数理科学実践研究レター　2023–18 October 11, 2023

A Generalization of Rotationally Invariant Estimator in Operator Algebras

by

朱 浩哲

UNIVERSITY OF TOKYO
GRADUATE SCHOOL OF MATHEMATICAL SCIENCES
KOMABA, TOKYO, JAPAN

A Generalization of Rotationally Invariant Estimator in Operator Algebras

Haozhe Zhu[1] (Graduate School of Mathematical Sciences, the University of Tokyo)

Abstract

In this paper, we investigate the rotationally invariant estimators of the population matrix **C** knowing the sample covariance matrix **E**. We give a possible generalization of rotationally invariant estimator from an operator algebraic point of view, and deduce the same formulae as in random matrix theory.

作用素環における回転不変推定量の一般化

朱浩哲 (東京大学数理科学研究科)

概要

本研究では、作用素環の観点から回転不変推定量の一般化の可能性を提案し、ランダム行列理論における同じ式を導いた。

1 Introduction

In many, very different fields such as quantitative finance, an important problem is to extract a true signal from noisy observations in extremely large data sets. More precisely, consider an empirical dataset $\{x_{i,t}\}_{1\le i\le N, 1\le t\le T}$ comprising the data of T observations of N quantities with $\frac{1}{T}\sum_{t=1}^{T} x_{i,t} = 0$, and consider the sample covariance matrix $\mathbf{E} = (E_{ij})_{1\le i,j\le N}$ defined by

$$E_{ij} = \frac{1}{T}\sum_{t=1}^{T} x_{i,t}x_{j,t}. \tag{1}$$

It is apparent that the sample matrix **E** is a noisy version of the "true" covariance matrix **C** that we aim to estimate (which might not even exist). For example, when building a portfolio, one may regard $x_{i,t}$ as daily returns of the i-th stock at time t, and estimates the "true" covariance matrix **C** based on **E** to predict future returns and minimum the risk.

Intuitively, when N is fixed and T is sufficiently large, i.e., $q = \frac{N}{T} \to 0$, the matrix **E** will be very close to **C**. However, in the general scenario where both N and T tend to infinity with a fixed ratio $q = \frac{N}{T}$, even if **C** exists, **E** can differ significantly from **C**. The well-known Kolmogorov limit implies that classical statistical tools are inadequate when dealing with this setting, and we need to study these covariance matrices of large size from another point of view, that is, the random matrix theory introduced by John Wishart. Roughly speaking, random matrix theory attempts to give universal properties of large random matrices through the spectral measure (or the density of eigenvalues), regardless of the specific form of the matrix. Some financial examples and applications of random matrix theory can be found in [2].

To construct estimators of **C** without prior knowledge of its properties, additional assumptions about **C** are necessary. A natural assumption is that the estimator $\Xi(\mathbf{E})$ is rotationally invariant, that is,

$$\Xi(\mathbf{O}\mathbf{E}\mathbf{O}^{\mathsf{T}}) = \mathbf{O}\Xi(\mathbf{E})\mathbf{O}^{\mathsf{T}}. \tag{2}$$

for any rotation matrix **O**. As mentioned before, **E** is a noisy version of **C**. There are two important cases:

- $\mathbf{E} = \mathbf{C} + \mathbf{X}$ (additive noise case),

[1] zhuhz@ms.u-tokyo.ac.jp

- $\mathbf{E} = \mathbf{C}^{1/2}\mathbf{X}\mathbf{C}^{1/2}$ (multiplicative noise case),

where $\mathbf{X}$ is the noise matrix and $\mathbf{C}$ and $\mathbf{X}$ are freely independent. In [3], the formulae of the optimal rotationally invariant estimator (RIE) in the above two cases (see Section 3 below) are derived by investigating the "overlaps" between random matrices $\mathbf{E}$ and $\mathbf{C}$.

In this paper, we study the rotationally invariant estimator $\Xi(\mathbf{E})$ of $\mathbf{C}$ within the framework of tracial von Neumann algebras, and deduce the same results as in [3]. In Section 2, we will first recall some important definitions and results in free probabilty theory, and a breif introduction of RIE of random matrices is given in Section 3. We will be discussing how to generalize rotationally invariant estimators from an operator algebraic point of view in Section 4.

2 Noncommutative probabilty space

In fact, the algebra of random matrices can be viewed as a special case of noncommutative probability space.

Definition 2.1 *A noncommutative probability space (A, φ) is a unital algebra A over $\mathbb{C}$ together with a unital linear functional $\varphi\colon A \to \mathbb{C}$.*

The following table shows the comparison between classical/free probability theory and random matrices.

classical probability	free probability	random matrices
probability space $(X, \Omega, \mathbb{P})$	noncommutative probability space (A, φ)	(A_N, φ_N)
random variable $f\colon \Omega \to \mathbb{C}$	noncommutative random variable $a \in A$	$N \times N$ matrix of random variables
expecatation $\mathbb{E}[f] = \int_\Omega f \mathrm{d}\mathbb{P}$	$\varphi(a)$	$\varphi_N(a) = \mathrm{tr} \otimes \mathbb{E}(a)$
independence	free independence	
Gaussian distribution	semi-circle	GUE

Here $(A_N, \varphi_N) = (\mathrm{M}_N(\mathbb{C}) \otimes L^{-\infty}(\Omega), \mathrm{tr} \otimes \mathbb{E})$ is the algebra of $N \times N$ random matrices, and $L^{-\infty}(\Omega) = \cap_{1 \le p < \infty} L^p(\Omega)$.

For self-adjoint $a \in A$, we have the following series.

- $G_a(z) = (z - a)^{-1}$ and Stieltjes transform $\mathfrak{g}_a(z) = \varphi(G_a(z)) = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \varphi(a^k)$;
- $T_a(z) = a(z - a)^{-1}$ and T-transform $\mathfrak{t}_a(z) = \varphi\left(a(z - a)^{-1}\right)$;
- R-transform $R_a(z) = \mathfrak{z}_a(z) - \frac{1}{z}$;
- S-transform $S_a(z) = \frac{z+1}{z\zeta_a(z)}$ (when $\varphi(a) \neq 0$).

Here $\mathfrak{z}_a(z)$ and $\zeta_a(z)$ are inverse functions of $\mathfrak{g}_a(z)$ and $\mathfrak{t}_a(z)$ (as formal series) respectively. Voiculescu's free probability theory gives a remarkable result of R- and S-transforms, stating that if $a, b \in A$ are positive and freely independent, then we have

$$R_{a+b}(z) = R_a(z) + R_b(z), \quad S_{ab}(z) = S_a(z) S_b(z). \tag{3}$$

Remark 2.1 Equivalently, the R-transform can be defined in terms of free cumulants as

$$R_a(z) = \sum_{n=0}^{\infty} \kappa_{n+1}^a z^n, \tag{4}$$

where κ_k^a is the k-th free cumulant of a. Actually, when a and b are self-adjoint and free, we have $\kappa_k^a + \kappa_k^b = \kappa_k^{a+b}$, from which the equality $R_{a+b} = R_a + R_b$ in (3) is deduced. Moreover, by using (4) and the definition of free cumulants, it can be shown that

$$\mathfrak{g}_{a+b}(z) R_a\left(\mathfrak{g}_{a+b}(z)\right) = \varphi\left(a\left(z - (a+b)\right)^{-1}\right). \tag{5}$$

The proof is cumbersome but is similar in spirit to that of the equivalence of the above two definitions of R-transform (for example, Theorem 5.2 [5]), as the case $b = 0$ is just the same as the fisrt definition of R-transform $R_a(z) = \mathfrak{z}_a(z) - \frac{1}{z}$.

We will not focus on the analytic properties of these series. More details of noncommutative probability spaces and R-/S-transforms can be found in [4] and [5].

3 Random matrices case: a brief review

In this section, we briefly review the results of RIE in [3] in terms of random matrices.

Let $\mathbf{E}$ and $\mathbf{C}$ be two self-adjoint $N \times N$ random matrices with N being sufficiently large. Let $c_1 \geq c_2 \geq \cdots \geq c_N$ (resp. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$) be the eigenvalues of $\mathbf{C}$ (resp. $\mathbf{E}$), and denote by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ (resp. $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N$) the corresponding normalized eigenvectors. The definition (2) implies that rotationally invariant estimator $\Xi(\mathbf{E})$ of $\mathbf{C}$ has the same eigenvectors as those of $\mathbf{E}$, namely

$$\Xi(\mathbf{E}) = \sum_{i=1}^{N} \xi_i \mathbf{u}_i \mathbf{u}_i^{\mathsf{T}}, \tag{6}$$

where ξ_i is a function of empirical eigenvalues $\{\lambda_i\}$. We are interested in the optimal choice of ξ_i, i.e., the RIE $\hat{\Xi}(\mathbf{E})$ that minimizes the following Euclidean norm (least-square error):

$$\hat{\Xi}(\mathbf{E}) = \underset{\Xi(\mathbf{E})}{\arg\min} \|\Xi(\mathbf{E}) - \mathbf{C}\|_{L^2} = \underset{\Xi(\mathbf{E})}{\arg\min} \operatorname{Tr}(\Xi(\mathbf{E}) - \mathbf{C})^2. \tag{7}$$

Apparently, the solution is $\hat{\Xi}(\mathbf{E}) = \sum_{i=1}^{N} \hat{\xi}_i \mathbf{u}_i \mathbf{u}_i^{\mathsf{T}}$, with

$$\hat{\xi}_i = \mathbf{u}_i \mathbf{C} \mathbf{u}_i^{\mathsf{T}} = \sum_{j=1}^{N} c_j \left(\mathbf{u}_i^{\mathsf{T}} \mathbf{v}_j\right)^2. \tag{8}$$

The so-called *self-averaging* property of $\hat{\xi}_i$ (for specific discussion, see [3] II-B or [4] chapter 2.1) yeilds

$$\hat{\xi}_i \approx \sum_{j=1}^{N} c_j \mathbb{E}\left[\left(\mathbf{u}_i^{\mathsf{T}} \mathbf{v}_j\right)^2\right], \tag{9}$$

so it suffies to calculate the *overlap* $\left(\mathbf{u}_i^{\mathsf{T}} \mathbf{v}_j\right)^2$, or equivalently, to calculate

$$O(\lambda_i, c_j) \stackrel{\text{def}}{=} N\mathbb{E}\left[\left(\mathbf{u}_i^{\mathsf{T}} \mathbf{v}_j\right)^2\right]. \tag{10}$$

If we denote by $\rho_{\mathbf{E}}$ the eigenvalue desity of $\mathbf{E}$, then it is shown that ([3] II.8 and IV) in the large N limit,

$$\operatorname{Im} \mathbf{v}_j^{\mathsf{T}} G_{\mathbf{E}}(\lambda - \mathrm{i}\eta) \mathbf{v}_j \approx \pi \rho_{\mathbf{E}}(\lambda) O(\lambda, c_j), \quad \text{for } 1 \gg \eta \gg N^{-1}, \tag{11}$$

and then

$$\hat{\xi}_i = \sum_{j=1}^{N} c_j \left(\mathbf{u}_i^\top \mathbf{v}_j\right)^2 = \frac{1}{N\pi\rho_{\mathbf{E}}(\lambda_i)} \lim_{\eta\to 0+} \operatorname{Im}\operatorname{Tr}(\mathbf{C}G_{\mathbf{E}}(\lambda_i - \mathrm{i}\eta)). \tag{12}$$

In the additive noise cases, the replica method indicates that $\mathbb{E}[G_{\mathbf{E}}(z)] = G_{\mathbf{C}}(z - R_{\mathbf{X}}(\mathfrak{g}_{\mathbf{E}}(z)))$. Finally, we deduce, as IV.1 and IV.2 in [3],

$$\hat{\xi}_i = \lambda_i - \frac{\lim_{\eta\to 0+} \operatorname{Im}\left(R_{\mathbf{X}}(\mathfrak{g}_{\mathbf{E}}(\lambda_i - \mathrm{i}\eta))\, \mathfrak{g}_{\mathbf{E}}(\lambda_i - \mathrm{i}\eta)\right)}{\pi\rho_{\mathbf{E}}(\lambda_i)}. \tag{13}$$

Similarly, in the multiplicative noise cases (IV.6 and IV.7),

$$\hat{\xi}_i = \frac{\lim_{\eta\to 0+} \operatorname{Im}\left(S_{\mathbf{X}}(\mathfrak{t}_{\mathbf{E}}(\lambda_i - \mathrm{i}\eta))\, \mathfrak{t}_{\mathbf{E}}(\lambda_i - \mathrm{i}\eta)\right)}{\pi\rho_{\mathbf{E}}(\lambda_i)}. \tag{14}$$

4 Optimal RIE in operator algebraic setting

Now we would like to study what (6) and (7) mean in a tracial von Neumann algebra. Actually, the definition (2) of RIE and (6) argue that if the rotationally invariant estimator $\Xi(\mathbf{E})$ of $\mathbf{C}$ can be defined, it should be a certain functional calculus of $\mathbf{E}$. Together with (7), it is natural to ask the following question.

Question 4.1 *Let (A, φ) be a tracial von Neumann algebra with a normal faithful tracial state φ, and let $\mathbf{C}, \mathbf{E} \in A$ be two positive elements. Find*

$$\hat{\Xi}(\mathbf{E}) = \operatorname*{arg\,min}_{f(\mathbf{E})} \|f(\mathbf{E}) - \mathbf{C}\|_{L^2}$$

over all possible bounded Borel functional calculus $f(\mathbf{E})$ of $\mathbf{E}$.

The positivity of $\mathbf{E}$ and $\mathbf{C}$ is due to the fact that covariance matrices are positive semi-definite. The answer to the above question is quite clear thanks to the existence of contional expectation in a tracial von Neumann algebra.

Theorem 1 *Let (A, φ) be a von Neumann algebra with a normal faithful tracial state φ, and let $B \subset A$ be a von Neumann subalgebra. There exists a unique trace preserving linear map $E_B \colon A \to B$ such that*
(1) $E_B(A_+) \subset B_+$;
(2) $E_B(b) = b$ for $b \in B$;
(3) $E_B(b_1 a b_2) = b_1 E_B(a) b_2$ for $b_1, b_2 \in B$ and $a \in A$.

The proof of this result is not provided here, but we refer to Theorem 9.1.2 in [1] for a complete proof. Indeed, it is shown that the conditional expectationE_B can be defined as the restriction of the projection $e_B \colon L^2(A, \varphi) \to L^2(B, \varphi|_B)$ to $A \subset L^2(A, \varphi)$. Consequently, for any element $x \in A$, the $E_B(x)$ corresponds to the unique element in B that minimizes the L^2-distance between x and B. It suffices to set $\hat{\Xi}(\mathbf{E}) = E_B(\mathbf{C})$, where B is the von Neumann subalgebra generated by $\mathbf{E}$. However, we still need to determine the function f or, equivalently, to find the moments of $E_B(\mathbf{C}) = f(\mathbf{E})$. While the approach is based on Chapter 19 of [4], we are considering a more general setting within operator algebraic framework.

Put $F(z) = \varphi\left(\mathbf{C}(z - \mathbf{E})^{-1}\right)$. The trace preserving property of E_B yields

$$F(z) = \varphi\left(\mathbf{C}(z - \mathbf{E})^{-1}\right) = \varphi\left(f(\mathbf{E})(z - \mathbf{E})^{-1}\right)$$

We denote by $\rho_{\mathbf{E}}$ the spectral measure of $\mathbf{E}$. Then we can recover f by using the Stieltjes inversion formula.

$$\pi\rho_{\mathbf{E}}(\lambda) f(\lambda) = \lim_{\eta\to 0+} \operatorname{Im}\varphi\left(\mathbf{C}(\lambda - \mathrm{i}\eta - \mathbf{E})^{-1}\right). \tag{15}$$

In the additive noise case $\mathbf{E} = \mathbf{C} + \mathbf{X}$, according to (5), we have

$$\begin{aligned}\varphi\left(\mathbf{C}(z-\mathbf{E})^{-1}\right) &= \mathfrak{g}_{\mathbf{E}}(z) R_{\mathbf{C}}\left(\mathfrak{g}_{\mathbf{E}}(z)\right) \\ &= \mathfrak{g}_{\mathbf{E}}(z) R_{\mathbf{E}}\left(\mathfrak{g}_{\mathbf{E}}(z)\right) - \mathfrak{g}_{\mathbf{E}}(z) R_{\mathbf{X}}\left(\mathfrak{g}_{\mathbf{E}}(z)\right) \\ &= -1 + \left(z - R_{\mathbf{X}}\left(\mathfrak{g}_{\mathbf{E}}(z)\right)\right) \mathfrak{g}_{\mathbf{E}}(z).\end{aligned}$$

Note that $\lim_{\eta\to 0+} \operatorname{Im} \mathfrak{g}_{\mathbf{E}}(\lambda - \mathrm{i}\eta) = \pi\rho_{\mathbf{E}}(\lambda)$. We immediately get

$$\pi\rho_{\mathbf{E}}(\lambda) f(\lambda) = \pi\lambda\rho_{\mathbf{E}}(\lambda) - \lim_{\eta\to 0+} \operatorname{Im}\left(R_{\mathbf{X}}\left(\mathfrak{g}_{\mathbf{E}}(\lambda - \mathrm{i}\eta)\right)\mathfrak{g}_{\mathbf{E}}(\lambda - \mathrm{i}\eta)\right).$$

The multiplicative noise case can be deduced similarly:

$$\pi\rho_{\mathbf{E}}(\lambda) f(\lambda) = \lim_{\eta\to 0+} \operatorname{Im}\left(S_{\mathbf{X}}\left(\mathfrak{t}_{\mathbf{E}}(\lambda - \mathrm{i}\eta)\right)\mathfrak{t}_{\mathbf{E}}(\lambda - \mathrm{i}\eta)\right).$$

Then we have the following theorem.

Theorem 2 *Let $\rho_{\mathbf{E}}$ be the compactly supported measure on $\mathbb{R}$ given by the eigenvalue distribution of $\mathbf{E}$ and let f be that of $\hat{\Xi}(\mathbf{E})$.*
(1) In the additive noise case $\mathbf{E} = \mathbf{C} + \mathbf{X}$, we have

$$f(\lambda) = \lambda - \frac{\lim_{\eta\to 0+} \operatorname{Im}\left(R_{\mathbf{X}}\left(\mathfrak{g}_{\mathbf{E}}(\lambda - \mathrm{i}\eta)\right)\mathfrak{g}_{\mathbf{E}}(\lambda - \mathrm{i}\eta)\right)}{\pi\rho_{\mathbf{E}}(\lambda)}. \tag{16}$$

(2) In the multiplicative noise case $\mathbf{E} = \mathbf{C}^{1/2}\mathbf{X}\mathbf{C}^{1/2}$, we have

$$f(\lambda) = \frac{\lim_{\eta\to 0+} \operatorname{Im}\left(S_{\mathbf{X}}\left(\mathfrak{t}_{\mathbf{E}}(\lambda - \mathrm{i}\eta)\right)\mathfrak{t}_{\mathbf{E}}(\lambda - \mathrm{i}\eta)\right)}{\pi\rho_{\mathbf{E}}(\lambda)}. \tag{17}$$

Remark 4.1 In the general setting of operator algebras, unlike random matrices, the equalities (8)-(12) may not make sense since not all points in the spectrum of $\mathbf{C}$ (or $\mathbf{E}$) are eigenvalues. Whether the calculation of $\mathbb{E}\left[G_{\mathbf{E}}(z)\right] = G_{\mathbf{C}}\left(z - R_{\mathbf{X}}\left(\mathfrak{g}_{\mathbf{E}}(z)\right)\right)$ in the additive noise case based on the replica method makes sense or not in a general tracial von Neumann algebra is not that clear.

So far, we have investigated what (optimal) rotationally invariant estimators should be in a tracial von Neumann algebra, and deduced the same fomulae as in [3]. Besides, some properties of optimal rotationally invariant estimators can be easily understood when regarding $\hat{\Xi}(\mathbf{E})$ as the conditional expectation of $\mathbf{C}$. For example, the following inequality

$$\mathrm{Tr}\ \hat{\Xi}(\mathbf{E})^2 \leq \mathrm{Tr}\ \mathbf{C}^2 \tag{18}$$

is an immediate result of $E_B(x)^* E_B(x) \leq E_B(x^* x)$.

Acknowledgement The author would like to express his thanks and appreciation to Yano Ryosuke, Yuta Kimura and Yuichiro Tanaka for their guidance and support during the study and research of WINGS-FMSP program.

References

[1] Claire Anantharaman and Sorin Popa. An introduction to II_1 factors. *preprint*, 8, 2017.

[2] Jean-Philippe Bouchaud and Marc Potters. Financial applications of random matrix theory: A short review. *arXiv preprint arXiv:0910.1205*, 2009.

[3] Joël Bun, Romain Allez, Jean-Philippe Bouchaud, and Marc Potters. Rotational invariant estimator for general noisy matrices. *IEEE Transactions on Information Theory*, 62(12):7475–7490, 2016.

[4] Marc Potters and Jean-Philippe Bouchaud. *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists*. Cambridge University Press, 2020.

[5] Dan-Virgil Voiculescu, Nicolai Stammeier, and Moritz Weber. *Free Probability and Operator Algebras*. European Mathematical Society, 2016.