An Efficient and Simple Algorithm for Estimating the Number of Sources via $\ell_{0.55}$-Norm

Dayan Adionel Guimarães
National Institute of Telecommunications - Inatel
Santa Rita do Sapucaí, MG, Brazil
Email: dayan@inatel.br

Giovanni Henrique Faria Floriano
National Institute of Telecommunications - Inatel
Santa Rita do Sapucaí, MG, Brazil
Email: giovanni@inatel.br

Rausley Adriano Amaral de Souza
National Institute of Telecommunications - Inatel
Santa Rita do Sapucaí, MG, Brazil
Email: rausley@inatel.br

***Abstract*—Recently, it has been proposed an empirical method for estimating the number of sources of signals impinging on multiple sensors, named norm-based (NB) algorithm. The algorithm computes the Euclidian norm of vectors whose elements are the normalized and nonlinearly scaled eigenvalues of the received signal covariance matrix, and the corresponding normalized indexes. Such norms are then used to discriminate the largest eigenvalues from the remaining ones, thus allowing for the estimation of the number of sources. In this paper we propose an improved norm-based (iNB) algorithm which uses the $\ell_{0.55}$-norm as a means for classifying the eigenvalues. Differently from the NB, the iNB algorithm does not use the nonlinear scaling and does not need to set an additional empirical constant that is crucial to the proper operation of the NB algorithm. Comparisons are made with the estimators MDL (minimum description length) and AIC (Akaike information criterion), and with a recently-proposed estimator based on the random matrix theory (RMT). It is shown that the iNB algorithm can outperform one or more of these estimators in several situations, and that it always outperforms the NB algorithm.**

I. Introduction

The estimation of the number of sources of signals impinging on multiple sensors is a fundamental problem in communications and signal processing. This number is important in itself in some cases, e.g. to determine the approximate number of neurons responding to some stimulus in medical applications. In other cases it is used as the input for subsequent procedures, e.g. for direction of arrival (DoA) estimation in antenna array processing applications. Common solutions to this problem adopt information theoretic approaches, such as the minimum description length (MDL) and the Akaike information criterion (AIC) [1], [2]. Recently, a random matrix theory (RMT) approach was proposed in [3], claiming high detection performance at low signal-to-noise ratio (SNR), similar to the AIC estimator, and near consistency at large sample sizes, similar to the MDL estimator.

In this paper we propose a new algorithm for improving the performance of the norm-based (NB) estimator discussed in [4]. We call it improved NB (iNB). Our algorithm enjoys lower complexity than the NB, yet exhibiting consistency and outperforming the MDL, the AIC and the RMT-based estimators in several situations. The performance of the iNB is always better than the performance achieved by the NB estimator alone.

II. Problem Formulation

Let an array with m sensors (antennas for example), each one collecting n samples of the received signal from p transmitters (sources). These samples are arranged in a matrix $\mathbf{Y} \in \mathbb{C}^{m\times n}$ and the samples from the p transmitters are arranged in a matrix $\mathbf{X} \in \mathbb{C}^{p\times n}$. Let $\mathbf{H} \in \mathbb{C}^{m\times p}$ be the channel matrix with elements $\{h_{ij}\}$, $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, p$, representing the channel gains between the j-th source and the i-th sensor. Finally, let $\mathbf{V} \in \mathbb{C}^{m\times n}$ be a matrix of additive Gaussian noise samples, distributed $\mathcal{N}(0, \sigma^2\mathbf{I}_m)$ and independent of the signal samples. The matrix of received samples is then $\mathbf{Y} = \mathbf{HX} + \mathbf{V}$.

We consider only those methods for estimating the number of sources that use the eigenvalues $\{\lambda_1 \geq \lambda_2 \geq \cdots \lambda_m\}$ of the received signal population covariance matrix $\mathbf{R}$, for which the maximum likelihood estimate is the sample covariance matrix $\hat{\mathbf{R}} = \mathbf{YY}^\dagger/n$, where $\dagger$ means complex conjugate and transpose. It can be shown [1] that $\mathbf{R}$ can be written as $\mathbf{R} = \mathbb{E}[\mathbf{YY}^\dagger] = \mathbf{HR}_x\mathbf{H}^\dagger + \sigma^2\mathbf{I}$, where $\mathbf{R}_x$ is the transmitted signal population covariance matrix, $\mathbf{I}$ is the identity matrix, σ^2 is the noise variance and $\mathbb{E}[\cdot]$ is the expectation operator. If $\mathbf{H}$ is full column rank, the rank of $\mathbf{HR}_x\mathbf{H}^\dagger$ is p, which means that the $m-p$ smallest eigenvalues of $\mathbf{HR}_x\mathbf{H}^\dagger$ are equal to zero. Therefore, the $m-p$ smallest eigenvalues of $\mathbf{R}$ are equal to σ^2. Then, it is possible to estimate the number of sources p from the multiplicity of the smallest eigenvalues of $\mathbf{R}$. Instead of $\mathbf{R}$, in practice its estimate $\hat{\mathbf{R}}$ is computed using a finite number of samples, and the resulting eigenvalues are all different with probability one. In this case the classification of the eigenvalues in two groups (the largest p and the smallest $m-p$) is not trivial, representing a challenge for the estimation of the number of sources.

III. AIC, MDL and RMT-based Estimators

The criteria for AIC and MDL are respectively [1]:

$$\text{AIC}(k) = -2n \ln \frac{(m-k)\prod_{i=k+1}^{m}\lambda_i}{\left(\sum_{i=k+1}^{m}\lambda_i\right)^{m-k}} + 2k\,(2m-k)\,,$$

$$\text{MDL}(k) = -n \ln \frac{(m-k)\prod_{i=k+1}^{m}\lambda_i}{\left(\sum_{i=k+1}^{m}\lambda_i\right)^{m-k}} + \frac{1}{2}k\,(2m-k)\ln n,$$

for $k = 0, 1, ..., m-1$. For both, the estimate $\hat{p}$ of the number of sources is the value of k which minimizes the criterion.

The RMT-based estimate can be written as [3] $\hat{p} = \arg\min_k \left\{\lambda_k < \hat{\sigma}^2 \left[\mu_{n,m-k} + s(\alpha)\xi_{n,m-k}\right]\right\} - 1$, where $\hat{\sigma}^2$ is the estimate of the noise variance, $\mu_{n,m-k}$ and $\xi_{n,m-k}$ are respectively determined from the centering and scaling parameters of the Tracy-Widom distribution, and $s(\alpha)$ is a function of the asymptotic false alarm (overestimation) probability α, which is obtained by inverting numerically the Tracy-Widom distribution (for more details on the RMT-based method, please refer to [3]). Notice that this method suffers from the need of computing a threshold that depends on an estimate of the noise variance, but on the other hand permits control of the overestimation probability, which can be desired in some application.

IV. THE ORIGINAL NB ALGORITHM

In the original NB algorithm [4], the ordered eigenvalues of $\hat{\mathbf{R}}$ and the corresponding indexes are normalized so that both are placed in the interval $[0, 1]$, that is, for $i = 1, ..., m$,

$$l_i = \frac{\lambda_i - \lambda_m}{\lambda_1 - \lambda_m}, \qquad i^{(\mathrm{N})} = \frac{i-1}{m-1}, \tag{1}$$

where l_i and $i^{(\mathrm{N})}$ are the i-th normalized eigenvalues and indexes, respectively. The normalized eigenvalues $\{l_i\}$ are further modified by a nonlinear operation, leading to

$$\lambda_i^{(\mathrm{N})} = \sqrt{1 - (1 - l_i)^E}. \tag{2}$$

The role of this nonlinear operation is explained in what follows, with the help of an example: Figure 1(a) shows the normalized eigenvalues $\{l_i\}$ and $\{\lambda_i^{(\mathrm{N})}\}$ for some values of the exponent E, assuming $m = 30$, $p = 0$ and $n = 5000$. Since $p = 0$, as n grows the curve for $\{l_i\}$ tends to become a straight line. The effect of (2) is a bending of this curve according to the value of E. For $E = 2$ and large n the curve for $\{\lambda_i^{(\mathrm{N})}\}$ tends to lay on the unit semicircumference quarter. For $E > 2$ the curves for $\{\lambda_i^{(\mathrm{N})}\}$ are further bent. Now define a vector $\mathbf{\Lambda}_i = [\lambda_i^{(\mathrm{N})} \; i^{(\mathrm{N})}]^T$. Notice in Figure 1(a) that for $E > 2$, $\|\mathbf{\Lambda}_1\| < \|\mathbf{\Lambda}_i\|$ for $i \neq 1$, where $\|\mathbf{\Lambda}_i\|$ is the Euclidean norm of $\mathbf{\Lambda}_i$. For $p > 0$ the smallest $m - p$ eigenvalues of $\hat{\mathbf{R}}$ tend to be equal to the noise variance, and it will become evident an inflection point in the curve, at the transition from the largest p and the smallest $m - p$ eigenvalues, as illustrated in Figure 1(b). This will tend to make $\|\mathbf{\Lambda}_{p+1}\| < \|\mathbf{\Lambda}_i\|$ for $i \neq p + 1$. Then, $\|\mathbf{\Lambda}_i\|$ can be used to estimate the number of sources. This is the essence of the NB algorithm. For a small number of sources, however, the placement of the inflection point more to the left will lead to a high chance of having $\|\mathbf{\Lambda}_i\| < \|\mathbf{\Lambda}_{p+1}\|$ for $i > p + 1$. To avoid this, only a subset of vectors of the set $\{\mathbf{\Lambda}_i\}$ have to be tested while searching for the smallest Euclidian norm. Figure 1(b) illustrates this heuristic for a subset size $K = 15$ for $p = 0$ and $p = 5$, assuming $m = 30$, $n = 5000$ and $E = 5$. Notice that searching within a subset of size smaller than m is equivalent to pushing the inflection point of the curves to the right, reducing the chance of overestimating of the number of sources. If the number of sources increases, the entire set with $K = m$ vectors is used; see an example for $p = 15$ in Figure 1(b). The need for determining the best value of K is the main drawback of the NB algorithm, since it is influenced by the expected maximum number of sources, an information that is not known a priori in most of the applications.

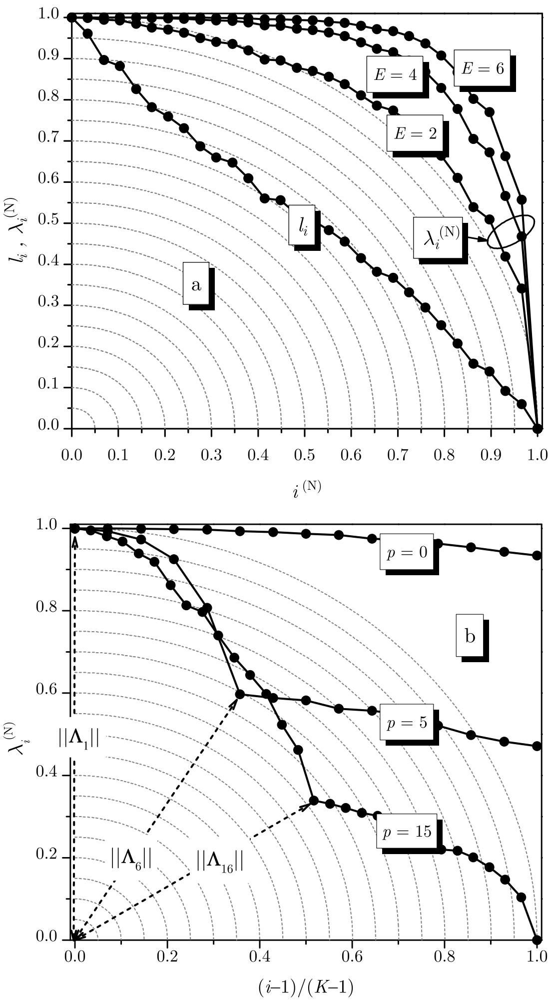

Fig. 1. (a) Graphical representation of the normalized eigenvalues and indexes (b) Normalized eigenvalues and indexes according to the NB algorithm

As stated in [4], the choice of the bending exponent E in (1) can be made via simulation, aiming at maximizing the probability of correct detection, P_{c}, which is the probability of correctly estimating the number of sources, for a specific set of system parameters. In a scenario of more practical significance, E can be found as the value which maximizes the average P_{c} over several sets of parameters. For example, combining the parameters $m = 10, 15, 20, 50$; $n = 50, 100, 200, 500, 1000, 50000$; SNR $= -5$ dB, 0 dB, 5 dB, 8 dB, 10 dB; $p = 2, 5, 10, 15$, the optimal exponent $E = 5$ was found in [4]. The numerical results shown in Section VI for the NB algorithm consider $E = 5$.

The original NB algorithm is summarized as follows [4]:

Algorithm 1 The NB Algorithm

for $i = 1 \ldots m$ **do**
  Compute $\lambda_i^{(\mathrm{N})}$ using (1) and (2)
end for
Do $K = m/2$
for $j = 1 \ldots K$ **do**
  Compute $\mathbf{\Lambda}_j = \left[\lambda_j^{(\mathrm{N})} \;\; \frac{j-1}{K-1}\right]^T$
end for
Compute $\hat{p} = \arg\min_j ||\mathbf{\Lambda}_j||_2 - 1$

V. The Proposed iNB Algorithm

As previously stated, the need for determining the value of K is the main drawback of the NB algorithm. However, a closer look at (2) allows us to interpret the nonlinear scaling of the eigenvalues as a change or distortion in the distance space of the normalized eigenvalues so that the Euclidian norm is computed in the sequel. This is intuitively reasonable, but can be viewed from another perspective: why do not keep unchanged the distance space of the normalized eigenvalues and change only the distance measure? But this new perspective for the problem at hand is nothing but a possible interpretation of the general ℓ_u-norm. Recalling, the ℓ_u-norm of a d-dimensional vector $\mathbf{x}$, which is usually denoted by $\|\mathbf{x}\|_u$, is defined by: $\|\mathbf{x}\|_u = (|x_1|^u + |x_2|^u + \cdots + |x_d|^u)^{1/u}$.

In order to see how the above reasoning works, Figure 2(a) shows the idealized asymptotic behavior of the eigenvalues and indexes, for $p = 0$, normalized according to (1), i.e. $\{l_i\}$ (balls), and according to (1) and (2), i.e. $\{\lambda_i^{(\mathrm{N})}\}$ (squares) for $E = 5$. Figure 2(b) shows the $\ell_{0.83}$-norm for all $\{l_i\}$ (balls) and the Euclidian norm (ℓ_2-norm) for all $\{\lambda_i^{(\mathrm{N})}\}$, for $p = 0$. From Figure 2(b) it is possible to see the effect of bending the normalized eigenvalues by the nonlinear scaling of (2) on the Euclidian norm for each $\lambda_i^{(\mathrm{N})}$, and compare this effect with the one produced by measuring the norm from each $\{l_i\}$ using the $\ell_{0.83}$-norm (the value 0.83 was chosen to keep the norms within the same range). Notice that, in both ways, what is being done is a sort of distortion in the distance space of the normalized eigenvalues. However, one can also notice that such distortions are different from each other, which is our main justification for the different performances between the NB and the iNB algorithm, as shown later on in this paper. The iNB operates with $\{l_i\}$ and uses the ℓ_u-norm with a value of u for optimized performance, whereas the NB operates with $\{\lambda_i^{(\mathrm{N})}\}$ and uses the ℓ_2-norm and a value of E for optimized performance. Moreover, the NB algorithm also demands the empirical choice of the parameter K, which does not exist in the iNB algorithm. In iNB, all the m normalized eigenvalues are considered in the search for the minimum norm.

Figure 3 shows the same parameters shown in Figure 2, but now for $p = 10$. This figure illustrates the effect of using the $\ell_{0.83}$-norm in a situation in which it is evident the failure of the NB algorithm. Notice in Figure 3(a) that the inflexion point is visible in the normalized eigenvalues $\{l_i\}$, whereas it is hardly

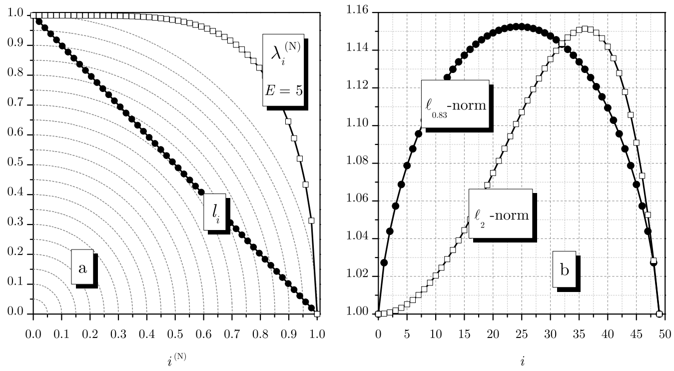

Fig. 2. (a) Idealized asymptotic behavior of the eigenvalues and indexes, for $p = 0$, normalized according to (1) (balls), and according to (1) and (2) (squares) for $E = 5$ (b) The $\ell_{0.83}$-norm for all $\{l_i\}$ (balls) and the Euclidian norm for all $\{\lambda_i^{(\mathrm{N})}\}$ (square), for $p = 0$

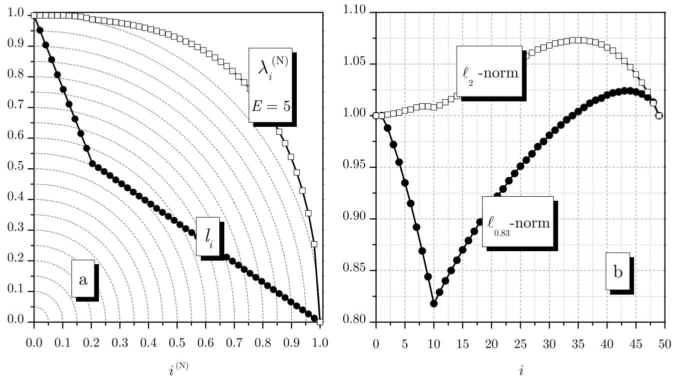

Fig. 3. (a) Idealized asymptotic behavior of the eigenvalues and indexes, for $p = 10$, normalized according to (1) (balls), and according to (1) and (2) (squares) for $E = 5$ (b) The $\ell_{0.83}$-norm for all $\{l_i\}$ (balls) and the Euclidian norm for all $\{\lambda_i^{(\mathrm{N})}\}$ (square), for $p = 10$

noticed in the normalized and nonlinear scaled $\{\lambda_i^{(\mathrm{N})}\}$. Notice also from Figure 3(a) that, if the normalized eigenvalues $\{l_i\}$ were used in the search for the minimum Euclidian norm, the estimated number of sources $\hat{p}$ would be around 13. In Figure 3(b) one can observe that the minimum $\ell_{0.83}$-norm indeed corresponds to $\hat{p} = 10$. However, the minimum Euclidian norm adopted by the NB algorithm corresponds to $\hat{p} \neq 10$, i.e. the NB algorithm would fail in estimating the number of sources in a situation similar to the one illustrated in Figure 3. The best value of u was found based on the results shown in Figure 4, which show the probability of correct detection (P_c) as a function of the norm value u combining the following sets of parameters: $m = 10, 15, 20, 30, 40$; $n = 50, 100, 250, 400$; $p = 2, 5, 8, 10, 20$ and $\mathrm{SNR} = 0, 2, 8, 10$ dB. From this figure it is apparent that the optimum value of u for maximum average P_c is around 0.55. Notice that finding u in the iNB algorithm is similar in process to finding E in the original NB algorithm.

The proposed iNB algorithm is summarized as follows. Notice that the parameter K of the NB algorithm is not used anymore, and that the iNB is less complex than the NB:

Algorithm 2 The Proposed iNB Algorithm

for $i = 1...m$ **do**
  Compute $\mathbf{\Lambda}_i = \begin{bmatrix} l_i & i^{(\mathrm{N})} \end{bmatrix}^T$ using (1)
end for
Compute $\hat{p} = \arg\min_i ||\mathbf{\Lambda}_i||_{0.55} - 1$

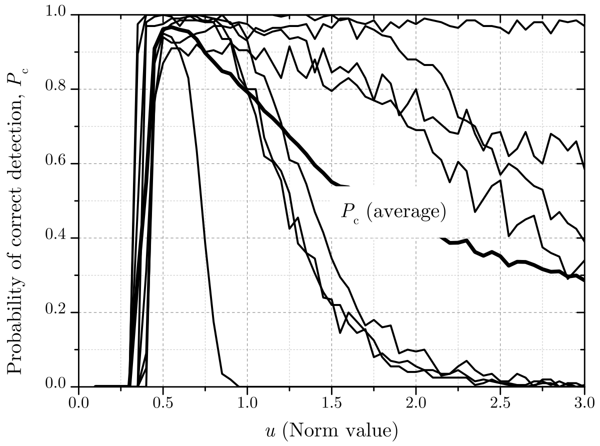

Fig. 4. Probability of correct detection as a function of the norm value u for several system parameters

VI. Numerical Results

In this section we present numerical results comparing the iNB, the NB, the MDL, the AIC, and the RMT-based estimators. Each point in the subsequent graphs was generated from 5000 Monte Carlo events. In each event a new matrix $\mathbf{Y} = \mathbf{HX}+\mathbf{V}$ was generated. The entries in $\mathbf{X}$ are independent and identically distributed (i.i.d.) complex Gaussian variates that simulate Gaussian-distributed and uncorrelated transmitted signals. This situation arises in wireless communications, where the envelope of most modulated signals are Gaussian-distributed. The entries in $\mathbf{H}$ are also i.i.d. complex Gaussian, to simulate a flat Rayleigh fading channel which is constant during each detection interval, changing independently from one interval to the next. The entries in $\mathbf{V}$ are also i.i.d. complex Gaussian, representing the additive thermal noise present at the receiver inputs. Assuming unitary total transmit power, the received SNR is given by $\mathrm{tr}[\mathbf{H}^\dagger\mathbf{H}]/\left(mp\sigma^2\right)$, where $\mathrm{tr}\left[\cdot\right]$ is the trace of the underlying matrix. To simulate an inaccurate noise variance estimate for the RMT-based algorithm, the noise variance is made a uniform random variable in $[\sigma^2-0.05\sigma^2, \sigma^2+0.05\sigma^2]$. This situation is denoted by RMT2 in the graphs. In order to simulate signal sources with variable power, we assume that the signal strengths are uniform random variables in $[0.4, 1]$. Additionally, for the RMT-based estimator the asymptotic false alarm (overestimation) probability is adjusted to $\alpha = 0.1\%$. In all figures, the probability of correct detection (P_c) and the probability of overestimation (P_oe) were considered as the performance measurements of the estimators.

Figure 5 shows results for P_c and P_oe as a function of the number of samples n collected by each sensor, for $m = 30$ sensors, $p = 3$ signals sources and SNR $= -10$ dB. Notice that the RMT-based estimator with perfectly known noise variance exhibits near consistency at large sample sizes, similar to the MDL (which is known to be consistent). However, the RMT-based estimator unveils severe inconsistence in the presence of noise uncertainty (RMT2 curves). The AIC estimator reaches higher detection performance with a smaller number of samples, but clearly it is inconsistent at large sample sizes, having a non-negligible P_oe for $n \gg 1$. It can be noted that the proposed iNB estimator performs better than the MDL estimator, and much better than the NB for all values of n, exhibiting a performance not far from the RMT (with perfectly known noise variance) and the AIC. The iNB also appears to be a consistent estimator, though we are not able to give a formal proof of this due to the empirical nature of the algorithm. It is worth mentioning that an increase in the SNR reduces the number of samples necessary for a target P_c (not shown here), which is a characteristic of all estimators.

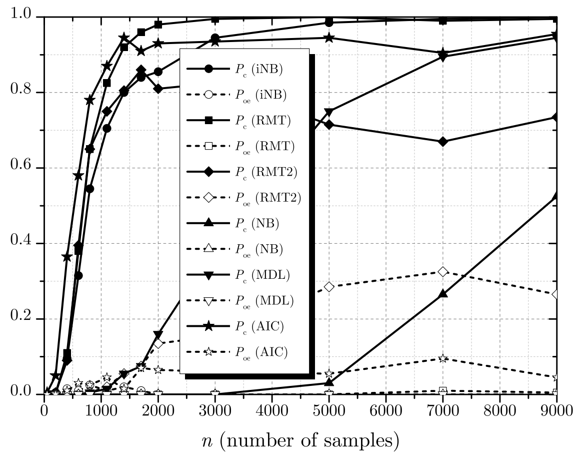

Fig. 5. P_c and P_oe against n for $p = 3$, $m = 30$, SNR $= -10$ dB

Figure 6 shows results for P_c and P_oe as a function of the SNR, for $p = 3$ signals sources, $m = 30$ sensors and $n = 1000$ samples. Again, this figure shows the superior detection capability of the iNB algorithm when compared with the NB and the MDL estimators, while exhibiting a performance very close to the RMT-based estimator (with perfect knowledge of the noise variance) and to the AIC in low SNR regimes. The advantages and drawbacks of the estimators as perceived from Figure 5 are also observable in Figure 6. Particularly, the MDL and the NB achieve $P_\mathrm{c} = 1$ for high values of SNR. The RMT-based estimator achieves $P_\mathrm{c} \approx 1$ in this situation; in fact it achieves $P_\mathrm{c} \approx 1 - \alpha$. However, when there is uncertainty in the noise variance (RMT2 curves), the RMT-based estimator produces a non-negligible P_oe, preventing it of achieving $P_\mathrm{c} \approx 1$. Due to inconsistence, the AIC also produces a non-negligible P_oe, leading to $P_\mathrm{c} \to (1 - P_\mathrm{oe})$ as SNR $\to \infty$.

Figure 7 depicts results for P_c and P_oe as a function of the number of sources p, considering $m = 30$ sensors, $n = 1000$ samples and SNR $= 8$ dB. Similar to the previous results,

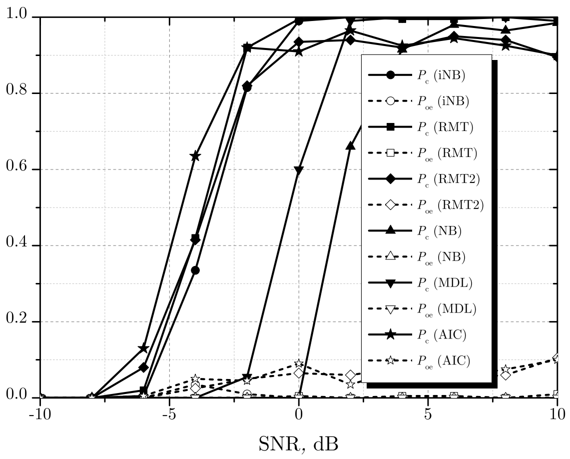

Fig. 6. P_{c} and P_{oe} against SNR for $m = 30$, $n = 1000$ and $p = 3$

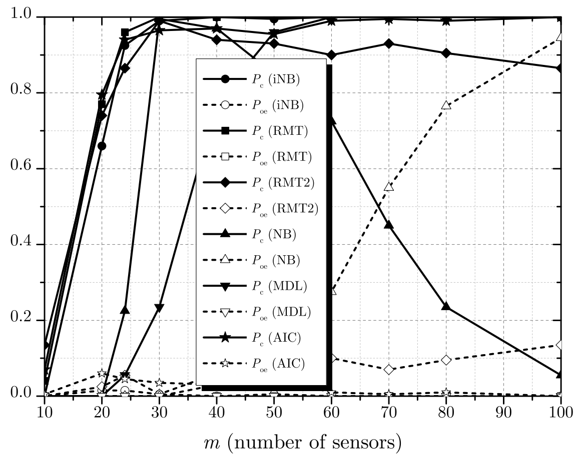

Fig. 8. P_{c} and P_{oe} against m for SNR $= -10$ dB, $n = 1000$ and $p = 4$

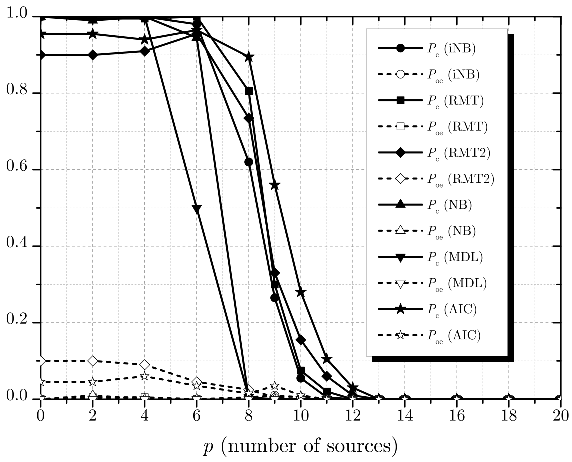

Fig. 7. P_{c} and P_{oe} against p for $m = 30$, $n = 1000$ and SNR $= 8$ dB

Figure 7 shows the superior performance of the iNB when compared with the MDL and NB. For a small number of sources the RMT-based estimator achieves $P_{\mathrm{c}} \approx 1$ when the noise variance is assumed to be perfectly known, but its performance is degraded if noise uncertainty takes place (RMT2 curves). The maximum estimated number of sources with high detection probability clearly varies from one estimator to another (and with the SNR; not shown), with advantage of the AIC, the RMT and the iNB, in this order.

In Figure 8, the performance of all estimators are plotted as a function of the number of sensors m, considering $n = 1000$ samples, SNR $= -10$ dB and $p = 4$ signals sources. Again we can see the superiority of the iNB when compared with the NB and the MDL. Notice in the case of the NB estimator that, as the value of m increases beyond 40, the probability of correct detection decreases. This is a consequence of the non optimality of the parameter K, which indicates that the NB algorithm suffers from a reduction in the capability of estimating a small number of sources when the number of sensors is large. The advantages and drawbacks of the AIC and the RMT-based estimators can again be observed, now with variable m.

VII. Conclusions

This paper proposed an empirical algorithm for estimating the number of sources of signals impinging on multiple sensors. The algorithm is a improvement of the one suggested in [4]. The improved NB (iNB) uses the $\ell_{0.55}$-norm as a means for classifying the eigenvalues. The main advantages of the iNB algorithm are: i) it is less complex than the NB algorithm; ii) it does not use a nonlinear scaling of the eigenvalues and does not need to set the empirical constant K that is crucial to the proper operation of the NB algorithm; iii) it outperforms one or more of the estimators MDL, AIC and RMT in several situations, and iv) it always outperforms the NB algorithm. In contrast to the AIC estimator and to the RMT-based estimator in the presence of noise variance uncertainty, the iNB showed to be consistent, although this was not supported by a formal proof. Being an empirical proposition, the iNB is not guaranteed to win, in all other situations, those estimators that were beat in the situations considered in this paper. We attest, however, that several other cases not shown here were analyzed, yet keeping the iNB algorithm in an advanced ranking when compared to the MDL and to the NB estimators, and in close ranking to the best results shown by the RMT and AIC estimators. Last but not least, it is worth mentioning that our results can be easily reproduced due to the low complexity of the iNB algorithm.

References

[1] 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.

[2] P. Stoica, and Y. Selén, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, pp. 36–47, Jul. 2004.

[3] S. Kritchman, and B. Nadler, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, Oct. 2009.

[4] P. M. Soares, "Um método empírico para estimação do número de fontes de sinais incidentes em um arranjo de sensores," Master's thesis, Instituto Nacional de Telecomunicações, Inatel, 2014, in Portuguese.