

Precise Performance Analysis of the Box-Elastic Net under Matrix Uncertainties

Ayed M. Alrashdi, Ismail Ben Atitallah, and Tareq Y. Al-Naffouri

Abstract—In this letter, we consider the problem of recovering an unknown sparse signal from noisy linear measurements, using an enhanced version of the popular Elastic-Net (EN) method. We modify the EN by adding a box-constraint, and we call it the Box-Elastic Net (Box-EN). We assume independent identically distributed (iid) real Gaussian measurement matrix with additive Gaussian noise. In many practical situations, the measurement matrix is not perfectly known, and so we only have a noisy estimate of it. In this work, we precisely characterize the mean squared error and the probability of support recovery of the Box-Elastic Net in the high-dimensional asymptotic regime. Numerical simulations validate the theoretical predictions derived in the paper and also show that the boxed variant outperforms the standard EN.

Index Terms—Elastic Net, squared error, measurement matrix uncertainties, probability of support recovery, box-constraint.

I. INTRODUCTION

The Elastic-Net (EN) [1] is a powerful method to recover an unknown signal \mathbf{x}_0 from noisy linear measurements $\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \mathbf{z}$ by solving the following optimization problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{x}\|_2^2, \quad (1)$$

where \mathbf{H} is the measurement matrix, \mathbf{z} is the noise vector, and $\lambda_1, \lambda_2 > 0$ are the regularization parameters. The EN reduces to the ridge regression for $\lambda_1 \rightarrow 0$, and to the LASSO [2] as $\lambda_2 \rightarrow 0$. It combines the good features of both of these approaches and overcomes many limitations of the LASSO [1]. This combination allows for learning a sparse model where few of the entries are non-zero like the LASSO, while still maintaining the regularization properties of the ridge regression. The EN has been used in many applications [3], [4], [5], [6], [7], [8]. In this paper, we refer to (1) as the *standard* EN, but we focus on a modified version that we call the *Box-EN* which solves the following optimization instead

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{B}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{x}\|_2^2, \quad (2)$$

where, $\mathcal{B} = [l, u]^n$, and $l \leq 0, u \geq 0 \in \mathbb{R}$.

The box constraint, which can be seen as an ℓ_∞ -norm constraint, is used to promote a boundedness constraint. For instance, if the EN aims to recover a digital bounded signal

A. M. Alrashdi is with the Computer, Electrical, and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia, and also with the Department of Electrical Engineering, University of Hail, Hail 55476, Saudi Arabia (e-mail: ayed.alrashdi@kaust.edu.sa).

Ismail Ben Atitallah is with the Department of Electrical Engineering, Harvard University, USA. (e-mail: benatitallah@g.harvard.edu).

T. Y. Al-Naffouri is with the Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia. (e-mail: tareq.alnaffouri@kaust.edu.sa).

through relaxation¹, then it is potentially useful to add the box constraint, thereby yielding the Box-EN. When $l = -u$, the Box-EN is equivalent to Regularized Least Squares, where we use three regularization norms: ℓ_1 , ℓ_2 and ℓ_∞ . This is used to solve simultaneously structured signals [9], for instance signals that are known to be both sparse and bounded, in a high-dimensional setting. Comparing (2) to (1), the only small variation is the "box-constraint" that has been added to (2). However, as we will see later, this small variation assures significant improvement in the performance in scenarios where the entries of \mathbf{x}_0 are bounded or approximately so. Examples of such signals are found in many applications such as wireless communication systems [10], image processing [11], etc. The Box-EN is not as popular as the standard EN, but we can find some references where different types of constraints were imposed on the algorithm [12], [13], [14].

In the last few years, various forms of precise analysis of the asymptotic estimation error of non-smooth regularized convex optimization problems (such as LASSO and related inverse problems) have been proved under the assumption of noisy iid Gaussian measurements. They mainly follow one of two parallel approaches. The first one is the Approximate Message Passing (AMP) framework that has been used in [15], [16], [17] to derive precise asymptotic analysis of the LASSO performance under the assumptions of iid Gaussian sensing matrix. The second approach is based on a recently developed framework that uses the Convex Gaussian Min-max Theorem (CGMT) [18]. It has been used in a series of works to also precisely evaluate the estimation performance of these problems under the same Gaussianity assumptions [19], [20], [21].

However, these results assume that the measurement matrix \mathbf{H} is perfectly known. In many practical applications it is reasonable to expect uncertainty in the measurement matrix due to, e.g., imperfections in the signal acquisition hardware, model mismatch, or estimation errors [22]. In this paper, we consider the *additive* uncertainty model: $\mathbf{A} = \sqrt{1 - \epsilon^2} \mathbf{H} + \epsilon \mathbf{\Omega}$, where \mathbf{H} is known, $\mathbf{\Omega}$ is an unknown error matrix, and $\epsilon^2 \in [0, 1]$ is the variance of the error. Such model is commonly used in communication theory and known as imperfect Channel State Information (CSI) [23], [24], [25].

In this work, we derive novel precise asymptotic (in the problem dimensions m and n) characterizations of the *mean squared error* and the *support recovery probability* of the Box-EN under the presence of uncertainties in the measurement matrix that has iid Gaussian entries using the CGMT

¹without the relaxation, having a discrete feasible set lead to a computationally prohibitive algorithm.

framework. Although our analysis is asymptotic in nature, numerical simulations show that our predictions are valid even for a few dozens of the problem dimensions. They also demonstrate that the Box-EN outperforms the standard one. These predictions can be used to optimally tune the involved parameters of the algorithm. To the best of our knowledge, the precise error analysis of the Box-EN under uncertainties has not been explicitly derived before. Finally, we note that following the same steps as in this paper, our results can be used to characterize the performance of the standard EN as well.

II. PROBLEM SETUP

A. System Model

We consider the problem of recovering an unknown signal $\mathbf{x}_0 \in \mathbb{R}^n$ from a noisy linear measurement vector $\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \mathbf{z}$. The unknown signal vector \mathbf{x}_0 is assumed to be k -sparse, i.e., only k of its entries are sampled iid from a distribution p_{X_0} which has zero mean and unit variance ($\mathbb{E}[X_0^2] = 1$), and the remaining entries are zeros. For the measurement matrix, we consider the following additive uncertainty model: $\mathbf{A} = \gamma\mathbf{H} + \epsilon\mathbf{\Omega}$, where \mathbf{H} , and $\mathbf{\Omega} \in \mathbb{R}^{m \times n}$ have iid entries $\mathcal{N}(0, 1/n)$, and $\epsilon^2 \in [0, 1]$ is the variance of the error such that $\gamma^2 + \epsilon^2 = 1$. The noise vector $\mathbf{z} \in \mathbb{R}^m$ has entries iid $\mathcal{N}(0, \sigma_z^2)$. The analysis is performed when the system dimensions (m , n and k) grow simultaneously large at fixed ratios: $\frac{m}{n} \rightarrow \delta \in (0, \infty)$, and $\frac{k}{n} \rightarrow \kappa \in (0, 1)$. Under these settings, the signal-to-noise ratio (SNR) becomes $\text{SNR} := \frac{\kappa}{\sigma_z^2}$.

B. Performance Metrics

We consider the following two performance metrics of the Box-Elastic Net:

Mean squared error (MSE): A natural and heavily used measure of performance is the reconstruction *mean squared error*, which measures the deviation of $\hat{\mathbf{x}}$ from the true signal \mathbf{x}_0 . Formally, the MSE is defined as $\text{MSE} := \frac{1}{n} \|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2$.

Support Recovery: In the problem of sparse recovery, a natural measure of performance that is used in many applications is support recovery, which is defined as identifying whether an entry of \mathbf{x}_0 is on the support (i.e., non-zero), or it is off the support (i.e., zero). The decision is based on the Box-EN solution $\hat{\mathbf{x}}$: we say the i^{th} entry of $\hat{\mathbf{x}}$ is on the support if $|\hat{x}_i| \geq \xi$, where $\xi > 0$ is a user-defined hard threshold on the entries of $\hat{\mathbf{x}}$. Formally, let

$$\Phi_{\xi, \text{on}}(\hat{\mathbf{x}}) = \frac{1}{k} \sum_{i \in S(\mathbf{x}_0)} \mathbb{1}_{\{|\hat{x}_i| \geq \xi\}}, \quad (3a)$$

$$\Phi_{\xi, \text{off}}(\hat{\mathbf{x}}) = \frac{1}{n-k} \sum_{i \notin S(\mathbf{x}_0)} \mathbb{1}_{\{|\hat{x}_i| \leq \xi\}}, \quad (3b)$$

where $\mathbb{1}_{\{\mathcal{A}\}}$ is the indicator function of a set \mathcal{A} , and $S(\mathbf{x}_0)$ is the support of \mathbf{x}_0 , i.e., the set of the non-zero entries of \mathbf{x}_0 . In Theorem 2, we precisely predict the *per-entry* rate of successful on-support and off-support recovery.

III. APPROACH AND SKETCH OF THE PROOF

In this section, we provide the asymptotic analysis of the MSE of the proposed Box-EN optimization. Our approach is based on the recently developed Convex Gaussian Min-max Theorem (CGMT) which is summarized in the next subsection.

A. Convex Gaussian Min-max Theorem (CGMT)

We first need to state the key ingredient of the analysis which is the Convex Gaussian Min-max Theorem CGMT. Here, we recall the statement of the theorem, and we refer the reader to [18] for the complete technical requirements. Consider the following two min-max problems, which we refer to as the Primary Optimization (PO) and Auxiliary Optimization (AO):

$$\Phi(\mathbf{G}) := \min_{\mathbf{w} \in \mathcal{S}_w} \max_{\mathbf{u} \in \mathcal{S}_u} \mathbf{u}^T \mathbf{G} \mathbf{w} + \psi(\mathbf{w}, \mathbf{u}), \quad (4a)$$

$$\phi(\mathbf{g}, \mathbf{h}) := \min_{\mathbf{w} \in \mathcal{S}_w} \max_{\mathbf{u} \in \mathcal{S}_u} \|\mathbf{w}\|_2 \mathbf{g}^T \mathbf{u} - \|\mathbf{u}\|_2 \mathbf{h}^T \mathbf{w} + \psi(\mathbf{w}, \mathbf{u}), \quad (4b)$$

where $\mathbf{G} \in \mathbb{R}^{m \times n}$, $\mathbf{g} \in \mathbb{R}^m$, $\mathbf{h} \in \mathbb{R}^n$, $\mathcal{S}_w \subset \mathbb{R}^m$, $\mathcal{S}_u \subset \mathbb{R}^n$ 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 of (4a) and (4b) respectively. Let $\mathcal{S}_w, \mathcal{S}_u$ be convex, and assume that $\psi(\mathbf{w}, \mathbf{u})$ is convex-concave continuous 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 $\lim_{n \rightarrow \infty} \mathbb{P}[\mathbf{w}_\phi \in \mathcal{S}] = 1$, it also holds that $\lim_{n \rightarrow \infty} \mathbb{P}[\mathbf{w}_\Phi \in \mathcal{S}] = 1$.

In a nutshell, we study the performance of the (PO) by analyzing its corresponding (AO) which is much easier to study, since it depends on the random vectors \mathbf{g} and \mathbf{h} instead of the large random matrix \mathbf{G} , simplifying the latter to a Scalar Optimization (SO), and finally studying the asymptotic performance of the (SO). To make use of the CGMT, the set \mathcal{S} should be properly chosen as the set in which the MSE concentrates. In the following subsection, we start the MSE analysis by specializing the CGMT to the Box-EN problem at hand and identifying its corresponding (PO) and (AO).

B. Identifying the (PO) and the (AO)

Under the imperfect measurements assumption, the Box-EN optimization in (2) becomes

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{B}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{x}\|_2^2. \quad (5)$$

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

$$\mathcal{B}' = \{\mathbf{w} \in \mathbb{R}^n : l - \mathbf{x}_{0,i} \leq \mathbf{w}_i \leq u - \mathbf{x}_{0,i}, i \in \{1, 2, \dots, n\}\}, \quad (6)$$

then the problem in (5) can be reformulated as

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathcal{B}'} \|\mathbf{H}\mathbf{w} + \frac{\epsilon}{\gamma} \mathbf{\Omega}(\mathbf{w} + \mathbf{x}_0) - \mathbf{z}\|_2^2 + \frac{\lambda_1}{\gamma} \|\mathbf{w} + \mathbf{x}_0\|_1 + \frac{\lambda_2}{\gamma^2} \|\mathbf{w} + \mathbf{x}_0\|_2^2.$$

This minimization is not in the (PO) form as it is missing the max part. So to go around this, let us express the loss function in its dual form through the Fenchel conjugate, $\|\mathbf{H}\mathbf{w} + \frac{\epsilon}{\gamma} \mathbf{\Omega}(\mathbf{w} + \mathbf{x}_0) - \mathbf{z}\|_2^2 = \max_{\mathbf{u}} \sqrt{n} \mathbf{u}^T (\mathbf{H}\mathbf{w} + \frac{\epsilon}{\gamma} \mathbf{\Omega}(\mathbf{w} + \mathbf{x}_0) - \mathbf{z})$

$\mathbf{x}_0) - \mathbf{z}) - \frac{n}{4} \|\mathbf{u}\|_2^2$. Hence, the problem above is equivalent to the following

$$\begin{aligned} \min_{\mathbf{w} \in \mathcal{B}'} \max_{\mathbf{u}} & \sqrt{n} \mathbf{u}^T \mathbf{H} \mathbf{w} + \frac{\sqrt{n} \epsilon}{\gamma} \mathbf{u}^T \boldsymbol{\Omega} (\mathbf{w} + \mathbf{x}_0) - \sqrt{n} \mathbf{u}^T \mathbf{z} \\ & - \frac{n}{4} \|\mathbf{u}\|_2^2 + \frac{\lambda_1}{\gamma} \|\mathbf{w} + \mathbf{x}_0\|_1 + \frac{\lambda_2}{\gamma^2} \|\mathbf{w} + \mathbf{x}_0\|_2^2. \end{aligned} \quad (7)$$

To reach the desired (PO) form, we introduce the variables

$$\begin{aligned} \mathbf{v} &= \begin{bmatrix} \mathbf{w} \\ \frac{\epsilon}{\gamma} (\mathbf{w} + \mathbf{x}_0) \end{bmatrix} \in \mathbb{R}^{2n}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{H} & \boldsymbol{\Omega} \end{bmatrix} \in \mathbb{R}^{m \times 2n} \text{ and} \\ \mathbf{C} &= \begin{bmatrix} \mathbf{O}_n & \frac{\gamma}{\epsilon} \mathbf{I}_n \end{bmatrix} \in \mathbb{R}^{n \times 2n}, \text{ where } \mathbf{O}_n \text{ and } \mathbf{I}_n \text{ are } n \times n \\ & \text{matrices that represent the all-zeros matrix and the identity} \\ & \text{matrix, respectively and use them to rewrite (7) in the desired} \\ & \text{(PO) form as} \end{aligned}$$

$$\begin{aligned} \min_{\mathbf{v} \in \mathcal{D}} \max_{\mathbf{u}} & \sqrt{n} \mathbf{u}^T \mathbf{G} \mathbf{v} - \sqrt{n} \mathbf{u}^T \mathbf{z} - \frac{n}{4} \|\mathbf{u}\|_2^2 \\ & + \frac{\lambda_1}{\gamma} \|\mathbf{C} \mathbf{v}\|_1 + \frac{\lambda_2}{\gamma^2} \|\mathbf{C} \mathbf{v}\|_2^2, \end{aligned}$$

where $\mathcal{D} = \{\mathbf{v}^T = [\mathbf{w}^T \quad \frac{\epsilon}{\gamma} (\mathbf{w} + \mathbf{x}_0)^T] : \mathbf{w} \in \mathcal{B}'\}$. The corresponding (AO) problem is thus given by

$$\begin{aligned} \min_{\mathbf{v} \in \mathcal{D}} \max_{\mathbf{u}} & \|\mathbf{v}\|_2 \mathbf{g}^T \mathbf{u} - \|\mathbf{u}\|_2 \boldsymbol{\mu}^T \mathbf{v} - \frac{n}{4} \|\mathbf{u}\|_2^2 \\ & - \sqrt{n} \mathbf{u}^T \mathbf{z} + \frac{\lambda_1}{\gamma} \|\mathbf{C} \mathbf{v}\|_1 + \frac{\lambda_2}{\gamma^2} \|\mathbf{C} \mathbf{v}\|_2^2, \end{aligned} \quad (8)$$

where $\boldsymbol{\mu} \in \mathbb{R}^{2n}$ and $\mathbf{g} \in \mathbb{R}^m$ are independent standard normal vectors.

C. Simplifying the (AO)

The next step is to simplify the (AO) to a scalar optimization (SO) problem. Since the vectors \mathbf{g} and \mathbf{z} are independent, $\|\mathbf{v}\|_2 \mathbf{g}^T \mathbf{u} - \sqrt{n} \mathbf{u}^T \mathbf{z}$ is equivalent in distribution to $\sqrt{\|\mathbf{v}\|_2^2 + n \sigma_z^2} \mathbf{g}^T \mathbf{u}$. Therefore, the (AO) can be rewritten as

$$\begin{aligned} \min_{\mathbf{v} \in \mathcal{D}} \max_{\mathbf{u}} & \sqrt{\|\mathbf{v}\|_2^2 + n \sigma_z^2} \mathbf{g}^T \mathbf{u} - \|\mathbf{u}\|_2 \boldsymbol{\mu}^T \mathbf{v} - \frac{n}{4} \|\mathbf{u}\|_2^2 \\ & + \frac{\lambda_1}{\gamma} \|\mathbf{C} \mathbf{v}\|_1 + \frac{\lambda_2}{\gamma^2} \|\mathbf{C} \mathbf{v}\|_2^2. \end{aligned} \quad (9)$$

Let us write (9) now in terms of the original \mathbf{w} variable

$$\begin{aligned} \min_{\mathbf{w} \in \mathcal{B}'} \max_{\mathbf{u}} & \sqrt{\|\mathbf{w}\|_2^2 + \frac{\epsilon^2}{\gamma^2} \|\mathbf{w} + \mathbf{x}_0\|_2^2 + n \sigma_z^2} \mathbf{g}^T \mathbf{u} - \frac{n}{4} \|\mathbf{u}\|_2^2 \\ & - \|\mathbf{u}\|_2 (\mathbf{h}_1^T \mathbf{w} + \frac{\epsilon}{\gamma} \mathbf{h}_2^T (\mathbf{w} + \mathbf{x}_0)) \\ & + \frac{\lambda_1}{\gamma} \|\mathbf{w} + \mathbf{x}_0\|_1 + \frac{\lambda_2}{\gamma^2} \|\mathbf{w} + \mathbf{x}_0\|_2^2, \end{aligned}$$

where $\mathbf{h}_1, \mathbf{h}_2 \in \mathbb{R}^n$ are independent standard normal vectors. Alternatively, since $\mathbf{w} = \gamma \mathbf{x} - \mathbf{x}_0$, we can express the above optimization over \mathbf{x} instead as

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{B}} \max_{\mathbf{u}} & \sqrt{\|\mathbf{x}\|_2^2 + \|\mathbf{x}_0\|_2^2 - 2\gamma \mathbf{x}_0^T \mathbf{x} + n \sigma_z^2} \mathbf{g}^T \mathbf{u} - \frac{n}{4} \|\mathbf{u}\|_2^2 \\ & - \|\mathbf{u}\|_2 (\gamma \mathbf{h}_1 + \epsilon \mathbf{h}_2)^T \mathbf{x} + \|\mathbf{u}\|_2 \mathbf{h}_1^T \mathbf{x}_0 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{x}\|_2^2. \end{aligned}$$

Fixing the norm of \mathbf{u} to $\beta := \|\mathbf{u}\|_2$, we can easily optimize over its direction by aligning it with \mathbf{g} further simplifying the (AO) to

$$\begin{aligned} \max_{\beta \geq 0} \min_{\mathbf{x} \in \mathcal{B}} & \sqrt{\frac{1}{n} (\|\mathbf{x}\|_2^2 + \|\mathbf{x}_0\|_2^2 - 2\gamma \mathbf{x}_0^T \mathbf{x}) + \sigma_z^2} \beta \|\mathbf{g}\|_2 \\ & - \beta (\gamma \mathbf{h}_1 + \epsilon \mathbf{h}_2)^T \mathbf{x} + \beta \mathbf{h}_1^T \mathbf{x}_0 - \frac{n \beta^2}{4} + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{x}\|_2^2. \end{aligned} \quad (10)$$

To have a separable optimization problem, we use the following variational form: $\sqrt{\chi} = \min_{\alpha > 0} \frac{\alpha}{2} + \frac{\chi}{2\alpha}$, where

$$\chi = \frac{1}{n} (\|\mathbf{x}\|_2^2 + \|\mathbf{x}_0\|_2^2 - 2\gamma \mathbf{x}_0^T \mathbf{x}) + \sigma_z^2. \quad (11)$$

which reduces (10) to

$$\begin{aligned} \min_{\tau > 0} \max_{\beta \geq 0} & \frac{\beta \tau \|\mathbf{g}\|_2^2}{2} + \frac{n \beta \sigma_z^2}{2\tau} - \frac{n \beta^2}{4} + \frac{\beta \|\mathbf{x}_0\|_2^2}{2\tau} \\ & + \beta \mathbf{h}_1^T \mathbf{x}_0 + \sum_{i=1}^n \left(\min_{l \leq \mathbf{x}_i \leq u} \left(\frac{\beta}{2\tau} + \lambda_2 \right) \mathbf{x}_i^2 \right. \\ & \left. - \beta \left(\mathbf{h}_i + \frac{\gamma \mathbf{x}_{0,i}}{\tau} \right) \mathbf{x}_i + \lambda_1 |\mathbf{x}_i| \right), \end{aligned} \quad (12)$$

where $\tau := \frac{\sqrt{n} \alpha}{\|\mathbf{g}\|_2}$, and $\theta = \frac{\beta}{2\tau} + \lambda_2$. We need to prove that the optimal β_* , denoted by β_n , is positive almost surely. To do so, suppose by contradiction that $\beta_n = 0$. First, it is possible to show using problem (10) that $\beta_n = 2 \max\{0, \sqrt{\chi} \frac{\|\mathbf{g}\|_2}{\sqrt{n}} - \frac{1}{n} \tilde{\mathbf{h}}^T \mathbf{x} + \frac{1}{n} \mathbf{h}_1^T \mathbf{x}_0\}$, where $\tilde{\mathbf{h}} = \gamma \mathbf{h}_1 + \epsilon \mathbf{h}_2$. If $\beta_n = 0$, then this is equivalent to say that

$$\sqrt{\chi} \frac{\|\mathbf{g}\|_2}{\sqrt{n}} - \frac{1}{n} \tilde{\mathbf{h}}^T \mathbf{x} + \frac{1}{n} \mathbf{h}_1^T \mathbf{x}_0 \leq 0. \quad (13)$$

If $\beta_n = 0$, the problem in (12) simplifies to the following: $\sum_{i=1}^n \min_{l \leq \mathbf{x}_i \leq u} \lambda_2 \mathbf{x}_i^2 + \lambda_1 |\mathbf{x}_i|$, which implies that the optimal \mathbf{x} is the zero vector. Consequently, $\sqrt{\chi} \frac{\|\mathbf{g}\|_2}{\sqrt{n}} - \frac{1}{n} \tilde{\mathbf{h}}^T \mathbf{x} + \frac{1}{n} \mathbf{h}_1^T \mathbf{x}_0 = \sqrt{\frac{1}{n} \|\mathbf{x}_0\|_2^2 + \sigma_z^2} \frac{\|\mathbf{g}\|_2}{\sqrt{n}} + \frac{1}{n} \mathbf{h}_1^T \mathbf{x}_0$. Recall that \mathbf{h}_1 is a standard Gaussian vector and then $\frac{1}{n} \mathbf{h}_1^T \mathbf{x}_0$ converges almost surely to zero. Hence, the latter quantity is almost surely strictly positive, which contradicts (13), i.e. the assumption $\beta_n = 0$. Thus, $\beta_n > 0$ almost surely. After some algebraic manipulations, we can write

$$\begin{aligned} \min_{\tau > 0} \max_{\beta > 0} & \frac{\beta \tau \|\mathbf{g}\|_2^2}{2} + \frac{n \beta \sigma_z^2}{2\tau} - \frac{n \beta^2}{4} + \beta \mathbf{h}_1^T \mathbf{x}_0 \\ & + \sum_{i=1}^n \left(\frac{\beta}{2\tau} - \frac{\beta^2 \gamma^2}{4\theta \tau^2} \right) \mathbf{x}_{0,i}^2 - \frac{\beta^2 \gamma}{2\theta \tau} \mathbf{h}_i \mathbf{x}_{0,i} - \frac{\beta^2}{4\theta} \mathbf{h}_i^2 \\ & + 2\theta \left(\sum_{i=1}^n \min_{l \leq \mathbf{x}_i \leq u} \frac{1}{2} (\mathbf{x}_i - \frac{\beta}{2\theta} (\frac{\gamma}{\tau} \mathbf{x}_{0,i} + \mathbf{h}_i))^2 + \frac{\lambda_1}{2\theta} |\mathbf{x}_i| \right). \end{aligned}$$

The optimization over \mathbf{x}_i can be solved in closed-form using the *saturated* soft-thresholding operator $\eta(a; \lambda, l, u) = \operatorname{argmin}_{l \leq x \leq u} \frac{1}{2} (x - a)^2 + \lambda |x|$ defined as:

$$\eta(a; \lambda, l, u) = \begin{cases} u & , \text{if } a > u + \lambda \\ a - \lambda & , \text{if } \lambda < a < u + \lambda \\ 0 & , \text{if } |a| \leq \lambda \\ a + \lambda & , \text{if } l - \lambda < a < -l \\ l & , \text{if } a \leq l - \lambda. \end{cases} \quad (14)$$

Also, let $e(a; \lambda, l, u) = \min_{l \leq x \leq u} \frac{1}{2} (x - a)^2 + \lambda |x|$ which is defined as:

$$e(a; \lambda, l, u) = \begin{cases} \frac{1}{2} (u - a)^2 + \lambda u & , \text{if } a > u + \lambda \\ \lambda a - \frac{1}{2} \lambda^2 & , \text{if } \lambda < a < u + \lambda \\ \frac{1}{2} a^2 & , \text{if } |a| \leq \lambda \\ -\lambda a - \frac{1}{2} \lambda^2 & , \text{if } l - \lambda < a < -l \\ \frac{1}{2} (l - a)^2 - \lambda l & , \text{if } a \leq l - \lambda. \end{cases} \quad (15)$$

Then, the above optimization problem finally simplifies to the following Scalar Optimization (SO) problem:

$$\begin{aligned} \min_{\tau > 0} \max_{\beta > 0} & \tilde{D}(\tau, \beta, \mathbf{g}, \mathbf{h}) := \frac{\beta \tau \|\mathbf{g}\|_2^2}{2} + \frac{n \beta \sigma_z^2}{2\tau} - \frac{n \beta^2}{4} \\ & + \sum_{i=1}^n \left(\frac{\beta}{2\tau} - \frac{\beta^2 \gamma^2}{4\theta \tau^2} \right) \mathbf{x}_{0,i}^2 + \left(\beta - \frac{\beta^2 \gamma}{2\theta \tau} \right) \mathbf{h}_i \mathbf{x}_{0,i} - \frac{\beta^2}{4\theta} \mathbf{h}_i^2 \\ & + 2\theta \sum_{i=1}^n e \left(\frac{\gamma \beta}{2\theta \tau} \mathbf{x}_{0,i} + \frac{\beta}{2\theta} \mathbf{h}_i; \frac{\lambda_1}{2\theta}, l, u \right). \end{aligned} \quad (16)$$

D. Probabilistic asymptotic analysis of the (SO) problem

After simplifying the (AO) to its form in (16), we are now in a position to analyze its limiting behavior. First, we need to properly normalize the objective function in (16) by dividing it by n . Then, using the weak law of large numbers (WLLN), we have²: $\frac{1}{n}\|\mathbf{g}\|_2^2 \xrightarrow{P} \delta$, $\frac{1}{n}\|\mathbf{h}\|_2^2 \xrightarrow{P} 1$, $\frac{1}{n}\|\mathbf{x}_0\|_2^2 \xrightarrow{P} \kappa$ and $\frac{1}{n}\mathbf{h}^T \mathbf{x}_0 \xrightarrow{P} 0$. Also, using the WLLN, it can be shown that for all $\tau > 0$ and $\beta > 0$, $\frac{1}{n} \sum_{i=1}^n e^{(\frac{\gamma\beta}{2\theta\tau}\mathbf{x}_{0,i} + \frac{\beta}{2\theta}\mathbf{h}_i; \frac{\lambda_1}{2\theta}, l, u)} \xrightarrow{P} \mathbb{E}[e^{(\frac{\gamma\beta}{2\theta\tau}X_0 + \frac{\beta}{2\theta}H; \frac{\lambda_1}{2\theta}, l, u)}]$, and $\frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{x}}_i \xrightarrow{P} \mathbb{E}[\eta(\frac{\gamma\beta}{2\theta\tau}X_0 + \frac{\beta}{2\theta}H; \frac{\lambda_1}{2\theta}, l, u)]$, where $\tilde{\mathbf{x}}$ is the solution of (AO) defined in (10) and H is a scalar standard normal random variable. Therefore, the point-wise convergence in τ and β of the objective function in (16) is $D(\tau, \beta) := \frac{\beta\tau\delta}{2} + \frac{\beta\sigma_z^2}{2\tau} - \frac{\beta^2}{4\theta}(\theta + 1) + (\frac{\beta}{2\tau} - \frac{\beta^2\gamma^2}{4\theta\tau^2})\kappa + 2\theta\mathbb{E}_{X_0, H}[e^{(\frac{\gamma\beta}{2\theta\tau}X_0 + \frac{\beta}{2\theta}H; \frac{\lambda_1}{2\theta}, l, u)}]$. Furthermore, it is possible to show that with probability one, the functions $\tau \mapsto \max_{\beta>0} \tilde{D}(\tau, \beta, \mathbf{g}, \mathbf{h})$ and $\tau \mapsto \max_{\beta>0} D(\tau, \beta)$ are convex in τ . Hence, it is possible to show using Theorem 2.7 in [26] that $\tau_n(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \tau_*$. Likewise, it can be similarly proved that $\beta_n(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \beta_*$.

E. Applying the CGMT

Now, we are in a position to study the convergence limit of the MSE of the Box-EN. First, recall from (11) that $\frac{1}{n}\|\tilde{\mathbf{x}} - \mathbf{x}_0\|_2^2 = \frac{\|\mathbf{g}\|_2^2}{4\theta} \tau_n^2(\mathbf{g}, \mathbf{h}) + \frac{2}{n}(\gamma - 1)\tilde{\mathbf{x}}^T \mathbf{x}_0 - \sigma_z^2$. Moreover, it can be shown that $\frac{1}{n}\tilde{\mathbf{x}}^T \mathbf{x}_0 \xrightarrow{P} \mathbb{E}[\eta(\frac{\gamma\beta_*}{2\theta_*\tau_*}X_0 + \frac{\beta_*}{2\theta_*}H; \frac{\lambda_1}{2\theta_*}, l, u)X_0] := \Upsilon(\tau_*, \beta_*)$. Using $\tau_n(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \tau_*$, and $\beta_n(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \beta_*$, we can show that:

$$\begin{aligned} \frac{1}{n}\|\tilde{\mathbf{x}} - \mathbf{x}_0\|_2^2 &\xrightarrow{P} \delta\tau_*^2 - \sigma_z^2 + 2(\gamma - 1)\Upsilon(\tau_*, \beta_*) \\ &:= M(\tau_*, \beta_*). \end{aligned} \quad (17)$$

The last step is to use the CGMT to prove that the quantities $\tilde{\mathbf{x}} - \mathbf{x}_0$ and $\tilde{\mathbf{x}} - \mathbf{x}_0$ are concentrated in the same set. Formally, for any fixed $\zeta > 0$, we define the set $\mathcal{S} = \{\mathbf{r} : |\frac{1}{n}\|\mathbf{r}\|_2^2 - M(\tau_*, \beta_*)| < \zeta\}$. Equation (17) proves that for any $\zeta > 0$, $\tilde{\mathbf{x}} - \mathbf{x}_0 \in \mathcal{S}$ with probability one. Then, we conclude using the CGMT that $\hat{\mathbf{x}} - \mathbf{x}_0 \in \mathcal{S}$ with probability one. The asymptotic prediction of the MSE is summarized in Theorem 1 given in the next section.

IV. MAIN RESULTS

This section summarizes our main results on the precise analysis of the mean squared error and the probability of support recovery of the Box-EN. We use standard notation $\text{plim } X_n = X$ to denote that a sequence of random variables X_n converges in probability towards a constant X .

Theorem 1 (Box-EN MSE). *Fix $\lambda_1, \lambda_2 > 0$, and let $\hat{\mathbf{x}}$ be a minimizer of the Box-EN problem in (5), where \mathbf{A}, \mathbf{z} and \mathbf{x}_0 satisfy the working assumptions of Section II-A. Then, in the limit of $m, n \rightarrow +\infty, m/n = \delta$, it holds:*

$$\begin{aligned} \text{plim } \frac{1}{n}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 &= \delta\tau_*^2 - \sigma_z^2 \\ &+ 2(\gamma - 1)\mathbb{E}_{\substack{X_0 \sim P_{X_0} \\ H \sim \mathcal{N}(0,1)}} \left[\eta \left(\frac{\gamma\beta_*}{2\theta_*\tau_*}X_0 + \frac{\beta_*}{2\theta_*}H; \frac{\lambda_1}{2\theta_*}, l, u \right) X_0 \right], \end{aligned} \quad (18)$$

where $\theta_* = \frac{\beta_*}{2\tau_*} + \lambda_2$, and (τ_*, β_*) is the unique solution to the following:

$$\begin{aligned} \min_{\tau>0} \max_{\beta>0} D(\tau, \beta) &:= \frac{\beta\tau\delta}{2} + \frac{\beta\sigma_z^2}{2\tau} - \frac{\beta^2}{4\theta}(\theta + 1) \\ &+ \left(\frac{\beta}{2\tau} - \frac{\beta^2\gamma^2}{4\theta\tau^2} \right) \kappa + 2\theta\mathbb{E}_{X_0, H} \left[e^{(\frac{\gamma\beta}{2\theta\tau}X_0 + \frac{\beta}{2\theta}H; \frac{\lambda_1}{2\theta}, l, u)} \right]. \end{aligned} \quad (19)$$

²We write “ \xrightarrow{P} ” to designate convergence in probability.

Remark 1 (Optimal Solutions). Note that τ_* and β_* can be efficiently computed by writing the first order stationary point conditions, i.e., $\nabla_{(\tau, \beta)} D(\tau, \beta) = 0$.

Remark 2 (Optimal Regularizers). Theorem 1 can be used to find the optimal pair of regularizers (λ_1, λ_2) that minimizes the MSE. **Remark 3**. For a sparse Bernoulli vector \mathbf{x}_0 , the expectation in (18) is given by:

$$\begin{aligned} \mathbb{E} \left[\eta \left(\frac{\gamma\beta_*}{2\theta_*\tau_*}X_0 + \frac{\beta_*}{2\theta_*}H; \frac{\lambda_1}{2\theta_*}, l, u \right) X_0 \right] &= \kappa u Q \left(\frac{2\theta_*(u - \xi) + \lambda_1}{\beta_*} \right) \\ &+ \kappa l Q \left(\frac{2\theta_*(\xi - l) + \lambda_1}{\beta_*} \right) + \kappa \int_{\frac{\lambda_1 - 2\theta_*\xi}{\beta_*}}^{\frac{2\theta_*(u - \xi) + \lambda_1}{\beta_*}} \left(\xi + \frac{\beta_*h - \lambda_1}{2\theta_*} \right) p(h) dh \\ &+ \kappa \int_{\frac{-2\theta_*\xi - \lambda_1}{\beta_*}}^{\frac{-2\theta_*\xi - \lambda_1}{\beta_*}} \left(\xi + \frac{\beta_*h + \lambda_1}{2\theta_*} \right) p(h) dh, \end{aligned} \quad (20)$$

where $p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is the pdf of a standard Gaussian random variable and $Q(x)$ is its associated Q-function. The expectation in (19) can be found in a similar way.

The following Theorem precisely characterizes the support recovery metrics introduced in (3).

Theorem 2 (Probability of support recovery). *Under the same settings of Theorem 1 and for any fixed $\xi > 0$, and in the limit of $m, n \rightarrow +\infty, m/n = \delta$, it holds that:*

$$\text{plim } \Phi_{\xi, \text{on}}(\hat{\mathbf{x}}) = \mathbb{P} \left[\left| \eta \left(\frac{\gamma\beta_*}{2\theta_*\tau_*}X_0 + \frac{\beta_*}{2\theta_*}H; \frac{\lambda_1}{2\theta_*}, l, u \right) \right| \geq \xi \right], \quad (21)$$

and

$$\text{plim } \Phi_{\xi, \text{off}}(\hat{\mathbf{x}}) = \mathbb{P} \left[\left| \eta \left(\frac{\beta_*}{2\theta_*}H; \frac{\lambda_1}{2\theta_*}, l, u \right) \right| \leq \xi \right]. \quad (22)$$

The proof of Theorem 2 is also based on the CGMT and largely follows the proof of Theorem 1 but is omitted for space limitations (see [19] for a similarly detailed treatment).

Remark 1 (Small/Large Regularizers). From Theorem 2, it can be seen that as λ_1 becomes large $\Phi_{\xi, \text{off}}(\hat{\mathbf{x}})$ converges to one while $\Phi_{\xi, \text{on}}(\hat{\mathbf{x}})$ converges to zero. Opposite behavior is observed when λ_1 takes values close to zero. This behavior is expected since large values of λ_1 parameter put more emphasis on the ℓ_1 -norm term, thus promoting sparser solution. This is clearly depicted in Figure 3.

Remark 2 (Optimal Regularizers). In order to trade-off between on- and off- support recovery probabilities a reasonable performance metric will be $\Phi_\xi = \omega\Phi_{\xi, \text{on}} + (1 - \omega)\Phi_{\xi, \text{off}}$ for $\omega \in [0, 1]$. Theorem 2 precisely characterizes the behavior of this as a function of (λ_1, λ_2) ; thus, it determines the optimal value of (λ_1, λ_2) that minimizes Φ_ξ [27].

Remark 3. The parameters l and u generally do not need to be tuned because the support of the signal is known (take the example of digital communication systems where the constellation is known by both the transmitter and the receiver). Hence, the choice of l and u is straightforward, i.e. $l = \min \mathbf{x}_0$ and $u = \max \mathbf{x}_0$.

Remark 4. For a sparse Bernoulli vector \mathbf{x}_0 , the expressions in (21), and (22) simplify to

$$\begin{aligned} \text{plim } \Phi_{\xi, \text{on}}(\hat{\mathbf{x}}) &= Q \left(\frac{2\theta_*\xi + \lambda_1}{\beta_*} + \frac{\gamma}{\tau_*} \right) + Q \left(\frac{2\theta_*l - \lambda_1}{\beta_*} - \frac{\gamma}{\tau_*} \right) \\ &+ Q \left(\frac{2\theta_*\xi + \lambda_1}{\beta_*} - \frac{\gamma}{\tau_*} \right) - 1, \end{aligned} \quad (23)$$

and

$$\text{plim } \Phi_{\xi, \text{off}}(\hat{\mathbf{x}}) = 1 - Q \left(\frac{2\theta_*\xi + \lambda_1}{\beta_*} \right). \quad (24)$$

Remark 5. Theorems 1 and 2 can also be used to characterize the performance of the standard EN by letting $l \rightarrow -\infty$, and $u \rightarrow \infty$, in which the saturated soft-thresholding operator reduces to the conventional soft-thresholding operator. To the best of our knowledge, the precise analysis of the standard EN under matrix uncertainties has not been studied before as well. However, we focus in this paper on the Box-EN since it generalizes the standard EN and gives better performance as compare to the standard one.

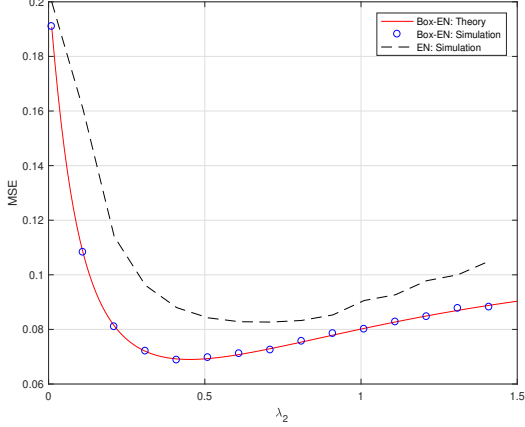


Fig. 1. MSE of the Box-EN and the EN. This figure shows that the former outperforms the latter. Theoretical prediction is based on Theorem 1 with $\lambda_1 = 0.1$. For the simulations, we used $\kappa = 0.1$, $\epsilon^2 = 0.1$, $\delta = 0.7$, $n = 500$, $\text{SNR} = 0.5$.

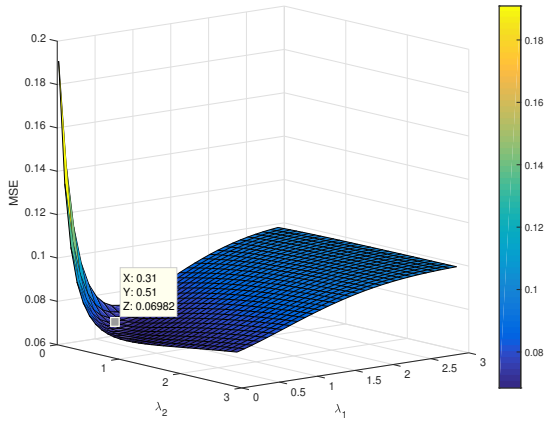


Fig. 2. MSE of the Box-EN as a function of λ_1 , λ_2 for a sparse Bernoulli signal. We used $\delta = 0.7$, $\kappa = 0.1$, $\epsilon^2 = 0.1$, $n = 500$, and $\text{SNR} = 0.5$.

V. NUMERICAL RESULTS

For illustration, we focus only on the case where \mathbf{x}_0 has entries that are sampled iid from a sparse-Bernoulli distribution with $\mathbb{P}[X_0 = 0] = 0.9$ and $\mathbb{P}[X_0 = 1] = 0.1$ (i.e. $\kappa = 0.1$). A natural choice for the box-constraint values in this situation is to set $l = 0$, and $u = 1$. Figure 3 shows the close agreement between the asymptotic mean squared error (MSE) of the Box-EN as predicted by Theorem 1 and numerical simulations. The simulation results are averages over 100 realizations of \mathbf{A} and \mathbf{z} with $n = 500$ and $\delta = 0.7$. This figure also demonstrates that the Box-EN outperforms the standard EN.

From Figure 3, we can see that as the regularizer λ_2 is varied, we notice a pronounced minimum for some $\lambda_2 > 0$. We also plotted the MSE as a function of both λ_1 , and λ_2 in Figure 2. The marked point is the point corresponding to the minimum value of MSE. As we can see from this figure, the optimal pair (λ_1, λ_2) that minimizes the MSE is non-zero for both λ_1 , and λ_2 .

- For ϵ^2 , which is the variance of the estimation error, as ϵ^2 increases we get worse and worse MSE values as shown in Figure 4 below. This is clear, since poor estimation of the measurement matrix would result in a poor signal recovery.
- For δ , which is defined as $\delta = \frac{m}{n}$, we plotted (in Figure 5) the MSE for a different values of δ while fixing all other parameters.

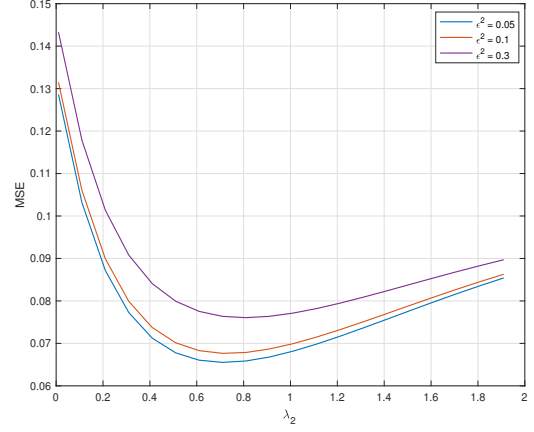


Fig. 3. MSE of the Box-EN for different values of ϵ^2 . We used $\lambda_1 = 0.1$, $\kappa = 0.1$, $\delta = 0.8$, $n = 500$, $\text{SNR} = 0.5$.

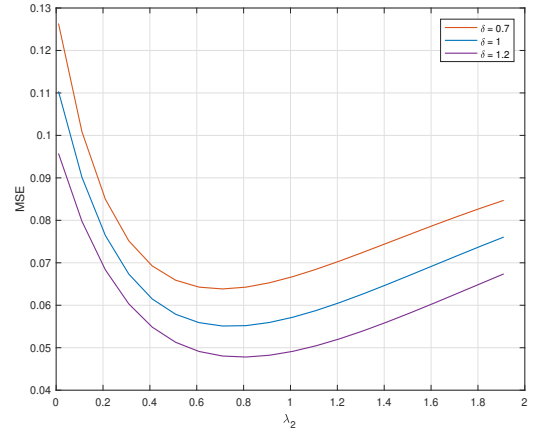


Fig. 4. MSE of the Box-EN for different values of δ . We used $\lambda_1 = 0.1$, $\kappa = 0.1$, $\epsilon^2 = 0.1$, $n = 500$, and $\text{SNR} = 0.5$.

The MSE is better as δ increases. For $\delta < 1$, we have an underdetermined system which makes the signal recovery more challenging than the case of an overdetermined system which corresponds to $\delta > 1$.

The prediction of Theorem 2 for the probability of support recovery is compared with the corresponding values obtained via numerical simulations and is plotted in Figure 5. This figure again shows the high accuracy of the proposed theoretical predictions.

VI. CONCLUSIONS

In this paper, we have used the recently developed Convex Gaussian Min-max Theorem (CGMT) framework to derive precise asymptotic characterizations of the mean squared error (MSE) and probability of support recovery of the Box-Elastic Net (Box-EN) under the presence of uncertainties in the measurement matrix. Numerical simulations show the close agreement to the proposed theoretical predictions. Also, we showed that the Box-EN outperforms standard Elastic-Net. These predictions can be used to optimally tune the involved parameters of the algorithm. Finally, we note that following the same steps as in this paper, our results can be used to characterize the performance of the standard EN.

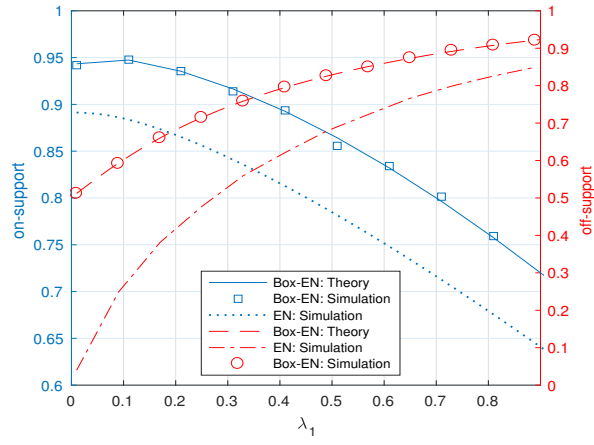


Fig. 5. Probability of successful support recovery of the Box-EN and the EN. This figure again shows that the former outperforms the latter. Theoretical prediction is based on Theorem 2 with $\lambda_2 = 0.1$, and $\xi = 10^{-3}$. For the simulations, we used $\kappa = 0.1$, $\epsilon^2 = 0.1$, $\delta = 0.7$, $n = 500$, $\text{SNR} = 0.5$.

REFERENCES

- [1] Hui Zou and Trevor Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
- [2] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [3] Trevor Hastie, Robert Tibshirani, and Martin Wainwright, *Statistical learning with sparsity: the lasso and generalizations*, CRC press, 2015.
- [4] Li Wang, Ji Zhu, and Hui Zou, "The doubly regularized support vector machine," *Statistica Sinica*, vol. 16, no. 2, pp. 589, 2006.
- [5] Weiwei Shen, Jun Wang, and Shiqian Ma, "Doubly regularized portfolio with risk minimization.," in *AAAI*, 2014, pp. 1286–1292.
- [6] Andrea Montanari and Phan-Minh Nguyen, "Universality of the elastic net error," in *Information Theory (ISIT), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 2338–2342.
- [7] Ali Khalajmehrabadi, Nikolaos Gatsis, Daniel J Pack, and David Akopian, "A joint indoor wlan localization and outlier detection scheme using lasso and elastic-net optimization techniques," *IEEE Transactions on Mobile Computing*, vol. 16, no. 8, pp. 2079–2092, 2017.
- [8] Chong You, Chun-Guang Li, Daniel P Robinson, and René Vidal, "Oracle based active set algorithm for scalable elastic net subspace clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3928–3937.
- [9] Samet Oymak, Amin Jalali, Maryam Fazel, Yonina C Eldar, and Babak Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices.," .
- [10] Jeyadeepan Jeganathan, Ali Ghayeb, Leszek Szczecinski, and Andres Ceron, "Space shift keying modulation for mimo channels," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3692–3703, 2009.
- [11] Michael Ting, Raviv Raich, and Alfred O Hero III, "Sparse image reconstruction for molecular imaging," *IEEE Transactions on Image Processing*, vol. 18, no. 6, pp. 1215–1227, 2009.
- [12] Emanuele Rodola, Andrea Torsello, Tatsuya Harada, Yasuo Kuniyoshi, and Daniel Cremers, "Elastic net constraints for shape matching," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1169–1176.
- [13] Yan Li, Bhanukiran Vinzamuri, and Chandan K Reddy, "Constrained elastic net based knowledge transfer for healthcare information exchange," *Data Mining and Knowledge Discovery*, vol. 29, no. 4, pp. 1094–1112, 2015.
- [14] Bo Liu, Liping Jing, Jian Yu, and Jia Li, "Robust graph learning via constrained elastic-net regularization," *Neurocomputing*, vol. 171, pp. 299–312, 2016.
- [15] David L Donoho, Arian Maleki, and Andrea Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.
- [16] Mohsen Bayati and Andrea Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, 2011.
- [17] Mohsen Bayati and Andrea Montanari, "The lasso risk for gaussian matrices," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 1997–2017, 2012.
- [18] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi, "Precise error analysis of regularized m-estimators in high-dimensions," *IEEE Transactions on Information Theory*, 2018.
- [19] 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.
- [20] Christos Thrampoulidis, Samet Oymak, and Babak Hassibi, "Regularized linear regression: A precise analysis of the estimation error," in *COLT*, 2015, pp. 1683–1709.
- [21] Ismail Ben Atitallah, Christos Thrampoulidis, Abla Kammoun, Tareq Y Al-Naffouri, Mohamed-Slim Alouini, and Babak Hassibi, "The box-lasso with application to gsk modulation in massive mimo systems," in *Information Theory (ISIT), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 1082–1086.
- [22] Mathieu Rosenbaum, Alexandre B Tsybakov, et al., "Sparse recovery under matrix uncertainty," *The Annals of Statistics*, vol. 38, no. 5, pp. 2620–2651, 2010.
- [23] Mohamed Ridha Zenaidi, Zouheir Rezki, Hamidou Tembine, and Mohamed-Slim Alouini, "Performance limits of energy harvesting communications under imperfect channel state information," in *Communications (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–6.
- [24] Ayed M Alrashdi, Ismail Ben Atitallah, Tareq Y Al-Naffouri, and Mohamed-Slim Alouini, "Precise performance analysis of the lasso under matrix uncertainties," in *Signal and Information Processing (GlobalSIP), 2017 IEEE Global Conference on*. IEEE, 2017, pp. 1290–1294.
- [25] Ayed M Alrashdi, Ismail Ben Atitallah, Tarig Ballal, Christos Thrampoulidis, 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.
- [26] Whitney K Newey and Daniel McFadden, "Large sample estimation and hypothesis testing," *Handbook of econometrics*, vol. 4, pp. 2111–2245, 1994.
- [27] Ehsan Abbasi, Christos Thrampoulidis, and Babak Hassibi, "General performance metrics for the lasso," in *Information Theory Workshop (ITW), 2016 IEEE*. IEEE, 2016, pp. 181–185.