PERFORMANCE ANALYSIS OF QUASI INTEGER LEAST SQUARES SOLVERS BASED ON SEMIDEFINITE RELAXATION*

MIKALAI KISIALIOU AND ZHI-QUAN LUO†

Abstract. We consider a random Integer Least Squares (ILS) problem. This NP-hard problem naturally arises in digital communications as the maximum-likelihood detection problem. We analyze two probabilistic quasi-ILS algorithms based on semidefinite relaxations: the SDR algorithm for binary variables and the PSK algorithm for constant modulus variables. Both algorithms are capable of delivering a near-optimal solution with a polynomial worst-case complexity. For a general class of random parameters, we prove that the SDR algorithm provides a constant factor approximation in terms of the objective value, and this constant factor remains bounded with increasing problem size. For the PSK algorithm we show that each local maximum of the low-rank semidefinite relaxation that is feasible for the ILS problem achieves at least a half of the minimum relative objective value, and for the binary case even yields an exact ILS solution. Our analysis shows that the ILS problem can be well approximated in polynomial time by the two semidefinite relaxation strategies.

Key words. Duality, integer least squares problem, performance analysis, random matrix theory, semidefinite relaxation.

AMS subject classifications.

1. Introduction. Consider the following Integer Least Squares (ILS) problem:

$$\mathbf{s}_{ils} = \arg\min_{\mathbf{s}\in\mathcal{S}^n} \|\mathbf{y} - \sqrt{\rho}\ \mathbf{H}\mathbf{s}\|^2. \tag{1.1}$$

This is a fundamental problem for many practical applications. For example, in digital communications the maximum likelihood detectors solve problem (1.1) where the parameters $\{\mathbf{H}, \mathbf{y}, \rho\}$ are defined by the vector channel with input $\mathbf{s} \in \mathbb{C}^n$ and output $\mathbf{y} \in \mathbb{C}^m$:

$$\mathbf{y} = \sqrt{\rho}\ \mathbf{H}\mathbf{s} + \mathbf{v}, \quad \mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}). \tag{1.2}$$

The channel matrix $\mathbf{H} \in \mathbb{C}^{m\times n}$ and noise vector $\mathbf{v} \in \mathbb{C}^m$ are appropriately scaled so that parameter ρ signifies the so-called signal-to-noise-ratio per each output dimension, i.e.:

$$\rho = \frac{E_{\mathbf{H},\mathbf{s}}\left\{\|\sqrt{\rho}\ \mathbf{H}\mathbf{s}\|^2\right\}}{E_{\mathbf{v}}\left\{\|\mathbf{v}\|^2\right\}},$$

where $E_{\mathbf{H},\mathbf{s}}\{\cdot\}$ ($E_{\mathbf{v}}\{\cdot\}$) denotes expectation w.r.t. the distribution of $\mathbf{H}$ and $\mathbf{s}$ (w.r.t. the distribution of $\mathbf{v}$).

For the most part of this work, we consider problem (1.1) with the feasible set $\mathcal{S} = \{\exp(j\phi) \mid \phi \in \{2\pi k/M, k = 0, \ldots, M-1\}\}$ and its special case $\mathcal{S} = \{-1, +1\}$. The ILS problem (1.1) is NP-hard due to the combinatorial structure in $\mathcal{S}^n$. For small problem size $\{n, m\}$, the exhaustive search can be used to solve (1.1). However, for large n, m or $|\mathcal{S}|$, this approach is prohibitively expensive.

Much of the past research was motivated by the desire to obtain an approximate solution to (1.1) with manageable computational complexity. Unfortunately,

*Manuscript received January, 2008. This research is supported in part by the National Science Foundation, Grant No. DMS-0610037. A preliminary version of this paper was published in the Proceedings of ICASSP'03, Hong Kong and the Proceedings of ICASSP'05, Philadelphia, U.S.A.

†M. Kisialiou and Z.-Q. Luo are with the Department of Electrical and Computer Engineering, University of Minnesota. Email address: {kisi0004,luozq@umn.edu}

all existing approaches either exhibit significant performance degradation, e.g. linear relaxation solvers, or suffer from exponential complexity, e.g. the so-called Sphere Decoding (SD) algorithm (see [4], [2], [6, Chapter 5] and [30]). For instance, for the binary feasible set the conventional linear solvers relax the integer constraint $\mathbf{s} \in \{-1, +1\}^n$ and impose a penalty term with a weight parameter σ:

$$\mathbf{s}_{lin} = \arg\min_{\mathbf{s}\in\mathbb{C}^n} \|\mathbf{y} - \sqrt{\rho}\ \mathbf{Hs}\|^2 + \sigma\|\mathbf{s}\|^2.$$

This unconstrained problem can be solved analytically. Rounding the solution to the integer feasible set $\mathbf{s} \in \{-1, +1\}^n$ leads to

$$\mathbf{s}_{lin} = \text{sign}\left(\left(\rho\,\mathbf{H}^\dagger\mathbf{H} + \sigma\,\mathbf{I}\right)^{-1}\mathbf{H}^\dagger\mathbf{y}\right), \tag{1.3}$$

where $(\cdot)^\dagger$ denotes matrix transpose (Hermitian transpose for complex matrices), and parameter $\sigma = 1, 0, \infty$ can be selected to specify various common types of solvers. While these linear solvers have $\mathcal{O}(n^3)$ complexity, they suffer from significant performance loss as compared to the performance of the exact ILS solver. To improve the solution of (1.1), one can combine the linear solvers with a limited exhaustive search, yielding the SD algorithm. More specifically, the SD algorithm first computes the unconstrained minimum:

$$\mathbf{s}_{zf} = (1/\sqrt{\rho})\left(\mathbf{H}^\dagger\mathbf{H}\right)^{-1}\mathbf{H}^\dagger\mathbf{y},$$

and then performs a local exhaustive search within an ellipsoid centered at $\mathbf{s}_{zf}$. When ρ is large the ILS solution is statistically close to $\mathbf{s}_{zf}$, and the SD algorithm is quite efficient. However, for small ρ or for large size problems, this hybrid strategy still suffers from exponential complexity [10].

Is there an algorithm that can deliver a *provably* near-optimal solution for problem (1.1) and yet still has a polynomial time complexity? In this paper we settle this question in the affirmative. More specifically, we analyze two polynomial time semidefinite relaxation strategies to solve the ILS problem: i) the SDR algorithm [17] which is based on a convex semi-definite relaxation of (1.1), and ii) the PSK algorithm which is based on a non-convex low-rank relaxation of (1.1). Both algorithms have a polynomial complexity ($O(n^{3.5})$ and $O(n^3)$ respectively).

For a general class of random models (1.2) whereby the entries of $\{\mathbf{H},\mathbf{v}\}$ are independent and identically distributed with zero-mean, we show that the SDR algorithm can achieve near-optimal performance with at most a constant objective value loss when $n, m \to \infty$ such that $m/n \to \gamma$ for some constant $\gamma \geq 1$.

Semidefinite relaxation is a powerful optimization technique that can generate high quality approximate solutions for nonconvex quadratic optimization problems. This technique is based on a careful relaxation of the original problem as a convex semidefinite program. The latter is efficiently solvable by the interior point methods. Semidefinite relaxation achieved a notable success in solving the NP-hard MAX-CUT problem [5] in combinatorial optimization, delivering a provable worst-case approximation quality of 87%. That is, for any weighted graph, the size of the cut generated by the semidefinite relaxation is at least 87% of the size of the maximum cut. Semidefinite relaxation can also be shown to provide strong (a constant factor of $2/\pi$) approximation for the *maximization* of a homogeneous convex quadratic function over the feasible set $\mathcal{S}^n = \{-1, +1\}^n$ [21]. Unfortunately, the ILS problem (1.1) involves the *minimization* (rather than *maximization*) of a convex quadratic function over $\mathcal{S}^n$. This renders the existing worst-case semidefinite relaxation bounds in the optimization literature inapplicable. In fact, one can construct instances to show that the worst-case approximation quality of the SDR algorithm can be arbitrarily bad for the problem (1.1). Will this imply a poor performance of the SDR algorithm for a typical practical application, e.g. maximum likelihood detection? The answer is no since, in digital communication, the performance of a detection algorithm is always *averaged* over all possible realizations of channel matrix $\mathbf{H}$ and noise $\mathbf{v}$. While the average performance is a standard, well-defined benchmark to compare detection algorithms, the worst-case performance is meaningless since it is always bad regardless of detectors. In this paper, we show that, for a general class of random channels (1.2), the SDR algorithm does allow a constant average approximation ratio:

$$\frac{\|\mathbf{y}-\sqrt{\rho}\,\mathbf{H}\mathbf{s}_{sdr}\|^2}{\|\mathbf{y}-\sqrt{\rho}\,\mathbf{H}\mathbf{s}_{ils}\|^2} \le c(\rho,\gamma), \tag{1.4}$$

where $\mathbf{s}_{sdr}$ is the output of the SDR algorithm and

$$c(\rho,\gamma) = 1 + \mathcal{O}(\rho^{1-\alpha}),\ \text{where } \alpha = \begin{cases} \alpha = 1/3, \ \gamma = 1, \\ \alpha = 1/2, \ \gamma > 1. \end{cases}$$

It is usually difficult to evaluate performance of an algorithm for a random problem when size n, m and ρ are finite and fixed. What is more tractable is to examine the algorithm's asymptotic behavior when either n, m or ρ, or both tend to infinity. There are two complementary approaches to the asymptotic performance analysis. The first is to fix n, m and let $\rho \to \infty$. The second approach is to fix the ρ and let $n, m \to \infty$. The second case is preferable since we can compare the *computational complexity* of various algorithms, as well as their asymptotic performance for large systems. This type of large-system analysis provides a natural way to study the complexity/performance tradeoff for an algorithm.

The existing large-system analysis of ILS solvers applied to channel detection was based on two analytic tools: random matrix theory [26, 28] and the replica method from statistical mechanics [7, 25]. In this paper we use the Marcenko-Pastur theorem from random matrix theory to analyze the asymptotic performance of the SDR algorithm. The result of Marcenko-Pastur characterizes the limit of the empirical eigenvalue distribution of the random matrix $\mathbf{H}^\dagger\mathbf{H}$ when n, m both tend to infinity while $m/n \to \gamma$ for some constant γ. It provides a key to establish the constant factor optimality of the SDR algorithm. Although our performance analysis is asymptotic, the exponentially fast convergence of the empirical eigenvalue distribution makes our results a useful analytical benchmark even for systems of practical size.

For the constant modulus feasible set of the form $\{\exp(j\phi)\}$, we will also consider an alternative algorithm based on a non-convex low-rank semidefinite relaxation [1], called the PSK algorithm [14] hereafter. This algorithm is implemented with the coordinate descent strategy on the homotopy of the feasible region of (1.1). When initialized with $\mathbf{H}^\dagger\mathbf{H}$, the remaining worst-case complexity is polynomial, $\mathcal{O}\left((\max\{n,m\})^2\right)$. Since the PSK algorithm uses a tighter relaxation than the SDR algorithm, it demonstrates superior empirical performance. However, the resulting tighter relaxation is not convex. In contrast to the SDR algorithm, theoretical results for this non-convex relaxation are scarce. In this paper we show that each local minimum of the low-rank semidefinite relaxation that is feasible for the ILS problem is within a factor of 2 of the minimum relative objective value, and for the binary case even yields an exact ILS solution.

2. Polynomial-time quasi-ILS solvers.

2.1. A convex semidefinite relaxation (SDR algorithm). Denote $\mathcal{S}_M^{n+1} \triangleq \{\exp(j\boldsymbol{\phi}) \mid \phi_i \in \{2\pi k/M, k = 0, \ldots, M-1\}, i = 1, \ldots, n+1\}$. For a general set $\mathcal{S}_M^{n+1}$, the ILS problem (1.1) allows an equivalent reformulation [17]:

$$\begin{aligned} f_{ils} \triangleq \min\ & \mathrm{Trace}(\mathbf{QX}) \\ \text{s.t.}\ & \mathbf{X} = \mathbf{x}\mathbf{x}^\dagger, \quad \mathbf{x} \in \mathcal{S}_M^{n+1}, \end{aligned} \tag{2.1}$$

where matrix $\mathbf{Q}$ and vector $\mathbf{x}$ are defined as follows

$$\mathbf{Q} = \begin{bmatrix} \rho\,\mathbf{H}^\dagger\mathbf{H} & -\sqrt{\rho}\,\mathbf{H}^\dagger\mathbf{y} \\ -\sqrt{\rho}\,\mathbf{y}^\dagger\mathbf{H} & \|\mathbf{y}\|^2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{s} \\ 1 \end{bmatrix}. \tag{2.2}$$

Notice that any feasible matrix $\mathbf{X}$ in (2.1) is positive semidefinite with all diagonal entries equal to 1. Let us use $\mathrm{diag}(\mathbf{X})$ to signify the vector with entries equal to the diagonal entries of matrix $\mathbf{X}$, use the notation $\mathbf{e}$ to represent the all-one vector and $\mathbf{X} \succeq 0$ to denote a Hermitian positive semidefinite matrix $\mathbf{X}$. The algorithm [17] relaxes the two constraints of (2.1) to $\{\mathrm{diag}(\mathbf{X}) = \mathbf{e}, \mathbf{X} \succeq 0\}$. This leads to the following semidefinite relaxation:

$$\begin{aligned} f_{sdp}(\mathbf{X}) \triangleq \min\ & \mathrm{Trace}(\mathbf{QX}) \\ \text{s.t.}\ & \mathrm{diag}(\mathbf{X}) = \mathbf{e}, \quad \mathbf{X} \succeq 0. \end{aligned} \tag{2.3}$$

Let $\mathbf{X}_{opt}$ be an optimal solution of (2.3). Various (often randomized) rounding procedures can be used to generate an estimate of the optimal solution $\mathbf{s}_{ils}$, see [16, 17]. In this work we analyze the following simple rounding procedure for the binary case, although a more general procedure for a general $\mathcal{S}_M^{n+1}$ set is possible [16]:

- Given $\mathbf{X}_{opt}$, define $\mathbf{Z} \in \mathbb{C}^{n \times n}$ and $\mathbf{z} \in \mathbb{C}^n$ such that

$$\mathbf{X}_{opt} = \begin{bmatrix} \mathbf{Z} & \mathbf{z} \\ \mathbf{z}^\dagger & 1 \end{bmatrix}. \tag{2.4}$$

- Define a Bernoulli distribution for the i-th entry of $\mathbf{x}$ (denoted by x_i), $i = 1, \ldots, n$:

$$\begin{aligned} P\{x_i = +1\} &= (1 + \mathrm{Re}\{z_i\})/2, \\ P\{x_i = -1\} &= (1 - \mathrm{Re}\{z_i\})/2, \end{aligned} \tag{2.5}$$

  where z_i is the i-th entry of $\mathbf{z}$ and $\mathrm{Re}\{\cdot\}$ denotes the real part of a complex number.
- Generate a fixed number D of i.i.d. vector samples $\mathbf{x}_d$, $\mathbf{x}_d \in \{-1, +1\}^{n+1}$, $d = 1, \ldots, D$, according to (2.5) and last bit set to $x_{n+1} = 1$.
- Pick $\mathbf{s}_{sdr} \triangleq \arg\min_d \mathbf{x}_d^\dagger \mathbf{Q} \mathbf{x}_d$ and set the best achieved objective value:

$$f_{sdr} \triangleq \mathbf{s}_{sdr}^\dagger \mathbf{Q} \mathbf{s}_{sdr}. \tag{2.6}$$

- Output the objective value f_{sdr} and the estimate $\mathbf{s}_{sdr}$ (discarding its last bit).

2.2. A non-convex semidefinite relaxation (PSK algorithm). Let us consider a tighter (non-convex) semidefinite relaxation of the ILS problem (2.1). This relaxation can deliver superior empirical performance in some cases, although a thorough theoretical analysis is substantially more difficult [14]. Notice that the constraint $\mathbf{x} \in \mathcal{S}_M^{n+1}$ implies that each entry $x_i, i = 1, \ldots, n+1$, can be represented as $x_i = \exp(j\phi_i)$, where $\phi_i \in \{2\pi k/M \mid k = 1, \ldots, M\}$. By allowing arbitrary phases, we can relax the constraint $\mathbf{x} \in \mathcal{S}_M^{n+1}$ to $\{|x_i| = 1,\ i = 1, 2, ..., n+1\}$. Define $\mathbf{X} \triangleq \mathbf{x}\mathbf{x}^\dagger$. Then, the relaxed feasible set can be written as $\{\mathrm{diag}(\mathbf{X}) = \mathbf{e}, \mathrm{rank}(\mathbf{X}) = 1, \mathbf{X} \succeq 0\}$. Thus, we obtain the following non-convex relaxation of (2.1):

$$\begin{array}{rl} \min & \mathrm{Trace}(\mathbf{QX}) \\ \text{s.t.} & \mathrm{diag}(\mathbf{X}) = \mathbf{e}, \quad \mathbf{X} \succeq 0, \\ & \mathrm{rank}(\mathbf{X}) = 1. \end{array} \tag{2.7}$$

When $\mathbf{X}$ is real the problem (2.7) is equivalent to the ILS problem (2.1). For the complex case, let $\mathbf{x}_R$ and $\mathbf{x}_I$ represent the real and imaginary parts of vector $\mathbf{x}$. For a real symmetric matrix $\mathbf{Q}$ the objective function in (2.7) is:

$$\mathrm{Trace}(\mathbf{Q}\mathbf{x}\mathbf{x}^\dagger) = \mathrm{Trace}\left(\mathbf{Q}(\mathbf{x}_R\mathbf{x}_R^\dagger + \mathbf{x}_I\mathbf{x}_I^\dagger)\right) = \mathrm{Trace}\left(\mathbf{Q}\bar{\mathbf{X}}\right),$$

where $\bar{\mathbf{X}} \triangleq \mathbf{x}_R\mathbf{x}_R^\dagger + \mathbf{x}_I\mathbf{x}_I^\dagger \succeq 0$. Therefore, the complex rank-1 constraint $\mathbf{X} = \mathbf{x}\mathbf{x}^\dagger$ is equivalent to the real-valued constraints $\left\{\mathrm{rank}(\bar{\mathbf{X}}) \le 2, \bar{\mathbf{X}} \succeq 0\right\}$. Thus, the relaxation of the ILS problem (1.1) with real matrix $\mathbf{Q}$ is given by:

$$\begin{array}{rl} \min & \mathrm{Trace}\left(\mathbf{Q}\bar{\mathbf{X}}\right) \\ \text{s.t.} & \mathrm{diag}(\bar{\mathbf{X}}) = \mathbf{e}, \quad \bar{\mathbf{X}} \succeq 0, \\ & \mathrm{rank}(\bar{\mathbf{X}}) \le 2. \end{array} \tag{2.8}$$

The above relaxation was first considered in [1] as a low-rank semidefinite relaxation of a Boolean quadratic maximization problem in the context of combinatorial optimization. For applications in digital communications, we shall be interested in this relaxation for general $\mathbf{x} \in \mathcal{S}_M^{n+1}$ set with complex or real $\mathbf{Q}$. For a complex matrix $\mathbf{Q}$ we can double the problem dimension and obtain a low-rank relaxation similar to (2.8).

In this work we use a different formulation when $\mathbf{Q}$ is complex. In particular, define the following set:

$$\mathbb{U}^{n+1} \triangleq \{\ \mathbf{x} \mid |x_i| = 1, \forall\, i = 1, \ldots, n+1\}.$$

The feasible set in (2.7) is equivalent to the set given by: $\left\{\mathbf{X} = \mathbf{x}\mathbf{x}^\dagger, \mathbf{x} \in \mathbb{U}^{n+1}\right\}$, hence, we can write (2.7) in the form

$$\min_{\mathbf{x} \in \mathbb{U}^{n+1}} \mathbf{x}^\dagger \mathbf{Q} \mathbf{x}. \tag{2.9}$$

The optimal $\mathbf{x}$ can be determined up to a common phase rotation of all entries of $\mathbf{x}$. We can remove this phase ambiguity by the condition $x_{n+1} = 1$.

The subsequent sections will be devoted to the theoretical performance analysis of the algorithms based on relaxations (2.3) and (2.9).

3. Performance analysis of the SDR algorithm. In this section we present the analysis of the probabilistic approximation ratio of objective values for the SDR algorithm and the exact ILS solver in a large-system limit with both fixed ρ and

$\rho \to \infty$. To facilitate the analysis, we specify model (1.2) with variables $\mathbf{s} \in \{-1,+1\}^n$ and assume the entries of $\mathbf{H}$ and $\mathbf{v}$ are i.i.d. with the following normalization

$$\begin{aligned} E\{s_k\} &= 0, \quad E\{|s_k|^2\} = 1, \quad \forall\, k, \\ E\{H_{ik}\} &= 0, \quad E\{|H_{ik}|^2\} = 1/n, \;\forall\, i,k, \\ E\{v_i\} &= 0, \quad E\{|v_i|^2\} = 1, \quad \forall\, i. \end{aligned} \tag{3.1}$$

3.1. Approximation ratio. Let f_{sdr} (f_{ils}) denote the objective value of the SDR algorithm (optimal objective value of the ILS problem), c.f. (2.1) and (2.6):

$$\begin{aligned} f_{sdr} &\triangleq \|\mathbf{y} - \sqrt{\rho}\,\mathbf{H}\mathbf{s}_{sdr}\|^2, \\ f_{ils} &\triangleq \|\mathbf{y} - \sqrt{\rho}\,\mathbf{H}\mathbf{s}_{ils}\|^2, \end{aligned} \tag{3.2}$$

where $\mathbf{s}_{sdr}$ and $\mathbf{s}_{ils}$ are the outputs of the SDR algorithm and the exact ILS solver respectively. The approximation ratio of the SDR algorithm f_{sdr}/f_{ils} is always lower bounded: $1 \le f_{sdr}/f_{ils}$. In the worst case, as $n, m \to \infty$ such that $m/n \to \gamma$ ($\gamma \ge 1$), the approximation ratio f_{sdr}/f_{ils} can grow unbounded. However, we will show that this ratio is bounded *in probability*, i.e. $f_{sdr}/f_{ils} \le c(\rho,\gamma)$ for some constant $c(\rho,\gamma)$.

The approximation ratio f_{sdr}/f_{ils} depends on the quality of the semidefinite relaxation (2.3) and the quality of the subsequent randomized rounding procedure. First, in Lemmas 3.1 and 3.2 we will analyze the unrounded estimate $\mathbf{z}$ in (2.4) produced by the semidefinite relaxation (2.3). Then, we will analyze the randomized rounding procedure which generates $\mathbf{s}_{sdr}$ from $\mathbf{z}$. Combining the results we can obtain the desired bound on the approximation ratio (Theorem 3.3).

Due to symmetry, we assume without loss of generality that the vector $\mathbf{s} = \mathbf{e}$ is the solution of the ILS problem (1.1). The quality of the relaxation (2.3) is measured in terms of the distance to the optimal solution: $\Delta\mathbf{z} \triangleq \mathbf{e} - \mathbf{z}$. When $\Delta\mathbf{z} = \mathbf{0}$, the randomized rounding procedure generates the correct estimate of the solution so that $\mathbf{s}_{sdr} = \mathbf{s}$ with probability 1. Whenever $\Delta\mathbf{z} \ne \mathbf{0}$ the randomized rounding procedure rounds the estimate $\mathbf{z}$ to $\mathbf{s}_{sdr} \in \{-1,+1\}^n$ according to the distribution (2.5). Let $\|\Delta\mathbf{z}\|_1 \triangleq \sum_i |\Delta z_i|$. We define the limiting average distance to the optimal solution as

$$\ell(\rho,\gamma) \triangleq \lim_{\substack{n,m\to\infty \\ m/n\to\gamma}} \frac{\|\Delta\mathbf{z}\|_1}{n}. \tag{3.3}$$

We are interested in the rate at which the limiting average distance $\ell(\rho,\gamma)$ vanishes with increasing ρ. In general, as $n, m \to \infty$, it is possible to select a deterministic sequence of parameter realizations $\{\mathbf{H},\mathbf{v}\}$ such that $\ell(\rho,\gamma) \nrightarrow 0$ as $\rho \to \infty$. We will prove that for random choices of $\{\mathbf{H},\mathbf{v}\}$ such event happens with probability 0.

The probabilistic analysis of $\|\Delta\mathbf{z}\|_1$ in the limit $n, m \to \infty$ is enabled by random matrix theory which offers a collection of useful results on the asymptotic behavior of a large random matrix. Consider the random matrix $\mathbf{H}^\dagger\mathbf{T}\mathbf{H}$, where $\mathbf{H}$ and $\mathbf{T}$ are statistically independent, $\mathbf{H}$ is normalized as in (3.1) and $\mathbf{T}$ is a diagonal matrix. The empirical eigenvalue distribution function for this random matrix is defined as

$$F_{\mathbf{H}^\dagger\mathbf{T}\mathbf{H}}(x) \triangleq \frac{1}{n}\sum_{i=1}^{n} 1\{\lambda_i(\mathbf{H}^\dagger\mathbf{T}\mathbf{H}) \le x\}, \tag{3.4}$$

where λ_i are the eigenvalues and $1\{\cdot\}$ is the indicator function. This distribution is useful for evaluating a random quadratic form $\mathbf{u}^\dagger\mathbf{H}^\dagger\mathbf{T}\mathbf{H}\mathbf{u}$, where $\mathbf{u}$ is a random vector, independent of $\mathbf{H}$ and $\mathbf{T}$. As $n, m \to \infty$ such that $\gamma \triangleq \lim_{n,m\to\infty} m/n$, this empirical eigenvalue distribution (3.4) converges almost surely [24] to a non-random limit. In particular, when $\mathbf{T} = \mathbf{I}$, this distribution allows a closed-form analytical density function called Marcenko-Pastur density [18]:

$$g_\gamma(\lambda) = (1-\gamma)^+ \, \delta(\lambda) + \frac{\sqrt{(\lambda - a)^+ (b-\lambda)^+}}{2\pi\lambda},$$

where $(x)^+ \triangleq \max(0, x)$, $a \triangleq (1-\sqrt{\gamma})^2$ and $b \triangleq (1+\sqrt{\gamma})^2$.

Random matrix theory enables us to analyze the behavior of the randomly generated SDP (2.3) when its size goes to infinity. To estimate $\ell(\rho, \gamma)$ we use a strategy similar to [8]. Specifically, we consider the following optimization problem (c.f. (2.3)) for some positive constants α and β to be specified later:

$$\begin{aligned} f_0(\rho, \alpha, \beta) \triangleq \min \;\; & \mathrm{Trace}(\mathbf{L}_0 \mathbf{X}) \\ \text{s.t.} \;\; & \mathrm{diag}(\mathbf{X}) = \mathbf{e}, \quad \mathbf{X} \succeq 0, \\ & \mathrm{Trace}(\mathbf{R}_0 \mathbf{X}) \geq \frac{\beta n}{\rho^\alpha}, \end{aligned} \tag{3.5}$$

where

$$\mathbf{L}_0 \triangleq \begin{bmatrix} \mathbf{H}^\dagger \mathbf{H} & -\mathbf{H}^\dagger \mathbf{H} \mathbf{e} \\ -\mathbf{e}^\dagger \mathbf{H}^\dagger \mathbf{H} & \mathbf{e}^\dagger \mathbf{H}^\dagger \mathbf{H} \mathbf{e} \end{bmatrix}, \quad \mathbf{R}_0 \triangleq \begin{bmatrix} \mathbf{I} & -\mathbf{e} \\ -\mathbf{e}^\dagger & n \end{bmatrix}. \tag{3.6}$$

The objective function in this problem is an equivalent of the objective function in (2.3) with $\mathbf{v} = \mathbf{0}$, scaled by $1/\rho$. Lemma 3.1 will provide a lower bound on the optimal objective value of (3.5) and Lemma 3.2 will use it to claim the desired bound on $\|\Delta \mathbf{z}\|_1$.

Lemma 3.1. *Suppose α and β in* (3.5) *are given by*

$$\alpha = \begin{cases} 1/3, & \text{if } \gamma = 1 \\ 1/2, & \text{if } \gamma > 1 \end{cases} \quad \beta = \begin{cases} 4\sqrt[3]{4}, & \text{if } \gamma = 1 \\ 4\sqrt{\gamma}/\sqrt{\gamma - 1}, & \text{if } \gamma > 1 \end{cases} \tag{3.7}$$

Then, for any $\gamma \geq 1$ and for any $\rho > 0$, the optimal objective value of the problem (3.5) *grows linearly with probability* 1:

$$P\left\{ \frac{4\gamma}{\rho} = \lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{4\|\mathbf{v}\|^2}{n\rho} < \lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{f_0(\rho, \alpha, \beta)}{n} \right\} = 1. \tag{3.8}$$

The proof of Lemma 3.1 is based on two technical tools: duality theory and random matrix theory. In particular, the lower bound on the primal objective $f_0(\rho, \alpha, \beta)$ is derived using Lagrangian duality:

$$f_d \leq f_0(\rho, \alpha, \beta), \tag{3.9}$$

where f_d is the dual objective function evaluated at any dual feasible point. With a fixed ρ, we consider the class of *randomly generated* semidefinite programs (3.5) with increasing size ($n, m \to \infty$). It turns out we can construct a sequence of *deterministic* dual feasible points for these random SDPs, using only 3 scalar parameters. We use random matrix theory to compute the asymptotic dual objective value achieved by this sequence of dual feasible points. The asymptotic lower bound on $f_0(\rho,\alpha,\beta)$ follows from (3.9). The details of the proof are provided in the appendix.

In Lemma 3.2 we relate the optimal solution $\mathbf{X}_{opt}$ of (2.3) with the lower bound derived in Lemma 3.1:

$$\mathrm{Trace}(\mathbf{L}_0\mathbf{X}_{opt}) \le 4\|\mathbf{v}\|^2/\rho < f_0(\rho,\alpha,\beta).$$

This inequality implies that the optimal solution $\mathbf{X}_{opt}$ of (2.3) can not be feasible for problem (3.5). Since $\mathbf{X}_{opt}$ satisfies the first two constraints in (3.5) it should violate the last constraint $\mathrm{Trace}(\mathbf{R}_0\mathbf{X}_{opt}) \ge \beta n\rho^{-\alpha}$. Notice that for $\mathbf{X}_{opt}$ we have

$$\mathrm{Trace}(\mathbf{R}_0\mathbf{X}_{opt}) = 2\sum_{i=1}^{n}(1-z_i) = 2\,\|\Delta\mathbf{z}\|_1,\ |z_i| \le 1. \tag{3.10}$$

This leads us to the desired bound $\|\Delta\mathbf{z}\|_1 \le \beta n\rho^{-\alpha}/2$. The details of the proof of Lemma 3.2 are relegated to the appendix.

LEMMA 3.2. *Consider a general class of linear models specified by* (1.2) *and* (3.1). *Suppose vector* $\mathbf{s}=\mathbf{e}$ *is the optimal solution, let* $\mathbf{z}$ *be specified in* (2.4), *and let* α *and* β *be defined in* (3.7). *Then, for any* $\gamma \ge 1$ *and for any* $\rho > 0$, $\mathbf{z}$ *satisfies*

$$P\left\{\ell(\rho,\gamma) \le \frac{\beta}{2\rho^\alpha}\right\} = P\left\{\lim_{\substack{m,n\to\infty\\ m/n\to\gamma}} \frac{\|\mathbf{e}-\mathbf{z}\|_1}{n} \le \frac{\beta}{2\rho^\alpha}\right\} = 1, \tag{3.11}$$

where $\ell(\rho,\gamma)$ *is defined by* (3.3).

Lemma 3.2 provides an estimate of the quality of SDP relaxation (2.3) in terms of the distance $\|\Delta\mathbf{z}\|_1$. Recall that for the binary feasible set, the Bernoulli distribution (2.5) ensures that z_i is equal to the expected value of the i-th bit. Thus, we can view the result of Lemma 3.2 as a bound on the average deviation $|\Delta z_i| = |1-z_i|$ of the mean z_i from the optimal solution $s_i = 1$.

To complete the analysis of the approximation ratio of the SDR algorithm we need to complement the bound (3.11) with an analysis of the randomized rounding procedure. At the core of this analysis is the following chain of inequalities which hold in probability for some constant $c(\rho,\gamma)$ when $n, m \to \infty$ such that $m/n \to \gamma$:

$$\frac{f_{sdr}}{m} \le \frac{E_D\{f_{sdr}\}}{m} \le \frac{f_{sdp}}{m} + \frac{(1+\sqrt{\gamma})^2\beta}{\gamma\rho^{\alpha-1}} \le c(\rho,\gamma)\frac{f_{ils}}{m},$$

where $E_D\{\cdot\}$ is expectation with respect to the randomized rounding procedure, f_{sdp} is defined in (2.3), and f_{sdr}, f_{ils} are given in (3.2). The first inequality holds for sufficiently large D. For the second inequality, we evaluate the expectation $E_D\{f_{sdr}\}$ and show that it can be expressed in terms of f_{sdp}, $\lambda_{\max}(\mathbf{H}^\dagger\mathbf{H})$, and $\|\Delta\mathbf{z}\|_1$. The bound in (3.11) and almost sure convergence of $\lambda_{\max}(\mathbf{H}^\dagger\mathbf{H})$ to $(1+\sqrt{\gamma})^2$ [24] allow us to claim the second inequality. The third inequality follows from the relaxation property $f_{sdp} \le f_{ils}$ and the lower bound [8]:

$$\lim_{\substack{m,n\to\infty\\ m/n\to\gamma}} P\left\{\frac{1}{2} \le \frac{f_{ils}}{m}\right\} = 1. \tag{3.12}$$

The result is summarized in the following theorem whose proof can be found in the appendix.

THEOREM 3.3. *Consider the general class of linear models specified by* (1.2) *and* (3.1)*. For any positive constants r, κ and $\gamma \geq 1$, the SDR algorithm provides a constant factor $c(\rho,\gamma)$ approximation for the ILS problem in probability:*

$$\lim_{\substack{m,n,D\to\infty \\ m/n\to\gamma, D=rn^{1+\kappa}}} P\left\{\frac{f_{sdr}}{f_{ils}} \leq c(\rho,\gamma)\right\} = 1,$$

where

$$c(\rho,\gamma) = 1 + \frac{2(1+\sqrt{\gamma})^2\beta}{\gamma\rho^{\alpha-1}}, \tag{3.13}$$

and α, β are given in (3.7)*.*

To help appreciate the bound (3.13), let us consider a naive random algorithm that simply generates a random vector $\mathbf{s}_{rnd} \in \mathcal{S}^n$ (independent of $\mathbf{H}$, $\mathbf{v}$) as an estimate of the solution. This algorithm achieves an objective value of $f_{rnd} \triangleq \|\mathbf{y} - \sqrt{\rho}\mathbf{H}\mathbf{s}_{rnd}\|^2$. Let $m, n \to \infty$ such that $m/n \to \gamma$ for some γ ($\gamma \geq 1$). It is straightforward to derive

$$\lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{f_{rnd}}{m} = (2\rho+1).$$

Using the lower bound (3.12), we conclude that such a random algorithm provides the approximation ratio $c(\rho,\gamma) = 4\rho+2$ in probability. This bound is worse than the SDP bound (3.13).

Theorem 3.3 implies that the approximation ratio of the SDR algorithm remains bounded for large problem sizes.

3.2. Large-ρ region for a large size ILS problem. So far we have fixed ρ and studied the asymptotic performance of the SDR algorithm as $n, m \to \infty$. In our next result, we evaluate the performance of the SDR algorithm by letting ρ grow with problem size. Define the following random variable:

$$\mathcal{A}(n,m,\rho) \triangleq \frac{\|\mathbf{H}^\dagger\mathbf{v}\|_2}{\sqrt{\rho}\,\lambda_{min}(\mathbf{H}^\dagger\mathbf{H})}. \tag{3.14}$$

It is shown in [8,9] that the semidefinite relaxation (2.3) is tight (rank-1 solution matrix) when $\mathcal{A}(n,m,\rho) < 1$, and in that case the ILS problem is solvable in polynomial time. The result below states the rate at which ρ needs to grow with $n, m \to \infty$ to ensure that the sufficient condition $\mathcal{A}(n,m,\rho) < 1$ holds with probability 1. The proof uses the limiting distribution of the minimum eigenvalue of a random matrix and is relegated to the appendix.

THEOREM 3.4. *Consider the linear model* (1.2) *and* (3.1) *specified with real or complex Rayleigh $\mathbf{H}$, i.e. $H_{ik} \sim \mathcal{N}(0,1/n)$, or $H_{ik} \sim \mathcal{CN}(0,1/n)$. Let*

$$\rho = \begin{cases} \Omega(m), & \textit{when } \gamma > 1, \\ \Omega(m^5 m^\epsilon),\ \forall\, \epsilon > 0, & \textit{when } \gamma = 1. \end{cases}$$

Then, the semidefinite program (2.3) *has a rank-*1 *solution:*

$$P\left\{\lim_{\substack{m,n\to\infty \\ m/n\to\gamma\geq 1}} \mathcal{A}(n,m,\rho) < 1\right\} = 1. \tag{3.15}$$

4. Performance analysis of the PSK algorithm. In this section, we present a theoretical analysis of the algorithm based on the low-rank semidefinite relaxation (2.9).

The relaxed problem (2.9) has a non-convex feasible set and is NP-hard [15]. In what follows, we analyze the structure of local minimizers for the relaxed problem (2.9). Problem (2.9) in the homogeneous form is given by

$$\hat{\mathbf{z}} = \arg \min_{\mathbf{x}\in\mathbb{U}^{n+1}} \text{Trace}(\mathbf{Q}\mathbf{x}\mathbf{x}^\dagger)$$

where $\mathbf{x}, \mathbf{Q}$ are defined in (2.2). Note that a vector $\mathbf{x}(\boldsymbol{\phi})$ in $\mathbb{U}^{n+1}$ can be parameterized by an angular vector $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{n+1})^\dagger \in \mathbb{R}^{n+1}$, i.e., $x_i = e^{j\,\phi_i}, \forall\, i$. Thus, we have

$$\left(\mathbf{x}(\boldsymbol{\phi})\mathbf{x}(\boldsymbol{\phi})^\dagger\right)_{ik} = e^{j\phi_i}(e^{j\phi_k})^* = e^{j(\phi_i-\phi_k)}$$

Hence, the original minimization problem (2.9) is equivalent to the minimization of

$$f(\boldsymbol{\phi}) \triangleq \text{Trace}\left(\mathbf{Q}\,\mathbf{x}(\boldsymbol{\phi})\mathbf{x}(\boldsymbol{\phi})^\dagger\right), \ \forall \boldsymbol{\phi} \in \mathbb{R}^{n+1}. \tag{4.1}$$

Let $f_{\max} = \max_{\boldsymbol{\phi}} f(\boldsymbol{\phi})$ and $f_{\min} = \min_{\boldsymbol{\phi}} f(\boldsymbol{\phi})$ denote the global maximum and minimum values of the objective function (4.1). A vector $\hat{\boldsymbol{\phi}}$ is called a δ-minimizer, $\delta \in [0,1]$, if

$$\frac{f(\hat{\boldsymbol{\phi}}) - f_{\min}}{f_{\max} - f_{\min}} \leq \delta.$$

When the above inequality holds, the function value $f(\hat{\boldsymbol{\phi}})$ is called a δ-minimum.

THEOREM 4.1. *Every local minimizer $\hat{\boldsymbol{\phi}}$ of (4.1) is a $\frac{1}{2}$-minimizer and $f(\hat{\boldsymbol{\phi}})$ is a $\frac{1}{2}$-minimum of (4.1). If in addition $\mathbf{x}(\hat{\boldsymbol{\phi}})$ is feasible for (2.1), then $\mathbf{x}(\hat{\boldsymbol{\phi}})$ is a $\frac{1}{2}$-maximizer for the ILS problem.*

To prove Theorem 4.1 we need the following lemma:

LEMMA 4.2. *For any complex vectors $\mathbf{z}_1, \mathbf{z}_2$, a Hermitian matrix $\mathbf{W}$ and $\mathbf{u} \in \mathbb{U}^{n+1}$, we have:*

$$(\mathbf{z}_2 \circ \mathbf{z}_1)^\dagger \mathbf{W} (\mathbf{z}_2 \circ \mathbf{z}_1) = \mathbf{z}_2^\dagger (\mathbf{W} \circ (\mathbf{z}_1\mathbf{z}_1^\dagger)^*)\mathbf{z}_2 \tag{4.2}$$

and

$$\mathbf{z}_1^\dagger \mathbf{W} \mathbf{z}_1 = \mathbf{u}^\dagger \text{Diag}\left(\left[\mathbf{W} \circ (\mathbf{z}_1\mathbf{z}_1^\dagger)^*\right]\mathbf{e}\right)\mathbf{u} = \mathbf{u}^\dagger \text{Diag}\left(\left[\mathbf{W}^* \circ (\mathbf{z}_1\mathbf{z}_1^\dagger)\right]\mathbf{e}\right)\mathbf{u} \tag{4.3}$$

The proof of this lemma is provided in the appendix.

Proof of Theorem 4.1: The Hessian matrix of (4.1) can be expressed in the form:

$$\nabla^2 f(\boldsymbol{\phi}) = \mathbf{Q} \circ (\mathbf{x}\mathbf{x}^\dagger)^* + \mathbf{Q}^* \circ (\mathbf{x}\mathbf{x}^\dagger) - \text{Diag}\left(\left[\mathbf{Q} \circ (\mathbf{x}\mathbf{x}^\dagger)^* + \mathbf{Q}^* \circ (\mathbf{x}\mathbf{x}^\dagger)\right]\mathbf{e}\right). \tag{4.4}$$

At a local minimum $\hat{\mathbf{x}} = \mathbf{x}(\hat{\boldsymbol{\phi}})$, the Hessian matrix must be positive semidefinite, so for any vector $\mathbf{u} \in \mathbb{U}^{n+1}$ we have:

$$\begin{aligned} &\mathbf{u}^\dagger \left(\mathbf{Q} \circ (\hat{\mathbf{x}}\hat{\mathbf{x}}^\dagger)^* + \mathbf{Q}^* \circ (\hat{\mathbf{x}}\hat{\mathbf{x}}^\dagger)\right)\mathbf{u} \\ &- \mathbf{u}^\dagger \left(\text{Diag}\left(\left[\mathbf{Q} \circ (\hat{\mathbf{x}}\hat{\mathbf{x}}^\dagger)^* + \mathbf{Q}^* \circ (\hat{\mathbf{x}}\hat{\mathbf{x}}^\dagger)\right]\mathbf{e}\right)\right)\mathbf{u} \geq 0. \end{aligned} \tag{4.5}$$

The first term of (4.5) can be written according to (4.2) as

$$\mathbf{u}^\dagger \left(\mathbf{Q} \circ (\hat{\mathbf{x}}\hat{\mathbf{x}}^\dagger)^* + \mathbf{Q}^* \circ (\hat{\mathbf{x}}\hat{\mathbf{x}}^\dagger)\right) \mathbf{u} = (\hat{\mathbf{x}} \circ \mathbf{u})^\dagger \mathbf{Q} (\hat{\mathbf{x}} \circ \mathbf{u}) + (\hat{\mathbf{x}} \circ \mathbf{u}^*)^\dagger \mathbf{Q} (\hat{\mathbf{x}} \circ \mathbf{u}^*).$$

By (4.3), the second term of (4.5) is equal to $2\hat{\mathbf{x}}^\dagger \mathbf{Q} \hat{\mathbf{x}}$. Thus, for any vector $\mathbf{u} \in \mathbb{U}^{m+1}$, we obtain

$$(\hat{\mathbf{x}} \circ \mathbf{u})^\dagger \mathbf{Q} (\hat{\mathbf{x}} \circ \mathbf{u}) + (\hat{\mathbf{x}} \circ \mathbf{u}^*)^\dagger \mathbf{Q} (\hat{\mathbf{x}} \circ \mathbf{u}^*) - 2\hat{\mathbf{x}}^\dagger \mathbf{Q} \hat{\mathbf{x}} \geq 0.$$

Define vectors $\mathbf{v}_1 = \hat{\mathbf{x}} \circ \mathbf{u} \in \mathbb{U}^{n+1}$, and $\mathbf{v}_2 = \hat{\mathbf{x}} \circ \mathbf{u}^* \in \mathbb{U}^{n+1}$. As $\mathbf{u}$ runs through all possible elements in $\mathbb{U}^{n+1}$, so does vector $\mathbf{v}_1$ ($\mathbf{v}_2$ is determined by $\mathbf{v}_1$ and $\mathbf{u}$). Thus, inequality (4.5) can be written as

$$\mathbf{v}_1^\dagger \mathbf{W} \mathbf{v}_1 + \mathbf{v}_2^\dagger \mathbf{W} \mathbf{v}_2 - 2\hat{\mathbf{x}}^\dagger \mathbf{W} \hat{\mathbf{x}} \geq 0, \ \forall\, \mathbf{v}_1 \in \mathbb{U}^{n+1}. \tag{4.6}$$

Specializing the above inequality (4.6) at the global minimum $\hat{\mathbf{v}}_1 = \arg\min_{\mathbb{U}^{n+1}} \mathbf{v}_1^\dagger \mathbf{W} \mathbf{v}_1$ and correspondingly $\hat{\mathbf{v}}_2 = \hat{\mathbf{v}}_1 \circ \hat{\mathbf{x}}^* \circ \hat{\mathbf{x}}^*$, we obtain

$$f(\mathbf{x}(\hat{\boldsymbol{\phi}})) = \hat{\mathbf{x}}^\dagger \mathbf{Q} \hat{\mathbf{x}} \leq \frac{1}{2}\left(\hat{\mathbf{v}}_1^\dagger \mathbf{Q} \hat{\mathbf{v}}_1 + \hat{\mathbf{v}}_2^\dagger \mathbf{Q} \hat{\mathbf{v}}_2\right) \leq \frac{1}{2}\left(f_{\min} + f_{\max}\right), \tag{4.7}$$

which implies that the following inequality holds:

$$\frac{f(\mathbf{x}(\hat{\boldsymbol{\phi}})) - f_{\min}}{f_{\max} - f_{\min}} \leq \frac{1}{2}.$$

This proves the theorem. □

For binary feasible set and a real symmetric matrix $\mathbf{Q}$, it is possible to strengthen the claim of Theorem 4.1: any local feasible minimizer is also a global minimizer [1]. This can be easily seen since for the real case minimizer $\hat{\mathbf{x}}$ and quadratic form $\hat{\mathbf{v}}_1^\dagger \mathbf{Q} \hat{\mathbf{v}}_1$ are real, and $\hat{\mathbf{v}}_1^* = \hat{\mathbf{v}}_2$. Thus, we have

$$\hat{\mathbf{v}}_2^\dagger \mathbf{Q} \hat{\mathbf{v}}_2 = (\hat{\mathbf{v}}_1^*)^\dagger \mathbf{Q} \hat{\mathbf{v}}_1^* = (\hat{\mathbf{v}}_1^*)^\dagger \mathbf{Q}^* \hat{\mathbf{v}}_1^* = \hat{\mathbf{v}}_1^\dagger \mathbf{Q} \hat{\mathbf{v}}_1.$$

Therefore, inequality (4.7) in the proof of Theorem 4.1 yields

$$f_{\min} \leq f(\mathbf{x}(\hat{\boldsymbol{\phi}})) \leq \hat{\mathbf{v}}_1^\dagger \mathbf{Q} \hat{\mathbf{v}}_1 = f_{\min},$$

which proves that $\hat{\boldsymbol{\phi}}$ is a global minimizer.

We caution that the presented analysis does not prove 1/2-optimality of the PSK algorithm. Our analysis is based on the assumption that the global minimizer belongs to the feasible set of the ILS problem. It is not clear to what extent this assumption can be relaxed. Further investigation is required to strengthen the analysis. Numerical simulations [13] indicate that the assumption holds with high probability. Based on this, the analysis presented in this paper provides a theoretical justification for the PSK algorithm.

5. Discussions.

The ILS problem is a difficult and yet fundamental problem for practical applications, such as maximum-likelihood detection in digital communication. This paper presents a theoretical analysis of two efficient suboptimal approaches to the ILS problem (1.1) based on semidefinite relaxation. For a general class of random linear models, we prove that the SDR algorithm provides a constant factor approximation ratio. For the PSK algorithm we show that every local minimizer of the low-rank semidefinite relaxation achieves at least a half of the minimum relative objective value, and for the binary case even yields an exact ILS solution.

The results in this paper indicate that the ILS problem can be efficiently approximated in polynomial time using semidefinite relaxation. Our analysis provides useful theoretical support for the good performance/complexity tradeoff reportedly achieved by the semidefinite relaxation based algorithms in practical applications [16, 17, 19, 20, 22, 31]. From a technical standpoint, the theoretical techniques developed in this work are likely to be useful in the probabilistic analysis of semidefinite relaxations for other NP-hard optimization problems, especially when the corresponding worst-case approximation quality is unbounded.

In a future paper [13], we will present an efficient implementation of the SDR algorithm based on an optimized dual-scaling interior-point method for the relaxed semidefinite program (2.3). We will also present a polynomial time implementation of the PSK algorithm using a coordinate descent strategy on the homotopy of the feasible region of (2.7). Parameter-sensitive improvements for both algorithm can be achieved by a dimension reduction strategy and warm start techniques. Extensive numerical simulations will show that both implementations can achieve near-optimal performance at a substantially reduced complexity when compared with either a general purpose interior point method implementation or the best available implementation of a sphere decoding algorithm.

Appendix A. Proof of Lemma 3.1.

The proof relies on two analytical tools: duality theory in optimization and random matrix theory [28]. The former allows us to obtain a lower bound for the primal objective value $f_0(\rho, \alpha, \beta)$ by the dual objective value at a dual feasible point, while the latter helps to find a good dual feasible point for a random semidefinite program (3.5).

Let us review two results from random matrix theory that will be needed in the ensuing analysis. If $n, m \to \infty$ such that $n/m \to \gamma$ for a random matrix $\mathbf{H}$ with i.i.d., zero-mean, variance $1/n$ entries, the empirical distribution of eigenvalues of $\mathbf{H}^\dagger \mathbf{H}$ converges almost surely [24] to a non-random distribution $F_{\mathbf{H}^\dagger \mathbf{H}}(\lambda)$ with Marcenko-Pastur density [18]:

$$g_\gamma(\lambda) = (1-\gamma)^+ \, \delta(\lambda) + \frac{\sqrt{(\lambda - a)^+ (b - \lambda)^+}}{2\pi\lambda}, \tag{A.1}$$

where $(x)^+ \triangleq \max(0, x)$, $a \triangleq (1 - \sqrt{\gamma})^2$ and $b \triangleq (1 + \sqrt{\gamma})^2$. For matrices of the form $\mathbf{M} = \mathbf{H}^\dagger \mathbf{T} \mathbf{H}$, where $\mathbf{T}$ is a random diagonal matrix independent of $\mathbf{H}$, the asymptotic empirical distribution of eigenvalues $F_{\mathbf{M}}(\lambda)$ can be described in terms of its Stieltjes transform [24], defined by:

$$S_{\mathbf{M}}(z) \triangleq \int \frac{dF_{\mathbf{M}}(\lambda)}{\lambda - z}. \tag{A.2}$$

It is shown in [24] that the Stieltjes transform of the asymptotic empirical eigenvalue distribution of $\mathbf{M} = \mathbf{H}^\dagger \mathbf{T} \mathbf{H}$ is a unique solution to the following functional equation:

$$S_{\mathbf{M}}(z) = \left(-z + \gamma \int \frac{t \, dF_{\mathbf{T}}(t)}{1 + t \, S_{\mathbf{M}}(z)} \right)^{-1}, \tag{A.3}$$

where $F_{\mathbf{T}}(t)$ is the asymptotic distribution of the diagonal entries of matrix $\mathbf{T}$.

For the primal semidefinite program (3.5) the dual problem with dual variables μ and $\mathbf{G}$ can be written as:

$$\begin{aligned} f_0(\rho,\alpha,\beta) \triangleq \max \;\; & \mathrm{Trace}(\mathbf{G}) + \mu\frac{\beta n}{\rho^{\alpha}} \\ \text{s.t.} \;\; & \mathbf{L}_0 - \mathbf{G} - \mu\mathbf{R}_0 \succeq 0, \\ & \mu \geq 0, \\ & \mathbf{G} \text{ is diagonal,} \end{aligned} \tag{A.4}$$

where we used the strong duality of semidefinite programs to claim the dual optimal objective value is given by $f_0(\rho,\alpha,\beta)$. We consider a sequence (as $n\to\infty$) of diagonal matrices $\mathbf{G}_{n+1}$ of dimension $n+1$ parameterized by two scalars τ and χ (to be chosen later to ensure dual feasibility):

$$\mathbf{G}_{n+1} = \begin{bmatrix} -\tau\mathbf{I}_n & \mathbf{0} \\ \mathbf{0}^{\dagger} & \chi n \end{bmatrix}. \tag{A.5}$$

The dual objective value of (A.4) at a point $\{\mathbf{G}_{n+1},\mu\}$ is defined as

$$f_0(\mathbf{G}_{n+1},\mu) \triangleq \mathrm{Trace}(\mathbf{G}_{n+1}) + \mu\frac{\beta n}{\rho^{\alpha}} = n(\chi-\tau) + \mu\frac{\beta n}{\rho^{\alpha}} \tag{A.6}$$

and serves as a lower bound on $f_0(\rho,\alpha,\beta)$ whenever $\{\mathbf{G}_{n+1},\mu\}$ is dual feasible. To ensure dual feasibility of $\{\mathbf{G}_{n+1},\mu\}$, the scalars $\{\tau,\chi\}$ must be chosen to satisfy $\mathbf{L}_0 - \mathbf{G}_{n+1} - \mu\mathbf{R}_0 \succeq 0$. By the definition (3.6), we obtain

$$\begin{bmatrix} \mathbf{H}^{\dagger}\mathbf{H} + (\tau-\mu)\mathbf{I} & -(\mathbf{H}^{\dagger}\mathbf{H} - \mu\mathbf{I})\mathbf{e} \\ -\mathbf{e}^{\dagger}(\mathbf{H}^{\dagger}\mathbf{H} - \mu\mathbf{I}) & \|\mathbf{He}\|^2 - (\chi+\mu)n \end{bmatrix} \succeq 0.$$

By Schur complement, this is equivalent to the following two conditions:

$$\begin{cases} \mathbf{H}^{\dagger}\mathbf{H} + (\tau-\mu)\mathbf{I} \succ 0, \\ \tau - \chi \geq \tau^2 \frac{1}{n}\mathbf{e}^{\dagger}\left(\mathbf{H}^{\dagger}\mathbf{H} + (\tau-\mu)\mathbf{I}\right)^{-1}\mathbf{e}. \end{cases} \tag{A.7}$$

Since $\mathbf{H}^{\dagger}\mathbf{H}$ is positive semidefinite, any choice of $\{\tau,\mu\}$ such that $\tau>\mu$ will satisfy the first constraint. For a random matrix $\mathbf{H}$ the right hand side of the second constraint in (A.7) is a random variable. Our goal is to select deterministic values for $\{\tau,\chi,\mu\}$ such that the second constraint is satisfied in probability as $n,m\to\infty$. This combination of τ and χ will specify the good dual feasible matrix $\mathbf{G}_{n+1}$ in (A.5) that will provide the lower bound claimed in Lemma 3.1. Notice that the limit $(n,m\to\infty)$ in the right hand side of (A.7) depends only on the asymptotic eigenvalue distribution of $\mathbf{H}^{\dagger}\mathbf{H}$. We can evaluate this limit using the Stieltjes transform (A.2):

$$\lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{1}{n}\mathbf{e}^{\dagger}\left(\mathbf{H}^{\dagger}\mathbf{H} + (\tau-\mu)\mathbf{I}\right)^{-1}\mathbf{e} = \int_0^{+\infty} \frac{dF_{\mathbf{H}^{\dagger}\mathbf{H}}(\lambda)}{\lambda - (\mu-\tau)} = S_{\mathbf{H}^{\dagger}\mathbf{H}}(\mu-\tau). \tag{A.8}$$

We calculate the Stieltjes transform $S_{\mathbf{H}^{\dagger}\mathbf{H}}(\mu-\tau)$ using functional equation (A.3):

$$S_{\mathbf{H}^{\dagger}\mathbf{H}}(z) = \left(-z + \frac{\gamma}{1+S_{\mathbf{H}^{\dagger}\mathbf{H}}(z)}\right)^{-1} = \frac{\gamma - 1 - z \pm \sqrt{(z+1-\gamma)^2 - 4z}}{2z},$$

where the second step follows from solving the first implicit equation for $S_{\mathbf{H}^\dagger\mathbf{H}}(z)$. Substitute $z = \mu - \tau$ to obtain:

$$S_{\mathbf{H}^\dagger\mathbf{H}}(\mu-\tau) = -\frac{1}{2} - \frac{\gamma-1}{2(\tau-\mu)} \pm \frac{1}{2}\sqrt{\left(1+\frac{\gamma-1}{\tau-\mu}\right)^2 + \frac{4}{\tau-\mu}}.$$

Since integral (A.8) is non-negative and dual feasibility requires $\tau > \mu$, only the plus sign gives a valid expression in the integral above. Substituting the limit into (A.7) we conclude that there $\exists\{n_0, m_0\} > 0$ such that $\forall n \geq n_0, \forall m \geq m_0$ dual feasibility of $\{\mathbf{G}_{n+1}, \mu\}$ is achieved with probability 1 if we select $\{\tau, \chi, \mu\}$ such that

$$\begin{aligned} &\tau > \mu \geq 0, \\ &\chi - \tau \leq \frac{\tau^2}{2}\left(1 + \frac{\gamma-1}{\tau-\mu} - \sqrt{\left(1+\frac{\gamma-1}{\tau-\mu}\right)^2 + \frac{4}{\tau-\mu}}\right). \end{aligned} \tag{A.9}$$

For χ satisfying the above condition with equality, the asymptotic dual objective value (A.6) grows linearly:

$$\begin{aligned} \lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{f_0(\rho,\alpha,\beta)}{n} &\geq \lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{f_0(\mathbf{G}_{n+1},\mu)}{n} \\ &= \chi - \tau + \frac{\mu\beta}{\rho^\alpha} = \frac{\mu\beta}{\rho^\alpha} + \frac{\tau^2}{2}\left(1 + \zeta - \sqrt{(1+\zeta)^2 + \frac{4}{\tau-\mu}}\right), \end{aligned}$$

where $\zeta \triangleq \frac{\gamma-1}{\tau-\mu}$. Now we specify the lower bound with constants (3.7). When $\gamma = 1$, we select $\{\alpha, \beta, \tau, \mu\}$ such that

$$\beta = 4\sqrt[3]{4},\ \alpha = \frac{2}{3},\ \tau = \frac{1}{16}\frac{\beta^2}{\rho^\alpha},\ \mu = \frac{3}{64}\frac{\beta^2}{\rho^\alpha} + \epsilon,\ \forall \epsilon > 0.$$

When $\gamma > 1$, we select $\{\alpha, \beta, \tau, \mu\}$ such that

$$\beta = \frac{4\sqrt{\gamma}}{\sqrt{\gamma-1}},\ \alpha = \frac{1}{2},\ \tau = \frac{\beta(\gamma-1)}{2\rho^\alpha},\ \mu = \tau + \epsilon,\ \forall \epsilon > 0.$$

For both cases $\tau > \mu$ satisfies (A.9), and for any fixed ρ the lower bound on the objective value is

$$\lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{f_0(\rho,\alpha,\beta)}{n} \geq \frac{4\gamma}{\rho} + \epsilon,\ \forall \epsilon > 0.$$

At the same time, by the strong law of large numbers

$$P\left\{\lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{4\|\mathbf{v}\|^2}{n\rho} = \frac{4\gamma}{\rho}\right\} = 1.$$

This establishes the desired lower bound with probability 1:

$$P\left\{\lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{4\|\mathbf{v}\|^2}{n\rho} < \lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{f_0(\rho,\alpha,\beta)}{n}\right\} = 1.$$

□

Appendix B. Proof of Lemma 3.2.

The proof follows in part the steps in the proof of Lemma 8.2 in [8]. For large n, m, we consider a randomly generated problem (2.3) whose optimal solution is represented in the block form (2.4):

$$\mathbf{X}_{opt} = \begin{bmatrix} \mathbf{Z} & \mathbf{z} \\ \mathbf{z}^\dagger & 1 \end{bmatrix}.$$

We will prove (3.11) by contradiction. In particular, if the desired lower bound (3.11) does not hold, then $\mathbf{X}_{opt}$ must belong to the feasible set of (3.5) with a non-zero probability. In light of Lemma 3.1, this in turn would imply that the limit of $\text{Trace}(\mathbf{L}_0\mathbf{X}_{opt})$ is strictly greater than the limit of $4\|\mathbf{v}\|^2/(n\rho)$ with a non-zero probability, which is a contradiction.

First we notice that for $\mathbf{R}_0$ defined in (3.6), there holds

$$\text{Trace}(\mathbf{R}_0\mathbf{X}_{opt}) = \text{Trace}\left(\begin{bmatrix} \mathbf{I} & -\mathbf{e} \\ -\mathbf{e}^\dagger & n \end{bmatrix}\begin{bmatrix} \mathbf{Z} & \mathbf{z} \\ \mathbf{z}^\dagger & 1 \end{bmatrix}\right) = 2\sum_{i=1}^{n}(1 - z_i) = 2\|\Delta\mathbf{z}\|_1$$

since $\text{diag}(\mathbf{Z}) = \mathbf{e}$ and $\Delta z_i = 1 - z_i \geq 0,\ i = 1, \ldots, n$.

Suppose that the claim in Lemma 3.2 is not true, i.e. under the assumptions of the lemma we have

$$P\left\{\lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{\|\Delta\mathbf{z}\|_1}{n} > \frac{\beta}{2\rho^\alpha}\right\} > 0, \tag{B.1}$$

where α and β are defined in (3.7). That is, $\exists n_0$ such that $\forall n \geq n_0$ we have $P\{\|\Delta\mathbf{z}\|_1/n > \beta/(2\rho^\alpha)\} > 0$. Therefore, according to (B.1) the probability that matrix $\mathbf{X}_{opt}$ is feasible for (3.5) is not 0:

$$P\left\{\lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{\text{Trace}(\mathbf{R}_0\mathbf{X}_{opt})}{n} > \frac{\beta}{\rho^\alpha}\right\} > 0.$$

Thus, from Lemma 3.1 we conclude that

$$P\left\{\lim_{\substack{m,n\to\infty \\ m/n\to\gamma}} \frac{\text{Trace}(\mathbf{L}_0\mathbf{X}_{opt})}{n} > \frac{4\|\mathbf{v}\|^2}{n\rho}\right\} > 0. \tag{B.2}$$

We will prove that for $\mathbf{X}_{opt}$ this probability should be 0, thus leading to a contradiction. The dual optimization problem for (2.3) can be written as

$$\begin{aligned} f_{sdp} \triangleq \max\ & \text{Trace}(\mathbf{G}) \\ \text{s.t.}\ & \mathbf{Q} - \mathbf{G} \succeq 0, \\ & \mathbf{G} \text{ is diagonal}, \end{aligned}$$

where $\mathbf{G}$ is a dual variable. Matrix $\mathbf{Q} - \mathbf{G}_{opt}$ has at least 1 non-zero eigenvalue with probability 1[1]. Thus, for the optimal $\{\mathbf{X}_{opt}, \mathbf{G}_{opt}\}$ the complementary slackness

[1]In fact, the matrix $\mathbf{Q} - \mathbf{G}_{opt}$ will have non-zero off-diagonal entries with probability 1.

condition $(\mathbf{Q}-\mathbf{G}_{opt})\mathbf{X}_{opt}=\mathbf{0}$ implies that $\mathbf{X}_{opt}$ is rank deficient. Hence, $\mathbf{X}_{opt}$ allows the following factorization:

$$\mathbf{X}_{opt}=\begin{bmatrix}\mathbf{W}^\dagger\\ \mathbf{u}^\dagger\end{bmatrix}[\mathbf{W}\ \ \mathbf{u}]=\begin{bmatrix}\mathbf{W}^\dagger\mathbf{W} & \mathbf{W}^\dagger\mathbf{u}\\ \mathbf{u}^\dagger\mathbf{W} & \mathbf{u}^\dagger\mathbf{u}\end{bmatrix}.$$

Let us express the optimal objective value of (2.3) in terms of $\mathbf{W}$ and $\mathbf{u}$:

$$\text{Trace}(\mathbf{Q}\mathbf{X}_{opt})=\|\sqrt{\rho}\,\mathbf{H}\mathbf{W}^\dagger-\mathbf{y}\mathbf{u}\|_F^2=\|\sqrt{\rho}\,\mathbf{H}(\mathbf{W}^\dagger-\mathbf{e}\mathbf{u}^\dagger)-\mathbf{v}\mathbf{u}\|_F^2,$$

where $\|\cdot\|_F$ is the Frobenius norm. Since $X_{n+1,n+1}=1$, we have $\|\mathbf{u}\|_2=1$, and

$$\sqrt{\text{Trace}(\mathbf{Q}\mathbf{X}_{opt})}\geq\|\sqrt{\rho}\,\mathbf{H}(\mathbf{W}^\dagger-\mathbf{e}\mathbf{u}^\dagger)\|_F-\|\mathbf{v}\|_2. \tag{B.3}$$

Notice that matrix $\mathbf{X}=\mathbf{e}\mathbf{e}^\dagger$ is feasible for (2.3), therefore we have

$$\text{Trace}(\mathbf{Q}\mathbf{X}_{opt})\leq\text{Trace}(\mathbf{Q}\mathbf{e}\mathbf{e}^\dagger)=\|\mathbf{v}\|_2^2.$$

This bound combined with the bound in (B.3) leads to

$$\|\sqrt{\rho}\,\mathbf{H}(\mathbf{W}^\dagger-\mathbf{e}\mathbf{u}^\dagger)\|_F^2\leq4\|\mathbf{v}\|_2^2.$$

On the other hand, the Frobenius norm in the expression above is related to the objective function of (3.5):

$$\begin{aligned}\|\sqrt{\rho}\,\mathbf{H}(\mathbf{W}^\dagger-\mathbf{e}\mathbf{u}^\dagger)\|_F^2 &= \rho\text{Trace}\Big((\mathbf{W}^\dagger-\mathbf{e}\mathbf{u}^\dagger)^\dagger\mathbf{H}^\dagger\mathbf{H}(\mathbf{W}^\dagger-\mathbf{e}\mathbf{u}^\dagger)\Big)\\ &= \rho\text{Trace}\Big(\mathbf{M}\mathbf{H}^\dagger\mathbf{H}\mathbf{M}^\dagger\mathbf{X}_{opt}\Big),\end{aligned}$$

where $\mathbf{M}\triangleq[\mathbf{I}\quad-\mathbf{e}^\dagger]$. Notice that $\mathbf{L}_0=\mathbf{M}\mathbf{H}^\dagger\mathbf{H}\mathbf{M}^\dagger$, thus

$$\frac{\|\sqrt{\rho}\,\mathbf{H}(\mathbf{W}^\dagger-\mathbf{e}\mathbf{u}^\dagger)\|_F^2}{n\rho}=\frac{\text{Trace}\left(\mathbf{L}_0\mathbf{X}_{opt}\right)}{n}\leq\frac{4\|\mathbf{v}\|_2^2}{n\rho}.$$

Taking the limit $n,m\to\infty, m/n\to\gamma$, we obtain a contradiction with (B.2). Thus, the assumption (B.1) is incorrect and the statement of the lemma follows. □

Appendix C. Proof of Theorem 3.3.

In what follows, $P\{\cdot\}$ and $E\{\cdot\}$ will denote probability and expectation evaluated with respect to random $\mathbf{H}$, $\mathbf{v}$, $\mathbf{s}$ as well as random sampling in the randomized rounding procedure. Define $P_D\{\cdot\}$ ($E_D\{\cdot\}$) as probability (expectation) with respect to the distribution of random samples $\bar{\mathbf{x}}_d$ in the randomized rounding procedure given a fixed realization of $\mathbf{H},\mathbf{v}$, and $\mathbf{s}$.

We will show that the following chain of inequalities holds in probability with respect to the distribution of $\mathbf{H},\mathbf{v}$, and $\mathbf{s}$, as $m,n,D\to\infty$ such that $m/n\to\gamma$ and $D=a\,n^{1+\kappa}$ with arbitrary positive constants $a,\kappa>0$:

$$\frac{f_{sdr}}{m}\leq\frac{E_D\{\mathbf{x}_d^\dagger\mathbf{Q}\mathbf{x}_d\}}{m}\leq\frac{f_{sdp}}{m}+\frac{(1+\sqrt{\gamma})^2\beta}{\gamma\rho^{\alpha-1}}\leq c(\rho,\gamma)\frac{f_{ils}}{m}, \tag{C.1}$$

where α and β are specified in (3.7), and

$$c(\rho,\gamma)=1+\frac{2(1+\sqrt{\gamma})^2\beta}{\gamma\rho^{\alpha-1}} \tag{C.2}$$

1) The randomized rounding procedure selects sample $\mathbf{x}_d$ that corresponds to the minimum objective value, that is

$$f_{sdr} = \min_{1 \le d \le D} \mathbf{x}_d^\dagger \mathbf{Q} \mathbf{x}_d.$$

Define $M_n \triangleq E_D\{\mathbf{x}_d^\dagger \mathbf{Q} \mathbf{x}_d\}$. Since vector samples $\mathbf{x}_d, d = 1, \ldots, D$, are independent, for any function $\ell(n)$ we have:

$$\begin{aligned} P_D\{f_{sdr} \ge \ell(n) M_n\} &= P_D\{\mathbf{x}_d^\dagger \mathbf{Q} \mathbf{x}_d \ge \ell(n) M_n, d = 1, \ldots, D\} \\ &= \left(P_D\left\{\mathbf{x}_1^\dagger \mathbf{Q} \mathbf{x}_1 \ge \ell(n) M_n\right\}\right)^D \le \left(\frac{M_n}{\ell(n) M_n}\right)^D = \big(\ell(n)\big)^{-D}, \end{aligned}$$

where we used Markov's inequality. Since the above inequality holds for all realizations of $\mathbf{H}$, $\mathbf{v}$, and $\mathbf{s}$, we obtain

$$P\{f_{sdr} \ge \ell(n) M_n\} \le \big(\ell(n)\big)^{-D}.$$

Select function $\ell(n)$:

$$\ell(n) \triangleq \left(1 - \frac{1}{n^{\kappa/2}}\right)^{-1} \to 1, \text{ as } n \to \infty.$$

For $D = r\, n^{1+\kappa}$ and $\ell(n)$ as selected above, the following series is convergent:

$$\sum_{n=1}^{\infty} \big(\ell(n)\big)^{-D} = \sum_{n=1}^{\infty} \left(1 - \frac{1}{n^{\kappa/2}}\right)^{r n^{1+\kappa}} < \infty.$$

Therefore, we can apply Borel-Cantelli lemma to obtain

$$P\left\{ \limsup_{\substack{m,n,D\to\infty \\ m/n\to\gamma, D=rn^{1+\kappa}}} \left\{ \frac{f_{sdr}}{m} \le \ell(n) \frac{E_D\{\mathbf{x}_d^\dagger \mathbf{Q} \mathbf{x}_d\}}{m} \right\} \right\} = 1$$

which further implies

$$P\left\{ \lim_{\substack{m,n,D\to\infty \\ m/n\to\gamma,\ D=rn^{1+\kappa}}} \frac{f_{sdr}}{m} \le \lim_{\substack{m,n,D\to\infty \\ m/n\to\gamma,\ D=rn^{1+\kappa}}} \frac{E_D\{\mathbf{x}_d^\dagger \mathbf{Q} \mathbf{x}_d\}}{m} \right\} = 1.$$

2) The second inequality in (C.1) is based on a bound on the performance of the randomized rounding procedure:

$$\begin{aligned} E_D\{\mathbf{x}_d^\dagger \mathbf{Q} \mathbf{x}_d\} &= E_D\left\{ \sum_{i,j=1, i\ne j}^{n+1} Q_{i,j} x_i x_j + \sum_{i=1}^{n+1} Q_{i,i} x_i^2 \right\} \\ &= \sum_{i,j=1, i\ne j}^{n+1} Q_{i,j} E_D\{x_i\} E_D\{x_j\} + \sum_{i=1}^{n+1} Q_{i,i} E_D\{x_i^2\} \\ &= \sum_{i,j=1, i\ne j}^{n} Q_{i,j} z_i z_j + 2 \sum_{i=1}^{n} Q_{i,n+1} z_i + \sum_{i=1}^{n+1} Q_{i,i} \\ &= \begin{bmatrix} \mathbf{z}^\dagger & 1 \end{bmatrix} \mathbf{Q} \begin{bmatrix} \mathbf{z} \\ 1 \end{bmatrix} + \sum_{i=1}^{n} Q_{i,i} \left(1 - z_i^2\right). \end{aligned} \tag{C.3}$$

Using the Schur complement of (2.4) we have $\mathbf{z}\mathbf{z}^\dagger \preceq \mathbf{Z}$, thus

$$\begin{bmatrix} \mathbf{z} \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{z}^\dagger & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{z}\mathbf{z}^\dagger & \mathbf{z} \\ \mathbf{z}^\dagger & 1 \end{bmatrix} \preceq \begin{bmatrix} \mathbf{Z} & \mathbf{z} \\ \mathbf{z}^\dagger & 1 \end{bmatrix} = \mathbf{X}_{opt}.$$

Hence, the first term in (C.3) can be bounded by

$$\begin{bmatrix} \mathbf{z}^\dagger & 1 \end{bmatrix} \mathbf{Q} \begin{bmatrix} \mathbf{z} \\ 1 \end{bmatrix} \leq \text{Trace}(\mathbf{Q}\mathbf{X}_{opt}) = f_{sdp}.$$

Since $-1 \leq z_i \leq 1$, and $\Delta z_i \triangleq 1 - z_i \geq 0$, the second term in (C.3) can be bounded as follows:

$$\begin{aligned} \frac{1}{m} \sum_{i=1}^{n} Q_{i,i} \left(1 - z_i^2\right) \leq \frac{\rho \lambda_{\max}(\mathbf{H}^\dagger \mathbf{H})}{m} \sum_{i=1}^{n} \left(1 - z_i^2\right) = \frac{\rho \lambda_{\max}(\mathbf{H}^\dagger \mathbf{H})}{m} \sum_{i=1}^{n} (1 - z_i)(1 + z_i) \\ \leq \frac{2\rho \lambda_{\max}(\mathbf{H}^\dagger \mathbf{H})}{\gamma} \frac{\|\Delta \mathbf{z}\|_1}{n}. \end{aligned}$$

As $n, m \to \infty,\ m/n \to \gamma$, the largest eigenvalue of $\mathbf{H}^\dagger \mathbf{H}$ converges almost surely to $(1 + \sqrt{\gamma})^2$, [3,24]. Lemma 3.2 allows us to bound $\|\Delta \mathbf{z}\|_1$. Therefore,

$$P \left\{ \lim_{\substack{m,n \to \infty \\ m/n \to \gamma}} \frac{E_D\{\mathbf{x}_d^\dagger \mathbf{Q} \mathbf{x}_d\}}{m} \leq \lim_{\substack{m,n \to \infty \\ m/n \to \gamma}} \frac{f_{sdp}}{m} + \frac{(1+\sqrt{\gamma})^2 \beta}{\gamma \rho^{\alpha - 1}} \right\} = 1.$$

3) Since the semidefinite program (2.3) is a relaxation of (2.1), we have $f_{sdp} \leq f_{ils}$. The third inequality in (C.1) follows from a lower bound on f_{ils} in probability [8]:

$$\lim_{\substack{m,n \to \infty \\ m/n \to \gamma}} P \left\{ \frac{1}{2} \leq \frac{f_{ils}}{m} \right\} = 1.$$

The convergence with probability 1 of previous inequalities implies convergence in probability, thus

$$\lim_{\substack{m,n,D \to \infty \\ m/n \to \gamma,\ D = rn^{1+\kappa}}} P \left\{ \frac{f_{sdr}}{m} \leq \left(1 + \frac{2(1+\sqrt{\gamma})^2 \beta}{\gamma \rho^{\alpha - 1}} \right) \frac{f_{ils}}{m} \right\} = 1.$$

This proves the approximation ratio claimed in Theorem 3.3. □

Appendix D. Proof of Theorem 3.4.

Let $\mathbf{H}$ be an $m \times n$ matrix and $\mathbf{v}$ be an $m \times 1$ vector independent of $\mathbf{H}$, both are real or complex, random, such that $E\{v_i\} = 0, E\{|v_i|^2\} = 1$, and $H_{ik} \sim \mathcal{N}(0, 1/n)$ or $H_{ik} \sim \mathcal{CN}(0, 1/n)$. Define a random variable $\mathcal{A}(n, m, \rho)$, parameterized by m, n, and ρ, as

$$\mathcal{A}(n, m, \rho) \triangleq \frac{\|\mathbf{H}^\dagger \mathbf{v}\|_2}{\sqrt{\rho}\, \lambda_{min}(\mathbf{H}^\dagger \mathbf{H})}.$$

The semidefinite relaxation is tight [9] when $\mathcal{A}(n, m, \rho) < 1$:

$$P\left\{\mathcal{A}(n, m, \rho) < 1\right\} = P \left\{ \frac{\frac{1}{m} \mathbf{v}^\dagger \mathbf{H} \mathbf{H}^\dagger \mathbf{v}}{\lambda_{min}^2(\mathbf{H}^\dagger \mathbf{H})} < \frac{\rho}{m} \right\}. \tag{D.1}$$

Consider the enumerator in the above inequality, $\frac{1}{m}\,\mathbf{v}^\dagger\mathbf{H}\mathbf{H}^\dagger\mathbf{v}$, as $n, m \to \infty$. The empirical distribution of eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$ converges almost surely to a deterministic distribution with density $f_\gamma(\lambda)$, c.f. (A.1):

$$f_\gamma(\lambda) \triangleq \left(1 - \frac{1}{\gamma}\right)^+ \delta(\lambda) + \frac{\sqrt{(\lambda-a)^+(b-\lambda)^+}}{2\pi\lambda\gamma}.$$

For independent $\mathbf{H}$ and $\mathbf{v}$ we can use this density to evaluate the limit

$$\lim_{\substack{m,n\to\infty\\ m/n\to\gamma}} \frac{1}{m}\,\mathbf{v}^\dagger\mathbf{H}\mathbf{H}^\dagger\mathbf{v} = \int_0^{+\infty} \lambda\, f_\gamma(\lambda)\delta\lambda = 1.$$

Thus, for any $\gamma > 0$ we have

$$P\left\{ \lim_{\substack{m,n\to\infty\\ m/n\to\gamma}} \frac{1}{m}\,\mathbf{v}^\dagger\mathbf{H}\mathbf{H}^\dagger\mathbf{v} = 1 \right\} = 1.$$

This almost sure convergence holds for both square, $\gamma = 1$, and rectangular, $\gamma > 1$, matrices $\mathbf{H}$, while the behavior of the minimum eigenvalue in (D.1) has to be considered separately.

For Wishart matrices $\mathbf{H}^\dagger\mathbf{H}$ with $m/n \to \gamma > 1$ the minimum eigenvalue of $\mathbf{H}^\dagger\mathbf{H}$ converges almost surely [3, 23] to:

$$P\left\{ \lim_{\substack{m,n\to\infty\\ m/n\to\gamma>1}} \lambda_{min}(\mathbf{H}^\dagger\mathbf{H}) = (\sqrt{\gamma}-1\,)^2 \right\} = 1.$$

Hence, for arbitrary small $\epsilon > 0$ and $\rho = \rho(m)$ such that

$$\rho(m) = \frac{1+\epsilon}{(\sqrt{\gamma}-1\,)^4}\, m = \Omega(m),$$

it follows from (D.1) that event $\{\mathcal{A}(n, m, \rho) < 1\}$ holds almost surely as $n, m \to \infty$ such that $m/n \to \gamma > 1$:

$$1 \geq P\left\{ \lim_{\substack{m,n\to\infty\\ m/n\to\gamma>1}} \mathcal{A}(n, m, \rho) < 1 \right\} \geq P\left\{\epsilon_1 > 0\right\} = 1.$$

For Wishart matrices $\mathbf{H}^\dagger\mathbf{H}$ with $m/n \to \gamma = 1$, we have the following almost sure convergence [3] for any $\epsilon > 0$:

$$P\left\{ \lim_{\substack{m,n\to\infty\\ m/n\to\gamma=1}} \frac{1}{m^{2+\epsilon/2}\lambda_{\min}(\mathbf{H}^\dagger\mathbf{H})} = 0 \right\} = 1.$$

Consider ρ that grows with m as $\rho(m) = m^5 m^\epsilon$, for some small $\epsilon > 0$. Taking limit in (D.1) as $n, m \to \infty$ such that $m/n \to \gamma = 1$, we obtain

$$\begin{aligned} 1 \geq P\left\{ \lim_{\substack{m,n\to\infty\\ m/n\to1}} \mathcal{A}(n, m, \rho) < 1 \right\} &= P\left\{ \lim_{\substack{m,n\to\infty\\ m/n\to1}} \frac{\|\mathbf{H}^\dagger\mathbf{v}\|_2^2}{\lambda^2_{\min}(\mathbf{H}^\dagger\mathbf{H})} < \rho \right\} \\ \geq P\left\{ \lim_{\substack{m,n\to\infty\\ m/n\to1}} \frac{\frac{1}{m}\,\mathbf{v}^\dagger\mathbf{H}\mathbf{H}^\dagger\mathbf{v}}{m^{4+\epsilon}\lambda^2_{\min}(\mathbf{H}^\dagger\mathbf{H})} < \frac{\rho}{m^5 m^\epsilon} \right\} &= P\left\{0 < 1\right\} = 1. \end{aligned}$$

Hence, the choice of $\rho(m) = \Omega(m^5 m^\epsilon)$ with arbitrary small $\epsilon > 0$ ensures that event $\{\mathcal{A}(n,m,\rho) < 1\}$ holds almost surely as $n, m \to \infty$ such that $m/n \to \gamma = 1$. □

Appendix E. Proof of Lemma 4.2.

Equation (4.2) follows immediately if we write both left and right hand sides in the scalar componentwise form. The left hand side of (4.2) is given by:

$$(\mathbf{z}_2 \circ \mathbf{z}_1)^\dagger \mathbf{W} (\mathbf{z}_2 \circ \mathbf{z}_1) = \sum_{i,k} (\mathbf{z}_1^* \circ \mathbf{z}_2^*)_i W_{i,k} (\mathbf{z}_1 \circ \mathbf{z}_2)_k = \sum_{i,k} z_{1i}^* z_{2i}^* W_{i,k} z_{1k} z_{2k}.$$

Writing the right hand side in the scalar form, we have:

$$\mathbf{z}_2^\dagger \left(\mathbf{W} \circ (\mathbf{z}_1 \mathbf{z}_1^\dagger)^* \right) \mathbf{z}_2 = \sum_{i,k} z_{2i}^* \left(\mathbf{W} \circ (\mathbf{z}_1 \mathbf{z}_1^\dagger)^* \right)_{ik} z_{2k} = \sum_{i,k} z_{2i}^* W_{i,k} z_{1i}^* z_{1k} z_{2k}.$$

The sums in (E.1) and in (E.1) are the same, which proves (4.2). To prove equation (4.3) we use the constraint on the entries of $\mathbf{u} \in \mathbb{U}^{n+1}$:

$$\begin{aligned}
\mathbf{u}^\dagger \mathrm{Diag} \left(\left[\mathbf{W} \circ (\mathbf{z}_1 \mathbf{z}_1^\dagger)^* \right] \mathbf{e} \right) \mathbf{u} &= \sum_i |u_i|^2 \left(\left[\mathbf{W} \circ (\mathbf{z}_1 \mathbf{z}_1^\dagger)^* \right] \mathbf{e} \right)_i \\
&= \sum_i \left(\left[\mathbf{W} \circ (\mathbf{z}_1 \mathbf{z}_1^\dagger)^* \right] \mathbf{e} \right)_i = \mathbf{e}^\dagger \left(\mathbf{W} \circ (\mathbf{z}_1 \mathbf{z}_1^\dagger)^* \right) \mathbf{e} \quad \text{(E.1)} \\
&= \mathbf{z}_1^\dagger \mathbf{W} \mathbf{z}_1,
\end{aligned}$$

where we applied (4.2) in the last step. Notice that (E.1) is a real number for a Hermitian matrix $\mathbf{W}$. Applying complex conjugation to (E.1) we obtain the second equality in (4.3):

$$\begin{aligned}
\mathbf{u}^\dagger \mathrm{Diag} \left(\left[\mathbf{W} \circ (\mathbf{z}_1 \mathbf{z}_1^\dagger)^* \right] \mathbf{e} \right) \mathbf{u} = \mathbf{e}^\dagger \left(\mathbf{W} \circ (\mathbf{z}_1 \mathbf{z}_1^\dagger)^* \right) \mathbf{e} &= \mathbf{e}^\dagger \left(\mathbf{W}^* \circ (\mathbf{z}_1 \mathbf{z}_1^\dagger) \right) \mathbf{e} \\
&= \mathbf{u}^\dagger \mathrm{Diag} \left(\left[\mathbf{W}^* \circ (\mathbf{z}_1 \mathbf{z}_1^\dagger) \right] \mathbf{e} \right) \mathbf{u}.
\end{aligned}$$

□

REFERENCES

[1] S. Burer, R.D.C. Monteiro, and Y. Zhang, Rank-two relaxation heuristics for Max-Cut and other binary quadratic programs, SIAM J. Optim., 12 (2001), no. 2, pp. 503 – 521.
[2] M.O. Damen, H. El Gamal, and G. Caire, On maximum-likelihood detection and the search for the closest lattice point, IEEE Trans. Inform. Theory, 49 (2003), no. 10, pp. 2389–2402.
[3] A. Edelman, Eigenvalues and condition numbers of random matrices, Ph.D. Thesis, Massachusetts Institute of Technology, Department of Mathematics, 1989.
[4] U. Fincke and M. Pohst, Improved methods for calculating vectors of short length in a lattice, including a complexity analysis, Math. Comp., 44 (1985), pp. 463 – 471.
[5] M.X. Goemans and D.P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problem using semi-definite programming, J. ACM, 42 (1995), pp. 1115 – 1145.
[6] G.B. Giannakis, Z. Liu, X. Ma, and S. Zhou, Space-time coding for broadband wireless communications, Wiley-Intersci., 1-st ed., 2003.
[7] D. Guo and S. Verdu, Randomly spread CDMA: asymptotics via statistical physics, IEEE Trans. Inform. Theory, 51 (2005), no. 6, pp. 1983 – 2010.
[8] J. Jalden, Detection for multiple input multiple output channels, Ph.D. Thesis, KTH, School of Electrical Engineering, Sweden, 2006.
[9] J. Jalden, C. Martin, and B. Ottersten, Semidefinite programming for detection in linear systems – optimality conditions and space-time decoding, Proc. ICASSP '03, 4 (2003), pp. IV-9 – IV-12.

[10] J. Jalden and B. Ottersten, An exponential lower bound on the expected complexity of sphere decoding, Proc. ICASSP '04, 4 (2004), pp. IV-393 – IV-396.

[11] J. Jalden and B. Ottersten, The diversity order of the semidefinite relaxation detector, IEEE Trans. Inform. Theory, submitted in 2006.

[12] M. Kisialiou and Z.-Q. Luo, Performance analysis of quasi-maximum-likelihood detector based on semi-definite programming, Proc. ICASSP '05, 3 (2005), pp. III 433 – III 436.

[13] M. Kisialiou and Z.-Q. Luo, Efficient implementation of quasi-maximum-likelihood detection based on semidefinite relaxation, submitted to IEEE Trans. Signal Process., 2008.

[14] Z.-Q. Luo, X. Luo, and M. Kisialiou, An efficient quasi-maximum-likelihood decoder for PSK signals, Proc. ICASSP '03, 6 (2003), pp. VI 561 – VI 564.

[15] Z.-Q. Luo, N.D. Sidiropoulos, P. Tseng, and S. Zhang, Approximation bounds for quadratic optimization with homogeneous quadratic constraints, SIAM J. Optim., 18 (2007), no. 1, pp. 1 – 28.

[16] W.K. Ma, P.C. Ching, and Z. Ding, Semidefinite relaxation based multiuser detection for M-ary PSK multiuser systems, IEEE Trans. Signal Process., 52 (2004), no. 10, pp. 2862–2872.

[17] W.K. Ma, T.N. Davidson, K.M. Wong, Z.-Q. Luo, and P.C. Ching, Quasi-maximum-likelihood multiuser detection using semi-definite relaxation, IEEE Trans. Signal Process., 50 (2002), no. 4, pp. 912 – 922.

[18] V.A. Marcenko and L.A. Pastur, Distribution of eigenvalues for some sets of random matrices, Math. USSR, Sb. 1 (1967), pp. 457 – 483.

[19] A. Mobasher and A.K. Khandani, Matrix-lifting semi-definite programming for decoding in multiple antenna systems, The 10th Canad. Workshop on Inform. Theory (CWIT'07), 2007.

[20] A. Mobasher, M. Taherzadeh, R. Sotirov, and A.K. Khandani, A near maximum likelihood decoding algorithm for MIMO systems based on semi-definite programming, IEEE Trans. Inform. Theory, submitted in 2007.

[21] Y.E. Nesterov, Quality of semidefinite relaxation for nonconvex quadratic optimization, CORE Discussion Paper 9719, 1997.

[22] N.D. Sidiropoulos and Z.-Q. Luo, A semidefinite relaxation approach to MIMO detection for higher-order QAM constellations, IEEE Signal Process. Letters, 13 (2006), pp. 525 – 528.

[23] J.W. Silverstein, The smallest eigenvalue of a large-dimensional Wishart matrix, Ann. Probab., 13 (1985), no. 4, pp. 1364 – 1368.

[24] J.W. Silverstein and Z.D. Bai, On the empirical distribution of eigenvalues of a class of large dimensional random matrices, J. Multivariate Anal., 54 (1995), no. 2, pp. 175 – 192.

[25] T. Tanaka, A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors, IEEE Trans. Inform. Theory, 48 (2002), no. 11, pp. 2888 – 2910.

[26] D.N.C. Tse and S.V. Hanly, Linear multiuser receivers: effective interference, effective bandwidth and user capacity, IEEE Trans. Inform. Theory, 45 (1999), no. 2, pp. 641 – 657.

[27] D.N.C. Tse and S. Verdu, Optimum asymptotic multiuser efficiency of randomly spread CDMA, IEEE Trans. Inform. Theory, 46 (2000), no. 7, pp. 2718–2722.

[28] A.M. Tulino and S. Verdu, Random matrix theory and wireless communications, Found. Trends Comm. Inform. Theory, 1 (2004).

[29] S. Verdu, Multiuser detection, Cambridge University Press, 1998.

[30] E. Viterbo and J. Bouros, A universal lattice code decoder for fading channels, IEEE Trans. Inform. Theory, 45 (1999), no. 5, pp. 1639–1642.

[31] A. Wiesel, Y. Eldar, and S. Shamai, Semidefinite relaxation for detection of 16-QAM signaling in MIMO channels, IEEE Signal Process. Letters, 12 (2005), no. 9, pp. 653–656.