{\delta^2} \left( (2C\delta\varepsilon)^{\frac{1}{2}} + 1 \right) \ge 0$$

which is satisfied for some $\varepsilon \ge \bar{\varepsilon}$. □

# 5 A globally convergent algorithm for solving problem (SDP)

In this section, we finally define an algorithm for solving problem (SDP) that makes use of the results stated in the previous sections.

In Section 2 we have seen that for $r \ge r_{\min}$ a global solution of problem ($\mathrm{NLP_r}$) provides a solution of problem (SDP). Moreover, Proposition 3.3 states a complete correspondence between problems ($\mathrm{NLP_r}$) and ($\mathrm{RQ_r}$). Finally, Proposition 3.4 ensures that we can find a stationary point of problem ($\mathrm{RQ_r}$) by applying any unconstrained minimization procedure satisfying Property A.

The value of $r_{\min}$ is not known. In principle, the only value of $r$ that can be calculated and that guarantees the correspondence between solutions of (SDP) and global solutions of ($\mathrm{NLP_r}$), is $\widehat{r}$ as defined in (3). However, this value is usually larger than the actual value needed to obtain a solution of problem (SDP). Hence, following the idea in [4] and [9], we choose $r << \widehat{r}$, and use the global optimality condition of Proposition 2.1 to prove optimality. Our algorithmic scheme is the same as in algorithm EXPA (see [9]) and is reported below.

**Regularized Quotient Algorithm** (`ReQuA`)

**Data.** $Q \in \mathcal{S}^n$.

**Inizialization.** Find the value $\widehat{r}$ given by (3). Set integers $2 \leq r^1 < r^2 < \ldots < r^p$ with $r^p \in [\widehat{r}, n]$. Choose $\varepsilon \geq \bar{\varepsilon}$.

**For** $j = 1, \ldots, p$

Find a stationary point $\hat{v} \in \mathbb{R}^{nr^j}$ of problem ($\mathrm{NLP_{r^j}}$) by using the equivalent formulation ($\mathrm{RQ}_{r^j}$) and Procedure UNC starting from a point $v^0 \in \mathbb{R}^{nr^j}$ feasible for problem ($\mathrm{NLP_{r^j}}$).

Compute the minimum eigenvalue $\mu_{\min}(\hat{v})$ of $Q + \mathrm{Diag}(\lambda(\hat{v}))$. If $\mu_{\min}(\hat{v}) = 0$, then **exit**.

**End**

**Return** $\hat{v} \in \mathbb{R}^{nr^j}$ and $\mu_{\min}(\hat{v})$

`ReQuA` returns $\hat{v} \in \mathbb{R}^{nr^j}$, and $\mu_{\min}(\hat{v})$. If $\mu_{\min}(\hat{v}) = 0$, then a solution for (SDP) is obtained as $X^* = \widehat{V}\widehat{V}^T$. If the optimality condition is not met, a bound can be easily computed on the optimal value of (SDP), that can be used in a branch and bound scheme. Indeed, the value $n\mu_{\min}\left[Q + \mathrm{Diag}(\lambda(\hat{v}))\right] + q_r(\hat{v})$, provides a lower bound on the solution of problem (SDP)(see e.g, [9]).

In practice, however, in all the computational experiments performed the stopping condition $\mu_{\min}(\bar{v}) = 0$ was always met with satisfactory accuracy, so that `ReQuA` always converged to a solution of (SDP), as we will illustrate in the next section.

# 6 Numerical Results

In this section, we describe our computational experience with algorithm `ReQuA`.

`ReQuA` is implemented in Fortran 90 and all the experiments have been run on a PC with processor Core2 DUO E6750 2.66Ghz, and RAM of 2.00 GB.

As unconstrained optimization procedure a Fortran 90 implementation of the non monotone Barzilai-Borwein gradient method proposed in [11] is used. This method satisfies Property A stated in the previous section.

The stopping criterion in the minimization procedure of $f_\varepsilon$ is

$$\|\nabla f_\varepsilon(v^k)\| \le tol \cdot (1 + |f_\varepsilon(v^k)|)$$

or

$$\|\nabla f_\varepsilon(v^k)\| \le 10 \cdot tol \cdot (1 + |f_\varepsilon(v^k)|) \;\text{ and }\; |f_\varepsilon(v^k) - f_\varepsilon{}^{best}| \le 10^{-4} \cdot tol \cdot (1 + |f_\varepsilon(v^k)|),$$

where $f_\varepsilon^{best}$ is the best value of $f_\varepsilon$ computed during the past iterations and $tol$ will be fixed depending on the desired accuracy, keeping in mind that the default value is $10^{-5}$.

The parameter $\varepsilon$ is set equal to $10^3$ for all the tests. This value has been chosen after some experiments for different values of $\varepsilon$.

As for the choice of the starting value $r^1$ of the rank of the solution we use the same values as in [9], reported in Table 1. We remark that on all the considered test problems

| n | $\le 200$ | $> 200, \; < 800$ | $\ge 800, < 1000$ | $\ge 1000, < 5000$ | $\ge 5000$ |
|---|---|---|---|---|---|
| $r^1$ | 8 | 10 | 15 | 18 | 20 |

Table 1: Values of $r^1$ depending on the dimension of the problem.

the output matrix $Q + \mathrm{Diag}(\lambda(\hat{v}))$ turned out to be positive semidefinite for $j = 1$ , so that the obtained point $\hat{v} \in \mathbb{R}^{nr^j}$ is actually a global solution of (SDP) (similar behavior was encountered in [9] for algorithm EXPA). In order to check positive semidefiniteness of $Q + \mathrm{Diag}(\lambda(\hat{v}))$, we use the ARPACK subroutines dsaupd and dseupd to compute the minimum eigenvalue of this matrix.

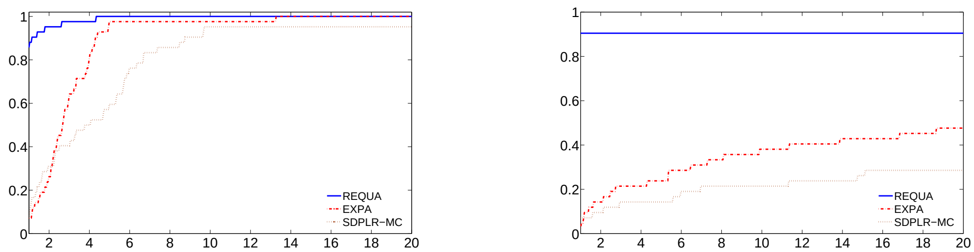
In order to evaluate the performance of our algorithm `ReQuA`, we test it on 42 instances of the Max-Cut problem. Indeed, the standard SDP relaxation of the Max-Cut problem is exactly problem (SDP), where the matrix $Q$ is equal to minus the Laplacian of the graph divided by four. The number of nodes and edges of the considered graphs range from 100 to 20000 and from 150 to 40000, respectively, with different degrees of sparsity (see [9] for more details on the test set).

We compare `ReQuA` with the best codes in literature in the two main classes of methods for solving (SDP) : interior point methods and low rank methods.

Up to our knowledge, the best low rank based methods are SDPLR−MC proposed by Burer and Monteiro in [4], which can be downloaded from the web page `http://dollar.biz.uiowa.edu/~ burer/software/SDPLR-MC`, and EXPA proposed in [9].

Both EXPA and SDPLR−MC have a structure similar to `ReQuA`. Indeed, the main scheme differs in the way of finding a stationary point for ($\mathrm{NLP_r}$). For any fixed value of $r$, EXPA uses the nonmonotone Barzilai-Borwein gradient proposed in [11] (the same one used in `ReQuA`) to minimize an exact penalty function for ($\mathrm{NLP_r}$). SDPLR−MC uses an L-BFGS method to obtain a stationary point of ($\mathrm{Q_r}$). Another difference is that SDPLR−MC does not certify global optimality of the produced solution, while EXPA (as `ReQuA`) checks the global optimality condition $Q + \mathrm{Diag}(\lambda(\hat{v})) \succeq 0$ at the end.

Among the interior point methods, we choose DSDP (version 5.8) proposed in [2] downloaded from the webpage `http://www-unix.mcs.anl.gov/DSDP/`. It is consid-




(a) time based comparison

(b) accuracy based comparison

Figure 1: Comparison among the low rank methods

ered the most efficient interior point method, especially for solving problems where the solution is known to be low rank (as for the $\{-1,1\}$ quadratic problem). DSDP has relatively low memory requirements for an interior-point method, and is indeed able to solve all our instances up to 10000 nodes. We use the feasible starting point version for all problems except for 8 graphs (mcp124-1, mcp250-1, mcp250-2, mcp500-1, mcp500-2, G55, G60 and G70) where we had to allow an infeasible starting point to get convergence.

`ReQuA`, EXPA and SDPLR$-$MC solve all the test problems, whereas DSDP runs out of memory on the two biggest problems (G77 and G81 of the Gset collection).

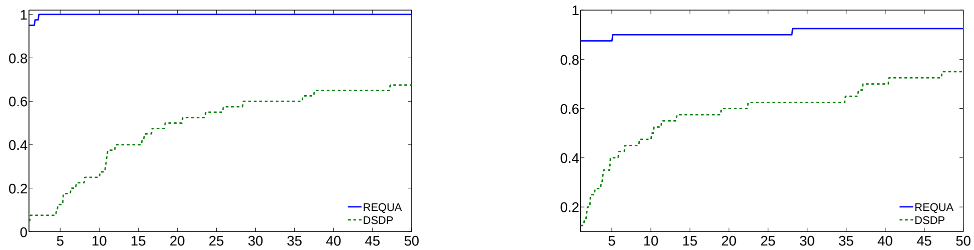
We compare the different codes on the basis of the computational time and the level of accuracy. We consider that the methods converge to the same solution whenever the relative difference between the objective function values is less than $\gamma = 10^{-4}$.

In order to have a better flavor of the results, following the approach proposed in [7], we draw the performance profile of the different methods with respect to the computational time and to the accuracy in the solution. In particular, when we want to evaluate the accuracy of the methods, we choose as a performance measure (keeping into account that all the objective values are negative)

$$|f_s^*(p) - f^*(p)| + \gamma$$

where $f^*(p)$ is the minimum objective value found by the best code on that problem, $f_s^*(p)$ is the objective value found by method s for problem p, and where $\gamma = 10^{-4}$. We recall that the higher the method in the profile, the better the performance.

In Figure 1, we report the comparison among the three low rank based methods on all the test problems, with respect to the computational time in Figure 1 (a), and with respect to the accuracy in Figure 1 (b). It emerges from the profiles that `ReQua` outperforms the other low rank methods with respect to both the performance measures. Interior point methods are well known for ensuring high levels of accuracy. For this reason, when we compare `ReQua` with DSDP, we require a higher level of accuracy, setting *tol* to $10^{-7}$. We report the obtained results in Figure 2, where we compare `ReQua` with DSDP, on all the problems except the two largest ones that DSDP could not solve. In Figure 2 (a), we compare the two methods with respect to the computational time, in Figure 2 (b) with respect to the accuracy. It emerges that in a significantly smaller




(a) time based comparison (b) accuracy based comparison

Figure 2: Comparison between `ReQuA` and DSDP

amount of time `ReQua` finds solutions that are more accurate than the ones produced by DSDP.

# 7 Concluding Remarks

In this paper, we have introduced a new merit function that allows us to recast the low rank reformulation of problem (SDP) as an unconstrained minimization problem. We have defined a globally convergent algorithm for solving problem (SDP), called `ReQuA`. An extensive numerical test showed that `ReQuA` outperforms the best available methods in terms of time, and it achieves an accuracy that favorably compares to the accuracy achieved by the interior point method DSDP. Further improvement could be achieved by defining a special unconstrained algorithm which further exploits the structure of the problem.

# References


[1] A. Barvinok. Problems of distance geometry and convex properties of quadratic maps. *Discrete Computational Geometry*, 13:189–202 (1995).

[2] S. J. Benson, Y. Ye, X. Zhang. Solving Large-Scale Sparse Semidefinite Programs for Combinatorial Optimization. *SIAM Journal on Optimization*, 10(2):443–461 (2000).

[3] Dimitri P. Bertsekas. Nonlinear Programming, Athena Scientific, 1999.

[4] S. Burer, R.D.C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, Ser. B 95:329–357 (2003).

[5] S. Burer, R.D.C. Monteiro. Local Minima and Convergence in Low-Rank Semidefinite Programming. *Mathematical Programming*, Ser. A, 103:427-444 (2005).

[6] C. Delorme, S. Poljak. Laplacian eigenvalues and the maximum cut problem. *Mathematical Programming*, 62(3):557–574 (1993).

[7] E.D. Dolan, J.J. Morè. Benchmarking optimization software with performance profile. *Mathematical Programming*, Ser. A, 91:201–213 (2002).

[8] M.X. Goemans, D.P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *J. Assoc. Comput. Mach.*, 42(6):1115–1145 (1995).

[9] L. Grippo, L. Palagi, V. Piccialli. An unconstrained minimization method for solving low rank SDP relaxations of the max cut problem. *Mathematical Programming*, online first, DOI: 10.1007/s10107-009-0275-8 (2009).

[10] L. Grippo, L. Palagi, V. Piccialli. Necessary and sufficient global optimality conditions for NLP reformulations of linear SDP problems. *Journal of Global Optimization*, online first, DOI: 10.1007/s10898-008-9328-4.

[11] L. Grippo, M. Sciandrone. Nonmonotone globalization techniques for the Barzilai-Borwein gradient method. *Computational Optimization and Applications*, 23:143–169 (2002).

[12] S. Homer, M. Peinado. Desing and performance of parallel and Distributed Approximation Algortihm for the maxcut. *J. of Parallel and Distributed Computing*, 46:48–61 (1997).

[13] M. Journée, F. Bach, P.A. Absil, R. Sepulchre. Low-rank optimization for semidefinite convex problems. Submitted to *Optimization and Control*, cite as arXiv:0807.4423v1, (2008).

[14] M. Laurent, S. Poljak, F. Rendl. Connections between semidefinite relaxations of the max-cut and stable set problems. *Mathematical Programming*, 77:225–246 (1997).

[15] G. Pataki. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Mathematics of Operations Research*, 23:339358 (1998).

[16] S. Poljak, F. Rendl. Solving the max-cut problem using eigenvalues. *Discrete Applied Mathematics*, 62(1-3):249–278 (1995).

[17] F. Rendl, G. Rinaldi, A. Wiegele. Solving Max-Cut to optimality by intersecting semidefinite and polyhedral relaxations. *Mathematical Programming*, 121(2):1436-4646 (2010).