

AN ALGORITHM TO IMPROVE AND VALIDATE CRAMÉR-RAO BOUNDS THROUGH THE FIBRE BUNDLE THEORY OF LOCAL EXPONENTIAL FAMILIES

Justin Dauwels

RIKEN Brain Science Institute
Amari Research Unit
Hirosawa 2-1, Wako-shi, Saitama, Japan
Email: justin@dauwels.com

ABSTRACT

An algorithm is proposed that can be used (i) to verify the validity of the asymptotic Cramér-Rao bound (CRB) for a finite number of observations; (ii) to improve the CRB. It is derived from the fibre bundle theory of local exponential families developed by Amari and by Barndorff-Nielsen et al. The algorithm is applicable to any regular estimation problem. As an illustration, the problem of estimating the parameters of an AR model with noisy observations is considered.

Index Terms— Asymptotic Estimation Theory, Exponential Family, Cramér-Rao Bound, Factor Graph, Message Passing

1. INTRODUCTION

The (asymptotic) Cramér-Rao bound (CRB) is a widely used lower bound on the mean square estimation error (MSE) [1]. It has been computed in a wide variety of contexts ranging from digital communications to signal, speech, and image processing, to computational neuroscience. In this paper, we address two questions:

1. how can the (asymptotic) CRB be improved?
2. how can one verify whether the (asymptotic) CRB is valid for a given (finite) number of observations and a given signal-to-noise ratio?

The CRB is well known to be often loose at low signal-to-noise ratios (SNR) and/or for a small number of observations; this explains our interest in the first question. In order to motivate the second question, we remind the reader of some well known facts. For many estimation problems, the CRB is only valid in the limit of an infinite number of observations (asymptotic CRB). Obviously, the number of observations is finite in practice, and often researchers apply the asymptotic CRB while silently assuming that the number of observations is sufficiently large. In practice, it is important to know for which number of observations and in which SNR region asymptotic CRBs are valid.

In earlier work [2], we have shown how the validity of asymptotic CRBs can be verified by means of higher-order asymptotic estimation theory, which is a theory that leads to systematic improvements of the asymptotic CRB (see, e.g., [3, pp. 81–100] and references therein). An alternative approach based on the Barankin bound is presented in [4]. In [2], we applied the higher-order asymptotic estimation theory for curved exponential families to the problem of phase and variance estimation. We obtained higher-order corrections on the asymptotic CRB, and we determined the SNR region in which the asymptotic CRB and its

corrections are valid. For the (toy) problem of joint phase and variance estimation, the higher-order bounds are tractable. For less trivial estimation problems (especially the ones that involve hidden random variables), however, this is usually not the case. In this paper, we propose an algorithm to compute intractable higher-order corrections on the asymptotic CRB. The algorithm is applicable to general (non-singular) statistical models, i.e., not only to curved exponential families (as in [2]); the algorithm remains practical for high-dimensional statistical models with hidden random variables. It is a natural extension of the method we proposed in [5] to compute intractable CRBs. The algorithm is based on the fibre bundle theory of local exponential families developed by Amari [6] and by Barndorff-Nielsen et al. [8], which is an extension of the higher-order estimation theory for curved exponential families.

This paper is structured as follows. First, we review higher-order estimation theory. In Section 3, we present our algorithm to compute higher-order extensions of Cramér-Rao bounds. In Section 4, we apply that algorithm to a generic state space model; in Section 5, we provide numerical results for a particular state space model, i.e., an autoregressive (AR) process with noisy observations.

2. REVIEW OF ASYMPTOTIC ESTIMATION THEORY

The asymptotic theory of estimation deals with the performance of estimators in the limit of an infinite number of observations. We first briefly review first-order asymptotic estimation theory, whose main result is the standard (asymptotic) CRB. Then we outline higher-order asymptotic estimation theory; we will closely follow [3]. First we consider i.i.d. observations, then non-i.i.d. observations (as in time series).

2.1. Independently Identically Distributed Observations

2.1.1. First-order asymptotic estimation theory

Let $p(y; u)$ be a statistical model with random variable Y and parameters $u = u_1, \dots, u_M$ with $u_i \in \mathbb{R}$ (for $i = 1, \dots, M$). We consider the problem of estimating the parameter vector u from N i.i.d. samples y_1, \dots, y_N of $p(y; u)$. Let the function $\hat{u}(y^N)$ be an estimator of u based on the observations $y^N \triangleq y_1, \dots, y_N$. We define the error matrix $E^{(N)}(u)$ as:

$$E^{(N)}(u) \triangleq \mathbb{E}_{Y^N; U} [(\hat{u}(Y^N) - u)(\hat{u}(Y^N) - u)^T]. \quad (1)$$

A key player in first-order estimation theory is the Fisher information matrix $F(u)$ given by [1]:

$$F(u) \triangleq \mathbb{E}_{Y;U} [\nabla_u \ell(Y; u) \nabla_u^T \ell(Y; u)], \quad (2)$$

where $\ell(y; u) \triangleq \log p(y; u)$, and the expectation is with respect to $p(y; u)$. The following inequality holds (under certain regularity conditions) for any *consistent* estimator (which in general may be biased):

$$\lim_{N \rightarrow \infty} N E^{(N)}(u) \succeq F(u)^{-1}. \quad (3)$$

The inequality (3) (“asymptotic Cramér-Rao bound”) is the main result of first-order asymptotic estimation theory [1].

2.1.2. Higher-order asymptotic estimation theory

Higher-order asymptotic estimation theory deals with expansions of the error matrix $E^{(N)}(u)$ as power series of $N^{-1/2}$:

$$E^{(N)}(u) = N^{-1} M^{(1)}(u) + N^{-3/2} M^{(2)}(u) + N^{-2} M^{(3)}(u) + \mathcal{O}(N^{-5/2}). \quad (4)$$

The matrices $M^{(i)}(u)$ ($i = 1, 2, \dots, 3$) depend on the estimator $\hat{u}(y_1^N)$. A consistent estimator is said to be first-order efficient when its first-order term $M^{(1)}(u)$ is minimal at all u among all other consistent estimators. In a similar fashion, one defines higher-order efficiency. Interestingly, a higher-order efficient estimator is not only optimal with respect to the MSE but with respect to any monotonic loss function [6]. For a first-order efficient estimator, the matrix $M^{(1)}(u)$ is given by the inverse of the Fisher information matrix $F(u)$ (cf. (3)). Expressions for the minimal higher-order matrices $M_{\min}^{(i)}(u)$ for $i = 2, 3$ have been derived first for curved exponential families [3] and later for general probabilistic models, based on the fibre bundle theory of local exponential families [6] [8].

The curved exponential family is a natural extension of the exponential family. The latter is defined as the set \mathcal{S} of probability models that can be written in the form

$$p(y; \theta) \triangleq \exp \left[C(y) + \sum_{i=1}^L \theta^i F_i(y) - \psi(\theta) \right], \quad (5)$$

where $\theta \in \mathbb{R}^L$, $C(y)$ and $F_i(y)$ are arbitrary functions of Y , and $\psi(\theta)$ ensures that $p(y; \theta)$ is normalized [3]. Many common probability models are exponential families, e.g., the univariate and multivariate normal distribution, the Poisson distribution, and the binomial distribution. A curved exponential family is a set of probability distributions that is smoothly embedded in an exponential family [3]. Let \mathcal{M} be a curved exponential family with coordinate system $u \in \mathbb{R}^M$; the components of u are denoted by u_a ($a = 1, \dots, M$). Since the probability distribution parameterized by u also belongs to the exponential family \mathcal{S} , we may write their coordinates as $\theta = \theta(u)$. The probability distributions in \mathcal{M} may be written as

$$p(y; u) \triangleq \exp \left[C(y) + \sum_{i=1}^L \theta^i(u) F_i(y) - \psi(\theta(u)) \right]. \quad (6)$$

In the following, we will use a new parametrization (“coordinate system”) $w = (u, v)$ for \mathcal{S} , where the mapping $\theta(w) \triangleq$

$\theta(u, v)$ is supposed to be differentiable. The coordinates v_κ ($\kappa = 1, \dots, L - M$) are chosen such that $\theta(u) = (u, 0)$.

It has been shown that $M_{\min}^{(2)}(u)$ is zero for any model $p(y; u)$, in other words, first-order efficiency implies second-order efficiency (but not third-order efficiency) [3]. For curved exponential families, the minimal third-order matrix $M_{3, \min}(u)$ may be decomposed as a sum of two positive semi-definite matrices [3, p. 95]:

$$M_{3, \min}(u) \triangleq \frac{1}{2} (G_M(u) + 2 G_E(u)), \quad (7)$$

with

$$[G_M(u)]_{ab} \triangleq \sum_{cdef} [H_M(u)]_{acd} [H_M(u)]_{bef} \cdot [F(u)^{-1}]_{ce} [F(u)^{-1}]_{df}, \quad (8)$$

$$[G_E(u)]_{ab} \triangleq \sum_{\kappa \lambda cdef} [H_E(u)]_{\kappa ce} [H_E(u)]_{\lambda df} [F(u, 0)]_{\kappa \lambda} \cdot [F(u)^{-1}]_{cd} [F(u)^{-1}]_{ea} [F(u)^{-1}]_{fb}, \quad (9)$$

where

$$[H_M(u)]_{acd} \triangleq \sum_b [F_M(u)]_{cdb} [F(u)^{-1}]_{ba}, \quad (10)$$

$$[H_E(u)]_{\kappa ce} \triangleq \sum_\lambda [F_E(u)]_{ce\lambda} [F(u, 0)^{-1}]_{\lambda \kappa}, \quad (11)$$

and

$$[F_M(u)]_{cdb} \triangleq \mathbb{E}_{Y;U} [\partial_{u_c} \ell(Y; u) \partial_{u_d} \ell(Y; u) \partial_{u_b} \ell(Y; u)] + \mathbb{E}_{Y;U} [\partial_{u_c} \partial_{u_d} \ell(Y; u) \partial_{u_b} \ell(Y; u)], \quad (12)$$

$$[F_E(u)]_{ce\lambda} \triangleq \mathbb{E}_{Y;W} [\partial_{u_c} \partial_{u_e} \ell(Y; w) \partial_{v_\lambda} \ell(Y; w)] \Big|_{w=(u,0)}. \quad (13)$$

The higher-order asymptotic theory for general models $p(y; u)$ is derived by first approximating the model $p(y; u)$ by a curved exponential family in the neighborhood of some point $\tilde{u} \in \mathbb{R}^M$. Next, one applies the theory of curved exponential families (cf. (7)–(13)), which leads to the following results [6] [8]: the higher-order matrix $M_{3, \min}(u)$ has again the form (7), and the term G_M in (7) is again given by (8), i.e., the expression (8) applies to general probabilistic models; the term G_E is obtained by applying the expression (9) to the curved-exponential approximation, amounting to:

$$[G_E(u)]_{ab} = \sum_{cdgh} [F(u)^{-1}]_{cd} [F(u)^{-1}]_{ag} [F(u)^{-1}]_{bh} \cdot \left([J_E(u)]_{gchd} - [F(u)]_{gc} [F(u)]_{hd} - \sum_{ef} [F_E(u)]_{gce} [F(u)^{-1}]_{ef} [F_E(u)]_{hdf} \right), \quad (14)$$

where

$$[F_E(u)]_{abc} \triangleq \mathbb{E}_{Y;U} [\partial_{u_a} \partial_{u_b} \ell(Y; u) \partial_{u_c} \ell(Y; u)] \quad (15)$$

$$[J_E(u)]_{abcd} \triangleq \mathbb{E}_{Y;U} [\partial_{u_a} \partial_{u_b} \ell(Y; u) \partial_{u_c} \partial_{u_d} \ell(Y; u)]. \quad (16)$$

If $p(y; u)$ is a curved exponential family, the expression (14) reduces to (9).

2.2. Non-i.i.d. Observations

So far, we have assumed that the observations y_1, \dots, y_N are i.i.d. An expansion similar to (4) has been derived in [9, Chapter 4] for non-i.i.d. observations (as in time series, see Section 4 and 5). That expansion involves certain limits as for example:

$$F_1(u) \triangleq \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{Y^N; U} [\nabla_u \ell(Y^N; u) \nabla_u^T \ell(Y^N; u)] \quad (17)$$

$$\triangleq \lim_{N \rightarrow \infty} \frac{1}{N} F^{(N)}(u). \quad (18)$$

Limits as (17) are hard to compute (even numerically). Therefore, we consider here an alternative expansion that does not involve such limits and can be computed efficiently:

$$E_{\min}^{(N)}(u) = M_{\min}^{(1, N)}(u) + M_{\min}^{(3, N)}(u) + \mathcal{O}(N^{-3}) \quad (19)$$

$$= [F^{(N)}(u)]^{-1} + \frac{1}{2} (G_M^{(N)}(u) + 2G_E^{(N)}(u)) + \mathcal{O}(N^{-3}), \quad (20)$$

where $F^{(N)}$, $G_M^{(N)}$, and $G_E^{(N)}$ are by obtained from (2), (8), and (14) respectively by replacing the loglikelihood $\ell(y; u)$ with $\ell(y^N; u)$.

3. GENERIC MONTE-CARLO ALGORITHM

We consider a stochastic model $p(x^N, y^N; u)$ with hidden random variables $X^N = X_1, \dots, X_N$, observed random variables $Y^N = Y_1, \dots, Y_N$, and parameters $u = u_1, \dots, u_M$. We wish to compute the terms $M_{\min}^{(i, N)}(u)$ ($i = 1, 3$) from the marginal $p(y^N; u)$ defined as

$$p(y^N; u) \triangleq \int_{x^N} p(x^N, y^N; u) dx^N. \quad (21)$$

We follow the same approach as in [5]. The two terms $M_{\min}^{(i, N)}(u)$ ($i = 1, 3$) involve expectations with respect to $p(y^N; u)$ (cf. (2)); we evaluate those expectations by Monte-Carlo simulations: we draw samples from $p(y^N; u)$ and compute the expectations by averaging over those samples. The two terms also involve the gradient $\nabla_u \ell(y^N; u)$ and Hessian $\nabla_u \nabla_u^T \ell(y^N; u)$. The former can be evaluated through the well-known equality

$$\nabla_u \ell(y^N; u) = \mathbb{E}_{X^N | Y^N; U} [\nabla_u \ell(X^N, y^N; u)], \quad (22)$$

where $\ell(x^N, y^N; u) \triangleq \log p(x^N, y^N; u)$. Also for computing the Hessian, we use the expression in the RHS of (22):

$$\nabla_u \nabla_u^T \ell(y^N; u) = \nabla_u \mathbb{E}_{X^N | Y^N; U} [\nabla_u \ell(X^N, y^N; u)]. \quad (23)$$

The expressions at the RHS of (22) and (23) can often be computed efficiently, as we will briefly outline in the following sections.

In summary, the terms $M_{1, \min}^{(N)}(u)$ and $M_{3, \min}^{(N)}(u)$ may be computed by the following generic Monte-Carlo algorithm:

1. Generate a list of n samples $\hat{y}^{(1)}, \hat{y}^{(2)}, \dots, \hat{y}^{(n)}$ from $p(y; u)$.
2. For $\ell = 1, 2, \dots, n$, evaluate the expressions

$$\mathbb{E}_{X^N | Y^N; U} [\nabla_u \ell(X^N, \hat{y}^{(\ell)}; u)] \quad (24)$$

and

$$\nabla_u \mathbb{E}_{X^N | Y^N; U} [\nabla_u \ell(X^N, \hat{y}^{(\ell)}; u)]. \quad (25)$$

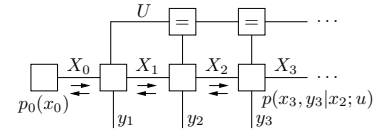


Fig. 1. Factor graph of the state space model (30).

3. Compute the estimate $\hat{F}^{(N)}(u)$ by averaging over the samples $\hat{y}^{(1)}, \hat{y}^{(2)}, \dots, \hat{y}^{(n)}$:

$$\hat{F}^{(N)}(u) \triangleq \frac{1}{n} \sum_{j=1}^n \left[\mathbb{E}_{X^N | Y^N; U} [\nabla_u \ell(X^N, \hat{y}^{(j)}; u)] \cdot \mathbb{E}_{X^N | Y^N; U} [\nabla_u \ell(X^N, \hat{y}^{(j)}; u)]^T \right]. \quad (26)$$

Similarly, compute estimates $\hat{F}_M^{(N)}(u)$, $\hat{F}_E^{(N)}(u)$, and $\hat{J}_E^{(N)}(u)$ of $F_M^{(N)}(u)$, $F_E^{(N)}(u)$ and $J_E^{(N)}(u)$ respectively.

4. Compute the estimate

$$[\hat{H}_M^{(N)}(u)]_{acd} \triangleq \sum_b [\hat{F}_M^{(N)}(u)]_{cdb} [\hat{F}^{(N)}(u)^{-1}]_{ba}. \quad (27)$$

5. Conclude with the estimates

$$\hat{M}_{1, \min}^{(N)}(u) \triangleq [\hat{F}^{(N)}(u)]^{-1} \quad (28)$$

and

$$\hat{M}_{3, \min}^{(N)}(u) \triangleq \frac{1}{2} (\hat{G}_M^{(N)}(u) + 2\hat{G}_E^{(N)}(u)), \quad (29)$$

where $\hat{G}_M^{(N)}(u)$ and $\hat{G}_E^{(N)}(u)$ are by obtained from (8) and (14) respectively by replacing F , H_M , F_E , and J_E with $\hat{F}^{(N)}$, $\hat{H}_M^{(N)}$, $\hat{F}_E^{(N)}$, and $\hat{J}_E^{(N)}$ respectively.

4. STATE SPACE MODEL

As an illustration, we apply the above algorithm to the (generic) parameterized state space model

$$p(x^N, y^N; u) \triangleq p_0(x_0) \prod_{k=1}^N p(x_k, y_k | x_{k-1}; u), \quad (30)$$

with parameter vector $u \in \mathbb{R}^M$. In many important applications (see, e.g., Section 5), the model (30) is a curved exponential family. The marginal distribution $p(y^N; u)$, however, is most often neither an exponential family nor a curved exponential family. Fig. 1 shows a factor graph of (30) (each node corresponds to factor, each edge to a variable; we refer to [10] for an introduction to factor graphs). Plugging the model (30) into the conditional expectation (24) amounts to:

$$\begin{aligned} & \mathbb{E}_{X^N | Y^N; U} [\nabla_u \ell(X^N, y^N; u)] \\ &= \sum_{k=1}^N \mathbb{E}_{X_{k-1} X_k | Y^N; U} [\nabla_u \ell(X_k, y_k | X_{k-1}; u)], \end{aligned} \quad (31)$$

where the expectations in the RHS are with respect to the marginal distributions $p(x_k, x_{k-1} | y; u)$. Those marginals may be obtained as [10] (see Fig. 1):

$$p(x_k, x_{k-1} | y^N; u) \propto \bar{\mu}(x_{k-1}) \bar{\mu}(x_k) p(x_k, y_k | x_{k-1}; u), \quad (32)$$

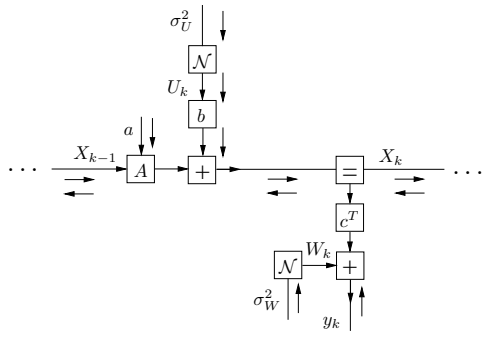


Fig. 2. Factor graph of state space model (37) (38).

where the messages $\vec{\mu}(x_k)$ (rightward arrows in Fig. 1) are computed by the sum-product forward recursion

$$\vec{\mu}(x_k) \propto \int_{x_{k-1}} \vec{\mu}(x_{k-1}) p(x_k, y_k | x_{k-1}; u) dx_{k-1}, \quad (33)$$

with $\mu_0(x_0) \triangleq p_0(x_0)$, and similarly, the messages $\overleftarrow{\mu}(x_k)$ (leftward arrows in Fig. 1) are determined by the sum-product backward recursion

$$\overleftarrow{\mu}(x_k) \propto \int_{x_{k+1}} \overleftarrow{\mu}(x_{k+1}) p(x_{k+1}, y_{k+1} | x_k; u) dx_{k+1}, \quad (34)$$

with $\overleftarrow{\mu}(x_N) \triangleq 1$ [10]. The gradient (25) of the conditional expectation (24) involves besides the messages $\vec{\mu}(x_k)$ and $\overleftarrow{\mu}(x_k)$ also the messages $\nabla_u \vec{\mu}(x_k)$ and $\nabla_u \overleftarrow{\mu}(x_k)$; the latter are computed by recursions similar to (33) and (34) (see the longer report [11] for more details).

5. NUMERICAL EXAMPLE

As in [5], we consider the auto-regressive (AR) process

$$X_k = a_1 X_{k-1} + a_2 X_{k-2} + \dots + a_K X_{k-K} + U_k, \quad (35)$$

where a_1, \dots, a_K are unknown real parameters, and U_1, U_2, \dots are i.i.d. real zero-mean Gaussian random variables with variance σ_U^2 . We observe the real random variable Y_k given by

$$Y_k = X_k + W_k, \quad (36)$$

where W_k are i.i.d. real zero-mean Gaussian random variables with variance σ_W^2 . From the observation $y = (y_1, \dots, y_N)$, one wishes to jointly estimate the coefficients a and the variances σ_U^2 and σ_W^2 . The expressions (35) and (36) can be rewritten as the state space model:

$$X_k = AX_{k-1} + bN_k \quad (37)$$

$$Y_k = c^T X_k + W_k, \quad (38)$$

where

$$X_k \triangleq [Z_k, \dots, Z_{k-K+1}]^T \quad (39)$$

$$A \triangleq \begin{bmatrix} a^T \\ I_{K-1} & 0_{K-1} \end{bmatrix} \quad (40)$$

$$b \triangleq c \triangleq [1 \quad 0_{K-1}^T]^T \quad (41)$$

$$a \triangleq [a_1, \dots, a_K]^T, \quad (42)$$

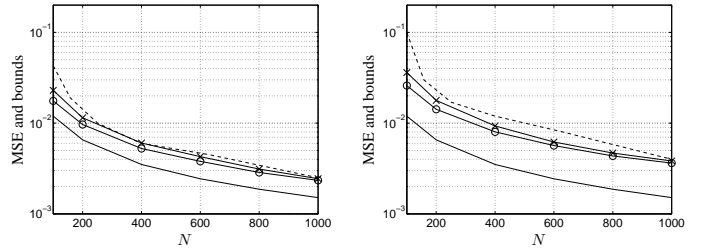


Fig. 3. MSE and bounds for coefficient estimation.

where I_n is the $n \times n$ identity matrix, and 0_n is a zero (column) vector of dimension n .

The factor graph of Fig. 2 represents the model (37) (38); it has been used in [12] to derive various algorithms for estimating the parameters $u \triangleq (a_1, \dots, a_K, \sigma_U^2, \sigma_W^2)$. The same graph can be used to compute the conditional expectation (24) and its gradient (25): the recursions (33) and (34) (indicated by the arrows in Fig. 2) are equivalent to Kalman recursions for mean vectors and covariance matrices [10]; the messages $\nabla_u \vec{\mu}(x_k)$ and $\nabla_u \overleftarrow{\mu}(x_k)$ are computed by similar Kalman recursions [11].

We have computed the CRB $M_{\min}^{(1,N)}(u)$ and its higher-order correction $M_{\min}^{(3,N)}(u)$ (cf. (19)) for the estimation of a , σ_U^2 and σ_W^2 . Fig. 3 summarizes the results for the coefficients a ; it shows: (i) the MSE of the linear predictive coding algorithm [13] (solid; left and right) with known $\sigma_N^2 = 0.1$ and $\sigma_W^2 = 0$; (ii) MSE of grid-based algorithm of [12] (dashed) with unknown $\sigma_N^2 = 0.1$ and unknown $\sigma_W^2 = 0.001$ (left) and $\sigma_W^2 = 0.01$ (right); (iii) CRB $M_{\min}^{(1,N)}(u)$ (circles) with unknown $\sigma_N^2 = 0.1$ and unknown $\sigma_W^2 = 0.001$ (left) and $\sigma_W^2 = 0.01$ (right); (iv) third-order bound (19) (crosses) with unknown $\sigma_N^2 = 0.1$ and $\sigma_W^2 = 0.001$ (left) and $\sigma_W^2 = 0.01$ (right). As can be seen from Fig. 3, the correction term $M_{\min}^{(3,N)}$ is significantly smaller than $M_{\min}^{(1,N)}$ (at most 30% of $M_{\min}^{(1,N)}$), and the third-order asymptotic expansion (19) can therefore be considered valid. Moreover, since the fifth- and higher-order corrections are likely to be negligible, the expansion (19) is expected to practically coincide with the minimum achievable MSE.

6. REFERENCES

- [1] M. J. Schervish, *Theory of Statistics*, Springer Verlag, 1995.
- [2] J. Dauwels, "Improving and validating Cramér-Rao Bounds Through Higher-Order Asymptotic Estimation Theory: a Case Study," *Proc. 2006 IEEE Int. Symp. on Information Theory and its Applications*, Seoul, Korea, Oct. 29–Nov. 1, to appear.
- [3] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Oxford university Press, 2000.
- [4] P. M. Schultheiss and A. Zeira, "Realizable Bounds for Time Delay Estimation: Part 2 - Threshold Phenomena," *IEEE Trans. Signal Processing*, vol. 42, May 1994, pp. 1001–1007.
- [5] J. Dauwels and S. Korf, "A numerical method to compute Cramér-Rao-type bounds for challenging estimation problems," *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006)*, Toulouse, France, May 15–19, 2006, pp. 717–720.
- [6] S. Amari, "Differential-Geometrical Theory of Statistics," Chapter II in [7], pp. 19–94, 1987.
- [7] S. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen, and C. R. Rao, *Differential Geometry in Statistical Inference*, IMS Lecture Notes: Monograph Series 10, Institute of Mathematical Statistics, Hayward, California, 1987.
- [8] O. E. Barndorff-Nielsen and P. E. Jupp, "Approximating Exponential Models," *Annals of the Institute of Statistical Mathematics*, 41:247–267, 1989.
- [9] M. Taniguchi and Y. Kakizawa, *Asymptotic Theory of Statistical Inference for Time Series*, Springer Series in Statistics, 2000.
- [10] H.-A. Loeliger, "An Introduction to Factor Graphs," *IEEE Signal Processing Magazine*, Jan. 2004, pp. 28–41.
- [11] J. Dauwels, "A Message-Passing Algorithm to Improve and Validate Cramér-Rao Bounds Through the Fibre Bundle Theory of Local Exponential Families," in preparation.
- [12] S. Korf, H.-A. Loeliger, A.-G. Lindgren, "AR Model Parameter Estimation: From Factor Graphs to Algorithms," *Proc. ICASSP'04*, vol. 5, Montreal, Canada, May 17–21, 2004, pp. 509–512.
- [13] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 2002.