



Research article

Square-root lasso under correlated regressors: Tight statistical analysis with a wireless communications application

Ayed M. Alrashdi^{1,*} and Masad A. Alrasheedi²

¹ Department of Electrical Engineering, College of Engineering, University of Ha'il, Ha'il 81441, Saudi Arabia

² Department of Management Information Systems, College of Business Administration, Taibah University, Al-Madinah Al-Munawara, 42353, Saudi Arabia

* **Correspondence:** Email: am.alrashdi@uoh.edu.sa.

Abstract: This paper provided a comprehensive analysis of sparse signal estimation from noisy and possibly underdetermined linear observations in the high-dimensional asymptotic regime. The focus was on the square-root lasso (sqrt-lasso), a popular convex optimization method used for sparse signal recovery. We analyzed its performance using several metrics, such as root-mean-squared error (r.m.s.e.), mean absolute error (m.a.e.), cosine similarity, and true/false recovery rates. The analysis assumed a normally distributed design matrix with left-sided correlation and Gaussian noise. In addition to theoretical contributions, we applied these results to a real-world wireless communications problem by examining the error performance of sqrt-lasso in generalized space shift keying (GSSK) modulation for multiple-input multiple-output (MIMO) systems. This application was particularly relevant, as the GSSK modulation generates sparse data symbols, making it an ideal scenario for sparse recovery techniques. Our study offered tight asymptotic approximations for the performance of sqrt-lasso in such systems. Beyond the wireless communications application, the results had broader implications for other high-dimensional applications, including compressed sensing, machine learning, and statistical inference. The analysis presented in this paper, supported by numerical simulations, provided practical insights into how sqrt-lasso behaved under correlated designs, offering useful guidelines for optimizing its use in real-world scenarios. The expressions and insights obtained from this study can be used to optimally choose the penalization parameter of the sqrt-lasso. By applying these results, one can make informed decisions about performance and fine-tuning the sqrt-lasso, considering the presence of correlated regressors in a high-dimensional context.

Keywords: square-root lasso; error performance; sparse signal estimation; high-dimensional statistics; MIMO; wireless communications

Mathematics Subject Classification: 60G35, 62E20, 62J05, 62J07

1. Introduction

1.1. Motivation

In the era of big data, where the volume and complexity of data can overwhelm traditional analytical methods, the concepts of compressed sensing and advanced regularization techniques have become increasingly important. Compressed sensing [1, 2] is a signal processing technique that leverages sparsity to reconstruct signals from a limited number of measurements, offering a powerful framework for data acquisition and analysis.

The problem of sparse signal recovery is vital in many modern applications because it allows for efficient processing and recovery of high-dimensional data where only a small number of signal components are significant [2]. This is especially important in applications like wireless communications [3], medical imaging [4], machine learning [5], and beyond, where data is often incomplete or noisy, but the underlying signals are sparse. Accurate sparse signal estimation enables the reconstruction of signals from limited measurements, reducing the need for excessive data collection, which can be costly or impractical. Identifying sparse structures helps in improving model interpretability and reducing computational complexity.

This imbalance necessitates the use of regularization techniques to prevent overfitting and enhance model performance. Among these techniques, the least absolute shrinkage and selection operator (lasso) [6] stands out for its dual role in variable selection and regularization. It is a widely used regularization technique in machine learning, data science, statistics, and beyond [5, 7]. A variant known as the square-root lasso (sqrt-lasso) [8] offers additional benefits for handling large-dimensional data with unknown noise variance, providing robust and efficient solutions for sparse signal estimation.

In this work, we consider the linear measurements model

$$y_j = \boldsymbol{\theta}_0^\top \mathbf{a}_j + \sigma w_j, \quad j = 1, 2, \dots, M, \quad (1.1)$$

where $y_j \in \mathbb{R}$ is the outcome (response) value, $\mathbf{a}_j \in \mathbb{R}^N$ is the regressor vector, and $w_j \in \mathbb{R}$ is independent and identically distributed (IID) Gaussian noise ($j = 1, 2, \dots, M$), with $\sigma > 0$ being its standard deviation. The vector of unknown true parameters, $\boldsymbol{\theta}_0 \in \mathbb{R}^N$, is assumed to be S -sparse, that is, it has at most S nonzero elements, where $S < M$. The sqrt-lasso [8] solves the following convex program:

$$\widehat{\boldsymbol{\theta}} := \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^N} \sqrt{\frac{1}{M} \sum_{j=1}^M (y_j - \mathbf{a}_j^\top \boldsymbol{\theta})^2} + \frac{\gamma}{M} \sum_{k=1}^N |\theta_k|, \quad (1.2)$$

where $\gamma > 0$ is the regularization coefficient.

The importance of the work presented in this paper lies in the fact that lasso-type problems, including sqrt lasso, are essential to many modern applications, including machine learning, signal processing, and wireless communications. Thus, a deeper understanding of their error performance is crucial for optimizing their use. Additionally, this work provides a theoretical framework for selecting the penalization parameter of sqrt-lasso based on derived performance metrics, as discussed in Section 5, rather than relying on traditional cross-validation methods. Cross-validation is both time-consuming and requires substantial data, which can be impractical in real-world applications.

By offering a more efficient and theoretically grounded method for parameter selection, this work significantly improves the practicality of using sqrt-lasso in high-dimensional settings. Furthermore, as discussed in Section 6, this study's findings allow for the design of practical communication systems, such as multiple-input multiple-output (MIMO) systems, in an optimal way, minimizing the probability of error when decoding data symbols. This makes the work particularly valuable in improving system reliability and performance in noisy and high-dimensional environments.

1.2. Literature survey on the applications of sqrt-lasso in engineering and beyond

The sqrt-lasso and related lasso-type problems have found applications in numerous fields due to its robustness and efficiency in large-dimensional data analysis. In medical imaging, such as magnetic resonance imaging (MRI) and computed tomography (CT) scans, a constrained version of the sqrt-lasso helps in reconstructing high-quality images from fewer measurements, reducing scan times and radiation exposure [4]. In astronomy, the sqrt-lasso was employed for reconstructing images from incomplete and noisy data, facilitating the study of celestial objects [9]. Moreover, in wireless communications, a box-constrained version of the lasso was used to enhance the signal recovery in compressed sensing-based modulation methods such as spatial modulation [3, 10]. Furthermore, the lasso was used in [11] to analyze genetic data for disease predictions by selecting the relevant genetic markers associated with certain diseases. In addition, [12] leverages the lasso method for its feature selection capabilities to improve the performance of the random subspace method in bearing fault diagnosis. By selecting a sparse and relevant set of features, lasso helps in enhancing the diagnostic accuracy and robustness of the model. The authors in [13] explored its role in sparse principal component analysis (PCA), to produce modified principal components with sparse loadings. A lasso-based approach was used in [14] for solar power generation forecasting, while [15, 16] used it for the analysis of large scale electrical power systems. Additionally, lasso techniques have been used in control engineering such as in system identification [17–19].

1.3. Related works on large-dimensional error performance characterization

Characterizing the error performance of the sqrt-lasso is crucial for understanding its theoretical and practical implications. In fact, a large body of research has been conducted on the analysis of lasso-type optimization problems under different assumptions and using various technical analysis tools, including approximate message passing (AMP) [2], the replica method [20], and the convex Gaussian minimax theorem (CGMT) [21], to name a few. First attempts to characterize the performance of the lasso and related compressed sensing problems provided some loose error bounds such as in [22–25]. In addition, the AMP framework was used in [26, 27] to derive large-dimensional error analysis of the standard lasso under IID Gaussian feature vectors. Furthermore, the replica method, a heuristic from statistical physics, was used in [28–30] to analyze related problems in compressed sensing.

The CGMT-based error analysis was motivated by the early work of Stojnic, who analyzed a constrained version of the sqrt-lasso in the noiseless case [31], and later considered the noisy case in the high signal-to-noise ratio (SNR) case [32] (i.e., when $\sigma \rightarrow 0$), all under the assumption of IID Gaussian design matrix. Shortly after, the authors in [33–36] extended the results of Stojnic to derive tight error bounds on the performance of the sqrt-lasso with general regularization functions in the high SNR regime. Then, the authors in [37, 38] extended the abovementioned findings to the

finite SNR case and obtained precise characterization of the squared-error of the generalized sqrt-lasso with IID Gaussian linear measurements. Additionally, [39] considered general performance measures of the sqrt-lasso beyond the squared-error, while [40] analyzed the sqrt-lasso using CGMT with non-linear measurements. Moreover, the box-lasso, a variant of the widely used lasso, was analyzed in [3] under the same IID Gaussian design matrix. All of the above references assumed the design matrix to be perfectly known. The CGMT-based analysis was extended to study the more difficult case of imperfect design matrices (i.e., when the design matrix has some errors) in several works including the lasso/box-lasso [10, 41, 42], box-elastic net [43], etc.

The literature that we provided above considered the design matrix to have IID Gaussian elements. However, it is important to study the effect of correlations in the design matrix on the large-dimensional error performance of the lasso-type problems discussed here. Unlike the case of IID elements, the correlated-design scenario poses many technical challenges and few references have studied it. In particular, [44] provided some loose bounds on how correlations affect the prediction of the lasso. Moreover, [45, 46] partially studied the performance of the standard lasso with correlated designs. Additionally, the CGMT framework was used to study the effect of correlations on the performance of non-sparse recovery algorithms such as the box-relaxation optimization (BRO) [47, 48] and regularized least square [49].

To the best of our knowledge, the asymptotic error performance of the sqrt-lasso in the presence of correlations in the design matrix has not been studied before in the large-dimensional setup. While previous studies have largely focused on IID designs, real-world systems, particularly in wireless communications, often involve correlated data. This motivates the study presented in this work which takes into account the presence of a left-sided correlation in the design matrix.

1.4. Summary of contributions

As discussed above, this paper derives a tight asymptotic analysis of the performance of the sqrt-lasso convex program in (1.2), with the assumption of a correlated Gaussian design matrix when used for the recovery of a sparse signal from noisy linear observations. In particular, we consider various performance measures of the sqrt-lasso such as the root-mean squared error (r.m.s.e.), mean absolute error (m.a.e.), cosine similarity, and false/true recovery rates and derive their asymptotic approximations. After that, we present an application of our results to study the large-dimensional limit of the probability of error performance of the sqrt-lasso in a modern modulation scheme called the generalized space-shift keying (GSSK) modulation [50], which is used to encode sparse data symbols in a MIMO wireless communication system. Numerical simulations are provided to support the theoretical results obtained in this work.

1.5. Paper organization

The paper is organized as follows: Section 2 introduces the problem setup and modeling assumptions. Section 3 presents the main theoretical results of the paper on the large-dimensional error performance characterization of sqrt-lasso. In Section 4, we discuss the sparse recovery rates of the sqrt-lasso. Moreover, in Section 5, we implement penalization parameter selection methods based on the derived performance measure. Section 6 discusses the application of our results in a wireless communications context, specifically in GSSK modulation for MIMO systems. Finally, Section 7

concludes the paper and suggests future research directions. All the proofs of the results of the paper are deferred to the appendices.

A flowchart that describes the general organizational structure of the paper is given in Figure 1.

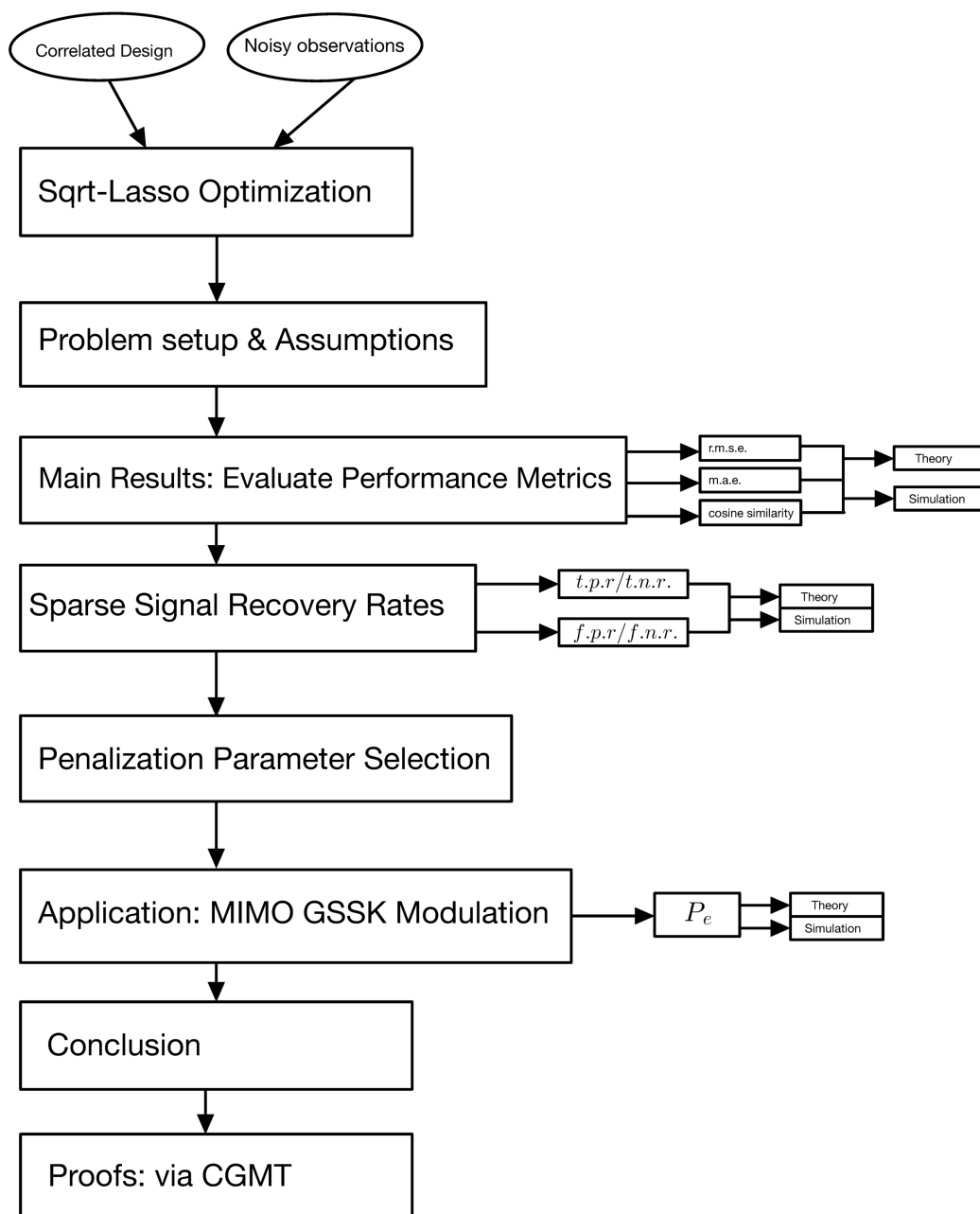


Figure 1. General organizational structure of the paper.

2. Problem setup

2.1. Notations

For the sake of simplicity, we will use the following notations from here on. We use boldface letters for vectors and matrices. We also use $\|\cdot\|_2$ and $\|\cdot\|_1$ to represent the Euclidean norm and the ℓ_1 -norm of a vector, respectively. In addition, $(\cdot)^\top$ is used to indicate the transpose operation. The notations $\mathbb{E}[\cdot]$ and $\mathbb{P}(\cdot)$ denote the expectation and probability, respectively. The notation $V_N \xrightarrow[N \rightarrow \infty]{P} c$ is used when a sequence of random variables V_N converges in probability to a constant c , as $N \rightarrow \infty$. The notation $V \sim p_V$ is used to indicate that the random variable V has a probability density function (PDF) p_V . In particular, $V \sim \mathcal{N}(0, 1)$ means that V is a standard Gaussian random variable. Furthermore, for $x \in \mathbb{R}$, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2} dz$ is the cumulative distribution function (CDF) of a standard Gaussian random variable. Finally, for any $x \in \mathbb{R}$ and $\lambda > 0$, $\eta(x; \lambda)$ is the wavelet shrinkage function defined as

$$\eta(x; \lambda) = [x - \lambda \operatorname{sign}(x)] \mathbb{1}_{\{|x| \geq \lambda\}}, \quad (2.1)$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function, which is defined as $\mathbb{1}_{\{\mathcal{S}\}} = 1$ if the statement \mathcal{S} holds true, otherwise $\mathbb{1}_{\{\mathcal{S}\}} = 0$.

2.2. Modeling assumptions

We consider the possibly underdetermined linear observations system given in (1.1), which can be compactly written as

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta}_0 + \sigma\mathbf{w}, \quad (2.2)$$

where $\mathbf{y} = [y_1, y_2, \dots, y_M]^\top \in \mathbb{R}^M$, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M]^\top \in \mathbb{R}^{M \times N}$, and $\mathbf{w} = [w_1, w_2, \dots, w_M]^\top \in \mathbb{R}^M$. Moreover, the vector of unknown parameters $\boldsymbol{\theta}_0 \in \mathbb{R}^N$ is assumed to be S -sparse, i.e., only S of its entries are sampled IID from a PDF $p_{\boldsymbol{\theta}_0}$, which has a zero mean and a finite second moment ($\mathbb{E}[\boldsymbol{\theta}_0^2] = \sigma_{\boldsymbol{\theta}_0}^2 < \infty$), while the remaining entries are zeros. The noise vector $\mathbf{w} \in \mathbb{R}^M$ is assumed to have IID entries $\mathcal{N}(0, 1)$, and $0 < \sigma < \infty$ is its standard deviation.

In addition, in this work, we consider a left-sided correlated Gaussian design matrix which can be modeled as

$$\mathbf{A} = \frac{1}{\sqrt{N}} \boldsymbol{\Psi}^{\frac{1}{2}} \mathbf{G}, \quad (2.3)$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{M \times M}$ is a known positive semi-definite left-sided correlation matrix, while $\mathbf{G} \in \mathbb{R}^{M \times N}$ is a Gaussian matrix with IID entries $\mathcal{N}(0, 1)$. This correlation-model is widely used in the wireless communications literature, and also known as the receive-side correlation which represents the spatial correlations between the receiver antennas [51]. Although this model is not particularly popular in statistics*, it is better suited for the GSSK system application discussed later in Section 6.

*The more popular model in statistics and data science is the right-sided correlation:

$$\mathbf{A} = \frac{1}{\sqrt{N}} \mathbf{G} \bar{\boldsymbol{\Psi}}^{\frac{1}{2}},$$

where $\bar{\boldsymbol{\Psi}}$ is the covariance matrix of the data. Our analysis applies to this model as well. However, we focus on the left-sided correlation for concreteness.

The analysis of this study is carried out in the asymptotic linear regime which requires dimensions of the problem (N , M , and S) to approach infinity at relative rates: $\frac{M}{N} \rightarrow \zeta \in (0, \infty)$, and $\frac{S}{N} \rightarrow \Delta \in (0, 1)$. Furthermore, the regularization coefficient, $\gamma > 0$, is assumed to be constant and independent of N .

2.3. Performance metrics

The large-dimensional error characterization of the sqrt-lasso considered in this work is based on deriving asymptotic limits of the following performance metrics.

- **r.m.s.e.:** is a widely used performance metric for evaluating the accuracy of a model's predictions. It measures the average magnitude of the errors between predicted and observed values by squaring the individual errors, averaging them, and then taking the square root of this average. This metric provides a clear and intuitive indication of how far, on average, the predictions are from the actual values, in the same units as the target variable. The r.m.s.e. is often preferred when large errors are particularly undesirable, as it penalizes these more heavily than smaller errors. Formally, it can be defined as follows:

$$\text{r.m.s.e.} \triangleq \sqrt{\frac{1}{N} \sum_{k=1}^n (\widehat{\theta}_k - \theta_{0,k})^2}, \quad (2.4)$$

where $\widehat{\theta}_k$ is the k^{th} element of the sqrt-lasso estimate in (1.2). The squaring process emphasizes larger errors, making the r.m.s.e. particularly sensitive to outliers and significant deviations.

- **m.a.e.:** is another common performance metric used to assess the accuracy of a model's predictions. Unlike the r.m.s.e., the m.a.e. measures the average magnitude of errors by taking the absolute values of the individual errors and averaging them. This results in a straightforward interpretation of the average error in the same units as the target variable. The m.a.e. treats all errors equally, providing a linear score that is less sensitive to outliers than r.m.s.e. This characteristic makes the m.a.e. a preferred choice when the impact of all errors, regardless of their magnitude, needs to be assessed uniformly. Formally, it can be defined as:

$$\text{m.a.e.} \triangleq \frac{1}{N} \sum_{k=1}^N |\widehat{\theta}_k - \theta_{0,k}|. \quad (2.5)$$

- **Cosine similarity:** often referred to as the cosine correlation, is a performance measure used to determine the similarity (correlation) between two nonzero vectors. In contrast to the aforementioned two metrics, this measure assesses how closely the two vectors are aligned, regardless of their magnitudes. This metric focuses on the orientation of the vectors, making it particularly useful in high-dimensional spaces where the direction of the vectors is more important than their magnitude. It is commonly used in text mining, information retrieval, and recommendation systems to compare the similarity of documents, user preferences, or other types of high-dimensional data [52]. Cosine similarity values range from -1 to 1 , with 1 indicating perfect alignment, 0 indicating no correlation, and -1 indicating perfect opposition. This measure does not account for the magnitude of the differences between the vectors, focusing solely on their directional relationship. For our sqrt-lasso algorithm, the cosine similarity between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ is

defined as:

$$\cos(\angle(\boldsymbol{\theta}_0; \widehat{\boldsymbol{\theta}})) \triangleq \frac{\boldsymbol{\theta}_0^\top \widehat{\boldsymbol{\theta}}}{\|\boldsymbol{\theta}_0\|_2 \|\widehat{\boldsymbol{\theta}}\|_2}. \quad (2.6)$$

3. Main results

In this section, we present the primary theoretical results of this work, specifically the tight analysis of the asymptotic performance of the sqrt-lasso in (1.2). We begin by examining the estimation performance using the r.m.s.e. metric, as detailed in Theorem 3.1 below, which accurately describes the large-dimensional asymptotic behavior of the error. Following this, we derive asymptotic approximations of the other considered performance measures, including the m.a.e. and cosine similarity. Numerical simulations are then provided to support the obtained theoretical results.

3.1. Large-dimensional characterization of sqrt-lasso

In this subsection, we state the main theoretical results of this work. Toward this goal, we define the eigenvalue decomposition of the correlation matrix $\boldsymbol{\Psi}$ as $\boldsymbol{\Psi} = \mathbf{P}\mathbf{Q}\mathbf{P}^\top$, where \mathbf{P} is an orthonormal matrix of eigenvectors, and \mathbf{Q} is a diagonal matrix of eigenvalues, such that $\mathbf{Q} = \text{diag}([q_1, q_2, \dots, q_M]^\top)$, where $q_1 \geq q_2 \geq \dots \geq q_M$ are the eigenvalues of $\boldsymbol{\Psi}$. All eigenvalues are required to be fixed and finite.

Theorem 3.1 (r.m.s.e. limiting behavior). Under the modeling assumptions of Section 2.2, the following convergence holds true for the r.m.s.e. of the sqrt-lasso:

$$\text{r.m.s.e.} \xrightarrow[N \rightarrow \infty]{P} \alpha_\star, \quad (3.1)$$

where α_\star is the unique minimizer to the following:

$$\min_{\substack{\alpha \geq 0 \\ t > 0}} \max_{\substack{\beta \geq 0 \\ \mu \geq 0 \\ \chi > 0}} C(\alpha, t, \beta, \chi, \mu), \quad (3.2)$$

and

$$\begin{aligned} C(\alpha, t, \beta, \chi, \mu) := & [(t - q_1^{-1})\beta^2 + 1]\mu - \frac{\alpha\chi}{2} + \frac{\chi\sigma_{\theta_0}^2}{2\alpha} + \frac{1}{4\mu N} \sum_{j=1}^M \frac{q_j\alpha^2 + \sigma^2}{1 + q_j(t - q_1^{-1})} \\ & - \frac{\chi}{2\alpha} \mathbb{E}_{\substack{\Theta_0 \sim p_{\Theta_0} \\ H \sim N(0,1)}} \left[\eta^2 \left(\Theta_0 - \frac{\alpha\beta}{\chi} H; \frac{\alpha\gamma}{\chi\sqrt{\zeta}} \right) \right], \end{aligned} \quad (3.3)$$

with q_j being the j^{th} eigenvalue of the matrix $\boldsymbol{\Psi}$, while q_1 is its maximum eigenvalue.

Proof. The asymptotic convergence of the cost function as in (3.2) is derived in Appendix A, and the proof of the r.m.s.e. convergence is given in Appendix B.1. \square

Remark 3.1. The optimal solutions $\alpha_\star, \beta_\star, \chi_\star$, among others, can be determined numerically by setting up the first-order optimality conditions, that is, by solving $\nabla_{(\alpha, t, \beta, \chi, \mu)} C(\alpha, t, \beta, \chi, \mu) = \mathbf{0}_5$. Although Theorem 3.1 primarily involves α_\star only, the roles of the other optimization variables, such as β_\star and χ_\star , will become evident in the prediction of the m.a.e. and cosine similarity, which are derived later.

Remark 3.2. For isotropic regressors, i.e., when $\Psi = \mathbf{I}$, then $q_j = 1 \quad \forall j = 1, 2, \dots, M$, and the scalar objective in (3.2) reduces to:

$$\min_{\substack{\alpha \geq 0 \\ \chi > 0}} \max_{\substack{0 \leq \beta \leq 1}} \beta \sqrt{\zeta} \sqrt{\alpha^2 + \sigma^2} - \frac{\alpha \chi}{2} + \frac{\chi \sigma_{\theta_0}^2}{2\alpha} - \frac{\chi}{2\alpha} \mathbb{E}_{\substack{\Theta_0 \sim p_{\Theta_0} \\ H \sim \mathcal{N}(0,1)}} \left[\eta^2 \left(\Theta_0 - \frac{\alpha \beta}{\chi} H; \frac{\alpha \gamma}{\chi \sqrt{\zeta}} \right) \right],$$

which is the well-known objective for the sqrt-lasso derived in [38], and [53]. In particular, see Eq (6.37) in [53], with the following map of variables therein: $\Theta_0 \mapsto X_0, \zeta \mapsto \delta, \chi \mapsto \tau_h, \frac{\gamma}{\sqrt{\zeta}} \mapsto \lambda$, and $\sigma_{\theta_0}^2 \rightarrow \sigma_x^2 = 1$.

Remark 3.3. Theorem 3.1 allows us to optimally adjust parameters of the sqrt-lasso problem, such as the regularization coefficient γ and the number of normalized measurements ζ , among others. Refer to Figure 2 for an illustration.

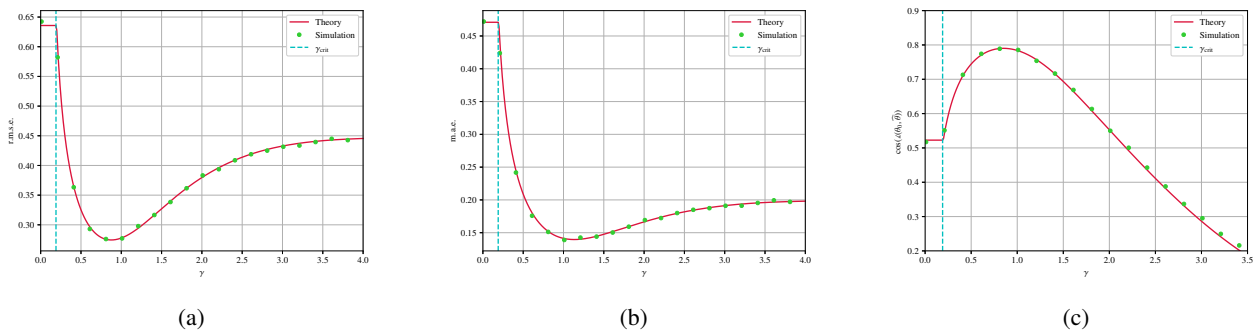


Figure 2. Performance of sqrt-lasso versus the regularization coefficient for $\zeta = 0.85$. (a) r.m.s.e.; (b) m.a.e.; (c) cosine similarity.

The following proposition presents a tight asymptotic evaluation of the m.a.e. of the sqrt-lasso.

Proposition 3.1 (Convergence of the m.a.e.). Under the same settings and assumptions of Theorem 3.1, the m.a.e. defined in (2.5) converges as

$$\text{m.a.e.} \xrightarrow[N \rightarrow \infty]{P} \mathbb{E}_{\Theta_0, H} \left[\left| \eta \left(\Theta_0 - \frac{\alpha_* \beta_*}{\chi_*} H; \frac{\gamma \alpha_*}{\chi_* \sqrt{\zeta}} \right) - \Theta_0 \right| \right], \quad (3.4)$$

where α_* , β_* , and χ_* are the optimal solutions of (3.2), and $\eta(\cdot; \cdot)$ is the wavelet shrinkage function defined in (2.1).

Proof. See Appendix B.2. □

In contrast to the aforementioned two metrics, we consider next the large-dimensional characterization of the cosine similarity metric which assesses how closely θ_0 and $\hat{\theta}$ are aligned, regardless of their magnitudes.

Proposition 3.2 (Cosine similarity convergence). *Under the setting of Theorem 3.1, and for the cosine similarity metric defined in (2.6), it holds that:*

$$\cos(\angle(\boldsymbol{\theta}_0; \widehat{\boldsymbol{\theta}})) \xrightarrow[N \rightarrow \infty]{P} \frac{\mathbb{E}_{\Theta_0, H} \left[\Theta_0 \eta \left(\Theta_0 - \frac{\alpha_* \beta_*}{\chi_*} H; \frac{\gamma \alpha_*}{\chi_* \sqrt{\zeta}} \right) \right]}{\sqrt{\sigma_{\theta_0}^2 \mathbb{E}_{\Theta_0, H} \left[\eta^2 \left(\Theta_0 - \frac{\alpha_* \beta_*}{\chi_*} H; \frac{\gamma \alpha_*}{\chi_* \sqrt{\zeta}} \right) \right]}}. \quad (3.5)$$

Proof. The proof closely follows the one provided in [42, App. A.2.5]; therefore, the details are omitted. \square

3.2. Numerical simulations

In this subsection, we conduct a numerical analysis of the sqrt-lasso performance across various settings. Specifically, we examine the three performance metrics introduced earlier: r.m.s.e., m.a.e., and cosine similarity. We then compare these numerical findings with the theoretical predictions discussed earlier in this section. As an example, we consider the so-called sparse binary distribution of the entries of $\boldsymbol{\theta}_0$:

$$p_{\Theta_0}(\theta) = \Delta \delta(\theta - E) + (1 - \Delta) \delta(\theta), \quad (3.6)$$

where $E > 0$, and $\delta(\cdot)$ is the Dirac delta function. This distribution is widely used in many applications, and of particular interest to us is its use in modeling GSSK signals, as we will discuss in Section 6.

Moreover, for the correlation matrix $\boldsymbol{\Psi}$, we consider the Toeplitz correlation matrix example [54] given by:

$$\Psi_{kj} = \rho^{|k-j|}, \quad (3.7)$$

for $\rho \in [0, 1)$. In all of our simulations, we used the CVXPY package [55] to numerically solve the sqrt-lasso optimization in (1.2).

3.2.1. Underdetermined (compressed measurements) case

First, we consider an underdetermined scenario of the problem, i.e., $M < N$ (equivalently $\zeta < 1$). Figure 2 depicts the behavior of the considered performance measures as functions of the regularization coefficient with $\zeta = 0.85$, $N = 256$, $\rho = 0.3$, $\sigma^2 = 0.05$, $E = 1$, and $\Delta = 0.2$. For each γ value, we used 100 independent Monte Carlo simulations. These figures illustrate the tight match of the theoretical expressions to numerical simulations. In addition, in all metrics, we can see that there is usually a value of the regularization coefficient γ that results in an optimal performance, e.g., minimum r.m.s.e./m.a.e., or maximum cosine similarity.

Moreover, from Figure 2, we can see that there exists a critical value of the regularizer[†] that we denote by γ_{crit} , for which the large-dimensional performance is constant when $\gamma \leq \gamma_{\text{crit}}$ and it is different for $\gamma > \gamma_{\text{crit}}$. This behavior is noticed only when $\zeta \leq 1$, otherwise $\gamma_{\text{crit}} = 0$ (see for example, [33, Section 8] and [40]).

[†]This behavior is only observed in the sqrt-lasso, while in the standard lasso $\gamma_{\text{crit}} = 0$ for both overdetermined and underdetermined systems.

3.2.2. Overdetermined case

We repeated the same simulations as in the above case, but we now consider the overdetermined case, i.e., when $M > N$. In Figure 3, we plotted the three performance metrics as functions of the regularization coefficient γ . We used $\zeta = 2, N = 256, \rho = 0.3, \sigma^2 = 0.05, E = 1$, and $\Delta = 0.2$. For each γ value, we performed 100 independent simulations. As expected, the overall performance of the overdetermined system is superior to an underdetermined one. Moreover, as discussed above, we can see from these figures that for $\zeta > 1$, indeed $\gamma_{\text{crit}} = 0$. Although our theoretical results are asymptotic, i.e., requiring $N \rightarrow \infty$, we can see from Figures 2 and 3 that there is a tight match between theory and simulations for finite values such as $N = 256$.

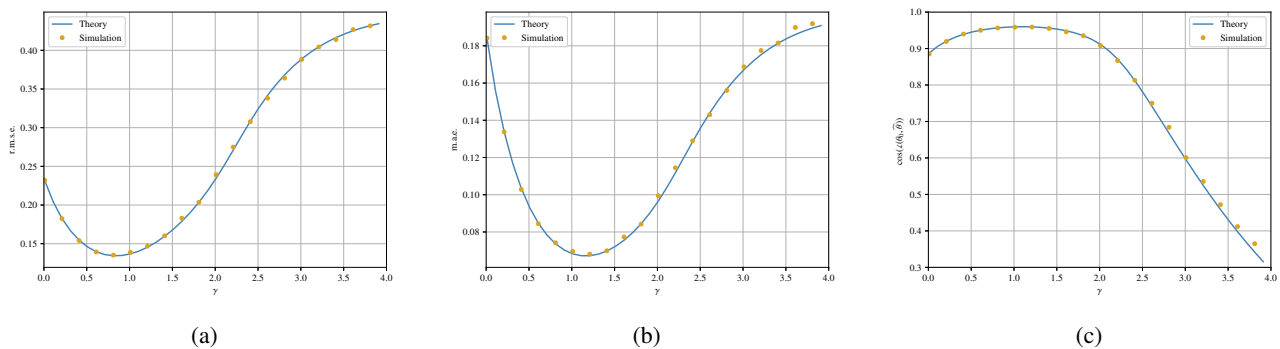


Figure 3. Performance of sqrt-lasso versus the regularization coefficient for $\zeta = 2$. (a) r.m.s.e.; (b) m.a.e.; (c) cosine similarity.

4. Sparse recovery rates

Sparse recovery rates are metrics used to evaluate the performance of algorithms designed to recover sparse signals from limited or noisy measurements such as the sqrt-lasso considered in this work. In the context of sparse signal estimation, sparse recovery amounts to determining how well an algorithm can identify the nonzero elements of a sparse vector (aka true elements) while minimizing the identification of zero elements as nonzero (false recovery). The key recovery rates include: true positive rate (t.p.r.), false positive rate (f.p.r.), true negative rate (t.n.r), and false negative rate (f.n.r.).

Mathematically, these rates are defined as follows (where $\hat{\theta}_i$ is the i^{th} element of the lasso estimate in (1.2) and $\theta_{0,i}$ is the i^{th} entry of the S -sparse vector):

$$\text{t.p.r.} = \frac{\sum_{k=1}^N \mathbb{1}_{\{\hat{\theta}_k \neq 0\}} \mathbb{1}_{\{\theta_{0,k} \neq 0\}}}{\sum_{k=1}^n \mathbb{1}_{\{\theta_{0,k} \neq 0\}}},$$

$$\text{f.p.r.} = \frac{\sum_{k=1}^N \mathbb{1}_{\{\hat{\theta}_k \neq 0\}} \mathbb{1}_{\{\theta_{0,k} = 0\}}}{\sum_{k=1}^N \mathbb{1}_{\{\theta_{0,k} = 0\}}},$$

$$\text{t.n.r.} = \frac{\sum_{k=1}^N \mathbb{1}_{\{\hat{\theta}_k = 0\}} \mathbb{1}_{\{\theta_{0,k} = 0\}}}{\sum_{k=1}^N \mathbb{1}_{\{\theta_{0,k} = 0\}}},$$

and

$$\text{f.n.r.} = \frac{\sum_{k=1}^N \mathbb{1}_{\{\hat{\theta}_k=0\}} \mathbb{1}_{\{\theta_{0,k} \neq 0\}}}{\sum_{k=1}^N \mathbb{1}_{\{\theta_{0,k} \neq 0\}}}.$$

Proposition 4.1 (Sparse recovery rates). *Under the settings of Theorem 3.1, the following convergences hold true:*

$$\begin{aligned} \text{t.p.r.} & \xrightarrow{P, N \rightarrow \infty} \mathbb{P}\left(\eta\left(\Theta_0 - \frac{\alpha_* \beta_*}{\chi_*} H; \frac{\gamma \alpha_*}{\chi_* \sqrt{\zeta}}\right) \neq 0\right), \\ \text{f.p.r.} & \xrightarrow{P, N \rightarrow \infty} \mathbb{P}\left(\eta\left(-\frac{\alpha_* \beta_*}{\chi_*} H; \frac{\gamma \alpha_*}{\chi_* \sqrt{\zeta}}\right) \neq 0\right) = 2\Phi\left(\frac{-\gamma}{\beta_* \sqrt{\zeta}}\right), \\ \text{t.n.r.} & \xrightarrow{P, N \rightarrow \infty} \mathbb{P}\left(\eta\left(-\frac{\alpha_* \beta_*}{\chi_*} H; \frac{\gamma \alpha_*}{\chi_* \sqrt{\zeta}}\right) = 0\right), \\ \text{f.n.r.} & \xrightarrow{P, N \rightarrow \infty} \mathbb{P}\left(\eta\left(\Theta_0 - \frac{\alpha_* \beta_*}{\chi_*} H; \frac{\gamma \alpha_*}{\chi_* \sqrt{\zeta}}\right) = 0\right), \end{aligned}$$

where α_* , β_* , and χ_* are the optimizers of the minimax optimization problem in (3.2), and the probabilities are with respect to $\Theta_0 \sim p_{\Theta_0}$ and $H \sim \mathcal{N}(0, 1)$.

Proof. The sparse recovery rate metrics defined above are not Lipschitz, so the CGMT framework cannot be directly used to prove Proposition 4.1. However, as shown in [27, 39], this can be proved via standard weak convergence arguments. The details are omitted to maintain a concise presentation. \square

Remark 4.1 (Special case: Sparse binary signals). *For the sparse binary distribution given in (3.6), the above sparse recovery rate expressions simplify to:*

$$\begin{aligned} \text{t.p.r.} & \xrightarrow{P, N \rightarrow \infty} \Phi\left(\frac{-\gamma}{\beta_* \sqrt{\zeta}} + \frac{\chi_*}{\alpha_* \beta_*} E\right) + \Phi\left(\frac{-\gamma}{\beta_* \sqrt{\zeta}} - \frac{\chi_*}{\alpha_* \beta_*} E\right), \\ \text{f.p.r.} & \xrightarrow{P, N \rightarrow \infty} 2\Phi\left(\frac{-\gamma}{\beta_* \sqrt{\zeta}}\right), \\ \text{t.n.r.} & \xrightarrow{P, N \rightarrow \infty} 1 - 2\Phi\left(\frac{-\gamma}{\beta_* \sqrt{\zeta}}\right), \end{aligned}$$

and

$$\text{f.n.r.} \xrightarrow{P, N \rightarrow \infty} \Phi\left(\frac{\gamma}{\beta_* \sqrt{\zeta}} - \frac{\chi_*}{\alpha_* \beta_*} E\right) - \Phi\left(\frac{-\gamma}{\beta_* \sqrt{\zeta}} - \frac{\chi_*}{\alpha_* \beta_*} E\right).$$

Note: It can be easily seen that t.n.r. = 1 - f.p.r. and f.n.r. = 1 - t.p.r.

Remark 4.2 (Numerical illustration). *For the sparse binary signal model in (3.6) with $E = 1$, we plotted in Figure 4 the positive recovery rates as functions of the regularization coefficient γ , with $\zeta = 0.85$, $N = 256$, $\rho = 0.3$, $\sigma^2 = 0.05$, and $\Delta = 0.2$.*

In addition, the negative recovery rates are illustrated in Figure 5 with the same problem parameters as in Figure 4. When compared to numerical simulations, both figures show the high accuracy of our theoretical expressions in predicting these metrics. Moreover, we can notice from these figures that as γ increases, the t.n.r. tends to 1, whereas the t.p.r. approaches zero. This behavior is not surprising since a large amount of regularization will result in sparser solutions.

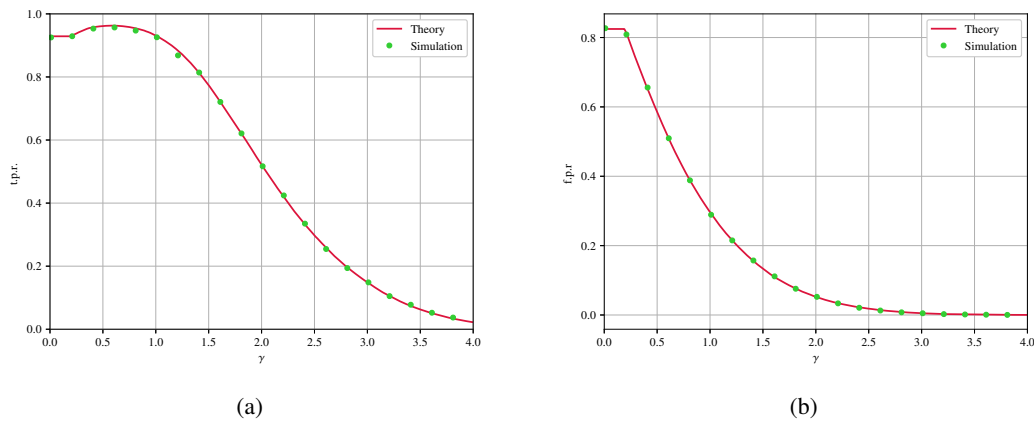


Figure 4. Positive recovery rates. (a) t.p.r.; (b) f.p.r.

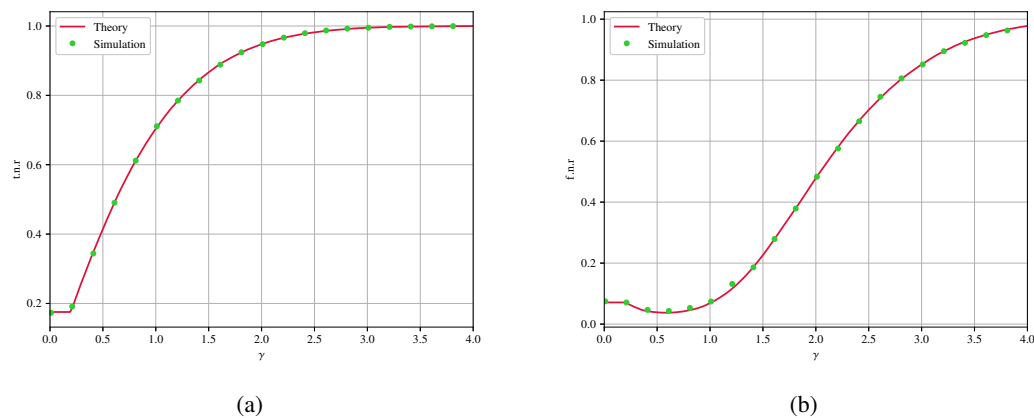


Figure 5. Negative recovery rates. (a) t.n.r.; (b) f.n.r.

5. Penalization parameter selection

The selection of the penalization parameter γ plays a critical role in the performance of sqrt-lasso as it directly influences the trade-off between model sparsity and the accuracy of signal recovery [7]. If too large, important features are missing; if too small, incorrect features are included. The appropriate choice of γ depends on the application at hand and the desired performance metrics, such as minimizing the r.m.s.e./m.a.e., reducing the f.p.r., or maximizing the cosine similarity.

Several methods have been proposed in the literature to choose γ optimally. Among these, the quantile universal threshold (QUT) method, as introduced by Bellec et al. in [56], provides a simple yet effective way to select γ for the sqrt-lasso. The QUT is designed for minimizing the false discovery rate (FDR) while maintaining good predictive performance, particularly when the noise level is unknown. This method sets the threshold such that the FDR is controlled at a desired level, making it especially useful in high-dimensional problems where false positives are a concern. The QUT threshold can be easily computed and incorporated into existing algorithms for sqrt-lasso, and it is particularly well-

suiting for sparse signal estimation tasks. In the context of regression problems where the number of observations M is equal to the number of predictors N , and the design matrix \mathbf{A} is orthonormal, the results of Mallat et al. [57] demonstrate another approach to parameter selection in the framework of sparse additive models. Their work, which explores the case of orthonormal design matrices and applies wavelets for dimensionality reduction, suggests parameter values that minimize the mean squared error (MSE). This is particularly relevant for problems with dense matrices or when the regression matrix exhibits orthogonality properties, which simplifies the selection of γ .

Given the theoretical approximations derived in this work, we implemented four different methods for selecting γ based on minimizing r.m.s.e., m.a.e., and f.p.r., or maximizing the cosine similarity. For the minimization of the r.m.s.e. metric, let us define the optimal penalization parameter, $\gamma_*^{\text{r.m.s.e.}}$ as

$$\gamma_*^{\text{r.m.s.e.}} := \arg \min_{\gamma > 0} \widetilde{\text{r.m.s.e.}}(\gamma), \quad (5.1)$$

where $\widetilde{\text{r.m.s.e.}}(\gamma)$ is the asymptotic limit of the r.m.s.e. as derived in (3.1). Moreover, we can similarly define $\gamma_*^{\text{m.a.e.}}$, $\gamma_*^{\text{f.p.r.}}$, and $\gamma_*^{\text{cos.}}$ as the values of the penalization parameter that optimize the m.a.e., f.p.r. and cosine similarity, respectively. These values can be obtained using any search methods such as the golden-section technique method and the ternary search method [58].

Tables 1 and 2 provide a comparison of different penalization parameter selection methods and their effects on performance metrics for underdetermined and overdetermined cases, respectively. We used a sparse-binary signal vector as defined in (3.6) to generate θ_0 .

Table 1. Comparison of different penalization parameter selection methods with performance metrics. We used $\zeta = 0.85, \rho = 0.3, \sigma^2 = 0.01, E = 1$, and $\Delta = 0.2$.

Method	γ_*	r.m.s.e.	m.a.e.	cos. sim.	f.p.r.
Min. r.m.s.e., $\gamma_*^{\text{r.m.s.e.}}$	0.81	0.1204	0.0712	0.9600	1.8×10^{-6}
Min. m.a.e., $\gamma_*^{\text{m.a.e.}}$	0.99	0.1356	0.0573	0.9605	3.15×10^{-6}
Max. cosine similarity, $\gamma_*^{\text{cos.}}$	0.92	0.1323	0.0684	0.9710	2.98×10^{-6}
Min. f.p.r., $\gamma_*^{\text{f.p.r.}}$	0.71	0.1334	0.0753	0.9569	1.5×10^{-6}
γ_*^{QUT}	0.74	0.1338	0.0772	0.9545	1.57×10^{-6}

Table 2. Comparison of different penalization parameter selection methods based on several performance metrics. We used $\zeta = 2, \rho = 0.3, \sigma^2 = 0.01, E = 1$, and $\Delta = 0.2$.

Method	γ_*	r.m.s.e.	m.a.e.	cos. sim.	f.p.r.
Min. r.m.s.e., $\gamma_*^{\text{r.m.s.e.}}$	0.84	0.0601	0.0756	0.9771	2.23×10^{-23}
Min. m.a.e., $\gamma_*^{\text{m.a.e.}}$	1.16	0.0639	0.0209	0.9837	4.04×10^{-26}
Max. cosine similarity, $\gamma_*^{\text{cos.}}$	1.29	0.0673	0.0319	0.9938	1.22×10^{-24}
Min. f.p.r., $\gamma_*^{\text{f.p.r.}}$	0.65	0.0617	0.0366	0.9851	4.14×10^{-29}
γ_*^{QUT}	0.69	0.0611	0.0357	0.9836	4.48×10^{-29}

In Table 1, the optimal value of the penalization parameter γ_* varies between 0.71 and 0.99, depending on the desired metric. Notably, the lowest r.m.s.e. of 0.1204 is achieved with $\gamma_*^{\text{r.m.s.e.}}$, while the lowest m.a.e. of 0.0573 is obtained with $\gamma_*^{\text{m.a.e.}}$. Interestingly, the method maximizing cosine

similarity achieves the highest value of 0.9710, though at the cost of a slightly higher r.m.s.e. and m.a.e. The f.p.r.-based method results in a moderately higher penalization, with $\gamma_* = 0.71$, showing balanced performance across the other metrics, particularly with a low f.p.r. of 1.5×10^{-6} . The QUT-based method has similar performance to the f.p.r.-based method as they minimize closely related metrics, namely FDR and f.p.r.

In Table 2, where $\zeta = 2$ indicates a setting with more measurements relative to parameters, we observe that the penalization parameters are generally higher, ranging from 0.65 to 1.29. The larger ζ typically allows for better estimation of the true signal. In particular, the reduced r.m.s.e. and m.a.e. values in Table 2 highlight how having more measurements leads to more accurate recovery of the true signal.

An important observation in Table 2 is the dramatic reduction in *false positive rates* across all methods. The $\gamma_*^{\text{f.p.r.}}$ method results in a minuscule f.p.r. of $4.14 \times 10^{-29} \approx 0$, indicating that the model is highly effective in suppressing irrelevant variables when the data provides more information. This is a clear advantage of having more measurements.

Furthermore, Figure 6 illustrates the variation of the optimal penalization parameter γ_* as a function of the correlation coefficient ρ . The two curves represent the values of γ_* that minimize the root mean squared error ($\gamma_*^{\text{r.m.s.e.}}$) and the false positive rate ($\gamma_*^{\text{f.p.r.}}$), respectively. As the correlation coefficient ρ increases, both penalization parameters exhibit a general downward trend. However, $\gamma_*^{\text{r.m.s.e.}}$ maintains higher values across the range of ρ , indicating a more conservative choice of γ when minimizing the error. On the other hand, $\gamma_*^{\text{f.p.r.}}$ decreases more rapidly, reflecting a tighter penalization required to control the false positive rate as the correlation increases. The problem parameters used for generating this plot are $\zeta = 0.85$, $\sigma^2 = 0.05$, $E = 1$, and $\Delta = 0.2$, for a sparse-binary signal vector.

As shown in the figure, both penalization parameters decrease with increasing ρ . This decrease can be explained by the fact that as correlation increases, the effective degrees of freedom in the model may reduce because highly correlated variables behave similarly. As a result, the model has fewer distinct variables to estimate, which in turn reduces the need for strong penalization, allowing the penalization parameter to be lower.

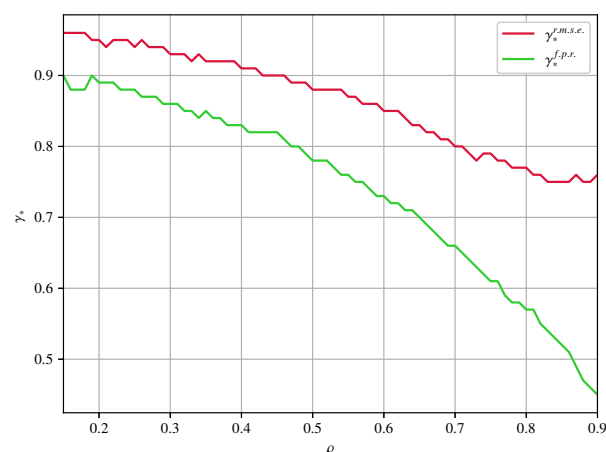


Figure 6. Optimal penalization parameter as a function of the correlation coefficient ρ . The problem parameters are $\zeta = 0.85$, $\sigma^2 = 0.05$, $E = 1$, and $\Delta = 0.2$.

6. Application: GSSK modulation in a MIMO wireless communications system

In this section, we consider applying the results derived so far in an application for a MIMO wireless communication system [59] that employs a modern spatial modulation scheme, such as the GSSK [50, 60], where information is conveyed solely through antenna location, with only a small subset of antennas active at any given time. In these systems, the transmitted constellation signal θ_0 is inherently sparse, promoting the use of sparse recovery algorithms such as the lasso and its variants to be used as decoding techniques as in [3, 10, 61]. Consequently, this section focuses on evaluating the large-dimensional performance of the sqrt-lasso as a detection method for GSSK-modulated MIMO systems.

In this MIMO setup, the signal at the receiving antenna satisfies $\mathbf{y} = \mathbf{A}\theta_0 + \sigma\mathbf{w}$, where now \mathbf{A} represents the channel matrix between the transmitter and receiver sides following the receive-side correlation model in (2.3), and \mathbf{w} is an additive white Gaussian noise vector with IID $\mathcal{N}(0, 1)$ components. Under these settings, the SNR is assumed to be constant and given as $\text{SNR} = \frac{\Delta E^2}{\sigma^2}$.

In GSSK modulation transmission, only a subset of antennas $\mathcal{K} \subset \{1, 2, \dots, N\}$ is active at any given time. Each active antenna transmits a symbol $\theta_{0,k} = E > 0, k \in \mathcal{K}$, while the other antennas stay idle, i.e., $\theta_{0,k} = 0, k \notin \mathcal{K}$. Consequently, the information is conveyed solely by the positions of the active antennas. Note that the transmitted signal θ_0 is well-modeled by the sparse binary distribution given in (3.6), where the cardinality of \mathcal{K} is assumed to be $|\mathcal{K}| = \Delta N$, for a normalized sparsity factor $\Delta \in (0, 1)$. This means that θ_0 is S -sparse on average.

After passing through the channel, we can decode the received signal by first obtaining $\widehat{\theta}$ as the solution of the sqrt-lasso in (1.2). Second, we obtain our final estimate as θ^* with elements:

$$\theta_k^* = \begin{cases} E, & \text{if } |\widehat{\theta}_k| \geq \xi, \\ 0, & \text{otherwise,} \end{cases} \quad (6.1)$$

where $\xi \in (0, E)$ is a fixed user-defined hard threshold.

To characterize the performance of the sqrt-lasso decoder, we will use the probability of error (P_e), defined below, as a performance metric,

$$P_e \triangleq \frac{1}{N} \sum_{k=1}^N \mathbb{P}(\theta_k^* \neq \theta_{0,k}). \quad (6.2)$$

In the next proposition, we provide the large-dimensional characterization of the probability of error for the considered GSSK-modulated MIMO system.

Proposition 6.1 (P_e convergence). *Under the setting of Theorem 3.1, the next convergence holds*

$$P_e \xrightarrow[N \rightarrow \infty]{P} \Phi\left(\frac{\gamma}{\beta_\star \sqrt{\zeta}} + \frac{(\xi - E)\chi_\star}{\alpha_\star \beta_\star}\right) - \Phi\left(\frac{-\gamma}{\beta_\star \sqrt{\zeta}} - \frac{(\xi + E)\chi_\star}{\alpha_\star \beta_\star}\right) + 2\Phi\left(\frac{-\gamma}{\beta_\star \sqrt{\zeta}} - \frac{\xi\chi_\star}{\alpha_\star \beta_\star}\right), \quad (6.3)$$

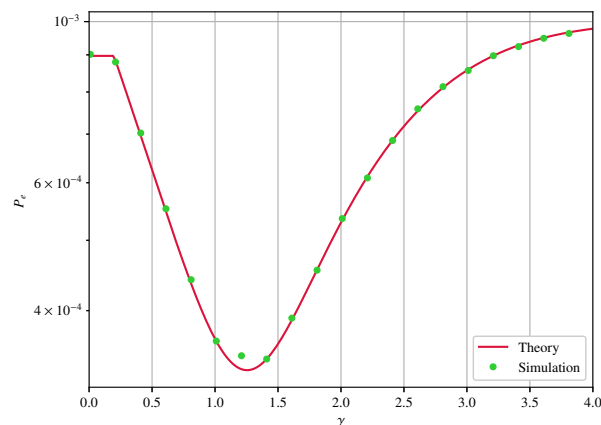
where $\alpha_\star, \beta_\star$, and χ_\star are solutions to the min-max optimization in (3.2), with p_{θ_0} as given by (3.6) therein.

Proof. It can be shown that the probability of error as defined in (6.2) converges as

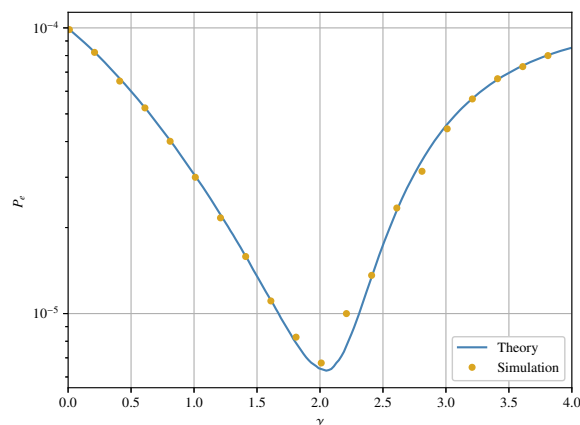
$$P_e \xrightarrow[N \rightarrow \infty]{P} \mathbb{P}\left(\left|\eta\left(E - \frac{\alpha_\star \beta_\star}{\chi_\star} H; \frac{\alpha_\star \gamma}{\chi_\star \sqrt{\zeta}}\right)\right| < \xi\right) + \mathbb{P}\left(\left|\eta\left(-\frac{\alpha_\star \beta_\star}{\chi_\star} H; \frac{\alpha_\star \gamma}{\chi_\star \sqrt{\zeta}}\right)\right| \geq \xi\right).$$

Using the definition of $\eta(\cdot; \cdot)$ in (2.1) and evaluating the above probabilities over the randomness of $H \sim \mathcal{N}(0, 1)$, the same convergence as in (6.3) can be achieved. \square

Figure 7 illustrates the P_e theoretical performance as predicted by (6.3) as a function of the regularization coefficient γ . This figure shows the tight agreement to the numerical simulations. Also, it indicates that there exists a value of γ for which the P_e is minimized. This allows the optimum tuning of the regularizer coefficient γ . We set $N = 256, \rho = 0.3, \sigma^2 = 0.01, E = 1, \Delta = 0.2$, and $\xi = 10^{-3}$, and we conducted 20 Monte Carlo simulations for each value of γ . This figure considers two cases for the GSSK-modulated MIMO system, namely: (a) an overloaded (under-determined) MIMO system with $\zeta = 0.85$, which has poorer performance when compared to the underloaded (overdetermined) system in (b).



(a)



(b)

Figure 7. Probability of error performance as a function of the regularization coefficient. (a) Overloaded MIMO: $\zeta = 0.85$. (b) Underloaded MIMO: $\zeta = 2$.

Furthermore, as expected, higher correlation between the regression parameters leads to poorer

performance, resulting in larger prediction errors. The impact of correlations is evident in Figure 8, where we plot the P_e of the same GSSK-modulated system as a function of the SNR and vary the correlation coefficient ρ . Similar to Section 5, we selected γ that minimizes the probability of error expression in (6.3), which we denote by $\gamma = \gamma_*^{P_e}$.

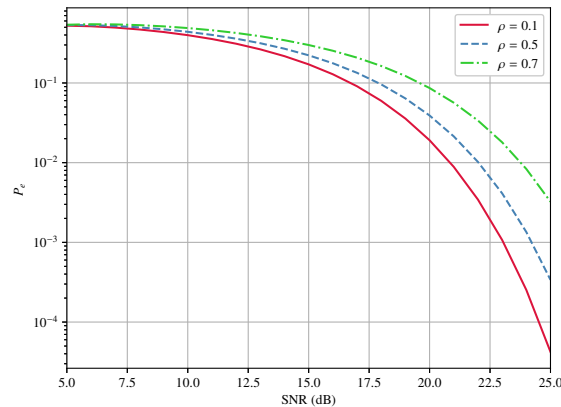


Figure 8. Probability of error performance as a function of the SNR. We used (6.3) with $\zeta = 2$, $E = 1$, $\Delta = 0.2$, $\gamma = \gamma_*^{P_e}$, and $\xi = 10^{-2}$.

7. Conclusions and future work

In this study, we conducted large-dimensional analysis of the error performance of the sqrt-lasso in the asymptotic linear regime with the assumption of a correlated Gaussian design matrix when used for the recovery of a sparse signal from noisy linear observations. Specifically, we derived tight limiting approximations of several measures including the r.m.s.e., m.a.e., cosine similarity, and sparse recovery rates. Numerical simulations were presented throughout the paper to support the obtained theoretical results. Finally, we presented an application of the results of this work to study the large-dimensional limit of the probability of error of the sqrt-lasso in modern modulation schemes such as the GSSK modulation that is used to encode sparse data symbols in a MIMO wireless communication system. Moreover, we observe that numerical simulations indicating our theoretical results remain accurate even for problems involving only a few hundred variables, despite being theoretically derived under the assumption of asymptotic problem dimensions.

It is essential to discuss some potential limitations of the current work.

- Single-sided correlation model:** The analysis in this paper is based on a left-sided correlated Gaussian design matrix, which simplifies modeling by assuming correlation only at the receiver side. While this model is relevant to many wireless communication applications, it may not fully capture the complexities of real-world systems where correlation might exist on both sides (i.e., the transmitter and the receiver). A more comprehensive model, such as a double-sided correlation model, may provide a better reflection of practical scenarios, particularly in MIMO systems. The extension of this work to include such correlation models involves additional technical challenges and is left as future work.

- **Gaussian design assumption:** The assumption of a Gaussian design matrix simplifies the mathematical analysis and is essential for proving the asymptotic results using the CGMT framework (refer to Appendix A). However, in many real-world applications, especially in wireless communications and machine learning, the design matrix may not strictly follow the Gaussian assumption.

Nevertheless, we expect the asymptotic results derived in this study to be robust and applicable to a broader class of random matrices. This expectation is supported by rigorous proofs found in [62–64], where it has been demonstrated that the asymptotic predictions exhibit a universal limit with respect to various types of random matrices. Future research could extend this analysis to non-Gaussian scenarios.

- **Asymptotic results:** The analysis in this paper is carried out in the asymptotic linear regime, assuming that the number of observations, regressors, and nonzero elements grow infinitely large. While this provides a mathematically tractable framework for deriving tight error bounds, real-world systems often involve finite dimensions, which may deviate from these asymptotic assumptions. As a result, the performance guarantees may be less tight for small to moderate-sized problems. Although our numerical simulations suggest that the results hold for problems of moderate size, further exploration is needed to understand the limitations of our approach in lower-dimensional regimes.
- **Perfect design matrix assumption:** Our analysis assumes that the design matrix is perfectly known and free from errors. However, in practical applications, especially in communications systems, the design matrix (i.e., the channel matrix in MIMO systems) may be estimated with errors, leading to imperfect knowledge of the system. Matrix imperfections due to channel estimation errors or noise can degrade the performance of the proposed sqrt-lasso algorithm. Extending the analysis to account for matrix imperfections and designing robust recovery algorithms is an important direction for future work.

Furthermore, this work could be extended by investigating more advanced modulation schemes, such as generalized spatial modulation (GSM).

Author contributions

Ayed M. Alrashdi: Conceptualization, formal analysis, investigation, methodology, validation, writing – original draft, review, edition; Masad A. Alrasheedi: Conceptualization, formal analysis, methodology, validation, writing – original draft, review, edition. All authors have read and approved the final version of the manuscript for publication.

Conflict of interest

The authors declare no conflict of interest.

References

1. D. L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory*, **52** (2006), 1289–1306. <https://doi.org/10.1109/TIT.2006.871582>

2. D. L. Donoho, A. Maleki, A. Montanari, Message-passing algorithms for compressed sensing, *Proc. Natl. Acad. Sci. U.S.A.*, **106** (2009), 18914–18919. <https://doi.org/10.1073/pnas.0909892106>
3. I. B. Atitallah, C. Thrampoulidis, A. Kammoun, T. Y. Al-Naffouri, M. Alouini, B. Hassibi, The BOX-LASSO with application to GSSK modulation in massive MIMO systems, In: *2017 IEEE International symposium on information theory (ISIT)*, Germany: IEEE, 2017, 1082–1086. <https://doi.org/10.1109/ISIT.2017.8006695>
4. M. Lustig, D. Donoho, J. M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, *Magn. Reson. Med.*, **58** (2007), 1182–1195. <https://doi.org/10.1002/mrm.21391>
5. M. I. Jordan, T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, *Science*, **349** (2015), 255–260. <https://doi.org/10.1126/science.aaa8415>
6. R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **58** (1996), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
7. P. Bühlmann, S. Van De Geer, *Statistics for high-dimensional data: Methods, theory and applications*, Heidelberg: Springer Berlin, 2011. <https://doi.org/10.1007/978-3-642-20192-9>
8. A. Belloni, V. Chernozhukov, L. Wang, Square-root lasso: Pivotal recovery of sparse signals via conic programming, *Biometrika*, **98** (2011), 791–806. <https://doi.org/10.1093/biomet/asr043>
9. Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, P. Vanderghenst, Compressed sensing imaging techniques for radio interferometry, *Mon. Not. Roy. Astron. Soc.*, **395** (2009), 1733–1742. <https://doi.org/10.1111/j.1365-2966.2009.14665.x>
10. A. M. Alrashdi, A. E. Alrashdi, A. Alghadhbhan, M. A. H. Eleiwa, Optimum GSSK transmission in massive MIMO systems using the Box-LASSO decoder, *IEEE Access*, **10** (2022), 15845–15859. <https://doi.org/10.1109/ACCESS.2022.3148329>
11. P. Waldmann, G. Mészáros, B. Gredler, C. Fuerst, J. Sölkner, Evaluation of the lasso and the elastic net in genome-wide association studies, *Front. Genet.*, **4** (2013), 270. <https://doi.org/10.3389/fgene.2013.00270>
12. Y. Chu, S. M. Ali, M. Lu, Y. Zhang, Incorporating heterogeneous features into the random subspace method for bearing fault diagnosis, *Entropy*, **25** (2023), 1194. <https://doi.org/10.3390/e25081194>
13. I. T. Jolliffe, N. T. Trendafilov, M. Uddin, A modified principal component technique based on the LASSO, *J. Comput. Graph. Statist.*, **12** (2003), 531–547. <https://doi.org/10.1198/1061860032148>
14. N. Tang, S. Mao, Y. Wang, R. M. Nelms, Solar power generation forecasting with a LASSO-based approach, *IEEE Internet Things J.*, **5** (2018), 1090–1099. <https://doi.org/10.1109/JIOT.2018.2812155>
15. M. Pawlak, J. Lv, Analysis of large scale power systems via lasso learning algorithms, In: *Artificial intelligence and soft computing. ICAISC 2019*, Cham: Springer, **11508** (2019), 652–662. https://doi.org/10.1007/978-3-030-20912-4_59
16. Y. Li, Y. Li, Y. Sun, Online static security assessment of power systems based on LASSO algorithm, *Appl. Sci.*, **8** (2018), 1442. <https://doi.org/10.3390/app8091442>
17. H. Ohlsson, L. Ljung, Identification of switched linear regression models using sum-of-norms regularization, *Automatica*, **49** (2013), 1045–1050. <https://doi.org/10.1016/j.automatica.2013.01.031>

18. A. Chiuso, G. Pillonetto, A bayesian approach to sparse dynamic network identification, *Automatica*, **48** (2012), 1553–1565. <https://doi.org/10.1016/j.automatica.2012.05.054>
19. S. L. Kukreja, J. Löfberg, M. J. Brenner, A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification, *IFAC Proc. Vol.*, **39** (2006), 814–819. <https://doi.org/10.3182/20060329-3-AU-2901.00128>
20. M. Mézard, G. Parisi, M. A. Virasoro, *Spin glass theory and beyond: An introduction to the replica method and its applications*, **9** (1986), 476. <https://doi.org/10.1142/0271>
21. C. Thrampoulidis, S. Oymak, B. Hassibi, The Gaussian min-max theorem in the presence of convexity, *arXiv:1408.4837*, 2014. <https://doi.org/10.48550/arXiv.1408.4837>
22. E. Candes, T. Tao, The dantzig selector: Statistical estimation when p is much larger than n , *Ann. Statist.*, **35** (2007), 2313–2351. <https://doi.org/10.1214/009053606000001523>
23. M. J. Wainwright, Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (Lasso), *IEEE Trans. Inform. Theory*, **55** (2009), 2183–2202. <https://doi.org/10.1109/TIT.2009.2016018>
24. P. J. Bickel, Y. Ritov, A. B. Tsybakov, Simultaneous analysis of Lasso and dantzig selector, *Ann. Statist.*, **37** (2009), 1705–1732. <https://doi.org/10.1214/08-AOS620>
25. G. M. James, C. L. Paulson, P. Rusevich, *The constrained Lasso*, 2012.
26. M. Bayati, A. Montanari, The Lasso risk for Gaussian matrices, *IEEE Trans. Inform. Theory*, **58** (2011), 1997–2017. <https://doi.org/10.1109/TIT.2011.2174612>
27. M. Bayati, J. Pereira, A. Montanari, The Lasso risk: Asymptotic results and real world examples, *Adv. Neural Inf. Process. Syst.*, **23** (2010).
28. M. Vehkaperä, Y. Kabashima, S. Chatterjee, Analysis of regularized LS reconstruction and random matrix ensembles in compressed sensing, *IEEE Trans. Inform. Theory*, **62** (2016), 2100–2124. <https://doi.org/10.1109/TIT.2016.2525824>
29. S. Rangan, V. Goyal, A. K. Fletcher, Asymptotic analysis of MAP estimation via the replica method and compressed sensing, *Adv. Neural Inf. Process. Syst.*, **22** (2009).
30. Y. Kabashima, T. Wadayama, T. Tanaka, Statistical mechanical analysis of a typical reconstruction limit of compressed sensing, In: *2010 IEEE international symposium on information theory*, Austin: IEEE, 2010, 1533–1537. <https://doi.org/10.1109/ISIT.2010.5513526>
31. M. Stojnic, Recovery thresholds for l_1 optimization in binary compressed sensing, In: *2010 IEEE international symposium on information theory*, Austin: IEEE, 2010, 1593–1597. <https://doi.org/10.1109/ISIT.2010.5513435>
32. M. Stojnic, A framework to characterize performance of LASSO algorithms, *arXiv:1303.7291*, 2013. <https://doi.org/10.48550/arXiv.1303.7291>
33. S. Oymak, C. Thrampoulidis, B. Hassibi, The squared-error of generalized LASSO: A precise analysis, In: *2013 51st Annual allerton conference on communication, control, and computing (Allerton)*, Monticello: IEEE, 2013, 1002–1009. <https://doi.org/10.1109/Allerton.2013.6736635>
34. C. Thrampoulidis, S. Oymak, B. Hassibi, Simple error bounds for regularized noisy linear inverse problems, In: *2014 IEEE international symposium on information theory*, Honolulu: IEEE, 2014, 3007–3011. <https://doi.org/10.1109/ISIT.2014.6875386>

35. C. Thrampoulidis, S. Oymak, B. Hassibi, Regularized linear regression: A precise analysis of the estimation error, In: *28th Conference on learning theory*, **40** (2015), 1683–1709.
36. C. Thrampoulidis, A. Panahi, B. Hassibi, Asymptotically exact error analysis for the generalized equation-LASSO, In: *2015 IEEE international symposium on information theory (ISIT)*, Hong Kong: IEEE, 2015, 2021–2025. <https://doi.org/10.1109/ISIT.2015.7282810>
37. C. Thrampoulidis, A. Panahi, D. Guo, B. Hassibi, Precise error analysis of the LASSO, In: *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, South Brisbane: IEEE, 2015, 3467–3471. <https://doi.org/10.1109/ICASSP.2015.7178615>
38. C. Thrampoulidis, E. Abbasi, B. Hassibi, Precise error analysis of regularized M -estimators in high dimensions, *IEEE Trans. Inform. Theory*, **64** (2018), 5592–5628. <https://doi.org/10.1109/TIT.2018.2840720>
39. E. Abbasi, C. Thrampoulidis, B. Hassibi, General performance metrics for the LASSO, In: *2016 IEEE information theory workshop (ITW)*, Cambridge: IEEE Trans, 2016, 181–185. <https://doi.org/10.1109/ITW.2016.7606820>
40. C. Thrampoulidis, E. Abbasi, B. Hassibi, Lasso with non-linear measurements is equivalent to one with linear measurements, *Adv. Neural Inf. Process. Syst.*, 2015, 3420–3428.
41. A. M. Alrashdi, I. B. Atitallah, T. Y. Al-Naffouri, M. S. Alouini, Precise performance analysis of the LASSO under matrix uncertainties, In: *2017 IEEE global conference on signal and information processing (GlobalSIP)*, Montreal: IEEE, 2017, 1290–1294. <https://doi.org/10.1109/GlobalSIP.2017.8309169>
42. A. M. Alrashdi, M. Alazmi, M. A. Alrasheedi, Generalized penalized constrained regression: Sharp guarantees in high dimensions with noisy features, *Mathematics*, **11** (2023), 3706. <https://doi.org/10.3390/math11173706>
43. A. M. Alrashdi, I. B. Atitallah, T. Y. Al-Naffouri, Precise performance analysis of the box-elastic net under matrix uncertainties, *IEEE Signal Process. Lett.*, **26** (2019), 655–659. <https://doi.org/10.1109/LSP.2019.2897215>
44. M. Hebiri, J. Lederer, How correlations influence Lasso prediction, *IEEE Trans. Inform. Theory*, **59** (2012), 1846–1854. <https://doi.org/10.1109/TIT.2012.2227680>
45. M. Celentano, A. Montanari, Y. Wei, The Lasso with general Gaussian designs with applications to hypothesis testing, *Anna. Statist.*, **51** (2023), 2194–2220. <https://doi.org/10.1214/23-AOS2327>
46. A. M. Alrashdi, H. Sifaou, A. Kammoun, M.-S. Alouini, T. Y. Al-Naffouri, Precise error analysis of the Lasso under correlated designs, *arXiv:2008.13033*, 2020. <https://doi.org/10.48550/arXiv.2008.13033>
47. A. M. Alrashdi, H. Sifaou, A. Kammoun, M. S. Alouini, T. Y. Al-Naffouri, Box-relaxation for bpsk recovery in massive MIMO: A precise analysis under correlated channels, In: *ICC 2020-2020 IEEE international conference on communications (ICC)*, Dublin: IEEE, 2020, 1–6. <https://doi.org/10.1109/ICC40277.2020.9149198>
48. A. M. Alrashdi, Large system analysis of box-relaxation in correlated massive MIMO systems under imperfect CSI, In: *2021 IEEE globecom workshops (GC Wkshps)*, Spain: IEEE, 2021, 1–6. <https://doi.org/10.1109/GCWkshps52748.2021.9682159>

49. A. M. Alrashdi, Asymptotic characterisation of regularised zero-forcing receiver for imperfect and correlated massive multiple-input multiple-output systems, *IET Signal Process.*, **16** (2022), 413–425. <https://doi.org/10.1049/si2.12105>
50. J. Jeganathan, A. Ghrayeb, L. Szczecinski, Generalized space shift keying modulation for MIMO channels, In: *2008 IEEE 19th international symposium on personal, indoor and mobile radio communications*, France: IEEE, 2008, 1–5. <https://doi.org/10.1109/PIMRC.2008.4699782>
51. A. Adhikary, J. Nam, J.Y. Ahn, G. Caire, Joint spatial division and multiplexing—the large-scale array regime, *IEEE Trans. Inform. Theory*, **59** (2013), 6441–6463. <https://doi.org/10.1109/TIT.2013.2269476>
52. P. Xia, L. Zhang, F. Li, Learning similarity with cosine similarity ensemble, *Inform. Sci.*, **307** (2015), 39–52. <https://doi.org/10.1016/j.ins.2015.02.024>
53. C. Thrampoulidis, *Recovering structured signals in high dimensions via non-smooth convex optimization: Precise performance analysis*, California Institute of Technology, 2016. <https://doi.org/doi:10.7907/Z998850V>
54. H. Shin, M. Z. Win, J. H. Lee, M. Chiani, On the capacity of doubly correlated MIMO channels, *IEEE Trans. Wirel. Commun.*, **5** (2006), 2253–2265. <https://doi.org/10.1109/TWC.2006.1687741>
55. S. Diamond, S. Boyd, CVXPY: A Python-embedded modeling language for convex optimization, *J. Mach. Learn. Res.*, **17** (2016), 1–5.
56. C. Giacobino, S. Sardy, J. Diaz-Rodriguez, N. Hengartner, Quantile universal threshold, *Electron. J. Statist.*, **11** (2017), 4701–4722. <https://doi.org/10.1214/17-EJS1366>
57. S. Sardy, X. Ma, Sparse additive models in high dimensions with wavelets, *Scand. J. Statist.*, **51** (2024), 89–108. <https://doi.org/10.1111/sjos.12680>
58. D. G. Luenberger, Y. Ye, *Linear and nonlinear programming*, Cham: Springer, 2021. <https://doi.org/10.1007/978-3-030-85450-8>
59. L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, R. Zhang, An overview of massive MIMO: Benefits and challenges, *IEEE J Selected Topics Signal Process.*, **8** (2014), 742–758. <https://doi.org/10.1109/JSTSP.2014.2317671>
60. J. Jeganathan, A. Ghrayeb, L. Szczecinski, A. Ceron, Space shift keying modulation for MIMO channels, *IEEE Trans. Wirel. Commun.*, **8** (2009), 3692–3703. <https://doi.org/10.1109/TWC.2009.080910>
61. C. M. Yu, S. H. Hsieh, H.W. Liang, C. S. Lu, W. H. Chung, S. Y. Kuo, et al., Compressed sensing detector design for space shift keying in MIMO systems, *IEEE Commun. Lett.*, **16** (2012), 1556–1559. <https://doi.org/10.1109/LCOMM.2012.091212.121319>
62. E. Abbasi, F. Salehi, B. Hassibi, Universality in learning from linear measurements, *Adv. Neural Inf. Process. Syst.*, **32** (2019).
63. H. Hu and Y. M. Lu, Universality laws for high-dimensional learning with random features, *IEEE Trans. Inform. Theory*, **69** (2022), 1932–1964. <https://doi.org/10.1109/TIT.2022.3217698>
64. F. Gerace, F. Krzakala, B. Loureiro, L. Stephan, L. Zdeborová, Gaussian universality of perceptrons with random labels, *Phys. Rev. E*, **109** (2024), 034305. <https://doi.org/10.1103/PhysRevE.109.034305>

65. W. Gander, G. H. Golub, U. von Matt, A constrained eigenvalue problem, *Linear Algebra Appl.*, **114-115** (1989), 815–839. [https://doi.org/10.1016/0024-3795\(89\)90494-1](https://doi.org/10.1016/0024-3795(89)90494-1)
66. P. D. Tao, L. T. H. An, A D.C. optimization algorithm for solving the trust-region subproblem, *SIAM J. Optim.*, **8** (1998), 476–505. <https://doi.org/10.1137/S1052623494274313>
67. S. Adachi, S. Iwata, Y. Nakatsukasa, A. Takeda, Solving the trust-region subproblem by a generalized eigenvalue problem, *SIAM J. Optim.*, **27** (2017), 269–291. <https://doi.org/10.1137/16M1058200>
68. O. Dhifallah, Y. M. Lu, A precise performance analysis of learning with random features, *arXiv:2008.11904*, 2020. <https://doi.org/10.48550/arXiv.2008.11904>
69. R. Couillet, M. Debbah, *Random matrix methods for wireless communications*, Cambridge: Cambridge University Press, 2011. <https://doi.org/10.1017/CBO9780511994746>

A. Outline of the proof of the main theoretical results

In this appendix, we prove the main results of this paper, as presented in Section 3. The primary technical tool used in the proof is the CGMT framework [21], [38, Theorem VI.1]. In this framework, the original optimization problem (such as the sqrt-LASSO in (1.2)) is first reformulated as a minimax optimization problem known as the primal optimization (PO) problem. Subsequently, a simpler optimization problem, called the auxiliary optimization (AO) problem, is associated with the PO problem, which is easier to analyze in the large-dimensional setting. The CGMT shows that if the AO cost concentrates around a certain value $c > 0$, then the PO cost does as well. Furthermore, under certain technical conditions regarding the AO solutions, the CGMT establishes that the concentration of Lipschitz functions of the AO solution implies a similar concentration for the PO solution. For the complete statement of the CGMT and its related technical assumptions, please refer to Appendix C and [38, Theorem VI.1], and [21].

In the following, we utilize the CGMT framework to assess the performance of the sqrt-lasso. As an initial step, we formulate the sqrt-lasso optimization as a PO problem.

A.1. Formulating the sqrt-lasso problem as a PO and an AO

PO Formulation. The sqrt-lasso problem in (1.2) can be written in a vector-form as follows

$$\widehat{\boldsymbol{\theta}} := \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^N} \frac{\|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2}{\sqrt{M}} + \frac{\gamma}{M} \|\boldsymbol{\theta}\|_1. \quad (\text{A.1})$$

Multiplying by \sqrt{M} , we get

$$\widehat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^N} \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2 + \frac{\gamma}{\sqrt{M}} \|\boldsymbol{\theta}\|_1. \quad (\text{A.2})$$

To directly address the error metrics introduced in (2.4) and (2.5), we consider changing the optimization variable to $\mathbf{e} := \boldsymbol{\theta} - \boldsymbol{\theta}_0$. With this substitution, the problem in (A.2) can be recast as follows:

$$\widehat{\mathbf{e}} := \arg \min_{\mathbf{e} \in \mathbb{R}^N} \|\mathbf{A}\mathbf{e} - \boldsymbol{\sigma}\mathbf{w}\|_2 + \frac{\gamma}{\sqrt{M}} \|\mathbf{e} + \boldsymbol{\theta}_0\|_1. \quad (\text{A.3})$$

By leveraging the invariance of the normal distribution under orthogonal transformations, we arrive at

$$\widehat{\mathbf{e}} = \arg \min_{\mathbf{e} \in \mathbb{R}^N} \|\mathbf{Q}^{\frac{1}{2}} \widetilde{\mathbf{H}} \mathbf{e} - \sigma \mathbf{w}\|_2 + \frac{\gamma}{\sqrt{M}} \|\mathbf{e} + \boldsymbol{\theta}_0\|_1, \quad (\text{A.4})$$

where $\widetilde{\mathbf{H}}$ is a random matrix with IID normal components $\widetilde{H}_{kj} \sim \mathcal{N}(0, \frac{1}{N})$. The loss function can be rewritten in its dual form through the Fenchel (convex) conjugate as follows:

$$\|\mathbf{Q}^{\frac{1}{2}} \widetilde{\mathbf{H}} \mathbf{e} - \sigma \mathbf{w}\|_2 = \max_{\|\mathbf{u}\|_2 \leq 1} \mathbf{u}^\top (\mathbf{Q}^{\frac{1}{2}} \widetilde{\mathbf{H}} \mathbf{e} - \sigma \mathbf{w}).$$

Using this, the problem in (A.4) can be formulated as a min-max optimization problem as follows:

$$\min_{\mathbf{e} \in \mathbb{R}^N} \max_{\|\mathbf{u}\|_2 \leq 1} \mathbf{u}^\top \mathbf{Q}^{\frac{1}{2}} \widetilde{\mathbf{H}} \mathbf{e} - \mathbf{u}^\top \sigma \mathbf{w} + \frac{\gamma}{\sqrt{M}} \|\mathbf{e} + \boldsymbol{\theta}_0\|_1. \quad (\text{A.5})$$

The problem above is not yet in the desired PO form because of the presence of the \mathbf{Q} matrix. To address this, we redefine \mathbf{u} as $\mathbf{u} = \mathbf{Q}^{\frac{1}{2}} \mathbf{u}$, which results in:

$$\min_{\mathbf{e} \in \mathbb{R}^N} \max_{\mathbf{u}^\top \mathbf{Q}^{-1} \mathbf{u} \leq 1} \mathbf{u}^\top \widetilde{\mathbf{H}} \mathbf{e} - \mathbf{u}^\top \mathbf{Q}^{-\frac{1}{2}} \sigma \mathbf{w} + \frac{\gamma}{\sqrt{M}} \|\mathbf{e} + \boldsymbol{\theta}_0\|_1, \quad (\text{A.6})$$

where $(\cdot)^{-1}$ is the inverse operator. We finally arrive at the following PO formulation

$$\min_{\mathbf{e} \in \mathbb{R}^N} \max_{\mathbf{u}^\top \mathbf{Q}^{-1} \mathbf{u} \leq 1} \frac{1}{\sqrt{N}} \mathbf{u}^\top \mathbf{H} \mathbf{e} - \mathbf{u}^\top \mathbf{Q}^{-\frac{1}{2}} \sigma \mathbf{w} + \frac{\gamma}{\sqrt{M}} \|\mathbf{e} + \boldsymbol{\theta}_0\|_1, \quad (\text{A.7})$$

where \mathbf{H} has IID standard normal elements. Additionally, we can factor out $\frac{1}{\sqrt{N}}$ resulting in

$$\frac{1}{\sqrt{N}} \min_{\mathbf{e} \in \mathbb{R}^N} \max_{\mathbf{u}^\top \mathbf{Q}^{-1} \mathbf{u} \leq 1} \mathbf{u}^\top \mathbf{H} \mathbf{e} - \sqrt{N} \mathbf{u}^\top \mathbf{Q}^{-\frac{1}{2}} \sigma \mathbf{w} + \sqrt{\frac{N}{M}} \gamma \|\mathbf{e} + \boldsymbol{\theta}_0\|_1. \quad (\text{A.8})$$

Recall that $\zeta = M/N$; thus, the PO problem is given by:

$$\frac{1}{\sqrt{N}} \min_{\mathbf{e} \in \mathbb{R}^N} \max_{\mathbf{u}^\top \mathbf{Q}^{-1} \mathbf{u} \leq 1} \mathbf{u}^\top \mathbf{H} \mathbf{e} - \sqrt{N} \mathbf{u}^\top \mathbf{Q}^{-\frac{1}{2}} \sigma \mathbf{w} + \frac{\gamma}{\sqrt{\zeta}} \|\mathbf{e} + \boldsymbol{\theta}_0\|_1. \quad (\text{A.9})$$

Note that the compactness of the constraint set over \mathbf{e} must be considered. This can be addressed rigorously in a way similar to that in [38, Section VI-A]. The details are omitted here, as they are not essential to our main objective.

AO Formulation. The optimization problem above is in PO format, and according to the CGMT framework outlined in [38, Theorem VI.1] and [21], we can formulate the following AO problem.

$$\frac{1}{\sqrt{N}} \min_{\mathbf{e} \in \mathbb{R}^N} \max_{\mathbf{u}^\top \mathbf{Q}^{-1} \mathbf{u} \leq 1} \|\mathbf{e}\|_2 \mathbf{g}^\top \mathbf{u} + \|\mathbf{u}\|_2 \mathbf{h}^\top \mathbf{e} - \sqrt{N} \mathbf{u}^\top \mathbf{Q}^{-\frac{1}{2}} \sigma \mathbf{w} + \frac{\gamma}{\sqrt{\zeta}} \|\mathbf{e} + \boldsymbol{\theta}_0\|_1, \quad (\text{A.10})$$

where $\mathbf{g} \in \mathbb{R}^m$, and $\mathbf{h} \in \mathbb{R}^n$ have IID standard normal entries. To further simplify, the AO can be represented as:

$$\frac{1}{\sqrt{N}} \min_{\mathbf{e} \in \mathbb{R}^N} \max_{\substack{\mathbf{u} \\ \mu \geq 0}} \|\mathbf{e}\|_2 \mathbf{g}^\top \mathbf{u} + \|\mathbf{u}\|_2 \mathbf{h}^\top \mathbf{e} - \sqrt{N} \mathbf{u}^\top \mathbf{Q}^{-\frac{1}{2}} \sigma \mathbf{w} + \sqrt{N} \mu (1 - \mathbf{u}^\top \mathbf{Q}^{-1} \mathbf{u}) + \frac{\gamma}{\sqrt{\zeta}} \|\mathbf{e} + \boldsymbol{\theta}_0\|_1, \quad (\text{A.11})$$

where $\tilde{\mu}$ is a Lagrange multiplier.

Given that $\|\mathbf{x}\|_1 = \max_{\|\mathbf{v}\|_\infty \leq 1} \mathbf{x}^\top \mathbf{v}$, for any vector \mathbf{x} , it follows that:

$$\begin{aligned} & \frac{1}{\sqrt{N}} \min_{\mathbf{e} \in \mathbb{R}^N} \max_{\substack{\mathbf{u} \\ \tilde{\mu} \geq 0}} \max_{\|\mathbf{v}\|_\infty \leq 1} \|\mathbf{e}\|_2 \mathbf{g}^\top \mathbf{u} + \|\mathbf{u}\|_2 \mathbf{h}^\top \mathbf{e} - \sqrt{N} \mathbf{u}^\top \mathbf{Q}^{-\frac{1}{2}} \sigma \mathbf{w} \\ & \quad + \sqrt{N} \tilde{\mu} (1 - \mathbf{u}^\top \mathbf{Q}^{-1} \mathbf{u}) + \frac{\gamma}{\sqrt{\zeta}} (\mathbf{e} + \boldsymbol{\theta}_0)^\top \mathbf{v}. \end{aligned} \quad (\text{A.12})$$

Setting $\alpha := \frac{\|\mathbf{e}\|_2}{\sqrt{N}}$, and defining $\bar{\mathbf{e}} := \frac{\mathbf{e}}{\|\mathbf{e}\|_2}$, the above problem can be represented as

$$\max_{\substack{\mathbf{u} \\ \tilde{\mu} \geq 0 \\ \|\mathbf{v}\|_\infty \leq 1}} \min_{\alpha \geq 0} \alpha \mathbf{g}^\top \mathbf{u} - \mathbf{u}^\top \mathbf{Q}^{-\frac{1}{2}} \sigma \mathbf{w} + \tilde{\mu} (1 - \mathbf{u}^\top \mathbf{Q}^{-1} \mathbf{u}) + \frac{\gamma}{\sqrt{M}} \boldsymbol{\theta}_0^\top \mathbf{v} + \min_{\|\bar{\mathbf{e}}\|_2=1} \alpha \left(\frac{\gamma}{\sqrt{\zeta}} \mathbf{v} + \|\mathbf{u}\|_2 \mathbf{h} \right)^\top \bar{\mathbf{e}}. \quad (\text{A.13})$$

The last minimization is straightforward and can be carried out as follows:

$$\min_{\|\bar{\mathbf{e}}\|_2=1} \alpha \left(\frac{\gamma}{\sqrt{\zeta}} \mathbf{v} + \|\mathbf{u}\|_2 \mathbf{h} \right)^\top \bar{\mathbf{e}} = -\alpha \left\| \frac{\gamma}{\sqrt{\zeta}} \mathbf{v} + \|\mathbf{u}\|_2 \mathbf{h} \right\|_2,$$

with an optimizer given by:

$$\bar{\mathbf{e}}_* = - \frac{\frac{\gamma}{\sqrt{\zeta}} \mathbf{v} + \|\mathbf{u}\|_2 \mathbf{h}}{\left\| \frac{\gamma}{\sqrt{\zeta}} \mathbf{v} + \|\mathbf{u}\|_2 \mathbf{h} \right\|_2}. \quad (\text{A.14})$$

After dividing by \sqrt{N} , we arrive at the following:

$$\begin{aligned} & \max_{\substack{\mathbf{u} \\ \tilde{\mu} \geq 0 \\ \|\mathbf{v}\|_\infty \leq 1}} \min_{\alpha \geq 0} \frac{\alpha}{\sqrt{N}} \mathbf{g}^\top \mathbf{u} - \frac{1}{\sqrt{N}} \mathbf{u}^\top \mathbf{Q}^{-\frac{1}{2}} \sigma \mathbf{w} + \frac{1}{\sqrt{N}} \tilde{\mu} (1 - \mathbf{u}^\top \mathbf{Q}^{-1} \mathbf{u}) \\ & \quad + \frac{\gamma}{\sqrt{NM}} \boldsymbol{\theta}_0^\top \mathbf{v} - \frac{\alpha}{\sqrt{N}} \left\| \frac{\gamma}{\sqrt{\zeta}} \mathbf{v} + \|\mathbf{u}\|_2 \mathbf{h} \right\|_2. \end{aligned} \quad (\text{A.15})$$

Define $\tilde{\mathbf{g}} := \alpha \mathbf{g} - \mathbf{Q}^{-\frac{1}{2}} \sigma \mathbf{w}$, $\mu = \frac{\tilde{\mu}}{\sqrt{N}}$, and with the norm of \mathbf{u} fixed to β (i.e., $\|\mathbf{u}\|_2 = \beta$), we obtain that

$$\min_{\alpha \geq 0} \max_{\substack{\beta \geq 0 \\ \|\bar{\mathbf{u}}\|_2=1 \\ \mu \geq 0 \\ \|\mathbf{v}\|_\infty \leq 1}} \frac{\beta}{\sqrt{N}} \tilde{\mathbf{g}}^\top \bar{\mathbf{u}} - \beta^2 \mu \bar{\mathbf{u}}^\top \mathbf{Q}^{-1} \bar{\mathbf{u}} + \mu + \frac{\gamma}{\sqrt{MN}} \boldsymbol{\theta}_0^\top \mathbf{v} - \frac{\alpha}{\sqrt{N}} \left\| \frac{\gamma}{\sqrt{\zeta}} \mathbf{v} + \beta \mathbf{h} \right\|_2. \quad (\text{A.16})$$

The optimization over $\bar{\mathbf{u}}$ is now separable, resulting in the following non-convex optimization problem (due to the norm equality constraint) that needs to be solved:

$$\max_{\|\bar{\mathbf{u}}\|_2=1} \frac{\beta}{\sqrt{N}} \tilde{\mathbf{g}}^\top \bar{\mathbf{u}} - \beta^2 \mu \bar{\mathbf{u}}^\top \mathbf{Q}^{-1} \bar{\mathbf{u}}. \quad (\text{A.17})$$

This is a well-known optimization problem referred to as the trust region subproblem, which has been thoroughly investigated in previous studies such as in [65–67]. It has been also shown in [67, 68]

that optimization problems of this type can be expressed in terms of scalar optimization problems as follows:

$$\min_{\|\mathbf{r}\|_2=r} \mathbf{c}^\top \mathbf{r} + \frac{1}{2} \mathbf{r}^\top \mathbf{B} \mathbf{r} = \sup_{t > -\vartheta_N} \left\{ -\frac{1}{2} \mathbf{c}^\top (\mathbf{B} + t\mathbf{I})^{-1} \mathbf{c} - \frac{t}{2} r^2 \right\},$$

where $r \geq 0$, ϑ_N is the minimum eigenvalue of matrix \mathbf{B} , and \mathbf{I} is the identity matrix. Therefore, using this, we have

$$\begin{aligned} & \max_{\|\bar{\mathbf{u}}\|_2=1} \frac{1}{\sqrt{N}} \beta \bar{\mathbf{g}}^\top \bar{\mathbf{u}} - \beta^2 \mu \bar{\mathbf{u}}^\top \mathbf{Q}^{-1} \bar{\mathbf{u}} = -2\mu\beta^2 \min_{\|\bar{\mathbf{u}}\|_2=1} \left\{ -\frac{1}{2\mu\beta\sqrt{N}} \bar{\mathbf{g}}^\top \bar{\mathbf{u}} + \frac{1}{2} \bar{\mathbf{u}}^\top \mathbf{Q}^{-1} \bar{\mathbf{u}} \right\} \\ & = -2\mu\beta^2 \sup_{t > -\vartheta_N} \left\{ \frac{-1}{8\mu^2\beta^2 N} \bar{\mathbf{g}}^\top (\mathbf{Q}^{-1} + t\mathbf{I})^{-1} \bar{\mathbf{g}} - \frac{t}{2} \right\} \\ & = 2\mu\beta^2 \min_{t > -\vartheta_N} \left\{ \frac{1}{8\mu^2\beta^2 N} \bar{\mathbf{g}}^\top (\mathbf{Q}^{-1} + t\mathbf{I})^{-1} \bar{\mathbf{g}} + \frac{t}{2} \right\} \\ & = \min_{t > -\vartheta_N} \left\{ \frac{1}{4\mu N} \bar{\mathbf{g}}^\top (\mathbf{Q}^{-1} + t\mathbf{I})^{-1} \bar{\mathbf{g}} + t\beta^2 \mu \right\}, \end{aligned} \quad (\text{A.18})$$

with ϑ_N being the minimum eigenvalue of the matrix \mathbf{Q}^{-1} , i.e., $\vartheta_N = 1/q_{\max} = 1/q_1$. The analysis above demonstrates that the AO problem presented in (A.16) can be equivalently expressed as:

$$\min_{\alpha \geq 0} \max_{\substack{\beta \geq 0 \\ \mu \geq 0 \\ \|\mathbf{v}\|_\infty \leq 1}} \min_{t > -\vartheta_N} \frac{1}{4\mu N} \bar{\mathbf{g}}^\top (\mathbf{Q}^{-1} + t\mathbf{I})^{-1} \bar{\mathbf{g}} + (t\beta^2 + 1)\mu + \frac{\gamma}{\sqrt{MN}} \boldsymbol{\theta}_0^\top \mathbf{v} - \frac{\alpha}{\sqrt{N}} \left\| \frac{\gamma}{\sqrt{\zeta}} \mathbf{v} + \beta \mathbf{h} \right\|_2. \quad (\text{A.19})$$

We continue by representing the ℓ_2 -norm in (A.19) by using the next identity

$$\|\mathbf{x}\|_2 = \min_{\chi > 0} \frac{\|\mathbf{x}\|_2^2}{2\chi} + \frac{\chi}{2}, \quad (\text{A.20})$$

with the minimizer $\widehat{\chi} = \|\mathbf{x}\|_2$. Thus, the AO can be reformulated as:

$$\begin{aligned} & \min_{\alpha \geq 0} \max_{\substack{\beta \geq 0 \\ \mu \geq 0 \\ \|\mathbf{v}\|_\infty \leq 1 \\ \chi > 0}} \min_{t > -\vartheta_N} \frac{1}{4\mu N} \bar{\mathbf{g}}^\top (\mathbf{Q}^{-1} + t\mathbf{I})^{-1} \bar{\mathbf{g}} + (t\beta^2 + 1)\mu \\ & + \frac{\gamma}{\sqrt{MN}} \boldsymbol{\theta}_0^\top \mathbf{v} - \frac{\alpha\chi}{2} - \frac{\alpha}{2\chi} \left\| \frac{1}{\sqrt{N}} \left(\frac{\gamma}{\sqrt{\zeta}} \mathbf{v} + \beta \mathbf{h} \right) \right\|_2^2. \end{aligned} \quad (\text{A.21})$$

Next, we carry out the optimization over \mathbf{v} given its separability. Thus, define

$$J_N(\alpha, \beta, \chi, \boldsymbol{\theta}_0, \mathbf{h}) := \frac{1}{N} \left[\max_{\|\mathbf{v}\|_\infty \leq 1} \frac{\gamma}{\sqrt{\zeta}} \boldsymbol{\theta}_0^\top \mathbf{v} - \frac{\alpha}{2\chi} \left\| \left(\frac{\gamma}{\sqrt{\zeta}} \mathbf{v} + \beta \mathbf{h} \right) \right\|_2^2 \right]. \quad (\text{A.22})$$

This can be rewritten as

$$J_N(\alpha, \beta, \chi, \boldsymbol{\theta}_0, \mathbf{h}) = -\frac{\alpha\gamma^2}{2\chi\zeta N} \sum_{k=1}^N \min_{|v_k| \leq 1} v_k^2 + \frac{2\beta\sqrt{\zeta}}{\gamma} h_k v_k - \frac{2\sqrt{\zeta}\chi}{\alpha\gamma} \theta_{0,k} v_k - \frac{1}{N} \sum_{k=1}^N \frac{\alpha\beta^2}{2\chi} h_k^2. \quad (\text{A.23})$$

Moreover, we can further rewrite $J_N(\alpha, \beta, \chi, \boldsymbol{\theta}_0, \mathbf{h})$ as

$$J_N(\alpha, \beta, \chi, \boldsymbol{\theta}_0, \mathbf{h}) = -\frac{\alpha\gamma^2}{2\chi\zeta N} \sum_{k=1}^N \min_{|v_k| \leq 1} \left(v_k - \left(\frac{\sqrt{\zeta}\chi}{\alpha\gamma} \theta_{0,k} - \frac{\sqrt{\zeta}\beta}{\gamma} h_k \right) \right)^2 + \frac{\alpha\gamma^2}{2\chi\zeta N} \left(\frac{\sqrt{\zeta}\chi}{\alpha\gamma} \theta_{0,k} - \frac{\sqrt{\zeta}\beta}{\gamma} h_k \right)^2 - \frac{1}{N} \sum_{k=1}^N \frac{\alpha\beta^2}{2\chi} h_k^2. \quad (\text{A.24})$$

Without much effort, it can be demonstrated that the above minimization yields the squared value of the wavelet shrinkage function defined in (2.1). Thus, we obtain

$$J_N(\alpha, \beta, \chi, \boldsymbol{\theta}_0, \mathbf{h}) = -\frac{\alpha\gamma^2}{2\chi\zeta N} \sum_{k=1}^N \eta^2 \left(\frac{\sqrt{\zeta}\chi}{\alpha\gamma} \theta_{0,k} - \frac{\sqrt{\zeta}\beta}{\gamma} h_k; 1 \right) + \frac{\alpha\gamma^2}{2\chi\zeta N} \sum_{k=1}^N \left(\frac{\sqrt{\zeta}\chi}{\alpha\gamma} \theta_{0,k} - \frac{\sqrt{\zeta}\beta}{\gamma} h_k \right)^2 - \frac{1}{N} \sum_{k=1}^N \frac{\alpha\beta^2}{2\chi} h_k^2. \quad (\text{A.25})$$

Using [38, Lemma A.10], it is possible to show, for $c \in \mathbb{R}$ and $\tau > 0$, that

$$\tau \sum_{k=1}^N \eta^2 \left(\frac{1}{\tau} \theta_{0,k} + \frac{c}{\tau} h_k; 1 \right) = \frac{1}{\tau} \sum_{k=1}^N \eta^2 (\theta_{0,k} + ch_k; \tau). \quad (\text{A.26})$$

Using the above result, with $\tau = \frac{\alpha\gamma}{\sqrt{\zeta}\chi}$, we get

$$J_N(\alpha, \beta, \chi, \boldsymbol{\theta}_0, \mathbf{h}) = -\frac{\chi}{2\alpha N} \sum_{k=1}^N \eta^2 \left(\theta_{0,k} - \frac{\alpha\beta}{\chi} h_k; \frac{\alpha\gamma}{\sqrt{\zeta}\chi} \right) + \frac{\alpha\gamma^2}{2\chi\zeta N} \sum_{k=1}^N \left(\frac{\sqrt{\zeta}\chi}{\alpha\gamma} \theta_{0,k} - \frac{\sqrt{\zeta}\beta}{\gamma} h_k \right)^2 - \frac{1}{N} \sum_{k=1}^N \frac{\alpha\beta^2}{2\chi} h_k^2. \quad (\text{A.27})$$

With $J_N(\alpha, \beta, \chi, \boldsymbol{\theta}_0, \mathbf{h})$ as defined above, the AO problem in (A.21) writes

$$\min_{\alpha \geq 0} \max_{\substack{\beta \geq 0 \\ \mu \geq 0 \\ \chi > 0}} \min_{t > -\vartheta_N} \frac{1}{4\mu N} \tilde{\mathbf{g}}^\top (\mathbf{Q}^{-1} + t\mathbf{I})^{-1} \tilde{\mathbf{g}} + (t\beta^2 + 1)\tilde{\mu} - \frac{\alpha\chi}{2} + J_N(\alpha, \beta, \chi, \boldsymbol{\theta}_0, \mathbf{h}). \quad (\text{A.28})$$

In the sequel, we study the large-dimensional asymptotic convergence of the AO problem as given above.

A.2. AO Asymptotic convergence

The next step involves transforming the AO into a scalar problem, i.e., a problem that consists of scalar variables only. To achieve this, start by noting that $\tilde{\mathbf{g}}$ has a multivariate Gaussian distribution with a zero mean vector and a covariance matrix $\Sigma_{\tilde{\mathbf{g}}}$ that is given by $\Sigma_{\tilde{\mathbf{g}}} = \alpha^2 \mathbf{I}_M + \sigma^2 \mathbf{Q}^{-1}$. To study the convergence of the first term of (A.28), we apply the trace lemma [69] from random matrix theory to get:

$$\frac{1}{N} \tilde{\mathbf{g}}^\top (\mathbf{Q}^{-1} + t\mathbf{I})^{-1} \tilde{\mathbf{g}} - \frac{1}{N} \text{tr} \left(\Sigma_{\tilde{\mathbf{g}}} (\mathbf{Q}^{-1} + t\mathbf{I})^{-1} \right) \xrightarrow{P} 0, \quad (\text{A.29})$$

where $\text{tr}(\cdot)$ is the trace operator. Furthermore, by applying the weak law of large numbers (WLLN), we can show the following convergences:

$$\frac{1}{N} \sum_{k=1}^N h_k \theta_{0,k} \xrightarrow{P} 0, \frac{1}{N} \sum_{k=1}^N h_k^2 \xrightarrow{P} 1, \text{ and } \frac{1}{N} \sum_{k=1}^N \theta_{0,k}^2 \xrightarrow{P} \sigma_{\theta_0}^2.$$

In addition, for all $\alpha \geq 0, \beta \geq 0$, and $\chi > 0$:

$$J_N(\alpha, \beta, \chi, \boldsymbol{\theta}_0, \mathbf{h}) \xrightarrow{P} \frac{-\chi}{2\alpha} \mathbb{E}_{\substack{\Theta_0 \sim p_{\Theta_0} \\ H \sim N(0,1)}} \left[\eta^2 \left(\Theta_0 - \frac{\alpha\beta}{\chi} H; \frac{\alpha\gamma}{\chi \sqrt{\zeta}} \right) \right] + \frac{\chi \sigma_{\theta_0}^2}{2\alpha}. \quad (\text{A.30})$$

Combining all of the above convergence results and using [38, Lemma 10], it can be shown that the AO problem in (A.28) converges to:

$$\begin{aligned} \min_{\substack{\alpha \geq 0 \\ t > 0}} \max_{\substack{\beta \geq 0 \\ \mu \geq 0 \\ \chi > 0}} \min_{t > -\vartheta_N} (t\beta^2 + 1)\mu - \frac{\alpha\chi}{2} + \frac{1}{4\mu N} \sum_{j=1}^M \frac{q_j \alpha^2 + \sigma^2}{1 + q_j t} \\ - \frac{\chi}{2\alpha} \mathbb{E}_{\substack{\Theta_0 \sim p_{\Theta_0} \\ H \sim N(0,1)}} \left[\eta^2 \left(\Theta_0 - \frac{\alpha\beta}{\chi} H; \frac{\alpha\gamma}{\chi \sqrt{\zeta}} \right) \right] + \frac{\chi \sigma_{\theta_0}^2}{2\alpha}, \end{aligned} \quad (\text{A.31})$$

where q_j is the j^{th} eigenvalue of matrix Ψ .

Using the change of variables $t \mapsto t + \vartheta_N$, and after flipping the order of min-max, we have

$$\begin{aligned} \min_{\substack{\alpha \geq 0 \\ t > 0}} \max_{\substack{\beta \geq 0 \\ \mu \geq 0 \\ \chi > 0}} [(t - \vartheta_N)\beta^2 + 1]\mu - \frac{\alpha\chi}{2} + \frac{\chi \sigma_{\theta_0}^2}{2\alpha} + \frac{1}{4\mu N} \sum_{j=1}^M \frac{q_j \alpha^2 + \sigma^2}{1 + q_j(t - \vartheta_N)} \\ - \frac{\chi}{2\alpha} \mathbb{E}_{\substack{\Theta_0 \sim p_{\Theta_0} \\ H \sim N(0,1)}} \left[\eta^2 \left(\Theta_0 - \frac{\alpha\beta}{\chi} H; \frac{\alpha\gamma}{\chi \sqrt{\zeta}} \right) \right]. \end{aligned} \quad (\text{A.32})$$

This optimization problem is exactly the scalar minimax problem in (3.2) of Theorem 3.1, with $\vartheta_N = 1/q_1$.

B. Large-dimensional error performance evaluation

B.1. R.M.S.E. Convergence analysis: Proof of Theorem 3.1

Having derived the scalar optimization problem in (A.32), which governs the large-dimensional behavior of the sqrt-lasso, we next proceed to prove the convergence of the r.m.s.e. as stated in Theorem 3.1.

Toward this goal, let $\tilde{\mathbf{e}}$ be the minimizer of the AO problem that was defined in (A.13). Moreover, let $\hat{\alpha}_N$ denote the optimizer of (A.13) in α . Then, recalling the definition of α introduced in the previous section, we can see that $\hat{\alpha}_N = \frac{\|\tilde{\mathbf{e}}\|_2}{\sqrt{N}}$. Using the asymptotic convergence of the AO problems derived earlier, one can show that $\hat{\alpha}_N \xrightarrow{P} \alpha_*$, where α_* is the unique minimizer of (3.2). This implies that

$$\frac{1}{\sqrt{N}} \|\tilde{\mathbf{e}}\|_2 \xrightarrow{P} \alpha_*. \quad (\text{B.1})$$

In order to utilize the CGMT framework [21], we need to consider the following set defined as:

$$\mathcal{S}_\epsilon = \left\{ \mathbf{x} \in \mathbb{R}^N : \left| \frac{1}{\sqrt{N}} \|\mathbf{x}\|_2 - \alpha_\star \right| < \epsilon \right\},$$

for some sufficiently small $\epsilon > 0$. From the convergence in (B.1), it follows immediately that $\lim_{N \rightarrow \infty} \mathbb{P}(\tilde{\mathbf{e}} \in \mathcal{S}_\epsilon) = 1$. Hence, according to the CGMT statement in [38, Theorem VI.1], it also holds that $\lim_{N \rightarrow \infty} \mathbb{P}(\widehat{\mathbf{e}} \in \mathcal{S}_\epsilon) = 1$, where $\widehat{\mathbf{e}} = \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ is the optimizer of the original sqrt-lasso in (A.3) (after the change of variables to \mathbf{e}). This establishes the proof of Theorem 3.1.

B.2. M.A.E. Convergence analysis: Proof of Proposition 3.1

We proceed now to the proof of the m.a.e. of the sqrt-lasso as given in Proposition 3.1. First, we change the concentration set to the following:

$$\check{\mathcal{S}}_\epsilon = \left\{ \mathbf{x} \in \mathbb{R}^N : \left| \frac{1}{N} \|\mathbf{x}\|_1 - \mathbb{E}_{\Theta_0, H} \left[\left| \eta \left(\Theta_0 - \frac{\alpha_\star \beta_\star}{\chi_\star} H; \frac{\gamma \alpha_\star}{\chi_\star \sqrt{\zeta}} \right) - \Theta_0 \right| \right] \right| < \epsilon \right\},$$

where $\alpha_\star, \beta_\star$, and χ_\star are the unique optimizers of (3.2).

Next, we note that based on (A.14), the AO solution satisfies:

$$\tilde{\mathbf{e}} = -\widehat{\alpha}_N \frac{\left(\frac{\gamma}{\sqrt{\zeta}} \mathbf{v}^* + \widehat{\beta}_N \mathbf{h} \right)}{\frac{1}{\sqrt{N}} \left\| \frac{\gamma}{\sqrt{\zeta}} \mathbf{v}^* + \widehat{\beta}_N \mathbf{h} \right\|_2}, \quad (\text{B.2})$$

and $\frac{1}{\sqrt{N}} \left\| \frac{\gamma}{\sqrt{\zeta}} \mathbf{v}^* + \widehat{\beta}_N \mathbf{h} \right\|_2 = \widehat{\chi}_N$, where $\widehat{\alpha}_N, \widehat{\beta}_N$, and $\widehat{\chi}_N$ are the solutions of (A.28). This means that:

$$\tilde{e}_k = -\widehat{\alpha}_N \frac{\left(\frac{\gamma}{\sqrt{\zeta}} v_k^* + \widehat{\beta}_N h_k \right)}{\widehat{\chi}_N}, \quad \text{for } k = 1, 2, \dots, N,$$

where v_k^* is the maximizer of (A.22). By standard differentiation, it is easy to show that:

$$v_k^* = \begin{cases} -1, & \text{if } \left(\frac{\chi \sqrt{\zeta}}{\alpha \gamma} \right)^2 \lambda_k < -1, \\ \left(\frac{\chi \sqrt{\zeta}}{\alpha \gamma} \right)^2 \lambda_k, & \text{if } -1 \leq \left(\frac{\chi \sqrt{\zeta}}{\alpha \gamma} \right)^2 \lambda_k \leq 1, \\ 1, & \text{if } \left(\frac{\chi \sqrt{\zeta}}{\alpha \gamma} \right)^2 \lambda_k > 1, \end{cases} \quad (\text{B.3})$$

where $\lambda_k = \frac{\gamma \alpha}{\chi \sqrt{\zeta}} \theta_{0,k} - \frac{\alpha^2 \beta \gamma}{\chi^2 \sqrt{\zeta}} h_k$. Recalling that $\tilde{\boldsymbol{\theta}}_k = \tilde{e}_k + \theta_{0,k}$, and after substituting the expressions of \tilde{e}_k and v_k^* , along with a bit of algebraic manipulation, it follows that

$$\tilde{\boldsymbol{\theta}}_k = \eta \left(\theta_{0,k} - \frac{\widehat{\alpha}_N \widehat{\beta}_N}{\widehat{\chi}_N} h_k; \frac{\gamma \widehat{\alpha}_N}{\widehat{\chi}_N \sqrt{\zeta}} \right).$$

Using the asymptotic convergence of the AO problems derived earlier, it holds that $\widehat{\alpha}_N \xrightarrow{P} \alpha_\star, \widehat{\beta}_N \xrightarrow{P} \beta_\star$, and $\widehat{\chi}_N \xrightarrow{P} \chi_\star$. Thus, using the WLLN, the following convergence holds true:

$$\frac{1}{N} \|\tilde{\mathbf{e}}\|_1 = \frac{1}{N} \sum_{k=1}^N |\tilde{e}_k| = \frac{1}{N} \sum_{k=1}^N |\tilde{\boldsymbol{\theta}}_k - \theta_{0,k}| \xrightarrow{P} \mathbb{E}_{\Theta_0, H} \left[\left| \eta \left(\Theta_0 - \frac{\alpha_\star \beta_\star}{\chi_\star} H; \frac{\gamma \alpha_\star}{\chi_\star \sqrt{\zeta}} \right) - \Theta_0 \right| \right]. \quad (\text{B.4})$$

This demonstrates that $\tilde{\mathbf{e}} \in \check{\mathcal{S}}_\epsilon$ with probability approaching 1. Consequently, by applying the CGMT, we can conclude that $\widehat{\mathbf{e}} \in \mathcal{S}_\epsilon$ with probability approaching 1, thereby proving Proposition 3.1.

C. The large-dimensional analysis framework: CGMT

In this section, we present an overview of the primary analytical framework employed to derive the theoretical results discussed in the paper, which is the Convex Gaussian Minimax Theorem (CGMT). For the reader's ease, we begin by revisiting the CGMT. This theorem is an extension of Gordon's comparison lemma, which has been applied to various large-dimensional inference problems. The CGMT was first introduced in [32] and further expanded in [38]. It leverages convexity to compare the min max values of given two Gaussian random processes.

To demonstrate the core concepts of this framework, let us begin by considering the next Gaussian processes:

$$\mathcal{X}_{\mathbf{e}, \mathbf{u}} := \mathbf{u}^\top \mathbf{H} \mathbf{e} + \phi(\mathbf{e}, \mathbf{u}), \quad (\text{C.1a})$$

$$\mathcal{Y}_{\mathbf{e}, \mathbf{u}} := \|\mathbf{e}\|_2 \mathbf{g}^\top \mathbf{u} + \|\mathbf{u}\|_2 \mathbf{h}^\top \mathbf{e} + \phi(\mathbf{e}, \mathbf{u}), \quad (\text{C.1b})$$

where $\mathbf{H} \in \mathbb{R}^{M \times N}$, $\mathbf{g} \in \mathbb{R}^M$, $\mathbf{h} \in \mathbb{R}^N$, and $\phi : \mathbb{R}^N \times \mathbb{R}^M \rightarrow \mathbb{R}$. Moreover, \mathbf{h} , \mathbf{g} and \mathbf{H} are all assumed to have IID standard normal components. For those random processes, let us consider the following minimax optimization problems, known as the primal optimization (PO) and the auxiliary optimization (AO):

$$F(\mathbf{H}) := \min_{\mathbf{e} \in \mathcal{S}_e} \max_{\mathbf{u} \in \mathcal{S}_u} \mathcal{X}_{\mathbf{e}, \mathbf{u}}, \quad (\text{C.2a})$$

$$f(\mathbf{g}, \mathbf{h}) := \min_{\mathbf{e} \in \mathcal{S}_e} \max_{\mathbf{u} \in \mathcal{S}_u} \mathcal{Y}_{\mathbf{e}, \mathbf{u}}, \quad (\text{C.2b})$$

where $\mathcal{S}_e \subset \mathbb{R}^N$ and $\mathcal{S}_u \subset \mathbb{R}^M$ are compact and convex sets. Furthermore, if the function $\phi(\mathbf{e}, \mathbf{u})$ is continuous and exhibits convex-concave properties on $\mathcal{S}_e \times \mathcal{S}_u$, then, following the CGMT framework outlined in [38, Theorem VI.1], and for all $\omega \in \mathbb{R}$ and $\varrho > 0$, it holds:

$$\mathbb{P}(|F(\mathbf{H}) - \omega| > \varrho) \leq 2\mathbb{P}(|f(\mathbf{g}, \mathbf{h}) - \omega| > \varrho). \quad (\text{C.3})$$

The result above indicates that if we can demonstrate that the optimal cost of the AO converges in probability as $f(\mathbf{g}, \mathbf{h}) \xrightarrow{P} d_\star$ asymptotically, where $d_\star \in \mathbb{R}$, then we can also conclude that the optimal cost of the PO converges in probability to $F(\mathbf{H}) \xrightarrow{P} d_\star$. The key idea is that analyzing the AO is typically much simpler than analyzing the PO. Moreover, the CGMT (Theorem VI.1 in [38]) establishes that if the optimal solution of the AO concentrates around a certain value, then the optimal solution of the PO will concentrate around the same value as well. Specifically, if the solutions of (C.2b) satisfy $\|\widehat{\mathbf{e}}_f(\mathbf{g}, \mathbf{h})\|_2 \xrightarrow{P} r_\star$, where $r_\star > 0$, then the same holds true for the solutions of (C.2a), meaning $\|\widehat{\mathbf{e}}_F(\mathbf{H})\|_2 \xrightarrow{P} r_\star$.

Additionally, we utilize the following theorem, which applies in the large-dimensional asymptotic regime.

Theorem C.1 (Asymptotic CGMT [38]). *With the same assumptions and notations as discussed earlier, let $\mathcal{S} \subset \mathcal{S}_e$ and $\mathcal{S}^c := \mathcal{S}_e / \mathcal{S}$. Moreover, let us define $F_{\mathcal{S}^c}(\mathbf{H})$ and $f_{\mathcal{S}^c}(\mathbf{g}, \mathbf{h})$ as the optimal costs in (C.2a) and (C.2b), respectively, with the constraint that the optimization is limited to $\mathbf{e} \in \mathcal{S}^c$. Also, given that there are constants $\bar{\varphi} < \bar{\varphi}_{\mathcal{S}^c}$, satisfying $f_{\mathcal{S}^c}(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \bar{\varphi}_{\mathcal{S}^c}$ and $f(\mathbf{g}, \mathbf{h}) \xrightarrow{P} \bar{\varphi}$, it follows that*

$$\lim_{N \rightarrow \infty} \mathbb{P}(\widehat{\mathbf{e}}_{F(\mathbf{H})} \in \mathcal{S}) = 1. \quad (\text{C.4})$$

For additional information on the CGMT framework, the reader is encouraged to refer to [38].



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)