distinguish between "long-period" motions and nonperiodic motions. The trajectories of a nonperiodic attractor do not conform to any simple geometrical form in the state space.

In most experimental situations, the "independent" coordinates required to construct a state space are not well defined, and establishing the "proper" state space is not an easy task. However, as discussed in the next section, certain systematic approaches are available for constructing such spaces.

7.4 PSEUDO–STATE SPACE

In the preceding section, we assumed that one can determine either theoretically or experimentally all of the variables $\mathbf{x}(t)$ that describe the behavior of the dynamical system of interest. However, in experiments, typically one observes only one or at best a few of the dynamical variables that govern the behavior of the desired system. Let us assume that we can only measure one component or, more generally, one scalar function $s_n = s(t) = s(t_0 + n\tau_s)$ of the state vector; that is,

$$s_n = s(t) = S[\mathbf{x}(t)] \tag{7.4.1}$$

where t_0 is some initial time and τ_s is the sampling time of the instrument used in the experiment. The observable scalar may be a voltage from a strain gauge or an accelerometer attached to a structure, a temperature or pressure in a fluid, a voltage or current in a nonlinear circuit, or a voltage from an optical sensor. Then, the question arises whether one can use this scalar or univariate observation to construct a multivariate state space that describes the attractor, especially in the absence of any a priori knowledge of the dimension of the state space required to describe the motion of the system. In the context of topology, this is an example of an **embedding problem**. The issue is to find a one-to-one mapping between points on the (original) attractor in the full system state space and the attractor in the reconstructed state space. This mapping is called an **embedding**. Here, the mapping should also preserve information about the derivatives of the flow. In 1936, Whitney showed that a smooth C^2 m-dimensional manifold may

be embedded in $\mathcal{R}^{2m+1}$ space. This result about the existence of an embedding is quite independent of the dynamics of the system.

Work done over the last decade in the area of dynamical systems theory suggests that a measurement of a scalar signal would suffice to carry out an embedding. From the works of Packard, Crutchfield, Farmer, and Shaw (1980) and Takens (1981), which were motivated by experiments in fluid mechanics, and the many publications that have followed them, it is now known that a state space can be reconstructed from a scalar time signal such that the dynamics in the reconstructed state space is equivalent to the original dynamics. Considering a deterministic three–dimensional autonomous system, Packard et al. (1980) demonstrated the equivalence through numerical simulations. Takens (1981) considered a deterministic finite–dimensional autonomous system and proved the equivalence with mathematical rigor, assuming that an infinite amount of noise–free data was available in the analysis. As a consequence of the equivalence, an attractor in the reconstructed state space has the same invariants, such as Lyapunov exponents and dimension, as the original attractor. Hence, measures such as the Lyapunov exponents and dimension can be obtained for the motion in the reconstructed state space.

When measures, such as Lyapunov exponents, are preserved, the differential structure of the original attractor is also said to be preserved in the reconstruction (Sauer, Yorke, and Casdagli, 1991). For this reason, the associated embedding is called a **differentiable embedding**. When there is only a one–to–one correspondence between the vectors in the reconstructed state space and the vectors in the full state space, the associated embedding is called a **topological embedding**.

Different methods have been proposed for construction of the "independent" coordinates for the state space. According to the method proposed by Packard et al. (1980), the time derivatives of a signal can be used along with it to construct a state space. (An analog differentiator [e.g., Horowitz and Hill, 1980] or a digital computer may be used to obtain the time derivatives.) Thus, using the time series, one can approximate the derivatives of $s(t)$ by using finite differences and hence generate the variables needed to describe the behavior of the system.

For example,

$$\frac{ds}{dt}(t_0 + n\tau_s) \approx \frac{s[t_0 + n\tau_s] - s[t_0 + (n-1)\tau_s]}{\tau_s} \tag{7.4.2}$$

$$\frac{d^2s}{dt^2}(t_0 + n\tau_s) \approx \frac{s[t_0 + (n+1)\tau_s] - 2s(t_0 + n\tau_s) + s[t_0 + (n-1)\tau_s]}{\tau_s^2} \tag{7.4.3}$$

Similarly, one can write down approximations to the higher derivatives. However, with finite τ_s and the presence of noise contamination, (7.4.2) is a crude approximation of the first derivative. Moreover, (7.4.3) represents an even poorer approximation of the second derivative. The quality of the approximation deteriorates even further for the higher derivatives.

Examining the finite-difference formulas for the derivatives, we note that, at each step of the differentiation, we are adding the new information already contained in the measurement at other time steps lagged or advanced by multiples of the sampling time τ_s. This observation led Packard et al. (1980), Ruelle (1989a), and Takens (1981) to conclude that one does not need the derivatives to form a coordinate system that describes the structure of orbits in phase space. Instead, one can use directly the time advanced variables $s(t + n\tau)$, where $n = 1, 2, \ldots, d$ and $\tau = k\tau_s$ is an appropriately chosen time delay, as discussed in Section 7.4.2. and define the so-called **delay-coordinate vectors**

$$\mathbf{y}_n = \{s(t_0 + n\tau_s) \;\; s(t_0 + n\tau_s + k\tau_s) \;\; \ldots \;\; s[t_0 + n\tau_s + k(d-1)\tau_s]\}^T$$

or

$$\mathbf{y}_n = [s_n \;\; s_{n+k} \;\; s_{n+2k} \;\; \ldots \;\; s_{n+kd-k}]^T \tag{7.4.4}$$

The space constructed by using the vectors $\mathbf{y}_n$ is called the **reconstructed space**. According to a theory of Takens (1981) and Mané (1981), the geometric structure of the dynamics of the system from which the s_n were measured can be observed in the reconstructed d-dimensional Euclidean space if $d \geq 2d_a + 1$, where d_a is the dimension of the attractor of interest. Recently, Sauer et al. (1991) extended the

work of Takens and showed that if d_a is the box-counting dimension, then choosing $d > 2d_a$ suffices for an embedding. The parameter τ is called **time delay**, the integer d is called the **embedding dimension**, the constructed coordinates are called **delayed coordinates**, and this method of constructing coordinates is called the **method of delays**.

Given the equations $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ describing the (deterministic) system dynamics and given the state $\mathbf{x}(t)$ of the system at $t = t_0$, one can, in principle, integrate these equations forward in time by an amount $k\tau_s$ and obtain $\mathbf{x}(t_0 + k\tau_s)$. In other words, $\mathbf{x}(t_0 + k\tau_s)$ is a unique function of $\mathbf{x}(t_0)$. Hence,

$$\mathbf{x}(t_0 + k\tau_s) = \mathbf{L}[\mathbf{x}(t_0)] \tag{7.4.5}$$

which, upon substitution into (7.4.1), yields

$$s(t_0 + k\tau_s) = S[\mathbf{x}(t_0 + k\tau_s)] = S\{\mathbf{L}[\mathbf{x}(t_0)]\} \tag{7.4.6}$$

Therefore, the reconstructed vector $\mathbf{y}$ is related to the state vector $\mathbf{x}$ by a smooth nonlinear relationship of the form

$$\mathbf{y} = \mathbf{H}(\mathbf{x}) \tag{7.4.7}$$

Consequently, the vector $\mathbf{y}$ can serve as a coordinate basis for the system dynamics because any smooth coordinate transformation can serve the purpose.

Abarbanel, Brown, Sidorowich, and Tsimring (1993) integrated the Lorenz equations (5.8.25)-(5.8.27) for $\rho = 45.92, \beta = 4.0$, and $\sigma = 16.0$ using a fourth-order Runge-Kutta scheme with the time step $\tau_s = 0.01$. In Figure 7.4.1a, we show a three-dimensional plot of the obtained chaotic attractor. Using a time delay $\tau = 20\tau_s = 0.2$, Abarbanel et al. (1993) reconstructed the orbit in phase space by using the time series of the variable x and an embedding dimension d of three; that is,

$$\mathbf{y} = [x(t) \;\; x(t + \tau) \;\; x(t + 2\tau)]^T \tag{7.4.8}$$

The reconstructed attractor in the pseudo-state space is shown in Figure 7.4.1b. Comparing Figure 7.4.1b with Figure 7.4.1a, we see that, although distorted as expected because of the nonlinear transformation, the reconstructed geometric object is similar in appearance to the original geometric object obtained by using the state variables.

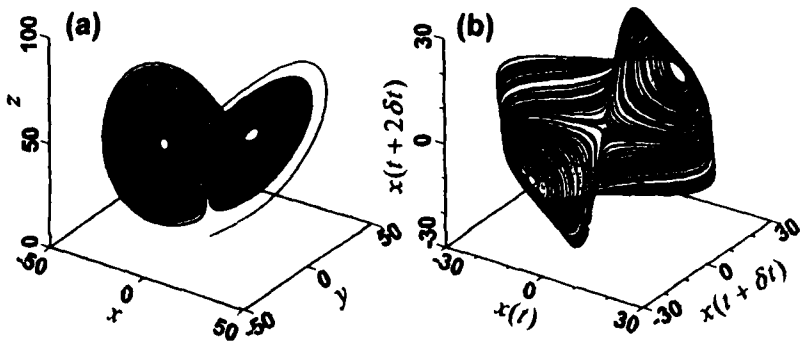

Figure 7.4.1: The Lorenz attractor in three-dimensional state space: (a) constructed from $x(t), y(t)$, and $z(t)$; and (b) constructed from $x(t), x(t+\tau)$, and $x(t+2\tau)$, where $\tau = 0.2$. Reprinted with permission from Abarbanel, Brown, Sidorowich, and Tsimring (1993).

In principle, if d is large enough and if the dimension of the attractor of interest is finite, then one can capture the dynamics by using the delay vectors $\mathbf{y}(t)$. The key idea in the reconstruction is that orbits generated by autonomous systems of equations do not intersect in the full state space. In fact, we have repeatedly noted that the intersections found in some of the plots in Chapters 3–5 are the result of projections onto a subset of the state space. Consequently, the dimension d of the pseudo-state space must be large enough so that the reconstructed orbit does not overlap with itself. When this happens, d is called an **embedding dimension.** In the next section, we discuss methods of choosing the embedding dimension.

In theory, the time delay can be arbitrary if one has an infinite amount of noise-free data. When this is so, it can be said that successive measurements contain new information for any nonzero time interval that separates them. But, in practice, one is limited by factors such as finite data, finite precision, and noise. Therefore, many considerations have to be taken into account in choosing the time delay. In principle, we require a time delay τ that will produce "independent" delayed coordinates. If τ is too small, the trajectories in the pseudo-state space of $\mathbf{y}(t)$ and $\mathbf{y}(t+\tau)$ stack up on the diagonal because the delayed coordinates are highly correlated. On the other hand, if τ is

too large, an artificial decorrelation is introduced and the delayed coordinates become uncorrelated. When τ is chosen to be equal to one of the basic periods of the system, one constructs a Poincaré section in the space of the delayed coordinates. So, if τ is close to any of the basic periods of the system, the corresponding periodic component will not be well represented in the space of the delayed coordinates. In Section 7.4.2, we discuss methods of choosing τ.

Assuming that a time delay has been chosen, one can proceed with the construction of the space of delayed coordinates. The issue of state space reconstruction in the presence of noise is addressed in detail by Casdagli, Eubank, Farmer, and Gibson (1991) and Kostelich and Schreiber (1993), among others.

7.4.1 Choosing the Embedding Dimension

The objective of the reconstruction is to find a Euclidean space $\mathcal{R}^d$ that is large enough so that the set of points of dimension d_a, which describe the attractor, can be unfolded without ambiguity. In other words, if two points of the set lie close to each other in some dimension d, they do so because of the property of the attractor rather than because of the small value of d in which the attractor is being examined. This means that the value of d should be large enough so that the asymptotic state of the motion can be captured (embedded) in this d-dimensional space. Based on the general existence theorem for embeddings in Euclidean space, given by Whitney (1936), that a smooth C^2 m-dimensional manifold may be embedded in $\mathcal{R}^{2m+1}$, in theory it is sufficient that $d \geq 2d_a + 1$, where d_a is the dimension associated with the observed motion. Sauer, Yorke, and Casdagli (1991) show that $d > 2d_a$ is sufficient when a delay coordinate embedding is used. For some cases, it has been found that $d \geq d_a$ is sufficient (e.g., Eckmann and Ruelle, 1985; Abarbanel, Brown, and Kadtke, 1990; Buzug, Reimers, and Pfister, 1990). In practice, d_a is not known a priori in most situations. Here, we discuss three approaches for estimating d.

Saturation of System Invariants

The basic idea underlying the saturation of the attractor of system in-

variants is that, if the attractor is unfolded by using a large enough embedding dimension d, then the invariant properties of the attractor, such as the Lyapunov exponents and dimension, calculated from the reconstructed trajectory do not change if one increases d. In other words, there is a dimension d beyond which all invariant properties of the desired attractor saturate.

To carry out this approach, we pick a value for d and then calculate one of the invariants of the attractor. Then, we increase d by one, recalculate the invariant, and compare the result with the preceding value. If the difference is within a specified tolerance, we conclude that d is the appropriate value. Otherwise, we repeat the process until there is relatively little change in the calculated invariant. Methods for calculating Lyapunov exponents and attractor dimensions are discussed in Sections 7.8 and 7.9. In this section we explain this approach by using the moments of the number density.

The number of points on the attractor within a sphere of radius r from the point $\mathbf{x}_m$ in the phase space is given by

$$P_m(r) = \frac{1}{N_0} \sum_{k=1}^{N_0} H(r - \mid \mathbf{x}_k - \mathbf{x}_m \mid) \tag{7.4.9}$$

where N_0 is the total number of sampled points and H is the Heaviside step function defined by

$$H(u) = \begin{cases} 0 & \text{if } u < 0 \\ 1 & \text{if } u \geq 0 \end{cases} \tag{7.4.10}$$

The average of powers of $P_m(r)$ over all points $\mathbf{x}_m$ yields the correlation function

$$C_q(r) = \frac{1}{N_0} \sum_{m=1}^{N_0} [P_m(r)]^{q-1} \tag{7.4.11}$$

where q is an integer. Measures, such as (7.4.11), are quite well known in statistics (e.g., Renyi, 1970). Furthermore, since these measures are independent of initial conditions, they can be used to characterize attractors (Abarbanel, Brown, Sidorowich, and Tsimring, 1993).

To determine the embedding dimension d, we calculate (Grassberger

and Procaccia, 1983a,b)

$$d_a = \frac{\log [C_q(r)]}{\log(r)} \tag{7.4.12}$$

as a function of d and determine when d_a becomes independent of d. (As discussed in Section 7.9, d_a is defined in the limit as r tends to zero. However, in practice, the finite number of data points and noise place a lower bound on the values of r that can be used to calculate d_a.) Therefore, one often plots $\log[C_q(r)]$ versus $\log(r)$ and estimates d_a from the slope. In Figure 7.4.2, we show variation of $\log[C_2(r)]$ with $\log(r)$ for data generated from the Lorenz equations as a function of d. Clearly, the slope of the plot and hence the value of d_a are independent

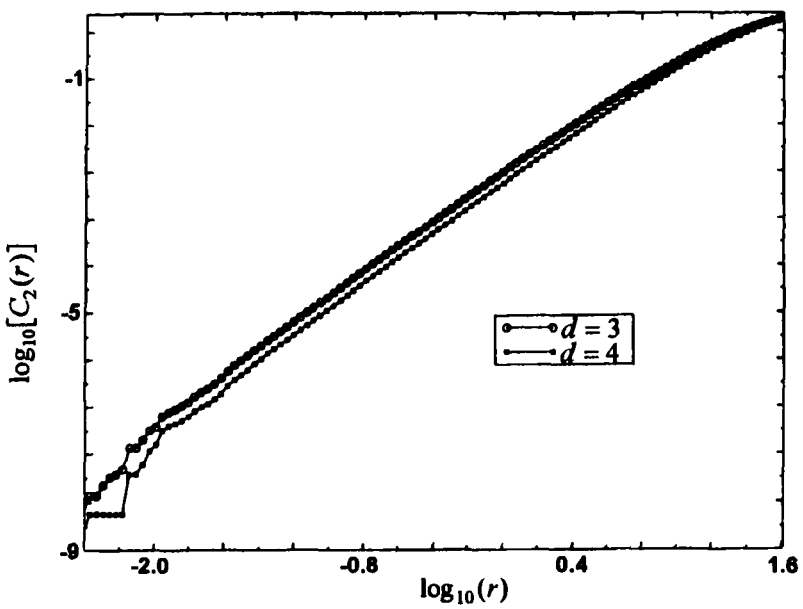

Figure 7.4.2: Variation of $\log[C_2(r)]$ with $\log(r)$ for $z_n = z(t) = z(t_0 + n\tau_s)$ data for the Lorenz attractor using the embedding dimension values $d = 3$ and 4. Reprinted with permission from Abarbanel (1995).

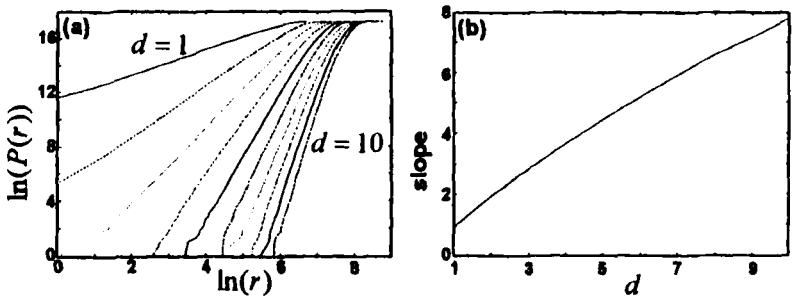

Figure 7.4.3: Illustration of the lack of saturation of the dimension for the dynamics associated with random noise.

of the value of d for $d \geq 3$. Hence, the embedding dimension according to the saturation criterion is 3, in contrast with the sufficient integer dimension of 5. According to the discussion in Section 7.9, the saturated slope is the attractor dimension d_a. In this case, $d_a = 2.06$, and hence the sufficient integer dimension is 5 because $d \geq 2d_a + 1$ according to the embedding theorem.

We note that for random noise, the correlation function continues to increase as the embedding dimension is increased; in other words, the dimension does not saturate. In Figure 7.4.3, we show variation of $\ln[P(r)]$ with $\ln(r)$ for random noise. (The definition for $P(r)$ is provided in Section 7.9.) Clearly, the slope of the plot does not saturate, and in fact it increases with increasing d.

Singular Value Analysis

Broomhead and King (1986) and Broomhead and Jones (1989) developed phase-space reconstructions by using the singular system approach, which is based on the Karhunen–Loeve theorem (Loeve, 1977). This technique was developed to deal with noise and errors that arise from finitely sampling data, and hence, it is ideally suited for experimental data. The key idea underlying this approach is that the mean-square distance between points on the reconstructed attractor should be maximized. The approach used by Landa and Rosenblum (1991), which is based on **Neymark algorithm**, is closely related to the sin-

gular system approach.

Using the sampled data s_j, choosing a large enough space dimension n, and assuming noise-contaminated data, we construct the **trajectory matrix** Y as

$$Y = \frac{1}{\sqrt{N}} \left[\mathbf{y}_1^T - \mathbf{y}_{av}^T \;\; \mathbf{y}_2^T - \mathbf{y}_{av}^T \cdots \;\; \mathbf{y}_N^T - \mathbf{y}_{av}^T \right]^T \tag{7.4.13}$$

where $N = N_0 - (n-1)$, N_0 is the total number of sampled data points,

$$\mathbf{y}_i = [s_i \;\; s_{i+1} \;\; s_{i+2} \cdots \;\; s_{i+n-1}]^T \tag{7.4.14}$$

and

$$\mathbf{y}_{av} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \tag{7.4.15}$$

Next, we calculate the dimension of the subspace of $\mathcal{R}^n$ that contains the reconstructed attractor. To accomplish this, we need to determine the number of linearly independent vectors that can be generated from the $\mathbf{y}_i - \mathbf{y}_{av}$ in the space $\mathcal{R}^n$. We note that

$$\sqrt{N} \mathbf{e}_i^T Y = \mathbf{y}_i^T - \mathbf{y}_{av}^T \tag{7.4.16}$$

where the $\mathbf{e}_i$ are the standard basis vectors in $\mathcal{R}^N$; they correspond to columns of the $N \times N$ identity matrix. Moreover, we note that any vector $\mathbf{w}$ in $\mathcal{R}^N$ can be expressed in terms of the $\mathbf{e}_i$ as

$$\mathbf{w}^T = \sqrt{N} \sum_{i=1}^{N} w_i \mathbf{e}_i^T \tag{7.4.17}$$

Hence,

$$\mathbf{w}^T Y = \sum_{i=1}^{N} w_i \left(\mathbf{y}_i^T - \mathbf{y}_{av}^T \right) \tag{7.4.18}$$

In other words, vectors in $\mathcal{R}^N$ give rise to linear combinations of the $\mathbf{y}_i^T - \mathbf{y}_{av}^T$.

We consider a set of vectors $\mathbf{b}_i \in \mathcal{R}^N$ that give rise to linearly independent vectors $\mathbf{c}_i$ for $i = 1, 2, \ldots, n$ in $\mathcal{R}^n$, which we assume, without loss of generality, to be orthonormalized. Thus,

$$\mathbf{b}_i^T Y = \sigma_i \mathbf{c}_i^T \tag{7.4.19}$$

where the σ_i are real constants that are chosen to fix the orthonormalization of the $\mathbf{c}_i$. Taking the transpose of (7.4.19) yields

$$Y^T\mathbf{b}_j = \sigma_j\mathbf{c}_j \tag{7.4.20}$$

Combining (7.4.19) and (7.4.20), we have

$$\mathbf{b}_i^T YY^T\mathbf{b}_j = \sigma_i\sigma_j\mathbf{c}_i^T\mathbf{c}_j = \sigma_i\sigma_j\delta_{ij} \tag{7.4.21}$$

on account of the orthonormality of the $\mathbf{c}_i$. Equation (7.4.21) can be solved by determining the eigenvectors of the $N \times N$ real, symmetric matrix YY^T; that is,

$$YY^T\mathbf{b}_i = \sigma_i^2\mathbf{b}_i \tag{7.4.22}$$

The σ_i^2 are the singular values of the so-called **structure matrix** YY^T; they are non-negative definite. It follows from (7.4.19) that, at most, n of the σ_i are nonzero. Therefore, the rank of the structure matrix is less than or equal to n. Next, we show that the associated eigenvalues are also the eigenvalues of the so-called covariance matrix Σ.

Multiplying (7.4.20) from the left with Y, we obtain

$$YY^T\mathbf{b}_j = \sigma_j Y\mathbf{c}_j \tag{7.4.23}$$

which, upon using (7.4.22), becomes

$$Y\mathbf{c}_j = \sigma_j\mathbf{b}_j \tag{7.4.24}$$

Multiplying (7.4.24) from the left with Y^T yields

$$Y^TY\mathbf{c}_j = \sigma_j Y^T\mathbf{b}_j = \sigma_j^2\mathbf{c}_j \tag{7.4.25}$$

on account of (7.4.20). The matrix $\Sigma = Y^TY$ is called the **covariance matrix**. It is an $n \times n$ real symmetric non-negative definite matrix. Hence, its eigenvalues σ_j^2 are non-negative definite. Multiplying (7.4.25) from the left with bfc_j^T and making use of the orthonormality of bfc_j leads to

$$(Y\mathbf{c}_j)^T(Y\mathbf{c}_j) = {\sigma_j}^2$$

In a perfect world, the number of nonzero eigenvalues σ_j^2 of Y^TY give the dimensionality of the subspace containing the embedded

attractor and hence the embedding dimension d. Moreover, the vectors $\mathbf{c}_i$ corresponding to these nonzero eigenvalues span the embedding subspace. (This approach of deriving a set of basis vectors based on the covariance matrix is the essence of the Karhunen–Loeve method used in the areas of signal processing and pattern recoginition.) However, as explained by Broomhead and King (1986), if the data are noisy and the variance of the noise is σ_η^2, then every singular value σ_i will be shifted by σ_η. This shifting means that the noise causes all the singular values of the covariance matrix to be nonzero. Consequently, noise will dominate any eigenvector $\mathbf{c}_j$ whose singular value σ_j is comparable to σ_η, and hence such vectors must be discarded. This results in the reduction of the dimension of the embedding subspace from n to $d = n - m$, where m is the number of singular values comparable to σ_η. Broomhead and King (1986) note that m increases as $\tau_w = n\tau_s$ increases, where τ_s is the sampling time. To limit the size of m, they suggest selecting $\tau_w = 2\pi/\omega^*$, where ω^* is the **band–limiting frequency**. The time τ_w corresponds to the first zero crossing of the second derivative of the autocorrelation function

$$C(\tau) = < s(t)s(t+\tau) > \tag{7.4.26}$$

To illustrate their approach, Broomhead and King (1986) integrated the Lorenz equations (5.8.25)–(5.8.27) for a particular set of parameters by using a Runge–Kutta scheme. The broadband character in the power spectrum of Figure 7.4.4 is indicative of the chaotic nature of $x(t)$.

Using $\tau_s = 0.009$, $\tau_w = 0.063$, and $n = 7$, they calculated the singular values of the covariance matrix $\Sigma = Y^T Y$ and their corresponding eigenvectors. The spectrum of the eigenvalues and the first three singular eigenvectors are shown in Figure 7.4.5. The spectrum of the eigenvalues has two distinct parts: one part, which can be associated with the noise floor, and a second part, which is associated with the deterministic component of the data. The noise floor can be distinguished by its magnitude and flatness; the magnitude corresponds to round–off errors in the computations. Thus, from the data in Figure 7.4.5, we conclude that the dynamics will be confined to a four–dimensional subspace of the embedding seven–dimensional space.

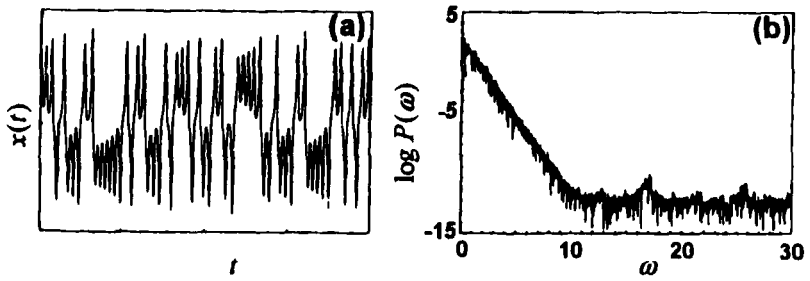

Figure 7.4.4: A time trace and associated power spectrum of the state variable $x(t)$ obtained by integrating the Lorenz equations (5.8.25)–(5.8.27) for $\sigma = 10$, $\beta = \frac{8}{3}$, and $\rho = 28$ by using a fourth–order Runge–Kutta scheme with a time step of 0.009 units. Reprinted with permission from Broomhead and King (1986).

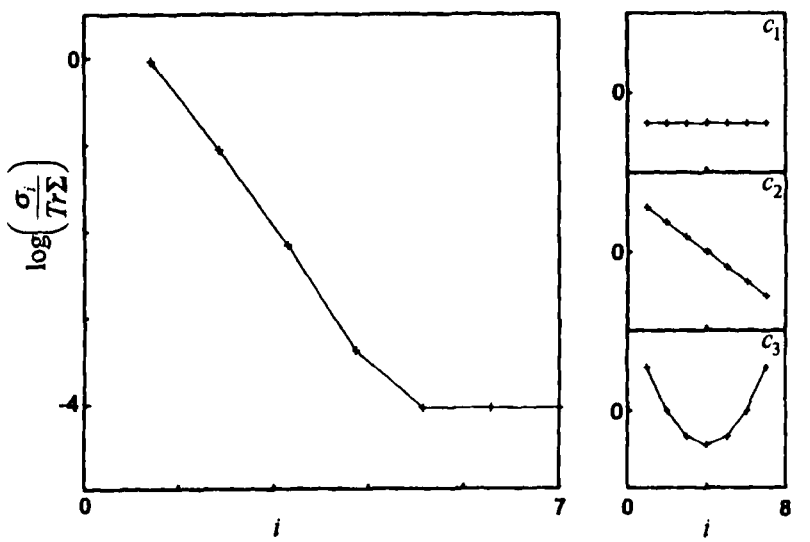

Figure 7.4.5: The first three singular values and their corresponding eigenvectors calculated for the covariance matrix constructed from the data in Figure 7.4.4. Reprinted with permission from Broomhead and King (1986).

Broomhead and Jones (1989) carried out the above analysis locally by considering the covariance matrix over a neighborhood of a point specified by the reconstructed state vector $\mathbf{y}_n$. We let N_B be the number of nearest neighbors $\mathbf{y}_n^{(r)}$ of $\mathbf{y}_n$. Then, we define the covariance matrix COV_n as

$$\mathrm{COV}_n = \frac{1}{N_B} \sum_{r=1}^{N_B} \left[\mathbf{y}_n^{(r)} - \mathbf{y}_a\right] \left[\mathbf{y}_n^{(r)} - \mathbf{y}_a\right]^T \tag{7.4.27}$$

where

$$\mathbf{y}_a = \frac{1}{N_B} \sum_{r=1}^{N_B} \mathbf{y}_n^{(r)} \tag{7.4.28}$$

The matrix COV_n will have d_L eigenvalues arising from the variation of the slightly contaminated real signal about its mean and $n - d_L$ eigenvalues due to the noise. The eigenvectors associated with the d_L eigenvalues can be used to construct a **local pseudo–state space**. The **local embedding dimension** d_L is less than or equal to the **global embedding dimension** d.

False Nearest Neighbors

The basic idea underlying the false nearest neighbors approach is to find an embedding space of dimension d in which all false crossings of the orbit with itself that arise because of the projection onto a low-dimension space are eliminated. When d is not large enough, points that are far apart in the full or original state space are brought close together in the reconstruction space, resulting in **false nearest neighbors**. To determine these neighbors, one needs to examine if two states are neighbors because of the dynamics or because of the projection onto a low-dimension space. Thus, by determining neighbors in increasing embedding dimensions, one can eliminate false neighbors and hence establish the embedding dimension. Kennel, Brown, and Abarbanel (1992) proposed using a kd-tree search routine for finding nearest neighbors among N points. This routine takes $N \log N$ operations.

In dimension d and time advance $k\tau_s$, we assume that the reconstructed vector $\mathbf{y}_n$ given by (7.4.4) has the nearest neighbor specified

by the vector

$$\hat{\mathbf{y}}_n = [\hat{s}_n \;\; \hat{s}_{n+k} \cdots \;\; \hat{s}_{n+kd-k}]^T \tag{7.4.29}$$

The Euclidean distance $R_n(d)$ between $\mathbf{y}_n$ and $\hat{\mathbf{y}}_n$ is given by

$$R_n^2(d) = \sum_{i=1}^{d} (\hat{s}_{n+ik-k} - s_{n+ik-k})^2 \tag{7.4.30}$$

This distance is assumed to be small. In dimension $d+1$, the distance between these two points becomes

$$R_n^2(d+1) = \sum_{i=1}^{d+1} (\hat{s}_{n+ik-k} - s_{n+ik-k})^2$$

or

$$R_n^2(d+1) = R_n^2(d) + (\hat{s}_{n+kd} - s_{n+kd})^2 \tag{7.4.31}$$

If $R_n(d+1)$ is large compared with $R_n(d)$, we can presume that this is so because $\mathbf{y}_n$ and $\hat{\mathbf{y}}_n$ are false neighbors in dimension d. In the calculations, we select a threshold R_T and decide whether the two neighbors $\mathbf{y}_n$ and $\hat{\mathbf{y}}_n$ are false nearest neighbors or not, depending on whether the following inequality is satisfied or not:

$$\frac{|\, \hat{s}_{n+kd} - s_{n+kd} \,|}{R_n(d)} > R_T \tag{7.4.32}$$

Abarbanel, Brown, Sidorowich, and Tsimring (1993) suggest values for R_T in the range $10 \leq R_T \leq 50$.

For the data in Figure 7.4.1, Abarbanel et al. (1993) used the aforementioned procedure to determine variation of the percentage of false nearest neighbors with the embedding dimension. The results are shown in Figure 7.4.6. Clearly, the number of false nearest neighbors drops to zero at $d = 3$, whereas the sufficient dimension from the embedding theorem is 5.

Again, the aforementioned procedure is good in a perfect world where an infinite amount of noise-free data are available. In the presence of noise, Abarbanel et al. (1993) suggest using the criterion

$$\frac{R_n(d+1)}{R_A} \geq 2 \tag{7.4.33}$$

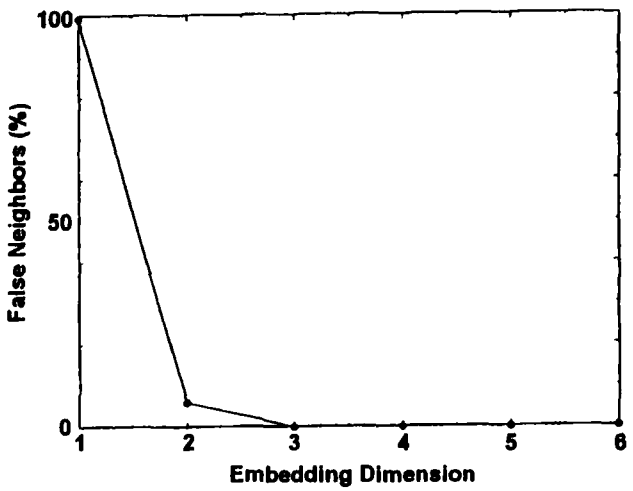

Figure 7.4.6: Variation of the percentage of false nearest neighbors with the embedding dimension calculated for the data in Figure 7.4.1. Reprinted with permission from Abarbanel et al. (1993).

for false nearest neighbors, where

$$R_A^2 = \frac{1}{N}\sum_{n=1}^{N}(s_n - s_{av})^2 \tag{7.4.34}$$

with s_{av} being the average of the s_n defined by

$$s_{av} = \frac{1}{N}\sum_{n=1}^{N} s_n \tag{7.4.35}$$

We note that R_A is a measure of the size of the attractor. In Figure 7.4.7, we show, after Abarbanel et al. (1993), the influence of adding uniform random numbers lying in the interval $[-L, L]$ to the x signal from the Lorenz system, shown in Figure 7.4.1. For this system, $R_A \approx 12$ and the different contamination levels L/R_A considered are indicated in the figure. Clearly, for values of L/R_A up to 0.5, a definite indication of a low-dimensional signal is discernible. When the contamination level is low, the residual percentage of false nearest neighbors provides an indication of the noise level. According to

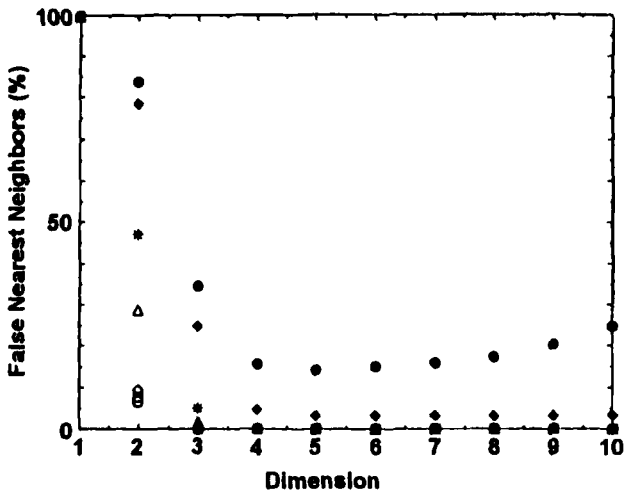

Figure 7.4.7: Influence of noise on variation of the percentage of false nearest neighbors with the embedding dimension calculated by for the data in Figure 7.4.1. The rms values of L/R_A corresponding to the open circles, open squares, open diamonds, open triangles, stars, solid diamonds, and solid circles are 0.0, 0.005, 0.01, 0.05, 0.1, 0.5, and 1.0, respectively. Reprinted with permission from Abarbanel et al. (1993).

Abarbanel et al. (1993), one may choose the embedding dimension to correspond to that at which false nearest neighbors drop to below, say, 1%.

Like the scheme discussed above, the schemes of Aleksic (1991) and Liebert, Pawelzik, and Schuster (1991) are based on topological considerations. Aleksic (1991) determines the minimum embedding dimension to be the dimension at which the dynamics in the reconstructed space is described by a continuous mapping (in a continuous mapping, images of close points are close). Liebert et al. (1991) provide a criterion for optimal embedding in terms of the distances between neighboring points on the reconstructed attractor for different embedding dimensions. Buzug and Pfister (1992) use what is called a **fill-factor method** to carry out a reconstruction. Roux, Simoyi, and Swinney (1983) have suggested that d can be systematically increased until the trajectories in the reconstructed state space no longer appear to cross or

intersect. Application of this criterion becomes difficult as d_a becomes large. Typically, we use the first method described in this section, saturation of system invariants. Thus, we calculate the dimension for a particular motion in the space of the delayed coordinates for different values of d and systematically increase d until the dimension values for the different d saturate to a common value.

7.4.2 Choosing the Time Delay

For an infinite amount of noise–free data, one can arbitrarily choose the time delay τ according to the embedding theorem of Mané (1981) and Takens (1981). However, using experimental data, Roux, Simoyi, and Swinney (1983) showed that the quality of the reconstructed portraits depends on the value of τ. For small τ, $s(t)$ and $s(t+\tau)$ are very close to each other in numerical value, and hence they are not independent of each other. On the other hand, for large values of τ, $s(t)$ and $s(t+\tau)$ are completely independent of each other, and any connection between them in the case of chaotic attractors is random because of the butterfly effect. Consequently, we need a criterion for an intermediate choice that is large enough so that $s(t)$ and $s(t+\tau)$ are independent but not so large that $s(t)$ and $s(t+\tau)$ are completely independent in a statistical sense. Moreover, the time delay must be a multiple of the sampling time τ_s because the data are available at these times only and any interpolation may introduce errors, as in the case of estimating the derivatives. For delay coordinate reconstructions with nonuniformly sampled data, we refer the reader to Breedon and Packard (1992). There are many systematic approaches for choosing the time delay. In this section, we discuss three of these approaches.

Autocorrelation Function

The autocorrelation function of the sampled data set $s_i = s(t_0 + i\tau_s)$, where τ_s is the sampling time and $i = 1, 2, \cdots, N_0$, is given by

$$C(\tau) = \frac{\sum_{k=1}^{N_0} [s(t_0 + k\tau_s + \tau) - s_{av}][s(t_0 + k\tau_s) - s_{av}]}{\sum_{k=1}^{N_0} [s(t_0 + k\tau_s) - s_{av}]^2} \qquad (7.4.36)$$

where

$$s_{av} = \frac{1}{N_0} \sum_{k=1}^{N_0} s(t_0 + k\tau_s) \tag{7.4.37}$$

Then, if the autocorrelation function $C(\tau)$ has a zero crossing at τ, the corresponding value of the time delay is chosen to be τ. Otherwise, the first local minimum of the autocorrelation function is used to specify τ.

We note that the autocorrelation function provides only a linear measure of the independence between the coordinates $s(t_0 + k\tau_s)$ and $s(t_0 + k\tau_s + \tau)$. To illustrate this, we assume that these coordinates are connected by the linear relation

$$s(t_0 + k\tau_s + \tau) - s_{av} = C(\tau)\,[s(t_0 + k\tau_s) - s_{av}] \tag{7.4.38}$$

Then, we determine $C(\tau)$ by minimizing the average of the square of the error over the observations; that is, by minimizing the mean-square error

$$e = \sum_{k=1}^{N_0} \{[s(t_0 + k\tau_s + \tau) - s_{av} - C(\tau)]\,[s(t_0 + k\tau_s) - s_{av}]\}^2 \tag{7.4.39}$$

Setting the derivative of e with respect to $C(\tau)$ equal to zero yields (7.4.36). Consequently, choosing τ to be the first zero of $C(\tau)$ would, on the average, make $s(t_0+k\tau_s+\tau)$ and $s(t_0+k\tau_s)$ linearly independent.

In Figure 7.4.8, we show variation of the autocorrelation function $C(\tau)$ with the time lag τ for the Lorenz data of Figure 7.4.1. Clearly, the first zero crossing occurs at $\tau \approx 30\tau_s$.

At the current time, the autocorrelation function is quite widely used to determine the time delay. In our limited experience, we have found the delay corresponding to the first zero crossing of the autocorrelation function or integer multiples of it to be adequate for the dimension calculations. We hasten to add that, when the autocorrelation function decays very slowly, there might be problems in choosing a delay (e.g., Gershenfeld, 1992). Moreover, as aforementioned, choosing τ to correspond to the first zero is the optimum linear choice from the point of view of the predictability in a least-squares sense of $s(t_0+k\tau_s+\tau)$ from a knowledge of $s(t_0+k\tau_s)$. Consequently, a number of researchers question its adequacy to determine the correlation due to the nonlinear process relating them (e.g., Abarbanel, 1995).

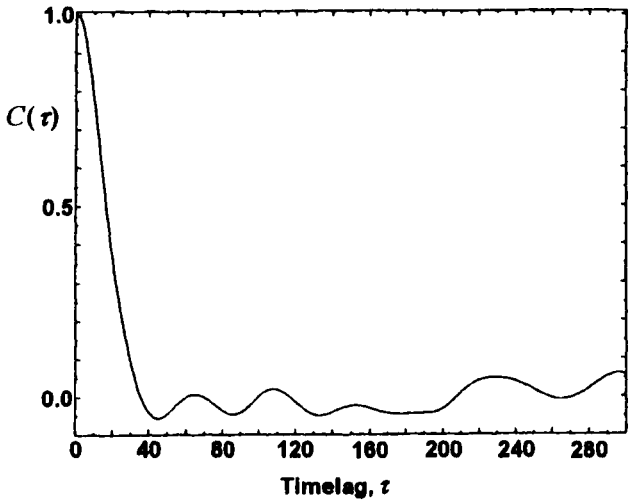

Figure 7.4.8: Variation of the autocorrelation function with the time lag calculated by for the data in Figure 7.4.1. Reprinted with permission from Abarbanel (1995).

Average Mutual Information

Fraser and Swinney (1986) used concepts of information theory and suggested that one should use the time delay corresponding to the first local minimum of the quantity called **mutual information**, which is a function of both linear and nonlinear dependencies between two variables. In the present context, the mutual information is a measure of the information (or predictability) that $s(t)$ can provide about $s(t+\tau)$. The two-dimensional approach of Fraser and Swinney was extended to higher dimensions by Fraser (1989a,b). He introduced a quantity called **redundancy** and suggested choosing a time delay corresponding to minimum redundancy.

Our present discussion follows along the lines of Abarbanel, Brown, Sidorowich, and Tsimring (1993). The idea underlying this approach is to identify how much information we can obtain about a measurement a_i drawn from a set A from a measurement b_j drawn from another set B. We assume that the probability of observing a_i out of the set A is $P_A(a_i)$, that the probability of observing b_j out of the set B is $P_B(b_j)$,

and that the joint probability of observing a_i from the set A and b_j from the set B is $P_{AB}(a_i, b_j)$. Then, according to Shannon's idea of mutual information (Gallager, 1968), the amount of information one learns in bits about a measurement of a_i from a measurement of b_j is given by

$$I_{AB}(a_i, b_j) = \log_2 \left[\frac{P_{AB}(a_i, b_j)}{P_A(a_i) P_B(b_j)} \right] \tag{7.4.40}$$

Then, the **average mutual information** I_{AB} between the sets of measurements A and B is given by

$$I_{AB} = \sum_{i,j} P_{AB}(a_i, b_j) \log_2 \left[\frac{P_{AB}(a_i, b_j)}{P_A(a_i) P_B(b_j)} \right] \tag{7.4.41}$$

To apply the mutual information theory to the data set $s_k = s(t_0 + k\tau_s)$, we take the set A to be made up of the measurements $s(t_0 + i\tau_s)$ and the set B to be made up of the measurements $s(t_0 + i\tau_s + \tau)$. Then, (7.4.41) becomes

$$\begin{aligned} I(\tau) = & \sum_i P\left[s(t_0 + i\tau_s), s(t_0 + i\tau_s + \tau)\right] \\ & \times \log_2 \left\{ \frac{P\left[s(t_0 + i\tau_s), s(t_0 + i\tau_s + \tau)\right]}{P\left[s(t_0 + i\tau_s)\right] P\left[s(t_0 + i\tau_s + \tau)\right]} \right\} \end{aligned} \tag{7.4.42}$$

with $I(\tau) \geq 0$. When τ is large, the measurements $s(t_0 + i\tau_s + \tau)$ and $s(t_0 + i\tau_s)$ are completely independent for a chaotic signal and hence

$$P\left[s(t_0 + i\tau_s + \tau), s(t_0 + i\tau_s)\right] = P\left[s(t_0 + i\tau_s + \tau)\right] P\left[s(t_0 + i\tau_s)\right] \tag{7.4.43}$$

Therefore, $I(\tau) \to 0$ as $\tau \to \infty$.

To evaluate $P[s(t_0 + i\tau_s)]$, we project the time series back onto the s axis. Then, the histogram formed by counting the frequency with which any of the values of s appears, when normalized, yields $P[s(t_0 + i\tau_s)]$. To evaluate $P[s(t_0 + i\tau_s + \tau)]$, we note that, if the time series is long and stationary, then

$$P[s(t_0 + i\tau_s + \tau)] = P[s(t_0 + i\tau_s)] \tag{7.4.44}$$

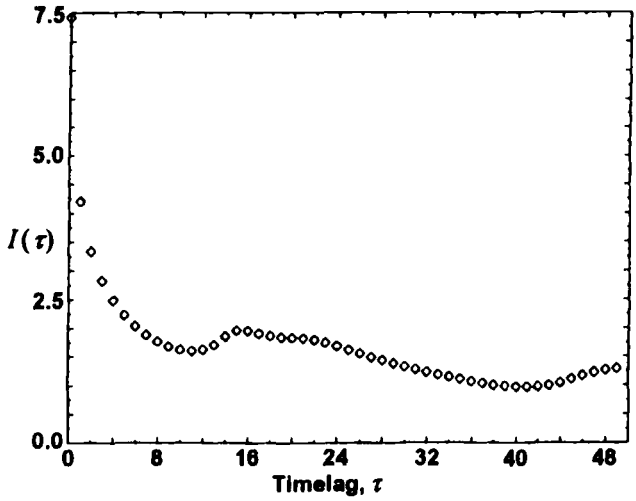

Figure 7.4.9: Variation of the average mutual information with the time lag calculated for the data in Figure 7.4.1. Reprinted with permission from Abarbanel (1995).

To evaluate the joint probability $P[s(t_0 + i\tau_s + \tau), s(t_0 + i\tau_s)]$, we form the two-dimensional histogram and count the number of times a box in the plane $s(t_0+i\tau_s+\tau)$ versus $s(t_0+i\tau_s)$ is occupied and then normalize this distribution.

Fraser and Swinney (1986) suggest selecting the value of τ that corresponds to the minimum of $I(\tau)$. In Figure 7.4.9, we show variation of the average mutual information $I(\tau)$ with the time lag τ for the Lorenz data of Figure 7.4.1. The first minimum of the mutual information occurs at $\tau = 10\tau_s$, whereas the first zero of the autocorrelation occurs at $\tau = 30\tau_s$. Fraser and Swinney (1986) found that the visual picture of the reconstructed attractor for low-dimensional systems, such as the Lorenz attractor, changes smoothly for variations around the minimum of $I(\tau)$.

There are situations in which $I(\tau)$ does not have a minimum. These situations include evolutions described by maps and evolutions of continuous-time systems determined with "large" integration steps or sampling times. Abarbanel, Brown, Sidorowich, and Tsimring (1993) suggest using $\tau = 1$ or 2 if the data comes from a map or choosing τ

so that $I(\tau) \approx \frac{1}{5} I(0)$.

Generalized Correlation Integral

Liebert and Schuster (1989) showed that the first minimum of the logarithm of the generalized correlation integral provides an easily calculable criterion for the proper choice of the time delay τ. Further, they examined the relationship between the correlation integral and the average mutual information. Their scheme requires a relatively smaller amount of data and seems to be easier to implement than the average mutual information scheme. As discussed later, the concept of the correlation integral is useful for certain types of dimension calculations.

The probability $P_m(r)$ of finding other d-dimensional states $\mathbf{x}_k$ within a sphere of radius r centered around the state $\mathbf{x}_m$ is given by (7.4.9). This probability is a function of τ because the reconstructed orbit depends on τ. Using (7.4.9), we define the so-called **generalized correlation integral** $C_1^d(r, \tau)$ as

$$C_1^d(r, \tau) = \lim_{q \to 1} \left[\frac{1}{M} \sum_{m=1}^{M} P_m^{q-1}(r, \tau) \right]^{1/(q-1)} \tag{7.4.45}$$

Using data from the Rössler system, the Mackey-Glass equation, and voltage measurements on barium sodium niobate crystals, Liebert and Schuster (1989) concluded that the first local minimum of the correlation integral with respect to τ together with the corresponding minimum of the mean-square deviations around the straight line defined by

$$\log C_1(r) = D_1 \ln(r) \quad \text{as} \quad r \to 0 \tag{7.4.46}$$

provide a practical and easy way to calculate a criterion for the best choice of the time delay needed to reconstruct an orbit in state space from a scalar time series.

7.4.3 Two or More Measured Signals

In some experimental situations, it may be possible to obtain two or more different signals from the system in question. These signals may be used as coordinates for construction of a state space. Often

in structural dynamics studies, signals from transducers mounted at different locations on a structure can be used to construct a state space. This construction may be explained as follows. Let $w(x,t)$ describe the displacement at time t and spatial location s on the structure. Further, let

$$w(x,t) = \sum_{i=1}^{n} \psi_i(x) v_i(t) \tag{7.4.47}$$

where n represents the number of modes used in the approximation, ψ_i represents the mode shape of the ith mode, and v_i represents the temporal function associated with the ith mode. We have implicitly assumed that the spatial and temporal informations are correlated. The n states v_j and their time derivatives represent the $2n$ state variables required to describe the motion of the structure. Let us suppose that a measurement of w taken at location $s = s_j$ yields the signal

$$w_j = \sum_{i=1}^{n} C_i \psi_i(x_j) v_i(t) \tag{7.4.48}$$

where the C_i are calibration constants that depend on the sensor. If the location $x = x_j$ corresponds to the node of a mode shape, there will not be any contribution from the corresponding mode to the signal w_j. The collection of n signals obtained from n different locations essentially represents a set of n coordinates obtained by applying a rotational transformation to the n coordinates v_i.

In structural dynamics, the idea of using different spatial locations to generate "independent" coordinates was first employed in the study of Nayfeh and Zavodney (1988). They used signals obtained from two strain gauges, mounted at different locations on a harmonically forced structure, to construct a state space. Following them, Balachandran (1990), Balachandran and Nayfeh (1991), and Anderson, Balachandran, and Nayfeh (1992) also employed this construction. In all of these studies, projections of motions onto a two–dimensional space were used to characterize the different motions. An example is provided in Figure 7.3.2. In the work of Guckenheimer and Buzyna (1983), simultaneous measurements in a fluid mechanics experiment were used to construct an embedding space and carry out dimension calculations.

Multiple measurements can often be exploited to reduce noise. Let us suppose that there are m measured signals w_m, where the measurement errors associated with the different signals are independent. Then, a linear combination of the m signals can be chosen so that the signal-to-noise ratio is high by using procedures similar to singular value analysis (e.g., Preisendorfer, 1988). Sauer, Yorke, and Casdagli (1991) provide a theoretical basis for state-space reconstruction by using m signals. The integer m should be greater than $2d_a$, where d_a is the box-counting or capacity dimension of the original attractor. Sauer et al. (1991) also discuss how a mixture of independent (measured) coordinates and delayed coordinates can be used to carry out a reconstruction.

7.5 FOURIER SPECTRA

The **Fourier** or **frequency spectra** help in distinguishing among periodic, quasiperiodic, and chaotic motions and are typically used to study stationary signals. The frequency spectrum can be either an **amplitude** or a **power spectrum**. In an amplitude spectrum, the Fourier amplitude is displayed at each frequency. On the other hand, in a power spectrum, the square of the Fourier amplitude per unit time is displayed at each frequency.

The **Fourier transform** of a signal $x(t)$ is defined as

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-2i\pi f t} dt \tag{7.5.1}$$

where f denotes the frequency and $X(f)$ is a complex quantity. In writing (7.5.1), we have assumed that $x(t)$ is integrable; that is,

$$\int_{-\infty}^{+\infty} | x(t) | \, dt < \infty \tag{7.5.2}$$

In theory, the Fourier transform can be used to determine the frequency content of a signal $x(t)$ if it is known for $-\infty < t < +\infty$ and is integrable. However, a stationary signal that exists for all t is not integrable. Besides, in practice, $x(t)$ is known for only a finite length