A SIMPLE FORMULA FOR THE MOMENTS OF UNITARILY INVARIANT MATRIX DISTRIBUTIONS

Stephen D. Howard[†] *and Ali Pezeshki*[‡]

[†]Defence Science and Technology Group; P.O. Box 1500, Edinburgh, SA 5111, Australia
[‡]Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA

ABSTRACT

We present a new formula for computing arbitrary moments of unitarily invariant matrix distributions. The Schur-Weyl duality is used to decompose the expected value of tensor powers of the random matrices as a linear combination of projection operators onto unitary irreducible representations. The coefficients in this combination, which are labeled by Young diagrams, are expectations of products of determinants of the random matrices. It is demonstrated in a number of important cases, including matrix gamma and matrix beta distributions, that these coefficients can be simply computed from a knowledge of the normalization factors of the distributions.

Index Terms— Unitary invariance, matrix gamma, matrix beta, moments, group representations.

1. INTRODUCTION

Complex multivariate probability distributions arise in many applications and have been the subject of considerable study [1, 2, 3]. In signal processing problems arising in applications such as radar, sonar and communications, the baseband signal measurements are modeled by multivariate complex Gaussian distributions. In constructing detectors and estimators in such applications, and in analysing their expected performance, the complex Wishart and complex multivariate beta distributions and their moments naturally arise, as they are the multivariate analogues of the χ^2 and beta distributions. The moments of the multivariate beta and inverse multivariate beta are also important in analysing the effect of compressive sensing on detection and estimation performance [4]. Techniques for computing moments of complex Wishart distributions have been studied in a number of articles [5, 6, 7]. Moments of complex multivariate beta distributions have been less well studied [8].

The purpose of this paper is to derive a simple general result which applies to all matrix distributions that can be transformed to a unitarily invariant distribution through conjugation by a fixed matrix. The closest previous work to ours is [6], where the authors exploit the relationship between the moments of the Wishart distribution and the symmetric groups. Here the simplicity and generality of our results is achieved by exploiting the relationship between the moments of unitarily invariant distributions and the joint action of the unitary and symmetric groups on the space $(\mathbb{C}^N)^{\otimes d}$. This is the realm of the Schur-Weyl duality, which intimately relates the irreducible representations of unitary and symmetric groups [9, 10, 11, 12]. Our approach has the advantage that it neatly separates combinatorial aspects of the moment calculation, which are essentially the same for all distributions in this class, from the calculation of a small number of specific distributionally dependent moments.

2. THE MOMENT FORMULA

Suppose that Q is a complex $N \times N$ matrix-valued random variable with probability distribution $F(Q)$. The distribution is unitarily invariant, if for all $V \in \mathcal{U}(N)$, the group of $N \times N$ unitary matrices, $VQV^\dagger$ has the same distribution as Q.

Our goal here is to consider the computation of the expected value $\mathsf{E}\left(Q^{\otimes d}\right)$ for any positive integer d. The notation $Q^{\otimes d}$ denotes the tensor power $Q \otimes Q \otimes \cdots \otimes Q$, with d factors. The elements of $\mathsf{E}\left(Q^{\otimes d}\right)$ include moments corresponding to all products of matrix elements of Q containing d factors.

In the usual way, we interpret Q as the matrix of a random linear operator on $\mathbb{C}^N$, with respect to the standard basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N$. Then $Q^{\otimes d}$ is a random linear operator on $(\mathbb{C}^N)^{\otimes d}$ and so $\mathsf{E}\left(Q^{\otimes d}\right)$ is a fixed linear operator on $(\mathbb{C}^N)^{\otimes d}$. With respect to the standard inner product $\langle \cdot, \cdot \rangle$, an orthonormal basis for $(\mathbb{C}^N)^{\otimes d}$ can be constructed as $\boldsymbol{e}_I = \boldsymbol{e}_{i_1} \otimes \cdots \otimes \boldsymbol{e}_{i_d}$ with multi-index $I = (i_1, \ldots, i_d)$, for all choices of $i_1, \ldots, i_d$ from the set $\{1, \ldots N\}$. The basis $\{\boldsymbol{e}_I\}$ is orthonormal with respect to the inner product on $(\mathbb{C}^N)^{\otimes d}$ defined by

$$\langle \boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_d, \boldsymbol{w}_1 \otimes \cdots \otimes \boldsymbol{w}_d \rangle = \prod_{j=1}^{d} \langle \boldsymbol{v}_j, \boldsymbol{w}_j \rangle.$$

For any two pure elements of the completely anti-symmetric power $\Lambda^d \mathbb{C}^N$, which is a subspace of $(\mathbb{C}^N)^{\otimes d}$, this inner product takes the form

$$\langle \boldsymbol{v}_1 \wedge \cdots \wedge \boldsymbol{v}_d, \boldsymbol{w}_1 \wedge \cdots \wedge \boldsymbol{w}_d \rangle = \det[\langle \boldsymbol{v}_i, \boldsymbol{w}_j \rangle]_{i,j=1,\ldots,d}. \quad (1)$$

Here $\wedge$ denotes the anti-symmetric tensor (or wedge) product.

A. Pezeshki is supported in part by AFOSR awards FA8750-20-2-0504 and FA9550-21-1-0235.

Two important group actions can be defined on $(\mathbb{C}^N)^{\otimes d}$. The first is a group action of the symmetric group $\mathcal{S}_d$ (group of permutations on d letters) defined by $L_\sigma(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_d) = \boldsymbol{v}_{\sigma(1)} \otimes \cdots \otimes \boldsymbol{v}_{\sigma(d)}$. The L_σ are unitary operators on $(\mathbb{C}^N)^{\otimes d}$ ($L_\sigma^\dagger = L_{\sigma^{-1}}$, where $\dagger$ denotes hermitian adjoint). In the $\{\boldsymbol{e}_I\}$ basis, we have matrix elements

$$(L_\sigma)_{IJ} = \langle \boldsymbol{e}_I, \boldsymbol{e}_{\sigma(J)} \rangle = \begin{cases} 1, & \text{if } I = \sigma(J), \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

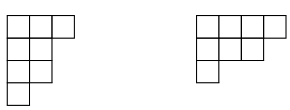
where the action $\sigma(J)$ permutes the entries of J according to σ. The other action is that of the unitary group $\mathcal{U}(N)$ on $(\mathbb{C}^N)^{\otimes d}$ defined by $U^{\otimes d}(\boldsymbol{v}_1 \otimes \cdots \otimes \boldsymbol{v}_d) = U\boldsymbol{v}_1 \otimes \cdots \otimes U\boldsymbol{v}_d$.

Under these two group actions $(\mathbb{C}^N)^{\otimes d}$ becomes a unitary representation of the product group $\mathcal{U}(N) \times \mathcal{S}_d$. This representation decomposes into irreducible representations

$$(\mathbb{C}^N)^{\otimes d} = \bigoplus_{\boldsymbol{\lambda}} W_{\boldsymbol{\lambda}} \tag{3}$$

where the sum is over all partitions $\boldsymbol{\lambda}$ of d with no more than N parts. The $\boldsymbol{\lambda}$ above denote partitions of d, that is, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_\ell)$ with the λ_i integers satisfying $\lambda_i \geq \lambda_{i+1}$ and $\lambda_1 + \cdots + \lambda_\ell = d$. What (3) means is that with an appropriate choice of basis for $(\mathbb{C}^N)^{\otimes d}$, the linear operators in $\mathcal{U}(N) \times \mathcal{S}_d$ can be simultaneously block diagonalized with the blocks labeled by the partitions $\boldsymbol{\lambda}$. The Schur-Weyl duality states that $W_{\boldsymbol{\lambda}} = V_{\boldsymbol{\lambda}} \otimes S_{\boldsymbol{\lambda}}$, where $V_{\boldsymbol{\lambda}}$ are non-equivalent unitary irreducible representations of $\mathcal{U}(N)$ and $S_{\boldsymbol{\lambda}}$ are non-equivalent unitary irreducible representations of $\mathcal{S}_d$.

In terms of partitions, we note that where the λ_i are repeated, such partitions are usually written as $\boldsymbol{\lambda} = (\lambda_1^{\alpha_1}, \ldots, \lambda_k^{\alpha_k})$ where $\lambda_i > \lambda_{i+1}$ and α_i denotes the number of times that λ_i is repeated in the partition. In addition, each partition $\boldsymbol{\lambda}$ can be associated with a Young diagram [13, 14]. For example, for $d = 8$ the following are two partitions and their corresponding Young diagrams:



$\boldsymbol{\lambda} = (3, 2^2, 1) \qquad \boldsymbol{\lambda}' = (4, 3, 1)$

Two partitions are called the transpose of each other if their Young diagrams are related through reflection in the main diagonal. This example shows such a pair.

Associated with each irreducible representation $W_{\boldsymbol{\lambda}}$ there is a Young orthogonal projection operator $Y_{\boldsymbol{\lambda}}$ onto this subspace. The Young projectors satisfy

$$\sum_{\boldsymbol{\lambda}} Y_{\boldsymbol{\lambda}} = I, \quad Y_{\boldsymbol{\lambda}_1} Y_{\boldsymbol{\lambda}_2} = \delta_{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2} Y_{\boldsymbol{\lambda}_1} \tag{4}$$

where δ is the Kronecker delta, and take the form

$$Y_{\boldsymbol{\lambda}} = \frac{\chi_{\boldsymbol{\lambda}}(e)}{d!} \sum_{\sigma \in \mathcal{S}_d} \chi_{\boldsymbol{\lambda}}(\sigma) L_\sigma. \tag{5}$$

In (5), the $\chi_{\boldsymbol{\lambda}}(\sigma)$ are real integer coefficients which comprise the irreducible *characters* of $\mathcal{S}_d$. These can be constructed by standard methods [9, 11, 12]. The characters satisfy the following relations:

$$\frac{1}{d!} \sum_{\sigma \in \mathcal{S}_d} \chi_{\boldsymbol{\lambda}_1}(\sigma\mu) \chi_{\boldsymbol{\lambda}_2}(\sigma^{-1}) = \delta_{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2} \frac{\chi_{\boldsymbol{\lambda}_1}(\mu)}{\chi_{\boldsymbol{\lambda}_1}(e)}$$

and $\frac{1}{d!} \sum_{\boldsymbol{\lambda}} \chi_{\boldsymbol{\lambda}}(e) \chi_{\boldsymbol{\lambda}}(\sigma) = \delta_{e,\sigma}$, which imply that (5) satisfies (4). For $d = 2$, with rows ordered as $\{2, 1^2\}$ and columns ordered as $\{e, (1,2)\}$, $\chi_{\boldsymbol{\lambda}}(\sigma)$ is

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

For $d = 3$, with rows ordered as $\{3, (2,1), 1^3\}$ and columns ordered as $\{e, (1,2), (1,2,3), (1,3,2), (2,3), (1,3)\}$, $\chi_{\boldsymbol{\lambda}}(\sigma)$ is

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 0 & -1 & -1 & 0 & 0 \\ 1 & -1 & 1 & 1 & -1 & -1 \end{pmatrix}.$$

For other values of d, the matrices of characters can be found in [11, 12], or computed with a computer algebra package for larger d.

Returning to the consideration of $\mathsf{E}(Q^{\otimes d})$, we first note that for all $\sigma \in \mathcal{S}_d$, $L_\sigma^\dagger \mathsf{E}(Q^{\otimes d}) L_\sigma = \mathsf{E}(L_\sigma^\dagger Q^{\otimes d} L_\sigma) = \mathsf{E}(Q^{\otimes d})$, since conjugation by L_σ just reorders the tensor product of Qs. Secondly, for any $U \in \mathcal{U}(N)$

$$\left(U^{\otimes d}\right)^\dagger \mathsf{E}(Q^{\otimes d}) U^{\otimes d} = \mathsf{E}((U^\dagger Q U)^{\otimes d}) = \mathsf{E}(Q^{\otimes d})$$

due to the unitary invariance of the distribution of Q. Thus, the linear operator $\mathsf{E}(Q^{\otimes d})$ commutes with the action of $\mathcal{U}(N) \times \mathcal{S}_d$ on $(\mathbb{C}^N)^{\otimes d}$. Now if $\mathsf{E}(Q^{\otimes d})$ is restricted to one of the irreducible representations $W_{\boldsymbol{\lambda}}$, then Schur's Lemma [11, p. 55] implies that $\mathsf{E}(Q^{\otimes d})$ acts as a multiple of the identity. Call this multiple, which is a coefficient moment of Q, $m_{\boldsymbol{\lambda}}(Q)$.

Overall, we have $\mathsf{E}\left(Q^{\otimes d}\right) = \sum_{\boldsymbol{\lambda}} m_{\boldsymbol{\lambda}}(Q)\, Y_{\boldsymbol{\lambda}}$, where the sum is over all partitions $\boldsymbol{\lambda}$ of d with no more than N parts. To determine the moments $m_{\boldsymbol{\lambda}}$ we just need to choose a convenient unit vector $\boldsymbol{w}_{\boldsymbol{\lambda}} \in W_\lambda$ in each irreducible representation and compute $m_{\boldsymbol{\lambda}} = \langle \boldsymbol{w}_{\boldsymbol{\lambda}}, \mathsf{E}\left(Q^{\otimes d}\right) \boldsymbol{w}_{\boldsymbol{\lambda}} \rangle$. To aid in the choice we first note that for any integer ℓ, the leading principal minor of order ℓ of Q, using (1), is

$$\begin{aligned} M_\ell &= \langle \boldsymbol{e}_1 \wedge \cdots \wedge \boldsymbol{e}_\ell, Q^{\otimes \ell} \boldsymbol{e}_1 \wedge \cdots \wedge \boldsymbol{e}_\ell \rangle \\ &= \langle \boldsymbol{e}_1 \wedge \cdots \wedge \boldsymbol{e}_\ell, (Q\boldsymbol{e}_1) \wedge \cdots \wedge (Q\boldsymbol{e}_\ell) \rangle \\ &= \det[\langle \boldsymbol{e}_i, Q\boldsymbol{e}_j \rangle]_{i,j=1,\ldots,\ell}. \end{aligned} \tag{6}$$

As discussed above, each partition $\boldsymbol{\lambda}$, and so each irreducible representation $W_{\boldsymbol{\lambda}}$, is associated with a Young diagram. To specify a basis vector in $W_{\boldsymbol{\lambda}}$ we select a standard Young tableau [11, p. 81], which specifies the symmetries of the vector under the action of $\mathcal{S}_d$, and in addition, a Weyl tableau indicating its structure in terms of the basis vectors $\boldsymbol{e}_1, \ldots \boldsymbol{e}_N$. Here the Young tableau is chosen such that its transpose is a canonical Young tableau [9, p. 45]. The Weyl tableau is chosen to have each column filled with consecutive integers starting with 1 at the top. This procedure gives a vector in $W_{\boldsymbol{\lambda}}$ with a particularly simple structure.

For $d = 2$:

$\boldsymbol{\lambda}$	Young tableau	Weyl tableau	$\boldsymbol{w}_{\boldsymbol{\lambda}}$
2	$\boxed{1}\boxed{2}$	$\boxed{1}\boxed{1}$	$\boldsymbol{e}_1 \otimes \boldsymbol{e}_1$
1^2	$\begin{array}{c}\boxed{1}\\\boxed{2}\end{array}$	$\begin{array}{c}\boxed{1}\\\boxed{2}\end{array}$	$\boldsymbol{e}_1 \wedge \boldsymbol{e}_2$

In this case we have

$$\begin{aligned} m_2 &= \mathsf{E}(\langle \boldsymbol{e}_1 \otimes \boldsymbol{e}_1, (Q \otimes Q)\, \boldsymbol{e}_1 \otimes \boldsymbol{e}_1 \rangle) = \mathsf{E}(q_{11}^2) \\ m_{1^2} &= \mathsf{E}(\langle \boldsymbol{e}_1 \wedge \boldsymbol{e}_2, Q \otimes Q\, \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \rangle) = \mathsf{E}\left(\begin{vmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{vmatrix} \right). \end{aligned}$$

For $d = 3$:

$\boldsymbol{\lambda}$	Young tableau	Weyl tableau	$\boldsymbol{w}_{\boldsymbol{\lambda}}$
3	$\boxed{1}\boxed{2}\boxed{3}$	$\boxed{1}\boxed{1}\boxed{1}$	$\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 \otimes \boldsymbol{e}_1$
$2, 1$	$\begin{array}{l}\boxed{1}\boxed{3}\\\boxed{2}\end{array}$	$\begin{array}{l}\boxed{1}\boxed{1}\\\boxed{2}\end{array}$	$(\boldsymbol{e}_1 \wedge \boldsymbol{e}_2) \otimes \boldsymbol{e}_1$
1^3	$\begin{array}{c}\boxed{1}\\\boxed{2}\\\boxed{3}\end{array}$	$\begin{array}{c}\boxed{1}\\\boxed{2}\\\boxed{3}\end{array}$	$\boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3$

For which we have

$$\begin{aligned} m_3 &= \mathsf{E}(\langle \boldsymbol{e}_1 \otimes \boldsymbol{e}_1 \otimes \boldsymbol{e}_1, (Q \otimes Q \otimes Q)\, \boldsymbol{e}_1 \otimes \boldsymbol{e}_1 \otimes \boldsymbol{e}_1 \rangle) \\ &= \mathsf{E}(q_{11}^3), \\ m_{1^2} &= \mathsf{E}(\langle (\boldsymbol{e}_1 \wedge \boldsymbol{e}_2) \otimes \boldsymbol{e}_1, (Q \otimes Q \otimes Q)\, (\boldsymbol{e}_1 \wedge \boldsymbol{e}_2) \otimes \boldsymbol{e}_1 \rangle) \\ &= \mathsf{E}\left(\langle \boldsymbol{e}_1 \wedge \boldsymbol{e}_2, (Q \otimes Q)\, \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \rangle \langle \boldsymbol{e}_1, Q \boldsymbol{e}_1 \rangle \right) \\ &= \mathsf{E}\left(\begin{vmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{vmatrix} q_{11} \right), \\ m_{1^3} &= \mathsf{E}(\langle (\boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3, (Q \otimes Q \otimes Q)\, \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \otimes \boldsymbol{e}_1 \rangle) \\ &= \mathsf{E}\left(\begin{vmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{vmatrix} \right). \end{aligned}$$

In general, we make the choice

$$\begin{aligned} \boldsymbol{w}_{\boldsymbol{\lambda}} =& (\boldsymbol{e}_1 \wedge \cdots \wedge \boldsymbol{e}_{\lambda'_1})^{\otimes \alpha'_1} \otimes (\boldsymbol{e}_1 \wedge \cdots \wedge \boldsymbol{e}_{\lambda'_2})^{\otimes \alpha'_2} \otimes \cdots \\ & \cdots \otimes (\boldsymbol{e}_1 \wedge \cdots \wedge \boldsymbol{e}_{\lambda'_k})^{\otimes \alpha'_k} \end{aligned}$$

where we recall the $\boldsymbol{\lambda}' = (\lambda_1'^{\alpha'_1}, \ldots, \lambda_k'^{\alpha'_k})$ denotes the transpose of $\boldsymbol{\lambda}$. Using (6) we get

$$\begin{aligned} m_{\boldsymbol{\lambda}} &= \langle \boldsymbol{w}_{\boldsymbol{\lambda}}, \mathsf{E}\left(Q^{\otimes d}\right) \boldsymbol{w}_{\boldsymbol{\lambda}} \rangle \\ &= \mathsf{E}\left(\prod_{\ell=1}^{k} \langle \boldsymbol{e}_1 \wedge \cdots \wedge \boldsymbol{e}_{\lambda'_\ell}, Q^{\otimes \lambda'_\ell}\, \boldsymbol{e}_1 \wedge \cdots \wedge \boldsymbol{e}_{\lambda'_\ell} \rangle^{\alpha'_\ell} \right) \\ &= \mathsf{E}\left(\prod_{\ell=1}^{k} M_{\lambda'_\ell}^{\alpha'_\ell} \right). \end{aligned}$$

To summarise, we have our **main result**: *The d^{th} order moments of Q are given by*

$$\mathsf{E}\left(Q^{\otimes d}\right) = \sum_{\boldsymbol{\lambda}} m_{\boldsymbol{\lambda}} Y_{\boldsymbol{\lambda}} \tag{7}$$

where the sum is over all partitions $\boldsymbol{\lambda}$ of d with no more than N parts and the $Y_{\boldsymbol{\lambda}}$ are the Young projectors onto the irreducible representations associated with the action of $\mathcal{U}(N) \times \mathcal{S}_d$ on $(\mathbb{C}^N)^{\otimes d}$. The coefficient moments $m_{\boldsymbol{\lambda}}(Q)$ are given by

$$m_{\boldsymbol{\lambda}} = \mathsf{E}\left(M_{\lambda'_1}^{\alpha_1} M_{\lambda'_2}^{\alpha_2} \ldots M_{\lambda'_k}^{\alpha_k} \right) \tag{8}$$

where $\boldsymbol{\lambda}' = (\lambda_1'^{\alpha_1}, \lambda_2'^{\alpha_2}, \ldots, \lambda_k'^{\alpha_k})$ is the transpose to the partition $\boldsymbol{\lambda}$ and M_ℓ denotes the leading principal minor of order ℓ for Q, i.e., the determinant of the top left $\ell \times \ell$ block of Q.

Finally we have an important application of this result: suppose that X is a complex matrix-valued random variable with the property that $Q = G^{-1}X(G^{-1})^\dagger$ has a unitarily invariant probability distribution for some fixed non-singular matrix G. Then

$$\mathsf{E}(X^{\otimes d}) = R^{\otimes d} \sum_{\boldsymbol{\lambda}} m_{\boldsymbol{\lambda}}(Q)\, Y_{\boldsymbol{\lambda}} \tag{9}$$

where $R = GG^\dagger$.

3. COMPUTING MOMENTS

In this section we consider how to use (7) to compute specific moments. Suppose that we want to calculate a particular d^{th} order moment of the elements of Q:

$$\begin{aligned} \mathsf{E}(q_{i_1 j_1} \cdots q_{i_d j_d}) &= \mathsf{E}(Q^{\otimes d})_{IJ} \\ &= \sum_{\boldsymbol{\lambda}} \widetilde{m}_{\boldsymbol{\lambda}} \sum_{\sigma \in \mathcal{S}_d} \chi_{\boldsymbol{\lambda}}(\sigma)\, (L_\sigma)_{IJ} \\ &= \sum_{\boldsymbol{\lambda}} \widetilde{m}_{\boldsymbol{\lambda}} \sum_{\sigma : I = \sigma(J)} \chi_{\boldsymbol{\lambda}}(\sigma) \end{aligned}$$

where $I = (i_1, \ldots, i_d)$ and $J = (j_1, \ldots, j_d)$ and we have used (5) along with the normalization $\widetilde{m}_{\boldsymbol{\lambda}} = \frac{\chi_{\boldsymbol{\lambda}}(e)}{d!} m_{\boldsymbol{\lambda}}$. Note that if I and J have repeated entries there may be more than one σ such that $I = \sigma(J)$.

If $Q = G^{-1}X(G^{-1})^\dagger$ has a unitarily invariant distribution, then, using (9) and (2), we have

$$\begin{aligned} \mathsf{E}(x_{i_1 j_1} \cdots x_{i_d j_d}) &= \mathsf{E}(X^{\otimes d})_{IJ} \\ &= \sum_{\boldsymbol{\lambda}} \widetilde{m}_{\boldsymbol{\lambda}}(Q) \sum_{\sigma \in \mathcal{S}_d} \chi_{\boldsymbol{\lambda}}(\sigma) \left(R^{\otimes d}\right)_{IK} (L_\sigma)_{KJ} \\ &= \sum_{\boldsymbol{\lambda}} \widetilde{m}_{\boldsymbol{\lambda}}(Q) \sum_{\sigma} \chi_{\boldsymbol{\lambda}}(\sigma) \left(R^{\otimes d}\right)_{I, \sigma(J)}. \end{aligned} \tag{10}$$

Equation (10) gives a basis for computing many types of moments. For example, it is straightforward to compute the expected value of products of arbitrary linear combinations of the elements of X. A general linear combination of these elements takes the form $\mathrm{Tr}(AX)$ for some matrix A. Multiplying $\mathsf{E}(X^{\otimes d})$ by $A_1 \otimes \cdots \otimes A_d$ and taking the trace gives

$$\mathsf{E}(\mathrm{Tr}(A_1 X) \cdots \mathrm{Tr}(A_d X)) = \mathrm{Tr}\left((A_1 \otimes \cdots \otimes A_d)\mathsf{E}(X^{\otimes d})\right)$$
$$= \sum_{\boldsymbol{\lambda}} \widetilde{m}_{\boldsymbol{\lambda}}(Q) \sum_{\sigma} \chi_{\boldsymbol{\lambda}}(\sigma) \sum_{I} \left((A_1 R) \otimes \cdots \otimes (A_d R)\right)_{I,\sigma(I)}.$$

Another application is to compute the expected values of powers of X, i.e., $\mathsf{E}(X^d)_{ij}$, using the relation $\mathsf{E}(X^d)_{ij} = \sum_K \mathsf{E}\left(X^{\otimes d}\right)_{(i,K),(K,j)}$, where the sum is over all multi-indexes K of size $d-1$.

4. MINOR MOMENTS FOR SOME DISTRIBUTIONS

We now consider the computation of the coefficient moments (8) for a number of important multivariate distributions.

Matrix Beta Distributions: Suppose that Q has a complex matrix beta distribution of type I, $\mathbb{C}B_N^I(a,b)$ with parameters $a, b > N-1$, i.e.,

$$dF(Q) = \frac{1}{B_N(a,b)} |Q|^{a-N} |I-Q|^{b-N} dQ$$

where dQ denotes the Lebesgue measure on $\mathbb{C}^{N\times N}$. Here $B_N(a,b) = \Gamma_N(a)\Gamma_N(b)/\Gamma_N(a+b)$ with

$$\Gamma_N(a) = \pi^{N(N-1)/2} \prod_{j=1}^{N} \Gamma(a-j+1).$$

An important property of the distribution $\mathbb{C}B_N^I(a,b)$ is the marginal distribution of the top left $\ell \times \ell$ submatrix of Q is $\mathbb{C}B_\ell^I(a,b)$ [15, p. 24]. Thus, denoting the top left $\ell \times \ell$ submatrix of Q by Q_ℓ, we can write

$$m_{\boldsymbol{\lambda}} = \frac{1}{B_N(a,b)} \int \prod_{\ell=1}^{k} |Q_{\lambda'_\ell}|^{\alpha'_\ell} |Q|^{a-N} |I-Q|^{b-N} dQ$$
$$= \frac{1}{B_{\lambda'_1}(a,b)} \int \prod_{\ell=2}^{k} |Q_{\lambda'_\ell}|^{\alpha'_\ell} \, |Q_{\lambda'_1}|^{a+\alpha'_1-\lambda'_1} |I-Q_{\lambda'_1}|^{b-\lambda'_1} dQ_{\lambda'_1}$$
$$= \frac{B_{\lambda'_1}(a+\alpha'_1,b)}{B_{\lambda'_1}(a,b)} \frac{1}{B_{\lambda'_2}(a+\alpha'_1,b)} \times$$
$$\int \prod_{\ell=3}^{k} |Q_{\lambda'_\ell}|^{\alpha'_\ell} \, |Q_{\lambda'_2}|^{a+\alpha'_1+\alpha'_2-\lambda'_2} |I-Q_{\lambda'_2}|^{b-\lambda'_2} dQ_{\lambda'_2}.$$

Continuing in this way we end up with

$$m_{\boldsymbol{\lambda}} = \prod_{\ell=1}^{k} \frac{B_{\lambda'_\ell}(a+\sum_{j=1}^{\ell}\alpha'_j, b)}{B_{\lambda'_\ell}(a+\sum_{j=1}^{\ell-1}\alpha'_j, b)}.$$

Similar results can be computed for the type II and inverse complex matrix beta distribution.

Wishart and Matrix Gamma Distributions Suppose that X is an hermitian positive random matrix having a complex gamma distribution $\mathbb{C}G_N(a,\Sigma)$:

$$dF(X) = \frac{|X|^{a-N}}{\Gamma_N(a)|\Sigma|^a} \exp(-\mathrm{Tr}(\Sigma^{-1}X))$$

where $a > N-1$ and $\Sigma > 0$. We begin by making the change of variable $Q = \Sigma^{-1/2} X \Sigma^{-1/2}$ so

$$dF'(Q) = \frac{|Q|^{a-N}}{\Gamma_N(a)} \exp(-\mathrm{Tr}(Q)).$$

The transformed distribution is unitarily invariant. Using the fact that if $Q \sim \mathbb{C}G_N(a,I)$ then marginal distribution of the top left $\ell\times\ell$ submatrix of Q is $Q_\ell \sim \mathbb{C}G_\ell(a,I)$. A similar calculation to the matrix beta distribution gives

$$m_{\boldsymbol{\lambda}} = \prod_{\ell=1}^{k} \frac{\Gamma_{\lambda'_\ell}(a+\sum_{j=1}^{\ell}\alpha'_j)}{\Gamma_{\lambda'_\ell}(a+\sum_{j=1}^{\ell-1}\alpha'_j)}$$

Again a corresponding result can be computed for the inverse complex gamma/Wishart distributions.

Unitarily Invariant Vector Distributions: Another interesting application of our result is the following. Suppose $\boldsymbol{z}$ is a unitarily invariant vector-valued random variable. The expected value of any product of elements of $\boldsymbol{z}$ and their complex conjugates is non-zero only if the number of conjugate and non-conjugate terms are balanced. This means the non-zero moments of order d are contained in $\mathsf{E}(Q^{\otimes d})$ where $Q = \boldsymbol{z}\boldsymbol{z}^\dagger$ (outer product). This Q and all of its principal submatrices Q_ℓ are rank one and as a consequence $m_{\boldsymbol{\lambda}}$ is zeros unless $\boldsymbol{\lambda} = (d)$. All the other moments $m_{\boldsymbol{\lambda}}$ include principal minors which are zero. As $m_d = \mathsf{E}(|z_1|^{2d})$, we have

$$\mathsf{E}\left((\boldsymbol{z}\boldsymbol{z}^\dagger)^{\otimes d}\right) = \mathsf{E}(|z_1|^{2d}) Y_d = \frac{1}{d!}\mathsf{E}(|z_1|^{2d}) \sum_{\sigma\in\mathcal{S}_d} L_\sigma.$$

If $\boldsymbol{z} = R^{-1/2}\boldsymbol{y}$, with R a positive-definite hermitian matrix,

$$\mathsf{E}(y_{i_1}\overline{y}_{j_1} \cdots y_{i_d}\overline{y}_{j_d}) = \mathsf{E}\left((\boldsymbol{y}\boldsymbol{y}^\dagger)^{\otimes d}\right)_{IJ}$$
$$= \frac{1}{d!}\mathsf{E}(|z_1|^{2d}) \sum_{\sigma\in\mathcal{S}_d} \left(R^{\otimes d}\right)_{I\sigma(J)}. \quad (11)$$

The simplicity of this result should be compared with that in [16, 17] for complex Gaussian distributed variables. Equation (11) can be seen as a generalization of Isserlis' Theorem [18] to unitarily invariant vector distributions.

5. CONCLUSION

The formula (7) for the moments of unitarily invariant distributions derived here provides a powerful tool for dealing with the combinatorial aspect of moment computation and in separating them from a small number of specific distribution dependent moments, for an important set of multivariate distributions. It is apparent that this approach could be generalized to compute moments of other classes of distributions.

6. REFERENCES

[1] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics. Wiley, 2003.

[2] R.J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley Series in Probability and Statistics. Wiley, 1982.

[3] A.K. Gupta and D.K. Nagar, *Matrix Variate Distributions*, Monographs and Surveys in Pure and Applied Mathematics. Taylor & Francis, 1999.

[4] Pooria Pakrooh, Ali Pezeshki, Louis L. Scharf, Douglas Cochran, and Stephen D. Howard, "Analysis of Fisher information and the Cramér-Rao bound for nonlinear parameter estimation after random compression," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6423–6428, 2015.

[5] D. Maiwald and D. Kraus, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices," *IEE Proceedings on Radar, Sonar and Navigation*, vol. 147, pp. 162–168, 09 2000.

[6] Piotr Graczyk, Gérard Letac, and Hélène Massam, "The complex Wishart distribution and the symmetric group," *Ann. Statist.*, vol. 41, February 2003.

[7] Daya K. Nagar and Arjun K. Gupta, "Expectations of functions of complex Wishart matrix," *Acta Applicandae Mathematicae*, vol. 113, pp. 265–288, 2011.

[8] Arjun K. Gupta, Daya K. Nagar, and Astrid Marissa Velez-Carvajal, "Unitary invariant and residual independent matrix distributions," *Computational & Applied Mathematics*, vol. 28, pp. 63–86, 2009.

[9] William Fulton and Joe Harris, *Representation theory : a first course*, Springer, 2004.

[10] Ambar N. Sengupta, *Representing Finite Groups, A Semimsimple Introduction.*, Springer, 2012.

[11] Shlomo Sternberg, *Group Theory and Physics*, Cambridge University Press, 1994.

[12] Morton Hamermesh, *Group Theory and its Application to Physical Problems*, Addison-Wesley, 1962.

[13] W. Fulton, *Young Tableaux: With Applications to Representation Theory and Geometry*, London Mathematical Society Student Texts. Cambridge University Press, 1997.

[14] Donald Knuth, *The Art Of Computer Programming, vol. 3: Sorting And Searching*, Addison-Wesley, 1973.

[15] D.K. Nagar and A.K. Gupta, *Contributions to complex matrix variate distributions theory*, Colección Ciencia y Tecnología. University of Antioquia Publisher, 2009.

[16] I. Reed, "On a moment theorem for complex Gaussian processes," *IRE Transactions on Information Theory*, vol. 8, no. 3, pp. 194–195, 1962.

[17] Shagufta A. Sultan and Derrick S. Tracy, "Moments of the complex multivariate normal distribution," *Linear Algebra and its Applications*, vol. 237-238, pp. 191–204, 1996.

[18] L. Isserlis, "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika*, , no. 1-2, pp. 134–139, November 1918.