

# Nonparametric Detection of Signals by Information Theoretic Criteria: Performance Analysis and an Improved Estimator

Boaz Nadler

**Abstract**—Determining the number of sources from observed data is a fundamental problem in many scientific fields. In this paper we consider the nonparametric setting, and focus on the detection performance of two popular estimators based on information theoretic criteria, the Akaike information criterion (AIC) and minimum description length (MDL). We present three contributions on this subject. First, we derive a new expression for the detection performance of the MDL estimator, which exhibits a much closer fit to simulations in comparison to previous formulas. Second, we present a random matrix theory viewpoint of the performance of the AIC estimator, including approximate analytical formulas for its overestimation probability. Finally, we show that a small increase in the penalty term of AIC leads to an estimator with a very good detection performance and a negligible overestimation probability.

**Index Terms**—Information theoretic criteria, performance analysis, random matrix theory, source enumeration.

## I. INTRODUCTION

**D**ETECTION of the number of sources measured by an array of passive sensors is a fundamental problem in statistical signal processing, which has received considerable attention in the past 20 years, see [3], [5], [24], [27], [29], [32], and [33]. In the nonparametric setting, where no assumptions on the array manifold or on the desired signal waveforms are made, two of the most common estimators for this problem are the AIC and the MDL estimators, both derived from information theoretic considerations [29]. Information theoretic criteria have also been used to derive estimators in parametric settings, where the array manifold is assumed known, as in [28] and [30], or when other information is available, such as explicit knowledge of the waveform of the desired signal or some shift invariance properties of the sensor array response, as in [9] and [10]. Most of these estimators require the eigendecomposition of the sample covariance matrix. We note that in certain parametric settings, this step can be avoided, leading to more computationally efficient estimators, see, for example, [11].

In this paper, we focus on the detection performance of the original AIC and MDL estimators as proposed in [29]. In particular, we analyze the performance of these estimators in the nonparametric setting, where no assumptions are made neither on

the array manifold, nor on the desired signal waveforms. While some of the techniques employed in our analysis could potentially be used to analyze more sophisticated estimators, such analysis is beyond the scope of the current paper.

### A. Problem Formulation

We consider the standard model for signals impinging on an array of  $p$  sensors,

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \sigma\mathbf{n} \quad (1)$$

where  $\mathbf{x} \in \mathbb{C}^p$  is the complex observation vector received at the  $p$  sensors,  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_q]$  is the steering matrix, composed of  $q$  linearly independent vectors,  $\mathbf{s}$  is a  $q \times 1$  vector containing the zero mean random signals and  $\mathbf{n}$  is a  $p$ -dimensional complex Gaussian white noise vector. The parameter  $\sigma^2$  is the unknown noise variance.

We assume the signals have a full rank covariance matrix (e.g., there are no fully correlated signals), and denote by  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q > 0$  the  $q$  noise-free population signal eigenvalues. That is, the eigenvalues of the population covariance matrix of  $\mathbf{x}$  are

$$(\lambda_1 + \sigma^2, \dots, \lambda_q + \sigma^2, \sigma^2, \dots, \sigma^2). \quad (2)$$

For simplicity of the analysis and without loss of generality, we assume that  $\sigma = 1$  for the rest of this paper. Furthermore, for the analysis of the MDL estimator we assume that the smallest signal eigenvalue has multiplicity one ( $\sigma^2 < \lambda_q + \sigma^2 < \lambda_{q-1} + \sigma^2$ , or equivalently  $0 < \lambda_q < \lambda_{q-1}$ ), and as described explicitly below, it is well separated from the next eigenvalue. The smallest signal eigenvalue will have multiplicity one, for example, whenever the emitting sources are at markedly different distances from the sensors. Our analysis can be generalized to the case of multiplicity larger than one, or to two signal eigenvalues which are not well separated, but these extensions are beyond the scope of this paper.

Let  $\mathbf{x}_1, \dots, \mathbf{x}_n$  denote  $n$  i.i.d. observations from the model (1). The problem of interest is to determine the number of sources  $q$ . In the nonparametric setting, where no knowledge of the desired signal waveforms is assumed, and no assumptions on the steering matrix are made beyond it being of rank  $q$ , most methods to perform this task use the eigenvalues  $\ell_1 \geq \ell_2 \geq \dots \geq \ell_p$  of the sample covariance matrix  $\mathbf{S}_n = \sum_i \mathbf{x}_i \mathbf{x}_i^H$  of the  $n$  observations, see [31] for a review. Some of the most common nonparametric estimators for the number of signals are based on information theoretic criteria

Manuscript received January 22, 2009; revised December 29, 2009. Date of publication February 05, 2010; date of current version April 14, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xavier Mestre.

The author is with the Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, 76100, Israel (e-mail: boaz.nadler@weizmann.ac.il).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2010.2042481

[29], [32]. In this paper, we consider the MDL and AIC estimators, given by

$$\begin{aligned} \hat{k}_{\text{MDL}} &= \arg \min_k \text{MDL}(k) \\ &= \arg \min_k \left( -\log \mathcal{L}_k + \frac{k(2p-k)}{2n} \log n \right) \end{aligned} \quad (3)$$

and

$$\begin{aligned} \hat{k}_{\text{AIC}} &= \arg \min_k \text{AIC}(k) \\ &= \arg \min_k \left( -\log \mathcal{L}_k + \frac{k(2p-k)}{n} \right) \end{aligned} \quad (4)$$

where

$$\mathcal{L}_k = \frac{\prod_{j=k+1}^p \ell_j}{\left( \frac{\sum_{j=k+1}^p \ell_j}{(p-k)} \right)^{p-k}}$$

Note that these two estimators differ only by their penalty terms  $k(2p-k)/n$  for AIC, and  $k(2p-k) \log n/2n$  for MDL. Also note that our definition of  $\text{MDL}(k)$  and  $\text{AIC}(k)$  is slightly different from the standard notation in the literature, as for convenience later on, we divide these terms by  $n$  (which does not change the estimate  $\hat{k}$  of the number of sources).

The main focus of this paper is on the performance analysis of these two estimators

$$\Pr[\hat{k}_{\text{MDL}} \neq q] \quad \text{and} \quad \Pr[\hat{k}_{\text{AIC}} \neq q].$$

In particular, we focus on the *underestimation* probability of MDL and *overestimation* probability of AIC by exactly one signal, as these are the leading terms governing their respective detection performances for large sample sizes.

### B. Previous Work

Various works analyzed the performance of these two estimators, see [7], [20], [34], and [35]. The MDL estimator was proven to be strongly consistent [36]

$$\lim_{n \rightarrow \infty} \Pr[\hat{k}_{\text{MDL}} = q] = 1.$$

For finite sample sizes, it was observed empirically that the main source of error in the MDL estimator is *underestimation* of the number of signals by exactly one signal. Following this observation, both [7] and [35] studied the properties of

$$\begin{aligned} \Delta_{\text{MDL}} &= \text{MDL}(q-1) - \text{MDL}(q) \\ &= -\log \ell_q - (p-q) \log \left( \frac{\sum_{i=q+1}^p \ell_i}{p-q} \right) \\ &\quad + (p+1-q) \log \left( \frac{\sum_{i=q}^p \ell_i}{p+1-q} \right) \\ &\quad - \frac{2p-2q+1}{2} \frac{\log n}{n}. \end{aligned} \quad (5)$$

Asymptotically as  $n \rightarrow \infty$  this random variable follows a Gaussian distribution with some mean  $\mu_{\text{MDL}}$  and standard deviation  $\sigma_{\text{MDL}}$ , where both of these quantities depend on the parameters  $p, n$ , and  $\lambda_q$ . It readily follows that

$$\Pr[\hat{k}_{\text{MDL}} \neq q] \approx \Pr[\Delta_{\text{MDL}} < 0] = 1 - \phi \left( -\frac{\mu_{\text{MDL}}}{\sigma_{\text{MDL}}} \right) \quad (6)$$

where  $\phi(z) = 1/\sqrt{2\pi} \int_{-\infty}^z e^{-x^2/2} dx$  is the cumulative distribution function of a standard Gaussian. Both references derived the following asymptotic approximations for the parameters  $\mu_{\text{MDL}}$  and  $\sigma_{\text{MDL}}$

$$\begin{aligned} \mu_{\text{MDL}} &= -\log(1 + \lambda_q) + (p-q+1) \log \left( 1 + \frac{\lambda_q}{p-q+1} \right) \\ &\quad - \frac{2p-2q+1}{2} \frac{\log n}{n} \end{aligned} \quad (7)$$

$$\sigma_{\text{MDL}}^2 = \frac{1}{n} \left( 1 + \frac{1}{p-q} \right) \frac{(p-q)^2 \lambda_q^2}{(\lambda_q + p - q + 1)^2}. \quad (8)$$

In [35, Fig. 4] and in [7, Fig. 2] it is suggested that these formulas are in very close agreement with simulation results.

The performance of the AIC estimator was also studied in various works, see [15], [35] and references therein. It was shown that its main source of error is model order *overestimation* by exactly one signal

$$\Pr[\hat{k}_{\text{AIC}} \neq q] \approx \Pr[\Delta_{\text{AIC}} > 0] \quad (9)$$

where  $\Delta_{\text{AIC}} = \text{AIC}(q) - \text{AIC}(q+1)$  is given by

$$\begin{aligned} \Delta_{\text{AIC}} &= -\log(\ell_{q+1}) + (p-q) \log \left( \frac{T_{p-q}}{p-q} \right) \\ &\quad - (p-q-1) \log \left( \frac{T_{p-q-1}}{p-q-1} \right) - \frac{2p-2q-1}{n} \end{aligned} \quad (10)$$

and  $T_{p-k} = \sum_{j=k+1}^p \ell_j$ . In [34] and [35], expressions for this overestimation probability were derived, which require numerical evaluation of a possibly high dimensional integral (although [34] also presented approximations to these high dimensional integrations). Both simulations and theory show that for number of sensors  $p \approx 10$  the overestimation probability of AIC for complex valued signals is nonnegligible, of the order of 5%–10%.

## II. MAIN RESULTS

In this paper we present three main contributions related to the AIC and MDL estimators. First, regarding the MDL estimator, we show that despite the above results, (6) for MDL's detection performance with  $\mu_{\text{MDL}}$  given by (7) is not as accurate as claimed. We explain the source of this discrepancy and derive a new more accurate formula which provides a much better fit to simulations. Our first main result is the following lemma.

*Lemma 1:* Assume  $\lambda_q$  has multiplicity one,  $\lambda_{q-1} \gg 1$  and  $(\lambda_{q-1} - \lambda_q)/(1 + \lambda_q) \gg \sqrt{1/n}$ . Then, an accurate expression for the detection performance of the MDL estimator is given by (6), but with a modified expression for  $\mu_{\text{MDL}}$  as follows [see

(11) at the bottom of the page], where  $\beta = 1$  for real valued signals and noise, and  $\beta = 2$  in the complex valued case.

Note that the difference between (7) and the new formula (11) are the additional  $O(1/n)$  terms, neglected in the calculation of the expectation  $\mu_{\text{MDL}} = \mathbb{E}[\Delta_{\text{MDL}}]$  in previous works. We note that independent of our work, a similar observation was made recently in [8]. While the underlying analysis is similar, our results are different as we obtain an explicit formula for the correction terms, (11), whereas [8] require the numerical solution of some nonlinear equations.

We now show that for practical values of  $p$ ,  $n$  and signal eigenvalues, these  $O(1/n)$  terms are nonnegligible. To this end, assume  $n$  is large and perform a Taylor expansion of the logarithms in (11). This gives, up to  $o(1/n)$  terms

$$\mu'_{\text{MDL}} - \mu_{\text{MDL}} = \frac{p-q + \frac{1}{\beta}}{n} + \frac{1}{n\beta} \left( 1 - \frac{1 + \frac{2\lambda_q + \lambda_q^2}{p-q+1}}{\left(1 + \frac{\lambda_q}{p-q+1}\right)^2} \right). \quad (12)$$

Note that the second term in (12) is  $O(\lambda_q^2/n(p-q+1))$ , which is small with respect to the first term if  $\lambda_q^2 \ll (p-q)$ . Hence, for reasonable values of  $n$  (say  $n < 5000$ , for which  $\log n < 10$ ), the difference  $\mu'_{\text{MDL}} - \mu_{\text{MDL}}$  is approximately  $(p-q-1/\beta)/n$ , which is nonnegligible in comparison to the difference in the penalty term of MDL,  $(p-q-1/2)\log n/n$ . Further, for  $\lambda_q \ll p-q$ , from (8) it follows that  $\sigma_{\text{MDL}} \approx \lambda_q/\sqrt{n}$ . Hence, the change in the number of standard deviations due to these terms is  $(\mu'_{\text{MDL}} - \mu_{\text{MDL}})/\sigma_{\text{MDL}} = O((p-q)/\sqrt{n}\lambda_q)$ , which again may be significant for small signal strengths, see Section V for a numerical example.

The second contribution of this paper is a random matrix view of the overestimation probability of the AIC estimator. We present a relation between this probability and the distribution of the largest eigenvalue of a Wishart matrix with identity covariance, and obtain the following result.

*Lemma 2:* Let  $\mathbf{x}_1, \dots, \mathbf{x}_n$  be  $n$  i.i.d. observations from (1) with  $q$  signals, noise variance  $\sigma^2 = 1$ , and noise-free signal eigenvalues  $\lambda_1 \geq \lambda_2, \dots \geq \lambda_q > 0$ . Let  $\xi_{n,p-q}$  be the solution of

$$-\log(1+\xi) - (p-q-1) \log\left(1 - \frac{\xi}{p-q-1}\right) = \frac{2p-2q-1}{n} \quad (13)$$

and define

$$u = \frac{1 + \xi_{n,p-q} - \mu_{n,p-q}}{\sigma_{n,p-q}}. \quad (14)$$

In addition, let  $W$  be a Wishart matrix with identity covariance matrix and parameters  $n$  and  $p-q$ , and let  $\ell_1(W)$  denote its largest eigenvalue.

Then, for  $\lambda_q = O(1)$ , in the joint limit as both  $n, p \rightarrow \infty$  with  $p/n \rightarrow c$ , an approximate expression for the overestimation probability of the AIC estimator in the presence of  $q$  signals is given by

$$\begin{aligned} \Pr[\hat{k}_{\text{AIC}} > q] &\approx \Pr[\Delta_{\text{AIC}} > 0] \\ &= \Pr\left[\frac{\ell_1(W)}{\frac{\text{Tr}(W)}{p-q}} > 1 + \xi_{n,p-q}\right] + O\left(\frac{1}{n}\right) \\ &= 1 - F_\beta(u) + \frac{1}{\beta n p} \left(\frac{\mu_{n,p-q}}{\sigma_{n,p-q}}\right)^2 \\ &\quad \times F''_\beta(u) + O(p^{-2/3}) \end{aligned} \quad (15)$$

where  $F_\beta$  is the Tracy-Widom distribution of order  $\beta$ , and  $\beta = 1$  for real valued noise or  $\beta = 2$  for complex valued noise. The centering and scaling parameters  $\mu_{n,p}$  and  $\sigma_{n,p}$  are given by (25), (26) below.

There are two key points in (15), beyond its possible use as an approximate expression for the overestimation probability of the AIC estimator. First, it shows that the overestimation probability is asymptotically *independent* of the signal eigenvalues. Second, and perhaps more importantly, it provides a novel insight, from a random matrix point of view, of why this overestimation probability is nonnegligible for practical values of  $p, n$ . Consider, for example the case of no signals,  $q = 0$ , and assume  $n$  is large, so that  $\xi_{n,p} \ll 1$ . A Taylor expansion with respect to  $\xi$  in (13) gives that

$$\xi_{n,p} = 2\sqrt{\frac{p-\frac{1}{2}}{n}} \sqrt{1 - \frac{1}{p}} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right). \quad (16)$$

According to (14) and (15), the condition for AIC to overestimate the number of signals is

$$\frac{\ell_1(W)}{\frac{\text{Tr}(W)}{p-q}} > 1 + \xi_{n,p} \approx 1 + 2\sqrt{\frac{p-\frac{1}{2}}{n}}.$$

But, as is well known from random matrix theory [13], the largest eigenvalue of a pure noise covariance matrix is  $O((1 + \sqrt{p/n})^2) \approx 1 + 2\sqrt{p/n}$ , which is comparable to the overestimation threshold  $1 + \xi_{n,p}$ . Thus, for finite small number of sensors  $p = O(10)$ , the penalty term of AIC is not sufficiently large for this event to have negligible probability. We stress that the

$$\begin{aligned} \mu'_{\text{MDL}} &= -\log(1+\lambda_q) - \log\left(1 + \frac{p-q}{n} \frac{1}{\lambda_q} - \frac{q-1}{n}\right) + \frac{2}{n\beta} - (p-q) \log\left(1 - \frac{1}{n} \frac{1+\lambda_q}{\lambda_q} - \frac{q-1}{n}\right) \\ &\quad + (p-q+1) \log\left(1 + \frac{\lambda_q}{p-q+1} \left(1 - \frac{q-1}{n}\right) - \frac{q-1}{n}\right) \\ &\quad - \frac{1}{n\beta} \frac{1}{p-q+1} \frac{(1+\lambda_q)^2 + p-q}{\left(1 + \frac{\lambda_q}{p-q+1}\right)^2} - \frac{2p-2q+1}{2} \frac{\log n}{n} \end{aligned} \quad (11)$$

above is only an approximate analysis, whose purpose is to provide insight, rather than exact results. In Lemma 3, we provide an explicit nonasymptotic bound on the overestimation probability for a family of modified AIC estimators.

Since the AIC estimator has a nonnegligible model order overestimation probability, an interesting question is what is a sufficient penalty term that would yield a negligible probability of overestimation for large sample sizes  $n$  and typical number of sensors,  $p = O(10)$ . As already discussed in the original paper by Wax and Kailath [29], (see also [36] for a corrected analysis), any penalty term of the form  $\alpha(n)k/n$  with an increasing function  $\alpha(n)$  such that  $\alpha(n)/\log \log n \rightarrow \infty$  but  $\alpha(n)/n \rightarrow 0$  as sample size  $n \rightarrow \infty$ , leads to an asymptotically consistent estimator. One solution is of course offered by the MDL estimator with penalty term  $k(2p - k)\log n/2n$ . However, this penalty is too large and leads to poor detection performance at low SNR, see [4], [17], and [35].

In the third contribution of this paper, we show that even a small increase in the penalty term of the original AIC estimator is sufficient to obtain a new nonparametric estimator for the number of sources, with both a high detection performance at low SNR and a negligible overestimation probability. In more detail, we consider a *modified* AIC estimator of the form

$$\text{AIC}'(k) = -\log \mathcal{L}_k + \frac{2Ck}{n} \left( p + 1 - \frac{(k+1)}{2} \right) \quad (17)$$

whose penalty term depends on a yet unspecified constant  $C$ . This specific form is chosen so that the difference in the penalty terms in  $\text{AIC}'(k) - \text{AIC}'(k+1)$  is  $2C(p-k)/n$ , which will turn out to be convenient for the analysis later on. Note that for  $C = 1$  this estimator has a penalty term comparable to that of the original AIC estimator, and so we naturally focus on values  $C > 1$ . The following lemma provides a nonasymptotic quantitative bound for the overestimation probability of this modified AIC estimator, and shows that a penalty constant as small as  $C = 2$  is sufficient to obtain a practically zero overestimation probability.

*Lemma 3:* Consider the modified AIC estimator (17) with constant  $C > 1$ . In the presence of pure noise with no signals ( $q = 0$ ), this estimator has a negligible overestimation probability which can be bounded by an explicitly computable expression. Analysis of the resulting expression for large  $n$  gives the following approximate bound which holds for any value of  $s < 2\sqrt{p}(\sqrt{C} - 1)$

$$\Pr[\hat{k}_{\text{AIC}'} \neq 0] \lesssim \exp\left(-\frac{ps^2}{2}\right) + \exp\left(-\frac{8\sqrt{2}p}{3} \left(\sqrt{C} - 1 - \frac{s}{2\sqrt{p}}\right)^{3/2}\right). \quad (18)$$

In addition, for a single signal to be detected with probability at least 1/2 by this estimator, its noise-free eigenvalue must be larger than  $\lambda_{\text{AIC}'}$ , where

$$\lambda_{\text{AIC}'} \approx \sqrt{2(2C-1)} \sqrt{\frac{p}{n}} \sqrt{1 + \frac{1}{p-1}}. \quad (19)$$

For comparison purposes, to detect the same signal with probability at least 1/2 by the MDL estimator, its eigenvalue must be larger than  $\lambda_{\text{MDL}}$ , where

$$\lambda_{\text{MDL}} \approx \sqrt{\frac{2p-1}{n} \log n - \frac{2p}{n}} \sqrt{1 + \frac{1}{p-1}}. \quad (20)$$

Hence, with a penalty coefficient as small as  $C = 2$  this modified AIC estimator enjoys both a much improved detection performance in comparison to MDL, together with a negligible overestimation probability. For example, for  $p = 15$ ,  $C = 2$  and  $s = 1$  the bound (18) is of the order of  $2\exp(-p/2) \approx 10^{-3}$ . In practice, this bound is highly conservative and the actual overestimation probability is orders of magnitude smaller, see Section V for simulation results. Finally, we mention that if one would like the estimator to also be strongly consistent ( $\Pr[\hat{k} = k] \rightarrow 1$  as  $n \rightarrow \infty$ ), then following the analysis in [29], this can be achieved with a sample-dependent penalty term, such as  $C_n = 2 + 0.001 \cdot \log n$ .

### III. MATHEMATICAL PRELIMINARIES

To derive approximate expressions for the overestimation and underestimation probabilities of the AIC and MDL estimators, we shall use the following results regarding the distribution of sample eigenvalues of both “signal plus noise” covariance matrices or of pure noise covariance matrices.

The following theorem characterizes the asymptotic variance and bias of sample eigenvalues. The expression for the variance is well known, see, e.g., Anderson’s book [1, ch. 13]. The expression for the bias, due to interaction terms with other eigenvalues, was first derived by Lawley [18].

*Theorem 1:* Let  $\ell_1, \dots, \ell_p$  denote the sample eigenvalues of a covariance matrix  $\mathbf{S}_n$  of  $n$  multivariate real or complex valued Gaussian observations  $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ . Let  $\alpha_1, \dots, \alpha_p$  denote the (population) eigenvalues of  $\mathbf{\Sigma}$ . For each  $i$  such that  $\alpha_i$  has multiplicity one, asymptotically in  $n$

$$\text{Var}[\ell_i] = \frac{2}{n\beta} \alpha_i^2 + o\left(\frac{1}{n}\right) \quad (21)$$

$$\mathbb{E}[\ell_i] = \alpha_i + \frac{1}{n} \sum_{j \neq i} \frac{\alpha_i \alpha_j}{\alpha_i - \alpha_j} + o\left(\frac{1}{n}\right) \quad (22)$$

with  $\beta = 1, 2$  for real or complex valued samples, respectively.

The following theorem, proven in [12] and [13] for complex or real valued noise matrices, respectively, characterizes the asymptotic distribution of the largest eigenvalue of a pure noise matrix.

*Theorem 2:* Let  $\mathbf{S}_n$  denote the sample covariance matrix of  $n$  pure noise observation vectors distributed  $\mathcal{N}(0, \mathbf{I}_p)$ . In the joint limit  $p, n \rightarrow \infty$ , with  $p/n \rightarrow c \in [0, \infty)$ , the distribution of the largest eigenvalue of  $\mathbf{S}_n$  converges to a Tracy-Widom distribution. That is, for every  $x \in \mathbb{R}$

$$\Pr\left[\frac{\ell_1 - \mu_{n,p}}{\sigma_{n,p}} < x\right] \rightarrow F_\beta(x) \quad (23)$$

where  $\beta = 1$  for real valued noise and  $\beta = 2$  for complex-valued noise. The centering and scaling parameters,  $\mu_{n,p}$  and  $\sigma_{n,p}$ , respectively, are functions of  $n$  and  $p$  only.

Furthermore, with carefully constructed expressions for  $\mu_{n,p}$  and  $\sigma_{n,p}$ , the convergence rate to the Tracy-Widom distribution is  $O(\min(p, n)^{-2/3})$ . In particular, for  $n > p$

$$\left| \Pr \left[ \frac{\ell_1 - \mu_{n,p}}{\sigma_{n,p}} < x \right] - F_\beta(x) \right| = O(p^{-2/3}). \quad (24)$$

For real valued observations, the following formulas provide  $O(p^{-2/3})$  convergence rate, see [21]

$$\mu_{n,p} = \frac{1}{n} \left( \sqrt{n - \frac{1}{2}} + \sqrt{p - \frac{1}{2}} \right)^2 \quad (25)$$

$$\sigma_{n,p} = \sqrt{\frac{\mu_{n,p}}{n}} \left( \frac{1}{\sqrt{n - \frac{1}{2}}} + \frac{1}{\sqrt{p - \frac{1}{2}}} \right)^{1/3}. \quad (26)$$

For complex valued observations, similar, though more involved expressions appear in [6], Theorem 2.

The last result we shall use is the following theorem, which provides a nonasymptotic bound on large deviations of the largest eigenvalue of a pure noise matrix for finite values of  $p, n$  (see [19, Eq. (2.4)]):

*Theorem 3:* Let  $\ell_1$  be the largest eigenvalue of a covariance matrix  $\mathbf{S}_n$  of complex multivariate Gaussian observations distributed  $\mathcal{N}(0, \mathbf{I}_p)$ . Then the following nonasymptotic bound holds for all values of  $p, n$ ,

$$\Pr \left[ \ell_1 > \left( 1 + \sqrt{\frac{p}{n}} \right)^2 + \varepsilon \right] \leq e^{-nJ_{\text{LAG}}(\varepsilon)} \quad (27)$$

where

$$J_{\text{LAG}}(\varepsilon) = \int_1^x (x-y) \frac{(1+c)y + 2\sqrt{c}}{(y+B)^2} \frac{dy}{\sqrt{y^2-1}} \quad (28)$$

with  $c = p/n$ ,  $x = 1 + \varepsilon/2\sqrt{c}$ , and  $B = 1 + c/2\sqrt{c}$ .

#### IV. ANALYSIS OF MDL AND AIC DETECTION PERFORMANCE

In this section, we prove Lemmas 1–3 from Section II. Following [7] and [35], we assume the performance of the MDL estimator is governed by the underestimation probability by exactly one signal. Thus, we start our analysis with the random variable  $\Delta_{\text{MDL}} = \text{MDL}(q-1) - \text{MDL}(q)$ , given by (5). We prove the following result, from which Lemma 1 follows.

*Lemma 4:* Let  $\mathbf{x}_1, \dots, \mathbf{x}_n$  denote  $n$  i.i.d. observations from (1) with  $q$  signals, noise variance  $\sigma^2 = 1$  and noise-free signal eigenvalues  $\lambda_1 \geq \lambda_2, \dots \geq \lambda_{q-1} > \lambda_q$ . Define

$$A = \sum_{j=1}^{q-1} \frac{1 + \lambda_j}{\lambda_j - \lambda_q}, \quad B_q = \sum_{j=1}^q \frac{1 + \lambda_j}{\lambda_j},$$

$$B_{q-1} = \sum_{j=1}^{q-1} \frac{1 + \lambda_j}{\lambda_j}.$$

Then, up to an  $o(1/n)$  error [see (29) at the bottom of the page]. This expression depends on all signal eigenvalues. However, assuming  $\lambda_{q-1} \gg \max(\lambda_q, 1)$ , and approximating

$$\frac{\lambda_j + 1}{\lambda_j - \lambda_q} \approx 1 \quad \text{and} \quad \frac{1 + \lambda_j}{\lambda_j} \approx 1 \quad (30)$$

gives the expression  $\mu'_{\text{MDL}}$  in (11) in Lemma 1.

*Remark 1:* By a similar analysis it is possible to derive the  $O(1/n)$  correction term to  $\sigma_{\text{MDL}}^2$  in (8). However, this requires considerable more tedious algebraic calculations and is hence omitted. Also, note that the  $O(1/n)$  correction term to  $\sigma_{\text{MDL}}^2$  does not change the value of  $n$  where  $\Pr[\hat{k}_{\text{MDL}} \neq q] = 1/2$ , since according to (6), this value is determined by the condition  $\mathbb{E}[\Delta_{\text{MDL}}] = 0$ .

*Remark 2:* The proof of this lemma uses Theorem 1 regarding the bias of sample eigenvalues. Hence, while (22) and (29) are correct asymptotically as  $n \rightarrow \infty$ , they provide a good approximation for finite  $n$  only if the eigenvalue with multiplicity one is sufficiently separated from the other eigenvalues, see [22]. A necessary condition is that no eigenvalue crossover (or subspace swap) occurs. Since the sample eigenvalue corresponding to a population eigenvalue  $\alpha$  has asymptotically Gaussian fluctuations with variance  $2\alpha^2/n\beta$ , the condition for a negligible probability of eigenvalue crossover is  $|\alpha_j - \alpha_i|/\alpha_i \gg \sqrt{2/(\beta n)}$ . In our case  $\alpha_j = 1 + \lambda_j$ , and so (29) is a good approximation if  $(\lambda_{q-1} - \lambda_q)/(1 + \lambda_q) \gg \sqrt{2/(\beta n)}$ .

*Proof of Lemma 4:* Equation (5) contains terms of the form  $\log(X_n)$  for various random variables  $X_n$ , whose distribution depends on the number of samples  $n$ , and whose variance is  $O(1/n)$ . Let

$$\mu_0 = \lim_{n \rightarrow \infty} \mathbb{E}[X_n], \quad \mu_1 = \lim_{n \rightarrow \infty} \mathbb{E}[n(X_n - \mu_0)].$$

$$\begin{aligned} \mathbb{E}[\Delta_{\text{MDL}}] &= -\log(\lambda_q + 1) - \log \left( 1 + \frac{p-k}{n\lambda_q} - \frac{1}{n}A \right) + \frac{1}{n\beta} \\ &\quad - (p-q) \left[ \log \left( 1 - \frac{1}{n}B_q \right) - \frac{1}{n\beta(p-q)} \right] \\ &\quad + (p+1-q) \log \left( 1 + \frac{\lambda_q}{p+1-q} - \frac{(\lambda_q+1)}{n(p+1-q)}A - \frac{p-q}{n(p+1-q)}B_{q-1} \right) \\ &\quad - \frac{1}{n\beta} \frac{1}{p+1-q} \frac{(1+\lambda_q)^2 + p-q}{\left(1 + \frac{\lambda_q}{p+1-q}\right)^2} + \frac{2p-2q+1}{2n} \log n. \end{aligned} \quad (29)$$

For each  $X_n$  we make the following change of variables

$$X_n = \mu_0 + \frac{1}{n}\mu_1 + \frac{\eta_m}{\sqrt{n}}$$

where  $\eta_m = \sqrt{n}(X_n - \mu_0 - \mu_1/n)$  is a random variable with some density  $f_n(\eta)$ , which in general depends on the number of samples  $n$ . Nonetheless, by construction  $\mathbb{E}[\eta_m] = o(1/\sqrt{n})$  so that  $\mathbb{E}[X_n] = \mu_0 + \mu_1/n + o(1/n)$ . Further, note that  $\text{Var}[X_n] = \mathbb{E}[\eta_m^2]/n(1 + o(1/n))$ .

Assuming  $n$  is large, we compute the asymptotic expansion in  $n$  of the quantity  $\mathbb{E}[\log X_n]$  via a Taylor expansion of the logarithm function, as follows:

$$\begin{aligned} \mathbb{E}[\log X_n] &= \int \log \left( \mu_0 + \frac{1}{n}\mu_1 + \frac{1}{\sqrt{n}}\eta_m \right) f_n(\eta) d\eta \\ &= \log \left( \mu_0 + \frac{1}{n}\mu_1 \right) - \frac{1}{2n} \frac{\mathbb{E}[\eta_m^2]}{\mu_0^2} + o\left(\frac{1}{n}\right) \\ &= \log(\mathbb{E}[X_n]) - \frac{1}{2} \frac{\text{Var}[X_n]}{\mathbb{E}[X_n]^2} + o\left(\frac{1}{n}\right). \end{aligned} \quad (31)$$

To apply (31) to the various random variables appearing in  $\Delta_{\text{MDL}}$  in (5) we thus need to know, up to  $o(1/n)$  terms, the means and variances of the following three random variables:  $\ell_q$ ,  $\sum_{j=q}^p \ell_j$  and  $\sum_{j=q+1}^p \ell_j$ . Applying (21) and (22) of Theorem 1 gives the equation at the bottom of the page, and

$$\text{Var} \left[ \frac{\sum_{i=q}^p \ell_i}{p+1-q} \right] = \frac{1}{(p+1-q)^2} \frac{2}{n\beta} [(1 + \lambda_q)^2 + p - q],$$

where  $\beta = 1$  for real valued observations and  $\beta = 2$  for complex valued observations. Combining these expressions with (31) and (5) yields the required result, (29).  $\square$

We now turn our attention to the performance analysis of the AIC and modified AIC estimators.

*Proof of Lemma 2:* Following [34] and [35], we assume that the signals present are sufficiently strong so that the misdetection performance of AIC is governed by its overestimation probability by exactly one source. That is

$$\text{Pr}[\hat{k}_{\text{AIC}} \neq q] \approx \text{Pr}[\Delta_{\text{AIC}} > 0] \quad (32)$$

with  $\Delta_{\text{AIC}}$  given by (10).

Our first goal is to show that under the conditions of the lemma, as  $n \rightarrow \infty$  this probability is to leading order in  $n$  independent of the  $q$  signal eigenvalues and depends only on the eigenvalues of a  $(p - q) \times (p - q)$  random Wishart matrix with identity covariance. To this end, we denote the signal subspace by  $V_s = \text{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_q\}$ , and its orthogonal complement by  $V_s^\perp$ , such that  $\mathbb{C}^p = V_s \oplus V_s^\perp$ . Let  $\tilde{\mathbf{W}} = [\mathbf{v}_1, \dots, \mathbf{v}_q, \tilde{\mathbf{w}}_{q+1}, \dots, \tilde{\mathbf{w}}_p]$  be a basis of  $\mathbb{C}^p$  whose first  $q$  vectors are the eigenvectors of the population covariance matrix corresponding to the signal eigenvalues  $\{\lambda_j + \sigma^2\}_{j=1}^q$ , and the remaining vectors diagonalize the projection of the sample covariance matrix onto the noise subspace  $V_s^\perp$ . That is

$$\tilde{\mathbf{W}}^H \mathbf{S}_n \tilde{\mathbf{W}} = \left( \begin{array}{cccc|cc} \rho_{11} & \rho_{12} & \cdots & \rho_{1q} & & \\ \vdots & \ddots & & \vdots & & \frac{1}{\sqrt{n}} \mathbf{B}^H \\ \rho_{q1} & \rho_{q2} & \cdots & \rho_{qq} & & \\ \hline & & & & \tilde{\ell}_{q+1} & 0 \\ & & & & & \ddots \\ \frac{1}{\sqrt{n}} \mathbf{B} & & & & 0 & \tilde{\ell}_p \end{array} \right).$$

The diagonal entries  $\{\rho_{jj}\}_{j=1}^q$  are the sample variances in the (unknown) directions  $\mathbf{v}_j$  that correspond to the population signal eigenvalues  $\lambda_j + 1$ . Hence,  $\rho_{jj} \sim (\lambda_j + 1)\chi_{\beta n}^2/\beta n$ . The quantities  $\tilde{\ell}_{q+1}, \dots, \tilde{\ell}_p$  are the eigenvalues of the projection of  $\mathbf{S}_n$  onto the noise subspace  $V_s^\perp$ , of dimension  $p - q$ . Note that the projection of  $\mathbf{S}_n$  onto the noise subspace gives a  $(p - q) \times (p - q)$  sub-matrix which is independent of its projection onto the signal subspace. Since this submatrix contains no signal contributions, it is a random realization of a Wishart matrix with identity covariance.

The matrix  $1/\sqrt{n}\mathbf{B}$  captures the interactions between the signal and the noise subspaces. We consider the case  $n \gg 1$ , view this matrix as a small perturbation, and expand the eigenvalues  $\ell_j$  and eigenvectors  $\mathbf{w}_j$  of  $\mathbf{S}_n$ , which correspond to the noise subspace, in terms of the matrix  $1/\sqrt{n}\mathbf{B}$  (see [22, The-

$$\begin{aligned} \mathbb{E}[\ell_q] &= (\lambda_q + 1) \left[ 1 + \frac{p-q}{n\lambda_q} - \frac{1}{n} \sum_{j=1}^{q-1} \frac{(\lambda_j + 1)}{\lambda_j - \lambda_q} \right], \quad \text{Var}[\ell_q] = \frac{2}{n\beta} (1 + \lambda_q)^2 \\ \mathbb{E} \left[ \frac{\sum_{i=q+1}^p \ell_i}{p-q} \right] &= 1 - \frac{1}{n} \sum_{j=1}^q \frac{1 + \lambda_j}{\lambda_j}, \quad \text{Var} \left[ \frac{\sum_{i=q+1}^p \ell_i}{p-q} \right] = \frac{2}{n\beta} \frac{1}{(p-q)} \\ \mathbb{E} \left[ \frac{\sum_{i=q}^p \ell_i}{p+1-q} \right] &= \frac{1}{p+1-q} \left[ p+1-q + \lambda_q - \frac{1}{n} \sum_{j=1}^{q-1} \frac{(\lambda_j + 1)(\lambda_q + 1)}{\lambda_j - \lambda_q} - \frac{p-q}{n} \sum_{j=1}^{q-1} \frac{1 + \lambda_j}{\lambda_j} \right] \end{aligned}$$

orem 2.2] for a similar analysis). We thus write for  $j = q + 1, \dots, p$

$$\begin{aligned} \ell_j &= \tilde{\ell}_j + \frac{1}{\sqrt{n}}a_j + \frac{1}{n}b_j + o_P\left(\frac{1}{n}\right) \\ \mathbf{w}_j &= \tilde{\mathbf{w}}_j + \frac{1}{\sqrt{n}}\mathbf{w}_j^{(1)} + \frac{1}{n}\mathbf{w}_j^{(2)} + o_P\left(\frac{1}{n}\right) \end{aligned}$$

where the expansion coefficients  $a_j, b_j$  and  $\mathbf{w}_j^{(1)}, \mathbf{w}_j^{(2)}$  are quantities yet to be determined. Inserting these expressions into the eigenvalue/eigenvector system  $\mathbf{S}_n \mathbf{w}_j = \ell_j \mathbf{w}_j$ , and equating powers of  $1/\sqrt{n}$  gives linear equations for the unknown quantities above. Solving these equations gives that for each  $k = 1, \dots, p - q$

$$\ell_{q+k} = \tilde{\ell}_{q+k} - \frac{1}{n} \sum_{j=1}^q \frac{|b_{j,q+k}|^2}{\rho_{jj} - \tilde{\ell}_{q+k}} + o_P\left(\frac{1}{n}\right). \quad (33)$$

The random variable  $b_{j,q+k}$  is given by

$$b_{j,q+k} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \langle \mathbf{x}_i, \mathbf{v}_j \rangle \overline{\langle \mathbf{x}_i, \tilde{\mathbf{w}}_{q+k} \rangle}.$$

It captures, to leading order in  $n$ , the interaction between the signal subspace direction  $\mathbf{v}_j$  with variance  $\rho_{jj}$  and the direction  $\tilde{\mathbf{w}}_{q+k}$  corresponding to the noise subspace eigenvalue  $\tilde{\ell}_{q+k}$ .

Since  $1/n \sum_i |\langle \mathbf{x}_i, \mathbf{v}_j \rangle|^2 = \rho_{jj}$  and  $1/n \sum_i |\langle \mathbf{x}_i, \tilde{\mathbf{w}}_{q+k} \rangle|^2 = \tilde{\ell}_{q+k}$ , we can write

$$b_{j,q+k} = \sqrt{\rho_{jj} \tilde{\ell}_{q+k}} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \alpha_i \beta_i \right) = \sqrt{\rho_{jj} \tilde{\ell}_{q+k}} \eta_{j,q+k} \quad (34)$$

where  $\alpha_i = \langle \mathbf{x}_i, \mathbf{v}_j \rangle / \sqrt{\rho_{jj}}$  and  $\beta_i = \langle \mathbf{x}_i, \tilde{\mathbf{w}}_{q+k} \rangle / \sqrt{\tilde{\ell}_{q+k}}$ . Hence, the sum in (33) can be written as

$$\ell_{q+k} = \tilde{\ell}_{q+k} \left( 1 - \frac{1}{n} \sum_{j=1}^q \frac{\rho_{jj}}{\rho_{jj} - \tilde{\ell}_{q+k}} \eta_{j,q+k}^2 \right) + o_P\left(\frac{1}{n}\right).$$

Next, we study the properties of the random variables  $\eta_{j,q+k}$ . By definition, each such variable is proportional to the correlation coefficient between the projection of the data onto a fixed direction  $\mathbf{v}_j$  and its projection onto the direction  $\tilde{\mathbf{w}}_{q+k}$ . Since the directions  $\mathbf{v}_j$  and  $\mathbf{w}_{q+k}$  are orthogonal, the random variables  $\alpha_i$  are independent of  $\beta_i$ . Furthermore, since the direction  $\mathbf{v}_j$  is fixed and independent of the signal and noise realizations, under the assumption of Gaussian sources and noise, the vector  $(\alpha_1, \dots, \alpha_n) / \sqrt{n}$  is uniformly distributed on the  $n$ -dimensional unit sphere. In particular, we have that  $\mathbb{E}[\alpha_i] = 0$ ,  $\mathbb{E}[|\alpha_i|^2] = 1$ ,  $\mathbb{E}[|\alpha_i|^2 - 1]^2] = O(1)$  and  $\mathbb{E}[\alpha_i \alpha_j] = 0$  for  $i \neq j$ . Similarly, the properties of the random variable  $\eta_{j,q+k}$  relevant to our proof are that  $\mathbb{E}[\eta_{j,q+k}] = 0$ ,  $\mathbb{E}[|\eta_{j,q+k}|^2] = 1$  and  $\mathbb{E}[|\eta_{j,q+k}|^2 - 1]^2] = O(1)$ . Furthermore, since  $\rho_{jj} \sim (\lambda_j + 1) \chi_{\beta n}^2 / \beta n$ , we have that  $\rho_{jj} = (\lambda_j + 1)(1 + O_P(1/\sqrt{n}))$  whereas for the noise eigenvalues  $\tilde{\ell}_{q+k} = 1 + O_P(1/\sqrt{n})$ . Hence

$$\ell_{q+k} = \tilde{\ell}_{q+k} \left( 1 - \frac{A_q}{n} + \frac{\sqrt{q}}{n} Z_{q+k} \right) + o_P\left(\frac{1}{n}\right) \quad (35)$$

where  $A_q = \sum_{j=1}^q (\lambda_j + 1) / \lambda_j$ , and the zero mean random variable  $Z_{q+k}$  is defined as follows:

$$Z_{q+k} = \frac{1}{\sqrt{q}} \sum_{j=1}^q \frac{\lambda_j + 1}{\lambda_j} (|\eta_{j,q+k}|^2 - 1).$$

Note that under the conditions of the lemma that  $\lambda_j = O(1)$ , it follows that  $A_q = O(q)$  and  $\text{Var}[Z_{q+k}] = O(1)$ . Similarly, we obtain up to  $o_P(1/n)$  terms

$$\begin{aligned} T_{p-q} &= \tilde{T}_{p-q} \left( 1 - \frac{A_q}{n} - \frac{\sqrt{q}}{n \tilde{T}_{p-q}} \sum_{j=q+1}^p \tilde{\ell}_j Z_j \right) \\ T_{p-q-1} &= \tilde{T}_{p-q-1} \left( 1 - \frac{A_q}{n} - \frac{\sqrt{q}}{n \tilde{T}_{p-q-1}} \sum_{j=q+2}^p \tilde{\ell}_j Z_j \right) \end{aligned}$$

where  $T_{p-q} = \sum_{j=q+1}^p \ell_j$  and  $\tilde{T}_{p-q} = \sum_{j=q+1}^p \tilde{\ell}_j$ . Inserting all of the above expressions into (10) the terms  $A_q/n$  cancel out, and up to  $o_P(1/n)$  terms we obtain

$$\begin{aligned} \Delta_{\text{AIC}} &= -\log(\tilde{\ell}_{q+1}) + (p-q) \log\left(\frac{\tilde{T}_{p-q}}{p-q}\right) \\ &\quad - (p-q-1) \log\left(\frac{\tilde{T}_{p-q-1}}{p-q-1}\right) - \frac{2p-2q-1}{n} + \frac{\sqrt{q}}{n} Z \end{aligned} \quad (36)$$

where the random variable  $Z$  is a linear combination of  $Z_j$ , for  $j = q+1, \dots, p$ , explicitly given by

$$\begin{aligned} Z &= \left( 1 - \frac{p-q}{\tilde{T}_{p-q}} \tilde{\ell}_{q+1} \right) Z_{q+1} \\ &\quad + \frac{\tilde{\ell}_{q+1}}{\tilde{T}_{p-q}} \left( \frac{p-q-1}{\tilde{T}_{p-q-1}} - 1 \right) \sum_{j=q+2}^p \tilde{\ell}_j Z_j. \end{aligned} \quad (37)$$

The key point thus far is that to leading order in  $n$ , the effect of the signals is captured solely by the fluctuations in the random variable  $Z$ , as all other terms in (36) depend only on the eigenvalues of the noise subspace  $\tilde{\ell}_j$ ,  $j = q+1, \dots, p$ . A similar conclusion was also reached by [35], where the random variable  $Z$  was neglected altogether. We now show that asymptotically as  $n \rightarrow \infty$ , the random variable  $Z$  can indeed be neglected. To this end, note that both  $1 - [(p-q)/\tilde{T}_{p-q}] \tilde{\ell}_{q+1} = O_P(\sqrt{(p-q)/n})$  and  $1 - (p-q-1)/\tilde{T}_{p-q-1} = O_P(\sqrt{(p-q-1)/n})$ . Hence,  $\sqrt{q}/n Z = O_P(\sqrt{q(p-q)/n\sqrt{n}})$  is indeed negligible with respect to the term  $(2p-2q-1)/n$  in (36).

Our analysis departs from [35] by introducing the random variable  $\xi = (p-q)\tilde{\ell}_{q+1}/\tilde{T}_{p-q} - 1$ , so that

$$\tilde{\ell}_{q+1} = \frac{\tilde{T}_{p-q}}{p-q} (1 + \xi). \quad (38)$$

Recall that  $\tilde{T}_{p-q-1} = \tilde{T}_{p-q} - \tilde{\ell}_{q+1}$ . Hence, by definition

$$\frac{\tilde{T}_{p-q-1}}{p-q-1} = \frac{\tilde{T}_{p-q}}{p-q} \left( 1 - \frac{\xi}{p-q-1} \right).$$

Inserting these expressions into (36) gives that the condition for AIC to overestimate the number of signals is, up to  $o_P(1/n)$

$$-\log(1+\xi) - (p-q-1) \log\left(1 - \frac{\xi}{p-q-1}\right) > \frac{2p-2q-1}{n}. \tag{39}$$

Let  $\xi_{n,p-q}$  be the solution with equality in (39). Then, asymptotically in  $n$ , overestimation occurs when

$$\tilde{\ell}_{q+1} > \frac{\tilde{T}_{p-q}}{p-q} (1 + \xi_{n,p-q}). \tag{40}$$

Thus, asymptotically, the overestimation probability of the AIC estimator depends on the tail statistics of the ratio

$$U_{n,p-q} = \frac{\ell_1(W)}{\frac{\text{Tr}(W)}{(p-q)}}$$

of the largest eigenvalue of a Wishart matrix with identity covariance and parameters  $n$  and  $p-q$  to the mean of all its eigenvalues. For small values of  $p$  (say  $p \leq 6$ ), the tail distribution of this random variable was tabulated, see [25], and in principle could be numerically calculated for larger values of  $p$  with the aid of a computer.

To derive an approximate explicit expression for this overestimation probability, we first note that similar to Theorem 2, in the joint limit as both  $p, n \rightarrow \infty$ , it can be proven that with the same centering and scaling parameters,  $\mu_{n,p}, \sigma_{n,p}$  as for the largest eigenvalue

$$\Pr\left[\frac{U_{n,p} - \mu_{n,p}}{\sigma_{n,p}} < s\right] \rightarrow TW_\beta(s).$$

However, as discussed in [23], a more accurate expression for tail probabilities of the ratio, with an error  $O(p^{-2/3})$  is

$$\Pr\left[\frac{U_{n,p} - \mu_{n,p}}{\sigma_{n,p}} > s\right] = 1 - TW_\beta(s) + \frac{1}{\beta np} \left(\frac{\mu_{n,p}}{\sigma_{n,p}}\right)^2 TW''_\beta(s).$$

Combining this expression with (40) proves the lemma.  $\square$

*Proof of Lemma 3:* We perform an analysis similar to the one for the AIC estimator. Note that for the case of no signals, there are no interactions between signal and noise eigenvalues. Hence, the exact condition for  $\Delta_{AIC'} > 0$  is

$$\ell_1 > \frac{\sum_{j=1}^p \ell_j}{p} (1 + \xi_{n,p}) \tag{41}$$

where  $\xi_{n,p}$  is the solution of the following equation, which depends on the penalty constant  $C$ ,

$$-\log(1+\xi) - (p-1) \log\left(1 - \frac{\xi}{p-1}\right) = 2C \frac{p}{n}. \tag{42}$$

For future use, we note that similar to the analysis made in (16), for large  $n$  we have that  $\xi_{n,p} = 2\sqrt{C}\sqrt{p/n} + o(1/\sqrt{n})$ . Also, recall that the random variable  $\sum_{j=1}^p \ell_j/p$  is the trace of a pure noise covariance matrix divided by its size, and is thus distributed as a  $\chi^2_{2np}/(2np)$  random variable.

Let  $s$  be some positive number. Then, [see (43) at the bottom of the page]. In principle, both terms on the right-hand side (RHS) above can be computed numerically for any value of  $p, n, s$ , and  $C$ , and one can also find the value of  $s$  where the bound attains a minimum. As we now show, the key point is that both terms in (43) are exponentially small in  $p$ .

For the first term, we use the following inequality (see [14]):

$$\Pr[\chi_k^2 \leq k(1-\varepsilon)] \leq \exp\left(\frac{-k\varepsilon^2}{4}\right). \tag{44}$$

Applying this bound with  $k = 2np$  and  $\varepsilon = s/\sqrt{n}$ , gives that the first term in (43) is bounded by  $\exp(-ps^2/2)$ . As for the second term, according to Theorem 3 it is bounded by (27), with  $\varepsilon = (1 - s/\sqrt{n})(1 + \xi_{n,p}) - (1 + \sqrt{p/n})^2$ . To understand the dependence of this bound on the number of sensors  $p$  and on the penalty coefficient  $C$ , we consider in more detail the probability of an event of the form  $\{\ell_1 > (1 + \sqrt{p/n})^2 + \delta\sqrt{p/n}\}$ . According to (27), this probability is bounded by  $\exp(-nJ_{\text{LAG}}(\delta\sqrt{p/n}))$ , where

$$\begin{aligned} nJ_{\text{LAG}}\left(\delta\sqrt{\frac{p}{n}}\right) &= 4p \int_1^{1+\delta/2} \frac{\frac{1+\delta}{2-y}}{\sqrt{y^2-1}} \frac{\left(\frac{1+p}{n}\right)y + \sqrt{\frac{p}{n}}}{\left(1 + 2y\sqrt{\frac{p}{n}} + \frac{p}{n}\right)^2} dy \\ &= 4p \int_1^{1+\delta/2} \frac{y\left(\frac{1+\delta}{2-y}\right)}{\sqrt{y^2-1}} \left(1 + O\left(\sqrt{\frac{p}{n}}\right)\right) dy. \end{aligned}$$

Evaluating the integral gives that for any  $\delta > 0$ , [see (45) at the bottom of the next page]. Hence

$$nJ_{\text{LAG}}\left(\delta\sqrt{\frac{p}{n}}\right) \geq \frac{4p}{3} \delta^{3/2} \left(1 + O\left(\sqrt{\frac{p}{n}}\right)\right). \tag{46}$$

In our case,  $\xi_{n,p} \approx 2\sqrt{C}\sqrt{p/n}$ , and so  $\delta \approx 2(\sqrt{C} - 1 - s/2\sqrt{p})$ . Hence, combining (44), (46), and (43) gives (18).

Finally, to prove (19), we note that the condition for the modified estimator to detect a single signal with probability at least

$$\begin{aligned} \Pr[\Delta_{AIC'} > 0] &= \Pr\left[\Delta_{AIC'} > 0 \cap \frac{\sum_{j=1}^p \ell_j}{p} \leq 1 - \frac{s}{\sqrt{n}}\right] + \Pr\left[\Delta_{AIC'} > 0 \cap \frac{\sum_{j=1}^p \ell_j}{p} > 1 - \frac{s}{\sqrt{n}}\right] \\ &\leq \Pr\left[\frac{\chi^2_{2np}}{2np} < 1 - \frac{s}{\sqrt{n}}\right] + \Pr\left[\ell_1 > \left(1 - \frac{s}{\sqrt{n}}\right) (1 + \xi_{n,p})\right]. \end{aligned} \tag{43}$$

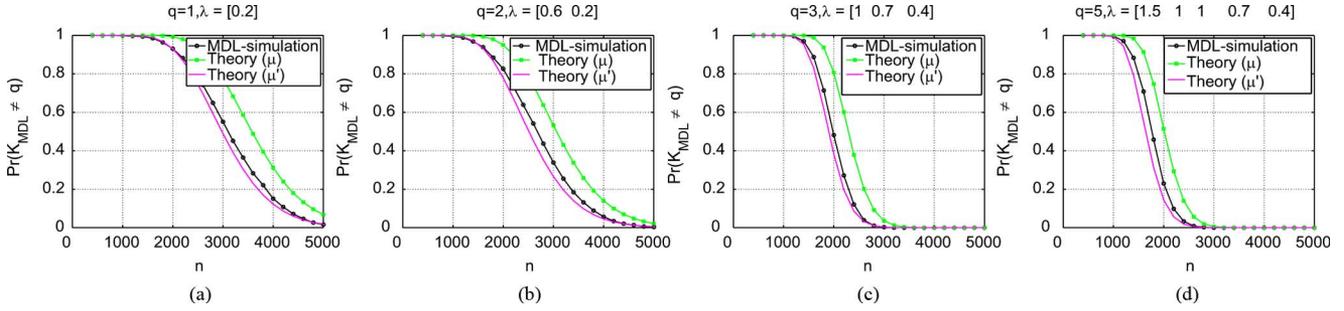


Fig. 1. Comparison of empirical performance of the MDL estimator to the theoretical formula (7) and to the improved formula (11). Parameters:  $\sigma = 1$ , complex valued signals and noise (a)  $p = 7$  one signal. (b)  $p = 7$  two signals. (c)  $p = 20$  three signals. (d)  $p = 20$ , five signals.

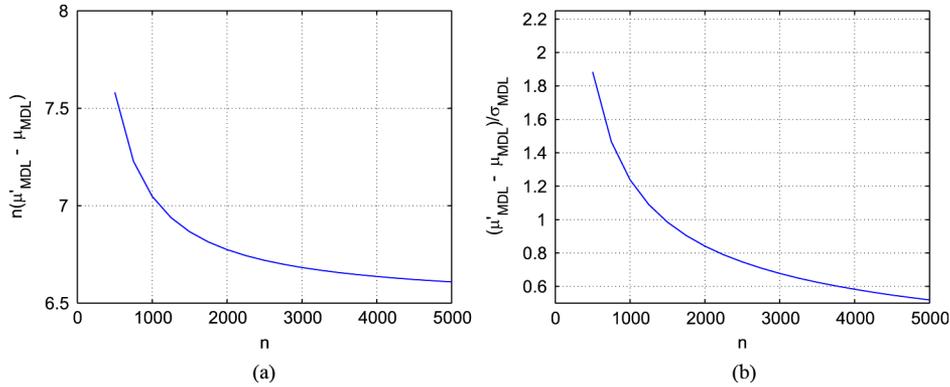


Fig. 2. Analysis of the difference between (7) and (11) for  $E[\Delta_{MDL}]$ . (a) Difference  $n(\mu'_{MDL} - \mu_{MDL})$  as function of sample size for  $p = 7, q = 1, \lambda_1 = 0.2$ . (b) Difference in standard deviations  $(\mu'_{MDL} - \mu_{MDL})/\sigma_{MDL}$  as function of sample size.

$1/2$  is  $E[AIC'(0) - AIC'(1)] > 0$ . Similar to the analysis of the MDL estimator, this gives

$$-\log(1 + \lambda) + p \log \left( 1 + \frac{\lambda}{p} \right) + \frac{p}{n} > 2C \frac{p}{n}.$$

For large  $n$  and small  $\lambda$ , a Taylor expansion of the logarithms gives the approximate condition

$$\frac{\lambda^2}{2} \left( 1 - \frac{1}{p} \right) > (2C - 1) \frac{p}{n}$$

from which (19) follows.  $\square$

### V. SIMULATIONS

In this section, we compare simulation results to our theoretical predictions regarding the probability of misdetection of both the MDL, the AIC and the modified AIC estimators. We start with the MDL estimator. In Fig. 1 we show the result of computer simulations for the misdetection probability of the MDL estimator, in comparison to the theoretical prediction of (6) for  $p = 7$  or  $p = 20$  sensors and for various numbers of signals, either with the original expression for  $\mu$ , (7), or with our modified expression, (11). As shown in the figure, there is a considerable discrepancy between the original theoretical curve com-

puted from (7) and the empirical results, whereas the theoretical curve computed with the modified expression (11) is more accurate.

Next, we show that the difference  $\mu'_{MDL} - \mu_{MDL}$  is nonnegligible for practical values of  $p, n$ . Consider  $p = 7$  sensors, and a single signal ( $q = 1$ ) of strength  $\lambda_1 = 0.2$ , as in the upper left simulation in Fig. 1. In Fig. 2(a), we plot the difference  $n(\mu'_{MDL} - \mu_{MDL})$  as a function of sample size  $n$ . As expected from (12), this difference converges to  $p - 1/2$ . In Fig. 2(b), we plot  $(\mu'_{MDL} - \mu_{MDL})/\sigma_{MDL}$  as a function of sample size  $n$ . Note that the difference is of the order of one standard deviation for the values of  $n$  considered. This nonnegligible difference explains the discrepancy in Fig. 1 between the empirical simulation results and the theoretical performance curve based on the original expression for  $\mu_{MDL}$ , (7).

Next, we consider the misdetection probability of the AIC estimator. In Fig. 3, we present the result of computer simulations of the probability of overestimation of the AIC estimator, in comparison to the theoretical prediction of (15), for  $n = 200$  or  $n = 1500$  samples as a function of the number of sensors  $p$ , and for different number of sources,  $q = 0, 2, 4$ . In all cases the theoretical prediction is quite accurate with an error of up to 1%–2%.

$$\int_1^{1+\delta/2} \frac{y \left( \frac{1+\delta}{2-y} \right)}{\sqrt{y^2 - 1}} dy = \frac{1}{4} \sqrt{\delta(\delta + 4)} \left( 1 + \frac{\delta}{2} \right) - \frac{1}{2} \log \left( 1 + \frac{\delta}{2} + \frac{1}{2} \sqrt{\delta(\delta + 4)} \right) \geq \frac{\delta^{3/2}}{3}. \quad (45)$$

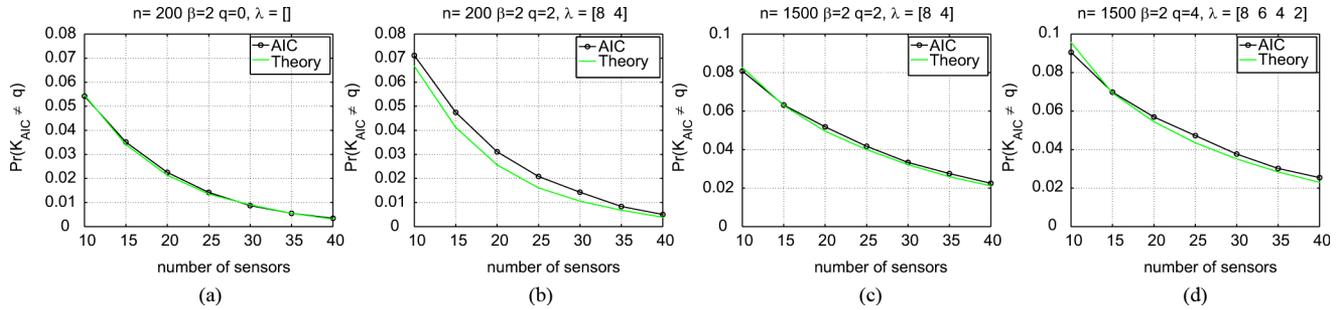


Fig. 3. Comparison of empirical misdetection probability of the AIC estimator to the theoretical formula (15). Parameters:  $\sigma = 1$ , complex valued signals and noise, each point is the average of 100 000 simulations (a)  $n = 200$ , no signals. (b)  $n = 200$ , two signals. (c)  $n = 1500$ , two signals. (d)  $n = 1500$ , four signals.

TABLE I  
MISDETECTION PROBABILITY OF MODIFIED AIC ESTIMATOR WITH PENALTY  $C = 2$ . EMPIRICAL AVERAGE RESULTS OF 15 000 000 SIMULATIONS, INCLUDING ONE STANDARD DEVIATION CONFIDENCE INTERVALS.

	$\Pr[\hat{k} \neq q], p = 8$	$\Pr[\hat{k} \neq q], p = 12$
$n = 200, q = 0$	$(5 \pm 2) \cdot 10^{-7}$	0
$n = 1000, q = 0$	$(6 \pm 2) \cdot 10^{-7}$	0
$n = 200, q = 2, \lambda = [50, 40]$	$(6.5 \pm 0.6) \cdot 10^{-6}$	0
$n = 1000, q = 2, \lambda = [50, 40]$	$(7 \pm 0.6) \cdot 10^{-6}$	0

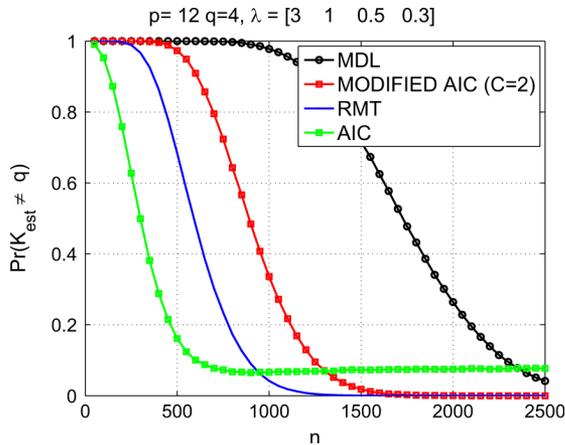


Fig. 4. Comparison of misdetection probability of various algorithms as a function of sample size  $n$ .

Finally, we consider the misdetection probability of the modified AIC estimator. Table I presents simulation results for the probability of misdetection of the modified AIC estimator with a penalty constant  $C = 2$ , various values for  $p, n$  and different number of sources. As predicted theoretically, the overestimation probability is extremely small, and it slightly increases with the number of signals  $q$ , as this decreases the noise subspace dimension from  $p$  to  $p - q$ . Since the penalty term of the modified AIC estimator is significantly smaller than that of the MDL estimator, it enjoys a much improved signal detection performance at low SNR, compared to the MDL estimator. We illustrate this in Fig. 4, where we compare the misdetection probability of the MDL, AIC, and modified AIC estimator with  $C = 2$ , as well as the algorithm recently developed in [16], with a confidence level of  $\alpha = 10^{-4}$ . The latter algorithm performs a sequence of hypothesis tests for the significance of the largest eigenvalue as arising from a signal, using results from random matrix theory,

and we thus denote it as RMT. Note that even though the modified-AIC and the RMT algorithms are not consistent, their overestimation probabilities are so small that on the scale of the figure, their misdetection error appears to reach a value of zero.

## VI. SUMMARY AND CONCLUSION

In this paper, we presented a statistical performance analysis of the MDL and AIC estimators, and of a modified AIC-type estimator with a penalty slightly larger than that of the AIC estimator but significantly smaller than that of the MDL estimator. We showed that this modified AIC estimator has a much better detection performance than MDL with a negligible probability of overestimation.

The two key quantities used in our analysis were: i) the distribution of the largest eigenvalue of a pure noise matrix; and ii) the interaction of signal and noise eigenvalues. These quantities also played a key role in our previous work [16], [17], where we developed a source enumeration estimator based on a sequence of hypothesis tests and analyzed its performance and that of two other recent estimators [24], [26].

Since source enumeration involves choosing amongst many composite hypothesis, in general there is no universally optimal detection procedure. Indeed, comparing the detection performance of the modified AIC estimator, (19), with that of the RMT estimator, (25) in [17], no estimator dominates the other over all parameter ranges. In particular, the modified AIC algorithm may be useful in legacy systems where only minimal changes to an existing code can be made.

In this paper, we focused on analysis in the nonparametric setting and relied on recent results from random matrix theory. In a similar fashion, both analysis and development of improved estimators in a parametric setting, with an explicit knowledge of the array manifold, are possible, this time using results from extreme value theory on the maxima of random processes and fields. This analysis will be described in a separate publication [2].

## ACKNOWLEDGMENT

The author would like to thank the three anonymous referees for valuable suggestions which greatly improved the exposition of the paper.

## REFERENCES

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed. New York: Wiley, 2003.

- [2] N. Arkind and B. Nadler, Parametric joint detection-estimation of the number of sources in array processing, unpublished.
- [3] J. F. Böhme, "Array processing," in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1991, pp. 1–63.
- [4] W. Chen, K. M. Wong, and J. P. Reilly, "Detection of the number of signals: A predicted eigen-threshold approach," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1088–1098, 1991.
- [5] P. Chen, M. C. Wicks, and R. S. Adve, "Development of a statistical procedure for detecting the number of signals in a radar measurement," *Inst. Elect. Eng. Proc. Radar Sonar Navig.*, vol. 148, no. 4, pp. 219–226, 2001.
- [6] N. El Karoui, "A rate of convergence result for the largest eigenvalue of complex white Wishart matrices," *Ann. Prob.*, vol. 36, no. 6, pp. 2077–2117, 2006.
- [7] E. Fishler, M. Grossmann, and H. Messer, "Detection of signals by information theoretic criteria: General asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1027–1036, 2002.
- [8] F. Haddadi, M. Malek-Mohamaddi, M. M. Nayebi, and M. M. Aref, "Statistical performance analysis of MDL source enumeration in array processing," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 452–457, 2010.
- [9] L. Huang and S. Wu, "Low-complexity MDL method for accurate source enumeration," *IEEE Signal Process. Lett.*, vol. 14, no. 9, pp. 581–584, 2007.
- [10] L. Huang, S. Wu, and X. Li, "Reduced-rank MDL method for source enumeration in high-resolution array processing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5658–5667, 2007.
- [11] L. Huang, T. Long, and S. Wu, "Source enumeration for high-resolution array processing using improved Gershgorin radii without eigendecomposition," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5916–5925, 2008.
- [12] K. Johansson, "Shape fluctuations and random matrices," *Comm. Math. Phys.*, vol. 209, no. 12, pp. 437–474, 2000.
- [13] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal component analysis," *Ann. Stat.*, vol. 29, pp. 295–327, 2001.
- [14] I. M. Johnstone and A. Lu, "On consistency and sparsity for principal component analysis in high dimensions," *J. Amer. Stat. Assoc.*, vol. 104, pp. 682–693, 2009.
- [15] M. Kaveh, H. Wang, and H. Hung, "On the theoretical performance of a class of estimators of the number of narrow-band sources," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 35, pp. 1350–1352, 1987.
- [16] S. Kritchman and B. Nadler, "Determining the number of components in a factor model from limited noisy data," *Chem. Intell. Lab. Syst.*, vol. 94, pp. 19–32, 2008.
- [17] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals, hypothesis tests and random matrix theory," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, 2009.
- [18] D. N. Lawley, "Tests of significance for the latent roots of covariance and correlation matrices," *Biometrika*, vol. 43, pp. 128–136, 1956.
- [19] M. Ledoux, "Deviation inequalities on largest eigenvalues, in geometric aspects of functional analysis," in *Lecture Notes in Mathematics*, V. D. Milman and G. Schechtman, Eds. New York: Springer, 2007, vol. 1910.
- [20] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1689–1695, 2001.
- [21] Z. Ma, "Accuracy of the Tracy-Widom limit for the extreme eigenvalues in white Wishart matrices," submitted for publication.
- [22] B. Nadler, "Finite sample approximation results for principal component analysis: A matrix perturbation approach," *Ann. Stat.*, vol. 36, no. 6, pp. 2791–2817, 2008.
- [23] B. Nadler, "On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix," submitted for publication.
- [24] N. R. Rao and A. Edelman, "Sample eigenvalue based detection of high dimensional signals in white noise using relatively few samples," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2625–2638, 2008.
- [25] F. J. Schuurmann, P. R. Krishnaiah, and A. K. Chattopadhyah, "On the distribution of ratios of the extreme roots to the trace of the Wishart matrix," *J. Mult. Anal.*, vol. 3, pp. 445–453, 1973.
- [26] J. Schott, "A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix," *J. Mult. Anal.*, vol. 97, no. 4, pp. 827–843, 2006.
- [27] M. O. Ulfarsson and V. Solo, "Dimension estimation in noisy PCA with sure and random matrix theory," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 5804–5816, 2008.
- [28] S. Valaee and P. Kabal, "An information theoretic approach to source enumeration in array signal processing," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1171–1178, 2004.
- [29] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.
- [30] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 37, no. 8, pp. 1190–1196, 1989.
- [31] D. B. Williams, "Detection: Determining the number of sources," in *Digital Signal Processing Handbook*, V. K. Madisetti and D. B. Williams, Eds. Boca Raton, FL: CRC, 1999.
- [32] K. M. Wong, Q.-T. Zhang, J. P. Reilly, and P. C. Yip, "On information theoretic criteria for determining the number of signals in high resolution array processing," *IEEE Trans. Signal Process.*, vol. 38, no. 11, pp. 1959–1971, 1990.
- [33] J. Xin, N. Zheng, and A. Sano, "Simple and efficient nonparametric method for estimating the number of signals without eigendecomposition," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1405–1420, 2007.
- [34] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. Signal Process.*, vol. 43, pp. 1413–1426, 1995.
- [35] Q. T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 10, pp. 1557–1567, 1989.
- [36] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "Remarks on certain criteria for detection of number of signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 129–132, 1987.

**Boaz Nadler** was born in Israel in 1971. He received the B.Sc. degree in mathematics and physics (*cum laude*), the M.Sc. degree in applied mathematics (*summa cum laude*), and the Ph.D. degree in applied mathematics, all from Tel Aviv University (TAU), Tel Aviv, Israel.

From 2002 to 2005, he was a Gibbs Instructor/Assistant Professor with the Department of Mathematics, Yale University, New Haven, CT. He is currently a Senior Research Scientist with the Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. His research interests are in mathematical statistics, machine learning, stochastic processes, and their applications in chemometrics and in signal processing.