Spectral moments of correlated Wishart matrices

Zdzisław Burda,* Jerzy Jurkiewicz,† and Bartłomiej Wacław‡

*Mark Kac Center for Complex Systems Research and Marian Smoluchowski Institute of Physics, Jagellonian University,
ul. Reymonta 4, 30-059 Kraków, Poland*

We present an analytic method to determine the spectral properties of the covariance matrices constructed of correlated Wishart random matrices. The method gives, in the limit of large matrices, exact analytic relations between the spectral moments and the eigenvalue densities of the covariance matrices and their estimators. The results can be used in practice to extract the information about genuine correlations from the given experimental realization of random matrices.

Wishart random matrices play an important role in the multivariate statistical analysis [1,2]. They are useful in some problems of fundamental physics [6], communication and information theory [3–5], internet trading [7], and quantitative finance [8–10].

A Wishart ensemble of correlated random matrices is defined by a Gaussian probability measure:

$$P(\mathbf{X})D\mathbf{X} = \mathcal{N}^{-1} \exp\left[-\frac{1}{2}\text{Tr } \mathbf{X}^\tau \mathbf{C}^{-1}\mathbf{X}\mathbf{A}^{-1}\right] \prod_{i,\alpha=1}^{N,T} dX_{i\alpha}, \quad (1)$$

where $\mathbf{X}=(X_{i\alpha})$ is a real rectangular matrix of dimension $N \times T$. It has two types of indices: an N-type index running over the set $i=1,\dots,N$ and a T-type index over $\alpha=1,\dots,T$. Throughout the paper the N-type indices will be denoted by Latin letters and the T-type by Greek ones. $\mathbf{X}^\tau$ denotes the transpose of $\mathbf{X}$. The matrices $\mathbf{C}=(C_{ij})$ and $\mathbf{A}=(A_{\alpha\beta})$ are symmetric square matrices of dimensions $N \times N$ and $T \times T$, respectively. They are positive definite. $\mathcal{N}$ is a normalization constant:

$$\mathcal{N}=(2\pi)^{NT/2}(\det \mathbf{C})^{T/2}(\det \mathbf{A})^{N/2}, \quad (2)$$

chosen to have $\int P(\mathbf{X})D\mathbf{X}=1$.

Let $Q(\mathbf{X})$ be a quantity depending on $\mathbf{X}$. The average of Q over random matrix ensemble (1) is defined as

$$\langle Q \rangle = \int Q(\mathbf{X})P(\mathbf{X})D\mathbf{X}. \quad (3)$$

In particular, the two-point correlation function is

$$\langle X_{i\alpha}X_{j\beta} \rangle = C_{ij}A_{\alpha\beta}, \quad (4)$$

as directly follows from the Gaussian integration. In this paper we are interested in the spectral behavior, the eigenvalue distribution and the spectral moments of the following random matrices:

$$\mathbf{c} = \frac{1}{T}\mathbf{X}\mathbf{X}^\tau, \quad \mathbf{a} = \frac{1}{N}\mathbf{X}^\tau\mathbf{X}. \quad (5)$$

These matrices can be used as estimators of the correlation matrices $\mathbf{C}$ and $\mathbf{A}$ if some realization of random matrices $\mathbf{X}$ are given. We will refer to $\mathbf{c}$ and $\mathbf{a}$ as to covariance matrices or statistically dressed correlation matrices. We will present an analytic method to determine the eigenvalue distribution and the spectral moments of $\mathbf{c}$ and $\mathbf{a}$ in the limit of large matrix size. Another method of calculating the eigenvalue density of correlated Wishart matrices has been recently discussed in [11].

In parallel to Eq. (1) one can define a Wishart ensemble of correlated complex matrices:

$$\mathcal{P}(\mathbf{X})D\mathbf{X} = \mathcal{N}^{-1} \exp[-\text{Tr } \mathbf{X}^\dagger \mathbf{C}^{-1}\mathbf{X}\mathbf{A}^{-1}]\prod_{i,\alpha} dX_{i\alpha}^{re}dX_{i\alpha}^{im}. \quad (6)$$

The matrices $\mathbf{C}$ and $\mathbf{A}$ are now Hermitian and positive definite. $\mathbf{X}^\dagger$ denotes the Hermitian conjugate of $\mathbf{X}$. The normalization constant is now $\mathcal{N}=(\pi)^{NT}(\det \mathbf{C})^T(\det \mathbf{A})^N$. In the analysis of the complex ensemble the estimators (5) of the correlation matrices are replaced correspondingly by

$$\mathbf{c} = \frac{1}{T}\mathbf{X}\mathbf{X}^\dagger, \quad \mathbf{a} = \frac{1}{N}\mathbf{X}^\dagger\mathbf{X}. \quad (7)$$

Notice that the factor one half in front of the trace in the measure for real matrices (1) is dropped in Eq. (6). With this choice of the measure the two-point correlations take the similar form as for real matrices (4):

$$\langle X_{i\alpha}X_{j\beta}^* \rangle = C_{ij}A_{\alpha\beta}. \quad (8)$$

The star stands for the complex conjugation. Additionally we also have

$$\langle X_{i\alpha}X_{j\beta} \rangle = \langle X_{i\alpha}^*X_{j\beta}^* \rangle = 0. \quad (9)$$

As a consequence, as we shall discuss towards the end of the paper, the matrices $\mathbf{c}$ and $\mathbf{a}$ (5) in the real ensemble have an identical large N behavior as the corresponding matrices (7) in the complex ensemble. Since we are interested here only in the large N behavior it is sufficient to consider one of the two ensembles and draw conclusions for the other. We will focus the presentation on the ensemble of real matrices.

_____

*Email address: burda@th.if.uj.edu.pl
†Email address: jjurkiew@th.if.uj.edu.pl
‡Email address: bwaclaw@th.if.uj.edu.pl

026111-1

An example of a problem which can be formulated in terms of Wishart random matrices (1) is the following. Imagine that we probe a statistical system of N correlated degrees of freedom by doing T measurements. We store the measured values of the ith degree of freedom in the αth measurement in a rectangular matrix $\mathbf{X}=(X_{i\alpha})$. The degrees of freedom as well as the measurements may be correlated. This is expressed by Eq. (4), which tells us that the covariance matrix for the correlations between degrees of freedom in the system is $\mathbf{C}=(C_{ij})$ and for the (auto)correlation between measurements is $\mathbf{A}=(A_{\alpha\beta})$. Note that in general the correlations between $X_{i\alpha}$ and $X_{j\beta}$ may have a more complicated form: $\langle X_{i\alpha}X_{j\beta}\rangle=\mathcal{C}_{i\alpha,j\beta}$, where the matrix $\mathcal{C}$ has double indices. Such a situation takes place if the autocorrelations are different for various degrees of freedom. We will not discuss this case here. Another assumption is that non-Gaussian effects can be neglected. Such effects might come about if there were collective correlations between many degrees of freedom or heavy tails in the distribution of an individual degree of freedom.

A perfect example of the situation described above is the problem of optimal portfolio assessment—one of the fundamental problems of quantitative finance. The portfolio assessment is based on the knowledge of the covariance matrix $\mathbf{C}$ for stocks' returns [12]. In practice, the covariance matrix is estimated from the historical data which are stored in a rectangular matrix representing T historical values of N stocks. Fluctuations of returns are well described by the Gaussian ensemble (1). The estimator of the covariance matrix is given by Eq. (5). Another problem of modern financial analysis which can be directly cast into the form (1) is the problem of taste matching [7]. This problem is encountered for instance in the large-scale internet trading.

It is worth mentioning that the random matrix framework may also be used in a statistical description of data generated in Monte Carlo simulations for a system with many degrees of freedom, in particular of data concerning the correlation functions. One frequently encounters such a problem in Monte Carlo simulations of lattice field theory, where the field is represented by correlated numbers distributed on a lattice. Usually, one is forced to use a dynamical Monte Carlo algorithm to sample such a system. The basic idea standing behind a dynamical algorithm is to generate a Markov chain—a sort of a random walk—in the space of configurations. The degrees of freedom on the lattice as well as the successive configurations are usually correlated. Outside a critical region no long range correlations are observed and the fluctuations can be treated as Gaussian.

Complex random matrices are useful for instance in telecommunication or information theory [3–5].

Let us come back to the ensemble of real matrices (1). As we mentioned the matrices (5) can be treated as estimators of the correlation matrices $\mathbf{C}$ and $\mathbf{A}$. Indeed from Eq. (4) we see that

$$\langle c_{ij}\rangle = \left\langle \frac{1}{T}\sum_{\alpha} X_{i\alpha}X_{j\alpha} \right\rangle = M_{\mathbf{A}1}C_{ij}, \tag{10}$$

$$\langle a_{\alpha\beta}\rangle = \left\langle \frac{1}{N}\sum_{i} X_{i\alpha}X_{i\beta} \right\rangle = M_{\mathbf{C}1}A_{\alpha\beta}, \tag{11}$$

where $M_{\mathbf{C}1}=(1/N)\mathrm{Tr}\,\mathbf{C}$ and $M_{\mathbf{A}1}=(1/T)\mathrm{Tr}\,\mathbf{A}$. This notation will be explained later. The last equation tells us that measuring the average of $\mathbf{c}$ over the ensemble (1) we obtain the matrix $\mathbf{C}$ up to a constant. In other words, having a realization of random matrices $\mathbf{X}$ we can use Eq. (5) to estimate $\mathbf{C}$. Similarly, we can use $\mathbf{a}$ to estimate $\mathbf{A}$. Notice that the measure (1) is invariant under the transformation $\mathbf{C}\rightarrow b\mathbf{C}$ and $\mathbf{A}\rightarrow b^{-1}\mathbf{A}$, where b is an arbitrary positive real number. In particular $\langle c_{ij}\rangle$ and $\langle a_{\alpha\beta}\rangle$ are independent of the rescaling factor b. This independence is ensured by the presence of the factors $M_{\mathbf{A}1}$ and $M_{\mathbf{C}1}$ in Eqs. (10) and (11). In practical calculations, if $\mathrm{Tr}\,\mathbf{A}$ and $\mathrm{Tr}\,\mathbf{C}$ are not specified, one can remove the redundancy with respect to the rescaling by b, setting $(1/T)\mathrm{Tr}\,\mathbf{A}=(1/N)\mathrm{Tr}\,\mathbf{C}$. In this case the constants $M_{\mathbf{A}1}=M_{\mathbf{C}1}$ can be determined from the data by evaluating the traces of $\mathbf{c}$ or $\mathbf{a}$:

$$M_{\mathbf{C}1}=M_{\mathbf{A}1}=\sqrt{\frac{1}{N}\mathrm{Tr}\langle\mathbf{c}\rangle}=\sqrt{\frac{1}{T}\mathrm{Tr}\langle\mathbf{a}\rangle}.$$

While considering the covariance matrices for the Wishart ensemble we can formulate two reciprocal problems, which we shall call *direct* and *inverse problems*. In the direct problem we want to learn as much as possible about the probability distribution of the estimators $\mathbf{c}$ and $\mathbf{a}$ (5) assuming that the matrices $\mathbf{C}$ and $\mathbf{A}$ are given. In particular, we want to calculate the eigenvalue density functions:

$$\rho_{\mathbf{c}}(\lambda)=\left\langle \frac{1}{N}\sum_{i=1}^{N}\delta(\lambda-\lambda_i)\right\rangle,$$

$$\rho_{\mathbf{a}}(\lambda)=\left\langle \frac{1}{T}\sum_{\alpha=1}^{T}\delta(\lambda-\lambda_\alpha)\right\rangle,$$

where λ_i and λ_α are eigenvalues of $\mathbf{c}$ and $\mathbf{a}$, respectively. The determination of the eigenvalue density functions is equivalent to the determination of all their spectral moments:

$$m_{\mathbf{c}k}=\int d\lambda\,\rho_{\mathbf{c}}(\lambda)\lambda^k=\left\langle \frac{1}{N}\mathrm{Tr}\,\mathbf{c}^k\right\rangle,$$

$$m_{\mathbf{a}k}=\int d\lambda\,\rho_{\mathbf{a}}(\lambda)\lambda^k=\left\langle \frac{1}{T}\mathrm{Tr}\,\mathbf{a}^k\right\rangle.$$

The moments $m_{\mathbf{c}k}$, $m_{\mathbf{a}k}$ are related to each other

$$m_{\mathbf{a}k}=r^{1-k}m_{\mathbf{c}k}, \tag{12}$$

where $r=N/T$, as follows from the cyclicity of the trace:

$$\mathrm{Tr}\,\mathbf{X}\mathbf{X}^\tau\cdots\mathbf{X}\mathbf{X}^\tau=\mathrm{Tr}\,\mathbf{X}^\tau\mathbf{X}\cdots\mathbf{X}^\tau\mathbf{X}.$$

In the inverse problem we want to learn as much as possible about the genuine correlations in the system, which are given by $\mathbf{C}$ and $\mathbf{A}$, using a measured sample of random matrices $\mathbf{X}$. We can do this by computing the estimators $\mathbf{c}$ and $\mathbf{a}$ (5) and relating them to matrices $\mathbf{C}$, $\mathbf{A}$. In particular we would like to

estimate the eigenvalue distributions and the moments of **C**, **A**:

$$M_{\mathbf{C}k} = \frac{1}{N}\mathrm{Tr}\,\mathbf{C}^k = \frac{1}{N}\sum_{i=1}^{N} \Lambda_i^k,$$

$$M_{\mathbf{A}k} = \frac{1}{T}\mathrm{Tr}\,\mathbf{A}^k = \frac{1}{T}\sum_{\alpha=1}^{T} \Lambda_\alpha^k,$$

where Λ_i and Λ_α are eigenvalues of **C** and **A**, respectively. The inverse problem is very important for practical applications, since in practice it is very common to reconstruct the properties of the underlying system from the experimental data.

In the analysis of the spectral properties of the matrices **a** and **c** it is convenient to apply the Green's function technique. One can define Green's functions for the correlation matrix **C** and its statistically fluctuating counterpart **c**:

$$\mathbf{G_C}(z) = \frac{1}{z\mathbf{1}_N - \mathbf{C}}, \tag{13}$$

$$\mathbf{g_c}(z) = \left\langle \frac{1}{z\mathbf{1}_N - \mathbf{c}} \right\rangle, \tag{14}$$

and correspondingly $\mathbf{G_A}(z)$ and $\mathbf{g_a}(z)$ for **A** and **a**. The symbol $\mathbf{1}_N$ stands for the $N \times N$ identity matrix. A corresponding symbol $\mathbf{1}_T$ appears in the definition of $\mathbf{G_A}(z)$ and $\mathbf{g_a}(z)$. The Green's functions are related to the generating functions for the moments:

$$M_{\mathbf{C}}(z) = \sum_{k=1}^{\infty} \frac{M_{\mathbf{C}k}}{z^k} = \frac{1}{N}\mathrm{Tr}(z\mathbf{G_C}(z)) - 1, \tag{15}$$

$$m_{\mathbf{c}}(z) = \sum_{k=1}^{\infty} \frac{m_{\mathbf{c}k}}{z^k} = \frac{1}{N}\mathrm{Tr}(z\mathbf{g_c}(z)) - 1, \tag{16}$$

or inversely

$$\frac{1}{N}\mathrm{Tr}\,\mathbf{G_C}(z) = \frac{1 + M_{\mathbf{C}}(z)}{z},$$

$$\frac{1}{N}\mathrm{Tr}\,\mathbf{g_c}(z) = \frac{1 + m_{\mathbf{c}}(z)}{z}. \tag{17}$$

The analogous relations exist for $\mathbf{G_A}(z)$ and $\mathbf{g_a}(z)$. The Green's functions can be used for finding the densities of eigenvalues:

$$\rho_{\mathbf{c}}(\lambda) = -\frac{1}{\pi}\,\mathrm{Im}\,\frac{1}{N}\mathrm{Tr}\,\mathbf{g_c}(\lambda + i0^+) = -\frac{1}{\pi}\,\mathrm{Im}\,\frac{1 + m_{\mathbf{c}}(\lambda + i0^+)}{\lambda + i0^+}, \tag{18}$$

and similarly for $\rho_{\mathbf{a}}(\lambda)$. The eigenvalue densities $\rho_{\mathbf{a}}(\lambda)$ are not independent. As follows from Eq. (12) the corresponding generating functions (16) fulfill the equation

$$m_{\mathbf{a}}(z) = r m_{\mathbf{c}}(rz). \tag{19}$$

Combining the last equation with Eq. (17) we obtain

$$\frac{1}{T}\mathrm{Tr}\,\mathbf{g_a}(z) = r^2\frac{1}{N}\mathrm{Tr}\,\mathbf{g_c}(rz) + \frac{1-r}{z}. \tag{20}$$

Applying now Eq. (18) we have

$$\rho_{\mathbf{a}}(\lambda) = r^2\rho_{\mathbf{c}}(r\lambda) + (1-r)\delta(\lambda). \tag{21}$$

The meaning of the last term on the right-hand side of this equation is that there are $T - N$ zero modes in the matrix **a** if $T > N$. The zero modes disappear when $N = T$. Moving the term containing the delta function to the other side of equation, dividing both sides of the equation by r^2 and substituting the parameter r by $s = T/N = 1/r$ we obtain

$$\rho_{\mathbf{c}}(\lambda) = s^2\rho_{\mathbf{a}}(s\lambda) + (1-s)\delta(\lambda). \tag{22}$$

Therefore, for $T < N$ the zero modes appear in the spectrum $\rho_{\mathbf{c}}(\lambda)$. In this case it is more convenient to use the parameter $s = 1/r$ instead of r. The zero modes appear in the eigenvalue distribution of either **a** or **c**. The two equations (21) and (22) are dual to each other. For $r = s = 1$ they are identical. Because of the duality it is sufficient to solve the problem for $r \leqslant 1$. We will present a solution for the limit $r = N/T = \mathrm{const} \leqslant 1$ and $N \to \infty$ neglecting effects of the order $1/N$, expanding Green's function in $1/N$. The large N limit and the $1/N$ expansion are well defined if the spectra of the matrices **A** and **C** are bounded from above by a constant independent of N. Otherwise the moments' expansion and the generating functions do not exist in this limit.

Using a diagrammatic method [13–15] one can write down a close set of equations for the Green's function $\mathbf{g_c}(z)$ (14):

$$\mathbf{g_c}(z) = \frac{1}{z\mathbf{1}_N - \boldsymbol{\Sigma_c}(z)}, \quad \mathbf{g_{*c}}(z) = \frac{1}{T\mathbf{1}_T - \boldsymbol{\Sigma_{*c}}(z)},$$

$$\boldsymbol{\Sigma_c}(z) = \mathbf{C}\,\mathrm{Tr}(\mathbf{A}\mathbf{g_{*c}}(z)), \quad \boldsymbol{\Sigma_{*c}}(z) = \mathbf{A}\,\mathrm{Tr}(\mathbf{C}\mathbf{g_c}(z)). \tag{23}$$

The set contains four equation for four unknown matrices including $\mathbf{g_c}(z)$ which we want to calculate, and three auxiliary ones: $\mathbf{g_{*c}}(z)$, $\boldsymbol{\Sigma_c}(z)$, and $\boldsymbol{\Sigma_{*c}}(z)$ (see Appendix A). Each of them can be interpreted in terms of a generating function for appropriately weighted diagrams with two external lines: $\mathbf{g_c}(z)$, $\mathbf{g_{*c}}(z)$ for all diagrams and $\boldsymbol{\Sigma_c}(z)$, $\boldsymbol{\Sigma_{*c}}(z)$ for one-line-irreducible diagrams [13–15] (see Appendix A). In the limit $N \to \infty$ the weights of nonplanar diagrams vanish at least as $1/N$. Thus in this limit only planar diagrams give a contribution to the Green's function. Therefore, the large N limit is alternatively called the planar limit. The diagrammatic equations (23) hold only in this limit. An analogous set of equations can be written for the Green's function $\mathbf{g_a}(z)$. The equations are identical to those of Eq. (23) if one exchanges $\mathbf{a} \leftrightarrow \mathbf{c}$, $\mathbf{A} \leftrightarrow \mathbf{C}$ and $T \leftrightarrow N$. The two sets can be solved independently of each other. However, as follows from the duality (20) it is sufficient to solve only one of them and deduce the solution of the other.

Equations (23) can be solved for $\mathbf{g_c}(z)$ by a successive elimination of $\mathbf{g_{*c}}(z)$, $\boldsymbol{\Sigma_c}(z)$, and $\boldsymbol{\Sigma_{*c}}(z)$. However, the resulting equation for $\mathbf{g_c}(z)$ is very entangled [14]:

$$\mathbf{g_c}(z) = \left(z\mathbf{1}_N - \mathbf{C}\,\mathrm{Tr}\frac{\mathbf{A}}{\mathbf{1}_T T - \mathbf{A}\,\mathrm{Tr}(\mathbf{C}\mathbf{g_c}(z))}\right)^{-1}, \qquad (24)$$

and cannot be easily used in practice for analytical calculations of the moments $m_{\mathbf{c}k}$ or spectral density $\rho_{\mathbf{c}}(\lambda)$. If one takes trace on both sides of the equation one does not get a close formula for $\mathrm{Tr}\,\mathbf{g_c}(z)$ since the matrix $\mathbf{g_c}(z)$ appears in the combination $\mathrm{Tr}\,\mathbf{C}\mathbf{g_c}(z)$ on the right-hand side of the equation.

Below we will present an alternative method of solving Eq. (23), along the lines proposed in [15], which gives a close scalar equation for $\mathrm{Tr}\,\mathbf{g_c}(z)$ and thus also equations for the moments and spectral density of the matrix $\mathbf{c}$. This method relies on introducing a new complex variable Z conjugate to z which is defined by the equation

$$m_{\mathbf{c}}(z) = M_{\mathbf{C}}(Z). \qquad (25)$$

At the first glance this equation looks useless because it refers to an unknown function $m_{\mathbf{c}}(z)$ which we actually want to determine. Quite contrary to this, as we shall see, the introduction of the conjugate variable Z allows us to write down a close functional equation for $m_{\mathbf{c}}(z)$. First, let us illustrate how the method works for $\mathbf{A} = \mathbf{1}_T$ [15]. In this case, the elimination of the auxiliary functions (23) leads to

$$Z = \frac{z}{1 + rm_{\mathbf{c}}(z)}, \qquad (26)$$

or equivalently to

$$z = Z[1 + rM_{\mathbf{C}}(Z)]. \qquad (27)$$

Suppose we solve the direct problem. In this case we know the matrix $\mathbf{C}$ and hence also the generating function $M_{\mathbf{C}}(Z)$. Inserting Eq. (26) to Eq. (25) we obtain a close compact functional relation for $m_{\mathbf{c}}(z)$:

$$m_{\mathbf{c}}(z) = M_{\mathbf{C}}\left(\frac{z}{1 + rm_{\mathbf{c}}(z)}\right). \qquad (28)$$

If $M_{\mathbf{C}}(Z)$ has a simple form, one can solve the equation for $m_{\mathbf{c}}(z)$ analytically [15]. All the cases considered there concern the matrix $\mathbf{C}$ having one, two or three distinct eigenvalues. In general, for $\mathbf{C}$ having more than three eigenvalues one can write a numerical program to calculate the eigenvalue density $\rho_{\mathbf{c}}(\lambda)$ from the last equation for any given $M_{\mathbf{C}}(Z)$ and thus for any correlation matric $\mathbf{C}$. This is easier than solving Eq. (24) because we have only one equation with one unknown variable $m_{\mathbf{c}}$. In case of solving the inverse problem, we assume that we can determine moments $m_{\mathbf{c}k}$ from the data and hence that we can approximate the generating function $m_{\mathbf{c}}(z)$. Then we can insert Eq. (27) to Eq. (25) and obtain a functional equation for $M_{\mathbf{C}}(Z)$:

$$M_{\mathbf{C}}(Z) = m_{\mathbf{c}}(Z[1 + rM_{\mathbf{C}}(Z)]). \qquad (29)$$

The problem is solved in principle. However, in practical terms the inverse problem is much more difficult, because one cannot compute all experimental moments $m_{\mathbf{c}k}$ with an arbitrary accuracy, unless one has an infinitely long series of measurements. But one never has. In practice one can estimate only a few lower moments $m_{\mathbf{c}k}$ with a good accuracy.

Because of this practical limitation one cannot entirely solve the inverse problem. However, as we discussed in [10] the inverse problem can be partially solved even in specific practical applications using a moments method. Let us sketch this method below.

We can gain some insight into the spectral properties of the correlation matrix $\mathbf{C}$ by determining the relation between the moments $M_{\mathbf{C}k}$ and $m_{\mathbf{c}k}$. Expanding the functions $M_{\mathbf{C}}(z)$ and $m_{\mathbf{c}}(z)$ in Eq. (28) in $1/z$ using Eqs. (15) and (16) and comparing the coefficients at $1/z^k$ we obtain

$$m_{\mathbf{c}1} = M_{\mathbf{C}1},$$

$$m_{\mathbf{c}2} = M_{\mathbf{C}2} + rM_{\mathbf{C}1}^2,$$

$$m_{\mathbf{c}3} = M_{\mathbf{C}3} + 3rM_{\mathbf{C}1}M_{\mathbf{C}2} + r^2M_{\mathbf{C}1}^3,$$

$$m_{\mathbf{c}4} = M_{\mathbf{C}4} + 2r(M_{\mathbf{C}2}^2 + 2M_{\mathbf{C}1}M_{\mathbf{C}3}) + 6r^2M_{\mathbf{C}1}^2M_{\mathbf{C}2} + r^3M_{\mathbf{C}1}^4,$$

$$\dots \; . \qquad (30)$$

We can also invert the equations for $M_{\mathbf{C}k}$. The result of inversion gives a set of equations which can be directly obtained from the $1/Z$ expansion of the functions in Eq. (29) which is the inverse transform of Eq. (28). We can also determine the corresponding relations for "negative" moments $m_{\mathbf{c}k}$ and $M_{\mathbf{C}k}$, that is for $k < 0$, or determine the spectral density $\rho_{\mathbf{c}}(\lambda)$ [15]. Using a computer tool for symbolic calculations one can easily write a program which successively generates the relations between spectral moments (30) from Eq. (28).

The calculations get more complicated in the general case when both $\mathbf{C}$ and $\mathbf{A}$ are arbitrary. The guiding principle is the same, though. We introduce the conjugate variable Z (25) and, using it, write down the solution of Eqs. (23). In the direct problem we assume that the generating functions $M_{\mathbf{C}}(Z)$ and $M_{\mathbf{A}}(Z)$ are known. We will show that in this case the solution of Eq. (23) takes a form of an explicit equation for $z = z(Z)$, where the function $z(Z)$ depends on the functions $M_{\mathbf{C}}(Z)$ and $M_{\mathbf{A}}(Z)$. Inserting this solution back to Eq. (23) we eventually obtain a functional equation $m_{\mathbf{c}}(z(Z)) = M_{\mathbf{C}}(Z)$ from which we can extract the function $m_{\mathbf{c}}(z)$.

The solution of Eq. (23) takes the form:

$$\frac{z}{Z} = \frac{1}{T}\sum_{\alpha=1}^{T}\frac{\Lambda_{\alpha}}{1 - \Lambda_{\alpha}r\dfrac{Z}{z}M_{\mathbf{C}}(Z)}, \qquad (31)$$

where Λ_{α} are eigenvalues of $\mathbf{A}$. It can be rewritten as

$$rM_{\mathbf{C}}(Z) = M_{\mathbf{A}}\left(\frac{z}{rZM_{\mathbf{C}}(Z)}\right), \qquad (32)$$

and can be formally solved for z:

$$z = ZrM_{\mathbf{C}}(Z)M_{\mathbf{A}}^{-1}(rM_{\mathbf{C}}(Z)), \qquad (33)$$

where $M_{\mathbf{A}}^{-1}$ is the inverse function of $M_{\mathbf{A}}$. Thus we have obtained an explicit equation for $z = z(Z)$ in terms of the known functions $M_{\mathbf{A}}$ and $M_{\mathbf{C}}$. One can easily check that for

$\mathbf{A}=\mathbf{1}_T$ the last equation reduces to Eq. (27). In this case $M_{\mathbf{A}}(z)=1/(z-1)$ and $M_{\mathbf{A}}^{-1}(z)=1+1/z$.

Combining the equation for $z=z(Z)$ given by Eq. (33) with Eq. (25) we arrive at a close equation for the generating function $m_{\mathbf{c}}(Z)$. It can be used for example to calculate the moments $m_{\mathbf{c}k}$'s (see Appendix B). The calculations yield a set of equations expressing $m_{\mathbf{c}k}$'s in terms of the bare moments $M_{\mathbf{A}k}$ and $M_{\mathbf{C}k}$:

$$m_{\mathbf{c}1} = M_{\mathbf{C}1}M_{\mathbf{A}1},$$

$$m_{\mathbf{c}2} = M_{\mathbf{C}2}M_{\mathbf{A}1}^2 + rM_{\mathbf{C}1}^2M_{\mathbf{A}2},$$

$$m_{\mathbf{c}3} = M_{\mathbf{C}3}M_{\mathbf{A}1}^3 + 3rM_{\mathbf{C}1}M_{\mathbf{C}2}M_{\mathbf{A}1}M_{\mathbf{A}2} + r^2M_{\mathbf{C}1}^3M_{\mathbf{A}3},$$

$$m_{\mathbf{c}4} = M_{\mathbf{C}4}M_{\mathbf{A}1}^4 + 2r(M_{\mathbf{C}2}^2 + 2M_{\mathbf{C}1}M_{\mathbf{C}3})M_{\mathbf{A}1}^2M_{\mathbf{A}2}$$
$$+ 2r^2M_{\mathbf{C}1}^2M_{\mathbf{C}2}(M_{\mathbf{A}2}^2 + 2M_{\mathbf{A}1}M_{\mathbf{A}3}) + r^3M_{\mathbf{C}1}^4M_{\mathbf{A}4},$$

$$\dots . \tag{34}$$

The equations reduce to the form (30) for $\mathbf{A}=\mathbf{1}_T$.

Using the relations (12) we can also determine the moments of the matrix $\mathbf{a}$. It is more convenient to write them using the variable $s=r^{-1}$—the dual counterpart of r—instead of r itself:

$$m_{\mathbf{a}1} = M_{\mathbf{A}1}M_{\mathbf{C}1},$$

$$m_{\mathbf{a}2} = M_{\mathbf{A}2}M_{\mathbf{C}1}^2 + sM_{\mathbf{A}1}^2M_{\mathbf{C}2},$$

$$m_{\mathbf{a}3} = M_{\mathbf{A}3}M_{\mathbf{C}1}^3 + 3sM_{\mathbf{A}1}M_{\mathbf{A}2}M_{\mathbf{C}1}M_{\mathbf{C}2} + s^2M_{\mathbf{A}1}^3M_{\mathbf{C}3},$$

$$m_{\mathbf{a}4} = M_{\mathbf{A}4}M_{\mathbf{C}1}^4 + 2s(M_{\mathbf{A}2}^2 + 2M_{\mathbf{A}1}M_{\mathbf{A}3})M_{\mathbf{C}1}^2M_{\mathbf{C}2}$$
$$+ 2s^2M_{\mathbf{A}1}^2M_{\mathbf{A}2}(M_{\mathbf{C}2}^2 + 2M_{\mathbf{C}1}M_{\mathbf{C}3}) + s^3M_{\mathbf{A}1}^4M_{\mathbf{C}4},$$

$$\dots . \tag{35}$$

The equations are completely symmetric to Eq. (34) with respect to the change $r\leftrightarrow s$ (which amounts to $N\leftrightarrow T$), and $\mathbf{c}\leftrightarrow\mathbf{a}$, $\mathbf{C}\leftrightarrow\mathbf{A}$. Using the method one can obtain Eqs. (34) and (35) to an arbitrary order.

The above relations are useful for computing the dressed moments $m_{\mathbf{a}k}, m_{\mathbf{c}k}$ for given matrices $\mathbf{A},\mathbf{C}$ or inversely, the genuine moments $M_{\mathbf{A}k}, M_{\mathbf{C}k}$ from the experimental data. As mentioned, the spectral moments give us in principle full information about the eigenvalue distribution. In practice the reconstruction of the eigenvalue density may be difficult, because to do it we would need to know all moments with a very good precision. Usually, in practical applications one can accurately evaluate only a few lower moments.

In some special cases if we can make some extra assumption about the form of the matrices $\mathbf{C}$ or $\mathbf{A}$ we can improve significantly the reconstruction of the eigenvalue density. In the previous work [15] we have analyzed the case of $\mathbf{A}=\mathbf{1}_T$ and of the matrix $\mathbf{C}$ which had only a few distinct eigenvalues. In this case the Green's function $g_{\mathbf{c}}(z)$ is given by an algebraic equation of the order which is equal to the number of distinct eigenvalues. It can be analytically solved when

this number is less or equal to four. If it is larger the problem can be handled numerically. The duality tells us that the solution also holds when we change the roles of $\mathbf{A}$ and $\mathbf{C}$.

Below we will discuss the case of exponential autocorrelations. Exponential correlations are encountered in many situations. The general solution, which we have discussed so far, simplifies in this case to a more compact relation for the Green's function, which allows us to find analytically an approximate form of the eigenvalue density of the random matrices $\mathbf{c}$ and $\mathbf{a}$. The approximation becomes exact in the large N limit. We consider purely exponential autocorrelations given by the autocorrelation matrix

$$A_{\alpha\beta} = \exp[-|\alpha-\beta|/t], \tag{36}$$

where t controls the range of autocorrelations. The inverse of the matrix $\mathbf{A}$ reads

$$\mathbf{A}^{-1} = \frac{1}{2s_1}\begin{bmatrix} e_1, & -1, & & & \\ -1, & 2c_1, & -1, & & \\ & & \dots & \dots & \dots \\ & & & -1, & 2c_1, & -1 \\ & & & & -1, & e_1 \end{bmatrix}. \tag{37}$$

We have introduced here a shorthand notation $e_1=\exp(1/t)$, $c_1=\cosh(1/t)$ and $s_1=\sinh(1/t)$. The spectrum of this matrix can be approximated by the spectrum of a matrix $\mathbf{M}$:

$$\mathbf{M} = \frac{1}{2s_1}\begin{bmatrix} 2c_1, & -1, & & & -1 \\ -1, & 2c_1, & -1, & & \\ & & \dots & \dots & \dots \\ & & & -1, & 2c_1, & -1 \\ -1, & & & & -1, & 2c_1 \end{bmatrix}, \tag{38}$$

whose eigenvalues can be found analytically:

$$\mu_\alpha = [c_1 + \cos(\pi\alpha/T)]/s_1.$$

The corresponding eigenvectors are given by the Fourier modes. The matrix $2s_1\cdot\mathbf{M}$ can be viewed as of a sum: $(2c_1-2)\mathbf{1}_T+\Delta_T$, of a unity matrix multiplied by a constant and a discretized one-dimensional Laplacian Δ_T for a cyclic chain of length T. The matrix $\mathbf{A}^{-1}$ can be obtained from $\mathbf{M}$ by adding to it a perturbation $\mathbf{P}$: $\mathbf{A}^{-1}=\mathbf{M}+\mathbf{P}$, where $\mathbf{P}$ has only four nonvanishing elements: $P_{11}=P_{TT}=e_1^{-1}$ and $P_{1T}=P_{T1}=1$. The first order corrections to the eigenvalues of $\mathbf{A}^{-1}$, which stem from the perturbation $\mathbf{P}$, behave as $1/T$. The perturbation $\mathbf{P}$ can be viewed as a change of a boundary condition of the Laplacian. As usual, boundary conditions affect mostly the longest (small momentum) modes. Indeed, a careful analysis shows that the two diagonal terms of the perturbation matrix, $P_{11}=P_{TT}$, introduce a constant correction independent of T of the lowest eigenvalues which does not vanish when T goes to infinity. However, since the differences between unperturbed eigenvalues of $\mathbf{M}$ and the corresponding perturbed eigenvalues of $\mathbf{A}^{-1}$ disappear for all other eigenvalues, we expect that for $t\ll T$ the spectral properties of $\mathbf{A}$ can be well approximated by the eigenvalues of $\mathbf{M}^{-1}$:

TABLE I. The moments M_{A2}, M_{A3} and M_{A4} of the matrix **A** (36) for three values of the autocorrelation length $t = 1, 5, 10$ calculated numerically for finite size $T = 20, \ldots, 500$ and by the analytic formula (45) which corresponds to $T = \infty$. The finite size values approach the values given by Eq. (45) as T/t tends to infinity.

	$t=1$				$t=5$				$t=10$		
T	M_{A2}	M_{A3}	M_{A4}	T	M_{A2}	M_{A3}	M_{A4}	T	M_{A2}	M_{A3}	M_{A4}
20	1.29493	2.01479	3.48620	20	4.44996	28.6455	204.107	20	7.58726	79.6134	891.336
50	1.30579	2.05757	3.60838	50	4.81980	34.2544	271.932	50	9.03668	120.509	1784.56
100	1.30941	2.07183	3.64911	100	4.94314	36.1292	294.733	100	9.53497	135.501	2146.93
200	1.31123	2.07896	3.66947	200	5.00482	37.0666	306.133	200	9.78414	143.001	2328.48
500	1.31231	2.08324	3.68169	500	5.04182	37.6290	312.973	500	9.93364	147.501	2437.40
∞	1.31304	2.08609	3.68983	∞	5.06649	38.0040	317.534	∞	10.0333	150.501	2510.02

$$\Lambda_\alpha \approx \frac{1}{\mu_\alpha} = \frac{s_1}{c_1 + \cos(\pi\alpha/T)}. \qquad (39)$$

In this limit we can also approximate the sum (31) by an integral:

$$\frac{z}{Z} \approx \frac{s_1}{\pi} \int_0^\pi \frac{d\tau}{(c_1 - s_1 F) + \cos \tau}, \qquad (40)$$

where $\tau = \pi\alpha/T$ and the symbol $F \equiv (Z/z) r m_c(z)$ is introduced for brevity. Note that in the definition of F we have replaced $M_C(Z)$, which would rather be dictated by Eq. (31), by $m_c(z)$. This change is legitimate due to Eq. (25). The integral (40) can be done:

$$\frac{z}{Z} = \frac{s_1}{\sqrt{(c_1 - s_1 F)^2 - 1}}. \qquad (41)$$

Setting back $F = (Z/z) r m_c(z)$ we eventually obtain

$$Z = z \frac{c_1 r m_c(z) - \sqrt{s_1^2 + r^2 m_c^2(z)}}{s_1 [r^2 m_c^2(z) - 1]}. \qquad (42)$$

This is an explicit equation for $Z = Z(z)$ which can be now inserted to $m_c(z) = M_C(Z)$ giving us a compact relation for $m_c(z)$ in the presence of the exponential autocorrelations (36):

$$m_c(z) = M_C\left(z \frac{c_1 r m_c(z) - \sqrt{s_1^2 + r^2 m_c^2(z)}}{s_1 [r^2 m_c^2(z) - 1]}\right). \qquad (43)$$

In the limit $t \to 0$, the parameters $s_1 = \sinh(1/t)$, $c_1 = \cosh(1/t)$ and $e_1 = \exp(1/t)$ increase to infinity and $c_1/e_1 \approx s_1/e_1 \approx 1$. As a consequence, the form of Eq. (42) simplifies to Eq. (26), which corresponds to the case without autocorrelations, as expected. Using Eq. (43) we can recursively generate equations for the consecutive moments:

$$m_{c1} = M_{C1},$$

$$m_{c2} = M_{C2} + r M_{C1}^2 \cdot t_1,$$

$$m_{c3} = M_{C3} + 3r M_{C1} M_{C2} \cdot t_1 + r^2 M_{C1}^3 \left(\frac{3}{2} t_1^2 - \frac{1}{2}\right),$$

$$m_{c4} = M_{C4} + 2r(M_{C2}^2 + 2M_{C1} M_{C3}) t_1 + 2r^2 M_{C1}^2 M_{C2} (4t_1^2 - 1)$$
$$+ r^3 M_{C1}^4 \left(\frac{5}{2} t_1^3 - \frac{3}{2} t_1\right),$$

$$\ldots, \qquad (44)$$

where $t_1 = c_1/s_1 = \coth(1/t)$. The coefficients on the right-hand side, which depend on t_1, can be directly expressed in terms of the moments M_{Ak} of the matrix **A**. Approximating again a sum by an integral in the large T limit we can write

$$M_{Ak} = \frac{1}{T} \sum_{\alpha=1}^T \Lambda_\alpha^k \approx \frac{s_1^k}{\pi} \int_0^\pi \frac{d\tau}{(c_1 + \cos \tau)^k}.$$

The integrals can be calculated yielding

$$M_{A1} = 1,$$

$$M_{A2} = t_1,$$

$$M_{A3} = \frac{3}{2} t_1^2 - \frac{1}{2},$$

$$M_{A4} = \frac{5}{2} t_1^3 - \frac{3}{2} t_1,$$

$$\ldots. \qquad (45)$$

We see that if we insert these coefficients into Eqs. (34) we obtain Eq. (44). This is a consistency check for the approximation which we use here. The quality of this approximation can also be checked by comparing the moments M_{Ak} of the matrix (36) for finite T with the result (45) which corresponds to $T = \infty$. We expect that for $t \ll T$ the numerical values shall approach the result (45). The results of this comparison
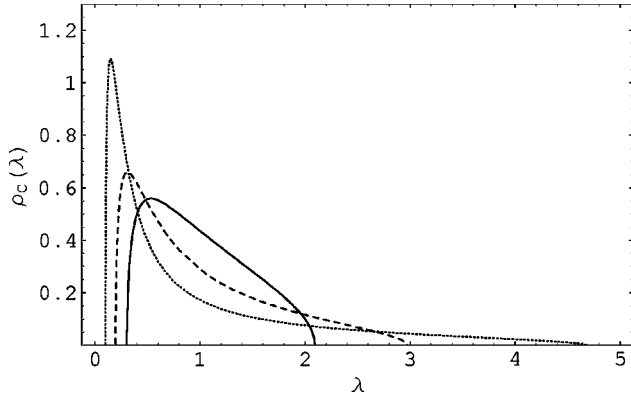
FIG. 1. The density of eigenvalues $\rho_{\mathbf{c}}(\lambda)$ for exponential matrix $\mathbf{A}$, $\mathbf{C}=\mathbf{1}_N$, $r=0.2$ and for three different autocorrelation times t: $t=0$ (solid line), $t=2$ (dashed line), and $t=5$ (dotted line).

confirm our expectations (see Table I). Thus we see that the formula (44) for $m_{\mathbf{c}}(z)$ in the presence of the exponential autocorrelations becomes exact in the limit $T=\infty$. This formula allows us to compute the eigenvalue distribution of the random matrices $\mathbf{c}$ and $\mathbf{a}$ (5). Let us illustrate this on the simplest example of the system which has no correlations: that means that $\mathbf{C}=\mathbf{1}_N$ and $M_{\mathbf{C}}(Z)=1/(Z-1)$. In this case Eq. (43) for $m_{\mathbf{c}}(z)$ takes the form

$$t_1 x^2 - \frac{1}{z}(x^2-1)(x+r) - x\sqrt{1+x^2/s_1^2} = 0, \qquad (46)$$

where we used the notation $x(z)=rm_{\mathbf{c}}(z)$. For $t\to 0$ ($\mathbf{A}\to\mathbf{1}_T$) this equation reduces to

$$(x-1)\left[x - \frac{1}{z}(x+1)(x+r)\right] = 0, \qquad (47)$$

which has a solution

$$x(z) = \frac{1}{2}(-1-r-i\sqrt{(\mu_+ - z)(z-\mu_-)} + z), \qquad (48)$$

where $\mu_\pm = (1\pm\sqrt{r})^2$, which leads to the well-known result for the uncorrelated Wishart ensemble:

$$\rho_{\mathbf{c}}(\lambda) = \frac{1}{2\pi r}\frac{\sqrt{(\mu_+ - \lambda)(\lambda - \mu_-)}}{\lambda}. \qquad (49)$$

For $t>0$ it is still possible to find analytically a solution of Eq. (46). Let us rewrite Eq. (46) as a polynomial equation:

$$-z^2 x^2(1+x^2/s_1^2) + (t_1 z x^2 - (r+x)(x^2-1))^2 = 0.$$

It has two trivial solutions $x=\pm 1$. Dividing out the polynomial $(x-1)(x+1)$ we get

$$x^4 + 2x^3(r-t_1 z) + x^2(-1+r^2-2t_1 rz+z^2) - 2rx - r^2 = 0. \qquad (50)$$

This is a quartic equation which can be solved analytically by the Ferrari method. We will not present the formal solution which is neither transparent nor informative. Instead, we show in Fig. 1 the eigenvalue density functions $\rho_{\mathbf{c}}(\lambda)$, for different t, resulting from this solution. The lower part of the

distribution approaches zero when t increases, but zero modes do not appear in the distribution as long as $r<1$. The formula (43) applies to any correlation matrix $\mathbf{C}$ but in the general case one has to use a numerical procedure to calculate from it the density function.

Let us stop here the presentation of results for the ensemble of real matrices. As we mentioned all results in the large N limit hold also in the ensemble of complex matrices if the covariance matrices (5) are replaced by Eq. (7). The reason why it is so is related to the fact that the moments of $\mathbf{c}=(1/T)\mathbf{X}\mathbf{X}^\dagger$ in the ensemble of complex matrices (6) are equal to the moments of $\mathbf{c}=(1/T)\mathbf{X}\mathbf{X}^\tau$ in the real ensemble (1) up to a $1/N$ corrections which disappear in the large N limit:

$$\frac{1}{N}\left\langle\left(\frac{1}{T}\mathbf{X}\mathbf{X}^\dagger\right)^k\right\rangle_{\text{complex}} = \frac{1}{N}\left\langle\left(\frac{1}{T}\mathbf{X}\mathbf{X}^\tau\right)^k\right\rangle_{\text{real}} + O(1/N). \qquad (51)$$

Let us illustrate this by explicit calculations of the second moment. Using the Wick's theorem for Gaussian integrals and Eq. (4) for the two-point correlation function, we have

$$\frac{1}{N}\left\langle\left(\frac{1}{T}\mathbf{X}\mathbf{X}^\tau\right)^2\right\rangle = \frac{1}{NT^2}\langle X_{i\alpha}X_{j\alpha}X_{j\beta}X_{i\beta}\rangle$$

$$= \frac{1}{NT^2}\{\langle X_{i\alpha}X_{j\alpha}\rangle\langle X_{j\beta}X_{i\beta}\rangle + \langle X_{i\alpha}X_{i\beta}\rangle\langle X_{j\alpha}X_{j\beta}\rangle$$

$$+ \langle X_{i\alpha}X_{j\beta}\rangle\langle X_{j\alpha}X_{i\beta}\rangle\}$$

$$= \frac{1}{NT^2}\{C_{ij}A_{\alpha\alpha}C_{ji}A_{\beta\beta} + C_{ii}A_{\alpha\beta}C_{jj}A_{\alpha\beta}$$

$$+ C_{ij}A_{\alpha\beta}C_{ji}A_{\alpha\beta}\}$$

$$= M_{\mathbf{C}2}M_{\mathbf{A}1}^2 + rM_{\mathbf{C}1}^2 M_{\mathbf{A}2} + \frac{r}{N}M_{\mathbf{C}2}M_{\mathbf{A}2}.$$

The corresponding calculations for the complex ensemble read

$$\frac{1}{N}\left\langle\left(\frac{1}{T}\mathbf{X}\mathbf{X}^\dagger\right)^2\right\rangle = \frac{1}{NT^2}\langle X_{i\alpha}X_{j\alpha}^* X_{j\beta}X_{i\beta}^*\rangle$$

$$= \frac{1}{NT^2}\{\langle X_{i\alpha}X_{j\alpha}^*\rangle\langle X_{j\beta}X_{i\beta}^*\rangle + \langle X_{i\alpha}X_{i\beta}^*\rangle\langle X_{j\alpha}X_{j\beta}^*\rangle$$

$$+ \langle X_{i\alpha}X_{j\beta}\rangle\langle X_{j\alpha}^* X_{i\beta}^*\rangle\}$$

$$= \frac{1}{NT^2}\{C_{ij}A_{\alpha\alpha}C_{ji}A_{\beta\beta} + C_{ii}A_{\alpha\beta}C_{jj}A_{\alpha\beta}\}$$

$$= M_{\mathbf{C}2}M_{\mathbf{A}1}^2 + rM_{\mathbf{C}1}^2 M_{\mathbf{A}2}.$$

The difference between the two calculations appears in the third term which in the real ensemble gives a contribution of the order $1/N$ while in the complex ensemble disappears by virtue of Eq. (9). We recognize that $M_{\mathbf{C}2}M_{\mathbf{A}1}^2 + rM_{\mathbf{C}1}^2 M_{\mathbf{A}2}$ which are the leading terms in the $1/N$ expansion are identical as in the second equation in the set (34). Generally one can show that the leading contributions which correspond to the planar diagrams in the expansion of $\mathbf{g}_{\mathbf{c}}(z)$ are identical for both ensembles. Nonplanar diagrams are different but they

contribute in the subleading orders: it turns out that in the diagrammatic expansion of the Green's function $g_c(z)$ for the complex matrix ensemble (6), which would be a counterpart of Eq. (A1) in Appendix A, all diagrams which contain a double arc whose dashed and solid lines cross each other are identically equal zero, since such an arc corresponds to the propagator $\langle X_{i\alpha}X_{j\beta}\rangle$ or $\langle X_{i\alpha}^*X_{j\beta}^*\rangle$. A crossing of two arcs is however allowed and leads to a factor $1/N^2$.

To summarize, in the paper we considered a Wishart ensemble of correlated random matrices. We obtained in the limit of large matrices a close set of equations relating the Green's function or equivalently the moments' generating functions $m_{\mathbf{c}}(z)$ and $m_{\mathbf{a}}(z)$ for statistically dressed correlations to the generating functions for genuine correlation matrices $M_{\mathbf{C}}(z)$ and $M_{\mathbf{A}}(z)$. The equations in the large N limit are the same for the ensemble of real and complex matrices. Using these equations we can write down exact relations between genuine and experimental spectral moments of correlation functions of an arbitrary order. The relations can be used in practical problems to learn about correlations in the studied system from the experimental samples. In the case of exponential correlations we have also found an explicit form the spectral density function of the covariance matrix. A natural generalization of the work presented here is to consider a more general type of time correlations than purely

exponential (36). If the correlations are of the form which depends on the time difference $A_{\alpha,\beta}=A(|\alpha-\beta|)$ then one can apply Fourier transform to determine in the large T limit an approximate spectrum of the matrix $\mathbf{A}$ and approximate values of its spectral moments. Some details are given in Appendix C.

Another interesting issue which can be addressed in the future is the determination of the probability distribution for individual elements of the covariance matrices $\mathbf{c}$ and $\mathbf{a}$ similarly as it was done for the uncorrelated case [16].

APPENDIX A

For completeness we recall here the graphical representation of the Green's function. The details of the diagrammatic method can be found in [13–15]. The Green's function $\mathbf{g_c}(z)$ can be represented as a sum over diagrams:

$$\tag{A1}$$

where the N-type and T-type indices of $\mathbf{X}$ are denoted by filled and empty circles, respectively. The matrix $\mathbf{X}$ is denoted by an ordered pair of neighboring filled and empty circles, while $\mathbf{X}^\tau$ is drawn as an pair of such circles in the reverse order. A horizontal solid line stands for $\mathbf{1}_N/z$, a dashed line for $\mathbf{1}_T/T$, a solid arc for $\mathbf{C}$ and a dashed arc for $\mathbf{A}$. The two point function (4) is drawn as a double arc. Matrices on a line are multiplied in the order of appearance on this line. If a line is closed, the trace is taken.

In the thermodynamical limit only planar diagrams give contribution to $\mathbf{g_c}$. In particular the last term in Eq. (A1) vanishes. The Green's function $\mathbf{g_{c*}}(z)$ is represented by an identical set of diagrams with dashed and solid lines exchanged. It is convenient to introduce one-line irreducible diagrams and corresponding generating functions Σ_c and Σ_{*c}. The Green's functions can be expressed in terms of Σ_c and Σ_{*c} as follows:

$$\tag{A2}$$

$$\tag{A3}$$

In the planar limit there are two additional equations which relate the sums over one-line irreducible diagrams to the Green's functions:

$$\tag{A4}$$

Analogous diagrammatic equations can be written For $\mathbf{g_a}$ with the only difference that the solid line shall denote the propagator $\mathbf{1}_T/z$ and the dashed shall line denote $\mathbf{1}_N/N$.

APPENDIX B

In this appendix we use Eqs. (33) and (25) to determine the relations (34). As for the case $\mathbf{A}=\mathbf{1}_T$ we shall do this using $1/z$ expansion. The function $M_\mathbf{A}(Z)$ is given by the series

$$M_\mathbf{A}(Z) = \frac{M_{\mathbf{A}1}}{Z} + \frac{M_{\mathbf{A}2}}{Z^2} + \frac{M_{\mathbf{A}3}}{Z^3} + \cdots. \qquad (B1)$$

Let us determine the expansion for the inverse function $M_\mathbf{A}^{-1}$ as a series around zero:

$$M_\mathbf{A}^{-1}(y) = M_{\mathbf{A}1}y^{-1}(1 + \mu_1 y + \mu_2 y^2 + \cdots). \qquad (B2)$$

The coefficients of the series can be directly calculated from the condition

$$y = M_\mathbf{A}(M_\mathbf{A}^{-1}(y)), \qquad (B3)$$

which gives us

$$\mu_1 = \frac{M_{\mathbf{A}2}}{(M_{\mathbf{A}1})^2}, \quad \mu_2 = \frac{M_{\mathbf{A}3}M_{\mathbf{A}1} - (M_{\mathbf{A}2})^2}{(M_{\mathbf{A}1})^4}, \quad \ldots. \quad (B4)$$

Equation (33) takes the form

$$z = M_{\mathbf{A}1}Z[1 + \mu_1 r M_\mathbf{C}(Z) + \mu_2 r^2 M_\mathbf{C}^2(Z) + \cdots]. \qquad (B5)$$

or, if written for $1/z$,

$$\frac{1}{z} = \frac{1}{M_{\mathbf{A}1}} \frac{1}{Z}[1 - \mu_1 r M_\mathbf{C}(Z) + (\mu_1^2 - \mu_2)r^2 M_\mathbf{C}^2(Z) + \cdots], \qquad (B6)$$

where

$$M_\mathbf{C}(Z) = \frac{M_{\mathbf{C}1}}{Z} + \frac{M_{\mathbf{C}2}}{Z^2} + \frac{M_{\mathbf{C}3}}{Z^3} + \cdots. \qquad (B7)$$

Thus we have expressed $1/z$ as a series of $1/Z$. Inserting this series to Eq. (25) and comparing coefficients at $1/Z^k$ we eventually obtain Eq. (34).

APPENDIX C

Let $\mathbf{A}$ be a $T \times T$ symmetric square matrix such that $A_{\alpha\beta} = A_{\beta\alpha} = A(|\alpha - \beta|)$. Assume also that $A(0) > A(1) > A(2) > \cdots > A(T-1) > 0$. One can show that under this assumption the matrix $\mathbf{A}$ is positive definite. The eigenproblem for this matrix can be written in the form:

$$\sum_{s=-T/2+1}^{T/2} A(t - s)x_\alpha(s) = \Lambda_\alpha x_\alpha(t). \qquad (C1)$$

In the equation above it is convenient to choose the set of T values, which the index s may assume, to be $-T/2+1$, $-T/2+2, \ldots, T/2$. This choice is appropriate if T is even, while if T is odd one should rather choose $-(T-1)/2$,

$-(T-1)/2+1, \ldots, (T-1)/2$. The same holds for the index α which enumerates the eigenvalues. We are interested in the spectrum of the matrix $\mathbf{A}$ in the limit of $T \to \infty$. In this limit the eigenvectors can be approximated by $e^{i\omega t}$:

$$\sum_{s=-\infty}^{\infty} A(t - s)e^{i\omega s} = \sum_{s=-\infty}^{\infty} A(t - s)e^{-i\omega(t-s)}e^{i\omega t}$$

$$= \sum_{k=-\infty}^{\infty} A(k)e^{-i\omega k}e^{i\omega t}$$

$$= \Lambda(\omega)e^{i\omega t}, \qquad (C2)$$

where

$$\Lambda(\omega) = \sum_{k=-\infty}^{\infty} A(k)e^{-i\omega k}. \qquad (C3)$$

For finite T the range of summation over the index $k=t-s$ depends on t: $t-T/2 \leq k < t+T/2$. Only in the limit $T \to \infty$ it becomes independent of t: $k \in (-\infty, \infty)$. The limit $T \to \infty$ can be safely taken only if $A(t)$ vanishes fast enough for $t \to \pm\infty$. For example, if $A(t) \sim 1/|t|^\alpha$ for $\alpha \leq 1$ the sum on the right-hand side in Eq. (C3) is divergent for $\omega=0$. In other words, in this case one cannot blindly take the large T limit since the results explicitly depend on T. From here on we assume that $A(t)$ falls off fast enough when t goes to infinity.

The frequency ω: $-\pi < \omega \leq \pi$ from Eq. (C3) is related to the index α from Eq. (C1) as $\omega=2\pi\alpha/T$. For $T \to \infty$ the frequency ω becomes a continuous variable. The inverse transform to Eq. (C3) reads

$$A(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\omega \, \Lambda(\omega)e^{i\omega k}. \qquad (C4)$$

Actually Eq. (C3) holds only for infinite T. In this case the eigenvalues are degenerate $\Lambda(\omega)=\Lambda(-\omega)$ as follows from inserting $A(k)=A(-k)$ to Eq. (C3). For finite T the boundary conditions affect the spectrum. The diagonal striplike structure of the matrix $A_{\alpha\beta}=A(|\alpha - \beta|)$ is broken at the corners of the matrix since for example the upper leftmost element of the matrix has no left neighbor. It turns out that the breaking of this structure is sufficient to lift the degeneration of eigenvalues. In this case the index ω is better mapped into the range $0 < \omega=\pi\alpha/T \leq \pi$. These effects have to be studied carefully if one wants to consider corrections to the large T formula. We will discuss here only the leading order approximation given by Eq. (C3). In this approximation Eq. (31) takes the form

$$\frac{z}{Z} = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\omega \frac{\Lambda(\omega)}{1 - \Lambda(\omega)F} = \frac{1}{\pi} \int_0^{\pi} d\omega \frac{\Lambda(\omega)}{1 - \Lambda(\omega)F}. \qquad (C5)$$

We recall that $F=(Z/z)rm_\mathbf{c}(z)$. In the second step we used $\Lambda(\omega)=\Lambda(-\omega)$. Thus the problem of determining the map $Z=Z(z)$ is reduced to calculating the Fourier transform $\Lambda(\omega)$ of $A(k)$ (C3). Similarly the kth moment of the spectrum of the matrix $\mathbf{A}$ is approximated by

$$M_{\mathbf{A}k} = \frac{1}{\pi} \int_0^\pi \Lambda^k(\omega) d\omega. \tag{C6}$$

We will now present some examples of applying this technique. Let us begin with multiexponential correlations: $A(t) = \Sigma_i a_i \exp(-|t|/t_i)$. The Fourier transform (C3) can be exactly calculated:

$$\Lambda(\omega) = \sum_i a_i \frac{s_i}{c_i - \cos(\omega)}, \tag{C7}$$

where $c_i = \cosh(1/t_i)$ and $s_i = \sinh(1/t_i)$. Using Eq. (C5) we obtain

$$\frac{z}{Z} = \sum_i a_i \frac{s_i}{\sqrt{(c_i - s_i F)^2 - 1}}. \tag{C8}$$

This algebraic equation can be solved analytically for $Z = Z(z)$ for purely exponential correlations $A(t) = \exp(-|t|/t_c)$ which lead to a quadratic equation (42) as discussed in the main body of the text or for double exponential correlations $A(t) = a_1 \exp(-|t|/t_1) + a_2 \exp(-|t|/t_2)$ which leads to a Ferrari equation for Z.

Let us now discuss correlations with a Gaussian autocorrelation function: $A(t) = \exp(-t^2/(2t_c^2))$. In this case one cannot analytically determine the sum (C3) but one can approximate it by an integral:

$$\Lambda(\omega) \approx \int_{-\infty}^\infty dk \, A(k) e^{-i\omega k}, \tag{C9}$$

which can be done analytically as a Gaussian integral, yielding

$$\Lambda(\omega) = \sqrt{2\pi} t_c e^{-\omega^2 t_c^2/2}. \tag{C10}$$

If we replace the sum (C3) by the integral (C9) in the Fourier transform then k becomes a continuous variable and we have to correspondingly replace the range of integration over ω in the inverse transform. Instead of Eq. (C4) we have

$$A(k) = \frac{1}{2\pi} \int_{-\infty}^\infty d\omega \, \Lambda(\omega) e^{i\omega k}, \tag{C11}$$

and correspondingly

$$M_{\mathbf{A}k} = \frac{1}{\pi} \int_0^\infty \Lambda^k(\omega) d\omega = \frac{(2\pi)^{(k-1)/2}}{\sqrt{k}} t_c^{k-1}. \tag{C12}$$

This result is in a very good agreement with the numerical computations for large T and sufficiently small t_c. If we

change the integration limits in the integral (C5) we obtain an equation

$$\frac{z}{Z} = \frac{1}{\pi} \int_0^\infty \frac{d\omega}{(\sqrt{2\pi} t_c)^{-1} e^{\omega^2 t_c^2/2} - F}, \tag{C13}$$

which yields

$$\frac{z}{Z} = \frac{1}{\sqrt{2\pi} t_c F} \sum_{n=1}^\infty \frac{1}{n^{1/2}} (\sqrt{2\pi} t_c F)^n = \frac{1}{\sqrt{2\pi} t_c F} \mathrm{Li}_{1/2}(\sqrt{2\pi} t_c F), \tag{C14}$$

where $\mathrm{Li}_\alpha(x)$ is polylogarithm (or actually its generalization to fractional-index α):

$$\mathrm{Li}_\alpha(x) = \sum_{n=1}^\infty \frac{x^n}{n^\alpha}. \tag{C15}$$

Inserting $F = (Z/z) r m_c(z)$ we can solve this equation for Z by inverting the function $\mathrm{Li}_{1/2}$:

$$Z = \frac{1}{\sqrt{2\pi} t_c r m_{\mathbf{c}}(z)} \mathrm{Li}_{1/2}^{-1}(\sqrt{2\pi} t_c r m_{\mathbf{c}}(z)). \tag{C16}$$

The coefficients a_n of series corresponding to the inverse function $\mathrm{Li}_{1/2}^{-1}(y) = \Sigma_{n=1}^\infty a_n y^n$ can be found recursively by inverting the series (C15) to an arbitrary order.

Similarly one can treat the correlations of the type $A(t) = 1/(|t/t_c|^2 + 1)$. One obtains

$$\Lambda(\omega) = \pi t_c e^{-\omega t_c}, \quad M_{\mathbf{A}k} = \frac{(\pi t_c)^{k-1}}{k}, \tag{C17}$$

which agree with numerical results. The equation for $Z = Z(z)$ reads

$$\frac{z}{Z} = \frac{1}{\pi} \int_0^\infty \frac{d\omega}{(t_c \pi)^{-1} e^{\omega t_c} - F} = \frac{1}{\pi F} \int_0^\infty d\omega \sum_{n=1}^\infty (\pi t_c F e^{-\omega t_c})^n$$

$$= \frac{1}{\pi F t_c} \sum_{n=1}^\infty \frac{1}{n} (\pi t_c F)^n = \frac{1}{\pi F t_c} \mathrm{Li}_1(\pi t_c F)$$

$$= -\frac{1}{\pi F t_c} \ln(1 - \pi t_c F). \tag{C18}$$

It can be solved for Z:

$$Z = \frac{z}{\pi t_c r m_{\mathbf{c}}(z)} [1 - e^{-\pi t_c r m_{\mathbf{c}}(z)}]. \tag{C19}$$

[1] J. Wishart, Biometrika **A20**, 32 (1928).
[2] T. W. Anderson, *Introduction to Multivariate Statistical Analysis* (Wiley, New York, 1958).
[3] A. L. Moustakas *et al.*, Science **287**, 287 (2000).
[4] A. M. Sengupta and P. P. Mitra, physics/0010081.
[5] S. E. Skipetrov, Phys. Rev. E **67**, 036621 (2003).
[6] Y. Demasure and R. A. Janik, Phys. Lett. B **553**, 105 (2003).
[7] S. Maslov and Y. C. Zhang, Phys. Rev. Lett. **87**, 248701 (2001).
[8] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).
[9] V. Plerou *et al.*, Phys. Rev. Lett. **83**, 1471 (1999).

[10] Z. Burda and J. Jurkiewicz, Physica A **344**, 67 (2004).

[11] S. H. Simon and A. L. Moustakas, math-ph/0401038.

[12] Because of the multiplicative nature of the price changes instead of the changes $X_{it} = P_{it+1} - P_{it}$ themselves, in the financial analysis one rather uses returns $X_{it} = \log P_{it+1}/P_{it}$ as random variables.

[13] J. Feinberg and A. Zee, J. Stat. Phys. **87**, 473 (1997).

[14] A. M. Sengupta and P. P. Mitra, Phys. Rev. E **60**, 3389 (1999).

[15] Z. Burda, A. Goerlich, A. Jarosz, and J. Jurkiewicz, Physica A **343**, 295 (2004).

[16] R. A. Janik and M. A. Nowak, J. Phys. A **36**, 3629 (2003).