Applied and Computational Harmonic Analysis 79 (2025) 101802

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

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

Letter to the Editor

Sparse free deconvolution under unknown noise level via eigenmatrix

Lexing Ying [1]

Department of Mathematics, Stanford University, Stanford, CA 94305, United States of America

ARTICLE INFO

Communicated by Radu Balan

MSC:
46L54
62H12
65R32

Keywords:
Free deconvolution
Sparse spectral measure
Eigenmatrix

ABSTRACT

This note considers the spectral estimation problems of sparse spectral measures under unknown noise levels. The main technical tool is the eigenmatrix method for solving unstructured sparse recovery problems. When the noise level is determined, the free deconvolution reduces the problem to an unstructured sparse recovery problem to which the eigenmatrix method can be applied. To determine the unknown noise level, we propose an optimization problem based on the singular values of an intermediate matrix of the eigenmatrix method. Numerical results are provided for both the additive and multiplicative free deconvolutions.

1. Introduction

This note considers free deconvolution problems of sparse spectral measures under unknown noise levels. In the additive setting, assume that A is a real $N \times N$ symmetric matrix with an unknown sparse spectral measure μ_A. Let B be a Wigner matrix with an unknown noise level σ, i.e., the off-diagonal and diagonal entries are Gaussian with variance $\frac{\sigma^2}{N}$ and $\frac{2\sigma^2}{N}$, respectively. The spectral measure μ_B of B follows the semicircle law with parameter σ in the large N limit. Given the spectral measure μ_C of $C = A + B$, the task is to recover σ and μ_A.

In the multiplicative setting, assume that A is a real $N \times N$ symmetric positive definite matrix with an unknown sparse spectral measure μ_A. Let B be a Wishart matrix with an unknown dimension-to-sample size ratio q, i.e., B is statistically equivalent to $\frac{1}{T}\sum_{t=1}^{T} X_t X_t^{\mathsf{T}}$ with $q = N/T$ with $X_t \sim \mathcal{N}(0, I_{N\times N})$. The spectral measure μ_B of B follows the Marchenko-Pastur law with parameter q in the large N limit. Given the spectral measure μ_C of $C = \sqrt{A}B\sqrt{A}$, the task is to recover q and μ_A.

These two problems have many applications in statistics and data science. The first one, sometimes referred to as the deformed Wigner model, comes up when the data matrix A is polluted by entrywise independent noise with unknown variance. The second problem appears in the estimation of the covariance matrix A, where q is not known due to either the lack of information about T or the dependence between the samples $\{X_t\}$.

1.1. Related work

Spectral estimation has been an active field of research over the past two decades. For covariance matrix estimation (the multiplicative setting), many methods have been proposed over the years, including linear shrinkage [7], methods based on optimizing

E-mail address: lexing@stanford.edu.

[1] This work is partially supported by NSF grant DMS-2208163. The author thanks Xiucai Ding for helpful discussions.

https://doi.org/10.1016/j.acha.2025.101802
Received 20 January 2025; Received in revised form 21 July 2025; Accepted 8 August 2025

Available online 14 August 2025

over the Marcenko-Pastur equation directly [3], the highly successful nonlinear shrinkage methods [6,8–10], and the moment-based method [5]. For the deformed Wigner model (the additive setting), there has also been a significant body of work, mostly for nonlinear shrinkage methods, including [2,4,11].

There is little work on using the free deconvolution directly. In [1], the subordination method reduces the task to a classical deconvolution problem, which is then solved via Tikhonov regularization with carefully chosen regularization parameters.

In most of the work mentioned above, the noise level σ or q is given. The case with an unknown noise level is much less explored.

1.2. Contributions

In this note, we consider the more challenging case where the noise level σ or q is unknown. Without prior information on the spectral measure μ_A of A, the problems are ill-defined. In order to work with a well-defined setting, we assume that μ_A is sparsely supported.

Assume for a moment that the noise level is already determined. By using the R-transform and S-transform [15,16] from free probability [12,13], the spectral estimation problem can be cast into a classical inverse problem of the Cauchy integral with observations at unstructured locations. To solve this inverse problem, we leverage the recently proposed eigenmatrix method [17,18] for solving the unstructured sparse recovery problem. The word "unstructured" refers to the fact that the sample locations can be arbitrary, as will be the case for the deconvolution problems considered in this context.

To determine the unknown noise level, we propose a novel optimization problem in which the loss function is based on the singular value of an intermediate matrix used in the eigenmatrix method. Optimizing this objective function gives an accurate estimate of the noise level.

The rest of the note is organized as follows. Section 2 reviews some basic facts of the free probability theory and the eigenmatrix method. Section 3 describes the additive case. Section 4 considers the multiplicative case. Section 5 concludes with a discussion for future work.

2. Background

2.1. Facts from free probability theory

In the study of free probability [12,13,15,16], a non-commutative probability space $(\mathcal{A}, \tau)$ consists of a tracial $*$-algebra $\mathcal{A}$ of non-commutative random variables and a trace operator $\tau : \mathcal{A} \to \mathbb{C}$ that maps unity to itself. Two random variables A and B are free if for any set of traceless polynomials $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$ the following holds

$$\tau(p_1(A)q_1(B)p_2(A)q_2(B)\ldots p_n(A)q_n(B)) = 0. \tag{1}$$

For any A, the Stieltjes (or Cauchy) transform, defined as $g_A(z) = \tau((z - A)^{-1})$, provides a correspondence between the complex planes z and g. Its inverse map from g to z is well-defined for sufficiently small values of g and is denoted by $z_A(g)$. The related R-transform is defined by

$$r_A(g) = z_A(g) - g^{-1}.$$

One can further introduce the complex domains $t = zg - 1$ and $s = \frac{t+1}{tz}$. The map from t to s, well-defined for sufficiently small values of t, is the S-transform, denoted by $s_A(t)$. Suppose that A and B are free with respect to each other.

- For $C = A + B$, $r_C(g) = r_A(g) + r_B(g)$ [15].
- For $C = AB$, $s_C(t) = s_A(t)s_B(t)$ [16].

To make the connection with large matrices, consider A and B to be large $N \times N$ matrices and $\tau(A) = \mathsf{E}\frac{1}{N}\mathsf{Tr}(A)$. The key result is that two large symmetric matrix ensembles whose eigenbases are uniformly random with respect to each other are asymptotically free, i.e., the left-hand side of (1) approaches zero as N goes to infinity. In fact, it is expected that, due to self-averaging, this holds for most realizations without taking the expectation. For example, the Wigner and Wishart models both have random orthonormal eigenbases. Therefore, if B comes from either ensemble, it is free with respect to any deterministic matrix A. In the matrix case, since the trace operator $\tau(A) = \mathsf{E}\frac{1}{N}\mathsf{Tr}(A)$ depends only on the spectrum measure of the matrix, we can write the R-transform and S-transform relations in the large N limit in terms of the spectrum measures

- For $C = A + B$, $r_{\mu_C}(g) = r_{\mu_A}(g) + r_{\mu_B}(g)$.
- For $C = AB$, $s_{\mu_C}(t) = s_{\mu_A}(t)s_{\mu_B}(t)$.

For a Wigner matrix B with parameter σ, the spectrum density follows the semi-circle law in the large N limit, and the R-transform is given by

$$r_{\mu_B}(g) = \sigma^2 g.$$

Similarly, for a Wishart matrix B with parameter q, the spectrum density follows the Marcenko-Pastur law in the large N limit, and the S-transform is equal to

$$s_{\mu_B}(t) = \frac{1}{1+qt}.$$

2.2. Eigenmatrix

Here we provide a short review of the eigenmatrix method for unstructured sparse recovery problems. Let X be the parameter space. $G(z,x)$ is a kernel function defined for $x \in X$ at sample z, and is assumed to be analytic in x. In this note, X is an interval of $\mathbb{R}$, $z \in \mathbb{C}$, and $G(z,x) = \frac{1}{z-x}$. Suppose that

$$f(x) = \sum_{k=1}^{n} w_k \delta(x - x_k) \tag{2}$$

is an unknown sparse signal, where $\{x_k\}_{1\le k\le n}$ are the spike locations and $\{w_k\}_{1\le k\le n}$ are the spike weights. The observed quantity is defined via the function

$$u(z) := \int_X G(z,x) f(x) dx = \sum_{k=1}^{n} G(z,x_k) w_k. \tag{3}$$

Let $\{z_j\}_{1\le j\le n_z}$ be a set of n_z unstructured samples. Suppose that we are given the noisy observations $u_j \approx u(z_j)$. The task is to recover the spikes $\{x_k\}$ and weights $\{w_k\}$ from $\{u_j\}$.

Define for each x the column vector

$$\mathbf{b}_x := \left[G(z_j,x)\right]_{1\le j\le n_z}$$

in $\mathbb{C}^{n_z}$. Notice that $\mathbf{b}_x$ is analytic in terms of x.

The first step is to construct a matrix $M \in \mathbb{C}^{n_z\times n_z}$ such that $M\mathbf{b}_x \approx x\mathbf{b}_x$ for $x \in X$, i.e., M is the matrix with $(x,\mathbf{b}_x)$ as approximate eigenpairs for $x \in X$. Numerically, it is more robust to use the normalized vector $\hat{\mathbf{b}}_x = \mathbf{b}_x/\|\mathbf{b}_x\|$ since the norm of $\mathbf{b}_x$ can vary significantly depending on x. The condition then becomes

$$M\hat{\mathbf{b}}_x \approx x\hat{\mathbf{b}}_x, \quad x \in X.$$

To construct M numerically, we choose a Chebyshev grid $\{c_t\}_{1\le t\le n_c}$ of size n_c on the interval X, where n_c is sufficiently large yet the vectors $\{\hat{\mathbf{b}}_{c_t}\}$ are numerically linearly independent. We then enforce this condition on this grid, i.e.,

$$M\hat{\mathbf{b}}_{c_t} \approx c_t\hat{\mathbf{b}}_{c_t}.$$

By defining the $n_z \times n_c$ matrix

$$\hat{B} = \left[\hat{\mathbf{b}}_{c_1} \quad \dots \quad \hat{\mathbf{b}}_{c_{n_c}}\right]$$

and the $n_c \times n_c$ diagonal matrix $\Lambda = \mathrm{diag}(c_t)$, the previous condition can be written in a matrix form as

$$M\hat{B} \approx \hat{B}\Lambda.$$

Because the columns of $\hat{B}$ are numerically linearly independent, we set the eigenmatrix as

$$M := \hat{B}\Lambda\hat{B}^{+}, \tag{4}$$

where the pseudoinverse $\hat{B}^{+}$ is computed by thresholding the singular values of $\hat{B}$. In practice, the thresholding value is chosen so that the norm of M is bounded by a small constant.

Once M is ready, the rest follows, for example, the ESPRIT algorithm [14]. Define the vector

$$\mathbf{u} = \left[u_j\right]_{1\le j\le n_z}$$

in $\mathbb{C}^{n_z}$, where u_j are the noisy observations. Notice that $\mathbf{u} \approx \sum_k \mathbf{b}_{x_k} w_k$. Consider the matrix

$$T \equiv \left[\mathbf{u} \quad M\mathbf{u} \quad \dots \quad M^{n_l}\mathbf{u}\right] \tag{5}$$

with $n_l > n$, obtained from applying M repetitively to $\mathbf{u}$. Since $\mathbf{u} \approx \sum_k \mathbf{b}_{x_k} w_k$ and $M\mathbf{b}_x \approx x\mathbf{b}_x$,

$$T = \left[\mathbf{u} \quad M\mathbf{u} \quad \dots \quad M^{n_l}\mathbf{u}\right] \approx \left[\mathbf{b}_{x_1} \quad \dots \quad \mathbf{b}_{x_n}\right] \begin{bmatrix} w_1 & & \\ & \ddots & \\ & & w_n \end{bmatrix} \begin{bmatrix} 1 & x_1 & \dots & (x_1)^{n_l} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & (x_n)^{n_l} \end{bmatrix}.$$

Let USV^* be the rank-n truncated SVD of T. The matrix V^* then satisfies

$$V^* \approx P \begin{bmatrix} 1 & x_1 & \dots & (x_1)^{n_l} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & (x_n)^{n_l} \end{bmatrix},$$

where P is an unknown non-degenerate $n \times n$ matrix. Let Z_L and Z_H be the submatrices obtained by excluding the last column and the first column of V^*, respectively, i.e.,

$$Z_L \approx P \begin{bmatrix} 1 & \dots & (x_1)^{n_l-1} \\ \vdots & \ddots & \vdots \\ 1 & \dots & (x_n)^{n_l-1} \end{bmatrix}, \quad Z_H \approx P \begin{bmatrix} x_1 & \dots & (x_1)^{n_l} \\ \vdots & \ddots & \vdots \\ x_n & \dots & (x_n)^{n_l} \end{bmatrix}.$$

By forming $Z_H(Z_L)^+$ and noticing

$$Z_H(Z_L)^+ \approx P \begin{bmatrix} x_1 & & \\ & \ddots & \\ & & x_n \end{bmatrix} P^{-1},$$

one obtains the estimates for $\{x_k\}$ by computing the eigenvalues of $Z_H(Z_L)^+$.

With the estimates for $\{x_k\}$ available, the least square solution of

$$\min_{w_k} \sum_j \left| \sum_k G(s_j, x_k) w_k - u_j \right|^2$$

gives the estimators for $\{w_k\}$.

3. Additive deconvolution

To address the additive case, we leverage the R-transform [15]. Since $C = A + B$, in the large dimension limit

$$r_{\mu_C}(g) = r_{\mu_A}(g) + r_{\mu_B}(g).$$

3.1. Known σ

Assume for now that σ is determined. Then $r_{\mu_B}(g) = \sigma^2 g$ from the semicircle law. Due to its sparsity, $\mu_A = \sum_k \delta_{x_k} w_k$. The task is to recover x_k and w_k.

Given μ_C, we choose $\{z_j\}$ to be a set of points on an ellipsis around μ_C and compute

$$g_j = \int \frac{1}{z_j - x} d\mu_C(x),$$

i.e., $g_{\mu_C}(z_j) = g_j$ and $z_{\mu_C}(g_j) = z_j$.

From $r_{\mu_C}(g) = r_{\mu_A}(g) + r_{\mu_B}(g)$ and $r_{\mu_B}(g) = \sigma^2 g$, we have

$$r_{\mu_A}(g_j) = r_{\mu_C}(g_j) - \sigma^2 g_j \quad z_{\mu_A}(g_j) = z_{\mu_C}(g_j) - \sigma^2 g_j = z_j - \sigma^2 g_j.$$

Define $z'_j = z_j - \sigma^2 g_j$. Then $z'_j = z_{\mu_A}(g_j)$ and $g_{\mu_A}(z'_j) = g_j$, i.e., (z'_j, g_j) are samples for the Stieltjes transform of μ_A:

$$g_j = \int \frac{1}{z'_j - x} d\mu_A(x) = \sum_k \frac{1}{z'_j - x_k} w_k.$$

Since the locations $\{z'_j\}$ are not a priori controlled, recovering $\{x_k\}$ and $\{w_k\}$ is a sparse, unstructured recovery problem.

Next, we apply the eigenmatrix method. More specifically, set X to be the shortest interval that covers μ_C. Treat $\{z'_j\}$ as the samples and $\{g_j\}$ as the observed data. Define $\mathbf{b}_x = \left[\frac{1}{z'_j - x} \right]_j$. Choose a Chebyshev grid $\{c_t\}$ and construct M such that $M\mathbf{b}_{c_t} \approx c_t \mathbf{b}_{c_t}$. Then, define $\mathbf{u} = \left[g_j \right]_j$ and form the T matrix (5) to recover $\{x_k\}$ and $\{w_k\}$.

3.2. Unknown σ

Let us consider how to determine σ. For a value of σ, since the T matrix is a function of σ, we denote it by $T(\sigma)$. The main observation is the following: at the correct σ value, $T(\sigma)$ has a numerical rank equal to n, while for other values of σ, the rank is higher.

Consider an example where $\sigma = 0.75$ and μ_A has $n = 3$ spikes. Fig. 3.1 plots the singular values of $T(\sigma)$ in logarithmic scale as a function of σ, i.e., the i-th curve from the top stands for the i-th singular value. The left plot shows the large N limit. At $\sigma = 0.75$, starting from the 4th curve, the singular value drops numerically to zero. This plot indicates that the right σ and n can be easily detected in the large N limit.

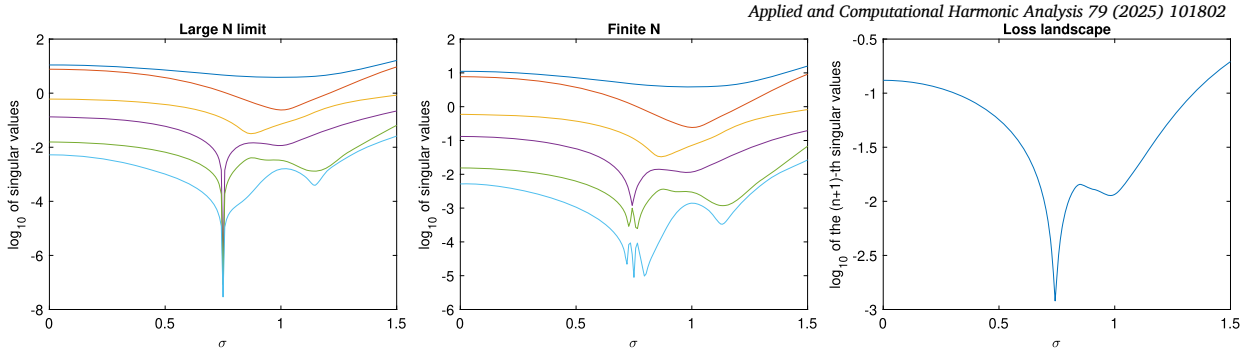

Fig. 3.1. Logarithm of singular values of $T(\sigma)$ as a function of σ. Left: large N limit. Middle: $N = 1024$. Right: loss landscape.

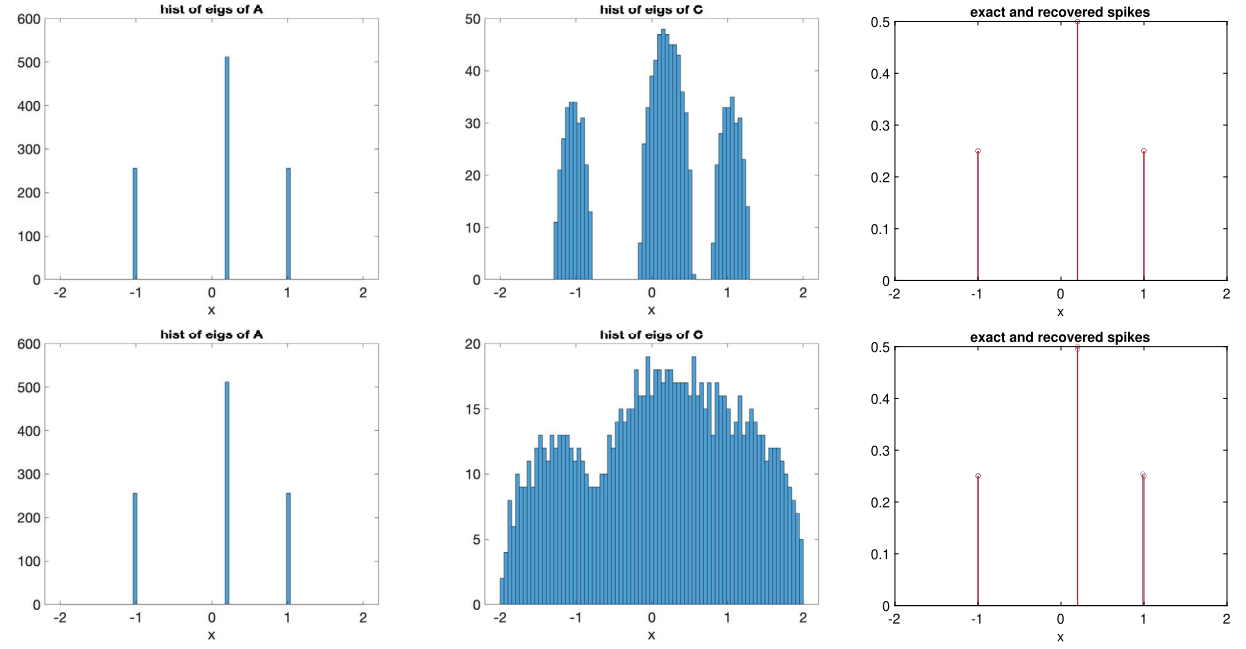

Fig. 3.2. Top: $\sigma = 0.25$. Bottom $\sigma = 0.75$. In each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

The middle plot shows the situation at $N = 1024$. The overall trends remain the same. However, due to the finiteness of N, it is harder to find an automatic procedure to identify σ and n together. In what follows, we assume n is given. Here, the $(n+1)$-th (i.e., 4th) singular value curve clearly shows a global minimum at the right σ.

To automate this process, we formulate it as a minimization problem

$$\hat{\sigma} = \mathrm{argmin}_{\sigma\in\Omega} \log(s_{n+1}(T(\sigma))),$$

where $s_{n+1}(\cdot)$ refers to the $(n+1)$-th singular value and Ω is the region for which the eigenmatrix method produces positive weights $\{w_k\}$. The right plot shows that this loss landscape can be quite non-convex, which can cause local optimization to get stuck. Since this is a one-dimensional optimization problem, we adopt a two-step procedure. The first step performs a grid search on a coarse grid, providing a good initial guess for σ_{init}. Starting from σ_{init}, the second step performs a local search to identify the optimal $\hat{\sigma}$. Finally, given $\hat{\sigma}$, we can compute the (z'_j, g_j) pairs of μ_A as in Section 3.1 to recover $\{x_k\}$ and $\{w_k\}$ via eigenmatrix.

Below, we present a few numerical examples.

Example 3.1. This example considers μ_A with three spikes and studies the performance for different values of σ. Here, $N = 1024$ and $\mu_A = \frac{1}{4}\delta_{-1} + \frac{1}{2}\delta_{0.2} + \frac{1}{4}\delta_1$. σ takes two values 0.25 and 0.75.

Fig. 3.2 summarizes the results. The top and bottom rows are for $\sigma = 0.25$ and $\sigma = 0.75$, respectively. Within each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right). Both reconstructions are very accurate.

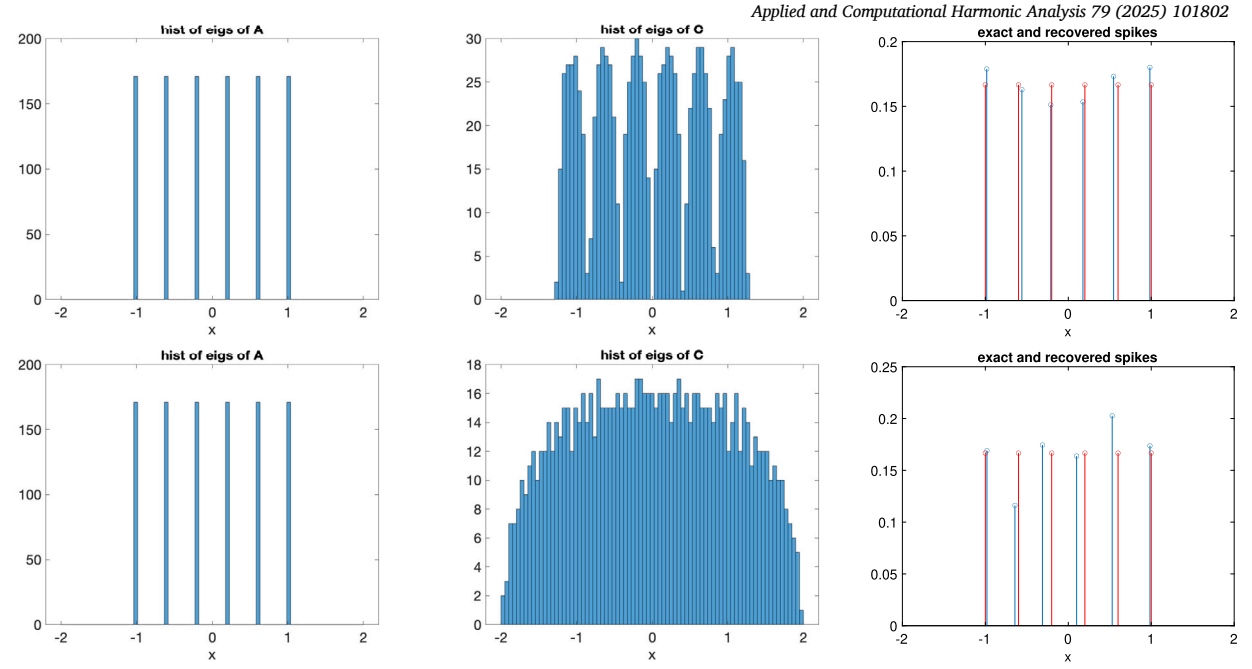

Fig. 3.3. Top: $\sigma = 0.25$. Bottom $\sigma = 0.75$. In each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right).

Example 3.2. This example considers μ_A with more spikes. Here, $N = 1024$ and μ_A is a union of 5 equally-spaced spikes in $[-1, 1]$. σ takes two values 0.25 and 0.75.

Fig. 3.3 summarizes the results. The top and bottom rows are for $\sigma = 0.25$ and $\sigma = 0.75$, respectively. Within each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right). While the reconstruction for small σ is quite accurate, the one for large σ shows significant error.

Example 3.3. In this example, μ_A is approximate sparse: μ_A is close to the first example, but with 25% of the eigenvalues randomly distributed in $[-1, 1]$. $N = 1024$ and σ takes two values 0.25 and 0.75.

Fig. 3.4 summarizes the results. The top and bottom rows are for $\sigma = 0.25$ and $\sigma = 0.75$, respectively. Within each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right). The reconstructions are quite accurate in both cases.

Example 3.4. This example performs a deterministic perturbation on μ_A: for each eigenvalue of μ_A from the first example, one perturbs the value by ± 0.05. $N = 1024$ and σ takes two values 0.25 and 0.75.

Fig. 3.5 summarizes the results. The top and bottom rows are for $\sigma = 0.25$ and $\sigma = 0.75$, respectively. Within each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right). The reconstructions are also quite accurate in both cases.

4. Multiplicative deconvolution

To address the multiplicative case, we leverage the S-transform [16]. Since $C = \sqrt{A}B\sqrt{A}$, in the large dimension limit

$$s_{\mu_C}(t) = s_{\mu_A}(t) s_{\mu_B}(t).$$

4.1. Known q

Assume for now that q is determined. $s_{\mu_B}(t) = \frac{1}{1+qt}$ from the Marchenko-Pastur law. Due to its sparsity, $\mu_A = \sum_k \delta_{x_k} w_k$. The task is to recover $\{x_k\}$ and $\{w_k\}$.

Given μ_C, we choose $\{z_j\}$ to be a set of points on an ellipsis around μ_C, compute

$$g_j = \int \frac{1}{z_j - x} d\mu_C(x)$$

and further define

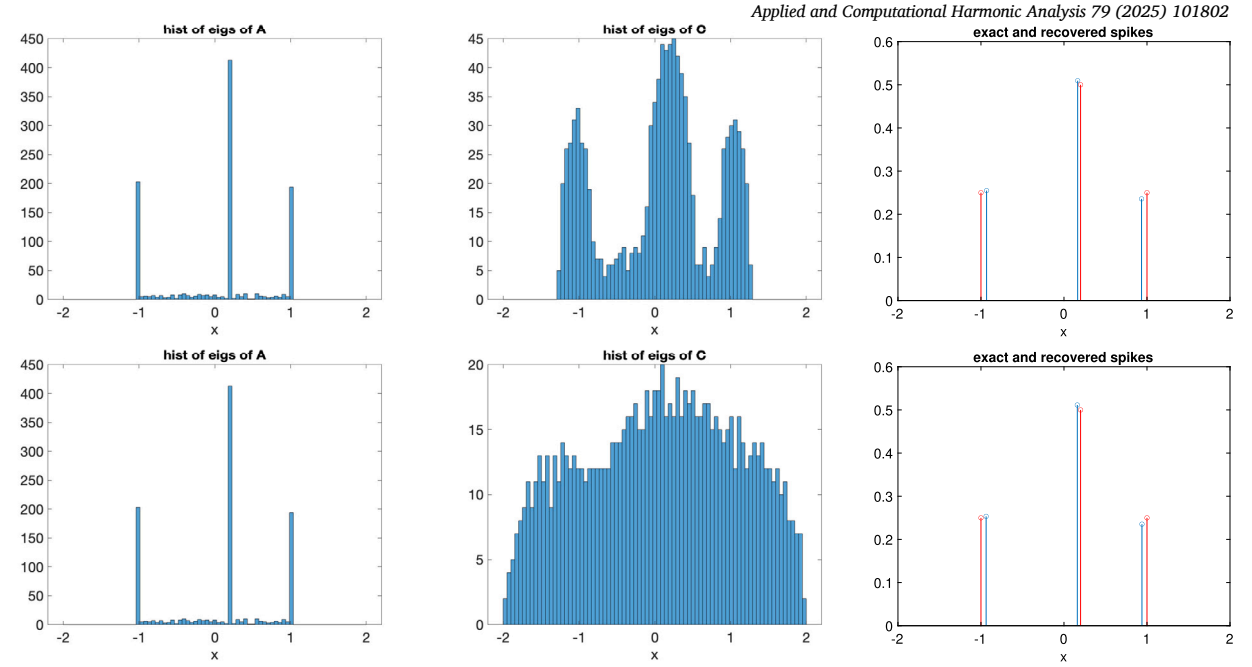

Fig. 3.4. Top: $\sigma = 0.25$. Bottom $\sigma = 0.75$. In each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right).

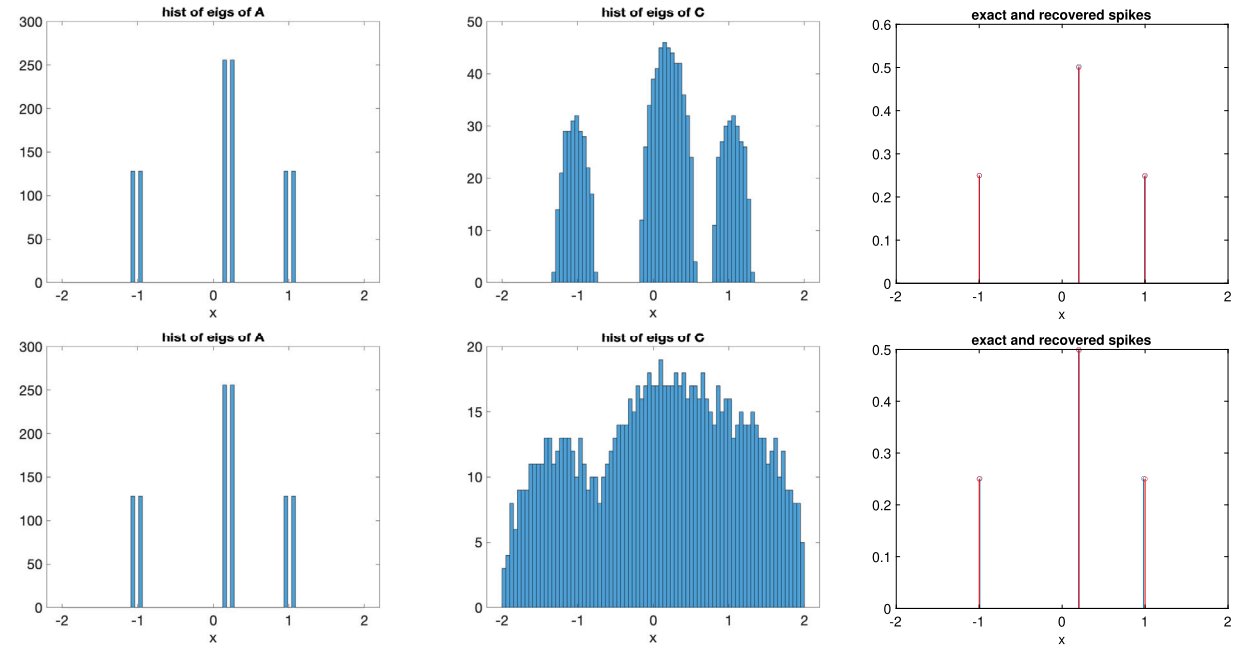

Fig. 3.5. Top: $\sigma = 0.25$. Bottom $\sigma = 0.75$. In each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right).

$$t_j = z_j g_j - 1, \quad s_j = \frac{t_j + 1}{t_j z_j},$$

i.e. $s_{\mu_C}(t_j) = s_j$.

From $s_{\mu_C}(t) = s_{\mu_A}(t) s_{\mu_B}(t)$ and $s_{\mu_B}(t) = \frac{1}{1+qt}$, we have

$$s_{\mu_A}(t_j) = s_{\mu_C}(t_j)/s_{\mu_B}(t_j) = s_{\mu_C}(t_j)(1 + qt_j) = s_j(1 + qt_j).$$

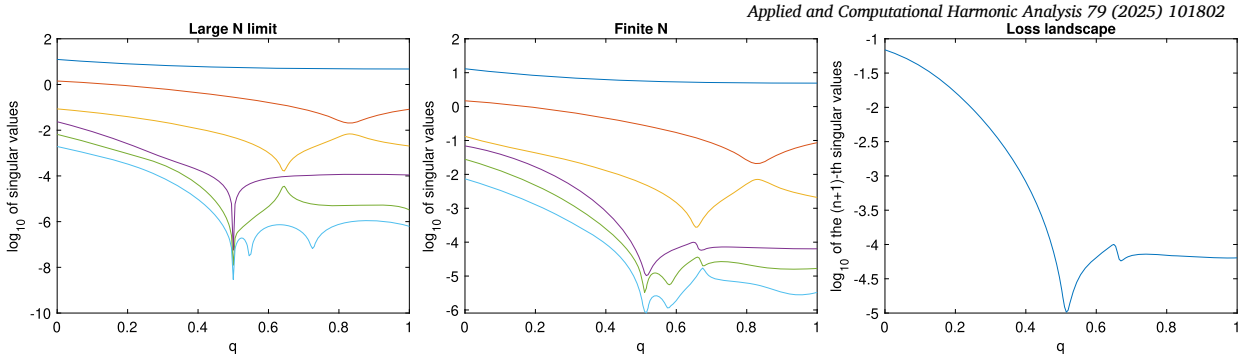

Fig. 4.1. Logarithm of singular values of $T(q)$ as a function of q. Left: large N limit. Middle: $N = 1024$. Right: loss landscape.

Introduce $s'_j = s_j(1 + qt_j)$. Then (t_j, s'_j) are samples of $s_{\mu_A}(t)$. i.e., $s_{\mu_A}(t_j) = s'_j$. Now further define

$$z'_j = \frac{t_j + 1}{t_j s'_j}, \quad g'_j = \frac{t_j + 1}{z'_j}.$$

The resulting (z'_j, g'_j) are samples of the Stieltjes transform of μ_A, i.e.,

$$g'_j = \int \frac{1}{z'_j - x} d\mu_A(x) = \sum_k \frac{1}{z'_j - x_k} w_k.$$

Since the locations $\{z'_j\}$ are not a priori controlled, recovering $\{x_k\}$ and $\{w_k\}$ is a sparse, unstructured recovery problem.

Next, we apply the eigenmatrix method. First, set X to be the shortest interval that covers the spectrum of C. Treat $\{z'_j\}$ as the samples and $\{g'_j\}$ as the observed data. Define $\mathbf{b}_x = \left[\frac{1}{z'_j - x}\right]_j$. Choose a Chebyshev grid $\{c_l\}$ and construct M such that $M\mathbf{b}_{c_l} \approx c_l \mathbf{b}_{c_l}$. Finally, define $\mathbf{u} = \left[g'_j\right]_j$ and form the T matrix to recover $\{x_k\}$ and $\{w_k\}$.

4.2. Unknown q

Let us discuss now how to determine q. For a value of q, the T matrix, as a function of q, will be denoted by $T(q)$. The main observation is similar to the additive case: for the correct q value, $T(q)$ has a numerical rank equal to n, while for other values of q, the rank is higher.

Consider an example where μ_A has $n = 3$ spikes and $q = 0.5$. Fig. 4.1 plots the singular values of $T(q)$ in a logarithmic scale as a function of q. The left plot shows the large N limit. At $q = 0.5$, starting from the 4th curve, the singular value drops numerically to zero. This plot indicates that the right q and n can be easily detected in the large N limit.

The middle plot shows the case $N = 1024$. The overall trends remain the same. However, due to the finiteness of N, it is harder to find an automatic procedure to identify q and n at the same time. In what follows, we assume n is given. In this case, the $(n+1)$-th (i.e., 4th) singular value curve clearly has a global minimum at the right q.

To automate this process, we formulate it again as a minimization problem

$$\hat{q} = \mathrm{argmin}_{q \in \Omega} \log(s_{n+1}(T(q))),$$

where $s_{n+1}(\cdot)$ refers to the $(n+1)$-th singular value and Ω is the region for which the eigenmatrix method produces positive weights $\{w_k\}$. The right plot highlights that this loss landscape is not convex, so local optimization can get stuck. Since this is still a one-dimensional optimization problem, we again take a two-step procedure similar to the additive case. First, we perform a coarse grid search to obtain a good initial guess q_{init}. Second, starting from q_{init}, a local search refines to the optimal $\hat{q}$. Finally, given $\hat{q}$, we can compute the (z'_j, g'_j) pairs of μ_A as in Section 4.1 to find $\{x_k\}$ and $\{w_k\}$.

Below, we give a few numerical examples.

Example 4.1. This example considers μ_A with three spikes and studies the performance for different values of σ. Here, $N = 1024$ and μ_A is supported at three spikes equally spaced on a logarithmic scale. q takes two values 0.1 and 0.5.

Fig. 4.2 summarizes the results. The top and bottom rows are for $q = 0.1$ and $q = 0.5$, respectively. Within each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right). Both reconstructions are accurate.

Example 4.2. This example considers μ_A with more spikes. Here, $N = 1024$ and μ_A is a union of 5 equally-spaced spikes. q takes two values 0.1 and 0.5.

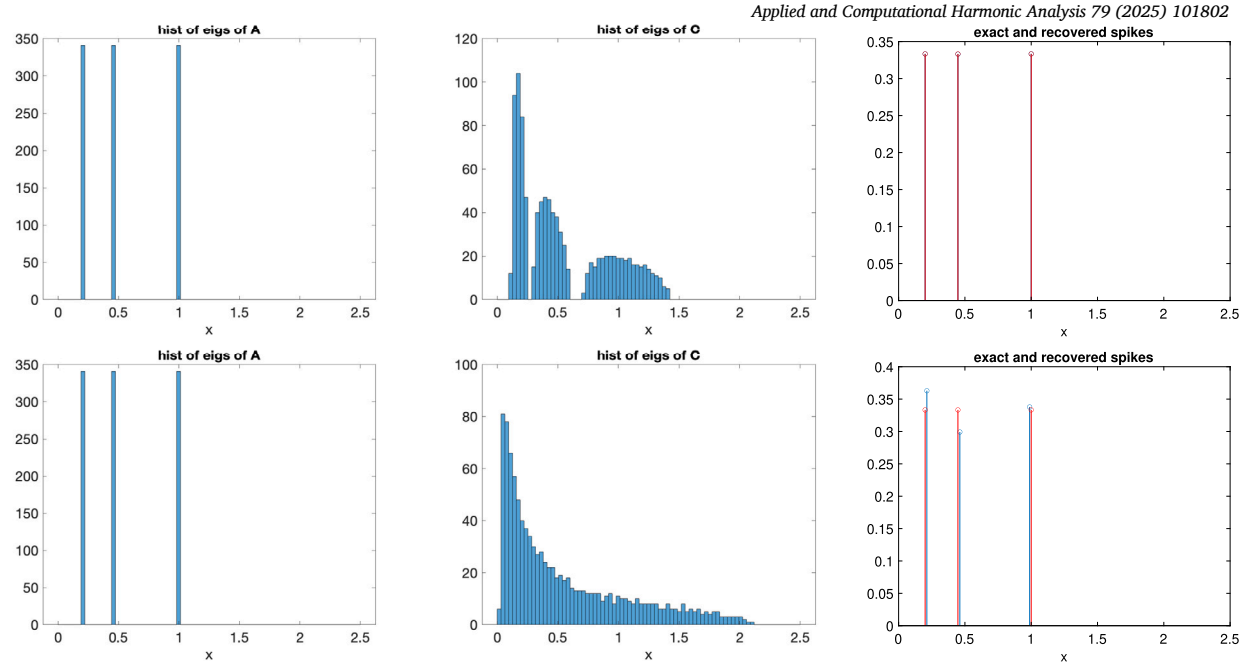

Fig. 4.2. Top: $q = 0.1$. Bottom $q = 0.5$. In each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right).

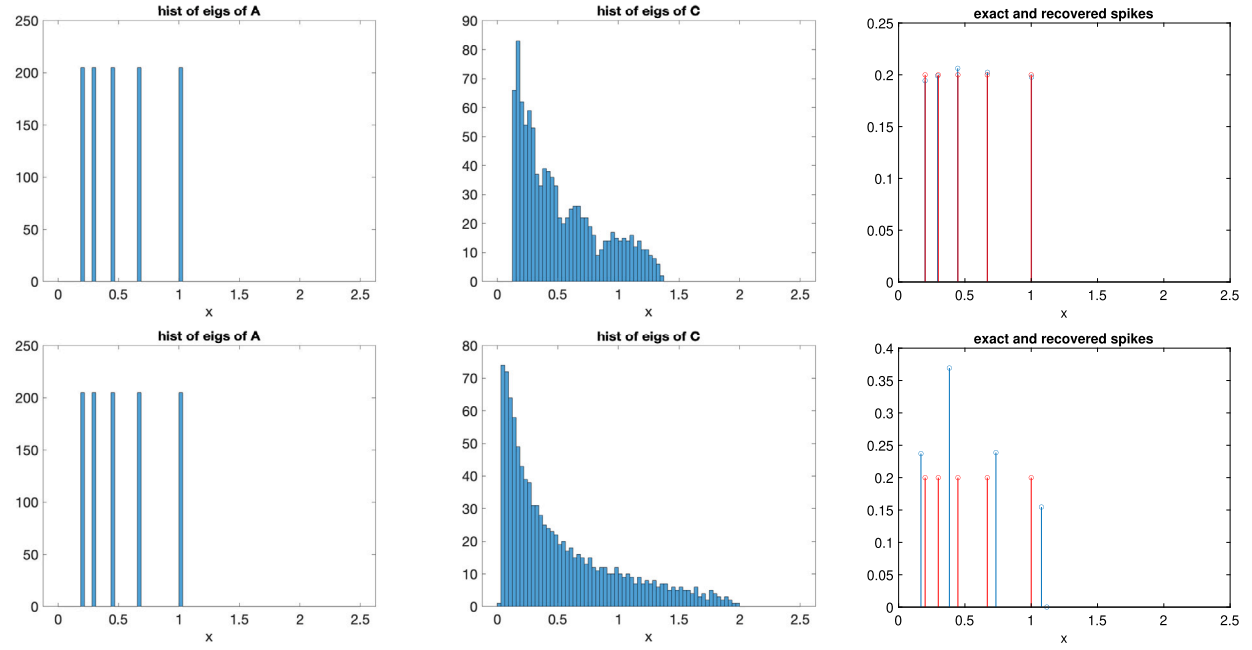

Fig. 4.3. Top: $q = 0.1$. Bottom $q = 0.5$. In each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right).

Fig. 4.3 summarizes the results. The top and bottom rows are for $q = 0.1$ and $q = 0.5$, respectively. Within each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right). While the reconstruction for small σ is quite accurate, the one for large q shows significant error.

Example 4.3. In this example, μ_A is approximate sparse: μ_A is close to the first example, but with 25% of the eigenvalues randomly distributed in $[0.2, 1]$. $N = 1024$ and q takes two values 0.1 and 0.5.

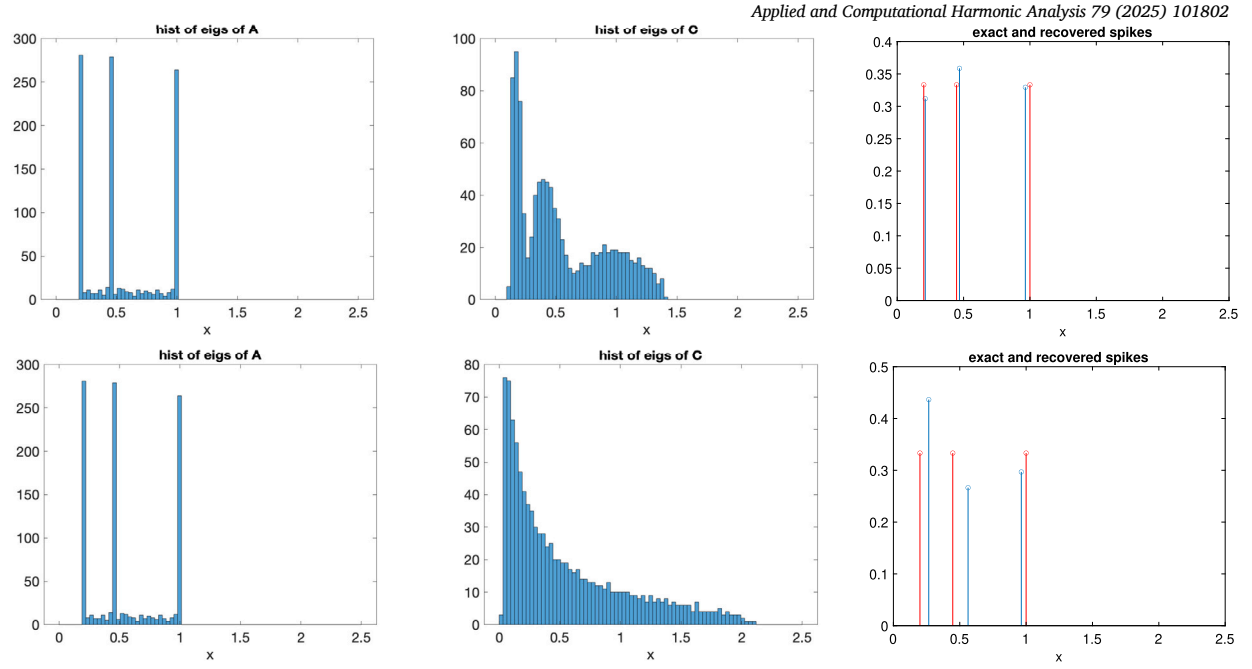

Fig. 4.4. Top: $q = 0.1$. Bottom $q = 0.5$. In each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right).

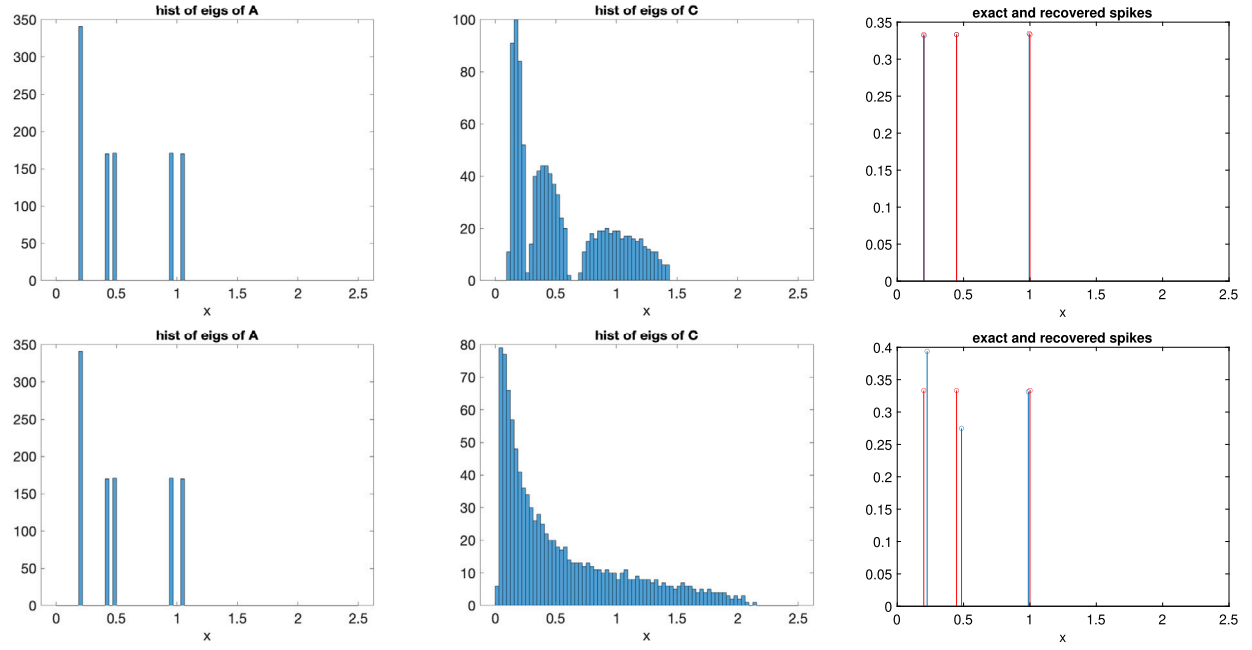

Fig. 4.5. Top: $q = 0.1$. Bottom $q = 0.5$. In each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right).

Fig. 4.4 summarizes the results. The top and bottom rows are for $q = 0.1$ and $q = 0.5$, respectively. Within each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right). The reconstruction of large q shows some clear errors.

Example 4.4. This example performs a deterministic perturbation on μ_A: for each eigenvalue of μ_A from the first example, one perturbs multiplicatively by ± 0.05. $N = 1024$ and q takes two values 0.1 and 0.5.

Fig. 4.5 summarizes the results. The top and bottom rows are for $q = 0.1$ and $q = 0.5$, respectively. Within each row, the plots are the histogram of the eigenvalues of A (left), the histogram of the eigenvalues of C (middle), and the exact (red) and reconstructed (blue) μ_A (right). The reconstruction of large q shows some errors in the weights.

5. Discussion

This section discusses several directions for future work. Here, we consider the most common scenarios for B: the Wigner matrix in the additive setting and the Wishart matrix multiplicative setting. If relevant to applications, other parametric families for B can also be considered.

Second, in the numerical examples, we assume the sparsity n is known. As we have seen, discovering the noise level and n together robustly from the eigenvalue decay of the T matrix is a non-trivial task. This involves a better understanding of the error of the eigenmatrix method and the finite effect of N.

Third, the current approach leverages the asymptotic relationship of the R-transform and the S-transform. For finite N, these relationships are approximate for spectral measures. As a result, the current approach has a systematic bias. An important direction is how to incorporate the N-dependent corrections for better estimations.

Fourth, the sparsity assumption of the spectral measure of A may not be appropriate for certain applications. For a specific application, if the spectral measure of A arises from some other low-complexity models controlled by a small number of parameters, this approach based on an eigenmatrix might also be useful.

Data availability

No data was used for the research described in the article.

References

[1] Octavio Arizmendi, Pierre Tarrago, Carlos Vargas, Subordination methods for free deconvolution, Ann. Inst. Henri Poincaré Probab. Stat. (2020) 2565–2590.
[2] Joël Bun, Romain Allez, Jean-Philippe Bouchaud, Marc Potters, Rotational invariant estimator for general noisy matrices, IEEE Trans. Inf. Theory 62 (12) (2016) 7475–7490.
[3] Noureddine El Karoui, Spectrum estimation for large dimensional covariance matrices using random matrix theory, Ann. Stat. 36 (1) (2008) 2757–2790.
[4] Matan Gavish, David L. Donoho, The optimal hard threshold for singular values is $4/\sqrt{3}$, IEEE Trans. Inf. Theory 60 (8) (2014) 5040–5053.
[5] Weihao Kong, Gregory Valiant, Spectrum estimation from samples, Ann. Stat. 45 (2017) 5.
[6] Olivier Ledoit, Sandrine Péché, Eigenvectors of some large sample covariance matrix ensembles, Probab. Theory Relat. Fields 151 (1) (2011) 233–264.
[7] Olivier Ledoit, Michael Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivar. Anal. 88 (2) (2004) 365–411.
[8] Olivier Ledoit, Michael Wolf, Nonlinear shrinkage estimation of large-dimensional covariance matrices, Ann. Stat. 40 (2) (2012) 1024–1060.
[9] Olivier Ledoit, Michael Wolf, Spectrum estimation: a unified framework for covariance matrix estimation and pca in large dimensions, J. Multivar. Anal. 139 (2015) 360–384.
[10] Olivier Ledoit, Michael Wolf, Analytical nonlinear shrinkage of large-dimensional covariance matrices, Ann. Stat. 48 (5) (2020) 3043–3065.
[11] Panagiotis Lolas, Lexing Ying, Shrinkage estimation of functions of large noisy symmetric matrices, arXiv preprint arXiv:2106.05183, 2021.
[12] James A. Mingo, Roland Speicher, Free Probability and Random Matrices, vol. 35, Springer, 2017.
[13] Marc Potters, Jean-Philippe Bouchaud, A First Course in Random Matrix Theory: for Physicists, Engineers and Data Scientists, Cambridge University Press, 2020.
[14] Richard Roy, Thomas Kailath, Esprit-estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust. Speech Signal Process. 37 (7) (1989) 984–995.
[15] Dan Voiculescu, Addition of certain non-commuting random variables, J. Funct. Anal. 66 (3) (1986) 323–346.
[16] Dan Voiculescu, Multiplication of certain non-commuting random variables, J. Oper. Theory (1987) 223–235.
[17] Lexing Ying, Eigenmatrix for unstructured sparse recovery, Appl. Comput. Harmon. Anal. 71 (2024) 101653.
[18] Lexing Ying, Multidimensional unstructured sparse recovery via eigenmatrix, Appl. Comput. Harmon. Anal. 74 (2025) 101725.