Probabilistic State Estimation with Discrete-Time Chaotic Systems

RLE Technical Report No. 571

Michael D. Richard

March 1992

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139-4307

This work was supported in part by the Defense Advanced Research Projects Agency monitored by the U.S. Navy Office of Naval Research under Grant N00014-89-J-1489, in part by the U.S. Air Force Office of Scientific Research under Grant AFOSR-91-0034-A, and in part by a subcontract from Lockheed Sanders, Inc. under the U.S. Navy Office of Naval Research Grant N00014-91-C-0125.

Contents

List of Figures

List of Tables

1 Introduction

Chaotic systems and their properties have received much attention in the mathematics and physics communities in the last two decades, and are receiving increasing attention in various engineering disciplines as well. Research has traditionally focused on possible causes of or transitions to chaos, universal properties shared by chaotic systems, and various topological and ergodic properties of chaotic systems. Recently however, practical applications of chaotic systems have been proposed for several disciplines including control, communication, and signal processing.

This report discusses probabilistic state estimation with chaotic systems, and as a result of the discussion and analysis reveals that distinguishing properties of chaotic systems may render them useful for practical engineering applications. The report begins with a brief review of probabilistic approaches to state estimation. The section continues by introducing the Kalman filter, a recursive, minimum-error-variance state estimator for linear systems. A recursive state estimator for nonlinear systems, the extended Kalman filter (EKF), is derived in Section 3 and shown to be the exact Kalman filter for linearizations of nonlinear systems. A related state estimator for nonlinear systems, the second order filter (EKF2), is also briefly discussed in the section.

The report continues with a quantitative and qualitative assessment of the EKF and EKF2 for performing state estimation with chaotic systems. Section 4 provides the quantitative assessment and Section 5, the qualitative assessment. In particular, Section 4 presents experimental performance results obtained with the EKF and EKF2 when used for state estimation with three discrete-time chaotic systems: the Henon, Ushiki, and Ikeda maps. Section 5 interprets these results primarily by focusing on the *a posteriori* state density and likelihood function for a chaotic system and contrasting it with the *a posteriori* state density and likelihood function for a linear system.

Section 6 briefly reviews linear smoothing terminology and techniques and introduces the extended Kalman smoother (EKS), an extension of the EKF, which exploits both past and future observations. The section presents experimental performance results obtained with the EKS on the same chaotic systems used with the EKF. Section 7 continues the analysis begun in Section 5 by interpreting the performance results for the EKS and comparing them with those for the EKF. The analysis focuses on the likelihood function for

a chaotic system that incorporates both past and future observations. The discussion reveals an interesting, distinguishing, potentially useful property of chaotic systems—the simultaneous existence of stable and unstable manifolds. This property is shown later in the report to strongly influence the Cramer-Rao error bound on state estimators for chaotic systems.

Section 8 considers two simpler problems than estimation, which are state and parameter discrimination among known, finite sets of values. Experimental results presented in the section suggest that robust state and parameter discrimination with chaotic systems is possible even with extremely low signal-to-noise (SNR) ratios. The section concludes by briefly discussing how this robust discrimination property of chaotic systems might be exploited to provide secure transmission of information.

Section 9 extends the informal analysis presented in Sections 5 and 7 by deriving and interpreting the Cramer-Rao bound on the error covariance matrix for unbiased estimators for each of three state estimation problems with chaotic systems. The analysis supports the assertion made in Section 7 that the simultaneous existence of both stable and unstable manifolds at each point on a chaotic attractor is an extremely important and useful property for state estimation. Section 9 concludes with a brief discussion of the relevance of the the error bounds.

Section 10 briefly discusses fundamental differences between linear and chaotic systems and the reasons that traditional engineering beliefs and assumptions are inappropriate when dealing with chaotic systems. Finally, Section 11 concludes the report with a brief summary.

2 State Estimation: Preliminaries

This section presents definitions and concepts used throughout this report. It introduces the general system model used in all derivations and examples, and briefly discusses and compares the three state estimation techniques considered in this report: maximum likelihood, maximum *a posteriori*, and minimum error variance. The section concludes with a discussion of the Kalman filter, a recursive, minimum-error-variance state estimator for linear, possibly time-varying dynamical systems.

2.1 The System Model

The following, general system model is used throughout the report:

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

$$y_n = h_n(x_n) + v_n. \quad (2)$$

The first equation is the *state equation.* In this equation, x_n is the $\mathcal{N}$-dimensional *state vector* we wish to estimate; $f_n(x_n)$ is a known, discrete-time, chaotic system; $g_n(x_n)$ is a known, possibly time-varying, and possibly nonlinear function of the state; and w_n, the *driving noise*, is an $\mathcal{N}$-dimensional, zero-mean, Gaussian, white-noise process. The second equation above is the *observation equation.* In this equation, y_n is the $\mathcal{P}$-dimensional *observation vector* we use to estimate x_n; $h_n(x_n)$ is a known, possibly time-varying, and possibly non-linear function of the state; and v_n, the *observation noise*, is a $\mathcal{P}$-dimensional, zero-mean, Gaussian, white-noise process. We assume that w_n and v_n are uncorrelated with each other and with the initial state x_0.

The notational convention used in the above equations and throughout the report is that the first subscript on a variable represents the time index. Therefore, x_n denotes the state at time n. In addition, a second subscript denotes a scalar component of a vector. Therefore, $x_{n,i}$ denotes the i^{th} component of the $\mathcal{N}$-dimensional vector x_n, or equivalently $x_n = [x_{n,1}, \cdots, x_{n,\mathcal{N}}]^T$.

2.2 Probabilistic Approaches to State Estimation

For the system model given by (1) and (2), the goal of state estimation is to estimate x_n at each time n in some "optimal" sense given a set of observations

$\{y_m\}$. If the set of observations used to estimate x_n for each n includes only observations y_m for times $m < n$, the estimation problem is known as a *prediction* problem. If the set of observations includes only observations for times $m \leq n$, the estimation problem is known as a *filtering* problem. Finally, if the set of observations includes at least some observations for times $m > n$, the estimation problem is known as a *smoothing* problem.

There are many approaches to perform state estimation with linear and nonlinear systems. In this report, we focus on three, related, probabilistic approaches: maximum likelihood (ML), maximum *a posteriori* (MAP), and minimum error variance (MEV). The relationship between the three approaches is best understood by considering an arbitrary set of observations $Y = \{y_m\}$ and the objective of estimating x_n for some fixed time n. The conditional density of x_n given Y, denoted $p(x_n|Y)$, is known as the *a posteriori* state density and can be expressed using Bayes rule as follows:

$$p(x_n|Y) = \frac{p(Y|x_n)p(x_n)}{p(Y)}. \tag{3}$$

In this equation, $p(Y|x_n)$ is the conditional density of the observation set Y given the state x_n and is known as the *likelihood function* when considered as a function of x_n for a fixed Y. Also, $p(x_n)$ is the unconditional or *a priori* state density, and $p(Y)$ is the unconditional density of the observation set.

With ML estimation, one chooses x_n to maximize $p(Y|x_n)$; with MAP estimation, one chooses x_n to maximize $p(x_n|Y)$; (as discussed in the next subsection) with MEV estimation, one chooses x_n to maximize the conditional mean $E(x_n|Y)$ given by

$$E(x_n|Y) = \int x_n p(x_n|Y)\, dx_n. \tag{4}$$

As indicated by (3), all three state-estimation approaches implicitly or explicitly use the likelihood function $p(Y|x_n)$.

2.3 Recursive State Estimation and the Kalman Filter

For many applications, the observations y_m are observed sequentially in time and one seeks to estimate x_n for each n using only observations for which $m \leq$

n. With the taxonomy introduced earlier, one thus has a *filtering* problem. In addition, it is desirable and sometimes necessary (for computational reasons) to recursively compute the "state estimate", hereafter denoted $\hat{x}_n$. That is for each n, one seeks to calculate $\hat{x}_n$ using only $\hat{x}_{n-1}$ and the current observation y_n. This is known as *recursive state estimation* and is the focus of this section and the next.

The problem of recursive state estimation simplifies when the functions $f_n(x_n)$ and $h_n(x_n)$ in (1) and (2), respectively, are linear functions of the state x_n, and $g_n(x_n)$ in (1) is linear (i.e., a matrix) and independent of the state. For this special case, (1) and (2) can be expressed

$$x_{n+1} = F_n x_n + G_n w_n \tag{5}$$

$$y_n = H_n x_n + v_n, \tag{6}$$

where F_n and G_n are $(\mathcal{N} \times \mathcal{N})$-matrices and H_n is a $(\mathcal{P} \times \mathcal{N})$-matrix. For this special case, the *a posteriori* state density $p(x_n|Y_0^n)$, where $Y_0^n = \{y_i\}_{i=0}^n$, is Gaussian [7]. A Gaussian density is completely characterized by a mean vector and a covariance matrix. Therefore, to recursively compute the density $p(x_n|Y_0^n)$, one need only update two finite sets of parameters at each time n: a mean vector and a covariance matrix.

A recursive state estimation problem in which only a finite number of parameters need to be updated to recursively compute the *a posteriori* state density is often referred to as a *finite dimensional* estimation problem. In contrast, a recursive state estimation problem in which an infinite number of parameters need to be updated to recursively compute the *a posteriori* state density is referred to as an *infinite dimensional* estimation problem. In general, when the functions $f_n(x_n)$, $g_n(x_n)$, and $h_n(x_n)$ are nonlinear functions of the state (such as when $f_n(x_n)$ is a chaotic system), the recursive estimation problem is infinite-dimensional and approximations are inevitably needed to recursively compute the *a posteriori* density.

It is well-known in the estimation literature that for the assumed system model and the special conditions on $f_n(x_n)$, $g_n(x_n)$, and $h_n(x_n)$ given above, the maximum *a posteriori* (MAP) and minimum-error-variance (MEV) recursive state estimates are identical. To see this equivalence, first note that with MEV recursive state estimation one chooses $\hat{x}_n$ to minimize the conditional error variance $E\left[(x_n - \hat{x}_n)^T(x_n - \hat{x}_n)|Y_0^n\right]$. However, the conditional

error variance is minimized when $\hat{x}_n$ equals the conditional mean,

$$\hat{x}_n = E(x_n|Y_0^n) = \int x_n p(x_n|Y_0^n)\, dx_n. \tag{7}$$

As mentioned earlier, $p(x_n|Y_0^n)$ is Gaussian and is thus a unimodal density centered about its mean. Since by definition the MAP recursive state estimate is the value of x_n for which $p(x_n|Y_0^n)$ has its maximum, the MAP estimate is the conditional mean. Therefore, both the MAP and MEV state estimates are identical for this special filtering problem with the system model given by (5) and (6)

The discrete-time Kalman filter, hereafter referred to as the Kalman filter, is a popular, practical, recursive, MEV (and MAP) state estimator for this special filtering problem The Kalman filter uses a two-step procedure for recursively computing two quantities—the state estimate $\hat{x}_n$ and the error covariance matrix $P_n \equiv E\left[(x_n - \hat{x}_n)(x_n - \hat{x}_n)^T)|Y_0^n\right]$. In the first step, known as the *prediction step*, the state estimate and covariance matrix for time $n+1$ are computed using only the final state estimate and covariance matrix for time n and observations through time n (i.e., Y_0^n). The state estimate and error covariance matrix computed in the prediction step for time $n+1$ are typically denoted $\hat{x}_{n+1|n}$ and $P_{n+1|n}$, respectively, to emphasize that only observations through time n are used to calculate them. In the second step, known as the *measurement (or observation) update* step, the quantities $\hat{x}_{n+1|n}$ and $P_{n+1|n}$ calculated in the first step are updated using the "new" observation y_{n+1}. The updated quantities are typically denoted $\hat{x}_{n+1|n+1}$ and $P_{n+1|n+1}$. The equations for the two steps of the Kalman filter are given in Figure 1 and are applicable to the system model given by (5) and (6) with $w_n \sim N(0, Q_n)$, $v_n \sim N(0, R_n)$, and $x_0 \sim N(m_0, P_0)$.

The Kalman filter has been successively used in many practical applications, perhaps most notably the Apollo program in the 1960's. Unfortunately, the Kalman filter is applicable only to the system model given by (5) and (6) and not the more general model given by (1) and (2). As discussed in the next section, the *extended* Kalman filter is a recursive state estimator, based on the Kalman filter, which is applicable to the system model given by (1) and (2). However, unlike the Kalman filter which is optimal in the sense that the state estimate is the MEV and MAP estimate, the extended Kalman filter is not optimal in this sense (or in any other usual sense of optimality). As discussed in the next section and illustrated with examples in

Prediction Step

$$\hat{x}_{n+1|n} = F_n \hat{x}_{n|n} \tag{8}$$

$$P_{n+1|n} = F_n P_{n|n} F_n^T + G_n Q_n G_n^T \tag{9}$$

Measurement Update Step

$$\hat{x}_{n+1|n+1} = \hat{x}_{n+1|n} + K_{n+1}[y_{n+1} - H_{n+1}\hat{x}_{n+1|n}] \tag{10}$$

$$K_{n+1} = P_{n+1|n} H_{n+1}^T \left[H_{n+1} P_{n+1|n} H_{n+1}^T + R_{n+1}\right]^{-1} \tag{11}$$

$$P_{n+1|n+1} = [I - K_{n+1} H_{n+1}]\, P_{n+1|n} \tag{12}$$

Initialization

$$\hat{x}_{0|-1} = m_0 \tag{13}$$

$$P_{0|-1} = P_0. \tag{14}$$

Figure 1: The Kalman filter

Section 4, because of this lack of optimality, one can not determine *a priori* the performance of the extended Kalman filter for a specific problem. As aptly remarked in [7] in reference to the extended Kalman filter and related nonlinear filters, "...our approximations are *ad hoc* and must be tested by simulation."

3 The Extended Kalman Filter (EKF)

The extended Kalman filter (EKF) is a recursive state estimator for the general system model given by (1) and (2). Unlike the Kalman filter, the EKF imposes no restrictions on the functions $f_n(x_n)$, $g_n(x_n)$, and $h_n(x_n)$ in (1) and (2) except that they be differentiable functions of the state x_n. Thus, whereas the Kalman filter is applicable only to linear system models (i.e., those of the form of (5) and (6)), the EKF is applicable to nonlinear models as well. However the EKF is a heuristically derived algorithm which does not in general yield either the minimum-error-variance (MEV) or the maximum *a posteriori* (MAP) state estimate.

3.1 Motivation

As mentioned in Section 2, the estimate of x_n based only on Y_0^n (i.e., the set of observations $\{y_i\}_{i=0}^n$) which minimizes the error variance (for both linear and nonlinear systems) is the conditional mean

$$E(x_n|Y_0^n) = \int x_n \, p(x_n|Y_0^n) \, dx_n. \tag{15}$$

Also as pointed out in Section 2, for the system model given by (5) and (6), the *a posteriori* density $p(x_n|Y_0^n)$ is Gaussian and is thus completely specified by two sets of parameters—a mean vector and a covariance matrix. Thus, to recursively update $p(x_n|Y_0^n)$, one need only recursively update these two sets of parameters. This suggests why only two sets of parameters—the state estimate (the conditional mean) and the error covariance matrix—are updated at each two-step iteration of the Kalman filter. In contrast, for most nonlinear systems, the *a posteriori* density requires an infinite set of parameters for its specification. As a result, MEV recursive state estimation for most nonlinear systems requires that an infinite set of parameters be updated at each time n. Since updating an infinite set of parameters is not feasible, any practical recursive state estimator for most nonlinear systems can provide at best only an approximation to the MEV state estimate.

In the past, there have been two basic approaches for deriving approximate MEV state estimators for nonlinear systems. The first approach involves approximating the nonlinear functions $f_n(x_n)$, $g_n(x_n)$, and $h_n(x_n)$ in (1) and (2). Typically, this is done by expanding the functions in Taylor

series and truncating all but the lowest-order terms. The second basic approach involves approximating $p(x_n|Y_0^n)$ with a finite set of parameters and updating only these parameters at each time instant.

3.2 Derivation

The derivation of the EKF uses the first approach for deriving approximate MEV state estimators. An underlying assumption of this approach is that the functions $f_n(x_n)$, $g_n(x_n)$, and $h_n(x_n)$ are "sufficiently smooth" or differentiable so that they have Taylor series expansions.

For the EKF, at each time n the functions are expanded about the current state estimate (either $\hat{x}_{n|n-1}$ or $\hat{x}_{n|n}$) where the subscripts have the same interpretation as in Section 2 for the Kalman filter. (That is, $\hat{x}_{n|n-1}$ is the estimate of x_n based on the observation set $Y_0^{n-1} = \{y_i\}_{i=0}^{n-1}$. Similarly, $\hat{x}_{n|n}$ is the estimate of x_n based on the observation set Y_0^n). Specifically, the Taylor series expansions are the following:

$$f_n(x_n) = f_n(\hat{x}_{n|n}) + F_n(x_n - \hat{x}_{n|n}) + \cdots \tag{16}$$

$$g_n(x_n) = G_n + \cdots \tag{17}$$

$$h_n(x_n) = h_n(\hat{x}_{n|n-1}) + H_n(x_n - \hat{x}_{n|n-1}) + \cdots, \tag{18}$$

where

$$F_n = \left.\frac{\partial f_n(x)}{\partial x}\right|_{x=\hat{x}_{n|n}} \tag{19}$$

$$G_n = g_n(\hat{x}_{n|n}) \tag{20}$$

$$H_n = \left.\frac{\partial h_n(x)}{\partial x}\right|_{x=\hat{x}_{n|n-1}}. \tag{21}$$

Retaining only those terms explicitly shown in the above expansions, yields the following approximations to (1) and (2):

$$x_{n+1} = f_n(\hat{x}_{n|n}) + F_n(x_n - \hat{x}_{n|n}) + G_n w_n \tag{22}$$

$$= F_n x_n + G_n w_n + \left[f_n(\hat{x}_{n|n}) - F_n \hat{x}_{n|n}\right] \tag{23}$$

$$y_n = h_n(\hat{x}_{n|n-1}) + H_n(x_n - \hat{x}_{n|n-1}) + v_n \tag{24}$$

$$= H_n x_n + v_n + \left[h_n(\hat{x}_{n|n-1}) - H_n \hat{x}_{n|n-1}\right]. \tag{25}$$

In (23) and (25), F_n and H_n are matrices, and the bracketed expressions can be evaluated. Thus, these equations are similar to the state and observation equations given by (5) and (6), with the addition of deterministic input terms. However, the Kalman filter equations can easily be modified to account for deterministic inputs in the state and observation equations. One simply incorporates the deterministic input terms in the equations for $\hat{x}_{n+1|n}$ and $\hat{x}_{n+1|n+1}$ given by (8) and (10), respectively.

In light of this, the Kalman filter (modified to account for deterministic inputs) is applicable to the system model given by (23) and (25). The resulting filtering equations, provided in Figure 2, constitute the *extended Kalman filter*.

Prediction Step

$$\hat{x}_{n+1|n} = f_n(\hat{x}_{n|n}) \tag{26}$$

$$P_{n+1|n} = F_n P_{n|n} F_n^T + G_n Q_n G_n^T \tag{27}$$

Measurement Update Step

$$\hat{x}_{n+1|n+1} = \hat{x}_{n+1|n} + K_{n+1}\left[y_{n+1} - h_{n+1}(\hat{x}_{n+1|n})\right] \tag{28}$$

$$K_{n+1} = P_{n+1|n} H_{n+1}^T \left[H_{n+1} P_{n+1|n} H_{n+1}^T + R_{n+1}\right]^{-1} \tag{29}$$

$$P_{n+1|n+1} = \left[I - K_{n+1} H_{n+1}\right] P_{n+1|n} \tag{30}$$

Initialization

$$\hat{x}_{0|-1} = m_0 \tag{31}$$

$$P_{0|-1} = P_0. \tag{32}$$

Figure 2: The extended Kalman filter (EKF)

A comparison of the equations provided in Figures 1 and 2 reveals the similarity of the Kalman and extended Kalman filters. However, whereas the Kalman filter applies to a "linear" system, the EKF applies to a "linearized" system. This analogy suggests that the effectiveness of the EKF depends largely on the accuracy of the linear approximations of the nonlinear functions $f_n(x_n)$, $g_n(x_n)$, and $h_n(x_n)$. If the neglected higher-order terms in the

Taylor series expansions of these functions are not negligible, the EKF may perform poorly. In addition, whereas the Kalman filter is a MEV recursive state estimator, the EKF in general is not a MEV recursive state estimator for the original nonlinear system.

One other interesting aspect of the EKF concerns the state estimate equation in the prediction step (26). This equation follows from (22),

$$\begin{aligned} \hat{x}_{n+1|n} &= E(x_{n+1}|Y_0^n) && (33)\\ &= f_n(\hat{x}_{n|n}) + F_n\left(E(x_n|Y_0^n) - \hat{x}_{n|n}\right) + E(G_n w_n) && (34)\\ &= f_n(\hat{x}_{n|n}), && (35) \end{aligned}$$

since $E(x_n|Y_0^n) = \hat{x}_{n|n}$ and $E(G_n w_n) = 0$. This implies

$$E(x_{n+1}|Y_0^n) = E(f(x_n)|Y_0^n) = f(E(x_n|Y_0^n)), \tag{36}$$

which although true for linear systems is not true in general for nonlinear systems.

3.3 The Second Order Filter (EKF2)

An alternative to the EKF known as the *second order filter*, and hereafter denoted the EKF2, has a similar derivation. However, both first-order and second-order terms in the Taylor series expansions of $f_n(x_n)$ and $h_n(x_n)$ are retained in the filtering equations. The derivation of the EKF2 is tedious and unrevealing. In light of this, only the resulting equations for the EKF2 are provided here.

First define the following [17]:

$$\begin{aligned} f_n(\cdot) &= [f_n^1(\cdot), \cdots, f_n^{\mathcal{N}}(\cdot)]^T && (37)\\ F_n &= \left.\frac{\partial f_n(x)}{\partial x}\right|_{x=\hat{x}_{n|n}} && (38)\\ E_n^i &= \left.\frac{\partial^2 f_n^i(x)}{\partial x^2}\right|_{x=\hat{x}_{n|n}} && (39)\\ h_n(\cdot) &= [h_n^1(\cdot), \cdots, h_n^{\mathcal{P}}(\cdot)]^T && (40)\\ H_n &= \left.\frac{\partial h_n(x)}{\partial x}\right|_{x=\hat{x}_{n|n-1}} && (41) \end{aligned}$$

$$M_n^i = \left. \frac{\partial^2 h_n^i(x)}{\partial x^2} \right|_{x=\hat{x}_{n|n-1}} . \tag{42}$$

(Note E_n^i and M_n^i are $(\mathcal{N} \times \mathcal{N})$-matrices). With these definitions, the filtering equations for the EKF2 are as shown in Figure 3. In the equations, the symbol $Tr(\cdot)$ denotes the trace operator.

Prediction Step

$$\hat{x}_{n+1|n} = f_n(\hat{x}_{n|n}) + \frac{1}{2}\sum_{i=1}^{\mathcal{N}} e_i \, Tr\left(E_n^i P_{n|n}\right) \tag{43}$$

$$P_{n+1|n} = F_n P_{n|n} F_n^T + G_n Q_n G_n^T + T_n \tag{44}$$

$$T_n = [T_n^{ij}] \tag{45}$$

$$= \left[\frac{1}{2} Tr\left(F_n^i P_{n|n} F_n^j P_{n|n}\right)\right] \tag{46}$$

Measurement Update Step

$$\hat{x}_{n+1|n+1} = \hat{x}_{n+1|n} - \frac{1}{2} K_{n+1} \sum_{i=1}^{\mathcal{N}} e_i \, Tr\left(M_{n+1}^i P_{n+1|n}\right)$$
$$+ K_{n+1}\left[y_{n+1} - h_{n+1}(\hat{x}_{n+1|n})\right] \tag{47}$$

$$K_{n+1} = P_{n+1|n} H_{n+1}^T \left[H_{n+1} P_{n+1|n} H_{n+1}^T + R_{n+1} + S_{n+1}\right]^{-1} \tag{48}$$

$$S_{n+1} = [S_n^{ij}] \tag{49}$$

$$= \left[\frac{1}{2} Tr\left(M_{n+1}^i P_{n+1|n} M_{n+1}^j P_{n+1|n}\right)\right] \tag{50}$$

$$P_{n+1|n+1} = \left[I - K_{n+1} H_{n+1}\right] P_{n+1|n} \tag{51}$$

Initialization

$$\hat{x}_{0|-1} = m_0 \tag{52}$$

$$P_{0|-1} = P_0. \tag{53}$$

Figure 3: The second order filter (EKF2)

4 Experimental Performance of the EKF

This section evaluates the extended Kalman filter (EKF) and second-order filter (EKF2) in performing state estimation with three, discrete-time, chaotic systems: the Henon, Ushiki, and Ikeda maps. The section begins by introducing these maps and continues by presenting experimental performance results for the EKF and EKF2.

4.1 The Henon, Ushiki, and Ikeda maps

The Henon, Ushiki, and Ikeda maps are two-dimensional, discrete-time, chaotic systems or *maps*. With the notational convention that the components of the two-dimensional state vector x_n are denoted $x_{n,1}$ and $x_{n,2}$ (and thus $x_n = [x_{n,1}, x_{n,2}]^T$), the three maps have the following functional forms ($x_{n+1} = f(x_n)$):

Henon Map

$$x_{n+1,1} = 1 - 1.4x_{n,1}^2 + x_{n,2} \tag{54}$$

$$x_{n+1,2} = .3x_{n,1} \tag{55}$$

Ushiki Map

$$x_{n+1,1} = 3.7x_{n,1} - x_{n,1}^2 - .1x_{n,1}x_{n,2} \tag{56}$$

$$x_{n+1,2} = 3.7x_{n,2} - x_{n,2}^2 - .15x_{n,1}x_{n,2} \tag{57}$$

$$\tag{58}$$

Ikeda Map

$$x_{n+1,1} = 1 + .9\left[x_{n,1} \cos \alpha_n - x_{n,2} \sin \alpha_n\right] \tag{59}$$

$$x_{n+1,2} = .9\left[x_{n,1} \sin \alpha_n + x_{n,2} \cos \alpha_n\right] \tag{60}$$

$$\alpha_n = .4 - \frac{6}{1 + x_{n,1}^2 + x_{n,2}^2}. \tag{61}$$

Other values of the constants used in the above equations are also permissible. The Ushiki map differs from the Henon and Ikeda maps in that it is not invertible, while the other two maps are diffeomorphisms (and thus invertible).

The Ikeda map can also be succinctly expressed with complex notation:

$$x_{n+1} = f(x_n) = 1 + .9\, x_n \, \exp\left\{ j \left[.4 - \frac{6}{1 + \|x_n\|^2} \right] \right\} \tag{62}$$

where the two-dimensional state vector x_n is now treated as a complex variable, with $x_{n,1}$ and $x_{n,2}$ corresponding to its real and imaginary parts, respectively. Figures 4, 5, and 6 depict the chaotic attractors for the three maps.

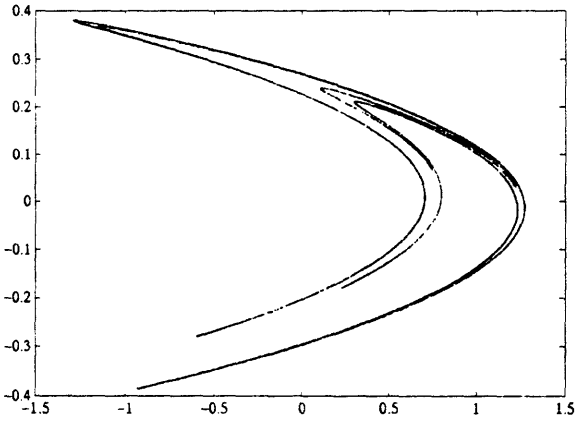

Figure 4: Chaotic attractor for the Henon map

4.2 Computer Experiments

Two sets of experiments were conducted to assess the performance of the EKF and EKF2. Both sets of experiments used the following system model, a special case of the general model given by (1) and (2):

$$x_{n+1} = f(x_n) \tag{63}$$

$$y_n = x_n + v_n, \tag{64}$$

where $f(x_n)$ refers to one of the three maps and v_n was a zero-mean, Gaussian, white-noise process with constant covariance matrix R given by

$$R = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}. \tag{65}$$

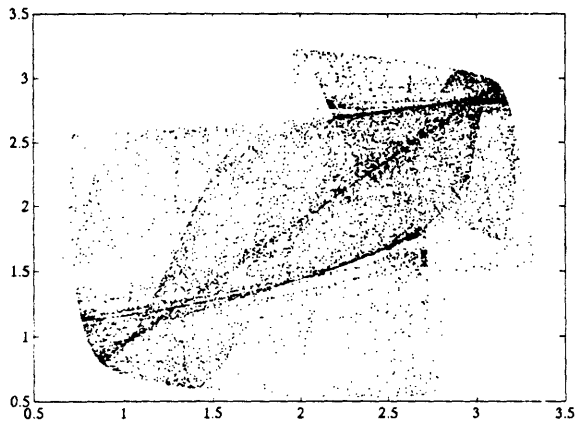

Figure 5: Chaotic attractor for the Ushiki map

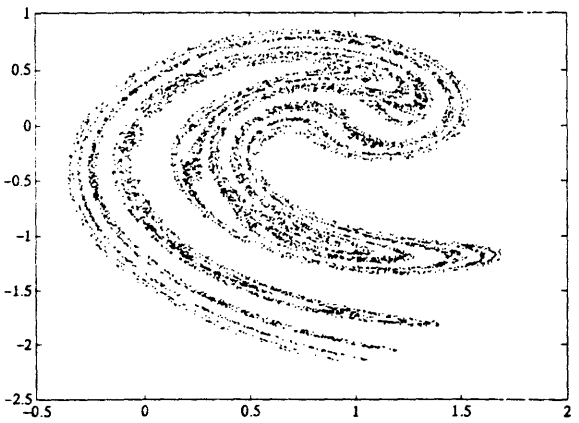

Figure 6: Chaotic attractor for the Ikeda map

As indicated by (63), driving noise was absent in the state equation; thus, the state was deterministically generated.

The EKF was tested on the Henon, Ushiki, and Ikeda maps, while the EKF2 was tested only on the Henon map. Each experiment used 1000 observations, $\{y_i\}_{i=0}^{999}$. Each performance result was obtained by averaging the results for five independent trials. The measure of performance was the improvement in signal-to-noise ratio (SNR) achieved for each component of x_n, where the initial and final SNRs for $x_{n,j}$, the j^{th} component of x_n (where $j = 1$ or 2), were calculated as

$$\text{Initial SNR} = 10\log_{10}\left[\frac{\sum_{i=0}^{999}(x_{i,j}-\bar{x}_j)^2}{\sum_{i=0}^{999}(x_{i,j}-y_{i,j})^2}\right] \tag{66}$$

$$\text{Final SNR} = 10\log_{10}\left[\frac{\sum_{i=0}^{999}(x_{i,j}-\bar{x}_j)^2}{\sum_{i=0}^{999}(x_{i,j}-\hat{x}_{(i|i),j})^2}\right], \tag{67}$$

where $y_{i,j}$ denotes the j^{th} component of the i^{th} observation, $\hat{x}_{(i|i),j}$ denotes the estimate of the j^{th} component of the state at time i. and where

$$\bar{x}_j \equiv \frac{1}{1000}\sum_{i=0}^{999} x_{i,j}. \tag{68}$$

For all three maps, the signal variance $\sum_{i=0}^{999}(x_{i,j}-\bar{x}_j)^2$ for the two components of x_n differs. Because of this, achieving the same initial SNR for the two components of x_n required that σ_1^2 and σ_2^2, the variances of the two components of v_n, be given different values.

A mean vector m_0 and covariance matrix P_0 were needed to initialize the EKF and EKF2. The same values of m_0 and P_0 and different values for σ_1 and σ_2 were used for each experiment. The mean vector m_0 was chosen as the actual initial point x_0 with an added, zero-mean, Gaussian, white-noise component. P_0 was chosen as the covariance matrix of this noise component, which for both sets of experiments was the following :

$$P_0 = \begin{bmatrix} 10^{-6} & 0 \\ 0 & 10^{-6} \end{bmatrix}. \tag{69}$$

Additional experiments showed that the initial error covariance P_0 had little influence on performance of the EKF and the extended Kalman smoother (introduced later in this report).

Table 1 contains the performance results for the first set of experiments. As indicated by the table, the variable in the experiments was the initial SNR, which was chosen to be the same for the two components of x_n. Each value shown in the columns under the heading "SNR Improvement" denotes the difference between the final and initial SNRs at the indicated noise level for the indicated component of the indicated map. A positive value denotes an increase in SNR, while a negative value denotes a decrease. For example, the value of 4.28 as the first entry of the first column means that with an initial SNR of 40 dB for the first component of the Henon map, the final SNR (as averaged over five independent trials) was 44.28 dB. In contrast, the value of -1.6 as the final entry of the first column means that with an initial SNR of 6 dB for the first component of the Henon map, the final SNR (as averaged over five independent trials) was 4.4 dB.

Initial SNR (dB)	SNR Improvement (dB)							
	Henon Map				Ushiki Map		Ikeda Map	
	EKF		EKF2		EKF		EKF	
	$x_{n,1}$	$x_{n,2}$	$x_{n,1}$	$x_{n,2}$	$x_{n,1}$	$x_{n,2}$	$x_{n,1}$	$x_{n,2}$
40	4.28	7.85	4.06	7.81	-29.52	-28.26	-11.89	-10.53
20	-1.37	1.08	3.94	7.70	-13.71	-13.40	-6.79	-6.25
10	-2.34	-1.56	3.43	6.65	-6.65	-6.19	-5.66	-4.96
6	-1.60	-.44	3.30	6.81	-4.25	-3.63	-5.59	-5.00

Table 1: Performance results for first set of experiments

Perhaps surprisingly, even though m_0, the estimate of the initial state, was close to the actual value, the improvement in SNR was small and in many cases the SNR deteriorated.

One reason for this was that since driving noise was absent from the system model, the Kalman gain K_n, given by (29) for the EKF and (48) for the EKF2, was small for both the EKF and EKF2, and the observations were essentially ignored. One way to avoid this undesirable behavior of the EKF and EKF2 was to "assume" that the state equation (63) had a small, driving noise term when specifying the EKF and EKF2 equations. This entailed assigning nonzero values to elements of G_n and Q_n in the EKF and EKF2 equations given in Section 3.

With G_n chosen as the identity matrix and Q_n chosen as

$$Q_n = \begin{bmatrix} .001 & 0 \\ 0 & .001 \end{bmatrix}, \tag{70}$$

considerable improvement in performance occurred with the EKF but not with the EKF2, as indicated in Table 2 for this second set of experiments.

Initial SNR (dB)	SNR Improvement (dB)							
	Henon Map				Ushiki Map		Ikeda Map	
	EKF		EKF2		EKF		EKF	
	$x_{n,1}$	$x_{n,2}$	$x_{n,1}$	$x_{n,2}$	$x_{n,1}$	$x_{n,2}$	$x_{n,1}$	$x_{n,2}$
40	.33	.04	.33	.03	.34	.34	.17	.44
20	2.81	2.54	2.74	2.53	3.30	3.83	2.12	3.44
10	3.05	6.08	3.36	6.21	3.48	4.82	.88	1.92
6	1.43	4.47	3.24	6.72	2.23	3.75	-1.07	-.30

Table 2: Performance results for second set of experiments (assumed driving noise)

Other choices of the parameters used by the EKF were also tested as well as other variations of the EKF, including the *iterated* EKF. Only a marginal improvement in performance resulted at best with any of these alternatives.

5 Performance Analysis of the EKF

As indicated by the performance results in the preceding section, the EKF and EKF2 perform poorly (or mediocre at best) for state estimation with at least three chaotic systems. An important question is why? This section attempts to answer or at least shed light on this question by providing an initial intuitive interpretation of theoretical results presented later in this report. In the process, the section reveals an interesting property of all chaotic systems that is potentially useful when performing state estimation with these systems. Although simple and revealing, the intuitive analysis presented in this section is applicable only to diffeomorphisms, and is thus relevant to the Henon and Ikeda maps but not to the Ushiki map (since it is not invertible).

5.1 Derivation of the *a Posteriori* State Density and the Likelihood Function

Since the EKF is in fact the Kalman filter for a linearized system, one would expect its performance to depend upon how "close" certain properties of the corresponding nonlinear system match those of a linear system. One property of the linear system model given by (5) and (6) is that the *a posteriori* state density $p(x_n|Y_0^n)$ is Gaussian. As discussed in Section 2, this is a fundamental factor in the derivation and optimality of the Kalman filter. This suggests that for the EKF to perform well, the *a posteriori* state density for the corresponding nonlinear system should be approximately Gaussian.

The *a posteriori* state density for the general system model given by (1) and (2) can be recursively defined as follows [7]:

$$p(x_{n+1}|Y_0^{n+1}) = \frac{p(y_{n+1}|x_{n+1}) \int p(x_{n+1}|x_n)\, p(x_n|Y_0^n)\, dx_n}{\int\int p(y_{n+1}|x_{n+1})\, p(x_{n+1}|x_n)\, p(x_n|Y_0^n)\, dx_n\, dx_{n+1}}. \tag{71}$$

The derivation of this formidable-looking expression is straightforward. Using Bayes rule, we can express $p(x_{n+1}|Y_0^{n+1})$ as

$$p(x_{n+1}|Y_0^{n+1}) = \frac{p(Y_0^{n+1}|x_{n+1})\, p(x_{n+1})}{p(Y_0^{n+1})}. \tag{72}$$

Since $Y_0^{n+1} = \{y_i\}_{i=0}^{n+1} = Y_0^n \cup y_{n+1}$, we can also express (72) as

$$p(x_{n+1}|Y_0^{n+1}) = \frac{p(y_{n+1}|x_{n+1}, Y_0^n)\, p(x_{n+1}|Y_0^n)}{p(y_{n+1}|Y_0^n)}. \tag{73}$$

By assumption, the driving noise w_n in (1) and the observation noise v_n in (2) are white, and uncorrelated both with each other and with the initial state. Therefore, it follows from (2) and the Markov property of (1) that

$$p(y_{n+1}|Y_0^n, x_{n+1}) = p(y_{n+1}|x_{n+1}) \tag{74}$$

$$p(x_{n+1}|x_n, Y_0^n) = p(x_{n+1}|x_n). \tag{75}$$

Also, we can write $p(y_{n+1}|Y_0^n)$ as

$$p(y_{n+1}|Y_0^n) = \int p(y_{n+1}, x_{n+1}|Y_0^n)\, dx_{n+1} \tag{76}$$

$$= \int p(y_{n+1}|x_{n+1}, Y_0^n)\, p(x_{n+1}|Y_0^n)\, dx_{n+1} \tag{77}$$

$$= \int p(y_{n+1}|x_{n+1})\, p(x_{n+1}|Y_0^n)\, dx_{n+1}, \tag{78}$$

where the last equality uses (74). Similarly, we can write $p(x_{n+1}|Y_0^n)$ as

$$p(x_{n+1}|Y_0^n) = \int p(x_{n+1}, x_n|Y_0^n)\, dx_n \tag{79}$$

$$= \int p(x_{n+1}|x_n, Y_0^n)\, p(x_n|Y_0^n)\, dx_n \tag{80}$$

$$= \int p(x_{n+1}|x_n)\, p(x_n|Y_0^n)\, dx_n. \tag{81}$$

where the last equality uses (75). Combining (73), (74), (78), and (81) yields (71).

In general, it is difficult to evaluate (71) for specific values of x_{n+1}. However, (71) simplifies and becomes easy to evaluate, when the restricted system model given by (63) and (64) (which omits driving noise) is used. The model is repeated here for convenience:

$$x_{n+1} = f(x_n) \tag{82}$$

$$y_n = x_n + v_n. \tag{83}$$

For this model,

$$p(x_{n+1}|x_n) = \delta\left(x_{n+1} - f(x_n)\right) \tag{84}$$

where $\delta(\cdot)$ is the dirac delta function. Substituting (84) in (71) yields

$$p(x_{n+1}|Y_0^{n+1}) = C_1\, p(y_{n+1}|x_{n+1}) \int \delta\left(x_{n+1} - f(x_n)\right)\, p(x_n|Y_0^n)\, dx_n, \quad (85)$$

where C_1 is a normalizing constant independent of x_{n+1}. (Actually, C_1 is the reciprocal of the denominator in (71)). After applying a change of variables and evaluating the integral, (85) reduces to

$$\begin{aligned} p(x_{n+1}|Y_0^{n+1}) &= C_1\, p(y_{n+1}|x_{n+1})\, p\left([x_n = f^{-1}(x_{n+1})]|Y_0^n\right) \\ &\quad \times \|D\{f^{-1}(x_{n+1})\}\|, \end{aligned} \quad (86)$$

where $\|\cdot\|$ is the determinant operator and $D\{f^{-1}(x_{n+1})\}$ is the derivative of $f^{-1}(\cdot)$ evaluated at x_{n+1} (or equivalently the the Jacobian of $f^{-1}(\cdot)$ evaluated at x_{n+1}). Therefore, $\|D\{f^{-1}(x_{n+1})\}\|$ is the determinant of the Jacobian of $f^{-1}(\cdot)$ evaluated at x_{n+1}. Iterating the recursion implicit in (86) and simplifying yields

$$\begin{aligned} p(x_{n+1}|Y_0^{n+1}) &= C_1\, p\left(x_0 = f^{-(n+1)}(x_{n+1})\right)\, \|D\{f^{-(n+1)}(x_{n+1})\}\| \\ &\quad \times \prod_{i=0}^{n+1} p\left(y_i|f^{-(n+1-i)}(x_{n+1})\right), \end{aligned} \quad (87)$$

where $f^{-(n+1)}(\cdot)$ means $f^{-1}(\cdot)$ iterated $n+1$ times and $D\{f^{-(n+1)}(x_{n+1})\}$ is the determinant of the Jacobian of $f^{-(n+1)}(\cdot)$ evaluated at x_{n+1}.

Since v_n in (83) is white and $v_n \sim N(0, R)$ by assumption, it follows that

$$p(y_n|x_n) = p\left([v_n = y_n - x_n]|x_n\right) \sim N(x_n, R). \quad (88)$$

Finally, combining (87) and (88) yields

$$\begin{aligned} p(x_{n+1}|Y_0^{n+1}) &= C_2\, p\left(x_0 = f^{-(n+1)}(x_{n+1})\right)\, \|D\{f^{-(n+1)}(x_{n+1})\}\| \\ &\times \exp\left\{-\sum_{i=0}^{n+1} \left[y_i - f^{-(n+1-i)}(x_{n+1})\right]^T R^{-1} \left[y_i - f^{-(n+1-i)}(x_{n+1})\right]\right\} \end{aligned} \quad (89)$$

where C_2 is a normalizing constant. A comparison of (72) and (89) reveals that $p(x_{n+1})$, the unconditional probability of the state x_{n+1}, is given by

$$p(x_{n+1}) = p\left(x_0 = f^{-(n+1)}(x_{n+1})\right)\, \|D\{f^{-(n+1)}(x_{n+1})\}\|, \quad (90)$$

and $p(Y_0^n|x_n)$, the likelihood function (or equivalently the conditional probability of the observation set), is given by

$$p(Y_0^n|x_n) \propto$$
$$\exp\left\{-\sum_{i=0}^{n+1}\left[y_i - f^{-(n+1-i)}(x_{n+1})\right]^T R^{-1}\left[y_i - f^{-(n+1-i)}(x_{n+1})\right]\right\}. \quad (91)$$

The next subsection considers the properties of $p(Y_0^n|x_n)$, and as a consequence shows that the *a posteriori* state density $p(x_n|Y_0^n)$ is not Gaussian for the Henon and Ikeda maps.

5.2 Properties of the Likelihood Function

The relation between the *a posteriori* density $p(x_n|Y_0^n)$ and the likelihood function $p(Y_0^n|x_n)$ given by (72) suggests that if the *a posteriori* density is Gaussian, then the likelihood function (as a function of x_n) should have a Gaussian "shape" as well. (Note that $p(Y_0^n|x_n)$ considered as a function of x_n for fixed Y_0^n is not a probability density). This argument is slightly flawed, since the *a priori* density $p(x_n)$ enters (72) as well. However for the Henon map, if the initial state density $p(x_0)$ is Gaussian, then $p(x_n)$ has a Gaussian "shape" as well, because the determinant of the Jacobian, $\|D\{f^{-1}(\cdot)\}\|$, is constant.

Figures 7 and 8 are contour plots of the likelihood function $p(Y_0^n|x_n)$ at a single time n for the Henon and Ikeda maps, respectively, for the system model given by (63) and (64) with SNRs of 6 dB. The figures depict the relative values of $p(Y_0^n|x_n)$ as a function of x_n (for a fixed Y_0^n). The set of observations used in obtaining the figures was not the entire set $\{y_m\}_{m=0}^n$ but the smaller set $\{y_m\}_{m=n-10}^n$, that is the eleven consecutive observations prior to and including y_n. The center point in each figure corresponds to the actual value of the state x_n. In the figures, nesting of contours indicates increasing values of the likelihood functions. Alternative graphical representations of the data are provided in Figures 9 and 10 which are mesh plots of the same data used for Figures 7 and 8, respectively.

Clearly, the plotted likelihood functions do not have Gaussian shapes. Instead each has a distinct, thin, elongated region of high functional values with peaks and valleys scattered throughout this region. A first-order

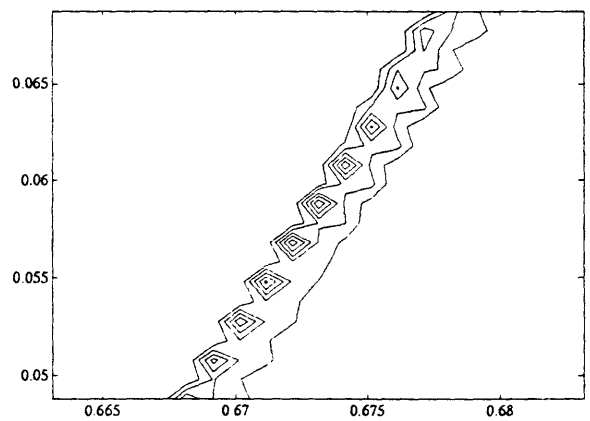

Figure 7: Contour plot of $p(Y_0^n|x_n)$ as a function of x_n for the Henon map

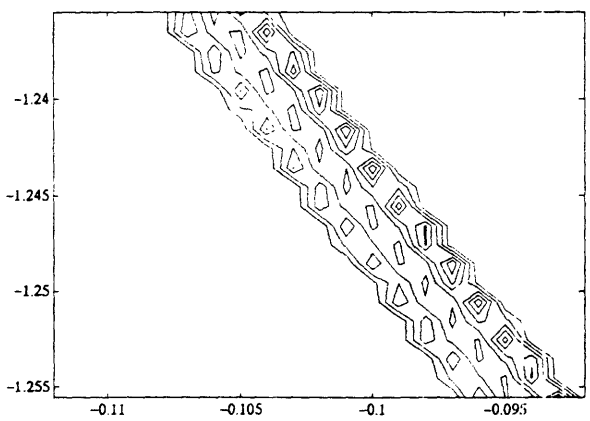

Figure 8: Contour plot of $p(Y_0^n|x_n)$ as a function of x_n for the Ikeda map

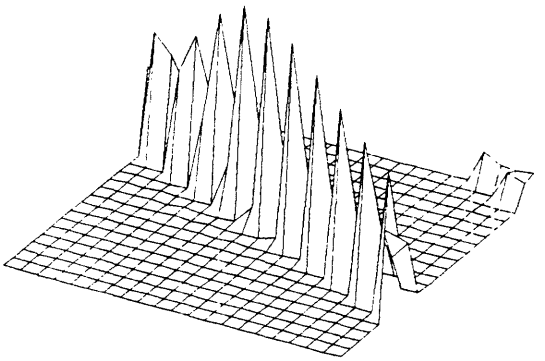

Figure 9: Mesh plot of $p(Y_0^n|x_n)$ as a function of x_n for the Henon map

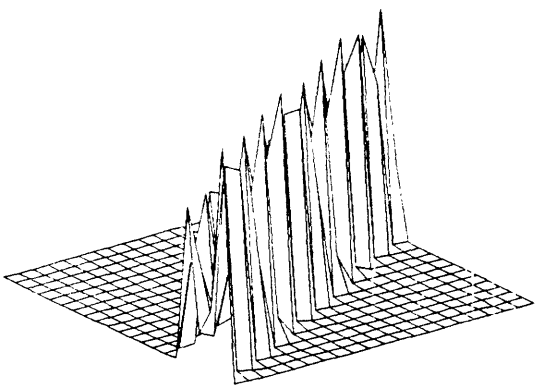

Figure 10: Mesh plot of $p(Y_0^n|x_n)$ as a function of x_n for the Ikeda map

analysis provides insight into this interesting property of the likelihood functions. Consider a single term S_i of the sum in the exponent of the likelihood function (91):

$$S_i(x_n) = \left[y_i - f^{-(n-i)}(x_n)\right]^T R^{-1} \left[y_i - f^{-(n-i)}(x_n)\right], \tag{92}$$

where for notational convenience, x_n rather than x_{n+1} (in (91)) is the state now under consideration. For a small deviation δ_n from the actual value of x_n, the following holds to first-order:

$$f^{-(n-i)}(x_n + \delta_n) \approx f^{-(n-i)}(x_n) + D\{f^{-(n-i)}(x_n)\}\delta_n, \tag{93}$$

where $D\{f^{-(n-i)}(x_n)\}$ is the derivative of $f^{-(n-i)}(\cdot)$ with respect to x_n (or equivalently, the Jacobian of $f^{-(n-i)}(\cdot)$ evaluated at x_n. Replacing x_n by $x_n + \delta_n$ in (92) and using (93) yields

$$\begin{aligned} S_i(x_n + \delta_n) \approx & \left[y_i - f^{-(n-i)}(x_n) - D\{f^{-(n-i)}(x_n)\}\delta_n\right]^T \\ & \times R^{-1} \left[y_i - f^{-(n-i)}(x_n) - D\{f^{-(n-i)}(x_n)\}\delta_n\right]. \end{aligned} \tag{94}$$

Since

$$y_n = x_n + v_n \tag{95}$$

and as in Section 4

$$R = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}, \tag{96}$$

then (94) reduces to

$$S_i(x_n + \delta_n) \approx \sum_{k=1}^{2} \frac{\left\{v_{n,k} - \left[D\{f^{-(n-i)}(x_n)\}\delta_n\right]_k\right\}^2}{\sigma_k^2} \tag{97}$$

$$= \sum_{k=1}^{2} \frac{v_{n,k}^2 - 2v_{n,k}\left[D\{f^{-(n-i)}(x_n)\}\delta_n\right]_k + \left[D\{f^{-(n-i)}(x_n)\}\delta_n\right]_k^2}{\sigma_k^2}, \tag{98}$$

where

$$v_n = [v_{n,1}, v_{n,2}]^T \tag{99}$$

$$D\{f^{-(n-i)}(x_n)\}\delta_n = \left[[D\{f^{-(n-i)}(x_n)\}\delta_n]_1, [D\{f^{-(n-i)}(x_n)\}\delta_n]_2\right]^T \tag{100}$$

Since v_n is zero-mean by assumption, the expected value of $S_i(x_n + \delta_n)$ is simply

$$E\left(S_i(x_n + \delta_n)\right) \approx \sum_{k=1}^{2} \frac{\sigma_k^2 + \left[D\{f^{-(n-i)}(x_n)\}\delta_n\right]_k^2}{\sigma_k^2}, \tag{101}$$

where $E(\cdot)$ is the expectation operator. Now the likelihood function $p(Y_0^n|x_n)$ is large if each term in the sum in the exponential is small, or equivalently (on average) if (101) is small for each i. Thus, the value of the likelihood function depends on the magnitudes of

$$D\{f^{-(n-i)}(x_n)\}\delta_n, \quad i = (0, \cdots, N). \tag{102}$$

A distinguishing property of an invertible chaotic system is the presence of a *stable* and an *unstable* manifold [3]. (Section 9 discusses these manifolds in detail). Basically, any small perturbation δ_n tangent to the stable manifold of a point x_n (on the chaotic attractor) remains small as the point and perturbed point propagate through the system dynamics, or more precisely [3],

$$\lim_{i\to\infty} \frac{1}{i} \log d(f^{-i}(x_n), f^{-i}(x_n + \delta_n)) < 0 \tag{103}$$

for δ_n on the stable manifold of x_n for the inverse system $f^{-1}(\cdot)$. If δ_n is small, an equivalent condition is the following:

$$\lim_{i\to\infty} |D\{f^{-i}(x_n)\}\delta_n| = 0 \tag{104}$$

where $|\cdot|$ means the magnitude.

Similarly, if δ_n is tangent to the unstable manifold of x_n for the inverse system $f^{-1}(\cdot)$, the perturbation grows in magnitude as x_n and $x_n + \delta_n$ propagate through the system dynamics. (However, the growth is bounded since dissipative chaotic systems such as the Henon and Ikeda maps have bounded attractors). By definition, the unstable manifold of a system is the stable manifold of the inverse of that system.

In light of the above and its relevance to (101), it appears that the thin, elongated regions in Figure 7 and 8 correspond to points near or along the stable manifolds of the inverses of the chaotic systems, which by definition are the unstable manifolds of the forward systems. The peaks and valleys

indicated in the figures are possibly due to higher-order effects not considered in the simple, first-order analysis performed above.

In summary, the likelihood function for the system model given by (63) and (64) with either the Henon or Ikeda map does not have a Gaussian shape. As a result, the *a posteriori* state density is far from Gaussian as well, which may explain why the extended Kalman filter performed poorly at state estimation in the experiments summarized in Section 4.

6 Smoothing

With many applications, improved state estimation occurs if one uses not only past observations but future observations as well, when estimating the state at a given time. With respect to the taxonomy given in Section 2.2, the use of future as well as past observations transforms the state-estimation problem from a filtering problem to a smoothing problem.

This section introduces and experimentally evaluates the *extended Kalman smoother* (EKS), a nonlinear smoother that combines the EKF with the Rauch-Tung-Striebel linear smoother. The section also experimentally evaluates the EKS when the system dynamics are initially unknown, but a "clean" reference orbit is available.

6.1 Smoothing Problems

Historically, researchers have focused on three classes of smoothing problems. The first, *fixed-point* smoothing, involves estimating the state vector x_n based on the observation set $Y_0^m = \{y_i\}_{i=0}^m$ for a "fixed" time n and increasing m (where $m > n$). The second, *fixed-lag* smoothing, involves estimating the state vector x_{n-L} based on the observation set $Y_0^n = \{y_i\}_{i=0}^n$ for each time n and a "fixed" lag L. The third, *fixed-interval* smoothing, involves estimating the state vector x_n based on the observation set $Y_0^N = \{y_i\}_{i=0}^N$ for all times n satisfying $0 \leq n \leq N$ (for a fixed N). This report considers only fixed-interval smoothing.

6.2 Optimal Fixed-Interval Smoothing

There are many approaches and associated algorithms for optimal (i.e., minimum error variance (MEV)), linear smoothing in general and optimal, fixed-interval, linear smoothing in particular [1, 9, 17]. Many of these approaches combine the output of a Kalman filter which sequentially processes the observations forward in time, with the output of a recursive filter which sequentially processes the observations backward in time. One such computationally efficient approach is the *Rauch-Tung-Striebel* smoother [9, 13]. Figure 11 provides the equations for this smoother (applicable to the system model given by (5) and (6)).

Forward Filter

Prediction Step

$$\hat{x}_{n+1|n} = F_n \hat{x}_{n|n} \tag{105}$$

$$P_{n+1|n} = F_n P_{n|n} F_n^T + G_n Q_n G_n^T \tag{106}$$

Measurement Update Step

$$\hat{x}_{n+1|n+1} = \hat{x}_{n+1|n} + K_{n+1}[y_{n+1} - H_{n+1}\hat{x}_{n+1|n}] \tag{107}$$

$$K_{n+1} = P_{n+1|n} H_{n+1}^T \left[H_{n+1} P_{n+1|n} H_{n+1}^T + R_{n+1}\right]^{-1} \tag{108}$$

$$P_{n+1|n+1} = [I - K_{n+1} H_{n+1}]\, P_{n+1|n} \tag{109}$$

Initialization

$$\hat{x}_{0|-1} = m_0 \tag{110}$$

$$P_{0|-1} = P_0. \tag{111}$$

Backward Filter and Smoothed State Estimator

Backward Filter Update Step

$$L_n = P_{n|n} F_n^T P_{n+1|n}^{-1} \tag{112}$$

$$P_n^s = P_{n|n} - L_n (P_{n+1|n} - P_{n+1}^s) L_n^T \tag{113}$$

Smoothed State Estimate

$$\hat{x}_{n|N} = \hat{x}_{n|n} + L_n (\hat{x}_{n+1|N} - \hat{x}_{n+1|n}) \tag{114}$$

Initialization

$$\hat{x}_{N|N} = \hat{x}_{N|N} \tag{115}$$

$$P_N^s = P_{N|N}. \tag{116}$$

Figure 11: The Rauch-Tung-Striebel fixed-interval smoother

As indicated by the smoothing equations, the "forward filter" is simply a Kalman filter. The "backward filter and smoothed state estimator" have little resemblance to a Kalman filter. However, this part of the smoother actually incorporates an information form of a Kalman filter, the output of which is combined with the output of the forward filter.

6.3 The Extended Kalman Smoother (EKS)

One can combine the EKF with the Rauch-Tung-Striebel smoother to produce a simple, (albeit non-optimal), nonlinear, fixed-interval smoother, hereafter referred to as the *extended Kalman smoother* (EKS). The combination entails first replacing the "forward filter" of the Rauch-Tung-Striebel smoother (given by (105)–(109)) with an EKF. Next. one exploits the fact that for the nonlinear system model given by (1) and (2), the EKF is simply a Kalman filter in which the matrices used in the filtering equations (F_n, G_n, and H_n in (105)–(109)) are leading terms in the Taylor series expansions of the nonlinear functions f_n, g_n, and h_n. As a result. the "backward filter and smoothed state estimator" of the Rauch-Tung-Striebel can be included without modification in the EKS. Figure 12 provides the equations for the EKS, which are applicable to the system model given by (1) and (2) with F_n, G_n, and H_n given by (19)–(21).

6.4 Computer Experiments

The EKS was evaluated on the same data as were used to evaluate the EKF. As with the EKF, the EKS performed better with an appropriately chosen driving-noise, covariance matrix Q in the smoothing equations. Table 3 depicts the performance results obtained with an appropriately chosen matrix Q for each noise level.

A comparison of Tables 2 and 3 reveals that the EKS performs much better than the EKF on the Henon and Ushiki maps and slightly better on the Ikeda map. The nonrigorous analysis in Section 7 and the interpretation of the Cramer-Rao error bounds in Section 9 explain why the use of past and future observations, rather than past observations alone, is essential when performing state estimation with chaotic systems.

Forward Filter

Prediction Step

$$\hat{x}_{n+1|n} = f_n(\hat{x}_{n|n}) \tag{117}$$

$$P_{n+1|n} = F_n P_{n|n} F_n^T + G_n Q_n G_n^T \tag{118}$$

Measurement Update Step

$$\hat{x}_{n+1|n+1} = \hat{x}_{n+1|n} + K_{n+1}\left[y_{n+1} - h_{n+1}(\hat{x}_{n+1|n})\right] \tag{119}$$

$$K_{n+1} = P_{n+1|n} H_{n+1}^T \left[H_{n+1} P_{n+1|n} H_{n+1}^T + R_{n+1}\right]^{-1} \tag{120}$$

$$P_{n+1|n+1} = \left[I - K_{n+1} H_{n+1}\right] P_{n+1|n} \tag{121}$$

Initialization

$$\hat{x}_{0|-1} = m_0 \tag{122}$$

$$P_{0|-1} = P_0. \tag{123}$$

Backward Filter and Smoothed State Estimator

Backward Filter Update Step

$$L_n = P_{n|n} F_n^T P_{n+1|n}^{-1} \tag{124}$$

$$P_n^s = P_{n|n} - L_n(P_{n+1|n} - P_{n+1}^s)L_n^T \tag{125}$$

Smoothed State Estimate

$$\hat{x}_{n|N} = \hat{x}_{n|n} + L_n(\hat{x}_{n+1|N} - \hat{x}_{n+1|n}) \tag{126}$$

Initialization

$$\hat{x}_{N|N} = \hat{x}_{N|N} \tag{127}$$

$$P_N^s = P_{N|N}. \tag{128}$$

Figure 12: The extended Kalman smoother (EKS)

Initial SNR (dB)	SNR Improvement (dB)					
	Henon Map		Ushiki Map		Ikeda Map	
	EKS		EKS		EKS	
	$x_{n,1}$	$x_{n,2}$	$x_{n,1}$	$x_{n,2}$	$x_{n,1}$	$x_{n,2}$
40	27.07	27.16	24.51	21.81	21.25	21.48
20	16.58	15.68	22.83	18.67	9.18	10.58
10	11.50	10.44	9.94	8.79	8.95	10.25
6	9.19	8.88	8.41	8.52	3.89	5.60
3	7.74	7.08	6.58	6.37	2.71	2.96

Table 3: Performance results with extended Kalman smoothing

6.5 EKS with Unknown Dynamics

All results presented thus far were obtained with known system dynamics. The question arises as to the usefulness of the EKS, when the system dynamics are initially unknown. This subsection considers the problem of unknown system dynamics for the special case in which not only the noisy set of observations (to which one wishes to apply the EKS) are available, but a noise-free reference orbit is available as well. By assumption, the reference orbit and noisy set of observations are both generated by the same chaotic system, but with different initial conditions.

With a given noise-free reference orbit and unknown system dynamics, we used the following procedure to apply the EKS to a noisy set of observations:

1. Given the present state estimate $\hat{x}_{n|n}$ obtained with the forward filter portion of the EKS, find the K nearest neighbors (in a Euclidean sense) to $\hat{x}_{n|n}$ in the reference orbit.

2. Using the K nearest neighbors and their immediate successors in the reference orbit, determine the parameters of an affine mapping which minimizes the one-step, squared, prediction error from the K neighbors to their successors.

3. Use the parameters of the affine mapping as an estimate of the system dynamics at $\hat{x}_{n|n}$.

4. Apply the forward filter portion of the EKS using these estimated dynamics to estimate $\hat{x}_{n+1|n}$ and subsequently $\hat{x}_{n+1|n+1}$.

5. Repeat the procedure until all observations have been processed

6. Apply the backward smoother and combined state estimator portions of the EKS using the estimates $\{\hat{x}_{n|n}\}$ and the estimated dynamics.

Since an affine mapping is simply a linear mapping plus an offset, the EKS reduces to a "linear", fixed-interval, Rauch-Tung-Striebel smoother in this special case of unknown system dynamics.

The use of an affine mapping (calculated using nearest neigbors) as an estimate of the dynamics at each point on a chaotic orbit is generally referred to as "locally linear prediction" and was apparently first proposed in [3] and later applied extensively in [4]. Two alternative state-estimation algorithms which use locally linear prediction are introduced in [4, 8]; both alternative algorithms treat state estimation as a constrained optimization problem. The application of locally linear prediction to other signals, including speech, along with extensions of the technique are discussed in [14, 15].

The state-estimation algorithms discussed in [10, 12] also use a reference orbit to estimate certain algorithm parameters. With these algorithms, the "phase space" is first quantized into a finite set of states with the system dynamics approximated as obeying a first-order Markov process in this quantized phase space. The reference orbit is used by both algorithms to estimate the transition probabilities between states; and it is used by the algorithm discussed in [12] to estimate the state output probabilities as well

Table 4 contains the performance results for the Henon and Ushiki maps that were obtained with a 5000-point reference orbit for each map and with the use of 10-nearest neighbors for estimating each set of affine parameters. A comparison of Tables 3 and 4 indicates that for the Henon map, the performance of the EKS with unknown dynamics is only slightly worse than the performance with known dynamics at each noise level. This is not surprising in light of the fact that the EKF used for the forward filter portion of the EKS is simply a linear Kalman filter applied to a "linearization" of the system dynamics. For the Ushiki map, the performance of the EKS with unknown dynamics is inexplicably much worse than the performance with known dynamics at low noise levels (high SNRs), but comparable at high noise levels (low SNRs).

Initial SNR (dB)	SNR Improvement (dB)			
	Henon Map		Ushiki Map	
	EKS		EKS	
	$x_{n,1}$	$x_{n,2}$	$x_{n,1}$	$x_{n,2}$
40	22.65	22.43	7.31	7.41
20	15.82	14.94	16.14	14.21
10	10.21	9.66	10.45	8.90
6	9.01	8.82	8.45	8.01
3	7.68	7.03	6.16	6.42

Table 4: Performance results with extended Kalman smoothing and unknown dynamics

7 Performance Analysis of the EKS

Section 5 strongly suggested that the likelihood function $p(x_n|Y_n)$ for an invertible chaotic system is ridgelike. It attributed this property to the unstable manifold of the system and suggested that this property was partially responsible for the poor performance of the EKF. This section continues the analysis begun in Section 5 by considering the likelihood functions that result when only future observations are used, and when both past and future observations are used. The discussion reveals that when both past and future observations are used, the likelihood function rapidly becomes impulse-like, which helps explain the superior performance of the EKS over the EKF for chaotic systems. This section also briefly discusses a heuristic state estimator that exploits the singular property of the likelihood function. Section 9 formalizes the intuitive interpretations provided in this section by deriving and interpreting Cramer-Rao bounds on the error covariance matrices of unbiased state estimators for chaotic systems.

7.1 The Likelihood Function with Future Observations

As briefly mentioned in Section 5, an invertible chaotic system has both a stable and an unstable manifold, and by definition the stable (unstable) manifold of the system is the unstable (stable) manifold of its inverse.

In light of this, consider a set of "future" observations Y_n^{n+m},

$$Y_n^{n+m} = \{y_i\}_{i=n}^{n+m} \tag{129}$$

where n is the "present" time.

Using a similar approach as that used earlier in deriving the *a posteriori* state density $p(x_n|Y_0^n)$ and likelihood function $p(Y_0^n|x_n)$, one can easily derive the likelihood function $p(Y_n^{n+m}|x_n)$ for the system model given by

$$x_{n+1} = f(x_n) \tag{130}$$

$$y_n = x_n + v_n, \tag{131}$$

where v_n is a zero-mean, Gaussian, white-noise process with constant, diagonal, covariance matrix R. The resulting likelihood function has the following form:

$$p(Y_n^{n+m}|x_n) \propto \exp\left\{-\sum_{i=n}^{n+m}\left[y_i - f^{(i-n)}(x_n)\right]^T R^{-1}\left[y_i - f^{(i-n)}(x_n)\right]\right\}. \quad (132)$$

In contrast to the likelihood function $p(Y_0^n|x_n)$ given by (91) which has terms involving iterates of the inverse function $f^{-1}(\cdot)$, the above likelihood function has terms involving iterates of $f(\cdot)$. Given the discussion and results presented in Section 5. one might expect a contour plot of $p(Y_n^{n+m}|x_n)$ for the Henon or Ikeda map to be similar to that of Figure 7 or 8, but with the thin, elongated region now corresponding to the stable manifold of $f(x_n)$, or equivalently the unstable manifold of $f^{-1}(x_n)$. Figures 13 and 14 confirm this expectation. The figures are contour plots of $p(Y_n^{n+m}|x_n)$ as a function of x_n at a single time n (for $m = 15$) for the Henon and Ikeda maps. The

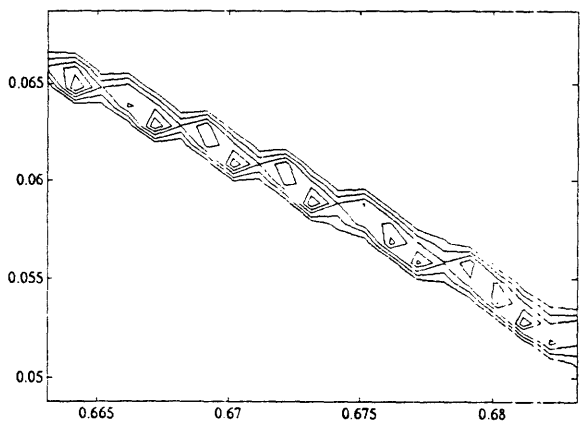

Figure 13: Contour plot of $p(Y_n^{n+m}|x_n)$ as a function of x_n for the Henon map

center point in each figure corresponds to the actual value of the state x_n.

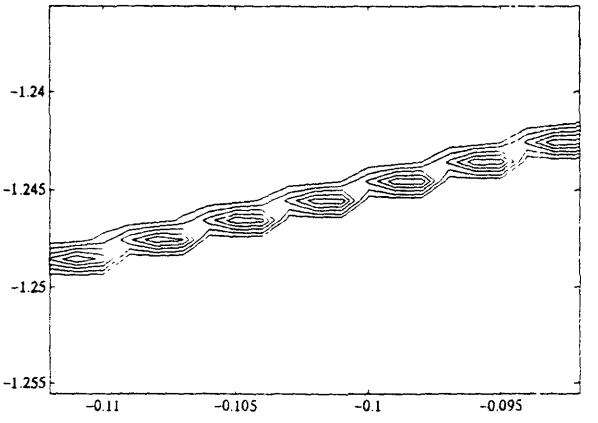

Figure 14: Contour plot of $p(Y_n^{n+m}|x_n)$ as a function of x_n for the Ikeda map

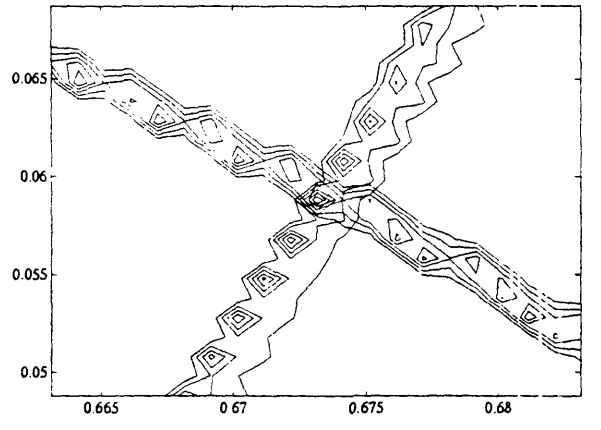

Figure 15: Superposition of Figures 7 and 13

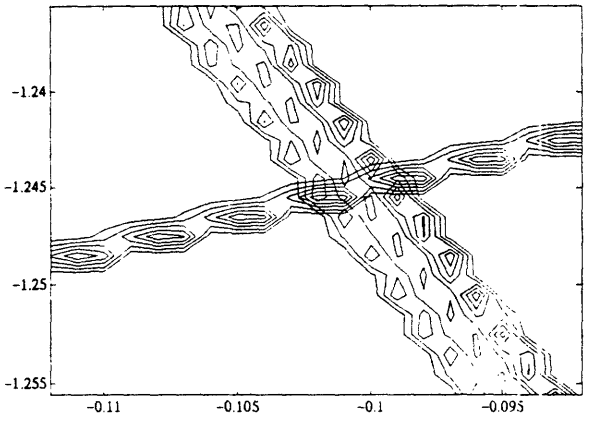

Figure 16: Superposition of Figures 8 and 14

Figure 15 is the superposition of Figures 7 and 13, and Figure 16 is the superposition of Figures 8 and 14. As indicated by Figures 15 and 16, the state x_n lies in the intersection of the two elongated regions in each figure. One would expect this, since the stable and unstable manifolds of a point on a chaotic attractor intersect at the point (and possibly elsewhere as well).

In [5, 6], an iterative state-estimation algorithm for two-dimensional, invertible, chaotic maps is discussed which exploits the known presence of a stable and unstable manifold at each point on a chaotic orbit. The algorithm first takes a fixed interval of noisy observation, $Y_0^N = \{y_i\}_{i=0}^N$, and calculates the one-step prediction error, $y_n - f(y_{n-1})$, for each pair of points in the interval. Each error is then decomposed into a component along the stable manifold and a component along the unstable manifold at the corresponding point. Finally, the algorithm uses a recursion running forward in time to reduce the error component at each point along the stable manifold and a recursion running backward in time to reduce the error component along the unstable manifold. The algorithm is then iterated using the resulting state estimates as the new observations. The algorithm appears to work reasonably well when the noise level is low (i.e., the initial SNR is large).

7.2 The Likelihood Function with Past and Future Observations

The question arises as to the properties of the likelihood function that results when both "past" and "future" observations are used, that is the likelihood function $p(Y_{n-r}^{n+m}|x_n)$ where

$$Y_{n-r}^{n+m} = \{y_i\}_{i=n-r}^{n+m} \tag{133}$$

Use of a similar approach as that used earlier to derive the *posteriori* state density $p(x_n|Y_0^n)$ and likelihood function $p(Y_0^n|x_n)$ yields the following expression for $p(Y_{n-r}^{n+m}|x_n)$ for the system model given by (130) and (131):

$$p(Y_{n-r}^{n+m}|x_n) \propto \exp\left\{-\sum_{i=n-r}^{n+m}\left[y_i - f^{(i-n)}(x_n)\right]^T R^{-1}\left[y_i - f^{(i-n)}(x_n)\right]\right\}. \tag{134}$$

Figures 17 and 18 are contour plots of this likelihood function for the Henon map with two sets of values for m and r, and Figures 19 and 20 are similar contour plots for the Ikeda map. The center point in each figure is

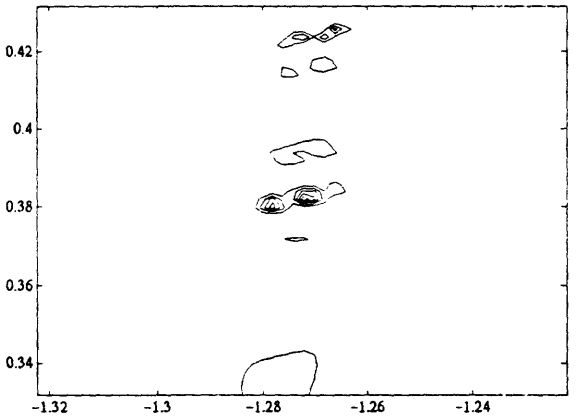

Figure 17: Contour plot of $p(Y_{n-r}^{n+m}|x_n)$ as a function of x_n for the Henon map (r=3, m=12)

the value of the actual state x_n. As indicated by Figures 18 and 20, the likelihood function rapidly becomes impulse-like.

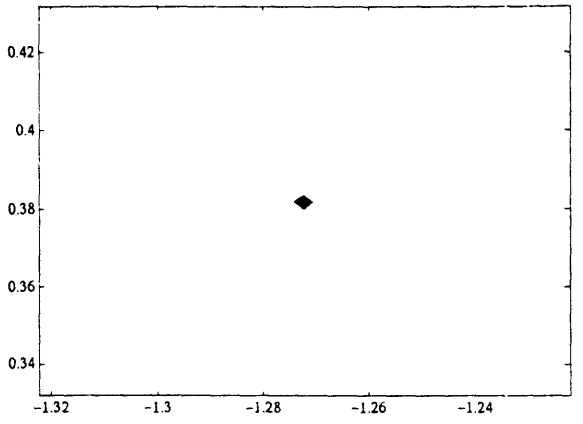

Figure 18: Contour plot of $p(Y_{n-r}^{n+m}|x_n)$ as a function of x_n for the Henon map (r=5, m=20)

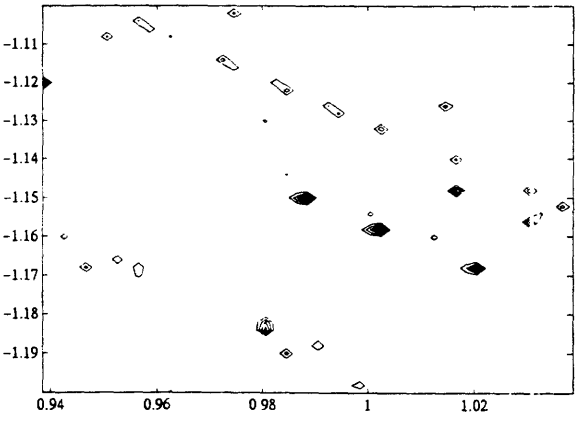

Figure 19: Contour plot of $p(Y_{n-r}^{n+m}|x_n)$ as a function of x_n for the Ikeda map (r=3, m=15)

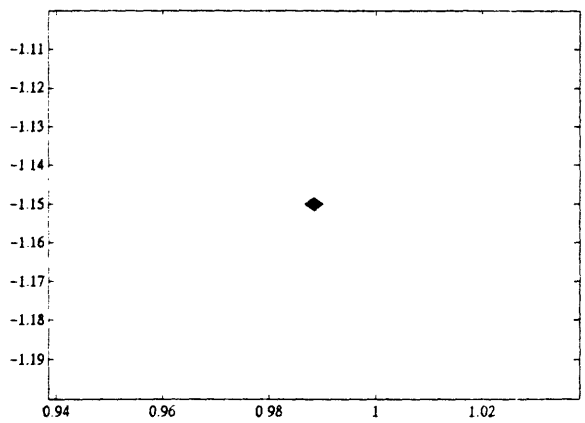

Figure 20: Contour plot of $p(Y_{n-r}^{n+m}|x_n)$ as a function of x_n for the Ikeda map (r=6, m=25)

By construction, a contour plot depicts the values of a function at only a finite set of grid points. Figures 21 and 22 depict the likelihood function $p(Y_{n-r}^{n+m}|x_n)$ for the same values of m and r as used for Figures 18 and 20, but with the center point now the perturbed state vector $x_n + [1.5 \times 10^{-4}, 1.5 \times 10^{-4}]^T$ and the actual state not one of the plotted points. (The grid spacing in both figures, as in the four earlier figures is .002). The figures further emphasize the impulse-like behavior of the likelihood function.

The impulse-like behavior of the likelihood function $p(Y_{n-r}^{n+m}|x_n)$ for chaotic systems suggests that state estimation with almost arbitrary precision is theoretically possible with these systems. Unfortunately, the likelihood function is nonlinear, and for smaller values of m and r, the function has multiple local minima. Furthermore, for larger values of m and r, the impulse-like behavior of $p(Y_{n-r}^{n+m}|x_n)$ precludes the use of popular, nonlinear, optimization methods to maximize the likelihood function. We are currently investigating alternative optimization methods for maximizing the likelihood function and will report the results in a future report.

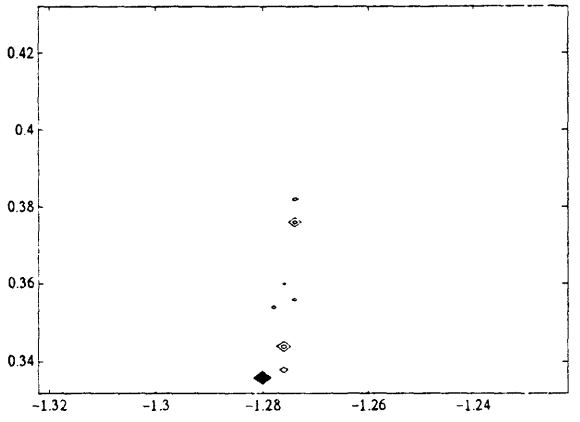

Figure 21: Contour plot of $p(Y_{n-r}^{n+m}|x_n)$ as a function of x_n for the Henon map with value for actual state not plotted (r=5, m=20)

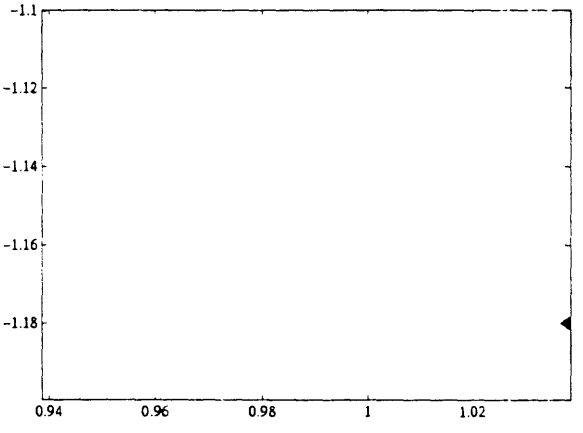

Figure 22: Contour plot of $p(Y_{n-r}^{n+m}|x_n)$ as a function of x_n for the Ikeda map with value for actual state not plotted (r=6, m=25)

7.3 Maximum Likelihood State Estimation

This subsection briefly considers practical heuristic approaches for exploiting the likelihood function $p(Y_{n-r}^{n+m}|x_n)$ to perform maximum-likelihood state estimation with chaotic systems. Several practical considerations must be dealt with when attempting maximum-likelihood state estimation. First, one must maximize a likelihood function that is nonlinear and rapidly becomes impulse-like as the number of observations before and after the present time increases. Second, given a finite sequential set of observations and the goal of estimating the state at each time for which there is an observation, in theory one need only maximize the likelihood function at a single time, since the system dynamics are completely deterministic. That is, given the maximum-likelihood state estimate $\hat{x}_n$ at time n, the maximum-likelihood state estimate at time m is simply

$$\hat{x}_m = f^{m-n}(\hat{x}_n). \tag{135}$$

However, because of finite-precision arithmetic and round-off errors inherent to all computers, the system dynamics are not truly deterministic in practice, and thus (135) is not true in practice.

Perhaps the simplest, albeit suboptimal way to deal with these two considerations is to calculate a separate likelihood function for each time n, to only use observations occurring near time n when determining the likelihood function, and to estimate the state at time n using only the corresponding likelihood function. The use of observations only near time n circumvents the problem of an impulse-like likelihood function. However, as shown by the figures in the preceding subsection, the resulting likelihood function will be multimodal. In light of this, a simple, potentially useful, but suboptimal approach to maximize the likelihood function is a "brute-force" grid search. That is, one evaluates the likelihood function at a finite set of points and then uses the point yielding the maximum value as the estimated state.

We applied this simple, heuristic, maximum-likelihood state estimation approach to the Henon and Ushiki maps. In particular, the state at time n was estimated by first calculating the likelihood function based on the six observations immediately before time n, the observation at time n, and the six observations immediately after time n. Next the likelihood function was evaluated at a finite set of test points. The test points were those on a four-thousand point orbit with arbitrarily selected initial starting point. We used

the test point that maximized the likelihood function as the state estimate at time n. In light of the expression for $p(Y_{n-r}^{n+m}|x_n)$ given by (134) and the method for choosing the test points, the estimation procedure essentially entailed finding for each time n, the segment of the four-thousand point orbit that best "fit" (in a least squares sense) the noisy observations near time n.

Table 5 indicates the improvement in SNR obtained with this estimation approach on one hundred point segments of the Henon and Ushiki maps for various noise levels. Surprisingly, despite its simplicity, the heuristic ap-

ORIGINAL SNR (dB)	SNR Improvement (dB)			
	Henon Map		Ushiki Map	
	EKS		EKS	
	$x_{n,1}$	$x_{n,2}$	$x_{n,1}$	$x_{n,2}$
20	12.48	14.88	10.46	4.98
10	12.98	13.17	9.17	11.98
6	11.82	6.85	11.82	9.81
3	14.02	11.48	13.26	9.76

Table 5: Performance results for heuristic, maximum likelihood state estimation with the Henon and Ushiki maps

proach works reasonably well, with the results superior to those obtained with the extended Kalman smoother at most SNR levels. By itself, this simple approach probably does not constitute a useful, robust, state-estimation algorithm. However, it illustrates the value of the likelihood function $p(Y_{n-r}^{n+m}|x_n)$ for performing state estimation. Also, it might be a useful component of an effective state-estimation algorithm.

An alternative technique for using the likelihood function for state estimation is discussed in [4]. This technique uses a linear approximation of the system dynamics, resulting in a closed-form solution to the problem of maximizing the likelihood function. The technique purportedly works well when the initial SNR is large.

Because of the Gaussian assumption on the additive white noise, the summation in the exponential of the likelihood function also arises with a constrained least-squares formulation of the state-estimation problem [5].

Unfortunately, this formulation of the problem offers little additional insight into maximizing the likelihood function, or equivalently minimizing the constrained least-squares cost function in [5]. As a consequence, the state-estimation algorithm discussed in [5] uses a linear approximation of the system dynamics (similar to the approximation used in [4]) to simplify the minimization problem.

8 Discrimination with Chaotic Systems

As the preceding sections implicitly suggest, state estimation with chaotic systems remains a challenging, unsolved problem even when the system dynamics are known. In contrast, this section shows that discrimination among a "finite" number of known states is relatively simple, even when the number of states is large and the SNR is low. The section also shows that discrimination among a "finite" number of known systems is also relatively simple, even with a large number of almost identical systems. In fact, the same properties of chaotic systems that complicate the problem of state estimation, simplify the problems of state and parameter discrimination.

The section begins by demonstrating that nearly perfect discrimination among a finite set of initial states is possible even with SNRs as low as -20 dB. It continues by demonstrating that nearly perfect discrimination among a finite set of similar systems, given a finite known set of initial states, is also possible even with extremely low SNRs. Next, the section shows that robust state discrimination is still possible even with multiplicative noise. The section concludes by discussing potential, practical applications of these results. A comprehensive analysis of both the experimental results presented in this section, as well as the general problems of state and parameter discrimination, will be provided in a future technical report.

8.1 State Discrimination

As suggested by the increasingly impulse-like behavior of the likelihood function $p(Y_{n-r}^{m+n}|x_n)$ as the number of observations grows, a chaotic system exhibits extreme sensitivity to the state x_n. As a result, robust discrimination among a finite number of states is possible even with extremely low SNRs. In fact, this subsection and the next experimentally confirm that chaotic orbits are like "snowflakes", in that no two orbits are identical and are readily distinguishable given "enough" points on the orbits.

Perhaps surprisingly, the "dynamical noise" caused by round-off errors and finite precision arithmetic on all computers actually assists in state and parameter and discrimination. The reason is simple. Theoretically, a stable manifold is associated with each point on a chaotic attractor. As explained earlier, the orbits starting from this point and from each point on its stable manifold eventually converge. However, in light of computer round-off errors,

the experimental existence of a stable manifold for any point on a chaotic attractor is questionable. The practical implication is that experimentally for "invertible" chaotic systems, the orbits evolving from any two starting points may become "close", but they will not "remain" close.

Table 6 indicates the robust state discrimination achievable with chaotic systems. The table provides the probability of correctly discriminating among 4000 initial states for the Henon map at various SNR levels for orbit lengths of 100, 1000, and 4000 points.

SNR (db)	Probability of Correct Decision		
	100 Points	1000 Points	4000 Points
-3	1.00	1.00	1.00
-6	.80	1.00	1.00
-10	.20	1.00	1.00
-15	.05	.85	1.00
-20	0	.30	1.00

Table 6: Probability of correctly discriminating among 4000 initial states (average for 20 independent trials at each SNR level)

The results were generated as follows. First, the 4000 initial states were selected by generating a 5000 point Henon orbit with an initial state of (0,0) and dropping the first 1000 points. Since the Henon map is a two-dimensional map, each of these points was a two-element vector. Next, a small perturbation of .001 was subtracted from the first component of each these points, to ensure that most (if not all) of the orbits emanating from these points were distinct. These perturbed 4000 points constituted the set of possible initial states.

For each trial, one of the 4000 points was selected at random, an orbit generated with that point as the initial state, and Gaussian, white noise added to the orbit. The first component of this noise-corrupted orbit constituted the "observation" data. Next, a likelihood test was applied to the "observation" data to choose among the initial states. As discussed earlier in this paper,

with additive, Gaussian, white noise, the likelihood for each initial state was simply the sum of squared errors between observation points and corresponding points on the orbit associated with that initial state. (Since the "observation" data consisted only of the first component of each two-element orbit point, only the first component of the each orbit point associated with each initial state was used in the likelihood calculation). The likelihood test consisted of simply choosing the initial state among the 4000 for which the corresponding sum of squared errors was smallest. Each probability shown in the table is the average number of correct decisions for twenty independent trials at the corresponding SNR. (Note that the probability of choosing the correct initial state by simply guessing is .00025).

The table shows that discrimination at extremely low SNRs is possible if "enough" data points are available. (A rigorous analytic definition of "enough" will be provided in a future report). The results support the earlier assertion that distinct chaotic orbits remain distinct.

8.2 Parameter Discrimination

As with state discrimination, robust discrimination among a finite number of known, similar, chaotic systems is also possible if there are a finite number of known initial states. The results reported in this section were obtained with the family of Henon maps given by

$$x_{n+1,1} = 1 - \alpha x_{n,1}^2 + x_{n,2} \quad (136)$$

$$x_{n+1,2} = .3x_{n,1}, \quad (137)$$

in which α was the variable parameter which distinguished the different systems. Table 7 depicts the results of discriminating among one hundred equally-spaced values of α between 1.39 and 1.4 with orbit lengths of 100, 1000, and 4000 points. The macroscopic properties of the orbits and attractors corresponding to many (if not all) of these parameter values are almost indistinguishable. The results shown in the columns for the 100 and 1000 point length orbits were obtained using the same 4000 initial states as were used for obtaining the results in Table 6. Because of computational considerations, the results shown in the column for the 4000 point length orbit were obtained using only the first 2000 of these possible initial states.

Likelihood tests, similar to those described in the preceding subsection, were used to discriminate among the parameter values. For each trial, a

SNR (db)	Probability of Correct Decision		
	100 Points	1000 Points	4000 Points
-3	.90	1.00	1.00
-6	.15	1.00	1.00
-10	0	1.00	1.00
-15	0	.50	1.00
-20	0	.10	.85

Table 7: Probability of correctly discriminating among 100 parameter values (average for 20 independent trials at each SNR level)

parameter value and initial state were selected at random, an orbit generated, Gaussian, white noise added to the orbit, and the first component of the noise-corrupted orbit retained as the "observation" data. Next, for each possible parameter value, a likelihood test was conducted to determine the most likely initial state for that parameter value. Finally, the parameter value with the "most likely" initial state (i.e., smallest sum of squared errors) was chosen as the parameter for that trial. Each probability shown in the table is the average number of correct decisions for twenty independent trials at the corresponding SNR. (Note that the probability of choosing the correct parameter value by simply guessing is .01). As suggested by the performance results in the table, discrimination among the 100 closely spaced parameter values is possible even with SNRs as low as -20 dB.

8.3 Discrimination with Multiplicative Noise

Preliminary results suggest that robust state and parameter discrimination with chaotic systems is possible not only with additive noise, but with multiplicative noise as well. Table 8 depicts the results of discriminating among the same 4000 initial states as were used for Table 6 based on a 100 point length orbit. In light of the confusion in defining SNR with multiplicative noise, the performance is indexed by the standard deviation of the multiplicative noise. (For comparison, the standard deviation of the first component

Std. Dev. of Noise	Probability of Correct Decision 100 Points
.71	1.00
2.89	1.00
7.79	.85
74.9	.10

Table 8: Probability of correctly discriminating among 4000 initial states (average for 20 independent trials at each noise level) with multiplicative noise

of the Henon map is approximately .71).

We will provide details of the experimental procedure in a future report. Basically, we used the same likelihood test procedure as discussed earlier, but applied it in the "log" domain. To avoid the need for complex logarithms, we added a constant positive bias to the uncorrupted orbit data so that all values would be positive. Each multiplicative noise term was the sum of a small positive bias and the squared value of a sample from a Gaussian, white-noise process.

Additional experiments suggest that robust state and parameter discrimination is still possible even when both additive and multiplicative noise is present.

8.4 Applications

In light of the encouraging discrimination results presented in the preceding three subsections, chaotic systems may be useful for several communication applications. For example, one might exploit the robust parameter discrimination capability to provide secure binary or m-ary communication over noisy channels. Each of m parameter values for a family of chaotic systems would correspond to one of m different signals to be transmitted. One would choose the parameter values so that the macroscopic properties of the orbits and attractor corresponding to each were nearly identical. In addition, the values for N allowable initial states would also be chosen. To transmit each

signal value, one would generate a fixed length orbit segment with the appropriate parameter value. The initial state for the segment would be the one of N allowable initial states that was "closest" to the final state of the preceding segment. If N were sufficiently large and the N known states properly distributed on the chaotic attractor, no discernible discontinuities would occur at the intersection of these segments. In addition, white noise could also be added to each orbit segment to thwart unwanted interception of the transmitted information. As long as the m possible parameter values and N possible initial states were known at the receiver, one could apply the simple likelihood test (discussed earlier) to each received noisy segment to determine the correct parameter value and thus the signal value corresponding to that segment.

The discussion and preliminary results in the preceding subsection suggest that chaotic systems might be useful as modulating carriers. For example, to transmit a signal sequence, one might use the sequence to modulate contiguous orbit segments from a known chaotic system. The initial state for each segment would be the allowable initial state closest to the final state of the preceding segment. As discussed above, if N were sufficiently large and the N known states properly distributed on the chaotic attractor, no discernible discontinuities would occur at the intersection of these segments. One could apply the simple likelihood test mentioned earlier (in the log domain) to each received segment to determine the initial state of the chaotic carrier and thus every value of the carrier for that segment. The original signal sequence could then be recovered by exponentiating the difference of the logarithm of the received signal and the logarithm of the chaotic carrier at each point. The use of different orbit segments of fixed length rather than a single orbit would enable the receiver to resynchronize to the chaotic carrier at the beginning of each segment.

9 Cramer-Rao Bounds

This section extends the informal analysis in Sections 5 and 7. Specifically, the section derives bounds on the error covariance matrix for an unbiased estimator for each of three estimation problems with chaotic systems. The analysis supports the earlier assertion that the simultaneous existence of both stable and unstable manifolds, at each point on the attractor of an invertible chaotic system, is an extremely important and useful property for state estimation.

9.1 Preliminaries

Throughout this section we use the following system model, which notationally is slightly different than the model used in Sections 5 and 7.

$$x_{n+1} = f(x_n) \tag{138}$$
$$y_n = h(x_n) \tag{139}$$
$$z_n = y_n + v_n. \tag{140}$$

The first equation is the *state equation.* In this deterministic equation, x_n is the $\mathcal{N}$-dimensional *state vector* and $f(x_n)$ is a known, discrete-time, chaotic system. The second equation is the *output equation.* In this equation, y_n is the $\mathcal{P}$-dimensional *output vector*, and $h(x_n)$ is a known, possibly nonlinear function of the state. The third equation is the *observation equation.* In this equation, z_n is the $\mathcal{P}$-dimensional *observation vector* used for estimating the state and output, and and v_n is a $\mathcal{P}$-dimensional, zero-mean, Gaussian white-noise process known as the *observation noise*, which by assumption is independent of the initial state x_0.

We seek error bounds for three related state-estimation problems. The first consists of using a set of $N+1$ sequential noisy observations $\{z_n\}_{n=0}^N$ to estimate the initial state x_0. The second consists of using the observations $\{z_n\}_{n=0}^N$ to estimate the output y_m for a specific time m where $0 \leq m \leq N$. We also consider a special case of this problem which consists of using the observations $\{z_n\}_{n=0}^N$ to estimate the state x_m for a specific time m where $0 \leq m \leq N$. The third problem consists of using the observations $\{z_n\}_{n=0}^N$ to estimate all of the outputs $\{y_i\}_{i=0}^N$.

The derivation of all three error bounds uses $p(Z_0^N|x_0)$, where $Z_0^N \equiv \{z_i\}_{i=0}^N$. This is the conditional probability density of a set of sequential observations given the initial state x_0. As noted in an earlier section, with state (or output) estimation one often treats $p(Z_0^N|x_0)$ as a function of the conditioning variable x_0. With this treatment, $p(Z_0^N|x_0)$ is known as a *likelihood function*. For the model given by (138)–(140), the maximum likelihood estimate of x_0 is that value of x_0 which maximizes $p(Z_0^N|x_0)$ for a given Z_0^N. Also, since the state equation (138) is deterministic, if $\hat{x}_{ML}$ denotes the maximum likelihood estimate of x_0 given the observations Z_0^N, then the the maximum likelihood estimate of x_m given the observations Z_0^N is simply $f^m(\hat{x}_{ML})$, and the maximum likelihood estimate of y_m given Z_0^N is simply $h(f^m(\hat{x}_{ML}))$.

The error bounds derived in the next subsection apply only to unbiased estimators, $\hat{x}_0$ and $\hat{y}_m$, that is those for which the following conditions hold:

$$E(\hat{x}_0|x_0) = \int \hat{x}_0 \, p(Z_0^N|x_0) \, dZ_0^N = x_0 \tag{141}$$

$$E(\hat{y}_m|x_0) = \int \hat{y}_m \, p(Z_0^N|x_0) \, dZ_0^N = y_m. \tag{142}$$

The above integrals make sense since the estimators $\hat{x}_0$ and $\hat{y}_m$ are, in general, functions of the observations Z_0^N. Also, both expectations are conditioned on x_0. For the second expectation, this conditioning on x_0 is more restrictive than conditioning on the actual output y_m (since $y_m = h(f^m(x_0))$; but the error bound on $\hat{y}_m$ is more useful and amenable to interpretation with the more restrictive conditioning.

We define the conditional error covariance matrices, $P_N(\hat{x}_0)$ and $P_N(\hat{y}_n)$, as follows:

$$P_N(\hat{x}_0) \equiv E\left[(\hat{x}_0 - x_0)(\hat{x}_0 - x_0)^T|x_0\right] \tag{143}$$

$$P_N(\hat{y}_m) \equiv E\left[(\hat{y}_m - y_m)(\hat{y}_m - y_m)^T|x_0\right]. \tag{144}$$

We are particularly interested in the traces of $P_N(\hat{x}_0)$ and $P_N(\hat{y}_m)$ (i.e., the sum of the diagonal elements). These traces are the sum of the conditional error variances for the individual components of the estimators.

The next two subsections derive and interpret lower bounds on both these error covariance matrices and their traces. The Cramer-Rao inequality provides one such lower bound, which is commonly referred to as the "Cramer-Rao bound". One advantage of the Cramer-Rao bound over other bounds is

the ease in explicitly deriving it for certain estimation problems (including the problems considered in this paper). In addition, an unbiased estimator for which the error covariance matrix satisfies this bound with equality, is also the maximum-likelihood estimator, that is the one that maximizes the appropriate likelihood function.

The general form of the Cramer-Rao inequality for $P_N(\hat{x}_0)$ is the following [16]:

$$P_N(\hat{x}_0) \geq J_N^{-1}(x_0), \tag{145}$$

where $J_N(x_0)$ is known as the *Fisher information matrix* and is given by

$$J_N(x_0) = E\left\{ D_{x_0}^T \left\{ \ln p(Z_0^N | x_0) \right\} D_{x_0} \left\{ \ln p(Z_0^N | x_0) \right\} \middle| x_0 \right\}, \tag{146}$$

and where $D_{x_0}\{\cdot\}$ denotes the derivative of the bracketed argument with respect to the vector x_0. (If the bracketed argument is a scalar, the result is an $\mathcal{N}$-element row vector, where $\mathcal{N}$ is the dimension of x_0).

The next subsection derives and interprets the Cramer-Rao bound on $P_N(\hat{x}_0)$ by calculating the Fisher information matrix $J_N(x_0)$. The following subsection then uses this bound to derive a lower bound on $P_N(\hat{y}_m)$. The subsections establish a close relation between these bounds, the Lyapunov exponents of the chaotic system $f(\cdot)$ in (138), and the stable and unstable manifolds of $f(\cdot)$. To aid in the interpretation, the subsections also consider error bounds for the following linear model closely related to the model given by (138)—(140):

$$x_{n+1} = Fx_n \tag{147}$$

$$y_n = Hx_n \tag{148}$$

$$z_n = y_n + v_n. \tag{149}$$

where the constant $(\mathcal{N} \times \mathcal{N})$-matrix F replaces the nonlinear function $f(\cdot)$ and the constant $(\mathcal{P} \times \mathcal{N})$-matrix H replaces the nonlinear function $h(\cdot)$.

9.2 The Cramer-Rao Bound on $P_N(\hat{x}_0)$

This subsection establishes the Cramer-Rao bound on the error covariance matrix $P_N(\hat{x}_0)$ for unbiased estimators $\hat{x}_0$ of the initial state x_0 given a set of $N+1$ sequential observations $Z_0^N = \{z_i\}_{i=0}^N$. It also provides a qualitative interpretation of the bounds it establishes.

As noted in the previous subsection, the Cramer-Rao inequality for the error covariance matrix $P_N(\hat{x}_0)$ of any unbiased estimator $\hat{x}_0$ for the initial state x_0 is the following:

$$P_N(\hat{x}_0) \geq J_N^{-1}(x_0), \tag{150}$$

where $J_N(x_0)$, the *Fisher information matrix*, is given by

$$J_N(x_0) = E\left\{ D_{x_0}^T \left\{ \ln p(Z_0^N|x_0) \right\} D_{x_0} \left\{ \ln p(Z_0^N|x_0) \right\} \middle| x_0 \right\}, \tag{151}$$

and where D_{x_0} denotes the derivative with respect to the vector x_0. Since $\ln p(Z_0^N|x_0)$ is a scalar, $D_{x_0}\{\ln p(Z_0^N|x_0)\}$ is an $\mathcal{N}$-element row vector. Establishing the Cramer-Rao bound on $P_N(\hat{x}_0)$ requires that $J_N(x_0)$ be expressed in terms of the parameters of the system model given by (138)—(140).

Since $\{v_n\}$ is a zero-mean, Gaussian, white-noise process, then for any observation z_n,

$$p(z_n|y_n) = p\left([v_n = z_n - y_n]|y_n\right) = N(y_n, R), \tag{152}$$

where $N(m, \Sigma)$ indicates the normal density with mean vector m and covariance matrix Σ. However, since

$$y_n = h(x_n) = h(f^n(x_0)), \tag{153}$$

it follows that

$$p(z_n|x_0) = p\left(z_n|[y_n = h(f^n(x_0)]\right) = N\left([h(f^n(x_0))], R\right). \tag{154}$$

Therefore,

$$\ln p(z_n|x_0) \;=\; C_n - \frac{1}{2}\left[z_n - h(f^n(x_0))\right]^T R^{-1}\left[z_n - h(f^n(x_0))\right], \tag{155}$$

where C_n is a normalizing constant and R is the covariance matrix of v_n. Therefore, we can express $\ln p(Z_0^N|x_0)$ as

$$\ln p(Z_0^N|x_0) = C_N - \frac{1}{2}\sum_{i=0}^{N}\left[z_i - h(f^i(x_0))\right]^T R^{-1}\left[z_i - h(f^i(x_0))\right], \tag{156}$$

where C_N is a normalizing constant. Note that (156) is a sum of weighted squared-error terms in which each error is the difference between an observation vector z_i and the corresponding output vector y_i (for a given value of

x_0), and the weighting matrix is the inverse of the covariance matrix of the noise vector v_i.

Applying the rules of vector calculus and using the fact that R^{-1} is symmetric, we can express $D_{x_0}\{\ln p(Z_0^N|x_0)\}$ as

$$D_{x_0}\{\ln p(Z_0^N|x_0)\} = \sum_{i=0}^{N}\left[z_i - h(f^i(x_0))\right]^T R^{-1} D_{x_0}\{h(f^i(x_0))\} \quad (157)$$

$$= \sum_{i=0}^{N}\left[z_i - h(f^i(x_0))\right]^T R^{-1} \times D\{h(f^i(x_0))\} D_{x_0}\{f^i(x_0)\} \quad (158)$$

$$= \sum_{i=0}^{N}\left[z_i - h(f^i(x_0))\right]^T R^{-1} \times D\{h(f^i(x_0))\}\, T_{x_0}^i \quad (159)$$

where

$$T_{x_0}^i \equiv \begin{cases} I_{\mathcal{N}}, \ \text{(the } (\mathcal{N}\times\mathcal{N})\text{-identity matrix)} & i = 0 \\ \prod_{j=1}^{i} D\{f(f^{i-j}(x_0))\}, & i > 0 \\ \prod_{j=i+1}^{0} D\{f^{-1}(f^j(x_0))\}, & i < 0 \end{cases} \quad (160)$$

and where $D\{\cdot\}$ without a subscript indicates the derivative of the function that follows with respect to the argument of the function. For example, $D\{h(f^i(x_0))\}$ indicates the derivative of $h(\cdot)$, with respect to $f^i(x_0)$.

Substituting (159) in (151) yields the following expression for the Fisher information matrix $J_N(x_0)$:

$$J_N(x_0) = E\left\{\sum_{i=0}^{N}\left[T_{x_0}^i\right]^T D^T\{h(f^i(x_0))\}R^{-1}\left[z_i - h(f^i(x_0))\right] \times \sum_{k=0}^{N}\left[z_k - h(f^k(x_0))\right]^T R^{-1} D\{h(f^k(x_0))\}T_{x_0}^k \,\middle|\, x_0\right\} \quad (161)$$

$$= \sum_{i=0}^{N}\sum_{k=0}^{N}\left[T_{x_0}^i\right]^T D^T\{h(f^i(x_0))\}R^{-1} \times E\left\{\left[z_i - h(f^i(x_0))\right]\left[z_k - h(f^k(x_0))\right]^T \,\middle|\, x_0\right\} \times R^{-1} D\{h(f^k(x_0))\}T_{x_0}^k. \quad (162)$$

Since $\{v_n\}$ is a zero-mean, white-noise sequence

$$E\left\{\left[z_i - h(f^i(x_0))\right]\left[z_k - h(f^k(x_0))\right]^T \middle| x_0\right\} = \delta_{i,k}\, R, \tag{163}$$

where

$$\delta_{i,k} = \begin{cases} 1, & i = k \\ 0, & \text{otherwise} \end{cases} \tag{164}$$

Applying (163) to (162) yields

$$J_N(x_0) = \sum_{i=0}^{N} \left[T_{x_0}^i\right]^T D^T\{h(f^i(x_0))\} R^{-1} D\{h(f^i(x_0))\}\, T_{x_0}^i \tag{165}$$

$$= \sum_{i=0}^{N} D_{x_0}^T\{h(f^i(x_0))\} R^{-1} D_{x_0}\{h(f^i(x_0))\}. \tag{166}$$

The above equation for $J_N(x_0)$ has a revealing structure when the function $h(\cdot)$ in (139) and R, the covariance matrix of v_n in (140), have the following forms:

$$h(x) = x \tag{167}$$

$$R = \sigma^2 I_{\mathcal{N}}, \text{ where } I_{\mathcal{N}} \text{ is the } (\mathcal{N} \times \mathcal{N})\text{-identity matrix.} \tag{168}$$

With these restrictions on $h(\cdot)$ and R, (165) reduces to

$$J_N(x_0) = \frac{1}{\sigma^2} \sum_{i=0}^{N} \left[T_{x_0}^i\right]^T T_{x_0}^i. \tag{169}$$

The above expression for $J_N(x_0)$ is closely related to the expression used to define the *Lyapunov* or *characteristic* exponents of the system $f(\cdot)$. In particular, the Lyapunov exponents of $f(\cdot)$ are defined as the natural logarithms of the eigenvalues of the following matrix [3]:

$$\Lambda = \lim_{i \to \infty} \left(\left[T_{x_0}^i\right]^T T_{x_0}^i\right)^{\frac{1}{2i}}. \tag{170}$$

A comparison of (169) and (170) reveals that the Fisher information matrix $J_N(x_0)$ consists of a sum of partial products of the infinite product of matrices which determines the Lyapunov exponents of the chaotic system $f(\cdot)$. As a result, the Lyapunov exponents of $f(\cdot)$ strongly influence both the eigenvalues

of $J_N(x_0)$ for each N, as well as the asymptotic properties of these eigenvalues as N goes to infinity.

The Lyapunov exponents determine the "local" expansion rates along different directions at each point x on a chaotic attractor. Local expansion refers to the change in size that a small perturbation of x undergoes when acted on by the system. Positive Lyapunov exponents cause a perturbation of x (in certain directions) to increase in size, whereas negative Lyapunov cause a perturbation to decrease in size when acted on by the system. That is, given an infinitesimal perturbation δx_0 of x_0, the perturbation δ_n after n iterates of the system (i.e., the distance from x_n) is given by

$$\delta x_n \approx T^n_{x_0} \delta x_0 \approx \delta x_0 \exp(n\lambda), \tag{171}$$

where λ is the appropriate Lyapunov exponent for the direction of δx_0.

Associated with the negative Lyapunov exponents at each point x on a chaotic attractor is a nonlinear manifold V^s_x known as the *stable manifold* and defined as [3]

$$V^s_x = \left\{ y : \lim_{i \to \infty} \frac{1}{i} \log d(f^i(x), f^i(y)) < 0 \right\}, \tag{172}$$

where $d(\cdot, \cdot)$ is the Euclidean metric. Intuitively, V^s_x is the set of points with orbits that eventually converge to the orbit of x. Perhaps surprisingly, this nonlinear manifold is tangent at x to the linear manifold (i.e., subspace) spanned by the eigenvectors of Λ in (170) corresponding to the negative Lyapunov exponents (i.e., the eigenvalues of Λ with magnitudes less than unity).

Similarly, if $f(\cdot)$ is invertible, associated with the positive Lyapunov exponents is a nonlinear manifold V^u_x defined as

$$V^u_x = \left\{ y : \lim_{i \to \infty} \frac{1}{i} \log d(f^{-i}(x), f^{-i}(y)) < 0 \right\}. \tag{173}$$

Intuitively, V^u_x is the set of points with orbits that eventually converge to the orbit of x backward in time. This nonlinear manifold is **NOT** tangent at x to the linear manifold spanned by the eigenvectors of Λ in (170) corresponding to the positive Lyapunov exponents (i.e., the eigenvalues of Λ with magnitudes greater than unity). However, it is tangent at x to the

linear manifold spanned by the eigenvectors corresponding to the negative Lyapunov exponents of the following matrix

$$\Lambda^- = \lim_{i\to\infty}\left(\left[T_{x_0}^{-i}\right]^T T_{x_0}^{-i}\right)^{\frac{1}{2i}} \tag{174}$$

In fact, the Lyapunov exponents of Λ^- are the negatives of the Lyapunov exponents of Λ. Equivalently, the eigenvalues of Λ are the reciprocals of the eigenvalues of Λ^-.

For discrete-time linear systems (such as the one given by (147)), the Lyapunov exponents are simply the logarithms of the eigenvalues of the one-step, state transition matrix (F in (147)). The "unstable manifold" is the subspace spanned by the eigenvectors corresponding to eigenvalues with magnitudes greater than one, while the "stable manifold" is the subspace spanned by the eigenvectors corresponding to eigenvalues with magnitudes less than one.

For both linear and nonlinear systems, the stable and unstable manifolds are invariant. This simply means that if y is any point lying on either the stable (unstable) manifold of x, then $f(y)$ (or Fy for the linear case), is on the stable (unstable) manifold of $f(x)$ (or Fx for the linear case). Similarly, $f^{-1}(y)$ (or $F^{-1}y$ for the linear case) is on the stable (unstable) manifold of $f^{-1}(x)$ (or $F^{-1}x$ for the linear case). As shown later in this section, this invariance of the stable and unstable manifolds has important implications for the asymptotic behavior of the Cramer-Rao bound on $P_N(\hat{x}_0)$. In addition, if there are no infinite or zero-valued Lyapunov exponents, the stable and unstable manifolds span $\mathcal{R}^{\mathcal{N}}$ (where $\mathcal{N}$ is the dimension of the system) for linear systems; and the linear subspaces (mentioned above) tangent to the stable and unstable manifolds at "most" points on a chaotic attractor span $\mathcal{R}^{\mathcal{N}}$. Therefore, any perturbation δ_x can be decomposed into a component along the stable manifold and a component along the unstable manifold of x for most points on a chaotic attractor.

As claimed earlier, the Lyapunov exponents of the chaotic system $f(\cdot)$ influence the eigenvalues of the Fisher information matrix $J_N(x_0)$ and thus the Cramer-Rao bound on $P_N(\hat{x}_0)$. This influence is direct and most easily understood for the Fisher information matrix for the linear model given by (147)–(149), in the special case in which H is the $\mathcal{N}\times\mathcal{N}$-identity matrix, R is given by (168), and F is a diagonal matrix, with diagonal elements $\{\lambda_i\}_{i=1}^{\mathcal{N}}$

that are real, distinct, nonzero, and have non-unity magnitudes:

$$F = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{\mathcal{N}} \end{bmatrix} \tag{175}$$

For this system, (169) reduces to the following:

$$J_N(x_0) = \frac{1}{\sigma^2} \sum_{i=0}^{N} F^{2i} \tag{176}$$

$$= \frac{1}{\sigma^2} \begin{bmatrix} S_{N,1} & 0 & \cdots & 0 \\ 0 & S_{N,2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & S_{N,\mathcal{N}} \end{bmatrix} \tag{177}$$

where

$$S_{N,j} = \sum_{i=0}^{N} \lambda_j^{2i} \tag{178}$$

$$= \frac{1 - \lambda_j^{2N+2}}{1 - \lambda_j^2}. \tag{179}$$

Thus, the Cramer-Rao inequality for $P_N(\hat{x}_0)$ is the following:

$$P_N(\hat{x}_0) \geq J_N^{-1}(x_0) = \sigma^2 \begin{bmatrix} S_{N,1}^{-1} & 0 & \cdots & 0 \\ 0 & S_{N,2}^{-1} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & S_{N,\mathcal{N}}^{-1} \end{bmatrix} \tag{180}$$

where from (179),

$$S_{N,j}^{-1} = \frac{1 - \lambda_j^2}{1 - \lambda_j^{2N+2}}. \tag{181}$$

The error variance for each component of $\hat{x}_0$ is bounded below by the corresponding diagonal element of $J_N^{-1}(x_0)$. For $N = 0$ (i.e., only 1 observation), each element of $J_N^{-1}(x_0)$ has the value σ^2, which is the variance of each component of v_n. As indicated by (181), for each component $\hat{x}_{0,j}$ of $\hat{x}_0$ (where

$\hat{x}_0 = [x_{0,1}, \cdots, x_{0,\mathcal{N}}]^T$) for which the diagonal element λ_j of F has magnitude greater than one, the lower bound of $\sigma^2 S_{N,j}^{-1}$ on the error variance decreases with each additional observation, with the rate of decrease being almost exponential for large N as indicated by the following approximation:

$$\sigma^2 S_{N,j}^{-1} = \sigma^2 \frac{1-\lambda_j^2}{1-\lambda_j^{2N+2}} \approx \sigma^2 \frac{\lambda_j^2 - 1}{\lambda_i^{2N+2}}, \quad \text{for large N.} \tag{182}$$

In contrast, for each component $\hat{x}_{0,j}$ of $\hat{x}_0$ for which the diagonal element λ_j of F has magnitude less than one, the lower bound of $\sigma^2 S_{N,j}^{-1}$ on the error variance asymptotically approaches a nonzero limit given by

$$\sigma^2 S_{Nj}^{-1} = \sigma^2 \frac{1-\lambda_j^2}{1-\lambda_j^{2N+2}} \rightarrow \sigma^2(1-\lambda_j^2) \ \ (\text{as } N \rightarrow \infty). \tag{183}$$

Since F is diagonal, its diagonal elements $\{\lambda_i\}_{i=1}^{\mathcal{N}}$ are also its eigenvalues. In addition, the direction of the eigenvector corresponding to λ_j is simply e_j, the unit vector along the j^{th} standard coordinate axis in $\mathcal{R}^{\mathcal{N}}$ (i.e., the vector of all zeros except for a one as the j^{th} component). The logarithms of these eigenvalues are the Lyapunov exponents of this linear system. As such, for each component $\hat{x}_{0,j}$ of $\hat{x}_0$ for which the eigenvalue λ_j of F is greater than one in magnitude, hence the Lyapunov exponent is positive, the lower bound on the error variance decreases almost exponentially with each additional observation (and asymptotically approaches zero). In contrast, for each component $\hat{x}_{0,j}$ of $\hat{x}_0$ for which λ_j is less than one in magnitude, hence the Lyapunov exponent is negative, the lower bound on the error variance decreases only slightly with each additional observation and has a positive asymptotic value of $\sigma^2(1-\lambda_j^2)$.

This relation between the magnitudes of the eigenvalues (or equivalently, the signs of the Lyapunov exponents) and the behavior of the Cramer-Rao bound along the corresponding eigenvector directions has a simple intuitive interpretation. Consider the m^{th} observation z_m. From (148) and (149), and since $H = I_{\mathcal{N}}$ by assumption,

$$z_m = x_m + v_m. \tag{184}$$

Since F is diagonal and invertible,

$$F^{-m}(z_m) = F^{-m}(x_m + v_m) = x_0 + \sum_{j=1}^{\mathcal{N}} \lambda_j^{-m} v_{m,j} e_j, \tag{185}$$

where $v_m = [v_{m,1}, \cdots, v_{m,\mathcal{N}}]^T$. For each λ_j with magnitude greater than one

$$|\lambda_j^{-m} v_{m,j}| < |v_{m,j}|, \tag{186}$$

whereas for each λ_j with magnitude less than one

$$|\lambda_j^{-m} v_{m,j}| > |v_{m,j}|. \tag{187}$$

Thus, applying the inverse system, F^{-1}, m times to z_m reduces the noise along each component $z_{m,j}$ for which $|\lambda_j| > 1$ and yields a "less" noisy observation of $x_{0,j}$ than that similarly obtained with each previous observation. In contrast, applying the inverse system, F^{-1}, m times to z_m increases the noise along each component $z_{m,j}$ for which $|\lambda_j| < 1$, and yields a noisier observation of $x_{0,j}$ than that similarly obtained with each previous observation. Equivalently, the signal-to-noise ratio "increases" with time along components $z_{m,j}$ of z_m for which $|\lambda_j| > 1$ and "decreases" with time along components $z_{m,j}$ for which $|\lambda_j| < 1$. Therefore, an improvement of the estimate of each component $x_{0,j}$ of x_0 for which $|\lambda_j| > 1$ is possible with each additional observation. As a result, the minimum error variance of unbiased estimators for these components decreases with time. In contrast, almost no improvement in the estimate of each component $x_{0,j}$ of x_0 for which $|\lambda_j| < 1$ is possible with each additional observation since the signal-to-noise ratio decreases with time for observations of these components. As a result, the minimum error variance of unbiased estimators for these components remains nearly constant.

A useful and important quantity for many applications is a lower bound on the trace of $P_N(\hat{x}_0)$ (i.e., the sum of the diagonal elements), which is the sum of the conditional error variances for the components of $\hat{x}_0$. From the Cramer-Rao inequality, it follows that

$$\begin{aligned} Tr\{P_N(\hat{x}_0)\} &\equiv \text{Trace of } P_N(\hat{x}_0) \\ &\geq Tr\{J_N^{-1}(\hat{x}_0)\}. \end{aligned} \tag{188}$$

An important fact from linear algebra is that the trace of a matrix is the sum of its eigenvalues. For the specific linear model being considered here, the Cramer-Rao bound on the trace reduces to the following (after substituting (180)):

$$Tr\{P_N(\hat{x}_0)\} \geq \sigma^2 \sum_{i=1}^{\mathcal{N}} \frac{1-\lambda_i^2}{1-\lambda_i^{2N+2}}. \tag{189}$$

As noted earlier, with a single observation ($N = 0$) the Cramer-Rao bound on the error variance for each component of $\hat{x}_0$ is σ^2. Therefore, the trace of $P_0(\hat{x}_0)$ is bounded below by $\mathcal{N}\sigma^2$. However as implied by (189), this bound decreases with each additional observation with an asymptotic value given by

$$Tr\{P_\infty(\hat{x}_0)\} \equiv \lim_{N\to\infty} Tr\{P_N(\hat{x}_0)\} \tag{190}$$

$$\geq \lim_{N\to\infty} Tr\{J_N^{-1}(x_0)\} \tag{191}$$

$$= \sigma^2 \sum_{\substack{i=1 \\ |\lambda_i|<1}}^{\mathcal{N}} (1-\lambda_i^2) \tag{192}$$

where the sum is over those i for which $|\lambda_i|$ is less than one. As indicated by (192), only eigenvalues of F with magnitudes less than one make nonzero contributions to the asymptotic bound on the trace of $P_N(\hat{x}_0)$.

If F is not diagonal but is real and symmetric, similar results apply since F can be diagonalized with a unitary matrix. However, for an arbitrary matrix F, such a simple analysis is not possible. Nor is such a simple analysis possible for the original nonlinear system given by (138)—(140). However, for real linear systems F, for which all eigenvalues are real, nonzero, distinct, and non-unity in magnitude, and for the discrete-time chaotic system $f(\cdot)$, "experimental" results indicate that the eigenvalues of $J_N^{-1}(x_0)$ behave similarly with increasing N as do the eigenvalues of $J_N^{-1}(x_0)$ for a linear system with diagonal matrix F. In particular, for a chaotic system $f(\cdot)$, the same number of eigenvalues of $J_N^{-1}(x_0)$ as the number of negative Lyapunov exponents of $f(\cdot)$ have positive asymptotic values, while the same number of eigenvalues of $J_N^{-1}(x_0)$ as the number of positive Lyapunov exponents of $f(\cdot)$ decrease with additional observations with asymptotic values of zero. This behavior of the eigenvalues directly influences the Cramer-Rao bound on the trace of $P_N(\hat{x}_0)$, since the trace is bounded below by the trace of $J_N^{-1}(x_0)$, and the trace of a matrix equals the sum of its eigenvalues.

(Unfortunately, for a chaotic system or a linear system with eigenvalues both greater and less than one in magnitude, $J_N(x_0)$ and $J_N^{-1}(x_0)$ quickly become ill-conditioned with increasing N. Because of this ill-conditioning, it is difficult to accurately determine the eigenvalues of either $J_N(x_0)$ or $J_N^{-1}(x_0)$ for even moderate values of N).

We can use a similar argument as that used above, to intuitively explain the observed relation between the the eigenvalues of $J_N^{-1}(x_0)$ and the Lyapunov exponents of $f(\cdot)$ for chaotic systems and of F for linear systems. However, unlike the diagonal case, we have yet to find a direct relation between the eigenvalues of $J_N^{-1}(x_0)$ and the Lyapunov exponents of $f(\cdot)$ or F. That is, although we can qualitatively explain the asymptotic behavior of the eigenvalues of $J_N^{-1}(x_0)$, we cannot analytically express each of these eigenvalues in terms of the Lyapunov exponents of $f(\cdot)$ or F.

For linear systems, we apply the argument used above to the eigenvector directions rather than to the directions of the coordinate axes. Given the above restrictions on the eigenvalues, these directions are invariant subspaces and provide a direct sum decomposition of $\mathcal{R}^N$. Therefore, one can decompose each observation into components along the eigenvector directions. Since (by definition) these eigenvector directions are invariant, the earlier explanation (specifically the discussion beginning with (184)) is applicable with the coordinate axis directions e_j replaced by the eigenvector directions.

For chaotic systems, we apply the argument used for diagonal systems to the linear subspaces tangent to the stable and unstable manifolds at x_0 and at each point of the orbit $\{f^i(x_0)\}$. In particular, at x_0 we use the subspace spanned by the eigenvectors corresponding to negative Lyapunov exponents for Λ (given by (170)) and the subspace spanned by the eigenvectors corresponding to negative Lyapunov exponents for Λ^- (given by (174)). At most points on a chaotic attractor, these subspaces span $\mathcal{R}^N$. At other points on the orbit of x_0, we calculate similar Lyapunov-exponent defining matrices and choose the appropriate subspaces. For small perturbations, these subspaces are invariant; a small perturbation along the subspace corresponding to the negative Lyapunov exponents of Λ at x_0 gets mapped by $f(\cdot)$ along the subspace corresponding to the negative Lyapunov exponents of Λ at $f(x_0)$, with the expansion (contraction) rate bounded above by $\exp(\lambda_{max})$ where λ_{max} is the largest negative Lyapunov exponent. In light of this, we can apply the earlier argument to these subspaces by using the "linearized" dynamics of $f(\cdot)$ (given by the Jacobian) at each point of the orbit of x_0. Thus, for chaotic systems, this intuitive explanation for the relation between the eigenvalues of $J_N^{-1}(x_0)$ and the Lyapunov exponents of $f(\cdot)$ is only a first-order approximation valid for observations that are infinitesimally small perturbations from the actual orbit of x_0. However, the behavior of the eigenvalues

of the Fisher information matrix is actually independent of the observation noise level.

Figure 23 experimentally supports this intuitive argument. The figure shows the eigenvalues of $J_N^{-1}(\hat{x}_0)$ as a function of N. Note that one eigenvalue

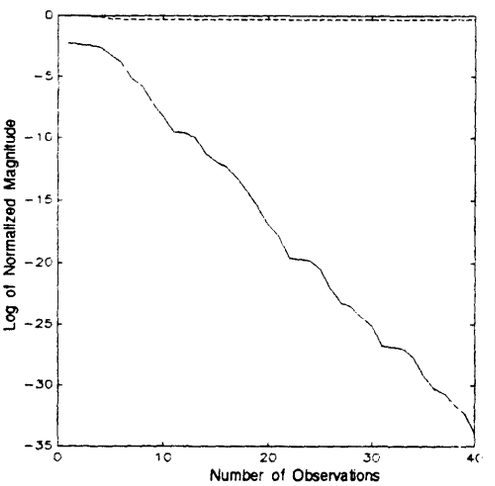

Figure 23: Eigenvalues of $J_N^{-1}(\hat{x}_0)$ as a function of N for the Henon map

rapidly approaches zero while the other decreases slowly. Unfortunately, because $J_N^{-1}(\hat{x}_0)$ becomes increasingly ill-conditioned with increasing N, the eigenvalues of $J_N^{-1}(\hat{x}_0)$ could not be accurately calculated for values of N larger than those shown.

9.3 The Cramer-Rao Bound on $P_N(\hat{y}_m)$

This subsection uses the Fisher information matrix $J_N(x_0)$ derived in the previous subsection to derive a lower bound on $P_N(\hat{y}_m)$ defined by (144) and repeated here for convenience:

$$P_N(\hat{y}_m) \equiv E\left[(\hat{y}_m - y_m)(\hat{y}_m - y_m)^T | x_0\right]. \tag{193}$$

As shown in the appendix, if

$$s = t(x) \tag{194}$$

for some arbitrary differentiable function t, then

$$P(\hat{s}) \geq D\{t(x)\}J^{-1}(x)D^T\{t(x)\}. \tag{195}$$

where $\hat{s}$ is any unbiased estimator for s,

$$P(\hat{s}) \equiv E\left[(\hat{s} - t(x))(\hat{s} - t(x))^T | x\right], \tag{196}$$

and where $J(x)$ is the Fisher information matrix for x.

Since,

$$y_m = h(f^m(x_0)), \tag{197}$$

it follows that

$$\begin{aligned} P_N(\hat{y}_m) &\geq D_{x_0}\{h(f^m(x_0))\} J_N^{-1}(x_0) D_{x_0}^T\{h(f^m(x_0))\} && (198) \\ &= D\{h(f^m(x_0))\} D\{f^m(x_0)\} J_N^{-1}(x_0) \\ &\quad \times D^T\{f^m(x_0)\} D^T\{h(f^m(x_0))\}. && (199) \end{aligned}$$

Also, note that since

$$x_m = f^m(x_0), \tag{200}$$

it follows that

$$P_N(\hat{x}_m) \geq D\{f^m(x_0)\} J_N^{-1}(x_0) D^T\{f^m(x_0)\}, \tag{201}$$

where

$$P_N(\hat{x}_m) \equiv E\left[(\hat{x}_m - x_m)(\hat{x}_m - x_m)^T | x_0\right]. \tag{202}$$

As with the expression for $J_N(x_0)$ given by (165), the above expression for $J_N(y_m)$ has a revealing structure when $f(\cdot)$ is a diffeomorphism and $h(\cdot)$ and R are given by

$$\begin{aligned} h(x) &= x && (203) \\ R &= \sigma^2 I_{\mathcal{N}}. && (204) \end{aligned}$$

With this restriction on $h(\cdot)$, $y_m = x_m$, and thus (198) reduces to

$$\begin{aligned} P_N(\hat{y}_m) &\geq D\{f^m(x_0)\} J_N^{-1}(x_0) D^T\{f^m(x_0)\} && (205) \\ &\equiv J_N^{-1}(\hat{y}_m). && (206) \end{aligned}$$

Since $f(\cdot)$ is a diffeomorphism, $D\{f^m(x_0)\}$ is invertible and thus

$$J_N(\hat{y}_m) = D^{-T}\{f^m(x_0)\} J_N(x_0) D^{-1}\{f^m(x_0)\}, \tag{207}$$

where

$$D^{-T}\{f^m(x_0)\} \equiv \left\{D^T\{f^m(x_0)\}\right\}^{-1}. \tag{208}$$

Substituting (160) and (169) in (207) yields

$$J_N(\hat{y}_m) = \frac{1}{\sigma^2}\sum_{i=0}^{N} D^{-T}\{f^m(x_0)\}\left[T^i_{x_0}\right]^T T^i_{x_0} D^{-1}\{f^m(x_0)\}, \tag{209}$$

where $T^i_{x_0}$ is given by (160) which is repeated here for convenience:

$$T^i_{x_0} \equiv \begin{cases} I_{\mathcal{N}}, \ \text{(the } \mathcal{N}\times\mathcal{N}\text{-identity matrix)} & i = 0 \\ \prod_{j=1}^{i} D\{f(f^{i-j}(x_0))\}, & i > 0 \\ \prod_{j=i+1}^{0} D\{f^{-1}(f^j(x_0))\}, & i < 0 \end{cases} \tag{210}$$

Using the following identity from vector calculus, valid for an arbitrary invertible function $g(\cdot)$ at a specific point (vector) x,

$$D_x^{-1}\{g(x)\} = D_{g(x)}\{g^{-1}(g(x))\}, \tag{211}$$

one can show that (209) is equivalent to

$$J_N(y_m) = \frac{1}{\sigma^2}\sum_{i=-m}^{N-m}\left[T^i_{x_m}\right]^T T^i_{x_m} \tag{212}$$

$$= \frac{1}{\sigma^2}\sum_{i=-m}^{-1}\left[T^i_{x_m}\right]^T T^i_{x_m} + \sum_{i=0}^{N-m}\left[T^i_{x_m}\right]^T T^i_{x_m}. \tag{213}$$

The expressions for $J_N(x_0)$ and $J_N(y_m)$ given by (169) and (213), respectively, are similar with the only difference being the limits on the summations. This apparently minor difference has profound implications since (as indicated by (160)) the terms in the first summation of (213) involve products of iterates of the Jacobian of the inverse function $f^{-1}(\cdot)$. As a result, whereas the Lyapunov exponents of $f(\cdot)$ influence the eigenvalues of $J_N(x_0)$, the Lyapunov exponents of both $f(\cdot)$ and $f^{-1}(\cdot)$ influence the eigenvalues of $J_N(y_m)$. However, the Lyapunov exponents of $f^{-1}(\cdot)$ are the negatives of the Lyapunov exponents of $f(\cdot)$. Equivalently, whereas the Lyapunov exponents of $f(\cdot)$ are the the logarithms of the eigenvalues of the following matrix

$$\Lambda = \lim_{i\to\infty}\left(\left[T^i_{x_m}\right]^T T^i_{x_m}\right)^{\frac{1}{2i}}, \tag{214}$$

the Lyapunov exponents of $f^{-1}(\cdot)$ are the logarithms of the reciprocals of these eigenvalues.

In the preceding subsection, we argued that the bound on the error variance of estimators for x_0 should decrease asymptotically to zero in directions tangent to the unstable manifold of $f(\cdot)$ at x_0, as the number of observations increased. Clearly, the same result holds for the bound on the error variance of estimators for y_m, as the number of observations "after" time m increases. In addition, it is shown below for diagonal linear systems and is argued for other linear systems as well as for chaotic systems, that as m increases and thus the number of observations "before" time m increases, the bound on the error variance of estimators for y_m in directions tangent to the unstable manifold of $f^{-1}(\cdot)$ decreases asymptotically to zero. However, (by definition) the unstable manifold of $f^{-1}(\cdot)$ is the stable manifold of $f(\cdot)$. Therefore, as the number of observations "before" time m increases, the bound on the error variance in directions tangent to the stable manifold of $f(\cdot)$ decreases asymptotically to zero. As a consequence, as the number of observations before time m and after time m both go to infinity, the bound on the error variance in directions tangent to the stable and unstable manifolds of $f(\cdot)$ and consequently in all directions goes to zero. As a result, the lower bound on the trace of $P_N(y_m)$ given by the Cramer-Rao inequality goes to zero as well.

We now explicitly show this result for the linear model given by (147)—(149) with diagonal matrix F. As in the previous subsection, let F be given by

$$F = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{bmatrix}, \tag{215}$$

where the diagonal elements λ_i are real, distinct, nonzero, and have non-unity magnitudes.

For this system, (212) reduces to the following:

$$J_N(y_m) = \frac{1}{\sigma^2} \sum_{i=-m}^{N-m} F^{2i} \tag{216}$$

$$= \frac{1}{\sigma^2}\begin{bmatrix} S_{N,1} & 0 & \cdots & 0 \\ 0 & S_{N,2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & S_{N,\mathcal{N}} \end{bmatrix}, \tag{217}$$

where

$$S_{N,j} = \sum_{i=-m}^{N-m} \lambda_j^{2i} \tag{218}$$

$$= \frac{1-\lambda_j^{2N+2}}{\lambda_j^{2m}(1-\lambda_j^2)}. \tag{219}$$

Thus, the Cramer-Rao inequality for $P_N(y_m)$ is the following:

$$P_N(\hat{y}_m) \geq J_N^{-1}(y_m) = \sigma^2 \begin{bmatrix} S_{N,1}^{-1} & 0 & \cdots & 0 \\ 0 & S_{N,2}^{-1} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & S_{N,\mathcal{N}}^{-1} \end{bmatrix}, \tag{220}$$

where from (219),

$$S_{N,j}^{-1} = \frac{\lambda_j^{2m}(1-\lambda_j^2)}{1-\lambda_j^{2N+2}}. \tag{221}$$

A comparison of (181) and (221) reveals that the diagonal elements of $J_N^{-1}(x_0)$ and $J_N^{-1}(y_m)$ are nearly identical, with the only difference being the inclusion of the factors λ_j^{2m} in the diagonal elements of $J_N^{-1}(y_m)$, but not of $J_N^{-1}(x_0)$.

Equation (221) reveals that for a fixed m, for those diagonal elements $S_{N,j}^{-1}$ of $J_n^{-1}(y_m)$ for which $|\lambda_j| > 1$, then

$$\lim_{N\to\infty} S_{N,j}^{-1} = \lim_{N\to\infty} \frac{\lambda_j^{2m}(1-\lambda_j^2)}{1-\lambda_j^{2N+2}} = 0 \tag{222}$$

This result is analogous to that given earlier for the diagonal elements of $J_N^{-1}(x_0)$ and has a similar interpretation. That is, for each component $\hat{y}_{m,j}$ of $\hat{y}_m$ (where $\hat{y}_m = [\hat{y}_{m,1}, \cdots, \hat{y}_{m,\mathcal{N}}]^T$) for which the diagonal element (or equivalently the eigenvalue) λ_j of F has magnitude greater than one, the lower

bound on the error variance given by the Cramer-Rao inequality asymptotically decreases to zero as the number of observations "after" time m goes to infinity. The reason for this asymptotic behavior is identical to that discussed in the previous subsection for the asymptotic behavior of elements of $J_N^{-1}(x_0)$ for which the corresponding diagonal elements of F had magnitudes greater than one.

However, a certain type of asymptotic behavior of diagonal elements of $J_N^{-1}(y_m)$ for which the corresponding elements of F have magnitudes less than one differs dramatically than the behavior observed earlier for the same diagonal elements of $J_N^{-1}(x_0)$. In particular, as the time of interest m goes to infinity, or equivalently the number of observations before time m goes to infinity, then for those diagonal elements $S_{N,j}^{-1}$ of $J_N^{-1}(y_m)$ for which $|\lambda_j| < 1$,

$$\lim_{m\to\infty} S_{N,j}^{-1} = \lim_{m\to\infty} S_{N,j}^{-1} = \frac{\lambda_j^{2m}(1-\lambda_j^2)}{1-\lambda_j^{2N+2}} = 0. \tag{223}$$

Therefore, the lower bound on the error variance for each component $\hat{y}_{m,j}$ of $\hat{y}_m$, for which the diagonal element λ_j of F has magnitude less than one, goes to zero as the number of observations "before" time m goes to infinity. The reason for this asymptotic behavior is similar to that for the components $\hat{y}_{m,j}$ for which $|\lambda_j| > 1$, but with the inverse system F^{-1} used in the analysis. This follows from the fact that the diagonal elements (or eigenvalues) of F^{-1} are the reciprocals of the diagonal elements of F. Therefore, the diagonal elements of F^{-1} with magnitudes greater than one are the diagonal elements λ_j of F with magnitudes less than one. In addition, the condition that the number of observations "after" time m goes to infinity for the system F^{-1} is the same as the condition that the number of observations "before" time m goes to infinity for the system F.

The overall result is that as the number of observations both "before" and "after" time m go to infinity, the lower bound on the error variance for each component of $\hat{y}_m$ goes to zero and thus the lower bound on the trace of $P_N(\hat{y}_m)$ goes to zero. This result on the asymptotic behavior of the Cramer-Rao bound on $P_N(\hat{y}_m)$ can be explained with a similar argument in terms of stable and unstable manifolds and signal-to-noise ratios as was used in Section 9.2 to explain the asymptotic behavior of the Cramer-Rao bound on $P_N(\hat{x}_0)$. In fact, the exact same argument applies to the decrease in error bound along the unstable manifold (the subspace spanned by the

eigenvectors of F corresponding to eigenvalues with magnitudes greater than one) with each additional observation after time m. The same argument also accounts for the decrease in error bound along the stable manifold (the subspace spanned by the eigenvectors of F corresponding to eigenvalues with magnitudes less than one) with each additional observation before time m if one applies the argument to the inverse system F^{-1}.

This argument also intuitively explains the experimentally observed decrease in all eigenvalues of $J_N^{-1}(y_m)$ for other linear systems and for chaotic systems, as the numbers of observations both before and after time m increase. In particular, the same number of eigenvalues of $J_N^{-1}(y_m)$ as positive Lyapunov exponents of $f(\cdot)$ experimentally decrease asymptotically to zero as the number of observations after time m increases. Also, the same number of eigenvalues of $J_N^{-1}(y_m)$ as negative Lyapunov exponents of $f(\cdot)$ experimentally decrease asymptotically to zero as the number of observations before time m increases. For linear systems, we apply the above argument to the eigenvector directions rather than to the coordinate axis directions. For chaotic systems, we apply the argument to the linear subspaces tangent to the stable and unstable manifolds at x_m and at each point of the orbit of x_m, both forward and backward in time.

Figures 24 and 25 depict this behavior of the eigenvalues of $J_N^{-1}(y_m)$ with increasing number of observations before and after time m. Figure 24 shows the eigenvalues of $J_N^{-1}(y_m)$ plotted as a function of the number of past observations with no future observations. Figure 25 shows the trace of $J_N^{-1}(y_m)$, which is the sum of its eigenvalues, as a function of the number of past and future observations. That is, each value on the horizontal axis indicates the number of past observations as well as the number of future observations. In both figures, the N in $J_N^{-1}(y_m)$ loses its meaning, since the time of interest m is fixed but the number of observations prior to this time is increasing.

9.4 Practical Implications

In practice, one does not have an infinite number of observations. The practical implication of the preceding analysis is that for a finite number of observations $\{z_m\}_{m=0}^N$, the lower bound on the trace of $P_N(\hat{y}_m)$ given by the trace of $J_N^{-1}(y_m)$, is largest at times m near the beginning (i.e., $m \approx 0$) and end (i.e., $m \approx N$) of the observation set and smallest for times m in the middle

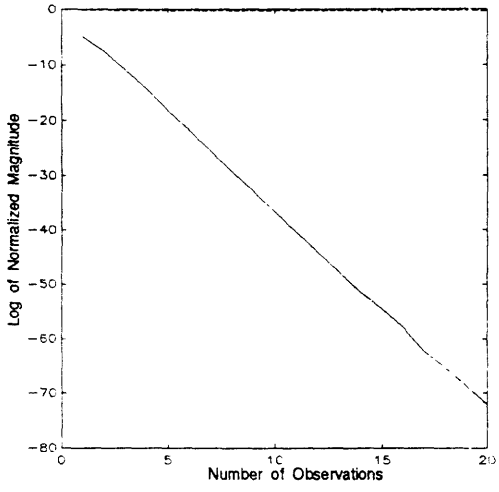

Figure 24: Eigenvalues of $J_N^{-1}(\hat{y}_m)$ as a function of the number of past observations

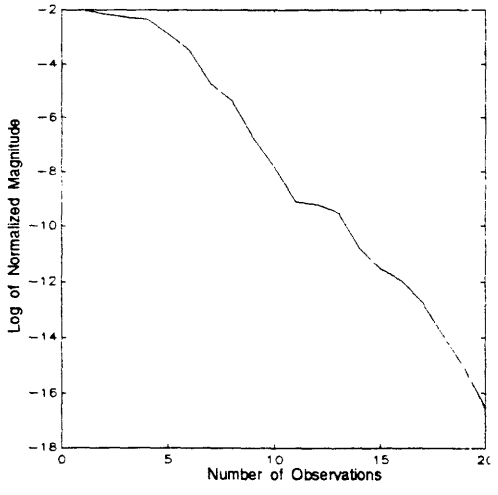

Figure 25: Sum of eigenvalues of $J_N^{-1}(\hat{y}_m)$ as a function of the number of past and future observations

of the observation set. As suggested by the preceding analysis, for times m near the beginning of the observation set, the bound on the error variance is small in directions tangent to the unstable manifold of x_m, but near its initial value (i.e., the value when only the observation at time m is available) in directions tangent to the stable manifold of x_m. Similarly, for times m near the end of the observation set, the bound on the error variance is small in directions tangent to the stable manifold of x_m, but near its initial value in directions tangent to the unstable manifold of x_m. However, for times near the middle of the observation set, the bound on the error variance is small in all directions. Figure 26 experimentally confirms this predicted behavior on the trace of $P_N(\hat{y}_m)$ for the Henon map with 20 observations. The figure shows the trace of $P_N(\hat{y}_m)$ as a function of m for $1 \leq m \leq 20$. (Note the initial observation occurs at time 1 and not time 0).

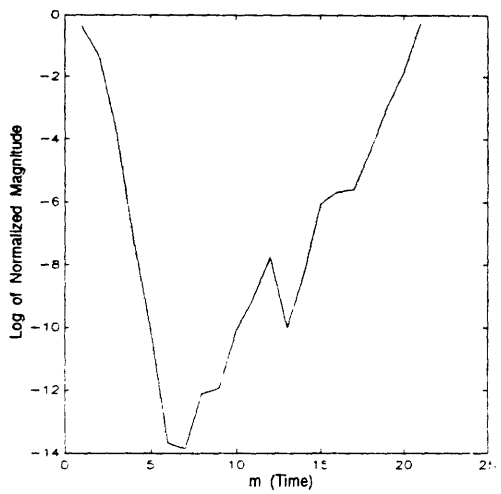

Figure 26: Trace of $J_N^{-1}(\hat{y}_m)$ as a function of m for the Henon map with 20 observations

One practical constraint ignored in the error bound derivation is that the precision available for representing and processing data in all computers is finite. As a result, finite-precision arithmetic and round-off error are inevitable; and the assumption expressed by (138) that the state x_n evolves deterministically is never true in practice. One can account for round-off error by including a driving noise component in the state equation. The inclusion of this component complicates the derivation of Cramer-Rao bounds on the

initial state x_0 and output y_m. For example, as was discussed in Section 5, whereas the state transition probability density $p(x_n|x_{n-1})$ is simply an impulse in the absence of driving noise, it satisfies the discrete-time "forward Kolmogorov" or "Fokker-Planck equation" in the presence of driving noise. The complexity of the resulting error bounds precludes any straightforward intuitive interpretation. We will analyze the complicated omnipresent influence of round-off error and finite-precision arithmetic on state and output estimation with chaotic systems in a future report.

The discussion in the preceding subsection also helps explain the ridgelike properties of the the likelihood functions $p(Y_0^n|x_n)$ and $p(Y_n^{n+m}|x_n)$ apparent in the figures in Sections 5 and 7. For example, since $p(Y_0^n|x_n)$ uses only past observations, the uncertainty in the estimate of x_n has decreased considerably along the stable manifold of x_n, but decreased little along the unstable manifold. Thus, the observed ridge of high likelihood values probably corresponds to points along or near the unstable manifold of x_n. The finite width of this ridge probably reflects the remaining but much smaller uncertainty along the stable manifold. Similarly, since $p(Y_n^{m+n}|x_n)$ uses only future observations, the observed ridge of high likelihood values probably corresponds to points along or near the stable manifold of x_n, with the finite width of the ridge reflecting the remaining uncertainty along the unstable manifold.

9.5 Bound on the Sum of the Traces of the ML Estimates $\{y_i\}_{i=0}^N$

It is straightforward to derive a lower bound on the following quantity

$$Tr\left\{\sum_{m=0}^{N} P_N(\hat{y}_m)\right\} = \sum_{m=0}^{N} Tr\{P_N(\hat{y}_m)\}, \tag{224}$$

when R, the covariance matrix of the observation noise v_n, equals $\sigma^2 I_{\mathcal{N}}$. The above expression is the sum of the traces of the error covariance matrices of unbiased output estimators for all observation times. The resulting sum is a measure of the accumulated, summed, error variances of the components of the estimators $\hat{y}_m$ for each m where $0 \leq m \leq N$. From (198), we know that

$$P_N(\hat{y}_m) \;\geq\; D_{x_0}\{h(f^m(x_0))\}J_N^{-1}(x_0)D_{x_0}^T\{h(f^m(x_0))\}. \tag{225}$$

It follows that

$$Tr\{P_N(\hat{y}_m)\} \geq Tr\left\{D_{x_0}\{h(f^m(x_0))\}J_N^{-1}(x_0)D_{x_0}^T\{h(f^m(x_0))\}\right\}. \quad (226)$$

Therefore, a lower bound on (224) is given by

$$\sum_{m=0}^{N} Tr\{P_N(\hat{y}_m)\} \geq \sum_{m=0}^{N} Tr\left\{D_{x_0}\{h(f^m(x_0))\}J_N^{-1}(x_0)D_{x_0}^T\{h(f^m(x_0))\}\right\}. \quad (227)$$

The following fact from linear algebra, valid for any two $\mathcal{N} \times \mathcal{N}$-matrices A and B,

$$Tr\{AB\} = Tr\{BA\}, \quad (228)$$

allows the following identity to be established:

$$Tr\left\{D_{x_0}\{h(f^m(x_0))\}J_N^{-1}(x_0)D_{x_0}^T\{h(f^m(x_0))\}\right\} = Tr\left\{J_N^{-1}(x_0)D_{x_0}^T\{h(f^m(x_0))\}D_{x_0}\{h(f^m(x_0))\}\right\}. \quad (229)$$

Substituting (229) in (226) yields the following:

$$\sum_{m=0}^{N} Tr\{P_N(\hat{y}_m)\} \geq Tr\{\sum_{m=0}^{N} J_N^{-1}(x_0)D_{x_0}^T\{h(f^m(x_0))\} \times D_{x_0}\{h(f^m(x_0))\}\} \quad (230)$$

$$= Tr\{J_N^{-1}(x_0)\sum_{m=0}^{N} D_{x_0}^T\{h(f^m(x_0))\} \times D_{x_0}\{h(f^m(x_0))\}\}. \quad (231)$$

But, from (165) and since $R = \sigma^2 I_{\mathcal{N}}$,

$$J_N(x_0) = \sum_{i=0}^{N} D_{x_0}^T\{h(f^i(x_0))\}R^{-1}D_{x_0}\{h(f^i(x_0))\} \quad (232)$$

$$= \sigma^{-2}\sum_{i=0}^{N} D_{x_0}^T\{h(f^i(x_0))\}D_{x_0}\{h(f^i(x_0))\}. \quad (233)$$

Thus, (231) reduces to the following:

$$\sum_{m=0}^{N} Tr\{P_N(\hat{y}_m)\} \geq Tr\{J_N^{-1}(x_0)\sigma^2 J_N(x_0)\} \quad (234)$$

$$= \mathcal{N}\,\sigma^2. \quad (235)$$

As shown earlier, the covariance matrix of z_m, when conditioned on y_m, is $R = \sigma^2 I_{\mathcal{N}}$, the covariance matrix of v_m, and the trace of this covariance matrix is $\mathcal{N}\sigma^2$. Note that this quantity is the trace of the error covariance matrix for the maximum likelihood estimate of y_m given only the observation z_m. (The ML estimate of y_m given only z_m is in fact z_m). As a result, it is also the bound given by the Cramer-Rao inequality on the trace of the error covariance matrix of any unbiased estimator for y_m, given only the observation z_m. We denote this quantity the "*a priori* processing error". Since the observation time m is arbitrary, it follows that the the "total *a priori* processing error" defined to be the sum of the *a priori* processing errors for all observation times m, where $0 \leq m \leq N$, is $\mathcal{N}(N+1)\sigma^2$.

Using the same line of reasoning, we can interpret the trace of $J_N^{-1}(y_m)$ as the "*a posteriori* processing error" for time m, since it is the bound given by the Cramer-Rao inequality on the trace of the error covariance matrix of any unbiased estimator for y_m, given all the observation $\{z_m\}_{n=0}^{N}$. Similarly, we can interpret the following sum

$$\sum_{m=0}^{N} Tr\{J_N^{-1}(y_m)\}, \quad (236)$$

as the "total *a posteriori* processing error", which from (235) is $\mathcal{N}\sigma^2$.

Therefore, the "processing gain", defined as the ratio of the total *a priori* processing error to the total *a posteriori* processing error is simply $N+1$, the number of observations. Perhaps surprisingly, this is the same processing gain that results for a linear system.

10 Discussion

The preceding section showed that the Cramer-Rao bounds on the error covariance matrices of unbiased estimators for the initial state x_0 and output y_m have similar properties for both linear and chaotic systems. The section also indicated that the total processing gain for unbiased estimation of the output y_m for all observation times was identical for both linear and chaotic systems. In light of this, the question arises as to what advantages, if any, chaotic systems offer over linear systems for performing state estimation and thereby reducing the effect of additive observation noise on a signal.

The answer to this question requires consideration of a fundamental practical issue ignored in the analysis in Section 9. Specifically, the discussion of linear systems considered the existence of eigenvalues both greater and less than one in magnitude. However, a linear system with at least one eigenvalue greater than one in magnitude is an unstable system, since for initial conditions and inputs with components along the corresponding eigenvector direction, the output grows without bound.

In contrast, certain nonlinear systems, in particular chaotic systems, have both positive and negative Lyapunov exponents (the equivalent of eigenvalues greater and less than one in magnitude for discrete-time linear systems), but nonzero, bounded, steady-state outputs (i.e., chaotic attractors) over a range of initial conditions. Locally, these systems are unstable, in that a small perturbation about a point on the attractor increases under iterations of the system; but, the increase is bounded by the size of the attractor. In fact as noted earlier, this local instability is responsible for the impulse-like property of the likelihood function $p(Y_{n-r}^{n+m}|x_n)$ considered in Section 7 and the difficulty in performing probabilistic state estimation with chaotic systems. However, this local instability is also responsible for the "snowflake" property of chaotic systems exploited in Section 8 for performing state and parameter discrimination.

Because of the simultaneous presence of positive and negative Lyapunov exponents and the concomitant existence of stable and unstable manifolds with chaotic systems, many traditional engineering attitudes and assumptions are not applicable to such systems. For example, most engineering applications emphasize the zero-state-response (ZSR) of a system, with the zero-input-response (ZIR) treated as either an undesirable nuisance (as in power systems) or something to be ignored (as suggested by the often used

statement, "... the response after all transients have died away." In contrast, as this report has shown, with chaotic systems the initial state is often the single most important quantity, and the ZIR often a non-transient phenomenon with many interesting, potentially useful properties.

In addition, as suggested by the performance of the EKF in Section 5 and the Cramer-Rao bound derivation in the preceding section, the use of recursive filtering (i.e., recursive state estimation with only past observations) prevalent in the engineering community has little, if any, practical value for chaotic systems. With most "useful", discrete-time, linear systems, all eigenvalues have magnitudes less than one (i.e., negative Lyapunov exponents) and thus these systems have "only" a stable manifold. In light of this, as the analysis in the preceding section showed, "past" observations alone are often sufficient (and necessary) for performing accurate state estimation. However, because chaotic systems have both positive and negative Lyapunov exponents, accurate state estimation requires that both past and future observations be used.

11 Summary

This report has discussed probabilistic state estimation with chaotic systems, and through the discussion revealed interesting, potentially useful properties of such systems. The report began by introducing and experimentally evaluating the extended Kalman filter (EKF), which is a nonlinear recursive filter closely related to the Kalman filter. Experimental results revealed that the EKF is a poor state estimator for chaotic systems. The report also introduced and experimentally evaluated the extended Kalman smoother (EKS), a nonlinear smoother which combined the EKF and Rauch-Tung-Striebel linear smoother. The performance of the EKS was shown to be much better than the EKF on two chaotic systems, and slightly better on a third system. The report provided a nonrigorous analysis of the performance of the EKF and EKS, primarily by investigating the properties of relevant likelihood functions. The analysis revealed that the presence of both positive and negative Lyapunov exponents and associated stable and unstable manifolds with chaotic systems, has an important impact on these likelihood functions and on any probabilistic state estimation technique applied to chaotic systems.

The report also derived and interpreted Cramer-Rao error bounds for several related estimation problems involving chaotic systems. The derivation revealed a close relation between these bounds and the Lyapunov exponents of the systems. In addition, the asymptotic behavior of these bounds was found to be similar for linear and nonlinear systems.

Two simpler problems, closely related to state estimation, were also briefly considered. These were state and parameter discrimination with chaotic systems using a a finite known set of possible states and parameter values. Experimental results confirmed that accurate discrimination is possible even at extremely low SNRs and with both additive and multiplicative noise. The report speculated on potential practical applications of state and parameter discrimination with chaotic systems in the area of secure communications. Finally, the report briefly discussed similarities and differences between linear and chaotic systems and pointed out that traditional engineering beliefs and assumptions are inappropriate when dealing with chaotic systems.

State estimation with chaotic systems remains a challenging problem. As shown in this report, chaotic systems have unique properties that simultaneously aid and hinder accurate state estimation. A distinguishing property of a deterministic chaotic system is that perfect knowledge of the initial state or

the state at any time reveals everything about past and future states; imperfect knowledge reveals almost nothing. As noted in this paper, this property may render chaotic systems useful for several practical applications. However, it remains to be seen whether chaotic systems offer superior alternatives to existing techniques for these or any other applications.

A Appendix

This appendix proves the following proposition stated without proof in [16].

Definitions:

Let x denote an unknown $\mathcal{N}$-dimensional parameter vector, z an observation vector with a probabilistic dependence on x, and $p(z|x)$ the conditional density which reflects the probabilistic relation between z and x. Also, let $\hat{x}$ denote an "unbiased" estimator for x based on the observation z, that is

$$\int \hat{x}\, p(z|x)dz = x. \tag{237}$$

Finally, let s denote a $\mathcal{P}$-dimensional vector defined as

$$s = t(x), \tag{238}$$

where $t(\cdot)$ is a differentiable function; and let $\hat{s}$ denote an unbiased estimator for s based on the observation z as conditioned on x. That is

$$\int \hat{s}\, p(z|x)dz = s = t(x). \tag{239}$$

Proposition:

The matrix

$$P(\hat{s}) - D\{t(x)\}J^{-1}(x)D^T\{t(x)\} \tag{240}$$

is positive semidefinite, where $P(\hat{s})$ is the error-covariance matrix for $\hat{s}$ given by

$$P(\hat{s}) \equiv E\left[(\hat{s} - t(x))(\hat{s} - t(x))^T \middle| x\right]. \tag{241}$$

and $J(x)$ is the Fisher information matrix for x

$$J(x) = E\left\{ D_x^T \left\{\ln p(z|x)\right\} D_x \left\{\ln p(Z|x)\right\} \middle| x \right\}. \tag{242}$$

Therefore, a lower bound on $P(\hat{s})$ is the following:

$$P(\hat{s}) \geq D\{t(x)\}J^{-1}(x)D^T\{t(x)\}. \tag{243}$$

Proof:

The following proof is a generalization of a proof given in [16] which establishes the Cramer-Rao inequality.

Since $\hat{s}$ is an unbiased estimator,

$$s_i = \int \hat{s}_i \; p(z|x)dz \tag{244}$$

where s_i and $\hat{s}_i$ are the i^{th} components of s and $\hat{s}$ respectively. Differentiating both sides of (244) with respect to x_j, the j^{th} component of x, yields

$$\frac{\partial s_i}{\partial x_j} = \frac{\partial}{\partial x_j} \int \hat{s}_i \; p(z|x)dz \tag{245}$$

$$= \int \hat{s}_i \frac{\partial p(z|x)}{\partial x_j} dz \tag{246}$$

$$= \int \hat{s}_i \frac{\partial \ln p(z|x)}{\partial x_j} p(z|x)dz. \tag{247}$$

Also note that

$$\int s_i \frac{\partial p(z|x)}{\partial x_j} dz = 0. \tag{248}$$

Now define the vector y as follows:

$$y = \begin{bmatrix} \hat{s}_1 - s_1 \\ \vdots \\ \hat{s}_\mathcal{P} - s_\mathcal{P} \\ \frac{\partial \ln p(z|x)}{\partial x_1} \\ \vdots \\ \frac{\partial \ln p(z|x)}{\partial x_\mathcal{N}} \end{bmatrix} \tag{249}$$

A straightforward, albeit tedious computation yields:

$$E\left\{yy^T\right\} = \left[\begin{array}{c|c} P(\hat{s}) & D\{t(x)\} \\ \hline D^T\{t(x)\} & J(x) \end{array}\right]. \tag{250}$$

A useful fact from linear algebra [2] is that given a matrix M partitioned as

$$M = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right], \tag{251}$$

where A, B, C, and D are appropriately sized matrices, then if D is invertible,

$$\|M\| = \|A - BD^{-1}C\|, \tag{252}$$

where $\|\cdot\|$ is the determinant operator.

Applying this fact to (250) yields

$$\left\|E\left\{yy^T\right\}\right\| = \left\|\begin{array}{c|c} P(\hat{s}) & D\{t(x)\} \\ \hline D^T\{t(x)\} & J(x) \end{array}\right\| \tag{253}$$

$$= \left\|P(\hat{s}) - D\{t(x)\}J^{-1}(x)D^T\{t(x)\}\right\|. \tag{254}$$

Since $E\left\{yy^T\right\}$ is a covariance matrix, it is nonnegative definite. Therefore, its determinant and the determinant of each of its submatrices is nonnegative. It is easy to show that the determinant of each submatrix of $E\left\{yy^T\right\}$ is the same as the determinant of the submatrix of $P(\hat{s}) - D\{t(x)\}J^{-1}(x)D^T\{t(x)\}$ formed by omitting the same rows and columns. Therefore, the determinant of $P(\hat{s}) - D\{t(x)\}J^{-1}(x)D^T\{t(x)\}$ and each of its submatrices is nonnegative. Therefore, $P(\hat{s}) - D\{t(x)\}J^{-1}(x)D^T\{t(x)\}$ is nonnegative definite.

References

[1] B. Anderson and J. Moore, *Optimal Filtering.* Englewood Cliffs: Prentice Hall, 1979.

[2] W. Brogan, *Modern Control Theory.* Englewood Cliffs:Prentice Hall, 1985, pp. 85.

[3] J. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Reviews of Modern Physics*, Volume 57, Number 3, Part 1, July 1985, pp. 617–656.

[4] J. Farmer and J. Sidorowich, "Exploiting chaos to predict the future and reduce noise," Los Alamos National Laboratory Technical Report, February 1988.

[5] J. Farmer and J. Sidorowich, "Optimal shadowing and noise reduction," *Physica D*, Volume 47, Number 3, January 1991, pp. 373-392.

[6] S. Hammel, "A noise reduction method for chaotic systems," *Physics Letters A*, Volume 148, Number 9, 3 September 1990, pp. 421–428.

[7] A. Jazwinski, *Stochastic Processes and Filtering Theory.* New York: Academic Press, 1970.

[8] E. Kostelich and J. Yorke, "Noise reduction: finding the simplest dynamical system consistent with the data," *Physica D*, Volume 41, Number 2, March 1990, pp. 183–196.

[9] F. Lewis, *Optimal Estimation with an Introduction to Stochastic Control Theory.* New York: John Wiley and Sons, 1986.

[10] P. Marteau and H. Abarbanel, "Noise reduction in chaotic time series using scaled probabilistic methods," To appear in *Journal of Nonlinear Science.*

[11] C. Myers, S. Kay, and M. Richard, "Signal separation for nonlinear dynamical systems," *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing*, San Francisco, March 1992.

[12] C. Myers, A. Singer, B. Shin, and E. Church, "Modeling chaotic systems with hidden markov models," *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing*, San Francisco, March 1992.

[13] H. Rauch, F. Tung, and C. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, Volume 3, Number 8, August 1965, pp. 1445–1450.

[14] A. Singer, "Codebook prediction: a nonlinear signal modeling paradigm," *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing*, San Francisco, March 1992.

[15] A. Singer, *Codebook prediction*, Master of Science Thesis, MIT, January 1992.

[16] H. Van Trees, *Detection, Estimation, and Modulation Theory, Part I.* New York: John Wiley and Sons, 1968.

[17] A. Willsky, *Course Notes for 6.433 "Recursive Estimation"*, Unpublished, 1989.