Systems & Control Letters 12 (1989) 281–286
North-Holland

Rate distortion lower bound for a special class of nonlinear estimation problems

R.B. WASHBURN, Jr.
ALPHATECH, Inc., 111 Middlesex Turnpike, Burlington, MA 01803-4901, U.S.A.

D. TENEKETZIS
Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122, U.S.A.

Received 3 October 1988

Abstract: A rate distortion lower bound of minimum mean square error is presented for a special class of discrete time nonlinear filtering problems which have measurements that can be expressed as a memoryless nonlinear function of a Gaussian distributed space process with additive Gaussian noise. The lower bound is exactly and practically computable for a large class of nonlinearities and is proved to be asymptotically tighter than Cramér–Rao type bounds in the limit of low signal-to-noise ratio.

Keywords: Nonlinear estimation, Minimum mean square error, Rate distortion bound, Cramér–Rao bound.

1. Introduction

The Cramér–Rao method [1–4] is an example of a successful approach to bounding optimal performance in nonlinear estimation problems. The approach applies to a large class of problems, it can be computed exactly and is relatively easy to compute in many cases, and it gives practically useful performance estimates in many cases. However, the method is not perfect. The Cramér–Rao method tends to underestimate the minimum square error in nonlinear estimation problems with low signal-to-noise ratio (SNR). This defect of the Cramér–Rao approach has led many scientists to investigate other techniques for estimating nonlinear estimation performance.

Several previous authors [5–12] have studied the use of rate distortion theory or related information theory methods to analyze the average error in estimation problems. This previous work either makes restrictive assumptions about the estimation problem or requires difficult computations (such as Monte Carlo simulation) to obtain a lower bound on the estimation error. This paper studies a simple rate distortion approach to analyze mean square error for a special class of nonlinear estimation problems in which the state is Gaussian distributed and the measurement is a memoryless nonlinear function of the state with an additive Gaussian white noise. For this special class of estimation problems we have obtained a rate distortion lower bound on the mean square error in estimating any component of the state. This bound is exactly and practically computable for a large class of measurement nonlinearities (including polynomial, exponential, and trigonometric functions), and the bound is tighter (i.e., larger) than Cramér–Rao type bounds for the class of problems where the measurements depend nonlinearly on the component of interest and the noise covariance is sufficiently large.

The paper is organized as follows. Section 2 gives the necessary rate distortion theory and presents the lower bound for the class of nonlinear estimation problems described above. In Section 3 we theoretically analyze the lower bound and compare it to Cramér–Rao type lower bounds for the same class of problems. Section 4 presents a discrete time nonlinear filtering example to illustrate the application of the bound.

2. Rate distortion inequality

Consider the problem of estimating a Gaussian random variable x in $\mathbb{R}^N$ given the measurement y in $\mathbb{R}^M$ for which

$$y = C(x) + v. \tag{1}$$

Here v is a Gaussian random variable in $\mathbb{R}^M$ and is independent of x, and C may be a nonlinear function of x. Let $C(x)$ have a covariance matrix

0167-6911/89/$3.50 © 1989, Elsevier Science Publishers B.V. (North-Holland)

Γ and let v have a covariance matrix R. If P_{11} is the variance of the component x_1 of x, then we will show

$$E\{[x_1 - \hat{x}_1(y)]^2\} \geq P_{11} \frac{\det(R)}{\det(\Gamma + R)} \tag{2}$$

for any estimator $\hat{x}_1(y)$ of x_1 which depends only on y. The right side of the inequality (2) is our rate distortion lower bound (RDB). Section 3 compares this bound to the corresponding Cramér–Rao bound for nonlinear estimation problems and Section 4 discusses how to compute the RDB efficiently and presents a discrete time nonlinear filtering example to illustrate the method. In the remainder of this section we prove inequality (2) using rate distortion theory.

Proof of Inequality (2). Let D denote the minimum mean square error of estimating x_1 given y. Rate distortion theory [13] tells us that D is related to the mutual information $I(x_1;\, y)$ between x_1 and y by

$$R(D) \leq I(x_1;\, y)$$

where $R(D)$ is the rate distortion function of the Gaussian random variable x_1. If P_{11} denotes the variance of x_1, then $R(D)$ is given by

$$R(D) = \frac{1}{2}\log\left[\frac{P_{11}}{D}\right].$$

We cannot compute $I(x_1;\, y)$ exactly for general nonlinear C, but we can approximate it as follows. Since x_1 is a function (i.e., component) of x, we have

$$I(x_1;\, y) \leq I(x;\, y).$$

We can express the mutual information $I(x;\, y)$ as the difference

$$I(x;\, y) = h(y) - h(y \mid x)$$

of the differential entropy $h(y)$ and the conditional differential entropy $h(y \mid x)$. We can compute $h(y \mid x)$ exactly as

$$h(y \mid x) = h(v) = \tfrac{1}{2}M \log\left(2\pi e[\det R]^{1/M}\right)$$

where M is the dimension of the measurement vector and R is the covariance of the measurement noise v. We cannot compute $h(y)$ in general, but we can bound it as follows [13]:

$$h(y) \leq \tfrac{1}{2}M \log\left(2\pi e[\operatorname{cov}(y)]^{1/M}\right)$$

where $\operatorname{cov}(y)$ denotes the covariance matrix of the measurement vector y. Noting that $\operatorname{cov}(y) = \Gamma + R$ and combining the results above gives the inequality (2). □

3. Theoretical comparison of rate distortion and Cramér–Rao bounds

The Cramér–Rao lower bound (CRB) for the estimation problem of (1) is given by

$$E\{[x - \hat{x}(y)][x - \hat{x}(y)]^{\mathrm{T}}\} \geq [J + P^{-1}]^{-1}, \tag{3}$$

see [1], where P is the covariance of x, and J is the information matrix defined by

$$J = E\{C'(x)^{\mathrm{T}} R^{-1} C'(x)\}, \tag{4}$$

in which R is the covariance of the noise v in (2), $C'(x)$ is the derivative of $C(x)$ with respect to x, and the superscript T denotes matrix or vector transposition. The following result shows that in the case of scalar state and measurement, it is always true that RDB is larger than CRB with equality only in the case of linear functions $C(x)$.

Comparison of RDB and CRB in scalar case. *If $C(x)$ is continuously differentiable in x, and the expectations $E\{C(x)^2\}$ and $E\{C'(x)^2\}$ are both finite, then* CRB $\leq$ RDB *with equality if and only if $C(x) = ax + b$ for some constants a, b.*

Proof. Assume first that x has 0 mean and variance P, and define

$$\Phi(x) = [C(x) - C(0)]^2 x^{-1} \exp(-x^2/2P)$$

for $x \neq 0$, and define $\Phi(0) = 0$. Note that Φ is continuously differentiable and the derivative is given by

$$\Phi'(x) = \{2[C(x) - C(0)]C'(x)x^{-1} - [C(x) - C(0)]^2 x^{-2} - [C(x) - C(0)]^2/P\} \exp(-x^2/2P)$$

for $x \neq 0$ and $\Phi'(0) = C'(0)^2$. Given that both

$E\{C(x)^2\}$ and $E\{C'(x)^2\}$ are finite, we have that

$$\int_{-\infty}^{\infty} \Phi'(x)\,\mathrm{d}x = \lim_{x\to\infty} [\Phi(x) - \Phi(-x)] = 0$$

which gives

$$E\{C'(x)^2\} = P^{-1}E\{[C(x)-C(0)]^2\} + E\{[C'(x) - (C(x)-C(0))/x]^2\}.$$

Let $\bar{C} = E\{C(x)\}$. Then

$$E\{C'(x)^2\} = P^{-1}E\{[C(x)-\bar{C}]^2\} + P^{-1}[\bar{C} - C(0)]^2 + E\{[C'(x) - (C(x)-C(0))/x]^2\}$$

and it follows that

$$E\{C'(x)^2\} \geq P^{-1}E\{[C(x)-\bar{C}]^2\} \tag{5}$$

with equality if and only if $\bar{C} = C(0)$ and

$$C'(x) = [C(x) - C(0)]/x.$$

This is possible if and only if $C(x) = ax + b$ for constants a, b.

Let Γ denote the covariance of $C(x)$ and let J be as in (4). Then we have proven

$$J \cdot R \geq \Gamma \cdot P^{-1}.$$

Since

$$\mathrm{CRB} = [J + P^{-1}]^{-1}$$

and

$$\mathrm{RDB} = [\Gamma P^{-1} R^{-1} + P^{-1}]^{-1},$$

this proves that CRB $\leq$ RDB, at least for the case of 0 mean. The general case follows easily. □

Simple examples show that the relation CRB $\leq$ RDB is not true in the general vector case. However, this relation is true if the signal-to-noise ratio is sufficiently low and the measurement is nonlinear in the component of interest. More precisely, consider the model of (1) and assume (without loss of generality) that the covariance matrix R of the noise is a constant multiple r of the M dimensional identity matrix I_M. Consider the CRB of equation (3), (4) as a function of r. Let CRB(r) denote the 1, 1 component of this matrix which is the Cramér–Rao lower bound of the mean square error for estimating x_1. Let RDB(r) be the analogous rate distortion bound of equation (2). Then we prove the following result.

Low SNR comparison ($r \to \infty$). *If C is continuously differentiable and if C and its partial derivatives have finite second moments, then*

$$\mathrm{CRB}(r) \leq \mathrm{RDB}(r) + \mathrm{O}(r^{-2})$$

where $\mathrm{O}(r^{-2})$ is a term that vanishes like r^{-2} as $r \to \infty$. Furthermore, for sufficiently large r,

$$\mathrm{CRB}(r) < \mathrm{RDB}(r),$$

unless there is a component x_2 of x which is independent of x_1 and such that

$$C(x) = a(x_2)x_1 + b(x_2)$$

for some functions a and b of x_2.

Proof. The proof of this result follows from the expansion of CRB(r) and RDB(r) in powers of r^{-1} and the scalar inequality (5). First consider the expansion of CRB(r). Let Γ_C denote

$$\Gamma_C = E\{C'(x)^{\mathrm{T}} C'(x)\}.$$

The matrix lower bound of equation (3) has the expansion

$$[r^{-1}\Gamma_C + P^{-1}]^{-1} = P \cdot [I_N + r^{-1}\Gamma_C \cdot P]^{-1} = P - r^{-1}P \cdot \Gamma_C \cdot P + \mathrm{O}(r^{-2})$$

where I_N is the N dimensional identity matrix, N being the dimension of the state x, and $\mathrm{O}(r^{-2})$ denotes a generic term that vanishes as r^{-2} as $r \to \infty$. The 1, 1 component of this expansion of the matrix bound gives

$$\mathrm{CRB}(r) = P_{11} - r^{-1}(P_{11})^2 \sum_{i=1}^{M} E\{[C_{i,1}(x)]^2\} + \mathrm{O}(r^{-2}) \tag{6}$$

where $C_{i,1}$ denotes the partial derivative of C_i (the i-th component of C) with respect to x_1.

Now consider the expansion of RDB(r). Let Γ denote the covariance of $C(x)$. Then the right hand side of (2) can be expressed as

$$\mathrm{RDB}(r) = P_{11}\det([I_M + r^{-1}\Gamma]^{-1}) = P_{11}\det([I_M - r^{-1}\Gamma + \mathrm{O}(r^{-2})]).$$

Expanding the determinant above for large r gives

$$\mathrm{RDB}(r) = P_{11}\left(1 - r^{-1}\sum_{i=1}^{M} E\{[C_i(x) - E\{C_i(x)\}]^2\} + \mathrm{O}(r^{-2})\right). \quad (7)$$

Let P_1 be the column of the covariance matrix P corresponding to the component x_1, and define x_2 as

$$x_2 = x - P_{11}^{-1} \cdot P_1 \cdot x_1.$$

The random variables x_1 and x_2 are uncorrelated and therefore independent (because they are jointly Gaussian). In this case the variance of x_1, namely P_{11}, is also the conditional variance of x_1 given x_2, and the scalar inequality (5) proved above implies that

$$E\{[C_{i,1}(x)]^2 \mid x_2\} \geq P_{11}^{-1} E\{[C_i(x) - E\{C_i(x) \mid x_2\}]^2 \mid x_2\}. \quad (8)$$

for almost all values of x_2. The inequality is strict for a value x_2 unless C_i is a linear function of x_1, that is unless

$$C_i(x) = a_i(x_2)x_1 + b_i(x_2)$$

for all values of x_1.

Rearranging the right side of the inequality (8) gives

$$E\{[C_{i,1}(x)]^2 \mid x_2\} \geq P_{11}^{-1}\left(E\{[C_i(x)]^2 \mid x_2\} - [E\{C_i(x) \mid x_2\}]^2\right).$$

Averaging over x_2 and using Jensen's inequality gives

$$E\{[C_{i,1}(x)]^2\} \geq P_{11}^{-1}\left(E\{[C_i(x)]^2\} - [E\{C_i(x)\}]^2\right) = P_{11}^{-1} E\{[C_i(x) - E\{C_i(x)\}]^2\}.$$

Summing both sides of this inequality over i from 1 to M gives

$$\sum_{i=1}^{M} E\{[C_{i,1}(x)]^2\} \geq \sum_{i=1}^{M} P_{11}^{-1} E\{[C_i(x) - E\{C_i(x)\}]^2\}. \quad (9)$$

Equality occurs in (9) only if equality occurs in (8) for each i and almost all values of x_2. Thus, the inequality in (9) is strict unless $C(x)$ is a linear function of x_1, that is unless

$$C(x) = a(x_2)x_1 + b(x_2).$$

Comparing the inequality (9) to the asymptotic expansions (6) and (7), we see that the order $\mathrm{O}(r^{-1})$ term of RDB(r) is at least as large as the comparable term of CRB(r). Thus, CRB(r) can exceed RDB(r) only by the order $\mathrm{O}(r^{-2})$ terms, that is

$$\mathrm{CRB}(r) \leq \mathrm{RDB}(r) + \mathrm{O}(r^{-2}).$$

Furthermore, unless equality occurs in (9), then the $\mathrm{O}(r^{-1})$ term of RDB(r) is strictly larger than the corresponding term of CRB(r). In this case, RDB(r) must be strictly larger than CRB(r) if r is sufficiently large. □

How low SNR needs to be for the results above depends on the specific measurement nonlinearities considered. In practice, the rate distortion bound will significantly improve upon the Cramér–Rao type bound only when the measurement nonlinearities are significant. We will illustrate this by an example in the next section. However, in specific cases it is easiest to compute both bounds and choose the largest rather than try to predict which is better apriori. The next section explains how to compute the RDB efficiently for a large class of nonlinear estimation problems.

4. Computation of rate distortion bound

In this section we discuss how to compute the RDB efficiently for nonlinear estimation problems of the type defined in (1). The computational problem is one of computing the covariance Γ and then the determinants in (2). Computing Γ requires calculating the Gaussian expectations $E\{C(x)\}$ and $E\{C(x)C(x)^{\mathrm{T}}\}$. These expectations can be computed in closed form in terms of elementary functions if the nonlinearities in $C(x)$ are sums of products of polynomial, exponential, sine, and cosine functions of the components of x.

Computing the determinant of $\Gamma + R$ requires the order of M^3 operations and is usually the most difficult calculation in obtaining the bound

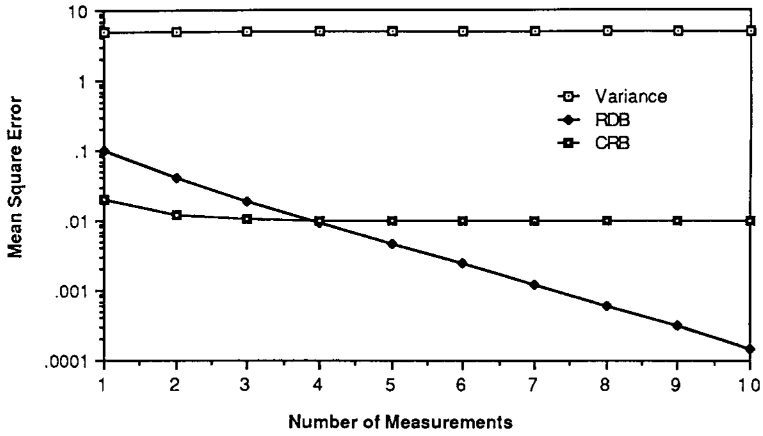

Fig. 1. Sine sensor, $r = 0.01$.

in (2). If the problem is derived from a discrete time filtering problem with measurement dimension m and time periods $s = 1, 2, \ldots, t$, then $M = mt$. In this case it is best to compute this determinant without pivoting as this allows one to obtain the bounds on the filter estimation error for each time $s = 1, \ldots, t$ using only one mt dimensional determinant calculation. Note that the lack of pivoting does not create a serious numerical problem in this case because the matrix $\Gamma + R$ is usually positive definite and well conditioned (e.g., R is large, positive definite, and block diagonal).

To illustrate the rate distortion bound and compare it to the Cramér–Rao bound, consider the scalar filtering problem defined by

$$x(t+1) = Ax(t) + w(t),$$
$$y(t) = \sin(x(t)) + v(t),$$

where $E\{x(1)\} = 0$ and w and v are Gaussian white noise processes with variance q and r respectively. To compute the rate distortion bound we need only the formula

$$E\{\sin(x)\,\sin(y)\} = \exp\left(-\tfrac{1}{2}\left[P_x + P_y\right]\right) \cdot \sinh(P_{xy})$$

for 0 mean, jointly Gaussian random variables x and y with variance and covariance P_x, P_y, and P_{xy}. For comparison, we also compute the Cramér–Rao bound for this class of problems, as described in [2] (see also [3,4]).

We fix $A = 1$, $q = 0.01$, and $E\{x(1)^2\} = 5$, and vary the noise variance r from 1 to 10^{-2}. Figures 1, 2, and 3 show the rate distortion bound and the Cramér–Rao bound versus the number t of measurements for values of r. In addition, we have plotted the variance of $x(t)$ in each figure as an upper bound on estimation error variance.

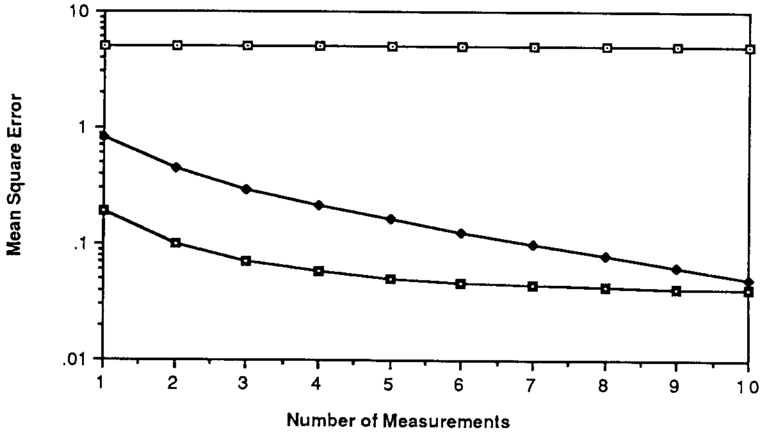

Fig. 2. sine Sensor, $r = 0.1$.

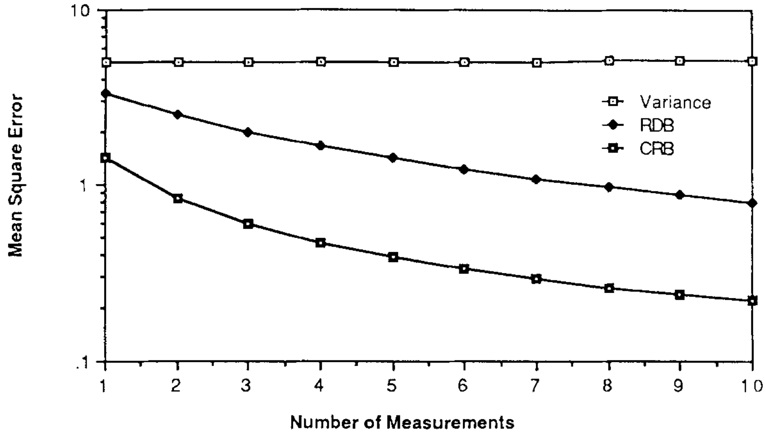

Fig. 3. Sine sensor, $r = 1.0$.

Note that the rate distortion bound in each figure is always greater than the Cramér–Rao bound at time period 1. This is a theoretical consequence of the results of Section 3. For vector states and measurements (in particular, for dynamic filtering problems) the rate distortion bound need not exceed the Cramér–Rao bound unless the signal-to-noise ratio is sufficiently low. In this example we see that a measurement variance of $r = 10^{-1}$ is required (Figure 2) before the rate distortion bound exceeds the Cramér–Rao bound at each time period between 1 and 10.

References

[1] H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I* (J. Wiley & Sons, New York, 1968).

[2] D.L. Snyder and I.B. Rhodes, Filtering and control performance bounds with implications on asymptotic separation, *Automatica* **8** (1972) 747–753.

[3] B.Z. Bobrovsky and M. Zakai, A lower bound on the estimation error for certain diffusion processes, *IEEE Trans. Inform. Theory* **22** (1976) 45–52.

[4] J.I. Galdos, A Cramér–Rao bound for multidimensional discrete-time dynamical systems, *IEEE Trans. Automat. Control* **25** (1980) 117–119.

[5] H.L. Weidemann and E.B. Stear, Entropy analysis of estimating systems, *IEEE Trans. Inform. Theory* **16** (1970) 264–270.

[6] L.P. Seidman, Performance limitations and error calculations for parameter estimation, *Proc. IEEE* **58** (1970) 644–652.

[7] H.L. Van Trees, *Detection, Estimation, and Modulation Theory: Part II, Nonlinear Modulation Theory* (J. Wiley & Sons, New York, 1968).

[8] M. Zakai and J. Ziv, Lower and upper bounds on the optimal filtering error of certain diffusion processes, *IEEE Trans. Inform. Theory* **18** (1972) 325–331.

[9] R.S. Bucy and J. Pages, A priori error bounds for the cubic sensor problems, *IEEE Trans. Automat. Control* **23** (1978) 88–91.

[10] R.S. Bucy, Information and filtering, *Inform. Sci.* **18** (1979) 179–187.

[11] J.I. Galdos, A lower bound on filtering error with application to phase demodulation, *IEEE Trans. Inform. Theory* **25** (1979) 452–462.

[12] J.I. Galdos, A rate distortion theory lower bound on desired function filtering error, *IEEE Trans. Inform. Theory* **27** (1981) 366–368.

[13] T. Berger, *Rate Distortion Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1971).