Journal of Multivariate Analysis 174 (2019) 104531

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Random matrix-improved estimation of covariance matrix distances[☆]

Romain Couillet [a,b,*], Malik Tiomoko [a], Steeve Zozor [a], Eric Moisan [a]

[a] GIPSA-lab, University Grenoble Alpes, France
[b] L2S, CentraleSupélec, University of Paris Saclay, France

ARTICLE INFO

Article history:
Received 11 February 2019
Received in revised form 27 June 2019
Accepted 27 June 2019
Available online 22 July 2019

AMS 2010 subject classifications:
primary 60B20
secondary 62M45

Keywords:
Random matrix theory
Covariance estimation
Distances and divergences

ABSTRACT

Given two sets $x_1^{(1)}, \ldots, x_{n_1}^{(1)}$ and $x_1^{(2)}, \ldots, x_{n_2}^{(2)} \in \mathbb{R}^p$ (or $\mathbb{C}^p$) of random vectors with zero mean and positive definite covariance matrices C_1 and $C_2 \in \mathbb{R}^{p\times p}$ (or $\mathbb{C}^{p\times p}$), respectively, this article provides novel estimators for a wide range of distances between C_1 and C_2 (along with divergences between some zero mean and covariance C_1 or C_2 probability measures) of the form $\frac{1}{p}\sum_{i=1}^{n} f(\lambda_i(C_1^{-1}C_2))$ (with $\lambda_i(X)$ the eigenvalues of matrix X). These estimators are derived using recent advances in the field of random matrix theory and are asymptotically consistent as $n_1, n_2, p \to \infty$ with non trivial ratios $p/n_1 < 1$ and $p/n_2 < 1$ (the case $p/n_2 > 1$ is also discussed). A first "generic" estimator, valid for a large set of f functions, is provided under the form of a complex integral. Then, for a selected set of atomic functions f which can be linearly combined into elaborate distances of practical interest (namely, $f(t) = t$, $f(t) = \ln(t)$, $f(t) = \ln(1 + st)$ and $f(t) = \ln^2(t)$), a closed-form expression is provided. Besides theoretical findings, simulation results suggest an outstanding performance advantage for the proposed estimators when compared to the classical "plug-in" estimator $\frac{1}{p}\sum_{i=1}^{n} f(\lambda_i(\hat{C}_1^{-1}\hat{C}_2))$ (with $\hat{C}_a = \frac{1}{n_a}\sum_{i=1}^{n_a} x_i^{(a)} x_i^{(a)\top}$), and this even for very small values of n_1, n_2, p. A concrete application to kernel spectral clustering of covariance classes supports this claim.

1. Introduction

In a host of statistical signal processing and machine learning methods, distances between covariance matrices are regularly sought for. These are notably exploited to estimate centroids or distances between clusters of data vectors mostly distinguished through their second order statistics. We may non exhaustively cite brain graph signal processing and machine learning (from EEG datasets in particular) which is a field largely rooted in these approaches [12], hyperspectral and synthetic aperture radar (SAR) clustering [7,25], patch-based image processing [15], etc. For random independent p-dimensional real or complex data vectors $x_1^{(a)}, \ldots, x_{n_a}^{(a)}$, $a \in \{1, 2\}$, having zero mean and covariance matrix C_a, and for a distance (or divergence) $D(X, Y)$ between covariance matrices X and Y (or probability measures associated to random variables with these covariances), the natural approach is to estimate $D(C_1, C_2)$ through the "plug-in" substitute $D(\hat{C}_1, \hat{C}_2)$. For well-behaved functions D, this generally happens to be a consistent estimator as $n_1, n_2 \to \infty$ in the sense that

☆ Couillet's work is supported by ANR RMT4GRAPH (ANR-14-CE28-0006) and IDEX GSTATS Chair at University Grenoble Alpes.
* Corresponding author at: GIPSA-lab, University Grenoble Alpes, France.
E-mail addresses: romain.couillet@gipsa-lab.grenoble-inp.fr (R. Couillet), malik.tiomoko@lss.centralesupelec.fr (M. Tiomoko), steeve.zozor@gipsa-lab.grenoble-inp.fr (S. Zozor), eric.moisan@gipsa-lab.grenoble-inp.fr (E. Moisan).

https://doi.org/10.1016/j.jmva.2019.06.009
0047-259X/

$D(\hat{C}_1, \hat{C}_2) \to D(C_1, C_2)$ almost surely as $n_1, n_2 \to \infty$ while p remains fixed. This is particularly the case for all subsequently introduced distances and divergences.

However, in many modern applications, one cannot afford large n_a ($a \in \{1, 2\}$) values or, conversely, p may be commensurable, if not much larger, than the n_a's. When $n_a, p \to \infty$ in such a way that p/n_a remains away from zero and infinity, it has been well documented in the random matrix literature, starting with the seminal works of Marčenko and Pastur [17], that the operator norm $\|\hat{C}_a - C_a\|$ no longer vanishes. Consequently, this entails that the aforementioned estimator $D(\hat{C}_1, \hat{C}_2)$ for $D(C_1, C_2)$ is likely inconsistent as $p, n_1, n_2 \to \infty$ at a commensurable rate.

This said, it is now interesting to note that many standard matrix distances $D(C_1, C_2)$ classically used in the literature can be written under the form of functionals of the eigenvalues of $C_1^{-1}C_2$ (assuming at least C_1 is invertible). A first important example is the squared Fisher distance between C_1 and C_2 [8] given by

$$D_{\mathrm{F}}(C_1, C_2)^2 = \frac{1}{p}\left\|\ln(C_1^{-\frac{1}{2}}C_2C_1^{-\frac{1}{2}})\right\|_F^2 = \frac{1}{p}\sum_{i=1}^{p}\ln^2(\lambda_i(C_1^{-1}C_2))$$

with $\|\cdot\|_F$ the Frobenius norm, $X^{\frac{1}{2}}$ the nonnegative definite square root of X, and $\ln(X) \equiv U\ln(\Lambda)U^\top$ for symmetric $X = U\Lambda U^\top$ (and $(\cdot)^\top$ the matrix transpose) in its spectral composition. This estimator arises from information geometry and corresponds to the length of the geodesic between C_1 and C_2 in the manifold of positive definite matrices.

Another example is the Bhattacharyya distance [6] between two real Gaussian distributions with zero mean and covariances C_1 and C_2, respectively (i.e., $\mathcal{N}(0, C_1)$ and $\mathcal{N}(0, C_2)$), which reads

$$D_{\mathrm{B}}(C_1, C_2) = \frac{1}{2p}\ln\det\left(\frac{1}{2}[C_1 + C_2]\right) - \frac{1}{4p}\ln\det C_1 - \frac{1}{4p}\ln\det C_2$$

which can be rewritten under the form

$$\begin{aligned} D_{\mathrm{B}}(C_1, C_2) &= \frac{1}{2p}\ln\det(I_p + C_1^{-1}C_2) - \frac{1}{4p}\ln\det(C_1^{-1}C_2) - \frac{1}{2}\ln(2) \\ &= \frac{1}{2p}\sum_{i=1}^{p}\ln(1 + \lambda_i(C_1^{-1}C_2)) - \frac{1}{4p}\sum_{i=1}^{p}\ln(\lambda_i(C_1^{-1}C_2)) - \frac{1}{2}\ln(2). \end{aligned}$$

In a similar manner, the Kullback–Leibler divergence [5] of the Gaussian distribution $\mathcal{N}(0, C_2)$ with respect to $\mathcal{N}(0, C_1)$ is given by

$$D_{\mathrm{KL}} = \frac{1}{2p}\mathrm{tr}\,(C_1^{-1}C_2) - \frac{1}{2} + \frac{1}{2p}\ln\det(C_1^{-1}C_2) = \frac{1}{2p}\sum_{i=1}^{p}\lambda_i(C_1^{-1}C_2) - \frac{1}{2} + \frac{1}{2p}\sum_{i=1}^{p}\ln(\lambda_i(C_1^{-1}C_2)).$$

More generally, the Rényi divergence [4] of $\mathcal{N}(0, C_2)$ with respect to $\mathcal{N}(0, C_1)$ reads, for $\alpha \in \mathbb{R} \setminus \{1\}$,

$$D_{\mathrm{R}}^{\alpha} = -\frac{1}{2(\alpha-1)}\frac{1}{p}\sum_{i=1}^{p}\ln\left(\alpha + (1-\alpha)\lambda_i(C_1^{-1}C_2)\right) + \frac{1}{2p}\sum_{i=1}^{p}\ln\left(\lambda_i(C_1^{-1}C_2)\right) \tag{1}$$

(one can check that $\lim_{\alpha\to 1} D_{\mathrm{R}}^{\alpha} = D_{\mathrm{KL}}$).

Revolving around recent advances in random matrix theory, this article provides a generic framework to consistently estimate such functionals of the eigenvalues of $C_1^{-1}C_2$ from the samples in the regime where p, n_1, n_2 are simultaneously large, under rather mild assumptions. In addition to the wide range of potential applications, as hinted at above, this novel estimator provides in practice a dramatic improvement over the conventional covariance matrix "plug-in" approach, as we subsequently demonstrate on a synthetic (but typical) example in Table 1. Here, for $C_1^{-\frac{1}{2}}C_2C_1^{-\frac{1}{2}}$ (having the same eigenvalues as $C_1^{-1}C_2$) a Toeplitz positive definite matrix, we estimate the (squared) Fisher distance $D_{\mathrm{F}}(C_1, C_2)^2$ (averaged over a large number of realizations of zero mean Gaussian $x_i^{(a)}$'s), for $n_1 = 1\,024$, $n_2 = 2\,048$ and p varying from 2 to 512. A surprising outcome, despite the theoretical request that p must be large for our estimator to be consistent, is that, already for $p = 2$ (while $n_1, n_2 \sim 10^3$), our proposed estimator largely outperforms the classical approach; for $n_a/p \sim 10$ or less, the distinction in performance between both methods is dramatic with the classical estimator extremely biased. Our main result, Theorem 1, provides a consistent estimator for functionals f of the eigenvalues of $C_1^{-1}C_2$ under the form of a complex integral, valid for all functions $f : \mathbb{R} \to \mathbb{R}$ that have natural complex analytic extensions on given bounded regions of $\mathbb{C}$. This estimator however assumes a complex integral form which we subsequently express explicitly for a family of elementary functions f in a series of corollaries (Corollaries 1 to 4). Distances and divergences of practical interest (including all metrics listed above) being linear combinations of these atomic functions, their resulting estimates follow from merely "linearly combining" these corollaries. Table 2 provides the mapping between distances and functions f. While Corollaries 1–3 provide an exact calculus of the form provided in Theorem 1, for the case $f(t) = \ln^2(t)$, covered in Corollary 4, the exact calculus leads to an expression involving dilogarithm functions which are not elementary functions. For this reason, Corollary 4 provides a large p approximation of Theorem 1, thereby leading to another (equally valid) consistent estimator. This explains why Table 1 and the figures to come (Figs. 1 and 2) display two different sets of

Table 1
Estimation of the Fisher distance $D_F(C_1, C_2)$. Simulation example for real $x_i^{(a)} \sim \mathcal{N}(0, C_a)$ (top part) and complex $x_i^{(a)} \sim \mathcal{CN}(0, C_a)$ (bottom part), $[C_1^{-\frac{1}{2}} C_2 C_1^{-\frac{1}{2}}]_{ij} = 0.3^{|i-j|}$, $n_1 = 1\,024$, $n_2 = 2\,048$, as a function of p. The "±" values correspond to one standard deviation of the estimates. Best estimates stressed in **boldface** characters.

p	2	4	8	16	32	64	128	256	512
$D_F(C_1, C_2)^2$	0.0980	0.1456	0.1694	0.1812	0.1872	0.1901	0.1916	0.1924	0.1927
Proposed est. Theorem 1, $x_i^{(a)} \in \mathbb{R}^p$	0.0993 ± 0.0242	0.1470 ± 0.0210	0.1708 ± 0.0160	0.1827 ± 0.0120	0.1887 ± 0.0089	0.1918 ± 0.0067	0.1933 ± 0.0051	0.1941 ± 0.0045	0.1953 ± 0.0046
Proposed est. Corollary 4, $x_i^{(a)} \in \mathbb{R}^p$	**0.0979** ± 0.0242	**0.1455** ± 0.0210	**0.1693** ± 0.0160	**0.1811** ± 0.0120	**0.1871** ± 0.0089	**0.1902** ± 0.0067	**0.1917** ± 0.0051	**0.1926** ± 0.0045	**0.1940** ± 0.0046
Classical est. $x_i^{(a)} \in \mathbb{R}^p$	0.1024 ± 0.0242	0.1529 ± 0.0210	0.1826 ± 0.0160	0.2063 ± 0.0120	0.2364 ± 0.0089	0.2890 ± 0.0068	0.3954 ± 0.0052	0.6339 ± 0.0048	1.2717 ± 0.0056
Proposed est. Theorem 1, $x_i^{(a)} \in \mathbb{C}^p$	**0.0982** ± 0.0171	**0.1455** ± 0.0145	**0.1691** ± 0.0114	**0.1811** ± 0.0082	**0.1877** ± 0.0063	**0.1901** ± 0.0046	**0.1917** ± 0.0037	**0.1922** ± 0.0028	**0.1924** ± 0.0028
Proposed est. Corollary 4, $x_i^{(a)} \in \mathbb{C}^p$	0.0968 ± 0.0171	0.1441 ± 0.0145	0.1675 ± 0.0114	0.1796 ± 0.0082	0.1861 ± 0.0063	0.1886 ± 0.0046	0.1903 ± 0.0037	0.1913 ± 0.0028	0.1931 ± 0.0028
Classical est. $x_i^{(a)} \in \mathbb{C}^p$	0.1012 ± 0.0171	0.1515 ± 0.0146	0.1809 ± 0.0114	0.2048 ± 0.0082	0.2354 ± 0.0064	0.2873 ± 0.0047	0.3937 ± 0.0038	0.6318 ± 0.0030	1.2679 ± 0.0034

Table 2
Distances and divergences, and their corresponding $f(z)$ and related atomic functions.

Metrics	$f(z)$	Required atomic functions			
		z	$\ln(z)$	$\ln(1+sz)$	$\ln^2(z)$
D_F^2	$\ln^2(z)$				×
D_B	$-\frac{1}{4}\ln(z) + \frac{1}{2}\ln(1+z) - \frac{1}{2}\ln(2)$		×	×	
D_{KL}	$\frac{1}{2}z - \frac{1}{2}\ln(z) - \frac{1}{2}$	×	×		
D_R^α	$\frac{-1}{2(\alpha-1)}\ln(\alpha + (1-\alpha)z) + \frac{1}{2}\ln(z)$		×	×	

estimates. In passing, it is worth noticing that in the complex Gaussian case, the estimators offer a noticeable improvement on average over their real counterparts; this fact is likely due to a bias in the second-order fluctuations of the estimators, as discussed in Section 5.

Technically speaking, our main result unfolds from a three-step approach: (i) relating the limiting (as $p, n_1, n_2 \to \infty$ with $p/n_a = O(1)$) eigenvalue distribution of the sometimes called Fisher matrix $\hat{C}_1^{-1}\hat{C}_2$ to the limiting eigenvalue distribution of $C_1^{-1}C_2$ by means of a functional identity involving their respective Stieltjes transforms (see definition in Section 3), then (ii) expressing the studied matrix distance as a complex integral featuring the Stieltjes transform and proceeding to successive change of variables to exploit the functional identity of (i), and finally (iii) whenever possible, explicitly evaluating the complex integral through complex analysis techniques. This approach is particularly reminiscent of the eigenvalue and eigenvector projection estimates proposed by Mestre in 2008 in a series of seminal articles [18,19]. In [18], Mestre considers a single sample covariance matrix $\hat{C}$ setting and provides a complex integration approach to estimate the individual eigenvalues as well as eigenvector projections of the population covariance matrix C; there, step (i) follows immediately from the popular result [23] on the limiting eigenvalue distribution of large dimensional sample covariance matrices; step (ii) was then the (simple yet powerful) key innovation and step (iii) followed from a mere residue calculus. In a wireless communication-specific setting, the technique of Mestre was then extended in [11] in a model involving the product of Gram random matrices; there, as in the present work, step (i) unfolds from two successive applications of the results of [23], step (ii) follows essentially the same approach as in [18] (yet largely simplified) and step (iii) is again achieved from residue calculus. As the random matrix $\hat{C}_1^{-1}\hat{C}_2$ may be seen as a single sample covariance matrix conditionally on $\hat{C}_1$, with $\hat{C}_1$ itself a sample covariance, in the present work, step (i) is obtained rather straightforwardly from applying twice the results from [23,24]; the model $\hat{C}_1^{-1}\hat{C}_2$ is actually reminiscent of the so-called multivariate F-matrices, of the type $\hat{C}_1^{-1}\hat{C}_2$ but with $C_1 = C_2$, extensively studied in [3,22,32]; more recently, the very model under study here (that is, for $C_1 \neq C_2$) was analyzed in [28,31]. Step (i) of the present analysis is in particular consistent with Theorem 2.1 of [31], yet our proposed formulation is here more convenient to our purposes.

The main technical difficulty of the present contribution lies in steps (ii) and (iii) of the analysis. First, as opposed to [11,18], the complex integrals under consideration here involve rather non-smooth functions, and particularly complex logarithms. Contour integrals of complex logarithms can in general not be treated through mere residue calculus. Instead, we shall resort here to an in-depth analysis of the so-called branch-cuts, corresponding to the points of discontinuity of

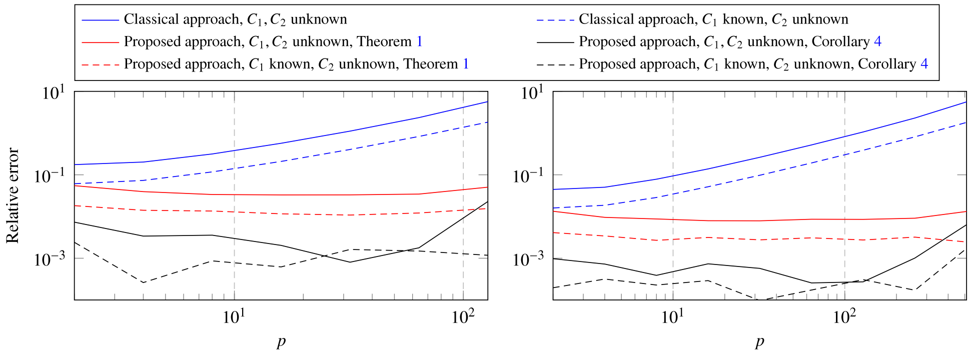

Fig. 1. Estimation of the squared Fisher distance $D_{\mathrm{F}}^2(C_1, C_2)$. Relative estimation error for $x_i^{(a)} \sim \mathcal{N}(0, C_a)$ with $[C_1^{-\frac{1}{2}} C_2 C_1^{-\frac{1}{2}}]_{ij} = .3^{|i-j|}$; (left) $n_1 = 256$, $n_2 = 512$, and (right) $n_1 = 1024$, $n_2 = 2048$, varying p.

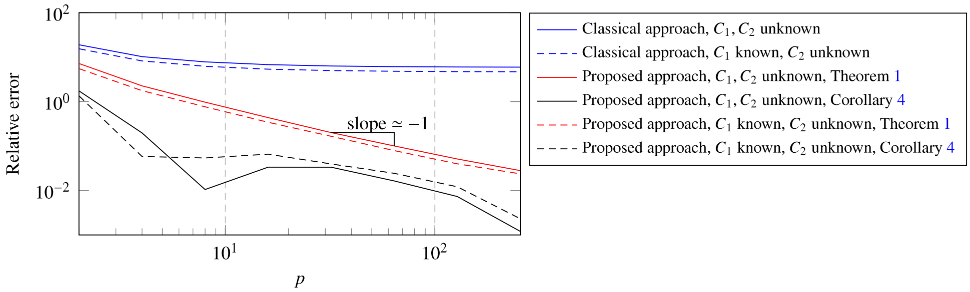

Fig. 2. Estimation of the squared Fisher distance $D_{\mathrm{F}}^2(C_1, C_2)$. Relative estimation error for $x_i^{(a)} \sim \mathcal{N}(0, C_a)$ with $[C_1^{-\frac{1}{2}} C_2 C_1^{-\frac{1}{2}}]_{ij} = .3^{|i-j|}$, $\frac{p}{n_1} = \frac{1}{4}$, $\frac{p}{n_2} = \frac{1}{2}$, varying p.

the complex logarithm, as well as to the conditions under which valid integration contours can be defined. Once integrals are properly defined, an elaborate contour design then turns the study of the complex integral into that of real integrals. As already mentioned, in the particular case of the function $f(t) = \ln^2(t)$, these real integrals result in a series of expressions involving the so-called dilogarithm function (see e.g., [30]), the many properties of which will be thoroughly exploited to obtain our final results.

A second difficulty arises in the complex contour determination which involves two successive variable changes. Working backwards from a contour surrounding the sample eigenvalues to reach a contour surrounding the population eigenvalues, one must thoroughly examine the conditions under which the expected mapping is achieved. The scenarios where $p > n_1$ or $p > n_2$ notably lead to non-trivial contour changes. This particularly entails the impossibility to estimate some functionals of the eigenvalues of $C_1^{-1}C_2$ under these conditions.

The remainder of the article presents our main result, that is the novel estimator, first under the form of a generic complex integral and then, for a set of functions met in the aforementioned classical matrix distances, under the form of a closed-form estimator. To support the anticipated strong practical advantages brought by our improved estimators, Section 4 concretely applies our findings in a statistical learning context.

2. Model

For $a \in \{1, 2\}$, let $x_1^{(a)}, \ldots, x_{n_a}^{(a)} \in \mathbb{R}^p$ (or $\mathbb{C}^p$) be n_a independent and identically distributed vectors of the form $x_i^{(a)} = C_a^{\frac{1}{2}} \tilde{x}_i^{(a)}$ with $\tilde{x}_i^{(a)} \in \mathbb{R}^p$ (respectively $\mathbb{C}^p$) a vector of i.i.d. zero mean, unit variance, and finite fourth order moment entries, and $C_a \in \mathbb{R}^{p\times p}$ (respectively $\mathbb{C}^{p\times p}$) positive definite. We define the sample covariance matrix

$$\hat{C}_a \equiv \frac{1}{n_a} X_a X_a^{\top}, \quad X_a = [x_1^{(a)}, \ldots, x_{n_a}^{(a)}].$$

We will work under the following set of assumptions.

Assumption 1 (*Growth Rate*)**.** For $a \in \{1, 2\}$,

(i) denoting $c_a \equiv \frac{p}{n_a}$, $c_a < 1$ and $c_a \to c_a^{\infty} \in (0, 1)$ as $p \to \infty$. The reader must here keep in mind that c_a is a function of p; yet, for readability and since this has little practical relevance, we do not make any explicit mention of this dependence. One may in particular suppose that $c_a = c_a^{\infty}$ for all valid p.

(ii) $\limsup_p \max(\|C_a^{-1}\|, \|C_a\|) < \infty$ with $\|\cdot\|$ the operator norm;
(iii) there exists a probability measure ν such that $\nu_p \equiv \frac{1}{p}\sum_{i=1}^p \boldsymbol{\delta}_{\lambda_i(C_1^{-1}C_2)} \to \nu$ weakly as $p \to \infty$ (with $\lambda_i(A)$ the eigenvalues of matrix A and $\boldsymbol{\delta}_x$ the atomic mass at x). Here again, in practice, one may simply assume that $\nu = \nu_p$, the discrete empirical spectral distribution for some fixed dimension p.

The main technical ingredients at the core of our derivations rely on an accurate control of the eigenvalues of $\hat{C}_1^{-1}\hat{C}_2$. In particular, we shall demand that these eigenvalues remain with high probability in a compact set. Item (ii) and the finiteness of the $2+\varepsilon$ moments of the entries of X_1 and X_2 enforce this request, through the seminal results of Bai and Silverstein on sample covariance matrix models [1,23]. Item (ii) can be relaxed but is mathematically convenient, in particular to ensure that $\mu_p \equiv \frac{1}{p}\sum_{i=1}^p \boldsymbol{\delta}_{\lambda_i(\hat{C}_1^{-1}\hat{C}_2)}$ has an almost sure limit, and practically inconsequential.

Item (i) deserves a deeper comment. Requesting that $p < n_1$ and $p < n_2$ is certainly demanding in some practical scarce data conditions. Yet, as we shall demonstrate subsequently, our proof approach relies on some key changes of variables largely involving the signs of $1 - c_1$ and $1 - c_2$. Notably, working under the assumption $c_1 > 1$ would demand a dramatic change of approach which we leave open to future work, starting with the fact that $\hat{C}_1^{-1}$ is no longer defined. The scenario $c_2 > 1$ is more interesting though. As we shall point out in a series of remarks, for some functionals f having no singularity at zero, the value $\int f d\nu_p$ can still be reliably estimated; however, dealing with $f(t) = \ln(t)$ or $f(t) = \ln^2(t)$, unfortunately at the core of all aforementioned distances and divergences D_{F}, D_{B}, D_{KL}, and D_{R}^{α}, will not be possible under our present scheme.

3. Main results

3.1. Preliminaries

For f some complex-analytically extensible real function, our objective is to estimate

$$\int f d\nu_p, \quad \nu_p \equiv \frac{1}{p}\sum_{i=1}^p \boldsymbol{\delta}_{\lambda_i(C_1^{-1}C_2)}$$

from the samples $x_i^{(1)}$'s and $x_i^{(2)}$'s in the regime where n_1, n_2, p are all large and of the same magnitude. In particular, given the aforementioned applications, we will be interested in considering the cases where $f(t) = t$, $f(t) = \ln(t)$, $f(t) = \ln(1+st)$ for some $s > 0$, or $f(t) = \ln^2(t)$. Note in passing that, since $C_1^{-1}C_2$ has the same eigenvalues as $C_1^{-\frac{1}{2}}C_2C_1^{-\frac{1}{2}}$ (by Sylvester's identity), the eigenvalues of $C_1^{-1}C_2$ are all real positive.

It will be convenient in the following to define

$$\mu_p \equiv \frac{1}{p}\sum_{i=1}^p \boldsymbol{\delta}_{\lambda_i}, \quad \lambda_i \equiv \lambda_i(\hat{C}_1^{-1}\hat{C}_2), \quad 1 \le i \le p.$$

By the law of large numbers, it is clear that, as $n_1, n_2 \to \infty$ with p fixed, $\mu_p \xrightarrow{\text{a.s.}} \nu_p$ (in law) and thus, up to some support boundedness control for unbounded f, $\int f d\mu_p - \int f d\nu_p \xrightarrow{\text{a.s.}} 0$. The main objective here is to go beyond this simple result accounting for the fact that n_1, n_2 may not be large compared to p. In this case, under Assumption 1, $\mu_p \not\to \nu$ and it is unlikely that for most f, the convergence $\int f d\mu_p - \int f d\nu_p \xrightarrow{\text{a.s.}} 0$ would still hold.

Our main line of arguments follows results from the random matrix theory and complex analysis. We will notably largely rely on the relation linking the Stieltjes transform of several measures (such as ν_p and μ_p) involved in the model. The Stieltjes transform m_μ of a measure μ is defined, for $z \in \mathbb{C} \setminus \mathrm{Supp}(\mu)$ (with $\mathrm{Supp}(\mu)$ the support of μ), as

$$m_\mu(z) \equiv \int \frac{d\mu(t)}{t - z}$$

which is complex-analytic on its definition domain and in particular has complex derivative $m'_\mu(z) \equiv \int \frac{d\mu(t)}{(t-z)^2}$. For instance, for a discrete measure $\mu \equiv \sum_{i=1}^p \alpha_i \boldsymbol{\delta}_{\lambda_i}$, $m_\mu(z) = \sum_{i=1}^p \frac{\alpha_i}{\lambda_i - z}$ and $m'_\mu(z) = \sum_{i=1}^p \frac{\alpha_i}{(\lambda_i - z)^2}$.

3.2. Generic results

With the notations from the section above at hand, our main technical result is as follows.

Theorem 1 (*Estimation via Contour Integral*). *Let Assumption 1 hold and $f : \mathbb{C} \to \mathbb{C}$ be analytic on $\{z \in \mathbb{C}, \Re[z] > 0\}$. Take also $\Gamma \subset \{z \in \mathbb{C}, \Re[z] > 0\}$ a (positively oriented) contour strictly surrounding $\cup_{p=1}^{\infty}\mathrm{Supp}(\mu_p)$ (this set is known to be almost surely compact). For $z \in \mathbb{C} \setminus \mathrm{Supp}(\mu_p)$, define the two functions*

$$\varphi_p(z) \equiv z + c_1 z^2 m_{\mu_p}(z), \quad \psi_p(z) \equiv 1 - c_2 - c_2 z m_{\mu_p}(z).$$

Then, the following result holds

$$\int f d\nu_p - \frac{1}{2\pi\imath}\oint_{\Gamma} f\left(\frac{\varphi_p(z)}{\psi_p(z)}\right)\left(\frac{\varphi_p'(z)}{\varphi_p(z)} - \frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi_p(z)}{c_2}dz \xrightarrow{\text{a.s.}} 0.$$

Proof Idea. From Cauchy's integral theorem, $\int f d\nu_p$ can be related to the Stieltjes transform m_{ν_p} through the identity $\int f d\nu_p = \frac{1}{2\pi\imath}\oint_{\Gamma_\nu} f(z) m_{\nu_p}(z)dz$ for Γ_ν a contour surrounding the eigenvalues of $\mathrm{Supp}(\nu_p)$. The idea is to connect the population-related m_{ν_p} to the sample-related m_{μ_p}. We proceed here similarly to [28] by applying twice the results from [23], thereby implying two complex variable changes. A notable difference to [28] is that, in the process, we need to track "backwards" the images of the contour Γ (from the theorem statement) by the induced variable changes. The conditions $c_1^\infty, c_2^\infty \in (0,1)$ are here fundamental to ensure that Γ is indeed mapped onto a contour Γ_ν enclosing $\mathrm{Supp}(\nu_p)$ with no extra singularity (such as 0). □

Remark 1 (*Known C_1*)**.** For C_1 known, Theorem 1 is particularized by taking the limit $c_1 \to 0$, i.e.,

$$\int f d\nu_p - \frac{1}{2\pi\imath}\oint_{\Gamma} f\left(\frac{z}{\psi_p(z)}\right)\left(\frac{1}{z} - \frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi_p(z)}{c_2}dz \xrightarrow{\text{a.s.}} 0$$

where now $m_{\mu_p}(z) = \frac{1}{p}\sum_{i=1}^p \frac{1}{\lambda_i(C_1^{-1}\hat{C}_2)-z}$ and $m_{\mu_p}'(z) = \frac{1}{p}\sum_{i=1}^p \frac{1}{(\lambda_i(C_1^{-1}\hat{C}_2)-z)^2}$. Basic algebraic manipulations allow for further simplification, leading up to

$$\int f d\nu_p - \frac{1}{2\pi\imath c_2}\oint_{\Gamma} f\left(\frac{-1}{m_{\tilde{\mu}_p}(z)}\right) m_{\tilde{\mu}_p}'(z) z dz \xrightarrow{\text{a.s.}} 0$$

where $\tilde{\mu}_p = c_2\mu_p + (1-c_2)\delta_0$ is the eigenvalue distribution of $\frac{1}{n_2}X_2^\top C_1^{-1}X_2$ (and thus $m_{\tilde{\mu}_p}(z) = c_2 m_{\mu_p}(z) - (1-c_2)/z$). Letting $g(z) = f(1/z)$ and $G(z)$ such that $G'(z) = g(z)$, integration by parts of the above expression further gives

$$\int f d\nu_p - \frac{1}{2\pi\imath c_2}\oint_{\Gamma} G\left(-m_{\tilde{\mu}_p}(z)\right) dz \xrightarrow{\text{a.s.}} 0.$$

For instance, for $f(z) = \ln^2(z)$, $G(z) = z(\ln^2(z) - 2\ln(z) + 2)$.

Remark 2 (*Extension to the $c_2 > 1$ Case*)**.** Theorem 1 extends to the case $c_2 > 1$ for all $f : \mathbb{C} \to \mathbb{C}$ analytic on the whole set $\mathbb{C}$ with $f(0) = 0$. For f defined in 0 with $f(0) \neq 0$, one can obviously write $\int f d\nu_p = f(0) + \int (f - f(0))d\nu_p$ and apply the result to $f - f(0)$. This excludes notably $f(z) = \ln^k(z)$ for $k \geq 1$, but also $f(z) = \ln(1+sz)$ for $s > 0$. Yet, the analyticity request on f can be somewhat relaxed. More precisely, Theorem 1 still holds when $c_2 > 1$ if there exists a Γ as defined in Theorem 1 such that f is analytic on the interior of the contour described by $(\varphi/\psi)(\Gamma)$, where φ and ψ are the respective almost sure limits of φ_p and ψ_p (see Appendix A for details). The main issue with the case $c_2 > 1$, as thoroughly detailed in Appendix C, is that, while $\Gamma \subset \{z \in \mathbb{C}, \Re[z] > 0\}$, the interior of $(\varphi/\psi)(\Gamma)$ necessarily contains zero. This poses dramatic limitations to the applicability of our approach for $f(z) = \ln^k(z)$ for which we so far do not have a workaround. For $f(z) = \ln(1+sz)$ though, we will show that there exist sufficient conditions on $s > 0$ to ensure that $-1/s$ (the singularity of $z \mapsto \ln(1+sz)$) is not contained within $(\varphi/\psi)(\Gamma)$, thereby allowing for the extension of Theorem 1 to $f(t) = \ln(1+st)$.

3.3. Special cases

While Theorem 1 holds for all well-behaved f on Γ, a numerical complex integral is required in practice to estimate $\int f d\nu_p$. It is convenient, when feasible, to assess the approximating complex integral in closed form, which is the objective of this section. When f is analytic in the inside of Γ, the integral can be estimated merely through a residue calculus. This is the case notably of polynomials $f(t) = t^k$. If instead f exhibits singularities in the inside of Γ, as for $f(t) = \ln^k(t)$, more advanced contour integration arguments are required.

Importantly, Theorem 1 is linear in f. Consequently, the contour integral calculus for elaborate functions f, such as met in most metrics of practical interest, can be reduced to the integral calculus of its linear components. Table 2 lists these components for the distances and divergences introduced in Section 1.

In the remainder of this section, we focus on the integral calculus for these atomic functions f.

Corollary 1 (*Case $f(t) = t$*)**.** *Under the conditions of Theorem 1,*

$$\int t d\nu_p(t) - (1-c_1)\int t d\mu_p(t) \xrightarrow{\text{a.s.}} 0.$$

and in the case where $c_1 \to 0$, this is simply $\int t d\nu_p(t) - \int t d\mu_p(t) \xrightarrow{\text{a.s.}} 0$.

As such, the classical sample covariance matrix estimator $\int t d\mu_p(t)$ needs only be corrected by a product with $(1-c_1)$. This results unfolds from Theorem 1 via a simple residue calculus.

Corollary 2 (*Case $f(t)=\ln(t)$*). *Under the conditions of Theorem 1,*

$$\int \ln(t)d\nu_p(t)-\left[\int \ln(t)d\mu_p(t)-\frac{1-c_1}{c_1}\ln(1-c_1)+\frac{1-c_2}{c_2}\ln(1-c_2)\right]\xrightarrow{\text{a.s.}}0.$$

When $c_1\to 0$, $\int \ln(t)d\nu_p(t)-\left[\int \ln(t)d\mu_p(t)+\frac{1-c_2}{c_2}\ln(1-c_2)+1\right]\xrightarrow{\text{a.s.}}0.$

Note that for $f(t)=\ln(t)$ and $c_1=c_2$, the standard estimator is asymptotically p,n_1,n_2-consistent. This is no longer true though for $c_1\neq c_2$ but only a fixed bias is induced. This result is less immediate than Corollary 1 as the complex extension of the logarithm function is multi-valued, causing the emergence of branch-cuts inside the contour. We evaluate the integral here, and in the subsequent corollaries, by means of a careful contour deformation subsequent to a thorough study of the function $\ln(\varphi_p(z)/\psi_p(z))$ and of its branch-cut locations.

Corollary 3 (*Case $f(t)=\ln(1+st)$*). *Under the conditions of Theorem 1, let $s>0$ and denote κ_0 the unique* negative *solution to $1+s\frac{\varphi_p(x)}{\psi_p(x)}=0$. Then we have*

$$\int \ln(1+st)d\nu_p(t)-\left[\frac{c_1+c_2-c_1c_2}{c_1c_2}\ln\left(\frac{c_1+c_2-c_1c_2}{(1-c_1)(c_2-sc_1\kappa_0)}\right)+\frac{1}{c_2}\ln\left(-s\kappa_0(1-c_1)\right)+\int \ln\left(1-\frac{t}{\kappa_0}\right)d\mu_p(t)\right]$$
$$\xrightarrow{\text{a.s.}}0.$$

In the case where $c_1\to 0$, this is simply

$$\int \ln(1+st)d\nu_p(t)-\left[(1+s\kappa_0+\ln(-s\kappa_0))/c_2+\int \ln\left(1-t/\kappa_0\right)d\mu_p(t)\right]\xrightarrow{\text{a.s.}}0.$$

The proof of Corollary 3 follows closely the proof of Corollary 2, yet with a fundamental variation on the branch-cut locations as the singularities of $\ln(1+s\varphi_p(z)/\psi_p(z))$ differ from those of $\ln(\varphi_p(z)/\psi_p(z))$.

Remark 3 (*Limit when $s\to\infty$ and $s\to 0$*). It is interesting to note that, as $s\to\infty$, κ_0 behaves as $-(1-c_2)/s$. Plugging this into the expression above, we find that

$$\int \ln(1+st)d\nu_p(t)-\ln(s)\sim_{s\to\infty}\int \ln(t)d\mu_p(t)+\frac{1-c_2}{c_2}\ln(1-c_2)-\frac{1-c_1}{c_1}\ln(1-c_1)$$

therefore recovering, as one expects, the result from Corollary 2.

Checking similarly the case where $s\to 0$ demands a second-order expansion of κ_0. It is first clear that κ_0 must grow unbounded as $s\to 0$, otherwise there would be a *negative* solution to $\psi_p(x)=0$, which is impossible. This said, we then find that $\kappa_0\sim-\frac{1}{(1-c_1)s}+\frac{c_1+c_2-c_1c_2}{1-c_1}\frac{1}{p}\sum_{i=1}^p\lambda_i+o(1)$. Plugging this into Corollary 3, we find

$$\frac{1}{s}\int \ln(1+st)d\nu_p(t)\sim_{s\to 0}(1-c_1)\int td\mu_p(t)$$

therefore recovering the results from Corollary 1, again as expected.

Remark 4 (*Location of κ_0*). For numerical purposes, it is convenient to easily locate κ_0. By definition,

$$\kappa_0=\frac{1-c_2-c_2\kappa_0 m_{\mu_p}(\kappa_0)}{-s(1+c_1\kappa_0 m_{\mu_p}(\kappa_0))}.$$

Along with the bound $-1<xm_{\mu_p}(x)<0$, for $x<0$, we then find that

$$-(s(1-c_1))^{-1}<\kappa_0<0.$$

As such, κ_0 may be found by a dichotomy search on the set $(-1/(s(1-c_1)),0)$.

Remark 5 (*The case $c_2>1$*). For $f(t)=\ln(1+st)$, Theorem 1 and Corollary 3 may be extended to the case $c_2>1$, however with important restrictions. Precisely, let $\mu^-\equiv\inf\{\mathrm{Supp}(\mu)\}$ for μ the almost sure weak limit of the empirical measure μ_p, and let $x^-\equiv\lim_{x\uparrow\mu^-}\frac{\varphi(x)}{\psi(x)}$, for $\varphi(z)\equiv z(1+c_1^\infty zm_\mu(z))$ and $\psi(z)\equiv 1-c_2^\infty-c_2^\infty zm_\mu(z)$ the respective almost sure functional limits of φ_p and ψ_p. Also redefine κ_0 in Corollary 3 as the smallest real (non-necessarily negative) solution to $1+s\frac{\varphi_p(x)}{\psi_p(x)}=0$. Then, for all $s>0$ satisfying $1+sx^->0$, we have the following results

(i) Theorem 1 holds true, however for a contour Γ having a leftmost real crossing within the set (κ_0,μ^-);

(ii) Corollary 3 extends to

$$\int \ln(1+st)d\nu_p(t) - \left[\frac{c_1+c_2-c_1c_2}{c_1c_2}\ln\left(\frac{c_1+c_2-c_1c_2}{(1-c_1)|c_2-sc_1\kappa_0|}\right) + \frac{1}{c_2}\ln|-s\kappa_0(1-c_1)| + \int \ln\left|1-\frac{t}{\kappa_0}\right| d\mu_p(t)\right] \xrightarrow{\text{a.s.}} 0.$$

For instance, for $C_1 = C_2$, i.e., $C_1^{-1}C_2 = I_p$, we have

$$\mu^- = \left(\frac{1-\sqrt{c_1^\infty + c_2^\infty - c_1^\infty c_2^\infty}}{1-c_1^\infty}\right)^2, \quad x^- = \frac{1-\sqrt{c_1^\infty + c_2^\infty - c_1^\infty c_2^\infty}}{1-c_1^\infty}$$

so that $x^- < 0$ when $c_2^\infty > 1$; so, there, Theorem 1 and Corollary 3 hold true as long as

$$0 < s < (1-c_1^\infty)/\left(\sqrt{c_1^\infty + c_2^\infty - c_1^\infty c_2^\infty} - 1\right). \tag{2}$$

It is clear from the second part of the remark that, as c_2^∞ increases, the range of valid s values vanishes. Perhaps surprisingly though, as $c_1^\infty \uparrow 1$, this range converges to the fixed set $(0, 2/(c_2^\infty - 1))$ and thus does not vanish.

As opposed to the previous scenarios, for the case $f(t) = \ln^2(t)$, the exact form of the integral from Theorem 1 is non-trivial and involves dilogarithm functions (see its expression in (B.4) in the appendix) which originate from numerous real integrals of the form $\int \ln(x-a)/(x-b)dx$ appearing in the calculus. This involved expression can nonetheless be significantly simplified using a large-p approximation, resulting in an estimate only involving usual functions, as shown subsequently.

Corollary 4 (*Case $f(t) = \ln^2(t)$*). *Let $0 < \eta_1 < \cdots < \eta_p$ be the eigenvalues of $\Lambda - \frac{\sqrt{\lambda}\sqrt{\lambda}^\top}{p-n_1}$ and $0 < \zeta_1 < \cdots < \zeta_p$ the eigenvalues of $\Lambda - \frac{\sqrt{\lambda}\sqrt{\lambda}^\top}{n_2}$, where $\lambda = (\lambda_1, \ldots, \lambda_p)^\top$, $\Lambda = \operatorname{diag}(\lambda)$, and $\sqrt{\lambda}$ is understood entry-wise. Then, under the conditions of Theorem 1,*

$$\int \ln^2(t)d\nu_p(t) - \left[\frac{1}{p}\sum_{i=1}^p \ln^2((1-c_1)\lambda_i) + 2\frac{c_1+c_2-c_1c_2}{c_1c_2}\left\{\left(\Delta_\zeta^\eta\right)^\top M\left(\Delta_\lambda^\eta\right) + (\Delta_\lambda^\eta)^\top r\right\}\right.$$
$$\left. -\frac{2}{p}\left(\Delta_\zeta^\eta\right)^\top N 1_p - 2\frac{1-c_2}{c_2}\left\{\frac{1}{2}\ln^2((1-c_1)(1-c_2)) + (\Delta_\zeta^\eta)^\top r\right\}\right] \xrightarrow{\text{a.s.}} 0$$

where we defined Δ_a^b the vector with $(\Delta_a^b)_i = b_i - a_i$ and, for $i,j \in \{1,\ldots,p\}$, $r_i = \frac{\ln((1-c_1)\lambda_i)}{\lambda_i}$ and

$$M_{ij} = \begin{cases} \frac{\frac{\lambda_i}{\lambda_j} - 1 - \ln\left(\frac{\lambda_i}{\lambda_j}\right)}{(\lambda_i-\lambda_j)^2} & , i \neq j \\ \frac{1}{2\lambda_i^2} & , i = j \end{cases}, \quad N_{ij} = \begin{cases} \frac{\ln\left(\frac{\lambda_i}{\lambda_j}\right)}{\lambda_i - \lambda_j} & , i \neq j \\ \frac{1}{\lambda_i} & , i = j. \end{cases}$$

In the limit $c_1 \to 0$ (i.e., for C_1 known), this becomes

$$\int \ln^2(t)d\nu_p(t) - \left[\frac{1}{p}\sum_{i=1}^p \ln^2(\lambda_i) + \frac{2}{p}\sum_{i=1}^p \ln(\lambda_i) - \frac{2}{p}\left(\Delta_\zeta^\lambda\right)^\top Q 1_p - 2\frac{1-c_2}{c_2}\left\{\frac{1}{2}\ln^2(1-c_2) + \left(\Delta_\zeta^\lambda\right)^\top q\right\}\right] \xrightarrow{\text{a.s.}} 0$$

with

$$Q_{ij} = \begin{cases} \frac{\lambda_i \ln\left(\frac{\lambda_i}{\lambda_j}\right) - (\lambda_i - \lambda_j)}{(\lambda_i - \lambda_j)^2} & , i \neq j, \\ \frac{1}{2\lambda_i} & , i = j, \end{cases} \quad \text{and } q_i = \frac{\ln(\lambda_i)}{\lambda_i}.$$

4. Experimental results

4.1. Validation of theoretical results

This section presents experimental verifications, completing Table 1 in Section 1. In all cases presented below, we consider the squared Fisher distance (i.e., the case $f(t) = \ln^2(t)$) between C_1 and C_2, where $C_1^{-\frac{1}{2}}C_2C_1^{-\frac{1}{2}}$ is a Toeplitz matrix with entry (i,j) equal to $.3^{|i-j|}$. Fig. 1 displays the normalized estimation error (i.e., the absolute difference between genuine and estimated squared distance over the genuine squared distance) for $x_i^{(a)}$ real $\mathcal{N}(0, C_a)$, for varying values of the size p of C_1 and C_2 but for n_1 and n_2 fixed ($n_1 = 256$ and $n_2 = 512$ in the left-hand display, $n_1 = 1024$, $n_2 = 2048$ in the right-hand display); results are averaged over 10 000 independent realizations. We compare the classical n_1, n_2-consistent estimator to the proposed n_1, n_2, p-consistent estimators obtained from both Theorem 1, or equivalently here from the

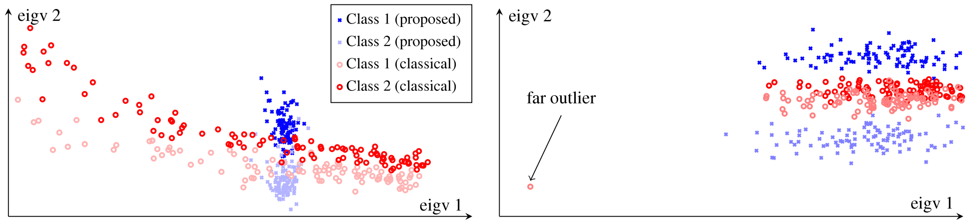

Fig. 3. First and second eigenvectors of K for the traditional estimator (red circles) versus the proposed one (blue crosses); (left) random number of snapshots n_i; (right) $n_1 = \cdots = n_{m-1} = 512$ and $n_m = 256$.

closed-form expression (B.4) from the appendix, and from Corollary 4. We also add in dashed lines the distance estimate for C_1 a priori known.

It is seen from the figures, provided in log scale, that the relative error of the standard n_1, n_2-consistent approach diverges with p, while the error of the proposed estimator remains at a low, rather constant, value. As already observed in Table 1, the large-p approximation formula from Corollary 4 even exhibits a better behavior than the exact integral expression from Theorem 1 in this real Gaussian setup.

In the same setting, Fig. 2 subsequently shows the performances in terms of relative error for *simultaneously* growing n_1, n_2, p, with a constant ratio $p/n_1 = 1/4$ and $p/n_2 = 1/2$. As expected from our theoretical analysis, the proposed estimators display a vanishing error as $n_1, n_2, p \to \infty$. The standard estimator, on the opposite, shows a high saturating relative error.

4.2. Application to spectral clustering

In this section, we develop a practical application of our theoretical findings to the machine learning context of kernel spectral clustering [26]. In several application domains, such as in brain signal processing [21] or hyperspectral imaging [7], the relevant discriminative data "features" are the population covariance matrices of the data vectors (p-sensor brain activities, p-frequency spectra). Classification in these contexts is thus performed by comparing the population covariance distances for each data-pair. Not being directly accessible, the population covariances are classically substituted by their sample estimates. We will show here that this method has strong limitations that our proposed improved distance estimates overcome.

Specifically, we consider $m = 200$ data $X_1, \ldots, X_m$ to be clustered. Datum X_i is a $p \times n_i$ independent generation of n_i independent p-dimensional zero mean Gaussian samples $X_i = [x_1^{(i)}, \ldots, x_{n_i}^{(i)}] \in \mathbb{R}^{p \times n_i}$. Half of the samples have covariance $\mathrm{E}[x_j^{(i)} x_j^{(i)\top}] = \mathrm{E}[\frac{1}{n_i} X_i X_i^\top] = C_1$ (they will be said to belong to class 1) and half have covariance C_2 (they belong to class 2). For clarity, let us say that $\mathrm{E}[x_j^{(i)} x_j^{(i)\top}] = C_1$ for $i \leq m/2$ and $\mathrm{E}[x_j^{(i)} x_j^{(i)\top}] = C_2$ for $i > m/2$. We then define the kernel matrix $K \in \mathbb{R}^{m \times m}$ by

$$K_{ij} \equiv \exp\left(-\frac{1}{2}\hat{D}_{\mathrm{F}}(X_i, X_j)^2\right)$$

with $\hat{D}_{\mathrm{F}}$ either the classical (naive) estimator of D_{F} or our proposed random matrix-improved estimator. The purpose of spectral clustering is to retrieve the mapping between data indices and classes. It can be shown (see e.g., [27] but more fundamentally [9] for large dimensional data) that, for sufficiently distinct classes (so here covariance matrices), the eigenvectors v_1 and v_2 of K associated to its largest two eigenvalues are structured according to the classes. Thus, the classes can be read out of a two-dimensional display of v_2 versus v_1. This is the procedure we follow.

A fundamental, but at first unsettling, outcome arises as we let $n_1 = \cdots = n_m$. In this case, for all tested values of p and m, the eigenvectors of K for either choice of $\hat{D}_{\mathrm{F}}$ are extremely similar. Consequently, spectral clustering performs the same, despite the obvious inaccuracies in the estimations of D_{F} by the classical estimator. This result can be explained from the following fact: spectral clustering is of utmost interest when C_1 and C_2 are close matrices; in this case, the classical estimator for D_{F} is systematically biased by a constant, almost irrespective of the covariance (since both are essentially the same). This constant bias does affect K but not its dominant eigenvectors.

This observation collapses when the values of the n_i's differ. This is depicted in Fig. 3. The left display provides a scatter plot of the dominant eigenvectors v_2 versus v_1 of K under the same conditions as above but now with n_i chosen uniformly at random in $[2p, 4p]$, with $p = 128$, $m = 200$, $C_1 = I_p$ and $[C_2]_{ij} = .05^{|i-j|}$. There, for the classical estimator, we observe a wide spreading of the eigenvector entries and a smaller inter-class spacing. This suggests poor clustering performance. On the opposite, the well-centered eigenvectors achieved by the proposed estimator imply good clustering performances. In a likely more realistic setting in practice, the right display considers the case where $n_1 = \cdots = n_{m-1} = 512$ and $n_m = 256$. This situation emulates a data retrieval failure for one observation (only half of the samples are seen). In this scenario, the classical estimator isolates one entry-pair in (v_1, v_2) (corresponding to their last entries). This is quite expected. However, more surprisingly, the presence of this outlier strongly alters the possibility to resolve the other data. This effect is further exacerbated when adding more outliers, which are not displayed here. This most likely follows from an adversarial effect between the outliers and the genuine clusters, which all tend to "drive" the dominant eigenvectors.

5. Concluding remarks

As pointed out in the introduction, the series of estimators derived in this article allow the practitioner to assess a large spectrum of covariance matrix distances or divergences based on sample estimators when the number n_1, n_2 of available data is of similar order of magnitude as the size p of the individual samples. For instance, the Rényi divergence (1) between $\mathcal{N}(0, C_2)$ and $\mathcal{N}(0, C_1)$ can be expressed as a linear combination of the estimator of Corollary 3 with $s = \frac{1-\alpha}{\alpha}$ and of the estimator of Corollary 2. In particular, the results from Corollaries 1–4 are sufficient to cover all matrix distances and divergences discussed in the introductory Section 1. For other distances involving "smooth" linear functionals of the eigenvalues of $C_1^{-1}C_2$, the generic result from Theorem 1 can be used to retrieve (at least) a numerical estimate of the sought-for distance or divergence. As exemplified by kernel spectral clustering in Section 4.2, applications of the present estimators to specific contexts of signal and data engineering at large, and specifically to the areas of statistical signal processing, artificial intelligence, and machine learning, are numerous. Further note that, while only proposed for metrics involving the eigenvalues of $C_1^{-1}C_2$, the methodology devised here can be flexibly adapted to distances based on the eigenvalues of other matrix models, such as C_1C_2. This encompasses the Wasserstein distance between two centered Gaussian as well as the Frobenius distance.

The analysis of the slope in the log–log plot of Fig. 2 of the relative error for the proposed estimator reveals that, for fixed ratios p/n_1 and p/n_2, the error behaves as p^{-1}, or equivalent as n_1^{-1} or n_2^{-1}. This suggests that, as opposed to classical n-consistent estimators for which optimal fluctuations of the estimators are usually expected to be of order $n^{-\frac{1}{2}}$, the proposed n_1, n_2, p-consistent estimator exhibits a quadratically faster convergence speed. This observation is in fact consistent with previous findings, such as [29] which demonstrated a central limit theorem with speed p^{-1} for the eigenvalue functional estimators proposed by Mestre in [18]. This is also reminiscent from the numerous central limit theorems of eigenvalue functionals derived in random matrix theory [14,20,31] since the early findings from Bai and Silverstein [2]. This fast convergence speed partly explains the strong advantage of the proposed estimator over standard large-n alone estimates, even for very small ratios p/n, as shown in our various simulations.

For practical purposes, it would be convenient to obtain a result similar to [29] in the present context, that is a central limit theorem for the fluctuations of the estimator of Theorem 1. This would allow practitioners to access both a consistent estimator for their sought-for matrix distance *as well as* a confidence margin. This investigation demands even more profound calculi (as can be seen from the detailed derivations of [29]) and is left to future work.

Also, recall from Table 1 that the estimator from Theorem 1 has a much better behavior on average in the complex Gaussian rather than in the real Gaussian case. This observation is likely due to a systematic bias of order $O(1/p)$ which is absent in the complex case. Indeed, the results from [31] show that the difference $p(m_{\mu_p}(z) - m_\mu(z))$ satisfies a central limit theorem with in general non-zero mean apart from the complex Gaussian case (or, to be more exact, apart from the cases where the $x_i^{(a)}$'s have complex independent entries with zero mean, unit variance, and zero kurtosis). Coupling this result with [29] strongly suggests that the proposed estimators in the present article exhibit a systematic order-p^{-1} bias but in the complex Gaussian case. As this bias can likely be itself estimated from the raw data, this observation opens the door to a further improvement of the estimator in Theorem 1 that would discard the bias. This would not change the order of magnitude of the error (still of order $1/p$) but reduces its systematic part.

A last important item to be discussed at this point lies in the necessary condition $n_1 > p$ and $n_2 > p$ in the analysis. We have shown in the proof of Theorem 1 (in Appendices A and C) that the requirements $c_1 < 1$ and $c_2 < 1$ are both mandatory for our estimation approach to remain valid on a range of functions f analytic on $\{z \in \mathbb{C}, \Re[z] > 0\}$ (which notably includes here logarithm functions). Yet, as discussed throughout the article, while $n_1 > p$ is mandatory for our proof approach to remain valid, the constraint $n_2 > p$ can be relaxed to some extent. Yet, this excludes functions f that are not analytic in a neighborhood of zero, thereby excluding functions such as powers of $1/z$ or of $\ln(z)$. More advanced considerations, and possibly a stark change of approach, are therefore demanded to retrieve consistent estimators when $p > n_2$ for these functions. If one resorts to projections, dimension reductions, or regularization techniques to obtain an invertible estimate for $\hat{C}_1$, one may even allow for $p > n_1$, but this would dramatically change the present analysis. As such, the quest for n_1, n_2, p-consistent estimators of the matrix distances when either $n_1 < p$ or $n_2 < p$ also remains an interesting open research avenue.

Appendices

We provide here the technical developments for the proof of Theorem 1 and Corollaries 1–4.

The appendix is structured as follows: Appendix A provides the proof of Theorem 1 following the same approach as in [11], relying mostly on the results from [23,24]. Appendix B then provides the technical details of the calculi behind Corollaries 1–4; this is undertaken through a first thorough characterization of the singular points of φ_p and ψ_p and functionals of these (these singular points are hereafter denoted λ_i, η_i, ζ_i and κ_i), allowing for a proper selection of the integration contour, and subsequently through a detailed calculus for all functions $f(t)$ under study. Appendix C discusses in detail the question of the position of the complex contours when affected by change of variables. Finally, Appendix D provides some analysis of the extension of Theorem 1 and the corollaries to the $c_2 > 1$ scenario.

Appendix A. Integral form

A.1. Relating m_ν to m_μ

We start by noticing that we may equivalently assume the following setting: (i) $x_1^{(1)},\ldots,x_{n_1}^{(1)}\in\mathbb{R}^p$ are vectors of i.i.d. zero mean and unit variance entries; (ii) $x_1^{(2)},\ldots,x_{n_2}^{(2)}\in\mathbb{R}^p$ are of the form $x_i^{(2)}=C^{\frac12}\tilde{x}_i^{(2)}$ with $\tilde{x}_i^{(2)}\in\mathbb{R}^p$ a vector of i.i.d. zero mean and unit variance entries, where $C\equiv C_1^{-\frac12}C_2C_1^{-\frac12}$.

Indeed, with our initial notations (Section 2), $\hat{C}_1^{-1}\hat{C}_2=\frac{1}{n_1}C_1^{-\frac12}\tilde{X}_1\tilde{X}_1^\top C_1^{-\frac12}\frac{1}{n_2}C_2^{\frac12}\tilde{X}_2\tilde{X}_2^\top C_2^{\frac12}$ (here $\tilde{X}_a=[\tilde{x}_1^{(a)},\ldots,\tilde{x}_{n_a}^{(a)}]$), the eigenvalue distribution of which is the same as that of the matrix $(\frac{1}{n_1}\tilde{X}_1\tilde{X}_1^\top)(\frac{1}{n_2}C_1^{-\frac12}C_2^{\frac12}\tilde{X}_2\tilde{X}_2^\top C_2^{\frac12}C_1^{-\frac12})$ and we may then consider that the $x_i^{(1)}$'s actually have covariance I_p, while the $x_i^{(2)}$'s have covariance $C=C_1^{-\frac12}C_2C_1^{-\frac12}$, without altering the spectra under study.

We then from now on redefine $\hat{C}_a=\frac{1}{n_a}X_aX_a^\top$ with $X_a=[x_1^{(a)},\ldots,x_{n_a}^{(a)}]\in\mathbb{R}^{p\times n_a}$ for $x_i^{(a)}$ given by (i) and (ii) above. With these definitions, we first condition with respect to the $x_i^{(2)}$'s, and study the spectrum of $\hat{C}_1^{-1}\hat{C}_2$, which is the same as that of $\hat{C}_2^{\frac12}\hat{C}_1^{-1}\hat{C}_2^{\frac12}$. A useful remark is the fact that $\hat{C}_2^{\frac12}\hat{C}_1^{-1}\hat{C}_2^{\frac12}$ is the "inverse spectrum" of $\hat{C}_2^{-\frac12}\hat{C}_1\hat{C}_2^{-\frac12}$, which is itself the same spectrum as that of $\frac{1}{n_1}X_1^\top\hat{C}_2^{-1}X_1$ except for n_1-p additional zero eigenvalues.

Denoting $\tilde{\mu}_p^{-1}$ the eigenvalue distribution of $\frac{1}{n_1}X_1^\top\hat{C}_2^{-1}X_1$, we first know from [23] that, under Assumption 1, as $p\to\infty$, $\tilde{\mu}_p^{-1}\xrightarrow{\text{a.s.}}\tilde{\mu}^{-1}$, where $\tilde{\mu}^{-1}$ is the probability measure with Stieltjes transform $m_{\tilde{\mu}^{-1}}$ defined as the unique (analytical function) solution to

$$m_{\tilde{\mu}^{-1}}(z)=\left(-z+c_1^\infty\int\frac{td\xi_2^{-1}(t)}{1+tm_{\tilde{\mu}^{-1}}(z)}\right)^{-1}$$

with ξ_2 the almost sure limiting spectrum distribution of $\hat{C}_2$ and m_{ξ_2} its associated Stieltjes transform; note importantly that, from [23] and Assumption 1, ξ_2 has bounded support and is away from zero. Recognizing a Stieltjes transform from the right-hand side integral, this can be equivalently written

$$m_{\tilde{\mu}^{-1}}(z)=\left(-z+\frac{c_1^\infty}{m_{\tilde{\mu}^{-1}}(z)}-\frac{c_1^\infty}{m_{\tilde{\mu}^{-1}}(z)^2}m_{\xi_2^{-1}}\left(-\frac{1}{m_{\tilde{\mu}^{-1}}(z)}\right)\right)^{-1}. \tag{A.1}$$

Accounting for the aforementioned zero eigenvalues, $\tilde{\mu}^{-1}$ relates to μ^{-1}, the (a.s.) limiting spectrum distribution of $\hat{C}_2^{-1/2}\hat{C}_1\hat{C}_2^{-1/2}$, through the relation $\tilde{\mu}^{-1}=c_1^\infty\mu^{-1}+(1-c_1^\infty)\boldsymbol{\delta}_0$ with $\boldsymbol{\delta}_x$ the Dirac measure at x and we have $m_{\tilde{\mu}^{-1}}(z)=c_1^\infty m_{\mu^{-1}}(z)-(1-c_1^\infty)/z$. Plugging this last relation in (A.1) leads then to

$$m_{\xi_2^{-1}}\left(z/(1-c_1^\infty-c_1^\infty zm_{\mu^{-1}}(z))\right)=m_{\mu^{-1}}(z)\left(1-c_1^\infty-c_1^\infty zm_{\mu^{-1}}(z)\right). \tag{A.2}$$

Now, with the convention that, for a probability measure θ, θ^{-1} is the measure defined through $\theta^{-1}([a,b])=\theta\left([\frac1a,\frac1b]\right)$, we have the Stieltjes transform relation $m_{\theta^{-1}}(z)=-1/z-m_\theta(1/z)/z^2$. Using this relation in (A.1), we get

$$zm_\mu(z)=(z+c_1^\infty z^2m_\mu(z))m_{\xi_2}(z+c_1^\infty z^2m_\mu(z))=\varphi(z)m_{\xi_2}(\varphi(z)) \tag{A.3}$$

where we recall that $\varphi(z)=z(1+c_1^\infty zm_\mu(z))$. It is useful for later use to differentiate this expression along z to obtain

$$m'_{\xi_2}(\varphi(z))=\frac{1}{\varphi(z)}\left(\frac{m_\mu(z)+zm'_\mu(z)}{\varphi'(z)}-m_{\xi_2}(\varphi(z))\right)=\frac{1}{\varphi(z)}\left(-\frac{\psi'(z)}{c_2^\infty\varphi'(z)}-m_{\xi_2}(\varphi(z))\right). \tag{A.4}$$

We next determine m_{ξ_2} as a function of ν. Since $\hat{C}_2$ is itself a sample covariance matrix, we may apply again the results from [23]. Denoting $\tilde{\xi}_2$ the almost sure limiting spectrum distribution of $\frac{1}{n_2}\tilde{X}_2^\top C\tilde{X}_2$, we first have

$$m_{\tilde{\xi}_2}(z)=\left(-z+c_2^\infty\int\frac{td\nu(t)}{1+tm_{\tilde{\xi}_2}(z)}\right)^{-1}. \tag{A.5}$$

Similar to previously, we have the relation $m_{\tilde{\xi}_2}(z)=c_2^\infty m_{\xi_2}(z)-\frac{(1-c_2^\infty)}{z}$ which, plugged in (A.5), yields

$$m_\nu\left(-z/(c_2^\infty zm_{\xi_2}(z)-(1-c_2^\infty))\right)=-m_{\xi_2}(z)\left(c_2^\infty zm_{\xi_2}(z)-(1-c_2^\infty)\right). \tag{A.6}$$

Relations (A.3) and (A.5) will be instrumental to relating $\int fd\nu$ to the observation measure μ_p, as described next.

Remark 6 (*The case $c_2 > 1$*)**.** The aforementioned reasoning carries over to the case $c_2 > 1$. Indeed, since Eq. (A.1) is now meaningless (as the support of ξ_2 contains the atom $\{0\}$), consider the model $\hat{C}_1^{-1}(\hat{C}_2 + \varepsilon I_p) = \hat{C}_1^{-1}\hat{C}_2 + \varepsilon\hat{C}_1^{-1}$ for some small $\varepsilon > 0$. Then (A.3) holds with now ξ_2 the limiting empirical spectral distribution of $\hat{C}_2 + \varepsilon I_p$. Due to ε, Eq. (A.5) now holds with $m_{\xi_2}(z)$ replaced by $m_{\xi_2}(z+\varepsilon)$. By continuity in the small ε limit, we then have that (A.3) and (A.6) still hold in the small ε limit. Now, since $\hat{C}_1^{-1}(\hat{C}_2 + \varepsilon I_p) - \hat{C}_1^{-1}\hat{C}_2 = \varepsilon\hat{C}_1^{-1}$, the operator norm of which almost surely vanishes as $\varepsilon \to 0$ from the almost sure boundedness of $\limsup_p \|\hat{C}_1^{-1}\|$, we deduce that $\mu_p \to \mu$ defined through (A.3) and (A.6), almost surely, also for $c_2^\infty > 1$.

A.2. Integral formulation over m_ν

With the formulas above, we are now in position to derive the proposed estimator. We start by using Cauchy's integral formula to obtain

$$\int f d\nu = -\frac{1}{2\pi\imath}\oint_{\Gamma_\nu} f(z)m_\nu(z)dz$$

for Γ_ν a complex contour surrounding the support of ν but containing no singularity of f in its inside. This contour is carefully chosen as the image of the mapping $\omega \mapsto z = -\omega/(c_2^\infty\omega m_{\xi_2}(\omega) - (1-c_2^\infty))$ of another contour Γ_{ξ_2} surrounding the limiting support of ξ_2; the details of this (non-trivial) contour change are provided in Appendix C (where it is seen that the assumption $c_2^\infty < 1$ is crucially exploited). We shall admit here that this change of variable is licit and that $\Gamma_\nu \subset \{z \in \mathbb{C} \mid \Re[z] > 0\}$.

Operating the aforementioned change of variable gives

$$\int f d\nu = \frac{1}{2\pi\imath}\oint_{\Gamma_{\xi_2}} \frac{f\left(\frac{-\omega}{c_2^\infty\omega m_{\xi_2}(\omega)-(1-c_2^\infty)}\right) m_{\xi_2}(\omega)\left(c_2^\infty\omega^2 m_{\xi_2}'(\omega) + (1-c_2^\infty)\right)}{c_2^\infty\omega m_{\xi_2}(\omega) - (1-c_2^\infty)} d\omega \tag{A.7}$$

where we used (A.6) to eliminate m_ν.

To now eliminate m_{ξ_2} and obtain an integral form only as a function of m_μ, we next proceed to the variable change $u \mapsto \omega = \varphi(u) = u + c_1^\infty u^2 m_\mu(u)$. Again, this involves a change of contour, which is valid as long as Γ_{ξ_2} is the image by φ of a contour Γ_μ surrounding the support of μ, which is only possible if $c_1^\infty < 1$ (see Appendix C for further details). With this variable change, we can now exploit the relations (A.3) and (A.4) to obtain, after basic algebraic calculus (using in particular the relation $um_\mu(u) = (-\psi(u) + 1 - c_2^\infty)/c_2^\infty$)

$$\begin{aligned}\int f d\nu &= \frac{-1}{2\pi\imath}\oint_{\Gamma_\mu} f\left(\frac{\varphi(u)}{\psi(u)}\right)\frac{1}{c_2^\infty\varphi(u)}\left[\varphi(u)\psi'(u) - \psi(u)\varphi'(u)\right]du \\ &= \frac{1}{2\pi\imath}\oint_{\Gamma_\mu} f\left(\frac{\varphi(u)}{\psi(u)}\right)\frac{\psi(u)}{c_2^\infty}\left[\frac{\varphi'(u)}{\varphi(u)} - \frac{\psi'(u)}{\psi(u)}\right]du - \frac{1-c_2^\infty}{c_2^\infty}\frac{1}{2\pi\imath}\oint_{\Gamma_\mu} f\left(\frac{\varphi(u)}{\psi(u)}\right)\left[\frac{\varphi'(u)}{\varphi(u)} - \frac{\psi'(u)}{\psi(u)}\right]du.\end{aligned}$$

Performing the variable change backwards ($u \mapsto z = \varphi(u)/\psi(u)$), the rightmost term is simply

$$\frac{1}{2\pi\imath}\oint_{\Gamma_\mu} f\left(\frac{\varphi(u)}{\psi(u)}\right)\left[\frac{\varphi'(u)}{\varphi(u)} - \frac{\psi'(u)}{\psi(u)}\right]du = \frac{1}{2\pi\imath}\oint_{\Gamma_\mu} f\left(\frac{\varphi(u)}{\psi(u)}\right)\frac{\psi(u)}{\varphi(u)}\left(\frac{\varphi(u)}{\psi(u)}\right)' du = \frac{1}{2\pi\imath}\oint_{\Gamma_\nu}\frac{f(z)}{z}dz = 0$$

since $\Gamma_\nu \subset \{z \in \mathbb{C} \mid \Re[z] > 0\}$. Note in passing that, if Γ_ν were to contain 0 (which occurs when $c_2^\infty > 1$), then an additional residue equal to $-(1-c_2^\infty)f(0)/c_2^\infty$ would have to be accounted for; this justifies the need to impose $f(0) = 0$ in Remark 2. We thus conclude that

$$\int f d\nu = \frac{1}{2\pi\imath}\oint_{\Gamma_\mu} f\left(\frac{\varphi(u)}{\psi(u)}\right)\frac{\psi(u)}{c_2^\infty}\left[\frac{\varphi'(u)}{\varphi(u)} - \frac{\psi'(u)}{\psi(u)}\right]du$$

as requested. It then remains to use the convergence $\nu_p \to \nu$ and $m_{\mu_p} \xrightarrow{\text{a.s.}} m_\mu$, along with the fact that the eigenvalues $\hat{C}_1^{-1}\hat{C}_2$ almost surely do not escape the limiting support μ as $p \to \infty$ (this is ensured from [1], Item List (ii) of Assumption 1 and the analyticity of the involved functions), to retrieve Theorem 1 by uniform convergence on the compact contour (see also [11] for a similar detailed derivation).

Remark 7 (*Case C_1 Known*)**.** The case where C_1 is known is equivalent to setting $c_1^\infty \to 0$ above, leading in particular to $m_\mu = m_{\xi_2}$ and to the unique functional equation

$$m_\nu\left(z/(1-c_2^\infty - c_2^\infty z m_\mu(z))\right) = m_\nu(z)\left(1 - c_2^\infty - c_2^\infty z m_\nu(z)\right).$$

In particular, if $C_1 = C_2$, this reduces to $-m_\mu(z)(z-\psi(z)) = 1$ with $\psi(z) = 1 - c_2^\infty - c_2^\infty z m_\mu(z)$, which is the functional Stieltjes-transform equation of the Marčenko–Pastur law [17].

Appendix B. Integral calculus

To compute the complex integral, note first that, depending on f, several types of singularities in the integral may arise. Of utmost interest (but not always exhaustively, as we shall see for $f(t) = \ln(1 + st)$) are: (i) the eigenvalues λ_i of $\hat{C}_1^{-1}\hat{C}_2$, (ii) the values η_i such that $\varphi_p(\eta_i) = 0$, (iii) the values ζ_i such that $\psi_p(\zeta_i) = 0$.

In the following, we first introduce a sequence of intermediary results of interest for most of the integral calculi.

B.1. Rational expansion

At the core of the subsequent analysis is the function $\left(\frac{\varphi_p'(z)}{\varphi_p(z)} - \frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi_p(z)}{c_2}$. As this is a mere rational function, we first obtain the following important expansion, that will be repeatedly used in the sequel:

$$\left(\frac{\varphi_p'(z)}{\varphi_p(z)} - \frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi_p(z)}{c_2} = \left(\frac{1}{p} - \frac{c_1+c_2-c_1c_2}{c_1c_2}\right)\sum_{j=1}^{p}\frac{1}{z-\lambda_j} + \frac{1-c_2}{c_2}\frac{1}{z} + \frac{c_1+c_2-c_1c_2}{c_1c_2}\sum_{j=1}^{p}\frac{1}{z-\eta_j}. \tag{B.1}$$

This form is obtained by first observing that the λ_j's, η_j's and 0 are the poles of the left-hand side expression. Then, pre-multiplying the left-hand side by $(z-\lambda_j)$, z, or $(z-\eta_j)$ and taking the limit when these terms vanish, we recover the right-hand side, using in particular the following estimates, which easily entail from the definitions of φ_p and ψ_p:

$$\varphi_p(z) = \frac{c_1}{p}\frac{\lambda_i^2}{\lambda_i - z} - 2c_1\frac{\lambda_i}{p} + \lambda_i + \frac{c_1}{p}\sum_{j\neq i}\frac{\lambda_i^2}{\lambda_j-\lambda_i} + O(\lambda_i - z), \qquad \varphi_p'(z) = \frac{c_1}{p}\frac{\lambda_i^2}{(\lambda_i-z)^2} + O(1),$$

$$\psi_p(z) = -\frac{c_2}{p}\frac{\lambda_i}{\lambda_i - z} + \frac{c_2}{p} + 1 - c_2 - \frac{c_2}{p}\sum_{j\neq i}\frac{\lambda_i}{\lambda_j-\lambda_i} + O(\lambda_i - z), \qquad \psi_p'(z) = -\frac{c_2}{p}\frac{\lambda_i}{(\lambda_i-z)^2} + O(1)$$

in the vicinity of λ_i, along with $\psi_p(\eta_i) = \frac{c_1+c_2-c_1c_2}{c_1}$ and $\psi_p(0) = 1 - c_2$.

From this expression, we have the following immediate corollary.

Remark 8 (*Residue for f Analytic at λ_i*)**.** If $f\circ(\varphi_p/\psi_p)$ is analytic in a neighborhood of λ_i, i.e., if f is analytic in a neighborhood of $-(c_1/c_2)\lambda_i$, then λ_i is a first order pole for the integrand, leading to the residue

$$\mathrm{Res}(\lambda_i) = -f\left(-\frac{c_1}{c_2}\lambda_i\right)\left[\frac{c_1+c_2-c_1c_2}{c_1c_2} - \frac{1}{p}\right].$$

B.2. Characterization of η_i and ζ_i, and φ_p/ψ_p

First note that the η_i (the zeros of $\varphi_p(z)$) and ζ_i (the zeros of $\psi_p(z)$) are all real as one can verify that, for $\Im[z] \neq 0$, $\Im[\varphi_p(z)]\Im[z] > 0$ and $\Im[\psi_p(z)]\Im[z] < 0$.

Before establishing the properties of φ_p and ψ_p in the vicinity of η_i and ζ_i, let us first locate these values. A study of the function $M_p : \mathbb{R} \to \mathbb{R}$, $x \mapsto x m_{\mu_p}(x)$ (see Fig. B.4) reveals that M_p is increasing, since $x/(\lambda_i - x) = -1 + 1/(\lambda_i - x)$, and has asymptotes at each λ_i with $\lim_{x\uparrow\lambda_i} M_p(x) = \infty$ and $\lim_{x\downarrow\lambda_i} M_p(x) = -\infty$. As a consequence, since $\varphi_p(x) = 0 \Leftrightarrow M_p(x) = -\frac{1}{c_1} < -1$, there exists exactly one solution to $\varphi_p(x) = 0$ in the set $(\lambda_i, \lambda_{i+1})$. This solution will be subsequently called η_i. Since $M_p(x) \to -1$ as $x \to \infty$, there exists a last solution to $\varphi_p(x) = 0$ in (λ_p, ∞), hereafter referred to as η_p. Similarly, $\psi_p(x) = 0 \Leftrightarrow M_p(x) = (1-c_2)/c_2 > 0$ and thus there exists exactly one solution, called ζ_i in $(\lambda_{i-1}, \lambda_i)$. When $x \to 0$, $M_p(x) \to 0$ so that a further solution is found in $(0, \lambda_1)$, called ζ_1. Besides, due to the asymptotes at every λ_i, we have that $\zeta_1 < \lambda_i < \eta_1 < \zeta_2 < \cdots < \eta_p$.

As such, the set Γ defined in Theorem 1 exactly encloses all η_i, λ_i, and ζ_i, for $i = 1,\ldots,p$, possibly to the exception of the leftmost ζ_1 and the rightmost η_p, as those are not comprised in a set of the form $[\lambda_{i+1}, \lambda_i]$. To ensure that the latter do asymptotically fall within the interior of Γ, one approach is to exploit Theorem 1 for the elementary function $f(t) = 1$. There we find that

$$\frac{1}{2\pi\imath}\oint_{\Gamma_\nu} m_\nu(z)dz - \frac{1}{2\pi\imath}\oint_{\Gamma}\left(\frac{\varphi_p'(z)}{\varphi_p(z)} - \frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi_p(z)}{c_2}dz \xrightarrow{\text{a.s.}} 0.$$

The left integral is easily evaluated by residue calculus and equals -1 (each $\lambda_i(C_1^{-1}C_2)$, $1 \le i \le p$, is a pole with associated residue $-1/p$), while the right integral can be computed from (B.1) again by residue calculus and equals $-1 + \frac{c_1+c_2-c_1c_2}{c_1c_2}(p - \#\{\eta_i \in \Gamma^\circ\})$ with Γ° the "interior" of Γ. As such, since both integrals are (almost surely) arbitrarily close in the large p limit, we deduce that $\#\{\eta_i \in \Gamma^\circ\} = p$ for all large p and thus, in particular, η_p is found in the interior of Γ. To obtain the same result for ζ_1, note that, from the relation $\psi_p(z) = \frac{c_1+c_2-c_1c_2}{c_1} - \frac{c_2}{c_1}\frac{\varphi_p(z)}{z}$ along with the fact that $\frac{\varphi_p'(z)}{\varphi_p(z)} - \frac{\psi_p'(z)}{\psi_p(z)}$ is an exact derivative (of $\ln(\varphi_p/\psi_p)$), the aforementioned convergence can be equivalently written

$$\frac{1}{2\pi\imath}\oint_{\Gamma_\nu} m_\nu(z)dz - \frac{-1}{2\pi\imath}\oint_{\Gamma}\left(\frac{\varphi_p'(z)}{\varphi_p(z)} - \frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\varphi_p(z)}{zc_1}dz \xrightarrow{\text{a.s.}} 0.$$

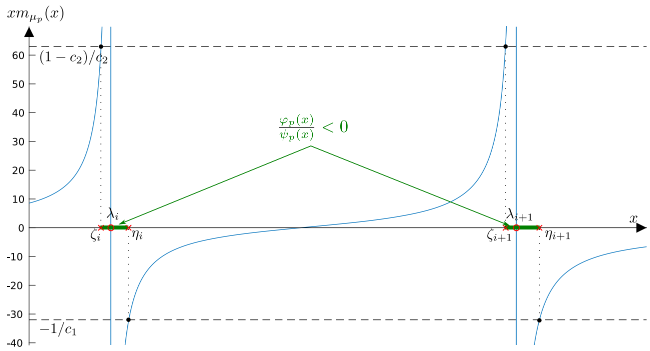

Fig. B.4. Visual representation of $x \mapsto M_p(x) = xm_{\mu_p}(x)$; here for $p = 4$, $n_1 = 8$, $n_2 = 16$. Solutions to $M_p(x) = -1/c_1$ (i.e., η_i's) and to $M_p(x) = (1-c_2)/c_2$ (i.e., ζ_i's) indicated in red crosses. Green solid lines indicate sets of negative φ_p/ψ_p.

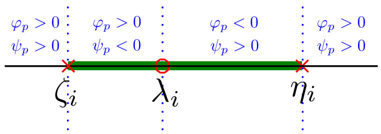

Fig. B.5. Visual representation of the signs of φ_p and ψ_p around singularities.

Reproducing the same line of argument, with an expansion of $(\frac{\varphi_p'(z)}{\varphi_p(z)} - \frac{\psi_p'(z)}{\psi_p(z)})\frac{\varphi_p(z)}{zc_1}$ equivalent to (B.1), the same conclusion arises and we then proved that both ζ_1 and η_p, along with all other ζ_i's and η_i's, are asymptotically found within the interior of Γ.

One can also establish that, on its restriction to $\mathbb{R}^+$, φ_p is everywhere positive but on the set $\cup_{i=1}^p(\lambda_i, \eta_i)$. Similarly, ψ_p is everywhere positive but on the set $\cup_{i=1}^p(\zeta_i, \lambda_i)$. As a consequence, the ratio φ_p/ψ_p is everywhere positive on $\mathbb{R}^+$ but on the set $\cup_{i=1}^p(\zeta_i, \eta_i)$. These observations are synthesized in Fig. B.5.

In terms of monotonicity, since $\psi_p(x) = 1 - c_2\int \frac{t}{t-x}d\mu_p(t)$, ψ_p is decreasing. As for φ_p, note that

$$\varphi_p'(x) = 1 + 2c_1\int \frac{x}{t-x}d\mu_p(t) + c_1\int \frac{x^2}{(t-x)^2}d\mu_p(t) = \int \frac{t^2 - 2(1-c_1)xt + (1-c_1)x^2}{(t-x)^2}d\mu_p(t).$$

Since $c_1 < 1$, we have $1 - c_1 > (1-c_1)^2$, and therefore $\varphi_p'(x) > \int (t-(1-c_1)x)^2/(t-x)^2 d\mu_p(t) > 0$, ensuring that φ_p is increasing on its restriction to $\mathbb{R}$.

Showing that $x \mapsto \varphi_p(x)/\psi_p(x)$ is increasing is important for the study of the case $f(t) = \ln(1+st)$ but is far less direct. This unfolds from the following remark, also of key importance in the following.

Remark 9 (*Alternative form of φ_p and ψ_p*)**.** It is interesting to note that, in addition to the zero found at $z = 0$ for φ_p, we have enumerated all zeros and poles of the rational functions φ_p and ψ_p (this can be ensured from their definition as rational functions) and it thus comes that

$$\varphi_p(z) = (1-c_1)z\frac{\prod_{j=1}^p(z-\eta_j)}{\prod_{j=1}^p(z-\lambda_j)}, \quad \psi_p(z) = \frac{\prod_{j=1}^p(z-\zeta_j)}{\prod_{j=1}^p(z-\lambda_j)}, \quad \frac{\varphi_p(z)}{\psi_p(z)} = (1-c_1)z\frac{\prod_{j=1}^p(z-\eta_j)}{\prod_{j=1}^p(z-\zeta_j)}. \tag{B.2}$$

where the constants $1 - c_1$ and 1 unfold from $\varphi_p(x)/x \to 1 - c_1$ and $\psi_p(x) \to 1$, as $z = x \in \mathbb{R} \to \infty$.

A further useful observation is that the η_i's are the eigenvalues of $\Lambda - \sqrt{\lambda}\sqrt{\lambda}^{\top}/(p-n_1)$, where $\Lambda = \mathrm{diag}(\{\lambda_i\}_{i=1}^p)$ and $\lambda = (\lambda_1, \ldots, \lambda_p)^{\top}$. Indeed, these eigenvalues are found by solving

$$\begin{aligned} 0 &= \det\left(\Lambda - \frac{\sqrt{\lambda}\sqrt{\lambda}^{\top}}{p-n_1} - xI_p\right) = \det(\Lambda - xI_p)\det\left(I_p - (\Lambda - xI_p)^{-1}\frac{\sqrt{\lambda}\sqrt{\lambda}^{\top}}{p-n_1}\right) \\ &= \det(\Lambda - xI_p)\left(1 - \frac{1}{p-n_1}\sqrt{\lambda}^{\top}(\Lambda - xI_p)^{-1}\sqrt{\lambda}\right) = \det(\Lambda - xI_p)\left(1 - \frac{1}{p-n_1}\sum_{i=1}^p \frac{\lambda_i}{\lambda_i - x}\right) \end{aligned}$$

which, for x away from the λ_i (not a solution to $\varphi_p(x) = 0$), reduces to $\frac{1}{p}\sum_{i=1}^p \frac{\lambda_i}{\lambda_i - x} = 1 - \frac{1}{c_1}$, which is exactly equivalent to $m_{\mu_p}(x) = -\frac{1}{c_1 x}$, i.e., $\varphi_p(x) = 0$. Similarly, the ζ_i's are the eigenvalues of the matrix $\Lambda - \sqrt{\lambda}\sqrt{\lambda}^{\top}/n_2$.

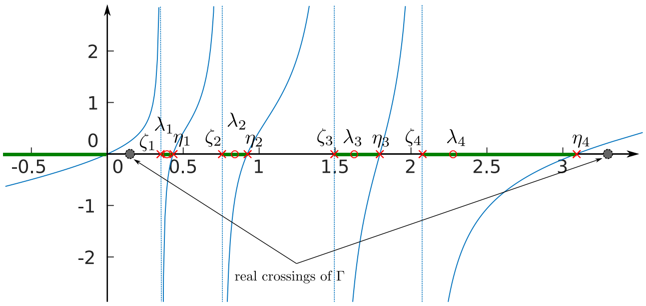

Fig. B.6. Example of visual representation of $\varphi_p/\psi_p : \mathbb{R} \to \mathbb{R}$, $x \mapsto \varphi_p(x)/\psi_p(x)$; here for $p = 4$, $n_1 = 8$, $n_2 = 16$. In green solid lines are stressed the sets over which $\varphi_p(x)/\psi_p(x) < 0$ (which correspond to branch cuts in the study of $f(z) = \ln^k(z)$). Possible real crossings of the contour Γ are indicated, notably showing that no branch cut is passed through when $f(z) = \ln^k(z)$.

These observations allow for the following useful characterization of φ_p/ψ_p:

$$\frac{\varphi_p(z)}{\psi_p(z)} = (1-c_1)z\frac{\det\left(zI_p - \Lambda - \frac{1}{n_1-p}\sqrt{\lambda}\sqrt{\lambda}^{\top}\right)}{\det\left(zI_p - \Lambda + \frac{1}{n_2}\sqrt{\lambda}\sqrt{\lambda}^{\top}\right)} = (1-c_1)z\left(1 - \frac{n_1+n_2-p}{n_2(n_1-p)}\sqrt{\lambda}^{\top}\left(zI_p - \Lambda + \frac{1}{n_2}\sqrt{\lambda}\sqrt{\lambda}^{\top}\right)^{-1}\sqrt{\lambda}\right)$$

after factoring out the matrix in denominator from the determinant in the numerator. The derivative of this expression is

$$\left(\frac{\varphi_p(z)}{\psi_p(z)}\right)' = (1-c_1)\left(1 + \frac{n_1+n_2-p}{n_2(n_1-p)}\sqrt{\lambda}^{\top}Q\left(\Lambda - \frac{1}{n_2}\sqrt{\lambda}\sqrt{\lambda}^{\top}\right)Q\sqrt{\lambda}\right).$$

for $Q = (zI_p - \Lambda + \frac{1}{n_2}\sqrt{\lambda}\sqrt{\lambda}^{\top})^{-1}$. Since $\Lambda - \frac{1}{n_2}\sqrt{\lambda}\sqrt{\lambda}^{\top}$ is positive definite (its eigenvalues being the ζ_i's), on the real axis the derivative is greater than $1 - c_1 > 0$ and the function $x \mapsto \varphi_p(x)/\psi_p(x)$ is therefore increasing. Fig. B.6 displays the behavior of φ_p/ψ_p when restricted to the real axis.

Since we now know that the contour Γ from Theorem 1 encloses exactly all η_i's and ζ_i's, it is sensible to evaluate the residues for these values when $f(z)$ is analytic in their neighborhood.

Remark 10 (*Residue for f Analytic at η_i and ζ_i*). If f is analytic with no singularity at zero, then the integral has a residue at η_i easily found to be $\mathrm{Res}(\eta_i) = f(0)(c_1 + c_2 - c_1c_2)/c_2$. Similarly, if $f(\omega)$ has a well defined limit as $|\omega| \to \infty$, then no residue is found at ζ_i.

As a consequence of Remarks 8 and 10, we have the following immediate corollary.

Remark 11 (*The case $f(t) = t$*). In the case where $f(t) = t$, a singularity appears at ζ_i, which is nonetheless easily treated by noticing that the integrand then reduces to

$$f\left(\frac{\varphi_p(z)}{\psi_p(z)}\right)\left(\frac{\varphi_p'(z)}{\varphi_p(z)} - \frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi_p(z)}{c_2} = \frac{\varphi_p'(z)}{c_2} - \frac{\psi_p'(z)\varphi_p(z)}{c_2\psi_p(z)}$$

and thus, with $\psi_p(z) = (z-\zeta_i)\psi_p'(\zeta_i) + O((z-\zeta_i)^2)$, we get $\mathrm{Res}_{\{f(t)=t\}}(\zeta_i) = -\frac{1}{c_2\varphi_p(\zeta_i)} = -\zeta_i(c_1 + c_2 - c_1c_2)/c_2^2$. Together with Remarks 8 and 10, along with the fact that Γ encloses all η_i and λ_i, for $i = 1, \ldots, p$, we then find

$$\int t d\nu_p(t) - \left[\frac{c_1 + c_2 - c_1c_2}{c_2^2}\sum_{i=1}^{p}(\lambda_i - \zeta_i) - \frac{c_1}{c_2}\frac{1}{p}\sum_{i=1}^{p}\lambda_i\right] \xrightarrow{\text{a.s.}} 0.$$

Noticing that $\sum_i \zeta_i = \mathrm{tr}\,(\Lambda - \frac{1}{n_2}\sqrt{\lambda}\sqrt{\lambda}^{\top}) = (1 - c_2/p)\sum_i \lambda_i$, we retrieve Corollary 1.

B.3. Development for $f(t) = \ln(t)$

The case $f(t) = \ln(t)$ leads to an immediate simplification since, then, $\ln\det(\hat{C}_1^{-1}\hat{C}_2) = \ln\det(\hat{C}_2) - \ln\det(\hat{C}_1)$; one may then use previously established results from the random matrix literature (e.g., the G-estimators in [13] or more recently [16]) to obtain the sought-for estimate. Nonetheless, the full explicit derivation of the contour integral in this case is quite instructive and, being simpler than the subsequent cases where $f(t) = \ln^2(t)$ or $f(t) = \ln(1 + st)$ that rely on the same key ingredients, we shall here conduct a thorough complex integral calculus.

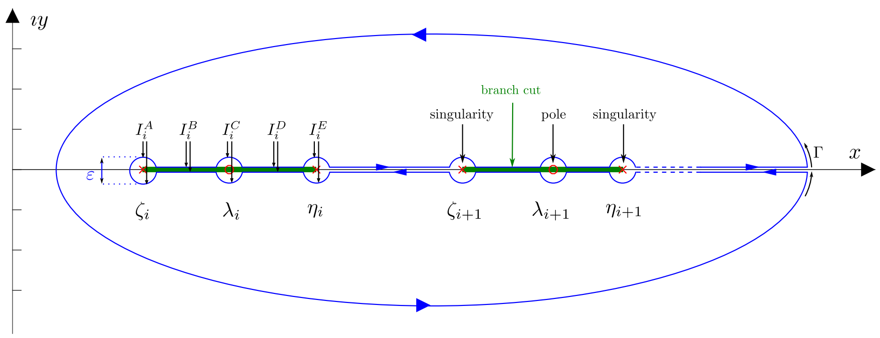

Fig. B.7. Chosen integration contour. The set I_i^B is the disjoint union of the segments $[\zeta_i+\varepsilon+0^+\imath, \lambda_i-\varepsilon+0^+\imath]$ and $[\zeta_i+\varepsilon+0^-\imath, \lambda_i-\varepsilon+0^-\imath]$. Similarly the set I_i^D is the disjoint union of the segments $[\lambda_i+\varepsilon+0^+\imath, \eta_i-\varepsilon+0^+\imath]$ and $[\lambda_i+\varepsilon+0^-\imath, \eta_i-\varepsilon+0^-\imath]$. The sets I_i^A, I_i^C and I_i^E are the disjoint unions of semi-circles (in the upper- or lower-half complex plane) of diameters ε surrounding ζ_i, λ_i and η_i respectively.

For $z\in\mathbb{C}$, define first $f(z)=\ln(z)$ where $\ln(z)=\ln(|z|)e^{i\arg(z)}$, with $\arg(z)\in(-\pi,\pi]$. For this definition of the complex argument, since $\varphi_p(x)/\psi_p(x)$ is everywhere positive but on $\cup_{i=1}^p(\zeta_i,\eta_i)$, we conclude that $\arg(\varphi_p(z)/\psi_p(z))$ abruptly moves from π to $-\pi$ as z moves from $x+0^+\imath$ to $x+0^-\imath$ for all $x\in\cup_{i=1}^p(\zeta_i,\eta_i)$. This creates a set of p branch cuts $[\zeta_i,\eta_i]$, $i=1,\ldots,p$ as displayed in Fig. B.6. This naturally leads to computing the complex integral estimate of $\int f d\nu$ based on the contour displayed in Fig. B.7, which avoids the branch cuts.

This contour encloses no singularity of the integrand and therefore has a null integral. With the notations of Fig. B.7, the sought-for integral (over Γ) therefore satisfies

$$0=\oint_\Gamma+\sum_{i=1}^p\left(\int_{I_i^A}+\int_{I_i^B}+\int_{I_i^C}+\int_{I_i^D}+\int_{I_i^E}\right).$$

We start by the evaluation of the integrals over I_i^B and I_i^D, which can be similarly handled. To this end, note that, since $\arg(\frac{\varphi_p}{\psi_p})$ moves from π to $-\pi$ across the branch cut, we have

$$\begin{aligned}\frac{1}{2\pi\imath}\int_{I_i^B}&=\frac{1}{2\pi\imath}\int_{\zeta_i+\varepsilon}^{\lambda_i-\varepsilon}\left[\ln\left(-\frac{\varphi_p(x)}{\psi_p(x)}\right)+\imath\pi-\ln\left(-\frac{\varphi_p(x)}{\psi_p(x)}\right)+\imath\pi\right]\left(\frac{\varphi_p'(x)}{\varphi_p(x)}-\frac{\psi_p'(x)}{\psi_p(x)}\right)\frac{\psi_p(x)}{c_2}dx\\&=\int_{\zeta_i+\varepsilon}^{\lambda_i-\varepsilon}\left(\frac{\varphi_p'(x)}{\varphi_p(x)}-\frac{\psi_p'(x)}{\psi_p(x)}\right)\frac{\psi_p(x)}{c_2}dx.\end{aligned}$$

We first exploit the rational form expansion (B.1) of $(\frac{\varphi_p'(z)}{\varphi_p(z)}-\frac{\psi_p'(z)}{\psi_p(z)})\frac{\psi_p(x)}{c_2}$ to obtain the integral over I_i^B

$$\begin{aligned}\frac{1}{2\pi\imath}\int_{I_i^B}&=\left(\frac{1}{p}-\frac{c_1+c_2-c_1c_2}{c_1c_2}\right)\left(\sum_{j\neq i}\ln\left|\frac{\lambda_i-\lambda_j}{\zeta_i-\lambda_j}\right|+\ln\left|\frac{\varepsilon}{\zeta_i-\lambda_i}\right|\right)\\&\quad+\frac{1-c_2}{c_2}\ln\frac{\lambda_i}{\zeta_i}+\frac{c_1+c_2-c_1c_2}{c_1c_2}\sum_{j=1}^p\ln\left|\frac{\lambda_i-\eta_j}{\zeta_i-\eta_j}\right|+o(\varepsilon).\end{aligned}$$

The treatment is similar for the integral over I_i^D which results, after summation of both integrals, to

$$\begin{aligned}\frac{1}{2\pi\imath}\int_{I_i^B\cup I_i^D}&=\left(\frac{1}{p}-\frac{c_1+c_2-c_1c_2}{c_1c_2}\right)\sum_{j=1}^p\ln\left|\frac{\eta_i-\lambda_j}{\zeta_i-\lambda_j}\right|+\frac{1-c_2}{c_2}\ln\frac{\eta_i}{\zeta_i}\\&\quad+\frac{c_1+c_2-c_1c_2}{c_1c_2}\left(\sum_{j\neq i}\ln\left|\frac{\eta_i-\eta_j}{\zeta_i-\eta_j}\right|+\ln\left|\frac{\varepsilon}{\zeta_i-\eta_i}\right|\right)+o(\varepsilon).\end{aligned}$$

Note here the asymmetry in the behavior of the integrand in the neighborhood of ζ_i $(+\varepsilon)$ and η_i $(-\varepsilon)$; in the former edge, the integral is well defined while in the latter it diverges as $\ln\varepsilon$ which must then be maintained.

Summing now over $i\in\{1,\ldots,p\}$, we recognize a series of identities. In particular, from the product form (B.2),

$$\sum_{j=1}^p\sum_{i=1}^p\ln\left|\frac{\eta_i-\lambda_j}{\zeta_i-\lambda_j}\right|=\sum_{j=1}^p\ln\left|\frac{\frac{\psi_p}{\varphi_p}(\lambda_j)}{(1-c_1)\lambda_j}\right|=\ln\left(\frac{c_1}{c_2(1-c_1)}\right),$$

$$\sum_{j=1}^{p} \ln \frac{\eta_i}{\zeta_i} = \lim_{z\to 0} \ln\left(\frac{\frac{\psi_p}{\varphi_p}(z)}{(1-c_1)z}\right) = -\ln\left((1-c_1)(1-c_2)\right)$$

$$\sum_{i=1}^{p}\sum_{j\neq i} \ln\left|\frac{\eta_i-\eta_j}{\zeta_i-\eta_j}\right| + \sum_{i=1}^{p}\ln\left|\frac{1}{\zeta_i-\eta_i}\right| = \lim_{z\to\eta_i}\sum_{j=1}^{p}\ln\left|\frac{\frac{\psi_p}{\varphi_p}(z)}{(1-c_1)z(z-\eta_j)}\right| = \sum_{j=1}^{p}\ln\left|\frac{\left(\frac{\psi_p}{\varphi_p}\right)'(\eta_j)}{(1-c_1)\eta_j}\right|.$$

As such, we now find that

$$\begin{aligned}\frac{1}{2\pi\imath}\sum_{i=1}^{p}\int_{I_i^B\cup I_i^D} &= \ln\left(\frac{c_1}{c_2(1-c_1)}\right) + \frac{c_1+c_2-c_1c_2}{c_1c_2} p\ln\varepsilon \\ &\quad - \frac{1-c_2}{c_2}\ln\left((1-c_1)(1-c_2)\right) - \frac{c_1+c_2-c_1c_2}{c_1c_2}\left(p\ln\left(\frac{c_1}{c_2(1-c_1)}\right) - \sum_{j=1}^{p}\ln\left|\frac{\left(\frac{\psi_p}{\varphi_p}\right)'(\eta_j)}{(1-c_1)\eta_j}\right|\right) + o(\varepsilon).\end{aligned}$$

The diverging term in $\ln\varepsilon$ is compensated by the integral over I_i^E. Indeed, letting $z=\eta_i+\varepsilon e^{\imath\theta}$, we may write

$$\begin{aligned}&\frac{1}{2\pi\imath}\int_{I_i^E}\ln\left(\frac{\varphi_p(z)}{\psi_p(z)}\right)\left(\frac{\varphi_p'(z)}{\varphi_p(z)}-\frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi(z)}{c_2}dz \\ &= \frac{\varepsilon}{2\pi c_2}\left[\int_{\pi}^{0^+}+\int_{0^-}^{-\pi}\right]\ln\left((1-c_1)(\eta_i+\varepsilon e^{\imath\theta})\frac{\prod_{j=1}^{p}(\eta_i-\eta_j+\varepsilon e^{\imath\theta})}{\prod_{j=1}^{p}(\eta_i-\zeta_j+\varepsilon e^{\imath\theta})}\right) \\ &\quad\times\left(\sum_{j=1}^{p}\frac{\frac{1}{p}-\frac{c_1+c_2-c_1c_2}{c_1c_2}}{\eta_i+\varepsilon e^{\imath\theta}-\lambda_j}+\frac{\frac{1-c_2}{c_2}}{\eta_i+\varepsilon e^{\imath\theta}}+\sum_{j=1}^{p}\frac{\frac{c_1+c_2-c_1c_2}{c_1c_2}}{\eta_i+\varepsilon e^{\imath\theta}-\eta_j}\right)e^{\imath\theta}d\theta.\end{aligned}$$

To evaluate the small ε limit of this term, first remark importantly that, for small ε, the term in the logarithm equals

$$(1-c_1)\frac{\eta_i}{\eta_i-\zeta_i}\frac{\prod_{j\neq i}(\eta_i-\eta_j)}{\prod_{j\neq i}(\eta_i-\zeta_j)}\varepsilon e^{\imath\theta}+o(\varepsilon)$$

the argument of which equals that of θ. As such, on the integral over $(\pi,0)$, the log term reads $\ln|\cdot|+\imath\theta+o(\varepsilon)$, while on $(0,-\pi)$, it reads $\ln|\cdot|-\imath\theta+o(\varepsilon)$. With this in mind, keeping only the non-vanishing terms in the small ε limit (that is: the term in $\ln\varepsilon$ and the term in $\frac{1}{\varepsilon}$) leads to

$$\frac{1}{2\pi\imath}\int_{I_i^E}\ln\left(\frac{\varphi_p(z)}{\psi_p(z)}\right)\left(\frac{\varphi_p'(z)}{\varphi_p(z)}-\frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi(z)}{c_2}dz = \frac{c_1+c_2-c_1c_2}{c_1c_2}\ln\varepsilon + \frac{c_1+c_2-c_1c_2}{c_1c_2}\ln\left|\left(\frac{\varphi_p}{\psi_p}\right)'(\eta_i)\right| + o(\varepsilon)$$

where we used the fact that $\lim_{\varepsilon\to 0}\frac{1}{\varepsilon e^{\imath\theta}}\left(\frac{\varphi_p}{\psi_p}\right)(\eta_i+\varepsilon e^{\imath\theta}) = \left(\frac{\varphi_p}{\psi_p}\right)'(\eta_i)$.

We proceed similarly to handle the integral over I_i^C

$$\begin{aligned}&\frac{1}{2\pi\imath}\int_{I_i^C}\ln\left(\frac{\varphi_p(z)}{\psi_p(z)}\right)\left(\frac{\varphi_p'(z)}{\varphi_p(z)}-\frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi(z)}{c_2}dz \\ &= \frac{\varepsilon}{2\pi c_2}\left[\int_{\pi}^{0^+}+\int_{0^-}^{-\pi}\right]\ln\left((1-c_1)(\lambda_i+\varepsilon e^{\imath\theta})\frac{\prod_{j=1}^{p}(\lambda_i-\eta_j+\varepsilon e^{\imath\theta})}{\prod_{j=1}^{p}(\lambda_i-\zeta_j+\varepsilon e^{\imath\theta})}\right) \\ &\quad\times\left(\sum_{j=1}^{p}\frac{\frac{1}{p}-\frac{c_1+c_2-c_1c_2}{c_1c_2}}{\lambda_i+\varepsilon e^{\imath\theta}-\lambda_j}+\frac{\frac{1-c_2}{c_2}}{\lambda_i}+\varepsilon e^{\imath\theta}+\sum_{j=1}^{p}\frac{\frac{c_1+c_2-c_1c_2}{c_1c_2}}{\lambda_i+\varepsilon e^{\imath\theta}-\eta_j}\right)e^{\imath\theta}d\theta.\end{aligned}$$

Here, for small ε, the angle of the term in the argument of the logarithm is that of

$$\frac{\varphi_p}{\psi_p}(\lambda_i)+\left(\frac{\varphi_p}{\psi_p}\right)'(\lambda_i)\varepsilon e^{\imath\theta}+o(\varepsilon) = -\frac{c_1}{c_2}\lambda_i+\varepsilon e^{\imath\theta}\frac{c_1}{c_2}\left(p\frac{c_1+c_2-c_1c_2}{c_1c_2}-1\right)+o(\varepsilon).$$

That is, for all large p, the argument equals $\pi+o(\varepsilon)<\pi$ uniformly on $\theta\in(0,\pi)$ and $-\pi+o(\varepsilon)>-\pi$ uniformly on $\theta\in(-\pi,0)$; thus the complex logarithm reads $\ln|\cdot|+\imath\theta+o(\varepsilon)$ on $(\pi,0)$, while on $(0,-\pi)$, it reads $\ln|\cdot|-\imath\theta+o(\varepsilon)$. Proceeding as previously for the integral over I_i^E, we then find after calculus that

$$\frac{1}{2\pi\imath}\int_{I_i^C}\ln\left(\frac{\varphi_p(z)}{\psi_p(z)}\right)\left(\frac{\varphi_p'(z)}{\varphi_p(z)}-\frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi(z)}{c_2}dz = \left(\frac{c_1+c_2-c_1c_2}{c_1c_2}-\frac{1}{p}\right)\ln\left(\frac{c_1}{c_2}\lambda_i\right).$$

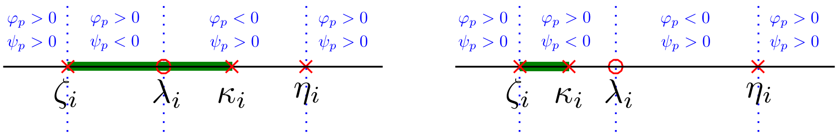

Fig. B.8. Visual representation of the signs of φ_p and ψ_p around singularities for the function $f(t) = \ln(1+st)$. Left: case where $\kappa_i > \lambda_i$. Right: case where $\kappa_i > \lambda_i$.

Note that this expression is reminiscent of a "residue" at λ_i (with negatively oriented contour), according to Remark 8, however for the function $\ln|\cdot|$ and not for the function $\ln(\cdot)$, due to the branch cut passing through λ_i.

The final integral over I_i^A is performed similarly. However, here, it is easily observed that the integral is of order $O(\varepsilon \ln \varepsilon)$ in the small ε limit, and thus vanishes.

Finally, summing up all contributions, we have

$$\begin{aligned}
\frac{1}{2\pi\imath}\oint_{\Gamma} &= -\sum_{i=1}^{p}\frac{1}{2\pi\imath}\left(\int_{I_i^A}+\int_{I_i^B}+\int_{I_i^C}+\int_{I_i^D}+\int_{I_i^E}\right)\\
&= -\ln\left(\frac{c_1(1-c_2)}{c_2}\right)+\frac{\ln((1-c_1)(1-c_2))}{c_2}+\frac{1}{p}\sum_{i=1}^{p}\ln\left(\frac{c_1}{c_2}\lambda_i\right)\\
&\quad -\frac{c_1+c_2-c_1c_2}{c_1c_2}\left(p\ln(1-c_1)+\sum_{i=1}^{p}\ln\lambda_i-\sum_{i=1}^{p}\ln((1-c_1)\eta_i)\right)\\
&= -\ln(1-c_2)+\frac{1}{c_2}\ln((1-c_1)(1-c_2))+\frac{1}{p}\sum_{i=1}^{p}\ln\lambda_i+\frac{c_1+c_2-c_1c_2}{c_1c_2}\sum_{i=1}^{p}\ln\left(\frac{\eta_i}{\lambda_i}\right)\\
&= \frac{1}{p}\sum_{i=1}^{p}\ln\lambda_i+\frac{1-c_2}{c_2}\ln(1-c_2)-\frac{1-c_1}{c_1}\ln(1-c_1)
\end{aligned}$$

where in the last equality we used, among other algebraic simplifications, the fact that $\sum_{i=1}^{p}\ln(\frac{\eta_i}{\lambda_i}) = \lim_{x\to 0}\ln(\frac{\psi_p(x)}{(1-c_1)x}) = -\ln(1-c_1)$. This is the sought-for result.

B.4. Development for $f(t) = \ln(1+st)$

The development for $f(t) = \ln(1+st)$ is quite similar to that of $f(t) = \ln(t)$, with some noticeable exceptions with respect to the position of singularity points.

A few important remarks are in order to start with this scenario. First note from Fig. B.5 and the previous discussions that the function $z \mapsto \ln(1+s\varphi_p(z)/\psi_p(z))$ has a singularity at $z = \kappa_i$, $i = 1,\ldots,p$, for some $\kappa_i \in (\zeta_i, \eta_i)$ solution to $1+s\varphi_p(x)/\psi_p(x) = 0$ (indeed, $\varphi_p(x)/\psi_p(x)$ is increasing on (ζ_i, η_i) with opposite asymptotes and thus κ_i exists and is uniquely defined). In addition, $\ln(1+s\varphi_p(z)/\psi_p(z))$ has a further singularity satisfying $1+s\varphi_p(x)/\psi_p(x) = 0$ in the interval $(-\infty, 0)$ which we shall denote κ_0.

A few identities regarding κ_i are useful. Using the relation between φ_p and ψ_p, we find in particular that

$$\varphi_p(\kappa_i) = -\frac{1}{s}\frac{c_1+c_2-c_1c_2}{c_2}\frac{\kappa_i}{-\frac{1}{s}+\frac{c_1}{c_2}\kappa_i},\quad \psi_p(\kappa_i) = \frac{c_1+c_2-c_1c_2}{c_2}\frac{\kappa_i}{-\frac{1}{s}+\frac{c_1}{c_2}\kappa_i},\quad (\psi_p+s\varphi_p)\left(\frac{c_2}{c_1 s}\right) = \frac{c_1+c_2-c_1c_2}{c_1}.$$

With the discussions above, we also find that

$$1+s\frac{\varphi_p(z)}{\psi_p(z)} = (1-c_1)s(z-\kappa_0)\frac{\prod_{i=1}^{p}(z-\kappa_i)}{\prod_{i=1}^{p}(z-\zeta_i)},\quad \psi_p(z)+s\varphi_p(z) = s(1-c_1)(z-\kappa_0)\frac{\prod_{i=1}^{p}(z-\kappa_i)}{\prod_{i=1}^{p}(z-\lambda_i)}. \tag{B.3}$$

Note now importantly that $\lambda_i > \frac{c_1}{c_2 s}$ is equivalent to $-\frac{c_2}{c_1}\lambda_i < -\frac{1}{s}$ which is also $\varphi_p(\lambda_i)/\psi_p(\lambda_i) < \varphi_p(\kappa_i)/\psi_p(\kappa_i)$; then, as φ_p/ψ_p is increasing, $\lambda_i > \frac{c_1}{c_2 s}$ is equivalent to $\lambda_i < \kappa_i$. On the opposite, for $\lambda_i < \frac{c_1}{c_2 s}$, we find $\lambda_i > \kappa_i$. As such, to evaluate the contour integral in this setting, one must isolate two sets of singularities (see Fig. B.8): (i) those for which $\kappa_i > \lambda_i$ (which are all the largest indices i for which $\lambda_i > \frac{c_1}{c_2 s}$) and (ii) those for which $\kappa_i < \lambda_i$. This affects the relative position of the branch cut with respect to λ_i and therefore demands different treatments. In particular, the integrals over I_i^B and I_i^D may be restricted to integrals over shorter (possibly empty) segments. Nonetheless, the calculus ultimately reveals that, since the branch cut does not affect the local behavior of the integral around λ_i, both cases entail the same result. In particular, in case (i) where $\lambda_i > \kappa_i$, recalling (B.1), one only has to evaluate

$$\int_{\zeta_i+\varepsilon}^{\kappa_i-\varepsilon} \left(\frac{\varphi_p'(x)}{\varphi_p(x)} - \frac{\psi_p'(x)}{\psi_p(x)}\right) \frac{\psi_p(x)}{c_2} dx$$
$$= \int_{\zeta_i+\varepsilon}^{\kappa_i-\varepsilon} \left(\frac{1}{p} - \frac{c_1+c_2-c_1c_2}{c_1c_2}\right) \sum_{j=1}^p \frac{1}{x-\lambda_j} + \frac{1-c_2}{c_2}\frac{1}{x} + \frac{c_1+c_2-c_1c_2}{c_1c_2} \sum_{j=1}^p \frac{1}{x-\eta_j} dx$$
$$= \frac{1}{p}\sum_{j=1}^p \ln\left|\frac{\kappa_i-\lambda_j}{\zeta_i-\lambda_j}\right| + \frac{c_1+c_2-c_1c_2}{c_1c_2}\sum_{j=1}^p \left(\ln\left|\frac{\kappa_i-\eta_j}{\kappa_i-\lambda_j}\right| - \ln\left|\frac{\zeta_i-\eta_j}{\zeta_i-\lambda_j}\right|\right) + \frac{1-c_2}{c_2}\ln\left|\frac{\kappa_i}{\zeta_i}\right| + o(\varepsilon).$$

In case (ii), subdividing the integral as $\int_{\zeta_i+\varepsilon}^{\lambda_i-\varepsilon} + \int_{\lambda_i+\varepsilon}^{\kappa_i-\varepsilon}$ brings immediate simplification of the additional terms in λ_i and thus the result remains the same.

The integral over I_i^C is slightly more delicate to handle. In case (i), in the limit of small ε,

$$1 + s\frac{\varphi_p}{\psi_p}(\lambda_i + \varepsilon e^{i\theta}) = 1 - s\frac{c_1}{c_2}\lambda_i + \varepsilon s\frac{c_1}{c_2}\left(p\frac{c_1+c_2-c_1c_2}{c_1c_2} - 1\right)e^{i\theta} + o(\varepsilon)$$

the angle of which is $0 + o(\varepsilon)$ uniformly on $\theta \in (-\pi, \pi]$ (since $1 - s\frac{c_1}{c_2}\lambda_i > 0$). As such, for all small ε, the sum of the integrals over $(-\pi, 0)$ and $(0, \pi]$ reduces to the integral over $(-\pi, \pi]$, leading up to a mere residue calculus, and

$$\frac{1}{2\pi\imath}\oint_{I_i^C} \ln\left(1 + s\frac{\varphi_p(z)}{\psi_p(z)}\right)\left(\frac{\varphi_p'(z)}{\varphi_p(z)} - \frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi_p(z)}{c_2}dz = \ln\left(1 - s\frac{c_1}{c_2}\lambda_i\right)\left(\frac{c_1+c_2-c_1c_2}{c_1c_2} - \frac{1}{p}\right) + o(\varepsilon).$$

In case (ii), $1 - s\frac{c_1}{c_2}\lambda_i < 0$ and thus the angle of $1 + s\frac{\varphi_p}{\psi_p}(\lambda_i + \varepsilon e^{i\theta})$ is close to π; for $\theta \in (0, \pi)$, this leads to an argument equal to $\pi + o(\varepsilon) < \pi$ and for $\theta \in (-\pi, 0)$ to an argument equal to $-\pi + o(\varepsilon) > -\pi$. All calculus made, we then find that in either case (i) or (ii)

$$\frac{1}{2\pi\imath}\oint_{I_i^C} \ln\left(1 + s\frac{\varphi_p(z)}{\psi_p(z)}\right)\left(\frac{\varphi_p'(z)}{\varphi_p(z)} - \frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi_p(z)}{c_2}dz = \ln\left|1 - s\frac{c_1}{c_2}\lambda_i\right|\left(\frac{c_1+c_2-c_1c_2}{c_1c_2} - \frac{1}{p}\right) + o(\varepsilon).$$

As in the case of $f(t) = \ln(t)$, the integral over I_i^A is of order $o(\varepsilon)$ and vanishes. As a consequence, summing over $i \in \{1, \ldots, p\}$, we find that

$$\frac{1}{2\pi\imath}\oint_\Gamma = -\frac{1}{p}\sum_{i,j=1}^p \ln\left|\frac{\kappa_i-\lambda_j}{\zeta_i-\lambda_j}\right| - \frac{1-c_2}{c_2}\sum_{i=1}^p \ln\frac{\kappa_i}{\zeta_i} + \frac{c_1+c_2-c_1c_2}{c_1c_2}\sum_{i,j=1}^p\left(\ln\left|\frac{\zeta_i-\eta_j}{\zeta_i-\lambda_j}\right| - \ln\left|\frac{\kappa_i-\eta_j}{\kappa_i-\lambda_j}\right|\right)$$
$$- \left(\frac{c_1+c_2-c_1c_2}{c_1c_2} - \frac{1}{p}\right)\sum_{i=1}^p \ln\left|1 - s\frac{c_1}{c_2}\lambda_j\right| + o(\varepsilon).$$

Before reaching the final result, note that, from (B.3),

$$\sum_{i=1}^p \frac{1}{p}\sum_{j=1}^p \ln\frac{|\kappa_i-\lambda_j|}{|\zeta_i-\lambda_j|} = \frac{1}{p}\sum_{j=1}^p \ln\left|\left(1 + s\frac{\varphi_p(\lambda_j)}{\psi_p(\lambda_j)}\right)\frac{1}{\lambda_j-\kappa_0}\frac{1}{(1-c_1)s}\right|$$
$$= \frac{1}{p}\sum_{j=1}^p \ln\left|1 - \frac{c_1}{c_2}s\lambda_i\right| - \frac{1}{p}\sum_{j=1}^p \ln(\lambda_j - \kappa_0) - \ln((1-c_1)s)$$

and similarly

$$\sum_{i,j=1}^p \ln\left|\frac{\zeta_i-\eta_j}{\zeta_i-\lambda_j}\right| = \sum_{i=1}^p \ln\left|\frac{\varphi_p(\zeta_i)}{(1-c_1)\zeta_i}\right| = p\ln\left|\frac{c_1+c_2-c_1c_2}{c_2(1-c_1)}\right|,$$
$$\sum_{i=1}^p \ln\frac{\kappa_i}{\zeta_i} = \ln\left(\frac{1 + s\frac{\varphi_p}{\psi_p}(0)}{-(1-c_1)s\kappa_0}\right) = -\ln\left(-(1-c_1)s\kappa_0\right)$$
$$\sum_{i,j=1}^p \ln\left|\frac{\kappa_i-\eta_j}{\kappa_i-\lambda_j}\right| = \sum_{i=1}^p \ln\left|\frac{\varphi_p(\kappa_i)}{(1-c_1)\kappa_i}\right| = \sum_{i=1}^p \ln\left|\frac{c_1+c_2-c_1c_2}{c_2(1-c_1)}\frac{1}{1 - \frac{c_1}{c_2 s}\kappa_i}\right|$$

Using now (B.3) (right), we find that

$$\sum_{i=1}^p \ln\left(\frac{1 - s\frac{c_1}{c_2}\lambda_i}{1 - s\frac{c_1}{c_2}\kappa_i}\right) = \sum_{i=1}^p \ln\left(\frac{\frac{c_2}{c_1 s} - \lambda_i}{\frac{c_2}{c_1 s} - \kappa_i}\right) = \ln\left(\frac{\psi_p\left(\frac{c_2}{c_1 s}\right) + s\varphi_p\left(\frac{c_2}{c_1 s}\right)}{s(1-c_1)\left(\frac{c_2}{c_1 s} - \kappa_0\right)}\right) = \ln\left(\frac{c_1+c_2-c_1c_2}{sc_1(1-c_1)\left(\frac{c_2}{c_1 s} - \kappa_0\right)}\right).$$

Combining the previous results and remarks then leads to

$$\frac{1}{2\pi\imath}\oint_{\Gamma}=\frac{c_1+c_2-c_1c_2}{c_1c_2}\ln\left(\frac{c_1+c_2-c_1c_2}{(1-c_1)(c_2-sc_1\kappa_0)}\right)+\frac{1-c_2}{c_2}\ln\left(-s\kappa_0(1-c_1)\right)+\ln((1-c_1)s)+\frac{1}{p}\sum_{i=1}^{p}\ln(\lambda_i-\kappa_0)$$
$$=\frac{c_1+c_2-c_1c_2}{c_1c_2}\ln\left(\frac{c_1+c_2-c_1c_2}{(1-c_1)(c_2-sc_1\kappa_0)}\right)+\frac{1}{c_2}\ln\left(-s\kappa_0(1-c_1)\right)+\frac{1}{p}\sum_{i=1}^{p}\ln\left(1-\frac{\lambda_i}{\kappa_0}\right).$$

This concludes the proof for the case $c_1>0$. In the limit where $c_1\to 0$, it suffices to use the Taylor expansion of the leftmost logarithm in the small c_1 limit, i.e., $\ln(c_2(1-c_1)+c_1)\sim\ln(c_2(1-c_1))+c_1/(c_2(1-c_1))$ and $\ln(c_2(1-c_1)-sc_1\kappa_0(1-c_1))\sim\ln(c_2(1-c_1))-sc_1\kappa_0/c_2$.

B.5. Development for $f(t)=\ln^2(t)$

The function $f(t)=\ln^2(t)$ is at the core of the Fisher distance and is thus of prime importance in many applications. The evaluation of the complex integral in Theorem 1 for this case is however quite technical and calls for the important introduction of the dilogarithm function. We proceed with this introduction first and foremost.

B.5.1. The dilogarithm function

The (real) dilogarithm is defined as the function $\mathrm{Li}_2(x)=-\int_0^x\{\ln(1-u)/u\}du$ for $x\in(-\infty,1]$. The dilogarithm function intervenes in many instances of the evaluation of the contour integral of Theorem 1, through the subsequently defined function $F(X,Y;a)$. This function assumes different formulations depending on the relative position of X,Y,a on the real axis.

Lemma 1 (*Dilogarithm Integrals*)**.** *We have the following results and definition*

$$(X,Y\geq a>0)\quad\int_Y^X\frac{\ln(x-a)}{x}dx\equiv F(X,Y;a)=\mathrm{Li}_2\left(\frac{a}{X}\right)-\mathrm{Li}_2\left(\frac{a}{Y}\right)+\frac{1}{2}\left[\ln^2(X)-\ln^2(Y)\right]$$
$$(X,Y>0>a)\quad\int_Y^X\frac{\ln(x-a)}{x}dx\equiv F(X,Y;a)=-\mathrm{Li}_2\left(\frac{X}{a}\right)+\mathrm{Li}_2\left(\frac{Y}{a}\right)+\ln\left(\frac{X}{Y}\right)\ln(-a)$$
$$(a>X,Y,0\ \&\ XY>0)\quad\int_Y^X\frac{\ln(a-x)}{x}dx\equiv F(-X,-Y;-a)=-\mathrm{Li}_2\left(\frac{X}{a}\right)+\mathrm{Li}_2\left(\frac{Y}{a}\right)+\ln\left(\frac{X}{Y}\right)\ln(a)$$
$$(X,Y>0)\quad\int_Y^X\frac{\ln(x)}{x}dx\equiv F(X,Y;0)=\frac{1}{2}\ln^2(X)-\frac{1}{2}\ln^2(Y).$$

Lemma 2 (*Properties of Dilogarithm Functions [30, Section I-2]*)**.** *The following relations hold*

$$(x<0)\quad\mathrm{Li}_2\left(\frac{1}{x}\right)+\mathrm{Li}_2(x)=-\frac{1}{2}\ln^2(-x)-\frac{\pi^2}{6}$$
$$(0<x<1)\quad\mathrm{Li}_2(1-x)+\mathrm{Li}_2(x)=-\ln(x)\ln(1-x)+\frac{\pi^2}{6}$$
$$(0<x<1)\quad\mathrm{Li}_2(1-x)+\mathrm{Li}_2\left(1-\frac{1}{x}\right)=-\frac{1}{2}\ln^2(x)$$
$$(x<1,\varepsilon\ll 1)\quad\mathrm{Li}_2(x+\varepsilon)=\mathrm{Li}_2(x)-\varepsilon\frac{\ln(1-x)}{x}+\varepsilon^2\frac{(1-x)\ln(1-x)+x}{2(1-x)x^2}+O(\varepsilon^3).$$

B.5.2. Integral evaluation

As in the case where $f(t)=\ln(t)$, we shall evaluate the complex integral based on the contour displayed in Fig. B.7. The main difficulty here arises in evaluating the real integrals over the segments I_i^B and I_i^D.

Again, we start from Eq. (B.1). In particular, the integral over I_i^B reads

$$\frac{1}{2\pi\imath}\int_{I_i^B}\ln^2\left(\frac{\varphi_p(z)}{\psi_p(z)}\right)\left(\frac{\varphi_p'(z)}{\varphi_p(z)}-\frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi_p(z)}{c_2}dz$$
$$=2\int_{\zeta_i+\varepsilon}^{\lambda_i-\varepsilon}\ln\left(-\frac{\varphi_p(x)}{\psi_p(x)}\right)\left(\frac{\varphi_p'(x)}{\varphi_p(x)}-\frac{\psi_p'(x)}{\psi_p(x)}\right)\frac{\psi_p(x)}{c_2}dx$$

$$= 2\int_{\zeta_i+\varepsilon}^{\lambda_i-\varepsilon}\left(\ln(1-c_1)+\ln(x)+\sum_{l<i}\ln(x-\eta_l)+\sum_{l>i}\ln(\eta_l-x)+\ln(\eta_i-x)-\sum_{l\leq i}\ln(x-\zeta_l)-\sum_{l>i}\ln(\zeta_l-x)\right)$$
$$\times\left(\left(\frac{1}{p}-\frac{c_1+c_2-c_1c_2}{c_1c_2}\right)\sum_{j=1}^{p}\frac{1}{x-\lambda_j}+\frac{1-c_2}{c_2}\frac{1}{x}+\frac{c_1+c_2-c_1c_2}{c_1c_2}\sum_{j=1}^{p}\frac{1}{x-\eta_j}\right)dx.$$

Note that we have chosen to write the logarithms in such a way that every integral is a well-defined real integral.

Using now the fact that

$$\int_Y^X\frac{\ln(x-a)}{x-b}dx=F(X-b,Y-b;a-b),\qquad\int_Y^X\frac{\ln(a-x)}{x-b}dx=F(b-X,b-Y;b-a)$$

which we apply repetitively, and very carefully, to the previous equality, we find after simplification and addition of the contributions from I_i^D, which is treated together with I_i^B, along with those from I_i^A, I_i^C and I_i^E (details are provided in Supplementary Material)

$$\begin{aligned}
&\frac{1}{2\pi\imath}\oint_\Gamma\ln^2\left(\frac{\varphi_p(z)}{\psi_p(z)}\right)\left(\frac{\varphi_p'(z)}{\varphi_p(z)}-\frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi_p(z)}{c_2}dz\\
&=\frac{c_1+c_2-c_1c_2}{c_1c_2}\left[\sum_{i=1}^{p}\left\{\ln^2\left((1-c_1)\eta_i\right)-\ln^2\left((1-c_1)\lambda_i\right)\right\}\right.\\
&\left.+2\sum_{1\leq i,j\leq p}\left\{\mathrm{Li}_2\left(1-\frac{\zeta_i}{\lambda_j}\right)-\mathrm{Li}_2\left(1-\frac{\eta_i}{\lambda_j}\right)+\mathrm{Li}_2\left(1-\frac{\eta_i}{\eta_j}\right)-\mathrm{Li}_2\left(1-\frac{\zeta_i}{\eta_j}\right)\right\}\right]\\
&-\frac{1-c_2}{c_2}\left[\ln^2(1-c_2)-\ln^2(1-c_1)+\sum_{i=1}^{p}\left\{\ln^2(\eta_i)-\ln^2(\zeta_i)\right\}\right]\\
&-\frac{1}{p}\left[2\sum_{1\leq i,j\leq p}\left\{\mathrm{Li}_2\left(1-\frac{\zeta_i}{\lambda_j}\right)-\mathrm{Li}_2\left(1-\frac{\eta_i}{\lambda_j}\right)\right\}-\sum_{i=1}^{p}\ln^2\left((1-c_1)\lambda_i\right)\right]
\end{aligned}\tag{B.4}$$

which provides an exact, yet rather impractical, final expression for the integral since the expression involves the evaluation of $O(p^2)$ dilogarithm terms which may be computationally intense for large p.

At this point, it is not easy to fathom why the retrieved expression would remain of order $O(1)$ with respect to p. In order to both simplify the expression and retrieve a visually clear $O(1)$ estimate, we next proceed to a large p Taylor expansion of the above result. In particular, from the last item in Lemma 2, we perform a (second order) Taylor expansion of all $\mathrm{Li}_2(1-X)$ terms above in the vicinity of λ_i/λ_j. This results in

$$\sum_{i,j}\mathrm{Li}_2\left(1-\frac{\zeta_i}{\lambda_j}\right)-\mathrm{Li}_2\left(1-\frac{\eta_i}{\lambda_j}\right)+\mathrm{Li}_2\left(1-\frac{\eta_i}{\eta_j}\right)-\mathrm{Li}_2\left(1-\frac{\zeta_i}{\eta_j}\right)=(\Delta_\zeta^\eta)^TM(\Delta_\lambda^\eta)+o_p(1)$$
$$\frac{1}{p}\sum_{i,j}\mathrm{Li}_2\left(1-\frac{\zeta_i}{\lambda_j}\right)-\mathrm{Li}_2\left(1-\frac{\eta_i}{\lambda_j}\right)=-\frac{1}{p}(\Delta_\zeta^\eta)^TN1_p+o_p(1)$$

with Δ_a^b, M and N defined in the statement of Corollary 4. With these developments, we deduce the final approximation

$$\begin{aligned}
&\frac{1}{2\pi\imath}\oint_\Gamma\ln^2\left(-\frac{\varphi_p(z)}{\psi_p(z)}\right)\left(\frac{\varphi_p'(z)}{\varphi_p(z)}-\frac{\psi_p'(z)}{\psi_p(z)}\right)\frac{\psi_p(z)}{c_2}dz\\
&=2\frac{c_1+c_2-c_1c_2}{c_1c_2}\left(\left(\Delta_\zeta^\eta\right)^TM\left(\Delta_\lambda^\eta\right)+\sum_i\frac{\ln\left((1-c_1)\lambda_i\right)}{\lambda_i}(\eta_i-\lambda_i)\right)\\
&-\frac{2}{p}\left(\Delta_\zeta^\eta\right)^TN1_p+\frac{1}{p}\sum_i\ln^2\left((1-c_1)\lambda_i\right)\\
&-2\frac{1-c_2}{c_2}\left(\frac{1}{2}\ln^2(1-c_2)-\frac{1}{2}\ln^2(1-c_1)+\sum_i(\eta_i-\zeta_i)\frac{\ln(\lambda_i)}{\lambda_i}\right)+o_p(1).
\end{aligned}$$

For symmetry, it is convenient to finally observe that $\ln(1-c_1)\sum_i(\eta_i-\zeta_i)/\lambda_i\sim\ln(1-c_1)\sum_i\ln(\eta_i/\zeta_i)=-\ln^2(1-c_1)$; replacing in the last parenthesis provides the result of Corollary 4 for $c_1>0$.

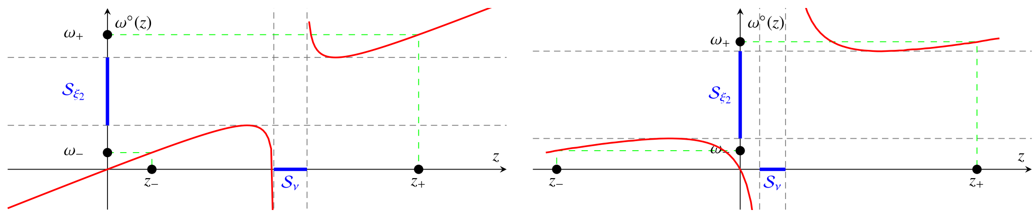

Fig. C.9. Variable change $z \mapsto \omega^\circ(z) = z + c_2^\infty \int \frac{zd\nu(t)}{z-t}$ for $c_2^\infty < 1$ (left) and $c_2^\infty > 1$ (right). $\mathcal{S}_\theta$ is the support of the probability measure θ. For $0 < \omega_- = \omega(z_-) < \inf \mathcal{S}_\nu$, the pre-image z_- is necessarily negative for $c_2^\infty > 1$.

To determine the limit as $c_1 \to 0$, it suffices to remark that in this limit $\eta_i = \lambda_i + \frac{c_1}{p}\lambda_i + o(c_1)$ (this can be established using the functional relation $\varphi_p(\eta_i) = 0$ in the small c_1 limit). Thus it suffices to replace in the above expression the vector $\eta - \zeta$ by the vector $\eta - \lambda$, the vector $\frac{c_1+c_2-c_1c_2}{c_1c_2}(\eta - \lambda)$ by the vector $\frac{1}{p}\lambda$, and taking $c_1 = 0$ in all other instances (where the limits for $c_1 \to 0$ are well defined).

Appendix C. Integration contour determination

This section details the complex integration steps sketched in Appendix A. These details rely heavily on the works of [24] and follow similar ideas as in, e.g., [11].

Our objective is to ensure that the successive changes of variables involved in Appendix A move any complex contour closely encircling the support of μ onto a valid contour encircling the support of ν; we will in particular be careful that the resulting contour, in addition to encircling the support of ν, does not encircle additional values possibly bringing undesired residues (such as 0). We will proceed in two steps, first showing that a contour encircling μ results on a contour encircling ξ_2 and a contour encircling ξ_2 results on a contour encircling ν.

Let us consider a first contour Γ_{ξ_2} closely around the support of ξ_2 (in particular not containing 0). We have to prove that any point ω of this contour is mapped to a point of a contour Γ_ν closely around the support of ν.

The change of variable performed in (A.6) reads, for all $\omega \in \mathbb{C} \setminus \mathrm{Supp}(\xi_2)$,

$$z \equiv z(\omega) = \frac{-\omega}{-(1-c_2^\infty) + c_2^\infty \omega m_{\xi_2}(\omega)} = \frac{-1}{m_{\tilde{\xi}_2}(\omega)}$$

where we recall that $\tilde{\xi}_2 = c_2^\infty \xi_2 + (1-c_2^\infty)\delta_0$. Since $\Im[\omega]\Im[m_{\tilde{\xi}_2}(\omega)] > 0$ for $\Im[\omega] \neq 0$, we already have that $\Im[z]\Im[\omega] > 0$ for all non-real ω.

It therefore remains to show that real ω's outside the support of ξ_2 project onto properly located real z's, i.e., on either side of the support of ν. This conclusion follows from the seminal work [24] on the spectral analysis of sample covariance matrices. The essential idea is to note that, due to (A.5), the relation $z(\omega) = -1/m_{\tilde{\xi}_2}(\omega)$ can be inverted as

$$\omega \equiv \omega(z) = -\frac{1}{m_{\tilde{\xi}_2}} + c_2^\infty \int \frac{td\nu(t)}{1 + tm_{\tilde{\xi}_2}} = z + c_2^\infty \int \frac{td\nu(t)}{1 - \frac{t}{z}}.$$

In [24], it is proved that the image by $\omega(\cdot)$ of $z(\mathbb{R} \setminus \mathrm{Supp}(\xi_2))$ coincides with the increasing sections of the function $\omega^\circ : \mathbb{R} \setminus \mathrm{Supp}() \to \mathbb{R}$, $z \mapsto \omega(z)$. The latter being an explicit function, its functional analysis is simple and allows in particular to properly locate the real pairs (ω, z). Details of this analysis are provided in [24] as well as in [10], which shall not be recalled here. The function ω° is depicted in Fig. C.9; we observe and easily prove that, for $c_2^\infty < 1$, any two values $z_- < \inf(\mathrm{Supp}(\nu)) \leq \sup(\mathrm{Supp}(\nu)) < z_+$ have respectively images ω_- and ω_+ satisfying $w_- < \inf(\mathrm{Supp}(\xi_2)) \leq \sup(\mathrm{Supp}(\xi_2)) < w_+$ as desired. This is however not the case for $c_2^\infty > 1$ where $\{z_-, z_+\}$ enclose not only $\mathrm{Supp}(\nu)$ but also 0 and therefore do not bring a valid contour. This essentially follows from the fact that $(\varphi_p/\psi_p)'(0)$ is positive for $c_2^\infty < 1$ and negative for $c_2^\infty > 1$.

The same reasoning now holds for the second variable change. Indeed, note that here

$$\omega = u(1 + c_1^\infty u m_\mu(u)) = u\left(1 - c_1^\infty - \frac{c_1^\infty}{u} m_{\mu^{-1}}\left(\frac{1}{u}\right)\right) = -m_{\tilde{\mu}^{-1}}\left(\frac{1}{u}\right).$$

Exploiting (A.1) provides, as above, a functional inverse given here by

$$u \equiv u(\omega) = \left(\frac{1}{\omega} + c_1^\infty \int \frac{d\xi_2(t)}{t - \omega}\right)^{-1}$$

and the analysis follows the same arguments as above (see display in Fig. C.10 of the extension to $u^\circ(\omega) = u(\omega)$ for all $\omega \in \mathbb{R} \setminus \mathrm{Supp}(\xi_2)$).

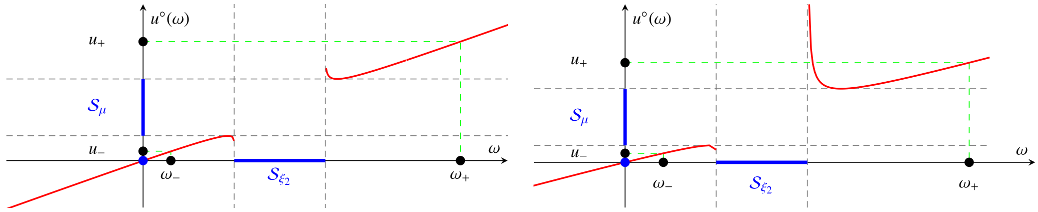

Fig. C.10. Variable change $u^{\circ}(\omega) = (\frac{1}{\omega} + c_1^\infty \int \frac{1}{t-\omega} d\xi_2(t))^{-1}$ for $c_2^\infty < 1$ (left) and $c_2^\infty > 1$ (right). $\mathcal{S}_\theta$ is the support of the probability measure θ.

Appendix D. The case $c_2^\infty > 1$

Allowing for $c_2^\infty > 1$ brings along some key difficulties. First recall from Appendix C that, for a contour Γ surrounding $\mathrm{Supp}(\mu)$, if $c_2^\infty > 1$, the image $(\varphi/\psi)(\Gamma)$ necessarily surrounds $\mathrm{Supp}(\nu) \cup \{0\}$, while for $c_2^\infty < 1$, 0 is excluded from the interior of $(\varphi/\psi)(\Gamma)$ if 0 is not contained within Γ. This implies that, if $f(z)$ has a singularity at $z = 0$, the relation $\int f d\nu(t) = \frac{1}{2\pi\imath} \oint_{(\varphi/\psi)(\Gamma)} f(z) m_\nu(z) dz$ no longer holds. For $f(z)$ a polynomial in z, e.g., for $f(z) = z$, this poses no problem, thereby implying the validity of Corollary 1 for all $c_2^\infty > 0$. For $f(z) = \ln(z)^k$ ($k \geq 1$) and similarly for $f(z) = 1/z$, to the best of our knowledge, there is no recovering from this technical difficulty. Notably, for $f(z) = \ln^k(z)$, one cannot pass a closed path around $\mathrm{Supp}(\nu) \cup \{0\}$ without crossing a branch cut for the logarithm. This problem is reminiscent of the simpler-posed, yet still open problem, consisting in evaluating $\frac{1}{p} \mathrm{tr}\, C^{-1}$ based on samples, say, $x_1, \ldots, x_p \sim \mathcal{N}(0, C)$, for $p > n$. While a consistent so-called G-estimator [13] does exist for all $p < n$, this is not the case when $p > n$.

The case $f(z) = \ln(1+sz)$ is more interesting. As the singularity for $\ln(1+sz)$ is located at $z = -1/s < 0$, one can pass a contour around $\mathrm{Supp}(\nu) \cup \{0\}$ with no branch cut issue. However, one must now ensure that there exists a contour Γ surrounding $\mathrm{Supp}(\mu)$ such that the leftmost real crossing of $(\varphi/\psi)(\Gamma)$ is located within $(-1/s, \inf\{\mathrm{Supp}(\nu)\})$. This cannot always be guaranteed. Precisely, one must show that there exists $u^- < \inf\{\mathrm{Supp}(\mu)\}$ such that $1 + s(\varphi/\psi)(u^-) > 0$. In the case $c_2^\infty < 1$ where $(\varphi/\psi)(0) = 0$, the increasing nature of φ/ψ ensures that for all $u^- \in (0, \inf\{\mathrm{Supp}(\mu)\})$, $1 + s(\varphi/\psi)(u^-) > 1$ and the condition is fulfilled; however, for $c_2^\infty > 1$, it is easily verified that $(\varphi/\psi)(0) < 0$. As a consequence, a valid u^- exists if and only if $1 + s \lim_{u \uparrow \inf\{\mathrm{Supp}(\mu)\}} (\varphi/\psi)(u) > 0$. When this condition is met, a careful calculus reveals that the estimators of Corollary 3 are still valid when $c_2^\infty > 1$, with additional absolute values in the logarithm arguments (those were discarded in the proof derivation of Corollary 3 for $c_2^\infty < 1$ as the arguments can be safely ensured to be positive). This explains the conclusion drawn in Remark 5. For generic ν, $\inf\{\mathrm{Supp}(\mu)\}$ is usually not expressible in explicit form (it can still be obtained numerically though by solving the fundamental equations (A.3) and (A.6), or estimated by $\min_i\{\lambda_i, \lambda_i > 0\}$ in practice). However, for $\nu = \delta_1$, i.e., for $C_1 = C_2$, [28, Proposition 2.1] provides the exact form of $\inf\{\mathrm{Supp}(\mu)\}$ and of $m_\mu(z)$; there, a simple yet cumbersome calculus leads to $\lim_{u \uparrow \inf\{\mathrm{Supp}(\mu)\}} (\varphi/\psi)(u) = (1 - \sqrt{c_1^\infty + c_2^\infty - c_1^\infty c_2^\infty})/(1 - c_1^\infty)$, which completes the results mentioned in Remark 5. Pursuing on the comments above on the estimation of $\mathrm{tr}\, C^{-1}$, the fact that there exists a maximal value for s allowing for a consistent estimate of $\int \ln(1 + st) d\nu(t)$, and thus not of $\int \ln(t) d\nu(t)$ which would otherwise be retrieved in a large s approximation, is again reminiscent of the fact that consistent estimators for $\frac{1}{p} \mathrm{tr}\, (C + sI_p)^{-1}$ are achievable when $p > n$, however for not-too small values of s, thereby not allowing for taking $s \to 0$ in the estimate. This problem is all the more critical that p/n is large, which we also do observe from Eq. (2) for large values of c_2.

Appendix E. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2019.06.009.

References

[1] Z.D. Bai, J.W. Silverstein, No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices, Ann. Probab. 26 (1998) 316–345.
[2] Z.D. Bai, J.W. Silverstein, CLT Of linear spectral statistics of large dimensional sample covariance matrices, Ann. Probab. 32 (2004) 553–605.
[3] Z.D. Bai, Y.Q. Yin, P.R. Krishnaiah, On the limiting empirical distribution function of the eigenvalues of a multivariate f matrix, Theory Probab. Appl. 32 (1988) 490–500.
[4] E. Basor, Y. Chen, M.R. McKay, Perturbed laguerre unitary ensembles, Painlevé V and information theory, 2013, ArXiv preprint, arXiv:1303.0773.
[5] M. Basseville, Divergence measures for statistical data processing – an annotated bibliography, Signal Process. 93 (2013) 621–633.
[6] A. Bhattacharyya, On a measure of divergence between two statistical populations defined by their probability distributions, Bull. Calcutta Math. Soc. 35 (1943) 99–109.
[7] C.I. Chang, Hyperspectral Imaging: Techniques for Spectral Detection and Classification, vol. 1, Springer Science & Business Media, 2003.
[8] S.I. Costa, S.A. Santos, J.E. Strapasson, Fisher Information distance: a geometrical reading, Discrete Appl. Math. 197 (2015) 59–69.
[9] R. Couillet, F. Benaych-Georges, Kernel spectral clustering of large dimensional data, Electron. J. Stat. 10 (2016) 1393–1454.
[10] R. Couillet, W. Hachem, Analysis of the limit spectral measure of large random matrices of the separable covariance type, Random Matrix Theory Appl. 3 (2014) 1–23.

[11] R. Couillet, J.W. Silverstein, Z.D. Bai, M. Debbah, Eigen-inference for energy estimation of multiple sources, IEEE Trans. Inform. Theory 57 (2011) 2420–2439.
[12] F.D.V. Fallani, J. Richiardi, M. Chavez, S. Achard, Graph analysis of functional brain networks: practical issues in translational neuroscience, Philos. Trans. R. Soc. B 369 (2014) 20130521.
[13] V.L. Girko, Introduction to general statistical analysis, Theory Probab. Appl. 32 (1987) 229–242.
[14] W. Hachem, P. Loubaton, J. Najim, A CLT for information theoretic statistics of gram random matrices with a given variance profile, Ann. Probab. 18 (2008) 2071–2130.
[15] A. Houdard, C. Bouveyron, J. Delon, High-Dimensional Mixture Models for Unsupervised Image Denoising (hdmi), 2017.
[16] A. Kammoun, R. Couillet, J. Najim, M. Debbah, Performance of mutual information inference methods under unknown interference, IEEE Trans. Inform. Theory 59 (2012).
[17] V.A. Marčenko, L.A. Pastur, Distribution of eigenvalues for some sets of random matrices, Math USSR Sb. 1 (1967) 457–483.
[18] X. Mestre, Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates, IEEE Trans. Inform. Theory 54 (2008) 5113–5129.
[19] X. Mestre, On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices, IEEE Trans. Signal Process. 56 (2008) 5353–5368.
[20] J. Najim, J.F. Yao, GaussIan fluctuations for linear spectral statistics of large random covariance matrices, 2013, ArXiv preprint, arXiv:1309.3728.
[21] P.L.C. Rodrigues, M. Congedo, C. Jutten, Multivariate time-series analysis via manifold learning, in: 2018 IEEE Statistical Signal Processing Workshop, SSP, IEEE, 2018, pp. 573–577.
[22] J.W. Silverstein, The limiting eigenvalue distribution of a multivariate f matrix, SIAM J. Math. Anal. 16 (1985) 641–646.
[23] J.W. Silverstein, Z.D. Bai, On the empirical distribution of eigenvalues of a class of large dimensional random matrices, J. Multivariate Anal. 54 (1995) 175–192.
[24] J.W. Silverstein, S. Choi, Analysis of the limiting spectral distribution of large dimensional random matrices, J. Multivariate Anal. 54 (1995) 295–309.
[25] G. Vasile, J.P. Ovarlez, F. Pascal, C. Tison, Coherency matrix estimation of heterogeneous clutter in high-resolution polarimetric sar images, IEEE Trans. Geosci. Remote Sens. 48 (2010) 1809–1826.
[26] U. Von Luxburg, A tutorial on spectral clustering, Stat. Comput. 17 (2007) 395–416.
[27] U. Von Luxburg, M. Belkin, O. Bousquet, Consistency of spectral clustering, Ann. Statist. (2008) 555–586.
[28] Q. Wang, J. Yao, et al., Extreme eigenvalues of large-dimensional spiked fisher matrices with application, Ann. Statist. 45 (2017) 415–460.
[29] J. Yao, R. Couillet, J. Najim, M. Debbah, Fluctuations of an improved population eigenvalue estimator in sample covariance matrix models, IEEE Trans. Inform. Theory 59 (2012) 1149–1163.
[30] D. Zagier, The dilogarithm function, in: Frontiers in Number Theory, Physics, and Geometry II, Springer, 2007, pp. 3–65.
[31] S. Zheng, Z. Bai, J. Yao, et al., Clt for eigenvalue statistics of large-dimensional general fisher matrices with applications, Bernoulli 23 (2017) 1130–1178.
[32] S. Zheng, et al., Central limit theorems for linear spectral statistics of large dimensional f-matrices, in: Annales de L'Institut Henri PoincarÉ, ProbabilitÉS Et Statistiques, Institut Henri Poincaré, 2012, pp. 444–476.