NEW METHODS FOR HANDLING SINGULAR SAMPLE COVARIANCE MATRICES

GABRIEL H. TUCCI AND KE WANG

Abstract. The estimation of a covariance matrix from an insufficient amount of data is one of the most common problems in fields as diverse as multivariate statistics, wireless communications, signal processing, biology, learning theory and finance. In [17], a new approach to handle singular covariance matrices was suggested. The main idea was to use dimensionality reduction in conjunction with an average over the Stiefel manifold. In this paper we continue with this research and we consider some new approaches to solve this problem. One of the methods is called the *Ewens* estimator and uses a randomization of the sample covariance matrix over all the permutation matrices with respect to the Ewens measure. The techniques used to attack this problem are broad and run from random matrix theory to combinatorics.

1. Introduction

The estimation of a covariance matrix from an insufficient amount of data is one of the most common problems in fields as diverse as multivariate statistics, wireless communications, signal processing, biology, learning theory and finance. For instance, the covariation between asset returns plays a crucial role in modern finance. The covariance matrix and its inverse are the key statistics in portfolio optimization and risk management. Many recent financial innovations involve complex derivatives, like exotic options written on the minimum, maximum or difference of two assets, or some structured financial products, such as CDOs. All of these innovations are built upon, or in order to exploit, the correlation structure of two or more assets. In the field of wireless communications, covariance estimates allows us to compute the direction of arrival (DOA), which is a critical task in smart antenna systems since it enables accurate mobile location. Another application is in the field of biology and involves the interactions between proteins or genes in an organism and the joint time evolution of their interactions.

Typically the covariance matrix of a multivariate random variable is not known but has to be estimated from the data. Estimation of covariance matrices then deals with the question of how to approximate the actual covariance matrix on the basis of samples from the multivariate distribution. Simple cases, where the number of observations is much greater than the number of variables, can be dealt with by using the sample covariance matrix. In this case, the sample covariance matrix is an unbiased and efficient estimator of the *true* covariance matrix. However, in many practical situations we would like to estimate the covariance matrix of a set of variables from an insufficient amount of data. In this case the sample covariance matrix is singular (non–invertible) and therefore a fundamentally bad estimate. More specifically, let X be a random vector $X = (X_1, \dots, X_m)^T \in \mathbb{C}^{m\times 1}$ and assume for simplicity that X is centered. Then the true covariance matrix is given by

$$\Sigma = \mathbb{E}(XX^*) = (\mathrm{cov}(X_i, X_j))_{1\le i,j\le m}. \tag{1.1}$$

Consider n independent samples or realizations $x_1, \dots, x_n \in \mathbb{C}^m$ and form the $m\times n$ data matrix $M = (x_1, \dots, x_n)$. Then the sample covariance matrix is an $m \times m$ non–negative definite matrix defined as

$$K = \frac{1}{n} MM^*. \tag{1.2}$$

If $n \to +\infty$ and m is fixed, then the sample covariance matrix K converges (entrywise) to Σ almost surely. Whereas, as we mentioned before, in many empirical problems, the number of measurements is less than the dimension ($n < m$), and thus the sample covariance matrix is singular. Our objective in this paper is to recover the true covariance matrix Σ from K under the condition $n < m$.

The conventional treatment of covariance singularity artificially converts the singular sample covariance matrix into an invertible (positive definite) covariance by the simple expedient of adding a positive diagonal matrix, or more generally, by taking a linear combination of the sample covariance and the identity matrix. This procedure is variously called "diagonal loading" or "ridge regression" [21, 7]. This one is defined as $\alpha K + \beta I_m$ where α and β are called loading parameters. The resulting matrix is positive definite, invertible and preserves the eigenvectors of the sample covariance. The eigenvalues of $\alpha K + \beta I_m$ are a uniform rescaling and shift of the eigenvalues of K. There are many methods in choosing the optimum loading parameters, see [14], [18] and [19]. On the other hand, if the true covariance matrix is assumed to have some level of sparsity, several works have been established, such as the banding and thresholding methods studied by Bickel and Levina [3, 4], Wu and Pourahmadi [26], El Karoui [8] and Rothman et al. [22], to mention a few. In more recent works, Cai, Zhang and Zhou [27] and Cai and Zhou [28] derive the optimal rate of convergence for estimating the true covariance matrix and its inverse under operator norm, Frobenius norm and l_1 norm, for a large range of sparse covariance matrices.

In Marzetta, Tucci and Simon's paper [17] a new approach to handle singular covariance matrices was suggested. Let $p \leq n$ be a parameter, to be estimated later, and consider the set of all $p \times m$ one-sided unitary matrices

$$\Omega_{p,m} = \{\Phi \in \mathbb{C}^{p\times m} \; : \; \Phi\Phi^* = I_p\}. \tag{1.3}$$

This set has a manifold structure and is called the Stiefel manifold. We also endow this manifold with the *Haar measure*, that is, the uniform distribution on the set $\Omega_{p,m}$. They define the operators

$$\mathrm{cov}_p(K) = \mathbb{E}(\Phi^*(\Phi K \Phi^*)\Phi), \tag{1.4}$$

and

$$\mathrm{invcov}_p(K) = \mathbb{E}(\Phi^*(\Phi K \Phi^*)^{-1}\Phi), \tag{1.5}$$

where the expectation is taken with respect to the Haar measure. It was found that

$$\mathrm{cov}_p(K) = \frac{p}{(m^2-1)m}\Big((mp-1)K + (m-p)\mathrm{Tr}(K)I_m\Big),$$

which is the same as diagonal loading. Moreover, they investigated the properties of $\mathrm{invcov}_p(K)$. If K is decomposed as $K = UDU^*$, with $D = \mathrm{diag}(d_1, \ldots, d_n, 0, \ldots, 0)$, then

$$\mathrm{invcov}_p(K) = U\mathrm{invcov}_p(D)U^*,$$

and

$$\mathrm{invcov}_p(D) = \mathrm{diag}(\lambda_1, \ldots, \lambda_n, \mu, \ldots, \mu).$$

In other words, $\mathrm{invcov}_p(K)$ preserves the eigenvectors of K, and transforms all the zero eigenvalues to a non–zero constant value. They also provided formulas to compute the values of λ_i and μ, and studied their asymptotic behavior using techniques from free probability.

In this paper, we investigate new methods to estimate singular covariance matrices. In Section 2, we continue to work on the operator invcov_p, suggested in [17], and show that $\mathrm{invcov}_p(K)$ has a very simple algebraic structure, i.e. it is a polynomial in K. An explicit formula for computing $\mathbb{E}\big(\Phi(\Phi^* D_n \Phi)^l \Phi^*\big)$ is given. In Section 4, we consider a new approach, called the *Ewens* estimator, to estimate Σ. In this one, the average is taken over the set of all $m \times m$ permutation matrices with respect to the *Ewens measure*. The explicit formula for the *Ewens* estimator is computed using combinatorial techniques. In Section 5, we combine the ideas of the first two methods. We first extend the definition of permutation matrices to get $p \times m$ unitary matrices and define two new operators

$$K_{\theta,m,p} := \mathbb{E}\Big(V_\sigma^T(V_\sigma K V_\sigma^T)V_\sigma\Big),$$

$$\tilde{K}_{\theta,m,p} := \mathbb{E}\Big(V_\sigma^T(V_\sigma K V_\sigma^T)^+ V_\sigma\Big)$$

to estimate Σ and Σ^{-1} respectively. We provide an explicit formula for $K_{\theta,m,p}$ and an inductive formula to compute $\tilde{K}_{\theta,m,p}$. In Section 6, it is assumed that Σ has some special form, i.e. tridiagonal Toeplitz matrix or power Toeplitz matrix and we study its asymptotic behavior under the *Ewens* estimator. In this Section, we also present some simulations under the different methods to test the effect of the parameters. We also compare our methods with other methods used in the literature.

Notation: Throughout this paper, $\mathbf{1}_S$ is the indicator function of a set S. We sometimes use $[n]$ to present the set $\{1, 2, \ldots, n\}$, and $\mathrm{Tr}(A)$ is the trace of a matrix A. For a vector $v = (v_1, \ldots, v_m)$, we use the Euclidean norm $\|v\|_2 = \sqrt{\sum_{i=1}^m |v_i|^2}$ and for an $m \times m$ matrix A, we use the (normalized) Frobenius norm $\|A\| = \frac{1}{\sqrt{m}}\sqrt{\mathrm{Tr}(AA^*)}$. We use the notation $\lambda \vdash n$ to indicate that λ is an integer partition of the positive integer n.

2. Some Properties of the invcov_p Estimator

We first collect some preliminaries about Schur polynomials that will be needed later in studying the properties of the invcov_p estimator.

2.1. Schur Polynomials Preliminaries.

A symmetric polynomial is a polynomial $P(x_1, x_2, \ldots, x_n)$ in n variables such that if any of the variables are interchanged one obtains the same polynomial. Formally, P is a symmetric polynomial if for any permutation σ of the set $\{1, 2, \ldots, n\}$ one has that

$$P(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}) = P(x_1, x_2, \ldots, x_n).$$

Symmetric polynomials arise naturally in the study of the relation between the roots of a polynomial in one variable and its coefficients, since the coefficients can be given by a symmetric polynomial expressions in the roots. Symmetric polynomials also form an interesting structure by themselves. The resulting structures, and in particular the ring of symmetric functions, are of great importance in combinatorics and in representation theory (see for instance [10, 20, 16, 23] for more on details on this topic).

The Schur polynomials are certain symmetric polynomials in n variables. This class of polynomials is also very important in representation theory since they are the characters of irreducible representations of the general linear groups. The Schur polynomials are indexed by partitions. A partition of a positive integer n, also called an integer partition, is a way of writing n as a sum of positive integers. Two partitions that differ only in the order of their summands are considered to be the same partition. Therefore, we can always represent a partition λ of a positive integer n as a sequence of n non-increasing and non-negative integers d_i such that

$$\sum_{i=1}^{n} d_i = n \quad \text{with} \quad d_1 \geq d_2 \geq d_3 \geq \ldots \geq d_n \geq 0.$$

Notice that some of the d_i could be zero. Integer partitions are usually represented by the so called Young's tableaux (also known as Ferrers' diagrams). A Young tableaux is a finite collection of boxes, or cells, arranged in left–justified rows, with the row lengths weakly decreasing (each row has the same or shorter length than its predecessor). Listing the number of boxes on each row gives a partition λ of a non-negative integer n, the total number of boxes of the diagram. The Young diagram is said to be of shape λ, and it carries the same information as that partition. For instance, in Figure 1 we can see the Young tableaux corresponding to the partition $(5, 4, 1)$ of the number 10.

Given a partition λ of n

$$n = d_1 + d_2 + \cdots + d_n \ : \ d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$$

FIGURE 1. Young tableaux representation of the partition $(5, 4, 1)$ (left). The same tableaux with the corresponding Hook's lengths (right).

the following functions are alternating polynomials (in other words they change sign under any transposition of the variables):

$$a_{(d_1,\ldots,d_n)}(x_1,\ldots,x_n) = \det \begin{bmatrix} x_1^{d_1} & x_2^{d_1} & \ldots & x_n^{d_1} \\ x_1^{d_2} & x_2^{d_2} & \ldots & x_n^{d_2} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{d_n} & x_2^{d_n} & \ldots & x_n^{d_n} \end{bmatrix} = \sum_{\sigma\in S_n} \epsilon(\sigma) x_{\sigma(1)}^{d_1} \cdots x_{\sigma(n)}^{d_n} \tag{2.1}$$

where S_n is the permutation group of the set $\{1, 2, \ldots, n\}$ and $\epsilon(\sigma)$ the sign of σ. Since they are alternating, they are all divisible by the Vandermonde determinant

$$\Delta(x_1,\ldots,x_n) = \prod_{1\le j<k\le n} (x_j - x_k).$$

The Schur polynomial associated to λ is defined as the ratio:

$$s_\lambda(x_1, x_2, \ldots, x_n) = \frac{a_{(d_1+n-1,d_2+n-2,\ldots,d_n+0)}(x_1,\ldots,x_n)}{\Delta(x_1,\ldots,x_n)}.$$

This is a symmetric function because the numerator and denominator are both alternating, and a polynomial since all alternating polynomials are divisible by the Vandermonde determinant (see [10, 16, 23] for more details here). For instance,

$$s_{(2,1,1)}(x_1, x_2, x_3) = x_1\, x_2\, x_3\, (x_1 + x_2 + x_3)$$

and

$$s_{(2,2,0)}(x_1, x_2, x_3) = x_1^2\, x_2^2 + x_1^2\, x_3^2 + x_2^2\, x_3^2 + x_1^2\, x_2\, x_3 + x_1\, x_2^2\, x_3 + x_1\, x_2\, x_3^2. \tag{2.2}$$

Another definition we need for the next Section is the so called Hook length, hook(x), of a box x in Young diagram of shape λ. This is defined as the number of boxes that are in the same row to the right of it plus those boxes in the same column below it, plus one (for the box itself). For instance, in Figure 1 we can see the hook lengths of the partition $(5, 4, 1)$. The product of the hook's length of a partition is the product of the hook lengths of all the boxes in the partition. We recommend the interested reader to consult [10, 16, 23] for more details and examples on this topic.

2.2. **Properties of the** ***$invcov_p$*** **estimator.** For an Hermitian matrix K, one can decompose $K = UDU^*$ where U is unitary and $D = \mathrm{diag}(d_1, \ldots, d_m)$. It was showed in [17] that $\mathrm{invcov}_p(K) = U\mathrm{invcov}_p(D)U^*$ where $\mathrm{invcov}_p(D)$ is diagonal. Thus it is enough to study the properties of $\mathrm{invcov}_p(D)$. Let $\mathcal{A}(D)$ be the algebra generated by the matrices D and the $m \times m$ identity matrix I_m. By the Cayley–Hamilton Theorem it is clear that

$$\mathcal{A}(D) = \Big\{\alpha_{m-1}D^{m-1} + \alpha_{m-2}D^{m-2} + \ldots + \alpha_1 D + \alpha_0 I_m \;\;:\;\; \alpha_i \in \mathbb{C}\Big\}. \tag{2.3}$$

We define $\mathcal{D}_m$ as the set of all $m \times m$ diagonal matrices.

Lemma 2.3. *Let $D = \mathrm{diag}(d_1, \ldots, d_m)$ be an $m \times m$ diagonal matrix. If $d_i \ne d_j$ for $i \ne j$ then $\mathcal{A}(D) = \mathcal{D}_m$. If $d_i = d_j$ for some $i \ne j$ then*

$$\mathcal{A}(D) = \{\mathrm{diag}(b_1, \ldots, b_i, \ldots, b_i, \ldots, b_m) \;:\; b_k \in \mathbb{C}\},$$

the set of all diagonal matrices where the i–th and j–th entries are equal.

Proof. It is clear to see $\mathcal{A}(D) \subset \mathcal{D}_m$. On the other hand, for any $B = \mathrm{diag}(b_1, \ldots, b_m) \in \mathcal{D}_m$, we form a system of linear equations,

$$\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} 1 & d_1 & d_1^2 & \ldots & d_1^{m-1} \\ \vdots & & \ldots & & \vdots \\ 1 & d_m & d_m^2 & \ldots & d_m^{m-1} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_{m-1} \end{pmatrix} := V \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_{m-1} \end{pmatrix}.$$

The matrix V is a Vandermonde matrix with $\det(V) = \prod_{i<j}(d_i - d_j)$. The matrix V is invertible by our assumption. Thus we can find a vector $(\alpha_0, \ldots, \alpha_{m-1})$ such that

$$B = \alpha_0 I_m + \alpha_1 D + \ldots + \alpha_{m-1} D^{m-1} \in \mathcal{A}(D).$$

This completes the proof.

To prove the second part we use essentially the same approach as before. ☐

Theorem 2.4. *The matrix* $\mathrm{invcov}_p(D)$ *belongs to the algebra* $\mathcal{A}(D)$.

Proof. It has been shown in Equation (39) of [17], that if the matrix D is equal to $D = \mathrm{diag}(D_n, 0_{m-n})$ where $D_n = (d_1, \ldots, d_n)$, then $\mathrm{invcov}_p(D) = \mathrm{diag}(\Lambda_L(D_n), \mu I_{m-n})$ where $\Lambda_L(D_n) = (\lambda_1, \ldots, \lambda_n)$. From Lemma 2.3, it is enough to show that if $d_i = d_j$ for some $i \neq j$, then $\lambda_i = \lambda_j$.

In part B of Section VI in [17], it was shown that

$$\lambda_k = \frac{\partial}{\partial d_k} \int_{\Omega_{p,n}} \mathrm{Tr}(\log(\Phi^* D_n \Phi)) d\phi = \frac{\partial F(d_1, \ldots, d_n)}{\partial d_k},$$

where $F(d_1, \ldots, d_n) := \int_{\Omega_{p,n}} \mathrm{Tr}(\log(\Phi^* D_n \Phi)) d\phi$. It was proved in Equation (77) in [17] that for any integer $l \geq 1$

$$\int_{\Omega_{p,n}} \mathrm{Tr}(\Phi^* D_n \Phi)^l d\phi = \sum_{k=0}^{p-1} (-1)^k c_k^{(n,p)} s_{(l-k,1^k)}(D_n),$$

where $s_{(l-k,1^k)}(D_n)$ are the Schur polynomials. By linearity and continuity, $F(d_1, \ldots, d_n)$ is symmetric. Hence assuming $d_i = d_j$, $\partial F/\partial d_i = \partial F/\partial d_j$, which implies $\lambda_i = \lambda_j$.This completes the proof. ☐

2.5. Formulas for computing $\mathbb{E}(\Phi(\Phi^* D_n \Phi)^l \Phi^*)$.

Using Lemma 1 in [17] we observe that

$$\Big(\mathbb{E}(\Phi(\Phi^* D_n \Phi)^l \Phi^*)\Big)_{ii} = \left(\int_{\Omega_{p,n}} \Phi(\Phi^* D_n \Phi)^l \Phi^* d\phi \right)_{ii} = \frac{\partial}{\partial d_i} \int_{\Omega_{p,n}} \frac{1}{l+1} \mathrm{Tr}(\Phi^* D_n \Phi)^{l+1} d\phi.$$

Using Equations (69) and (70) from [17] we see that

$$\begin{aligned} \int_{\Omega_{p,n}} \mathrm{Tr}(\Phi^* D_n \Phi)^N d\phi &= \sum_{j=0}^{p-1} (-1)^j \frac{s_{(N-j,1^j)}(I_p)}{s_{(N-j,1^j)}(I_n)} s_{(N-j,1^j)}(D_n) \\ &= \sum_{j=0}^{p-1} (-1)^j \frac{(N+p-(j+1))!(n-(j+1))!}{(N+n-(j+1))!(p-(j+1))!} s_{(N-j,1^j)}(D_n). \end{aligned} \tag{2.4}$$

Given a partition $\mu = (\mu_1, \ldots, \mu_N)$ of N. The Schur polynomial of shape μ in the variables $(d_1, \ldots, d_n)$ is defined as

$$s_\mu(D_n) = s_\mu(d_1, \ldots, d_n) = \frac{\det(d_i^{n+\mu_j-j})_{i,j=1}^n}{\det(d_i^{n-j})_{i,j=1}^n}.$$

By the Murnaghan–Nakayama rule (see Corollary 7.17.5 in [24]),

$$s_{(N-j,1^j)}(D) = \sum_{\rho=(1^{r_1},2^{r_2},\ldots,N^{r_N})\vdash N} \chi^{(N-j,1^j)}(\rho) \prod_{l=1}^{N} \frac{\mathrm{Tr}(D^l)^{r_l}}{l^{r_l} r_l!}, \tag{2.5}$$

where $\chi^{\mu}(\rho) = \sum_T (-1)^{\mathrm{ht}(T)}$ summed over all border-strip tableaux of shape μ and type ρ and $\mathrm{ht}(T)$ is the *height* of a border-strip tableaux (see Section 7.17 in [24] for more details and we will show a small example to compute (2.5) soon). Thus

$$\begin{aligned} \frac{\partial s_{(N-j,1^j)}(D_n)}{\partial d_i} &= \sum_{k=1}^{N} d_i^{k-1} \sum_{\rho=(1^{r_1},2^{r_2},\ldots,N^{r_N})\vdash N} \chi^{\lambda_j}(\rho) \frac{r_k \mathrm{Tr}(D^k)^{r_k-1}}{k^{r_k-1} r_k!} \prod_{l\neq k} \frac{\mathrm{Tr}(D^l)^{r_l}}{l_{r_l} r_l!} \\ &:= \sum_{k=1}^{N} c_{k-1} d_i^{k-1} = \sum_{k=0}^{N-1} c_k d_i^k. \end{aligned} \tag{2.6}$$

Therefore

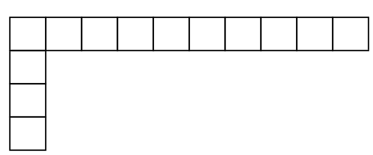
$$\begin{aligned} \big(\mathbb{E}(\Phi(\Phi^* D\Phi)^l \Phi^*)\big)_{ii} &= \frac{1}{l+1} \sum_{j=0}^{p-1} (-1)^j \frac{(l+1+p-(j+1))!(n-(j+1))!}{(l+1+n(j+1))!(p-(j+1))!} \frac{\partial s_{(N-j,1^j)}(D_n)}{\partial d_i} \\ &= \sum_{k=0}^{l} \left(\frac{c_k}{l+1} \sum_{j=0}^{p-1} (-1)^j \frac{(l+p-j)!(n-j-1))!}{(l+n-j)!(p-j-1)!} \right) d_i^k := \sum_{k=0}^{l} a_k d_i^k \end{aligned} \tag{2.7}$$

The coefficients a_k depend only on D_n, p and l. Thus we are able to show $\mathbb{E}(\Phi(\Phi^* D_n \Phi)^l \Phi^*)$ is a polynomial in D_n of degree l,

$$\mathbb{E}\Big(\Phi(\Phi^* D_n \Phi)^l \Phi^*\Big) = \sum_{k=0}^{l} a_k D_n^k.$$

Next we provide a small dimensional example to show how to apply the derived formula for computation.

2.5.1. *Small dimensional examples:* Let $\lambda_j = (N-j, 1^j)$ be the partition of N with j ones. This one has a hook shape with $N-j$ blocks in the row and $j+1$ blocks in the column.



For $l = 1$, it was shown in [17] that

$$\mathbb{E}(\Phi(\Phi^* D_n \Phi)\Phi^*) = \frac{p(np-1)}{n(n^2-1)} D_n + \frac{p(n-p)}{n(n^2-1)} \mathrm{Tr}(D_n) I_n.$$

For $l = 2$ and $\rho = (1,1,1), (1,2), (3) \vdash 3$, we list all border–strip tableaux of shape λ_j and type ρ in the following table.

	$\rho=(1,1,1)$	$\rho=(1,2)$	$\rho=(3)$
$\lambda_0=(3)$	1 2 3	1 2 2	1 1 1
$\lambda_1=(2,1)$	1 2 3 & 1 3 2	Does not exist	1 1 1
$\lambda_2=(1,1,1)$	1 2 3	1 2 2	1 1 1

Thus,

$\chi^{\lambda_j}(\rho)$	$\rho=(1,1,1)$	$\rho=(1,2)$	$\rho=(3)$
$\lambda_0=(3)$	1	1	1
$\lambda_1=(2,1)$	2	0	-1
$\lambda_2=(1,1,1)$	1	-1	1

$$s_{\lambda_0}(D)=\frac{\mathrm{Tr}(D)^3}{3!}+\frac{\mathrm{Tr}(D)\mathrm{Tr}(D^2)}{2}+\frac{\mathrm{Tr}(D^3)}{3},\qquad \frac{\partial s_{\lambda_0}}{\partial d_i}=d_i^2+\mathrm{Tr}(D)d_i+\frac{\mathrm{Tr}(D)^2+\mathrm{Tr}(D^2)}{2}$$
$$s_{\lambda_1}(D)=2\frac{\mathrm{Tr}(D)^3}{3!}-\frac{\mathrm{Tr}(D^3)}{3},\qquad \frac{\partial s_{\lambda_1}}{\partial d_i}=-d_i^2+\mathrm{Tr}(D)^2$$
$$s_{\lambda_2}(D)=\frac{\mathrm{Tr}(D)^3}{3!}-\frac{\mathrm{Tr}(D)\mathrm{Tr}(D^2)}{2}+\frac{\mathrm{Tr}(D^3)}{3},\qquad \frac{\partial s_{\lambda_2}}{\partial d_i}=d_i^2-\mathrm{Tr}(D)d_i+\frac{\mathrm{Tr}(D)^2-\mathrm{Tr}(D^2)}{2}.$$

Furthermore,

$$\begin{aligned}\big(\mathbb{E}(\Phi(\Phi^*D\Phi)^2\Phi^*)\big)_{ii}&=\frac{1}{3}\sum_{j=0}^{2}(-1)^j\frac{(2+p-j)!(n-j-1)!}{(2+n-j)!(p-j-1)!}\frac{\partial s_{\lambda_j}(D)}{\partial d_i}\\&=(c_0+c_1+c_2)d_i^2+(c_0-c_2)\mathrm{Tr}(D)d_i+c_0\frac{\mathrm{Tr}(D)^2+\mathrm{Tr}(D^2)}{2}-c_1\\&\quad+c_2\frac{\mathrm{Tr}(D)^2-\mathrm{Tr}(D^2)}{2},\end{aligned}$$

where

$$c_0=\frac{1}{3}\frac{(2+p)!(n-1)!}{(2+n)!(p-1)!},\quad c_1=\frac{1}{3}\frac{(1+p)!(n-2)!}{(1+n)!(p-2)!},\quad c_2=\frac{1}{3}\frac{p!(n-3)!}{n!(p-3)!}.$$

Finally,

$$\begin{aligned}\mathbb{E}\Big(\Phi(\Phi^*D\Phi)^2\Phi^*\Big)&=(c_0+c_1+c_2)D^2+(c_0-c_2)\mathrm{Tr}(D)D\\&\quad+\left(c_0\frac{\mathrm{Tr}(D)^2+\mathrm{Tr}(D^2)}{2}-c_1\mathrm{Tr}(D)^2+c_2\frac{\mathrm{Tr}(D)^2-\mathrm{Tr}(D^2)}{2}\right)I_n.\end{aligned}$$

3. The Ewens Estimator

Let S_m be the set of permutations of the set $[m]:=\{1,2,\dots,m\}$. The Ewens measure is a probability measure on the set of permutations that depends on a parameter $\theta>0$. In this measure, each permutation has a weight proportional to its total number of cycles. More specifically, for each permutation σ in S_m its probability is equal to

$$p_{\theta,m}(\sigma)=\frac{\theta^{K(\sigma)}}{\theta(\theta+1)\dots(\theta+m-1)},$$

where $\theta > 0$ and $K(\sigma)$ is the number of cycles in σ. The case $\theta = 1$ corresponds to the uniform measure. This measure has recently appeared in mathematical physics models (see e.g. [2] and [9]) and one has only recently started to gain insight into the cycle structures of such random permutations.

Let σ be a permutation in S_m, the corresponding permutation matrix M_σ is the $m \times m$ matrix defined as $M_\sigma(i,j) = \mathbf{1}_{\sigma(i)}(j)$. If we denote e_i to be a $1 \times m$ vector such that the i–th entry is equal to 1 and all the others entries are 0, then

$$M_\sigma = \begin{pmatrix} e_{\sigma(1)} \\ \vdots \\ e_{\sigma(m)} \end{pmatrix},$$

which is, of course, a unitary matrix. Given the sample covariance matrix K we define the new estimator for Σ as

$$K_\theta := \mathbb{E}(M_\sigma K M_\sigma^*), \tag{3.1}$$

where the expectation is taken with respect to the Ewens measure of parameter θ.

Theorem 3.1. *Let $K = (a_{ij})$ be an $m \times m$ matrix in $\mathbb{C}^{m\times m}$. Then $K_\theta = \mathbb{E}(M_\sigma K M_\sigma^*)$ is an $m \times m$ matrix such that the diagonal terms satisfy*

$$(K_\theta)_{ii} = \frac{\theta - 1}{\theta + m - 1} a_{ii} + \frac{1}{\theta + m - 1} \mathrm{Tr}(K), \tag{3.2}$$

and the non–diagonal terms ($i \neq j$) satisfy

$$\begin{aligned} (K_\theta)_{ij} &= \frac{1}{(\theta+m-2)(\theta+m-1)} \left(\theta^2 a_{ij} + (\theta-1)a_{ji} + \theta \sum_{k\neq i,j} (a_{ik} + a_{kj}) + \sum_{\substack{l\neq i, k\neq j \\ k \neq l}} a_{lk} \right) \\ &= \frac{1}{(\theta+m-2)(\theta+m-1)} \left((\theta^2 - 1)a_{ij} + (\theta-1)a_{ji} + (\theta - 1) \sum_{k\neq i,j} (a_{ik} + a_{kj}) + \sum_{l \neq k} a_{lk} \right). \end{aligned} \tag{3.3}$$

Remark 3.2. If $\theta = 1$ then

$$K_1 = \alpha \frac{ee^T}{m} + \beta(I_m - \frac{ee^T}{m}) \text{ where } \alpha = \frac{eKe^T}{m} = \frac{\sum_{i,j} a_{ij}}{m}, \beta = \frac{\mathrm{Tr}(K) - \alpha}{m-1},$$

and $e = (1, 1, \ldots, 1)$. This result already been shown in Prop. 2.2 of [25]. If $K = D = \mathrm{diag}(d_1, \ldots, d_m)$, then

$$K_\theta = \frac{\theta - 1}{\theta + m - 1} D + \frac{\mathrm{Tr}(D)}{\theta + m - 1} I_m,$$

which corresponds to diagonal loading.

Proof. First,

$$M_\sigma K M^* = \begin{pmatrix} e_\sigma(1) \\ \vdots \\ e_\sigma(m) \end{pmatrix} K \begin{pmatrix} e_\sigma^*(1) & \cdots & e_\sigma^*(m) \end{pmatrix} = \left(\sum_{l=1}^m \sum_{k=1}^m a_{kl} e_\sigma^k(i) e_\sigma^l(j) \right) = (a_{\sigma(i)\sigma(j)})_{1 \le i,j \le m}.$$

For diagonal terms,

$$
\begin{aligned}
(K_\theta)_{ii} &= \sum_{\sigma\in S_m} p_{\theta,m}(\sigma)a_{\sigma(i)\sigma(i)} = a_{ii}\sum_{\substack{\sigma\in S_m\\ \sigma(i)=i}} p_{\theta,m}(\sigma) + \sum_{\neq i} a_{ll}\sum_{\substack{\sigma\in S_m\\ \sigma(i)=l}} p_{\theta,m}(\sigma)\\
&= a_{ii}\frac{\theta}{\theta+m-1}\sum_{\tilde\sigma\in S_{m-1}} p_{\theta,m-1}(\tilde\sigma) + \sum_{l\neq i}\frac{a_{ll}}{\theta+m-1}\sum_{\hat\sigma(l)} p_{\theta,m-1}(\hat\sigma(l))\\
&= \frac{\theta}{\theta+m-1}a_{ii} + \frac{1}{\theta+m-1}\sum_{l\neq i} a_{ll} = \frac{\theta-1}{\theta+m-1}a_{ii} + \frac{1}{\theta+m-1}\mathrm{Tr}(K).
\end{aligned}
$$

Now we compute the off–diagonal terms $(K_\theta)_{ij}$ $(i\neq j)$. For $\sigma\in S_m$, if $\sigma(i)=i$ and $\sigma(j)=j$ then $\sigma=(i)(j)\sigma_1$ with $\sigma_1\in S_{m-2}$, $K(\sigma)=K(\sigma_1)+2$ and

$$
p_{\theta,m}(\sigma) = \frac{\theta^2}{(\theta+m-2)(\theta+m-1)} p_{\theta,m-2}(\sigma_1).
$$

If $\sigma(i)=j$ and $\sigma(j)=i$ we erase i and j from σ to obtain $\sigma_2\in S_{m-2}$, and

$$
p_{\theta,m}(\sigma) = \frac{\theta}{(\theta+m-2)(\theta+m-1)} p_{\theta,m-2}(\sigma_2).
$$

If $\sigma(i)=i$ and $\sigma(j)=k\neq i,j$ then $\sigma=(i)\hat\sigma$ with $\hat\sigma\in S_{m-1}$ and $K(\sigma)=K(\hat\sigma)+1$. Furthermore, we can erase j from $\hat\sigma$ to get a new permutation $\sigma_3(k)\in S_{m-2}$ such that $K(\sigma_3(k))=K(\hat\sigma)$ and finally

$$
p_{\theta,m}(\sigma) = \frac{\theta}{(\theta+m-2)(\theta+m-1)} p_{\theta,m-2}(\sigma_3(k)).
$$

Notice that $\sum_{\sigma_3(k)} p_{\theta,m-2}(\sigma_3(k)) = 1$.

If $\sigma(i)=l\neq i,j$ and $\sigma(j)=j$ then as above we can have $\sigma_4(l)\in S_{m-2}$ such that

$$
p_{\theta,m}(\sigma) = \frac{\theta}{(\theta+m-2)(\theta+m-1)} p_{\theta,m-2}(\sigma_4(l))
$$

and again $\sum_{\sigma_4(l)} p_{\theta,m-2}(\sigma_4(l)) = 1$.

If $\sigma(i)=l\neq i$ and $\sigma(j)=k\neq j$ $(k\neq l)$ we exclude the case that $\sigma(i)=j,\sigma(j)=i$ and we erase i and j from σ to obtain $\sigma_5(l,k)\in S_{m-2}$. Thus

$$
p_{\theta,m}(\sigma) = \frac{1}{(\theta+m-2)(\theta+m-1)} p_{\theta,m-2}(\sigma_5(l,k))
$$

and $\sum_{\sigma_5(l,k)} p_{\theta,m-2}(\sigma_5(l,k)) = 1$.

Therefore, for $i \neq j$

$$
\begin{aligned}
(K_\theta)_{ij} &= \sum_{\sigma \in S_m} p_{\sigma,m}(\sigma) a_{\sigma(i)\sigma(j)} \\
&= a_{ij} \frac{\theta^2}{(\theta+m-2)(\theta+m-1)} \sum_{\sigma_1 \in S_{m-2}} p_{\theta,m-2}(\sigma_1) \\
&+ a_{ji} \frac{\theta}{(\theta+m-2)(\theta+m-1)} \sum_{\sigma_2 \in S_{m-2}} p_{\theta,m-2}(\sigma_2) \\
&+ \sum_{k \neq i,j} a_{ik} \frac{\theta}{(\theta+m-2)(\theta+m-1)} \sum_{\sigma_3(k) \in S_{m-2}} p_{\theta,m-2}(\sigma_3(k)) \\
&+ \sum_{l \neq i,j} a_{lj} \frac{\theta}{(\theta+m-2)(\theta+m-1)} \sum_{\sigma_4(l) \in S_{m-2}} p_{\theta,m-2}(\sigma_4(l)) \\
&+ \sum_{\substack{l \neq i, k \neq j, k \neq l \\ \text{without } l=j, k=i}} \sum_{\sigma_5(k,l) \in S_{m-2}} a_{lk} \frac{1}{(\theta+m-2)(\theta+m-1)} p_{\theta,m-2}(\sigma_5(k,l)) \\
&= \frac{1}{(\theta+m-2)(\theta+m-1)} \left(\theta^2 a_{ij} + (\theta-1) a_{ji} + \theta \sum_{k \neq i,j} (a_{ik} + a_{kj}) + \sum_{\substack{l \neq i, k \neq j \\ k \neq l}} a_{lk} \right).
\end{aligned}
$$

□

4. Hybrid Method

In this Section, we combine the ideas of the first two methods to create a third hybrid method. First, we extend the definition of a permutation. For an integer $p \leq m$, let

$$S_{p,m} := \Big\{ \sigma \ : \sigma \text{ an injection from } \{1, 2, \ldots, p\} \text{ to } \{1, 2, \ldots m\} \Big\}.$$

The size of the set $S_{p,m}$ is $\frac{m!}{(m-p)!}$ and it is clear that $S_{m,m}$ is the set of all permutations on $[m]$. For $\sigma \in S_{p,m}$, the associated $p \times m$ matrix takes the form

$$V_\sigma := \begin{pmatrix} e_{\sigma(1)} \\ e_{\sigma(2)} \\ \vdots \\ e_{\sigma(p)} \end{pmatrix},$$

where $e_{\sigma(i)} = (e^1_{\sigma(i)}, e^2_{\sigma(i)}, \ldots, e^m_{\sigma(i)})$ is a $1 \times m$ row vector with the $\sigma(i)$–th entry 1 and all others 0. Notice

$$V_\sigma V_\sigma^{\mathrm{T}} = I_p, \tag{4.1}$$

and

$$P_\sigma := V_\sigma^{\mathrm{T}} V_\sigma = \operatorname{diag}(p_1, \ldots, p_m), \tag{4.2}$$

where

$$p_i = \sum_{l=1}^{p} (e^i_{\sigma(l)})^2 = \begin{cases} 1 & \text{if } i \in \{\sigma(1), \ldots, \sigma(p)\}, \\ 0 & \text{otherwise.} \end{cases}$$

Next, we use the *Ewens measure* on the permutation sets to define a probability on the set $S_{p,m}$. For each $\sigma \in S_{p,m}$, consider the set

$$\Omega_\sigma := \Big\{ \tilde{\sigma} \in S_m \ : \ \tilde{\sigma}_{\{1,\ldots,p\}} = \sigma \Big\}.$$

In other words, Ω_σ is the set of all permutations in S_m whose restriction to the set $\{1, 2, \ldots, p\}$ is equal to σ. Recall that $p_{\theta,m}$ is the *Ewens measure* on S_m with parameter θ. Define the probability measure on $S_{p,m}$ for $\sigma \in S_{p,m}$ as

$$\mu_{\theta,m,p}(\sigma) := p_{\theta,m}(\Omega_\sigma) = \sum_{\tilde{\sigma}\in\Omega_\sigma} p_{\theta,m}(\tilde{\sigma}). \tag{4.3}$$

Now we are ready to introduce two new operators

$$K_{\theta,m,p} := \mathbb{E}\Big(V_\sigma^T (V_\sigma K V_\sigma^T) V_\sigma\Big) \tag{4.4}$$

$$\tilde{K}_{\theta,m,p} := \mathbb{E}\Big(V_\sigma^T (V_\sigma K V_\sigma^T)^+ V_\sigma\Big), \tag{4.5}$$

where $(V_\sigma K V_\sigma^T)^+$ is the Moore–Penrose pseudo–inverse of the matrix $V_\sigma K V_\sigma^T$. We use $K_{\theta,m,p}$ as an estimate for Σ and $\tilde{K}_{\theta,m,p}$ for Σ^{-1}. Now we show a few results on these new estimators.

Theorem 4.1. *Let A be an $m \times m$ complex matrix. Then $K_{\theta,m,p}$ is an $m \times m$ matrix such that the diagonal entries are equal to*

$$(K_{\theta,m,p})_{ii} = \begin{cases} \frac{\theta+p-1}{\theta+m-1} a_{ii}, & \text{if } 1 \le i \le p, \\ \\ \frac{p}{\theta+m-1} a_{ii}, & \text{if } p+1 \le i \le m. \end{cases}$$

and the non–diagonal entries, assuming $i < j$ (if $j < i$ exchange i and j in the following expression) are equal to

$$(K_{\theta,m,p})_{ij} = \begin{cases} \frac{(\theta+p-1)(\theta+p-2)}{(\theta+m-1)(\theta+m-2)} a_{ij}, & \text{if } 1 \le i < j \le p, \\ \\ \frac{(p-1)(\theta+p-1)}{(\theta+m-1)(\theta+m-2)} a_{ij}, & \text{if } 1 \le i \le p < j \le m, \\ \\ \frac{p(p-1)}{(\theta+m-1)(\theta+m-2)} a_{ij}, & \text{if } p < i < j \le m. \end{cases}$$

Remark 4.2. In the particular case that A is a diagonal matrix $A = \mathrm{diag}(d_1, \ldots, d_m)$, then

$$K_{\theta,m,p} = \frac{p}{\theta+m-1} A + \frac{\theta-1}{\theta+m-1} \mathrm{diag}(d_1, \ldots, d_p, 0, \ldots, 0).$$

For instance, if $p = 1$ and $m = 3$ then

$$K_{\theta,3,1} = \frac{1}{\theta+2} \mathrm{diag}(\theta a_{11}, a_{22}, a_{33}).$$

Remark 4.3. In the general case with $p = 2$ and $m = 3$ then

$$K_{\theta,3,2} = \frac{1}{\theta+2} \begin{pmatrix} (\theta+1)a_{11} & \theta a_{12} & a_{13} \\ \theta a_{21} & (\theta+1)a_{22} & a_{23} \\ a_{31} & a_{32} & 2a_{33} \end{pmatrix}.$$

Proof. Recall from Equation (4.2) that

$$P_\sigma = V_\sigma^{\mathrm{T}} V_\sigma = \mathrm{diag}(p_1^\sigma, \ldots, p_m^\sigma),$$

thus $V_\sigma^T (V_\sigma A V_\sigma^T) V_\sigma = (p_i^\sigma p_j^\sigma a_{ij})_{1 \le i,j \le m}$, where

$$p_i = \sum_{l=1}^{p} (e_{\sigma(l)}^i)^2 = \begin{cases} 1 & \text{if } i \in \{\sigma(1), \ldots, \sigma(p)\}, \\ 0 & \text{otherwise.} \end{cases}$$

For the diagonal entries, if $1 \le i \le p$,

$$\begin{aligned}
(K_{\theta,m,p})_{ii} &= \sum_{\sigma\in S_{m,p}} \mu_{\theta,m,p}(\sigma)(p_i^\sigma)^2 a_{ii} = a_{ii}\sum_{l=1}^{p}\sum_{\sigma\in S_{m,p},\sigma(l)=i}\mu_{\theta,m,p}(\sigma)\\
&= a_{ii}\left(\sum_{\sigma\in S_{m,p},\sigma(i)=i}\mu_{\theta,m,p} + \sum_{l\neq i}\sum_{\sigma\in S_{m,p},\sigma(l)=i}\mu_{\theta,m,p}\right)\\
&= a_{ii}\left(\frac{\theta}{\theta+m-1}\sum_{\sigma'\in S_{m-1,p-1}}\mu_{\theta,m-1,p-1} + \frac{p-1}{\theta+m-1}\sum_{\sigma'\in S_{m-1,p-1}}\mu_{\theta,m-1,p-1}\right)\\
&= \frac{\theta+p-1}{\theta+m-1}a_{ii}.
\end{aligned}$$

If $p+1 \le i \le m$,

$$\begin{aligned}
(K_{\theta,m,p})_{ii} &= \sum_{\sigma\in S_{m,p}} \mu_{\theta,m,p}(\sigma)(p_i^\sigma)^2 a_{ii} = a_{ii}\sum_{l=1}^{p}\sum_{\sigma\in S_{m,p},\sigma(l)=i}\mu_{\theta,m,p}(\sigma)\\
&= a_{ii}\left(\frac{p}{\theta+m-1}\sum_{\sigma'\in S_{m-1,p-1}}\mu_{\theta,m-1,p-1}\right)\\
&= \frac{p}{\theta+m-1}a_{ii}.
\end{aligned}$$

For non-diagonal entries, if $1 \le i < j \le p$, which turns out to be the most complicated case, $p_i^\sigma p_j^\sigma a_{ij}$ is non zero if $i, j \in \{\sigma(1), \dots, \sigma(p)\}$. Thus

$$(K_{\theta,m,p})_{ij} = a_{ij}\sum_{s,t\in[p],s\neq t}\ \sum_{\substack{\sigma\in S_{m,p},\\ \sigma(s)=i,\sigma(t)=j}}\mu_{\theta,m,p}(\sigma).$$

We divide the previous sum into five parts:

(1) If $\sigma(i) = i$ and $\sigma(j) = j$ we “erase” i and j from the sets $[p]$ and $[m]$ to get a new injection σ_1 from $[p]\backslash\{i,j\}$ to $[m]\backslash\{i,j\}$ with $K(\sigma) = K(\sigma_1) + 2$

(2) If $\sigma(s) = i$ for some $s \in [p]\backslash\{i,j\}$ and $\sigma(j) = j$ we “erase” j from the sets $[p]$ and $[m]$ and consider s and i as one number $\tilde{s}$. Then we get a new injection $\sigma_2 : [p] \cup \tilde{s}\backslash\{i,j,s\} \to [m] \cup \tilde{s}\backslash\{i,j,s\}$ with $K(\sigma) = K(\sigma_2) + 1$

(3) If $\sigma(t) = j$ for some $t \in [p]\backslash\{i,j\}$ and $\sigma(i) = i$ then, similarly to case (2), by exchanging the roles of i and j we can get a new injection σ_3 with $K(\sigma) = K(\sigma_3) + 1$

(4) If $\sigma(s) = i$ and $\sigma(t) = j$ with $s \neq t$ for some $s \in [p]\backslash\{i\}$ and $t \in [p]\backslash\{j\}$ then we consider s and i as a new number $\tilde{s}$ and t and j as a new number $\tilde{t}$ to get a new injection $\sigma_4 : [p]\cup\tilde{s},\tilde{t}\backslash\{i,j,s,t\} \to [m]\cup\tilde{s},\tilde{t}\backslash\{i,j,s,t\}$ with $K(\sigma) = K(\sigma_4)$

(5) If $\sigma(i) = j$ and $\sigma(j) = i$ we “erase” i and j to get a new injection $\sigma_5 : [p]\backslash\{i,j\} \to [m]\backslash\{i,j\}$ with $K(\sigma) = K(\sigma_5) + 1$.

$$\begin{aligned}(K_{\theta,m,p})_{ij} &= a_{ij}\frac{\theta^2}{(\theta+m-1)(\theta+m-2)}\sum_{\sigma_1\in S_{m-2,p-2}}\mu_{\theta,m-2,p-2}(\sigma_1)\\ &+\frac{a_{ij}\theta(p-2)}{(\theta+m-1)(\theta+m-2)}\sum_{\sigma_2\in S_{m-2,p-2}}\mu_{\theta,m-2,p-2}(\sigma_2)\\ &+\frac{a_{ij}\theta(p-2)}{(\theta+m-1)(\theta+m-2)}\sum_{\sigma_3\in S_{m-2,p-2}}\mu_{\theta,m-2,p-2}(\sigma_3)\\ &+a_{ij}\frac{(p-2)^2+(p-2)}{(\theta+m-1)(\theta+m-2)}\sum_{\sigma_4\in S_{m-2,p-2}}\mu_{\theta,m-2,p-2}(\sigma_4)\\ &+\frac{a_{ij}\theta}{(\theta+m-1)(\theta+m-2)}\sum_{\sigma_5\in S_{m-2,p-2}}\mu_{\theta,m-2,p-2}(\sigma_5)\\ &=\frac{(\theta+p-1)(\theta+p-2)}{(\theta+m-1)(\theta+m-2)}a_{ij}.\end{aligned}$$

For $1\le i\le p<j\le m$ we only need consider two cases: $s=i$ and $s\neq i$,

$$\begin{aligned}(K_{\theta,m,p})_{ij} &= a_{ij}\frac{\theta(p-1)}{(\theta+m-1)(\theta+m-2)}\sum_{\sigma_1\in S_{m-2,p-2}}\mu_{\theta,m-2,p-2}(\sigma_1)\\ &+a_{ij}\frac{(p-1)^2}{(\theta+m-1)(\theta+m-2)}\sum_{\sigma_2\in S_{m-2,p-2}}\mu_{\theta,m-2,p-2}(\sigma_2)\\ &=a_{ij}\frac{(p-1)(p+\theta-1)}{(\theta+m-1)(\theta+m-2)}.\end{aligned}$$

For $p<i<j\le m$,

$$(K_{\theta,m,p})_{ij}=a_{ij}\frac{p(p-1)}{(\theta+m-1)(\theta+m-2)}.$$

□

Now we consider the estimate $\tilde{K}_{\theta,m,p}$ as in Equation (4.5). First we analyze the case when K is diagonal.

Theorem 4.4. *Let $D=D_m=\operatorname{diag}(d_1,\dots,d_n,0,\dots,0)$, then for $p\le n$,*

$$\tilde{K}_{\theta,m,p}=\mathbb{E}\Big(V_\sigma^T(V_\sigma DV_\sigma^T)^+V_\sigma\Big)=\frac{\theta+p-1}{\theta+m-1}D^+-\frac{\theta-1}{\theta+m-1}\operatorname{diag}(d_1^{-1},\dots,d_p^{-1},0,\dots,0),$$

where $D^+=\operatorname{diag}(d_1^{-1},\dots,d_n^{-1},0,\dots,0)$ by definition.

Proof. First we notice that $W_\sigma:=V_\sigma DV_\sigma^T=(\sum_{i=1}^n d_l e^l_{\sigma(i)}e^l_{\sigma(j)})_{1\le i,j\le p}$ is a diagonal matrix. For $1\le i\le p$,

$$(W_\sigma)_{ii}=\sum_{l=1}^n d_l(e^l_{\sigma(i)})^2=\begin{cases} d_{\sigma(i)} & \text{if } \sigma(i)\in[n],\\ 0 & \text{otherwise.}\end{cases}$$

Thus

$$W_\sigma=\operatorname{diag}(d_{\sigma(1)}\mathbf{1}_{\sigma(1)\in[n]},\dots,d_{\sigma(p)}\mathbf{1}_{\sigma(p)\in[n]})$$

and

$$W_\sigma^+=\operatorname{diag}\left((d_{\sigma(1)}\mathbf{1}_{\sigma(1)\in[n]})^+,\dots,(d_{\sigma(p)}\mathbf{1}_{\sigma(p)\in[n]})^+\right).$$

Next $V_\sigma^T W^+ V_\sigma = \sum_{l=1}^p (d_{\sigma(l)} \mathbf{1}_{\sigma(l)\in[n]})^+$ is still a diagonal matrix where for $1 \le i \le m$

$$(V_\sigma^T W^+ V_\sigma)_{ii} = \begin{cases} (d_{\sigma(l)} \mathbf{1}_{\sigma(l)\in[n]})^+ & \text{if } i \in \{\sigma(1), \dots, \sigma(p)\}, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $\tilde{K}_{\theta,m,p}$ is also diagonal and

$$(\tilde{K}_{\theta,m,p})_{ii} = \sum_{l=1}^{p} \sum_{\substack{\sigma\in S_{m,p}, \\ \sigma(l)=i}} \mu_{\theta,m,p}(\sigma)(d_i \mathbf{1}_{i\in[n]})^+.$$

For $1 \le i \le n$,

$$(\tilde{K}_{\theta,m,p})_{ii} = d_i^{-1} \sum_{\substack{\sigma\in S_{m,p}, \\ \sigma(l)=i}} \mu_{\theta,m,p}(\sigma) = \begin{cases} d_i^{-1} \frac{p}{\theta+m-1}, & \text{if } 1 \le i \le p, \\ \\ d_i^{-1} \frac{\theta+p-1}{\theta+m-1}, & \text{if } p+1 \le i \le n. \end{cases}$$

For $n+1 \le i \le m$, $(\tilde{K}_{\theta,m,p})_{ii} = 0$. □

Obtaining a close form expression for Equation (4.5) in the general case seems to be much more challenging. However, we are able to obtain an inductive formula with the help of a result of Kurmayya and Sivakumar's result [13].

Theorem 4.5 (Theorem 3.2, [13])**.** *Let $M = [A\ a] \in \mathbb{R}^{m\times n}$ be a block matrix, with $A \in \mathbb{C}^{m\times(n-1)}$ and $a \in \mathbb{C}^m$ being written as a column vector. Let $B = M^*M$ and $s = \|a\|^2 - a^*AA^+a$. Then if $s \neq 0$*

$$B^+ = \begin{pmatrix} (AA^*)^+ + s^{-1}(A^+a)(A^+a)^* & -s^{-1}(A^+a) \\ -s^{-1}(A^+a)^* & s^{-1} \end{pmatrix},$$

and if $s = 0$,

$$B^+ = \begin{pmatrix} (AA^*)^+ + \|b\|^2(A^+a)(A^+a)^* - (A^+a)(A^+b)^* - (A^+b)(A^+a)^* & -\|b\|^2 A^+a + A^+b \\ -\|b\|^2(A^+a)^* + (A^+b)^* & \|b\|^2 \end{pmatrix},$$

where

$$b = (A^*)^+(I + A^+a(A^+a)^*)^{-1}A^+a.$$

For a non–negative definite matrix K, one can decompose

$$K = UDU^* = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix} \begin{pmatrix} d_1 & & & \\ & d_1 & & \\ & & \ddots & \\ & & & d_m \end{pmatrix} \begin{pmatrix} u_1^* & u_2^* & \dots & u_m^* \end{pmatrix},$$

where U is a unitary matrix. Then

$$W_\sigma = V_\sigma K V_\sigma^T = \begin{pmatrix} u_{\sigma(1)} \\ u_{\sigma(2)} \\ \vdots \\ u_{\sigma(p)} \end{pmatrix} \begin{pmatrix} d_1 & & & \\ & d_1 & & \\ & & \ddots & \\ & & & d_m \end{pmatrix} \begin{pmatrix} u_{\sigma(1)}^* & u_{\sigma(2)}^* & \cdots & u_{\sigma(p)}^* \end{pmatrix}$$

$$= \begin{pmatrix} \tilde{u}_{\sigma(1)} \\ \tilde{u}_{\sigma(2)} \\ \vdots \\ \tilde{u}_{\sigma(p)} \end{pmatrix} \begin{pmatrix} \tilde{u}_{\sigma(1)}^* & \tilde{u}_{\sigma(2)}^* & \dots & \tilde{u}_{\sigma(p)}^* \end{pmatrix} := M^*M,$$

where

$$\tilde{u}_i = (\sqrt{d_1} u_i^i, \dots, \sqrt{d_m} u_i^m).$$

Let $M = [M_1 \ a]$ with $M_1 = \begin{pmatrix} \tilde{u}^*_{\sigma(1)} & \tilde{u}^*_{\sigma(2)} & \cdots & \tilde{u}^*_{\sigma(p-1)} \end{pmatrix}$ and $a = \tilde{u}^*_{\sigma(p)}$. Let $s = \|a\|^2 - a^* M_1 M_1^+ a$ and $b = (M_1^*)^+ (I + M_1^+ a (M_1^+ a)^*)^{-1} M_1^+ a$. By Theorem 4.5,

$$(M^* M)^+ = \begin{pmatrix} (M_1 M_1^*)^+ & 0 \\ 0 & 0 \end{pmatrix} + E_\sigma$$

where the matrix $E_\sigma =$

$$\begin{cases} \begin{pmatrix} s^{-1}(M_1^+ a)(M_1^+ a)^* & -s^{-1}(M_1^+ a) \\ -s^{-1}(M_1^+ a)^* & s^{-1} \end{pmatrix} & \text{if } s \neq 0, \\ \\ \begin{pmatrix} \|b\|^2 (M_1^+ a)(M_1^+ a)^* - (M_1^+ a)(M_1^+ b)^* - (M_1^+ b)(M_1^+ a)^* & -\|b\|^2 M_1^+ a + M_1^+ b \\ -\|b\|^2 (A^+ a)^* + (A^+ b)^* & \|b\|^2 \end{pmatrix} & \text{if } s = 0. \end{cases} \tag{4.6}$$

Therefore,

$$\tilde{K}_{\theta,m,p} = \mathbb{E}(V_\sigma^T \begin{pmatrix} (M_1 M_1^*)^+ & 0 \\ 0 & 0 \end{pmatrix} V_\sigma) + \mathbb{E}(V_\sigma^T E_\sigma V_\sigma) = \tilde{K}_{\theta,m,p-1} + \mathbb{E}(V_\sigma^T E_\sigma V_\sigma). \tag{4.7}$$

5. Performance and Simulations

In this Section, we study the performance of our estimators and we compare them with other traditional methods. We first focus on the case where the true covariance matrix has a Toeplitz structure. More specifically, we focus on the following two types of Toeplitz matrices.

5.1. Tridiagonal Toeplitz Matrix.

Consider an $m \times m$ symmetric tridiagonal Toeplitz matrix of the form

$$B = \begin{pmatrix} 1 & b & & & \\ b & 1 & b & & \\ & \ddots & \ddots & \ddots & \\ & & b & 1 & b \\ & & & b & 1 \end{pmatrix}.$$

Proposition 5.1.1 ([5]). *The eigenvalues and corresponding eigenvectors of B are given by*

$$\lambda_j = 1 + 2b \cos\left(\frac{\pi j}{m+1}\right),$$

and

$$v_j = \left(\sin\left(\frac{\pi j}{m+1}\right), \sin\left(\frac{2\pi j}{m+1}\right), \ldots, \sin\left(\frac{m\pi j}{m+1}\right) \right)^T \quad \textit{where } j = 1, 2, \ldots, m.$$

We are interested in the case when B is non–negative definite and the entries of B are non–negative. Therefore, it is not hard to see that b should belong to the set $\left[0, \frac{1}{2\cos(\pi/(m+1))}\right]$ for this to hold.

5.2. Power Toeplitz matrix.

An $m \times m$ power Toeplitz matrix is given by

$$A_\alpha = \begin{pmatrix} 1 & \alpha & \alpha & \ldots & \alpha^{m-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{m-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \alpha^{m-2} & \alpha^{m-3} & \cdots & 1 & \alpha \\ \alpha^{m-1} & \alpha^{m-2} & \cdots & \alpha & 1 \end{pmatrix} = \left(\alpha^{|i-j|}\right)_{1 \leq i,j \leq m}.$$

Proposition 5.2.1. *Let A_α as before then*

(1) $\det(A_\alpha) = (1 - \alpha^2)^{m-1}$.

(2) $A_\alpha \geq 0$ *if and only if* $|\alpha| \leq 1$.
(3) *For* $\alpha \neq 1$,

$$A_\alpha{}^{-1} = \frac{1}{1-\alpha^2}\begin{pmatrix} 1 & -\alpha & & & \\ -\alpha & 1+\alpha^2 & -\alpha & & \\ & \ddots & \ddots & \ddots & \\ & & -\alpha & 1+\alpha^2 & -\alpha \\ & & & -\alpha & 1 \end{pmatrix}.$$

Proof. For (1), use induction. (2) follows directly from (1). To prove (3), use the matrix inverse formula and (1). □

For our practical purposes, we consider the case when $\alpha \in [0,1)$.

5.3. **Preliminaries on the asymptotic behavior of large Toeplitz matrices.** We first collect some basic definitions and theorems regarding large Toeplitz matrices from Albrecht Böttcher and Bernd Silbermann's book [6]. For an infinite Toeplitz matrix of the form

$$A = (a_{j-k})_{j,k=0}^{\infty} = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \dots \\ a_1 & a_0 & a_{-1} & \dots \\ a_2 & a_1 & a_0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

define the *symbol* of the matrix A as

$$a = a(e^{i\varphi}) = \sum_{n=-\infty}^{\infty} a_n e^{i\varphi n},$$

for $\varphi \in [0, 2\pi]$.

Let A_m be the $m \times m$ principal minor of the matrix A. Given a Borel subset $E \subset \mathbb{C}$ we define the measures

$$\mu_m(E) = \frac{1}{m}\sum_{j=1}^{m} \mathbf{1}_E(\lambda_j^{(m)}), \tag{5.1}$$

and

$$\mu(E) = \frac{1}{2\pi}\int_0^{2\pi} \mathbf{1}_E(a(e^{i\varphi}))d\varphi, \tag{5.2}$$

where $\mathbf{1}_E$ is the characteristic function of the set E and $\{\lambda_j^{(m)}\}_{j=1}^m$ are the eigenvalues of A_m. The following classical result holds.

Theorem 5.1 (Corollary 5.12 in [6])**.** *If* $a \in L^\infty$ *is real-valued, then the measures* μ_m *given by* (5.1) *converge weakly to the measure* μ *defined by* (5.2).

5.4. **Asymptotic Behavior of Toeplitz Matrices under Ewens Estimator.** For the symmetric tridiagonal Toeplitz matrix B its symbol is

$$a(e^{i\varphi}) = 1 + be^{i\varphi} + be^{-i\varphi} = 1 + 2b\cos\varphi,$$

where $\varphi \in [0, 2\pi]$. By Theorem 1.2 in [6], the spectrum of B as m tends to infinity is supported on the interval $[1-2b, 1+2b]$. On the other hand, by Theorem 3.1, we have that

$$\begin{aligned} B_\theta &:= \mathbb{E}(M_\sigma B M_\sigma^*) \\ &= I_m + \frac{\theta^2+\theta-2}{(\theta+m-2)(\theta+m-1)} L_m + \frac{b(\theta-1)}{(\theta+m-2)(\theta+m-1)} T_m \\ &+ \frac{2b(m-1)}{(\theta+m-2)(\theta+m-1)} \left(ee^T - I_m\right), \end{aligned} \tag{5.3}$$

where

$$T_m := \begin{pmatrix} 0 & 1 & 3 & 3 & \cdots & 3 & 2 \\ 1 & 0 & 2 & 4 & \cdots & 4 & 3 \\ 3 & 2 & 0 & 2 & \cdots & 4 & 3 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 3 & 4 & 4 & \cdots & 0 & 2 & 3 \\ 3 & 4 & 4 & \cdots & 2 & 0 & 1 \\ 2 & 3 & 3 & \cdots & 3 & 1 & 0 \end{pmatrix}$$

and

$$L_m := \begin{pmatrix} 0 & b & & & \\ b & 0 & b & & \\ & \ddots & \ddots & \ddots & \\ & & b & 0 & b \\ & & & b & 0 \end{pmatrix}.$$

If θ is a fixed constant greater than 1 then as $m \to \infty$,

$$\frac{b(\theta-1)}{(\theta+m-2)(\theta+m-1)} \|T_m\| \le \frac{4m}{m^2} \to 0, \tag{5.4}$$

and

$$\frac{\theta^2+\theta-1}{(\theta+m-2)(\theta+m-1)} \|L_m\| \to 0, \tag{5.5}$$

as $m \to \infty$. Therefore, B_θ and $(1-\frac{2}{m})I_m + \frac{2}{m}ee^T$ are asymptotically equivalent sequences (see Chapter 2 of [12]) and by Theorem 2.6 in [12]

$$\lim_{m\to\infty} \mu_m^{B_\theta} = \lim_{m\to\infty} \mu_m^{(1-\frac{2}{m})I_m + \frac{2}{m}ee^T},$$

which is a rank-1 perturbation of identity matrix. Therefore,

$$\lim_{m\to\infty} \mu_m^{B_\theta} = \delta_1,$$

where δ_t is the Dirac measure at the point t. A more interesting situation happens when $\theta = \beta m$ for a fixed constant β. In this case,

$$B_\theta = I_m + \frac{\beta^2}{(\beta+1)^2} L_m + \frac{\beta}{(\beta+1)^2}\frac{1}{m} T_m + \frac{2b}{(\beta+1)^2}\frac{1}{m}(ee^T - I_m).$$

Since

$$\frac{1}{m}\mathrm{Tr}\left(\frac{\beta}{(\beta+1)^2}\frac{1}{m}T_m\right)^2 \le \frac{1}{m^3}16m^2 \to 0,$$

and

$$\frac{1}{m}\mathrm{Tr}\left(\frac{2b}{(\beta+1)^2}\frac{1}{m}(ee^T - I_m)\right)^2 \le \frac{4b^2}{m^3}m^2 \to 0,$$

as $m \to \infty$. By Lemma 2.3 in [1] the Levy metric of the empirical distributions of two $m \times m$ Hermitian matrix A, B satisfies

$$L(\mu_m^A, \mu_m^B) \leq \left(\frac{1}{m}\mathrm{Tr}(A-B)(A-B)^*\right)^{1/3}.$$

It is known (see Theorem 6, Section 4.3, [11]) that the distribution functions μ_m converges weakly to μ if and only if the Levy metric $L(\mu_m, \mu) \to 0$. Therefore

$$\lim_{m\to\infty} \mu_m^{B_\theta} = \lim_{m\to\infty} \mu_m^{I_m + (\frac{\beta}{\beta+1})^2 L_m}.$$

Thus,

$$I_m + \frac{\beta^2}{(\beta+1)^2}L_m = \frac{\beta^2}{(\beta+1)^2}B + \left(1 - \frac{\beta^2}{(\beta+1)^2}\right)I_m,$$

which is still a tridiagonal Toeplitz matrix with symbol

$$a(e^{i\varphi}) = 1 + 2b\frac{\beta^2}{(\beta+1)^2}\cos\varphi.$$

Hence the limit eigenvalue distribution is supported on the interval $\left[1 - 2b\frac{\beta^2}{(\beta+1)^2}, 1 + 2b\frac{\beta^2}{(\beta+1)^2}\right]$. The Figure below shows the estimated density function for the spectrum as θ changes.

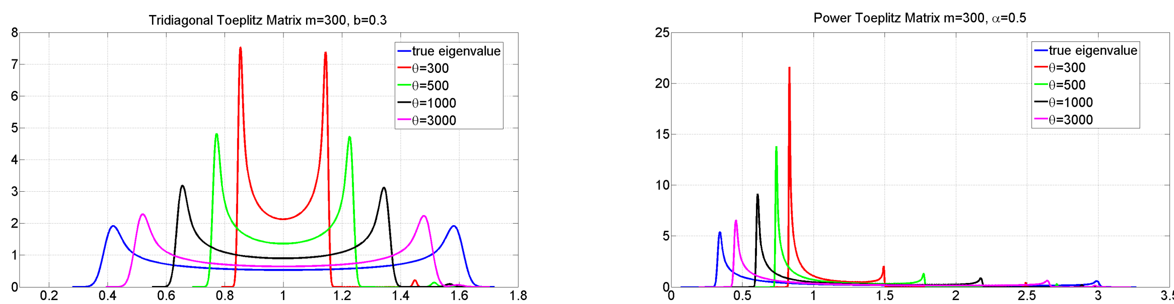


FIGURE 2. Density functions of the empirical spectral distribution of 300 × 300 tridiagonal Toeplitz matrix B with $b = 0.3$ and those of $\mathbb{E}(M_\sigma B M_\sigma^*)$ for different values of θ (left). Estimated density functions of the empirical spectral distribution of 300 × 300 power Toeplitz matrix $A_{0.5}$ and those of $\mathbb{E}(M_\sigma A_{0.5} M_\sigma^*)$ for different values of θ (right).

For the power Toeplitz matrix A_α,

$$a(e^{i\varphi}) = 1 + \frac{\alpha}{e^{i\varphi} - \alpha} + \frac{\alpha}{e^{-i\varphi} - \alpha} = 1 + 2\alpha\frac{\cos\varphi - \alpha}{(\cos\varphi - \alpha)^2 + \sin^2\varphi}.$$

Thus the spectrum of A_α as m tends to infinity is supported on $[\frac{1-\alpha}{1+\alpha}, \frac{1+\alpha}{1-\alpha}]$.

By Theorem 3.1, one can get

$$\begin{aligned} A_\theta &= \mathbb{E}(M_\sigma A_\alpha M_\sigma^*) \\ &= I_m + \frac{\theta^2 + \theta - 1}{(\theta + m - 2)(\theta + m - 1)}(A_\alpha - I_m) + \frac{\alpha(a^m - ma + m - 1 - (\theta - 1)(a - 1))}{(1-\alpha)^2(\theta + m - 2)(\theta + m - 1)}(ee^T - I_m) \\ &\quad - \frac{1}{1-\alpha}\frac{\theta - 1}{(\theta + m - 2)(\theta + m - 1)}J_m, \end{aligned} \tag{5.6}$$

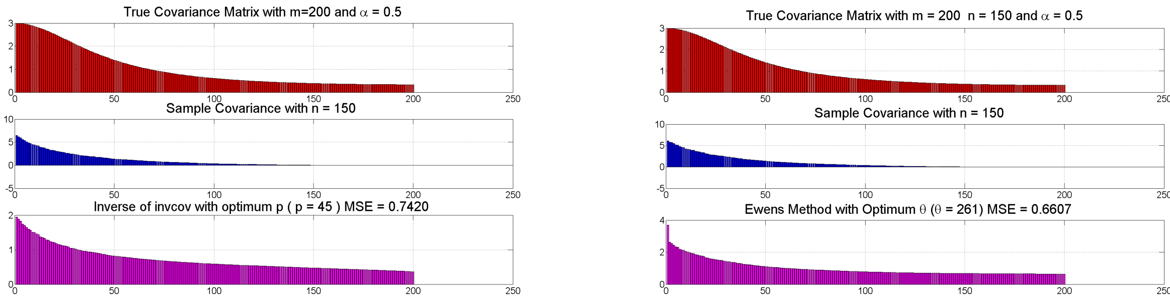


FIGURE 3. Comparison of the eigenvalue distributions of the true covariance matrix and sample covariance matrix and the *invcov* estimator vs. the *Ewens* estimator.

where $J_m = (l_{ij})$ is the matrix with diagonal entries $l_{ii} = 0$ and non–diagonal entries $l_{ij} = \alpha^i + \alpha^j + \alpha^{m+1-i} + \alpha^{m+1-j}$. In the case $\theta = \beta m$ then

$$\begin{aligned} \frac{1}{m}\mathrm{Tr}\left(\frac{1}{1-\alpha}\frac{\theta-1}{(\theta+m-2)(\theta+m-1)}J_m\right)^2 &\le \frac{1}{m^3(1-\alpha)^2}\sum_{i,j=1}^{m}(\alpha^i+\alpha^j+\alpha^{m+1-i}+\alpha^{m+1-j})^2 \\ &\le \frac{16}{m(1-\alpha)} = o(1). \end{aligned} \tag{5.7}$$

Similarly, we can show that

$$\lim_{m\to\infty}\mu_m^{A_\theta} = \lim_{m\to\infty}\mu_m^{I_m+\frac{\beta^2}{(\beta+1)^2}(A_\alpha-I_m)}.$$

For the matrix

$$I_m + \frac{\beta^2}{(\beta+1)^2}(A_\alpha - I_m) = \frac{\beta^2}{(\beta+1)^2}A_\alpha + \Big(1-\frac{\beta^2}{(\beta+1)^2}\Big)I_m,$$

one has

$$a(e^{i\varphi}) = 1 + \frac{\beta^2}{(\beta+1)^2}\Big(\frac{\alpha}{e^{i\varphi}-\alpha}+\frac{\alpha}{e^{-i\varphi}-\alpha}\Big) = 1+\frac{\beta^2}{(\beta+1)^2}\frac{2\alpha(\cos\varphi-\alpha)}{(\cos\varphi-\alpha)^2+\sin^2\varphi}.$$

Thus the limiting spectrum is supported on the interval

$$\left[1-\frac{\beta^2}{(\beta+1)^2}\frac{\alpha}{1+\alpha}, 1+\frac{\beta^2}{(\beta+1)^2}\frac{\alpha}{1-\alpha}\right].$$

5.5. **Simulations.** In this subsection, we present some simulations to test the performance of our estimators. Let A_α be an $m\times m$ Toeplitz covariance matrix with entries $a_{i,j} = \alpha^{|i-j|}$. Assume that we take n measurements and we want to recover Σ to the best of our knowledge. After performing the measurements we construct the sample covariance matrix K and proceed to recover A_α in terms of the operators $\mathrm{invcov}_p(K)$ and $\mathbb{E}(M_\sigma K M_\sigma^*)$. First we look at the eigenvalue distributions under invcov_p and *Ewens* estimators. In Figure 3, we can observe a realization of this experiment with $\alpha = 0.5$, $m = 200$ and $n = 150$. We see that the eigenvalues of A_α range roughly from 1/3 to 3. For the sample covariance matrix K, 50 eigenvalues are precisely zero. Both, the inverse of invcov_p and *Ewens* estimators give non–zero eigenvalues. The eigenvalues under the inverse of invcov_p ($p = 45$) range from 0.4 to 2 and those under *Ewens* (with $\theta = 261$) estimate from 0.6 to 2.7. Similar results were observed for other parameter values.

In Figure 4, we show the performance of the estimators for different values of p and θ. It was observed in [17] that the estimator invcov_p outperforms the more standard and classical estimator of diagonal loading for optimal loading parameters as in Ledoit and Wolf [14] by computing the Frobenius norm (MSE) $\|A_\alpha - \frac{p}{m}\mathrm{invcov}_p(K)^{-1}\|_2$

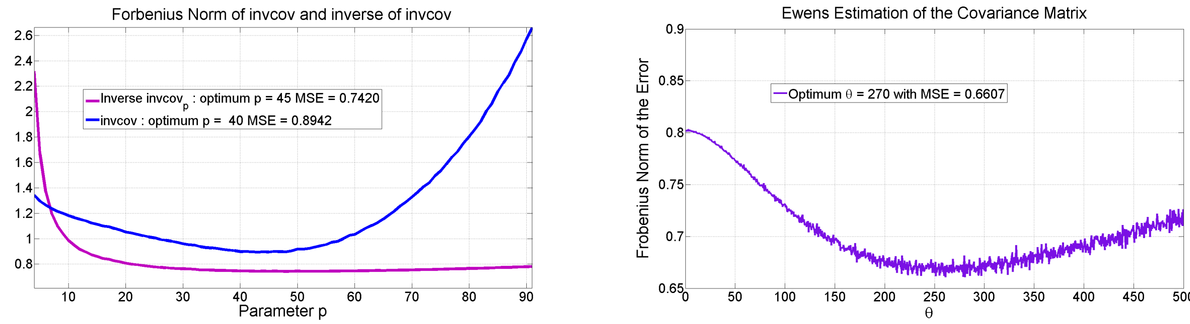


FIGURE 4. The functions f and g for $m = 200$, $n = 150$ and $\alpha = 0.5$ as functions of p (left). The function F for $m = 200$, $n = 150$ and $\alpha = 0.5$ as functions of θ (right).

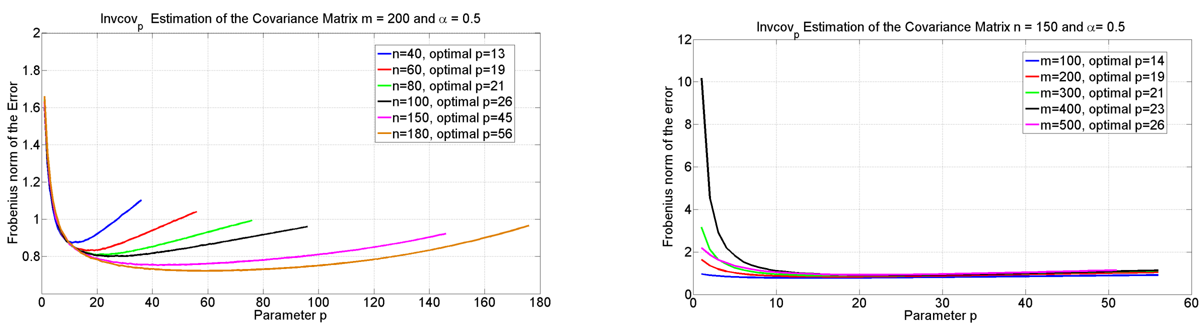


FIGURE 5. $f(m, n, \alpha, p) = \|A_\alpha - \text{invcov}_p(K)^{-1}\|_2$

for the different values of p and then computing $\|A_\alpha - K_{LW}\|_2$. The same type of experiments were performed on a variety of different scenarios as well. Let A_α, m, p, n, K and θ as before and define the functions

$$f(m, n, \alpha, p) = \|A_\alpha - (p/m)\text{invcov}_p(K)^{-1}\|_2,$$
$$g(m, n, \alpha, p) = \|A_\alpha^{-1} - (m/p)\text{invcov}_p(K)\|_2,$$
$$F(m, n, \alpha, \theta) = \|A_\alpha - \mathbb{E}(M_\sigma K M_\sigma^*)\|_2.$$

We can observe how the Ewens estimator outperforms the invcov_p estimator for the optimum values of p and θ. The next Figures show the behavior of the previous functions for different parameter values α, m, n, p and θ.

5.6. **Comparison.** Now we carry out numerical simulations of the proposed *invcov* estimator and *Ewens* estimator, and compare the performance with those of Ledoit and Wolf's method [15] and the tapering estimator established in Cai, Zhang and Zhou's paper [27]. The tapering estimator depends on the choice of parameter $1 \leq k \leq m$ which is an even number. Normally distributed random variables are used in the simulation. The experiment is conducted with the matrix size $m = 100$ and the sample size $n = 60$. Another two methods are considered here. The first is the linear shrinkage method introduced by Ledoit and Wolf [15]. We use $\hat{\Sigma}$ to denote the estimation of true covariance matrix Σ and $\hat{\Sigma}^{-1}$ as the estimation of Σ^{-1} under different methods. Three models of covariance matrices are considered. The average errors are taken over 50 replications.

Model 1: We take Σ to be a power Toeplitz matrix where $\Sigma_{ij} = a^{|i-j|}$ with $a = 0.5$. This model is studied by Cai, Zhang and Zhou [27] and their method in estimating Σ gives the smallest error.

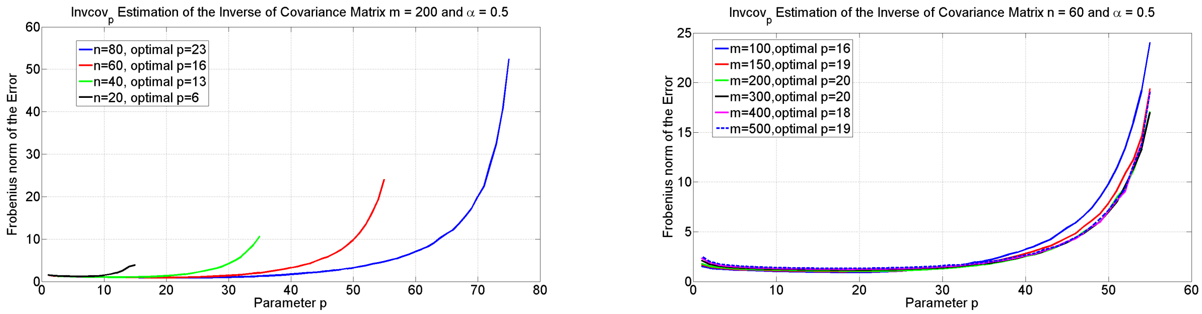


FIGURE 6. $g(m,n,\alpha,p) = \|A_\alpha^{-1} - \text{invcov}_p(K)\|_2$

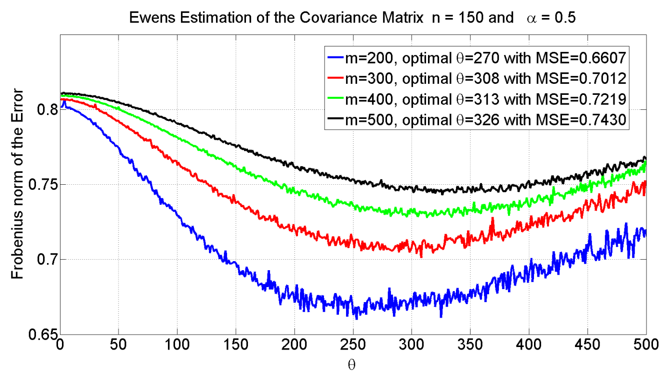


FIGURE 7. $F(m,n,\alpha,\theta) = \|A_\alpha - \mathbb{E}(M_\sigma K M_\sigma^*)\|_2$

Estimator	LW	CZZ	Invcov	Ewens
$\|\Sigma - \hat{\Sigma}\|_2$	0.6933	0.3937 ($k=4$)	0.7767 ($p=19$)	0.6752 ($\theta=117$)
$\|\Sigma - \hat{\Sigma}\|_{op}$	1.9521	1.0349 ($k=4$)	2.0235 ($p=7$)	1.7921 ($\theta=97$)

Estimator	LW	CZZ	Invcov	Ewens
$\|\Sigma^{-1} - \hat{\Sigma}^{-1}\|_2$	0.9843	1.0797 ($k=4$)	0.9261 ($p=19$)	0.8797 ($\theta=280$)
$\|\Sigma^{-1} - \hat{\Sigma}^{-1}\|_{op}$	1.8164	4.7511 ($k=4$)	0.9272 ($p=20$)	1.6536 ($\theta=253$)

Model 2: We take Σ to be a tridiagonal Toeplitz matrix where $\Sigma_{ii} = 1$, $\Sigma_{i,i+1} = \Sigma_{i-1,i} = b$ with $b = 0.2$ and others zero.

Estimator	LW	CZZ	Invcov	Ewens
$\|\Sigma - \hat{\Sigma}\|_2$	0.2772	0.2604 ($k=2$)	0.4106 ($p=10$)	0.2777 ($\theta=20$)
$\|\Sigma - \hat{\Sigma}\|_{op}$	0.4540	0.6892 ($k=2$)	0.7804 ($p=9$)	0.4052 ($\theta=4$)

Estimator	LW	CZZ	Invcov	Ewens
$\|\Sigma^{-1} - \hat{\Sigma}^{-1}\|_2$	0.3323	0.3579 ($k=2$)	0.5338 ($p=10$)	0.3275 ($\theta=23$)
$\|\Sigma^{-1} - \hat{\Sigma}^{-1}\|_{op}$	0.6671	1.2622 ($k=2$)	1.1928 ($p=8$)	0.6214 ($\theta=15$)

Model 3: We take Σ to be the long dependence matrix (see [3]) of the form

$$\Sigma_{ij} = \frac{1}{2}[(|i-j|+1)^{2H} - 2|i-j|^{2H} + (|i-j|-1)^{2H}]$$

with $H = 0.8$.

Estimator	LW	CZZ	Invcov	Ewens
$\|\Sigma - \hat{\Sigma}\|_2$	1.0448	1.0627 ($k = 46$)	1.0939 ($p = 61$)	0.7267 ($\theta = 124$)
$\|\Sigma - \hat{\Sigma}\|_{op}$	7.0422	5.1496 ($k = 56$)	7.6684 ($p = 61$)	3.1621 ($\theta = 190$)

Estimator	LW	CZZ	Invcov	Ewens
$\|\Sigma^{-1} - \hat{\Sigma}^{-1}\|_2$	0.9169	1.5193 ($k = 8$)	0.8109 ($p = 15$)	0.7308 ($\theta = 147$)
$\|\Sigma^{-1} - \hat{\Sigma}^{-1}\|_{op}$	1.8116	6.9894 ($k = 10$)	1.5363 ($p = 13$)	1.3936 ($\theta = 109$)

From these simulations, it seems the *Ewens* estimator outperforms other estimators in many cases. And it is an interesting question how to select the parameter θ to minimize the error.

References

[1] Z. D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica*, vol. 9, no. 3, pp. 611–677, 1999.

[2] V. Betz, D, Ueltschi and Y. Velenik. Random permutations with cycle weights *Ann. Appl. Probab.*, vol. 21, no. 1, pp. 312331, 2011.

[3] P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, vol. 36, no. 1, pp. 199–227, 2008.

[4] P. J. Bickel and E. Levina. Covariance regularization by thresholding. *The Annals of Statistics*, vol. 36, no. 6, pp. 2577–2604, 2008.

[5] A. Böttcher and S. M. Grudsky. Spectral properties of banded Toeplitz matrices. Society for Industrial Mathematics, 2005.

[6] A. Böttcher and B. Silbermann. Introduction to large truncated Toeplitz matrices. Springer Verlag, 1999.

[7] N. R. Draper and H. Smith. Applied Regression Analysis (Wiley Series in Probability and Statistics). Wiley-Interscience, 1998.

[8] N. El Karoui. Operator norm consistent estimation of large-dimensional sparse covariance matrices. *The Annals of Statistics*, pp. 2717–2756, 2008.

[9] N. Ercolani and D. Ueltschi. Cycle structure of random permutations with cycle weights, 2011.

[10] H. Fulton, Representation Theory, *Springer*, 1991.

[11] J. Galambos. *Advanced probability theory*, volume 10. CRC, 1995.

[12] R. M. Gray. *Toeplitz and circulant matrices: A review.* Information Systems Laboratory, Stanford University, 1971.

[13] T. Kurmayya and K. C. Sivakumar. Moore-penrose inverse of a gram matrix and its nonnegativity. *Journal of Optimization Theory and Applications*, vol. 139, no. 1, pp.201–207, 2008.

[14] O. Ledoit and M. Wolf. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Annals of statistics*, pp. 1081–1102, 2002.

[15] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[16] I. Macdonald, Symmetric functions and Hall Polynomials *Clarendon Press*, Oxford University Press, New York, 1995.

[17] T. Marzetta, G. Tucci, and S. Simon. A random matrix–theoretic approach to handling singular covariance estimates, IEEE Transactions on Information Theory, vol. 57, no. 9, pp. 6256–6271, 2011.

[18] X. Mestre. Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *Information Theory, IEEE Transactions on*, vol. 54, no. 11, pp. 5113–5129, 2008.

[19] X. Mestre and M. A. Lagunas. Diagonal loading for finite sample size beamforming: an asymptotic approach. *Robust adaptive beamforming*, pp. 201–257, 2006.

[20] R. Muirhead. Aspects of Multivariate Statistical Theory. *John Wiley & Sons*, New York, 1982.

[21] C. D. Richmond, R. Rao Nadakuditi, and A. Edelman. Asymptotic mean squared error performance of diagonally loaded capon–mvdr processor. In *Signals, Systems and Computers, 2005. Conference Record of the Thirty-Ninth Asilomar Conference on*, pp. 1711–1716, 2005.

[22] A. J. Rothman, P.J. Bickel, E. Levina and J. Zhou. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, vol. 2, pp. 494–515, 2008.

[23] B. Sagan. The Symmetric Group: Representations. *Combinatorial Algorithms, and Symmetric Functions*, Springer, 2nd edition, 2010.

[24] R. P. Stanley et al. Enumerative Combinatorics: Volume 2 *Cambridge university press Cambridge*, 1999.

[25] M. A. G. Viana. The covariance structure of random permutation matrices. *Algebraic methods in statistics and probability: AMS Special Session on Algebraic Methods and Statistics, April 8–9, 2000, University of Notre Dame, Notre Dame, Indiana*, pp. 287–303, 2001.

[26] W. B. Wu and M. Pourahmadi. Banding sample autocovariance matrices of stationary processes *Statistica Sinica*, vol. 19, no. 4, pp. 1755, 2009.

[27] T. T. Cai, C. H. Zhang, and H. H. Zhou. Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, vol. 38, no. 4, pp. 2118–2144, 2010.

[28] T. T. Cai and H. H. Zhou. Minimax estimation of large covariance matrices under l_1 norm, *Statistica Sinica*, 2011.

Gabriel H. Tucci is with Bell Labs, Alcatel–Lucent, 600 Mountain Ave, Murray Hill, NJ 07974.

E-mail address: `gabriel.tucci@alcatel-lucent.com`

Ke Wang is with the Math Department at Rutgers University, Busch Campus, Piscataway, NJ

E-mail address: `wkelucky@math.rutgers.edu`