On Eigenvalues of Wishart Matrix for Analysis of Compressive Sensing Systems

Oliver James and Heung-No Lee, Senior Member, IEEE

Abstract—We consider the problem of analyzing compressive sensing (CS) systems using the eigenvalues of Wishart matrices. Towards this goal, we first aim to derive new eigenvalue distributions for Wishart matrices. We next discuss the usefulness of the distributions in the context of sparse signal recovery in CS systems. Sparse signal recovery necessitates sensing matrices with a good restricted isometry constant (RIC). We address the problem of existence of a Gaussian sensing matrix with a prescribed RIC. In this regard, we first motivate and propose a new framework that hinges on the relation between the RIC and the eigenvalues of Gaussian sensing matrices. We then derive a condition in terms of undersampling ratio for the existence of an ensemble of Gaussian matrices. Adopting our framework, we determine the sufficient undersampling ratios for various RIC conditions prescribed for orthogonal matching pursuit (OMP) algorithm. We demonstrate that for a given RIC condition there exists a sharp threshold on the undersampling ratio above which the probability to find a good Gaussian matrix is 1.

Index Terms — Compressive Sensing, Eigenvalues, Orthogonal Matching Pursuit (OMP), Restricted Isometry Constant (RIC), Sensing matrices, Wishart distribution.

I. INTRODUCTION

Eigenvalues of random matrices summarize the macroscopic properties of the engineering and scientific systems in the form of probability distributions. Hence, many researches aim to find the probability distributions of the eigenvalues of the random matrices. In multivariate statistics [1] and in wireless MIMO communications [2], the eigenvalue distributions of random Wishart matrices are widely used for analysis. Let \( A_x \) be an \( M \times K \) real or complex-valued Gaussian matrix. Then, the \( K \times K \) matrix \( W = A_x^H A_x \) is called a Wishart matrix, where \( A_x^H \) denotes the conjugate transpose of \( A_x \).

In many multivariate applications such as financial data analysis, genetic studies, search engines, and climate studies, the study of sample covariance matrices is fundamental and real-valued Wishart matrices are good models of the sample covariance matrices [3]. Hence, the eigenvalues of the real-valued Wishart matrices play a crucial role in multivariate applications. On the other hand, the applications such as MIMO communications require eigenvalues of complex-valued Wishart matrices in order to analyze information theoretic capacity.

In general, a marginal eigenvalue distribution is obtained by integrating the joint eigenvalue distribution over all other eigenvalues except the required one. The joint eigenvalue distribution for the real [4, eq. (58)] and the complex Wishart matrices [4, eq. (95)] have been known for nearly five decades. However, the marginal distributions are known, or can be expressed in tractable form only for some special cases, such as the largest or the smallest eigenvalues.

The first known largest eigenvalue distribution for complex Wishart matrices was derived by Khatri [5] in terms of the product of beta integrals. In [6], Krishnaiah and Chang derived the smallest eigenvalue distribution (for the complex case) in terms of infinite series of Zonal polynomials. An explicit and usable form of Zonal polynomials is available only for small polynomial orders [4, p. 498]. In addition, Zonal polynomials are notoriously difficult to compute. It has been shown in [6] that a series that involves Zonal polynomials converges very slowly. These problems lead to other directions for deriving the eigenvalue distributions during the late 60s. A notable direction was to find a cumulative distribution rather than the probability distribution. Khatri [7] derived a cumulative distribution function for the largest and smallest eigenvalues in terms of a matrix determinant. Ming and Alouini [8] constructed a probability distribution from the cumulative distribution using the idea of derivative of a determinant. However, it has been shown [8, p.1418] that the closed-form probability distribution can be derived only for the special case of a \( 2 \times 2 \) matrix.

Just like the complex-valued case, the extreme eigenvalues of the real-valued Wishart matrices are also derived in terms of Zonal polynomials by Sugiyama and Fukutomi [9]. In [10], Edelman derived the smallest and largest eigenvalue distributions in terms of Tricomi functions. Edelman gave the exact distributions only for a certain matrix sizes leaving behind a recursive expression for computing the distributions [10, p. 45]. The complexity of calculating the recursive functions is questionable for large matrix sizes. In order to alleviate the burden of computing the close-form expression for eigenvalue distributions, limit distributions of the eigenvalues are usually derived [10-14]. These limit distributions show the behavior of the eigenvalues only when the system size tends to infinity and they may not necessarily capture the effects of finite size matrices that occur in practice. Thus, there arises a need to derive novel, tractable eigenvalue distributions for the eigenvalue based analysis of systems.

In [15], Chiani, Win and Zanella attempted to derive new, closed-form expressions for the marginal distributions of eigenvalues for complex Wishart matrices. Their idea is to first express the joint eigenvalue distribution as a product of two determinants. They then derived the marginal distributions by evaluating the multi-dimensional integration.
of the product of determinants. In [16], Chiani, Win and Zanella again used this same idea to derive the distributions for any arbitrary eigenvalue. Until now, the distributions given by Chiani, Win and Zanella [15, 16] are the best available closed-form expressions for the complex Wishart matrices. Due to the simplicity and tractability of these distributions, they are conveniently used to evaluate the information theoretic capacity of MIMO systems [16].

It is well-known [1, 10] that the joint eigenvalue distributions for both complex and real-valued Wishart matrices are two different real-valued functions. The way the difference arrived between these two joint eigenvalue distributions from their Wishart matrix distribution is succinctly summarized in Theorem 3.1 [10, p. 33] and Theorem 3.2 [10, p. 36]. Theorem 3.1 discusses the transformations between the volume elements of a Wishart matrix and their eigenvalues. Using these transformations, Theorem 3.2 presents a way to convert the distribution of the Wishart matrices into their joint eigenvalue distribution. Thus, the joint eigenvalue distribution of real-valued Wishart matrices cannot be obtained simply from their complex-valued counterpart. While Chiani, Win and Zanella [16] derived the eigenvalue distributions for the complex-valued case, we derive them, in this paper, for the real-valued case.

In this paper, we aim to derive new eigenvalue distributions for the real-valued Wishart matrices. We derive the distributions borrowing the tools from the theory of skew-symmetric matrices. In particular, we show the eigenvalue distributions in terms of Pfaffian of a skew-symmetric matrix. We observe that unlike the complex-valued case [15, 16], the distributions of eigenvalues for the real-valued case are different for odd and even matrix orders. We show that our distributions can be readily calculated for large matrix orders.

We next illustrate the usefulness of the derived distributions in the context of compressive sensing (CS). Especially, we discuss a new framework by using the relation between the eigenvalues of a Gaussian sensing matrix and its restricted isometry constant (RIC). Adopting our new framework, we obtain sufficient undersampling ratios for the existence of Gaussian sensing matrices with a prescribed RIC. We evaluate the sufficient undersampling ratios for various RIC-based conditions advised for the OMP algorithm. We finally demonstrate an interesting tradeoff that exists between the undersampling ratio and the RIC-based condition.

This paper is organized as follows. Section II reviews the joint distribution of the eigenvalues of Wishart matrices and Section III presents the derivation of the extreme eigenvalue distributions. A new eigenvalue framework for the analysis of CS systems is developed in Section IV and Section V concludes the paper.

Notations: Capital letters represent matrices while bold face small letter denote vectors. The symbol $|A|$ denotes the matrix determinant if $A$ is a matrix and it denotes the set cardinality if $A$ is a set. The symbol $A'$ represents the transpose of $A$, $\text{diag}(a_1, a_2, \ldots, a_K)\,$ denotes a $K \times K$ diagonal matrix with diagonal entries $a_1, a_2, \ldots, a_K$.

II. PRELIMINARIES

We first review the joint distribution of eigenvalues of the real-valued Wishart matrix [1], which is used to derive the new eigenvalue distributions in the next section.

Let $A$ be an $M \times N$ real-valued Gaussian sensing matrix and $A_k$ be a random collection of $K$ columns of $A$. We assume $K < M < N$. Then, we note that $W := A_k^*A_k$ is a $K \times K$ real-valued Wishart matrix, where $A_k$ is an $M \times K$ real-valued Gaussian matrix. We note that $M > K$ corresponds to a class of full-rank Wishart matrices and we deal with these matrices in this paper. The covariance matrix of each column of $A_k^*$ is given by $\Sigma = E[a_k a_k^*]$. In this paper, we consider an uncorrelated Gaussian matrix, i.e., $\Sigma = \rho I_K$, where $\rho$ is a scalar and the determinant $|E| = \rho^K$.

Let $\lambda_1 > \lambda_2 > \cdots > \lambda_K$ represent the ordered eigenvalues of the Wishart matrix $W$. We define $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_K)$ and thus, $|\Lambda| = \prod_{i=1}^K \lambda_i$. The joint distribution of the eigenvalues of a real-valued Wishart matrix is then given by [1, p. 107]

$$f(\lambda) = \frac{\pi^{K/2}}{2\Gamma_K(\frac{M}{2})\Gamma_K(\frac{N}{2})} (\prod_{i,j}^{K} \lambda_{ij}) \cdot |\Sigma|^{\frac{1}{2}} |\Lambda|^{-M-N-1} \prod_{i,j}^{K} (\lambda_i - \lambda_j),$$

(1)

where $\Gamma_p(a)$ is a multivariate gamma function [1] given by

$$\Gamma_p(a) = \pi^{p(p-3)/4} \prod_{i=1}^{p} \Gamma(a - \frac{1}{2}(i-1)).$$

In (1), the term $\varphi F_0(-\frac{1}{2}\Sigma^{-1}, \Lambda)$ represents the hyper-geometric function defined as [1, p. 107]

$$\varphi F_0(-\frac{1}{2}\Sigma^{-1}, \Lambda) = \prod_{i=1}^{K} \exp(-\frac{\lambda_i}{2}).$$

(2)

Equation (2) is valid only for the case $\Sigma = \rho I_K$. When $\Sigma \neq \rho I_K$, that is, for a correlated case, it is difficult to derive $\varphi F_0(-\frac{1}{2}\Sigma^{-1}, \Lambda)$ [17, p. 445]. However, for some special case of correlated $\Sigma$, (2) can be expressed in terms of an infinite series of zonal polynomials [17]. Since closed-form expression for the hyper-geometric function for correlated case is unavailable, deriving the closed-form marginal distributions of the real-valued correlated Wishart matrices still remains an open problem. On the other hand, for the correlated complex-valued Wishart matrices, the term $\varphi F_0(-\frac{1}{2}\Sigma^{-1}, \Lambda)$ can be evaluated using Harish-Chandra–Itzykson–Zuber integral [17]. In (1), the product term $\prod_{i,j}^{K} (\lambda_i - \lambda_j)$ represents the determinant $|V(\Lambda)|$ of a Vandermonde matrix whose $(i,j)$th entry is $v_{ij}(\lambda_i) = \lambda_i^{k-j}$. The matrix $V(\Lambda)$ is given below.
By substituting (2) into (1), the joint density can be written in
the form of a matrix determinant as

$$f(\lambda) = \frac{\pi^{K^2/2} \cdot \Gamma(K/2)^2}{\Gamma(K)^2} \left| V(\lambda) \right| \prod_{i=1}^{K-1} \lambda_i^{(i-M-K)/2} e^{-\frac{\lambda_i}{\lambda_i+\lambda_{i+1}}}$$

$$= c \left| V(\lambda) \right| \prod_{i=1}^{K} \xi(\lambda_i).$$

(4)

In (4), we arrange the joint distribution as a product of a
constant, a Vandermonde determinant and the product of
functions $\xi(\lambda_i)$. We arrange the joint distribution in this
form for the ease of deriving the extreme eigenvalue
distributions as discussed in the next section. We note that
the joint distribution for the complex-valued case [16, p. 1051] is different from the joint distribution (4) for the
real-valued case.

III. DERIVATION OF EXTREME EIGENVALUE DISTRIBUTIONS

In this section, we aim to derive the largest and the smallest
eigenvalue distributions of a Wishart matrix from the
joint distribution. In our approach, we first expand the
determinant in (4) by using the generalized determinant
expansion. This expansion results in a multi-dimensional
integration of a determinant. We perform the multi-dimensional integration using the results from the
theory of skew-symmetric matrices.

Let $f(\lambda_i)$ and $f(\lambda_K)$ represent the largest and the
smallest eigenvalue distributions of the Wishart matrix,
respectively. These distributions can be written in terms of
$f(\lambda)$ as the following $K-1$ dimensional ordered integrals:

$$f(\lambda_i) = \int_{0}^{\lambda_i} \cdots \int_{0}^{\lambda_i} f(\lambda) \prod_{j<i} \lambda_j \prod_{j>i} \lambda_j d\lambda_j d\lambda_i.$$  

(5)

$$f(\lambda_K) = \int_{\lambda_K}^{\infty} \cdots \int_{\lambda_K}^{\infty} f(\lambda) \prod_{j<i} \lambda_j \prod_{j>i} \lambda_j d\lambda_j d\lambda_i.$$  

(6)

In order to derive $f(\lambda_i)$ and $f(\lambda_K)$, we substitute $f(\lambda)$
from (4) into (5) and (6) and obtain

$$f(\lambda_i) = c \int_{0}^{\lambda_i} \cdots \int_{0}^{\lambda_i} \left| V(\lambda) \right| \prod_{j<i} \xi(\lambda_j) \prod_{j>i} \lambda_j d\lambda_j d\lambda_i,$$  

(7)

$$f(\lambda_K) = c \int_{\lambda_K}^{\infty} \cdots \int_{\lambda_K}^{\infty} \left| V(\lambda) \right| \prod_{j<i} \xi(\lambda_j) \prod_{j>i} \lambda_j d\lambda_j d\lambda_i.$$  

(8)

In [9], the $(K-1)$ dimensional integrals in (7) and (8) are
evaluated using the theory of Zonal polynomials. In
particular, the integrals are transformed into a slowly
decaying infinite series of Zonal polynomials. As mentioned
in the introduction, an explicit usable form of Zonal
polynomials is available only for orders up to 6 [4, p. 498]
that makes the distributions highly difficult to compute. In
[10], the $(K-1)$ dimensional integrals are evaluated using the
theory of differential equations. Specifically, the integrals
are shown [10, p. 41] to satisfy the solution of the Tricomi
differential equations and thus, the distributions are given in
terms of Tricomi functions. However, the exact distributions
are given only for the values of $M = K + 1, K + 2, K + 3$. For
$M > K + 3$, no exact distribution can be found [10, p. 45].
In [9], we are able to give the exact distributions for all
values of $K$ and $M$, and our distributions can be computed
readily.

The integration in (7) is over the variables $\lambda_2, \cdots, \lambda_K$.
However, the integrand, which contains the determinant
$\left| V(\lambda) \right|$ and the product $\prod_{i=1}^{K} \xi(\lambda_i)$, is a function of
$\lambda_1, \lambda_2, \cdots, \lambda_K$. In (7), in order to integrate conveniently, we
separate the variable $\lambda_i$ both from the determinant as well as
from the product. It is easy to separate $\lambda_i$ from the product.
In order to separate $\lambda_i$ from the determinant, we use the
generalized determinant expansion provided in Lemma 1. A
similar procedure can be followed to separate the variable $\lambda_K$
from the integrand in (8), because the integration in (8) is
over the variables $\lambda_1, \lambda_2, \cdots, \lambda_{K-1}$.

Lemma 1 [17]: Let $V(\lambda)$ be a $K \times K$ matrix with the
$(i, j)$th element denoted by $V_{ij}(\lambda)$, a function of $\lambda_i$. The
determinant $\left| V(\lambda) \right|$ can be expanded along the $k$th row,
$k = 1, 2, \cdots, K$ as

$$\left| V(\lambda) \right| = \sum_{n=0}^{K} (-1)^n V_{nk}(\lambda) \left| V_{nk}(\lambda) \right|.$$  

(9)

where $V_{nk}(\lambda)$ is an $(K-1) \times (K-1)$ sub-matrix obtained by
deleting the $k$th row and $n$th column of $V(\lambda)$. The
sub-matrices $V_{nk}(\lambda)$, $n = 1, 2, \cdots, K$, are independent of $\lambda_k$.

In order to understand Lemma 1, we consider an example
with $K = 3$ and $V_{ij}(\lambda) = \lambda_i^{K-1}$. The matrix $V(\lambda)$ and its
determinant expansion along the first row ($k = 1$) are given
respectively as

$$V(\lambda) = \begin{bmatrix} \lambda_1^2 & \lambda_1 & 1 \\ \lambda_2^3 & \lambda_2 & 1 \\ \lambda_3^3 & \lambda_3 & 1 \end{bmatrix},$$  

(10)

And

$$\left| V(\lambda) \right| = \lambda_1^2 \lambda_2 + \lambda_1 \lambda_3 + 1 = \sum_{n=1}^{K} (-1)^n \lambda_n^{K-n}.$$  

(10)
In (10), the sub-matrices $V_n(\lambda)$, $n = 1, 2, 3$ are functions of the variables $\lambda_1$ and $\lambda_2$ only and independent of $\lambda_3$. Thus, we pulled the variable $\lambda_3$ out from the determinant expansion. Using Lemma 1, we can pull $\lambda_1$ or $\lambda_2$ out from the determinant expansion. Expanding $|V(\lambda)|$ in (7) along the first row for the largest eigenvalue $\lambda_1$, and by moving $\xi(\lambda_1)$ outside of the product, we obtain

$$f(\lambda) = \epsilon \sum_{n=1}^{K-n} (-1)^{n+1} \lambda_1^{K-n} \xi(\lambda_1)$$

$$\cdot |V_n(\lambda)| \prod_{i=2}^{K} \xi(\lambda_i) \ d\lambda_2 \cdots d\lambda_3 \ d\lambda_2.$$ (11)

In (11), for each $n$, we can combine the product $|V_n(\lambda)| \prod_{i=2}^{K} \xi(\lambda_i)$ into a single term by using the scaling property of the determinant, i.e.,

$$|B_n(\lambda)| = |V_n(\lambda)| \prod_{i=2}^{K} \xi(\lambda_i).$$ (12)

The scaling property dictates that for two matrices, $B$ and $V$, and a constant $\xi$, if $|B| = \xi |V|$, then the matrix $B$ can be obtained by multiplying a row or a column of $V$ by the same constant $\xi$. In the R.H.S of (12), we have a determinant which is multiplied by a product. Using the scaling property, the $(K-1)\times(K-1)$ matrix $B_n(\lambda)$ can then be obtained by multiplying $k$th row of the matrix $V_n(\lambda)$ with the $k$th term of the product. We illustrate (12) with the example in (10).

The product $|V_n(\lambda)| \prod_{i=2}^{K} \xi(\lambda_i)$ for $n = 1$ is given by

$$|B_1(\lambda)| = |V_1(\lambda)| \prod_{i=2}^{K} \xi(\lambda_i).$$ (13)

From (10), $|V_1(\lambda)| = \begin{bmatrix} \lambda_2 & 1 \\ \lambda_3 & 1 \end{bmatrix}$ and thus

$$|B_1(\lambda)| = \begin{bmatrix} \lambda_2 & 1 \\ \lambda_3 & 1 \end{bmatrix} \begin{bmatrix} \xi(\lambda_2) \xi(\lambda_3) \\ \xi(\lambda_3) \xi(\lambda_3) \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_2 \xi(\lambda_2) \xi(\lambda_3) \\ \lambda_3 \xi(\lambda_3) \xi(\lambda_3) \end{bmatrix} \begin{bmatrix} \lambda_2 \xi(\lambda_2) \\ \lambda_3 \xi(\lambda_3) \\ \lambda_3 \xi(\lambda_3) \end{bmatrix},$$ (14)

where in the second step of (14) the terms $\xi(\lambda_2)$ and $\xi(\lambda_3)$ multiply the first and second rows of $V_1(\lambda)$, respectively. Using (12), we rewrite (11) after interchanging the summation and integration as

$$f(\lambda) = \epsilon \sum_{n=1}^{K-n} (-1)^{n+1} \lambda_1^{K-n} \xi(\lambda_1)$$

$$\cdot \prod_{i=2}^{K} \xi(\lambda_i) \ d\lambda_2 \cdots d\lambda_3 \ d\lambda_2.$$ (15)

We will continue to work with (15), the distribution of the largest eigenvalue. For the smallest eigenvalue $\lambda_k$, we follow the similar procedure later on. In (15), for each value of $n$, we need to find a $K-1$ dimensional integration of the determinant of an $(K-1) \times (K-1)$ matrix $B_n(\lambda)$. This $K-1$ dimensional integration for an $n$ can be evaluated using a result from the theory of skew-symmetric matrices provided in the following lemma. A square matrix $A$ is skew-symmetric if $A = -A^t$. This implies that the $(i, j)$th element of $A$ satisfies $a_{ij} = -a_{ji}$, and the diagonal elements are zero, i.e., $a_{ii} = 0$. A $K \times K$ matrix is also called a matrix of order $K$.

**Lemma 2 [18, N. G. De Bruijn]:** Let $\theta(\lambda_1)$ denote the $(i, j)$th entry of the $(K-1) \times (K-1)$ matrix $B_n(\lambda)$. Then,

$$\int \prod_{i=2}^{K} \xi(\lambda_i) \ d\lambda_2 \cdots d\lambda_3 \ d\lambda_2 = \text{PF}(B_n),$$ (16)

where $\text{PF}(B_n)$ is the Pfaffian of a $(K-1) \times (K-1)$ skew-symmetric matrix $B_n$. The entries of $B_n$ is calculated using the diagonal entries of the matrix $B_n(\lambda)$. This calculation is different for odd and even $K$.

**Odd $K$:** In this case, the matrix $B_n$ is of order $K-1$. The $(i, j)$th element of $B_n$ is given by

$$b_{i,j} = \int \theta(\lambda_1) \theta(\lambda_j) \ d\lambda_i \ d\lambda_j,$$ (17)

for $i, j = 1, 2, \ldots, K-1$.

**Even $K$:** In this case, the matrix $B_n$ is of order $K+1$. In addition to (17), $B_n$ has two more rows at the bottom and two more columns at the right whose non-zero values are calculated as

$$b_{i,k} = 1; \quad i = 1, 2, \ldots, K-1$$

$$b_{i,k+1} = \int \theta(\lambda_1) \ d\lambda_i; \quad i = 1, 2, \ldots, K$$ (18)

From Lemma 2, we note that the $K-1$ dimensional integration of $|B_n(\lambda)|$ is reduced to the calculation of Pfaffian of the skew-symmetric matrix $B_n$, whose entries can be obtained by using (17) and (18). The $|\text{PF}(B_n)|$ can be obtained using the relation $|\text{PF}(B_n)| = \sqrt{\det B_n}$ as $\det B_n \geq 0$ for skew-symmetric matrices. Thus, applying Lemma 2 to (15), we can find the distribution of the largest eigenvalue in terms of $\text{PF}(B_n)$. Similarly, an expression for the distribution of smallest eigenvalue $\lambda_k$ can be obtained by expanding the determinant in (8) along the last row and then defining a determinant $|D_n(\lambda)| = |V_{nk}(\lambda)| \prod_{i=2}^{K} \xi(\lambda_i)$ as
\[
    f(\lambda_k) = c \sum_{n=1}^{K} (-1)^{n+K} \lambda_k^{K-n} e^{\frac{\lambda_k}{\nu}} \xi(\lambda_k) \]

where \( c = \frac{\sqrt{\pi}}{\Gamma(K/2)} \) is a normalization constant. We can evaluate the integral for \( \lambda_k \) in the interval from \( -\infty \) to \( \infty \).

To find the \( K-1 \) dimensional integration in (19), Lemma 2 can still be used except that the interval of the integration in Lemma 2 ranges from \( \lambda_k \) to \( \infty \). In the subsequent discussions, we find the exact distribution for the largest and smallest eigenvalues.

A. Largest eigenvalue distribution

Using Lemma 2 in (15), the distribution of the largest eigenvalue \( f(\lambda) \) is given by

\[
    f(\lambda) = c \sum_{n=1}^{K} (-1)^{n+1} \lambda^{K-n} e^{\frac{\lambda}{\nu}} \xi(\lambda) \]

where \( c = \frac{\sqrt{\pi}}{\Gamma(K/2)} \) is a normalization constant. We can evaluate the integral for \( \lambda_k \) in the interval from \( -\infty \) to \( \infty \).

In (20), in order to arrive at the second step, we substitute the expression \( \xi(\lambda_k) = \lambda_k^{K-n} e^{\frac{\lambda_k}{\nu}} \). To evaluate \( \text{PF}(B_k) \) for each value of \( n \), we need to first construct \( B_k \) by calculating its entries \( b_{ij} \) given in (17) and (18). Evaluation of \( b_{ij} \) needs \( \theta(\lambda) \), the diagonal element of \( B(\lambda) \). The diagonal element is given by \( \theta(\lambda_i) = \lambda_i^{1+K/n} e^{\frac{\lambda_i}{\nu}} \).

By substituting the value of \( \theta(\lambda_i) \) in (17), we obtain

\[
    b_{ij} = \int_{\lambda_i}^{\lambda_j} \lambda_i^{1+K/n} e^{\frac{\lambda_i}{\nu}} \lambda_j^{1+K/n} e^{\frac{\lambda_j}{\nu}} d\lambda_i = \frac{\Gamma\left(r_{ij} + \frac{M-K+1}{2}, \frac{\lambda_j}{\nu}\right)}{\Gamma\left(r_{ij} + \frac{M-K+1}{2}, \frac{\lambda_i}{\nu}\right)} e^{-\frac{\lambda_i}{\nu}} \gamma\left(r_{ij} + \frac{M-K+1}{2}, \frac{\lambda_j}{\nu}\right) \]

where \( \gamma(v, \mu) = \int_{0}^{\infty} x^{v-1} \mu^\mu e^{-\mu x} dx \) is the lower incomplete gamma integral. Similarly, substituting \( \theta(\lambda_i) \) in (18), we get

\[
    b_{ij} = \int_{\lambda_i}^{\lambda_j} \lambda_i^{1+K/n} e^{\frac{\lambda_i}{\nu}} \lambda_j^{1+K/n} e^{\frac{\lambda_j}{\nu}} d\lambda_i = \frac{\Gamma\left(r_{ij} + \frac{M-K+1}{2}, \frac{\lambda_j}{\nu}\right)}{\Gamma\left(r_{ij} + \frac{M-K+1}{2}, \frac{\lambda_i}{\nu}\right)} e^{-\frac{\lambda_i}{\nu}} \gamma\left(r_{ij} + \frac{M-K+1}{2}, \frac{\lambda_j}{\nu}\right) \]

By evaluating the integrals in (23) and (24), we obtain the entries of the matrix \( B_n \). The value of \( \text{PF}(B_n) \) can then be calculated using the relation \( \text{PF}(B_n) = \sqrt{\text{det} B_n} \).

B. Smallest eigenvalue distribution

Similarly, in order to derive \( f(\lambda) \), we use Lemma 2 in (19) and obtain

\[
    f(\lambda) = c \sum_{n=1}^{K} (-1)^{n+1} \lambda^{K-n} e^{\frac{\lambda}{\nu}} \xi(\lambda) \]

where \( \xi(\lambda) = \lambda^{K-n} e^{\frac{\lambda}{\nu}} \). To evaluate the integral for \( \lambda_k \) in the interval from \( -\infty \) to \( \infty \), we need to first construct \( B_k \) by calculating its entries \( b_{ij} \) given in (17) and (18). Evaluation of \( b_{ij} \) needs \( \theta(\lambda) \), the diagonal element of \( B(\lambda) \). The diagonal element is given by \( \theta(\lambda_i) = \lambda_i^{K-n} e^{\frac{\lambda_i}{\nu}} \).

By substituting the value of \( \theta(\lambda_i) \) in (17), we obtain

\[
    b_{ij} = \int_{\lambda_i}^{\lambda_j} \lambda_i^{K-n} e^{\frac{\lambda_i}{\nu}} \lambda_j^{K-n} e^{\frac{\lambda_j}{\nu}} d\lambda_i = \frac{\Gamma\left(r_{ij} + \frac{M-K+1}{2}, \frac{\lambda_j}{\nu}\right)}{\Gamma\left(r_{ij} + \frac{M-K+1}{2}, \frac{\lambda_i}{\nu}\right)} e^{-\frac{\lambda_i}{\nu}} \gamma\left(r_{ij} + \frac{M-K+1}{2}, \frac{\lambda_j}{\nu}\right) \]

By evaluating the integrals in (23) and (24), we obtain the entries of the matrix \( B_n \). The value of \( \text{PF}(B_n) \) can then be calculated using the relation \( \text{PF}(B_n) = \sqrt{\text{det} B_n} \).

In summary, we have pursued the following steps in order to obtain the largest eigenvalue distribution in (20) from (11):

1. Find \( B_n(\lambda) \) in (12) using the scaling property of determinants.
2. Obtain \( B_n \) using the diagonal entries of \( B_n(\lambda) \) via (17) and (18).
3. Obtain the result of the \( K-1 \) dimensional integration of \( \left| B_n(\lambda) \right| \) as a Pfaffian of \( B_n \) via \( \text{PF}(B_n) = \sqrt{\text{det} B_n} \).

We followed the similar steps in order to obtain the smallest eigenvalue distribution in (25) from (19).
C. Verification and Remarks

In this section, we aim to verify the derived expressions (20) and (25), and provide two remarks.

Verification of the closed-form expressions: The numerical evaluation of the largest eigenvalue distribution in (20) are plotted in Fig. 1 for $K = 51$ and for various values of $M$ with $\rho = \frac{1}{5}$. Similar plots for the smallest eigenvalue distribution using (25) are shown in Fig. 2. In order to verify the exactness of plotted distributions, we compare them with the empirical distributions calculated as follows:

1. We generate an iid Gaussian matrix $A$ of size $M \times 5000$ for $M = 300, 500, \text{and} 700$.
2. We randomly draw a sub-matrix $A_k$ of size $M \times K$ from $A$ and find the largest and smallest eigenvalues of the matrix $A_k^T A_k$ using the MATLAB command $\text{eig}$. We call these values as the empirical values.
3. We repeat step 2 for 10000 times and obtain 10000 empirical values from which we compute the empirical probability distributions.

We report here that the plots of the empirical distributions exactly coincide with the numerical evaluation plots shown in Figs. 1 and 2. The empirical plots are not included in the figures in order not to make the figures busy.

**Confirmation of the large deviation result:** The large deviation result [12] says that as $M$ increases, 1) the largest and smallest eigenvalues are found near to 1, and 2) the shapes of the eigenvalues get sharper. This is confirmed from the plots in Figs. 1 and 2 in which as $M$ increase the eigenvalues move closer to 1 while their shapes get sharper.

**Remark 1.** Our results are readily computable compared to previous results in [10]. The distributions in [10, p. 53] are plotted for $K = 3$ and $M = 27$. However, we plot them for $K = 51$ and $M = 700$. As per our knowledge, we have not seen the eigenvalue distributions computed and plotted for the range of values we have shown in these figures. For values of $K$ and $M$ much larger than that of those shown in the figures, the computation of the distributions is still difficult due to the numerical evaluation steps, i.e., (23)-(24) and (26)-(27).

**Remark 2.** In order to be able to handle the distributions for large values of $K$ and $M$ that typically occur in CS, we take the direction of obtaining a tight upper bound for the distribution by using the Hadamard maximum determinant theorem (see Appendix). By this upper bound, we can tap on the probability of finding a well-conditioned Gaussian matrix (see Section IV.F).

**Fig. 2. Smallest eigenvalue distribution of Wishart matrix for various $M$.**

IV. APPLICATIONS OF THE EIGENVALUE DISTRIBUTIONS IN COMPRESSIVE SENSING

A. Background

In the traditional signal processing approach, signal samples are acquired first and then the acquired samples are compressed to a fewer number of samples by exploiting the redundancy. Recently, researchers have asked the question: Why should we acquire redundant samples when they have to be discarded during compression [19]? This question leads to a new signal processing paradigm called compressive sensing (CS). In CS, signal samples are compressed during the procedure of signal acquisition. As a result, with CS it is easy to acquire only sufficient number of (compressed) samples or measurements during the acquisition. Thus, CS combines the acquisition and the compression in a single step. This one step process is performed using a suitable sensing matrix.

CS theory dictates that a signal can be reconstructed from its compressed samples, provided the signal is sparse. A sparse signal has a fewer number of non-zero values. Let $x$ denote an $N \times 1$ signal vector. The vector $x$ is said to be $K$-sparse, if it contains exactly $K$ non-zero entries. If a signal vector is not sparse, it can be represented as sparse in a certain transform basis. In CS, compressed samples, that is, the measurement vector is obtained using a sensing matrix. Let $A$ denote an $M \times N$ sensing matrix with $M < N$ and $\frac{M}{N}$ is called undersampling ratio. Then, the acquired $M \times 1$ measurement vector can be modeled as $y = Ax$.

In order to reconstruct the sparse signal vector $x$ from $y$, either greedy-type or $L_1$-norm minimization-based algorithms are employed in CS. The recovery performance of these algorithms depends on the type of the sensing matrix used [20]. Thus, to quantify the goodness of a sensing matrix in signal recovery, Candès and Tao [21] introduced the restricted isometry property (RIP), and its associated metric called the restricted isometry constant (RIC).

B. Restricted Isometry Property (RIP)

In this section, we introduce the RIP and the RIC and discuss their importance in CS. Let $x$ be a $K$-sparse signal vector. The set of locations of the non-zero entries of $x$ is
called the support set of \( x \) and it is denoted by \( \mathcal{X} \). For \( K \)-sparse signals, \( |\mathcal{X}| = K \).

**Definition 3 Restricted isometry property [21, Candes and Tao]**: Let \( x \) be a \( K \)-sparse signal, and \( A \) be an \( M \times N \) deterministic matrix with unit normalized columns and \( M < N \). Let \( A_x \) denote an \( M \times K \) sub-matrix whose column indices \( j \in \mathcal{X} \). Let \( x_j \) denote a \( K \times 1 \) sub-vector taken from \( x \). For a given integer \( K \), let \( \delta_k \) be the smallest quantity such that \( A_x \) obeys

\[
(1 - \delta_k) \leq \frac{||A_x x||}{||x||} \leq (1 + \delta_k)
\]

for all \( \mathcal{X} \) and for all real values of \( x \).

The ratio in (28) is bounded between the two numbers, \( 1 - \delta_k \) and \( 1 + \delta_k \), where the scalar \( \delta_k \) is the RIC. If a sensing matrix \( A \) satisfies (28) with \( \delta_k < 1 \) then \( A \) is said to obey the RIP of order \( K \) with the RIC \( \delta_k \). A sensing matrix with a prescribed \( \delta_k \) guarantees perfect reconstruction of a sparse signal. Smaller the \( \delta_k \), better is the reconstruction accuracy.

We note from (28) that determining the RIC for a given matrix \( A \) is tedious. First, the ratio \( \frac{||A_x x||}{||x||} \) should be calculated for all \( \binom{N}{K} \) support sets; this is an NP-hard problem. Second, the ratio depends on the signal amplitude \( x \) and hence it should be calculated for all possible signal values.

We observe from (28) that when the matrix \( A \) is random, the RIC \( \delta_k \) becomes a random variable. For the random matrices, probabilistic statements about the RIP are provided in [21, 22]. Such statements allow us to bypass the burden of calculating the RIC. Candes [21] and Baraniuk et al. [22] showed that Gaussian matrices satisfy (28) with overwhelming probability as the sensing matrix dimensions tends to infinity.

**C. RIC-based performance conditions**

In this section, we discuss the conditions given in terms of the RIC for the stable recovery of sparse signals by the CS algorithms. In CS, there are two categories of recovery algorithms, namely, \( L_1 \)-norm minimization-based algorithms and greedy algorithms.

1) **RIC conditions for \( L_1 \)-algorithms**

Assuming a deterministic sensing matrix, Candes and Tao [21, Lemma 1.2] illustrated that a sensing matrix with RIC \( \delta_k < 1 \) guarantees a unique \( L_0 \) solution. Candes [23] showed that if \( \delta_k < 0.41 \), then \( L_1 \) minimization exactly recovers every \( K \)-sparse signal. Foucart and Lai [24] improved the condition on RIC for the recovery of \( K \)-sparse signals to \( \delta_k < 0.4531 \). By improvement on RIC, Foucart and Lai mean lesser strict condition than that of Candes with the same recovery guarantee. Several such recent improvements on RIC have been summarized in [25].

2) **RIC conditions for greedy algorithms**

Just like the \( L_1 \)-norm-based algorithms, the sufficient RIC conditions for the greedy algorithms have also been reported in the literature [26-30]. Davenport and Wakin [26, Theorem 3.1] established that if a matrix satisfies the RIC \( \delta_{k+1} < \frac{1}{\sqrt{K}} \), then the OMP exactly recovers every \( K \)-sparse signals in \( K \)-steps. Huang and Zhu [27] improved this bound to \( \delta_{k+1} < \frac{1}{\sqrt{2K}} \), which was further improved to \( \delta_{k+1} < \frac{1}{\sqrt{3K}} \) by Liu and Temlyakov [28]. Recently, Mo and Shen [29], Wang and Shim [30] independently derived the much improved condition \( \delta_{k+1} < \frac{1}{\sqrt{3K}} \) for the successful recovery by the OMP algorithm. Again, the term improvement in these works mean lesser strict conditions on the RIC. Similar RIC conditions have been reported for other greedy algorithms, such as the subspace pursuit [31], the iterative hard-thresholding [32], and the CoSaMP [33].

In all these works, an explanation for the big assumption that a matrix exists with the suggested RIC is not discussed and it still remains an open problem. In this paper, we aim to address this issue by proposing a new eigenvalue framework. For this purpose, we first review the relation between the RIC and the eigenvalues of \( A_x^T A_x \).

**D. Eigenvalues and their relation to RIC**

We note that the ratio in (28) is a Rayleigh quotient (RQ) given by

\[
\text{RQ}(\mathcal{X}) = \frac{||A_x x||^2}{||x||^2} .
\]

It is well-known that the RQ is bounded [34] as

\[
\lambda_{\min} (A_x^T A_x) \leq \frac{||A_x x||^2}{||x||^2} \leq \lambda_{\max} (A_x^T A_x) .
\]

When the matrix \( A_x \) is iid Gaussian, then \( \lambda_{\min} (A_x^T A_x) \) and \( \lambda_{\max} (A_x^T A_x) \) denote the smallest and largest eigenvalues of the \( K \times K \) Wishart matrix \( A_x^T A_x \). In [21, p. 4208], Candes and Tao used the following inequality relation for a Gaussian matrix \( A_x \) based on the results from [35]:

\[
(1 - \delta_k) \leq \lambda_{\min} (A_x^T A_x) \leq \lambda_{\max} (A_x^T A_x) \leq (1 + \delta_k)
\]

Equation (31) says that the extreme value of the smallest and the largest eigenvalues of the matrix \( A_x^T A_x \) lie within the numbers \( 1 - \delta_k \) and \( 1 + \delta_k \) for all possible \( \mathcal{X} \). Candes and Tao [21, p. 4209] showed that for the Gaussian sensing matrices, (31) holds with the constant \( \delta_k \approx 2 \frac{\sqrt{K}}{\sqrt{M}} + \frac{K}{M} \) when \( \frac{M}{N} \rightarrow \gamma < 1 \) as \( M \rightarrow \infty \) (for a fixed \( \mathcal{X} \)). That is, (31) is valid only in the limit of \( \frac{M}{N} \) approaches a small value close to zero.

When \( \frac{M}{N} \) approaches a value close to one, Blanchard, Cartis, and Tanner [24, Fig. 2.1] observed that (31) does not hold. In particular, they have noted the following:

1. The maximum of the largest eigenvalue and minimum of the smallest eigenvalue are bounded between
1 − δₖ and 1 + δₖ (Δₖ < 1) only when $\frac{1}{M}$ approaches a value close to zero (or $M \to \infty$).

2. When $\frac{1}{M}$ approaches 1, the eigenvalues exhibit an asymmetry around 1. That is, the maximum value of $\lambda_{\min}$ moves beyond 2 and thereby breaches the 1 + δₖ upper bound of (31). However, the minimum of the smallest eigenvalue is still confined between 0 and 1 and thus, behaves like 1 − δₖ. Thus, the same constant δₖ on both sides of (31) behaves differently when $\frac{1}{M}$ approaches 1.

This observation motivated Blanchard, Cartis, and Tanner to propose two different RICs one for each side of (31). Thus, it is apparent that (31) establishes a large deviation relation between δₖ and the extreme eigenvalues of the Wishart matrices. Since the minimum of the smallest eigenvalue behaves like 1 − δₖ, we can estimate δₖ from the smallest eigenvalue alone. This intuition is crucial for our new eigenvalue framework proposed in the subsequent section.

E. Motivations for new eigenvalue framework

The motivations behind developing the eigenvalue-based criterion are as follows:

1. **Smallest eigenvalue and δₖ**: From the previous section, we note that the minimum value of the smallest eigenvalue behaves like 1 − δₖ. Thus, we can relate the smallest eigenvalue to δₖ. For example, CS theory asserts that smaller the δₖ, better is the sensing matrix. In terms of smallest eigenvalue this statement translates to: farther the smallest eigenvalue from zero, better is the sensing matrix. Thus, the statements involving δₖ can be equivalently stated in terms of the smallest eigenvalue.

2. **RIP of order K**: If a sensing matrix $A$ satisfies the ratio in (28) with δₖ < 1, then $A$ is said to obey the RIP of order $K$ with the RIC $\delta_k$. We can assert this statement equivalently in terms of the minimum of the smallest eigenvalue as follows. The RIP of order $K$ implies that sub-matrices $A_k$ for all possible $\mathcal{X}$ are well-conditioned. Equivalently, the gram-matrix $A_k^\dagger A_k$ is positive definite. This implies that the smallest eigenvalue of the matrix $A_k^\dagger A_k$ for all possible $\mathcal{X}$ must be non-zero. Since the minimum of the smallest eigenvalue is lower bounded by 1 − δₖ, we can deduce the RIP of order $K$ from the well-conditioning of sub-matrices, which can be evaluated in terms of their smallest eigenvalues. We use this insight in the next subsection in order to find a sufficient condition for signal recovery using our proposed framework.

F. The proposed framework

In this section, we aim to define an equivalent statement for the RIP of order $K$ (that is a well-conditioned matrix) using the smallest eigenvalue.

From (28) and (29), we infer that for a given deterministic matrix $A$ with δₖ < 1 and if 1 − δₖ ≤ RQ($\mathcal{X}$) for all possible $\mathcal{X}$, then $A$ is said to obey RIP of order $K$ with the constant δₖ. This property can be stated equivalently in terms of the smallest eigenvalue as follows.

Let $A$ be an $M \times N$ iid Gaussian matrix. For a given $\mathcal{X}$, $\lambda_{\min}^\mathcal{X}(K,M,\Sigma)$ is the smallest eigenvalue distribution of a Wishart matrix $A_k^\dagger A_k$ that depends on $\Sigma$, the second order statistics of a single column of $A_k^\dagger$. For each $\lambda_\mathcal{X}$, the distribution $\lambda_{\min}^\mathcal{X}(K,M,\Sigma)$ stays identical, because, $\Sigma$ remains the same. We now define the well-conditioned matrix as follows:

**Definition 4** Well-conditioned Matrix: An $M \times N$ iid Gaussian matrix $A$ is said to be well-conditioned if for a given user defined constant $a > 0$ and for every small $\eta > 0$, there exists an integer $N_0 < \infty$ such that, for every $N > N_0$, we have

$$\Pr \left\{ \min_{\lambda_{\min}^\mathcal{X}(K,M,\Sigma)}(a) > a \right\} > 1 - \eta.$$  (32)

From Definition 4, we note that in order to establish the RIP of order $K$ using the smallest eigenvalues we need to evaluate $\Pr \left\{ \min_{\lambda_{\min}^\mathcal{X}(K,M,\Sigma)}(a) > a \right\}$. We now set

$$1 - \delta_k = \min_{\lambda_{\min}^\mathcal{X}(K,M,\Sigma)}.$$

Recalling our remarks from Section IV. E that the minimum of the smallest eigenvalue behaves like 1 − δₖ. Since the distribution of $\lambda_{\min}^\mathcal{X}(K,M,\Sigma)$ is the same for each $\mathcal{X}$, and we have $\binom{N}{K}$ number of such $\mathcal{X}$, we can write the probability in (32) in terms of $\delta_k$ as

$$\Pr \left\{ \min_{\lambda_{\min}^\mathcal{X}(K,M,\Sigma)}(a) > a \right\} = \left[ \Pr \lambda_{\min}^\mathcal{X}(K,M,\Sigma) > a \right]^{\binom{N}{K}}$$

Then, the complementary probability, $\Pr \{1 - \delta_k \leq a\}$ is upper bounded by

$$\Pr \left\{ 1 - \delta_k \leq a \right\} = \Pr \left\{ \min_{\lambda_{\min}^\mathcal{X}(K,M,\Sigma)}(a) \right\} \leq \left( \frac{\binom{N}{K}}{\binom{K}{a}} \right)^{\Pr \lambda_{\min}^\mathcal{X}(K,M,\Sigma) \leq a} \leq \binom{N}{K} P_0$$

(a) is the union bound and $P_0$ denotes an upper-bound on

$\Pr \lambda_{\min}^\mathcal{X}(K,M,\Sigma) \leq a$, which is a function of the parameters $a$, $K$, and $M$. In order to obtain $P_0$, we first upper bound the distribution $\lambda_{\min}^\mathcal{X}(K,M,\Sigma)$ using the Hadamard maximum determinant theorem (see Appendix). The upper bound for the distribution is then used to derive the $P_0$, which is given in (44) and used in (37).

Equation (33) provides an upper bound for $\Pr \{1 - \delta_k \leq a\}$. Our aim is to calculate $\Pr \{1 - \delta_k > a\}$, the probability of finding the well-conditioned matrix, i.e.,
\[
\Pr \{1 - \delta_K > a\} = 1 - \Pr \{1 - \delta_K \leq a\} \\
\geq 1 - \left(\frac{N}{K}\right) P_U .
\]  
(34)

We state this probability in Theorem 5.

**Theorem 5:** Let \( a \) be a user-defined constant and \( \delta_K \) is the RIC of an \( M \times N \) iid Gaussian matrix. The probability of finding a well-conditioned matrix, \( \Pr \{1 - \delta_K > a\} \), is lower bounded by \( 1 - e^{-N E_k} \) where the exponent \( E_k \) is given by
\[
E_k = \frac{M}{N} \frac{K}{N} \log \frac{N}{K} - c_1 \frac{K}{N} + o(N).
\]

**Proof.**

From (33), we have
\[
\Pr \{1 - \delta_K \leq a\} \leq \left(\frac{N}{K}\right) P_U 
\]
(35)

The probability in (35) enables us to evaluate how the constant \( \delta_K \) for an \( M \times N \) Gaussian matrix behaves for fixed \( M \) and \( N \). The binomial coefficient in (35) is upper bounded as \( \left(\frac{N}{K}\right) \leq e^{N H(K/N)} \) where \( H(K/N) = -\frac{K}{N} \log \frac{K}{N} -(1-\frac{K}{N}) \log (1-\frac{K}{N}) \) is the binary entropy function. For \( 0 < \frac{K}{N} < 1 \), it can be shown that \(-1 - \frac{K}{N} \log (1-\frac{K}{N}) \leq \frac{K}{N} \). By substituting the binomial bound into (35), we obtain
\[
\Pr \{1 - \delta_K \leq a\} \leq e^{N H(K/N)} P_U \\
\leq e^{N \left[\frac{-K}{N} \log \frac{K}{N} + c_1\right]} P_U \\
\leq e^{K \log \frac{K}{N} + \frac{c_1}{2}} P_U
\]
(36)

Using the \( P_U \) from Appendix, we have
\[
\Pr \{1 - \delta_K \leq a\} \leq e^{K \log \frac{K}{N} + \frac{c_1}{2}} P_U \\
\leq e^{K \log \frac{K}{N} + \frac{c_1}{2}} e^{-\left(\frac{M}{N} - \frac{K}{N}\right) \log \frac{N}{K} - c_1 \frac{K}{N} + o(N)} \\
\leq e^{-N E_k}
\]
where \( E_k \) is the exponent given by
\[
E_k = \frac{1}{c_1} \frac{M}{N} - \frac{K}{N} \log \frac{N}{K} - c_1 \frac{K}{N} + o(N) \quad \text{with} \quad c_1 = \frac{3}{2} \quad \text{and} \quad c_2 = \left(1 + \frac{2}{3}\right).
\]

We note that as \( N \) increases, the R.H.S of (37) decreases to zero as long as \( E_k \) remains positive. We say that an ensemble of Gaussian matrices well-conditioned if the exponent in (37) stays positive. The exponent is a function of \( K \), the matrix dimensions \( M \) and \( N \), and the constant \( a \). Thus, the positivity of \( E_k \), i.e., \( E_k > 0 \) yields a condition for well-conditioned matrices in terms of matrix dimensions. For the exponent to stay positive, it is straightforward to note that the following inequality must hold,
\[
\frac{M}{N} > c_1 \log \frac{K}{N} + c_2 \frac{K}{N} + o(N)
\]
(38)

where \( c_1 = c_2 \). We rewrite (38) as a function of the sparsity ratio \( \varepsilon := \frac{K}{N} \) and the undersampling ratio \( \theta := \frac{M}{N} \) as,
\[
\theta > c_1 \varepsilon \log \frac{1}{\theta} + c_2 \varepsilon + o(N).
\]
(39)

In this paper, we call the undersampling ratio the row-column ratio as well, since, it is the ratio between the number of rows and that of the columns of a matrix. The value of \( a \) can be set depending on a specific recovery condition as explained in Section IV. G.

In Fig. 3, we plot the R.H.S of (39) for \( a = 0.92 \) as a function of \( \varepsilon \) in order to find the region on which (39) holds. That is, we would like to find a region where the exponent is positive. We call such a region as the region of well-conditioned matrices. In Fig. 3, the well-conditioned matrix region is the region above the curve. From Fig. 3, we note that for a given value of \( \varepsilon \), there exists a \( \theta \) above which it is possible to find the well-conditioned matrices.

**G. Gaussian matrix with a prescribed RIC**

In this section, we aim to address the question of finding an \( M \times N \) Gaussian matrix with a given RIC. From Theorem 1, we note that \( \Pr \{1 - \delta_K > a\} \geq 1 - e^{-N E_k} \), which can be rearranged as \( \Pr \{\delta_K < 1 - a\} \geq 1 - e^{-N E_k} \). We use this lower bound to compute the probability of finding a Gaussian matrix with \( \delta_K < 1 - a \). This investigation is important, because, as explained in Section IV. C, in OMP-based studies the existence of matrices with a specific RIC is taken for granted and the evidence for the existence of such matrices or conditions for obtaining such matrices remain undiscussed.

In [26], Davenport and Wakin advised that a matrix with the RIC \( \delta_{K+1} < \frac{1}{3\sqrt{K}} \) helps OMP in order to find a \( K \)-sparse signal in \( K \)-steps. We now aim to find the probability that a Gaussian matrix exists with such an RIC prescription. Towards this goal, we set \( a = \frac{\sqrt{2\pi}}{3}\varepsilon \) and replace \( \delta_K \) with \( \delta_{K+1} \) in the above inequality. With these settings, the above probability can be rewritten as
\[
\Pr \{\delta_{K+1} < \frac{1}{3\sqrt{K}}\} \geq 1 - e^{-N E_{K+1}}
\]
(40)
where $E_{k,1}$ is the exponent with $K$ replaced by $K+1$. Equation (40) states that the probability that a Gaussian matrix with $\delta_{k,1} < \frac{1}{N^\varepsilon}$ is lower bounded by $1-e^{-N^\varepsilon E_{k,1}}$. Thus, (40) acts as a useful tool with which we can find the existence of well-conditioned matrices for sparse signal recovery.

In order to illustrate the usefulness of (40), in Fig. 4, we plot the lower bound $1-e^{-N^\varepsilon E_{k,1}}$ in (40) for $\varepsilon = 0.1$ and for various values of $N$. From Fig. 4, we note that for a given sparsity ratio $\varepsilon$ there occurs a threshold phenomenon in the lower bound in terms of $\theta$ for large values of $N$ (solid curve). In this phenomenon, the lower bound abruptly jumps from 0 to 1 at a certain value of $\theta$, denoted as $\theta_0^\varepsilon$, and termed as threshold. This abrupt jump occurs due to the change of polarity in the exponent $E_{k,1}$ from negative to positive. This polarity change makes the probability of the event that an arbitrarily chosen matrix from a Gaussian ensemble satisfies $\delta_{k,1} < \frac{1}{N^\varepsilon}$ to reach 1.

**Definition 6 Threshold $\theta_0^\varepsilon$**: Let $A_1, A_2, \ldots$ be an ensemble of $M \times N$ iid Gaussian sensing matrices with the prescribed RIC $\delta_{k,1} < \delta$ for a given $\delta$. Let $\Pr \{ \delta_{k,1} < \delta \}$ denote the probability that any arbitrarily chosen matrix from the ensemble satisfy $\delta_{k,1} < \delta$. This probability is a function of three variables, namely, $\varepsilon = \frac{\delta}{\delta}$, $\theta = \frac{\varepsilon}{\delta}$ and $\delta$. For fixed $\varepsilon$ and $\delta$, we say the threshold $\theta_0^\varepsilon$ is the value of $\theta$ for which

$$\lim_{N \to \infty} \Pr \{ \delta_{k,1} < \delta \} \to 1.$$  

For example, for the desired RIC $\delta_{k,1} < \frac{1}{N^\varepsilon}$ with $\varepsilon = 0.1$, from Fig. 4, we find that $\theta_0^\varepsilon = 0.56$. This means that any $M \times N$ matrix chosen arbitrarily from the Gaussian ensemble has an RIC $\delta_{k,1} < \frac{1}{N^\varepsilon}$ if the value of $\theta > \theta_0^\varepsilon$. We also verify $\theta_0^\varepsilon$ from (39). In Fig. 4, for $\varepsilon = 0.1$ and $N = 10^4$ the value of $a = 0.92$. Then, from (39) with $c_1 = 1.5$ and $c_2 = 2.19$, we get $\theta_0^\varepsilon = 0.56$. For small values of $N$, we define $\theta_0^\varepsilon$ is the value of $\theta$ at which the lower bound is just above zero. For $N = 10^2$, in Fig. 4, we note that $\theta_0^\varepsilon = 0.58$.

We observe from Table I that for $N = 100$ as the condition on the RIC improves, there appears a slight increase in the value of $\theta_0^\varepsilon$. The improved conditions on RIC actually negotiate with the value of $\theta_0^\varepsilon$. Thus, we note that improving the conditions on RIC establishes an interesting tradeoff with $\theta_0$ when $N$ is small. That is, there is a choice to be made between the relaxed RIC condition and the value of $\theta_0$. However, this tradeoff becomes a strict win in favor of the improved RIC conditions, because, the rate of increase of $\theta_0^\varepsilon$ is slow. We also observe that when $N$ approaches a large
value, the value $\theta_n$ of converges to $\theta_0 = 0.56$ irrespective of the RIC conditions. We report here that we obtained similar results as in Table 1 for the RIC conditions [23-25] suggested for the $L_1$ algorithms.

In summary, in this section, we have shown a way of finding the existence of a Gaussian ensemble with a prescribed RIC. We have revealed that there exists a threshold (in terms of the undersampling ratio) above which it is possible to find a Gaussian matrix with a specific RIC. We have evaluated this threshold for the OMP and the $L_1$ algorithms. We have found that this threshold approaches to a particular constant for large matrix dimensions irrespective of the RIC. For small matrix dimensions, our study has revealed a fundamental tradeoff that exists between the improved RIC conditions and the threshold. Relaxing the conditions on RIC, increases the value of threshold, and thereby affects the amount of signal undersampling. An RIC condition $\delta_k < \delta$ implies that $\theta > \theta_n$ and thus, $M > \theta_n N$ number of measurements are sufficient for any $M \times N$ number of measurements to recover a $K$-sparse signal. Thus, our new eigenvalue-based analysis exposes the largely unknown impact of RIC-based conditions on CS signal recovery.

V. CONCLUSIONS

In this paper, we have derived new closed-form eigenvalue distributions for real-valued Wishart matrices. The distributions are derived using the theory of skew-symmetric matrices. The usefulness of the distributions is illustrated in the context of compressive sensing (CS) where matrix dimensions are typically large, unlike the case of MIMO whose sizes are on the scale of a few tens. We have then proposed a new, eigenvalue-based framework in order to calculate the probability of finding a Gaussian matrix with a prescribed restricted isometry constant (RIC). We have demonstrated that there exists a threshold in terms of the undersampling ratio above which a Gaussian matrix that satisfies a specific RIC is very easy.

APPENDIX

DERIVATION OF THE UPPER BOUND $P_1$ USED IN (37)

We need to find an upper bound for the probability $Pr\{\lambda_{\min}(K,M,\Sigma) \leq a\}$, where $\lambda_{\min}(K,M,\Sigma)$ is the smallest eigenvalue distribution of a Wishart matrix. This distribution, given in (25), can be written for $\rho = \frac{1}{2}$ as

$$f(\lambda) = c \sum_{n=1}^{K} (-1)^{n+K} \lambda^{n+1} e^{-\frac{\lambda}{2}} Pr(D_\lambda).$$

We first upper bound $f(\lambda)$ by using Lemma 7 and the fact that $Pr(D_\lambda) = \sqrt{det D_\lambda}$ for skew-symmetric matrices.

**Lemma 7** Hadamard maximum determinant theorem [37];

Let $D_{\lambda}$ be a real-valued matrix of size $K \times K$. If the entries of $D_{\lambda}$ are bounded by a constant, i.e., $|d_{ij}| \leq \beta$, then

$$|det D_{\lambda}| \leq \left(\frac{\beta}{2}\right)^K (K+1)^{K+1}.$$  

The value of $\beta$ for the matrix $D_{\lambda}$ in (41) can be approximated as $\beta = \frac{\theta_0}{2}$ for all values of $n$, where $0 < \alpha < 1$.

Therefore, $\sqrt{det D_{\lambda}} \leq \left(\frac{\theta_0}{2}\right)^{K/2} (K+1)^{K/2}$, which when used in (41) gives an upper bound for the smallest eigenvalue distribution as

$$f(\lambda) \leq c \left(\frac{\alpha}{2\pi}\right)^{K/2} (K+1)^{(K+1)/2} \sum_{n=1}^{K} (-1)^{n+K} \lambda^{n+1} e^{-\frac{\lambda}{2}}.$$  

$$\leq c \left(\frac{\alpha}{2\pi}\right)^{K/2} (K+1)^{(K+1)/2} \lambda^{(M-K-1)/2} e^{-\frac{\lambda}{2}}.$$  

(42)

To obtain the last step of (42), we note that the maximum value for the summation occurs at $n = K$, and thus, the summation is replaced with $K$ times its value at $n = K$.

Now, we aim to find the probability $Pr\{\lambda \leq a\}$, which is given as

$$Pr\{\lambda \leq a\} = \int_{0}^{a} f(\lambda) d\lambda \leq \int_{0}^{a} c \left(\frac{\alpha}{2\pi}\right)^{K/2} (K+1)^{(K+1)/2} \lambda^{(M-K-1)/2} e^{-\frac{\lambda}{2}} d\lambda \leq c \left(\frac{\alpha}{2\pi}\right)^{K/2} \lambda^{(M-K-1)/2} e^{-\frac{\lambda}{2}}.$$  

(43)

where $\gamma(v, \mu) = \int_{0}^{\mu} x^{v-1} e^{-x} dx$ is the incomplete lower gamma integral, which can be upper bounded as

$$\gamma\left(M-K-1, \frac{\mu}{2}\right) \leq \left(\frac{\mu}{2}\right)^{M-K-1} e^{-M/2}.$$  

On substituting the above upper bound and simplifying (43), we obtain

$$Pr\{\lambda \leq a\} \leq c \left(\frac{\alpha}{2\pi}\right)^{K/2} \lambda^{(M-K-1)/2} e^{-\frac{\lambda}{2}} \leq f(M, K, a) e^{-\frac{M}{2}}.$$  

The function $f(M, K, a)$ can be upper bounded by an exponentially decaying function, i.e., $f(M, K, a) \leq b e^{-\frac{1}{2}(\rho-\alpha)K}$, where $b > 0$ is a constant and thus,

$$Pr\{\lambda \leq a\} \leq b e^{-\frac{1}{2}(\rho-\alpha)K}.\quad (44)$$

REFERENCES

[1] R. J. Muirhead, Aspects of Multivariate Statistical Theory, John Wiley & Sons Inc., 1982.
[2] Antonia Maria Tulino and Sergio Verdu, Random Matrix Theory and Wireless Communications. Now Publishers Inc, 2004.
M. Y. Mo, “Rank-1 real Wishart spiked model,” *Communications on Pure and Applied Mathematics*, vol. 65, no. 11, pp. 1528-1638, Nov. 2012.

Alan T. James, “Distributions of matrix variates and latent roots derived from normal sample,” *Annals of Mathematical Statistics*, vol. 35, no. 2, pp. 475-501, 1964.

C. G. Khatri, “Non-central distribution of i-th largest characteristic roots of three matrices concerning complex multivariate multivariate normal populations,” *Journal of Institute of Ann. Statistical Math.*, vol. 21, pp. 23-32, 1969.

P. R. Krishnaprasad and T. C. Chang, “On the exact distribution of the smallest root of the Wishart matrix using zonal polynomials,” *Ann. Math. Stat.*, vol. 23, pp. 293-295, 1971.

C. G. Khatri, “Distribution of the largest or the smallest characteristic root under null hypothesis concerning complex multivariate normal populations,” *Ann. Math. Stat.*, vol. 35, pp. 1807-1810, Dec. 1964.

Ming Kang and Mohamed-Slim Alouini, “A comparative study on the performance of MIMO MRC systems with and without cochannel interference,” *IEEE Trans. on Comm.*, vol. 52, no. 8, pp. 1417-1425, 2004.

Sugiyama, T. and Fukutomi, K. “On the distribution of the extreme characteristic roots of matrices in multivariate analysis,” *Repts. Stat. Appl. Res., Union of Japanese Scientists and Engineers*, vol. 13, 1966.

Alan Edelman, “Eigenvalues and conditional number of random matrices,” Ph.D Thesis, MIT, 1989.

Dag Jonsson, “Some limit theorems for the eigenvalues of a sample covariance matrix,” *Journal of Multivariate Analysis*, vol. 12, pp. 1-38, 1982.

S. Geman, “A limit theorem for the norm of random matrices,” *Ann. Probab.*, vol. 8, pp. 252-261, 1980.

J. W. Silverstein, “The smallest eigenvalue of a large dimensional Wishart matrix,” *Ann. Probab.*, vol. 13, pp. 1364-1368, 1985.

Z. D. Bai and Y. Q. Yin, “Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix,” *Ann. Probab.*, vol. 21, pp. 1275-1294, 1993.

Marco Chiani, Moe Z. Win, and Alberto Zanella, “On the capacity of spatially correlated MIMO Rayleigh-fading channel,” *IEEE Trans. on Info. Theory*, vol. 49, no. 10, pp. 2363-2371, 2003.

Alberto Zanella, Marco Chiani, and Moe Z. Win, “On the marginal distribution of the eigenvalues of Wishart matrices,” *IEEE Trans. on Comm.*, vol. 57, no. 4, pp. 1050-1060, 2009.

John Harnad, *Random Matrices, Random Processes, and Integrable Systems*, CRM Series in Mathematical Physics, Springer, 2011.

N. G. De Bruijn, “On some multiple integrals involving determinants,” *Journal of Indian Mathematical Society*, vol. 19, pp. 133-151, 1955.

David L. Donoho, “Compressed Sensing,” *IEEE Trans. on Info. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

Richard Baraniuk, “Lecture Notes: Compressive Sensing,” *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118-121, 2007.

Emmanuel Candès and Terence Tao, “Decoding by linear programming,” *IEEE Trans. on Info. Theory*, vol. 51, no. 12, pp. 4203 - 4215, 2005.

R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Construct. Approx.*, vol. 29, no. 3, pp. 253-263, 2008.

Emmanuel J. Candès, “The restricted isometry property and its implications for compressed sensing,” *C. R. Acad. Sci. Paris, Ser. I*, 346, pp. 589-592, 2008.

Simon Foucart and Ming-Jun Lai, “Sparsest solutions of underdetermined linear systems via $L_0$ minimization for $0 < q \leq 1$, *App. Comp. Harmon. Anal.*, vol. 26, no. 3, pp. 395-407, May 2009.

Qun Mo and Song Li, “New bounds on the restricted isometry constant $\delta_k$,” *Appl. Comput. Harmon. Anal.*, vol. 31, pp. 460-468, 2011.

Mark A. Davenport and Michael B. Wakin, “Analysis of Orthogonal Matching Pursuit Using the Restricted Isometry Property,” *IEEE Trans. on Info. Theory*, vol. 56, no. 9, September 2010.

Shisheng Huang and Jubo Zha, “Recovery of sparse signals using OMP and its variants: convergence analysis based on RIP,” *Inverse problems*, vol. 27, no. 3, pp. 1-14, 2011.

Entao Liu and Vladimir N. Temlyakov, “The orthogonal super greedy algorithm and applications in compressive sensing,” *IEEE Trans. on Info. Theory*, vol. 58, no.4, 2012.

Qun Mo and Yi Shen, “A remark on the restricted isometry property in orthogonal matching pursuit,” *IEEE Trans. on Info. Theory*, vol. 58, no. 6, June 2012.