

Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/OJCOMS.2022.1234567

# Asymptotic Performance of Box-RLS Decoders under Imperfect CSI with Optimized Resource Allocation

Ayed M. Alrashdi\*, Member, IEEE, Abla Kammoun<sup>†</sup>, Member, IEEE, Ali H. Muqaibel<sup>‡</sup>, Senior Member, and Tareq Y. Al-Naffouri<sup>†</sup>, Senior Member, IEEE

<sup>1</sup>Department of Electrical Engineering, College of Engineering, University of Ha'il, P.O. Box 2440, Ha'il, 81441, Saudi Arabia

<sup>2</sup>Computer, Electrical, and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

<sup>3</sup> Department of Electrical Engineering, King Fahd University of Petroleum and Minerals, Dhahran 3126, Saudi Arabia

CORRESPONDING AUTHOR: Ayed M. Alrashdi (e-mail: am.alrashdi@uoh.edu.sa).

A preliminary version of this work was presented in [1]. This work was supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research under Award OSR-CRG2019-4041.

**ABSTRACT** This paper considers the problem of symbol detection in massive multiple-input multiple-output (MIMO) wireless communication systems. We consider hard-thresholding preceded by two variants of the regularized least squares (RLS) decoder; namely the unconstrained RLS and the RLS with a box constraint, which is called Box-RLS. For all schemes, we focus on the evaluation of the mean squared error (MSE) and the symbol error probability (SEP) for  $M$ -ary pulse amplitude modulation ( $M$ -PAM) symbols transmitted over a massive MIMO system when the channel is estimated using linear minimum mean squared error (LMMSE) estimator. Under such circumstances, the channel estimation error is Gaussian which allows for the use of the convex Gaussian min-max theorem (CGMT) to derive asymptotic approximations for the MSE and SEP when the system dimensions and the coherence duration grow large with the same pace. The obtained expressions are then leveraged to derive the optimal power distribution between pilot and data under a total transmit energy constraint. In addition, we derive an asymptotic approximation of the goodput for all schemes which is then used to jointly optimize the number of training symbols and their associated power. Numerical results are presented to support the accuracy of the theoretical results.

**INDEX TERMS** Resource allocation, channel estimation, mean squared error, symbol error probability, goodput, constrained least squares.

## I. INTRODUCTION

THE use of multiple-input multiple-output (MIMO) systems has been recognized as an efficient technology to meet the ever-increasing demand in spectral efficiency. It is indeed known since the early works of Telatar [2] and Foschini [3] that the mutual information scales with the minimum of the number of transmit and receive antennas. In practice, however, the spectral efficiency of a wireless link depends not only on how many antennas are deployed at the transmit and receive sides but also on the channel estimation accuracy, the detection procedure as well as the distribution of the power resources, all of which have a direct bearing on the end-to-end signal-to-noise-ratio. At the receiver side,

accurate channel estimation and symbol detection are crucial to reap the gains promised by the additional degrees of freedom offered by MIMO systems. Channel estimation is performed by allocating a training period during which the transmitter sends known *pilot* symbols to the receiver for it to acquire an estimate of the channel state information (CSI) [4], [5]. During the data transmission phase, this estimate is leveraged by the receiver to equalize the channel and recover the data symbols. It is worth mentioning that it is important for the success of the data recovery step to acquire accurate channel estimates because otherwise the error in the channel estimation would propagate to the symbol recovery, crippling the overall performance even under state-of-the-art detection

strategies. Resource allocation in terms of power and time is also an essential part of the design of wireless systems. Increasing the duration of the transmitted pilot sequence would lead to enhanced channel estimation quality but at the cost of spectral efficiency losses since less time is spent to transmit useful data. Moreover, as the total power allocated to data and training transmission is fixed, we cannot increase the power allocated to data without affecting the channel accuracy estimation and vice versa.

The problem of finding the optimal power allocation between pilot and data has received a lot of attention over the last decades. It has been applied to different contexts with the aim of realizing various communication objectives. In [4], [6] and [7], the authors proposed power designs that optimizes bounds on the average channel capacity. A different line of research works in [8]–[11] considered the post-equalization signal to interference and noise ratio as a target metric to determine the optimal power allocation. Depending on the application at hand, different other metrics such as the bit error rate (BER) [12], [13], the symbol error rate (SER) [14], mean squared error (MSE)-related indexes [15], [16], max-min fairness utilities [17], energy and spectral efficiency [18], [19], or bounds on the received signal-to-noise ratio (SNR) [20] have been used in several existing papers. In [21], the optimal power allocation of a non-orthogonal multiple access (NOMA)-enabled device-to-device communication was proposed as a sum data rate maximization problem. In addition, the works in [22] and [23] considered power allocation problem for a cooperative THz MIMO-NOMA system, and a wireless-powered full-duplex cooperative NOMA system, respectively. Moreover, following the recent advances in deep learning/reinforcement learning, various power allocation schemes have been proposed, such as in [24]–[33]. Of interest in these works are a wide range of contexts including classical MIMO systems [4], mutli-carrier systems [34]–[39], amplify-and-forward relaying [12], cognitive radio systems [40] and very recently massive MIMO systems in both single-cell [41] and multi-user multi-cell settings [15], [17], [42]–[45].

Most of the aforementioned works, despite looking at the power allocation problem from different angles, present the common denominator of relying on linear receivers and optimizing some related performance metrics. To achieve optimal performance, it is well known that non-linear detectors are required, but they are often not implemented as they require a prohibitively high computation complexity. In this work, we consider optimization of the power allocation of a non-linear decoder coined Box-regularized least squares (Box-RLS) decoder, the idea behind which has been proposed in [46], [47] and earlier in [48]. As explained in [48], this decoder is rooted in the formulation of the maximum-likelihood (ML) problem. At the one hand, the ML decoder is known to be NP-hard since its solution is constrained to belong to a finite discrete set, at the other hand, the least squares (LS) decoder or its regularized version

referred to as regularized least squares (RLS) decoder [49] are linear decoders, obtained by solving the unconstrained ML problem. The Box-RLS decoder falls between the ML and the RLS decoders in that the solution is constrained to lie within a closed convex set. As a consequence, unlike the RLS decoder, the solution of the Box-RLS cannot be expressed in closed-form but is numerically computed using standard convex-optimization tools. Both the RLS and Box-RLS decoders are more computationally efficient than the ML decoder. Besides they are also shown to outperform many heuristic algorithms such as zero-forcing (ZF), successive interference cancellation and decision-feedback [46]–[48]. Performance analysis of the Box-RLS has been carried out in [46], [47], [50] but these studies assume unrealistic scenarios in which the channel is perfectly known. This lies behind the main motivation of this work. More specifically, to the best of our knowledge, this work is the first to account for the impact of channel estimation error in the study of Box-RLS decoders. Underlying our consideration of imperfect channel knowledge is the question of how to share available power between training and data transmissions under a fixed power budget. We show how our analysis can lead to addressing this question. More specifically, assuming a linear minimum mean square error estimator for channel estimation, we derive sharp approximations of several performance metrics of interest ranging from MSE, symbol error probability (SEP) to goodput for the RLS and Box-RLS decoders used to recover  $M$ -ary pulse amplitude modulation ( $M$ -PAM) signaling by applying hard-thresholding on their outputs. Our analysis, based on the assumption that the MIMO channel and the noise are independent and follow standard Gaussian distributions, builds upon the CGMT framework put forth in [51]–[53]. As compared to previous works dealing with the use of CGMT in high-dimensional regression problems, our consideration of imperfect CSI poses technical challenges towards assessing the performance of the Box-RLS. Particularly, contrary to previous studies, the application of the CGMT in the asymptotic regime leads to a non-convex optimization problem; the uniqueness of its solution which is an important step in the analysis becomes thus extremely challenging. This required us to develop new techniques to break these difficulties, which while being accommodated to this specific scenario, may be of independent interest. We refer the interested reader to Appendix B devoted to the detailed proofs of our main results.

To summarize, the main contributions of this work can be listed as follows:

- 1) We derive sharp characterizations of the MSE, SEP and goodput expressions for the RLS, LS and Box-RLS under imperfect CSI. Our expressions shed light on interesting relationships between MSE and SEP for  $M$ -PAM modulation and under imperfect CSI.
- 2) We determine the optimal power allocation between training and data symbols when SEP or MSE are used as target criteria.

- 3) We optimize the power and the number of pilot symbols to maximize the goodput for all studied decoders.

To the best of our knowledge, none of the above was previously derived for the  $M$ -PAM case in the presence of imperfect CSI.

### A. Paper ORGANIZATION

This paper is organized as follows. The system model is presented in Section II. In Section III, we discuss channel estimation, the properties of the estimator and the channel estimation error, as well as symbol estimation. The MSE/SEP of symbol estimation is derived in Section IV by applying the CGMT. These expressions are then validated through an assortment of numerical results and leveraged to find optimal power strategies in Section V. The key ingredient of the analysis which is the CGMT is reviewed in Appendix A. The proofs for the derived MSE and SEP are presented in Appendix B and Appendix C for the Boxed RLS and unboxed RLS, respectively.

### B. NOTATIONS

Scalars are denoted by lower-case letters (e.g.,  $\alpha$ ), column vectors are represented by boldface lowercase letters (e.g.,  $\mathbf{x}$ ), whereas matrices are denoted by boldface upper-case letters (e.g.,  $\mathbf{X}$ ). The notations  $(\cdot)^T$  and  $(\cdot)^{-1}$  denote the transpose and inversion operators, respectively. The  $j$ -th element of vector  $\mathbf{x}$  will be denoted by  $x_j$ . The symbol  $\mathbf{I}_N$  is used to represent the identity matrix of dimension  $N \times N$ . We use the standard notation  $\mathbb{P}[\cdot]$  and  $\mathbb{E}[\cdot]$  to denote probability and expectation. We write  $X \sim p_X$  to denote that a random variable  $X$  has a probability density function (pdf)  $p_X$ . In particular,  $G \sim \mathcal{N}(\mu, \sigma^2)$  implies that  $G$  has a Gaussian (normal) distribution of mean  $\mu$  and variance  $\sigma^2$ . Furthermore,  $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  and  $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$  denote the pdf of a standard normal distribution and its associated  $Q$ -function, respectively. Finally,  $\|\cdot\|$  indicates the Euclidean norm (i.e., the  $\ell_2$ -norm) of a vector and  $\|\cdot\|_\infty$  represents its  $\ell_\infty$ -norm.<sup>1</sup>

## II. System Model

We consider a flat block-fading massive MIMO system [54], [55] with  $K$  transmitter antennas and  $N$  receiver antennas. The transmission consists of  $T$  symbols that occur in a time interval within which the channel is assumed to be static. A number  $T_p$  pilot symbols (for channel estimation) occupy the first part of the transmission interval with power,  $\rho_p$ . The remaining part is devoted for transmitting  $T_d = T - T_p$  data symbols with power,  $\rho_d$ . Figure 1 illustrates the system model. It implies from conservation of time and energy that:

$$\rho_p T_p + \rho_d T_d = \rho T, \quad (1)$$

where  $\rho$  is the expected average power. Alternatively, we have

$$\rho_d T_d = \alpha \rho T, \quad (2)$$

<sup>1</sup>For a vector  $\mathbf{x}$ ,  $\|\mathbf{x}\|_\infty = \max_j |x_j|$ .

TABLE 1: Summary of Main Variables

| Symbol                        | Meaning   |
|-------------------------------|---|
| $T_p$                         | number of pilot symbols                         |
| $T_d$                         | number of data symbols                          |
| $T$                           | total number of symbols                         |
| $M$                           | size of the PAM signal                          |
| $\mathcal{E}$                 | average power of non-normalized $M$ -PAM signal |
| $\rho_p$                      | average training power                          |
| $\rho_d$                      | average data power                              |
| $\rho$                        | average received power per receive antenna      |
| $\rho_{\text{eff}}$           | effective SNR                                   |
| $\alpha$                      | data power ratio                                |
| $K$                           | Number of transmit (Tx) antennas                |
| $N$                           | Number of receive (Rx) antennas                 |
| $\delta$                      | ratio of Rx to Tx antennas $N/K$                |
| $\lambda$                     | regularization coefficient                      |
| $t$                           | box-constraint threshold                        |
| $\mathbf{H}$                  | channel matrix                                  |
| $\hat{\mathbf{H}}$            | estimated channel matrix                        |
| $\Delta$                      | channel estimation error matrix                 |
| $\sigma_\Delta^2$             | variance of error matrix                        |
| $\sigma_{\hat{\mathbf{H}}}^2$ | variance of estimated channel matrix            |
| $\mathbf{X}_p$                | pilot symbols matrix                            |
| $\mathbf{x}_0$                | transmitted data symbol vector                  |
| $\hat{\mathbf{x}}$            | estimated data symbol vector                    |
| $\mathbf{y}$                  | received data symbol vector                     |
| $\mathbf{z}$                  | noise vector                                    |

where  $\alpha \in (0, 1)$  is the ratio of the energy allocated to the data, so that  $\rho_p T_p = (1 - \alpha)\rho T$ . The received signal model for the *data* transmission phase is given by

$$\mathbf{y} = \sqrt{\frac{\rho_d}{K}} \mathbf{H} \mathbf{x}_0 + \mathbf{z}, \quad (3)$$

where  $\mathbf{y} \in \mathbb{R}^N$  is the received data symbol vector,  $\mathbf{x}_0 \in \mathbb{R}^K$  is the transmitted data symbol vector,  $\mathbf{H} \in \mathbb{R}^{N \times K}$  is a channel matrix with i.i.d. Gaussian elements  $h_{ij} \sim \mathcal{N}(0, 1)$ , and  $\mathbf{z} \in \mathbb{R}^N$  stands for the additive Gaussian noise at the receiver with i.i.d. elements of mean 0 and variance 1. It is assumed that  $\mathbf{x}_0$  has i.i.d.  $M$ -PAM symbols normalized to have unit variance ( $\mathbb{E}[\mathbf{x}_0 \mathbf{x}_0^T] = \mathbf{I}_K$ ), such that each transmit antenna sends a data symbol  $x_{0,j}$  that takes values (with equal probability  $1/M$ ) in the set:

$$x_{0,j} \in \mathcal{C} := \left\{ \pm \frac{1}{\sqrt{\mathcal{E}}}, \pm \frac{3}{\sqrt{\mathcal{E}}}, \dots, \pm \frac{(M-1)}{\sqrt{\mathcal{E}}} \right\}, j = 1, 2, \dots, K, \quad (4)$$

where  $\mathcal{E} = \frac{M^2-1}{3}$  is the average power of the non-normalized  $M$ -PAM signal,  $M = 2^b$  being the modulation order and  $b$  the number of bits carried by each symbol.

As the channel matrix  $\mathbf{H}$  is unknown to the receiver, a training phase during which the transmitter sends  $T_p \geq K$  *pilot* symbols is dedicated. The received signals correspond-

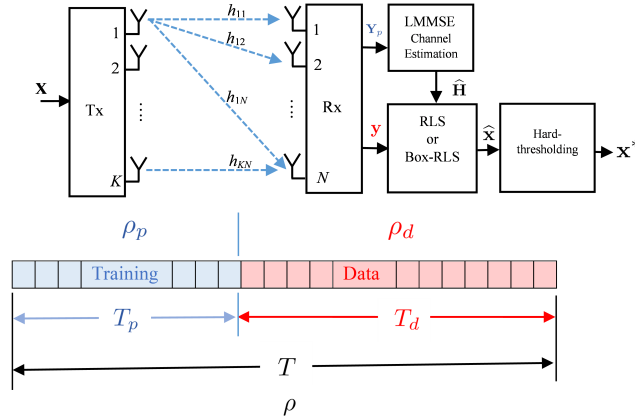


FIGURE 1: A massive MIMO system with a training-based transmission. Here, Tx is the transmitter and Rx is the receiver.

ing to this phase can be modeled as

$$\mathbf{Y}_p = \sqrt{\frac{\rho_p}{K}} \mathbf{H} \mathbf{X}_p + \mathbf{Z}_p, \quad (5)$$

where  $\mathbf{Y}_p \in \mathbb{R}^{N \times T_p}$  is the received signal matrix,  $\mathbf{X}_p \in \mathbb{R}^{K \times T_p}$  is the matrix of transmitted pilot symbols, and  $\mathbf{Z}_p \in \mathbb{R}^{N \times T_p}$  stands for the additive Gaussian noise with  $\mathbb{E}[\mathbf{Z}_p \mathbf{Z}_p^T] = T_p \mathbf{I}_N$ .

For the reader convenience, we summarize in Table 1 the notation symbols of the parameters used in this paper.

**Beyond point-to-point communications.** As an important note, we would like to mention that our results can be applied to studying the performance of uplink transmissions between multiple users and a base station (BS). In this case, both the BS and the users are perfectly synchronized: during pilot transmission, users send orthogonal pilot sequences; then the BS estimates the users' channels to use them later to decode the signals stemming from all users during data transmission.

### III. MIMO Symbols Detection under LMMSE Channel Estimation

#### A. LMMSE Channel Estimation

Based on the knowledge of  $\mathbf{Y}_p$  from (5), the LMMSE channel estimate is given by [56]

$$\begin{aligned} \hat{\mathbf{H}} &= \sqrt{\frac{K}{\rho_p}} \mathbf{Y}_p \mathbf{X}_p^T \left( \frac{K}{\rho_p} \mathbf{I}_K + \mathbf{X}_p \mathbf{X}_p^T \right)^{-1}, \\ &= \mathbf{H} - \mathbf{\Delta}, \end{aligned} \quad (6)$$

where  $\mathbf{\Delta}$  is the zero-mean channel estimation error matrix, which is independent of  $\hat{\mathbf{H}}$ , as per the orthogonality principle of the LMMSE estimation [4], [56]. For MIMO channels with i.i.d. entries, it has been proved that the optimal  $\mathbf{X}_p$  that minimizes the estimation mean squared error under a total power constraint satisfies [4]

$$\mathbf{X}_p \mathbf{X}_p^T = T_p \mathbf{I}_K. \quad (7)$$

For the above condition to hold, the number of training symbols should be greater than or equal to  $K$ . Moreover, under (7), the channel estimate  $\hat{\mathbf{H}}$  has i.i.d. zero-mean Gaussian entries with mean 0 and variance  $\sigma_{\hat{\mathbf{H}}}^2 = 1 - \sigma_{\mathbf{\Delta}}^2$  [4], with

$$\sigma_{\mathbf{\Delta}}^2 = \frac{1}{1 + \frac{\rho_p}{K} T_p}, \quad (8)$$

being the variance of each element in  $\mathbf{\Delta}$ . It appears from (8) that channel estimation error decreases with the pilot energy given by  $\rho_p T_p$ .

#### B. Symbol Detection under Imperfect Channel Estimation

With the channel estimate  $\hat{\mathbf{H}}$  at hand, the receiver can proceed to the recovery of the transmitted symbols. The optimal decoder which minimizes the probability of error is the maximum likelihood (ML) decoder under perfect channel knowledge which is given by:

$$\hat{\mathbf{x}}_{\text{ML}} = \arg \min_{\mathbf{x} \in \mathcal{C}^K} \left\| \mathbf{y} - \sqrt{\frac{\rho_d}{K}} \mathbf{H} \mathbf{x} \right\|^2. \quad (9)$$

As can be seen, the ML decoder involves a combinatorial optimization problem. It presents thus a prohibitively high computational complexity, especially when the system dimensions become large as envisioned by current communication systems. To overcome this issue, suboptimal strategies that require less computational complexity are in general used. Among these strategies are *linear decoders* that relax the discrete set constraint used in ML detection. They often proceed in two steps. First, a real-valued approximation of the transmitted symbol is obtained. This estimate is then hard-thresholded in a second step to produce the final estimate.

One of the simplest and commonly used linear decoders in massive MIMO systems is the LS decoder, which relaxes the discrete-alphabet constraint on  $\mathbf{x}$  to  $\mathbb{R}^K$ , i.e.,

$$\hat{\mathbf{x}}_{\text{LS}} = \arg \min_{\mathbf{x} \in \mathbb{R}^K} \left\| \mathbf{y} - \sqrt{\frac{\rho_d}{K}} \mathbf{H} \mathbf{x} \right\|^2. \quad (10)$$

However, the LS solution is known for its sensitivity to perturbations of the measurements vector  $\mathbf{y}$  or the channel matrix  $\mathbf{H}$ . This sensitivity is manifested by a significant variation of the solution upon any small perturbation of the measurement vector due to noise or imperfect knowledge of the channel matrix  $\mathbf{H}$ . To overcome these limitations, one frequently used approach is to use regularization techniques. The most common form of regularization is Tikhonov regularization [49], which is also known as RLS or ridge regression. In this work, we are interested in the study of two regularization-based decoders, namely RLS and Box-RLS decoders.

The RLS decoder is based on regularizing the cost in (9) and relaxing the finite-alphabet constraint, thus leading to:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^K} \left\| \mathbf{y} - \sqrt{\frac{\rho_d}{K}} \hat{\mathbf{H}} \mathbf{x} \right\|^2 + \lambda \rho_d \|\mathbf{x}\|^2, \quad (11a)$$

$$= (\mathbf{A}^T \mathbf{A} + \lambda \rho_d \mathbf{I}_K)^{-1} \mathbf{A}^T \mathbf{y}, \quad (11b)$$

$$x_j^* = \arg \min_{s \in \mathcal{C}} \left| \frac{\hat{x}_j}{B} - s \right|, j = 1, 2, \dots, K, \quad (11c)$$

where  $\lambda \geq 0$  is the regularization coefficient,  $\mathbf{A} = \sqrt{\frac{\rho_d}{K}} \hat{\mathbf{H}}$  and  $B$  is a normalization constant, the value of which will be suggested from our analysis so as to remove the bias of the decoder [57]<sup>2</sup>. Note that the optimization in (11c) simply selects the symbol value  $s$  that is closest to the solution  $\frac{\hat{x}_j}{B}$  among a total of  $M$  possible choices. As can be seen from (11b), the elements of the solution  $\hat{\mathbf{x}}$  may take large values in high-noise conditions or poor channel estimation scenarios. This motivates the Box-RLS decoder [48], [58], [59] given by

$$\hat{\mathbf{x}} = \arg \min_{\|\mathbf{x}\|_\infty \leq t} \left\| \mathbf{y} - \sqrt{\frac{\rho_d}{K}} \hat{\mathbf{H}} \mathbf{x} \right\|^2 + \lambda \rho_d \|\mathbf{x}\|^2, \quad (12a)$$

$$x_j^* = \arg \min_{s \in \mathcal{C}} \left| \frac{\hat{x}_j}{B} - s \right|, j = 1, 2, \dots, K, \quad (12b)$$

which is based on relaxing the finite-alphabet constraint to the convex constraint  $\mathbf{x} \in [-t, t]^K$ , where  $t > 0$  is a fixed threshold that can be optimally tailored according to the propagation scenario.

### Computational Complexity

Although the Box-RLS decoder does not possess an explicit formulation, it shares the same computational complexity cost as the RLS decoder. Indeed, both decoders rely on solving quadratic programs, which can be solved with a cubic complexity of  $K$ , that is  $O(K^3)$ . For the Box-RLS decoder, the underlying quadratic program can be solved using iterative methods such as the alternating direction method of multipliers (ADMM) algorithm. On the other hand, computing the explicit formulation of the RLS decoder requires inverting a matrix, an operation presenting similarly a cubic complexity.

### C. Performance Metrics

This work considers the performance evaluation of the RLS and the Box-RLS decoders in terms of three different performance metrics, which are:

**Mean Squared Error:** A natural and heavily used measure of performance is the reconstruction *mean squared error* (MSE), which measures the deviation of  $\hat{\mathbf{x}}$  from the true

<sup>2</sup>It follows from our analysis that  $\hat{x}_j$  behaves as  $B x_{0,j}$  plus some independent Gaussian noise, where  $B$  is some constant depending on the regularization factor, the data and channel estimation powers. It is thus sensible to divide  $\hat{x}_j$  by  $B$  to remove the induced bias.

signal  $\mathbf{x}_0$ . This assesses the performance of the *first* step of the decoding algorithms. Formally, the MSE is defined as

$$\text{MSE} := \frac{1}{K} \|\hat{\mathbf{x}} - \mathbf{x}_0\|^2. \quad (13)$$

**Symbol Error Probability:** The symbol error rate (SER) characterizes the performance of the detection process and is defined as:

$$\text{SER} := \frac{1}{K} \sum_{j=1}^K \mathbf{1}_{\{x_j^* \neq x_{0,j}\}}, \quad (14)$$

where  $\mathbf{1}_{\{\cdot\}}$  indicates the indicator function.

In relation to the SER is the symbol error probability (SEP) which is defined as the expectation of the SER averaged over the noise, the channel and the constellation. Formally, the symbol error probability denoted by SEP is given by:

$$\text{SEP} := \mathbb{E}[\text{SER}] = \frac{1}{K} \sum_{j=1}^K \mathbb{P}[x_j^* \neq x_{0,j}]. \quad (15)$$

**Goodput:** The goodput is a performance measure that accounts for the amount of useful data transmitted, divided by the time it takes to *successfully* transmit it. The amount of data considered excludes protocol overhead bits as well as retransmitted data packets [60]. In our context, it can be defined as

$$G := \left( \frac{T - T_p}{T} \right) (1 - \text{SEP}). \quad (16)$$

Goodput and throughput are connected performance parameters in that the throughput can be obtained by simply dividing the goodput by the data transmission rate.

## IV. Analysis of the Mean Squared Error (MSE) and Symbol-Error Probability (SEP)

In this section, we derive asymptotic expressions of the MSE and SEP for the RLS and Box-RLS decoders. Particularly, we show that these metrics can be approximated by deterministic quantities that involve the power and time devoted for data and training transmissions. Our analysis builds upon the CGMT framework. For the RLS decoder, the same results could have been obtained using tools from random matrix theory as the decoder possesses a closed-form expression. However, since the use of the CGMT framework is more adapted to the Box-RLS decoder that cannot be expressed in closed-form, we rely in this work on the CGMT framework for both decoders for the sake of a unified presentation.

Prior to stating our main results, we shall introduce the following assumptions which describe the considered growth rate regime:

### A. Technical Assumptions

**Assumption 1.** We consider the asymptotic regime in which the system dimensions  $K$  and  $N$  grow simultaneously to infinity at a fixed ratio

$$\delta := \frac{N}{K} \in (0, \infty).$$



*Assumption 2.* We assume a fixed normalized coherence interval

$$\tau := \frac{T}{K} \in (1, \infty).$$

and that the pilot and data symbols grow proportionally with  $K$ , where:

$$\tau_p := \frac{T_p}{K} \in [1, \infty),$$

and

$$\tau_d := \frac{T_d}{K},$$

are fixed and denote the normalized number of pilot and data symbols, respectively.

In the sequel, we leverage the statistical distribution of the channel and the channel estimate as well as the asymptotic regime specified in Assumption 1 and 2 to provide asymptotic approximations of the MSE and SEP for RLS and Box-RLS. We use the standard notation  $\text{plim}_{n \rightarrow \infty} X_n = X$  to denote that a sequence of random variables  $X_n$  converges in probability towards a constant  $X$ .

### B. MSE and SEP Analysis for RLS

We provide herein asymptotic approximations of the MSE and SEP for RLS under imperfect channel state information. The derived closed form expressions are given in Theorem 1, and Theorem 2 while the proof is given in Appendix C.

**Theorem 1.** (*MSE for RLS*): Fix  $\lambda > 0$ ,  $\delta > 0$ , and let  $\hat{\mathbf{x}}$  be a minimizer of the RLS problem in (11a). Define

$$\Upsilon(\lambda, \delta) = \frac{-(\delta - \frac{\lambda}{\sigma_H^2} - 1) + \sqrt{(\delta - \frac{\lambda}{\sigma_H^2} - 1)^2 + 4\frac{\lambda}{\sigma_H^2}\delta}}{2\delta},$$

and

$$\theta_\star = \sqrt{\frac{\rho_d \sigma_H^2 \left( \frac{\Upsilon(\lambda, \delta)}{1 + \Upsilon(\lambda, \delta)} \right)^2 + \rho_d \sigma_\Delta^2 + 1}{\delta - \frac{1}{(1 + \Upsilon(\lambda, \delta))^2}}}.$$

Then, under Assumption 1 and Assumption 2, it holds:

$$\text{plim}_{K \rightarrow \infty} \text{MSE} = \frac{1}{\rho_d \sigma_H^2} \left( \delta \theta_\star^2 - \rho_d \sigma_\Delta^2 - 1 \right). \quad (17)$$

*Proof:*

The proof of Theorem 1 is given in Appendix C. ■

It is worth mentioning that the above formula is not restricted to  $\mathbf{x}_0$  belonging to  $M$ -PAM constellation and is valid for  $\mathbf{x}_0$  from any distribution provided that  $\mathbf{x}_0$  is normalized to have unit-variance. However, assuming that  $\mathbf{x}_0$  is drawn from  $M$ -PAM constellations, the SEP can be approximated as:

**Theorem 2.** (*SEP of RLS*): Under the same setting of Theorem 1, the SEP under  $M$ -PAM modulation of the RLS decoder employing the normalization constant

$$B = \frac{\sigma_H^2 \frac{\beta_\star}{\theta_\star}}{\sigma_H^2 \frac{\beta_\star}{\theta_\star} + 2\lambda},$$

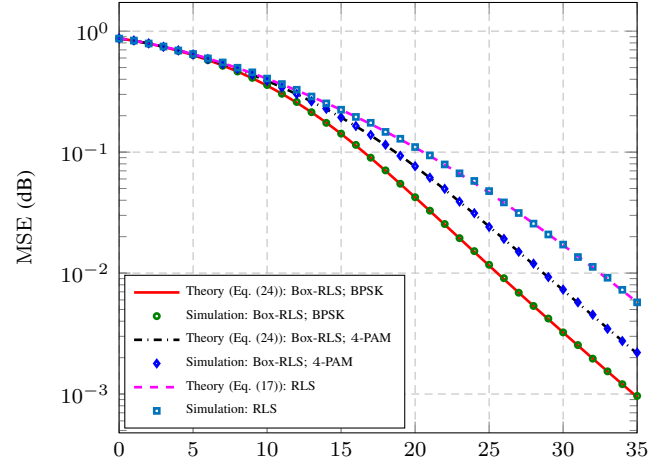


FIGURE 2: MSE performance of both RLS and Box-RLS decoders.

with  $\beta_\star = \frac{2\lambda\theta_\star}{\sigma_H^2 \Upsilon(\lambda, \delta)}$ , converges to:

$$\text{plim}_{K \rightarrow \infty} \text{SEP} = \widetilde{\text{SEP}}_{\text{RLS}}, \quad (18)$$

where

$$\widetilde{\text{SEP}}_{\text{RLS}} = 2 \left( 1 - \frac{1}{M} \right) Q \left( \frac{\sqrt{\rho_d \sigma_H^2}}{\sqrt{\mathcal{E}} \theta_\star} \right). \quad (19)$$

where  $\theta_\star$  is as defined in Theorem 1.

*Proof:*

A sketch of the proof is provided in Appendix C. ■

Before proceeding further, we validate the approximations provided in Theorem 1 and Theorem 2. To this end, we report in Figure 2 and 3 the MSE and SEP for the RLS decoder when  $K = 400$ ,  $\delta = 1.2$ ,  $T_p = 456$ ,  $T = 1000$ ,  $\alpha = 0.5$  and  $M = 2$  (corresponding to Binary Phase Shift Keying (BPSK) modulation), as a function of the average power  $\rho$ . As seen, the simulation results, averaged over 500 realizations of the channel, show a perfect agreement with the theoretical results.

**Corollary 1.** (Optimal regularization coefficient for RLS in MSE and SEP senses): Let  $\lambda_\star$  denote the optimal regularization coefficient that minimizes the limit in (17) or in (18). Then,

$$\lambda_\star = \frac{1}{\rho_d} + \sigma_\Delta^2. \quad (20)$$

*Proof:*

Note that in both (17) and (18), the regularization coefficient  $\lambda$  appears through  $\theta_\star$  only. Then,  $\lambda_\star = \arg \min_{\lambda \geq 0} \theta_\star$ . Taking the derivative of  $\theta_\star$  with respect to  $\lambda$ , setting it to zero and solving completes the proof of the corollary. ■

**Remark 1.** It is worth mentioning that the optimal regularization coefficient in (20) minimizes both the MSE and SEP. Moreover, it can be written in terms of the so called effective

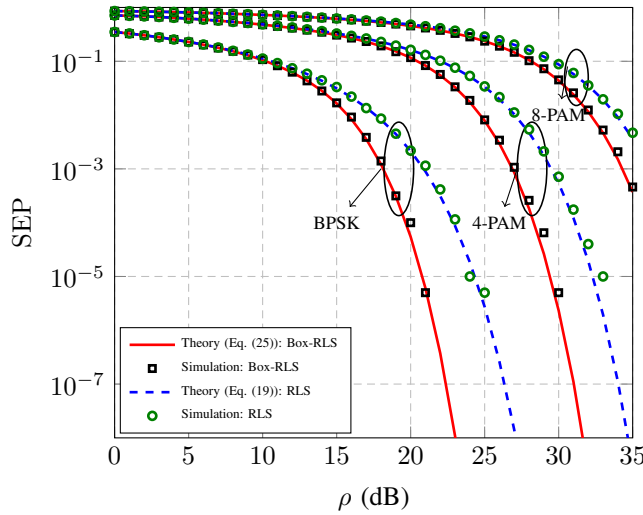


FIGURE 3: Symbol error probability of the RLS and Box-RLS decoders.

SNR of the system [4] as  $\lambda_\star = \frac{\sigma_H^2}{\rho_{\text{eff}}}$ , where

$$\rho_{\text{eff}} := \frac{\rho_a \sigma_H^2}{1 + \rho_a \sigma_\Delta^2}. \quad (21)$$

*Remark 2.* In Appendix H, we show that the RLS detector with optimal regularization coefficient is equivalent to the LMMSE detector. The later is known by definition to minimize the MSE, but it turns out according to Corollary 1 that it also minimizes the asymptotic SEP among all other choices of  $\lambda$ . In the perfect CSI case,  $\sigma_\Delta^2 = 0$ , hence the optimal regularization coefficient becomes  $\lambda_\star = \frac{1}{\rho_a}$  which is clearly equivalent to the LMMSE decoder. This shows that in both perfect and imperfect settings, the RLS with optimal regularization coefficient turns out to be the LMMSE decoder. Such a finding is appealing due to the fundamental importance of the LMMSE decoder in many applications.

### C. MSE and SEP Analysis for Box-RLS

In this subsection, we study the asymptotic performance of the Box-RLS decoder in terms of the MSE and SEP. We first present the MSE result in the following theorem.

**Theorem 3.** (*MSE of Box-RLS*): Fix  $\lambda > 0$ ,  $\delta > 0$ , and let  $\hat{\mathbf{x}}$  be a minimizer of the Box-RLS problem in (12a). Let  $\beta_\star$  and  $\theta_\star$  be the unique solutions in  $\beta$  and  $\theta$  to the following max-min problem:

$$\sup_{\beta > 0} \min_{\theta \geq 0} D(\theta, \beta) := \frac{\beta \delta \theta}{2} + \frac{\beta}{2\theta} (1 + \rho_a) - \frac{\beta^2}{4} + \frac{1}{M} \sum_{i=\pm 1, \pm 3, \dots, \pm(M-1)} \Psi(\theta, \beta; i), \quad (22)$$

with

$$\begin{aligned} \Psi(\theta, \beta; i) := & t(c_i Q(-\ell_i) + d_i Q(\mu_i)) \\ & - \beta \sqrt{\rho_a \sigma_H^2} t(p(\ell_i) + p(\mu_i)) \\ & - \frac{\beta^2}{2\rho_a \sigma_H^2 \theta + 4\lambda \rho_a} \int_{\ell_i}^{\mu_i} \left( \frac{\rho_a \sigma_H^2 i}{\sqrt{\mathcal{E}\theta}} + \sqrt{\rho_a \sigma_H^2} h \right)^2 p(h) dh, \end{aligned} \quad (23)$$

where  $\ell_i = -t\left(\frac{\sqrt{\rho_a \sigma_H^2}}{\theta} + \frac{2\lambda \rho_a}{\sqrt{\rho_a \sigma_H^2 \beta}}\right) - \frac{\sqrt{\rho_a \sigma_H^2} i}{\theta \sqrt{\mathcal{E}}}$ ,  $\mu_i = t\left(\frac{\sqrt{\rho_a \sigma_H^2}}{\theta} + \frac{2\lambda \rho_a}{\sqrt{\rho_a \sigma_H^2 \beta}}\right) - \frac{\sqrt{\rho_a \sigma_H^2} i}{\theta \sqrt{\mathcal{E}}}$ ,  $c_i = \frac{-\beta \sqrt{\rho_a \sigma_H^2}}{2} \ell_i + \frac{\beta \rho_a \sigma_H^2 i}{2\theta \sqrt{\mathcal{E}}}$ ,  $d_i = \frac{\beta \sqrt{\rho_a \sigma_H^2}}{2} \mu_i - \frac{\beta \rho_a \sigma_H^2 i}{2\theta \sqrt{\mathcal{E}}}$ , and  $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ . Then, under Assumption 1 and Assumption 2, it holds:

$$\text{plim}_{K \rightarrow \infty} \text{MSE} = \frac{1}{\rho_a \sigma_H^2} \left( \delta \theta_\star^2 - \rho_a \sigma_\Delta^2 - 1 \right), \quad (24)$$

*Proof:*

The proof of this theorem is deferred to Appendix B. ■

*Remark 3.* Theorem 3 is the key result of this paper, from which all other findings concerning the Box-RLS decoder have been derived. A crucial step in the proof of this theorem is to find the deterministic max-min problem in (22) which involves only the two scalar variables  $\beta$  and  $\theta$ . As shown by our analysis, it has a unique solution and its optimal cost corresponds to the asymptotic limit of the optimal cost of problem (12a) when scaled by  $1/K$ . While for the RLS decoder analyzed in Theorem 1, the asymptotically equivalent problem can be simplified to yield a closed-form expression for  $\theta_\star$ , the solution in the variable  $\theta$  and  $\beta$  does not have an explicit form.

*Remark 4.* It is worth mentioning that contrary to previous works based on the framework of the CGMT, the optimization problem in (22) which resulted from the asymptotic analysis is not convex-concave in the variables  $\theta$  and  $\beta$ . Indeed, it is concave in  $\beta$  but not convex in  $\theta$ . This poses a major challenge to prove the uniqueness of the solutions in  $\theta$  of (22), which is a crucial step that is required to ensure convergence results. The reader can refer to Appendix B for more details of the technical arguments developed to show the uniqueness of the solutions in (22).

*Remark 5.* If the optimal values  $\theta_\star$  and  $\beta_\star$  are strictly positive, then they satisfy the following first-order stationarity conditions:

$$\begin{aligned} \frac{\partial D(\theta, \beta)}{\partial \theta} &= 0, \\ \frac{\partial D(\theta, \beta)}{\partial \beta} &= 0, \end{aligned}$$

which can be exploited in practice to facilitate their numerical evaluation.

The following theorem provides the asymptotic expression of the SEP for the Box-RLS decoder.

**Theorem 4.** (*SEP for Box-RLS*): Under the same settings of Theorem 3, assuming that the Box-RLS decoder

uses a normalization constant given by  $B = \frac{\sigma_H^2 \beta_\star}{\sigma_H^2 \beta_\star + 2\lambda}$  where  $\beta_\star$  is the solution to (22) in  $\beta$ , and that  $t \notin \left\{ \frac{B_i}{\sqrt{\mathcal{E}}}, i = 1, 3, \dots, M-1 \right\}$ , it holds that:

$$\text{plim}_{K \rightarrow \infty} \text{SEP} = \widetilde{\text{SEP}}_{\text{Box-RLS}},$$

where  $\widetilde{\text{SEP}}_{\text{Box-RLS}}$  is given in (25).

If  $t \geq \frac{M-1}{\sqrt{\mathcal{E}}}$ , then  $\widetilde{\text{SEP}}_{\text{Box-RLS}}$  is simplified to:

$$\widetilde{\text{SEP}}_{\text{Box-RLS}} = 2 \left( 1 - \frac{1}{M} \right) Q \left( \frac{\sqrt{\rho_d \sigma_H^2}}{\sqrt{\mathcal{E} \theta_\star}} \right). \quad (26)$$

*Proof:*

The proof of Theorem 4 is also based on the CGMT framework and is given in Appendix B. ■

*Remark 6.* Figure 2 and Figure 3 reveal a perfect match between the analytical expressions of MSE and SEP given by Theorem 3 and Theorem 4 and the numerical simulations. It also shows that the Box-RLS outperforms the ordinary RLS. Figure 2 also suggests that as  $M$  increases, the performance of the Box-RLS approaches that of the un-boxed RLS. This is because in this figure  $t = \frac{M-1}{\sqrt{\mathcal{E}}}$  and as such as  $M \rightarrow \infty$ , the box-constraint  $[-t, t]$  tends to  $(-\infty, \infty)$  which is the whole real line  $\mathbb{R}$ , thereby reducing the Box-RLS to the RLS.

*Remark 7.* It should be noted that when  $t \geq \frac{M-1}{\sqrt{\mathcal{E}}}$ , the MSE and SEP expressions take the same form as in the RLS case, with the single difference that  $\theta_\star$  and  $\beta_\star$  do not admit a closed-form expression. Moreover, as for the RLS, the optimal regularization coefficient for Box-RLS is given by  $\lambda_\star = \arg \min_{\lambda \geq 0} \theta_\star$ , since  $\lambda$  appears in the expressions for the MSE and SEP only through  $\theta_\star$ . However, in contrast to the RLS, the optimal regularization coefficient cannot be obtained in closed-form, but could be retrieved by invoking any bisection algorithm. Moreover, as opposed to the RLS, its value depends on  $M$ .

*Remark 8.* Figure 4 plots the optimal regularization coefficient computed using a bisection algorithm as a function of  $\rho_d$  for RLS and Box-RLS for different values of  $M$ . As a first observation, we note that the optimal regularization coefficient for Box-RLS becomes zero starting from moderate values of  $\rho_d$ . Moreover, the Box-RLS needs less regularization, due to its achieved improvement over the RLS. On the other hand, in low SNR regions corresponding to low  $\rho_d$  values, the optimal regularization coefficient for both RLS and Box-RLS are higher than  $\frac{1}{\rho_d}$  which coincides with the optimal regularization coefficient in the perfect CSI case. This can be explained by the fact, under imperfect CSI cases, more regularization is needed in low SNR regions, because of the degradation caused by channel estimation errors.

*Remark 9.* Similar to the regularization coefficient, we can set the threshold  $t$  to the optimal value that minimizes the MSE and SEP expressions, that is  $t_\star = \arg \min_{t > 0} \theta_\star$ . Figure 5 shows the optimal box-threshold as a function of  $\rho_d$  when the regularization coefficient is already optimized as

well. As can be seen, for practical SNR regions, the optimal threshold coincides with  $\frac{M-1}{\sqrt{\mathcal{E}}}$ , which is the maximum value of  $x_0$ . For this reason, we will use in the subsequent simulations this value for the threshold  $t$ .

## V. Optimal Data Power Allocation and Optimal Training Duration Allocation

In this section, we leverage the asymptotic expressions of the MSE and SEP derived thus far for the RLS and Box-RLS to determine the optimal power distribution between the training and data symbols. Particularly, we show that for all considered decoders, the optimal allocation schemes boils down to maximizing the effective SNR of the system  $\rho_{\text{eff}}$  defined in (21). Additionally, we derive for each decoder the asymptotic expression of the goodput and derive the optimal fraction of power allocated to the pilot transmission as well as the training duration (i.e.,  $(\tau_p, \alpha)$  that maximizes the goodput). In this respect, we illustrate that, while the optimal power allocation remains to be the one that maximizes the effective SNR, the optimal number of training symbols coincides with the number of transmitting antennas  $K$ , which is also the minimum number of training symbols that needs to be employed to satisfy orthogonality between pilot sequences.

### A. Simplifying the Decoders' Expressions

#### 1) LS decoder

To begin with, we consider first the LS decoder for which  $\lambda = 0$  and  $\delta > 1$ . Hence, from (17),  $\Upsilon(0, \delta) = 0$ , and as such:

$$\theta_\star = \sqrt{\frac{\rho_d \sigma_\Delta^2 + 1}{\delta - 1}},$$

The MSE limit in (17) reduces thus to

$$\begin{aligned} \widetilde{\text{MSE}}_{\text{LS}} &:= \text{plim}_{K \rightarrow \infty} \text{MSE}_{\text{LS}} = \frac{\delta(\rho_d \sigma_\Delta^2 + 1)}{(\delta - 1)\rho_d \sigma_H^2} - \frac{\rho_d \sigma_\Delta^2 + 1}{\rho_d \sigma_H^2} \\ &= \frac{\delta}{(\delta - 1)\rho_{\text{eff}}} - \frac{1}{\rho_{\text{eff}}} \\ &= \frac{1}{(\delta - 1)\rho_{\text{eff}}}, \end{aligned} \quad (27)$$

where  $\rho_{\text{eff}}$  is the effective SNR defined in (21). The result in (27) recovers the well-known formula of the MSE of LS with the difference being that  $\rho_d$  which stands for the SNR in the perfect CSI case is replaced by  $\rho_{\text{eff}}$ . Similarly, for  $\delta > 1$ , and from (18), the SEP of the LS decoder can also be expressed in terms of  $\rho_{\text{eff}}$  as follows

$$\begin{aligned} \text{plim}_{K \rightarrow \infty} \text{SEP}_{\text{LS}} &= 2 \left( 1 - \frac{1}{M} \right) Q \left( \sqrt{\frac{(\delta - 1)}{\mathcal{E}} \rho_{\text{eff}}} \right) \\ &= 2 \left( 1 - \frac{1}{M} \right) Q \left( \sqrt{\frac{1}{\mathcal{E} \cdot \widetilde{\text{MSE}}_{\text{LS}}}} \right). \end{aligned} \quad (28)$$

Again, the first equation in (28) parallels the well-known result for the LS and BPSK signaling but under perfect CSI [61], (in which case the BER converges in probability to  $\text{BER}_{\text{LS}} = Q(\sqrt{(\delta - 1)\rho_d})$ ) in that it takes the same



$$\begin{aligned} \widetilde{\text{SEP}}_{\text{Box-RLS}} = & \frac{4}{M} \sum_{i=1,3,\dots,M-3} \mathbf{1}_{\{\frac{t}{B} \geq \frac{i+1}{\sqrt{\mathcal{E}}}\}} Q\left(\frac{\sqrt{\rho_d \sigma_{\hat{H}}^2}}{\sqrt{\mathcal{E}} \theta_*}\right) + \frac{2}{M} \sum_{i=1,3,\dots,M-3} \left\{ \mathbf{1}_{\{\frac{i-1}{\sqrt{\mathcal{E}}} \leq \frac{t}{B} \leq \frac{i+1}{\sqrt{\mathcal{E}}}\}} Q\left(\frac{\sqrt{\rho_d \sigma_{\hat{H}}^2}}{\sqrt{\mathcal{E}} \theta_*}\right) + \mathbf{1}_{\{\frac{t}{B} \leq \frac{i-1}{\sqrt{\mathcal{E}}}\}} \right\} \\ & + \frac{2}{M} \mathbf{1}_{\{\frac{t}{B} \geq \frac{M-2}{\sqrt{\mathcal{E}}}\}} Q\left(\frac{\sqrt{\rho_d \sigma_{\hat{H}}^2}}{\sqrt{\mathcal{E}} \theta_*}\right) + \frac{2}{M} \mathbf{1}_{\{\frac{t}{B} \leq \frac{M-2}{\sqrt{\mathcal{E}}}\}}. \end{aligned} \quad (25)$$

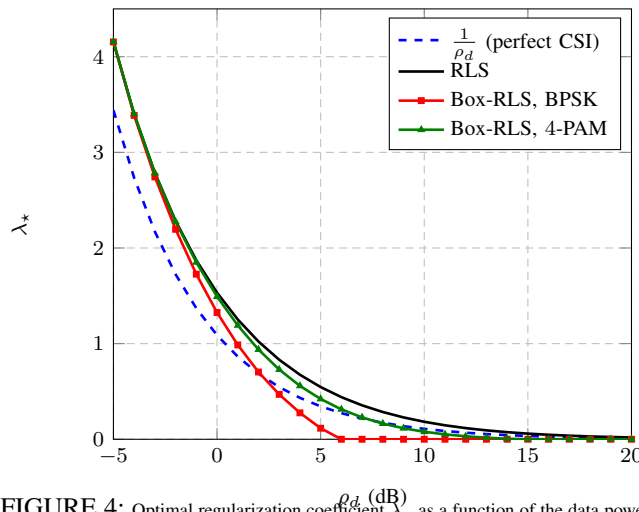


FIGURE 4: Optimal regularization coefficient  $\lambda_*$  as a function of the data power. We used  $\delta = 1.2$ ,  $K = 400$ ,  $\alpha = 0.5$ ,  $T_p = 400$ , and  $T = 1000$ .

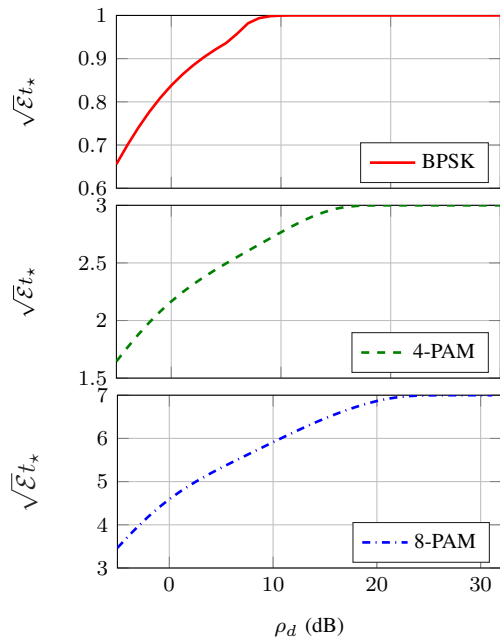


FIGURE 5: The optimal normalized box-threshold  $\sqrt{\mathcal{E}}t_*$  as a function of the data power. We used  $\delta = 1.2$ ,  $K = 400$ ,  $T_p = 400$ ,  $\alpha = 0.5$ , and  $T = 1000$ .

form with  $\rho_{\text{eff}}$  replacing  $\rho_d$ . Hence, our result generalizes [61] to encompass  $M$ -PAM modulation and imperfect CSI scenarios.

## 2) RLS decoder

We proceed now with the RLS decoder. The MSE expression in (17) can also be written in terms of  $\rho_{\text{eff}}$  as

$$\widetilde{\text{MSE}}_{\text{RLS}} := \text{plim}_{K \rightarrow \infty} \text{MSE}_{\text{RLS}} = \frac{\delta \theta_*^2}{\rho_d \sigma_{\hat{H}}^2} - \frac{1}{\rho_{\text{eff}}}, \quad (29)$$

from which it follows that  $\frac{\rho_d \sigma_{\hat{H}}^2}{\theta_*^2} = \frac{\delta}{\widetilde{\text{MSE}}_{\text{RLS}} + \frac{1}{\rho_{\text{eff}}}}$ . This yields the following interesting relationship between the MSE and SEP for the RLS decoder:

$$\widetilde{\text{SEP}}_{\text{RLS}} := \text{plim}_{K \rightarrow \infty} \text{SEP}_{\text{RLS}} = 2 \left(1 - \frac{1}{M}\right) Q\left(\sqrt{\frac{\delta}{\mathcal{E} \cdot (\widetilde{\text{MSE}}_{\text{RLS}} + \frac{1}{\rho_{\text{eff}}})}}\right). \quad (30)$$

Such an expression holds for any  $\lambda > 0$ , and not necessarily  $\lambda_*$ . But when  $\lambda = \lambda_*$ , with  $\lambda_* = \frac{\sigma_{\hat{H}}^2}{\rho_{\text{eff}}}$ , we obtain after some algebraic manipulations the following expression for the MSE:

$$\text{plim}_{K \rightarrow \infty} \text{MSE}_{\text{RLS}} = \frac{1}{2} \left( -(\delta - 1 + \frac{1}{\rho_{\text{eff}}}) + \sqrt{(\delta - 1 + \frac{1}{\rho_{\text{eff}}})^2 + \frac{4}{\rho_{\text{eff}}}} \right). \quad (31)$$

Note that in the perfect CSI case for which the optimal regularization coefficient is  $\frac{1}{\rho_d}$ , the right-hand side of (31) is exactly the minimum mean squared error (MMSE) estimator (see [62, Theorem 8]), where  $\rho_{\text{eff}}$  is replaced by  $\rho_d$ .

## 3) Box-RLS decoder

In a similar way, for the Box-RLS decoder, we have the same asymptotic relationships between MSE and SEP:

$$\widetilde{\text{MSE}}_{\text{Box-RLS}} := \text{plim}_{K \rightarrow \infty} \text{MSE}_{\text{Box-RLS}} = \frac{\delta \theta_*^2}{\rho_d \sigma_{\hat{H}}^2} - \frac{1}{\rho_{\text{eff}}}, \quad (32)$$

and, for  $t \geq \frac{M-1}{\sqrt{\mathcal{E}}}$ :

$$\begin{aligned} \widetilde{\text{SEP}}_{\text{Box-RLS}} &:= \text{plim}_{K \rightarrow \infty} \text{SEP}_{\text{Box-RLS}} \\ &= 2 \left(1 - \frac{1}{M}\right) Q\left(\sqrt{\frac{\delta}{\mathcal{E} \cdot (\widetilde{\text{MSE}}_{\text{Box-RLS}} + \frac{1}{\rho_{\text{eff}}})}}\right), \end{aligned} \quad (33)$$

which again reveals that minimizing the MSE is equivalent to minimizing the SEP.

### B. Optimal Power Allocation in MSE and SEP Sense

For the RLS decoder, we prove in Appendix E that both  $\widehat{\text{MSE}}_{\text{RLS}}$ , and  $\widehat{\text{SEP}}_{\text{RLS}}$  are monotonically increasing functions in  $\frac{1}{\rho_{\text{eff}}}$ . Hence, minimizing the MSE or SEP is equivalent to maximizing  $\rho_{\text{eff}}$ . This can be easily seen to be the case of the LS decoder. However, for the Box-RLS decoder, such a statement could not be checked analytically as  $\theta_*$  does not possess a closed-form expressions. However, based on extensive simulations, we conjecture that both  $\widehat{\text{MSE}}_{\text{Box-RLS}}$  and  $\widehat{\text{SEP}}_{\text{Box-RLS}}$  increase with  $\frac{1}{\rho_{\text{eff}}}$ . All these considerations suggest that the optimal power allocation is the one that maximizes  $\rho_{\text{eff}}$  over  $\alpha$ , i.e.,

$$\alpha_* = \arg \max_{0 < \alpha < 1} \rho_{\text{eff}}. \quad (34)$$

Recall that  $\rho_{\text{eff}} = \frac{\rho_d \sigma_H^2}{1 + \rho_d \sigma_\Delta^2}$ . Substituting the expressions for  $\sigma_H^2$  and  $\sigma_\Delta^2$  gives

$$\rho_{\text{eff}} = \frac{\tau_p \rho_p \rho_d}{(1 + \rho_d) + \tau_p \rho_p}. \quad (35)$$

Further, upon using  $\rho_p = \frac{(1-\alpha)\rho\tau}{\tau_p}$ , and  $\rho_d = \frac{\alpha\rho\tau}{\tau_d}$ , the effective SNR becomes

$$\rho_{\text{eff}} = \frac{\rho\tau}{\tau_d - 1} \cdot \frac{\alpha(1-\alpha)}{-\alpha + \frac{1+\rho\tau}{\rho\tau(1-\frac{1}{\tau_d})}}.$$

With this expression at hand, we determine in the following Theorem the optimal power allocation that maximizes the effective SNR:

**Theorem 5. (Optimal Power Allocation):** The optimal power allocation  $\alpha_*$  that maximizes the effective SNR in a training-based system is given by

$$\alpha_* = \begin{cases} \vartheta - \sqrt{\vartheta(\vartheta - 1)}, & \text{if } \tau_d > 1, \\ \frac{1}{2}, & \text{if } \tau_d = 1, \\ \vartheta + \sqrt{\vartheta(\vartheta - 1)} & \text{if } \tau_d < 1, \end{cases} \quad (36)$$

where  $\vartheta = \frac{1+\rho\tau}{\rho\tau(1-\frac{1}{\tau_d})}$ .

*Proof:*

The proof of this theorem is given in Appendix F. ■

It is worth mentioning that the use of this power allocation has already been proposed in the early work of [4] as the one that maximizes a lower bound on the capacity. Interestingly, we retrieve the same power allocation scheme which we prove to be optimum in the MSE/SEP sense for RLS and LS decoders and conjecture that it is also optimum for the Box-RLS decoder.

*Remark 10.* • At high SNR ( $\rho \gg 1$ ),  $\vartheta \approx \frac{\tau_d}{\tau_d - 1}$ , then

$$\alpha_* \approx \frac{\sqrt{\tau_d}}{1 + \sqrt{\tau_d}}.$$

- At low SNR ( $\rho \ll 1$ ),  $\vartheta \approx \frac{\tau_d}{\rho\tau(\tau_d - 1)}$ , and  $\alpha_* \approx \frac{1}{2}$ . This means that, at low SNR, *half* of the transmit energy should be devoted to training and the other half to data transmission.

*Remark 11 (Numerical Illustration).* The asymptotic predictions of the MSE and the SEP are plotted as functions of the data power ratio  $\alpha$  in Figure 6 and Figure 7 when  $\delta = 2, K = 256, T = 1000, T_p = 256, M = 2$  and  $\rho = 15$

dB. As can be seen, the optimal power allocation,  $\alpha_*$ , is the same in the MSE and SEP sense for the different decoders considered here, namely LS, RLS and Box-RLS. The same conclusion has been found for other settings confirming the conjecture that for Box-RLS, the optimal power allocation is obtained by maximizing the effective SNR.

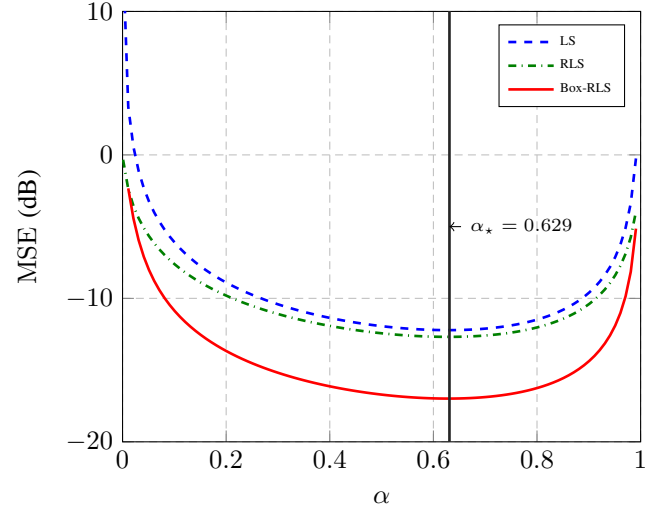


FIGURE 6: MSE as a function of the data power ratio  $\alpha$ .

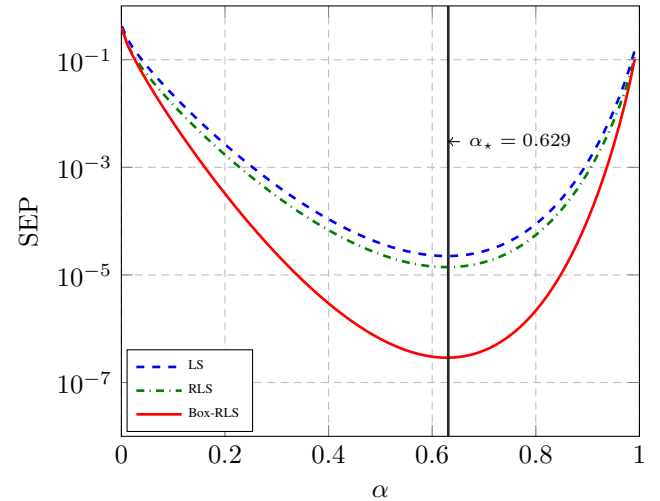


FIGURE 7: SEP as a function of the data power ratio  $\alpha$ .

### C. Joint Optimization of Power Allocation and Training Duration in the Goodput Sense

We now consider the goodput metric for the joint optimization of the power allocation and the training duration. From its definition in (16), its asymptotic value can be written as:

$$\text{plim}_{K \rightarrow \infty} G = \left(1 - \frac{\tau_p}{\tau}\right) \left(1 - \text{plim}_{K \rightarrow \infty} \text{SEP}\right), \quad (37)$$

where the limit in the right hand side is given by (18) for RLS, and by (25) for Box-RLS. The above expression can

be used to find the optimal pair  $(\tau_p^*, \alpha_*)$  that maximizes the goodput limit in (37). The result is summarized below.

**Proposition 1.** (*Joint optimization in goodput sense*): The optimal pair  $(\tau_p^*, \alpha_*)$  that maximizes the goodput limit in (37) is given by:  $\tau_p^* = 1$  (or  $T_p^* = K$ ), and  $\alpha_*$  is the same as in (36) for all  $\rho$  and  $\tau$  (or  $T$ ).

*Proof:*

The proof of this proposition is given in Appendix G. ■

*Remark 12.* A major outcome of the above result is that the optimal number of training symbols that maximizes the goodput is given by the minimum number of the required training symbols that is the number of transmit antennas,  $K$ . This result differs from the finding of [4] in which it has been proven that in case of equal distribution of power between training and data, the optimal number of training symbols may be larger than the number of transmit antennas.

## VI. Conclusions

A large body of the literature has thus far focused on the theoretical analysis of decoders with explicit formulation, like LS and RLS decoders. The prospect of achieving nearly optimal performance has led researchers in the field of wireless communication to consider non-linear decoders. Such decoders are in general based on implicit formulations and thus are more difficult to analyze. The recent progress made on new tools like the CGMT has rekindled interest in the performance analysis of techniques involving problems lacking explicit formulations. Falling within this context, our work aims to study the Box-RLS decoder when used to recover  $M$ -ary constellation signaling under imperfect channel knowledge. This decoder, recently proposed in former works, is based on solving the same optimization problem as the RLS decoder but while constraining the solution to lie in a zero-centered interval  $[-t, t]$ . Our work differs from previous works that used the CGMT to analyze the Box-RLS decoder in the following ways. Firstly, considering the LMMSE channel estimation error, it is, to the best of our knowledge, the first work to account for the impact of channel estimation errors in the study of the Box-RLS decoder. Secondly, it derives tight approximations of several metrics ranging from MSE, SEP to goodput for RLS, LS, and Box-RLS decoders. Based on these metrics, it determines optimal ways to share the power between training and data to minimize MSE and SEP, as well as the number of pilot symbols to maximize the goodput. Apart from its practical relevance, consideration of imperfect CSI posed several technical challenges and brought us to develop novel tools to rigorously prove convergence results. Such tools can be of independent interest and may be used in the future to further facilitate carrying out rigorous analysis based on the use of the CGMT framework.

## Acknowledgment

We would like to thank Prof. Ahmed-Sultan Salem for the very helpful comments, discussions and suggestions. We also would like to thank Housseem Sifaou for fruitful discussions.

## Appendix A

### Gaussian Min-max Theorem

The key ingredient of the analysis is the Convex Gaussian Min-max Theorem (CGMT), a concrete formulation for it can be found in [51]. The CGMT is a tool that allows analyzing the behavior of solutions of stochastic optimization problems that can be cast into the following form:

$$\Phi(\mathbf{G}) := \min_{\mathbf{w} \in \mathcal{S}_w} \max_{\mathbf{u} \in \mathcal{S}_u} \mathbf{u}^T \mathbf{G} \mathbf{w} + \psi(\mathbf{w}, \mathbf{u}), \quad (38)$$

where  $\mathbf{G} \in \mathbb{R}^{N \times K}$  with i.i.d. standard normal entries,  $\mathcal{S}_w$  and  $\mathcal{S}_u$  are sets of  $\mathbb{R}^K$  and  $\mathbb{R}^N$  and  $\psi : \mathbb{R}^K \times \mathbb{R}^N \rightarrow \mathbb{R}$  is continuous convex-concave function on  $\mathcal{S}_w \times \mathcal{S}_u$ . Problem (38) is referred to as Primary Problem (PO) and its analysis is in general not tractable. The CGMT associates with it an Auxiliary Optimization (AO) problem given by

$$\phi(\mathbf{g}, \mathbf{s}) := \min_{\mathbf{w} \in \mathcal{S}_w} \max_{\mathbf{u} \in \mathcal{S}_u} \|\mathbf{w}\| \mathbf{g}^T \mathbf{u} - \|\mathbf{u}\| \mathbf{s}^T \mathbf{w} + \psi(\mathbf{w}, \mathbf{u}), \quad (39)$$

where  $\mathbf{g} \in \mathbb{R}^N$ , and  $\mathbf{s} \in \mathbb{R}^K$  have i.i.d. standard Gaussian entries. The initial formulation of the CGMT establishes that the (AO) has the same asymptotic behavior as the (PO) in the regime in which  $N$  and  $K$  grow simultaneously with the same pace under the condition that the sets  $\mathcal{S}_w$  and  $\mathcal{S}_u$  are convex and compact. Particularly, if for some  $\nu \in \mathbb{R}$ , the optimal cost of the (AO) concentrates around  $\nu$  in the sense that

$$\mathbb{P} [|\phi(\mathbf{g}, \mathbf{s}) - \nu|] \rightarrow 0,$$

then the optimal cost of the PO concentrates also around  $\nu$ , satisfying similarly:

$$\mathbb{P} [|\Phi(\mathbf{G}) - \nu|] \rightarrow 0.$$

Recently in [63], the compactness of  $\mathcal{S}_u$  is shown to be possibly relaxed provided that the order of the min-max in (39) can be inverted, that is  $\phi(\mathbf{g}, \mathbf{s})$  is also given by:

$$\phi(\mathbf{g}, \mathbf{s}) := \max_{\mathbf{u} \in \mathcal{S}_u} \min_{\mathbf{w} \in \mathcal{S}_w} \|\mathbf{w}\| \mathbf{g}^T \mathbf{u} - \|\mathbf{u}\| \mathbf{s}^T \mathbf{w} + \psi(\mathbf{w}, \mathbf{u}). \quad (40)$$

More formally, we have the following result:

**Theorem 6** (CGMT [51]). Let  $\mathcal{S}$  be any arbitrary open subset of  $\mathcal{S}_w$ , and  $\mathcal{S}^c = \mathcal{S}_w \setminus \mathcal{S}$ . Denote  $\phi_{\mathcal{S}^c}(\mathbf{g}, \mathbf{s})$  the optimal cost of the optimization in (39), when the minimization over  $\mathbf{w}$  is constrained over  $\mathbf{w} \in \mathcal{S}^c$ . Assume that  $\mathcal{S}_u$  is convex while  $\mathcal{S}_w$  is convex and compact. Assume also that (40) holds true. Consider the regime  $K, N \rightarrow \infty$  such that  $\frac{N}{K} \rightarrow \delta$ , which will be denoted by  $K \rightarrow \infty$ . Suppose that there exist constants  $\bar{\phi}$  and  $\eta > 0$  such that in the limit as  $K \rightarrow +\infty$  it holds with probability approaching one: (i)  $\phi(\mathbf{g}, \mathbf{s}) \leq \bar{\phi} + \eta$ , and, (ii)  $\phi_{\mathcal{S}^c}(\mathbf{g}, \mathbf{s}) \geq \bar{\phi} + 2\eta$ . Let  $\mathbf{w}_\Phi$  and  $\mathbf{w}_\phi$  denote respectively the solutions in  $\mathbf{w}$  to the (PO) and the (AO). Then,  $\lim_{K \rightarrow \infty} \mathbb{P}[\mathbf{w}_\phi \in \mathcal{S}] = 1$ , and  $\lim_{K \rightarrow \infty} \mathbb{P}[\mathbf{w}_\Phi \in \mathcal{S}] = 1$ .

*Remark 13.* It is worth mentioning that the result in Theorem 6 goes beyond the asymptotic equivalence between the costs of the (AO) and the (PO) to the localization of the (PO) and (AO) solutions. More specifically, one can easily see that conditions (i) and (ii) in Theorem 6 imply that the solution of the (AO) lies in the set  $\mathcal{S}$  with probability approaching 1. Theorem 6 allows us to carry over this property to the solution of the (PO), that is  $\mathbf{w}_\Phi$  is in  $\mathcal{S}$  with probability approaching 1.

*Remark 14.* To satisfy (i) and (ii) in Theorem 6, one can prove that  $\phi(\mathbf{g}, \mathbf{s})$  converges to  $\bar{\phi}$  while  $\phi_{\mathcal{S}^c}(\mathbf{g}, \mathbf{s})$  is lower-bounded by a quantity that converges to  $\bar{\phi}_{\mathcal{S}^c}$  with

$$\bar{\phi}_{\mathcal{S}^c} > \bar{\phi}. \quad (41)$$

In practice, it is usually the case that  $\bar{\phi}$  and  $\bar{\phi}_{\mathcal{S}^c}$  represent optimal costs of the same optimization problem but with the solution of the latter being constrained to be away from the optimal solution of the former. Under this setting, showing that the optimization problem whose optimal cost is  $\bar{\phi}$  admits a unique solution directly implies (41).

## Appendix B Proofs of Box-RLS

In this appendix we prove Theorem 3 and Theorem 4. For simplicity, we will divide the steps of the proof into subsections.

### A. Identifying the (PO) and the (AO)

For convenience, we consider the error vector  $\mathbf{w} := \mathbf{x} - \mathbf{x}_0$ , and also the box set:

$$\mathcal{B} = \left\{ \mathbf{w} \in \mathbb{R}^K \mid -t - x_{0,j} \leq w_j \leq t - x_{0,j}, j \in \{1, 2, \dots, K\} \right\}, \quad (42)$$

With this notation, the problem in (12a) can be reformulated as

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathcal{B}} \left\| \sqrt{\frac{\rho_d}{K}} \hat{\mathbf{H}} \mathbf{w} - \sqrt{\frac{\rho_d}{K}} \Delta \mathbf{x}_0 - \mathbf{z} \right\|^2 + \rho_d \lambda \|\mathbf{x}_0 + \mathbf{w}\|^2. \quad (43)$$

To bring the problem in (43) to the form of (38) required by the CGMT, we express the loss function of (43) in its dual form through the Fenchel's conjugate as

$$\begin{aligned} & \left\| \sqrt{\frac{\rho_d}{K}} \hat{\mathbf{H}} \mathbf{w} - \sqrt{\frac{\rho_d}{K}} \Delta \mathbf{x}_0 - \mathbf{z} \right\|^2 \\ &= \max_{\mathbf{u}} \mathbf{u}^T \left( \sqrt{\frac{\rho_d}{K}} \hat{\mathbf{H}} \mathbf{w} - \sqrt{\frac{\rho_d}{K}} \Delta \mathbf{x}_0 - \mathbf{z} \right) - \frac{\|\mathbf{u}\|^2}{4}. \end{aligned}$$

Hence, the problem in (43) is equivalent to the following:

$$\begin{aligned} & \min_{\mathbf{w} \in \mathcal{B}} \max_{\mathbf{u}} \mathbf{u}^T \sqrt{\frac{\rho_d}{K}} \hat{\mathbf{H}} \mathbf{w} - \mathbf{u}^T \sqrt{\frac{\rho_d}{K}} \Delta \mathbf{x}_0 \\ & \quad - \mathbf{u}^T \mathbf{z} - \frac{\|\mathbf{u}\|^2}{4} + \rho_d \lambda \|\mathbf{x}_0 + \mathbf{w}\|^2, \quad (44) \end{aligned}$$

To reach the desired PO form, we introduce the variables

$$\begin{aligned} \mathbf{v} &= \begin{bmatrix} \sigma_{\hat{H}} \sqrt{\rho_d} \mathbf{w} \\ -\sigma_{\Delta} \sqrt{\rho_d} \mathbf{x}_0 \end{bmatrix} \in \mathbb{R}^{2K}, \quad \mathbf{G} = \begin{bmatrix} \tilde{\mathbf{H}} & \tilde{\Delta} \end{bmatrix} \in \mathbb{R}^{N \times 2K} \text{ and} \\ \mathbf{C} &= \begin{bmatrix} \frac{1}{\sigma_{\hat{H}}} \mathbf{I}_K & \frac{1}{\sigma_{\Delta}} \mathbf{I}_K \end{bmatrix} \in \mathbb{R}^{K \times 2K}, \text{ where } \tilde{\mathbf{H}} \text{ and } \tilde{\Delta} \text{ are } N \times \end{aligned}$$

$K$  independent matrices with i.i.d. standard normal entries. Now, using these variables, and after normalization by  $1/K$ , the above problem can be written as:

$$\begin{aligned} \Phi &:= \frac{1}{\sqrt{K}} \min_{\mathbf{v} \in \mathcal{D}} \max_{\mathbf{u}} \frac{1}{K} \mathbf{u}^T \mathbf{G} \mathbf{v} - \frac{1}{\sqrt{K}} \mathbf{u}^T \mathbf{z} \\ & \quad - \frac{\|\mathbf{u}\|^2}{4\sqrt{K}} + \frac{\lambda}{\sqrt{K}} \|\mathbf{C} \mathbf{v}\|^2, \quad (45) \end{aligned}$$

where

$$\mathcal{D} = \left\{ \mathbf{v} \in \mathbb{R}^{2K} \mid \mathbf{v} = \begin{bmatrix} \sigma_{\hat{H}} \sqrt{\rho_d} \mathbf{w} \\ -\sigma_{\Delta} \sqrt{\rho_d} \mathbf{x}_0 \end{bmatrix}, \text{ and } \mathbf{w} \in \mathcal{B} \right\}.$$

We note that (45) is in the form of the (PO) and associate with it the following (AO)

$$\begin{aligned} \phi &:= \frac{1}{\sqrt{K}} \min_{\mathbf{v} \in \mathcal{D}} \max_{\mathbf{u}} \|\mathbf{v}\| \frac{1}{K} \mathbf{g}^T \mathbf{u} - \frac{1}{K} \|\mathbf{u}\| \mathbf{q}^T \mathbf{v} - \frac{1}{\sqrt{K}} \mathbf{u}^T \mathbf{z} \\ & \quad - \frac{1}{4\sqrt{K}} \|\mathbf{u}\|^2 + \frac{1}{\sqrt{K}} \lambda \|\mathbf{C} \mathbf{v}\|^2, \quad (46) \end{aligned}$$

where  $\mathbf{q} \in \mathbb{R}^{2K}$  and  $\mathbf{g} \in \mathbb{R}^N$  are independent standard normal vectors. Note that for the moment, we relate the (PO) to the unbounded (AO) as  $\mathcal{S}_u = \mathbb{R}^N$  is not compact. In the sequel, we check that (40) holds true, which gives support to considering the unbounded (AO) according to Theorem 6.

### B. Scalarizing the (AO)

The next step is to simplify the (AO) as it appears in (46) into an optimization problem involving only scalar variables. Since the vectors  $\mathbf{g}$  and  $\mathbf{z}$  are independent and have i.i.d. Gaussian entries, then,  $\|\mathbf{v}\| \mathbf{g} - \sqrt{K} \mathbf{z}$  has i.i.d. entries  $\mathcal{N}(0, \|\mathbf{v}\|^2 + K)$ . Hence, for our purposes and using some abuse of notation so that  $\mathbf{g}$  continues to denote a vector with i.i.d. standard normal entries, the corresponding terms in (46) can be combined as  $\sqrt{\|\mathbf{v}\|^2 + K} \mathbf{g}^T \mathbf{u}$ , instead. Therefore, (46) is equivalent to

$$\begin{aligned} \phi &= \frac{1}{\sqrt{K}} \min_{\mathbf{v} \in \mathcal{D}} \max_{\mathbf{u}} \mathbf{g}^T \mathbf{u} \sqrt{\frac{1}{K} \|\mathbf{v}\|^2 + 1} - \frac{1}{K} \|\mathbf{u}\| \mathbf{q}^T \mathbf{v} \\ & \quad - \frac{1}{4\sqrt{K}} \|\mathbf{u}\|^2 + \frac{1}{\sqrt{K}} \lambda \|\mathbf{C} \mathbf{v}\|^2. \quad (47) \end{aligned}$$

Expressing the above problem in terms of the original  $\mathbf{w}$  variable:

$$\begin{aligned} \phi &= \frac{1}{\sqrt{K}} \min_{\mathbf{w} \in \mathcal{B}} \max_{\mathbf{u}} \mathbf{g}^T \mathbf{u} \sqrt{\sigma_{\hat{H}}^2 \rho_d \frac{1}{K} \|\mathbf{w}\|^2 + \sigma_{\Delta}^2 \rho_d \frac{1}{K} \|\mathbf{x}_0\|^2 + 1} \\ & \quad - \frac{1}{4\sqrt{K}} \|\mathbf{u}\|^2 - \frac{\|\mathbf{u}\|}{\sqrt{K}} \sqrt{\rho_d} (\sigma_{\hat{H}} \mathbf{w}^T \mathbf{q}^1 - \sigma_{\Delta} \mathbf{x}_0^T \mathbf{q}^2) \\ & \quad + \frac{1}{\sqrt{K}} \lambda \rho_d \|\mathbf{x}_0 + \mathbf{w}\|^2, \quad (48) \end{aligned}$$

where  $\mathbf{q}^1, \mathbf{q}^2 \in \mathbb{R}^K$  are independent standard normal vectors.

Fixing the norm of  $\frac{\mathbf{u}}{\sqrt{K}}$  to  $\beta := \frac{\|\mathbf{u}\|}{\sqrt{K}}$ , it is easy to see that its optimal direction should be aligned with  $\mathbf{g}$ . Working with  $\mathbf{x}$  instead of  $\mathbf{w}$  results into the following optimization

problem:

$$\begin{aligned} \phi = & \min_{\substack{-t \leq x_j \leq t \\ j=1, \dots, K}} \max_{\beta > 0} \frac{\beta \|\mathbf{g}\|}{\sqrt{K}} \sqrt{\frac{\rho_d}{K} (\sigma_{\tilde{H}}^2 \|\mathbf{x}\|^2 - 2\sigma_{\tilde{H}}^2 \mathbf{x}_0^T \mathbf{x} + \|\mathbf{x}_0\|^2) + 1} \\ & - \frac{\beta^2}{4} - \beta \sqrt{\rho_d} \sigma_{\tilde{H}} \frac{1}{K} \mathbf{x}^T \mathbf{q}^1 + \frac{1}{K} \beta \sqrt{\rho_d} \mathbf{x}_0^T \tilde{\mathbf{h}} + \lambda \rho_d \frac{1}{K} \|\mathbf{x}\|^2, \end{aligned} \quad (49)$$

where  $\tilde{\mathbf{h}} = \sigma_{\tilde{H}} \mathbf{q}^1 + \sigma_{\Delta} \mathbf{q}^2$ . At this point, it is worth mentioning that (49) is convex in  $\mathbf{x}$  and concave in  $\beta$ . Based on this, we can prove that (40) holds true. Indeed using [64, Cor. 37.3.2], we can flip the order of  $\min_{\mathbf{x}} \max_{\beta}$ . Now, consider the following problem:

$$\begin{aligned} \max_{\beta} \max_{\substack{\tilde{\mathbf{u}} \\ \|\tilde{\mathbf{u}}\|=1}} \min_{\substack{-t \leq x_j \leq t \\ j=1, \dots, K}} \beta \mathbf{g}^T \tilde{\mathbf{u}} \sqrt{\frac{\rho_d}{K} (\sigma_{\tilde{H}}^2 \|\mathbf{x}\|^2 - 2\sigma_{\tilde{H}}^2 \mathbf{x}_0^T \mathbf{x} + \|\mathbf{x}_0\|^2) + 1} \\ - \frac{\beta^2}{4} - \beta \sqrt{\rho_d} \sigma_{\tilde{H}} \frac{1}{K} \mathbf{x}^T \mathbf{q}^1 + \frac{1}{K} \beta \sqrt{\rho_d} \mathbf{x}_0^T \tilde{\mathbf{h}} + \lambda \rho_d \frac{1}{K} \|\mathbf{x}\|^2. \end{aligned} \quad (50)$$

Then, based on [65, Lemma 8] and setting  $\mathbf{u} = \sqrt{K} \beta \tilde{\mathbf{u}}$  we can prove that (50) is the same as (49) in which the order of the min-max is inverted. This completes the proof of (40), which as aforementioned, allows us to extend the scope of the CGMT for optimization problems in which the variable  $\mathbf{u}$  is constrained to lie in a non-compact set.

Now, getting back to the optimization problem in (49) and flipping the order of  $\min_{\mathbf{x}} \max_{\beta}$  results into the following optimization problem:

$$\begin{aligned} \phi = & \max_{\beta > 0} \min_{\substack{-t \leq x_j \leq t \\ j=1, \dots, K}} \\ \hat{\mathcal{H}}(\beta, \mathbf{x}) := & \frac{\beta \|\mathbf{g}\|}{\sqrt{K}} \sqrt{\frac{\rho_d}{K} (\sigma_{\tilde{H}}^2 \|\mathbf{x}\|^2 - 2\sigma_{\tilde{H}}^2 \mathbf{x}_0^T \mathbf{x} + \|\mathbf{x}_0\|^2) + 1} \\ & - \frac{\beta^2}{4} - \beta \sqrt{\rho_d} \sigma_{\tilde{H}} \frac{1}{K} \mathbf{x}^T \mathbf{q}^1 + \beta \frac{1}{K} \sqrt{\rho_d} \mathbf{x}_0^T \tilde{\mathbf{h}} + \lambda \rho_d \frac{1}{K} \|\mathbf{x}\|^2. \end{aligned} \quad (51)$$

Prior to proceeding further, we shall first check that the optimization over  $\beta$  of the above problem is not achieved in the limit  $\beta \rightarrow 0$  and more specifically, there exists  $\tilde{\delta} > 0$  such that taking the supremum over  $\beta > \tilde{\delta}$  instead of  $\beta > 0$  would almost surely not change the optimal cost of (51). Towards this goal, first note that

$$\left| \min_{\substack{-t \leq x_j \leq t \\ j=1, \dots, K}} \hat{\mathcal{H}}(\beta, \mathbf{x}) \right| \leq \left| \beta \frac{\|\mathbf{g}\|}{\sqrt{K}} - \frac{\beta^2}{4} + \beta \left| \frac{1}{K} \sqrt{\rho_d} \mathbf{x}_0^T \tilde{\mathbf{h}} \right| \right|,$$

and the function  $\beta \mapsto \beta \left( \frac{\|\mathbf{g}\|}{\sqrt{K}} + \frac{1}{K} \sqrt{\rho_d} \left| \mathbf{x}_0^T \tilde{\mathbf{h}} \right| \right) - \frac{\beta^2}{4}$  is increasing and positive for all  $\beta \in [0, 2 \left( \frac{\|\mathbf{g}\|}{\sqrt{K}} + \frac{1}{K} \sqrt{\rho_d} \left| \mathbf{x}_0^T \tilde{\mathbf{h}} \right| \right)]$ . As  $2 \left( \frac{\|\mathbf{g}\|}{\sqrt{K}} + \frac{1}{K} \sqrt{\rho_d} \left| \mathbf{x}_0^T \tilde{\mathbf{h}} \right| \right)$  converges in probability to  $2\sqrt{\delta}$ , with probability approaching one, for all  $\tilde{\delta} \in (0, \sqrt{\delta})$ , and all  $\beta \in (0, \tilde{\delta})$ ,

$$\left| \min_{\substack{-t \leq x_j \leq t \\ j=1, \dots, K}} \hat{\mathcal{H}}(\beta, \mathbf{x}) \right| \leq 4\tilde{\delta}\sqrt{\delta}. \quad (52)$$

To conclude it suffices to prove that there exists a  $\beta_0$  such that with probability approaching one,

$$\min_{\substack{-t \leq x_j \leq t \\ j=1, \dots, K}} \hat{\mathcal{H}}(\beta_0, \mathbf{x}) > \Theta, \quad (53)$$

where  $\Theta$  is a some positive constant. Indeed, if (53) is satisfied then almost surely,

$$\sup_{\beta > 0} \min_{\substack{-t \leq x_j \leq t \\ j=1, \dots, K}} \hat{\mathcal{H}}(\beta_0, \mathbf{x}) > \Theta.$$

Setting  $\tilde{\delta} = \min(\frac{\Theta}{4\sqrt{\delta}}, \sqrt{\delta})$  in (52), we conclude thus that the supremum over  $\beta$  could not be attained in the interval  $[0, \tilde{\delta}]$ . To keep the flow of the proof, (53) is proved in Appendix D. With this result at hand, we are now ready to proceed to the optimization of the (AO). Let

$$\chi = \frac{\rho_d}{K} (\sigma_{\tilde{H}}^2 \|\mathbf{x}\|^2 - 2\sigma_{\tilde{H}}^2 \mathbf{x}_0^T \mathbf{x} + \|\mathbf{x}_0\|^2) + 1. \quad (54)$$

To make the above optimization problem separable, we express the term in the square root in a variational form using the identity:

$$\sqrt{\chi} = \min_{r > 0} \frac{1}{2r} + \frac{r\chi}{2}. \quad (55)$$

Note that at optimum,  $r_* = \frac{1}{\sqrt{\chi}}$ . Using the fact that  $\chi \geq 1$  and as such bigger than any small positive constant, we also have

$$\sqrt{\chi} = \min_{0 < r \leq C'} \frac{1}{2r} + \frac{r\chi}{2},$$

where  $C'$  is any constant greater than 1. Similarly, as  $\chi$  is almost surely bounded by some constant, we can also argue that:

$$\sqrt{\chi} = \min_{\epsilon' < r \leq C'} \frac{1}{2r} + \frac{r\chi}{2}, \quad (56)$$

where  $\epsilon'$  is a sufficiently small positive constant. Using this relation, the optimization problem (50) becomes

$$\begin{aligned} \phi = & \max_{\beta \geq 0} \min_{\epsilon' < r \leq C'} \beta \left( \frac{\|\mathbf{g}\|}{2r\sqrt{K}} + \frac{r\|\mathbf{g}\|}{2\sqrt{K}} + \frac{\rho_d r \|\mathbf{g}\| \|\mathbf{x}_0\|^2}{2\sqrt{K}K} \right) \\ & + \sqrt{\rho_d} \frac{1}{K} \tilde{\mathbf{h}}^T \mathbf{x}_0 + \frac{1}{K} \sum_{j=1}^K \left[ \min_{-t \leq x_j \leq t} \left( \frac{\beta \sigma_{\tilde{H}}^2 \rho_d r \|\mathbf{g}\|}{2\sqrt{K}} + \lambda \rho_d \right) x_j^2 \right. \\ & \left. - \beta \left( \frac{\rho_d \sigma_{\tilde{H}}^2 r \|\mathbf{g}\|}{\sqrt{K}} x_{0,j} + \sqrt{\rho_d} \sigma_{\tilde{H}} q_j^1 \right) x_j \right] - \frac{\beta^2}{4}. \end{aligned} \quad (57)$$

For ease of notation, define  $\tilde{\theta} := \frac{r\|\mathbf{g}\|}{\sqrt{K}}$ . As  $\frac{r\|\mathbf{g}\|}{\sqrt{K}}$  is almost surely bounded above and below, the variable  $\tilde{\theta}$  is almost surely bounded above by a constant  $C$  which we shall assume as large as needed and also bounded below by some positive constant  $\epsilon$ . Introducing this notation leads to:

$$\begin{aligned} \phi = & \max_{\beta > 0} \min_{\epsilon \leq \tilde{\theta} \leq C} \frac{\beta \|\mathbf{g}\|^2}{2\tilde{\theta}K} + \frac{\beta\tilde{\theta}}{2} - \frac{\beta^2}{4} + \frac{\beta\tilde{\theta}\rho_d \|\mathbf{x}_0\|^2}{2K} \\ & + \sqrt{\rho_d} \beta \frac{1}{K} \mathbf{x}_0^T \tilde{\mathbf{h}} + \frac{1}{K} \sum_{j=1}^K \left[ \min_{-t \leq x_j \leq t} \left( \frac{\beta\tilde{\theta}\rho_d \sigma_{\tilde{H}}^2}{2} + \lambda \rho_d \right) x_j^2 \right. \\ & \left. - \beta \left( \rho_d \sigma_{\tilde{H}}^2 x_{0,j} \tilde{\theta} + \sqrt{\rho_d} \sigma_{\tilde{H}} q_j^1 \right) x_j \right]. \end{aligned} \quad (58)$$



For  $\beta > 0$ , the optimal solution in the variables  $x_j$ ,  $j = 1, \dots, K$  of (58) is given by:

$$\tilde{x}_j = \begin{cases} -t, & \text{if } q_j^1 < x_0^-(\tilde{\theta}, \beta, x_{0,j}), \\ t, & \text{if } q_j^1 > x_0^+(\tilde{\theta}, \beta, x_{0,j}), \\ \beta \left( \frac{\rho_d \sigma_{\tilde{H}}^2 x_{0,j} \tilde{\theta} + \sqrt{\rho_d} \sigma_{\tilde{H}} q_j^1}{\rho_d \sigma_{\tilde{H}}^2 \tilde{\theta} + 2\lambda \rho_d} \right), & \text{otherwise,} \end{cases} \quad (59)$$

where

$$x_0^-(\tilde{\theta}, \beta, x_{0,j}) = -t \left( \tilde{\theta} \sqrt{\rho_d} \sigma_{\tilde{H}} + \frac{2\lambda \rho_d}{\beta \sqrt{\rho_d} \sigma_{\tilde{H}}} \right) - \sqrt{\rho_d} \sigma_{\tilde{H}} x_{0,j} \tilde{\theta}, \quad (60)$$

$$x_0^+(\tilde{\theta}, \beta, x_{0,j}) = t \left( \tilde{\theta} \sqrt{\rho_d} \sigma_{\tilde{H}} + \frac{2\lambda \rho_d}{\sqrt{\rho_d} \sigma_{\tilde{H}} \beta} \right) - \sqrt{\rho_d} \sigma_{\tilde{H}} x_{0,j} \tilde{\theta}. \quad (61)$$

To simplify notation, define  $\xi = \sqrt{\rho_d} \sigma_{\tilde{H}}$ , then  $x_0^-(\tilde{\theta}, \beta, x_{0,j}) = -t(\xi \tilde{\theta} + \frac{2\lambda \rho_d}{\xi \beta}) - \xi x_{0,j} \tilde{\theta}$ , and  $x_0^+(\tilde{\theta}, \beta, x_{0,j}) = t(\xi \tilde{\theta} + \frac{2\lambda \rho_d}{\xi \beta}) - \xi x_{0,j} \tilde{\theta}$ . With these notations at hand, the above optimization problem reduces to the following SO

$$\begin{aligned} \max_{\beta > 0} \min_{\epsilon < \tilde{\theta} < C} \tilde{D}(\tilde{\theta}, \beta, \mathbf{g}, \mathbf{q}^1) &:= \frac{\beta \|\mathbf{g}\|^2}{2\tilde{\theta}K} + \frac{\beta \tilde{\theta}}{2} - \frac{\beta^2}{4} \\ &+ \frac{\beta \tilde{\theta} \rho_d \|\mathbf{x}_0\|^2}{2K} + \sqrt{\rho_d} \beta \frac{1}{K} \tilde{\mathbf{h}}^T \mathbf{x}_0 + \frac{1}{K} \sum_{j=1}^K v(\tilde{\theta}, \beta; q_j^1), \end{aligned} \quad (62)$$

where

$$v(\tilde{\theta}, \beta; q_j^1) = \begin{cases} t(\tilde{c}_j + \beta \xi q_j^1), & \text{if } q_j^1 < x_0^-(\tilde{\theta}, \beta, x_{0,j}), \\ t(\tilde{d}_j - \beta \xi q_j^1), & \text{if } q_j^1 > x_0^+(\tilde{\theta}, \beta, x_{0,j}), \\ -\frac{\beta^2 (\xi^2 x_{0,j} \tilde{\theta} + \xi q_j^1)^2}{2\xi^2 \beta \tilde{\theta} + 4\lambda \rho_d}, & \text{otherwise,} \end{cases} \quad (63)$$

where  $\tilde{c}_j = \frac{-\beta \xi}{2} x_0^-(\tilde{\theta}, \beta, x_{0,j}) + \frac{\beta \tilde{\theta} \xi^2}{2} x_{0,j}$ , and  $\tilde{d}_j = \frac{\beta \xi}{2} x_0^+(\tilde{\theta}, \beta, x_{0,j}) - \frac{\beta \tilde{\theta} \xi^2}{2} x_{0,j}$ .

### C. Asymptotic analysis of the SO problem

After simplifying the (AO) as in (62), we are now in a position to analyze its limiting behavior. Using the Weak Law of Large Numbers (WLLN) <sup>3</sup>,  $\frac{1}{K} \|\mathbf{g}\|^2 \xrightarrow{P} \frac{N}{K} := \delta$ ,  $\frac{1}{K} \mathbf{x}_0^T \tilde{\mathbf{h}} \xrightarrow{P} 0$ , and  $\frac{1}{K} \|\mathbf{x}_0\|^2 \xrightarrow{P} 1$ .

To analyze the behavior of the summand, recall that each  $x_{0,j}$  takes values  $\pm 1/\sqrt{\mathcal{E}}, \pm 3/\sqrt{\mathcal{E}}, \dots, \pm(M-1)/\sqrt{\mathcal{E}}$  with equal probability  $1/M$ . Let  $i = \pm 1, \pm 3, \dots, \pm(M-1)$  and denote by  $\ell_i = -t(\xi \tilde{\theta} + \frac{2\lambda \rho_d}{\xi \beta}) - \frac{\xi i \tilde{\theta}}{\sqrt{\mathcal{E}}}$ ,  $\mu_i = t(\xi \tilde{\theta} + \frac{2\lambda \rho_d}{\xi \beta}) - \frac{\xi i \tilde{\theta}}{\sqrt{\mathcal{E}}}$ ,  $c_i = \frac{-\beta \xi}{2} \ell_i + \frac{\beta \xi^2 i \tilde{\theta}}{2\sqrt{\mathcal{E}}}$ , and  $d_i = \frac{\beta \xi}{2} \mu_i - \frac{\beta \xi^2 i \tilde{\theta}}{2\sqrt{\mathcal{E}}}$ .

Hence, it can be shown that for all  $\tilde{\theta} > 0$  and  $\beta > 0$ ,

$$\frac{1}{K} \sum_{j=1}^K v(\tilde{\theta}, \beta; q_j^1) \xrightarrow{P} Y(\tilde{\theta}, \beta), \quad (64)$$

where

$$Y(\tilde{\theta}, \beta) := \frac{1}{M} \sum_{i=\pm 1, \pm 3, \dots, \pm(M-1)} \tilde{Y}(\tilde{\theta}, \beta, i), \quad (65)$$

<sup>3</sup>We write  $\xrightarrow{P}$  to denote convergence in probability as  $K \rightarrow \infty$ .

with

$$\begin{aligned} \tilde{Y}(\tilde{\theta}, \beta, i) &:= \mathbb{E}_{h \sim \mathcal{N}(0,1)} [v(\tilde{\theta}, \beta; h, \ell_i, \mu_i)] \\ &= -\frac{\beta^2}{2\xi^2 \beta \tilde{\theta} + 4\lambda \rho_d} \int_{\ell_i}^{\mu_i} \left( \frac{\xi^2 i \tilde{\theta}}{\sqrt{\mathcal{E}}} + \xi h \right)^2 p(h) dh \\ &+ \int_{-\infty}^{\ell_i} t(c_i + \beta \xi h) p(h) dh + \int_{\mu_i}^{\infty} t(d_i - \beta \xi h) p(h) dh \\ &= t(c_i Q(-\ell_i) + d_i Q(\mu_i)) - \beta \xi t(p(\ell_i) + p(\mu_i)) \\ &- \frac{\beta^2}{2\xi^2 \beta \tilde{\theta} + 4\lambda \rho_d} \int_{\ell_i}^{\mu_i} \left( \frac{\xi^2 i \tilde{\theta}}{\sqrt{\mathcal{E}}} + \xi h \right)^2 p(h) dh \\ &= t(c_i Q(-\ell_i) + d_i Q(\mu_i)) - \beta \xi t(p(\ell_i) + p(\mu_i)) \\ &- \frac{\beta^2}{2\xi^2 \beta \tilde{\theta} + 4\lambda \rho_d} \left( \left( \xi^2 + \frac{\xi^4 i^2 \tilde{\theta}^2}{\mathcal{E}} \right) (Q(\ell_i) - Q(\mu_i)) \right. \\ &\left. + \xi \left( \xi \ell_i + \frac{2\xi^2 i \tilde{\theta}}{\sqrt{\mathcal{E}}} \right) p(\ell_i) - \xi \left( \xi \mu_i + \frac{2\xi^2 i \tilde{\theta}}{\sqrt{\mathcal{E}}} \right) p(\mu_i) \right). \end{aligned} \quad (66)$$

For a given  $\beta$ , consider the sequence of functions

$$\varphi_K : \tilde{\theta} \mapsto \frac{1}{K} \sum_{j=1}^K v(\tilde{\theta}, \beta; q_j^1).$$

We can easily see that this sequence of functions is concave in  $\tilde{\theta}$  since it has been derived by taking the infimum of linear functions in  $\tilde{\theta}$  (cf. (58)). Hence,  $\tilde{\theta} \mapsto Y(\tilde{\theta}, \beta)$  is concave in  $\tilde{\theta}$ . Since the convergence of concave functions is uniform over compact sets [66, Theorem.II.1],  $\tilde{\theta} \mapsto \varphi_K(\tilde{\theta})$  converges uniformly to  $\tilde{\theta} \mapsto Y(\tilde{\theta}, \beta)$ . Moreover, it is easy to prove that  $\tilde{\theta} \mapsto \frac{\beta \|\mathbf{g}\|^2}{2\tilde{\theta}K} + \frac{\beta \tilde{\theta}}{2} + \frac{\beta \tilde{\theta} \rho_d \|\mathbf{x}_0\|^2}{2K}$  converges uniformly to  $\tilde{\theta} \mapsto \frac{\beta \delta}{2\tilde{\theta}} + \frac{\beta \tilde{\theta}}{2} + \frac{\beta \tilde{\theta} \rho_d}{2}$  on the compact set  $[\epsilon, C]$ . Combining both results yields that  $\tilde{\theta} \mapsto \tilde{D}(\tilde{\theta}, \beta, \mathbf{g}, \mathbf{q}^1)$  converges uniformly to  $\tilde{\theta} \mapsto \bar{D}(\tilde{\theta}, \beta)$ , where

$$\bar{D}(\tilde{\theta}, \beta) := \frac{\beta \delta}{2\tilde{\theta}} - \frac{\beta^2}{4} + \frac{\beta \tilde{\theta}}{2} + \frac{\beta \tilde{\theta} \rho_d}{2} + Y(\tilde{\theta}, \beta). \quad (67)$$

As a consequence,

$$\min_{\epsilon < \tilde{\theta} < C} \tilde{D}(\tilde{\theta}, \beta, \mathbf{g}, \mathbf{q}^1) \rightarrow \min_{\epsilon < \tilde{\theta} < C} \bar{D}(\tilde{\theta}, \beta). \quad (68)$$

We need now to prove that the supremum over  $\beta$  converges to the supremum of the right-hand side of the above equation.

To this end, note that the function  $\beta \mapsto \min_{\epsilon < \tilde{\theta} < C} \tilde{D}(\tilde{\theta}, \beta, \mathbf{g}, \mathbf{q}^1)$  is concave and converges pointwise to  $\beta \mapsto \min_{\epsilon < \tilde{\theta} < C} \bar{D}(\tilde{\theta}, \beta)$ .

Moreover, it is easy to check that  $\lim_{\beta \rightarrow \infty} \min_{\epsilon < \tilde{\theta} < C} \tilde{D}(\tilde{\theta}, \beta) = -\infty$ . Using Lemma 10 in [51], we conclude that:

$$\phi = \sup_{\beta \geq 0} \min_{\epsilon < \tilde{\theta} < C} \tilde{D}(\tilde{\theta}, \beta, \mathbf{g}, \mathbf{q}^1) \xrightarrow{P} \bar{\phi} := \sup_{\beta \geq 0} \min_{\epsilon < \tilde{\theta} < C} \bar{D}(\tilde{\theta}, \beta). \quad (69)$$

### D. Proof of the uniqueness of $(\tilde{\theta}_*, \beta_*)$ : Solving

$\sup_{\beta \geq 0} \min_{\epsilon < \tilde{\theta} < C} \bar{D}(\tilde{\theta}, \beta)$

Based on the above convergence, it follows from the CGMT that the optimal cost of the (PO) converges to the asymptotic limit of the (AO), which is given by

$$\sup_{\beta > 0} \min_{\epsilon \leq \tilde{\theta} < C} \bar{D}(\tilde{\theta}, \beta).$$

However, our interest does not directly concern the characterization of the asymptotic limit of the (PO) but that of functionals of the vector  $\mathbf{w} = \mathbf{x} - \mathbf{x}_0$  that can be linked to some important metrics like MSE or SEP. As explained in Remark 14, proving that the max-min problem in (69) admits a unique solution  $(\beta_*, \tilde{\theta}_*)$  would allow us to transfer any property of the solution of the (AO) to that of the (PO). Unfortunately, the objective in the max-min problem (69) is not convex in  $\tilde{\theta}$ , and hence the same approach pursued in [51] could not be used here. A new approach to handle this problem is thus proposed. To begin with, we notice that since  $\beta \mapsto -\frac{\beta^2}{4}$  is strictly concave,  $\beta \mapsto \min_{\epsilon < \tilde{\theta} < C} \overline{D}(\tilde{\theta}, \beta)$  is strictly concave in  $\beta$ . It thus has a unique maximum as it satisfies  $\lim_{\beta \rightarrow \infty} \min_{\epsilon < \tilde{\theta} < C} \overline{D}(\tilde{\theta}, \beta) = -\infty$ . Denote by  $\beta_*$  such a maximum. Let us prove that there exists a unique  $\theta_*$  that minimizes function  $h$  defined as  $h : \tilde{\theta} \mapsto \overline{D}(\tilde{\theta}, \beta_*)$ . The proof of this result will be carried out into the following steps:

- 1) First, we prove that the minimum should be in the interior domain of  $(\epsilon, C)$  for  $C$  sufficiently large and  $\epsilon$  sufficiently small.
- 2) Next, we establish that  $Y(\tilde{\theta}, \beta_*)$  satisfies:

$$\left| \tilde{\theta} \frac{\partial^3 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^3} \right| < 3 \left| \frac{\partial^2 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^2} \right|. \quad (70)$$

- 3) Starting from the observation that  $\tilde{\theta}_*$  is in the interior domain of the optimization set and based on the previously established results, we prove that  $h$  admits a unique minimum.

We start by establishing the first statement. It is obvious that the optimum could not be reached when  $\tilde{\theta}$  is in the vicinity of zero since  $\lim_{\tilde{\theta} \rightarrow 0^+} \overline{D}(\tilde{\theta}, \beta_*) = \infty$ . Similarly, to prove that the minimum is not reached when  $\tilde{\theta}$  grows to infinity, it suffices to check that  $\lim_{\tilde{\theta} \rightarrow \infty} \overline{D}(\tilde{\theta}, \beta_*) = \infty$ . Simple calculations lead to:

$$\begin{aligned} Y(\tilde{\theta}, \beta_*) \underset{\tilde{\theta} \rightarrow \infty}{\sim} -\frac{\beta_*^2 \xi^2 (M-1)^2}{M\mathcal{E}} \tilde{\theta} &= -\frac{3\beta_* \xi^2 (M-1)^2}{M(M^2-1)} \tilde{\theta} \\ &= -\frac{3\beta_* \rho_d \sigma_{\tilde{H}^2} (M-1)}{M(M+1)} \tilde{\theta}. \end{aligned}$$

Using this approximation, we thus have:

$$\overline{D}(\tilde{\theta}, \beta_*) \underset{\tilde{\theta} \rightarrow \infty}{\sim} \tilde{\theta} \beta_* \left( -\frac{3\rho_d \sigma_{\tilde{H}^2} (M-1)}{M(M+1)} + \frac{1}{2} + \frac{\rho_d}{2} \right). \quad (71)$$

It is easy to check that for  $M \geq 2$ ,  $\frac{3(M-1)}{M(M+1)} \leq \frac{1}{2}$ . As  $\sigma_{\tilde{H}^2} < 1$ , we thus have  $\lim_{\tilde{\theta} \rightarrow \infty} \overline{D}(\tilde{\theta}, \beta_*) = \infty$ .

To prove (70), we need to compute the first three derivatives of the function  $\tilde{\theta} \mapsto Y(\tilde{\theta}, \beta_*)$ . After simple calculations,

we can establish that:

$$\begin{aligned} \frac{\partial Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}} &= \frac{1}{M} \sum_{\substack{i=\pm 1, \\ \dots, \pm(M-1)}} \left[ \int_{\ell_i}^{\mu_i} 2\beta_*^3 \xi^2 \left( \frac{\xi^2 i \tilde{\theta}}{\sqrt{\mathcal{E}}} + \xi h \right)^2 p(h) dh \right. \\ &\quad \left. - \frac{2\beta_*^2 \xi^2 i}{\sqrt{\mathcal{E}}} \int_{\ell_i}^{\mu_i} \left( \frac{\xi^2 i \tilde{\theta}}{\sqrt{\mathcal{E}}} + \xi h \right) p(h) dh \right. \\ &\quad \left. + \int_{-\infty}^{\ell_i} t^2 \beta_* \frac{\xi^2}{2} p(h) dh + \int_{\mu_i}^{\infty} t^2 \beta_* \frac{\xi^2}{2} p(h) dh \right]. \end{aligned} \quad (72)$$

Based on this expression, we compute the second and third derivatives of  $Y(\tilde{\theta}, \beta_*)$  as:

$$\frac{\partial^2 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^2} = -2\beta_*^2 \frac{1}{M} \sum_{\substack{i=\pm 1, \dots, \\ \pm(M-1)}} \int_{\ell_i}^{\mu_i} \frac{(4\lambda \rho_d \frac{\xi^2 i}{\sqrt{\mathcal{E}}} - 2\xi^3 \beta_* h)^2}{(2\xi^2 \beta_* \tilde{\theta} + 4\lambda \rho_d)^3} p(h) dh, \quad (73)$$

$$\frac{\partial^3 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^3} = 6\beta_*^2 \frac{1}{M} \sum_{\substack{i=\pm 1, \dots, \\ \pm(M-1)}} \int_{\ell_i}^{\mu_i} \frac{2\xi^2 \beta_* (4\lambda \rho_d \frac{\xi^2 i}{\sqrt{\mathcal{E}}} - 2\xi^3 \beta_* h)^2}{(2\xi^2 \beta_* \tilde{\theta} + 4\lambda \rho_d)^4} p(h) dh. \quad (74)$$

Leveraging (73) and (74), it is easy to check that:

$$\begin{aligned} \tilde{\theta} \frac{\partial^3 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^3} &= -3 \frac{\partial^2 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^2} \\ &\quad - 24\beta_*^2 \frac{1}{M} \sum_{\substack{i=\pm 1, \\ \dots, \pm(M-1)}} \int_{\ell_i}^{\mu_i} \frac{4\lambda \rho_d (4\lambda \rho_d \frac{\xi^2 i}{\sqrt{\mathcal{E}}} - 2\xi^3 \beta_* h)^2}{(2\xi^2 \beta_* \tilde{\theta} + 4\lambda \rho_d)^4} p(h) dh, \end{aligned}$$

from which we deduce that

$$\left| \tilde{\theta} \frac{\partial^3 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^3} \right| < 3 \left| \frac{\partial^2 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^2} \right|. \quad (75)$$

With this result at hand, we are now ready to prove that function  $h$  admits a unique minimum. We already proved that any minimum should lie in the interior domain of  $(\epsilon, C)$ . Assume that there exists two minimizers of  $h$  which we denote by  $\tilde{\theta}_{*,1}$  and  $\tilde{\theta}_{*,2}$  such that  $\tilde{\theta}_{*,1} < \tilde{\theta}_{*,2}$ . The first order and second order conditions imply that:

$$\frac{dh}{d\tilde{\theta}} \Big|_{\tilde{\theta}=\tilde{\theta}_{*,i}} = 0 \quad \text{and} \quad \frac{d^2 h}{d\tilde{\theta}^2} \Big|_{\tilde{\theta}=\tilde{\theta}_{*,i}} \geq 0, \quad i = 1, 2. \quad (76)$$

Hence, there exists  $\tilde{\theta}_3 \in (\tilde{\theta}_{*,1}, \tilde{\theta}_{*,2})$  such that

$$\frac{d^2 h}{d\tilde{\theta}^2} \Big|_{\tilde{\theta}=\tilde{\theta}_3} = 0. \quad (77)$$

We will prove that this will lead to contradiction unless  $\tilde{\theta}_{*,1} = \tilde{\theta}_{*,2}$ . To this end, first we notice that:

$$\frac{d^2 h}{d\tilde{\theta}^2} = -\beta_* \delta \tilde{\theta}^{-3} + \frac{\partial^2 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^2}. \quad (78)$$

Hence, from (76) and (77) we obtain the following relations

$$\beta_* \tilde{\delta} + \tilde{\theta}_{*,1}^3 \frac{\partial^2 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^2} \Big|_{\tilde{\theta}=\tilde{\theta}_{*,1}} \geq 0, \quad (79)$$

$$\beta_* \tilde{\delta} + \tilde{\theta}_{*,2}^3 \frac{\partial^2 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^2} \Big|_{\tilde{\theta}=\tilde{\theta}_{*,2}} \geq 0, \quad (80)$$

$$\beta_* \tilde{\delta} + \tilde{\theta}_{*,3}^3 \frac{\partial^2 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^2} \Big|_{\tilde{\theta}=\tilde{\theta}_{*,3}} = 0. \quad (81)$$

Consider function  $k : \tilde{\theta} \mapsto \beta\tilde{\delta} + \tilde{\theta}^3 \frac{\partial^2 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^2}$ , the derivative of  $k$  with respect to  $\tilde{\theta}$  is given by:

$$k'(\tilde{\theta}) = \tilde{\theta}^2 \left( 3 \frac{\partial^2 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^2} + \tilde{\theta} \frac{\partial^3 Y(\tilde{\theta}, \beta_*)}{\partial \tilde{\theta}^3} \right). \quad (82)$$

From (70),  $k'(\tilde{\theta}) < 0$  and as such  $k$  is decreasing. Hence, the relations (79), (80) and (81) could not simultaneously hold. Hence,  $\tilde{\theta}_{*,1} = \tilde{\theta}_{*,2}$ , and as a consequence  $h$  admits a unique minimizer which we denote by  $\tilde{\theta}_*$ . Now, since for any  $\beta > 0$ ,  $\tilde{\theta} \mapsto \overline{D}(\tilde{\theta}, \beta)$  goes to infinity when  $\tilde{\theta}$  approaches zero or grows to infinity, we thus have:

$$\min_{\epsilon \leq \tilde{\theta} \leq C} \overline{D}(\tilde{\theta}, \beta) = \inf_{\tilde{\theta} \geq 0} \overline{D}(\tilde{\theta}, \beta), \quad \forall \beta > 0. \quad (83)$$

The above relation holds for all  $\beta > 0$ . We already proved in Section B that the optimal solution in  $\beta$  is almost surely away from zero. Thus,

$$\sup_{\beta > 0} \min_{\epsilon \leq \tilde{\theta} \leq C} \overline{D}(\tilde{\theta}, \beta) = \sup_{\beta > 0} \inf_{\tilde{\theta} > 0} \overline{D}(\tilde{\theta}, \beta). \quad (84)$$

### E. Asymptotic behavior of metrics depending on the solution of the (PO)

So far, we proved that the PO cost converges to the asymptotic cost of the (AO). We prove now that the uniqueness of the minimizer  $\tilde{\theta}_*$  allows us to carry over this convergence to metrics depending on the solution of the (PO). The recipe is as follows. Let  $\eta > 0$  and define

$$\mathcal{S}_\eta = \left\{ \mathbf{v} \mid \mathbf{v} \in \mathcal{D}, \frac{1}{K} \|\mathbf{v}\|^2 \in \left( \frac{\delta}{\tilde{\theta}_*^2} - 1 - \eta, \frac{\delta}{\tilde{\theta}_*^2} - 1 + \eta \right) \right\}. \quad (85)$$

Consider the ‘‘perturbed’’ version of the (AO) in (46) as follows:

$$\phi_\eta := \frac{1}{\sqrt{K}} \min_{\mathbf{v} \notin \mathcal{S}_\eta} \max_{\mathbf{u}} \frac{1}{K} \|\mathbf{v}\| \mathbf{g}^T \mathbf{u} - \frac{1}{K} \|\mathbf{u}\| \mathbf{q}^T \mathbf{v} - \frac{1}{\sqrt{K}} \mathbf{u}^T \mathbf{z} - \frac{1}{4\sqrt{K}} \|\mathbf{u}\|^2 + \frac{1}{\sqrt{K}} \lambda \|\mathbf{C}\mathbf{v}\|^2. \quad (86)$$

Following the same analysis carried out previously, we lower bound  $\phi_\eta$  by

$$\phi_\eta \geq \max_{\beta \geq 0} \min_{\epsilon \leq \tilde{\theta} \leq C} \tilde{D}(\tilde{\theta}, \beta, \mathbf{g}, \mathbf{q}^1). \quad \left| \frac{\|\mathbf{g}\|^2}{\tilde{\theta}^2 K} - \frac{\delta}{\tilde{\theta}_*^2} \right| > \eta$$

It is easy to note that if  $\tilde{\theta}$  in  $(\epsilon, C)$  is such that  $\left| \frac{\|\mathbf{g}\|^2}{K\tilde{\theta}^2} - \frac{\delta}{\tilde{\theta}_*^2} \right| > \eta$ , then

$$\left| \tilde{\theta} - \tilde{\theta}_* \frac{\|\mathbf{g}\|}{\sqrt{K}\sqrt{\delta}} \right| \geq \frac{\eta \tilde{\theta}^2 \tilde{\theta}_*^2}{\sqrt{\delta} (\tilde{\theta}_* \frac{\|\mathbf{g}\|}{\sqrt{K}} + \tilde{\theta} \sqrt{\delta})} \geq \frac{\eta \epsilon^2 \tilde{\theta}_*^2}{\sqrt{\delta} (C\sqrt{\delta} + \tilde{\theta}_* \frac{\|\mathbf{g}\|}{\sqrt{K}})},$$

or also

$$\left| \tilde{\theta}_* - \tilde{\theta} \right| > \frac{\eta \epsilon \tilde{\theta}_*}{\frac{1}{\epsilon} + \frac{\sqrt{\delta}}{\tilde{\theta}_*}} - \tilde{\theta} \left| 1 - \frac{\|\mathbf{g}\|}{\sqrt{K}\sqrt{\delta}} \right|.$$

Since  $\left| 1 - \frac{\|\mathbf{g}\|}{\sqrt{K}\sqrt{\delta}} \right|$  converges to zero almost surely, there exists  $\tilde{\eta} > 0$  such that with probability approaching 1,

$$\epsilon \leq \tilde{\theta} \leq C \text{ and } \left| \frac{\|\mathbf{g}\|^2}{K\tilde{\theta}^2} - \frac{\delta}{\tilde{\theta}_*^2} \right| \geq \eta \implies \left| \tilde{\theta} - \tilde{\theta}_* \right| \geq \tilde{\eta}.$$

Hence, with probability approaching 1,

$$\phi_\eta \geq \sup_{\beta \geq 0} \min_{\substack{\epsilon \leq \tilde{\theta} \leq C \\ |\tilde{\theta} - \tilde{\theta}_*| \geq \tilde{\eta}}} \tilde{D}(\tilde{\theta}, \beta, \mathbf{g}, \mathbf{q}^1). \quad (87)$$

Following the same asymptotic analysis as in Subsection C, we can prove similarly that

$$\sup_{\beta \geq 0} \min_{\substack{\epsilon \leq \tilde{\theta} \leq C \\ |\tilde{\theta} - \tilde{\theta}_*| \geq \tilde{\eta}}} \tilde{D}(\tilde{\theta}, \beta, \mathbf{g}, \mathbf{q}^1) \xrightarrow{P} \bar{\phi}_\eta := \sup_{\beta \geq 0} \inf_{\substack{\tilde{\theta} \geq 0 \\ |\tilde{\theta} - \tilde{\theta}_*| \geq \tilde{\eta}}} \overline{D}(\tilde{\theta}, \beta).$$

Clearly,

$$\bar{\phi}_\eta \geq \inf_{\substack{\tilde{\theta} \geq 0 \\ |\tilde{\theta} - \tilde{\theta}_*| \geq \tilde{\eta}}} \overline{D}(\tilde{\theta}, \beta_*).$$

As  $\tilde{\theta}_*$  is the unique minimizer of  $\inf_{\tilde{\theta} \geq 0} \overline{D}(\tilde{\theta}, \beta_*)$ ,

$$\bar{\phi}_\eta > \bar{\phi}.$$

Based on the CGMT in Theorem 6 and recalling Remark 14, we thus have:

$$\lim_{K \rightarrow \infty} \mathbb{P}[\mathbf{v}_{\text{PO}} \in \mathcal{S}_\eta] = 1, \quad (88)$$

where  $\mathbf{v}_{\text{PO}}$  is the solution in  $\mathbf{v}$  of the (PO) in (45), or equivalently:

$$\frac{1}{K} \|\mathbf{v}_{\text{PO}}\|^2 \xrightarrow{P} \frac{\delta}{\tilde{\theta}_*^2} - 1 = \delta \theta_*^2 - 1, \quad (89)$$

where  $\theta_* = \frac{1}{\tilde{\theta}_*}$ .

### F. Convergence of Lipschitz functions of the estimated vector $\hat{\mathbf{x}}$

The objective here is to study the asymptotic behavior of Lipschitz functions of the solution to the PO which we denote by  $\hat{\mathbf{x}}$ . As will be seen in the next section, such a result is fundamental for the asymptotic analysis of the symbol error rate and can be of independent interest to analyze any other performance metric. Let  $\beta_*$  and  $\tilde{\theta}_*$  be the unique solutions of the optimization problem  $\sup_{\beta > 0} \inf_{\tilde{\theta} > 0} \overline{D}(\tilde{\theta}, \beta)$ . Recall that  $\mathbf{x}_0$  represents the transmitted vector, whose elements are drawn with equal probability from the  $M$ -PAM constellation.

*Lemma 1.* For  $j = 1, \dots, K$ , and for fixed  $\tilde{\theta}$  and  $\beta$ , define  $\kappa_j(\cdot; \tilde{\theta}, \beta, x_{0,j}) : \mathbb{R} \rightarrow \mathbb{R}$  as:

$$\kappa_j(x; \tilde{\theta}, \beta, x_{0,j}) = \begin{cases} -t, & \text{if } x < x_0^-(\tilde{\theta}, \beta, x_{0,j}) \\ t, & \text{if } x > x_0^+(\tilde{\theta}, \beta, x_{0,j}) \\ \frac{\beta(\rho_d \sigma_H^2 x_{0,j} \tilde{\theta} + \sqrt{\rho_d} \sigma_H x)}{\rho_d \sigma_H^2 \beta \tilde{\theta} + 2\lambda \rho_d}, & \text{otherwise,} \end{cases}$$

where  $x_0^-(\tilde{\theta}, \beta, x_{0,j})$  and  $x_0^+(\tilde{\theta}, \beta, x_{0,j})$  are given by (60) and (61). Let  $\hat{\mathbf{x}}$  be the solution of the (PO). Then for all for all Lipschitz functions  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  with Lipschitz constant  $L$ , it holds:

$$\frac{1}{K} \sum_{i=1}^K \psi(\hat{x}_i) - \frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \psi(\kappa_j(q_j; \tilde{\theta}_*, \beta_*, x_{0,j})) \right] \xrightarrow{P} 0, \quad (90)$$

where  $q_1, \dots, q_K$  are independent and identically distributed standard Gaussian random variables and  $(\beta_*, \tilde{\theta}_*)$  is the unique solution to the optimization problem  $\sup_{\beta > 0} \inf_{\tilde{\theta} > 0} \overline{D}(\tilde{\theta}, \beta)$ .

*Proof:*

To avoid heavy notations, we will remove  $x_{0,j}$  from the notations of  $\kappa_j(\tilde{\theta}, \beta, x_{0,j})$  and that of  $x_0^-(\tilde{\theta}, \beta, x_{0,j})$  and  $x_0^+(\tilde{\theta}, \beta, x_{0,j})$  as it will not play any role in the proof. To prove (90), we consider the set:

$$\mathcal{S}_\epsilon = \left\{ \mathbf{x} \mid \left| \frac{1}{K} \sum_{i=1}^K \psi(x_i) - \frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \psi(\kappa_j(q_j; \tilde{\theta}_*, \beta_*)) \right] \right| < \epsilon \right\}. \quad (91)$$

Then, in view of the CGMT, for (90) to hold true, it suffices to show that with probability approaching 1,

$$\max_{\beta > 0} \min_{\substack{-t < x_j < t \\ j=1, \dots, K \\ \mathbf{x} \notin \mathcal{S}_\epsilon}} \hat{\mathcal{H}}(\beta, \mathbf{x}) \geq \bar{\phi}, \quad (92)$$

where we recall that  $\hat{\mathcal{H}}(\beta, \mathbf{x})$  is the objective of the optimization problem in (50). Noting that

$$\max_{\beta > 0} \min_{\substack{-t < x_j < t \\ j=1, \dots, K \\ \mathbf{x} \notin \mathcal{S}_\epsilon}} \hat{\mathcal{H}}(\beta, \mathbf{x}) \geq \min_{\substack{-t < x_j < t \\ j=1, \dots, K \\ \mathbf{x} \notin \mathcal{S}_\epsilon}} \hat{\mathcal{H}}(\beta_*, \mathbf{x}), \quad (93)$$

the proof of (92) boils down to proving that

$$\min_{\substack{-t < x_j < t \\ j=1, \dots, K \\ \mathbf{x} \notin \mathcal{S}_\epsilon}} \hat{\mathcal{H}}(\beta_*, \mathbf{x}) > \bar{\phi}. \quad (94)$$

A key step towards showing (94) is to analyze the asymptotic behavior of the following optimization problem:

$$\min_{\substack{-t < x_j < t \\ j=1, \dots, K}} \hat{\mathcal{H}}(\beta_*, \mathbf{x}). \quad (95)$$

Particularly, the following statements will be shown in the sequel:

(i) The following convergence holds true:

$$\min_{\substack{-t < x_j < t \\ j=1, \dots, K}} \hat{\mathcal{H}}(\beta_*, \mathbf{x}) \xrightarrow{P} \bar{\phi} = \sup_{\beta \geq 0} \min_{\tilde{\theta} \geq 0} \bar{D}(\tilde{\theta}, \beta) \quad (96)$$

(ii) Let  $\hat{\mathbf{x}} = [\hat{x}_1, \dots, \hat{x}_K]^T$  be the solution to the optimization problem in (95). Then, it holds with probability approaching 1,

$$\forall \mathbf{x} \in [-t, t]^K \setminus \mathcal{S}_\epsilon, \quad \frac{1}{\sqrt{K}} \|\mathbf{x} - \hat{\mathbf{x}}\| \geq \frac{\epsilon}{2L}. \quad (97)$$

(iii) For any  $\mathbf{x} = [x_1, \dots, x_K]^T \in [-t, t]^K$ , there exists a constant  $\bar{C}$  such that:

$$\hat{\mathcal{H}}(\beta_*, \mathbf{x}) \geq \hat{\mathcal{H}}(\beta_*, \hat{\mathbf{x}}) + \frac{\bar{C}}{2K} \|\mathbf{x} - \hat{\mathbf{x}}\|^2. \quad (98)$$

Prior to proving the above statements, let us see how they lead to the desired inequality (94). Putting together (96) and (98) shows that for any  $\ell > 0$ , with probability approaching 1,

$$\hat{\mathcal{H}}(\beta_*, \mathbf{x}) \geq \bar{\phi} - \ell + \frac{\bar{C}}{2K} \|\mathbf{x} - \hat{\mathbf{x}}\|^2. \quad (99)$$

From (97), we have for any  $\mathbf{x} \in [-t, t]^K \setminus \mathcal{S}_\epsilon$ ,

$$\frac{1}{K} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \geq \frac{\epsilon^2}{4L}. \quad (100)$$

Now, setting  $\ell = \frac{\bar{C}\epsilon^2}{16L}$  in (99), and using (100) we get:

$$\hat{\mathcal{H}}(\beta_*, \mathbf{x}) \geq \bar{\phi} + \frac{\bar{C}\epsilon^2}{16L}. \quad (101)$$

The above inequality holds for any  $\mathbf{x} \in [-t, t]^K \setminus \mathcal{S}_\epsilon$  hence,

$$\min_{\substack{-t < x_j < t \\ j=1, \dots, K \\ \mathbf{x} \notin \mathcal{S}_\epsilon}} \hat{\mathcal{H}}(\beta_*, \mathbf{x}) \geq \bar{\phi} + \frac{\bar{C}\epsilon^2}{16L} > \bar{\phi}, \quad (102)$$

thus proving (94).

Proof of (96): Following the same calculations used in the analysis of the (AO), we can prove that for sufficiently small  $\epsilon$  and large constant  $C$ , with probability approaching 1,

$$\min_{\substack{-t \leq x_j \leq t \\ j=1, \dots, K}} \hat{\mathcal{H}}(\beta_*, \mathbf{x}) = \min_{\epsilon \leq \tilde{\theta} \leq C} \bar{D}(\tilde{\theta}, \beta_*, \mathbf{g}, \mathbf{q}^1). \quad (103)$$

Based on similar asymptotic analysis to that carried out in Section B, we can prove that  $\tilde{\theta} \mapsto \bar{D}(\tilde{\theta}, \beta_*, \mathbf{g}, \mathbf{q}^1)$  converges uniformly to  $\tilde{\theta} \mapsto \bar{D}(\tilde{\theta}, \beta_*)$  and hence,

$$\min_{\substack{-t \leq x_j \leq t \\ j=1, \dots, K}} \hat{\mathcal{H}}(\beta_*, \mathbf{x}) \xrightarrow{P} \bar{\phi} = \min_{\epsilon \leq \tilde{\theta} \leq C} \bar{D}(\tilde{\theta}, \beta_*). \quad (104)$$

Proof of (97): Let  $\hat{\theta} \in \left\{ \arg \min_{\epsilon \leq \tilde{\theta} \leq C} \bar{D}(\tilde{\theta}, \beta_*, \mathbf{g}, \mathbf{q}^1) \right\}$ . In view of (59), for  $j = 1, \dots, K$ ,  $\left\{ \hat{x}_j = \kappa_j(q_j^1, \hat{\theta}, \beta_*) \right\}$  is a solution to (95). On the other hand, as the minimum of  $\tilde{\theta} \mapsto \bar{D}(\tilde{\theta}, \beta_*)$  is unique, we conclude that  $\hat{\theta}$  converges in probability to  $\tilde{\theta}_*$ . From this convergence, we argue that

$$\max_{1 \leq j \leq K} \left| \kappa_j(q_j^1, \hat{\theta}, \beta_*) - \kappa_j(q_j^1, \tilde{\theta}_*, \beta_*) \right| \xrightarrow{P} 0. \quad (105)$$

Prior to showing (105), let us explain how it leads to (97). Indeed from the Lipschitz assumption it holds that:

$$\max_{1 \leq j \leq K} \left| \psi(\kappa_j(q_j^1, \hat{\theta}, \beta_*)) - \psi(\kappa_j(q_j^1, \tilde{\theta}_*, \beta_*)) \right| \xrightarrow{P} 0. \quad (106)$$

Moreover, from the law of large numbers,

$$\frac{1}{K} \sum_{j=1}^K \psi(\kappa_j(q_j^1, \tilde{\theta}_*, \beta_*)) - \mathbb{E} \frac{1}{K} \sum_{j=1}^K \psi(\kappa_j(q_j^1, \tilde{\theta}_*, \beta_*)) \xrightarrow{P} 0,$$

which combined with (106) yields:

$$\frac{1}{K} \sum_{j=1}^K \psi(\kappa_j(q_j^1, \hat{\theta}, \beta_*)) - \mathbb{E} \frac{1}{K} \sum_{j=1}^K \psi(\kappa_j(q_j^1, \tilde{\theta}_*, \beta_*)) \xrightarrow{P} 0.$$

Hence, for  $\epsilon > 0$ , with probability approaching one,

$$\left| \frac{1}{K} \sum_{j=1}^K \psi(\kappa_j(q_j^1, \hat{\theta}, \beta_*)) - \mathbb{E} \frac{1}{K} \sum_{j=1}^K \psi(\kappa_j(q_j^1, \tilde{\theta}_*, \beta_*)) \right| \leq \epsilon.$$

By definition of the set  $\mathcal{S}_\epsilon$  and the triangle inequality, it holds with probability approaching 1 that for all  $\mathbf{x} \in [-t, t]^K \setminus \mathcal{S}_\epsilon$

$$\left| \frac{1}{K} \sum_{i=1}^K \psi(x_i) - \frac{1}{K} \sum_{j=1}^K \psi(\hat{x}_j) \right| \geq \frac{\epsilon}{2}. \quad (107)$$

Then, based on the Lipschitz property of  $\psi$ , it holds that:

$$\left\| \mathbf{x} - \hat{\mathbf{x}} \right\| \geq \frac{\epsilon}{2L},$$

which shows (97).

Now, to prove (105), note that since  $\hat{\theta} \xrightarrow{P} \tilde{\theta}_*$ , for any  $\eta > 0$ , with probability approaching one,

$$-\eta \leq \hat{\theta} - \tilde{\theta}_* \leq \eta, \quad (108)$$

from which we deduce that:

$$\max_{1 \leq j \leq K} \left| x_0^+(\hat{\theta}, \beta_*) - x_0^+(\tilde{\theta}_*, \beta_*) \right| \leq \bar{C}\eta,$$

where  $\bar{C} = \sqrt{\rho_d \sigma_{\hat{H}}} (\max_{1 \leq j \leq K} |x_{0,j}| + t)$  and similarly,

$$\left| x_0^-(\hat{\theta}, \beta_*) - x_0^-(\tilde{\theta}_*, \beta_*) \right| \leq \bar{C} \eta.$$

Using the fact that  $x_0^-(\tilde{\theta}_*, \beta_*) < x_0^+(\tilde{\theta}_*, \beta_*)$ , and choosing  $\eta$  sufficiently small, we conclude that if for some  $j = 1, \dots, K$ ,  $q_j^1 \leq x_0^-(\tilde{\theta}_*, \beta_*)$ , then  $q_j^1 \leq x_0^+(\tilde{\theta}, \beta_*)$ , and similarly, if  $q_j^1 \geq x_0^+(\tilde{\theta}_*, \beta_*)$ , then  $q_j^1 \geq x_0^-(\tilde{\theta}, \beta_*)$ . As a consequence,

$$\begin{aligned} & \max_{1 \leq j \leq K} \left| \kappa_j(q_j^1, \hat{\theta}, \beta_*) - \kappa_j(q_j^1, \tilde{\theta}_*, \beta_*) \right| \\ & \leq \left| t + \frac{\beta_* (\rho_d \sigma_{\hat{H}}^2 x_{0,j} \hat{\theta} + \sqrt{\rho_d} \sigma_{\hat{H}} q_j^1)}{\rho_d \sigma_{\hat{H}}^2 \beta_* \hat{\theta} + 2\lambda \rho_d} \right| \mathbf{1}_{\{x_0^-(\hat{\theta}, \beta_*) \leq q_j^1 \leq x_0^-(\tilde{\theta}_*, \beta_*)\}} \\ & + \left| t + \frac{\beta_* (\rho_d \sigma_{\hat{H}}^2 x_{0,j} \tilde{\theta}_* + \sqrt{\rho_d} \sigma_{\hat{H}} q_j^1)}{\rho_d \sigma_{\hat{H}}^2 \beta_* \tilde{\theta}_* + 2\lambda \rho_d} \right| \mathbf{1}_{\{x_0^-(\tilde{\theta}_*, \beta_*) \leq q_j^1 \leq x_0^-(\hat{\theta}, \beta_*)\}} \\ & + \left| t - \frac{\beta_* (\rho_d \sigma_{\hat{H}}^2 x_{0,j} \hat{\theta} + \sqrt{\rho_d} \sigma_{\hat{H}} q_j^1)}{\rho_d \sigma_{\hat{H}}^2 \beta_* \hat{\theta} + 2\lambda \rho_d} \right| \mathbf{1}_{\{x_0^+(\tilde{\theta}_*, \beta_*) \leq q_j^1 \leq x_0^+(\hat{\theta}, \beta_*)\}} \\ & + \left| t - \frac{\beta_* (\rho_d \sigma_{\hat{H}}^2 x_{0,j} \tilde{\theta}_* + \sqrt{\rho_d} \sigma_{\hat{H}} q_j^1)}{\rho_d \sigma_{\hat{H}}^2 \beta_* \tilde{\theta}_* + 2\lambda \rho_d} \right| \mathbf{1}_{\{x_0^+(\hat{\theta}, \beta_*) \leq q_j^1 \leq x_0^+(\tilde{\theta}_*, \beta_*)\}} \\ & + \left| \frac{\beta_* (\rho_d \sigma_{\hat{H}}^2 x_{0,j} \tilde{\theta}_* + \sqrt{\rho_d} \sigma_{\hat{H}} q_j^1)}{\rho_d \sigma_{\hat{H}}^2 \beta_* \tilde{\theta}_* + 2\lambda \rho_d} - \frac{\beta_* (\rho_d \sigma_{\hat{H}}^2 x_{0,j} \hat{\theta} + \sqrt{\rho_d} \sigma_{\hat{H}} q_j^1)}{\rho_d \sigma_{\hat{H}}^2 \beta_* \hat{\theta} + 2\lambda \rho_d} \right| \\ & \times \mathbf{1}_{\{\max(x_0^-(\tilde{\theta}_*, \beta_*), x_0^-(\hat{\theta}, \beta_*)) \leq q_j^1 \leq \min(x_0^+(\hat{\theta}, \beta_*), x_0^+(\tilde{\theta}_*, \beta_*))\}}. \quad (109) \end{aligned}$$

Each term of the right-hand side of (109) can be bounded by a linear function of  $\eta$  using (108), thereby proving (105). Proof of (98): It can be checked that  $\mathbf{x} \mapsto \hat{\mathcal{H}}(\beta_*, \mathbf{x})$  is strongly convex, and its Hessian satisfies  $\nabla^2 \hat{\mathcal{H}}(\beta_*, \mathbf{x}) \succeq \frac{2\lambda}{K} \rho_d \mathbf{I}_K$ . Hence, for any  $\mathbf{x} \in [-t, t]^K$ ,

$$\hat{\mathcal{H}}(\beta_*, \mathbf{x}) \geq \hat{\mathcal{H}}(\beta_*, \hat{\mathbf{x}}) + \frac{2\lambda}{K} \rho_d \left\| \mathbf{x} - \hat{\mathbf{x}} \right\|^2. \quad (110)$$

### G. From Lipschitz to the indicator function of solutions to the (PO)

*Lemma 2.* Let  $\hat{\mathbf{x}}$  be the solution to the (PO). Let  $c \in \mathbb{R}$  such that  $c \notin \{-t, t\}$ . Then,

$$\frac{1}{K} \sum_{j=1}^K \mathbf{1}_{\{\hat{x}_j \leq c\}} - \frac{1}{K} \sum_{j=1}^K \mathbb{P} \left[ \kappa_j(q, \tilde{\theta}_*, \beta_*, x_{0,j}) \leq c \right] \xrightarrow{P} 0, \quad (111)$$

where  $q$  is assumed to be drawn from a standard normal distribution.

*Proof:*

Similar to the proof of Lemma 1, for  $j = 1, \dots, K$ , to easy notations, we shall remove  $x_{0,j}$  from the notation  $\kappa_j(q, \tilde{\theta}_*, \beta_*, x_{0,j})$ . Let  $\eta > 0$ , and consider the following functions parametrized by  $\eta$ ,

$$\bar{\psi}_\eta(\alpha) := \begin{cases} 1, & \alpha \leq c \\ 1 - \frac{1}{\eta}(\alpha - c), & c \leq \alpha \leq c + \eta \\ 0, & \alpha \geq c + \eta, \end{cases}$$

and

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

Both functions are Lipschitz, with a Lipschitz constant  $\frac{1}{\eta}$ . Moreover, for any  $\alpha$ ,

$$\underline{\psi}_\eta(\alpha) \leq \mathbf{1}_{\{\alpha \leq c\}} \leq \bar{\psi}_\eta(\alpha).$$

Also, define the function:  $\tilde{\psi}_\eta(\alpha) = \bar{\psi}_\eta(\alpha) - \underline{\psi}_\eta(\alpha)$ . Then, it is easy to see that:

$$\tilde{\psi}_\eta(\alpha) \leq \mathbf{1}_{\{c - \eta \leq \alpha \leq c + \eta\}}. \quad (112)$$

Hence,

$$\frac{1}{K} \sum_{j=1}^K \underline{\psi}_\eta(\hat{x}_j) \leq \frac{1}{K} \sum_{j=1}^K \mathbf{1}_{\{\hat{x}_j \leq c\}} \leq \frac{1}{K} \sum_{j=1}^K \bar{\psi}_\eta(\hat{x}_j), \quad (113)$$

and

$$\begin{aligned} & \frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \underline{\psi}_\eta(\kappa_j(q, \tilde{\theta}_*, \beta_*)) \right] \leq \frac{1}{K} \sum_{j=1}^K \mathbb{P} \left[ \kappa_j(q, \tilde{\theta}_*, \beta_*) \leq c \right] \\ & \leq \frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \bar{\psi}_\eta(\kappa_j(q, \tilde{\theta}_*, \beta_*)) \right]. \quad (114) \end{aligned}$$

Using (113) and (114), it follows that:

$$\begin{aligned} & \frac{1}{K} \sum_{j=1}^K \underline{\psi}_\eta(\hat{x}_j) - \frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \bar{\psi}_\eta(\kappa_j(q, \tilde{\theta}_*, \beta_*)) \right] \\ & \leq \frac{1}{K} \sum_{j=1}^K \mathbf{1}_{\{\hat{x}_j \leq c\}} - \frac{1}{K} \sum_{j=1}^K \mathbb{P} \left[ \kappa_j(q, \tilde{\theta}_*, \beta_*) \leq c \right] \\ & \leq \frac{1}{K} \sum_{j=1}^K \bar{\psi}_\eta(\hat{x}_j) - \frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \bar{\psi}_\eta(\kappa_j(q, \tilde{\theta}_*, \beta_*)) \right]. \quad (115) \end{aligned}$$

From Lemma 1, for  $\epsilon > 0$ , with probability approaching one:

$$\left| \frac{1}{K} \underline{\psi}_\eta(\hat{x}_j) - \frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \underline{\psi}_\eta(\kappa_j(q, \tilde{\theta}_*, \beta_*)) \right] \right| \leq \epsilon, \quad (116)$$

and

$$\left| \frac{1}{K} \bar{\psi}_\eta(\hat{x}_j) - \frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \bar{\psi}_\eta(\kappa_j(q, \tilde{\theta}_*, \beta_*)) \right] \right| \leq \epsilon. \quad (117)$$

Hence,

$$\begin{aligned} & -\epsilon - \frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \tilde{\psi}_\eta(\kappa_j(q, \tilde{\theta}_*, \beta_*)) \right] \\ & \leq \frac{1}{K} \sum_{j=1}^K \mathbf{1}_{\{\hat{x}_j \leq c\}} - \frac{1}{K} \sum_{j=1}^K \mathbb{P} \left[ \kappa_j(q, \tilde{\theta}_*, \beta_*) \leq c \right] \\ & \leq \epsilon + \frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \tilde{\psi}_\eta(\kappa_j(q, \tilde{\theta}_*, \beta_*)) \right]. \quad (118) \end{aligned}$$

From (112),

$$\begin{aligned} & \lim_{\eta \rightarrow 0} \frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \tilde{\psi}_\eta(\kappa_j(q, \tilde{\theta}_*, \beta_*)) \right] \\ & \leq \lim_{\eta \rightarrow 0} \max_{1 \leq j \leq K} \mathbb{P} \left[ c - \eta \leq \kappa_j(q, \tilde{\theta}_*, \beta_*) \leq c + \eta \right]. \end{aligned}$$

As  $c \notin \{-t, t\}$ , the right-hand side of the above inequality converges to zero. Hence, there exists  $\eta_0$  such that for all  $\eta \leq \eta_0$ ,

$$\frac{1}{K} \sum_{j=1}^K \mathbb{E} \left[ \tilde{\psi}_\eta(\kappa_j(q, \tilde{\theta}_*, \beta_*)) \right] \leq \epsilon. \quad (119)$$



Combining (118) and (119), we get the desired result, that is that for  $\epsilon > 0$ , with probability approaching 1,

$$\left| \frac{1}{K} \sum_{j=1}^K \mathbf{1}_{\{\hat{x}_j \leq c\}} - \frac{1}{K} \sum_{j=1}^K \mathbb{P} \left[ \kappa_j(q, \tilde{\theta}_*, \beta_*) \leq c \right] \right| \leq 2\epsilon. \quad (120)$$

**Corollary 2.** Consider the setting of Lemma 2. We thus have:

$$\frac{1}{K} \sum_{j=1}^K \mathbb{P} [\hat{x}_j \leq c] - \frac{1}{K} \sum_{j=1}^K \mathbb{P} \left[ \kappa_j(q, \tilde{\theta}_*, \beta_*, x_{0,j}) \leq c \right] \rightarrow 0.$$

*Proof:*

The proof follows straightforwardly by applying the dominated convergence theorem. ■

### H. Applying the CGMT: MSE of Box-RLS

Let  $\hat{\mathbf{x}}$  be the solution of (12a). Recall that the MSE is given by:

$$\text{MSE} = \frac{1}{K} \|\hat{\mathbf{x}} - \mathbf{x}_0\|^2,$$

which can be also written as:

$$\text{MSE} = \frac{1}{\rho_d \sigma_H^2} \left( \frac{1}{K} \|\mathbf{v}\|^2 - \frac{\rho_d}{K} \sigma_\Delta^2 \|\mathbf{x}_0\|^2 \right). \quad (121)$$

As  $\frac{1}{K} \|\mathbf{v}\|^2 \xrightarrow{P} \delta \theta_*^2 - 1$ , and  $\frac{1}{K} \|\mathbf{x}_0\|^2 \xrightarrow{P} 1$ , we thus have:

$$\text{MSE} \xrightarrow{P} \frac{1}{\rho_d \sigma_H^2} (\delta \theta_*^2 - 1 - \rho_d \sigma_\Delta^2). \quad (122)$$

### I. Applying the CGMT: SEP for Box-RLS

In this subsection, we study the limiting behavior of the SEP defined in (15). For  $j = 1, \dots, K$ , consider the output of the (PO) problem which denoted by  $\hat{\mathbf{x}}$ . Recall the expression of the SEP

$$\text{SEP} = \frac{1}{K} \sum_{j=1}^K \mathbb{P} \left[ \hat{s} \neq x_{0,j}, \text{ with } \hat{s} = \arg \min_{s \in \mathcal{C}} \left| \frac{\hat{x}_j}{B} - s \right| \right].$$

In PAM-constellations, we distinguish inner symbols when  $x_{0,j}$  belongs to  $\frac{1}{\sqrt{\mathcal{E}}} \{\pm 1, \pm 3, \dots, \pm(M-3)\}$  from edge symbols when  $x_{0,j} = \pm \frac{M-1}{\sqrt{\mathcal{E}}}$ . Let us consider first the case where  $x_{0,j}$  is a certain inner symbol. In this case, there is an error if and only if  $|\frac{\hat{x}_j}{B} - x_{0,j}| > \frac{1}{\sqrt{\mathcal{E}}}$ , where  $B = \frac{\xi^2 \beta_* \tilde{\theta}_*}{\xi^2 \beta_* \tilde{\theta}_* + 2\lambda \rho_d}$ . Similarly, considering the edge points  $x_{0,j} = \pm \frac{M-1}{\sqrt{\mathcal{E}}}$ , we deduce that there is an error if  $|\frac{\hat{x}_j}{B} - x_{0,j}| > \frac{1}{\sqrt{\mathcal{E}}}$  when  $x_{0,j} = \frac{M-1}{\sqrt{\mathcal{E}}}$  and if  $|\frac{\hat{x}_j}{B} - x_{0,j}| < -\frac{1}{\sqrt{\mathcal{E}}}$  when  $x_{0,j} = -\frac{M-1}{\sqrt{\mathcal{E}}}$ . The SEP thus becomes:

$$\begin{aligned} \text{SEP} &= \frac{1}{K} \sum_{j=1}^K \frac{1}{M} \sum_{i=\pm 1, \pm 3, \dots, \pm(M-2)} \mathbb{P} \left[ \left| \frac{\hat{x}_j}{B} - x_{0,j} \right| > \frac{1}{\sqrt{\mathcal{E}}} \mid x_{0,j} = \frac{i}{\sqrt{\mathcal{E}}} \right] \\ &+ \frac{1}{K} \sum_{j=1}^K \frac{1}{M} \mathbb{P} \left[ \frac{\hat{x}_j}{B} - x_{0,j} < -\frac{1}{\sqrt{\mathcal{E}}} \mid x_{0,j} = \frac{M-1}{\sqrt{\mathcal{E}}} \right] \\ &+ \frac{1}{K} \sum_{j=1}^K \frac{1}{M} \mathbb{P} \left[ \frac{\hat{x}_j}{B} - x_{0,j} > \frac{1}{\sqrt{\mathcal{E}}} \mid x_{0,j} = -\frac{M-1}{\sqrt{\mathcal{E}}} \right]. \end{aligned}$$

Based on Corollary 2, we have

$$\text{SEP} - \widetilde{\text{SEP}} \rightarrow 0,$$

where  $\widetilde{\text{SEP}}$  follows after some tedious but straightforward calculations and is given in (123).

### Appendix C Un-Boxed RLS Proofs

The analysis of the Un-Boxed RLS scheme is similar to the analysis of the Box-RLS. Below we provide a brief sketch of the proof. Following the same analysis as before, we identify the same (PO) and (AO) with the single difference that the constraints on  $\{x_j\}_{j=1}^K$  are now removed. Particularly, the (AO) associated with the RLS writes as:

$$\begin{aligned} \phi &= \max_{\beta > 0} \min_{0 \leq \tilde{\theta} \leq C} \frac{\beta \|\mathbf{g}\|^2}{2\tilde{\theta}K} + \frac{\beta \tilde{\theta}}{2} - \frac{\beta^2}{4} + \frac{\beta \tilde{\theta} \rho_a \|\mathbf{x}_0\|^2}{2K} \\ &+ \sqrt{\rho_d} \beta \frac{1}{K} \mathbf{x}_0^T \tilde{\mathbf{h}} + \frac{1}{K} \sum_{j=1}^K \left[ \min_{x_j \in \mathbb{R}} \left( \frac{\beta \tilde{\theta} \rho_a \sigma_H^2}{2} + \lambda \rho_a \right) x_j^2 \right. \\ &\left. - \beta \left( \rho_a \sigma_H^2 x_{0,j} \tilde{\theta} + \sqrt{\rho_d} \sigma_H q_j^1 \right) x_j \right], \quad (124) \end{aligned}$$

which is similar to (58) with the difference that the optimization over  $x_j$  is on the whole real line. Optimizing over the variables  $x_j, j = 1, \dots, K$ , we thus obtain:

$$\begin{aligned} \phi &= \max_{\beta > 0} \min_{0 \leq \tilde{\theta} \leq C} \frac{\beta \|\mathbf{g}\|^2}{2\tilde{\theta}K} + \frac{\beta \tilde{\theta}}{2} - \frac{\beta^2}{4} + \frac{\beta \tilde{\theta} \rho_a \|\mathbf{x}_0\|^2}{2K} \\ &+ \sqrt{\rho_d} \beta \frac{1}{K} \mathbf{x}_0^T \tilde{\mathbf{h}} - \frac{1}{K} \sum_{j=1}^K \frac{\beta^2 (\xi^2 x_{0,j} \tilde{\theta} + \xi q_j^1)^2}{2\xi^2 \beta \tilde{\theta} + 4\lambda \rho_a}. \quad (125) \end{aligned}$$

Using the same approach as that in the Box-RLS, we can prove that  $\phi$  converges to:

$$\phi \xrightarrow{P} \bar{\phi} := \sup_{\beta > 0} \inf_{\tilde{\theta} > 0} \frac{\beta \delta}{2\tilde{\theta}} + \frac{\beta \tilde{\theta}}{2} - \frac{\beta^2}{4} + \frac{\beta \tilde{\theta} \rho_a}{2} - \frac{\beta^2 \xi^2 \tilde{\theta}^2 + \xi^2}{2\xi^2 \beta \tilde{\theta} + 4\lambda \rho_a}. \quad (126)$$

Using the change of variable  $\theta = \frac{1}{\tilde{\theta}}$ , this brings us to solve the following max-min problem:

$$\bar{\phi} = \sup_{\beta > 0} \inf_{\theta > 0} \beta \theta \frac{\delta}{2} + \frac{\beta}{2\theta} (1 + \rho_a) - \frac{\beta^2}{4} - \frac{\beta^2 \sigma_H^2}{2\beta \sigma_H^2 + 4\lambda} \left( 1 + \frac{\rho_a \sigma_H^2}{\theta^2} \right),$$

for which we prove that there exists a unique solution  $(\theta_*, \beta_*)$ . As a matter of fact, writing the first order optimality conditions, we obtain:

$$\delta \beta - \frac{\beta}{\theta^2} - \frac{\rho_a \sigma_\Delta^2 \beta}{\theta^2} - \frac{\beta \sigma_H^2 (\beta^2 \sigma_H^2 + 4\rho_a \lambda^2)}{(\beta \sigma_H^2 + 2\lambda \theta)^2} = 0, \quad (127a)$$

$$\delta \theta + \frac{1}{\theta} - \beta + \frac{\rho_a \sigma_\Delta^2}{\theta} - \frac{\sigma_H^2 \theta (\beta^2 \sigma_H^2 + 4\lambda \theta \beta - 4\rho_a \lambda^2)}{(\beta \sigma_H^2 + 2\lambda \theta)^2} = 0. \quad (127b)$$

Combining the two equations together  $(\frac{1}{\beta} (127a) + \frac{1}{\theta} (127b))$ , gives

$$\delta - \frac{\beta}{2\theta} - \frac{\beta \sigma_H^2}{\beta \sigma_H^2 + 2\lambda \theta} = 0, \quad (128)$$

$$\begin{aligned} \widetilde{\text{SEP}} &= \frac{4}{M} \sum_{i=1,3,\dots,M-3} \mathbf{1}_{\{\frac{i}{B} \geq \frac{i+1}{\sqrt{\mathcal{E}}}\}} Q\left(\frac{\xi \tilde{\theta}_*}{\sqrt{\mathcal{E}}}\right) + \frac{2}{M} \sum_{i=1,3,\dots,M-3} \left\{ \mathbf{1}_{\{\frac{i-1}{\sqrt{\mathcal{E}}} \leq \frac{i}{B} \leq \frac{i+1}{\sqrt{\mathcal{E}}}\}} Q\left(\frac{\xi \tilde{\theta}_*}{\sqrt{\mathcal{E}}}\right) + \mathbf{1}_{\{\frac{i}{B} \leq \frac{i-1}{\sqrt{\mathcal{E}}}\}} \right\} \\ &+ \frac{2}{M} \mathbf{1}_{\{\frac{i}{B} \geq \frac{M-2}{\sqrt{\mathcal{E}}}\}} Q\left(\frac{\xi \tilde{\theta}_*}{\sqrt{\mathcal{E}}}\right) + \frac{2}{M} \mathbf{1}_{\{\frac{i}{B} \leq \frac{M-2}{\sqrt{\mathcal{E}}}\}}. \end{aligned} \quad (123)$$

which is equivalent to a second order polynomial in  $\theta$ , the solution of which is given by:

$$\theta = \frac{1}{2\lambda'} \Upsilon(\lambda', \delta) \beta, \quad (129)$$

where,  $\Upsilon(\lambda', \delta) = \frac{-(\delta - \lambda' - 1) + \sqrt{(\delta - \lambda' - 1)^2 + 4\lambda'\delta}}{2\delta}$ , and  $\lambda' = \frac{\lambda}{\sigma_{\hat{H}}^2}$ . Substituting (129) into (127) and after some algebraic manipulations, the unique solution  $(\theta_*, \beta_*)$  of the system is

$$\theta_* = \sqrt{\frac{\rho_d \sigma_{\hat{H}}^2 \kappa^2 + \rho_d \sigma_{\Delta}^2 + 1}{\delta - (1 - \kappa)^2}}, \quad (130a)$$

$$\beta_* = 2 \left( (\delta - \lambda' - 1) + \delta \Upsilon(\lambda', \delta) \right) \theta_*, \quad (130b)$$

where  $\kappa = \frac{\Upsilon(\lambda', \delta)}{1 + \Upsilon(\lambda', \delta)}$ . Substituting  $\kappa$  back gives the same expression as  $\theta_*$  in Theorem 1. Similar to the Box-RLS case, we can obtain the same MSE and SEP expressions as before. The only difference is that  $\theta_*$  and  $\beta_*$  are now given by the closed form expressions in (130a) and (130b), respectively.

#### Appendix D

##### Proof of (53)

To begin with, we perform the change of variable  $\mathbf{s} = \sigma_{\hat{H}} \mathbf{x} - \mathbf{x}_0$ . Function  $\hat{\mathcal{H}}(\beta, \mathbf{x})$  can be lower-bounded as:

$$\begin{aligned} \hat{\mathcal{H}}(\beta, \mathbf{x}) &\geq \beta \frac{\|\mathbf{g}\|}{\sqrt{K}} \sqrt{\rho_d} - \frac{\tilde{\beta}^2}{4} - \tilde{\beta} \sqrt{\rho_d} \frac{\mathbf{s}^T \mathbf{q}^1}{K} - \beta \sqrt{\rho_d} \frac{\mathbf{x}_0^T \mathbf{q}^1}{K} \\ &+ \beta \sqrt{\rho_d} \frac{1}{K} \mathbf{x}_0^T \tilde{\mathbf{h}} + \frac{\lambda \rho_d}{\sigma_{\hat{H}}^2} \frac{1}{K} (\|\mathbf{s}\|^2 + \|\mathbf{x}_0\|^2 + 2\mathbf{s}^T \mathbf{x}_0). \end{aligned}$$

Denote by  $\kappa$  the norm of  $\frac{1}{\sqrt{K}} \|\mathbf{s}\|$ . Minimizing the lower-bound over  $\mathbf{s}$ , we obtain:

$$\begin{aligned} \hat{\mathcal{H}}(\beta, \mathbf{x}) &\geq \min_{\kappa \geq 0} \tilde{\beta} \frac{\|\mathbf{g}\|}{\sqrt{K}} \sqrt{\rho_d} - \frac{\beta^2}{4} - \beta \sqrt{\rho_d} \frac{\mathbf{x}_0^T \mathbf{q}^1}{K} + \tilde{\beta} \sqrt{\rho_d} \frac{1}{K} \mathbf{x}_0^T \tilde{\mathbf{h}} \\ &+ \frac{\lambda \rho_d}{\sigma_{\hat{H}}^2} \kappa^2 + \frac{\lambda \rho_d}{\sigma_{\hat{H}}^2} \|\mathbf{x}_0\|^2 - \kappa \tilde{\beta} \sqrt{\rho_d} \frac{1}{\sqrt{K}} \mathbf{q}^1 - 2 \frac{\lambda \rho_d}{\sqrt{K} \sigma_{\hat{H}}^2} \mathbf{x}_0 \|\mathbf{s}\| \\ &= \beta \frac{\|\mathbf{g}\|}{\sqrt{K}} \sqrt{\rho_d} - \frac{\beta^2}{4} - \beta \sqrt{\rho_d} \frac{\mathbf{x}_0^T \mathbf{q}^1}{K} + \beta \sqrt{\rho_d} \frac{1}{K} \mathbf{x}_0^T \tilde{\mathbf{h}} + \frac{\lambda \rho_d}{\sigma_{\hat{H}}^2} \|\mathbf{x}_0\|^2 \\ &- \frac{\sigma_{\hat{H}}^2}{4\lambda \rho_d K} \|2 \frac{\lambda \rho_d}{\sigma_{\hat{H}}^2} \mathbf{x}_0 - \beta \sqrt{\rho_d} \mathbf{q}^1\|^2. \end{aligned}$$

In the asymptotic regime with  $K$  and  $N$  tending to infinity, and taking  $\beta$  as fixed, the right-hand side of the above equation converges to:

$$\beta \sqrt{\rho_d} \sqrt{\delta} - \frac{\beta^2}{4} - \beta^2 \frac{\sigma_{\hat{H}}^2}{4\lambda}, \quad (131)$$

which simplifies to  $\frac{\rho_d \delta}{1 + \frac{\sigma_{\hat{H}}^2}{4\lambda}}$  when  $\beta$  is replaced by  $\beta_0 :=$

$\frac{2\sqrt{\rho_d} \sqrt{\delta}}{1 + \frac{\sigma_{\hat{H}}^2}{4\lambda}}$ . We thus obtain almost surely

$$\hat{\mathcal{H}}(\beta_0, \mathbf{x}) \geq \frac{1}{2} \frac{\rho_d \delta}{1 + \frac{\sigma_{\hat{H}}^2}{4\lambda}},$$

and as such, almost surely,

$$\min_{\substack{-t \leq x_j \leq t \\ j=1, \dots, K}} \hat{\mathcal{H}}(\beta_0, \mathbf{x}) \geq \frac{1}{2} \frac{\rho_d \delta}{1 + \frac{\sigma_{\hat{H}}^2}{4\lambda}}. \quad (132)$$

Setting  $\Theta = \frac{1}{2} \frac{\rho_d \delta}{1 + \frac{\sigma_{\hat{H}}^2}{4\lambda}}$  ends up the proof of (53).

#### Appendix E

##### Monotonicity of the MSE and SEP of the RLS decoder

We showed that for the LS case, optimizing the power allocation in MSE sense is equivalent to optimizing the SEP and it boils down to maximizing  $\rho_{\text{eff}}$ .

In this appendix, we will show that this holds also true for the RLS decoder that employs optimal regularization coefficient. Towards this goal, we proceed with the following change of variables  $J = \frac{1}{\rho_{\text{eff}}}$ ,  $c_1 = 2(1+\delta)$ , and  $c_2 = (1-\delta)^2$ . Then, the MSE and SEP write as:

$$\widetilde{\text{MSE}}_{\text{RLS}} = \frac{1}{2} \left( -J + (1 - \delta) + \sqrt{J^2 + c_1 J + c_2} \right), \quad (133)$$

$$\begin{aligned} \widetilde{\text{SEP}}_{\text{RLS}} &= 2 \left( 1 - \frac{1}{M} \right) \\ &\times Q \left( \sqrt{\frac{\delta}{\mathcal{E} \left( \frac{1}{2} \left( J + (1 - \delta) + \sqrt{J^2 + c_1 J + c_2} \right) \right)}} \right). \end{aligned} \quad (134)$$

It appears from (133) and (134) that to minimize the MSE or the SEP, it suffices to minimize for all  $i \in \{0, 1\}$ , function  $\mathcal{F}_i(J) = (-1)^i J + \sqrt{J^2 + c_1 J + c_2}$ . We can check easily that the first derivative of  $\mathcal{F}_i$  is given by:  $\mathcal{F}'_i(J) = (-1)^i + \frac{2J + c_1}{2\sqrt{J^2 + c_1 J + c_2}}$  which is strictly positive for all  $i \in \{0, 1\}$ . It follows thus that both  $\widetilde{\text{MSE}}_{\text{RLS}}$  and  $\widetilde{\text{SEP}}_{\text{RLS}}$  are increasing functions of  $\frac{1}{\rho_{\text{eff}}}$  and hence minimizing them amounts to maximizing  $\rho_{\text{eff}}$ . Then,

$$\widetilde{\text{MSE}}_{\text{RLS}} = \frac{1}{2} \left( -J + (1 - \delta) + \sqrt{J^2 + c_1 J + c_2} \right).$$

On the other hand, from (30), it is clear that to minimize the SEP we need to maximize the argument of the  $Q$ -function, which means to minimize  $\widetilde{\text{MSE}}_{\text{RLS}} + J$ .

## Appendix F

### Optimal Power Allocation Derivation: Proof of Theorem 5

Here, we derive the optimal power allocation given in Theorem 5. First, rewrite  $\rho_{\text{eff}}$  as follows

$$\begin{aligned}\rho_{\text{eff}} &= \frac{\alpha \rho \frac{\tau}{\tau_d} \cdot (1 - \alpha) \rho \tau}{(1 + \alpha \rho \frac{\tau}{\tau_d}) + (1 - \alpha) \rho \tau} \\ &= \frac{(\rho \tau)^2}{\tau_d} \cdot \frac{\alpha(1 - \alpha)}{1 + \rho \tau - \alpha \rho \tau (1 - \frac{1}{\tau_d})} \\ &= \frac{\rho \tau}{\tau_d - 1} \cdot \frac{\alpha(1 - \alpha)}{-\alpha + \frac{1 + \rho \tau}{\rho \tau (1 - \frac{1}{\tau_d})}}.\end{aligned}$$

We need to maximize  $\rho_{\text{eff}}$  over  $0 < \alpha < 1$ . To do so, we consider the following cases:

1)  $\tau_d = 1$ :

$$\rho_{\text{eff}} = \frac{(\rho \tau)^2}{1 + \rho \tau} \alpha(1 - \alpha).$$

In this case,  $\alpha_* = \frac{1}{2}$ .

2)  $\tau_d > 1$ :

$$\rho_{\text{eff}} = \frac{\rho \tau}{\tau_d - 1} \cdot \frac{\alpha(1 - \alpha)}{-\alpha + \vartheta}, \quad \vartheta = \frac{1 + \rho \tau}{\rho \tau (1 - \frac{1}{\tau_d})} > 1.$$

Hence, the optimal  $\alpha_*$  that maximizes  $\rho_{\text{eff}}$  is given by:

$$\begin{aligned}\alpha_* &= \arg \max_{\substack{0 < \alpha < 1 \\ \vartheta > 1}} \frac{\alpha(1 - \alpha)}{-\alpha + \vartheta} \\ &= \vartheta - \sqrt{\vartheta(\vartheta - 1)}.\end{aligned}$$

3)  $\tau_d < 1$ :

$$\rho_{\text{eff}} = \frac{\rho \tau}{1 - \tau_d} \cdot \frac{\alpha(1 - \alpha)}{\alpha - \vartheta}, \quad \vartheta = \frac{1 + \rho \tau}{\rho \tau (1 - \frac{1}{\tau_d})} < 0.$$

In this case, we have

$$\begin{aligned}\alpha_* &= \arg \max_{\substack{0 < \alpha < 1 \\ \vartheta < 0}} \frac{\alpha(1 - \alpha)}{\alpha - \vartheta} \\ &= \vartheta + \sqrt{\vartheta(\vartheta - 1)}.\end{aligned}$$

## Appendix G

### Optimal Power and Training Time Allocation Derivation based on Goodput

In this section, we determine the optimal power and training time allocation that optimizes the asymptotic value of the goodput for LS decoder which we denote here by  $\tilde{G}_{LS}$ .

To this end, we proceed with the change of variable  $\gamma = \frac{T_p}{T}$ . Then,  $\tau_p = \frac{T_p}{K} = \frac{T_p}{T} \cdot \frac{T}{K} = \gamma \tau$ ,  $\delta' = \frac{\delta - 1}{\varepsilon}$ ,  $\rho_p = \frac{(1 - \alpha) \rho}{\gamma}$ ,

and  $\rho_d = \frac{\alpha \rho}{1 - \gamma}$ . The SEP for LS case is

$$\begin{aligned}\widetilde{\text{SEP}}_{LS} &= 2(1 - \frac{1}{M}) Q \left( \sqrt{\frac{\delta - 1}{\varepsilon} \rho_{\text{eff}}} \right) \\ &= 2(1 - \frac{1}{M}) Q \left( \sqrt{\frac{\delta - 1}{\varepsilon} \cdot \frac{\rho_d \rho_p \tau_p}{1 + \rho_d + \rho_p \tau_p}} \right) \\ &= 2(1 - \frac{1}{M}) Q \left( \sqrt{\frac{\delta - 1}{\varepsilon} \cdot \frac{\frac{\alpha \rho}{1 - \gamma} \cdot \frac{(1 - \alpha) \rho}{\gamma} \gamma \tau}{1 + \frac{\alpha \rho}{1 - \gamma} + \frac{(1 - \alpha) \rho}{\gamma} \gamma \tau}} \right) \\ &= 2(1 - \frac{1}{M}) Q \left( \sqrt{\frac{\delta' \alpha (1 - \alpha) \rho^2 \tau}{1 + \alpha \rho + (1 - \alpha) \rho \tau - \gamma (1 + (1 - \alpha) \rho \tau)}} \right) \\ &= 2(1 - \frac{1}{M}) Q \left( \sqrt{\frac{\frac{\delta' \alpha (1 - \alpha) \rho^2 \tau}{1 + (1 - \alpha) \rho \tau}}{1 + \alpha \rho + (1 - \alpha) \rho \tau - \gamma}} \right) \\ &= \tilde{M} Q \left( \sqrt{\frac{b(\alpha)}{a(\alpha) - \gamma}} \right),\end{aligned}$$

where  $\tilde{M} = 2(1 - \frac{1}{M})$ ,  $a(\alpha) = \frac{1 + \alpha \rho + (1 - \alpha) \rho \tau}{1 + (1 - \alpha) \rho \tau} > 1$ , and  $b(\alpha) = \frac{\delta' \alpha (1 - \alpha) \rho^2 \tau}{1 + (1 - \alpha) \rho \tau} > 0$ . The asymptotic goodput becomes

$$\begin{aligned}\tilde{G}_{LS}(\alpha, \gamma) &= (1 - \gamma)(1 - \widetilde{\text{SEP}}_{LS}) \\ &= (1 - \gamma) \left( 1 - \tilde{M} Q \left( \sqrt{\frac{b(\alpha)}{a(\alpha) - \gamma}} \right) \right).\end{aligned}$$

The power allocation problem amounts thus to solving:

$$(\alpha_*, \gamma_*) = \arg \max_{\substack{0 < \alpha < 1 \\ \frac{1}{\tau} \leq \gamma < 1}} \tilde{G}_{LS}(\alpha, \gamma), \quad (135)$$

Recall that we need  $T_p \geq K$  or  $\tau_p \geq 1$ , but  $\tau_p = \gamma \tau$ , hence  $\gamma \geq \frac{1}{\tau}$ . We also require  $T_p < T$ , hence  $\gamma < 1$ . This justifies the constraint imposed on  $\gamma$  above.

To begin with, it is easy to see that the optimal  $\alpha$  is the one that maximizes  $\rho_{\text{eff}}$  and this is for any  $\gamma$ . It remains thus to optimize the goodput in terms of  $\gamma$ . We will solve this by proving that the goodput is a decreasing function with respect to  $\gamma$ . To proceed, let us make the change of variables:  $x = \sqrt{\frac{b}{a - \gamma}}$ , then  $\gamma = a - \frac{b}{x^2}$ . Hence, for any  $0 < \alpha < 1$ , the goodput is

$$\tilde{G}_{LS}(\alpha, \gamma) = \bar{G}_{LS}(x) := \left( 1 - a + \frac{b}{x^2} \right) \left( 1 - \tilde{M} Q(x) \right).$$

Taking the derivative of  $\bar{G}_{LS}(x)$  with respect to  $x$  yields

$$\begin{aligned}\bar{G}'_{LS}(x) &= \frac{-2b}{x^3} \left( 1 - \tilde{M} Q(x) \right) + \left( 1 - a + \frac{b}{x^2} \right) \left( \frac{\tilde{M}}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right) \\ &= \tilde{M} e^{-\frac{x^2}{2}} \left[ \frac{-2b}{x^3} e^{\frac{x^2}{2}} \left( \frac{1}{M} - Q(x) \right) + \frac{1}{\sqrt{2\pi}} \left( 1 - a + \frac{b}{x^2} \right) \right].\end{aligned}$$

We need to study the sign of  $\bar{G}'_{LS}(x)$ . To do so, first write the Taylor series expansion of  $e^{\frac{x^2}{2}} \left( \frac{1}{M} - Q(x) \right)$  as:

$$\begin{aligned}e^{\frac{x^2}{2}} \left( \frac{1}{M} - Q(x) \right) &= \frac{1}{2(M-1)} + \frac{x}{\sqrt{2\pi}} + \frac{x^2}{4(M-1)} + \dots, \text{ then} \\ \bar{G}'_{LS}(x) &= \tilde{M} e^{-\frac{x^2}{2}} \left[ \frac{-b}{(M-1)x^3} - \frac{b}{\sqrt{2\pi}x^2} \right. \\ &\quad \left. - \frac{b}{4(M-1)x} + \dots + \frac{1}{\sqrt{2\pi}}(1 - a) \right].\end{aligned}$$

Recall that  $a > 1$  and note that all the terms in  $\bar{G}'_{LS}(x)$  expression above are negative. Hence,  $\bar{G}'_{LS}(x) < 0$  for all  $\alpha$ .

Now, by the chain rule:  $\frac{d\bar{G}_{LS}}{dx} = \frac{d\bar{G}_{LS}}{d\gamma} \cdot \frac{d\gamma}{dx}$ , but  $\frac{d\gamma}{dx} = \frac{2b}{x^3} > 0$ , then  $\frac{d\bar{G}_{LS}}{d\gamma} < 0$ .

Hence,  $\frac{\partial \bar{G}_{LS}(\alpha, \gamma)}{\partial \alpha} < 0$  for any  $\alpha$ . This suggests that we choose  $\gamma$  as small as possible to maximize  $\bar{G}_{LS}(\alpha, \gamma)$ .

So,  $\gamma_* = \frac{1}{\tau}$ . Or,  $\tau_p^* = \gamma_* \tau$ , then  $\tau_p^* = 1$ .

The proof above is done for LS but the same conclusions hold for both RLS and also for the Box-RLS under the conjecture that for Box-RLS, the optimal  $\alpha$  is the one that maximizes  $\rho_{\text{eff}}$ . We omit details for brevity.

## Appendix H Comparison with the LMMSE Decoder

In this appendix we will show that the LMMSE estimator of  $\mathbf{x}_0$  is equivalent to an RLS estimator with the optimal regularizer  $\lambda_* = \frac{1}{\rho_d} + \sigma_\Delta^2$ . The LMMSE estimate of  $\mathbf{x}_0$  is given by [56]

$$\hat{\mathbf{x}}_{\text{LMMSE}} = \mathbf{C}_{xy} \mathbf{C}_{yy}^{-1} \mathbf{y}, \quad (136)$$

where  $\mathbf{C}_{xy} = \mathbb{E}[(\mathbf{x}_0 - \mathbb{E}[\mathbf{x}_0])(\mathbf{y} - \mathbb{E}[\mathbf{y}])^T]$ , and  $\mathbf{C}_{yy} = \mathbb{E}[(\mathbf{y} - \mathbb{E}[\mathbf{y}])(\mathbf{y} - \mathbb{E}[\mathbf{y}])^T]$ . It can be shown that  $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{x}_0] = \mathbf{0}$ , and

$$\mathbf{C}_{xy} = \sqrt{\frac{\rho_d}{K}} \hat{\mathbf{H}}^T. \quad (137)$$

To find  $\mathbf{C}_{yy}$ , let us write  $\mathbf{y}$  as

$$\begin{aligned} \mathbf{y} &= \sqrt{\frac{\rho_d}{K}} (\hat{\mathbf{H}} + \Delta) \mathbf{x}_0 + \mathbf{z}, \\ &= \sqrt{\frac{\rho_d}{K}} \hat{\mathbf{H}} \mathbf{x}_0 + \tilde{\mathbf{z}}, \end{aligned} \quad (138)$$

where  $\tilde{\mathbf{z}} \triangleq \sqrt{\frac{\rho_d}{K}} \Delta \mathbf{x}_0 + \mathbf{z}$  which is a zero-mean vector, with  $\mathbf{C}_{\tilde{z}\tilde{z}} = \frac{\rho_d}{K} \mathbb{E}_{\mathbf{x}_0, \Delta} [\Delta \mathbf{x}_0 \mathbf{x}_0^T \Delta^T] + \mathbf{C}_{zz} = \frac{\rho_d}{K} \mathbb{E}_\Delta [\Delta \mathbb{E}_{\mathbf{x}_0} [\mathbf{x}_0 \mathbf{x}_0^T] \Delta^T] + \mathbf{C}_{zz} = \frac{\rho_d}{K} \mathbb{E}_\Delta [\Delta \mathbf{C}_{xx} \Delta^T] + \mathbf{C}_{zz} = \frac{\rho_d}{K} \mathbb{E}_\Delta [\Delta \Delta^T] + \mathbf{I}_N = (\rho_d \sigma_\Delta^2 + 1) \mathbf{I}_N$ . Then,

$$\mathbf{C}_{yy} = \frac{\rho_d}{K} \hat{\mathbf{H}} \hat{\mathbf{H}}^T + (\rho_d \sigma_\Delta^2 + 1) \mathbf{I}_N. \quad (139)$$

Note that we used  $\mathbf{C}_{xx} = \mathbf{I}_K$ ,  $\mathbf{C}_{zz} = \mathbf{I}_N$ , and  $\mathbb{E}[\Delta \Delta^T] = K \sigma_\Delta^2 \mathbf{I}_N$ . Hence,

$$\begin{aligned} \hat{\mathbf{x}}_{\text{LMMSE}} &= \sqrt{\frac{\rho_d}{K}} \hat{\mathbf{H}}^T \left( \frac{\rho_d}{K} \hat{\mathbf{H}} \hat{\mathbf{H}}^T + (\rho_d \sigma_\Delta^2 + 1) \mathbf{I}_N \right)^{-1} \mathbf{y}, \\ &= \sqrt{\frac{\rho_d}{K}} \left( \frac{\rho_d}{K} \hat{\mathbf{H}}^T \hat{\mathbf{H}} + (\rho_d \sigma_\Delta^2 + 1) \mathbf{I}_K \right)^{-1} \hat{\mathbf{H}}^T \mathbf{y}, \\ &= (\mathbf{A}^T \mathbf{A} + (1 + \rho_d \sigma_\Delta^2) \mathbf{I}_K)^{-1} \mathbf{A}^T \mathbf{y}, \end{aligned} \quad (140)$$

where  $\mathbf{A} = \sqrt{\frac{\rho_d}{K}} \hat{\mathbf{H}}$ , and the second equality follows from the *matrix inversion Lemma*. The LMMSE estimate in (140) is equivalent to the RLS solution as given in (11b) with  $\lambda \rho_d = \lambda_* \rho_d = 1 + \rho_d \sigma_\Delta^2$ . This shows that the RLS with optimal regularizer is nothing but the popular LMMSE decoder. Finally, the LMMSE estimate can be written in terms of  $\hat{\mathbf{H}}$  as

$$\begin{aligned} \hat{\mathbf{x}}_{\text{LMMSE}} &= \sqrt{\frac{K}{\rho_d}} \left( \hat{\mathbf{H}}^T \hat{\mathbf{H}} + \frac{\sigma_\Delta^2}{\rho_{\text{eff}}} K \mathbf{I}_K \right)^{-1} \hat{\mathbf{H}}^T \mathbf{y} \\ &= \hat{\mathbf{x}}_{\text{RLS}}(\lambda_*). \end{aligned} \quad (141)$$

## REFERENCES

- [1] Ayed M Alrashdi, Ismail Ben Atitallah, Tarig Ballal, Christos Thramoulidis, Anas Chaaban, and Tareq Y Al-Naffouri, "Optimum Training for MIMO BPSK Transmission," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2018, pp. 1–5.
- [2] I. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *Euro-pean Trans. Telecommun.*, vol. 10, pp. 585 – 595, 1999.
- [3] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wirel. Pers. Commun.*, vol. 6, no. 3, pp. 311–335, 1998.
- [4] Babak Hassibi and Bertrand M Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?," *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.
- [5] Ribhu Chopra, Chandra R Murthy, Himal A Suraweera, and Erik G Larsson, "Blind channel estimation for downlink massive MIMO systems with imperfect channel reciprocity," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3132–3145, 2020.
- [6] Arun P Kannu and Philip Schniter, "Capacity Analysis of MMSE Pilot-Aided Transmission for Doubly Selective Channels," in *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications, 2005*. IEEE, 2005, pp. 801–805.
- [7] VK Varma Gottumukkala and Hlaing Minn, "Optimal Pilot Power Allocation for OFDM Systems With Transmitter and Receiver IQ Imbalances," in *GLOBECOM 2009-2009 IEEE Global Telecommunications Conference*. IEEE, 2009, pp. 1–5.
- [8] Michal Simko, Stefan Pendl, Stefan Schwarz, Qi Wang, Josep Colom Ikuno, and Markus Rupp, "Optimal Pilot Symbol Power Allocation in LTE," in *2011 IEEE Vehicular Technology Conference (VTC Fall)*. IEEE, 2011, pp. 1–5.
- [9] Michal Šimko, Qi Wang, and Markus Rupp, "Optimal Pilot Symbol Power Allocation under Time-Variant Channels," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, pp. 225, 2012.
- [10] Jun Wang, Oliver Yu Wen, Hongyang Chen, and Shaoqian Li, "Power allocation between pilot and data symbols for MIMO systems with MMSE detection under MMSE channel estimation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, pp. 785437, 2011.
- [11] Ye Zhang and Wei-Ping Zhu, "Energy-efficient pilot and data power allocation in massive MIMO communication systems based on MMSE channel estimation," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 3571–3575.
- [12] Kezhi Wang, Yunfei Chen, Mohamed-Slim Alouini, and Feng Xu, "BER and Optimal Power Allocation for amplify-and-forward relaying using pilot-aided maximum likelihood estimation," *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3462–3475, 2014.
- [13] Gongpu Wang and Chintia Tellambura, "Super-imposed pilot-aided channel estimation and power allocation for relay systems," in *2009 IEEE Wireless Communications and Networking Conference*. IEEE, 2009, pp. 1–6.
- [14] Xiaodong Cai and Georgios B Giannakis, "Error probability minimizing pilots for OFDM with M-PSK modulation over Rayleigh-fading channels," *IEEE transactions on vehicular technology*, vol. 53, no. 1, pp. 146–155, 2004.
- [15] Pei Liu, Shi Jin, Tao Jiang, Qi Zhang, and Michail Matthaiou, "Pilot power allocation through user grouping in multi-cell massive MIMO systems," *Ieee transactions on communications*, vol. 65, no. 4, pp. 1561–1574, 2016.
- [16] Peiyue Zhao, Gábor Fodor, György Dán, and Miklós Telek, "A game theoretic approach to setting the pilot power ratio in multi-user MIMO systems," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 999–1012, 2017.
- [17] Trinh Van Chien, Emil Björnson, and Erik G Larsson, "Joint pilot design and uplink power allocation in multi-cell massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 2000–2015, 2018.
- [18] Hongyuan Gao, Yumeng Su, Shibo Zhang, and Ming Diao, "Antenna selection and power allocation design for 5G massive MIMO uplink networks," *China Communications*, vol. 16, no. 4, pp. 1–15, 2019.
- [19] Trinh Van Chien, Emil Björnson, and Erik G Larsson, "Joint power allocation and load balancing optimization for energy-efficient cell-



- free massive MIMO networks,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6798–6812, 2020.
- [20] Xiangyun Zhou, Tharaka A Lamahewa, Parastoo Sadeghi, and Salman Durrani, “Two-way Training: Optimal Power Allocation for Pilot and Data Transmission,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 564–569, 2010.
- [21] Shanshan Yu, Wali Ullah Khan, Xiaoqing Zhang, and Ju Liu, “Optimal power allocation for NOMA-enabled D2D communication with imperfect SIC decoding,” *Physical Communication*, vol. 46, pp. 101296, 2021.
- [22] Omar Elkhartouty, Engy Maher, Ahmed El-Mahdy, and Falko Dressler, “Optimal power allocation in cooperative MIMO-NOMA with FD/HD relaying in THz Communications,” in *2020 9th IFIP International Conference on Performance Evaluation and Modeling in Wireless Networks (PEMWN)*. IEEE, 2020, pp. 1–6.
- [23] Shizhao Yang, Yuan Ren, and Guangyue Lu, “Optimal power allocation for wireless-powered full-duplex cooperative NOMA system with partial CSI,” in *2020 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2020, pp. 1–5.
- [24] Yasar Sinan Nasir and Dongning Guo, “Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2239–2250, 2019.
- [25] Khai Nguyen Doan, Mojtaba Vaezi, Wonjae Shin, H Vincent Poor, Hyundong Shin, and Tony QS Quek, “Power allocation in cache-aided NOMA systems: Optimization and deep reinforcement learning approaches,” *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 630–644, 2019.
- [26] Fan Meng, Peng Chen, and Lenan Wu, “Power allocation in multi-user cellular networks with deep Q learning approach,” in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.
- [27] Fan Meng, Peng Chen, Lenan Wu, and Julian Cheng, “Power allocation in multi-user cellular networks: Deep reinforcement learning approaches,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6255–6267, 2020.
- [28] Mark Eisen and Alejandro Ribeiro, “Large scale wireless power allocation with graph neural networks,” in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2019, pp. 1–5.
- [29] Chaofan He, Yang Hu, Yan Chen, and Bing Zeng, “Joint power allocation and channel assignment for NOMA with deep reinforcement learning,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2200–2210, 2019.
- [30] Fei Liang, Cong Shen, Wei Yu, and Feng Wu, “Towards optimal power control via ensembling deep neural networks,” *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1760–1776, 2019.
- [31] Ranning Wang, Animesh Yadav, Esraa A Makled, Octavia A Dobre, Ruiqin Zhao, and Pramod K Varshney, “Optimal power allocation for full-duplex underwater relay networks with energy harvesting: A reinforcement learning approach,” *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 223–227, 2019.
- [32] Nguyen Quang Hieu, Dinh Thai Hoang, Dusit Niyato, and Dong In Kim, “Optimal power allocation for rate splitting communications with deep reinforcement learning,” *IEEE Wireless Communications Letters*, vol. 10, no. 12, pp. 2820–2823, 2021.
- [33] Yunyu Sun, Ye Wang, Jian Jiao, Shaohua Wu, and Qinyu Zhang, “Deep learning-based long-term power allocation scheme for NOMA downlink system in S-LoT,” *IEEE Access*, vol. 7, pp. 86288–86296, 2019.
- [34] François Rottenberg, François Horlin, Eleftherios Kofidis, and Jérôme Louveaux, “Generalized optimal pilot allocation for channel estimation in multicarrier systems,” in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2016, pp. 1–5.
- [35] Zhikun Xu, Geoffrey Ye Li, Chenyang Yang, Shunqing Zhang, Yan Chen, and Shugong Xu, “Energy-efficient power allocation for pilots in training-based downlink OFDMA systems,” *IEEE Transactions on Communications*, vol. 60, no. 10, pp. 3047–3058, 2012.
- [36] Xiaoli Ma, Georgios B Giannakis, and Shuichi Ohno, “Optimal training for block transmissions over doubly selective wireless fading channels,” *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1351–1366, 2003.
- [37] Mahdi Khosravi and Saeed Mashhadi, “Joint pilot power & pattern design for compressive OFDM channel estimation,” *IEEE Communications Letters*, vol. 19, no. 1, pp. 50–53, 2014.
- [38] Jiming Chen, Youxi Tang, and Shaoqian Li, “Pilot power allocation for OFDM systems,” in *The 57th IEEE Semiannual Vehicular Technology Conference, 2003. VTC 2003-Spring*. IEEE, 2003, vol. 2, pp. 1283–1287.
- [39] Rafael Montalban, José A López-Salcedo, Gonzalo Seco-Granados, and A Lee Swindlehurst, “Power allocation method based on the channel statistics for combined positioning and communications OFDM systems,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 4384–4388.
- [40] Heejung Yu, “Optimal primary pilot power allocation and secondary channel sensing in cognitive radios,” *IET Communications*, vol. 10, no. 5, pp. 487–494, 2016.
- [41] Hei Victor Cheng, Emil Björnson, and Erik G Larsson, “Optimal Pilot and Payload Power Control in Single-cell Massive MIMO Systems,” *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2363–2378, 2016.
- [42] Kusi Ankras Bonsu, Weiwei Zhou, Su Pan, and Yan Yan, “Optimal power allocation with limited feedback of channel state information in multi-user MIMO systems,” *China Communications*, vol. 17, no. 2, pp. 163–175, 2020.
- [43] Hieu Trong Dao and Sunghwan Kim, “Pilot power allocation for maximising the sum rate in massive MIMO systems,” *IET Communications*, vol. 12, no. 11, pp. 1367–1372, 2018.
- [44] Hien Quoc Ngo, Michail Matthaiou, and Erik G Larsson, “Massive MIMO with optimal power and training duration allocation,” *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 605–608, 2014.
- [45] Kaifeng Guo, Yan Guo, and Gerd Ascheid, “Energy-efficient uplink power allocation in multi-cell MU-massive-MIMO systems,” in *Proceedings of European Wireless 2015; 21th European Wireless Conference*. VDE, 2015, pp. 1–5.
- [46] Ismail Ben Atitallah, Christos Thrampoulidis, Abba Kammoun, Tareq Y Al-Naffouri, Babak Hassibi, and Mohamed-Slim Alouini, “BER Analysis of Regularized Least Squares for BPSK Recovery,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 4262–4266.
- [47] Christos Thrampoulidis, Ehsan Abbasi, Weiyu Xu, and Babak Hassibi, “BER Analysis of the Box Relaxation for BPSK Signal Recovery,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 3776–3780.
- [48] Peng Hui Tan, Lars K Rasmussen, and Teng J Lim, “Constrained maximum-likelihood detection in CDMA,” *IEEE Transactions on Communications*, vol. 49, no. 1, pp. 142–153, 2001.
- [49] AN TIKHONOV, “Solution of incorrectly formulated problems and the regularization method,” in *Dokl. Akad. Nauk.*, 1963, vol. 151, pp. 1035–1038.
- [50] Christos Thrampoulidis, Weiyu Xu, and Babak Hassibi, “Symbol Error Rate Performance of Box-relaxation Decoders in Massive MIMO,” *IEEE Transactions on Signal Processing*, vol. 66, no. 13, pp. 3377–3392, 2018.
- [51] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi, “Precise Error Analysis of Regularized M-Estimators in High-Dimensions,” *IEEE Transactions on Information Theory*, 2018.
- [52] Ayed M Alrashdi, Ismail Ben Atitallah, Tareq Y Al-Naffouri, and Mohamed-Slim Alouini, “Precise performance analysis of the LASSO under matrix uncertainties,” in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2017, pp. 1290–1294.
- [53] Ayed M Alrashdi, Ismail Ben Atitallah, and Tareq Y Al-Naffouri, “Precise performance analysis of the box-elastic net under matrix uncertainties,” *IEEE Signal Processing Letters*, vol. 26, no. 5, pp. 655–659, 2019.
- [54] Fredrik Rusek, Daniel Persson, Buon Kiong Lau, Erik G Larsson, Thomas L Marzetta, Ove Edfors, and Fredrik Tufvesson, “Scaling up MIMO: Opportunities and Challenges with Very Large Arrays,” *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2012.
- [55] Thomas L Marzetta, *Fundamentals of Massive MIMO*, Cambridge University Press, 2016.
- [56] Steven M Kay, *Fundamentals of Statistical Signal Processing*, Prentice Hall PTR, 1993.



[57] T. Brown, Elisabeth D. C, and P. Kyritsi, *Practical Guide to the MIMO Radio Channel with MATLAB Examples: with MATLAB Examples*, Wiley, 1 edition, 2012.

[58] Aylin Yener, Roy D Yates, and Sennur Ulukus, "CDMA multiuser detection: A nonlinear programming approach," *IEEE Transactions on Communications*, vol. 50, no. 6, pp. 1016–1024, 2002.

[59] Wing-Kin Ma, Timothy N Davidson, Kon Max Wong, Zhi-Quan Luo, and Pak-Chung Ching, "Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 912–922, 2002.

[60] Alex Grote, Walter Grote, and Rodolfo Feick, "IEEE 802.11 Goodput Analysis for Mixed Real Time and Data Traffic," in *Home Networking*, pp. 15–28. Springer, 2008.

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

[62] Yihong Wu and Sergio Verdú, "Optimal phase transitions in compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6241–6263, 2012.

[63] Zeyu Deng, Abla Kammoun, and Christos Thrampoulidis, "A model of double descent for high-dimensional binary linear classification," *Information and Inference: A Journal of the IMA*, vol. 11, no. 2, pp. 435–495, 2022.

[64] Ralph Tyrell Rockafellar, *Convex Analysis*, Princeton University Press, 2015.

[65] Abla Kammoun and Mohamed-Slim Alouini, "On the precise error analysis of support vector machines," *arXiv preprint arXiv:2003.12972*, 2020.

[66] R. D. Gill and P. K. Andersen, "Cox's regression model for counting processes : a large sample study," *Ann. Statist.*, vol. 10, pp. 1100–1120, 1982.



**AYED M. ALRASHDI** (Member, IEEE) received the B.S. degree in Electrical Engineering (with first honors) from the University of Ha'il, Ha'il, Saudi Arabia, in 2014, and the M.S. degree in Electrical Engineering from King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, in 2016. He received the Ph.D. degree in Electrical and Computer Engineering from KAUST in 2021.

He joined the University of Ha'il in 2014, where he is currently an Assistant Professor in the Electrical Engineering Department. From 2017 to 2021, he was a Research Assistant with the Information System Lab (ISL) at KAUST. His research interests are in the areas of statistical signal processing, high-dimensional statistics, compressed sensing, statistical learning, and wireless communications.



**ABLA KAMMOUN** (Member, IEEE) was born in Sfax, Tunisia. She received the Engineering degree in signal and systems from Tunisia Polytechnic School, La Marsa, Tunisia, and the masters and Ph.D. degrees in digital communications from Telecom Paris Tech, Paris, France (formerly, Ecole

Nationale Supereure des T). From 2010 to 2012, she was a Postdoctoral Researcher with the TSI Department, Telecom ParisTech. She was with Supelec, Alcatel-Lucent Chair on Flexible Radio until 2013. She is currently a Research Scientist with the King Abdullah University of Science and Technology, Saudi Arabia. Her research interests include performance analysis of wireless communication systems, random matrix theory, and statistical signal processing.



**ALI H. MUQAIBEL** (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 1996 and 1999, respectively, and the Ph.D. degree from the Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, in 2003. During his study at Virginia Tech, he was with the Time Domain and RF Measurements Laboratory and the Mobile and Portable Radio Research Group. He was a Visiting Associate Professor with the Center of Advanced Communications, Villanova University, Villanova, PA, USA, in 2013, a Visiting Professor with the Georgia Institute of Technology, in 2015, and a Visiting Scholar with the King Abdullah University for Science and Technology (KAUST), Thuwal, Saudi Arabia, from 2018 to 2019. He is currently a Professor with the Electrical Engineering Department, KFUPM. He is also the Director of the Center for Communication Systems and Sensing. He has authored two book chapters and over 130 articles. His research interests include direction of arrival estimation, through-wall-imaging, localization, channel characterization, and ultra-wideband signal processing. He was a recipient of many awards in the excellence in teaching, advising, and instructional technology.



**TAREQ Y. AL-NAFFOURI** (Senior Member, IEEE) received the B.S. degrees in mathematics and electrical engineering (with first Hons.) from the King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, the M.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, Georgia, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2004. He was a Visiting Scholar with the California Institute of Technology, Pasadena, CA, in 2005 and summer 2006. He was a Fulbright Scholar with the University of Southern California, Los Angeles, CA, in 2008. He is currently a Professor with Electrical Engineering Department, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. His research interests include sparse, adaptive, and statistical signal processing and their applications to wireless communications and localization, machine learning, and network information theory. He has more than 350 publications in journal and conference proceedings and 24 issued/pending patents. Dr. Al-Naffouri was the recipient of the IEEE Education Society Chapter Achievement Award in 2008 and Al-Marai Award for innovative research in communication in 2009. Dr. Al-Naffouri was an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING during 2013-2018.