Spectra of empirical auto-covariance matrices

R. Kühn and P. Sollich
Department of Mathematics, King's College London, Strand, London WC2R 2LS, U.K.

(Dated: June 20, 2012)

We compute spectral densities of large sample auto-covariance matrices of stationary stochastic processes at fixed ratio $\alpha = N/M$ of matrix dimension N and sample size M. We find a remarkable scaling relation which expresses the spectral density $\rho_\alpha(\lambda)$ of sample auto-covariance matrices for processes *with* correlations as a continuous superposition of copies of the spectral density $\rho_\alpha^{(0)}(\lambda)$ for a sequence of *uncorrelated* random variables at the same value of α, rescaled in terms of the Fourier transform $\hat{C}(q)$ of the true auto-covariance function. We also obtain a closed-form approximation for the scaling function $\rho_\alpha^{(0)}(\lambda)$. Our results are in excellent agreement with numerical simulations using auto-regressive processes, and processes exhibiting a power-law decay of correlations.

PACS numbers: 02.50.-r,05.10.-a

Introduction When analyzing the properties of stationary stochastic processes [1], one typically begins by concentrating on low-order statistics, in particular the mean and the auto-covariance as a function of time-lag. One way to characterize the latter further consists in looking at the auto-covariance matrix obtained by sampling the auto-covariance function on a grid of (equidistant) time-lags, and in particular at its spectrum. For second order stationary processes the auto-covariance function depends only on time-lags, so that the auto-covariance matrix is on average Toeplitz. However, due to the finiteness of the samples used in practice to estimate the auto-covariance function, empirically determined auto-covariance matrices will exhibit random fluctuations about their average Toeplitz form.

A key issue then is to judge how the spectral properties of empirical auto-covariance matrices will be affected by these random fluctuations. Clearly, having a theory that would quantify such effects analytically could be useful for the empirical analysis of stochastic processes. However, such theoretical understanding is at present almost entirely lacking – in marked contrast to the situation for the closely related problem of sample covariance matrices of a multi-dimensional data set estimated from finitely many independent measurements.

From an abstract point of view, both problem classes belong to random matrix theory [2, 3]. In the case of sample covariance matrices, the random matrix ensemble in question is the well known Wishart-Laguerre ensemble [4], which has been widely studied for several decades, and for which numerous results are available. The spectral problem for example was solved in the 60s by Marčenko and Pastur [5]; typical fluctuations of the largest eigenvalue of Wishart matrices were shown [6] to follow a Tracy-Widom distribution [7], and large deviation properties of both the largest [8] and smallest [9] eigenvalue have recently been characterized. Numerous variants of the original Wishart-Laguerre ensemble have been studied in the literature over the years (e.g. [10–13]), and applications have been formulated in a variety of fields, including multivariate statistics [14], wireless communication [15] and the analysis of cross-correlations in financial data [16, 17]. For a more complete recent overview, we refer to [3].

Due to the temporal structure of the underlying signals in the problem of sample auto-covariance matrices of time series, the ensemble of random matrices describing this problem is radically different from the Wishart-Laguerre ensemble, and *much* less is known about their spectral properties. The *existence* of the limiting spectral density of sample auto-covariance matrices of moving-average processes [1] with i.i.d. driving (of both finite and infinite order) has in fact been established only very recently [18]; the corresponding existence proof for the closely related problem of random Toeplitz matrices with i.i.d. entries is also only a few years old [19]. We are not aware of closed form expressions for limiting spectral densities for these cases – whether exact, or approximate but of a quality that would allow meaningful use for e.g. time series analysis. The purpose of the present letter is to report recent progress that fills this gap.

Ensemble definition and spectral density We consider stationary zero-mean processes $(x_n)_{n\in\mathbb{Z}}$. These could be discrete-time processes to begin with, or sampled from continuous-time processes at discrete equidistant time steps $\Delta\tau$, in which case $x_n \equiv x(n\Delta\tau)$. We are interested in the spectrum of $N \times N$ empirical auto-covariance matrices C, which are estimated by measurements on sequences of M samples. There are several (non-equivalent) ways to define the elements of C. Our choice is

$$C_{k\ell} = \frac{1}{M}\sum_{m=0}^{M-1} x_{m+k}x_{m+\ell}\ , \quad 1 \le k,\ell \le N\ . \qquad (1)$$

Sample auto-covariance matrices C of this form constitute randomly perturbed Toeplitz matrices [20]. They are not Toeplitz themselves, but their averages are, and fluctuations about these averages decrease with increasing sample size M. We note that our choice differs from the ones looked at in [18], where sample auto-covariance matrices were constructed as random Toeplitz matrices from the start.

Our main results are the following. In the large N limit, spectral properties depend on the 'aspect ratio' $\alpha = N/M$, i.e. on the ratio of matrix dimension N and the size M of the samples used to define empirical averages, which we take to be fixed. We find a remarkable scaling relation which expresses the spectral density $\rho_\alpha(\lambda)$ of sample auto-covariance matrices for processes *with* dynamical correlations as a continuous superposition of rescaled copies of the spectral density $\rho_\alpha^{(0)}(\lambda)$ for a sequence of *uncorrelated*, i.i.d. random variables. We also obtain a simple closed form expression for $\rho_\alpha^{(0)}$ that provides an excellent approximation to numerically simulated spectra.

The spectral density of a matrix C is evaluated in terms of its resolvent as

$$\rho_N(\lambda;C) = \frac{1}{\pi N}\mathrm{Im}\ \mathrm{Tr}\ \left[\lambda_\varepsilon \mathbb{I} - C\right]^{-1} . \tag{2}$$

Here $\mathbb{I}$ is the $N\times N$ unit matrix and $\lambda_\varepsilon = \lambda - i\varepsilon$, the limit $\varepsilon \to 0^+$ being understood. We follow Edwards and Jones [21] and express the trace of the resolvent, averaged over the matrix ensemble, in terms of a Gaussian integral as

$$\rho_N(\lambda) = -\frac{2}{\pi}\lim_{\varepsilon\to 0}\mathrm{Im}\,\frac{\partial}{\partial\lambda}\,\frac{1}{N}\,\langle \ln Z_N\rangle\ , \tag{3}$$

with

$$Z_N = \int \prod_{i=k}^{N} \frac{\mathrm{d}u_k}{\sqrt{2\pi/\mathrm{i}}}\ \exp\Big\{-\frac{\mathrm{i}}{2}\sum_{k,\ell} u_k(\lambda_\varepsilon\delta_{k\ell} - C_{k\ell})u_\ell\Big\}\ . \tag{4}$$

The angled brackets in Eq. (3) indicate an ensemble average, which can be evaluated using replicas. Analogous calculations in random matrix theory [21] suggest that the final results will exhibit the structure of a replica-symmetric high-temperature solution, and hence that an annealed calculation (which replaces $\langle \ln Z_N\rangle$ by $\ln\langle Z_N\rangle$ in (3)) will provide exact results. Indeed, both the Wigner semi-circle law for spectral densities of Gaussian random matrices and the Marčenko Pastur law for spectral densities of Wishart matrices can be obtained from such an annealed calculation, and this is the approach we adopt here.

Inserting the definition Eq. (1) into Eq. (4), one notes that Z_N depends on the disorder, i.e. on the $\{x_n\}$, only through the M variables

$$z_i = \frac{1}{\sqrt{N}}\sum_{k=1}^{N} x_{i+k}u_k\ ,\quad 0 \le i < M\ . \tag{5}$$

Assuming that the true auto-covariance $\bar{C}(k) = \langle x_n x_{n+k}\rangle$ is absolutely summable, we can argue from the central limit theorem (CLT) for weakly dependent random variables that the z_i will be correlated Gaussian variables with $\langle z_i\rangle = 0$, and covariance matrix Q whose elements are given in terms of $\bar{C}$ as

$$Q_{ij} = \langle z_i z_j\rangle = \frac{1}{N}\sum_{k\ell}\bar{C}(i-j+k-\ell)\,u_k u_\ell\ . \tag{6}$$

The disorder average $\langle\ldots\rangle$ is thus a Gaussian integral, which can be performed to give

$$\begin{aligned}\langle Z_N\rangle\ =\ \int\prod_k \frac{\mathrm{d}u_k}{\sqrt{2\pi/\mathrm{i}}}\ \exp\Big\{-\frac{\mathrm{i}}{2}\lambda_\varepsilon\sum_k u_k^2 \\ -\frac{1}{2}\ln\det(\mathbb{I}-\mathrm{i}\alpha Q)\Big\}\ .\end{aligned} \tag{7}$$

The matrix Q being Toeplitz, we will use Szegö's theorem [20] to evaluate the 'spectral sum' $\ln\det(\mathbb{I}-\mathrm{i}\alpha Q)=$ Tr $\ln(\mathbb{I}-\mathrm{i}\alpha Q)$ in Eq. (7).

Briefly, Szegö's theorem states that for Toeplitz matrices A which have matrix elements a_{ij} depending only on the difference of their arguments, $a_{ij} = a_{i-j}$, the spectral density in the limit of large matrix dimension N can be expressed in terms of the Fourier transform or "symbol" $\hat{a}(q)$ of the sequence $(a_n)_{n\in\mathbb{Z}}$ as

$$\rho_N(\lambda;A) \to \int_{-\pi}^{\pi}\frac{\mathrm{d}q}{2\pi}\,\delta(\lambda-\hat{a}(q))\ ,\quad N\to\infty\ , \tag{8}$$

provided the a_n decrease sufficiently rapidly with $|n|$. The convergence is understood in the weak sense.

Given that our sequence of Q-matrices doesn't fully fit the assumptions of the standard theory in that the matrix elements are *themselves* dependent on the dimension M, we expect this to be only an approximation; it should, however, become exact in the limit $\alpha\to 0$.

To proceed, we need Fourier representations of Q, and we will have to keep track of finite-M, finite-N expressions in what follows. Assuming M to be odd, we have

$$Q_{ij} = \frac{1}{M}\sum_{\mu=-(M-1)/2}^{(M-1)/2} \mathrm{e}^{-\mathrm{i}q_\mu(i-j)}Q_\mu \tag{9}$$

for the ($\{u_k\}$ dependent) elements of Q, with

$$\begin{aligned}Q_\mu\ &=\ \frac{1}{N}\sum_{k\ell}\hat{C}(q_\mu)\mathrm{e}^{-\mathrm{i}q_\mu(k-\ell)}\,u_k u_\ell \\ &=\ \hat{C}(q_\mu)|\hat{u}(q_\mu)|^2 \equiv Q(q_\mu)\end{aligned} \tag{10}$$

where $q_\mu = \frac{2\pi}{M}\,\mu$ and $\hat{u}(q_\mu) = \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\mathrm{e}^{\mathrm{i}q_\mu k}u_k$. Here

$$\hat{C}(q_\mu) = \sum_{n=-(M-1)/2}^{(M-1)/2}\bar{C}(n)\mathrm{e}^{\mathrm{i}q_\mu n} \tag{11}$$

is the Fourier transform of the true auto-covariance function of the underlying process. Truncating the sum in Eq. (11) at $|n| \le (M-1)/2$, as we have done, will create negligible errors in the large M limit if $\sum_{n=-\infty}^{\infty}|\bar{C}(n)|$ exists, as was required already when appealing to the CLT for the z_i statistics above. The fact that we have restricted the q_μ values to the discrete grid with spacing $2\pi/M$ approximates Q by its cyclified version. In Szegö's terminology, the matrix Q has $Q_\mu = Q(q_\mu)$ as its (M-grid) symbol, and Szegö's approximation for the spectral sum reads

$$\ln\det(\mathbb{I} - \mathrm{i}\alpha Q) \simeq \sum_{\mu=-(M-1)/2}^{(M-1)/2} \ln\big(1 - \mathrm{i}\alpha Q_\mu\big) \ . \tag{12}$$

The symmetry $\bar{C}(n) = \bar{C}(-n)$ entails $\hat{C}(q_\mu) = \hat{C}(-q_\mu)$, thus $Q(q_\mu) = Q(-q_\mu)$. Next, one extracts the $\{u_k\}$ dependence (via the $\{Q_\mu\}$) from the evaluation of (12), using δ-functions and their Fourier representations. The $\{u_k\}$ integrals then become Gaussian, and we can express $\langle Z_N\rangle$ as

$$\begin{aligned}\langle Z_N\rangle \ &= \int \prod_{\mu=0}^{(M-1)/2} \frac{\mathrm{d}\hat{Q}_\mu \mathrm{d}Q_\mu}{2\pi}\ \exp\Big\{-\sum_{\mu=0}^{(M-1)/2} \mathrm{i}\hat{Q}_\mu Q_\mu \\ &- \sum_{\mu=0}^{(M-1)/2} \ln(1 - \mathrm{i}\alpha Q_\mu) - \frac{1}{2}\ln\det(\lambda_\varepsilon\mathbb{I} - R)\Big\}. \end{aligned} \tag{13}$$

The elements of the $N \times N$ matrix R in (13) are given by $R_{k\ell} = \frac{2}{N}\sum_{\mu=0}^{(M-1)/2} \hat{Q}_\mu \hat{C}(q_\mu)\cos(q_\mu(k-\ell))$, with $1 \le k,\ell \le N$. We have combined modes with μ and $-\mu$ and neglected as sub-leading the fact that the resulting prefactors differ for the $\mu = 0$ mode.

We next use residues to evaluate the Q_μ integrals in (13):

$$\int \frac{\mathrm{d}Q_\mu}{2\pi}\, \frac{\mathrm{e}^{-\mathrm{i}\hat{Q}_\mu Q_\mu}}{1 - \mathrm{i}\alpha Q_\mu} = \begin{cases} \alpha^{-1}\,\mathrm{e}^{-\hat{Q}_\mu/\alpha} & ;\ \hat{Q}_\mu > 0\ , \\ 0 & ;\ \text{else}\ . \end{cases}$$

After rescaling $\hat{Q}_\mu/\alpha \to \hat{Q}_\mu$ this yields

$$\langle Z_N\rangle \ = \ \Big\langle \exp\Big\{ -\tfrac{1}{2}\ln\det(\lambda_\varepsilon\mathbb{I} - R)\Big\}\Big\rangle_{\{\hat{Q}_\mu\}} \tag{14}$$

with now

$$R_{k\ell} = \frac{2}{M}\sum_{\mu=0}^{(M-1)/2} \hat{Q}_\mu \hat{C}(q_\mu)\cos(q_\mu(k-\ell)) \ . \tag{15}$$

In Eq. (14) we have introduced the short-hand

$$\langle \dots \rangle_{\{\hat{Q}_\mu\}} = \int_0^\infty \prod_{\mu=0}^{(M-1)/2} \Big\{\mathrm{d}\hat{Q}_\mu \mathrm{e}^{-\hat{Q}_\mu}\Big\}\,(\dots) \tag{16}$$

for the $\hat{Q}_\mu$-integrals. As the notation indicates, these amount to averages over exponentially distributed random variables of unit mean. Hence within our Szegö-approximation, the original spectral problem for sample auto-covariance matrices C is equivalent to that for random Toeplitz matrices R given by (15).

To make progress we use the fact that the matrices R, too, are Toeplitz, and approximate the spectral sum $\ln\det(\lambda_\varepsilon\mathbb{I} - R)$ appearing in (14) in terms of Szegö's theorem,

$$\ln\det(\lambda_\varepsilon\mathbb{I} - R) \simeq \sum_{\nu=-(N-1)/2}^{(N-1)/2} \ln\big(\lambda_\varepsilon - R_\nu\big)\ , \tag{17}$$

with

$$\begin{aligned} R_\nu \ &= \ \frac{1}{M}\sum_{\mu=0}^{(M-1)/2} \hat{Q}_\mu \hat{C}(q_\mu) \sum_{n=-(N-1)/2;\sigma=\pm1}^{(N-1)/2} \mathrm{e}^{\mathrm{i}(p_\nu + \sigma q_\mu)n} \\ &= \ \sum_{\mu=0}^{(M-1)/2} \hat{Q}(q_\mu)\hat{C}(q_\mu) S_{\nu\mu} \equiv R(p_\nu)\ , \end{aligned} \tag{18}$$

for $p_\nu = \frac{2\pi}{N}\,\nu$ defined on a grid of spacing $2\pi/N$, and the S-kernel given by

$$S_{\nu\mu} = \frac{1}{M}\sum_{\sigma=\pm1} \frac{\sin(N(p_\nu + \sigma q_\mu)/2)}{\sin((p_\nu + \sigma q_\mu)/2)}\ . \tag{19}$$

Next one extracts the $\hat{Q}_\mu$ dependence from the spectral sum (17) by enforcing the R_ν-definitions using δ-functions and their Fourier representations. This enables one to perform the R_ν integrals using residues much as in the case of the Q_μ integrals above, giving

$$\langle Z_N\rangle = \Big\langle \prod_{\nu=0}^{(N-1)/2} F_\nu \Big\rangle_{\{\hat{Q}_\mu\}} \tag{20}$$

with

$$F_\nu = \mathrm{i}\int_0^\infty \mathrm{d}\hat{R}_\nu\, \mathrm{e}^{-\mathrm{i}\hat{R}_\nu\left(\lambda_\varepsilon - \sum_{\mu=0}^{(M-1)/2}\hat{Q}_\mu\hat{C}(q_\mu)S_{\nu\mu}\right)}\ . \tag{21}$$

The coupling via the S-kernel entails that the F_ν for different ν are correlated. To proceed, we exploit the property that the S-kernel is rapidly oscillating, and sharply peaked at $|p_\nu \pm q_\mu| \simeq \mathcal{O}(1/N)$. The dominant contributions to the exponential in (21) at fixed ν therefore lie in the interval $I_\nu = \{\mu : |\nu \pm \alpha\mu| \le 1\}$. We therefore approximate the S-kernel by a rectangular window on I_ν. The height of this window is set by the requirement that $\sum_{\mu=0}^{(M-1)/2} S_{\nu\mu} = 1$. Given that I_ν contains $2/\alpha$ indices μ, the height must be chosen as $\alpha/2$. Using also smoothness of $\hat{C}(q_\mu)$ on the q_μ-scale, we thus approximate

$$\sum_{\mu=0}^{(M-1)/2} \hat{Q}_\mu\hat{C}(q_\mu)S_{\nu\mu} \simeq \frac{\alpha}{2}\hat{C}(p_\nu)\sum_{\mu\in I_\nu}\hat{Q}_\mu\ . \tag{22}$$

As the I_ν are overlapping, the F_ν in (21) remain correlated. As a last step we ignore these residual correlations

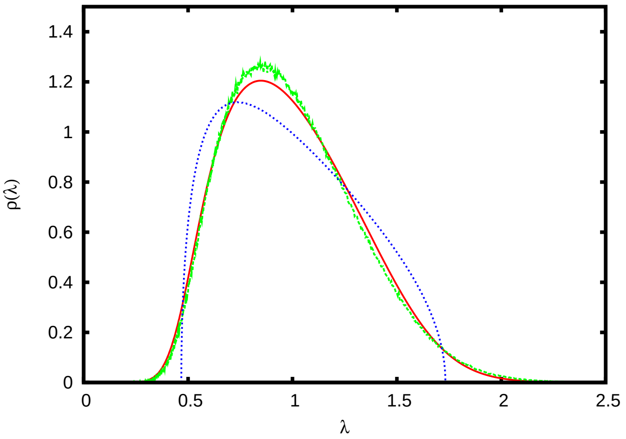

FIG. 1: (Colour online) Spectral density for sample auto covariance matrices of i.i.d. signals $x_n \sim \mathcal{N}(0,1)$ at $\alpha = 0.1$ (green dashed curve); analytic approximation, Eq. (25) for $\rho_\alpha^{(0)}(\lambda)$ (red curve). The Marčenko-Pastur law (blue dotted curve) for the same α is also shown for comparison.

and substitute $y = \alpha \hat{R}_\nu \hat{C}(q_\mu)/2$ in Eq. (21) to arrive at a closed form approximation for $\langle Z_N \rangle$:

$$\langle Z_N\rangle \simeq \prod_{\nu=0}^{(N-1)/2}\left\{\frac{2\,\mathrm{i}}{\alpha\hat{C}(p_\nu)}\int_0^\infty \mathrm{d}y\,\frac{\mathrm{e}^{-\mathrm{i}y\lambda_\varepsilon 2/(\alpha\hat{C}(p_\nu))}}{\left(1-\mathrm{i}y\right)^{2/\alpha}}\right\} \qquad (23)$$

For the spectral density (3) in the thermodynamic limit $N \to \infty$, $M \to \infty$, keeping the aspect ratio $\alpha = N/M$ fixed, we then get

$$\begin{aligned}\rho_\alpha(\lambda) &= -\frac{2}{\pi}\lim_{\varepsilon\to 0}\mathrm{Im}\frac{\partial}{\partial\lambda}\lim_{N\to\infty}\frac{1}{N}\ln\langle Z_N\rangle \\ &= \int_0^\pi \frac{\mathrm{d}q}{\pi}\,\frac{1}{\hat{C}(q)}\,\rho_\alpha^{(0)}(\lambda/\hat{C}(q)) \qquad (24)\end{aligned}$$

in which

$$\rho_\alpha^{(0)}(\lambda) = -\lim_{\varepsilon\to 0}\frac{1}{\pi}\mathrm{Im}\frac{\partial}{\partial\lambda}\ln I_\alpha\left(\frac{2}{\alpha}\lambda_\varepsilon\right) \qquad (25)$$

with I_α obtained from (23) in terms of an incomplete Γ-function: for $\mathrm{Im}\, x < 0$,

$$\begin{aligned}I_\alpha(x) &\equiv \int_0^\infty \mathrm{d}y\,\mathrm{e}^{-\mathrm{i}yx}\left(1-\mathrm{i}y\right)^{-2/\alpha} \\ &= \mathrm{i}\,(-x)^{-1+2/\alpha}\,\mathrm{e}^{-x}\Gamma(1-2/\alpha,-x)\ . \qquad (26)\end{aligned}$$

Note that Eq. (24) implies that the scaling function $\rho_\alpha^{(0)}(\lambda)$ has an independent meaning as the spectral density of the empirical auto-covariance matrix (at the same value of α) for a sequence of *uncorrelated* data, for which $\hat{C}(q) \equiv 1$. Eq. (24) thus constitutes a remarkable scaling relation relating the spectral density of sample auto-covariance matrices for processes with dynamical correlation to the spectral density of sample auto-covariance matrices for i.i.d. sequences of random data.

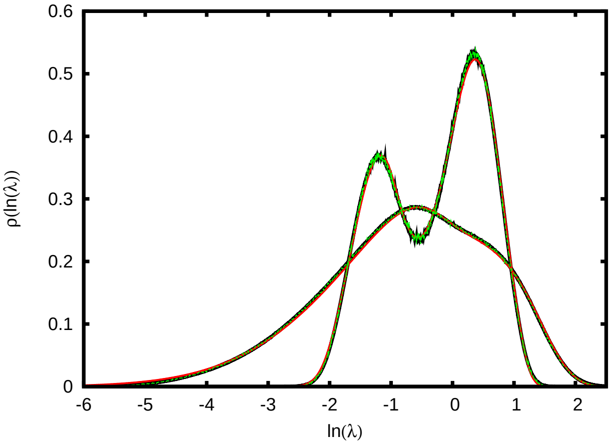

FIG. 2: (Colour online) Test of scaling for spectra of sample auto-covariance matrices of an AR2 process as described in the main text at $\alpha = 0.1$ (narrower set of curves) and $\alpha = 0.8$ (wider set of curves). For both α's we show simulation results (green dashed curve) and scaling results using either the *empirical* scaling function (black full curve), or our analytic approximation (25) (red full curve).

We note in passing that the approximations introduced to evaluate the spectral sums for the R matrix (15), which are needed in the general case to deal with the coupling of terms through the S-kernel, would not be required at $\alpha = 1$ where the R matrix is diagonalized exactly in terms of Fourier modes. Taking the exponential distribution of the $\hat{Q}_\mu$ in (15) into account, one would obtain a scaling function for this case that is itself exponential, $\rho_1^{(0)}(\lambda) = \mathrm{e}^{-\lambda}$. This is different from but qualitatively and quantitatively very close to the $\alpha \to 1$ limit of the approximate result (25), (26).

Numerical tests We checked our results using simulations of 800×800 matrices, taking averages over 5000 realizations. Fig. 1 compares results for a sequence of i.i.d. variables with our prediction (25) for the scaling function $\rho_\alpha^{(0)}$, and the Marčenko-Pastur law at $\alpha = 0.1$. The figure clearly shows that our analytic prediction, while not exact, captures the salient features of the spectra of auto-covariance matrices for sequences of i.i.d. variables reasonably well, including in particular the peak position, as well as the fact that the spectra have non-compact support. It also shows that spectral properties of the ensemble of sample auto-covariance matrices are qualitatively different from those of Wishart matrices.

Fig. 2 checks our scaling prediction for an AR2 process of the form $x_n + \frac{1}{2}x_{n-1} + \frac{5}{16}x_{n-2} = \sigma\xi_n$, with i.i.d. $\xi_n \sim \mathcal{N}(0,1)$, and σ chosen to ensure that $\bar{C}(0) = \langle x_n^2 \rangle = 1$. Figure 3 does the same for a process P with a power-law decay of its auto-correlations of the form $\bar{C}(k) = 1/(1 + (k/2)^2)$. Results are shown as probability density functions for logarithms of eigenvalues so as to better resolve tails at small and large λ. The spectra

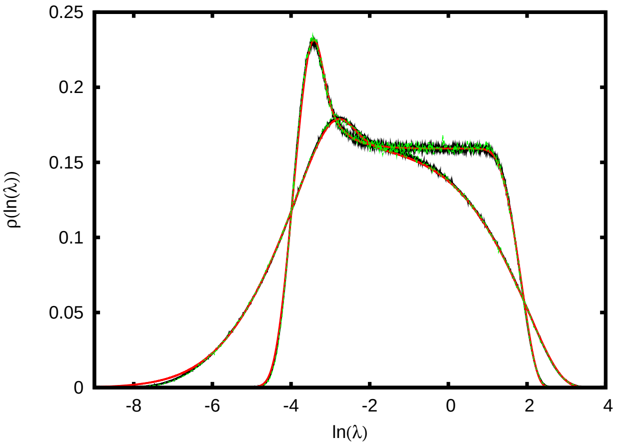

FIG. 3: (Colour online) Test of scaling for spectra of sample auto-covariance matrices of a process P with power-law decay of its auto-covariance function at $\alpha = 0.1$ (narrower set of curves) and $\alpha = 0.8$ (wider set of curves). For both α's we show simulation results (green dashed curve) and scaling results using either the *empirical* scaling function (black full curve), or our analytic approximation (25) (red full curve).

in Figs. 2 and 3 are both evaluated at two values of α. In each case we compare simulations with scaling based either on our analytic approximation (25) for $\rho_\alpha^{(0)}$, or on an empirical scaling function determined via simulation, with an $\alpha = 0.1$-example shown in Fig. 1. Agreement is excellent throughout: apart from small discrepancies in the tails we find that the curves representing simulation results and empirical and analytic scaling predictions lie virtually on top of each other. Larger noise levels in the empirical scaling prediction for the process with a power-law decay of its auto-correlation are due to the existence of $\hat{C}(q) \ll 1$-values in its Fourier transform; these lead to relatively large contributions to the scaling integral (24) from λ's in the *tails* of the scaling function $\rho_\alpha^{(0)}$, where our empirical estimates of $\rho_\alpha^{(0)}(\lambda)$ are necessarily poorer due to scarcity of events.

Summary and discussion In summary, we have computed spectra of sample auto-covariance matrices in the limit of large matrix dimension N, at fixed value of the aspect ratio $\alpha = N/M$ of the number N of time-lags included in the matrix and the size M of the samples used to define empirical averages. We find a remarkable scaling relation which expresses the spectral density $\rho_\alpha(\lambda)$ of sample auto-covariance matrices for processes *with* dynamical correlations as a continuous superposition of rescaled copies of the spectral density $\rho_\alpha^{(0)}(\lambda)$ for a sequence of *uncorrelated*, i.i.d. random variables, with the rescaling factors given in terms of the Fourier transform $\hat{C}(q)$ of the true auto-covariance function. We also obtain a simple closed form expression for $\rho_\alpha^{(0)}$ that provides an excellent approximation to numerically simulated spectra.

Our analytical calculations are based on a number of approximations, viz. (i) the annealed approximation for the computation of free energies, (ii) the use of Szegö's theorem for the evaluation of spectral sums involving the Q and R matrices in Eqs. (12) and (17), respectively, (iii) a rectangular window approximation for describing couplings via the S-kernel in (22), and finally (iv) a decorrelation approximation used to obtain the final product representation (23) of the partition function.

Experience with matrices from the Gaussian orthogonal ensemble and the Wishart-Laguerre ensemble suggest that the annealed approximation is indeed exact. While the use of Szegö's theorem for the evaluation of the spectral sum (12) involving the Q matrix can be argued to be exact in the $\alpha \to 0$-limit, and thus to constitute an approximation that can be considered as controlled, it is harder to assess the accuracy of this approximation in the case of the R matrix in (17). Finally, both the rectangular window approximation (iii) and the decorrelation approximation (iv) are largely uncontrolled. Thus all elements of our analysis clearly deserve further scrutiny, and improvement where possible. For the time being they derive their ultimate justification mainly through the excellent results they produce – even, as we have seen, for α as large as 0.8. A fuller account of our results, including in particular more detailed results on the quality of our approximations in the tail regions of large and small eigenvalues, as well as a proper quenched calculation will appear in a longer journal article.

Our numerical evidence very strongly suggests that the scaling result (24) itself is indeed exact, as long as the true functional form for the spectral density $\rho_\alpha^{(0)}$ of sample auto-covariances for i.i.d. random variables is used, rather than our analytical approximation (25). We have elements of an independent proof of scaling which we intend to publish in a forthcoming paper.

In the $\alpha \to 0$ limit, random fluctuations in sample auto-covariance matrices will be suppressed and these matrices will thus be arbitrarily close to Toeplitz form, with their symbol given by the Fourier transform $\hat{C}(q)$ of the true auto-covariance functions. For such matrices Szegö's theorem can be invoked to describe the limiting spectral density in terms of the integral representation (8) with $\hat{a}(q) \equiv \hat{C}(q)$. Using properties of the Dirac δ-function (and $\hat{C}(q) = \hat{C}(-q)$) the result can be written in terms of the scaling form (24), with

$$\rho_0^{(0)}(\lambda) = \delta(\lambda - 1) \ . \tag{27}$$

This form of $\rho_0^{(0)}$ is indeed the correct $\alpha \to 0$ limit of the spectral density for auto-covariance matrices of i.i.d. data (of zero-mean and unit variance), which establishes that scaling is exact in the $\alpha \to 0$-limit. Our scaling result (24) could thus be thought of as a *generalization* of Szegö's theorem for randomly perturbed Toeplitz matrices; for $\alpha \to 0$ it recovers the Szegö result because the scaling function (25) then converges to (27). We have checked that scaling as in (24) also holds for sample auto-covariance matrices which, as in [18], are constructed as (random) Toeplitz matrices from the start, albeit with different forms for the scaling functions.

Judging from the impact which results for spectral properties of (Wishart-Laguerre) sample covariance matrices have had, we believe our results to hold *significant* potential for applications in a wide variety of fields, including time-series analysis, information theory, signal processing, or finance.

Specifically for time-series analysis [1] our results could be used to provide estimates for parameters governing auto-regressive processes, including in particular reliable estimates for the order of the process generating a given data sequence. Alternatively, they could be used as an independent systematic tool to correct for finite sample effects in estimating Fourier transforms $\hat{C}(q)$ of auto-covariance functions, which could then feed into similarly correcting estimated bounds for one-step prediction errors for stationary processes, and so on.

In information theory [22], spectral properties of auto-covariance matrices are used to estimate both, entropy rates and Shannon rate-distortion functions of stationary Gaussian processes, and as in the case of time-series analysis our tools could conceivably help to systematically correct for errors in these estimates that are induced by finite sample effects.

In finance one could contemplate a translation of Markowitz portfolio optimization into the time domain, looking at optimal liquidation strategies for a single asset across a given time window, and utilizing knowledge about spectral properties of sample auto-covariance matrices in a manner analogous to the uses of spectral theory of Wishart matrices for the standard portfolio optimization advocated in [16, 17].

Finally, the natural next methodical step is to generalize our results to sample covariance matrices for multiple time-series, for which sample covariance matrices will be randomly perturbed block-Toeplitz matrices, and we have indeed been able to make first promising steps in that direction [23].

Acknowledgement: It is a pleasure to thank K. Anand, L. Dall'Asta and P. Vivo for illuminating discussions on the occasion of a visit of RK to the ICTP at Trieste, which triggered the present investigation.

[1] J. D. Hamilton, *Time Series Analysis* (Princeton University Press, Princeton, NJ, 1994).

[2] M. L. Mehta, *Random Matrices, 3rd Edition* (Elsevier, Amsterdam, 2004).

[3] G. Akemann, J. Baik, and P. D. Francesco, eds., *The Oxford Handbook of Random Matrix Theory* (Oxford University Press, Oxford, 2011).

[4] J. Wishart, Biometrika **20 A**, 32 (1928).

[5] V. A. Marčenko and L. A. Pastur, Math. USSR-Sb. **1**, 457 (1967).

[6] I. M. Johnstone, Ann. Stat. **29**, 295 (2001).

[7] C. A. Tracy and H. Widom, Comm. Math. Phys. **177**, 727 (1996).

[8] P. Vivo, S. N. Majumdar, and O. Bohigas, J. Phys. A **40**, 4317 (2007).

[9] E. Katzav and I. Pérez Castillo, Phys. Rev. E **82**, 040104 (4pp) (2010).

[10] Z. Burda, J. Jurkiewicz, and B. Waclaw, Acta Phys. Polon. B **36**, 2641 (2005).

[11] G. Akemann and P. Vivo, J. Stat. Mech. **2008**, P09002 (31pp) (2008).

[12] C. Biely and S. Thurner, Quant. Fin. **8**, 705 (2008).

[13] C. Recher, M. Kieburg, and T. Guhr, Phys. Rev. Lett. **105**, 244101 (2010).

[14] R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).

[15] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications* (now Publishers Inc, Hanover, MA, 2004).

[16] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).

[17] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amaral, and H. E. Stanley, Phys.Rev. Lett. **83**, 1471 (1999).

[18] A. Basak, A. Bose, and S. Sen, preprint arXiv:1108.3147 **[math.PR]**, 41pp (2011).

[19] W. Bryc, A. Dembo, and T. Jiang, Ann. Prob. **34**, 1 (2006).

[20] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications* (American Mathematical Society, Providence, RI, 1984).

[21] S. F. Edwards and R. C. Jones, J. Phys. A **9**, 1595 (1976).

[22] T. A. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Edition* (Wiley, Hoboken, NJ, 2006).

[23] R. Kühn and P. Sollich, *Spectra of Sample Covariance Matrices for Multiple Time-Series*, in preparation (2012).