An SDP Relaxation for the Sparse Integer Least Square Problem*

Alberto Del Pia † Dekun Zhou ‡

January 5, 2023

Abstract

In this paper, we study the polynomial approximability or solvability of *sparse integer least square problem* (SILS), which is the NP-hard variant of the least square problem, where we only consider sparse $\{0, \pm 1\}$-vectors. We propose an ℓ_1-based SDP relaxation to SILS, and introduce a randomized algorithm for SILS based on the SDP relaxation. In fact, the proposed randomized algorithm works for a broader class of binary quadratic program with cardinality constraint, where the objective function can be possibly non-convex. Moreover, when the sparsity parameter is fixed, we provide sufficient conditions for our SDP relaxation to solve SILS. The class of data input which guarantee that SDP solves SILS is broad enough to cover many cases in real-world applications, such as privacy preserving identification, and multiuser detection. To show this, we specialize our sufficient conditions to two special cases of SILS with relevant applications: the *feature extraction problem* and the *integer sparse recovery problem*. We show that our SDP relaxation can solve the feature extraction problem with sub-Gaussian data, under some weak conditions on the second moment of the covariance matrix. We also show that our SDP relaxation can solve the integer sparse recovery problem under some conditions that can be satisfied both in high and low coherence settings.

Key words: Semidefinite relaxation, Sparsity, Integer least square problem, ℓ_1 relaxation

1 Introduction

The Integer Least Square problem is a fundamental NP-hard optimization problem which arises from many real-world applications, including communication theory, lattice design, Monte Carlo second-moment estimation, and cryptography. We refer readers to the comprehensive survey [1] and references therein. In the *integer least square problem*, we are given an $n \times d$ matrix M, a d-vector b, and we seek the closest point to b, in the lattice spanned by the columns of M. The ILS problem can be formulated as the following optimization problem:

$$\begin{aligned} \min\ & \frac{1}{n} \|Mx - b\|_2^2 \\ \text{s.t. } & x \in \mathbb{Z}^d. \end{aligned} \tag{ILS}$$

In many scenarios, one is only interested in sparse solutions to (ILS), i.e., vectors x with a large fraction of entries equal to zero. This is primarily motivated by the need to recover a sparse signal [31, 53], or the need to improve the efficiency of data structure representation [39].

*The authors are partially funded by ONR grant N00014-19-1-2322. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Office of Naval Research.

†Department of Industrial and Systems Engineering & Wisconsin Institute for Discovery, University of Wisconsin-Madison. E-mail: delpia@wisc.edu

‡Department of Industrial and Systems Engineering, University of Wisconsin-Madison. E-mail: dzhou44@wisc.edu

Applications include cyber security [31], array signal processing [53], and sparse code multiple access [13]. In this sparse setting, the feasible region is often further restricted to the set $\{0, \pm 1\}^d$. Applications can be found in multiuser detection, where user terminals transmit binary symbols in a code-division multiple access (CDMA) system [57], in sensor networks, where sensors with low duty cycles are either silent (transmit 0) or active (transmit ± 1) [44], and in privacy preserving identification, where a sparse vector in $\{0, \pm 1\}$ is employed to approximate the 'content' of feature data [39]. In this paper, we study this version of (ILS), where we only consider sparse solutions with entries in $\{0, \pm 1\}$. Formally, in the *sparse integer least square problem*, an instance consists of an $n \times d$ matrix M, a vector $b \in \mathbb{R}^n$, and a positive integer $\sigma \leq d$. Our task is to find a vector x which solves the optimization problem (SILS) or its variant (SILS'), defined as follows:

$$\begin{aligned} \min \quad & \frac{1}{n} \|Mx - b\|_2^2 \\ \text{s.t.} \quad & x \in \{0, \pm 1\}^d, \\ & \|x\|_0 \leq \sigma, \end{aligned} \qquad \text{(SILS)} \qquad \begin{aligned} \min \quad & \frac{1}{n} \|Mx - b\|_2^2 \\ \text{s.t.} \quad & x \in \{0, \pm 1\}^d, \\ & \|x\|_0 = \sigma. \end{aligned} \qquad \text{(SILS')}$$

One can interpret (SILS') as (SILS) with extra information or belief on the optimal choice of sparsity of the optimal solution.

As we will see in Theorem 1 in Section 2, Problem (SILS) and (SILS') are NP-hard in their full generality. In this paper, we are interested in polynomial running time algorithms that either obtain an approximated optimal solution to Problem (SILS), or obtain an exact optimal solution to Problem (SILS') provided some assumptions on the data input are satisfied. To be best of our knowledge, the only known result is in [5]. The authors propose a sparse sphere decoding algorithm which returns an optimal solution to Problem (SILS). They also show that this algorithm has an expected running time which is polynomial in d, in the case where M has i.i.d. standard Gaussian entries and there exists a sparse integer vector $z^* \in \{0, \pm 1\}^d$ such that the residual vector $b - Mz^*$ is comprised of i.i.d. Gaussian entries. However, this algorithm results in an exponential running time at the presence of a nonsparse z^*. Algorithms for Problem (SILS) with a non-polynomial running time include sparsity-exploiting sphere decoding-based MUD [57] and integer quadratic optimization algorithms (see, e.g., [7] and references therein). Efficient algorithms for Problem (SILS) with no theoretical guarantee on the quality of the solution can be found in [57, 44], where the authors proposed sparsity-exploiting decision-directed MUD, Lasso-based convex relaxation methods, and CoSaMP. Problem (SILS') has aroused many interests as well. Practical algorithms for Problem (SILS') include SF-OMP [45], and discrete valued sparse ADMM algorithm [43]. However, these two algorithm also do not have guarantees on the quality of their solutions.

Our contribution. In this paper, we further the understanding of the limits of computations for Problem (SILS) and Problem (SILS'). We provide a randomized algorithm that finds a feasible solution to (SILS) with high probability, and shows an approximation gap. Then, we obtain a broad class of data input which guarantee that (SILS') can be solved efficiently. To be concrete, in Section 3 we give an ℓ_1-based semidefinite relaxation to (SILS) and (SILS'), denoted by (SILS-SDP) and (SILS'-SDP), respectively. These two relaxations only differ at one single linear matrix inequality constraint. It is known that semidefinite programming (SDP) problems can be solved in polynomial time up to an arbitrary accuracy, by means of the ellipsoid algorithm and interior point methods [47, 29]. Recent studies have witnessed great success of SDP relaxations in (i) solving structured integer quadratic optimization problems in polynomial time, and (ii) finding the hidden sparse structure of a given mathematical object. Examples include clustering [26], sparse principal component analysis [2], sparse support vector machine [11], sub-Gaussian mixture model [18], community detection problem [25], and so on. Note that problems (SILS) and (SILS') are also by nature integer quadratic optimization problems with a sparsity constraint, so it is natural and well-motivated to seek for an effective SDP relaxations. In Section 4, we proposed a randomized algorithm, Algorithm 1, for (SILS). In fact, our randomized algorithm will not only work for (SILS), but for any binary quadratic programs with cardinality constraint (SBQP), provided that the coefficient matrix of the quadratic function has non-negative diagonal entries. Thus, Algorithm 1 can handle some non-convex quadratic objectives as well. The input of Algorithm 1 consists of an (approximated) optimal solution to (SILS-SDP), and two threshold constant T and C; the output is a d-vector $\bar{x}$ in $\{0, \pm 1\}$. We show in Theorem 2 that $\bar{x}$ is feasible to (SILS) with high probability, and the expected objective function is a $1/T^2$ multiple of the optimal value, after substracting an additional term that depends on T, C and the input data (M, b, σ). It can be shown that when $\sigma \ll T$, the additional term will diminish as $(\sigma, T) \to \infty$, and hence Algorithm 1 is an asymptotic $1/T^2$-approximation algorithm. To the best of our knowledge, Algorithm 1 is the first known randomized algorithm for (SILS) that has a theoretical guarantee. Then, we focus on (SILS'). One can conceive (SILS') to be (SILS) where the optimal sparsity parameter is known. We found that in this case, our proposed SDP relaxation shows stronger power than one will expect. In particular, we show that (SILS'-SDP) is able to solve (SILS') under several diverse sets of input (M, b, σ), suggesting that it is a very flexible relaxation. We give both theoretical and computational evidence, aiming to explain the flexibility of (SILS'-SDP). In Theorems 3 and 4 in Section 5, we provide sufficient conditions for (SILS'-SDP) to find a unique optimal solution to (SILS'). To the best of our knowledge, our results are the first ones that study the polynomial solvability of (SILS') in its full generality. Furthermore, we illustrate that our proposed sufficient conditions can be easily verified in some practical situation, based on a quantity known as matrix coherence. To be formal, we define the *coherence* of a positive semidefinite matrix Ψ to be

$$\mu(\Psi) := \max_{i \neq j} \frac{|\Psi_{ij}|}{\sqrt{|\Psi_{ii}\Psi_{jj}|}}, \tag{1}$$

where we assume $0/0 = 0$ if necessary. Recently, matrix coherence has aroused much attention in compressed sensing [17] and in sparsity-aware learning [46], thanks to its ease of computation and connection to the ability to recover a sparse optimal solution [37]. In this paper, we say that a model has a *high coherence* if we have $\mu(M^\top M) = \omega(1/\sigma)$, while it has a *low coherence* if we have $\mu(M^\top M) = \mathcal{O}(1/\sigma)$. In particular, in Theorem 5, we give sufficient conditions for (SILS'-SDP) to solve (SILS'), which are tailored to low coherence models.

Next, in Sections 6 and 7, we showcase the power and flexibility of (SILS'-SDP), by showing that it is able to nicely solve two problems of interest related to (SILS'): the *feature extraction problem* and the *integer sparse recovery problem*. All these results will be consequences of Theorems 3 and 4. The input to both the feature extraction problem and the integer sparse recovery problem is the same as the input to (SILS'): we are given an $n \times d$ matrix M, a vector $b \in \mathbb{R}^n$, and a positive integer $\sigma \leq d$. One key difference with general (SILS') is that the data input in these two problems satisfies

$$b = Mz^* + \epsilon, \tag{LM}$$

for some *ground truth* vector $z^* \in \mathbb{R}^d$ and for some small noise vector $\epsilon \in \mathbb{R}^n$. Note that, in this setting, z^* and ϵ are unknown, i.e., they are not part of the input of the problem. The linear model assumption is often present in real-world problems and has been considered in several works in the literature, including [39, 44, 57]. Next, we discuss in detail the feature extraction problem and the integer sparse recovery problem.

Feature extraction problem. The *feature extraction problem* is defined as Problem (SILS'), where (LM) holds (for a general vector z^*). A version of this problem, where the sparsity constraint is replaced with a penalty term in the objective, was studied in [39], where an optimal solution to the problem serves as a public storage of the feature vector z^*. This justifies the name of the problem, since it can be viewed as a way to extract features from z^*.

A closely related problem was considered in [52] to design an illumination-robust descriptor in face recognition. More generally, the idea of obtaining a sparse estimator from a general vector z^* arises in several areas of research, including subset selection, statistical learning, and face recognition. Some advantages of finding a sparse estimator even if the ground truth z^* is not necessarily sparse are reducing the cost of collecting data, improving prediction accuracy when variables are highly correlated, reducing model complexity, avoiding overfitting, and enhancing robustness [34, 51, 54]. As previously discussed, methods proposed in [5, 7, 44, 57] can also solve the feature extraction problem. However, they neither have a polynomial running time in the case where z^* is a nonsparse vector, nor they give a quality guarantee on the quality of the obtained solution.

Since the feature extraction problem is a special case of (SILS'), Theorems 3 and 4 already show that our semidefinite relaxation (SILS'-SDP) can efficiently solve this problem under certain conditions. Next, in Theorem 6, we specialize Theorem 4 to the feature extraction problem, in the case where M and ϵ have sub-Gaussian entries. The reason we are interested in this setting is that, in many fields of modern research such as compressed sensing [4], computer vision [38], and high dimensional statistics [32], it is more and more common to assume (sub-)Gaussianity in real-world data distributions. In particular, in Theorem 6, we derive a user-friendly version of our sufficient conditions when the second moment information of M is known.

Next, in Model 1, we give a concrete data model where the rows of M are i.i.d. standard Gaussian vectors. We prove in Theorem 7 that, for this model, (SILS'-SDP) can solve the feature extraction problem with high probability. We also provide numerical results showcasing the empirical probability that (SILS'-SDP) solves the feature extraction problem in this model.

Integer sparse recovery problem. In the *integer sparse recovery problem*, our input satisfies (LM), for some $z^* \in \{0, \pm 1\}^d$ with cardinality σ, and our goal is to recover z^* correctly. This is a well-known problem that arises in many fields, including sensor network [44], digital fingerprints [31], array signal processing [53], compressed sensing [27], and multiuser detection [57, 42]. In this paper, we show that we can often efficiently recover z^* by solving (SILS'-SDP). Intuitively, if ϵ is a sufficiently small vector, then z^* will be the unique optimal solution to (SILS'). Furthermore, if (SILS'-SDP) has a rank-one optimal solution W^*, then such solution is optimal also for (SILS'), and hence we recover z^* by checking the first column of W^*. Therefore, we study when (SILS'-SDP) can recover z^* correctly, and thus Theorems 3 and 4 naturally apply to the integer sparse recovery problem as well. We first apply Theorem 3 to this problem and obtain Theorem 8, where we provide sufficient conditions that do not depend on the coherence of $M^\top M$. This indicates that our proposed (SILS'-SDP) has the potential to withstand high coherence. We show that this is indeed true, both theoretically and computationally, by studying a concrete data model given by Model 2, where the rows of M have highly correlated random variables (which implies that $M^\top M$ admits high coherence). In Theorem 9, we prove that (SILS'-SDP) can recover z^* correctly with high probability in this model, thanks to Theorem 8. For low coherence models, we see that (SILS'-SDP) is able to recover z^* if we specialize Theorem 5 to the integer sparse recovery problem, even though this corollary has implications for more general problems. We study a low coherence concrete data model, given by Model 3, that is well studied in the literature, where the rows of M are i.i.d. standard Gaussian vectors. We show in Theorem 10 that (SILS'-SDP) can solve the integer sparse recovery problem with high probability in this model, thanks to the generality of Theorem 6.

We note that the integer sparse recovery problem is, in fact, a special case of the sparse recovery problem, which is a fundamental problem that has aroused much attention from different fields of research in the past decades, including compressed sensing [10, 16], high dimensional statistical analysis [9, 50], and wavelet denoising [14]. In the *sparse recovery problem*, our input satisfies (LM), for some $z^* \in \mathbb{R}^d$ with cardinality σ, and our goal is to recover the signed support of z^*. For details on the sparse recovery problem, we refer interesting readers to the excellent review [15]. Observe that, under the assumptions of the integer sparse recovery problem, i.e., $z^* \in \{0, \pm 1\}^d$, determining the signed support of z^* is equivalent to determining z^* itself. A large number of algorithms for sparse recovery problem have been introduced and studied in the literature [15, 19]. Since our SDP relaxation (SILS'-SDP) is by nature an ℓ_1-based convex relaxation, in Section 7, we compare our method with well-known ℓ_1-based convex relaxation algorithms. In particular we consider Lasso [2] and Dantzig Selector [9] (definitions are given in Section 7), and we see how they compare with (SILS'-SDP) in solving the integer sparse recovery problem theoretically and numerically. Theoretical guarantees for Lasso and Dantzig Selector have been extensively studied in the literature. For Lasso, a condition known as *mutual incoherence* [50] or *irrepresentable criterion* [55], is necessary and sufficient for the recovery of the signed support of z^*. In [30], the authors show that when the coherence of matrix $M^\top M$ is less than $1/(4\sigma)$, Lasso converges to z^*, provided some additional assumptions are met. Similarly, it was studied in [33] that when the coherence of $M^\top M$ is of order $\mathcal{O}(1/\sigma)$, Dantzig Selector is guaranteed to converge to z^* as well. For high coherence models, there have been several studies conducted for Lasso and Dantzig Selector. For example, the *restricted isometry property (RIP)* or the *null space property (NSP)* guarantee that Lasso and Dantzig Selector obtain a relatively good convergence to z^*. We refer interested readers to [56] and references therein for details and more sufficient conditions. As discussed in [39], however, all these assumptions are often violated in many real-world applications, and oftentimes these convex relaxation techniques do not attain a satisfactory performance under high coherence models [2, 41, 21]. In this paper, we show computationally that, under Model 2, Lasso and Dantzig Selector perform poorly, yet (SILS'-SDP) recovers z^* with high probability. The fact that, for this model, (SILS'-SDP) recovers z^* with high probability is implied by Theorem 9, thus the sufficient conditions in Theorem 8 cannot imply any (known or unknown) sufficient condition for the sparse recovery problem.

Organization of this paper In Section 2, we show that (SILS) and (SILS') is NP-hard. In Section 3, we present our SDP relaxation (SILS-SDP) and (SILS'-SDP). In Section 4, we give a randomized algorithm for (SILS), and deliver an optimality gap of this algorithm. In Section 5, we provide our general sufficient conditions for (SILS'-SDP) to solve (SILS'). In Section 6, we apply these sufficient conditions to the scenarios where (LM) holds, and discuss the implications for the feature extraction problem and the integer sparse recovery problem. In Section 7, we present the numerical results. To streamline the presentation, we defer some proofs to Sections 8 to 10. We conclude the introduction with the notation that will be used in this paper.

Notation: constants. In this paper, we say that a number in $\mathbb{R}$ is a *constant* if it only depends on the input of the problem, including its dimension. We say that a number in $\mathbb{R}$ is an *absolute constant* if it is a fixed number that does not depend on anything at all.

Notation: vectors. 0_d denotes the d-vector of zeros, 1_d denotes the d-vector of ones. For any positive integer d, we define $[d] := \{1, 2, \ldots, d\}$. Let x be a d-vector. The *support* of x is the set $\operatorname{Supp}(x) := \{i \in [d] : x_i \neq 0\}$. We denote by $\operatorname{diag}(x)$ the diagonal $d \times d$ matrix with diagonal entries equal to the components of x. For an index set $\mathcal{I} \subseteq [d]$, we denote by $x_{\mathcal{I}}$ the subvector of x whose entries are indexed by $\mathcal{I}$. We say that x is a *unit vector* if $\|x\|_2 = 1$, and we define the *unit sphere* in $\mathbb{R}^d$ as $\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$. For $1 \leq p \leq \infty$, we denote the *p-norm* of x by $\|x\|_p$. The 0-*(pseudo)norm* of x is $\|x\|_0 := |\operatorname{Supp}(x)|$.

Notation: matrices. I_n denotes the $n \times n$ identity matrix. O_n denotes the $n \times n$ zero matrix, and $O_{m\times n}$ denotes the $m \times n$ zero matrix. We denote by $\mathcal{S}^n$ the set of all $n \times n$ symmetric matrices. Let M be a $m \times n$ matrix. Given two index sets $\mathcal{I} \subseteq [m]$, $\mathcal{J} \subseteq [n]$, we denote by $M_{\mathcal{I},\mathcal{J}}$ the submatrix of M consisting of the entries in rows $\mathcal{I}$ and columns $\mathcal{J}$. We denote by $|M|$ the matrix obtained from M by taking the absolute values of the entries. We denote the rows of M by $m_1, m_2, \cdots, m_n$, and its columns by $M_1, M_2, \cdots, M_d$. For two $m \times n$ matrices M and N, we write $M \le N$ when each entry of M is at most the corresponding entry of N. If $M, N \in \mathcal{S}^n$, we use $M \succeq N$ to denote that $M - N$ is a positive semidefinite matrix. Let R be a $m \times m$ positive semidefinite matrix. We denote by $\lambda_i(R)$ the i-th smallest eigenvalue of R, and by $v_i(R)$ the (right) eigenvector corresponding to $\lambda_i(R)$. The minimum eigenvalue is also denoted by $\lambda_{\min}(R) := \lambda_1(R)$. We denote by $\text{diag}(R)$ the m-vector $(R_{11}, R_{22}, \ldots, R_{mm})^\top$. If X is a matrix, we denote by $X^\dagger$ the Moore-Penrose generalized inverse of X. Let $f(X) : \mathbb{R}^{n\times n} \to \mathbb{R}$ be a convex function and let $X_0 \in \mathbb{R}^{n\times n}$. We denote by $\partial f(X_0)$ the *subdifferential* (which is the set of subgradients) of f at X_0, i.e., $\partial f(X_0) := \{G \in \mathbb{R}^{n\times n} : f(Y) \ge f(X_0) + \text{tr}(G(Y - X_0)),\ \forall Y \in \mathbb{R}^{n\times n}\}$. The *p-to-q norm* of a matrix P, where $1 \le p,\ q \le \infty$, is defined as $\|P\|_{p\to q} := \min_{\|x\|_p=1} \|Px\|_q$. The *2-norm* of a matrix P is defined by $\|P\|_2 = \|P\|_{2\to 2}$. The *infinity norm*, also known as *Chebyshev norm*, of P is defined by $\|P\|_\infty := \max_{i,j} |P_{ij}|$. For a rank-one matrix $P = uv^\top$, clearly $\|P\|_\infty = \|u\|_\infty \|v\|_\infty$.

Notation: probability. We denote the expected value by $\mathbb{E}(\cdot)$. For a random event A, we denote the indicator variable for A to be $\mathbb{1}_A$. The expected value of a random variable X on a given random event A by $\mathbb{E}[X; A] = \mathbb{E}[X\mathbb{1}_A]$. A random vector $X \in \mathbb{R}^d$ is *centered* if $\mathbb{E}(X) = 0_d$. We denote the (multivariate) Gaussian distribution by $\mathcal{N}(\theta, \Sigma)$, where θ is the *mean* and Σ is the *covariance matrix*. We abbreviate 'independent and identically distributed' with 'i.i.d.', and 'with high probability' with 'w.h.p.', meaning with probability at least $1-\mathcal{O}(1/d)-\mathcal{O}(\exp(-c\sigma))$ for some absolute constant $c > 0$ in this paper. We say that a random variable $X \in \mathbb{R}$ is *sub-Gaussian* with parameter L if $\mathbb{E}\exp\{t(X - \mathbb{E}X)\} \le \exp\left(t^2L^2/2\right)$, for every $t \in \mathbb{R}$, and we write $X \sim \mathcal{SG}(L^2)$. We say a centered random vector $X \in \mathbb{R}^d$ is *sub-Gaussian* with parameter L if $\mathbb{E}\exp\left(tX^\top x\right) \le \exp\left(t^2L^2/2\right)$, for every $t \in R$ and for every x such that $\|x\|_2 = 1$. With a little abuse of notation, we also write $X \sim \mathcal{SG}(L^2)$. We say that a random variable $X \in \mathbb{R}$ is *sub-exponential* if $\mathbb{E}\exp(t|X|) \le \exp\{Kt\}$, for every $|t| \le 1/K$, for some constant K. For a sub-exponential random variable X, the *Orlicz norm* of X is defined as $\|X\|_{\psi_1} := \inf\{t > 0 : \mathbb{E}\exp(|X|/t) \le 2\}$. For more details, and for properties of sub-Gaussian and sub-exponential random variables (or vectors), we refer readers to the book [49].

Notation: optimality gap. Denote w^* to be the optimal solution to a optimization problem $\mathcal{P}$ with objective function f and input D. We say a randomized algorithm $\mathcal{A}$ is an r-approximation algorithm to the optimization problem, if $\mathcal{A}$ can output a random vector $\bar{w}$ with input D such that $\mathbb{E}f(\bar{w}) \ge 1/r \cdot f(w^*)$ if $\mathcal{P}$ is a maximization problem, and $\mathbb{E}f(\bar{w}) \le r \cdot f(w^*)$ if $\mathcal{P}$ is a minimization problem.

2 NP-hardness

In this section, we show that (SILS) is NP-hard. The proof of NP-hardness for (SILS') is almost identical, and hence we omit it here. To prove NP-hardness, we give a polynomial reduction from *Exact Cover by 3-sets (X3C)*. An instance of this decision problem consists of a set S and a collection C of 3-element subsets of S. The task is to decide whether C contains an *exact cover* for S, i.e., a sub-collection $\hat{C}$ of C such that every element of S occurs exactly once in $\hat{C}$. See [20] for details.

Theorem 1. *Problem* (SILS) *is NP-hard.*

Proof. First, we define the decision problem *SILS0*. An instance consists of the same data as in (SILS), and our task is that of deciding whether there exists $x \in \mathbb{R}^d$ such that

$$\begin{aligned} &Mx = b \\ &x \in \{0, \pm 1\}^d \\ &\|x\|_0 \le \sigma. \end{aligned} \tag{SILS0}$$

(SILS0) can be trivially solved by (SILS) since (SILS0) is feasible if and only if the optimal value of (SILS) is zero. Hence, to prove the theorem it is sufficient to show that (SILS0) is NP-hard. In the remainder of the proof, we show that (SILS0) is NP-hard by giving a polynomial reduction from X3C.

We start by showing how to transform an instance of X3C to an instance of (SILS0). Consider an instance of X3C given by a set $S = \{s_1, s_2, \dots, s_n\}$ and a collection $C = \{c_1, c_2, \dots, c_d\}$ of 3-element subsets of S. Without loss of generality we can assume that n is a multiple of 3, since otherwise there is trivially no exact cover. Let b be the vector 1_n of n ones. The matrix M has d column vectors, one for each set in C. Specifically, the jth column of M has entries $(z_1, z_2, \dots, z_n)$ where $z_i = 1$ if $s_i \in c_j$ and $z_i = 0$ otherwise. Finally, we set $\sigma := n/3$. To conclude the proof we show that the constructed instance of (SILS0) is feasible if and only if C contains an exact cover for S.

If C contains an exact cover for S, say $\hat{C}$, then consider the vector $\bar{x} \in \mathbb{R}^d$, where $\bar{x}_j = 1$ if $c_j \in \hat{C}$, and $\bar{x}_j = 0$ otherwise. Then we have $M\bar{x} = b$ and $\|\hat{x}\|_0 = n/3 = \sigma$, thus (SILS0) is feasible.

Conversely, assume the (SILS0) is feasible and let $\bar{x}$ be a feasible solution. Now consider the subcollection $\hat{C}$ of C, consisting of those sets c_j such that $\bar{x}_j$ is nonzero. We wish to prove that $\hat{C}$ is an exact cover for S. $Mx = b$ implies that each element of S is contained in at least one set in $\hat{C}$, and hence $\|\bar{x}\|_0 \geq \sigma$, thus $\|\bar{x}\|_0 = \sigma$. Since $\bar{x}$ has exactly $n/3$ nonzero entries, we have that $\hat{C}$ contains exactly $n/3$ subsets of S. Therefore each element of S is contained in exactly one set in $\hat{C}$ and so $\hat{C}$ is an exact cover for S. □

3 Semidefinite programming relaxations

In this section, we introduce our SDP relaxation of problems (SILS) and (SILS'). We define the $n \times (1+d)$ matrix $A := \begin{pmatrix} -b & M \end{pmatrix}$. We are now ready to define our SDP relaxations:

$$\begin{aligned} \min \quad & \frac{1}{n}\operatorname{tr}(A^\top A W) \\ \text{s.t.} \quad & W \succeq 0, \\ & W_{11} = 1, \\ & \operatorname{tr}(W_x) \leq \sigma, \\ & 1_d^\top \, |W_x| \, 1_d \leq \sigma^2, \\ & \operatorname{diag}(W_x) \leq 1_d. \end{aligned} \qquad \text{(SILS-SDP)}$$

$$\begin{aligned} \min \quad & \frac{1}{n}\operatorname{tr}(A^\top A W) \\ \text{s.t.} \quad & W \succeq 0, \\ & W_{11} = 1, \\ & \operatorname{tr}(W_x) = \sigma, \\ & 1_d^\top \, |W_x| \, 1_d \leq \sigma^2, \\ & \operatorname{diag}(W_x) \leq 1_d. \end{aligned} \qquad \text{(SILS'-SDP)}$$

In these models, the decision variables both $(1+d) \times (1+d)$ matrix of variables W. The matrix W_x is the submatrix of W obtained by dropping its first row and column. It is clear that the only difference of (SILS-SDP) and (SILS'-SDP) are whether or not $\operatorname{tr}(W_x)$ is strictly equal to σ.

In the next proposition, we show that (SILS-SDP) is indeed a relaxation of (SILS). The proof of (SILS'-SDP) being a valid relaxation of (SILS') is almost identical, and hence we omit it here.

Proposition 1. *Problem* (SILS-SDP) *is an SDP relaxation of Problem* (SILS). *Precisely:*

(i) Let x be a feasible solution to Problem (SILS)*, let w be obtained from x by adding a new first component equal to one, and let $W := ww^\top$. Then, W is feasible to Problem* (SILS-SDP) *and has the same cost as x.*

(ii) Let W be a feasible solution to Problem (SILS-SDP)*, and let x be obtained from the first column of W by dropping the first entry. If* $\operatorname{rank}(W) = 1$ *and $x \in \{0, \pm 1\}^d$, then x is feasible to Problem* (SILS) *and has the same cost as W.*

Proof. (i). Let x, w, W be as in the statement. To show that W is feasible to (SILS-SDP), we first see $W = ww^\top \succeq 0$ and $W_{11} = 1 \cdot 1 = 1$. Then, by direct calculation, $\mathrm{tr}(W_x) = \mathrm{tr}(xx^\top) = \|x\|_2^2 \leq \sigma$ and $\mathrm{diag}(W_x) \leq 1_d$ hold true. L astly, for a σ-dimensional vector z, we have $\|z\|_1 \leq \sqrt{\sigma}\,\|z\|_2$ by Cauchy-Schwartz inequality. Thus $\|x\|_0 \leq \sigma$ implies $\|x\|_1 \leq \sqrt{\sigma}\,\|x\|_2$, and we obtain

$$1_d^\top |W_x| 1_d = \|x\|_1^2 \leq \sigma\,\|x\|_2^2 \leq \sigma^2.$$

Regarding the costs of the solutions, we have

$$\frac{1}{n}\|Mx - b\|_2^2 = \frac{1}{n}\|Aw\|_2^2 = \frac{1}{n}\mathrm{tr}(w^\top A^\top Aw) = \frac{1}{n}\mathrm{tr}(A^\top Aww^\top) = \frac{1}{n}\mathrm{tr}(A^\top AW). \tag{2}$$

(ii). Let W and x be as in the statement and assume $\mathrm{rank}(W) = 1$ and $x \in \{0, \pm 1\}^d$. We write $W = ww^\top$ for some $(d+1)$-vector w. Given $W_{11} = 1$, we either have $w_1 = 1$ or $w_1 = -1$. In the case $w_1 = 1$ we have $W = \begin{pmatrix} 1 \\ x \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix}^\top$. In the case $w_1 = -1$ we have $W = \begin{pmatrix} -1 \\ -x \end{pmatrix}\begin{pmatrix} -1 \\ -x \end{pmatrix}^\top = \begin{pmatrix} 1 \\ x \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix}^\top$. From $x \in \{0, \pm 1\}^d$ and $\mathrm{tr}(W_x) \leq \sigma$, we obtain in both cases $\|x\|_0 \leq \sigma$, and so x is feasible to (SILS). Regarding the costs of the solutions, in both cases we have that (2) holds. □

4 A randomized algorithm for (SILS)

In this section, we present a novel randomized algorithm for the following binary quadratic optimization problem with sparsity constraint:

$$\begin{aligned} \min \quad & x^\top Px - 2c^\top x \\ \text{s.t.} \quad & x \in \{0, \pm 1\}^d, \\ & \|x\|_0 \leq \sigma, \end{aligned} \tag{SBQP}$$

where we assume that the input matrix $P \in \mathbb{R}^{d \times d}$ satisfies $P_{ii} \geq 0$, $\forall i \in [d]$, i.e., all its diagonal entries are non-negative, thus the objective function is not necessarily convex. Note that the optimal value of (SBQP) is non-positive, due to the feasibility of 0_d. Moreover, if one takes $P = M^\top M$ and $c = M^\top b$, then (SILS) is equivalent to (SBQP) by ignoring a constant $b^\top b$. To the best of our knowledge, this is the first randomized algorithm for solving a binary quadratic optimization problem with cardinality constraint. Our proposed randomized algorithm is inspired by [12], where the authors presented a $\mathcal{O}(\log d)$-approximation algorithm for maximizing a quadratic function $x^\top Px$ over $\{\pm 1\}^d$. In their setting, the authors assume that $P_{ii} = 0$, as x_i^2 must be one. However, in (SBQP), such assumption is not reasonable due to the cardinality constraint. This issue also prevents one from applying their algorithm directly, as one cannot obtain a sparse vector. In fact, [12] introduced a specific random variable that decides a chosen entry is ± 1. The idea depends on the fact that the u_i's forming square root of the (approximated) optimal solution are unit vectors, which is not true in Algorithm 1. Moreover, we have an additional linear term $-2c^\top x$. In this section, we show that, all these problems can all be solved by choosing a distribution that carefully handles sparsity, at a cost of an additional addictive term in the approximation gap.

Let the matrix $Q(c, P) := \begin{pmatrix} 0 & -c^\top \\ -c & P \end{pmatrix}$. Denote $\mathrm{SDP}(c, P)$ to be the optimization problem by replacing the objective function $1/n \cdot \mathrm{tr}(A^\top AW)$ by $\mathrm{tr}(Q(c, P)W)$ in (SILS-SDP). Following the proof idea of Proposition 1, it is clear that $\mathrm{SDP}(c, P)$ is indeed a relaxation of (SBQP). We define a threshold function $h(x)$ which takes value 1 if $x > 1$, x if $-1 < x < 1$, and -1 if $x < -1$. Now, we present the detailed randomized algorithm in Algorithm 1.

An approximation gap of Algorithm 1 is stated as follows, and the proof is left in Section 8.

Algorithm 1 Randomized Algorithm for (SBQP)

Input: An ϵ-approximated optimal solution $W^* \in \mathbb{R}^{(d+1)\times(d+1)}$ to SDP(c, P), threshold constants $0 < C \leq 1$ and $T > 0$.

Output: A vector (0-indexed) $\bar{x}$ in $\{0, \pm 1\}^d$

$U := (u_0, u_1, \ldots, u_d) \in \mathbb{R}^{(d+1)\times(d+1)} \leftarrow \sqrt{W^*}$
Generate a random vector $g \sim \mathcal{N}(0_{d+1}, I_{d+1})$
$z_0 \leftarrow u_0^\top g$, $y_0 \leftarrow h(z_0/T)$
Sample $x_0 = 1$ with probability $(1 + y_0)/2$, and $x_0 = -1$ with probability $(1 - y_0)/2$
for $k = 1, 2, \ldots, d$ **do**
    $p_k \leftarrow 2/3 \cdot \|u_i\|_2^2$ if $\|u_i\|_2 > C$, and $p_k \leftarrow 0$ if otherwise
    Sample $\epsilon_k = 1$ with probability p_k and $\epsilon_k = 0$ with probability $1 - p_k$, independent of k and g
    $\tilde{u}_k \leftarrow \epsilon_k \cdot u_k / p_k$ (where we assume $0/0 = 0$), $z_k \leftarrow \tilde{u}_k^\top g_k$, $y_k \leftarrow h(z_k/T)$
    Sample $x_k = \text{sign}(y_k)$ with probability $|y_k|$, and $x_k = 0$ with probability $1 - |y_k|$
end for
return $\bar{x} := \text{sign}(x_0) \cdot (x_1, \ldots, x_d)^\top$

Theorem 2. *Assume P is a $d \times d$ symmetric matrix with nonnegative diagonal entries, and c is a d-vector. Denote W^* to be an ϵ-optimal solution to* SDP(c, P)*, x^* to be the optimal solution to* (SBQP)*. Let $\bar{x}$ be the output of Algorithm 1, with input W^* and threshold constants $0 < C < 1$ and $T > 0$. Define $B := \|Q(c, P)\|_\infty$. Then, we have*

$$\mathbb{E}(\bar{x}^\top P \bar{x} - 2c^\top \bar{x}) - B \cdot \left[f(T, C, \sigma, d) + \frac{1}{T^2}(3\sigma + \sigma^2) + \frac{\sqrt{3}}{\sqrt{2}T} \min\left\{ d, \frac{\sigma}{C^2} \right\} \right]$$
$$\leq \frac{1}{T^2} \cdot \text{tr}(Q(c, P) W^*) \leq \frac{1}{T^2} \cdot \left[(x^*)^\top P x^* - 2c^\top x^* + \epsilon \right]$$

where $f(T, C, \sigma, d) := \mathcal{O}\left(\sigma e^{-C^2 T^2} [\min\{d, \sigma/C^2\}/(CT) + T/C] \right)$, and we omit possibly a constant scaling of T in the Big-O notation. Furthermore, with high probability, $\bar{x}$ is feasible to (SBQP).

Remark. We first observe that, in the case where $\sigma \ll T$ and $B, C > 0$ are fixed, the term $g(B, T, C, \sigma, d) := B \cdot \left[f(T, C, \sigma, d) + \frac{1}{T^2}(3\sigma + \sigma^2) + \frac{\sqrt{3}}{\sqrt{2}T} \min\left\{ d, \frac{\sigma}{C^2} \right\} \right]$ in Theorem 2 is diminishing as $(\sigma, T) \to \infty$, and thus we can obtain a solution $\bar{x}$ with an expected objective value that is an asymptotically $1/T^2$ multiple of $(x^*)^\top P x^* - 2c^\top x^* + \epsilon$. In [12], the authors take $T = 4\sqrt{\log(d)}$ and obtain a $\mathcal{O}(\log(d))$-approximation algorithm for maximization binary quadratic problems. We can obtain a similar result by taking the same value for such T, and if we further fix $0 < C < 1$, at the cost of an additional term $g(B, T, C, \sigma, d)$. If we further assume that $\sigma \ll \sqrt{\log(d)}$ and B is fixed, then we obtain an asymptotic $\mathcal{O}(1/\log(d))$-approximation algorithm. Finally, for different input $Q(c, P)$ and σ, one can accordingly choose different values for T and C to obtain a acceptable trade-off between the term $g(B, T, C, \sigma, d)$ and the multiplicative factor $1/T^2$.

In Section 7.1, we will demonstrate some numerical results of Algorithm 1.

5 Sufficient conditions for recovery

In this section, we study (SILS'). Note that one can interpret solving (SILS') as solving (SILS) given an optimal choice of σ. For the ease of illustration, starting from this section, we say that (SILS'-SDP) *recovers* x^*, if $x^* \in \{0, \pm 1\}^d$, and (SILS'-SDP) admits a unique rank-one optimal solution $W^* := \begin{pmatrix} 1 \\ x^* \end{pmatrix} \begin{pmatrix} 1 \\ x^* \end{pmatrix}^\top$. Due to Proposition 1, the vector x^* is then optimal to (SILS'), and hence we also say that (SILS'-SDP) *solves* (SILS') if there exists a vector $x^* \in \{0, \pm 1\}^d$ such that (SILS'-SDP) recovers x^*. We remark that, if (SILS'-SDP) solves (SILS'), then (SILS') can be indeed solved in polynomial time by solving (SILS'-SDP), because we can obtain x^* by checking the first column of W^*.

We present Theorems 3 and 4, which are two of the main results of this section. In both theorems, we provide sufficient conditions for (SILS'-SDP) to solve (SILS'), which are primarily focused on the input $A = (M, -b)$ and σ. The statements require the existence of two parameters μ_2^* and δ, and in Theorem 3 we also require the existence of a decomposition of a specific matrix Θ. Therefore, both theorems below can help us identify specific classes of problem (SILS') that can be solved by (SILS'-SDP). As a corollary to Theorem 4, we then obtain Theorem 5, where we show that in a low coherence model, (SILS') can be solved by (SILS'-SDP) under certain conditions.

It is worth to note that, although the linear model assumption (LM) is often present in the literature in integer least square problems (see, e.g., [5]), in this section we consider the general setting where we do not make this assumption. To help readers understand better the complicated geometry, we will split the section into two parts. In the first part, we discuss KKT conditions, and state Lemma 1 based on KKT conditions, along with a stronger assumption that two specific parameters μ_2^* and δ exist. In the second part, we leave the statements of two theorems, and discuss the conditions semantically. The proofs can be found in Section 9.

5.1 KKT conditions

In this section, we study the Karush–Kuhn–Tucker (KKT) conditions [28]. We start by studying the dual of (SILS'-SDP), and list KKT conditions when (SILS'-SDP) admits an optimal solution W^*. Based on KKT conditions, we then provide a cleaner sufficient conditions for recovering a sparse vector $x^* \in \{0, \pm 1\}^d$ in Lemma 1.

The dual problem of (SILS'-SDP) is

$$\begin{aligned} \max_{\mu_1, \mu_2, \mu_3, Y, p} \quad & -\mu_1 - \sigma\mu_2 - \sigma^2\mu_3 - p^\top 1_d \\ \text{s.t.} \quad & \left| \frac{1}{n} A^\top A - Y + \begin{pmatrix} \mu_1 & \\ & \operatorname{diag}(p) + \mu_2 I_d \end{pmatrix} \right| \le \mu_3 1_d 1_d^\top, \\ & Y \succeq 0, \\ & p \ge 0_d. \end{aligned} \tag{SILS'-SDP-dual}$$

Define a convex function $f : \mathbb{R}^{(1+d)\times(1+d)} \to \mathbb{R}$ by $f(Z) := (0, 1_d^\top)|Z| \begin{pmatrix} 0 \\ 1_d \end{pmatrix}$. Note that the subdifferential of f at Z is exactly

$$\partial f(Z) = \left\{ U \in \mathbb{R}^{(1+d)\times(1+d)} : U_{ij} = \begin{cases} 0, & \text{if at least one of } i, j \le 1, \\ \operatorname{sign}(Z_{ij}), & \text{if both of } i, j \ge 2 \text{ and } Z_{ij} \ne 0, \\ \in [-1, 1], & \text{otherwise.} \end{cases} \right\}. \tag{3}$$

Then, KKT conditions state that $W^* = \begin{pmatrix} 1 \\ x^* \end{pmatrix} \begin{pmatrix} 1 \\ x^* \end{pmatrix}^\top$ is optimal to (SILS'-SDP) if and only if there exist dual variables $Y^* = \begin{pmatrix} Y_{11}^* & (y^*)^\top \\ y^* & Y_x^* \end{pmatrix}$, μ_1^*, p^*, μ_2^*, and μ_3^* feasible to (SILS'-SDP-dual) such that:

$$O_{d+1} \in \left\{ \frac{1}{n} A^\top A - Y^* + \begin{pmatrix} \mu_1^* & \\ & \operatorname{diag}(p^*) + \mu_2^* I_d \end{pmatrix} \right\} + \mu_3^* \partial f(W^*), \tag{KKT-1}$$

$$Y^*W^* = O_{1+d} \Longleftrightarrow Y^* \begin{pmatrix} 1 \\ x^* \end{pmatrix} = 0_{1+d}, \tag{KKT-2}$$

$$(p^*)^\top (\mathrm{diag}(W_x^*) - 1_d) = 0, \tag{KKT-3}$$

where we apply Minkowski sum in (KKT-1). If we focus our attention on the block matrix that contains $1/n \cdot (M^\top M)_{S,S}$ in (KKT-1), we obtain that

$$-\mu_3^* x_S^* (x_S^*)^\top = \left[\frac{1}{n} M^\top M - Y_x^* + \mathrm{diag}(p^* + \mu_2^* 1_d)\right]_{S,S}. \tag{4}$$

Moreover, insert $(Y_x^*)_{S,S}$ in (4) into (KKT-2), one observe that

$$\mathrm{diag}(p_S^*) x_S^* = -\frac{1}{n}(M^\top M)_{S,S} x_S^* - \sigma \mu_3^* x_S^* - y_S^* - \mu_2^* x_S^*. \tag{5}$$

Note that (5) uniquely determines the vector p_S^* if other dual variables are determined. The constraint $p_S^* \geq 0_\sigma$ is then implied by the following two stronger conditions:

$$\mu_2^* \leq -\lambda_{\min}\Big(\frac{1}{n}(M^\top M)_{S,S}\Big) + \delta, \tag{6}$$

$$\mu_3^* := \frac{1}{\sigma}\Big\{\lambda_{\min}\Big(\frac{1}{n}(M^\top M)_{S,S}\Big) - \delta + \min_{i \in S}\Big[-y^* - \frac{1}{n}(M^\top M)_{S,S} x_S^*\Big]_i / x_i^*\Big\}. \tag{7}$$

Here, the minimum eigenvalue of the matrix $1/n \cdot (M^\top M)_{S,S}$ introduced in (6) and (7) helps guarantee that, a block matrix $H_{S,S}$ defined in the statement of Lemma 1, is positive semidefinte, which is a necessary condition for $Y^* \succeq 0$. The details will be made clear in the proof of Lemma 1, in Section 9.

Together with all these intuitions, we are ready to state Lemma 1, about block structures of dual variables that guarantee recovery of x^*:

Lemma 1. *Let $x^* \in \{0, \pm 1\}^d$, define $S := \mathrm{Supp}(x^*)$, and assume $|S| = \sigma$. Define $y^* := -M^\top b/n$, $Y_{11}^* := -(y_S^*)^\top x_S^*$, and assume $Y_{11}^* > 0$. Let $\delta > 0$, μ_2^* satisfy* (6), *μ_3^* be defined by* (7), *$p^* \in \mathbb{R}^d$ be a vector with $p_{S^c}^* := 0_{d-\sigma}$ and p_S^* satisfying* (5). *Let $Y_x^* \in \mathbb{R}^{d \times d}$ be a matrix that satisfies* (4), *and let $H := Y_x^* - \frac{1}{Y_{11}^*} y^* (y^*)^\top$. Then we have $p^* \geq 0_d$, $\lambda_2(H_{S,S}) \geq \delta$, and $H_{S,S} \succeq 0$.*

Assume, in addition, that the following conditions are satified:

1A. $H_{S^c,S^c} \succeq H_{S^c,S} H_{S,S}^\dagger H_{S^c,S}^\top$;

1B. $H_{S^c,S} x_S^* = 0_{d-\sigma}$;

1C. $\big\|(\frac{1}{n} M^\top M - Y_x^*)_{S^c,S}\big\|_\infty \leq \mu_3^*$;

1D. $\big\|(\frac{1}{n} M^\top M - Y_x^*)_{S^c,S^c} + \mu_2^* I_{d-\sigma}\big\|_\infty \leq \mu_3^*$.

Then $W^* = w^*(w^*)^\top$, where $w^* = \begin{pmatrix} 1 \\ x^* \end{pmatrix}$, is an optimal solution to (SILS'-SDP)*. Furthermore, if we also assume that $\lambda_2(H) > 0$, then W^* is the unique optimal solution to* (SILS'-SDP)*.*

Remark. In this remark, we draw attention to the fact that the assumption $Y_{11}^* > 0$ in Lemma 1 is actually natural, given that $\sigma \geq 1$ is indeed the optimal support size of (SILS). Indeed, for any optimal solution x^* to (SILS'), one must have $\|Mx^* - b\|_2^2 = (x^*)^\top M^\top M x^* - 2b^\top M x^* + \|b\|_2^2 < \|b\|_2^2$, since otherwise we choose $x^* = 0_d$. This implies $0 \leq \|Mx^*\|_2^2 < 2b^\top M x^* = n \cdot Y_{11}^*$. Finally, we point out that the optimality of σ in (SILS) is not necessarily required in Lemma 1 - all that is required are the assumptions made there.

5.2 Main theorems for recovery

In this section, we state the main theorems for recovery. In a nutshell, we take different candidates of $(Y_x^*)_{S^c,S}$ in Lemma 1, and present the corresponding sufficient conditions for recovery. Note that in Lemma 1, our choice of $(Y_x^*)_{S,S}$ is fixed (which is implied by (4)). Thus, it would be well-motivated if we further fixed $(Y_x^*)_{S^c,S}$ to be a specific determined matrix, and then construct $(Y_x^*)_{S^c,S^c}$ accordingly. Particularly, in Theorem 3, we assign $(Y_x^*)_{S^c,S}$ to be the solution to the optimization problem

$$\min \left\| \frac{1}{n} M^\top M - (Y_x)_{S^c,S} \right\|_F \qquad \text{s.t.} \qquad (Y_x)_{S^c,S} x_S^* = -y_S^*, \tag{8}$$

where we relax the max norm of the matrix in **1C** by its Frobenius norm, and combined with **1A**, so that we can obtain a closed-form solution. In Theorem 4, we we assign $(Y_x^*)_{S^c,S}$ to be a even simpler matrix - a rank-one matrix $-y_{S^c}^*(x_S^*)^\top/\sigma$.

We note here, although these candidates for $(Y_x^*)_{S^c,S}$ might not make perfect sense for general data inputs (M, b, σ), we found that they fit well in (sub-)Gaussian data matrix M and the linear model assumption (LM). We leave these theorems here as they might still be of interest for some other specific data inputs. Further discussion on (sub-)Gaussianity and (LM) can be found in Section 6.

We state the first theorem in this section:

Theorem 3. *Let $x^* \in \{0, \pm 1\}^d$, define $S := \operatorname{Supp}(x^*)$, and assume $|S| = \sigma$. Define $y^* := -M^\top b/n$, $Y_{11}^* := -(y_S^*)^\top x_S^*$, and assume $Y_{11}^* > 0$. Then,* (SILS'-SDP) *recovers x^*, if there exists a constant $\delta > 0$ such that the following conditions are satisfied:*

A1. *$\left\| \frac{1}{n\sigma}(M^\top M)_{S^c,S} x_S^* + \frac{1}{\sigma} y_{S^c}^* \right\|_\infty \le \mu_3^*$, where μ_3^* is defined by* (7)

A2. *There exists μ_2^* satisfying* (6) *such that the matrix*

$$\Theta := \frac{1}{n}(M^\top M)_{S^c,S^c} + \mu_2^* I_{d-\sigma} - \frac{1}{Y_{11}^*} y_{S^c}^* (y_{S^c}^*)^\top - R \frac{1}{\delta}\big(I_\sigma - \frac{1}{\sigma} x_S^* (x_S^*)^\top\big) R^\top$$

can be written as the sum of two matrices $\Theta_1 + \Theta_2$, with $\Theta_1 \succ 0$, $\|\Theta_2\|_\infty \le \mu_3^*$ or $\Theta_1 \succeq 0$, $\|\Theta_2\|_\infty < \mu_3^*$, where $R := \frac{1}{n}(M^\top M)_{S^c,S} - \frac{1}{Y_{11}^*} y_{S^c}^* (y_S^*)^\top$.

Remark. We first remark that condition **A1** would not be a very restricted assumption, as we are optimizing the relaxed problem (8), and one can choose $\delta > 0$ in (7) wisely according to the optimal value of (8). Plus, condition **A2** in Theorem 3 is not as strong as it might seem. This condition asks for a decomposition of Θ into the sum of a positive definite Θ_1 and another matrix Θ_2 with infinity norm upper bounded by μ_3^*. To construct Θ_1, the following informal idea may be helpful. By Lemma 5 (which can be found in Section 9), $M^\top M \succeq 0$ implies

$$(M^\top M)_{S^c,S^c} \succeq (M^\top M)_{S^c,S} M_{S,S}^\dagger (M^\top M)_{S^c,S}^\top.$$

Therefore, if $(M^\top M)_{S^c,S^c}$ is large enough and δ is chosen wisely, the matrix

$$\frac{1}{n}(M^\top M)_{S^c,S^c} - \frac{1}{n}(M^\top M)_{S^c,S} \frac{1}{\delta}\big(I_\sigma - \frac{1}{\sigma} x_S^*(x_S^*)^\top\big) \frac{1}{n}(M^\top M)_{S,S^c}$$

is positive semidefinite and can be used to construct the positive semidefinite matrix Θ_1.

Numerically, we found that such decomposition $\Theta = \Theta_1 + \Theta_2$ often exists for several different instances; however, it can be challenging to write it down explicitly. A specific instance is given in Model 2 in Section 6. In particular, it is an interesting open problem to obtain a simple sufficient condition which guarantees the existence of such decomposition.

In the next theorem, the sufficient conditions are easier to check than those in Theorem 3. This is because the main idea of Theorem 4 depends on a simpler structure of $(Y_x^*)_{S^c,S}$, and hence the theorem statement only requires the existence of two parameters μ_2^* and δ.

Theorem 4. *Let $x^* \in \{0, \pm 1\}^d$, define $S := \mathrm{Supp}(x^*)$, and assume $|S| = \sigma$. Define $y^* := -M^\top b/n$, $Y_{11}^* := -(y_S^*)^\top x_S^*$, and assume $Y_{11}^* > 0$. Let $\theta := \arccos\left(\frac{(y_S^*)^\top x_S^*}{\sqrt{\sigma}\|y_S^*\|_2}\right)$. Then,* (SILS'-SDP) *recovers x^*, if there exists a constant $\delta > 0$ such that the following conditions are satisfied:*

B1. *$\left\|\frac{1}{n}(M^\top M)_{S,S^c} + \frac{1}{\sigma} y_{S^c}^* (x_S^*)^\top\right\|_\infty \le \mu_3^*$, where μ_3^* is defined by* (7)*;*

B2. *There exists μ_2^* satisfying* (6) *such that μ_3^* is strictly greater than*

$$\left\|\frac{1}{n}(M^\top M)_{S^c,S^c} + \mu_2^* I_{d-\sigma}\right\|_\infty + \left\|\frac{1}{Y_{11}^*} y_{S^c}^* (y_{S^c}^*)^\top\right\|_\infty + \frac{1-\cos^2(\theta)}{\sigma\delta\cos^2(\theta)} \|y_{S^c}^*\|_\infty^2$$

Next, we give a corollary to Theorem 4, which shows that the assumptions of Theorem 4 can be fulfilled in models with a low coherence.

Corollary 5. *Let $x^* \in \{0, \pm 1\}^d$, define $S := \mathrm{Supp}(x^*)$, and assume $|S| = \sigma$. Define $y^* := -M^\top b/n$, $Y_{11}^* := -(y_S^*)^\top x_S^*$, and assume $Y_{11}^* > 0$. Let $\theta := \arccos\left(\frac{(y_S^*)^\top x_S^*}{\sqrt{\sigma}\|y_S^*\|_2}\right)$. Denote $\Delta_1 := \min_{i\in S}(-y_i^*/x_i^*) - \|y_{S^c}^*\|_\infty$ and*

$$\Delta_2 := \min_{i\in S}(-y_i^*/x_i^*) - \sigma \|y_{S^c}^*\|_\infty^2 / Y_{11}^* + \frac{1-\cos^2(\theta)}{\delta\cos^2(\theta)} \|y_{S^c}^*\|_\infty^2 .$$

Suppose that the columns of M are normalized such that $\max_{i\in[d]} \|M_i\|_2 \le 1$. Then, (SILS'-SDP) *recovers x^*, if there exists a constant $\delta > 0$ such that the following conditions are satisfied:*

C1. *$\lambda_{\min}\left(\frac{1}{n}(M^\top M)_{S,S}\right) - \delta - \left\|\frac{1}{n}(M^\top M)_{S,S} x_S^*\right\|_\infty + \min_{j=1,2}\Delta_j \ge \Delta > 0$ for some constant Δ;*

C2. *There exists μ_2^* satisfying* (6) *such that*

$$\left\|\mathrm{diag}\left(\frac{1}{n}(M^\top M)_{S^c,S^c} + \mu_2^* I_{d-\sigma}\right)\right\|_\infty < \frac{\Delta}{\sigma};$$

C3. *$\mu(M^\top M) < \Delta/\sigma$, where $\mu(\cdot)$ is defined in* (1).

Proof. We define μ_3^* as in (7). From **C3**, we obtain that $\max_{i\neq j} |(M^\top M/n)_{ij}| \le \mu(M^\top M/n) = \mu(M^\top M) \le \frac{\Delta}{\sigma}$. Then, we observe that

$$\mu_3^* \ge \frac{1}{\sigma}\Big\{\lambda_{\min}\big(\frac{1}{n}(M^\top M)_{S,S}\big) - \delta + \min_{i\in S}(-y_i^*/x_i) - \left\|\frac{1}{n}(M^\top M)_{S,S} x_S^*\right\|_\infty\Big\}$$

and

$$\left\|\frac{1}{n}(M^\top M)_{S,S^c} + \frac{1}{\sigma} y_{S^c}^*(x_S^*)^\top\right\|_\infty \le \left\|\frac{1}{n}(M^\top M)_{S,S^c}\right\|_\infty + \frac{1}{\sigma}\|y_{S^c}^*\|_\infty .$$

Combining these facts with **C1**, we see that **B1** holds. If, in addition, **C2** holds, we obtain **B2**. □

Remark. Theorem 5 shows that, if the data matrix $M^\top M$ has a low coherence, (SILS'-SDP) can solve (SILS') well under conditions **C1** and **C2**. In this remark, we informally illustrate how these two conditions can be easily fulfilled in certain scenarios. Observe that **C1** and **C2** hold if $\min_{j=1,2} \Delta_j$ is sufficiently large, and it is indeed possible to obtain a large $\min_{j=1,2} \Delta_j$. Intuitively, a large Δ_1 can be obtained if, for example, there is a set S with cardinality σ such that $\min_{i\in S} |y_i^*| - \|y^*_{S^c}\|_\infty$ is large, and $x^*_S = \text{sign}(y^*_S)$. Also the requirement that Δ_2 is large is not as restrictive as it might seem. In particular, if $\cos(\theta)$ is close to one, we easily obtain a large Δ_2 if we secure a large Δ_1. Indeed, since $\sigma \|y^*_{S^c}\|^2_\infty / Y^*_{11} = \sigma \|y^*_{S^c}\|^2_\infty / (-\sum_{i\in S} y_i^* x_i^*)$, each term in the summation on the denominator is always greater than $\|y^*_{S^c}\|_\infty$ if Δ_1 is large. Thus, this term is in fact upper bounded by $\|y^*_{S^c}\|_\infty$. As another term $[1-\cos^2(\theta)]/\cos^2(\theta) \cdot \|y^*_{S^c}\|^2_\infty$ vanishes given that $\cos(\theta)$ is close to one, we thus obtain that $\Delta_2 \approx \Delta_1$, and so Δ_2 is also large.

While the above ideas on how **C1** and **C2** can be satisfied are not very precise, they can be further formalized and used in proofs for some concrete data models, including those given in the next section.

6 Consequences for linear data models

In this section, we showcase the power of Theorems 3 and 4, by presenting some of their implications for the feature extraction problem and the integer sparse recovery problem, as defined in Section 1. First, note that we can directly employ these two theorems and Theorem 5 in the specific settings of the two problems, in order to obtain corresponding sufficient conditions for (SILS'-SDP) to solve these problems. To avoid repetition, we do not present these specialized sufficient conditions, and we leave their derivation to the interested reader. Instead, we focus on the consequences of Theorems 3 and 4 for these two problems, that we believe are the most significant. In Section 6.1, we consider the feature extraction problem, where M and ϵ have sub-Gaussian entries. We specialize Theorem 4 to this setting, and thereby obtain Theorem 6, where we give user-friendly sufficient conditions based on second moment information. In Section 6.1.1, we then give a concrete data model for the feature extraction problem. In particular, the feature extraction problem under this data model can be solved by (SILS'-SDP) due to Theorem 6. Next, in Section 6.2, we consider the integer sparse recovery problem. We present Theorem 8, which is obtained by specializing Theorem 3 to this problem. We then consider two concrete data models for the integer sparse recovery problem, which can be solved by (SILS'-SDP). The first one, presented in Section 6.2.1, has a high coherence, while the second one, in Section 6.2.2, has a low coherence.

We note that, we will prove that (SILS'-SDP) works well for several probabilistic models, by showing that if the number of data points n is large enough, (SILS'-SDP) recovers a specific x^* with high probability. However, discussion on sample complexity is not the main focus of this paper. All these illustrations are intended to showcase the power and flexibility of (SILS'-SDP) solving (SILS').

6.1 Feature extraction problem with sub-Gaussian data

In this section, we consider the feature extraction problem, and we assume that M and ϵ have sub-Gaussian entries. Recall that the feature extraction problem is Problem (SILS'), where (LM) holds (for a general vector z^*). We first give a technical lemma, which gives high-probability upper bounds for metrics between some random variables and their means. This lemma is due to known results in probability and statistics.

Lemma 2. *Suppose that M consists of centered row vectors $m_i \overset{i.i.d.}{\sim} \mathcal{SG}(L^2)$ for some $L > 0$ and $i \in [n]$, and denote the covariance matrix of m_i by Σ. Assume the noise vector ϵ is a centered sub-Gaussian random vector independent of M, with $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{SG}(\varrho^2)$ for $i \in [n]$. Then, the following statements hold:*

2A. *Suppose $\sigma/n \to 0$. Then, there exists an absolute constant $c_1 > 0$ such that $\left\|\frac{1}{n}(M^\top M)_{S,S} - \Sigma_{S,S}\right\|_2 \leq c_1 L\sqrt{\sigma/n}$ holds w.h.p. as $(n, \sigma) \to \infty$;*

2B. *Suppose $\log(d)/n \to 0$ and let $F := \frac{1}{n}M^\top M - \Sigma$. Then, there exists an absolute constant B such that $\|F\|_\infty \leq BL^2\sqrt{\log(d)/n}$ holds w.h.p. as $(n, d) \to \infty$;*

2C. *Suppose $\log(d)/n \to 0$ and let $F := \frac{1}{n}M^\top M - \Sigma$. Let $x^* \in \{0, \pm 1\}^d$, define $S := \mathrm{Supp}(x^*)$, and assume $|S| = \sigma$. Then, there exists an absolute constant B_1 such that $\|Fx^*\|_\infty = \|F_{S,S}x^*_S\|_\infty \leq B_1 L^2\sqrt{\sigma\log(d)/n}$ holds w.h.p. as $(n, d) \to \infty$;*

2D. *Suppose $\log(d)/n \to 0$ and let $F := \frac{1}{n}M^\top M - \Sigma$. Let $z^* \in \mathbb{R}^d$. Then, there exists an absolute constant B_2 such that*

$$\left\|Fz^* + \frac{1}{n}M^\top \epsilon\right\|_\infty < B_2 L\sqrt{(\varrho^2 + L^2\|z^*\|_2^2)\log(d)/n}$$

holds w.h.p. as $(n, d) \to \infty$.

Proof. **2A** follows from Proposition 2.1 in [48].

To show **2B**, we first observe that each entry in F is a sub-exponential random variable with Orlicz norm upper bounded by an absolute constant multiple of L^2 (by Lemma 2.7.7 and Exercise 2.7.10 in [49]). Since $\log(d)/n \to 0$, by Bernstein inequality (see, e.g., Theorem 2.8.1 in [49]), we obtain that there exists an absolute constant $c > 0$ such that for $i, j \in [d]$,

$$\mathbb{P}(|F_{ij}| > t) \leq 2\exp\Big(-c\frac{nt^2}{L^4}\Big). \tag{9}$$

Then, using the union bound, we see that

$$\mathbb{P}(\|F\|_\infty > t) \leq 2d^2\exp\Big(-c\frac{nt^2}{L^4}\Big) = 2\exp\Big(2\log(d) - c\frac{nt^2}{L^4}\Big).$$

Taking $t = BL^2\sqrt{\log(d)/n}$ for some large absolute constant $B > 0$, we see that $\mathbb{P}(\|F\|_\infty > t) \leq \mathcal{O}(1/d)$. Note that, in the previous argument, although we cannot apply Theorem 2.8.1 directly because we do not know the exact Orlicz norms, the statement is true if we replace $\sum_{i=1}^n \|X_i\|_{\psi_1}^2$ and $\max_i \|X_i\|_{\psi_1}$ with their upper bounds, which are nK^4L^4 and K^2L^2, for an absolute constant $K > 0$. The reason is that the proof of Theorem 2.8.1 still works with such replacement, although we get a slightly worst bound. We will use Bernstein inequality similarly later in the proof.

To show **2C**, we observe that for any nonzero vector $x \in \mathbb{R}^d$, $\sum_{j=1}^d x_j m_{kj}$ falls into the distribution class $\mathcal{SG}(\|x\|_2^2 L^2)$. Then $(Fx)_i = \frac{1}{n}\sum_{j=1}^d\sum_{k=1}^n m_{ki}m_{kj}x_j - (\Sigma x)_i = \frac{1}{n}\sum_{k=1}^n m_{ki}\sum_{j=1}^d x_j m_{kj} - (\Sigma x)_i$, and we can view the first term in $(Fx)_i$ as a sum of independent sub-Gaussian products. Therefore, again by Lemma 2.7.7 and Exercise 2.7.10 in [49], we see that $(Fx)_i$ is the average of sub-exponential random variables that have Orlicz norms upper bounded by an absolute constant multiple of $\|x\|_2 L^2$. Hence, by Bernstein inequality,

$$\mathbb{P}\left(\left|\sum_{j=1}^d F_{ij}x_j\right| > t\right) \leq 2\exp\left(-c\frac{nt^2}{L^4\|x\|_2^2}\right).$$

Then, using the union bound,

$$\mathbb{P}(\|Fx\|_\infty > t) \leq 2d\exp\left(-c\frac{nt^2}{L^4\|x\|_2^2}\right). \tag{10}$$

Taking $x = x^*$ and $t = B_1 L^2 \sqrt{\sigma \log(d)/n}$, for some large absolute constant $B_1 > 0$, we obtain **2C**.

For **2D**, we first observe that each entry of $M^\top \epsilon$ is sub-exponential with Orlicz norm upper bounded by an absolute constant multiple of $L\varrho$. Then, we can show that there exist absolute constants $C_1, C_2 > 0$ such that $\left\|\frac{1}{n} M^\top \epsilon\right\|_\infty \le C_1 \sqrt{\varrho^2 L^2 \log(d)/n}$ and $\|Fz^*\|_\infty \le C_2 \sqrt{L^4 \|z\|_2^2 \log(d)/n}$ hold w.h.p., similarly to the proof of **2B** and **2C**. By taking a large enough absolute constant $B_2 > 0$, e.g., $B_2 = 2\max\{C_1, C_2\}$, we obtain **2D**. □

We are now ready to present our sufficient conditions for solving the feature extraction problem with sub-Gaussian data. The proof of Theorem 6 is based on Theorem 4. In short, we utilize the concentration bounds in Lemma 2 and replace the random variables in Theorem 6 by their means, and then add or subtract gaps of metrics between random variables and their means, so that the conditions in Theorem 6 are still true.

Theorem 6. *Let $x^* \in \{0, \pm 1\}^d$, define $S := \mathrm{Supp}(x^*)$, and assume $|S| = \sigma$. Assume* (LM) *holds. In addition, suppose that M consists of centered row vectors $m_i \overset{i.i.d.}{\sim} \mathcal{SG}(L^2)$ for some $L > 0$ and $i \in [n]$, and we denote the covariance matrix of m_i by Σ. Assume the noise vector ϵ is a centered sub-Gaussian random vector independent of M, with each $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{SG}(\varrho^2)$ for $i \in [n]$. Let the constants c_1, B, B_1, B_2 be the same as in Lemma 2. Define $\hat{y}^* := -\Sigma z^*$, $\hat{Y}_{11}^* := -(\hat{y}_S^*)^\top x_S^*$, $\hat{\theta} := \arccos\left(\frac{(\hat{y}_S^*)^\top x_S^*}{\sqrt{\sigma}\|y_S^*\|_2}\right)$, and assume $\hat{Y}_{11}^* > 0$ and $\frac{1}{\sigma}\hat{Y}_{11}^* = \Omega(1)$. Suppose there exist $\delta > 0$ such that the following conditions are satisfied:*

- ***D1.*** *The function $f_n(x) := \sqrt{\frac{\|x\|_2^2}{(x^\top x_S^*)^2} - \frac{1}{\sigma}}$ is $\frac{\ell_n}{\sqrt{\sigma}}$-Lipschitz continuous at the point $\hat{y}_S^*$ for some constant ℓ_n;*

- ***D2.*** *$\left\|\Sigma_{S,S^c} + \frac{1}{\sigma}\hat{y}_{S^c}^*(x_S^*)^\top\right\|_\infty + BL^2\sqrt{\log(d)/n} + \frac{1}{\sigma}\lambda_n \le \hat{\mu}_3^*$ holds, where $\lambda_n := B_2 L \sqrt{(\varrho^2 + L^2 \|z^*\|_2^2)\log(d)/n}$ and*

$$\hat{\mu}_3^* := \frac{1}{\sigma}\Big\{\lambda_{\min}(\Sigma_{S,S}) - \delta + \min_{i \in S}\frac{-\hat{y}_i^* - (\Sigma x^*)_i}{x_i^*} - \lambda_n - B_1 L^2 \sqrt{\frac{\sigma \log(d)}{n}} - c_1 L \sqrt{\frac{\sigma}{n}}\Big\};$$

- ***D3.*** *There exists $\hat{\mu}_2^* \in (-\infty, -\lambda_{\min}(\Sigma_{S,S}) - c_1 L\sqrt{\frac{\sigma}{n}} + \delta]$ such that the inequality $\|\Sigma_{S^c,S^c} + \hat{\mu}_2^* I_{d-\sigma}\|_\infty + BL^2\sqrt{\frac{\log(d)}{n}} + \frac{\left(\|\hat{y}_{S^c}^*\|_\infty + \lambda_n\right)^2}{\hat{Y}_{11}^* - \sigma\lambda_n} + \gamma_n/\delta \le \hat{\mu}_3^*$ holds, where $\gamma_n := (f_n(\hat{y}_S^*) + \ell_n \lambda_n)^2 \left(\|\hat{y}_{S^c}^*\|_\infty + \lambda_n\right)^2/\delta$.*

Then, there exists a constant $C = C(\Sigma, z^*, x^*, \sigma)$ such that when

$$n \ge CL^2(\varrho^2 + L^2\|z^*\|_2^2 + \sigma)\log(d),$$

(SILS'-SDP) *recovers x^* w.h.p. as $(n, \sigma, d) \to \infty$.*

Proof. It is sufficient to check that all assumptions in this theorem imply all assumptions in Theorem 4. We define Y_{11}^*, y_S^*, μ_3^*, $\cos(\theta)$ as in Theorem 4. We define μ_3^* with the same δ here, and we take $\mu_2^* = \hat{\mu}_2^*$. Throughout the proof, we take $n \ge CL^2(\varrho^2 + L^2\|z^*\|_2^2 + \sigma)\log(d)$ for some constant C, so $\sigma/n \to 0$, $\log(d)/n \to 0$, and we can apply Lemma 2. Consequently, in the rest of the proof, we assume that **2A - 2D** in Lemma 2 hold.

We now check $Y_{11}^* > 0$. We have

$$\frac{1}{\sigma}|Y_{11}^* - \hat{Y}_{11}^*| = \frac{1}{\sigma}| - (y_S^* - \hat{y}_S^*)^\top x_S^*| \le \|y_S^* - \hat{y}_S^*\|_\infty \overset{\textbf{2D}}{<} B_2 L \sqrt{\frac{(\varrho^2 + L^2\|z^*\|_2^2)\log(d)}{n}}. \tag{11}$$

Note that the RHS is exactly λ_n. Thus, if we take $C = C(\Sigma, z^*, x^*, \sigma)$ large enough such that $\lambda_n \le \hat{Y}_{11}^*/\sigma = \Omega(1)$, we see Y_{11}^*/σ and $\hat{Y}_{11}^*/\sigma$ have the same sign, while $\hat{Y}_{11}^*/\sigma > 0$ is guaranteed by $\hat{Y}_{11}^* > 0$.

Next, we show **D2** implies **B1**. We see that

$$\begin{aligned} &\left\| \frac{1}{n}(M^\top M)_{S,S^c} + \frac{1}{\sigma} y_{S^c}^* (x_S^*)^\top - \Sigma_{S,S^c} - \frac{1}{\sigma} \hat{y}_{S^c}^* (x_S^*)^\top \right\|_\infty \\ &\le \left\| \frac{1}{n}(M^\top M)_{S,S^c} - \Sigma_{S,S^c} \right\|_\infty + \frac{1}{\sigma} \left\| y_{S^c}^* - \hat{y}_{S^c}^* \right\|_\infty < BL^2 \sqrt{\frac{\log(d)}{n}} + \frac{1}{\sigma}\lambda_n, \end{aligned} \tag{12}$$

where we use **2B** and **2D** at the last inequality. To show that **B1** is true, we first obtain

$$\begin{aligned} &\frac{1}{\sigma}\Bigg(\left| \lambda_{\min}(\Sigma_{S,S}) - \lambda_{\min}(\frac{1}{n}(M^\top M)_{S,S}) \right| \\ &+ \left| \min_{i\in S} [\hat{y}^* - \Sigma_{S,S} x_S^*]_i / x_i^* - \min_{i \in S} [-y^* - \frac{1}{n}(M^\top M)_{S,S} x_S^*]_i / x_i^* \right| \Bigg) \\ &\le \frac{c_1}{\sigma} L \sqrt{\frac{\sigma}{n}} + \frac{1}{\sigma}\Big(\max_{i\in S} |\hat{y}_i^* - y_i^*| + \max_{i\in S} |[\Sigma_{S,S} x_S^* - \frac{1}{n}(M^\top M)_{S,S} x_S^*]_i / x_i^*]_i| \Big) \\ &< \frac{c_1 L}{\sigma} \sqrt{\frac{\sigma}{n}} + \frac{1}{\sigma}\lambda_n + \frac{1}{\sigma} B_1 L^2 \sqrt{\frac{\sigma \log(d)}{n}}, \end{aligned}$$

where we use **2A** and the fact that $|\min_{i\in S}(a_i+b_i) - \min_{i\in S}(c_i+d_i)| \le \max_{i\in S}|a_i+b_i-(c_i+d_i)| \le \max_{i\in S}|a_i - c_i| + \max_{i\in S}|b_i - d_i|$ in the first inequality, and **2C** and **2D** in the last inequality. Thus,

$$\begin{aligned} \hat{\mu}_3^* &= \frac{1}{\sigma}\Bigg\{ \lambda_{\min}(\Sigma_{S,S}) - \delta + \min_{i\in S} \left[-\hat{y}_i^* - (\Sigma_{S,S} x_S^*)_i \right] / x_i^* \\ &\quad - BL\sqrt{\frac{\|z^*\|_2^2 \log(d)}{n}} - B_1 L \sqrt{\frac{\sigma \log(d)}{n}} - c_1 L \sqrt{\frac{\sigma}{n}} \Bigg\} \\ &\le \frac{1}{\sigma}\Big\{ \lambda_{\min}(\frac{1}{n}(M^\top M)_{S,S}) - \delta + \min_{i\in S} [-y^* - \frac{1}{n}(M^\top M)_{S,S} x_S^*]_i / x_i^* \Big\} = \mu_3^*. \end{aligned}$$

From the triangle inequality and (12), we obtain

$$\left\| \frac{1}{n}(M^\top M)_{S,S^c} + \frac{1}{\sigma} y_{S^c}^* (x_S^*)^\top \right\|_\infty < \left\| \Sigma_{S,S^c} + \frac{1}{\sigma} \hat{y}_{S^c}^* (x_S^*)^\top \right\|_\infty + BL\sqrt{\frac{\log(d)}{n}} + \frac{1}{\sigma}\lambda_n,$$

and we observe that the RHS is upper bounded by $\mu_3^* \le \mu_3^*$.

Next, we show that **D1** and **D3** lead to **B2**. From **2D**, we obtain that $\|\hat{y}_{S^c}^* - y_{S^c}^*\|_\infty < \lambda_n$, which implies

$$\left\| y_{S^c}^* (y_{S^c}^*)^\top \right\|_\infty = \|y_{S^c}^*\|_\infty^2 \le \big(\|\hat{y}_{S^c}^*\|_\infty + \|y_{S^c}^* - \hat{y}_{S^c}^*\|_\infty \big)^2 < \Big(\|\hat{y}_{S^c}^*\|_\infty + \lambda_n \Big)^2.$$

Combining with (11), we see

$$\left\| \frac{1}{Y_{11}^*} y_{S^c}^* (y_{S^c}^*)^\top \right\|_\infty \le \frac{\|y_{S^c}^* (y_{S^c}^*)^\top\|_\infty}{\hat{Y}_{11}^* - |Y_{11}^* - \hat{Y}_{11}^*|} < \frac{\Big(\|\hat{y}_{S^c}^*\|_\infty + \lambda_n \Big)^2}{\hat{Y}_{11}^* - \sigma\lambda_n}.$$

Next, note that $\sqrt{\frac{1-\cos^2(\hat{\theta})}{\cos^2(\hat{\theta})}} = \sqrt{\sigma} f_n(\hat{y}_S^*)$. Lipschitzness of f_n in **D1**, together with **2D**, yield

$$\sqrt{\sigma}|f_n(y_S^*) - f(\hat{y}_S^*)| \le \ell_n \|y_S^* - \hat{y}_S^*\|_2 < \sqrt{\sigma}\ell_n \lambda_n,$$

and **2B** yields

$$\left\|\Sigma_{S^c,S^c} - \frac{1}{n}(M^\top M)_{S^c,S^c}\right\|_\infty \le BL^2\sqrt{\frac{\log(d)}{n}}.$$

Combining the above three inequalities, we obtain that

$$\begin{aligned}
&\left\|\frac{1}{n}(M^\top M)_{S^c,S^c} + \mu_2^* I_{d-\sigma}\right\|_\infty + \left\|\frac{1}{Y_{11}^*} y_{S^c}^*(y_{S^c}^*)^\top\right\|_\infty + \frac{1}{\sigma}\frac{1-\cos^2(\theta)}{\cos^2(\theta)}\|y_{S^c}^*\|_\infty^2 \\
&< \|\Sigma_{S^c,S^c} + \hat\mu_2^* I_{d-\sigma}\|_\infty + BL\sqrt{\frac{\log(d)}{n}} + \frac{\Big(\|\hat y_{S^c}^*\|_\infty + \lambda_n\Big)^2}{\hat Y_{11}^* - \sigma\lambda_n} \\
&\quad + \frac{1}{\delta\sigma}\Big(\sqrt{\sigma} f(\hat y_S^*) + \sqrt{\sigma}\ell_n\lambda_n\Big)^2\Big(\|\hat y_{S^c}^*\|_\infty + \lambda_n\Big)^2 \overset{\textbf{D3}}{\le} \hat\mu_3^* \le \mu_3^*.
\end{aligned}$$

We have shown that the second part of **B2** is true. $\mu_2^* \in (-\infty, -\lambda_{\min}\big(\frac{1}{n}(M^\top M)_{S,S}\big) + \delta]$ follows from **2A**, and hence the first part of **B2** is also true. □

Remark. Condition **D1** is not very restrictive. In fact, in some cases, it can be easily fulfilled. For example, in the case where $x_S^* = \mathrm{sign}(\hat y_S^*)$, the assumption $\frac{1}{\sigma}\hat Y_{11}^* = \frac{1}{\sigma}(-\hat y_S^*)^\top x_S^* = \Omega(1)$ in Theorem 6 guarantees condition **D1**. Indeed, we see

$$\nabla_i f(x) = \frac{1}{2\sqrt{\frac{\|x\|_2^2}{[x^\top x_S^*]^2} - \frac{1}{\sigma}}} \cdot \frac{2x_i[x^\top x_S^*]^2 - 2x_i^*[x^\top x_S^*]\,\|x\|_2^2}{[x^\top x_S^*]^4} = \frac{x_i[x^\top x_S^*] - x_i^*\,\|x\|_2^2}{[x^\top x_S^*]^3\sqrt{\frac{\|x\|_2^2}{[x^\top x_S^*]^2} - \frac{1}{\sigma}}},$$

and hence $\|\nabla f_n(x)\|_2 = \frac{\sqrt{\sigma}\|x\|_2}{[x^\top x_S^*]^2}$. Using Taylor's expansion, there exists some $\eta \in [0,1]$ such that $|f_n(\hat y_S^*) - f_n(\hat y_S^*)| \le \|\nabla f_n(\hat y_S^* + \eta(\hat y_S^* - y_S^*))\|_2\,\|\hat y_S^* - y_S^*\|_2$. As long as $\|y_S^* - \hat y_S^*\|_\infty$ is sufficiently small such that $\mathrm{sign}(y_S^*) = \mathrm{sign}(\hat y_S^*)$ and $\frac{1}{\sigma}[\hat y_S^* + \eta(\hat y_S^* - y_S^*)]^\top x_S^* = \Omega(1)$, we have

$$\begin{aligned}
\|\nabla f_n(\hat y_S^* + \eta(\hat y_S^* - y_S^*))\|_2 &= \frac{\sqrt{\sigma}}{|[\hat y_S^* + \eta(\hat y_S^* - y_S^*)]^\top x_S^*|} \cdot \frac{\|\hat y_S^* + \eta(\hat y_S^* - y_S^*)\|_2}{\|\hat y_S^* + \eta(\hat y_S^* - y_S^*)\|_1} \\
&\le \frac{\sqrt{\sigma}}{|[\hat y_S^* + \eta(\hat y_S^* - y_S^*)]^\top x_S^*|} = \mathcal{O}(\frac{1}{\sqrt{\sigma}}),
\end{aligned}$$

and hence we obtain **D1**.

In the opposite case, where $x_S^* \ne \mathrm{sign}(\hat y_S^*)$, some additional but realistic conditions can be assumed to guarantee **D1**. A possible case is that the function $g(x) := \frac{\|x\|_2}{|x^\top x_S^*|}$ is upper bounded by some absolute constant $c > 0$ at $x = \hat y_S^*$, and $\min_{i\in S}|\hat y_i^*| = \Omega(1)$. Intuitively, the first assumption is equivalent to saying that the unit direction vector of $\hat y_S^*$ is not nearly orthogonal to x_S^*, and the second assumption is equivalent to saying that the vector $\Sigma_{S,[d]}z^*$ is bounded away from zero. Since $\min_{i\in S}|\hat y_i^*| = \Omega(1)$, when $\|y_S^* - \hat y_S^*\|_\infty$ is sufficiently small, then $[\hat y_S^* + \eta(\hat y_S^* - y_S^*)]^\top x_S^* \ge \frac{1}{2}|(\hat y_S^*)^\top x_S^*|$ and $\|\hat y_S^* + \eta(\hat y_S^* - y_S^*)\|_2 \le 2\|\hat y_S^*\|_2$ hold. Combining the assumption $\frac{1}{\sigma}\hat Y_{11}^* = \Omega(1)$, we obtain **D1** from the fact

$$\begin{aligned}
\|\nabla f_n(\hat y_S^* + \eta(\hat y_S^* - y_S^*))\|_2 &= \frac{\sqrt{\sigma}}{|[\hat y_S^* + \eta(\hat y_S^* - y_S^*)]^\top x_S^*|} \cdot g(\hat y_S^* + \eta(\hat y_S^* - y_S^*)) \\
&\le \frac{2\sqrt{\sigma}}{|(\hat y_S^*)^\top x_S^*|} \cdot 4\frac{\|\hat y_S^*\|_2}{|(\hat y_S^*)^\top x_S^*|} \le \frac{8c}{\Omega(\sqrt{\sigma})}.
\end{aligned}$$

6.1.1 A data model for the feature extraction problem

In this section, we study a concrete data model for the feature extraction problem and we show that it can be solved by (SILS'-SDP) with high probability, due to Theorem 6. We now define our first data model, in which the m_i's are standard Gaussian vectors.

Model 1. Assume that (LM) holds, where the input matrix M consists of i.i.d. centered random entries drawn from $\mathcal{SG}(1)$, and where the noise vector ϵ is centered and is sub-Gaussian independent of M, with $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{SG}(\varrho^2)$. We assume the ground truth vector z^* satisfies $\|z^*\|_\infty \leq u$ for some absolute constant $u > 0$. We additionally assume $|z_1^*| \geq |z_2^*| \geq \cdots \geq |z_d^*|$, and that $|z_\sigma^*| \geq 1 + g$, and $|z_{\sigma+1}^*| < 1$ for some absolute constants $g > 0$. Finally, we assume z^* satisfies

$$\sigma \sum_{i=1}^{\sigma} |z_i^*|^2 \leq \left(\frac{g^2}{2(g+1)} + 1 \right) \left(\sum_{i=1}^{\sigma} |z_i^*| \right)^2 \tag{13}$$

Model 1 can be viewed as follows: M is a normalized real-world sub-Gaussian data matrix (for each entry of the real-world data matrix, we subtract the column mean and then divide by the column standard deviation) with independent columns, and z^* is a feature vector, with the σ most significant features having feature significance that is at least $g > 0$ more than those $d-\sigma$ less significant features. Lastly, (13) can be seen as a reversed Cauchy Schwartz inequality, which guarantees that the most significant σ components do not 'spread' too far away from each other. One can see that (13) holds if g is sufficiently large. In computer vision, we can view a Gaussian M as an image, which is a simplified yet natural assumption [38], and we view the vector z^* as the relationship among the center pixel and the pixels around [52]. It is worthy pointing out that, existing algorithms generally take an exponential running time [5, 57] due to the fact that z^* is not sparse.

Note that, in Model 1, it is not realistic to assume that the largest components of z^* are all in the first σ components. Rather, we should consider the more general model where the components of z^* are arbitrarily permuted. However, this assumption on z^* in the model can be made without loss of generality. In fact, (SILS'-SDP) can solve Model 1 if and only if it can solve the more general model. This is because both (SILS'-SDP) and the model are invariant under permutation of variables. A similar note applies to Models 2 and 3 that will be considered later. In addition, it is not realistic to assume that all less significant features are less than or equal to one, but this can be done by a proper scaling of input (M, b), at the cost of a scaling of noise variance ρ.

In our next theorem, we show that (SILS'-SDP) solves (SILS') with high probability provided that n is sufficiently large. The numerical performance of (SILS'-SDP) under Model 1 will be demonstrated and discussed in Section 7.2.

Theorem 7. *Consider the feature extraction problem under Model 1. Then, there exists an absolute constant C such that when*

$$n \geq C\big(\sigma^2 + d + \varrho^2\big) \log(d),$$

(SILS'-SDP) *solves* (SILS') *w.h.p. as* $(n, d) \to \infty$.

Proof. Let $x_i^* = \begin{cases} \text{sign}(z_i^*),\ i \leq \sigma, \\ \quad 0, \text{otherwise,} \end{cases}$ and $S := [\sigma]$. In this proof, we employ Theorem 6 to prove that (SILS) recovers x^* when n is large enough, by checking all the assumptions therein. We observe that $L = 1$ when $\Sigma = I_d$. We also have $\hat{y}_S^* = -z_S^*$ and $\hat{Y}_{11}^*/\sigma = (x_S^*)^\top I_d z_S^*/\sigma \geq g + 1 = \Omega(1)$. Throughout the proof, we take $n \geq C(\|z^*\|_2^2 + \sigma^2 + \varrho^2)\log(d)$ for some absolute constant $C > 0$. For brevity, we say that n is sufficiently large if we take a sufficiently large C.

For **D1**, we first show that $l_n = \mathcal{O}(1/\sqrt{\sigma})$ if n is large enough. By Section 6.1, we see that for some $\eta \in [0,1]$,

$$l_n \le \frac{\sigma}{|[\hat{y}_S^* + \eta(\hat{y}_S^* - y_S^*)]^\top x_S^*|} \cdot \frac{\|\hat{y}_S^* + \eta(\hat{y}_S^* - y_S^*)\|_2}{\left\|\hat{y}_S^* + \eta(\hat{y}_S^* - y_S^*)\right\|_1}.$$

For ease of notation, we denote $\lambda_n := B_2\sqrt{(\varrho^2 + \|z^*\|_2^2)\log(d)/n}$. From **2D**, we obtain

$$\begin{aligned}|[\hat{y}_S^* + \eta(\hat{y}_S^* - y_S^*)]^\top x_S^*| &\ge |(\hat{y}_S^*)^\top z_S^*| - |(\hat{y}_S^* - y_S^*)^\top x_S^*| \ge 2(1+g)\sigma - \sigma\,\|\hat{y}_S^* - y_S^*\|_\infty \\ &\ge \sigma\left(2(1+g) - \lambda_n\right).\end{aligned}$$

Using **2D** again, we have

$$\begin{aligned}\frac{\|\hat{y}_S^* + \eta(\hat{y}_S^* - y_S^*)\|_2}{\left\|\hat{y}_S^* + \eta(\hat{y}_S^* - y_S^*)\right\|_1} &\le \frac{\|\hat{y}_S^*\|_2 + \|\hat{y}_S^* - y_S^*\|_2}{\left\|\hat{y}_S^*\right\|_1 - \left\|\hat{y}_S^* - y_S^*\right\|_1} \le \frac{\sqrt{2u\sigma} + \sqrt{\sigma}\,\|\hat{y}_S^* - y_S^*\|_\infty}{2(1+g)\sigma - \sigma\left\|\hat{y}_S^* - y_S^*\right\|_\infty} \\ &\le \frac{1}{\sqrt{\sigma}} \cdot \frac{\sqrt{2u} + \lambda_n}{2(1+g) - \lambda_n}.\end{aligned}$$

Combining the above two inequalities, we see $l_n = \mathcal{O}(1/\sqrt{\sigma})$ when n is sufficiently large.

For **D2**, we set $\delta = g/2$. We obtain that

$$\hat{\mu}_3^* \ge \frac{1}{\sigma}\left(1 - \frac{g}{2} + g - \lambda_n - B_1\sqrt{\frac{\sigma\log(d)}{n}} - c_1\sqrt{\frac{\sigma}{n}}\right) > \frac{1}{\sigma}\left(1 + \frac{g}{4}\right)$$

if n is sufficiently large. Since we have $|\hat{y}_{S^c}^*| \le 1_{d-\sigma}$ and $\Sigma_{S,S^c} = O_{\sigma\times(d-\sigma)}$, we see that **D2** is true for a sufficiently large n.

To show **D3**, we set $\hat{\mu}_2^* = -1$ and we see that $\hat{\mu}_2^* = -1 \le -1 + \delta - c_1\sqrt{\sigma/n}$ holds for large n. Therefore, $\Sigma_{S^c,S^c} + \hat{\mu}_2^* I_{d-\sigma} = O_{(d-\sigma)\times(d-\sigma)}$. Moreover, (13) implies $f_n(\hat{y}_S^*)^2 = \frac{\|\hat{y}_S^*\|_2^2}{[(\hat{y}_S^*)^\top x_S^*]^2} - \frac{1}{\sigma} \le \frac{g^2}{2\sigma(g+1)}$, and hence

$$\gamma_n = (f_n(\hat{y}_S^*) + l_n\lambda_n)^2\left(\|\hat{y}_{S^c}\|_\infty + \lambda_n\right)^2 \cdot \frac{1}{\delta} \le \frac{g^2}{2\sigma(g+1)} \cdot 1 \cdot \frac{2}{g} = \frac{g}{\sigma(g+1)},$$

for sufficiently large n, where we absorb the diminishing term brought by $l_n\lambda_n$ into the term $(\|\hat{y}_{S^c}\|_\infty + \lambda_n)^2$, as $\|\hat{y}_{S^c}\|_\infty = \|z_{S^c}^*\|_\infty < 1$. It remains to check $B\sqrt{\frac{\log(d)}{n}} + \frac{(\|\hat{y}_{S^c}\|_\infty + \lambda_n)^2}{\hat{Y}_{11}^* - \sigma\lambda_n} + \frac{g}{\sigma(g+1)} \le \hat{\mu}_3^*$. By absorbing the diminishing term brought by λ_n into $\|\hat{y}_{S^c}\|_\infty < 1$, we obtain that

$$B\sqrt{\frac{\log(d)}{n}} + \frac{1}{\sigma(g+1)} + \frac{g}{\sigma(g+1)} = B\sqrt{\frac{\log(d)}{n}} + \frac{1}{\sigma}\cdot 1 < \frac{1}{\sigma}\left(1 + \frac{g}{4}\right) < \mu_3^*$$

for a sufficiently large n. Finally, we observe that $\|z^*\|_2^2 \le d + \sigma u^2$, which concludes the proof. □

In the proof of Theorem 7, we showed that, if $n \ge C\big(\sigma^2 + d + \varrho^2\big)\log(d)$, then (SILS'-SDP) solves (SILS') by recovering a special x^*, which is supported on $[\sigma]$. As we will see in Section 7.2, we observe from numerical tests that (SILS'-SDP) solves (SILS') even for smaller values of n, and the recovered sparse integer vector is not necessarily supported on $[\sigma]$. A possible explanation of this phenomenon is that the upper bounds given in Lemma 2, and used in the proof, can be large when n is not sufficiently large. The terms related to n in conditions **D2** - **D3** in Theorem 6 will no longer vanish and may become the dominating terms, causing the support set S of the optimal solution to possibly change.

6.2 Integer sparse recovery problem

In the realm of communications and signal processing, reconstruction of sparse signals has become a prominent and essential subject of study. In this section, we aim to solve the integer sparse recovery problem. Recall that, in this problem, our input M, b, σ satisfies (LM), for some $z^* \in \{0, \pm 1\}^d$ with cardinality σ, and our goal is to recover z^* correctly. As mentioned in Section 1, assuming $z^* \in \{0, \pm 1\}^d$, solving the integer sparse recovery problem is equivalent to solving the well-known sparse recovery problem.

We first give sufficient conditions for (SILS'-SDP) to recover z^*. For brevity, we denote by $H^0 := I_\sigma - z_S^*(z_S^*)^\top/\sigma$ and define

$$
\begin{aligned}
\Theta := & \frac{1}{n}(M^\top M)_{S^c,S^c} + \mu_2^* I_{d-\sigma} - \frac{(M^\top \epsilon)_{S^c}(y_S^*)^\top}{\delta n Y_{11}^*} H^0 \big(I_\sigma + \frac{y_S^*(z_S^*)^\top}{Y_{11}^*}\big)\frac{1}{n}(M^\top M)_{S,S^c} \\
& - \frac{1}{n}(M^\top M)_{S^c,S}\big(I_\sigma + \frac{z_S^*(y_S^*)^\top}{Y_{11}^*}\big)H^0 \frac{y_S^*(M^\top\epsilon)_{S^c}^\top}{\delta n Y_{11}^*} - \frac{1}{Y_{11}^*}(\frac{1}{n}M^\top\epsilon)_{S^c}(\frac{1}{n}M^\top\epsilon)_{S^c}^\top \\
& - \frac{1}{Y_{11}^*}(\frac{1}{n}M^\top\epsilon)_{S^c}\big(\frac{1}{n}(M^\top M)_{S^c,S} z_S^*\big)^\top - \frac{1}{Y_{11}^*}\big(\frac{1}{n}(M^\top M)_{S^c,S} z_S^*\big)(\frac{1}{n}M^\top\epsilon)_{S^c}^\top \\
& - \frac{1}{\delta(nY_{11}^*)^2}(M^\top\epsilon)_{S^c}(y_S^*)^\top H^0 y_S^*(M^\top\epsilon)_{S^c}^\top \\
& - \frac{1}{n^2}(M^\top M)_{S^c,S}\Big([I_\sigma + \frac{1}{Y_{11}^*} z_S^*(y_S^*)^\top]\frac{1}{\delta}H^0[I_\sigma + \frac{1}{Y_{11}^*}y_S^*(z_S^*)^\top] + \frac{z_S^*(z_S^*)^\top}{Y_{11}^*}\Big)(M^\top M)_{S,S^c}.
\end{aligned} \tag{14}
$$

In light of Theorem 3 and the model assumption (LM), we are able to derive the following sufficient conditions for recovering z^*.

Theorem 8. *Consider the integer sparse recovery problem. We denote that* $S := \mathrm{Supp}(z^*)$, $y^* := -M^\top b/n$, $Y_{11}^* := -(y_S^*)^\top z_S^*$, *and assume* $Y_{11}^* > 0$. *Then* (SILS'-SDP) *recovers* z^*, *if there exists a constant* $\delta > 0$ *such that the following conditions are satisfied:*

- ***E1.*** $\frac{1}{n\sigma}\left\|(M^\top\epsilon)_{S^c}\right\|_\infty \le \mu_3^* := \frac{1}{\sigma}\{\lambda_{\min}\big(\frac{1}{n}(M^\top M)_{S,S}\big) - \delta + \min_{i\in S}(\frac{1}{n}M^\top\epsilon)_i/z_i^*\}$;
- ***E2.*** *There exists* $\mu_2^* \in (-\infty, -\lambda_{\min}\big(\frac{1}{n}(M^\top M)_{S,S}\big) + \delta]$ *such that the matrix* Θ *defined in* (14) *can be written as the sum of two matrices* $\Theta_1 + \Theta_2$, *with* $\Theta_1 \succ 0$, $\|\Theta_2\|_\infty \le \mu_3^*$ *or* $\Theta_1 \succeq 0$, $\|\Theta_2\|_\infty < \mu_3^*$.

Proof. We intend to use Theorem 3 with $x^* = z^*$, hence we need to prove that conditions **E1** - **E2** imply **A1** - **A2**. Recall that we have $b = Mz^* + \epsilon$, and $|S| = |\mathrm{Supp}(z^*)| = \sigma$. To show **A1**, we only need to observe that

$$
-y^* - \frac{1}{n}(M^\top M)_{S,S}x_S^* = \frac{1}{n}M^\top(Mx_S^* + \epsilon) - \frac{1}{n}(M^\top M)_{S,S}x_S^* = \frac{1}{n}M^\top\epsilon,
$$

so **A1** coincides with **E1** in this setting. Then, a direct calculation shows that Θ in this theorem coincides with the one in Theorem 3 by expanding $y_{S^c}^*$. □

We observe that the assumptions in Theorem 8 do not imply that $M^\top M$ has a low coherence, the RIP, the NSP, or any other property which guarantees that Lasso or Dantzig Selector solve the sparse recovery problem. This will be evident from our computational results in Section 7.3.

Remark. The assumptions of Theorem 8 can be easily fulfilled in some scenarios. We start by claiming that **E1** is essentially weak and natural. It is met in the case where ϵ is a random noise vector independent of M when n is large, and $\lambda_{\min}((M^\top M/n)_{S,S})$ is lower bounded by some positive constant. In addition, **E1** is quite similar to the constraint in the definition of Dantzig Selector (DS), but here we only require this type of constraint for the S^c block of $M^\top\epsilon/n$. Next, Condition **E2** asks to construct Θ_1 in a way such that $\|\Theta_2\|_\infty$ is small. Note that, although **E2** is complicated and sometimes it can be challenging to give such decomposition of Θ, this assumption holds in an ideal scenario, where $(M^\top M)_{S^c,S^c}$ is large enough such that $\lambda_{\min}(\Theta) \ge 0$.

6.2.1 A data model with a high coherence for the integer sparse recovery problem

In this section, we introduce a data model for the integer sparse recovery problem that admits high coherence. The reason why we look into data models with high coherence is straightforward: by Theorem 5 and Section 5.2, (SILS'-SDP) is not expected to misidentify a certain active user with a silent user in the case where they both have low correlation, i.e., in the low coherence case. Hence, one may ask whether (SILS'-SDP) tend to make mistake when data coherence becomes higher. We will prove that our SDP relaxation (SILS'-SDP) can solve the integer sparse recovery problem under a simple yet fundamental high coherence model with high probability, as a consequence of Theorem 8. To be concrete, we study the following data model.

Model 2. Assume that (LM) holds, where the rows $m_1, m_2, \cdots, m_n$ of the input matrix M are random vectors drawn from i.i.d. $\mathcal{N}(0_d, \Sigma)$, with

$$\Sigma := \begin{pmatrix} cI_\sigma & 1_\sigma 1_{d-\sigma}^\top \\ 1_{d-\sigma} 1_\sigma^\top & c'\sigma 1_{d-\sigma} 1_{d-\sigma}^\top \end{pmatrix} + \begin{pmatrix} O_\sigma & \\ & c'' I_{d-\sigma} \end{pmatrix} := \Sigma_1 + \Sigma_2$$

for $c > 1$, $c' > 1$ and $c'' > 0$. The ground truth vector is $z^* = \begin{pmatrix} a \\ 0_{d-\sigma} \end{pmatrix}$, with $a \in \{\pm 1\}^\sigma$, and the noise vector ϵ is centered and is sub-Gaussian independent of M, with $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{SG}(\varrho^2)$.

We can interpret Model 2 as follows: the first σ independent variables (active users) sends out signal a, while the remaining variables (silent users) do nothing. Those $d - \sigma$ silent users have high correlations with the active ones, and even higher correlations among themselves. The part explained by Σ_2 states that the silent users are not the same, so the model does not reduce to a trivial model in which repeated users happen to be involved in the data set.

Though seemed to be a bit simplified and restrictive, Model 2 is in fact a baseline model for us to understand how algorithms perform under data with high coherence. A perceptual reasoning is that, one can always split a set of variables into two groups having the following property: group 1 has variables with a covariance matrix that admits a low coherence; and once any one of variables in group 2 is added to group 1, the corresponding covariance matrix of group 1 will admit a high coherence. In Model 2, we can assign the first σ active users to group 1, and assign the remaining $(d - \sigma)$ highly correlated silent users to group 2. In particular, we study the simplest case, where correlations among two different users in the same group are exactly the same, and where correlations among two users in different groups are also exactly the same. We further limit our focus to the case when users in group 1 are independent, i.e., two different users in group 1 have correlation zero, in order to quickly verify that the proposed model is valid, i.e., the covariance matrix Σ is positive semidefinite. Indeed, Lemma 5 in Section 9 and the fact that $\Sigma_{S^c,S^c} \succeq (\sigma/c) 1_\sigma 1_{d-\sigma}^\top$ together imply that $\Sigma \succeq 0$.

As Model 2 is a model with highly correlated users, and $\mu(M^\top M) = \Omega(1)$ when n is sufficiently large, we see that Model 2 does not have a low coherence. Moreover, Model 2 does not satisfy the mutual incoherence property, since $\|(M^\top M)_{S^c,S}(M^\top M)_{S,S}^{-1}\|_{\infty\to\infty} = \Omega(\sigma) > 1$ when n is sufficiently large. The above two facts follow from **2A** and **2B** in Lemma 2. The aforementioned properties are known to be crucial for ℓ_1-based convex relaxation algorithms like Dantzig Selector and Lasso to recover z^*. Though the intuition behind Model 2 may seem naive, we find that numerically, these two algorithms indeed give a high prediction error in this model, as we will discuss in Section 7.3. However, the following theorem shows that our semidefinite relaxation (SILS'-SDP) can recover z^* with high probability.

Theorem 9. *Consider the integer sparse recovery problem under Model 2. Then, there exists a constant $C = C(c, c', c'')$ such that when*

$$n \geq C\sigma^2 \varrho^2 \log(d),$$

(SILS-SDP) *recovers z^* w.h.p. as $(n, \sigma, d) \to \infty$.*

The proof of Theorem 9 is given in Section 10 and the numerical performance of (SILS-SDP) under Model 2 is presented in Section 7.3.

6.2.2 A data model with a low coherence for the integer sparse recovery problem

In this section, we show that (SILS'-SDP) can solve the integer sparse recovery problem also under some low coherence data models. Here, we focus on the following data model, which is a generalized version of the model studied in [40].

Model 3. Assume that (LM) holds, where the input matrix M consist of i.i.d. random entries drawn from $\mathcal{SG}(1)$, the ground truth vector is $z^* = \begin{pmatrix} a \\ 0_{d-\sigma} \end{pmatrix}$, with $a \in \{\pm 1\}^\sigma$, and the noise vector ϵ is centered and is sub-Gaussian independent of M, with $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{SG}(\varrho^2)$.

From **2A** and **2B** in Lemma 2, we can see that when $n = \Omega(\sigma^2 \log(d))$, the mutual incoherence property holds in Model 3, i.e., $\|(M^\top M)_{S^c,S}(M^\top M)^{-1}_{S,S}\|_{\infty\to\infty}$ is indeed strictly less than one. At the same time, Model 3 admits a low coherence, so it is known that algorithms like Lasso and Dantzig Selector can recover z^* efficiently [50, 30]. As a similar result, we show in the next theorem that (SILS'-SDP) can recover z^* when $n = \Omega((\sigma^2 + \varrho^2)\log(d))$. While this result can be proven using Theorem 5 or Theorem 8, in our proof we use Theorem 6 instead. This is because, although Theorem 6 is tailored to the feature extraction problem, it leads to a cleaner proof. In Section 7.4, we will demonstrate the numerical performance of (SILS'-SDP) under Model 3 and we will compare it with (Lasso) and (DS).

Theorem 10. *Consider the integer sparse recovery problem under Model 3. There exists an absolute constant C such that when*

$$n \geq C\big(\sigma^2 + \varrho^2\big)\log(d),$$

(SILS'-SDP) *recovers z^* w.h.p. as $(n, d) \to \infty$.*

Proof. We prove this proposition using Theorem 6. Note that $L = \mathcal{O}(1)$ when $\Sigma = I_d$. We first see $\hat{y}^*_S = -z^*_S$ and then $\hat{Y}^*_{11}/\sigma = (z^*_S)^\top I_d z^*_S/\sigma = 1$. Throughout the proof, we take $n \geq C\big(\sigma^2 + \varrho^2\big)\log(d)$ for some absolute constant C. For brevity, we say that n is sufficiently large if we take a sufficiently large C.

For condition **D1**, we can show that $l_n = \mathcal{O}(1/\sqrt{\sigma})$ if n is large enough, in a similar way to the proof of Theorem 7.

For **D2**, we set $\delta = 1/2$, and obtain

$$\hat{\mu}^*_3 = \frac{1}{\sigma}\Big(\frac{1}{2} - B_2\sqrt{\frac{(\varrho^2 + \|z^*\|_2^2)\log(d)}{n}} - B_1\sqrt{\frac{\sigma\log(d)}{n}} - c_1\sqrt{\frac{\sigma}{n}}\Big\}\Big) > \frac{1}{4\sigma}$$

when n is sufficiently large. Since $\hat{y}^*_{S^c} = 0_{d-\sigma}$ and $\Sigma_{S,S^c} = O_{\sigma\times(d-\sigma)}$, **D2** indeed holds for n sufficiently large.

To show **D3**, we first set $\hat{\mu}^*_2 = -1$. This is indeed a valid choice because we require $\hat{\mu}^*_2 \leq -1/2 - c_1\sqrt{\sigma/n}$, and when n is sufficiently large we can enforce $c_1\sqrt{\sigma/n} \leq 1/4$. Therefore, we see $\Sigma_{S^c,S^c} + \hat{\mu}^*_2 I_{d-\sigma} = O_{d-\sigma}$. Furthermore, since $x^*_S = z^*_S$, we have $\cos(\hat{\theta}) = 1$, and it remains to check whether $B\sqrt{\frac{\log(d)}{n}} + \frac{\lambda_n^2}{\hat{Y}^*_{11} - \sigma\lambda_n} + 2\ell_n^2\lambda_n^4 \leq \hat{\mu}^*_3$. It is clear that

$$B\sqrt{\frac{\log(d)}{n}} + \frac{B_2^2\frac{(\varrho^2+\sigma)\log(d)}{n}}{\sigma\big(1 - B_2\sqrt{\frac{(\varrho^2+\sigma)\log(d)}{n}}\big)} + 2\ell_n^2 B_2^4\frac{(\varrho^2+\sigma)^2\log^2(d)}{n^2} \leq \frac{1}{4\sigma} < \mu^*_3$$

is indeed true for a sufficiently large n. □

An algorithm proposed in [36] shows that it is possible to recover z^* efficiently when the entries of M are i.i.d. standard Gaussian random variables at sample complexity $n = \Omega(\sigma \log(ed/\sigma) + \varrho^2 \log(d))$. In Theorem 10, we show that we need $n = \Omega((\sigma^2 + \varrho^2)\log(d))$ many samples. The differences between these results are that: (1) we recover the integer vector z^* exactly, while [36] recovers an estimator of z^*; (2) our method is more general, since theirs may not extend to the sub-Gaussian setting. We view such difference of sample complexity as a trade off to obtain integrality and a more general setting.

7 Numerical tests

In this section, we discuss the numerical performance of our SDP relaxations (SILS-SDP) and (SILS'-SDP). We first report the performance of Algorithm 1, given an (approximated) optimal solution to (SILS-SDP), under a binary quadratic optimization benchmark [6]. We then report the numerical performance of (SILS'-SDP) under the data models which are studied in Section 6. We also compare the performance of (SILS'-SDP) with other known convex relaxation algorithms. Recall that, in Section 6, we considered two problems: the feature extraction problem and the integer sparse recovery problem. For the feature extraction problem, we are not aware of other convex relaxation algorithms, and hence we solely report the performance of (SILS'-SDP) for Model 1. For the integer sparse recovery problem, we report the performance of (SILS'-SDP) for Model 2 and Model 3, and we compare its performance with Lasso and Dantzig Selector, which are the most studied convex relaxation algorithms for the sparse recovery problem. Lasso and Dantzig Selector are defined by

$$z^{Lasso} := \arg\min \frac{1}{2n} \|Mz - b\|_2^2 + \lambda \|z\|_1, \qquad \text{(Lasso)}$$

$$z^{DS} := \mathop{\arg\min}_{\text{s.t. } \left\|M^\top (Mz-b)\right\|_\infty \leq \eta} \|z\|_1, \qquad \text{(DS)}$$

where λ and η are parameters to be chosen. All calculations of convex programs are made via CVX v2.2, a package for solving convex optimization problems [23] implemented in Matlab, with solver Mosek 9.2 [3]. All Mixed Integer quadratic programs are solved via Gurobi 10.0 [24] with its Matlab interface.

7.1 Performance of Algorithm 1

In this section, we present the performance of (SILS-SDP) by showing the numerical test results of Algorithm 1 - our proposed randomized algorithm in Section 4. We test the performance of Algorithm 1 under a Binary Quadratic Programming (BQP) benchmark maintained by J E Beasley [6]. We need to clarify that the benchmark is not initially intended for (SILS) or (SBQP), but we believe using the data therein will provide interested readers a sense of how Algorithm 1 performs under real-world data sets. We utilize the symmetric matrices provided therein as P in (SBQP), and zero out all negative entries on diagonal to keep aligned with the assumptions in Theorem 2. Note that the matrix P is not necessarily positive semidefinite, and hence (SBQP) can be a non-convex problem. Since the vector c is not provided by the benchmark data set, we generate it as a random vector $c \sim \mathcal{N}(0_d, I_d)$. Due to the large number of testing problems, we only report the performance on the first two benchmark data sets, for different sets of σ.

We summarize the results in Tables 1 to 3, where we take $T = \sqrt{\log d}$ and $C = 0.1$ as input threshold constants in Algorithm 1. After solving an (approximated) optimal solution to (SILS-SDP), we run Algorithm 1 for a thousand times, and report the mean value of objective value for $\bar{x}$ in (SBQP) (mean val), and also report the best $\bar{x}$ that is feasible to (SBQP) and that achieves the minimum objective value (best val). Since we need to find out the optimal value of

(SBQP) and SDP(c, P), we also report the running time of these two programs for interested readers. The time limit for (SBQP) is 45000 seconds, and we report the MIP gap generated by Gurobi as well. It should be pointed out that running time comparison is not the main focus of this paper, as the main focus of this paper is the approximability and even the solvability of (SILS) and (SILS') in polynomial time. It can be seen that the optimality gap indeed holds in Theorem 2. Moreover, we are surprised to see that best value obtained by Algorithm 1 seems to differ from the true optimal value by a constant multiple, which suggests that Algorithm 1 is more practical than what Theorem 2 states.

	(SBQP)			(SILS-SDP)		Algorithm 1	
	optval	time	mipgap	optval	time	mean val	best val
$\sigma = 2$	-197.26	0.35	0	-201.42	2.14	-1.59	-185.09
	-200.73	0.13	0	-213.89	2.19	-2.89	-186.04
$\sigma = 5$	-830.42	0.17	0	-936.11	2.41	-13.25	-778.68
	-935.56	0.17	0	-1002.38	3.30	-14.04	-661.71
$\sigma = 10$	-1743.66	1.12	0	-2112.93	3.98	-21.15	-1113.5
	-2327.86	0.21	0	-2509.01	3.57	-30.78	-1362.18
$\sigma = 20$	-3692.45	4.64	0	-4324.59	3.15	-58.67	-2576.74
	-4902.50	0.30	0	-5356.67	3.06	-77.17	-3530.11

Table 1: Performance under BQP50 ($d = 50$)

	(SBQP)			(SILS-SDP)		Algorithm 1	
	optval	time	mipgap	optval	time	mean val	best val
$\sigma = 2$	-202.11	1.33	0	-253.88	31.27	-3.19	-198.54
	-205.17	1.00	0	-218.26	38.59	-1.49	-196.10
$\sigma = 5$	-1062.05	5.41	0	-1225.52	52.14	-15.63	-588.83
	-944.07	12.36	0	-1052.31	56.01	-9.17	-584.16
$\sigma = 10$	-2470.12	255.83	0	-2897.89	74.83	-32.09	-1535.83
	-2254.50	472.24	0	-2647.47	53.38	-18.81	-905.34
$\sigma = 20$	-5445.56	44254.76	0	-6457.71	54.35	-56.92	-2308.37
	-5146.21	40999.62	0	-6175.32	42.63	-54.23	-1899.83

Table 2: Performance under BQP100 ($d = 100$)

7.2 Performance under Model 1

In this part, we show the numerical performance of (SILS'-SDP) in the feature extraction problem under Model 1, as studied in Section 6.1.1. We assume that the entries of M in Model 1 are i.i.d. standard Gaussian, and $\epsilon \sim \mathcal{N}(0_d, \varrho^2 I_d)$. For simplicity, we take the first σ entries of z^* to be ± 2, and the remaining entries to be ± 1. Note that (13) indeed holds in this case. In Figure 1, we first validate Theorem 7 numerically, by plotting the *empirical probability of recovery*, i.e., the percentage of times (SILS'-SDP) solves (SILS') over 100 instances, for each $n = \lceil cd \log(d) \rceil$, with control parameter c ranging from 0.25 to 4. Note that, here, $d \log(d)$ is the dominating term in the lower bound on n in Theorem 7. As discussed after Theorem 7, for small values of n, the recovered sparse integer vector is not necessarily the vector x^* in the proof of Theorem 7. In Figure 2, we then plot the *empirical probability of recovery of* x^*, i.e., the percentage of times (SILS'-SDP) recovers x^* over 100 instances. The instances considered in Figure 2 are identical to those considered in Figure 1. As shown in Figures 1 and 2, both

	(SBQP)			(SILS-SDP)		Algorithm 1	
	optval	time	mipgap	optval	time	mean val	best val
$\sigma = 2$	-205.73	16.60	0	-261.00	2772.25	-3.72	-200.41
	-207.92	8.54	0	-245.64	2877.39	-4.75	-197.25
$\sigma = 5$	-1250.38	2866.28	0	-1353.48	3335.86	-15.15	-728.86
	-1202.20	4366.75	0	-1287.54	4854.88	-13.02	-980.39
$\sigma = 10$	-3037.26	45000	46.9%	-3599.40	4891.06	-17.82	-1333.36
	-2919.55	45000	55.3%	-3741.36	4285.94	-23.92	-1051.18
$\sigma = 20$	-7363.80	45000	48.3%	-8970.73	4368.48	-41.21	-2717.95
	-6871.36	45000	56.9%	-8648.22	3615.02	-37.63	-2258.16

Table 3: Performance under BQP250 ($d = 250$)

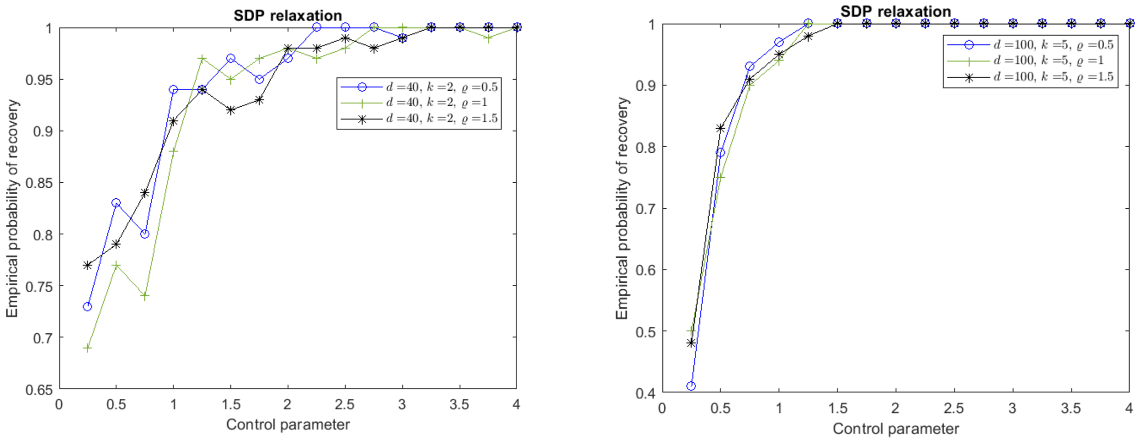

Figure 1: Performance of (SILS'-SDP) under Model 1: empirical probability of recovery.

the empirical probability of recovery and the empirical probability of recovery of x^* go to 1 as c grows larger. However, the empirical probability of recovery is much closer to one also for small values of c.

7.3 Performance under Model 2

In this part, we show how (SILS'-SDP) performs numerically in the integer sparse recovery problem under Model 2, as studied in Section 6.2.1. We take $c = 1.2$, $c' = 1.05$, and $c'' = 1$ in the covariance matrix Σ, and we take $\epsilon \sim \mathcal{N}(0_d, \varrho^2 I_d)$.

In Figure 3, we study the setting where $z^* = \begin{pmatrix} a \\ 0_{d-\sigma} \end{pmatrix}$ with a uniformly drawn in $\{\pm 1\}^\sigma$. We plot the empirical probability of recovery of z^* for each $n = \lceil c\varrho^2\sigma^2 \log(d) \rceil$, with control parameter c ranging from 1 to 15. As predicted in Theorem 9, when c is large enough, the empirical probability of recovery of z^* goes to 1 as the control parameter c increases. Empirically, we also observe there is a transition to failure of recovery when the control parameter c is sufficiently small.

In Figure 4, we restrict ourselves to the setting where $z^* = \begin{pmatrix} 1_\sigma \\ 0_{d-\sigma} \end{pmatrix}$, and we compare the performance of (SILS'-SDP), (Lasso), and (DS). We are particularly interested in this setting as it is explicitly shown in [50] that Lasso is not guaranteed to perform well. This is still a high coherence model and no guarantee on the performance of Dantzig Selector is known for this

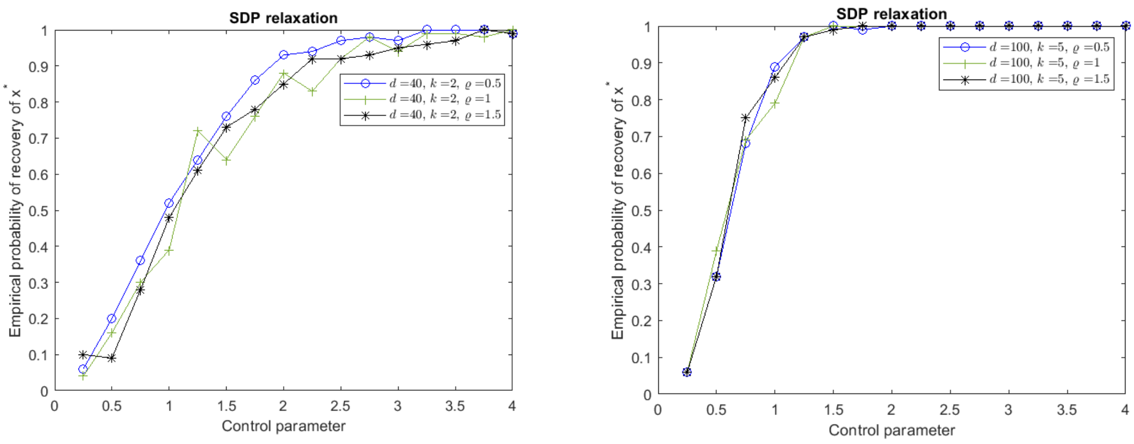

Figure 2: Performance of (SILS'-SDP) under Model 1: empirical probability of recovery of x^*.

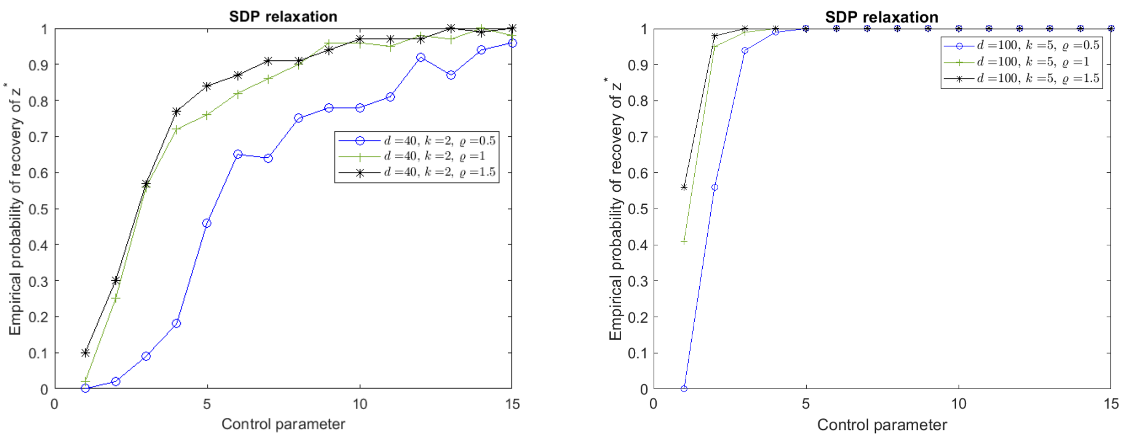

Figure 3: Performance of (SILS'-SDP) under Model 2: empirical probability of recovery of z^*.

model. The parameters λ in (Lasso) and η in (DS) are determined via a 10-fold cross-validation on a held out validation set, as suggested in [7]. We report three significant quantities for sparse recovery problems, which evaluate the quality of the solution vector z returned by the algorithm. For (SILS'-SDP), the vector z that we evaluate is the vector w^* obtained from the first column of the optimal solution W^* to (SILS'-SDP), by deleting its first entry equal to one. The first quantity that we report is the *number of nonzeros*, which is $|\operatorname{Supp}(z)|$ and measures how sparse a solution is. The second quantity that we report is the *true positive rate*, defined as

$$\text{true positive rate}(z) := \frac{|\operatorname{Supp}(z^*) \cap \operatorname{Supp}(z)|}{|\operatorname{Supp}(z^*)|}.$$

This quantity measures how well z recovers the ground truth sparse vector z^* by evaluating how much their support sets overlap. The last quantity that we report, which is suggested in [7], is known as *prediction error*, which is defined as

$$\text{prediction error}(z) := \frac{\|M(z - z^*)\|_2^2}{\|Mz^*\|_2^2}.$$

As discussed in [7], the prediction error takes into account the correlation of features and is a meaningful measure of error for algorithms that do not have performance guarantee. We

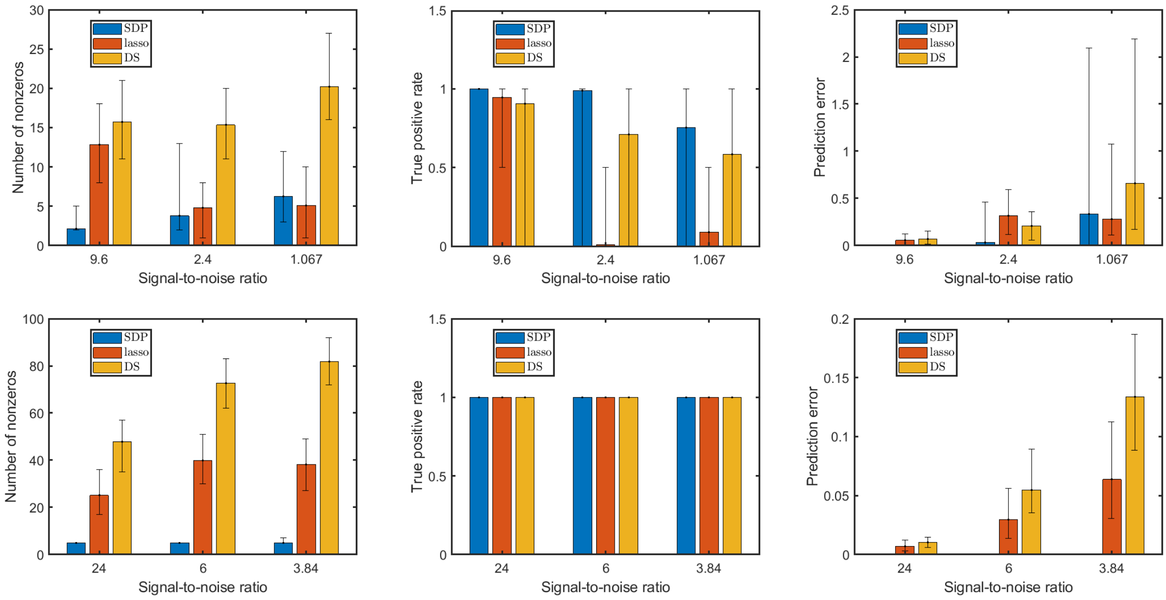

Figure 4: Performance of (SILS'-SDP), (Lasso), (DS) under Model 2, with $d = 40$, $\sigma = 2$, $n = \lceil 2\sigma^2 \log(d) \rceil = 30$ in the first row, and with $d = 100$, $\sigma = 5$, $n = \lceil 2\sigma^2 \log(d) \rceil = 231$ in the second row. 100 instances are considered with $\varrho \in \{0.5, 1, 1.5\}$. The average is reported in the histogram, and the minimum and maximum in the box plot.

report these three quantities under different *signal-to-noise ratios*, i.e.,

$$\text{signal-to-noise ratio} := \frac{\operatorname{Var}(m_i^\top z^*)}{\varrho^2} = \frac{\left\| \Sigma_{[d],S}^{\frac{1}{2}} z_S^* \right\|_2^2}{\varrho^2}.$$

In Figure 4, we study two sets of (d, σ), namely, $(d, \sigma) \in \{(100, 5), (40, 2)\}$, with $\varrho \in \{0.5, 1, 1.5\}$, and we fix our choice of n to be $\lceil 2\sigma^2 \log(d) \rceil$.

In an underdetermined system ($d > n$), plotted in the first row of Figure 4, we conclude that the probability that Lasso and Dantzig Selector recover the true support $[\sigma]$ of z^* is low, while (SILS'-SDP) nearly always recovers the true support, even when signal-to-noise ratio is low. In an overdetermined system ($d < n$), plotted in the second row of Figure 4, the true positive rates of Lasso and Dantzig Selector dramatically improve, however they are still inferior to (SILS'-SDP) in terms of number of nonzeros and prediction error.

We remark that, Model 2 is just one example of a high coherence model for the sparse recovery problem under which (SILS'-SDP) works better than (Lasso) and (DS). For instance, we observe the same behavior in a model introduced in [7] (see Example 1 therein for details). For this model, several methods including Lasso, tend to give a solution with an excessively large support set, and cannot provide a satisfactory prediction error (see Fig. 4. therein for details). On the other hand, for (SILS'-SDP), as n grows, the empirical probability of recovery of z^* tends to one, and the conditions in Theorem 8 can be satisfied.

7.4 Performance under Model 3

In this part, we study the numerical performance of (SILS'-SDP) in the integer sparse recovery problem under Model 3, as studied in Section 6.2.2. Note that Model 3 has a low coherence when $n \geq \sigma^2 \log(d)$. We restrict ourselves to the scenario where each entry of M is i.i.d. standard Gaussian, $z^* = \begin{pmatrix} a \\ 0_{d-\sigma} \end{pmatrix}$ with a uniformly drawn in $\{\pm 1\}^\sigma$, and $\epsilon \sim \mathcal{N}(0_d, \varrho^2 I_d)$.

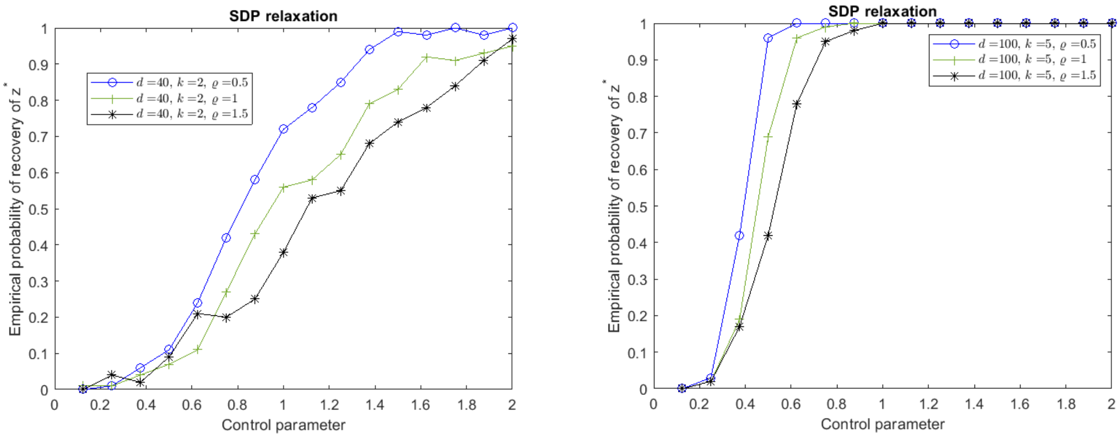

Figure 5: Performance of (SILS'-SDP) under Model 3: empirical probability of recovery of z^*.

In Figure 5, we plot the empirical probability of recovery of z^*, for each $n = \lceil c(\sigma^2 + \varrho^2)\log(d)\rceil$ with control parameter c ranging from $1/8$ to 2. As predicted in Proposition 10, when c grows, the probability that (SILS'-SDP) recovers z^* goes to 1. Empirically, we also observe that there is a transition to failure of recovery when the control parameter c is sufficiently small.

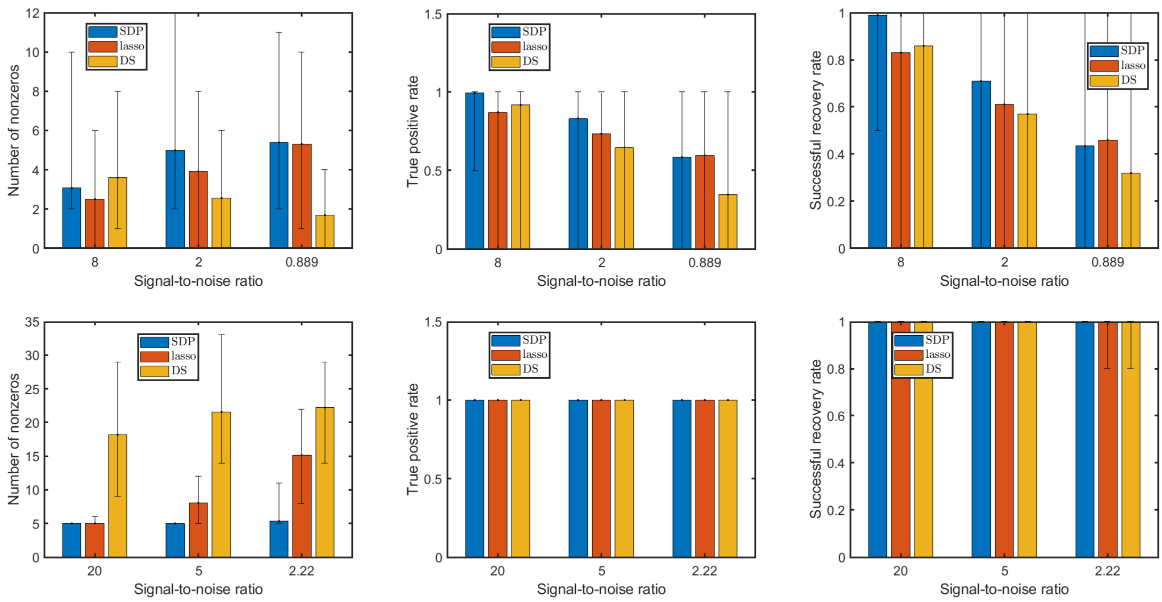

Figure 6: Performance of (SILS'-SDP), (Lasso), and (DS) under Model 3, with $d = 40$, $\sigma = 2$, $n = \lceil \sigma^2 \log(d)\rceil = 15$ in the first row, and with $d = 100$, $\sigma = 5$, $n = \lceil \sigma^2 \log(d)\rceil = 116$ in the second row. 100 instances are considered with $\varrho \in \{0.5, 1, 1.5\}$. The average is reported in the histogram, and the minimum and maximum in the box plot.

In Figure 6, we compare the numerical performance of (SILS'-SDP), (Lasso), and (DS). From [50] and [30], we know that also (Lasso) and (DS) converge to z^*, provided that we set $\lambda = 2\sqrt{\log(d)/n}$ in (Lasso) and $\eta = 2\varrho(5/4 + \sqrt{\log(d)})$ in (DS). Hence, we set the parameters λ and η to these values without performing cross-validation. In Figure 6, we report three significant quantities: the first two are the number of nonzeros and the true positive rate, as defined in Section 7.3. The third one is the *successful recovery rate*, defined as

$$\text{successful recovery rate}(z) := \frac{|\operatorname{Supp}(z^*) \cap S^{\sigma}_{\max}(z)|}{|\operatorname{Supp}(z^*)|},$$

where $S^{\sigma}_{\max}(z)$ is the set indices corresponding to the top σ entries of z having largest absolute values. The reason we consider here the successful recovery rate instead of the prediction error, considered for Model 2, is that in all three algorithms z converges to z^* in Model 3. Hence, for n large enough, $|z_i|$ is close to 0 when $z_i^* = 0$, and $|z_j|$ is close to one if $z_j^* = \pm 1$. Hence, we can recover z^* by simply looking at the σ largest entries of $|z|$. We conclude from Figure 6 that all three algorithms obtain great results in Model 3, and this is mainly due to the low coherence of the model. Since all three algorithms perform well, (Lasso) and (DS) should be preferred since they run significantly faster than (SILS'-SDP). In particular, (SILS'-SDP) can be solved in about one second with $d = 40$ and in about one minute with $d = 100$, while the other two can be solved in less than 0.1 second in both cases.

8 Proof of Theorem 2

In this section, we prove Theorem 2. To keep aligned with the notations in Section 4, throughout this section, we will keep using the same notations introduced in Algorithm 1 and Theorem 2. Moreover, we will assume the matrix $Q(c, P) = \begin{pmatrix} 0 & -c^\top \\ -c & P \end{pmatrix}$ is 0-indexed, and denote its (i, j)-th entry by q_{ij}, $0 \le i, j \le d$. As we will see later in the proofs, u_0 is in fact a special vector, so it is worthy to distinguish it from $u_1, u_2, \ldots, u_d$, with index zero. In other sections, we will continue to assume all matrices are 1-indexed.

Recall that the problem $\mathrm{SDP}(c, P)$ is defined by replacing the objective function $1/n \cdot \mathrm{tr}(A^\top A W)$ by $\mathrm{tr}(Q(c, P)W)$ in (SILS-SDP). We first show a nice property about the first column of any feasible solution to $\mathrm{SDP}(c, P)$:

Proposition 2. *Consider any feasible solution W to $\mathrm{SDP}(c, P)$. Let the first column of W be $(1, w_x^\top)^\top$, where $w_x \in \mathbb{R}^d$. Then, $\|w_x\|_1 \le \sigma$.*

Proof. Denote $\mathcal{F}$ to be the feasible region of $\mathrm{SDP}(c, P)$, we show that the optimal value of the optimization problem $\max_{W \in \mathcal{F}} \|w_x\|_1$ is exactly σ. By symmetry of $\mathcal{F}$, the problem is equivalent to $\max_{W \in \mathcal{F}} 1_d^\top w_x$. It is clear that by taking $W^* := uu^\top$ with $u := (1, \sigma/d, \sigma/d, \cdots, \sigma/d) \in \mathbb{R}^{1+d}$, we attain a cost of σ in this problem.

We conclude the proof by showing that σ can be attained by its dual. Denote $P_0 := \begin{pmatrix} 0 & 1_d^\top/2 \\ 1_d^\top/2 & O_d \end{pmatrix}$, the primal problem is then equivalent to $\max_{W \in \mathcal{F}} \mathrm{tr}(P_0 W)$, and its dual can be derived as follows:

$$\begin{aligned}
&\min_{\substack{Y \succeq 0, \mu_1 \in \mathbb{R}, \mu_2 \ge 0, \\ \mu_3 \ge 0, p \ge 0_d}} \max_{W} \quad \mathrm{tr}(P_0 W) - \mu_1(W_{11} - 1) - \mu_2(\mathrm{tr}(W_x) - \sigma) \\
&\qquad\qquad - \mu_3(1_d^\top |W_x| 1_d - \sigma^2) - p^\top[\mathrm{diag}(W_x) - 1_d] + \mathrm{tr}(YW) \\
= &\min_{\substack{Y \succeq 0, \mu_1 \in \mathbb{R}, \mu_2 \ge 0, \\ \mu_3 \ge 0, p \ge 0_d}} \max_{W} \mathrm{tr}\left(\left(P_0 - \mu_1 \begin{pmatrix} 1 & \\ & O_d \end{pmatrix} - \begin{pmatrix} 0 & \\ & \mu_2 I_d + p \end{pmatrix} + Y\right) W\right) \\
&\qquad\qquad - \mu_3 \,\mathrm{tr}(1_d 1_d^\top |W_x|) + \mu_1 + \sigma\mu_2 + \sigma^2 \mu_3 + p^\top 1_d \\
= &\min_{\substack{Y \succeq 0, \mu_1 \in \mathbb{R}, \mu_2 \ge 0, \mu_3 \ge 0 \\ \|P_0 - R(\mu_1, \mu_2, p) + Y\|_\infty \le \mu_3}} \quad \mu_1 + \sigma\mu_2 + \sigma^2\mu_3 + p^\top 1_d,
\end{aligned}$$

where $R(\mu_1, \mu_2, p) := \begin{pmatrix} \mu_1 & \\ & \mu_2 I_d + p \end{pmatrix}$. It can be checked that the set of dual variables $\mu_1^* := \sigma/2$, $\mu_2^* := 0$, $\mu_3^* := 1/(2\sigma)$, $p^* := 0_d$, and $Y^* := vv^\top$ with $v := \sqrt{\sigma/2}(1, 1/\sigma, 1/\sigma, \cdots, 1/\sigma)^\top \in \mathbb{R}^{1+d}$ is indeed feasible to the dual problem with cost σ. □

The following lemma states some properties regarding some random variables that we introduce in Algorithm 1:

Lemma 3. *Consider Algorithm 1 and the variables therein. Denote $p_0 := 1$, then:*

3A. $\mathbb{E}z_i z_j = u_i^\top u_j$ *for those $0 \le i < j \le d$ such that $p_i, p_j > 0$.*

3B. $\mathbb{E}z_i^2 = \|u_i\|_2^2 / p_i$ *for those $1 \le i \le d$ such that $p_i > 0$.*

3C. $\mathbb{E}x_i^2 = \mathbb{E}|y_i|$ *for $1 \le i \le d$.*

3D. $\mathbb{E}x_i x_j = \mathbb{E}y_i y_j$*, for any $0 \le i < j \le d$.*

3E. *Define $P := \{i \in [d] : p_i > 0\}$, then $|P| \le \min\{d, \sigma/C^2\}$.*

3F. *For those $i, j \in P$, we have that $\mathbb{E}\left[\left|y_i y_j - z_i z_j / T^2\right|\right]$ is upper bounded by*

$$e^{-\frac{2C^2T^2}{9}} \left\{ \frac{2\|u_i\|_2^2 + 2\|u_j\|_2^2}{\sqrt{2\pi}CT} + \|u_i\|_2 \|u_j\|_2 \left[\frac{4}{\sqrt{2\pi}} \cdot \left(\frac{2T}{3} + \frac{3}{2CT} \right) + \frac{4}{\pi} \right] \right\}$$

3G. *For those $j \in P$, we have that $\mathbb{E}\left[\left|y_0 y_j - z_0 z_j / T^2\right|\right]$ is upper bounded by*

$$e^{-\frac{2C^2T^2}{9}} \left\{ \frac{2\|u_j\|_2^2}{\sqrt{2\pi}CT} + \|u_j\|_2 \left[\frac{2}{\sqrt{2\pi}} \cdot \left(\frac{2T}{3} + \frac{3}{2CT} \right) + \frac{2}{\pi} \right] \right\}$$
$$+ e^{-\frac{T^2}{2}} \left\{ \frac{2}{\sqrt{2\pi}} \frac{1}{T} + \|u_j\|_2 \cdot \left[\frac{2}{\sqrt{2\pi}} \left(T + \frac{2}{T} \right) + \frac{1}{\pi} \right] \right\}$$

Proof. **3A**, **3B**, and **3C** follow from direct calculation. We start with **3D**. We first consider the case $i = 0$. We observe that the conditional probability $\mathbb{E}[x_0 x_j | y_0, y_j]$ is exactly $y_0 y_j$. Indeed,

$$\mathbb{E}[x_0 x_j | y_0, y_j] = \left(1 \cdot \frac{1 + y_0}{2}\right) \cdot \operatorname{sign}(y_j)|y_j| + \left(-1 \cdot \frac{1 - y_0}{2}\right) \cdot \operatorname{sign}(y_j)|y_j| = y_0 y_j,$$

and then by law of total expectation we are done. Then, we assume that $i \ge 1$, and we see that

$$\mathbb{E}[x_i x_j | y_i, y_j] = (\operatorname{sign}(y_i) \cdot |y_i|) \cdot (\operatorname{sign}(y_j) \cdot |y_j|) = y_i y_j.$$

By law of total expectation, we again obtain the desired result.

To show **3E**, one only need to observe that

$$\sigma \ge \sum_{i=1}^{d} \|u_i\|_2^2 \ge \sum_{i \in P} \|u_i\|_2^2 \ge C^2 |P|.$$

Finally, we show **3F** and **3G**. The proof ideas are almost identical to that of Lemma 2 in [12], but for the completeness of the paper, we leave a proof here. We define the random event $S := \{g : |z_i| \le 1, |z_j| \le 1\}$, thus $\mathbb{E}[|z_i z_j / T^2 - y_i y_j|; S] = 0$. This implies that we only need to upper bound the difference $|z_i z_j / T^2 - y_i y_j|$ when $|z_i| > 1$ or $|z_j| > 1$. Next, due to rotational symmetry of g, we can assume WLOG that $u_i = \|u_i\|_2 \cdot (1, 0, \ldots, 0)^\top$ and $u_j = \|u_j\|_2 \cdot (a, b, 0, \ldots, 0)^\top$ with $a^2 + b^2 = 1$, without changing distributions of z_i and z_j. We first define $B := \{g : z_i > 1\}$, and denote $Z \sim \mathcal{N}(0, 1)$ to be a standard Gaussian variable. Since $g^\top u$ has the same distribution as $\|u\|_2 Z$, we see that for $i \ge 1$,

$$\mathbb{E}[|y_i y_j|; B] \le \mathbb{E}[1; B] = \mathbb{P}(B) = \mathbb{P}\left(z_i > T, \tilde{u}_i = \frac{u_i}{p_i}\right)$$
$$= p_i \mathbb{P}\left(\frac{\|u_i\|_2}{p_i} Z > T\right) = p_i \mathbb{P}\left(Z > \frac{p_i}{\|u_i\|_2} T\right) \le p_i \mathbb{P}\left(Z > \frac{2}{3} C \cdot T\right)$$

$$= p_i \int_{\frac{2}{3}CT}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathrm{d}x < p_i \cdot \frac{1}{\sqrt{2\pi}} \frac{1}{\left(\frac{2}{3}CT\right)} e^{-\frac{2C^2T^2}{9}} = \frac{\|u_i\|_2^2}{\sqrt{2\pi}} \frac{1}{CT} e^{-\frac{2C^2T^2}{9}}$$

where we use the fact that $\int_t^{+\infty} e^{-x^2/2}\mathrm{d}x < 1/t \cdot e^{-t^2/2}$ in the last line. Note that the above bound is similar when $i = 0$. We see that

$$\mathbb{E}[|y_0 y_j|; B] \leq \mathbb{E}[1; B] = \mathbb{P}(B) = \mathbb{P}(z_0 > T) = \mathbb{P}\left(g^\top u_0 > T\right)$$
$$= \mathbb{P}(Z > T) = \int_T^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathrm{d}x < \frac{1}{\sqrt{2\pi}} \frac{1}{T} e^{-\frac{T^2}{2}}$$

Then, we calculate $\mathbb{E}[|z_i z_j|; B]$. Since $\mathbb{E}[|z_i z_j|; B] = \mathbb{E}\mathbb{E}[|z_i z_j| \mathbb{1}_B | \tilde{u}_i, \tilde{u}_j]$, we calculate the conditional expectation first. We see that

$$\mathbb{E}\left[|z_i z_j| \mathbb{1}_B \middle| \tilde{u}_i = \frac{u_i}{p_i}, \tilde{u}_j = \frac{u_j}{p_j}\right] = \frac{\|u_i\|_2 \|u_j\|_2}{2\pi p_i p_j} \int_{-\infty}^{+\infty} \int_{\frac{p_i T}{\|u_i\|_2}}^{+\infty} |s(as + bt)| e^{-\frac{s^2}{2}} e^{-\frac{t^2}{2}} \mathrm{d}s \mathrm{d}t.$$

Since

$$\int_{\frac{p_i T}{\|u_i\|_2}}^{+\infty} s^2 e^{-\frac{s^2}{2}} \mathrm{d}s = -s e^{-\frac{s^2}{2}} \Bigg|_{\frac{p_i T}{\|u_i\|_2}}^{+\infty} + \int_{\frac{p_i T}{\|u_i\|_2}}^{+\infty} e^{-\frac{s^2}{2}} \mathrm{d}s$$
$$< \frac{p_i T}{\|u_i\|_2} e^{-\frac{p_i^2 T^2}{2\|u_i\|_2^2}} + \frac{1}{\left(\frac{p_i}{\|u_i\|_2} T\right)} e^{-\frac{p_i^2 T^2}{2\|u_i\|_2^2}} \leq \frac{2T}{3} e^{-\frac{2C^2T^2}{9}} + \frac{3}{2CT} e^{-\frac{2C^2T^2}{9}}$$

and

$$\int_{\frac{p_i T}{\|u_i\|_2}}^{+\infty} |s| e^{-\frac{s^2}{2}} \mathrm{d}s = e^{-\frac{p_i^2 T^2}{2\|u_i\|_2^2}} \leq e^{-\frac{2C^2T^2}{9}},$$

we obtain that

$$\mathbb{E}\left[|z_i z_j| \mathbb{1}_B\right] \leq \|u_i\|_2 \|u_j\|_2 \cdot e^{-\frac{2C^2T^2}{9}} \cdot \left[\frac{1}{\sqrt{2\pi}} \cdot \left(\frac{2T}{3} + \frac{3}{2CT}\right) + \frac{1}{\pi}\right]$$

Note that for $i = 0$, the calculation is similar, and one can obtain that

$$\mathbb{E}\left[|z_0 z_j| \mathbb{1}_B \middle| \tilde{u}_j = \frac{u_j}{p_j}\right] = \frac{\|u_j\|_2}{p_j} \int_{-\infty}^{+\infty} \int_T^{+\infty} |s(as + bt)| \frac{1}{2\pi} e^{-\frac{s^2}{2}} e^{-\frac{t^2}{2}} \mathrm{d}s \mathrm{d}t$$
$$\leq \frac{\|u_j\|_2}{p_j} \cdot \left[\frac{1}{\sqrt{2\pi}} \left(T + \frac{1}{T}\right) e^{-\frac{T^2}{2}} + \frac{1}{\pi} e^{-\frac{T^2}{2}}\right],$$

and hence

$$\mathbb{E}\left[|z_0 z_j| \mathbb{1}_B\right] \leq \|u_0\|_2 \|u_j\|_2 \cdot \left[\frac{1}{\sqrt{2\pi}} \left(T + \frac{1}{T}\right) + \frac{1}{\pi}\right] e^{-\frac{T^2}{2}}$$

By symmetry, similar upper bounds hold for random events $\{g : z_i < -1\}$, $\{g : z_j > 1\}$, and $\{g : z_j < -1\}$. Therefore, we obtain **3F** and **3G**. □

We are now ready to prove Theorem 2:

Proof of Theorem 2. We first show the approximation gap. The second inequality is due to relaxation and ϵ-optimality, and we only need to show the first. We denote $U := (u_0, u_1, \ldots, u_d) = \sqrt{W^*}$, as in Algorithm 1. We observe that $\bar{x}^\top P \bar{x} - 2c^\top \bar{x} = \sum_{i,j=0}^d q_{ij} \bar{x}_i \bar{x}_j$, and $\mathrm{tr}(Q(c, P)W^*) = \sum_{i,j=0}^d q_{ij} u_i^\top u_j$. We will split the proof into two parts:

(i) (Non-diagonal entries, i.e., $i < j$) We first assume $p_i, p_j > 0$, where p_i's are defined in Algorithm 1. By **3A**, **3D**, and the fact that $\bar{x}$ differs x only by possibly flipping a sign in Algorithm 1, we observe that

$$\begin{aligned}\frac{1}{T^2}\cdot q_{ij}u_i^\top u_j &= q_{ij}\mathbb{E}y_iy_j + q_{ij}\left(\frac{1}{T^2}\mathbb{E}z_iz_j - \mathbb{E}y_iy_j\right)\\ &= q_{ij}\mathbb{E}\bar{x}_i\bar{x}_j + q_{ij}\left(\frac{1}{T^2}\mathbb{E}z_iz_j - \mathbb{E}y_iy_j\right)\\ &\geq q_{ij}\mathbb{E}\bar{x}_i\bar{x}_j - |q_{ij}|\cdot\mathbb{E}\left[\left|y_iy_j - z_iz_j\cdot\frac{1}{T^2}\right|\right].\end{aligned}$$

For the case where, WLOG, $p_i = 0$. By the definition of p_i in Algorithm 1, it must be the case $\|u_i\|_2 \leq C$. Therefore, we obtain a trivial bound (note that $\mathbb{E}\bar{x}_i\bar{x}_j = 0$)

$$\frac{1}{T^2}\cdot q_{ij}u_i^\top u_j \geq q_{ij}\mathbb{E}\bar{x}_i\bar{x}_j - \frac{1}{T^2}\cdot|q_{ij}|\cdot|u_i^\top u_j|.$$

(ii) (Diagonal entries, i.e., $i = j$) We first study the case $p_i > 0$ ($i \geq 1$). By **3B**, **3C**, and the facts that $q_{ii} \geq 0$, $\|u_i\|_2 \geq C$, and $p_i = 2/3\cdot\|u_i\|_2^2$, we see that

$$\begin{aligned}q_{ii}\mathbb{E}\bar{x}_i^2 = q_{ii}\mathbb{E}|y_i| \leq q_{ii}\sqrt{\mathbb{E}|y_i|^2} \leq q_{ii}\sqrt{\mathbb{E}\frac{1}{T^2}|z_i|^2} = \frac{q_{ii}}{T\sqrt{p_i}}\|u_i\|_2 = \frac{\sqrt{3}q_{ii}}{\sqrt{2}T}\\ = \frac{q_{ii}}{T^2}u_i^\top u_i + q_{ii}\left(\frac{\sqrt{3}}{\sqrt{2}T} - \frac{1}{T^2}u_i^\top u_i\right)\end{aligned}$$

For the case $p_i = 0$, we again use the trivial inequality

$$\frac{1}{T^2}\cdot q_{ii}u_i^\top u_i \geq q_{ii}\mathbb{E}\bar{x}_i^2 - \frac{1}{T^2}\cdot q_{ii}\cdot u_i^\top u_i.$$

Denote the set $P := \{i \in [d] : p_i > 0\}$ the same as in **3E**, and define $g(C,T) := 1/\sqrt{2\pi}\cdot(2T/3 + 3/(2T)) + 1/\pi$. Putting (i), (ii), **3F**, and **3G** together, we see that $\mathrm{tr}(Q(c,P)W^*)/T^2$ is lower bounded by

$$\begin{aligned}&\sum_{i,j=0}^{d} q_{ij}\mathbb{E}\bar{x}_i\bar{x}_j - \sum_{\substack{i\neq j,\\ i,j\in P}}|q_{ij}|e^{-\frac{2C^2T^2}{9}}\left\{\frac{2\|u_i\|_2^2 + 2\|u_j\|_2^2}{\sqrt{2\pi}CT} + 4g(C,T)\|u_i\|_2\|u_j\|_2\right\}\\ &- 2\sum_{j\in P}|q_{0j}|\left\{e^{-\frac{2C^2T^2}{9}}\left\{\frac{2\|u_j\|_2^2}{\sqrt{2\pi}CT} + 2g(C,T)\|u_j\|_2\right\}\right.\\ &\left.+ e^{-\frac{T^2}{2}}\left\{\frac{2}{\sqrt{2\pi}}\frac{1}{T} + \|u_j\|_2\cdot\left[\frac{2}{\sqrt{2\pi}}\left(T + \frac{2}{T}\right) + \frac{1}{\pi}\right]\right\}\right\}\\ &- \frac{1}{T^2}\sum_{\substack{(i,j)\notin P\times P,\\ 0\leq i,j\leq d,(i,j)\neq(0,0)}}|q_{ij}|\cdot|u_i^\top u_j| - \sum_{i\in P}q_{ii}\left|\frac{\sqrt{3}}{\sqrt{2}T} - \frac{1}{T^2}u_i^\top u_i\right|\\ \geq &\sum_{i,j=0}^{d} q_{ij}\mathbb{E}\bar{x}_i\bar{x}_j - B\cdot e^{-\frac{2C^2T^2}{9}}\left\{\frac{4\sigma\min\{d,\sigma/C^2\}}{\sqrt{2\pi}CT} + \frac{4\sigma}{\sqrt{2\pi}CT} + 4g(C,T)(\sigma+1)\right\}\\ &- B\cdot e^{-\frac{T^2}{2}}\left\{\frac{2|P|}{\sqrt{2\pi}T} + \frac{\sigma}{C}\cdot\left[\frac{2}{\sqrt{2\pi}}\left(T + \frac{2}{T}\right) + \frac{1}{\pi}\right]\right\} - \frac{1}{T^2}B(3\sigma + \sigma^2) - \frac{\sqrt{3}B}{\sqrt{2}T}|P|,\end{aligned}$$

where we use Hölder's inequality with $(\infty, 1)$-norm, together with the following facts:

- $\sum_{i=1}^{d} \|u_i\|_2^2 = \mathrm{tr}(W_x^*) \le \sigma$, $\|u_0\|_2^2 = W_{11}^* = 1$,
- $\sum_{i\in P} \|u_i\|_2 \le \sum_{i\in P} \|u_i\|_2^2 / C \le \sigma/C$,
- $\sum_{0\le i,j\le d} \|u_i\|_2 \|u_j\|_2 \le \sum_{i=0}^{d} \|u_i\|_2^2 \le \sigma + 1$,
- $\sum_{0\le i,j\le d,(i,j)\neq(0,0)} |u_i^\top u_j| \le 2\sum_{i=1}^{d} |u_0^\top u_i| + \sum_{i,j=1}^{d} |u_i^\top u_j| \le 2\sigma + \sigma^2$, where we use Proposition 2 and $1_d^\top |W_x^*| 1_d \le \sigma^2$ in the last inequality.

Lastly, by **3E** we obtain our desired inequality.

To conclude the proof, we remains to show that $\bar{x}$ is feasible to (SBQP) with high probability. we only need to show that $\|\bar{x}\|_0 \le \sigma$ holds with probability at least $1 - \exp\{-c\sigma\}$ for some (absolute) constant $c > 0$. Since $\mathbb{E}\,\|\bar{x}\|_0 \le \sum_{i=1}^{d} p_i \le 2/3 \cdot \sigma$, by multiplicative Chernoff bound equipped with an upper bound for the expectation (see, e.g., Theorem 4.4 and the remark after Corollary 4.6 in [35]), we have

$$\mathbb{P}\left(\|\bar{x}\|_0 \ge \sigma\right) \le \mathbb{P}\left(\|\bar{x}\|_0 \ge \left(1 + \frac{1}{2}\right) \cdot \frac{2}{3}\sigma\right) \le e^{-\frac{\sigma}{18}}.$$

□

9 Proofs of Theorems 3 and 4

In this section, we prove Theorem 3 and Theorem 4, stated in Section 5.

We first prove Lemma 1, and then we use it to prove prove Theorem 3 in Section 9.1 and Theorem 4 in Section 9.2. To show Lemma 1, we need two lemmas.

Lemma 4 ([22], Section 5)**.** *Let $D = \mathrm{diag}(d_i)$ be a diagonal matrix of order n, and let $C = D + auu^\top$ with $a < 0$ and u being an n-vector. Denote the eigenvalues of C by $\lambda_1, \lambda_2, \cdots, \lambda_n$ and assume $\lambda_i \le \lambda_{i+1}$, $d_i \le d_{i+1}$. We have $d_1 + a\|u\|_2^2 \le \lambda_1 \le d_1$, and $d_{i-1} \le \lambda_i \le d_i$ for $i \ge 2$.*

Lemma 5 ([8], Appendix A.5.5)**.** *Let P be a symmetric matrix written as a 2×2 block matrix $P = \begin{pmatrix} P_{11} & P_{12} \\ P_{12}^\top & P_{22} \end{pmatrix}$. The following are equivalent:*

(1) $P \succeq 0$.

(2) $P_{11} \succeq 0$, $(I - P_{11}P_{11}^\dagger)P_{12} = O$, *and* $P_{22} \succeq P_{12}^\top P_{11}^\dagger P_{12}$.

We are now ready to prove Lemma 1.

Proof of Lemma 1. We divide the proof into three steps. In Step A, we show $p^* \ge 0_d$, $\lambda_2(H_{S,S}) \ge \delta$, and $H_{S,S} \succeq 0$. In Step B, we show that if in addition, **1A - 1D** hold, then W^* is optimal to (SILS'-SDP). In Step C, we show that if furthermore $\lambda_2(H) > 0$ holds, then W^* is the unique optimal solution to (SILS'-SDP).

Step A. We first show $p^* \ge 0_d$. Since $p_{S^c}^* = 0_{d-\sigma}$, it suffices to prove $p_S^* \ge 0_\sigma$. We have

$$\begin{aligned} \min_{i\in S} p_i^* &= -\sigma\mu_3^* - \mu_2^* + \min_{i\in S}\left[(-\frac{1}{n}(M^\top M)_{S,S}x_S^*)_i - y_i^*\right]/x_i^* \\ &\overset{(7)}{=} -\lambda_{\min}\big(\frac{1}{n}(M^\top M)_{S,S}\big) + \delta - \mu_2^* \ge 0, \end{aligned} \tag{15}$$

where the last inequality is due to (6). Next, we show $\lambda_2(H_{S,S}) \ge \delta$. To see this, (4) gives

$$H_{S,S} = \frac{1}{n}(M^\top M)_{S,S} + \mu_3^* x_S^*(x_S^*)^\top + \mathrm{diag}(p_S^* + \mu_2^* 1_\sigma) - \frac{1}{Y_{11}^*} y_S^*(y_S^*)^\top.$$

By (5), x_S^* is an eigenvector of $H_{S,S}$ corresponding to the zero eigenvalue. Therefore, to show $\lambda_2(H_{S,S}) \geq \delta$, it is sufficient to show that for any unit vector $a \in \mathrm{Span}(\{x_S^*\})^\perp$, we have $a^\top H_{S,S} a \geq \delta$. We obtain

$$a^\top H_{S,S} a = a^\top \Big(\frac{1}{n}(M^\top M)_{S,S} + \mu_3^* x_S^* (x_S^*)^\top + \mathrm{diag}(p_S^* + \mu_2^* 1_\sigma) - \frac{1}{Y_{11}^*} y_S^* (y_S^*)^\top\Big) a$$
$$= a^\top \Big(\frac{1}{n}(M^\top M)_{S,S} + \mathrm{diag}(p_S^* + \mu_2^* 1_\sigma) - \frac{1}{Y_{11}^*} y_S^* (y_S^*)^\top\Big) a.$$

We then define the following two auxiliary matrices:

$$R := \frac{1}{n}(M^\top M)_{S,S} + \mu_2^* I_\sigma + \mathrm{diag}(p_S^*) - \frac{1}{Y_{11}^*} y_S^* (y_S^*)^\top,$$
$$P := \frac{1}{n}(M^\top M)_{S,S} + \mu_2^* I_\sigma + \mathrm{diag}(p_S^*).$$

To prove $a^\top H_{S,S} a \geq \delta$, it is sufficient to show $\lambda_{\min}(P) \geq \delta$. Indeed, by Lemma 4, we see $\lambda_2(R) \geq \lambda_{\min}(P) \geq \delta$. From (5), x_S^* is an eigenvector of R corresponding to eigenvalue $-\sigma\mu_3^* \leq 0$, so it is an eigenvector corresponding to the smallest eigenvalue of R, which then implies $a^\top H_{S,S} a = a^\top R a \geq \delta$. We now check $\lambda_{\min}(P) \geq \delta$. Recall again $\min_{i\in S} p_i^* = -\lambda_{\min}\big(\frac{1}{n}(M^\top M)_{S,S}\big) + \delta - \mu_2^*$ by (15). We have

$$P = \frac{1}{n}(M^\top M)_{S,S} + \mu_2^* I_\sigma + \mathrm{diag}(p_S^*) \succeq \Big(\lambda_{\min}\big(\frac{1}{n}(M^\top M)_{S,S}\big) + \mu_2^* + \min_{i\in S} p_i^*\Big) I_\sigma = \delta I_\sigma.$$

This concludes the proof that $\lambda_{\min}(P) \geq \delta$, and therefore $\lambda_2(H_{S,S}) \geq \delta$.

Finally, $H_{S,S} \succeq 0$ follows easily if one observes that $\lambda_{\min}(H_{S,S}) = 0$. Indeed, direct calculation and (5) gives $H_{S,S} x_S^* = 0_\sigma$, which gives our desired property.

Step B. In this part, we show W^* is optimal by checking (KKT-1) - (KKT-3). We first show that $H \succeq 0$. From Lemma 5, it suffices to show the following three facts: (i) $H_{S,S} \succeq 0$, (ii) $(I_\sigma - H_{S,S} H_{S,S}^\dagger) H_{S,S^c} = O_{\sigma\times(d-\sigma)}$, and (iii) $H_{S^c,S^c} \succeq H_{S^c,S} H_{S,S}^\dagger H_{S^c,S}^\top$. Note that (i) holds by part (a) and (iii) holds by **1A**, so it remains to show (ii). We see

$$(I_\sigma - H_{S,S} H_{S,S}^\dagger) H_{S,S^c} = \frac{1}{\sigma} x_S^* (x_S^*)^\top H_{S,S^c} \overset{\mathbf{1B}}{=} \frac{1}{\sigma} x_S^* 0_{d-\sigma}^\top = O_{\sigma\times(d-\sigma)},$$

where we have used the facts $\lambda_2(H_{S,S}) \geq \delta > 0$ from part (a) and $H_{S,S} x_S^* = 0_\sigma$ in the first equality.

We define $Y^* := \begin{pmatrix} Y_{11}^* & (y^*)^\top \\ y^* & Y_x^* \end{pmatrix}$ and $\mu_1^* := Y_{11}^* - 1/n \cdot b^\top b$. Observe that $Y^* \succeq 0$ again by Lemma 5, due to the facts $H \succeq 0$ and $Y_{11}^* > 0$. (KKT-1) is equivalent to

$$-\mu_3^* x_S^* (x_S^*)^\top = \Big[\frac{1}{n} M^\top M - Y_x^* + \mathrm{diag}(p^* + \mu_2^* 1_d)\Big]_{S,S}, \tag{16}$$

$$\mu_3^* 1_d 1_d^\top \geq \Big|\frac{1}{n} M^\top M - Y_x^* + \mathrm{diag}(p^* + \mu_2^* 1_d)\Big|. \tag{17}$$

We see that (16) coincides with (4), and (17) is implied by (4), the fact that $p_{S^c}^* = 0_{d-\sigma}$, **1C**, and **1D**. (KKT-2) and (KKT-3) hold clearly by definition.

Step C. Finally, we show that W^* is the unique optimal solution if we additionally assume $\lambda_2(H) > 0$. First, note that $\lambda_2(H) > 0$ implies $\lambda_2(Y^*) > 0$ due to the fact that

$$Y^* = \begin{pmatrix} Y_{11}^* & (y^*)^\top \\ y^* & Y_x^* \end{pmatrix} = \begin{pmatrix} 1 & \\ \frac{1}{Y_{11}^*} y^* & I_d \end{pmatrix} \begin{pmatrix} Y_{11}^* & \\ & Y_x^* - \frac{1}{Y_{11}^*} y^* (y^*)^\top \end{pmatrix} \begin{pmatrix} 1 & \\ \frac{1}{Y_{11}^*} y^* & I_d \end{pmatrix}^\top.$$

We define the Lagrangian function $\mathcal{L} : \mathbb{R}^{(1+d)\times(1+d)} \to \mathbb{R}$ as follows:

$$\mathcal{L}(W) := \frac{1}{n}\operatorname{tr}(A^\top AW) - \operatorname{tr}(Y^*W) + \mu_1^*(W_{11} - 1) + \mu_2^*(\operatorname{tr}(W_x) - \sigma) + \mu_3^*(1_d^\top |W_x| 1_d - \sigma^2) + \operatorname{tr}(\operatorname{diag}(p^*)(W_x - I)).$$

Then, for any optimal solution W_0 to (SILS'-SDP), we show $W^* = W_0$. It is clear that

$$\frac{1}{n}\operatorname{tr}(A^\top AW_0) \geq \mathcal{L}(W_0) \geq \mathcal{L}(W^*) = \frac{1}{n}\operatorname{tr}(A^\top AW^*),$$

where the second inequality is due to (KKT-1), which states that O_{1+d} lies in the subdifferential of $\mathcal{L}(W^*)$. By the optimality of W_0, it is clear that, from the second term $-\operatorname{tr}(Y^*W_0)$ to the last term $\operatorname{tr}(\operatorname{diag}(p^*)((W_0)_x - I))$ in $\mathcal{L}(W_0)$, are all zero, as they are always non-positive. In particular, $0 = \operatorname{tr}(Y^*W^*) = \operatorname{tr}(Y^*W_0)$ holds. This implies that W_0 must be a scaling of W^* since $\lambda_2(Y^*) > 0$. Again by optimality of W_0, we see $W_0 = W^*$.

□

In the remainder of the section we prove Theorems 3 and 4. We start with a useful lemma, which introduces the Schur complement of a positive semidefinite matrix. This result follows from Lemma 5.

Lemma 6. *For a positive semidefinite matrix $H \in \mathbb{R}^{d\times d}$, and a set of indices $S \subseteq [d]$. Denote $P_1 := \begin{pmatrix} I_\sigma & \\ H_{S,S^c}^\top H_{S,S}^\dagger & I_{d-\sigma} \end{pmatrix}$, we have*

$$\begin{pmatrix} H_{S,S} & H_{S,S^c} \\ H_{S,S^c}^\top & H_{S^c,S^c} \end{pmatrix} = P_1 \cdot \begin{pmatrix} H_{S,S} & \\ & H_{S^c,S^c} - H_{S,S^c}^\top H_{S,S}^\dagger H_{S,S^c} \end{pmatrix} \cdot P_1^\top.$$

9.1 Proof of Theorem 3

In this proof we intend to use Lemma 1, thus we check that all assumptions in Lemma 1 are satisfied. In particular, we take $(Y_x^*)_{S,S}$ as per (4), p_S^* as per (5), and $p_{S^c}^* = 0_{d-\sigma}$, as in the statement of Lemma 1. Note that since Y_x^* is not completely determined, we also need to define its missing parts, i.e., its (S^c, S) and (S^c, S^c) blocks. For brevity, we denote $H^0 := I_\sigma - x_S^*(x_S^*)^\top/\sigma$ and $P := (M^\top M)_{S,S^c}/n - y_S^*(y_{S^c}^*)^\top/Y_{11}^*$. We take

$$(Y_x^*)_{S^c,S} := \frac{1}{n}(M^\top M)_{S^c,S} - \left[\frac{1}{n\sigma}(M^\top M)_{S^c,S}x_S^* - \frac{1}{Y_{11}^*\sigma}y_{S^c}^*(y_S^*)^\top x_S^*\right](x_S^*)^\top, \tag{18}$$

$$(Y_x^*)_{S^c,S^c} := \Theta_1 + \nu I_{d-\sigma} + \frac{1}{Y_{11}^*}y_{S^c}^*(y_{S^c}^*)^\top + \frac{1}{\delta}P^\top H^0 P, \tag{19}$$

where we set $\nu := \mu_3^* - \|\Theta_2\|_\infty \geq 0$. As in Lemma 1 we define $H := Y_x^* - y^*(y^*)^\top/Y_{11}^*$.

Next, we show that **1A - 1D** are implied due to our choice of p^* and Y_x^*, and conditions **A1 - A2**. This will show that $W^* := \begin{pmatrix} 1 \\ x^* \end{pmatrix}\begin{pmatrix} 1 \\ x^* \end{pmatrix}^\top$ is optimal to (SILS'-SDP). After that, we show that **A2** automatically implies $\lambda_2(H) > 0$, which additionally guarantees the uniqueness of W^*, and we conclude that (SILS-SDP) recovers x^*.

We now check that **1A** holds. By direct calculation,

$$H_{S^c,S^c} \succeq \frac{1}{\delta}P^\top H^0 P \succeq H_{S^c,S}H_{S,S}^\dagger H_{S^c,S}^\top,$$

where the last inequality is due to the facts that $P = H^0 H_{S^c,S}^\top$, $(H^0)^2 = H^0$, and $H_{S,S} \succeq \delta\big(I_\sigma - x_S^*(x_S^*)^\top/\sigma\big)$. The last fact is due to $\lambda_2(H_{S,S}) \geq \delta$ and $H_{S,S}x_S^* = 0_\sigma$.

Next, we prove that **1B** is satisfied. From (18), we obtain

$$H_{S^c,S}x^*_S = [\frac{1}{n}(M^\top M)_{S^c,S} - \frac{1}{Y^*_{11}}y^*_{S^c}(y^*_S)^\top][I_\sigma - \frac{1}{\sigma}x^*_S(x^*_S)^\top]x^*_S$$
$$= [\frac{1}{n}(M^\top M)_{S^c,S} - \frac{1}{Y^*_{11}}y^*_{S^c}(y^*_S)^\top]0_\sigma = 0_{d-\sigma}.$$

1C is true because we have

$$\left\|(\frac{1}{n}M^\top M - Y^*_x)_{S^c,S}\right\|_\infty = \left\|[\frac{1}{n\sigma}(M^\top M)_{S^c,S}x^*_S - \frac{1}{Y^*_{11}\sigma}y^*_{S^c}(y^*_S)^\top x^*_S](x^*_S)^\top\right\|_\infty \overset{\mathbf{A1}}{\leq} \mu^*_3.$$

Consider now **1D**. Due to (19),

$$\left\|(\frac{1}{n}M^\top M - Y^*_x)_{S^c,S^c} + \mu^*_2 I_{d-\sigma}\right\|_\infty = \|-\nu I_{d-\sigma} + \Theta_2\|_\infty \leq \nu + \|\Theta_2\|_\infty \overset{\mathbf{A2}}{=} \mu^*_3.$$

Finally, we will show **A2** implies $\lambda_2(H) > 0$. Lemma 6 shows that $\lambda_2(H) > 0$ is equivalent to $H_{S^c,S^c} - H^\top_{S,S^c}H^\dagger_{S,S}H_{S,S^c} \succ 0$, due to the facts $\lambda_{\min}(H_{S,S}) = 0$ and $\lambda_2(H_{S,S}) \geq \delta > 0$. Finally, we observe that $H_{S^c,S^c} - H^\top_{S,S^c}H^\dagger_{S,S}H_{S,S^c} \succeq H_{S^c,S^c} - H^\top_{S,S^c}(1/\delta)\cdot\big(I_\sigma - x^*_S(x^*_S)^\top/\sigma\big)H_{S,S^c} = \Theta_1 + \nu I_{d-\sigma} \succ 0$ as desired. □

9.2 Proof of Theorem 4

In this proof we use Lemma 1, thus we check that all assumptions in Lemma 1 are satisfied. We fix $(Y^*_x)_{S,S}$ as per (4), p^*_S as per (5), and $p^*_{S^c} = 0_{d-\sigma}$. Note that we still need to define the missing parts of Y^*_x, namely, its (S^c, S) and (S^c, S^c) blocks. We take

$$(Y^*_x)_{S^c,S} := -\frac{1}{\sigma}y^*_{S^c}(x^*_S)^\top, \tag{20}$$
$$(Y^*_x)_{S^c,S^c} := \nu I_{d-\sigma} + \frac{1}{Y^*_{11}}y^*_{S^c}(y^*_{S^c})^\top + H_{S^c,S}H^\dagger_{S,S}H^\top_{S^c,S}. \tag{21}$$

With a little abuse of notation, we denote by $\nu > 0$ the slack in the inequality introduced in **B2**. As in Lemma 1 we define $H := Y^*_x - y^*(y^*)^\top/Y^*_{11}$.

Next, we check **1A - 1D**, and $\lambda_2(H) > 0$. Similarly to the proof of Theorem 3, we show that **1A - 1D** are implied by our choice of p^* and Y^*_x, and conditions **B1 - B2**. This will show that $W := \begin{pmatrix}1\\x^*\end{pmatrix}\begin{pmatrix}1\\x^*\end{pmatrix}^\top$ is optimal to (SILS'-SDP). After that, we show that **B2** implies $\lambda_2(H) > 0$, which additionally guarantees the uniqueness of W^*, and we conclude that (SILS'-SDP) recovers x^*.

We now check that **1A** holds. From (21),

$$H_{S^c,S^c} = \nu I_{d-\sigma} + H_{S^c,S}H^\dagger_{S,S}H^\top_{S^c,S} \succeq H_{S^c,S}H^\dagger_{S,S}H^\top_{S^c,S}.$$

Next, we prove that **1B** is satisfied. From (20), we obtain

$$H_{S^c,S}x^*_S = [-\frac{1}{\sigma}y^*_{S^c}(x^*_S)^\top - \frac{1}{Y^*_{11}}y^*_{S^c}(y^*_S)^\top]x^*_S = -y^*_{S^c} + y^*_{S^c} = 0_{d-\sigma}.$$

1C is true because

$$\left\|(\frac{1}{n}M^\top M - Y^*_x)_{S^c,S}\right\|_\infty = \left\|\frac{1}{n}(M^\top M)_{S^c,S} + \frac{1}{\sigma}y^*_{S^c}(x^*_S)^\top\right\|_\infty \overset{\mathbf{B1}}{\leq} \mu^*_3.$$

Consider now **1D**. From (21), $\left\|(\frac{1}{n}M^\top M - Y_x^*)_{S^c,S^c} + \mu_2^* I_{d-\sigma}\right\|_\infty$ is then upper bounded by

$$\begin{aligned}
&\left\|\frac{1}{n}(M^\top M)_{S^c,S^c} + \mu_2^* I_{d-\sigma}\right\|_\infty + \left\|H_{S^c,S}H_{S,S}^\dagger H_{S^c,S}^\top\right\|_\infty + \left\|\frac{1}{Y_{11}^*}y_{S^c}^*(y_{S^c}^*)^\top\right\|_\infty + \nu \\
\leq &\left\|\frac{1}{n}(M^\top M)_{S^c,S^c} + \mu_2^* I_{d-\sigma}\right\|_\infty + \left\|\frac{1}{Y_{11}^*}y_{S^c}^*(y_{S^c}^*)^\top\right\|_\infty + \nu + \frac{1}{\delta}\|\frac{1}{\sigma}x_S^* + \frac{1}{Y_{11}^*}y_S^*\|_2^2 \|y_{S^c}^*\|_\infty^2 \\
= &\left\|\frac{1}{n}(M^\top M)_{S^c,S^c} + \mu_2^* I_{d-\sigma}\right\|_\infty + \left\|\frac{1}{Y_{11}^*}y_{S^c}^*(y_{S^c}^*)^\top\right\|_\infty + \nu + \frac{1-\cos^2(\theta)}{\delta\sigma\cos^2(\theta)}\|y_{S^c}^*\|_\infty^2 \overset{\mathbf{B2}}{=} \mu_3^*,
\end{aligned}$$

where we used the triangle inequality in the first inequality, the fact that $\left\|H_{S,S}^\dagger\right\|_2 \leq \frac{1}{\delta}$ in the second inequality, and the fact that

$$\left\|-\frac{1}{\sigma}x_S^* - \frac{1}{Y_{11}^*}y_S^*\right\|_2^2 = \frac{1}{\sigma} + \frac{2(x_S^*)^\top y_S^*}{Y_{11}^*\sigma} + \left\|\frac{1}{Y_{11}^*}y_S^*\right\|_2^2 = -\frac{1}{\sigma} + \left\|\frac{1}{Y_{11}^*}y_S^*\right\|_2^2 = \frac{1-\cos^2(\theta)}{\sigma\cos^2(\theta)}$$

in the penultimate equality.

Finally, we show that **B2** implies $\lambda_2(H) > 0$. From Lemma 6, it suffices to show $\lambda_{\min}(H_{S^c,S^c} - H_{S,S^c}^\top H_{S,S}^\dagger H_{S,S^c})$ is positive. By definition of H_{S^c,S^c}, we obtain that $H_{S^c,S^c} - H_{S,S^c}^\top H_{S,S}^\dagger H_{S,S^c} = \nu I_{d-\sigma} \succ 0$. □

10 Proof of Theorem 9

Before proving Theorem 9, we need some detailed analysis of our covariance matrix Σ and some useful probabilistic inequalities. We will use them to evaluate norms of some matrices, which are used for the construction of the decomposition $\Theta = \Theta_1 + \Theta_2$ in Theorem 8.

Throughout the section, we use the same definitions as in the statement of Theorem 8, i.e., $S := \mathrm{Supp}(z^*)$, $y^* := -M^\top b/n$, $Y_{11}^* := -(y_S^*)^\top z_S^*$, and $\mu_3^* = 1/\sigma \cdot \{\lambda_{\min}\big((M^\top M/n)_{S,S}\big) - \delta + \min_{i\in S}[M^\top \epsilon]_i/(nx_i^*)\}$. Furthermore, we use the notation introduced in Model 2 and we introduce some additional notation that is specific for it. Let $y_i', y_i'' \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0_d, I_d)$. We observe that m_i has the same distribution as another random vector $\Sigma_1^{\frac{1}{2}}y_i' + \Sigma_2^{\frac{1}{2}}y_i''$. For the ease of notation, we write $M_1^\top := \Sigma_1^{\frac{1}{2}}(y_1', \cdots, y_n')$ and $M_2^\top := \Sigma_2^{\frac{1}{2}}(y_1'', \cdots, y_n'')$. Hence we assume $M = M_1 + M_2$. Observe that $\Sigma_2^{\frac{1}{2}} = \begin{pmatrix} O_\sigma & \\ & \sqrt{c''}I_{d-\sigma}\end{pmatrix}$, so M_2 is an $n \times d$ matrix with the first σ columns being zero.

In Lemma 7 below, we show that $\Sigma_1^{\frac{1}{2}}$ has a simple structure.

Lemma 7. *In Model 2, we have* $\Sigma_1^{\frac{1}{2}} = \begin{pmatrix} A_{11} & a1_\sigma 1_{d-\sigma}^\top \\ a1_{d-\sigma}1_\sigma^\top & b1_{d-\sigma}1_{d-\sigma}^\top\end{pmatrix}$ *for some matrix* $A_{11} \in \mathbb{R}^{\sigma\times\sigma}$ *and* $a, b \in \mathbb{R}$.

Proof. Since the characteristic polynomial of Σ_1 is $(x-c)^{\sigma-1}x^{d-\sigma-1}[x^2 - (c + c'\sigma(d-\sigma))x + \sigma(cc'-1)(d-\sigma)]$, we conclude that Σ_1 has eigenvalue c with multiplicity $\sigma - 1$, and has eigenvalues λ_1, λ_2 with multiplicity 1, where λ_1 and λ_2 are the two distinct roots of $x^2 - (c + c'\sigma(d-\sigma))x + \sigma(cc'-1)(d-\sigma)$. It is clear that every eigenvector corresponding to c is a σ-sparse vector supported on $[\sigma]$, since the equation $(\Sigma_1 - cI_\sigma)w = 0_d$ forces $1_{d-\sigma}^\top w_{[\sigma]^c} = 0$ and $(1_\sigma^\top w_{[\sigma]})1_{d-\sigma} - cw_{[\sigma]^c} = 0_{d-\sigma}$, which implies $w_{[\sigma]^c} = 0_{d-\sigma}$. Furthermore, direct calculation shows that the structure of eigenvectors corresponding to λ_i $(i = 1, 2)$ must be $u_i = (a_i 1_\sigma^\top, b_i 1_{d-\sigma}^\top)^\top$, for some constants a_i, b_i such that $\sigma a_i^2 + (d-\sigma)b_i^2 = 1$. Therefore, the $([\sigma], [\sigma]^c)$ block of Σ is solely contributed by the corresponding block of $\sqrt{\lambda_1}u_1u_1^\top + \sqrt{\lambda_2}u_2u_2^\top$, and the coresponding entries are all equal to $a = \sqrt{\lambda_1}a_1b_1 + \sqrt{\lambda_2}a_2b_2$. Similarly, we can show $b = \sqrt{\lambda_1}b_1^2 + \sqrt{\lambda_2}b_2^2$. □

By Lemma 7, we observe that $(M_1^\top M_1)_{S^c,S}$, $(M_1^\top M_2)_{S^c,S^c}$, and $(M_1^\top M_1)_{S^c,S^c}$ are rank-one matrices. In fact, there exist vectors $u \in \mathbb{R}^\sigma$, $v \in \mathbb{R}^{d-\sigma}$, and a scalar c_1 such that $(M_1^\top M_1)_{S^c,S}/n = 1_{d-\sigma}u^\top$, $(M_1^\top M_2)_{S^c,S^c}/n = 1_{d-\sigma}v^\top$, and $(M_1^\top M_1)_{S^c,S^c}/n = c_1 1_{d-\sigma}1_{d-\sigma}^\top$. In the next lemma, we provide some probabilistic upper bounds.

Lemma 8. *Consider Model 2 and suppose* $\log(d)/n \to 0$ *and* $(n, d, \sigma) \to \infty$. *Let* u, v, c_1 *be as defined above. Then, the following properties hold with probability at least* $1 - \mathcal{O}(1/d)$*:*

8A. $\exists$ *constant* $C_1 = C_1(c, c'')$ *such that* $\left\|(M_2^\top M_1/n)_{S^c,S}\right\|_\infty \le C_1\sqrt{\log(d)/n}$;

8B. $\exists$ *constant* $C_2 = C_2(c, c'')$ *such that* $\left\|(M_2^\top M_1)_{[d],S}z_S^*/n\right\|_\infty \le C_2\sqrt{\sigma\log(d)/n}$;

8C. $\exists$ *constant* $C_3 = C_3(c, c', c'')$ *such that* $\left\|(M^\top\epsilon/n)_{S^c}\right\|_\infty \le C_3\sqrt{\varrho^2\sigma\log(d)/n}$;

8D. $\exists$ *constant* $C_4 = C_4(c)$ *such that* $\left\|(M^\top\epsilon/n)_S\right\|_\infty \le C_4\sqrt{\varrho^2\log(d)/n}$;

8E. $\exists$ *constant* $C_5 = C_5(c', c'')$ *such that* $\|v\|_\infty \le C_5\sqrt{\sigma\log(d)/n}$

8F. $\exists$ *constant* $C_6 = C_6(c, c'')$ *such that* $\|(M_2^\top M_1/n)_{S^c,S}\|_{2\to\infty} \le C_6(\sqrt{\log(d)} + \sqrt{\sigma})/\sqrt{n}$;

8G. $\exists$ *constant* $C_7 = C_7(c, c')$ *such that* $\|u - 1_\sigma\|_\infty \le C_7\sqrt{\sigma\log(d)/n}$;

8H. $\exists$ *constant* $C_8 = C_8(c')$ *such that* $|c_1 - c'\sigma| \le C_8\sigma\sqrt{\log(d)/n}$.

Proof. We note that most of these properties are due to Bernstein inequality or similar derivations in Lemma 2 and in [2]. To avoid repetition, we only show why Bernstein inequality can be applied, provide upper bounds for the Orlicz norm (see [49] for definition) of some sub-exponential random variables that are of interest, and then briefly discuss how **8A** - **8H** can be shown.

We first show that the entries of $M_2^\top M_1$, $M_1^\top M_1$, and $M^\top\epsilon$ are sums of sub-exponential random variables. Indeed, this is due to the fact that the product of two sub-Gaussian random variables is sub-exponential (see, e.g., Lemma 2.7.7 in [49]). Then, we can safely apply Bernstein inequality. Next, we upper bound the Orlicz norms of their entries, and discuss the derivations of **8A** - **8H**. We start by showing **8A**, and then we illustrate that the proofs of **8B** - **8H** can be obtained in a similar way.

The proof of **8A** is very similar to the proof of **2B** in Lemma 2. We first notice that $(M_2^\top M_1)_{[d],S} = \Sigma_2^{\frac{1}{2}}(y_1'', \cdots, y_n'')(y_1', \cdots, y_n')^\top(\Sigma_1^{\frac{1}{2}})_{[d],S}$, thus every entry in $(M_2^\top M_1)_{[d],S}$ is the sum of products of two independent centered Gaussian variables with variance upper bounded by c'' and c, respectively (since $\Sigma_2^{\frac{1}{2}}y_i'' \sim \mathcal{N}(0_d, c''I_d)$ and $(\Sigma_1^{\frac{1}{2}})_{S,[d]}y_i' \sim \mathcal{N}(0_\sigma, cI_\sigma)$). Then, from Lemma 2.7.7 in [49], we see that $(M_2^\top M_1)_{[d],S}$ has entries that are the sums of sub-exponential random variables with Orlicz norm upper bounded by a constant multiple of $\sqrt{cc''}$. Next, applying Bernstein inequality, for each entry of $(M_2^\top M_1)_{S^c,S}$, applying a union bound argument to upper bound the probability of the random event $R_1(t) := \{\left\|\frac{1}{n}(M_2^\top M_1)_{S^c,S}\right\|_\infty > t\}$, and then by setting $t = C_1'\sqrt{cc''}\sqrt{\log(d)/n}$, for some large absolute constant $C_1' > 0$, we obtain $\mathbb{P}(R_1(t)) \le \mathcal{O}(1/d)$.

For **8B**, we have shown that each entry in $(M_2^\top M_1)_{[d],S}$ is the sum of sub-exponential variables with Orlicz norm upper bounded by an absolute constant multiple of $\sqrt{c'c''}$. Mirroring the proof of **2C** in Lemma 2, we can show that there exists an absolute constant $c_0 > 0$ such that for every nonzero vector $x \in \mathbb{R}^\sigma$,

$$\mathbb{P}\left(\left\|\frac{1}{n}(M_2^\top M_1)_{S^c,S}x\right\|_\infty > t\right) \le 2(d-\sigma)\exp\left(-c_0\frac{nt^2}{cc'\left\|x\right\|_2^2}\right), \tag{22}$$

and hence we obtain **8B** by taking a sufficiently large absolute constant C_2. For **8C**, **8D**, **8E**, **8G**, and **8H**, we can similarly obtain that entries of $(M^\top\epsilon)_{S^c}$, $(M^\top\epsilon)_S$, $(M_1^\top M_2)_{S^c,S^c}$, $(M_1^\top M_1)_{S^c,S}$, and $(M_1^\top M_1)_{S^c,S^c}$, are sums of sub-exponential variables with Orlicz norm upper bounded by an absolute constant multiple of $\varrho\sqrt{c+c'\sigma+c''}$, $\varrho\sqrt{c}$, $\sqrt{c'c''\sigma}$, $\sqrt{cc'\sigma}$, and $c'\sigma$, respectively. Then, we can apply Bernstein inequality and the union bound to derive these five properties, similarly to the proof of **8A**. For **8F**, the derivation is the same as the proof of Lemma 15 in [2], if we replace eq. (82) therein by (22), and proceed with the arguments after Lemma 16 therein. □

In the following, we define some matrices that will be used in the proof of Theorem 9 for the construction of Θ_1 and Θ_2 in Theorem 8. Recall that $H^0 = I_\sigma - z_S^*(z_S^*)^\top/\sigma$. For simplicity, we denote $B := [I_\sigma + z_S^*(y_S^*)^\top/Y_{11}^*](1/\delta)H^0[I_\sigma + y_S^*(z_S^*)^\top/Y_{11}^*] + z_S^*(z_S^*)^\top/Y_{11}^*$, and we define

$$
\begin{aligned}
\Theta_2^A &:= -\frac{1}{Y_{11}^*}(\frac{1}{n}M^\top\epsilon)_{S^c}(\frac{1}{n}M^\top\epsilon)_{S^c}^\top,\\
\Theta_1^B &:= \Big(\sqrt{\tilde{c}}1_{d-\sigma} + \frac{u^\top x_S^*}{Y_{11}^*\sqrt{\tilde{c}}}(\frac{1}{n}M^\top\epsilon)_{S^c}\Big)\Big(\sqrt{\tilde{c}}1_{d-\sigma} + \frac{u^\top x_S^*}{Y_{11}^*\sqrt{\tilde{c}}}(\frac{1}{n}M^\top\epsilon)_{S^c}\Big)^\top,\\
\Theta_2^B &:= -\frac{1}{Y_{11}^*}(\frac{1}{n}M^\top\epsilon)_{S^c}\big(\frac{1}{n}(M_2^\top M_1)_{S^c,S}z_S^*\big)^\top - \frac{1}{Y_{11}^*}\big(\frac{1}{n}(M_2^\top M_1)_{S^c,S}z_S^*\big)(\frac{1}{n}M^\top\epsilon)_{S^c}^\top\\
&\quad - \frac{(u^\top z_S^*)^2}{(Y_{11}^*)^2\tilde{c}}(\frac{1}{n}M^\top\epsilon)_{S^c}(\frac{1}{n}M^\top\epsilon)_{S^c}^\top,\\
\Theta_2^C &:= -\frac{1}{\delta n^2 Y_{11}^*}\Bigg[(M^\top M)_{S^c,S}\big(I_\sigma + \frac{z_S^*(y_S^*)^\top}{Y_{11}^*}\big)H^0 y_S^*(M^\top\epsilon)_{S^c}^\top\\
&\quad - (M^\top\epsilon)_{S^c}(y_S^*)^\top H^0\big(I_\sigma + \frac{y_S^*(z_S^*)^\top}{Y_{11}^*}\big)(M^\top M)_{S,S^c}\Bigg],\\
\Theta_2^D &:= \frac{1}{\delta(nY_{11}^*)^2}(M^\top\epsilon)_{S^c}(y_S^*)^\top H^0 y_S^*(M^\top\epsilon)_{S^c}^\top,\\
\Theta_1^E &:= \hat{c}1_{d-\sigma}1_{d-\sigma} - \frac{1}{n}(M_1^\top M_1)_{S^c,S}B\frac{1}{n}(M_1^\top M_1)_{S,S^c}\\
&\quad + \Big(\sqrt{\bar{c}}1_{d-\sigma} - \frac{1}{\sqrt{\bar{c}}}\frac{1}{n}(M_2^\top M_1)_{S^c,S}Bu\Big)\Big(\sqrt{\bar{c}}1_{d-\sigma} - \frac{1}{\sqrt{\bar{c}}}\frac{1}{n}(M_2^\top M_1)_{S^c,S}Bu\Big)^\top,\\
\Theta_2^E &:= -\frac{1}{\bar{c}}\frac{1}{n}(M_2^\top M_1)_{S^c,S}Au\Big(\frac{1}{n}(M_2^\top M_1)_{S^c,S}Bu\Big)^\top - \frac{1}{n}(M_2^\top M_1)_{S^c,S}B\frac{1}{n}(M_2^\top M_1)_{S,S^c},\\
\Theta_1^F &:= \Big(\frac{1}{n}M_1^\top M_1\Big)_{S^c,S^c} - (\bar{c}+\hat{c}+\tilde{c}+\check{c})1_{d-\sigma}1_{d-\sigma}^\top + \check{c}\Big(1_{d-\sigma} + \frac{1}{\check{c}}v\Big)\Big(1_{d-\sigma} + \frac{1}{\check{c}}v\Big)^\top,\\
\Theta_2^F &:= -\frac{1}{\check{c}}vv^\top + \Big(\frac{1}{n}M_2^\top M_2\Big)_{S^c,S^c} + \mu_2^* I_{d-\sigma},
\end{aligned}
$$

for some proper positive constants $\bar{c}$, $\hat{c}$, $\tilde{c}$ and $\check{c}$ such that Θ_1^B, Θ_1^E and Θ_1^F are positive semidefinite matrices. The high-level idea in the proof of Theorem 9 is to take $\Theta_1 = \Theta_1^B + \Theta_1^E + \Theta_1^F$ and $\Theta_2 = \Theta_2^A + \Theta_2^B + \Theta_2^C + \Theta_2^D + \Theta_2^E + \Theta_2^F$, and to directly check that such Θ_1 and Θ_2 add up to Θ in Theorem 8. Before proving Theorem 9, we need two lemmas: Lemma 9 gives some useful results that will be used repeatedly in the proofs of Lemma 10 and Theorem 9, and Lemma 10 gives upper bounds on the infinity norms of the matrices defined above that contribute to Θ_2.

Lemma 9. *There exists a constant $C = C(c, c', c'') > 0$ such that when $n \geq C\varrho^2\sigma^2\log(d)$, the following properties hold w.h.p. as $(n, \sigma, d) \to \infty$:*

9A. $Y_{11}^* \geq \sigma/2$*;*

9B. $\|-y_S^* - z_S^*\|_2 \leq 1/2$*;*

9C. $\left\|u^\top(I_\sigma + y_S^*(z_S^*)^\top/Y_{11}^*)\right\|_2 \le 6\sqrt{\sigma}$;

9D. $\left\|H^0 y_S^*\right\|_2 \le 1/2$.

Proof. For brevity, in this proof, we say that n is sufficiently large if we take a sufficiently large C.

For **9A**, observe $Y_{11}^* = -(z_S^*)^\top y_S^* = (z_S^*)^\top(M^\top M/n)_{S,S} z_S^* - (z_S^*)^\top(M^\top\epsilon/n)_S$, and hence from **2A** and **8D**, $Y_{11}^* \ge \sigma - c_1\sigma\sqrt{\sigma/n} - (z_S^*)^\top(M^\top\epsilon/n)_S \ge \sigma(1 - c_1\sqrt{\sigma/n} - C_4\sqrt{\varrho^2\log(d)/n}) \ge \sigma/2$, for sufficiently large n.

For **9B**, observe $\|-y_S^* - z_S^*\|_2 = \left\|((M^\top M)_{S,S}/n - I_\sigma)z_S^* + (M^\top\epsilon/n)_S\right\|_2 \le \left\|(M^\top M)_{S,S}/n - I_\sigma\right\|_2 \cdot \|z_S^*\|_2 + \left\|(M^\top\epsilon/n)_S\right\|_2$. From **2A** and **8D**, we see that this quantity is upper bounded by $c_1\sqrt{\sigma^2/n} + \sqrt{\sigma}C_4\sqrt{\varrho^2\log(d)/n}$, which is less than 1/2, for sufficiently large n.

For **9C**, we have $\left\|u^\top(I_\sigma + y_S^*(z_S^*)^\top/Y_{11}^*)\right\|_2 \le \left\|I_\sigma + z_S^*(y_S^*)^\top/Y_{11}^*\right\|_2 \|u\|_2 \le (1 + \|z_S^*\|_2\|y_S^*\|_2/Y_{11}^*)\|u\|_2$, and hence by **9A**, **8G**, and **9B**, we obtain that it is upper bounded by $[1 + 2\cdot(1/2 + 1/(2\sqrt{\sigma}))]\cdot\sqrt{\sigma}(1 + C_7\sqrt{\sigma\log(d)/n}) \le 6\sqrt{\sigma}$ for sufficiently large n.

Finally, for **9D**, we observe that $H^0 z_S^* = (I_\sigma - z_S^*(z_S^*)^\top/\sigma)z_S^* = 0_\sigma$, thus $\left\|H^0 y_S^*\right\|_2 = \left\|H^0(y_S^* - z_S^*)\right\|_2 \le \left\|H^0\right\|_2\|y_S^* - z_S^*\|_2 \le 1/2$ by the fact that $\left\|H^0\right\|_2 = 1$ and **9B**. □

Lemma 10. *There exists a constant $C = C(c, c', c'') > 0$ such that when $n \ge C\varrho^2\sigma^2\log(d)$, the following properties hold w.h.p. as $(n, \sigma, d) \to \infty$:*

10A. $\left\|\Theta_2^A\right\|_\infty = \mathcal{O}\Big(\varrho^2\log(d)/n\Big)$;

10B. $\left\|\Theta_2^B\right\|_\infty = \mathcal{O}\Big(\varrho\sigma\log(d)/n + \varrho^2\sigma\log(d)/(\tilde{c}n)\Big)$;

10C. $\left\|\Theta_2^C\right\|_\infty = \mathcal{O}\Big(\sqrt{\varrho^2\log(d)/n}/\delta + \sqrt{\sigma}\varrho\log(d)/(\delta n)\Big)$;

10D. $\left\|\Theta_2^D\right\|_\infty = \mathcal{O}\Big(\varrho^2\log(d)/(n\sigma\delta)\Big)$;

10E. $\left\|\Theta_2^E\right\|_\infty = \mathcal{O}\Big((\sqrt{\sigma\log(d)} + \sigma)^2/(\bar{c}\delta^2 n) + \sigma\log(d)/(\delta n)\Big)$.

Proof. For brevity, in this proof, we say that n is sufficiently large if we take a sufficiently large C. In the proof, we will repeatedly use the fact that for a rank-one matrix $P = ab^\top$, $\|P\|_\infty = \|a\|_\infty\|b\|_\infty$.

(**10A**). The statement simply follows from **9A** and **8C**.

(**10B**). observe that, among the three terms in Θ_2^B, the first term is the transpose of the second term, so it is sufficient to upper bound the infinity norm of the first term, since the same bound holds for the second. From **8B**, **8C**, and **9A**, we have that $\left\|1/Y_{11}^* \cdot (M^\top\epsilon/n)_{S^c}((M_2^\top M_1/n)_{S^c,S} z_S^*)^\top\right\|_\infty$ is upper bounded by $2C_2C_3\varrho\sigma\log(d)/n$. Then, from **8C**, **8G**, and **9A**, we conclude to the fact that $\left\|(u^\top z_S^*)^2/[(Y_{11}^*)^2\tilde{c}]\cdot(M^\top\epsilon/n)_{S^c}(M^\top\epsilon/n)_{S^c}^\top\right\|_\infty \le 4C_3^2\varrho^2\sigma\log(d)/(\tilde{c}n)$.

(**10C**). Note that the first term in the definition of Θ_2^C is the transpose of the second term, thus it is sufficient to upper bound the infinity norm of the first term. We write $(M^\top M/n)_{S^c,S}(I_\sigma + z_S^*(y_S^*)^\top/Y_{11}^*)H^0 y_S^*(M^\top\epsilon/n)_{S^c}^\top = 1_{d-\sigma}u^\top(I_\sigma + z_S^*(y_S^*)^\top/Y_{11}^*)H^0 y_S^*(M^\top\epsilon/n)_{S^c}^\top + (M_2^\top M_1)_{S^c,S}(I_\sigma + z_S^*(y_S^*)^\top/Y_{11}^*)H^0 y_S^*(M^\top\epsilon/n)_{S^c}^\top := P_1 + P_2$, since $(M_1^\top M_2)_{S^c,S} = (M_2^\top M_2)_{S^c,S} = O_{(d-\sigma)\times\sigma}$. $\|P_1\|_\infty = |u^\top(I_\sigma + y_S^*(z_S^*)^\top/Y_{11}^*)H^0 y_S^*|\left\|(M^\top\epsilon/n)_{S^c}\right\|_\infty$, by **9C**, **9D**, and **8C**, we see $\|P_1\|_\infty \le 3C_3\sqrt{\varrho^2\sigma^2\log(d)/n}$.

Next, $\|P_2\|_\infty = \left\|(M_2^\top M_1)_{S^c,S}(I_\sigma + z_S^*(y_S^*)^\top/Y_{11}^*)H^0 y_S^*\right\|_\infty\left\|(M^\top\epsilon/n)_{S^c}\right\|_\infty$, thus from **8F** and **8C**, we obtain that $\|P_2\|_\infty \le C_3C_6\sqrt{\varrho^2\sigma\log(d)/n}(\sqrt{\log(d)} + \sqrt{\sigma})/\sqrt{n}\cdot\left\|(I_\sigma + z_S^*(y_S^*)^\top/Y_{11}^*)H^0 y_S^*\right\|_2$. By **9D**, **9B**, and **9A**, we obtain that $\left\|(I_\sigma + z_S^*(y_S^*)^\top/Y_{11}^*)H^0 y_S^*\right\|_2 \le \left\|H^0 y_S^*\right\|_2 + \|x_S^*\|_2\|y_S^*\|_2\left\|H^0 y_S^*\right\|_2/Y_{11}^* \le 2$. Hence, we see $\|P_2\|_\infty \le 2C_3C_6\sqrt{\varrho^2\sigma\log(d)/n}(\sqrt{\log(d)} + \sqrt{\sigma})/\sqrt{n}$.

Finally, from **9A**, we obtain $\left\|\Theta_2^C\right\|_\infty = \mathcal{O}\Big(\sqrt{\varrho^2\log(d)/n}/\delta + \sqrt{\sigma}\varrho\log(d)/(\delta n)\Big)$.

(**10D**). Since $H^0 z_S^* = 0_\sigma$, we obtain that $(y_S^*)^\top H^0 y_S^* = (y_S^* - z_S^*)^\top H^0(y_S^* - z_S^*)$. By **9B** and $\left\|H^0\right\|_2 = 1$, we see $(y_S^*)^\top H^0 y_S^* \le 1/4$. We are done by combining the above conclusion, **8C**, and **9A**.

(**10E**). We start by estimating the infinity norm of the first term in the definition of Θ_2^E. To do so, we first provide an upper bound on $\|B\|_2$. Write $B = [H^0/\delta + z_S^*(z_S^*)^\top/Y_{11}^* + (y_S^*)^\top H^0 y_S^* z_S^*(z_S^*)^\top/(Y_{11}^*)^2] + [z_S^*(y_S^*)^\top H^0 + H^0 y_S^*(x_S^*)^\top]/(\delta Y_{11}^*) := B_1 + B_2$, and we will upper bound $\|B_1\|_2$ and $\|B_2\|_2$. For B_1, recall that $H^0 = I_\sigma - z_S^*(z_S^*)^\top/\sigma$, thus $B_1 = (1/\delta)I_\sigma + [1/Y_{11}^* + (y_S^*)^\top H^0 y_S^*/(Y_{11}^*)^2 - 1/(\sigma\delta)]x_S^*(x_S^*)^\top$. From $H^0 = (H^0)^2$, **9A**, and **9D**, we see $(y_S^*)^\top H^0 y_S^*/(Y_{11}^*)^2 \le 1/\sigma^2$, and thus $\|B_1\|_2 \le 1/\delta + 2 + 1/\sigma + 1/\delta \le 3 + 2/\delta$. For B_2, we only need to upper bound $z_S^*(y_S^*)^\top H^0/(\delta Y_{11}^*)$, since the other term is symmetric. From **9A** and **9D**, $\left\|z_S^*(y_S^*)^\top H^0/(\delta Y_{11}^*)\right\|_2 \le 2\left\|x_S^*\right\|_2\left\|H^0 y_S^*\right\|_2/(\delta\sigma) \le 1/\delta$. Thus, $\|B\|_2 \le 3 + 3/\delta$. Combining this and **8F**, **8G**, we obtain $\left\|(M_2^\top M_1/n)_{S^c,S}Bu\right\|_\infty = \mathcal{O}\Big((\sqrt{\sigma\log(d)} + \sigma)/(\delta\sqrt{n})\Big)$.

For the second term in the definition of Θ_2^E, we write $B := H^0/\delta + B_3$, and we give upper bounds on the infinity norms of $(M_2^\top M_1/n)_{S^c,S}H^0(M_2^\top M_1/n)_{S,S^c}$ and $(M_2^\top M_1/n)_{S^c,S}B_3(M_2^\top M_1/n)_{S,S^c}$. We know that the diagonal entries of H^0 are $1 - 1/\sigma$, and the off-diagonal entries have an absolute value of $1/\sigma$, thus, along with **8A** we see $\left\|(M_2^\top M_1/n)_{S^c,S}H^0(M_2^\top M_1/n)_{S,S^c}\right\|_\infty \le \left\|(M_2^\top M_1/n)_{S^c,S}\right\|_\infty^2[\sigma\cdot(1-1/\sigma)+\sigma(\sigma-1)\cdot 1/\sigma] = \mathcal{O}(\sigma\log(d)/n)$. Next, by **9A** and **9D**, each entry in B_3 is upper bounded by $\mathcal{O}(1/\sigma+1/(\delta\sigma))$. Together with **8A**, we obtain $\big\|(M_2^\top M_1/n)_{S^c,S}B_3(M_2^\top M_1/n)_{S,S^c}$ $\sigma^2\left\|(M_2^\top M_1/n)_{S^c,S}\right\|_\infty^2\cdot\mathcal{O}(1/\sigma + 1/(\delta\sigma)) = \mathcal{O}(\sigma\log(d)/(\delta n))$. Using the triangle inequality, the second term in Θ_2^E has infinity norm upper bounded by $\mathcal{O}(\sigma\log(d)/(\delta n))$. $\square$

We are now ready to prove Theorem 9 using Theorem 8.

Proof of Theorem 9. We use Theorem 8 to prove this proposition. In the proof, We take $n \ge C\varrho^2\sigma^2\log(d)$ for some constant $C = C(c, c', c'') > 0$. For brevity, we say n is sufficiently large if we take a sufficiently large C. Recall that we take $\mu_3^* = 1/\sigma\cdot\{\lambda_{\min}\big((M^\top M/n)_{S,S}\big) - \delta + \min_{i\in S}[M^\top\epsilon]_i/(nx_i^*)\}$. We now check the remaining conditions required in Theorem 8. Note that the assumption $Y_{11}^* > 0$ is automatically true by **9A**. Next, we take $\delta := 1 + \max\{\lambda_{\min}\big((M^\top M/n)_{S,S}\big) - 1 - c'', 0\} \ge 1$. μ_3^* is indeed nonnegative due to **8D** and **2A** with $L^2 = c$, because $\mu_3^* \ge (c-1)/(2\sigma) > 0$ (if $\delta = 1$) or $\mu_3^* \ge c''/(2\sigma) > 0$ (if $\delta > 1$) for sufficiently large n. From **8C**, **E1** is true for sufficiently large n. Next, we focus on **E2**. We first take $\mu_2^* := -c''$, and now we show that it is a valid choice by checking $\mu_2^* \in (-\infty, -\lambda_{\min}\big((M^\top M/n)_{S,S}\big)+\delta]$. Note that if $\delta = 1$, we have $\lambda_{\min}\big((M^\top M/n)_{S,S}\big) - 1 - c'' \le 0$, and therefore $-\lambda_{\min}\big((M^\top M/n)_{S,S}\big) + \delta \ge -c''$; on the contrary, if $\delta > 1$, we have $-\lambda_{\min}\big((M^\top M/n)_{S,S}\big) + \delta = -c''$. This implies that we can take $\mu_2^* = -c''$ in both cases.

Next, we construct Θ_1 and Θ_2 as required in **E2**. We take $\Theta_1 = \Theta_1^B + \Theta_1^E + \Theta_1^F$ and $\Theta_2 = \Theta_2^A + \Theta_2^B + \Theta_2^C + \Theta_2^D + \Theta_2^E + \Theta_2^F$. It still remains to (a) give valid choices for the constants $\bar{c}$, $\hat{c}$, $\tilde{c}$, and $\check{c}$ in Θ_1^B, Θ_1^E and Θ_1^F such that these three matrices are positive semidefinite; (b) show that $\Theta = \Theta_1 + \Theta_2$; and (c) prove that $\|\Theta_2\|_\infty < \mu_3^*$.

For (a), it suffices to show that we can take $\bar{c}$, $\hat{c}$, $\tilde{c}$, and $\check{c}$ in a way such that the first two terms in the definition of Θ_1^E sum up to a positive semidefinite matrix, and the first two terms in the definition of Θ_1^F sum up to a positive semidefinite matrix. From **8H**, we obtain $(M_1^\top M_1/n)_{S^c,S^c} \succeq (c'\sigma - C_8\sigma\sqrt{\log(d)/n})1_{d-\sigma}1_{d-\sigma}^\top$, so it suffices to give some choices of these constants such that $\hat{c}1_{d-\sigma}1_{d-\sigma} - 1_{d-\sigma}u^\top Bu1_{d-\sigma}^\top \succeq 0$ and $c'\sigma - C_8\sigma\sqrt{\log(d)/n} - (\bar{c}+\hat{c}+\tilde{c}+\check{c}) \ge 0$. We first take $\hat{c} = u^\top Bu$, where the definition of B can be found after the proof of Lemma 8. We then validate the choice by showing $u^\top Bu = \sigma + \mathcal{O}(\sqrt{\sigma})$. Indeed, since $c' > 1$, this shows $u^\top Bu < c'\sigma$ for some moderately large σ, making it possible to attain a nonnegative

$c'\sigma - Bc'\sigma\sqrt{\log(d)/n} - (\bar{c}+\hat{c}+\tilde{c}+\check{c})$, for sufficiently large n. Observe that

$$u^\top Bu = u^\top\Big(H^0 + \frac{1}{Y_{11}^*}x_S^*(x_S^*)^\top\Big)u + 2\frac{u^\top x_S^*}{Y_{11}^*}(y_S^*)^\top H^0 u + \frac{(u^\top x_S^*)^2}{(Y_{11}^*)^2}(y_S^*)^\top H^0 y_S^*.$$

By **2A** and **8D**, we have $Y_{11}^* \geq \sigma\Big(1 - c_1\sqrt{\sigma/n} - C_4\sqrt{\varrho^2\log(d)/n}\Big)$. Thus, when n is large enough, we obtain $1/Y_{11}^* \cdot \leq 1/\sigma(1+2c_1\sqrt{\sigma/n} + 2C_4\sqrt{\varrho^2\log(d)/n})$. Recall that $H^0 = I_\sigma - x_S^*(x_S^*)^\top/\sigma$, we then have

$$\begin{aligned}
&u^\top\Big(H^0 + \frac{1}{Y_{11}^*}x_S^*(x_S^*)^\top\Big)u \\
&\overset{\mathbf{8G}}{\leq} \sigma\Big(1 + C_7\sqrt{\frac{\sigma\log(d)}{n}}\Big)^2 + \sigma\cdot\Big(2c_1\sqrt{\frac{\sigma}{n}} + 2C_4\sqrt{c\frac{\varrho^2\log(d)}{n}}\Big) = \sigma + \mathcal{O}(\sqrt{\sigma}),
\end{aligned}$$

when n is sufficiently large. For the remaining terms, we see that

$$\begin{aligned}
&\Big|\frac{u^\top x_S^*}{Y_{11}^*}(y_S^*)^\top H^0 u\Big| \\
&\overset{\mathbf{8G}}{\leq} \sigma\Big(1 + C_7\sqrt{\frac{\sigma\log(d)}{n}}\Big)\cdot\frac{1}{\sigma}\Big(1 + 2c_1\sqrt{\frac{\sigma}{n}} + 2C_4\sqrt{\frac{\varrho^2\log(d)}{n}}\Big)\cdot\left\|H^0 y_S^*\right\|_2\|u\|_2 \\
&\overset{\mathbf{9D,8G}}{\leq} \mathcal{O}(\sqrt{\sigma}),
\end{aligned}$$

$$\begin{aligned}
&\frac{(u^\top x_S^*)^2}{(Y_{11}^*)^2}(y_S^*)^\top H^0 y_S^* \\
&\overset{\mathbf{8G,9B,9D}}{\leq} \Big(1 + C_7\sqrt{\frac{\sigma\log(d)}{n}}\Big)^2\Big(1 + 2c_1\sqrt{\frac{\sigma}{n}} + 2C_4\sqrt{\frac{\varrho^2\log(d)}{n}}\Big)^2\cdot\frac{1}{4} \leq \mathcal{O}(1).
\end{aligned}$$

Finally, we take $0 < \tilde{c}, \check{c} \ll 1$ small enough, and $\bar{c} = c'\sigma - \hat{c} - C_8\sigma\sqrt{\log(d)/n} - \tilde{c} - \check{c}$, to enforce $c'\sigma - C_8\sigma\sqrt{\log(d)/n} - (\bar{c}+\hat{c}+\tilde{c}+\check{c}) \geq 0$. We can verify that $\bar{c} > 0$ if n and σ are sufficiently large and $\tilde{c}, \check{c}$ are chosen to be sufficiently small.

Checking the validity of (b) is straightforward by direct calculation. For (c), we first show $\left\|\Theta_2^F\right\|_\infty = \mathcal{O}\Big(\sigma\log(d)/n + \sqrt{\log(d)/n}\Big)$, which is indeed true because $\left\|\Theta_2^F\right\|_\infty \leq \left\|vv^\top/\check{c}\right\|_\infty + \left\|(M_2^\top M_2/n)_{S^c,S^c} - c'' I_{d-\sigma}\right\|_\infty = \mathcal{O}\Big(\sigma\log(d)/n + \sqrt{\log(d)/n}\Big)$, where the last equality is due to **8E** and **2B** with $L^2 = c''$. Combing this fact and Lemma 10, we obtain that

$$\begin{aligned}
\|\Theta_2\|_\infty &\leq \left\|\Theta_2^A\right\|_\infty + \left\|\Theta_2^B\right\|_\infty + \left\|\Theta_2^C\right\|_\infty + \left\|\Theta_2^D\right\|_\infty + \left\|\Theta_2^E\right\|_\infty + \left\|\Theta_2^F\right\|_\infty \\
&\leq \mathcal{O}\Big(\frac{\varrho^2\log(d)}{n}\Big) + \mathcal{O}\Big(\frac{\varrho\sigma\log(d)}{n} + \frac{\varrho^2\sigma\log(d)}{n}\Big) + \mathcal{O}\Big(\sqrt{\frac{\varrho^2\log(d)}{n}} + \frac{\sqrt{\sigma}\varrho\log(d)}{n}\Big) \\
&\quad + \mathcal{O}\Big(\frac{\varrho^2\log(d)}{n\sigma}\Big) + \mathcal{O}\Big(\frac{(\sqrt{\sigma\log(d)}+\sigma)^2}{n} + \frac{\sigma\log(d)}{n}\Big) + \mathcal{O}\Big(\frac{\sigma\log(d)}{n} + \sqrt{\frac{\log(d)}{n}}\Big) \\
&\leq \frac{1}{4\sigma}\min\{c-1, c''\} < \frac{1}{2\sigma}\min\{c-1, c''\} \leq \mu_3^*.
\end{aligned}$$

w.h.p. when $n \geq C\varrho^2\sigma^2\log(d)$, for some large constant $C = C(c, c', c'') > 0$. □

References

[1] Erik Agrell, Thomas Eriksson, Alexander Vardy, and Kenneth Zeger. Closest point search in lattices. *IEEE Transactions on Information Theory*, 48(8):2201–2214, 2002.

[2] Arash A Amini and Martin J Wainwright. High-dimensional analysis of semidefinite relaxations for sparse principal components. In *IEEE International Symposium on Information Theory*, pages 2454–2458, 2008.

[3] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 9.2.*, 2020.

[4] Ulaş Ayaz and Holger Rauhut. Nonuniform sparse recovery with subgaussian matrices. *arXiv preprint arXiv:1007.2354*, 2010.

[5] Somsubhra Barik and Haris Vikalo. Sparsity-aware sphere decoding: Algorithms and complexity analysis. *IEEE Transactions on Signal Processing*, 62(9):2212–2225, 2014.

[6] John E Beasley. Or-library: distributing test problems by electronic mail. *Journal of the operational research society*, 41(11):1069–1072, 1990.

[7] Dimitris Bertsimas, Angela King, and Rahul Mazumder. Best subset selection via a modern optimization lens. *The annals of statistics*, 44(2):813–852, 2016.

[8] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[9] Emmanuel Candes and Terence Tao. The dantzig selector: Statistical estimation when p is much larger than n. *Annals of statistics*, 35(6):2313–2351, 2007.

[10] Emmanuel J Candes and Terence Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[11] Antoni B Chan, Nuno Vasconcelos, and Gert RG Lanckriet. Direct convex relaxations of sparse svm. In *Proceedings of the 24th international conference on Machine learning*, pages 145–153, 2007.

[12] Moses Charikar and Anthony Wirth. Maximizing quadratic programs: Extending grothendieck's inequality. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 54–60. IEEE, 2004.

[13] Guangjin Chen, Jincheng Dai, Kai Niu, and Chao Dong. Sparsity-inspired sphere decoding (si-sd): A novel blind detection algorithm for uplink grant-free sparse code multiple access. *IEEE Access*, 5:19983–19993, 2017.

[14] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

[15] Elaine Crespo Marques, Nilson Maciel, Lírida Naviner, Hao Cai, and Jun Yang. A review of sparse recovery algorithms. *IEEE Access*, 7:1300–1322, 2019.

[16] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[17] David L Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

[18] Yingjie Fei and Yudong Chen. Hidden integrality of sdp relaxations for sub-gaussian mixture models. In *Conference On Learning Theory*, pages 1931–1965. PMLR, 2018.

[19] Axel Flinth and Gitta Kutyniok. Promp: A sparse recovery approach to lattice-valued signals. *Applied and Computational Harmonic Analysis*, 45(3):668–708, 2018.

[20] Michael R. Garey and David S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness.* W. H. Freeman & Co., USA, 1990.

[21] Huanmin Ge and Peng Li. The dantzig selector: recovery of signal via ℓ_1 - $\alpha\ell_2$ minimization. *Inverse Problems*, 38(1):015006, 2021.

[22] Gene H Golub. Some modified matrix eigenvalue problems. *Siam Review*, 15(2):318–334, 1973.

[23] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1, March 2014.

[24] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2022.

[25] Bruce Hajek, Yihong Wu, and Jiaming Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016.

[26] Takayuki Iguchi, Dustin G. Mixon, Jesse Peterson, and Soledad Villar. On the tightness of an sdp relaxation of k-means, 2015.

[27] Sandra Keiper, Gitta Kutyniok, Dae Gwan Lee, and Götz E Pfander. Compressed sensing for finite-valued signals. *Linear Algebra and its Applications*, 532:570–613, 2017.

[28] Harold W Kuhn and Albert W Tucker. Nonlinear programming. In *Traces and emergence of nonlinear programming*, pages 247–258. Springer, 2014.

[29] Monique Laurent and Franz Rendl. Semidefinite programming and integer programming. In K. Aardal, G. Nemhauser, and R. Weismantel, editors, *Handbook on Discrete Optimization*, pages 393–514. Elsevier B.V., December 2005.

[30] Peng Li and Wengu Chen. Signal recovery under mutual incoherence property and oracle inequalities. *Frontiers of Mathematics in China*, 13(6):1369–1396, 2018.

[31] Zang Li and W. Trappe. Collusion-resistant fingerprints from wbe sequence sets. pages 1336 – 1340 Vol. 2, 06 2005.

[32] Po-Ling Loh. Statistical consistency and asymptotic normality for high-dimensional robust m-estimators. *The Annals of Statistics*, 45(2):866–896, 2017.

[33] Karim Lounici. Sup-norm convergence rate and sign concentration property of lasso and dantzig estimators. *Electronic Journal of statistics*, 2:90–102, 2008.

[34] Alan Miller. *Subset selection in regression.* CRC Press, 2002.

[35] Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis.* Cambridge university press, 2017.

[36] Mohamed Ndaoud and Alexandre B Tsybakov. Optimal variable selection and adaptive noisy compressed sensing. *IEEE Transactions on Information Theory*, 66(4):2517–2532, 2020.

[37] Richard Obermeier and Jose Angel Martinez-Lorenzo. Sensing matrix design via mutual coherence minimization for electromagnetic compressive imaging applications. *IEEE Transactions on Computational Imaging*, 3(2):217–229, 2017.

[38] Simon J. D. Prince. *Computer Vision: Models, Learning, and Inference.* Cambridge University Press, USA, 1st edition, 2012.

[39] Behrooz Razeghi, Slava Voloshynovskiy, Dimche Kostadinov, and Olga Taran. Privacy preserving identification using sparse approximation with ambiguization. In *2017 IEEE Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2017.

[40] Galen Reeves, Jiaming Xu, and Ilias Zadik. The all-or-nothing phenomenon in sparse linear regression. In *Conference on Learning Theory*, pages 2652–2663. PMLR, 2019.

[41] M Ross Kunz and Yiyuan She. Multivariate calibration maintenance and transfer through robust fused lasso. *Journal of Chemometrics*, 27(9):233–242, 2013.

[42] Hampei Sasahara, Kazunori Hayashi, and Masaaki Nagahara. Multiuser detection based on map estimation with sum-of-absolute-values relaxation. *IEEE Transactions on Signal Processing*, 65(21):5621–5634, 2017.

[43] Nuno MB Souto and Hugo André Lopes. Efficient recovery algorithm for discrete valued sparse signals using an admm approach. *IEEE Access*, 5:19562–19569, 2017.

[44] Susanne Sparrer and Robert FH Fischer. Adapting compressed sensing algorithms to discrete sparse signals. In *WSA 2014; 18th International ITG Workshop on Smart Antennas*, pages 1–8. VDE, 2014.

[45] Susanne Sparrer and Robert FH Fischer. Soft-feedback omp for the recovery of discrete-valued sparse signals. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 1461–1465. IEEE, 2015.

[46] Sergios Theodoridis. Chapter 9 - sparsity-aware learning: Concepts and theoretical foundations. In Sergios Theodoridis, editor, *Machine Learning*, pages 403–448. Academic Press, Oxford, 2015.

[47] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996.

[48] Roman Vershynin. How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability*, 25, 04 2010.

[49] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[50] Martin J Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (lasso). *IEEE Transactions on Information Theory*, 55(5):2183–2202, 2009.

[51] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2):210–227, 2008.

[52] Zuodong Yang, Yong Wu, Wenteng Zhao, Yicong Zhou, Zongqing Lu, Weifeng Li, and Qingmin Liao. A novel illumination-robust local descriptor based on sparse linear regression. *Digital Signal Processing*, 48:269–275, 2016.

[53] Tarik Yardibi, Jian Li, Peter Stoica, and Louis N. Cattafesta III. Sparse representations and sphere decoding for array signal processing. *Digital Signal Processing*, 22(2):253–262, 2012.

[54] Tong Zhang. Adaptive forward-backward greedy algorithm for learning sparse representations. *IEEE transactions on information theory*, 57(7):4689–4708, 2011.

[55] Peng Zhao and Bin Yu. On model selection consistency of lasso. *The Journal of Machine Learning Research*, 7:2541–2563, 2006.

[56] Yun-Bin Zhao and Duan Li. A theoretical analysis of sparse recovery stability of dantzig selector and lasso. *arXiv preprint arXiv:1711.03783*, 2017.

[57] Hao Zhu and Georgios B. Giannakis. Exploiting sparse user activity in multiuser detection. *IEEE Transactions on Communications*, 59(2):454–465, 2011.