FINITE SAMPLE PERFORMANCE OF LEAST SQUARES ESTIMATION IN SUB-GAUSSIAN NOISE

Michael Krikheli *

Department of Mathematics,
Bar-Ilan University 52900,
Ramat-Gan Israel

Amir Leshem

Faculty of Engineering,
Bar-Ilan University 52900,
Ramat-Gan Israel

ABSTRACT

In this paper we analyze the finite sample performance of the least squares estimator. In contrast to standard performance analysis which uses bounds on the mean square error together with asymptotic normality, our bounds are based on large deviation and concentration of measure results. This allows for accurate bounds on the tail of the estimator. We show the fast exponential convergence of the number of samples required to ensure accuracy with high probability. We analyze a sub-Gaussian setting with fixed or random mixing matrix of the least squares problem. We provide probability tail bounds on the L infinity norm of the error of the finite sample approximation of the true parameter. Our method is simple and uses simple analysis for L infinity type bounds of the estimation error. The tightness of the bound is studied through simulations.

Index Terms— Estimation; least squares; non Gaussian; finite sample; large deviations; confidence bounds

1. INTRODUCTION

Since least squares estimation is generally consistent, it is used in many fields. Image processing applications such as soft-decision interpolation were proposed in [1] and [2]. The former used linear least squares and the latter used weighted linear least squares estimation [3]. Weighted linear least squares is also used in the field of medical imaging [4]. In the field of security research least squares estimation was used in [5] to estimate user profiles in a network. The field of communications uses least squares estimation for molecular communication, for example [6]. Other implementations in communications include least squares estimation in mobile orthogonal frequency-division multiplexing for rapidly time-varying channels [7]. Another use for least squares estimation in signal processing is for source localization [3]. The standard analysis of estimation problems employs the Cramer-Rao bound (CRB), as in [8] and [9]. While the theory of least squares estimation is well understood in the Gaussian setting, the non-Gaussian setting is much more challenging. For example, in many applications a Gaussian mixture rather than the Gaussian noise model is used. For instance, in [10] a Gaussian mixture model of a time-varying autoregressive process was assumed and analyzed. The Gaussian mixture model was used to model noise in underwater communication systems in [11]. Wiener filters in Gaussian mixture signal estimation was analyzed in [12]. In [13] a likelihood based algorithm for Gaussian mixture noise was devised and analyzed in the terms of the CRB. In [14] a robust detection technique using Maximum-Likelihood estimation was proposed for an impulsive noise modeled as a Gaussian mixture. In this paper we consider sub-Gaussian noise, which is a general non-Gaussian noise framework. Here, Gaussian mixture model is sub-Gaussian and our results are valid for this model. Since least squares estimation in non-Gaussian settings has many applications it is interesting to achieve finite sample performance results. Finite sample performance results such as these can be used to analyze massive MIMO channels [15].

There have been numerous asymptotic studies on the least squares problem for both the linear and the non-linear setting. The strong consistency of the estimator was examined for example in [16]. An exponential convergence rate for linear and non-linear was shown in [17] for independent random variables and a polynomial convergence rate was described in [18] for dependent random variables. Asymptotic normality and consistency were also studied in [19]. The law of large deviation results for non-linear least squares were given in [20] and [21]. These results were asymptotic in nature.

One approach to performance analysis of the least squares estimator is to use the asymptotic normality of the estimator. However, this is insufficient for large deviation bounds since by the Berry-Esseen theorem [22], [23] [24] the error of the normal approximation decays as $\frac{1}{\sqrt{N}}$.

In the past few years, the finite sample behavior of regression problems have been studied in [25–28]. An optimal convergence rate analysis was given in [29]. The bounds derived in this work are tighter on the one hand and simpler to derive on the other. In contrast to [25–29] our results also provide box confidence bounds using the L^∞ norm.

The structure of the paper is as follows: in section 2 we formulate the problem. In section 3 we formulate and prove the main theorem. In section 4 we present an extension of the main theorem in the case of a random mixing matrix. We show simulation results in 5 and conclude the results in 6.

2. PROBLEM FORMULATION

Consider a linear model with additive noise

$$\mathbf{X} = \mathbf{A}\theta_0 + \mathbf{V}, \tag{1}$$

where, $\mathbf{X} \in \mathbb{R}^{N\times 1}$ is our output, $\mathbf{A} \in \mathbb{R}^{N\times p}$ is the mixing matrix, $\theta_0 \in \mathbb{R}^p$ is the estimated parameter and $\mathbf{V} \in \mathbb{R}^{N\times 1}$ is a noise vector with independent and sub-Gaussian elements. We can write $\mathbf{A} = [\mathbf{a}_1^T, \dots, \mathbf{a}_N^T]^T$ where $\mathbf{a}_i \in \mathbb{R}^{p\times 1}$.[1]

Many real world noise models are sub-Gaussian. These include

*This work is part of the first author's M.Sc. thesis. The work of Amir Leshem is partially supported by ISF grant 903/2013.

[1] For simplicity we only consider the real case. The complex case is similar with minor modifications.

978-1-4673-7802-4/16/$31.00 ©2016 IEEE

Gaussian distributions, finite Gaussian mixtures, all kinds of bounded variables, and any combination of these.
The least squares estimator is given by solving the problem

$$\hat{\theta}^N = \arg\min J^N(\theta, \mathbf{X}), \tag{2}$$

where,

$$J^N(\theta, \mathbf{X}) = (\mathbf{X} - \mathbf{A}\theta)^T(\mathbf{X} - \mathbf{A}\theta) \tag{3}$$

The minimum for J^N is given by [16]

$$\hat{\theta}^N = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{X}. \tag{4}$$

In zero mean noise the least squares estimator is unbiased.

We want to study the tail distribution of $\left\|\hat{\theta}^N - \theta_0\right\|_\infty$ or more specifically obtain bounds of the form

$$P\left(\left\|\hat{\theta}^N - \theta_0\right\|_\infty > r\right) < \varepsilon \tag{5}$$

as a function of N, r and ε. We analyze two interesting cases. The first is a constant mixing matrix $\mathbf{A}$ with dimensions $N \times p$. The second is a random matrix $\mathbf{A}$ with bounded elements.

Definition 2.1. Throughout the paper we use the following mathematical notations:

1. The logarithmic moment generating function is: $\Lambda_{\mathbf{X}}(s) := \log E[e^{s\mathbf{X}}]$.
2. Let $\mathbf{A}$ be a matrix. $\lambda_{max}(\mathbf{A})$ and $\lambda_{min}(\mathbf{A})$ are the maximal and minimal eigenvalues of $\mathbf{A}$ respectively.
3. The spectral norm for matrices is given by $\|\mathbf{A}\| \doteq \sqrt{\lambda_{max}(\mathbf{A^T A})}$.

We make the following assumptions on our model.

A1: $E(v_n) = 0 \quad \forall 1 \le n \le N$.

A2: $\operatorname{rank}\left(\mathbf{A}^T\mathbf{A}\right) = p$.

A3: $|a_{ni}| \le \alpha \quad \forall n = 1 \dots N \quad \forall i = 1 \dots p$.

A4: $\mathbf{A}$ is chosen from a family of matrices G which is the set of all matrices $X \in \mathbb{R}^{N \times p}$ such that $\exists \mu_1, \mu_2 \ge 0$ that satisfy $\mu_1 N \le \lambda_{min}\left(\mathbf{X}^T\mathbf{X}\right) \le \mu_2 N$.

A5: $\forall 1 \le n \le N \quad v_n$ are independent sub-Gaussian random variables with parameter R.

Assumptions A1-A2 are standard in least squares theory. Assumption A3 is mild and is achievable through normalizing each row of the mixing matrix a_n with proper scaling of the sub-Gaussian parameter. Assumption A4 holds for any random i.i.d matrix. Assumption A5 is the sub-Gaussian model assumption.

3. MAIN RESULT

In this section we formulate and prove the main result of this paper. The main theorem provides the number of samples needed so that the maximal error in any dimension between the true parameter and the finite sample solution will be at most r with a probability of at least $1 - \varepsilon$.

Theorem 3.1. (Main theorem)
Let $\mathbf{X}$ be defined as in 1. Assume that our model admits assumptions A1 - A5. Let $\hat{\theta}^N$ be the least squares estimator and θ_0 the true parameter. Furthermore, let $\varepsilon > 0$ and $r > 0$ be given; then there exists $N(r, \varepsilon)$ such that $\forall N > N(r, \varepsilon)$

$$P\left(\left\|\hat{\theta}^N - \theta_0\right\|_\infty > r\right) < \varepsilon, \tag{6}$$

Where

$$N(r, \varepsilon) = \max\{N_1(r), N_2(r, \varepsilon), N_3(r, \varepsilon)\} \tag{7}$$

and

$$N_1(r) = \frac{\alpha^2 R^2}{\mu_1^2 r^2}, \tag{8}$$

$$N_2(r, \varepsilon) = \inf_{0 < s < \frac{1}{2\alpha^2 R^2}} \left\{\frac{1}{\mu_1^2 r^2 s}\left(2\beta(s) + \mu_2 r\sqrt{2s\log\frac{2p}{\varepsilon}}\right)\right\}, \tag{9}$$

$$N_3(r, \varepsilon) \le \inf_{0 < s < \frac{1}{\alpha^2 R^2}} \left\{\sqrt{\frac{\log\frac{2p}{\varepsilon}}{\frac{\mu_1^2 r^2}{2} - \gamma(s)}}\right\}, \tag{10}$$

$$\beta(s) \doteq R^2\alpha^2 s + \frac{R^4\alpha^4 s^2}{1 - 2R^2\alpha^2 s}, \tag{11}$$

$$\gamma(s) \doteq \frac{R^4\alpha^4 s^2}{2} + \frac{R^8\alpha^8 s^4}{4(1 - \alpha^4 s^2 R^4)}. \tag{12}$$

3.1. Discussion

The importance of this result is that it provides very sharp bounds on the finite sample performance of linear least squares problems as well as a very simple proof for a very general noise model. The result shows that in the fixed mixing matrix setting $N \sim O\left(\frac{1}{r^2}\sqrt{\log\frac{1}{\varepsilon}}\right)$.
In section 4 we show that in the random mixing matrix setting $N \sim O\left(\frac{1}{r^2}\log\frac{1}{\varepsilon}\right)$.
This theorem gives results in the L^∞ norm. Using the facts that $\|\mathbf{x}\|_2 \le \sqrt{p}\|\mathbf{x}\|_\infty$, $\|\mathbf{x}\|_1 \le p\|\mathbf{x}\|_\infty$ we can achieve L^2, L^1 norm bounds as well. These will be the same bounds multiplied by p and p^2 respectively.

3.2. Concentration result for the least squares problem

In order to prove the main theorem, we begin by formulating and proving a concentration result for the linear least squares model.

Lemma 3.2. *Let $\mathbf{x}$ be defined as in (1). Assume that v_n has a moment generating function for each n and assumptions A1-A4. Furthermore, let $\hat{\theta}^N$ be defined in (4). Let $\varepsilon > 0$ and $r > 0$; then there exists $N(r, \varepsilon)$ such that $\forall N > N(r, \varepsilon)$ such that for all $1 \le i \le p$*

$$P\left(\left|\left(\hat{\theta}^N - \theta_0\right)_i\right| > r\right) \le \varepsilon \tag{13}$$

where

$$N(r, \varepsilon) = \max\{N_1(r), N_2(r, \varepsilon), N_3(r, \varepsilon)\}, \tag{14}$$

$$N_1(r) = \frac{\alpha^2 E(v^2)}{\mu_1^2 r^2}, \tag{15}$$

$$N_2(r,\varepsilon) = \inf_{s>0} \left\{ \frac{1}{\mu_1^2 r^2 s} \left(2\Lambda_{v^2}(\alpha^2 s) + \mu_2 r \sqrt{2s \log \frac{2}{\varepsilon}}\right)\right\}, \tag{16}$$

$$N_3 = \inf_{s>0} \left\{ \sqrt{\frac{\log \frac{2}{\varepsilon}}{\left(\frac{\mu_1^2 r^2}{2} - \Lambda_{v_1 v_2}(\alpha^2 s)\right)}} \right\}, \tag{17}$$

Proof. We are interested in studying the term

$$\hat{\theta}^N = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{x}_n = \left(\frac{1}{N}\sum_{n=1}^{N}\mathbf{a}_n\mathbf{a}_n^T\right)^{-1}\frac{1}{N}\sum_{n=1}^{N}\mathbf{a}_n^T\mathbf{x}_n \tag{18}$$

and its relationship with the true parameter θ_0. Recalling the definition of $\mathbf{X}$ we obtain,

$$\left|\left(\hat{\theta}^N - \theta_0\right)_i\right| \leq \lambda_{max}\left(\frac{1}{N}\sum_{n=1}^{N}\mathbf{a}_n\mathbf{a}_n^T\right)^{-1}\frac{1}{N}\sum_{n=1}^{N}a_{ni}v_n \tag{19}$$

Define $\omega(\mathbf{A}) \doteq \left(\lambda_{min}\left(\frac{1}{N}\sum_{n=1}^{N}\mathbf{a}_n\mathbf{a}_n^T\right)\right)^{-1}$. Using the fact that $\lambda_{max}\left(\mathbf{A}^{-1}\right) = \frac{1}{\lambda_{min}(\mathbf{A})}$ we obtain

$$P\left(\left|\left(\hat{\theta^N} - \theta_0\right)_i\right| > r\right) \leq P\left(\left(\frac{1}{N}\sum_{n=1}^{N}a_{ni}v_n\right)^2 > \frac{r^2}{\omega(\mathbf{A})^2}\right). \tag{20}$$

We want to use Chernoff bound to bound the probability (20). In order to use Chernoff bound we need

$$E\left(\left(\frac{1}{N}\sum_{n=1}^{N}a_{ni}v_n\right)^2\right) < \frac{r^2}{\omega(\mathbf{A})^2}. \tag{21}$$

Developing the term $\left(\frac{1}{N}\sum_{n=1}^{N}a_{ni}v_n\right)^2$ yields

$$\left(\frac{1}{N}\sum_{n=1}^{N}a_{ni}v_n\right)^2 = \frac{1}{N^2}\sum_{n=1}^{N}a_{ni}^2v_n^2 + \frac{1}{N^2}\sum_{n=1}^{N}\sum_{\substack{l=1\\l\neq n}}^{N}a_{ni}v_n a_{li}v_l. \tag{22}$$

Using the fact that $|a_{ni}| \leq \alpha$ and v_n are zero mean i.i.d we achieve

$$E\left(\left(\frac{1}{N}\sum_{n=1}^{N}a_{ni}v_n\right)^2\right) \leq \frac{\alpha^2}{N}E\left(v^2\right) \tag{23}$$

Assigning this to inequality (21), rearranging and using assumption A4 to bound $\omega(\mathbf{A})$ we achieve the first condition for N.

$$N > \frac{\alpha^2 E(v^2)\,\omega(\mathbf{A})^2}{r^2} > \frac{\alpha^2 E(v^2)}{\mu_1^2 r^2} = N_1(r). \tag{24}$$

We now evaluate the two terms of equation (22) using Chernoff-like bounds. $\forall s > 0$

$$P\left(\frac{1}{N^2}\sum_{n=1}^{N}a_{ni}^2v_n^2 > \frac{r^2}{2\omega(\mathbf{A})^2}\right) \leq \exp\left(-\frac{r^2N^2s}{2\omega(\mathbf{A})^2}\right)E\left(\exp\left(s\alpha^2v^2\right)\right)^N. \tag{25}$$

We utilized the bound on a_{ni} and the independence of v_n.

To ensure that $\forall\varepsilon > 0$

$$P\left(\frac{1}{N^2}\sum_{n=1}^{N}a_{ni}^2v_n^2 > \frac{r^2}{2\omega(\mathbf{A})^2}\right) < \frac{\varepsilon}{2}, \tag{26}$$

we use the bound we achieved and solve the inequality for N. We also take infimum over s and use assumption A4 to bound $\omega(\mathbf{A})$. This gives the bound

$$N > \inf_{s>0}\left\{\frac{1}{\mu_1^2r^2s}\left(2\Lambda_{v^2}(\alpha^2 s) + \mu_2 r\sqrt{2s\log\frac{2}{\varepsilon}}\right)\right\} = N_2(r,\varepsilon). \tag{27}$$

We now evaluate the second term of (22)

Define

$$\Gamma = \left\{X : \frac{1}{N^2}\sum_{n=1}^{N}\sum_{\substack{l=1\\l\neq n}}^{N}a_{ni}a_{li}v_nv_l > \frac{r^2}{2\omega(\mathbf{A})^2}\right\}. \tag{28}$$

$$P(\Gamma) < \exp\left(-\frac{N^2r^2}{2\omega(\mathbf{A})^2}\right)E\left(\exp\left(s\alpha^2v_1v_2\right)\right)^{N(N-1)}, \tag{29}$$

where v_1 and v_2 are two i.i.d random variables distributed according to the law of v. In order to ensure that

$$P(\Gamma) < \frac{\varepsilon}{2} \tag{30}$$

we use the achieved inequality, take the logarithm on both sides, solve for N, take the infimum over s and bound $\omega(\mathbf{A})$. We get

$$N > \inf_{s>0}\left\{\sqrt{\frac{\log\frac{2}{\varepsilon}}{\left(\frac{\mu_1^2r^2}{2} - \Lambda_{v_1v_2}(s\alpha^2)\right)}}\right\} = N_3(r,\varepsilon). \tag{31}$$

Choosing $N > \max\{N_1(r), N_2(r,\varepsilon), N_3(r,\varepsilon)\}$ and using the union bound ends the proof. □

3.3. Proof of the main theorem

In this subsection we prove the main theorem 4.1. We use lemma 3.2 and provide the bounds on $N_1(r)$, $N_2(r,\varepsilon)$ and $N_3(r,\varepsilon)$ derived in this theorem. Furthermore, to complete the proof we use a union bound for the p elements of the vector $\hat{\theta}_\lambda^N$.

Using the fact that v is sub-Gaussian and using a one dimensional version of remark 2.2 in [30] we get

$$E\left(v^2\right) \leq R^2. \tag{32}$$

Assigning this bound to (15) we get

$$N_1(r) \leq \frac{\alpha^2R^2}{\mu_1^2r^2}. \tag{33}$$

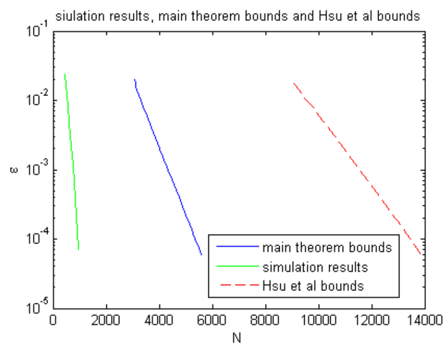

Fig. 1: Main theorem bounds, Hsu et al. bounds and simulation results for uniform + Gaussian noise with $R = 0.1$, $r = 0.01$ and $p = 8$.

We turn to bound the moment generating function of the term v^2. By using this bound and assigning it to (16) we achieve a bound on (16). Using remark 2.3 in [30] we obtain $\forall 0 < s < \frac{1}{2\alpha^2 R^2}$

$$E\left(e^{s\alpha^2 v^2}\right) \leq \exp\left(R^2\alpha^2 s + \frac{R^4\alpha^4 s^2}{1-2R^2\alpha^2 s}\right) = \exp\left(\beta\left(s\right)\right). \tag{34}$$

Substituting into (16) we obtain

$$N_2\left(r,\varepsilon\right) \leq \inf_{0<s<\frac{1}{2\alpha^2R^2}} \left\{\frac{1}{\mu_1^2 r^2 s}\left(2\beta\left(s\right)\right. \right. \left.\left. + \mu_2 r\sqrt{2s\log\frac{2}{\varepsilon}}\right)\right\}. \tag{35}$$

We next evaluate the term $\Lambda_{v_1 v_2}\left(s\right)$. Using bounds from [31] and [30] together with smoothing we get $\forall 0 < s < \frac{1}{\alpha^2 R^2}$

$$\Lambda_{v_1 v_2}\left(\alpha^2 s\right) \leq \frac{R^4\alpha^4 s^2}{2} + \frac{R^8\alpha^8 s^4}{4\left(1-\alpha^4 s^2 R^4\right)} = \gamma\left(s\right). \tag{36}$$

Using the bound on $\Lambda_{v_1 v_2}\left(\alpha^2 s\right)$ we get

$$N_3\left(r,\varepsilon\right) \leq \inf_{0<s<\frac{1}{\alpha^2R^2}} \left\{\sqrt{\frac{\log\frac{2}{\varepsilon}}{\frac{\mu_1^2 r^2}{2} - \gamma\left(s\right)}}\right\}. \tag{37}$$

We thus obtain bounds for N_1, N_2 and N_3. Using the union bound yields bounds for every element in the vector $\hat{\theta}^N - \theta_0$. Using union bound again on the elements of this vector to make sure that every element in the vector is less than r and assigning $\varepsilon = \frac{\varepsilon}{p}$ makes sure that with a probability of $1-\varepsilon$, $\left\|\hat{\theta}^N - \theta_0\right\|_\infty < r$. This gives us the desired result and completes the proof.

4. RANDOM MIXING MATRIX EXTENSION

In the previous sections we established and proved a finite sample result for the least squares problem with a fixed mixing matrix. In this section we consider an extension to the case where the mixing matrix is random.

Theorem 4.1. (Random mixing matrix)
Assume our model admits the assumptions A1-A5 where $\mathbf{A}$ *is a random matrix satisfying* $\forall 1 \leq n \leq N, 1 \leq l \leq p$ $E\left(A_{nl}\right) = 0$, $E\left(A_{nl}^2\right) = 1$ *and* $\left|A_{nl}\right| \leq \alpha$. *Assume there exist* $\varepsilon, r > 0$. *Then,* $\forall N > \max\left\{N\left(r,\frac{\varepsilon}{2}\right), N_{rand}\left(\frac{\varepsilon}{2}\right)\right\}$.

$$P\left(\left\|\hat{\theta}^T - \theta_0\right\|_\infty > r\right) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \tag{38}$$

where

$$N_{rand}\left(\frac{\varepsilon}{2}\right) = \frac{56p\alpha^2}{3}\log\frac{2p}{\varepsilon}, \tag{39}$$

$\omega\left(\mathbf{A}\right)$ *is bounded by* 2 *and* $N\left(r,\frac{\varepsilon}{2}\right)$ *is defined in 7.*

Proof. The proof for the main theorem remains valid in this setting. However, we can no longer assume that we know the minimal eigenvalue of $\frac{1}{N}\sum_{n=1}^{N} E\left(\mathbf{a}_n\mathbf{a}_n^T\right)$. In order to prove this theorem we will use random matrix inequalities to bound $\omega\left(\mathbf{A}\right)$. Using the bound on $\left|\mathbf{A}_{nl}\right|$ we obtain $\left\|\mathbf{a}_n\right\| \leq \sqrt{p}\alpha$. We denote $\mathbf{B}_n \doteq \frac{1}{N}\left(\mathbf{a}_n\mathbf{a}_n^T - \mathbf{I}\right)$. Following the proof of theorem 5.41 in [32] we have

$$\lambda_{max}\left(\mathbf{B}_n\right) \leq \frac{2p\alpha^2}{N}, \quad \left\|\sum_{n=1}^{N} E\left(\mathbf{B}_n\right)\right\| \leq \frac{2p\alpha^2}{N} \tag{40}$$

Using this result we can assign to matrix Bernstein inequality [33] and achieve

$$P\left(\lambda_{min}\left(\frac{1}{N}\sum_{n=1}^{N}\mathbf{a}_n\mathbf{a}_n^T\right) \leq \frac{1}{2}\right) \leq p\exp\left(-\frac{3N}{56p\alpha^2}\right) \tag{41}$$

Choosing $N > N_{rand}\left(\frac{\varepsilon}{2}\right)$ ensures that

$$P\left(\lambda_{min}\left(\frac{1}{N}\sum_{n=1}^{N}\mathbf{a}_n\mathbf{a}_n^T\right) \leq \frac{1}{2}\right) \leq \frac{\varepsilon}{2} \tag{42}$$

and therefore, $\omega\left(\mathbf{A}\right) \leq 2$ with a probability of at least $1-\frac{\varepsilon}{2}$. Using this bound we can write $\omega\left(\mathbf{A}\right) = 2$ in the context of the main theorem. The proof of the main theorem remains the same. The union bound for the results completes the proof. □

5. SIMULATION RESULTS

In this section we show simulation results to demonstrate the bounds. We chose a model matrix with elements generated from uniform distribution in $[-\frac{\sqrt{12}}{2}, \frac{\sqrt{12}}{2}]$. The chosen noise model was gaussian plus uniform noise which admits the sub-Gaussian assumption. The simulations were made using 100000 samples for each value of N. ϵ was estimated by dividing the number of outliers with the number of simulations. The simulation results shown in Fig 1. show that the overall performance of the bound is similar to the performance of the simulations and that the bounds in this paper are tighter than the bounds generated by [26]. It is interesting to see that the bound is not tight and try to tighten the bound to achieve better results.

6. CONCLUSION

In this paper we gave simple bounds on L^∞ error of least squares estimator with finitely many samples and non Gaussian noise. We show very fast convergence of the number of samples required for a given accuracy and outage probability. In the fixed mixing matrix setting we show that $N \sim O\left(\frac{1}{r^2}\sqrt{\log\frac{1}{\varepsilon}}\right)$. We also prove an extension for random mixing matrix setting. In that case we see that $N \sim O\left(\frac{1}{r^2}\log\frac{1}{\varepsilon}\right)$. This result has significant impact on the analysis of least squares solutions in communications and signal processing.

7. REFERENCES

[1] Xiangjun Zhang and Xiaolin Wu, "Image interpolation by adaptive 2-d autoregressive modeling and soft-decision estimation," *Image Processing, IEEE Transactions on*, vol. 17, no. 6, pp. 887–896, 2008.

[2] Kwok-Wai Hung and Wan-Chi Siu, "Robust soft-decision interpolation using weighted least squares," *Image Processing, IEEE Transactions on*, vol. 21, no. 3, pp. 1061–1069, 2012.

[3] Hing Cheung So and Lanxin Lin, "Linear least squares approach for accurate received signal strength based source localization," *Signal Processing, IEEE Transactions on*, vol. 59, no. 8, pp. 4035–4040, 2011.

[4] Jelle Veraart, Jan Sijbers, Stefan Sunaert, Alexander Leemans, and Ben Jeurissen, "Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls," *Neuroimage*, vol. 81, pp. 335–346, 2013.

[5] Fernando Perez-Gonzalez, Carmela Troncoso, and Simon Oya, "A least squares approach to the static traffic analysis of high-latency anonymous communication systems," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 9, pp. 1341–1355, 2014.

[6] A Ozan Bicen and Ian F Akyildiz, "System-theoretic analysis and least-squares design of microfluidic channels for flow-induced molecular communication," *Signal Processing, IEEE Transactions on*, vol. 61, no. 20, pp. 5000–5013, 2013.

[7] Jia-Chin Lin, "Least-squares channel estimation for mobile ofdm communication on time-varying frequency-selective fading channels," *Vehicular Technology, IEEE Transactions on*, vol. 57, no. 6, pp. 3538–3550, 2008.

[8] Amir Leshem and Alle-Jan Van der Veen, "Direction-of-arrival estimation for constant modulus signals," *Signal Processing, IEEE Transactions on*, vol. 47, no. 11, pp. 3125–3129, 1999.

[9] Petre Stoica and Nehorai Arye, "MUSIC, maximum likelihood, and Cramer-Rao bound," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 5, pp. 720–741, 1989.

[10] Petar M Djurić, Jayesh H Kotecha, Fabien Esteve, and Etienne Perret, "Sequential parameter estimation of time-varying non-Gaussian autoregressive processes," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 1, pp. 865–875, 2002.

[11] Sean Banerjee and Meena Agrawal, "On the performance of underwater communication system in noise with Gaussian mixture statistics," in *Communications (NCC), 2014 Twentieth National Conference on*. IEEE, 2014, pp. 1–6.

[12] Jin Tan, Dror Baron, and Liyi Dai, "Wiener filters in Gaussian mixture signal estimation with-norm error," *Information Theory, IEEE Transactions on*, vol. 60, no. 10, pp. 6626–6635, 2014.

[13] Vimal Bhatia and Bernard Mulgrew, "Non-parametric likelihood based channel estimator for Gaussian mixture noise," *Signal Processing*, vol. 87, no. 11, pp. 2569–2586, 2007.

[14] Xiaodong Wang and H Vincent Poor, "Robust multiuser detection in non-Gaussian channels," *Signal Processing, IEEE Transactions on*, vol. 47, no. 2, pp. 289–305, 1999.

[15] Erik Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas Marzetta, "Massive MIMO for next generation wireless systems," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 186–195, 2014.

[16] Tze Leung Lai, Herbert Robbins, and Ching Zong Wei, "Strong consistency of least squares estimates in multiple regression," in *Herbert Robbins Selected Papers*, pp. 510–512. Springer, 1985.

[17] B.L.S Prakasa Rao, "On the exponential rate of convergence of the least squares estimator in the nonlinear regression model with Gaussian errors," *Statistics & probability letters*, vol. 2, no. 3, pp. 139–142, 1984.

[18] B.L.S Prakasa Rao, "The rate of convergence of the least squares estimator in a non-linear regression model with dependent errors," *Journal of Multivariate Analysis*, vol. 14, no. 3, pp. 315–322, 1984.

[19] Chien-Fu Wu, "Asymptotic theory of nonlinear least squares estimation," *The Annals of Statistics*, pp. 501–513, 1981.

[20] Arthur Sieders and Kacha Dzhaparidze, "A large deviation result for parameter estimators and its application to nonlinear regression analysis," *The Annals of Statistics*, pp. 1031–1049, 1987.

[21] Hu Shuhe, "A large deviation result for the least squares estimators in nonlinear regression," *Stochastic processes and their applications*, vol. 47, no. 2, pp. 345–352, 1993.

[22] Carl-Gustaf Esseen, *On the Liapounoff limit of error in the theory of probability*, Almqvist & Wiksell, 1942.

[23] Andrew C Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Transactions of the american mathematical society*, vol. 49, no. 1, pp. 122–136, 1941.

[24] I.G. Shevtsova, "An improvement of convergence rate estimates in the Lyapunov theorem," in *Doklady Mathematics*. Springer, 2010, vol. 82, pp. 862–864.

[25] Roberto Imbuzeiro Oliveira, "The lower tail of random quadratic forms, with applications to ordinary least squares and restricted eigenvalue properties," *arXiv preprint arXiv:1312.2903*, 2013.

[26] Daniel Hsu, Sham M Kakade, and Tong Zhang, "Random design analysis of ridge regression," *Foundations of Computational Mathematics*, vol. 14, no. 3, pp. 569–600, 2014.

[27] Jean-Yves Audibert and Olivier Catoni, "Robust linear regression through PAC-Bayesian truncation," *Preprint, URL http://arxiv. org/abs/1010.0072*, vol. 38, pp. 60, 2010.

[28] Jean-Yves Audibert and Olivier Catoni, "Robust linear least squares regression," *The Annals of Statistics*, pp. 2766–2794, 2011.

[29] Andrea Caponnetto and Ernesto De Vito, "Optimal rates for the regularized least-squares algorithm," *Foundations of Computational Mathematics*, vol. 7, no. 3, pp. 331–368, 2007.

[30] Daniel Hsu, Sham M Kakade, and Tong Zhang, "A tail inequality for quadratic forms of subgaussian random vectors," *Electron. Commun. Probab*, vol. 17, no. 52, pp. 6, 2012.

[31] V Buldygin and E Pechuk, "Inequalities for the distributions of functionals of sub-Gaussian vectors," *Theory of Probability and Mathematical Statistics*, vol. 80, pp. 25–36, 2010.

[32] Roman Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *arXiv preprint arXiv:1011.3027*, 2010.

[33] Joel A. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of Computational Mathematics*, vol. 12, no. 4, pp. 389–434, 2012.