Cleaning large-dimensional covariance matrices for correlated samples

Zdzislaw Burda* and Andrzej Jarosz†
AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, al. Mickiewicza 30, 30-059 Krakow, Poland

A non-linear shrinkage estimator of large-dimensional covariance matrices is derived in a setting of auto-correlated samples, thus generalizing the recent formula by Ledoit-Péché. The calculation is facilitated by random matrix theory. The result is turned into an efficient algorithm, and an associated Python library, `shrinkage`, with help of Ledoit-Wolf kernel estimation technique. An example of exponentially-decaying auto-correlations is presented.

Keywords: random matrix theory, Marčenko-Pastur equation, free probability, shrinkage estimators

In experimental research, when working with large datasets, one often faces a generic problem of determining correlations between multiple entities of interest, on the basis of observed data. The problem, in its minimalistic form, can be formulated as follows: One has a statistical system with N degrees of freedom (indexed henceforth by $i, j = 1, \ldots, N$), and one performs T measurements of its state (indexed henceforth by $t, s = 1, \ldots, T$), collecting these observations in an $N \times T$ matrix $\boldsymbol{Y} = [Y_{it}]$. The challenge lies in estimating from this data the underlying two-point correlations $\boldsymbol{C} = [C_{ij}] = \langle Y_i Y_j \rangle$ (assume standardized samples, for simplicity of discussion), the so-called *population* (or "true", "signal") correlation matrix.

This setting is very general, commonly encountered in physics, finance, genomics and bioinformatics, signal processing and acoustics, image recognition, speech recognition, cancer research, climatology, neuroscience, and many other areas. For instance, in lattice QCD, one may consider [1] neutron-antineutron two- and three-point correlation functions, described using N ($\sim$ a few dozen) parameters related to the number of dynamics time steps; one simulates gauge field and fermions on the lattice, obtaining thereby T samples (in one version of the procedure, one sample per gauge field configuration, thus $T \sim$ a few dozen) for these correlation functions; these samples are used to estimate the statistical correlation matrix between the N degrees of freedom, necessary for a least-squares fit to the correlation functions. In cosmology [2], when investigating cosmic shear, the distortion of images of distant galaxies by weak gravitational lensing caused by the large-scale structure, the basic observable is the shear two-point correlation function, computed from products of ellipticities of galaxy pairs located within given angular bins, $N \sim$ a hundred to thousand; this correlation function can be estimated from mock simulations, only $T \sim$ a few thousands, as they are computationally expensive. In finance, one needs to know correlations between N investment assets, e.g. S&P500 stocks ($N \sim$ a few hundred), in order to construct optimal, well-diversified portfolios, in the spirit of Markowitz theory [3, 4]; we estimate these correlations from T daily stock returns ($T \sim$ hundreds to thousands, due to stationarity requirements). In genomics [5], one may analyze expression profiles of N genes ($N \sim$ thousands, from which we may select a subset of differentially expressed genes, say $N \sim$ a hundred) in order to describe similarities and functional groupings of them; each gene is sampled T times in a microarray experiment ($T \sim$ a few observation times), and the resulting $\boldsymbol{Y}$, having chosen some suitable gene expression measure, is called a gene expression matrix. In signal processing and acoustics [6], one considers an array of N ($\sim$ a dozen) directional sensors, and performs so-called beamforming, combining elements of an antenna array in such a way as to have constructive/destructive interference along given directions; but (frequency-domain) beamforming algorithms, such as the Capon beamformer, require the knowledge of the correlation matrix between the array outputs, which in turn is estimated from T ($\sim$ a few dozen) samples. In anomaly detection in hyperspectral images [7], one has a sensor "image" comprised of N spectral bands ($\sim$ a hundred), i.e., collected narrow ranges of the electromagnetic spectrum; when the goal is detecting anomalous pixels in such an image, one needs the Mahalanobis distance between each (N-dimensional) pixel and their background mean; that in turn requires the correlation matrix between the bands, estimated from a local sliding window of pixels, of length T which is not too large in order to cover a homogeneous area of space ($\sim$ hundreds).

This problem of estimating the correlation matrix is thus of paramount importance in very diverse fields of science, and has accordingly merited a large body of research. Indeed, when $N \ll T$, the setting is that of classical multivariate statistics, and the usual *sample* estimator,

$$\boldsymbol{E} = \frac{1}{T} \boldsymbol{Y} \boldsymbol{Y}^\top, \tag{1}$$

converges almost surely to $\boldsymbol{C}$, making it a strongly consistent estimator of correlations. In particular, the expected value of the Frobenius norm of their difference, $\|\boldsymbol{E} - \boldsymbol{C}\|^2 = \mathrm{Tr}\,(\boldsymbol{E} - \boldsymbol{C})^2$ (i.e., the mean squared error, MSE) is close to zero. Yet quite often, like in the above examples, one rather faces a different limit, that of both N and T large and of comparable size, $q = N/T \sim 1$;

sometimes (e.g. in genomics) even $q \gg 1$. In this *big data regime*, the simultaneous estimation of $\sim N^2$ elements of $\boldsymbol{C}$ from $T \sim N$ samples leads to substantial noise (variance) in the sample estimator $\boldsymbol{E}$, rendering the whole process close to meaningless. This challenge has been met with considerable effort to construct other estimators $\boldsymbol{\Xi}$ of $\boldsymbol{C}$ with lower MSE; one way to achieve this is via *shrinkage*: estimators with higher bias, but much lower variance, so that the MSE (= bias^2 + variance) is reduced.

This is reminiscent of the James-Stein (JS) estimator [8], a biased estimator of the mean $\boldsymbol{\mu}$ in dimension $N \geq 3$, with better MSE than the least-squares (maximum likelihood) estimator. The JS estimator can be naturally understood within the *empirical Bayes* approach (cf. [3] for an extensive discussion), and a similar Bayeasian reasoning (with the so-called *inverse Wishart* prior) leads to the first successful attempt at constructing a shrinkage estimator of correlations, the *Ledoit-Wolf linear shrinkage* [9], $\boldsymbol{\Xi} = \alpha_s \boldsymbol{E} + (1-\alpha_s)\boldsymbol{1}$, interpolating (with some α_s estimated from data) between the sample estimator and the null hypothesis (here: the unit matrix, but it can encode a more specific prior belief). This simple result has been the workhorse of large-dimensional estimation, and can be found in all the applications mentioned in the introduction. The choice of the inverse Wishart conjugate pair is, however, dictated more by computational tractability than any insight gleaned from data, and in fact the observed sample estimator $\boldsymbol{E}$ strongly constraints the prior distribution of $\boldsymbol{C}$; indeed, through the Marčenko-Pastur law, which we discuss below.

The recent strand, initiated by [10], of research [3, 11–14] thus approaches the problem of constructing $\boldsymbol{\Xi}$ in greater generality and with marked attention to data: First, consider the eigendecomposition of the sample estimator, $\boldsymbol{E} = \sum_{i=1}^{N} \lambda_i |u_i\rangle\langle u_i|$; we have already established that due to the curse of dimensionality both the eigenvalues and eigenvectors have high variance, "concealing" behind this noise the underlying nature of $\boldsymbol{C}$. Second, suppose we have no prior belief on the eigenvectors of $\boldsymbol{C}$; that our goal is merely to "clean" the eigenvalues. In the absence of such a bias in any particular direction, the only available basis is that of $\boldsymbol{E}$; in other words, we wish to seek a MSE-optimal estimator of $\boldsymbol{C}$ in the space of matrices with the eigenvectors $|u_i\rangle$,

$$\boldsymbol{\Xi} = \sum_{i=1}^{N} \xi_i |u_i\rangle\langle u_i|. \tag{2}$$

This is called a *rotationally-invariant estimator* (RIE) [3], because it can alternatively be obtained from a Bayesian argument with a prior on $\boldsymbol{C}$ invariant under orthogonal similarity transformations. Note that the linear shrinkage estimator also has the same sample eigenvectors, so it is a RIE; its eigenvalues are simple linear functions of the sample eigenvalues, $\xi_i = \alpha_s \lambda_i + 1 - \alpha_s$; on the other hand, we will find out that the MSE-optimal *shrunk eigenvalues* ξ_i are nonlinear functions of λ_i, making it an example of *nonlinear shrinkage*.

Minimizing the MSE in this space is a quadratic problem, and so its solution is straightforward, $\xi_i = \langle u_i | \boldsymbol{C} | u_i \rangle$, which may be written as,

$$\xi_i = \frac{1}{N} \sum_{j=1}^{N} c_j \mathcal{O}(\lambda_i, c_j) = \int \mathrm{d}c \rho_{\boldsymbol{C}}(c) c \mathcal{O}(\lambda_i, c). \tag{3}$$

Here we introduced the eigendecomposition of the true correlation matrix, $\boldsymbol{C} = \sum_{i=1}^{N} c_i |v_i\rangle\langle v_i|$, as well as defined the *overlap* between the population and sample eigenvectors, $\mathcal{O} = N \langle u_i | v_j \rangle^2$. The final equality holds in the large-N limit, with $\rho_{\boldsymbol{C}}(c)$ the *density of eigenvalues* of $\boldsymbol{C}$.

This is however a solution only in principle: it depends on the knowledge of $\boldsymbol{C}$, which is not available; after all, the whole purpose of this research is uncovering from data as much information about $\boldsymbol{C}$ as possible. Since it requires foresight into the unknown, this solution is called the *oracle estimator*. It is however a known phenomenon that a MSE-optimal estimator may in fact be independent of the estimated parameter; this happens e.g. when MSE-minimizing for the prefactor of the variance estimator: it is independent of the true σ. And indeed, a similar miracle takes place here in the large-N limit; this we will now demonstrate with help of *random matrix theory* (RMT). Some definitions are first in order. For a random matrix $\boldsymbol{E}$, say our sample estimator, the fundamental object is the *resolvent*; it is a matrix function of a complex argument z, $\boldsymbol{G}_{\boldsymbol{E}}(z) = (z\boldsymbol{1} - \boldsymbol{E})^{-1}$. One might suppose that the randomness of $\boldsymbol{E}$ makes this a random function, but the claim is that at large N, it is self-averaging: independent of any realization of $\boldsymbol{E}$, converging to a deterministic matrix, $\boldsymbol{G}_{\boldsymbol{E}}(z) = \langle \boldsymbol{G}_{\boldsymbol{E}}(z) \rangle$. Its normalized trace, $G_{\boldsymbol{E}}(z) = \frac{1}{N} \mathrm{Tr}\, \boldsymbol{G}_{\boldsymbol{E}}(z) = \int \mathrm{d}\lambda \rho_{\boldsymbol{E}}(\lambda) \frac{1}{z-\lambda}$, is called the *Stieltjes transform*, and encodes information about the spectrum of $\boldsymbol{E}$. In particular, evaluating it near the real axis yields the eigenvalue density as its imaginary part, and so-called *Hilbert transform*, $h_{\boldsymbol{E}}(\lambda) = \mathrm{pv} \int \mathrm{d}\lambda' \rho_{\boldsymbol{E}}(\lambda') \frac{1}{\lambda - \lambda'}$, as the real part, $G_{\boldsymbol{E}}(\lambda - \mathrm{i}\epsilon) = h_{\boldsymbol{E}}(\lambda) + \mathrm{i}\pi\rho_{\boldsymbol{E}}(\lambda)$; here and in the following the limit $\epsilon \to 0^+$ is implicitly assumed. A closely related notion is that of the *M-transform* (or "moment generating function"), in both the matrix form, $\boldsymbol{M}_{\boldsymbol{E}}(z) = z\boldsymbol{G}_{\boldsymbol{E}}(z) - 1$, and analogous scalar form, $M_{\boldsymbol{E}}(z)$, i.e. the normalized trace of the above; it is called so because its series in $1/z$ is comprised of the moments of $\boldsymbol{E}$. Finally, what will become a key ingredient of our main formula below, we have the Voiculescu's χ*-transform* (cf. eq. (3.16) in [15]): you take the scalar M-transform, invert it functionally, and then invert the result numerically, $M_{\boldsymbol{E}}(1/\chi_{\boldsymbol{E}}(z)) = 1/\chi_{\boldsymbol{E}}(M_{\boldsymbol{E}}(z)) = z$. These are the building blocks of the RMT calculation below.

It is a simple step from the definition of the resolvent

to express the eigenvector overlaps as,

$$\mathcal{O}(\lambda, c_j) = \frac{1}{\pi \rho_{\boldsymbol{E}}(\lambda)} \langle v_j | \mathrm{Im}\, \boldsymbol{G}_{\boldsymbol{E}}(\lambda - \mathrm{i}\epsilon) | v_j \rangle. \qquad (4)$$

This is an intermediate step: the overlaps still depend on the knowledge of $\boldsymbol{C}$ (through its eigenvectors, $|v_j\rangle$).

Now comes the central piece of reasoning. We wish to model the relationship between the underlying correlation $\boldsymbol{C}$ and the observed data $\boldsymbol{Y}$, in the hope of capturing essential features of the data-generating process. The simplest assumption is that of Gaussian distribution with cross-correlation $\boldsymbol{C}$, i.e., $\boldsymbol{Y}$ normal with $\langle Y_{it} Y_{js} \rangle = C_{ij}\delta_{ts}$. One may go one step further—and this is the topic of this letter—and introduce also *auto-correlations*, by which we mean correlations between the samples, crucially assumed identical for all the entities, i.e. decoupled from the cross-correlations, $\langle Y_{it} Y_{js} \rangle = C_{ij} A_{ts}$. Alternatively, $\boldsymbol{Y} = \sqrt{\boldsymbol{C}}\boldsymbol{X}\sqrt{\boldsymbol{A}}$, where $\boldsymbol{X}$ is $N \times T$ standard normal. Here, both $\boldsymbol{C}$, of shape $N \times N$, and $\boldsymbol{A}$, of shape $T \times T$, are positive semi-definite real symmetric matrices. This model is simple enough to allow for an analytical solution, yet, as in the central limit theorem, it is believed to describe in the large-dimensional limit a broader class of distributions, only perhaps without fat tails. In fact, even fat tails may be accounted for: it is known that the maximum likelihood estimator of correlations for fat-tailed random variables is the robust Maronna estimator, and it turns out it can be expressed in the form under consideration with a proper choice of $\boldsymbol{A}$ [16].

For this model, RMT can be used [3, 17] to derive the following key equation relating the matrix M-transform of the sample estimator, $\boldsymbol{M}_{\boldsymbol{E}}(z)$, where its scalar version (normalized trace) we denote $M \equiv M_{\boldsymbol{E}}(z)$ for brevity, to that of the parameter matrices $\boldsymbol{C}$ and $\boldsymbol{A}$,

$$\boldsymbol{M}_{\boldsymbol{E}}(z) = \boldsymbol{M}_{\boldsymbol{C}}(Z), \qquad \frac{Z(z)}{z} = \frac{\chi_{\boldsymbol{A}}(qM)}{qM}. \qquad (5)$$

This equation has a very suggestive form: the matrix M-transforms of the sample estimator $\boldsymbol{E}$ and the true correlation matrix $\boldsymbol{C}$ are equal, but at different complex arguments; this complex transformation $z \to Z$ depends on the scalar M-transform of $\boldsymbol{E}$, and does so through the χ-transform of the auto-correlation matrix $\boldsymbol{A}$. This equation has first been derived in [17] using diagrammatic methods; [3] presents another derivation (for $\boldsymbol{A} = \boldsymbol{1}$ only, though the general case is completely analogous) via the replica trick and a low-rank Harish-Chandra-Itzykson-Zuber integral; in its scalar form, that is after taking normalized trace of both sides, it is discussed in [18] within the framework of Voiculescu's free probability theory [15]. One more remark: in the simplest case of both $\boldsymbol{C}$ and $\boldsymbol{A}$ unit matrices, this equation in the scalar version can be solved for M (our shorthand for the scalar M-transform of $\boldsymbol{E}$), which in turn, evaluated near the real axis, gives the density of sample eigenvalues as, $\rho_{\boldsymbol{E}}(\lambda) = \frac{1}{2\pi q \lambda}\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}$, with $\lambda_\pm \equiv (1 \pm \sqrt{q})^2$; this is the famous Marčenko-Pastur law [19], and it beautifully demonstrates the very problem we are trying to solve: the true correlation eigenvalue is 1 (with multiplicity N), but the observed eigenvalues λ_i are scattered around 1, smeared by noise the more the greater $q = N/T$ is, i.e. the more we are into the "big data regime" (and irrespective of how many samples T we have collected!); the underlying eigenvalue 1 is hidden inside this "blob" of sample eigenvalues.

Now it remains to insert the master equation (5) to the overlaps formula (4), and then to the oracle solution (3) for the shrunk eigenvalues, to obtain the main result of this letter, an expression for the MSE-optimal RIE shrinkage in the presence of auto-correlations $\boldsymbol{A}$, dependent only on observable quantities (the λ_i's), and not on the unknown $\boldsymbol{C}$,

$$\frac{\xi_i}{\lambda_i} = \frac{\mathrm{Im}\, \chi_{\boldsymbol{A}}(u_i)}{\mathrm{Im}\, u_i}, \qquad u_i \equiv \alpha_i + \mathrm{i}\beta_i. \qquad (6)$$

We denote $\alpha_i \equiv q(\lambda_i h_{\boldsymbol{E}}(\lambda_i) - 1)$ and $\beta_i \equiv q\pi\lambda_i \rho_{\boldsymbol{E}}(\lambda_i)$. To understand the logic behind this formula: first, we need to assume some model of auto-correlations, $\boldsymbol{A}$; second, when we observe sample eigenvalues λ_i in the data, we need some way of estimating the Hilbert transform $h_{\boldsymbol{E}}(\lambda)$ and density of eigenvalues $\rho_{\boldsymbol{E}}(\lambda)$ of the sample estimator; from these ingredients, we readily calculate the MSE-optimal RIE shrunk eigenvalues ξ_i.

The simplest model of auto-correlations is their absence, $\boldsymbol{A} = \boldsymbol{1}$; in this case, the scalar M-transform, $M_{\boldsymbol{A}}(z) = 1/(z-1)$, and so the χ-transform, $\chi_{\boldsymbol{A}}(z) = z/(z+1)$. Eq. (6) reduces to the *Ledoit-Péché (LP) shrinkage* [10],

$$\frac{\xi_i}{\lambda_i} = \frac{1}{(\alpha_i + 1)^2 + \beta_i^2}, \qquad (7)$$

historically the first nonlinear shrinkage formula.

As a more non-trivial example, consider a model of exponentially-decaying auto-correlations, in which $\boldsymbol{A}$ is a Toeplitz matrix, $A_{ts} = a(|t-s|)$, with $a(t) = \mathrm{e}^{-t/\tau}$, for a given parameter τ. The relevant χ-transform is readily calculated [18] as, $\chi_{\boldsymbol{A}}(z) = 1/(\gamma + \sqrt{\gamma^2 - 1 + 1/z^2})$, with $\gamma \equiv \coth(1/\tau)$. We may even take this a step further, to more complicated auto-correlations generated by VARMA models; for some of them, the χ-transform turns out to be analytically tractable. Notice also that exp-decay is a particular case of VAR(1).

As mentioned, the other critical requirement of our shrinkage formula (6) is a method of estimating the Hilbert transform and density of eigenvalues of $\boldsymbol{E}$. The recent work [14] provides just this missing component, and does so in the form of a very simple and computationally straightforward prescription: both quantities are approximated as a sum of N *kernels* set up around each λ_i. The density, for instance, is approximated as

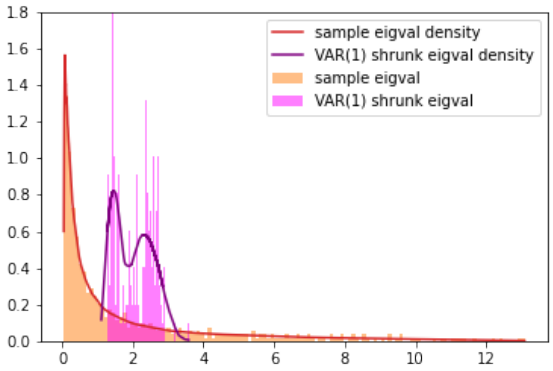

FIG. 1. Gaussian data, $\tau = 3$, and exp-decay shrinkage.

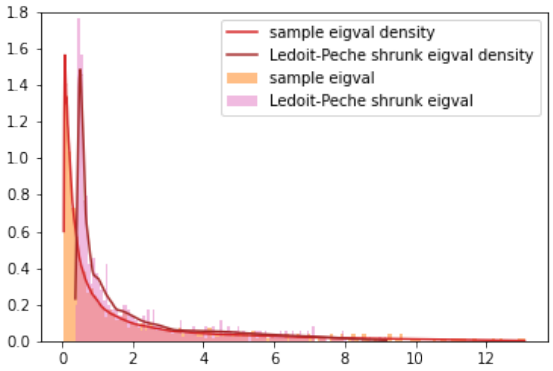

FIG. 2. Gaussian data, $\tau = 3$, and LP shrinkage.

$\rho_{\mathbf{E}}(\lambda) \approx \frac{1}{N}\sum_{i=1}^{N} \mathfrak{r}((\lambda-\lambda_i)/\mathfrak{b}\lambda_i)/\mathfrak{b}\lambda_i$ (analogously for the Hilbert transform), where the *Epanechnikov kernel* is used, $\mathfrak{r}(x) \equiv (3/4\sqrt{5})(1-x^2/5)^+$, and these little parabolas around each sample eigenvalue have a certain width scale, chosen in [14] to be $\mathfrak{b} = T^{-1/3}$ (which we retain). We remark that until this work, the most comprehensive numerical solution consisted of quite a complicated scheme termed *inverse QuEST* [13].

We have developed a Python module `shrinkage`, available at `https://github.com/yedrek/shrinkage`, that contains the outlined functionalities, including kernel estimation of $h_{\boldsymbol{E}}(\lambda)$, $\rho_{\boldsymbol{E}}(\lambda)$, then computing the shrunk eigenvalues (6) for a number of models of $\boldsymbol{A}$, all in a versatile, extensible, and, we hope, easy to use form [20].

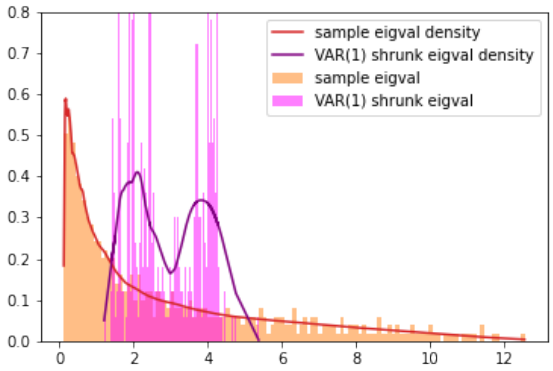

FIG. 3. Student-t data, VMA(1), and exp-decay shrinkage.

Fig. 1 represents an application of the library: We work here with synthetic data, so that knowing the underlying correlations $\boldsymbol{C}$ we may verify how efficient our shrinkage is. We generate $\boldsymbol{Y}$, with $N = 500$, $T = 1000$, from a Gaussian with $\boldsymbol{C}$ containing two eigenvalues, 1 and 3, in 50% proportions, and exp-decay $\boldsymbol{A}$ with $\tau = 3$. All these numbers are in principle unknown; but suppose we know τ, and we use it in our exp-decay shrinkage (6). We observe how the broad expansion of sample eigenvalues is efficiently shrunk to two peaks close to 1 and 3, the true eigenvalues. For comparison, the LP shrinkage does not reveal this much, cf. Fig. 2. More quantitatively, for a given estimator $\boldsymbol{\Xi}$ we may calculate its *Frobenius ratio*, defined as the MSE of $(\boldsymbol{\Xi} - \boldsymbol{C})$ to that of $(\boldsymbol{E} - \boldsymbol{C})$; the lower the ratio, the better the cleaning. Our estimator has it at about 12%, the LP one at 53%.

But τ is really unknown, and moreover the data distribution won't often fit this nicely with the assumed model. In Fig. 3 we have data from the same $\boldsymbol{C}$ as above, but instead VMA(1)-type auto-correlations, and Student-t noise with $\mu = 5$ degrees of freedom. Still, one may use the simple exp-decay shrinkage by searching for τ which yields the lowest error. The Frobenius ratio of such an optimal exp-decay shrinkage estimator is 24%, compared to 33% of LP. Despite model mismatch, the mere fact of incorporating auto-correlations into the shrinkage formula significantly improves the quality of eigenvalue cleaning. This bodes well for applications in real-world data, with genuinely unknown underlying processes.

We thank Artur Swiech and Christopher Wells for interesting discussions.

* zdzislaw.burda@agh.edu.pl; 0000-0002-9656-9570
† jedrekjarosz@gmail.com; corresponding author, 0000-0002-5857-9241

[1] E. Rinaldi, et al., Phys. Rev. D **99**, 074510 (2019).
[2] B. Joachimi, arXiv:1612.00752.
[3] J. Bun, J.-P. Bouchaud, and M. Potters, Phys. Rep. **666**, 1 (2017).
[4] O. Ledoit and M. Wolf, Rev. Financ. Stud., **30**, 4349 (2017).
[5] J. Schäfer and K. Strimmer, Stat. Appl. Genet. Mol. Biol. **4**, 32 (2005).
[6] Y. Zhang, D. Sun, and D. Zhang, Appl. Acoust. **70**, 1029 (2009).
[7] L.R. Bachega, J. Theiler, and C.A. Bouman, 2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 1 (2011).
[8] W. James and C. Stein, Proc. Fourth Berkeley Symp. Math. Statist. Prob. **1**, 361 (1961).
[9] O. Ledoit and M. Wolf, J. Multivar. Anal. **88**, 365 (2004).

[10] O. Ledoit and S. Péché, Probab. Theory Relat. Fields **151**, 233 (2011).
[11] O. Ledoit and M. Wolf, Ann. Stat. **40**, 1024 (2012).
[12] O. Ledoit and M. Wolf, J. Multivar. Anal. **139**, 360 (2015).
[13] O. Ledoit and M. Wolf, Comput. Stat. Data Anal. **115**, 199 (2017).
[14] O. Ledoit and M. Wolf, Ann. Stat. **48**, 3043 (2020).
[15] D.V. Voiculescu, K.J. Dykema, and A. Nica, *Free random variables*, CRM Monograph Series 1, Providence, RI: American Mathematical Society, 1992.
[16] R. Couillet, A. Kammoun, and F. Pascal, J. Multivar. Anal., **143** (2016).
[17] Z. Burda, J. Jurkiewicz, and B. Waclaw, Phys. Rev. E **71**, 026111 (2005).
[18] Z. Burda, A. Jarosz, J. Jurkiewicz, M.A. Nowak, G. Papp, and I. Zahed, Quant. Financ. **11**, 1103 (2011).
[19] V.A. Marčenko and L.A. Pastur, Mat. Sb. **72**, 507 (1967).
[20] A good entry point to the library is the *User's Guide* present in the repository.