An Innovative Approach for Analysing Rank Deficient Covariance Matrices

Gabriel H. Tucci
Bell Labs, Alcatel–Lucent
600 Mountain Ave, Murray Hill, NJ 07974
Email: gabriel.tucci@alcatel-lucent.com

Ke Wang
Math Department, Rutgers University
Busch Campus, Piscataway, NJ
E-mail: wkelucky@math.rutgers.edu

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 [13], a new approach to handle rank deficient 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 further continue in this direction and consider a few innovative methods that show considerable improvements with respect to more traditional ones such as diagonal loading. 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.

I. 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 allow 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 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 \( X = (X_1, \ldots, X_m)^T \in \mathbb{C}^{m \times 1} \) and assume for simplicity that \( X \) is centered. Then the true covariance matrix is given by

\[
\Sigma = \mathbb{E}(XX^*) = (\text{cov}(X_i, X_j))_{1 \leq i, j \leq m}. \tag{1}
\]

Consider \( n \) independent samples or realizations \( x_1, \ldots, x_n \in \mathbb{C}^{m \times 1} \) and form the \( m \times n \) data matrix \( M = (x_1, \ldots, x_n) \). Then the sample covariance matrix is an \( m \times m \) non-negative definite matrix defined as

\[
K = \frac{1}{n} MM^*. \tag{2}
\]

If \( n \to +\infty \) and \( m \) is fixed, then the sample covariance matrix \( K \) converges (entry-wise) to \( \Sigma \) 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 almost surely rank deficient (singular). Our target in this paper is to recover the true covariance matrix \( \Sigma \) from \( K \) under the condition \( n < m \). The conventional treatment of covariance singularity artificially converts the rank deficient 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” [5], [17]. Consider \( \alpha K + \beta I_m \) as an estimate of \( \Sigma \), where \( \alpha, \beta \) 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 uniformly rescaled and shifted versions of the eigenvalues of \( K \). There are many methods of choosing the optimum loading parameters, see [11], [14] and [15].

In Marzetta, Tucci and Simon’s paper [13], 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

\[
\Omega_{p,m} = \{ \Phi \in \mathbb{C}^{p \times m} : \Phi \Phi^* = I_p \}. \tag{3}
\]
The topology on $\Omega_{p,m}$ is the subspace topology inherited from $\mathbb{C}^{p\times m}$. With this topology $\Omega_{p,m}$ is a compact manifold, called the Stiefel manifold, whose dimension is $\dim(\Omega_{p,m}) = 2mp - p^2$. Endow $\Omega_{p,m}$ with the Haar measure, that is, the uniform distribution on the set $\Omega_{p,m}$. Then define the operators

$$
\text{cov}_{p}(K) = E(\Phi^*(\Phi K \Phi^*) \Phi),
$$

and

$$
\text{invcov}_{p}(K) = E(\Phi^*(\Phi K \Phi^*)^{-1} \Phi),
$$

where the expectation is taken with respect to the Haar measure. Surprisingly, it was found that

$$
\text{cov}_{p}(K) = \frac{p}{(m^2 - 1)m} \left( (mp - 1)K + (m - p)\text{Tr}(K)I_m \right),
$$

which is the same as diagonal loading. Moreover, the properties of $\text{invcov}_{p}(K)$ were investigated. In particular, it was shown that if $K$ is decomposed as $K = UDU^*$, with $D = \text{diag}(d_1, \ldots, d_n, 0, \ldots, 0)$, then

$$
\text{invcov}_{p}(K) = U\text{invcov}_{p}(D)U^*.
$$

In other words, $\text{invcov}_{p}(K)$ preserves the eigenvectors of $K$, and transforms all the zero eigenvalues to a non-zero constant. Formulas to compute the values of $\lambda_i$ and $\mu$ were provided and the asymptotic behavior of $\text{invcov}_{p}(D)$ was studied using techniques from free probability.

In this paper, we investigate new methods to estimate rank deficient covariance matrices. In Section II, we consider a new approach, called the Ewens estimator, to estimate $\Sigma$. 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 by a combinatorial argument. In Section III, we combine the ideas of the first two methods, extend the definition of permutation matrices to get $p \times m$ unitary matrices, and define two new operators $K_\theta,m,p$ and $K_\Phi,m,p$ to estimate $\Sigma$ and $\Sigma^{-1}$ respectively. We provide an explicit formula for $K_\Phi,m,p$ and an inductive formula to compute $K_\theta,m,p$. In Section IV, it is assumed that $\Sigma$ has some special values, 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.

II. THE EWENS ESTIMATOR

Let $S_m$ be the set of permutations of the set $[m] := \{1, \ldots, m\}$. 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 $\sigma$ in $S_m$, its probability is equal to

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

where $\theta > 0$ and $K(\sigma)$ is the number of cycles in $\sigma$. The case $\theta = 1$ corresponds to the uniform measure. This measure has recently appeared in mathematical physics models (see e.g. [2] and [6]) and has recently started to gain insight into the cycle structures of random permutations.

Let $\sigma$ be a permutation in $S_m$, the corresponding permutation matrix $M_\sigma$ is an $m \times m$ matrix defined as $M_\sigma(i,j) = 1_{\sigma(i)(j)}$. If we denote $e_i$ to be the $1 \times m$ vector with the $i$-th entry equal to 1 and all others entries equal to 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 $\Sigma$ as

$$
K_\theta := E(M_\sigma K M_\sigma^*),
$$

where the expectation is taken with respect to the Ewens measure of parameter $\theta$.

Theorem 1. Let $K = (a_{ij})$ be an $m \times m$ complex matrix. Then $K_\theta = E(M_\sigma K M_\sigma^*)$ is an $m \times m$ matrix such that the diagonal terms satisfy

$$
\begin{align*}
K_{\theta}^{(i)} &= \frac{\theta - 1}{\theta + m - 1} a_{ii} + \frac{1}{\theta + m - 1} \text{Tr}(K), \\
\end{align*}
$$

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

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

Remark 1. If $\theta = 1$ then $K_1 = \frac{1}{m} \text{Tr}(K) I_m - \frac{e^T e}{m}$ where $e = (1, 1, \ldots, 1)$. This has already been computed in [20], Prop. 2.2. If $K = D = \text{diag}(d_1, \ldots, d_m)$, then

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

which corresponds to a diagonal loading.

III. 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

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

The size of the set $S_{p,m}$ is $\frac{m!}{(m-p)!p!}$ and in the case $p = m$, $S_{m,m}$ 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 $e_{\sigma(i)} = (e_{\sigma(1)}^1, e_{\sigma(2)}^2, \ldots, e_{\sigma(m)}^m)$ is a $1 \times m$ row vector with the $\sigma(i)$-th entry 1 and all others equal to 0. Notice that $V_\sigma V_\sigma^T = I_p$ and

$$P_{\sigma} := V_{\sigma}^T V_\sigma = \text{diag}(p_1, \ldots, p_m), \quad (8)$$

where

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

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

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

In other words, $\Omega_\sigma$ is the set of all permutations in $S_m$ whose restriction to the set $\{1, 2, \ldots, p\}$ is equal to $\sigma$. Recall that $\mu_{\theta,m,p}$ is the Ewens measure on $S_m$ with parameter $\theta$. Now we define the probability of $\sigma \in S_{p,m}$ as

$$\mu_{\theta,m,p}(\sigma) := \mu_{\theta,m,p} (\Omega_\sigma) = \sum_{\tilde{\sigma} \in \Omega_\sigma} p_{\theta,m}(\tilde{\sigma}). \quad (9)$$

Now we are ready to introduce two new operators

$$K_{\theta,m,p} := E\left(V_{\sigma}^T (V_{\sigma} K V_{\sigma}^T) V_{\sigma}\right), \quad (10)$$

and

$$\tilde{K}_{\theta,m,p} := E\left(V_{\sigma}^T (V_{\sigma} K V_{\sigma}^T)^+ V_{\sigma}\right), \quad (11)$$

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

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

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

Now we consider the estimate $\tilde{K}_{\theta,m,p}$ as in Equation (11). First we analyze the case when $K$ is diagonal.

**Theorem 3.** Let $D = \text{diag}(d_1, \ldots, d_p, 0, \ldots, 0)$, then

$$\tilde{K}_{\theta,m,p} = E\left(V_{\sigma}^T (V_{\sigma} D V_{\sigma}^T)^+ V_{\sigma}\right) = \frac{\theta + p - 1}{\theta + m - 1} D^+.$$ 

Obtaining a close form expression for Equation (11) in the general case seems to be much more challenging. However, we are able to give an inductive formula with the help of a result of Kurmayya and Sivakumar’s [10]. Unfortunately, we omit these results for space limitations.

### IV. Performance and Simulations

In this Section we study the performance of our estimators and compare them with traditional methods. We 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.

**A. Tridiagonal Toeplitz Matrix**

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

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

**B. Power Toeplitz matrix**

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

$$A_\alpha = \begin{pmatrix} 1 & \alpha & \alpha & \ldots & \alpha^{m-1} \\ \alpha & 1 & \alpha & \ldots & \alpha^{m-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \alpha^{m-2} & \alpha^{m-3} & \ldots & 1 & \alpha \\ \alpha^{m-1} & \alpha^{m-2} & \ldots & \alpha & 1 \end{pmatrix}.$$ 

**C. 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 [4]. For an infinite Toeplitz matrix of the form

$$A = (a_{j-k})_{k=0}^\infty = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \ldots \\ a_1 & a_0 & a_{-1} & \ldots \\ a_2 & a_1 & a_0 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$
define the symbol of the matrix $A$ as
\[
a = a(e^{i\varphi}) = \sum_{n=-\infty}^{\infty} a_n e^{i\varphi n},
\]
for $0 \in [0, 2\pi]$. 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
\[
\mu_m(E) = \frac{1}{m} \sum_{j=1}^{m} \chi_E(\lambda_j^{(m)}), \tag{12}
\]
and
\[
\mu(E) = \frac{1}{2\pi} \int_{0}^{2\pi} \chi_E(a(e^{i\varphi})) d\varphi, \tag{13}
\]
where $\chi_E$ 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 4** (Corollary 5.12 in [4]). If $a \in L^{\infty}$ is real-valued, then the measures $\mu_m$ given by (12) converge weakly to the measure $\mu$ defined by (13).

**D. Asymptotic Behavior of Toeplitz Matrices under Ewens Estimator**

For the symmetric tridiagonal Toeplitz matrix $B$ its symbol is
\[
a(e^{i\varphi}) = 1 + b e^{i\varphi} + \bar{b} e^{-i\varphi} = 1 + 2b \cos \varphi,
\]
where $\varphi \in [0, 2\pi]$. By Theorem 1.2 in [4], the spectrum of $B$ as $m$ tends to infinity is supported on the interval $[1 - 2b, 1 + 2b]$. On the other hand, by Theorem 1, we have that
\[
B_{\theta} := E(M_{\theta}BM_{\theta}^*)
= I_m + \frac{\theta^2 + \theta - 2}{(\theta + m - 2)(\theta + m - 1)} L_m
+ \frac{b(\theta - 1)}{(\theta + m - 2)(\theta + m - 1)} T_m
+ \frac{2b(m - 1)}{(\theta + m - 2)(\theta + m - 1)} (ee^T - I_m), \tag{14}
\]
where
\[
\begin{pmatrix}
0 & 1 & 3 & 3 & \cdots & 3 & 2 \\
1 & 0 & 2 & 4 & \cdots & 4 & 3 \\
3 & 2 & 0 & 2 & \cdots & 4 & 3 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
3 & 4 & 4 & \cdots & 0 & 2 & 3 \\
3 & 4 & 4 & \cdots & 2 & 0 & 1 \\
2 & 3 & 3 & \cdots & 3 & 1 & 0
\end{pmatrix},
\]
\[
T_m := \begin{pmatrix}
0 & b & 0 & b \\
b & 0 & b & 0 \\
\ddots & \ddots & \ddots & \ddots \\
b & 0 & b & 0 \\
b & 0 & b & 0
\end{pmatrix},
\]
and
\[
L_m := \begin{pmatrix}
0 & b & 0 & b \\
\vdots & \ddots & \ddots & \ddots \\
b & 0 & b & 0 \\
b & 0 & b & 0
\end{pmatrix}.
\]

If $\theta$ is a fixed constant greater than $1$ then as $m \to \infty$,
\[
\frac{b(\theta - 1)}{(\theta + m - 2)(\theta + m - 1)} \|T_m\|_{\infty} \leq \frac{4m}{m^2} \to 0, \tag{15}
\]
and
\[
\frac{\theta^2 + \theta - 1}{(\theta + m - 2)(\theta + m - 1)} \|L_m\|_{\infty} \to 0. \tag{16}
\]

Therefore, $B_{\theta}$ and $(1 - \frac{2}{m})I_m + \frac{2}{m}ee^T$ are asymptotically equivalent matrices (see Chapter 2, [9]) and by Theorem 2.6 in [9],
\[
\lim_{m \to \infty} \mu_{B_{\theta}} = \lim_{m \to \infty} \mu_{(1 - \frac{2}{m})I_m + \frac{2}{m}ee^T},
\]
which is a rank one perturbation of identity matrix. Therefore,
\[
\lim_{m \to \infty} \mu_{B_{\theta}} = \delta_1,
\]
where $\delta_1$ is the Dirac measure at the point $1$. A more interesting situation happens when $\theta = \beta m$ for a fixed constant $\beta$. In this case,
\[
\lim_{m \to \infty} \mu_{B_{\theta}} = \lim_{m \to \infty} \mu_{I_m + (\frac{\beta^2}{\pi^2})L_m}.
\]

Hence,
\[
I_m + \frac{\beta^2}{(\beta + 1)^2} L_m = \frac{\beta^2}{(\beta + 1)^2} B + \left(1 - \frac{\beta^2}{(\beta + 1)^2}\right)I_m,
\]
which is still a tridiagonal Toeplitz matrix with symbol
\[
a(e^{i\varphi}) = 1 + 2b \frac{\beta^2}{(\beta + 1)^2} \cos \varphi.
\]

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

![Fig. 1. This Figure shows the density functions of the empirical spectral distribution of $300 \times 300$ tridiagonal Toeplitz matrix $B$ with $b = 0.3$ and those of $B(M_{\theta}BM_{\theta}^*)$ for different values of $\theta$.](image)

Similar results were obtained for the power Toeplitz matrix $A_{\alpha}$.

**E. Simulations**

In this Section we present some simulations to test the performance of our estimators. Let $A_{\alpha}$ be an $m \times m$ Toeplitz matrix with entries $a_{i,j} = a^{(i-j)}$. Given $n$ observations we want to recover the matrix $A_{\alpha}$. We construct the sample covariance matrix $K$ and proceed to recover $A_{\alpha}$ in terms of the operators...
The functions $f$ and $g$ for $m = 200$, $n = 150$ and $\alpha = 0.5$ as functions of $p$.

In Figures 2 and 3, we show the performance of the estimators for different values of $p$ and $\theta$. It was observed in [13] that the estimator $\text{invcov}_p$ outperforms the more standard and classical estimator of diagonal loading for optimal loading parameters as in Ledoit and Wolf [11] by computing $f(m, n, \alpha, p)$, for the different values of $p$, and comparing them with $\|A_{\alpha} - K_{LW}\|_2$.

The same type of experiments were performed on a variety of different scenarios as well. We can observe how the Ewens estimator outperforms the $\text{invcov}_p$ estimator for the optimum values of $p$ and $\theta$. The next Figures show the behavior of the previous functions for the different parameter values.

REFERENCES