New bounds for subset selection from conic relaxations

Walid Ben-Ameur, José Neto

New bounds for subset selection from conic relaxations

Walid Ben-Ameur[a], José Neto[a,*]

[a]*Samovar, Télécom SudParis, Institut Polytechnique de Paris, 19 place Marguerite Perey, 91120 Palaiseau, France*

Abstract

New bounds are proposed for the subset selection problem which consists in minimizing the residual sum of squares subject to a cardinality constraint on the maximum number of non-zero variables. They rely on new convex relaxations providing both upper and lower bounds that are compared with others present in the literature. The performance of these methods is illustrated through computational experiments.

Keywords: Combinatorial optimization, subset selection, convex relaxation

1. Introduction

This paper is dedicated to the well-known *subset selection problem* (denoted by SSP) which may be presented as follows. Let $A \in \mathbb{R}^{m\times n}$ and $y \in \mathbb{R}^m$ be some given matrix and vector respectively. Let k represent a positive integer value. Then, SSP can be formulated as

$$(SSP1)\begin{cases} Z^\star = & \min & \|y - Ax\|_2^2 \\ & s.t. & \|x\|_0 \leq k \\ & & x \in \mathbb{R}^n \end{cases}$$

where, for any vector $x \in \mathbb{R}^n$, $\|x\|_0 = |\{i\colon x_i \neq 0\}|$ (i.e., the number of non-zero entries in x, or equivalently, the cardinality of its support), while $\|x\|_2 =$

*Corresponding author
Email addresses: walid.benameur@telecom-sudparis.eu (Walid Ben-Ameur), jose.neto@telecom-sudparis.eu (José Neto)

$\sqrt{\sum_{i=1}^{n} x_i^2}$ is the Euclidean norm of x. In words, the problem consists in determining a vector in a subspace generated by at most k column vectors of A, and that is closest to the vector y with respect to Euclidean distance.

SSP is known to be NP-hard (see Natarajan (1995)). It has many applications, e.g., for (sparse) decomposition of audio signals (see Gribonval & Bacry (2003)), video coding (Neff & Zakhor (1997)), image denoising (see Elad & Aharon (2006); Mairal et al. (2008)), shape representation and recognition (Mendels et al. (2006)), health monitoring (Chakraborty et al. (2009)) and machine learning (Girosi (1998)). SSP is also related to the feature selection problem (also called "variable selection" or "attribute selection" problem) arising in the data mining area where the objective is to determine a subset of relevant features to build simple models providing a more compact representation of an important amount of available information. Classical applications include the analysis of written texts (Wichmann & Kamholz (2008)) and DNA microarray data (Davis et al. (2006)).

Related to SSP is the following problem:

$$\min_{x \in \mathbb{R}^n} \|y - Ax\|_2^2 + p(x), \tag{1}$$

where $p(\cdot)$ stands for some penalty function. For the particular case $p(x) = \lambda\|x\|_1$, where $\|x\|_1$ denotes the L_1 norm and $\lambda \in \mathbb{R}_+$, (1) becomes a convex quadratic optimization problem that can be solved efficiently (see Efron et al. (2004)) and corresponds to the Lagrangian form of the *lasso* (*'least absolute shrinkage and selection operator'*) estimate (see Tibshirani (1996)). However, *lasso* has also some drawbacks, such as biased regression or a large support of the solution (see, e.g., Bertsimas et al. (2016) and the references therein). To overcome these, there has also been some interest to consider non convex penalties in (1) (see e.g. Fan & Li (2001); Zhang (2010)). Methods based on non convex penalty functions such as in Mazumder et al. (2011), look for a local minimizer and come with no guarantee on the approximation of the solution found. And in general, differently from SSP, a (local or global) minimizer of (1) may not comply exactly with some desired level of sparsity. Shen et al. (2013)

draw a comparison between the formulations $(SSP1)$ and (1) with $p(x) = \lambda\|x\|_0$, $\lambda \in \mathbb{R}_+$; see also Tropp (2006) and the references therein.

The methods just mentioned have also been widely used to derive approximate solutions heuristically for SSP, in particular with greedy-type procedures which basically consist in iteratively selecting or eliminating variables, the decision at each step being based on a criterion related to the objective value of a subproblem of the same form as SSP, see e.g. Miller (2002); Moghaddam et al. (2006). Theoretical bounds on the performance of greedy algorithms appear in Das & Kempe (2011).

Approaches to solve SSP (exactly or approximately) with branch and bound algorithms include Furnival & Wilson (1974); Hand (1981); Moghaddam et al. (2008); Narendra & Fukunaga (1977). More recently, Bertsimas et al. (2016) present an approach for solving SSP, which is based on a reformulation of the problem as a mixed-integer quadratic optimization program (MIQP). A different approach, which we explore further in this paper, has consisted in looking for new convex relaxations for SSP. In particular, this work may be viewed in the continuity of Bach et al. (2010) and Atamtürk & Gómez (2019). (For clarity of the presentation, we postpone a description of their and other formulations present in the literature to the next section). Our contributions are new convex relaxations providing lower bounds together with procedures relying on them for producing approximate solutions. We also illustrate the performance of our approach compared to Bach et al. (2010) and Atamtürk & Gómez (2019) through computational experiments.

The paper is organized as follows. In Section 2 we survey formulations of SSP present in the literature.

In Section 3, we prove some preliminary results and show connections between the approaches of Bach et al. (2010), Atamtürk & Gómez (2019) and ours. New lower bounds relying on different convex relaxations are then introduced in Section 4. Heuristics based on the solutions of such relaxations are then presented in Section 5 to obtain upper bounds. Computational experiments illustrating the performance of the different approaches are reported in Section

6, before we conclude in Section 7.

2. Formulations and convex relaxations from the literature

Atamtürk & Gómez (2019) derived convex relaxations for sparse regression from convex formulations of rank one quadratic terms with indicator variables. Among the different formulations they present, there is namely a family of semidefinite relaxations that we report just after introducing some notation. Given a vector $v \in \mathbb{R}^n$, a matrix $B \in \mathbb{R}^{n\times n}$, and a subset $T \subset \{1, 2, \ldots, n\}$, let v_T (resp. B_T) denote the subvector of v induced by T (resp. the submatrix of B induced by T). The set of symmetric matrices of order n is denoted by $\mathcal{S}^n$, and the set of positive semidefinite matrices of order n is denoted by $\mathcal{S}^n_+$. Atamtürk and Gómez's convex relaxation for SSP may then be expressed as follows where $Tr(.)$ denotes the trace of a square matrix:

$$(SDP)_r \begin{cases} \min & \|y\|_2^2 - 2y^\top Ax + Tr\left(BA^\top A\right) & \\ s.t. & \sum_{i=1}^n z_i \leq k & \\ & 0 \leq w_T \leq \min\left\{1, \sum_{i\in T} z_i\right\} & \forall T \subseteq \{1, 2, \ldots, n\} \colon |T| \leq r \\ & w_T B_T - x_T x_T^\top \in \mathcal{S}_+^{|T|} & \forall T \subseteq \{1, 2, \ldots, n\} \colon |T| \leq r \\ & B - xx^\top \in \mathcal{S}_+^n & \\ & x \in \mathbb{R}^n, z \in [0, 1]^n, B \in \mathbb{R}^{n\times n} & \end{cases} \tag{2}$$

for some $r \in \mathbb{N}$, $r \leq n$. These relaxations provide bounds with increasing quality for an increasing value of r, but at the expense of a very rapidly increasing computational work to solve them. In particular, the bound from $(SDP)_1$ coincides with the one from other works of Dong et al. (2015), Zhang (2010). In the computational experiments reported in Atamtürk & Gómez (2019) (see this reference for further details), whereas the bounds of $(SDP)_1$ may be poor for some instances, $(SDP)_2$ appears to provide substantial improvements and offer a good trade-off between computational effort and quality of the bound.

Another line of research was initiated by Moghaddam et al. (2008) who proved that SSP is in fact equivalent to a rank-1 sparse generalized eigenvalue

problem. This is made explicit in the following expression of the optimal objective of SSP derived by Bach et al. (2010):

$$Z^\star = \|y\|_2^2 - \rho^\star, \text{ with } \rho^\star = \max_{\substack{u\in\{0,1\}^n \\ ||u||_1\leq k}} \max_{\substack{\|x\|_2=1 \\ ADiag(u)x\neq 0}} \frac{\left(y^\top ADiag(u)x\right)^2}{\|ADiag(u)x\|_2^2} \tag{3}$$

where $Diag(u)$ is the diagonal matrix whose diagonal is u. Given $M \in \mathcal{S}^n$ and a positive integer k, its k sparse maximum eigenvalue, denoted by $\lambda_{\max}^k(M)$, is defined by $\lambda_{\max}^k(M) = \max_{\substack{\|x\|_2=1 \\ \|x\|_0\leq k}} x^\top Mx$. It is shown in Bach et al. (2010) that condition $\rho \geq \rho^\star$ is equivalent to $\lambda_{\max}^k\left(A^\top\left(yy^\top - \rho I_n\right)A\right) \leq 0$. D'Aspremont et al. (2007) derived the following convex relaxation to get an upper bound on $\lambda_{\max}^k(M)$, for any given $M \in \mathcal{S}^n$,

$$\left\{\begin{array}{lll} \widehat{\lambda_{\max}^k}(M) = & \max & Tr\left(MX\right) \\ & s.t. & \|X\|_1 \leq k \\ & & Tr(X) = 1 \\ & & X \in \mathcal{S}_+^n \end{array}\right. \tag{4}$$

where $\|X\|_1$ denotes the standard L_1 norm of a matrix ($\|X\|_1 = \sum_{i=1}^n \sum_{j=1}^p |X_{ij}|$).

Using duality, Bach et al. (2010) proved that $\widehat{\lambda_{\max}^k}(M) = \min_{Y\in\mathcal{S}^n} \lambda_{\max}\left(M+Y\right) + k\|Y\|_\infty$ where $\|Y\|_\infty = \max_{i,j}|Y_{ij}|$. This implies that $Z^\star \geq \|y\|_2^2 - \widehat{\rho}$ where

$$\left\{\begin{array}{lll} \widehat{\rho} = & \min & \rho \\ & s.t. & \rho A^\top A - k\mu I_n + Y - A^\top yy^\top A \in \mathcal{S}_+^n \\ & & |Y_{ij}| \leq \mu, \forall i,j \\ & & (\rho,\mu) \in \mathbb{R}^2, Y \in \mathcal{S}^n \end{array}\right. \tag{5}$$

and I_n denotes the identity matrix with order n.

3. Preliminary results

We now introduce a reformulation of SSP from which convex relaxations can be derived that allow us to draw some connections between the works from Bach et al. (2010), Atamtürk & Gómez (2019) and ours.

Let $\mathcal{V}_k = \{xx^\top : x \in \mathbb{R}^n, \|x\|_2 = 1, \|x\|_0 \leq k\}$. Let $\text{conv}(\mathcal{V}_k)$ (resp. $\text{cone}(\mathcal{V}_k)$) denote the convex hull (resp. conic hull) of $\mathcal{V}_k$. We obviously have $\text{cone}(\mathcal{V}_k) = \text{cone}(\{xx^\top : x \in \mathbb{R}^n, \|x\|_0 \leq k\})$ and $V \in \text{cone}(\mathcal{V}_k)$, if and only if, $V \in Tr(V) \cdot \text{conv}(\mathcal{V}_k)$.

Let us first prove that SSP can be formulated as follows.

$$\begin{cases} \min & \|y\|_2^2 - Tr(XA^\top yy^\top A) \\ s.t. & Tr(XA^\top A) = 1 \\ & X \in \text{cone}(\mathcal{V}_k). \end{cases} \tag{6}$$

Proposition 1. *The formulations* $(SSP1)$ *and* (6) *have the same optimal objective value* $Z^\star$.

Proof. Observe that

$$\max_{\substack{X \in \text{cone}(\mathcal{V}_k): \\ Tr(XA^\top A) \neq 0}} \frac{Tr(XA^\top yy^\top A)}{Tr(XA^\top A)} = \max_{\substack{\|x\|_0 \leq k: \\ Tr(xx^\top A^\top A) \neq 0}} \frac{Tr(xx^\top A^\top yy^\top A)}{Tr(xx^\top A^\top A)} \tag{7}$$

$$= \max_{\substack{\|x\|_0 \leq k \\ Ax \neq 0}} \frac{(y^\top Ax)^2}{x^\top A^\top Ax} \tag{8}$$

$$= \rho^\star \tag{9}$$

To write (7) we use the fact that the maximum is achieved at extreme rays of the cone (which is a straightforward consequence of the standard result $(a + b)/(c + d) \leq \max(a/c, b/d)$ for positive numbers). It can be easily checked that (8) is equivalent to (3). Finally, to maximize the ratio $\frac{Tr(XA^\top yy^\top A)}{Tr(XA^\top A)}$, one can impose that $Tr(XA^\top A) = 1$ while $Tr(XA^\top yy^\top A)$ is maximized. □

By considering tractable convex relaxations of $\text{cone}(\mathcal{V}_k)$, we can compute lower bounds for $Z^\star$. Let us consider the following relaxation.

$$\begin{cases} \min & \|y\|_2^2 - Tr(XA^\top yy^\top A) \\ s.t. & Tr(XA^\top A) = 1 \\ & X \in \mathcal{S}_+^n \\ & \|X\|_1 \leq k \cdot Tr(X). \end{cases} \tag{10}$$

The lower bound given by (10) is equal to the bound obtained from (5) (i.e. $\|y\|_2^2 - \widehat{\rho}$). We will loosely say that (10) is *equivalent to* (5).

Proposition 2. *Formulation* (10) *is equivalent to* (5).

Proof. Problem (10) (or more precisely, the maximization of $Tr(XA^\top yy^\top A)$ under the same constraints) is a dual of (5) where ρ is the multiplier corresponding to constraint $Tr(XA^\top A) = 1$ while μ is the multiplier of constraint $\|X\|_1 \leq k \cdot Tr(X)$. Strong duality holds since Slater's conditions are satisfied. □

The lower bounds that will be presented in Section 4 are based on relaxations of (6). Proposition 2 gives the connection with the bound of Bach et al. (2010) and shows that any relaxation of cone($\mathcal{V}_k$) that is tighter than the one considered in (10) can potentially lead to a better lower bound.

Observe that the lower bound obtained from (10) is already exact (i.e, equal to the optimal solution $Z^\star$) when $k = 1$. This is easy to see from formulation (10) since $\|X\|_1 \leq Tr(X)$ is only possible if X is a diagonal matrix implying that $X = \sum_i v^i {v^i}^\top$ where each vector v^i has at most one component (the i^{th} one) different from 0. In other words, $X \in$ cone($\mathcal{V}_1$) and the bound is tight (from Proposition 1).

Let us now look for a connection with the bounds of Atamtürk & Gómez (2019). Let $\mathcal{R}_k$ be any cone containing $\mathcal{V}_k$ (so it is a relaxation of cone($\mathcal{V}_k$)). Consider the formulation

$$\begin{cases} \min & \|y\|_2^2 - 2y^\top Ax + Tr\left(BA^\top A\right) \\ s.t. & B - xx^\top \in \mathcal{S}_+^n \\ & B \in \mathcal{R}_k. \end{cases} \tag{11}$$

For the case $r = 0$, formulation (11) corresponds to (2) with $\mathcal{R}_k = \mathcal{S}^n$. So, by considering cones $\mathcal{R}_k$ satisfying: cone($\mathcal{V}_k$) $\subseteq \mathcal{R}_k \subset \mathcal{S}^n$, (11) can be seen as a strengthened form of (2) for $r = 0$. We are going to prove that (11) is equivalent to the formulation (12) given below whose expression fits with the one of the new proposed formulation (see (32) introduced later).

$$\begin{cases} \min & \|y\|_2^2 - Tr(XA^\top yy^\top A) \\ s.t. & Tr(XA^\top A) = 1 \\ & X \in \mathcal{S}_+^n \cap \mathcal{R}_k. \end{cases} \tag{12}$$

Proposition 3. *Formulation* (12) *is equivalent to* (11).

Proof. Firstly, one can easily check that the result holds if $y^\top A = 0$. So in what follows, we assume $y^\top A \neq 0$. Since $\mathcal{R}_k$ is a cone, (12) is equivalent to

$$\min_{\substack{X \in \mathcal{S}_+^n \cap \mathcal{R}_k : \\ Tr(XA^\top A) \neq 0}} \|y\|_2^2 - \frac{Tr(XA^\top yy^\top A)}{Tr(XA^\top A)}. \tag{13}$$

Let X^* be an optimal solution of (13). Let us add to (11) the constraint $B \in \text{cone}(X^*)$ and consider an optimal solution (B^*, x^*) of the new restricted problem

$$\begin{cases} \min\limits_{B,x} & \|y\|_2^2 - 2y^\top Ax + Tr\left(BA^\top A\right) \\ s.t. & B - xx^\top \in \mathcal{S}_+^n, B \in \text{cone}(X^*). \end{cases} \tag{14}$$

Notice that B^* is also an optimal solution of (13). Multiplying x^* (resp. B^*) by λ (resp. λ^2) leads to a feasible solution of (14) for any positive λ. This means that $\max\limits_{\lambda \geq 0} 2\lambda y^\top Ax^* - \lambda^2 Tr(B^* A^\top A)$ is obtained for $\lambda = 1$. Writing that the derivative is 0 for $\lambda = 1$ implies that $Tr(B^* A^\top A) = y^\top Ax^*$. The optimal objective value in (14) will then be equal to $\|y\|_2^2 - y^\top Ax^*$.

Let us now set B to B^* in (14) and consider the more restricted problem

$$\begin{cases} \min\limits_{x} & \|y\|_2^2 - y^\top Ax \\ s.t. & B^* - xx^\top \in \mathcal{S}_+^n. \end{cases} \tag{15}$$

We know that x^* is an optimal solution of (15). Writing the dual of (15) (see Appendix for the full derivation), we get

$$\begin{cases} \max\limits_{Z,\gamma} & \|y\|_2^2 - Tr(ZB^*) - \gamma \\ s.t. & Z - \frac{1}{4\gamma} A^\top yy^\top A \in \mathcal{S}_+^n \\ & Z \in \mathcal{S}^n, \gamma > 0. \end{cases} \tag{16}$$

Note that for any feasible solution (Z, γ) of (16) we have $Z = C + \frac{1}{4\gamma} A^\top yy^\top A$ for some $C \in \mathcal{S}_+^n$. It follows that $Tr(ZB^*) = Tr(CB^*) + \frac{1}{4\gamma} Tr\left(B^* A^\top yy^\top A\right)$ Since $B^* \in \mathcal{S}_+^n$, we have $Tr(CB^*) \geq 0$ and we can deduce that (16) admits an optimal solution satisfying $Z = \frac{1}{4\gamma} A^\top yy^\top A$. Consequently, (16) has the same optimal objective value as the following problem:

$$\max_{\gamma > 0} \|y\|_2^2 - \frac{1}{\gamma} Tr(B^* \frac{1}{4} A^\top yy^\top A) - \gamma.$$

The maximum is achieved for $\gamma = \frac{1}{2}\sqrt{Tr(B^*A^\top yy^\top A)}$. Since strong duality holds here, we can write that $y^\top Ax^* = Tr(B^*A^\top A) = \sqrt{Tr(B^*A^\top yy^\top A)}$. Hence, we have $\frac{Tr(B^*A^\top yy^\top A)}{Tr(B^*A^\top A)} = Tr(B^*A^\top A) = y^\top Ax^* = 2y^\top Ax^* - Tr(B^*A^\top A)$. In other words, starting from an optimal solution of (12) (or equivalently (13)), we get a feasible solution of (11) having the same objective value.

Let us now consider an optimal solution (B^*, x^*) of (11). The approach used above to prove that $y^\top Ax^* = Tr(B^*A^\top A)$ still applies here. Using $B^* - x^*x^{*\top} \in \mathcal{S}^n_+$, we get that $Tr(B^*A^\top yy^\top A) \geq (y^\top Ax^*)^2$. This implies that $\frac{Tr(B^*A^\top yy^\top A)}{Tr(B^*A^\top A)} \geq y^\top Ax^* = 2y^\top Ax^* - Tr(B^*A^\top A)$. Hence, there is a feasible solution of (12) whose objective value is less than or equal than the optimal objective value of (11). □

4. New lower bounds

We now give approximations of $\text{conv}(\mathcal{V}_k)$ that will allow us to compute better lower bounds. Given any vector $b \in \mathbb{R}^n$ and any integer number $l \leq n$, let $f_l(b)$ denote the square root of the sum of the l largest squares of the components of b. In other words, we have $f_l(b) = \sqrt{\max_{0\leq z_i\leq 1, \sum z_i = l} \sum_{i=1}^n z_i b_i^2}$. We will also use $b_{\setminus i}$ to denote the vector obtained from b by setting the i^{th} component of b to the value zero.

One basic idea underlying the proposed approximations of $\text{conv}(\mathcal{V}_k)$ is (in addition to trace and positive semidefiniteness constraints) to try to generalize the constraint $\|X\|_1 \leq k$ used in (4). Given $X = xx^T$ belonging to $\mathcal{V}_k$, constraint $\|X\|_1 \leq k$ is a consequence of constraints $\sum_{i=1}^n |x_i| \leq \sqrt{k}$ (just by squaring). Observe that the last constraint is of type $\sum_{i=1}^n b_i|x_i| \leq f_k(b)$ where b is the all-one vector of size n. Generalizing this idea to different non-negative b vectors, leads to new formulations.

Lemma 1. *conv*$(\mathcal{V}_k) \subseteq \mathcal{P}_k$ *where*

$$\mathcal{P}_k = \left\{ \begin{array}{ll} X \in \mathcal{S}_+^n \text{ such that:} & \\ \quad Tr(X) = 1, & \text{(17a)} \\ \quad \exists u \in \mathbb{R}_+^n, w \in \mathbb{R}_+^n : & \\ \quad X_{ii} \geq u_i^2, (u_i - X_{ii})(u_i + X_{ii}) \geq w_i^2, \qquad \forall i \in \{1, ..., n\}, & \text{(17b)} \\ \quad \sum_{i=1}^n b_i u_i \leq f_k(b), \qquad \forall b \in \mathbb{R}_+^n, & \text{(17c)} \\ \quad \sum_{j \in \{1,...,n\} \setminus \{i\}} b_j |X_{ij}| \leq f_{k-1}(b_{\setminus i}) w_i, \qquad \forall i \in \{1, ..., n\}, b \in \mathbb{R}_+^n. & \text{(17d)} \end{array} \right.$$

Proof. One can easily check that, for any matrix $X \in \mathcal{V}_k$, there exist vectors $(u, w) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$ satisfying (17b): take $u_i = \sqrt{X_{ii}}$ and $w_i = 0$ for all $i \in \{1, ..., n\}$. Then let us first prove the validity of inequalities (17c) for any matrix $X = xx^\top$ where $\|x\|_2 = 1$ and $\|x\|_0 \leq k$. Note that (17b) implies $u_i \leq \sqrt{X_{ii}} = |x_i|$. It follows that for any $b \in \mathbb{R}_+^n$, we have $b^\top u \leq \sum_{i=1}^n b_i |x_i|$ which is bounded by $\|x\|_2 \cdot \|b\|_2$. Since x has at most k non-zeros components, then at most k components of b are considered. The upper bound $\|x\|_2 \cdot \|b\|_2$ can then be strengthened and replaced by $\|x\|_2 \cdot f_k(b) = f_k(b)$.

Let us now consider the sum $\sum_{j \in \{1,...,n\} \setminus \{i\}} b_j |X_{ij}|$. Focusing again on matrices X written as $X = xx^\top$, we get that $\sum_{j \in \{1,...,n\} \setminus \{i\}} b_j |X_{ij}| = |x_i| \sum_{j \in \{1,...,n\} \setminus \{i\}} b_j |x_j|$. Notice that $x_i = 0$ leads to $X_{ii} = 0$, $u_i = 0$ and $\sum_{j \in \{1,...,n\} \setminus \{i\}} b_j |X_{ij}| = 0$ showing (17d). Let us then focus on the case $x_i \neq 0$. Observe that $x_{\setminus i}$ will then have at most $k - 1$ non-zero components. Using the same argument as above, we get that $\sum_{j \in \{1,...,n\} \setminus \{i\}} b_j |x_j| \leq f_{k-1}(b_{\setminus i}) \|x_{\setminus i}\|_2 = f_{k-1}(b_{\setminus i}) \sqrt{1 - x_i^2}$. Then, $\sum_{j \in \{1,...,n\} \setminus \{i\}} b_j |X_{ij}| \leq f_{k-1}(b_{\setminus i}) \sqrt{x_i^2 - x_i^4}$. By taking $u_i = \sqrt{X_{ii}}$ and $w_i = \sqrt{u_i^2 - X_{ii}^2}$, the constraints (17b) are satisfied and we get (17d). □

Notice that all constraints used to define $\mathcal{P}_k$ are convex since they are either of linear type or of second-order-cone type. Observe that variable u_i is supposed to represent $\sqrt{X_{ii}}$ while w_i represents $\sqrt{X_{ii} - X_{ii}^2}$ when $X \in \mathcal{V}_k$. Since this is non-convex, constraints (17b) are considered to express a relaxed version of $u_i = \sqrt{X_{ii}}$ and $w_i = \sqrt{X_{ii} - X_{ii}^2}$.

Let us consider the constraints $\|X\|_1 \leq k$ introduced by D'Aspremont et al. (2007) and recalled in Section 2 to get (4). First observe that they can be generalized into constraints

$$\sum_{i,j\in\{1,\ldots,n\}} b_i b_j |X_{ij}| \leq \left(f_k\left(b\right)\right)^2, \forall b \in \mathbb{R}_+^n. \tag{18}$$

The validity of (18) for any matrix $X = xx^\top \in \mathcal{V}_k$ is obvious, since then $\sum_{i,j\in\{1,\ldots,n\}} b_i b_j |X_{ij}| = \left(\sum_{i=1}^n b_i |x_i|\right)^2$. And using the same argument as above, $\sum_{i=1}^n b_i |x_i| \leq f_k(b)$. Observe that by taking $b_i = 1$ for each i, we get the constraint $\|X\|_1 \leq k$. The next lemma shows that constraints (18) are already implied by those of $\mathcal{P}_k$.

Lemma 2. *Constraints* (18) *are satisfied for each matrix X belonging to $\mathcal{P}_k$.*

Proof. Let us consider inequality $\sum_{j\in\{1,\ldots,n\}\setminus\{i\}} b_j |X_{ij}| \leq f_{k-1}(b_{\setminus i}) w_i$ for some $i \in \{1, ..., n\}$ and $b \in \mathbb{R}_+^n$. Let us first show that this constraint is stronger than constraint $\sum_{j\in\{1,\ldots,n\}} b_j |X_{ij}| \leq f_k(b) u_i$. Using the definition of f_k, one can write that $\left(f_k\left(b\right)\right)^2 - \left(f_{k-1}\left(b_{\setminus i}\right)\right)^2 - \left(b_i\right)^2 \geq 0$. Multiplying both sides by $\left(f_{k-1}\left(b_{\setminus i}\right)\right)^2$, we get the following inequality:
$\left(f_k\left(b\right)\right)^2 \left(f_{k-1}\left(b_{\setminus i}\right)\right)^2 \geq \left(f_{k-1}\left(b_{\setminus i}\right)\right)^2 \left[\left(b_i\right)^2 + \left(f_{k-1}\left(b_{\setminus i}\right)\right)^2\right]$. Regrouping terms we obtain $\left[(f_k(b))^2 - (f_{k-1}(b_{\setminus i}))^2\right] \left[(b_i)^2 + (f_{k-1}(b_{\setminus i}))^2\right] \geq (b_i)^2 (f_k(b))^2$ leading to

$$2u_i \sqrt{(f_k(b))^2 - (f_{k-1}(b_{\setminus i}))^2} X_{ii} \sqrt{(b_i)^2 + (f_{k-1}(b_{\setminus i}))^2} \geq 2 b_i f_k(b) u_i X_{ii}. \tag{19}$$

The trivial non-negativity constraint of a squared term:

$$\left[u_i \sqrt{(f_k(b))^2 - (f_{k-1}(b_{\setminus i}))^2} - X_{ii} \sqrt{(b_i)^2 + (f_{k-1}(b_{\setminus i}))^2}\right]^2 \geq 0$$

leads to

$$\begin{aligned} & u_i^2 \left[(f_k(b))^2 - (f_{k-1}(b_{\setminus i}))^2\right] + X_{ii}^2 \left[(b_i)^2 + (f_{k-1}(b_{\setminus i}))^2\right] \\ & \geq 2u_i \sqrt{(f_k(b))^2 - (f_{k-1}(b_{\setminus i}))^2} X_{ii} \sqrt{(b_i)^2 + (f_{k-1}(b_{\setminus i}))^2} \end{aligned} \tag{20}$$

With the lower bound on the right-hand side of (20) given by (19), we get that $u_i^2 \left[(f_k(b))^2 - (f_{k-1}(b_{\setminus i}))^2\right] + X_{ii}^2 \left[(b_i)^2 + (f_{k-1}(b_{\setminus i}))^2\right] - 2 b_i f_k(b) u_i X_{ii} \geq$

0. Reorganizing the terms leads to $(f_{k-1}(b_{\setminus i}))^2(u_i^2 - X_{ii}^2) \leq (f_k(b))^2 u_i^2 + b_i^2 X_{ii}^2 - 2b_i f_k(b) u_i X_{ii}$. Taking the square roots of both sides implies that $f_{k-1}(b_{\setminus i})\sqrt{u_i^2 - X_{ii}^2} \leq |f_k(b)u_i - b_i X_{ii}|$. Observe that constraints (17b) imply that $u_i \geq X_{ii}$, while $f_k(b) \geq b_i$ from the definition of f_k. Consequently, $|f_k(b)u_i - b_i X_{ii}| = f_k(b)u_i - b_i X_{ii}$ holds. Using $w_i \leq \sqrt{u_i^2 - X_{ii}^2}$, we deduce that $f_{k-1}(b_{\setminus i})w_i + b_i X_{ii} \leq f_k(b)u_i$ proving that $\sum_{j\in\{1,\ldots,n\}} b_j|X_{ij}| \leq f_k(b)u_i$ is dominated by $\sum_{j\in\{1,\ldots,n\}\setminus\{i\}} b_j|X_{ij}| \leq f_{k-1}(b_{\setminus i})w_i$. To finish the proof, we multiply each inequality $\sum_{j\in\{1,\ldots,n\}} b_j|X_{ij}| \leq f_k(b)u_i$ by b_i and we sum up over i leading to $\sum_{i,j\in\{1,\ldots,n\}} b_i b_j|X_{ij}| \leq f_k(b)\sum_{i=1}^n b_i u_i$. Now using the constraint $\sum_{i=1}^n b_i u_i \leq f_k(b)$, we get the wanted inequality $\sum_{i,j\in\{1,\ldots,n\}} b_i b_j|X_{ij}| \leq (f_k(b))^2$. □

The previous lemma implies that the lower bounds that will be proposed in the rest of the paper are tighter than the bound of Bach et al. (2010) which was shown to be equivalent to (10) (Proposition 2).

Observe that the separation of inequalities (18) does not seem to be easy since it is equivalent to maximize a quadratic form under some convex constraints. However, inequalities (18) are dominated by those defining $\mathcal{P}_k$ which can be separated in polynomial time as stated in next proposition.

Proposition 4. *Inequalities defining $\mathcal{P}_k$ can be separated in polynomial time.*

Proof. Let us start with the separation of the inequalities $\sum_{i=1}^n b_i u_i \leq f_k(b)$ for fixed u. Note that these inequalities are satisfied if and only if $\sum_{i=1}^n u_i b_i \leq 1$, for all $b \in \mathbb{R}_+^n$ such that $f_k(b) \leq 1$. While the maximum violation over all the inequalities $\sum_{i=1}^n b_i u_i \leq f_k(b)$ may be unbounded (if there exists $b \in \mathbb{R}_+^n$ for which $\sum_{i=1}^n u_i b_i > 1$, then multiplying b by an arbitrarily large value one gets another violated inequality with an arbitrarily large violation), this observation shows that the original separation problem is equivalent to the following one with bounded optimal objective value and solution (due to the bounding constraint on the right-hand side):

$$\begin{cases} \max & \sum_{i=1}^n b_i u_i \\ s.t. & f_k(b) \leq 1, b \in \mathbb{R}_+^n. \end{cases} \tag{21}$$

So all the inequalities $\sum_{i=1}^{n} b_i u_i \leq f_k(b)$ are satisfied if and only if the optimal objective of (21) is less than or equal to one.

Observe that condition $f_k(b) \leq 1$ is equivalent to $(f_k(b))^2 \leq 1$. Moreover, one can express $(f_k(b))^2$ as follows:

$$\begin{cases} (f_k(b))^2 = & \max & \sum_{i=1}^{n} b_i^2 z_i \\ & s.t. & \sum_{i=1}^{n} z_i = k \\ & & 0 \leq z_i \leq 1, \forall i \in \{1, ..., n\}. \end{cases} \tag{22}$$

The dual of (22) reads as follows

$$\begin{cases} Z_{D1}^{\star} = & \min & k\gamma + \sum_{i=1}^{n} \alpha_i \\ & s.t. & \gamma + \alpha_i \geq b_i^2, \forall i \in \{1, ..., n\} \\ & & \gamma \in \mathbb{R}, \alpha \in \mathbb{R}_+^n, \end{cases} \tag{23}$$

where variable γ (resp. α_i) is associated to inequality $\sum_{i=1}^{n} z_i = k$ (resp. $z_i \leq 1$) in (22). Using the fact that for any optimal solution $(\gamma^{\star}, \alpha^{\star})$ of (23) we have $\gamma^{\star} = \max_{i \in \{1,2,...,n\}} \left(b_i^2 - \alpha_i^{\star}\right)$, we deduce

$$Z_{D1}^{\star} = \min_{\alpha \in \mathbb{R}_+^n} \left(k \cdot \max_{i \in \{1,2,...,n\}} \left(b_i^2 - \alpha_i\right) + \sum_{j=1}^{n} \alpha_j \right)$$

From the last equation, we deduce that the constraint $(f_k(b))^2 \leq 1$ can be expressed by the following set of convex constraints:

$$\begin{array}{rll} k b_i^2 - 1 - (k-1)\alpha_i + \sum_{j \in \{1,...,n\} \setminus \{i\}} \alpha_j & \leq 0, \forall i \in \{1, ..., n\} \\ \alpha_i & \geq 0, \forall i \in \{1, ..., n\}. \end{array}$$

This immediately implies that separation can be done by solving the convex problem (in fact a second-order-cone program, abbreviated by SOCP):

$$\begin{cases} \max & \sum_{i=1}^{n} b_i u_i & \\ s.t. & k b_i^2 - 1 - (k-1)\alpha_i + \sum_{j \in \{1,...,n\} \setminus \{i\}} \alpha_j \leq 0, & \forall i \in \{1, ..., n\} \\ & \alpha_i \geq 0, b_i \in \mathbb{R}_+, & \forall i \in \{1, ..., n\}. \end{cases} \tag{24}$$

If the objective value is larger than 1, we get a violated inequality.

Inequalities $\sum_{j \in \{1,...,n\} \setminus \{i\}} b_j |X_{ij}| \leq f_{k-1}(b_{\setminus i}) w_i$ can be separated in a similar way

by solving for each i the following problem:

$$\begin{cases} \max & \sum\limits_{j\in\{1,\dots,n\}\setminus\{i\}} b_j|X_{ij}| \\ s.t. & (k-1)b_j^2 - 1 - (k-2)\alpha_j + \sum_{l\in\{1,\dots,n\}\setminus\{i,j\}} \alpha_l \leq 0, \quad \forall j \in \{1,...,n\}\setminus\{i\} \\ & \alpha_j \geq 0, b_j \in \mathbb{R}_+, \qquad\qquad \forall j \in \{1,...,n\}\setminus\{i\}. \end{cases} \tag{25}$$

If the obtained objective value is greater than w_i, we get a violated inequality.

□

The previous result suggests a simple iterative cutting-plane algorithm to get a lower bound. However, at each iteration, one has to solve SOCP problems. An implementation of this algorithm has shown that this approach is not efficient.

To avoid adding cuts, we consider the dual problems of (24) and (25). Let us relax the constraints of (24) in a Lagrangean way. This leads to:

$$\min_{\substack{\phi\in\mathbb{R}^n_+ \\ \lambda\in\mathbb{R}^n_+}} \max_{\substack{b\in\mathbb{R}^n_+ \\ \alpha\in\mathbb{R}^n}} \sum_{i=1}^n b_i u_i - \sum_{i=1}^n \lambda_i \left(kb_i^2 - 1 - (k-1)\alpha_i + \sum_{j\in\{1,\dots,n\}\setminus\{i\}} \alpha_j \right) + \sum_{i=1}^n \phi_i\alpha_i.$$

Reorganizing terms we get

$$\min_{\substack{\phi\in\mathbb{R}^n_+ \\ \lambda\in\mathbb{R}^n_+}} \max_{\substack{b\in\mathbb{R}^n_+ \\ \alpha\in\mathbb{R}^n}} \sum_{i=1}^n \alpha_i \left(\phi_i + (k-1)\lambda_i - \sum_{j\in\{1,\dots,n\}\setminus\{i\}} \lambda_j \right) + \sum_{i=1}^n b_i u_i - k\sum_{i=1}^n \lambda_i b_i^2 + \sum_{i=1}^n \lambda_i.$$

Writing that the maximum if finite, we must have $\phi_i + (k-1)\lambda_i - \sum_{j\in\{1,\dots,n\}\setminus\{i\}} \lambda_j = 0$, for all $i \in \{1,...,n\}$. So the inner maximization problem reduces to

$$\begin{cases} \max & \sum_{i=1}^n b_i u_i - k\sum_{i=1}^n \lambda_i b_i^2 + \sum_{i=1}^n \lambda_i \\ s.t. & (k-1)\lambda_i - \sum_{j\in\{1,\dots,n\}\setminus\{i\}} \lambda_j \leq 0, \forall i \in \{1,...,n\} \\ & \lambda \in \mathbb{R}^n_+, b \in \mathbb{R}^n_+. \end{cases} \tag{26}$$

Maximizing the quadratic function leads to $b_i = \frac{u_i}{2k\lambda_i}$. The dual problem of (24) becomes:

$$\begin{cases} \min & \sum_{i=1}^n \left(\frac{u_i^2}{4k\lambda_i} + \lambda_i \right) \\ s.t. & (k-1)\lambda_i - \sum_{j\in\{1,\dots,n\}\setminus\{i\}} \lambda_j \leq 0, \forall i \in \{1,...,n\} \\ & \lambda \in \mathbb{R}^n_+. \end{cases} \tag{27}$$

Since (24) satisfies Slater's conditions and is bounded, strong duality holds. Requiring that the optimal objective value in (24) is upper bounded by 1 is of course equivalent to saying that (27) has a feasible solution whose objective value is less than 1. So (17c) is equivalent to the following system of constraints:

$$\begin{cases} \sum_{i=1}^{n} \left(\frac{\beta_i}{4k} + \lambda_i \right) \leq 1 & \\ (k-1)\lambda_i - \sum_{j \in \{1,\dots,n\} \setminus \{i\}} \lambda_j \leq 0, & \forall i \in \{1,\dots,n\} \\ \beta_i \geq \frac{u_i^2}{\lambda_i}, & \forall i \in \{1,\dots,n\} \\ \lambda \in \mathbb{R}_+^n, \beta \in \mathbb{R}_+^n. & \end{cases} \tag{28}$$

where we introduced the variable β^i to represent any value larger than or equal to $\frac{u_i^2}{\lambda_i}$ (and linearize in this way the objective in (27)).

Proceeding in a similar way using formulation (25), we get that (17d) is equivalent to the following system of constraints:

$$\begin{cases} \sum\limits_{j \in \{1,\dots,n\} \setminus \{i\}} \left(\frac{\beta_j^i}{4} + (k-1)\lambda_j^i \right) \leq (k-1)w_i, & \forall i \in \{1,\dots,n\} \\ (k-2)\lambda_j^i - \sum\limits_{l \in \{1,\dots,n\} \setminus \{i,j\}} \lambda_l^i \leq 0, & \forall i,j \in \{1,\dots,n\}, j \neq i \\ \beta_j^i \geq \frac{X_{ij}^2}{\lambda_j^i} & \forall i,j \in \{1,\dots,n\}, j \neq i \\ \lambda^i \in \mathbb{R}_+^n, \beta^i \in \mathbb{R}_+^n. & \end{cases} \tag{29}$$

where λ_j^i are the dual variables associated with the first set of inequalities in (25) and β_j^i represents any value larger than or equal to $\frac{X_{ij}^2}{\lambda_j^i}$ (so as to make the first inequality linear).

Gathering the results just obtained leads to another way of expressing $\mathcal{P}_k$.

$$\mathcal{P}_k = \left\{ \begin{array}{lll} X \in \mathcal{S}_+^n & \text{such that:} & \\ & Tr(X) = 1 & \\ & \exists u \in \mathbb{R}_+^n, w \in \mathbb{R}_+^n, \lambda \in \mathbb{R}_+^n, \beta \in \mathbb{R}_+^n, \lambda^i \in \mathbb{R}_+^{n-1}, \beta^i \in \mathbb{R}_+^{n-1} & \forall i \in \{1,...,n\} \\ & u_i^2 \leq X_{ii}, w_i^2 \leq (u_i - X_{ii})(u_i + X_{ii}) & \forall i \in \{1,...,n\} \\ & u_i^2 \leq \lambda_i \beta_i & \forall i \in \{1,...,n\} \\ & \sum_{i=1}^n \left(\frac{\beta_i}{4k} + \lambda_i \right) \leq 1 & \\ & (k-1)\lambda_i - \sum\limits_{j \in \{1,...,n\} \setminus \{i\}} \lambda_j \leq 0 & \forall i \in \{1,...,n\} \\ & X_{ij}^2 \leq \lambda_j^i \beta_j^i & \forall i,j \in \{1,...,n\}, j \neq i \\ & \sum\limits_{j \in \{1,...,n\} \setminus \{i\}} \left(\frac{\beta_j^i}{4} + (k-1)\lambda_j^i \right) \leq (k-1) w_i & \forall i \in \{1,...,n\} \\ & (k-2)\lambda_j^i - \sum\limits_{l \in \{1,...,n\} \setminus \{i,j\}} \lambda_l^i \leq 0 & \forall i,j \in \{1,...,n\}, j \neq i. \end{array} \right. \tag{30}$$

We now introduce some inequalities that may be added to strengthen this relaxation. The main idea is to try to use again the fact that for any $X \in \mathcal{V}_k$, by taking $u_i = \sqrt{X_{ii}}$ and $w_i = \sqrt{X_{ii} - X_{ii}^2}$, there exists a set of values for vectors λ, β, λ^i and β^i leading to a feasible solution of $\mathcal{P}_k$.

We will prove later that the following constraints can be added in (30) to get a stronger relaxation of $\text{conv}\,(\mathcal{V}_k)$: $\sum\limits_{i \in \{1,...,n\}} \lambda_i = 1/2$, $\sum_{i=1}^n \frac{\beta_i}{4k} = 1/2$, $\sum\limits_{j \in \{1,...,n\} \setminus \{i\}} \lambda_j^i = w_i/2$, $\sum\limits_{j \in \{1,...,n\} \setminus \{i\}} \beta_j^i = 2(k-1)w_i$, in addition to inequalities $2k\lambda_i X_{ii} \geq u_i^2$. To ease presentation we will replace variables λ_i by $z_i = 2k\lambda_i$.

This leads to the formulation (31) defining the set $\tilde{\mathcal{P}}_k$.

$$\tilde{\mathcal{P}}_k = \left\{ \begin{array}{lll} X \in \mathcal{S}_+^n & \text{such that:} & \\ & Tr(X) = 1 & \\ & \exists u \in \mathbb{R}_+^n, w \in \mathbb{R}_+^n, z \in \mathbb{R}_+^n, \lambda^i \in \mathbb{R}_+^{n-1}, \beta^i \in \mathbb{R}_+^{n-1} & \forall i \in \{1,...,n\} \\ & u_i^2 \leq z_i X_{ii} & \forall i \in \{1,...,n\} \\ & w_i^2 \leq (u_i - X_{ii})(u_i + X_{ii}) & \forall i \in \{1,...,n\} \\ & \sum_{i=1}^n z_i = k & \\ & z_i \leq 1 & \forall i \in \{1,...,n\} \\ & X_{ij}^2 \leq \lambda_j^i \beta_j^i & \forall i,j \in \{1,...,n\}, j \neq i \\ & \sum\limits_{j \in \{1,...,n\} \setminus \{i\}} \lambda_j^i = \frac{w_i}{2} & \forall i \in \{1,...,n\} \\ & \sum\limits_{j \in \{1,...,n\} \setminus \{i\}} \beta_j^i = 2(k-1) w_i & \forall i \in \{1,...,n\} \\ & 2(k-1)\lambda_j^i \leq w_i & \forall i,j \in \{1,...,n\}, j \neq i. \end{array} \right. \tag{31}$$

Proposition 5. *conv*$(\mathcal{V}_k) \subset \tilde{\mathcal{P}}_k \subset \mathcal{P}_k$.

Proof. Since some constraints have been added to (30) to get (31), $\tilde{\mathcal{P}}_k \subset \mathcal{P}_k$ obviously holds. Let us then focus on the other inclusion.

Firstly, note that from the way the formulation (28) was obtained, for any $X \in \mathcal{V}_k$ there exists $(u, \lambda, \beta) \in \mathbb{R}_+^n \times \mathbb{R}_+^n \times \mathbb{R}_+^n$ such that $\beta_i = \frac{u_i^2}{\lambda_i}$, $\sum_{i=1}^n \left(\frac{u_i^2}{4k\lambda_i} + \lambda_i \right) \leq 1$, and $(k-1)\lambda_i - \sum\limits_{j \in \{1,...,n\} \setminus \{i\}} \lambda_j \leq 0, \forall i \in \{1,2,...,n\}$. The last inequality implies that $\lambda_i \leq \frac{\sum_{j \in \{1,...,n\}} \lambda_j}{k}$. And using the latter with $\sum_{i=1}^n \left(\frac{u_i^2}{4k\lambda_i} + \lambda_i \right) \leq 1$ leads to $\sum_{i=1}^n \left(\frac{u_i^2}{4\sum_{j \in \{1,...,n\}} \lambda_j} + \lambda_i \right) \leq 1$, which is equivalent to $\frac{1}{4\sum_{j \in \{1,...,n\}} \lambda_j} \sum_{i=1}^n u_i^2 + \sum_{j \in \{1,...,n\}} \lambda_j \leq 1$.

For each matrix $X = xx^\top$ with $\|x\|_2 = 1$ and $\|x\|_0 \leq k$, it is enough to impose feasibility of $\mathcal{P}_k$ when $u_i = \sqrt{X_{ii}}$ and $w_i = \sqrt{X_{ii} - X_{ii}^2}$. This implies that $\sum_{i=1}^n u_i^2 = 1$. Then we get $\frac{1}{4\sum_{j \in \{1,...,n\}} \lambda_j} + \sum_{j \in \{1,...,n\}} \lambda_j \leq 1$, which is however possible only if $\sum_{j \in \{1,...,n\}} \lambda_j = 1/2$, and consequently $\sum_{i=1}^n \beta_i = 2k$. The equation $\sum_{j \in \{1,...,n\}} \lambda_j = 1/2$ together with the inequality $\lambda_i \leq \frac{\sum_{j \in \{1,...,n\}} \lambda_j}{k}$ also imply that we can assume that $\lambda_i \leq \frac{1}{2k}$. In fact, considering

how we get $\frac{1}{4\sum_{j\in\{1,\dots,n\}}\lambda_j} + \sum_{j\in\{1,\dots,n\}}\lambda_j \leq 1$, the only way to get equality is to impose that $\beta_i = 2ku_i^2 = 2kX_{ii}$ for all i, $\lambda_i = \frac{1}{2k}$ when $X_{ii} \neq 0$. Constraints $u_i^2 \leq \lambda_i\beta_i$ can then be replaced by $u_i^2 \leq 2k\lambda_i X_{ii}$. By eliminating variables β_i and replacing λ_i by $z_i = 2k\lambda_i$, the set of constraints $u_i^2 \leq X_{ii}$, $u_i^2 \leq \lambda_i\beta_i$, $\sum_{i=1}^n \left(\frac{\beta_i}{4k} + \lambda_i\right) \leq 1$, $(k-1)\lambda_i - \sum_{j\in\{1,\dots,n\}\setminus\{i\}} \lambda_j \leq 0$ can be simply replaced by $u_i^2 \leq z_i X_{ii}$, $\sum_{i=1}^n z_i = k$ and $0 \leq z_i \leq 1$. Using the fact that $\sum_{j\in\{1,\dots,n\}\setminus\{i\}} X_{ij}^2 = w_i^2$, the same kind of arguments allows to prove the validity of $\sum_{j\in\{1,\dots,n\}\setminus\{i\}} \lambda_j^i = \frac{w_i}{2}$, $\sum_{j\in\{1,\dots,n\}\setminus\{i\}} \beta_j^i = 2(k-1)w_i$ and $2(k-1)\lambda_j^i \leq w_i$. □

Let us now give a formulation to determine a lower bound using the relaxation (12) where $\mathcal{R}_k$ is equal to cone($\tilde{\mathcal{P}}_k$). Remember that requiring $X \in$ cone($\tilde{\mathcal{P}}_k$) is equivalent to impose that $X \in Tr(X)\cdot\tilde{\mathcal{P}}_k$. Using (31) instead of $X \in \mathcal{S}_+^n \cap \mathcal{R}_k$ in formulation (12) leads to (32).

$$
\left\{
\begin{array}{lll}
\min & ||y||_2^2 - Tr(XA^\top yy^\top A) & \\
s.t. & X \in \mathcal{S}_+^n & \\
& Tr(XA^\top A) = 1 & \\
& u_i^2 \leq z_i X_{ii} & \forall i \in \{1,...,n\} \\
& w_i^2 \leq (u_i - X_{ii})(u_i + X_{ii}) & \forall i \in \{1,...,n\} \\
& \sum_{i=1}^n z_i = k \cdot Tr(X) & \\
& z_i \leq Tr(X) & \forall i \in \{1,...,n\} \\
& X_{ij}^2 \leq \lambda_j^i \beta_j^i & \forall i,j \in \{1,...,n\}, j \neq i \\
& \sum_{j\in\{1,\dots,n\}\setminus\{i\}} \lambda_j^i = \frac{w_i}{2} & \forall i \in \{1,...,n\} \\
& \sum_{j\in\{1,\dots,n\}\setminus\{i\}} \beta_j^i = 2(k-1)w_i & \forall i \in \{1,...,n\} \\
& 2(k-1)\lambda_j^i \leq w_i & \forall i,j \in \{1,...,n\}, j \neq i \\
& u \in \mathbb{R}_+^n, w \in \mathbb{R}_+^n, z \in \mathbb{R}_+^n, \lambda^i \in \mathbb{R}_+^{n-1}, \beta^i \in \mathbb{R}_+^{n-1} & \forall i \in \{1,...,n\}.
\end{array}
\right. \tag{32}
$$

To close this section, we show that by adding constraints $\frac{z_i}{Tr(X)} \in \{0,1\}$ to (32), we get $Z^\star$ as an optimal objective value (i.e., the optimal objective value of SSP).

Proposition 6. *Imposing constraints* $\frac{z_i}{Tr(X)} \in \{0,1\}$ *in* (32) *leads to an optimal solution of* SSP.

Proof. From $\sum_{i=1}^{n} z_i = k \cdot Tr(X)$ and $\frac{z_i}{Tr(X)} \in \{0,1\}$ we get that $n-k$ variables among the z_i are equal to 0. If $z_i = 0$, then from $u_i^2 \leq z_i X_{ii}$ we get that $u_i = 0$. As a consequence, from constraint $w_i^2 \leq (u_i - X_{ii})(u_i + X_{ii})$ we deduce that $w_i = 0$. Equality $\sum_{j \in \{1,\ldots,n\} \setminus \{i\}} \lambda_j^i = \frac{w_i}{2}$ implies that all λ_j^i are equal to 0. Using $X_{ij}^2 \leq \lambda_j^i \beta_j^i$ we get that $X_{ij} = 0$ for any j. In other words, the matrix X will have $n-k$ zero columns. X can then be written as a sum of at most k rank-one matrices: $X = \sum_{i=1}^{k} v_i {v_i}^\top$. Since there are at most k non-zero terms on the diagonal of X, each vector v_i in the decomposition of X will also have at most k non-zero components.

Let us now remember that an optimal solution of (32) is an optimal solution of (13) when $\mathcal{R}_k = \text{cone}(\tilde{\mathcal{P}}_k)$. Using that $\frac{Tr\left(XA^\top yy^\top A\right)}{Tr(XA^\top A)} \leq \max_i \frac{Tr\left(v_i {v_i}^\top A^\top yy^\top A\right)}{Tr(v_i {v_i}^\top A^\top A)}$ and $v_i {v_i}^\top \in \text{cone}(\mathcal{V}_k)$, we deduce that there exists an index i such that $\frac{Tr\left(XA^\top yy^\top A\right)}{Tr(XA^\top A)} = \frac{Tr\left(v_i {v_i}^\top A^\top yy^\top A\right)}{Tr(v_i {v_i}^\top A^\top A)}$ and $\|v_i\|_0 \leq k$ ending the proof. □

5. Upper bounds

Starting from (32), a simple upper bound can be obtained as follows. Given a solution of (32), we compute the boolean vector $x \in \{0,1\}^n$ having k non-zero components minimizing the distance with $z/Tr(X)$ (i.e., $\|x - z/Tr(X)\|_2$). This can be done in a greedy way by sorting the components of z in descending order according to z_i. Then we just compute the objective value of the orthogonal projection of y on the vector space spanned by the set of columns A^i for which $x_i = 1$.

When (10) is solved, we can also get an upper bound using the greedy approach. We consider the vector whose components are $X_{ii}/Tr(X)$ and then we compute the closest boolean vector x having k non-zero components. Then we project y on the vector space spanned by the set of columns A^i for which $x_i = 1$, and we compute the objective value of the obtained vector to get an upper bound.

6. Numerical experiments

All formulations have been solved with MOSEK 8.1 solver. Computations are done on a laptop having a 2.3 GHz Intel Core i5 processor and 16 GB of RAM. Notice that in all conducted experiments, the columns of A and the vector y are normalized (i.e., $||A^i||_2 = 1$ and $||y||_2 = 1$). This assumption seems to reduce numerical difficulties. We also tried other solvers (COSMO and SCS). Some preliminary experiments have shown that the results given by MOSEK (in a comparable time) are more robust than those obtained by COSMO and SCS. Furthermore, in addition to the new formulation (32), we also implemented a dual formulation and checked that both formulations give the same bound. However, the computing time presented in the paper will only integrate the solution of the primal problem (32).

6.1. Normally-distributed data

We start reporting the results obtained on randomly generated 'gaussian instances', where both y and the columns of A are normally distributed. The new bounds are compared with those of Bach et al. (2010) and Atamtürk & Gómez (2019) since the latter already compare favorably with previous bounds (see Atamtürk & Gómez (2019); Bach et al. (2010) for details).

To compare our lower bounds with the bounds (2) and (10), we show on Table 1 the lower bounds obtained using (10), (32) and (2) (for $r = 1, 2, 3$). We also give the cpu time in each case. Notice that the cpu times shown in the tables include the total time needed to load the problem, transform it into the right conic model and solve it. The values reported in each row of Table 1 correspond to averaged results over 20 instances. One can see that the bound from (32) seems to be higher than the one obtained from (2) (with $r = 2$). However, this does not hold for all instances. Observe that the computing times related to (32) is comparable to the one related to (2) (with $r = 2$). The bounds obtained when $r = 3$ can be slightly better than the new bound for large values of k. Solving (2) with $r = 3$ is however much more time consuming.

Sizes			(10)		(32)		(2): $r=1$		(2): $r=2$		(2): $r=3$	
m	n	k	Low	Time(s)	Low	Time(s)	Low	Time(s)	Low	Time(s)	Low	Time(s)
100	40	5	0.7900	0.40	0.8026	1.05	0.7733	0.45	0.7992	1.33	0.8012	29.22
100	40	15	0.6403	0.45	0.6582	1.21	0.6503	0.43	0.6587	1.32	0.6593	27.58
100	40	25	0.5830	0.47	0.5951	1.22	0.5946	0.43	0.5960	1.36	0.5961	29.23
100	60	5	0.7383	1.59	0.7551	4.49	0.6542	1.86	0.7291	5.31	0.7471	144.23
100	60	15	0.5570	2.27	0.5750	4.23	0.5248	1.75	0.5664	4.98	0.5731	143.42
100	60	25	0.4569	2.35	0.4744	4.44	0.4645	1.73	0.4792	5.34	0.4813	150.10
100	80	5	0.7228	5.97	0.7375	15.72	0.4886	7.09	0.6317	17.95	0.6873	510.85
100	80	15	0.4723	8.18	0.4857	13.53	0.3508	6.69	0.4339	17.19	0.4600	482.22
100	80	25	0.2856	7.98	0.2977	15.40	0.2558	6.51	0.3003	17.21	0.3089	486.19

Table 1: Normally-distributed data: $n \in \{40, 60, 80\}$

Larger values of n are considered in Table 2. The $r = 3$ variant is not considered in Table 2 since its computing time becomes quite large (as already shown in Table 1). Table 2 contains also the upper bounds obtained as described in Section 5. The upper bounds related to (2) are obtained by the procedure described in Atamtürk & Gómez (2019). Each row in Table 2 gives the values of the parameters m, n and k, and the values (averaged over 20 instances) for the lower and upper bounds from the solution of (32), (10), and (2) the gap defined by 100·(upper bound - lower bound)/(lower bound) and the cpu time (in seconds). For the case when for some fixed triplet (m, n, k), lower bounds less than $< 10^{-4}$ have been obtained for some of the instances, we indicate in parentheses in the field 'Gap' the number of instances for which the lower bound was $> 10^{-4}$. The given gap is the average of the gaps on this set of instances.

Observe again that the computing time mainly depends on n. Problems with $n = 150$ are solved in around 6 minutes.

The gaps generally decrease when the ratio m/n increases. The gaps related to the new bounds (from (32)) are sensitively smaller than those provided by (10) and also tend to improve over (2) for $r = 1, 2$.

To better understand the dependence of the gap on k and m/n, we show in Figures 1 and 2 the average values of the gap (related to the new bound (32)) for $n = 50$ and $n = 30$ (over 100 random gaussian instances). Different values of m are considered: 75, 100, 200 and 300. Figures 1 and 2 clearly confirm that the gap decreases when m/n increases. We also see that the gap seems to be

Sizes			(10)				(32)				(2): $r = 1$				(2): $r = 2$			
m	n	k	Low	Up	Gap(%)	Time(s)	Low	Up	Gap (%)	Time(s)	Low	Up	Gap (%)	Time(s)	Low	Up	Gap (%)	Time(s)
75	100	5	0.6348	0.6711	5.78	20.10	0.6576	0.6607	0.47	57.47	0.0000	0.8402	- (0)	22.22	0.0000	0.8385	- (0)	45.08
75	100	10	0.3697	0.5501	48.90	21.28	0.3811	0.5262	38.07	36.04	0.0000	0.7198	- (0)	20.49	0.0000	0.7224	- (0)	43.22
75	100	15	0.0000	0.7723	- (0)	20.59	0.0000	0.8028	- (0)	40.84	0.0000	0.5945	- (0)	18.81	0.0000	0.6020	- (0)	39.26
300	100	5	0.9055	0.9104	0.54	18.25	0.9099	0.9100	0.01	35.81	0.8947	0.9108	1.79	26.18	0.9083	0.9103	0.21	56.41
300	100	15	0.8013	0.8163	1.88	22.63	0.8134	0.8151	0.20	51.78	0.7951	0.8178	2.86	23.66	0.8107	0.8159	0.64	52.75
300	100	25	0.7388	0.7587	2.73	22.69	0.7536	0.7576	0.53	45.20	0.7428	0.7589	2.17	24.23	0.7532	0.7576	0.58	50.91
500	100	5	0.9437	0.9465	0.29	16.62	0.9463	0.9463	0.00	37.45	0.9429	0.9466	0.39	27.87	0.9462	0.9464	0.02	68.32
500	100	15	0.8759	0.8842	0.94	22.45	0.8830	0.8835	0.05	50.64	0.8789	0.8839	0.56	24.37	0.8828	0.8835	0.08	60.76
500	100	25	0.8477	0.8594	1.38	23.36	0.8576	0.8584	0.09	54.05	0.8548	0.8589	0.48	24.06	0.8575	0.8585	0.10	58.21
300	150	5	0.8929	0.8993	0.73	154.23	0.8984	0.8985	0.01	326.74	0.8470	0.9019	6.50	179.49	0.8852	0.9006	1.73	463.90
300	150	15	0.7756	0.7917	2.08	180.70	0.7866	0.7905	0.49	421.66	0.7212	0.8008	11.03	172.41	0.7654	0.7941	3.73	404.75
300	150	25	0.6797	0.7086	4.26	225.99	0.6927	0.7060	1.91	350.22	0.6461	0.7142	10.58	172.99	0.6807	0.7076	3.96	385.59

Table 2: Results for normally-distributed data: $n \in \{100, 150\}$

larger for k between $\frac{n}{4}$ and $\frac{3n}{4}$ than on the rest of the interval $[1, n]$. This does not hurt intuition since there is some combinatorial explosion between $\frac{n}{4}$ and $\frac{3n}{4}$.

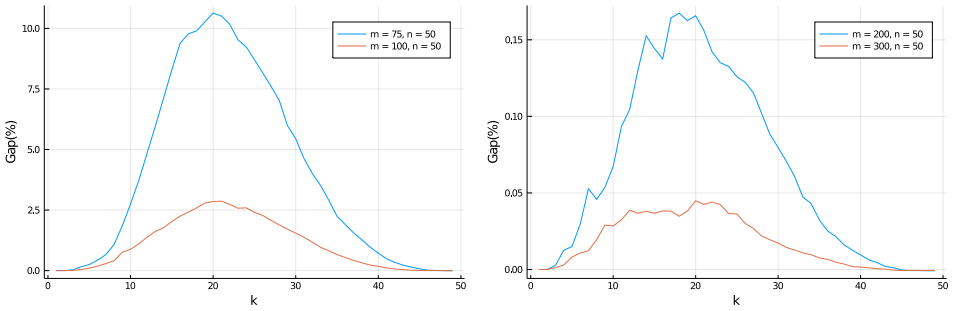

Figure 1: Optimality gap related to bound (32) (average values over 100 instances) as a function of k for a fixed value of $n = 50$ and different values of m: 75, 100, 200 and 300

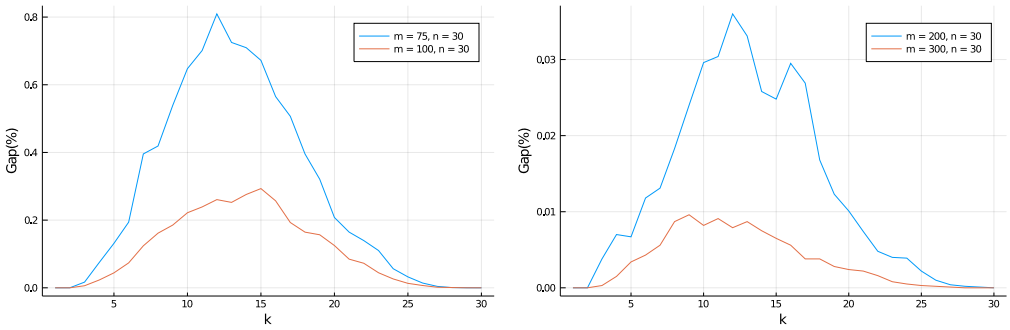

Figure 2: Optimality gap related to bound (32) (average values over 100 instances) as a function of k for a fixed value of $n = 30$ and different values of m: 75, 100, 200 and 300

6.2. Further random instances

We consider here two types of random instances: "Gamma-distributed instances" where each component of y and A is distributed according to Gamma distribution $\Gamma(1, 1)$, and "Uniform instances" where components are uniformly distributed in $[-1, 1]$. In all cases the vector y is normalized such that $\|y\|_2 = 1$. To better understand the behavior of the 4 lower bounds : (10), (32) and (2)

with $r \in \{1, 2, 3\}$, Figures 3-5 illustrate the results obtained for the two types of instances described above (Gamma-distributed, and Uniform). The number of columns n is set to 40 while three values of m are considered: a relatively high value ($m = 80$), a value close to n ($m = 45$) and a value smaller than n ($m = 35$). Computing times are about 0.45 seconds for (10) and also for the $r = 1$ variant of (2), 1.2 seconds for (32), 1.3 seconds for the $r = 2$ case, and 30 seconds for $r = 3$. Several observations can be made here. First, even if the values of the lower bounds depend on the instance type, the curves seem to be quite similar (for a given value of n and m). Second, for the large values of m, we see that the bounds (2): $r = 2$ and $r = 3$ are very close to the bound (32). When m is close to n ($m = 45$) and k is small, (32) becomes much better than relaxation (2) with $r = 2$ and also better than (2) with $r = 3$. Observe however, that for larger values of k, (2) for $r = 2$ and $r = 3$ are still slightly better than (32). The situation for $m = 35$ is more clear, the bounds (2) obtained for $r = 2$ or $r = 3$ are not efficient since they are equal to 0 while (32) still gives non-zero bounds for small values of k but also fails for larger values.

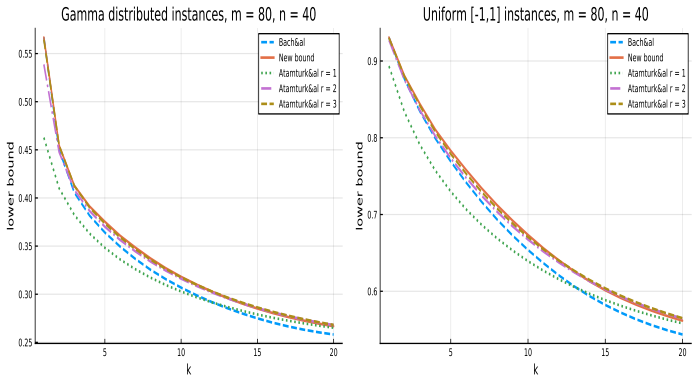

Figure 3: Comparing lower bounds for Gamma-distributed instances and uniformly-distributed instances (average values over 100 instances) for $n = 40$, $m = 80$.

6.3. Some real instances

Four real datasets are considered here. They are chosen in such a way that it is still possible to solve the relaxation (2) with $r = 3$ in a reasonable

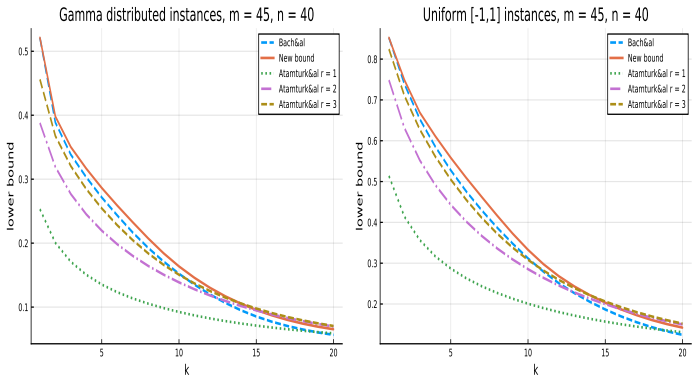

Figure 4: Comparing lower bounds for Gamma-distributed instances and uniformly-distributed instances (average values over 100 instances) for $n = 40$, $m = 45$.

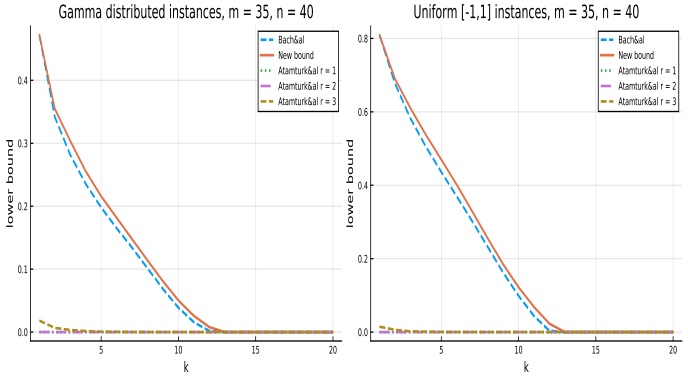

Figure 5: Comparing lower bounds for Gamma-distributed instances and uniformly-distributed instances (average values over 100 instances) for $n = 40$, $m = 35$.

Instance	Number of rows (m)	Number of columns (n)
Auto MPG	392	25
Forest Fires	517	29
Residential Building	100	70
Appliances Energy prediction	500	26

Table 3: Characteristics of real instances

time. "Auto MPG" is used in Miyashiro & Takano (2015) and several other articles. Instance "Forest Fires" comes from Cortez & Morais (2007) and was also used in Miyashiro & Takano (2015). The "Residential Building" instance was defined in Rafiei & Adeli (2016) (we considered the first 100 rows of their file). Finally, the "Appliances Energy Prediction" instance is extracted from Candanedo et al. (2017) by considering the first 500 rows. The four datasets are available on request.

The five lower bounds are given in addition to the upper bound obtained after solving (32). As in our other computations, the columns of A and the vector y are normalized. We can see that the new lower bound is better than the other bounds for small values of k. When k becomes bigger, then the situation depends on the instance. For example, in the "Energy" instance, the new bound stays close to the bound obtained from (2) with $r = 3$. The situation is different for the three other instances where the bound related to $r = 3$ becomes significantly better than the new bound when k increases. Of course, this comes at the cost of a higher computational effort as already shown on Table 1.

6.4. Synthetic instances

Using a similar setup as in (Atamtürk & Gómez, 2019, Section 5.2) (see also Bertsimas et al. (2016); Hastie et al. (2020)) we generate instances as follows. For given dimensions m, n, sparsity s, predictor autocorrelation ρ, and signal-to-noise ratio SNR:

- The rows of the predictor matrix $A \in \mathbb{R}^{m\times n}$ are drawn from i.i.d. distributions $\mathcal{N}_p(0, \Sigma)$, where $\Sigma \in \mathbb{R}^{n\times n}$ such that $\Sigma_{i,j} = \rho^{|i-j|}$.

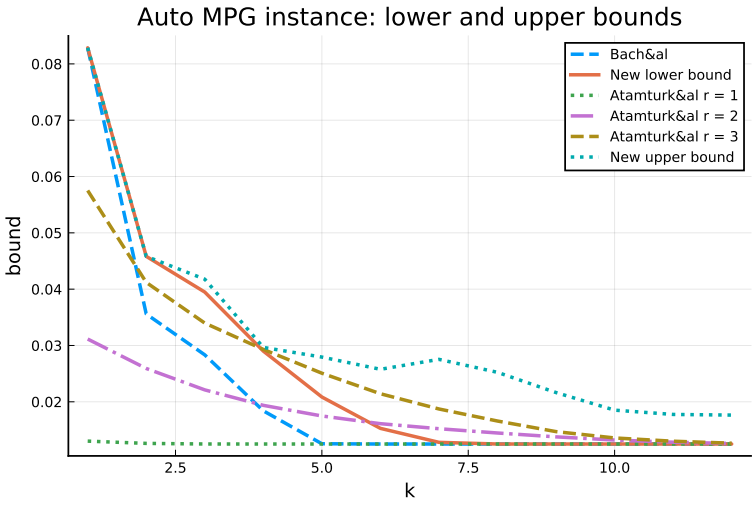

Figure 6: Comparing bounds on the instance 'auto MPG'

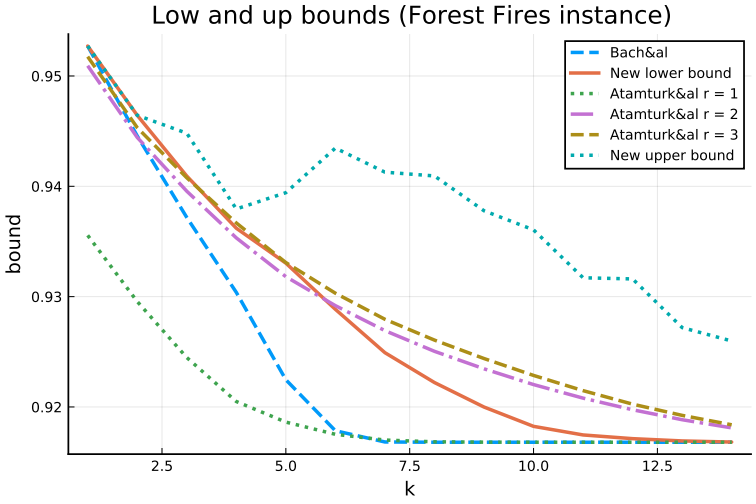

Figure 7: Comparing bounds on the instance 'Forest Fires'

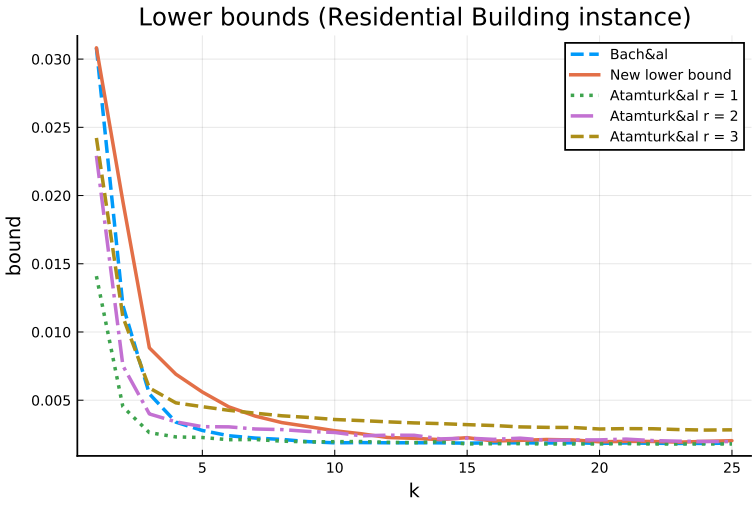

Figure 8: Comparing bounds on the instance 'Residential Building'

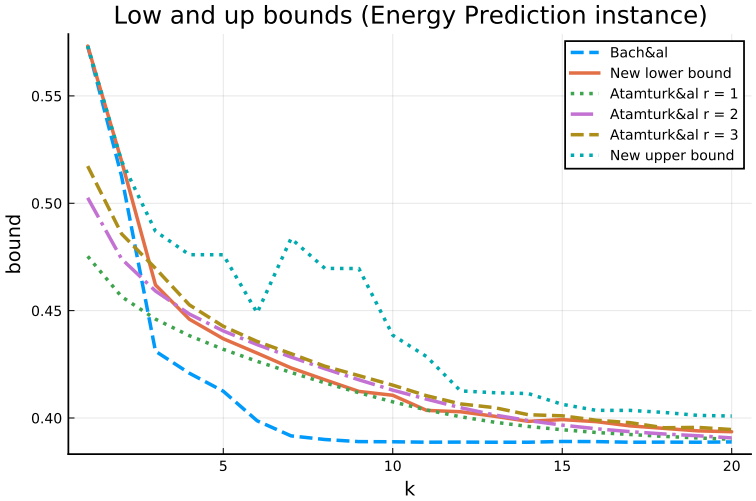

Figure 9: Comparing bounds on the instance 'Appliances Energy Prediction'

- The "right" vector y is equal to $A\beta^0$ where β^0 is a vector of size n whose first s components are equal to one and the rest is equal to zero.
- The response vector y is drawn from $\mathcal{N}_p\left(A\beta^0, \sigma^2 I\right)$, where I stands for the identity matrix and $\sigma^2 = \left(\beta^0\right)^\top \Sigma\beta^0/SNR$.

In our experiments, we used $m = 300$, $n = 50$, SNR$\in \{0.01, 1, 100\}$. ρ varies from 0.1 to 0.9 and $s = k$ with $k \in \{5, 10\}$. The vector y is normalized.
The computing time for (32) and the $r = 2$ variant of (2) is approximately equal to 2 seconds.

The results are displayed in Figure 10 for $s = k = 5$ and in Figure 11 for $s = k = 10$.

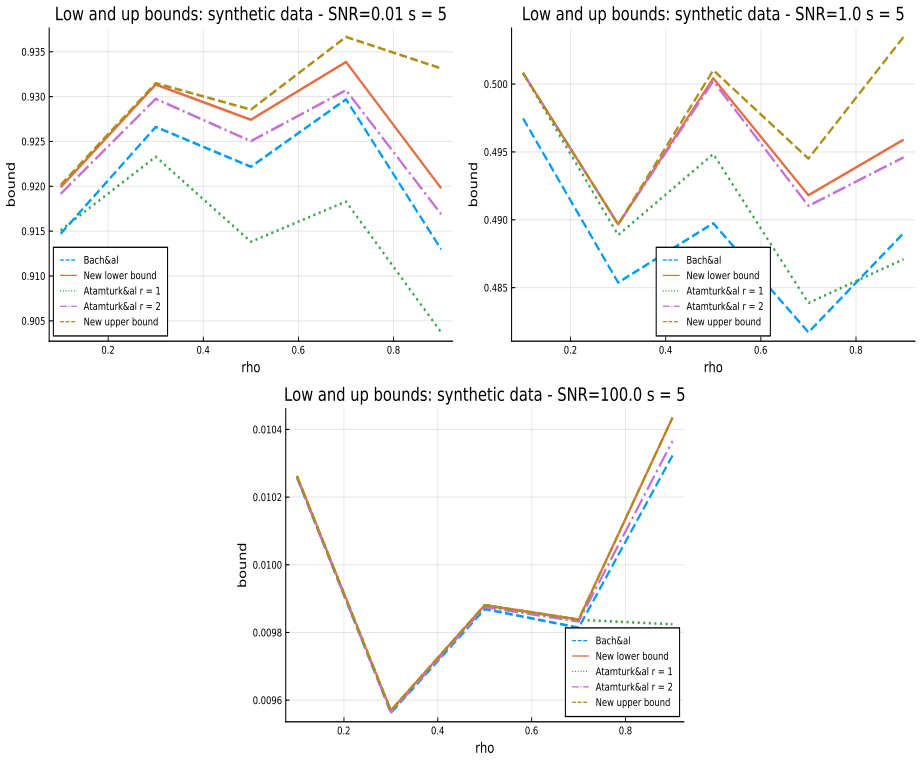

Figure 10: Results on synthetic instances with with $s = k = 5$

The obtained results clearly show that the bound is competitive with other bounds from the literature for this family of instances. In both cases (i.e. $s =$

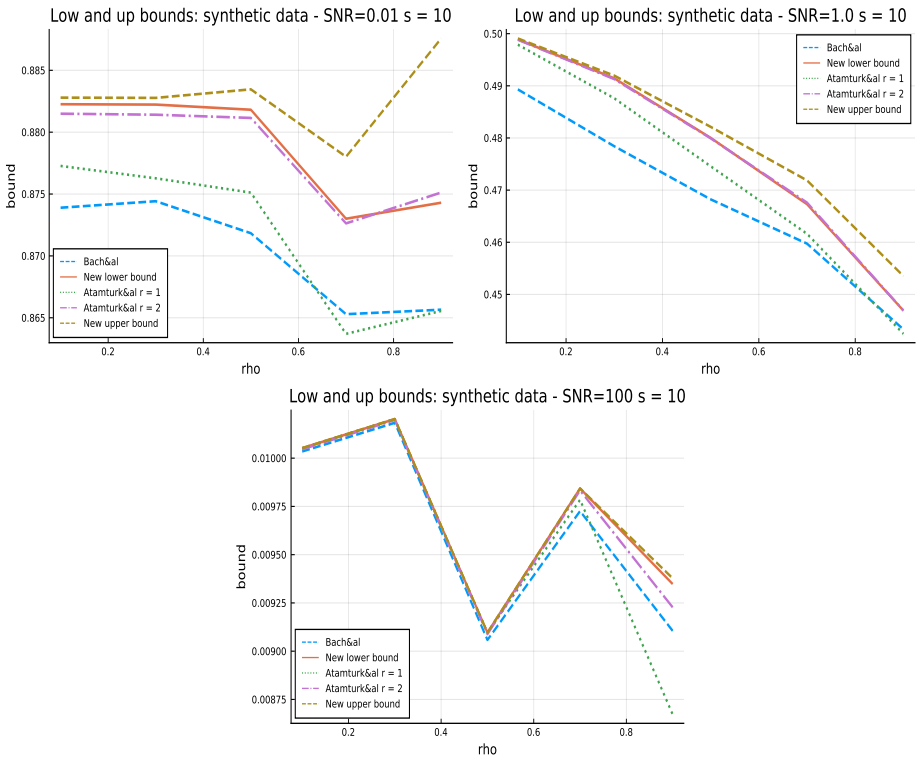

Figure 11: Results on synthetic instances with $s = k = 10$

$k = 5$ and $s = k = 10$) it is interesting to observe that for a low value of the SNR and large value of ρ (≥ 0.6) the proposed bound seems to provide larger improvements over other formulations.

In addition to comparing the values of the bounds stemming from the different formulations, we also considered studying the proportion of correctly identified variables, i.e. the number of nonzero entries in the solution found by the algorithm described in Section 4 and that were used to build y, divided by k. This ratio is called the *support reconstruction ratio* in what follows. In our experiments we set $\rho = 0.5$, $n = 50$ and $s \in \{5, 10\}$. The results (averaged over 50 instances) are displayed as a function of the SNR for $m \in \{50, 100, 150\}$ in Figure 12. As expected, the reconstruction ratio tends to 1 when the SNR increases and the greater m, the easier the reconstruction is. Also, we see that for a low SNR the reconstruction ratio is about 0.1, which comes down to choosing uniformly randomly 5 columns among 50. The convergence of the ratio to the value 1 seems to be more rapid when $s = 5$.

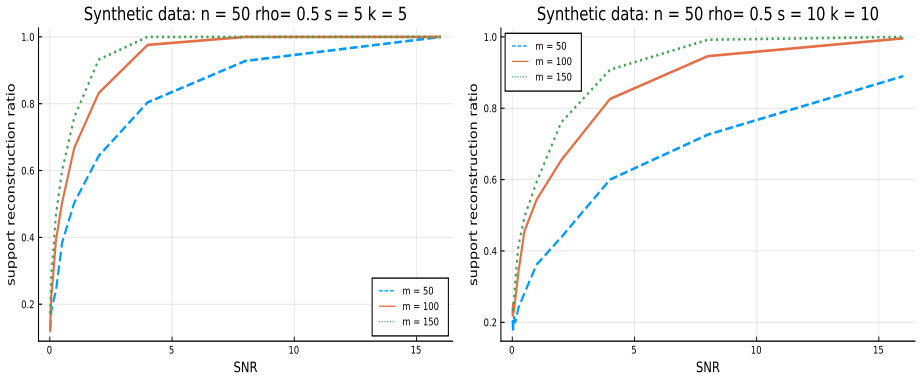

Figure 12: Support reconstruction ratio

7. Conclusion and further research directions

The new proposed bounds are tighter than those of Bach et al. (2010) but more difficult to compute. They are generally tighter than or very close to the second-order bounds of Atamtürk & Gómez (2019) ($r = 2$). When the number

of columns is not too large, the third-order formulation of Atamtürk & Gómez (2019) can be solved leading to a bound that is generally higher than the new bound for large values of k. However, this requires a much greater computational effort.

For large values of the ratio m/n, the obtained gaps are almost equal to 0. When m/n is less than or equal to 1, even if the new bounds seem to be better than previous bounds, they are still not good for large values of k.

Finally, the bounds proposed in this paper are mainly based on an approximation of the cone $\mathcal{C} = \text{cone}(\{xx^\top : x \in \mathbb{R}^n, \|x\|_0 \leq k\})$. We can easily establish that the dual cone $\mathcal{C}^\star$ of $\mathcal{C}$ is the set of symmetric matrices whose principal submatrices of size k are positive semidefinite. This may suggest other approaches for approximating $\mathcal{C}$ by making use of inequalities of the form $Tr(X \cdot Y) \geq 0$, which are valid for $\mathcal{C}$, for all $Y \in \mathcal{C}^\star$. Future work will be dedicated to this other line of research.

Acknowledgments

Both authors warmly thank Andrés Gómez for making his code from Atamtürk & Gómez (2019) available to them. They also wish to thank the anonymous referees for many valuable comments and suggestions that have led to a substantially improved paper.

References

Atamtürk, A., & Gómez, A. (2019). Rank-one convexification for sparse regression. BCOL Research Report 19.01. University of California, Berkeley, USA. `https://arxiv.org/pdf/1901.10334.pdf`.

Bach, F., Ahipaşaoğlu, S. D., & d'Aspremont, A. (2010). Convex relaxations for subset selection. Available at arXiv:1006.3601.

Bertsimas, D., King, A., & Mazumder, R. (2016). Best Subset Selection via a Modern Optimization Lens. *Ann. Statist.*, *44*, 813–852.

Candanedo, L., Feldheim, V., & Deramaix, D. (2017). Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, *140*, 81–97.

Chakraborty, D., Kovvali, N., Wei, J., Papandreou-Suppappola, A., Cochran, D., & Chattopadhyay, A. (2009). Damage Classification Structural Health Monitoring in Bolted Structures Using Time-frequency Techniques. *Journal of Intelligent Material Systems and Structures*, *20*, 1289–1305.

Cortez, P., & Morais, A. (2007). A data mining approach to predict forest fires using meteorological data. In Proceedings of the 13th EPIA 2007: Portuguese Conference on Artificial Intelligence.

Das, A., & Kempe, D. (2011). Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. In Proceedings of the International Conference on Machine Learning (ICML) (2011).

Davis, C., Gerick, F., Hintermair, V., Friedel, C., Fundel, K., Küffner, R., & Zimmer, R. (2006). Reliable gene signatures for microarray classification: Assessment of stability and performance. *Bioinformatics*, *22*, 2356–2363.

Dong, H., Chen, K., & Linderoth, J. (2015). Regularization vs. relaxation: A conic optimization perspective of statistical variable selection. arXiv preprint arXiv:1510.06083.

D'Aspremont, A., El Ghaoui, L., Jordan, M., & Lanckriet, G. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Review*, *49*, 434–448.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression (with discussion). *Annals of Statistics*, *32*, 407–499.

Elad, M., & Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, *15*, 3736–3745.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*, 1348–1360.

Furnival, G., & Wilson, R. (1974). Regressions by Leaps and Bounds. *Technometrics*, *16*, 499–511.

Girosi, F. (1998). An equivalence between sparse approximation and support vector machines. *Neural Comput.*, *10*, 1455–1480.

Gribonval, R., & Bacry, E. (2003). Harmonic decomposition of audio signals with matching pursuit. *IEEE Trans. Signal Process.*, *51*, 101–111.

Hand, D. (1981). Branch and Bound in statistical data analysis. *The Statistician*, (pp. 1–13).

Hastie, T., Tibshirani, R., & Tibshirani, R. (2020). Best subset, forward stepwise, or lasso? Analysis and recommendations based on extensive comparisons. *Statistical Science*, *35*, 579–592.

Mairal, J., Sapiro, G., & Elad, M. (2008). Learning multiscale sparse representations for image and video restoration. *SIAM Multiscale Modeling and Simulation*, *7*, 214–241.

Mazumder, R., Friedman, J., & Hastie, T. (2011). Sparsenet: Coordinate descent with non-convex penalties. *Journal of the American Statistical Association*, *117*, 1125–1138.

Mendels, F., Vandergheynst, P., & Thiran, J. (2006). Matching pursuit-based shape representation and recognition using scale-space. *International Journal of Imaging Systems and Technology*, *16*, 162–180.

Miller, A. (2002). *Subset selection in regression*. CRC Press.

Miyashiro, R., & Takano, Y. (2015). Mixed integer second-order cone programming formulations for variable selection in linear regression. *European Journal of Operational Research*, *247*, 721–731.

Moghaddam, B., Gruber, A., Weiss, Y., & Avidan, S. (2008). Sparse regression as a sparse eigenvalue problem. In *Information Theory and Applications Workshop 2008* (pp. 121–127).

Moghaddam, B., Weiss, Y., & S., A. (2006). Spectral Bounds for Sparse PCA: Exact & Greedy Algorithms. *Neural Information Processing Systems*, *18*.

Narendra, P., & Fukunaga, K. (1977). A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers*, *100*, 917–922.

Natarajan, B. (1995). Sparse approximate solutions to linear systems. *SIAM J. Comput.*, *24*, 227–234.

Neff, R., & Zakhor, A. (1997). Very low bit-rate video coding based on matching pursuit. *IEEE Transactions on Circuits and Systems for Video Technology*, *7*, 158–171.

Rafiei, M., & Adeli, H. (2016). A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management*, *142*, 04015066.

Shen, X., Pan, W., Zhu, Y., & Zhou, H. (2013). On constrained and regularized high-dimensional regression. *Annals of the Institute of Statistical Mathematics*, *65*, 807–832.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, *58*, 267–288.

Tropp, J. (2006). Just Relax: Convex Programming Methods for Identifying Sparse Signals in Noise. *IEEE Transactions on Information Theory*, *52*, 1030–1051.

Wichmann, S., & Kamholz, D. (2008). A stability metric for typological features. *STUF- Language Typology and Universals Sprachtypologie und Universalienforschung*, *61*, 251–262.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, *38*, 894–942.

Appendix A. Justification for the formulation of the dual problem (16)

Problem (15) can be expressed as follows.

$$\begin{cases} \min \quad \|y\|_2^2 + Tr\left(\left[\begin{array}{c|c} 0 & -\frac{1}{2}y^\top A \\ \hline -\frac{1}{2}A^\top y & 0_n \end{array}\right]\left[\begin{array}{c|c} 1 & x^\top \\ \hline x & Q \end{array}\right]\right) \\ s.t. \\ \qquad Q - B^* = 0_n \\ \qquad \left[\begin{array}{c|c} 1 & x^\top \\ \hline x & Q \end{array}\right] \in \mathcal{S}_+^n, x \in \mathbb{R}^n, Q \in \mathcal{S}^n \end{cases} \tag{A.1}$$

where 0_n represents the $n \times n$ matrix with zero entries only.

Let us now associate a matrix dual variable $W \in \mathbb{R}^{n \times n}$ to the first constraint $Q - B^* = 0_n$ and the following matrix dual variable

$$\left[\begin{array}{c|c} \gamma & u^\top \\ \hline u & Z \end{array}\right]$$

to the positive semidefiniteness constraint, where $\gamma \in \mathbb{R}$, $u \in \mathbb{R}^n$ and $Z \in \mathcal{S}^n$. Then, the conic Lagrangian of (A.1) has the following expression.

$$\mathcal{L}(x, Q, W, \gamma, u, Z) = \|y\|^2 - \left(y^\top A + 2u^\top\right) x + Tr\left(W\left(Q - B^*\right)\right) - Tr(ZQ) - \gamma.$$

The corresponding dual function is

$$\begin{aligned} g(W, \gamma, u, Z) &= \min_{x \in \mathbb{R}^n, Q \in \mathcal{S}^n} \mathcal{L}(x, Q, W, \gamma, u, Z) \\ &= \begin{cases} \|y\|_2^2 - Tr\left(ZB^*\right) - \gamma & \text{if } W = Z \text{ and } u = -\frac{1}{2}A^\top y \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

The dual problem then writes

$$\begin{cases} \max \quad \|y\|_2^2 - Tr\left(ZB^*\right) - \gamma \\ s.t. \\ \qquad \left[\begin{array}{c|c} \gamma & -\frac{1}{2}y^\top A \\ \hline -\frac{1}{2}A^\top y & Z \end{array}\right] \in \mathcal{S}_+^n, Z \in \mathbb{S}^n, \gamma \in \mathbb{R}. \end{cases} \tag{A.2}$$

Recall that we assume $y^\top A \neq 0$ (see beginning of the proof of Proposition 3). Together with the positive semidefiniteness constraint in (A.2), this implies $\gamma > 0$, which in turn implies

$$\left[\begin{array}{c|c} \gamma & -\frac{1}{2}y^\top A \\ \hline -\frac{1}{2}A^\top y & Z \end{array}\right] \in \mathcal{S}_+^n \iff Z - \frac{1}{4\gamma}A^\top y y^\top A \in \mathcal{S}_+^n.$$

This directly leads to the formulation (16).