Teze 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

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

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

where 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. $\forall\theta \in \Theta; \forall m = 1, \ldots, M;$ $\qquad 0 = \frac{\partial}{\partial\theta_m} \int f_\theta(Y)dY = \int \frac{\partial f_\theta(Y)}{\partial\theta_m} dY$

3. $\forall\theta \in \Theta; \forall m = 1, \ldots, M;$ $\frac{\partial}{\partial\theta_m}\int \hat{\theta}(Y) f_\theta(Y) dY = \int \hat{\theta}(Y)\frac{\partial f_\theta(Y)}{\partial\theta_m} dY$

4. $F(\theta)$ in (1) exists and is *invertible* .

Then, the celebrated Cramér-Rao inequality holds,

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

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

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

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

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

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 is 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,

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

obeys the condition

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

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. $\ldots y_{n-2}, y_{n-1}, y_n$. 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 $\ldots y_{n-2}, y_{n-1}, y_n$, 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 distribution [1]. The task is to find 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 $d_1 \times d_2 \times \ldots \times d_N$, where N is called the tensor order. Each element of the tensor has N indices, say $t_{i_1,\ldots i_N}$. 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) $\mathbf{A}_j$, $j = 1, \ldots, N$ of the size $d_j \times R$ with elements $a_{j,i,r}$, $i = 1, \ldots, d_j$, $r = 1, \ldots R$, and R scalars

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

$\lambda_1, \ldots, \lambda_R$ such that

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

for all $i_j = 1, \ldots, d_j$, $j = 1, \ldots N$. 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 of CANDECOMP, 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.