[6] Siouris, G. M. (1974)
Comparison between proportional and augmented proportional guidance.
Nachrichtentechnische Zeitschrift, (Jul. 1974), 278–280.

[7] Singer, R. A. (1970)
Estimating optimal tracking performance for manned maneuvering targets.
IEEE Transactions on Aerospace and Electronic Systems, **AES-6**, 4 (Jul. 1970), 473–483.

[8] Song, T. L. (1996)
Observability of target tracking with bearings-only measurements.
IEEE Transactions on Aerospace and Electronic Systems, **32**, 4 (Oct. 1996), 1468–1472.

[9] Smith, F. G. (1993)
Atmospheric Propogation of Radiation.
IR and EO Systems Handbook, Vol. 2.
SPIE Optical Engineering Press, 1993.

Performance Bounds for Nonlinear Filters

An error in a well-known lower bound on the performance of nonlinear filters based on dynamic statistical models is described. The error occurs in the final step of the proof and deletion of this step results in a valid lower bound which, however, requires evaluation of a conditional expectation in function space. For the limited, but interesting, class of linear Gaussian process models, computational methods are described.

I. INTRODUCTION

In the review by Kerr [1] of lower bounds on the performance of nonlinear filters based on dynamic statistical models, a rate-distortion bound due to Galdos [2] is prominently discussed (e.g., [1, pp. 596, 597]). However, the original paper [2] has an error which invalidates the bound. The purpose of this note is to point out the error, to point out that an intermediate result in [2] is correct and provides a bound but that the bound is difficult to evaluate because it requires the computation of expectations in function space, and to present methods for evaluating this new bound in the case where the process model is linear and Gaussian.

In a later paper concerning performance bounds based on rate-distortion bounds [3], Galdos considers bounds on functions of the state rather than just the state itself and uses Monte Carlo simulation in function space to evaluate the necessary conditional expectations. The Monte Carlo theory in [3] focuses on the case in which the process model is linear and Gaussian and, for the same reasons, when we discuss evaluation of the bound (Section III), we only consider the case in which the process model is linear and Gaussian.

Manuscript received June 8, 1995; revised April 20, 1996.

IEEE Log No. T-AES/33/1/01102.

This work was supported by U.S. National Science Foundation Grant MIP-9110919, a Whirlpool Faculty Fellowship, and the School of Electrical and Computer Engineering, Purdue University.

0018-9251/97/$10.00 © 1997 IEEE

II. THE BOUND

In [2], a lower bound on the minimum mean square error attainable by nonlinear filters using rate distortion theory and the Bucy–Mortensen–Duncan representation Theorem is presented. Theorems 2 (continuous time) and 3 (discrete time) have parallel proofs. An error occurs in the final step of the common proof between equations (A9) and (A10) in [2, Appendix A] and is of the following general type. Let r and s be scalar random variables on the same probability space and let $z \in R$ be a deterministic parameter. Let h be a scalar random variable derived from r and s. Define two functions $f(s) = (E\{h(r,z)\})|_{z=s}$ and $g(s) = E\{h(r,s) \mid s\}$. In general, $f(s) \neq g(s)$. Furthermore, while $E\{h(r,s)\} = E\{E\{h(r,s) \mid s\}\} = E\{g(s)\}$, in general it is not true that $E\{h(r,s)\}$ equals $E\{f(s)\}$. The error leading up to [2, Eq. (A10)] is an assertion of the type that $E\{h(r,s)\} = E\{f(s)\}$. A concrete example is given in the Appendix. It is also possible for $E\{h(r,s)\} \neq E\{E\{h(r,s) \mid s\}\}$ and an example, based on the Cauchy distribution, is given in [4].

We focus on the discrete-time case where for each equation we give both the abstract form in terms of expectations and the concrete form in terms of integrals and probability density functions (which we assume exist). The model is

$$x_{t+1} = a(x_t,t) + b(x_t,t)w_t$$

$$y_t = g(x_t,t) + N(t)v_t$$

where $x_t \in R^n$, $y_t \in R^m$, w_t is independent and identically distributed (IID) $\mathcal{N}(0,I_n)$, v_t is IID $\mathcal{N}(0,I_m)$, x_0 is $\mathcal{N}(\overline{x}_0,\Sigma_0)$, and w, v, and x_0 are all independent. ($\mathcal{N}(m,Q)$ is the Gaussian distribution with mean m and covariance Q). Notation: $x_n^m = \{x_n, x_{n+1}, \ldots, x_m\}$ and likewise for y_n^m, v_n^m, and w_n^m. $R(j) = N(j)N(j)^T$. E is expectation and E^{x_t} is conditional expectation given x_t.

The Bucy–Mortensen–Duncan representation theorem in discrete time is

$$p(x_t \mid y_0^t) = \frac{[E^{x_t}\{\exp(\zeta_t(x_0^t, z_0^t))\}]_{z_0^t=y_0^t}\, p(x_t)}{[E\{\exp(\zeta_t(x_0^t, z_0^t))\}]_{z_0^t=y_0^t}}$$

where z_0^t is a deterministic vector of the same dimensions as y_0^t and

$$\zeta_t(x_0^t, y_0^t) = \sum_{j=0}^{t} g(x_j, j)^T R^{-1}(j) y_j - \frac{1}{2}\sum_{j=0}^{t} g(x_j, j)^T R^{-1}(j) g(x_j, j)$$

$$= \frac{1}{2}\sum_{j=0}^{t} g(x_j, j)^T R^{-1}(j) g(x_j, j) + \sum_{j=0}^{t} g(x_j, j)^T R^{-1}(j) N(j) v_j(x_j, y_j). \quad (1)$$

(There is a typographical error in [2, eq. (26)]: the "$N(j)$" factor in (1) is missing). More explicitly, the Bucy–Mortensen–Duncan representation theorem is

$$p(x_t \mid y_0^t) = \frac{\left[\int_{x_0^{t-1}} p(x_0^{t-1} \mid x_t) \exp(\zeta_t(x_0^t, y_0^t))\, dx_0^{t-1}\right] p(x_t)}{\left[\int_{x_0^t} p(x_0^t) \exp(\zeta_t(x_0^t, y_0^t))\, dx_0^t\right]}.$$

Equation (A10) of [2] is a bound on the mutual information $I(x_t; y_0^t)$ based on the Bucy–Mortensen–Duncan representation theorem. The derivation is correct through

$$I(x_t; y_0^t) = E\{\log[E^{x_t}\{\exp(\zeta_t(x_0^t, z_0^t))\}]_{z_0^t = y_0^t} - \log[E\{\exp(\zeta_t(x_0^t, z_0^t))\}]_{z_0^t = y_0^t}\} \quad (2)$$

$$\leq \log E\{[E^{x_t}\{\exp(\zeta_t(x_0^t, z_0^t))\}]_{z_0^t = y_0^t}\} - E\{\log[E\{\exp(\zeta_t(x_0^t, z_0^t))\}]_{z_0^t = y_0^t}\}. \quad (3)$$

(The second step is Jensen's inequality). The final step leading to [2, eq. (A10)] is the assertion that $E\{[E^{x_t}\{\exp(\zeta_t(x_0^t, z_0^t))\}]_{z_0^t = y_0^t}\}$ equals $E\{\exp(\zeta_t(x_0^t, y_0^t))\}$ which, unlike $E\{E^{x_t}\{\exp(\zeta_t(x_0^t, y_0^t))\}\} = E\{\exp(\zeta_t(x_0^t, y_0^t))\}$, is incorrect. More explicitly, the bound is

$$I(x_t; y_0^t) = \int_{x_t, y_0^t} p(x_t, y_0^t) \times \log\left\{\int_{x_0^{t-1}} p(x_0^{t-1} \mid x_t) \exp(\zeta_t(x_0^t, y_0^t))\, dx_0^{t-1}\right\} dx_t\, dy_0^t - \int_{y_0^t} p(y_0^t)\left\{\log \int_{x_0^t} p(x_0^t) \exp(\zeta_t(x_0^t, y_0^t))\, dx_0^t\right\} dy_0^t$$

$$\leq \log \int_{x_0^t, y_0^t} p(x_t, y_0^t) p(x_0^{t-1} \mid x_t) \exp(\zeta_t(x_0^t, y_0^t))\, dx_0^t\, dy_0^t - \int_{y_0^t} p(y_0^t)\left\{\log \int_{x_0^t} p(x_0^t) \exp(\zeta_t(x_0^t, y_0^t))\, dx_0^t\right\} dy_0^t$$

and the incorrect assertion is that

$$\int_{x_0^t, y_0^t} p(x_t, y_0^t) p(x_0^{t-1} \mid x_t) \exp(\zeta_t(x_0^t, y_0^t))\, dx_0^t\, dy_0^t$$

equals

$$\int_{x_0^t, y_0^t} p(x_0^t, y_0^t) \exp(\zeta_t(x_0^t, y_0^t))\, dx_0^t\, dy_0^t.$$

The incorrect assertion is important because, if it were true, then there would be no conditional function space expectations which are more difficult to evaluate than unconditional expectations. In the absence of the assertion, (2) provides a tighter bound than (3) and, at least in a Monte Carlo approach, requires essentially the same amount of computation.

III. EVALUATION OF THE BOUND

In the remainder of this note we describe a Monte Carlo method for evaluating (2), more specifically, a method for sampling from the conditional distribution (in the Gaussian case) and an importance sampling method which accelerates the convergence of the Monte Carlo sums. Though the Gaussian restriction is limiting, there exist applications, such as analog angle modulation systems [5] and speech processing [6], where the state process is linear and Gaussian but the measurement equation is nonlinear, and these methods are oriented toward such problems.

Any practical sampling algorithm must operate by transforming a set of IID $\mathcal{N}(0,1)$ samples. The sampling method we propose is based on the following observation. If x is $\mathcal{N}(\overline{x}, \Sigma)$ and x is partioned $x = (x_a', x_b')'$ (and likewise for $\overline{x}$ and Σ) then [7, p. 321] $p_{x_a|x_b}(x_a \mid x_b) = \mathcal{N}(\overline{x}_{a|b}(x_b), \Sigma_{a|b})(x_a)$ where $\overline{x}_{a|b}(x_b) = \overline{x}_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \overline{x}_b)$ and $\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ab}'$. The conditional mean estimator of x_a based on x_b, denoted by $\hat{x}_{a|b}$, is $\hat{x}_{a|b} = \overline{x}_{a|b}$ and the error, denoted by $\tilde{x}_{a|b}$, is defined to be $\tilde{x}_{a|b} = x_a - \hat{x}_{a|b}$ and has the properties that $E\{\tilde{x}_{a|b}\} = 0$ and $\mathrm{var}\{\tilde{x}_{a|b}\} = \Sigma_{a|b}$. Therefore, samples can be drawn from $\mathcal{N}(\overline{x}_{a|b}(x_b), \Sigma_{a|b})$ for a particular value of x_b (denoted by x_b^*) by using the following algorithm.

1) Choose x_b^*.
2) Pick a realization of x.
3) Compute $\hat{x}_{a|b}$.
4) Compute $\tilde{x}_{a|b}$. $\tilde{x}_{a|b}$ is distributed $\mathcal{N}(0, \Sigma_{a|b})$.
5) Compute $m_a = \overline{x}_{a|b}(x_b^*)$. (This would actually be done only once during an initialization procedure).
6) Compute $z = \tilde{x}_{a|b} + m_a$. z is distributed $\mathcal{N}(\overline{x}_{a|b}(x_b^*), \Sigma_{a|b})$ as desired.

In the case where x is the sequence of states of a dynamical system (specifically, x_a is the first M time-steps and x_b is the $(M+1)$st time-step), this process can be done efficiently because 1) the process noise w is a white sequence and therefore is easy to sample from, and the realization of x can easily be computed from the realization of w; 2) the conditional mean $\hat{x}_{a|b}$ can easily be computed by a Kalman smoother once an observation equation, with measurement matrix H and measurement noise covariance matrix R, is defined by

$$H_k = \begin{cases} 0, & k = 0,1,\ldots,M-1 \\ I, & k = M \end{cases}$$

$$R_k = \begin{cases} I, & k = 0,1,\ldots,M-1 \\ 0, & k = M \end{cases}$$

A forward-backward two-filter algorithm [7, p. 189] is attractive because the forward pass is very simple.

Let χ be a random variable with probability density function p. Computation of $I = E[f(\chi)]$ using Monte Carlo and importance sampling [8, Sect. 2.5] requires the computation of $\hat{I}_q = (1/L)\sum_{i=0}^{L-1} f(\gamma_i)p(\gamma_i)/q(\gamma_i)$ where γ_i are IID samples from the probability density function q. We apply these ideas to the computation of the conditional expectation in (2). (The complete computation of the first term of (2) requires Monte Carlo for the outer expectation also). Though the optimal q is known, we use a simple suboptimal choice: p is Gaussian and we choose q to be Gaussian with the same covariance but a different mean. The new mean is chosen as a compromise between the x_0^t trajectory that maximizes $\exp(\zeta_t(x_0^t, z_0^t))$ (which corresponds to f in $I = E[f(\chi)]$ above) and the mean of p, which maximizes p. The compromise is a time-step by time-step convex combination of the two trajectories, which are denoted m_ζ and m_p, respectively. The weight in the convex combination at step i on the mean of p is $\sigma_\zeta^2/(\sigma_\zeta^2 + \sigma_p^2)$ where σ_ζ is the width, as a function of x_i, of the maximum of $\exp(\zeta_t(x_0^t, z_0^t))$ at the trajectory m_ζ and σ_p^2 is the ith diagonal element of the covariance of p. The width is defined as the value of δx_i such that $\zeta_t(m_\zeta, z_0^t) - \zeta_t(\delta m_\zeta, z_0^t) = 2$ where $(\delta m_\zeta)_k = (m_\zeta)_k$ for $k \neq i$ and $(\delta m_\zeta)_k = (m_\zeta)_k + \delta x_i$ for $k = i$ [$(\cdot)_k$ means the value at time step k]. Because of the form of ζ, which is due to the fact that the observation noise is white, it is easy to find the trajectory x_0^t that maximizes $\exp(\zeta_t)$ because it can be done time-step by time-step.

In order to demonstrate these ideas, we have examined a scalar linear-Gaussian example: $x_{k+1} = ax_k + bw_k$ and $y_k = gx_k + Nv_k$ where $a = 0.5$, $b = 0.5$, $g = 2$, $N = 1$, and x_0 is $\mathcal{N}(0,20)$. The exact mean-squared error at $k = 50$, computed using the Kalman filter, is $\Sigma_{50|50} = 0.1328$. Pencil and paper evaluation of the rate distortion bound (i.e., (2) or the equation preceding (A10) in [2, Appendix]) gives the same result. The Monte Carlo value based on $L = 1000$ terms in each inner integral and in the outer integral and without using importance sampling is 0.0438 ± 0.0358 which is clearly far from convergence. (The sample mean of the bound plus/minus 1 sample standard deviation of the bound based on 10 runs is reported). Finally, if the calculation is unaltered except for the use of importance sampling, then the value is 0.1044 ± 0.0370 which is substantially closer to the exact value.

APPENDIX

Let $r \in \{0,1\}$, $s \in \{0,1\}$, $p_{r,s}(0,0) = p_{0,0}$, $p_{r,s}(0,1) = p_{0,1}$, $p_{r,s}(1,0) = p_{1,0}$, $p_{r,s}(1,1) = 1 - p_{0,0} - p_{0,1} - p_{1,0}$, and $h(r,s) = rs$. Then $f(s) = (E\{h(r,z)\})|_{z=s} = s(1 - p_{0,0} - p_{0,1})$ and $g(s) = E\{h(r,s) \mid s\} = s(1 - p_{0,0} - p_{0,1} - p_{1,0})/(1 - p_{0,0} - p_{1,0})$. Therefore, $E[g(s)] = 1 - p_{0,0} - p_{0,1} - p_{1,0} = E[h(r,s)]$ and $E[f(s)] = (1 - p_{0,0} - p_{0,1})(1 - p_{0,0} - p_{1,0})$ so that $E[f(s)] \neq E[g(s)] = E[h(r,s)]$.

SHAN LU
Comsat
22300 Comsat Drive
Clarksburg, MD 20871

PETER C. DOERSCHUK
School of Electrical and Computer Engineering
Purdue University
1285 Electrical Engineering Bldg.
W. Lafayette, IN 47907-1285

REFERENCES

[1] Kerr, T. H. (1989)
Status of CR-like lower bounds for nonlinear filtering.
IEEE Transactions on Aerospace and Electronic Systems, **25**, 5 (Sept. 1989), 590–601.

[2] Galdos, J. I. (1979)
A lower bound on filtering error with application to phase demodulation.
IEEE Transactions on Information Theory, **25**, 4 (July 1979), 452–462.

[3] Galdos, J. I. (1981)
A rate distortion theory lower bound on desired function filtering error.
IEEE Transactions on Information Theory, **27**, 3 (May 1981), 366–368.

[4] Enis, P. (1973)
On the relation $E(X) = E\{E(X \mid Y)\}$.
Biometrika, **60**, 2 (Aug. 1973), 432–433.

[5] Viterbi, A. J. (1966)
Principles of Coherent Communication.
New York: McGraw-Hill, 1966.

[6] Lu, S., and Doerschuk, P. C. (1996)
Nonlinear modeling and processing of speech based on sums of AM-FM formant models.
IEEE Transactions on Signal Processing, **44**, 3 (Apr. 1996).

[7] Anderson, B. D. O., and Moore, J. B. (1979)
Optimal Filtering.
Englewood Cliffs, NJ: Prentice-Hall, 1979.

[8] Bratley, P., Fox, B. L., and Schrage, L. E. (1983)
A Guide to Simulation.
New York: Springer-Verlag, 1983.