Rank detection thresholds for Hankel or Toeplitz data matrices

van der Veen, A.J.; Romme, J.P.A.; Cui, Ye

Publication date
2020

Document Version
Final published version

Published in
28th European Signal Processing Conference (EUSIPCO 2020)

Citation (APA)
van der Veen, A. J., Romme, J. P. A., & Cui, Y. (2020). Rank detection thresholds for Hankel or Toeplitz data matrices. In *28th European Signal Processing Conference (EUSIPCO 2020)* (pp. 1911-1915). Eurasip.

RANK DETECTION THRESHOLDS FOR HANKEL OR TOEPLITZ DATA MATRICES

Alle-Jan van der Veen[†]*, Jac Romme*[‡]*, Ye Cui*[†]

[†] TU Delft, Delft, The Netherlands
[‡] IMEC, Eindhoven, The Netherlands

ABSTRACT

In Principal Component Analysis (PCA), the dimension of the signal subspace is detected by counting the number of eigenvalues of a covariance matrix that are above a threshold. Random matrix theory provides accurate estimates for this threshold if the underlying data matrix has independent identically distributed columns. However, in time series analysis, the underlying data matrix has a Hankel or Toeplitz structure, and the columns are not independent. Using an empirical approach, we observe that the largest eigenvalue is fitted well by a Generalized Extreme Value (GEV) distribution, and we obtain accurate estimates for the thresholds to be used in a sequential rank detection test. In contrast to AIC or MDL, this provides a parameter that controls the probability of false alarm. Also a lower bound is presented for the rank detection rate of threshold-based detection for rank-1 problems.

Index Terms— PCA, structured Wishart matrix, rank detection, Generalized Extreme Value

1. INTRODUCTION

A fundamental problem in array processing is the detection of the number of source signals present in a data matrix. Consider an antenna array with M elements and N temporal samples, resulting in an $M \times N$ data matrix $\mathbf{X}$ and a related covariance matrix $\hat{\mathbf{R}} = (1/N)\mathbf{X}\mathbf{X}^H$. Under narrowband conditions, each source gives a rank-1 contribution to the matrix. If the background noise is white, zero mean, i.i.d. with a Gaussian distribution, then the noise covariance matrix is $\mathbf{R}_n = \sigma_n^2\mathbf{I}$ with a Wishart distribution. To detect the number of sources, one common technique is to count the number of eigenvalues of $\hat{\mathbf{R}}$ that are significantly above σ_n^2. The required threshold depends on M and N and the desired probability of false alarm. We consider the noise to have a complex Gaussian distribution; alternative techniques that use eigenvalues are AIC and MDL [1].

For the rank-0 hypothesis (no sources), we essentially need to know the distribution of the largest eigenvalue of a random matrix with a Wishart distribution. The first results date back a century, and currently we know that this distribution converges to a Tracy-Widow (TW) distribution [2], resulting in accurate thresholds that are straightforward to compute. With some loss of accuracy, these results can be translated to the detection of higher ranks [3].

Subspace-based techniques for e.g. system identification [4], delay estimation [5] or sinusoidal signal detection/harmonic retrieval (e.g., MUSIC, ESPRIT) [6] proceed by forming an $M \times N$ Toeplitz or Hankel data matrix $\mathbf{X}$ with entries $x_{ij} = x[i-j]$, for which $M + N - 1$ samples are needed. In the antenna array context, a similar data repetition technique is known as spatial smoothing (applicable to uniform linear arrays) [7].

As we show in this paper, the existing thresholds for unstructured matrices are not accurate for covariance matrices derived from Toeplitz/Hankel data matrices, as the columns of $\mathbf{X}$ are not independent. As a result, significantly higher thresholds need to be used. Our objective is to derive such thresholds. To this end, we first describe the distribution of the largest eigenvalue of $\hat{\mathbf{R}}$. We will do this using an empirical approach, where observed distributions for the largest eigenvalue from simulations for a range of (M, N) and rank d are parametrically fitted to theoretical distributions. This results in new detection thresholds that are straightforward to compute and use.

2. THRESHOLDS FOR COVARIANCE MATRICES FROM UNSTRUCTURED DATA

We review the results summarized in [3, 8, 9] on bounds of the largest eigenvalue of Wishart matrices. Consider a zero mean complex white Gaussian noise random vector $\mathbf{x}[n]$ of dimension M, with unit variance entries. If we have N samples, the sample covariance matrix is $\hat{\mathbf{R}} = \sum_{n=1}^{N} \mathbf{x}[n]\mathbf{x}[n]^H$, note that the scaling by $1/N$ is omitted to simplify subsequent expressions. The largest eigenvalue of $\hat{\mathbf{R}}$ is denoted by $\hat{\lambda}_1$.

Let $M, N \to \infty$, $\gamma := \frac{M}{N}$ constant and $M \leq N$. Define centering and scaling constants as[1]

$$\mu_{M,N} = (\sqrt{M} + \sqrt{N})^2 = N(1+\sqrt{\gamma})^2\,, \tag{1}$$

$$\sigma_{M,N} = (\sqrt{M} + \sqrt{N})\left(\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right)^{\frac{1}{3}} = N^{\frac{1}{2}} M^{-\frac{1}{6}} (1+\sqrt{\gamma})^{\frac{4}{3}}$$

Then [10]

$$s := \frac{\hat{\lambda}_1 - \mu}{\sigma} \quad \xrightarrow{\mathcal{D}} \quad F_2(s)$$

where $F_2(s)$ is the TW distribution of order 2 [2]. A useful approximation of this distribution in terms of a Gamma distribution is given in [8], and we use this approximation in subsequent plots.

For the largest eigenvalue, we are interested in a threshold u for which the probability that $\hat{\lambda}_1 > u$ is $1-p$, a desired false alarm rate (p is the confidence level). Compute $f_2 = F_2^{-1}(p)$ (where F_2^{-1} denotes the inverse Cumulative Distribution Function (CDF); with the approximation in [8], it can be found from an inverse Gamma distribution), and set

$$u = \mu + \sigma f_2\,.$$

Then

$$\begin{aligned} \mathrm{P}(\hat{\lambda}_1 > u) = 1-p \quad &\Leftrightarrow \quad \mathrm{P}\big(\frac{\hat{\lambda}_1 - \mu}{\sigma} < \frac{u-\mu}{\sigma}\big) = p \\ &\Leftrightarrow \quad \frac{u-\mu}{\sigma} = F_2^{-1}(p) \end{aligned}$$

so that, indeed, u is the required threshold.

[1] For real-valued matrices, slight modifications are needed [8].

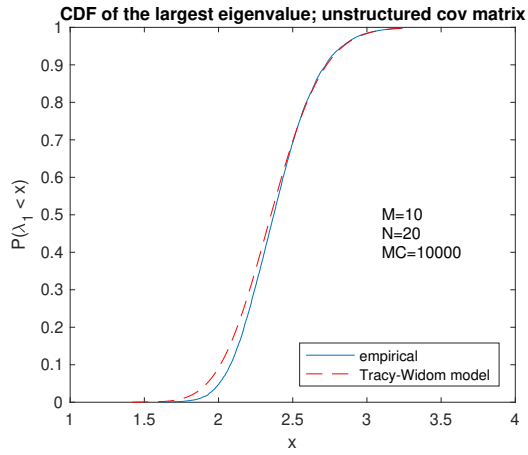

Fig. 1. CDF of the largest eigenvalue of an unstructured covariance matrix. $M = 10$, $N = 20$. Here and in subsequent plots, the covariance was scaled by $1/N$.

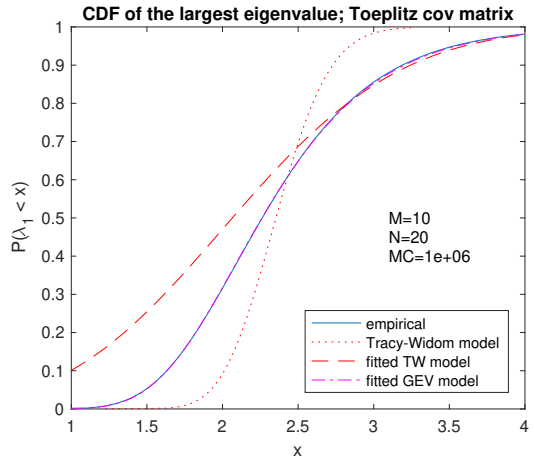

Fig. 2. CDF of the largest eigenvalue of a Toeplitz-structured covariance matrix, compared to the fitted TW distribution and the GEV distribution. $M = 10$, $N = 20$.

Fig. 1 shows a simulation where an empirical CDF $\hat{F}_\lambda(x)$ for $\hat{\lambda}_1$ is obtained using Monte-Carlo (MC) runs, and compared to its model, $F_2(\frac{x-\mu}{\sigma})$. It is seen that the fit is tight; the deviation for low probabilities is due to the approximation [8].

3. THRESHOLDS FOR COVARIANCE MATRICES FROM HANKEL-STRUCTURED DATA MATRICES

3.1. Proposed distribution

Moving now to Hankel-structured data matrices, consider a zero mean complex white Gaussian noise random vector $\mathbf{x}$ with unit variance. Stack $M + N - 1$ samples into an $M \times N$ Hankel matrix $\mathbf{X} = (x[i + j])_{i,j}$. The sample covariance matrix is $\hat{\mathbf{R}} = \mathbf{X}\mathbf{X}^H$ (omitting the scaling by $1/N$), for large N it converges to a Toeplitz matrix.

Fig. 2 shows a simulation where the empirical CDF of the largest eigenvalue (solid line) is compared to the same TW model as before, $F_2(\frac{x-\mu}{\sigma})$ (dotted line). It is seen that in this case the fit is poor. The reason is that the columns of $\mathbf{X}$ are not independent. Selecting new values for μ and σ to fit the right tail (dashed line) shows that the TW distribution is not a good model.

To obtain a better approximation, we use the matlab `fitdist` function to fit to a GEV distribution, it is seen in Fig. 2 that this fit is excellent (dash-dot line, overlaps completely the empirical curve). The CDF of the GEV distribution is given by [11]

$$F(s;\xi) = e^{-(1+\xi s)^{-1/\xi}}, \quad \text{where} \quad s = \frac{x-\mu}{\sigma}. \tag{2}$$

Other values for centering μ and scale σ are needed, and the GEV distribution has an additional shape parameter ξ that essentially controls the slope of the right tail of the distribution (which is of interest to us as the threshold for small false alarm rates is found here).

Although we make no theoretical claims, a fitting to the GEV distribution is motivated as it is the limit distribution of the maximum of a sequence of independent and identically distributed random variables [11], in this case the eigenvalues. This would require M and N to be sufficiently large.

From (2), the GEV distribution has a closed-form expression for the inverse CDF,

$$Q(p;\mu,\sigma,\xi) = \mu + \sigma\frac{(-\ln(p))^{-\xi} - 1}{\xi}. \tag{3}$$

Thus, given values for μ, σ, ξ and a desired false alarm rate $P_{FA} = 1 - p$, we can easily compute the required threshold.

3.2. GEV distribution parameter estimates

We obtained best-fitting parameters for (μ, σ, ξ) for a range of M, N. Although these could be used in a table look-up, it is often more convenient to have "friendly" polynomial expressions. Using an empirical approach, we obtained the following approximations:

$$\begin{aligned}
\hat{\mu} &= N\big(a_1 MN^{a_5} + a_2 M^{1/2}N^{-1/2} + a_3 N^{a_6} + a_4\big)\\
\hat{\sigma} &= N^{1/2}M^{-1/6}\big(b_1 MN^{-0.1138} + b_2 M^{1/2}N^{-0.2721}\\
&\qquad + b_3 N^{-0.6820} + b_4\big)\\
\hat{\xi} &= c_1 M^{-0.2}\exp\big(c_2\frac{M}{N}\big)
\end{aligned} \tag{4}$$

$$\begin{aligned}
\mathbf{a} &= [0.4003 \;\; 1.5601 \;\; -3.8142 \;\; 0.9748 \;\; -0.6521 \;\; -0.6391]\\
\mathbf{b} &= [0.1888 \;\; 1.5819 \;\; -5.6885 \;\; 0.6129]\\
\mathbf{c} &= [-0.2189 \;\; -3.1877]\,.
\end{aligned}$$

Starting from (1), these expressions were obtained by fitting the coefficients of similar polynomials in M and N, over a range of values: $M = 2, \cdots, 120$ and $N = 20, \cdots, 500$, with $M < N$. Matlab `fminsearch` was used for the fitting, and in total 1110 pairs (M, N) were used.

To test the resulting model (3), (4) for the CDF, we took a desired $P_{FA} = 0.01$, computed the threshold based on the model, and evaluated the resulting P_{FA}. The results for a range of M and N are shown in Fig. 3. The computed thresholds are decently accurate for $M > 3$: the effective P_{FA} is mostly between 0.09 and 0.011 (with mean = 0.010, std = 0.0011 over the range $M = 4, \cdots, 120$, $N = 20, \cdots, 500$, $M < N$, total 1030 pairs (M, N)).

4. SEQUENTIAL RANK DETECTION

We now consider a data model $\mathbf{X} = \mathbf{X}_s + \mathbf{N}$, where $\mathbf{X}_s$ is a rank-d deterministic signal matrix, and $\mathbf{N}$ is a noise matrix as before. To detect the number of signals from the corresponding covariance matrix

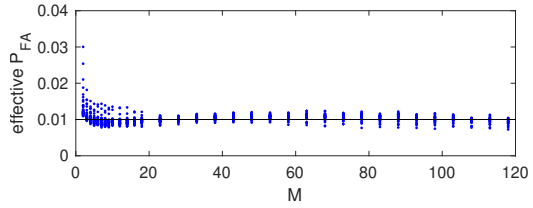

Fig. 3. Effective P_{FA} when the desired $P_{FA} = 0.01$ is inserted into the model (3), (4), for a range of M and N. Each dot represents the result for a specific (M, N). The apparent outliers for $4 < M < 20$ occur for $N = 300, 400, 500$.

$\hat{\mathbf{R}}$, we select a threshold and compute the number of eigenvalues $\hat{\lambda}_i$ that are larger than it. The threshold should be selected such that a certain false alarm rate P_{FA} is obtained. In more detail, we consider the "null hypothesis" $\mathcal{H}_k$: the signal rank is k, and we test it against $\mathcal{H}'_k$: the signal rank is larger than k. The threshold T_k depends on k and defines $P_{FA}(k)$. In a sequential test, we start with $k = 0$, and if we accept $\mathcal{H}'_0$, increase k and test again until we accept $\mathcal{H}_k$.

Input: $\{\hat{\lambda}_i\}$,M, N, P_{FA}
Output: $\hat{d}$, the estimated signal rank
$\hat{d} = 0$;
repeat
  $k = \hat{d}$;
  $T_k = \text{tresh}(k, M, N, P_{FA})$; `// see eqn. (6)`
  $\hat{d} = \#\{\hat{\lambda}_i > T_k\}$;
until $\hat{d} = k$;

Assuming the tests are independent (not quite true), the overall false alarm rate is $P_{FA} = P_{FA}(0)(1 + P_{FA}(1))(\cdots) \approx P_{FA}(0)$. The probability of correctly detecting the signal rank depends on the SNR, but will not be larger than $1-P_{FA}(k)$, since we only stop once the null hypothesis $\mathcal{H}_k$ is accepted. Thus, the test has a tendency to slightly overestimate the signal rank.

It remains to determine the value of the kth threshold T_k, under the hypothesis $\mathcal{H}_k$, to achieve a desired $P_{FA}(k)$. For the unstructured case, [3, prop.1.2] states that the eigenvalues of the noise subspace of a covariance matrix with a rank-k signal subspace is dominated by the distribution of a noise-only covariance matrix of dimension $M - k$ by N. This result can be made slightly more precise: also N has to be replaced by $N - k$:

Proposition 1. *Let $\mathbf{X} = \mathbf{X}_s + \mathbf{N}$, and suppose that the rank of $\mathbf{X}_s$ is k and that the non-zero singular values of $\mathbf{X}_s$ are much larger than those of $\mathbf{N}$. Then the noise eigenvalues of $\hat{\mathbf{R}}$ converge to the distribution of a noise-only covariance matrix of size $M-k$ by $N-k$.*

Proof. Let $\mathbf{X}_s = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^H$ be the 'economy-size' SVD of $\mathbf{X}_s$ (i.e., $\mathbf{U}_s$ and $\mathbf{V}_s$ have k columns), and complete $\mathbf{U}_s$ and $\mathbf{V}_s$ to square unitary matrices $\mathbf{U}$, $\mathbf{V}$. If the singular values of $\mathbf{X}_s$ are sufficiently large, then $\mathbf{N}$ will hardly disturb the directions of the corresponding singular vectors, and these are the dominant singular vectors of $\mathbf{X}$. Consider $\mathbf{X}' := \mathbf{U}^H\mathbf{X}\mathbf{V}$. With the same arguments as in the recursive definition of the SVD [12], we find

$$\mathbf{X}' = \begin{bmatrix} \mathbf{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{N}' \end{bmatrix}.$$

$\mathbf{N}'$ has size $M - k \times N - k$, while the rotations do not disturb the distribution of the entries, i.e., $\mathbf{N}'$ has zero-mean i.i.d. Gaussian entries. As a result, the distribution of the $k+1$-st eigenvalue of $\hat{\mathbf{R}}$ is equal to the distribution of the largest eigenvalue of a noise-only matrix of size $M - k$ by $N - k$. □

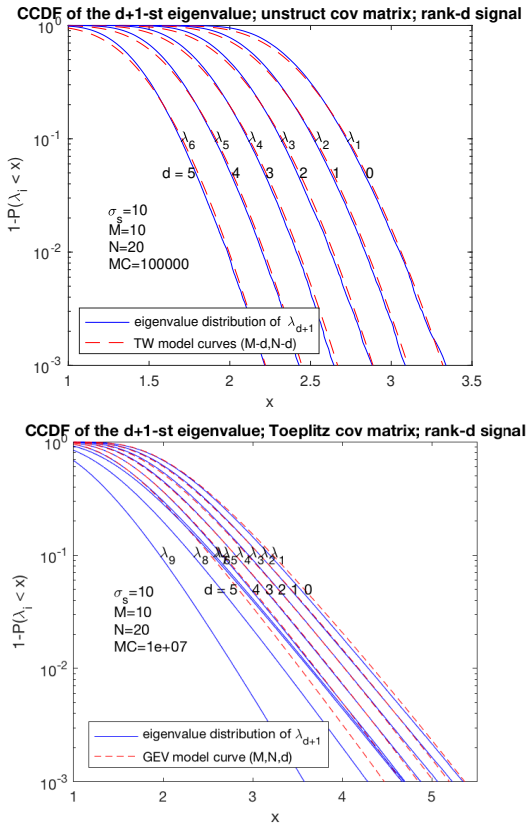

Fig. 4. CCDF of the $d+1$-st eigenvalue of a covariance matrix with d signal components, compared to the corresponding model with parameters $(M-d, N-d)$, for various d. (a) Unstructured data and TW model; (b) Hankel-structured data and GEV model $(M = 10, N = 20)$.

Following this proposition, the threshold for detection of rank k in the sequential detection algorithm should be based on Eqn. (1), but using $(M-k, N-k)$ as parameters. To demonstrate this, Fig. 4(a) shows the empirical Complementary CDF (CCDF) of the $d+1$-st eigenvalue of the covariance matrix of a data matrix consisting of a rank-d source matrix in white noise, compared to the theoretical TW curve with parameters $(M-d, N-d)$. It is seen that the fit is excellent.

Unfortunately, this proof does not carry over to a Hankel-structured data matrix with a rank-d signal component. Indeed, simulations indicate that in this case the distribution of λ_{d+1} does not follow the distribution of the largest eigenvalue of an $(M-d) \times (N-d)$ size noise-only matrix. The GEV model still fits, but the parameters depend on d in a more complicated way. Fig. 4(b) (solid lines) shows the same curves as Fig. 4(a), but for Hankel-structured data. For d not too large, say $d < \frac{1}{2}M$, the curves are regularly spaced and have a constant slope as function of d. For $d \approx \frac{1}{2}M$, the curves become very close, and for larger d, the slopes get much more steep and the curves are widely spaced.

To capture the more regular behavior for $d < \frac{1}{2}M$ in a model, we took the previous model and inserted modifications for d, for

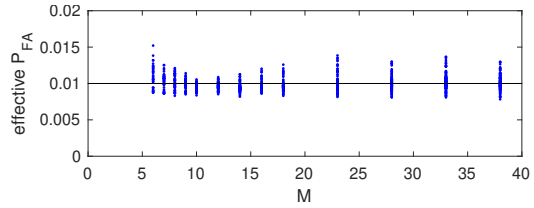

Fig. 5. Effective P_{FA} when the desired $P_{FA} = 0.01$ is inserted into the model (6), (5), for a range of M, N and d. Each dot represents the result for a specific (M, N, d).

each M and N in the model. We derived approximations based on M ranging from 6 to 38, N from 15 to 100 in steps of 5, and d from 0 to 20, with $d < M/2$ and $M < N$. In total 1110 data points (M, N, d) were generated; the empirical distributions were obtained using 5×10^5 Monte Carlo runs. The resulting model is as follows:

$$\begin{aligned}
\hat{\mu} &= N\big(a_1(M - d_1k)(N - d_3k)^{a_5} + \\
&\quad a_2(M - d_2k)^{1/2}(N - d_4k)^{-1/2} + a_3(N - d_5k)^{a_6} + a_4\big) \\
\hat{\sigma} &= N^{1/2}M^{-1/6}\big(b_1(M - e_1k)N^{-0.1138} + \\
&\quad b_2(M - e_2k)^{1/2}N^{-0.2721} + b_3N^{-0.6820} + b_4\big) \\
\hat{\xi} &= c_1(M - f_1k)^{-0.2}\exp\big(c_2\frac{M - f_2k}{N - f_3k}\big) \qquad (5)
\end{aligned}$$

$\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ as in (4)

$\mathbf{d} = [-0.0982 \;\; 1.2442 \;\; 0.2531 \;\; -2.0890 \;\; 0.6487]$

$\mathbf{e} = [-0.3360 \;\; 1.0001]$

$\mathbf{f} = [-0.3424 \;\; -0.0060 \;\; 0.5902]$.

In summary, we propose to take the value of the kth threshold in the sequential detection algorithm as

$$T_k = Q(1 - P_{FA}; \hat{\mu}, \hat{\sigma}, \hat{\xi}), \quad \text{for } k < \tfrac{1}{2}M \qquad (6)$$

where Q is shown in (3) and the values of the parameters are defined in (5). Note that unnormalized covariances were assumed ($\hat{\mathbf{R}} = \mathbf{X}\mathbf{X}^H$); otherwise, adjust $\hat{\mu}, \hat{\sigma}$ accordingly. For $d \geq \frac{1}{2}M$, we currently do not have an accurate model.

The dashed lines in Fig. 4(b) shows the model fit for $M = 10$, $N = 20$ and a range of d. For $d < \frac{1}{2}M$, the fit is very good.

To demonstrate the accuracy of the model over the range of (M, N, d) for which the model was fitted, the effective P_{FA} for a requested $P_{FA} = 0.01$ is shown in Fig. 5. The fit is of similar quality as for Fig. 3. The mean of the effective P_{FA} is 0.0100 and the standard deviation is 0.0010, as estimated over 1110 data points. This accuracy should be adequate in practice.

5. SINGLE SOURCE DETECTION RATE

Apart from the false alarm rate, the probability of detection P_D is also of interest. A detection is called successful if $\hat{d} = d$. For simplicity, we focus on the rank-1 case. In this case, the matrix $\mathbf{X}$ is a superposition of the (deterministic) signal matrix $\mathbf{X}_s$ and $\mathbf{N}$, where $\mathbf{X}_s = \mathbf{u}_s\sqrt{N\lambda_s}\mathbf{v}_s^H$. To compute the detection rate, the CDF of the largest eigenvalue $\hat{\lambda}_1$ of $(1/N)\mathbf{X}\mathbf{X}^H$ is required, and such a result is (to our knowledge) unknown. Let us define

$$L_1 = \frac{1}{N}\mathbf{u}_s^H\mathbf{X}\mathbf{X}^H\mathbf{u}_s, \qquad (7)$$

such that $\hat{\lambda}_1 \geq L_1$ and $\mathrm{P}(\hat{\lambda}_1 > T_1) \geq \mathrm{P}(L_1 > T_1)$. We will derive the CDF of L_1, as a function of $\mathbf{X}_s$. Thus write L_1 as

$$L_1 = \frac{1}{N}\mathbf{u}_s^H\mathbf{X}\mathbf{X}^H\mathbf{u}_s = \frac{1}{N}\mathbf{t}^H\mathbf{t} \qquad (8)$$

where $\mathbf{t} := \mathbf{X}^H\mathbf{u}_s = \sqrt{N\lambda_s}\mathbf{v}_s + \mathbf{t}_n$ with $\mathbf{t}_n = \mathbf{N}^H\mathbf{u}_s$. The entries of $\mathbf{t}_n$ are linear combinations of Gaussian Random Variables (RVs), therefore they are Gaussian distributed as well, but colored due to the Hankel-structure of $\mathbf{N}$. Thus pose

$$\mathbf{t} \sim \mathcal{CN}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) \qquad (9)$$

where $\boldsymbol{\mu}_t = \sqrt{N\lambda_s}\mathbf{v}_s$, and $\boldsymbol{\Sigma}_t$ is calculated below. As a result, L_1 is the Hermitian quadratic form of a colored, complex Gaussian random vector, which can be reformulated as a linear combination of N independent non-central χ^2 distributed RVs [13]. In other words,

$$L_1 = \sum_{k=1}^{N}\alpha_k X_k \qquad (10)$$

where X_k is an independent χ^2 distributed RV with 2 degrees of freedom (for complex data) and non-centrality parameter β_k^2, while the weight of X_k is denoted by α_k. Let $\boldsymbol{\Sigma}_t = \mathbf{C}\mathbf{D}\mathbf{C}^H$ be an eigenvalue decomposition of $\boldsymbol{\Sigma}_t$, then [13]

$$\boldsymbol{\alpha} = \tfrac{1}{2}\mathrm{diag}(\mathbf{D}), \qquad \boldsymbol{\beta} = \sqrt{\frac{2}{N}}\mathbf{D}^{-\frac{1}{2}}\mathbf{C}^H\boldsymbol{\mu}_t\,. \qquad (11)$$

To compute $\boldsymbol{\Sigma}_t$ for some $\mathbf{u}_s$, we use that the random vector $\mathbf{t}_n = \mathbf{N}^H\mathbf{u}_s$ (where $\mathbf{N}$ has a Hankel structure) can be viewed as a convolution/filtering operation in matrix notation. Equivalently, we can view $\mathbf{t}_n$ as the output of a filter with coefficients $\mathbf{u}_s$ and input a complex white Gaussian noise signal with unit variance. It follows that $\boldsymbol{\Sigma}_t$ can be written as

$$\boldsymbol{\Sigma}_t[k, l] = \begin{cases} \mathbf{u}_s^H\mathbf{Z}^{k-l}\mathbf{u}_s\,, & k > l \\ \mathbf{u}_s^H\left(\mathbf{Z}^T\right)^{k-l}\mathbf{u}_s\,, & \text{otherwise} \end{cases} \qquad (12)$$

where $\mathbf{Z}$ is a shift (delay) matrix, containing ones on the first diagonal below the main diagonal and zeros otherwise.

At this point, all ingredients are available to compute the closed-form CDF of L_1 using e.g. [14]. The rank-1 detection rate P_D is now simply the probability that $\hat{\lambda}_1$ exceeds the respective threshold, while $\hat{\lambda}_2$ remains smaller than the threshold, i.e.

$$P_D \geq \mathrm{P}\{L_1 > T_1, \hat{\lambda}_2 < T_2\} \approx \mathrm{P}\{L_1 > T_1\}(1 - P_{FA}) \qquad (13)$$

where for the approximation we have assumed independence, which is strictly speaking not true.

To verify the modeling, we show in Fig. 6 the empirical CCDF of λ_1 and L_1, and the theoretically derived CCDF of L_1, for a single, unit-circle, complex exponential source in the presence of noise.[2] The simulations show a perfect fit between the model and the empirical distribution of L_1. The CCDF of L_1 is indeed a lower bound for the CCDF of λ_1 and their convergence with increasing SNR and x is confirmed. Therefore, we expect the theoretical detection rate to be a conservative lower bound for the empirical detection rate, converging with increasing SNR and decreasing P_{FA}.

[2]For higher signal ranks, the model provides accurate results if the signal eigenvalues are well separated.

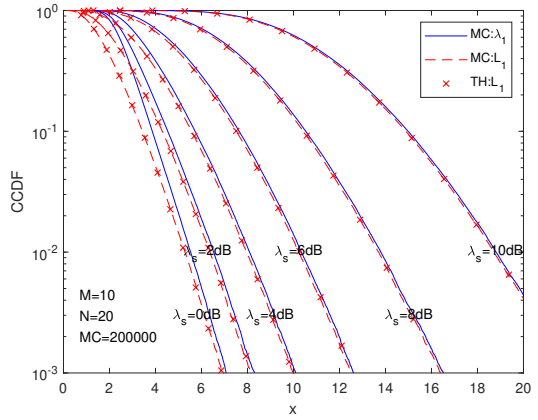

Fig. 6. Comparison between empirical (MC) and theoretical (TH) CCDF of λ_1 and L_1.

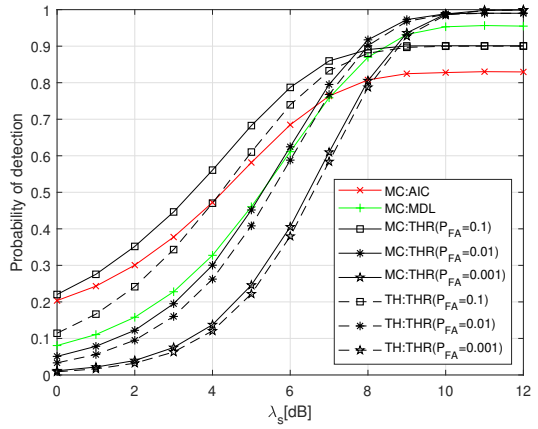

Fig. 7. Performance comparison between different rank estimation detection algorithms in the presence of noise. $M = 10, N = 20$.

6. SIMULATIONS–COMPARISON TO AIC AND MDL

In this section, we compare the detection rate of the threshold-based rank detector with the widely-used AIC [15, Eqn. (7.506)] and MDL [15, Eqn. (7.508)] for a rank-1 situation, as used in the previous section, with varying SNR, or equivalently λ_s.

In Fig. 7, we observe at high SNR that MDL has a 0.045 probability of incorrect estimating the signal-rank compared to 0.17 for AIC. On the other hand, AIC has a higher detection rate at low SNR. Both observations are in line with [15] and illustrate the inherent tradeoff in performance in low SNR vs high SNR regimes. Using the proposed threshold-based rank detector (THR), a similar tradeoff can be made simply by changing the (desired) false alarm rate P_{FA}. For $P_{FA} = 0.01$, approximately the same detection rate as MDL is obtained at lower SNR, but the detection rate is improved at high SNR. When the false alarm rate is increased ($P_{FA} = 0.1$), the detection rate at low SNR is slightly better than AIC, but with improved detection rate at high SNR. For very small false alarm rates ($P_{FA} = 0.001$), the performance at low SNR is reduced, but with almost perfect detection at high SNR.

7. CONCLUSIONS

We proposed rank detection thresholds valid for Hankel or Toeplitz data matrices, as occurs in time series analysis, delay estimation, or harmonic retrieval. The thresholds are straightforward to evaluate in closed form, and are accurate for ranks up to half the matrix size. Compared to AIC and MDL, the detection performance is improved, with the benefit of a parameter that controls the false alarm rate. On the side, we improved on a result on the distribution of the $d + 1$st eigenvalue of a rank-d covariance matrix in the unstructured case.

8. REFERENCES

[1] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Tr. Acoustics, Speech, Signal Proc.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.

[2] C. A. Tracy and H. Widom, "Level-spacing distributions and the Airy kernel," *Communications in Mathematical Physics*, vol. 159, pp. 151–174, 1994.

[3] I. M. Johnstone, "On the distribution of the largest eigenvalue in Principal Components Analysis," *The Annals of Statistics*, vol. 29, no. 2, pp. 295–327, 2001.

[4] P. Van Overschee and B. D. Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75 – 93, 1994.

[5] A. J. van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Tr. Signal Proc.*, vol. 46, no. 2, pp. 405–418, Feb. 1998.

[6] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, Inc., New York, NY, USA, 1996.

[7] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Tr. Acoustics, Speech, Signal Proc.*, vol. 33, no. 4, pp. 806–811, Aug. 1985.

[8] M. Chiani, "Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy–Widom distribution," *Journal of Multivariate Analysis*, vol. 129, pp. 69–81, 2014.

[9] I. Johnstone and D. Paul, "PCA in high dimensions: An orientation," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1277–1292, Aug. 2018.

[10] K. Johansson, "Shape fluctuations and random matrices," *Communications in Mathematical Physics*, vol. 209, pp. 437–476, 2000.

[11] M. Charras-Garrido and P. Lezaud, "Extreme value analysis: an introduction," *Journal de la Société Française de Statistique*, vol. 154, no. 2, pp. 66–97, 2013.

[12] G. Golub and C. van Loan, *Matrix Computations*, Johns Hopkins University Press, 1986.

[13] H.-T. Ha and S. B. Provost, "An accurate approximation to the distribution of a linear combination of non-central chi-square random variables," *REVSTAT Statistical Journal*, vol. 11, no. 3, pp. 231–254, November 2013.

[14] R. B. Davies, "The distribution of a linear combination of chi-square random variables," *Applied Statistics*, vol. 29, pp. 69–81, 1980.

[15] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, Wiley, 2002.