Random Matrices: Theory and Applications
Vol. 4, No. 4 (2015) 1550019 (36 pages)

DOI: 10.1142/S2010326315500197

Efficient computation of limit spectra of sample covariance matrices

Edgar Dobriban

Department of Statistics, Stanford University
Stanford, CA 94305, USA
dobriban@stanford.edu

Received 6 July 2015
Accepted 1 September 2015
Published 20 October 2015

Models from random matrix theory (RMT) are increasingly used to gain insights into the behavior of statistical methods under high-dimensional asymptotics. However, the applicability of the framework is limited by numerical problems. Consider the usual model of multivariate statistics where the data is a sample from a multivariate distribution with a given covariance matrix. Under high-dimensional asymptotics, there is a deterministic map from the distribution of eigenvalues of the population covariance matrix (the population spectral distribution or PSD), to the of empirical spectral distribution (ESD). The current methods for computing this map are inefficient, and this limits the applicability of the theory.

We propose a new method to compute numerically the ESD from an arbitrary input PSD. Our method, called Spectrode, finds the support and the density of the ESD to high precision; we prove this for finite discrete distributions. In computational experiments Spectrode outperforms existing methods by orders of magnitude in speed and accuracy. We apply it to compute expectations and contour integrals of the ESD, which are often central in applications.

We also illustrate that Spectrode is directly useful in statistical problems, such as estimation and hypothesis testing for covariance matrices. Our proposal, implemented in open source software, may broaden the use of RMT in high-dimensional data analysis.

Keywords: Limiting spectral distribution; sample covariance matrix; Stieltjes transform; numerical computation; high-dimensional statistics.

Mathematics Subject Classification 2010: 65C50, 15B52

1. Introduction

Large data matrices are now commonly analyzed in science and engineering. Models from random matrix theory (RMT) are becoming increasingly used to understand the behavior of popular statistical methods on such matrices. RMT is particularly applicable to analyze statistical methods which depend on the sample covariance matrix of the data: for instance principal component analysis (PCA), classification, hypothesis testing of high-dimensional means, and independence tests, see e.g. the monographs [35, 30, 9, 37].

Concretely, consider an $n \times p$ matrix $\mathbf{X}$, whose rows $\mathbf{x}_i$ are independent and identically distributed random vectors. Suppose that $\mathbf{x}_i$ are mean zero, and their covariance matrix is the $p \times p$ matrix $\mathbf{\Sigma} = \mathbb{E}\mathbf{x}_i\mathbf{x}_i^\top$. To estimate $\mathbf{\Sigma}$, we form the sample covariance matrix $\widehat{\mathbf{\Sigma}} = n^{-1}\mathbf{X}^\top\mathbf{X}$. In the asymptotic model classically used in statistics, when p is fixed and $n \to \infty$, the sample covariance matrix is a good estimator of the population covariance [1].

However, if n and p are of comparable size, then $\widehat{\mathbf{\Sigma}}$ deviates substantially from the true covariance. The asymptotic theory of random matrices describes the behavior of the eigenvalues of $\widehat{\mathbf{\Sigma}}$ as n, p grow large proportionally, see [4]. If the distribution of the eigenvalues of $\mathbf{\Sigma}$ tends to a limit population spectral distribution (PSD) as $n, p \to \infty$ and the aspect ratio $p/n \to \gamma$, then under mild conditions the random eigenvalue distribution of $\widehat{\mathbf{\Sigma}}$ also tends to a deterministic limit empirical spectral distribution (ESD) [23, 32].

The "fundamental theorem of applied statistics" is the statement that often the limit of the empirical distribution function is the population distribution. This theorem applies in numerous settings, see e.g. [31], but not here. When $n \to \infty$ but $p/n \to \gamma > 0$, the limit empirical spectrum differs from the true spectrum, because the number of samples is only a constant multiple of the dimension. This is very different from the case where p is fixed and $n \to \infty$, in which case the sample spectrum converges to the true spectrum. The difference between the population and empirical eigenvalues has fundamental implications for high-dimensional statistical inference, see e.g. [20, 37]. It becomes essential to understand the relationship between the population and sample eigenvalues. This understanding should help adjust classical statistical methods to the high-dimensional setting.

However, the relationship between the population and sample spectrum is complex, implicit and nonlinear; it is described by a fixed-point equation — often called the Marchenko–Pastur equation or Silverstein equation — for the Stieltjes transform of the limit ESD. As a consequence, the ESD is not available in closed form, except for very special cases. The implicit description of the sample spectrum can be somewhat hard to understand, as well as hard to use in any practical setting, including data analysis.

Reliable, precise, and efficient computational methods are needed to understand the relationship between the population and sample eigenvalues. Perhaps surprisingly, research focused on delivering robust software tools for numerically computing large classes of limit ESDs has received relatively little attention. While there are important contributions to related problems (see Sec. 6), none of them are fully suitable for our problem.

The main method for computing the limit ESD is a fixed point algorithm (FPA) which directly iterates the Silverstein equation. Since the algorithm is immediately suggested by the fixed-point characterization of the ESD, the history of this algorithmic approach is apparently lost in the prehistory of the subject. FPA has appeared recently in various forms, e.g. [5, 16, 10, 36, 18]. Further, FPA is recommended as the default method for computing the ESD in the monograph of Yao, Zheng

and Bai [37]. FPA is a good method for computing the density of the ESD at a single point. However, usually the density of the ESD must be computed on a dense grid on the real line. When this is the case, we show that FPA is inefficient for high-precision computations.

We propose the new method SPECTRODE to compute the limit empirical spectrum of covariance matrices from the limit population spectrum. SPECTRODE improves on FPA by exploiting the smoothness of the ESD, via an ordinary differential equation (ODE). We show in computational experiments on dense grids that our new method is dramatically faster and more accurate than FPA and other methods. For instance, on natural test problems in Sec. 3.3, SPECTRODE is up to 1000 times faster than FPA while achieving the same accuracy! Finally, for atomic PSDs, i.e. weighted mixtures of point masses, we prove its convergence to the correct answer.

SPECTRODE is publicly available in an open source MATLAB implementation at `https://github.com/dobriban/eigenedge`, or from the author. This software package also has the code to reproduce all computational results of our paper (see Sec. 7).

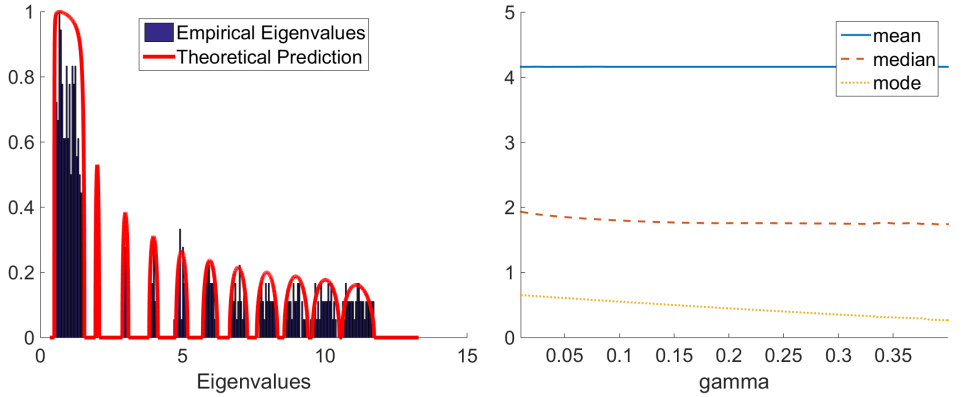
In the remainder of this section, we showcase example computations with SPECTRODE, and highlight the key aspects of the method. Then we state its properties more precisely.

1.1. *Two examples*

We illustrate SPECTRODE by computing two quantities of interest. In this example the PSD is an equal mixture of two components: (1) a mixture of 10 point masses at $2, 3, \ldots, 11$, with weights forming an arithmetic progression with step $r = 0.005$ as follows: $0.0275, 0.0325, \ldots, 0.0725$; and (2) a uniform distribution — or a "boxcar" — on $[0.5, 1.5]$, with mixture weight $1/2$. The weights sum to one. We use the aspect ratio $\gamma = 0.01$.

Figure 1 shows the example computations. In subplot (a), we show the density of the limit ESD, the key output of SPECTRODE. The computation takes 0.5 s on a desktop computer. *A priori* it is not obvious how many disjoint clusters there are in the ESD, or what their shape is. Several insights can be derived from the computation: there are 11 clusters in total, so all population clusters separate. Each population cluster in the PSD, in this instance, creates a distinct component of the ESD. Further, the two rightmost clusters nearly touch; and the height of the clusters decreases while the width of the point-mass clusters increases.

As a second example, in subplot (b) we compute three functionals of the ESD as a function of γ: the mean, median, and the mode. Such functionals are important in statistical applications: for instance, the median is used for optimal singular value shrinkage in [13]; see also Sec. 5.1. As expected, the mean does not depend on γ. However, the behavior of the median and the mode is not obvious. Using SPECTRODE, one can get insight into their behavior: the mode decreases as a function of γ, and the median is greater than the mode.



(a) (b)

Fig. 1. Boxcar + point mass mixture example: SPECTRODE computes (a) the density of the limit ESD, normalized to have maximum equal to one for display purposes, and (b) its mean, median, and mode as a function of γ.

1.2. *Highlights*

To summarize and expand on the above argument, we highlight the following aspects of our method.

(1) It provides ready access to a wide variety of new examples of limit spectra of covariance matrices. There has been, until now, no convenient tool for precisely calculating the ESD for such large collections of examples.
(2) SPECTRODE computes several important functionals of the limit empirical spectrum, namely:
   (a) **The edges of the support:** The ESD is typically supported on a union of compact intervals. The edges of the support are of interest in the study of phase transitions in spiked covariance models [7], and in designing optimal singular value shrinkers for matrix denoising [26].
   (b) **Moments of the ESD:** Numerous applications depend on moments of the ESD [35, 9, 37]. We show that general moments $\mathbb{E}_F[h(\lambda)]$ can be computed with SPECTRODE. The polynomial moments $\mathbb{E}_F[\lambda^k]$ can be computed alternatively via challenging free probability calculations [28]. However, this does not hold for more general moments $\mathbb{E}_F[h(\lambda)]$ for arbitrary h. Therefore, our method could extend the applicability of existing techniques by providing a unified way to compute nearly all moments of interest (Sec. 5.1).
   (c) **Contour integrals of the ESD's Stieltjes transform:** Contour integrals of the Stieltjes transform appear crucially in the central limit theorem for linear spectral statistics of covariance matrices due to Bai and Silverstein [3]. In applications of this powerful result to multivariate statistics, calculating the contour integral formulas for the mean and the variance is

a key step, see e.g. [37]. These moments are known in closed form only in a few cases. The current approach is to calculate them using residue theory from complex analysis. This analytic approach can require substantial effort, and is limited to the cases where the ESD is known in closed form. SPECTRODE enables us to compute such contour integrals numerically instead (Sec. 5.2). High precision numerical results may suffice in many applications.

(3) SPECTRODE is directly useful in statistical applications. We give two key examples where SPECTRODE could improve significantly on current statistical methodology:

(a) **Estimating the covariance matrix $\mathbf{\Sigma}$:** A problem of considerable interest in statistics is to estimate the unobserved covariance matrix $\mathbf{\Sigma}$ based on the observed data. When the number of samples n is comparable to the dimension p, covariance estimation is a challenging problem.

A recent series of methods due to Ledoit and Wolf [21, 22] assumes that one can accurately compute the ESD for any proposal PSD. Their estimation method consists of solving a nonlinear least squares problem, matching the quantiles of the output ESD — their so-called QuEST function — against the observed eigenvalues. This is solved by an iterative optimization method, and the ESD computation is repeatedly invoked at each step of the iteration. Unfortunately, the whole framework is limited by the accuracy of the ESD computation. We suggest that SPECTRODE may be used to upgrade this procedure, replacing the existing implementation of QuEST.

(b) **Hypothesis tests on the covariance matrix:** Testing statistical hypotheses on the covariance matrix can be approached by using the CLT for the linear spectral statistics, see e.g. [2, 37]. We suggest that the mean and variance in the CLT could be computed numerically (Sec. 5.2). Our approach, implemented with open source software, might be significantly more convenient than traditional analytic calculations. In addition, it may lead to entirely new test statistics, whose analysis was not possible via pre-existing methodology.

There may be of course many other ways that our efficient computational framework will be useful to the statistics and engineering communities.

1.3. *Properties of* SPECTRODE

1.3.1. *Background*

To state precisely the properties enjoyed by SPECTRODE, we first set up the formal background. A more thorough presentation will be given in Sec. 2. Consider a sequence of problems indexed by p, with deterministic $p \times p$ covariance matrices $\mathbf{\Sigma}_p$. Let $\tau_1, \ldots, \tau_p$ be the eigenvalues of $\mathbf{\Sigma}_p$ and H_p be their cumulative distribution

function $H_p(x) = p^{-1}\sum_i I(\tau_i \le x)$. For each p, draw $n := n_p$ independent samples $\mathbf{x}_{ip}$ of the form $\mathbf{x}_{ip} = \mathbf{\Sigma}_p^{1/2}\mathbf{y}_{ip}$, where $\mathbf{y}_{ip}$ is a p-dimensional random vector with independent and identically distributed, mean zero, variance one entries.

Arrange the vectors $\mathbf{x}_{ip}$ into the rows of the $n \times p$ data matrix $\mathbf{X}_p$, and form the sample covariance matrix $\widehat{\mathbf{\Sigma}}_p = {n_p}^{-1}\mathbf{X}_p^\top\mathbf{X}_p$. Let $\lambda_1, \ldots, \lambda_p$ be the eigenvalues of $\widehat{\mathbf{\Sigma}}_p$, and F_p their cumulative distribution function $F_p(x) = p^{-1}\sum_i I(\lambda_i \le x)$.

Consider the high-dimensional limit where $p, n_p \to \infty$ such that $p/n_p \to \gamma > 0$. Suppose the eigenvalue distributions H_p converge to a limit PSD H, i.e. $H_p \Rightarrow H$ in distribution. Then a cornerstone result in RMT, the Marchenko–Pastur theorem, states that the empirical eigenvalue distributions F_p also converge, almost surely, to a limit ESD F [23, 32].

We consider the computation of F from H. The method we propose is general and well-defined for all PSDs H. Our analysis considers *atomic* PSDs H, which are finite mixtures of point masses, but see Sec. 4.2 for the extension to general distributions. Let $H = \sum_{i=1}^J w_i\delta_{t_i}$, where δ_t is the point mass at t, $w_i > 0$ are the component masses with $\sum_i w_i = 1$, and $t_i > 0$ are the population eigenvalues. We exclude the case $\gamma = 1$ for technical reasons, such as the potentially unbounded density of the ESD at $x = 0$.

In pioneering work, Silverstein and Choi [33] study the limit ESD corresponding to general H. They show that the limit ESD F has a continuous density $f(x)$ for $x \neq 0$. The density $f(x)$ exists at $x = 0$ if $\gamma < 1$, but not if $\gamma > 1$. Instead F has a point mass of weight $1 - \gamma^{-1}$ at $x = 0$. For atomic distributions, it follows from the results in [33] that the distribution is supported on a union of K disjoint compact intervals $[l_k, u_k]$, where l_k is the lower endpoint and u_k is the upper endpoint of the kth interval for $1 \le k \le K$. The endpoints are such that $0 \le l_1 < u_1 < \cdots < l_K < u_K$. The number of sample intervals K is at most the number of population components J. If the aspect ratio γ is sufficiently close to 1, then some population components can "merge" in the sample spectrum, and $K < J$ will occur. Finally, it is shown in [33] that f is analytic in the neighborhood of all points where the density is positive.

1.3.2. *Input and output of* Spectrode

Given the aspect ratio γ, a population spectrum H (for instance an atomic distribution) and a user-specified precision control parameter $\varepsilon > 0$, Spectrode produces an approximation of F consisting of the following numerical estimates:

(1) The number of intervals in the support of F: $\hat{K} = \hat{K}(\varepsilon)$.
(2) The endpoints of the support intervals $[\hat{l}_k(\varepsilon), \hat{u}_k(\varepsilon)]$, for $k = 1, \ldots, \hat{K}$.
(3) The density $\hat{f}(x, \varepsilon)$ for all real x.

For the reader's convenience, the input and output of Spectrode are summarized in Table 1.

Table 1. Input and output of SPECTRODE.

SPECTRODE: Input and Output
Input:
$H \leftarrow$ population spectrum (e.g. atomic measure: eigenvalues $t_1, \ldots, t_J$ and masses $w_1, \ldots, w_J$) $\gamma \leftarrow$ aspect ratio $\varepsilon \leftarrow$ precision control parameter
Output:
$\hat{K}(\varepsilon) \leftarrow$ number of intervals in the support $[\hat{l}_k(\varepsilon), \hat{u}_k(\varepsilon)] \leftarrow$ endpoints of intervals in the support $\hat{f}(x, \varepsilon) \leftarrow$ density of the spectrum, for any x

1.3.3. *Correctness of* SPECTRODE

Our main theoretical results, given in Sec. 2, demonstrate the correctness of our proposed method. As the user-specified precision control parameter $\varepsilon \to 0$, SPECTRODE has the following performance characteristics:

(1) Correctness of the number of disjoint intervals of the support:

$$\lim_{\varepsilon \to 0} \hat{K}(\varepsilon) = K. \tag{1.1}$$

(2) Accuracy of the endpoints of the support:

$$\lim_{\varepsilon \to 0} \hat{l}_k(\varepsilon) = l_k \quad \text{and} \quad \lim_{\varepsilon \to 0} \hat{u}_k(\varepsilon) = u_k. \tag{1.2}$$

(3) Accuracy of the density[a]:

$$\lim_{\varepsilon \to 0} \sup_{x \in \mathbb{R} \setminus \{0\}} |\hat{f}(x, \varepsilon) - f(x)| = 0. \tag{1.3}$$

Claims (1.1)–(1.2) are proved in Theorem 4.1, while claim (1.3) is proved in Theorem 4.2. The claims are verified in reproducible computational experiments in Sec. 3 (see also Sec. 7).

As a consequence of these results, we show in Corollary 5.1 that the moments of the limit ESD can be approximated by integrals against the numerical estimate of the density. Finally, we adapt SPECTRODE to compute contour integrals involving the Stieltjes transform of the limit ESD in Sec. 5.2.

The computational framework used by SPECTRODE is applicable to general population distributions H, not just atomic distributions. Indeed, we already showed an example involving a uniform distribution on an interval in Fig. 1. However, our current software implementation of SPECTRODE assumes that H is a finite mixture of uniform distributions and point masses. Moreover, the proof of convergence that we supply in this paper only holds for atomic distributions. Therefore, we will work

[a]A more precise statement is: For $\gamma < 1$, the convergence is uniform over all $x \in \mathbb{R}$. For $\gamma > 1$, the density does not exist at $x = 0$, but is equal to $f(x) = 0$ on some intervals $\mathcal{I} = (-\delta, 0) \cup (0, \delta)$, with $\delta > 0$. Then, the convergence is uniform over the closed set $\mathbb{R} \setminus (\mathcal{I} \cup 0)$.

with atomic distributions through most of this paper. This issue is further discussed in Sec. 4.2.

In the rest of the paper, we present SPECTRODE in Sec. 2. We validate our claims with computational experiments in Sec. 3, and with a convergence proof in Sec. 4. After giving some applications and extensions in Sec. 5, we describe related literature in Sec. 6. The available software and the tools to reproduce our computational results are described in Sec. 7.

2. The Method

2.1. *Background*

In this section we explain our method. We start with some background about limiting spectral distributions of large covariance matrices. Chapter 7 of Couillet and Debbah's monograph [9] provides a good summary of the material presented here. Recall the model presented in Sec. 1: $\mathbf{X}$ is $n \times p$, of the form $\mathbf{X} = \mathbf{Y}\mathbf{\Sigma}_p^{1/2}$, where the entries of $\mathbf{Y}$ are iid with mean zero and variance one. We take a sequence of such problems with p, n growing to infinity such that $p/n \to \gamma > 0$. The PSD of the deterministic $\mathbf{\Sigma}_p$ converges to the limit PSD H.

The Marchenko–Pastur theorem (see [23, 32]) states that the ESD of the sample covariance matrix $\widehat{\mathbf{\Sigma}} = n^{-1}\mathbf{X}^\top\mathbf{X}$ converges almost surely to a distribution F. Denote the imaginary part of $z \in \mathbb{C}$ by $\mathrm{Imag}(z)$ and the upper half of the complex plane by $\mathbb{C}^+ = \{z \in \mathbb{C} : \mathrm{Imag}(z) > 0\}$. If $m(z)$ denotes the Stieltjes transform of F, defined for $z \in \mathbb{C}^+$ as:

$$m(z) = \int \frac{dF(x)}{x - z}, \tag{2.1}$$

and $v(z)$ is the companion Stieltjes transform defined on $\mathbb{C}^+$ by the equation

$$\gamma(m(z) + 1/z) = v(z) + 1/z, \tag{2.2}$$

then it is shown in [33] that $v(z)$ is the unique solution with positive imaginary part of the Silverstein equation:

$$-\frac{1}{v(z)} = z - \gamma \int \frac{tdH(t)}{1 + tv(z)}, \quad z \in \mathbb{C}^+. \tag{2.3}$$

This equation links the limit PSD H to the limit ESD F. The function v is analytic in the upper half $\mathbb{C}^+$ of the complex plane. Marchenko and Pastur [23] obtained a more general, but more complicated, form of this equation. The present form is due to Silverstein and appeared in [34].

Our problem is to compute F from H. For this it is enough to find $v(z)$, and thus $m(z)$, for z on a grid close to the real axis. Since F has a density f [33], by the inversion formula for Stieltjes transforms

$$f(x) = \frac{1}{\pi} \lim_{\varepsilon \to 0^+} \mathrm{Imag}\{m(x + i\varepsilon)\}. \tag{2.4}$$

This limit is valid at all points x where the density $f(x)$ exists.[b] For $\gamma < 1$, the density exists for all x, while for $\gamma > 1$ it exists for all x except for $x = 0$, where F has a point mass. Numerically it is natural to use the approximation $\hat{f} = \text{Imag}\{m(x + i\varepsilon)\}/\pi$ for some small $\varepsilon > 0$. There may be more accurate methods for interpolating $m(x + i\varepsilon)$, but those are beyond our scope.

In general there are many solutions to (2.3) with non-positive imaginary part. For a finite mixture of point masses, $H = \sum_{i=1}^{J} w_i \delta_{t_i}$, the Silverstein equation becomes

$$-\frac{1}{v(z)} = z - \gamma \sum_{i=1}^{p} \frac{w_i t_i}{1 + t_i v(z)}, \quad z \in \mathbb{C}^+. \tag{2.5}$$

This is generally equivalent to a polynomial equation of degree $p + 1$, and hence it has $p + 1$ complex roots, compare [27]. The desired solution will "track" one of the roots as a function of z. However, finding the right solution by root tracking is not feasible in general for large p. There does not appear to be a way to efficiently compute the coefficients of the polynomial. We will take a different approach.

2.2. *Our approach*

We differentiate the fixed-point equation (2.5) in z, and solve for v'. These steps yield the following ODE for v:

$$\frac{dv}{dz} = \mathcal{F}(v) := \frac{1}{\frac{1}{v^2} - \gamma \sum_{i=1}^{J} \frac{w_i t_i^2}{(1+t_i v)^2}}, \quad \hat{v}(z_0) = \hat{v}_0. \tag{2.6}$$

A high-accuracy starting point for the ODE can be found by running the fixed-point algorithm once, at a point $z_0 = x_0 + i\varepsilon$ near the real axis. The ESD can be computed at other real values x by solving the ODE on the line $x + i\varepsilon, x \in \mathbb{R}$. Solving the ODE is much more convenient than solving the original equation repeatedly for each new point $x + i\varepsilon$. The reason is that the limit spectral density is smooth, and the Stieltjes transform provides further smoothing. Our ODE uses this smoothness for efficient computation. This is in contrast to FPA, which re-runs the entire fixed-point iteration at each nearby point and does not exploit the smoothness. Using the smoothness via the ODE is the key inspiration behind our approach.[c]

Since the ODE was obtained by differentiating (2.5), it has at least one solution. We will show that there is only one solution in the range of interest. Once we obtain a numerical solution $\hat{v}(x)$ to the ODE, we could define $\hat{f}$ directly based on the

[b]The result of Silverstein and Choi [33] is stronger. It also states that the density $f(x)$ is the imaginary part of $m(x)$, defined as the limit of the Stieltjes transform as $z \to x$; and the associated $v(x)$ is in fact the solution of the Silverstein equation (2.3) with $z = x$. While this is in fact an exact equation for the density of the ESD, we will not use it in this paper. We will instead rely on the Silverstein equation on a grid close to the real axis. The reason is that we use FPA as a starting point of our method, and FPA is only known to converge for $z \in \mathbb{C}^+$, in the interior of the upper half-plane.

[c]This "spectral ODE" was also the source of the name SPECTRODE.

explicit formulas (2.2) and (2.4). This direct method already leads to a relatively good solution which provably converges to the right answer as $\varepsilon \to 0$.

However, for small ε this direct method has numerical problems caused by irregularities in the density at the edges of the support. It is well known, and also shown in Fig. 1, that the density exhibits a square root behavior at the edges of the support. If implemented naively, these square root irregularities can cause difficulties to the ODE solver. We will avoid them by finding the edges of the spectrum via the theory of Silverstein and Choi [33], and solving the ODE only within the support of F.

In brief, Silverstein and Choi find the support of the spectrum in the following way. They consider the Silverstein equation (2.3), which defines the companion Stieltjes transform v implicitly as a function of z. They observe that the same equation defines a function $z(v)$:

$$z(v) = -\frac{1}{v} + \gamma \sum_{i=1}^{J} \frac{w_i t_i}{1 + t_i v}. \tag{2.7}$$

They prove that the support can be obtained by analyzing the monotonicity of z. The real intervals $v \in (a, b)$ where $z(v)$ is increasing, i.e. $z'(v) > 0$, are precisely those whose image under $z(v)$ (i.e. $(z(a), z(b))$) is the complement of the support of the distribution $\underline{F}$. Here $\underline{F}$ is the limit ESD of $n^{-1}\mathbf{X}\mathbf{X}^\top$. Since $\underline{F}$ is directly related to F, see (4.1), in theory this is enough to find the support. We will give an accurate method to do so.

3. Computational Results

In addition to theoretical results, we validate our performance claims by computational experiments. We show the correctness of the support in Sec. 3.1; the correctness of the density in Sec. 3.2; and the computational efficiency of SPECTRODE in Sec. 3.3. The experiments are reproducible (Sec. 7).

3.1. *Correctness of the support*

3.1.1. *The comb model*

To show that our algorithm identifies the support (claims (1.1)–(1.2)), we consider the following *comb model* for eigenvalues. Here the eigenvalues and the weights are each defined in terms of arithmetic progressions $H = \sum_{j=0}^{J-1}(a + jb)\delta_{c+jd}$.

The eigenvalues are placed at $c + jd$, for some $c > 0$ and $d \in \mathbb{R}$ such that $c + jd > 0$ for all j. They have weights $a + jb$ for some $a, b > 0$. The constants a, b are constrained so that the sum of the weights is one, leaving, say, b a free parameter. This is a flexible model governed by only three parameters.

The comb model is a useful test for the support identification problem. If $\gamma \to 0$ while other parameters are fixed, then $F_\gamma \to H$ [33]. Intuitively, the number of samples is much larger than the dimension, $n \gg p$, so the ESD converges to the atomic PSD. Now as γ increases, the sharp atoms spread out into density bumps. If

the original atoms are sufficiently close to each other, then at some point bumps will start merging. The precise moment when this happens is a complicated function of J, b, c, d and γ, but can be determined precisely with SPECTRODE.

3.1.2. *Testing our method*

Since in most cases there is no closed form for the density, we compare our algorithm against FPA; see its description preceding Lemma 4.6. FPA is empirically slow for dense grid evaluation (Sec. 3.3), but converges, as shown in a more general setting in [10]. The convergence rate of FPA is not known, hence one cannot guarantee its exact accuracy. We have validated FPA separately on simpler test cases where the closed form expression was known (data not shown).

Our numerical test has the following framework: For given problem parameters, and an accuracy control parameter ε, we run SPECTRODE to produce numerical approximations $\hat{K}(\varepsilon)$, $\hat{l}_k(\varepsilon)$, $\hat{r}_k(\varepsilon)$. The method also returns a dense grid of x_i. On this grid we compute the density approximations $\hat{f}_{\text{fp}}(x_i, \varepsilon_0)$ of FPA, with an accuracy control parameter ε_0. Here the parameter ε_0 is smaller than ε, so that FPA's solution can be reliably used as a basis of comparison for ε-accurate computations.

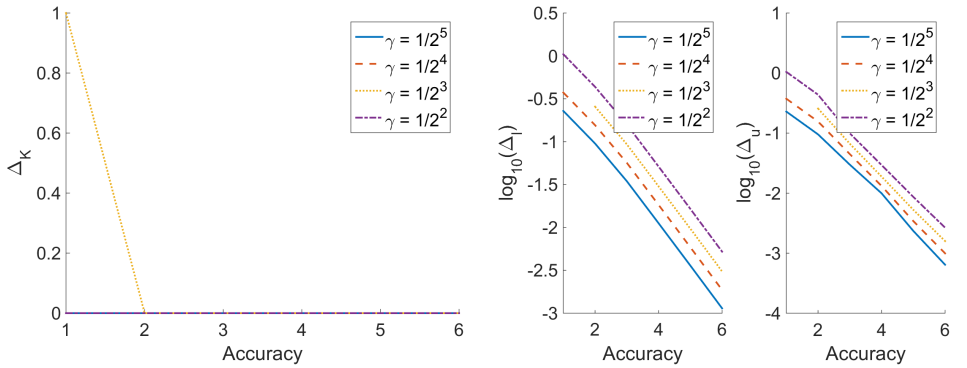
We then define the gold standard approximation to the support as the connected components of the grid x_i where the density $\hat{f}_{\text{fp}}(x_i, \varepsilon_0) > \varepsilon_0$. This step thresholds the density at level ε_0, because FPA was tuned to have accuracy of the order ε_0. This prescription produces approximations $\hat{K}_{\text{fp}}(\varepsilon_0)$, $\hat{l}_{\text{fp},k}(\varepsilon_0)$, $\hat{r}_{\text{fp},k}(\varepsilon_0)$.

We evaluate SPECTRODE by calculating the error in the number of clusters: $\Delta_K(\varepsilon) = |\hat{K}(\varepsilon) - \hat{K}_{\text{fp}}(\varepsilon_0)|$, where ε_0 is suppressed for brevity. For the support endpoints we proceed similarly. If the number of intervals is not computed correctly, then we set this error to $+\infty$. Even if the number of intervals is correct, so that $\hat{K} = K$, we take into account the finite precision of the grid x_i — as explained below.

Consider a lower endpoint for one of the clusters. Suppose that among the grid points $x_1, \ldots, x_M$, the two methods output the grid elements x_i and x_j, with $i \le j$, as numerical approximations for the lower endpoint. Due to finite grid precision, $|x_i - x_j|$ can be an underestimate of the actual error. For instance if $x_i = x_j$, it is clear that our accuracy bound cannot, in general, be better than the size of grid spacings $|x_i - x_{i-1}|$, $|x_j - x_{j+1}|$. Generally, a conservative estimate of the accuracy can be obtained by adding these neighboring grid spacings to $|x_i - x_j|$ to get $\Delta_k^l(\varepsilon) = |x_{i-1} - x_{j+1}|$, where we recall that $x_{i-1} < x_i \le x_j < x_{j+1}$.

Finally, the approximation error for lower endpoints is defined as the average of all errors for lower endpoints: $\Delta_l(\varepsilon) = \sum_{k=1}^{\hat{K}} \Delta_k^l(\varepsilon)/\hat{K}$. The approximation error for upper endpoints $\Delta_u(\varepsilon)$ is defined analogously. Note again: if $\hat{K} \neq K$, we set the error to be ∞.

The comb model in this test has $J = 6$ clusters spaced evenly between $1/2$ and 10, and a gap in the sequence of weights $b = 0.01$, leading to nearly equal weights.



(a) $\Delta_K(\varepsilon)$: error in the number of clusters (b) $\Delta_l(\varepsilon), \Delta_u(\varepsilon)$: error in the cluster endpoints

Fig. 2. SPECTRODE (a) correctly identifies the number of disjoint intervals in a comb model; and (b) accurately computes the lower and upper endpoints.

The aspect ratio γ takes four values between $1/2^5$ and $1/2^2$. We fix $\varepsilon_0 = 10^{-7}$ and vary the accuracy $\varepsilon = 10^{-m}$, $m = 1, \ldots, 6$.

3.1.3. *Results*

We show the results of the experiment in Fig. 2. In panel (a), we show the error in the number of clusters $\Delta_K(\varepsilon)$ for the four different aspect ratios γ, as a function of the accuracy. SPECTRODE makes at most one error in the number of clusters. For sufficiently high accuracy the number of clusters is correct.

In panel (b), we show the approximation error for the endpoints: $\Delta_l(\varepsilon)$ (left), and $\Delta_u(\varepsilon)$ (right), on a logarithmic scale. For the experiments where $\Delta_K(\varepsilon) > 0$, we leave blanks. The approximation improves with higher accuracy. This convergence appears nearly linear in ε, i.e. the number of correct digits is approximately linearly related to $-\log_{10}(\varepsilon)$, with a slope of approximately 1/2. These experiments provide evidence that SPECTRODE correctly identifies the support (claims (1.1)–(1.2)).

3.2. *Accurate computation of the density*

We validate our claim (1.3) that SPECTRODE accurately computes the limit density. To test the accuracy up to several digits, we rely on examples where the density f can be found exactly in an alternative way.

3.2.1. *Test problems*

For our first test, called `MP`, the population spectrum H is a point mass at 1. The ESD has the density:

$$f(x;\gamma) = \frac{\sqrt{(\gamma_+ - x)(x - \gamma_-)}}{2\pi\gamma x} I(x \in [\gamma_-, \gamma_+]), \tag{3.1}$$

where $\gamma_\pm = (1 \pm \sqrt{\gamma})^2$. The distribution of eigenvalues has a point mass at $x = 0$ if $\gamma > 1$.

For the second test, called `TwoPoint`, H is a mixture of two point masses at $x = 1$ and t, with weights q and $1 - q$. The Silverstein equation (2.3) reads

$$-\frac{1}{v(z)} = z - \gamma\left(\frac{q}{1+v(z)} + \frac{(1-q)t}{1+tv(z)}\right).$$

This is equivalent to a polynomial equation in v of degree at most three:

$$ztv^3 + (zt + z + t - t\gamma)v^2 + [z + t + 1 - \gamma(q + (1-q)t)]v + 1 = 0. \tag{3.2}$$

When $z, t \neq 0$, as is always the case for us, this is a cubic equation that can be solved exactly. The theory of Silverstein and Choi [33] guarantees that for real x within the support of the spectrum, Eq. (3.2) has exactly one root with positive imaginary part, which leads to the correct density. For x outside the spectrum, Eq. (3.2) has three real roots. This distinguishes the inside from the outside.

For each grid point x_i and accuracy ε, SPECTRODE produces a numerical approximation $\hat{f}(x_i, \varepsilon)$ to the true density $f(x_i)$. To test SPECTRODE, we compute the error in the density:

$$\Delta(x_i, \varepsilon) = \log_{10} |\hat{f}(x_i, \varepsilon) - f(x_i)|. \tag{3.3}$$

We set $\gamma = 1/2$, and vary the global accuracy parameter ε in powers of ten as $10^{-4}, 10^{-6}, 10^{-8}$. In addition, for the two-point mixture model we set a fraction $q = 1/2$ of the eigenvalues to $t = 8$.

3.2.2. *Results*

The results are shown in Fig. 3. For both test problems, the error in the density decreases uniformly as the tuning parameter $\varepsilon \to 0$. Furthermore, SPECTRODE

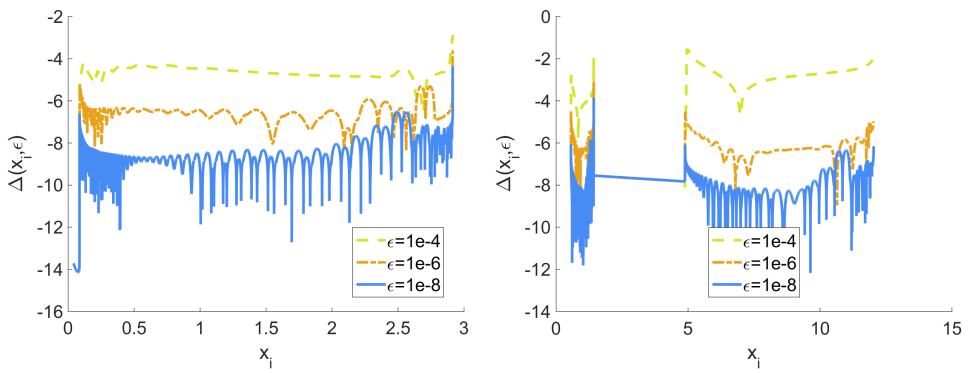


Fig. 3. Accurate computation of the density (Sec. 3.2) in two test problems. (a) `MP`. (b) `TwoPoint`. We display the error in the density $\Delta(x_i, \varepsilon)$ for three different values of ε, $10^{-4}, 10^{-6}, 10^{-8}$.

produces approximately the required accuracy: for instance the average precision for $\varepsilon = 10^{-8}$ is approximately eight digits. These experiments provide empirical evidence for claim (1.3): SPECTRODE computes the density of the limit ESD with uniform accuracy over all x.

3.3. *Computational efficiency*

We now establish that SPECTRODE is computationally efficient. We compare running times with FPA and find that for high precision problems on dense grids, SPECTRODE significantly outperforms FPA.

3.3.1. *Test problems and parameters*

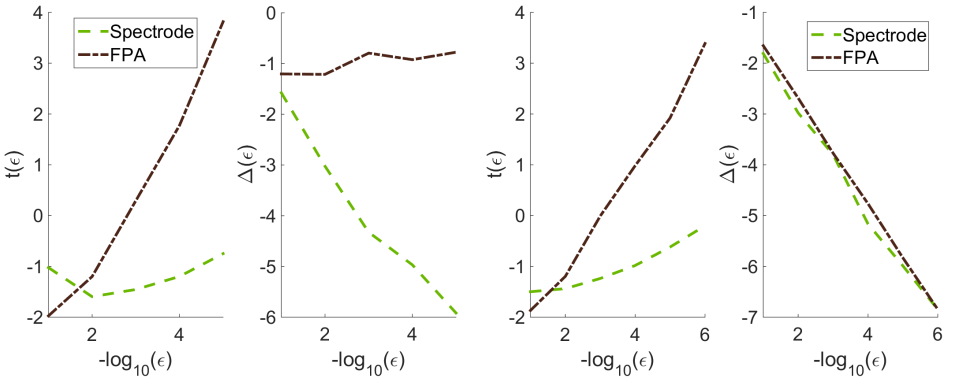
We use the same test problems, `MP` and `TwoPoint`, and the same parameters as in the previous section. For a specified set of inputs H, γ, and accuracy ε, SPECTRODE produces density estimates $\hat{f}(x_i, \varepsilon)$ on a grid x_i, $i = 1, \ldots, I$. We record the running time $t(\varepsilon, H, \gamma)$ of the algorithm, defined as the base ten logarithm of seconds to completion. Times were measured on an Intel i7 2.4 GHz PC. The relative running times are relevant more generally for other systems. We also record the average accuracy in the density, defined as: $\bar{\Delta}(\varepsilon) = -\log_{10}(\sum_{i=1}^{I} |\hat{f}(x_i, \varepsilon) - f(x_i)|/I)$. Here $f(x_i)$ is the true density which is available in both cases.

We repeat this experiment for FPA, which is described later in Algorithm 2 from Sec. 4.1.3. We record the running time $t_{\mathrm{fp}}(\varepsilon, H, \gamma)$ and accuracy $\bar{\Delta}_{\mathrm{fp}}(\varepsilon)$. To ensure comparability, we use the same grid x_i that was produced by SPECTRODE. We set the accuracy parameter η to $\eta = \varepsilon$. We emphasize that η limits the precision due to the smoothing property of Stieltjes transforms (see Lemma 4.9). Therefore, it should be of the same order as ε to get the desired precision; this motivates our choice $\eta = \varepsilon$. Further, we apply an early stopping rule to FPA, due to its long running time. For each grid point, we stop after $1/\varepsilon$ iterations. For this reason, FPA does not always achieve the required accuracy.

3.3.2. *Results*

Figure 4 shows the results, for `MP` in the left panel and for `TwoPoint` in the right panel. The running time $t(\varepsilon, H, \gamma)$ and the accuracy $\bar{\Delta}(\varepsilon)$ of the two methods are displayed as a function of the number of significant digits requested $-\log_{10}(\varepsilon)$.

For the test problem `MP` in subplot (a), the number of significant digits requested varies from one to five, i.e. $\varepsilon = 10^{-1}, \ldots, 10^{-5}$. The running time of SPECTRODE is below 0.5 s, regardless of the accuracy requested, and produces the required average accuracy. The running time of FPA increases approximately linearly in $1/\varepsilon$, reaching ~ 5000 s for $\varepsilon = 10^{-5}$. At the same time, the average accuracy is always about one digit. In this example SPECTRODE is faster and more accurate at the same time. Had we stopped later, FPA would have taken even longer to converge.



(a) `MP` (b) `TwoPoint`

Fig. 4. Running time (base 10 logarithm) versus accuracy on two test problems (Sec. 3.3). We show the log-running time ($t(\varepsilon, H, \gamma)$, left subplot in (a) and (b)) and average accuracy ($\bar{\Delta}(\varepsilon)$, right subplot in (a) and (b)) of the methods as a function of the number of correct significant digits k in the precision parameter $\varepsilon = 10^{-k}$. Methods: SPECTRODE — dashed, fixed-point — dash-dotted.

This result is worth emphasizing: SPECTRODE is 1000 times faster and 1000 times more accurate than FPA, at least for the highest precision $\varepsilon = 10^{-5}$.

For the test problem `TwoPoint` in subplot (b), the number of significant digits requested now varies from one to six. The running time of SPECTRODE is below 1 s, and it produces the required accuracy. The running time of FPA increases as $\varepsilon \downarrow 0$, reaching about 2000 s for the largest accuracy. FPA also gives approximately the required accuracy. In this case, SPECTRODE is faster than FPA (by three orders of magnitude for $\varepsilon = 10^{-6}$) while producing the same accuracy. These two examples show that SPECTRODE is fast and accurate, and compares favorably to FPA.

4. Theoretical Results

In this section we prove the convergence of our method. Algorithm 1 has the pseudocode.

4.1. *Correctness of* SPECTRODE

SPECTRODE has an user-adjustable accuracy parameter $\varepsilon > 0$. Here we show that as $\varepsilon \to 0$, the output of the algorithm converges to the correct limiting values.

SPECTRODE has the following two main steps:

(1) Find the support of the distribution, as a union of compact intervals.
(2) Compute an approximation of the density on the intervals inside the spectrum.

We will analyze these two parts separately, and give our main results in Theorems 4.1 and 4.2. We will focus on the case $\gamma < 1$; the case $\gamma > 1$ is similar and therefore omitted.

Algorithm 1 SPECTRODE: computation of the limit ESD

procedure SPECTRODE
input
  $t_1, \dots, t_J \leftarrow$ positive eigenvalues
  $w_1, \dots, w_J \leftarrow$ weights $w_i > 0, \sum w_i = 1$
  $\gamma \leftarrow$ aspect ratio $\gamma \neq 1$
  $\varepsilon \leftarrow$ precision parameter
begin
  With accuracy $\varepsilon > 0$, find all intervals $(\hat{a}_k, \hat{b}_k)$ where $z(v)$ (2.7) is increasing $(\hat{a}_k < \hat{b}_k < \hat{a}_{k+1})$
  Define the support intervals $[\hat{l}_k(\varepsilon), \hat{u}_k(\varepsilon)]$:
  if $\gamma < 1$ **then**
    set $\hat{l}_k(\varepsilon) = z(\hat{b}_k)$ and $\hat{u}_k(\varepsilon) = z(\hat{a}_{k+1})$ for all $k \leq J-1$
  else
    set $\hat{l}_1(\varepsilon) = z(\hat{b}_J)$, $\hat{l}_k(\varepsilon) = z(\hat{b}_{k-1})$ for all $2 \leq k \leq J-1$, and $\hat{u}_k(\varepsilon) = z(\hat{a}_k)$ for all $k \leq J-1$
  Set $\hat{K}(\varepsilon)$ to the number of intervals $[\hat{l}_k(\varepsilon), \hat{u}_k(\varepsilon)]$
  for intervals $[\hat{l}_k(\varepsilon), \hat{u}_k(\varepsilon)]$ **do**
    Approximate by $\hat{v}_k$ the value $v_k = v(\hat{l}_k(\varepsilon) + i\delta)$; $\delta = \varepsilon^2$ using FPA (Algorithm 2) with accuracy $\eta = \varepsilon$.
    Define a uniform grid $\hat{l}_k = x_{k0} < \cdots < x_{kM} = \hat{u}_k$ with $\lceil \varepsilon^{-1/2} \rceil$ elements
    Solve the ODE (2.6) starting at $\hat{v}_k$ to find the values $\hat{v}(x_{kj} + i\delta)$
    Compute $\hat{f}(x_{kj}, \varepsilon) = \text{Imag}\{\hat{m}(x_{kj} + i\delta)\}/\pi$, with $\hat{m}$ from (2.2)
return $\hat{K}(\varepsilon)$; support intervals $[\hat{l}_k(\varepsilon), \hat{u}_k(\varepsilon)]$. Within estimated support intervals, define $\hat{f}(x)$ by linear interpolation. Outside the estimated support define $\hat{f}(x) = 0$.

4.1.1. *Summary of Silverstein and Choi's results*

We rely in an essential way on the results of Silverstein and Choi in [33]. For the reader's convenience, we summarize below the required results. For any cumulative distribution function G on the real line, define the complement of the support of G, S_G^c, by $S_G^c = \{x \in \mathbb{R} : \text{there is an open neighborhood } N \text{ of } x, \text{ such that } G(x) \text{ is constant on } N\}$.

The support of a distribution function G is defined as $S_G = \mathbb{R} \backslash S_G^c$. The companion distribution function $\underline{F}$ is the limit ESD of $n^{-1}\mathbf{X}\mathbf{X}^\top$, and satisfies

$$\underline{F} = \gamma F + (1-\gamma) I_{[0,\infty)}. \tag{4.1}$$

We summarize the needed results, some of which were stated informally earlier in the paper.

Lemma 4.1 (Silverstein and Choi [33]). *Let F be the limit ESD of covariance matrices with limit PSD H, and aspect ratio $\gamma < 1$. It holds that*:

(1) *F has a continuous density $f(x)$ for all x.*
(2) *f is analytic in the neighborhood of any point x such that $f(x) > 0$.*
(3) *Let $B = \{m \in \mathbb{R} : m \neq 0, -m^{-1} \in S_H^c\}$. Then m belongs to B if and only if $z(m)$ belongs to $S_{\underline{F}}^c$ and $z'(m) > 0$. This characterizes the support of $\underline{F}$ and thus also that of F.*
(4) *Suppose the PSD is an atomic distribution with J point masses. The number of disjoint intervals (a, b) in S_F^c such that $a, b \in S_F$ is at most $J - 1$. Therefore the support is the union of at most J disjoint compact intervals.*

The following is a restatement of [4, Lemma 6.2], see also [9, Theorem 7.5].

Lemma 4.2 (consequences of [33], see [4, Lemma 6.2]). *Suppose the PSD H is an atomic distribution with $t_i > 0$ for all $i = 1, \ldots, J$, and $\gamma < 1$. Then*

(1) *F is compactly supported.*
(2) *The density $f(x)$ equals zero in some right-neighborhood $(0, a)$ of 0.*
(3) *Define $D = -1/\min_j t_j < 0$. There is a finite constant $b_1 < D$, such that $z'(b_1) = 0$, while $z'(v) > 0$ for $v \in (-\infty, b_1)$ and $z'(v) < 0$ for $v \in (b_1, D)$. Then $l_1 = z(b_1)$ is the lowest endpoint in the support of F.*
(4) *Suppose the t_i are sorted: $t_1 < t_2 < \cdots < t_J$. Then within each interval $(-1/t_i, -1/t_{i+1})$, $z''(v)$ has a unique zero c_i. There are at most two points within each such interval where $z'(v) = 0$. If these points exist, we denote them by $a_j \leq b_j$, indexing them in an increasing order by j; including $a_1 = -\infty$ and b_1 defined above. Then $a_j \leq c_i \leq b_j$ and z is increasing on $[a_j, b_j]$.*

Therefore, the support of F is the union of some intervals $[l_i, u_i]$, where $0 < l_1 < u_1 < \cdots < l_K < u_K$. The density equals $f(x) = 0$ on $(0, l_1]$, on $[u_i, l_{i+1}]$ for all i, and on $[u_K, \infty)$. Within the intervals $[l_i, u_i]$, the density $f(x)$ is usually strictly positive. However, there are cases in which the density $f(x) = 0$ at isolated points within $[l_i, u_i]$. This can happen if $[l_i, u_i]$ arises from a "merge" when two intervals corresponding to neighboring population eigenvalues "just touch". We emphasize that in this case we consider $[l_i, u_i]$ as one component of the support of F.

The roots of z' are connected to the edges by $z(b_k) = l_k$, for all $1 \leq k \leq K$, and $z(a_{k+1}) = u_k$, for all $1 \leq k \leq K - 1$. It also follows from the references above that the upper boundary u_K equals $z(a_{K+1})$, where a_{K+1} is the unique root of $z'(v) = 0$ on $v \in (-1/t_1, 0)$.

4.1.2. *Correctness of the support*

Here we explain in detail the steps to find the support of F, and prove their correctness.

Theorem 4.1. ***Correctness of $\hat{K}(\varepsilon)$, $\hat{l}_k(\varepsilon)$, and $\hat{u}_k(\varepsilon)$:*** *Consider the numerical approximations $\hat{K}(\varepsilon)$, $\hat{l}_k(\varepsilon)$, $\hat{u}_k(\varepsilon)$ outlined in Algorithm 1 and described in detail below. Then, as $\varepsilon \to 0$:*

(1) *The number of disjoint intervals is correctly identified*: $\lim_{\varepsilon\to 0} \hat{K}(\varepsilon) = K$.
(2) *The endpoints of the support are accurately approximated*: $\lim_{\varepsilon\to 0} \hat{l}_k(\varepsilon) = l_k$, $\lim_{\varepsilon\to 0} \hat{u}_k(\varepsilon) = u_k$.

This theorem is a consequence of Lemmas 4.3 and 4.4. Our strategy, following Lemmas 4.1 and 4.2, is to find the intervals where z is increasing, and the points where it switches monotonicity. Once we find such a switch point $\hat{a}$, we will define $z(\hat{a})$ as an approximate support endpoint. In detail, by Lemma 4.2(4), $l_1 = z(b_1)$ is the lowest endpoint in the support of F, where b_1 is the largest point such that z is increasing on $(-\infty, b_1]$. Therefore, in our Algorithm 1, line 4, we set the leftmost interval where z is increasing as $(\hat{a}_1, \hat{b}_1)$, with $a_1 = -\infty$, and $\hat{b}_1(\varepsilon)$ is the numerical estimate of b_1 found below.

To find $\hat{b}_1(\varepsilon)$, recall $D = -1/\min_j t_j < 0$ and consider an interval $[L(\varepsilon), D]$ depending on ε. For $L(\varepsilon)$ small enough, $b_1 \in [L(\varepsilon), D]$. To solve the equation $z'(b_1) = 0$, one may think of using Newton's method. However, Newton's method is not guaranteed to converge in general, therefore we will in this paper consider Brent's method [8], which has guaranteed convergence. We run Brent's method on $[L(\varepsilon), D]$, with precision $h(\varepsilon)$ to be specified, to find a value $\hat{b}_1(\varepsilon)$ such that $z'(\hat{b}_1(\varepsilon)) \approx 0$. Then we define $\hat{l}_1(\varepsilon) = z(\hat{b}_1(\varepsilon))$ as a numerical approximation of $l_1 = z(b_1)$. The following lemma ensures the convergence of this procedure.

Lemma 4.3. ***Correctness of $\hat{l}_1(\varepsilon)$:*** *Choose the function $h(\varepsilon)$ such that $\lim_{\varepsilon\to 0} h(\varepsilon) = 0$, and choose the function L such that $\lim_{\varepsilon\to 0} L(\varepsilon) = -\infty$. Then $\lim_{\varepsilon\to 0} \hat{l}_1(\varepsilon) = l_1$.*

Proof. By the condition on $L(\varepsilon)$, for small ε the true solution b_1 of $z'(v) = 0$ will belong to the desired interval: $L(\varepsilon) < b_1 < D$. Using the differentiability of z on $(L(\varepsilon), D)$, the convergence of the precision $h(\varepsilon) \to 0$, and the uniqueness of b_1 — from Lemma 4.2(4) — it follows that Brent's method produces a solution such that $\hat{b}_1(\varepsilon) \to b_1$. Since z is continuous on $(-\infty, D)$, we have $\hat{l}_1(\varepsilon) = z(\hat{b}_1(\varepsilon)) \to z(b_1) = l_1$. □

Since z has a singularity at D, in practice we work with intervals of the form $[L(\varepsilon), D - \delta(\varepsilon)]$, and we employ an iterative doubling strategy to ensure that the interval contains the desired solution. The distance from $L(\varepsilon)$ to D is iteratively doubled and $\delta(\varepsilon)$ is iteratively halved until the continuous function $z'(v)$ changes sign on $[L(\varepsilon), D - \delta(\varepsilon)]$.

We have shown that the numerical approximation to the lowest endpoint of F converges. Similarly, we define the remaining endpoints for each interval $(-1/t_i, -1/t_{i+1})$. By Lemma 4.2(4), there is a unique point $c_i \in (-1/t_i, -1/t_{i+1})$

such that $z''(c_i) = 0$. By inspection the function $v^2 z''(v)$ is monotonic on that interval. We use Brent's method on $(-1/t_i, -1/t_{i+1})$ to solve the equation $v^2 z''(v) = 0$ with accuracy $h(\varepsilon)$, and denote the approximation by $\hat{c}_i(\varepsilon)$.

Then we solve the equation $z'(v) = 0$ on $(-1/t_i, c_i)$ (respectively, $(c_i, -1/t_{i+1})$) using Brent's method, and denote the approximated solution by $\hat{a}_k(\varepsilon)$ (respectively, $\hat{b}_k(\varepsilon)$), where we index the solutions by k in an increasing order for different intervals. Brent's method will return an indicator variable $Q_i = 0$ if there is no solution in at least one of the two halves within the allowed error margin. As described in line 5 of Algorithm 1, we set $\hat{l}_k(\varepsilon) = z(\hat{b}_k(\varepsilon))$ for $2 \le k \le J-1$ and $\hat{u}_k(\varepsilon) = z(\hat{a}_{k+1}(\varepsilon))$ for all $k \le J-1$. Finally, we set $\hat{K}(\varepsilon)$ as the number of intervals $(\hat{a}_k, \hat{b}_k)$ constructed. The lemma below ensures the convergence of this procedure.

Lemma 4.4. ***Correctness of $\hat{K}(\varepsilon)$, $\hat{l}_k(\varepsilon)$ $k \ge 2$, and $\hat{u}_k(\varepsilon)$:*** *Choose the function $h(\varepsilon)$ such that $\lim_{\varepsilon\to 0} h(\varepsilon) = 0$. Then the number of intervals is correctly identified in the high-precision limit*: $\lim_{\varepsilon\to 0} \hat{K}(\varepsilon) = K$. *Further, in addition to l_1, the approximations to the other endpoints of the support converge to the right answers*: $\lim_{\varepsilon\to 0} \hat{l}_k(\varepsilon) = l_k$ *for all $k \ge 2$, and* $\lim_{\varepsilon\to 0} \hat{u}_k(\varepsilon) = u_k$ *for all k.*

Proof. This is analogous to the previous proposition. By Lemma 4.1(3), we must find all intervals where $z(v)$ is increasing. By Lemma 4.2(4), there is at most one increasing interval within every $(-1/t_i, -1/t_{i+1})$. We already found $\hat{l}_1$ in the previous lemma, so we exclude the interval $(-\infty, -1/t_1)$.

Since the accuracy $h(\varepsilon) \to 0$, and since $v^2 z''(v) = 0$ has a unique solution c_i on $(-1/t_i, -1/t_{i+1})$, Brent's method produces a solution that converges: $\hat{c}_i(\varepsilon) \to c_i$. Next, by Lemma 4.2(4), the equation $z'(v) = 0$ has at most one solution a_j on $(-1/t_i, c_i)$. Since z is differentiable on this interval, for sufficiently high precision Brent's method will find all intervals where there is a solution, and will return no solution otherwise. Furthermore, by the uniqueness of the solution, Brent's method produces an approximation that converges: $\hat{a}_k(\varepsilon) \to a_k$. Therefore, for all values a_k we have $\hat{a}_k(\varepsilon) \to a_k$; similarly $\hat{b}_k(\varepsilon) \to b_k$.

Now z is continuous at a_{k+1}, because the only points of discontinuity are at the values $-1/t_i$, and at 0. Thus $\lim_{\varepsilon\to 0} z(\hat{a}_{k+1}(\varepsilon)) = z(a_{k+1}) = u_k$. Similarly $\lim_{\varepsilon\to 0} \hat{l}_k(\varepsilon) = l_k$ for all $k \ge 2$, and $\lim_{\varepsilon\to 0} \hat{u}_K(\varepsilon) = u_K$. □

Lemmas 4.3 and 4.4 together imply Theorem 4.1.

4.1.3. *Correctness of the density*

In this section, we prove that our method computes the density accurately.

Theorem 4.2. ***Correctness of $\hat{f}(x, \varepsilon)$:*** *Consider the numerical approximation to the density $\hat{f}(x, \varepsilon)$, outlined in Algorithm* 1 *and described in detail below. Then, as $\varepsilon \to 0$, the approximation converges to the true density uniformly over all x :* $\sup_{x\in\mathbb{R}} |\hat{f}(x, \varepsilon) - f(x)| \to 0$.

Table 2. Definitions used in the proof of Theorem 4.2.

Name	Definition	Defined in equations	Analyzed in lemmas
$x_j(\varepsilon)$	Grid	(4.4)	
$f(x)$	True density	(2.4)	4.5
$f(x_j(\varepsilon), \varepsilon)$	Solution to exact ODE	(4.5)	4.6
$\tilde{f}(x_j(\varepsilon), \varepsilon)$	Solution to inexact start ODE	(4.6)	4.7, 4.8
$\hat{f}(x_j(\varepsilon), \varepsilon)$	Euler's method approximation	(4.8)	4.5, 4.8

This theorem is proved in at the end of this section, using the tools developed in the following lemmas. The approximation to the density is defined in several stages. For the reader's convenience, we provide Table 2 summarizing the notation and definitions for each stage.

In the first stage, outside the support intervals $[\hat{l}_k(\varepsilon), \hat{u}_k(\varepsilon)]$ we define $\hat{f}(x, \varepsilon) = 0$. We also set $\hat{f}(\hat{l}_k(\varepsilon), \varepsilon) = \hat{f}(\hat{u}_k(\varepsilon), \varepsilon) = 0$. Since the approximated support converges, we see that the estimated density converges to zero uniformly outside of the support. It remains to handle the support intervals.

Consider a support interval $[l, u]$, where $l < u$ are the true endpoints of a connected component of the support of F. Denote the estimated support by $[\hat{l}, \hat{u}]$, and let $\hat{l} = x_0 < x_1 < \cdots < x_M = \hat{u}$ be a uniformly spaced grid $x_i = x_i(\varepsilon)$ of length $M = M(\varepsilon)$ depending on ε, on which we will approximate the density $\hat{f}(x_j)$. In Algorithm 1, we specified $M = \lceil \varepsilon^{-1/2} \rceil$ for concreteness. We will see in the proof that more general choices of grids work. For this reason, we will not specify the grid at the moment, and instead only require that the spacings tend to zero: $|x_{i+1} - x_i| \leq h(\varepsilon)$, and $\lim_{\varepsilon \to 0} h(\varepsilon) = 0$.

Next, we reduce the approximation problem to the grid x_i. As explained in Algorithm 1, for x within the estimated support and not necessarily on the grid, define the linear interpolation $\hat{f}(x, \varepsilon) = \alpha \hat{f}(x_i, \varepsilon) + (1 - \alpha)\hat{f}(x_{i+1}, \varepsilon)$, where $x_i \leq x < x_{i+1}$, and $x = \alpha x_i + (1-\alpha)x_{i+1}$. This ensures that the estimated density $\hat{f}(\cdot, \varepsilon)$ is continuous. With these definitions, we reduce uniform convergence over all x to uniform convergence only on the grid.

Lemma 4.5. *To show the convergence in Theorem* 4.2, *it is enough to show that the density approximations converge uniformly on the grid* $x_i = x_i(\varepsilon)$. *Equivalently, if we have the convergence on the grid*

$$\lim_{\varepsilon \to 0} \max_{0 \leq i \leq M(\varepsilon)} |\hat{f}(x_i(\varepsilon), \varepsilon) - f(x_i(\varepsilon))| = 0, \tag{4.2}$$

then the convergence over all x *follows*:

$$\lim_{\varepsilon \to 0} \sup_{x \in \mathbb{R}} |\hat{f}(x, \varepsilon) - f(x)| = 0. \tag{4.3}$$

Proof. It is easy to check that by construction, $l_i \le \hat{l}_i \le \hat{u}_i \le u_i$ for all support intervals. Therefore, we have for any x

$$\hat{f}(x,\varepsilon) - f(x) = \begin{cases} 0 & \text{if } x \notin [l_i, u_i], \text{ for any } i, \\ -f(x) & \text{if } l_i \le x \le \hat{l}_i, \text{ or } \hat{u}_i \le x \le u_i \text{ for some } i, \\ \hat{f}(x,\varepsilon) - f(x) & \text{if } \hat{l}_i \le x \le \hat{u}_i, \text{ for some } i. \end{cases}$$

The convergence claim made by Lemma 4.5 is clear for the first case. For the second case, note that there are only finitely many support intervals. Therefore it is enough to show $\lim_{\varepsilon\to 0} \sup_{x: l_i \le x \le \hat{l}_i} |f(x)| = 0$ for all i, and the analogous statement for upper endpoints. We showed earlier in Proposition 4.1 that $\hat{l}_i \to l_i$. By continuity of f, this shows the desired claim $\sup_{x: l_i \le x \le \hat{l}_i} |f(x)| \to 0$ for the second case, when x is in $[l_i, \hat{l}_i]$.

The third case is the most interesting one. Consider any x such that $\hat{l}_i \le x \le \hat{u}_i$. There are two neighbors in the grid such that $x_i(\varepsilon) \le x < x_{i+1}(\varepsilon)$. By the triangle inequality, we can bound

$$|\hat{f}(x,\varepsilon) - f(x)| \le |\hat{f}(x,\varepsilon) - \hat{f}(x_i,\varepsilon)| + |\hat{f}(x_i,\varepsilon) - f(x_i)| + |f(x_i) - f(x)|.$$

Recall that the maximum spacing was bounded: $|x_{i+1} - x_i| \le h(\varepsilon)$. Let us denote a modulus of continuity for a function g by ω. This function ω enjoys $|g(x) - g(y)| \le \omega(|x-y|, g)$ for any x, y. Taking the maximum over all $x \in S_i$ (where $S_i = [\hat{l}_i, \hat{u}_i]$) in the previous display, we obtain:

$$\sup_{x\in S_i} |\hat{f}(x,\varepsilon) - f(x)| \le \omega(h(\varepsilon), \hat{f}) + \max_{0\le i\le M(\varepsilon)} |\hat{f}(x_i,\varepsilon) - f(x_i)| + \omega(h(\varepsilon), f).$$

Since $f, \hat{f}$ are continuous and compactly supported, they are uniformly continuous. Therefore, as $h(\varepsilon) \to 0$, we get $\omega(h(\varepsilon), f) \to 0$, and similarly for $\hat{f}$. Assuming (4.2), we hence obtain the desired claim that the density converges uniformly for all x:

$$\lim_{\varepsilon\to 0} \sup_{x\in S_i} |\hat{f}(x,\varepsilon) - f(x)| = 0. \qquad \square$$

We will now focus on showing the convergence on the grid. Our algorithm relies on an ODE, whose starting point is obtained via the fixed-point algorithm. In [10] FPA is presented for a more general class of problems; for the reader's convenience, we describe the special case needed in Algorithm 2.

The fixed-point algorithm is a method for solving the Silverstein equation. It is based on the observation that (2.5) is a fixed-point equation $v = -1/h(v;z)$, for any given z. Then one defines a starting point $v_0 = -1/z$, and iterates $v_{n+1} = -1/h(v_n;z)$ until the convergence criterion $|1/v_n + h(v_n;z)| \le \eta$ is met. Let $\hat{v}(z,\eta)$ be the solution produced by the FPA, with accuracy control parameter $\eta > 0$. It was shown in [10] that, for any fixed $z \in \mathbb{C}^+$, $\hat{v}(z,\eta)$ converges to the unique solution of the Silverstein equation (2.5) with positive imaginary part, as $\eta \to 0$: $\hat{v}(z,\eta) \to v(z)$.

Algorithm 2 FPA: Fixed-point algorithm

procedure FPA
input
  $H \leftarrow$ population eigenvalue distribution
  $\gamma \leftarrow$ aspect ratio
  $\eta \leftarrow$ accuracy parameter (> 0)
  $z \leftarrow$ complex argument $\in \mathbb{C}^+$. If the argument x is real, then $z \leftarrow x + i\eta^2$
initialize:
  $v_0 \leftarrow -1/z$
  $n \leftarrow 0$
  $h(v) := z - \gamma \int \frac{t}{1+tv}\, dH(t)$
while $|1/v_n + h(v_n)| > \eta$
  $v_{n+1} \leftarrow -1/h(v_n)$
  $n \leftarrow n + 1$.
end;
  $m_n \leftarrow \gamma^{-1} v_n + (\gamma^{-1} - 1)/z$
return $\hat{v}(z, \eta) = v_n$; $\hat{f}(x, \eta) = \text{Imag}(m_n)/\pi$, where $x = \text{Re}(z)$

The accuracy parameter η in FPA is important. If the algorithm is used to approximate the density of the ESD at x, with accuracy η — as in the section comparing the different methods — we run FPA at $z = x + i\eta^2$. The scaling η^2 is motivated by Lemma 4.9. The proof of this lemma shows that the true solution at imaginary part ε guarantees an accuracy $\varepsilon^{1/2}$. Therefore, to get accuracy η, we go down to imaginary part η^2. Further, we use the same threshold η in the stopping criterion. In principle, these two parameters could be decoupled, but this simple choice suffices for our purposes.

Therefore, to find the starting point, we use FPA for $z = \hat{l} + i\delta(\varepsilon)$, where $\hat{l}$ is the approximation to the lower endpoint of the current support interval, and $\delta(\varepsilon) = \varepsilon^2$. The accuracy parameter η is set to $\eta = \varepsilon$. This gives a starting point $\hat{v}_0 = \hat{v}(\hat{l} + i\delta(\varepsilon))$. We will first analyze the case of an exact starting point $v_0 = v(\hat{l} + i\delta(\varepsilon))$, in Lemma 4.6. That is, we argue about the case when the solution of the Silverstein equation has been found *exactly* by FPA. In Lemma 4.7 we extend the argument to the inexact starting point $\hat{v}_0$.

We call Eq. (2.6) started at the true solution $v_0 = v(\hat{l} + i\delta(\varepsilon))$, the *exact ODE*. This exact ODE has the right solution over the whole relevant interval.

Lemma 4.6. ***Correctness of the exact ODE*:** *Consider the ODE* (2.6) *for the complex-valued function r of a real variable x*

$$\frac{dr}{dx} = \frac{1}{\frac{1}{r^2} - \gamma \sum_{i=1}^{J} \frac{w_i t_i^2}{(1+t_i r)^2}}, \quad r(x_0) = v_0.$$

Let the starting point be the exact solution $v_0 = v(x_0 + i\delta(\varepsilon))$. Then this equation has a unique solution on $[x_0, \hat{u}]$, and this solution is $r(x) = v(x + i\delta(\varepsilon))$.

Proof. The ODE was obtained by differentiating the Silverstein equation (2.5) for v. Since $r(x) = v(x + i\delta(\varepsilon))$ obeys the Silverstein equation and also satisfies the starting condition, it is clearly a solution.

To show that the solution is unique, consider the general ODE $y' = g(y), y(x_0) = y_0$. It is well known (e.g. [17, Theorem 7.4, p. 39]) that the solution is unique on the open set $(x, y) \in U \subset \mathbb{R} \times \mathbb{C}$ where $g(y), g'(y)$ are continuous. If started at any point $(x_0, y_0) \in U$, the solution can be continued to the boundary of U.

In our case $g(r)$ is continuous on the entire image $v(\mathbb{C}^+) = \{y : y = v(z), \text{ for some } z \in \mathbb{C}^+\}$. Indeed, let y_0 be an arbitrary element of $v(\mathbb{C}^+)$, so that $y_0 = v(z_0)$ for some $z_0 \in \mathbb{C}^+$. Then by the definition of the ODE, $v'(z_0) = g(y_0)$. Now v is analytic on $\mathbb{C}^+$, so clearly $v'(z)$ is well-defined and continuous at z_0. By the expression for $v'(z)$, $v'(z) = 1/k(v)$ for some complex function k, we see that $v'(z) \neq 0$. Hence, by the inverse function theorem, v is invertible near y_0, so that locally $z = v^{-1}(y)$ in a neighborhood of y_0. Therefore, locally near y_0, $g(y) = v'(v^{-1}(y))$. This shows that g is continuous near y_0. Hence, $g(y)$ is continuous on the entire image $v(\mathbb{C}^+)$. By a similar argument $g'(y)$ is continuous on the entire image $v(\mathbb{C}^+)$.

This shows that for our problem U contains $\mathbb{R} \times v(\mathbb{C}^+)$. Clearly we start at a point (x_0, v_0) in U. By the result cited above, the solution to the ODE is unique on the entire set $x \in \mathbb{R}$, and in particular on $[x_0, \hat{u}]$, finishing the proof. □

Next, we will argue that even with an inexact starting point for the ODE, the solutions are still nearly exact. Suppose that FPA produces an estimate $\tilde{v}_0 = \hat{v}(z, \eta)$ of v_0. We call the ODE (2.6) started at $\tilde{v}_0$ the *inexact start ODE*. The difference $c(\varepsilon) = \tilde{v}_0 - v_0$ can be made arbitrarily small by taking η sufficiently close to zero in FPA. The following lemma ensures that the solution to the inexact start ODE stays close to the true solution.

Lemma 4.7. ***Correctness of the inexact start ODE*:** *Consider the ODE* (2.6) *as given in Lemma* 4.6, *but started at $\tilde{v}_0 = v_0 + c(\varepsilon)$, where $\tilde{v}_0$ is the solution produced by FPA for starting point $z = \hat{l} + i\delta(\varepsilon)$ and for sufficiently small η. Then, for sufficiently small ε, this inexact start ODE has a unique solution $\tilde{r}$ on $[\hat{l}, \hat{u}]$, which obeys $|\tilde{r}(x) - r(x)| = O(c(\varepsilon))$ uniformly over all $x \in [\hat{l}, \hat{u}]$.*

Proof. First we will show the uniqueness of the solution. By [10, Proposition 1], Algorithm 2 started at $z = \hat{l} + i\delta(\varepsilon)$, and with accuracy η, produces a solution $\tilde{v}_0 = \hat{v}(z, \eta)$ such that $\hat{v}(z, \eta) \to v(z)$ as $\eta \to 0$. Therefore, for sufficiently small η, we can ensure that $\tilde{v}_0 - v_0 = O(c(\varepsilon))$ for an arbitrary small $c(\varepsilon)$.

By the open mapping theorem, $v(\mathbb{C}^+)$ is an open set containing the true solution curve $r(x) = v(x + i\delta(\varepsilon))$. Hence, for sufficiently small $c(\varepsilon)$, one has $\tilde{v}_0 \in v(\mathbb{C}^+)$. Therefore, $\tilde{v}_0 = v(a + ib)$ for some a and $b > 0$. Then, by the same argument as in

Lemma 4.6, the solution $\tilde{r}$ exists and is unique, and is given by $\tilde{r}(x) = v((x - x_0) + a + ib)$. This solution exists for all x and belongs to $v(\mathbb{C}^+)$.

Next, we will use a general inequality for inexact start ODEs for the quantitative bound. The "fundamental lemma" [17, Theorem 10.2, p. 58] states: Consider the ODE $y' = g(y)$, and let $y, \tilde{y}$ be two solutions with starting points $y(x_0), \tilde{y}(x_0)$. Suppose $|g'(y)| \leq L$ on a connected set K_0 containing the two solution curves $y, \tilde{y}$ for $x \in [x_0, x_M]$. Then the solutions $y, \tilde{y}$ are close to each other for all $x \in [x_0, x_M]$, specifically: $|y(x) - \tilde{y}(x)| \leq |y(x_0) - \tilde{y}(x_0)| \exp((x - x_0)L)$.

We have shown in the proof of Lemma 4.6 that $f'(y)$ is continuous on the image $v(\mathbb{C}^+)$. Let K_0 be a compact connected subset of $v(\mathbb{C}^+)$ containing the solution curves $r(x), \tilde{r}(x)$ for $x \in [\hat{l}, \hat{u}]$ (which exist by the above argument). Then f' is bounded by some constant L on K_0. By the fundamental lemma, we have $|\tilde{r}(x) - r(x)| \leq c(\varepsilon) \exp((\hat{u} - \hat{l})L) = O(c(\varepsilon))$, as required. This finishes the lemma. □

Next we introduce notations for the solutions of the two ODEs. As we discussed earlier, the grid $x_i = x_i(\varepsilon)$ is uniformly spaced on $[\hat{l}, \hat{u}]$:

$$\hat{l} = x_0 < \cdots < x_M = \hat{u}. \tag{4.4}$$

The solutions to the exact ODE (2.6) in Lemma 4.6 on the grid are $v(x_j + i\delta(\varepsilon))$. We then define $m(x_j + i\delta(\varepsilon))$ according to (2.2), and

$$f(x_j(\varepsilon), \varepsilon) = \text{Imag}(m(x_j + i\delta(\varepsilon)))/\pi. \tag{4.5}$$

Similarly, with the solutions $\tilde{v}(x_j + i\delta(\varepsilon))$ to the inexact start ODE analyzed in Lemma 4.7 on the same grid define $\tilde{m}(x_j + i\delta(\varepsilon))$ using (2.2), and

$$\tilde{f}(x_j(\varepsilon), \varepsilon) = \text{Imag}(\tilde{m}(x_j + i\delta(\varepsilon)))/\pi. \tag{4.6}$$

Lemma 4.7 shows that

$$|\tilde{f}(x_j(\varepsilon), \varepsilon) - f(x_j(\varepsilon), \varepsilon)| = O(c(\varepsilon)). \tag{4.7}$$

In particular, this bound highlights that the approximation $\tilde{f}$ is uniformly accurate over the grid x_j, as required by Lemma 4.5.

We show next that Euler's method for discretizing the inexact start ODE on the grid x_i produces numerical approximations that converge as $\varepsilon \to 0$. In practice we use a higher order ODE solver, but for simplicity here we consider Euler's method.

Let $\hat{v}_j$ be the sequence produced by Euler's method for the inexact start ODE (2.6) on the discretization $\hat{l} = x_0 < \cdots < x_M = \hat{u}$. Define $\hat{m}_j = \gamma^{-1}\hat{v}_j + (\gamma^{-1} - 1)/z_j$, where $z_j = x_j + i\delta(\varepsilon)$. Also define the density approximations

$$\hat{f}(x_j, \varepsilon) = \text{Imag}(\hat{m}_j)/\pi. \tag{4.8}$$

Then we have the following.

Lemma 4.8. ***Euler's method approximates the true solution to the inexact start ODE:*** *Consider a fixed $\varepsilon > 0$, and suppose that $\max_i |x_{i+1} - x_i| \leq h$. Then $|\hat{f}(x_j(\varepsilon), \varepsilon) - \tilde{f}(x_j(\varepsilon), \varepsilon)| = O(h)$ for sufficiently small h.*

Proof. This lemma is a direct consequence of the well-known error estimate for Euler's method. Consider the general ODE $y' = g(y)$, $y(x_0) = y_0$. Theorem 7.5 on p. 40 of [17] states that the bounds $|g| \leq A$, $|g'| \leq L$ in a neighborhood of the solution imply that the Euler polygons $y_h(x)$, on a grid x_i with maximum spacing at most h, obey $|y_h(x) - y(x)| \leq A(\exp(L(x - x_0)) - 1)h$. In Lemmas 4.6 and 4.7, we have shown for the inexact start ODE (2.6) started at $\tilde{v}_0$ that g, g' are continuous in a neighborhood of the solution, so the bounds A, L exist. We only consider the finite interval $[\hat{l}, \hat{u}]$, therefore the exponential term is bounded. We get $|y_h(x) - y(x)| = O(h)$, as required. □

If we take $h = h(\varepsilon) \to 0$ as $\varepsilon \to 0$, then the above lemma shows that $|\hat{f}(x_j(\varepsilon), \varepsilon) - \tilde{f}(x_j(\varepsilon), \varepsilon)| \to 0$ uniformly over all $x_j(\varepsilon)$. Comparing with Lemmas 4.5 and 4.7 (specifically bound (4.7)), all that remains to show for our main result is that the true density f is well approximated by the Stieltjes-smoothed density $f(\cdot, \varepsilon)$, i.e.: $f(x_j(\varepsilon), \varepsilon) - f(x_j(\varepsilon)) \to 0$. Since the $x_j(\varepsilon)$ depend on ε, it is necessary to show that $f(x, \varepsilon) - f(x) \to 0$ uniformly in x, where recall that $f(x, \varepsilon) = \text{Imag}(m(x + i\delta(\varepsilon)))/\pi$.

Lemma 4.9. ***Approximation of a density by its Stieltjes transform:*** *Let f be a bounded probability density function. Denote by $m(z)$ its Stieltjes transform: $m(z) = \int f(x)/(x - z)dx$. Suppose f is uniformly continuous. Then as $\varepsilon \to 0$:*

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{\pi} \text{Imag}(m(x + i\varepsilon)) - f(x) \right| \to 0,$$

Proof. A simple calculation reveals

$$f(x, \varepsilon) := \frac{1}{\pi} \text{Imag}(m(x + i\varepsilon)) = \frac{1}{\pi} \int \frac{\varepsilon f(t) dt}{(t - x)^2 + \varepsilon^2}.$$

Clearly

$$1 = \frac{1}{\pi} \int \frac{\varepsilon dt}{(t - x)^2 + \varepsilon^2}, \tag{4.9}$$

therefore we can define g such that

$$f(x, \varepsilon) - f(x) = \frac{1}{\pi} \int \frac{\varepsilon (f(t) - f(x)) dt}{(t - x)^2 + \varepsilon^2} = \int g(t) dt.$$

We will bound this by breaking it down into two integrals: one near x and another far from x. Let $\eta > 0$ be the parameter determining the split to be chosen later. We bound first the integral of g near x:

$$\left| \int_{|t-x| \leq \eta} g(t)\, dt \right| \leq \frac{1}{\pi} \int_{|t-x| \leq \eta} \frac{\varepsilon dt}{(t - x)^2 + \varepsilon^2} \cdot \sup_{t : |t-x| \leq \eta} |f(t) - f(x)| \leq \omega(\eta, f).$$

Above we used (4.9) to upper bound the integral term; and we introduced ω, a modulus of continuity for f. This gives a bound for the integral of g near x.

Next we bound the integral away from x:

$$\left|\int_{|t-x|>\eta} g(t)dt\right| \leq 2|f|_\infty \frac{1}{\pi}\int_{|t-x|>\eta} \frac{\varepsilon dt}{(t-x)^2+\varepsilon^2}.$$

The integral term in the upper bound can be evaluated explicitly as $\pi - 2\arctan(\eta/\varepsilon)$, and can be bounded above by $\pi\varepsilon/\eta$.

Putting everything together, we find the bound

$$|f(x,\varepsilon) - f(x)| \leq \omega(\eta, f) + \frac{2|f|_\infty \varepsilon}{\eta}.$$

Choosing $\eta = \varepsilon^\alpha$ with $\alpha \in (0,1)$ yields a bound that tends to zero uniformly over x, when $\varepsilon \to 0$. The modulus of continuity tends to zero because f is uniformly continuous. Thus we have shown the desired claim and finished the proof.

Note that, if f is continuously differentiable in the neighborhood of a point x, then we obtain an optimal bound on $|f(x,\varepsilon) - f(x)|$ by taking $\eta = c\varepsilon^{1/2}$. This guarantees $|f(x,\varepsilon) - f(x)| = O(\varepsilon^{1/2})$. The square root scaling in ε motivates us to work on the line with imaginary part $\delta = \varepsilon^2$ throughout the paper, to get accuracy of order ε. □

The previous results can be put together to prove Theorem 4.2.

Proof of Theorem 4.2. We shall show the uniform convergence of the density approximation $\hat{f}$. By Lemma 4.5, we only need to show the uniform convergence on the grid $x_i(\varepsilon)$ within the support intervals. Let $[l,u]$ be such an interval, let $[\hat{l},\hat{u}]$ be the approximation produced by SPECTRODE for some ε. Further, $x_i(\varepsilon)$ is the uniformly spaced grid $\hat{l} = x_0 < x_1 < \cdots < x_M = \hat{u}$ of length $M = \lceil \varepsilon^{-1/2} \rceil$. Lemma 4.8 is applicable to this grid, because the grid width scales as $\propto \varepsilon^{1/2} \to 0$. By this lemma, $\max_j |\hat{f}(x_j(\varepsilon),\varepsilon) - \tilde{f}(x_j(\varepsilon),\varepsilon)| \to 0$.

Further, Lemma 4.7 (specifically bound (4.7)) applies if ε is sufficiently small; and if for fixed ε, the accuracy parameter $\eta = \eta(\varepsilon)$ in FPA is sufficiently small. By bound (4.7), and because $c(\varepsilon) \to 0$, we get $\max_j |\tilde{f}(x_j(\varepsilon),\varepsilon) - f(x_j(\varepsilon),\varepsilon)| \to 0$.

On the other hand, f is continuous and compactly supported, hence uniformly continuous. Therefore, by Lemma 4.9: $\sup_x |f(x,\varepsilon) - f(x)| \to 0$. Putting everything together: $\max_j |\hat{f}(x_j(\varepsilon),\varepsilon) - f(x_j(\varepsilon))| \to 0$, which gives the result by Lemma 4.5. □

4.2. *Non-atomic measures*

Our algorithm makes sense for general non-atomic limit PSDs. Indeed, for general PSD H the ODE (2.6) takes the form

$$\frac{dv}{dx} = \mathcal{F}(v) := \frac{1}{\frac{1}{v^2} - \gamma\int \frac{t^2 dH(t)}{(1+tv)^2}}, \quad v(x_0) = v_0.$$

This ODE can be implemented and solved efficiently as long as the integral in the denominator is convenient to compute. Lemma 4.1 provides the means to find

the support, and FPA converges to a starting point even at this level of generality. Therefore the ODE approach is more generally applicable than the atomic measures discussed in this paper.

In our current implementation of SPECTRODE, we go beyond mixtures of point masses and also allow mixtures of uniform distributions, so $H = \sum_{i=1}^{J} w_i \delta_{t_i} + \sum_{t=1}^{T} w_t^* U_{a_t,b_t}$, where $U_{a,b}$ is a uniform distribution on $[a, b]$. An example was shown in Fig. 1. To efficiently support uniform distributions, we compute in closed form the integrals that appear in FPA iteration in the function $h(v)$ in Algorithm 2, and in the ODE. We use the formulas:

$$\int_{-\infty}^{\infty} \frac{t dU_{a,b}(t)}{1+tv} = \frac{1}{v} - \frac{\log \frac{bv+1}{av+1}}{(b-a)v^2},$$

$$\int_{-\infty}^{\infty} \frac{t^2 dU_{a,b}(t)}{(1+tv)^2} = \frac{1}{v^2} - \frac{2\log \frac{bv+1}{av+1} + \frac{1}{bv+1} - \frac{1}{av+1}}{(b-a)v^3}.$$

Armed with these, we obtain efficient computation with arbitrary finite mixtures of uniform distributions.

In addition, a large part of the analysis holds true. Specifically, the convergence of FPA and the analysis of the ODE do not use the atomic structure directly. Instead, the atomic structure of the PSD is used through the structure of the support of the ESD F as a union of compact intervals; and the behavior of z' characterizing the support (Lemma 4.1(3)).

To our knowledge, these claims are currently not known to hold for more general PSDs. Indeed, in the very recent related work [15] examining the fluctuations of the eigenvalues at the edges of the support, the authors work conditionally, assuming that the edges are regular in a certain sense. Extending the validity of our algorithm would presumably require developing a better understanding of the support of ESDs for general PSDs. This could be an interesting direction for future research.

5. Applications

In this section, we apply SPECTRODE to compute moments of the limit ESD and contour integrals of its Stieltjes transform.

5.1. *Moments of the ESD*

The uniform convergence of the approximated density allows us to compute general moments of the ESD. These moments have many applications, see [35, 9, 37].

Obtaining the moments of the ESD is in general difficult. The polynomial moments $\mathbb{E}_F X^k$ of the ESD can be computed using free probability. However, there seem to be no general rules for calculating more general moments such as $\mathbb{E}_F \log(X)$ or $\mathbb{P}_F(X \le c)$. In contrast, they can be computed conveniently with our method.

Corollary 5.1. *Let $\gamma < 1$, and $h : \mathbb{R} \to \mathbb{R}$ be bounded on compact intervals and Riemann-integrable on compact intervals. Then the integral of h computed against*

the density approximation $\hat{f}(\cdot, \varepsilon)$ produced by SPECTRODE *converges to the moment of h under the limit ESD F:*

$$\lim_{\varepsilon \to 0} \int h(x)\hat{f}(x, \varepsilon)dx = \int h(x)f(x)dx.$$

The same holds for $\gamma > 1$ if we account for the point mass at $x = 0$.

Proof. Let M be an arbitrary upper bound on the support of F. Then for sufficiently small ε, f is zero outside $[0, M]$, and so is $\hat{f}$ by Theorem 4.2. This shows that the integrals in the theorem are well-defined. Further,

$$\left| \int h(x)(\hat{f}(x, \varepsilon) - f(x))dx \right| \leq \int_0^M |h(x)|dx \sup_{x \in [0,M]} |\hat{f}(x, \varepsilon) - f(x)| \to 0.$$

The convergence to zero follows because the first term is bounded by the assumptions on h, while the second term tends to zero by Theorem 4.2. The case $\gamma > 1$ is analogous. □

It is also possible to prove the convergence of Riemann sums $\sum_i h(x_i)\ \hat{f}(x_i, \varepsilon)\ \Delta(x_i)$, but this will not be pursued here.

As an example, in Table 3 we show the results of computing three moments of the standard MP distribution, with $\gamma = 1/2$. The three functions are $h(x) = x$, $\log(x)$, and $\log^2(x)$. The true value of the expectation for x is 1, for $\log(x)$ is $\log(2) - 1$ (see e.g. [37]), while for $\log^2(x)$ it is unknown. The numerical values computed with SPECTRODE, for a precision parameter $\varepsilon = 10^{-8}$, are very accurate in the known cases. In addition, SPECTRODE also approximates the integral of $\log^2(x)$, which is not known in closed form.

5.2. *Contour integrals of the Stieltjes transform of the ESD*

SPECTRODE can be adapted to compute contour integrals involving the Stieltjes transform of the limit ESD. Such integrals appear in Bai and Silverstein's central limit theorem for linear spectral statistics of the covariance matrix [3].

Let Γ be a smooth contour in the complex plane that does not intersect the support of the limit ESD F. Let $c : [0, 1] \to \mathbb{C}$ be a parametrization of the contour, and $v(z)$ be the Stieltjes transform of F. Note that $v(z)$ can be defined by the same formulas (2.1)–(2.2) for all z outside the support of F, and in particular for all $z \in \Gamma$.

Table 3. Moments of standard MP law with $\gamma = 1/2$.

Function	True value	Numerical value	Accuracy
x	1	1	2.3308e−06
$\log(x)$	−0.30685	−0.30684	1.4483e−05
$\log^2(x)$	Unknown	0.81724	

Suppose that for a smooth function G, we want to calculate the following integral over the clockwise oriented contour Γ:

$$\mathcal{I} = \oint_{\Gamma} G(z, v(z))dz. \tag{5.1}$$

An example is the mean in the CLT for linear spectral statistics $\sum_i g(\lambda_i)$ of $\widehat{\Sigma}$ (for a smooth function g), which involves the formula (see [3]):

$$\mathcal{J}(g, H, \gamma) = -\frac{1}{2\pi i}\oint_{\Gamma} g(z)\frac{\gamma\int \frac{v(z)^3 t^2 dH(t)}{(1+tv(z))^3}}{[1-\gamma\int \frac{v(z)^2 t^2 dH(t)}{(1+tv(z))^2}]^2}dz. \tag{5.2}$$

This is a special case of our general problem (5.1). To compute the general integral $\mathcal{I}$, we perform the change of variables $z = c(t)$, and rewrite $\mathcal{I}$ in the form $\mathcal{I} = \int_{t\in[0,1]} G\{c(t), v(c(t))\}c'(t)dt$. We assume $G(\cdot)$, $c(t)$ and $c'(t)$ are conveniently computable. Then the key problem in approximating this integral is obtaining $v(c(t))$ for the entire range $t \in [0, 1]$. This is where the ideas used in SPECTRODE will help.

We will obtain $h(t) := v(c(t))$ by exhibiting an ODE for it. Specifically, by using the chain rule $h'(t) = v'(c(t))c'(t)$, and recalling the ODE (2.6), which states $v'(z) = \mathcal{F}(v(z))$, we get the following new ODE for h: $h'(t) = \mathcal{F}(h)c'(t)$.

The starting point $h(0)$ (i.e. $v(c(0))$) can again be found using FPA, provided the starting point of the curve, $c(0)$ has non-zero imaginary part. Indeed, this follows from the general properties of FPA if the imaginary part of $c(0)$ is positive. Moreover, the Stieltjes transform enjoys $\bar{v}(z) = v(\bar{z})$ ($\bar{z}$ denotes complex conjugation), so FPA also converges for z with negative imaginary part. Now, if the curve Γ lies entirely on the real line, then a starting point be obtained using the function (2.7), which becomes the explicit inverse of the map $c \to v(c)$, as shown in [33].

The new ODE can be integrated numerically. The obtained values for $\hat{h}$ can be used to approximate the contour integral $\mathcal{I}$ using standard quadrature methods.

5.2.1. *Example*

In Table 4 we show an example. Consider the sample covariance matrix $\widehat{\mathbf{\Sigma}} = n^{-1}\mathbf{X}^\top\mathbf{X}$, where x_{ij} are real random variables with $\mathbb{E}x_{ij} = 0$, $\mathbb{E}x_{ij}^2 = 1$, and $\mathbb{E}x_{ij}^4 = 3$. Let λ_i be the eigenvalues of the sample covariance matrix, and let F_γ be the standard MP law with index γ. Consider a sequence of such problems with $n, p \to \infty$, $\gamma_p := p/n \to \gamma$. Let F_p be discrete spectral distribution of $\widehat{\mathbf{\Sigma}}$. For a

Table 4. Mean of normalized LSS for identity covariance.

Linear statistic	True value	Numerical value	Accuracy
x	0	−7.1292e−12	7.1292e−12
$\log(x)$	−0.34657	−0.34655	2.2214e−05
$\log^2(x)$	Unknown	1.2111	

function g that is analytic in a neighborhood of the support of F_γ, let $X_p(g)$ be the linear spectral statistic $X_p(g) := p\{F_p(g) - F_{\gamma_p}(g)\} = \sum_{i=1}^p g(\lambda_i) - p\mathbb{E}_{F_{\gamma_p}}[g(\lambda)]$.

Then Bai and Silverstein [3] proved that $X_p(g)$ is asymptotically normal with mean given in (5.2), with $H = \delta_1$ and Γ an arbitrary contour enclosing the support of the ESD. It is known (see for instance [37]) that:

$$\mathcal{J}(x, \delta_1, \gamma) = 0, \quad \mathcal{J}(\log x, \delta_1, \gamma) = \frac{1}{2}\log(1-\gamma).$$

We use our method, outlined above, to compute the integral (5.2). We take $\gamma = 1/2$ and compare against the closed form solutions for $g(x) = x$ and $g(x) = \log(x)$. We also compute the integral for $g(x) = \log^2(x)$, for which no closed form solution appears to be known. The results, displayed in Table 4, show the excellent performance of our method.

In this experiment, we used the circle contour $c(t) = a/2 + a/2 \cdot e^{2\pi i t}$, with $a = 1.1 \cdot (1 + \gamma^{1/2})^2$. This contour encloses the support of the ESD. The starting point of the ODE, $v(c(0)) = v(a)$, was found using the function $z(v)$ (2.7). Using Brent's method we found the unique solution to the equation $z(v) = a$ on the interval $v \in (-1, 0)$ such that $z'(v) > 0$.

6. Related Work

Problems related to computing the ESD of covariance matrices have been discussed in several important works. Here we examine the strengths and weaknesses of related and alternative methods.

6.1. *Monte Carlo*

Monte Carlo (MC) simulation can be used to approximate the eigenvalue density of large covariance matrices via a smoothed empirical histogram of eigenvalues. It was proved in [19] that this method consistently estimates the ESD. However, we show in a simple simulation that MC is prohibitively slow when more than three digits of accuracy are required.

6.1.1. *Experiment setup and parameters*

We use the `MP` test problem when the covariance matrix $\mathbf{\Sigma}_p = I_p$ and $H = \delta_1$. For a pair n, p we sample random matrices $\mathbf{X}$ with iid standard normal entries, and compute the eigenvalues of $\widehat{\mathbf{\Sigma}} = \mathbf{X}^\top\mathbf{X}/n$. We fit a kernel density estimate to the eigenvalues, using an Epanechnikov kernel with automatically chosen kernel width in MATLAB. The kernel density estimate is averaged over the n_{MC} independent MC trials to get a final estimate $\hat{f}_{\mathrm{MC}}(x_i)$ of the density. We use the following parameters: $n_{\mathrm{MC}} = 1000$, $\gamma = 1/2$, and p takes the values $10, 10^2, 10^3$.

We compare against the true limit f from Eq. (3.1) and report the error in the density from Eq. (3.3): $\Delta_{\mathrm{MC}}(x_i) = \log_{10}|\hat{f}_{\mathrm{MC}}(x_i) - f(x_i)|$.

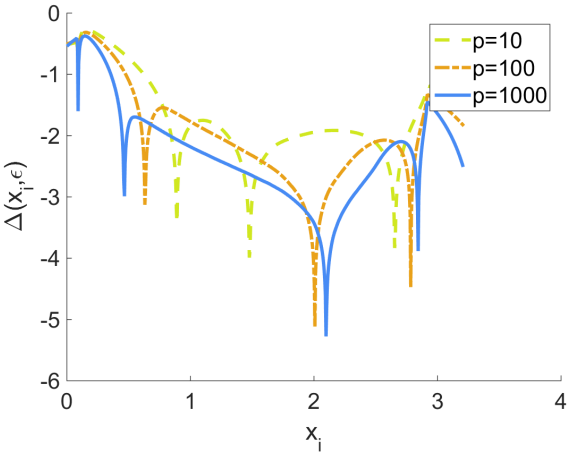


Fig. 5. Accuracy of the MC method, discussed in Sec. 6. We sample $n_{\mathrm{MC}} = 1000$ independent random matrices with iid Gaussian entries and aspect ratio $\gamma = 1/2$. We fit a kernel density estimator to the histogram of eigenvalues of each sample, and average over independent samples. We display the pointwise error for p in the range 10, $10^2, 10^3$.

6.1.2. *Results*

In the results in Fig. 5, we see that number of correct significant digits is two or three. We get only 1/2 extra digit when we move from $p = 100$ to $p = 1000$. The experiment for $p = 1000$ takes about 10 min on the same hardware described in Sec. 3.3. Computing the eigenvalues using the SVD takes $\Theta(p^3)$ steps. For 10 times larger $p = 10^4$, such an experiment would take about 10^5 min, or 1600 h, which is prohibitively slow.

Based on this experiment, we conclude that the MC simulation for computing the ESD — if implemented in a straightforward way — is not suitable for getting more than three digits of accuracy.

6.2. *Method of successive approximation*

While Marchenko and Pastur [23] do not place emphasis on numerically solving their equation, they mention that its solution can be found by the method of successive approximation (SA). SA was also proposed by Girko [14] for several variants of the Marchenko–Pastur equation. We will argue that SA is not an efficient method for computing the ESD.

Marchenko and Pastur in [23] consider a more general model than this paper, allowing for an additive term $\mathbf{Y} = \mathbf{A} + n^{-1}\mathbf{X}^\top\mathbf{T}\mathbf{X}$. They denote by $\tau(\xi) : [0, 1] \to \mathbb{R}$ the quantile function of the PSD H, which is the limit of the spectrum of $\mathbf{T}$; and $c = \lim p/n$, which was called γ in our paper. They call $m_0(z)$ the Stieltjes transform of the limit spectrum of $\mathbf{A}$, writing the equation for the Stieltjes transform of the

ESD of $\mathbf{Y}$ as (see Eq. (1.13) in the English translation of [23]):

$$u(z,t) = m_0\left(z - c\int_0^t \frac{\tau(\xi)d\xi}{1+\tau(\xi)u(z,t)}\right). \tag{6.1}$$

The unknown function is $u(z,t)$. With the initial condition $u(z,0) = m_H(z)$, there is a unique solution $u(z,t)$ analytic in $z \in \mathbb{C}^+$ and continuous in $t \in [0,1]$ of (6.1). Then, $u(z,1)$ is the Stieltjes transform $m_F(z)$ of the limit ESD F.

In our special case $\mathbf{A} = \mathbf{0}$, so $m_0(z) = -1/z$ and the equation simplifies to $u(z,t) = -1/(z - c\int_0^t \frac{\tau(\xi)d\xi}{1+\tau(\xi)u(z,t)})$. Denoting $v(z) := u(z,1)$, and switching from the integral over $t \in [0,1]$ to the integral against dH, we see that this reduces to the Silverstein equation (2.3).

The method of SA starts with an arbitrary function $u_0(z,t)$ obeying the smoothness and continuity conditions above. It defines the sequence of functions $u_n(z,t)$ inductively for $n \in \mathbb{N}$ by:

$$u_{n+1}(z,t) = m_0\left(z - c\int_0^t \frac{\tau(\xi)d\xi}{1+\tau(\xi)u_n(z,t)}\right). \tag{6.2}$$

For this method it is crucial that one maintains bivariate functions $u_n(z,t)$. The definition of u_{n+1} depends on the integral of u_n against t, even if we are interested only in the values $u_{n+1}(z,1)$, for $t = 1$. Therefore, the method of SA relies on an additional time dimension. However, this extra dimension seems costly in the special case when $\mathbf{A} = \mathbf{0}$. There are more efficient methods, such as FPA and SPECTRODE, that do not rely on additional dimensions.

6.3. *Fixed-point algorithm*

FPA appears to be the standard approach that most researchers use to compute the ESD of sample covariance matrices in the Marchenko–Pastur asymptotic regime. Versions of FPA have been developed in many different areas, including free probability, wireless communications, signal processing, and mathematical statistics. The existing techniques for analyzing it fall into the following categories: (1) complex analytic methods; (2) contraction by deterministic equivalents of random matrices; and (3) interference functions.

Belinschi and Bercovici in [5] explain subordination results in free probability, with the Marchenko–Pastur–Silverstein equation as a special case. They employ a complex analytic approach, which involves the study of Denjoy–Wolff points. As a special case of their results, it follows that FPA converges. This was not explicitly stated by the authors, but it is indeed an immediate consequence of their results.

In many random matrix models, FPA is often invoked implicitly, to show the uniqueness of the solution to the fixed-point equation governing the spectrum. For instance, this is the approach in the analysis of deterministic equivalents for random matrices by Hachem, Loubaton and Najim [16]. Here the authors show uniqueness of solutions to their equations by exhibiting bounds on the size of successive iterates, which is equivalent to the convergence of FPA in that context.

In the wireless communications literature, explicit fixed point algorithms have been developed for computing the ESDs of very general random matrix models in several papers: [10, 9, 11]. Couillet, Debbah and Silverstein in [10] give a fixed-point algorithm for a random matrix model that is a sum of arbitrary covariance matrices. They prove its convergence by showing that the iteration is a contraction for complex points z with large $\mathrm{Imag}(z)$, similarly to the earlier work [16]. Then the convergence extends to all points outside the support of the ESD using Vitali's theorem.

In contrast, the authors in [11] take a different approach to proving the convergence of FPA. They show that for negative arguments $z \in (-\infty, 0]$, their equations are fixed points of interference functions, and they rely on general convergence results of such functions [38]. Again, Vitali's theorem extends the convergence to other points. In the model of [11] the convergence of FPA cannot be established for all complex numbers; which is a counterexample showing that FPA is not expected to converge in all circumstances. However, the interference function approach for proving convergence of FPA is powerful and general, see for instance [12, 24].

FPA has appeared in other papers as well. Yao in [36] has a fixed-point algorithm for a time series problem, but without convergence guarantees. Hendrikse and his coauthors [18] discover the fixed point algorithm for the limit ESD of sample covariance matrices, and claim to prove convergence for the argument z with sufficiently large imaginary part $\mathrm{Imag}(z) > c$, using elementary arguments. However, their arguments appear to be incomplete; as they do not show that the iterates z_t remain in the region $\mathrm{Imag}(z) > c$ for all t.

In a free probability context, Belinschi, Mai and Speicher in the paper [6] generalize the FPA to compute the limit ESD of arbitrary polynomials $P(X_{n1}, \ldots, X_{nk})$, given the limit ESD of random matrices $X_{n1}, \ldots, X_{nk}$. This important paper has the advantage of generality. As we showed, however, for our problem FPA can unfortunately be slow for high-precision computations.

6.4. *Other methods*

Special cases of limit ESDs have been computed in a case-by-case fashion. In the analysis of wireless networks, [25] develops a computational procedure for a special case that involves a fourth-order polynomial equation.

Further, Nadakuditi and Edelman in the influential work [27] developed a polynomial method as a general framework for computing limit spectra of ensembles whose Stieltjes transform is algebraic. As noted by the authors (see [27, Sec. 7]), these methods in general do not lead to an automated way to compute the limit density. Indeed, this is presented as an open problem in [27]. SPECTRODE addresses a narrower setting and shows that the ESD can be computed reliably in that setting.

Olver and Nadakuditi in [29] present an interesting approach for calculating the additive, multiplicative and compressive convolution in free probability. The map

we consider, from population to sample spectra $H \to F$, corresponds to free multiplicative convolution with the identity Marchenko–Pastur distribution. However, their method is not generally applicable to our problem. It requires that the support of the LSD F be precisely one compact interval, because it relies on specific series expansions (see [29, Sec. 4]). In our case this is often not the case. We see SPECTRODE as complementary to their method, for the case of multiple intervals in the support.

7. Software

A software companion for this paper is available at https://github.com/dobriban/eigenedge. It contains implementations of

(1) methods to compute the sample spectrum: SPECTRODE and the fixed-point method,
(2) methods to compute arbitrary moments and quantiles of the ESD,
(3) MATLAB scripts to reproduce all computational results of this paper,
(4) detailed documentation with examples.

The package is user-friendly. Once the appropriate environment is installed, the ESD of a uniform mixture of four-point masses at `t = [1; 2; 3; 4]`, and with aspect ratio `gamma = 1/2` requires three lines of code:

```
 t = [1; 2; 3; 4];
gamma = 1/2;
[grid, density] = spectrode(t, gamma); %compute limit ESD
```


Acknowledgments

We are grateful to David L. Donoho for posing the problem and reviewing the manuscript. We are obliged to Romain Couillet for numerous references and comments, especially on FPA. We thank Iain M. Johnstone, Matthew McKay, Art B. Owen and Jack W. Silverstein for helpful comments; and Tobias Mai for discussions about the paper [6]. Financial support has been provided by NSF DMS 1418362.

We thank the referee for many helpful comments which improved the paper, and for suggesting to solve the equations $z''(v) = 0$ and $z'(v) = 0$ on each interval $(-1/t_i, -1/t_{i+1})$ separately.

References

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (Wiley, New York, 2003).
[2] Z. Bai, D. Jiang, J.-F. Yao and S. Zheng, Corrections to LRT on large-dimensional covariance matrix by RMT, *Ann. Statist.* **37**(6B) (2009) 3822–3840.
[3] Z. Bai and J. W. Silverstein, CLT for linear spectral statistics of large-dimensional sample covariance matrices, *Ann. Probab.* **32**(1A) (2004) 553–605.

[4] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Springer Series in Statistics (Springer, New York, 2009).
[5] S. Belinschi and H. Bercovici, A new approach to subordination results in free probability, *J. Anal. Math.* **101**(1) (2007) 357–365.
[6] S. Belinschi, T. Mai and R. Speicher, Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem, preprint (2013), arXiv:1303.3196.
[7] F. Benaych-Georges and R. R. Nadakuditi, The singular values and vectors of low rank perturbations of large rectangular random matrices, *J. Multivariate Anal.* **111** (2012) 120–135.
[8] R. P. Brent, An algorithm with guaranteed convergence for finding a zero of a function, *Comput. J.* **14**(4) (1971) 422–425.
[9] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications* (Cambridge University Press, Cambridge, MA, 2011).
[10] R. Couillet, M. Debbah and J. W. Silverstein, A deterministic equivalent for the analysis of correlated MIMO multiple access channels, *IEEE Trans. Inform. Theory* **57**(6) (2011) 3493–3514.
[11] R. Couillet, J. Hoydis and M. Debbah, Random beamforming over quasi-static and fading channels: A deterministic equivalent approach, *IEEE Trans. Inform. Theory* **58**(10) (2012) 6392–6425.
[12] R. Couillet, F. Pascal and J. W. Silverstein, The random matrix regime of Maronnas M-estimator with elliptically distributed samples, *J. Multivariate Anal.* **139** (2015) 56–78.
[13] M. Gavish and D. Donoho, The optimal hard threshold for singular values is $4/\sqrt{3}$, *IEEE Trans. Inform. Theory* **60**(8) (2014) 5040–5053, doi:10.1109/TIT.2014.2323359.
[14] V. L. Girko, *Theory of Stochastic Canonical Equations* (Springer, New York, 2001).
[15] W. Hachem, A. Hardy and J. Najim, Large complex correlated Wishart matrices: Fluctuations and asymptotic independence at the edges, preprint (2014), arXiv:1409.7548.
[16] W. Hachem, P. Loubaton and J. Najim, Deterministic equivalents for certain functionals of large random matrices, *Ann. Appl. Probab.* **17**(3) (2007) 875–930, doi:10.1214/105051606000000925.
[17] E. Hairer, S. P. Norsett and G. Wanner, *Solving Ordinary Differential Equations I: Nonstiff Problems* (Springer, New York, 2009).
[18] A. Hendrikse, R. Veldhuis and L. Spreeuwers, Smooth eigenvalue correction, *EURASIP J. Adv. Signal Process.* **2013**(1) (2013) 1–16.
[19] B.-Y. Jing, G. Pan, Q.-M. Shao and W. Zhou, Nonparametric estimate of spectral density functions of sample covariance matrices: A first step, *Ann. Statist.* **38**(6) (2010) 3724–3750.
[20] I. M. Johnstone, High dimensional statistical inference and random matrices, in *Proc. Int. Congr. Mathematicians*, Vol. 1 (European Mathematical Society, Zürich, 2007), pp. 307–333, doi:10.4171/022-1/13.
[21] O. Ledoit and M. Wolf, Optimal estimation of a large-dimensional covariance matrix under Stein's loss, Working Paper 122, Department of Economics, University of Zurich (2013).
[22] O. Ledoit and M. Wolf, Spectrum estimation: A unified framework for covariance matrix estimation and {PCA} in large dimensions, *J. Multivariate Anal.* **139** (2015) 360–384.
[23] V. A. Marchenko and L. A. Pastur, Distribution of eigenvalues for some sets of random matrices, *Mat. Sb.* **114**(4) (1967) 507–536.

[24] D. Morales-Jimenez, R. Couillet and M. R. McKay, Large dimensional analysis of robust M-estimators of covariance with outliers, preprint (2015), arXiv:1503.01245.
[25] V. Morgenshtern and H. Bolcskei, Crystallization in large wireless networks, *IEEE Trans. Inform. Theory* **53**(10) (2007) 3319–3349, doi:10.1109/TIT.2007.904789.
[26] R. R. Nadakuditi, OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage, *IEEE Trans. Inform. Theory* **60**(5) (2014) 3002–3018, doi:10.1109/TIT.2014.2311661.
[27] R. R. Nadakuditi and A. Edelman, The polynomial method for random matrices, *Found. Comput. Math.* **8**(6) (2008) 649–702.
[28] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability* (Cambridge University Press, Cambridge, 2006).
[29] S. Olver and R. R. Nadakuditi, Numerical computation of convolutions in free probability theory, preprint (2012), arXiv:1203.1958.
[30] V. I. Serdobolskii, *Multiparametric Statistics* (Elsevier, Amsterdam, 2007).
[31] R. J. Serfling, *Approximation Theorems of Mathematical Statistics* (John Wiley & Sons, New York, 2009).
[32] J. W. Silverstein, Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices, *J. Multivariate Anal.* **55**(2) (1995) 331–339.
[33] J. W. Silverstein and S.-I. Choi, Analysis of the limiting spectral distribution of large dimensional random matrices, *J. Multivariate Anal.* **54**(2) (1995) 295–309.
[34] J. W. Silverstein and P. L. Combettes, Signal detection via spectral theory of large dimensional random matrices, *IEEE Trans. Signal Process.* **40**(8) (1992) 2100–2105.
[35] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, Vol. 1 (Now Publishers, 2004), pp. 1–182.
[36] J. Yao, A note on a Marčenko–Pastur type theorem for time series, *Statist. Probab. Lett.* **82**(1) (2012) 22–28.
[37] J. Yao, Z. Bai and S. Zheng, *Large Sample Covariance Matrices and High-Dimensional Data Analysis* (Cambridge University Press, 2015).
[38] R. D. Yates, A framework for uplink power control in cellular radio systems, *IEEE J. Sel. Areas Commun.* **13**(7) (1995) 1341–1347.