

# A SHORT PROOF OF LEDOIT-PÉCHÉ’S RIE FORMULA FOR COVARIANCE MATRICES

FLORENT BENAYCH-GEORGES

**ABSTRACT.** This is a short proof of Ledoit-Péché’s RIE formula for covariance matrices. The proof is based on the Stein formula, which gives a very simple way to derive the result. One of the advantages of this approach is that it shows that the only really needed hypothesis, for the machinery to work, is that the mean of the eigenvalues of the true covariance matrix and the largest of them have the same order.

**Notation:** For  $M$  a real matrix,  $M'$  denotes the transpose of  $M$ ,  $\|M\|_F$  denotes the Frobenius norm of  $M$  defined at (2) and  $\|M\|$  denotes the operator norm of  $M$  (with respect to the canonical Euclidian norms). For  $Z$  a random variable,  $\mathbb{E} Z$  denotes the expectation of  $Z$ .

## 1. MODEL

Let  $n$  and let  $X \in \mathbb{R}^n$  be a centered Gaussian (column) random vector with covariance  $\Sigma$ . Let

$$\mathbf{X} = (X(1) \quad \cdots \quad X(T)) \in \mathbb{R}^{n \times T}$$

be a collection of independent copies of  $X$ . The problem is to estimate the true covariance  $\Sigma$  out of the *empirical estimator*

$$\mathbf{E} = \frac{1}{T} \mathbf{X} \mathbf{X}'.$$

We are specifically looking for a Rotationally Invariant Estimator<sup>1</sup>, i.e. an estimator of  $\Sigma$  which has the same eigenvectors as  $\mathbf{E}$  (thus differs from it only by the eigenvalues). We want this estimator to be *optimal*, i.e. to be solution of

$$\operatorname{argmin}_{\text{estimators}} \|\text{Estimator} - \Sigma\|_F \tag{1}$$

among the estimators whose eigenvectors are those of the empirical estimator  $\mathbf{E}$ . Here,  $\|\cdot\|_F$  denotes the *Frobenius norm*, i.e. the standard Euclidean norm on matrices:

$$\|M\|_F := \sqrt{\operatorname{Tr} MM'}. \tag{2}$$

---

*Date:* May 2, 2023.

<sup>1</sup>Such estimators arise naturally in the Bayesian framework where  $\Sigma$  has been chosen at random with a distribution of which we only know that it is invariant under the action of the orthogonal group by conjugation.

Let  $\lambda_1, \dots, \lambda_n$  denote the eigenvalues of  $\mathbf{E}$ , with associated eigenvectors  $\mathbf{u}_1, \dots, \mathbf{u}_n$ . It is obvious, by invariance of the Frobenius norm under the action of the orthogonal group by conjugation, that the solution of (1) is the real symmetric matrix with the eigenvectors  $\mathbf{u}_1, \dots, \mathbf{u}_n$  and respective associated eigenvalues  $\mathbf{u}'_k \Sigma \mathbf{u}_k$  ( $1 \leq k \leq n$ ).

It follows that the optimal RIE is the one given by the formula

$$\widehat{\mathbf{E}} := \sum_{k=1}^n \widehat{\lambda}_k \mathbf{u}_k \mathbf{u}'_k \quad \text{for } \widehat{\lambda}_k := \mathbf{u}'_k \Sigma \mathbf{u}_k. \quad (3)$$

Introducing the measures

$$\sum_{k=1}^n \delta_{\lambda_k}$$

and

$$\sum_{k=1}^n \mathbf{u}'_k \Sigma \mathbf{u}_k \delta_{\lambda_k},$$

we can express the  $\widehat{\lambda}_k$  of (3) as Radon-Nikodym derivatives, and, using (10), we have that for any  $\varepsilon > 0$  such that  $[\lambda_k - \varepsilon, \lambda_k + \varepsilon] \cap \{\lambda_1, \dots, \lambda_n\} = \{\lambda_k\}$ ,

$$\widehat{\lambda}_k = \lim_{\eta \rightarrow 0} \frac{\int_{\lambda_k - \varepsilon}^{\lambda_k + \varepsilon} \Im L(x + i\eta) dx}{\int_{\lambda_k - \varepsilon}^{\lambda_k + \varepsilon} \Im G(x + i\eta) dx}, \quad (4)$$

where  $L$  and  $G$  are the meromorphic functions on the complex plane defined by

$$L(z) := \frac{1}{T} \operatorname{Tr} \mathbf{G} \Sigma, \quad G(z) := \frac{1}{T} \operatorname{Tr} \mathbf{G},$$

for  $\mathbf{G} = \mathbf{G}(z) := (z - \mathbf{E})^{-1}$  the resolvent matrix of  $\mathbf{E}$ .

The main results of [5] can be summed-up in the following one (with main consequence Equation (6) below):

**Theorem 1.1.** *For any fixed  $z \in \mathbb{C} \setminus \mathbb{R}$ , we have*

$$L(z) = 1 - \frac{1}{1 - q + zG(z)} + o(1), \quad (5)$$

where  $o(1)$  denotes a random variable tending to zero in probability as  $n, T$  tend to infinity in such a way that  $q := n/T$  stays bounded as well as the operator norm of  $\Sigma$  and  $T/\operatorname{Tr} \Sigma$ .

**Remark 1.2.** Note that as  $z$  tends to  $\lambda$  in (5),

$$\frac{\Im L(z)}{\Im G(z)} \sim \frac{-\Im \frac{1}{1-q+zG(z)}}{\Im G(z)} = \frac{1}{|1-q+zG(z)|^2} \frac{\Im(zG(z))}{\Im G(z)} \sim \frac{\lambda}{|1-q+\lambda G(z)|^2},$$

which, given  $G$  is normalized by  $T$  and not by  $n$ , is the well known RIE formula one can find e.g. in [4]. Note that in the finite dimensional case where this formula is meant to be applied,  $G(z)$  is singular at any eigenvalue  $\lambda$ , hence one has to slightly regularize the denominator, i.e. use  $G(z)$  for  $z = \lambda + i\eta$  with a positive spectral resolution  $\eta = T^{-\alpha}$

(any exponent  $\alpha \in (0, 1)$ , typically  $\alpha = 1/2$ , should work). To sum up, for such a spectral resolution  $\eta$ , the formula of the cleaned eigenvalue  $\widehat{\lambda}$  associated to an eigenvalue  $\lambda$  is given by

$$\widehat{\lambda} = \frac{\lambda}{|1 - q + \lambda G(\lambda + i\eta)|^2}. \quad (6)$$

**Remark 1.3.** One of the advantages of this approach is that it shows that the only really needed hypothesis, for the machinery to work, is that the mean  $\text{Tr } \Sigma/n$  of the eigenvalues of the true covariance matrix and the largest of them (i.e. the operator norm  $\|\Sigma\|$  of  $\Sigma$ ) have the same order.

Let us now show how concentration of measure and the Stein formula for Gaussian random vectors allow to recover the above result very easily. In the case where the random vector  $\mathbf{X}$  is not Gaussian, what follows also works with a bit more of work, using e.g. [3, Lem. 1.13.9] instead of the Stein formula and the McDiarmid inequality instead of Gaussian measure concentration. The idea of the proof is the following one: one starts with the formula  $z\mathbf{G} - I_n = \mathbf{GE}$ , pass to the expectation and expand the RHT using Stein formula for Gaussian vectors. The expansion makes products of traces appear, each of which concentrates around its expectation by the concentration of measure principle. At the end, we obtain the desired relation between  $G(z)$  and  $L(z)$ .

*Proof.* Let

$$H := \frac{1}{T} \text{Tr } \mathbf{GE}.$$

Using Corollary 2.2,

$$\begin{aligned} \mathbb{E} \text{Tr } \mathbf{GE} &= \mathbb{E} \frac{1}{T} \sum_t \text{Tr } \mathbf{GX}(t)X(t)' \\ &= \frac{1}{T} \sum_t \mathbb{E} X(t)' \mathbf{GX}(t) \\ &= \frac{1}{T} \sum_t \mathbb{E} \text{Tr } \mathbf{G}\Sigma + \frac{1}{T} \sum_t \sum_{k=1}^n \mathbb{E} \mathbf{e}'_k \Sigma \left( \frac{\partial}{\partial X(t)_k} \mathbf{G} \right) X(t) \end{aligned}$$

Note that for any  $t$ , at  $X(1), \dots, \widetilde{X(t)}, \dots, X(T)$  fixed, the differential, at  $X(t) \in \mathbb{R}^n$ , of the function  $X(t) \mapsto \mathbf{G}$  is the function

$$x \in \mathbb{R}^n \mapsto \frac{1}{T} \mathbf{G} (X(t)x' + xX(t)') \mathbf{G},$$

so that

$$\frac{\partial}{\partial X(t)_k} \mathbf{G} = \frac{1}{T} \mathbf{G} (X(t)\mathbf{e}'_k + \mathbf{e}_k X(t)') \mathbf{G}$$

and

$$\begin{aligned}
\mathbb{E} \operatorname{Tr} \mathbf{G}\mathbf{E} &= \frac{1}{T} \sum_t \mathbb{E} \operatorname{Tr} \mathbf{G}\Sigma + \frac{1}{T^2} \sum_t \sum_{k=1}^n \mathbb{E} \mathbf{e}'_k \Sigma (\mathbf{G} (X(t)\mathbf{e}'_k + \mathbf{e}_k X(t)') \mathbf{G}) X(t) \\
&= \mathbb{E} \operatorname{Tr} \mathbf{G}\Sigma + \frac{1}{T^2} \sum_t \sum_{k=1}^n \mathbb{E} (\mathbf{e}'_k \Sigma \mathbf{G} X(t) \mathbf{e}'_k \mathbf{G} X(t) + \mathbf{e}'_k \Sigma \mathbf{G} \mathbf{e}_k X(t)' \mathbf{G} X(t)) \\
&= \mathbb{E} \operatorname{Tr} \mathbf{G}\Sigma + \frac{1}{T^2} \sum_t \mathbb{E} (X(t)' \mathbf{G} \Sigma \mathbf{G} X(t) + \operatorname{Tr}(\Sigma \mathbf{G}) X(t)' \mathbf{G} X(t)) \\
&= \mathbb{E} \operatorname{Tr} \mathbf{G}\Sigma + \frac{1}{T} \mathbb{E} \operatorname{Tr} \mathbf{G}\Sigma \mathbf{G}\mathbf{E} + \frac{1}{T} \mathbb{E} \operatorname{Tr}(\Sigma \mathbf{G}) \operatorname{Tr}(\mathbf{G}\mathbf{E})
\end{aligned}$$

Dividing by  $T$  and using Proposition 2.3 with [2, Lem. B.2], we find

$$H = L + LH + o(1). \quad (7)$$

Now, we are close to the conclusion but have to divide by  $1+H$ . Writing  $\mathbf{G} \times (z - \mathbf{E}) = I_n$ , we have

$$z\mathbf{G} - I_n = \mathbf{G}\mathbf{E},$$

so that

$$H = zG - q = \frac{q}{n} \sum_{i=1}^n \frac{\lambda_i}{z - \lambda_i}.$$

Note that for  $x = \Re z$ ,  $\eta = \Im z$ , we have

$$|\Im H| = \frac{q|\eta|}{n} \sum_{i=1}^n \frac{\lambda_i}{(x - \lambda_i)^2 + \eta^2} \geq \frac{|\eta| \operatorname{Tr} \mathbf{E}/T}{2x^2 + 2\|\mathbf{E}\|^2 + \eta^2}. \quad (8)$$

By (8) and Lemma 1.4, there is  $c > 0$  such that with probability tending to 1,  $|1+H| > c$ .

Hence by (7),

$$L = \frac{H}{1+H} + o(1) = 1 - \frac{1}{1+H} + o(1).$$

□

**Lemma 1.4.** *As  $n, T$  tend to infinity in such a way that  $q = n/T$  stays bounded and  $\Sigma$  stays bounded in operator norm, there is a constant  $C > 0$  (depending on the bounds on  $q$  and on  $\|\Sigma\|$ ) such that with probability tending to 1, the operator norm  $\|\mathbf{E}\|$  of  $\mathbf{E}$  and its trace  $\operatorname{Tr} \mathbf{E}$  satisfy*

$$\|\mathbf{E}\| \leq q\|\Sigma\|C \quad \text{and} \quad \operatorname{Tr} \mathbf{E} \geq \operatorname{Tr} \Sigma/C$$

*Proof.* Let  $\mathbf{Y} \in \mathbb{R}^{n \times T}$  be a matrix with independent standard Gaussian entries. Then  $\mathbf{X}$  has the same distribution as  $\Sigma^{1/2}\mathbf{Y}$  and  $T\|\mathbf{E}\|$  has the same distribution as  $\|\Sigma\mathbf{Y}\mathbf{Y}'\|$ . Thus to prove the part about the operator norm, it suffices to prove that there is a universal constant  $C_0$  such that with probability tending to 1,  $\|\mathbf{Y}\mathbf{Y}'\| \leq (n+T)C_0$ . This is a very well known fact, that can be obtained for example as follows: given the eigenvalues of  $\mathbf{Y}\mathbf{Y}'$

are, up to the null eigenvalue, the squares of the eigenvalues of  $\begin{pmatrix} 0 & \mathbf{Y} \\ \mathbf{Y}' & 0 \end{pmatrix}$ , it follows from [3, Th. 1.13.17]. The part about the trace follows from the fact that  $\text{Tr } \mathbf{E}$  has the same law as  $T^{-1} \text{Tr } \Sigma \mathbf{Y} \mathbf{Y}'$  which has expectation  $\text{Tr } \Sigma$  and variance  $\leq 2T^{-1} \text{Tr } \Sigma^2$ .  $\square$

## 2. APPENDIX

**2.1. Stieltjes transform inversion.** Any signed measure  $\mu$  on  $\mathbb{R}$  can be recovered out of its Stieltjes transform

$$g_\mu(z) := \int \frac{d\mu(t)}{z - t}, \quad z \in \mathbb{C} \setminus \mathbb{R} \quad (9)$$

by the formula

$$\mu = -\frac{1}{\pi} \lim_{\eta \rightarrow 0^+} (\Im g_\mu(x + i\eta) dx), \quad (10)$$

where the limit holds in the weak topology (see e.g. [1, Th. 2.4.3] and use the decomposition of any signed measure as a difference of finite positive measures).

## 2.2. Stein formula for real Gaussian random vectors.

**Proposition 2.1.** *Let  $X = (X_1, \dots, X_d)$  be a real centered Gaussian vector with covariance  $\Sigma$  and  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a  $C^1$  function with gradient having at most polynomial growth at infinity. Then for all  $i_0 = 1, \dots, d$ ,*

$$\mathbb{E} X_{i_0} f(X_1, \dots, X_d) = \sum_{k=1}^d \Sigma_{i_0 k} \mathbb{E} (\partial_k f)(X_1, \dots, X_d).$$

**Corollary 2.2.** *With the same notation, considering  $X$  as a column vector, for  $F : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  a matrix-valued function with gradient having at most polynomial growth at infinity. we have*

$$\mathbb{E} X' F(X) X = \text{Tr } \Sigma \mathbb{E} F(X) + \sum_{k=1}^d (\mathbb{E} \Sigma (\partial_k F)(X) X)_k. \quad (11)$$

*Proof.* We have, by Proposition 2.1,

$$\begin{aligned} \mathbb{E} X' F(X) X &= \sum_{ij} \mathbb{E} X_i X_j F(X)_{ij} \\ &= \sum_{ijk} \mathbb{E} \Sigma_{ik} \frac{\partial}{\partial X_k} X_j F(X)_{ij} \\ &= \sum_{ijk} \mathbb{E} \Sigma_{ik} (\delta_{j=k} F(X)_{ij} + X_j (\partial_k F)(X)_{ij}) \\ &= \text{Tr } \Sigma \mathbb{E} F(X) + \sum_{ijk} \mathbb{E} \Sigma_{ik} X_j (\partial_k F)(X)_{ij} \end{aligned}$$

$$= \text{Tr} \Sigma \mathbb{E} F(X) + \sum_k (\mathbb{E} \Sigma(\partial_k F)(X) X)_k$$

□

**2.3. Concentration of measure for Gaussian vectors.** The following proposition can be found e.g. in [1, Sec. 4.4.1] or [6, Th. 5.2.2].

**Proposition 2.3.** *Let  $X = (X_1, \dots, X_d)$  be a standard real Gaussian vector and  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a  $\mathcal{C}^1$  function with gradient  $\nabla f$ . Then we have*

$$\text{Var}(f(X)) \leq \mathbb{E} \|\nabla f(X)\|^2, \quad (12)$$

where  $\|\cdot\|$  denotes the standard Euclidean norm.

Besides, if  $f$  is  $k$ -Lipschitz, then for any  $t > 0$ , we have

$$\mathbb{P}(|f(X) - \mathbb{E} f(X)| \geq t) \leq 2e^{-\frac{t^2}{2k^2}}, \quad (13)$$

i.e.  $f(X) - \mathbb{E} f(X)$  is Sub-Gaussian with Sub-Gaussian norm  $\leq k$ , up to a universal constant factor.

## REFERENCES

- [1] Anderson, G., Guionnet, A., Zeitouni, O. *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics, 118 (2009).
- [2] Benaych-Georges, F., Couillet, R. *Spectral analysis of the Gram matrix of mixture models*, ESAIM Probab. Statist., Vol. 20 (2016), 217–237.
- [3] Benaych-Georges, F., Knowles, A. *Local semicircle law for Wigner matrices*. Advanced topics in random matrices, 1–90, Panor. Synthèses, 53, Soc. Math. France, Paris, 2017.
- [4] Bun, J., Bouchaud, J.-P., Potters, M. *Cleaning large correlation matrices: Tools from Random Matrix Theory*, Physics Reports Volume 666, Review article, 1–109, 2017.
- [5] Ledoit, O., Péché, S. *Eigenvectors of some large sample covariance matrix ensembles* Probability Theory and Related Fields, 2011, 151.1, 233–264
- [6] Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Series in Statistical and Probabilistic Mathematics, 2018.

CFM, 23 RUE DE L'UNIVERSITÉ, 75007 PARIS, FRANCE

Email address: florent.benaych@gmail.com