On the endmembers detection for hyperspectral imaging in the high-dimensional regime

H. Jeannin[1], F. Bouchard[2], P. Vallet[1], G. Ginolhac[3], A. Giremus[1]
[1]Université Bordeaux, CNRS, Bordeaux INP, *IMS*, Talence (France)
[2]Université Paris-Saclay, CNRS, CentraleSupélec, *L2S*, Gif-sur-Yvette (France)
[3]Université Savoie Mont-Blanc, *LISTIC*, Annecy (France)
{hugo.jeannin, pascal.vallet, audrey.giremus}@u-bordeaux.fr, florent.bouchard@cnrs.fr, guillaume.ginolhac@univ-smb.fr

***Abstract*—This paper addresses the estimation of the number of endmembers in a hyperspectral image composed of p spectral bands. Each pixel is modelled as a Dirichlet linear mixture of k endmembers corrupted by an additive Gaussian noise correlated across the spectral bands and having a Toeplitz covariance matrix. Assuming than n i.i.d. pixels are observed, we leverage on the spiked model techniques in random matrix theory to study the behaviour of the largest eigenvalues of the non-centered sample covariance matrice in the high-dimensional asymptotic regime in which p, n converge to infinity at the same rate while k is kept fixed. The results are further specified in the case where the noise covariance follows a first-order autoregressive model, and an estimate of k is developed. Numerical experiments on real world datasets show promising results compared to alternative approaches of the literature.**

***Index Terms*—hyperspectral imaging, intrinsic dimension, high-dimensional regime, random matrix theory**

I. Introduction

Hyperspectral imaging encompasses many different tasks, among which the estimation of the number of spectrally distinct signatures, often termed as *virtual dimensionality* (VD), which is a preliminary step for more involved tasks including spectral unmixing and pixel classification/identification.

Although the precise definition of VD has been subject to some variations in the literature, the classical methods, e.g. [1]–[3] (see also [4] for a comprehensive review), assume that the observed signal is a linear mixture of the material spectral signatures corrupted by some additive noise, and rely on a signal/noise subspace decomposition to define the VD. Let us consider an observed pixel $\mathbf{y}$ composed of p spectral bands, which is modelled as

$$\mathbf{y} = \sum_{i=1}^{k} \mathbf{a}_i s_i + \mathbf{v}, \tag{1}$$

where $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{R}^p$ are the spectral signature vectors, assumed deterministic, $s_1, \ldots, s_k \geq 0$ are the fractional abundances, which can be modelled either as deterministic or random, and $\mathbf{v}$ some random additive noise with zero-mean and covariance matrix $\mathbb{V}[\mathbf{v}] := \mathbb{E}[\mathbf{v}\mathbf{v}^\top] = \mathbf{\Sigma}$. In most of the aforementioned approaches, it is assumed that the noise is white across the spectral bands, i.e. $\mathbf{\Sigma} = \sigma^2\mathbf{I}$, and that a sample $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of (1) is available to estimate the eigenvalues and eigenvectors of the covariance matrix of $\mathbf{y}$, from which an estimate of k can be obtained. In particular, those methods often rely on the sample covariance matrix (SCM):

$$\hat{\mathbf{R}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{y}_i\mathbf{y}_i^\top, \tag{2}$$

and/or its centered version

$$\tilde{\mathbf{R}}_n = \frac{1}{n}\sum_{i=1}^{n} (\mathbf{y}_i - \overline{\mathbf{y}}_n)(\mathbf{y}_i - \overline{\mathbf{y}}_n)^\top, \tag{3}$$

with $\overline{\mathbf{y}}_n = n^{-1}\sum_{i=1}^{n} \mathbf{y}_i$, which are consistent estimates in the asymptotic regime where the dimension p is fixed and the sample size $n \to \infty$, and under a large class of statistical models for $\mathbf{y}$.

In practice, the reliability of such estimators requires that $n \gg p$, which is often unrealistic in the context of hyperspectral imaging since p is large (usually of the order of a hundred bands), and the number of "sample" pixels n is limited to preserve the stationarity across $\mathbf{y}_1, \ldots, \mathbf{y}_n$. In that case, it is more reasonable to assume that p and n are of the same order of magnitude, in which case $\hat{\mathbf{R}}_n$ and $\tilde{\mathbf{R}}_n$ are known to be poor estimates. To retrieve a situation where $p \ll n$, dimension reduction techniques can be used, but they rely on a delicate choice of certain hyperparameters, which may also be related to an accurate estimation of the unknown covariance matrices.

An alternative approach to model the situation where p and n are of the same order of magnitude consists in using the *high-dimensional regime* in which $p, n \to \infty$ such that $\frac{p}{n}$ converges to a positive constant, while k is fixed. In that case, results from *random matrix theory* and the so-called *spiked models* [5] can be exploited to study the behaviour of the largest eigenvalues of $\hat{\mathbf{R}}_n$ and $\tilde{\mathbf{R}}_n$. This direction was explored in [6]–[8] in which estimates of k based on thresholding each eigenvalue of $\hat{\mathbf{R}}_n$, or the difference between successive eigenvalues, were provided under the general scenario where $\mathbf{\Sigma}$ is not necessarily equal to $\sigma^2\mathbf{I}$. However, both above-cited approaches rely on estimates of $\mathbf{\Sigma}$ which are not proved to be consistent in the high-dimensional regime. In practice, the poor estimation of $\mathbf{\Sigma}$ due to the high-dimensionality of the model

Identify applicable funding agency here. If none, delete this.

leads to a severe performance degradation of VD estimation techniques [9].

In this paper, we pursue the path taken in [7] and provide, in section II, a study of the largest eigenvalues of $\hat{\mathbf{R}}_n$ in the high-dimensional regime and in the not too restrictive case where $\mathbf{\Sigma}$ is assumed Toeplitz. The study relies on the technique developed in [10] for spiked models as well as the results of [11]. In section III, the results are further specified in the case of an AR(1) covariance matrix, and an estimate of the VD is developed. Providing an upper bound on the total number of endmembers is available as in [7], [8], and relying on a conjecture on the fluctuations of the largest noise eigenvalue of $\hat{\mathbf{R}}_n$, this estimate is also proved to be consistent in the high-dimensional regime. Finally, in section IV, numerical experiments are performed on the real datasets ([12], [13]) for which a ground truth is available, and shown to be in agreement with the theoretical guarantee of the proposed method as well as competitive with alternative approaches. The code for our experiments is available at [14].

II. Spectrum of the SCM in the high-dimensional regime

We first detail the statistical model used in this paper. Let us rewrite (1) as

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{v}, \tag{4}$$

where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$ and $\mathbf{s} = (s_1, \ldots, s_k)^\top$. We assume in the following that $\mathbf{v} \sim \mathcal{N}_{\mathbb{R}^p}(\mathbf{0}, \mathbf{\Sigma})$ and that $\mathbf{s}$ follows a Dirichlet distribution with parameter $\boldsymbol{\alpha} \in (0,\infty)^k$ for which we recall that $\mathbb{E}[\mathbf{s}] = \|\boldsymbol{\alpha}\|_1^{-1}\boldsymbol{\alpha}$ and

$$\mathbf{\Gamma} = \mathbb{V}[\mathbf{s}] = \frac{1}{1+\|\boldsymbol{\alpha}\|_1}\left(\mathrm{dg}\left(\frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|_1}\right) - \frac{\boldsymbol{\alpha}\boldsymbol{\alpha}^\top}{\|\boldsymbol{\alpha}\|_1^2}\right),$$

where $\|.\|_1$ is the ℓ^1 norm and $\mathrm{dg}(\mathbf{u})$ is the diagonal matrix having the entries of vector $\mathbf{u}$ on its diagonal. The noise covariance matrix is assumed to be Toeplitz with the following assumption on the generating covariance sequence.

***Assumption** 1: We have $\mathbf{\Sigma} = (\rho_{i-j})_{i,j=1,\ldots,p}$ where $(\rho_\ell)_{\ell\in\mathbb{Z}}$ satisfies $\sum_{\ell\in\mathbb{Z}} |\ell|^2|\rho_\ell| < \infty$. In addition, the associated spectral density*

$$s(t) := \sum_{\ell\in\mathbb{Z}} \rho_\ell \mathrm{e}^{-\mathrm{i}2\pi\ell t},$$

satisfies $\mathsf{s}_- := \min_{t\in[0,1]} s(t) > 0$.

In particular, Assumption 1 ensures that the spectral density s associated with $(\rho_\ell)_{\ell\in\mathbb{Z}}$ is twice continuously differentiable on $[0,1]$. By denoting $\mathsf{s}_+ := \max_{t\in[0,1]} s(t)$, we further have

$$0 < \mathsf{s}_- \leq \lambda_p(\mathbf{\Sigma}) \leq \ldots \leq \lambda_1(\mathbf{\Sigma}) \leq \mathsf{s}_+ < \infty,$$

where for a $p \times p$ symmetric matrix $\mathbf{B}$, $\lambda_1(\mathbf{B}), \ldots, \lambda_p(\mathbf{B})$ denotes its eigenvalues in decreasing order. We also assume that i.i.d. copies of $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of (4) are available. The next assumption is related to the high-dimensional regime used in the study.

***Assumption** 2: $p = p(n)$ is a function of n such that $\frac{p}{n} \to c > 0$ as $n \to \infty$, while k, $\boldsymbol{\alpha}$ and s are independent of n. Moreover, if $\mathbf{\Xi} = \mathbf{\Gamma} + \|\boldsymbol{\alpha}\|_1^{-2}\boldsymbol{\alpha}\boldsymbol{\alpha}^\top$, then for all $i = 1, \ldots, k$, $\lambda_i\left(\mathbf{A}\mathbf{\Xi}\mathbf{A}^\top + \mathbf{\Sigma}\right) \to \gamma_i$ as $n \to \infty$, with $\gamma_1 \geq \ldots \geq \gamma_k > \mathsf{s}_+$.*

By leveraging model (4) and Assumptions 1-2, we are now in position to state the results regarding the asymptotic behaviour of the eigenvalues of $\hat{\mathbf{R}}_n$ defined in (2). Since k is assumed fixed, we are in the context of the so-called spiked models, and the global asymptotic behaviour of the eigenvalues only depends on c and the spectral density s.

Let us consider the *empirical eigenvalue distribution* of $\mathbf{\Sigma}$ defined as the probability measure $\nu_n = p^{-1}\sum_{i=1}^p \delta_{\lambda_i(\mathbf{\Sigma})}$, where δ_x is the Dirac measure at point x. From Assumption 1 and Szegö theorem [15], it is well-known that

$$\nu_n \xrightarrow[n\to\infty]{w} \nu := \tau \circ s^{-1},$$

i.e. the sequence (ν_n) converges weakly to the image measure of the Lebesgue measure τ on $[0,1]$ by s. Moreover, the support of ν is given by $\mathrm{supp}(\nu) = [\mathsf{s}_-, \mathsf{s}_+]$.

Let us now denote by $\hat{\mu}_n$ the empirical eigenvalue distribution of $\hat{\mathbf{R}}_n$. Then one can prove using [16] that almost surely (a.s.), $\hat{\mu}_n \xrightarrow{w} \mu$ as $n \to \infty$, where μ is a deterministic probability distribution whose *Stieltjes transform* $z \mapsto m(z) = \int_{\mathbb{R}}(\lambda - z)^{-1}\mathrm{d}\mu(\lambda)$, defined on $\mathbb{C}\backslash\mathrm{supp}(\mu)$, verifies the following equation

$$m(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\nu(\lambda)}{\lambda(1-c-czm(z)) - z}. \tag{5}$$

Moreover, using Assumption 1 and the results of [17, Th. 1.1] (see also [11]), we have the following decomposition

$$\mu(\mathrm{d}\lambda) = f_\mu(\lambda)\mathrm{d}\lambda + \left(1 - \frac{1}{c}\right)^+ \delta_0(\mathrm{d}\lambda), \tag{6}$$

where the density f_μ has a compact interval as support and is given by $f_\mu(\lambda) = \pi^{-1}\lim_{y\downarrow 0}\mathrm{Im}\left(m\left(\lambda + \mathrm{i}y\right)\right)$. It turns out that the support of f_μ can be further characterized using again the results of [17, Th. 4.1-4.3] which we summarize below. Define for all $w \in \mathbb{R}\backslash\mathrm{supp}(\nu)$,

$$\phi(w) = w\left(1 - c - cw\eta(w)\right). \tag{7}$$

where $w \mapsto \eta(w) = \int_{\mathbb{R}}(\lambda - w)^{-1}\mathrm{d}\nu(\lambda)$ is the Stieltjes transform of ν. From Assumption 1, it can be verified that $\eta(w) \to \pm\infty$ as $w \to \mathsf{s}_\mp$, $\eta(w) \to \pm 0$ as $w \to \pm\infty$ and that η is increasing on $\mathbb{R}\backslash\mathrm{supp}(\nu)$. As a consequence, we have $\phi(w) \to \pm\infty$ as $w \to \mathsf{s}_\pm$ and $w \to \pm\infty$, and ϕ admits a unique local maximum at point w_- (resp. a unique local minimum at point w_+), with $\mathsf{w}_- \in (-\infty, \mathsf{s}_-)$ and $\mathsf{w}_+ \in (\mathsf{s}_+, +\infty)$. Finally, denoting $\mathsf{x}_\pm := \phi(\mathsf{w}_\pm)$, we have $\mathrm{supp}(f_\mu) = [\mathsf{x}_-, \mathsf{x}_+]$.

Moreover, we have $\mathsf{s}_+ < \mathsf{w}_+ < \mathsf{x}_+ \leq \mathsf{s}_+(1+\sqrt{c})^2$. In practice, the computation of η can be done via some numerical integration, but in some special cases such as the AR(1) model described in the next section, a closed-form expression can be obtained.

As we have seen, the fixed rank perturbation generated by $\mathbf{A}\mathbf{s}$ in (4) does not impact the global asymptotic behaviour of the eigenvalues of $\hat{\mathbf{R}}_n$ described in (6), which only depends on c and the noise spectral density s. However, if we focus

on the individual behaviour of the k largest eigenvalues, more can be said, as described by the next result.

***Theorem** 1: Under Assumptions 1 and 2, we have for all $i = 1, \ldots, k$,*

$$\lambda_i\left(\hat{\mathbf{R}}_n\right) \xrightarrow[n\to\infty]{\text{a.s.}} \begin{cases} \phi\left(\gamma_i\right) & \text{if } \gamma_i > \mathsf{w}_+ \\ \mathsf{x}_+ & \text{if } \gamma_i \leq \mathsf{w}_+ \end{cases},$$

while $\lambda_{k+1}(\hat{\mathbf{R}}_n) \to \mathsf{x}_+$ and $\lambda_p(\hat{\mathbf{R}}_n) \to \mathsf{x}_-$ a.s. as $n \to \infty$. Additionally, if $\gamma_i > \mathsf{w}_+$, we have

$$\lambda_i\left(\hat{\mathbf{R}}_n\right) = \phi(\gamma_i) + \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}}).$$

Proof: The proof, which follows verbatim the steps of [10], is omitted. We note however that our model (4) does not strictly satisfies the hypothesis formulated in [10] (termed as "i.i.d". or "orthonormalized" models) due to the strong dependency between the entries of $\mathbf{s}$. To deal with this special case, we use a standard concentration inequality [18] for sub-Gaussian random vectors which allows one to study random matrices of the type $\mathbf{S}\left(n^{-1}\mathbf{V}^\top\mathbf{V} - z\mathbf{I}\right)^{-1}\mathbf{S}^\top$, $\mathbf{A}^\top\left(n^{-1}\mathbf{V}\mathbf{V}^\top - z\mathbf{I}\right)^{-1}\mathbf{V}\mathbf{S}^\top$ for $z \in \mathbb{C}\backslash\mathbb{R}$, with $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_n]$, $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$. Those matrices appear in [10, Lemma 4.1] and are the keystone of the proof. ∎

Thus, providing that the limit eigenvalues $\gamma_1, \ldots, \gamma_k$ are above the threshold w_+, a phase transition phenomenon occurs and we observe that the corresponding largest eigenvalues $\lambda_1(\hat{\mathbf{R}}_n), \ldots, \lambda_k(\hat{\mathbf{R}}_n)$ escape from the bulk of the "noise" eigenvalues $\lambda_{k+1}(\hat{\mathbf{R}}_n), \ldots, \lambda_p(\hat{\mathbf{R}}_n)$ which converge into the interval $[\mathsf{x}_-, \mathsf{x}_+]$.

It is also worth noting that the limit eigenvalue γ_i defined in Assumption 2 depends on both the signal covariance $\mathbf{A}\boldsymbol{\Xi}\mathbf{A}^\top$ and the noise covariance $\boldsymbol{\Sigma}$ which do not share the same eigenbasis in general. Nevertheless, using Weyl's inequality, if the condition

$$\liminf_{n\to\infty} \lambda_i\left(\mathbf{A}\boldsymbol{\Xi}\mathbf{A}^\top\right) > \mathsf{w}_+ - \mathsf{s}_-, \tag{8}$$

is verified, then we have $\gamma_i > \mathsf{w}_+$. Although not optimal, the sufficient condition (8) involves a decoupling between the eigenvalues of the signal part $\mathbf{A}\boldsymbol{\Xi}\mathbf{A}^\top$ and the noise covariance $\boldsymbol{\Sigma}$.

***Remark** 1: In the well-known white noise case, $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, we have $\phi(w) = w(\sigma^2 - w)^{-1}(\sigma^2(1-c) - w)$ so that $\mathsf{w}_+ = \sigma^2(1+\sqrt{c})$ and $\mathsf{x}_+ = \sigma^2(1+\sqrt{c})^2$, and the condition $\gamma_i > \mathsf{w}_+$ rewrites $\lambda_i\left(\mathbf{A}\boldsymbol{\Xi}\mathbf{A}^\top\right) \gtrapprox \sigma^2\sqrt{c}$. Therefore, in that case, the sufficient condition (8) appears sharp.*

III. The special case of an AR(1) profile

In this section, we specify the previous results in the special case where the covariance sequence $(\rho_\ell)_{\ell\in\mathbb{Z}}$ is such that

$$\rho_\ell = \frac{\sigma^2\theta^{|\ell|}}{1-\theta^2},$$

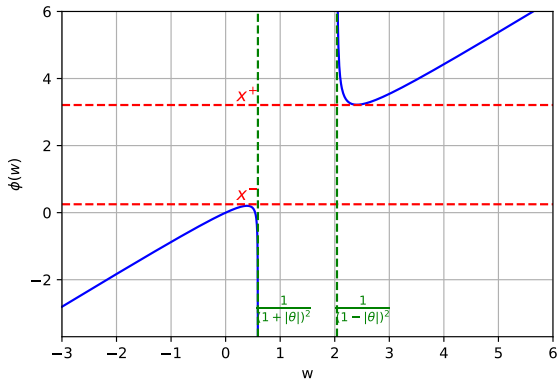

Fig. 1. Function ϕ for $\sigma^2 = 1$, $\theta = 0.3$, $c = 0.25$.

for some $\sigma^2 > 0$, $\theta \in (-1, 1)$. Using the associated spectral density $s(t) = \sigma^2|1 - \theta\mathrm{e}^{-\mathrm{i}2\pi t}|^{-2}$, a straightforward computation provides

$$\phi(w) = w\left(1 + c\,\epsilon(w)\sqrt{\frac{1}{(1-\frac{w}{\mathsf{s}_+})(1-\frac{w}{\mathsf{s}_-})}}\right), \tag{9}$$

with $\mathsf{s}_\pm = \sigma^2(1 \mp |\theta|)^{-2}$ and where $\epsilon(w) = -1$ if $w < \mathsf{s}_-$ and 1 if $w > \mathsf{s}_+$. A typical illustration of ϕ is given in Figure 7, together with the edge points $\mathsf{x}_\pm$ of $\mathrm{supp}(f_\mu)$ defined in the previous section. In particular, we have from (7) that $\mathsf{x}_+ \to +\infty$ as $|\theta| \to 1$ and the same holds for the critical threshold w_+ in Theorem 1. In other words, from (8), for highly correlated noise ($|\theta|$ close to 1), the signal strength represented by the eigenvalues $\lambda_i(\mathbf{A}\boldsymbol{\Xi}\mathbf{A}^\top)$, $i = 1, \ldots, k$, must be also large for the corresponding eigenvalues of $\hat{\mathbf{R}}_n$ to escape the bulk of the noise eigenvalues $[\mathsf{x}_-, \mathsf{x}_+]$.

Providing an upper bound $\mathsf{k}_{\max}$, *independent of* n, on the number k of endmembers is available, it is also interesting to note that estimates of the noise covariance parameters σ^2 and θ, consistent in the high-dimensional regime, can be obtained. Indeed, denote by

$$\hat{\alpha}_n = \frac{1}{p - \mathsf{k}_{\max}} \sum_{i=\mathsf{k}_{\max}+1}^{p} \lambda_i\left(\hat{\mathbf{R}}_n\right),$$
$$\hat{\beta}_n = \frac{1}{p - \mathsf{k}_{\max}} \sum_{i=\mathsf{k}_{\max}+1}^{p} \lambda_i\left(\hat{\mathbf{R}}_n\right)^2.$$

Then we have by straightforward computations

$$\hat{\alpha}_n = \frac{1}{p}\mathrm{tr}\boldsymbol{\Sigma} + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right),$$
$$\hat{\beta}_n = \frac{1}{pn}\left(\mathrm{tr}\boldsymbol{\Sigma}\right)^2 + \frac{1}{p}\mathrm{tr}\boldsymbol{\Sigma}^2 + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right),$$

as well as,

$$\frac{1}{p}\mathrm{tr}\boldsymbol{\Sigma} = \frac{\sigma^2}{1-\theta^2},\ \frac{1}{p}\mathrm{tr}\boldsymbol{\Sigma}^2 = \frac{\sigma^4(1+\theta^2)}{(1-\theta^2)^3} + \mathcal{O}\left(\frac{1}{n}\right).$$

Putting these estimates together, let

$$\hat{\theta}_n^2 = \frac{|\hat{\beta}_n - (1+\frac{p}{n})\hat{\alpha}_n^2|}{\hat{\beta}_n + (1-\frac{p}{n})\hat{\alpha}_n^2},\ \hat{\sigma}_n^2 = \hat{\alpha}_n|1 - \hat{\theta}_n^2|. \tag{10}$$

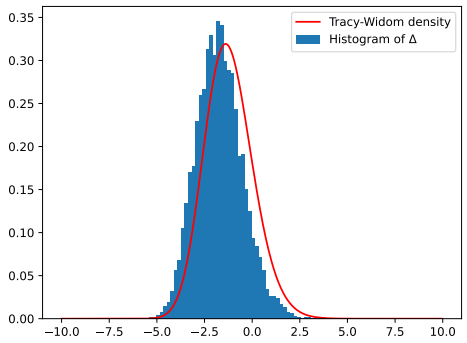

Fig. 2. Histogram of Δ (5000 draws) and Tracy-Widom density

Then we have $\hat{\theta}_n^2 = \theta^2 + \mathcal{O}_{\mathbb{P}}\left(n^{-1}\right)$ and $\hat{\sigma}_n^2 = \sigma^2 + \mathcal{O}_{\mathbb{P}}\left(n^{-1}\right)$.

Let us now denote by $\hat{s}_{\pm,n} = \hat{\sigma}_n^2(1 \mp |\hat{\theta}_n|)^{-2}$ the associated estimates of $\mathsf{s}_\pm$, as well as $\hat{\phi}_n$ the corresponding estimate of ϕ given in (9) where $\mathsf{s}_\pm$ are replaced by $\hat{s}_{\pm,n}$. Then for any compact interval $I \subset (\mathsf{s}_+, +\infty)$, we have $\sup_{w\in I} \left|\hat{\phi}_n(w) - \phi(w)\right| \to 0$ a.s. as $n \to \infty$. This in turn implies that the unique local minimum of $\hat{\phi}_n$ on $(\hat{s}_{+,n}, +\infty)$, which we denote by $\hat{x}_n$ verifies

$$\hat{x}_n = \mathsf{x}_+ + \mathcal{O}_{\mathbb{P}}(n^{-1}). \tag{11}$$

The estimate $\hat{x}_n$ of the right edge x_+ can be used to build a consistent estimate of the rank k, providing the following conjecture on the fluctuations of the largest noise eigenvalue $\lambda_{k+1}(\hat{\mathbf{R}}_n)$.

***Conjecture** 1: If $\frac{p}{n} = c + \mathcal{O}(n^{-1})$ and $\gamma_k > \mathsf{w}_+$, then*

$$\lambda_{k+1}(\hat{\mathbf{R}}_n) = \mathsf{x}_+ + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{2}{3}}\right).$$

We note that Conjecture 1 has been proved for $\mathbf{s}$ modelled as a complex or real Gaussian vector [19, Th. 4] [20]. Owing to the universality principle in random matrix theory, we expect Conjecture 1 to hold for a broader class of distributions, including the Dirichlet distribution used in this work, although the proof would be beyond the scope of this paper. We also add that Conjecture 1 should be formally stated as $\lambda_{k+1}(\hat{\mathbf{R}}_n) = x_{+,n} + \mathcal{O}_{\mathbb{P}}(n^{-\frac{2}{3}})$ where $x_{+,n}$ is defined as the largest local minimum of the function ϕ_n, the non-asymptotic version of ϕ in (7) with $w \mapsto \eta(w)$ replaced by $w \mapsto p^{-1}\mathrm{tr}(\mathbf{A}\mathbf{\Xi}\mathbf{A}^T + \mathbf{\Sigma} - w\mathbf{I})^{-1}$. But from Assumptions 1 and 2, it can be proved that $x_{+,n} = \mathsf{x}_+ + \mathcal{O}(n^{-1})$. Finally, we compare in Figure 2 the empirical distribution function of

$$\Delta = n^{2/3}\left(\frac{2}{\mathsf{w}_+^4\phi''(\mathsf{w}_+)}\right)^{\frac{1}{3}}\left(\lambda_{k+1}(\hat{\mathbf{R}}_n) - \mathsf{x}_+\right),$$

together with the density of the GOE Tracy-Widom [20], for $p = 200$, $n = 400$, $\theta = 0.5$, $\sigma = 1$, $\boldsymbol{\alpha} = (1, \ldots, 1)^\top$ $\mathbf{A} = \sqrt{10}[\mathbf{e}_1, \ldots, \mathbf{e}_k]$ with $\mathbf{e}_1, \ldots, \mathbf{e}_k$ the first k elements of the standard basis of $\mathbb{R}^p$. The observed result tends to validate Conjecture 1 empirically.

Using Conjecture 1 together with (11) yields

$$\lambda_{k+1}(\hat{\mathbf{R}}_n) = \hat{x}_n + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{2}{3}}\right),$$

while from Theorem 1, we have $\lim \lambda_k(\hat{\mathbf{R}}_n) > \mathsf{x}_+$ a.s. Therefore, we deduce the following result.

***Proposition** 1: Let $(\kappa_n)_{n\geq 1}$ a deterministic sequence such that $C_1 n^{-\delta} \leq \kappa_n \leq C_2 n^{-\delta}$ for some constants $C_1, C_2 > 0$ and $0 < \delta < \frac{2}{3}$, and define*

$$\hat{k}_n := \max\{i : \lambda_i(\hat{\mathbf{R}}_n) > \hat{x}_n + \kappa_n\}.$$

Then under Assumptions 1, 2 and Conjecture 1, and if $\gamma_k > \mathsf{w}_+$, we have $\mathbb{P}\left(\hat{k}_n = k\right) \to 1$ as $n \to \infty$.

***Remark** 2:* This method can be generalized to other models such as MA(1).

IV. Simulations

We first illustrate the convergence of the empirical eigenvalue distribution $\hat{\mu}_n$ to μ described in Section II on a real dataset [14]. We use the Urban hyperspectral scene [12], which is an aerial image of size 307×307 pixels composed of 210 wavelengths uniformly spaced between 400–2500 nm. After excluding channels 1-4, 76, 87, 101-111, 136-153, and 198-210 due to dense water vapor and atmospheric effects, a total of $p = 162$ wavelengths remain. This preprocessing step is commonly used in hyperspectral unmixing analyses. More precisely, we use the red-marked homogeneous region represented in Figure 3 composed of 1024 pixels and which we split into four parts of equal sizes, giving four samples of size $n = 256$ so that $\frac{p}{n} \approx 0.63$.

Fig. 3. Representation of the 70th band of the Urban hyperspectral image. An homogeneous area has been outlined in red.

In Figure 4, we plot the histogram of the eigenvalues $\lambda_{k_{max}+1}, \ldots, \lambda_p$ over the four samples. We also plot the density f_μ computed through its Stieltjes transform m given in (5), for the AR(1) model case described in Section III whose parameters σ^2 and θ are estimated using (10) and a bound $\mathsf{k}_{\max} = 10$ which give $\hat{\sigma}_n^2 \approx 1.1$ and $\hat{\theta}_n \approx 0.81$. We observe a good match between the histogram and the estimated density, which tends to validate the AR(1) model chosen for the noise correlation across the spectral bands. We also mention that

similar results have been observed for experiments performed on the datasets AIRIS, ROSIS, and HYDICE ([12], [21], [22]), although the results are not included due to lack of space.

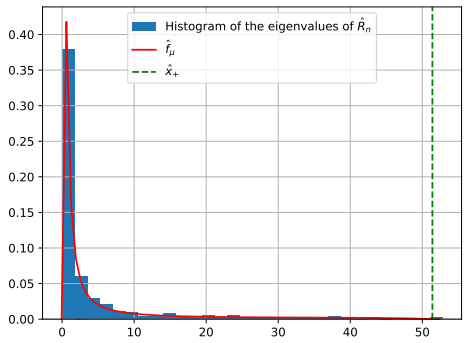

Fig. 4. Histogram of noise eigenvalues of $\hat{\mathbf{R}}_n$ and estimated density f_μ.

We now evaluate the proposed estimator $\hat{k}_n$ given in Proposition 1 and compute in Figure 5 the probability of correct detection $\mathbb{P}(\hat{k}_n = k)$ against the signal-to-noise ratio defined as SNR = $10 \log \frac{1}{\sigma^2}$. For this simulation, we generate semi-synthetic data by selecting $k = 4$ endmembers from the Cuprite hyperspectral scene [13], after removing noisy and water absorption channels ($p = 188$). The fractional abundances and AR(1) noise are then synthesized according to the model (4), with $\boldsymbol{\alpha} = (0.1, 0.15, 0.18, 0.57)^\top$, $\theta = 0.4$ and the sample size $n = 376$.

We also compare the proposed estimator to the one of [7], which is given by

$$\hat{k}_n^{(T)} = \underset{i \in [\![1, k_{max}]\!]}{\operatorname{argmin}} \left(\frac{\lambda_i(\hat{\mathbf{R}}_n)}{\lambda_{i+1}(\hat{\mathbf{R}}_n)} > 1 + \kappa_n \right),$$

and where κ_n is a threshold chosen according to the Fig. 2 given in [7]. In particular, we observe that the proposed estimator performs signficantly better. Indeed, $\hat{k}_n^{(T)}$ tends to underestimate the number of endmembers when γ_i values are close which occurs in the simulation. In this case, by Theorem 1 and the continuity of ϕ, spike positions are also close to each other, leading to an underestimation.

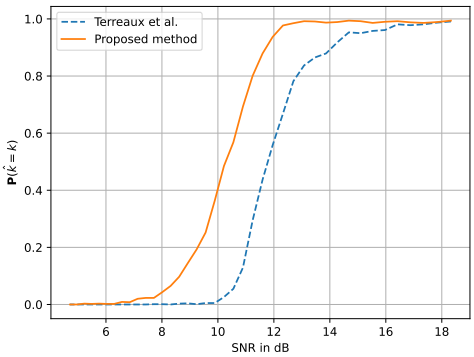

Fig. 5. Probability of correct detection as a function of the SNR. $p = 188$, $c = 0.5$, $\rho = 0.4$, $k = 4$

REFERENCES

[1] J. Harsanyi, W. Farrand, and C. Chang, "Detection of subpixel spectral signatures in hyperspectral image sequences," in *Proceedings of American Society of Photogrammetry & Remote Sensing*, 1994, pp. 236–247.

[2] C. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, 2004.

[3] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, 2008.

[4] C. Chang, "A review of virtual dimensionality for hyperspectral imagery," *IEEE J. Sel. Top. Appl. Earth Obs.*, vol. 11, no. 4, pp. 1285–1305, 2018.

[5] Z. Bai and J.W. Silverstein, *Spectral analysis of large dimensional random matrices*, Springer Verlag, 2010.

[6] K. Cawse-Nicholson, S. Damelin, A. Robin, and M. Sears, "Determining the intrinsic dimension of a hyperspectral image using random matrix theory," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1301–1310, 2012.

[7] E. Terreaux, J.-P. Ovarlez, and F. Pascal, "Anomaly detection and estimation in hyperspectral imaging using random matrix theory tools," in *IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, 2015, pp. 169–172.

[8] A. Halimi, P. Honeine, M. Kharouf, C. Richard, and J.-Y. Tourneret, "Estimating the intrinsic dimension of hyperspectral images using a noise-whitened eigengap approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3811–3821, 2016.

[9] K. Cawse-Nicholson, A. Robin, and M. Sears, "The effect of correlation on determining the intrinsic dimension of a hyperspectral image," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 6, no. 2, pp. 482–487, 2013.

[10] F. Benaych-Georges and R.R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices," *J. Multivariate Anal.*, vol. 111, no. 0, pp. 120–135, 2012.

[11] J. Yao, "A note on a Marchenko-Pastur type theorem for time series," *Stat. Probab. Lett.*, vol. 82, no. 1, pp. 22–28, 2012.

[12] Linda S. Kalman and Edward M. Bassett III, "Classification and material identification in an urban environment using HYDICE hyperspectral data," in *Imaging Spectrometry III*, Michael R. Descour and Sylvia S. Shen, Eds. International Society for Optics and Photonics, 1997, vol. 3118, pp. 57 – 68, SPIE.

[13] NASA Jet Propulsion Laboratory, "Aviris hyperspectral data – cuprite, nevada," Available from NASA AVIRIS Data Portal, 1995, Accessed: 2025-03-04.

[14] H. Jeannin, "Eusipco 2025 code repository," https://gitub.u-bordeaux.fr/hujeannin/eusipco2025.git, 2025.

[15] U. Grenander, *Toeplitz forms and their applications*, AMS Chelsea Publishing Company, 1984, 2nd edition.

[16] J. Silverstein and Z. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivar. Anal.*, vol. 54, no. 2, pp. 175–192, 1995.

[17] J.W. Silverstein and S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," *J. Multivar. Anal.*, vol. 52, no. 2, pp. 175–192, 1995.

[18] M. Rudelson and R. Vershynin, "Hanson-Wright inequality and sub-Gaussian concentration," *Electron. Commun. Probab.*, vol. 18, 2013.

[19] W. Hachem, A. Hardy, and J. Najim, "Large complex correlated Wishart matrices: Fluctuations and asymptotic independence at the edges," *Ann. Probab.*, vol. 44, no. 3, pp. 2264 – 2348, 2016.

[20] Z. Fan and I. Johnstone, "Tracy-Widom at each edge of real covariance and MANOVA estimators," *Ann. Appl. Probab.*, vol. 32, no. 4, pp. 2967, 2022.

[21] Purdue University, "Indian pines hyperspectral dataset," Online, 1992.

[22] Jing Yao, Danfeng Hong, Jocelyn Chanussot, Deyu Meng, Xiaoxiang Zhu, and Zongben Xu, "Dataset: Pavia university dataset," Online, 2024.