



Akademie věd České republiky

Disertace
k získání vědeckého titulu “doktor věd”
ve skupině věd FYZIKÁLNĚ-MATEMATICKÝCH

CRAMÉR–RAO LOWER BOUNDS IN SIGNAL PROCESSING APPLICATIONS
název práce

Komise pro obhajobu doktorských disertací v oboru: INFORMATIKA A KYBERNETIKA

Jméno uchazeče: ING. PETR TICHAVSKÝ, CSc.

Pracoviště uchazeče: ÚSTAV TEORIE INFORMACE A AUTOMATIZACE V.V.I, AV ČR

Místo a datum: PRAHA, 2015

Contents

1	Introduction	4
2	Research Articles in the Dissertation	5
2.1	CRLB for the Adaptive Harmonic Retrieval [1]	5
2.2	CRLB for Nonlinear Filtering [2], [3]	5
2.3	CRLB for DOA Estimation Using a Single Hydrophone [4]	6
2.4	CRLB for DOA Estimation Using a Towed Array [5]	6
2.5	CRLB for Independent Component Analysis [6]	6
2.6	CRLB for Canonical Polyadic Tensor Decomposition [7]	7
3	Conclusions	7
4	Reprints	9
4.1	Paper No. 1 – Posterior Cramér–Rao bounds for adaptive harmonic retrieval	10
4.2	Paper No. 2 – Posterior Cramér–Rao bounds for discrete–time nonlinear filtering	14
4.3	Paper No. 2 – Posterior Cramér–Rao bounds for discrete–time nonlinear filtering	25
4.4	Paper No. 4 – Near-Field/Far-Field Azimuth & Elevation Angle Estimation Using a Single Vector-Hydrophone	39
4.5	Paper No. 5 – Quasi-fluid-mechanics-based quasi-Bayesian Cramer–Rao bounds for deformed towed-array direction finding	52
4.6	Paper No. 6 – Performance Analysis of the FastICA Algorithm and Cramér– Rao Bounds for Linear Independent Component Analysis	64
4.7	Paper No. 7 – Cramér–Rao-Induced Bounds for CANDECOMP/PARAFAC tensor decomposition	81

Resumé

The Cramér–Rao Lower Bound (CRLB) is a lower bound on the covariance matrix of the error of unbiased vector parameter estimators. It represents a bound on an information content of data about an unknown parameter in a statistical model of the given data. CRLB is a classical tool stemming from the works of Cramér, Rao [8], [9], and other researchers in 1950’s. CRLB has many extensions and modifications: a Bayesian CRLB for random parameters [10], [11], a hybrid CRLB for a mixture of random and deterministic parameters [12], a CRLB for biased estimates [13].

Several other more accurate lower bounds were derived, e.g. Barankin bound [14], Bhattacharyya bound [15], and (in the Bayesian context) a Ziv-Zakai bound [16], to name a few. However, CRLB remains the most frequently used lower bound in a very wide variety of signal processing problems thanks to its mathematical tractability. The bound is used as a performance gauge for all existing parameter estimators, indicating whether the estimators utilize the available information about the estimated parameter efficiently or not, and in what extent. The CRLB itself is subject of a theoretical research up to now, see, e.g., [13].

The dissertation consists of seven scientific articles on computing different variants of the CRLB in different applications:

- [1] P. Tichavský, “Posterior Cramer-Rao bounds for adaptive harmonic retrieval”, *IEEE Trans. on Signal Processing* vol. 43, no.5, pp. 1299-1302, May 1995.
- [2] P. Tichavský, C. Muravchik and A. Nehorai, “Posterior Cramér–Rao bounds for discrete-time nonlinear filtering”, *IEEE Tr. on Signal Processing*, vol. 46, no. 5, pp. 1386-1396, May 1998.
- [3] M. Šimandl, J. Královec and P. Tichavský, “Filtering, predictive, and smoothing Cramér–Rao bounds for discrete-time nonlinear dynamic systems”, *Automatica*, vol. 37, no. 11, pp. 1703-1716, November 2001.
- [4] P. Tichavský, K.T. Wong and M.D. Zoltowski, “Near-Field/Far-Field Azimuth & Elevation Angle Estimation Using a Single Vector-Hydrophone”, *IEEE Tr. on Signal Processing*, vol. 49, no. 11, pp. 2498-2510, November 2001.
- [5] P. Tichavský and K.T. Wong, “Quasi-fluid-mechanics-based quasi-Bayesian Cramer- Rao bounds for deformed towed-array direction finding”, *IEEE Tr. on Signal Processing*, vol. 52, no.1, pp. 36-47, January 2004.
- [6] P. Tichavský, Z. Koldovský, and E. Oja, “Performance Analysis of the FastICA Algorithm and Cramér-Rao Bounds for Linear Independent Component Analysis”, *IEEE Tr. on Signal Processing*, vol. 54, no. 4, pp.1189–1203, April 2006. Corrections: vol. 56, no. 4, pp. 1715–1716, April 2008.
- [7] P. Tichavský, A.H. Phan, Z. Koldovský, “Cramér-Rao-Induced Bounds for CANDECOMP/PARAFAC tensor decomposition”, *IEEE Trans. Signal Processing*, vol. 61, no. 8, pp. 1986–1997, April 2013.

These papers include the Bayesian CRLB derived for the recursive system identification, and the deterministic and the hybrid CRLB for the recursive sinusoidal frequency estimation, for nonlinear filtering, for the direction-of-arrival estimation, for the independent component analysis, and for the canonical polyadic tensor decomposition, respectively. Although the concept of the theory of the CRLB is well known, in practical applications its computation might be quite complicated, and the computation of this bound in particular applications

is novel and important contribution to understanding the relation between the data and the estimated parameter. Sometimes, analysis of the CRLB leads to a derivation of new estimators. For example, the performance analysis of the algorithm FastICA for the independent component analysis and the computation of the corresponding CRLB [6] has led to a derivation of the algorithm EFICA [18].

1 Introduction

Classical Cramér-Rao lower bound is a bound on covariance matrix of error of *unbiased* estimates of an unknown *deterministic* parameter.

Assume we are given a family of distribution functions of the N -dimensional vector X , indexed by a vector of parameters θ . X represents the random data and θ is the unknown deterministic parameter. The range Θ of θ is assumed to be a subset of \mathbb{R}^M , so θ is a real-valued vector of dimension M . Let $f_\theta(X)$ be the probability density of X given $\theta \in \Theta$. Assume that such probability density exists and is twice differentiable with respect to θ . The Fisher information, if exists, is defined as

$$F(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2 \log f_\theta(X)}{\partial \theta \partial \theta^T} \right] \quad (1)$$

where \mathbb{E}_θ is the expectation operator with respect to the density $f_\theta(X)$. Let $\hat{\theta}(X)$ be an *unbiased* estimate of θ , and assume that

1. support of the density $f_\theta(X)$, i.e. the set of $X \in \mathbb{R}^N$, where $f_\theta(X) > 0$, is independent of θ
2. $\forall \theta \in \Theta; \forall m = 1, \dots, M; \quad 0 = \frac{\partial}{\partial \theta_m} \int f_\theta(Y) dY = \int \frac{\partial f_\theta(Y)}{\partial \theta_m} dY$
3. $\forall \theta \in \Theta; \forall m = 1, \dots, M; \quad \frac{\partial}{\partial \theta_m} \int \hat{\theta}(Y) f_\theta(Y) dY = \int \hat{\theta}(Y) \frac{\partial f_\theta(Y)}{\partial \theta_m} dY$
4. $F(\theta)$ in (1) exists and is *invertible*.

Then, the celebrated Cramér-Rao inequality holds,

$$\mathbb{E}_\theta \left[(\hat{\theta}(X) - \theta)(\hat{\theta}(X) - \theta)^T \right] \geq [F(\theta)]^{-1}. \quad (2)$$

The matrix inequality in (2) means that the difference between the left-hand side and right-hand side of (2) is a positive semi-definite matrix.

The classical CRB is very well known. For example, it is known that equality in the CRB inequality can be achieved if and only if the probability distribution $f_\theta(X)$ belongs to the family of exponential distributions. If a maximum likelihood estimator of parameter θ exists, its variance attains the CRLB asymptotically.

In comparison to the classical CRLB, the Bayesian CRLB is much less frequently studied. The set-up is different. It is assumed that the parameter θ is random, and a joint probability density $f_{\theta, X}$ of the pair (θ, X) exists. The Cramer-Rao inequality reads

$$\mathbb{E} \left[(\hat{\theta}(X) - \theta)(\hat{\theta}(X) - \theta)^T \right] \geq F^{-1} \quad (3)$$

where the expectation is taken with respect to the pair (θ, X) , and F is the information matrix defined as

$$F = -\mathbb{E} \left[\frac{\partial^2 \log f_{\theta, X}(\theta, X)}{\partial \theta \partial \theta^T} \right]. \quad (4)$$

Indeed, in the case of random parameter θ , the optimum estimator $\hat{\theta}(X)$ that minimizes the left-hand side of (2) exists: it is the conditional mean of θ given the data X . Covariance matrix of this conditional mean is, in general, a tighter bound on covariance of all other estimators than the inverse of the Fisher information matrix in (4). A disadvantage of the exact (tight) bound is that it may not be mathematically tractable, unlike the CRLB.

The technical assumptions of the CRLB in (3) to be valid are different than the assumptions of the classical CR inequality. First of all, the estimators $\hat{\theta}(X)$ need not be unbiased, their bias can be nonzero, and the bias conditioned by given θ ,

$$B(\theta) = \int (\hat{\theta}(X) - \theta) f_{x|\theta}(x|\theta) dX \quad (5)$$

obeys the condition

$$\lim_{\theta_m \rightarrow \infty} B(\theta) f_{\theta}(\theta) = \lim_{\theta_m \rightarrow -\infty} B(\theta) f_{\theta}(\theta) = 0 \quad (6)$$

for $m = 1, \dots, M$.

The model of the parameter θ can also be hybrid: a part of θ can be deterministic and another part random [12]. A typical example is a direction-of-arrival (DOA) estimation using the sensor array. In this application it is assumed that there is a number of plane acoustic or electromagnetic waves impinging on an array of sensors. The main task is the estimation of directions of arrival of the plane waves, which are the main deterministic parameters of the model. Usually, there are some other deterministic nuisance parameters as well, e.g. signal amplitudes, phases, etc. On top of it, there might be random parameters that describe random fluctuations of the sensor position and the orientation from their nominal position, random fluctuations of the sensor gains, and others. Although the nuisance parameters need not be estimated, absence of their knowledge and the presence of the random parameters influence the estimation of the parameters of the interest and its accuracy. An example of the analysis of the model uncertainty can be found in the papers [4] and [5].

2 Research Articles in the Dissertation

2.1 CRLB for the Adaptive Harmonic Retrieval [1]

The first paper [1] deals with the computation of CRLB for the adaptive harmonic retrieval. Here, received data is modeled as a cisoid (complex-valued sinusoid) which has a frequency that randomly drifts in the interval $(0, 2\pi)$. Frequency increments are modeled as independent Gaussian random variables with the zero mean and a small variance. In addition, the data contain a complex-Gaussian random noise. The goal is, given variance of the frequency increments and variance of the additive noise, to estimate the lowest possible mean square error of a tracking algorithm estimating the instantaneous frequency. Here, “tracking” means a recursive estimation of the instantaneous frequency at time t given the history of the signal up to time t . The estimated parameter (the instantaneous frequency) is random, therefore a Bayesian CRLB is derived. We computed the bound in a closed form and showed that the bound is attained by certain frequency tracking algorithms [17]. These algorithms were proved to be statistically efficient in this way.

2.2 CRLB for Nonlinear Filtering [2], [3]

The second paper [2] (from 1998) is a generalization of the former one to a very general scenario of nonlinear filtering. This paper became very popular in the system identification community and received hundreds of citations in SCI. Assume that we are given a nonlinear

system represented by a state vector x_n which evolves in time through a possibly nonlinear function f_n as $x_{n+1} = f_n(x_n, w_n)$, where w_n is a random Gaussian noise that enters in the state evolution equation. The function can be, for example, linear or simply additive, $x_{n+1} = x_n + w_n$. The challenge is that we cannot observe the state x_n directly but only through a nonlinear observation, as $y_n = g_n(x_n, v_n)$, where g_n is nonlinear function and v_n is another random noise that enters in the system. In the special case, the latter noise can be additive, $y_n = g_n(x_n) + v_n$. The goal is to derive a CRLB on covariance matrix of errors $\hat{x}_n - x_n$ where \hat{x}_n is a function of the observations up to time n , i.e. $\dots y_{n-2}, y_{n-1}, y_n$. In this paper, the bound is derived in a recursive form. It has been found useful in many applications. The nonlinear filtering algorithms are often realized through particle filters. As the computational power of modern computers grows, the particle filters become more popular. It is, however, not known a priori, how many particles have to be used to get close to the best possible performance. The CRLB helps to answer this question.

The following paper [3] by Královec, Šimandl and Tichavský derives a similar CRLB for nonlinear prediction and smoothing. Given the measurements $\dots y_{n-2}, y_{n-1}, y_n$, the goal is to estimate x_{n+m} with $m > 1$ (prediction) or x_{n-m} (smoothing).

2.3 CRLB for DOA Estimation Using a Single Hydrophone [4]

An application of CRLB in underwater statistics is studied in [4]. In particular, an accuracy of Direction-of-Arrival (DOA) estimation using a single vector hydrophone is analyzed. A vector hydrophone is composed of two or three spatially co-located but orthogonally oriented velocity hydrophones plus another optional co-located pressure hydrophone. It is no longer a tracking scenario, but a stationary scenario with an unknown deterministic parameter. The CR bound is used to compare performance of complete and incomplete vector hydrophones. In the latter case, one or more velocity hydrophones are absent. The analysis helps to quantify the tradeoff between the estimation accuracy and complexity (cost) of the hardware.

2.4 CRLB for DOA Estimation Using a Towed Array [5]

The fifth paper [5] studies the accuracy of the DOA estimation using an array of classical hydrophones that are placed on a cable towed by a vessel. The shape of the array is subject to random deformations due to the towing vessel's varying speed and transverse motion, by the array's non-neutral buoyance and nonuniform density changes, and by hydrodynamic effects plus oceanic swells and currents. The inaccuracy of the array geometry is modeled using physical considerations. In particular, transverse deformation/vibration of a thin flexible cylinder, towed by a vessel, is known to obey a fourth-order partial differential equation known as the Paidoussis equation. This equation describes the mechanical propagation of the array-deformation down the array's length. The equation was used to derive the covariance matrix of random deviations of the array from its nominal position, which is further used in expressions for CRLB for the DOA estimation using a randomly curved array.

2.5 CRLB for Independent Component Analysis [6]

The sixth paper [6] is related to the independent component analysis (ICA) and the blind source separation. In the paper we study the task of analysis of an $N \times N$ linear mixture of N independent non-stationary signals. Each of the signals is modeled as a series of independent realizations of a random variable having a non-Gaussian distribution¹. The task is to find

¹To be accurate, at most one signal in the mixture is allowed to have Gaussian distribution, the other signals must be non-Gaussian.

a mixing matrix of the size $N \times N$ that represents the mixture without any other prior information about the separated signals. In the literature several popular algorithms to solve the ICA problem were proposed. In the paper, one of the most successful ones (FastICA) is studied and its performance is analyzed in terms of the Interference-to-Signal Ratio (ISR) of the separated signals. In the same paper, the theoretical CRLB-based bound on accuracy of the separation is derived and compared to performance of FastICA. The performance and the CRLB depend namely on the probability distributions of the separated signals and their length. The analysis was used to propose a novel variant of FastICA, called EFICA [18].

2.6 CRLB for Canonical Polyadic Tensor Decomposition [7]

The seventh paper [7] is related to a different area (tensor decompositions), but can be related to the ICA model in a sense. The statistical estimation problem is related to stability of canonical-polyadic (CP) tensor decomposition. The word “tensor” here means a rectangular array of real or complex numbers. In general it can have a size $d_1 \times d_2 \times \dots \times d_N$, where N is called the tensor order. Each element of the tensor has N indices, say t_{i_1, \dots, i_N} . The goal of the CP decomposition is to find the smallest possible integer R (called rank of the tensor) and N matrices (called factor matrices) \mathbf{A}_j , $j = 1, \dots, N$ of the size $d_j \times R$ with elements $a_{j,i,r}$, $i = 1, \dots, d_j$, $r = 1, \dots, R$, and R scalars $\lambda_1, \dots, \lambda_R$ such that

$$t_{i_1, \dots, i_N} = \sum_{r=1}^R \lambda_r a_{1,i_1,r} \cdot \dots \cdot a_{N,i_N,r}$$

for all $i_j = 1, \dots, d_j$, $j = 1, \dots, N$. Without any loss in generality it can be assumed that all columns of all factor matrices have the unit Euclidean norm.

The CP decomposition, also known under the acronyms PARAFAC or CANDECOP, was found useful namely in several applications as chemometrics, biomedical signal processing, and others.

The CRLB derived in the paper helps to study the stability of the CP decomposition. It reveals how the small perturbations of the tensor elements translate in the accuracy of the factor matrices’ estimates. The result has led to derivation of a novel CP decomposition algorithm for high-order tensors, see [19].

3 Conclusions

The presented dissertation summarizes author’s contribution to different areas of statistical signal processing in the last twenty years. The underlying theme linking the collection of seven publications that comprise the dissertation is the computation of the Cramér–Rao bound. The computation of the bound has helped to understand the relation between the available data and its information content about estimated parameters of the models in the sinusoidal frequency estimation with slowly varying parameters, in nonlinear filtering, smoothing and tracking, in the underwater DOA estimation, in the independent component analysis and in the canonical polyadic tensor decomposition. A high interest of the research community in these areas is proved by a significant impact of the presented collection of articles, which is about 589 citations according to the Thomson Reuters citation index (with self-citations included, for simplicity).

References

- [8] H. Cramér, “A Contribution to the Theory of Statistical Estimation,” *Aktuariestidskrift*, vol. 29, pp. 458–463, 1946.
- [9] C.R. Rao, “Information and the Accuracy Attainable in the Estimation of Statistical Parameters”, *Bull. Calcutta Math. Soc.*, 37, pp. 465–471, 1978.
- [10] H. L. van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968.
- [11] H. L. van Trees and K. L. Bell (Eds), *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, Wiley, 2007.
- [12] Y. Rockah and P. Schultheiss, “Array shape calibration using sources in unknown locations—Part I: Far-field sources”, *IEEE Tr. Signal Processing* 35, no.3, pp. 286–299, 1987.
- [13] Y. C. Eldar, “Rethinking Biased Estimation: Improving Maximum Likelihood and the Cramér-Rao Bound”, *Foundations and Trends in Signal Processing*, vol. 1, no.4, pp. 305–449, 2008.
- [14] E.W. Barankin, “Locally Best Unbiased Estimates”, *Ann. Math. Statist.*, vol. 20, pp.477–501, 1949.
- [15] A. Bhattacharyya, “On some Analogues to the Amount of Information and Their Uses in Statistical Estimation”, *Sankhya* 8, pp. 1–14, 1946.
- [16] J. Ziv, M. Zakai, M., “Some lower bounds on signal parameter estimation”. *IEEE Transactions on Information Theory*, vol. 15, no.3, pp. 386–391, 1969.
- [17] P. Tichavský and P. Händel, “Two algorithms for adaptive retrieval of slowly varying multiple cisoids in noise”, *IEEE Trans. on Signal Processing*, vol. 43, no.5, pp. 1116–1127, May 1995.
- [18] Z. Koldovský, P. Tichavský, and E. Oja, “Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cramér-Rao Lower Bound”, *IEEE Tr. Neural Networks*, vol. 17, no. 5, pp. 1265–1277, September 2006.
- [19] A.H. Phan, P. Tichavský and A. Cichocki, “CANDECOMP/PARAFAC Decomposition of High-order Tensors Through Tensor Reshaping”, *IEEE Tr. Signal Processing*, vol. 61, no. 19, October 2013, pp. 4847–4860.

4 Reprints

Posterior Cramér–Rao Bound for Adaptive Harmonic Retrieval

Petr Tichavský

Abstract—The problem of adaptive parameter estimation for a single nonstationary noisy cisoid, where the sinusoidal frequency evolves according to Gaussian random walks, is studied. The lower bound on the minimum mean-square estimation error, which was derived by van Trees, is evaluated for the problem. It is shown that two estimation methods attain this lower bound.

I. INTRODUCTION

The problem of recursive estimation of parameters (frequencies and amplitudes) of multiple sine waves in noise has received a great deal of interest in the literature. The problem can be alternatively formulated as adaptive line enhancement of the signal or as elimination of a sinusoidal interference. Several algorithms for solving this task have been derived, and their properties have been analyzed. The most important procedures include the adaptive notch filter (ANF) in many variants and modifications; cf. [1], [2] and the references therein. Another method is the multiple frequency tracker (MFT) [3], [4].

For stationary signals, the minimal variance of any unbiased estimator is given by the Cramér–Rao lower bound (CRB); cf. [5]. In particular, for the above-mentioned estimation problem, the bound is proportional to $1/n^3$ and $1/n$ (where n denotes the size of the data sample) for the frequency and the amplitude estimates, respectively; cf. [6]. Thus, for a statistically efficient recursive estimator, the variance of the estimation error tends to zero as $n \rightarrow \infty$, implying that the update in the parameter estimates tends to zero or, loosely speaking, that the algorithm's gain tends to zero.

In on-line applications, however, it is often desired that the gain be nonzero in order to preserve the algorithm's tracking ability under sudden parameter changes or drift in the measured data. In these cases, the variance of the parameter estimates depends on the effective memory length of the algorithm, and thus, for sufficiently long data, it no longer depends on the actual length of the signal processed. Some general results on the asymptotic distribution of exponentially weighted prediction error estimators are given in [7]. The tracking scenario, where the drift in the parameter vector is modeled as a random walk, is studied in [8].

The case of slow drift in the sinusoidal amplitudes and frequencies is highly relevant in many signal processing applications, where the classical cisoid-in-noise model is too conservative to accurately describe the underlying data generation process. Some analytical expressions for the variance of the frequency estimation error for different algorithms have, among other things, been derived in [4], [9], and [10]. In common for the variance expressions in the papers cited above is that the frequency estimation variance consists of two terms: one corresponding to the additive measurement noise and one term related to the random walk modeled frequency drift. A question of considerable theoretical interest concerns the lower bound on the estimation accuracy for this problem.

Manuscript received June 20, 1994; revised November 28, 1994. This work was supported by Grant 201/93/0233 of the Grant Agency of the Czech Republic. The associate editor coordinating the review of this paper and approving it for publication was Dr. R. D. Preuss.

The author is with Institute of Information Theory and Automation, Prague, Czech Republic.

IEEE Log Number 9410289.

The assumption that the parameters of the signal evolve according to a random walk basically means that these parameters are considered to be random. This implies that the classical CRB is not applicable in this case. Fortunately, a lower bound on the minimum mean-square error in estimating a random parameter also exists; see pp. 72–73 of [5]. Since the bound has, in principle, similar form to the standard CRB, we call it, for easy reference, the posterior Cramér–Rao bound (PCRB).

In this correspondence, we consider a signal consisting of a single nonstationary complex-valued sinusoid (cisoid), embedded in additive Gaussian white noise with zero mean and a known variance. It is assumed that the instantaneous frequency of the cisoid evolves according to random walk, namely, that the increments in the frequency are independent Gaussian random variables with zero mean and known variances. In addition, it is assumed that the sinusoid has a known constant magnitude, initial frequency, and phase. For this model, we calculate the PCRB for estimation of the instantaneous frequency at time instant $n > 0$ and find $\lim_{n \rightarrow \infty}$ PCRB, which we call LPCRB.

The LPCRB is compared with the analytical expression characterizing the large sample performance of the ANF and MFT in estimating the slowly varying frequency in the given model (cf. [4]). It is shown that both the procedures attain the LPCRB so that they achieve the minimum attainable mean square estimation error.

II. POSTERIOR CRAMÉR-RAO BOUND

In the classical parameter estimation, it is common to consider an observation (random) vector $\mathbf{x} = (x_1, \dots, x_n)^T$ with joint probability density $p_{\mathbf{x}|\theta}(\mathbf{X}|\Theta)$, where $\theta = (\theta_1, \dots, \theta_r)^T$ is the parameter vector. In this correspondence, it is assumed that θ is a random vector with given *a priori* probability function $p_\theta(\Theta)$, as in the Bayesian statistics. Then, the joint probability function of the pair (\mathbf{x}, θ) is given as $p_{\mathbf{x},\theta}(\mathbf{X}, \Theta) = p_{\mathbf{x}|\theta}(\mathbf{X}|\Theta)p_\theta(\Theta)$.

Consider the task of estimating the parameter θ by means of a function of \mathbf{x} , $\hat{\theta} = g(\mathbf{x})$, and assume first that θ is a scalar $r = 1$. It is well known, cf. [5], that the best estimator of θ in the sense of the least mean square error is the conditioned expectation $E(\theta|\mathbf{x})$. In other words, $\text{var}\{E(\theta|\mathbf{x})\}$ is the universal lower bound for mean square error of estimating θ by functions of \mathbf{x} .

However, it may occur that the estimator $E(\theta|\mathbf{x})$ is represented by a complex expression, and thus, it may be difficult to calculate its variance. The alternative lower bound, which was derived in [5] and denoted here by PCRB, can sometimes be more easily evaluated.

In the case of the *multiple* parameter estimation ($r > 1$) the bound has the form

$$\mathbf{V} \triangleq E\left\{[g(\mathbf{x}) - \theta][g(\mathbf{x}) - \theta]^T\right\} \geq \mathbf{J}^{-1} \quad (1)$$

where \mathbf{J} is the $r \times r$ matrix with the elements

$$\mathbf{J}_{ij} = E\left[-\frac{\partial^2 \log p_{\mathbf{x},\theta}(\mathbf{X}, \Theta)}{\partial \Theta_i \partial \Theta_j}\right] \quad i, j = 1, \dots, r \quad (2)$$

and the inequality in (1) means that the difference $\mathbf{V} - \mathbf{J}^{-1}$ is a positive semidefinite matrix. Note that \mathbf{J} is a counterpart to the Fisher information matrix; for easy reference, it is called the *posterior* information matrix. The matrix can be decomposed into two parts:

$$\mathbf{J} = \mathbf{J}_D + \mathbf{J}_F \quad (3)$$

where \mathbf{J}_D stands for the standard Fisher information matrix, representing the information obtained from the data, and the matrix \mathbf{J}_F

represents the *a priori* information matrix

$$\mathbf{J}_{D_{ij}} = \mathbf{E} \left[-\frac{\partial^2 \log p_{\mathbf{x}|\theta}(\mathbf{X}|\Theta)}{\partial \Theta_i \partial \Theta_j} \right] \quad i, j = 1, \dots, r \quad (4)$$

$$\mathbf{J}_{P_{ij}} = \mathbf{E} \left[-\frac{\partial^2 \log p_{\theta}(\Theta)}{\partial \Theta_i \partial \Theta_j} \right] \quad i, j = 1, \dots, r. \quad (5)$$

It is also worth noting here that, in contrast with the standard CRB, $g(\mathbf{x})$ need not be an *unbiased* estimator of θ .

The full statement of the theorem follows.

Theorem 1: Let θ be a random r -dimensional vector, and let \mathbf{x} , which is the observation vector, have the joint probability function $p_{\mathbf{x},\theta}(\mathbf{X}, \Theta)$. Let $\hat{\theta} = g(\mathbf{x})$ be an arbitrary estimator of the parameter θ . Assume the following

- 1) The matrix \mathbf{J} , which is defined in (2), exists and is regular.
- 2) $(\partial p_{\mathbf{x},\theta}(\mathbf{X}, \Theta)/\partial \Theta_i)$ is absolutely integrable with respect to \mathbf{X} and Θ for all $i = 1, \dots, r$.
- 3) $(\partial^2 p_{\mathbf{x},\theta}(\mathbf{X}, \Theta)/\partial \Theta_i \partial \Theta_j)$ is absolutely integrable with respect to \mathbf{X} and Θ for all $i, j = 1, \dots, r$.
- 4) The conditional expectation of the error, given \mathbf{X} , is

$$B(\Theta) = \int_{-\infty}^{\infty} [g(\mathbf{X}) - \Theta] p_{\mathbf{x}|\theta}(\mathbf{X}|\Theta) d\mathbf{X}. \quad (6)$$

It is assumed that

$$\lim_{\Theta_i \rightarrow -\infty} B(\Theta) p_{\mathbf{x}|\theta}(\mathbf{X}|\Theta) = 0 \quad (7)$$

$$\lim_{\Theta_i \rightarrow -\infty} B(\Theta) p_{\mathbf{x}|\theta}(\mathbf{X}|\Theta) = 0 \quad (8)$$

for $i = 1, \dots, r$.

Then, the inequality (1) holds.

Proof: See [5]. ■

In the special case, when only the i th component of the parameter vector θ is estimated, $\hat{\theta}_i = g_i(\mathbf{x})$, the PCRb, similarly to the standard CRB, is given as

$$\mathbf{E}[(g_i(\mathbf{x}) - \theta_i)^2] \geq [\mathbf{J}^{-1}]_{ii} \quad (9)$$

where $[\cdot]_{ii}$ denotes the (i, i) th element.

III. CALCULATION OF THE PCRb

In this section, the PCRb is calculated for the estimation problem of a single noisy cisoid with slowly varying frequency. The signal model has the form

$$y_t = m_0 e^{i\varphi_t} + v_t \quad t = 0, 1, 2, \dots, n \quad (10)$$

where m_0 is the magnitude, φ_t is the instantaneous phase of cisoid at time instant t , and $\{v_t\}$ is the noise. The instantaneous frequency is defined as the one-step increment of φ_t

$$\omega_t = \varphi_t - \varphi_{t-1} \quad t = 1, 2, \dots, n. \quad (11)$$

Put

$$e_t = \omega_t - \omega_{t-1} \quad t = 1, 2, \dots, n \quad (12)$$

and assume that

A1: $\{v_t\}$ and $\{e_t\}$ are independent sequences of independent random variables with zero mean values and variances σ^2 and γ^2 , respectively; $\{e_t\}$ is Gaussian, and $\{v_t\}$ is complex circular Gaussian.

A2: The initial parameters of the signal, i.e., φ_0 , ω_0 , and m_0 are known constants.

The main problem is to estimate the parameter vector

$$\theta = (\omega_1, \dots, \omega_n) \quad (13)$$

on the base of the data $\mathbf{x} = (y_1, \dots, y_n)^T$.

Assumption **A1** implies that the evolution of the sinusoidal frequency is described as a random walk with increments of given variance. Assumption **A2** is not too restrictive if one is interested in estimating the signal parameters at time instants $t \gg 0$; since the frequency is time varying, the exact knowledge of φ_t and ω_t for $t = 0$ is irrelevant for large t .

The log-likelihood function for our model is given as

$$-\log p_{\mathbf{x},\theta}(\mathbf{X}, \Theta) = C + \frac{1}{\sigma^2} \sum_{j=1}^n |y_j - m_0 e^{i\varphi_j}|^2 + \frac{1}{2\gamma^2} \sum_{j=1}^n (\omega_j - \omega_{j-1})^2 \quad (14)$$

where C is a constant that does not play a role in further calculations.¹

Note from (11) that the sinusoidal phase at the time instant t can be written as

$$\varphi_t = \varphi_0 + \sum_{j=1}^t \omega_j \quad t = 1, 2, \dots, n. \quad (15)$$

By aid of (15) it can be easily shown that

$$\begin{aligned} -\frac{\partial \log p_{\mathbf{x},\theta}(\mathbf{X}, \Theta)}{\partial \omega_k} &= \frac{(-2)}{\sigma^2} \sum_{j=k}^n \operatorname{Re} (i y_j^* m_0 e^{i\varphi_j}) \\ &\quad + \frac{1}{\gamma^2} (2\omega_k - \omega_{k-1} - \omega_{k+1}) \\ -\frac{\partial^2 \log p_{\mathbf{x},\theta}(\mathbf{X}, \Theta)}{\partial \omega_k \partial \omega_{k'}} &= \frac{2}{\sigma^2} \sum_{j=\max(k, k')}^n \operatorname{Re} (y_j^* m_0 e^{i\varphi_j}) \\ &\quad + \frac{1}{\gamma^2} r_{kk'} \end{aligned} \quad (16)$$

for $k, k' = 2, \dots, n-1$, where $\delta_{kk'}$ is the Kronecker delta function, and

$$r_{kk'} = \begin{cases} 2 & \text{for } k' = k \\ -1 & \text{for } k' = k \pm 1 \\ 0 & \text{elsewhere} \end{cases} \quad (17)$$

Similar expressions are obtained for $k, k' = 1$ and $k, k' = n$, but they are omitted for brevity. Taking the expectations of the second-order derivatives, we get the following result.

Proposition 1: The posterior information matrix for the estimation problem (10)–(13), under assumptions **A1** and **A2**, is given by

$$\mathbf{J}_n = 2 \frac{m_0^2}{\sigma^2} \mathbf{H}_n + \frac{1}{\gamma^2} \mathbf{G}_n \quad (18)$$

where \mathbf{H}_n and \mathbf{G}_n are the $(n \times n)$ matrices

$$\mathbf{H}_n = \begin{pmatrix} n & n-1 & n-2 & \dots & 1 \\ n-1 & n-1 & n-2 & \dots & 1 \\ n-2 & n-2 & n-2 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix} \quad (19)$$

$$\mathbf{G}_n = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}. \quad (20)$$

¹Since $\omega_j \in (-\pi, \pi]$, the Gaussian assumption imposed on e_t and used in (14) is realistic only for $\gamma^2 \ll 1$.

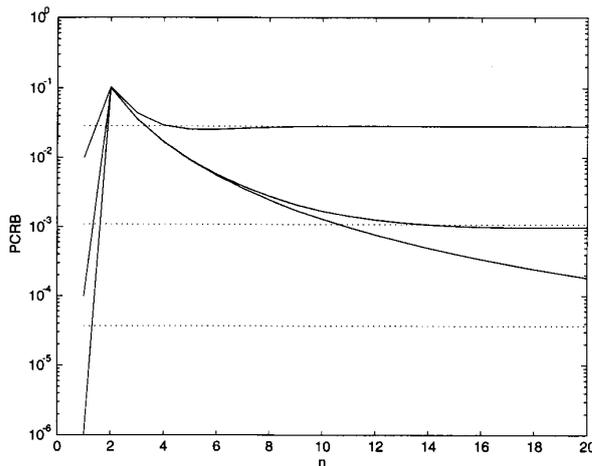


Fig. 1. Illustration of the convergence of the PCRB ($\hat{\omega}_n$) (solid lines), to the corresponding limit PCRB (dotted lines), for SNR = 0 dB and $\gamma = 0.1, 0.01$ and 0.001, respectively (from the top down).

In particular, for estimating the n th instantaneous frequency, we obtain

$$E\{[\hat{\omega}_n - \omega_n]^2\} \geq [(\mathbf{J}_n)^{-1}]_{nn} \triangleq \text{PCRB}(\hat{\omega}_n). \quad (21)$$

Note that the same PCRB can be obtained for the more general model $y_t = m_t e^{i\varphi_t} + v_t$, $t > 0$, where the magnitudes $\{m_t\}$ are i.i.d. Gaussian r.v.'s, independent of $\{v_t\}$ and $\{e_t\}$, if $E\{m_t^2\} = m_0^2$. This observation is consistent with the idea of estimating frequency only from the phase of the data [11].

The main contribution of this correspondence is the analytic formula for $\lim_{n \rightarrow \infty} \text{PCRB}(\hat{\omega}_n)$, i.e., LPCRB($\hat{\omega}_n$).

Proposition 2: The PCRB for the estimation of ω_n in the model (10)–(13), under assumptions **A1** and **A2**, has the limit

$$\begin{aligned} \lim_{n \rightarrow \infty} E\{(\hat{\omega}_n - \omega_n)^2\} &\geq \text{LPCRB}(\hat{\omega}_n) \\ &= \frac{\gamma^2}{w} (-w + \sqrt{w^2 + 4w}) \end{aligned} \quad (22)$$

where

$$w = \Gamma + \sqrt{\Gamma^2 + 8\Gamma} \quad (23)$$

$$\Gamma = \gamma^2 \frac{m_0^2}{\sigma^2}. \quad (24)$$

Proof (Outline): The right-hand side of (21) is rewritten as

$$[(\mathbf{J}_n)^{-1}]_{nn} = \frac{\det \mathbf{J}_{n,1:n-1}}{\det \mathbf{J}_n} \quad (25)$$

where $\mathbf{J}_{n,1:n-1}$ is the left-upper $(n-1) \times (n-1)$ submatrix of \mathbf{J}_n . Using some elementary row and column operations, the determinants in (25) can be written as determinants of five-diagonal matrices, and finally, by expansion with respect to their last column or row, they can be rewritten into a recursive fashion; analysis of the obtained recursions together with (25) implies the statement. See [12] for details. ■

It can be seen from the proof of the proposition that the same LPCRB would be obtained under the assumption that φ_0 and ω_0 are Gaussian random variables, independent of $\{e_n\}$ and $\{v_n\}$.

IV. EFFICIENCY OF TWO METHODS

The algorithms MFT and ANF that solve the estimation problem (10)–(13) for multiple cisoids have been studied in [4]. The

algorithms have been analyzed for a single cisoid by aid of an approximating linear filter (ALF) technique, assuming high signal-to-noise ratio (SNR = $m_0^2/\sigma^2 \gg 1$), slow evolution of the signal parameters ($\gamma^2 \ll 1$), and a proper initialization of the procedures. It has been shown that for both methods

$$\begin{aligned} \lim_{n \rightarrow \infty} E\{(\hat{\omega}_n - \omega_n)^2\} &= \frac{\sigma^2}{m_0^2} \frac{(1-\lambda)(1-\rho)^2}{1+3\lambda+\rho(1-\lambda)} \\ &+ \gamma^2 \frac{\rho^2(1-\lambda)^2 + 2\lambda(\rho+\lambda-2\rho\lambda)}{[1+3\lambda+\rho(1-\lambda)](1-\lambda)(1-\rho)} \end{aligned} \quad (26)$$

where λ and ρ are two forgetting factors in the MFT or the pole contraction factor and forgetting factor in the ANF, $\rho \in (0, 1)$. Further, it can be proved that (26) achieves its minimum value for

$$\lambda = \lambda_\Gamma \triangleq 1 - \frac{\kappa}{2} \quad (27)$$

$$\rho = \rho_\Gamma \triangleq 1 - \frac{\kappa}{4 - \kappa} \quad (28)$$

where

$$\kappa = -w + \sqrt{w^2 + 4w} \quad (29)$$

and w was defined in (23). The minimum value of (26) is

$$\lim_{n \rightarrow \infty} E\{(\hat{\omega}_n - \omega_n)^2 | \lambda = \lambda_\Gamma, \rho = \rho_\Gamma\} = \gamma^2 \frac{\kappa}{w}. \quad (30)$$

Comparing the last result with (22) implies that the MFT and the ANF are statistically efficient estimators of the slowly varying sinusoidal frequency under the assumptions of the ALF approximation.

V. CONCLUSIONS

The PCRB for the problem of estimation of instantaneous frequency of a single nonstationary cisoid in noise has been calculated. It has been assumed that the evolution of the frequency is modeled by a Gaussian random walk and that the additive noise is Gaussian and white. We showed that the two estimation procedures (the MFT and the ANF) can, under certain assumptions, asymptotically attain this lower bound.

ACKNOWLEDGMENT

The author is thankful to Dr. R. Kulhavý for helpful discussion on the PCRB, to Dr. P. Händel for his contribution to the Introduction, and to Prof. P. Stoica for [5].

REFERENCES

- [1] A. Nehorai, "A minimal parameter adaptive notch filter with constrained poles and zeros," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-33, pp. 983–996, Aug. 1985.
- [2] B. S. Chen, T. Y. Yang, and B. H. Lin, "Adaptive notch filter by direct frequency estimation," *Signal Processing*, vol. 27, no. 2, pp. 161–167, May 1992.
- [3] P. Tichavský and P. Händel, "Efficient tracking of multiple sinusoids with slowly varying parameters," in *Proc. 1993 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Minneapolis, MN, Apr. 1993, pp. 368–371, Pt. III.
- [4] —, "Two algorithms for adaptive retrieval of slowly time-varying multiple cisoids in noise," *IEEE Trans. Signal Processing*, this issue, pp. 1116–1127.
- [5] H. L. van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968.
- [6] S. M. Kay, *Modern Spectral Estimation. Theory and Application*. Englewood Cliffs, NJ: Prentice Hall, 1988.
- [7] P. Stoica and A. Nehorai, "On the asymptotic distribution of exponentially weighted prediction error estimators," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 136–139, Jan. 1988.
- [8] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification—A survey," *Automatica*, vol. 26, no. 1, pp. 7–21, Jan. 1990.

- [9] P. Händel and A. Nehorai, "Tracking analysis of an adaptive notch filter with constrained poles and zeros," *IEEE Trans. Signal Processing*, vol. 42, no. 2, pp. 281–291, Feb. 1994.
- [10] N. Cho and S. Lee, "On the tracking properties of the adaptive lattice notch filter," in *Proc. 1993 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Minneapolis, MN, Apr. 1993, pp. 396–399, Pt. III.
- [11] S. Kay, "A fast and accurate single frequency estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 12, pp. 1987–1990, Dec. 1989.
- [12] P. Tichavský, "A posteriori Cramér–Rao bound for adaptive harmonic retrieval," Research Report no. 1805, Inst. of Information Theory and Automation, Prague, Czech Republic, June 1994.

Identification of Quadratic Volterra Systems Driven by Non-Gaussian Processes

A. M. Zoubir

Abstract—A nonlinear and time-invariant system representable by a Volterra series up to second order is considered. Closed-form expressions for the generalized transfer functions of first and second order are derived for non-Gaussian stationary input processes whose trispectrum vanishes. It is shown that the parameters obtained are optimum in the mean square sense. Once the system is identified, a closed-form expression for the quadratic coherence is derived. This expression simplifies to well-known results when the system is linear or its input is Gaussian. The quadratic coherence is validated using simulated data as input to a known second-order Volterra filter with known statistic.

I. INTRODUCTION

The Volterra series provides an important type of representation for nonlinear models and has been used in many applications. In this representation, a nonlinear system is characterized by a set of functions called the Volterra kernels, which are determined by "cross-correlating" the system response with its input. A difficulty encountered in this approach is to find simple analytical expressions for the Volterra kernels in terms of the cumulants or equivalently for the generalized transfer functions in terms of higher order spectra. There is one special case when one can derive a generalization of the basic result for linear systems. This is the case where the Volterra series contains just one term of any order and the input is a Gaussian process, as studied in [1]. In many applications, it is more desirable to identify a system with a non-Gaussian input.

The purpose of this paper is to derive analytical expressions for the generalized transfer functions of Volterra filters of second order in terms of higher order spectra for a class of non-Gaussian input processes. This class consists of stationary processes whose fourth-order cumulant function vanishes. An outline of the paper follows.

In Section II, a formulation of the problem is given. Section III discusses the identification of quadratic Volterra filters. Analytical expressions for the generalized transfer functions of first and second order are given. It is shown that the derived parameters are optimum in the mean square sense. In Section IV, a closed-form expression for the quadratic coherence is given. In Section V, we discuss the

Manuscript received December 21, 1992; revised September 27, 1994. The associate editor coordinating the review of this paper and approving it for publication was Prof. Douglas Williams.

A. M. Zoubir is with the Signal Processing Research Centre, Queensland University of Technology, Brisbane Q4001, Australia.

IEEE Log Number 9410302.

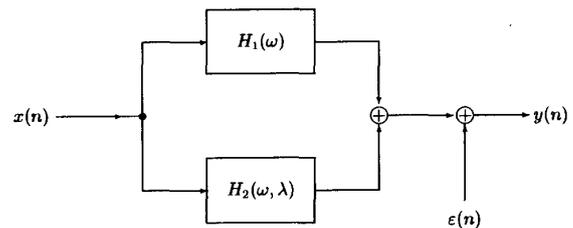


Fig. 1. Signal flow diagram of the considered nonlinear system.

estimation of the generalized transfer functions and the quadratic coherence. Simulation results are illustrated and discussed before concluding.

II. PROBLEM FORMULATION

Let $x(n)$ be a zero-mean stationary random process with discrete time parameter $n = 0, \pm 1, \pm 2, \dots$. Assume the process $y(n)$ to be generated by

$$y(n) = \sum_{n_1=-\infty}^{\infty} h_1(n-n_1)x(n_1) + \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} h_2(n-n_1, n-n_2)x(n_1)x(n_2) + \varepsilon(n) \quad (1)$$

where $\varepsilon(n)$ is zero-mean stationary modelling noise, and $\varepsilon(n)$ and $x(n)$ are independent. For the description of the time-invariant filter, we have used the notations $h_1(n_1)$ and $h_2(n_1, n_2)$, $n_1, n_2 = 0, \pm 1, \pm 2, \dots$ for the real Volterra kernels of first and second order, respectively. Without loss of generality, we assume that $h_2(n_1, n_2)$ is symmetric in its arguments.

A signal flow diagram of the model is given in Fig. 1, where $H_1(\omega) = \sum_{n_1=-\infty}^{\infty} h_1(n_1)e^{-j\omega n_1}$ and $H_2(\omega, \lambda) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} h_2(n_1, n_2)e^{-j(\omega n_1 + \lambda n_2)}$ are, respectively, called the generalized linear and quadratic transfer functions.

By identification of the time-invariant second-order Volterra filter, we mean determination of its generalized transfer functions $H_1(\omega)$ and $H_2(\omega, \lambda)$ from second- and third-order spectra of the input and output series. The identification of such a filter has been studied by Tick [6], who derived closed-form expressions for $H_1(\omega)$ and $H_2(\omega, \lambda)$, assuming $x(n)$ to be a Gaussian process. Kim and Powers [4] have concentrated on the non-Gaussian case. They propose a digital method based on a linear regression derived from (1) that is solved with respect to $H_1(\omega)$ and $H_2(\omega, \lambda)$ at a set of discrete frequencies.

In this correspondence, we derive closed-form expressions for the system describing generalized transfer functions $H_1(\omega)$ and $H_2(\omega, \lambda)$ in terms of second- and third-order spectra of the input and output process for non-Gaussian stationary zero-mean input processes whose fourth-order cumulant function vanishes. Processes with this property may arise in practice, when they are generated as the sum of independent processes whose fourth-order cumulants cancel. Another example is when the process is singly or doubly truncated Gaussian with a parameter that reduces the fourth order cumulant to zero. In other applications, we may assume a vanishing trispectrum because the shape parameter of the distribution is large and fourth-order cumulants fall faster to zero than third-order cumulants, such as in the Gamma or K-distributions often used in radar signal processing.

Posterior Cramér–Rao Bounds for Discrete-Time Nonlinear Filtering

Petr Tichavský, *Member, IEEE*, Carlos H. Muravchik, *Member, IEEE*, and Arye Nehorai, *Fellow, IEEE*

Abstract—A mean-square error lower bound for the discrete-time nonlinear filtering problem is derived based on the Van Trees (posterior) version of the Cramér–Rao inequality. This lower bound is applicable to multidimensional nonlinear, possibly non-Gaussian, dynamical systems and is more general than the previous bounds in the literature. The case of singular conditional distribution of the one-step-ahead state vector given the present state is considered. The bound is evaluated for three important examples: the recursive estimation of slowly varying parameters of an autoregressive process, tracking a slowly varying frequency of a single cisoid in noise, and tracking parameters of a sinusoidal frequency with sinusoidal phase modulation.

Index Terms—Adaptive estimation, Kalman filtering, nonlinear filters, time-varying systems, tracking filters.

I. INTRODUCTION

DISCRETE-TIME nonlinear filtering or the associated problem of adaptive system identification arise in various applications such as adaptive control, analysis, and prediction of nonstationary time series. As is well known, the optimal estimator for this problem cannot be built in general, and it is necessary to turn to one of the large number of existing suboptimal filtering techniques [1]. Assessing the achievable performance may be difficult, and we have to resort to simulations and comparing proximity to lower bounds corresponding to optimum performance. Lower bounds give an indication of performance limitations, and consequently, they can also be used to determine whether imposed performance requirements are realistic or not.

In time-invariant statistical models, a commonly used lower bound is the Cramér–Rao bound (CRB), given by the inverse of the Fisher information matrix. In the time-varying systems context we deal with here, the estimated parameter vector has to be considered random since it corresponds to an underlying nonlinear, randomly driven model. A lower bound that is

Manuscript received December 3, 1996; revised September 12, 1997. A short version of this paper was presented at the First European Conference on Signal Analysis and Prediction ECSAP-97, June 1997, Prague, Czech Republic. This work was supported in part by Grant 102/97/0466 of the Grant Agency of the Czech Republic, CICIPBA and UNLP, the Air Force Office of Scientific Research under Grant F49620-97-1-0481, the National Science Foundation under Grant MIP-9615590, and the Office of Naval Research under Grant N00014-96-1-1078.

The associate editor coordinating the review of this paper and approving it for publication was Prof. James A. Bucklew.

P. Tichavský is with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic (e-mail: tichavsk@utia.cas.cz).

C. Muravchik is with LEICI, Departamento Electrotecnia, Universidad Nacional de La Plata, La Plata, Argentina.

A. Nehorai is with the Department of Electrical Engineering and Computer Sciences, University of Illinois at Chicago, Chicago, IL 60607-7053 USA.

Publisher Item Identifier S 1053-587X(98)03256-5.

analogous to the CRB for random parameters was derived in [11]; this bound is usually referred to as the Van Trees version of the CRB, or posterior CRB (PCRB) [16]. Some properties of the PCRB are summarized in Section II.

Several lower bounds for nonlinear dynamical systems have appeared in the literature; see the overview in [6]. However, the continuous-time case has received heavy emphasis but not the discrete-time case, which is of greater practical importance. Bobrovsky and Zakai [2] were the first to apply the Cramér–Rao theory to scalar discrete-time systems. The bound was later improved and generalized to the multidimensional case by Galdos [3]. Both of these bounds were obtained by comparing the information matrix of the original system with an information matrix of a suitable Gaussian system. The bound in [3] is already quite general, but it still has some limitations (see the discussion in [6]), i.e., the assumption that the dimension of the system and measurements are identical. Recently, the approach by Galdos has been generalized for nonlinear p th-order autoregressive processes driven by additive Gaussian noise with state-dependent gain [4].

In Section III of this paper, a novel and simple derivation of the posterior CRB for the discrete-time multidimensional nonlinear filtering problem that avoids any Gaussian assumptions is presented. The derivation is obtained from first principles and differs from other approaches that instead consider comparison of the original nonlinear system with an appropriate linear Gaussian system. We present an example of a linear Gaussian system (which is different from those in [2] and [3]) that has the same associated information matrix as the original system. In Section IV, the lower bound is extended for a frequently occurring case of nonlinear filtering, where the conditional distribution of the state one step ahead, given the current state, is singular. Note that a special case of a similar extension was proposed in [3]. Section V illustrates an application of the bound for three important examples:

- recursive estimation of slowly varying parameters of an autoregressive process;
- tracking of a slowly varying frequency of a single cisoid in noise (a new alternate derivation of the lower bound in [16]);
- tracking parameters of a varying frequency that is modulated by a sinusoid [17].

Conclusions are drawn in Section VI.

II. PROPERTIES OF THE PCRB (REVIEW)

Let x represent a vector of measured data, let θ be an r -dimensional estimated random parameter, let $p_{x,\theta}(X, \Theta)$ be

the joint probability density of the pair (x, θ) , and let $g(x)$ be a function of x , which is an estimate of θ . The PCRB on the estimation error has the form

$$P \triangleq \mathbb{E}\{[g(x) - \theta][g(x) - \theta]^T\} \geq J^{-1} \quad (1)$$

where J is the $r \times r$ (Fisher) information matrix with the elements

$$J_{ij} = \mathbb{E}\left[-\frac{\partial^2 \log p_{x,\theta}(X, \Theta)}{\partial \Theta_i \partial \Theta_j}\right] \quad i, j = 1, \dots, r \quad (2)$$

provided that the derivatives and expectations in (1) and (2) exist. The superscript “ T ” in (1) denotes the transpose of a matrix, and the inequality in (1) means that the difference $P - J^{-1}$ is a positive semidefinite matrix. The proof given in [10] or [11] holds under the additional condition of

$$\lim_{\Theta_i \rightarrow -\infty} B(\Theta)p_\theta(\Theta) = 0, \quad \lim_{\Theta_i \rightarrow \infty} B(\Theta)p_\theta(\Theta) = 0 \quad (3)$$

$$i = 1, \dots, r$$

where $B(\Theta)$ is the estimation bias conditioned by $\theta = \Theta$, and

$$B(\Theta) = \int_{-\infty}^{\infty} [g(X) - \Theta] p_{x|\theta}(X|\Theta) dX. \quad (4)$$

Let ∇ and δ be operators of the first and second-order partial derivatives, respectively

$$\nabla_\Theta = \left[\frac{\partial}{\partial \Theta_1}, \dots, \frac{\partial}{\partial \Theta_r} \right]^T \quad (5)$$

$$\Delta_\Psi^\Theta = \nabla_\Psi \nabla_\Theta^T. \quad (6)$$

Using this notation, (2) can be written as

$$J = \mathbb{E}[-\Delta_\Theta^\Theta \log p_{x,\theta}(X, \Theta)]. \quad (7)$$

Since $p_{x,\theta}(X, \Theta) = p_{x|\theta}(X|\Theta) \cdot p_\theta(\Theta)$, it can easily be seen that J can be decomposed into two additive parts:

$$J = J_D + J_P \quad (8)$$

where J_D represents the information obtained from the data, and J_P represents the *a priori* information

$$J_D = \mathbb{E}\{-\Delta_\Theta^\Theta \log p_{x|\theta}(X|\Theta)\} \quad (r \times r) \quad (9)$$

$$J_P = \mathbb{E}\{-\Delta_\Theta^\Theta \log p_\theta(\Theta)\} \quad (r \times r) \quad (10)$$

provided that the expectations in (9) and (10) exist. Note that J_D is an expectation of the standard Fisher information matrix over the *a priori* distribution of Θ .

An alternative expression for the information matrix can be derived from the equality $p_{x,\theta}(X, \Theta) = p_{\theta|x}(\Theta|X) \cdot p_x(X)$. Since $p_x(X)$ is an integral of $p_{x,\theta}(X, \Theta)$ over Θ , it does not depend any longer on Θ ; therefore, we have

$$J = \mathbb{E}\{-\Delta_\Theta^\Theta \log p_{\theta|x}(\Theta|X)\}. \quad (11)$$

For example, if the posterior distribution of θ conditioned on the data vector x is Gaussian with mean $\bar{\theta}_x$ and a (regular) covariance matrix Σ_x

$$-\log p_{\theta|x}(\Theta|X) = c_0 + \frac{1}{2}(\Theta - \bar{\theta}_x)^T \Sigma_x^{-1}(\Theta - \bar{\theta}_x) \quad (12)$$

holds, where c_0 denotes a constant independent of Θ . Then, the information matrix in (11) reads

$$J = \mathbb{E}\{\Sigma_x^{-1}\}. \quad (13)$$

If θ is estimated by $g(x) = \mathbb{E}(\theta|x)$, then (1) is satisfied with equality. This is exactly the case for the Kalman filter when performing the task of linear filtering.

Assume now that the parameter θ is decomposed into two parts as $\theta = [\theta_\alpha^T, \theta_\beta^T]^T$, and the information matrix J is correspondingly decomposed into blocks

$$J = \begin{bmatrix} J_{\alpha\alpha} & J_{\alpha\beta} \\ J_{\beta\alpha} & J_{\beta\beta} \end{bmatrix}. \quad (14)$$

It can easily be shown that the covariance of estimation of θ_β is lower bounded by the right-lower block of J^{-1} , i.e.,

$$P_\beta \triangleq \mathbb{E}\{[g_\beta(x) - \theta_\beta][g_\beta(x) - \theta_\beta]^T\} \geq [J_{\beta\beta} - J_{\beta\alpha} J_{\alpha\alpha}^{-1} J_{\alpha\beta}]^{-1} \quad (15)$$

assuming that $J_{\alpha\alpha}^{-1}$ exists. In the following, the matrix $J_{\beta\beta} - J_{\beta\alpha} J_{\alpha\alpha}^{-1} J_{\alpha\beta}$ will be called the *information submatrix* for parameter θ_β .

III. A LOWER BOUND FOR THE NONLINEAR FILTERING PROBLEM

Consider the nonlinear filtering problem

$$x_{n+1} = f_n(x_n, w_n) \quad (16)$$

$$z_n = h_n(x_n, v_n) \quad (17)$$

where

- x_n system state at time n ;
- $\{z_n\}$ measurement process;
- $\{w_n\}$ and $\{v_n\}$ independent white processes (i.e., sequences of mutually independent random variables or vectors);
- f_n and h_n (in general) nonlinear functions.

The functions f_n and h_n may depend on time n . Further assume that the initial state x_0 has a known probability density function $p(x_0)$. Let the dimension of the states $\{x_n\}$ be r .

Equations (16) and (17) together with $p(x_0)$ determine unambiguously the joint probability distribution of $X_n = (x_0, \dots, x_n)$ and $Z_n = (z_0, \dots, z_n)$ for an arbitrary n [2]

$$p(X_n, Z_n) = p(x_0) \prod_{j=1}^n p(z_j|x_j) \prod_{k=1}^n p(x_k|x_{k-1}). \quad (18)$$

In (18) as well as in the sequel, $p(\cdot)$'s refer to (unconditional and conditional) probability densities of the variables depicted in the argument of p 's. The conditional probability densities $p(x_k|x_{k-1})$ and $p(z_k|x_k)$ follow from (16) and (17), respectively, under suitable hypotheses.

Let $J(X_n)$ be the $(nr \times nr)$ information matrix of X_n derived from the above joint distribution. The problem that we wish to solve in this section is the computation of the information submatrix for estimating x_n , which is denoted J_n , which is given as the inverse of the $(r \times r)$ right-lower block of $[J(X_n)]^{-1}$. The matrix J_n^{-1} will provide a lower bound on the

mean square error of estimating x_n . In the sequel, $p(X_n, Z_n)$ is denoted by p_n for brevity.

Decompose X_n as $X_n = [X_{n-1}^T, x_n^T]^T$ and $J(X_n)$ correspondingly as

$$J(X_n) = \begin{bmatrix} A_n & B_n \\ B_n^T & C_n \end{bmatrix} \triangleq \begin{bmatrix} E\{-\Delta_{X_{n-1}}^{x_n} \log p_n\} & E\{-\Delta_{X_{n-1}}^{x_n} \log p_n\} \\ E\{-\Delta_{x_n}^{X_{n-1}} \log p_n\} & E\{-\Delta_{x_n}^{X_{n-1}} \log p_n\} \end{bmatrix} \quad (19)$$

provided that the derivatives and the expectations exist. Comparison of (16) and (20) gives

$$J_n = C_n - B_n^T A_n^{-1} B_n. \quad (20)$$

Thus, computation of the $(r \times r)$ matrix J_n involves either calculation of the inverse of $[(n-1)r \times (n-1)r]$ matrix A_n or inverse of the full $(nr \times nr)$ matrix $J(X_n)$.

The following proposition gives a recipe for computing J_n recursively without manipulating large matrices such as A_n or $J(X_n)$. In particular, an efficient method for computing the limit of J_n for $n \rightarrow \infty$ follows from the recursion.

Proposition 1: The sequence $\{J_n\}$ of posterior information submatrices for estimating state vectors $\{x_n\}$ obeys the recursion

$$J_{n+1} = D_n^{22} - D_n^{21}(J_n + D_n^{11})^{-1} D_n^{12} \quad (21)$$

where

$$D_n^{11} = E\{-\Delta_{x_n}^{x_n} \log p(x_{n+1}|x_n)\} \quad (r \times r) \quad (22)$$

$$D_n^{12} = E\{-\Delta_{x_n}^{x_{n+1}} \log p(x_{n+1}|x_n)\} \quad (r \times r) \quad (23)$$

$$D_n^{21} = E\{-\Delta_{x_{n+1}}^{x_n} \log p(x_{n+1}|x_n)\} = [D_n^{12}]^T \quad (24)$$

$$D_n^{22} = E\{-\Delta_{x_{n+1}}^{x_{n+1}} \log p(x_{n+1}|x_n)\} + E\{-\Delta_{x_{n+1}}^{z_{n+1}} \log p(z_{n+1}|x_{n+1})\} \quad (r \times r). \quad (25)$$

Proof: The joint probability function of X_{n+1} and Z_{n+1} can be written as

$$\begin{aligned} p_{n+1} &\triangleq p(X_{n+1}, Z_{n+1}) \\ &= p(X_n, Z_n) \cdot p(x_{n+1}|X_n, Z_n) \\ &\quad \cdot p(z_{n+1}|x_{n+1}, X_n, Z_n) \\ &= p_n \cdot p(x_{n+1}|x_n) \cdot p(z_{n+1}|x_{n+1}). \end{aligned} \quad (26)$$

Using (26) and the notations in (19) and (22)–(25), the posterior information matrix for X_{n+1} can be written in block form as

$$J(X_{n+1}) = \begin{bmatrix} A_n & B_n & 0 \\ B_n^T & C_n + D_n^{11} & D_n^{12} \\ 0 & D_n^{21} & D_n^{22} \end{bmatrix} \quad (27)$$

where 0's stand for zero blocks of appropriate dimensions.

The information submatrix J_{n+1} can be found as an inverse of the right-lower $(r \times r)$ submatrix of $[J(X_{n+1})]^{-1}$

$$\begin{aligned} J_{n+1} &= D_n^{22} - [0 \quad D_n^{21}] \begin{bmatrix} A_n & B_n \\ B_n^T & C_n + D_n^{11} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ D_n^{12} \end{bmatrix} \\ &= D_n^{22} - D_n^{21} [C_n + D_n^{11} - B_n^T A_n^{-1} B_n]^{-1} D_n^{12}. \end{aligned} \quad (28)$$

Using the definition of J_n in (20), we obtain the desired formula (21). ■

Note that the recursion in (21) involves computations with matrices of dimension $(r \times r)$. The initial information submatrix J_0 can be calculated from the *a priori* probability function $p(x_0)$

$$J_0 = E\{-\Delta_{x_0}^{x_0} \log p(x_0)\}. \quad (29)$$

A few remarks follow to elucidate special cases.

A. Additive Gaussian Noise

Assume that the nonlinear filtering problem in (16) and (17) has the form

$$x_{n+1} = f_n(x_n) + w_n \quad (30)$$

$$z_n = h_n(x_n) + v_n \quad (31)$$

and that the noises $\{w_n\}$ and $\{v_n\}$ are Gaussian with zero mean and invertible covariance matrices Q_n and R_n , respectively. From these assumptions, it follows that

$$\begin{aligned} -\log p(x_{n+1}|x_n) &= c_1 + \frac{1}{2} [x_{n+1} - f_n(x_n)]^T Q_n^{-1} [x_{n+1} - f_n(x_n)] \end{aligned} \quad (32)$$

$$\begin{aligned} -\log p(z_{n+1}|x_{n+1}) &= c_2 + \frac{1}{2} [z_{n+1} - h_{n+1}(x_{n+1})]^T \\ &\quad \cdot R_{n+1}^{-1} [z_{n+1} - h_{n+1}(x_{n+1})] \end{aligned} \quad (33)$$

where c_1 and c_2 are constants, and

$$D_n^{11} = E\{[\nabla_{x_n} f_n^T(x_n)] Q_n^{-1} [\nabla_{x_n} f_n^T(x_n)]^T\} \quad (34)$$

$$D_n^{12} = -E\{[\nabla_{x_n} f_n^T(x_n)] Q_n^{-1} \nabla_{x_n} h_{n+1}^T(x_{n+1})\} \quad (35)$$

$$D_n^{22} = Q_n^{-1} + E\{[\nabla_{x_{n+1}} h_{n+1}^T(x_{n+1})] \cdot R_{n+1}^{-1} [\nabla_{x_{n+1}} h_{n+1}^T(x_{n+1})]^T\}. \quad (36)$$

The well-known solution of the problem in the linear case [with linear functions f_n and g_n in (30) and (31)] is the Kalman filter. This is an algorithm that computes parameters of the conditional distribution of the state x_n given the data Z_n . The distribution is Gaussian, and its mean and covariance matrix are usually denoted $\hat{x}_{n|n}$ and $\Sigma_{n|n}$, respectively. It can easily be shown that the recursion (21) for J_n is identical to those that are usually derived for $\Sigma_{n|n}^{-1}$ from the Kalman filter equations [1].

In order to compare the result (21) with the PCRb computations in [2] and [3], we find matrices \tilde{F}_n , \tilde{H}_n , \tilde{Q}_n , and \tilde{R}_n such that the linear system

$$\tilde{x}_{n+1} = \tilde{F}_n \tilde{x}_n + \tilde{w}_n \quad (37)$$

$$\tilde{z}_n = \tilde{H}_n \tilde{x}_n + \tilde{v}_n \quad (38)$$

has the same information matrix as the original nonlinear system; in (37) and (38), $\{\tilde{w}_n\}$ and $\{\tilde{v}_n\}$ are independent white Gaussian noises with zero means and covariance matrices \tilde{Q}_n and \tilde{R}_n , respectively. The matrices \tilde{F}_n , \tilde{H}_n , \tilde{Q}_n , and \tilde{R}_n can be determined by comparing the matrices D_n^{11} , D_n^{12} , and D_n^{22} of the original system, which are obtained from (34)–(37) to those of the linear system in (37) and (38), yielding

$$\tilde{D}_n^{11} = \tilde{F}_n^T \tilde{Q}_n^{-1} \tilde{F}_n \quad (39)$$

$$\tilde{D}_n^{12} = -\tilde{F}_n^T \tilde{Q}_n^{-1} \tilde{H}_n \quad (40)$$

$$\tilde{D}_n^{22} = \tilde{Q}_n^{-1} + \tilde{H}_n^T \tilde{R}_n^{-1} \tilde{H}_n. \quad (41)$$

One possible solution of the above system of equations is

$$\tilde{F}_n = -(D_n^{12})^{-1} D_n^{11} \quad (42)$$

$$\tilde{Q}_n = D_n^{21} (D_n^{11})^{-1} D_n^{12} \quad (43)$$

$$\tilde{R}_n = R_n \quad (44)$$

$$\tilde{H}_n = R_n^{1/2} [D_{n-1}^{22} - (D_{n-1}^{12})^{-1} D_{n-1}^{11} (D_{n-1}^{12})^{-1}]^{1/2} \quad (45)$$

where $A^{1/2}$ denotes the square root of a positive semidefinite matrix A , assuming that the requisite inverses in (42)–(45) exist. Note that the above linear filter is different from those proposed in [2] and [3].

B. A Generalization

Consider the generalization of the nonlinear system in (16) and (17) as

$$x_{n+1} = f_n(x_n, w_n) \quad (46)$$

$$z_n = h_n(x_n, v_n, z_{n-1}, \dots, z_{n-m}) \quad (47)$$

where m is an integer. It can easily be seen that for the generalized system, the whole derivation of (21) can be repeated en masse, with only two small differences: First, in the initialization, it has to be assumed that z_{-1}, \dots, z_{-m} are known constants and second that $p(z_{n+1}|x_{n+1}, X_n, Z_n)$ in (26) cannot be reduced to $p(z_{n+1}|x_{n+1})$ but merely to $p(z_{n+1}|x_{n+1}, z_n, \dots, z_{n-m+1})$. The latter term will also replace the former one in (25).

C. Time-Invariant Solutions

Now, assume that the functions $f_n(\cdot, \cdot)$ and $h_n(\cdot)$ are time invariant (independent of n). It can easily be seen that the matrices $D_n^{11}, \dots, D_n^{22}$ also do not depend on n . It can be shown that for $n \rightarrow \infty$, the matrix J_n converges to a matrix J_∞ , which is given as a solution to the equation

$$J_\infty = D_n^{22} - D_n^{21} (J_\infty + D_n^{11})^{-1} D_n^{12}. \quad (48)$$

Note that (48) is a discrete-time algebraic Riccati equation. A more common form of the Riccati equation is obtained if the recursion (21) is equivalently written as

$$\begin{aligned} J_{n+1} = & D_n^{21} (D_n^{11})^{-1} J_n (D_n^{11})^{-1} D_n^{12} - D_n^{21} (D_n^{11})^{-1} \\ & \cdot J_n (J_n + D_n^{11})^{-1} J_n (D_n^{11})^{-1} D_n^{12} \\ & + D_n^{22} - D_n^{21} (D_n^{11})^{-1} D_n^{12} \end{aligned} \quad (49)$$

which can be easily proved by simple algebraic manipulations. Then, put $J_{n+1} = J_n = J_\infty$.

Two popular methods for solving the Riccati equation are derived in [5] and [8], respectively; for a more comprehensive survey, see [7]. In addition, note that there is an available software for solving the equation in Matlab, namely, a function DARESOLV or an older function DLQR.

IV. A FREQUENT SINGULAR CASE

Computation of the information submatrix J_n , as described in the previous section, fails if the conditional distribution of x_{n+1} , given x_n is singular, and therefore, the probability density $p(x_{n+1}|x_n)$ is not defined. In the case of the

additive Gaussian noise considered in the previous section, this happens when the matrix Q_n is singular. In order to deal with these cases, consider the following modification of the original problem.

Assume that the state vector x_n can be written in block form as

$$x_n = \begin{bmatrix} x_n^{(1)} \\ x_n^{(2)} \end{bmatrix} \quad (50)$$

where $x_n^{(j)}$ has the length r_j , $j = 1, 2$, with $r_1 + r_2 = r$. The filtering is described by the set of equations

$$x_{n+1}^{(1)} = f_n(x_n, w_n) \quad (51)$$

$$x_{n+1}^{(2)} = g_n(x_n, x_{n+1}^{(1)}) \quad (52)$$

$$z_n = h_n(x_n, v_n) \quad (53)$$

where f_n , g_n , and h_n are (in general) nonlinear functions. Again, the task is to calculate the information submatrix J_n for x_n . The partitioning restriction (51)–(53) of the problem is somewhat general and includes, among others, the case $x_{n+1}^{(2)} = x_n^{(2)}$, which means that the second part of the state vector is constant in time, and it can be considered for use when there are unknown constant parameters in the model. Note that in [3], the case was considered when g_n is only a function of $x_n^{(2)}$.

In this section, we present first an explicit solution—a recursive equation for J_n —for a special case of the system (51)–(53) with a linear function g_n and then a conceptual solution for general g_n .

Case 1—Linear g_n :

Proposition 2: Consider the linear filtering in (51)–(53), and assume that the function g_n is linear so that (52) can be written as

$$x_{n+1}^{(2)} = G_n^{(1)} x_n^{(1)} + G_n^{(2)} x_n^{(2)} + G_n^{(3)} x_{n+1}^{(1)}. \quad (54)$$

In addition, assume that $G_n^{(2)}$ is invertible for all n . Put

$$X_n^{(1)} = \begin{bmatrix} x_1^{(1)} \\ \vdots \\ x_n^{(1)} \end{bmatrix}. \quad (55)$$

Let $J(X_n^{(1)}, x_n^{(2)})$ be an information matrix derived from the joint probability density $p(X_n^{(1)}, x_n^{(2)}, Z_n)$, and let S_n and J_n be the information submatrices for $[x_{n-1}^{(1)}, x_n]$ and for x_n , respectively. Then, S_n and J_n obey the recursions

$$\begin{aligned} S_{n+1} & \triangleq \begin{bmatrix} S_{n+1}^{11} & S_{n+1}^{12} & S_{n+1}^{13} \\ S_{n+1}^{21} & S_{n+1}^{22} & S_{n+1}^{23} \\ S_{n+1}^{31} & S_{n+1}^{32} & S_{n+1}^{33} \end{bmatrix} \\ & = M_n^{-T} \begin{bmatrix} J_n^{11} + H_n^{11} & J_n^{12} + H_n^{12} & H_n^{13} \\ (J_n^{12} + H_n^{12})^T & J_n^{22} + H_n^{22} & H_n^{23} \\ (H_n^{13})^T & (H_n^{23})^T & H_n^{33} \end{bmatrix} M_n^{-1} \end{aligned} \quad (56)$$

$$J_{n+1} = \begin{bmatrix} S_{n+1}^{22} & S_{n+1}^{23} \\ S_{n+1}^{32} & S_{n+1}^{33} \end{bmatrix} - \begin{bmatrix} S_{n+1}^{21} \\ S_{n+1}^{31} \end{bmatrix} [S_{n+1}^{11}]^{-1} [S_{n+1}^{12} \quad S_{n+1}^{13}] \quad (57)$$

where

$$J_n = \begin{bmatrix} J_n^{11} & J_n^{12} \\ J_n^{21} & J_n^{22} \end{bmatrix} \quad (r \times r) \quad (58)$$

$$M_n = \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & I \\ G_n^{(1)} & G_n^{(2)} & G_n^{(3)} \end{bmatrix} \quad [(r+r_1) \times (r+r_1)] \quad (59)$$

$$H_n^{11} = E\{-\Delta_{x_n^{(1)}} \log \bar{p}_n\} \quad (r_1 \times r_1) \quad (60)$$

$$H_n^{12} = E\{-\Delta_{x_n^{(2)}} \log \bar{p}_n\} \quad (r_1 \times r_2) \quad (61)$$

$$H_n^{13} = E\{-\Delta_{x_n^{(1)}} \log \bar{p}_n\} \quad (r_1 \times r_1) \quad (62)$$

$$H_n^{22} = E\{-\Delta_{x_n^{(2)}} \log \bar{p}_n\} \quad (r_2 \times r_2) \quad (63)$$

$$H_n^{23} = E\{-\Delta_{x_n^{(2)}} \log \bar{p}_n\} \quad (r_2 \times r_1) \quad (64)$$

$$H_n^{33} = E\{-\Delta_{x_{n+1}^{(1)}} \log \bar{p}_n\} \quad (r_1 \times r_1) \quad (65)$$

and

$$\bar{p}_n = p[x_{n+1}^{(1)}|x_n] \cdot p[z_{n+1}|x_n, x_{n+1}^{(1)}] \quad (66)$$

provided that the above derivatives and expectations exist. In (59), the I 's and 0 's stand for identity and zero matrices of appropriate dimensions.

Proof: See the Appendix.

Note that the conditional probability function $p[z_{n+1}|x_n, x_{n+1}^{(1)}]$ in (66) is obtained from $p(z_{n+1}|x_{n+1})$ by substituting for $x_{n+1}^{(2)}$ from (54).

A stationary solution for J_n would be obtained by inserting J_∞ for J_n and J_{n+1} . Note that the resulting equation no longer has the form of a Riccati equation, unlike (48) in the previous section.

For example, consider the above-mentioned case when $x_n^{(2)}$ is a constant unknown parameter. Comparing the equation $x_{n+1}^{(2)} = x_n^{(2)}$ with (54), we have $G_n^{(1)} = 0$, $G_n^{(2)} = I$, and $G_n^{(3)} = 0$, where 0 's and I 's stand again for zero and identity matrices of appropriate dimensions. Utilizing the special form of the matrix M_n in (59), from (56) and (57), the recursions

$$J_{n+1}^{11} = H_n^{33} - (H_n^{13})^T [J_n^{11} + H_n^{11}]^{-1} H_n^{13} \quad (67)$$

$$J_{n+1}^{12} = (J_{n+1}^{21})^T = (H_n^{23})^T - (H_n^{13})^T [J_n^{11} + H_n^{11}]^{-1} (J_n^{12} + H_n^{12}) \quad (68)$$

$$J_{n+1}^{22} = J_n^{22} + H_n^{22} - (J_n^{12} + H_n^{12})^T \cdot [J_n^{11} + H_n^{11}]^{-1} (J_n^{12} + H_n^{12}), \quad (69)$$

can be derived. Note that in the stationary case, where $H_n^{11}, \dots, H_n^{33}$ do not depend on n , the matrix sequence $\{J_n^{11}\}$ converges for $n \rightarrow \infty$ to the solution of the Riccati-type of equation

$$J_\infty^{11} = H_n^{33} - (H_n^{13})^T [J_\infty^{11} + H_n^{11}]^{-1} H_n^{13}. \quad (70)$$

The sequence $\{J_n^{12}\}$ either converges to a constant matrix

$$J_\infty^{12} = [I + (H_n^{13})^T (J_\infty^{11} + H_n^{11})^{-1}]^{-1} \cdot [(H_n^{23})^T - (H_n^{13})^T (J_\infty^{11} + H_n^{11})^{-1} H_n^{12}] \quad (71)$$

or diverges to infinity when at least one of the eigenvalues of $(H_n^{13})^T (J_\infty^{11} + H_n^{11})^{-1}$ has magnitude larger or equal to one.

The matrices J_n^{22} in (69) grow without any bound in general. If this happens, then the limit PCRB for estimating $x_n^{(1)}$ for $n \rightarrow \infty$ is the same as if $x_n^{(2)}$ were known. Indeed, these results can be expected because if the data bear any information about the parameter $x_n^{(2)}$, this information is accumulated as the time n goes to infinity.

Another example of application of Proposition 2 is given in Example 2 in the next section.

Case 2—Nonlinear g_n : The main idea for handling the singular case of the nonlinear filter in (51)–(53) is to “regularize” the filter, e.g., to replace (52) by a perturbed equation

$$x_{n+1}^{(2)} = g(x_n, x_{n+1}^{(1)}) + w_n^{(2)} \quad (72)$$

where $\{w_n^{(2)}\}$ is a sequence of pairwise independent Gaussian random vectors with zero mean and covariance matrix εI , independent of $\{w_n\}$ and $\{v_n\}$, with ε close to 0. For the modified system, it is possible to apply the result (21) from Section III.

Let $p_\varepsilon(\cdot)$'s and E_ε denote probability densities and the expectation operator induced by the perturbed system (51), (53), and (72). Note that

$$p_\varepsilon(x_{n+1}|x_n) = p(x_{n+1}^{(1)}|x_n) \cdot p_\varepsilon(x_{n+1}^{(2)}|x_n, x_{n+1}^{(1)}) \quad (73)$$

where $p(x_{n+1}^{(1)}|x_n)$ is determined by (51), and

$$-\log p_\varepsilon(x_{n+1}^{(2)}|x_n, x_{n+1}^{(1)}) = c_3 + \frac{1}{2\varepsilon} \|x_{n+1}^{(2)} - g[x_n, x_{n+1}^{(1)}]\|^2 \quad (74)$$

where c_3 is a constant. The matrices $D_{\varepsilon,n}^{11}, \dots, D_{\varepsilon,n}^{22}$ for the regularized system can be written as

$$D_{\varepsilon,n}^{ij} = \bar{D}_{\varepsilon,n}^{ij} + \frac{1}{\varepsilon} K_{\varepsilon,n}^{ij} \quad i, j = 1, 2 \quad (75)$$

where $\bar{D}_{\varepsilon,n}^{ij}$ is given as an E_ε -expectation of the second-order derivative of $-\log p(x_{n+1}^{(1)}|x_n)$ w.r.t. x_n and x_{n+1} , as in (22)–(25), $\bar{D}_{\varepsilon,n}^{22}$ contains, in addition, an E_ε -expectation of the second-order derivative of $-\log p(z_{n+1}|x_{n+1})$ w.r.t. x_{n+1} , and $K_{\varepsilon,n}^{ij}$, $i, j = 1, 2$ are given as an E_ε -expectation of the same derivatives of $\frac{1}{2}\|x_{n+1}^{(2)} - g(x_n, x_{n+1}^{(1)})\|^2$. In particular

$$K_{\varepsilon,n}^{11} = E_\varepsilon\{[\nabla_{x_n} g^T][\nabla_{x_n} g^T]^T\} \quad (76)$$

$$K_{\varepsilon,n}^{12} = [E_\varepsilon\{[\nabla_{x_n} g^T][\nabla_{x_{n+1}^{(1)}} g^T]^T\} \quad E_\varepsilon\{\nabla_{x_n} g^T\}] \quad (77)$$

$$K_{\varepsilon,n}^{22} = \begin{bmatrix} E_\varepsilon\{[\nabla_{x_{n+1}^{(1)}} g^T][\nabla_{x_{n+1}^{(1)}} g^T]^T\} & -E_\varepsilon\{\nabla_{x_{n+1}^{(1)}} g^T\} \\ -E_\varepsilon\{[\nabla_{x_{n+1}^{(1)}} g^T]^T\} & I \end{bmatrix} \quad (78)$$

where the arguments of g are omitted for brevity. The information submatrix for the original system will be obtained from the result (21) in the limit $\varepsilon \rightarrow 0$

$$J_{n+1} = \lim_{\varepsilon \rightarrow 0} \left[\bar{D}_{\varepsilon,n}^{22} + \frac{1}{\varepsilon} K_{\varepsilon,n}^{22} - \left(\bar{D}_{\varepsilon,n}^{21} + \frac{1}{\varepsilon} K_{\varepsilon,n}^{21} \right) \cdot \left(J_n + \bar{D}_{\varepsilon,n}^{11} + \frac{1}{\varepsilon} K_{\varepsilon,n}^{11} \right)^{-1} \left(\bar{D}_{\varepsilon,n}^{12} + \frac{1}{\varepsilon} K_{\varepsilon,n}^{12} \right) \right]. \quad (79)$$

An example of application of (79) is given in Example 3 in the following section.

V. EXAMPLES

Example 1—AR Process with Time-Varying Parameters: Consider a scalar-valued random process $\{z_n\}$ and introduce the notation

$$\underline{z}_n = [z_n, z_{n-1}, \dots, z_{n-r+1}]^T. \quad (80)$$

Let z_n obey the recursion

$$z_{n+1} = x_{n+1}^T \underline{z}_n + v_n \quad (81)$$

where x_{n+1} is a vector of instantaneous autoregressive coefficients at time instant n , and $\{v_n\}$ is a Gaussian white noise with zero mean and variance σ^2 . Further, assume that x_n has Gaussian random increments

$$x_{n+1} = x_n + w_n \quad (82)$$

where $\{w_n\}$ is white, independent of $\{v_n\}$, zero mean, and has covariance matrices $\{Q_n\}$.

The system (81) and (82) has the form of (46) and (47). The information submatrix J_n can be obtained by a straightforward application of (21) and (34)–(37). The result is

$$D_n^{11} = -D_n^{12} = Q_n^{-1} \quad (83)$$

$$D_n^{22} = Q_n^{-1} + Z_n \quad (84)$$

where

$$Z_n = \sigma^{-2} \mathbb{E}\{\underline{z}_n \underline{z}_n^T\} \quad (85)$$

so that

$$J_{n+1} = Q_n^{-1} + Z_n - Q_n^{-1} [J_n + Q_n^{-1}]^{-1} Q_n^{-1}. \quad (86)$$

Note that the optimum estimate of x_n from the data Z_n in the mean-of-square sense is the Kalman filter; the conditional distribution of x_n given Z_n is Gaussian. Let $\hat{x}_{n|n}$ and $\Sigma_{n|n}$ denote parameters of this distribution, namely, the mean and the covariance matrix. As mentioned in the introduction, the PCRB is tight in this case, and J_n is equal to the expected value of $\Sigma_{n|n}^{-1}$. Note that $\Sigma_{n|n}^{-1}$ in the Kalman filter obeys the same recursion as J_n with the exception that Z_n in (85) is replaced by $\sigma^{-2} \underline{z}_n \underline{z}_n^T$ without the expectation operator.

In order to achieve practical conclusions from the above theory, assume that drift of the autoregressive parameter is slow, i.e., that the trace of Q_n is much lower than 1, and

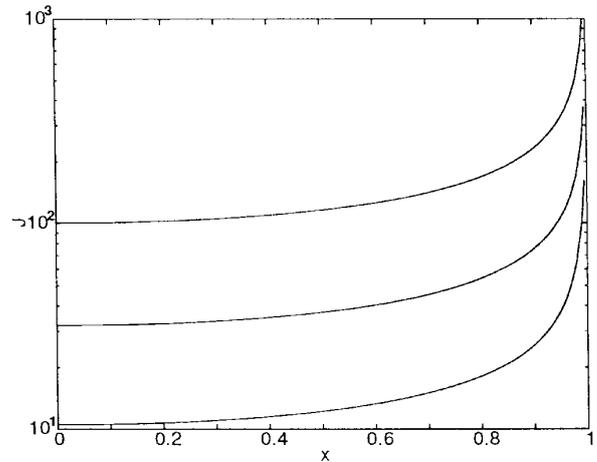


Fig. 1. Fisher information for slowly varying parameter of an AR(1) process as a function of this parameter for $Q = 10^{-2}$, 10^{-3} , and 10^{-4} (from the bottom up), respectively.

that x_n fluctuates around a mean value \bar{x} for a considerably long period of time. Then, the covariance function of $\{z_n\}$ is approximately equal to the covariance function of an AR process with parameter \bar{x} . The matrix Z_n in (85) can be replaced by a covariance matrix \bar{Z} of the above process, which is a function of \bar{x} . Note that \bar{Z} is independent of the variance of innovations σ^2 . Some methods for calculating the covariance matrix of an AR process are presented, e.g., in [14]. For example, for the first-order autoregressive process [abbreviated as AR(1) in the sequel]

$$\bar{Z} = \frac{1}{1 - \bar{x}^2} \quad (87)$$

holds. Here, \bar{x} is restricted to the interval $(-1, 1)$ to assure stability of the model. If, in addition, the matrix sequence $\{Q_n\}$ is constant, $Q_n \equiv Q$, and it is possible to calculate the limit information matrix (which is a scalar, in the case of $r = 1$) from the equation

$$J_\infty = Q^{-1} + \bar{Z} - Q^{-1} [J_\infty + Q^{-1}]^{-1} Q^{-1}. \quad (88)$$

In particular, for the AR(1) process, we obtain the solution

$$J_\infty = \frac{1}{2(1 - \bar{x}^2)} \left[1 + \sqrt{1 + 4Q^{-1}(1 - \bar{x}^2)} \right]. \quad (89)$$

Numerical values of (89) for $Q = 10^{-2}$, 10^{-3} , and 10^{-4} are plotted in Fig. 1. It is shown that the information about the parameter increases rapidly if the pole approaches unity. For the pole well separated from ± 1 , i.e., $\bar{x} \approx 0$, it holds that $J_\infty \approx Q^{-1/2}$.

The matrix J_∞ in (88) [or the corresponding scalar in (89) in the special case] describes the information content that the AR process bears about the fluctuating AR parameter. This information content depends on the actual value of the estimated parameter. If it happens that J_∞ is small and, consequently, that the limit PCRB J_∞^{-1} is large, it indicates that the assumed data model might not be appropriate.

Example 2—Sinusoidal Frequency Estimation: In this subsection, the developed methodology is applied to computation of the posterior CRB for tracking parameters of a single noisy cisoid with slowly varying frequency. This computation is easier than those recently presented in [16]. Second, as a special case of a single time-invariant frequency, the well known Cramér–Rao bound by Rife and Boorstyn [9] is derived.

The signal is assumed to have the form

$$z_n = m_0 e^{i\varphi_n} + v_n \quad n = 0, 1, 2, \dots \quad (90)$$

where

- m_0 magnitude;
- φ_n instantaneous phase of cisoid at time instant n ;
- $\{v_n\}$ noise.

The instantaneous frequency (denoted ω_n) is defined as the one-step increment of φ_n . Thus, the signal with randomly varying frequency can be described by the state vector

$$x_n = (\omega_n, \varphi_n)^T \quad (91)$$

and time update of x_n is given by the pair of the equations

$$\omega_{n+1} = \omega_n + e_n \quad (92)$$

$$\varphi_{n+1} = \varphi_n + \omega_{n+1} = \varphi_n + \omega_n + e_n. \quad (93)$$

It is assumed that $\{v_n\}$ and $\{e_n\}$ are independent sequences of independent random variables with zero mean values and variances σ^2 and γ^2 , respectively; $\{e_n\}$ is Gaussian, and $\{v_n\}$ is complex circular Gaussian (i.e., the real and imaginary parts of $\{v_n\}$ are independent normally distributed with zero means and equal variances $\sigma^2/2$). Next, assume that the probability distribution of the initial instantaneous phase and frequency is known.

Obviously, in the standard formulation, the covariance matrix of the system noise $w_n = (e_n, e_n)$ is not invertible, and the conditional probability $p(x_{n+1}|x_n)$ is singular. The calculation of the information submatrix as in Section II fails, but it is possible to apply the approach developed in Section III with $x_n^{(1)} = \omega_n$ and $x_n^{(2)} = \varphi_n$. Comparing (93) with (54), we get $G_n^{(1)} = 0$, and $G_n^{(2)} = G_n^{(3)} = 1$. The assumed probability distributions of the noise $\{e_n\}$ and $\{v_n\}$ imply that

$$\begin{aligned} -\log p(x_{n+1}^{(1)}|x_n) &= -\log p(\omega_{n+1}|\omega_n) \\ &= c_4 + \frac{1}{2\gamma^2} (\omega_n - \omega_{n+1})^2 \end{aligned} \quad (94)$$

$$\begin{aligned} -\log p(z_{n+1}|x_n, x_{n+1}^{(1)}) &= -\log p(z_{n+1}|\varphi_n, \omega_{n+1}) \\ &= c_5 + \frac{1}{\sigma^2} |m_0 e^{i(\varphi_n + \omega_{n+1})} - z_{n+1}|^2 \end{aligned} \quad (95)$$

where c_4 and c_5 are normalization constants. A straightforward calculation of (60)–(65) gives

$$H_n^{11} = \frac{1}{\gamma^2}, \quad H_n^{12} = 0, \quad H_n^{13} = -\frac{1}{\gamma^2} \quad (96)$$

$$H_n^{22} = H_n^{23} = \frac{2m_0^2}{\sigma^2}, \quad H_n^{33} = \frac{2m_0^2}{\sigma^2} + \frac{1}{\gamma^2}. \quad (97)$$

Inserting the above relations into (56) and (57) and we get, after some simplifications, (98), shown at the bottom of the page, where

$$d_n = \det J_n = J_n^{\omega\omega} J_n^{\varphi\varphi} - [J_n^{\omega\varphi}]^2. \quad (99)$$

In (98) and (99), $J_n^{\omega\omega}$, $J_n^{\omega\varphi}$, and $J_n^{\varphi\varphi}$ denote elements of the matrix J_n .

The stationary solution of (98) can be found by putting $J_n = J_{n+1} = J_\infty$. After excluding the terms $J_\infty^{\omega\varphi}$ and $J_\infty^{\varphi\omega}$, a fourth-order polynomial equation for $J_\infty^{\omega\omega}$ is obtained. This equation can be shown to have only one positive real-valued root. The final result is

$$J_\infty = \frac{1}{2\gamma^2} \begin{bmatrix} h & -\frac{h^2}{4+h} \\ -\frac{h^2}{4+h} & \frac{2h^3}{(4+h)^2} \end{bmatrix} \quad (100)$$

where

$$h = w + \sqrt{w^2 + 4w} \quad (101)$$

$$w = \Gamma + \sqrt{\Gamma^2 + 8\Gamma} \quad (102)$$

$$\Gamma = \gamma^2 \frac{m_0^2}{\sigma^2}. \quad (103)$$

The limit PCRb on the instantaneous frequency is equal to the left-upper corner element of J_∞^{-1} , i.e.,

$$\begin{aligned} \text{LPCRB}(\hat{\omega}_n) &= [J_\infty^{\omega\omega} - (J_\infty^{\omega\varphi})^2/J_\infty^{\varphi\varphi}]^{-1} = \gamma^2 \frac{4}{h} \\ &= \frac{\gamma^2}{w} (-w + \sqrt{w^2 + 4w}) \end{aligned} \quad (104)$$

which coincides with the result derived in [16].

Finally, let us consider estimation of stationary frequency, i.e., put $\gamma^2 = 0$. Then, (98) is reduced to

$$\begin{aligned} J_{n+1} &= \begin{bmatrix} J_{n+1}^{\omega\omega} & J_{n+1}^{\omega\varphi} \\ J_{n+1}^{\omega\varphi} & J_{n+1}^{\varphi\varphi} \end{bmatrix} \\ &= \begin{bmatrix} J_n^{\omega\omega} - 2J_n^{\omega\varphi} + J_n^{\varphi\varphi} & J_n^{\omega\varphi} - J_n^{\varphi\omega} \\ J_n^{\omega\varphi} - J_n^{\varphi\omega} & J_n^{\varphi\varphi} + \frac{2m_0^2}{\sigma^2} \end{bmatrix}. \end{aligned} \quad (105)$$

For $J_0 = 0$ (no *a priori* information about the frequency and phase), the recursion (105) has a solution

$$J_n^{\varphi\varphi} = n \frac{2m_0^2}{\sigma^2} \quad (106)$$

$$J_n^{\omega\varphi} = -\sum_{k=1}^{n-1} J_k^{\varphi\omega} = -n(n-1) \frac{m_0^2}{\sigma^2} \quad (107)$$

$$J_n^{\omega\omega} = \sum_{k=1}^{n-1} (J_k^{\varphi\varphi} - 2J_k^{\omega\varphi}) = n(n-1)(2n-1) \frac{m_0^2}{3\sigma^2}. \quad (108)$$

$$J_{n+1} = \frac{1}{1 + \gamma^2 J_n^{\omega\omega}} \cdot \begin{bmatrix} \gamma^2 d_n + J_n^{\omega\omega} - 2J_n^{\omega\varphi} + J_n^{\varphi\varphi} & -\gamma^2 d_n + J_n^{\omega\varphi} - J_n^{\varphi\omega} \\ -\gamma^2 d_n + J_n^{\omega\varphi} - J_n^{\varphi\omega} & \gamma^2 d_n + J_n^{\varphi\varphi} + \frac{2m_0^2}{\sigma^2} (1 + \gamma^2 J_n^{\omega\omega}) \end{bmatrix} \quad (98)$$

The PCRFB on the frequency is equal to the left-upper corner element of J_n^{-1} , i.e.,

$$\begin{aligned} \text{PCRFB}(\hat{\omega}_n) &= [J_n^{\omega\omega} - (J_n^{\omega\varphi})^2 / J_n^{\varphi\varphi}]^{-1} \\ &= \frac{6}{n(n^2 - 1)} \frac{\sigma^2}{m_0^2} \end{aligned} \quad (109)$$

which coincides with the CRB for the problem [9].

Example 3—Sinusoidal Signal with Sinusoidal Phase Modulation: Consider a sinusoidal signal as in (90), define the instantaneous frequency ω_n of the carrier as a one-step backward difference of the instantaneous phase φ_n as in (93), and assume that the frequency evolves in time like a sinusoid within the range $(-\pi, \pi)$. We refer to this sinusoid as a message and assume that the frequency of the message evolves like a random walk. Note that an algorithm for tracking parameters of signals of this kind was proposed in [17].

At each time instant, the signal can be characterized by a state vector with three components

$$x_n = [\nu_n, \phi_n, \varphi_n]^T \quad (110)$$

where

- φ_n instantaneous phase of the carrier;
- ϕ_n instantaneous phase of the message;
- ν_n frequency of the message.

Assume that the instantaneous frequency of the carrier equals

$$\omega_n = \omega_c + \eta \cos \phi_n \quad (111)$$

where ω_c is the central frequency of the carrier, and η is the maximum deviation of the carrier frequency from ω_c .

The time update of the state vector is given by the set of equations

$$\nu_{n+1} = \nu_n + c_n \quad (112)$$

$$\phi_{n+1} = \phi_n + \nu_{n+1} \quad (113)$$

$$\begin{aligned} \varphi_{n+1} &= \varphi_n + \omega_c + \eta \cos \phi_{n+1} \\ &= \varphi_n + \omega_c + \eta \cos(\phi_n + \nu_{n+1}). \end{aligned} \quad (114)$$

As in the previous subsection, assume that $\{c_n\}$ is a Gaussian white noise with variance γ^2 . The filtering in (112)–(114) and (90) is an example of the singular case from Section IV with nonlinear function g_n and

$$x_n^{(1)} = \nu_n \quad (115)$$

$$x_n^{(2)} = [\phi_n, \varphi_n]^T \quad (116)$$

$$-\log p(x_{n+1}^{(1)} | x_n) = c_0 + \frac{1}{2\gamma^2} (\nu_{n+1} - \nu_n)^2 \quad (117)$$

and

$$\begin{aligned} \frac{1}{2} \|x_{n+1}^{(2)} - g(x_n, x_{n+1}^{(1)})\|^2 \\ = \frac{1}{2} (\phi_{n+1} - \phi_n - \nu_{n+1})^2 \\ + \frac{1}{2} [\varphi_{n+1} - \varphi_n - \omega_c - \eta \cos(\phi_n + \nu_{n+1})]^2. \end{aligned} \quad (118)$$

A straightforward calculation gives

$$\bar{D}_{\varepsilon, n}^{11} = -\bar{D}_{\varepsilon, n}^{12} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (119)$$

$$\bar{D}_{\varepsilon, n}^{22} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{2m_0^2}{\sigma^2} \end{bmatrix} \quad (120)$$

$$K_{\varepsilon, n}^{11} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & a_{\varepsilon, n} & b_{\varepsilon, n} \\ 0 & b_{\varepsilon, n} & 1 \end{bmatrix} \quad (121)$$

$$K_{\varepsilon, n}^{12} = \begin{bmatrix} 0 & 0 & 0 \\ a_{\varepsilon, n} & -1 & -b_{\varepsilon, n} \\ b_{\varepsilon, n} & 0 & -1 \end{bmatrix} \quad (122)$$

$$K_{\varepsilon, n}^{22} = \begin{bmatrix} a_{\varepsilon, n} & -1 & -b_{\varepsilon, n} \\ -1 & 1 & 0 \\ -b_{\varepsilon, n} & 0 & 1 \end{bmatrix} \quad (123)$$

where

$$a_{\varepsilon, n} = 1 + \eta^2 E_{\varepsilon} \{\sin^2(\phi_n + \nu_{n+1})\} \quad (124)$$

$$b_{\varepsilon, n} = -\eta E_{\varepsilon} \{\sin(\phi_n + \nu_{n+1})\}. \quad (125)$$

An available but tedious method of computing an approximate value of J_n is to choose a small fixed ε , do a number of independent simulations of the data according to the “regularized” model, and replace the expectations in (124) and (125) by corresponding sample averages. Then, evaluate J_n as in (79).

Another approach for computing J_n can be utilized in cases when the rate of evolution of ν_n , i.e., the variance γ^2 , and the variance of the observation noise are small. Consider sequences $\{\bar{\nu}_n\}$, $\{\bar{\phi}_n\}$, $\{\bar{\varphi}_n\}$ that obey (112)–(114) with $\{c_n \equiv 0\}$ (this is called an “equilibrium state” in [15]), and assume that the probability densities of $\{\nu_n\}$, $\{\phi_n\}$, $\{\varphi_n\}$ are concentrated in neighborhoods of $\{\bar{\nu}_n\}$, $\{\bar{\phi}_n\}$, and $\{\bar{\varphi}_n\}$. Then, $a_{\varepsilon, n}$ and $b_{\varepsilon, n}$ are approximated by

$$a_{\varepsilon, n} \approx \bar{a}_n \triangleq 1 + \eta^2 \sin^2(\bar{\phi}_n + \bar{\nu}_{n+1}) \quad (126)$$

$$b_{\varepsilon, n} \approx \bar{b}_n \triangleq -\eta \sin(\bar{\phi}_n + \bar{\nu}_{n+1}). \quad (127)$$

Using the above approximation the limit in (79) can be evaluated analytically. The result, which is obtained with the aid of symbolic Mathematica, is

$$J_{n+1}^{\nu\nu} = (J_n^{\nu\nu} + J_n^{\phi\phi} - 2J_n^{\nu\phi} + \gamma^2 d_1) / (1 + \gamma^2 J_n^{\nu\nu}) \quad (128)$$

$$\begin{aligned} J_{n+1}^{\nu\phi} &= [J_n^{\nu\phi} - J_n^{\phi\phi} - \gamma^2 d_1 + \gamma^2 d_2 \bar{b}_n + (J_n^{\phi\varphi} - J_n^{\nu\varphi}) \bar{b}_n] \\ &\quad / (1 + \gamma^2 J_n^{\nu\nu}) \end{aligned} \quad (129)$$

$$J_{n+1}^{\nu\varphi} = (J_n^{\nu\varphi} - J_n^{\phi\varphi} - \gamma^2 d_2) / (1 + \gamma^2 J_n^{\nu\nu}) \quad (130)$$

$$\begin{aligned} J_{n+1}^{\phi\phi} &= [J_n^{\phi\phi} + \bar{b}_n^2 J_n^{\varphi\varphi} - 2\bar{b}_n J_n^{\phi\varphi} + \gamma^2 (d_1 - 2\bar{b}_n d_2 + \bar{b}_n^2 d_3)] \\ &\quad / (1 + \gamma^2 J_n^{\nu\nu}) \end{aligned} \quad (131)$$

$$\begin{aligned} J_{n+1}^{\phi\varphi} &= [J_n^{\phi\varphi} - \bar{b}_n J_n^{\varphi\varphi} + \gamma^2 (d_2 - \bar{b}_n d_3)] \\ &\quad / (1 + \gamma^2 J_n^{\nu\nu}) \end{aligned} \quad (132)$$

$$J_{n+1}^{\varphi\varphi} = (J_n^{\varphi\varphi} + \gamma^2 d_3) / (1 + \gamma^2 J_n^{\nu\nu}) + \frac{2m_0^2}{\sigma^2}. \quad (133)$$

where

$$d_1 = J_n^{\nu\nu} J_n^{\phi\phi} - [J_n^{\nu\phi}]^2 \quad (134)$$

$$d_2 = J_n^{\nu\nu} J_n^{\phi\varphi} - J_n^{\nu\phi} J_n^{\nu\varphi} \quad (135)$$

$$d_3 = J_n^{\nu\nu} J_n^{\varphi\varphi} - [J_n^{\nu\varphi}]^2 \quad (136)$$

and $J_n^{\nu\nu}, \dots, J_n^{\varphi\varphi}$ are the elements of J_n .

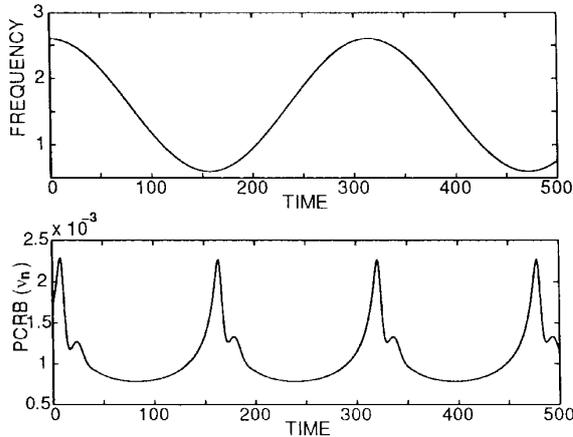


Fig. 2. Instantaneous frequency of the carrier of a sinusoidal signal and the PCRB on the signal frequency as functions of time in the model considered in Example 3.

To illustrate the above result, consider a signal of the length $N = 500$ with the following parameters: $m_0^2/\sigma^2 = 1/2$, $\bar{\nu} = 0.02$, $\omega_c = 1.6$, $\eta = 1$, $\gamma^2 = 10^{-4}$, $\phi_0 = 0$. Fig. 2 shows the posterior CRB on parameter ν_n , which was derived from J_n , as a function of time. Simultaneously, the instantaneous frequency of the carrier is plotted. Note that the nonlinear character of the signal model implies that the PCRB does not converge to any limit value for $n \rightarrow \infty$, but it is periodic in time with the frequency that is twice greater than the frequency of the message ν_n . In particular, if the frequency of the carrier is close to its minimum or maximum and its rate of change is low, the amount of information that the signal bears about the possible changes of ν_n is small, the PCRB increases, and vice versa.

VI. CONCLUSIONS

A simple and straightforward derivation of the posterior Cramér-Rao lower bound for the discrete-time nonlinear filtering problem was presented. Explicit realizations of this lower bound were calculated for three important examples.

- 1) tracking a slowly varying AR parameter;
- 2) tracking a slowly varying sinusoidal frequency;
- 3) tracking a slowly varying frequency that is modulated by a sinusoid.

The derived lower bound can be used for evaluating the performance of existing suboptimal methods of nonlinear filtering. It is believed that a similar bound can be derived for a more general model of nonlinear autoregressive systems as well.

APPENDIX

PROOF OF PROPOSITION 2

The proof of Proposition 2 utilizes the following lemma.

Lemma 1: Consider the problem of estimating a random vector x from an observation vector z . Let $p(x, z)$ be the joint probability density of (x, z) , and assume that information

matrix

$$J(x) = E\{-\Delta_x^2 \log p(x, z)\}. \quad (137)$$

exists. Let $y = Mx$, where M is a constant invertible matrix. Then, the probability density $p(y, z)$ exists, and the corresponding information matrix for estimating y is given by

$$J(y) = M^{-T} J(x) M^{-1}. \quad (138)$$

Proof: The proof is based on the well-known rule for change of coordinates of the estimated parameters (see, e.g., [13]), is straightforward, and is therefore omitted here. ■

Proof of Proposition: Let p_n denote the probability density of the triplet $[X_n^{(1)}, x_n^{(2)}, Z_n]$

$$p_n \triangleq p(X_n^{(1)}, x_n^{(2)}, Z_n) = p(X_{n-1}^{(1)}, x_n^{(1)}, x_n^{(2)}, Z_n). \quad (139)$$

It will be shown by induction that p_n exists.

The information matrix that corresponds to the triplet $[X_{n-1}^{(1)}, x_n^{(1)}, x_n^{(2)}]$ can be written in block form as

$$J(X_{n-1}^{(1)}, x_n^{(1)}, x_n^{(2)}) = \begin{bmatrix} A_n & B_n & C_n \\ B_n^T & D_n & E_n \\ C_n^T & E_n^T & F_n \end{bmatrix} \quad (140)$$

where the blocks A_n, B_n, \dots, F_n are obtained as expectations of the second-order derivatives of $-\log p_n$ with respect to $X_{n-1}^{(1)}, x_n^{(1)}$, and $x_n^{(2)}$.

The information submatrix for the state vector x_n can be obtained as the inverse of the right-lower submatrix of $\{J(X_{n-1}^{(1)}, x_n^{(1)}, x_n^{(2)})\}^{-1}$, i.e.,

$$\begin{aligned} J_n &\triangleq \begin{bmatrix} J_n^{11} & J_n^{12} \\ J_n^{21} & J_n^{22} \end{bmatrix} \\ &= \begin{bmatrix} D_n & E_n \\ E_n^T & F_n \end{bmatrix} - \begin{bmatrix} B_n^T \\ C_n^T \end{bmatrix} A_n^{-1} [B_n, C_n] \\ &= \begin{bmatrix} D_n - B_n^T A_n^{-1} B_n & E_n - B_n^T A_n^{-1} C_n \\ E_n^T - C_n^T A_n^{-1} B_n & F_n - C_n^T A_n^{-1} C_n \end{bmatrix}. \end{aligned} \quad (141)$$

Consider the probability density of the quartet $[X_n^{(1)}, x_n^{(2)}, x_{n+1}^{(1)}, Z_{n+1}]$, denoted by \tilde{p}_{n+1} . Note that two vectors

$$\begin{aligned} \begin{bmatrix} X_{n-1}^{(1)} \\ x_n^{(1)} \\ x_{n+1}^{(1)} \\ x_{n+1}^{(2)} \end{bmatrix} &= \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & I \\ 0 & G_n^{(1)} & G_n^{(2)} & G_n^{(3)} \end{bmatrix} \begin{bmatrix} X_{n-1}^{(1)} \\ x_n^{(1)} \\ x_n^{(2)} \\ x_{n+1}^{(1)} \end{bmatrix} \\ &\triangleq \tilde{M}_n \begin{bmatrix} X_{n-1}^{(1)} \\ x_n^{(1)} \\ x_n^{(2)} \\ x_{n+1}^{(1)} \end{bmatrix} \end{aligned} \quad (142)$$

obey the linear relationship. Since $G_n^{(2)}$ is assumed to be regular, it follows that \tilde{M}_n is regular as well. Applying Lemma 1, it follows by induction that p_n in (139) exists for each n , and

$$J(X_n^{(1)}, x_{n+1}^{(1)}, x_{n+1}^{(2)}) = \tilde{M}_n^{-T} J(X_n^{(1)}, x_n^{(2)}, x_{n+1}^{(1)}) \tilde{M}_n^{-1}. \quad (143)$$

Using conditional densities, \tilde{p}_{n+1} can be written as the product

$$\begin{aligned}\tilde{p}_{n+1} &= p_n p(x_{n+1}^{(1)} | X_n^{(1)}, x_n^{(2)}, Z_n) \\ &\quad \cdot p(z_{n+1} | X_n^{(1)}, x_n^{(2)}, x_{n+1}^{(1)}, Z_n) \\ &= p_n p(x_{n+1}^{(1)} | x_n) p(z_{n+1} | x_n, x_{n+1}^{(1)}).\end{aligned}\quad (144)$$

The second equality in (144) follows from the formulation of the filtering problem. From (140) and (144), it follows that

$$\begin{aligned}J(X_n^{(1)}, x_n^{(2)}, x_{n+1}^{(1)}) \\ = \begin{bmatrix} A_n & B_n & C_n & 0 \\ B_n^T & D_n + H_n^{11} & E_n + H_n^{12} & H_n^{13} \\ C_n^T & (E_n + H_n^{12})^T & F_n + H_n^{22} & H_n^{23} \\ 0 & (H_n^{13})^T & (H_n^{23})^T & H_n^{33} \end{bmatrix}\end{aligned}\quad (145)$$

where H_n^{ij} , $i, j = 1, 2, 3$ were defined in (60)–(65). The information submatrix for $[x_n^{(1)}, x_n^{(2)}, x_{n+1}^{(1)}]$ then equals

$$\begin{aligned}\tilde{S}_{n+1} &\triangleq \begin{bmatrix} D_n + H_n^{11} & E_n + H_n^{12} & H_n^{13} \\ (E_n + H_n^{12})^T & F_n + H_n^{22} & H_n^{23} \\ (H_n^{13})^T & (H_n^{23})^T & H_n^{33} \end{bmatrix} \\ &\quad - \begin{bmatrix} B_n^T \\ C_n^T \\ 0 \end{bmatrix} A_n^{-1} [B_n, C_n, 0].\end{aligned}\quad (146)$$

This can be rewritten using (141) as

$$\tilde{S}_{n+1} = \begin{bmatrix} J_n^{11} + H_n^{11} & J_n^{12} + H_n^{12} & H_n^{13} \\ (J_n^{12} + H_n^{12})^T & J_n^{22} + H_n^{22} & H_n^{23} \\ (H_n^{13})^T & (H_n^{23})^T & H_n^{33} \end{bmatrix}.\quad (147)$$

Combining (143) and the form of \tilde{M}_n in (59) and (142) implies

$$S_{n+1} = M_n^{-T} \tilde{S}_{n+1} M_n^{-1}.\quad (148)$$

The statement follows. ■

REFERENCES

- [1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [2] B. Z. Bobrovsky and M. Zakai, "A lower bound on the estimation error for Markov processes," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 785–788, Dec. 1975.
- [3] J. I. Galdos, "A Cramér–Rao bound for multidimensional discrete-time dynamical systems," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 117–119, Feb. 1980.
- [4] P. C. Doerschuk, "Cramer–Rao bounds for discrete-time nonlinear filtering problems," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1465–1469, Aug. 1995.
- [5] G. A. Hewer, "An iterative technique for the computation of the steady state gains for the discrete optimal regulator," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 382–384, Feb. 1971.
- [6] T. H. Kerr, "Status of CR-like lower bounds for nonlinear filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, pp. 590–600, Sept. 1989.
- [7] V. Kučera, "Factorization of rational spectral matrices: A survey of methods," in *Proc. IEE Int. Conf. Contr.*, Edinburgh, U.K., 1991, pp. 1074–1078.
- [8] T. Pappas, A. J. Laub, and N. R. Sandel, "On the numerical solution of the discrete-time algebraic Riccati equation," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 631–641, June 1980.
- [9] D. C. Rife and R. R. Boorstyn, "Multiple tone parameter estimation from discrete time observation," *Bell Syst. Tech. J.*, pp. 1389–1410, Nov. 1976.
- [10] D. J. Sakrison *Notes on Analog Communication*. Princeton, NJ: Van Nostrand, 1970.
- [11] H. L. van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968.
- [12] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, "Some classes of global Cramér–Rao bounds," *Ann. Statist.*, vol. 15, no. 4, pp. 1421–1438, 1987.
- [13] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [14] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. New York: Springer-Verlag, 1987.
- [15] P. Tichavský and P. Händel, "Two algorithms for adaptive retrieval of slowly varying multiple cisoids in noise," *IEEE Trans. Signal Processing*, vol. 43, pp. 1116–1127, May 1995.
- [16] P. Tichavský, "Posterior Cramér–Rao bounds for adaptive harmonic retrieval," *IEEE Trans. Signal Processing*, vol. 43, pp. 1299–1302, May 1995.
- [17] P. Tichavský and P. Händel, "Recursive estimation of linearly or harmonically modulated frequencies of multiple cisoids in noise," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, München, Germany, Apr. 1997, pt. III, pp. 1925–1928.
- [18] P. Tichavský, C. Muravchik, and A. Nehorai, "Posterior C–R bounds for performance of adaptive parameter estimation," in *Signal Anal. Prediction I, Proc. First Euro. Conf. Signal Anal. Prediction*, A. Procházka, J. Uhlř and P. Sovka, Eds. Prague, Czech Republic, June 24–27, 1997, pp. 153–156.



Petr Tichavský (M'98) graduated in 1987 from the Czech Technical University, Prague, Czechoslovakia. He received the Ph.D. degree in theoretical cybernetics from the Czechoslovak Academy of Sciences, Prague, in 1992.

He is now with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague. In 1994, he received the Fulbright grant for a ten-month fellowship at the Department of Electrical Engineering, Yale University, New Haven, CT. His current research interests

include adaptive parameter estimation and spectrum analysis.



Carlos H. Muravchik (S'81–M'83) graduated in 1973 as an electronics engineer from the National University of La Plata, La Plata, Argentina. He received the M.Sc. degree in statistics in 1983 and the M.Sc. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1980 and 1983, respectively.

He is currently a Professor with the Department of Electrical Engineering, National University of La Plata, and a member of its Industrial Electronics, Control, and Instrumentation Laboratory. He is also a member of the Comisión de Investigaciones Científicas de la Pacífica, Buenos Aires, Argentina. He was a Visiting Professor at Yale University, New Haven, CT, in 1983 and 1994 and at the University of Illinois, Chicago, in 1996. His research interests are in the areas of statistical signal and array processing with biomedical, control, and communications applications and in nonlinear control systems.



Arye Nehorai (S'80–M'83–SM'90–F'94) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, in 1976 and 1979, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1983.

After graduation, he was a Research Engineer at Systems Control Technology, Inc., Palo Alto, CA. From 1985 to 1995, he was with the Department of Electrical Engineering, Yale University, New Haven, CT, where he became an Associate Professor in 1989. In 1995, he joined the Department of Electrical Engineering and Computer Science, The University of Illinois, Chicago, as a Full Professor.

Dr. Nehorai has been an Associate Editor of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING and IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION. He has also been an Associate Editor of *Circuits, Systems, and Signal Processing*, and *The Journal of the Franklin Institute*. He served as the Chairman of the Connecticut IEEE Signal Processing Chapter from 1986 to 1995 and was co-recipient of the 1989 IEEE Signal Processing Society's Senior Award for the Best Paper in Statistical Signal and Array Processing. He is a Fellow of the Royal Statistical Society.



PERGAMON

Automatica 37 (2001) 1703–1716

automatica

www.elsevier.com/locate/automatica

Filtering, predictive, and smoothing Cramér–Rao bounds for discrete-time nonlinear dynamic systems[☆]

Miroslav Šimandl^{a,*}, Jakub Královec^a, Petr Tichavský^b

^aDepartment of Cybernetics, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 30614 Pilsen, Czech Republic

^bInstitute for Information Theory and Automation, Academy of Sciences of the Czech Republic, Box 18, 18208 Prague, Czech Republic

Received 8 December 1999; revised 24 November 2000; received in final form 6 April 2001

Abstract

Cramér–Rao lower bounds for the discrete-time nonlinear state estimation problem are treated. The Cramér–Rao bound for the mean-square error matrix of a state estimate is particularly important for quality evaluation of nonlinear state estimators as it represents a limit of cognizability of the state. Recursive relations for filtering, predictive, and smoothing Cramér–Rao bounds are derived to establish a unifying framework for several previously published derivation procedures and results. Lower bounds for systems with unknown parameters are newly provided. Computation of filtering, predictive, and smoothing Cramér–Rao bounds, their mutual comparison and utilization for quality evaluation of some nonlinear filters are shown in numerical examples. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Nonlinear systems; Stochastic systems; Nonlinear state estimation; Cramér–Rao bound; Mean-square error; Filtering; Prediction; Smoothing

1. Introduction

A recursive state estimator for a nonlinear stochastic discrete-time system can be designed directly using the Bayesian approach. As the closed form solution of the Bayesian recursive relations is restricted to a few single cases, it is necessary to provide numerical or analytical approximations (Sorenson, 1974; Kulhavý, 1996). Nonlinear filters based on such approximations generate estimates which are more or less affected by these approximations and deviate from the ideal exact solution. Quality evaluation of the nonlinear filters is one of the most complex problems in the area of nonlinear estimation. The knowledge of a lower bound for the mean-square error of an estimate can give an indication of estimator performance limitations, and consequently it

can be used to determine whether imposed performance requirements are realistic or not.

As is well known, the Cramér–Rao (CR) bound, defined as the inverse of the Fisher information matrix, represents an objective lower limit of cognizability of parameters in constant parameter estimation. The CR bound methodology was extended for random parameters estimation by Van Trees (1968). More recent discussions and extensions of the bound can be found in Bobrovsky, Mayer-Wolf, and Zakai (1987). The idea of the CR bound was successfully applied in state estimation for discrete-time nonlinear stochastic dynamic systems by Bobrovsky and Zakai (1975) and Galdos (1980). These works are based on a certain kind of “equivalence” between probability density functions of the original nonlinear stochastic system and an auxiliary linear Gaussian system. A survey and a detailed critical discussion of this approach to CR lower bounds for nonlinear filtering was presented by Kerr (1989). The approach by Galdos (1980) was generalized for nonlinear p th order autoregressive processes driven by additive Gaussian noise with state-dependent gain (Doerschuk, 1995).

An alternative approach to computation of the CR bound (called also posterior CR bound) for the filtering problem in discrete-time nonlinear systems was proposed

[☆]This is an expanded version of the paper presented at the 14th IFAC World Congress, 5–9 July 1999, Beijing. This paper was recommended for publication in revised form by Associate Editor Håkan Hjalmarsson under the direction of Editor Torsten Söderström.

*Corresponding author. Tel.: +420-19-7491171; fax: +420-19-279050.

E-mail address: simandl@kky.zcu.cz (M. Šimandl).

by Tichavský, Muravchik, and Nehorai (1998). The idea of this approach is to regard the state history as a random parameter vector. The CR bound for the state of the system is obtained as the lower right block of the CR bound for the complete state history. In the basic form of the bound it is assumed that the state transition probability density function (pdf) $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ exists and is twice differentiable with respect to both its arguments. This is rather restrictive assumption because the state disturbance may easily have lower dimension than the state, what follows that $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ need not exist or is not continuous. The same problem with singularity was encountered in the former approach as well, see Galdos (1980). One of the special cases, where this singularity occurs, is the case of nonlinear system with unknown constant parameters. The CR bound for filtering in such a system was derived already in the mentioned paper by Tichavský et al. (1998). Another special case is a system with linear state evolution but with nonlinear measurement. The posterior CR bound for one-step ahead prediction in these systems was derived by Bergman (1999).

The recursive relations for posterior CR bounds, based on derivation techniques using either filtering or one-step predictive pdf's, were alternatively derived in the form of time and measurement update steps and extended for the general multi-step prediction problem in Šimandl, Královec, and Tichavský (1999). Similar CR bound for smoothing problem was derived by Bergman (1999). The CR bound methodology was used for analysis of nonlinear filters, e.g. point-mass filter in a navigation and tracking application (Bergman, Ljung, & Gustafsson, 1997), and also for synthesis of the nonlinear Gaussian-sum filter (Šimandl & Královec, 1998).

The aim of the paper is to derive recursive relations for filtering, predictive, and smoothing CR bounds in order to present the previous results of Tichavský et al. (1998), Šimandl et al. (1999), and Bergman (1999) in a structurally unified form, and to make a further extension of the CR bounds for nonlinear stochastic systems with unknown parameters, keeping the established framework.

The paper is organized as follows. After a detailed specification of the main goal of the paper in Section 2, a derivation of the recursive relations for the filtering and one-step predictive CR bounds (in the form of time and measurement update steps) for nonlinear discrete-time stochastic system will be shown in Section 3. This derivation is crucial for expressing structurally unified recursive relations for multi-step predictive and smoothing CR bounds in Sections 4 and 5, respectively. Each of the Sections 3–5 will be completed by a solution of the CR bound problem for the special case of system with unknown parameters. Practical aspects of computation of the bounds and effects of additive Gaussian disturbances on a simplification of the relation for the CR bounds will be discussed in Section 6. Two numerical examples in

Section 7 present a comparison of filtering, predictive, and smoothing CR bounds and the use of the CR bound for quality evaluation of several nonlinear filters.

2. Problem statement

Consider the problem of estimating a vector of random parameters $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_n]^T$ from a set of measured data $\mathbf{z}^N = [\mathbf{z}_1^T \ \mathbf{z}_2^T \ \dots \ \mathbf{z}_N^T]^T$. The joint pdf $p(\mathbf{z}^N, \boldsymbol{\theta})$ is supposed to be known. To simplify notations, the nabla operator will be used

$$\nabla_{\boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial}{\partial \theta_1} & \frac{\partial}{\partial \theta_2} & \dots & \frac{\partial}{\partial \theta_n} \end{bmatrix}.$$

The Fisher information matrix (FIM) $\mathbf{J}(\boldsymbol{\theta})$ for the parameter vector $\boldsymbol{\theta}$ is defined as follows

$$\mathbf{J}(\boldsymbol{\theta}) = -E\{\nabla_{\boldsymbol{\theta}}[\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{z}^N, \boldsymbol{\theta})]^T\} \quad (1)$$

provided that the derivatives and expectation exist. An alternative formula for the FIM is

$$\mathbf{J}(\boldsymbol{\theta}) = E\{\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{z}^N, \boldsymbol{\theta})^T \nabla_{\boldsymbol{\theta}} \ln p(\mathbf{z}^N, \boldsymbol{\theta})\}. \quad (2)$$

Let $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{z}^N)$ be an estimate of the parameter vector $\boldsymbol{\theta}$. The mean-square error matrix (MSEM) defined as $\boldsymbol{\Pi}(\hat{\boldsymbol{\theta}}) = E\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\}$ is bounded by an inverse of the FIM

$$\boldsymbol{\Pi}(\hat{\boldsymbol{\theta}}) \geq \mathbf{J}^{-1}(\boldsymbol{\theta}). \quad (3)$$

The inverse $\mathbf{J}^{-1}(\boldsymbol{\theta})$ is called the posterior (alternatively global (Bobrovsky et al., 1987) or Bayesian (Gill & Levit, 1995) Cramér–Rao (CR) bound and it will be denoted as $\mathbf{C}(\boldsymbol{\theta})$. In contrast to CR bound for deterministic (nonrandom) parameters, it is not required that $\hat{\boldsymbol{\theta}}$ must be unbiased. The only necessary assumption is that both sides of (3) must exist; see e.g. Van Trees (1968).

The idea of the CR bound for random parameters $\boldsymbol{\theta}$ can be applied to the state estimation problem for a nonlinear dynamic system (Kerr, 1989; Tichavský et al., 1998; Doerschuk, 1995).

Consider the discrete-time nonlinear stochastic dynamic system

$$\mathbf{x}_{k+1} = \boldsymbol{\varphi}_k(\mathbf{x}_k, \mathbf{w}_k), \quad k = 0, 1, 2, \dots, \quad (4)$$

$$\mathbf{z}_k = \boldsymbol{\gamma}_k(\mathbf{x}_k, \mathbf{v}_k), \quad k = 0, 1, 2, \dots, \quad (5)$$

where k is a time index, \mathbf{x}_k and \mathbf{z}_k with $\dim(\mathbf{x}_k) = n$ and $\dim(\mathbf{z}_k) = r$ represent the state and measurement vectors, respectively, $\boldsymbol{\varphi}_k(\mathbf{x}_k, \mathbf{w}_k)$ and $\boldsymbol{\gamma}_k(\mathbf{x}_k, \mathbf{v}_k)$ are known vector functions, $\{\mathbf{w}_k\}$ and $\{\mathbf{v}_k\}$ are mutually independent white sequences, with $\dim(\mathbf{w}_k) = n$ and $\dim(\mathbf{v}_k) = r$, which are described by known pdf's $p(\mathbf{w}_k)$ and $p(\mathbf{v}_k)$, respectively. The noises are independent of the initial state \mathbf{x}_0 which is described by the known pdf $p(\mathbf{x}_0)$.

Suppose that the state transition pdf $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ exists and is twice differentiable with respect to both its

arguments. Similarly, suppose that the measurement pdf $p(\mathbf{z}_k|\mathbf{x}_k)$ exists and is twice differentiable with respect to \mathbf{x}_k .

The aim is to find a recursive algorithm for lower bounds of mean-square error matrices of three different types of state estimates. The MSEM of a state estimate $\hat{\mathbf{x}}_{\ell|k}$ is defined as

$$\mathbf{\Pi}_{\ell|k} = E\{(\mathbf{x}_{\ell} - \hat{\mathbf{x}}_{\ell|k})(\mathbf{x}_{\ell} - \hat{\mathbf{x}}_{\ell|k})^T\}, \quad (6)$$

where $\hat{\mathbf{x}}_{\ell|k}$ is an estimate of the state \mathbf{x}_{ℓ} using the measurements $\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_k$. The estimate $\hat{\mathbf{x}}_{\ell|k}$ is called *filtering* state estimate for $\ell = k$, *predictive* state estimate for $\ell > k$, and *smoothing* state estimate for $\ell < k$.

Since the assumption of two-fold differentiability of the transition pdf $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is not fulfilled for an important practical case of systems with unknown parameters, the second main goal of the paper is to solve this case as well.

3. Filtering Cramér–Rao bound

A derivation of the CR bound for nonlinear filtering problem was presented by Tichavský et al. (1998) and Šimandl et al. (1999). The results have the form of recursive relations for the CR bound setting a lower limit for the MSEM of a state estimate at each time. This section is based mainly on Šimandl et al. (1999) and its aim is to show the derivation of recursive relations for CR bounds of both filtering and one-step predictive MSEMs. The derivation process and its results will then serve as a starting point for solving the multi-step prediction and smoothing problems in Sections 4 and 5, respectively.

A significant attention is paid to a special case of a nonlinear stochastic system with unknown parameters. In design of estimation algorithms, the unknown system parameters can be treated as a part of the system state. However, a CR bound for such extended system state cannot be derived directly and has to be treated as a special singular case. The recursive formulae for filtering in a system with unknown constant parameters was derived as a special case of a more general result in Tichavský et al. (1998). Section 3.3 presents an alternative derivation of the CR bound in this case, and its extension to prediction and smoothing is given in Sections 4.2 and 5.2, respectively.

3.1. Fisher information matrix for state history

The main idea of the derivation of the CR bounds for nonlinear state estimation is to regard the whole state history as an unknown vector quantity. Let the complete state and measurement histories up to time instant k be denoted as $\mathbf{x}^k = [\mathbf{x}_0^T \mathbf{x}_1^T \dots \mathbf{x}_k^T]^T$ and $\mathbf{z}^k = [\mathbf{z}_0^T \mathbf{z}_1^T \dots \mathbf{z}_k^T]^T$, respectively. Then the state history \mathbf{x}^k may be interpreted as a vector of parameters of a random measured vector \mathbf{z}^k . First, a lower bound for a

MSEM of the whole state history will be derived and the obtained results are going to be used for derivation of bounds for filtering and one-step predictive estimates at time k .

First, the filtering estimate is considered. The joint pdf of the state and measurement histories $p(\mathbf{x}^k, \mathbf{z}^k)$ may be written as $p(\mathbf{x}^k, \mathbf{z}^k) = p(\mathbf{z}^k|\mathbf{x}^k)p(\mathbf{x}^k)$. Respecting the properties of the stochastic system (4) and (5), the logarithm of this pdf can be expressed as

$$\begin{aligned} \ln p(\mathbf{x}^k, \mathbf{z}^k) &= \sum_{i=0}^k \ln p(\mathbf{z}_i|\mathbf{x}_i) + \ln p(\mathbf{x}_0) \\ &\quad + \sum_{i=1}^k \ln p(\mathbf{x}_i|\mathbf{x}_{i-1}). \end{aligned} \quad (7)$$

The $(k+1)n \times (k+1)n$ FIM for the state history \mathbf{x}^k can now be computed according to (1)

$$\mathbf{J}_{k|k}(\mathbf{x}^k) = -E\{\nabla_{\mathbf{x}^k}[\nabla_{\mathbf{x}^k} \ln p(\mathbf{x}^k, \mathbf{z}^k)]^T\} \quad (8)$$

provided that the expectation and derivatives exist.

To simplify the computation of $\mathbf{J}_{k|k}(\mathbf{x}^k)$ the following notation for $n \times n$ matrices is introduced

$$\mathbf{K}_{i+1}^i = E\{-\nabla_{\mathbf{x}_i}[\nabla_{\mathbf{x}_i} \ln p(\mathbf{x}_{i+1}|\mathbf{x}_i)]^T\}, \quad (9)$$

$$\mathbf{K}_{i+1}^{i,i+1} = E\{-\nabla_{\mathbf{x}_{i+1}}[\nabla_{\mathbf{x}_i} \ln p(\mathbf{x}_{i+1}|\mathbf{x}_i)]^T\}, \quad (10)$$

$$\mathbf{K}_{i+1}^{i+1} = E\{-\nabla_{\mathbf{x}_{i+1}}[\nabla_{\mathbf{x}_{i+1}} \ln p(\mathbf{x}_{i+1}|\mathbf{x}_i)]^T\}, \quad (11)$$

$$\mathbf{L}_i^i = E\{-\nabla_{\mathbf{x}_i}[\nabla_{\mathbf{x}_i} \ln p(\mathbf{z}_i|\mathbf{x}_i)]^T\}, \quad (12)$$

with $\mathbf{K}_{i+1}^{i+1,i} = [\mathbf{K}_{i+1}^{i,i+1}]^T$, $i = 0, 1, \dots, k$,

and

$$\mathbf{K}_0^0 = E\{-\nabla_{\mathbf{x}_0}[\nabla_{\mathbf{x}_0} \ln p(\mathbf{x}_0)]^T\}. \quad (13)$$

The indexes in the \mathbf{K} matrices have the following meaning. The lower index is the time instant of the state described by the transition pdf. The upper index expresses the states for which the derivatives of the transition pdf are performed.

Note that if no information about the initial state is available, which can be mathematically expressed by a pdf with an infinitely small information content, i.e. with an infinite covariance matrix of \mathbf{x}_0 , then $\mathbf{K}_0^0 = 0$. Hence, such situation does not restrict finding the CR bound.

Using (7) and (9)–(12) in (8), it follows that the FIM for filtering is a block-tridiagonal matrix, having

$$\mathbf{D}_i = \mathbf{L}_i^i + \mathbf{K}_i^i + \mathbf{K}_{i+1}^i, \quad i = 0, 1, \dots, k-1, \quad (14)$$

and $\mathbf{L}_k^k + \mathbf{K}_k^k$ on its main diagonal, and $\mathbf{K}_{i+1}^{i+1,i}$ and $[\mathbf{K}_{i+1}^{i+1,i}]^T$ for $i = 0, 1, \dots, k-1$ on its lower and upper side diagonals, respectively. Let $\mathbf{J}_{k|k}(\mathbf{x}^k)$ be decomposed

into four blocks as

$$\mathbf{J}_{k|k}(\mathbf{x}^k) = \begin{bmatrix} \mathbf{D}_0 & \mathbf{K}_1^{01} & & & \\ \mathbf{K}_1^{10} & \ddots & \ddots & & \\ & \ddots & \mathbf{D}_{k-1} & & \\ & & & \mathbf{K}_k^{k-1,k} & \\ \hline & & & \mathbf{K}_k^{k,k-1} & \mathbf{L}_k^k + \mathbf{K}_k^k \end{bmatrix} = \begin{bmatrix} \mathbf{J}_{k|k}^{11} & \mathbf{J}_{k|k}^{12} \\ \mathbf{J}_{k|k}^{21} & \mathbf{J}_{k|k}^{22} \end{bmatrix}, \quad (15)$$

where zero block elements have been left empty. For $k = 0$ it holds $\mathbf{J}_{0|0}(\mathbf{x}^0) = \mathbf{J}_{0|0}^{22} = \mathbf{L}_0^0 + \mathbf{K}_0^0$.

The FIM $\mathbf{J}_{k|k-1}(\mathbf{x}^k)$ for the pdf $p(\mathbf{x}^k, \mathbf{z}^{k-1})$, with $p(\mathbf{x}^0, \mathbf{z}^{-1}) = p(\mathbf{x}_0)$, can be derived analogically to the previous procedure. Since it holds that

$$p(\mathbf{x}^k, \mathbf{z}^k) = p(\mathbf{x}^k, \mathbf{z}^{k-1})p(\mathbf{z}_k|\mathbf{x}^k, \mathbf{z}^{k-1}) = p(\mathbf{x}^k, \mathbf{z}^{k-1})p(\mathbf{z}_k|\mathbf{x}_k), \quad (16)$$

it can be easily seen that $\mathbf{J}_{k|k-1}(\mathbf{x}^k)$ is equal to $\mathbf{J}_{k|k}(\mathbf{x}^k)$, except for the lower-right corner block, which is \mathbf{K}_k^k instead of $\mathbf{L}_k^k + \mathbf{K}_k^k$. Thus it holds $\mathbf{J}_{k|k-1}^{11} = \mathbf{J}_{k|k}^{11}$, $\mathbf{J}_{k|k-1}^{12} = \mathbf{J}_{k|k}^{12}$, and $\mathbf{J}_{k|k-1}^{22} = \mathbf{K}_k^k$, and the FIMs $\mathbf{J}_{k|k}(\mathbf{x}^k)$, $\mathbf{J}_{k+1|k}(\mathbf{x}^{k+1})$ may be expressed recursively as

$$\mathbf{J}_{k|k}(\mathbf{x}^k) = \begin{bmatrix} \mathbf{J}_{k|k-1}^{11} & \mathbf{J}_{k|k-1}^{12} \\ \mathbf{J}_{k|k-1}^{21} & \mathbf{J}_{k|k-1}^{22} + \mathbf{L}_k^k \end{bmatrix} \quad (17)$$

$$\mathbf{J}_{k+1|k}(\mathbf{x}^{k+1}) = \begin{bmatrix} \mathbf{J}_{k|k}^{11} & \mathbf{J}_{k|k}^{12} & 0 \\ \mathbf{J}_{k|k}^{21} & \mathbf{J}_{k|k}^{22} + \mathbf{K}_{k+1}^k & \mathbf{K}_{k+1}^{k,k+1} \\ \hline 0 & \mathbf{K}_{k+1}^{k+1,k} & \mathbf{K}_{k+1}^{k+1} \end{bmatrix} \quad (18)$$

starting from $\mathbf{J}_{0|-1}(\mathbf{x}_0) = \mathbf{J}_{0|-1}^{22} = \mathbf{K}_0^0$. Note that the dimensions of the FIMs $\mathbf{J}_{k|k}(\mathbf{x}^k)$ and $\mathbf{J}_{k+1|k}(\mathbf{x}^{k+1})$ increase at each iteration and therefore these matrices are not acceptable as a final result.

3.2. Fisher information matrix for filtering estimate

In this section, recursive relations for FIMs $\mathbf{J}_{k|k}(\mathbf{x}_k)$ and $\mathbf{J}_{k|k-1}(\mathbf{x}_k)$ belonging to the state \mathbf{x}_k are derived. Let

$$\hat{\mathbf{x}}^{\ell|k} = [\hat{\mathbf{x}}_{0|k}^T \quad \hat{\mathbf{x}}_{1|k}^T \quad \dots \quad \hat{\mathbf{x}}_{\ell|k}^T]^T \quad (19)$$

be an estimate of the state trajectory up to time ℓ given measurements up to time k . Let the state trajectory estimate error and the state estimate error be denoted as $\tilde{\mathbf{x}}^{\ell|k} = \mathbf{x}^\ell - \hat{\mathbf{x}}^{\ell|k}$, $\tilde{\mathbf{x}}_{\ell|k} = \mathbf{x}_\ell - \hat{\mathbf{x}}_{\ell|k}$, respectively. For $\ell = k$, the MSEM of the estimate trajectory $\hat{\mathbf{x}}^{k|k}$ will be denoted $\mathbf{\Pi}^{k|k}$ and by (3) it is bounded as

$$\mathbf{\Pi}^{k|k} = E\{\tilde{\mathbf{x}}^{k|k}[\tilde{\mathbf{x}}^{k|k}]^T\} \geq \mathbf{J}_{k|k}^{-1}(\mathbf{x}^k). \quad (20)$$

Applying (17) to (20) yields

$$\mathbf{\Pi}^{k|k} = E \begin{bmatrix} \tilde{\mathbf{x}}^{k-1|k}[\tilde{\mathbf{x}}^{k-1|k}]^T & \tilde{\mathbf{x}}^{k-1|k}\tilde{\mathbf{x}}_{k|k}^T \\ \tilde{\mathbf{x}}_{k|k}[\tilde{\mathbf{x}}^{k-1|k}]^T & \tilde{\mathbf{x}}_{k|k}\tilde{\mathbf{x}}_{k|k}^T \end{bmatrix} \geq \begin{bmatrix} \mathbf{J}_{k|k-1}^{11} & \mathbf{J}_{k|k-1}^{12} \\ \mathbf{J}_{k|k-1}^{21} & \mathbf{L}_k^k + \mathbf{K}_k^k \end{bmatrix}^{-1}. \quad (21)$$

By comparing lower-right blocks of the matrices on both sides of the inequality it is possible to formulate an inequality for the MSEM of a filtering state estimate at time k :

$$\mathbf{\Pi}_{k|k} = E\{\tilde{\mathbf{x}}_{k|k}\tilde{\mathbf{x}}_{k|k}^T\} \geq \mathbf{C}_{k|k} = [\mathbf{J}_{k|k}^{-1}(\mathbf{x}^k)]_{22}, \quad (22)$$

where $\mathbf{C}_{k|k} = \mathbf{J}_{k|k}^{-1}(\mathbf{x}^k)$ is the CR bound and $\mathbf{J}_{k|k}(\mathbf{x}^k)$ is the FIM for an estimate $\hat{\mathbf{x}}_{k|k}$, and $[\mathbf{J}_{k|k}^{-1}(\mathbf{x}^k)]_{22}$ denotes the $n \times n$ lower-right block of the inverse of the matrix $\mathbf{J}_{k|k}(\mathbf{x}^k)$.

The FIM $\mathbf{J}_{k|k}(\mathbf{x}_k) = \mathbf{C}_{k|k}^{-1}$, which will be called the filtering FIM, can be obtained from $\mathbf{J}_{k|k}^{-1}(\mathbf{x}^k)$ taken in the form on the right-hand side of the inequality (21) using the block-matrix inverse (A.2) and (A.3) (see Appendix A). Hence

$$\mathbf{C}_{k|k}^{-1} = \mathbf{L}_k^k + \mathbf{K}_k^k - \mathbf{J}_{k|k-1}^{21}[\mathbf{J}_{k|k-1}^{11}]^{-1}\mathbf{J}_{k|k-1}^{12}. \quad (23)$$

The same routine can be followed for a one-step predictive estimate $\hat{\mathbf{x}}_{k+1|k}$ as well. The MSEM $\mathbf{\Pi}^{k+1|k} = E\{\tilde{\mathbf{x}}^{k+1|k}[\tilde{\mathbf{x}}^{k+1|k}]^T\}$ is bounded below by $\mathbf{J}_{k+1|k}^{-1}(\mathbf{x}^{k+1})$ and from (3) it holds

$$\mathbf{\Pi}^{k+1|k} = E \begin{bmatrix} \tilde{\mathbf{x}}^{k|k}[\tilde{\mathbf{x}}^{k|k}]^T & \tilde{\mathbf{x}}^{k|k}\tilde{\mathbf{x}}_{k+1|k}^T \\ \tilde{\mathbf{x}}_{k+1|k}[\tilde{\mathbf{x}}^{k|k}]^T & \tilde{\mathbf{x}}_{k+1|k}\tilde{\mathbf{x}}_{k+1|k}^T \end{bmatrix} \geq \begin{bmatrix} \mathbf{J}_{k+1|k}^{11} & \mathbf{J}_{k+1|k}^{12} \\ \mathbf{J}_{k+1|k}^{21} & \mathbf{K}_{k+1}^k \end{bmatrix}^{-1} = \mathbf{J}_{k+1|k}^{-1}(\mathbf{x}^{k+1}). \quad (24)$$

For the MSEM of the predictive estimate $\hat{\mathbf{x}}_{k+1|k}$ the following inequality, analogous to (22), is obtained:

$$\mathbf{\Pi}_{k+1|k} = E\{\tilde{\mathbf{x}}_{k+1|k}\tilde{\mathbf{x}}_{k+1|k}^T\} \geq \mathbf{C}_{k+1|k}, \quad (25)$$

where $\mathbf{C}_{k+1|k} = \mathbf{J}_{k+1|k}^{-1}(\mathbf{x}_{k+1})$ is the CR bound for a one-step predictive estimate $\hat{\mathbf{x}}_{k+1|k}$, and $\mathbf{J}_{k+1|k}(\mathbf{x}_{k+1})$ is the FIM for $\hat{\mathbf{x}}_{k+1|k}$, equal to the lower-right block of an inverse of $\mathbf{J}_{k+1|k}(\mathbf{x}^{k+1})$.

The FIM $\mathbf{J}_{k+1|k}(\mathbf{x}_{k+1})$, i.e. the inverse of the CR bound $\mathbf{C}_{k+1|k}$, can be expressed as

$$\mathbf{C}_{k+1|k}^{-1} = \mathbf{K}_{k+1}^{k+1} - \mathbf{J}_{k+1|k}^{21}[\mathbf{J}_{k+1|k}^{11}]^{-1}\mathbf{J}_{k+1|k}^{12}. \quad (26)$$

Using (18), (A.2), (A.3), (15), and (23), the following recursion for the one-step predictive FIM is obtained:

$$\mathbf{C}_{k+1|k}^{-1} = \mathbf{K}_{k+1}^{k+1} - \mathbf{K}_{k+1}^{k+1,k}(\mathbf{K}_{k+1}^k + \mathbf{C}_{k|k}^{-1})^{-1}\mathbf{K}_{k+1}^{k,k+1}. \quad (27)$$

Finally, relation (23) for the filtering FIM $\mathbf{C}_{k|k}^{-1}$ will be treated using (26), then

$$\mathbf{C}_{k|k}^{-1} = \mathbf{C}_{k|k-1}^{-1} + \mathbf{L}_k^k. \quad (28)$$

Relations (28) and (27) describe a recursive computation of the CR bounds for filtering and one-step predictive state estimates. Initial conditions for this recursion are given by $\mathbf{C}_{0| -1}^{-1} = \mathbf{K}_0^0$. Equation (27) represents transformation of the information matrix of the system state from time k to time $k + 1$, without any new information. The matrix \mathbf{L}_k^k in (28) represents the new independent information about the system state contained in the observation \mathbf{z}_k .

Note that a generally different CR bound for the state \mathbf{x}_k could be theoretically obtained from the marginal pdf $p(\mathbf{x}_k, \mathbf{z}^k)$ derived from $p(\mathbf{x}^k, \mathbf{z}^k)$. It follows from Proposition 1 in Bobrovsky et al. (1987) that such a bound would be tighter, in general. However, it is extremely difficult to compute the marginal pdf and consequently the whole latter bound.

3.3. CR bound for filtering with unknown parameters

Consider the following extension of the nonlinear stochastic system in (4), (5):

$$\mathbf{x}_{k+1} = \boldsymbol{\varphi}_k(\mathbf{x}_k, \mathbf{a}, \mathbf{w}_k), \quad k = 0, 1, 2, \dots, \quad (29)$$

$$\mathbf{z}_k = \boldsymbol{\gamma}_k(\mathbf{x}_k, \mathbf{a}, \mathbf{v}_k), \quad k = 0, 1, 2, \dots, \quad (30)$$

where \mathbf{a} is a random vector parameter with $\dim(\mathbf{a}) = m$, independent of the initial state \mathbf{x}_0 , $p(\mathbf{x}_0) = p(\mathbf{x}_0|\mathbf{a})$, and with a known, twice differentiable pdf $p(\mathbf{a})$. The task is to compute the filtering CR bound for the pair $(\mathbf{x}_k, \mathbf{a})$. A valid CR bound can be computed also in a case when no prior information (no $p(\mathbf{a})$) is available for the same reason as in case of \mathbf{x}_0 ; see the note after (13).

The nonlinear stochastic system in (29) and (30) can be treated in the same way as the system in (4) and (5), if the task is to construct a nonlinear filter for the system. It is sufficient to consider an extended system state vector $\boldsymbol{\xi}_k = [\mathbf{x}_k^T \ \mathbf{a}_k^T]^T$ and let $\mathbf{a}_{k+1} = \mathbf{a}_k$ for all k with $\mathbf{a}_0 = \mathbf{a}$. However, this approach fails if one wants to compute the corresponding CR bound. Recall that computation of the CR bound assumes existence of the transition pdf and its second-order differentiability. However, these conditions are not fulfilled for the transition pdf of the extended state $p(\boldsymbol{\xi}_{k+1}|\boldsymbol{\xi}_k)$. The reason is that the parameters \mathbf{a}_k are not affected by any disturbances. Thus the transition pdf $p(\mathbf{a}_{k+1}|\mathbf{a}_k)$ is the Dirac function and the pdf of the extended state is not differentiable with respect to \mathbf{a}_k . In this sense, computation of the CR bound for the stochastic system (29) and (30) is a nontrivial generalization of the recursions (27) and (28).

First, a recursion for the filtering CR bound $\mathbf{C}_{k|k}$ for $(\mathbf{x}_k, \mathbf{a})$ will be derived. Let the inverse of this matrix be decomposed in blocks as

$$\mathbf{C}_{k|k}^{-1} = \begin{bmatrix} \mathbf{J}_{k|k}^{xx} & \mathbf{J}_{k|k}^{xa} \\ \mathbf{J}_{k|k}^{ax} & \mathbf{J}_{k|k}^{aa} \end{bmatrix}. \quad (31)$$

Again, the derivation will start with the expression for the logarithm of the pdf of the joint state and measurement histories,

$$\begin{aligned} \ln p(\mathbf{x}^k, \mathbf{a}, \mathbf{z}^k) &= \sum_{i=0}^k \ln p(\mathbf{z}_i|\mathbf{x}_i, \mathbf{a}) + \ln p(\mathbf{x}_0) + \ln p(\mathbf{a}) \\ &\quad + \sum_{i=1}^k \ln p(\mathbf{x}_i|\mathbf{x}_{i-1}, \mathbf{a}). \end{aligned} \quad (32)$$

Let matrices \mathbf{K}_{i+1}^i , \mathbf{K}_{i+1}^{i+1} , \mathbf{K}_{i+1}^{i+1} , and \mathbf{L}_i^i be defined as in (9)–(12) with the exception that $p(\mathbf{x}_{i+1}|\mathbf{x}_i)$ and $p(\mathbf{z}_i|\mathbf{x}_i)$ are replaced by $p(\mathbf{x}_{i+1}|\mathbf{x}_i, \mathbf{a})$ and $p(\mathbf{z}_i|\mathbf{x}_i, \mathbf{a})$, respectively. Let the additional notations analogous to (9)–(12) be introduced.

$$\mathbf{K}_{i+1}^{ai} = E\{-\nabla_{\mathbf{x}_i}[\nabla_{\mathbf{a}} \ln p(\mathbf{x}_{i+1}|\mathbf{x}_i, \mathbf{a})]^T\}, \quad (33)$$

$$\mathbf{K}_{i+1}^{a,i+1} = E\{-\nabla_{\mathbf{x}_{i+1}}[\nabla_{\mathbf{a}} \ln p(\mathbf{x}_{i+1}|\mathbf{x}_i, \mathbf{a})]^T\}, \quad (34)$$

$$\mathbf{K}_{i+1}^a = E\{-\nabla_{\mathbf{a}}[\nabla_{\mathbf{a}} \ln p(\mathbf{x}_{i+1}|\mathbf{x}_i, \mathbf{a})]^T\}, \quad (35)$$

$$\mathbf{L}_i^{ai} = E\{-\nabla_{\mathbf{x}_i}[\nabla_{\mathbf{a}} \ln p(\mathbf{z}_i|\mathbf{x}_i, \mathbf{a})]^T\}, \quad (36)$$

$$\mathbf{L}_i^a = E\{-\nabla_{\mathbf{a}}[\nabla_{\mathbf{a}} \ln p(\mathbf{z}_i|\mathbf{x}_i, \mathbf{a})]^T\}, \quad (37)$$

$$\mathbf{A}_0 = E\{-\nabla_{\mathbf{a}}[\nabla_{\mathbf{a}} \ln p(\mathbf{a})]^T\}, \quad (38)$$

$$\mathbf{A}_k = \mathbf{A}_0 + \mathbf{L}_0^a + \sum_{i=1}^{k-1} (\mathbf{L}_i^a + \mathbf{K}_i^a) + \mathbf{K}_k^a, \quad (39)$$

$$\mathbf{G}_i = \mathbf{L}_i^{ai} + \mathbf{K}_i^{ai} + \mathbf{K}_{i+1}^{ai}, \quad (40)$$

where $i = 0, 1, \dots, k$, and $\mathbf{L}_i^{ia} = [\mathbf{L}_i^{ai}]^T$, $\mathbf{K}_i^{ia} = [\mathbf{K}_i^{ai}]^T$, $\mathbf{K}_{i+1}^{ia} = [\mathbf{K}_{i+1}^{ai}]^T$. The matrices \mathbf{K}_i^{ai} , \mathbf{K}_{i+1}^{ai} , \mathbf{L}_i^{ai} , and \mathbf{G}_i have the size $m \times n$, \mathbf{K}_i^{ia} , \mathbf{L}_i^a , and \mathbf{A}_k are $m \times m$ matrices, and \mathbf{K}_0^{a0} appearing in (40) for $i = 0$ is an $m \times n$ zero matrix.

Now, the FIM for the pair $(\mathbf{x}^k, \mathbf{a})$, defined as

$$\mathbf{J}_{k|k}(\mathbf{x}^k, \mathbf{a}) = -E\{\nabla_{\mathbf{x}^k, \mathbf{a}}[\nabla_{\mathbf{x}^k, \mathbf{a}} \ln p(\mathbf{x}^k, \mathbf{a}, \mathbf{z}^k)]^T\}, \quad (41)$$

provided the expectation and the derivatives exist, can be written as

$$\mathbf{J}_{k|k}(\mathbf{x}^k, \mathbf{a}) = \begin{bmatrix} \mathbf{J}_{k|k}^{11} & \mathbf{J}_{k|k}^{12} & [\mathbf{G}^{k-1}]^T \\ \mathbf{J}_{k|k}^{21} & \mathbf{J}_{k|k}^{22} & \mathbf{L}_k^{ka} + \mathbf{K}_k^{ka} \\ \mathbf{G}^{k-1} & \mathbf{L}_k^{ak} + \mathbf{K}_k^{ak} & \mathbf{L}_k^a + \mathbf{A}_k \end{bmatrix} \quad (42)$$

where $\mathbf{G}^i = [\mathbf{G}_0 \ \mathbf{G}_1 \ \dots \ \mathbf{G}_i]$.

The filtering CR bound $\mathbf{C}_{k|k}$ for $(\mathbf{x}_k, \mathbf{a})$ is found as the lower-right $(n+m) \times (n+m)$ block of $\mathbf{J}_{k|k}^{-1}(\mathbf{x}^k, \mathbf{a})$. The detailed derivation is shown in Appendix B. The block elements of $\mathbf{C}_{k|k}^{-1}$, as denoted in (31), obey the following recursive relations

$$\mathbf{J}_{k|k}^{xx} = \mathbf{J}_{k|k-1}^{xx} + \mathbf{L}_k^k, \quad (43)$$

$$\mathbf{J}_{k|k}^{xa} = \mathbf{J}_{k|k-1}^{xa} + \mathbf{L}_k^{ka}, \quad (44)$$

$$\mathbf{J}_{k|k}^{aa} = \mathbf{J}_{k|k-1}^{aa} + \mathbf{L}_k^a, \quad (45)$$

where the one-step predictive FIMs are given as

$$\mathbf{J}_{k+1|k}^{xx} = \mathbf{K}_{k+1}^{k+1} - \mathbf{K}_{k+1}^{k+1,k} \Delta_{kk}^{-1} \mathbf{K}_{k+1}^{k,k+1}, \quad (46)$$

$$\mathbf{J}_{k+1|k}^{xa} = \mathbf{K}_{k+1}^{k+1,a} - \mathbf{K}_{k+1}^{k+1,k} \Delta_{kk}^{-1} \Delta_{kk}^{xa}, \quad (47)$$

$$\mathbf{J}_{k+1|k}^{aa} = \mathbf{J}_{k|k}^{aa} + \mathbf{K}_{k+1}^a - \Delta_{kk}^{ax} \Delta_{kk}^{-1} \Delta_{kk}^{xa}, \quad (48)$$

where $\Delta_{\ell k} = \mathbf{J}_{\ell|k}^{xx} + \mathbf{K}_{\ell+1}^{\ell}$, $\Delta_{\ell k}^{ax} = \mathbf{J}_{\ell|k}^{ax} + \mathbf{K}_{\ell+1}^{a\ell}$ and $\Delta_{\ell k}^{xa} = [\Delta_{\ell k}^{ax}]^T$. In (46)–(48) it holds that $\ell = k$.

The initial conditions for the recursive relations (43)–(48) are $\mathbf{J}_{0|-1}^{xx} = \mathbf{K}_0^0$, $\mathbf{J}_{0|-1}^{xa} = \mathbf{K}_0^{0a} = 0$, and $\mathbf{J}_{0|-1}^{aa} = \mathbf{A}_0$. The relations describe a recursive block-wise computation of the CR bound for filtering with unknown parameters. Note that (43) and (46) are exactly as (28) and (27), respectively.

4. Predictive Cramér–Rao bound

4.1. Derivation of multi-step predictive lower bound

The recursions for filtering and one-step predictive CR bounds derived in Section 3.2 can be easily generalized for the general prediction problem (Šimandl et al., 1999). In particular, it is possible to derive the CR bound for a multi-step predictive estimate $\hat{\mathbf{x}}_{\ell|k}$ with $\ell > k$, using the joint pdf $p(\mathbf{x}^{\ell}, \mathbf{z}^k)$, in the similar way as for the filtering and one-step prediction problems. First, the derivation of the two-step predictive FIM, i.e. $\ell = k + 2$, will be shown and then this case will be generalized for a multi-step prediction ($\ell \geq k + 2$) by the mathematical induction.

The logarithm of the joint pdf $p(\mathbf{x}^{k+2}, \mathbf{z}^k)$ can be expressed as

$$\ln p(\mathbf{x}^{k+2}, \mathbf{z}^k) = \sum_{i=0}^k \ln p(\mathbf{z}_i | \mathbf{x}_i) + \ln p(\mathbf{x}_0) + \sum_{i=1}^{k+2} \ln p(\mathbf{x}_i | \mathbf{x}_{i-1}) \quad (49)$$

and it could be rewritten using the one-step predictive joint pdf

$$\ln p(\mathbf{x}^{k+2}, \mathbf{z}^k) = \ln p(\mathbf{x}^{k+1}, \mathbf{z}^k) + \ln p(\mathbf{x}_{k+2} | \mathbf{x}_{k+1}).$$

Thus, the two-step predictive FIM represents an expansion of the one-step predictive FIM, which is known from (18), by three additional blocks,

$$\mathbf{J}_{k+2|k}(\mathbf{x}^{k+2}) = \left[\begin{array}{cc|c} \mathbf{J}_{k+1|k}^{11} & \mathbf{J}_{k+1|k}^{12} & 0 \\ \mathbf{J}_{k+1|k}^{21} & \mathbf{J}_{k+1|k}^{22} + \mathbf{K}_{k+2}^{k+1} & \mathbf{K}_{k+2}^{k+1,k+2} \\ \hline 0 & \mathbf{K}_{k+2}^{k+2,k+1} & \mathbf{K}_{k+2}^{k+2} \end{array} \right]. \quad (50)$$

Obviously, it is possible to follow the same steps which lead from (18) to the relation for the one-step predictive FIM (27). Then

$$\mathbf{C}_{k+2|k}^{-1} = \mathbf{K}_{k+2}^{k+2} - \mathbf{K}_{k+2}^{k+2,k+1} (\mathbf{K}_{k+2}^{k+1} + \mathbf{C}_{k+1|k}^{-1})^{-1} \mathbf{K}_{k+2}^{k+1,k+2}. \quad (51)$$

The derivation of the two-step predictive FIM can be inductively applied to a general $(\ell - k)$ -step prediction,

yielding the following recursion

$$\mathbf{C}_{\ell|k}^{-1} = \mathbf{K}_{\ell}^{\ell} - \mathbf{K}_{\ell}^{\ell,\ell-1} (\mathbf{K}_{\ell}^{\ell-1} + \mathbf{C}_{\ell-1|k}^{-1})^{-1} \mathbf{K}_{\ell}^{\ell-1,\ell} \quad (52)$$

for $\ell = k + 1, k + 2, \dots$. The recursion starts from the filtering FIM $\mathbf{C}_{k|k}^{-1}$ and the \mathbf{K} matrices are given by (9)–(11) without any modifications.

The relation (52) represents the final result of the multi-step predictive CR bound problem.

4.2. Predictive CR bound for system with unknown parameters

It is easy to see that predictive CR bound for systems (29) and (30) with unknown parameters obeys the same recursions as the one-step predictive CR bound derived in Section 3.3. Let the predictive FIM $\mathbf{C}_{\ell|k}^{-1}$ be composed of four blocks $\mathbf{J}_{\ell|k}^{xx}$, $\mathbf{J}_{\ell|k}^{xa}$, $\mathbf{J}_{\ell|k}^{ax}$, and $\mathbf{J}_{\ell|k}^{aa}$ in the same way as in (31). Then, similarly to (46)–(48), it holds that

$$\mathbf{J}_{\ell|k}^{xx} = \mathbf{K}_{\ell}^{\ell} - \mathbf{K}_{\ell}^{\ell,\ell-1} \Delta_{\ell-1,k}^{-1} \mathbf{K}_{\ell}^{\ell-1,\ell}, \quad (53)$$

$$\mathbf{J}_{\ell|k}^{xa} = \mathbf{K}_{\ell}^{\ell a} - \mathbf{K}_{\ell}^{\ell,\ell-1} \Delta_{\ell-1,k}^{-1} \Delta_{\ell-1,k}^{xa}, \quad (54)$$

$$\mathbf{J}_{\ell|k}^{aa} = \mathbf{J}_{\ell-1|k}^{aa} + \mathbf{K}_{\ell}^a - \Delta_{\ell-1,k}^{ax} \Delta_{\ell-1,k}^{-1} \Delta_{\ell-1,k}^{xa} \quad (55)$$

for $\ell = k + 1, k + 2, \dots$.

5. Smoothing Cramér–Rao bound

5.1. Derivation of smoothing lower bound

Consider again the system (4), (5). The derivation of the smoothing CR bound for a smoothing state estimate $\hat{\mathbf{x}}_{\ell|k}$, $0 \leq \ell \leq k - 1$, is based on the FIM $\mathbf{J}_{k|k}(\mathbf{x}^k)$, introduced in (15), and on a suitable decomposition of this matrix (Bergman, 1999). Note that $\mathbf{J}_{k|k}(\mathbf{x}^k)$ is the lower bound of the MSEM $\mathbf{\Pi}^{k|k}$ of the estimate trajectory $\hat{\mathbf{x}}^{k|k}$ from (19) consisting of the smoothing estimates $\hat{\mathbf{x}}_{0|k}$, $\hat{\mathbf{x}}_{1|k}, \dots, \hat{\mathbf{x}}_{k-1|k}$, and the filtering estimate $\hat{\mathbf{x}}_{k|k}$. Thus, smoothing CR bounds should be implicitly contained in $\mathbf{J}_{k|k}(\mathbf{x}^k)$. The decomposition divides $\mathbf{J}_{k|k}(\mathbf{x}^k)$ into blocks which correspond to time instants $0, 1, \dots, \ell$, and $\ell + 1, \ell + 2, \dots, k$, respectively, as

$$\mathbf{J}_{k|k}(\mathbf{x}^k) = \left[\begin{array}{c|c} \Gamma_{\ell|\ell} & \mathbf{S}_{\ell,k} \\ \hline \mathbf{S}_{\ell,k}^T & \Psi_{\ell,k} \end{array} \right] = \left[\begin{array}{cc|ccc} \mathbf{D}_0 & \mathbf{K}_1^{01} & & & \\ \mathbf{K}_1^{10} & \ddots & \ddots & & \\ & \ddots & \mathbf{D}_{\ell} & \mathbf{K}_{\ell+1}^{\ell,\ell+1} & \\ \hline & & \mathbf{K}_{\ell+1}^{\ell+1,\ell} & \ddots & \ddots \\ & & & \ddots & \mathbf{D}_{k-1} & \mathbf{K}_k^{k-1,k} \\ & & & & \mathbf{K}_k^{k,k-1} & \mathbf{L}_k^k + \mathbf{K}_k^k \end{array} \right], \quad (56)$$

where zero blocks have been left empty, and the block $\Gamma_{\ell|k}$ can be also expressed as

$$\Gamma_{\ell|k} = \begin{bmatrix} \mathbf{J}_{\ell|k}^{11} & \mathbf{J}_{\ell|k}^{12} \\ \mathbf{J}_{\ell|k}^{21} & \mathbf{J}_{\ell|k}^{22} + \mathbf{K}_{\ell+1}^{\ell} \end{bmatrix}. \quad (57)$$

The matrix $\Gamma_{\ell|k}$ is almost identical with the FIM $\mathbf{J}_{\ell|k}(\mathbf{x}^{\ell})$, except for the matrix $\mathbf{K}_{\ell+1}^{\ell}$ in the lower-right block of $\Gamma_{\ell|k}$. Recall that the same structure appeared also in (18).

The inverse of $\mathbf{J}_{k|k}(\mathbf{x}^k)$ contains the smoothing CR bounds $\mathbf{C}_{\ell|k}$, $\ell = 0, 1, \dots, k-1$, and the filtering CR bound $\mathbf{C}_{k|k}$ on its main diagonal. Let the inverse $\mathbf{J}_{k|k}^{-1}(\mathbf{x}^k)$ be decomposed as follows

$$\mathbf{J}_{k|k}^{-1}(\mathbf{x}^k) = \begin{bmatrix} \mathbf{C}_{0|k} & & & \\ & \ddots & & \\ & & \mathbf{C}_{\ell|k} & \\ & & & \ddots \\ & & & & \mathbf{C}_{k|k} \end{bmatrix} = \begin{bmatrix} [\mathbf{J}_{k|k}^{-1}]_{11} & [\mathbf{J}_{k|k}^{-1}]_{12} \\ [\mathbf{J}_{k|k}^{-1}]_{21} & [\mathbf{J}_{k|k}^{-1}]_{22} \end{bmatrix}, \quad (58)$$

where only the diagonal blocks have been indicated. Since only the lower-right block $\mathbf{C}_{\ell|k}$ of $[\mathbf{J}_{k|k}^{-1}]_{11}$ is of interest, it will be extracted by

$$\mathbf{C}_{\ell|k} = [0 \quad \mathbf{I}_n][\mathbf{J}_{k|k}^{-1}]_{11}[0 \quad \mathbf{I}_n]^T, \quad (59)$$

where \mathbf{I}_n is an $n \times n$ identity matrix and 0 denotes a zero matrix of appropriate size.

To obtain $[\mathbf{J}_{k|k}^{-1}]_{11}$, the block-matrix inverse (A.2), (A.3) will be applied to (56). Furthermore suppose that all smoothing CR bounds $\mathbf{C}_{\ell+1|k}$, $\mathbf{C}_{\ell+2|k}$, \dots , $\mathbf{C}_{k-1|k}$, and the filtering bound $\mathbf{C}_{k|k}$ in the block $[\mathbf{J}_{k|k}^{-1}]_{22}$ (see (58)) are known. Then using (A.2) yields

$$[\mathbf{J}_{k|k}^{-1}]_{11} = \Gamma_{\ell|k}^{-1} + \Gamma_{\ell|k}^{-1} \mathbf{S}_{\ell,k} [\mathbf{J}_{k|k}^{-1}]_{22} \mathbf{S}_{\ell,k}^T \Gamma_{\ell|k}^{-1}. \quad (60)$$

Before the prescribed extraction (59) is performed for (60), the following intermediate computation should be pointed out

$$[0 \quad \mathbf{I}_n] \Gamma_{\ell|k}^{-1} [0 \quad \mathbf{I}_n]^T = (\mathbf{J}_{\ell|k}^{22} + \mathbf{K}_{\ell+1}^{\ell} - \mathbf{J}_{\ell|k}^{21} [\mathbf{J}_{\ell|k}^{11}]^{-1} \mathbf{J}_{\ell|k}^{12})^{-1} = (\mathbf{C}_{\ell|k}^{-1} + \mathbf{K}_{\ell+1}^{\ell})^{-1}, \quad (61)$$

where the inverse of $\Gamma_{\ell|k}$ in the form of (57) was realized by (A.2), (A.3), and the final expression in (61) was obtained utilizing the relation (23) for the filtering FIM.

Now it is possible to substitute (60) into (59). Using (61) and substituting for $\mathbf{S}_{\ell,k}$ from (56), the following relation for the smoothing CR bound is obtained:

$$\mathbf{C}_{\ell|k} = (\mathbf{C}_{\ell|k}^{-1} + \mathbf{K}_{\ell+1}^{\ell})^{-1} + (\mathbf{C}_{\ell|k}^{-1} + \mathbf{K}_{\ell+1}^{\ell})^{-1} \mathbf{K}_{\ell+1}^{\ell+1} \times \mathbf{C}_{\ell+1|k} \mathbf{K}_{\ell+1}^{\ell+1} (\mathbf{C}_{\ell|k}^{-1} + \mathbf{K}_{\ell+1}^{\ell})^{-1} \quad (62)$$

for $\ell = k-1, k-2, \dots, 0$. Note that $\mathbf{C}_{\ell|k}$ depends only on the previous smoothing CR bound $\mathbf{C}_{\ell+1|k}$ and on the filtering bound $\mathbf{C}_{\ell|k}$. Thus, the relation (62) explicitly describes a backward recursive computation of the smoothing CR bound $\mathbf{C}_{\ell|k}$, $\ell < k$, while both predictive and filtering bounds were described through their inverses. The inverse of the filtering CR bound $\mathbf{C}_{\ell|k}$ also acts in (62) and it can be taken directly from the forward recursions (27) and (28).

However, it might be useful to unify the relation for the smoothing CR bound with the structures of the predictive and filtering bounds. Thus, (62) will be further modified using the matrix inverse (A.1); (see Appendix A).

Obviously, the form of the left-hand side of (A.1) can be found on the right-hand side of (62). Hence the inverse of the smoothing CR bound, i.e. the smoothing FIM, obeys this recursive relation

$$\mathbf{C}_{\ell|k}^{-1} = \mathbf{C}_{\ell|k}^{-1} + \mathbf{K}_{\ell+1}^{\ell} - \mathbf{K}_{\ell+1}^{\ell+1} [\mathbf{K}_{\ell+1}^{\ell+1} (\mathbf{C}_{\ell|k}^{-1} + \mathbf{K}_{\ell+1}^{\ell})^{-1} \times \mathbf{K}_{\ell+1}^{\ell+1} + \mathbf{C}_{\ell+1|k}^{-1}]^{-1} \mathbf{K}_{\ell+1}^{\ell+1}. \quad (63)$$

Finally, compare the bracketed inverted expression on the right-hand side of (63) with (27). Employing the latter relation, the final relation for the smoothing CR bound is obtained

$$\mathbf{C}_{\ell|k}^{-1} = \mathbf{C}_{\ell|k}^{-1} + \mathbf{K}_{\ell+1}^{\ell} - \mathbf{K}_{\ell+1}^{\ell+1} \times (\mathbf{K}_{\ell+1}^{\ell+1} + \mathbf{C}_{\ell+1|k}^{-1} - \mathbf{C}_{\ell+1|k}^{-1})^{-1} \mathbf{K}_{\ell+1}^{\ell+1}, \quad (64)$$

for $\ell = k-1, k-2, \dots, 0$. The initial condition for this backward recursion is given by the filtering FIM $\mathbf{C}_{k|k}^{-1}$.

5.2. CR bound for smoothing with unknown parameters

The derivation of the smoothing CR bound for a smoothing state estimate $\hat{\mathbf{x}}_{\ell|k}$, $0 \leq \ell \leq k-1$, in the extended nonlinear system (29), (30) proceeds similarly as in the previous subsection.

Firstly, consider the FIM $\mathbf{J}_{k|k}(\mathbf{x}^k, \mathbf{a})$ from (42) in the form analogous to (56),

$$\mathbf{J}_{k|k}(\mathbf{x}^k, \mathbf{a}) = \begin{bmatrix} \Gamma_{\ell|k} & \tilde{\mathbf{S}}_{\ell,k} \\ \tilde{\mathbf{S}}_{\ell,k}^T & \tilde{\Psi}_{\ell,k} \end{bmatrix}. \quad (65)$$

In (65), $\Gamma_{\ell|k}$ was introduced in (57),

$$\tilde{\mathbf{S}}_{\ell,k} = [\mathbf{S}_{\ell,k} \quad [\mathbf{G}^{\ell}]^T] = \begin{bmatrix} 0 & [\mathbf{G}^{\ell-1}]^T \\ \tilde{\mathbf{K}}_{\ell} & \mathbf{G}_{\ell}^T \end{bmatrix}, \quad (66)$$

where $\tilde{\mathbf{K}}_{\ell} = [\mathbf{K}_{\ell+1}^{\ell+1} \ 0]$, and $\tilde{\Psi}_{\ell,k}$ is the lower-right $(n(k-\ell) + m) \times (n(k-\ell) + m)$ block of $\mathbf{J}_{k|k}(\mathbf{x}^k, \mathbf{a})$.

The smoothing CR bound for the pair $(\mathbf{x}_{\ell}, \mathbf{a})$ given the measurements up to time k can be decomposed as

$$\mathbf{C}_{\ell|k} = \begin{bmatrix} \mathbf{C}_{\ell|k}^{xx} & \mathbf{C}_{\ell|k}^{xa} \\ \mathbf{C}_{\ell|k}^{ax} & \mathbf{C}_{\ell|k}^{aa} \end{bmatrix}, \quad (67)$$

where, indeed, $\mathbf{C}_{0|k}^{aa} = \mathbf{C}_{1|k}^{aa} = \dots = \mathbf{C}_{k|k}^{aa}$.

The detailed derivation of recursive relations for the remaining blocks of $\mathbf{C}_{\ell|k}$ is shown in Appendix C and the final result is the following:

$$\begin{aligned} \mathbf{C}_{\ell|k}^{xx} &= \Delta_{\ell\ell}^{-1} + \Delta_{\ell\ell}^{-1}(\mathbf{K}_{\ell+1}^{\ell,\ell+1})^T \mathbf{C}_{\ell+1|k}^{xx} \mathbf{K}_{\ell+1}^{\ell,\ell+1} \\ &\quad + \Delta_{\ell\ell}^{xa} \mathbf{C}_{\ell+1|k}^{ax} \mathbf{K}_{\ell+1}^{\ell,\ell+1} + \mathbf{K}_{\ell+1}^{\ell,\ell+1} \mathbf{C}_{\ell+1|k}^{xa} \Delta_{\ell\ell}^{ax} \\ &\quad + \Delta_{\ell\ell}^{aa} \mathbf{C}_{\ell+1|k}^{aa} \Delta_{\ell\ell}^{-1}, \end{aligned} \quad (68)$$

$$\mathbf{C}_{\ell|k}^{xa} = -\Delta_{\ell\ell}^{-1}(\mathbf{K}_{\ell+1}^{\ell,\ell+1})^T \mathbf{C}_{\ell+1|k}^{xa} + \Delta_{\ell\ell}^{xa} \mathbf{C}_{\ell+1|k}^{aa}, \quad (69)$$

$$\mathbf{C}_{\ell|k}^{aa} = \mathbf{C}_{\ell+1|k}^{aa}, \quad (70)$$

where $\ell = k - 1, k - 2, \dots, 0$.

6. Computation of Cramér–Rao bounds

The previous sections were focused on theoretical derivations of CR bounds for all basic types of state estimation. Now, practical possibilities of computation of the CR bounds will be discussed for system (4) and (5), but analogous results can be obtained also for system (29) and (30) with unknown parameters.

Obviously, the expectations in (9)–(12) are crucial for complexity of the CR bounds. Computational complexity of these expectations depends on the structure of the system (4), (5) and on properties of the random variables \mathbf{x}_0 , \mathbf{w}_k , and \mathbf{v}_k . Hence, after a general discussion two significant special cases of the system will be analyzed separately: additive Gaussian noise in state and measurement equations, and linear Gaussian system.

6.1. General case

A necessary condition for calculation of the \mathbf{K} and \mathbf{L} matrices in (9)–(13) is to find the transition pdf's $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ and the measurement pdf's $p(\mathbf{z}_k|\mathbf{x}_k)$ for $k = 0, 1, 2, \dots$, from system (4) and (5), and the derivatives in (9)–(13). Of course, these relations cannot be used without a further specification of the system structure and the random variables.

As it is not generally possible to compute the expectations in (9)–(13) analytically, a simulation Monte Carlo method can be used for their estimation. With the estimated matrices, the filtering, predictive, and smoothing CR bounds can be generated by (28), (52), and (64).

6.2. Additive Gaussian noise

An important special case of general system description (4) and (5) is assuming additive Gaussian state and measurement noises as follows:

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{w}_k, \quad k = 0, 1, 2, \dots, \quad (71)$$

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \quad k = 0, 1, 2, \dots, \quad (72)$$

where $\mathbf{f}_k(\mathbf{x}_k)$ and $\mathbf{h}_k(\mathbf{x}_k)$ are known vector functions and the state and measurement noises \mathbf{w}_k , \mathbf{v}_k have zero means and positive definite covariance matrices \mathbf{Q}_k and \mathbf{R}_k , respectively, thus $p(\mathbf{w}_k) = \mathcal{N}(\mathbf{w}_k; 0, \mathbf{Q}_k)$ and $p(\mathbf{v}_k) = \mathcal{N}(\mathbf{v}_k; 0, \mathbf{R}_k)$. The initial state \mathbf{x}_0 is described as $p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0; \mathbf{m}_0, \mathbf{M}_0)$.

Using the assumptions of the additive Gaussian noises, the transition and measurement pdf's can be expressed as follows:

$$p(\mathbf{x}_{i+1}|\mathbf{x}_i) = \mathcal{N}(\mathbf{x}_{i+1}; \mathbf{f}_i(\mathbf{x}_i), \mathbf{Q}_i), \quad (73)$$

$$p(\mathbf{z}_i|\mathbf{x}_i) = \mathcal{N}(\mathbf{z}_i; \mathbf{h}_i(\mathbf{x}_i), \mathbf{R}_i) \quad (74)$$

and then the relations (9)–(12) can be rewritten as

$$\mathbf{K}_{i+1}^i = E\{[\nabla_{\mathbf{x}_i} \mathbf{f}_i(\mathbf{x}_i)]^T \mathbf{Q}_i^{-1} \nabla_{\mathbf{x}_i} \mathbf{f}_i(\mathbf{x}_i)\}, \quad (75)$$

$$\mathbf{K}_{i+1}^{i,i+1} = -E\{[\nabla_{\mathbf{x}_i} \mathbf{f}_i(\mathbf{x}_i)]^T\} \mathbf{Q}_i^{-1}, \quad (76)$$

$$\mathbf{K}_{i+1}^{i+1} = \mathbf{Q}_i^{-1}, \quad (77)$$

$$\mathbf{L}_i^i = E\{[\nabla_{\mathbf{x}_i} \mathbf{h}_i(\mathbf{x}_i)]^T \mathbf{R}_i^{-1} \nabla_{\mathbf{x}_i} \mathbf{h}_i(\mathbf{x}_i)\}. \quad (78)$$

Since $\nabla_{\mathbf{x}_0} \ln p(\mathbf{x}_0) = -(\mathbf{x}_0 - \mathbf{m}_0)^T \mathbf{M}_0^{-1}$, it follows from the definition (13) of \mathbf{K}_0^0 that $\mathbf{K}_0^0 = \mathbf{M}_0^{-1}$.

By carrying out M simulation experiments for system (71), (72) with $k = 0, 1, \dots, N$, one obtains M realizations of the state trajectory $\{\mathbf{x}_k\}_{k=0}^N$ and corresponding measurements $\{\mathbf{z}_k\}_{k=0}^N$.

The CR bounds can be computed using the relation (28) for filtering, (52) for prediction, and (64) for smoothing with the matrices \mathbf{K}_{k+1}^k , $\mathbf{K}_{k+1}^{k,k+1}$, and \mathbf{L}_k^k replaced by their estimates; e.g. the estimate of $\mathbf{K}_{k+1}^{k,k+1}$ is computed as

$$\hat{\mathbf{K}}_{k+1}^{k,k+1} = \frac{1}{M} \sum_{j=1}^M [-\nabla_{\mathbf{x}_k} \mathbf{f}_k(\mathbf{x}_k)|_{\mathbf{x}_k = \mathbf{x}_k(j)}]^T \mathbf{Q}_k^{-1},$$

where $\{\mathbf{x}_k(j)\}_{j=1}^M$ is a j th realization of the state trajectory, $j = 1, 2, \dots, M$. The quality of Monte Carlo estimates increases with M , but even a high number of simulation experiments need not guarantee that the estimates are satisfactory. Such situation is illustrated in Example 2, Fig. 10.

6.3. Linear Gaussian case

On the contrary to the previous cases, the CR bounds can be found analytically for linear Gaussian system with $\mathbf{f}_k(\mathbf{x}_k) = \mathbf{F}_k \mathbf{x}_k$ in (71) and $\mathbf{h}_k(\mathbf{x}_k) = \mathbf{H}_k \mathbf{x}_k$ in (72) where \mathbf{F}_k and \mathbf{H}_k are known $n \times n$ and $r \times n$ matrices, respectively.

Note that in this special case, the matrices in (75)–(78) become deterministic. Thus the recursive relations for all types of CR bounds (28), (27), (52), and (64), which generate inverses of the CR bounds, could be rearranged particularly using (A.1), and explicit relations for the CR bounds can be obtained.

After the rearrangement, the relations for the filtering CR bound are identical with the recursive equations for

covariance matrices in the Kalman filter (Anderson & Moore, 1979):

$$\mathbf{C}_{k|k} = \mathbf{C}_{k|k-1} - \mathbf{C}_{k|k-1} \mathbf{H}_k^T \times (\mathbf{H}_k \mathbf{C}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \mathbf{H}_k \mathbf{C}_{k|k-1}, \quad (79)$$

$$\mathbf{C}_{k+1|k} = \mathbf{F}_k \mathbf{C}_{k|k} \mathbf{F}_k^T + \mathbf{Q}_k. \quad (80)$$

The relation for the multi-step predictive CR bound $\mathbf{C}_{\ell|k}$, $\ell > k$, is given as

$$\mathbf{C}_{\ell|k} = \mathbf{F}_{\ell-1} \mathbf{C}_{\ell-1|k} \mathbf{F}_{\ell-1}^T + \mathbf{Q}_{\ell-1} \quad (81)$$

corresponding to the recursion for the covariance matrix of the optimal linear predictor.

The recursive relation for the smoothing CR bound $\mathbf{C}_{\ell|k}$, $0 \leq \ell < k$, can be rearranged to the relation for the conditional covariance matrix in the Rauch–Tung–Striebel optimal linear smoother (Lewis, 1986):

$$\mathbf{C}_{\ell|k} = \mathbf{C}_{\ell|\ell} - \mathbf{C}_{\ell|\ell} \mathbf{F}_{\ell} \mathbf{C}_{\ell+1|\ell}^{-1} \times (\mathbf{C}_{\ell+1|\ell} - \mathbf{C}_{\ell+1|k}) \mathbf{C}_{\ell+1|\ell}^{-1} \mathbf{F}_{\ell}^T \mathbf{C}_{\ell|\ell}. \quad (82)$$

7. Numerical examples

The goal of this section is to demonstrate computation and behaviour of the CR bounds for filtering, prediction, and smoothing both for the system (4) and (5) and for the system with unknown parameters (29) and (30).

Example 1. The following discrete-time nonlinear stochastic system is considered

$$\mathbf{x}_{k+1} = [x_{1,k} x_{2,k}]^T + \mathbf{w}_k, \quad z_k = x_{1,k} x_{2,k} + v_k,$$

where $\mathbf{x}_k = [x_{1,k} \ x_{2,k}]^T$. The covariances of the zero-mean white Gaussian noises $\{\mathbf{w}_k\}$, $\{v_k\}$ are $\mathbf{Q} = \text{diag}\{0.25, 10^{-5}\}$, $R = 0.01$, respectively.

Simulations of the system were carried out for the initial conditions $p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0; [10 \ -0.85]^T,$

$\text{diag}\{0.1, 10^{-3}\})$ for $k = 0, 1, \dots, N$, $N = 40$ and $M = 200$. The CR bounds were calculated for the following state estimation problems: filtering, 1-step to 7-step prediction, and 1-step to 3-step fixed-lag smoothing. Also, the fixed-interval smoothing CR bound $\mathbf{C}_{k|N}$ was computed.

The results are presented in Figs. 1–3. The two diagonal components of the matrices are displayed separately. In Figs. 1 and 2, the predictive CR bounds are depicted apart from the others and the comparison of all three types of CR bounds is shown in Fig. 3. Obviously, the predictive bounds, which lie above the filtering CR bound, exhibit a growth with increasing prediction step. Naturally, the smoothing bounds are lower than the corresponding filtering bounds since they are based on both past and future information.

Example 2. The second example demonstrates the computation of the CR bounds for a system with an unknown parameter and shows using the CR bounds for quality evaluation of nonlinear filters. The following stochastic system from Chui, Chen, and Chui (1990) is considered

$$\mathbf{x}_{k+1} = \mathbf{F}(a)\mathbf{x}_k + \mathbf{w}_k, \quad z_k = x_{1,k} + v_k,$$

where $\mathbf{x}_k = [x_{1,k} \ x_{2,k}]^T$ and a is an unknown random parameter of the matrix $\mathbf{F}(a)$ whose first row is $[1 \ a]$ and second row is $[-0.1 \ 1]$. The initial state is given by the Gaussian pdf $p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0; [20 \ 20]^T, 300 \mathbf{I}_2)$. Covariances of zero-mean white Gaussian the noises $\{\mathbf{w}_k\}$, $\{v_k\}$ are $\mathbf{Q} = 0.1 \mathbf{I}_2$, $R = 0.01$, respectively. The prior pdf of parameter a is Gaussian, $p(a) = \mathcal{N}(a; 0.1, 10^{-3})$.

As the unknown parameter is not affected by a noise, the filtering, predictive, and smoothing CR bounds must be computed by (43)–(48), (53)–(55), and (68)–(70), respectively. The \mathbf{K} , \mathbf{L} , and \mathbf{A} matrices are given by definitions (9)–(12), (33)–(39) but thanks to the assumption that \mathbf{w}_k and v_k are additive Gaussian noises, the simplifications described in Section 6.2 may be employed. The computation of \mathbf{K}_{i+1}^i , \mathbf{K}_{i+1}^{i+1} , \mathbf{K}_{i+1}^{i+1} , \mathbf{L}_i^i is straightforward

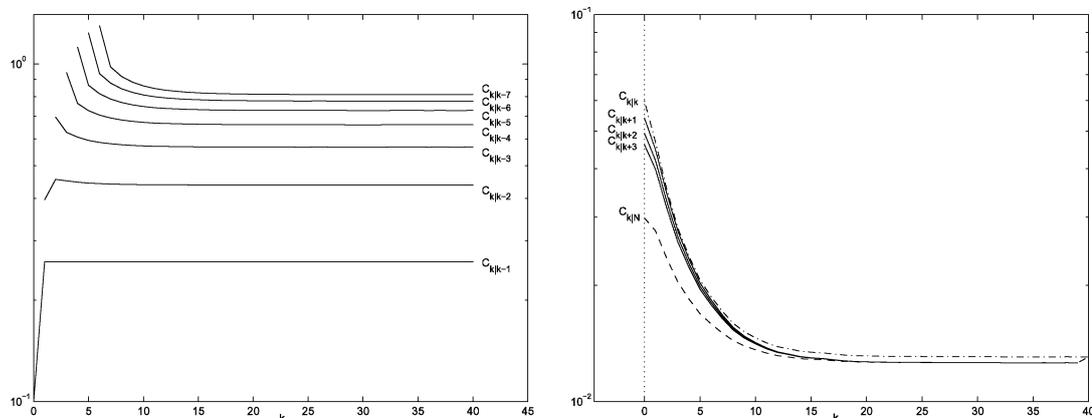


Fig. 1. Predictive CR bound (left graph) and filtering with smoothing CR bounds (right graph) for the first state component in Example 1.

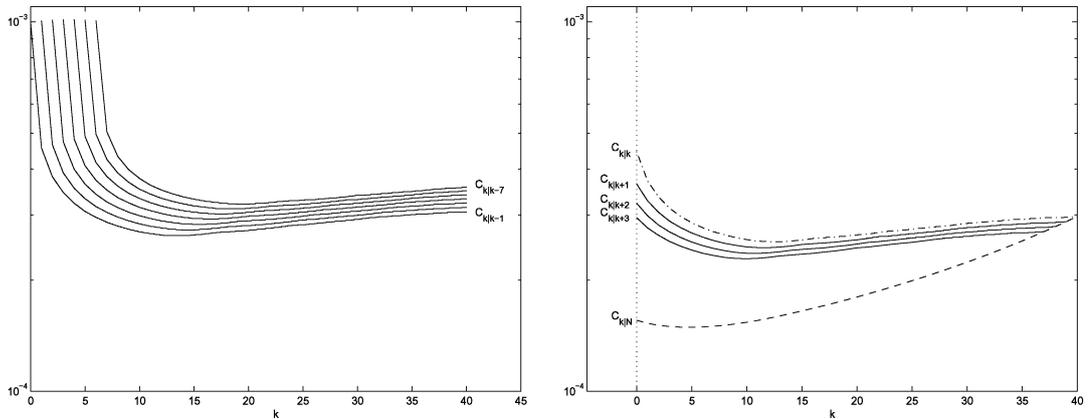


Fig. 2. Predictive CR bound (left graph) and filtering with smoothing CR bounds (right graph) for the second state component in Example 1.

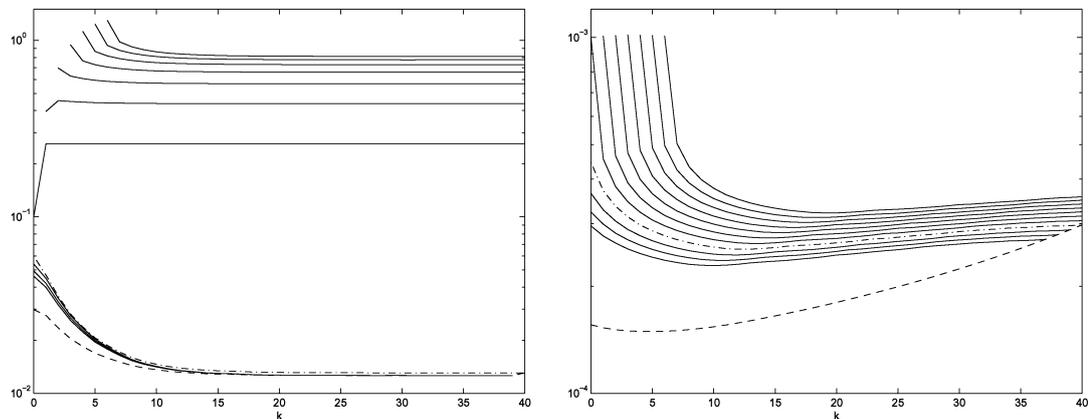


Fig. 3. Filtering CR bound (dot-and-dash line) lies above smoothing CR bounds (solid lines for fixed-lag and dashed line for fixed-interval smoothing) and below predictive CR bounds (solid lines). Left graph: first state component; right graph: second state component in Example 1.

from (75)–(78) with $\nabla_{\mathbf{x}_i} \mathbf{f}_i(\mathbf{x}_i) = \mathbf{F}(a)$ and $\nabla_{\mathbf{x}_i} \mathbf{h}_i(\mathbf{x}_i) = [1 \ 0]$. Note that all these \mathbf{K} and \mathbf{L} matrices can be enumerated analytically. For this example, the relations (33)–(35) can be expressed as $\mathbf{K}_{i+1}^{a_i} = [10E\{x_{2,i}\} \ 10E\{ax_{2,i}\}]$, $\mathbf{K}_{i+1}^{a_i+1} = [-10E\{x_{2,i}\} \ 0]$, $K_{i+1}^a = 10E\{x_{2,i}^2\}$. To enumerate these relations, the expectations $E\{x_{2,k}\}$, $E\{ax_{2,k}\}$, and $E\{x_{2,k}^2\}$ must be numerically computed using Monte Carlo simulations. Finally, $\mathbf{L}_i^{a_i}$ and L_i^a equal to $[0 \ 0]$ and 0, respectively, because the measurement pdf does not depend on the parameter a .

Simulations of the system were carried out for $k = 0, 1, \dots, N$, $N = 100$, with M Monte Carlo experiments, yielding M realizations of state and measurement trajectories $\{\mathbf{x}_k(j)\}_{k=0}^N$, $\{\mathbf{z}_k(j)\}_{k=0}^N$, and M realizations of the parameter $a(j)$ with $j = 1, 2, \dots, M$.

The time behaviour of the filtering, five-step predictive, five-step smoothing, and fixed-interval smoothing CR bounds, computed with $M = 20,000$, is depicted in Figs. 4–6 for each diagonal element of the bound

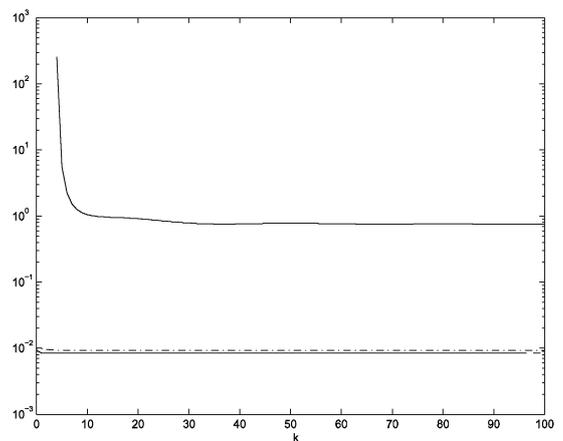


Fig. 4. Filtering CR bound $C_{k|k}^{11}$ (dot-and-dash line), predictive CR bound $C_{k|k-5}^{11}$ (upper solid line), fixed-lag smoothing CR bound $C_{k|k+5}^{11}$ (lower solid line), and fixed-interval smoothing CR bound $C_{k|N}^{11}$ (dashed line; coincides with $C_{k|k+5}^{11}$) for the first state component in Example 2.

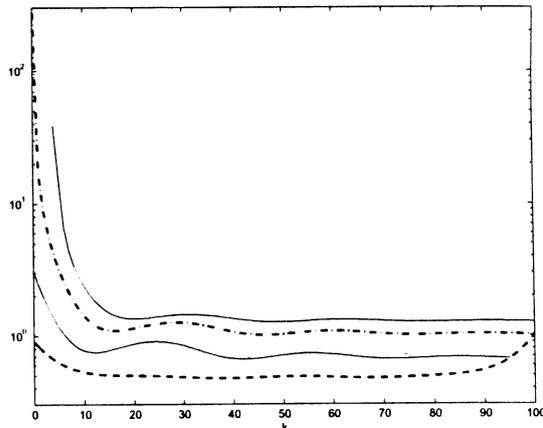


Fig. 5. Filtering CR bound $C_{k|k}^{22}$ (dot-and-dash line), predictive CR bound $C_{k|k-5}^{22}$ (upper solid line), fixed-lag smoothing CR bound $C_{k|k+5}^{22}$ (lower solid line), and fixed-interval smoothing CR bound $C_{k|N}^{22}$ (dashed line) for the second state component in Example 2.

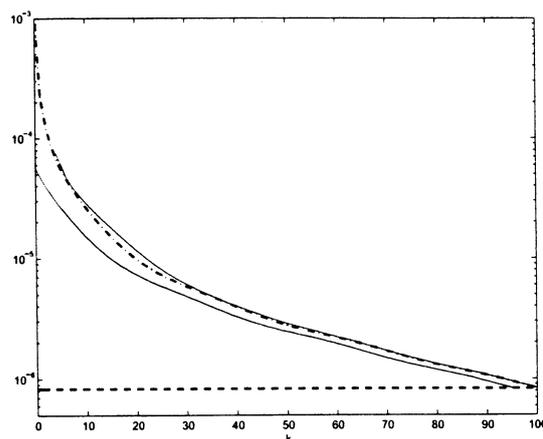


Fig. 6. Filtering CR bound $C_{k|k}^{aa}$ (dot-and-dash line), predictive CR bound $C_{k|k-5}^{aa}$ (upper solid line), fixed-lag smoothing CR bound $C_{k|k+5}^{aa}$ (lower solid line), and fixed-interval smoothing CR bound $C_{k|N}^{aa}$ (dashed line) for parameter a .

separately. The diagonal components of the CR bound for the state are denoted as $C_{k|k}^{11}$, $C_{k|k}^{22}$ and the CR bound for the parameter is denoted as $C_{k|k}^{aa}$. The filtering, predictive, and fixed-lag smoothing CR bounds for state variables tend to nonzero values with $k \rightarrow \infty$ (see Figs. 4 and 5), while the bounds for the parameter a tend to zero (see Fig. 6), because the parameter is not influenced by any noise.

For the purpose of the combined state/parameter estimation, the state of the system is extended as $\xi_k = [x_k^T a_k]^T$, where $a_{k+1} = a_k$ with $a_0 = a$. The extended state ξ_k will be estimated by three nonlinear filters: extended Kalman filter (EKF) (Anderson & Moore, 1979), modified extended Kalman filter (MEKF)

(Chui et al., 1990), and nine-term Gaussian-sum filter (GSF) (Sorenson & Alspach, 1971; Šimandl & Flidr, 1997).

The initial conditions for EKF and MEKF are given by the Gaussian priors $p(x_0)$, $p(a)$. The nine-term GSF is started from the following prior pdf:

$$p(\xi_0|z^{-1}) = \sum_{i=1}^9 \alpha_0^{(i)} \mathcal{N}(\xi_0; \hat{\xi}_0^{(i)}, \mathbf{P}_0^{(i)}),$$

where $\alpha_0^{(i)} = 1/9$, $\mathbf{P}_0^{(i)} = \text{diag}\{60, 60, 0.001\}$ for all $i = 1, 2, \dots, 9$, and

$$\begin{bmatrix} \hat{\xi}_0^{(1)} & \dots & \hat{\xi}_0^{(9)} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 20 & 20 & 20 & 39 & 39 & 39 \\ 1 & 20 & 39 & 1 & 20 & 39 & 1 & 20 & 39 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}$$

to maintain the first two moments of $p(x_0)$. For a j th realization of the measurement trajectory $\{z_k(j)\}_{k=0}^N$, $j = 1, 2, \dots, M$, where $M = 550,000$, each of the filters generates a trajectory of extended state estimates $\{\hat{\xi}_{k|k}(j)\}_{k=0}^N$. These estimates are utilized for numerical computation of the mean-square error matrix given by (6)

$$\hat{\mathbf{\Pi}}_{k|k} = \frac{1}{M} \sum_{j=1}^M [\xi_k(j) - \hat{\xi}_{k|k}(j)][\xi_k(j) - \hat{\xi}_{k|k}(j)]^T$$

for $k = 0, 1, \dots, N$, where $\{\xi_k(j)\}_{k=0}^N$ is the j th simulated state trajectory.

Time behaviours of the filtering CR bound $C_{k|k}$ and of the MSE estimates $\hat{\mathbf{\Pi}}_{k|k}$ for the three filters are displayed in Figs. 7–9 for each component ξ_k . As can be seen in the figures, the GSF and the MEKF generate estimates with much higher quality than those produced by the standard EKF. The best results are obtained by the GSF.

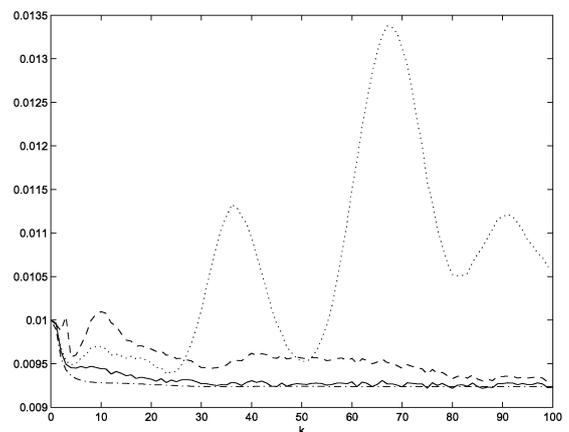


Fig. 7. Filtering CR bound $C_{k|k}^{11}$ (dot-and-dash line) and mean-square errors of the $x_{1,k}$ estimates for EKF (dotted line), MEKF (dashed line), and GSF (solid line).

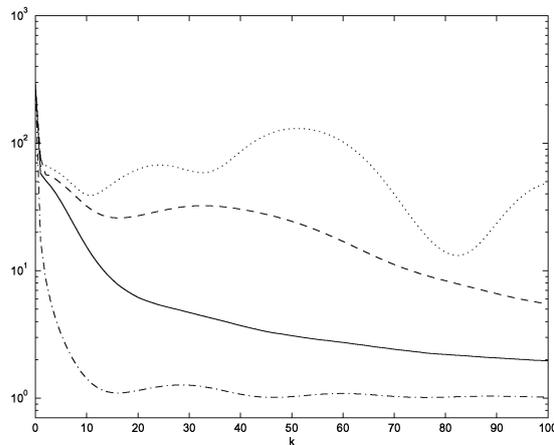


Fig. 8. Filtering CR bound $C_{k|k}^{22}$ (dot-and-dash line) and mean-square errors of the $x_{2,k}$ estimates for EKF (dotted line), MEKF (dashed line), and GSF (solid line).

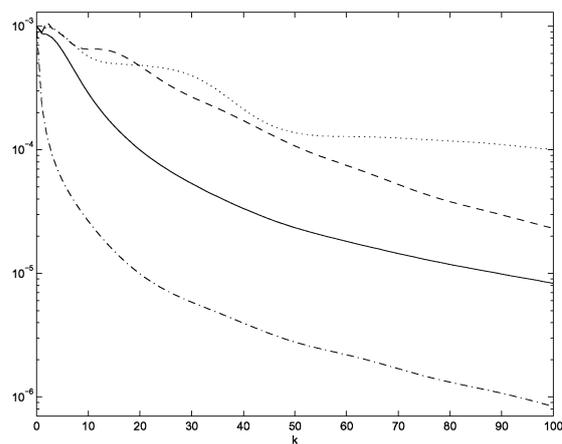


Fig. 9. Filtering CR bound $C_{k|k}^{aa}$ (dot-and-dash line) and mean-square errors of the parameter estimates for EKF (dotted line), MEKF (dashed line), and GSF (solid line).

For computation of the MSEM, it was necessary to set the number of simulations as $M = 550,000$, because for $M = 20,000$, which was used for computation of the CR bounds, the mean-square error $\hat{\Pi}_{k|k}^{11}$ of the GSF estimate of $x_{1,k}$ is so inaccurate that it significantly falls under the corresponding lower bound $C_{k|k}^{11}$, as illustrated in Fig. 10.

8. Conclusions

The Cramér–Rao bound for discrete-time nonlinear stochastic systems represents a lower limit of cognizability of the system state. It can be used as a gauge for performance evaluation of nonlinear estimators. The paper presents a unified derivation of filtering,

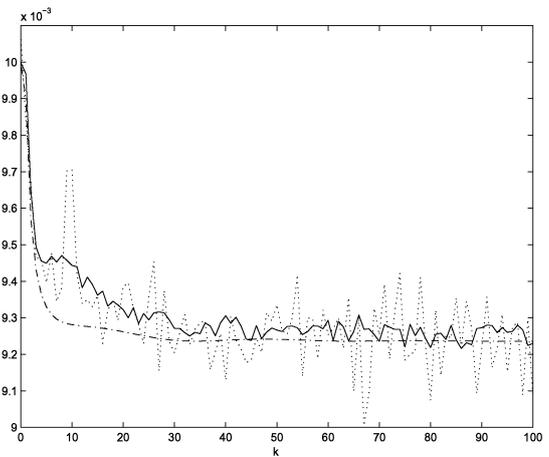


Fig. 10. Comparison of mean-square errors of GSF estimates of $x_{1,k}$ for $M = 20,000$ (dotted line) and $M = 550,000$ (solid line). The CR bound $C_{k|k}^{11}$ is depicted as dot-and-dash line.

predictive, and smoothing Cramér–Rao bounds for the state in systems that may depend on unknown parameters.

Utilization of the filtering Cramér–Rao bounds for performance evaluation of nonlinear filters and comparison of the filtering, predictive, and smoothing bounds were illustrated in numerical examples.

Acknowledgements

This work was supported by the Grant Agency of the Czech Republic under Grant No. 102/01/0021 and by the Ministry of Education of the Czech Republic under Grant No. 2352/00004. The authors would also like to acknowledge the reviewers of the paper for their suggestions.

Appendix A: Inverse of matrices

The inverse of the matrix expression $\mathbf{A} + \mathbf{BCD}$ is given by the following relation:

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B} \times (\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1} \quad (\text{A.1})$$

provided that \mathbf{A}^{-1} and \mathbf{C}^{-1} exist.

The inverse of a 2×2 block matrix can be expressed as

$$\begin{bmatrix} \mathbf{A} & \mathbf{D} \\ \mathbf{C} & \mathbf{B} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{\Phi} & -\mathbf{A}^{-1}\mathbf{D}\mathbf{\Delta}^{-1} \\ -\mathbf{\Delta}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{\Delta}^{-1} \end{bmatrix}, \quad (\text{A.2})$$

$$\begin{aligned} \mathbf{\Phi} &= \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{D}\mathbf{\Delta}^{-1}\mathbf{C}\mathbf{A}^{-1}, \\ \mathbf{\Delta} &= \mathbf{B} - \mathbf{C}\mathbf{A}^{-1}\mathbf{D} \end{aligned} \quad (\text{A.3})$$

provided that \mathbf{A}^{-1} exists.

Appendix B: Derivation of recursive relations for filtering CR bound with unknown parameters

The filtering CR bound $\mathbf{C}_{k|k}$ for the pair $(\mathbf{x}_k, \mathbf{a})$ is found as the lower-right $(n+m) \times (n+m)$ block of $\mathbf{J}_{k|k}^{-1}(\mathbf{x}^k, \mathbf{a})$ in (42) and can be expressed by aid of the block-matrix inverse (A.2) and (A.3) as

$$\mathbf{C}_{k|k}^{-1} = \begin{bmatrix} \mathbf{J}_{k|k}^{22} & \mathbf{L}_k^{ka} + \mathbf{K}_k^{ka} \\ \mathbf{L}_k^{ak} + \mathbf{K}_k^{ak} & \mathbf{L}_k^a + \mathbf{A}_k \end{bmatrix} - \begin{bmatrix} \mathbf{J}_{k|k}^{21} \\ \mathbf{G}^{k-1} \end{bmatrix} \begin{bmatrix} \mathbf{J}_{k|k}^{11} \\ \mathbf{G}^{k-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{J}_{k|k}^{12} \\ \mathbf{G}^{k-1} \end{bmatrix}^T. \quad (\text{B.1})$$

Using the notations in (15) and (31), the following relations can be written:

$$\mathbf{J}_{k|k}^{xx} = \mathbf{L}_k^k + \mathbf{K}_k^k - \mathbf{J}_{k|k}^{21} [\mathbf{J}_{k|k}^{11}]^{-1} \mathbf{J}_{k|k}^{12}, \quad (\text{B.2})$$

$$\mathbf{J}_{k|k}^{xa} = \mathbf{L}_k^{ka} + \mathbf{K}_k^{ka} - \mathbf{J}_{k|k}^{21} [\mathbf{J}_{k|k}^{11}]^{-1} [\mathbf{G}^{k-1}]^T, \quad (\text{B.3})$$

$$\mathbf{J}_{k|k}^{aa} = \mathbf{L}_k^a + \mathbf{A}_k - \mathbf{G}^{k-1} [\mathbf{J}_{k|k}^{11}]^{-1} [\mathbf{G}^{k-1}]^T. \quad (\text{B.4})$$

The one-step predictive FIM $\mathbf{J}_{k+1|k}(\mathbf{x}^{k+1}, \mathbf{a})$ has the form similar to (18),

$$\begin{bmatrix} \mathbf{J}_{k|k}^{11} & \mathbf{J}_{k|k}^{12} & 0 & [\mathbf{G}^{k-1}]^T \\ \mathbf{J}_{k|k}^{21} & \mathbf{J}_{k|k}^{22} + \mathbf{K}_{k+1}^k & \mathbf{K}_{k+1}^{k,k+1} & \mathbf{G}_k^T \\ 0 & \mathbf{K}_{k+1}^{k+1,k} & \mathbf{K}_{k+1}^{k+1} & \mathbf{K}_{k+1}^{k+1,a} \\ \mathbf{G}^{k-1} & \mathbf{G}_k & \mathbf{K}_{k+1}^{a,k+1} & \mathbf{A}_{k+1} \end{bmatrix} \quad (\text{B.5})$$

and with notation (57) the predictive FIM for $(\mathbf{x}_k, \mathbf{a})$ is

$$\mathbf{C}_{k+1|k}^{-1} = \begin{bmatrix} \mathbf{J}_{k+1|k}^{xx} & \mathbf{J}_{k+1|k}^{xa} \\ \mathbf{J}_{k+1|k}^{ax} & \mathbf{J}_{k+1|k}^{aa} \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{k+1}^{k+1} & \mathbf{K}_{k+1}^{k+1,a} \\ \mathbf{K}_{k+1}^{a,k+1} & \mathbf{A}_{k+1} \end{bmatrix} - \begin{bmatrix} 0 & \mathbf{K}_{k+1}^{k+1,k} \\ \mathbf{G}^{k-1} & \mathbf{G}_k \end{bmatrix} \Gamma_{k|k}^{-1} \begin{bmatrix} 0 & [\mathbf{G}^{k-1}]^T \\ \mathbf{K}_{k+1}^{k,k+1} & \mathbf{G}_k^T \end{bmatrix}. \quad (\text{B.6})$$

Apply the matrix inverse (A.2) and (A.3) for $\Gamma_{k|k}^{-1}$, then

$$\Gamma_{k|k}^{-1} = \begin{bmatrix} \Phi_k & -[\mathbf{J}_{k|k}^{11}]^{-1} \mathbf{J}_{k|k}^{12} \Delta_{kk}^{-1} \\ -\Delta_{kk}^{-1} \mathbf{J}_{k|k}^{21} [\mathbf{J}_{k|k}^{11}]^{-1} & \Delta_{kk}^{-1} \end{bmatrix}, \quad (\text{B.7})$$

where (cf. (B.2))

$$\begin{aligned} \Delta_{kk} &= \mathbf{J}_{k|k}^{22} + \mathbf{K}_{k+1}^k - \mathbf{J}_{k|k}^{21} [\mathbf{J}_{k|k}^{11}]^{-1} \mathbf{J}_{k|k}^{12} \\ &= \mathbf{J}_{k|k}^{xx} + \mathbf{K}_{k+1}^k, \end{aligned} \quad (\text{B.8})$$

$$\Phi_k = [\mathbf{J}_{k|k}^{11}]^{-1} + [\mathbf{J}_{k|k}^{11}]^{-1} \mathbf{J}_{k|k}^{12} \Delta_{kk}^{-1} \mathbf{J}_{k|k}^{21} [\mathbf{J}_{k|k}^{11}]^{-1}. \quad (\text{B.9})$$

Thus, from (B.6) and (B.7) one gets

$$\mathbf{J}_{k+1|k}^{xx} = \mathbf{K}_{k+1}^{k+1} - \mathbf{K}_{k+1}^{k+1,k} \Delta_{kk}^{-1} \mathbf{K}_{k+1}^{k,k+1}, \quad (\text{B.10})$$

$$\begin{aligned} \mathbf{J}_{k+1|k}^{xa} &= \mathbf{K}_{k+1}^{k+1,a} + \mathbf{K}_{k+1}^{k+1,k} \Delta_{kk}^{-1} \mathbf{J}_{k|k}^{21} [\mathbf{J}_{k|k}^{11}]^{-1} [\mathbf{G}^{k-1}]^T \\ &\quad - \mathbf{K}_{k+1}^{k+1,k} \Delta_{kk}^{-1} \mathbf{G}_k^T, \end{aligned} \quad (\text{B.11})$$

$$\begin{aligned} \mathbf{J}_{k+1|k}^{aa} &= \mathbf{A}_{k+1} - \mathbf{G}^{k-1} \Phi_k [\mathbf{G}^{k-1}]^T + \mathbf{G}_k \Delta_{kk}^{-1} \mathbf{J}_{k|k}^{21} \\ &\quad \times [\mathbf{J}_{k|k}^{11}]^{-1} [\mathbf{G}^{k-1}]^T + \mathbf{G}^{k-1} [\mathbf{J}_{k|k}^{11}]^{-1} \mathbf{J}_{k|k}^{12} \\ &\quad \times \Delta_{kk}^{-1} \mathbf{G}_k^T - \mathbf{G}_k \Delta_{kk}^{-1} \mathbf{G}_k^T. \end{aligned} \quad (\text{B.12})$$

After some simplifications, using (B.2)–(B.4), (40) and (39), relations (B.11) and (B.12) can be rewritten as

$$\mathbf{J}_{k+1|k}^{xa} = \mathbf{K}_{k+1}^{k+1,a} - \mathbf{K}_{k+1}^{k+1,k} \Delta_{kk}^{-1} \Delta_{kk}^{xa}, \quad (\text{B.13})$$

$$\mathbf{J}_{k+1|k}^{aa} = \mathbf{J}_{k|k}^{aa} + \mathbf{K}_{k+1}^a - \Delta_{kk}^{ax} \Delta_{kk}^{-1} \Delta_{kk}^{xa}, \quad (\text{B.14})$$

where $\Delta_{kk}^{ax} = \mathbf{J}_{k|k}^{ax} + \mathbf{K}_{k+1}^{ak}$ and $\Delta_{kk}^{xa} = [\Delta_{kk}^{ax}]^T$.

Comparing (B.2)–(B.4) with (B.10), (B.13) and (B.14), respectively, for time $k+1$ it is evident that the relations differ only by \mathbf{L}_{k+1}^{k+1} , $\mathbf{L}_{k+1}^{k+1,a}$, and \mathbf{L}_{k+1}^a which represent a new information about the system state \mathbf{x}_k and the parameter \mathbf{a} that is contained in the observation \mathbf{z}_{k+1} . Thus, filtering FIMs $\mathbf{J}_{k|k}^{xx}$, $\mathbf{J}_{k|k}^{xa}$, and $\mathbf{J}_{k|k}^{aa}$ are given as

$$\mathbf{J}_{k|k}^{xx} = \mathbf{J}_{k|k-1}^{xx} + \mathbf{L}_k^k, \quad (\text{B.15})$$

$$\mathbf{J}_{k|k}^{xa} = \mathbf{J}_{k|k-1}^{xa} + \mathbf{L}_k^{ka}, \quad (\text{B.16})$$

$$\mathbf{J}_{k|k}^{aa} = \mathbf{J}_{k|k-1}^{aa} + \mathbf{L}_k^a. \quad (\text{B.17})$$

Appendix C: Derivation of recursive relations for smoothing CR bound with unknown parameters

The inverse of $\mathbf{J}_{k|k}(\mathbf{x}^k, \mathbf{a})$ in (65) contains the smoothing CR bounds $\mathbf{C}_{\ell|k}^{xx}$, for $\ell = 0, 1, \dots, k-1$, and the filtering CR bounds $\mathbf{C}_{\ell|k}^{xx}$, $\mathbf{C}_{k|k}^{aa}$ on its main diagonal. Let the matrix be decomposed as follows:

$$\mathbf{J}_{k|k}^{-1}(\mathbf{x}^k, \mathbf{a}) = \begin{bmatrix} \mathbf{C}_{0|k}^{xx} & & & & & \mathbf{C}_{0|k}^{xa} \\ & \ddots & & & & \vdots \\ & & \mathbf{C}_{\ell|k}^{xx} & & & \mathbf{C}_{\ell|k}^{xa} \\ & & & \ddots & & \vdots \\ & & & & \mathbf{C}_{k|k}^{xx} & \mathbf{C}_{k|k}^{xa} \\ \mathbf{C}_{0|k}^{ax} & \dots & \mathbf{C}_{\ell|k}^{ax} & \dots & \mathbf{C}_{k|k}^{ax} & \mathbf{C}_{k|k}^{aa} \end{bmatrix} = \begin{bmatrix} [\mathbf{J}_{k|k}^{-1}]_{11} & [\tilde{\mathbf{J}}_{k|k}^{-1}]_{12} \\ [\tilde{\mathbf{J}}_{k|k}^{-1}]_{21} & [\tilde{\mathbf{J}}_{k|k}^{-1}]_{22} \end{bmatrix}, \quad (\text{C.1})$$

where all blocks except for the diagonal ones, the last row, and the last column have been left empty. The blocks that are of interest can be extracted similarly to (59) as

$$\mathbf{C}_{\ell|k}^{xx} = [0 \quad \mathbf{I}_n] [\mathbf{J}_{k|k}^{-1}]_{11} [0 \quad \mathbf{I}_n]^T, \quad (\text{C.2})$$

$$\mathbf{C}_{\ell|k}^{xa} = [0 \quad \mathbf{I}_n] [\tilde{\mathbf{J}}_{k|k}^{-1}]_{12} [0 \quad \mathbf{I}_m]^T. \quad (\text{C.3})$$

Then using (A.2) and the notation from (65) yields

$$[\mathbf{J}_{k|k}^{-1}]_{11} = \Gamma_{\ell'}^{-1} + \Gamma_{\ell'}^{-1} \tilde{\mathbf{S}}_{\ell',k} [\tilde{\mathbf{J}}_{k|k}^{-1}]_{22} \tilde{\mathbf{S}}_{\ell',k}^T \Gamma_{\ell'}^{-1} \quad (\text{C.4})$$

and

$$[\tilde{\mathbf{J}}_{k|k}^{-1}]_{12} = -\Gamma_{\ell'}^{-1} \tilde{\mathbf{S}}_{\ell',k} [\tilde{\mathbf{J}}_{k|k}^{-1}]_{22}. \quad (\text{C.5})$$

Note that the inverse of $\Gamma_{\ell'}$ has been already computed in (B.7) and $\tilde{\mathbf{S}}_{\ell',k}$ was introduced in (66). Hence using (B.3) and (40), it holds that

$$\begin{aligned} & [0 \quad \mathbf{I}_n] \Gamma_{\ell'}^{-1} \tilde{\mathbf{S}}_{\ell',k} \\ &= \begin{bmatrix} -\mathbf{J}_{\ell'}^{12} [\mathbf{J}_{\ell'}^{11}]^{-1} \Delta_{\ell'}^{-1} \\ \Delta_{\ell'}^{-1} \end{bmatrix}^T \begin{bmatrix} 0 & [\mathbf{G}^{\ell'}]^{-1} \\ \tilde{\mathbf{K}}_{\ell'} & \mathbf{G}_{\ell'}^T \end{bmatrix} \\ &= \Delta_{\ell'}^{-1} [\tilde{\mathbf{K}}_{\ell'} \quad \Delta_{\ell'}^{-1} \mathbf{G}_{\ell'}^T - \mathbf{J}_{\ell'}^{21} [\mathbf{J}_{\ell'}^{11}]^{-1} [\mathbf{G}^{\ell'}]^{-1}] \\ &= \Delta_{\ell'}^{-1} [\tilde{\mathbf{K}}_{\ell'} \quad \mathbf{L}_{\ell'}^{a} + \mathbf{K}_{\ell'+1}^{a} + \mathbf{K}_{\ell'+1}^{a} - (\mathbf{L}_{\ell'}^{a} + \mathbf{K}_{\ell'}^{a} - \mathbf{J}_{\ell'}^{xa})] \\ &= \Delta_{\ell'}^{-1} [\mathbf{K}_{\ell'+1}^{\ell'+1} \quad 0 \quad \Delta_{\ell'}^{xa}]. \end{aligned} \quad (\text{C.6})$$

Now, (C.2) and (C.3) can be expressed using (C.4), (C.5), and (C.6)

$$\begin{aligned} \mathbf{C}_{\ell'|k}^{xx} &= [0 \quad \mathbf{I}_n] (\Gamma_{\ell'}^{-1} + \Gamma_{\ell'}^{-1} \tilde{\mathbf{S}}_{\ell',k} [\tilde{\mathbf{J}}_{k|k}^{-1}]_{22} \tilde{\mathbf{S}}_{\ell',k}^T \Gamma_{\ell'}^{-1}) [0 \quad \mathbf{I}_n]^T \\ &= \Delta_{\ell'}^{-1} + \Delta_{\ell'}^{-1} [\tilde{\mathbf{K}}_{\ell'} \quad \Delta_{\ell'}^{xa}] [\tilde{\mathbf{J}}_{k|k}^{-1}]_{22} [\tilde{\mathbf{K}}_{\ell'} \quad \Delta_{\ell'}^{xa}]^T \Delta_{\ell'}^{-1} \\ &= \Delta_{\ell'}^{-1} + \Delta_{\ell'}^{-1} (\mathbf{K}_{\ell'+1}^{\ell'+1} \mathbf{C}_{\ell'+1|k}^{xx} \mathbf{K}_{\ell'+1}^{\ell'+1} \\ &\quad + \Delta_{\ell'}^{xa} \mathbf{C}_{\ell'+1|k}^{ax} \mathbf{K}_{\ell'+1}^{\ell'+1} + \mathbf{K}_{\ell'+1}^{\ell'+1} \mathbf{C}_{\ell'+1|k}^{xa} \Delta_{\ell'}^{ax} \\ &\quad + \Delta_{\ell'}^{xa} \mathbf{C}_{k|k}^{aa} \Delta_{\ell'}^{ax}) \Delta_{\ell'}^{-1} \end{aligned} \quad (\text{C.7})$$

and

$$\begin{aligned} \mathbf{C}_{\ell'|k}^{xa} &= -[0 \quad \mathbf{I}_n] \Gamma_{\ell'}^{-1} \tilde{\mathbf{S}}_{\ell',k} [\tilde{\mathbf{J}}_{k|k}^{-1}]_{22} [0 \quad \mathbf{I}_m]^T \\ &= -\Delta_{\ell'}^{-1} [\mathbf{K}_{\ell'+1}^{\ell'+1} \quad \Delta_{\ell'}^{xa}] \begin{bmatrix} \mathbf{C}_{\ell'+1|k}^{xx} & \mathbf{C}_{\ell'+1|k}^{xa} \\ \mathbf{C}_{\ell'+1|k}^{ax} & \mathbf{C}_{k|k}^{aa} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{I}_m \end{bmatrix} \\ &= -\Delta_{\ell'}^{-1} (\mathbf{K}_{\ell'+1}^{\ell'+1} \mathbf{C}_{\ell'+1|k}^{xa} + \Delta_{\ell'}^{xa} \mathbf{C}_{k|k}^{aa}). \end{aligned} \quad (\text{C.8})$$

References

- Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering*. Englewood Cliffs, NJ: Prentice-Hall.
- Bergman, N. (1999). *Recursive Bayesian estimation: Navigation and tracking applications*. Ph.D. thesis, Linköping University, Sweden.
- Bergman, N., Ljung, L., & Gustafsson, F. (1997). Point-mass filter and Cramér–Rao bound for terrain-aided navigation. *Proceedings of the 36th IEEE conference on decision and control*, San Diego, CA, pp. 565–570.
- Bobrovsky, B. Z., Mayer-Wolf, E., & Zakai, M. (1987). Some classes of global Cramér–Rao bounds. *Annals of Statistics*, 15, 1421–1438.
- Bobrovsky, B. Z., & Zakai, M. (1975). A lower bound on the estimation error for Markov processes. *IEEE Transactions on Automatic Control*, 20(6), 785–788.
- Chui, C. K., Chen, G., & Chui, H. C. (1990). Modified extended Kalman filtering and a real-time parallel algorithm for system parameter identification. *IEEE Transactions on Automatic Control*, 35(1), 100–104.
- Doerschuk, P. C. (1995). Cramér–Rao bounds for discrete-time nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 40(8), 1465–1469.
- Galdos, J. I. (1980). A Cramér–Rao bound for multi-dimensional discrete-time dynamical systems. *IEEE Transactions on Automatic Control*, 25(1), 117–119.

- Gill, R. D., & Levit, B. Y. (1995). Applications of the van Trees inequality: A Bayesian Cramér–Rao bound. *Bernoulli*, 1(1/2), 59–79.
- Kerr, T. H. (1989). Status of CR-like lower bounds for nonlinear filtering. *IEEE Transactions on Aerospace and Electronic Systems*, 25, 590–601.
- Kulhavý, R. (1996). *Recursive nonlinear estimation: A geometric approach*. London: Springer.
- Lewis, F. L. (1986). *Optimal estimation*. New York: Wiley.
- Šimandl, M., & Flidr, M. (1997). Nonlinear nonnormal dynamic models: State estimation and software. In K. Warwick, & M. Kárný (Eds.), *Computer intensive methods in control and signal processing* (pp. 195–207). Boston: Birkhäuser.
- Šimandl, M., & Královec, J. (1998). Using CR bound for performance evaluation of Gaussian sum filter. *Proceedings of IASTED conference on signal and image processing*. Las Vegas, NV, pp. 665–669.
- Šimandl, M., Královec, J., & Tichavský, P. (1999). Predictive and filtering lower bounds for nonlinear filters. *Preprints of 14th IFAC World Congress, vol. H*. Beijing, China, pp. 43–48.
- Sorenson, H. W. (1974). On the development of practical nonlinear filters. *Information Science*, 7, 230–270.
- Sorenson, H. W., & Alspach, D. L. (1971). Recursive Bayesian estimation using Gaussian sums. *Automatica*, 7(4), 465–479.
- Tichavský, P., Muravchik, C., & Nehorai, A. (1998). Posterior Cramér–Rao bounds for discrete-time nonlinear filtering. *IEEE Transactions on Signal Processing*, 46, 1386–1396.
- Van Trees, H. L. (1968). *Detection, estimation and modulation theory*. New York: Wiley.



Miroslav Šimandl was born in Ledce, Czechoslovakia, in 1954. He received the M.Sc. (Ing.) degree in Control Engineering in 1978 and the Ph.D. (CSc.) degree in Technical Cybernetics in 1984, both from the Institute of Technology in Pilsen, Czechoslovakia. In the period 1978–1992 he held various research and teaching positions at the Institute of Technology in Pilsen. Since 1993 he has been an Associate Professor at the Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia in Pilsen, Czech Republic. His main research interests are in the fields of nonlinear estimation, adaptive control, and system identification.



Jakub Královec was born in Pilsen, Czechoslovakia, in 1974. He received the M.Sc. (Ing.) degree in Cybernetics and Automatic Control from the Faculty of Applied Sciences, University of West Bohemia in Pilsen, Czech Republic in 1997. Now he is a Ph.D. student at the Department of Cybernetics, University of West Bohemia. His main research interests are in the fields of nonlinear state estimation and system identification.



Petr Tichavský graduated in 1987 from the Czech Technical University, Prague, Czechoslovakia. He received the Ph.D. degree in theoretical cybernetics from the Czechoslovak Academy of Sciences in 1992. He is with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic in Prague. In 1994 he received the Fulbright grant for a 10 month fellowship at Yale University, Department of Electrical Engineering, in New Haven, USA. His current research interests include adaptive parameter estimation, sensor array processing, wireless communications, and independent component analysis.

Near-Field/Far-Field Azimuth and Elevation Angle Estimation Using a Single Vector Hydrophone

Petr Tichavský, *Member, IEEE*, Kainam Thomas Wong, *Senior Member, IEEE*, and Michael D. Zoltowski, *Fellow, IEEE*

Abstract—This paper introduces a new underwater acoustic eigenstructure ESPRIT-based algorithm that yields closed-form direction-of-arrival (DOA) estimates using a single vector hydrophone. A vector hydrophone is composed of two or three spatially co-located but orthogonally oriented velocity hydrophones plus another optional co-located pressure hydrophone. This direction finding algorithm may (under most circumstances) resolve up to four uncorrelated monochromatic sources impinging from the near field or the far field, but it assumes that all signal frequencies are distinct. It requires no *a priori* knowledge of the signals' frequencies, suffers no frequency-DOA ambiguity, and pairs automatically the *x*-axis direction cosines with the *y*-axis direction cosines. It significantly outperforms an array of spatially displaced pressure hydrophones of comparable array-manifold size and computational load but may involve more complex hardware. This work also derives new Cramér–Rao bounds (CRBs) for various vector hydrophone constructions of arrival angle estimates for the incident uncorrelated sinusoidal signals corrupted by spatio-temporally correlated additive noise.

Index Terms—Acoustic interferometry, acoustic signal processing, acoustic velocity measurement, array signal processing, blind estimation, direction-of-arrival estimation, phased arrays, sonar arrays, sonar signal processing, underwater acoustic arrays.

I. INTRODUCTION

A new algorithm called the *univector hydrophone ESPRIT* is herein proposed and analyzed for eigenstructure-based closed-form azimuth/elevation direction-of-arrival (DOA) estimation for multiple sinusoidal sources incident from either the near-field or the far-field using a single vector hydrophone.

A vector hydrophone consists of two or three orthogonally oriented velocity hydrophones plus an optional pressure hydrophone, all co-located in space. Each velocity hydrophone measures one Cartesian component of the impinging underwater acoustic particle velocity vector field. A four-component vector hydrophone would thus measure all three Cartesian components

Manuscript received May 1, 2000; revised August 21, 2001. P. Tichavský was supported by the Grant Agency of the Czech Republic through Grants 102/97/0466 and 102/01/0021. P. Tichavský and K. T. Wong were supported by Direct Grants 2050187 and 2050247 as well as Mainline Grant 44M5010, all three from Hong Kong's Research Grant Council. K. T. Wong and M. Zoltowski were supported by U.S. National Science Foundation under Grant MIPS-9320890, the U.S. Air Force Office of Scientific Research under Contract F49620-95-1-0367, and U.S. Army Research Office Focused Research Initiative Grant DAAH04-95-1-0246. Part of this paper was presented at the MTS/IEEE Oceans96 Conference. The associate editor coordinating the review of this paper and approving it for publication was Dr. Steven Smith.

P. Tichavský is with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic.

K. T. Wong is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1.

M. D. Zoltowski is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285 USA.

Publisher Item Identifier S 1053-587X(01)09597-6.

of the acoustic velocity vector field plus the overall pressure scalar field; both the azimuth angles and the elevation angles may be estimated and automatically matched with only one vector hydrophone. Velocity hydrophone technology has been used in underwater acoustics for some time [1] and currently attracts reinvigorated attention [18]. Diverse types of velocity hydrophones are commercially available and have been constructed using a variety of technologies (see references cited in [10]).

The present scheme differs from most other direction finding methods in its recognition of the vector character of the impinging underwater acoustic wavefields and in how it uses the eigenvalues and the eigenvectors of the data covariance matrix. Customary arrays of spatially displaced pressure hydrophones typically encapsulate the arrival angle information in the spatial phase offsets among spatially displaced pressure hydrophones. In contrast, the arrival angle information here is embedded only in the intrinsic directionality of each constituent component of the vector hydrophone. Thus, the present method requires no *a priori* information of the signals' frequencies (but it is necessary that no two sources have the same frequency)¹ because the array manifold is entirely independent of signal frequency due to the spatial co-location of its constituent sensors. The complicating effects of a near-field wave front's curvature is also avoided because of the spatial co-location of the univector hydrophone array's constituent sensors. The present algorithm may be adapted to handle frequency-hopped signals of unknown hop sequences [24].

Underwater acoustic vector hydrophones have been used by D'Spain *et al.* [9] in linearly constrained minimum-variance (LCMV) beamforming toward predetermined directions. Shchurov *et al.* [12] have also deployed similar arrays to measure ambient noises but not for source localization. Nehorai and Paldi [10] first introduce the vector hydrophone measurement model to the signal processing research community; they also propose a scalar performance measure [the mean square angular error (MSAE)] and derive an expression and a bound for the MSAE for the vector hydrophone. Hochwald and Nehorai [15] investigate identifiability issues associated with vector hydrophones. Hawkes and Nehorai [21] adapt the Capon method of spectrum estimation to arrays of vector hydrophones. Hawkes and Nehorai [22] analyzed array geometry design with vector hydrophones as array elements. The vector

¹This assumption may hold in active sonar wherein the reflectors have distinct Dopplers in reference to the receiver. This assumption would be violated, say, when multipaths (with the same Doppler) arrive from a same source. The steering vectors of monochromatic sources at the same frequency cannot be separated by the present temporal-invariance implementation of ESPRIT. The frequency separation required of the monochromatic signals is investigated in Section V.

hydrophone's beam pattern is analyzed by Wong and Chi in [29]. Hawkes and Nehorai [28] investigate the performance of vector hydrophones mounted on rigid-pressure surfaces or pressure-releasing surfaces.

The *Uni-Vector-Hydrophone ESPRIT* algorithm herein proposed represents the first *eigenstructure (subspace)* method that estimates the directions-of-arrival of multiple uncorrelated monochromatic sources using only a single vector hydrophone. Eigenstructure-based (also called subspace-based) direction-finding methods such as *ESPRIT* [5], though sub-optimal, have supplanted optimal methods such as the maximum likelihood (ML) method because eigenstructure methods require only the second-order statistics of the additive noise, require lighter computation loads but still offer comparable performance as the optimal methods at low-SNR and/or few-snapshot scenarios. The direction-finding approach of normalizing the vector hydrophone steering vectors is first adapted to the eigenstructure method by Wong and Zoltowski [19] to multiple arbitrarily spaced vector-sensors at possibly unknown locations. Wong and Zoltowski [20] extend the intersensor spacing with a sparse regular array of vector hydrophones while avoiding ambiguity in the direction-cosine estimates. Wong and Zoltowski [23] also advanced a Root-MUSIC-based direction-finding algorithm applicable to vector hydrophones. Wong and Zoltowski [25] present still another direction-finding algorithm allowing irregularly spaced vector hydrophones that adaptively steers null beams in the underwater acoustic particle velocity vector-field and that self-initiates a subsequent iterative search while requiring no *a priori* source information.

In the algorithm proposed herein, a matrix pencil pair is formed out of two temporally-displaced data sets collected from the single vector hydrophone. This proposed algorithm, unlike most direction-finding applications of the popular eigenstructure-based parameter estimation algorithm *ESPRIT*, forms only a *temporal* invariance but no *spatial* invariance (which involves two displaced but otherwise identical subarrays in the overall array geometry).²

II. MATHEMATICAL MODEL OF THE FOUR-COMPONENT VECTOR HYDROPHONE MANIFOLD

The present signal model involves multiple uncorrelated monochromatic longitudinal underwater acoustic waves of distinct frequencies, having traveled through a homogeneous isotropic medium and impinging upon a single four-component vector hydrophone. The k th impinging underwater acoustic wave front would have the following 4×1 array manifold at the vector hydrophone [10]:

$$\mathbf{a}_k \stackrel{\text{def}}{=} \begin{bmatrix} \sin\psi_k \cos\phi_k \\ \sin\psi_k \sin\phi_k \\ \cos\psi_k \\ 1 \end{bmatrix} = \begin{bmatrix} u(\psi_k, \phi_k) \\ v(\psi_k, \phi_k) \\ w(\psi_k) \\ 1 \end{bmatrix}. \quad (1)$$

²A direction-finding algorithm specifically for vector hydrophones exploiting spatial invariances may be found in [20]; a direction-finding algorithm exploiting the invariances among various Cartesian components of the underwater acoustic velocity field may be found in [19]. An electromagnetic version of the present algorithm has also been proposed by the present authors [16] using three orthogonally oriented electric dipoles and three orthogonally oriented magnetic poles, all co-located in space.

The first, second, and third component above corresponds to the velocity hydrophone aligned along, respectively, the x-axis, the y-axis, and the z-axis. The last component corresponds to the pressure hydrophone. ψ_k may range between $0 \leq \psi_k < \pi$ (instead of $0 \leq \psi_k < \pi/2$) because the pressure hydrophone helps to distinguish between acoustic compressions and dilation. This is important because acoustic particle motion sensors (such as a velocity hydrophone), by themselves, suffer a 180° ambiguity, with their plane-wave response given by the "Fig. 8" curve.

There exist several essential observations about the vector hydrophone array manifold. First, one single vector hydrophone measurement yields a 4×1 steering vector. Thus, a single vector hydrophone effectively embodies a four-element array in and of itself. Second, this vector hydrophone array manifold contains no time-delay phase factor. That is, the vector hydrophone array manifold, unlike that of a spatially displaced array, is *independent* of the impinging signals' frequency spectra. This frequency independence is due to the spatial co-location of the four component sensors that comprise the vector hydrophone. Third, the Frobenius norm of the first three components of any source's steering vector always equal to its fourth component, regardless of source parameters; the first three components of \mathbf{a}_k gives the three Cartesian direction cosines. Thus, if the steering vectors of all impinging sources can be estimated from the received data, then the signal-of-interests' DOAs can be estimated by normalizing each steering vector to have norm equal to $\sqrt{2}$ and to have the last component to equal to 1.

III. ESTIMATION OF AZIMUTH AND ELEVATION ANGLES

A. Uni-Vector-Hydrophone Data Model

Uni-Vector-Hydrophone ESPRIT forms a temporal invariance via two time-delayed data sets collected from one vector hydrophone. That is, the k th monochromatic signal impinging on the vector hydrophone produces two $4 \times N$ time-delayed (and possibly overlapping) data sets

$$\begin{aligned} \mathbf{a}_k s(t_n, f_k), & \quad \text{for } n = 1, \dots, N \\ \mathbf{a}_k s(t_n + \Delta_T, f_k) &= \mathbf{a}_k s(t_n, f_k) e^{j2\pi f_k \Delta_T} \end{aligned}$$

where the vector hydrophone steering vector \mathbf{a}_k is defined as in (1), and

$$s(t_n, f_k) \stackrel{\text{def}}{=} b_k e^{j(2\pi f_k t_n + \varphi_k)} \quad k = 1, \dots, K \quad (2)$$

where

- b_k k th source's amplitude;
- f_k k th signal's frequency;
- φ_k k th signal's uniformly distributed random phase;
- Δ_T constant time delay between the two sets of time samples.

Note that the invariance $e^{j2\pi f_k \Delta_T}$ does *not* depend on the arrival angles but only on the signal frequency and the time delay Δ_T . Δ_T may be completely arbitrary and is *not* constrained by the Nyquist sampling rate to be twice the highest signal frequency as long as a set of distinct phase delays $\{e^{j2\pi f_k \Delta_T}, k = 1, \dots, K\}$ are preserved. It is necessary that $|f_j - f_k| \neq l/\Delta_T$. Section V will investigate the spectral separation needed for source resolution. The present algorithm

requires no prior knowledge of the value of Δ_T , although that information is typically available for all $j \neq k$. The above sinusoid model of the uncorrelated incident signals differ from that in [10], [19]–[25], and [27], which models the set of all incident signals as a set of statistically independent and temporally uncorrelated zero-mean complex-Gaussian random sequences. Section V will derive and analyze the applicable nonasymptotic and asymptotic Cramér–Rao bounds (CRBs) for the uncorrelated sinusoidal model used here.

With a total of K impinging signals³ and additive complex-valued zero-mean, possibly spatio-temporally correlated, noise at each constituent sensor of the vector hydrophone

$$\begin{aligned} \mathbf{z}(t_n) &\stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \mathbf{s}(t_n) + \mathbf{n}(t_n) \\ &= \sum_{k=1}^K \begin{bmatrix} \mathbf{a}_k s(t_n, f_k) \\ \mathbf{a}_k e^{j2\pi f_k \Delta_T} s(t_n, f_k) \end{bmatrix} + \mathbf{n}(t_n) \end{aligned} \quad (3)$$

where

$$\mathbf{A}_1 \stackrel{\text{def}}{=} [\mathbf{a}_1, \dots, \mathbf{a}_K] \quad (4)$$

$$\mathbf{A}_2 \stackrel{\text{def}}{=} [\mathbf{a}_1 e^{j2\pi f_1 \Delta_T}, \dots, \mathbf{a}_K e^{j2\pi f_K \Delta_T}] = \mathbf{A}_1 \Phi \quad (5)$$

$$\begin{aligned} \mathbf{s}(t_n) &\stackrel{\text{def}}{=} \begin{bmatrix} s(t_n, f_1) \\ \vdots \\ s(t_n, f_K) \end{bmatrix}; \quad \mathbf{n}(t_n) \stackrel{\text{def}}{=} \begin{bmatrix} n_1(t_n) \\ \vdots \\ n_8(t_n) \end{bmatrix} \\ \Phi &\stackrel{\text{def}}{=} \begin{bmatrix} e^{j2\pi f_1 \Delta_T} & & \\ & \ddots & \\ & & e^{j2\pi f_K \Delta_T} \end{bmatrix}. \end{aligned} \quad (6)$$

The entire $8 \times N$ set of collected data measurements is

$$\mathbf{Z} \stackrel{\text{def}}{=} [\mathbf{z}(t_1) \cdots \mathbf{z}(t_N)] = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} \quad (7)$$

where \mathbf{Z}_1 represents the $4 \times N$ data set sampled at $\{t_1, \dots, t_N\}$, and \mathbf{Z}_2 represents the $4 \times N$ data set sampled at $\{t_1 + \Delta_T, \dots, t_N + \Delta_T\}$. The present direction finding problem⁴ is to determine $\{\psi_k, \phi_k, k = 1, \dots, K\}$ from the $8 \times N$ data set above without any *a priori* knowledge of $\{\Delta_T, f_k, b_k, \varphi_k, k = 1, \dots, K\}$.

In practice, the observations in \mathbf{Z}_1 and \mathbf{Z}_2 may overlap. The CRB performance analysis in Section V will assume a single $4 \times N$ data set $\hat{\mathbf{Z}}$ sampled at $t_n = (n-1)/f_s, n = 1, \dots, N$, where f_s is the sampling frequency. From $\hat{\mathbf{Z}}$, one data subset

³Subspace-based parameter estimation algorithms, such as ESPRIT, attempt to separate the signal and the noise, respectively, into a signal subspace and a noise subspace. This leads to the requirement that the number of incident signals must be less than the maximally achievable rank of the data covariance matrix, which in the present case equals 4 for the four-component vector hydrophone. The number of resolvable sources is further limited by the identifiability of the sources' steering vectors [15]. It is found in [15] that only up to two arbitrarily oriented sources can always be uniquely identified, although in many situations, the number of resolvable sources can be larger. However, if N is sufficiently large (say, $N = 100$), the signal frequencies may first be estimated, and then, the sources may be separated from each other through a spectral comb filter. The spectrally separated data may then be processed by this processed algorithm. In such case, an infinite number of uncorrelated sources with distinct frequencies may potentially be resolved as N approaches infinity.

⁴Although the proposed algorithm is to be developed below for the batch processing mode, real-time adaptive implementations of this present algorithm may be readily realized for nonstationary environments, using the fast recursive eigendecomposition updating methods such as that in [13].

\mathbf{Z}_1 is formed containing data sampled at $\{t_1, \dots, t_{N-L}\}$, and a second data subset \mathbf{Z}_2 is formed containing data sampled at $\{t_{L+1}, \dots, t_N\}$, where $L \geq 1$ is an integer constant. Analogous to the case of applying ESPRIT to a uniformly spaced linear array of identical sensors [7], the proposed algorithm's performance may be best for some $L \neq 1$.

B. Adapting ESPRIT and TLS-ESPRIT to One Vector Hydrophone

In this subsection, the ESPRIT [5] and the *total least squares* (TLS) ESPRIT [7] parameter estimation algorithms are adapted to one vector hydrophone.

Let $\mathbf{E}_s = [\mathbf{E}_1^T, \mathbf{E}_2^T]^T$ denote the $8 \times K$ signal-subspace eigenvector matrix for the matrix $\mathbf{Z}\mathbf{Z}^H$, whose K columns are the principal eigenvectors of $\mathbf{Z}\mathbf{Z}^H$ associated with the K largest eigenvalues of the matrix. The partitioning of \mathbf{E}_s is such that \mathbf{E}_1 and \mathbf{E}_2 , respectively, represent the top and bottom $4 \times K$ submatrices of \mathbf{E}_s .

For analytical purposes, first consider the noiseless case. It follows from the model in (3)–(7) that an alternative basis of the signal subspace is that spanned by the columns of the matrix $[\mathbf{A}_1^T, \mathbf{A}_2^T]^T$. This implies the existence of a unique $K \times K$ nonsingular matrix \mathbf{T} such that

$$\mathbf{E}_1 = \mathbf{A}_1 \mathbf{T} \quad \text{and} \quad \mathbf{E}_2 = \mathbf{A}_2 \mathbf{T} = \mathbf{A}_1 \Phi \mathbf{T} \quad (8)$$

Introducing the notation $\Psi = \mathbf{T}^{-1} \Phi \mathbf{T}$

$$\mathbf{E}_1 \Psi = \mathbf{E}_2. \quad (9)$$

Because Φ is a diagonal matrix, its diagonal elements $\{[\Phi]_{k,k} = e^{j2\pi f_k \Delta_T}, k = 1, \dots, K\}$ equal the eigenvalues of Ψ . The columns of \mathbf{T}^{-1} represent the corresponding right eigenvectors.

In the noisy case, when \mathbf{E}_1 and \mathbf{E}_2 are, respectively, replaced by the estimates $\hat{\mathbf{E}}_1$ and $\hat{\mathbf{E}}_2$, the equality in (9) becomes an approximation. The least-square fit to (9) is determined by

$$\hat{\Psi}_{LS} = (\hat{\mathbf{E}}_1^H \hat{\mathbf{E}}_1)^{-1} \hat{\mathbf{E}}_1^H \hat{\mathbf{E}}_2. \quad (10)$$

The TLS fit to the (9) is given by [7]

$$\hat{\Psi}_{TLS} = -\mathbf{V}_{1,2} \mathbf{V}_{2,2}^{-1} \quad (11)$$

where the $K \times K$ matrices $\mathbf{V}_{1,2}$ and $\mathbf{V}_{2,2}$ are implicitly defined by the eigen-decomposition of

$$\begin{aligned} &\begin{bmatrix} \hat{\mathbf{E}}_1^H \\ \hat{\mathbf{E}}_2^H \end{bmatrix} \begin{bmatrix} \hat{\mathbf{E}}_1 & \hat{\mathbf{E}}_2 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{V}_{1,1} & \mathbf{V}_{1,2} \\ \mathbf{V}_{2,1} & \mathbf{V}_{2,2} \end{bmatrix} \mathbf{L} \begin{bmatrix} \mathbf{V}_{1,1}^H & \mathbf{V}_{2,1}^H \\ \mathbf{V}_{1,2}^H & \mathbf{V}_{2,2}^H \end{bmatrix} \end{aligned} \quad (12)$$

where $\mathbf{L} \stackrel{\text{def}}{=} \text{diag}(l_1, \dots, l_{2K})$, whose diagonal elements are the eigenvalues of $[\hat{\mathbf{E}}_1 \hat{\mathbf{E}}_2]^H [\hat{\mathbf{E}}_1 \hat{\mathbf{E}}_2]$ ordered nonincreasingly.

The matrices $\hat{\Phi}$ and $\hat{\mathbf{T}}$ are computed by eigen-decomposition of $\hat{\Psi}_{TLS}$.

Because in the noiseless case the steering vectors obey the relation

$$\begin{aligned} \mathbf{A}_1 &= [\mathbf{a}_1, \dots, \mathbf{a}_K] = \mathbf{E}_1 \mathbf{T}^{-1} = \mathbf{E}_2 \mathbf{T}^{-1} \Phi^{-1} \\ &= \frac{1}{2} \{ \mathbf{E}_1 \mathbf{T}^{-1} + \mathbf{E}_2 \mathbf{T}^{-1} \Phi^{-1} \} \end{aligned} \quad (13)$$

they can, in general, be estimated as

$$\hat{\mathbf{A}}_1 = [\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_K] = \frac{1}{2} \{ \hat{\mathbf{E}}_1 \hat{\mathbf{T}}^{-1} + \hat{\mathbf{E}}_2 \hat{\mathbf{T}}^{-1} \hat{\mathbf{\Phi}}^{-1} \}. \quad (14)$$

The $\hat{\mathbf{\Phi}}^{-1}$ factor in the above expression facilitates *coherent* summation of the two sets of signal-subspace eigenvectors (which differ by the phase factors $\{[\mathbf{\Phi}]_{k,k}, k = 1, \dots, K\}$) and is pivotal to the proposed algorithm's performance.

From $\hat{\mathbf{a}}_k$, the direction cosines may be estimated as follows:

$$\hat{u}_k \stackrel{\text{def}}{=} \frac{[\hat{\mathbf{a}}_k]_1}{\sqrt{[\hat{\mathbf{a}}_k]_1^2 + [\hat{\mathbf{a}}_k]_2^2 + [\hat{\mathbf{a}}_k]_3^2}} \quad (15)$$

$$\hat{v}_k \stackrel{\text{def}}{=} \frac{[\hat{\mathbf{a}}_k]_2}{\sqrt{[\hat{\mathbf{a}}_k]_1^2 + [\hat{\mathbf{a}}_k]_2^2 + [\hat{\mathbf{a}}_k]_3^2}} \quad (16)$$

$$\hat{w}_k \stackrel{\text{def}}{=} \frac{[\hat{\mathbf{a}}_k]_3}{\sqrt{[\hat{\mathbf{a}}_k]_1^2 + [\hat{\mathbf{a}}_k]_2^2 + [\hat{\mathbf{a}}_k]_3^2}}. \quad (17)$$

From the above direction-cosine estimates, the k th signal's arrival angles may be estimated as

$$\hat{\psi}_k = \arcsin \left(\sqrt{\hat{u}_k^2 + \hat{v}_k^2} \right) = \arccos(\hat{w}_k) \quad (18)$$

$$\hat{\phi}_k = \angle(\hat{u}_k + j\hat{v}_k). \quad (19)$$

Azimuth-elevation direction finding has thus been performed *without any a priori* knowledge of the signal frequencies while using just one solitary vector hydrophone and no planar arrays. The azimuth angle estimates and the elevation angle estimates are automatically matched without any additional processing.

The values of $\{f_k, k = 1, \dots, K\}$ need not be known *a priori* for *univector hydrophone ESPRIT*, which also incurs no frequency-DOA ambiguity, as would an array of spatially displaced pressure hydrophones. Such a spatially displaced array estimates the DOAs through the phase factors $e^{j2\pi(f_k d/c) \sin \psi_k \cos \phi_k}$ and $e^{j2\pi(f_k d/c) \sin \psi_k \sin \phi_k}$ (where d denotes the intersensor spacing, and c symbolizes the propagation speed); thus, f_k must be precisely known *a priori* or otherwise estimated through extra computation in order to estimate the DOAs *unambiguously*. In contrast, *Uni-Vector-Hydrophone ESPRIT* estimates the DOAs by performing a normalization on each source's frequency-independent steering-vector estimate and, thus, suffers no frequency-DOA ambiguity.

This present delay-sampling construction of a temporal invariance would not be useful with an array of spatially displaced pressure hydrophones. Such a spatially displaced array's steering vectors can still be estimated, but with no prior knowledge of each signals' frequency, no closed-form DOA-estimation solution for using such an *arbitrary* array of spatially displaced pressure hydrophones is yet known. Iterative search methods (such as *MUSIC*) would become necessary, resulting in much heavier computational costs.

IV. ALTERNATE CONSTRUCTIONS OF THE VECTOR HYDROPHONE

The present method will also work if any one of the four constituent hydrophones is taken out from the previously defined vector hydrophone configuration, rendering the 4×1

array manifold in (1) to become a 3×1 array manifold; the maximum resolvable number of sources may decrease. A procedure similar to that in Section IV-B produces $\hat{\mathbf{a}}_k$. If it is the pressure hydrophone that is removed, (15)–(17) still hold, but ψ_k now may range only between $0 \leq \psi_k < \pi/2$ instead of $0 \leq \psi_k < \pi$ because acoustic compressions and dilation can no longer be distinguished.

If the x-axis velocity hydrophone is removed, then (15) and (16) become

$$\hat{u}_k \stackrel{\text{def}}{=} \sqrt{1 - \left(\frac{[\hat{\mathbf{a}}_k]_1}{[\hat{\mathbf{a}}_k]_3} \right)^2 - \left(\frac{[\hat{\mathbf{a}}_k]_2}{[\hat{\mathbf{a}}_k]_3} \right)^2} \quad \text{and} \quad \hat{v}_k \stackrel{\text{def}}{=} \frac{[\hat{\mathbf{a}}_k]_1}{[\hat{\mathbf{a}}_k]_3}. \quad (20)$$

If the y-axis velocity hydrophone is removed, the corresponding equations are

$$\hat{u}_k \stackrel{\text{def}}{=} \frac{[\hat{\mathbf{a}}_k]_1}{[\hat{\mathbf{a}}_k]_3} \quad \text{and} \quad \hat{v}_k \stackrel{\text{def}}{=} \sqrt{1 - \left(\frac{[\hat{\mathbf{a}}_k]_1}{[\hat{\mathbf{a}}_k]_3} \right)^2 - \left(\frac{[\hat{\mathbf{a}}_k]_2}{[\hat{\mathbf{a}}_k]_3} \right)^2}. \quad (21)$$

If the z-axis velocity hydrophone is removed, then (15) and (16) become

$$\hat{u}_k \stackrel{\text{def}}{=} \frac{[\hat{\mathbf{a}}_k]_1}{[\hat{\mathbf{a}}_k]_3} \quad \text{and} \quad \hat{v}_k \stackrel{\text{def}}{=} \frac{[\hat{\mathbf{a}}_k]_2}{[\hat{\mathbf{a}}_k]_3}. \quad (22)$$

For cases without either the x-axis or the y-axis velocity hydrophone (ψ_k, ϕ_k) cannot be distinguished from $(\pi - \psi_k, \pi + \phi_k)$. To avoid problems with this ambiguity, either ϕ_k needs to be restricted to $[0, \pi)$ (instead of $[0, 2\pi)$), or ψ_k needs to be confined to $[0, \pi/2)$ (instead of $[0, \pi)$) for all $\{k = 1, \dots, K\}$. The omission of the vertical velocity hydrophone avoids direct measurement of the vertical component of the underwater acoustical particle motion, thereby allowing actual ocean acoustics to be better modeled as rectilinear. Moreover, ambient oceanic noise tends to be vertically directional, and therefore, the vertically oriented velocity hydrophone's noise level may likely exceed those at the other component hydrophones. An example of such a construction of the vector hydrophone is the cardioid [11].

V. NEW CRAMÉR–RAO BOUND EXPRESSIONS FOR PERFORMANCE ANALYSIS

The CRB analysis in [10], [21], [22], and [27] models the set of all incident signals as a set of statistically independent temporally and uncorrelated zero-mean complex-Gaussian random sequences. That model is inapplicable to the present scheme where the set of incident signals are uncorrelated pure sinusoids. Moreover, [10], [21], [22], and [27] assume the noise to be spatio-temporally uncorrelated, which typically is not the case in underwater acoustics [28]. This section derives and analyzes the nonasymptotic and asymptotic CRBs for each of the vector hydrophone constructions considered earlier in the presence of spatio-temporally correlated or uncorrelated noise. Alternate theoretical performance bounds for a vector hydrophone, other than the CRB, are discussed in [26].

A. For Noise of Arbitrary Spatio-Temporal Correlation

All noise samples $\{n_k(t_n), k = 1, \dots, 4, \text{ and } n = 1, \dots, N\}$ are herein modeled as zero-mean, circular complex Gaussian

with known spatio-temporal covariance matrix $\mathbf{\Gamma}$, drawing its elements from the spatio-temporal covariance function

$$c(k, l, n, m) = E[n_k^*(t_n)n_l(t_m)]. \quad (23)$$

The subsequent exposition will first consider the general case of $\mathbf{\Gamma}$ having arbitrary spatio-temporal correlation, to be followed by two special cases of $\mathbf{\Gamma}$:

- 1) noise that is first-order auto-regressively correlated in time but has a time-invariant spatial covariance matrix $\mathbf{\Gamma}_0 = \text{diag}(\sigma_v^2, \sigma_v^2, \sigma_v^2, \sigma_p^2)$, where σ_v^2 and σ_p^2 , respectively, are variances of errors of the velocity and pressure hydrophones;
- 2) temporally white noise with time-invariant spatial covariance matrix $\mathbf{\Gamma}_0$.

With the AR(1) parameter $|\rho| < 1$

$$c(k, l, n, m) = [\mathbf{\Gamma}_0]_{k,l} \rho^{|n-m|}. \quad (24)$$

For temporally uncorrelated noise, $\rho = 0$. For temporally correlated noise, $\rho \neq 0$; $c(k, l, n, m)$ drops to 50% of its peak value at time lag equal to $\log(0.5)/\log(\rho)$ and to 10% at $\log(0.1)/\log(\rho)$. From (24)

$$\mathbf{\Gamma} = \mathbf{\Gamma}_0 \otimes \mathbf{M}_\rho \quad (25)$$

where

$$\mathbf{M}_\rho = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \dots & \dots & 1 \end{bmatrix}. \quad (26)$$

Note that $\mathbf{\Gamma}^{-1} = \mathbf{\Gamma}_0^{-1} \otimes \mathbf{M}_\rho^{-1}$, where \mathbf{M}_ρ^{-1} is a tridiagonal matrix

$$\mathbf{M}_\rho^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & & & 0 \\ -\rho & 1+\rho^2 & -\rho & & \\ & \ddots & \ddots & \ddots & \\ & & & 1+\rho^2 & -\rho \\ 0 & & & -\rho & 1 \end{bmatrix}. \quad (27)$$

Define $\theta \stackrel{\text{def}}{=} [\theta_1^T, \dots, \theta_K^T]^T$, where $\theta_k \stackrel{\text{def}}{=} [b_k, \psi_k, \phi_k, \varphi_k, f_k]^T$. [Note that φ_k , herein defined as the k th signal's phase in the middle of the given time interval $(1, N)$, differs from the φ_k of (2).] Let $\tilde{\mathbf{z}} = \text{vec}(\tilde{\mathbf{Z}})$ be a vector consisting of all collected time samples; thus, $\tilde{\mathbf{z}} \sim N_c(\mu(\theta), \mathbf{\Gamma})$, where $\mu(\theta) \stackrel{\text{def}}{=} \sum_{k=1}^K b_k \mathbf{a}_k \otimes \mathbf{s}_k$. \otimes denotes the Kronecker product, and

$$\mathbf{s}_k \stackrel{\text{def}}{=} \mathbf{s}(f_k, \varphi_k) \stackrel{\text{def}}{=} \begin{bmatrix} e^{j[2\pi f_k(1-\frac{N+1}{2})+\varphi_k]} \\ \vdots \\ e^{j[2\pi f_k(N-\frac{N+1}{2})+\varphi_k]} \end{bmatrix}^T.$$

The Fisher information matrix equals [14]

$$\mathbf{J}(\theta) = 2 \text{Re} \left[\left(\frac{\partial \mu(\theta)}{\partial \theta} \right)^H \mathbf{\Gamma}^{-1} \frac{\partial \mu(\theta)}{\partial \theta} \right] \quad (28)$$

where elements of the vector $\partial \mu(\theta)/\partial \theta$ are

$$\frac{\partial \mu(\theta)}{\partial b_k} = \mathbf{a}_k \otimes \mathbf{s}_k \quad (29)$$

$$\frac{\partial \mu(\theta)}{\partial \psi_k} = b_k \frac{\partial \mathbf{a}_k}{\partial \psi_k} \otimes \mathbf{s}_k \quad (30)$$

$$\frac{\partial \mu(\theta)}{\partial \phi_k} = b_k \frac{\partial \mathbf{a}_k}{\partial \phi_k} \otimes \mathbf{s}_k \quad (31)$$

$$\frac{\partial \mu(\theta)}{\partial \varphi_k} = j b_k \mathbf{a}_k \otimes \mathbf{s}_k \quad (32)$$

$$\frac{\partial \mu(\theta)}{\partial f_k} = 2\pi b_k \mathbf{a}_k \otimes \tilde{\mathbf{s}}_k \quad (33)$$

with

$$\tilde{\mathbf{s}}_k \stackrel{\text{def}}{=} j \left[1 - \frac{(N+1)}{2}, 2 - \frac{(N+1)}{2}, \dots, N - \frac{(N+1)}{2} \right]^T \odot \mathbf{s}_k$$

and \odot represents the element-wise (Hadamard) product operator. Representing $\mathbf{J}(\theta)$ in a block matrix form

$$\mathbf{J}(\theta) = \begin{bmatrix} \mathbf{J}_{1,1} & \dots & \mathbf{J}_{1,K} \\ \vdots & & \vdots \\ \mathbf{J}_{K,1} & \dots & \mathbf{J}_{K,K} \end{bmatrix} \quad (34)$$

the (k, l) th block equals

$$\mathbf{J}_{k,l} = \begin{bmatrix} J_{b_k, b_l} & J_{b_k, \psi_l} & J_{b_k, \phi_l} & J_{b_k, \varphi_l} & J_{b_k, f_l} \\ J_{\psi_k, b_l} & J_{\psi_k, \psi_l} & J_{\psi_k, \phi_l} & J_{\psi_k, \varphi_l} & J_{\psi_k, f_l} \\ J_{\phi_k, b_l} & J_{\phi_k, \psi_l} & J_{\phi_k, \phi_l} & J_{\phi_k, \varphi_l} & J_{\phi_k, f_l} \\ J_{\varphi_k, b_l} & J_{\varphi_k, \psi_l} & J_{\varphi_k, \phi_l} & J_{\varphi_k, \varphi_l} & J_{\varphi_k, f_l} \\ J_{f_k, b_l} & J_{f_k, \psi_l} & J_{f_k, \phi_l} & J_{f_k, \varphi_l} & J_{f_k, f_l} \end{bmatrix}. \quad (35)$$

For arbitrary but size-compatible vectors \mathbf{a} , \mathbf{b} , \mathbf{e} , and \mathbf{f} and matrices \mathbf{B} and \mathbf{C} , it holds that $(\mathbf{a} \otimes \mathbf{b})^H (\mathbf{B} \otimes \mathbf{C})(\mathbf{e} \otimes \mathbf{f}) = \mathbf{a}^H \mathbf{B} \mathbf{e} \mathbf{b}^H \mathbf{C} \mathbf{f}$. Hence, using (25) and (28)–(33)

$$J_{b_k, b_l} = \frac{2}{\sigma_v^2} \mathbf{a}_k^T \mathbf{\Gamma}_0^{-1} \mathbf{a}_l \text{Re} [\mathbf{s}_k^H \mathbf{M}_\rho^{-1} \mathbf{s}_l] \quad (36)$$

$$J_{b_k, \psi_l} = \frac{2}{\sigma_p^2} b_l \mathbf{a}_k^T \mathbf{\Gamma}_0^{-1} \frac{\partial \mathbf{a}_l}{\partial \psi_l} \text{Re} [\mathbf{s}_k^H \mathbf{M}_\rho^{-1} \mathbf{s}_l]$$

$$k, l = 1, \dots, K.$$

$$\vdots \quad (37)$$

The asymptotic behavior of $\{\mathbf{J}_{k,l}, k, l = 1, \dots, K\}$ will next be studied, assuming N to be large and the signal frequencies to be fixed and distinct. The off-diagonal elements of these matrices, after appropriate normalization, will be shown to be asymptotically negligible. The asymptotic behaviors of the above-mentioned matrices are determined by the products $\mathbf{s}_k^H \mathbf{M}_\rho^{-1} \mathbf{s}_l$, $\mathbf{s}_k^H \mathbf{M}_\rho^{-1} \tilde{\mathbf{s}}_l$, and $\tilde{\mathbf{s}}_k^H \mathbf{M}_\rho^{-1} \tilde{\mathbf{s}}_l$.

For temporally white noise (i.e., $\rho = 0$ and \mathbf{M}_ρ is an identity matrix), the definition of $\{\mathbf{s}_k, k = 1, \dots, K\}$ gives the equa-

tion at the bottom of the page, where $\Delta_{k,l} \stackrel{\text{def}}{=} \pi(f_l - f_k)$. As a result, for large N , $k \neq l$, and $k, l = 1, \dots, K$

$$\mathbf{J}_{k,l} = \begin{bmatrix} O(1) & O(1) & O(1) & O(1) & O(N) \\ \vdots & & & \vdots & \vdots \\ O(1) & \dots & & O(1) & O(N) \\ O(N) & O(N) & O(N) & O(N) & O(N^2) \end{bmatrix}. \quad (38)$$

For $k = 1, \dots, K$

$$\mathbf{J}_{k,k} = \begin{bmatrix} O(N) & O(N) & O(N) & 0 & 0 \\ O(N) & O(N) & O(N) & 0 & 0 \\ O(N) & O(N) & O(N) & 0 & 0 \\ 0 & 0 & 0 & O(N) & 0 \\ 0 & 0 & 0 & 0 & O(N^3) \end{bmatrix}. \quad (39)$$

Define the 5×5 normalizing matrix $\mathbf{D}_N = \text{diag}(N^{1/2}, N^{1/2}, N^{1/2}, N^{1/2}, N^{3/2})$ and the $5K \times 5K$ normalizing matrix $\tilde{\mathbf{D}}_N = \text{diag}(\mathbf{D}_N, \dots, \mathbf{D}_N)$. All elements in the matrices $\{\mathbf{D}_N^{-1} \mathbf{J}_{k,l} \mathbf{D}_N^{-1}, \text{all } k \neq l\}$ are asymptotically $O(1/N)$, and hence, the matrix $\tilde{\mathbf{D}}_N^{-1} \mathbf{J}(\theta) \tilde{\mathbf{D}}_N^{-1}$ is asymptotically block diagonal. The diagonal of $\tilde{\mathbf{D}}_N^{-1} \mathbf{J}(\theta) \tilde{\mathbf{D}}_N^{-1}$ consists of the elements $N^{-1} J_{\varphi_k, \varphi_k}$, $N^{-1} J_{f_k, f_k}$ and the blocks

$$N^{-1} \mathbf{J}(b_k, \psi_k, \phi_k) = \frac{1}{N} \begin{bmatrix} J_{b_k, b_k} & J_{b_k, \psi_k} & J_{b_k, \phi_k} \\ J_{\psi_k, b_k} & J_{\psi_k, \psi_k} & J_{\psi_k, \phi_k} \\ J_{\phi_k, b_k} & J_{\phi_k, \psi_k} & J_{\phi_k, \phi_k} \end{bmatrix} \quad (40)$$

for $k = 1, \dots, K$ that have the order $O(1)$, whereas all other elements have the order $O(1/N)$. This means, in other words, that the parameters f_k , φ_k , and the triplets (b_k, ψ_k, ϕ_k) for $k = 1, \dots, K$ are asymptotically decoupled. CRBs for these parameters, which are defined as proper diagonal elements and diagonal blocks of \mathbf{J} , are approximately equal to the inverse of J_{φ_k, φ_k} , J_{f_k, f_k} , and $\mathbf{J}(b_k, \psi_k, \phi_k)$, respectively. The approximate CRB expression for f_k is proportional to N^{-3} , as is usual in the frequency estimation, and CRB expressions for the other parameters (including the angles of arrival) is proportional to N^{-1} .

It appears that asymptotic behavior of the CRB is very similar in the case of temporarily colored noise, which is modeled as AR(1) with parameter ρ . For $\rho \neq 0$

$$\begin{aligned} \mathbf{M}_\rho^{-1} \mathbf{s}_k &\approx \frac{1 + \rho^2}{1 - \rho^2} \mathbf{s}_k - \frac{\rho}{1 - \rho^2} (e^{j2\pi f_k} + e^{-j2\pi f_k}) \mathbf{s}_k \\ &\approx \frac{1 - 2\rho \cos(2\pi f_k) + \rho^2}{1 - \rho^2} \mathbf{s}_k. \end{aligned} \quad (41)$$

The approximation in (41) is exact for the entire vector, except for its first and the last elements; the first and the last elements on both sides differ by quantities of order $O(1)$. Because the vector has length N , the above approximation error is of order $O(1)$ in the vector's norm. It follows that $\mathbf{s}_k^H \mathbf{M}_\rho^{-1} \mathbf{s}_l$ and $\mathbf{s}_k^H \mathbf{M}_\rho^{-1} \tilde{\mathbf{s}}_l$, respectively, have the same asymptotic behavior for large N as $\mathbf{s}_k^H \mathbf{s}_l$ and $\mathbf{s}_k^H \tilde{\mathbf{s}}_l$ up to the constant multiplicative factors $c_\rho(f_k) = [1 - 2\rho \cos(2\pi f_k) + \rho^2]/(1 - \rho^2)$. Hence, for $\rho \neq 0$ as compared with $\rho = 0$, elements of the k th source's Fisher information matrix that correspond to angles of arrival are approximately $c_\rho(f_k)$ times as large and the corresponding CRB approximately $1/c_\rho(f_k)$ times as large.

The mean square angular estimation error (MSAE) [10], [26] represents an alternative estimation performance metric; its lower bound may be related to the CRB as follows:

$$\begin{aligned} \text{MSAE}_k^{CR} &\stackrel{\text{def}}{=} \left\| \frac{\partial \mathbf{a}_k}{\partial \psi_k} \right\|^2 \text{CRB}(\psi_k) + \left\| \frac{\partial \mathbf{a}_k}{\partial \phi_k} \right\|^2 \text{CRB}(\phi_k) \\ &= \text{CRB}(\psi_k) + \sin^2 \psi_k \text{CRB}(\phi_k). \end{aligned} \quad (42)$$

Here, $\text{CRB}(\psi_k)$ and $\text{CRB}(\phi_k)$ stand for the CRB for ψ_k and ϕ_k , respectively, and they can be found as the second and the third diagonal elements of $[\mathbf{J}(b_k, \psi_k, \phi_k)]^{-1}$.

The CRB of the Cartesian direction cosines may be derived using [14, Th. 3.4]: "Let $\mathbf{g}(\theta)$ be a differentiable function of θ and have the Jacobian $\mathbf{G}(\theta)$, and (2), and let $\mathbf{J}^{-1}(\theta)$ be the CRB for estimating θ . Then, the CRB for estimating $\mathbf{g}(\theta)$ is $\text{CRB}(\mathbf{g}(\theta)) = \mathbf{G}(\theta) \text{CRB}(\theta) \mathbf{G}^T(\theta) = \mathbf{G}(\theta) \mathbf{J}^{-1}(\theta) \mathbf{G}^T(\theta)$." For the Cartesian direction cosines, this means that $\mathbf{G}(\psi_k, \phi_k) = [\partial \mathbf{a}_k / \partial \psi_k, \partial \mathbf{a}_k / \partial \phi_k]$ and $\text{CRB}(\mathbf{a}_k) = \mathbf{G}(\psi_k, \phi_k) \text{CRB}(\psi_k, \phi_k) (\mathbf{G}(\psi_k, \phi_k))^H$, where $\text{CRB}(\psi_k, \phi_k)$ is a proper submatrix of $\text{CRB}(\theta)$.

B. For Spatio-Temporally Uncorrelated Noise

Hereafter in Section V, the additive noise is assumed to be spatially and temporarily uncorrelated, all velocity hydrophones to have equal noise variance σ_v^2 , and the pressure-hydrophone to have noise variance σ_p^2 . Then, $\mathbf{\Gamma} = \mathbf{\Gamma}_0 \otimes \mathbf{I}$, where $\mathbf{\Gamma}_0 = \text{diag}(\sigma_v^2, \sigma_v^2, \sigma_v^2, \sigma_p^2)$. The case of unequal noise variances at the velocity hydrophones has already been treated in the more general expressions in (36) and (37).

Given the above form of $\mathbf{\Gamma}_0$, $\mathbf{J}(\theta) = \mathbf{J}^v(\theta) + \mathbf{J}^p(\theta)$. $\mathbf{J}^v(\theta)$ is proportional to $1/\sigma_v^2$ and represents the information provided by the constituent velocity hydrophones. $\mathbf{J}^p(\theta)$ is proportional to $1/\sigma_p^2$ and represents the information provided by the pressure

$$\begin{aligned} \mathbf{s}_k^H \mathbf{s}_l &= \begin{cases} e^{j(\varphi_l - \varphi_k)} \frac{\sin(N\Delta_{k,l})}{\sin\Delta_{k,l}}, & \text{for } f_l \neq f_k \\ N e^{j(\varphi_l - \varphi_k)}, & \text{for } f_l = f_k \end{cases} \\ \mathbf{s}_k^H \tilde{\mathbf{s}}_l &= \begin{cases} j e^{j(\varphi_l - \varphi_k)} \left[\frac{N}{2} \frac{\cos(N\Delta_{k,l})}{\sin\Delta_{k,l}} - \frac{1}{2} \frac{\sin(N\Delta_{k,l}) \cos\Delta_{k,l}}{\sin^2\Delta_{k,l}} \right], & \text{for } f_l \neq f_k \\ 0, & \text{for } f_l = f_k \end{cases} \\ \mathbf{s}_k^H \tilde{\tilde{\mathbf{s}}}_l &= \begin{cases} e^{j(\varphi_l - \varphi_k)} \left[\frac{N^2}{4} \frac{\sin(N\Delta_{k,l})}{\sin\Delta_{k,l}} + \frac{N}{2} \frac{\cos(N\Delta_{k,l}) \cos\Delta_{k,l}}{\sin^2\Delta_{k,l}} - \frac{1}{4} \frac{\sin(N\Delta_{k,l})(1 + \cos^2\Delta_{k,l})}{\sin^3\Delta_{k,l}} \right], & \text{for } f_l \neq f_k \\ \frac{N(N^2 - 1)}{12} e^{j(\varphi_l - \varphi_k)}, & \text{for } f_l = f_k \end{cases} \end{aligned}$$

hydrophone, if present. For example, the upper-left blocks of these two matrices contain the elements

$$J_{b_k, b_l}^v = \frac{2}{\sigma_v^2} (\mathbf{a}'_k)^T \mathbf{a}'_l \operatorname{Re} [\mathbf{s}_k^H \mathbf{s}_l] \quad (43)$$

$$J_{b_k, b_l}^p = \frac{2}{\sigma_p^2} \operatorname{Re} [\mathbf{s}_k^H \mathbf{s}_l] \quad k, l = 1, \dots, K \quad (44)$$

where \mathbf{a}'_k contains the elements of \mathbf{a}_k corresponding to the velocity hydrophones present in the vector hydrophone. If all three velocity hydrophones are present, then \mathbf{a}'_k consists of the first three elements of \mathbf{a}_k .

For large N , $\mathbf{J}^v(\theta)$ and $\mathbf{J}^p(\theta)$ exhibit the same asymptotic behavior as $\mathbf{J}(\theta)$. For the case where the vector hydrophone contains all three velocity hydrophones with equal error variances, the three vectors \mathbf{a}_k , $\partial \mathbf{a}_k / \partial \psi_k$, and $\partial \mathbf{a}_k / \partial \phi_k$ are pairwise orthogonal. Hence, $\mathbf{J}_{k,k}$ and $\mathbf{J}(b_k, \psi_k, \phi_k)$ are diagonal matrices.

C. CRB for One Vector Hydrophone Consisting of Only Three Velocity Hydrophones

The diagonal elements of the nonasymptotic information matrix $\mathbf{J}(\theta) = \mathbf{J}^v(\theta)$ become

$$J_{b_k, b_k}^v = \frac{2}{\sigma_v^2} (\mathbf{a}'_k)^T \mathbf{a}'_k \operatorname{Re} [\mathbf{s}_k^H \mathbf{s}_k] = \frac{2N}{\sigma_v^2} \quad (45)$$

$$J_{\psi_k, \psi_k}^v = \frac{2}{\sigma_v^2} b_k^2 \left(\frac{\partial \mathbf{a}_k}{\partial \psi_k} \right)^T \left(\frac{\partial \mathbf{a}_k}{\partial \psi_k} \right) \operatorname{Re} [\mathbf{s}_k^H \mathbf{s}_k] = \frac{2N b_k^2}{\sigma_v^2} \quad (46)$$

$$J_{\phi_k, \phi_k}^v = \frac{2}{\sigma_v^2} b_k^2 \left(\frac{\partial \mathbf{a}_k}{\partial \phi_k} \right)^T \left(\frac{\partial \mathbf{a}_k}{\partial \phi_k} \right) \operatorname{Re} [\mathbf{s}_k^H \mathbf{s}_k] = \frac{2N b_k^2 \sin^2 \psi_k}{\sigma_v^2} \quad (47)$$

$$J_{\varphi_k, \varphi_k}^v = \frac{2}{\sigma_v^2} b_k^2 (\mathbf{a}'_k)^T \mathbf{a}'_k \operatorname{Re} [\mathbf{s}_k^H \mathbf{s}_k] = \frac{2N b_k^2}{\sigma_v^2} \quad (48)$$

$$J_{f_k, f_k}^v = \frac{8\pi^2}{\sigma_v^2} b_k^2 (\mathbf{a}'_k)^T \mathbf{a}'_k \operatorname{Re} [\hat{\mathbf{s}}_k^H \hat{\mathbf{s}}_k] = \frac{2\pi^2 b_k^2 N (N^2 - 1)}{3\sigma_v^2} \quad (49)$$

As previously shown, the asymptotic CRBs of the signal parameters (for large N and well-separated frequencies $2\pi|f_k - f_l| \gg 1/N$) are inversely related to the above elements. That is

$$\operatorname{CRB}(b_k) = \frac{\sigma_v^2}{2N} + O(N^{-2}) \quad (50)$$

$$\operatorname{CRB}(\psi_k) = \frac{\sigma_v^2}{2N b_k^2} + O(N^{-2}) \quad (51)$$

$$\operatorname{CRB}(\phi_k) = \frac{\sigma_v^2}{2N b_k^2 \sin^2 \psi_k} + O(N^{-2}) \quad (52)$$

$$\operatorname{CRB}(\varphi_k) = \frac{\sigma_v^2}{2N b_k^2} + O(N^{-2}) \quad (53)$$

$$\operatorname{CRB}(f_k) = \frac{3\sigma_v^2}{2\pi^2 N^3 b_k^2} + O(N^{-4}). \quad (54)$$

Noting that

$$\operatorname{CRB}(\psi_k, \phi_k) \approx \frac{\sigma_v^2}{2N b_k^2} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\sin^2 \psi_k} \end{bmatrix} \quad (55)$$

the asymptotic CRBs for the Cartesian directional cosines equal (56), shown at the bottom of the page. The asymptotic lower bound for the mean square angular error becomes

$$\operatorname{MSAE}_k^{CR} = \operatorname{CRB}(\psi_k) + \sin^2 \psi_k \operatorname{CRB}(\phi_k) = \frac{\sigma_v^2}{N b_k^2} \quad (57)$$

D. CRB for One Vector Hydrophone Consisting of Three Velocity Hydrophones Plus a Pressure Hydrophone

The Fisher information matrix $\mathbf{J}_p(\theta)$ for the pressure hydrophone data has the (nonasymptotic) diagonal elements

$$J_{b_k, b_k}^p = \frac{2}{\sigma_p^2} \operatorname{Re} [\mathbf{s}_k^H \mathbf{s}_k] = \frac{2N}{\sigma_p^2} \quad (58)$$

$$J_{\psi_k, \psi_k}^p = J_{\phi_k, \phi_k}^p = 0 \quad (59)$$

$$J_{\varphi_k, \varphi_k}^p = \frac{2}{\sigma_p^2} b_k^2 \operatorname{Re} [\mathbf{s}_k^H \mathbf{s}_k] = \frac{2N b_k^2}{\sigma_p^2} \quad (60)$$

$$J_{f_k, f_k}^p = \frac{8\pi^2}{\sigma_p^2} b_k^2 \operatorname{Re} [\hat{\mathbf{s}}_k^H \hat{\mathbf{s}}_k] = \frac{2\pi^2 N (N^2 - 1) b_k^2}{3\sigma_p^2} \quad (61)$$

In the single-source case (i.e., $K = 1$), the estimated parameters are decoupled, and measurements from the pressure hydrophone bring no useful information on the arrival angle. In the case of multiple sources (i.e., $K > 1$) and a large N , the preceding statement becomes approximately true. Again, the asymptotic CRBs of b_k , ψ_k , ϕ_k , φ_k , and f_k are the inverses of the corresponding asymptotic diagonal elements of $\mathbf{J}(\theta) = \mathbf{J}^v(\theta) + \mathbf{J}^p(\theta)$

$$\operatorname{CRB}(b_k) = \frac{\sigma_v^2 \sigma_p^2}{2N(\sigma_v^2 + \sigma_p^2)} + O(N^{-2}) \quad (62)$$

$$\operatorname{CRB}(\psi_k) = \frac{\sigma_v^2}{2N b_k^2} + O(N^{-2}) \quad (63)$$

$$\operatorname{CRB}(\phi_k) = \frac{\sigma_v^2}{2N b_k^2 \sin^2 \psi_k} + O(N^{-2}) \quad (64)$$

$$\begin{aligned} \operatorname{CRB}(u_k, v_k) &= \mathbf{G}(\psi_k, \phi_k) \operatorname{CRB}(\psi_k, \phi_k) \mathbf{G}^T(\psi_k, \phi_k) \approx \frac{\sigma_v^2}{2N b_k^2} \\ &\times \begin{bmatrix} \cos^2 \psi_k \cos^2 \phi_k + \sin^2 \phi_k & -\sin^2 \psi_k \cos \phi_k \sin \phi_k \\ -\sin^2 \psi_k \cos \phi_k \sin \phi_k & \cos^2 \psi_k \sin^2 \phi_k + \cos^2 \phi_k \end{bmatrix} \\ &\approx \frac{\sigma_v^2}{2N b_k^2} \begin{bmatrix} 1 - u_k^2 & -u_k v_k \\ -u_k v_k & 1 - v_k^2 \end{bmatrix} \end{aligned} \quad (56)$$

$$\text{CRB}(\varphi_k) = \frac{\sigma_v^2 \sigma_p^2}{2N b_k^2 (\sigma_v^2 + \sigma_p^2)} + O(N^{-2}) \quad (65)$$

$$\text{CRB}(f_k) = \frac{3\sigma_v^2 \sigma_p^2}{2\pi^2 N^3 b_k^2 (\sigma_v^2 + \sigma_p^2)} + O(N^{-4}). \quad (66)$$

Note that the asymptotic CRBs for ψ_k and ϕ_k do not depend on σ_p : the pressure-hydrophone data do not improve the estimation of ψ_k and ϕ_k obtained from the three velocity hydrophones' data. Moreover, as σ_p approaches infinity, the asymptotic CRBs of b_k , φ_k , and f_k converge to those for the three-velocity hydrophone vector hydrophone. Hence, the asymptotic CRB(u_k, v_k) and the asymptotic MSAE $_k^{CR}$ remain the same when the pressure hydrophone is absent.

E. Asymptotic CRB for One Vector Hydrophone Consisting of Two Velocity Hydrophones Plus a Pressure Hydrophone

In contrast to earlier cases where all three velocity hydrophones are present, the information from the pressure hydrophone now becomes necessary for the asymptotic (as N approaches infinity) estimation of the arrival angles. The estimates of ψ_k and ϕ_k are still asymptotically decoupled from the estimates of φ_k and f_k but not necessarily from the estimate of the signal amplitude b_k .

The asymptotic Fisher information matrix of the triad (b_k, ψ_k, ϕ_k) corresponding to the velocity hydrophones has the general form in (67), shown at the bottom of the page, where \mathbf{a}'_k is composed of those elements of \mathbf{a}_k that are present in the vector hydrophone construction.

If the x-axis velocity hydrophone is absent, we have (68), shown at the bottom of the page. Because $\mathbf{J}_x^v(b_k, \psi_k, \phi_k)$ is non-invertible, the asymptotic CRBs (and, consequently, the variance) of unbiased estimators are infinite if the information from

the pressure-hydrophone is absent. The asymptotic information matrix corresponding to the pressure-hydrophone equals

$$\mathbf{J}^p(b_k, \psi_k, \phi_k) = \frac{2N}{\sigma_p^2} \text{diag}(1, 0, 0).$$

Hence, we have (69), shown at the bottom of the page. Note that $\text{CRB}_x(\psi_k) = [\text{CRB}_x(b_k, \psi_k, \phi_k)]_{(2,2)}$ and $\text{CRB}_x(\phi_k) = [\text{CRB}_x(b_k, \psi_k, \phi_k)]_{(3,3)}$.

Similarly, if the y-axis velocity hydrophone is absent, we have (70), shown at the bottom of the page. As the y-axis velocity hydrophone, instead of the x-axis velocity hydrophone is now absent, the asymptotic CRB $_y(b_k, \psi_k, \phi_k)$ may be obtained from the asymptotic CRB $_x(b_k, \psi_k, \phi_k)$ by simply interchanging $\sin^2 \psi_k$ and $\cos^2 \psi_k$. Moreover, $\text{CRB}_x(\psi_k) = \text{CRB}_y(\psi_k)$.

If the z-axis velocity hydrophone is absent

$$\text{CRB}_z(b_k, \psi_k, \phi_k) = \frac{\sigma_p^2}{2N b_k^2} \cdot \begin{bmatrix} b_k^2 & -b_k \tan \psi_k & 0 \\ -b_k \tan \psi_k & \frac{\sin^2 \psi_k + \frac{\sigma_p^2}{\sigma_v^2}}{\cos^2 \psi_k} & 0 \\ 0 & 0 & \frac{\frac{\sigma_v^2}{\sigma_p^2}}{\sin^2 \psi_k} \end{bmatrix}. \quad (71)$$

The asymptotic lower bounds for the MSAEs of these three constructions equal, respectively

$$\begin{aligned} \text{MSAE}_x^{CR} &= \text{CRB}_x(\psi_k) + \sin^2 \psi_k \text{CRB}_x(\phi_k) \\ &= \frac{\sigma_v^2}{2N b_k^2} \left(\frac{\sigma_v^2}{\sigma_p^2} - 1 + \frac{\frac{\sigma_v^2}{\sigma_p^2} + 1}{\sin^2 \psi_k \cos^2 \phi_k} \right) \end{aligned} \quad (72)$$

$$\mathbf{J}^v(b_k, \psi_k, \phi_k) = \frac{2N}{\sigma_v^2} \begin{bmatrix} \mathbf{a}'_k{}^T \mathbf{a}'_k & b_k \mathbf{a}'_k{}^T \left(\frac{\partial \mathbf{a}'_k}{\partial \psi_k} \right) & b_k \mathbf{a}'_k{}^T \left(\frac{\partial \mathbf{a}'_k}{\partial \phi_k} \right) \\ b_k \left(\frac{\partial \mathbf{a}'_k}{\partial \psi_k} \right)^T \mathbf{a}'_k & b_k^2 \left(\frac{\partial \mathbf{a}'_k}{\partial \psi_k} \right)^T \left(\frac{\partial \mathbf{a}'_k}{\partial \psi_k} \right) & b_k^2 \left(\frac{\partial \mathbf{a}'_k}{\partial \psi_k} \right)^T \left(\frac{\partial \mathbf{a}'_k}{\partial \phi_k} \right) \\ b_k \left(\frac{\partial \mathbf{a}'_k}{\partial \phi_k} \right)^T \mathbf{a}'_k & b_k^2 \left(\frac{\partial \mathbf{a}'_k}{\partial \phi_k} \right)^T \left(\frac{\partial \mathbf{a}'_k}{\partial \psi_k} \right) & b_k^2 \left(\frac{\partial \mathbf{a}'_k}{\partial \phi_k} \right)^T \left(\frac{\partial \mathbf{a}'_k}{\partial \phi_k} \right) \end{bmatrix} \quad (67)$$

$$\mathbf{J}_x^v(b_k, \psi_k, \phi_k) = \frac{2N b_k^2}{\sigma_v^2} \begin{bmatrix} b_k^{-2} (\sin^2 \psi_k \sin^2 \phi_k + \cos^2 \psi_k) & -b_k^{-1} \sin \psi_k \cos \psi_k \cos^2 \phi_k & b_k^{-1} \sin^2 \psi_k \sin \phi_k \cos \phi_k \\ -b_k^{-1} \sin \psi_k \cos \psi_k \cos^2 \phi_k & \cos^2 \psi_k \sin^2 \phi_k + \sin^2 \psi_k & \sin \psi_k \cos \psi_k \sin \phi_k \cos \phi_k \\ b_k^{-1} \sin^2 \psi_k \sin \phi_k \cos \phi_k & \sin \psi_k \cos \psi_k \sin \phi_k \cos \phi_k & \sin \psi_k \cos^2 \phi_k \end{bmatrix} \quad (68)$$

$$\begin{aligned} \text{CRB}_x(b_k, \psi_k, \phi_k) &= [\mathbf{J}_x^v(b_k, \psi_k, \phi_k) + \mathbf{J}_p(b_k, \psi_k, \phi_k)]^{-1} \\ &= \frac{\sigma_p^2}{2N b_k^2} \begin{bmatrix} b_k^2 & b_k \cot \psi_k & -b_k \frac{\tan \phi_k}{\sin^2 \psi_k} \\ b_k \cot \psi_k & \frac{\cos^2 \psi_k + \frac{\sigma_p^2}{\sigma_v^2}}{\sin^2 \psi_k} & -\frac{(\frac{\sigma_p^2}{\sigma_v^2} + 1) \cos \psi_k \tan \phi_k}{\sin^3 \psi_k} \\ -b_k \frac{\tan \phi_k}{\sin^2 \psi_k} & -\frac{(\frac{\sigma_p^2}{\sigma_v^2} + 1) \cos \psi_k \tan \phi_k}{\sin^3 \psi_k} & \frac{\frac{\sigma_p^2}{\sigma_v^2} + \sin^2 \phi_k - \frac{\sigma_p^2}{\sigma_v^2} \cos^2 \psi_k \cos^2 \phi_k}{\sin^4 \psi_k \cos^2 \phi_k} \end{bmatrix} \end{aligned} \quad (69)$$

$$\begin{aligned} \text{CRB}_y(b_k, \psi_k, \phi_k) &= \frac{\sigma_p^2}{2N b_k^2} \begin{bmatrix} b_k^2 & b_k \cot \psi_k & b_k \frac{\cot \phi_k}{\sin^2 \psi_k} \\ b_k \cot \psi_k & \frac{\cos^2 \psi_k + \frac{\sigma_p^2}{\sigma_v^2}}{\sin^2 \psi_k} & \frac{(\frac{\sigma_p^2}{\sigma_v^2} + 1) \cos \psi_k \cot \phi_k}{\sin^3 \psi_k} \\ b_k \frac{\cot \phi_k}{\sin^2 \psi_k} & \frac{(\frac{\sigma_p^2}{\sigma_v^2} + 1) \cos \psi_k \cot \phi_k}{\sin^3 \psi_k} & \frac{\frac{\sigma_p^2}{\sigma_v^2} + \cos^2 \phi_k - \frac{\sigma_p^2}{\sigma_v^2} \cos^2 \psi_k \sin^2 \phi_k}{\sin^4 \psi_k \sin^2 \phi_k} \end{bmatrix} \end{aligned} \quad (70)$$

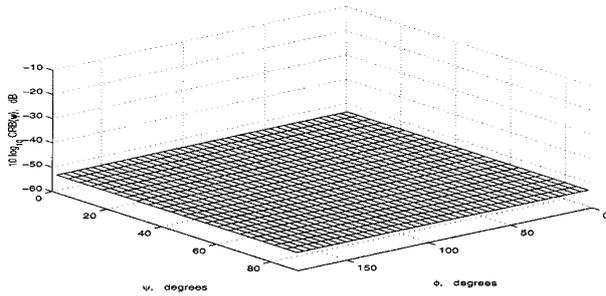


Fig. 1. Asymptotic CRB(ψ) with all three velocity hydrophones with or without a pressure hydrophone.

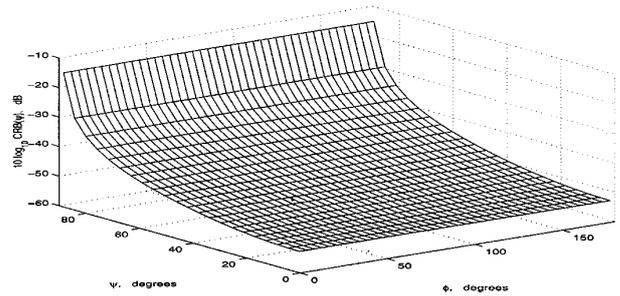


Fig. 4. Asymptotic CRB(ψ) without the z-axis velocity hydrophone but with a pressure hydrophone.

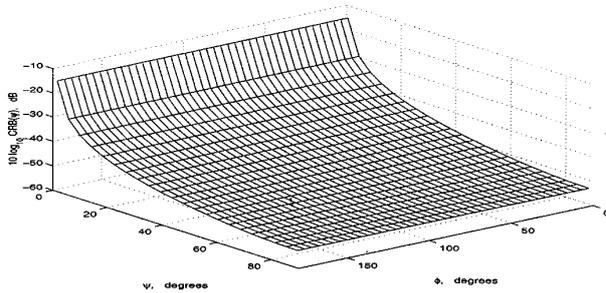


Fig. 2. Asymptotic CRB(ψ) without the x-axis velocity hydrophone but with a pressure hydrophone.

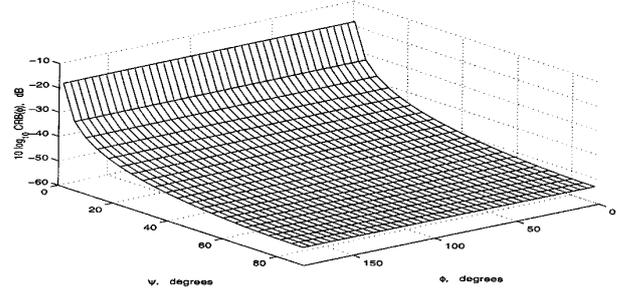


Fig. 5. Asymptotic CRB(ϕ) with all three velocity hydrophones with or without a pressure hydrophone.

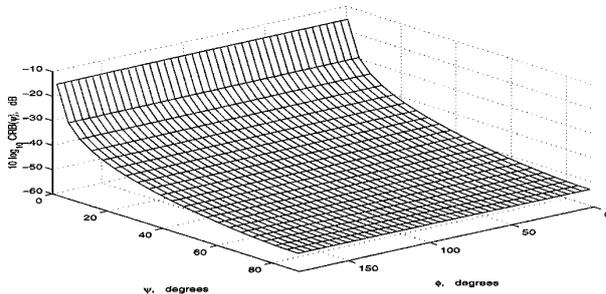


Fig. 3. Asymptotic CRB(ψ) without the y-axis velocity hydrophone but with a pressure hydrophone.

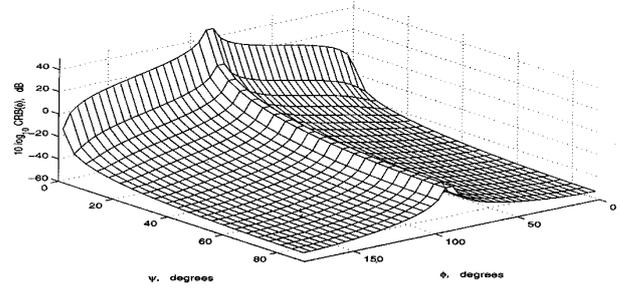


Fig. 6. Asymptotic CRB(ϕ) without the x-axis velocity hydrophone but with a pressure hydrophone.

$$\begin{aligned} \text{MSAE}_y^{CR} &= \text{CRB}_y(\psi_k) + \sin^2 \psi_k \text{CRB}_y(\phi_k) \\ &= \frac{\sigma_v^2}{2N b_k^2} \left(\frac{\sigma_v^2}{\sigma_p^2} - 1 + \frac{\frac{\sigma_v^2}{\sigma_p^2} + 1}{\sin^2 \psi_k \sin^2 \phi_k} \right) \end{aligned} \quad (73)$$

$$\begin{aligned} \text{MSAE}_z^{CR} &= \text{CRB}_z(\psi_k) + \sin^2 \psi_k \text{CRB}_z(\phi_k) \\ &= \frac{\sigma_v^2}{2N b_k^2} \left(\frac{\sigma_v^2}{\sigma_p^2} - 1 + \frac{\frac{\sigma_v^2}{\sigma_p^2} + 1}{\cos^2 \psi_k} \right). \end{aligned} \quad (74)$$

Figs. 1–4 plot, in decibels of radians, the asymptotic CRB(ψ_k), respectively, for all five aforementioned vector hydrophone constructions, and Figs. 5–8 plot the corresponding asymptotic CRB(ϕ_k). Note that the horizontal axes in Fig. 4 are oriented differently from the other seven figures. The additive noise is spatio-temporally uncorrelated.

The following qualitative trends may be observed from these eight figures.

- 1) ϕ_k does not affect the asymptotic CRB(ψ_k) for any vector hydrophone construction, as ϕ_k does not affect the fraction of signal energy distributed on the x-plane or parallel to the vertical axis.
- 2) ϕ_k does not affect the asymptotic CRB(ϕ_k) for those vector hydrophone constructions with both an x-axis and a y-axis velocity hydrophone, for the same reason given above. However, when one of the two horizontal velocity hydrophones is absent, the asymptotic CRB(ϕ_k) increases significantly when ϕ_k becomes aligned along the Cartesian coordinate axis without a velocity hydrophone. In such a case, the horizontal component of the incident signal's energy become unobservable by the vector hydrophone. Hence, the estimation of the azimuth angle becomes impossible, as indicated by the very large asymptotic CRB(ϕ_k) value for such cases in Figs. 6 and 7. Note that Fig. 6 is identical with Fig. 7 when ϕ_k is shifted by 90° .

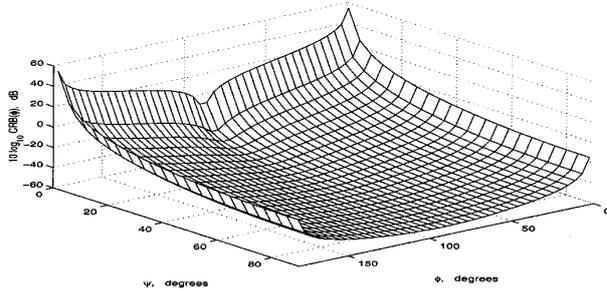


Fig. 7. Asymptotic CRB(ϕ) without the y-axis velocity hydrophone but with a pressure hydrophone.

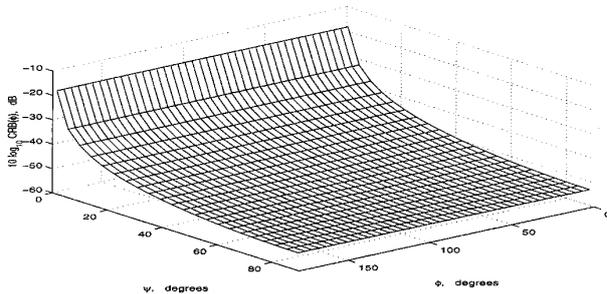


Fig. 8. Asymptotic CRB(ϕ) without the z-axis velocity hydrophone but with a pressure hydrophone.

- 3) The asymptotic CRB(ϕ_k) increases as ψ_k decreases for all five vector hydrophone constructions. This is because a smaller ψ_k implies a small fraction of signal energy lies on the x-y plane, thereby decreasing the signal-to-noise ratio on the horizontal plane.
- 4) The asymptotic CRB(ψ_k) is independent of ψ_k when all three velocity hydrophones are present, with or without the pressure hydrophone. This is because the elevation angle information is encapsulated in the received data in two complementary ways:
 - a) the signal's energy along the vertical axis;
 - b) the signal's energy along the x-y plane to be subtracted from the signal's overall energy as measured by the pressure hydrophone or by all three velocity hydrophones.
- 5) The asymptotic CRB(ψ_k) increases as ψ_k decreases when either of the two horizontal velocity hydrophones is absent. In such cases, the elevation angle information is encapsulated only as in (a) through the nonlinear trigonometric function $\cos \psi_k$, which has a flat slope with respect to ψ_k when ψ_k approaches zero.
- 6) The asymptotic CRB(ψ_k) increases as ψ_k increases toward 90° when the vertical velocity hydrophone is absent. In such a case, the elevation angle information is encapsulated only as in (b) through the function $1 - \sin^2 \psi_k \sin^2 \phi_k - \sin^2 \psi_k \sin^2 \phi_k$, which must lie between 0 and 1. As ψ_k increases toward 90° , the $\sin^2 \psi$ in $1 - \sin^2 \psi_k \sin^2 \phi_k - \sin^2 \psi_k \sin^2 \phi_k$ has flatter slopes. This means that $1 - \sin^2 \psi_k \sin^2 \phi_k - \sin^2 \psi_k \sin^2 \phi_k$ becomes less accurately estimated as ψ_k increases toward 90° .

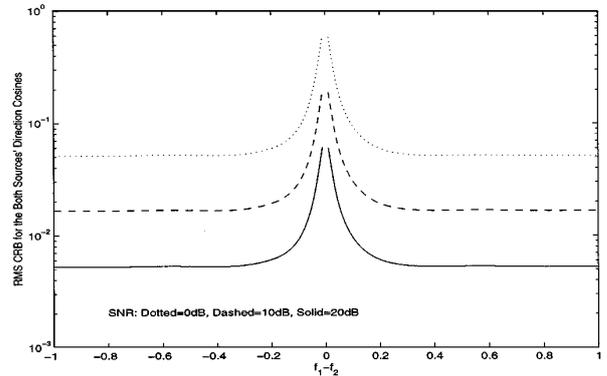


Fig. 9. RMS nonasymptotic CRB for the direction cosines versus the separation in the incident sources' digital frequencies: two closely spaced sources with $\{\psi_1, \psi_2\} = \{75^\circ, 80^\circ\}$, $\{\phi_1, \phi_2\} = \{35^\circ, 30^\circ\}$, $b_1 = b_2 = 1$, $\Delta_T = 0.1$, and 100 snapshots with uniform sampling frequency at 10 in each of 200 independent experiments.

Figs. 1–8 together suggest that the three-velocity hydrophone construction gives the best asymptotic CRBs and requires the least number of constituent hydrophones. However, when only the azimuth angle needs to be estimated, a comparable asymptotic CRB(ψ_k) may be obtained using the vector hydrophone construction with the two horizontal velocity hydrophones plus a pressure-hydrophone, with the advantage that the measured data will better conform to a rectilinear model for the oceanic dynamics. On the other hand, only the four hydrophone vector hydrophones can handle up to four incident sources using the proposed algorithm.

Fig. 9 shows the CRBs dependence on the incident sources' frequency separation in a two-source scenario at various SNRs. The *asymptotic* CRB cannot be used here because the signal frequencies are not well separated. Instead, the CRB is computed by inverting the exact information matrix in (28). Moreover, because this computation depends on the signals' initial phases [which appear in $\mathbf{s}_k^T \mathbf{s}_l$ in (37)], the data in Fig. 9 are averaged from 200 independent trials, with the initial phases uniformly distributed in $(0, 2\pi]$. The CRB for the four direction-cosine estimates are combined in the root-mean-square (RMS) CRB by taking the square root of the sum of the squares of the four individual CRBs. The additive noise is spatio-temporally uncorrelated.

VI. MONTE CARLO SIMULATIONS OF PROPOSED ALGORITHM

Fig. 10 plots the root-mean-square (RMS) standard deviations, and Fig. 11 plots the RMS biases of *univector hydrophone ESPRITs* direction-cosine estimates of Monte Carlo simulations in a scenario involving three uncorrelated monochromatic sources impinging on a single four-component vector hydrophone. A source's root-mean-square estimation standard deviation is the square root of the mean of square of the estimation standard deviations for that source's x-axis and y-axis direction cosines. A source's RMS estimation bias is similarly defined. The incident source's parameters are given in Fig. 9's caption. Their direction-cosines equal $u_1 = 0.47$, $u_2 = 0.67$, $u_3 = 0.87$, $v_1 = 0.88$, $v_2 = 0.68$, and $v_3 = 0.48$. The additive noise is spatio-temporally uncorrelated and

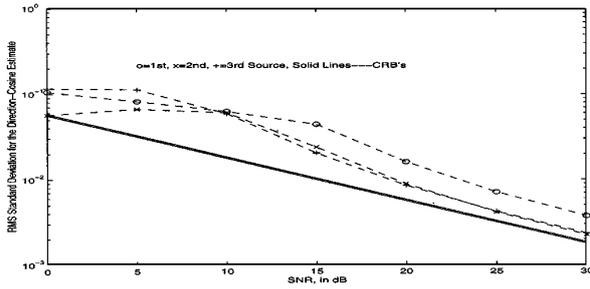


Fig. 10. *Univector hydrophone ESPRITs* RMS estimation standard deviations $\{\hat{u}_k, \hat{v}_k, k = 1, 2, 3\}$ at various SNRs. Three uncorrelated monochromatic sources $\{\psi_1, \psi_2, \psi_3\} = \{86.1^\circ, 72.7^\circ, 83.5^\circ\}$, $\{\phi_1, \phi_2, \phi_3\} = \{61.9^\circ, 45.4^\circ, 28.9^\circ\}$, and $b_1 = b_2 = b_3 = 1$, baseband digital frequencies $\{f_1, f_2, f_3\} = \{.55, .95, .15\}$ impinge on a vector hydrophone, SNR is relative to unity signal power, and $\Delta_T = 0.1$. Eighty snapshots with uniform sampling frequency of 10 in each of 500 independent experiments.

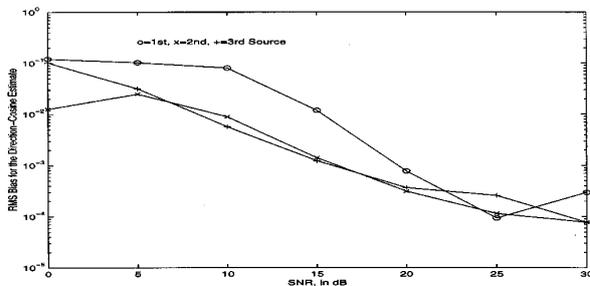


Fig. 11. *Univector hydrophone ESPRITs* RMS estimation bias of $\{\hat{u}_k, \hat{v}_k, k = 1, 2, 3\}$ at various SNRs. Same settings same as in Fig. 10.

complex Gaussian; the SNR is defined relative to each source. With the smallest difference among the u_k s and among the v_k s being 0.20, the proposed algorithm successfully resolves all three sources with high probability at SNRs above 0 dB. The nonasymptotic RMS CRB is also plotted in Fig. 10. Because the nonasymptotic CRB depends on the sources' temporal phases $\{\varphi_k, k = 1, 2, 3\}$, the CRBs plotted in Fig. 10 are the average of 500 Monte Carlo runs, each with its statistically independent $\{\varphi_k, k = 1, 2, 3\}$. Note that the estimation biases are more than an order of magnitude less than the standard deviations, thus supporting earlier claims that this algorithm yields asymptotically unbiased estimates.

Fig. 12 compares the performance of *univector hydrophone ESPRIT* (using a three-velocity hydrophone vector hydrophone) with that of *ESPRIT* using an L-shaped half-wavelength spaced⁵ array of variable number of pressure-hydrophones, half of which lie along the x-axis, and the other half along the y-axis. The signal scenario involves two uncorrelated monochromatic sources with spatio-temporally uncorrelated complex-Gaussian additive noise, as specified in the figure's caption. The two sets of Cartesian direction cosines are matched using *a priori* source information. That information is, of course, not available *a*

⁵The half-wavelength spacing is with respect to the transmission frequency, which is herein assumed to greatly exceed any difference in baseband digital frequencies between the incident sources.

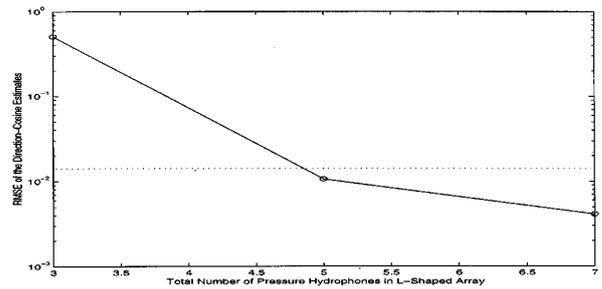


Fig. 12. *Univector hydrophone ESPRIT* versus *ESPRIT* on a half-wavelength spaced L-shaped array of pressure-hydrophones. Two incident sources with $u_1 = 0.2, u_2 = 0.1, v_1 = 0.1, v_2 = 0.2$, baseband digital frequencies $f_1 = 0.60, f_2 = 0.95, SNR = 20$ dB. SNR is relative to unity signal power $\Delta_T = 0.1$. One hundred snapshots with uniform sampling frequency at 10 in each of 500 independent experiments.

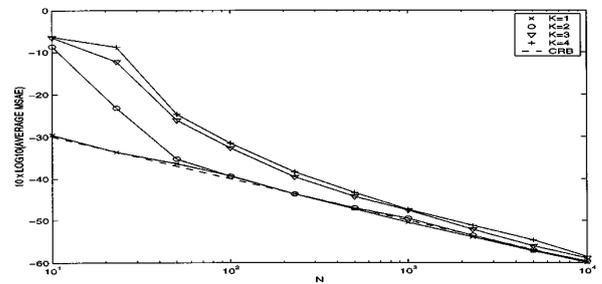


Fig. 13. CRB and proposed algorithm's MSAEs with one to four statistically independent sinusoidal sources in spatio-temporally uncorrelated noise, plotted against N .

priori; the pressure hydrophone scheme thus enjoys an advantage on this point over the proposed scheme.⁶ The vertical axis in Fig. 12 plots the root-mean-square error (RMSE) of each of the direction finding approaches. The RMSE is defined as the square root of the sum of the estimation variance and the square of the estimation bias. To match the proposed scheme's performance, the spatially displaced pressure-hydrophone array needs five elements. Fig. 12 also clearly shows that the spatially displaced pressure hydrophone array fails to resolve these two sources if limited to the same number of hydrophone elements (i.e., three) as in the proposed scheme.

Fig. 13 plots the one-source, two-source, three-source, and four-source MSAEs and the CRB versus N . The Cartesian direction cosines of the four equal-power sources (when all present) are $\{0.87, 0.32, 0.67, 0.5\}$ along the x-axis and $\{0.48, 0.81, 0.68, 0.5\}$ along the y-axis, with corresponding digital frequencies at $\{0.15, 0.35, 0.68, 0.95\}$. The additive noise is spatio-temporally uncorrelated with 20 dB SNR. The raw data, which is regularly time sampled at the vector hydrophone, are segmented into two temporally overlapping subsets, with Δ_T equal to ten sampling periods for $N > 10$ but

⁶This comparison between a vector hydrophone and a pressure hydrophone array needs to be understood by recognizing that a velocity hydrophone involves substantially more complex hardware than even a pair of pressure hydrophones. The robust and accurate suspension of a velocity hydrophone necessitates an intricate mechanical support mechanism. Co-locating two or more orthogonally oriented velocity hydrophones presents a very challenging hardware design problem.

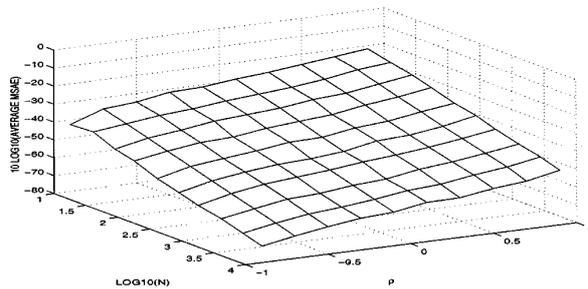


Fig. 14. Proposed algorithm's MSAE with one sinusoidal source in AR(1) temporarily correlated additive noise plotted against ρ and N .

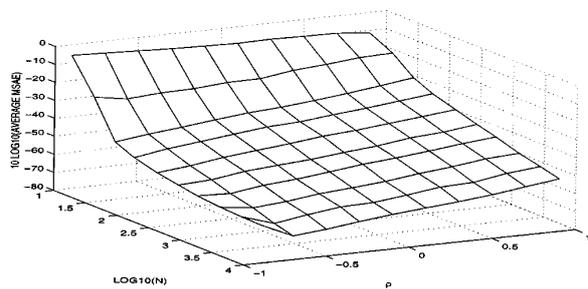


Fig. 15. Proposed algorithm's MSAE with two statistically independent sinusoidal sources in AR(1) temporarily correlated additive noise plotted against ρ and N .

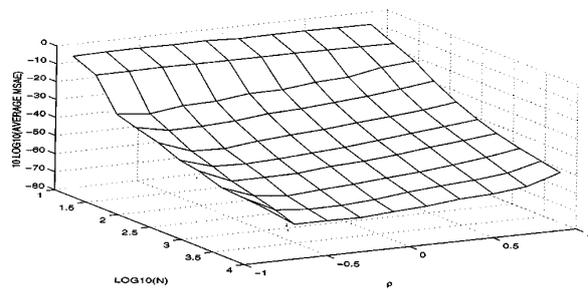


Fig. 16. Proposed algorithm's MSAE with three statistically independent sinusoidal sources in AR(1) temporarily correlated additive noise plotted against ρ and N .

one sampling period for $N = 10$. There exist 100 independent Monte Carlo runs for each MSAE data point. The MSAE of the proposed algorithm is near optimum (very close to the CRB) for all N in the single-source scenario. As the number of sources increase, the proposed algorithm approaches optimality as N increases.

Figs. 14–18 investigate the proposed algorithm's performance when the additive noise has AR(1) temporal correlation. The proposed algorithm's actual MSAEs are plotted in Figs. 14–17 versus N and ρ , respectively, for the cases involving one, two, three, or four statistically independent sinusoidal sources. The corresponding approximate CRB (i.e., the CRB for $\rho = 0$ multiplied by $1/c_p(f_k)$ with $c_p(f_k)$ defined in (41)), which is essentially identical for $K = 1, \dots, 4$, is plotted in Fig. 18. Referring to Fig. 18, the CRB for the AR(1) noise is found to be increasing function of ρ . The proposed

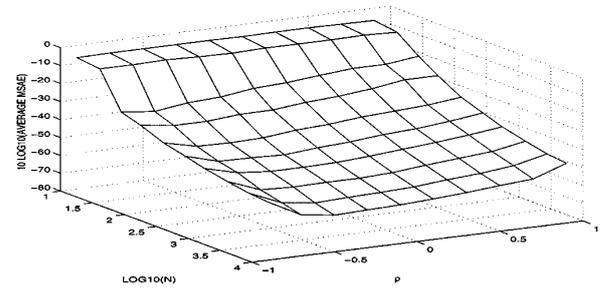


Fig. 17. Proposed algorithm's MSAE with four statistically independent sinusoidal sources in AR(1) temporarily correlated additive noise plotted against ρ and N .

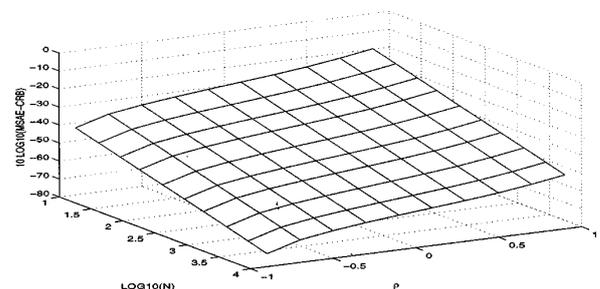


Fig. 18. CRB corresponding to the MSAE's in Figs. 14–17.

algorithm (see Figs. 14–17) very closely approximates the CRB (see Fig. 18) in the single-source case for all ρ and N . For multiple sources, the proposed algorithm is still very close to the CRB if the noise is only moderately colored with ρ near zero (cf. Fig. 13). The difference between the actual MSAE and the CRB is more apparent for highly correlated noise, for negative ρ ,⁷ and for cases with more sources.

REFERENCES

- [1] C. B. Leslie, J. M. Kendall, and J. L. Jones, "Hydrophone for measuring particle velocity," *J. Acoust. Soc. Amer.*, vol. 28, no. 4, pp. 711–715, July 1956.
- [2] C. L. LeBlanc, "Handbook of Hydrophone Element Design Technology," Newport, RI, Naval Underwater Syst. Cent. Tech. Rep. 5813, 1978.
- [3] M. A. Jossierand and C. Maerfeld, "PVF2 velocity hydrophone," *J. Acoust. Soc. Amer.*, vol. 78, no. 3, pp. 860–867, 1985.
- [4] V. J. Hughes, J. G. Boulton, J. M. Coles, T. R. Empson, and N. J. Kerry, "Why an optically-based hydrophone works better," *Sensor Rev.*, vol. 7, no. 3, pp. 123–126, July 1987.
- [5] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 984–995, July 1989.
- [6] G. L. D'Spain, W. S. Hodgkiss, and G. L. Edmonds, "The simultaneous measurement of infrasonic acoustic particle velocity and acoustic pressure in the ocean by freely drifting swallow floats," *IEEE J. Oceanic Eng.*, vol. 16, pp. 195–207, Apr. 1991.
- [7] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Trans. Signal Processing*, vol. 39, pp. 1122–1135, May 1991.
- [8] J. C. Nickles, G. Edmonds, R. Harriss, F. Fisher, W. S. Hodgkiss, J. Giles, and G. D'Spain, "A vertical array of directional acoustic sensors," in *Proc. IEEE Oceans Conf.*, 1992, pp. 340–345.

⁷A negative ρ would mean that much of the noise power resides in the high-frequency band near the Nyquist frequency. Indeed, any spectral peak in the noise may be mistaken as a signal frequency.

- [9] G. L. D'Spain, W. S. Hodgkiss, G. L. Edmonds, J. C. Nickles, F. H. Fisher, and R. A. Harris, "Initial analysis of the data from the vertical DIFAR array," in *Proc. IEEE Oceans Conf.*, 1992, pp. 346–351.
- [10] A. Nehorai and E. Paldi, "Acoustic vector sensor array processing," *IEEE Trans. Signal Processing*, vol. 42, pp. 2481–2491, Sept. 1994.
- [11] H. Cox and R. M. Zeskind, "Adaptive cardioid processing," in *Proc. 26th Asilomar Conf.*, 1992, pp. 1058–1061.
- [12] V. A. Shchurov, V. I. Ilyichev, and V. P. Kuleshov, "Ambient noise energy motion in the near surface layer in ocean wave-guide," *J. Phys.*, pt. 2, vol. 4, no. 5, pp. 1273–1276, May 1994.
- [13] B. Champagne, "Adaptive eigendecomposition of data covariance matrices based on first-order perturbations," *IEEE Trans. Signal Processing*, vol. 42, pp. 2758–2770, Oct. 1994.
- [14] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [15] B. Hochwald and A. Nehorai, "Identifiability in array processing models with vector-sensor applications," *IEEE Trans. Signal Processing*, vol. 44, pp. 83–95, Jan. 1996.
- [16] K. T. Wong and M. D. Zoltowski, "Uni-vector-sensor ESPRIT for multi-source azimuth-elevation angle-estimation," *IEEE Trans. Antennas Propagat.*, vol. 45, pp. 1467–1474, Oct. 1997.
- [17] —, "Orthogonal velocity hydrophone ESPRIT for sonar source localization," in *Proc. MTS/IEEE Oceans Conf.*, vol. 3, 1996, pp. 1307–1312.
- [18] M. J. Berliner and J. F. Lindberg, *Acoustic Particle Velocity Sensors: Design, Performance and Applications*. Woodbury, NY: AIP, 1996.
- [19] K. T. Wong and M. D. Zoltowski, "Closed-form underwater acoustic direction-finding with arbitrarily spaced vector hydrophones at unknown locations," *IEEE J. Oceanic Eng.*, vol. 22, pp. 566–575, July 1997.
- [20] —, "Extended-aperture underwater acoustic multi-source azimuth/elevation direction-finding using uniformly but sparsely spaced vector hydrophones," *IEEE J. Oceanic Eng.*, vol. 22, pp. 659–672, Oct. 1997.
- [21] M. Hawkes and A. Nehorai, "Acoustic vector-sensor beamforming and Capon direction estimation," *IEEE Trans. Signal Processing*, vol. 46, pp. 2291–2304, Sept. 1998.
- [22] —, "Effects of sensor placement on acoustic vector-sensor array performance," *IEEE J. Oceanic Eng.*, vol. 24, pp. 33–40, Jan. 1999.
- [23] K. T. Wong and M. D. Zoltowski, "Root-MUSIC-based azimuth-elevation angle-of-arrival estimation with uniformly spaced but arbitrarily oriented velocity hydrophones," *IEEE Trans. Signal Processing*, vol. 47, pp. 3250–3260, Dec. 1999.
- [24] K. T. Wong, "Adaptive source localization and blind beamforming for underwater acoustic wideband fast frequency-hop signals of unknown hop sequences & unknown arrival angles using a vector hydrophone," in *Proc. IEEE Wireless Commun. Networking Conf.*, 1999, pp. 664–668.
- [25] K. T. Wong and M. D. Zoltowski, "Self-initiating velocity-field beamspace MUSIC for underwater acoustic direction-finding with irregularly spaced vector hydrophones," *IEEE J. Oceanic Eng.*, vol. 25, pp. 262–273, Apr. 2000.
- [26] M. Hawkes and A. Nehorai, "Performance measures for estimating vector systems," *IEEE Trans. Signal Processing*, vol. 48, pp. 1737–1749, June 2000.
- [27] —, "Acoustic vector-sensor processing in the presence of a reflecting boundary," *IEEE Trans. Signal Processing*, vol. 48, pp. 2981–2993, Nov. 2000.
- [28] —, "Acoustic vector-sensor correlations in ambient noise," *IEEE J. Oceanic Eng.*, to be published.
- [29] K. T. Wong and H. Chi, "Beam patterns of an underwater acoustic vector hydrophone located away from any reflecting boundary," *IEEE J. Oceanic Eng.*, vol. 26, pp. 337–347, July 2001.

Kainam Thomas Wong (SM'01) received the B.S.E. degree in chemical engineering from the University of California, Los Angeles, in 1985, the B.S.E.E. from the University of Colorado, Boulder, in 1987, the M.S.E.E. degree from the Michigan State University, East Lansing, MI, in 1990, and the Ph.D. in electrical engineering from Purdue University, West Lafayette, IN, in 1996.

He was a manufacturing engineer with the General Motors Technical Center, Warren, MI, from 1990 to 1991 and a Senior Professional Staff Member at the Johns Hopkins University Applied Physics Laboratory, Laurel, MD, from 1996 to 1998. He was an Assistant Professor at Nanyang Technological University, Singapore in 1998 and the Chinese University of Hong Kong from 1998 to 2001. He has been an Assistant Professor at the University of Waterloo, Waterloo, ON, Canada since late 2001. He is a contributing author of about 70 articles for the inaugural edition of the *CRC Dictionary of Pure and Applied Physics*. He has authored or co-authored over a dozen IEEE journal papers. His research interests involve diverse aspects of signal processing for communications and sensor-array signal processing.

Dr. Wong has served on the Technical Program Committees of the 1999 and 2000 IEEE Wireless Communications and Networking Conference, the 2000 Spring IEEE Vehicular Technology Conference, and the Organizing Committee of the 2000 IEEE International Symposium on Circuits and Systems, among other conferences. He is listed in the *Marquis Who's Who in Science and Engineering*.

Michael D. Zoltowski (F'00) was born in Philadelphia, PA, on August 12, 1960. He received both the B.S. and M.S. degrees in electrical engineering with highest honors from Drexel University, Philadelphia, PA, in 1983 and the Ph.D. in systems engineering from the University of Pennsylvania, Philadelphia, in 1986.

In the fall of 1986, he joined the faculty of Purdue University, West Lafayette, IN, where he currently holds the position of Professor of electrical and computer engineering. His present research interests include space-time adaptive processing for all areas of mobile and wireless communications, GPS, and radar.

Dr. Zoltowski was the recipient of the IEEE Signal Processing Society's 1991 Paper Award (in the Statistical Signal and Array Processing area), "The Fred Eilersick MILCOM Award for Best Paper in the Unclassified Technical Program" at the IEEE Military Communications (MILCOM'98) Conference, and a Best Paper Award at the IEEE International Symposium on Spread Spectrum Techniques and Applications. He was a co-recipient of the IEEE Communications Society 2001 Leonard G. Abraham Prize Paper Award in the Field of Communications Systems. He is a contributing author to *Adaptive Radar Detection and Estimation* (New York: Wiley, 1991), *Advances in Spectrum Analysis and Array Processing, Vol. III* (Englewood Cliffs, NJ: Prentice-Hall, 1994), and *CRC Handbook on Digital Signal Processing* (Boca Raton, FL: CRC, 1996). He has served as an associate editor for both the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE COMMUNICATIONS LETTERS. Within the IEEE Signal Processing Society, he has been a member of the Technical Committee for the Statistical Signal and Array Processing Area and is currently a member of both the Technical Committee for Communications and the Technical Committee on DSP Education. In addition, he is currently a Member-at-Large of the Board of Governors and Secretary of the IEEE Signal Processing Society.

Petr Tichavský (M'98) graduated in 1987 from the Czech Technical University, Prague, Czechoslovakia. He received the Ph.D. degree in theoretical cybernetics from the Czechoslovak Academy of Sciences, Prague, in 1992.

Since that time he has been with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague. In 1994, he received the Fulbright grant for a 10-month fellowship at the Department of Electrical Engineering, Yale University, New Haven, CT. He is author and co-author of research papers in the area of sinusoidal frequency/frequency-rate estimation, adaptive filtering and tracking of time-varying signal parameters, and algorithm-independent bounds on achievable performance. His recent research interests also include independent component analysis and blind signal separation, electromagnetic and acoustic array processing, and signal processing for wireless communications.

Quasi-Fluid-Mechanics-Based Quasi-Bayesian Cramér–Rao Bounds for Deformed Towed-Array Direction Finding

Petr Tichavský, *Member, IEEE*, and Kainam Thomas Wong, *Senior Member, IEEE*

Abstract—New quasi-Bayesian (hybrid) Cramér–Rao bound (CRB) expressions are herein derived for far-field deep-sea direction-of-arrival (DOA) estimation with a nominally linear towed-array that 1) is deformed by *spatio-temporally correlated* oceanic currents, which have been previously overlooked in the towed-array shape-deformation statistical analysis literature, 2) is deformed by *temporally correlated* motion of the towing vessel, which is modeled only as temporally *uncorrelated* in prior literature, and 3) suffers gain-uncertainties and phase-uncertainties in its constituent hydrophones. This paper attempts to bridge an existing literature gap in deformed towed-array DOA-estimation performance analysis, by *simultaneously* a) incorporating several essential fluid-mechanics considerations to produce a shape-deformation statistical model physically more realistic than those previously used for DOA performance analysis and b) rigorously derive a mathematical analysis to characterize quantitatively and qualitatively the DOA estimation’s statistical performance. The derived CRB expressions are parameterized in terms of the towed-array’s physically measurable nonidealities for the single-source case. The new hybrid-CRB expressions herein derived are numerically more stable than those in the current literature.

Index Terms—Acoustical signal processing, array signal processing, direction-of-arrival estimation, marine telemetry, parameter estimation, sonar arrays, sonar signal processing, underwater acoustic arrays.

I. INTRODUCTION

A towed-array consists of an acoustically transparent and neutrally buoyant cable of hydrophones hauled behind a surface ship or a submerged vessel. A towed array may extend for several tens of meters to several hundreds of meters. The towed array’s nominally linear geometry may be arbitrarily distorted by the towing vessel’s varying speed and transverse motion, by the array’s non-neutral buoyance and nonuniform changes in density, and by hydrodynamic effects plus oceanic

swells and currents. The resulting snake-like deformation from the array’s nominal linearity can lead to critical degradation in the accuracy of arrival angle estimation/tracking, beamforming, and imaging because all these signal processing operations are predicated on a sufficiently accurate (*a priori* or estimated) model of the array’s inter-hydrophone spacings.

Towed-array deformity has been investigated by researchers from several complementary perspectives: Towed arrays’ geometric deformation has been empirically measured [13], [24], [35], computer simulated [39], and theoretically predicted based on fluid mechanics and oceanic physics [1]–[3], [7], [8], [11], [12]. A wealth of array-shape calibration algorithms have been devised using cooperative sources from known arrival-angles (in “aided calibration”) [23], [34], [38], by exploiting noncooperative sources from unknown arrival angles (in “self-calibration”) [9], [10], [14], [18], [22], [25], [29], [36], or by attaching on the towed-array nonacoustic positioning-devices (such as heading-sensors, depth-sensors and compasses to estimate the array’s displacements along the array-length axis, the vertical transverse axis and the horizontal transverse axis, respectively) [8], [16], [21], [24], [27]. The present work provides a quantitative analysis of bearing-estimation accuracy for deformed towed-arrays, assuming array shape-deformation information is available from neither cooperative calibration sources nor from nonacoustic positioning devices.

A. Literature on Modeling Towed-Array Shape-Deformation

The towed-array shape-deformation modeling literature generally falls into two categories: 1) fluid-mechanics-intensive models that are physically accurate but mathematically intractable for statistical signal parameter-estimation performance analysis and (2) mathematically simple models that overlooks most (if not all) fluid-mechanics-based considerations. The present manuscript aims to make one *initial* step toward bridging this crucial literature gap between 1) and 2) above by incorporating certain (admittedly, not all) essential fluid-mechanics considerations into the statistical measurement model, while drawing out in detail with rigorous mathematics a comprehensive (admittedly, not exhaustive) analysis of what this enhanced model implies in the statistical performance of direction-of-arrival (DOA) estimation.

1) *Fluid-Mechanics-Intensive Models*: Transverse deformation/vibration of a thin flexible cylinder, towed by a vessel, has been shown to obey a fourth-order partial differential equation known as the Paidoussis equation [1]–[3], [11], [12]. This equation, which was first applied in the towed-array

Manuscript received December 30, 2002; revised April 4, 2003. Part of this paper was presented at the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL. P. Tichavský was supported by the Grant Agency of the Czech Republic through Grant 102/01/0021. K. T. Wong was supported by Canada’s Natural Sciences and Engineering Research Council’s Individual Research Grant NSERC-RGPIN-249775-02. The associate editor coordinating the review of this paper and approving it for publication was Dr. Alex B. Gershman.

P. Tichavský is with the Institute of Information Theory and Automation, the Academy of Sciences of the Czech Republic, Prague, Czech Republic (e-mail: p.tichavsky@ieee.org).

K. T. Wong is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: ktwong@ieee.org).

Digital Object Identifier 10.1109/TSP.2003.820072

context in [7] and [8], describes the mechanical propagation of array-deformation down the array's length. The validity of this theoretical model was verified under field conditions [2].

Fluid-mechanics-based array deformation models have been used to investigate only array shape calibration in [21], [24], [36], and [38] but not for the present objective of DOA-estimation Cramér-Rao bound (CRB) analysis. Moreover, these models from [21], [24], [36], and [38] overlook the oceanic currents' statistical influence in the Paidoussis equation on array shape deformation. These earlier works also model the towing vessel's movement as temporally *uncorrelated*, which may be physically unrealistic for high time-sampling rates. In contrast, the present analysis offers more realism by allowing arbitrary temporal correlation (while assuming statistical stationarity) in the towing vessel's motion.

2) *Mathematically Simple Models With Little Fluid Mechanics*: A wealth of CRB analysis exists in the antenna-array signal-processing research literature on DOA estimation with uncertainties in the inter-antenna spacings. However, this antenna-array literature presumes *spatially*¹ *uncorrelated* and *spatially stationary* locational uncertainties from sensor to sensor. Unfortunately, such assumptions are manifestly invalid for a towed array, whose elements are strung up on a cable. Spatial *decorrelation*, in the towed array context, would imply rather implausibly that an upstream hydrophone's positional deviation has no effect on the downstream hydrophones' positional deviations. Spatial *stationarity* would unrealistically imply that the hydrophone secured at the tow point likely has a positional deviation comparable with those hydrophones at the tow cable's unsecured free end.

Among all DOA estimation CRB work accounting for spatial correlation among the sensors' dislocation (see [4], [5], [9], [10], [14], [17], [18], [20], [25], [26], [28], [29], [40]), none uses an array-deformation model rigorously derived from fluid mechanics. *Ad hoc* statistical models for array-shape deformation include [4], [9], [10], and [17], which assume as statistically *uncorrelated* the transverse and array-length axis positional perturbations. In [4], the transverse perturbations to be spatially correlated from hydrophone to hydrophone with a dependence inversely exponential to the cable length connecting the two hydrophones, but without rigorous justification, are modeled. In [9] and [10], the prior distribution for both the transverse uncertainties and the array-length axis uncertainties is assumed to be spatially uncorrelated Gaussian, which is an assumption that unrealistically implies that hydrophones near the tow point have positional variances comparable with those at the cable's free end. The rudimentary model of [17] postulates nothing beyond the aforementioned uncorrelated condition between the transverse and array-length axis positional perturbations. Another *ad hoc* deformation model is used in [28], without any physics-based justification, involving a transverse perturbation whose standard deviation increases quadratically downstream and a linearly increasing perturbation along the array-length axis. The spatial correlation and the transverse/array-length correlation in [28] are both 100% correlated.

The more sophisticated deterministic piecewise-linear model presumes the relative angles between adjoining piecewise-linear segments to be deterministic unknown constants. For example,

[22], [23], [29], [33], [34], and [38] use the deterministic piecewise linear shape deformation model for array shape calibration performance analysis. A stochastic piecewise-linear model, assuming the relative angle between adjoining piecewise-linear segments to be Gaussian and (implausibly² as) spatially uncorrelated, is used in [29] for DOA estimation.

Moreover, much of the above-mentioned deformed-array bearing-estimation literature (all except [17], [28], and [29] and unlike the present work) unrealistically assumes that no uncertainty exists in the hydrophone's gain and phase responses. This work attempts to be comprehensive in accounting simultaneously for diverse array nonidealities.

B. CRB Literature on Deformed-Array DOA Estimation

The DOA estimation lower bounds herein derived are quasi-Bayesian (hybrid) CRBs that characterize the best standard deviation obtainable using any unbiased estimator of a vector parameter. The CRB may serve as a performance metric in towed-array design with any required level of bearing estimation accuracy.

The terms "quasi-Bayesian" and "hybrid" aim to contrast against the standard CRB to signify that the vector-parameter here has a deterministic subvector and a random nuisance-parameter subvector. The former consists of the incident sources' unknown but to-be-estimated angles of arrival. The latter does not need to be estimated but characterizes the array-shape deformations, the phase/gain uncertainties of the individual acoustic sensors, and other factors. The "hybrid" (quasi-Bayesian) CRB may be defined as a proper submatrix of the overall vector-parameter's CRB matrix, which is equal to the inverse of the corresponding Fisher information matrix. The bound depends on the *a priori* distribution (uncertainty) of the random subvector. It depends only on the signal/noise statistical model but not the particular estimation algorithm method used; however, the quasi-Bayesian CRB may be attained by an maximum *a posteriori* (MAP) estimator of the parameter-estimation problem [29].

The present analysis allows a broad class of Bayesian-like statistical models parameterized with physically measurable quantities. For example, certain independent parameters [describing the tow-point induced (TPI) motion and oceanic currents] in the Paidoussis equation are herein characterized as stochastic with known prior distributions instead of as deterministic unknowns. This quasi-Bayesian approach is advantageous because the underlying fluid mechanical and oceanic physical processes (that cause the array's geometric deformation) can be neither exactly measured nor precisely estimated. Hence, they would best be modeled as stochastic phenomena. As oceanic engineers gather new data and update the statistics of such TPI-motion and oceanic current, the statistical properties of the Bayesian parameters may be estimated and substituted in the CRB formulas presented in this paper. Lower bounds of the deformed towed-array's DOA-estimation variance can then be obtained along with the general expressions for the quasi-Bayesian Cramér-Rao lower bound, which is derived in [29]. Among all prior work cited in the preceding paragraphs, only [4], [9], [10], [17], [28],

¹"Space" as spanned by the array's geometrical axes.

²See further discussion in the following paragraphs.

and [29] also use a Bayesian approach, but (as discussed in the preceding subsection) none of these papers model array shape deformation based on rigorous fluid mechanics, as in the present work.

This paper is partly based on the generic CRB expression derived in [29] for sensor-array estimation under array uncertainties. Because this expression might be numerically unstable when the data-length N approaches infinity or when the covariance matrices of array dislocations are small, this present paper derives an alternative expression that is numerically more stable. This new expression allows easy computation of the limit-CRB for N approaching infinity. As an example, the CRB is herein computed in the case of a single source. The main contribution of the paper, however, consists of computing physically meaningful covariance matrices of array dislocations, which was discussed in the previous subsection.

II. MATHEMATICAL DATA MODELS FOR FAR-FIELD SOURCES PROPAGATING THROUGH A DEEP-SEA CHANNEL

This section introduces the mathematical and statistical data models involved in far-field deep-sea direction finding. K far-field narrowband sources impinge on an L -hydrophone array as plane waves without time-delayed multipaths to produce at time t the measured data vector:

$$\mathbf{d}(t) = \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho}) \mathbf{s}(t) + \mathbf{n}(t). \quad (1)$$

The k th column $\mathbf{a}_k(\boldsymbol{\theta}, \boldsymbol{\rho})$ of the $L \times K$ matrix $\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho})$ represents the k th source's steering vector, which has as components³

$$[\mathbf{a}_k(\boldsymbol{\theta}, \boldsymbol{\rho})]_\ell = g_\ell e^{j\varphi_\ell} e^{j2\pi(x_\ell u_k + y_\ell v_k + z_\ell w_k)/\lambda}, \quad \ell = 1, \dots, L \quad (2)$$

where (u_k, v_k, w_k) represent the Cartesian direction-cosines of the k th incident source, g_ℓ and φ_ℓ , respectively, denote the ℓ th sensor's unity-mean gain-perturbation and zero-mean phase-perturbation, (x_ℓ, y_ℓ, z_ℓ) symbolize the three-dimensional (3-D) Cartesian position coordinates of the ℓ th sensor, and λ denotes the wavelength. The k th element in the $K \times 1$ vector $\mathbf{s}(t)$ represents the k th frequency-down-converted incident temporal signal and is modeled as a temporally uncorrelated zero-mean complex-valued circular-Gaussian stochastic process with $\mathbf{s}(t)$ having the *a priori* unknown covariance matrix $\boldsymbol{\Omega}_s$. The ℓ th element in the $L \times 1$ vector $\mathbf{n}(t)$ refers to the spatio-temporally uncorrelated complex-valued additive noise at the ℓ th hydrophone, with *a priori* unknown variance σ^2 .

The $M_\theta \times 1$ vector $\boldsymbol{\theta}$ contains as its elements the to-be-estimated unknown deterministic signal parameters, e.g., the azimuth and/or elevation angles or, equivalently, the Cartesian direction cosines. The $M_\rho \times 1$ nuisance vector $\boldsymbol{\rho}$ consists of the nuisance parameters — $\{g_\ell, \varphi_\ell, x_\ell, y_\ell, z_\ell, \ell = 1, \dots, L\}$. Further, define $\mathbf{g} = [g_1, \dots, g_L]^T$, $\boldsymbol{\varphi} = [\varphi_1, \dots, \varphi_L]^T$, $\mathbf{x} = [x_1, \dots, x_L]^T$, $\mathbf{y} = [y_1, \dots, y_L]^T$, and $\mathbf{z} = [z_1, \dots, z_L]^T$. These stochastic parameter vectors are modeled as mutually independent and real-valued Gaussian distributed, with *a priori* known nominal means \mathbf{g}_0 , $\boldsymbol{\varphi}_0$, \mathbf{x}_0 , \mathbf{y}_0 , \mathbf{z}_0 , and *a priori* known covariance matrices $\boldsymbol{\Omega}_{\Delta g}$, $\boldsymbol{\Omega}_{\Delta \varphi}$, $\boldsymbol{\Omega}_{\Delta x}$, $\boldsymbol{\Omega}_{\Delta y}$, and $\boldsymbol{\Omega}_{\Delta z}$. Hence, $\boldsymbol{\rho}$

may be represented as a real-valued, Gaussian, stochastic vector with *a priori* known mean $\boldsymbol{\rho}_0$ and *a priori* known covariance matrix $\boldsymbol{\Omega}_\rho$. To summarize, the present data model involves the unknown *stochastic* entities of $\boldsymbol{\rho}$, $\mathbf{s}(t)$ and $\mathbf{n}(t)$, plus the unknown *deterministic* entities of $\boldsymbol{\theta}$, $\boldsymbol{\Omega}_s$, and σ^2 . However, only $\boldsymbol{\theta}$ needs to be estimated.

III. NEW CRB EXPRESSION COMPATIBLE WITH VARIOUS TO-BE-SPECIFIED QUASI-BAYESIAN MODELS OF ARRAY NONIDEALITIES

In [29], there is a “generic” quasi-Bayesian CRB expression applicable to far-field deep-sea nonideal array direction finding. Building on [29], this section will develop CRB expressions that 1) are numerically more stable and applicable to any number of incident sources, 2) reveal the multisource CRBs asymptotic behavior as the data-length N approaches infinity, and 3) link to the physical quantities parameterizing various array nonidealities in the single-source case. From [29]:

$$\begin{aligned} E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T] &\geq \text{CRB}_\theta \\ &= \frac{1}{N} [\mathbf{C}_{\theta\theta} - \mathbf{C}_{\rho\theta}^T (\mathbf{C}_{\rho\rho} + (2N\boldsymbol{\Omega}_\rho)^{-1})^{-1} \mathbf{C}_{\rho\theta}]^{-1} \end{aligned} \quad (3)$$

where

$$\begin{aligned} \mathbf{C}_{\theta\theta} &= \text{Re}\{\mathbf{D}_\theta^H \mathbf{M} \mathbf{D}_\theta\} \\ &\quad (\text{real-valued, } M_\theta \times M_\theta \text{ in size}) \end{aligned} \quad (4)$$

$$\begin{aligned} \mathbf{C}_{\rho\theta} &= \text{Re}\{\mathbf{D}_\rho^H \mathbf{M} \mathbf{D}_\theta\} \\ &\quad (\text{real-valued, } M_\rho \times M_\theta \text{ in size}) \end{aligned} \quad (5)$$

$$\begin{aligned} \mathbf{C}_{\rho\rho} &= \text{Re}\{\mathbf{D}_\rho^H \mathbf{M} \mathbf{D}_\rho\} \\ &\quad (\text{real-valued, } M_\rho \times M_\rho \text{ in size}) \end{aligned} \quad (6)$$

$$\begin{aligned} \mathbf{D}_\theta &= \left[\frac{\partial \text{vec}[\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho})]}{\partial \boldsymbol{\theta}_1}, \dots, \frac{\partial \text{vec}[\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho})]}{\partial \boldsymbol{\theta}_{M_\theta}} \right] \\ &\quad (\text{complex-valued, } LK \times M_\theta \text{ in size}) \end{aligned} \quad (7)$$

$$\begin{aligned} \mathbf{D}_\rho &= \left[\frac{\partial \text{vec}[\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho})]}{\partial \rho_1}, \dots, \frac{\partial \text{vec}[\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\rho})]}{\partial \rho_{M_\rho}} \right] \\ &\quad (\text{complex-valued, } LK \times M_\rho \text{ in size}) \end{aligned} \quad (8)$$

$$\begin{aligned} \mathbf{M} &= \sigma^{-2} \{\boldsymbol{\Omega}_s \mathbf{A}^H(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) \mathbf{R}^{-1} \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) \boldsymbol{\Omega}_s\}^T \otimes \boldsymbol{\Pi}_A^\perp \\ &\quad (\text{complex-valued, } LK \times LK \text{ in size}) \end{aligned} \quad (9)$$

$$\begin{aligned} \mathbf{R} &= \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) \boldsymbol{\Omega}_s \mathbf{A}^H(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) + \sigma^2 \mathbf{I} \\ &\quad (\text{complex-valued, } L \times L \text{ in size}) \end{aligned} \quad (10)$$

$$\begin{aligned} \boldsymbol{\Pi}_A^\perp &= \mathbf{I} - \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) (\mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) \mathbf{A}^H(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0))^{-1} \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)^H \\ &\quad (\text{complex-valued, } L \times L \text{ in size}) \end{aligned} \quad (11)$$

where \otimes denotes the Kronecker product, $\boldsymbol{\theta}_0$ represents the theoretical value of $\boldsymbol{\theta}$, and the derivatives in (7) and (8) are evaluated for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\boldsymbol{\rho} = \boldsymbol{\rho}_0$.

The CRB $_\theta$ expression in (3) might be numerically unstable because $\boldsymbol{\Pi}_A^\perp$ is rank-deficient, as $\boldsymbol{\Pi}_A^\perp$ represents a projection operator such that $\boldsymbol{\Pi}_A^\perp \mathbf{A}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) = \mathbf{0}$. This means that $\mathbf{C}_{\rho\rho}$ and \mathbf{M} are both rank-deficient, and consequently, the to-be-inverted term $\mathbf{C}_{\rho\rho} + (2N\boldsymbol{\Omega}_\rho)^{-1}$ might be ill-conditioned for large $N\boldsymbol{\Omega}_\rho$.

³The index k will be dropped in the case of a single source.

Assume that \mathbf{A} has full column rank. Then, the projection operator $\mathbf{\Pi}_A^\perp$ has rank $L - K$ and

$$\mathbf{\Pi}_A^\perp = \mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \quad (12)$$

where \mathbf{B} consists of the $L - K$ linearly independent columns of $\mathbf{\Pi}_A^\perp$ that span the column space of $\mathbf{\Pi}_A^\perp$. Usually, \mathbf{B} can be formed from the first $L - K$ columns of $\mathbf{\Pi}_A^\perp$, i.e., $\mathbf{B} = \mathbf{\Pi}_A^\perp \mathbf{J}$, where

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_{(L-K) \times (L-K)} \\ \mathbf{0}_{K \times (L-K)} \end{bmatrix}. \quad (13)$$

Put

$$\mathbf{M}_0 = \sigma^{-2} \{ \mathbf{\Omega}_s \mathbf{A}^H(\theta_0, \rho_0) \mathbf{R}^{-1} \mathbf{A}(\theta_0, \rho_0) \mathbf{\Omega}_s \}^T \quad (14)$$

so that \mathbf{M} of (9) may be written as $\mathbf{M} = \mathbf{M}_0 \otimes \mathbf{\Pi}_A^\perp$. Then

$$\begin{aligned} \mathbf{C}_{\theta\theta} &= \text{Re}\{ \mathbf{D}_{\theta}^H \mathbf{M} \mathbf{D}_{\theta} \} = \text{Re}\{ \mathbf{D}_{\theta}^H (\mathbf{M}_0 \otimes \mathbf{\Pi}_A^\perp) \mathbf{D}_{\theta} \} \\ &= \text{Re}\{ \mathbf{D}_{\theta}^H [\mathbf{M}_0 \otimes (\mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H)] \mathbf{D}_{\theta} \} \\ &= \text{Re}\{ \mathbf{D}_{\theta}^H (\mathbf{I} \otimes \mathbf{B}) [\mathbf{M}_0 \otimes (\mathbf{B}^H \mathbf{B})^{-1}] (\mathbf{I} \otimes \mathbf{B})^H \mathbf{D}_{\theta} \} \\ &= [\text{Re}\{ \mathbf{D}_{\theta}^H (\mathbf{I} \otimes \mathbf{B}) \}, \text{Im}\{ \mathbf{D}_{\theta}^H (\mathbf{I} \otimes \mathbf{B}) \}] \\ &\quad \times \begin{bmatrix} \text{Re}\{ \mathbf{M}_0 \otimes (\mathbf{B}^H \mathbf{B})^{-1} \} & -\text{Im}\{ \mathbf{M}_0 \otimes (\mathbf{B}^H \mathbf{B})^{-1} \} \\ \text{Im}\{ \mathbf{M}_0 \otimes (\mathbf{B}^H \mathbf{B})^{-1} \} & \text{Re}\{ \mathbf{M}_0 \otimes (\mathbf{B}^H \mathbf{B})^{-1} \} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \text{Re}\{ \mathbf{D}_{\theta}^H (\mathbf{I} \otimes \mathbf{B}) \}^T \\ \text{Im}\{ \mathbf{D}_{\theta}^H (\mathbf{I} \otimes \mathbf{B}) \}^T \end{bmatrix} \\ &= \mathbf{W} \mathbf{M}_c \mathbf{W}^T \end{aligned} \quad (15)$$

where \mathbf{I} denotes the identity matrix of size $K \times K$, and

$$\mathbf{W} = [\text{Re}\{ \mathbf{D}_{\theta}^H (\mathbf{I} \otimes \mathbf{B}) \}, \text{Im}\{ \mathbf{D}_{\theta}^H (\mathbf{I} \otimes \mathbf{B}) \}] \quad (16)$$

$$\mathbf{M}_c = \begin{bmatrix} \text{Re}\{ \mathbf{M}_0 \otimes (\mathbf{B}^H \mathbf{B})^{-1} \} & -\text{Im}\{ \mathbf{M}_0 \otimes (\mathbf{B}^H \mathbf{B})^{-1} \} \\ \text{Im}\{ \mathbf{M}_0 \otimes (\mathbf{B}^H \mathbf{B})^{-1} \} & \text{Re}\{ \mathbf{M}_0 \otimes (\mathbf{B}^H \mathbf{B})^{-1} \} \end{bmatrix}. \quad (17)$$

Similarly, it can be shown that

$$\mathbf{C}_{\rho\rho} = \mathbf{S} \mathbf{M}_c \mathbf{W}^T \quad (18)$$

$$\mathbf{C}_{\rho\theta} = \mathbf{S} \mathbf{M}_c \mathbf{S}^T \quad (19)$$

where

$$\mathbf{S} = [\text{Re}\{ \mathbf{D}_{\rho}^H (\mathbf{I} \otimes \mathbf{B}) \}, \text{Im}\{ \mathbf{D}_{\rho}^H (\mathbf{I} \otimes \mathbf{B}) \}]. \quad (20)$$

Inserting (15), (18), and (19) in (3) and applying the matrix inversion lemma $(\mathbf{A} + \mathbf{B}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}$, which holds for any size-compatible matrices \mathbf{A} , \mathbf{B} , and \mathbf{C} , and providing all relevant inverses exist

$$\begin{aligned} & [\text{CRB}_{\theta}]^{-1} \\ &= N [\mathbf{W} \mathbf{M}_c \mathbf{W}^T - \mathbf{W} \mathbf{M}_c \mathbf{S}^T (\mathbf{S} \mathbf{M}_c \mathbf{S}^T + (2N \mathbf{\Omega}_{\rho})^{-1})^{-1} \mathbf{S} \mathbf{M}_c \mathbf{W}^T] \\ &= N \mathbf{W} [\mathbf{M}_c - \mathbf{M}_c \mathbf{S}^T ((2N \mathbf{\Omega}_{\rho})^{-1} + \mathbf{S} \mathbf{M}_c \mathbf{S}^T)^{-1} \mathbf{S} \mathbf{M}_c] \mathbf{W}^T \\ &= N \mathbf{W} (\mathbf{M}_c^{-1} + 2N \mathbf{S}^T \mathbf{\Omega}_{\rho} \mathbf{S})^{-1} \mathbf{W}^T \\ &= \frac{1}{2} \mathbf{W} \left(\mathbf{S}^T \mathbf{\Omega}_{\rho} \mathbf{S} + \frac{1}{2N} \mathbf{M}_c^{-1} \right)^{-1} \mathbf{W}^T. \end{aligned} \quad (21)$$

Note that \mathbf{M}_c of (17) is regular, provided that \mathbf{M}_0 in (14) is regular because $\mathbf{M}_0 \otimes (\mathbf{B}^H \mathbf{B})^{-1}$ would then be regular as well. Thus, the new CRB expression in (21) is numerically stable for large N , unlike (3).

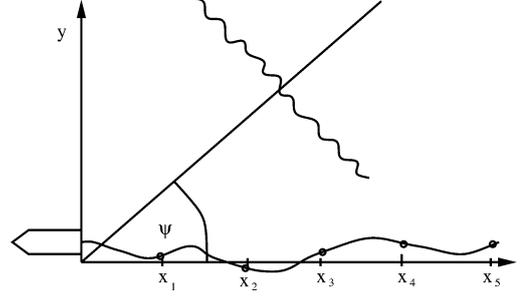


Fig. 1. Coordinate system for a towed-array of hydrophones with one far-field incident source.

The limit-CRB for N going to infinity easily follows:

$$\text{CRB}_{\theta}^{N \rightarrow \infty} \stackrel{\text{def}}{=} \lim_{N \rightarrow \infty} \text{CRB}_{\theta} = 2 [\mathbf{W} (\mathbf{S}^T \mathbf{\Omega}_{\rho} \mathbf{S})^{-1} \mathbf{W}^T]^{-1}. \quad (22)$$

The limit has the interpretation that it describes the best achievable residual variance of the DOA-estimate due to array uncertainties.

To detail the impact of array-shape uncertainties and of the gain-phase uncertainties on direction-finding accuracy, the subsequent analysis assumes a single incident source (i.e., $K = 1$) for mathematical simplicity. Without loss of generality, the nominally linear towed array is assumed to align along the x -axis; hence, the x -axis Cartesian direction-cosine needs to be estimated. That is, $\theta = u = \cos \psi$; see Fig. 1. From (2), $\mathbf{a}(\theta, \rho) = \mathbf{g} \odot e^{j\phi}$, where \odot denotes the element-wise product, and $\phi = \varphi + 2\pi/\lambda(\mathbf{x}u + \mathbf{y}v + \mathbf{z}w)$.

Further assume that uncertainties in the hydrophones' gains, phases, and locations are mutually independent. Then, random deviations of ρ around its nominal value ρ_0 have the *a priori* covariance matrix

$$\mathbf{\Omega}_{\rho} = \text{diag}(\mathbf{\Omega}_{\Delta g}, \mathbf{\Omega}_{\Delta \varphi}, \mathbf{\Omega}_{\Delta x}, \mathbf{\Omega}_{\Delta y}, \mathbf{\Omega}_{\Delta z}). \quad (23)$$

For the above-defined $\mathbf{a}(\theta, \rho)$

$$\mathbf{D}_{\theta} = \frac{\partial \mathbf{a}(\theta, \rho)}{\partial u} = j \frac{2\pi}{\lambda} (\mathbf{x} \odot \mathbf{a}(\theta_0, \rho_0)) \quad (24)$$

where \odot denotes the element-wise product, and

$$\begin{aligned} \mathbf{D}_{\rho} &= \frac{\partial \mathbf{a}(\theta, \rho)}{\partial \rho} \\ &= \left[\frac{\partial \mathbf{a}(\theta, \rho)}{\partial \mathbf{g}}, \frac{\partial \mathbf{a}(\theta, \rho)}{\partial \varphi}, \frac{\partial \mathbf{a}(\theta, \rho)}{\partial \mathbf{x}}, \frac{\partial \mathbf{a}(\theta, \rho)}{\partial \mathbf{y}}, \frac{\partial \mathbf{a}(\theta, \rho)}{\partial \mathbf{z}} \right] \\ &= [\text{diag}(e^{j\phi}), j\mathbf{m} \otimes \text{diag}(\mathbf{a})] \end{aligned} \quad (25)$$

where

$$\mathbf{m} = \left[1, \frac{2\pi}{\lambda} u, \frac{2\pi}{\lambda} v, \frac{2\pi}{\lambda} w \right]. \quad (26)$$

Note that \mathbf{M}_0 of (14) becomes the scalar M_0

$$\begin{aligned} M_0 &= \sigma^{-2} (\mathbf{\Omega}_s \mathbf{A}^H \mathbf{R}^{-1} \mathbf{A} \mathbf{\Omega}_s) \\ &= \frac{\Omega_s^2}{\sigma^2} \mathbf{a}^H (\sigma^2 \mathbf{I} + \Omega_s \mathbf{a} \mathbf{a}^H)^{-1} \mathbf{a} \\ &= \frac{\Omega_s^2}{\sigma^2} \mathbf{a}^H \left[\frac{1}{\sigma^2} \mathbf{I} - \frac{\Omega_s}{\sigma^4} (1 + \sigma^{-2} \Omega_s \|\mathbf{a}\|^2)^{-1} \mathbf{a} \mathbf{a}^H \right] \mathbf{a} \\ &= \frac{\Omega_s^2 \|\mathbf{g}\|^2}{\sigma^2 (\sigma^2 + \Omega_s \|\mathbf{g}\|^2)}. \end{aligned} \quad (27)$$

In (27), the matrix inversion lemma is used along with $\|\mathbf{a}\| = \|\mathbf{g}\|$. The matrix \mathbf{B} , containing the basis of the column-space of $\mathbf{\Pi}_A^\perp$, may be chosen arbitrarily. However

$$\mathbf{B} = \mathbf{\Pi}_A^\perp \text{diag}(e^{j\phi}) \mathbf{J} \quad (28)$$

would be a convenient choice,⁴ giving

$$\begin{aligned} \mathbf{D}_\rho^H \mathbf{B} &= \begin{bmatrix} \text{diag}(e^{-j\phi}) \\ -j\mathbf{m}^T \otimes \text{diag}(\mathbf{a}^*) \end{bmatrix} \begin{bmatrix} \mathbf{I} - \frac{\mathbf{a}\mathbf{a}^H}{\|\mathbf{a}\|^2} \\ \mathbf{0} \end{bmatrix} \text{diag}(e^{j\phi}) \mathbf{J} \\ &= \begin{bmatrix} \mathbf{\Pi}_g^\perp \mathbf{J} \\ -j\mathbf{m}^T \otimes (\text{diag}(\mathbf{g})\mathbf{\Pi}_g^\perp \mathbf{J}) \end{bmatrix} = \begin{bmatrix} \mathbf{\Pi}_g^\perp \mathbf{J} \\ -j\mathbf{m}^T \otimes \mathbf{H} \end{bmatrix} \end{aligned} \quad (29)$$

where

$$\mathbf{\Pi}_g^\perp = \mathbf{I} - \frac{\mathbf{g}\mathbf{g}^H}{\|\mathbf{g}\|^2} \quad (30)$$

$$\mathbf{H} = \text{diag}(\mathbf{g})\mathbf{\Pi}_g^\perp \mathbf{J}. \quad (31)$$

After some algebra, (20), (17), and (16) can be rewritten as

$$\mathbf{S} = \begin{bmatrix} \mathbf{\Pi}_g^\perp \mathbf{J} & \mathbf{0} \\ \mathbf{0} & -\mathbf{m}^T \otimes \mathbf{H} \end{bmatrix} \quad (32)$$

$$\mathbf{W} = \begin{bmatrix} 0, -\frac{2\pi}{\lambda} \mathbf{x}^T \mathbf{H} \end{bmatrix} \quad (33)$$

$$\mathbf{M}_c = M_0 \begin{bmatrix} (\mathbf{J}\mathbf{\Pi}_g^\perp \mathbf{J})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{J}\mathbf{\Pi}_g^\perp \mathbf{J})^{-1} \end{bmatrix}. \quad (34)$$

Combining (32) and (23)

$$\mathbf{S}^T \mathbf{\Omega}_\rho \mathbf{S} = \begin{bmatrix} * & * \\ * & \mathbf{H}^T \mathbf{\Omega}_{\Delta\phi} \mathbf{H} \end{bmatrix} \quad (35)$$

where the blocks denoted by the asterisks are not displayed, and

$$\mathbf{\Omega}_{\Delta\phi} = \mathbf{\Omega}_{\Delta\varphi} + \left(\frac{2\pi}{\lambda}\right)^2 [u^2 \mathbf{\Omega}_{\Delta x} + v^2 \mathbf{\Omega}_{\Delta y} + w^2 \mathbf{\Omega}_{\Delta z}]. \quad (36)$$

Inserting (32)–(35) in (21) gives

$$\begin{aligned} [\text{CRB}_\theta]^{-1} &= \frac{1}{2} \mathbf{W} \left(\mathbf{S}^T \mathbf{\Omega}_\rho \mathbf{S} + \frac{1}{2N} \mathbf{M}_c^{-1} \right)^{-1} \mathbf{W}^T \\ &= \frac{2\pi^2}{\lambda^2} \mathbf{x}^T \mathbf{H} \left(\mathbf{H}^T \mathbf{\Omega}_{\Delta\phi} \mathbf{H} + \frac{M_0}{2N} \mathbf{J}\mathbf{\Pi}_g^\perp \mathbf{J} \right)^{-1} \mathbf{H}^T \mathbf{x}. \end{aligned} \quad (37)$$

The CRB depends on the hydrophones' gains by means of the matrix \mathbf{H} defined in (31). It is independent of the uncertainties in the hydrophones' gain $\mathbf{\Omega}_{\Delta g}$ but depends only on $\mathbf{\Omega}_{\Delta\phi}$, which combines the hydrophones' phase uncertainties and location uncertainties. This is because all arrival-angle information is contained in the phase of the data; cf. (2). For multiple sources (i.e., $K > 1$), these uncertainties have more complex interactions; and the general formula in (21) would be necessary.

$N \rightarrow \infty$ gives the asymptotic CRB

$$\text{CRB}_\theta^{N \rightarrow \infty} = \frac{\lambda^2}{2\pi^2} [\mathbf{x}^T \mathbf{H} (\mathbf{H}^T \mathbf{\Omega}_{\Delta\phi} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}]^{-1}. \quad (38)$$

⁴The exponential term in (28) is allowed because the scaling of the columns of \mathbf{B} is arbitrary due to the projection in (12)

As expected, the limit-CRB is independent of the additive noise's variance σ^2 , which drops out along with M_0 of (27). The CRB does depend on the array geometry not only through the vector of sensor coordinates \mathbf{x} but through the error covariance matrix $\mathbf{\Omega}_{\Delta\phi}$ as well, which incorporates the sensor location uncertainties.

IV. NEW FLUID-MECHANICS-BASED STATISTICAL MODELS OF A NOMINALLY LINEAR TOWED ARRAY'S SHAPE DEFORMATION

The derivation of the Paidoussis equation in [21] is herein revisited in order to incorporate a new term for the fluid flow's transverse speed and normal speed caused by oceanic streams and swells. The fluid flow's instantaneous normal speed is modeled as an homogenous stationary Gaussian random field with known space-time correlation structure, which may be measured offline and tabulated for different field conditions. This Gaussian assumption is for mathematical simplicity and is not unreasonable because statistical distributions with longer "tails" (thereby implying a higher probability for very high fluid-flow speeds) can hardly be observed here due to fluid viscosity.

The towing vessel transversal motion represents another cause of array deformation. Already accounted for in the original Paidoussis equation, the towing vessel's transversal motion is herein assumed to be due to the vessel's small random maneuvers and is modeled as a Gaussian random field with known space-time correlation structure (which may be measured and tabulated off-line) and as stochastically independent of the fluid flows along the array.

The Paidoussis equation is discretized both in time and in space and consequently used to derive physically meaningful covariance matrices of the sensor location uncertainties, $\mathbf{\Omega}_{\Delta x}$, $\mathbf{\Omega}_{\Delta y}$, and $\mathbf{\Omega}_{\Delta z}$ for use in the quasi-Bayesian CRB. The following developments will consider only small array-shape deformations. Since the tow-cable is assumed neutrally buoyant, the horizontal deformation and the vertical deformation obeys the same differential equation (but possibly with different constants for the two directions). Hence, with no loss in generality, the subsequent analysis will express only the horizontal deformation's $\mathbf{\Omega}_{\Delta y}$ in terms of physically measurable constants. Subsequent simulation examples will assume that $\mathbf{\Omega}_{\Delta z} = \mathbf{\Omega}_{\Delta y}$. Finally, relative longitudinal contractions of the array can be neglected thanks to the small array-shape deformation assumption. It follows that $\mathbf{\Omega}_{\Delta x} \approx \mathbf{0}$.

A. Generic Model of Towed Array Fluid Mechanics

Two causes exist for towed-array deformation: 1) the towing-vessel's transverse motion or varying speed and 2) oceanic swells and currents. The Paidoussis equation [1]–[3], [7], [8], [11], [12] describes the fluid mechanics through which the two above-mentioned factors affect the shape of a towed array. More precisely, the Paidoussis equation describes the dynamical behavior of a flexible and cross-sectionally thin cable towed through a certain fluid:

$$\begin{aligned} m \frac{\partial^2 \mathbf{y}(t, x)}{\partial t^2} - \frac{\partial}{\partial x} \left(T(x) \frac{\partial \mathbf{y}(t, x)}{\partial x} \right) + f_A(t, x) + f_B(t, x) \\ + f_N(t, x) - f_T(t, x) \frac{\partial \mathbf{y}(t, x)}{\partial x} + B \frac{\partial^4 \mathbf{y}(t, x)}{\partial x^4} = 0. \end{aligned} \quad (39)$$

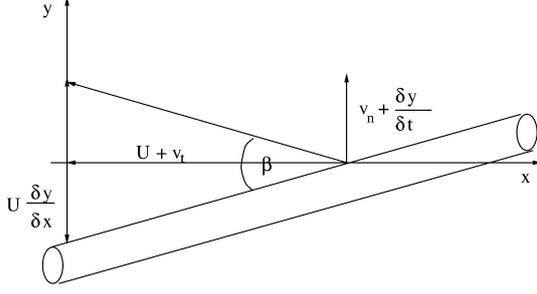


Fig. 2. Enlarged segment of the deformed towed array.

The notation is as follows:

- $y(t, x)$ towed-array's transverse displacement at time t and location x along the array's length;
- m towed-array's per-unit-length mass;
- $T(x)$ tow-cable's spatially variable tension;
- $f_A(t, x)$ inviscid force due to the acceleration of the tow-cable's virtual mass;
- $f_N(t, x)$ per-unit-length viscous force acting on the tow-cable in the normal direction;
- $f_T(t, x)$ similar force acting in the tangential direction;
- B tow-cable's bending stiffness.

From [11] and [38],

$$f_A(t, x) = M \left[\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right]^2 y(t, x) \quad (40)$$

$$f_N(t, x) = \frac{M[V(t, x)]^2}{D} C_n \sin \beta(t, x) \quad (41)$$

$$f_T(t, x) = \frac{M[V(t, x)]^2}{D} C_t \cos \beta(t, x) \quad (42)$$

$$\begin{aligned} T(x) &= T(L) - \int_x^L f_T(t, x) dx \\ &= C_t' MU^2 + \int_x^L f_T(t, x) dx \end{aligned} \quad (43)$$

with these notations:

- M displaced fluid's per-unit-length mass;
- D towed-array's cross-sectional diameter;
- C_n towed-array's normal drag coefficient;
- C_t towed-array's tangential drag coefficient;
- C_t' form drag at the trailing end (= 0 for a free end);
- U tow-vessel's speed along the positive x -axis;
- $V(t, x)$ tow-cable's speed relative to distant fluid;
- $\beta(t, x)$ angle between the relative velocity of the surrounding fluid flow and the local tangent of the cable.

$\beta(t, x)$ depends on U but also on the fluid flow's transverse speed $v_t(t, x)$ and normal speed $v_n(t, x)$ due to the ocean streams and swells (see Fig. 2).

To summarize, the inputs to the fourth-order partial differential equation in (39) are U , $v_t(t, x)$, and $v_n(t, x)$, under the TPI-motion defined boundary condition $y(t, 0)$, and the output is $y(t, x)$: the towed-array's space-time deformed geometry.

Prior researchers (to the best of the authors' knowledge) have overlooked the statistical influence of oceanic currents on array shape deformation in the Paidoussis equation. One contribution

of the present work is to rigorously characterize the statistical effects of $v_t(t, x)$ and $v_n(t, x)$ in the Paidoussis equation on the hydrophones' dislocation.

Under the reasonable assumption that the tow-speed greatly exceeds the surrounding water's flow velocity (i.e., $U \gg v_t(t, x)$, $U \gg v_n(t, x)$), it holds that $V(t, x) \approx U$ and $\beta(t, x) \approx 0$. The latter approximation means that hydrophone dislocation is substantial only perpendicular to, but not along, the array axis. Referring to Fig. 2

$$\begin{aligned} \sin \beta(t, x) &\approx \beta(t, x) \\ &\approx \frac{1}{U} \left[\frac{\partial y(t, x)}{\partial t} + U \frac{\partial y(t, x)}{\partial x} - v_n(t, x - tU) \right] \end{aligned} \quad (44)$$

$$\cos \beta(t, x) \approx 1 \quad (45)$$

$$T(x) \approx \left(C_t \frac{L-x}{D} + C_t' \right) MU^2. \quad (46)$$

Assuming that the tow-cable is sufficiently flexible to neglect the bending stiffness term $B(\partial^4 y(t, x)/\partial x^4)$ in (39), the fourth-order Paidoussis equation in (39) may be reduced to second-order:

$$\begin{aligned} m \frac{\partial^2 y(t, x)}{\partial t^2} + M \left[\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right]^2 y(t, x) \\ - \frac{\partial}{\partial x} \left[\left(C_t \frac{L-x}{D} + C_t' \right) MU^2 \frac{\partial y(t, x)}{\partial x} \right] \\ - \frac{MU^2}{D} C_t \frac{\partial y(t, x)}{\partial x} \\ + C_n \frac{MU}{D} \left[\frac{\partial y(t, x)}{\partial t} + U \frac{\partial y(t, x)}{\partial x} - v_n(t, x - tU) \right] = 0. \end{aligned} \quad (47)$$

Moreover, because $D \ll L$, the subsequent analysis will ignore those terms in the above equations that are not inversely proportional to D . This second approximation results in the *small-diameter Paidoussis equation*⁵ [21]:

$$\begin{aligned} (x-L) \frac{\partial^2 y(t, x)}{\partial x^2} + \frac{C_n}{C_t U} \left[\frac{\partial y(t, x)}{\partial t} \right. \\ \left. + U \frac{\partial y(t, x)}{\partial x} - v_n(t, x - tU) \right] = 0. \end{aligned} \quad (48)$$

The subsequent analysis will model $v_n(t, x)$ as a two-dimensional random field of Gaussian distribution, with zero mean and an *a priori* known spatio-temporal covariance function. This random field is modeled as statistically stationary over space (i.e., array length) and time; hence, $E[v_n(t, x)v_n(t', x')] = c_v(t-t', x-x')$, the mathematical form of which reflects oceanic conditions and may be empirically determined. Moreover, this random field may be statistically correlated over time; an illustrative case of AR(1) will be analyzed in detail.

B. Discretizing the Small-Diameter Paidoussis Equation

Toward solving the above partial differential equation, the space-time discretization in [21] and [38] is herein adopted to represent array shape deformation as a finite-dimensional state-

⁵The small-diameter Paidoussis equation in [21] does not include the $-v_n(t, x - tU)$ term in (48) because [21] neglects the effects of oceanic currents.

space, with the towed-point induced temporal motion and the oceanic currents' space-time behavior as the system's driving inputs. Referring to Appendix A for details

$$\underbrace{\begin{bmatrix} y((i+1)H_t, 0) \\ y((i+1)H_t, H_x) \\ \vdots \\ y((i+1)H_t, (M_L-1)H_x) \end{bmatrix}}_{\mathbf{y}(i+1)} = \mathbf{F} \underbrace{\begin{bmatrix} y(iH_t, 0) \\ y(iH_t, H_x) \\ \vdots \\ y(iH_t, (M_L-1)H_x) \end{bmatrix}}_{\mathbf{y}(i)} + \underbrace{\begin{bmatrix} u_0(iH_t) \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{u}(i)} + H_t \underbrace{\begin{bmatrix} v_n(iH_t, -iH_x) \\ v_n(iH_t, -(i-1)H_x) \\ \vdots \\ v_n(iH_t, -(i-M_L+1)H_x) \end{bmatrix}}_{\mathbf{v}(i)} \quad (49)$$

where \mathbf{F} represents the transition matrix, i denotes the discrete-time index, M_L symbolizes the number of discretization steps along the array's length, and $H_x = L/(M_L - 1)$ and $H_t = H_x/U$, respectively, represent the discretization step sizes in space and time. The stochastic vector $\mathbf{u}(i)$ stands for tow-point induced (TPI) motion, where $\{u_0(iH_t)\}_{i=0}^{\infty}$ is a statistically stationary random sequence. Although [21] and [38] model $\{u_0(iH_t)\}_{i=0}^{\infty}$ as a temporally *uncorrelated* noise sequence, this present work will model $\{u_0(iH_t)\}_{i=0}^{\infty}$ in the mathematically more general and physically more realistic form of a temporally *correlated* random sequence. The stochastic vector $\mathbf{v}(i)$, which is not included in [21] and [38]⁶ but is newly introduced in this present work, represents the sea water's spatio-temporally correlated currents.

This above discretization scheme serves only as a mathematical technique to solve the partial differential equation in (47) but imposes no presumption on the physical behavior of the towed array. This discretization is to be distinguished from the piecewise linear model of array deformation [22], [23], [29], [33], [34], [40]. The latter array model assumes the towed array to behave like a concatenation of rigid linear segments, jointed at arbitrary angles. The above discretization makes no piecewise assumption regarding array deformation.

Although the transition matrix \mathbf{F} is strictly speaking a tridiagonal matrix (see Appendix A), empirical researchers [21], [24], [36] find it useful to use the first-order approximation

$$\mathbf{F} = \alpha(H_x)\mathbf{L} \quad (50)$$

where $\alpha(H_x)$ (with $0 \leq \alpha(H_x) \leq 1$) denotes the damping over a length H_x for TPI motion propagating down the array's length, and

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 \\ 1 & 0 & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 1 & 0 \end{bmatrix}. \quad (51)$$

With respect to the formulation of (50) and (51) developed in [21], [36], and [38], the present paper offers the following new

⁶Instead, [21] and [38] have a statistically stationary and spatio-temporally *uncorrelated* driving input as a "catch-all" function to include all modeling errors.

insight: The length-dependent damping coefficient α is physically related to H_x such that $\alpha(H_x + H'_x) = \alpha(H_x)\alpha(H'_x)$ for any arbitrary positive numbers H_x and H'_x ; hence, $\alpha(H_x)$ must take on the mathematical form of an exponential function of H_x , i.e.,

$$\alpha(H_x) = e^{-C_\alpha H_x} \quad (52)$$

when C_α represents an empirically measurable constant dependent only on the sea water's and the array's physical properties, namely, on C_t and C_n . A smaller C_α means less damping of towed-point induced or ocean-induced transverse motion along the array's length. For notational simplicity, the argument H_x will be omitted from α wherever possible.

C. Solving for the Towed-Array's Space-Time Shape Deformation

This subsection advances an original solution to (49) for $\mathbf{u}(i)$ and $\mathbf{v}(i)$ that is stationary and stochastically independent. Note that the tow-point's transverse displacement has been empirically determined to propagate down the array at close to the tow-boat's speed with little damping [21], [36]. Equation (49) has a bounded (in the least square sense) solution $\mathbf{y}(i)$, provided that $\|\mathbf{F}\| < 1$ for some matrix norm $\|\cdot\|$. Hence

$$\mathbf{y}(i) = \underbrace{\sum_{j=0}^{\infty} \mathbf{F}^j \mathbf{u}(i-j)}_{\mathbf{u}_F(i)} + \underbrace{\sum_{j=0}^{\infty} \mathbf{F}^j \mathbf{v}(i-j)}_{\mathbf{v}_F(i)}. \quad (53)$$

For the \mathbf{F} in (50), the condition $\|\mathbf{F}\| < 1$ is equivalent to $\alpha < 1$.

From the independence assumption for $\mathbf{u}(i)$ and $\mathbf{v}(i)$, the hydrophones' location uncertainties have the following spatio-temporal covariance

$$\underbrace{\text{cov}[\mathbf{y}(i)]}_{\mathbf{C}_y} = \underbrace{\text{cov}[\mathbf{u}_F(i)]}_{\mathbf{C}_{u_F}} + \underbrace{\text{cov}[\mathbf{v}_F(i)]}_{\mathbf{C}_{v_F}}. \quad (54)$$

The hydrophones' dislocation covariances may thus be determined once the specific form of the above two right-hand-side terms are known, perhaps from empirical measurements or databases. The above entities are not functions of i because of the temporal stationarity assumption.

The towing-vessel's motion and the oceanic currents represent statistically independent inputs to the towed-array system in (47) and (53); hence, the system's output equals (as expected) a sum of the system outputs due separately to either input. The main problem solved in the section is to express the above terms \mathbf{C}_{u_F} and \mathbf{C}_{v_F} in terms of the covariance of the TPI motion \mathbf{C}_u and covariance matrices of instantaneous fluid speeds \mathbf{C}_v , which are assumed to be known.

For hydrophones nominally at $\{x_1, \dots, x_L\}$ from the tow-point, where $x_m = \xi_m H_x$ for integers $\{\xi_m, m = 1, \dots, L\}$, define $[\mathbf{C}_y]_{\xi_\ell, \xi_m} \stackrel{\text{def}}{=} \text{cov}[y(t, x_\ell), y(t, x_m)]$ for $\ell, m = 1, \dots, L$. Notice that $[\mathbf{C}_y]_{\xi_\ell, \xi_m}$ is the (ℓ, m) th element of the desired matrix $\mathbf{\Omega}_{\Delta y}$. Under assumption (54), $[\mathbf{C}_y]_{\xi_\ell, \xi_m}$ may be expressed as $[\mathbf{C}_{u_F}]_{\xi_\ell, \xi_m} + [\mathbf{C}_{v_F}]_{\xi_\ell, \xi_m}$, where the first term is due to TPI motion, and the second term is due to oceanic currents. The

form of $[\mathbf{C}_{\mathbf{u}_F}]_{\xi_\ell, \xi_m}$ remains to be derived from the spatio-temporal statistics of \mathbf{u}_F as well as $[\mathbf{C}_{\mathbf{v}_F}]_{\xi_\ell, \xi_m}$ from the spatio-temporal statistics of \mathbf{v}_F . Each term is to be studied separately below, with detailed attention given to the particular illustrative case of first-order autoregressive auto-covariances for the TPI motion and of oceanic current velocities. The resulting covariances $[\mathbf{C}_y]_{\xi_\ell, \xi_m}$ turn out to be largely independent of the space-time discretization, assuming the discretization to be sufficiently fine.

D. New Statistical Modeling of Tow-Point Induced Towed-Array Shape-Deformation

The towing-vessel's motion has been modeled in [21] as temporally uncorrelated, which may be physically unrealistic for high time-sampling rates. Instead, the analysis here allows arbitrary temporal correlation (but requires statistical stationarity) in the towing-vessel's motion. The following expresses $\mathbf{C}_{\mathbf{u}_F}$ in terms of the spatio-temporal statistics of \mathbf{u}_F , first for the general case of any statistically stationary (but otherwise arbitrary) spatio-temporal covariance and then for the special case of a first-order auto-regressive temporal covariance. All subsequent expressions will turn out to be independent of the discretization grid used earlier.

Assume that $\{u_0(iH_t)\}_{i=1}^\infty$ is statistically stationary, zero-mean, and with a Toeplitz covariance matrix \mathbf{C}_u containing elements $[\mathbf{C}_u]_{i,j} = E[u_0(iH_t)u_0(jH_t)]$ for $i, j = 1, \dots, M_L$. Hence

$$\begin{aligned} \mathbf{u}_F(i) &= \sum_{j=0}^{\infty} \mathbf{F}^j \mathbf{u}(i-j) \\ &= [u_0(iH_t), \alpha u_0((i-1)H_t), \dots, \alpha^{M_L-1} u_0((i-M_L+1)H_t)]^T \end{aligned} \quad (55)$$

and $\mathbf{u}_F(i)$ has the spatial covariance matrix $\mathbf{C}_{\mathbf{u}_F} = \mathbf{F}_\alpha \mathbf{C}_u \mathbf{F}_\alpha^T$, where $\mathbf{F}_\alpha = \text{diag}[1, \alpha, \alpha^2, \dots, \alpha^{M_L-1}]$. Thus

$$[\mathbf{C}_{\mathbf{u}_F}]_{\xi_\ell, \xi_m} = \alpha^{\xi_\ell + \xi_m - 2} [\mathbf{C}_u]_{\xi_\ell, \xi_m}, \quad \ell, m = 1, \dots, L. \quad (56)$$

The above equation constitutes this subsection's main contribution, relating the positional uncertainty's covariance function to the TPI motion's spatio-temporal covariance function.

For the special case where the TPI motion may be represented as an AR(1) temporally random process,⁷

$$[\mathbf{C}_u]_{\xi_\ell, \xi_m} = \sigma_u^2 e^{-C_u |\xi_\ell - \xi_m| H_t}, \quad \ell, m = 1, \dots, L \quad (57)$$

where C_u and σ_u^2 represent constants that may be empirically measured: σ_u^2 denotes the variance of the TPI motion, and C_u characterizes the time correlation of the TPI motion. Given C_u , the time delay Δt [in which the correlation between $u_0(t)$ and $u_0(t + \Delta t)$ decays to 1/10] equals $\log 10 / C_u$. Combining (52), (56), and (57)

$$\begin{aligned} [\mathbf{C}_{\mathbf{u}_F}]_{\xi_\ell, \xi_m} &= [e^{-C_u H_x}]^{(\xi_\ell + \xi_m - 2)} \sigma_u^2 e^{-C_u |\xi_\ell - \xi_m| H_t} \\ &= \sigma_u^2 e^{-C_u (x_\ell + x_m) - C_u |x_\ell - x_m| / U} e^{-2C_u H_x} \\ &\rightarrow \sigma_u^2 e^{-C_u (x_\ell + x_m) - C_u |x_\ell - x_m| / U}, \quad H_x \rightarrow 0 \end{aligned} \quad (58)$$

and may be substituted into (54) to give the towed-array's space-time shape-deformation covariance. The expression on

⁷For covariance functions of the general AR(i) form or the general ARMA(i, j) form, see [37].

the right-hand side of (58) is the covariance element for an infinitely fine discretization grid.

E. Statistical Modeling of Ocean-Induced Array Shape Deformation

The following will express $\mathbf{C}_{\mathbf{v}_F}$ in terms of the spatio-temporal statistics of $v_n(t, x)$, first for the general case stipulating only statistical stationarity and then for the special case where the auto-correlation of $v_n(t, x)$ is AR(1) in both space and time.

Define $\mathbf{C}_v(j) \stackrel{\text{def}}{=} E\{\mathbf{v}(t)\mathbf{v}^T(t+j)\}$, for $j = 0, \pm 1, \pm 2, \dots$. With the spatio-temporal covariance function $c_v(t, x)$ of the space-time random field $v_n(t, x)$ characterizing fluid flow, (49) implies $[\mathbf{C}_v(j)]_{\ell, m} = H_t^2 c_v(jH_t, (m - \ell - j)H_x)$ for $\ell, m = 1, \dots, M_L$. Define

$$\check{\mathbf{C}}_v(j) = \sum_{m=0}^{\infty} \mathbf{F}^m \mathbf{C}_v(j) (\mathbf{F}^m)^T = \sum_{m=0}^{M_L-1} \mathbf{F}^m \mathbf{C}_v(j) (\mathbf{F}^m)^T. \quad (59)$$

The last equality holds because $\mathbf{F}^m = 0$ for $m \geq M_L$, given \mathbf{F} defined in (50). As $\mathbf{C}_v(j)$ is Toeplitz, some straightforward manipulation gives

$$[\check{\mathbf{C}}_v(j)]_{\ell, m} = [\mathbf{C}_v(j)]_{\ell, m} \frac{1 - \alpha^{\min\{\ell, m\}}}{1 - \alpha}, \quad \ell, m = 1, \dots, M_L. \quad (60)$$

Referring to (50), (53), and (59)

$$\begin{aligned} \mathbf{C}_{\mathbf{v}_F} &= \sum_{\ell=0}^{\infty} \sum_{m=0}^{\infty} \mathbf{F}^\ell E[\mathbf{v}(i-\ell)\mathbf{v}^T(i-m)] (\mathbf{F}^m)^T \\ &= \sum_{\ell=0}^{\infty} \sum_{m=0}^{\infty} \mathbf{F}^\ell \mathbf{C}_v(\ell-m) (\mathbf{F}^m)^T \\ &= \check{\mathbf{C}}_v(0) + \sum_{m=1}^{M_L-1} [\mathbf{F}^m \check{\mathbf{C}}_v(m) + \check{\mathbf{C}}_v(-m) (\mathbf{F}^m)^T]. \end{aligned} \quad (61)$$

The above equation represents this subsection's key contribution, relating the positional uncertainty's covariance function to the oceanic flow's space-time covariance function.

Consider the illustrative case of $v_n(t, x)$ being AR(1) in both space and time

$$c_v(t, x) = \sigma_v^2 e^{-C_{vt}t - C_{vx}x} \quad (62)$$

where σ_v^2 , C_{vt} , and C_{vx} are physical constants that may be empirically measured. σ_v^2 is the total variance of oceanic flows, whereas C_{vt} and C_{vx} determine the correlation length of the random field $v_n(t, x)$ in reference to, respectively, time and space; cf. (57).

Referring to Appendix B for details, for $\xi_\ell = x_\ell/H_x$, $\xi_m = x_m/H_x$, $x_\ell < x_m$, and $H_x \rightarrow 0$,

$$\begin{aligned} & [\mathbf{C}_{\mathbf{v}_F}]_{\xi_\ell, \xi_m} \\ & \approx \frac{\sigma_v^2 e^{-C_{vx}|x_\ell - x_m|}}{C_\alpha U} \left\{ \frac{1}{C_{vt}} \left[e^{-(C_\alpha + C_{vt}/U)x_\ell} + e^{-(C_\alpha + C_{vt}/U)x_m} \right. \right. \\ & \quad \left. \left. - e^{-C_\alpha x_\ell} - e^{-(C_\alpha + C_{vt}/U)(x_m - x_\ell)} \right] \right. \\ & \quad \left. + \frac{1}{C_\alpha U + C_{vt}} \left[2e^{-(C_\alpha + C_{vt}/U)x_\ell} - e^{-(C_\alpha + C_{vt}/U)x_m} \right. \right. \\ & \quad \left. \left. - e^{-C_\alpha x_\ell} (1 - e^{-(C_\alpha + C_{vt}/U)(x_m - x_\ell)}) \right] \right\} \quad (63) \end{aligned}$$

and may be substituted into (54) to give the towed-array's space-time shape-deformation covariance.

F. Summary of Notations of Constants

For easy reference, the following summarizes the notations used in the above two subsections to describe the AR(1,1) oceanic-current model and the AR(1) TPI-motion model.

σ_u^2	variance of the TPI motion;
C_u	constant characterizing the temporal correlation of the TPI motion (57);
σ_v^2	variance of the oceanic flows;
C_{vt}	constant characterizing the temporal correlation of the oceanic flows (62);
C_{vx}	constant characterizing the per-unit-length correlation of the TPI motion (62);
H_α	constant in the exponential dumping model (52).

V. NUMERICAL EXAMPLES

A. Example 1: Variance of Positional Deviation Along the Towed-Array

Fig. 3 plots $[\mathbf{C}_{\mathbf{u}_F}]_{x,x}$ and $[\mathbf{C}_{\mathbf{v}_F}]_{x,x}$ along the array's length x when the towing-vessel's motion is an AR(1) temporal stochastic process and when the oceanic currents may be modeled as a spatio-temporal AR(1) space-time stochastic process. The simulation parameters are as follows: The towed-array has the damping parameter $C_\alpha = 0.008 \text{ (m}^{-1}\text{)}$ (corresponding to a damping factor $\alpha = 0.95$ per 6.25 m of tow-array length [24]), the TPI-motion has the variance $\sigma_u^2 = 1 \text{ (m}^2\text{)}$ and $C_u = 1 \text{ (s}^{-1}\text{)}$, the ocean-induced motion has variance $\sigma_v^2 = 0.01 \text{ (m}^2\text{)}$ and $C_{vx} = 1 \text{ (m}^{-1}\text{)}$ and $C_{vt} = 1 \text{ (s}^{-1}\text{)}$.

As x increases (i.e., further from the tow point), the TPI motion becomes less significant, but the oceanic flow becomes more important. A faster tow speed does not affect $[\mathbf{C}_y]_{x,x}$, which is intuitively reasonable as the angle $\beta(t, x)$ between the array and surrounding fluid's relative velocity also decreases with increasing tow-speed, thereby diminishing the influence of the oceanic currents.

The following examples illustrate the dependence of the CRB of u on various physical parameters in the far-field deep-sea single-source scenario, where the towed-array has uniform half-wavelength spaced hydrophones moving along the x -axis.

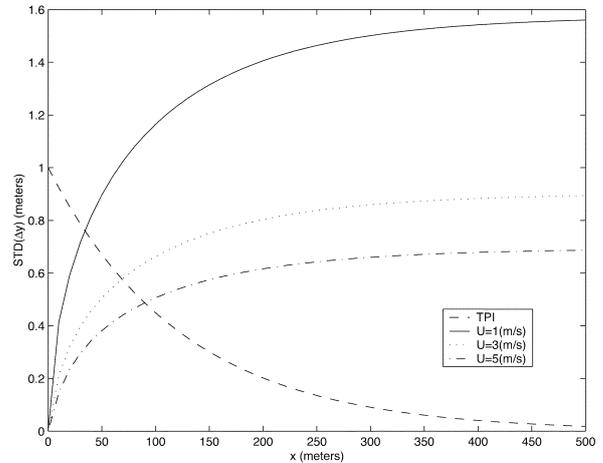


Fig. 3. Standard deviation of positional deviation along the towed-array. Dashed line: Influence from TPI motion. Solid line: Influence from fluid-flows for tow-speeds shown. Referring to (52), (57), and (62), $C_\alpha = 0.008 \text{ (m}^{-1}\text{)}$, $\sigma_u^2 = 1 \text{ (m}^2\text{)}$, $C_u = 1 \text{ (s}^{-1}\text{)}$, $\sigma_v^2 = 0.01 \text{ (m}^2\text{)}$, $C_{vt} = 1 \text{ (s}^{-1}\text{)}$, and $C_{vx} = 1 \text{ (m}^{-1}\text{)}$. The tow speed is $U = 1, 3$ and 5 (m/s) , respectively.

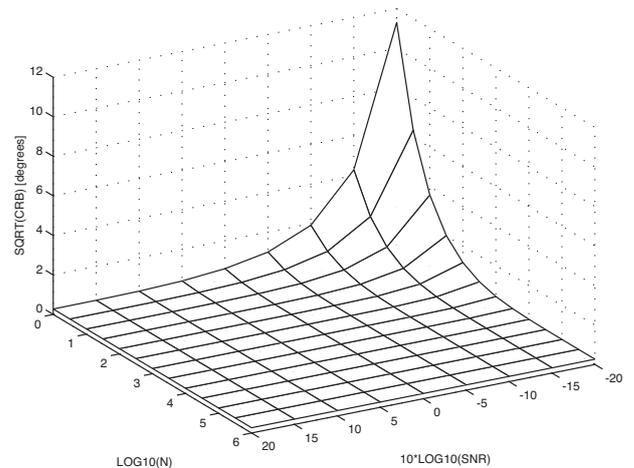


Fig. 4. Square root of CRB (in degrees) versus the number of observation snapshots N and versus the SNR Ω_s/σ_n^2 .

B. Example 2: CRB versus the SNR and Number of Snapshots n

Fig. 4 plots the square root of the CRB (in degrees) versus the SNR $= \Omega_s/\sigma_n^2$ and the number of snapshots N . The array has $L = 25$ hydrophones, equispaced at 6.25 m (corresponding to a frequency of 120 Hz [24]) and towed with speed $U = 3 \text{ m/s}$. The hydrophones' phase uncertainties are uncorrelated with a 3° standard deviation, implying $\mathbf{\Omega}_{\Delta\varphi} = (3\pi/180)^2 \mathbf{I} \text{ (rad}^2\text{)}$. All other simulation parameters remain the same as in Example 1. Fig. 4 shows that for SNR $> 0 \text{ dB}$, the CRB approximates the large- N limit-CRB, even at $N = 1$. Moreover, the limit-CRB does not depend on the SNR.

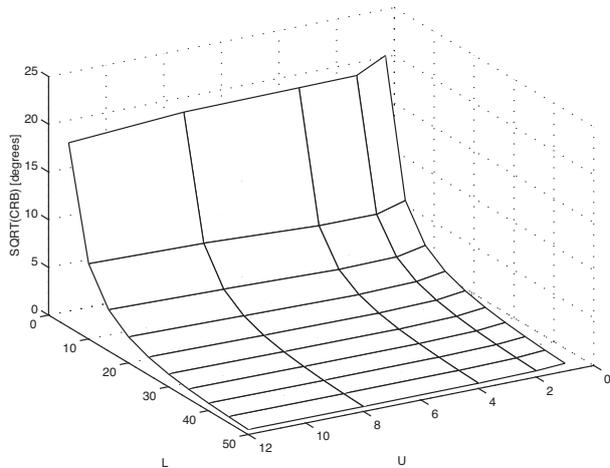


Fig. 5. Square root of the limit-CRB (with $N \rightarrow \infty$) versus the number of hydrophones L and versus the tow-speed U .

C. Example 3: CRB Versus Number of Hydrophones l and Tow-Speed u

Fig. 5 plots the large- N limit-CRB versus the towed-array's number of constituent hydrophones L and the tow-speed U (m/s^{-1}). All other simulation parameters remain identical as in Example 2. As expected, the CRB in Fig. 5 decreases (i.e., the potentially achievable accuracy improves) with more hydrophones and a faster tow-speed.

With at least 30 hydrophones in the towed-array and at low tow-speed, adding more hydrophones to the towed-array (while maintaining the towed-array's half-wavelength inter-hydrophone spacing) will offer more improvement in the direction-finding's CRB when the tow-speed is faster than when the tow-speed is slower.

D. Example 4: CRB versus the TPI Parameters

Fig. 6 plots the large- N limit-CRB versus the TPI-motion parameter C'_u (s^{-1}) and the damping parameter C'_α (m^{-1}) in the absence of oceanic currents and hydrophone gain/phase uncertainties. The TPI motion is statistically independent but identically distributed along the y - and z -axes. The limit-CRB, which is plotted in Fig. 6, is proportional to the TPI transversal motion's variance σ_u^2 but is independent of the SNR. The CRB also depends significantly on C_α but only slightly on C_u . Recall that $C_u \approx 0$ (m^{-1}) means low-frequency TPI-motions (say, due to the towing vessel's slow maneuvers) and a moderate C_u corresponds to TPI-motions similar to white noise. For fixed C_α , the CRB has a broad plateau with respect to C_u .

E. Example 5: CRB Versus the Oceanic-Current Parameters

Fig. 7 plots the large- N limit-CRB in the absence of TPI motion and hydrophone gain/phase uncertainties versus the oceanic current parameters C_{vx} and C_{vt} for oceanic current that is statistically independent and identically distributed in the vertical and horizontal directions. All other simulation parameters remain the same, as in the previous example. The CRB, which is plotted in Fig. 7, decreases with increasing C_{vx} and with increasing C_{vt} , as expected. For $C_{vx} \approx 0$ (m^{-1}) and $C_{vt} \approx 0$

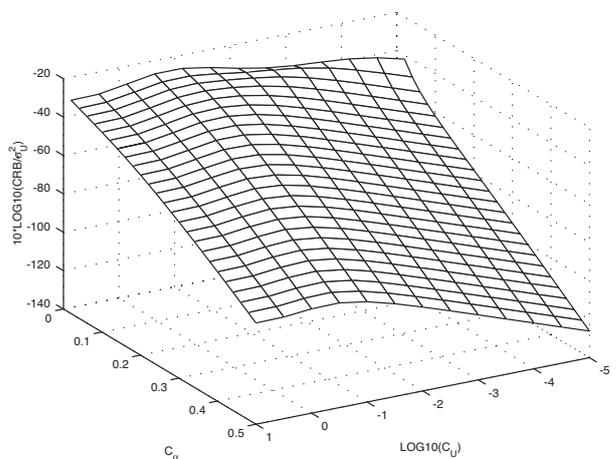


Fig. 6. Limit-CRB (with $N \rightarrow \infty$) per unit TPI-motion's variance (in the absence of oceanic currents) versus the TPI-motion parameter C_u and versus the towed-array transversal movement's damping parameter C_α .

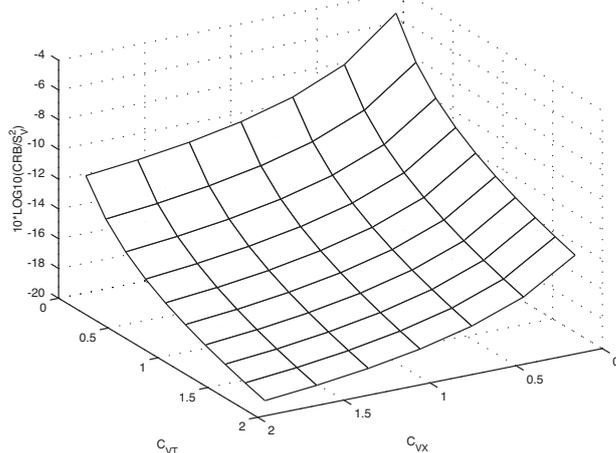


Fig. 7. Limit-CRB (with $N \rightarrow \infty$) per unit oceanic current's variance (assuming no TPI motion) plotted versus oceanic motion's AR(1) space-time parameters C_{vt} and C_{vx} .

(s^{-1}), the oceanic current's velocities are more correlated over space and over time, implying that the instantaneous velocities $v_n(t, x)$ and $v_n(t + \Delta t, x + \Delta x)$ are more likely to have the same sign, and array deformation would consequentially have a larger variance.

VI. CONCLUSION

This work represents an initial step to bridge a serious literature gap in deformed towed-array direction-finding performance analysis by incorporating into the statistical measurement model several essential fluid-mechanics considerations while deriving mathematically rigorous quantitative expressions and qualitative insights into how DOA estimation may depend on physically measurable sources of array deformation. Among various derived properties of the far-field deep-water single-source CRB, especially noteworthy is its independence from the hydrophones' gain uncertainties.

$$\begin{aligned}
[\mathbf{C}_{\mathbf{v}_F}]_{k\ell} &= [\mathbf{C}_{\mathbf{v}_2}(0)]_{k\ell} + \sum_{m=1}^{M_L-1} [\mathbf{F}^k \mathbf{C}_{\mathbf{v}_2}(m)] + \mathbf{C}_{\mathbf{v}_2}(-m)(\mathbf{F}^m)^T]_{k\ell} \\
&= [\mathbf{C}_{\mathbf{v}_2}(0)]_{k\ell} + \sum_{m=1}^{k-1} \alpha^m [\mathbf{C}_{\mathbf{v}_2}(m)]_{k-m,\ell} + \sum_{m=1}^{\ell-1} \alpha^m [\mathbf{C}_{\mathbf{v}_2}(-k)]_{k,\ell-m} \\
&= H_t^2 \left\{ \mathbf{C}_{\mathbf{v}}(0, (\ell-k)H_x) + \sum_{m=1}^{k-1} \alpha^m \mathbf{C}_{\mathbf{v}}(mH_t, (\ell-k)H_x) \frac{1-\alpha^{k-m}}{1-\alpha} \right. \\
&\quad \left. + \sum_{m=1}^{\ell-1} \alpha^m \mathbf{C}_{\mathbf{v}}(mH_t, (\ell-k)H_x) \frac{1-\alpha^{\min(\ell-m,k)}}{1-\alpha} \right\} \\
&= H_t^2 \left\{ \mathbf{C}_{\mathbf{v}}(0, (\ell-k)H_x) + \sum_{m=1}^{k-1} \alpha^m \mathbf{C}_{\mathbf{v}}(mH_t, (\ell-k)H_x) \frac{1-\alpha^{k-m}}{1-\alpha} \right. \\
&\quad \left. + \sum_{m=1}^{\ell-k} \alpha^m \mathbf{C}_{\mathbf{v}}(mH_t, (\ell-k)H_x) \frac{1-\alpha^k}{1-\alpha} \right. \\
&\quad \left. + \sum_{m=\ell-k+1}^{\ell} \alpha^m \mathbf{C}_{\mathbf{v}}(mH_t, (\ell-k)H_x) \frac{1-\alpha^{\ell-m}}{1-\alpha} \right\} \\
&= H_t^2 \sigma_v^2 e^{-C_{vx}(\ell-k)H_x} \left\{ 1 + \sum_{m=1}^{k-1} \alpha^m \frac{1-\alpha^{k-m}}{1-\alpha} e^{-C_{vt}mH_t} \right. \\
&\quad \left. + \sum_{m=1}^{\ell-k} \alpha^m \frac{1-\alpha^k}{1-\alpha} e^{-C_{vt}mH_t} + \sum_{m=\ell-k+1}^{\ell} \alpha^m \frac{1-\alpha^{\ell-m}}{1-\alpha} e^{-C_{vt}mH_t} \right\}. \tag{70}
\end{aligned}$$

APPENDIX A

Discretization of the Paidoussis equation is achieved [21], [36] by substituting $t = iH_t$ and $x = mH_x$ in the Euler approximation

$$\frac{\partial y(t, x)}{\partial t} \approx \frac{\mathbf{y}_m(i+1) - \mathbf{y}_m(i)}{H_t} \tag{64}$$

$$\frac{\partial y(t, x)}{\partial x} \approx \frac{\mathbf{y}_m(i) - \mathbf{y}_{m-1}(i)}{H_x} \tag{65}$$

$$\frac{\partial^2 y(t, x)}{\partial x^2} \approx \frac{\mathbf{y}_m(i) - 2\mathbf{y}_{m-1}(i) + \mathbf{y}_{m-2}(i)}{H_x^2} \tag{66}$$

where $\mathbf{y}_m(i)$ denotes the $(m+1)$ th element of $\mathbf{y}(i)$ in (49). After some manipulations

$$\begin{aligned}
\mathbf{y}_m(i+1) &= \mathbf{y}_{m-1}(i) + H_t v_n(iH_t, mH_x - iH_t)U \\
&\quad + \frac{C_t}{C_n} \frac{L - mH_x}{H_x} [\mathbf{y}_m(i) - 2\mathbf{y}_{m-1}(i) + \mathbf{y}_{m-2}(i)]. \tag{67}
\end{aligned}$$

The last equation may be written in matrix form as (49), where

$$\mathbf{F} = \mathbf{L} + \mathbf{D}_D(\mathbf{I} - 2\mathbf{L} + \mathbf{L}^2) \tag{68}$$

where \mathbf{L} is defined in (51), and $\mathbf{D}_D = \text{diag}(\delta_1, \delta_2, \dots, \delta_{M_L})$ is a diagonal matrix with

$$\delta_m = \frac{C_t}{C_n} \frac{L - mH_x}{H_x}, \quad m = 1, \dots, M_L. \tag{69}$$

In [21], [24], and [36] replacing (68) by $\mathbf{F} = \alpha\mathbf{L}$ with a scalar correcting factor α is suggested.

APPENDIX B

Let $x_1 = kH_x$ and $x_k = \ell H_x$, $k \leq \ell$. Then, we have (70), shown at the top of the page. Further simplification is obtained by the Taylor series expansion, which is valid for small H_x

$$1 - \alpha = 1 - e^{-C_\alpha H_x} \approx C_\alpha H_x = C_\alpha U H_t$$

and for an arbitrary $|\gamma| < 1$, it holds that

$$\sum_{m=1}^{k-1} \gamma^m = \gamma \frac{1 - \gamma^{k-1}}{1 - \gamma}.$$

REFERENCES

- [1] M. P. Paidoussis, "Dynamics of flexible slender cylinder in axial flow, Part I: Theory," *J. Fluid Mech.*, vol. 26, pp. 717-736, 1966.
- [2] —, "Dynamics of flexible slender cylinder in axial flow, Part II: Experiment," *J. Fluid Mech.*, vol. 26, pp. 737-751, 1966.
- [3] —, "Dynamics of cylindrical structures subject to axial flow," *J. Sound Vibr.*, vol. 29, pp. 365-385, 1973.
- [4] M. J. Hinich, "Bearing estimation using a perturbed linear array," *J. Acoust. Soc. Amer.*, vol. 61, no. 6, pp. 1540-1544, June 1977.
- [5] P. M. Schultheiss and J. P. Ianniello, "Optimum range & bearing estimation with randomly perturbed arrays," *J. Acoust. Soc. Amer.*, vol. 68, no. 1, pp. 167-173, July 1980.
- [6] N. L. Owsley, "Shape estimation for a flexible underwater cable," in *Proc. IEEE EASCON Conf.*, 1981.
- [7] R. M. Kennedy, "Crosstrack dynamics of a long cable towed in the ocean," in *Proc. IEEE Oceans Conf.*, 1981, pp. 966-981.
- [8] R. M. Kennedy and E. S. Strahan, "A linear theory of transverse cable dynamics at low frequencies," Naval Underwater Syst. Center, New London, CT, NUSC Tech. Rep. 6463, 1981.
- [9] Y. Rockah and P. M. Schultheiss, "Array shape calibration using sources in unknown locations—Part I: Far-field sources," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 286-299, Mar. 1987.

- [10] —, "Array shape calibration using sources in unknown locations—Part II: Near-field sources and estimator implementation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 724–735, June 1987.
- [11] A. P. Dowling, "The dynamics of towed flexible cylinders, Part I: Neutrally buoyant elements," *J. Fluid Mech.*, vol. 187, pp. 507–532, 1988.
- [12] —, "The dynamics of towed flexible cylinders, Part II: Negatively buoyant elements," *J. Fluid Mech.*, vol. 187, pp. 533–571, 1988.
- [13] E. C. van Ballegoijen, G. W. M. van Mierlo, C. van Schooneveld, P. P. M. van der Zalm, A. T. Parsons, and N. H. Field, "Measurement of towed array position, shape and altitude," *IEEE J. Ocean. Eng.*, vol. 14, pp. 375–383, Oct. 1989.
- [14] A. J. Weiss and B. Friedlander, "Array shape calibration using sources in unknown locations — A maximum likelihood approach," *IEEE Trans. Signal Processing*, vol. 37, pp. 1958–1966, Dec. 1989.
- [15] B. Wahlberg, B. Ottersten, and M. Viberg, "Robust signal parameter estimation in the presence of array perturbations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1991, pp. 3277–3280.
- [16] B. E. Howard and J. M. Syck, "Calculation of the shape of a towed underwater acoustic array," *IEEE J. Ocean. Eng.*, vol. 17, pp. 193–203, Apr. 1992.
- [17] A. L. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors, Part I: The MUSIC algorithm," *IEEE Trans. Signal Processing*, vol. 40, pp. 1758–1773, July 1992.
- [18] S. Marcos, "Calibration of a distorted towed array using a propagator operator," *J. Acoust. Soc. Amer.*, pt. 1, vol. 93, no. 4, pp. 1987–1994, Apr. 1993.
- [19] S. Li and P. M. Schultheiss, "Depth measurement of remote sources using multipath propagation," *IEEE J. Ocean. Eng.*, vol. 18, pp. 379–387, Oct. 1993.
- [20] B. H. Maranda and J. A. Fawcett, "The localization accuracy of a horizontal array observing a narrowband target with partial coherence," *IEEE J. Ocean. Eng.*, vol. 18, pp. 466–473, Oct. 1993.
- [21] D. A. Gray, B. D. O. Anderson, and R. R. Bitmead, "Towed array shape estimation using kalman filters — Theoretical models," *IEEE J. Ocean. Eng.*, vol. 18, pp. 543–556, Oct. 1993.
- [22] B. G. Quinn, R. F. Barrett, P. J. Kootsookos, and S. J. Searle, "The estimation of the shape of an array using a hidden Markov model," *IEEE J. Ocean. Eng.*, vol. 18, pp. 557–564, Oct. 1993.
- [23] B. G. Ferguson, "Remedying the effects of array shape distortion on the spatial filtering of acoustic data from a line array of hydrophones," *IEEE J. Ocean. Eng.*, vol. 18, pp. 565–571, Oct. 1993.
- [24] J. L. Riley and D. A. Gray, "Towed array shape estimation using kalman filters—Experimental investigations," *IEEE J. Ocean. Eng.*, vol. 18, pp. 572–581, Oct. 1993.
- [25] D. E. Wahl, "Towed array shape estimation using frequency-wavenumber data," *IEEE J. Ocean. Eng.*, vol. 18, pp. 582–590, Oct. 1993.
- [26] G. S. Edelson and D. W. Tufts, "Passive localization using a spatially-referenced towed array," in *Proc. IEEE Oceans Conf.*, vol. 2, 1993, pp. 218–223.
- [27] N. C. Wyeth, "Methods of array element localization for a towed underwater acoustic array," *IEEE J. Ocean. Eng.*, vol. 19, pp. 128–133, Jan. 1994.
- [28] M. Viberg and A. L. Swindlehurst, "Analysis of the combined effects of finite samples and model errors on array processing performance," *IEEE Trans. Signal Processing*, vol. 42, pp. 3073–3083, Nov. 1994.
- [29] —, "A bayesian approach to auto-calibration for parametric array signal processing," *IEEE Trans. Signal Processing*, vol. 42, pp. 3495–3507, Dec. 1994.
- [30] S. Narasimhan and J. L. Krolik, "A Cramér-Rao bound for source range estimation in a random ocean waveguide," in *Proc. IEEE Workshop Stat. Signal Array Processing*, 1994, pp. 309–312.
- [31] B. Porat, *Digital Processing of Random Signals, Theory & Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [32] J. Tabrikjan and H. Messer, "Three-dimensional source localization in a waveguide," *IEEE Trans. Signal Processing*, vol. 44, pp. 1–13, Jan. 1996.
- [33] J. J. Smith, Y. H. Leung, and A. Cantoni, "The Cramér-Rao lower bound for towed array shape estimation with a single source," *IEEE Trans. Signal Processing*, vol. 44, pp. 1033–1036, Apr. 1996.
- [34] —, "The partitioned eigenvector method for towed array shape estimation," *IEEE Trans. Signal Processing*, vol. 44, pp. 2273–2283, Sept. 1996.
- [35] S. K. Srivastava and C. Ganapathy, "Experimental investigations on loop maneuver of underwater towed cable array system," *Ocean Eng.*, vol. 25, no. 1, pp. 85–102, Jan. 1998.
- [36] J. M. Goldberg, "Joint direction-of-arrival and array-shape tracking for multiple moving targets," *IEEE J. Ocean. Eng.*, vol. 23, pp. 118–126, Apr. 1998.
- [37] T. Söderström, J. Ježek, and V. Kučera, "An efficient and versatile algorithm for computing the covariance function of an ARMA process," *IEEE Trans. Signal Processing*, vol. 46, pp. 1591–1600, June 1998.
- [38] N. V. Nikitakos, A. K. Leros, and S. K. Katsikas, "Towed array shape estimation using multimodel partitioning filters," *IEEE J. Ocean. Eng.*, vol. 23, no. 4, pp. 118–126, Oct. 1998.
- [39] D. E. Calkins, "Metamodel-based towed system simulation," *Ocean Eng.*, vol. 26, no. 11, pp. 1183–1247, Nov. 1999.
- [40] A. Jakoby, J. Goldberg, and H. Messer, "Source localization in shallow water in the presence of sensor location uncertainty," *IEEE J. Ocean. Eng.*, vol. 25, pp. 331–336, July 2000.
- [41] F. Gini and R. Reggiannini, "On the use of Cramér-Rao-like bounds in the presence of random nuisance parameters," *IEEE Trans. Commun.*, vol. 48, pp. 2120–2126, Dec. 2000.



Petr Tichavský (M'98) graduated in 1987 from the Czech Technical University, Prague, Czechoslovakia. He received the Ph.D. degree in theoretical cybernetics from the Czechoslovak Academy of Sciences, Prague, in 1992.

Since that time, he has been with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague. He is author and co-author of research papers in the areas of sinusoidal frequency/frequency-rate estimation, adaptive filtering, and tracking of time-varying signal parameters and algorithm-independent bounds on achievable performance. His recent research interests include independent component analysis and blind signal separation and signal processing for wireless communications.

Dr. Tichavský received the Fulbright grant for a ten-month fellowship at the Department of Electrical Engineering, Yale University, New Haven, CT, in 1994. In 2002, he received the Otto Wichterle Award from Academy of Sciences of the Czech Republic. Since 2002, he has served as associate editor of the *IEEE SIGNAL PROCESSING LETTERS*.



Kainam Thomas Wong (SM'01) received the B.S.E. (Chem. E.) degree from the University of California, Los Angeles, in 1985, the B.S.E.E. degree from the University of Colorado, Boulder, in 1987, the M.S.E.E. degree from Michigan State University, East Lansing, in 1990, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1996.

He was a manufacturing engineer at the General Motors Technical Center, Warren, MI, from 1990 to 1991 and a Senior Professional Staff Member at the Johns Hopkins University Applied Physics Laboratory, Laurel, MD, from 1996 to 1998. He was an Assistant Professor with Nanyang Technological University, Singapore, in 1998 and with the Chinese University of Hong Kong from 1998 to 2001. He has been with the University of Waterloo, Waterloo, ON, Canada, since 2001. He was a contributing author of about 70 articles for the telecommunications section of the inaugural edition of the *CRC Dictionary of Pure and Applied Physics*. His research interest is signal processing for communications and sensor-array signal processing.

Dr. Wong has served on the Organizing Committees of the 2000 IEEE International Symposium on Circuits and Systems (ISCAS) and the 2004 IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP). He also serves on the Technical Program Committees of the 2002–2004 IEEE International Conference on Communications (ICC), the 2002 and 2003 IEEE Global Telecommunications Conference (GlobeCom), plus a dozen other conferences. He received the Premier's Research Excellence Award by the government of Ontario in 2003. He is listed in Marquis' *Who's Who in the World*, *Who's Who in America*, and *Who's Who in Science and Engineering*.

Performance Analysis of the FastICA Algorithm and Cramér–Rao Bounds for Linear Independent Component Analysis

Petr Tichavský, *Senior Member, IEEE*, Zbyněk Koldovský, *Member, IEEE*, and Erkki Oja, *Fellow, IEEE*

Abstract—The FastICA or fixed-point algorithm is one of the most successful algorithms for linear independent component analysis (ICA) in terms of accuracy and computational complexity. Two versions of the algorithm are available in literature and software: a one-unit (deflation) algorithm and a symmetric algorithm. The main result of this paper are analytic closed-form expressions that characterize the separating ability of both versions of the algorithm in a local sense, assuming a “good” initialization of the algorithms and long data records. Based on the analysis, it is possible to combine the advantages of the symmetric and one-unit version algorithms and predict their performance. To validate the analysis, a simple check of saddle points of the cost function is proposed that allows to find a global minimum of the cost function in almost 100% simulation runs. Second, the Cramér–Rao lower bound for linear ICA is derived as an algorithm independent limit of the achievable separation quality. The FastICA algorithm is shown to approach this limit in certain scenarios. Extensive computer simulations supporting the theoretical findings are included.

Index Terms—Blind source separation, independent component analysis (ICA), Cramér–Rao lower bound.

I. INTRODUCTION

BLIND SOURCE separation (BSS), which consists of recovering original signals from their mixtures when the mixing process is unknown, has been a widely studied problem in signal processing for the last two decades (for a review, see [1]). Independent component analysis (ICA), a statistical method for signal separation [2], [3], is also a well-known issue in the community. Its aim is to transform the mixed random signals into source signals or components that are as mutually

independent as possible. There are a number of methods intended to solve related problems such as blind deconvolution and blind equalization [4]–[6].

One of the most widely used ICA algorithms for the linear mixing model is FastICA, a fixed-point algorithm first proposed by Hyvärinen and Oja [7], [8]. It is based on the optimization of a nonlinear contrast function measuring the non-Gaussianity of the sources. A widely used contrast function both in FastICA and in many other ICA algorithms is the kurtosis [9]–[11]. This approach can be considered as an extension of the algorithm by Shalvi and Weinstein [6].

There are two varieties of the FastICA algorithm: the deflation, or one-unit algorithm, and the symmetric algorithm. The deflation approach, which is common for many other ICA algorithms [9], estimates the components successively under orthogonality conditions. The symmetric algorithm estimates the components in parallel. This consists of parallel computation of the one-unit updates for each component, followed by subsequent symmetric orthogonalization of the estimated demixing matrix after each iteration. A version of FastICA for complex valued signals was proposed in [12].

An essential question is the convergence of the FastICA algorithm. This can be approached from two directions. First, assuming an ideal infinitely large sample, theoretical expectations for the contrast functions such as the kurtosis can be used in the analysis. Then, the contrast function and the algorithm itself become deterministic, and questions such as asymptotic stability of the extrema and the convergence speed can be discussed. For the kurtosis cost function and the one-unit algorithm, this analysis was done in [7], showing cubic convergence. For a general cost function, the convergence speed is at least quadratic, as shown in [8] (see also [3]). The monotonic convergence and the speed for a general cost function for the related gradient algorithm was considered in [13]. For the kurtosis cost function and the symmetric FastICA algorithm, the cubic convergence was proven in [14] (see also [15]). Different properties of the one-unit version have been illustrated by computer simulations in [16] where the accuracy is also shown to be very good in most cases.

The second question of convergence considers the behavior of the algorithm for a finite sample, which is the practical case. Then, the theoretical expectations in the contrast functions are replaced by sample averages. This results in errors in the estimator for the demixing matrix. A classical measure of the error is the asymptotic variance of the matrix elements. The goal of designing an ICA algorithm is then to make this error as small

Manuscript received November 26, 2004; revised May 19, 2005. The work of P. Tichavský was supported by Ministry of Education, Youth and Sports of the Czech Republic through the project 1M0572. Part of this paper (on the CRB) was presented at International Conference on Acoustics, Speech and Signal Processing (ICASSP), Philadelphia, PA, 2005, and another part (analysis of FastICA) was presented at the 13th IEEE/SP Statistical Signal Processing Workshop, Bordeaux, France, July, 17–20, 2005. The associate editor coordinating the review of this paper and approving it for publication was Prof. Jonathon Chambers.

P. Tichavský is with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, P 182 08 Prague 8, Czech Republic (e-mail: p.tichavsky@ieee.org; website: <http://si.utia.cas.cz/Tichavsky.html>).

Z. Koldovský was with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, 182 08 Prague 8, Czech Republic. He is now with the Department of Electronics and Signal Processing, Technical University of Liberec, 461 17 Liberec, Czech Republic (e-mail: Zbynek.Koldovsky@tul.cz; website: <http://itakura.kes.vslib.cz/zbynek/index.htm>).

E. Oja is with the Neural Networks Research Centre, Helsinki University of Technology, 02015 HUT, Finland (e-mail: Erkki.Oja@hut.fi; website: <http://www.cis.hut.fi/~oja/>).

Digital Object Identifier 10.1109/TSP.2006.870561

as possible. For the FastICA algorithm, such an asymptotic performance analysis for a general cost function was proposed in [17].

The Cramér–Rao lower bound (CRB) provides an algorithm independent bound for parameter estimation. In the context of ICA, a Cramér–Rao-like bound for intersignal interference is derived as asymptotic variance of a maximum-likelihood estimate in [24], [26]–[29], and [32]. A similar result is known for a related problem of blind deconvolution [30].

The purpose of the present paper is to look at the performance of the FastICA algorithm, both the one-unit and symmetric versions, in this latter sense of asymptotic error, and compare it with the exact CRB computed from its definition. The paper is organized as follows. In Section II, the linear ICA model and the FastICA algorithm are described. In addition, a novel check of saddle points of the FastICA cost function is proposed that allows to find the global minimum of the cost function in almost 100% simulation runs. Finally, the following criteria to characterize the performance of the algorithm are introduced: a gain matrix (variance of its elements) and a signal-to-interference ratio (SIR). In Section III, analytic expressions for the variance of the off-diagonal gain matrix elements are derived and discussed. These expressions are asymptotically valid for large data sets when a “good” initialization of the algorithm is assumed. Most of the details of the analysis are deferred to Appendixes. As an example of utilization of the analysis, a novel variant of FastICA is proposed, which combines the one-unit algorithm and the symmetric algorithm adaptively, depending on empirical distribution of the estimated signal components, to improve the performance.

In Section IV, the CRB on the variance of the off-diagonal gain matrix elements is computed via inverse of a Fisher information matrix. Section V compares the CRB with the asymptotic performance of FastICA and explains nonexistence of the CRB for signals with bounded magnitude (e.g., uniform distribution) and for some long-tailed distributions.

Section VI presents a number of computer simulations using artificial data that validate and support the theoretical analysis. The simulations also compare the algorithmic performance with the CRB derived in Section IV. Finally, Section VII summarizes the results and presents the conclusions.

II. DATA MODEL AND THE METHOD

Let \mathbf{S} represent a $d \times N$ data matrix, composed of d rows, where each row $\mathbf{s}_k^T, k = 1, \dots, d$ contains N independent realizations of a random variable s_k . Next assume that s_k has a distribution function $F_k(t) = P(s_k \leq t)$. In a typical case for ICA, the rows \mathbf{s}_k^T are called the source signals, and the d random variables s_k are mutually independent.

The standard linear ICA model of a given $d \times N$ data matrix is

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (1)$$

where \mathbf{A} is an unknown, nonsingular $d \times d$ mixing matrix. Thus, each row of \mathbf{X} is a linear mixture of the unknown independent signals \mathbf{s}_k^T . The goal of independent component analysis

is to estimate the matrix \mathbf{A} or, equivalently, the demixing matrix $\mathbf{W} = \mathbf{A}^{-1}$ or, equivalently, the original source signals \mathbf{S} . The following are well known:

- 1) the separation is unique only up to an unknown scaling and ordering of the components \mathbf{s}_k^T ;
- 2) the separation is possible only if at most one of the original source variables s_k has a Gaussian distribution.

Since the scale of the source signals cannot be retrieved, one can assume, without any loss in generality, that the sample variance of the estimated source signals is equal to one. Thus, instead of the original source signals \mathbf{S} , a normalized source signal matrix denoted \mathbf{U} can be estimated, where

$$\mathbf{U} = \mathbf{D}^{-1/2}(\mathbf{S} - \bar{\mathbf{S}}) \quad (2)$$

$$\mathbf{D} = \text{diag}[\hat{\sigma}_1^2, \dots, \hat{\sigma}_d^2] \quad (3)$$

$$\hat{\sigma}_k^2 = (\mathbf{s}_k - \bar{\mathbf{s}}_k)^T (\mathbf{s}_k - \bar{\mathbf{s}}_k) / N \quad (4)$$

$$\bar{\mathbf{s}}_k = (\mathbf{s}_k^T \cdot \mathbf{1}_N) \mathbf{1}_N / N, \quad k = 1, \dots, d \quad (5)$$

where $\mathbf{1}_N$ stands for $N \times 1$ vector of 1's.

A. Preprocessing

The first step of many variants of the ICA algorithms consists of removing the sample mean and a whitening (decorrelation and scaling), i.e., the transformation

$$\mathbf{Z} = \hat{\mathbf{C}}^{-1/2}(\mathbf{X} - \bar{\mathbf{X}}) \quad (6)$$

where

$$\hat{\mathbf{C}} = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T / N \quad (7)$$

is the sample covariance matrix, and $\bar{\mathbf{X}}$ is the sample mean, $\bar{\mathbf{X}} = \mathbf{X} \cdot \mathbf{1}_N \mathbf{1}_N^T / N$. The output \mathbf{Z} contains decorrelated and unit variance data in the sense that $\mathbf{Z}\mathbf{Z}^T / N = \mathbf{I}$ (identity matrix). Note that \mathbf{Z} can be rewritten using (1) and (2) as

$$\mathbf{Z} = \hat{\mathbf{C}}^{-1/2} \mathbf{A} \mathbf{D}^{1/2} \mathbf{U}. \quad (8)$$

The ICA problem can be formulated as the one to find a demixing matrix $\hat{\mathbf{W}}(\mathbf{Z})$ that separates the original signals from the mixture \mathbf{Z} , i.e., $\hat{\mathbf{U}} = \hat{\mathbf{W}}(\mathbf{Z}) \cdot \mathbf{Z}$.

B. FastICA Algorithm for One Unit

The fixed-point algorithm for one-unit estimates one row of the demixing matrix $\hat{\mathbf{W}}(\mathbf{Z})$ as a vector $\hat{\mathbf{w}}_z^T$ that is a stationary point (minimum or maximum) of the expression $E[G(\mathbf{w}^T \mathbf{Z})] \stackrel{\text{def}}{=} G(\mathbf{w}^T \mathbf{Z}) \mathbf{1}_N / N$ subject to $\|\mathbf{w}\| = 1$, where $G(\cdot)$ is a suitable nonlinear and nonquadratic function [3]. In the above expression, $G(\cdot)$ is applied elementwise.

Finding $\hat{\mathbf{w}}_z$ proceeds iteratively. Starting with a random initial unit norm vector \mathbf{w} , iterate

$$\mathbf{w}^+ \leftarrow \mathbf{Z}g(\mathbf{Z}^T \mathbf{w}) - \mathbf{w}g'(\mathbf{w}^T \mathbf{Z}) \mathbf{1}_N \quad (9)$$

$$\mathbf{w} \leftarrow \mathbf{w}^+ / \|\mathbf{w}^+\| \quad (10)$$

until convergence is achieved. In (9) and also elsewhere in the paper, in accord with the standard notation [3], $g(\cdot)$ and $g'(\cdot)$ denote the first and the second derivative of the function $G(\cdot)$.

The application of $g(\cdot)$ and $g'(\cdot)$ to the vector $\mathbf{w}^T \mathbf{Z}$ is elementwise. Classical widely used functions $g(\cdot)$ include “pow3,” i.e., $g(x) = x^3$ (then the algorithm performs kurtosis minimization), “tanh,” i.e., $g(x) = \tanh(x)$, and “Gauss,” $g(x) = x \exp(-x^2/2)$.

It is not known in advance which column of $\hat{\mathbf{W}}^T(\mathbf{Z})$ is being estimated: It largely depends on the initialization. Note that the recursion for some components might not converge. In the deflation method [9], which is not studied in this paper, this problem is solved by separating the components from the mixture one by one using orthogonal projections. Here, we shall assume that each signal component can be separated from the *original* signal mixture using suitable initializations. Assume that the separating vectors \mathbf{w} computed for all components are appropriately sorted [20] and summarized as rows in a matrix denoted $\hat{\mathbf{W}}^{1U}(\mathbf{Z})$. The rows in $\hat{\mathbf{W}}^{1U}(\mathbf{Z})$ may not be mutually orthogonal, in general.

C. Symmetric Fastica Algorithm

The symmetric FastICA proceeds similarly, the estimation of all independent components (or equivalently, of all rows of \mathbf{W}) proceeds in parallel, and each step is completed by a symmetric orthonormalization. Starting with a random unitary matrix \mathbf{W} , iterate

$$\mathbf{W}^+ \leftarrow g(\mathbf{W}\mathbf{Z})\mathbf{Z}^T - \text{diag}[g'(\mathbf{W}\mathbf{Z})\mathbf{1}_N] \mathbf{W} \quad (11)$$

$$\mathbf{W} \leftarrow (\mathbf{W}^+ + \mathbf{W}^{+T})^{-1/2} \mathbf{W}^+ \quad (12)$$

until convergence is achieved. The stopping criterion proposed in [14] is

$$1 - \min(\text{abs}(\text{diag}(\mathbf{W}^T \mathbf{W}_{\text{old}}))) < \epsilon \quad (13)$$

for a suitable constant ϵ .

The result of the symmetric FastICA (unlike in the one-unit algorithm without deflation) is a unitary matrix denoted $\hat{\mathbf{W}}^{\text{SYM}}(\mathbf{Z})$. As a consequence, sample correlations between the separated signals are exactly equal to zero.

D. Check of Saddle Points

In general, the global convergence of the symmetric FastICA is known to be quite good. Nevertheless, if it is run 10 000 times from random initial demixing matrices, on the average in 1–100 cases, the algorithm gets stuck at solutions that can be recognized by exceptionally low achieved SIR. The rate of these false solutions depends on the dimension of the model, on the stopping rule, and on the length of the data (see the example at the end of this subsection).

A detailed investigation of the false solutions showed that they contain one or more pairs of estimated components, say $(\hat{\mathbf{u}}_k, \hat{\mathbf{u}}_\ell)$, such that they are close to $(\mathbf{u}_k + \mathbf{u}_\ell)/\sqrt{2}$ and $(\mathbf{u}_k - \mathbf{u}_\ell)/\sqrt{2}$, respectively, where $(\mathbf{u}_k, \mathbf{u}_\ell)$ is the desired solution (see Fig. 1). Due to symmetry, the saddle points of the criterion function lie approximately halfway between two correct solutions that differ in the order of two of their components. Thus, an appropriate estimate of $(\mathbf{u}_k, \mathbf{u}_\ell)$ would be $(\hat{\mathbf{u}}'_k, \hat{\mathbf{u}}'_\ell)$, where

$$\hat{\mathbf{u}}'_k = (\hat{\mathbf{u}}_k + \hat{\mathbf{u}}_\ell)/\sqrt{2} \quad \text{and} \quad \hat{\mathbf{u}}'_\ell = (\hat{\mathbf{u}}_k - \hat{\mathbf{u}}_\ell)/\sqrt{2}.$$

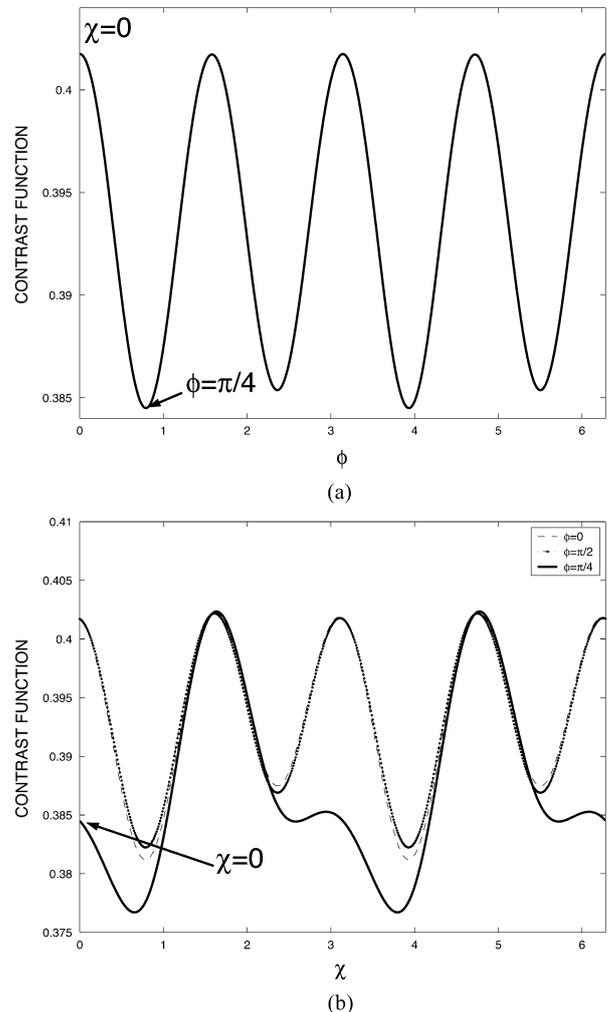


Fig. 1. Contrast function $E\{G(\cos \chi (\cos \varphi \cdot \mathbf{x}_1 + \sin \varphi \cdot \mathbf{x}_2) + \sin \chi \cdot \mathbf{x}_3)\}$ (a) as a function of φ for $\chi = 0$, and (b) as a function of χ for $\varphi = 0, \pi/4$, and $\pi/2$, respectively; $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ were generated as i.i.d. uniformly distributed in $[-\sqrt{3}, \sqrt{3}]$ with the length $N = 10\,000$, and $G(x) = \log \cosh(x)$. The point $[\varphi, \chi] = [\pi/4, 0]$ is a saddle point of the contrast function—it is its local minimum with regard to φ and a local maximum with regard to χ .

A selection between given candidates $(\hat{\mathbf{u}}_k, \hat{\mathbf{u}}_\ell), (\hat{\mathbf{u}}'_k, \hat{\mathbf{u}}'_\ell)$ for a better estimate of $(\mathbf{u}_k, \mathbf{u}_\ell)$ can be done by maximizing the criterion used in the very beginning of derivation of FastICA

$$c(\hat{\mathbf{u}}_k, \hat{\mathbf{u}}_\ell) = [G(\hat{\mathbf{u}}_k^T \mathbf{1}_N/N - G_0)]^2 + [G(\hat{\mathbf{u}}_\ell^T \mathbf{1}_N/N - G_0)]^2$$

where $G_0 = E[G(\xi)]$ and ξ is a standard normal random variable. In the case of the nonlinearity “tanh,” $G(x) = \log \cosh(x)$ and $G_0 \approx 0.3746$.

Thus, we suggest to complete the plain symmetric FastICA by the check of all $\binom{d}{2}$ pairs of the estimated independent components for a possible improvement via the saddle points. If the test for saddle point is positive, it is suggested to perform one or two additional iterations of the original algorithm, starting from the improved estimate.

TABLE I
NUMBER OF FAILURES OF SYMMETRIC FASTICA (tanh) AMONG 10 000 TRIALS

	N=200	N=500	N=1000	N=10000
$d = 2 \& \epsilon = 10^{-4}$	85	57	59	46
$d = 2 \& \epsilon = 10^{-5}$	49	16	15	12
$d = 2 \& \text{ stop } 3 \times$	41	4	1	2
$d = 2 \& \text{ s.p.check}$	0	0	0	0
$d = 3 \& \epsilon = 10^{-4}$	49	5	4	6
$d = 3 \& \epsilon = 10^{-5}$	43	0	1	0
$d = 3 \& \text{ stop } 3 \times$	45	0	0	0
$d = 3 \& \text{ s.p.check}$	0	0	0	0
$d = 4 \& \epsilon = 10^{-4}$	95	9	4	11
$d = 4 \& \epsilon = 10^{-5}$	85	2	0	5
$d = 4 \& \text{ stop } 3 \times$	90	1	0	1
$d = 4 \& \text{ s.p.check}$	5	0	0	0
$d = 5 \& \epsilon = 10^{-4}$	166	2	4	11
$d = 5 \& \epsilon = 10^{-5}$	151	1	2	2
$d = 5 \& \text{ stop } 3 \times$	157	1	2	0
$d = 5 \& \text{ s.p.check}$	17	0	0	0

The failure rates of the plain symmetric FastICA with three different stopping rules and of the improved FastICA with the check of the saddle points are compared in the following example. The first stopping rule was (13) with $\epsilon = 10^{-4}$, the second stopping rule was the same with $\epsilon = 10^{-5}$, and the third stopping rule required the former condition to be fulfilled in three consecutive steps. The improved algorithm used the first stopping rule and the test of the saddle points.

These four variants of the algorithm were applied to separate $d = 2, 3, 4,$ and 5 independent signals with uniform distribution and varying length in 10 000 independent trials with a randomly selected initial demixing matrix. The number of algorithmic failures that are detected by the condition that SIR of some of the separated components is smaller than 3 dB is displayed in Table I. The table shows zero rate of the improved algorithm except for the case of the data with the shortest length, $N = 200$. In the latest case, the rate of failures has significantly dropped compared to the former three variants.

E. Measure of the Separation Quality

The separation ability of ICA algorithms can be characterized by the relative presence of the k th source signal in the estimated i th source signal. It is possible, if the source signals are known. Due to the permutation and sign/phase uncertainty, the estimated sources need to be appropriately sorted to fit the original ones. In this paper, the method proposed in [20] is used. Formally, the estimated source signals can be written using (8) as

$$\begin{aligned} \hat{\mathbf{U}} &= \hat{\mathbf{W}}(\mathbf{Z}) \cdot \mathbf{Z} = \hat{\mathbf{W}}(\mathbf{Z}) \hat{\mathbf{C}}^{-1/2} \mathbf{A} \mathbf{D}^{1/2} \mathbf{U} \\ &= \mathbf{G} \mathbf{U} \end{aligned} \quad (14)$$

where $\mathbf{G} = \hat{\mathbf{W}}(\mathbf{Z}) \hat{\mathbf{C}}^{-1/2} \mathbf{A} \mathbf{D}^{1/2}$ and $\hat{\mathbf{W}}(\mathbf{Z})$ stands either for $\hat{\mathbf{W}}^{1U}(\mathbf{Z})$ or for $\hat{\mathbf{W}}^{\text{SYM}}(\mathbf{Z})$. Note that \mathbf{G} has the meaning of the estimated demixing matrix provided that $\mathbf{A} = \mathbf{D} = \mathbf{I}$. It will be called the *gain* matrix for brevity.

The relative presence of the k th source signal in the estimated i th source signal is represented by the (i, k) th element of \mathbf{G} , denoted \mathbf{G}_{ik} . Then, the total SIR of the k th source signal is defined as follows:

$$\text{SIR}_k = \frac{\mathbb{E}[\mathbf{G}_{kk}^2]}{\mathbb{E} \left[\sum_{\substack{k=1 \\ k \neq i}}^d \mathbf{G}_{ik}^2 \right]}. \quad (15)$$

It is important to note that the estimator $\hat{\mathbf{U}}$ is invariant with respect to orthogonal transformations of the decorrelated data \mathbf{Z} , or equivariant [10]. It is because the recursions (9) and (10) or (11) and (12) that represent the algorithm are equivalent to the same relations with \mathbf{Z} , \mathbf{W}^+ , and \mathbf{W} replaced by $\mathbf{Q}\mathbf{Z}$, $\mathbf{W}^+ \mathbf{Q}^{-1}$, and $\mathbf{W}\mathbf{Q}^{-1}$, respectively, where \mathbf{Q} is an arbitrary unitary (i.e., obeying $\mathbf{Q}^T = \mathbf{Q}^{-1}$) matrix. Then, the product

$$\hat{\mathbf{U}} = \mathbf{W} \cdot \mathbf{Z} = \mathbf{W}\mathbf{Q}^{-1} \cdot \mathbf{Q}\mathbf{Z}$$

remains independent of \mathbf{Q} . From these facts, it follows that the gain matrix \mathbf{G} and consequently the SIR are *independent* of the mixing matrix \mathbf{A} .¹

III. ANALYSIS

Due to the above-mentioned equivariant property of FastICA it can be assumed, without any loss in generality, that the recursions (9) and (10) or (11) and (12) begin with the decorrelated data of the form

$$\mathbf{Z} = \mathbf{R}^{-1/2} \mathbf{U} \quad (16)$$

where

$$\mathbf{R} = \frac{1}{N} \mathbf{U} \mathbf{U}^T. \quad (17)$$

The gain matrix of interest is now

$$\mathbf{G} = \hat{\mathbf{W}}(\mathbf{Z}) \cdot \mathbf{R}^{-1/2}. \quad (18)$$

Note that the gain matrix \mathbf{G} (and consequently the SIR as well) is a function of the normalized source signals \mathbf{U} and of the non-linear function $g(\cdot)$ used in the algorithm only.

The main result of this section can be summarized as follows.

Proposition 1: Assume that 1) all original independent components have zero mean and unit variance and are temporarily white, 2) the function g in algorithm FastICA is twice continuously differentiable, 3) the following expectations exist:

$$\mathbb{E}[s_k g(s_k)] \stackrel{\text{def}}{=} \mu_k \quad (19)$$

$$\mathbb{E}[g'(s_k)] \stackrel{\text{def}}{=} \rho_k \quad (20)$$

$$\mathbb{E}[g^2(s_k)] \stackrel{\text{def}}{=} \beta_k \quad (21)$$

¹To be exact, a change of the mixing matrix (or a change in the algorithm initialization) may cause a change of the order or sign of the components at the algorithm output. Here, however, we assume that the order and signs of the components are post-processed to fit the original signals [20].

TABLE II
SIR (IN DECIBELS) OF FastICA IN ITS MAIN SIX VARIANTS FOR TWO COMPONENTS WITH THE SAME DISTRIBUTION, AND THE CRAMÉR–RAO BOUND (DERIVED IN SECTION IV) FOR $N = 1000$. THE BEST SIR IS MARKED BY BOLD CHARACTERS

PDF	SYMMETRIC			ONE UNIT			CRB
	TANH	GAUSS	POW3	TANH	GAUSS	POW3	
uniform	32.3	32.2	33.3	31.6	31.5	33.7	∞
sinus	34.7	34.7	35.1	37.5	37.6	39.5	∞
bpsk	36.0	36.0	36.0	∞	∞	∞	∞
GG(4)	27.6	27.5	28.0	25.2	25.1	25.7	28.1
GG(3)	23.9	23.7	23.7	21.1	20.1	20.1	24.0
Laplace	29.0	29.4	24.9	27.0	27.4	22.2	31.8
GG(0.5)	33.3	33.9	24.9	33.6	35.0	22.2	∞
GG(0.1)	35.9	35.9	-3.1	47.8	50.7	-6.11	∞

for $k = 1, \dots, d$, and 4) the FastICA algorithm (in both variants) is started from the correct demixing matrix and stops after a single iteration.

Then, the normalized gain matrix elements $N^{1/2}\mathbf{G}_{k\ell}^{1U}$ and $N^{1/2}\mathbf{G}_{k\ell}^{\text{SYM}}$ for the one-unit FastICA and for symmetric FastICA, respectively, have asymptotically Gaussian distribution $\mathcal{N}(0, V_{k\ell}^{1U})$ and $\mathcal{N}(0, V_{k\ell}^{\text{SYM}})$, where

$$V_{k\ell}^{1U} = \frac{\beta_k - \mu_k^2}{(\mu_k - \rho_k)^2} \quad (22)$$

$$V_{k\ell}^{\text{SYM}} = \frac{\beta_k - \mu_k^2 + \beta_\ell - \mu_\ell^2 + (\mu_\ell - \rho_\ell)^2}{(|\mu_k - \rho_k| + |\mu_\ell - \rho_\ell|)^2} \quad (23)$$

for $k, \ell = 1, \dots, d, k \neq \ell$, provided that the denominators are nonzero.

Proof: See Appendix A. An expression similar to (22) can be found in [10] and [17], but (23) is novel.

The assumption 4 may look peculiar at the first glance, but it is not so restrictive as it seems to be. It reflects the fact that the presented analysis is “local” and assumes a “good” initialization of the algorithm. The algorithm itself may have good global convergence properties (see Section VI), but it is not a subject of this proposition. Once the algorithm is started from an initial \mathbf{W} that lies in a right domain of attraction, the resultant stationary point of the recursion, denoted $\hat{\mathbf{W}}$, is the same and is approximately equal to \mathbf{W}^+ obtained after one step from the ideal solution, due to the fact that the convergence is quadratic.²

Our numerical simulations presented in Section VII, and also other simulations that were skipped for lack of space, confirm the validity of the asymptotic variances (22) and (23) for the algorithm variant introduced in Section VI *working with arbitrary (random) initialization*. Namely, it is shown that $\text{var}[\mathbf{G}_{k\ell}^{1U}] \approx (1/N)V_{k\ell}^{1U}$ and $\text{var}[\mathbf{G}_{k\ell}^{\text{SYM}}] \approx (1/N)V_{k\ell}^{\text{SYM}}$. The expressions in (22) and (23) are functions of the probability distribution of s_k and of the nonlinear function $g(\cdot)$ via the expectations in (19)–(21). Given the distribution and the nonlinearity, these expressions can be evaluated.

²The quadratic convergence means that if the initial difference between the initial \mathbf{W} and $\hat{\mathbf{W}}$ is $\Delta\mathbf{W}$, the distance of \mathbf{W}^+ (that is \mathbf{W} after one iteration) is $O(\|\Delta\mathbf{W}\|^2)$.

Table II shows the theoretical SIR of the main six variants of FastICA for separation of two components with the same distribution, computed for a few distributions considered frequently in the literature, for sample size $N = 1000$. Here, the distribution “sinus” means the distribution of $\sqrt{2}\sin(u)$, where u is uniformly distributed in $(0, 2\pi)$, “bpsk” is the discrete distribution with values ± 1 , both with the probability 0.5, and $\text{GG}(\alpha)$ means the generalized Gaussian distribution with parameter α , described in Appendix F. Note that the latter distribution is standard Gaussian for $\alpha = 2$, the Laplace distribution for $\alpha = 1$, sub-Gaussian for $\alpha > 2$, approaching the uniform distribution for $\alpha \rightarrow \infty$, and super-Gaussian (spiky) for $\alpha \rightarrow 0^+$.

Note that for separation of $d > 2$ components, the SIR would be $(d-1) \times 3$ dB lower than in the table, and if N is increased/decreased ten times, the resultant theoretical SIR is increased/decreased by 10 dB compared with the table.

A. Example of Utilization

In this subsection, the previous analysis is used to derive a novel variant of the FastICA algorithm, which combines advantages of both previously discussed variants. For easy reference, it will be called “Smart FastICA.” This algorithm begins with applying symmetric FastICA with nonlinearity “tanh.” For each estimated component signal $\hat{\mathbf{u}}_k^T$, parameters μ_k, ρ_k , and β_k are computed as sample estimates of the expectations in (19)–(21), namely $\hat{\mu}_k = \hat{\mathbf{u}}_k^T g(\hat{\mathbf{u}}_k)/N$, $\hat{\rho}_k = \hat{\mathbf{1}}_N^T g'(\hat{\mathbf{u}}_k)/N$, $\hat{\beta}_k = \hat{\mathbf{1}}_N^T g^2(\hat{\mathbf{u}}_k)/N$, and then they are plugged in (22) and (23) and (15), namely

$$\hat{\text{SIR}}_k^{(1U)} = \frac{N}{\sum_{\ell=1, \ell \neq k}^d \hat{V}_{k\ell}^{(1U)}}$$

$$\hat{\text{SIR}}_k^{(\text{SYM})} = \frac{N}{\sum_{\ell=1, \ell \neq k}^d \hat{V}_{k\ell}^{(\text{SYM})}}.$$

If the obtained SIR_k for the one-unit algorithm is better than for the former estimate, the algorithm is performed, taking advantage of a more suitable nonlinearity g for each of particular cases: In the super-Gaussian case, defined by the condition $\hat{\mu}_k < \hat{\rho}_k$, the option “Gauss” is selected, and in the sub-Gaussian case

with $\hat{\mu}_k > \hat{\rho}_k$, “pow3” is applied (see the simulation section for a reason).

Then, μ_k, ρ_k, β_k , and SIR_k are computed again. If the new SIR_k is better than the previous one and if, at the same time, the scalar product between the former separating vector and the new one is higher in absolute value than a constant (we have used 0.75), then the one-unit refinement is accepted in favor of the former vector. The condition on the scalar product is intended to eliminate the cases where the one-unit algorithm converged to a wrong component. A further optimization of the algorithm exceeds the scope of the paper [33].

B. Optimum Nonlinearity G

It is interesting to know, which function $g(\cdot)$ would be optimal for given probability density function (pdf) of s_k . If all source signals have the same distribution, the answer is well known. It is the so-called score function of the distribution, defined as $\psi(x) = -f'(x)/f(x)$, where $f(x)$ is the underlying pdf. Introduce the notation

$$\kappa = E[\psi^2(\xi)] = \int_R \frac{f'^2(x)}{f(x)} dx \quad (24)$$

where ξ is a random variable with the pdf $f(\cdot)$. Note that if ξ has zero mean and variance one, it holds $\kappa \geq 1$, where the equality is attained if and only if the underlying distribution is standard Gaussian (see Appendix E). Thus, κ represents a measure of non-Gaussianity.

For the optimum nonlinearity $g_{\text{opt}}(x) = \psi(x)$, a straightforward computation gives $\mu_k = 1$ and $\rho_k = \beta_k = \kappa$, and consequently

$$\min_g V_{k\ell}^{1U} = \frac{1}{\kappa - 1} \quad (25)$$

$$\min_g V_{k\ell}^{\text{SYM}} = \frac{1}{4} + \frac{1}{2} \frac{1}{\kappa - 1}. \quad (26)$$

IV. CRAMÉR–RAO LOWER BOUND FOR ICA

Consider a vector of parameters θ being estimated from a data vector \mathbf{x} , having probability density $f_{\mathbf{x}|\theta}(\mathbf{x}|\theta)$, using some unbiased estimator $\hat{\theta}$. The CRB is the lower bound for the variance of $\hat{\theta}$. Assume that $f_{\mathbf{x}|\theta}$ is smooth and the following Fisher information matrix exists:

$$\mathbf{F}_\theta = E_\theta \left[\frac{1}{f_{\mathbf{x}|\theta}^2} \frac{\partial f_{\mathbf{x}|\theta}(\mathbf{x}|\theta)}{\partial \theta} \left(\frac{\partial f_{\mathbf{x}|\theta}(\mathbf{x}|\theta)}{\partial \theta} \right)^T \right]. \quad (27)$$

Then, under some mild regularity condition,[18],³ it holds

$$\text{cov} \hat{\theta} \geq \text{CRB}_\theta = \mathbf{F}_\theta^{-1}.$$

Next, if $\varphi = \varphi(\theta)$ is a differentiable function of θ , then the Fisher information matrix for φ exists as well and is equal to

$$\mathbf{F}_\varphi = \mathbf{J}_\theta^{-1} \mathbf{F}_\theta \mathbf{J}_\theta^{-T} \quad (28)$$

³1) Support of $f_{\mathbf{x}|\theta}$ is independent of θ ; 2) $\partial f_{\mathbf{x}|\theta}(\mathbf{x}|\theta)/\partial \theta$ exists for all θ from an open set; and 3) $E[\partial f_{\mathbf{x}|\theta}(\mathbf{x}|\theta)/\partial \theta] = 0$

where \mathbf{J} is the Jacobian of the mapping $\varphi(\theta)$. If the mapping is linear, or $\varphi(\theta) = \mathbf{M}\theta$ for some regular matrix \mathbf{M} , then $\mathbf{J}_\theta = \mathbf{M}^T$.

In the context of ICA, we first focus on deriving the CRB for estimation of the demixing matrix $\mathbf{W} = \mathbf{A}^{-1}$, i.e., the parameter vector is $\theta = \text{vec}[\mathbf{W}]$.

The following assumptions will be considered throughout this section:

$$E[s_i^2] = \int_R t^2 f_i(t) dt = 1 \quad (29)$$

$$\kappa_i \stackrel{\text{def}}{=} E[\psi_i^2(s_i)] = \int_R \psi_i^2(t) f_i(t) dt < +\infty \quad (30)$$

$$\eta_i \stackrel{\text{def}}{=} E[s_i^2 \psi_i(s_i)^2] = \int_R t^2 \psi_i^2(t) f_i(t) dt < +\infty \quad (31)$$

where $i = 1, \dots, d$ and ψ_i denotes the score function of the corresponding pdf, i.e., $\psi_i(x) = -(f'_i(x)/f_i(x)) \cdot \psi_i$ is assumed to have zero mean for all i , and $f_i(x) > 0$ for all i and x .

A. Fisher Information Matrix

From the independence of the original signals, it follows that their joint pdf is $f_{\mathbf{S}}(\mathbf{S}) = \prod_{i=1}^d \prod_{j=1}^N f_i(s_{ij})$. Then, using the transformation $\mathbf{X} = \mathbf{W}^{-1}\mathbf{S}$

$$f_{\mathbf{X}}(\mathbf{X}) = |\det \mathbf{W}| f_{\mathbf{S}}(\mathbf{W}\mathbf{X}). \quad (32)$$

Incorporating this density into (27), the m th element of the $d^2 \times d^2$ Fisher information matrix \mathbf{F}_θ , where $m = (i-1)d + j$, $n = (u-1)d + v$, and w_{ij} denotes the ij th element of the matrix \mathbf{W} , is

$$\mathbf{F}_{mn} = E \left[\frac{|\det \mathbf{W}|^{-2}}{f_{\mathbf{S}}^2(\mathbf{S})} \frac{\partial f_{\mathbf{X}}}{\partial w_{ij}} \frac{\partial f_{\mathbf{X}}}{\partial w_{uv}} \right]. \quad (33)$$

A straightforward computation (see Appendix C) gives

$$\mathbf{F}_{mn} = (N-1)^2 a_{ji} a_{vu} + N a_{ju} a_{vi} + \delta_{iu} N a_{ji} a_{vi} (\eta_i - 2) + \delta_{iu} N \kappa_i \sum_{\ell=1, \ell \neq u}^d a_{j\ell} a_{v\ell} \quad (34)$$

with κ_i, η_i defined in (30) and (31), δ_{iu} is the Kronecker's delta, and a_{ij} denotes the ij th element of the mixing matrix \mathbf{A} . It can be shown, using (28), that

$$\mathbf{F}_\theta = (\mathbf{A} \otimes \mathbf{I}) \mathbf{F}_\mathbf{I} (\mathbf{A}^T \otimes \mathbf{I}) \quad (35)$$

where $\mathbf{F}_\mathbf{I}$ stands for the Fisher information matrix derived for a case when $\mathbf{A} = \mathbf{I}$ (identity matrix); \otimes denotes the Kronecker product. Substituting $a_{ij} = \delta_{ij}$ into (34), it easily follows that

$$(\mathbf{F}_\mathbf{I})_{mn} = (N-1)^2 \delta_{ji} \delta_{vu} + N \delta_{ju} \delta_{vi} + N (\delta_{ji} \delta_{vu} \delta_{vi} (\eta_i - \kappa_i - 2) + \delta_{iu} \delta_{vj} \kappa_i). \quad (36)$$

Some properties of the matrix will be shown in Appendix D.

B. Accuracy of the Estimation of $\mathbf{G}_2 = \hat{\mathbf{W}}\mathbf{A}$

Let $\hat{\mathbf{W}}$ denote an estimator of the demixing matrix \mathbf{W} . Estimated signals $\hat{\mathbf{S}}$ are then $\hat{\mathbf{S}} = \hat{\mathbf{W}}\mathbf{X} = \hat{\mathbf{W}}\mathbf{A}\mathbf{S}$. It is interesting to compute the CRB for the elements of the gain matrix $\mathbf{G}_2 = \hat{\mathbf{W}}\mathbf{A}$, which is closely related to the gain matrix \mathbf{G}

defined in (14). A comparison of the definition relations gives $\mathbf{G} = \mathbf{G}_2 \mathbf{D}^{1/2}$, where \mathbf{D} contains, on its diagonal, sample variances of the original independent signal components. Asymptotically, \mathbf{D} converges to unity matrix, and hence any estimate of \mathbf{G} is at the same time an estimate of \mathbf{G}_2 , and vice versa. In addition, it follows from the analysis in Appendix A that the asymptotic distribution of nondiagonal elements of \mathbf{G} and those of \mathbf{G}_2 is the same.

To compute the CRB for \mathbf{G}_2 , note that the new parameter vector $\boldsymbol{\theta}_{\mathbf{G}} = \text{vec}[\mathbf{G}_2]$ is just a linear function of the parameter $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta}_{\mathbf{G}} = \text{vec}[\hat{\mathbf{W}}\mathbf{A}] = (\mathbf{A}^T \otimes \mathbf{I})\text{vec}[\hat{\mathbf{W}}] = (\mathbf{A}^T \otimes \mathbf{I})\boldsymbol{\theta}$. Then, using (28), the Fisher information matrix of $\boldsymbol{\theta}_{\mathbf{G}}$ is

$$\mathbf{F}_{\mathbf{G}} = (\mathbf{W} \otimes \mathbf{I})\mathbf{F}_{\boldsymbol{\theta}}(\mathbf{W}^T \otimes \mathbf{I}) = \mathbf{F}_{\mathbf{I}}. \quad (37)$$

Note that $\mathbf{F}_{\mathbf{G}}$ is independent of the mixing matrix \mathbf{A} . The CRB for the ij th element of \mathbf{G} is

$$\text{var}((\mathbf{G}_2)_{ij}) \geq \text{CRB}((\mathbf{G}_2)_{ij}) = (\mathbf{F}_{\mathbf{I}}^{-1})_{mm}$$

where $m = (i-1)d + j$ and $i \neq j$. In Appendix D, it is proved that for such m

$$(\mathbf{F}_{\mathbf{I}}^{-1})_{mm} = \frac{1}{N} \frac{\kappa_j}{\kappa_i \kappa_j - 1} \quad (38)$$

which gives us the desired lower bound

$$\text{CRB}((\mathbf{G}_2)_{ij}) = \frac{1}{N} \frac{\kappa_j}{\kappa_i \kappa_j - 1}. \quad (39)$$

The diagonal elements of \mathbf{G}_2 are not as important, they just reflect the accuracy of estimating the power of the components, or equivalently, the norm of rows of the demixing matrix.

V. DISCUSSION

A. Comparison of CRB With Performance of FastICA With Optimum G

The Cramér–Rao lower bound in (39) is compared with the asymptotic variance of FastICA in (25) and (26) in Fig. 2. We can see that for κ close to 1, the CRB is close to the variance of the symmetric FastICA with the optimum nonlinearity. In this case, however, the estimation may fail, because the variance of the estimator itself goes to infinity, and convergence of the algorithm may be slow.

In the opposite case, for $\kappa \gg 1$, the CRB asymptotically coincides with the variance of the one-unit FastICA with the optimum nonlinearity, because

$$\frac{\text{var}[\mathbf{G}_{kl}^{1U-\text{opt}}]}{\text{CRB}[\mathbf{G}_{kl}]} \approx \frac{N-1 V_{kl}^{1U-\text{opt}}}{\text{CRB}[\mathbf{G}_{kl}]} \rightarrow 1 \quad \text{for } \kappa \rightarrow \infty.$$

We conclude that the FastICA algorithm with the optimum nonlinearity is asymptotically efficient in two cases: 1) one-unit version for $\kappa_i \gg 1$ and 2) symmetric version for $\kappa_i \rightarrow 1^+$ provided that all components have the same distribution law.

B. Separation of Sources With the Generalized Gaussian Distribution

Properties of the generalized Gaussian distribution are listed for easy reference in Appendix F. Note that the score func-

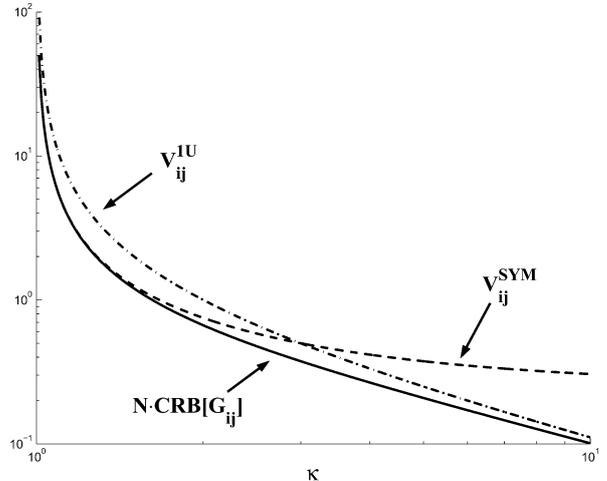


Fig. 2. Asymptotic performance of one-unit and symmetric FastICA and the CRB versus parameter κ .

tion of this distribution is proportional to $|x|^{\alpha-1}\text{sign}(x)$ so that $g(x) = |x|^{\alpha-1}\text{sign}(x)$ is the theoretically optimum nonlinearity for the distribution. However, only for $\alpha > 1$ is this function continuous and hence suitable nonlinearity for FastICA. For discontinuous g 's, the algorithm appears not to converge.

C. Distributions With Finite Support

The CRB does not exist (the bound is infinite) for the bounded magnitude distributions such as “uniform,” “sinus,” and “bpsk” in Table II. It happens because these distributions do not have infinite support, as required for existence of the CRB. Since the uniform distribution is a limit of the GGD(α) for α going to infinity, it is natural to study FastICA with nonlinearity $g_k(x) = |x|^k \text{sign}(x)$ with large k . It can be easily shown that the one-unit FastICA with this nonlinearity has asymptotic variance $V_{ij}^{1U}(k) \approx 3/(2k+1)$ that goes to zero for $k \rightarrow \infty$. Similar results can be obtained for the distribution “sinus.” In other words, the asymptotic variance of FastICA cannot be lower bounded by any bound of the form B/N . Implications of the above observation for an adaptive choice of the nonlinearity exceed the scope of this paper.

D. Distributions With Long Tails

The CRB does not exist for the GGD(α) distribution with parameter $\alpha \leq 1/2$ (cf. lines 7 and 8 in Table II). These distributions are sometimes called “long tailed”. Instead of the score function, let us consider the nonlinearity $g_{\alpha,k}(x) = x \exp[-(k|x|)^\alpha]$. This choice has the advantage, that the asymptotic variance of FastICA with this nonlinearity can be computed analytically. The result is $V_{ij}^{1U}(\alpha, k) \approx 2^{-3/\alpha} (\beta_\alpha/k)^{1-2\alpha}$ for large k and $\alpha \leq 1/2$, with β_α defined in (93). Again, $V_{ij}^{1U}(\alpha, k)$ goes to zero for $k \rightarrow \infty$ and all $0 < \alpha < 1/2$. This explains nonexistence of the CRB in this case. Design of an FastICA-based algorithm tailored for long-tailed distributions exceeds the scope of this paper.

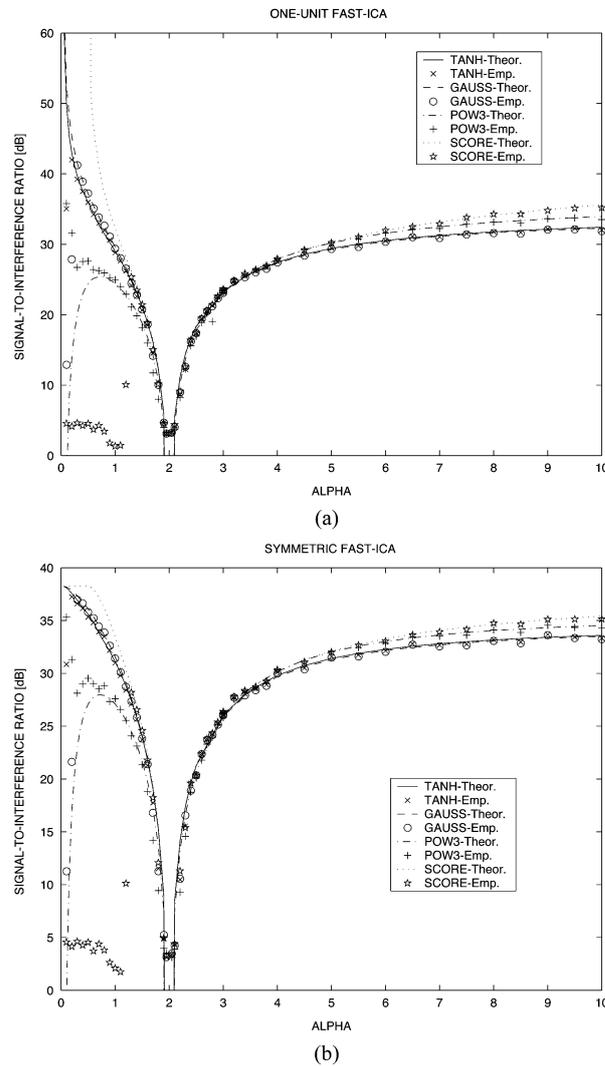


Fig. 3. Performance of (a) one-unit FastICA and (b) symmetric FastICA in separating signals with distribution $GG(\alpha)$ as a function of α .

VI. NUMERICAL RESULTS

Example 1: Four independent random signals with generalized Gaussian distribution (see Appendix C) with parameter α and length $N = 5000$ were generated in 100 independent trials. The signals were mixed with a matrix that was randomly generated in each trial, and demixed again by eight variants of the algorithm: the symmetric FastICA with nonlinearities tanh, Gauss, pow3, and with the score function (dependent on α), as well as the one-unit FastICA with the same nonlinearities, implemented like smart FastICA. The resulting theoretical and empirical SIR is plotted in Fig. 3(a) and (b). An erratic behavior of the empirical results is experienced for small α and nonlinearity pow3. Here, the convergence of sample estimates of the expressions in (19)–(21) to their expectations is slow. We can see that among the α -independent nonlinearities, the “pow3” performs best in the case of $\alpha > 2$ that corresponds to the sub-Gaussian

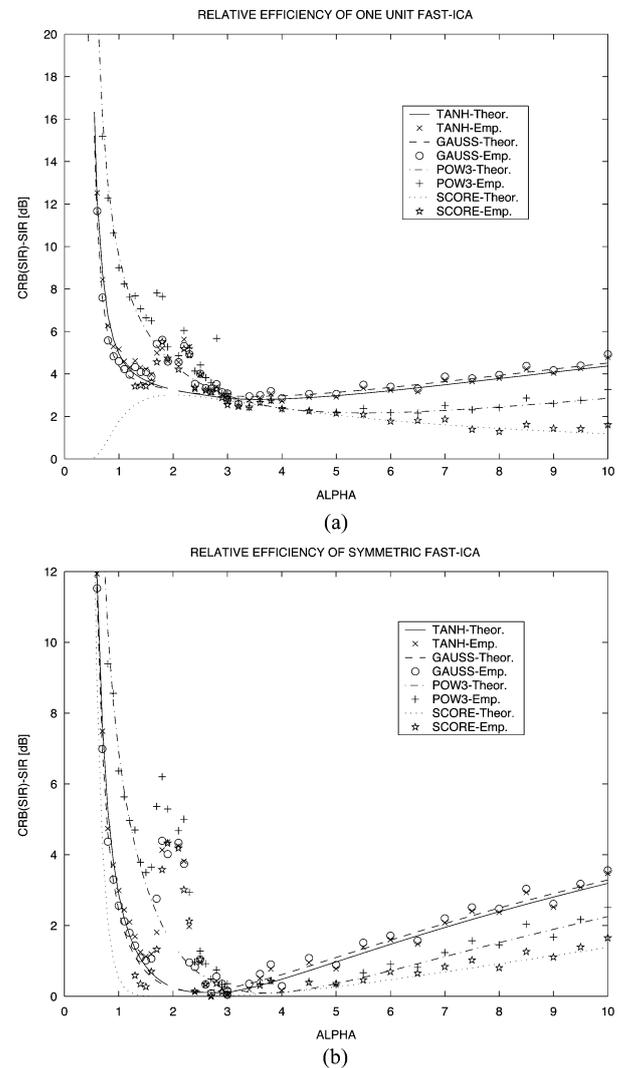


Fig. 4. Relative efficiency of (a) one-unit FastICA and (b) symmetric FastICA.

case, and “gauss” is the best one for $\alpha < 2$ where the distribution is super-Gaussian. FastICA with $g(\cdot)$ equal to the score function does not work properly (does not converge at all) for $\alpha \leq 1$, because the score function is not continuous for these α 's.

Fig. 4 is similar, showing the relative efficiency of the eight methods compared with the corresponding CRB.

Example 2: In the second experiment, we have generated three different components with Gaussian, $GG(\alpha)$, and Laplace distribution of the fixed length $N = 5000$ in 100 independent trials for each α . Signals were randomly mixed and separated by the symmetric FastICA and Smart FastICA with nonlinearity tanh. The resultant SIRs are shown in Fig. 5. Note that this example includes the situation where the mixture includes two Gaussian distributions for $\alpha = 2$. The empirical and theoretical SIR are shown to agree very well. The Smart FastICA outperforms the symmetric version for such α when the one-unit

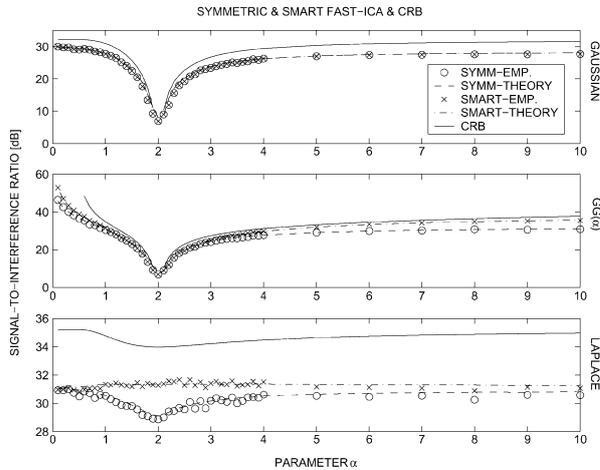


Fig. 5. Performance of symmetric FastICA and smart FastICA separating three different components using “tanh” nonlinearity.

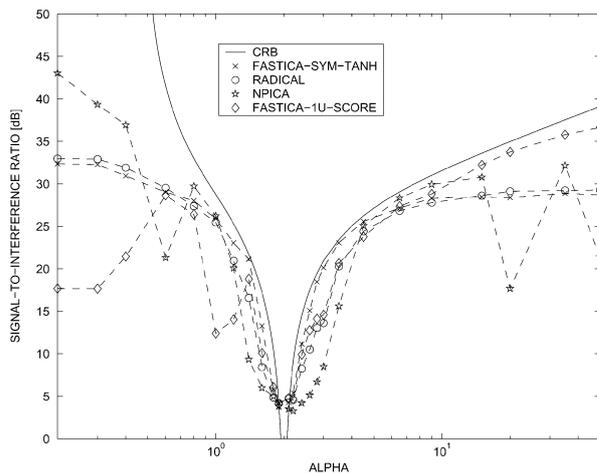


Fig. 6. Comparison of CRB with performance of four ICA techniques.

approach has better variance than the symmetric one, and gives the same result otherwise.

Example 3: In the last experiment, we studied performance of two computationally extensive algorithms that are claimed to be more accurate than older algorithms: RADICAL [22] and NPICA [23]. We tested implementations available on the Internet and compared their performance with the CRB. The simulations are obtained from 50 independent separations of a signal of length $N = 1000$ with $d = 3$ components, all having the same distribution function, GGD(α) (see Fig. 6). In the neighborhood of the point $\alpha = 2$, the symmetric FastICA appears to outperform the other techniques. In general, it appears to give stable results unlike the NPICA.

VII. CONCLUSION

In this paper, 1) a novel technique to improve stability of FastICA is proposed, 2) novel analytical expressions are derived

for the variance of gain matrix elements for one-unit and symmetric FastICA, with an arbitrary twice differentiable nonlinear function and arbitrary probability distribution with finite variance of the independent components in the linear mixture, and 3) the Cramér–Rao bound for the above ICA problem is computed. The CRB does not exist for sources with bounded magnitude and for sources with long-tailed distribution. It was shown that asymptotic variance of estimates produced by FastICA with properly selected nonlinearity can approach the CRB, if the CRB exists, or approach zero, if the CRB does not exist. Good general performance of this popular algorithm is confirmed and possibilities of its further improvements are indicated.

Computer simulations confirm very well the validity of the theoretical predictions.

APPENDIX A

PROOF OF PROPOSITION 1

A. Preliminaries

Invoking assumption (1) of the proposition, and the weak law of large numbers it follows that the sample variance of s_k defined in (4) converges to 1 in probability for N going to infinity, symbolically $\hat{\sigma}_k^2 \xrightarrow{p} 1$, or $\hat{\sigma}_k = 1 + o_p(1)$, where $o_p(\cdot)$ is the stochastic order symbol (see, e.g., Appendix C in [31]). Similarly, thanks to the assumption (3)

$$N^{-1} \mathbf{s}_k^T g(\mathbf{s}_k) \xrightarrow{p} \mu_k \quad (40)$$

$$N^{-1} g^T(\mathbf{s}_k) \mathbf{1}_N \xrightarrow{p} \rho_k. \quad (41)$$

In addition, due to the mutual independence of components, it holds for $\ell \neq k$

$$N^{-1} g^T(\mathbf{s}_k) (\mathbf{s}_\ell \odot \mathbf{s}_\ell) \xrightarrow{p} E[g'(\mathbf{s}_k)] E[\mathbf{s}_\ell^2] = \rho_k \quad (42)$$

where \odot denotes the elementwise product. It can be shown, that the same limits are obtained if $\mathbf{s}_k, \mathbf{s}_\ell$ in (40)–(42) are replaced by the normalized components $\mathbf{u}_k, \mathbf{u}_\ell$, where \mathbf{u}_k is the k th column of \mathbf{U}^T , $k = 1, \dots, d$. Note from (2) that $\mathbf{u}_k = (\mathbf{s}_k - \bar{\mathbf{s}}_k) / \hat{\sigma}_k$, $\bar{\mathbf{s}}_k = O_p(N^{-1/2})$, $\hat{\sigma}_k = 1 + o_p(1)$, consequently $\mathbf{u}_k = \mathbf{s}_k + o_p(1)$, $g(\mathbf{u}_k) = g(\mathbf{s}_k) + o_p(1)$, and

$$\begin{aligned} \mathbf{u}_k^T g(\mathbf{u}_k) &= [\mathbf{s}_k + o_p(1)]^T [g(\mathbf{s}_k) + o_p(1)] \\ &= \mathbf{s}_k^T g(\mathbf{s}_k) + o_p(N) = N\mu_k + o_p(N). \end{aligned} \quad (43)$$

Similarly, it can be shown that

$$g^T(\mathbf{u}_k) \mathbf{1}_N = N\rho_k + o_p(N) \quad (44)$$

$$g^T(\mathbf{u}_k) (\mathbf{u}_\ell \odot \mathbf{u}_\ell) = N\rho_k + o_p(N). \quad (45)$$

Moreover, using the asymptotic expression for \mathbf{R} , to be derived in the next subsection, it can be shown that the relations (40) and (41) hold true as well, if \mathbf{s}_k is replaced with \mathbf{z}_k , that is defined as the k th column of \mathbf{Z}^T , $k = 1, \dots, d$

$$\mathbf{z}_k^T g(\mathbf{z}_k) = N\mu_k + o_p(N) \quad (46)$$

$$g^T(\mathbf{z}_k) \mathbf{1}_N = N\rho_k + o_p(N). \quad (47)$$

B. Asymptotic Behavior of \mathbf{R}

As N goes to infinity, the matrix \mathbf{R} defined in (17) approaches identity matrix in the mean square sense. To see this, note that the diagonal elements of \mathbf{R} are equal to one by definition, and that the off-diagonal elements $\mathbf{R}_{k\ell}$ with $k \neq \ell$ have zero mean. Due to assumed independence of $(\tilde{\mathbf{s}}_k, \hat{\sigma}_k)$ and $(\tilde{\mathbf{s}}_\ell, \hat{\sigma}_\ell)$, it holds

$$\begin{aligned} \mathbb{E}[\mathbf{R}_{k\ell}^2] &= \mathbb{E}\left(\frac{\mathbf{u}_k^T \mathbf{u}_\ell}{N}\right)^2 = \frac{1}{N^2} \mathbb{E}\left(\frac{\tilde{\mathbf{s}}_k^T \tilde{\mathbf{s}}_\ell}{\hat{\sigma}_k \hat{\sigma}_\ell}\right)^2 \\ &= \frac{1}{N^2} \mathbb{E}\left\{\frac{\tilde{\mathbf{s}}_k^T}{\hat{\sigma}_k} \mathbb{E}\left(\frac{\tilde{\mathbf{s}}_\ell \tilde{\mathbf{s}}_\ell^T}{\hat{\sigma}_\ell^2}\right) \frac{\tilde{\mathbf{s}}_k}{\hat{\sigma}_k}\right\} \end{aligned} \quad (48)$$

where $\tilde{\mathbf{s}}_k = \mathbf{s}_k - \bar{\mathbf{s}}_k$. Let $\mathbf{S}^{(\ell)} = \mathbb{E}[\tilde{\mathbf{s}}_\ell \tilde{\mathbf{s}}_\ell^T / \hat{\sigma}_\ell^2]$. Since all elements of $\tilde{\mathbf{s}}_\ell$ have the same distribution, the diagonal elements of $\mathbf{S}^{(\ell)}$ have all the same value

$$\begin{aligned} \mathbf{S}_{nn}^{(\ell)} &= \mathbb{E}[\tilde{s}_{\ell n}^2 / \hat{\sigma}_\ell^2] = \frac{1}{N} \sum_{m=1}^N \mathbb{E}[\tilde{s}_{\ell m}^2 / \hat{\sigma}_\ell^2] \\ &= \frac{1}{N} \mathbb{E}[\tilde{\mathbf{s}}_\ell^T \tilde{\mathbf{s}}_\ell / \hat{\sigma}_\ell^2] = \mathbb{E}[1] = 1 \end{aligned} \quad (49)$$

for $n = 1, \dots, N$. The off-diagonal elements have all the same value as well

$$\begin{aligned} \mathbf{S}_{mn}^{(\ell)} &= \mathbb{E}[\tilde{s}_{\ell m} \tilde{s}_{\ell n} / \hat{\sigma}_\ell^2] = \frac{1}{N-1} \sum_{\substack{q=1 \\ q \neq n}}^N \mathbb{E}[\tilde{s}_{\ell q} \tilde{s}_{\ell n} / \hat{\sigma}_\ell^2] \\ &= -\frac{1}{N-1} \mathbb{E}[\tilde{s}_{\ell n}^2 / \hat{\sigma}_\ell^2] = -\frac{1}{N-1} \end{aligned} \quad (50)$$

for $m, n = 1, \dots, N$. Combining (48), (49), and (50) gives

$$\mathbf{S}^{(\ell)} = \frac{N}{N-1} \mathbf{I} - \frac{1}{N-1} \mathbf{1}_N \mathbf{1}_N^T \quad (51)$$

$$\begin{aligned} \mathbb{E}[\mathbf{R}_{k\ell}^2] &= \frac{1}{N^2} \mathbb{E}\left\{\frac{\tilde{\mathbf{s}}_k^T}{\hat{\sigma}_k} \mathbf{S}^{(\ell)} \frac{\tilde{\mathbf{s}}_k}{\hat{\sigma}_k}\right\} \\ &= \frac{1}{N(N-1)} \mathbb{E}\left\{\frac{\tilde{\mathbf{s}}_k^T \tilde{\mathbf{s}}_k}{\hat{\sigma}_k \hat{\sigma}_k}\right\} = \frac{1}{N-1}. \end{aligned} \quad (52)$$

It follows from (52) that

$$\Delta \mathbf{R} \stackrel{\text{def}}{=} \mathbf{R} - \mathbf{I} = O_p(N^{-1/2}) \quad (53)$$

where $O_p(\cdot)$ denotes a standard stochastic order symbol, or a matrix of stochastic order symbols of appropriate dimension. Using Lemma 1 in Appendix B, it can be derived that

$$\mathbf{R}^{-1/2} = \mathbf{I} - \frac{1}{2} \Delta \mathbf{R} + O_p(N^{-1}). \quad (54)$$

C. Approximation for $\mathbf{Z}, g(\mathbf{Z})$

Obviously, $\mathbf{U} = O_p(1)$ and

$$\begin{aligned} \mathbf{Z} &= \mathbf{R}^{-1/2} \mathbf{U} = \left(\mathbf{I} - \frac{1}{2} \Delta \mathbf{R} + O_p(N^{-1})\right) \mathbf{U} \\ &= \mathbf{U} - \frac{1}{2} \Delta \mathbf{R} \mathbf{U} + O_p(N^{-1}). \end{aligned} \quad (55)$$

A Taylor series expansion of function $g(\cdot)$ in a neighborhood of $\mathbf{Z} = \mathbf{U}$ gives

$$g(\mathbf{Z}) = g(\mathbf{U}) + g'(\mathbf{U}) \odot \Delta \mathbf{Z} + O_p(N^{-1}) \quad (56)$$

where \odot denotes the elementwise product and

$$\Delta \mathbf{Z} \stackrel{\text{def}}{=} \mathbf{Z} - \mathbf{U} = -\frac{1}{2} \Delta \mathbf{R} \mathbf{U} + O_p(N^{-1}). \quad (57)$$

Using (17), the k th column of $\Delta \mathbf{Z}^T$ is

$$\Delta \mathbf{z}_k = -\frac{1}{2N} \sum_{\substack{m=1 \\ m \neq k}}^d \mathbf{u}_k^T \mathbf{u}_m \mathbf{u}_m + O_p(N^{-1}). \quad (58)$$

D. Approximation for $\mathbf{W}^+, \hat{\mathbf{W}}$

Inserting $\mathbf{W} = \mathbf{I}$ in (11), the $k\ell$ th element of \mathbf{W}^+ reads

$$\mathbf{W}_{k\ell}^+ = \begin{cases} g(\mathbf{z}_k^T) \mathbf{z}_\ell & \text{for } k \neq \ell \\ g(\mathbf{z}_k^T) \mathbf{z}_k - g'(\mathbf{z}_k^T) \mathbf{1}_N & \text{for } k = \ell. \end{cases} \quad (59)$$

For $k = \ell$, we get using (46) and (47)

$$\mathbf{W}_{kk}^+ = N(\mu_k - \rho_k) + o_p(N). \quad (60)$$

For $k \neq \ell$, we get using (56)

$$\begin{aligned} \mathbf{W}_{k\ell}^+ &= g(\mathbf{z}_k^T) \mathbf{z}_\ell \\ &= [g(\mathbf{u}_k^T) + g'(\mathbf{u}_k^T) \odot \Delta \mathbf{z}_k^T + O_p(N^{-1})] [\mathbf{u}_\ell + \Delta \mathbf{z}_\ell] \\ &= \left[g(\mathbf{u}_k^T) - \frac{1}{2N} \sum_{\substack{m=1 \\ m \neq k}}^d g'(\mathbf{u}_k^T) \odot (\mathbf{u}_k^T \mathbf{u}_m \mathbf{u}_m^T) \right] \\ &\quad \cdot \left[\mathbf{u}_\ell - \frac{1}{2N} \sum_{\substack{m=1 \\ m \neq \ell}}^d \mathbf{u}_\ell^T \mathbf{u}_m \mathbf{u}_m \right] + O_p(1). \end{aligned} \quad (61)$$

The reminder term in (61) has the stochastic order $O_p(1)$ for the following reason. It holds that $\mathbf{u}_k = O_p(1)$, and the remainder in the expansion of $g(\mathbf{z}_k^T)$, that is $O_p(N^{-1})$, are N -element vectors. The stochastic order notation is valid uniformly over elements of these vectors. Hence, scalar product of these two vectors is $O_p(1)$. Similarly, $\Delta \mathbf{z}_\ell = O_p(N^{-1/2})$, $g'(\mathbf{u}_k^T) \odot \Delta \mathbf{z}_\ell = O_p(N^{-1/2})$, and $(g'(\mathbf{u}_k^T) \odot \Delta \mathbf{z}_k^T) \Delta \mathbf{z}_\ell = O_p(1)$.

In the following, let \mathbf{g}_k and \mathbf{g}'_k stand for $g(\mathbf{u}_k)$ and $g'(\mathbf{u}_k)$, respectively, $k = 1, \dots, d$. Note that, due to (21) and due to independence of $\mathbf{u}_k, \mathbf{u}_\ell$ for $k \neq \ell$, it holds

$$\begin{aligned} \mathbb{E}[(\mathbf{g}_k^T \mathbf{u}_\ell)^2] &= \mathbb{E}[\mathbf{g}_k^T \mathbf{u}_\ell \mathbf{u}_\ell^T \mathbf{g}_k] = \mathbb{E}[\mathbf{g}_k^T \mathbb{E}(\mathbf{u}_\ell \mathbf{u}_\ell^T) \mathbf{g}_k] \\ &= \mathbb{E}[\mathbf{g}_k^T \mathbf{g}_k] = N\beta_k. \end{aligned} \quad (62)$$

It follows from (19) and (62) that

$$\mathbf{g}_k^T \mathbf{u}_\ell = N\mu_k \delta_{k\ell} + o_p(N). \quad (63)$$

Similarly

$$\mathbf{u}_k^T \mathbf{u}_\ell = N\delta_{k\ell} + O_p(N^{1/2}). \quad (64)$$

Applying (63) and (64) and (43)–(45) in (61) gives

$$\begin{aligned} \mathbf{W}_{k\ell}^+ &= \mathbf{g}_k^T \mathbf{u}_\ell - \frac{1}{2N} \mathbf{g}_k^T \mathbf{u}_k \mathbf{u}_k^T \mathbf{u}_\ell \\ &\quad - \frac{1}{2N} [\mathbf{g}_k^T \odot (\mathbf{u}_k^T \mathbf{u}_\ell \mathbf{u}_\ell^T)] \mathbf{u}_\ell + O_p(1) \\ &= \mathbf{g}_k^T \mathbf{u}_\ell - \frac{1}{2N} (\mathbf{g}_k^T \mathbf{u}_k) \mathbf{u}_k^T \mathbf{u}_\ell \\ &\quad - \frac{1}{2N} (\mathbf{u}_k^T \mathbf{u}_\ell) \mathbf{g}_k^T (\mathbf{u}_\ell \odot \mathbf{u}_\ell) + O_p(1) \\ &= \mathbf{g}_k^T \mathbf{u}_\ell - \frac{\mu_k + \rho_k}{2} \mathbf{u}_k^T \mathbf{u}_\ell + o_p(N^{1/2}). \end{aligned} \quad (65)$$

E. Approximation for $\hat{\mathbf{W}}, \mathbf{G}$

Note that if $\hat{\mathbf{W}}_{kk}^+ < 0$ for some k , the k th diagonal element of the demixing matrices $\hat{\mathbf{W}}_{kk}^{1U}$ and $\hat{\mathbf{W}}_{kk}^{\text{SYM}}$ may have the wrong sign, i.e., it might be close to -1 instead of 1 . It corresponds to reversed sign of the k th estimated independent component. In the one-unit version of the algorithm, the sign can be corrected by replacing the normalization in (10) by an equivalent formula

$$\hat{\mathbf{W}}_{k\ell}^{1U} = \frac{\mathbf{W}_{k\ell}^+}{\mathbf{W}_{kk}^+} = \frac{\mathbf{W}_{k\ell}^+}{N(\mu_k - \rho_k)} + o_p(N^{-1/2}). \quad (66)$$

Similarly, using Lemma 2 in Appendix B, the asymptotically equivalent sign corrected expression for the estimated demixing matrix is

$$\hat{\mathbf{W}}_{k\ell}^{\text{SYM}} = \delta_{k\ell} + \frac{\mathbf{W}_{k\ell}^+ \text{sign}(\mathbf{W}_{kk}^+) - \mathbf{W}_{\ell k}^+ \text{sign}(\mathbf{W}_{\ell\ell}^+)}{|\mathbf{W}_{kk}^+| + |\mathbf{W}_{\ell\ell}^+|} + o_p(N^{-1/2}). \quad (67)$$

For both estimator variants, $\hat{\mathbf{W}}^{1U}$ and $\hat{\mathbf{W}}^{\text{SYM}}$ we can write

$$\Delta \mathbf{W} = \hat{\mathbf{W}} - \mathbf{I} = O_p(N^{-1/2}). \quad (68)$$

Since

$$\begin{aligned} \mathbf{G} &= \hat{\mathbf{W}} \mathbf{R}^{-1/2} = (\mathbf{I} + \Delta \mathbf{W}) \left(\mathbf{I} - \frac{1}{2} \Delta \mathbf{R} + O_p(N^{-1}) \right) \\ &= \mathbf{I} + \Delta \mathbf{W} - \frac{1}{2} \Delta \mathbf{R} + O_p(N^{-1}) \end{aligned} \quad (69)$$

the gain matrix off-diagonal elements read

$$\mathbf{G}_{k\ell} = \hat{\mathbf{W}}_{k\ell} - \frac{1}{2N} \mathbf{u}_k^T \mathbf{u}_\ell + O_p(N^{-1}). \quad (70)$$

For the one-unit variant, we get

$$\begin{aligned} N^{1/2} \mathbf{G}_{k\ell}^{1U} &= N^{1/2} \frac{\mathbf{W}_{k\ell}^+}{N(\mu_k - \rho_k)} - \frac{1}{2N^{1/2}} \mathbf{u}_k^T \mathbf{u}_\ell + o_p(1) \\ &= \frac{N^{-1/2}}{\mu_k - \rho_k} (\mathbf{g}_k^T \mathbf{u}_\ell - \mu_k \mathbf{u}_k^T \mathbf{u}_\ell) + o_p(1). \end{aligned} \quad (71)$$

Finally, we show that (71) can be rewritten in terms of $\mathbf{s}_k, \mathbf{s}_\ell$ in an asymptotically equivalent formula

$$N^{1/2} \mathbf{G}_{k\ell}^{1U} = \frac{N^{-1/2}}{\mu_k - \rho_k} (g(\mathbf{s}_k^T) \mathbf{s}_\ell - \mu_k \mathbf{s}_k^T \mathbf{s}_\ell) + o_p(1). \quad (72)$$

To see that, note that

$$\begin{aligned} \mathbf{u}_k^T \mathbf{u}_\ell &= \left(\frac{\mathbf{s}_k - \bar{\mathbf{s}}_k}{\hat{\sigma}_k} \right)^T \frac{\mathbf{s}_\ell - \bar{\mathbf{s}}_\ell}{\hat{\sigma}_\ell} = \frac{\mathbf{s}_k^T \mathbf{s}_\ell - \bar{\mathbf{s}}_k^T \bar{\mathbf{s}}_\ell}{\hat{\sigma}_k \hat{\sigma}_\ell} \\ &= \frac{\mathbf{s}_k^T \mathbf{s}_\ell - O_p(1)}{1 + o_p(1)} = \mathbf{s}_k^T \mathbf{s}_\ell + o_p(N^{1/2}). \end{aligned} \quad (73)$$

Similarly, it can be shown that

$$g(\mathbf{u}_k^T) \mathbf{u}_\ell = g(\mathbf{s}_k^T) \mathbf{s}_\ell + o_p(N^{1/2}). \quad (74)$$

Equation (74) concludes the proof of (72). Now, applying the central limit theorem to (72) implies that the distribution of $N^{1/2} \mathbf{G}_{k\ell}^{1U}$ is asymptotic normal with zero mean and variance equal to the variance of the leading term in (72). Using (62)–(64) gives

$$\begin{aligned} V_{k\ell}^{1U} &= \text{var} \left[\frac{N^{-1/2}}{\mu_k - \rho_k} (g(\mathbf{s}_k^T) \mathbf{s}_\ell - \mu_k \mathbf{s}_k^T \mathbf{s}_\ell) \right] \\ &= \frac{N^{-1}}{(\mu_k - \rho_k)^2} \text{var} [(g(\mathbf{s}_k^T) \mathbf{s}_\ell - \mu_k \mathbf{s}_k^T \mathbf{s}_\ell)] \\ &= \frac{\beta_k - \mu_k^2}{(\mu_k - \rho_k)^2}. \end{aligned} \quad (75)$$

Similarly, for symmetric FastICA, it holds using (67) that

$$\begin{aligned} N^{1/2} \mathbf{G}_{k\ell}^{\text{SYM}} &= N^{1/2} \hat{\mathbf{W}}_{k\ell}^{\text{SYM}} - \frac{1}{2N^{1/2}} \mathbf{u}_k^T \mathbf{u}_\ell + o_p(1) \\ &= \frac{\mathbf{W}_{k\ell}^+ \text{sign}(\mu_k - \rho_k) - \mathbf{W}_{\ell k}^+ \text{sign}(\mu_\ell - \rho_\ell)}{N^{1/2} (|\mu_k - \rho_k| + |\mu_\ell - \rho_\ell|)} \\ &\quad - \frac{1}{2N^{1/2}} \mathbf{u}_k^T \mathbf{u}_\ell + o_p(1). \end{aligned} \quad (76)$$

The variance of the leading term in (76) results, after some algebra using (63)–(65), in

$$V_{k\ell}^{\text{SYM}} = \frac{\beta_k - \mu_k^2 + \beta_\ell - \mu_\ell^2 + (\mu_\ell - \rho_\ell)^2}{(|\mu_k - \rho_k| + |\mu_\ell - \rho_\ell|)^2} \quad (77)$$

as desired.

APPENDIX B LEMMAS

Lemma 1: Let \mathbf{R}_0 and \mathbf{R} be positive definite matrices of the same dimension and $\Delta \mathbf{R} = \mathbf{R} - \mathbf{R}_0$. Then, for $\|\Delta \mathbf{R}\| \rightarrow 0$ (in any matrix norm), it holds

$$\mathbf{R}^{-1/2} = \mathbf{R}_0^{-1/2} + \Delta \mathbf{M} + O(\|\Delta \mathbf{R}\|^2) \quad (78)$$

where

$$\Delta \mathbf{M} = -\text{unvec} \left\{ \left[\mathbf{I} \otimes \mathbf{R}_0^{-1/2} + \mathbf{R}_0^{-1/2} \otimes \mathbf{I} \right]^{-1} \times \text{vec} \left(\mathbf{R}_0^{-1} \Delta \mathbf{R} \mathbf{R}_0^{-1} \right) \right\}. \quad (79)$$

Here, “vec” denotes the operation that reshapes columns of a matrix in one long column vector, and “unvec” is the corresponding inverse operation.

In the case that \mathbf{R}_0 is diagonal, $\mathbf{R}_0 = \text{diag}(r_1, \dots, r_d)$ is a diagonal matrix with $r_k > 0$ for $k = 1, \dots, d$, then $\Delta \mathbf{M}$ has elements

$$\Delta \mathbf{M}_{k\ell} = -\frac{\Delta \mathbf{R}_{k\ell}}{\sqrt{r_k} \sqrt{r_\ell} (\sqrt{r_k} + \sqrt{r_\ell})}. \quad (80)$$

In the case that $\mathbf{R}_0 = \mathbf{I}$, (80) gives $\Delta \mathbf{M} = -(1/2)\Delta \mathbf{R}$.

Proof: The identity

$$\mathbf{I} = \mathbf{R}(\mathbf{R}^{-1/2})^2 = (\mathbf{R}_0 + \Delta \mathbf{R})(\mathbf{R}_0^{-1/2} + \Delta \mathbf{M})^2 \quad (81)$$

leads, after neglecting higher than first-order terms in $\Delta \mathbf{R}$ and $\Delta \mathbf{M}$, to the relation

$$\Delta \mathbf{M} \mathbf{R}_0^{-1/2} + \mathbf{R}_0^{-1/2} \Delta \mathbf{M} = -\mathbf{R}_0^{-1} \Delta \mathbf{R} \mathbf{R}_0^{-1} \quad (82)$$

or, equivalently

$$\left[\mathbf{I} \otimes \mathbf{R}_0^{-1/2} + \mathbf{R}_0^{-1/2} \otimes \mathbf{I} \right] \text{vec} \Delta \mathbf{M} = -\text{vec} \left[\mathbf{R}_0^{-1} \Delta \mathbf{R} \mathbf{R}_0^{-1} \right].$$

The desired solution (79) follows. ■

Lemma 2: Let

$$\mathbf{W} = \mathbf{W}_0 + \Delta \mathbf{W} \quad (83)$$

where $\mathbf{W}_0 = \text{diag}(w_1, \dots, w_d)$ is a diagonal matrix, and let $w_k > 0$ for $k = 1, \dots, d$. Then, for $\|\Delta \mathbf{W}\| \rightarrow 0$ it holds

$$\mathbf{S} \stackrel{\text{def}}{=} (\mathbf{W} \mathbf{W}^T)^{-1/2} \mathbf{W} = \mathbf{I} + \Delta \mathbf{S} + O(\|\Delta \mathbf{W}\|^2) \quad (84)$$

where $\Delta \mathbf{S}$ has elements

$$\Delta \mathbf{S}_{k\ell} = \frac{\Delta \mathbf{W}_{k\ell} - \Delta \mathbf{W}_{\ell k}}{w_k + w_\ell}. \quad (85)$$

Proof: Using Lemma 1 gives

$$(\mathbf{W} \mathbf{W}^T)^{-1/2} = \mathbf{V}_0 + \Delta \mathbf{V} + O(\|\Delta \mathbf{W}\|^2) \quad (86)$$

where

$$\mathbf{V}_0 = (\mathbf{W}_0 \mathbf{W}_0^T)^{-1/2} = \text{diag} \left(\frac{1}{w_1}, \dots, \frac{1}{w_d} \right) \quad (87)$$

and $\Delta \mathbf{V}$ has as elements

$$\begin{aligned} \Delta \mathbf{V}_{k\ell} &= -\frac{(\mathbf{W} \mathbf{W}^T - \mathbf{W}_0 \mathbf{W}_0^T)_{k\ell}}{w_k w_\ell (w_k + w_\ell)} \\ &= -\frac{(\mathbf{W}_0 \Delta \mathbf{W}^T + \Delta \mathbf{W} \mathbf{W}_0^T)_{k\ell}}{w_k w_\ell (w_k + w_\ell)} + O(\|\Delta \mathbf{W}\|^2) \\ &= -\frac{w_k \Delta \mathbf{W}_{\ell k} + \Delta \mathbf{W}_{k\ell} w_\ell}{w_k w_\ell (w_k + w_\ell)} + O(\|\Delta \mathbf{W}\|^2). \end{aligned} \quad (88)$$

Then

$$\begin{aligned} \mathbf{S} &= (\mathbf{V}_0 + \Delta \mathbf{V} + O(\|\Delta \mathbf{W}\|^2))(\mathbf{W}_0 + \Delta \mathbf{W}) \\ &= \mathbf{I} + \mathbf{V}_0 \Delta \mathbf{W} + \Delta \mathbf{V} \mathbf{W}_0 + O(\|\Delta \mathbf{W}\|^2) \end{aligned} \quad (89)$$

and hence the leading term $\Delta \mathbf{S}$ has elements

$$\Delta \mathbf{S}_{k\ell} = \frac{1}{w_k} \Delta \mathbf{W}_{k\ell} + \Delta \mathbf{V}_{k\ell} w_\ell = \frac{\Delta \mathbf{W}_{k\ell} - \Delta \mathbf{W}_{\ell k}}{w_k + w_\ell}. \quad \blacksquare$$

APPENDIX C

COMPUTING FISHER INFORMATION MATRIX

Applying the fact that $(\partial \det \mathbf{W})/(\partial w_{ij}) = a_{ji} \det \mathbf{W}$, we get from (32)

$$\begin{aligned} \frac{\partial f_{\mathbf{x}}}{\partial w_{uv}} &= \frac{\partial f_{\mathbf{s}}(\mathbf{W} \mathbf{x}) |\det \mathbf{W}|}{\partial w_{uv}} \\ &= \frac{\partial |\det \mathbf{W}|}{\partial w_{uv}} f_{\mathbf{s}}(\mathbf{W} \mathbf{x}) + |\det \mathbf{W}| \frac{\partial f_{\mathbf{s}}(\mathbf{W} \mathbf{x})}{\partial w_{uv}} \\ &= |\det \mathbf{W}| a_{vu} f_{\mathbf{s}}(\mathbf{W} \mathbf{x}) + |\det \mathbf{W}| \\ &\quad \times \sum_{k=1}^d \sum_{l=1}^N \left(\prod_{\substack{i=1, \dots, d \\ j=1, \dots, N \\ -(i=k \wedge j=l)}} f_i((\mathbf{W} \mathbf{x})_{ij}) \right) \frac{\partial f_k((\mathbf{W} \mathbf{x})_{kl})}{\partial w_{uv}} \\ &= |\det \mathbf{W}| a_{vu} f_{\mathbf{s}}(\mathbf{W} \mathbf{x}) + |\det \mathbf{W}| \\ &\quad \times \sum_{k=1}^d \sum_{l=1}^N f_{\mathbf{s}}(\mathbf{W} \mathbf{x}) \frac{\partial}{\partial w_{uv}} \frac{f_k((\mathbf{W} \mathbf{x})_{kl})}{f_k((\mathbf{W} \mathbf{x})_{kl})}. \end{aligned}$$

Next

$$\begin{aligned} \frac{\partial f_k((\mathbf{W} \mathbf{x})_{kl})}{\partial w_{uv}} &= f'_k((\mathbf{W} \mathbf{x})_{kl}) \frac{\partial (\mathbf{W} \mathbf{x})_{kl}}{\partial w_{uv}} \\ &= f'_k((\mathbf{W} \mathbf{x})_{kl}) \sum_{\kappa=1}^d \frac{\partial w_{k\kappa}}{\partial w_{uv}} x_{\kappa l} \\ &= f'_k((\mathbf{W} \mathbf{x})_{kl}) \delta_{ku} x_{vl} \\ &= f'_k((\mathbf{W} \mathbf{x})_{kl}) \delta_{ku} \sum_{k=1}^d a_{vk} s_{kl}. \end{aligned}$$

Returning to the above formula, we get

$$\begin{aligned} \frac{\partial f_{\mathbf{x}}}{\partial w_{uv}} &= |\det \mathbf{W}| f_{\mathbf{s}}(\mathbf{W} \mathbf{x}) \\ &\quad \times \left[a_{vu} + \sum_{l=1}^N \sum_{k=1}^d \frac{f'_u((\mathbf{W} \mathbf{x})_{ul})}{f_u((\mathbf{W} \mathbf{x})_{ul})} a_{vk} s_{kl} \right]. \end{aligned}$$

From (1), it follows that $\mathbf{s} = \mathbf{A}^{-1} \mathbf{x} = \mathbf{W} \mathbf{x}$, and consequently

$$\frac{\partial f_{\mathbf{x}}}{\partial w_{uv}} = |\det \mathbf{W}| f_{\mathbf{s}}(\mathbf{s}) \left[a_{vu} + \sum_{l=1}^N \sum_{k=1}^d \frac{f'_u(s_{ul})}{f_u(s_{ul})} a_{vk} s_{kl} \right].$$

Using this, we can directly compute the mn th entry of the Fisher information matrix.

$$\begin{aligned} \mathbf{F}_{mn} &= \mathbb{E} \left[\left(\frac{|\det \mathbf{W}|^{-2}}{f^2} \right) \frac{\partial f_{\mathbf{x}}}{\partial w_{uv}} \frac{\partial f_{\mathbf{x}}}{\partial w_{pr}} \right] \\ &= a_{vu} a_{rp} + a_{rp} \mathbb{E} \left[\sum_{l=1}^N \sum_{k=1}^d \frac{f'_u(s_{ul})}{f_u(s_{ul})} a_{vk} s_{kl} \right] \\ &\quad + a_{vu} \mathbb{E} \left[\sum_{i=1}^N \sum_{j=1}^d \frac{f'_p(s_{pi})}{f_p(s_{pi})} a_{rj} s_{ji} \right] \\ &\quad + \mathbb{E} \left[\sum_{l=1}^N \sum_{i=1}^N \sum_{k=1}^d \sum_{j=1}^d \frac{f'_u(s_{ul})}{f_u(s_{ul})} \frac{f'_p(s_{pi})}{f_p(s_{pi})} s_{kl} s_{ji} a_{vk} a_{rj} \right]. \end{aligned}$$

The second and the third term are equal to $-N a_{vu} a_{rp}$, because $\mathbb{E}[(f'_u(s_{ul})/(f_u(s_{ul}))s_{kl})] = -\delta_{ku}$. To simplify the last term, we shall consider two cases:

1) $u \neq p$, then

$$\begin{aligned} \sum_{l=1}^N \sum_{i=1}^N \sum_{k=1}^d \sum_{j=1}^d \mathbb{E} \left[\underbrace{\frac{f'_u(s_{ul})}{f_u(s_{ul})} \frac{f'_p(s_{pi})}{f_p(s_{pi})} s_{kl} s_{ji}}_{\delta_{ku} \delta_{jp} + \delta_{kp} \delta_{ju} \delta_{il}} \right] a_{vk} a_{rj} \\ = N^2 a_{vu} a_{rp} + N a_{vp} a_{ru} \end{aligned}$$

2) $u = p$, then

$$\begin{aligned} \mathbb{E} \left[\sum_{l=1}^N \sum_{i=1}^N \sum_{k=1}^d \sum_{j=1}^d \frac{f'_u(s_{ul})}{f_u(s_{ul})} \frac{f'_u(s_{ui})}{f_u(s_{ui})} s_{kl} s_{ji} a_{vk} a_{rj} \right] \\ = \sum_{i=1}^N \sum_{k=1}^d \sum_{j=1}^d \mathbb{E} \left[\underbrace{\left[-\frac{f'_u(s_{ui})}{f_u(s_{ui})} \right]^2}_{\delta_{kj} \times \dots} s_{ki} s_{ji} \right] a_{vk} a_{rj} \\ + \sum_{\substack{i,l=1 \\ i \neq l}}^N \sum_{k=1}^d \sum_{j=1}^d \mathbb{E} \left[\underbrace{\frac{f'_u(s_{ul})}{f_u(s_{ul})} \frac{f'_u(s_{ui})}{f_u(s_{ui})} s_{kl} s_{ji}}_{\delta_{ku} \delta_{ju} \times \dots} \right] a_{vk} a_{rj} \\ = \sum_{i=1}^N \mathbb{E} \left[\underbrace{\left[-\frac{f'_u(s_{ui})}{f_u(s_{ui})} \right]^2}_{\mathbb{E}[\psi_u^2(\xi_u)]} \sum_{j=1, j \neq u}^d \underbrace{\mathbb{E}[s_{ji}^2]}_1 a_{vj} a_{rj} \right] \\ + a_{vu} a_{ru} \sum_{i=1}^N \mathbb{E} \left[\underbrace{\left[-\frac{f'_u(s_{ui})}{f_u(s_{ui})} s_{ui} \right]^2}_{\mathbb{E}[\psi_u^2(\xi_u) \xi_u^2]} \right] + \sum_{\substack{i,l=1 \\ i \neq l}}^N a_{vu} a_{ru} \\ = N \left(\mathbb{E}[\psi_u^2(\xi_u)] \sum_{\substack{j=1 \\ j \neq u}}^d a_{vj} a_{rj} \right. \\ \left. + (\mathbb{E}[\psi_u^2(\xi_u) \xi_u^2] + (N-1)) a_{vu} a_{ru} \right). \end{aligned}$$

Here, ξ_u denotes a random variable with pdf f_u , and ψ_u denotes its score function, i.e., $\psi_u(x) = -(f'_u(x)/f_u(x))$. After a few simplifications, (34) follows. ■

APPENDIX D COMPUTING MATRIX INVERSION OF $\mathbf{F}_{\mathbf{I}}$

Definition (36) can be rewritten as $\mathbf{F}_{\mathbf{I}} = (N-1)^2 \mathbf{F}_1 + N(\mathbf{P} + \mathbf{\Sigma})$, where mn th element of \mathbf{F}_1 , \mathbf{P} and $\mathbf{\Sigma}$ are $\delta_{ji} \delta_{vu} \delta_{vi} (\eta_i - \kappa_i - 2) + \delta_{iu} \delta_{vj} \kappa_i$, respectively, for $m = (i-1)d + j$ and $n = (u-1)d + v$. Note that \mathbf{F}_1 is a rank-one matrix, $\mathbf{F}_1 = \mathbf{e} \mathbf{e}^T$, where $\mathbf{e} = \text{vec}(\mathbf{I})$. Applying the matrix inversion lemma gives

$$\mathbf{F}_{\mathbf{I}}^{-1} = \frac{1}{N} \left[(\mathbf{P} + \mathbf{\Sigma})^{-1} - \frac{(\mathbf{P} + \mathbf{\Sigma})^{-1} \mathbf{e} \mathbf{e}^T (\mathbf{P} + \mathbf{\Sigma})^{-1}}{N(N-1)^{-2} + \mathbf{e}^T (\mathbf{P} + \mathbf{\Sigma})^{-1} \mathbf{e}} \right].$$

To compute the inversion $(\mathbf{P} + \mathbf{\Sigma})^{-1}$, note that $\mathbf{\Sigma}$ is diagonal

$$\mathbf{\Sigma} = \text{diag}(\underbrace{\eta_1 - 2, \kappa_1, \dots, \kappa_1}_d, \underbrace{\kappa_2, \eta_2 - 2, \kappa_2, \dots, \kappa_2}_d, \dots) \quad (90)$$

and \mathbf{P} is a special permutation matrix such that $\mathbf{P} \text{vec}(\mathbf{M}) = \text{vec}(\mathbf{M}^T)$ for any $d \times d$ matrix \mathbf{M} . Moreover, \mathbf{P} obeys $\mathbf{P} \mathbf{P} = \mathbf{I}$, and for any diagonal matrix $\mathbf{D} = \text{diag}(\mathbf{d})$ it holds that

$$\mathbf{P} \mathbf{D} = \mathbf{D}' \mathbf{P}$$

where $\mathbf{D}' = \text{diag}(\mathbf{P} \mathbf{d}) = \mathbf{P} \mathbf{D} \mathbf{P}$. These facts can be used to show that the inversion of $\mathbf{P} + \mathbf{\Sigma}$ can be written in the form $\mathbf{D}_1 + \mathbf{D}_2 \mathbf{P}$ for suitable diagonal matrices \mathbf{D}_1 and \mathbf{D}_2 . The equality

$$(\mathbf{P} + \mathbf{\Sigma})(\mathbf{D}_1 + \mathbf{D}_2 \mathbf{P}) = \mathbf{I}$$

is fulfilled for $\mathbf{\Sigma} \mathbf{D}_1 + \mathbf{D}'_2 = \mathbf{I}$ and $\mathbf{D}'_1 + \mathbf{\Sigma} \mathbf{D}_2 = \mathbf{0}$. Hence

$$\mathbf{D}_1 = (\mathbf{\Sigma}' \mathbf{\Sigma} - \mathbf{I})^{-1} \mathbf{\Sigma}' \quad \text{and} \quad \mathbf{D}_2 = -\mathbf{\Sigma}^{-1} \mathbf{D}'_1$$

where $\mathbf{\Sigma}' = \mathbf{P} \mathbf{\Sigma} \mathbf{P}$ and $\mathbf{D}'_1 = \mathbf{P} \mathbf{D}_1 \mathbf{P}$. Finally, it can be shown that $(\mathbf{F}_{\mathbf{I}}^{-1})_{mm} = N^{-1} (\mathbf{D}_1)_{mm}$ for $m = (i-1)d + j$, $i \neq j$. (38) easily follows.

APPENDIX E PROOF THAT $\kappa \geq 1$

Assume that f is a positive probability density function of a random variable with zero mean and variance 1, such that κ in (24) exists. Then, integration *per partes* and the Cauchy–Schwartz inequality gives

$$\begin{aligned} 1 &= \int_{\mathcal{R}} f(x) dx = - \int_{\mathcal{R}} x f'(x) dx \\ &\leq \sqrt{\int_{\mathcal{R}} x^2 f(x) dx} \sqrt{\int_{\mathcal{R}} \left(\frac{f'(x)}{f(x)} \right)^2 f(x) dx} \\ &= 1 \cdot \sqrt{\kappa}. \end{aligned} \quad (91)$$

The equality in (91) is attained if f'/f is proportional to x , which necessarily means that the distribution is Gaussian.

APPENDIX F GENERALIZED GAUSSIAN DISTRIBUTION FAMILY

Consider the generalized Gaussian density function with parameter α , zero mean and variance one, as [19]

$$f_{\alpha}(x) = \frac{\alpha \beta_{\alpha}}{2\Gamma(1/\alpha)} \exp\{-(\beta_{\alpha}|x|)^{\alpha}\} \quad (92)$$

where $\alpha > 0$ is a positive parameter that controls the distribution's exponential rate of decay, $\Gamma(\cdot)$ is the Gamma function, and

$$\beta_\alpha = \sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}. \quad (93)$$

This generalized Gaussian family encompasses the ordinary standard normal distribution for $\alpha = 2$, the Laplacean distribution for $\alpha = 1$, and the uniform distribution in the limit $\alpha \rightarrow \infty$.

The k th absolute moment for the distribution is

$$E_\alpha\{|x|^k\} = \int_{-\infty}^{\infty} |x|^k f_\alpha(x) dx = \frac{1}{\beta_\alpha^k} \frac{\Gamma(\frac{k+1}{\alpha})}{\Gamma(\frac{1}{\alpha})}. \quad (94)$$

The score function of the distribution is

$$\psi_\alpha(x) = -\frac{\partial f_\alpha(x)}{f_\alpha(x)} = \frac{|x|^{\alpha-1} \text{sign}(x)}{E_\alpha[|x|^\alpha]}. \quad (95)$$

Then, simple computations give

$$\begin{aligned} \kappa_\alpha &= E_\alpha[\psi_\alpha^2(x)] = \frac{E_\alpha[|x|^{2\alpha-2}]}{\{E_\alpha[|x|^\alpha]\}^2} \\ &= \begin{cases} \frac{\Gamma(2-\frac{1}{\alpha})\Gamma(\frac{3}{\alpha})}{[\Gamma(1+\frac{1}{\alpha})]^2} & \text{for } \alpha > 1/2 \\ +\infty & \text{otherwise.} \end{cases} \end{aligned} \quad (96)$$

REFERENCES

- [1] S.-I. Amari and A. Cichocki, *Adaptive Blind Signal and Image Processing*. New York: Wiley, 2002.
- [2] P. Comon, "Independent components analysis: A new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.
- [3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley-Interscience, 2001.
- [4] D. Donoho, "On minimum entropy deconvolution," in *Applied Time-Series Analysis II*. New York: Academic, 1981, pp. 565–609.
- [5] P. Comon, "Contrasts for multichannel blind deconvolution," *IEEE Signal Process. Lett.*, vol. 3, no. 7, pp. 209–211, Jul. 1996.
- [6] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 312–321, Mar. 1990.
- [7] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, pp. 1483–1492, 1997.
- [8] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
- [9] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Process.*, vol. 45, pp. 59–83, 1995.
- [10] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.
- [11] C. B. Papadias, "Globally convergent blind source separation based on a multiuser kurtosis maximization criterion," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3508–3519, Dec. 2000.
- [12] A. Hyvärinen and E. Bingham, "A fast fixed-point algorithm for independent component analysis of complex-valued signals," *Int. J. Neural Syst.*, vol. 10, no. 1, pp. 1–8, 2000.
- [13] P. A. Regalia and E. Kofidis, "Monotonic convergence of fixed-point algorithms for ICA," *IEEE Trans. Neural Netw.*, vol. 14, no. 4, pp. 943–949, Jul. 2003.
- [14] E. Oja, "Convergence of the symmetrical FastICA algorithm," in 9th Int. Conf. Neural Information Processing (ICONIP), Singapore, Nov. 18–22, 2002.
- [15] S. Douglas, "On the convergence behavior of the FastICA algorithm," in *Proc. 4th Symp. Independent Component Analysis Blind Source Separation*, Nara, Japan, Apr. 2003, pp. 409–414.
- [16] X. Giannakopoulos, J. Karhunen, and E. Oja, "Experimental comparison of neural algorithms for independent component analysis and blind separation," *Int. J. Neural Syst.*, vol. 9, pp. 651–656, 1999.
- [17] A. Hyvärinen, "One-unit contrast functions for independent component analysis: A statistical analysis," in *Proc. IEEE Neural Networks for Signal Processing (NNSP) Workshop*, Amelia Island, FL, 1997, pp. 388–397.
- [18] R. C. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley, 1973.
- [19] K. Waheed and F. M. Salam, "Blind source recovery using an adaptive generalized Gaussian score function," in *Proc. 45th Midwest Symp. Circuits Systems (MWSCAS)*, vol. 3, Aug. 4–7, 2002, pp. 656–659.
- [20] P. Tichavský and Z. Koldovský, "Optimal pairing of signal components separated by blind techniques," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 119–122, Feb. 2004.
- [21] Z. Koldovský, P. Tichavský, and E. Oja, "Cramér-Rao lower bound for linear independent component analysis," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. III, Philadelphia, PA, Mar. 2005, pp. 581–584.
- [22] E. G. Learned-Miller and J. W. Fisher III, "ICA using spacings estimates of entropy," *J. Mach. Learning Res.*, vol. 4, pp. 1271–1295, 2003.
- [23] R. Boscolo, H. Pan, and V. P. Roychowdhury, "Independent component analysis based on nonparametric density estimation," *IEEE Tran. Neural Netw.*, vol. 15, no. 1, pp. 55–65, Jan. 2004.
- [24] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proc. IEEE*, vol. 90, no. 8, pp. 2009–2026, Oct. 98.
- [25] —, "On the performance of orthogonal source separation algorithms," in *Proc. EUSIPCO*, Edinburgh, U.K., Sep. 1994, pp. 776–779.
- [26] D. T. Pham and P. Garat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1712–1725, Jul. 1997.
- [27] D. Yellin and B. Friedlander, "Multichannel system identification and deconvolution: performance bounds," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1410–1414, May 1999.
- [28] O. Shalvi and E. Weinstein, "Maximum likelihood and lower bounds in system identification with non-Gaussian inputs," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 328–339, Mar. 1994.
- [29] V. Vigneron and Ch. Jutten, "Fisher information in source separation problems," in *Proc. 5th Int. Conf. Independent Component Analysis (ICA 2004)*, C. G. Puntonet and A. Prieto, Eds., Granada, Spain, Sep. 2004, pp. 168–176.
- [30] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "QML, blind deconvolution: Asymptotic analysis," in *Proc. 5th Int. Conf. Independent Component Analysis (ICA 2004)*, C. G. Puntonet and A. Prieto, Eds., Granada, Spain, Sep. 2004, pp. 677–684.
- [31] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [32] Y. Lomnitz and A. Yeredor, "A blind-ML scheme for blind source separation," presented at the IEEE Workshop on Statistical Signal Processing (SSP'03), St. Louis, MO, Sep. 2003.
- [33] Z. Koldovský, P. Tichavský, and E. Oja, "Efficient variant of algorithm FastICA for independent component analysis attaining the Cramer-Rao lower bound," *IEEE Trans. Neural Netw.*, to be published.



Petr Tichavský (M'98–SM'04) graduated from the Czech Technical University, Prague, Czechoslovakia, in 1987 and received the C.Sc. (Ph.D.) degree in theoretical cybernetics from the Czechoslovak Academy of Sciences, Prague, in 1992.

Since that time, he has been with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague. He is author and coauthor of research papers in the area of sinusoidal frequency/frequency-rate estimation, adaptive filtering and tracking of time-varying signal parameters, algorithm-independent bounds on achievable performance, sensor-array processing, independent component analysis, and blind signal separation.

Dr. Tichavský received the Fulbright grant for a ten-month fellowship at the Department of Electrical Engineering, Yale University, New Haven, CT, in 1994 and the Otto Wichterle Award from the Academy of Sciences of the Czech Republic in 2002. He served as Associate Editor of the IEEE SIGNAL PROCESSING LETTERS from 2002 to 2004, and since 2005, has served as Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING.



Zbyněk Koldovský (S'03–M'04) was born in Jablonec nad Nisou, Czech Republic, in 1979. He received the M.S. and Ph.D. degrees from the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague, in 2002 and 2006, respectively.

In 2002, he joined the Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic, Prague. Since 2005, he has been with the Faculty of Mechatronic and Interdisciplinary studies, Technical University of Liberec, Czech Republic. His main research interests are in audio processing, independent component analysis, and blind deconvolution.



Erkki Oja (S'75–M'78–SM'90–F'00) received the Dr.Sc. degree from Helsinki University of Technology, Finland, in 1977.

He is Director of the Adaptive Informatics Research Centre and Professor of Computer Science at the Laboratory of Computer and Information Science, Helsinki University of Technology. He has been Research Associate at Brown University, Providence, RI, and visiting Professor at the Tokyo Institute of Technology, Tokyo, Japan. He is the author or coauthor of more than 280 articles and

book chapters on pattern recognition, computer vision, and neural computing, and three books: *Subspace Methods of Pattern Recognition* (New York: Research Studies Press and Wiley, 1983), which has been translated into Chinese and Japanese; *Kohonen Maps* (Amsterdam, The Netherlands: Elsevier, 1999); and *Independent Component Analysis* (New York: Wiley, 2001). His research interests are in the study of principal component and independent component analysis, self-organization, statistical pattern recognition, and applying artificial neural networks to computer vision and signal processing.

Prof. Oja is a member of the Finnish Academy of Sciences, Founding Fellow of the International Association of Pattern Recognition (IAPR), and Past President of the European Neural Network Society (ENNS). He is also a member of the editorial boards of several journals and has been in the program committees of several recent conferences, including the International Conference on Artificial Neural Networks (ICANN), the International Joint Conference on Neural Networks (IJCNN), and the International Conference on Neural Information Processing (ICONIP). He is the recipient of the 2006 IEEE Computational Intelligence Society Neural Networks Pioneer Award.

cardinal number of a set \mathcal{X} [9, p. 552]). This can be shown by using the fact that $b_k(t_n)$ and $b_{k'}(t_n)$ are independent of all other elements of \mathbf{B} due to Assumption 1. Hence, for every possible combination of $b_k(t_n)$ and $b_{k'}(t_n)$ it is possible to group terms corresponding to $b_{k,l}(t_n) = b_k(t_n)$ and $b_{k',l}(t_n) = b_{k'}(t_n)$ on the left-hand side of (39) into a sub sum which is equal to the marginal distribution of that combination. Finally, note that since $b_k(t_n), b_{k'}(t_n) \in \{0, 1\}$ and

$$\Pr\{b_k(t_n) = 1, b_{k'}(t_n) = 1\} = \begin{cases} (1 - p_k)(1 - p_{k'}), & \text{if } k \neq k' \\ (1 - p_k), & \text{if } k = k' \end{cases}$$

the proof follows.

REFERENCES

- [1] T. J. Abatzoglou, "Fast maximum likelihood joint estimation of frequency and frequency rate," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 6, pp. 708–715, 1986.
- [2] J. Angeby, "Estimating signal parameters using the nonlinear instantaneous least squares approach," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2721–2732, Oct. 2000.
- [3] S. Barbarossa, A. Scaglione, and G. B. Giannakis, "Product high-order ambiguity function for multicomponent polynomial phase signal modeling," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 691–708, Mar. 1998.
- [4] P. Bloomfield, "Spectral analysis with randomly missing observations," *J. Roy. Stat. Soc. B*, vol. 32, no. 3, pp. 369–380, 1970.
- [5] R. F. Brcich and A. M. Zoubir, "Parameter estimation of linear frequency modulated signals with missing observations," in *Proc. IEEE Int. Conf. Acous., Speech, Signal Process. (ICASSP)*, Apr. 2003, pp. 569–572.
- [6] J. A. Legg and D. A. Gray, "Performance bounds for polynomial phase parameter estimation with nonuniform and random sampling schemes," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 331–337, Feb. 2000.
- [7] M. Maksimov, *Radar Anti-Jamming Techniques*. Norwood, MA: Artech House, 1980.
- [8] S. Mirsaidi, G. A. Fleury, and J. Oksman, "LMS-Like AR modeling in the case of missing observations," *IEEE Trans. Signal Process.*, vol. 45, no. 6, pp. 1574–1583, Jun. 1997.
- [9] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*. New York: Springer, 2000.
- [10] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308–313, 1965.
- [11] S. Peleg and B. Porat, "The Cramér–Rao lower bound for signals with constant amplitude and polynomial phase," *IEEE Trans. Signal Process.*, vol. 39, no. 3, pp. 749–752, Mar. 1991.
- [12] D. S. Pham and A. M. Zoubir, "Analysis of multicomponent polynomial phase signals," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 56–65, Jan. 2007.
- [13] B. Porat, *Digital Processing of Random Signals: Theory & Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [14] B. Porat and B. Friedlander, "ARMA Spectral estimation of time series with missing observations," *IEEE Trans. Inf. Theory*, vol. 30, no. 6, pp. 823–831, Nov. 1984.
- [15] Y. Rosen and B. Porat, "The second-order moments of the sample covariances for time series with missing observations," *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 334–341, Mar. 1989.
- [16] R. Villarino and J. F. Böhme, "Pressure reconstruction and misfire detection from multichannel structure-borne sound," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2004, vol. 2, pp. 141–144.
- [17] M. Wagner, S. Carstens-Behrens, and J. F. Böhme, "In-cylinder pressure estimation using structural vibration measurements of spark ignition engines," in *Proc. IEEE Workshop Higher-Order Statistics*, 1999, pp. 174–177.
- [18] A. M. Zoubir and J. F. Böhme, "Bootstrap multiple tests applied to sensor location," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1386–1396, Jun. 1995.
- [19] A. M. Zoubir and D. R. Iskander, "Bootstrap modeling of a class of nonstationary signals," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 399–408, Feb. 2000.

Corrections to "Performance Analysis of the FastICA Algorithm and Cramér–Rao Bounds for Linear Independent Component Analysis"

Petr Tichavský, Zbyněk Koldovský, and Erkki Oja

Abstract—The derivation of the Cramér–Rao bound (CRB) in ["Performance Analysis of the FastICA Algorithm and Cramér–Rao Bounds for Linear Independent Component Analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 4, Apr. 2006, pp. 1189–1203] contains errors, which influence the matrix form of the CRB but not the CRB on variance of relevant off-diagonal elements of the demixing matrix. In this correspondence, we correct these errors.

I. THE FISHER INFORMATION MATRIX FOR ICA

The referenced paper considers a standard linear independent component analysis (ICA) model of a given $d \times N$ data matrix

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (1)$$

where \mathbf{A} is an unknown, nonsingular $d \times d$ mixing matrix. The joint probability density function (pdf) of the independent components is assumed to be $f_{\mathbf{S}}(\mathbf{S}) = \prod_{i=1}^d \prod_{j=1}^N f_i(s_{ij})$, where s_{ij} is the (i, j) th element of \mathbf{S} , $i = 1, \dots, d$, $j = 1, \dots, N$, and f_i is the pdf of s_{ij} .

The data matrix \mathbf{X} is obtained as a linear transformation of \mathbf{S} , $\mathbf{X} = \mathbf{W}^{-1}\mathbf{S}$, or equivalently, $\text{vec}[\mathbf{X}] = (\mathbf{I}_N \otimes \mathbf{W}^{-1})\text{vec}[\mathbf{S}]$, where $\mathbf{W} = \mathbf{A}^{-1}$, \mathbf{I}_N denotes the $N \times N$ identity matrix and \otimes is the Kronecker product. Therefore, the joint pdf of the data has the form

$$f_{\mathbf{X}|\theta}(\mathbf{X}|\theta) = |\det \mathbf{W}|^N f_{\mathbf{S}}(\mathbf{W}\mathbf{X}) \quad (2)$$

where θ is the unknown to-be-estimated vector parameter, $\theta = \text{vec}[\mathbf{W}]$. The error in [1] begins with the missing exponent N in the pdf expression above; cf. [1, eq. (32)].

A straightforward computation similar to that in Appendix C in [1] follows that (34) in [1] should be replaced with

$$\mathbf{F}_{mn} = N \left(a_{ju} a_{vi} + \delta_{iu} a_{ji} a_{vi} (\eta_i - 2) + \delta_{iu} \kappa_i \sum_{\ell=1, \ell \neq u}^d a_{j\ell} a_{v\ell} \right). \quad (3)$$

Recall for completeness that \mathbf{F}_{mn} is the mn th element of the $d^2 \times d^2$ Fisher information matrix \mathbf{F}_{θ} , where $m = (i-1)d + j$, $n = (u-1)d + v$, a_{ij} denotes the ij th element of the matrix \mathbf{A} , $\kappa_i \stackrel{\text{def}}{=} E[\psi_i^2(s_{ij})]$, $\eta_i \stackrel{\text{def}}{=} [s_i^2 \psi_i^2(s_{ij})]$, and $\psi_i \stackrel{\text{def}}{=} -f_i'/f_i$. A comparison of (3) with (34) in [1] shows that the correct Fisher information matrix element does not include any term proportional to $(N-1)^2$ but is proportional to N .

The derivation of (3) via Appendix C in [1] can be simplified by putting $N = 1$ and multiplying the resultant information matrix by N afterwards. The information matrix must be proportional to N , because

Manuscript received September 19, 2007; revised September 14, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jonathon Chambers. The work was supported by the Ministry of Education, Youth and Sports of the Czech Republic through the project 1 M0572.

P. Tichavský and Z. Koldovský are with the Institute of Information Theory and Automation, Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic (e-mail: tichavsk@utia.cas.cz).

E. Oja is with the Neural Networks Research Centre, Helsinki University of Technology, 02015 HUT, Finland.

Digital Object Identifier 10.1109/TSP.2007.910503

the observed data are composed of N independent observations of a random vector with the same distribution.

The computation proceeds by proving the formula

$$\mathbf{F}_\theta = (\mathbf{A}^T \otimes \mathbf{I}) \mathbf{F}_\mathbf{I} (\mathbf{A} \otimes \mathbf{I}) \quad (4)$$

where $\mathbf{F}_\mathbf{I}$ stands for the Fisher information matrix derived for a case when $\mathbf{A} = \mathbf{I}$ (identity matrix). Note the error in the matrix transpositions in (35) in [1] and also in (37). The latter equation should read $\mathbf{F}_\mathbf{G} = (\mathbf{W}^T \otimes \mathbf{I}) \mathbf{F}_\theta (\mathbf{W} \otimes \mathbf{I}) = \mathbf{F}_\mathbf{I}$. A similar typo exists in [2].

For the proof of (4), it was referred to (28) in [1], but it is not accurate. In fact, (4) was only inspired by this formula.

Substituting $a_{ij} = \delta_{ij}$ into (3), it easily follows that the mn th element of $\mathbf{F}_\mathbf{I}$ for $m = (i-1)d+j$ and $n = (u-1)d+v$ reads

$$(\mathbf{F}_\mathbf{I})_{mn} = N(\delta_{ju}\delta_{vi} + \delta_{ji}\delta_{vu}\delta_{vi}(\eta_i - \kappa_i - 2) + \delta_{iu}\delta_{vj}\kappa_i). \quad (5)$$

Again, there is no term proportional to $(N-1)^2$, unlike (36) in [1]. Thus, $\mathbf{F}_\mathbf{I}$ can be written as $\mathbf{F}_\mathbf{I} = N(\mathbf{P} + \mathbf{\Sigma})$, where \mathbf{P} is a permutation matrix and $\mathbf{\Sigma}$ is a diagonal matrix such that the mn th element of \mathbf{P} and $\mathbf{\Sigma}$ are $\delta_{ju}\delta_{vi}$, and $\delta_{iu}\delta_{vj}[\kappa_i + \delta_{ij}(\eta_i - \kappa_i - 2)]$, respectively, for $m = (i-1)d+j$ and $n = (u-1)d+v$.

To prove (4) rigorously, note that the mn th elements of $\mathbf{A} \otimes \mathbf{I}$ and $\mathbf{A}^T \otimes \mathbf{I}$ for $m = (i-1)d+j$ and $n = (u-1)d+v$ are equal to $(\mathbf{A} \otimes \mathbf{I})_{mn} = a_{iu}\delta_{jv}$ and $(\mathbf{A}^T \otimes \mathbf{I})_{mn} = a_{ui}\delta_{jv}$, respectively. Then, the mn th element of the product $(\mathbf{A}^T \otimes \mathbf{I}) \mathbf{F}_\mathbf{I} (\mathbf{A} \otimes \mathbf{I})$ is

$$[(\mathbf{A}^T \otimes \mathbf{I}) \mathbf{F}_\mathbf{I} (\mathbf{A} \otimes \mathbf{I})]_{mn} = \sum_{n', n''} (\mathbf{A}^T \otimes \mathbf{I})_{mn'} (\mathbf{F}_\mathbf{I})_{n'n''} (\mathbf{A} \otimes \mathbf{I})_{n''n}. \quad (6)$$

A straightforward computation gives that the matrix element in (6) is identical to that in (3).

Appendix D in [1] should be changed accordingly. Application of the matrix inversion lemma is not needed, and the rest of the Appendix and the final result in (38) are correct. Note that an alternative elegant method of inversion of a matrix similar to $\mathbf{F}_\mathbf{I}$ was used in [2].

Finally, note that one of assumptions of the Cramér–Rao inequality is that the support of the pdf $f_{\mathbf{x}|\theta}(\mathbf{X}|\theta)$ is independent of the estimated parameter θ . In the ICA scenario, this assumption is equivalent to the condition that $f_i(x) > 0$ for all i and finite x . In particular, the CRB is not defined for sources with a bounded (limited) support.

ACKNOWLEDGMENT

The authors would like to thank Dr. A. Yeredor for spotting the error in (34) and to an anonymous reviewer of these corrections for finding the error in (35) in [1].

REFERENCES

- [1] P. Tichavský, Z. Koldovský, and E. Oja, "Performance analysis of the FastICA algorithm and Cramér–Rao bounds for linear independent component analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1189–1203, Apr. 2006.
- [2] E. Doron, A. Yeredor, and P. Tichavský, "Cramér–Rao-induced bound for blind separation of stationary parametric Gaussian sources," *IEEE Signal Process. Letters.*, vol. 14, no. 6, pp. 417–420, Jun. 2007.

Hilbert Pair of Orthogonal Wavelet Bases: Revisiting the Condition

David B. H. Tay, *Member, IEEE*

Abstract—It is now well known that in order to have wavelet bases that form a Hilbert Transform pair, the corresponding low-pass conjugate quadrature filters (CQF) should ideally be related through a half sampled delay, i.e., $e^{-j\omega/2}$. In this correspondence we revisit this condition and examine some subtleties associated with this condition that were overlooked in previous work. We will show that there is a more general condition where the delay can be any "even-half" samples, i.e., $e^{-j(2N+1/2)\omega}$. More importantly we examine the implications in formulating design strategies for Hilbert pairs and its implementation.

Index Terms—Complex wavelet, dual-tree, Hilbert pair, orthonormal filter banks.

I. INTRODUCTION AND PRELIMINARIES

Overcomplete complex (valued) transforms that are based on the Hilbert pairs are becoming an increasingly important signal processing tool [1]. These complex transforms have the advantage of approximate shift-invariance over the critically sampled real (valued) wavelet transforms.

Orthogonal wavelets are usually associated or obtained from a low-pass conjugate quadrature filter (CQF) $H(z)$. A CQF satisfies $H(z)H(z^{-1}) + H(-z)H(-z^{-1}) = 1$. In the filter bank, the constituent filters, denoted by $H_0(z)$ (low-pass analysis), $H_1(z)$ (high-pass analysis), $F_0(z)$ (low-pass synthesis), and $F_1(z)$ (high-pass synthesis), are usually obtained from a CQF filter $H(z)$ as follows:

$$\begin{aligned} H_0(z) &= H(z) & H_1(z) &= z^{-1}H(-z^{-1}) \\ F_0(z) &= H(z^{-1}) & F_1(z) &= zH(-z). \end{aligned} \quad (1)$$

With (1), it can be verified that the aliasing function $A(z) \equiv H_0(-z)F_0(z) + H_1(-z)F_1(z) = 0$ and the reconstruction function $T(z) \equiv H_0(z)F_0(z) + H_1(z)F_1(z) = 1$, perfect reconstruction with zero delay.

In a Hilbert pair, the two wavelets corresponding to two CQFs are related through the Hilbert transform:

$$\Psi^g(\omega) = \begin{cases} -j\Psi^h(\omega) & \text{for } \omega > 0 \\ j\Psi^h(\omega) & \text{for } \omega < 0 \end{cases} \quad (2)$$

where $\Psi^h(\omega)$ and $\Psi^g(\omega)$ are the Fourier transforms of $\psi^h(t)$ and $\psi^g(t)$, respectively. By denoting the corresponding CQFs by $H^h(z)$ and $H^g(z)$ respectively, it was first shown in [2] that (2) is achieved if

$$H^g(e^{j\omega}) = e^{-j\omega/2}H^h(e^{j\omega}), \quad |\omega| < \pi \quad (3)$$

and is known as the *half sample delay condition*. Further analysis on the condition were presented in [3] and [4] using alternative formulations which are easier to manipulate analytically. The conclusion drawn in [3] and [4] are similar to that in [2], namely the half sample delay condition is required. The most general analysis appeared in [4] where no assumption on the relationship between the two CQFs were made and (3) is shown to be necessary and sufficient.

Manuscript received August 23, 2006; revised August 13, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Antonia Papandreou-Suppappola.

The author is with the Department of Electronic Engineering, LaTrobe University, Bundoora, Victoria 3086, Australia (e-mail: D.Tay@latrobe.edu.au).
Digital Object Identifier 10.1109/TSP.2007.909221

Cramér-Rao-Induced Bounds for CANDECOMP/PARAFAC Tensor Decomposition

Petr Tichavský, *Senior Member, IEEE*, Anh Huy Phan, and Zbyněk Koldovský, *Member, IEEE*

Abstract—This paper presents a Cramér-Rao lower bound (CRLB) on the variance of unbiased estimates of factor matrices in Canonical Polyadic (CP) or CANDECOMP/PARAFAC (CP) decompositions of a tensor from noisy observations, (i.e., the tensor plus a random Gaussian i.i.d. tensor). A novel expression is derived for a bound on the mean square angular error of factors along a selected dimension of a tensor of an arbitrary dimension. The expression needs less operations for computing the bound, $O(NR^6)$, than the best existing state-of-the-art algorithm, $O(N^3R^6)$ operations, where N and R are the tensor order and the tensor rank. Insightful expressions are derived for tensors of rank 1 and rank 2 of arbitrary dimension and for tensors of arbitrary dimension and rank, where two factor matrices have orthogonal columns.

The results can be used as a gauge of performance of different approximate CP decomposition algorithms, prediction of their accuracy, and for checking stability of a given decomposition of a tensor (condition whether the CRLB is finite or not). A novel expression is derived for a Hessian matrix needed in popular damped Gauss-Newton method for solving the CP decomposition of tensors with missing elements. Beside computing the CRLB for these tensors the expression may serve for design of damped Gauss-Newton algorithm for the decomposition.

Index Terms—Canonical polyadic decomposition, Cramér-Rao lower bound, multilinear models, stability, uniqueness.

I. INTRODUCTION

ORDER-3 and higher-order data arrays need to be analyzed in diverse research areas such as chemistry, astronomy, and psychology [1]–[3]. The analyses can be done through finding multi-linear dependencies among elements within the arrays. The most popular model is Parallel factor analysis (PARAFAC), also called Canonical decomposition (CANDECOMP) or CP, which is an extension of a low rank decomposition of matrices to higher-way arrays, usually called

tensors. In signal processing, the tensor decompositions have become popular for their usefulness in blind source separation [4].

Note that a best-fitting CP decomposition may not exist for some tensors. In that case, trying to find a best-fitting CP decomposition results in diverging factors [5], [6]. This paper is focussed on studying CP decompositions of a noisy observations of tensors, which admit an exact CP decomposition. The decomposition of the noiseless tensor is taken as a *ground truth* for computing errors.

An important issue is the essential uniqueness of CP decomposition as it entails identifiability of the model (the factor matrices) from the tensor. The adjective “essential” means that the model is unique up to a scale and permutation ambiguity, which is inherent to the problem. Initial works in the field can be traced back in 70’s in works of Harshman [7], [8]. A popular sufficient condition for the uniqueness was derived by Kruskal in [9]. Recently, the problem has been addressed again, namely by Stegeman, Ten Berge, De Lathauwer, Jiang, Sidiropoulos *et al.*; see [10]–[24].

This paper is focussed on stability of the CP decomposition rather than on the uniqueness. By stability we mean existence of a finite Cramér-Rao bound in a stochastic set-up, where tensor elements are corrupted by additive Gaussian-distributed noise. Relation of this kind of stability to a deterministic stability and to the uniqueness was studied in [25]. It is not true, in general, that stability of a solution of a nonlinear problem implies uniqueness of the solution. For example, there might always be a permutation or sign ambiguity. It is yet an open theoretical question if stability of the CP tensor decomposition problem implies its *essential* uniqueness. Regardless of the missing link to identifiability, the stability is an interesting concept which is worth to be studied, because different kind of noise is very common.

In general, in order to evaluate performance of a tensor decomposition, the approximation error between the data tensor and its approximate is sometimes used. Unfortunately, such measure does not imply quality of the estimated components. In practice, in some difficult scenarios such as decomposition of tensor with linear dependency among components of factor matrices, or large difference in magnitude between components [26], [27], most CP algorithms explained the data tensor at almost identical fit, but only few algorithms can accurately retrieve the hidden components from the tensor [26], [28]. In order to verify theoretically the quality of the estimated components and evaluate robustness of an algorithm, an appropriate measure is an essential prerequisite. The squared angular error between the estimated component and its original one is such a measure [29], [30]. Working with angular errors is practical,

Manuscript received August 01, 2012; revised November 19, 2012 and January 18, 2013; accepted January 26, 2013. Date of publication February 07, 2013; date of current version March 22, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Nicholas Sidiropoulos. This work was supported by the Grant Agency of the Czech Republic through projects 102/09/1278 and P103/11/1947.

P. Tichavský is with the Institute of Information Theory and Automation, Pod vodárenskou věží 4, 182 08 Prague 8, Czech Republic (e-mail: tichavsk@utia.cas.cz).

A. H. Phan is with the Brain Science Institute, RIKEN, Wakoshi, Japan (e-mail: phan@brain.riken.jp).

Z. Koldovský is with the Institute of Information Theory and Automation, Pod vodárenskou věží 4, 182 08 Prague 8, Czech Republic. He is also with the Faculty of Mechatronic and Interdisciplinary Studies, Technical University of Liberec, 461 17 Liberec, Czech Republic (e-mail: zbynek.koldovsky@tul.cz).

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.2013.2245660

because the scaling ambiguity does not play a role. Only the permutation ambiguity has to be solved in practical examples, because order of the factor can be quite arbitrary.

Cramér-Rao lower bound for CP decomposition was first studied in [31], and later, a more compact asymptotic expression was derived in [32] for tensors of order 3 appearing in wireless communications. A non-asymptotic (exact) CRLB-induced bound (CRIB) on squared angular deviation of columns of the factor matrices with respect to their nominal values has been studied in [29]. Similar results for symmetric tensors are derived in [33]. Nevertheless, the study is limited to the case of three-way tensors. In the general case, CRIB can be, indeed, calculated through the approximate Hessian which is often huge, and is impractical to directly invert. Note that such task normally costs $O(R^3 T^3)$ where $T = \sum_n I_n$. Seeking a cheaper method for CRIB is a challenge to make it applicable.

This paper presents new CRIB expressions for tensors of arbitrary dimension and rank, and specialized expressions for rank 1 and rank 2 tensors. The results rely on compact expressions for Hessian of the problem derived in [28]. Alternative expressions for the Hessian exist in [39]. Note, however, that unlike [28], this paper presents different expressions for inverse of the Hessian, which have lower computational complexity. In particular, complexity of inversion of the Hessian is reduced from $O(N^3 R^6)$ operations to $O(NR^6)$, where N and R are the tensor order and the tensor rank, respectively.

On basis of new discovered properties of the CRIB, we established connection between theoretical and practical results in CP decomposition (CPD):

- Stability of CPD for rank-1 and rank-2 tensors of arbitrary dimension.
- The work may serve as theoretical support for a novel CP decomposition algorithm through tensor reshaping [34], which was designed to decompose high-dimensional and high-order tensors. In particular, it appears that higher-order orthogonally constrained CPD [35]–[38] can be decomposed efficiently through tensor unfolding.
- Stability when factor matrices occur linear dependence problem and especially the rank-overlap problem [1], [23], [36]. The problem is related to a variant of CPD for linear dependent loadings which was investigated in chemometric data and in flow injection analysis [1], [36]. A partial uniqueness condition of the related model is discussed in [23].
- CP decomposition of tensors with missing entries, which is quite frequent in practice, is addressed. An approximate Hessian for this case is derived, which is the core for the damped Gauss-Newton algorithm for the decomposition.
- A maximum tensor rank, given dimension of the tensor, which admits a stable decomposition is discussed.

The paper is organized as follows. Section II presents the main result, the Cramér-Rao induced bound on angular error of one factor vector in full generality. In Section III, this result is specialized for tensors of rank 1 and rank 2, and for the case when two factor matrices have mutually orthogonal columns. Section IV is devoted to a possible application of the bound: investigation of loss of accuracy of the tensor decomposition

when the tensor is reshaped to a lower-dimensional form. Section V deals with the bound for tensors with missing entries, Section VI contains examples—CRIB computed for CP decomposition of a fluorescence tensor, stability of the tensor investigated by Brie *et al.*, and a discussion of a maximum stable rank given the tensor dimension. Section VII concludes the paper.

II. PRESENTATION OF THE CRIB

A. Cramér-Rao Bound for CP Decomposition

Let \mathcal{Y} be an N -way tensor of dimension $I_1 \times I_2 \times \dots \times I_N$. The tensor is said to be of rank R , if R is the smallest number of rank-one tensors which admit the decomposition of \mathcal{Y} of the form

$$\mathcal{Y} = \sum_{r=1}^R \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \dots \circ \mathbf{a}_r^{(N)} \quad (1)$$

where \circ denotes the outer vector product, $\mathbf{a}_r^{(n)}$, $r = 1, \dots, R$, $n = 1, \dots, N$ are vectors of the length I_n called factors. The tensor in (1) can be characterized by N factor matrices $\mathbf{A}_n = [\mathbf{a}_1^{(n)}, \mathbf{a}_2^{(n)}, \dots, \mathbf{a}_R^{(n)}]$ of the size $I_n \times R$ for $n = 1, \dots, N$. Sometimes (1) is referred to as a Kruskal form of a tensor [45].

In practice, CP decomposition of a given rank (R) is used as an approximation of a given tensor, which can be a noisy observation $\hat{\mathcal{Y}}$ of the tensor \mathcal{Y} in (1). Owing to the symmetry of (1), we can focus on estimating the first factor matrix \mathbf{A}_1 , without any loss of generality, and we can assume that all other factor matrices have columns of unit norm. Then the “energy” of the parallel factors is determined by the squared Euclidean norm of columns of \mathbf{A}_1 .

It is common to assume that the noise has a zero mean Gaussian distribution with variance σ^2 , and is independently added to each element of the tensor in (1).

Let a vector parameter $\boldsymbol{\theta}$ containing all parameters of our model be arranged as

$$\boldsymbol{\theta} = [(\text{vec} \mathbf{A}_1)^T, \dots, (\text{vec} \mathbf{A}_N)^T]^T. \quad (2)$$

The maximum likelihood solution for $\boldsymbol{\theta}$ consists in minimizing the least squares criterion

$$\mathcal{Q}(\boldsymbol{\theta}) = \left\| \hat{\mathcal{Y}} - \mathcal{Y}(\boldsymbol{\theta}) \right\|_F^2 \quad (3)$$

where $\| \cdot \|_F$ stands for the Frobenius norm.

We wish to compute the Cramér-Rao lower bound for estimating $\boldsymbol{\theta}$. In general, for this estimation problem, the CRLB is given as the inverse of the Fisher information matrix, which is equal to [29]

$$\mathbf{F}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \mathbf{J}^T(\boldsymbol{\theta}) \mathbf{J}(\boldsymbol{\theta}) \quad (4)$$

where $\mathbf{J}(\boldsymbol{\theta})$ is the Jacobi matrix (matrix of the first-order derivatives) of $\mathcal{Q}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. In other words, the Fisher information matrix is proportional to the approximate Hessian matrix of the criterion, $\mathbf{H}(\boldsymbol{\theta}) = \mathbf{J}^T(\boldsymbol{\theta}) \mathbf{J}(\boldsymbol{\theta})$ [39].

Let $\mathbf{\Gamma}_{nm}$ denote the Hadamard (elementwise) product of matrices $\mathbf{C}_k = \mathbf{A}_k^T \mathbf{A}_k$, $k \in \{1, \dots, N\} - \{n, m\}$,

$$\mathbf{\Gamma}_{nm} = \bigotimes_{k \neq n, m} \mathbf{C}_k, \quad \mathbf{C}_k = \mathbf{A}_k^T \mathbf{A}_k. \quad (5)$$

Theorem 1 [28]: The Hessian \mathbf{H} can be decomposed into low rank matrices under the form as

$$\mathbf{H} = \mathbf{G} + \mathbf{Z}\mathbf{K}\mathbf{Z}^T \quad (6)$$

where $\mathbf{K} = [\mathbf{K}_{nm}]_{n,m=1}^N$ contains submatrices \mathbf{K}_{nm} given by

$$\mathbf{K}_{nm} = (1 - \delta_{nm}) \mathbf{P}_R \text{dvec}(\mathbf{\Gamma}_{nm}) \quad (7)$$

\mathbf{P}_R is the permutation matrix of dimension $R^2 \times R^2$ defined in [28] such that $\text{vec}\mathbf{M} = \mathbf{P}_R \text{vec}(\mathbf{M}^T)$ for any $R \times R$ matrix \mathbf{M} , and δ_{nm} is the Kronecker delta, and $\text{dvec}(\mathbf{M})$ is a short-hand notation for $\text{diag}(\text{vec}(\mathbf{M}))$, i.e. a diagonal matrix containing all elements of a matrix \mathbf{M} on its main diagonal. Next,

$$\mathbf{G} = \text{bdiag}(\mathbf{\Gamma}_{nn} \otimes \mathbf{I}_{I_n})_{n=1}^N \quad (8)$$

and

$$\mathbf{Z} = \text{bdiag}(\mathbf{I}_R \otimes \mathbf{A}_n)_{n=1}^N \quad (9)$$

where \otimes denotes the Kronecker product, \mathbf{I}_{I_n} is an identity matrix of the size $I_n \times I_n$, and $\text{bdiag}(\cdot)$ is a block diagonal matrix with the given blocks on its diagonal. Note that the Hessian \mathbf{H} in (6) is rank deficient because of the scale ambiguity of columns of factor matrices [27], [41]. It has dimension $(R \sum_n I_n) \times (R \sum_n I_n)$ but its rank is at most $R \sum_n I_n - (N - 1)R$.

A regular (reduced) Hessian can be obtained from \mathbf{H} by deleting $(N - 1)R$ rows and corresponding columns in \mathbf{H} , because the estimation of one element in the vectors $\mathbf{a}_r^{(n)}$, $r = 1, \dots, R$, $n = 2, \dots, N$ can be skipped. The reduced Hessian may have the form

$$\mathbf{H}_E = \mathbf{E}\mathbf{H}\mathbf{E}^T \quad (10)$$

where

$$\mathbf{E} = \text{bdiag}(\mathbf{I}_{RI_1}, \mathbf{I}_R \otimes \mathbf{E}_2, \dots, \mathbf{I}_R \otimes \mathbf{E}_N) \quad (11)$$

and \mathbf{E}_n is an $(I_n - 1) \times I_n$ matrix of rank $I_n - 1$. For example, one can put $\mathbf{E}_n = [\mathbf{0}_{(I_n-1) \times 1} \mathbf{I}_{I_n-1}]$ for $n = 2, \dots, N$. With this definition of \mathbf{E}_n , \mathbf{H}_E is a Hessian for estimating the first factor matrix \mathbf{A}_1 and all other vectors $\mathbf{a}_r^{(n)}$, $r = 1, \dots, R$, $n = 2, \dots, N$ without their first elements. In the sequel, however, we use a different definition of \mathbf{E}_n . Note that each \mathbf{E}_n can be quite arbitrary, together facilitate a regular transformation of nuisance parameters, which does not influence CRLB of the parameter of interest.

The CRLB for the first column of \mathbf{A}_1 , denoted simply as \mathbf{a}_1 , is defined as σ^2 times the left-upper submatrix of \mathbf{H}_E^{-1} of the size $I_1 \times I_1$,

$$\text{CRLB}(\mathbf{a}_1) = \sigma^2 [\mathbf{H}_E^{-1}]_{1:I_1, 1:I_1}. \quad (12)$$

Substituting (6) in (10) gives

$$\mathbf{H}_E = \mathbf{G}_E + \mathbf{Z}_E \mathbf{K} \mathbf{Z}_E^T \quad (13)$$

where $\mathbf{G}_E = \mathbf{E}\mathbf{G}\mathbf{E}^T$ and $\mathbf{Z}_E = \mathbf{E}\mathbf{Z}$. Inverse of \mathbf{H}_E can be written using a Woodbury matrix identity [40] as

$$\mathbf{H}_E^{-1} = \mathbf{G}_E^{-1} - \mathbf{G}_E^{-1} \mathbf{Z}_E \mathbf{K} (\mathbf{I}_{NR^2} + \mathbf{Z}_E^T \mathbf{G}_E^{-1} \mathbf{Z}_E \mathbf{K})^{-1} \mathbf{Z}_E^T \mathbf{G}_E^{-1} \quad (14)$$

provided that the involved inverses exist.

Next,

$$\mathbf{G}_E = \text{bdiag}(\mathbf{\Gamma}_{11} \otimes \mathbf{I}_1, \mathbf{\Gamma}_{22} \otimes (\mathbf{E}_2 \mathbf{E}_2^T), \dots, \mathbf{\Gamma}_{NN} \otimes (\mathbf{E}_N \mathbf{E}_N^T)) \quad (15)$$

$$\mathbf{G}_E^{-1} = \text{bdiag}((\mathbf{\Gamma}_{11})^{-1} \otimes \mathbf{I}_1, \mathbf{\Gamma}_{22}^{-1} \otimes (\mathbf{E}_2 \mathbf{E}_2^T)^{-1}, \dots, \mathbf{\Gamma}_{NN}^{-1} \otimes (\mathbf{E}_N \mathbf{E}_N^T)^{-1}). \quad (16)$$

Put

$$\mathbf{\Psi} = \mathbf{Z}_E^T \mathbf{G}_E^{-1} \mathbf{Z}_E \quad (17)$$

$$\mathbf{B} = \mathbf{K} (\mathbf{I}_{NR^2} + \mathbf{\Psi} \mathbf{K})^{-1} \quad (18)$$

and let \mathbf{B}_0 be the upper-left $R^2 \times R^2$ submatrix of \mathbf{B} , symbolically $\mathbf{B}_0 = \mathbf{B}_{1:R^2, 1:R^2}$. Finally, let g_{11} and $\mathbf{g}_{1\cdot}$ be the upper-left element and the first row of $\mathbf{\Gamma}_{11}^{-1}$, respectively. Then

$$\begin{aligned} [\mathbf{H}_E^{-1}]_{1:I_1, 1:I_1} &= [\mathbf{G}_E^{-1}]_{1:I_1, 1:I_1} \\ &\quad + [\mathbf{G}_E^{-1} \mathbf{Z}_E]_{1:I_1, 1:R^2} \mathbf{B}_0 [\mathbf{G}_E^{-1} \mathbf{Z}_E]_{1:I_1, 1:R^2}^T \\ &= g_{11} \mathbf{I}_{I_1} + (\mathbf{g}_{1\cdot} \otimes \mathbf{A}_1) \mathbf{B}_0 (\mathbf{g}_{1\cdot} \otimes \mathbf{A}_1)^T. \end{aligned} \quad (19)$$

The CRLB represents a lower bound on the error covariance matrix $\text{E}[(\hat{\mathbf{a}}_1 - \mathbf{a}_1)(\hat{\mathbf{a}}_1 - \mathbf{a}_1)^T]$ for any unbiased estimator of \mathbf{a}_1 . The bound is asymptotically tight in the case of Gaussian noise and least squares estimator, which is equivalent to maximum likelihood estimator, under the assumptions that the permutation ambiguity has been solved out (order of the estimated factors was selected to match the original factors) and scaling of the estimator is in accord with the selection of the matrix \mathbf{E} .

B. Cramér-Rao-Induced Bound for Angular Error

CRLB(\mathbf{a}_1) considered in the previous subsection is a matrix. In applications it is practical to characterize the error of the factor \mathbf{a}_1 in the decomposition by a scalar quantity. In [30] it was proposed to characterize the error by an angle between the true and the estimated vector, and compute a Cramér-Rao-induced bound (CRIB) for the squared angle. The CRIB may serve a gauge of achievable accuracy of estimation/CP decomposition. Again, it is an asymptotically (in the sense of variance of the noise going to zero) tight bound on the angular error between an estimated and true factor.

The angle α_1 between the true factor \mathbf{a}_1 and its estimate $\hat{\mathbf{a}}_1$ obtained through the CP decomposition is defined through its cosine

$$\cos \alpha_1 = \frac{\mathbf{a}_1^T \hat{\mathbf{a}}_1}{\|\mathbf{a}_1\| \|\hat{\mathbf{a}}_1\|}. \quad (20)$$

The Cramér-Rao induced bound for the squared angular error α_1^2 [radians²] will be denoted $\text{CRIB}(\mathbf{a}_1)$ in the sequel. $\text{CRIB}(\mathbf{a}_1)$ in decibels (dB) is then defined as $-10 \log_{10}[\text{CRIB}(\mathbf{a}_1)]$ [dB].

Before computing $\text{CRIB}(\mathbf{a}_1)$ we present another interpretation of this quantity. Let the estimate $\hat{\mathbf{a}}_1$ be decomposed into a sum of a scalar multiple of \mathbf{a}_1 and a reminder, which is orthogonal to \mathbf{a}_1 ,

$$\hat{\mathbf{a}}_1 = \beta \mathbf{a}_1 + \mathbf{r}_1 \quad (21)$$

where $\beta = \mathbf{a}_1^T \hat{\mathbf{a}}_1 / \|\mathbf{a}_1\|^2$ and $\mathbf{r}_1 = \hat{\mathbf{a}}_1 - \beta \mathbf{a}_1$. Then, the Distortion-to-Signal Ratio (DSR) of the estimate $\hat{\mathbf{a}}_1$ can be defined as

$$\text{DSR}(\hat{\mathbf{a}}_1) = \frac{\|\mathbf{r}_1\|^2}{\beta^2 \|\mathbf{a}_1\|^2}. \quad (22)$$

A straightforward computation gives

$$\text{DSR}(\hat{\mathbf{a}}_1) = \frac{1 - \cos^2 \alpha_1}{\cos^2 \alpha_1} \approx \alpha_1^2. \quad (23)$$

The approximation in (23) is valid for small α_1^2 . We can see that $\text{CRIB}(\mathbf{a}_1)$ serves not only as a bound on the mean squared angular estimation error, but also as a bound on the achievable Distortion-to-Signal Ratio.

Theorem 2 [30]: Let $\text{CRLB}(\mathbf{a}_1)$ be the Cramér-Rao bound on covariance matrix of unbiased estimators of \mathbf{a}_1 . Then the Cramér-Rao-induced bound on the squared angular error between the true and estimated vector is

$$\text{CRIB}(\mathbf{a}_1) = \frac{\text{tr} \left[\Pi_{\mathbf{a}_1} \text{CRLB}^\perp(\mathbf{a}_1) \right]}{\|\mathbf{a}_1\|^2} \quad (24)$$

where

$$\Pi_{\mathbf{a}_1}^\perp = \mathbf{I}_{I_1} - \mathbf{a}_1 \mathbf{a}_1^T / \|\mathbf{a}_1\|^2 \quad (25)$$

is the projection operator to the orthogonal complement of \mathbf{a}_1 and $\text{tr}[\cdot]$ denotes trace of a matrix.

Proof: A sketch of a proof can be found in [30]. It is based on analysis of a mean square angular error of a maximum likelihood estimator, which is known to be asymptotically tight (achieving the Cramér-Rao bound). Note that a conceptually more straightforward but longer proof would be obtained through the formula for CRLB on a transformed parameter, see e.g., Theorem 3.4 in [44]. In particular,

$$\text{CRIB}(\mathbf{a}_1) = \mathbf{G}_a(\mathbf{a}_1) \text{CRLB}(\mathbf{a}_1) \mathbf{G}_a^T(\mathbf{a}_1) \quad (26)$$

where $\mathbf{G}_a(\hat{\mathbf{a}}_1)$ is the Jacobi matrix of the mapping representing the angular error as a function of the estimate $\hat{\mathbf{a}}_1$.

Theorem 3: The $\text{CRIB}(\mathbf{a}_1)$ can be written in the form

$$\text{CRIB}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \left\{ (I_1 - 1)g_{11} - \text{tr} \left[\mathbf{B}_0 \left((\mathbf{g}_{1,:}^T \mathbf{g}_{1,:}) \otimes \mathbf{X}_1 \right) \right] \right\} \quad (27)$$

where \mathbf{B}_0 is the submatrix of \mathbf{B} in (18), $\mathbf{B}_0 = \mathbf{B}_{1:R^2, 1:R^2}$,

$$\mathbf{X}_n = \mathbf{C}_n - \frac{1}{\mathbf{C}_{11}^{(n)}} \mathbf{C}_{:,1}^{(n)} \mathbf{C}_{:,1}^{(n)T} \quad (28)$$

for $n = 1, \dots, N$, $\mathbf{C}_{11}^{(n)}$ and $\mathbf{C}_{:,1}^{(n)}$ denote the upper-right element and the first column of \mathbf{C}_n , respectively, and Ψ in the definition of \mathbf{B} (18) takes, for a special choice of matrices \mathbf{E}_n , the form

$$\Psi = \text{bdiag} \left(\Gamma_{11}^{-1} \otimes \mathbf{C}_1, \Gamma_{22}^{-1} \otimes \mathbf{X}_2, \dots, \Gamma_{NN}^{-1} \otimes \mathbf{X}_N \right). \quad (29)$$

Proof: Substituting (12) and (19) into (24) gives, after some simplifications,

$$\begin{aligned} \text{CRIB}(\mathbf{a}_1) &= \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \text{tr} \left[\Pi_{\mathbf{a}_1}^\perp \left(g_{11} \mathbf{I}_{I_1} - (\mathbf{g}_{1,:} \otimes \mathbf{A}_1) \mathbf{B}_0 (\mathbf{g}_{1,:} \otimes \mathbf{A}_1)^T \right) \right] \\ &= \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \left\{ (I_1 - 1)g_{11} \right. \\ &\quad \left. - \text{tr} \left[\Pi_{\mathbf{a}_1}^\perp \left(\mathbf{g}_{1,:} \otimes \mathbf{A}_1 \right) \mathbf{B}_0 \left(\mathbf{g}_{1,:} \otimes \mathbf{A}_1 \right)^T \right] \right\} \\ &= \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \left\{ (I_1 - 1)g_{11} \right. \\ &\quad \left. - \text{tr} \left[\mathbf{B}_0 \left((\mathbf{g}_{1,:}^T \mathbf{g}_{1,:}) \otimes (\mathbf{A}_1^T \Pi_{\mathbf{a}_1}^\perp \mathbf{A}_1) \right) \right] \right\}. \quad (30) \end{aligned}$$

This is (27), because

$$\mathbf{A}_1^T \Pi_{\mathbf{a}_1}^\perp \mathbf{A}_1 = \mathbf{C}_1 - \frac{1}{\mathbf{C}_{11}^{(1)}} \mathbf{C}_{:,1}^{(1)} \mathbf{C}_{:,1}^{(1)T} = \mathbf{X}_1. \quad (31)$$

Next, assume that \mathbf{E} is defined as in (11), but \mathbf{E}_n are arbitrary full rank matrices of the dimension $(I_n - 1) \times I_n$. Then, combining (17), (9), (11) and (16) gives

$$\begin{aligned} \Psi &= \mathbf{Z}_E^T \mathbf{G}_E^{-1} \mathbf{Z}_E \\ &= \text{bdiag} \left(\Gamma_{11}^{-1} \otimes \mathbf{C}_1, \Gamma_{22}^{-1} \otimes \tilde{\mathbf{X}}_2, \dots, \Gamma_{NN}^{-1} \otimes \tilde{\mathbf{X}}_N \right) \quad (32) \end{aligned}$$

where

$$\tilde{\mathbf{X}}_n = \mathbf{A}_n^T \mathbf{E}_n^T (\mathbf{E}_n \mathbf{E}_n^T)^{-1} \mathbf{E}_n \mathbf{A}_n \quad (33)$$

for $n = 2, \dots, N$. Note that the expression $\mathbf{E}_n^T (\mathbf{E}_n \mathbf{E}_n^T)^{-1} \mathbf{E}_n$ is an orthogonal projection operator to the columnspace of \mathbf{E}_n^T . If \mathbf{E}_n is chosen as the first $(I_n - 1)$ rows of

$$\Pi_{\mathbf{a}_1^{(n)}}^\perp = \mathbf{I}_{I_n} - \mathbf{a}_1^{(n)} \mathbf{a}_1^{(n)T} / \|\mathbf{a}_1^{(n)}\|^2 \quad (34)$$

then $\mathbf{E}_n^T (\mathbf{E}_n \mathbf{E}_n^T)^{-1} \mathbf{E}_n = \Pi_{\mathbf{a}_1^{(n)}}^\perp$ and consequently $\tilde{\mathbf{X}}_n = \mathbf{A}_n^T \Pi_{\mathbf{a}_1^{(n)}}^\perp \mathbf{A}_n = \mathbf{X}_n$. \blacksquare

Note that the first row and the first column of \mathbf{X}_n are zero.

Theorem 4: Assume that all elements of the matrices \mathbf{C}_n in (5) are nonzero. Then, the matrix \mathbf{B}_0 in Theorem 3 can be written in the form

$$\mathbf{B}_0 = [-\mathbf{I}_{R^2} + \mathbf{V}(\mathbf{I}_{R^2} + \mathbf{V})^{-1}] \mathbf{Y} \quad (35)$$

where

$$\mathbf{V} = \mathbf{W} - \mathbf{Y} (\Gamma_{11}^{-1} \otimes \mathbf{C}_1) \quad (36)$$

$$\mathbf{W} = \mathbf{P}_R \sum_{n=2}^N \text{dvec}(\Gamma_{1n}) \mathbf{S}_n^{-1} (\Gamma_{nn}^{-1} \otimes \mathbf{X}_n) \text{dvec}(\mathbf{C}_1 \otimes \mathbf{C}_n) \quad (37)$$

$$\mathbf{Y} = \mathbf{P}_R \sum_{n=2}^N \text{dvec}(\mathbf{\Gamma}_{1n}) \mathbf{S}_n^{-1} (\mathbf{\Gamma}_{nn}^{-1} \otimes \mathbf{X}_n) \mathbf{P}_R \text{dvec}(\mathbf{\Gamma}_{1n}) \quad (38)$$

$$\mathbf{S}_n = \mathbf{I}_{R^2} - \left(\mathbf{\Gamma}_{nn}^{-1} \otimes \mathbf{X}_n \right) \text{dvec}(\mathbf{\Gamma}_{nn} \otimes \mathbf{C}_n) \mathbf{P}_R \quad (39)$$

for $n = 2, \dots, N$. In (37) and (39), “ \otimes ” stands for the element-wise division.

Proof: See Appendix B.

Note that in place of inverting the matrix \mathbf{B} of the size $NR^2 \times NR^2$, Theorem 4 reduces the complexity of the CRIB computation to N inversions of the matrices of the size $R^2 \times R^2$. The Theorem can be extended to computing the inverse of the whole Hessian in $O(NR^6)$ operations, see [48].

Finally, note that the assumption that elements of \mathbf{C}_n must not be zero is not too restrictive. Basically, it means that no pair of columns in the factor matrices must be orthogonal. The Cramér-Rao bound does not exhibit any singularity in these cases, and is continuous function of elements of \mathbf{C}_n . If some element of \mathbf{C}_n is closer to zero than say 10^{-5} , it is possible to increase its distance from zero to that value, and the resultant CRIB will differ from the true one only slightly.

Theorem 5: (Properties of the CRIB)

- 1) The CRIB in Theorems 3 and 4 depends on the factor matrices \mathbf{A}_n only through the products $\mathbf{C}_n = \mathbf{A}_n^T \mathbf{A}_n$.
- 2) The CRIB is inversely proportional to the signal-to-noise ratio (SNR) of the factor of the interest (i.e. $\|\mathbf{a}_1\|^2 / (\sigma^2 I_1)$) and independent of the SNR of the other factors, $\|\mathbf{a}_r\|^2 / (\sigma^2 I_r)$, $r = 2, \dots, R$.

Proof: Property 1 follows directly from Theorem 3. Property 2 is proven in Appendix C.

III. SPECIAL CASES

A. Rank 1 Tensors

In this case, the matrix \mathbf{X}_1 is zero, and

$$\text{CRIB}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2} (I_1 - 1) g_{11} = \frac{\sigma^2}{\|\mathbf{a}_1\|^2} (I_1 - 1). \quad (40)$$

In (40), $g_{11} = 1$ due to the convention that the factor matrices \mathbf{A}_n , $n \geq 2$, have columns of unit norm. The result (40) is in accord with Harshman’s early results on uniqueness of rank-1 tensor decomposition [8].

B. Rank 2 Tensors

Consider the scaling convention that all factor vectors except the first factor have unit norm. Let c_n , $|c_n| \leq 1$, be defined as

$$c_n = \begin{cases} \left(\mathbf{a}_1^{(n)} \right)^T \mathbf{a}_2^{(n)} & \text{for } n = 2, \dots, N \\ \left(\mathbf{a}_1^{(1)} \right)^T \mathbf{a}_2^{(1)} / \left(\left\| \mathbf{a}_1^{(1)} \right\| \left\| \mathbf{a}_2^{(1)} \right\| \right) & \text{for } n = 1. \end{cases} \quad (41)$$

It follows from Theorem 5 that the CRIB on \mathbf{a}_1 is a function of c_1, \dots, c_N multiplied by $\sigma^2 / \|\mathbf{a}_1\|^2$. It is symmetric function in c_2, \dots, c_N and possibly nonsymmetric in c_1 . A closed form

expression for the CRIB in the special case is subject of the following theorem.

Theorem 6: It holds for rank 2 tensors

$$\text{CRIB}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \frac{1}{1 - h_1^2} \left[I_1 - 1 + \frac{(1 - c_1^2) h_1^2 [y^2 + z - h_1^2 z(z + 1)]}{(1 - c_1 y - h_1^2(z + 1))^2 - h_1^2(y + c_1 z)^2} \right] \quad (42)$$

where

$$h_n = \prod_{2 \leq k \neq n}^N c_n \quad \text{for } n = 1, \dots, N \quad (43)$$

$$y = -c_1 \sum_{n=2}^N \frac{h_n^2 (1 - c_n^2)}{c_n^2 - h_n^2 c_1^2} \quad (44)$$

$$z = \sum_{n=2}^N \frac{1 - c_n^2}{c_n^2 - h_n^2 c_1^2}. \quad (45)$$

Proof: See Appendix D.

Note that the expressions (44), (45) contain, in their denominators, terms $c_n - h_n c_1$. If any of these terms goes to zero, then quantities y and z go to infinity. In spite of this, the whole CRIB remain finite, because y and z appear both in the numerator and denominator in (42).

For example, for order-3 tensors ($N = 3$) we get (using e.g., Symbolic Matlab or Mathematica)

$$\text{CRIB}_{N=3}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \frac{1}{1 - h_1^2} \left[I_1 - 1 + \frac{c_2^2}{1 - c_2^2} + \frac{c_3^2}{1 - c_3^2} \right]. \quad (46)$$

The above result coincides with the one derived in [29]. As far as the stability is concerned, the CRIB is finite unless either the second or third factor have co-linear columns. Note that the fact that the CRIB for \mathbf{a}_1 does not depend on c_1 can be linked to the uni-mode uniqueness conditions presented in [23].

For $N = 4$, the similar result is hardly tractable. Unlike the case $N = 3$, the result depends on c_1 . A closer inspection of the result shows that the CRIB, as a function of c_1 , achieves its maximum at $c_1 = 0$, and minimum at $c_1 = \pm 1$. Therefore we shall treat these two limit cases separately. We get [(47)-(48) at the bottom of the next page]. As far as the stability is concerned, we can see that the CRIB is always finite unless two of the factor matrices have co-linear columns.

Similarly, for a general N , we have for $c_1 = 0$

$$\text{CRIB}_{c_1=0}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \frac{1}{1 - h_1^2} \left[I_1 - 1 + \frac{h_1^2 z}{1 - h_1^2(z + 1)} \right]. \quad (49)$$

C. A Case With Two Factor Matrices Having Orthogonal Columns

This subsection presents a closed-form CRIB for a tensor of a general order and rank, provided that two of its factor matrices have mutually orthogonal columns. The result cannot be derived from Theorem 5, because assumptions of the theorem are not fulfilled.

Theorem 7: When the factor matrices \mathbf{A}_1 and \mathbf{A}_2 both have mutually orthogonal columns, it holds

$$\text{CRIB}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \left[I_1 - 1 + \sum_{r=2}^R \frac{\gamma_r^2}{1 - \gamma_r^2} \right] \quad (50)$$

where $\gamma_r = \prod_{n=3}^N (\mathbf{a}_1^{(n)})^T \mathbf{a}_r^{(n)}$ for $r = 2, \dots, R$.

Proof: See Appendix E.

Theorem 7 represents an important example when a tensor reshaping (see Section V-A. and [34] for more details) enables very efficient (fast) CP decomposition without compromising accuracy. It has close connection with orthogonally constrained CPD [36], [37], [38].

IV. CRIB FOR TENSORS WITH MISSING OBSERVATIONS

It happens in some applications, that tensors to be decomposed via CP have missing entries (some observations are simply missing). In this case, it is possible to treat stability of the decomposition through the CRIB as well. The only problem is that it is not possible to use expressions in Theorems 3–8 in such cases.

Assume that the tensor to be studied is given by its factor matrices $\mathbf{A}_1, \dots, \mathbf{A}_N$ and a 0–1 “indicator” tensor \mathcal{W} of the same dimension as \mathcal{Y} , which determines which tensor elements are available (observed). The task is to compute CRIB for columns of the factor matrices, like in the previous sections. The CRIB is computed through the Hessian matrix \mathbf{H} as in (12) and (20), but its fast inversion is no longer possible. The Hessian itself can be computed as in its earlier definition

$$\mathbf{H} = \mathbf{J}_W^T(\boldsymbol{\theta}) \mathbf{J}_W(\boldsymbol{\theta}), \quad \mathbf{J}_W(\boldsymbol{\theta}) = \frac{\partial \text{vec}(\mathcal{Y} \circledast \mathcal{W})}{\partial \boldsymbol{\theta}} \quad (51)$$

where $\boldsymbol{\theta}$ is the parameter of the model (2). More specific expressions for the Hessian can be derived in a straightforward manner.

Theorem 8: Consider the Hessian for tensor with missing data as an $N \times N$ partitioned matrix $\mathbf{H} = [\mathbf{H}^{(n,m)}]_{n=1,m=1}^{N,N}$ where $\mathbf{H}^{(n,m)} = [\mathbf{H}_{r,s}^{(n,m)}]_{r=1,s=1}^{R,R} \in \mathbb{R}^{RI_n \times RI_m}$. Then [see (52) at the bottom of the page], $\mathcal{Y} \bar{\times}_n \mathbf{u}_n$ denotes the mode- n tensor-vector product between \mathcal{Y} and \mathbf{u}_n [4], and

$$\mathcal{Y} \bar{\times}_{-n} \{\mathbf{u}\} = \mathcal{Y} \bar{\times}_1 \mathbf{u}_1 \cdots \bar{\times}_{n-1} \mathbf{u}_{n-1} \bar{\times}_{n+1} \mathbf{u}_{n+1} \cdots \bar{\times}_N \mathbf{u}_N. \quad (53)$$

Proof: See Appendix F.

Theorem 8 can be used either to compute the CRIB for tensors with missing elements, or for implementing damped Gauss-Newton method for finding the decomposition in difficult cases, where ALS converges poorly.

V. APPLICATION AND EXAMPLES

A. Tensor Decomposition Through Reshape

Assume that the tensor to-be decomposed is of dimension $N \geq 4$. The tensor can be reshaped to a lower dimensional tensor, which is computationally easier to decompose, so that the first factor matrix remains unchanged. The topic will be better elaborated in our next paper [34], in this paper we present only the main idea on two examples, to demonstrate usefulness of the CRIB.

In the first example, consider $N = 4$. The tensor in (1) can be reshaped to an order-3 tensor

$$\mathcal{Y}_{res} = \sum_{r=1}^R \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \left(\mathbf{a}_r^{(4)} \otimes \mathbf{a}_r^{(3)} \right). \quad (54)$$

Both the original and the re-shaped tensors have the same number of elements ($I_1 I_2 I_3 I_4$) and the same noise added to them.

The question is, what is the accuracy of the factor matrix of the reshaped tensor compared to the original one. The former accuracy should be worse, because a decomposition of the reshaped tensor ignores structure of the third factor matrix. The

$$\text{CRIB}_{N=4,c_1=0}(\mathbf{a}_1) = \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \frac{1}{1-h_1^2} \left[I_1 - 1 + \frac{c_2^2 c_3^2 + c_2^2 c_4^2 + c_3^2 c_4^2 - 3c_2^2 c_3^2 c_4^2}{2c_2^2 c_3^2 c_4^2 - c_2^2 c_3^2 - c_2^2 c_4^2 - c_3^2 c_4^2 + 1} \right] \quad (47)$$

$$\text{CRIB}_{N=4,c_1=\pm 1}(\mathbf{a}_1) = \begin{cases} \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \frac{I_1-1}{1-h_1^2} & \text{for } (|c_2| < 1) \ \& \ (|c_3| < 1) \ \& \ (|c_4| < 1) \\ \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \frac{1}{1-h_1^2} \left[I_1 - 1 + \frac{c_2^2 + c_3^2 - 2c_2^2 c_3^2}{(1-c_2^2)(1-c_3^2)} \right] & \text{for } |c_4| = 1 \\ \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \frac{1}{1-h_1^2} \left[I_1 - 1 + \frac{c_2^2 + c_4^2 - 2c_2^2 c_4^2}{(1-c_2^2)(1-c_4^2)} \right] & \text{for } |c_3| = 1 \\ \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \frac{1}{1-h_1^2} \left[I_1 - 1 + \frac{c_3^2 + c_4^2 - 2c_3^2 c_4^2}{(1-c_3^2)(1-c_4^2)} \right] & \text{for } |c_2| = 1. \end{cases} \quad (48)$$

$$\mathbf{H}_{r,s}^{(n,m)} = \begin{cases} \text{diag} \left(\mathcal{W} \bar{\times}_{-n} \left\{ \mathbf{a}_r^{(1)} \circledast \mathbf{a}_s^{(1)}, \dots, \mathbf{a}_r^{(N)} \circledast \mathbf{a}_s^{(N)} \right\} \right), & \text{for } n = m, \\ \left(\mathbf{a}_r^{(n)} \mathbf{a}_s^{(m)T} \right) \circledast \left(\mathcal{W} \bar{\times}_{-\{n,m\}} \left\{ \mathbf{a}_r^{(1)} \circledast \mathbf{a}_s^{(1)}, \dots, \mathbf{a}_r^{(N)} \circledast \mathbf{a}_s^{(N)} \right\} \right), & \text{for } n \neq m \end{cases} \quad (52)$$

TABLE I
ESTIMATED CRIBS [dB] ON BEST FIT CP COMPONENTS OF FLUORESCENCE TENSOR COMPUTED FOR ASSUMED RANK $R = 1, 2, 3, 4$

Factor n	$R = 1$		$R = 2$		$R = 3$			$R = 4$			
	1		1	2	1	2	3	1	2	3	4
1	44.43		44.44	41.87	64.76	61.34	64.98	65.78	60.96	65.77	38.17
2	27.44		30.28	27.71	53.15	50.17	49.60	54.33	51.39	50.87	23.29
3	32.67		36.23	33.66	58.96	55.75	54.87	60.25	56.28	54.27	25.74

question is, by how much worse. If the difference were negligible, then it is advised to decompose the simpler tensor (of lower dimension).

If the tensor has rank one, accuracy of both decompositions is the same. It is obvious from (40).

Let us examine tensors of rank 2. If the original tensor has scalar products of columns of the factor matrices c_1, c_2, c_3 and c_4 , the reshaped tensor has scalar products c_1, c_2 , and c_3c_4 , respectively. $\text{CRIB}(\mathbf{a}_1)$ of the reshaped tensor is independent of c_1 , while CRIB of the original tensor is dependent on c_1 , so there is a difference, in general. The difference will be smallest for $c_1 = 0$ (orthogonal factors) and largest for c_1 close to ± 1 (nearly or completely co-linear factors along the first dimension).

The smallest difference between $\text{CRIB}(\mathbf{a}_1)$ for the reshaped tensor and for the original one is

$$\frac{\sigma^2}{\|\mathbf{a}_1\|^2} \left[\frac{c_2^2 + c_3^2c_4^2 - 2c_2^2c_3^2c_4^2}{(1 - c_2^2)(1 - c_3^2c_4^2)} - \frac{c_2^2c_3^2 + c_2^2c_4^2 + c_3^2c_4^2 - 3c_2^2c_3^2c_4^2}{(1 - c_2^2c_3^2c_4^2)(2c_2^2c_3^2c_4^2 - c_2^2c_3^2 - c_2^2c_4^2 - c_3^2c_4^2 + 1)} \right]$$

and the largest difference is

$$\frac{\sigma^2}{\|\mathbf{a}_1\|^2} \left[\frac{c_2^2 + c_3^2c_4^2 - 2c_2^2c_3^2c_4^2}{(1 - c_2^2)(1 - c_3^2c_4^2)} \right] = \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \left[\frac{c_2^2}{1 - c_2^2} + \frac{c_3^2c_4^2}{1 - c_3^2c_4^2} \right].$$

We can see that the difference may be large if the second or third factor matrix of the reshaped tensor has nearly co-linear columns ($c_2^2 \approx 1$ or $c_3^2c_4^2 \approx 1$). For example, for a tensor with $I_1 = 5$, $c_1 = 0$, $c_2 = 0.99$, $c_3 = c_4 = 0.1$ the loss of accuracy in decomposing reshaped tensor in place of the original one is 11.22 dB. If c_1 is changed to 1, the loss is only slightly higher, 11.23 dB. If $c_1 = c_2 = 0$, the loss is 0 dB for any c_3, c_4 (compare Theorem 7). If $c_1 = 1$, $c_2 = 0$ and $c_3 = c_4 = 0.99$, the loss is 8.5 dB.

Another example is a tensor of an arbitrary order and rank considered in Theorem 7. Let this tensor be reshaped to the order-3 tensor of the size $I_1 \times I_2 \times (I_3 \dots I_N)$. Comparing the $\text{CRIB}(\mathbf{a}_1)$ of the original tensor and of the reshaped tensor shows that these two coincide. It follows that the decomposition based on reshaping is lossless in terms of accuracy.

B. Amino Acids Tensor

A data set consisting of five simple laboratory-made samples of fluorescence excitation-emission (5 samples \times 201 emission wavelengths \times 61 excitation wavelengths) is considered. Each sample contains different amounts of tryptophan, tyrosine, and phenylalanine dissolved in phosphate buffered water. The samples were measured by fluorescence on a spectrofluorometer

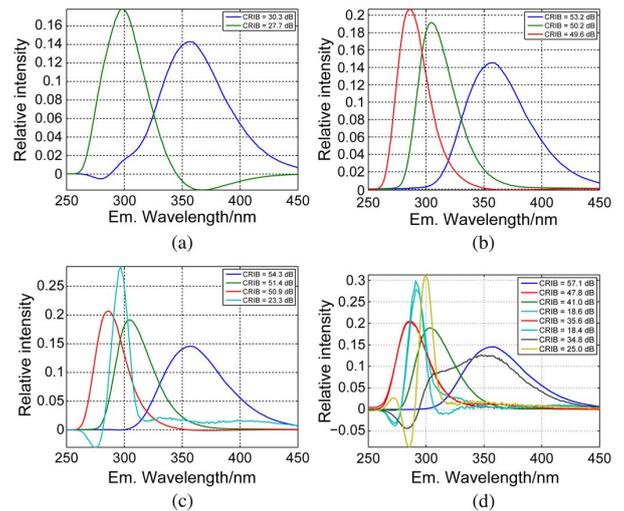


Fig. 1. Illustration for emission components from best-fit decompositions over 100 Monte Carlo runs for example VI-A. (a) Estimated components as $R = 2$. (b) Estimated components as $R = 3$. (c) Estimated components as $R = 4$. (d) Estimated components as $R = 8$.

[43]. Hence, a CP model with $R = 3$ is appropriate to the fluorescence data.

The tensor was factorized for several possible ranks R using the fLM algorithm [28]. CRIBs on the extracted components were then computed with the noise levels deduced from the error tensor $\mathcal{E} = \mathcal{Y} - \hat{\mathcal{Y}}$

$$\sigma^2 = \frac{\|\mathcal{Y} - \hat{\mathcal{Y}}\|_F^2}{\prod_n I_n}. \quad (55)$$

The resultant CRIB's are computed for all columns of all factor matrices and are summarized in Table I.

Note that due to the “ $-10 \log_{10}$ ” definition, high CRIB in dB means high accuracy, and vice versa. A CRIB of 50 dB means that the standard angular deviation (square root of mean square angular error) of the factor is cca 0.18° ; a CRIB of 20 dB corresponds to the standard deviation 5.7° .

The second mode to the decomposition, which represents intensity of the data versus the emission wavelength, for $R = 2, 3, 4$ and 8 is shown in Fig. 1. We can see that the CRIB allows to distinguish between strong/significant modes of the decomposition and possibly artificial modes due to over-fitting the model. The criterion is different in general than the plain “energy” of the factor; if a factor has a low energy, it will probably have high CRIB, but it might not hold true vice versa. A high energy component might have a high CRIB.

In the next experiment, we have studied how much the accuracy of the decomposition is affected in case that some data

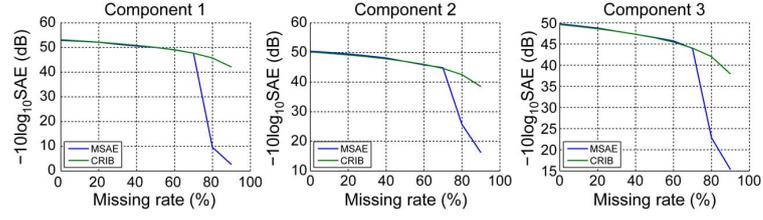


Fig. 2. CRIB for the second-mode components of CP decomposition of tensor in section VI.A with missing elements and mean square angular error obtained in simulations versus percentage of the missing elements.

are missing (not available). The decomposition with the correct rank $R = 3$ and σ^2 estimated as in (55) was taken as a ground truth; the 0-1 indicator tensor \mathcal{W} of the same size was randomly generated with a given percentage of missing values. The CRIB of the second mode factors was plotted in Fig. 2 as a function of this missing value rate. The figure also contains mean square angular error of the components obtained in simulations. Here an artificial Gaussian noise with zero mean and variance σ^2 was added to the “ground truth” tensor. The decomposition was obtained by a Levenberg-Marquardt algorithm [28] modified for tensors with missing entries.

A few observations can be made here.

- CRIB coincides with MSAE for the percentage of the missing entries smaller than 70%. If the percentage exceeds the threshold, CRIB becomes overly optimistic.
- In general, accuracy of the decomposition declines slowly with the number of missing entries. If the number of missing entries is about 20%, loss of accuracy of the decomposition is only about 1–2 dB.

C. Stability of the Decomposition of Brie’s Tensor

Brie *et al.* [20] presented an example of a four-way tensor of rank 3, which arises while studying the response of bacterial bio-sensors to different environmental agents. The tensor has co-linear columns in three of four modes and the main message of the paper is that its CP decomposition is still unique. In this subsection we verify stability of the decomposition.

The factor matrices of the tensor have the form

$$\begin{aligned} \mathbf{A}_1 &= [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3], & \mathbf{A}_2 &= [\mathbf{a}_4, \mathbf{a}_4, \mathbf{a}_5], \\ \mathbf{A}_3 &= [\mathbf{a}_6, \mathbf{a}_7, \mathbf{a}_6], & \mathbf{A}_4 &= [\mathbf{a}_8, \mathbf{a}_9, \mathbf{a}_9]. \end{aligned}$$

Assume for simplicity that all factors have unit norm, $\|\mathbf{a}_n\| = 1$, $n = 1, \dots, 9$. Due to Theorem 5 it holds that CRIB on \mathbf{a}_1 is a function of scalars $c_{11} = \mathbf{a}_1^T \mathbf{a}_2$, $c_{12} = \mathbf{a}_1^T \mathbf{a}_3$, $c_{13} = \mathbf{a}_2^T \mathbf{a}_3$, $c_2 = \mathbf{a}_4^T \mathbf{a}_5$, $c_3 = \mathbf{a}_6^T \mathbf{a}_7$, $c_4 = \mathbf{a}_8^T \mathbf{a}_9$ and I_1 , which is the dimension of \mathbf{a}_1 . Then, the matrices $\mathbf{C}_n = \mathbf{A}_n^T \mathbf{A}_n$, $n = 2, 3, 4$, have the form

$$\begin{aligned} \mathbf{C}_2 &= \begin{bmatrix} 1 & 1 & c_2 \\ 1 & 1 & c_2 \\ c_2 & c_2 & 1 \end{bmatrix}, & \mathbf{C}_3 &= \begin{bmatrix} 1 & c_3 & 1 \\ c_3 & 1 & c_3 \\ 1 & c_3 & 1 \end{bmatrix} \\ \mathbf{C}_4 &= \begin{bmatrix} 1 & c_4 & c_4 \\ c_4 & 1 & 1 \\ c_4 & 1 & 1 \end{bmatrix}. \end{aligned}$$

A straightforward usage of Theorem 4 is not possible, because some of the involved matrices become singular. The CRIB itself, however, is finite and can be computed using an artificial parameter ε as a limit. The limit CRIB is computed for modified matrices at $\varepsilon \rightarrow 0$,

$$\begin{aligned} \mathbf{C}_{2\varepsilon} &= \begin{bmatrix} 1 & 1 - \varepsilon & c_2 \\ 1 - \varepsilon & 1 & c_2 \\ c_2 & c_2 & 1 \end{bmatrix} \\ \mathbf{C}_{3\varepsilon} &= \begin{bmatrix} 1 & c_3 & 1 - \varepsilon \\ c_3 & 1 & c_3 \\ 1 - \varepsilon & c_3 & 1 \end{bmatrix} \\ \mathbf{C}_{4\varepsilon} &= \begin{bmatrix} 1 & c_4 & c_4 \\ c_4 & 1 & 1 - \varepsilon \\ c_4 & 1 - \varepsilon & 1 \end{bmatrix}. \end{aligned}$$

If any of the correlations c_2, c_3, c_4 is zero, it is also augmented by ε .

The limit CRIB can be shown to be independent of off-diagonal elements of \mathbf{C}_1 , unless \mathbf{C}_1 is singular. Assume that \mathbf{C}_1 is regular. The result, obtained by Symbolic Matlab, is

$$\begin{aligned} \text{CRIB}_{\varepsilon=0}(\mathbf{a}_1) &= \frac{\sigma^2}{\|\mathbf{a}_1\|^2} \frac{1}{2c_2^2 c_3^2 c_4^2 - c_2^2 c_3^2 - c_2^2 c_4^2 - c_3^2 c_4^2 + 1} \\ &\cdot \left[(I_1 - 1) (1 - c_2^2 c_3^2) - \frac{c_4^4 (c_2^2 + 1) - 3c_3^2 + 1}{1 - c_3^2} \right. \\ &\quad \left. - \frac{c_2^4 (c_3^2 + 1) - 3c_2^2 + 1}{1 - c_2^2} + \frac{2 - c_2^2 - c_3^2}{1 - c_4^2} \right]. \quad (56) \end{aligned}$$

It follows that the decomposition is stable, unless all three factors in some mode are collinear.

D. Maximum Stable Rank

A theoretically interesting question is, what is the maximum rank of a tensor of a given dimension which has a stable CP decomposition (with finite CRIB). For easy reference, we shall call it *maximum stable rank* and denote it $R_{smax}(I_1, \dots, I_N)$.

An upper bound for the maximum stable rank can be deduced from the requirement that the number of free parameters in the model, which is $R(\sum_{n=1}^N I_n - N + 1)$ in CP decomposition, cannot exceed dimension of the available data, which is $\prod_{n=1}^N I_n$. It follows that

$$R_{smax}(I_1, \dots, I_N) \leq \left\lfloor \frac{\prod_{n=1}^N I_n}{\sum_{n=1}^N I_n - N + 1} \right\rfloor \quad (57)$$

where $\lfloor x \rfloor$ denotes the lower integer part of x . It can be verified numerically that for many (and maybe all¹) tensor dimensions, an equality in (57) holds. In other words, it means that the CRIB computed, e.g., via Theorem 4 for a CP decomposition with rank $R = R_{smax}$ and some (e.g. random) factor matrices is finite. For example, the maximum stable rank is $R_{smax} = 2$ for $2 \times 2 \times 2$ tensors, and $R_{smax} = 3$ for $3 \times 3 \times 3$ tensors. For order-8 tensors of dimension $2 \times \dots \times 2$, ($8 \times$), it holds $R_{smax} = 28$.

It might be interesting to compare the maximum stable rank with the maximum rank and the maximum typical rank (to be explained below) for given tensor dimension, if they are known [46]. If the elements of a tensor are chosen randomly according to a continuous probability distribution, there is not a rank which occurs with probability 1 in general. Such rank, if exists, is called generic. Ranks which occur with strictly positive probabilities are called typical ranks. For example it was computed in [10] that probability for a real random Gaussian tensor of the size $2 \times 2 \times 2$ to be 2 and 3 is $\pi/4$, and $1 - \pi/4$, respectively. We can see that no tensor of the rank 3 and the dimension has a stable decomposition. For tensors of the dimension $3 \times 3 \times 3$ the typical rank is 5 [10], it is a generic rank—but no decomposition of these rank-5 tensors is stable, as $R_{smax} = 3$.

Next, it might be interesting to compare the maximum stable rank with the maximum rank for unique tensor decomposition, or prove that these two coincide. Liu and Sidiropoulos [11], [31] derived a necessary condition for uniqueness of the CP decomposition, which, according to a formulation in [45] reads

$$\min_{n=1, \dots, N} \text{rank}(\mathbf{A}_1 \odot \dots \odot \mathbf{A}_{n-1} \odot \mathbf{A}_{n+1} \odot \dots \odot \mathbf{A}_N) = R \quad (58)$$

where \odot means the Khatri-Rao product. The condition (58) is equivalent to the condition that the matrices $\Xi_n = \mathbf{A}_1 \odot \dots \odot \mathbf{A}_{n-1} \odot \mathbf{A}_{n+1} \odot \dots \odot \mathbf{A}_N$ have all full column rank, $n = 1, \dots, N$, which is further equivalent to the condition that the product $\Xi_n^T \Xi_n$ are regular for $n = 1, \dots, N$. Finally note that

$$\Xi_n^T \Xi_n = \Gamma_{nn}, \quad n = 1, \dots, N.$$

where Γ_{nn} was defined in (5) and appears in computation of the CRIB.

Unfortunately, it appears that the condition (58) is only necessary, but not sufficient for uniqueness. It is often fulfilled for R higher than R_{smax} . Thus a relation between the stability and uniqueness of the CP decomposition remains open question for now.

VI. CONCLUSIONS

Cramér-Rao bounds for CP tensor decomposition represent an important tool for studying accuracy and stability of the decomposition. The bounds derived in this manuscript serve as a theoretical support for a method of the decomposition through tensor reshaping [34]. As a side result, a novel method of inverting Hessian matrix, which is more computationally efficient, is derived for the problem. It enables a further improvement of speed of the fast Gauss-Newton for the problem

¹We do not have yet a formal proof that the equality in (57) holds for *all* tensor dimensions and orders.

[28], [48]. A novel expression for Hessian for CP decomposition of tensor with missing entries has been derived. It can serve for assessing accuracy of CP decomposition of these tensors without need of long Monte Carlo simulations, and for implementing a damped Gauss-Newton algorithm for CP decomposition of these tensors.

A direct link between stability and essential uniqueness remains to be an open theoretical question. In particular, it is not known yet for sure if stability implies the essential uniqueness.

CRB expressions similar to the ones derived in this paper can be also derived for other important special tensor decomposition models such as INDSCAL (where two or more factor matrices coincide) [16], [39], or for the PARALIND model, where the factor matrices have certain structure [23], and for block factorization methods.

APPENDIX A

Matrix Inversion Lemma (Woodbury identity): Let \mathbf{A} , \mathbf{X} , \mathbf{Y} , and \mathbf{R} are matrices of compatible dimensions such that the following products and inverses exist. Then

$$(\mathbf{A} + \mathbf{XRY})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}(\mathbf{R}^{-1} + \mathbf{Y}\mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{Y}\mathbf{A}^{-1}. \quad (59)$$

APPENDIX B

Proof of Theorem 4: Let the matrices \mathbf{K} and Ψ in (18) be partitioned as

$$\mathbf{K} = \begin{bmatrix} \mathbf{0} & \mathbf{K}_1 \\ \mathbf{K}_1^T & \mathbf{K}_2 \end{bmatrix}, \quad \Psi = \begin{bmatrix} \Psi_1 & \mathbf{0} \\ \mathbf{0} & \Psi_2 \end{bmatrix} \quad (60)$$

where the left-upper blocks have the size $R^2 \times R^2$. Then, using a formula for inverse of partitioned matrices, the left-upper block of \mathbf{B} in (18) can be written as

$$\begin{aligned} \mathbf{B}_0 &= \mathbf{K}_1 (\mathbf{I}_{(N-1)R^2} + \Psi_2 \mathbf{K}_2 - \Psi_2 \mathbf{K}_1^T \Psi_1 \mathbf{K}_1)^{-1} \Psi_2 \mathbf{K}_1^T \\ &\triangleq \mathbf{K}_1 \mathbf{K}_3^{-1} \Psi_2 \mathbf{K}_1^T. \end{aligned} \quad (61)$$

A key observation which enables a fast inversion of the term \mathbf{K}_3 is that

$$\mathbf{K} = \mathbf{K}_0 + \mathbf{DFD}^T \quad (62)$$

where

$$\mathbf{K}_0 = -\text{bdiag} \left(\mathbf{P}_R \mathbf{F} (\text{dvec}(1 \otimes \mathbf{C}_n))^2 \right)_{n=1}^N \quad (63)$$

$$\mathbf{F} = \mathbf{P}_R \prod_{n=1}^N \text{dvec}(\mathbf{C}_n) = \mathbf{P}_R \text{dvec}(\Gamma_{11} \otimes \mathbf{C}_1) \quad (64)$$

$$\mathbf{D} = [\text{dvec}(1 \otimes \mathbf{C}_1), \dots, \text{dvec}(1 \otimes \mathbf{C}_N)]^T. \quad (65)$$

Similarly,

$$\mathbf{K}_2 = \mathbf{K}_{02} + \mathbf{D}_2 \mathbf{F} \mathbf{D}_2^T \quad (66)$$

where

$$\mathbf{K}_{02} = -\text{bdiag} \left(\mathbf{P}_R \mathbf{F} (\text{dvec}(1 \otimes \mathbf{C}_n))^2 \right)_{n=2}^N \quad (67)$$

$$\mathbf{D}_2 = [\text{dvec}(1 \otimes \mathbf{C}_2), \dots, \text{dvec}(1 \otimes \mathbf{C}_N)]^T. \quad (68)$$

Then the matrix \mathbf{K}_3 in (61) can be written as

$$\begin{aligned}\mathbf{K}_3 &= \mathbf{I}_{(N-1)R^2} + \Psi_2 \mathbf{K}_2 - \Psi_2 \mathbf{K}_1^T \Psi_1 \mathbf{K}_1 \\ &= \mathbf{I}_{(N-1)R^2} + \Psi_2 (\mathbf{K}_{02} - \mathbf{K}_1^T \Psi_1 \mathbf{K}_1) + \Psi_2 \mathbf{D}_2 \mathbf{F} \mathbf{D}_2^T \\ &= \mathbf{Q} + \Psi_2 \mathbf{D}_2 \mathbf{F} \mathbf{D}_2^T\end{aligned}\quad (69)$$

where

$$\begin{aligned}\mathbf{Q} &= \text{bdiag}(\mathbf{Q}_n)_{n=2}^N \quad (70) \\ \mathbf{Q}_n &= \mathbf{I}_{R^2} - \left(\Gamma_{nn}^{-1} \otimes \mathbf{X}_n \right) \mathbf{P}_R \\ &\quad \times \left(\mathbf{F} (\text{dvec}(1 \otimes \mathbf{C}_n))^2 \right. \\ &\quad \left. + \text{dvec}(\Gamma_{1n}) \left(\Gamma_{11}^{-1} \otimes \mathbf{C}_1 \right) \text{dvec}(\Gamma_{1n}) \mathbf{P}_R \right).\end{aligned}\quad (71)$$

Now, \mathbf{K}_3 can be easily inverted using the matrix inversion lemma (59),

$$\begin{aligned}\mathbf{K}_3^{-1} &= \mathbf{Q}^{-1} \\ &\quad - \mathbf{Q}^{-1} \mathbf{D}_2^T \left(\mathbf{I}_{R^2} + \mathbf{D}_2^T \mathbf{Q}^{-1} \Psi_2 \mathbf{D}_2 \mathbf{F} \right)^{-1} \Psi_2 \mathbf{D}_2 \mathbf{F} \mathbf{Q}^{-1}.\end{aligned}\quad (72)$$

Inserting (72) in (61) gives, after some simplifications, the result (35). \blacksquare

APPENDIX C

Proof of Theorem 5: Consider the change of scale of columns of factor matrices up to their first columns. As in Section II assume that the scale change is realized in \mathbf{A}_1 , while the other factor matrices have columns of unit norm. The theorem claims that the substitution $\mathbf{A}_1 \leftarrow \mathbf{A}_1 \mathbf{D}$ into (27) where $\mathbf{D} = \text{diag}(1, \lambda_2, \dots, \lambda_R)$, $\lambda_r \neq 0$, has no influence on $\text{CRIB}(\mathbf{a}_1)$.

The substitution $\mathbf{A}_1 \leftarrow \mathbf{A}_1 \mathbf{D}$ leads to $\mathbf{C}_1 \leftarrow \mathbf{D} \mathbf{C}_1 \mathbf{D}$ and $\mathbf{X}_1 \leftarrow \mathbf{D} \mathbf{X}_1 \mathbf{D}$ while \mathbf{C}_n and \mathbf{X}_n , $n = 2, \dots, N$, remain the same. Consequently, Γ_{1n} , $n = 1, \dots, N$, remain unchanged while $\Gamma_{nn} \leftarrow \mathbf{D} \Gamma_{nn} \mathbf{D}$ for $n = 2, \dots, N$. Now, we can substitute into (35) assuming that the condition of Theorem 4 is satisfied.

Let $\tilde{\mathbf{S}}_n$ denote the matrix \mathbf{S}_n in (39) after the substitution $\mathbf{A}_1 \leftarrow \mathbf{A}_1 \mathbf{D}$. It can be shown that $(\mathbf{D} \otimes \mathbf{I}_R) \tilde{\mathbf{S}}_n = \mathbf{S}_n (\mathbf{D} \otimes \mathbf{I}_R)$ using the rules

$$\begin{aligned}(\mathbf{D} \Gamma_{nn} \mathbf{D})^{-1} \otimes \mathbf{X}_n &= (\mathbf{D}^{-1} \otimes \mathbf{I}_R) \left(\Gamma_{nn}^{-1} \otimes \mathbf{X}_n \right) \\ &\quad \times (\mathbf{D}^{-1} \otimes \mathbf{I}_R)\end{aligned}\quad (73)$$

$$\text{dvec}(\mathbf{D} \Gamma_{nn} \mathbf{D} \otimes \mathbf{C}_n) = (\mathbf{D} \otimes \mathbf{D}) \text{dvec}(\Gamma_{nn} \otimes \mathbf{C}_n)\quad (74)$$

$$(\mathbf{I}_R \otimes \mathbf{D}) \mathbf{P}_R = \mathbf{P}_R (\mathbf{D} \otimes \mathbf{I}_R)\quad (75)$$

and the fact that diagonal matrices commute. Using the same rules in further substitutions, after some computations, the independence of $\text{CRIB}(\mathbf{a}_1)$ on \mathbf{D} follows.

APPENDIX D

Proof of Theorem 6: Again, assume for simplicity that all factors have unit norms. It holds

$$\Gamma_{11} = \begin{bmatrix} 1 & h_1 \\ h_1 & 1 \end{bmatrix}, \quad \mathbf{X}_n = \begin{bmatrix} 0 & 0 \\ 0 & 1 - c_n^2 \end{bmatrix}, \quad n = 1, \dots, N.$$

and

$$g_{11} = \left[\Gamma_{11}^{-1} \right]_{11} = \frac{1}{1 - h_1^2}\quad (76)$$

$$\mathbf{g}_{1,:} = g_{11} [1, \quad -h_1].\quad (77)$$

The matrix Ψ in (32) can be decomposed as $\Psi = \mathbf{J} \Phi$ where

$$\mathbf{J} = \text{bdiag}(\mathbf{I}_4, \mathbf{I}_2 \otimes [0, 1]^T, \dots, \mathbf{I}_2 \otimes [0, 1]^T)\quad (78)$$

$$\begin{aligned}\Phi &= \text{bdiag} \left(\Gamma_{11}^{-1} \otimes \mathbf{C}_1, (1 - c_2^2) \Gamma_{22}^{-1} \otimes [0, 1], \dots, \right. \\ &\quad \left. (1 - c_N^2) \Gamma_{NN}^{-1} \otimes [0, 1] \right).\end{aligned}\quad (79)$$

Then the matrix \mathbf{B} in (18) can be rewritten using the Woodbury identity (59) as

$$\begin{aligned}\mathbf{B} &= \mathbf{K} (\mathbf{I}_{4N} + \mathbf{J} \Phi \mathbf{K})^{-1} \\ &= \mathbf{K} - \mathbf{K} \mathbf{J} (\mathbf{I}_{2N+2} + \Phi \mathbf{K} \mathbf{J})^{-1} \Phi \mathbf{K}.\end{aligned}\quad (80)$$

Now, put $\mathbf{B}_4 = \mathbf{I}_{2N+2} + \Phi \mathbf{K} \mathbf{J}$ and write it in the block form as

$$\mathbf{B}_4 = \mathbf{I}_{2N+2} + \Phi \mathbf{K} \mathbf{J} = \begin{bmatrix} \mathbf{B}_{41} & \mathbf{B}_{42} \\ \mathbf{B}_{43} & \mathbf{B}_{44} \end{bmatrix}\quad (81)$$

where \mathbf{B}_{41} has the size 4×4 . The bottom-right block \mathbf{B}_{44} of dimension $(2N - 2) \times (2N - 2)$ is easy to be inverted using the Woodbury identity again, because it can be written as

$$\mathbf{B}_{44} = \mathbf{B}_5 + \mathbf{s} \mathbf{f}^T\quad (82)$$

where

$$\mathbf{B}_5 = \text{bdiag}(\mathbf{B}_{52}, \dots, \mathbf{B}_{5N})\quad (83)$$

$$\mathbf{B}_{5n} = \begin{bmatrix} 1 & -\frac{h_n c_1 (1 - c_n^2)}{1 - h_n^2 c_1^2} \\ 0 & \frac{c_n^2 - h_n^2 c_1^2}{1 - h_n^2 c_1^2} \end{bmatrix}, \quad n = 2, \dots, N\quad (84)$$

$$\begin{aligned}\mathbf{s} &= \left[-\frac{h_2 c_1 (1 - c_2^2)}{1 - h_2^2 c_1^2}, \frac{(1 - c_2^2)}{1 - h_2^2 c_1^2}, \dots, \right. \\ &\quad \left. -\frac{h_N c_1 (1 - c_N^2)}{1 - h_N^2 c_1^2}, \frac{(1 - c_N^2)}{1 - h_N^2 c_1^2} \right]^T\end{aligned}\quad (85)$$

$$\mathbf{f} = [0, \quad 1, \quad 0, \quad 1, \quad \dots, \quad 1]^T.\quad (86)$$

After some computations, we receive the result (42). \blacksquare

APPENDIX E

Proof of Theorem 7: Under the assumption of the Theorem, it holds that the matrix $\mathbf{C}_1 = \mathbf{A}_1^T \mathbf{A}_1$ is diagonal and $\mathbf{C}_2 = \mathbf{I}_R$ (identity matrix). Thanks to Theorem 5 we can assume, without any loss of generality, that $\mathbf{C}_1 = \mathbf{I}_R$ as well. It can be shown for Γ_{mn} in (5) that $\Gamma_{mn} = \mathbf{I}_R$ for all pairs (m, n) , $(m, n) \neq (1, 2), (2, 1)$. Only Γ_{12} and $\Gamma_{21} = \Gamma_{12}$ are possibly different. Note that the first row of Γ_{12} is $(1, \gamma_2, \dots, \gamma_N)$.

It follows from these observations that all non-diagonal $R^2 \times R^2$ blocks \mathbf{K}_{mn} of \mathbf{K} in (6) with $(m, n) \neq (1, 2), (2, 1)$ are identical, diagonal, having 1 at positions (p, p) , $p = 1, R + 1, 2R + 2, \dots, R^2$ and 0 elsewhere. In other words, these \mathbf{K}_{mn} can be written as $\mathbf{K}_{mn} = \mathbf{Q} \mathbf{Q}^T$, where \mathbf{Q} is a 0-1 matrix of the

size $R^2 \times R$, the p -th column of \mathbf{Q} has the value 1 at position $(p-1)(R+1)+1$ and 0 elsewhere.

Computation of the CRIB can proceed from equation (61) by inserting the special form of the blocks of \mathbf{K}_1 and \mathbf{K}_2 and using the Woodbury identity (59). ■

APPENDIX F

Proof of Theorem 8: The following identities are used in this proof

$$\text{vec}(\mathbf{A} \circledast \mathbf{B}) = \text{dvec}(\mathbf{B}) \text{vec}(\mathbf{A}), \quad (87)$$

$$\mathbf{a}^T \text{diag}(\mathbf{b}) \mathbf{c} = (\mathbf{a} \circledast \mathbf{c})^T \mathbf{b}, \quad (88)$$

$$(\mathbf{a} \otimes \mathbf{b}) \circledast (\mathbf{c} \otimes \mathbf{d}) = (\mathbf{a} \circledast \mathbf{c}) \otimes (\mathbf{b} \circledast \mathbf{d}). \quad (89)$$

Here, dimensions of \mathbf{a} , \mathbf{b} , \mathbf{c} and \mathbf{d} are assumed to match accordingly.

The approximate Hessian in (51) is given by

$$\mathbf{H} = \mathbf{J}_W^T(\boldsymbol{\theta}) \mathbf{J}_W(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta})^T \text{dvec}(\mathcal{W}) \mathbf{J}(\boldsymbol{\theta}), \quad (90)$$

where $\mathbf{J}(\boldsymbol{\theta})$ is the Jacobian for the complete data.

We have

$$\frac{\partial \text{vec} \mathcal{Y}}{\partial \mathbf{a}_{ir}^{(n)}} = \left(\bigotimes_{k=n+1}^N \mathbf{a}_r^{(k)} \right) \otimes \mathbf{e}_i^{(n)} \otimes \left(\bigotimes_{k=1}^{n-1} \mathbf{a}_r^{(k)} \right), \quad (91)$$

where unit vector $\mathbf{e}_i^{(n)}$ for $i = 1, 2, \dots, I_n$ is the i -th column of the identity matrix of size $I_n \times I_n$.

An (i, j) entry of a sub matrix $\mathbf{H}_{r,s}^{(n,n)}$ for $i = 1, 2, \dots, I_n$, and $j = 1, 2, \dots, I_n$ is given by

$$\begin{aligned} \mathbf{H}_{r,s}^{(n,n)}(i, j) &= \left(\frac{\partial \text{vec} \mathcal{Y}}{\partial \mathbf{a}_{ir}^{(n)}} \right)^T \text{dvec}(\mathcal{W}) \left(\frac{\partial \text{vec} \mathcal{Y}}{\partial \mathbf{a}_{js}^{(n)}} \right) \\ &= \left(\left(\bigotimes_{k=n+1}^N \mathbf{a}_r^{(k)} \circledast \mathbf{a}_s^{(k)} \right) \otimes \left(\mathbf{e}_i^{(n)} \circledast \mathbf{e}_j^{(n)} \right) \right. \\ &\quad \left. \otimes \left(\bigotimes_{k=1}^{n-1} \mathbf{a}_r^{(k)} \circledast \mathbf{a}_s^{(k)} \right) \right)^T \text{vec} \mathcal{W} \quad (92) \end{aligned}$$

$$= \mathcal{W} \bar{\mathbf{x}}_{-n} \{\mathbf{b}^{(k)}\} \bar{\mathbf{x}}_n \delta_{ij} \mathbf{e}_i^{(n)} \quad (93)$$

where δ_{ij} is the Kronecker delta, $\mathbf{b}^{(n)} = \mathbf{a}_r^{(n)} \circledast \mathbf{a}_s^{(n)}$, for $n = 1, \dots, N$. This leads to that a diagonal sub-matrix $\mathbf{H}_{r,s}^{(n,n)}$ is a diagonal matrix as in Theorem IV.

For off-diagonal sub matrices $\mathbf{H}_{r,s}^{(n,m)}$ of size $I_n \times I_m$ ($1 \leq n < m \leq N$), we have

$$\begin{aligned} \mathbf{H}_{r,s}^{(n,m)}(i, j) &= \left(\frac{\partial \text{vec}(\mathcal{Y})}{\partial \mathbf{a}_{ir}^{(n)}} \right)^T \text{dvec}(\mathcal{W}) \left(\frac{\partial \text{vec} \mathcal{Y}}{\partial \mathbf{a}_{js}^{(m)}} \right) \\ &= \left(\left(\bigotimes_{k=m+1}^N \mathbf{a}_r^{(k)} \circledast \mathbf{a}_s^{(k)} \right) \right. \\ &\quad \left. \otimes \left(\mathbf{a}_r^{(m)} \circledast \mathbf{e}_j^{(m)} \right) \otimes \left(\bigotimes_{k=n+1}^{m-1} \mathbf{a}_r^{(k)} \circledast \mathbf{a}_s^{(k)} \right) \right)^T \end{aligned}$$

$$\begin{aligned} &\otimes \left(\mathbf{e}_i^{(n)} \circledast \mathbf{a}_s^{(n)} \right) \\ &\otimes \left(\bigotimes_{k=1}^{n-1} \mathbf{a}_r^{(k)} \circledast \mathbf{a}_s^{(k)} \right) \Big)^T \text{vec}(\mathcal{W}) \quad (94) \\ &= a_{jr}^{(m)} a_{is}^{(n)} \left(\mathcal{W} \bar{\mathbf{x}}_{-\{n,m\}} \left\{ \mathbf{b}^{(k)} \right\} \right. \\ &\quad \left. \bar{\mathbf{x}}_n \mathbf{e}_i^{(n)} \bar{\mathbf{x}}_m \mathbf{e}_j^{(m)} \right). \quad (95) \end{aligned}$$

This leads to the compact form in Theorem 8. ■

REFERENCES

- [1] R. Bro, "Multi-way analysis in the food industry models, algorithms, and applications," Univ. Amsterdam, Amsterdam, The Netherlands [Online]. Available: [http://www/models.life.ku.dk/research/theses/](http://www.models.life.ku.dk/research/theses/), 1998.
- [2] P. M. Kroonenberg, *Applied Multiway Data Analysis*. New York, NY, USA: Wiley, 2008.
- [3] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis: Applications in the Chemical Sciences*. New York, NY, USA: Wiley, 2004.
- [4] A. Cichocki, R. Zdunek, A. H. Phan, and S. I. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. New York, NY, USA: Wiley, 2009.
- [5] V. De Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM J. Matrix Anal. Appl.*, vol. 30, pp. 1084–1127, 2008.
- [6] W. P. Krijnen, T. K. Dijkstra, and A. Stegeman, "On the non-existence of optimal solutions and the occurrence of "degeneracy" in the Candecomp/Parafac model," *Psychometrika*, vol. 73, pp. 431–439, 2008.
- [7] R. A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an "explanatory" multimode factor analysis," *UCLA Working Papers Phonet.*, vol. 16, pp. 1–84, 1970.
- [8] R. A. Harshman, "Determination and proof of minimum uniqueness conditions for PARAFAC," *UCLA Working Papers Phonet.*, vol. 22, pp. 111–117, 1972.
- [9] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, pp. 95–138, 1977.
- [10] J. B. Kruskal, "Rank, decomposition, and uniqueness for 3-way and N-way arrays," in *Multivary Data Analysis*. Amsterdam, The Netherlands: North-Holland, 1989, pp. 7–18.
- [11] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *J. Chemometr.*, vol. 14, pp. 229–239, May 2000.
- [12] J. M. F. Ten Berge and N. D. Sidiropoulos, "On uniqueness in CANDECOP/PARAFAC," *Psychometrika*, vol. 67, no. 3, pp. 399–409, Sep. 2002.
- [13] T. Jiang and N. D. Sidiropoulos, "Kruskal's permutation lemma and the identification of Candecomp/Parafac and bilinear models with constant modulus constraints," *IEEE Trans. Signal Process.*, vol. 52, pp. 2625–2636, 2004.
- [14] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 28, pp. 642–666, 2006.
- [15] A. Stegeman and N. D. Sidiropoulos, "On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition," *Linear Algebra Appl.*, vol. 420, pp. 540–552, 2007.
- [16] A. Stegeman, "On uniqueness conditions for Candecomp/Parafac and Indscal with full column rank in one mode," *Linear Algebra Appl.*, vol. 431, no. 1-2, pp. 211–227, 2009.
- [17] A. Stegeman and A. L. F. de Almeida, "Uniqueness conditions for constrained three-way factor decompositions with linearly dependent loadings," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 3, pp. 1469–1490, Aug. 2009.
- [18] A. Stegeman, "On uniqueness of the n th order tensor decomposition into rank-1 terms with linear independence in one mode," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 5, pp. 2498–2516, 2010.
- [19] A. Stegeman, "On uniqueness of the canonical tensor decomposition with some form of symmetry," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 3, pp. 561–583, 2011.
- [20] D. Brie, S. Miron, F. Caland, and C. Mustin, "An uniqueness condition for the 4-way CANDECOP/PARAFAC model with collinear loadings in three modes," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2011, pp. 4112–4115.
- [21] I. Domanov and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part I: Basic results and uniqueness of one factor matrix," [Online]. Available: <http://arxiv.org/pdf/1301.4602v1.pdf>

- [22] I. Domanov and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part II: Uniqueness of the overall decomposition," [Online]. Available: <http://arxiv.org/pdf/1301.4603v1.pdf>
- [23] X. Guo, S. Miron, D. Brie, and A. Stegeman, "Uni-mode and partial uniqueness conditions for CANDECOMP/PARAFAC of three-way arrays with linearly dependent loadings," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 1, pp. 111–129, 2012.
- [24] A. Stegeman, "Candecomp/Parafac: From diverging components to a decomposition in block terms," *SIAM J. Matrix Anal. Appl.*, vol. 33, pp. 291–316, 2012.
- [25] S. Basu and Y. Bresler, "The stability of nonlinear least squares problems and the Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 48, pp. 3426–3436, Dec. 2000.
- [26] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the PARAFAC model," *Computat. Statist. Data Anal.*, vol. 50, no. 7, pp. 1700–1734, Apr. 2006.
- [27] P. Paatero, "A weighted non-negative least squares algorithm for three-way 'PARAFAC' factor analysis," *Chemometr. Intell. Lab. Syst.*, vol. 38, pp. 223–242, 1997.
- [28] A.-H. Phan, P. Tichavský, and A. Cichocki, "Low complexity damped Gauss-Newton algorithms for CANDECOMP/PARAFAC," *SIAM J. Matrix Anal. Appl.*, vol. 34, pp. 126–147, 2013.
- [29] P. Tichavský and Z. Koldovský, "Stability of CANDECOMP-PARAFAC tensor decomposition," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, 2011, pp. 4164–4167.
- [30] P. Tichavský and Z. Koldovský, "Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1037–1047, Mar. 2011.
- [31] X. Liu and N. D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, Sep. 2001.
- [32] T. Jiang and N. D. Sidiropoulos, "Blind identification of out of cell users in DS-CDMA," *EURASIP J. Appl. Signal Process. (Special Issue on Advances in Smart Antennas)*, no. 9, pp. 1212–1224, Aug. 2004.
- [33] Z. Koldovský, P. Tichavský, and A.H. Phan, "Stability analysis and fast damped-gauss-newton algorithm for INDSCAL tensor decomposition," in *Proc. IEEE Workshop Statist. Signal Process.*, Nice, France, Jun. 2011, pp. 581–584.
- [34] A.-H. Phan, P. Tichavský, and A. Cichocki, "CANDECOMP/PARAFAC decomposition of high-order tensors through tensor reshaping," *IEEE Trans. Signal Process.*, submitted for publication.
- [35] R. Bro, R. A. Harshman, N. D. Sidiropoulos, and M. E. Lundy, "Modeling multi-way data with linearly dependent loadings," *J. Chemometr.*, vol. 23, no. 7–8, pp. 324–340, 2009.
- [36] M. B. Dosse, J. M. F. Berge, and J. N. Tendeiro, "Some new results on orthogonally constrained candecomp," *J. Class.*, vol. 28, pp. 144–155, 2011.
- [37] M. Sorensen, L. De Lathauwer, P. Comon, S. Icart, and L. Deneire, "Canonical polyadic decomposition with a columnwise orthonormal factor matrix," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 4, pp. 1190–1213, 2012.
- [38] J. Chen and Y. Saad, "On the tensor SVD and the optimal low rank orthogonal approximation of tensors," *SIAM J. Matrix Anal. Appl.*, vol. 30, pp. 1709–1734, 2009.
- [39] J. Tendeiro, M. Bennani Dosse, and J. M. F. Ten Berge, "First and second-order derivatives for CP and INDSCAL," *Chemometr. Intell. Lab. Syst.*, vol. 106, pp. 27–36, 2011.
- [40] A. Householder, *The Theory of Matrices in Numerical Analysis*. New York, NY, USA: Blaisdell, 1964.
- [41] G. Tomasi and R. Bro, "PARAFAC and missing values," *Chemometr. Intell. Lab. Syst.*, vol. 75, pp. 163–180, 2005.
- [42] R. Bro, "Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis," *Chemometr. Intell. Lab. Syst.*, vol. 46, pp. 133–147, 1999.
- [43] R. Bro, "Multi-way Analysis in the Food Industry—Models, Algorithms, and Applications," Ph.D. thesis, Univ. Amsterdam, Amsterdam, The Netherlands, 1998.
- [44] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1994.
- [45] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Sep. 2009.
- [46] P. Comon, J. M. F. ten Berge, L. De Lathauwer, and J. Castaing, "Generic and typical ranks of multi-way arrays," *Linear Algebra Appl.*, vol. 430, no. 11, pp. 2997–3007, 2009, (2009).
- [47] T. G. Kolda, "Orthogonal tensor decompositions," *SIAM J. Matrix Anal. Appl.*, vol. 23, pp. 243–255, 2001.
- [48] P. Tichavský, A. H. Phan, and A. Cichocki, "A further improvement of a fast damped Gauss-Newton algorithm for CP tensor decomposition," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2013.



Petr Tichavský (M'98–SM'04) received the M.S. degree in mathematics in 1987 from the Czech Technical University, Prague, and the Ph.D. degree in theoretical cybernetics from the Czechoslovak Academy of Sciences in 1992.

Since that time, he has been with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague. In 1994, he received the Fulbright grant for a ten-month fellowship at with the Department of Electrical Engineering, Yale University, New Haven, CT. In

2002, he received the Otto Wichterle Award from the Academy of Sciences of the Czech Republic. He is an author and coauthor of research papers in the area of sinusoidal frequency/rate estimation, adaptive filtering and tracking of time varying signal parameters, algorithm-independent bounds on achievable performance, sensor array processing, independent component analysis, and blind source separation.

Dr. Tichavský served as Associate Editor of the IEEE SIGNAL PROCESSING LETTERS from 2002 to 2004, and as Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2005 to 2009 and from 2011 to the present. Since 2009, he is a member of the IEEE Signal Processing Society's Signal Processing Theory and Methods (SPTM) Technical Committee. He has also served as a General Co-Chair of the 36th IEEE Int. Conference on Acoustics, Speech and Signal Processing ICASSP 2011 in Prague.



Anh Huy Phan received the Master's degree from Hochiminh University of Technology, Vietnam, in 2005, and the Ph.D. degree from the Kyushu Institute of Technology, Japan, in 2011.

He worked as Deputy head of Research and Development Department, Broadcast Research and Application Center, Vietnam Television, and is currently a Research Scientist at the Laboratory for Advanced Brain Signal Processing, and a Visiting Research Scientist in Toyota Collaboration Center, Brain Science Institute, RIKEN. His research interests include multi-

linear algebra, tensor computation, blind source separation, and brain computer interface.



Zbyněk Koldovský (S'03–M'04) was born in Jablonec nad Nisou, Czech Republic, in 1979. He received the M.S. and Ph.D. degrees in mathematical modeling from the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague, in 2002 and 2006, respectively.

He is currently an Associate Professor with the Institute of Information Technology and Electronics, Technical University of Liberec. He has also been with the Institute of Information Theory and Automation, Academy of Sciences of the Czech

Republic, since 2002. His main research interests are focused on audio signal processing, blind source separation, statistical signal processing, and multi-linear algebra.