POLITECNICO DI TORINO
Repository ISTITUZIONALE

Enhancing low-rank solutions in semidefinite relaxations of Boolean quadratic problems

Original
Enhancing low-rank solutions in semidefinite relaxations of Boolean quadratic problems / Cerone, Vito; Fosson, Sophie; REGRUTO TOMALINO, Diego. - ELETTRONICO. - vol. 53 n.2:(2020), pp. 1894-1899. (Intervento presentato al convegno 21th IFAC World Congress tenutosi a Berlin, Germany (Virtual) nel July 12-17, 2020) [10.1016/j.ifacol.2020.12.2578].

Availability:
This version is available at: 11583/2847109 since: 2021-04-25T23:54:37Z

Publisher:
IFAC

Published
DOI:10.1016/j.ifacol.2020.12.2578

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

23 April 2024

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 53-2 (2020) 1894–1899

Enhancing low-rank solutions in semidefinite relaxations of Boolean quadratic problems

Vito Cerone * **Sophie M. Fosson** * **Diego Regruto** *

* *Department of Control and Computer Engineering, Politecnico di Torino, Italy (e-mail: sophie.fosson@polito.it).*

Abstract Boolean quadratic optimization problems occur in a number of applications. Their mixed integer-continuous nature is challenging, since it is inherently NP-hard. For this motivation, semidefinite programming relaxations (SDR's) are proposed in the literature to approximate the solution, which recasts the problem into convex optimization. Nevertheless, SDR's do not guarantee the extraction of the correct binary minimizer. In this paper, we present a novel approach to enhance the binary solution recovery. The key of the proposed method is the exploitation of known information on the eigenvalues of the desired solution. As the proposed approach yields a non-convex program, we develop and analyze an iterative descent strategy, whose practical effectiveness is shown via numerical results.

Copyright © 2020 The Authors. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0)

Keywords: Mixed-integer programming, binary optimization, semidefinite programming relaxations, low-rank semidefinite programming.

1. INTRODUCTION

Boolean quadratic optimization refers to those minimization problems with quadratic cost functional and binary variables. It is a long-time problem with a number of applications in different scientific areas. To mention some examples, it is encountered in maximum cut (MAX-CUT) problems in graphs, see, Palagi et al. (2012), knapsack problems, see Kellerer et al. (2004), hybrid model predictive control, see Naik and Bemporad (2017); Bemporad and Naik (2018), sensor selection, see Shekhar et al. (2014); Carmi and Gurfil (2013), medical imaging, see Schüle et al. (2005); Weber et al. (2015), and binary compressed sensing, see Fosson and Abuabiah (2019).

Boolean quadratic optimization is challenging as it is NP-hard, even when the cost functional is convex, due to the integer nature of the variables. To overcome this drawback, different relaxations are proposed in the literature that approximate the correct solution. The most popular approach consists in the semidefinite programming relaxation (SDR), known also as Shor relaxation, introduced by Shor (1987) and Lovász and Shrijver (1990). In a nutshell, the key idea at the basis of SDR is to embed the variable matrix xx^T, $x \in \mathbb{R}^n$, into the space of $n \times n$ symmetric positive semidefinite matrices. The rationale is that a complete description of the cone of symmetric positive semidefinite matrices is available, while this does not hold for the convex polytope of matrices xx^T, as illustrated, e.g., by Nesterov (1998). The tightness of SDR is analyzed, e.g., in Goemans and Williamson (1995); Nesterov (1998).

More recently, a substantial evolution on SDR has been provided by the Lasserre's theory, which states that the global optimum of a polynomial optimization problem can be achieved by solving a hierarchy of SDR's; we refer the reader to Lasserre (2001, 2015) for details, and to Lasserre (2000) for the specific case of quadratic functionals. Theorem 4.2 in Lasserre (2001) and further elaborations in Chapter 6 of Lasserre (2015) provide sufficient conditions to get the global minimum and a global minimizer of the original problem from an SDR of a certain relaxation order. As to the Shor relaxation (or equivalently, SDR of order 1) of a Boolean quadratic problem, if the SDR solution is 1-rank, then the global minimum is achieved and the desired binary minimizer can be extracted, as discussed in Fosson and Abuabiah (2019). For this motivation, methods to minimize or reduce the rank in semidefinite programming are relevant for SDR; we refer to Lemon et al. (2016) for a complete overview on the topic. Since the rank minimization is an NP-hard problem, the minimization of the nuclear norm is often exploited, which provides the tightest convex relaxation of the problem, as proven by Recht et al. (2010). As to symmetric positive semidefinite matrices, the nuclear norm is equal to the trace, and in turn to the sum of the eigenvalues: then, the minimization of the nuclear norm has a sparsifying effect on the eigenvalues vector, as the ℓ_1-norm is a suitable convex relaxation of the sparsity a vector, see, e.g., Tibshirani (1996). Then, a nuclear norm penalty encourages a low-rank solution. Nevertheless, this method does not apply to some Boolean quadratic problems, e.g., MAX-CUT, see Lasserre (2016), where the trace is constant. For these problems, an effective method is the log-det heuristic, proposed by Fazel et al. (2003).

In this paper, we propose a novel method to enhance SDR for Boolean quadratic optimization, by supporting 1-rank solutions. The key idea is to exploit the knowledge of the eigenvalues of the desired solution are known (in particular, only one eigenvalue is non-null). Then, we promote 1-rank solutions by maximizing the energy of the eigenvalues vector, which we prove to have a sparsifying effect on the

2405-8963 Copyright © 2020 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2020.12.2578

eigenvalues, when the trace is constant. We illustrate the proposed approach for two SDR's, with slightly different spectral properties. Furthermore, as the method yields a non-linear and non-convex program, a suitable iterative descent algorithm is developed to search the global minimum. The overall strategy is analyzed, and its effectiveness is illustrated through numerical simulations.

The paper is organized as follows. In Sec. 2, we formally state the problem. In Sec. 3, we present and analyze the proposed approach. In Sec. 4, we show some numerical results; finally, we draw some conclusions in Sec. 5.

2. PROBLEM STATEMENT

Let $\mathbb{S}^n$ be the space of $n \times n$ symmetric matrices, and $\mathbb{S}^n_+ \subset \mathbb{S}_n$ the subspace of positive semidefinite matrices. We consider Boolean quadratic problems of the kind

$$\min_{x\in\{0,1\}^n} x^T Cx + 2d^T x \text{ s. t. } Ax = b \tag{1}$$

where $C \in \mathbb{S}^n$, $d \in \mathbb{R}^n$, $A \in \mathbb{R}^{m,n}$, and $b \in \mathbb{R}^m$. As in Lasserre (2016), we exploit the Lagrangian formulation

$$\min_{x\in\{0,1\}^n} x^T Cx + 2d^T x + \mu\|Ax - b\|_2^2 \tag{2}$$

where $\mu > 0$. If μ is sufficiently large, problems (1) and (2) are equivalent, see Lasserre (2016). By adding a slack variable $x_0 = 1$ and by redefining $x = (1, x_1, \dots, x_n)^T$, we rewrite (2) as the augmented problem

$$\min_{x\in\{0,1\}^{n+1}} x^T Qx \tag{3}$$

where $Q = \left(\begin{array}{c|c} 0 & d^T \\ \hline d & C \end{array}\right) + \mu(-b|A)^T(-b|A)$. The solution to this NP-hard problem can be searched by solving the associated SDR, see Nesterov (1998) for details:

$$\begin{aligned} &\min_{X\in\mathbb{S}^{n+1}_+} \langle Q, X\rangle \\ &\text{s. t. } X_{0,0} = 1; \quad X_{i,i} = X_{0,i}, \quad i = 1, \dots, n. \end{aligned} \tag{4}$$

The constraint $X_{i,i} = X_{0,i}$ represents the constraint $x_i = x_i^2$, which holds for any $x_i \in \{0, 1\}$.

An alternative, MAX-CUT approach is studied in Lasserre (2016): given $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^{n+1}$, by substituting $z = 2x - \mathbf{1} \in \{-1, 1\}^{n+1}$ in (3), we obtain $x^T Qx = \frac{1}{4}(z + \mathbf{1})^T Q(z + \mathbf{1})$. Then, problem (3) is equivalent to

$$\min_{z\in\{-1,1\}^{n+1}} z^T Rz \tag{5}$$

where $R = Q + \left(\begin{array}{c|c} 0 & \mathbf{1}^T Q \\ \hline Q\mathbf{1} & \mathbf{0} \end{array}\right)$. Problem (5) is a MAX-CUT problem, and the associated SDR is

$$\begin{aligned} &\min_{Z\in\mathbb{S}^{n+1}_+} \langle R, Z\rangle \\ &\text{s. t. } Z_{i,i} = 1, \quad i = 0, \dots, n. \end{aligned} \tag{6}$$

The constraint $Z_{i,i} = 1$ represents the fact that $z_i^2 = 1$ whenever $z_i \in \{-1, 1\}$.

The aim of this paper is the recovery of the correct binary minimizers of problems (3) and (5) by starting from their SDR's (4) and (6). As described, e.g, in Lasserre (2016) and Lemon et al. (2016), the correct binary minimizers can be extracted if and only if the SDR solutions are 1-rank. Therefore, we develop a strategy to reduce, possibly minimize, the rank, by exploiting specific features of the Boolean setting.

3. CONCAVE PENALIZATION EXPLOITING INFORMATION ON THE EIGENVALUES

In this section, we develop the proposed strategy to promote 1-rank solutions to SDR's (4) and (6). Specifically, we propose a suitable cost functional, and we illustrate the algorithms used to minimize it.

3.1 Proposed cost functional

Let us consider SDR (6), and let us call $Z^\star \in \mathbb{S}^{n+1}_+$ the desired 1-rank solution. The rank of $Z^\star$ corresponds to the number of non-null eigenvalues; thus, the eigenvalues vector of $Z^\star$, denoted as v, is 1-sparse, i.e., $v = (v_0, 0, \dots, 0)$, $v_0 > 0$. We also remark that $Z^\star$ necessarily has all the components in $\{-1, 1\}$, and $\frac{1}{2}((Z^\star_{0,1}, \dots, Z^\star_{0,n})^T + \mathbf{1})$ is the exact minimizer of (2). Recht et al. (2010); Lemon et al. (2016) show that a sparsifying effect on v can be obtained by penalizing the trace of the variable Z, hereafter denoted as $\mathrm{tr}(Z)$, which corresponds to the ℓ_1-norm of the eigenvalues, and, in turn, the ℓ_1-norm is a suitable convex relaxation of the sparsity of a vector. Nevertheless, this approach is not beneficial for SDR (6), as $\mathrm{tr}(Z) = n + 1$ by construction; therefore, it makes no sense to penalize a constant quantity.

However, we observe that the information $\mathrm{tr}(Z^\star) = \sum_{i=0}^n v_i = n + 1$ can be exploited to state that $v_0 = n + 1$. In other terms, not only we know that v is 1-sparse, but also that its components belong to the binary alphabet $\{0, n + 1\}$. Then, we wonder how to force the solution to have $v = \{n + 1, 0, \dots, 0\}$.

The key idea is as follows. Given a vector $v \in [0, n + 1]^{n+1}$ with $\sum_{i=0}^n v_i = n + 1$, we can force 1-sparsity by maximizing its energy $\|v\|_2^2$. This is straightforward to check in the two-dimensional case: let us consider $(v_0, v_1) \in [0, 2]^2$ with $v_0 + v_1 = 2$; the maximum of $v_0^2 + v_1^2 = v_0^2 + (2 - v_0)^2$ is achieved at the boundaries, that is, at $(0, 2)$ or $(2, 0)$. This reasoning can be extended to any dimension.

Given this principle, we search a method to maximize the energy of v within SDR (6). We notice that $\|v\|_2^2 = \mathrm{tr}(ZZ) = \langle Z, Z\rangle$. Then, we propose to add a term that penalizes $-\langle Z, Z\rangle$ in the cost functional of (6), i.e.,

$$\begin{aligned} &\min_{Z\in\mathbb{S}^{n+1}_+} \langle R, Z\rangle - \lambda\langle Z, Z\rangle \\ &\text{s. t. } Z_{i,i} = 1, \quad i = 0, \dots, n \end{aligned} \tag{7}$$

where $\lambda > 0$ is a design parameter that can be assessed by cross-validation. Interestingly, if a global minimizer of (6) is binary, then, for any $\lambda > 0$, the global minimizer of (7) is exact, as illustrated in the following proposition.

Proposition 1. Let $\widetilde{z} \in \{-1, 1\}^{n+1}$ be the correct solution to problem (5). Let us assume that SDR (6) has the desired binary, 1-rank solution $\widetilde{z}\widetilde{z}^T$ among its global minimizers. Then, the minimizer of (7) is $Z^\star = \widetilde{z}\widetilde{z}^T$, for any $\lambda > 0$. Moreover, $\widetilde{z} = (1, Z^\star_{1,0}, \dots, Z^\star_{n,0})^T$.

Proof. Since $\mathrm{tr}(Z) = n + 1$, then $\langle Z, Z\rangle \leq (n + 1)^2$. In particular, the maximum $(n+1)^2$ is achieved if and only if the eigenvalues of Z are $v = (n + 1, 0, \dots, 0)$. In this case,

Z is 1-rank, and necessarily $Z = \widetilde{z}\widetilde{z}^T$. Moreover, since by assumption $\widetilde{z}\widetilde{z}^T$ minimizes $\langle R, Z\rangle$, then it is the unique global minimizer of (7). □

A similar approach can be applied to SDR (4), with a considerable difference: $\mathrm{tr}(X)$ is not priorly set by construction. Then, we proceed by considering two possible settings.

In the first setting, we assume to know the sparsity level k of the true solution of problem (3). This implies that the desired solution to SDR (4) has trace equal to $(k+1)^2$ and eigenvalues $v = (k+1, 0, \ldots, 0)$. The sparsity level is known in many applications, e.g., in compressed sensing, see (Foucart and Rauhut, 2013, Sec. III), or in sensor selection, where k is the number of used sensors, see Shekhar et al. (2014). In other cases, an unknown k can be estimated through ad hoc techniques, see Ravazzi et al. (2018). Given k, we propose to modify SDR (4) as follows to leverage the binary nature of the eigenvalues:

$$\begin{aligned} &\min_{X \in \mathbb{S}^n_+} \langle Q, X\rangle + \lambda\left[(k+1)\mathrm{tr}(X) - \langle X, X\rangle\right] \\ &\text{s. t. } X_{0,0} = 1; \quad X_{i,i} = X_{0,i}, \quad i = 1, \ldots, n. \end{aligned} \tag{8}$$

Proposition 2. Let $\widetilde{x} \in \{0,1\}^{n+1}$ be the correct solution to problem (3). Let us assume that SDR (4) has the desired binary, 1-rank solution $\widetilde{x}\widetilde{x}^T$ among its global minimizers. Then, the minimizer of problem (8) is $X^\star = \widetilde{x}\widetilde{x}^T$, for any $\lambda > 0$. Moreover, $\widetilde{x} = (1, X^\star_{1,0}, \ldots, X^\star_{n,0})^T$.

Proof. If v is the eigenvalues vector of X, we have

$$\mathrm{tr}(X) - \langle X, X\rangle = \sum_{i=1}^{n}[(k+1)v_i - v_i^2] \geq 0 \tag{9}$$

since $v_i \in [0, k+1]$ for each $i = 1, \ldots, n$. Moreover, $\mathrm{tr}(X) - \langle X, X\rangle = 0 \iff v_i \in \{0, k+1\}$ for each $i = 1, \ldots, n$. Then, we can conclude that the global minimum is achieved for $v \in \{0, k+1\}^{n+1}$. Eventually, since $\sum_{i=1}^n v_i = k+1$, we have $v = (k+1, 0, \ldots, 0)$, which implies the 1-rank solution, and $X^\star = \widetilde{x}\widetilde{x}^T$. □

In the second setting, we assume that k is unknown. In this case, we propose to replace k by $n \geq k$. This approximate procedure takes a larger weight on the term $\mathrm{tr}(X)$, which might be advantageous when $k \ll n$: in fact, in this setting, not only $\mathrm{tr}(X^\star) = \sum_{i=1}^{n+1} v_i$, but also $\mathrm{tr}(X^\star) = 1 + \sum_{i=1}^n x_i^\star$, where $x^\star \in [0,1]^n$ is the final estimation of $\widetilde{x}$, extracted from the diagonal of $X^\star$: $x^\star = (X^\star_{1,1}, \ldots, X^\star_{n,n})^T$. In conclusion, by penalizing $\mathrm{tr}(X)$ we obtain a sparsifying effect both on the eigenvalues and on the solution; for this motivation, we expect better performance when $k \ll n$.

Remark 1. The considered problem bears some similarities with phase retrieval from Fourier transform magnitude, see, e.g., Jaganathan et al. (2013, 2017). An effective approach to phase retrieval is the embedding of the unknown vector x into a higher dimensional space by the transformation $X = xx^T$. This approach, called lifting, shares the same principle of the SDR's illustrated in this paper. Similarly to the MAX-CUT problem, the penalization of the trace is not effective for phase retrieval, as the energy of x is fixed, see (Jaganathan et al., 2013, Sec. III). In Jaganathan et al. (2013), this issue is overcome by penalizing the term $\log\det(X + \epsilon I)$, where $\epsilon > 0$ is a small design parameter necessary for boundedness. This term is a concave surrogate of the rank, whose effectiveness is discussed in Fazel et al. (2003). Our approach consists in a concave penalization as well. However, differently from the log-det heuristic, it exploits the known binary eigenvalues; moreover, its practical implementation is computationally less complex, as illustrated in Remark 2.

3.2 Descent algorithms

Problems (8) and (7) are well posed, that is, their unique global minima are the correct binary solutions. However, they introduce the concave term $-\langle X, X\rangle$, which makes the problem non-linear and non-convex. For this motivation, we propose an iterative descent algorithm to search the minimum, which, although sub-optimal, is effective in practice. The idea is to replace the concave term $\langle X, X\rangle$ with $\langle G, X\rangle$, where G is an available estimate of X. By assuming G fixed, the penalty is linear in X, then the whole cost functional is linear. Then, we propose an iterative procedure: we start from an initial estimate of the solution, denoted by X_0, which can be assessed, for example, by solving the non-penalized problems (4) and (6); in turn, we solve the penalized problems until convergence is reached. The overall procedure is summarized in algorithms 1 and 2 for problems (8) and (7), respectively . In Algorithm 1, h is equal to $k+1$ when k is known, and $n+1$ otherwise.

Algorithm 1 Descent algorithm for Problem (8)

Input: $Q, \lambda > 0$;
1: $X_0 =$ solution of (4)
2: **for all** $t = 1, \ldots, T$ **do**
3: $X_t = \underset{X \in \mathbb{S}^{n+1}_+}{\mathrm{argmin}} \langle Q, X\rangle + \lambda\left[h\mathrm{tr}(X) - \langle X_{t-1}, X\rangle\right]$, s. t. $X_{0,0} = 1$, $X_{i,i} = X_{0,i}$, $i = 1, \ldots, n$
4: **end for**

Algorithm 2 Descent algorithm for Problem (7)

Input: Input: $R, \lambda > 0$;
1: $X_0 =$ solution of (6)
2: **for all** $t = 1, \ldots, T$ **do**
3: $X_t = \underset{X \in \mathbb{S}^{n+1}_+}{\mathrm{argmin}} \langle R, X\rangle - \lambda\langle X_{t-1}, X\rangle$, s. t. $X_{i,i} = 1$, $i = 0, \ldots, n$
4: **end for**

In this way, we solve a sequence of semidefinite programming problems. This is not guaranteed to get the global minimum, while it is guaranteed to provide a non-increasing cost functional sequence.

Proposition 3. Let us define $f(X) := \langle Q, X\rangle + \lambda h\mathrm{tr}(X)$ from (8) (respectively, $f(X) := \langle R, X\rangle$ in (7)), and $F(X_t) := f(X_t) - \frac{\lambda}{2}\langle X_t, X_t\rangle$.

By applying Algorithm 1 (respectively, Algorithm 2), $F(X_t)$ is a non-increasing function.

Proof. Since X_t is the minimizer,

$$f(X_t) - \lambda\langle X_{t-1}, X_t\rangle \leq f(X_{t-1}) - \lambda\langle X_{t-1}, X_{t-1}\rangle. \tag{10}$$

On the other hand, if $A, B \in \mathbb{S}^n_+$, then $2\langle A, B\rangle \leq \langle A, A\rangle + \langle B, B\rangle$, see, e.g., Zhou (2014). Therefore, for any t, $2\langle X_{t-1}, X_t\rangle - \langle X_{t-1}, X_{t-1}\rangle \leq \langle X_t, X_t\rangle$, which implies

$$\begin{aligned} F(X_t) &\leq f(X_t) - \lambda\langle X_{t-1}, X_t\rangle + \frac{\lambda}{2}\langle X_{t-1}, X_{t-1}\rangle \\ &\leq f(X_{t-1}) - \lambda\langle X_{t-1}, X_{t-1}\rangle + \frac{\lambda}{2}\langle X_{t-1}, X_{t-1}\rangle \\ &\leq F(X_{t-1}) - \frac{\lambda}{2}\langle X_{t-1}, X_{t-1}\rangle + \frac{\lambda}{2}\langle X_{t-1}, X_{t-1}\rangle \\ &= F(X_{t-1}). \end{aligned}$$

□

Furthermore, if the proposed approach achieves a binary solution, by uniqueness it is guaranteed that this is exactly the desired global minimum. Conversely, when the obtained solution is not 1-rank, the final solution is not binary and it is guaranteed that it is not the correct solution. This awareness about achieving the correct solution provides the possibility of running again the algorithm by suitably changing the initialization, which may yield a better solution.

***Remark** 2.* The proposed descent strategy can be interpreted as a reweighting algorithm. In Jaganathan et al. (2013); Fazel et al. (2003), the reweighting algorithm is derived as iterative minimization of the local linearization of the concave log-det term. The convergence of the log-det method to a local minimum is discussed in Fazel et al. (2003). We mention that the proposed descent algorithm can be described under the reweighting viewpoint as well, and results from Fazel et al. (2003); Fosson (2018a) might be leverage to rigorously prove its convergence. The convergence analysis and a complete comparison to log-det heuristic are beyond the scope of this paper; however, some numerical comparisons are proposed in Sec. 4.

4. NUMERICAL RESULTS

In this section, we illustrate some numerical results, that support the effectiveness of the proposed approach. We consider the following problem, as presented in Fosson and Abuabiah (2019): we aim to solve the underdetermined system

$$b = Ax, \quad x \in \{0,1\}^n, \;\; b \in \mathbb{R}^m, \;\; A \in \mathbb{R}^{m,n}, \;\; m < n \quad (11)$$

under the assumption that the solution is unique in $\{0,1\}^n$. This linear problem is encountered in a number of applications, ranging from compressed sensing, see Flinth and Kutyniok (2018); Fosson (2018b, 2019) to tomography, see Schüle et al. (2005); Weber et al. (2015).

In Fosson and Abuabiah (2019), problem (11) is tackled in case of sparse x, which recasts into binary compressed sensing, and a Shor SDR is proposed to solve it. In previous works, other methods for binary compressed sensing are proposed, namely, relaxation over the convex hull in Flinth and Kutyniok (2018), ℓ_1-reweighting algorithms in Fosson (2018b), alternating direction method of multipliers in Fosson (2019), and difference of convex functions in Schüle et al. (2005); Weber et al. (2015). SDR is shown to achieve better accuracy in Fosson and Abuabiah (2019).

Here, we recast problem (11) into Boolean quadratic optimization by replacing the constraint $Ax = b$ with the the cost functional $\|Ax-b\|_2^2$. As we know that $\widetilde{x} \in \{0,1\}^n$ is solution of $Ax = b$, the global minimum is null. Given, the cost functional $\|Ax-b\|_2^2$, we homogenize it by adding the slack variable $x_0 = 1$, so that $\|Ax-x_0b\|_2^2 = x^TQx$ with $Q = (-b|A)^T(-b|A)$, and $x = (x_0, x_1, \ldots, x_n)^T$. Then, we can apply the approach developed in Sec. 3.

Specifically, we compare the recovery accuracy of the proposed "known binary eigenvalues" (KBE) approach to the following known methods: SDR (4) and SDR (6), nuclear norm heuristic, and log-det heuristic, see Fazel et al. (2003); Lemon et al. (2016). The implemented nuclear norm and log-det algorithms read as follows, respectively:

$$\begin{aligned} &\min_{X \in \mathbb{S}_+^{n+1}} \langle Q, X\rangle + \lambda \mathrm{tr}(X) \\ &\qquad \text{s. t. } X_{0,0} = 1, \quad X_{i,i} = X_{0,i}, \quad i = 1, \ldots, n \end{aligned} \quad (12)$$

and, for $t = 1, \ldots, T$,

$$\begin{aligned} &X_t = \operatorname*{argmin}_{X \in \mathbb{S}_+^{n+1}} \langle Q, X\rangle + \lambda\langle (X_{t-1} + \epsilon I)^{-1}, X\rangle \\ &\qquad \text{s. t. } X_{0,0} = 1, \quad X_{i,i} = X_{0,i}, \quad i = 1, \ldots, n. \end{aligned} \quad (13)$$

The considered setting is a follows. We take binary vectors $\widetilde{x} \in \{0,1\}^n$, with $n = 50$, sparsity level $k \in [5, 45]$, and uniformly distributed support. A is a random Gaussian matrix with m rows, with $m \in [14, 34]$. For all the methods, $\lambda = 10^{-4}$, and $T = 3$. Algorithms 1 and 2, and log-det heuristic (13) are initialized with the solutions of SDR's (4) and (6). If a binary solution is not achieved, i.e., the algorithm stops in a local minimum or saddle point, we randomly reinitialize for a maximum of 5 times. In (13), we set $\epsilon = 10^{-6}$. To solve the involved SDR's, we use the Mosek C++ Fusion API, see MOSEK (2019), which guarantees fast solutions even for quite large dimensional problems. For reasons of space, a complete analysis of the complexity and large-scale experiments are left for future work. We just remark that the matrix inversion in the log-det heuristic yields an higher computational complexity when compared to Algorithms 1 and 2, and might be prohibitive in large-scale problems. The considered algorithms are compared in terms of exact recovery rate, i.e., the number of experiments where a perfect recovery is achieved. Results are averaged over 200 random runs. As SDR's (4) and (6) are different formulations of the same problem, they are conveyed in the same graph.

In Fig. 1, we show the case of unknown k. We observe that Algorithm 2 increases the success rate with respect to SDR's (4)-(6). Specifically, KBE - Algorithm 2 has a transition phase from low to high recovery rate (namely, from 60% to 90%, highlighted by red color in the figure) with less measurements than SDR's (4)-(6): $m \in [24, 26]$ instead of $m \in [26, 28]$. We remark that these two methods have approximately constant performance in k. In contrast, as discussed above, the nuclear norm and log-det heuristics as well as KBE - Algorithm 1 penalize the trace, hence they have a sparsifying effect that makes them favorable when k is small. Among them, the proposed KBE - Algorithm 1 generally achieves the best performance. In conclusion KBE - Algorithm 1 is the most reliable approach in non sparse problems, while KBE - Algorithm 2 is the most reliable approach for sparse problems. This attests that the KBE methodology is effective.

In Fig. 2, the same experiment is reported in case of known k. As discussed in Sec. 3, Algorithm 1 can specifically exploit the information on k. Moreover, the equation $\mathbf{1}^Tx = k$ is added to $Ax = b$ for all the methods. In

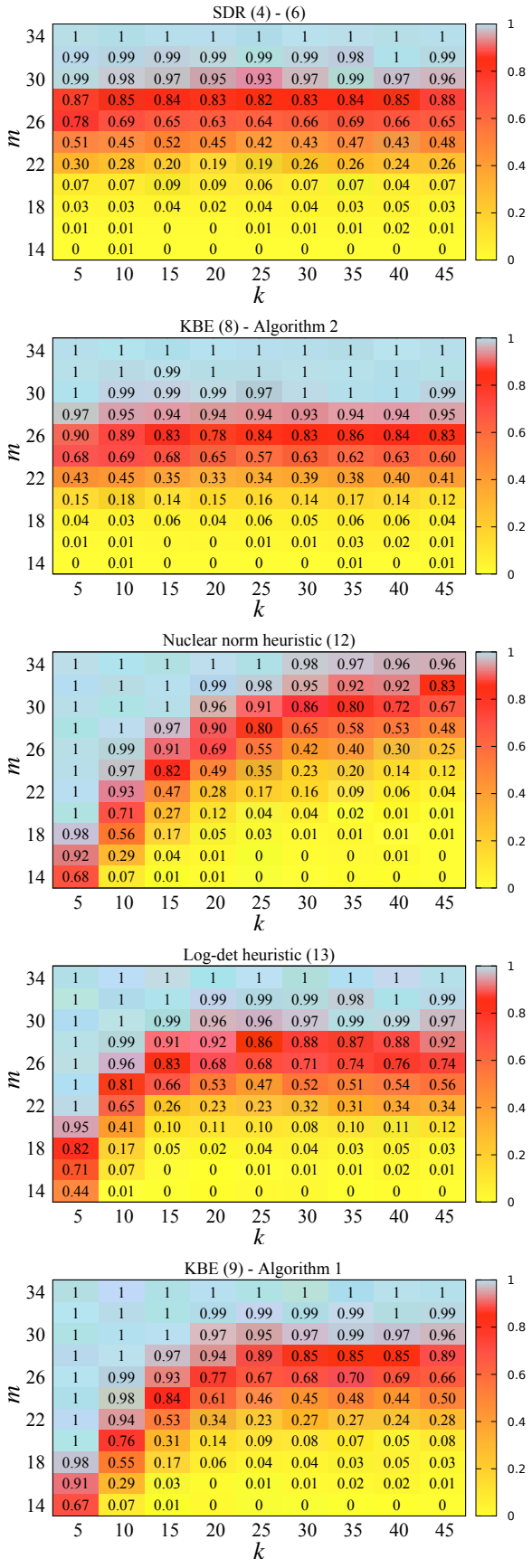

Figure 1. Experiment 1 (unknown k): exact recovery rate.

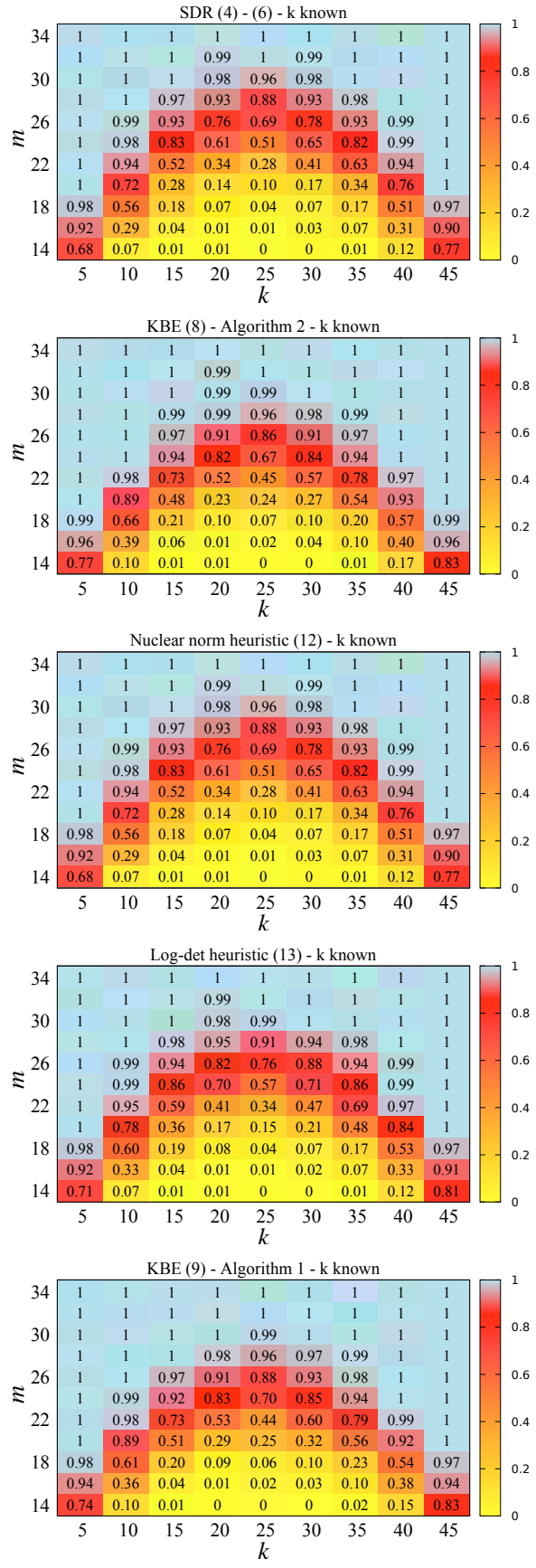

Figure 2. Experiment 2 (known k): exact recovery rate.

Fig. 2, we observe that, as expected, all the methods take advantage of the knowledge of k, in particular when k is either very small or very large k. Also in this experiment, we observe that the proposed KBE Algorithms 1-2 are more accurate than the other methods. In particular, they are successful in at least 99% of runs when $m \geq 30$, which is not achieved by the competitors.

Finally, we observe that, in our experiments, the run-time is between 0.1 and 0.8 seconds for KBE Algorithms 1-2, and between 0.2 and 1.4 seconds for the log-det heuristic.

5. CONCLUSION

Semidefinite programming relaxations of Boolean quadratic problems provide the right minimizer if a 1-rank solution is found; nevertheless, enforcing low-rank solutions is NP-hard. In this paper, we tackle this drawback by leveraging known information on the eigenvalues. The global minimum of the proposed cost functional is proven to be the exact solution. A low-complex descent algorithm is developed to estimate the minimum, and shown to be effective through numerical results. Future work will envisage a formal proof of the convergence of the proposed algorithms, and the development of different descent techniques, with particular focus on large scale problems.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Jean-Bernard Lasserre for his constructive comments on the manuscript.

REFERENCES

Bemporad, A. and Naik, V.V. (2018). A numerically robust mixed-integer quadratic programming solver for embedded hybrid model predictive control. In *Proc. IFAC NMPC*, 20, 412–417.

Carmi, A. and Gurfil, P. (2013). Sensor selection via compressed sensing. *Automatica*, 49(11), 3304–3314.

Fazel, M., Hindi, H., and Boyd, S. (2003). Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In *Proc. IEEE ACC*, volume 3, 2156–2162.

Flinth, A. and Kutyniok, G. (2018). PROMP: A sparse recovery approach to lattice-valued signals. *Appl. Comput. Harmon. Anal.*, 45(3), 668–708.

Fosson, S.M. (2018a). A biconvex analysis for Lasso ℓ_1 reweighting. *IEEE Signal Process. Lett.*, 25(12), 1795–1799.

Fosson, S.M. (2018b). Non-convex approach to binary compressed sensing. In *Asilomar Conf. Signals Syst. Comput.*, 1959–1963.

Fosson, S.M. (2019). Non-convex Lasso-kind approach to compressed sensing for finite-valued signals. *ArXiv*. URL `https://arxiv.org/pdf/1811.03864.pdf`.

Fosson, S.M. and Abuabiah, M. (2019). Recovery of binary sparse signals from compressed linear measurements via polynomial optimization. *IEEE Signal Process. Lett.*, 26(7), 1070–1074.

Foucart, S. and Rauhut, H. (2013). *A Mathematical Introduction to Compressive Sensing*. Springer.

Goemans, M.X. and Williamson, D.P. (1995). Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6), 1115–1145.

Jaganathan, K., Oymak, S., and Hassibi, B. (2013). Sparse phase retrieval: Convex algorithms and limitations. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 1022–1026.

Jaganathan, K., Oymak, S., and Hassibi, B. (2017). Sparse phase retrieval: Uniqueness guarantees and recovery algorithms. *IEEE Trans. Signal Process.*, 65(9), 2402–2410.

Kellerer, H., Pisinger, H., Pferschy, U., and Pisinger, D. (2004). *Knapsack Problems*. Springer.

Lasserre, J. (2001). Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3), 796–817.

Lasserre, J.B. (2000). Convergent lmi relaxations for nonconvex quadratic programs. In *Proc. IEEE CDC*, 5041–5046.

Lasserre, J.B. (2015). *An introduction to polynomial and semi-algebraic optimization*. Cambridge Univ. Press.

Lasserre, J.B. (2016). A MAX-CUT formulation of 0/1 programs. *Oper. Res. Lett.*, 44(2), 158–164.

Lemon, A., So, A.M.C., and Ye, Y. (2016). Low-rank semidefinite programming: Theory and applications. *Found. Trends Optim.*, 2(1-2), 1–156.

Lovász, L. and Shrijver, A. (1990). Cone of matrices and setfunctions. and 0-1 optimization. *SIAM J. Optim.*, 1, 166–190.

MOSEK (2019). *Fusion API for C++*. URL `www.mosek.com/documentation/`.

Naik, V.V. and Bemporad, A. (2017). Embedded mixed-integer quadratic optimization using accelerated dual gradient projection. In *Proc. IFAC World Congress*, 10723–10728.

Nesterov, Y. (1998). Semidefinite relaxation and nonconvex quadratic optimization. *Optim. Methods Softw.*, 9(1-3), 141–160.

Palagi, L., Piccialli, V., Rendl, F., Rinaldi, G., and Wiegele, A. (2012). *Computational Approaches to Max-Cut*, 821–847. Springer US, Boston, MA.

Ravazzi, C., Fosson, S.M., Bianchi, T., and Magli, E. (2018). Sparsity estimation from compressive projections via sparse random matrices. *EURASIP J. Adv. Signal Process.*, 2018(1).

Recht, B., Fazel, M., and Parrilo, P.A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3), 471–501.

Schüle, T., Schnörr, C., Weber, S., and Hornegger, J. (2005). Discrete tomography by convexconcave regularization and D.C. programming. *Discrete Appl. Math.*, 151, 229–243.

Shekhar, S., Patel, V., Nasrabadi, N., and Chellappa, R. (2014). Joint sparse representation for robust multimodal biometrics recognition. *IEEE Trans. Patt. Ana. & Mach. Intel.*, 36(1), 113–126.

Shor, N.Z. (1987). Quadratic optimization problems. *Sov. J. Comput. Syst. Sci*, 26.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. Royal. Statist. Soc. B*, 58, 267–288.

Weber, S., Schüle, T., and Schnörr, C. (2015). Prior learning and convex-concave regularization of binary tomography. *Elec. Notes Discrete Math.*, 20, 313–327.

Zhou, H. (2014). On some trace inequalities for positive definite Hermitian matrices. *J. Inequal. Appl.*, 64(1).