Comment on "Singular-value decomposition and embedding dimension"

D. S. Broomhead, R. Jones, and Gregory P. King*
Royal Signals and Radar Establishment, Great Malvern, Worcestershire WR14 3PS, United Kingdom
(Received 29 June 1987)

In this Comment we consider the conclusion drawn by Mees, Rapp, and Jennings [Phys. Rev. A **36**, 340 (1987)] concerning the usefulness of the singular-system analysis of a trajectory matrix (obtained by applying Takens's construction to time-series data). We point out that this is primarily due to their misinterpretation of our work, and to an assumption they apparently made concerning the conditions under which a map preserves the dimension of a set. We attempt to clarify the issues.

In their paper[1] Mees, Rapp, and Jennings (hereafter referred to as MRJ) address the question of how one should choose a dimension of the embedding space when applying the method of delays to time-series data. In particular, they consider rank estimation of the trajectory matrix (referred to as the embedding matrix in Ref. 1) constructed from a time series. They present various numerical results which lead them to the conclusion that the rank of the trajectory matrix does not give useful dynamical information.

In developing these ideas they draw upon previous work by us.[2,3] We wish to take this opportunity to comment on and perhaps clarify the issues raised.

In the following discussion we use the notation of Refs. 2 and 3: n is the embedding dimension, $\mathbf{X}$ denotes the $N\times n$ trajectory matrix obtained by constructing n-component vectors $(v_i, v_{i+1}, \ldots, v_{i+n-1})$ from a time series of measurements $\{v_k \mid k=1,\ldots,N+n-1\}$ using Takens's construction, $d \leq n$ is the rank of $\mathbf{X}$ or the number of linearly independent rows or columns of $\mathbf{X}$ and m is the dimension of the manifold to which the trajectory is confined.

In Ref. 2 the properties of d were investigated in some detail. In particular, care was taken to distinguish between d, the signal processing notion of *number of degrees of freedom*, and m, the dimension of the embedded manifold: while m is a topological invariant of the manifold preserved by the embedding, d is not. Indeed, it was pointed out that d depends on (1) the quality of the data (e.g., the precision of the measurement and the signal-to-noise ratio), (2) the time interval τ_s between measurements, and (3) the time span $\tau_w = n\tau_s$ of the window used to construct the embedding. Furthermore, d is sensitive to the nature of the measurement (i.e., what is measured and where in the system it is measured). In short, d is strongly dependent upon the characteristics of the particular time series under consideration.

MRJ consider the general use of d for the characterization of the dimension of a dynamical system. In light of the discussion in Ref. 2 summarized above, this use cannot be sustained. Rather, the importance of d lies in its role as an *effective* embedding dimension when $d < n$, since then it is d rather than n which should exceed $2m$ in order to satisfy Takens's embedding criterion.[4] We should understand, therefore, when we can expect d to be less than n. For the case where $\mathbf{X}$ is constructed from experimental data obtained by sampling a continuous, noisy, physical (band-limited) process two situations can exist: the process is either *undersampled* (sampled at a frequency less than the band limit) or *oversampled* (sampled at a frequency greater than the band limit). When the signal is undersampled, the measurements $v(t)$ and $v(t+\tau_s)$ are on average statistically independent; this will typically cause $\mathbf{X}$ to be full rank ($d=n$). (In any case, it is worth pointing out that, in general, undersampling a signal is undesirable, since it introduces aliasing effects which lead to data corruption and loss of high-frequency information.) On the other hand, when the signal is oversampled, the measurements $v(t)\approx v(t+\tau_s)$ and consequently adjacent column and/or row vectors of $\mathbf{X}$ will be nearly the same and therefore linearly dependent; this will cause $\mathbf{X}$ to be less than full rank ($d<n$). Indeed, as the sampling frequency is increased further above the band limit, so the rank deficiency of $\mathbf{X}$ will increase. Therefore, when $\mathbf{X}$ is constructed with an oversampled signal in the manner described in Refs. 2 and 3, $d<n$ should be obtained.

We should emphasize that our intention in Ref. 2 was not to propose a new dimension for the characterization of dynamical systems. Our purpose was to introduce the concept of, and limits for, the timescale τ_w. In Ref. 2 we argued that m and not d is an intrinsic property of the dynamical system and we feel that MRJ are really of this opinion too. We have since shown how m may be estimated using a local singular system analysis.[4-6]

When estimating d from the singular value decomposition of $\mathbf{X}$, MRJ state that one should look for an abrupt decrease between adjacent singular values. We do not agree with this. One has to establish the number of singular values which exceeds the magnitude of the error or noise associated with the elements of $\mathbf{X}$.[7] An abrupt decrease is of importance only if subsequent singular values are insignificant in this sense. Naturally, one would expect the precision of the computer and its algorithm to be adequate so that the dominant source of error is the noise in the data.

MRJ then claim that "the efficacy of singular-value decomposition...in reducing dimension calculations...

depends on an abrupt decrease in the magnitude of the singular values." The authors appear to base these comments on the mistaken assumption that it is necessary for a map to preserve lengths in order for it to preserve the dimension of sets. However, in proving the Whitney embedding theorem[8] one shows that generically, the projection of the embedded manifold onto *any* subspace of the embedding space which has dimension greater than $2m$ is an embedding, and therefore preserves the dimension of the manifold and even of subsets of the manifold (e.g., the attractor). The magnitudes of the singular values corresponding to directions orthogonal to the subspace do not enter this argument: it requires only that the singular vectors which span the subspace should correspond to significant (in the sense described above) singular values. The virtue of the singular system method is that it identifies the subspaces for which the signal to noise is maximized.

MRJ imply that Ref. 2 advocates "thinning" the data (i.e., by deleting alternate rows and columns of $\mathbf{X}$) to determine the optimum embedding dimension. We wish to emphasize that this was never advocated by us. "Thinned" data were used in Ref. 2 to demonstrate the dependence of the singular spectrum on the sampling interval τ_s. Indeed, "thinning" $\mathbf{X}$ corresponds to projecting the retained portion of the data onto a nonoptimum subspace.

Finally, we address the question of the numerical implementation of these ideas. Typically, the measured signal will be represented by a sequence of integers obtained from an analog-to-digital (AD) converter. The AD converter will have limited precision which will add quantization noise to the input signal, which itself will be noisy. These noise components result in the data being able to support only an imperfect representation of the system's dynamic properties. The computation will be carried out also to a limited precision which must be greater than the precision of the data. If this condition is satisfied, the computation will give d singular values which appear above a noise floor determined by the data precision. The corresponding singular vectors define a d-dimensional subspace which contains all the information that one is entitled to extract from the signal.

Within the above constraints there is no preferred algorithm for the singular-value decomposition. The choice of algorithm must depend upon the circumstances of its implementation and the source of data to be processed. MRJ advocate calculating the singular values of $\mathbf{X}$ using the Golub-Reinsch singular-value decomposition algorithm for a rectangular matrix. They support this with their Fig. 5 with which we agree (except that, from the evidence of their graph, the limitation of their DEC-10 computer and algorithm is 16 significant figures and not 18). It is clear from their figure that this algorithm is capable of processing data which is of a far higher precision than that which could be obtained from any realistic experiment. On the other hand, the diagonalization of $\mathbf{X}^T\mathbf{X}$ using similar precision computers (Data General 30 and VAX 8600) gives adequate precision for the processing of up to 24-bit data ($\sim$6 significant figures). This is far in excess of that which we might expect from a good AD converter (12 to 16 bits).

We note that in this approach the accumulation and diagonalization of $\mathbf{X}^T\mathbf{X}$ should be done in double precision. The result presented in Refs. 2 and 3 used data from the integration of the Lorenz model with a Runge-Kutta algorithm in single precision. This was erroneously accumulated to form $\mathbf{X}^T\mathbf{X}$ in single precision. This accounts in part for the different noise levels calculated by MRJ and Ref. 2.

---

*Also at Mathematics Department, Imperial College, London SW7 2BZ, United Kingdom.

[1]A. I. Mees, P. E. Rapp, and L. S. Jennings, Phys. Rev. A **36**, 340 (1987).

[2]D. S. Broomhead and G. P. King, Physica **20D**, 217 (1987).

[3]D. S. Broomhead and G. P. King, in *Nonlinear Phenomena and Chaos,* edited by S. Sarkar (Adam Hilger, Bristol, 1986), p. 113.

[4]G. P. King, R. Jones, and D. S. Broomhead, Nucl. Phys. B (Proc. Suppl.) **2**, 379 (1987).

[5]D. S. Broomhead, R. Jones, and G. P. King, J. Phys. A **20**, L563 (1987).

[6]D. S. Broomhead, R. Jones, G. P. King, and E. R. Pike, in *Chaos, Noise and Fractals,* edited by E. R. Pike and L. A. Lugiato (Adam Hilger, Bristol, 1987), pp. 15–27.

[7]G. H. Golub and C. F. van Loan, *Matrix Computations* (The John Hopkins University Press, Baltimore, 1985).

[8]M. W. Hirsch, *Differential Topology* (Springer, New York, 1976).