Journal of Sound and Vibration (2002) **250**(5), 877–901
doi:10.1006/jsvi.2001.3963, available online at http://www.idealibrary.com on IDEAL®

OPTIMAL TRACKING OF PARAMETER DRIFT IN A CHAOTIC SYSTEM: EXPERIMENT AND THEORY

A. CHATTERJEE

Mechanical Engineering, Indian Institute of Science, Bangalore 560012, *India.*
E-mail: anindya@mecheng.iisc.ernet.in

J. P. CUSUMANO

Engineering Science & Mechanics, Penn State University, University Park, PA 16802, *U.S.A.*
E-mail: jpc@crash.esm.psu.edu

AND

D. CHELIDZE

Mechanical Engineering & Applied Mechanics, University of Rhode Island, Kingston, RI 02881, *U.S.A.*
E-mail: chelidze@egr.uri.edu

(*Received* 17 *December* 1999, *and in final form* 26 *July* 2001)

We present a method of optimal tracking for chaotic dynamical systems with a slowly drifting parameter. The net drift in the parameter is assumed to be small: this makes detecting and tracking the drift more difficult. The method relies on the existence of underlying deterministic behavior in the dynamical system, yet neither requires a system model nor develops one. We begin by describing an experimental study where a heuristic optimality criterion gave good tracking performance: the tracking method there was based on maximizing smoothness and overall variation in the drift observer, which was found by solving an eigenvalue problem. We then develop a theory, based on simplifying assumptions about the chaotic dynamics, to explain the success of the tracking method for chaotic systems. For signals from deterministic systems that are sufficiently complex in a sense that we make precise, typical drift observers provide poor tracking performance and require the drift to be particularly slow. In contrast, our theory shows that the optimality criterion seeks out a special drift observer that both provides better tracking performance and allows the drift to be appreciably faster. For periodic or quasiperiodic systems (no chaos), good tracking is easily achievable and the present method is irrelevant. For stochastic systems (no determinism), the optimal tracking method does not asymptotically improve tracking performance. Exhaustive numerical simulations of a simple drifting chaotic map, first without and then with stochastic forcing, show agreement with theoretical predictions of tracking performance and validate the theory.

1. INTRODUCTION

We wish to track a slowly drifting parameter in a dynamic system using observations of only the fast variables in the system. Formally, we consider systems governed by maps of the form

$$x_{k+1} = F(x_k, \mu_k, k) + \xi_k \quad \text{and} \quad \mu_{k+1} = \mu_k + \varepsilon G(x_k, \mu_k, k), \tag{1}$$

where x is a vector of "fast" variables, μ is a slow scalar variable, $0 < \varepsilon \ll 1$ governs time scale separation between the fast and slow variables, F and G are unknown but well-defined functions that are as smooth as necessary, and ξ_k is a random variable that represents input noise to the system. In this paper, we will take $\xi_k = 0$ except for one demonstration in section 7.2.

We are interested in tracking small net changes in μ, using observations of x and no detailed knowledge of F and G. That is, we are interested in situations where we believe that an underlying deterministic model exists, but neither know what it is nor wish to find it: we are in the middle ground between deterministic, model-based, parameter estimation on the one hand and purely statistical data analysis on the other.

As motivation, in systems with evolving damage, "slow" damage variables (e.g., a crack length) act like drifting parameters in a "fast" system (e.g., a rotating shaft). Typically, the fast variables x (e.g., vibrations) are directly monitored while the slow variable μ must be indirectly estimated. Thus, monitoring slow damage evolution is equivalent to tracking slow parameter drift. Note that modelling and parameter estimation in many systems of practical interest† is not a trivial matter; as such, our focus on trying to track drift without actually developing a detailed model for the drifting system is in line with practical considerations.

The problem of tracking parameter drift in the context of machinery condition monitoring is an active applied research topic (with, e.g., regular conferences devoted to the topic [1, 2]). The conceptually simplest approach to tracking parameter drift is to repeatedly identify the parameters of the system as they slowly drift. The problem of non-linear system identification and parameter estimation is certainly not new (see, e.g., references [3–6]). In the literature, several different approaches to parameter estimation in dynamic systems have been reported (see, e.g., references [7–13]). All these methods can, in principle, be applied once the structure of a model for the system is available. There are also many papers that address the problem of damage detection in specific classes of systems. For some recent examples in the area of structural dynamics, see references [14–23]. Finally, there are many papers that focus on straightforward techniques of time–frequency domain data analysis or statistical methods to detect damage (e.g., references [24, 25]), as well as various techniques motivated by ideas and results from dynamical systems theory (e.g., references [26–31]) that provide varying levels of ability to detect and monitor slowly changing dynamics.

Here, we focus on tracking slow parameter drift in the presence of chaotic dynamics. We concentrate on systems that are difficult to model accurately, or for which a model structure is not known, so that direct estimation of relevant parameters is not feasible. We would like to avoid *ad hoc* methods and find a technique that has some demonstrably superior tracking performance. An issue which we have not seen discussed in the health monitoring literature, but which we address here, is the way in which chaotic dynamics (in a sense that we define in section 4) can make parameter tracking difficult when compared to systems with regular dynamics.

In this paper, we demonstrate that the statistical convergence properties of chaotic signals make it difficult, on the one hand, to track drift using typical drift observers, but on the other allow the existence of a special (optimal) drift observer whose performance is much better, asymptotically for small slow drift and large data set size, than that of typical drift observers. In addition, we show that this optimal drift observer can be found using a straightforward calculation that is well suited to direct experimental application.

To our knowledge, there is no prior work that provides a general, theoretical, and model-based explanation for the success of a simple, heuristic, and data-based tracking procedure for chaotic systems, where the success of the procedure is crucially dependent on the existence of deterministic chaotic dynamics: that, in a line, summarizes the contribution of this paper.

†Such as internal combustion engines, gearboxes, turbines, machine tools, polluted rivers, stock markets, and insect populations.

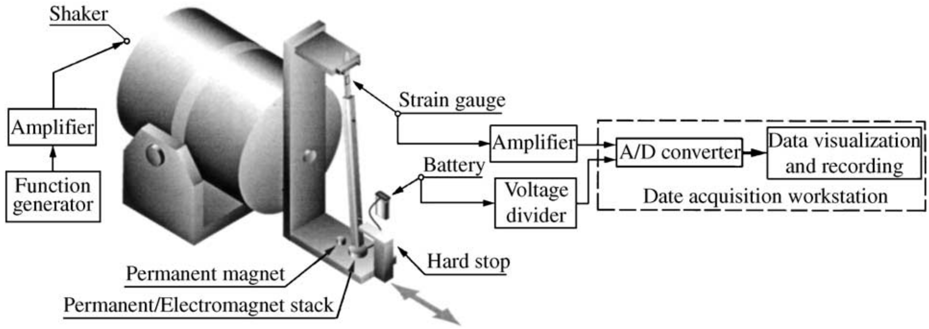

Figure 1. Experimental set-up.

2. BACKGROUND

In section 2.1, we define some terms used in the rest of this paper. A new optimal method for tracking parameter drift was developed and tested experimentally in reference [32]. Since the results from that work provide the main motivation for the theory presented later in this paper, in the remainder of this section we briefly describe the experiment and summarize the results.

2.1. TERMINOLOGY

The directly measured fast variable x is a *signal*. The signal is quantitatively characterized using some suitable statistical measures or *features*: these might be, e.g., its r.m.s. value, its autocorrelation, etc. Several features together are called a *feature vector*. Each feature vector is computed from a *record* consisting of a finite number of samples. As the system drifts, many records are collected, and as many feature vectors are computed. A scalar function of a feature vector is called a *tracking metric* or a *drift observer* if its graph versus the record-averaged graph of μ is one to one. Ideally, we would like this graph to be a straight line with non-zero slope.

2.2. EXPERIMENT

The experimental system studied in reference [32] was a cantilevered beam with stiffeners that roughly constrained it to one degree of freedom (d.o.f.), with a two-well potential created with permanent magnets. One magnet was augmented by a battery powered electromagnet. The system, mounted on a shaker, was forced at 5·6 Hz. The beam intermittently impacted a hard stop. The set-up is shown schematically in Figure 1.

For this system, the forcing amplitude was set to obtain apparently chaotic motions. The battery drained itself over several hours, weakening the electromagnet (thus affecting the system dynamics). Strain gauge output and battery voltage were sampled at 180 Hz, digitized and stored. In this system, the slowly draining battery strength acts as a drifting parameter. The strain gauge monitors the beam's deflection (the fast variable x).

2.3. NEW HEURISTIC TRACKING METHOD

Let $x(k)$ denote the discretely sampled strain gauge output. Our task is to process the data $x(k)$ in order to obtain, over time, a representation of the drift process (battery discharge).

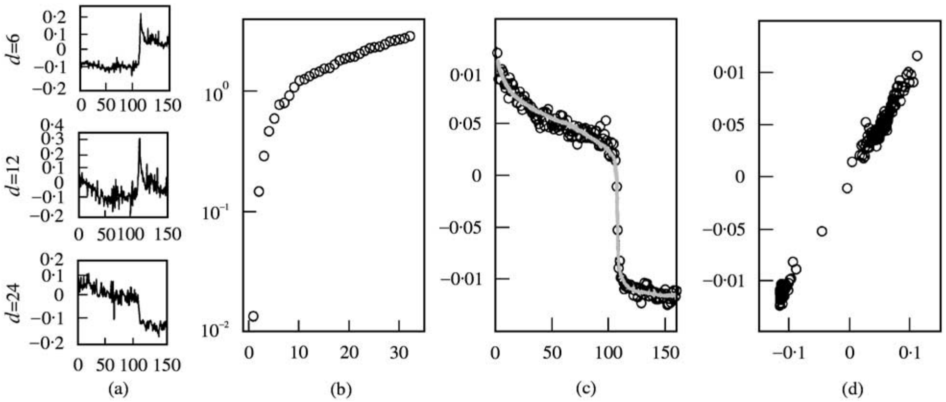

Figure 2. (a) Graphs of some columns of matrix Y ($d =$ column number, 6, 12 and 24). (b) Eigenvalues of (A_1, B_1). (c) Voltage (——) and tracking metric (○○○). (d) Calibration curve.

We divided (see reference [32]) the total data into 160 records (about 840 forcing periods each). For each record, we computed the autocorrelation of x, i.e., we averaged $x(k)x(k+d)$, for $d = 1, 2, 3, \ldots, 32$ ($d = 32$ is about one forcing period). This gave 32 vectors (160×1); from these vectors we subtracted their means, and scaled them to unit norm. We arranged these in a 160×32 matrix, called Y. Thus, element Y_{ij} represented the mean value over record i, of $x(k)x(k+j)$, mean-subtracted and scaled. Each row of Y was a feature vector, and each column represented the time history of one particular feature.

The graphs of the columns of Y were non-smooth, and unsuitable for tracking drift (three of them are shown in Figure 2(a)). Of them all, $d = 24$ provides best tracking, but that is known in this case only because, by experiment design, we separately monitored the battery voltage.

Thus, each feature itself was a poor tracking metric or drift observer. Next, we considered candidate drift observers that are linear combinations of the features, i.e., given by $v = Yc$. Note that since the underlying drift process is smoothly varying, good drift observers must vary smoothly as well. We therefore asked: can smoothness be used as a criterion for selecting a suitable drift observer?

To this end, we took the 159×160 discrete derivative matrix W given by

$$W = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix}.$$

We minimized, with respect to c, the ratio

$$g(v) := \frac{\|Wv\|^2}{\|v\|^2}. \tag{2}$$

The rationale was that if v is jagged or non-smooth, then $\|Wv\|$ is high. Meanwhile, since v has zero mean, $\|v\|$ is a measure of its total variation. In minimizing g, we maximize smoothness and overall variation, and expect to obtain a good tracking metric or drift observer. The minimum of g is the smallest generalized eigenvalue of (A_1, B_1), where $A_1 := (WY)^T(WY)$ and $B_1 := Y^TY$. The corresponding eigenvector gives the optimal c and hence, v (note that the minimization is with respect to the 32-dimensional c, which is unconstrained; the 160-dimensional v is constrained by the relation $v = Yc$).

Figure 2(b) shows the eigenvalues of (A_1, B_1) obtained in reference [32]. The smallest one is much smaller than the next, implying that the calculation is robust. The optimal metric, which happens to be 4·5 times better (in an r.m.s. sense) than the $d = 24$ column of Y, is plotted in Figure 2(c) with circles; the measured battery voltage (mean-subtracted and normalized) is plotted with a heavy gray line. Figure 2(d) shows these quantities plotted against each other: the curve is *linear*.

3. SCOPE OF THIS PAPER

Based on the experimental results summarized in the previous section, in this paper we develop an asymptotic theory for the performance of tracking metrics. These metrics are calculated by averaging suitable scalar functions of the measured fast variable x over a number S of records of size N each (S and N are both large). Over each record, the drifting parameter is treated as constant (i.e., the system is treated as stationary; the drift rate and net drift are both small). The values of the tracking metrics calculated for the different records are not constant, but change slightly from record to record. Tracking this change enables indirect tracking of the drifting parameter. Within this framework, the development of this paper may be summarized as follows:

1. On averaging stationary random quantities in samples of size N, the averages converge at least as fast as $1/\sqrt{N}$ (by the central limit theorem). But the central limit theorem does not preclude *faster* convergence for some signals from some systems. For example, averages of periodic or quasiperiodic signals converge like $1/N$.
2. Because of the statistical convergence properties of chaotic systems (as discussed in section 4), typical tracking metrics converge like $1/\sqrt{N}$, and tracking is poor. However, for certain *special* choices of tracking metrics, convergence is faster and tracking quality improves dramatically. In this paper, we show that these special tracking metrics are, in fact, asymptotically equivalent to the heuristically obtained tracking method of the previous section (which, as we have shown, can be found easily be solving an eigenvalue problem).
3. For periodic or quasiperiodic systems, convergence is already fast, and the above method is irrelevant or unnecessary.
4. For stochastic systems (whose statistical convergence properties are like those for chaotic systems, but which lack determinism), the tracking method does not provide asymptotically better tracking quality.
5. The asymptotic theory developed in this paper predicts items 2–4 above, and provides scaling rules for the tracking performance in these cases. These scaling rules have been verified with detailed numerical simulations of some simple systems.

4. STATISTICAL CONVERGENCE AND OUR DEFINITION OF CHAOS

In qualitative terms, chaos is associated with time series that can be thought of as lying midway between regular and stochastic motions. Technically, of course, chaos is associated with the existence of "strange" invariant sets in phase space, and with sensitive dependence on initial conditions. In fact, a commonly accepted working definition of chaos is a motion from a deterministic system possessing at least one positive Lyapunov exponent.

However, these characteristics of chaotic dynamics are not important for the theory developed in this paper. Rather, we make use of the fact that chaotic signals correspond to deterministic time series with a sort of statistical complexity that we now make precise.

Consider a stationary chaotic system at steady state. Let $x(k)$ be a discretely sampled state variable of this system. Consider any statistic $E[f(x)]$, where E represents the expected value. Say we collect a sample of size $N \gg 1$. Let the mean value of $f(x)$ for that sample be $\langle f(x_k)\rangle_N$. We assume (as is widely assumed in experimental work, and as is consistent with the convergence rate guaranteed by the central limit theorem) that

$$E[f(x)] - \langle f(x_k)\rangle_N = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right), \tag{3}$$

where the order notation is to be interpreted in a statistical sense, i.e.,

$$a = b + \mathcal{O}(\varepsilon) \quad \text{implies} \quad E[(a-b)^2] = \mathcal{O}(\varepsilon^2). \tag{4}$$

Now, for periodic signals, as well as for quasiperiodic signals, equation (3) still applies in the sense that the $\mathcal{O}(1/\sqrt{N})$ is actually the smaller or *faster* converging[‡] quality $\mathcal{O}(1/N)$. For this paper, we *define* a system to be chaotic if and only if it is deterministic but typical statistics converge *no faster* than $\mathcal{O}(1/\sqrt{N})$. This assumption represents our mathematical description of the "irregularity" of chaotic dynamics.

In this paper, we do not use sensitivity to initial conditions except in the sense that, due to the presence of minute unmodelled effects, it effectively gives the system a finite prediction horizon, with the following consequence. Let two large samples, each of size $N \gg 1$, be collected. Each sample consists of consecutive values of $x(k)$, and the two samples are themselves collected non-concurrently. Under these conditions, we assume that (due to the finite prediction horizon and the largeness of N), essentially all elements in the first sample are effectively uncorrelated with essentially all elements in the second sample. Under this assumption, the $\mathcal{O}(1/\sqrt{N})$ error in computing a typical statistic from the first sample is statistically independent of, and identically distributed as, the $\mathcal{O}(1/\sqrt{N})$ error in computing the same statistic from the second sample.

Finally, it is well known that, as a parameter is varied over an interval, chaotic systems rarely stay chaotic over the entire interval. However, in our theory, we assume the simplest case where the system is in fact chaotic over the entire range[§] of the parameter of interest. As evidenced by the foregoing experimental results with the vibroimpact system (where the system did pass through several periodic windows) and the simulations of a simple drifting map later in this paper, it appears that these periodic windows may in many cases not be important: the main conclusions derived from the theory appear to hold anyway. One way to view our approach is to note that it is, in some sense, similar to the neglect of dry friction in the vibration analysis of an airplane wing: it is not that the dry friction does not exist, but

[‡]Note, however, that for quasiperiodic signals with large numbers of frequencies (say of the order of 100 or more), N may have to be very large before this asymptotic convergence rate is achieved.

[§]Strictly speaking, the drifting system is not chaotic because it is not at steady state. However, since the drift is slow, we make the approximation that over intermediate time scales (long compared to fast dynamics, but short compared to drift dynamics) the difference is negligible between the actual fast dynamics and the dynamics of a system with the drifting parameter frozen at a representative current value.

rather that the simpler model of viscous friction enables satisfactory description of the phenomena of interest.

5. THEORY

To begin our theory of tracking, we first quantitatively define tracking quality via the tracking error Q, given by

$$Q^2 := \frac{\text{Mean square value of deviations from the correct underlying curve}}{\text{Mean square value of deviations of the underlying curve about its own mean value}}. \tag{5}$$

What the correct underlying curve is will become clear below. Obviously, a drift observer is useful only if Q is sufficiently smaller than unity.

5.1. NOTATION

We now fix some notation. In what follows, $x(k)$ represents the kth data point in some record. In the rest of this section, the index k will be used exclusively for this purpose. Each record has N points. The total number of records is denoted by S. The index n will refer to the record number in the rest of this section. For each record, a number of features will be computed. The index i will refer to the feature number. The drifting parameter, as mentioned earlier, is denoted by μ. The average value of μ over the entire experiment will be denoted by $\bar{\mu}$. The average value of μ over the time duration of record number n will be denoted by $\bar{\mu}_n$. The variance of the $\bar{\mu}_n$ will be denoted by σ_μ^2 (we do not assume here that μ varies stochastically; the σ_μ notation is merely for convenience in some later calculations). Angled brackets will be used to denote a sample average; when necessary, the sample size will be denoted by a right subscript, as in $\langle \cdot \rangle_N$.

5.2. ASSUMPTIONS, AVERAGING, AND SIMPLIFICATIONS

We now assume that the drift in the parameter μ is small at all times, and always comparable to σ_μ; that the number of records collected (S) is large; and that the size of each record (N) is also large.[¶] These assumptions allow the following simplifications.

First, because the net drift is small and the number of records is large, over any one record the drift is doubly small and neglected. Thus, over any one record, we treat the system as essentially a stationary system.[‖]

For a chaotic system, following the discussion of section 4, any averages computed over each record (for calculating the features) converge to their expected values like $1/\sqrt{N}$ and no faster. For example,

$$\langle x(k)\rangle_N = \langle x(k)\rangle_\infty + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right),$$

[¶] Note that smallness of σ_μ is no serious restriction, since large drift (or a large change in dynamics) is easier to detect. Largeness of S and N is allowed by slow drift, and is relevant for many systems of practical interest.

[‖] While passage through significant bifurcations can, in principle, lead to violations of this simplifying assumption, in our experimental studies as well as in the numerical simulations presented later in this paper, such bifurcations have not been significant. Here, for simplicity, we assume this difficulty away.

where $\langle x(k)\rangle_\infty$ is the expected value of $x(k)$ for that record. If we compute the averages of several functions $f_1(x(k)), f_2(x(k)), \ldots$, then for each such function $f_i(x(k))$

$$\langle f_i(x(k))\rangle_N = \langle f_i(x(k))\rangle_\infty + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \tag{6}$$

and no faster, due to chaos.

For periodic or quasiperiodic systems, the averages converge at an asymptotically faster rate, with

$$\langle f_i(x(k))\rangle_N = \langle f_i(x(k))\rangle_\infty + \mathcal{O}\left(\frac{1}{N}\right) \quad \text{if no chaos.} \tag{7}$$

For brevity, we now write

$$\langle f_i(x(k))\rangle_\infty = \bar{f}_i(\bar{\mu}_n), \tag{8}$$

where the explicit dependence on $x(k)$ has been dropped; and the implicit dependence on the value of the drifting parameter in the record of interest has been shown. We also assume that $\bar{f}_i(\mu)$ allows a good local linear approximation,††

$$\bar{f}_i(\bar{\mu}_n) = \bar{f}_i(\mu) + \bar{f}_i'(\bar{\mu})\cdot(\bar{\mu}_n - \bar{\mu}) + \mathcal{O}(\sigma_\mu^2).$$

Henceforth, the small $\mathcal{O}(\sigma_\mu^2)$ term will be dropped in comparison with $\mathcal{O}(\sigma_\mu)$ terms.

5.3. TYPICAL DRIFT OBSERVERS

Let us pick any function $f_i(x(k))$ as a candidate drift observer. What kind of tracking quality may we expect? Averaging f_i over record length N for the nth record, we find from equation (6) that

$$\langle f_i(x(k))\rangle_N(\bar{\mu}_n) = \bar{f}_i(\bar{\mu}) + \bar{f}_i'(\bar{\mu})\cdot(\bar{\mu}_n - \bar{\mu}) + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right), \tag{9}$$

where on the left-hand side, the dependence on record number has been explicitly shown. On the right-hand side of equation (9), the first term is a constant, and does not affect tracking quality (see equation (5)). The second term is the correct underlying curve, while the third term represents random, uncorrelated fluctuations about this curve that come from finite-sample averaging of a chaotic process (see section 4). Now the numerator of equation (5) is trivially seen to be $\mathcal{O}(1/N)$, while the denominator is seen to be $\{\bar{f}_i'(\bar{\mu})\}^2\sigma_\mu^2$, giving

$$Q^2 = \mathcal{O}\left(\frac{1}{\{\bar{f}_i'(\bar{\mu})\}^2\sigma_\mu^2 N}\right).$$

Assuming, as an observability condition, that $\{\bar{f}_i'(\bar{\mu})\}^2 = \mathcal{O}(1)$ (and strictly non-zero), we see that good tracking is possible only if

$$N \gg \frac{1}{\sigma_\mu^2}. \tag{10}$$

††We do *not* assume that $\bar{f}_i(\mu)$ is differentiable. We merely assume that we are in a parameter regime where, in some small μ-range of interest, $\bar{f}_i(\mu)$ allows a good linear approximation. The coefficient of the linear term, here denoted with a prime, can be taken to mean the derivative when a continuous derivative exists.

Moreover, for sufficiently large N,

$$Q \propto \frac{1}{\sigma_\mu \sqrt{N}}. \tag{11}$$

We do not yet make an explicit connection with the experimental study; note, however, that such poor convergence ($O(1/\sqrt{N})$) is expected to occur in calculations of autocorrelations as well, compromising tracking quality.

5.4. DETERMINISTIC DYNAMICS, AND AN OPTIMAL DRIFT OBSERVER

We return to equation (1) (with $\xi_k = 0$). Consider the first of these equations, and assume for simplicity that the evolution of $\mu(k)$, as given by the second equation, is given. In other words, consider a dynamical system of the form

$$z_{k+1} = F(z_k, \mu_k, k), \tag{12}$$

where μ_k is a well-defined (but unknown), slowly varying function of k with small overall variation.

In what follows, the explicit dependence of the function F on k can be dropped with no loss of generality.‡‡ We then have

$$z_{k+1} = F(z_k, \mu_k). \tag{13}$$

Since the overall variation in μ is small, we write equation (13) as

$$z_{k+1} = F(z_k, \bar{\mu}) + F_\mu(z_k, \bar{\mu})(\mu_k - \bar{\mu}) + \mathcal{O}(\sigma_\mu^2), \tag{14}$$

where F_μ is the partial derivative of F with respect to μ. For the nth record we write equation (14) as (dropping the small $\mathcal{O}(\sigma_\mu^2)$ term as before)

$$z_{k+1} = F(z_k, \bar{\mu}) + F_\mu(z_k, \bar{\mu})(\bar{\mu}_n - \bar{\mu}). \tag{15}$$

Since $\bar{\mu}$ is a constant for the entire experiment, we can consider F and F_μ simply as functions, of z_k. Let these functions be represented, to satisfactory accuracy, as linear combinations of a finite number of basis functions:

$$F(z, \bar{\mu}) = F(z) = \sum_{m=1}^{M_1} a_m g_m(z), \qquad F_\mu(z, \bar{\mu}) = F_\mu(z) = \sum_{m=1}^{M_2} b_m h_m(z),$$

where M_1 and M_2 are much smaller than either S or N, and are treated as $\mathcal{O}(1)$. Substituting the above into equation (15), we obtain

$$z_{k+1} = \sum_{m=1}^{M_1} a_m g_m(z_k) + \sum_{m=1}^{M_2} b_m h_m(z_k)(\bar{\mu}_n - \bar{\mu}). \tag{16}$$

‡‡This is possible since one can always write equation (12) in an autonomous form by defining a new variable $u_k = \{x_k^{\mathrm{T}}, k\}^{\mathrm{T}}$. In practice, this situation arises when such systems are studied using delay-co-ordinate embedding, in which case the explicit k-dependence disappears.

Averaging the above over one record, we obtain

$$\langle z_{k+1}\rangle_N = \sum_{m=1}^{M_1} a_m \langle g_m(z_k)\rangle_N + \sum_{m=1}^{M_2} b_m \langle h_m(z_k)\rangle_N (\bar{\mu}_n - \bar{\mu}). \tag{17}$$

Now, note that

$$\langle z_{k+1}\rangle_N = \langle z_k\rangle_N + \mathcal{O}\left(\frac{1}{N}\right),$$

because only the two endpoints in the sum are different. However, z_k itself is trivially a function of z_k, and so can be absorbed into the first sum on the right-hand side of equation (17), to yield

$$\sum_{m=0}^{M_1} a_m \langle g_m(z_k)\rangle_N + \sum_{m=1}^{M_2} b_m \langle h_m(z_k)\rangle_N (\bar{\mu}_n - \bar{\mu}) = \mathcal{O}\left(\frac{1}{N}\right), \tag{18}$$

where $a_0 = -1$ and $g_0(z) = z$. We now identify the functions g and h in equation (18) with the functions f in the formulas of equations (6)–(8). Accordingly,

$$\langle g_m(z_k)\rangle_N = \langle g_m(z_k)\rangle_\infty + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) = \bar{g}_m(\bar{\mu}_n) + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \tag{19}$$

and

$$\langle h_m(z_k)\rangle_N = \langle h_m(z_k)\rangle_\infty + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) = \bar{h}_m(\bar{\mu}_n) + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right), \tag{20}$$

where the dependence has been shown explicitly of the average on the value of μ in the record of interest.

Substituting equation (19) into the first sum in equation (18), we obtain

$$\sum_{m=0}^{M_1} a_m \langle g_m(z_k)\rangle_N = \sum_{m=0}^{M_1} a_m \bar{g}_m(\bar{\mu}_n) + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{21}$$

Similarly, substituting equation (20) into the second sum in equation (18), we obtain

$$\sum_{m=1}^{M_2} b_m \langle h_m(z_k)\rangle_N (\bar{\mu}_n - \bar{\mu}) = \sum_{m=1}^{M_2} b_m \bar{h}_m(\bar{\mu}_n)(\bar{\mu}_n - \bar{\mu}) + \mathcal{O}\left(\frac{\sigma_\mu}{\sqrt{N}}\right). \tag{22}$$

On directly taking the limit as $N \to \infty$ in equation (18), we obtain

$$\sum_{m=0}^{M_1} a_m \bar{g}_m(\bar{\mu}_n) + \sum_{m=1}^{M_2} b_m \bar{h}_m(\bar{\mu}_n)(\bar{\mu}_n - \bar{\mu}) = 0. \tag{23}$$

We now come to a crucial argument. Substituting equations (21)–(23) into equation (18), we find that the following relationship *must hold*:

$$\mathcal{O}\left(\frac{1}{\sqrt{N}}\right) + \mathcal{O}\left(\frac{\sigma_\mu}{\sqrt{N}}\right) = \mathcal{O}\left(\frac{1}{N}\right). \tag{24}$$

Equation (24) may seem contradictory, since the terms do not appear to balance. The apparent contradiction can be resolved, however, as follows. The first term on the left, which comes *only from* equation (21), though potentially much bigger than the other two, must balance them by *construction*. Thus it can be at most as big as the bigger of the two. Using this, we rewrite equation (21) as

$$\sum_{m=0}^{M_1} a_m \langle g_m(z_k)\rangle_N = \sum_{m=0}^{M_1} a_m \bar{g}_m(\bar{\mu}_n) + \mathcal{O}\left(\frac{\sigma_\mu}{\sqrt{N}}\right) + \mathcal{O}\left(\frac{1}{N}\right). \tag{25}$$

It is worth pausing to carefully examine equation (25), obtaining which was the aim of the preceding asymptotic arguments. It demonstrates a key idea of this paper: while typical averages of dynamical quantities converge slowly due to the presence of chaos, the presence of underlying deterministic dynamics allows the existence of a special quantity whose average converges much faster. The rest is straightforward.

Suppose that by chance, we pick the functions f_i to be the same as the functions g_m. And suppose that the linear combination of the f_i that we decide to use for our drift observer happens to be the same as the sum of the left-hand side of equation (25). Then, it only remains to write the right-hand side of equation (25) as (again dropping $\mathcal{O}(\sigma_\mu^2)$ terms)

$$\text{drift observer} = \sum_{m=0}^{M_1} a_m \bar{g}_m(\bar{\mu}) + \sum_{m=0}^{M_1} a_m \bar{g}'_m(\bar{\mu})(\bar{\mu}_n - \bar{\mu}) + \mathcal{O}\left(\frac{\sigma_\mu}{\sqrt{N}}\right) + \mathcal{O}\left(\frac{1}{N}\right). \tag{26}$$

Equation (26) is of the form

$$\text{drift observer} = A + B(\bar{\mu}_n - \bar{\mu}) + \mathcal{O}\left(\frac{\sigma_\mu}{\sqrt{N}}\right) + \mathcal{O}\left(\frac{1}{N}\right). \tag{27}$$

where A and B are constants for a given experiment. A has no influence on tracking quality. Assuming, as an observability condition, that $B = \mathcal{O}(1)$ and non-zero, we see that the denominator in equation (5) is simply $B^2\sigma_\mu^2$, while the numerator is on the order of

$$\frac{\sigma_\mu^2}{N} + \frac{1}{N^2} = \frac{1 + N\sigma_\mu^2}{N^2},$$

giving the order of magnitude estimate

$$Q^2 \sim \frac{1 + N^2\sigma_\mu^2}{N^2\sigma_\mu^2}. \tag{28}$$

Since we need $Q \ll 1$ for good tracking, and the numerator in equation (28) is at least as big as unity, tracking is not possible unless $N^2\sigma_\mu^2 \gg 1$ or $N \gg 1/\sigma_\mu$ (compare with equation (10) for typical drift observers). In the range

$$\frac{1}{\sigma_\mu} \ll N \ll \mathcal{O}\left(\frac{1}{\sigma_\mu^2}\right), \tag{29}$$

we also have

$$Q \propto \frac{1}{\sigma_\mu N}, \tag{30}$$

an $\mathcal{O}(\sqrt{N})$ improvement over typical drift observers (equation 11)).

The range $N = \mathcal{O}(1/\sigma_\mu^2)$ is a transition region. Finally, for $N \gg 1/\sigma_\mu^2$, the second term in the numerator of equation (28) dominates and we have

$$Q \propto \frac{1}{\sqrt{N}}, \tag{31}$$

which is an $\mathcal{O}(1/\sigma_\mu)$ improvement over typical drift observers (equation 11)).

Based on the foregoing analysis, we conclude the following. The presence of chaotic dynamics makes tracking difficult. For typical drift observers, tracking quality is poor and data requirements are high (N needs to be large). For a given sampling rate in the data acquisition system, "high" data requirements imply that for typical drift observers drift cannot be tracked unless it is very slow. In contrast, the underlying deterministic dynamics allows the existence of *at least one* drift observer which may be called optimal in that both tracking quality *and* data requirements are much lower for this drift observer than for typical drift observers.

5.5. CONNECTION WITH EXPERIMENT

At this stage, the connection between the experimental study and the foregoing theory has not yet been established. Some qualitative points are worth noting, however.

5.5.1. *Uniqueness*

In the experimental study, it was found that the smallest eigenvalue of the system was much smaller than the next smallest one, showing empirically that the choice of a smooth tracking metric with large overall variation was unequivocal. In the foregoing analysis, we find that there is at least one good tracking metric. Is there perhaps *only one* good tracking metric? Our experiment suggests that the answer is yes; but our theory does not address this question. We will draw some empirical conclusions about this question below.

5.5.2. *Equivalence*

In the foregoing analysis, we have studied a certain measure of tracking quality. Is maximizing this tracking quality the same as maximizing smoothness and overall variation as quantified by equation (2)? We will address this question below.

5.5.3. *Use of autocorrelations*

Finally, in the experimental study we used autocorrelations, while in the theory we used the state variables directly. Does the theory still apply to the experiment? We will address this question in Appendix A.

5.6. EQUIVALENCE OF TRACKING QUALITY AND HEURISTIC OPTIMALITY CRITERION

So far we have assumed that the number of records, S, is large and fixed; that it is *large* allowed us to treat the drifting parameter as constant over each record. Now, imagine that, we are free to hold the μ variation σ_μ as well as the record size N constant, while increasing the number of records,[§§] S.

Consider a candidate drift observer $y(n)$ (where, as before, n refers to record number). Let

$$y(n) = a + \bar{\mu}_n + e_n, \tag{32}$$

where a is an undetermined constant, $\bar{\mu}_n$ is the correct underlying drift process (or the true observer), and the e_n's represent zero mean, mutually independent (see section 4), random fluctuations that arise due to the finite-length averaging involved in computing the drift observer. Let the r.m.s. value of the e_n and σ_e. For simplicity, let $E(e_i e_j) = \delta_{ij}\sigma_e^2$, where $\delta_{ij} = 1$ if $i = j$, and zero otherwise. That is, let the expected size of the fluctuation be independent of the record number. Observe that

$$\sum_{n=1}^{S} e_n = \mathcal{O}(\sqrt{S\sigma_e}) \tag{33}$$

and

$$\sum_{n=1}^{S-1} e_n e_{n+1} = \mathcal{O}(\sqrt{S\sigma_e^2}). \tag{34}$$

Since we include the constant a in equation (32), we assume with no loss of generality that

$$\sum_{n-1}^{S} \bar{\mu}_n = 0. \tag{35}$$

The tracking quality for this case (see equation (5)) is

$$Q^2 = \frac{\sigma_e^2}{\sigma_\mu^2}. \tag{36}$$

Since the underlying locally averaged drift process is smooth, we write

$$\bar{\mu}_{n+1} = \bar{\mu}_n + \frac{\bar{\mu}_n'}{S} + \mathcal{O}\left(\frac{1}{S^2}\right), \tag{37}$$

where $\bar{\mu}_n'$ is a local rate of change of the drifting parameter that is independent of the mesh-refinement S. In other words, if the time duration of the entire experiment is scaled to unity, then the width of each record in that slow time is $1/S$; and $\bar{\mu}_n'$ is the derivative of the drifting parameter with respect to that slow time, independent of record width.

Since, in the heuristic tracking procedure of section 2.3, the columns of Y were shifted to zero mean, we have (by summing equation (32) using equations (33) and (35))

$$a = \mathcal{O}\left(\frac{\sigma_e}{\sqrt{S}}\right). \tag{38}$$

Finally,

$$\sum_{n=1}^{S} a e_n = \mathcal{O}(\sigma_e^2), \qquad \sum_{n=1}^{S} a\bar{\mu}_n = 0, \qquad \sum_{n=1}^{S} \bar{\mu}_n e_n = \mathcal{O}(\sqrt{S}\sigma_\mu\sigma_e). \tag{39–41}$$

[§§]In the context of our experiment this would correspond, for example, to having different batteries that drained over 7, 70, 700, 7000 h, …, while keeping the initial battery voltage and the size of each data record fixed. In principle, this could be achieved by using batteries with 10, 100, 1000, …, the capacity in amp-hours of the original battery. Such experiments are impractical, but imagining them helps us develop our asymptotic theory.

Now consider equation (2). The numerator is

$$\sum_{n=1}^{S-1} (\bar{\mu}_{n+1} e_{n+1} - \bar{\mu}_n - e_n)^2,$$

which after some manipulations is

$$2S\sigma_e^2 - e_1^2 - e_s^2 + \mathcal{O}(\sqrt{S}\sigma_e^2) + \sum_{n=1}^{S-1} \frac{(\bar{\mu}_n')^2}{S}. \tag{42}$$

Note that

$$\sum_{n=1}^{S-1} \frac{(\bar{\mu}_n')^2}{S} = \sum_{n=1}^{S} \frac{(\bar{\mu}_n')^2}{S} + \mathcal{O}\left(\frac{1}{S}\right) = \overline{(\bar{\mu}')^2} + \mathcal{O}\left(\frac{1}{S}\right),$$

for some constant $\overline{(\bar{\mu}')^2}$ that is fixed for the experiment. In equation (42), on dropping the smaller e^2 terms as negligible compared to the dominant $2S\sigma_e^2$, we therefore have

$$\overline{(\bar{\mu}_2')} + 2S\sigma_e^2. \tag{43}$$

The denominator of equation (2) is

$$\sum_{n=1}^{S} (a + \bar{\mu}_n + e_n)^2 = \sum_{n=1}^{S} \{a^2 + (\bar{\mu}_n)^2 + e_n^2 + 2ae_n + 2a\bar{\mu}_n + 2\bar{\mu}_n e_n\},$$

$$= \mathcal{O}(\sigma_e^2) + S\sigma_\mu^2 + S\sigma_e^2 + \mathcal{O}(\sigma_e^2) + 0 + \mathcal{O}(\sqrt{S}\sigma_\mu\sigma_e),$$

using equation (38), the definition of σ_μ, the definition of σ_e, equations (39), (40) and (41), respectively. Thus, the denominator of equation (2) is

$$S\sigma_\mu^2 + S\sigma_e^2 + \text{smaller terms.} \tag{44}$$

By equations (43) and (44), for large S, the heuristic optimality criterion is seen to minimize

$$\frac{\overline{(\bar{\mu}')^2} + 2S\sigma_e^2}{S\sigma_\mu^2 + S\sigma_e^2},$$

where the effect of the parameter variation rate $\bar{\mu}_n'$ is seen to become smaller and smaller for bigger and bigger S. On dropping this small term as well, it is finally clear that for large S the heuristic optimality criterion minimizes a function that approximately equals

$$\frac{\sigma_e^2}{\sigma_\mu^2 + \sigma_e^2}.$$

Minimizing the above quantity is the same as maximizing its reciprocal,

$$\frac{\sigma_\mu^2 + \sigma_e^2}{\sigma_e^2} = \frac{\sigma_\mu^2}{\sigma_e^2} + 1.$$

Dropping the constant, maximizing σ_μ^2/σ_e^2 is in turn the same as minimizing its reciprocal σ_e^2/σ_μ^2. Comparing this with equation (36) we conclude that, asymptotically for large S, the heuristic optimality criterion is in fact *identical* to optimal tracking quality as given by minimizing equation (5). At this point, we make an empirical observation about uniqueness (question 1, section 5.5).

Section 5.4 shows, by construction, that there is at least one drift observer whose tracking quality is much better than that of typical ones. This subsection shows that maximizing tracking quality is equivalent to the heuristic optimality criterion (section 5.5.2), assuming that using autocorrelations does not cause any difficulty (section 5.5.3). Therefore, pending resolution of the autocorrelations issue (section 5.5.3), we have shown *by construction* that there is at least one drift observer that satisfies the *heuristic criterion* much better than others. Meanwhile, in the numerical results of section 2.3, the smallest eigenvalue is much smaller than the others, and the choice of a good drift observer is therefore unequivocal and robust, in the sense that small changes in the choice of optimality criterion will only cause small changes in the choice of drift observer. Thus, theory shows that there is at least one really good drift observer, and numerics show that there is only one. Pending the resolution of the question in section 5.5.3, therefore, we can empirically conclude that the drift observer found from theory is effectively the same as the one found using heuristics.

This last remaining gap in the theory (see section 5.5.3) is addressed in Appendix A. In the next section, we proceed to numerically validate the theory.

6. STUDY OF A DRIFTING CHAOTIC MAP

The theory presented in the previous section provides a feasible explanation for why the tracking method works, based on strong simplifying assumptions about the dynamics and some asymptotic arguments based on those assumptions. In experiments, the asymptotic validity of the arguments used above cannot be verified easily because it is difficult to vary the drift rate over several orders of magnitude (see footnote † in section 5.6). We, therefore, attempt to validate the theory using extensive numerical simulations. To better understand the tracking method, we numerically study a slowly drifting logistic map, given by

$$x_{k+1} = \mu_k x_k (1 - x_k),$$

where μ_k is a slowly varying function of k (we arbitrarily take a half sinusoid of small amplitude, within a chaotic range). The features used are the record-averages of x_k and x_k^2, these being clearly relevant to the dynamics.

The drifting logistic map is similar to the experimental system in the following key ways. The graphs of the features are non-smooth and provide poor tracking; they have a smooth linear combination that is a good tracking metric; this tracking metric is found by solving the same eigenvalue problem as earlier. However, in this case, it is possible to independently control the key features of the system (net drift, and drift rate).

In the numerical study, the parameter μ drifted slowly, in a half-sinusoid, between the values of 3·805 and 3·820: apparently chaotic dynamics occurs in this range. The number of records S was fixed at 60. Several simulations were performed. The number of data points N in each record was changed from 81, through increasing powers of 3, to 59 049. In each run, for each record, the values of x and x^2 were averaged to obtain features. The method of section 2.3 was also used to obtain the optimal drift observer. The results are shown in Figures 3 and 4.

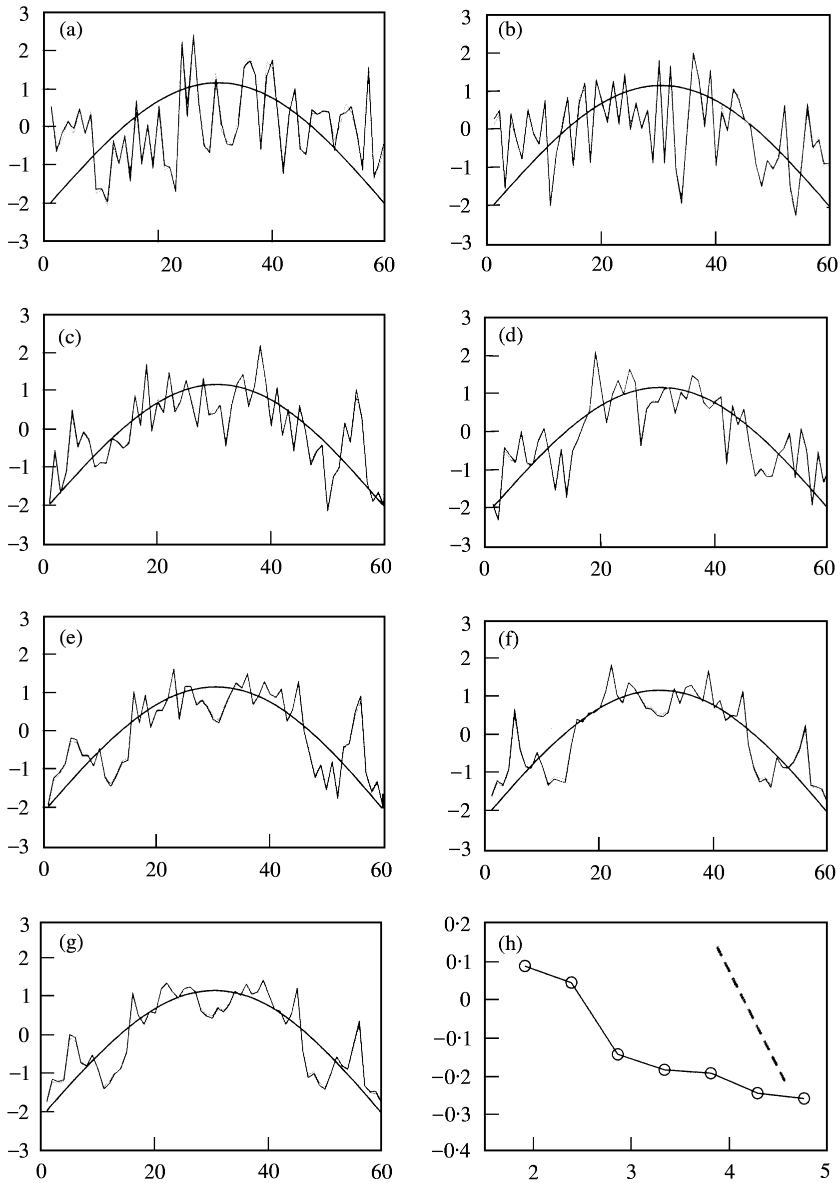

Figure 3. (a)–(g) "Typical" drift observers $\bar{x}$ and $\overline{x^2}$, mean-subtracted and scaled, plotted against record number. The variable quantity is record length N, mentioned in the individual figures. The two curves (/\/\/\/\) are nearly indistinguishable for each N. The smooth half sinusoid is the true parameter drift, shown for comparison. N = (a) 81, (b) 243, (c) 729, (d) 2187, (e) 6561, (f) 19·683, (g) 59 049 (h) log–log plot of tracking quality $\log_{10} Q$ versus $\log_{10} N$. Tracking is poor, at least in this range of N. The dashed gray line has slope $-\frac{1}{2}$.

In Figure 3, the "typical" drift observers $\bar{x}$ and $\overline{x^2}$ are shown, mean-subtracted and normalized as always, for different simulations. It is seen that good tracking is not possible even for fairly large N. The scaling is even worse than the $1/\sqrt{N}$ predicted by the theory for very large N; this may be because N is not large enough. For the same range of N, the optimal tracking method of section 2.3 is seen (see Figure 4) to provide excellent tracking.

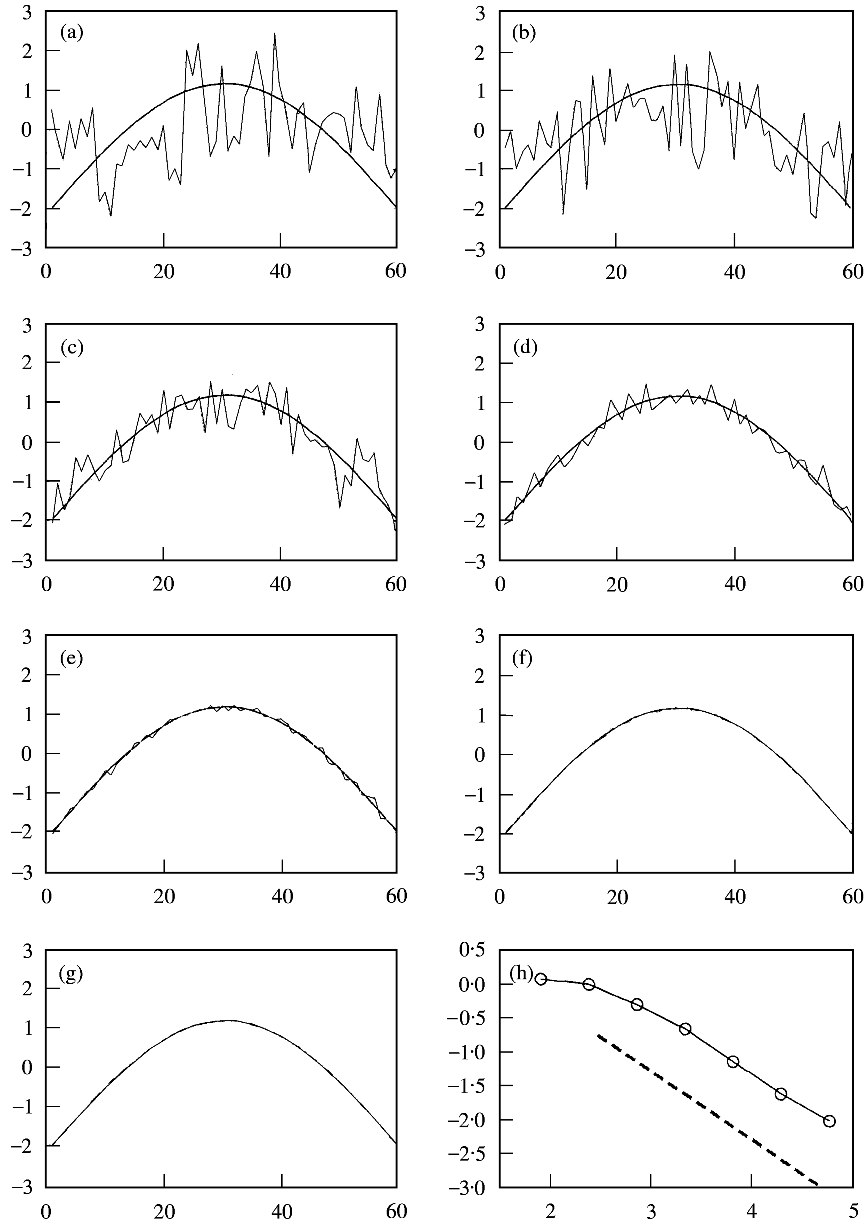

Figure 4. (a)–(g) Optimal drift observer of section 2.3, scaled, plotted against record number. The variable quantity is record length N, mentioned in the individual figures. The smooth half sinusoid is the true parameter drift, shown for comparison. N for (a)–(g) is the same as in Figure 3. (h) log–log plot of tracking quality $\log_{10} Q$ versus $\log_{10} N$. The dashed gray line has slope -1. Tracking is excellent, and tracking quality is proportional to $1/N$, as predicted by theory, and over a large scaling region.

Moreover, as predicted by the theory, the tracking quality is roughly proportional to $1/N$, for large N varying over two orders of magnitude.

7. IS CHAOS NECESSARY?

Is chaos necessary for the tracking method to succeed?

7.1. QUASIPERIODIC DYNAMICS

Our theory assumes deterministic chaos. However, what happens if the method of section 2.3 is applied directly to data from a (say) quasiperiodic system? By the assumptions of our theory, the method is *irrelevant* for quasiperiodic systems because statistics converge fast already. Computed features are all already smooth, with only $\mathcal{O}(1/N)$ fluctuations instead of the $\mathcal{O}(1/\sqrt{N})$ fluctuations encountered with chaotic dynamics. No special linear combination of the features stands out as dramatically smoother than all the rest. However, actually applying the algorithm to quasiperiodic data leads to an interesting result: the optimal tracking metric does *not*, in fact, track the drifting parameter well.[¶¶] To explain this apparent failure, we present some examples.

7.1.1. *Example* 1

First, imagine that in an experiment the parameter μ drifts linearly with time, such that (say) $\bar{\mu}_n = (n - 21)/400$. Let the features be computed to perfect accuracy. Let these features be nearly linear functions of time, which here is the same as μ. That is, we retain the small higher order terms that were previously neglected as small compared to the random errors that are now missing. For definiteness, assume that the features are the following arbitrarily selected[‖‖] polynomials:

$$p_1 = \mu + \exp(1{\cdot}5)\mu^2 - 6{\cdot}0\mu^3, \quad p_2 = \mu - \pi\mu^2 - \sqrt{5}\mu^3, \quad p_3 = \mu + \ln 7\mu^2 + \sqrt{2}\mu^3. \tag{45}$$

Let $n = 1, 2, \ldots, 41$. Applying the tracking method to the "features" p_1, p_2, p_3 above, we obtain the results shown in Figure 5(a, b).

In Figure 5(a), the features themselves are shown (mean-subtracted and scaled). Each makes a reasonable drift observer if the drift is small enough. In Figure 5(b), the results of applying the tracking method are shown. Superimposed on the plot is a nearly indistinguishable scaled plot of the cosine function over half a period.

7.1.2. *Example* 2

Now let the variation of μ not be linear with time. Accordingly, define the intermediate variable $\bar{\tau}_n = (n - 21)/400$, and let the computed features be the nearly, but not quite, identical quantities:

$$p_1 = \cos(20\tau) + \exp(1{\cdot}5)\tau^2 - 6{\cdot}0\tau^4, \quad p_2 = \cos(20\tau) - \pi\tau^2 - \sqrt{5}\tau^4,$$
$$p_3 = \cos(20\tau) + \ln 7\tau^2 + \sqrt{2}\tau^4. \tag{46}$$

Application of the tracking procedure yields the results shown in Figure 5(c, d). In Figure 5(c), note that p_1 through p_3 are almost indistinguishable. Nevertheless, the small but smooth differences between them are enough to let the minimization procedure pick out another cosine (a full period in this case), as shown in Figure 5(d).

[¶¶] For example, with vibration data from a gearbox condition monitoring experiment not described here, the results obtained appeared to be roughly like a cosine function. However, the vibration signature from a gearbox, with slowly accumulating gear tooth damage, should be steady for a long time and then change rapidly soon before failure.

[‖‖] Instead of arbitrarily selecting the coefficients of μ^2 and μ^3 in these polynomials, we could have randomly generated them, to obtain the same results. The choice made here ensures linear independence.

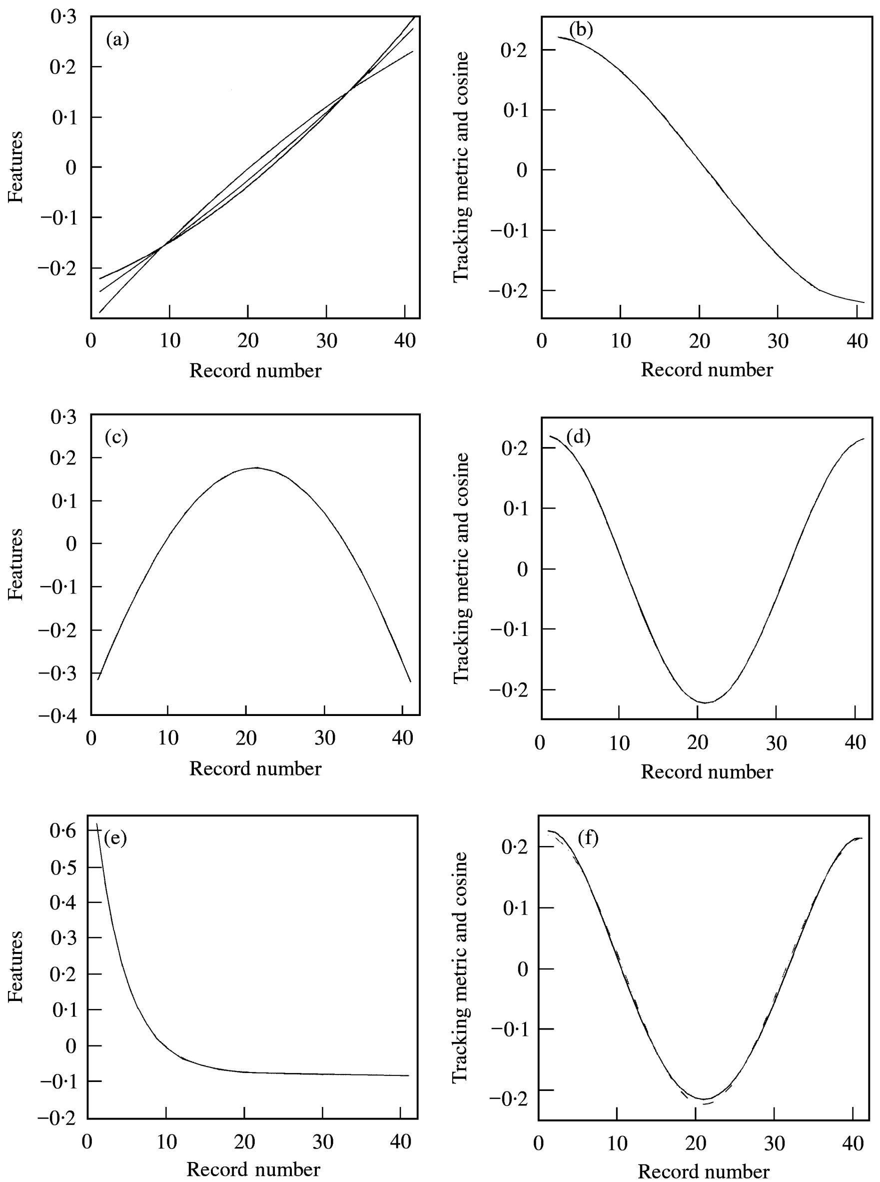

Figure 5. (a) Features for example 1, mean-subtracted and scaled. (b) Results from application of the tracking method. Also plotted is half a period of a cosine, scaled for comparison. (c) Features for example 2, mean-subtracted and scaled. (d) Results from application of the tracking method. Also plotted is a full period of a cosine, scaled for comparison. (e) Features for example 3, mean-subtracted and scaled. (f) Results from application of the tracking method. Also plotted is a full period of a cosine, scaled for comparison.

7.1.3. *Example* 3

Finally, let the computed features be the nearly, but not quite, identical quantities:

$$p_1 = \exp(-100\tau) + \exp(1{\cdot}5)\tau^2 - 6{\cdot}0\tau^4, \quad p_2 = \exp(-100\tau) - \pi\tau^2 - \sqrt{5}\tau^4,$$

$$p_3 = \exp(-100\tau) + \ln 7\tau^2 + \sqrt{2}\tau^4. \tag{47}$$

Application of the tracking procedure yields the results shown in Figure 5(e,f). In Figure 5(e), p_1–p_3 are almost indistinguishable. Nevertheless, the minimization procedure picks out a cosine yet again, as shown in Figure 5(h).

7.1.4. *Discussion*

We showed above that the tracking procedure can give misleading results, such as a cosine function regardless of what the actual variation in μ is, if there is no effective randomness in the computed statistics. That cosines might be obtained in this manner is not, in retrospect, surprising. The continuous version of the differential operator W^TW is nothing but $-(\mathrm{d}/\mathrm{d}x^2)$ with zero-slope boundary conditions; and the cosines are simply the eigenfunction of

$$-\frac{\mathrm{d}^2u}{\mathrm{d}x^2} = \lambda u, \quad u_x(0) = u_x(1) = 0.$$

Due to the excessive smoothness of the features, too many smooth functions are available to the minimization procedure, and the cosine is chosen.††† In contrast, for chaotic systems, where slowly converging statistics introduce random errors in the computed features, there is only one special choice of smooth function, and the minimization procedure is forced to choose it.

Thus, it is clear that deterministic chaos, at least during a substantial portion of the entire experiment, is a requirement for the tracking method to work. In this sense, the method is irrelevant as far as the problem of tracking drift in quasiperiodic systems is concerned. Note, however, that irrelevance to quasiperiodic systems is not a weakness. The very smoothness that makes the tracking method inapplicable for non-chaotic systems also ensures that for such systems almost *any* feature is a good drift observer. Conversely, the new tracking method directly exploits the very feature that makes it difficult to track drift in chaotic systems.

7.2. STOCHASTIC DYNAMICS

In the previous subsection, it was explained as to why the lack of "irregular" dynamics (see also section 4) in quasiperiodic systems can make the present tracking method give misleading results. In contrast to that situation, we now consider stochastically forced systems, whose dynamics is irregular but *not deterministic*. However, we expect that the lack of some especially smooth tracking metric will make the method perform poorly, with tracking quality scaling as $1/\sqrt{N}$ instead of $1/N$ (for sufficiently large N).

To this end, consider the stochastically forced drifting logistic map, given by

$$x_{k+1} = z_k + y(z_k)\eta_k,$$

where the second term represents the so far ignored ξ_k term of equation (1), z_k is the deterministic part of the map, given by

$$z_k = \mu_k x_k(1 - x_k),$$

$y(z_k)$ is defined as

$$y(z_k) = \tfrac{3}{4}\min(z_k, 1 - z_k)$$

†††In some cases, the results may not resemble cosines quite so strongly. However, they may still fail to resemble the actual underlying variation of μ.

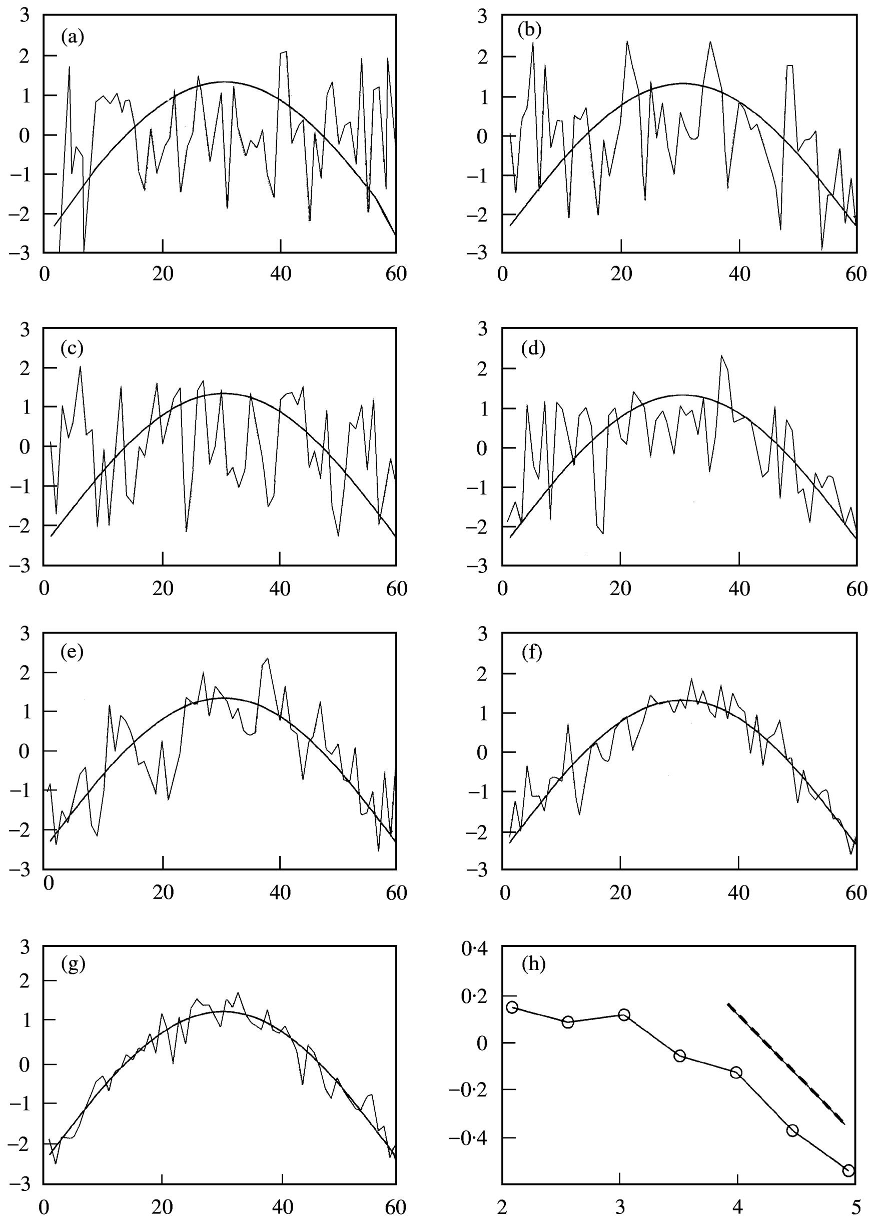

Figure 6. (a)–(g) Optimal drift observer of section 2.3, scaled, plotted against record number, for the stochastically forced drifting logistic map. As before, record length N is given in each figure. The smooth half sinusoid is the true parameter drift. $N =$ (a) 120, (b) 360, (c) 1080, (d) 3240, (e) 9720, (f) 29 160, (g) 87 480. (h) log–log plot of tracking quality $\log_{10} Q$ versus $\log_{10} N$. The dashed gray line has slope $-\frac{1}{2}$. Tracking is poorer, with tracking quality roughly proportional to $1/\sqrt{N}$ for large N, as predicted by theory.

to ensure boundedness of solutions, and η_k is a random uniformly distributed between -1 and 1. For this system, we redid the numerical experiment of section 6, with μ varying in exactly the same way as before (half sinusoid), with the same number of records (60), but with somewhat larger values of N.

The results are shown in Figure 6. It is seen in the figure that, for this stochastic system, parameter tracking is not as good as for the deterministic drifting logistic map of section 6.

In particular, for large N, the tracking quality (or error) improves roughly like $1/\sqrt{N}$ instead of $1/N$. Moreover, this slower $1/\sqrt{N}$ scaling itself shows up only for significantly larger N than the $1/N$ scaling does in the deterministic case. Both of these results are consistent with our theory.

8. CONCLUSIONS AND FUTURE WORK

We have presented a method of optimal tracking for chaotic dynamical systems with a parameter that drifts slowly and over a small range. The method relies on the existence of underlying deterministic behavior in the dynamical system, yet neither requires a system model nor develops one.

We have described an experimental study, where a heuristic optimality criterion gave good tracking performance. We have also developed a theory that explains the success of the tracking method for chaotic systems.

By our theory, for deterministic chaotic systems, typical drift observers provide poor tracking performance and require the drift to be particularly slow in order to be successfully detected. In contrast, the optimality criterion seeks out a special drift observer that both provides better tracking performance and allows the drift to be appreciably faster. For periodic or quasiperiodic systems (no chaos), good tracking is easily achievable and the present method is irrelevant. For stochastic systems (no determinism), the optimal tracking method does not asymptotically improve tracking performance.

Exhaustive numerical simulations of a simple drifting chaotic map, first without and then with stochastic forcing, have been used to validate the theory.

Future work may be directed towards broader experimental validation of the tracking method for a variety of drifting chaotic systems. On the theoretical front, the extreme simplifying assumption of negligible influence of bifurcations (such as passage through periodic windows) might be relaxed. In the experiment and simulations, such periodic windows do not appear to have seriously degraded the performance of the tracking method. Finally, it seems that in the determination of the tracking metric (which relies on the underlying determinism in the system), we are uncovering some information about the dynamics of the system; it is conceivable that a similar calculation might turn out to have applications in parameter estimation for non-linear systems as well.

ACKNOWLEDGMENTS

This work was partially supported by the Office of Naval Research MURI on Integrated Predictive Diagnostics, Grant #N0014-95-0461. Thanks to Mrinal Ghosh for a technical discussion, and to an anonymous reviewer for suggestions that led to substantial improvements in the presentation.

REFERENCES

1. H. C. PUSEY and S. C. PUSEY (editors) 1997 *Proceedings of a Joint Conference: the 51st meeting of the Society for Machinery Failure Prevention Technology (MEPT), and the 12th biennial conference on Reliability, Stress Analysis and Failure Prevention (RSAFP committee of ASME), Virginia Beach, VA.* A critical link: diagnosis to prognosis.
2. H. C. PUSEY and S. C. PUSEY (editors) 1998 *Proceedings of 52nd meeting of the Society for Machinery Failure Prevention Technology (MFPT), Virginia Beach, VA.* Prognosis of residual life of machinery and structures.
3. Y. BARD 1974 *Nonlinear Parameter Estimation.* Orlando, FL: Academic Press.

4. P. EYKHOFF 1974 *System Identification: Parameter and State Estimation*. London: John Wiley & Sons; reprinted 1977.
5. M. D. SRINATH, P. K. RAJASEKARAN and R. VISWANATHAN 1996 *Introduction to Statistical Signal Processing with Applications*. Englewood Cliffs, NJ: Prentice-Hall; Indian reprint. New Delhi: Prentice-Hall of India. 1999.
6. L. JEZEQUEL and C. H. LAMARQUE (editors) 1992 *Euromech* 280. Rotterdam, Netherlands: A. A. Balkema. *Proceedings of the International Symposium on Identification of Nonlinear Mechanical Systems from Dynamic Tests, Ecully, France*, 1991.
7. S. F. MASRI and T. K. CAUGHEY 1979 *American Society of Mechanical Engineers Journal of Applied Mechanics* **46**, 433–447. A nonparametric identification technique for nonlinear dynamic problems.
8. S. F. MASRI, R. K. MILLER, A. F. SAUD and T. K. CAUGHEY 1987 *American Society of Mechanical Engineers Journal of Applied Mechanics* **54**, 918–929. Identification of nonlinear vibrating structures: Part I—Formulation, and Part II—Applications.
9. K. S. MOHAMMAD, K. WORDEN and G. R. TOMLINSON 1992 *American Society of Mechanical Engineers Journal of Sound and Vibration* **152**, 471–499. Direct parameter estimation for linear and non-linear structures.
10. B. F. FEENY and C. M. YUAN 1999 *Proceedings of the* 1999 *ASME Design Engineering and Technical Conference, Las Vegas, Nevada*, ASME Paper no. DETC99/VIB-8361. Parametric identification of an experimental two-well oscillator.
11. D. J. MOOK 1989 *American Institute of Aeronautics and Astronautics Journal* **27**, 968–974. Estimation and identification of nonlinear dynamic systems.
12. A. CHATTERJEE and J. P. CUSUMANO 2001 *American Society of Mechanical Engineers Journal of Dynamic Systems, Measurement and Control*. Asymptotic parameter estimation via implicit averaging on a nonlinear extended system. (accepted).
13. A. MAYBHATE and R. E. AMRITKAR 2000 *Physical Review E* **61**, 6461–6470. Dynamic algorithm for parameter estimation and its applications.
14. P. F. PAI and S. JIN 2000 *Journal of Sound and Vibration* **231**, 1079–1110. Locating structural damage by detecting boundary effects.
15. S. L. TSYFANSKY and V. I. BERESNEVICH 2000 *Journal of Sound and Vibration* **236**, 49–60. Non-linear vibration method for detection of cracks in aircraft wings.
16. Y. XIA and H. HAO 2000 *Journal of Sound and Vibration* **236**, 89–104. Measurement selection for vibration-based structural damage identification.
17. Y. ZOU, L. TONG and G. P. STEVEN 2000 *Journal of Sound and Vibration* **230**, 357–378. Vibration-based model-dependent damage (delamination) identification and health monitoring for composite structures—a review.
18. P. CORNWELL, S. W. DOEBLING and C. R. FARRAR 1999 *Journal of Sound and Vibration* **224**, 359–374. Application of the strain energy damage detection method to plate-like structures.
19. W. GAWRONSKI and J. T. SAWICKI 2000 *Journal of Sound and Vibration* **229**, 194–198. Structural damage detection using modal norms.
20. M. M. ABDEL WAHAB and G. DE ROECK 1999 *Journal of Sound and Vibration* **226**, 217–236. Damage detection in bridges using modal curvatures: application to a real damage scenario.
21. R. RUOTOLO and C. SURACE 1999 *Journal of Sound and Vibration* **226**, 425–440. Using SVD to detect damage in structures with different operational conditions.
22. R. P. C. SAMPAIO, N. M. M. MAIA and J. M. M. SILVA 1999 *Journal of Sound and Vibration* **226**, 1029–1042. Damage detection using the frequency-response function curvature method.
23. L. R. RAY and L. TIAN 1999 *Journal of Sound and Vibration* **227**, 987–1002. Damage detection in smart structures through sensitivity enhancing feedback control.
24. K. WORDEN, G. MANSON and N. R. J. FIELLER 2000 *Journal of Sound and Vibration* **229**, 647–667. Damage detection using outlier analysis.
25. G. Y. LUO, D. OSIPIW and M. IRLE 2000 *Journal of Sound and Vibration* **236**, 413–430. Real-time condition monitoring by significant and natural frequencies analysis of vibration signal with wavelet filter and autocorrelation enhancement.
26. R. S. CHANCELLOR, R. M. ALEXANDER and T. S. NOAH 1996 *Journal of Vibration and Acoustics*, **118**, 375–383. Detecting parameter changes using experimental nonlinear dynamics and chaos.
27. J. P. CUSUMANO, D. CHELIDZE and A. CHATTERJEE 1997 *Emerging Technologies for Machinery Health Monitoring and Prognosis* (R. S. Cowan, editor), ASME, Vol. TRIB-7, 45–54. Experimental application of a method for hidden parameter tracking in a slowly changing, chaotic system.

28. D. GRIBKOV and V. GRIBKOVA 2000 *Physical Review E* **61**, 6538–6545. Learning dynamics from nonstationary time series: analysis of electroencephalograms.
29. M. B. KENNEL and A. I. MEES 2000 *Physical Review E* **61**, 2563–2568. Testing for general dynamical stationarity with a symbolic data compression technique.
30. C. CRAIG, R. D. NELSON and J. PENMAN 2000 *Journal of Sound and Vibration* **231**, 1–17. The use of correlation dimension in condition monitoring of systems with clearance.
31. J. D. JIANG, J. CHEN and L. S. QU 1999 *Journal of Sound and Vibration* **233**, 529–541. The application of correlation dimension in gearbox condition monitoring.
32. A. CHATTERJEE, J. P. CUSUMANO and D. CHELIDZE 1998 *Proceedings of the First International Symposium on Impact and Friction of Solids, Structures and Intelligent Machines, Series on Stability, Vibration and Control of Systems* (Series B, Vol. 14), 159–162, Singapore: World Scientific, Tracking Parameter Drift in a Vibro-impact System.
33. E. OTT 1993 *Chaos in Dynamical Systems*. New York: Cambridge University Press.

APPENDIX A: AUTOCORRELATIONS AS GENERIC CHOICES OF FEATURE VECTORS

For the logistic map, it is easy to guess that $\bar{x}$ and $\overline{x^2}$ will be useful for tracking drift. However, in a general problem where a model is not available *a priori*, how should feature vectors be selected? In particular, are autocorrelations expected to work in the context of the present theory (question 3, section 5.5)?

As is well known (see e.g., reference [33]), delay co-ordinate embedding can be used to reconstruct the phase spaces of generic non-linear systems. Given the strain gauge output time series $x(1), x(2), \ldots$ and a positive integer delay d, a delay-reconstructed vector of the form

$$\mathbf{Z}_k = \{x(k), x(k-d), x(k-2d), \ldots, x(k-nd)\}^{\mathrm{T}}$$

can be used as an $(n+1)$-dimensional state vector for the system. The dynamics of the system itself can then be looked upon as a map of the form

$$\mathbf{Z}_{k+1} = F(\mathbf{Z}_k).$$

We consider here a slight variant of the above, using the change of variables $\mathbf{X}_k = \mathbf{Z}_{k+1} - \mathbf{Z}_k$, and assume that there is a map of the form

$$\mathbf{X}_{k+1} = G(\mathbf{X}_k).$$

Since $\mathbf{X}$ is $(n+1)$-dimensional, the dimension of the map is also $n+1$. For the special case where the prediction time step is equal to the delay d, n components of the map are trivial: the pth element of $\mathbf{X}_k$ is mapped exactly to the $(p+1)$th element of $\mathbf{X}_{k+1}$, for $p = 1, 2, \ldots, n$. Thus, only one component of the map is non-trivial.

Since $\mathbf{X}_k = \mathbf{Z}_{k+1} - \mathbf{Z}_k$, we define the variable

$$x_k = z_k - z_{k-d},$$

noting that the expected value of x_k is zero. Now, let us assume that there is a second order polynomial relationship of the form

$$x_{p+1} = \alpha_0 + \sum_{m=0}^{n} \beta_m x_{p-md} + \sum_{r=0}^{n} \sum_{s=r}^{n} \gamma_{rs} x_{p-rd} x_{p-sd}, \tag{A.1}$$

for some constants α_0, β_m and γ_{rs}. If we average both sides, we find that

$$0 = \alpha_0 + \sum_{m=0}^{n+1} b_m \langle x(k)x(k-md)\rangle, \tag{A.2}$$

for some $b_0, b_1, b_2, \ldots, b_{n+1}$ that depend on the constant γ_{rs} (the first order terms average to zero).

For a drifting system with a parameter μ, we would then have

$$0 = \alpha_0(\mu) + \sum_{m=0}^{n+1} b_m(\mu) \langle x(k)x(k-md)\rangle.$$

Thus, this system follows the prescription of the analysis in this paper (except that some of the averaged quantities in equation (A.1) turn out to be identical, and are combined in equation (A.2)).

The foregoing discussion answers the question of section 5.5.3: we can expect autocorrelations, combined with the tracking method presented and analyzed in this paper, to work well in a variety of systems with chaotic dynamics. The success of the method will be limited by the accuracy to which a second order polynomial in several delayed variables can approximate the underlying dynamics of the system. As evidenced, at least, by the experimental study, this accuracy may be acceptable in many cases.