Stability Analysis of LASSO and Dantzig Selector via Constrained Minimal Singular Value of Gaussian Sensing Matrices

Oliver James
Post-Doctoral Researcher
School of Information and Communications
Gwangju Institute of Science and Technology, South Korea

ABSTRACT

In this paper, we introduce a new framework for interpreting the existing theoretical stability results of sparse signal recovery algorithms in practical terms. Our framework is built on the theory of constrained minimal singular values of Gaussian sensing matrices. Adopting our framework, we study the stability of two algorithms, namely LASSO and Dantzig selector. We demonstrate that for a given stability parameter (noise sensitivity), there exits a minimum undersampling ratio above which the recovery algorithms are guaranteed to be stable.

General Terms:

Algorithms, Mobile computing

Keywords:

Compressed Sensing, Constrained minimal singular value, Stability analysis, Convex algorithms, Undersampling analysis

1. INTRODUCTION

Recovery algorithms in compressed sensing (CS) aim to reconstruct a K-sparse signal $\boldsymbol{x} \in \mathbb{R}^N$ from its measurement $\boldsymbol{z}$ of length M, where $M < N$. The stability of a recovery algorithm indicates the resistance of the algorithm to noise. The noisy system model in CS is [3]

$$\boldsymbol{z} = A\boldsymbol{x} + \boldsymbol{w} \ . \tag{1}$$

In (1), $A \in \mathbb{R}^{M\times N}$ is an $M \times N$ sensing matrix, which can be deterministic or random [3]. The vector $\boldsymbol{w} \in \mathbb{R}^M$ denotes the zero-mean Gaussian noise. The signal $\boldsymbol{x}$ is K-sparse, i.e., it has *exactly* K non-zero values. Let Ψ denote the set of all K-sparse signals, and $\mathcal{K}_i$, $i = 1, \cdots, \binom{N}{K}$, stands for the i-th *support set* of a signal $\boldsymbol{x} \in \Psi$. Let $\delta = \frac{M}{N}$ denote the *undersampling ratio* and $\rho = \frac{K}{M}$ represent a sparsity measure.

The stability of various algorithms are investigated theoretically using the restricted isometry constant (RIC) [3] and the l_1-constrained minimal singular value (l_1-CMSV) [10] of a sensing matrix. However, till date, no framework has been proposed to interpret these theoretical stability guarantees in practical terms, because the computation of the constants are intractable. While the computation of the RIC is NP-hard, till date only the lower bounds of the l_1-CMSV are computable [10, Section IV.B]. The l_1-CMSV is simpler to calculate than the RIC; also, it is as tight as that of the RIC as demonstrated in [10, Fig. 4]. Thus, in this letter, we use the l_1-CMSV for analysing the stability of the algorithms. The l_1-CMSV is defined as follows

DEFINITION 1. *l_1-CMSV [10]: For any $K \in [1, N]$ and for a sensing matrix $A \in \mathbb{R}^{M\times N}$, the l_1-CMSV of A is defined as*

$$\alpha_K = \min_{\boldsymbol{x}\neq 0, \boldsymbol{x}\in S_K} \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \ , \tag{2}$$

*where the l_1-*sparsity set $S_K = \left\{\boldsymbol{x} \in \mathbb{R}^N : \frac{\|\boldsymbol{x}\|_1^2}{\|\boldsymbol{x}\|_2^2} \leq K\right\}$.

The l_1-sparsity set contains both the exactly sparse and the approximately sparse signals [12]. In this letter, we focus on the exactly sparse signal set Ψ, in which case the l_1-CMSV can be computed by considering the set $\Psi \subseteq S_K$.

The central goal in CS is to recover $\boldsymbol{x}$ from $\boldsymbol{z}$ given A. Two popular convex relaxation algorithms for achieving this goal are the LASSO and the Dantzig selector (DS) and they are given for the model (1) as

$$\text{LASSO} : \min_{\boldsymbol{x}\in\mathbb{R}^N} \tfrac{1}{2}\|\boldsymbol{z} - A\boldsymbol{x}\|_2^2 + \lambda\sigma\|\boldsymbol{x}\|_1 \ , \tag{3}$$

$$\text{DS} : \min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{x}\|_1 \quad \text{s.t.} \|A^T(\boldsymbol{z} - A\boldsymbol{x})\|_\infty \leq \lambda\sigma \ . \tag{4}$$

Here λ is a tuning parameter and σ is a measure of noise level. Stability analysis aims to derive bounds on the recovery error in terms of the noise level σ and the constants of the sensing matrix such as the RIC [2] or the l_1-CMSV [10, p. 5737]. The recovery error bounds for the LASSO and the DS in terms of the l_1-CMSV are given below.

LEMMA 2. *Let α_K be the l_1-CMSV in (2) and $\boldsymbol{x} \in \Psi$ be a K-sparse signal. If the noise $\boldsymbol{w}$ in the LASSO (3) obeys $\|A^T\boldsymbol{w}\|_\infty \leq \kappa\lambda\sigma$ for some $\kappa \in (0, 1)$, then the solution $\widehat{\boldsymbol{x}}$ to (3) obeys*

$$\|\widehat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{1+\kappa}{1-\kappa} \frac{2\sqrt{K}\lambda\sigma}{\alpha^2 \frac{4K}{(1-\kappa)^2}} \ . \tag{5}$$

Similarly, if the noise $\boldsymbol{w}$ in the DS (4) obeys $\|A^T\boldsymbol{w}\|_\infty \leq \lambda\sigma$, then the solution to (4) satisfies

$$\|\widehat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \frac{4\sqrt{K}\lambda\sigma}{\alpha_{4K}^2} . \quad (6)$$

We note that in both (5) and (6), the term $\alpha_{(.)}^2$ denotes the square of the l_1-CMSV (SCV). Thus, we consider the SCV than the l_1-CMSV. In this letter, we aim to address the two questions: 1. How to precisely compute the SCV for the widely-used Gaussian sensing matrices? 2. Using the SCV, how to interpret the theoretical stability results in (5) and (6) in practical terms? We first bring to forefront that for the Gaussian sensing matrices, the SCV is a random variable. Using the extreme value theory (EVT), we show that the SCV is Weibull distributed. Thus, by determining the probability distributions of the SCV, we precisely characterize the l_1-CMSV rather than using its lower bounds as in [10]. We demonstrate the benefit of these distributions for analysing the stability of the algorithms in terms of the minimum undersampling ratio.

2. THE SCV FOR GAUSSIAN ENSEMBLE

In this section, we characterize the SCV for Gaussian sensing matrices and in the next section we discuss how to use the SCV for stability analysis. The SCV is a constant for deterministic sensing matrices and it is a random variable for random sensing matrices. Before we describe the SCV, we consider the ratio in (7) and study its squared value for the Gaussian sensing matrices. For every K-sparse signal with the fixed support set $\mathcal{K}$, the squared ratio

$$R_{\mathcal{K}} = \frac{\|A\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \quad (7)$$

is a random variable, which is described in Lemma 3.

LEMMA 3. *Let A be a Gaussian sensing matrix whose elements are i.i.d. Gaussian random variables each with zero mean and variance $\sigma^2 = \frac{1}{M}$. For every $\boldsymbol{x} \in \Psi$, $R_{\mathcal{K}}$ in (7) is a central Chi-square random variable with M degrees of freedom.*

PROOF. Let $\boldsymbol{y} = A\boldsymbol{x}$ and $c = \frac{1}{\|\boldsymbol{x}\|_2^2}$. The ith entry y_i of $\boldsymbol{y}$ is a Gaussian random variable with mean 0 and variance $\sigma^2\|\boldsymbol{x}\|_2^2$. The random variables $\{y_i\}$, $i = 1, 2, \cdots, M$, are independent. Now, y_i^2 is a central Chi-square random variable with 1-degree of freedom. Then, $\|\boldsymbol{y}\|_2^2 = \|A\boldsymbol{x}\|_2^2$ is a central Chi-square distribution with M-degrees of freedom with mean $\frac{M\sigma^2}{c}$ and variance $\frac{2M\sigma^4}{c^2}$. Since $R_{\mathcal{K}} = c\|A\boldsymbol{x}\|_2^2$, it is also a Chi-square random variable with mean $M\sigma^2$ and variance $2M\sigma^4$. $\square$

We note that the statistics of $R_{\mathcal{K}}$ is identical for all support sets and it is independent of $\boldsymbol{x}$. Since $R_{\mathcal{K}}$ is a central Chi-square random variable, it is completely characterized by its degrees of freedom, M. We bring to the notice of the reader that a Lemma similar to Lemma 3 has been derived in [8] for the general case when the "entries of A are not necessarily independent." We now define the SCV for Gaussian sensing matrices via $R_{\mathcal{K}}$ as follows.

LEMMA 4. *SCV for Gaussian sensing matrix: Let $A \in \mathbb{R}^{M\times N}$ be a Gaussian sensing matrix. Let $N_S = \binom{N}{K}$ and for any integer $K \in [1, N]$ and for any K-sparse signal $\boldsymbol{x} \in S_K$, the SCV for A is defined as*

$$\begin{aligned}\alpha_K^2 &= \min_{\boldsymbol{x}\neq 0, \boldsymbol{x}\in S_K} \frac{\|A\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \\ &= \min_{i=1,2,\cdots,N_S} R_{\mathcal{K}_i}\end{aligned} \quad (8)$$

The second line in (8) is due to the fact that for every K-sparse signal $\boldsymbol{x} \in S_K$, the ratio $R_{\mathcal{K}}$ is a Chi-square random variable. Therefore, the minimization in (8) is over the total number of support sets. Thus, we define α_K^2 as the minimum of N_S number of i.i.d. Chi-square random variables. We note that the Lemma 8 is for the Gaussian ensemble. For non-Gaussian matrices, works need to be done in order to define the ratio in (7) and the SCV in (8).
We note that the SCV is defined in (8) with the assumption that the random variables $R_{\mathcal{K}_i}$ and $R_{\mathcal{K}_j}, i \neq j$ are i.i.d. These random variables, of course, are correlated, because, most of the support sets share common support elements. This correlation structure can be exploited to further improve the values predicted by the SCV. How far can we improve is still remains as an open problem. However, i.i.d. assumption is sufficient for the stability analysis considered in this paper. The proof for the sufficiency of the i.i.d. assumption for the stability analysis is beyond the scope of this paper.
In the next section, we derive the cumulative distribution function (CDF) and the probability distribution function (PDF) of the SCV in (8). While for small values of N_S the distributions can be obtained straightforwardly by using the minimum order statistics of N_S i.i.d. Chi-square random variables. For large values of N_S, the distributions are obtained utilizing the tools from the extreme value theory.

3. DISTRIBUTIONS OF SCV

In this section, we derive the CDF and the PDF of the SCV for two scenarios, namely, non-asymptotic and asymptotic. The criterion that differentiates between the two scenarios is the convergence of the minimum of i.i.d. random variables in (8). It is well-known in EVT that the CDF of the minimum of n i.i.d. Chi-square random variables converges (in distribution) to the Weibull CDF [4, Table 9.5]. The rate of this convergence [5, Theorem 3.2] is $\mathcal{O}((\log N_S)^{-2})$. Therefore, we refer to the scenarios where $(\log N_S)^{-2} > \epsilon$ as non-asymptotic and $(\log N_S)^{-2} < \epsilon$ as asymptotic scenarios for a small number $\epsilon > 0$.

3.1 Non-asymptotic scenario

THEOREM 5. *Let Ω denote an ensemble of Gaussian encoders and α_K^2 be the SCV defined in (8). The non-asymptotic CDF of the SCV is given by*

$$F_\alpha(u) = 1 - \left[1 - \frac{\gamma\left(\frac{M}{2}, \frac{Mu}{2}\right)}{\Gamma\left(\frac{M}{2}\right)}\right]^{\binom{N}{K}} , \quad 0 \leq u \leq 1 \quad (9)$$

and its corresponding PDF is given by

$$p_\alpha(u) = \frac{\left(\frac{M}{2}\right)^{\frac{M}{2}}\binom{N}{K}}{\Gamma\left(\frac{M}{2}\right)}\left[1 - \frac{\gamma(\frac{M}{2}, \frac{Mu}{2})}{\Gamma\left(\frac{M}{2}\right)}\right]^{\binom{N}{K}-1} u^{\frac{M}{2}-1}e^{-\frac{Mu}{2}} \quad (10)$$

where $\gamma(w, \mu z) = \mu^w \int_0^z x^{w-1}e^{-\mu x}dx$ is the lower incomplete gamma integral [7, eq. 3.381.1] and $\Gamma(x)$ is the value of the gamma function at x.

PROOF. The CDF, $F_X(x)$ of the minimum of n i.i.d. random variables with the common CDF $F(x)$ is [1, eq. 2.2.11] $F_X(x) = 1 - [1 - F(x)]^n$. Thus, we write the CDF of α_K^2, as $F_\alpha(u) = 1 - [1 - F_C(u)]^{N_S}$. The PDF of α_K^2 is then given by $p_\alpha(u) = \frac{dF_\alpha(u)}{du} = N_S[1 - F_C(u)]^{N_S-1}p_C(u)$. By substituting $F_C(u)$ and $p_C(u)$ from Appendix A into $F_\alpha(u)$ and $p_\alpha(u)$, we obtain the desired results stated in Theorem 5. $\square$

The distributions in Theorem 5 are, as expected, functions of K, M and N. While their dependence on M arrives solely from the ratio random variable, their dependence on N and K comes from the total number of support sets.

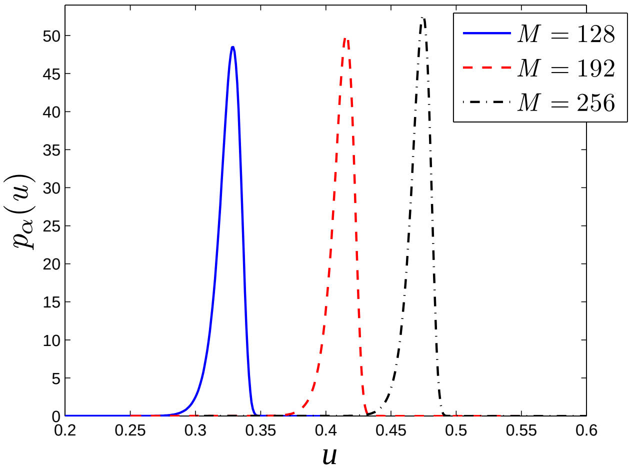

Fig. 1. PDF of SCV in (10) for $N = 512$, $K = 6$ and for various M.

In Fig. 1, we plot the PDF of the SCV for $N = 512$, $K = 6$ and for various M. We note from Fig. 1 that on increasing M (for a fixed K) the PDF moves towards one (better signal recovery). This confirms the results in [10, Section VI. B] that the values of the l_1-CMSV move towards one when increasing M. This observation also aligns with the fact that increasing the number of measurements results in better sparse signal recovery.

Next, we are interested in finding the asymptotic distributions of the SCV for practical systems where N_S is sufficiently large. For example, in a typical Gaussian matrix-based CS systems such as the turbid lens imaging [9], the problem sizes are $N = 2 \times 10^4$ and $K = 147$, in which case N_S is very large. It is well-known in EVT that for sufficiently large values of n, the CDF of the minimum of n i.i.d. Chi-square random variables converges (in distribution) to the Weibull CDF [4, Table 9.5]. We adopt this result from the EVT and tailored it for the SCV when N_S is sufficiently large. The asymptotic distribution of the SCV is defined in the next theorem.

3.2 Asymptotic scenario: Weibull

The distributions for this scenario are described below.

THEOREM 6. *Let Ω denote an ensemble of Gaussian encoders and α_K^2 is the SCV defined in (8). The asymptotic CDF of the SCV follows the Weibull distribution:*

$$F_\alpha^\infty(u) = 1 - \exp^{-\left(\frac{u}{q}\right)^\beta} \quad , \quad 0 \le u \le 1 \tag{11}$$

where $q = \frac{6}{M}\exp\left(\frac{2}{M}\left[\log\frac{M}{2} + \log\Gamma\left(\frac{M}{2}\right) - \log\binom{N}{K}\right]\right)$ *is the scaling constant and* $\beta = \frac{M}{2}$ *is the shape parameter. The corresponding asymptotic PDF of the SCV is given by*

$$p_\alpha^\infty(u) = \left(\frac{\beta}{q}\right)\left(\frac{u}{q}\right)^{\beta-1}\exp^{-\left(\frac{u}{q}\right)^\beta} \tag{12}$$

PROOF. Please see Appendix B. □

Thus, by finding the CDF and the PDF of the SCV, we precisely characterize the SCV for the Gaussian matrices.

4. STABILITY ANALYSIS OF LASSO AND DS

Stability analysis addresses the issue of robustness of a recovery algorithm to measurement noise. In this section, we aim to analyse the stability of the LASSO and the DS by using the distributions of the SCV in Lemma 6. In particular, we show the role of SCV for interpreting the theoretical stability results in (5) and (6) in terms of how aggressive a K-sparse signal can be undersampled in practice. Let $\xi := \sup_{\sigma>0}\frac{\|\widehat{\boldsymbol{x}}-\boldsymbol{x}\|_2}{\sigma}$ denote the sensitivity of a recovery algorithm to noise and hence it is called the *noise sensitivity*. A recovery is said to be *robust* if the noise sensitivity is finite [11]. First let us focus on the LASSO.

From (5), the noise sensitivity of the LASSO is given in terms of the SCV as $\xi_L := \frac{1+\kappa}{1-\kappa}\frac{2\sqrt{K}\lambda}{\alpha^2_{\frac{4K}{(1-\kappa)^2}}}$, where $\lambda = \sqrt{2\log N}$ for Gaussian noise [10, p. 5737]. We observe that ξ_L is a random variable. We now find, with noise sensitivity ξ_L stays finite (guaranteed stable recovery), what is the minimum undersampling ratio δ possible. To this end, we first give the minimum undersampling ratio for the LASSO.

THEOREM 7. *(Minimum undersampling for LASSO) Let α_K^2 be the SCV (8) and consider its CDF $F_\alpha(u)$ in (11) with the parameters q and β. Let $\widetilde{\xi}_L$ be the finite, desired noise sensitivity and $\epsilon > 0$ be a very small number. Then, the minimum undersampling of LASSO is given by*

$$\delta_L^*(\rho) = \inf\{\delta \ : \ \Pr(\xi_L \le \widetilde{\xi}_L) \to 1-\epsilon\} \ , \tag{13}$$

Equation (13) tells us that $\delta_L^*(\rho)$ is the minimum value of δ when the noise sensitivity ξ_L stays less than $\widetilde{\xi}_L$ with probability $1-\epsilon$. Thus, $\delta_L^*(\rho)$ determines how aggressive a signal can be undersampled along with a guaranteed stable recovery.

In order to find $\delta_L^*(\rho)$, we first derive a general expression for the upper bound $\widetilde{\xi}_L$ as a function of ρ and δ, $\widetilde{\xi}_L(\rho,\delta)$, by setting the probability (13) equal to $1-\epsilon$. We call $\widetilde{\xi}_L(\rho,\delta)$ the *noise sensitivity upper bound*. The minimum value of this upper bound, for a fixed ρ, yields $\delta_L^*(\rho)$. To find $\widetilde{\xi}_L(\rho,\delta)$, we set $\Pr(\xi_L \le \widetilde{\xi}_L) = 1-\epsilon$. Using the definition of ξ_L, we write $\Pr(\xi_L \le \widetilde{\xi}_L) = \Pr\left\{\alpha^2_{\frac{4K}{(1-\kappa)^2}} \ge \frac{1+\kappa}{1-\kappa}\frac{2\sqrt{K}\lambda}{\widetilde{\xi}_L}\right\} = 1 - F_\alpha\left(\frac{1+\kappa}{1-\kappa}\frac{2\sqrt{K}\lambda}{\widetilde{\xi}_L}\right)$. Now setting the probability to $1-\epsilon$ and solving for $\widetilde{\xi}_L$ gives an expression in terms of N, M and K. On substituting, $K = \rho M$, $\lambda = \sqrt{2\log N}$ and $M = \delta N$, we obtain the noise sensitivity upper bound for the LASSO as,

$$\widetilde{\xi}_L(\rho,\delta) = \frac{\frac{1+\kappa}{1-\kappa}\,2\sqrt{\rho\delta N}\sqrt{2\log N}}{q_L\left[\ln\frac{1}{1-\epsilon}\right]^{\frac{2}{\delta N}}} \tag{14}$$

where q_L is q in (11) with K replaced by $\frac{4K}{(1-\kappa)^2}$.

In a similar manner, we now derive the minimum undersampling of the DS. From (6), the noise sensitivity of the DS is $\xi_D = \frac{4\sqrt{K}\lambda}{\alpha^2_{4K}}$. With the finite noise sensitivity $\widetilde{\xi}_D$, the minimum undersampling ratio of the DS is given in the following theorem.

THEOREM 8. *(Minimum undersampling for DS) Let α_K^2 be the SCV (8) and consider its CDF $F_\alpha(u)$ in (11) with the parameters q and β. Let $\widetilde{\xi}_D$ be the finite, desired noise sensitivity and $\epsilon > 0$ be a very small number. Then, the minimum undersampling of DS is given by*

$$\delta_D^*(\rho) = \inf\{\delta \ : \ \Pr(\xi_D \le \widetilde{\xi}_D) \to 1-\epsilon\} \ , \tag{15}$$

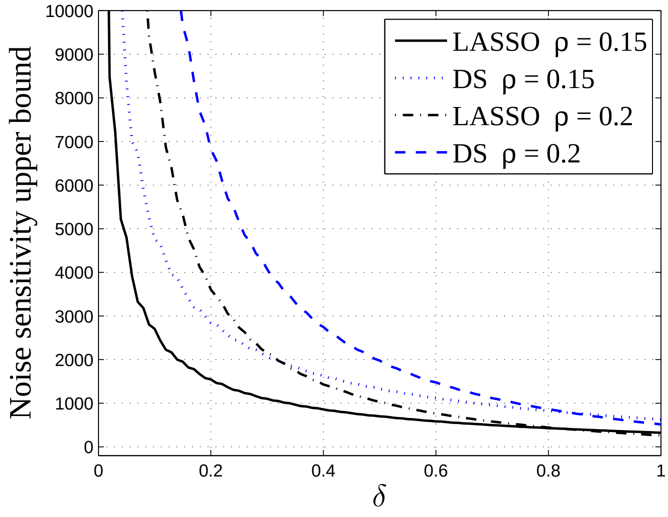

Fig. 2. Noise sensitivity upper bound of LASSO and DS for various ρ

Following the derivation steps for the LASSO, we find the noise sensitivity upper bound for the DS as

$$\widetilde{\xi}_D(\rho,\delta) = \frac{4\sqrt{\rho\delta N}\sqrt{2\log N}}{q_D\left[\ln\frac{1}{1-\epsilon}\right]^{\frac{2}{\delta N}}} \tag{16}$$

where q_D is q in (11) with K replaced by $4K$.
In Fig. 2, we plot the noise sensitivity upper bounds of the LASSO and the DS for various ρ. We note that for a given finite noise sensitivity there exists a minimum undersampling ratio above which the algorithms are guaranteed to be stable. For instance, for LASSO with $\rho = 0.2$ and for the noise sensitivity upper bound of 1000, the minimum undersampling ratio $\delta^*_L(\rho) = 0.5$. Thus, above the undersampling ratio of 0.5, the noise sensitivities of the LASSO are always bounded above by 1000. Comparing the LASSO and the DS, we note that the LASSO requires smaller undersampling ratios than the DS. For example, with $\rho = 0.15$ and for the noise sensitivity upper bound of 1000, the minimum undersampling required by the LASSO is $\delta^*_L(\rho) = 0.32$, where as for the DS it is $\delta^*_D(\rho) = 0.67$. We also note that for a fixed noise sensitivity, as ρ increases the minimum undersampling ratio increases as well. Thus, we illustrated that computing the precise values of the SCV enables us to investigate the minimum undersampling ability of the LASSO and the DS that were hidden in the existing theoretical stability results.

5. CONCLUSIONS

We analyzed the stability of two recovery algorithms in compressed sensing, namely, the LASSO and the Dangling selector (DS). We translated the existing theoretical stability results of the algorithms in terms of the minimum undersampling ratio by using the probability distributions of the squared constrained minimal singular values. We demonstrated that the LASSO requires smaller undersampling ratio than the DS for a given stability parameter.

APPENDIX

A. DISTRIBUTIONS OF CHI-SQUARE RANDOM VARIABLE

The PDF of a central Chi-square random variable with M degrees of freedom with $\sigma^2 = \frac{1}{M}$ is

$$p_C(x) = \begin{cases} \frac{\left(\frac{M}{2}\right)^{\frac{M}{2}}}{\Gamma\left(\frac{M}{2}\right)} x^{\frac{M}{2}-1} e^{-\frac{Mx}{2}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

and its corresponding CDF is given by $F_C(x) = \frac{\gamma\left(\frac{M}{2},\frac{Mu}{2}\right)}{\Gamma\left(\frac{M}{2}\right)}$.

B. PROOF OF THEOREM 6

As $n \to \infty$, the CDF of n i.i.d. Chi-square (special case of Gamma) random variables converges to the Weibull [4, Table 9.5]. It remain to find the (location, scale and shape) of the distribution using $F_C(u)$ in Appendix A.
Location constant, p_n: For the Weibull $p_n = \inf\{u : F_C(u) > 0\}$. Since $F_C(u)$ is supported on $[0,\infty)$, $p_n = 0, \forall n$.
Scale constant, q_n: For the Weibull, the scale constant is given by $q_n = F_C^{-1}(\frac{1}{n}) - p_n$. We find $F_C^{-1}(\frac{1}{n})$ by expanding the CDF $F_C(u)$ in Taylor series as

$$F_C(u) = \frac{\gamma\left(\frac{M}{2},\frac{Mu}{2}\right)}{\Gamma\left(\frac{M}{2}\right)} = \frac{1}{\Gamma\left(\frac{M}{2}\right)} \sum_{i=0}^{\infty} \frac{(-1)^i \left(\frac{Mu}{2}\right)^{\left(i+\frac{M}{2}\right)}}{i!\left(i+\frac{M}{2}\right)} \tag{18}$$

We need to find u such that $F_C(u) = \frac{1}{n}$. Since $\frac{1}{n}$ approaches a very small value as $n \to \infty$, the value of u must be very small as well. So, we approximate $F_C(u)$ by using only the first term of the series as $F_C(q_n) = \frac{1}{n} \approx \frac{\left(\frac{Mq_n}{2}\right)^{\left(\frac{M}{2}\right)}}{\frac{M}{2}\Gamma\left(\frac{M}{2}\right)}$ from which we determine $q_n = \frac{2}{M}\left[\frac{\frac{M}{2}\Gamma\left(\frac{M}{2}\right)}{n}\right]^{\frac{2}{M}}$. We observed via Newton iterative method that inclusion of more terms in the approximation scales q_n up to a factor of 3. Hence, we multiply q_n by 3. With $n = N_S = \binom{N}{K}$ and expressing q_n in terms of logarithms we obtain the scale parameter.
Shape constant, β: The shape parameter is the positive exponent of the result obtained by evaluating the following limit [6]:

$$\lim_{z\to\infty} \frac{F_C\left(p_n - \frac{1}{zu}\right)}{F_C\left(p_n - \frac{1}{z}\right)}. \tag{19}$$

The limit evaluates to $u^{-\frac{M}{2}}$ and thus, $\beta = \frac{M}{2}$.

C. REFERENCES

[1] Barry C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*. SIAM, Philadelphia.

[2] Emmanuel J. Candes. The restricted isometry property and its implications for compressed sensing. *C. R. Acad. Sci. Paris, Ser. I*.

[3] Emmanuel J. Candes and Terence Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*.

[4] Enrique Castillo, Ali S. Hadi, N. Balakrishnan, and Jose M. Sarabia. *Extreme Value and Related Models with Applications in Engineering and Science*. John Wiley and Sons, New York.

[5] L. Canto E Castro. Uniform rates of convergence in extreme-value theory: Normal and gamma models. *Annales de la Facult des sciences de l'Universit de Clermont, Srie Probabilits et applications*.

[6] Paul Embrechts, Claudia Kluppelberg, and Thomas Mikosch. *Modeling Extremal Events for Insurance and Finance*. Springer Verlag, New York.

[7] I.S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academics Press, San Diego.

[8] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Chapman and Hall, Florida.

[9] Hwan-Chol Jang, Chang-Hyeong Yoon, Eui-Heon Chung, Won-Shik Choi, and Heung-No Lee. Speckle suppression via sparse representation for wide-field imaging through turbid media. *Optics Express*.

[10] Gongguo Tang and Arye Nehorai. Performance analysis of sparse recovery based on constrained minimal singular values. *IEEE Transactions on Signal Processing*.

[11] Brendt Wohlberg. Noise sensitivity of sparse signal representations: Reconstruction error bounds for the inverse problem. *IEEE Transactions on Signal Processing*.

[12] Hui Zhang and Lizhi Cheng. On the constrained minimal singular values for sparse signal recovery. *IEEE Signal Processing Letters*.