BER ANALYSIS OF REGULARIZED LEAST SQUARES FOR BPSK RECOVERY

Ismail Ben Atitallah[1], *Christos Thrampoulidis*[2], *Abla Kammoun*[1],
Tareq Y. Al-Naffouri[1], *Babak Hassibi*[2] *and Mohamed-Slim Alouini*[1]

[1] CEMSE division, KAUST, Thuwal, KSA

[2] Department of Electrical Engeeniring, Caltech, Pasadena, USA

ABSTRACT

This paper investigates the problem of recovering an n-dimensional BPSK signal $\mathbf{x}_0 \in \{-1,1\}^n$ from m-dimensional measurement vector $\mathbf{y} = \mathbf{A}\mathbf{x}+\mathbf{z}$, where $\mathbf{A}$ and $\mathbf{z}$ are assumed to be Gaussian with *iid* entries. We consider two variants of decoders based on the regularized least squares followed by hard-thresholding: the case where the convex relaxation is from $\{-1,1\}^n$ to $\mathbb{R}^n$ and the box constrained case where the relaxation is to $[-1,1]^n$. For both cases, we derive an exact expression of the bit error probability when n and m grow simultaneously large at a fixed ratio. For the box constrained case, we show that there exists a critical value of the SNR, above which the optimal regularizer is zero. On the other side, the regularization can further improve the performance of the box relaxation at low to moderate SNR regimes. We also prove that the optimal regularizer in the bit error rate sense for the unboxed case is nothing but the MMSE detector.

Index Terms— BER analysis, box relaxation, regularized least squares, MMSE, high dimensions

1. INTRODUCTION

Estimating the linear least squares fit to data is a well-known problem in many applications throughout science and engineering. Of great interest is the setting where the variables are constrained to be in a discrete set. This integer least squares problem has many diverse applications such as decoding in multi-input-multi-output (MIMO) systems, high precision GNSS positioning [1] and many lattice problems in computer science [2]. In contrast to the continuous linear least squares problem where a closed form solution can be found, this problem is known to be NP-hard. In MIMO detection, Maximum of Likelihood boils down to an integer least squares problem. Many low-complexity detection algorithms are thus proposed such as zero forcing, MMSE, and decision feedback.

Of interest in this paper is BPSK recovery, i.e. recovering a sign vector in the set $\{-1,1\}^n$ corrupted by noise. The detection procedure is based on the regularized least squares (RLS) estimation followed by hard-thresholding. The RLS is a LASSO-type algorithm, where an ℓ_2 norm regularization is used to control the variance of the estimates and to avoid ill-conditioning issues. This regularization is known as ridge regression in statistics and machine learning contexts. The MMSE decoder is a special case of the RLS decoder when the regularizer λ is set to $\frac{1}{\mathrm{SNR}}$.

Another popular heuristic that enhances the performance of the detector is the box relaxation [3–5], *i.e.* solving the RLS problem when constraining the variables to be in $[-1,1]^n$ instead of $\{-1,1\}^n$.

The error probability of RLS, particularly the MMSE, can be found using standard results from random matrix theory and the asymptotic normality assumption of the SINR [6]. However, very little is known about the performance of the RLS with box relaxation. We use the Convex Gaussian Min-max Theorem (CGMT) to derive exact expressions of the bit-wise error probability for both decoders and answer questions about the optimal regularizer.

Setup. We consider the problem of recovering an n-dimensional BPSK signal $\mathbf{x}_0 \in \{-1,1\}^n$ from a noisy received measurement vector $\mathbf{y} = \mathbf{A}\mathbf{x}_0+\mathbf{z}$, where $\mathbf{A}$ is the MIMO channel matrix assumed to be known with iid entries $\mathcal{N}(0,\frac{1}{n})$, and $\mathbf{z}$ is the thermal noise vector $\mathcal{N}(0,\sigma^2)$. The normalization in the entries of $\mathbf{A}$ is made such that the SNR does not scale with n, i.e. $\mathrm{SNR} = \frac{1}{\sigma^2}$.

Regularized Least Squares with Box Relaxation Optimization (RLS-BRO). This heuristic detection scheme involves two steps. The first one consists in solving the relaxed optimization problem, by assuming that $\mathbf{x}$ is in $[-1,1]^n$ instead of $\{-1,1\}^n$, and thus yielding a convex problem. Then, the output $\widehat{\mathbf{x}}$ of the first step is hard-thresholded by a sign function to produce a sign vector $\mathbf{x}^*$ that estimates $\mathbf{x}_0$. Formally, denoting by $\|.\|$ the ℓ_2 norm, the two steps are as follows:

$$\widehat{\mathbf{x}} = \underset{-1\leq \mathbf{x}_i \leq 1}{\operatorname{argmin}} \ \|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2+\lambda\|\mathbf{x}\|^2 \tag{1a}$$

$$\mathbf{x}^* = \operatorname{sign}(\widehat{\mathbf{x}}) \tag{1b}$$

We also consider the ordinary convex relaxation in $\mathbb{R}^n$. We denote this scheme simply by **RLS**. We want to answer herein the following question: when the regularizer is optimally tuned to minimize the error probability, does this coincide with the

This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. URF/1/2221-01.

MMSE decoder?

Bit error probability We evaluate the performance of the detection algorithm using the bit error probability P_e defined as the expectation of the bit error rate (BER).

$$\mathrm{BER} := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{x}*_i \neq \mathbf{x}_{0,i}\}} \tag{2a}$$

$$P_e := \mathbb{E}[\mathrm{BER}] = \frac{1}{n}\sum_{i=1}^{n} \Pr(\mathbf{x}*_i \neq \mathbf{x}_{0,i}) \tag{2b}$$

The rest of this paper is organized as follows. In the second section, we present and discuss the results regarding the probabilities of error and the optimal regularizers. In the third section, we recall the Convex Gaussian Min-max Theorem (CGMT), and we give a proof outline that leads to the results in Theorem 1 and 3.

2. MAIN RESULTS

We analyze the asymptotic error probability for both RLS and RLS-BRO, denoted respectively by P_e and P_e^{BRO}, when the BPSK signal is unknown. The analysis is performed when the system dimensions m and n grow simultaneously large at a fixed ratio $\delta := \lim_{n\to\infty} \frac{m}{n} \in (0,\infty)$. The SNR is assumed to be constant. Let $Q(.)$ denote the Q-function associated with the standard normal density $p(h) = \frac{1}{\sqrt{2\pi}} e^{-h^2/2}$.

Theorem 1. *(P_e of RLS). Let $\lambda \geq 0$ and $\delta > 0$. Then,*

$$\lim_{n\to\infty} \mathrm{P_e} = Q\left(\sqrt{\frac{\delta - \frac{1}{(1+\Upsilon(\lambda,\delta))^2}}{\left(\frac{\Upsilon(\lambda,\delta)}{1+\Upsilon(\lambda,\delta)}\right)^2 + \frac{1}{\mathrm{SNR}}}}\right) \tag{3}$$

$$\text{where} \quad \Upsilon(\lambda,\delta) = \frac{1-\delta+\lambda+\sqrt{(1-\delta+\lambda)^2+4\lambda\delta}}{2\delta}.$$

This theorem provides a closed form expression of the bit-wise error probability of the RLS at high dimensions. It namely can serve to compute the asymptotic BER of the MMSE detector simply by setting λ to $\frac{1}{\mathrm{SNR}}$. Such closed form result is appealing because of the fundamental importance of the MMSE decoder. We focus now on tuning the regularizer λ that minimizes this error probability. This is the objective of the next proposition.

Proposition 2. *(Optimal regularizer of RLS) Let λ^* denote the optimal regularizer that minimizes the limit in* (3)*. Then, $\lambda^* = \frac{1}{\mathrm{SNR}}$.*

This proposition establishes that the RLS with optimal regularizer set to λ^* to minimize the error probability is nothing but the MMSE detector. The latter is well known (by definition) to minimize the MSE, but it turns out according to proposition 2 that it also minimizes the BER among all other choices of λ. In the case of the RLS, $\widehat{\mathbf{x}}$ has a closed form solution, that is $\widehat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A}+\lambda\mathbf{I})^{-1}\mathbf{A}^T\mathbf{y}$. Potentially, one could then show Theorem 1 relying on results in asymptotic random matrix theory. Instead, we prove this using the CGMT framework, which yields the result in a more natural way. Importantly, the CGMT approach can be used to analyze more complicated detectors where no closed form solution is available. One such example is the RLS-BRO (1) in next theorem.

Theorem 3. *(P_e^{BRO} of RLS-BRO). Let $\lambda, \delta \in (0,\infty)$. Then,*

$$\lim_{n\to\infty} \mathrm{P_e^{BRO}} = Q(\frac{1}{\tau_*}),$$

where τ_* is the unique solution to the following

$$\min_{\tau>0}\max_{\beta>0} D(\tau,\beta) := \beta\delta\tau + \frac{\beta}{\tau\mathrm{SNR}} - \frac{\beta^2\lambda}{2} + \frac{4\beta}{\tau}Q\left(\frac{2}{\tau}+\frac{2}{\beta}\right)$$
$$-4\beta p\left(\frac{2}{\tau}+\frac{2}{\beta}\right) - \frac{\beta^2}{\frac{\beta}{\tau}+2}\int_{-\frac{2}{\beta}-\frac{2}{\tau}}^{\frac{2}{\beta}} \left(h-\frac{2}{\beta}\right)^2 p(h)\mathrm{d}h \tag{4}$$

The objective function in (4) is strictly convex-concave. Hence, the unique stationary point (τ_*, β_*) can be computed numerically by writing the first order optimality conditions. Apart from predicting the error probability of the system, one major importance of this result lies in the fact that it allows an optimal setting of the regularizer. It is clear that the asymptotic probability of error is minimized when τ_* is minimized with respect to λ. Although the analytical expression of the optimal regularizer seems to be out of reach, it is possible to show that it is exactly zero at high SNR. This result is stated formally in the following proposition:

Proposition 4. *(Optimal regularizer of RLS-BRO at high SNR) Let $\lambda_*^{\mathrm{BRO}} := \mathrm{argmin}_{\lambda\geq 0}\, \tau_*$. Then, $\exists \overline{\mathrm{SNR}} \in \mathbb{R}_+$, such that, $\lambda_*^{\mathrm{BRO}} = 0$ for all $\mathrm{SNR} \in (\overline{\mathrm{SNR}}, \infty)$.*

This result can be proven by assuming that λ is small enough so that $\tau_* \to 0$ as $\mathrm{SNR} \to \infty$.

Numerical results. Figure 1 illustrates the accuracy of the results in theorems 1 and 3. An analytic comparison of the MMSE performance to that of the RLS-BRO, shows significant improvement on the BER performance by imposing the additional box constraint.

In Figure 2, we plot the optimal regularizer predicted by Theorem 3, for different values of δ using a bisection algorithm. The first deduced ascertainment is that the optimal regularizer stagnates at zero starting from moderate values of the SNR. It is also always below $\frac{1}{SNR}$. We can also notice that the optimal value of the regularizer is a decreasing function of δ

Theorem 1 and 3 predict the bit error probability at high dimensions. The two considered schemes belong to the LASSO-type class. Whilst the asymptotic square-error has been recently studied precisely for a wide range of such algorithms [7–11], the error probability has been limited so far only for the ordinary least squares with BRO [12]. It is worth mentioning that the bit-wise error probability is a more appropriate performance metric for the considered schemes because the square-error assesses the performance of the first step (1a) only, but not the overall scheme performance. The analysis of the MSE builds upon the Convex Gaussian Min-max Theorem (CGMT). The application of this theorem can go beyond the MSE to precisely characterize the BER. Details of the derivations are in the next section.

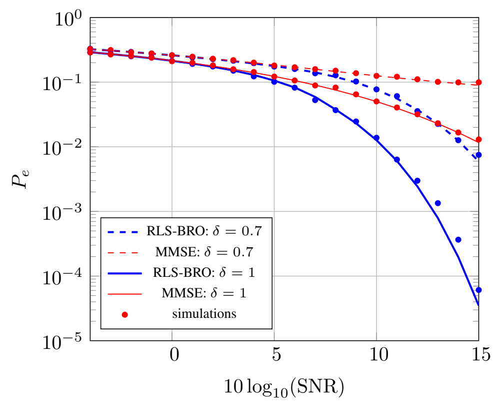

Fig. 1. Error probability of the MMSE and RLS-BRO with optimal regularizer. P_e as a function of SNR in dB for $\delta = 0.7$ and $\delta = 1$. The theoretical predictions follow Theorem 1 and 3. For the Monte-Carlo simulations, we use $n = 512$. for each iteration and each SNR value, we generate 32 independent realizations of the channel matrix and the noise vector.

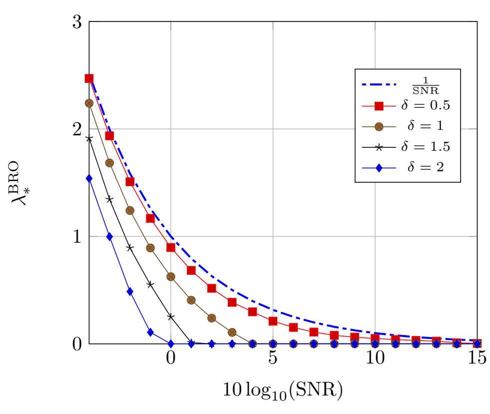

Fig. 2. Optimal regularizer λ_*^{BRO} as a function of the SNR. Different values of δ are considered.

3. PROOF OUTLINE

In this section, we prove Theorem 3. Theorem 1 can be proven in a similar fashion, and the result involves solving a min-max problem as well, but that can be solved in closed form expression. For convenience, we consider the error vector $\mathbf{w} = \mathbf{x} - \mathbf{x}_0$. Define also $\mathcal{B} := [-2, 0]^n$.
Then, the problem in (1a) can be reformulated in terms of $\mathbf{w}$.

$$\widehat{\mathbf{w}} := \underset{\mathbf{w}\in\mathcal{B}}{\operatorname{argmin}} \|\mathbf{z} - \mathbf{A}\mathbf{w}\|^2 + \lambda\|\mathbf{w} + \mathbf{x}_0\|^2 \tag{5}$$

Without loss of generality, we assume for the analysis that $\mathbf{x}_0 = \mathbf{1}_n = (1, 1, \cdots, 1)$. Henceforth, the BER can be expressed as: $\mathrm{BER} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathbb{1}_{\{\widehat{\mathrm{x}}_{\mathrm{i}} < 0\}}$.
CGMT. The key ingredient of the proof is the Convex Gaussian Min-max Theorem [10]. It associates with the Primary Optimization (PO) problem we are interested in an Auxiliary Optimization (AO) problem, from which we can tightly characterize the (PO). The behavior of the (AO) problem is often easier to analyze because it does not involve large random matrices but only random vectors, contrary to the PO that depends on the random measurement matrix $\mathbf{A}$.
In the following, we recall the statement of the CGMT, and we refer the reader to corollary 6.1 in [11], for the complete technical requirements of this theorem. Consider the following two min-max problems:

$$\Phi(\mathbf{G}) := \min_{\mathbf{w}\in\mathcal{S}_w} \max_{\mathbf{u}\in\mathcal{S}_u} \mathbf{u}^T\mathbf{G}\mathbf{w} + \psi(\mathbf{w}, \mathbf{u}), \tag{6a}$$

$$\phi(\mathbf{g}, \mathbf{h}) := \min_{\mathbf{w}\in\mathcal{S}_w} \max_{\mathbf{u}\in\mathcal{S}_u} \|\mathbf{w}\|\mathbf{g}^T\mathbf{u} - \|\mathbf{u}\|\mathbf{h}^T\mathbf{w} + \psi(\mathbf{w}, \mathbf{u}), \tag{6b}$$

where $\mathbf{G} \in \mathbb{R}^{m\times n}$, $\mathbf{g} \in \mathbb{R}^m$, $\mathbf{h} \in \mathbb{R}^n$, $\mathcal{S}_w \subset \mathbb{R}^n$, $\mathcal{S}_u \subset \mathbb{R}^m$ and $\psi : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$. Denote by $\mathbf{w}_\Phi := \mathbf{w}_\phi(\mathbf{G})$ and $\mathbf{w}_\phi := \mathbf{w}_\phi(\mathbf{g}, \mathbf{h})$ any optimal minimizers in (6a) and (6b) respectively.
In (6), let $\mathcal{S}_w$ and $\mathcal{S}_u$ be convex and compact sets, ψ be continuous and convex-concave on $\mathcal{S}_w \times \mathcal{S}_u$, and , $\mathbf{G}$, $\mathbf{g}$ and $\mathbf{h}$ all have *iid* standard normal entries. Let $\mathcal{S}$ be any arbitrary open subset of $\mathcal{S}_w$. Then, if $\lim_{n\to\infty} Pr(\mathbf{w}_\phi \in \mathcal{S}) = 1$, it also holds $\lim_{n\to\infty} Pr(\mathbf{w}_\Phi \in \mathcal{S}) = 1$.

Identifying the (PO) and the (AO) The first step of the analysis is transforming the optimization problem (5) in the form of a (PO). To do, we shall first rewrite (5) by expressing the squared norm loss function in its dual form through the Fenchel conjugate, $\|\mathbf{z} - \mathbf{A}\mathbf{w}\|^2 = \max_{\mathbf{u}} \sqrt{n}\mathbf{u}^T(\mathbf{z} - \mathbf{A}\mathbf{w}) - \frac{n}{4}\|\mathbf{u}\|^2$:

$$\min_{\mathbf{w}\in\mathcal{B}} \max_{\mathbf{u}} \sqrt{n}\mathbf{u}^T\mathbf{A}\mathbf{w} - \sqrt{n}\mathbf{u}^T\mathbf{z} - \frac{n}{4}\|\mathbf{u}\|^2 + \lambda\|\mathbf{x}_0 + \mathbf{w}\|^2. \tag{7}$$

The dual variable $\mathbf{u}$ is scaled by a factor $\sqrt{n}$ for an issue of convergence. Strictly speaking, the terms in the objective function should be all of the same order $\mathcal{O}_p(n)$. The above form in (7) satisfies the (PO) formulation of the CGMT. Hence, we can define the corresponding (AO) as:

$$\min_{\mathbf{w}\in\mathcal{B}} \max_{\mathbf{u}} \|\mathbf{w}\|\mathbf{g}^T\mathbf{u} - \|\mathbf{u}\|\mathbf{h}^T\mathbf{w} - \sqrt{n}\mathbf{u}^T\mathbf{z} - \frac{n}{4}\|\mathbf{u}\|^2 + \lambda\|\mathbf{x}_0 + \mathbf{w}\|^2. \tag{8}$$

Computing the BER **via the (AO).** For any fixed $\epsilon > 0$, define the set $\mathcal{S} = \left\{ \mathbf{v} : \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{v}_i \leq -1\}} - Q(1/\tau_*) \right| < \epsilon \right\}$, where τ_* is defined in theorem 3. Denote by $\widetilde{\mathbf{w}}$ the optimal solution of the (AO) problem. We will prove that $\widetilde{\mathbf{w}} \in \mathcal{S}$ with probability 1. Then, applying the CGMT to the set $\mathcal{S}$ suffices to conclude that $\widehat{\mathbf{w}} \in \mathcal{S}$ with probability 1, which establishes the expression of the asymptotic P_e^{BRO}.

Simplifying the AO Since vectors $\mathbf{g}$ and $\mathbf{z}$ are independent, $\|\mathbf{w}\|\mathbf{g}^T\mathbf{u} - \sqrt{n}\mathbf{u}^T\mathbf{z} \overset{d}{=} \sqrt{\|\mathbf{w}\|^2 + n\sigma^2}\mathbf{g}^T\mathbf{u}$, that is, (8) is equivalent to

$$\min_{\mathbf{w}\in\mathcal{B}} \max_{\mathbf{u}} \sqrt{\|\mathbf{w}\|^2 + n\sigma^2}\mathbf{g}^T\mathbf{u} - \|\mathbf{u}\|\mathbf{h}^T\mathbf{w} - \frac{n}{4}\|\mathbf{u}\|^2 + \lambda\|\mathbf{x}_0 + \mathbf{w}\|^2 \tag{9}$$

Fixing the magnitude of $\mathbf{u}$ to $\beta := \|\mathbf{u}\|$, we can optimize over its direction by aligning it with $\mathbf{g}$. Then, the (AO) simplifies to the following:

$$\max_{\beta\geq 0} \min_{\mathbf{w}\in\mathcal{B}} \sqrt{n}\beta\left(\sqrt{\frac{\|\mathbf{w}\|^2}{n} + \sigma^2}\|\mathbf{g}\| - \frac{\mathbf{h}^T\mathbf{w}}{\sqrt{n}}\right) - \frac{n\beta^2}{4} + \lambda\|\mathbf{x}_0 + \mathbf{w}\|^2 \tag{10}$$

To have a separable optimization problem, we use the identity $\sqrt{\chi} = \min_{\tau>0} \frac{\tau}{2} + \frac{\chi}{2\tau}$, where $\chi = \frac{\|\mathbf{w}\|^2}{n} + \sigma^2$. This yields the following form:

$$\min_{\tau>0} \max_{\beta\geq 0} \frac{\sqrt{n}\beta\tau\|\mathbf{g}\|}{2} + \frac{\sqrt{n}\beta\sigma^2\|\mathbf{g}\|}{2\tau} - \frac{n\beta^2}{4} + n\lambda + \sum_{i=1}^{n} \min_{-2\leq\mathbf{w}_i\leq 0} \left(\frac{\beta\|\mathbf{g}\|}{2\tau\sqrt{n}} + \lambda\right)\mathbf{w}_i^2 - (\beta\mathbf{h}_i - 2\lambda)\mathbf{w}_i. \tag{11}$$

Let τ_n and β_n denote the optimal solutions to the above min-max optimization problem. Then, if $\beta_n > 0$, the optimal $\widetilde{\mathbf{w}}_i$ satisfies

$$\widetilde{\mathbf{w}}_i = \begin{cases} 0 & \text{, if} \quad \mathbf{h}_{\mathrm{i}} \geq \frac{2\lambda}{\beta_{\mathrm{n}}} \\ \frac{\beta_n\mathbf{h}_i - 2\lambda}{\frac{\beta\|\mathbf{g}\|}{\tau_n\sqrt{n}} + 2\lambda} & \text{, if} \quad -2\left(\frac{\|\mathbf{g}\|}{\tau_{\mathrm{n}}\sqrt{\mathrm{n}}} + \frac{\lambda}{\beta_{\mathrm{n}}}\right) < \mathbf{h}_{\mathrm{i}} < \frac{2\lambda}{\beta_{\mathrm{n}}} \\ -2 & \text{, if} \quad \mathbf{h}_{\mathrm{i}} \leq -2\left(\frac{\|\mathbf{g}\|}{\tau_{\mathrm{n}}\sqrt{\mathrm{n}}} + \frac{\lambda}{\beta_{\mathrm{n}}}\right) \end{cases} \tag{12}$$

τ_n and β_n are thus the solution to the following:

$$\min_{\tau>0} \max_{\beta>0} \frac{\sqrt{n}\beta}{2}\left(\tau\|\mathbf{g}\| + \frac{\sigma^2\|\mathbf{g}\|}{\tau}\right) - \frac{n\beta^2}{4} + \sum_{i=1}^{n} v(\tau, \beta) \tag{13}$$

$$v(\tau, \beta) = \begin{cases} 0 & \text{, if} \quad \mathbf{h}_{\mathrm{i}} \geq \frac{2\lambda}{\beta} \\ -\frac{(\beta\mathbf{h}_i - 2\lambda)^2}{\frac{2\beta\|\mathbf{g}\|}{\tau\sqrt{n}} + 4\lambda} & \text{, if} \quad -2\left(\frac{\|\mathbf{g}\|}{\tau\sqrt{\mathrm{n}}} + \frac{\lambda}{\beta}\right) < \mathbf{h}_{\mathrm{i}} < \frac{2\lambda}{\beta} \\ \frac{2\beta\|\mathbf{g}\|}{\tau\sqrt{n}} + 2\beta\mathbf{h}_i & \text{, if} \quad \mathbf{h}_{\mathrm{i}} \leq -2\left(\frac{\|\mathbf{g}\|}{\tau\sqrt{\mathrm{n}}} + \frac{\lambda}{\beta}\right) \end{cases}$$

Convergence of the (AO). After simplifying the (AO) as in (13), we are now in position to analyse its limiting behaviour. First, we need to properly normalize the (AO) by dividing the objective function of the (AO) by n. Also, redefine $\tau := \frac{\tau}{\sqrt{\delta}}$. Using the WLLN, $\frac{\|\mathbf{g}\|}{\sqrt{n}} \xrightarrow{\mathrm{P}} \sqrt{\delta}$ and for all $\tau > 0$ and $\beta > 0$, $\frac{1}{n}v(\tau; \mathbf{h}_i; \|\mathbf{g}\|) \xrightarrow{\mathrm{P}} Y(\tau, \beta)$
$Y(\tau, \beta) := -\frac{1}{\frac{2\beta}{\tau} + 4\lambda} \int_{-\frac{2}{\tau} - \frac{2\lambda}{\beta}}^{\frac{2\lambda}{\beta}} (\beta h - 2\lambda)^2 p(h)\mathrm{d}h + \frac{2\beta}{\tau} Q\left(\frac{2}{\tau} + \frac{2\lambda}{\beta}\right) - 2\beta \int_{2(\frac{1}{\tau} + \frac{\lambda}{\beta})}^{\infty} hp(h)\mathrm{d}h$. Hence, The objective function in (τ, β) in (13) converges to $Y(\tau, \beta) + \frac{\beta\delta\tau}{2} + \frac{\beta}{2\tau\mathrm{SNR}} - \frac{\beta^2}{4}$. Multiplying the latter expression by $\frac{2}{\lambda}$ and redefining $\beta := \frac{\beta}{\lambda}$, we prove the point-wise convergence to $D(\tau, \beta)$ in (4). Furthermore, it is possible to show that for $\lambda \neq 0$, $\beta_* > 0$ with probability one.

The functions $\tau \mapsto \max_{\beta>0} \frac{\sqrt{n}\beta}{2}(\tau\|\mathbf{g}\| + \frac{\sigma^2\|\mathbf{g}\|}{\tau}) - \frac{n\beta^2}{4} + \sum_{i=1}^{n} v(\tau, \beta)$ and $\tau \mapsto \max_{\beta>0} D(\tau, \beta)$ are convex. Hence, we can show using theorem 2.7 in [13] that $\tau_n \xrightarrow{\mathrm{P}} \tau_*$. The latter convergence is crucial for the final step of the proof.

Proving $\widetilde{\mathbf{w}} \in \mathcal{S}$ Using the expression in (12),
$\mathbb{1}_{\{\widetilde{\mathbf{w}}_i \leq -1\}} = \mathbb{1}_{\{\beta\mathbf{h}_i - 2\lambda \leq -(\frac{\beta\|\mathbf{g}\|}{\tau\sqrt{n}} + 2\lambda)\}} = \mathbb{1}_{\{\mathbf{h}_i \leq -\frac{\|\mathbf{g}\|}{\sqrt{n}\sqrt{\delta}\tau}\}}$
recall that $\|\mathbf{g}\|/\sqrt{n} \xrightarrow{\mathrm{P}} \sqrt{\delta}$ and $\tau_n \xrightarrow{\mathrm{P}} \tau_*$. Putting all results together, it can be shown that $\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\widetilde{\mathbf{w}}_i \leq -1\}} \xrightarrow{\mathrm{P}} Q(\frac{1}{\tau_*})$.

Unboxed RLS (Theorem 1) For the the unboxed problem, it can be shown similarly that τ_* is solution to the following:

$$\min_{\tau>0} \max_{\beta>0} \beta\delta\tau + \frac{\beta}{\tau\mathrm{SNR}} - \frac{\beta^2}{2} - \frac{\beta^2 + 4\lambda^2}{\frac{\beta}{\tau} + 2\lambda} \tag{14}$$

The first order optimality conditions are :

$$\delta - \frac{1}{\tau^2\mathrm{SNR}} - \frac{\beta^2 + 4\lambda^2}{(\beta + 2\lambda\tau)^2} = 0 \tag{15a}$$

$$\delta\tau + \frac{1}{\tau\mathrm{SNR}} - \beta - \tau\frac{\beta^2 + 4\beta\lambda\tau - 4\lambda^2}{(\beta + 2\lambda\tau)^2} = 0. \tag{15b}$$

The unique positive solution in τ of the above system is $\sqrt{\frac{\left(\frac{\Upsilon(\lambda,\delta)}{1+\Upsilon(\lambda,\delta)}\right)^2 + \frac{1}{\mathrm{SNR}}}{\delta - \frac{1}{(1+\Upsilon(\lambda,\delta))^2}}}$. Similarly to the Box Relaxation case, we can show that $\mathrm{BER} := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\widetilde{\mathbf{w}}_i \leq -1\}} \xrightarrow{\mathrm{P}} Q(\frac{1}{\tau_*})$, which proves Theorem 1 .

4. CONCLUSION

In this paper, we leveraged the CGMT framework to conduct a precise analysis of the BER of the Regularized Least Squares algorithm, with and without Box Relaxation (BRO). When using the (BRO), we proved that there exists a critical value of the SNR, above which the optimal regularizer is zero. We also proved that the unboxed RLS with a regularizer tuned optimally in the BER sense yields the MMSE detector. This analysis might be extended to higher-order constellations, such as QAM and PSK. But, this requires a generalization of the CGMT and should be left to another occasion.

5. REFERENCES

[1] Peter JG Teunissen, "The least-squares ambiguity decorrelation adjustment: a method for fast gps integer ambiguity estimation," *Journal of geodesy*, vol. 70, no. 1-2, pp. 65–82, 1995.

[2] Jeffrey C Lagarias and Andrew M Odlyzko, "Solving low-density subset sum problems," *Journal of the ACM (JACM)*, vol. 32, no. 1, pp. 229–246, 1985.

[3] 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.

[4] 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.

[5] Charles Jeon, Arian Maleki, and Christoph Studer, "On the performance of mismatched data detection in large MIMO systems," *CoRR*, vol. abs/1605.02324, 2016.

[6] David N. C. Tse and Ofer Zeitouni, "Linear multiuser receivers in random environments," *IEEE Trans. Information Theory*, vol. 46, no. 1, pp. 171–205, 2000.

[7] Mihailo Stojnic, "A framework to characterize performance of LASSO algorithms," *CoRR*, vol. abs/1303.7291, 2013.

[8] Samet Oymak, Christos Thrampoulidis, and Babak Hassibi, "The squared-error of generalized LASSO: A precise analysis," in *51st Annual Allerton Conference on Communication, Control, and Computing, Allerton 2013, Allerton Park & Retreat Center, Monticello, IL, USA, October 2-4, 2013*, 2013, pp. 1002–1009.

[9] Christos Thrampoulidis, Ashkan Panahi, Daniel Guo, and Babak Hassibi, "Precise error analysis of the LASSO," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2015, South Brisbane, Queensland, Australia, April 19-24, 2015*, 2015, pp. 3467–3471.

[10] Christos Thrampoulidis, Samet Oymak, and Babak Hassibi, "Regularized linear regression: A precise analysis of the estimation error," in *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, 2015, pp. 1683–1709.

[11] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi, "Precise error analysis of regularized m-estimators in high-dimensions," *CoRR*, vol. abs/1601.06233, 2016.

[12] 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.

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