Journal of Multivariate Analysis 139 (2015) 124–146

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Hypergeometric functions of matrix arguments and linear statistics of multi-spiked Hermitian matrix models

Damien Passemier [a,*], Matthew R. McKay [a], Yang Chen [b]

[a] Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

[b] Department of Mathematics, University of Macau, Avenue Padre Tomás Pereira, Taipa Macau, China

ARTICLE INFO

Article history:
Received 3 June 2014
Available online 10 March 2015

AMS subject classification:
33C20
60B20
62H10

Keywords:
Random matrix theory
High-dimensional statistics
Spiked population model
Hypergeometric function
Wishart ensembles
F-matrix

ABSTRACT

This paper derives central limit theorems (CLTs) for general linear spectral statistics (LSS) of three important multi-spiked Hermitian random matrix ensembles. The first is the most common spiked scenario, proposed by Johnstone, which is a central Wishart ensemble with fixed-rank perturbation of the identity matrix, the second is a non-central Wishart ensemble with fixed-rank noncentrality parameter, and the third is a similarly defined non-central F ensemble. These CLT results generalize our recent work Passemier (2015) to account for multiple spikes, which is the most common scenario met in practice. The generalization is non-trivial, as it now requires dealing with hypergeometric functions of matrix arguments. To facilitate our analysis, for a broad class of such functions, we first generalize a recent result of Onatski (2014) to present new contour integral representations, which are particularly suitable for computing large-dimensional properties of spiked matrix ensembles. Armed with such representations, our CLT formulas are derived for each of the three spiked models of interest by employing the Coulomb fluid method from random matrix theory along with saddlepoint techniques. We find that for each matrix model, and for general LSS, the individual spikes contribute additively to yield a $O(1)$ correction term to the asymptotic mean of the linear statistic, which we specify explicitly, whilst having no effect on the leading order terms of the mean or variance.

1. Introduction

Linear spectral statistics (LSS) are of fundamental importance in multivariate analysis. Such statistics are characterized, quite generally, by summations of functions of the individual eigenvalues of random matrices. Of prime interest are sample covariance matrices, constructed based on m observations (samples) of a n-dimensional random vector (variables), or suitably defined F matrices. Many classical results are available which specify the asymptotic distribution of certain LSS as the number of observations m become asymptotically large, for fixed n (see., e.g., [2,33]). However, modern applications often deal with high-dimensional data sets, for which n and m have similar order, and thus classical asymptotics no longer apply. This has inspired a new wave of research, aimed at characterizing the limiting distributions of LSS in the double-asymptotic regime, with n and m both large, using tools from asymptotic random matrix theory. The asymptotic behavior is typically found to be markedly different from the classical asymptotic setting, whilst giving substantially improved accuracy for various practical applications.

* Corresponding author.
E-mail addresses: damien.passemier@gmail.com (D. Passemier), eemckay@ust.hk (M.R. McKay), yayangchen@umac.mo (Y. Chen).

http://dx.doi.org/10.1016/j.jmva.2015.03.001

Under double-asymptotics, central limit theorems (CLTs) for LSS of various random matrix ensembles have now been derived, providing generic asymptotic formulas for the limiting mean and variance (see, e.g., [16,26,50,78]). Much of this attention has focused on scenarios with identity population covariance (e.g., [16,50,3]), although some results for more general matrix models have also appeared [9].

In this paper, our main focus is on three related classes of so-called "spiked" Hermitian random matrix ensembles: (i) central Wishart with finite-rank perturbation of the identity (proposed in [38], which we refer to as "Johnstone's spiked model"), (ii) non-central Wishart with fixed-rank noncentrality, and (iii) similarly defined non-central F matrices. In our recent work [65], we demonstrated that each of these models shared a common feature, with their joint eigenvalue distributions being expressible in a similar contour-integral form. These were initially derived in [32,75,55,63,24] for Johnstone's spiked model and the non-central Wishart spike model respectively, whilst given as a new result in [65] for the spiked F model. If one disregards the contour integral, then the joint eigenvalue density in each case yields the same general structure as typical of "classical" Hermitian random matrix ensembles (i.e., the Gaussian, Laguerre and Jacobi unitary ensembles [54]), albeit with a modified weight function. This is particularly important, as it allows one to employ powerful methods designed for such classical ensembles in the study of spiked models. This was precisely the approach undertaken in [65], where we employed the framework [16], designed for non-spiked models based on Dyson's Coulomb fluid method [30] (see also [17,18,15,11,70,74,23]), as well as saddlepoint integration techniques, to derive CLTs for each of the three spiked models under consideration as the matrix dimensions grew large.

Our recent results in [65] assumed the presence of a single spike only, which for many practical applications may not be reasonable. Examples include [2] in psychology, [45,58,29,12,20,21] in signal processing, [59,44] in physics of mixture, [14,73] in finance, [22,69] in computational immunology/virology, [6,7] in statistics, in addition to others (e.g., [38,67]). For scenarios with multiple spikes, there are relatively few existing results concerning LSS, and the results which are available focus primarily on Johnstone's spiked model. These include [76], which derived a CLT for general LSS, expressing the limiting mean and variance in terms of contour integrals, as well as [62,68,77,64], which considered specific linear statistics (i.e., for specific applications). For alternative spiked models, such as the non-central Wishart and F scenarios, results concerning LSS are currently absent, beyond the single-spike scenario considered in [65]. Yet, CLTs of these two multi-spiked models have wide applications. In wireless communication, multi-spiked non-central Wishart matrices arises when considering the mutual information of multiple-input multiple-output (MIMO) systems with Rician fading [65,39,36,56,72], whereas multi-spike F matrices are encountered in the capacity of multiuser MIMO models [53,41,40]. This is also the case in the classical multiple sample significance test, where the likelihood statistic is a LSS of a non-central F matrix (see [65] for the single-spike case, [7] for the non-spiked case).

The primary objective of this paper is to close this gap by deriving CLTs for general LSS under all three spiked models indicated above, allowing for arbitrary numbers of spikes. This generalization is substantial, since one can no longer rely on the contour-integral-based joint eigenvalue densities in [75,55,63,24]. Thus, the first major step is to obtain expressions for the joint eigenvalue densities which, for the three spiked models, are classically expressed in terms of hypergeometric functions of matrix arguments [37]. Such functions are fundamental objects arising in multivariate analysis and random matrix theory, and are often difficult to handle. Quite generally, they are denoted

$$ {}_pF_q^{(\alpha)}\left(a_1,\ldots,a_p;b_1,\ldots,b_q;\mathbf{X},\mathbf{Y}\right), \tag{1} $$

where $\mathbf{X}$ and $\mathbf{Y}$ are $n\times n$ Hermitian matrix arguments. For $\alpha = 2, 1,$ and $1/2$, these functions are often described as solutions to matrix integrals over the orthogonal, unitary, or symplectic groups respectively (see, for example, [57,35,37]); alternatively, they may be described via infinite series expansions involving Jack polynomials. In addition, for the special case $\alpha = 1$ (only), they may also be written as $n \times n$ determinants of scalar ${}_pF_q$ functions [42,34]. Whilst computational algorithms have been developed recently (e.g., [43]), such representations are still difficult to describe when the matrix dimension n is large, which is often the realm of interest for problems in random matrix theory.

For each of our three spiked models of interest, the joint eigenvalue density involves a specific particularization of (1): For Johnstone's spiked model it is a ${}_0F_0^{(1)}$, for the non-central Wishart spiked model it is a ${}_0F_1^{(1)}$, whilst for the non-central F spiked model it is a ${}_1F_1^{(1)}$. However, it turns out that for all three cases, one of the matrix arguments in the associated hypergeometric function has rank $r \leq n$, with r denoting the number of spikes. For such a reduced-rank scenario, in some exciting and very recent work by Onatski [62], a new representation was derived for ${}_0F_0^{(\alpha)}$ in (1), which was expressed as a r-dimensional contour integral involving a ${}_0F_0^{(\alpha)}$ function of only $r \times r$ matrix arguments (rather than $n \times n$ matrix arguments). This is critical for facilitating the large dimensional analysis of spiked random matrix models, under the asymptotics $n \to \infty$ with r fixed, since the dimension of the matrix arguments does not explode under such asymptotics. Effectively, the result in [62] is a generalization of the previous contour-integral formulations in [32,55,75,63], which applied for the case $r = 1$. We also mention that, for the same case $r = 1$, an additional generalization of [32,55,75,63] was presented very recently in [25], which provided an analogous contour-integral formula for the case of ${}_pF_q^{(\alpha)}$ with general p and q.

In this paper, to facilitate analysis of LSS of multi-spiked random matrix models, we provide a necessary generalization of the result in [62] beyond the case $p = 0, q = 0$, by deriving a new r-dimensional contour-integral representation for the hypergeometric function (1), involving a reduced complexity hypergeometric function with $r \times r$ matrix arguments. This result applies for arbitrary r, arbitrary p and q, and under some mild conditions on α. We keep this discussion general