EIGENINFERENCE BASED ON ONE- AND TWO-POINT GREEN'S FUNCTIONS*

Zbigniew Drogosz

Institute of Theoretical Physics, Jagiellonian University
Łojasiewicza 11, 30-348, Kraków, Poland
zbigniew.drogosz@doctoral.uj.edu.pl

(Received December 31, 2019)

Two methods of eigeninference, using one-point and two-point Green's functions, are compared and tested on large samples of random matrices. We note that the first method is mostly robust and very fast, whereas the second method is based on a flawed assumption (the matrix appearing in an optimization problem being in many cases not positive-definite) and computationally complex, which limits its usability.

DOI:10.5506/APhysPolB.51.1655

1. Introduction

The Wishart ensemble is the ensemble of sample covariance matrices

$$S = \frac{1}{T} X X^{\dagger}\,, \tag{1}$$

where $X_{i\alpha}$, with $i = 1, \ldots, N$ and $\alpha = 1, \ldots, T$, are independent, Gaussian-distributed numbers. For $T \to \infty$ and $r = \text{const.} < 1$, the entries of a sample covariance matrix S_{ij} would tend to the values Σ_{ij} of the "true" covariance matrix. (The case of the parameter $r = \frac{N}{T}$ being larger than one is uninteresting because of the appearance of $T - N$ trivial eigenvalues, so henceforward we will consider only $r < 1$.)

The problem of determining the eigenvalues of Σ given the eigenvalues of S ("eigen-inference") is complex, and various ways of approaching it have been tried. The search for efficient methods of eigeninference was already active in the 1960s [1]. Since that time, a number of methods have been invented, including the method of characteristic polynomials [2] and Mestre's "G-estimation" [3].

* Presented at the conference *Random Matrix Theory: Applications in the Information Era*, Kraków, Poland, April 29–May 3, 2019.

Here we focus on two methods of eigeninference, a "statistical method" using two-point Green's function and numerical minimization [4, 5] and an "analytical method" using one-point Green's function and Padé approximants [6, 7]. We will describe the methods, discuss potential pitfalls in their application, and test the methods on samples of randomly generated matrices.

2. Preliminaries

The one-point Green's function for the Wishart ensemble,

$$G_S(z) = \frac{1}{N}\left\langle \mathrm{tr}\frac{1}{z\mathbf{1}_N - S}\right\rangle , \tag{2}$$

is a generating function for the spectral moments $\alpha_i^S = \int_L \rho_S(\lambda)\lambda^i \mathrm{d}\lambda$

$$G_S(z) = \sum_{i=0}^{\infty}\frac{1}{z^{i+1}}\left\langle \frac{1}{N}\mathrm{tr}S^i\right\rangle \equiv \sum_{i=0}^{\infty}\frac{\alpha_i^S}{z^{i+1}}, \tag{3}$$

and it also allows one to calculate the spectral distribution from the Sochocki–Plemelj formula

$$\rho_S(\lambda) = -\frac{1}{\pi}\lim_{\epsilon\to 0^+}\Im G_S(z)|_{z=\lambda+i\epsilon}. \tag{4}$$

The two-point Green's function for the Wishart ensemble is

$$G_S(z,w) = \frac{1}{N^2}\left\langle \mathrm{tr}\frac{1}{z\mathbf{1}_N - S}\mathrm{tr}\frac{1}{w\mathbf{1}_N - S}\right\rangle_{\mathrm{c}}, \tag{5}$$

where the subscript 'c' denotes the *connected part* defined as

$$\langle AB\rangle_{\mathrm{c}} \equiv \langle AB\rangle - \langle A\rangle\langle B\rangle . \tag{6}$$

Its double expansion in z and w around infinity yields the double spectral moments

$$\alpha_{i,j}^S = \left\langle \frac{1}{N}\mathrm{tr}\left(S^i\right)\frac{1}{N}\mathrm{tr}\left(S^j\right)\right\rangle_{\mathrm{c}}. \tag{7}$$

The two-point Green's function can be expressed in terms of one-point Green's functions (as a consequence of AJM universality [8])

$$\begin{aligned} G_S(z,w) &= \frac{1}{N^2}\partial_z\partial_w \ln\left[\frac{G_S(w)-G_S(z)}{z-w}\right] \\ &= \frac{1}{N^2}\left[\frac{\partial_z G_S(z)\partial_w G_S(w)}{[G_S(z)-G_S(w)]^2} - \frac{1}{(z-w)^2}\right]. \end{aligned} \tag{8}$$

By expanding all Green's functions in this equation in terms of spectral moments, we can express the sequence of double moments in terms of the single moments

$$\begin{aligned}
\alpha_{11} &= -\alpha_1^2 + \alpha_2\,, \\
\alpha_{12} &= \alpha_{21} = 2\alpha_1^3 - 4\alpha_1\alpha_2 + 2\alpha_3\,, \\
\alpha_{22} &= -6\alpha_1^4 + 16\alpha_1^2\alpha_2 - 6\alpha_2^2 - 8\alpha_1\alpha_3 + \alpha_4\,.
\end{aligned} \tag{9}$$

Analogically, we can also express generating functions for inverse spectral moments (dual moments) in terms of standard Green's functions

$$\begin{aligned}
G_{S^{-1}}\left(\frac{1}{z}\right) &= \left\langle \frac{1}{N}\mathrm{tr}\frac{1}{\frac{1}{z}\mathbf{1}_N - S^{-1}} \right\rangle \\
&= z\left\langle \frac{1}{N}\mathrm{tr}\,\mathbf{1}_N\left(1 - \frac{z}{z\mathbf{1}_N - S}\right)\right\rangle = z(1 - zG_S(z))\,, \qquad (10) \\
G_{S^{-1}}\left(\frac{1}{z},\frac{1}{w}\right) &= z^2w^2G_S(z,w)\,, \qquad (11)
\end{aligned}$$

and relate double-dual moments to single-dual moments

$$\begin{aligned}
\tilde{\alpha}_{11}\tilde{\alpha}_2^2 &= -\tilde{\alpha}_3^2 + \tilde{\alpha}_2\tilde{\alpha}_4\,, \\
\tilde{\alpha}_{12}\tilde{\alpha}_2^3 &= \tilde{\alpha}_{21}\tilde{\alpha}_2^3 = 2\tilde{\alpha}_3^3 - 4\tilde{\alpha}_2\tilde{\alpha}_3\tilde{\alpha}_4 + 2\tilde{\alpha}_2^2\tilde{\alpha}_5\,, \\
\tilde{\alpha}_{22}\tilde{\alpha}_2^4 &= 4\tilde{\alpha}_2^3\tilde{\alpha}_6 - 6\tilde{\alpha}_3^4 + 16\tilde{\alpha}_2\tilde{\alpha}_3^2\tilde{\alpha}_4 - 8\tilde{\alpha}_2^2\tilde{\alpha}_3\tilde{\alpha}_5 - 6\tilde{\alpha}_2^2\tilde{\alpha}_4^2\,,
\end{aligned} \tag{12}$$

where

$$\tilde{\alpha}_{i,j} \equiv \alpha_{i,j}^{S^{-1}} = \left\langle \frac{1}{N}\mathrm{tr}\left(S^{-i}\right)\frac{1}{N}\mathrm{tr}\left(S^{-j}\right)\right\rangle_{\mathrm{c}}\,. \tag{13}$$

Now, we use a unitary transformation to write down the unknown "true" covariance matrix as block diagonal: $\Sigma = U\Lambda U^\dagger$, where $\sum_{i=1}^{m_{\max}} n_i = N$ (we guess or choose in another way a value for $m_{\max}$). Now, we write down as one vector the spectral parameters we seek

$$\Theta = (\Lambda_1, \ldots, \Lambda_{m_{\max}}, p_1, \ldots, p_{m_{\max}-1}) \tag{14}$$

with $p_i = \frac{n_i}{N}$.

There exists a conformal mapping between the generating functions for the "true" moments of Σ and the measured moments of S, $M_S(1/x) = \sum_{i=1}\alpha_i^S x^i, M_\Sigma(1/x) = \sum_{i=1}\alpha_i^\Sigma x^i$, namely [9],

$$M_S(z) = M_\Sigma(Z) \tag{15}$$

with the arguments related by

$$Z = \frac{1}{1 + rM_S(z)}\,. \tag{16}$$

By expanding the generating functions in this relation in terms of the moments,

$$\sum_{k=1}^{\infty} \frac{\alpha_k^S}{z^k} = \sum_{k=1}^{\infty} \frac{\alpha_k^\Sigma}{z^k} \left(1 + r \sum_{l=1}^{\infty} \frac{\alpha_l^S}{z^l}\right)^k , \tag{17}$$

and then by collecting like terms, we obtain the relations between the moments themselves

$$\begin{aligned} \alpha_1^S &= \alpha_1^\Sigma\,, \\ \alpha_2^S &= \alpha_2^\Sigma + r\left(\alpha_1^\Sigma\right)^2\,, \\ \alpha_3^S &= \alpha_3^\Sigma + 3r\alpha_1^\Sigma\alpha_2^\Sigma + r^2\left(\alpha_1^\Sigma\right)^3\,, \\ &\dots \end{aligned} \tag{18}$$

These can easily be solved for α_1^Σ, α_2^Σ, α_3^Σ to find the reciprocal relations, which will also be of use.

3. The statistical method for eigeninference

Let us define an infinite-dimensional vector of fluctuations for the ensemble S

$$v_j = \mathrm{tr}S^j - \left\langle \mathrm{tr}S^j \right\rangle = \mathrm{tr}S^j - N\alpha_j^S\,. \tag{19}$$

Now, we utilize a theorem by Bai and Silverstein [10], which states that the statistical distribution of ν_Θ is multidimensional Gaussian with the elements of the dispersion matrix Q_Θ given by the double spectral moments

$$f(v_\Theta) \sim \frac{1}{\det Q_\Theta} \exp\left(-v_\Theta^\dagger Q_\Theta^{-1} v_\Theta\right)\,, \tag{20}$$

$$[Q_\Theta]_{l\,m} = \alpha_{l\,m}^S\,. \tag{21}$$

According to the maximum likelihood principle, the desired estimator Θ maximizes the pdf. This leads to a method for eigeninference, which can be used in practice in the following way:

— Truncate Q_Θ and ν_Θ to a dimension not lower than $2m_{\max} - 1$;

— Express the truncated Q_Θ in terms of α_j^S;

— Express the truncated ν_Θ in terms of α_j^S and of the moments of the measured matrix;
— Express α_j^S in terms α_j^Σ, and those, in turn, in terms of parameters Θ;
— Find Θ by maximizing f or, equivalently, by minimizing

$$g_\Theta = g_\Theta\left(\{\Lambda_i, p_i\}_{i=1}^K, S, r\right) = v_\Theta^\dagger Q_\Theta^{-1} v_\Theta + \ln \det Q_\Theta. \tag{22}$$

The same procedure can be pursued using from the outset the dual moments $\alpha^{S^{-1}}$ instead of α^S, which gives rise to the dual statistical method.

4. The analytical method for eigeninference

By definition,

$$G_\Sigma(z) = \sum_{i=1}^{m_{\max}} \frac{p_i}{z - \Lambda_i} = \frac{1}{z}\left(1 + \sum_{j=1}^{2m_{\max}-1} \frac{\alpha_j^\Sigma}{z^j}\right) = \frac{R(z)}{P(z)},$$

with R and P being polynomials whose coefficients are functions of $\{\Lambda_j\}$ and $\{p_j\}$. Let us write those polynomials explicitly and rearrange the terms

$$1 + \frac{R_1}{z} + \frac{R_2}{z^2} + \ldots + \frac{R_{m_{\max}-1}}{z^{m_{\max}-1}} = \left(1 + \frac{P_1}{z} + \frac{P_2}{z^2} + \ldots + \frac{P_{m_{\max}}}{z^{m_{\max}}}\right) \times \left(1 + \sum_{j=1}^{2m_{\max}-1} \frac{\alpha_j^\Sigma}{z^j}\right). \tag{23}$$

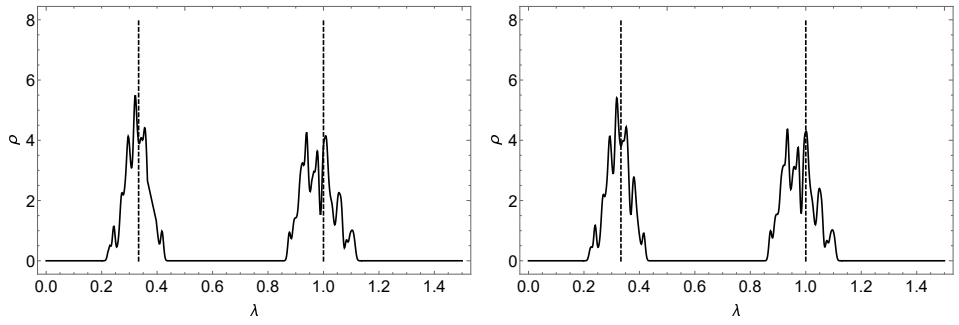
Now, a comparison of coefficients next to each power of z leads to a set of $2m_{\max} - 1$ linear equations for P_j and R_j. We also have to rephrase α_j^Σ in terms of measured moments α_j^S.

Eigenvalues of Σ correspond to zeroes of $P(z)$, and their degeneracies are $p_i = \frac{R(\Lambda_i)}{P'(\Lambda_i)}$.

5. Tests of the methods

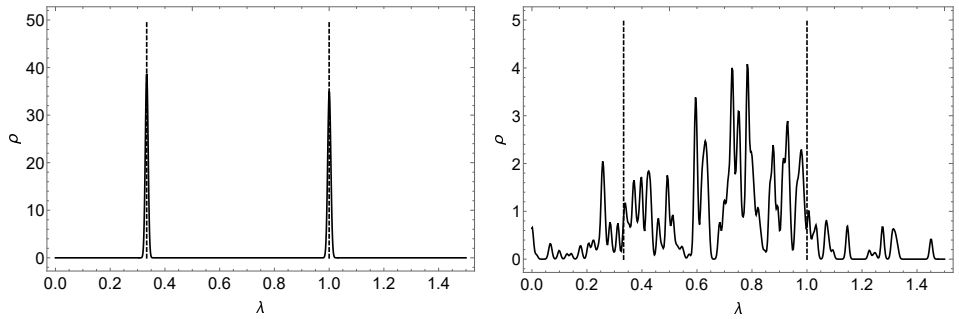
Both the methods were implemented and then tested on samples of random matrices. Some results are presented in Fig. 1. For more tests (including the dual methods and the statistical method with truncation of Q_Θ to size 4×4 or 5×5) and tables with detailed comparison of results, see [7].

Figure 1 shows that the analytical method performed better for small r, as expected, since the covariance matrices were built of longer series of data. It was moreover very fast and robust. The statistical method, however, yielded similar results to the analytical one for large $r = \frac{N}{T}$, but its performance deteriorated with decreasing r, which seems counter-intuitive. This is



(a) $r = 0.8$, analytical method

(b) $r = 0.8$, statistical method (3×3 truncation of Q_Θ)



(c) $r = 0.02$, analytical method

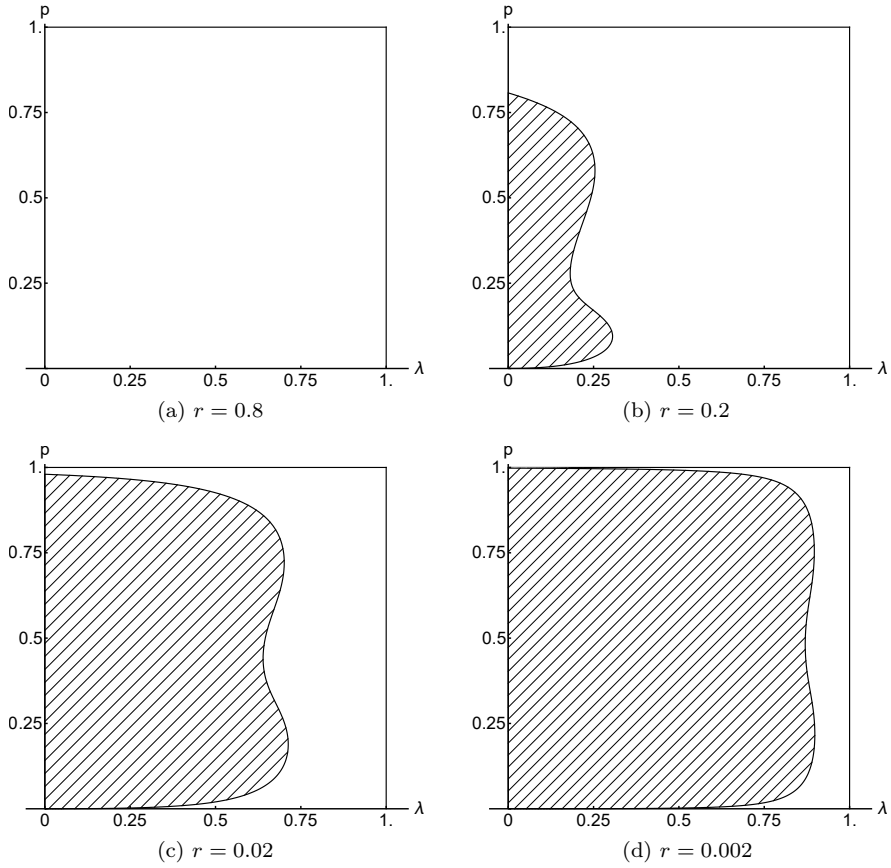
(d) $r = 0.02$, statistical method (3×3 truncation of Q_Θ)

Fig. 1. Estimated spectrum of the covariance matrix. The underlying exact covariance matrix has eigenvalues $\mu_1 = 1/3$, $\mu_2 = 1$ (shown as the dashed lines) with degeneracies $p_1 = 1/2$, $p_2 = 1/2$. The samples consisted of 100 covariance matrices built out of empirical matrices of size either 80×100 ($r = 0.8$) or 80×4000 ($r = 0.02$).

explained by the fact that a *truncated* matrix Q_Θ may cease to be positive-definite for small r, and then the minimization problem is no longer well-defined.

The expressions for the determinant of the matrix Q_Θ are homogeneous in the eigenvalues. Rescaling the eigenvalues by a constant factor so as to set the larger of the two eigenvalues Λ_{large} to 1 reduces the dimensionality of the problem by 1, thus making the two-eigenvalue problem two-dimensional. The only independent variables that remain in this case are the ratio $\Lambda_{\text{large}}/\Lambda_{\text{small}}$ (where Λ_{small} is the smaller eigenvalue) and degeneracy p_{large} of the larger eigenvalue.

In the plots of Fig. 2 the sign of $\det Q_\Theta$ truncated to size 3×3 is plotted for several values of r. It can be seen that as r decreases, the regions of non-positive $\det Q_\Theta$ become larger, filling most of the space for very small r. This causes the statistical method to be usable only if r is large (so that there is no region of non-positivity of $\det Q_\Theta$ in the plot).



(a) $r = 0.8$ (b) $r = 0.2$

(c) $r = 0.02$ (d) $r = 0.002$

Fig. 2. For $\lambda = \Lambda_{\text{small}}/\Lambda_{\text{large}}$ and $p = p_{\text{large}}$ in the meshed region, the determinant of the matrix Q_Θ truncated to size 3×3 is non-positive.

6. Conclusion

Both the methods are best suited to problems with a low number of distinct eigenvalues, whose degeneracies are of the same order.

The analytical method performed well in all the test cases and is very fast. The statistical method gives about as good results as the analytical one for large $r = \frac{N}{T}$ but may break down for small r because of the lack

of positive definiteness of truncated matrix Q_Θ appearing in a minimization problem, which is an inherent part of the method. This fact and the overall faster performance of the analytical method makes the application of the analytical method preferable.

REFERENCES

[1] G. Anderson, *Ann. Math. Statist.* **36**, 1153 (1965).

[2] D. Hydorn, R. Muirhead, *Commun. Statist. — Theory Meth.* **28**, 581 (1999).

[3] X. Mestre, *IEEE Trans. Inf. Theory* **54**, 5113 (2008).

[4] N. Raj Rao, «Applied Stochastic Eigen-Analysis», Ph.D. Thesis, MIT, 2007.

[5] N. Raj Rao, J.A. Mingo, R. Speicher, A. Edelman, *Ann. Statist.* **36**, 2850 (2008).

[6] G. Łukaszewski, «Macierzowe techniki fizyki teoretycznej w telekomunikacji i innych realnych układach złożonych», Ph.D. Thesis, Jagiellonian University, Kraków, 2012 (in Polish).

[7] Z. Drogosz, J. Jurkiewicz, G. Łukaszewski, M.A. Nowak, *Phys. Rev. E* **92**, 022111 (2015).

[8] J. Jurkiewicz, G. Łukaszewski, M.A. Nowak, *Acta Phys. Pol. B* **39**, 799 (2008).

[9] Z. Burda, A. Görlich, A. Jarosz, J. Jurkiewicz, *Physica A* **343**, 295 (2004).

[10] Z.D. Bai, J. Silverstein, *Ann. Probab.* **32**, 533 (2004).