Abstract

The inverse problem of extracting evolution equations from chaotic time series measured from continuous systems is considered. The resulting equations of motion form an autonomous system of nonlinear ordinary differential equations (ODEs). The vector fields are modeled in the manner of implicit Adams integration using a basis set of polynomials that are constructed to be orthonormal on the data. The fitting method uses the Rissanen minimum description length (MDL) criterion to determine the optimal polynomial vector field. It is then demonstrated that one can synchronize the model to an experimentally measured time series. In this case synchronization is used as a nontrivial test for the validity of the models.

Experimentally obtained time series measurements from chaotic systems have been the subject of a great deal of research in the last decade. Much of this work has been motivated by the observation that many systems in the world around us appear to behave in a chaotic manner. If one can learn how to model such systems reliably (i.e. to determine equations of motion), then the door is open to further applications such as short-time prediction and/or control. Local and global methods have been developed by a variety of authors to model the time evolution of the system under investigation [1]. We will focus on dynamical systems that are continuous in time and are naturally modeled by autonomous ordinary differential equations (ODEs), \( \frac{dx}{dt} = F(x) \).

The results presented in this Letter concern a new method for modeling a vector field from time series data. We call this method the Adams-MDL method for training the vector field. The results indicate that the Adams portion of the method is superior to the more common Euler method for modeling time evolution. In addition, the MDL portion of the method is known to be superior to the Akaike or Jaynes methods for choosing an optimal model [2]. We combine these two techniques, Adams and MDL, to produce the Adams-MDL method. We have implemented this procedure on a wide variety of experimentally measured data sets. The systems we have successfully modeled include two different electronic circuits, a chemical reaction, and a mechanical system consisting of a vibrating wire. In all cases we find that the model produced by the method is numerically stable under integration and produces an attractor that is, to the eye, the same as the one obtained by the embedded data.

We believe that this Letter is the first to demonstrate that it is possible to synchronize [3] a fitted model to a time series obtained via experimental observations. We will use synchronization as a nontrivial test for the validity of the fitted models. This application of synchronization is also new. We have used a modification of the Fujisaka and Yamada (FY) driving technique to achieve synchronization. We will show...
that, for our purposes, FY is more flexible than the PC method [4,5].

Experimental measurements are often in the form of a scalar time series, \( s_n = s(n \tau), \) \( n = 1, 2, \ldots, N_p, \) where \( \tau \) is the sampling interval. We will assume that some method has been used to reconstruct phase space vectors, \( y(n) \in \mathbb{R}^d, \) \( n = 1, \ldots, N, \) from the scalar time series [1].

The most common technique found in the literature for determining \( F \) models the time evolution by Euler integration, \( y(n + 1) = y(n) + \tau F(y(n)). \) This method works well only when \( \tau \) is very small. We will model the time evolution by implicit Adams integration,

\[
y(n + 1) = y(n) + \tau \sum_{j=0}^{M} a_j^{(M)} F(y(n + 1 - j)),
\]

which reduces to implicit Euler integration when \( M = 0. \) The \( a_j^{(M)} \) are the implicit Adams coefficients of order \( M \) [6]. It will be shown that, when compared to the Euler method, the Adams method permits one to model data with much larger sampling intervals. Our use of Eq. (1) is the opposite of numerical integration in the following sense: Numerical integration attempts to find an unknown \( y(n) \) from a known \( F, \) we attempt to find an unknown \( F \) from a known \( y(n). \)

Typically the only a priori information known to the modeler is the data, \( y(n). \) In particular the functional form of the vector field is unknown and one is forced to model it as a series expansion

\[
F(z) = \sum_{I=0}^{N_p} p^{(I)} \pi^{(I)}(z),
\]

where the \( \pi^{(I)}(z) \) denote a set of basis functions. The \( p^{(I)} \) are parameters whose values are selected so that Eq. (1) is as accurately as possible. In this Letter the \( \pi^{(I)} \) are constructed to be orthonormal on the attractor mapped out by the phase space vectors, \( y(n). \) \( I \) is an index vector used to identify an element of the basis set and its corresponding coefficient. \( N_p \) is an index vector that indicates the maximum order of the polynomial basis. A detailed discussion of how to construct this basis has previously been presented in Ref. [7].

The minimum description length (MDL) principle is an extension of the maximum likelihood (ML) principle. It has been developed by Rissanen into an objective criterion for choosing the best values of \( M \) and \( N_p \) in Eqs. (1) and (2) [2].

Let \( \Theta = \{a_0^{(M)}, \ldots, a_M^{(M)}, p^{(0)}, \ldots, p^{(N_p)}\} \) denote an \( M_p \) component vector of all parameters in the model and let \( Y \) denote the entire data set \( y(n), n = 1, \ldots, N. \) The ML principle selects a model by maximizing \( \chi^2_{\text{ML}} = -\ln[P(Y|\Theta)], \) where \( P(Y|\Theta) \) is the conditional probability of obtaining \( Y \) given \( \Theta [2,6]. \) The size of the model (i.e. the description length) is determined, in part, by the number of components in \( \Theta, \) which is assumed to be given. Therefore, ML cannot be used to determine the size of the model. On the other hand, the MDL principle selects a model by maximizing \( \chi^2_{\text{MDL}} = -\ln[P(Y, \Theta)], \) where \( P(Y, \Theta) = P(Y|\Theta)P(\Theta) \) is the joint probability of obtaining \( Y \) and \( \Theta, \) and \( P(\Theta) \) is the probability of obtaining \( \Theta. \) By using a joint probability the MDL principle is capable of determining the optimal size for the model. Another interpretation for the MDL principle says that the optimal model (within the class given by Eqs. (1) and (2)) is the one that produces the smallest encoding for the data and the vector field [2].

Minimizing

\[
\chi^2_{\text{MDL}} = \frac{1}{2\tau^2N} \sum_{n=1}^{N} \left| y(n + 1) - y(n) - \tau \sum_{j=0}^{M} a_j^{(M)} \sum_{I=0}^{N_p} p^{(I)} \pi^{(I)}(y(n + 1 - j)) \right|^2
\]

\[
- \tau \sum_{j=0}^{M} a_j^{(M)} \sum_{I=0}^{N_p} p^{(I)} \pi^{(I)}(y(n + 1 - j)) \right|^2
\]

\[
\ln(2\pi e N/M_p) + \ln(||\Theta||^2) \]

is the Adams–MDL criterion for fitting a model [2]. In this equation \( \sigma \) is the standard deviation of the noise in the data. The first term in Eq. (3) is the usual \( \chi^2_{\text{ML}} \) term that arises from the ML principle. The term proportional to \( M_p/N \) is associated with the size of the model. The metric tensor for \( ||\Theta||^2 \) is given by \( \delta^2 \chi^2_{\text{ML}}/\partial \Theta \partial \Theta. \) Typically, as one adds more parameters the first term decreases while the term proportional to \( M_p/N \) increases. This implies a minimum somewhere in the middle, which is chosen as the optimal model.
In principle, we should permit the values of the $a_j$ to vary and minimize $\chi_{\text{MDL}}$ over the $a_j$ as well as the $p$. It is important for our modeling procedure that $\chi_{\text{MDL}}$ is quadratic with respect to the fitting parameters. Therefore, we will assume that the optimal values of the $a_j$ are the ones originally dictated by the Adams integration method. It has been argued that "strong" modeling (where the model is a nonlinear function of the fitting parameters) is superior to models where the parameters appear linearly.

In practice the $\chi_{\text{MDL}}$ term usually dominates Eq. (3). Under these conditions we minimize $\chi_{\text{MDL}}$ over the $p$ for particular values of $Np$ and $M$ by inserting Eq. (2) into Eq. (1) and projecting the result onto $x$. This procedure results in a matrix inversion problem for the $p$, which we solve by a singular value decomposition [5]. Having found the $p$ one calculates the term proportional to $M_N/N$ and adds it to the first term to find $\chi_{\text{MDL}}$.

By finding the optimal value of $Np$ we are determining the optimal order for the polynomial. By finding the optimal value of $M$ we are determining which set of coefficients, $p$, associated with the optimal value of $Np$ and Eq. (1) produce the best vector field. A slightly different approach using the MDL principle has been investigate by Judd and Mees [8].

We present results obtained by fitting data from an electronic circuit and report results from a variety of other systems. The circuit consists of a nonlinear amplifier with linear feedback. It contains a series connection for low-pass filtering and a resonant LC circuit [5,9]. For this circuit we examined two values of amplifier gain, $a = 17.4$ and $a = 18.9$. When $a$ is below some critical value, $a_c$, the dynamics lives on one of two disjoint attractors which are related to each other by an inversion symmetry. At $a_c$ the two attractors merge via a symmetry increasing bifurcation [11]. For our experiments $17.4 < a_c < 18.9$. (For $a = 17.4$ only one attractor is recorded.)

The experimentally measured signals were sampled every 20 $\mu$s and digitized to form a scalar time series, $s(\tau)$. The phase space for the circuit was reconstructed using time delays, $f(n) = [s(n), s(n + T), \ldots, s(n + (d - 1)T)]$. The embedding time delay, $T = 10$, and dimension, $d = 3$, were obtained from mutual information and false near neighbors methods [10].

We tested the Adams–MDL technique by modeling $N = 3000$ vectors, $f$, obtained at $a = 17.4$. The results are shown in Fig. 1. A clear minimum can be seen.
at $N_p = 4$, $M = 3$ corresponding to the fourth order polynomial found when using a third order Adams integrator. Repeating the process with data taken at $\alpha = 18.9$ resulted in an optimal model when $N_p = 5$ and $M = 3$ [5]. This model is more complicated than the $\alpha = 17.4$ model, which should not be surprising since the attractor is more complicated [5]. We have also examined experimentally obtained data from the Belousov–Zhabotinski chemical reaction, a hysteretic circuit from NRL, and a mechanical systems (a vibrating wire) [5]. In all cases we were able to find low order polynomial models for the dynamics.

As a trivial test of the optimal $F$ we chose initial conditions on the attractors and integrated the appropriate ODEs forward in time. The resulting orbits traced out attractors that were the same, to the eye, as the ones mapped out by the observed $y(n)$. A non-trivial test involves synchronizing the optimal $F$ to a time series. Synchronization requires using an output from one dynamical system as a driving input to another system. The results of PC imply that it should be possible to synchronize models to an experimentally measured time series, and that synchronization will not occur if $F$ is not close to the true equations of motion [5].

The modified FY driving method is

$$\frac{dx}{dt} = F(x) - E \cdot (x - y), \quad (4)$$

where the coupling matrix has only one nonzero element, $E_{ij} = \epsilon$ when $y_j$ is used as the driving component. It can be shown that this method reduces to PC driving in the $\epsilon \to \infty$ limit [5].

A variable order, variable time step integrator was employed for our synchronization tests [12]. Linear interpolation was used to obtain the drive variable, $y$, at values of time not contained in the data set. The numerical experiments required that we generate two initial condition $y$ from $y$ due to the instability of chaotic orbits. The $z$ and $w$ orbits were generated using a data set, the modified FY driving method is

$$\frac{d\mathbf{x}}{dt} = F(\mathbf{x}) - E \cdot (\mathbf{x} - \mathbf{y}), \quad (6)$$

where the coupling matrix has only one nonzero element, $E_{ij} = \epsilon$ when $y_j$ is used as the driving component. It can be shown that this method reduces to PC driving in the $\epsilon \to \infty$ limit [5].

A variable order, variable time step integrator was employed for our synchronization tests [12]. Linear interpolation was used to obtain the drive variable, $y$, at values of time not contained in the data set. The numerical experiments required that we generate two different orbits. The first orbit, $z$, results when the initial condition $y(1)$ is numerically integrated forward in time using Eq. (4) with $\epsilon = 0$ (i.e., no driving). This orbit will begin near $y(1)$ but will diverge away from $y$ due to the instability of chaotic orbits. The other orbit, $w$, results when $E_{ij} = \epsilon$ is integrated using $\epsilon \neq 0$ (i.e., chaotic driving).

The $z$ and $w$ orbits were generated using a data set, and the optimal $F$, from $\alpha = 17.4$. We, arbitrarily, chose to drive with $y_3$ by setting $E_{33} = \epsilon = 20$ in Eq. (4). The results are shown in Fig. 2. The dashed line is $D_z(n) = |y(n) - z(n)|$ while the solid line is $D_w(n) = |y(n) - w(n)|$. The oscillations in $D_z$ occur because the two orbits, $y$ and $z$, are evolving independently on the same attractor. In contrast, $D_w$ starts with a value near 0 but soon decreases to $\sim 10^{-2}$. This implies that $w \sim y$ and the driven orbit, $w$, has become synchronized to the driving orbit, $y$.

We also examined the effect of driving different components of $F$ and found that synchronization occurred when driving with $y_2$ but not with $y_1$. In contrast PC driving resulted in synchronization only when $y_3$ was used. Finally, the PC driving method can only test $d - 1$ components of $F$ since one component is replaced by direct substitution of the variable.

Finally we determined the behavior of the modeling method as a function of the sampling rate, $\tau$ and the order of the Adams integrator, $M$. The numerical experiments were performed on the Lorenz system [13] using all three components as our time series, $y(n) = [x(n\tau), y(n\tau), z(n\tau)]$. We have fixed the total observation time, $T = N\tau$. (Hence, we vary $N$ as we vary $\tau$.)

Fig. 3 reports the relative prediction error

$$\chi^2 = \left( \frac{1}{N} \sum_{n=1}^{N} \frac{|\Delta y(n + 1; n)|^2}{|y(n + 1) - y(n)|^2} \right)^{1/2},$$

where $\Delta y(n + 1; n)$ is the difference between the left and right hand sides of Eq. (1). We see that $\chi^2$ often decreases by several orders of magnitude when $M > 0$. This indicates that one can obtain a more accurate ODE model by using Eq. (1) instead of the Euler method. The benefit of this is that $\tau$ can be greatly increased before the relative prediction error from the Adams method is comparable to that of the Euler method. The MDL criterion selected $N_p = 2$, $M = 5$ as the best model when $\tau = 0.004336$ [5]. Finally, the coefficients of $F$ obtained using data with $\tau = 0.01153$ are shown in Table 1. Except for the first row (which should all be zero) the $M = 9$ Adams-
Fig. 2. Results from driving $F$ with the experimentally measured data, $s(n)$ from the electronic circuit. The dashed line is $D_z(n) = |y(n) - z(n)|$ while the solid line is $D_w(n) = |y(n) - w(n)|$. Synchronization between $y$ and $w$ is clearly demonstrated.

Fig. 3. The relative prediction error for the Lorenz system ($N_p = 2$). Notice that the Adams–MDL method typically produces much smaller errors than the Euler method.
Table 1
The polynomial coefficients obtained by modeling the Lorenz system from data at \( \tau = 0.01153 \)

<table>
<thead>
<tr>
<th>( M = 0 )</th>
<th>( F_1 )</th>
<th>( F_2 )</th>
<th>( F_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.136</td>
<td>-4.144</td>
<td>33.92</td>
</tr>
<tr>
<td>1</td>
<td>-22.277</td>
<td>48.626</td>
<td>-0.113</td>
</tr>
<tr>
<td>2</td>
<td>17.942</td>
<td>-5.152</td>
<td>0.248</td>
</tr>
<tr>
<td>3</td>
<td>-0.019</td>
<td>0.419</td>
<td>-6.548</td>
</tr>
<tr>
<td>4</td>
<td>-0.004</td>
<td>0.073</td>
<td>-0.003</td>
</tr>
<tr>
<td>5</td>
<td>0.003</td>
<td>-0.061</td>
<td>1.237</td>
</tr>
<tr>
<td>6</td>
<td>0.104</td>
<td>-1.090</td>
<td>-0.002</td>
</tr>
<tr>
<td>7</td>
<td>0.000</td>
<td>0.005</td>
<td>-0.183</td>
</tr>
<tr>
<td>8</td>
<td>-0.009</td>
<td>0.144</td>
<td>-0.004</td>
</tr>
<tr>
<td>9</td>
<td>0.006</td>
<td>-0.011</td>
<td>0.039</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( M = 9 )</th>
<th>( F_1 )</th>
<th>( F_2 )</th>
<th>( F_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000</td>
<td>-0.022</td>
<td>-0.009</td>
</tr>
<tr>
<td>1</td>
<td>-16.000</td>
<td>45.920</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>16.000</td>
<td>-1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>0.000</td>
<td>-4.000</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>6</td>
<td>0.000</td>
<td>-1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The MDL method finds the exact values for the coefficients while the Euler method, \( M = 0 \), fails.

Our goal was to develop a method of generating, from a time series, an ODE that modeled the dynamics of the system that produced the time series. We have demonstrated that the Adams–MDL method allows one to choose the optimal model (within the polynomial class of models). The optimal model is usually not the one obtained by the implicit Euler method used by most researchers. We have demonstrated that the Adams–MDL method is superior to the Euler method because it allows one to use data with much larger sampling intervals.

We have also demonstrated that one can synchronize our models to an experimentally measured time series. Synchronization indicates that \( F \) is structurally close to the true vector field in regions of phase space near the attractor. Thus, synchronization is a nontrivial test of the validity of our models. We find that all our models pass this test. Possible applications for modeling and synchronization in the realm of fault detection and nondestructive testing can be found in our longer papers [5].

We would like to thank Drs. H.D.I. Abarbanel, Jon Wright, and Lev Tsimring for many useful discussions and encouragement. This research was supported in part by the University of California, Universitywide Energy Research Group and the College of William and Mary.

References

R. Brown, N.F. Rulkov and N.B. Tufillaro, Synchronization of chaotic systems: The effect of additive noise and drift in the dynamics of the driving, INLS preprint (June 1994);
see also K. Judd and A. Mees, On selecting models for nonlinear time series, preprint, University of Western Australia (1994).
R. Brown, Orthonormal polynomials as prediction functions in arbitrary phase space dimensions, INLS preprint (October 1993).