Symbol Error Rate Performance of Box-relaxation Decoders in Massive MIMO

Christos Thrampoulidis[*], Weiyu Xu[†], Babak Hassibi[‡]

Abstract—The maximum-likelihood (ML) decoder for symbol detection in large multiple-input multiple-output wireless communication systems is typically computationally prohibitive. In this paper, we study a popular and practical alternative, namely the Box-relaxation optimization (BRO) decoder, which is a natural convex relaxation of the ML. For iid real Gaussian channels with additive Gaussian noise, we obtain exact asymptotic expressions for the symbol error rate (SER) of the BRO. The formulas are particularly simple, they yield useful insights, and they allow accurate comparisons to the matched-filter bound (MFB) and to the zero-forcing decoder. For BPSK signals the SER performance of the BRO is within 3dB of the MFB for square systems, and it approaches the MFB as the number of receive antennas grows large compared to the number of transmit antennas. Our analysis further characterizes the empirical density function of the solution of the BRO, and shows that error events for any fixed number of symbols are asymptotically independent. The fundamental tool behind the analysis is the convex Gaussian min-max theorem.

I. INTRODUCTION

The problem of recovering an unknown vector of symbols that belong to a finite constellation from a set of noise corrupted linearly related measurements arises in numerous applications, and in particular in multiple-input multiple output (MIMO) wireless communication systems [1, 2, 3, 4]. As a result, a large host of exact and heuristic optimization algorithms have been proposed over the years. Exact algorithms, such as sphere decoding and its variants, become computationally prohibitive as the problem dimension grows, a scenario that is typical in modern massive MIMO systems, e.g., [2]. Heuristic algorithms such as zero-forcing, MMSE, decision-feedback, etc., [5, 6, 7, 8] have inferior performances that are often difficult to precisely characterize. One popular heuristic is the so called box-relaxation optimization decoder, which is a natural convex relaxation of the maximum-likelihood (ML) decoder, and which allows one to recover the signal via efficient convex optimization followed by hard thresholding, e.g., [9, 10, 11]. Despite its popularity, very little is known analytically about the decoding performance of this method. In this paper, we close this gap by deriving exact asymptotic error-rate characterizations under the assumption of real Gaussian wireless channel and additive Gaussian noise.

[*]Research Laboratory of Electronics, MIT, Cambridge, USA, [†]Department of ECE, University of Iowa, Iowa city, USA, [‡]Department of Electrical Engeeniring, Caltech, Pasadena, USA.

A. Problem formulation

We consider the problem of recovering an unknown vector $\mathbf{x}_0$ of n transmitted symbols each belonging to a finite constellation from the noisy multiple-input multiple-output relation, $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^m$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the MIMO channel matrix (assumed to be known) and $\mathbf{z} \in \mathbb{R}^m$ is the noise vector. We assume iid real Gaussian channel with additive Gaussian noise. In particular, $\mathbf{A}$ has entries iid $\mathcal{N}(0, 1/n)$ and $\mathbf{z}$ has entries iid $\mathcal{N}(0, \sigma^2)$. The normalization is such that the signal-to-noise ratio (SNR) varies inversely proportional to the noise variance σ^2. We are interested in the large-system limit, where both the number n of transmit antennas and the number m of receive antennas are large. For simplicity of exposition we assume, for the most part of the paper, that $\mathbf{x}_0$ is an n-dimensional BPSK vector, i.e., $\mathbf{x}_0 \in \{\pm 1\}^n$. Extensions to M-ary constellations are also provided.

Maximum-Likelihood decoder. The ML decoder for BPSK signal recovery, which maximizes the block error probability (assuming the $\mathbf{x}_{0,i}$ are equally likely) is given by $\min_{\mathbf{x} \in \{\pm 1\}^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$. Solving for the exact ML solution is often computationally intractable, especially when n is large, and therefore a variety of heuristics have been proposed (zero-forcing, mmse, decision-feedback, etc.) [12, 8].

Box-relaxation optimization decoder. The heuristic we consider in this paper is the box-relaxation optimization (BRO) decoder [9, 10, 11]. It consists of two steps. The first one involves solving a convex relaxation of the ML algorithm, where $\mathbf{x} \in \{\pm 1\}^n$ is relaxed to $\mathbf{x} \in [-1, 1]^n$. The output of the optimization is hard-thresholded in the second step to produce the final binary estimate. Formally, the algorithm outputs an estimate $\mathbf{x}^*$ of $\mathbf{x}_0$ given as

$$\hat{\mathbf{x}} = \arg \min_{-1 \leq \mathbf{x}_i \leq 1} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2, \quad (1a)$$

$$\mathbf{x}^* = \text{sign}(\hat{\mathbf{x}}), \quad (1b)$$

where the $\text{sign}(\cdot)$ function returns the sign of its input and acts element-wise on input vectors. The BRO decoder naturally extends to the case of recovering signals from higher-order constellations; see Section III.

Symbol error rate. We evaluate the performance of the decoder by the symbol error rate (SER), defined as

$$\text{SER} := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{x}_i^* \neq \mathbf{x}_{0,i}\}}, \quad (2)$$

with $\mathbb{1}_{\{\}}$ used to denote the indicator function. A closely related quantity that is also of interest is the symbol-error

probability P_e, which is defined as the expectation of the SER averaged over the noise, over the channel, and over the constellation. Formally,

$$P_e := \mathbb{E}\left[\text{SER}\right] = \frac{1}{n}\sum_{i=1}^{n}\text{Pr}\left(\mathbf{x}_i^* \neq \mathbf{x}_{0,i}\right). \tag{3}$$

B. Contribution and related work

In this paper, we derive the first rigorous precise characterization of the SER for the BRO decoder in the large-system limit, where the numbers m and n of receive and transmit antennas grow proportionally large at a fixed rate $\delta = m/n$. We complement the precise error formulas with closed-form, tight, upper and lower bounds that are simple functions of the SNR and of δ. These bounds allow useful insights on the decoding performance of the BRO, and they allow a quantitative comparison to the matched-filter bound (MFB) and to the zero-forcing (ZF) decoder. As a concrete example, for BPSK signals the SER of the BRO at high-SNR is $Q(\sqrt{(\delta - 1/2)\text{SNR}})$, where the Q-function is the tail probability of the standard normal distribution. This value is within 3dB of the MFB for square systems, and it approaches the MFB as m approaches n. Finally, we evaluate the large-system empirical distribution of the output of the BRO, and we show that error events for any fixed number of symbol-errors are asymptotically independent.

To the best of our knowledge, a precise formula for the SER was unknown for the BRO. We remark that the replica method developed in statistical mechanics can be used to give formulas for the SER of various detectors in multiuser detection for code-division multiple access (CDMA) or massive MIMO systems. However, the replica method involves a set of conjectured assumptions that remain mostly unverified by rigorous means; please see [13, 8, 14] and references therein. In contrast, our analysis is rigorous, and the techniques used are fundamentally different. They are based on recent advances in comparison inequalities for Gaussian processes; in particular, the convex Gaussian min-max theorem [15, 16].

The present paper is a significantly extended version of our conference paper in [17] [1]. In a related recent line of work [20, 21, 22], the authors have proposed and have investigated the performance of a new class of iterative decoding methods for signal detection in large MIMO systems, which rely on approximate message passing (AMP) [23]. The decoding methods that these papers discuss are different than the BRO decoder, and the analysis tools used are also different than the ones presented here. Interestingly, after our paper [17] appeared, the authors of [22] used our results to show that their proposed algorithm achieves the same error-rate performance as the BRO decoder.

C. Paper Organization

In Section II we analyze the performance of the BRO for BPSK signals. The main theorem of this section, namely

---

[1]The analysis framework that we present here is general and can be used to analyze the performance of other decoders as well. For example, see our recent papers with co-authors [18, 19], which build upon the framework of this work.

Theorem II.1, characterizes the SER and leads to an accurate comparison of the BRO to the MFB and to the ZF decoder. We extend the results to M-PAM constellations in Section III. Section IV includes the main technical result of the paper, namely Theorem IV.1, as well as its detailed proof. The paper concludes in Section V with a discussion on future research directions. Finally, some technical proofs are deferred to the Appendix.

II. THE BRO DECODER FOR BPSK SIGNALS

We precisely analyze the error-rate performance of the BRO decoder for BPSK signals. Our main Theorem II.1 in Section II-A evaluates its symbol error rate, and simple, closed-form (upper and lower) bounds are computed in Section II-B. In Sections II-C and II-D we use these bounds to compare the BRO decoder to the matched-filter bound, and to the zero-forcing decoder, respectively.

A. Precise SER performance

Our main result explicitly characterizes the limiting behavior of the SER of the BRO in (1), under a large-system limit in which $m, n \to +\infty$ at a proportional (constant) rate $\delta > 0$. The SNR is assumed constant; in particular, it does not scale with n. Note that for $\mathbf{x}_0 \in \{\pm 1\}^n$, $\text{SNR} = 1/\sigma^2$.

We use standard notation $\text{plim}_{n \to \infty} X_n = X$ to denote that a sequence of random variables X_n converges *in probability* towards a constant X. All limits will be taken in the regime $m, n \to +\infty, m/n = \delta$; to keep notation short we simply write $n \to \infty$. Finally, we use $Q(\cdot)$ denote the Q-function associated with the standard normal density $p(h) = \frac{1}{\sqrt{2\pi}}e^{-h^2/2}$.

Theorem II.1 (SER for BPSK signals). *Let* SER *denote the symbol-error-rate of the box-relaxation optimization decoder in* (1)*, for some fixed but unknown BPSK signal* $\mathbf{x}_0 \in \{\pm 1\}^n$. *Fix a constant* SNR *and a constant* $\delta \in (\frac{1}{2}, +\infty)$. *Then, in the limit of* $m, n \to +\infty$, $m/n = \delta$, *it holds:*

$$\text{plim}_{n \to \infty} \text{SER} = Q\left(\frac{1}{\tau_*}\right),$$

where τ_* *is the unique positive minimizer of the strictly convex function* $F : (0, +\infty) \to \mathbb{R}$ *defined as:*

$$F(\tau) := \tau\left(\delta - \frac{1}{2}\right) + \frac{1/\text{SNR}}{\tau} + \left(\tau + \frac{4}{\tau}\right)Q\left(\frac{2}{\tau}\right) - \sqrt{\frac{2}{\pi}}e^{-\frac{2}{\tau^2}}. \tag{4}$$

The theorem explicitly characterizes the high-probability limit of the SER over the randomness of the channel matrix $\mathbf{A}$, and of the noise vector $\mathbf{z}$. The function $F(\tau)$ in (4) is deterministic, strictly convex, and is parametrized by the value of the SNR and by the proportionality factor δ.

The proof of the theorem uses the convex Gaussian min-max theorem (CGMT) [15, 16], which has thus far found major use in precisely quantifying the squared-error performance of regularized M-estimators in high-dimensions, such as the LASSO [16]. In this paper we extend the applicability of the CGMT to the characterization of the SER performance, to arrive to Theorem II.1. More than that, along the way we
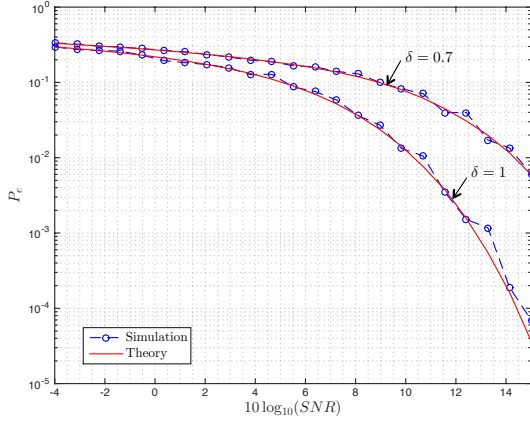
Fig. 1: Symbol-error probability of the BRO as a function of SNR for different values of the ratio δ of receive to transmit antennas. The theoretical prediction follows from Theorem II.1. For the simulations, we used $n = 512$. The data are sample averages of the SER over 20 independent realizations of the channel matrix and of the noise vector for each value of the SNR.

prove a number of even stronger statements regarding the error performance of the BRO. We:

(i) establish the large-system error performance of the BRO for a wide class of performance metrics; this class includes the squared-error and the SER as special cases.

(ii) explicitly characterize the limiting empirical distribution of the output $\hat{\mathbf{x}}$ of (1a).

(iii) show that error events for any fixed number of bits are asymptotically independent.

Please refer to Theorem IV.1 and to Corollary IV.1 for the formal statements of these results. The detailed proof of Theorem II.1 is also deferred to Section IV.

Some further remarks on Theorem II.1 are given below.

1) On $\delta > \frac{1}{2}$: The theorem holds as long as the ratio of proportionality δ is (strictly) greater than 1/2. To begin with, note that this allows for the number of receive antennas to be less than the number of transmit antennas, and as low as (almost) half of them. When $\delta < 1$ the system of linear equations $\mathbf{y} = \mathbf{A}\mathbf{x}$ is underdetermined; hence, recovering the true solution is generally ill-posed even in the the absence of noise. However, in the problem of interest it is a-priori known that the true solution $\mathbf{x}_0$ only takes values $\{\pm 1\}^n$. The BRO decoder uses that information by enforcing an ℓ_∞-norm constraint in (1a). Of course, this idea of using convex optimization with (typically non-smooth) constraints that promote the particular structure of the unknown signal $\mathbf{x}_0$ to solve underdetermined system of equations, is one of the core ideas that emerged from the Compressed Sensing literature (e.g. [24]). In fact, it is by now well-understood that in the noiseless case the program in (1a) successfully recovers the true $\mathbf{x}_0 \in \{\pm 1\}^n$ with high probability over the randomness of $\mathbf{A}$ if and only if $\delta > 1/2$ ([24, 25]). The same necessary condition naturally arises out of our proof of

Theorem II.1.

2) Probability of error: Recall from (3) that the symbol-error probability is given as $P_e = \mathbb{E}[\text{SER}]$. Also, the SER is bounded between 0 and 1. Thus, using Theorem II.1 we show in Appendix A1 that P_e converges (deterministically) to the same value $Q(1/\tau_*)$.

Corollary II.1 (P_e). *Under the setting of Theorem II.1, let P_e denote the symbol-error probability of the BRO and τ_* be the minimizer of (4). Then,*

$$\lim_{n \to \infty} P_e = Q\left(1/\tau_*\right).$$

3) Solving for τ_*: In order to evaluate the large-system limit of the SER, one needs to compute the unique positive minimizer of $F(\tau)$ in (4). The function F is strictly convex, hence this can be done numerically in an efficient way. Due to convexity, τ_* can also be described as the unique solution to the first order optimality conditions of the minimization program (see Lemma A.2). By further analyzing the properties of τ_*, we derive in Section II-B simple closed-form (upper and lower) bounds on the quantity of interest, namely $Q(1/\tau_*)$.

4) Numerical illustration: Figure 1 illustrates the accuracy of the prediction of Theorem II.1. Note that although the theorem requires $n \to \infty$, the prediction is already accurate for n on the scale of a few hundreds.

B. Simple bounds and high-SNR regime

We derive simple, closed-form upper and lower bounds on $Q(1/\tau_*)$, the limiting value of the SER. We further show that these bounds are tight. The proof is deferred to Appendix A2.

Theorem II.2 (Closed-form bounds). *Let τ_* be the unique minimizer of (4). Then, for all values of $\delta > 1/2$ and all values of $\text{SNR} > 0$, it holds,*

$$Q(\sqrt{\delta \cdot \text{SNR}}) < Q(1/\tau_*) \leq Q(\sqrt{(\delta - 1/2) \cdot \text{SNR}}). \quad (5)$$

Furthermore, the upper bound becomes tight as $\text{SNR} \to +\infty$.

In view of Theorem II.1, the statement in (5) directly establishes upper and lower bounds on the (asymptotic) SER performance of the BRO. These bounds are given in closed-form and are simple functions of δ and of SNR.

As stated in the theorem, the upper bound in (5) becomes tight in the high-SNR regime. Hence, for $\text{SNR} \gg 1$, in the limit of $n \to \infty$,

$$\text{SER} \approx Q(\sqrt{(\delta - 1/2) \cdot \text{SNR}}). \quad (6)$$

A formal statement of this result is given in Theorem A.1 in Appendix A2. The fact that $\tau_* \approx 1/\sqrt{(\delta - 1/2)SNR}$ when $\text{SNR} \gg 1$, can be intuitively understood as follows: at high-SNR we expect τ_* to be going to zero (correspondingly SER or $Q(1/\tau_*)$ to be small). When this is the case, the last two summands in (4) are negligible; then, τ_* is the solution to $\min_{\tau > 0} \tau \left(\delta - \frac{1}{2}\right) + \frac{1/\text{SNR}}{\tau}$, which gives the derided result.

For illustration, in Figure 2 we have plotted the high-SNR expression for the SER in (6) versus its exact value as predicted by Theorem II.1. It is seen that, as already discussed,
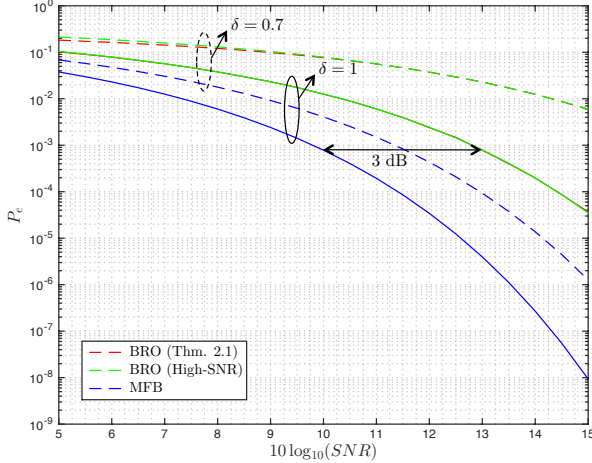
Fig. 2: Symbol-error probability of the BRO (in red, see Theorem II.1) in comparison to high-SNR approximation (in green, see (6)) and to the matched filter bound (in blue, see (7)), for $\delta = 0.7$ (dashed lines) and $\delta = 1$ (solid lines). Theorem II.2 successfully predicts that the red curves are sandwiched between the corresponding green (upper-bound) and blue ones (lower-bound).

the high-SNR expression is an upper bound, and in fact a good proxy, for the true probability of error at all values of SNR. The approximation becomes better with increasing δ.

Finally, in Section II-C, we show that the lower bound $Q(\sqrt{\delta \cdot \text{SNR}})$ has an operational meaning: it is equal to the bit error probability of an isolated bit transmission over the channel, which is also known as the *matched filter bound* in digital communications.

C. Comparison to the matched filter bound

Here, we compare the performance of the BRO to an idealistic case, where all $n-1$, but 1, bits of $\mathbf{x}_0$ are known to us. As is customary in the field, we refer to the symbol error probability of this case as the *matched filter bound* (MFB) and denote it by P_e^{MFB}. The MFB corresponds to the probability of error in detecting (say) $\mathbf{x}_{0,n} \in \{\pm 1\}$ from: $\tilde{\mathbf{y}} = \mathbf{x}_{0,n}\mathbf{a}_n + \mathbf{z}$, where $\tilde{\mathbf{y}} = \mathbf{y} - \sum_{i=1}^{n-1} \mathbf{x}_{0,i}\mathbf{a}_i$ is assumed known, and, $\mathbf{a}_i$ denotes the ith column of $\mathbf{A}$. (This can be equivalently thought of as the error probability of an isolated transmission of only the last bit over the channel.) The ML estimate is equal to the sign of the projection of the vector $\tilde{\mathbf{y}}$ to the direction of $\mathbf{a}_n$. Without loss of generality we assume that $\mathbf{x}_{0,n} = +1$. Then, the output of the matched filter becomes $\text{sign}(\tilde{X})$, where

$$\tilde{X} = \|\mathbf{a}_n\|^2 + \sigma^2 \tilde{\mathbf{z}}_n,$$

and $\tilde{\mathbf{z}}_n \sim \mathcal{N}(0,1)$. Recall that the entries of the m-dimensional vector $\mathbf{a}_n$ are iid $\mathcal{N}(0,1/n)$, so it holds $\text{plim}_{n\to\infty} \|\mathbf{a}_n\| = \delta$. Hence,

$$\lim_{n\to\infty} P_e^{MFB} = \lim_{n\to\infty} \mathbb{P}(\tilde{X} < 0) = Q(\sqrt{\delta \cdot \text{SNR}}). \quad (7)$$

First, observe that this formula coincides with the lower bound on the probability of error of the BRO derived in

Theorem II.2. Combined, they establish formally that the MFB is (strictly) better than the BRO. Of course, this is naturally expected since the former is an idealistic scheme.

Next, when compared to the upper-bound on the probability of error of the BRO derived in Theorem II.2, the formula in (7), leads to the following conclusion:

The BRO achieves a desired symbol-error probability at a higher SNR value by at most $10 \log_{10} \frac{\delta}{\delta - 1/2} dB$ than that predicted by the MFB.

In particular, in the square case ($\delta = 1$), where the number of receive and transmit antennas are the same, the BRO is 3dB off the MFB (cf., Figure 2). When the number of receive antennas is much larger, i.e., when $\delta \gg 1$, then the performance of the BRO approaches the MFB.

D. Box-relaxation vs Zero-forcing

In this section, we use Theorem II.1 to compare the performance of the BRO to another widely used decoder, namely the *zero-forcing* (ZF) decoder. The ZF decoder obtains an estimate $\mathbf{x}_{ZF}^*$ of $\mathbf{x}_0$ as follows

$$\hat{\mathbf{x}}_{ZF} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2, \quad (8a)$$

$$\mathbf{x}_{ZF}^* = \text{sign}(\hat{\mathbf{x}}_{ZF}). \quad (8b)$$

Observe that this is very similar to the BRO, only that in (8a) the minimization is *unconstrained*. Therefore, in contrast to the BRO, for the ZF decoder we require $\delta > 1$, i.e., the number of receive antennas be larger than the number of transmit antennas. When this is the case and n is large, $\mathbf{A}$ is full column-rank with probability one, and (8a) has a unique closed-form solution:

$$\hat{\mathbf{x}}_{ZF} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}. \quad (9)$$

In particular, it is a well-known result in the literature how to use standard tools from random matrix theory to derive the symbol-error probability of the ZF decoder (e.g. [7]). For convenience of the reader, we briefly summarize the main idea here. Without loss of generality, consider the last bit $\mathbf{x}_n$ of $\mathbf{x}$. Further let $\mathbf{A} = \mathbf{Q}\mathbf{R}$ be the QR decomposition of $\mathbf{A}$, such that $\mathbf{Q} \in \mathbb{R}^{m\times n}$ is a matrix with orthogonal columns and $\mathbf{R} \in \mathbb{R}^{n\times n}$ is upper triangular. Define $\tilde{\mathbf{y}} := \mathbf{Q}^T\mathbf{y}$ and $\tilde{\mathbf{z}} := \mathbf{Q}^T\mathbf{z}$ and note that

$$\tilde{\mathbf{y}}_n = \mathbf{R}_{nn}\mathbf{x}_n + \tilde{\mathbf{z}}_n,$$

where $\mathbf{R}_{nn}$ is the nth diagonal element of $\mathbf{R}$. From the rotational invariance of the Gaussian distribution, it holds $\tilde{\mathbf{z}}_n \sim \mathcal{N}(0,\sigma^2)$. Next, we use the following well-known facts, e.g., [7, Lem. 1]: (i) $\mathbf{Q}$ and $\mathbf{R}$ are independent matrices. Hence, $\tilde{\mathbf{z}}_n$ is independent of $\mathbf{R}_{nn}$; (ii) $\mathbf{R}_{nn}$ is such that $n\mathbf{R}_{nn}^2$ is χ^2 random variable with $(m-n+1)$ degrees of freedom. Thus, by the corresponding formula for BPSK single-input single-output (SISO) Gaussian channel, the symbol-error probability of the zero-forcing decoder is

$$P_e^{ZF} = E_{\gamma_1,\dots,\gamma_{m-n+1}}\left[Q\left(\sqrt{\frac{\frac{1}{n}\sum_{i=1}^{m-n+1}\gamma_i^2}{\sigma^2}}\right)\right],$$

where γ_i's are iid standard Gaussians $\mathcal{N}(0,1)$. But, $\text{plim}_{n\to\infty} \frac{\sum_{i=1}^{m-n+1} \gamma_i^2}{n} = (\delta - 1)$, giving

$$\lim_{n\to\infty} P_e^{ZF} = Q(\sqrt{(\delta-1)\cdot\text{SNR}}). \qquad (10)$$

Comparing this formula to the upper bound on the probability of error of the BRO derived in Theorem II.2, we formally quantify the superiority of the BRO over the ZF decoder:

The BRO achieves the same performance as the ZF decoder at a lower SNR value by at least $10\log_{10}\left(\frac{\delta-\frac{1}{2}}{\delta-1}\right)dB$.

This holds for $\delta > 1$. However, Theorem II.1 further shows that the BRO can successfully decode even when $\delta < 1$, and in particular as low as $1/2$.

Above, we derived formula (10) using tools from random matrix theory. Alternatively, we can obtain the same result using the CGMT, and the proof technique is very similar to that of Theorem II.1. The use of random-matrix-theory tools for the analysis of the ZF decoder is in large possible because the minimizer $\hat{x}_{ZF}$ of (8a) can be expressed in closed-form as a function of $\mathbf{A}$ and $\mathbf{z}$ (see (9)). On the contrary, this is not the case with the BRO decoder and the use of the CGMT is critical for establishing Theorem II.1.

III. EXTENSION TO M-PAM CONSTELLATIONS

A. Setting

Each transmit antenna sends a symbol $\mathbf{x}_{0,i}$ that take values

$$\mathbf{x}_{0,i} \in \mathcal{C} := \{\pm 1, \pm 3, \ldots, \pm(M-1)\},$$

for some $M = 2^b$ and b a positive integer. When each antenna transmits a single bit, i.e. $b = 1$, then $\mathbf{x}_0 \in \{\pm 1\}^n$ and the setting is the same as in Section II. As always, we assume additive Gaussian noise of variance σ^2.

The ML decoder is given by $\min_{\mathbf{x}\in\mathcal{C}^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$, but it is often computationally intractable for large number of receive/transmit antennas. We consider, the natural extension of the box-relaxation decoder for BPSK in (1). Specifically, for M-PAM symbol transmission, the BRO outputs an estimate $\mathbf{x}^*$ of $\mathbf{x}_0$ as follows:

$$\hat{\mathbf{x}} = \arg\min_{-(M-1)\leq\mathbf{x}_i\leq(M-1)} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2, \qquad (11a)$$

$$\mathbf{x}_i^* = \arg\min_{c\in\mathcal{C}} |\hat{\mathbf{x}}_i - c|. \qquad (11b)$$

The optimization in (11a) is convex, and (11b) simply selects the symbol value c that is closest to the solution $\hat{\mathbf{x}}_i$ among a total of M choices: $\{\pm 1, \pm 3, \ldots, \pm(M-1)\}$. Therefore, the proposed decoder is computationally efficient. In the next section, we evaluate its error-rate performance.

B. SER performance

Theorem III.1 below precisely characterizes the large-system limit of the SER of the BRO in (11) under an M-PAM transmission. We assume that a *typical sequence* of symbols is sent over the channel, i.e., each transmitted symbol $\mathbf{x}_{0,i}$ takes values $\{\pm 1, \pm 3, \ldots, \pm(M-1)\}$ with equal probability $1/M$. The result extends to other distributions over the constellation,
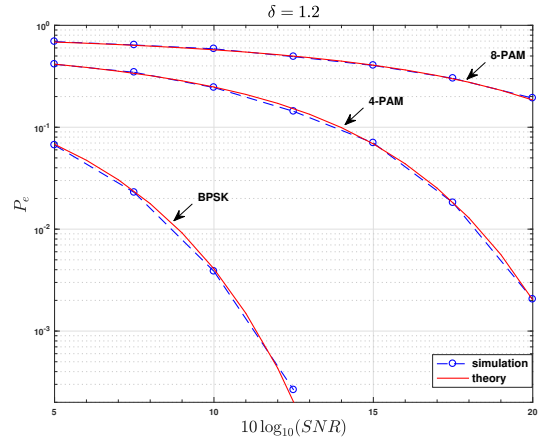

Fig. 3: Symbol error probability of the Box Relaxation Optimization (BRO) in (11) as a function of the SNR for BPSK, 4-PAM and 8-PAM signals. The theoretical prediction follows from Theorem III.1. For the simulations, we used n = 512 and $\delta = 1.2$. The data are averages over 20 independent realizations of the channel matrix and of the noise vector for each value of the SNR.

but for simplicity we focus on this typical case. For a typical sequence, the average power of the transmitted vector $\mathbf{x}_0$ is

$$\mathbb{E}[\mathbf{x}_{0,i}^2] = (2/M) \sum_{i=1,3\ldots,M-1} i^2 = (M^2 - 1)/3.$$

Therefore, the SNR of the system becomes

$$\text{SNR} = (M^2 - 1)/3\sigma^2. \qquad (12)$$

Theorem III.1 (SER for M-PAM). *Let* SER *denote the symbol error rate of the detection scheme in* (11), *for a typical transmitted signal* $\mathbf{x}_0$ *such that each symbol* $\mathbf{x}_{0,i}$ *takes values* $\{\pm 1, \pm 3, \ldots, \pm(M-1)\}$ *with equal probability* $1/M$. *Fix a constant noise variance* σ^2 *(eqv., a constant* SNR *as in* (12)) *and a constant* $\delta \in (1 - \frac{1}{M}, +\infty)$. *Then, in the limit of* $m, n \to \infty$, $m/n = \delta$, *it holds:*

$$\text{plim}_{n\to\infty} \text{SER} = 2\left(1 - \frac{1}{M}\right) Q\left(\frac{1}{\tau_*}\right),$$

where τ_* *is the unique positive minimizer of the strictly convex function* $F_M : (0, +\infty) \to \mathbb{R}$ *defined as:*

$$F_M(\tau) := \frac{\tau}{2}\left(\delta - \frac{M-1}{M}\right) + \frac{\sigma^2}{2\tau} + \frac{1}{M}\sum_{k=2,4,\ldots,2(M-1)} S(\tau;k), \qquad (13)$$

with,

$$S(\tau;k) := \left(\tau + \frac{k^2}{\tau}\right) Q\left(\frac{k}{\tau}\right) - \frac{k}{\sqrt{2\pi}} e^{-\frac{k^2}{2\tau^2}}. \qquad (14)$$

Theorem III.1 generalizes Theorem II.1, and the former reproduces the latter for $M = 2$. Figure 3 illustrates the accuracy of the prediction. The proof of the theorem is defered to Appendix C.

Most of the remarks that followed the statement of Theorem II.1 in Section II, are readily extended to general M-PAM constellations. The guarantees of Theorem III.1 hold as long as the ratio of transmit to receive antennas δ is larger than $1 - 1/M$. Thus, successful transmission is possible with fewer number of receive than transmit antennas. The minimum allowed ratio increases for higher-order constellations. Similar to Theorem II.2, we can show the following simple upper bound on probability of error P_e for all values of SNR:

$$\lim_{n \to \infty} P_e \le 2\Big(1 - \frac{1}{M}\Big) Q\left(\sqrt{\Big(\delta - 1 + \frac{1}{M}\Big)\Big(\frac{3}{M^2 - 1}\Big)\mathrm{SNR}}\right).$$
(15)

Moreover, the bound is *tight* at high-SNR. Of course, for $M = 2$, this coincides with the upper bound in (5).

IV. PROOF OF MAIN RESULT

This section includes the proof of Theorem II.1. In fact, towards proving the theorem, we obtain a more general result which is stated as Theorem IV.1 below.

For simplicity, we make use of the following notation onwards. We say that an event $\mathcal{E}$ holds with probability approaching 1 (*w.p.a.1*) if $\lim_{n \to \infty} \mathbb{P}(\mathcal{E}) = 1$. Also, we use the following shorthands: $X_n \xrightarrow{P} X$ to denote convergence in probability; $X \overset{d}{=} Y$ to denote that the random variables X and Y have the same distribution; and, $\|\cdot\|$ to denote the n-dimensional Euclidean norm.

A. Main technical result

As far as the performance is concerned, we can assume without loss of generality that $\mathbf{x}_0 = +\mathbf{1}_n = (1, 1, \ldots, 1)$. Also, it is convenient to re-write (1a) by changing the variable to the *error vector* $\mathbf{w} := \mathbf{x} - \mathbf{x}_0 = \mathbf{x} - \mathbf{1}$:

$$\hat{\mathbf{w}} := \arg \min_{-2 \le \mathbf{w}_i \le 0} \|\mathbf{z} - \mathbf{A}\mathbf{w}\|.$$
(16)

Then, observe that the SER defined in (2) is written in terms of the error vector $\mathbf{w}$ as:

$$\mathrm{SER} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\hat{\mathbf{w}}_i \le -1\}}.$$
(17)

The following theorem characterizes the limit of the empirical distribution of the optimal solution $\hat{\mathbf{w}}$ in (16), and yields Theorem II.1 as a corollary.

Theorem IV.1 (Lipschitz metrics and empirical distribution). *Recall the definition of τ_* in Theorem II.1, and assume, without loss of generality, that $\mathbf{x}_0 = +\mathbf{1}$. Let $\hat{\mathbf{w}}$ be as in (16) and consider its (normalized) empirical density function*

$$\mu_{\hat{\mathbf{w}}} := n^{-1} \sum_{i=1}^n \delta_{\hat{\mathbf{w}}_i}.$$

Further, consider the function $\theta : \mathbb{R} \to [-2, 0]$:

$$\theta(\gamma) := \begin{cases} 0 & , \textit{if } \gamma \ge 0, \\ \tau_* \gamma & , \textit{if } -\frac{2}{\tau_*} \le \gamma < 0, \\ -2 & , \textit{if } \gamma < -\frac{2}{\tau_*}, \end{cases}$$
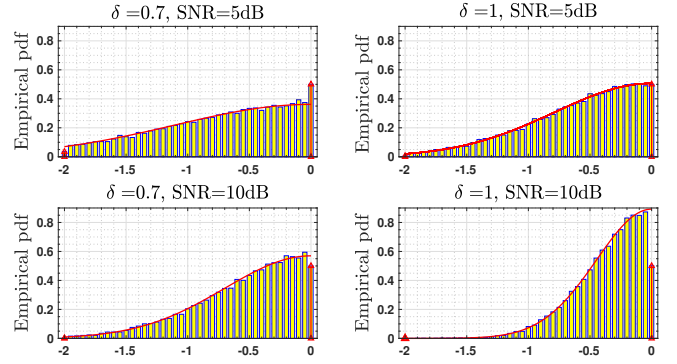

Fig. 4: Empirical distribution of the error vector $\mathbf{w} := \hat{\mathbf{x}} - \mathbf{x}_0$ (conditioned on $\mathbf{x}_0 = +\mathbf{1}$) for the solution $\hat{\mathbf{x}}$ of the BRO. The empirical histograms shown are averages over 200 realizations of the channel matrix and of the noise vector for $n = 256$ number of transmit antennas. They are compared to the asymptotic limiting distribution predicted by Theorem IV.1, see (18). The limiting density is supported in the interval $[-2, 0]$ and has point masses both at -2 and 0. Different values of δ and of SNR are shown.

and let μ_W be the density measure of a random variable W

$$W \overset{d}{=} \theta(\mathcal{N}(0, 1)).$$
(18)

The following are true:

(a). $\mu_{\tilde{\mathbf{w}}}$ converges weakly in probability to μ_W. [2]

(b). For all Lipschitz functions $\psi : \mathbb{R} \to \mathbb{R}$ with Lipschitz constant L (independent of n), it holds

$$\frac{1}{n} \sum_{i=1}^n \psi(\hat{\mathbf{w}}_i) \xrightarrow{P} \mathbb{E}_W[\psi(W)].$$

Theorem IV.1 is the main technical result of this paper. In Section IV-B we show how it can be used to prove Theorem II.1. Next, in Section IV-C we rely again on Theorem IV.1 to prove that error events for any fixed number of bits are asymptotically independent. The rest of Section IV is devoted to the proof of Theorem IV.1.

B. Proof of Theorem II.1

On the one hand, by (17), it suffices to prove that $\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\hat{\mathbf{w}}_i \le -1\}} \xrightarrow{P} Q(1/\tau_*)$. On the other hand, it is easily checked that $\mathbb{E}_W[\mathbb{1}_{\{W \le -1\}}] = \mathbb{E}_{\gamma \sim \mathcal{N}(0,1)}[\mathbb{1}_{\{\gamma \le -1/\tau_*\}}] = Q(1/\tau_*)$. Note that the indicator function $\mathbb{1}_{\{W \le -1\}}$ is not Lipschitz, so we cannot directly apply Theorem IV.1(b). However, since the discontinuity point (i.e., -1) of the indicator function has μ_W-measure zero, and also W is a *continuous* random variable, one can appropriately approximate the indicator with Lipschitz functions and conclude the desired based on Theorem IV.1(b). This is a somewhat standard argument, but

[2]Note that $\mu_{\tilde{\mathbf{w}}}$ defines a (sequence of) *random* probability measure(s); on the other hand, μ_W is a deterministic measure. We use terminology that is standard in the theory of random matrices and say that a sequence of random measures μ_n converges weakly to a deterministic measure μ if for every continuous compactly supported ψ: $\int \psi \mathrm{d}\mu_n \xrightarrow{P} \int \psi \mathrm{d}\mu$ (see for example [26, pg. 160]).

we reproduce a detailed proof of the claim in Lemma A.3 in Appendix B for completeness.

C. Independence of Error Events

Here, we obtain as a corollary of Theorem IV.1 that error events for any fixed number of bits are *asymptotically* independent. We defer the proof of the corollary to Appendix B2.

Corollary IV.1 (Independence of error events)**.** *Under the notation and definition of Theorem IV.1, let $\psi_i : \mathbb{R} \to \mathbb{R}, i = 1, \ldots, k$ be bounded Lipschitz functions for fixed $k \geq 2$. Then, it holds*

$$n^{-k} \sum_{1 \leq i_1, \ldots, i_k \leq n} \psi_1(\hat{\mathbf{w}}_{i_1}) \cdots \psi_k(\hat{\mathbf{w}}_{i_k}) \xrightarrow{P} \prod_{\ell=1}^{k} \mathbb{E}[\psi_\ell(W_\ell)],$$

where the expectations of the right-hand side are with respect to $W_1, \ldots, W_k$ that are iid random variables distributed as $\theta(\mathcal{N}(0,1))$. Moreover, it holds

$$n^{-k} \sum_{1 \leq i_1, \ldots, i_k \leq n} \mathbb{1}_{\{\hat{\mathbf{w}}_{i_1} \leq -1, \ldots, \hat{\mathbf{w}}_{i_k} \leq -1\}} \xrightarrow{P} (Q(1/\tau_*))^k.$$

D. The convex Gaussian min-max theorem

The fundamental tool behind our analysis is the convex Gaussian min-max theorem (CGMT) [16], [15]. The CGMT associates with a primary optimization (PO) problem a simplified auxiliary optimization (AO) problem from which we can tightly infer properties of the original (PO), such as the optimal cost, the optimal solution, etc.. In particular, the (PO) and (AO) problems are defined respectively as follows:

$$\Phi(\mathbf{G}) := \min_{\mathbf{w} \in \mathcal{S}_\mathbf{w}} \max_{\mathbf{u} \in \mathcal{S}_\mathbf{u}} \mathbf{u}^T \mathbf{G} \mathbf{w} + \psi(\mathbf{w}, \mathbf{u}), \tag{19a}$$

$$\phi(\mathbf{g}, \mathbf{h}) := \min_{\mathbf{w} \in \mathcal{S}_\mathbf{w}} \max_{\mathbf{u} \in \mathcal{S}_\mathbf{u}} \|\mathbf{w}\| \mathbf{g}^T \mathbf{u} - \|\mathbf{u}\| \mathbf{h}^T \mathbf{w} + \psi(\mathbf{w}, \mathbf{u}), \tag{19b}$$

where $\mathbf{G} \in \mathbb{R}^{m \times n}, \mathbf{g} \in \mathbb{R}^m, \mathbf{h} \in \mathbb{R}^n, \mathcal{S}_\mathbf{w} \subset \mathbb{R}^n, \mathcal{S}_\mathbf{u} \subset \mathbb{R}^m$ and $\psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. Denote $\mathbf{w}_\Phi := \mathbf{w}_\Phi(\mathbf{G})$ and $\mathbf{w}_\phi := \mathbf{w}_\phi(\mathbf{g}, \mathbf{h})$ any optimal minimizers in (19a) and (19b), respectively. Further let $\mathcal{S}_\mathbf{w}, \mathcal{S}_\mathbf{u}$ be convex and compact sets, ψ be continuous and convex-concave on $\mathcal{S}_\mathbf{w} \times \mathcal{S}_\mathbf{u}$, and, $\mathbf{G}, \mathbf{g}$ and $\mathbf{h}$ all have entries iid standard normal.

Theorem IV.2 (CGMT, [16])**.** *Let $\mathcal{S}$ be an arbitrary open subset of $\mathcal{S}_\mathbf{w}$ and $\mathcal{S}^c = \mathcal{S}_\mathbf{w}/\mathcal{S}$. Denote $\phi_{\mathcal{S}^c}(\mathbf{g}, \mathbf{h})$ the optimal cost of the optimization in (19b), when the minimization over $\mathbf{w}$ is now constrained over $\mathbf{w} \in \mathcal{S}^c$. Suppose there exist constants $\overline{\phi}$ and $\eta > 0$ such that in the limit of $n \to +\infty$ it holds w.p.a.1: (i) $\phi(\mathbf{g}, \mathbf{h}) \leq \overline{\phi} + \eta$, and, (ii) $\phi_{\mathcal{S}^c}(\mathbf{g}, \mathbf{h}) \geq \overline{\phi} + 2\eta$. Then, $\lim_{n \to \infty} \Pr(\mathbf{w}_\Phi \in \mathcal{S}) = 1$.*

It is not hard to argue that the conditions (i) and (ii) regarding the optimal cost of the (AO) imply the following for its solution: $\mathbf{w}_\phi \in \mathcal{S}$ w.p.a.1. The non-trivial and powerful part of the theorem is that the same conclusion is true for the optimal solution $\mathbf{w}_\Phi$ of the (PO) as well. The CGMT builds upon a classical result due to Gordon [27]. Gordon's original result is classically used to establish non-asymptotic probabilistic lower bounds on the minimum singular value of

Gaussian matrices [28], and has a number of other applications in high-dimensional convex geometry [29], [30]. The idea of combining the GMT with convexity is attributed to Stojnic [31]. Thrampoulidis et. al. built and significantly extended on this idea arriving at the CGMT as it appears in [16, Thm. 6.1].

E. Proof of Theorem IV.1

1) Strategy: We will first prove Theorem IV.1(b); Part (a) will then follow by standard arguments from the theory of weak convergence.

As mentioned the proof is based on the use of the CGMT. The first step is to identify the (PO) and the (AO), such that $\hat{\mathbf{w}}$ is optimal for the (PO). Then, our goal is to apply Theorem IV.2 to the following set

$$\mathcal{S}_\epsilon := \{\mathbf{v} \ : \ \left|n^{-1} \sum_{i=1}^{n} \psi(\mathbf{v}_i) - \mathbb{E}_W[\psi(W)]\right| < \epsilon\}, \tag{20}$$

where $\epsilon > 0$ is arbitrary. To see that this is desired note that if for all $\epsilon > 0$ it holds $\mathbf{w} \in \mathcal{S}_\epsilon$ w.p.a.1, then $n^{-1} \sum_{i=1}^{n} \psi(\mathbf{w}_i) \xrightarrow{P} \mathbb{E}_W[\psi(W)]$. Thus, the bulk of the proof amounts to checking that the conditions of Theorem IV.2 are satisfied for $\mathcal{S}_\epsilon$ in (20). For the rest of the proof, we fix $\epsilon > 0$ and denote $\mathcal{S} := \mathcal{S}_\epsilon$, for convenience

2) Identifying the (PO) and the (AO): Using the CGMT for the analysis of the SER, requires as a first step expressing the optimization in (1a) in the form of a (PO) as it appears in (19a). It is easy to see that (16) is equivalent to

$$\frac{1}{\sqrt{n}} \min_{-2 \leq \mathbf{w}_i \leq 0} \max_{\|\mathbf{u}\| \leq 1} \mathbf{u}^T \mathbf{A} \mathbf{w} - \mathbf{u}^T \mathbf{z}. \tag{21}$$

Observe that the constraint sets above are both convex and compact; also, the objective function is convex in $\mathbf{w}$ and concave in $\mathbf{u}$. These are consistent with the requirements of the CGMT. The corresponding (AO) problem becomes:

$$\phi(\mathbf{g}, \mathbf{h}) := \frac{1}{n} \min_{-2 \leq \mathbf{w}_i \leq 0} \max_{\|\mathbf{u}\| \leq 1} (\|\mathbf{w}\| \mathbf{g} - \sqrt{n}\mathbf{z})^T \mathbf{u} - \|\mathbf{u}\| \mathbf{h}^T \mathbf{w}. \tag{22}$$

Note the normalization to account for the variance of the entries of $\mathbf{A}$. Onwards, we refer to the optimization in (22) as the (AO) problem.

3) Simplifying the (AO): We begin by simplifying the (AO) problem as it appears in (22). First, since both $\mathbf{g}$ and $\mathbf{z}$ have entries iid Gaussian, then, the vector $\|\mathbf{w}\| \mathbf{g} - \sqrt{n}\mathbf{z}$ has entries iid $\mathcal{N}(0, \sqrt{\|\mathbf{w}\|^2 + n\sigma^2})$. Hence, for our purposes and using some abuse of notation so that $\mathbf{g}$ continues to denote a vector with iid standard normal entries, the first term in (22) can be treated as $\sqrt{\|\mathbf{w}\|^2 + n\sigma^2} \mathbf{g}^T \mathbf{u}$, instead. As a next step, fix the norm of $\mathbf{u}$ to say $\|\mathbf{u}\| = \beta$. Optimizing over its direction is now straightforward, and gives

$$\min_{-2 \leq \mathbf{w}_i \leq 0} \max_{0 \leq \beta \leq 1} \frac{\beta}{n} \left(\sqrt{\|\mathbf{w}\|^2 + n\sigma^2} \|\mathbf{g}\| - \mathbf{h}^T \mathbf{w}\right).$$

In fact, it is easy to now further optimize over β as well: its optimizing value is 1 if the term in the parenthesis is non-

negative, and, is 0 otherwise. With this, the (AO) simplifies to the following:

$$\phi(\mathbf{g}, \mathbf{h}) = \min_{-2 \leq \mathbf{w}_i \leq 0} \left(\sqrt{\frac{\|\mathbf{w}\|^2}{n} + \sigma^2} \frac{\|\mathbf{g}\|}{\sqrt{n}} - \frac{1}{n} \mathbf{h}^T \mathbf{w} \right)_+, \quad (23)$$

where we used the notation $(\cdot)_+ := \max\{\cdot, 0\}$.

In order to perform the optimization over $\mathbf{w}$, we will express the "square-root term" $\chi := \chi(\mathbf{w}) := \sqrt{\|\mathbf{w}\|^2/n + \sigma^2}$ in a variational form. First, observe that all $\mathbf{w} \in [-2, 0]^n$ satisfy $\sigma^2 \leq \chi \leq 4 + \sigma^2 := T$. Hence, we can write

$$\chi = \min_{0 \leq \tau \leq T} \frac{\tau}{2} + \frac{\chi^2}{2\tau}.$$

With this trick, the minimization over the entries of $\mathbf{w}$ becomes separable as follows:

$$\min_{0 \leq \tau \leq T} \frac{\tau \|\mathbf{g}\|}{2\sqrt{n}} + \frac{\sigma^2 \|\mathbf{g}\|}{2\tau\sqrt{n}} + \frac{1}{n} \sum_{i=1}^n \min_{-2 \leq \mathbf{w}_i \leq 0} \left\{ \frac{\|\mathbf{g}\|}{2\tau\sqrt{n}} \mathbf{w}_i^2 - \mathbf{h}_i \mathbf{w}_i \right\}. \quad (24)$$

In particular, the optimal $\tilde{\mathbf{w}}_i := \tilde{\mathbf{w}}_i(\mathbf{g}, \mathbf{h})$ of (22) satisfies

$$\tilde{\mathbf{w}}_i = \begin{cases} 0 & , \text{if } \mathbf{h}_i \geq 0, \\ \frac{\tilde{\tau}\sqrt{n}}{\|\mathbf{g}\|} \mathbf{h}_i & , \text{if } -\frac{2\|\mathbf{g}\|}{\tilde{\tau}\sqrt{n}} \leq \mathbf{h}_i < 0, \\ -2 & , \text{if } \mathbf{h}_i < -\frac{2\|\mathbf{g}\|}{\tilde{\tau}\sqrt{n}}. \end{cases} \quad (25)$$

where, $\tilde{\tau} := \tilde{\tau}(\mathbf{g}, \mathbf{h})$ is the solution to the following:

$$\phi(\mathbf{g}, \mathbf{h}) = \left(\min_{0 \leq \tau \leq T} \frac{\tau \|\mathbf{g}\|}{2\sqrt{n}} + \frac{\sigma^2 \|\mathbf{g}\|}{2\tau\sqrt{n}} + \frac{1}{n} \sum_{i=1}^n v_n \left(\frac{\tau\sqrt{n}}{\|\mathbf{g}\|}; \mathbf{h}_i \right) \right)_+, \quad (26)$$

with, v_n :

$$v_n(\alpha; h) := \begin{cases} 0 & , \text{if } h \geq 0, \\ -\frac{\alpha}{2} h^2 & , \text{if } -\frac{2}{\alpha} \leq h < 0, \\ \frac{2}{\alpha} + 2h & , \text{if } h \leq -\frac{2}{\alpha}, \end{cases}$$

for all $\alpha > 0$ and $h \in \mathbb{R}$. We remark that the minimization in (26) is convex. (The easiest way to see this is noting that the objective function in (24) is jointly convex in $\mathbf{w}$ and τ.)

4) Convergence properties of the (AO): Now that the (AO) is simplified as in (26), we study here its behavior in the limit of $m, n \to \infty$ with $m/n = \delta$.

First, we compute the point-wise (in τ) limit of the objective function in (26). Clearly,

$$\|\mathbf{g}\|/\sqrt{n} \xrightarrow{P} \sqrt{\delta}. \quad (27)$$

Also, conditioned on the value of $n^{-1/2}\|\mathbf{g}\|$, the random variable $\sum_{i=1}^n v_n(\tau\sqrt{n}/\|\mathbf{g}\|; \mathbf{h}_i)$ is a sum of absolutely integrable iid random variables. Hence, combining the WLLN with (27) it follows that, for all $\tau > 0$,

$$\frac{1}{n} \sum_{i=1}^n v_n \left(\frac{\tau\sqrt{n}}{\|\mathbf{g}\|}; \mathbf{h}_i \right) \xrightarrow{P} Y \left(\frac{\tau}{\sqrt{\delta}} \right)$$

where,

$$Y(\alpha) := -\frac{\alpha}{2} \int_0^{\frac{2}{\alpha}} h^2 p(h) \mathrm{d}h + \frac{2}{\alpha} Q \left(\frac{2}{\alpha} \right) - 2 \int_{\frac{2}{\alpha}}^\infty h p(h) \mathrm{d}h$$

$$= -\frac{\alpha}{4} + \frac{\alpha}{2} \int_{\frac{2}{\alpha}}^\infty \left(h - \frac{2}{\alpha} \right)^2 p(h) \mathrm{d}h. \quad (28)$$

Next, the point-wise convergence implies uniform convergence, thanks to convexity. This follows from [32, Cor.. II.1], which is also known in the literature of estimation theory as the convexity lemma: point wise convergence of convex functions implies uniform convergence in compact subsets (see also [33, Lem. 7.75]). Hence, the random optimization in (26) converges to the following deterministic optimization (for convenience we rescale the optimization variable τ as follows: $\tau := \frac{\tau}{\sqrt{\delta}}$):

$$\overline{\phi} := \min_{0 \leq \tau \leq (T/\sqrt{\delta})} \frac{\tau\delta}{2} + \frac{\sigma^2}{2\tau} + Y(\tau). \quad (29)$$

Expanding the square in the second summand in (28) and applying integration by parts, it can be checked that the objective function in (29) is exactly $2F(\tau)$, where $F(\tau)$ is defined in (4).

When $\delta > 1/2$, all summands in the objective function in (29) are non-negative for all $\tau > 0$. Thus, $\overline{\phi} \geq 0$, and consequently (recall (26)),

$$\phi(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \overline{\phi}. \quad (30)$$

We remark that the objective function in (29) is strictly convex in the optimization variable τ. (Its convexity follows directly as it is the point-wise limit of convex functions in (26), which is known to be convex. Alternatively, and to further check strict convexity, it can be shown that the second derivative is positive.) Hence, there is a unique minimizer, call it τ_*. With these, it only takes a standard argument (e.g., see [34, Thm. 2.1]) to further conclude that the minimizer $\tilde{\tau}(\mathbf{g}, \mathbf{h})$ of (26) converges in probability to $\tau_* \sqrt{\delta}$, i.e.

$$\delta^{-1/2} \tilde{\tau}(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \tau_*. \quad (31)$$

5) The optimal solution of the (AO): We now have all the tools necessary to study the properties of the optimal solution $\tilde{\mathbf{w}}$ of the (AO). The lemma below establishes that for Lipschitz functions, $\tilde{\mathbf{w}} \in \mathcal{S}$ (recall the definition of $\mathcal{S}$ in (20)).

Lemma IV.1 (Lipschitz convergence of the (AO)). *Let $\psi : \mathbb{R} \to \mathbb{R}$ be L-Lipschitz, $\tilde{\mathbf{w}} = \tilde{\mathbf{w}}(\mathbf{g}, \mathbf{h})$ as in (25), and random variable W as in the statement of Theorem IV.1. It holds, $\frac{1}{n} \sum_{i=1}^n \psi(\tilde{\mathbf{w}}_i) \xrightarrow{P} \mathbb{E}_W[\psi(W)]$.*

Proof. For $i = 1, \ldots, n$, define $\mathbf{v}_i := \theta(\mathbf{h}_i)$ (recall the definition of θ in the statement of Theorem IV.1). The WLLN gives

$$n^{-1} \sum_{i=1}^n \psi(\mathbf{v}_i) \xrightarrow{P} \mathbb{E}_{\gamma \sim \mathcal{N}(0,1)}[\psi(\theta(\gamma))] = \mathbb{E}_W[\psi(W)], \quad (32)$$

where we also used the Gaussianity of $\mathbf{h}_i$ and (18). Hence, it will sufficec for the proof to show that $|n^{-1} \sum_{i=1}^n (\psi(\tilde{\mathbf{w}}_i) - \psi(\mathbf{v}_i))| \xrightarrow{P} 0$. We show this using the Lipschitz assumption and (31). First, by the Lipschitz property:

$$|\psi(\tilde{\mathbf{w}}_i) - \psi(\mathbf{v}_i)| \leq L|\tilde{\mathbf{w}}_i - \mathbf{v}_i|. \quad (33)$$

Next, the expression of $\tilde{\mathbf{w}}$ in (25), along with (27) and with (31), they can be used to show that the RHS in (33) is appropriately small. Formally, writing $\xi := \xi(\mathbf{g}, \mathbf{h}) = \frac{\tilde{\tau}\sqrt{n}}{\|\mathbf{g}\|}$ for simplicity, it follows from the continuous mapping theorem

that for some $\eta > 0$ (the value of which to be chosen later) we have w.p.a.1: $|\xi - \tau_*| \leq \eta$, and, $|\frac{2}{\xi} - \frac{2}{\tau_*}| \leq \eta$. Hence, w.p.a.1:

$$|\tilde{\mathbf{w}}_i - \mathbf{v}_i| \leq \max \Big\{ |\tau_* - \xi| |\mathbf{h}_i| \mathbb{1}_{\{\mathbf{h}_i \geq \max\{-2/\tau_*, -2/\xi\}\}},$$
$$|\tau_* \mathbf{h}_i + 2| \mathbb{1}_{\{-2/\tau_* \leq \mathbf{h}_i \leq -2/\xi\}},$$
$$|\xi \mathbf{h}_i + 2| \mathbb{1}_{\{-2/\xi \leq \mathbf{h}_i \leq -2/\tau_*\}} \Big\}$$
$$\leq \eta(\eta + 2/\tau_*) + \eta + \eta(\eta + \tau_*).$$

For any $\zeta > 0$, choose $\eta = \min\{\frac{\sqrt{\zeta}}{2} , \frac{\zeta}{4}(\frac{1}{\tau_*} + \frac{1+\tau_*}{2})\}$, such that in view of (33) $|\psi(\tilde{\mathbf{w}}_i) - \psi(\mathbf{v}_i)| \leq L\zeta$, which completes the proof. $\square$

6) Satisfying the conditions of the CGMT: The following result uses Lemma IV.1 and strong-convexity of the (AO) to show that the optimal cost of the (AO) strictly increases when the optimization is constrained outside the set $\mathcal{S}$ defined in (20). The proof is deferred to Appendix B3.

Lemma IV.2 (Strong convexity of the (AO)). *Let $\psi : \mathbb{R} \to \mathbb{R}$ be L-Lipschitz, W a random variable as in the statement of Theorem IV.1, and $\mathcal{S} := \mathcal{S}_\epsilon$ the set defined in (20). Finally, denote $f(\mathbf{w}) := f(\mathbf{w}; \mathbf{g}, \mathbf{h})$ the objective function in (23). There exists constant $C > 0$, such that the following statement holds w.p.a.1,*

$$\min_{\substack{\mathbf{w} \in [-2,0]^n \\ \mathbf{w} \in \mathcal{S}^c}} f(\mathbf{w}; \mathbf{g}, \mathbf{h}) \geq \phi(\mathbf{g}, \mathbf{h}) + \frac{C\epsilon}{L}.$$

The lemma above essentially verifies conditions (i) and (ii) of the CGMT Theorem IV.2. To be specific, let C as in the statement of Lemma IV.2, $\bar{\phi}$ as in (29), and, choose $\eta := \frac{C\epsilon}{3L}$. From (30) it holds w.p.a.1: $|\phi(\mathbf{g}, \mathbf{h}) - \bar{\phi}| \leq \eta$. Combine this with Lemma IV.2 to conclude that $\phi_{\mathcal{S}^c}(\mathbf{g}, \mathbf{h}) \geq \bar{\phi} + 2\eta$ w.p.a.1, as desired.

7) Completing the proof: At the end of last section we showed that the conditions of the CGMT Theorem IV.2 are satisfied. Hence, its application yields that any minimizer $\hat{\mathbf{w}}$ of the (PO) in (16) satisfies $\hat{\mathbf{w}} \in \mathcal{S}_\epsilon$ w.p.a.1. This proves part (b) of Theorem IV.1. It remains to prove Part (a). Recall the note in Footnote 2: it suffices to prove that

$$\frac{1}{n} \sum_{i=1}^n \psi(\hat{\mathbf{w}}_i) \xrightarrow{P} \mathbb{E}_W[\psi(W)], \qquad (34)$$

for all continuous functions with compact support. Of course, the statement in (34) is true for Lipschitz continuous functions from part (b) of the theorem. But, continuous compactly supported functions are also bounded. The implication from Lipschitz bounded functions to continuous bounded functions is standard and is part of what is known in the literature as the Portmanteau Theorem; see for example [35, Thm. 13.16].

V. DISCUSSION AND FUTURE WORK

In this paper we have used the recently developed CGMT framework in [15, 16] to precisely compute the large-system error-rate performance of the popular box-relaxation optimization method for recovering signals from M-ary constellations, when the channel matrix and additive noise are both iid real Gaussians. The derived formulas were previously unknown. Also, the CGMT was previously only used to analyze squared-error performance; here, we illustrate for the first time its use to analyze the error-rate performance of convex optimization-based massive MIMO decoders.

In future work, we seek to extend the analysis to complex Gaussian channels with symbols originating from complex-valued constellations. At its core, this task requires extending the CGMT to complex-valued Gaussian matrices, an extension that is currently unavailable; thus, it poses a challenging, yet practically important, research direction. What appears more accessible is establishing the universality of our results for iid channels beyond Gaussians. We believe that this is possible by combining the ideas of our paper for extended use of the CGMT with the techniques in [36], where the universality property has been proven for the squared-error (rather than for the symbol-error-rate).

For BPSK signal recovery using the BRO, we proved in Corollary IV.1 that error events for any fixed number of bits in the solution of the BRO are iid. This fact has potentially significant consequences to be explored. For example, it implies that, when a block of data is in error, only a few of its bits are. This means that the output of the BRO can be used by various local methods to further reduce the SER. We are planning to explore such implications further in future work.

REFERENCES

[1] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser mimo systems," *Communications, IEEE Transactions on*, vol. 61, no. 4, pp. 1436–1449, 2013.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[3] C.-K. Wen, J.-C. Chen, K.-K. Wong, and P. Ting, "Message passing algorithm for distributed downlink regularized zero-forcing beamforming with cooperative base stations," *Wireless Communications, IEEE Transactions on*, vol. 13, no. 5, pp. 2920–2930, 2014.

[4] T. L. Narasimhan and A. Chockalingam, "Channel hardening-exploiting message passing (chemp) receiver in large-scale mimo systems," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 8, no. 5, pp. 847–860, 2014.

[5] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric algorithms and combinatorial optimization*. Springer Science & Business Media, 2012, vol. 2.

[6] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell labs technical journal*, vol. 1, no. 2, pp. 41–59, 1996.

[7] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm i. expected complexity," *Signal Processing, IEEE Transactions on*, vol. 53, no. 8, pp. 2806–2818, 2005.

[8] D. Guo and S. Verdú, "Multiuser detection and statistical mechanics," *KLUWER INTERNATIONAL SERIES IN ENGINEERING AND COMPUTER SCIENCE*, pp. 229–278, 2003.

[9] P. H. Tan, L. K. Rasmussen, and T. J. Lim, "Constrained maximum-likelihood detection in cdma," *Communications, IEEE Transactions on*, vol. 49, no. 1, pp. 142–153, 2001.

[10] A. Yener, R. D. Yates, and S. Ulukus, "Cdma multiuser detection: A nonlinear programming approach," *Communications, IEEE Transactions on*, vol. 50, no. 6, pp. 1016–1024, 2002.

[11] 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 with application to synchronous cdma," *Signal Processing, IEEE Transactions on*, vol. 50, no. 4, pp. 912–922, 2002.

[12] S. Verdu, *Multiuser detection*. Cambridge university press, 1998.

[13] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Transactions on Information Theory*, vol. 48, no. 11, pp. 2888–2910, Nov 2002.

[14] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1983–2010, 2005.

[15] C. Thrampoulidis, S. Oymak, and B. Hassibi, "Regularized linear regression: A precise analysis of the estimation error," in *Proceedings of The 28th Conference on Learning Theory, 2015*, 2015.

[16] C. Thrampoulidis, E. Abbasi, and B. Hassibi, "Precise error analysis of regularized M-estimators in high-dimensions," *arXiv preprint*, 2016.

[17] C. Thrampoulidis, E. Abbasi, W. Xu, and B. Hassibi, "Ber analysis of the box relaxation for bpsk signal recovery," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 3776–3780.

[18] I. B. Atitallah, C. Thrampoulidis, A. Kammoun, T. Al-Naffouri, B. Hassibi, and M.-S. Alouini, "Ber analysis of regularized least squares for bpsk recovery," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017.

[19] I. B. Atitallah, C. Thrampoulidis, A. Kammoun, T. Y. Al-Naffouri, M.-S. Alouini, and B. Hassibi, "The box-lasso with application to gssk modulation in massive mimo systems," in *Information Theory (ISIT), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 1082–1086.

[20] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimality of large mimo detection via approximate message passing," in *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 1227–1231.

[21] R. Ghods, C. Jeon, A. Maleki, and C. Studer, "Optimal large-mimo data detection with transmit impairments," in *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on*. IEEE, 2015, pp. 1211–1218.

[22] C. Jeon, A. Maleki, and C. Studer, "On the performance of mismatched data detection in large mimo systems," in *Information Theory (ISIT), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 180–184.

[23] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[24] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[25] D. Amelunxen, M. Lotz, M. B. Mccoy, and J. A. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference*, vol. 3, no. 3, pp. 224–294, Sep. 2014.

[26] T. Tao, *Topics in random matrix theory*. American Mathematical Society, 2012, vol. 132, Graduate Studies in Mathematics.

[27] Y. Gordon, "Some inequalities for gaussian processes and applications," *Israel Journal of Mathematics*, vol. 50, no. 4, pp. 265–289, 1985.

[28] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *arXiv preprint arXiv:1011.3027*, 2010.

[29] Y. Gordon, "On milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$," 1988.

[30] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: isoperimetry and processes*. Springer, 1991, vol. 23.

[31] M. Stojnic, "A framework to characterize performance of lasso algorithms," *arXiv preprint arXiv:1303.7291*, 2013.

[32] P. K. Andersen and R. D. Gill, "Cox's regression model for counting processes: a large sample study," *The annals of statistics*, pp. 1100–1120, 1982.

[33] K.-J. Miescke and F. Liese, *Statistical Decision Theory: Estimation, Testing, and Selection*. Springer New York, 2008.

[34] W. K. Newey and D. McFadden, "Large sample estimation and hypothesis testing," *Handbook of econometrics*, vol. 4, pp. 2111–2245, 1994.

[35] A. Klenke, *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.

[36] S. Oymak and J. A. Tropp, "Universality laws for randomized dimension reduction, with applications," *arXiv preprint arXiv:1511.09433*, 2015.

APPENDIX

A. Supplementary proofs for Section II

1) Corollary II.1: The corollary follows from Theorem II.1 when combined with the following statement, which we prove here:"If $\mathrm{SER}(\mathbf{A}, \mathbf{z}) \xrightarrow{P} c$, for some deterministic constant c, then, $P_e \to c$. "

For convenience, let us define the random variable $X := X(\mathbf{A}, \mathbf{z}) := \mathrm{SER}(\mathbf{A}, \mathbf{z})$. With this notation, $P_e = \mathbb{E}_{\mathbf{A}, \mathbf{z}}[X]$. Thus, for any $\epsilon > 0$,

$$
\begin{aligned}
P_e \leq\ & \mathbb{E}\left[X \mid |X - c| \leq \epsilon\right] + \\
& \mathbb{E}\left[X \mid |X - c| > \epsilon\right] \cdot \mathbb{P}\left(|X - c| > \epsilon\right). \\
\leq\ & (c + \epsilon) + \mathbb{P}(|X - c| > \epsilon),
\end{aligned}
$$

where in the second inequality we used the fact that $X \leq 1$. Notice that $(c + \epsilon) + P(|X - c| > \epsilon) \to (c + \epsilon)$ as $n \to \infty$, since $X \xrightarrow{P} c$, by assumption. In a similar vein,

$$
\begin{aligned}
P_e \geq\ & \mathbb{E}\left[X \mid |X - c| \leq \epsilon\right] \cdot \mathbb{P}\left(|X - c| \leq \epsilon\right) \\
\geq\ & (c - \epsilon) \cdot \mathbb{P}(|X - c| > \epsilon),
\end{aligned}
$$

where, again, $(c - \epsilon)\mathbb{P}(|X - c| \leq \epsilon) \to (c - \epsilon)$ as $n \to \infty$, since $X \xrightarrow{P} c$. Since the above hold for all ϵ, we have shown that $P_e \to c$, as desired.

2) Proof of Theorem II.2: Here, we prove the first part of the theorem, namely the lower and upper bounds on $Q(1/\tau_*)$. The tightness of the upper bound at high-SNR is shown later in Section A3. Due to the decreasing nature of the function Q, it suffices to prove that

$$
\sqrt{(\delta - 1/2) \cdot \mathrm{SNR}} < \tau_*^{-1} < \sqrt{\delta \cdot \mathrm{SNR}}. \tag{35}
$$

This is shown in Lemma A.2(b) below. The proof of the lemma builds on understanding the behavior of the function $F(\tau)$ in (4). The function F is composed of 4 additive terms. The first is linear in τ and the second is simply $1/\tau$. We view the remaining terms as a single function of τ, namely $S_2(\tau) := \left(\tau + \frac{4}{\tau}\right)Q\left(\frac{2}{\tau}\right) - \sqrt{\frac{2}{\pi}}e^{-\frac{2}{\tau^2}}$, and we gather its properties in Lemma A.1 below.

Lemma A.1. *Fix a positive integer $\ell > 0$ and consider the function $S : (0, \infty) \to \mathbb{R}$ defined as follows:*

$$
S_\ell(\alpha) := S(\alpha; \ell) := \left(\alpha + \frac{\ell^2}{\alpha}\right)Q\left(\frac{\ell}{\alpha}\right) - \frac{1}{\sqrt{2\pi}}\ell e^{-\frac{\ell^2}{2\alpha^2}}. \tag{36}
$$

The following statements are true.

(a) The first two derivatives $S'_\ell(\alpha)$ and $S''_\ell(\alpha)$ are given as follows

$$S'_\ell(\alpha) = \frac{\ell}{\alpha\sqrt{2\pi}}e^{-\frac{\ell^2}{2\alpha^2}} + \left(1 - \frac{\ell^2}{\alpha^2}\right)Q\left(\frac{\ell}{\alpha}\right). \quad (37)$$

$$S''_\ell(\alpha) = 2\frac{\ell^2}{\alpha^2}Q\left(\frac{\ell}{\alpha}\right).$$

(b) The function $S_\ell(\alpha)$ is strictly convex.

(c) The derivative $S'_\ell(\alpha)$ is strictly increasing. Moreover

$$\lim_{\alpha\to 0^+} S_\ell(\alpha) = 0 < S'_\ell(\alpha) < \frac{1}{2} = \lim_{\alpha\to+\infty} S_\ell(\alpha).$$

Proof. Statement (a) follows easily by direct calculations. It can be readily observed that $S''_\ell(\alpha)$ is strictly greater than 0 for all $\alpha > 0$. This proves statement (b). For the last statement, we argue as follows: $S'_\ell(\alpha)$ is strictly increasing by strict convexity of $S_\ell(\alpha)$. Thus, it suffices to compute the limits of $S'_\ell(\alpha)$ at 0 and $+\infty$. Easily,

$$\lim_{\alpha\to+\infty} S'_\ell(\alpha) = \lim_{\alpha\to+\infty} Q(\ell/\alpha) = 1/2.$$

For the limit $\alpha \to 0^+$, use the following facts: (i) in the limit of $x \to +\infty$: $Q(x) \sim p(x)/x$, and, (ii) $\lim_{x\to+\infty} xe^{-x^2/2} = 0$, to conclude with the desired. $\square$

Observe that $F(\tau)$ in (4) can be written as

$$F(\tau) = \tau\left(\delta - \frac{1}{2}\right) - \frac{1/SNR}{\tau} + S(\tau; 2). \quad (38)$$

We are now ready to state and prove Lemma A.2.

Lemma A.2. *[Properties of τ_*] Let τ_* be defined as in Theorem II.1, i.e. the (unique) positive minimizer of the function $F(\tau)$ in (4). The following hold.*

(a) τ_* is the unique positive solution of the equation

$$\delta - \frac{1}{2} - \frac{1/SNR}{\tau^2} + G(\tau^{-1}) = 0, \quad (39)$$

where

$$G(u) := \sqrt{(2/\pi)}ue^{-2u^2} + (1 - 4u^2)\cdot Q(2u). \quad (40)$$

(b) τ_* satisfies (35).

Proof. Recall from Theorem II.1 that the function $F(\tau)$ in (4) is strictly convex. Hence, τ_* is the unique positive solution to the first-order optimality condition: $F'(\tau) := \frac{d}{d\tau}F(\tau) = 0$. It is convenient for the rest of the proof to define a function $H : (0, \infty) \to \mathbb{R}$ as follows:

$$H(u) := F'(u^{-1}).$$

Also, note from (37) that G in (40) satisfies

$$G(u) = S'_2(u^{-1}). \quad (41)$$

In particular, properties of G to be used later in the proof follow from Lemma A.1.

Starting with (38) and using Lemma A.1(a) and (41):

$$H(u) := \delta - \frac{1}{2} - \frac{u^2}{SNR} + \underbrace{\sqrt{\frac{2}{\pi}}ue^{-2u^2} + (1 - 4u^2)Q(2u)}_{:=G(u)}.$$

This proves the first statement. Moreover, since $F(\tau)$ is strictly convex, we have that $F'(\tau)$ is strictly increasing, and equivalently that $H(u)$ is a decreasing function of u.

Next, we prove that,

$$\tau_*^{-1} \geq \sqrt{(\delta - 1/2)SNR} =: \tau_0^{-1}. \quad (42)$$

From Lemma A.1(c) and (41),

$$G(u) > 0, \quad \text{for all } u > 0.$$

Hence, $H(\tau_0^{-1}) = G(\tau_0^{-1}) > 0$. But, $H(u)$ is decreasing and τ_*^{-1} is its unique zero, from which (42) follows.

Finally, we show that

$$\tau_*^{-1} < \sqrt{\delta \cdot SNR} := \tau_1^{-1}. \quad (43)$$

Note that,

$$H(\tau_1^{-1}) = -\frac{1}{2} + G(\tau_1^{-1}).$$

Again, from Lemma A.1(c) and (41), it follows that $G(u) < 1/2$. Therefore, $H(\tau_1^{-1}) < 0$. Combine this with the fact that $H(u)$ is decreasing and τ_*^{-1} is its unique zero, to conclude with (43), as desired. $\square$

3) High-SNR regime: Theorem A.1 below formalizes and proves (6).

Theorem A.1 (High-SNR regime)**.** *As in the statement of Theorem II.1, fix $\delta \in (\frac{1}{2}, \infty)$ and let SER denote the bit error probability of the detection scheme in (1) for some fixed but unknown BPSK signal $\mathbf{x}_0 \in \{\pm 1\}^n$. For any $\epsilon > 0$, there exists constant $\overline{SNR} := \overline{SNR}(\epsilon)$ such that for all values $SNR > \overline{SNR}$, it holds*

$$\lim_{\substack{m,n\to\infty \\ m/n\to\delta}} \mathbb{P}\left(\left|\frac{SER}{Q\left(\sqrt{(\delta - 1/2)SNR}\right)} - 1\right| > \epsilon\right) = 0.$$

Proof. Fix any $\epsilon > 0$. Recall $\tau_* := \tau_*(SNR)$, the minimizer of (4), and define for convenience:

$$\tau_0 := \tau_0(SNR) = \left(\sqrt{(\delta - 1/2)SNR}\right)^{-1}. \quad (44)$$

We will prove that there exists $\overline{SNR}(\epsilon)$, such that

$$\left|\frac{Q\left(\tau_*^{-1}\right)}{Q\left(\tau_0^{-1}\right)} - 1\right| \leq \frac{\epsilon}{2}, \quad (45)$$

for all $SNR \geq \overline{SNR}(\epsilon)$. This would suffice to complete the proof of the theorem. To see this, write

$$\left|\frac{SER}{Q(\tau_0^{-1})} - 1\right| = \left|\frac{SER - Q(\tau_*^{-1})}{Q(\tau_0^{-1})} + \frac{Q(\tau_*^{-1})}{Q(\tau_0^{-1})} - 1\right|$$

$$\leq \frac{|SER - Q(\tau_*^{-1})|}{Q(\tau_0^{-1})} + \left|\frac{Q(\tau_*^{-1})}{Q(\tau_0^{-1})} - 1\right|,$$

and observe the following. (a) The last term above is further upper bounded by $\epsilon/2$ using (45) for large enough $SNR > \overline{SNR}(\epsilon)$. (b) From Theorem II.1, for all values of SNR, there exist large enough m, n such that the nominator of the first term is upper bounded by $(\epsilon/2)Q(\tau_0^{-1})$ with probability 1.

In what follows, we show (45), which is a deterministic statement about the minimizer $\tau_* := \tau_*(SNR)$ of (4). We use Lemma A.2.

From (35), we have that

$$\lim_{\text{SNR}\to+\infty} \tau_*^{-1} = +\infty. \tag{46}$$

Also, recall from (44) that $(\delta - 1/2) = \frac{\tau_0^{-2}}{\text{SNR}}$. Substituting this in (39) we find that

$$0 \le \tau_*^{-2} - \tau_0^{-2} = \text{SNR} \cdot G(\tau_*^{-1}) \tag{47}$$

for G as in (40) (also, recall (41)). The non-negativity above follows from the lower bound in (35). From Lemma A.1(c) and (41), G is decreasing in $(0, \infty)$. Using this, and applying the lower bound in (35) once more, (47) leads to the following:

$$0 \le \tau_*^{-2} - \tau_0^{-2} \le \text{SNR} \cdot G(\tau_0^{-1}) = \text{SNR} \cdot G(\sqrt{(\delta - 1/2)\text{SNR}}). \tag{48}$$

But, from Lemma A.1(c) the limit of the right-hand side as $\text{SNR} \to +\infty$ is equal to 0. Combining,

$$\lim_{\text{SNR}\to+\infty} (\tau_*^{-2} - \tau_0^{-2}) = 0. \tag{49}$$

Next, write $\tau_*^{-2} - \tau_0^{-2} = \tau_*^{-2}(1 - \frac{\tau_*^2}{\tau_0^2})$ and combine (46) with (49) to further show that

$$\lim_{\text{SNR}\to+\infty} \frac{\tau_*}{\tau_0} = 1. \tag{50}$$

We are now ready to prove (45). For simplicity, we write $f(x) \sim g(x)$ instead of $\lim_{x\to+\infty} \frac{f(x)}{g(x)} = 1$. It is well known that $Q(x) \sim p(x)/x$. Therefore,

$$\frac{Q(\tau_*^{-1})}{Q(\tau_0^{-1})} \sim \frac{p(\tau_*^{-1})}{p(\tau_0^{-1})} \frac{\tau_0}{\tau_*} = \frac{\tau_0}{\tau_*} \exp\left(-\frac{\tau_*^{-2} - \tau_0^{-2}}{2}\right)$$
$$\sim 1,$$

where the second line follows from (49) and (50). $\quad\square$

B. Supplementary proofs for Section IV

1) From Lipschitz to the indicator function:

Lemma A.3 (Approximating the indicator). *Let μ be a continuous measure on the real line such that $c \in \mathbb{R}$ is a point of measure zero. Further let $\{\mu_n\}$ be a sequence of random measures indexed by n such that as $n \to +\infty$,*

$$\int \psi \mathrm{d}\mu_n \xrightarrow{P} \int \psi \mathrm{d}\mu,$$

for all Lipschitz functions $\psi : \mathbb{R} \to \mathbb{R}$. For the indicator function $\chi_c(\alpha) := \mathbb{1}_{\{\alpha \le c\}}$ it holds that,

$$\int \chi_c \mathrm{d}\mu_n \xrightarrow{P} \int \chi_c \mathrm{d}\mu.$$

Proof. Fix any $\epsilon, \zeta > 0$ and consider the random variable $X = \left|\int \chi_c \mathrm{d}\mu_n - \int \chi_c \mathrm{d}\mu\right|$. Note that is random since the measures μ_n are random. It will suffice to show that there exists N_* such that for all $n > N_*$: $\mathbb{P}(X > \epsilon) \le \zeta$.

Let $\eta > 0$, the exact value of which to be determined later, and, consider the following functions parametrized by η:

$$\overline{\psi}_\eta(\alpha) := \begin{cases} 1, & \alpha \le c \\ 1 - \frac{1}{\eta}(\alpha - c), & c \le \alpha \le c + \eta \\ 0, & \alpha \ge c + \eta, \end{cases}$$

and

$$\underline{\psi}_\eta(\alpha) := \begin{cases} 1, & \alpha \le c - \eta \\ -\frac{1}{\eta}(\alpha - c), & c - \eta \le \alpha \le c \\ 0, & \alpha \ge c. \end{cases}$$

These functions are both Lipschitz with Lipschitz constant $1/\eta$. Define, the random variable Y_η as

$$Y_\eta := \max\{\left|\int \overline{\psi}_\eta \mathrm{d}\mu_n - \int \overline{\psi}_\eta \mathrm{d}\mu\right|, \left|\int \underline{\psi}_\eta \mathrm{d}\mu_n - \int \underline{\psi}_\eta \mathrm{d}\mu\right|\}.$$

From the assumption of the lemma there is $N(\epsilon, \zeta, \eta)$ such that for all $n \ge N(\epsilon, \zeta, \eta)$:

$$\mathbb{P}(Y_\eta > \epsilon/2) \le \zeta. \tag{51}$$

Moreover, $\underline{\psi}_\eta(\alpha) \le \chi_c(\alpha) \le \overline{\psi}_\eta(\alpha)$. Thus,

$$X \le Y_\eta + \int |\overline{\psi}_\eta - \underline{\psi}_\eta| \mathrm{d}\mu \le Y_\eta + \mu\{[c - \eta, c + \eta]\}, \tag{52}$$

where for the second inequality we further used the fact that $|\overline{\psi}_\eta - \underline{\psi}_\eta|$ is upper bounded by 1 and has support $[c - \eta, c + \eta]$.

Finally, from continuity of μ and the fact that c is μ-measure zero, we can choose $\eta = \eta_*(\epsilon)$ such that

$$\mu\{[c - \eta, c + \eta]\} \le \epsilon/2. \tag{53}$$

Combining, (51)–(53), we conclude, as desired, that there is $N_* := N(\epsilon, \zeta, \eta_*(\epsilon))$ such that for all $n > N_*$ it holds

$$\mathbb{P}(X > \epsilon) \le \mathbb{P}(Y_\eta > \epsilon/2) \le \zeta.$$

$\quad\square$

2) Proof of Corollary IV.1: On the one hand, by Theorem IV.1(b), it holds for all $\ell = 1, \ldots, k$ that

$$\overline{\psi_\ell}(\hat{\mathbf{w}}) := n^{-1} \sum_{i=1}^n \psi_\ell(\hat{\mathbf{w}}_i) \xrightarrow{P} \mathbb{E}_{W_\ell}[\psi_\ell(W_\ell)].$$

On the other hand, for some constant $C > 0$

$$\left|\prod_{\ell=1}^k \overline{\psi_\ell}(\hat{\mathbf{w}}) - n^{-k} \sum_{1 \le i_1, \ldots, i_k \le n} \psi_1(\hat{\mathbf{w}}_{i_1}) \cdots \psi_k(\hat{\mathbf{w}}_{i_k})\right| \le \frac{C}{n}.$$

To see this, expand the product term on the left-hand side and use the boundedness of the functions ψ_ℓ.

Combining the above proves the first statement of the corollary. The second statement follows with the exact same argument starting from Theorem II.1 and observing that $\mathbb{1}_{\{\hat{\mathbf{w}}_{i_1} \le -1, \ldots, \hat{\mathbf{w}}_{i_k} \le -1\}} = \prod_{\ell=1}^k \mathbb{1}_{\{\hat{\mathbf{w}}_{i_\ell}\}}$.

3) Proof of Lemma IV.2: Denote, $\overline{\psi} := \mathbb{E}_W[\psi(W)]$. From Lemma IV.1, it holds w.p.a.1: $|\frac{1}{n} \sum_{i=1}^n \psi(\tilde{\mathbf{w}}_i) - \overline{\psi}| \le \epsilon/2$. Hence, by definition of the set $\mathcal{S}$ and the triangle inequality, it holds w.p.a.1 that for all $\mathbf{w} \in \mathcal{S}^c$: $|\frac{1}{n} \sum_{i=1}^n \psi(\mathbf{w}_i) - \frac{1}{n} \sum_{i=1}^n \psi(\tilde{\mathbf{w}}_i)| \ge \epsilon/2$. Then, the Lipschitz property of ψ guarantees that

$$\frac{\|\mathbf{w} - \tilde{\mathbf{w}}\|}{\sqrt{n}} \ge \frac{\epsilon}{2L}. \tag{54}$$

In what follows we show that $n \cdot f(\mathbf{w})$ is C-strongly convex for appropriate constant $C > 0$. In view of (54) and recalling $\phi(\mathbf{g}, \mathbf{h}) = f(\tilde{\mathbf{w}})$, this will suffice to complete the proof.

It can be checked that the Hessian $\nabla^2 f(\mathbf{w})$ satisfies $n\nabla^2 f(\mathbf{w}) \succcurlyeq \frac{\|\mathbf{g}\|_2}{\sqrt{n}} \frac{\sigma^2}{\sqrt{\frac{\|\mathbf{w}\|^2}{n} + \sigma^2}} \mathbf{I}$. Further use the fact that $\|\mathbf{g}\|_2/\sqrt{n} \geq \sqrt{\delta}/2$ w.p.a.1 and $\|\mathbf{w}\|^2 \leq 4n$, to conclude that w.p.a.1 F is $\frac{C}{n}$-strongly convex with $C := \frac{\sigma^2\sqrt{\delta}}{2\sqrt{\sigma^2+4}}$, or $f(\mathbf{w}) \geq f(\tilde{\mathbf{w}}) + \frac{C}{2}\frac{\|\mathbf{w}-\tilde{\mathbf{w}}\|}{\sqrt{n}}$.

C. Proof of Theorem III.1

The proof of the theorem requires repeating, mutatis mutandis, the line of arguments detailed in Section IV for the proof of Theorem II.1. We omit most of the details for brevity, and only show the necessary calculations that yield to function F_M in (13). The idea is the same as in Section IV: thanks to the CGMT, it suffices to analyze a corresponding Auxiliary Optimization (AO) instead of the original optimization in (11a). Repeating the steps in Section IV-E3, the corresponding (AO) becomes (compare to Eqn. (24)):

$$\min_{\tau \geq 0} \frac{\tau\|\mathbf{g}\|}{2\sqrt{n}} + \frac{\sigma^2\|\mathbf{g}\|}{2\tau\sqrt{n}} + \frac{1}{n}\sum_{i=1}^{n} \min_{\mathbf{x}_{0,i}^- \leq \mathbf{w}_i \leq \mathbf{x}_{0,i}^+} \left\{ \frac{\|\mathbf{g}\|}{2\tau\sqrt{n}}\mathbf{w}_i^2 - \mathbf{h}_i\mathbf{w}_i \right\},$$

where, as always $\mathbf{w} = \mathbf{x}_0 - \mathbf{x}$ denotes the "error-vector" and we further defined

$$\mathbf{x}_{0,i}^- := -(M-1) - \mathbf{x}_{0,i} \quad \text{and} \quad \mathbf{x}_{0,i}^+ := (M-1) - \mathbf{x}_{0,i}.$$

For simplicity in notation, further denote $A = \frac{\|\mathbf{g}\|}{\tau\sqrt{n}}$. Then, the optimal $\tilde{\mathbf{w}}_i := \tilde{\mathbf{w}}_i(\mathbf{g},\mathbf{h},\mathbf{x}_0)$ satisfies

$$\tilde{\mathbf{w}}_i = \begin{cases} \mathbf{x}_{0,i}^- & \text{, if } \mathbf{h}_i < A\mathbf{x}_{0,i}^-, \\ \frac{1}{A}\mathbf{h}_i & \text{, if } A\mathbf{x}_{0,i}^- \leq \mathbf{h}_i \leq A\mathbf{x}_{0,i}^+, \\ \mathbf{x}_{0,i}^+ & \text{, if } \mathbf{h}_i > A\mathbf{x}_{0,i}^+. \end{cases} \quad (55)$$

where, $\tilde{\tau} := \tilde{\tau}(\mathbf{g},\mathbf{h},\mathbf{x}_0)$ is the solution to the following:

$$\left(\min_{\tau>0} \frac{\tau\|\mathbf{g}\|}{2\sqrt{n}} + \frac{\sigma^2\|\mathbf{g}\|}{2\tau\sqrt{n}} + \frac{1}{n}\sum_{i=1}^{n} v_n\left(\frac{\tilde{\tau}\sqrt{n}}{\|\mathbf{g}\|}; \mathbf{h}_i, \mathbf{x}_{0,i}^-, \mathbf{x}_{0,i}^+ \right) \right)_+,$$

$$(56)$$

with

$$v_n(\alpha; h, \ell, u) := \begin{cases} \frac{1}{2\alpha}\ell^2 - h\ell & \text{, if } \alpha h < \ell, \\ -\frac{\alpha}{2}h^2 & \text{, if } \ell \leq \alpha h \leq u, \\ \frac{1}{2\alpha}u^2 - hu & \text{, if } \alpha h > u. \end{cases}$$

This is of course very similar to Equation (26). Next, we follow the same steps as in Section IV-E4 and study the convergence of the (AO) in (56). For the first two summands in (56), we use the fact that $\frac{\|\mathbf{g}\|}{\sqrt{n}} \xrightarrow{P} \sqrt{\delta}$. For the third summand, recall that each $\mathbf{x}_{0,i}$ takes values $\pm 1, \pm 3, \ldots, \pm(M-1)$ with equal probability $1/M$. Let $j = 1, 3, \ldots, M-1$ and denote,

$$\ell_j := (M-1) - j \quad \text{and} \quad u_j := (M-1) + j.$$

Then, the pairs $(\mathbf{x}_{0,i}^-, \mathbf{x}_{0,i}^+)$ take values $(-u_j, \ell_j)$ and $(-\ell_j, u_j)$ with equal probability $1/M$ each. With these, $\frac{1}{n}\sum_{i=1}^{n} v_n\left(\frac{\tau\sqrt{n}}{\|\mathbf{g}\|}; \mathbf{h}_i, \mathbf{x}_{0,i}^-, \mathbf{x}_{0,i}^+ \right) \xrightarrow{P} Y\left(\frac{\tau}{\sqrt{\delta}} \right)$, where

$$Y(\alpha) := \frac{1}{M} \sum_{j=1,3,\ldots,M-1} \mathbb{E}_{h\sim\mathcal{N}(0,1)}[v_n(\alpha; h, -u_j, \ell_j)]$$

$$+ \frac{1}{M} \sum_{j=1,3,\ldots,M-1} \mathbb{E}_{h\sim\mathcal{N}(0,1)}[v_n(\alpha; h, -\ell_j, u_j)]. \quad (57)$$

Simple calculations show that

$$\mathbb{E}_{h\sim\mathcal{N}(0,1)}[v_n(\alpha; h, \ell, u)] =$$
$$-\frac{\alpha}{2} + \frac{\alpha}{2}\int_{\frac{\ell}{\alpha}}^{\infty} (h - \frac{\ell}{\alpha})^2 p(h)\mathrm{d}h + \frac{\alpha}{2}\int_{\frac{u}{\alpha}}^{\infty} (h - \frac{u}{\alpha})^2 p(h)\mathrm{d}h.$$

For convenience, define (see also Lemma A.1)

$$S(\alpha; \ell) := \alpha\int_{\frac{\ell}{\alpha}}^{\infty} (h - \frac{\ell}{\alpha})^2 p(h)\mathrm{d}h$$

$$= \left(\alpha + \frac{\ell^2}{\alpha} \right) Q\left(\frac{\ell}{\alpha} \right) - \frac{1}{\sqrt{2\pi}}\ell e^{-\frac{\ell^2}{2\alpha^2}}. \quad (58)$$

Putting all these together with (57) and grouping terms we find that

$$Y(\alpha) = \frac{1}{M} \sum_{j=1,3,\ldots,M-3} \left(-\alpha + S(\alpha; \ell_j) + S(\alpha; u_j) \right)$$

$$+ \frac{1}{M}\left(-\frac{\alpha}{2} + S(\alpha; u_{M-1}) \right) = -\frac{\alpha}{2}\left(\frac{M-1}{M} \right)$$

$$+ \frac{1}{M} \sum_{j=1,3,\ldots,M-3} \{S(\alpha; \ell_j) + S(\alpha; u_j)\} + \frac{1}{M}S(\alpha; u_{M-1}).$$

Observe that $Y(\alpha)$ is nonnegative for $\alpha > 0$ as long as $\delta > \frac{M-1}{M}$. Therefore, we can repeat the technical arguments of Section IV-E4, to conclude that the random optimization in (56) converges to the following deterministic optimization (where, for convenience, we have rescaled the optimization variable τ as follows $\tau := \frac{\tau}{\sqrt{\delta}}$):

$$\min_{\tau>0} \frac{\tau\delta}{2} + \frac{\sigma^2}{2\tau} + Y(\tau). \quad (59)$$

The objective function in (59) can be identified with the function $F_M(\tau)$ in the statement of the theorem. From Lemma A.1(b) the second derivative of $F_M(\tau)$ is strictly positive for $\tau > 0$, hence (59) has a unique minimizer, which we denote τ_*. With arguments same as in the end of Section IV-E4, we can show that $\sqrt{\delta}\tilde{\tau}(\mathbf{g},\mathbf{h},\mathbf{x}_0) \xrightarrow{P} \tau_*$.

Finally, we sketch how all these leads to the desired, namely:

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{x}_i^* \neq \mathbf{x}_{0,i}\}} \xrightarrow{P} 2\left(1 - \frac{1}{M} \right)Q(\tau_*^{-1}).$$

First, consider the case: $\mathbf{x}_{0,i} \in \{\pm 1, \pm 3, \ldots, \pm(M-3)\}$. Then, the thresholding rule (11b) implies that there is an error iff $|\tilde{\mathbf{w}}_i| > 1$. Equivalently, in view of (55), and noting that $\mathbf{x}_{0,i}^+ \geq 2$, it follows that and error occurs iff $|\mathbf{h}_i| > A$. Next, consider the case(s) $\mathbf{x}_{0,i} = M-1$ (or, $\mathbf{x}_{0,i} = -(M-1)$). Then the error event corresponds to $\tilde{\mathbf{w}}_i < -1$ (or, $\tilde{\mathbf{w}}_i > 1$), which in view of (55) translates to $\mathbf{h}_i < -A$ (or $\mathbf{h}_i > A$). Putting these together and conditioning on the high-probability events $\|\mathbf{g}\|/\sqrt{n} \xrightarrow{P} \sqrt{\delta}$ and $\tilde{\tau} \xrightarrow{P} \tau_*$, we find that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\arg\min_{s\in\mathcal{C}} |\mathbf{x}_{0,i} + \tilde{\mathbf{w}}_i - s| \neq \mathbf{x}_{0,i}\}} \xrightarrow{P}$$

$$\frac{2}{M}\left((M-2)Q(\tau_*^{-1}) + Q(\tau_*^{-1}) \right) = 2\left(1 - \frac{1}{M} \right)Q(\tau_*^{-1}).$$