PRECISE PERFORMANCE ANALYSIS OF THE LASSO UNDER MATRIX UNCERTAINTIES

Ayed M. Alrashdi[1,2], *Ismail Ben Atitallah*[1], *Tareq Y. Al-Naffouri*[1], *and Mohamed-Slim Alouini*[1]

[1] Computer, Electrical, Mathematical Sciences and Engineering (CEMSE) Division
King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[2]Electrical Engineering Department, University of Hail, Hail, Saudi Arabia

ABSTRACT

In this paper, we consider the problem of recovering an unknown sparse signal $\mathbf{x}_0 \in \mathbb{R}^n$ from noisy linear measurements $\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^m$. A popular approach is to solve the ℓ_1-norm regularized least squares problem which is known as the LASSO. In many practical situations, the measurement matrix $\mathbf{H}$ is not perfectely known and we only have a noisy version of it. We assume that the entries of the measurement matrix $\mathbf{H}$ and of the noise vector $\mathbf{z}$ are iid Gaussian with zero mean and variances $1/n$ and $\sigma_{\mathbf{z}}^2$. In this work, an imperfect measurement matrix is considered under which we precisely characterize the limiting behavior of the mean squared error and the probability of support recovery of the LASSO. The analysis is performed when the problem dimensions grow simultaneously to infinity at fixed rates. Numerical simulations validate the theoretical predictions derived in this paper.

Index Terms— LASSO, mean squared error, CGMT, measurement matrix uncertainties, probability of support recovery

1. INTRODUCTION

The Least Absolute Shrinkage and Selection Operator (LASSO) [1] is a powerfull method to recover a *k-sparse* unknown signal $\mathbf{x}_0 \in \mathbb{R}^n$ from noisy linear measurements: $\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \mathbf{z} \in \mathbb{R}^m$, where $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the measurement matrix, and $\mathbf{z} \in \mathbb{R}^m$ is the noise vector. In this paper, we assume that $\mathbf{H}$ is not perfectly known, and we only have a noisy version of it that is denoted by $\mathbf{A}$. Then, the LASSO solves the following convex optimization proplem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{A}\mathbf{x}||^2 + \lambda||\mathbf{x}||_1, \tag{1}$$

where $||\cdot||$ and $||\cdot||_1$ denote the ℓ_2-norm and the ℓ_1-norm respectively, and $\lambda \geq 0$ is the regularization parameter that balances between the deviation of $\mathbf{A}\hat{\mathbf{x}}$ from the observations $\mathbf{y}$ on one side, and the sparsity of the solution as promoted by the ℓ_1-norm on the other side. Problems of the form of (1) have many different diverse applications in science and engineering such as image processing [2], machine learning [3], wireless communications [4], etc.. The LASSO has been studied from different prespectives over the years. In recent years, the asymptotic exact characterization of the estimation performance gained a lot of interest. General performance metrics have been introduced such as the mean squared error and the probability support recovery. The first well-known bounds on the estimation performance of the lasso were order-wise in nature [5, 6, 7, 8]. The Approximate Message Passing (AMP) framework has been used in [9, 10, 11] to derive precise asymptotic analysis of the LASSO performance under the assumptions of iid Gaussian sensing matrix $\mathbf{A}$. A recently developed framework, that is based on the Convex Gaussian Min-max Theorem (CGMT) [12], has been used in a series of works to precisely evaluate the estimation performance of non-smooth regularized convex estimators under noisy iid Gaussian measurements (including the LASSO) [12] - [16].

However, these results assume that the measurement matrix $\mathbf{A}$ is perfectly known. In many practical applications it is reasonable to expect uncertainty in the linear measurement matrix $\mathbf{A}$ due to, e.g., imperfections in the signal acquisition hardware, model mismatch, estimation errors [17]. 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 and $\mathbf{\Omega}$ 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) [18].

In this work, we derive precise asymptotic predictions of the *mean squared error* and the *support recovery* of the LASSO under the presence of uncertainties in the measurement matrix that has iid Gaussian entries (both $\mathbf{H}$ and $\mathbf{\Omega}$ have iid Gaussian entries). The Gaussianity assumption of the entries of $\mathbf{A}$ is met in a wide range of applications such as MIMO application for Rayleigh fading model. The analysis is based on the CGMT framework and is performed when the problem dimensions m, n and k all grow simultaneously to infinity at fixed rates. Although our analysis is asymptotic in nature, numerical simulations show that our theoretical predictions are valid even for a few dozens of the problem dimensions.

2. PROBLEM SETUP

2.1. Performance Metrics

Finding a good estimate is an application dependent, since different applications require different desired properties of $\hat{\mathbf{x}}$. This results in a need for a variety of different performance metrics. Here we discuss some of them.

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 MSE $:= \frac{1}{n}||\hat{\mathbf{x}} - \mathbf{x}_0||^2$.

Support Recovery: In the problem of sparse recovery, a natural measure of performance that is used in many applications (e.g. parameter selection in regression, sparse approximation, structure estimation in graphical models [19]) is the 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 decison is based on the LASSO solution $\hat{\mathbf{x}}$: we say the i^{th} entry of $\hat{\mathbf{x}}$ is on the support if $|\hat{\mathbf{x}}_i| \geq \xi$, where $\xi > 0$ is a user-defined hard threshold on the entries on $\hat{\mathbf{x}}$. In Theorem 2, we precisely predict the *per-entry* rate of successful on-support and off-support recovery. Formaly, let

$$\Phi_{\xi,\text{on}}(\hat{\mathbf{x}}) = \frac{1}{k}\sum_{i\in S(\mathbf{x}_0)} \mathbb{1}_{\{|\hat{\mathbf{x}}_i|\geq\xi\}} \tag{2a}$$

$$\Phi_{\xi,\text{off}}(\hat{\mathbf{x}}) = \frac{1}{n-k}\sum_{i\notin S(\mathbf{x}_0)} \mathbb{1}_{\{|\hat{\mathbf{x}}_i|\leq\xi\}}, \tag{2b}$$

where $\mathbb{1}_{\{\mathcal{B}\}}$ is the indicator function of a set $\mathcal{B}$, and $S(\mathbf{x}_0)$ is the support of $\mathbf{x}_0$, i.e. the set of the non-zero entries of $\mathbf{x}_0$.

2.2. Working Assumptions

The unkown signal $\mathbf{x}_0 \in \mathbb{R}^n$ is a k-sparse signal, 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 entires are zeros. For the measurement matrix $\mathbf{A}$, we consider the following additive uncertainty model: $\mathbf{A} = \gamma\mathbf{H} + \epsilon\mathbf{\Omega}$, where $\mathbf{H}, \mathbf{\Omega} \in \mathbb{R}^{m\times n}$ both have entries iid $\mathcal{N}(0, 1/n)$, 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 $\mathcal{N}(0, \sigma_{\mathbf{z}}^2)$. The analysis is performed when the system dimensions (m, n and k) grow simultaneously large at fixed ratios: $\frac{m}{n} \longrightarrow \delta \in (0,\infty)$, and $\frac{k}{n} \longrightarrow \kappa \in (0, 1)$. Under these settings, the Signal to Noise Ratio (SNR) becomes SNR $:=\kappa/\sigma_{\mathbf{z}}^2$.

2.3. Notation

Throughout this paper, we use boldface letters to represent vectors and matrices. 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/mass function p_X. In particular, $H \sim \mathcal{N}(\mu, \sigma^2)$ implies that H has Gaussian distribution of mean μ and variance σ^2. $\phi(x)$ and $Q(x)$ denote the pdf of a standard normal distribution and its associated Q-function respectively. For $a, \lambda \in \mathbb{R}$, such that $\lambda > 0$, we define the following functions:

The soft-thresholding operator: $\eta(a; \lambda) = \arg\min_x \frac{1}{2}(x - a)^2 + \lambda|x|$, which can be written:

$$\eta(a;\lambda) = \begin{cases} a-\lambda & \text{, if } a > \lambda \\ 0 & \text{, if } |a| \leq \lambda \\ a+\lambda & \text{, if } a < -\lambda. \end{cases} \tag{3}$$

and its optimal value $e(a; \lambda) = \min_x \frac{1}{2}(x - a)^2 + \lambda|x|$

$$e(a;\lambda) = \begin{cases} \lambda a - \frac{1}{2}\lambda^2 & \text{, if } a > \lambda \\ \frac{1}{2}a^2 & \text{, if } |a| \leq \lambda \\ -\lambda a - \frac{1}{2}\lambda^2 & \text{, if } a < -\lambda. \end{cases} \tag{4}$$

Finaly, we write "$\xrightarrow{P}$" to designate convergence in probability.

3. 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 LASSO.

Theorem 1 (LASSO MSE) *Fix $\lambda > 0$, and let $\hat{\mathbf{x}}$ be a minimizer of the LASSO problem in (1), where $\mathbf{A}, \mathbf{z}$ and $\mathbf{x}_0$ satisfy the working assumptions of Section 2.2. Then it holds in probability:*

$$\lim_{n\to\infty}\frac{1}{n}||\hat{\mathbf{x}} - \mathbf{x}_0||^2 = \delta\tau_*^2 - \sigma_{\mathbf{z}}^2 + 2(\gamma-1)\mathbb{E}_{\substack{X_0\sim p_{X_0}\\ H\sim\mathcal{N}(0,1)}}\left[\eta\left(\gamma X_0 + \tau_* H; \frac{2\lambda\tau_*}{\beta_*}\right)X_0\right], \tag{5}$$

where (τ_*, β_*) is the unique solution to the following:

$$\min_{\tau>0}\max_{\beta>0} D(\tau,\beta) := \frac{\beta\tau}{2}(\delta-1) + \frac{\beta\sigma_{\mathbf{z}}^2}{2\tau} - \frac{\beta^2}{4} + \frac{\beta\epsilon^2\kappa}{2\tau} + \frac{\beta}{\tau}\cdot\mathbb{E}_{X_0,H}\left[e\left(\gamma X_0 + \tau H; \frac{2\lambda\tau}{\beta}\right)\right]. \tag{6}$$

τ_* and β_* can be efficiently computed by writing the first order optimality conditions, i.e. $\nabla_{(\tau,\beta)}D(\tau, \beta)$. The proof of Theorem 1 is based on the CGMT framework and is deferred to Section 5.

The following Theorem precisely characterizes the support recovery metrics introduced in (2).

Theorem 2 (Probability of support recovery) *Under the same settings of Theorem 1 and for any fixed $\xi > 0$, it holds in probability that:*

$$\lim_{n\to\infty}\Phi_{\xi,on}(\hat{\mathbf{x}}) = \mathbb{P}[\left|\eta(\gamma X_0 + \tau_* H; \frac{2\lambda\tau_*}{\beta_*})\right| \geq \xi],$$

and

$$\lim_{n\to\infty}\Phi_{\xi,\mathit{off}}(\hat{\mathbf{x}}) = \mathbb{P}[\left|\eta(\tau_* H; \frac{2\lambda\tau_*}{\beta_*})\right| \leq \xi] = 1 - 2Q\left(\frac{\xi}{\tau_*} + \frac{2\lambda}{\beta_*}\right).$$

The proof of Theorem 2 is also based on the CGMT and largely follows the proof of Theorem 1 and is omitted for space limitations.

4. NUMERICAL RESULTS

For illustration, we focus only on the case where $\mathbf{x}_0$, has enties that are sampled from sparse Bernoulli distribution. i.e. most of the entries of $\mathbf{x}_0$ are zeros and few are equal to 1. The mean squared error of the LASSO is predicted by Theorem 1, and the particular term $\mathbb{E}[e(\gamma X_0 + \tau H; \chi)]$, for $\tau > 0$ in (6) can be expressed as:

$$\kappa\int e(\gamma + \tau h; \chi)\phi(h)dh + (1-\kappa)\int e(\tau h; \chi)\phi(h)dh.$$

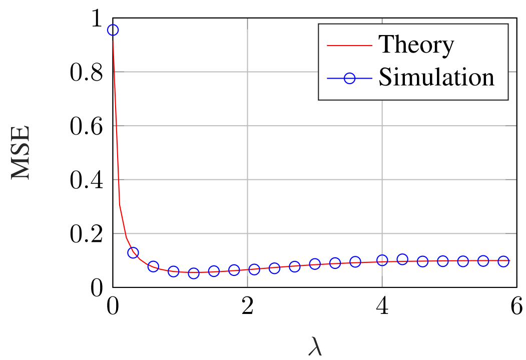

Fig. 1. The MSE performance of the LASSO. Theoretical prediction from Theorem 1. For simulations $\kappa = 0.1, \epsilon^2 = 0.1, \delta = 0.8, n = 256$, SNR = 0.5, and the data are averaged over 50 independent realizations of problem.

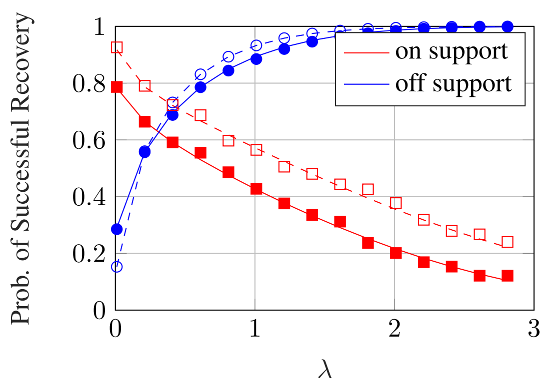

Fig. 2. Probability of scucessful on-support and off-support entries for two problem setup. The theoretical prediction (Solid and dashed lines) comes from Theorem 2. For the simulations (Squares and Circles), we used $n = 256$, SNR= 0.5, $\xi = 10^{-3}$, $\kappa = 0.1, \epsilon^2 = 0.2$, and the data are averaged over 50 independent realizations of problem. For solid lines and squares and circles, we used $\delta = 0.8$, while for dashed lines and empty squares and circles $\delta = 1.2$.

Figure 1 shows the accuracy of the mean squared error of the LASSO as predicted by Theorem 1.

Remark (Optimal Tuning): from Figure 1, we can see that there is a value of regularizer λ for which the MSE is minimized.

The prediction of theorem 2 for the support recovery compaired with the numerical simulations is shown in Figure 2. For the on-support recovery, the term $\mathbb{P}[|\eta(\gamma X_0 + \tau_* H; \frac{2\lambda\tau_*}{\beta_*})| \geq \xi] = Q(\frac{\xi+\gamma}{\tau_*} + \frac{2\lambda}{\beta_*}) + Q(\frac{\xi-\gamma}{\tau_*} + \frac{2\lambda}{\beta_*})$ for the sparse Bernoulli case. Both figures show the high accuracy of our predictions.

5. PROOF OUTLINE

In this section, we provide a proof outline of Theorem 1. For clarity, the steps of the proof are in divided into different subsections.

5.1. 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 just recall the statement of the theorem, and we refer the reader to [12] for the complete technical requirements. Consider the following two min-max problems, which we refer to as the Primary Optimization (PO) and the Auxiliary Optimization (AO) problems:

$$\Phi(\mathbf{G}) := \min_{\mathbf{w}\in\mathcal{S}_w} \max_{\mathbf{u}\in\mathcal{S}_u} \mathbf{u}^T\mathbf{G}\mathbf{w} + \psi(\mathbf{w},\mathbf{u}), \tag{7a}$$

$$\phi(\mathbf{g},\mathbf{h}) := \min_{\mathbf{w}\in\mathcal{S}_w} \max_{\mathbf{u}\in\mathcal{S}_u} ||\mathbf{w}||\mathbf{g}^T\mathbf{u} - ||\mathbf{u}||\mathbf{h}^T\mathbf{w} + \psi(\mathbf{w},\mathbf{u}), \tag{7b}$$

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 of (7a) and (7b) respectively. Let $\mathcal{S}_w, \mathcal{S}_u$ be convex, $\psi(\mathbf{w},\mathbf{u})$ be 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 $\lim_{n\to\infty} \mathbb{P}[\mathbf{w}_\phi \in \mathcal{S}] = 1$, it also holds $\lim_{n\to\infty} \mathbb{P}[\mathbf{w}_\Phi \in \mathcal{S}] = 1$.

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

For convenience, we consider the vector $\mathbf{w} := \gamma\mathbf{x} - \mathbf{x}_0$, then the problem in (1) can be reformulated in terms of $\mathbf{w}$ as:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} ||\mathbf{H}\mathbf{w} + \frac{\epsilon}{\gamma}\mathbf{\Omega}(\mathbf{w}+\mathbf{x}_0) - \mathbf{z}||^2 + \frac{2\lambda}{\gamma}||\mathbf{w}+\mathbf{x}_0||_1. \tag{8}$$

The problem in (8) is still not a form of a (PO) of the CGMT, so first we need to write it in form that suits the CGMT. To do so, we first express the loss function of (8) in its dual form through the Fenchel conjugate, $||\mathbf{H}\mathbf{w} + \frac{\epsilon}{\gamma}\mathbf{\Omega}(\mathbf{w}+\mathbf{x}_0) - \mathbf{z}||^2$ $= \max_{\mathbf{u}} \sqrt{n}\mathbf{u}^T(\mathbf{H}\mathbf{w} + \frac{\epsilon}{\gamma}\mathbf{\Omega}(\mathbf{w}+\mathbf{x}_0) - \mathbf{z}) - \frac{n}{4}||\mathbf{u}||^2$. The dual variable $\mathbf{u}$ is scaled by a factor $\sqrt{n}$ to have a proper normalization that guarantees the convergence afterwards. Hence, the problem in (8) is equivalent to the following:

$$\min_{\mathbf{w}} \max_{\mathbf{u}} \sqrt{n}\mathbf{u}^T\mathbf{H}\mathbf{w} + \frac{\sqrt{n}\epsilon}{\gamma}\mathbf{u}^T\mathbf{\Omega}(\mathbf{w}+\mathbf{x}_0) - \sqrt{n}\mathbf{u}^T\mathbf{z} - \frac{n}{4}||\mathbf{u}||^2 + \frac{2\lambda}{\gamma}||\mathbf{w}+\mathbf{x}_0||_1. \tag{9}$$

The above problem is in the form of a (PO) of the CGMT. Therefore, we can define its corresponding (AO) as:

$$\min_{\mathbf{w}} \max_{\mathbf{u}} ||\mathbf{w}||\mathbf{g}^T\mathbf{u} - ||\mathbf{u}||\mathbf{h}^T\mathbf{w} + \frac{\sqrt{n}\epsilon}{\gamma}\mathbf{u}^T\mathbf{\Omega}(\mathbf{w}+\mathbf{x}_0) - \frac{n}{4}||\mathbf{u}||^2 - \sqrt{n}\mathbf{u}^T\mathbf{z} + \frac{2\lambda}{\gamma}||\mathbf{w}+\mathbf{x}_0||_1. \tag{10}$$

5.3. Simplifying the (AO)

The next step is to show that the (AO1) as it appears in (10) can be transformed to a Scalar Optimization (SO) problem. Since the vectors $\mathbf{g}$ and $\mathbf{h}$ are independent, $||\mathbf{w}||\mathbf{g}^T\mathbf{u} - \sqrt{n}\mathbf{u}^T\mathbf{z} \overset{d}{=} \sqrt{||\mathbf{w}||^2 + n\sigma_{\mathbf{z}}^2}\mathbf{g}^T\mathbf{u}$. Therefore, (10) is equivalent to

$$\min_{\mathbf{w}} \max_{\mathbf{u}} \sqrt{||\mathbf{w}||^2 + n\sigma_{\mathbf{z}}^2}\mathbf{g}^T\mathbf{u} - ||\mathbf{u}||\mathbf{h}^T\mathbf{w} - \frac{n}{4}||\mathbf{u}||^2 + \frac{\sqrt{n}\epsilon}{\gamma}\mathbf{u}^T\mathbf{\Omega}(\mathbf{w}+\mathbf{x}_0) + \frac{2\lambda}{\gamma}||\mathbf{w}+\mathbf{x}_0||_1. \tag{11}$$

Now, it is more convenient to work with $\mathbf{x}$ instead of $\mathbf{w}$,

$$\min_{\mathbf{x}} \max_{\mathbf{u}} \sqrt{n}\epsilon \mathbf{u}^T \mathbf{\Omega}\mathbf{x} + \sqrt{||\gamma\mathbf{x} - \mathbf{x}_0||^2 + n\sigma_{\mathbf{z}}^2}\mathbf{g}^T\mathbf{u} - ||\mathbf{u}||\mathbf{h}^T(\gamma\mathbf{x} - \mathbf{x}_0) - \frac{n}{4}||\mathbf{u}||^2 + 2\lambda||\mathbf{x}||_1. \quad (12)$$

The optimization problem in (12) can be seen as another primary optimization problem (PO2). Hence, we can define another auxiliary optimization problem (AO2) that corresponds to the new (PO2). First, let $\mathbf{r} \in \mathbb{R}^m$ and $\mathbf{s} \in \mathbb{R}^n$ be standard Gaussian vectors, then the (AO2) can be defined as:

$$\min_{\mathbf{x}} \max_{\mathbf{u}} \epsilon||\mathbf{x}||\mathbf{r}^T\mathbf{u} - \epsilon||\mathbf{u}||\mathbf{s}^T\mathbf{x} + \sqrt{||\gamma\mathbf{x} - \mathbf{x}_0||^2 + n\sigma_{\mathbf{z}}^2}\mathbf{g}^T\mathbf{u} - ||\mathbf{u}||\mathbf{h}^T(\gamma\mathbf{x} - \mathbf{x}_0) - \frac{n}{4}||\mathbf{u}||^2 + 2\lambda||\mathbf{x}||_1. \quad (13)$$

Since $\mathbf{r}$ and $\mathbf{g}$ are independent standard Gaussian vectors, with abuse of notation, we have the following:

$$\epsilon||\mathbf{x}||\mathbf{r}^T\mathbf{u} + \sqrt{||\gamma\mathbf{x} - \mathbf{x}_0||^2 + n\sigma_{\mathbf{z}}^2}\mathbf{g}^T\mathbf{u}$$

$$\stackrel{d}{=} \sqrt{||\mathbf{x}||^2 + ||\mathbf{x}_0||^2 - 2\gamma\mathbf{x}_0^T\mathbf{x} + n\sigma_{\mathbf{z}}^2}\mathbf{g}^T\mathbf{u}.$$

Therefore, the (AO2) becomes:

$$\min_{\mathbf{x}} \max_{\mathbf{u}} \sqrt{||\mathbf{x}||^2 + ||\mathbf{x}_0||^2 - 2\gamma\mathbf{x}_0^T\mathbf{x} + n\sigma_{\mathbf{z}}^2}\mathbf{g}^T\mathbf{u} - \frac{n}{4}||\mathbf{u}||^2 - ||\mathbf{u}||(\epsilon\mathbf{s} + \gamma\mathbf{h})^T\mathbf{x} + ||\mathbf{u}||\mathbf{h}^T\mathbf{x}_0 + 2\lambda||\mathbf{x}||_1. \quad (14)$$

Fixing the norm of $\mathbf{u}$ to $\beta := ||\mathbf{u}||$, we can easily optimize over its direction by aligning it with $\mathbf{g}$. Then the (AO2) simplifies to:

$$\max_{\beta\geq 0} \min_{\mathbf{x}} \sqrt{n}\beta\sqrt{\frac{1}{n}(||\mathbf{x}||^2 + ||\mathbf{x}_0||^2 - 2\gamma\mathbf{x}_0^T\mathbf{x}) + \sigma_{\mathbf{z}}^2}||\mathbf{g}|| - \beta(\epsilon\mathbf{s} + \gamma\mathbf{h})^T\mathbf{x} + \beta\mathbf{h}^T\mathbf{x}_0 - \frac{n\beta^2}{4} + 2\lambda||\mathbf{x}||_1. \quad (15)$$

To have a separable optimization problem, we use the following identity: $\sqrt{\chi} = \min_{\alpha>0} \frac{\alpha}{2} + \frac{\chi}{2\alpha}$, where $\chi = \frac{1}{n}(||\mathbf{x}||^2 + ||\mathbf{x}_0||^2 - 2\gamma\mathbf{x}_0^T\mathbf{x}) + \sigma_{\mathbf{z}}^2$. Also, define $\tau := \frac{\sqrt{n}\alpha}{||\mathbf{g}||}$, and $\tilde{\mathbf{h}} := \epsilon\mathbf{s} + \gamma\mathbf{h}$. This yields the following optimization problem:

$$\min_{\tau>0} \max_{\beta>0} \frac{\beta\tau||\mathbf{g}||^2}{2} + \frac{n\beta\sigma_{\mathbf{z}}^2}{2\tau} - \frac{n\beta^2}{4} + \frac{\beta}{\gamma}(\tilde{\mathbf{h}} - \epsilon\mathbf{s})^T\mathbf{x}_0 + \frac{\beta}{\tau}\left(\sum_{i=1}^n \frac{\epsilon^2}{2}\mathbf{x}_{0,i}^2 - \gamma\tilde{\mathbf{h}}_i\mathbf{x}_{0,i} - \frac{\tau^2}{2}\tilde{\mathbf{h}}_i^2\right) + \frac{\beta}{\tau}\left(\sum_{i=1}^n \min_{\mathbf{x}_i} \frac{1}{2}(\mathbf{x}_i - \gamma\mathbf{x}_{0,i} - \tau\tilde{\mathbf{h}}_i)^2 + \frac{2\lambda\tau}{\beta}|\mathbf{x}_i|\right). \quad (16)$$

The optimization over $\mathbf{x}_i$ can be solved in a closed-form expression using the soft-thresholding operator, which is exactly the function defined in (3). Then, the above optimization problem simplifies to the following Scalar Optimization (SO) problem:

$$\min_{\tau>0} \max_{\beta>0} \tilde{D}(\tau, \beta, \mathbf{g}, \mathbf{h}) := \frac{\beta\tau||\mathbf{g}||^2}{2} + \frac{n\beta\sigma_{\mathbf{z}}^2}{2\tau} - \frac{n\beta^2}{4} + \frac{\beta}{\tau}\left(\sum_{i=1}^n \frac{\epsilon^2}{2}\mathbf{x}_{0,i}^2 - \gamma\mathbf{h}_i\mathbf{x}_{0,i} - \frac{\tau^2}{2}\mathbf{h}_i^2\right) + \frac{\beta}{\gamma}\mathbf{h}^T\mathbf{x}_0 + \frac{\beta}{\tau}\sum_{i=1}^n e\left(\gamma\mathbf{x}_{0,i} + \tau\mathbf{h}_i; \frac{2\lambda\tau}{\beta}\right). \quad (17)$$

5.4. Probabilistic asymptotic analysis of the (SO) problem

After simplifiying the (AO2) as in (17), we are now in a position to analyze its limiting behavior. First, we need to properly normalize the objective function in (17) by dividing it by n. Then, using the WLLN, we have: $\frac{1}{n}||\mathbf{g}||^2 \xrightarrow{P} \delta$, $\frac{1}{n}||\mathbf{h}||^2 \xrightarrow{P} 1$, $\frac{1}{n}||\mathbf{x}_0||^2 \xrightarrow{P} \kappa$ 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 $\beta > 0$, $\frac{1}{n}\sum_{i=1}^n e(\gamma\mathbf{x}_{0,i} + \tau\mathbf{h}_i; \frac{2\lambda\tau}{\beta}) \xrightarrow{P} \mathbb{E}[e(\gamma X_0 + \tau H; \frac{2\lambda\tau}{\beta})]$, and $\frac{1}{n}\sum_{i=1}^n \tilde{\mathbf{x}}_i \xrightarrow{P} \mathbb{E}[\eta(\gamma X_0 + \tau H; \frac{2\lambda\tau}{\beta})]$, where $\tilde{\mathbf{x}}$ is the solution of (AO2) defined in (14). Therefore, the point-wise convergence in τ and β of the objective function in (17) is the quantity $D(\tau, \beta)$ defined in Theorem 1. 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 [20] that $\tau_n(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \tau_*$.

5.5. Applying the CGMT

We prove that the quantities $\hat{\mathbf{x}} - \mathbf{x}_0$ and $\tilde{\mathbf{x}} - \mathbf{x}_0$ are concentarted in the same set. Formally, for any fixed $\zeta > 0$, we define the set: $\mathcal{S} = \left\{\mathbf{v} : \left|\frac{1}{n}||\mathbf{v}||^2 - M(\tau_*, \beta_*)\right| < \zeta\right\}$, where $M(\tau_*, \beta_*) = \delta\tau_*^2 - \sigma_{\mathbf{z}}^2 + 2(\gamma-1)\mathbb{E}[\eta(\gamma X_0 + \tau_* H; \frac{2\lambda\tau_*}{\beta_*})X_0]$, and τ_* and β_* are as defined in Theorem 1. Let $\check{\mathbf{x}}$ be the solution of (AO1) defined in (12). The error can be written as: $||\tilde{\mathbf{x}} - \mathbf{x}_0||^2 = ||\tilde{\mathbf{w}}||^2 + 2(\gamma - 1)\tilde{\mathbf{x}}^T\mathbf{x}_0$. Recall that $\tau_n(\mathbf{g}, \mathbf{h}) = \frac{\sqrt{n}}{||\mathbf{g}||}\sqrt{\frac{1}{n}||\tilde{\mathbf{w}}||^2 + \sigma_{\mathbf{z}}^2}$. Using $\tau_n(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \tau_*$, we find $\frac{||\tilde{\mathbf{w}}||^2}{n} \xrightarrow{P} \delta\tau_*^2 - \sigma_{\mathbf{z}}^2$. Also, it can be shown that $\frac{1}{n}\tilde{\mathbf{x}}^T\mathbf{x}_0 \xrightarrow{P} \mathbb{E}[\eta(\gamma X_0 + \tau_* H; \frac{2\lambda\tau_*}{\beta_*})X_0]$. Putting all the results together, it can be shown that $\frac{1}{n}||\tilde{\mathbf{x}} - \mathbf{x}_0||^2 \xrightarrow{P} M(\tau_*, \beta_*)$. This proves that for any ζ, $\tilde{\mathbf{x}} - \mathbf{x}_0 \in \mathcal{S}$ with probability one. Then, we conclude using the CGMT that $\check{\mathbf{x}} - \mathbf{x}_0 \in \mathcal{S}$ with probability one. A second application of the CGMT is needed to conclude that $\hat{\mathbf{x}} - \mathbf{x}_0 \in \mathcal{S}$ with probability one and is omitted for space considerations. This completes the proof of Theorem 1.

6. CONCLUSION

In this paper, we proposed a precise asymptotic analysis of the MSE and the probability of support recovery of the LASSO under imperfect Gaussian measurement matrix assumptions. Although our analysis is asymptotic in nature, numerical simulations show that our theoretical predictions are valid even for a few dozens of the problem dimensions.

7. REFERENCES

[1] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[2] 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.

[3] Christopher M Bishop, "Pattern recognition," *Machine Learning*, vol. 128, pp. 1–58, 2006.

[4] Guan Gui, Wei Peng, and Ling Wang, "Improved sparse channel estimation for cooperative communication systems," *International Journal of Antennas and Propagation*, vol. 2012, 2012.

[5] Emmanuel J Candes, Justin K Romberg, and Terence Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[6] Emmanuel Candes and Terence Tao, "The dantzig selector: Statistical estimation when p is much larger than n," *The Annals of Statistics*, pp. 2313–2351, 2007.

[7] Peter J Bickel, Ya'acov Ritov, and Alexandre B Tsybakov, "Simultaneous analysis of lasso and dantzig selector," *The Annals of Statistics*, pp. 1705–1732, 2009.

[8] Sahand Negahban, Bin Yu, Martin J Wainwright, and Pradeep K Ravikumar, "A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers," in *Advances in Neural Information Processing Systems*, 2009, pp. 1348–1356.

[9] 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.

[10] 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.

[11] Mohsen Bayati and Andrea Montanari, "The lasso risk for gaussian matrices," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 1997–2017, 2012.

[12] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi, "Precise error analysis of regularized m-estimators in high-dimensions," *arXiv preprint arXiv:1601.06233*, 2016.

[13] Christos Thrampoulidis, Samet Oymak, and Babak Hassibi, "Regularized linear regression: A precise analysis of the estimation error.," in *COLT*, 2015, pp. 1683–1709.

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

[15] Christos Thrampoulidis, Ashkan Panahi, Daniel Guo, and Babak Hassibi, "Precise error analysis of the lasso," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3467–3471.

[16] 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.

[17] Mathieu Rosenbaum, Alexandre B Tsybakov, et al., "Sparse recovery under matrix uncertainty," *The Annals of Statistics*, vol. 38, no. 5, pp. 2620–2651, 2010.

[18] 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.

[19] Martin J Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using -constrained quadratic programming (lasso)," *IEEE transactions on information theory*, vol. 55, no. 5, pp. 2183–2202, 2009.

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

[21] Mihailo Stojnic, "Recovery thresholds for 1 optimization in binary compressed sensing," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*. IEEE, 2010, pp. 1593–1597.

[22] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi, "Lasso with non-linear measurements is equivalent to one with linear measurements," in *Advances in Neural Information Processing Systems*, 2015, pp. 3420–3428.