BROADBAND ML ESTIMATION UNDER MODEL ORDER UNCERTAINTY

Pei-Jung Chung[1)], *Mats Viberg*[2)], *Christoph F. Mecklenbräuker*[3)]

[1)]Institute for Digital Communications, The University of Edinburgh, UK
[2)]Dept. of Signals and Systems, Chalmers University of Technology, Sweden
[3)]Inst. of Communications and RF Engineering, Vienna University of Technology, Austria
p.chung@ed.ac.uk, viberg@chalmers.se, cfm@nt.tuwien.ac.at

ABSTRACT

The number of signals plays a crucial role in array processing. The performance of most direction finding algorithms relies strongly on a correctly specified number of signals. When this information is not available, conventional approaches apply information theoretic criteria or multiple hypothesis tests to simultaneously estimate model order and parameter. These methods are usually computationally intensive, since ML estimates are required for a hierarchy of nested models. In the previous work [1], we proposed a computationally efficient solution to avoid this full search procedure and demonstrated its feasibility by extensive simulations. Here we extend [1] to broadband data, and address issues unique to the broadband case. Our max-search approach computes ML estimates only for the maximally hypothesized number of signals, and selects relevant components through hypothesis testing. Another novelty of this work is the reduction of indistinguishable components caused by overparameterization. Our approach is based on the rank of the estimated steering matrix. Numerical experiments show that despite an unknown number of signals, the proposed method achieves comparable estimation and detection accuracy as standard methods, but at much lower computational expense.

Index Terms— broadband signals, maximum likelihood estimation, direction of arrival, unknown number of signals, overparameterized models

1. INTRODUCTION

Direction of arrival (DOA) estimation is a key issue in array processing. Among existing methods the maximum likelihood (ML) approach is characterized by excellent statistical properties and robustness against small sample numbers, signal coherence and closely located sources. In contrast to subspace methods, which are typically designed for narrow band signals, ML is applicable to both narrow band and broadband data.

Standard ML assumes the number of signals, m, to be known, and maximizes the likelihood function over an m-dimensional parameter space. When the number of signals is unknown, conventional approaches, such as methods based on information-theoretic criteria [2, 3], or multiple hypothesis test procedures [4, 5], jointly estimate the number of signals and DOA parameters. These methods are computationally intensive since ML estimates are required for a hierarchy of nested models.

In the previous work [1], we suggested a computationally attractive procedure that computes ML estimates *only* for the maximally hypothesized model. This max-search approach is motivated by the fact that the ML estimator derived from an overparameterized model contains relevant components that coincide with the true parameters [6]. The relevant components can be selected through simple hypothesis tests [1], or by their contribution to the likelihood function [7].

Here we extend [1] to broadband data and address related challenges. Unlike the narrow band case, the test statistic used to identify relevant components has an unknown distribution. We apply the Cornish-Fisher expansion [8] to approximate the test threshold. Moreover, we discuss the identification problem caused by overparameterization and introduce a criterion to reduce indistinguishable components.

The proposed procedure is different from the aforementioned methods [2, 3, 4, 5] in that it is aimed at extracting useful information about DOA parameters regardless of whether the number of signals is correctly specified or not. However, as a byproduct, the number of relevant components can be considered as an estimate for the actual number of signals. As it will be shown later, the proposed max-search procedure is computationally more attractive than the full search methods.

In the following, we give a brief description of the broadband signal model. In Section 3, we develop the broadband ML estimation procedure and derive the Fisher-Cornish approximation to the test threshold. Simulation results are presented in Section 4, whereas Section 5 concludes the paper.

2. PROBLEM FORMULATION

Consider an array of n sensors receiving m broadband signals emitted by far-field sources located at $\boldsymbol{\theta} = [\,\theta_1,\ldots,\theta_m]^T$. The array output $\boldsymbol{x}^{(k)}(t)$, $(t = 0,\ldots,T-1)$ within the kth observation interval (or *snapshot*) is short time Fourier-transformed

$$\boldsymbol{X}^{(k)}(\omega) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} w(t)\boldsymbol{x}^{(k)}(t)e^{-j\omega t} \qquad (1)$$

where $\{w(t)\}_{t=0}^{T-1}$ is a window function. For large number of samples T, we can describe the frequency domain data approximately by the following relation

$$\boldsymbol{X}^{(k)}(\omega) = \boldsymbol{H}_m(\omega;\boldsymbol{\theta}_m)\boldsymbol{S}^{(k)}(\omega) + \boldsymbol{U}^{(k)}(\omega) \qquad (2)$$

where the matrix $\boldsymbol{H}_m(\omega,\boldsymbol{\theta}_m) = [\boldsymbol{d}_1(\omega)\cdots\boldsymbol{d}_i(\omega)\cdots\boldsymbol{d}_m(\omega)] \in \mathbb{C}^{n\times m}$ consists of m steering vectors with the ith column $\boldsymbol{d}_i(\omega)$, corresponding to the ith incoming wave. The signal waveform $\boldsymbol{S}^{(k)}(\omega)$ is considered to be unknown and deterministic. The noise $\boldsymbol{U}^{(k)}(\omega)$ results only from sensors. According to the asymptotic Fourier transform theory, $\boldsymbol{X}^{(k)}(\omega_j)$, $(k=1,\ldots,K, j=1,\ldots,J)$ are independent, identically complex normally distributed with mean $\boldsymbol{H}_m(\omega_j;\boldsymbol{\theta})\boldsymbol{S}^{(k)}(\omega_j)$ and covariance matrix $\nu(\omega_j)\boldsymbol{I}$, where $\nu(\omega_j)$ is the unknown noise spectral parameter and $\boldsymbol{I}$ is an identity matrix of conformable dimension.

When the true number of signals is known, i.e. $m=m_0$, the ML estimate is obtained by minimizing the *negative* log-likelihood function:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_m &= \arg\min_{\boldsymbol{\theta}_m} l_T(\boldsymbol{\theta}_m),\\ l_T(\boldsymbol{\theta}_m) &= \sum_{j=1}^{J} \operatorname{tr}[(\boldsymbol{I}-\boldsymbol{P}(\omega_j;\boldsymbol{\theta}_m))\hat{\boldsymbol{R}}(\omega_j)] \qquad (3)\end{aligned}$$

where $\hat{\boldsymbol{R}}(\omega_j) = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}^{(k)}(\omega_j)\boldsymbol{X}^{(k)}(\omega_j)^H$ represents a non-parametric power spectral estimate of sensor outputs over K snapshots, and $\boldsymbol{P}(\omega_j;\boldsymbol{\theta}_m)$ is the projection matrix onto the subspace spanned by the columns of $\boldsymbol{H}_m(\omega_j;\boldsymbol{\theta}_m)$.

When m is unknown, the most successful approaches jointly estimate the DOA parameters and the model order. Given a maximally hypothesized number of signals, the ML estimate is derived for a hierarchy of nested models

$$\mathcal{M}_1 \subset \mathcal{M}_2 \subset \cdots \subset \mathcal{M}_M.$$

Such procedures are computationally demanding, due to the multi-dimensional nonlinear minimization required by each candidate model. In contrary to this full search idea, we suggest a computationally simple solution that considers the maximally hypothesized model $\mathcal{M}_M$ and extracts DOA information from the corresponding estimate.

3. BROADBAND ML ESTIMATION FOR UNKNOWN NUMBERS OF SIGNALS

From the analysis of ML estimation under misspecified numbers of signals [6], we know that for an overparameterized model, $m>m_0$, the ML estimate $\boldsymbol{\theta}_m^*$ obtained from large sample size is characterized by the following property

$$\operatorname{sp}(\boldsymbol{H}_m(\omega;\boldsymbol{\theta}_m^*)) \supset \operatorname{sp}(\boldsymbol{H}_{m_0}(\omega;\boldsymbol{\theta}_0)) \qquad (4)$$

where $\operatorname{sp}(\boldsymbol{H}_m(\omega;\boldsymbol{\theta}_m^*))$ and $\operatorname{sp}(\boldsymbol{H}_{m_0}(\omega;\boldsymbol{\theta}_0))$ denote the signal subspaces corresponding to $\boldsymbol{\theta}_m^*$ and $\boldsymbol{\theta}_0$, respectively. Furthermore, $\boldsymbol{\theta}_m^*$ contains m_0 components equal to those of the true parameter $\boldsymbol{\theta}_0$. Although this result is derived for narrow band signals, it is not difficult to extend it to the broadband case.

Motivated by the above observation, we suggest to compute the ML estimate for the maximal possible number of signals, $m=M$, and select relevant components that are associated with the true parameters. More specifically, the proposed algorithm minimizes the negative log-likelihood function (3) over an M-dimensional space

$$\hat{\boldsymbol{\theta}}_M = \arg\min_{\boldsymbol{\theta}_M} l_T(\boldsymbol{\theta}_M). \qquad (5)$$

Since $M \geq m_0$, the $M\times 1$ vector $\hat{\boldsymbol{\theta}}_M = [\hat{\theta}_1,\ldots,\hat{\theta}_M]^T$ contains more elements than the $m_0\times 1$ true parameter vector $\boldsymbol{\theta}_0$. The elements of $\hat{\boldsymbol{\theta}}_M$ that are associated with those of $\boldsymbol{\theta}_0$, are referred to as *relevant* components. The remaining $(M-m_0)$ components of $\hat{\boldsymbol{\theta}}_M$ are the *redundant* components.

Clearly, the key step in the proposed max-search procedure is identification of relevant components. In [7], these components are selected by thresholding the likelihood function, because the redundant components do not change the value of the likelihood function significantly. Here we extend the hypothesis test suggested in [1] to broadband signals to validate the ith component:

$$\begin{aligned} H_i &: \ \boldsymbol{X}(\omega) = \boldsymbol{H}_{M-1}(\omega;\tilde{\boldsymbol{\theta}}_i)\tilde{\boldsymbol{S}}_{M-1}(\omega)+\boldsymbol{U}(\omega)\\ A_i &: \ \boldsymbol{X}(\omega) = \boldsymbol{H}_M(\omega;\hat{\boldsymbol{\theta}}_M)\boldsymbol{S}_M(\omega)+\boldsymbol{U}(\omega) \qquad (6)\end{aligned}$$

where H_i and A_i represent the null hypothesis and the alternative, respectively. The $(M-1)\times 1$ vector

$$\tilde{\boldsymbol{\theta}}_i = [\hat{\theta}_1\cdots\hat{\theta}_{i-1}\,\hat{\theta}_{i+1}\cdots\hat{\theta}_M]^T \qquad (7)$$

contains all elements of $\hat{\boldsymbol{\theta}}_M$ except the ith component. The $n\times(M-1)$ matrix $\boldsymbol{H}_{M-1}(\tilde{\boldsymbol{\theta}}_i)$ contains steering vectors corresponding to the DOA parameters in $\hat{\boldsymbol{\theta}}_{M-1}$. Given the estimate $\hat{\boldsymbol{\theta}}_M$, the hypothesis test (6) decides whether the signal $S_i(\omega)$ associated with the ith component is zero. Note that $S_i(\omega)=0$ implies $\hat{\theta}_i$ does not correspond to any actual signal source.

The test statistic $T_i,\ i=1,2,\cdots,M$ derived from the likelihood ratio $l_T(\tilde{\boldsymbol{\theta}}_i)-l_T(\hat{\boldsymbol{\theta}}_M)$ is given by

$$\begin{aligned} T_i &= \frac{1}{J}\sum_{j=1}^{J}\log\left(\frac{\operatorname{tr}[(\boldsymbol{I}-\boldsymbol{P}(\omega_j;\tilde{\boldsymbol{\theta}}_i))\hat{\boldsymbol{R}}(\omega_j)]}{\operatorname{tr}[(\boldsymbol{I}-\boldsymbol{P}(\omega_j;\hat{\boldsymbol{\theta}}_M))\hat{\boldsymbol{R}}(\omega_j)]}\right)\\ &= \frac{1}{J}\sum_{j=1}^{J}\log\left(1+\frac{n_1}{n_2}F_i(\omega_j)\right). \qquad (8)\end{aligned}$$

The component $\hat{\theta}_i$ is *relevant* if H_i is rejected. Given the significance level α, H_i is rejected if T_i exceeds a threshold t_α. Consequently,

$$\hat{\theta}_i \text{ is relevant if } T_i \geq t_\alpha. \qquad (9)$$

The output of the algorithm is the *relevant* vector that contains all components

$$\hat{\boldsymbol{\theta}}_0 = [\hat{\theta}_{(1)},\cdots,\hat{\theta}_{(k)}]^T. \qquad (10)$$

As a byproduct of the proposed algorithm, the number of relevant components provides an estimate for the number of signals. However, we emphasize that the primary concern of the proposed approach is parameter estimation. Unlike methods designed for model order determination, whose performance is measured solely by correctness of the number of signals, the output of the proposed algorithm $\hat{\boldsymbol{\theta}}_0$ still contains useful information about the true parameters $\boldsymbol{\theta}_0$ even when the number of signals is misspecified.

3.1. Cornish-Fisher approximation for t_α

Under the null hypothesis H_i, the statistic $F_i(\omega_j)$ is F_{n_1,n_2}–distributed with the degrees of freedom n_1, n_2 given by [4]

$$n_1 = 3K, \quad n_2 = K(2n-2M-1). \qquad (11)$$

As (8) shows, in the narrow band case, $J = 1$, the test can be equivalently conducted with the statistic F_i whose distribution is known. However, for broadband signals, T_i has no closed form distribution under H_i. To overcome this difficulty, we suggest the Cornish-Fisher expansion to approximate the threshold t_α.

Note that the summands in (8) are i.i.d. samples from the random variable

$$\mathcal{T}_i = \log(1 + \frac{n_1}{n_2} F_i) \tag{12}$$

where F_i has an F_{n_1,n_2}-distribution. The i.i.d. property follows immediately from the asymptotic independence of the Fourier-transformed data $\boldsymbol{X}^{(k)}(\omega_j),\ j = 1, \cdots, J$. Since the degrees of freedom, n_1 and n_2, are the same for each F_i, $\mathcal{T}_i, i = 1, \cdots, M$ all have the same distribution. The threshold t_α needs to be computed only once for a pre-specified test level α.

Let $\mu_T = \mathsf{E}\,[\mathcal{T}_i]$ and $\sigma_T^2 = \mathsf{E}\,[\mathcal{T}_i - \mu_T]^2$ denote the mean and variance of $\mathcal{T}_i$. Then the distribution of the normalized test statistic

$$\bar{T}_i = \frac{T_i - \mu_T}{\sigma_T} \tag{13}$$

can be approximated by the Edgeworth expansion [9]. The normalized threshold $\bar{t}_\alpha$, the α-quantile of $\bar{T}_i$'s distribution, can be approximated by the Cornish-Fisher expansion [8] as follows:

$$\bar{t}_\alpha \approx z_\alpha + \frac{1}{\sqrt{J}} p_{11}(z_\alpha) + \frac{1}{J} p_{21}(z_\alpha), \tag{14}$$

where z_α is the α-quantile of the standard normal distribution and

$$\begin{aligned} p_{11}(z) &= \frac{\kappa_3}{6}(z^2 - 1), \\ p_{21}(z) &= \frac{\kappa_3^2}{18} z^3 (z^2 - 1) + z \left[\frac{\kappa_4}{24}(z^2 - 3) + \frac{\kappa_3^2}{72}(z^4 - 10z + 15) \right]. \end{aligned}$$

The rth cumulant κ_r of $\mathcal{T}_i$ depends on n_1, n_2 through the following formula

$$\kappa_r = (-1)^r \left[\Psi^{(r-1)}(\frac{n_1}{2}) - \Psi^{(r-1)}(\frac{n_1}{2} + \frac{n_2}{2}) \right], \tag{15}$$

where $\Psi^{(r-1)}(s) = \frac{\mathrm{d}^{\mathrm{r}}}{\mathrm{d}s^r}(\log \Gamma(s))$ denotes the rth derivative of the logarithm of the gamma function. The mean μ_T and the variance σ_T^2 are related to the cumulants as follows

$$\mu_T = \kappa_1, \qquad \sigma_T^2 = \kappa_2. \tag{16}$$

More details can be found in [10].

3.2. Identification of indistinguishable components

One situation may occur in the proposed algorithm is that the estimate $\hat{\boldsymbol{\theta}}_M$ contains the indistinguishable components, which leads to a rank deficient steering matrix $\boldsymbol{H}(\omega; \hat{\boldsymbol{\theta}}_M)$. Note that property (4) permits such parameters because the underlying model is overparameterized. In this case, both the relevant vector $\hat{\boldsymbol{\theta}}_0$ and the test (6) no longer provide satisfying results.

To overcome difficulties caused by overparameterization, we suggest to examine the smallest singular value $\hat{\sigma}_M$ of the steering matrix $\boldsymbol{H}(\omega; \hat{\theta}_M)$. If $|\sigma_M| \leq \delta$ where δ is a pre-specified small positive number, then the procedure computes the ML estimate for the next reduced model $\mathcal{M}_{M-1}$. If the resulting $\boldsymbol{H}(\omega; \hat{\theta}_{M-1})$ is full rank, then the relevant components are chosen from $\hat{\boldsymbol{\theta}}_{M-1}$. Otherwise, the same step is repeated for the next reduced model. In the simulation, we observe that this modified step is carried out only by 5% of all trials. The computational cost is still much lower than the full search approach.

4. SIMULATION

In the simulation, a uniform linear array of 10 sensors with inter-element spacing of half a wavelength is employed. The narrow band signals are generated by $m_0 = 2$ uncorrelated signals, located at $\boldsymbol{\theta}_0 = [28^\circ\ 36^\circ]$ with respect to array broadside, and of various strengths. The number of selected frequency bins is $J = 11$ and the number of snapshots is $K = 50$. The difference of signal strengths is $[\,1\ \ 0\,]$ dB, where 0 dB corresponds to the reference signal. The signal to noise ratio (SNR) varies from -10 to 10 dB in a 2 dB step. We consider two upper bounds on the number of signals: $M = 3$ and $M = 4$. The latter represents a larger mismatch in the model order. Each experiment consists of 200 Monte Carlo trials. The test level α is chosen to be 0.05. The parameter $\delta = 0.02$ is used to test whether the steering matrix is rank deficient or not.

Fig. 1 shows the sample mean of the relevant components $\hat{\theta}_{(1)}$, $\hat{\theta}_{(2)}$, respectively. For both $M = 3$ and $M = 4$, the bias is less than 0.06 degree over the entire SNR range. Since the results are obtained from finite samples, we conjecture that $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(2)}$ are asymptotically bias free.

The empirical variance of $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(2)}$ is presented in Fig. 2. For comparison, we also show results obtained from the no-mismatch case with $M = m_0 = 2$. All three curves decline with increasing SNR. The estimates obtained from $M = m_0 = 2$ have the smallest variance. For both relevant components $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(2)}$, $M = 4$ results in a larger variance than $M = 3$. This suggests that the variance increases with a larger degree of mismatch.

In Fig. 3, we compare the probability of correct detection of the proposed algorithm with the multiple hypothesis testing procedure [5]. By "correct detection", we mean that the number of relevant components equals the true number of signals. At the low SNR region, -10 to -4 dB, both $M = 3$ and $M = 4$ have a higher probability of correct detection. For SNR -6 to 10 dB, the multiple test procedure achieves 100% probability of correct detection, while $M = 3, 4$ increase from 90% to 95%. The curve associated with $M = 3$ shows a slightly higher probability of correct detection than that of $M = 4$.

In summary, simulation results demonstrate that relevant components convey useful information about the true parameters even when the correct number of signals is unknown or misspecified. The price paid for model uncertainty is increased variance. Furthermore, the comparison between the proposed algorithm and the full search procedure [5] showed that without computing ML estimates for each candidate model, the former has a lower SNR threshold and achieves good detection performance at high SNRs. The computational time required by the suggested max-search procedure is on average 35% less than the full search procedure.

5. CONCLUSION

We have developed a broadband ML estimation procedure for unknown numbers of signals. The suggested algorithm computes ML

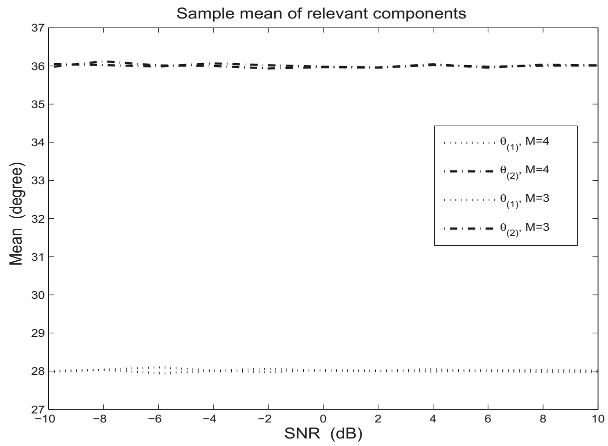

Fig. 1. Sample mean of relevant components. $M = 3, 4$. The true DOA parameter $\boldsymbol{\theta}_0 = [28^\circ\ 36^\circ]$, SNR $= [-10 : 2 : 10]$ dB.

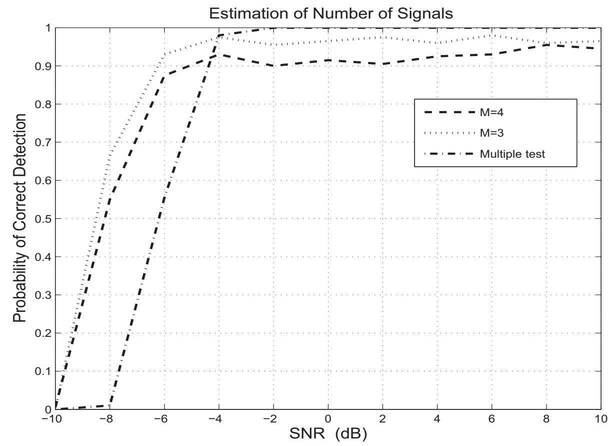

Fig. 3. Probability of correct detection. Comparison of multiple test procedure [5], $M = 3$, $M = 4$.

estimates only for the maximally hypothesized number of signals. The relevant components associated with the true DOA parameters are selected by the simple hypothesis tests. The test threshold is approximated by the Cornish-Fisher expansion. We also introduce a criterion to reduce indistinguishable components caused by overparameterization.

Compared to traditional methods for joint parameter estimation and signal detection, the proposed max-search approach avoids the full search process through a series of nested models, which leads to significant improvement in computational efficiency. Numerical results showed that the proposed algorithm achieves comparable estimation accuracy as the standard ML approach does. The number of signals can also be accurately determined by the number of relevant components. The proposed algorithm provides a computationally attractive alternative to existing methods, particularly in the low SNR region.

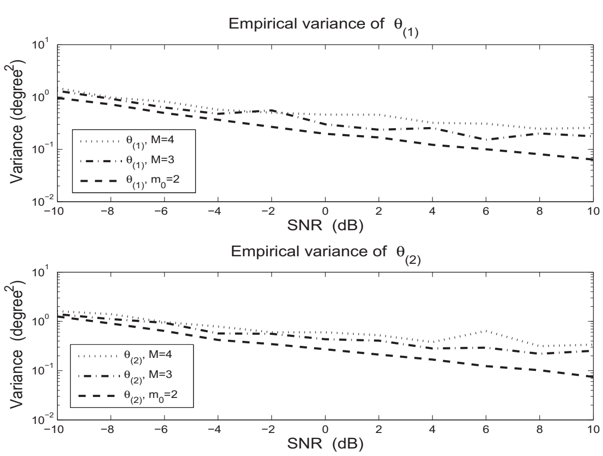

Fig. 2. Empirical variance of $\hat{\theta}_{(i)},\ i = 1, 2$.

6. REFERENCES

[1] P.-J. Chung and C. F. Mecklenbräuker, "Deterministic ML estimation for unknown numbers of signals," in *Proc. European Signal Processing Conference*, Lausanne, Switzerland, August 2008.

[2] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE ASSP*, vol. 37, no. 8, pp. 1190–1196, August 1989.

[3] M. Wong, Q. T. Zou, and J. P. Reilly, "On information theoretic criterion for determining the number of signals in high resolution array processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 11, pp. 1959–1971, November 1990.

[4] D. Maiwald, *Breitbandverfahren zur Signalentdeckung und –ortung mit Sensorgruppen in Seismik– und Sonaranwendungen*, Dr.–Ing. Dissertation, Dept. of Electrical Engineering, Ruhr–Universität Bochum, Shaker Verlag, Aachen, 1995.

[5] P.-J.Chung, J. F. Böhme, C. F. Mecklenbräuker, and A. O. Hero, "Detection of the number of signals using the Benjamini-Hochberg procedure," *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 2497–2508, June 2007.

[6] P.-J. Chung, "ML estimation under misspecified number of signals," in *Proc. of the 39th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Oct 30 - Nov 2, 2005.

[7] P.-J. Chung, "Robust ML estimation for unknown numbers of signals," in *Proc. European Signal Processing Conference*, Poznań, Poland, September 2007.

[8] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, Dover Publications, 9th edition, 1972.

[9] P. Hall, *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York, 1992.

[10] Christoph F. Mecklenbräuker, Peter Gerstoft, Johann F. Böhme, and Pei-Jung Chung, "Hypothesis testing for geoacoustic environmental models using likelihood ratio," *J. Acoust. Soc. Am.*, vol. 105(3), pp. 1738–1748, March 1999.