Kybernetika

Rakshith Jagannath; Neelesh S. Upadhye
The LASSO estimator: Distributional properties

Kybernetika, Vol. 54 (2018), No. 4, 778–797

Persistent URL: `http://dml.cz/dmlcz/147424`

Terms of use:

© Institute of Information Theory and Automation AS CR, 2018

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* `http://dml.cz`

KYBERNETIKA — VOLUME *54* (2018), NUMBER 4, PAGES 778–797

THE LASSO ESTIMATOR: DISTRIBUTIONAL PROPERTIES

Rakshith Jagannath and Neelesh S. Upadhye

The least absolute shrinkage and selection operator (LASSO) is a popular technique for simultaneous estimation and model selection. There have been a lot of studies on the large sample asymptotic distributional properties of the LASSO estimator, but it is also well-known that the asymptotic results can give a wrong picture of the LASSO estimator's actual finite-sample behaviour. The finite sample distribution of the LASSO estimator has been previously studied for the special case of orthogonal models. The aim in this work is to generalize the finite sample distribution properties of LASSO estimator for a real and linear measurement model in Gaussian noise.

In this work, we derive an expression for the finite sample characteristic function of the LASSO estimator, we then use the Fourier slice theorem to obtain an approximate expression for the marginal probability density functions of the one-dimensional components of a linear transformation of the LASSO estimator.

Keywords: linear regression, LASSO, characteristic function, finite sample probability distribution function, Fourier-Slice theorem, Cramer–Wold theorem

Classification: 62E15, 62J05, 62G05, 60E05

1. INTRODUCTION AND MOTIVATION

LASSO (least absolute shrinkage and selection operator) [7, 24] has been developed as a tool to find sparse solutions of the linear regression problem. It has been used extensively in an expanding field of applications from statistics to estimation scenarios with remarkably good results. It is used extensively in the parameter estimation framework to estimate the unknown parameter(s) with guarantees for the estimation error (model fit) [3, 9]. As a robust approximation of the well known Maximum Likelihood (ML) estimator, the LASSO can be realized robustly and efficiently through convex Second Order Cone (SOC) programming techniques [4] and unlike the efficient subspace methods [17], the LASSO technique is reliable even with one data measurement realization (single snapshot) [22].

The regularization parameter (sparsity threshold parameter) in the LASSO is a mathematical tool to implement the compromise between model fit (measure of estimation error) and the estimated model order, which is the number of nonzero entries in the

DOI: 10.14736/kyb-2018-4-0778

estimate. As the regularization parameter evolves, the LASSO solution changes continuously, forming a continuous trajectory in a very high dimensional space which is referred to as the LASSO path [10, 22].

The LASSO algorithm, in general either assumes the knowledge of sparsity of the unknown parameter or the optimum regularization parameter for estimating the unknown parameter. However, these are generally unknown and need to be estimated.

In the detection framework, the focus is to propose detection tests to estimate the optimal sparsity threshold parameter, so that the number of non-zero entries (or the sparsity) and their corresponding locations (indices) in the estimate is same as in the actual parameter. The performance evaluation of the detection tests is done using the p-values or the probability of correct detection. As sparsity plays an important role in the estimation performance, this issue has been recognized as a significant gap between theory and practice by several authors [5, 11]. The problem of estimating the sparsity of the parameter from the measurement data is fundamental to many other applications such as estimating the critical number of measurements for successful recovery, design of the sensing matrix e.t.c [20]. The exploration of the detection framework of the LASSO is fairly recent (e. g. [19] and its citations). Sparsity and model order estimation techniques like statistical cross-validation, Mallow's Cp selection, Stein's unbiased risk estimator and Bayesian information criteria (BIC) and its variants have been proposed in [1, 5, 20] for estimation of sparsity (or sparsity threshold parameter). Bayesian based LASSO estimation, wherein the linear model is interpreted from a Bayesian perspective and the sparsity threshold parameter is modeled as a hyper-prior for estimation, have been proposed in [2]. Hence we see that an accurate estimate of the sparsity or the sparsity threshold parameter is critical for enhancing the performance of parameter estimation using LASSO. However, the performance of the above techniques depends on the initial guess of the sparsity threshold parameter or the choice of a grid for the sparsity threshold parameter. These techniques also use the distribution of the ML estimator instead of the distribution of the LASSO estimator for deriving the tests for model order estimation. Hence, we note that the detection framework requires the distribution of the LASSO estimator for proposing reliable detection tests and evaluating their performance. So, there is a need for studying the distributional properties of the LASSO estimator. Hence, in this work we explore the distribution of the LASSO estimator.

There has been a lot of related work in the understanding of the asymptotic distributional properties of LASSO estimator, e. g., [16], but it is also well-known that the asymptotic results can give a wrong picture of the LASSO estimator's actual finite-sample behavior [14, 18, 23]. In particular, [23] studies the finite sample LASSO distribution for the case of orthogonal models and shows that the asymptotic results do not provide a reliable assessment for the finite sample distribution. Therefore there is a need for studying the finite sample distributional properties of the LASSO estimator. Hence, in this work, we will study the finite sample characteristic function (cf) of the LASSO estimator for any general model matrix.

1.1. Major contributions

We now highlight the major contributions of this work.

- We find a closed form approximation for the marginal pdfs of the LASSO estimator as a function of the regularization parameter, τ. To the best of our knowledge, there are no closed form approximations for the marginal pdfs of the LASSO estimator in the literature.

- We use the approach of characteristic functions and the Fourier slice theorem tool for deriving the marginal pdfs. While this approach and the tool existed separately in the literature, our contribution lies in applying these in conjunction to obtain the closed form expressions of pdf of the LASSO estimator for non-orthogonal and singular models as described in the paper.

This article is organized as follows. In Section 2, we discuss the notations and state without proof, some well known theorems that will be used in this work. In Section 3, we give the details of our results on the finite sample cf and approximate probability density function (pdf) of the LASSO estimator for a fixed $N \in \mathbb{N}$. In Section 4, we perform numerical simulations to verify the results discussed in Section 3. In Section 5, we discuss the conclusions and some possibilities for future work. We give the details of the proofs of the theorems stated in Section 3 in Appendix A.

2. PRELIMINARIES

In this section, we discuss the notations, measurement model, LASSO estimator and some well known theorems used in this work.

2.1. Notations

We use bold lower case letters to represent vectors ($\mathbf{x}$), bold upper case letters to represent matrices ($\mathbf{A}$) and scripted letters to represent sets ($\mathcal{J}$, generally finite index sets). $\mathcal{J} \setminus \mathcal{I}$ denotes the set-difference operation between sets $\mathcal{J}$ and $\mathcal{I}$. For a given matrix (vector) $\mathbf{A}$, $\mathbf{A}^T$ denotes the regular transpose, $|\mathbf{A}|$ and $\mathbf{A}^{-1}$ denote the determinant and inverse of the square matrix $\mathbf{A}$ and $\mathbf{A}^\dagger$ denotes the Moore-Penrose pseudo inverse of $\mathbf{A}$. For a vector $\mathbf{x}$, $\|\mathbf{x}\|_0$, $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$ denote the l_0 pseudo norm which is equal to the number of non-zero elements in $\mathbf{x}$, l_1 and l_2 norms respectively. $\mathbb{E}$ denotes expectation, $\mathbb{P}$ denotes probability. We reserve $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ for random vectors, $\hat{x}_k$ and $\hat{z}_k$ for their kth components and $\mathbf{t}$ for the realizations of $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ and t_k for the realizations of $\hat{x}_k$ and $\hat{z}_k$. We use $\mathbf{v}$ to denote a Gaussian random vector. For a random vector $\hat{\mathbf{x}}$, $f_{\hat{\mathbf{x}}}(\mathbf{t})$ denotes its pdf. Similarly for a random variable $\hat{x}$, $f_{\hat{x}}(t_k)$ denotes its pdf. Convolution of $f([x_1, x_2, \ldots, x_N])$ with $g(x_1)$ is denoted equivalently by $f(\mathbf{x}) \star g(x_1)$ or $f(\overline{x}_1, x_2, \ldots, x_N)$ or $f(\overline{x}_1, \mathbf{x}^-)$, which is defined as the following integral

$$f(\overline{x}_1, \mathbf{x}^-) = f(\mathbf{x}) \star g(x_1) \quad = \quad \int_{\mathbb{R}} f(x_1 - u, \mathbf{x}^-) g(u) \, \mathrm{d}u \tag{1}$$

where $\mathbf{x}^- = [x_2, \ldots, x_N]$. Similarly $f(\overline{\mathbf{x}}_{\mathcal{I}}, \mathbf{x}_{\mathcal{J} \setminus \mathcal{I}})$ denotes the $|\mathcal{I}|$ dimensional extension of convolution across the dimensions of $\mathbf{x}$ given by $\mathcal{I}$ and finally $\mathbb{S}(\mathbf{u})$ denotes the sign of elements of $\mathbf{u}$, we have

$$\mathbb{S}(u_i) \quad = \quad \begin{cases} -1 & \text{if } u_i < 0, \\ \;\;1 & \text{if } u_i > 0. \end{cases} \tag{2}$$

$\mathbb{S}(u_i)$ is arbitrary for $u_i = 0$.

Let $f : \mathbb{R}^N \to \mathbb{R}$ be a function. We define the following operations on f.

1. **Integral Projection:** $\mathcal{I}_{\mathbf{x}}$ is the projection operator that reduces an N dimensional function to M dimensions by integrating out $N - M$ dimensions:

$$\mathcal{I}_{\mathbf{x}}[f](x_1, x_2, \ldots, x_M) = \int f(x_1, \ldots, x_N)\, \mathrm{d}x_{M+1} \ldots \mathrm{d}x_N.$$

   In the above equation, the $N - M$ dimensions can be integrated out in any order and need not necessarily be the last $N - M$ dimensions, i. e., we assume that the order of integration is permutation invariant.

2. **Slicing:** $\mathcal{S}_{\mathbf{x}}$ is the slicing operator that reduces an N dimensional function to an M dimensional function, by zeroing out $N - M$ dimensions: $\mathcal{S}_{\mathbf{x}}[f](x_1, \ldots, x_M) = f(x_1, \ldots, x_M, 0, \ldots, 0)$. The $N - M$ dimensions can be sliced out in any order as described in definition of the projection operation.

3. **Change of Basis:** Let $\mathbf{B}$ denote a full rank $N \times N$ matrix and let $\mathbf{x}$ denote an N dimensional vector. Then $\mathbf{B}[f](\mathbf{x}) = f(\mathbf{B}^{-1}\mathbf{x})$.

The action of multiple operators on the function is denoted by $\odot$. For example, $\mathcal{I}_{\mathbf{Bx}} \odot \mathbf{B}^{-1}[f]$ denotes the change of basis followed by projection operation on the function $f(\mathbf{x})$.

2.2. Measurement model

We consider the following linear regression model,

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{v}, \tag{3}$$

where $\mathbf{b} = [b_1, b_2, \ldots, b_M]$ denotes the measurement vector of length M, $\mathbf{A}$ denotes the model matrix of size $M \times N$, $\mathbf{v}$ denotes the white Gaussian noise with zero mean and covariance matrix, $\sigma^2\mathbf{I}$ and $\mathbf{x} = [x_1, x_2, \ldots, x_N]$ denotes the deterministic but unknown and sparse parameter vector of length N and sparsity K (number of non-zero entries in $\mathbf{x}$), which needs to be estimated. We also assume that columns of $\mathbf{A}$ have unit norm. We focus on the LASSO estimator [7, 24], which estimates the sparse parameter $\mathbf{x}$ in (3) by solving the following convex optimization problem,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathbb{R}^N} \tau \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2, \tag{4}$$

where $\tau > 0$ is the sparsity thresholding parameter, which controls the sparsity of $\hat{\mathbf{x}}$. We observe that (4) is a convex relaxation of the following combinatorial problem

$$\min \|\mathbf{x}\|_0 \ \text{ subject to } \ \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \sigma. \tag{5}$$

It has been shown in [6] that the LASSO estimator is an exact relaxation of (5) if the model matrix, $\mathbf{A}$ satisfies the restricted isometry property (R.I.P) defined below in Definition 2.1, and hence gives the sparsest estimate to the linear regression problem of (3) (see Theorem 2.1 below).

Definition 2.1. For each integer $K = 1, 2, \ldots$, we define the restricted isometry constant δ_K of a matrix $\mathbf{A}$ as the smallest number such that

$$(1 - \delta_K)\|\mathbf{y}\|_2^2 \le \|\mathbf{A}\mathbf{y}\|_2^2 \le (1 + \delta_K)\|\mathbf{y}\|_2^2 \tag{6}$$

holds for all K sparse vectors, $\mathbf{y}$. A vector is said to be K sparse if it has at most K nonzero entries.

Theorem 2.1. Assume that $\delta_{2K} < \sqrt{2} - 1$ and $\|\mathbf{v}\|_2 < \sigma$. Then the solution $\hat{\mathbf{x}}$ to (4) obeys

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \le C_0 \frac{1}{\sqrt{K}} \|\mathbf{x} - \mathbf{x}^K\|_1 + C_1 \sigma \tag{7}$$

for some constants C_0 and C_1. In particular, if $\mathbf{x}$ is K-sparse, the recovery is exact. Here $\mathbf{x}^K$ is the best sparse approximation one could obtain if one knew exactly the locations and amplitudes of the K-largest entries of $\mathbf{x}$.

Proof. See [6]. □

We now state some well known definitions and theorems required for this work.

Definition 2.2. The cf of a random vector $\mathbf{y}$ is,

$$\begin{aligned} \mathbb{F}_{\mathbf{y}}(\mathbf{u}) &= \mathbb{E}_{\mathbf{y}}\{\exp(i\mathbf{u}^T\mathbf{y})\} \\ &= \int_{-\infty}^{\infty} f_{\mathbf{y}}(\mathbf{y}) \exp(i\mathbf{u}^T\mathbf{y})\, \mathrm{d}\mathbf{y}. \end{aligned} \tag{8}$$

Clearly, cf is the Fourier transform of the pdf of $\mathbf{y}$ with $\mathbf{u}$ as the variable in the Fourier domain. The cf has the properties like, cf is a uniformly continuous, bounded and hermitian function with guaranteed existence, $\mathbb{F}_{\mathbf{y}}(\mathbf{0}) = 1$ and cf is a bijection with probability distributions, i.e, for any two random variables X_1 and X_2, both have the same probability distribution if and only if $\mathbb{F}_{X_1} = \mathbb{F}_{X_2}$.

Theorem 2.2. A Borel probability measure $\mathbb{P}$ on $\mathbb{R}^N$ is uniquely determined by its one dimensional projections, i. e. a probability measure on Euclidean space is uniquely determined by the values it gives to half-spaces.

Proof. See [8]. □

Theorem 2.3. Let f be an N dimensional function, let $\mathbb{F}$, $\mathbf{B}$, $\mathcal{I}$, $\mathcal{S}$ represent the Fourier transform, change of basis, projection and slicing operations as explained above, then we have

$$\mathbb{F} \odot \mathcal{I}_{\mathbf{B}^{-1}\mathbf{x}} \odot \mathbf{B}[f] \;=\; \mathcal{S}_{\mathbf{B}^T\mathbf{u}} \odot \frac{\mathbf{B}^{-T}}{|\mathbf{B}^{-T}|} \odot \mathbb{F}[f] \tag{9}$$

where $\mathbf{B}^{-T}$ stands for $(\mathbf{B}^{-1})^T$.

Proof. See [21]. □

Next, we discuss the main results of this work in the following section.

3. MAIN RESULTS

In this section, we first derive an expression for the cf of the LASSO estimator. The expression for cf is appropriately sliced to derive the one dimensional projections of a linear transformation of the LASSO estimator. The one dimensional projections yield the approximate marginal pdfs of components of the linear transformation of the LASSO estimator.

Theorem 3.1. For a fixed N, the cf of the LASSO estimator, $\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{u})$ as a function of the true parameter, $\mathbf{x}$ is given by the following implicit relationship

$$\mathbb{F}_{\mathbf{W}\hat{\mathbf{x}}+\tau\mathbb{S}(\hat{\mathbf{x}})}(\mathbf{u}) = \sum_{\mathcal{I}\in\mathcal{P}}\Big(\prod_{\substack{k\in\mathcal{I}\\ j\in\mathcal{J}\setminus\mathcal{I}}}\sin(\tau u_k)\cos(\tau u_j)\Big)\mathbb{F}_{\hat{\mathbf{x}}}(\overline{\mathbf{c}}_{\mathcal{I}}, \mathbf{c}_{\mathcal{J}\setminus\mathcal{I}}) = \exp\left(i\mathbf{u}^T\mathbf{W}\mathbf{x} - \frac{\sigma^2}{2}\mathbf{u}^T\mathbf{W}\mathbf{u}\right) \tag{10}$$

where $\mathcal{J} = \{1, 2, \ldots, N\}$ is an index set, $\mathcal{P}$ is the power set of $\mathcal{J}$, $\mathcal{I}$ is an element of $\mathcal{P}$ or equivalently $\mathcal{I} \subseteq \mathcal{J}$, $\mathbf{W} = \mathbf{A}^T\mathbf{A}$, $\mathbf{w}_j$ is the jth column of $\mathbf{W}$, $\mathbf{c}_{\mathcal{I}} = \mathbf{u}^T\mathbf{w}_j, j \in \mathcal{I}$ similarly $\mathbf{c}_{\mathcal{J}\setminus\mathcal{I}} = \mathbf{u}^T\mathbf{w}_k, k \in \mathcal{J}\setminus\mathcal{I}$ and $\mathbb{F}_{\hat{\mathbf{x}}}(\overline{\mathbf{c}}_{\mathcal{I}}, \mathbf{c}_{\mathcal{J}\setminus\mathcal{I}})$ denotes the $|\mathcal{I}|$ dimensional convolution of $\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{c}_{\mathcal{I}}, \mathbf{c}_{\mathcal{J}\setminus\mathcal{I}})$ with $g(\mathbf{c}_{\mathcal{I}}) = \prod_{j\in\mathcal{I}}\frac{-1}{\pi c_j}$.

Remark: We note that (10) is an implicit relationship between the estimator $\hat{\mathbf{x}}$ and measurements $\mathbf{b}$. Although, the right hand side seems independent of τ, there is an inherent relationship which is given by the KKT conditions and discussed in the proof of the Theorem 3.1. Example-3.1 below illustrates Theorem 3.1

Example 3.1. When $\hat{\mathbf{x}}$ has $N = 3$ elements, we have $\mathcal{J} = \{1, 2, 3\}$, the power set, $\mathcal{P} = \{\emptyset, 1, 2, 3, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\}$ and $\mathcal{I}$ is one of the elements of $\mathcal{P}$. Hence, the relationship between $\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{u})$ and $\mathbf{x}$ is given by

$$\begin{aligned}
&\mathbb{F}_{\mathbf{W}\hat{\mathbf{x}}+\tau\mathbb{S}(\hat{\mathbf{x}})}(\mathbf{u}) = \\
&\qquad \cos(\tau u_1)\cos(\tau u_2)\cos(\tau u_3)\mathbb{F}_{\hat{\mathbf{x}}}(c_1, c_2, c_3) + \sin(\tau u_1)\cos(\tau u_2)\cos(\tau u_3)\mathbb{F}_{\hat{\mathbf{x}}}(\overline{c}_1, c_2, c_3) \\
&\qquad + \cos(\tau u_1)\sin(\tau u_2)\cos(\tau u_3)\mathbb{F}_{\hat{\mathbf{x}}}(c_1, \overline{c}_2, c_3) + \cos(\tau u_1)\cos(\tau u_2)\sin(\tau u_3)\mathbb{F}_{\hat{\mathbf{x}}}(c_1, c_2, \overline{c}_3) \\
&\qquad + \sin(\tau u_1)\sin(\tau u_2)\cos(\tau u_3)\mathbb{F}_{\hat{\mathbf{x}}}(\overline{c}_1, \overline{c}_2, c_3) + \cos(\tau u_1)\sin(\tau u_2)\sin(\tau u_3)\mathbb{F}_{\hat{\mathbf{x}}}(c_1, \overline{c}_2, \overline{c}_3) \\
&\qquad + \sin(\tau u_1)\cos(\tau u_2)\sin(\tau u_3)\mathbb{F}_{\hat{\mathbf{x}}}(\overline{c}_1, c_2, \overline{c}_3) + \sin(\tau u_1)\sin(\tau u_2)\sin(\tau u_3)\mathbb{F}_{\hat{\mathbf{x}}}(\overline{c}_1, \overline{c}_2, \overline{c}_3) \\
&= \exp\left(i\mathbf{u}^T\mathbf{W}\mathbf{x} - \frac{\sigma^2}{2}\mathbf{u}^T\mathbf{W}\mathbf{u}\right)
\end{aligned}$$

where $\mathbb{F}_{\hat{\mathbf{x}}}(\overline{c}_j, \ldots,)$ represents convolution as defined in (1) with $g(c_j) = \frac{-1}{\pi c_j}$.

Remarks: We make the following observations from Theorem 3.1.

1. We observe that the relationship between $\mathbb{F}_{\mathbf{W}\hat{\mathbf{x}}+\tau\mathbb{S}(\hat{\mathbf{x}})}(\mathbf{u})$ and $\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{c} = \mathbf{W}\mathbf{u})$ has 2^N terms in total, which does not simplify the evaluation of the pdf of LASSO estimator except for the special case of orthogonal model matrix ($\mathbf{W} = \mathbf{I}$), where the entries of $\hat{\mathbf{x}}$ become independent and hence the pdf of $\hat{\mathbf{x}}$ can be obtained easily (see Corollary 3.1). Proof of Corollary 3.1 (Appendix A.2) is an alternate way ([23]) to obtain pdf of $\hat{\mathbf{x}}$, when $\mathbf{W}$ is orthogonal.

2. We observe that the one dimensional projections of the cf can be evaluated from (10) by slicing.

3. We then invoke the Cramer–Wold theorem (Theorem 2.2) to conclude that the one-dimensional projections are sufficient for the evaluation of the joint distribution of the LASSO estimator.

4. We note that for $\tau = 0$, we obtain the cf relationship for the maximum likelihood (ML) estimator as a special case and hence the corresponding pdf of the ML estimator [15] can be easily evaluated.

Next, we evaluate the one dimensional projections of the cf, $\mathbb{F}_{\hat{\mathbf{x}}}$ for different cases of $\mathbf{W}$. We first start with the case of $\mathbf{W} = \mathbf{I}$.

Corollary 3.1. If $\mathbf{W} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, or any diagonal matrix $\mathbf{D}$ (in general), then the estimates $\hat{x}_k, k = 1, 2, \ldots, N$ are all independent and hence the cf of the lasso estimator is given by,

$$\mathbb{F}_{\hat{x}_k}(u_k)\cos(\tau u_k) + \mathbb{F}_{\hat{x}_k}(\overline{u}_k)\sin(\tau u_k) = \exp\big(iu_kx_k - \frac{\sigma^2}{2}u_k^2\big), k = 1, 2, \ldots, N \tag{11}$$

and the marginal pdf of the individual components, $\hat{x}_i, i = 1, 2, \ldots, N$ is obtained by simply applying the inversion theorem to (11) as,

$$f_{\hat{x}_k}(t_k) \quad = \quad \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\big(-\frac{(t_k+\tau-x_k)^2}{2\sigma^2}\big) & \text{if } t_k > 0, \\ \frac{1}{\sqrt{2\pi\sigma^2}}\exp\big(-\frac{(t_k-\tau-x_k)^2}{2\sigma^2}\big) & \text{if } t_k < 0. \end{cases} \tag{12}$$

Remarks:

- The expression (12) is similar to the expression (5) of [23] (we note that [23] have derived the pdf of the estimation error, $\hat{x} - x$).

- We note that when $x_k = 0$, the pdf of positive values of $\hat{x}_k$ is given by the tail of a normal distribution with mean $-\tau$ and the pdf of negative values of $\hat{x}_k$ is given by the tail of a normal distribution with mean τ.

- We note that the joint pdf of $\hat{\mathbf{x}}$ is just the product of the marginal pdfs of $\hat{x}_k, k = 1, 2, \ldots, N$ obtained in (12).

Now, we evaluate the cf for the case when $\mathbf{W}$ is a full rank matrix.

Corollary 3.2. Let $\mathbf{W}$ be any full rank matrix, $\hat{\mathbf{z}} = \mathbf{W}\hat{\mathbf{x}}$, $\mathbf{w}_j$ be the jth column of $\mathbf{W}$ and $\hat{z}_j$ be the jth element of $\hat{\mathbf{z}}$. Then the cf of the one dimensional projections, $\hat{z}_k$ for components k corresponding to large $|\hat{x}_k|$ can be approximated as,

$$\cos(\tau u)\mathbb{F}_{\hat{z}_k}(u) + \sin(\tau u)\mathbb{F}_{\hat{z}_k}(\overline{u}) = \exp(-u^2\frac{\sigma^2}{2}w_{kk})\exp\Big(ju(\mathbf{w}_k^T\mathbf{x})\Big) \tag{13}$$

and the marginal pdfs of the components $\hat{z}_k$, corresponding to the large $|\hat{x}_k|$ is obtained by simply applying the inversion theorem to (13) as,

$$f_{\hat{z}_k}(t_k) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2 w_{kk}}} \exp\big(-\frac{(t_k+\tau-\mathbf{w}_k^T\mathbf{x})^2}{2\sigma^2 w_{kk}}\big) & \text{if } t_k > 0, \\ \frac{1}{\sqrt{2\pi\sigma^2 w_{kk}}} \exp\big(-\frac{(t_k-\tau-\mathbf{w}_k^T\mathbf{x})^2}{2\sigma^2 w_{kk}}\big) & \text{if } t_k < 0. \end{cases} \tag{14}$$

Remarks:

1. In Corollary 3.2, we make the approximation $\mathbb{S}(\mathbf{h}_k^T\hat{\mathbf{z}}) \approx \mathbb{S}(\hat{z}_k)$ to evaluate the expression, $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{Wu})\star(\frac{-1}{\pi c_k})]$ (more details in Appendix A.3). Here $\mathbf{h}_k$ is a column of $\mathbf{H} = \mathbf{W}^{-1}$. This approximation is valid whenever $|h_{kk}\hat{z}_k| \gg \sum_{i=1,i\neq k}^{N} |h_{ki}\hat{z}_i|$. Since $\mathbf{W}$ is a symmetric positive definite (pd) matrix, $\mathbf{H}$ is also symmetric and pd, hence the diagonal entries of $\mathbf{H}$ are positive. So, if $\mathbf{H}$ is diagonally dominant and the index k corresponds to a large entries of $|\hat{x}_k|$, then the approximation works well. From the simulations (Section 4) also, we observe that the approximation works well for the index k corresponding to the large non-zero entries of $|\hat{x}_k|$.

2. The evaluation of exact expression for $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{Wu})\star(\frac{-1}{\pi c_k})]$ requires some prior knowledge or assumptions on the pdf $f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}})$. In Appendix A.5, we derive an expression for $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{Wu})\star(\frac{-1}{\pi c_k})]$ by assuming the inherent distribution of $f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}})$ to be multivariate Gaussian. This assumption is justified because the expression in (10), which represents some operations on cf of $\hat{\mathbf{x}}$ is equal to cf of Gaussian pdf. Also, for the special case of orthogonal $\mathbf{W}$, we obtain a Gaussian pdf for $f(\hat{\mathbf{x}})$. Hence, we can make the assumption that $f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}})$ is multivariate Gaussian. From the resulting expression for $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{Wu})\star(\frac{-1}{\pi c_k})]$, we then justify the use of the approximation $\mathbb{S}(\mathbf{h}_k^T\hat{\mathbf{z}}) \approx \mathbb{S}(\hat{z}_k)$ for k corresponding to large $|\hat{x}_k|$.

Further, we evaluate the cf for any general case of $\mathbf{W}$.

Corollary 3.3. Let $\mathbf{W}$ be any general matrix, $\hat{\mathbf{z}} = \mathbf{W}\hat{\mathbf{x}}$, $\mathbf{w}_j$ be the jth column of $\mathbf{W}$ and $\hat{z}_j$ be the jth element of $\hat{\mathbf{z}}$. Then the cf of the one dimensional projections, $\hat{z}_k$ for components k corresponding to large $|\hat{x}_k|$ can be approximated as,

$$\cos(\tau u)\mathbb{F}_{\hat{z}_k}(u) + \sin(\tau u)\mathbb{F}_{\hat{z}_k}(\overline{u}) = \exp(-u^2\frac{\sigma^2}{2}w_{kk})\exp\Big(ju(\mathbf{w}_k^T\mathbf{x})\Big) \tag{15}$$

and the marginal pdfs of the components $\hat{z}_k$, corresponding to the large $|\hat{x}_k|$ is obtained by simply applying the inversion theorem to (15) as,

$$f_{\hat{z}_k}(t_k) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2 w_{kk}}} \exp\big(-\frac{(t_k+\tau-\mathbf{w}_k^T\mathbf{x})^2}{2\sigma^2 w_{kk}}\big) & \text{if } t_k > 0, \\ \frac{1}{\sqrt{2\pi\sigma^2 w_{kk}}} \exp\big(-\frac{(t_k-\tau-\mathbf{w}_k^T\mathbf{x})^2}{2\sigma^2 w_{kk}}\big) & \text{if } t_k < 0. \end{cases} \tag{16}$$

Remarks:

1. In Corollary 3.3, we again make the approximation $\mathbb{S}(\mathbf{h}_k^T\hat{\mathbf{z}}) \approx \mathbb{S}(\hat{z}_k)$ to evaluate $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})]$ (more details in Appendix A.3). Now, $\mathbf{h}_k$ is a column of $\mathbf{H} = \mathbf{W}^{\dagger}$. The approximation is valid whenever $|h_{kk}\hat{z}_k| \gg \sum_{i=1, i\neq k}^{N} |h_{ki}\hat{z}_i|$. Since $\mathbf{W}$ is a symmetric positive semi definite (psd) matrix, $\mathbf{H}$ is also symmetric and psd, hence the diagonal entries of $\mathbf{H}$ are non-negative. So, if the matrix is diagonally dominant and the index k corresponds to the non-zero entry of $\hat{\mathbf{x}}$, then the approximation works well. From the simulations also we observe that the approximation seems to work for the index k corresponding to non-zero entries of $\hat{\mathbf{x}}$.

2. Again the evaluation of exact expression for $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})]$ requires some prior knowledge or assumptions on the pdf $f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}})$ and the exact expression confirms the feasibility of the approximation $\mathbb{S}(\mathbf{h}_k^T\hat{\mathbf{z}}) \approx \mathbb{S}(\hat{z}_k)$ for the components k corresponding to large $|\hat{x}_k|$.

3.1. Discussions

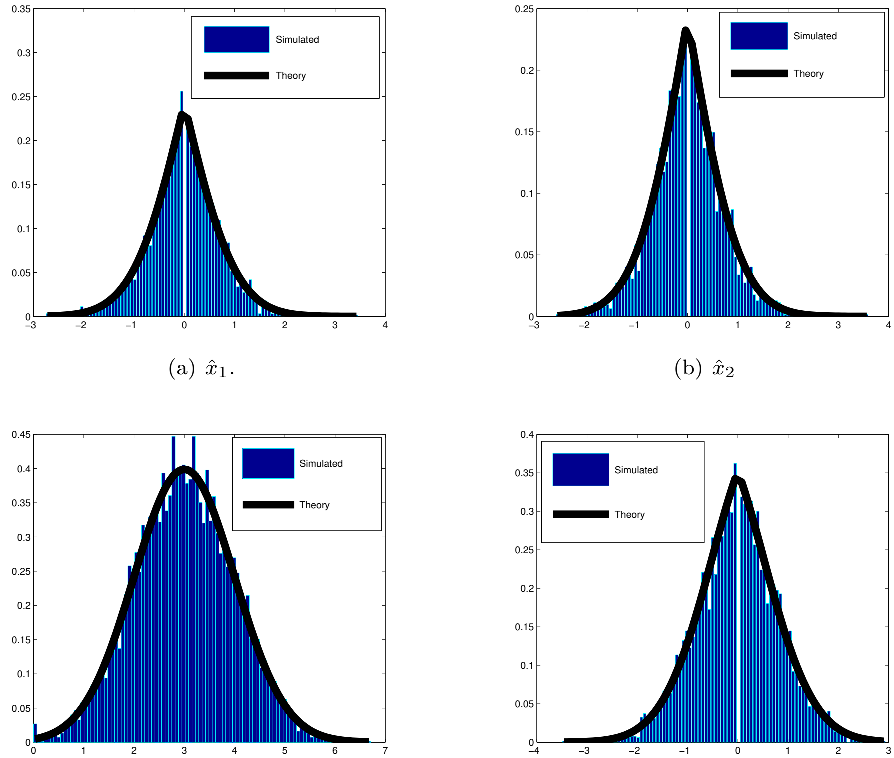
Now, we comment on the results obtained in the context of the well known oracle properties of the LASSO estimator.

We observe that the finite-sample pdf of the LASSO estimator derived in equations (12), (13) and (15) are multimodal and are characterized by the absolutely continuous normal distribution, whenever the marginals of the estimate are non-zero. For the orthogonal models, the pdf consists of a mixture of a singular normal distributions (Dirac measure), whenever the marginals of the estimate are zero. The large-sample behavior of the LASSO distributions are well studied in the literature and are characterized by the well-known oracle properties of the LASSO [12, 16]. The oracle properties of the LASSO depend on the choice of the regularization parameter, which can be chosen in the following two ways.

- In the first case, the regularization parameters are chosen such that the LASSO performs consistent model selection. However, LASSO does not satisfy the oracle property whenever the regularization parameter is chosen for consistent model selection [23, 25].

- In the second case, the regularization parameters are chosen for the LASSO performs conservative model selection. It is well known that the LASSO estimator satisfies the oracle property for the case of conservative model selection (or equivalently, the estimator's finite sample and the large sample pdfs match for the conservative selection).

We note that the pdfs derived in this section can be proven to be compatible with the oracle properties in a manner similar to the Theorems 5 and 10 of [23] by using the appropriate regularization parameters for consistent or conservative model selection, respectively.



(a) $\hat{x}_1$.

(b) $\hat{x}_2$

(c) $\hat{x}_3$

(d) $\hat{x}_4$

Fig. 1: pdf of the LASSO estimator for orthogonal model matrices ($\mathbf{W} = \mathbf{I}$).

4. NUMERICAL SIMULATIONS

In this section, we perform simulations to validate the theoretical results for cf. The simulations are performed using the linear regression model of (3). In the simulation setup, we choose a measurement vector, $\mathbf{b}$ of length $M = 4$. The model matrix $\mathbf{A}$ is chosen as a Hadamard matrix for orthogonal case and a random matrix of size $M \times N$ consisting of entries from Bernoulli distribution for non-orthogonal case. The columns of the model matrix $\mathbf{A}$ are always normalized to have unit norm. $N = 4$ for orthogonal and full rank models and $N = 8$ for singular model. The parameter vector $\mathbf{x}$ is chosen such that its sparsity $K = 1$ and its $\frac{N}{2} + 1$ entry is non-zero ($|x_3| = 4$ when $N = 4$ and $|x_5| = 8$ when $N = 8$). The noise is generated as a multivariate Gaussian random vector noise covariance variance matrix $\sigma^2\mathbf{I}$, where $\sigma = 1$. In the following, we use Monte-Carlo simulations for $L = 10000$ noisy realizations to obtain L noisy estimate vectors. We then run the LASSO algorithm using the MATLAB CVX package [13] to solve the optimization problem of (4) to obtain the LASSO estimate $\hat{\mathbf{x}}$ for each realization. We choose $\tau = 1$ for orthogonal and full rank models and $\tau = 2$ for the singular model. In

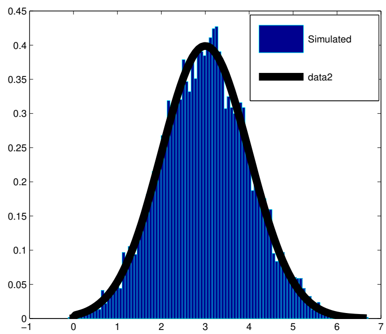


Fig. 2: pdf of the $\hat{z}_3$ for full rank W.

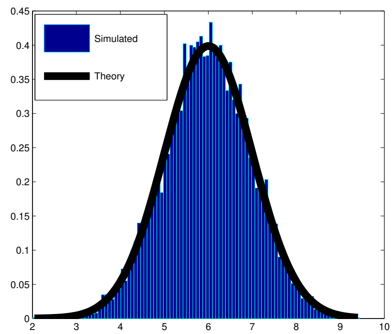


Fig. 3: pdf of $\hat{z}_5$ for singular $\mathbf{W}$.

Figure 1, we show the normalized histogram (normalized to make the total area one) of the components of the estimate vector $\hat{x}_k, k = 1, 2, 3, 4$ for the case of orthogonal models and compare it with the theoretical expression for the pdf obtained in (12). We observe that the estimate $\hat{x}_3$ for non-zero parameter x_3 has a normal distribution with mean $x_3 - \tau = 4 - 1 = 3$ as seen in Figure 1(c) and as explained in the second remark below Corollary 3.1 and as expected from (12). The other estimates $\hat{x}_k, k = 1, 2, 4$ follow the Dirac measure whenever $\hat{x}_k = 0$. This is indicated by the cut in the histogram plots at $\hat{x}_k = 0$. The non-zero values of $\hat{x}_k$ follow the normal pdf with mean -1 for negative $\hat{x}_k$ and positive $\hat{x}_k$ respectively as expected from (12).

In Figure 2, we show the normalized histogram (normalized to make the total area one) of the components of the estimate vector $\hat{z}_k, k = 3$ for the case of full rank model and compare it with the theoretical expression for the pdf obtained in (13). The simulated pdf does not match with (13) for the other components of $\hat{\mathbf{z}}$. We observe that the estimate $\hat{z}_3$ for the non-zero parameter x_3 has a normal distribution with mean $\pm(\tau - \mathbf{w}_3^T\mathbf{x})$ for negative and positive values $\hat{z}_3$ as seen in Figure 1(c) and as expected from (13). The pdf for the $\hat{z}_k, k = 1, 2, 4$ do not match the pdf of (13) as the approximation of the pdf expression is not valid for small values of $|\hat{x}|$.

In Figure 3, we show the normalized histogram (normalized to make the total area one) of the components of the estimate vector $\hat{z}_k, k = 3$ for the case of full rank model and compare it with the theoretical expression for the pdf obtained in (15). The simulated pdf does not match with (15) for the other components of $\hat{\mathbf{z}}$. Again, we observe that the estimate $\hat{z}_3$ for the non-zero parameter x_3 has a normal distribution with mean $\pm(\tau - \mathbf{w}_3^T\mathbf{x})$ for negative and positive values $\hat{z}_3$ as seen in Figure 1(c) and as expected from (15). The pdf for the $\hat{z}_k, k = 1, 2, 4$ do not match the pdf of (15) as the approximation of the pdf expression is not valid for small values of $|\hat{x}|$.

We observe from the figures that the theoretical pdfs follow the simulated pdfs closely for all the components in case of orthogonal models and for k corresponding to non-zero $\mathbf{x}$ in case of the other models.

Although, we have shown the simulations only for single source scenarios ($K = 1$), the same simulations work for multiple source ($K > 1$) scenarios and also in case of strong source-weak source scenarios, i. e. when $x_{\max} \ll x_{\min}$.

5. CONCLUDING REMARKS

In this work, we have derived the cf of the LASSO estimator with an aim to get some insight on the distributional properties of the LASSO estimator. The expression of the cf contains derived in (10) contains 2^N summations, and hence does not simplify the evaluation of the distribution of the LASSO estimator except for the special case of orthogonal model matrix ($\mathbf{W} = \mathbf{I}$). So we use the Fourier-Slice theorem to calculate the one-dimensional projections of a linear transformation of the LASSO estimator. The approximate pdf of the one dimensional components this linear transformation is then found from the projections by using the inversion theorem.

It is well-known from the Cramer–Wold theorem that the all the one dimensional projections of a distribution are sufficient for the evaluation of the overall cf [8] and hence the joint pdf of the LASSO estimator can be evaluated by its one dimensional projections, which is an interesting future work. As an application in statistical detection theory, the distribution of the estimator or (any of its functions, or test statistics) plays an important role to make decisions based on hypothesis. Hence, it may be an interesting future work to use the pdfs of the one-dimensional projections to propose test statistics for hypothesis testing. Many engineering applications like wireless communications and signal processing work with complex measurement models. Hence, a generalization of the distribution function for complex measurement models can be another interesting future work.

A. PROOFS

In this section, we present the proofs of theorems and corollaries discussed in Section 3.

A.1. Proof of Theorem 3.1

The solution to the problem in (4) can be obtained in a straightforward manner by using the KKT conditions which results in the following implicit relation for $\hat{\mathbf{x}}$.

$$\begin{aligned} \mathbf{W}\hat{\mathbf{x}} + \tau\boldsymbol{\gamma} &= \mathbf{A}^T\mathbf{b} = \tilde{\mathbf{b}} \\ \gamma_k &\in \begin{cases} \mathbb{S}(\hat{x}_k) & \text{if } \hat{x}_k \neq 0, \\ [-1,1] & \text{if } \hat{x}_k = 0 \end{cases} \end{aligned} \tag{17}$$

where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N] = \mathbf{A}^T\mathbf{A}$. Let $\mathcal{J} = \{1, 2, \ldots, N\}$ and $\mathcal{E} = \{k \in \mathcal{J} : |\gamma_k| = 1\}$, then any LASSO solution $\hat{\mathbf{x}}$ satisfies

$$\hat{\mathbf{x}}_{\mathcal{J}\setminus\mathcal{E}} = \mathbf{0};\ \mathbf{W}_{\mathcal{E}}\hat{\mathbf{x}}_{\mathcal{E}} + \tau\boldsymbol{\gamma}_{\mathcal{E}} = \mathbf{A}_{\mathcal{E}}^T\mathbf{b} = \breve{\mathbf{b}}. \tag{18}$$

Our aim is now to calculate the cf of $\hat{\mathbf{x}}$ using (18). Since, $\hat{\mathbf{x}}_{\mathcal{J}\setminus\mathcal{E}} = \mathbf{0}$, it does not contribute in the evaluation of the cf. Hence, it is enough to consider the first part of (17).

Now, we first evaluate the cf of right hand side (R.H.S) of (17). We observe that $\tilde{\mathbf{b}}$ is a multivariate Gaussian random variable with mean $\mathbf{W}\mathbf{x}$ and variance $\sigma^2\mathbf{W}$. Hence, the cf of $\tilde{\mathbf{b}}$, by definition is,

$$\mathbb{F}_{\tilde{\mathbf{b}}}(\mathbf{u}) = \exp\big(i\mathbf{u}^T\mathbf{W}\mathbf{x} - \frac{\sigma^2}{2}\mathbf{u}^T\mathbf{W}\mathbf{u}\big). \tag{19}$$

Now, we need to calculate the cf of the left hand side (L.H.S) of (17). We first define $c_j = \mathbf{u}^T\mathbf{w}_j$, $\hat{\mathbf{x}}^- = [\hat{x}_2, \hat{x}_3, \ldots, \hat{x}_N]$ and $\mathrm{d}\hat{\mathbf{x}}^- = \mathrm{d}\hat{x}_2\mathrm{d}\hat{x}_2\ldots\mathrm{d}\hat{x}_N$. We have,

$$
\begin{aligned}
&\mathbb{F}_{\mathbf{W}\hat{\mathbf{x}}+\tau\mathbb{S}(\hat{\mathbf{x}})}(\mathbf{u}) = \mathbb{E}_{\mathbf{W}\hat{\mathbf{x}}+\tau\mathbb{S}(\hat{\mathbf{x}})}\left\{\exp\left(i\mathbf{u}^T\left(\mathbf{W}\hat{\mathbf{x}}+\tau\mathbb{S}(\hat{\mathbf{x}})\right)\right)\right\}\\
=\ & \int\limits_{\mathbb{R}^N} f_{\mathbf{W}\hat{\mathbf{x}}+\tau\mathbb{S}(\hat{\mathbf{x}})}(\mathbf{W}\mathbf{t}+\tau\mathbb{S}(\mathbf{t}))\exp\left(i\mathbf{u}^T\left(\mathbf{W}\mathbf{t}+\tau\mathbb{S}(\mathbf{t})\right)\right)\mathrm{d}\mathbf{t}\\
=\ & \int\limits_{\mathbb{R}^N} f_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{t}+\tau\mathbb{S}(\mathbf{t}))\exp\left(i\mathbf{u}^T\left(\mathbf{W}\mathbf{t}+\tau\mathbb{S}(\mathbf{t})\right)\right)\mathrm{d}\mathbf{t}\\
=\ & \mathbb{E}_{\hat{\mathbf{x}}}\left\{\exp\left(i\sum_{j=1}^{N}\left(\hat{x}_j\mathbf{u}^T\mathbf{w}_j+\tau u_j\mathbb{S}(\hat{x}_j)\right)\right)\right\}\\
=\ & \int\limits_{\mathbb{R}^N} f_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})\exp\left(i\sum_{j=1}^{N}\left(\hat{x}_j\mathbf{u}^T\mathbf{w}_j+\tau u_j\mathbb{S}(\hat{x}_j)\right)\right)\mathrm{d}\hat{\mathbf{x}}\\
=\ & \int\limits_{\mathbb{R}^N} f_{\hat{\mathbf{x}}}(\hat{x}_1,\ldots,\hat{x}_N)\exp\left(i\sum_{j=1}^{N}\left(\hat{x}_j\mathbf{u}^T\mathbf{w}_j+\tau u_j\mathbb{S}(\hat{x}_j)\right)\right)\mathrm{d}\hat{\mathbf{x}}\\
=\ & \int\limits_{\mathbb{R}^N} f_{\hat{\mathbf{x}}}(\hat{x}_1,\ldots,\hat{x}_N)\exp\left(i\sum_{j=1}^{N}\left(\hat{x}_jc_j+\tau u_j\mathbb{S}(\hat{x}_j)\right)\right)\mathrm{d}\hat{\mathbf{x}}\\
=\ & \int\limits_{\mathbb{R}^{N-1}}\left(\int\limits_{\mathbb{R}} f_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})\exp(i(\hat{x}_1c_1+\tau u_1\mathbb{S}(\hat{x}_1)))\mathrm{d}\hat{x}_1\right)\exp\left(i\sum_{j=2}^{N}\left(\hat{x}_jc_j+\tau u_j\mathbb{S}(\hat{x}_j)\right)\right)\mathrm{d}\hat{\mathbf{x}}^-\\
=\ & \int\limits_{\mathbb{R}^{N-1}}\mathcal{J}(c_1,u_1,\hat{\mathbf{x}}^-)\exp\left(i\sum_{j=2}^{N}\left(\hat{x}_jc_j+\tau u_j\mathbb{S}(\hat{x}_j)\right)\right)\mathrm{d}\hat{\mathbf{x}}^-. \qquad (20)
\end{aligned}
$$

We first evaluate $\mathcal{J}(c_1,u_1,\hat{\mathbf{x}}^-)$. Defining $H(x)$ as the Heaviside step function, we have

$$
\begin{aligned}
\mathcal{J}(c_1,u_1,\hat{\mathbf{x}}^-) &= \int_{-\infty}^{\infty} f_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})\exp(i(\hat{x}_1c_1+\tau u_1\mathbb{S}(\hat{x}_1)))\mathrm{d}\hat{x}_1\\
&= \int_{-\infty}^{0} f_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})e^{i(\hat{x}_1c_1-\tau u_1)}\mathrm{d}\hat{x}_1 + \int_{0}^{\infty} f_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})e^{i(\hat{x}_1c_1+\tau u_1)}\mathrm{d}\hat{x}_1\\
&= e^{-i\tau u_1}\int_{-\infty}^{0} f_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})e^{i(\hat{x}_1c_1)}\mathrm{d}\hat{x}_1 + e^{i\tau u_1}\int_{0}^{\infty} f_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})e^{i(\hat{x}_1c_1)}\mathrm{d}\hat{x}_1\\
&= e^{-i\tau u_1}\int_{-\infty}^{\infty} f_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})H(-\hat{x}_1)e^{i(\hat{x}_1c_1)}\mathrm{d}\hat{x}_1 + e^{i\tau u_1}\int_{-\infty}^{\infty} f_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})H(\hat{x}_1)e^{i(\hat{x}_1c_1)}\mathrm{d}\hat{x}_1 \qquad (21)
\end{aligned}
$$

$$= e^{-i\tau u_1}\Big(\mathbb{G}_{\hat{x}_1}(c_1,\hat{\mathbf{x}}^-)\star\mathbb{F}_H(-c_1)\Big)+e^{i\tau u_1}\Big(\mathbb{G}_{\hat{x}_1}(c_1,\hat{\mathbf{x}}^-)\star\mathbb{F}_H(c_1)\Big) \tag{22}$$

$$= e^{-i\tau u_1}\left(\mathbb{G}_{\hat{x}_1}(c_1,\hat{\mathbf{x}}^-)\star\left(\frac{1}{2\pi j c_1}+\frac{\delta(c_1)}{2}\right)\right)$$

$$+e^{i\tau u_1}\left(\mathbb{G}_{\hat{x}_1}(c_1,\hat{\mathbf{x}}^-)\star\left(\frac{-1}{2\pi j c_1}+\frac{\delta(c_1)}{2}\right)\right) \tag{23}$$

$$= \mathbb{G}_{\hat{x}_1}(c_1,\hat{\mathbf{x}}^-)\cos(\tau u_1)+\left(\mathbb{G}_{\hat{x}_1}(c_1,\hat{\mathbf{x}}^-)\star\left(\frac{-1}{\pi c_1}\right)\right)\sin(\tau u_1)$$

$$= \mathbb{G}_{\hat{x}_1}(c_1,\hat{\mathbf{x}}^-)\cos(\tau u_1)+\mathbb{G}_{\hat{x}_1}(\overline{c}_1,\hat{\mathbf{x}}^-)\sin(\tau u_1). \tag{24}$$

Here, $\mathbb{G}_{\hat{x}_1}(c_1,\hat{\mathbf{x}}^-)$ denotes the one-dimensional Fourier transform along c_1 and $\mathbb{G}_{\hat{x}_1}(\overline{c}_1,\hat{\mathbf{x}}^-) = \left(\mathbb{G}_{\hat{x}_1}(c_1,\hat{\mathbf{x}}^-)\star\left(\frac{-1}{\pi c_1}\right)\right)$, denotes the one-dimensional Hilbert transform along c_1. Similarly, the Fourier and Hilbert transforms over each dimension is absorbed into the notation of the cf. We have used the Fourier-transform convolution theorem to obtain (22) from (21) and we have used the Fourier transform of the Heaviside step function in (23). Now, using the value of $\mathcal{J}(c_1,u_1,\hat{\mathbf{x}}^-)$ obtained from (24) in (20), we have

$$\begin{aligned}
\mathbb{F}_{\mathbf{W}\hat{\mathbf{x}}+\tau\mathbb{S}(\hat{\mathbf{x}})}(\mathbf{u}) &= \int_{\mathbb{R}^{N-1}}\mathcal{J}(c_1,u_1,\hat{\mathbf{x}}^-)\exp\left(i\sum_{j=2}^{N}\Big(\hat{x}_jc_j+\tau u_j\mathbb{S}(\hat{x}_j)\Big)\right)\mathrm{d}\hat{\mathbf{x}}^- \\
&= \cos(\tau u_1)\int_{\mathbb{R}^{N-1}}\mathbb{F}_{\hat{x}_1}(c_1,\hat{\mathbf{x}}^-)\exp\left(i\sum_{j=2}^{N}\Big(\hat{x}_jc_j+\tau u_j\mathbb{S}(\hat{x}_j)\Big)\right)\mathrm{d}\hat{\mathbf{x}}^- \\
&\quad+\sin(\tau u_1)\int_{\mathbb{R}^{N-1}}\mathbb{F}_{\hat{x}_1}(\overline{c}_1,\hat{\mathbf{x}}^-)\exp\left(i\sum_{j=2}^{N}\Big(\hat{x}_jc_j+\tau u_j\mathbb{S}(\hat{x}_j)\Big)\right)\mathrm{d}\hat{\mathbf{x}}^-.
\end{aligned} \tag{25}$$

Now, integrating over $\hat{x}_2,\hat{x}_3,\dots\hat{x}_N$ in the same manner, we obtain

$$\begin{aligned}
\mathbb{F}_{\mathbf{W}\hat{\mathbf{x}}+\tau\mathbb{S}(\hat{\mathbf{x}})}(\mathbf{u}) = \;& \cos(\tau u_1)\cos(\tau u_2)\dots\cos(\tau u_N)\mathbb{F}_{\hat{\mathbf{x}}}(c_1,\dots,c_N) \\
& +\sin(\tau u_1)\cos(\tau u_2)\dots\cos(\tau u_N)\mathbb{F}_{\hat{\mathbf{x}}}(\overline{c}_1,c_2,\dots,c_N) \\
& +\cos(\tau u_1)\sin(\tau u_2)\dots\cos(\tau u_N)\mathbb{F}_{\hat{\mathbf{x}}}(c_1,\overline{c}_2,\dots,c_N) \\
& \dots+(\text{All combinations of 1 sine term and } N-1 \text{ cosine terms}) \\
& \dots+\sin(\tau u_1)\sin(\tau u_2)\cos(\tau u_3)\dots\cos(\tau u_N)\mathbb{F}_{\hat{\mathbf{x}}}(\overline{c}_1,\overline{c}_2,c_3,\dots,c_N) \\
& \dots+\cos(\tau u_1)\sin(\tau u_2)\sin(\tau u_3)\dots\cos(\tau u_N)\mathbb{F}_{\hat{\mathbf{x}}}(c_1,\overline{c}_2,\overline{c}_3,c_4,\dots,c_N) \\
& \dots+(\text{All combinations of 2 sine terms and } N-2 \text{ cosine terms}) \\
& \vdots \\
& \dots+(\text{All combinations of } k \text{ sine terms and } N-k \text{ cosine terms}) \\
& \vdots \\
& \dots+\sin(\tau u_1)\dots+\sin(\tau u_N)\mathbb{F}_{\hat{\mathbf{x}}}(\overline{c}_1,\dots,\overline{c}_N)
\end{aligned} \tag{26}$$

which is equal to expression (10)

A.2. Proof of Corollary 3.1

If $\mathbf{W} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, or any diagonal matrix $\mathbf{D}$ (in general), then the estimates $\hat{x}_k, k = 1, 2, \ldots, N$ are all independent and hence the cf of the LASSO estimator is given by,

$$\prod_{k=1}^{N} \Bigg(\mathbb{F}_{\hat{x}_k}(u_k)\cos(\tau u_k) + \mathbb{F}_{\hat{x}_k}(\overline{u}_k)\sin(\tau u_k) \Bigg) = \prod_{k=1}^{N} \exp\big(iu_k x_k - \frac{\sigma^2}{2}u_k^2\big) \tag{27}$$

which can be expressed as,

$$\mathbb{F}_{\hat{x}_k}(u_k)\cos(\tau u_k) + \mathbb{F}_{\hat{x}_k}(\overline{u}_k)\sin(\tau u_k) = \exp\big(iu_k x_k - \frac{\sigma^2}{2}u_k^2\big), k = 1, 2, \ldots, N. \tag{28}$$

Hence, the marginal probability distribution functions of the individual components, $\hat{x}_k, k = 1, 2, \ldots, N$ is given by simply applying the inversion theorem to the (28) as,

$$\begin{aligned}
\frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(t_k - x_k)^2}{2\sigma^2}\big) &= \frac{1}{2} f_{\hat{x}_k}(t_k) \star \big[\delta(t_k - \tau) + \delta(t_k + \tau)\big] \\
&\quad + \frac{1}{2}\mathbb{S}(t_k) f_{\hat{x}_k}(t_k) \star \big[\delta(t_k - \tau) - \delta(t_k + \tau)\big] \\
&= \frac{1}{2}\big[f_{\hat{x}_k}(t_k - \tau) + f_{\hat{x}_k}(t_k + \tau)\big] + \frac{1}{2}\mathbb{S}(t_k - \tau) f_{\hat{x}_k}(t_k - \tau) \\
&\quad - \frac{1}{2}\mathbb{S}(t_k + \tau) f_{\hat{x}_k}(t_k + \tau)
\end{aligned} \tag{29}$$

which gives,

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(\hat{x}_k - x_k)^2}{2\sigma^2}\big) = \begin{cases} f_{\hat{x}_k}(\hat{x}_k - \tau) & \text{if } \hat{x}_k > \tau, \\ f_{\hat{x}_k}(\hat{x}_k + \tau) & \text{if } \hat{x}_k < -\tau. \end{cases} \tag{30}$$

Hence,

$$f_{\hat{x}_k}(t_k) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(t_k + \tau - x_k)^2}{2\sigma^2}\big) & \text{if } t_k > 0, \\ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(t_k - \tau - x_k)^2}{2\sigma^2}\big) & \text{if } t_k < 0. \end{cases} \tag{31}$$

A.3. Proof of Corollary 3.2

If $\mathbf{W}$ is any full rank matrix, we first perform slicing by substituting $\mathbf{u} = u\mathbf{e}_k$ in equation (26) to obtain

$$\mathbb{F}_{\mathbf{W}\hat{\mathbf{x}} + \tau\mathbb{S}(\hat{\mathbf{x}})}(u) = \cos(\tau u)\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u})] + \sin(\tau u)\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})]. \tag{32}$$

Now, we define $\hat{\mathbf{z}} = \mathbf{W}\hat{\mathbf{x}}$ and $\hat{\mathbf{x}} = \mathbf{H}\hat{\mathbf{z}}$, where $\mathbf{H} = \mathbf{W}^{-1}$. Let $\mathbf{h}_k$ be the kth column of $\mathbf{H}$, we have $f(\hat{\mathbf{z}}) = |\mathbf{W}|^{-1} f(\hat{\mathbf{x}})$ and $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u})] = \mathcal{S}_{\mathbf{u}} \odot \mathbf{W}^{-1} \odot \mathbb{F}_{\hat{\mathbf{x}}} = \mathcal{S}_{\mathbf{u}} \odot \mathbb{F}_{\hat{\mathbf{z}}} \odot \mathbf{W} = \mathbb{F}_{\hat{z}_k} \odot \mathcal{I}_{\hat{z}_k} \odot \mathbf{W} = \mathbb{F}_{\hat{z}_k}(u)$, where $\hat{z}_k$ is the kth component of $\hat{\mathbf{z}}$ and we obtain the last equation by applying the generalized Fourier slice theorem (Theorem 2.3). We evaluate the term, $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})]$ below by noting that $\frac{-1}{\pi u_k}$ is the Fourier transform of $i\mathbb{S}(\hat{x}_k)$,

$$\begin{aligned}
\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\tfrac{-1}{\pi c_k}) &= i \int\limits_{\mathbb{R}^N} f_{\hat{\mathbf{x}}}(\hat{\mathbf{x}}) \mathbb{S}(x_k) \exp(i\mathbf{u}^T \mathbf{W}\hat{\mathbf{x}})\, \mathrm{d}\hat{\mathbf{x}} \\
&= i \int\limits_{\mathbb{R}^N} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}}) \mathbb{S}(\mathbf{h}^T \hat{\mathbf{z}}) \exp(i\mathbf{u}^T \hat{\mathbf{z}}) \mathrm{d}\hat{\mathbf{z}} \qquad (33) \\
\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\tfrac{-1}{\pi c_k})] &= i \int\limits_{\mathbb{R}^N} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}}) \mathbb{S}(\mathbf{h}_k^T \hat{\mathbf{z}}) \exp(iu\hat{z}_k)\, \mathrm{d}\hat{\mathbf{z}}.
\end{aligned}$$

We now make an approximation that $\mathbb{S}(\mathbf{h}_k^T \hat{\mathbf{z}}) \approx \mathbb{S}(\hat{z}_k)$ for k corresponding to large $|\hat{x}_k|$ as explained in Theorem 3.2. Hence, the term $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})] \approx \mathbb{F}_{\hat{z}_k}(\overline{u})$. Substituting for $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u})]$ and the approximated expression of $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})]$ for k corresponding to large $|\hat{x}_k|$ and equating the above expression to the slice (w.r.t to u) of the cf of $\tilde{\mathbf{b}}$, we have

$$\cos(\tau u)\mathbb{F}_{\hat{z}_k}(u) + \sin(\tau u)\mathbb{F}_{\hat{z}_k}(\overline{u}) = \exp(-u^2 \frac{\sigma^2}{2} w_{kk}) \exp\Big(ju(\mathbf{w}_k^T \mathbf{x})\Big). \qquad (34)$$

Hence, the marginal pdf of the individual components, $\hat{z}_k$ for k corresponding to large $|\hat{x}_k|$ is given by simply applying the inversion theorem to the (34) as,

$$f_{\hat{z}_k}(t_k) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2 w_{kk}}} \exp\big(-\frac{(t_k+\tau-\mathbf{w}_k^T\mathbf{x})^2}{2\sigma^2 w_{kk}}\big) & \text{if } t_k > 0, \\ \frac{1}{\sqrt{2\pi\sigma^2 w_{kk}}} \exp\big(-\frac{(t_k-\tau-\mathbf{w}_k^T\mathbf{x})^2}{2\sigma^2 w_{kk}}\big) & \text{if } t_k < 0. \end{cases} \qquad (35)$$

A.4. Proof of Corollary 3.3

For any general $\mathbf{W}$, we again perform slicing by substituting $\mathbf{u} = u\mathbf{e}_k$ in equation (26) to obtain (32). Again, we define $\hat{\mathbf{z}} = \mathbf{W}\hat{\mathbf{x}}$, $\hat{\mathbf{x}} = \mathbf{H}\hat{\mathbf{z}}$, where $\mathbf{H}$ is now $\mathbf{H} = \mathbf{W}^\dagger$ and let $\mathbf{h}_k$ be the kth column of $\mathbf{H}$.

Now, as in Proof A.3, we have $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u})] = \mathcal{S}_{\mathbf{u}} \odot \mathbb{F}_{\hat{\mathbf{z}}} \odot \mathbf{W} = \mathbb{F}_{\hat{z}_k} \odot \mathcal{I}_{\hat{z}_k} \odot \mathbf{W} = \mathbb{F}_{\hat{z}_k}(u)$, where $\hat{z}_k$ is the kth element of $\hat{\mathbf{z}}$ and we obtain the last equation by applying the generalized Fourier slice theorem (Theorem 2.3). The term, $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})]$ is again equal to,

$$\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})] = i \int\limits_{\mathbb{R}^N} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}}) \mathbb{S}(\mathbf{h}_k^T \hat{\mathbf{z}}) \exp(ju\hat{z}_k)\, \mathrm{d}\hat{\mathbf{z}}. \qquad (36)$$

We again make the approximation $\mathbb{S}(\mathbf{h}_k^T \hat{\mathbf{z}}) \approx \mathbb{S}(\hat{z}_k)$ for k corresponding to large $|\hat{x}_k|$. Substituting for $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u})]$ and the approximated expression of $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})]$ for k corresponding to large $|\hat{x}_k|$ in equation (19) and equating the above expression to the slice (w.r.t to u) of the cf of $\tilde{\mathbf{b}}$, we have

$$\cos(\tau u)\mathbb{F}_{\hat{z}_k}(u) + \sin(\tau u)\mathbb{F}_{\hat{z}_k}(\overline{u}) = \exp(-u^2 \frac{\sigma^2}{2} w_{kk}) \exp\Big(ju(\mathbf{w}_k^T \mathbf{x})\Big). \qquad (37)$$

Hence, the marginal pdf of the individual components, $\hat{z}_k$ for k corresponding to large $|\hat{x}_k|$ is given by simply applying the inversion theorem to the (37) as,

$$f_{\hat{z}_k}(t_k) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2 w_{kk}}} \exp\big(-\frac{(t_k+\tau-\mathbf{w}_k^T\mathbf{x})^2}{2\sigma^2 w_{kk}}\big) & \text{if } t_k > 0, \\ \frac{1}{\sqrt{2\pi\sigma^2 w_{kk}}} \exp\big(-\frac{(t_k-\tau-\mathbf{w}_k^T\mathbf{x})^2}{2\sigma^2 w_{kk}}\big) & \text{if } t_k < 0. \end{cases} \tag{38}$$

A.5. Evaluation of $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})]$

In this section, we evaluate $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})]$ with the assumption that $\hat{\mathbf{z}}$ has a multivariate Gaussian distribution of with mean $\mathbf{m}$ and co-variance matrix $\mathbf{R}$. We use $\phi(\hat{\mathbf{z}}_N, \mathbf{m}_N, \mathbf{R}_N)$ to denote that the random vector $\hat{\mathbf{z}}_N$ of length N has a multivariate Gaussian distribution with mean vector $\mathbf{m}_N$ of length N and co-variance matrix $\mathbf{R}_N$ of size $N \times N$ and Φ is used to denote the cumulative distribution function (cdf) of the normal distribution. Let $\mathbf{H} = \mathbf{W}^\dagger$ and $\mathbf{h}_k$ as the kth column of $\mathbf{H}$. We have,

$$\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})] = i \int_{\mathbb{R}^N} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}}) \exp(ju\hat{z}_k)\mathbb{S}(\mathbf{h}_k^T\hat{\mathbf{z}})\, \mathrm{d}\hat{\mathbf{z}}. \tag{39}$$

Without loss of generality, we choose $k = 1$. We partition $\mathbf{m}_N$, $\hat{\mathbf{z}}_N$ and $\mathbf{R}_N$ as $\mathbf{m} = [\mathbf{m}_{N-1}, m_N]^T$, $\hat{\mathbf{z}} = [\hat{\mathbf{z}}_{N-1}, \hat{z}_N]^T$, $\mathbf{R} = \begin{bmatrix} \mathbf{R}_{N-1} & \mathbf{r}_N \\ \mathbf{r}_N^T & r_{NN} \end{bmatrix}$ respectively. Let $\mathbf{s} = \frac{-1}{h_{1N}}[h_{11}, h_{12}, \ldots, h_{1N}]^T$, then $\mathbb{S}(\mathbf{h}_1^T\hat{\mathbf{z}}) = \mp 1$ depending on $\hat{z}_N \lessgtr \mathbf{s}^T\hat{\mathbf{z}}_{N-1}$. We have,

$$\begin{aligned} \mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\tfrac{-1}{\pi c_k})] &= i \int_{\mathbb{R}^N} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}}) \exp(ju\hat{z}_1)\mathbb{S}(\mathbf{h}_1^T\hat{\mathbf{z}})\, \mathrm{d}\hat{\mathbf{z}} \\ &= i \int_{\mathbb{R}^{N-1}} \Bigg(- \int_{-\infty}^{\mathbf{s}^T\hat{\mathbf{z}}_{N-1}} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}})\mathrm{d}\hat{z}_N + \int_{\mathbf{s}^T\hat{\mathbf{z}}_{N-1}}^{\infty} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}})\mathrm{d}\hat{z}_N \Bigg) \exp(ju\hat{z}_1)\, \mathrm{d}\mathbf{z}_{N-1} \\ &= i \int_{\mathbb{R}^{N-1}} \Bigg(- 2 \int_{-\infty}^{\mathbf{s}^T\hat{\mathbf{z}}_{N-1}} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}})\mathrm{d}\hat{z}_N + \int_{-\infty}^{\infty} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}})\mathrm{d}\hat{z}_N \Bigg) \exp(ju\hat{z}_1)\, \mathrm{d}\hat{\mathbf{z}}_{N-1} \\ &= -2i \int_{\mathbb{R}^{N-1}} \Bigg(\int_{-\infty}^{\mathbf{s}^T\hat{\mathbf{z}}_{N-1}} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}}_{N-1}, \hat{z}_N)\, \mathrm{d}\hat{z}_N \Bigg) \exp(ju\hat{z}_1)\, \mathrm{d}\hat{\mathbf{z}}_{N-1} + i\mathbb{F}_{\hat{z}_1}(u). \end{aligned} \tag{40}$$

We now evaluate the first integral below.

$$J = -2 \int_{\mathbb{R}^{N-1}} \Bigg(\int_{-\infty}^{\mathbf{s}^T\hat{\mathbf{z}}_{N-1}} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}}_{N-1}, \hat{z}_N)\, \mathrm{d}\hat{z}_N \Bigg) \exp(ju\hat{z}_1)\, \mathrm{d}\hat{\mathbf{z}}_{N-1}. \tag{41}$$

Let us consider the inner integral, defining $p_N = m_N + \mathbf{r}_N^T \mathbf{R}_{N-1}^{\dagger}(\hat{\mathbf{z}}_{N-1} - \mathbf{m}_{N-1})$ and $q_N = r_{NN} - \mathbf{r}_N^T \mathbf{R}_{N-1}^{\dagger} \mathbf{r}_N$ and using the fact that $f(\hat{\mathbf{z}})$ is multivariate Gaussian, we have

$$\begin{aligned}\int_{-\infty}^{\mathbf{s}^T\hat{\mathbf{z}}_{N-1}} f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}}_{N-1}, \hat{z}_N)\,\mathrm{d}\hat{z}_N &= \int_{-\infty}^{\mathbf{s}^T\hat{\mathbf{z}}_{N-1}} \phi(\hat{\mathbf{z}}_{N-1}, \mathbf{m}_{N-1}, \mathbf{R}_{N-1}) N(\hat{z}_N, p_N, q_N)\mathrm{d}\hat{z}_N \\ &= \phi(\hat{\mathbf{z}}_{N-1}, \mathbf{m}_{N-1}, \mathbf{R}_{N-1})\Phi\left(\frac{\mathbf{s}^T\hat{\mathbf{z}}_{N-1} - p_N}{\sqrt{q_N}}\right) \qquad (42) \\ &= \phi(\hat{\mathbf{z}}_{N-1}, \mathbf{m}_{N-1}, \mathbf{R}_{N-1})\Phi\left(\frac{\mathbf{g}_{N-1}^T\hat{\mathbf{z}}_{N-1} - k_{N-1}}{\sqrt{q_N}}\right)\end{aligned}$$

where $\mathbf{g}_{N-1}^T = \mathbf{s}^T - \mathbf{r}_N^T \mathbf{R}_{N-1}^{\dagger}$ and $k_{N-1} = \mathbf{r}_N^T \mathbf{R}_{N-1}^{\dagger}\mathbf{m}_{N-1} - m_N$. We partition $\mathbf{g}_{N-1}$ as $g_{N-1} = [\mathbf{g}_{N-2}, g_N]^T$. Now, substituting the inner integral in (42), we have

$$J = -2\int_{\mathbb{R}^{N-1}} \phi(\hat{\mathbf{z}}_{N-1}, \mathbf{m}_{N-1}, \mathbf{R}_{N-1})\Phi\left(\frac{\mathbf{g}_{N-1}^T\hat{\mathbf{z}}_{N-1} - k_{N-1}}{\sqrt{q_N}}\right)\exp(ju\hat{z}_1)\mathrm{d}\hat{\mathbf{z}}_{N-1} \qquad (43a)$$

$$= -2\int_{\mathbb{R}^{N-1}} \phi(\hat{\mathbf{z}}_{N-2}, \mathbf{m}_{N-2}, \mathbf{R}_{N-2})\phi(\hat{z}_{N-1}, p_{N-1}, q_{N-1}) \Phi\left(\frac{g_{N-1}\hat{z}_{N-1} + \mathbf{g}_{N-2}^T\hat{\mathbf{z}}_{N-2} - k_{N-1}}{\sqrt{q_N}}\right)\exp(ju\hat{z}_1)\mathrm{d}\hat{\mathbf{z}}_{N-1}$$

$$= -2\int_{\mathbb{R}^{N-2}} \Upsilon(\hat{\mathbf{z}}_{N-2})\left(\int_{-\infty}^{\infty}\phi(\hat{z}_{N-1}, p_{N-1}, q_{N-1})\Phi\left(\frac{g_{N-1}\hat{z}_{N-1} + \mathbf{g}_{N-2}^T\hat{\mathbf{z}}_{N-2} - k_{N-1}}{\sqrt{q_N}}\right)\mathrm{d}\hat{z}_{N-1}\right)\mathrm{d}\hat{\mathbf{z}}_{N-2}$$

where $\Upsilon(\hat{\mathbf{z}}_{N-2}) = \phi(\hat{\mathbf{z}}_{N-2}, \mathbf{m}_{N-2}, \mathbf{R}_{N-2})\exp(u\hat{z}_1)$, $p_{N-1} = m_{N-1} + \mathbf{r}_{N-1}^T \mathbf{R}_{N-2}^{\dagger}(\hat{\mathbf{z}}_{N-2} - \mathbf{m}_{N-2})$ and $q_{N-1} = r_{N-1N-1} - \mathbf{r}_{N-1}^T \mathbf{R}_{N-2}^{\dagger}\mathbf{r}_{N-1}$. To evaluate the inner integral, we make the transformation $\hat{t}_{N-1} = \frac{\hat{z}_{N-1} - p_{N-1}}{\sqrt{q_{N-1}}}$ to make it a standard integral, so we have

$$J = -2\int_{\mathbb{R}^{N-2}} \Upsilon(\hat{\mathbf{z}}_{N-2})\Phi\left(\frac{\alpha_{N-2}}{\sqrt{1+\beta_{N-1}^2}}\right)\mathrm{d}\hat{\mathbf{z}}_{N-2} \qquad (44)$$

where $\beta_{N-1} = g_{N-1}\sqrt{\frac{q_{N-1}}{q_N}}$ and $\alpha_{N-2} = \frac{\mathbf{g}_{N-2}^T\hat{\mathbf{z}}_{N-2} + g_{N-1}p_{N-1} - k_{N-1}}{\sqrt{q_N}}$. Similarly evaluating the integral $N-3$ times, we have

$$J = -2\int_{-\infty}^{\infty} \phi(\hat{z}_1, m_1, r_{11})\Phi(\beta_1\hat{z}_1 + \alpha_1)\exp(ju\hat{z}_1)\mathrm{d}\hat{z}_1 \qquad (45)$$

which is equal to the Fourier transform of the function $\hat{f}(\hat{z}_1) = -2f(\hat{z}_1)\Phi(\beta_1\hat{z}_1 + \alpha_1)$, where α_1 and β_1 are related to the entries of $\mathbf{s}$, $\mathbf{m}$ and $\mathbf{R}$. Hence, $\mathcal{S}_{\mathbf{u}}[\mathbb{F}_{\hat{\mathbf{x}}}(\mathbf{W}\mathbf{u}) \star (\frac{-1}{\pi c_k})]$

is equal to the Fourier transform of $f(\hat{z}_1) + \hat{f}(\hat{z}_1) = f(\hat{z}_1)[1 - 2\Phi(\beta_1 \hat{z}_1 + \alpha_1)]$. So, we can observe from that when $\hat{z}_k$ is large and positive, then $\Phi(\beta \hat{z}_k + \alpha_k)$ tends to zero and when $\hat{z}_k$ is large and negative, $\Phi(\beta \hat{z}_k + \alpha_k)$ tends to one. So $(f(\hat{z}_k) + \hat{f}(\hat{z}_k)) \approx f(\hat{z}_k)\mathbb{S}(\hat{z}_k)$ for large $|\hat{z}_k|$, which justifies the use of this approximation in Corollaries 3.2 and 3.3.

ACKNOWLEDGMENT

We are thankful to the referees for their many helpful comments and suggestions which certainly improved the presentation of the paper.

(Received December 1, 2016)

REFERENCES

[1] C. D. Austin, R. Moses, J. Ash, and E. Ertin: On the relation between sparse reconstruction and parameter estimation with model order selection. IEEE J. Selected Topics Signal Process. *4* (2010), 3, 560–570. DOI:10.1109/jstsp.2009.2038313

[2] S. Babacan, R. Molina, and A. Katsaggelos: Bayesian compressive sensing using laplace priors. IEEE Trans. Image Process. *19* (2010), 1, 53–63. DOI:10.1109/tip.2009.2032894

[3] R. Baraniuk, E. Candes, R. Nowak, R., M. and Vetterli: Compressive sampling. IEEE Signal Processing Magazine *25* (2008), 2, 12–13. DOI:10.1109/msp.2008.915557

[4] A. Ben-Tal and A. S. Nemirovskiaei: Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. Society for Industrial and Applied Mathematics, Philadelphia 2001. DOI:10.1137/1.9780898718829

[5] P. Boufounos, M. F. Duarte, and R. G. Baraniuk: Sparse signal reconstruction from noisy compressive measurements using cross validation. In: IEEE/SP 14th Workshop on Statistical Signal Processing, 2007, pp. 299–303. DOI:10.1109/ssp.2007.4301267

[6] E. Candes: The restricted isometry property and its implications for compressed sensing. Comptes Rendus Mathematique *346* (2008), 9–10, 589–592. DOI:10.1016/j.crma.2008.03.014

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders: Atomic decomposition by basis pursuit. SIAM Rev. *43* (2001), 1, 129–159. DOI:10.1137/s003614450037906x

[8] J. A. Cuesta-Albertos, R. Fraiman, and T. Ransford: A sharp form of the Cramér–wold theorem. J. Theoret. Probab. *20* (2007), 2, 201–209. DOI:10.1007/s10959-007-0060-7

[9] D. Donoho: Compressed sensing. IEEE Trans. Inform. Theory *52* (2006), 4, 1289–1306. DOI:10.1109/tit.2006.871582

[10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani: Least angle regression. Ann. Statist. *32* (2004), 407–499. DOI:10.1214/009053604000000067

[11] Y. Eldar: Generalized sure for exponential families: Applications to regularization. IEEE Trans. Signal Process. *57* (2009), 2, 471–481. DOI:10.1109/tsp.2008.2008212

[12] J. Fan and R. Li: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. *96* (2001), 456, 1348–1360. DOI:10.1198/016214501753382273

[13] M. Grant and S. Boyd: CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, 2014.

[14] P. Kabaila: The effect of model selection on confidence regions and prediction regions. Econometr. Theory *11* (1995), 537–549. DOI:10.1017/s0266466600009403

[15] S. M. Kay: Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice-Hall, Inc., Upper Saddle River, NJ 1993.

[16] K. Knight and W. Fu: Asymptotics for lasso-type estimators. Ann. Statist. *28* (2000), 5, 1356–1378. DOI:10.1214/aos/1015957397

[17] H. Krim and M. Viberg: Two decades of array signal processing research: the parametric approach. IEEE Signal Processing Magazine *13* (1996), 4, 67–94. DOI:10.1109/79.526899

[18] H. Leeb and B. M. Pötscher: Model selection and inference: Facts and fiction. Econometr. Theory *21* (2005), 21–59. DOI:10.1017/s0266466605050036

[19] R. Lockhart, J. Taylor, R. J. Tibshirani, and R. Tibshirani: A significance test for the lasso. Ann. Statist. *42* (2014), 2, 413–468. DOI:10.1214/13-aos1175

[20] M. E. Lopes: Estimating unknown sparsity in compressed sensing. CoRR 2012, abs/1204.4227.

[21] R. Ng: Fourier slice photography. ACM Trans. Graph. *24* (2005), 3, 735–744.

[22] A. Panahi and M. Viberg: Fast candidate points selection in the lasso path. IEEE Signal Process. Lett. *19* (2012), 2, 79–82. DOI:10.1109/lsp.2011.2179534

[23] B. M. Pötscher and H. Leeb: On the distribution of penalized maximum likelihood estimators: The lasso, scad, and thresholding. J. Multivar. Anal. *100* (2009), 9, 2065–2082. DOI:10.1016/j.jmva.2009.06.010

[24] R. Tibshirani: Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc., Ser. B *58* (1994), 267–288.

[25] H. Zou: The adaptive lasso and its oracle properties. J. Amer. Statist. Assoc. *101* (2006), 476, 1418–1429. DOI:10.1198/016214506000000735

Rakshith Jagannath, Indian Institute of Technology, Madras, Sardar Patel Road, Adyar, Chennai. Tamil Nadu-600036.
e-mail: rejarr@gmail.com

Neelesh S. Upadhye, Indian Institute of Technology, Madras, Sardar Patel Road, Adyar, Chennai. Tamil Nadu-600036.
e-mail: neelesh@iitm.ac.in