Identification of continuous, spatiotemporal systems

H. Voss,[1] M. J. Bünner,[2] and M. Abel[1]

[1]*Institut für Theoretische Physik, Universität Potsdam, Am Neuen Palais 10, D-14415, Potsdam, Germany*
[2]*Max-Planck-Institut für Physik Komplexer Systeme, Nöthnitzer Strasse 38, D–01187 Dresden, Germany*

We present a method for the identification of continuous, spatiotemporal dynamics from experimental data. We use a model in the form of a partial differential equation and formulate an optimization problem for its estimation from data. The solution is found as a multivariate nonlinear regression problem using the alternating conditional expectation algorithm. The procedure is successfully applied to data, obtained by simulation of the Swift-Hohenberg equation. There are no restrictions on the dimensionality of the investigated system, allowing for the analysis of high-dimensional chaotic as well as transient dynamics. The demands on the experimental data are discussed as well as the sensitivity of the method towards noise. [S1063-651X(98)10603-7]

The unstable dynamics observed in spatially extended systems attracted huge experimental and theoretical research activity in the last decades (see [1,2] and references therein). Progress has been achieved by describing the dynamics in the vicinity of bifurcations with the help of universal amplitude equations, vastly reducing the complexity of the involved models. Additionally, the research has concentrated on the classification of the observed instabilities and the resulting patterns, and the investigation of scaling laws and intermittency effects [1]. For most investigations the models for spatiotemporal systems arise from mainly theoretical considerations and their validity is affirmed by the comparison with experimental measurements. Here we address the problem of finding a model, which describes the dynamics of an observed system, directly from experimental data. For systems that exhibit temporal low-dimensional chaotic motion this was accomplished with the help of nonlinear maps [3]. Other general methods rely on some sort of mode expansion and were successfully applied [4]. A different approach consists in a nonparametric model identification as proposed for systems with a time delay feedback [5,6].

In this paper we extend that approach to the identification of the underlying evolution equation of spatially extended systems. First we formulate an optimization problem for finding a model equation from the data. Then we rewrite the equations in the form of a multivariate nonlinear regression problem. As a last step, we use a different kind of numerical algorithm for solving the problem. Our approach does not include any parameter dependencies; rather, those are delivered as a byproduct. We discuss the identification of homogeneous and autonomous partial differential equations (PDE's), but emphasize that the ideas are quite general and can be applied to other problems such as finite-difference equations, coupled-map lattices, and integrodifferential equations.

We assume the dynamics of the system under consideration to be governed by a PDE of the form

$$\mathcal{F}[\vec{\nabla}, \partial_t, \vec{u}(\vec{x}, t)] = 0, \qquad (1)$$

where $\vec{u}$ is the field variable with N components, $\mathcal{F}$ is an N-dimensional function of $\vec{u}(\vec{x}, t)$, and its spatial and temporal derivatives, with t and $\vec{x}$ the temporal and spatial variables, respectively.

To simplify the analysis and the representation, we only consider systems with a single component u and at most two spatial dimensions. We note that the following considerations are valid, in principle, also for multicomponent systems and higher dimensions in space.

In the following, we discuss the procedure for estimating a PDE of the form (1) from experimental data. Therefore, we distinguish between the solution $u(\vec{x}, t)$ and the data $v(\vec{x}, t)$. For the sake of simplicity, we denote both the continuous space-time variables of the model field and the discrete space-time variables of the data by the same symbols $\vec{x}$ and t. Considering the data v as random variables, we act on a probability space and denote all entities estimated from the data by a hat $\hat{\ }$.

Since one can get only an estimate of the true function $\mathcal{F}$ from the data v, all one can achieve is to estimate the analogue of Eq. (1),

$$\hat{\mathcal{F}}[\hat{\vec{\nabla}}, \hat{\partial}_t, v(\vec{x}, t)] = 0, \qquad (2)$$

where $\hat{\mathcal{F}}$ is the estimate of $\mathcal{F}$ and the derivatives have been replaced by estimates computed from the data v.

To obtain $\hat{\mathcal{F}}$, we formulate from Eq. (1) the following optimization problem [7]:

$$\inf_{\mathcal{F}} \|\mathcal{F}\| = e^2, \qquad (3)$$

where the norm has to be specified according to the respective problem. The optimization lies in varying $\mathcal{F}$ until e^2 converges to the minimum.

The function $\mathcal{F}$ is defined as a function of operators, e.g., ∇^2, ∂_t, id, ... We denote the set of constituent operators by $\{\mathcal{O}_i\}_{i=1, \ldots, K}$. Note that in our definition of the differential operators $\mathcal{O}_i$, any product terms like $u^2 \partial_{xx}$, $\partial_t \partial_x$ are also included. Since we consider a nonlinear problem, it is useful

to split the function $\mathcal{F}$ into subfunctions Φ_i, which have as arguments $\mathcal{O}_i u$, e.g., $\Phi(\partial_{xx}u)=(\partial_{xx}u)^\alpha$ or $\Phi(\mathrm{id}u)=u^\beta$. These Φ_i are elements of some class of functions S, which has to be specified according to the problem. Thus, the final representation of Eq. (1) reads

$$\mathcal{F}=\sum_{i=0}^{K}\Phi_i(\mathcal{O}_i u)=0. \tag{4}$$

We want to determine the constituents $\mathcal{O}_i u$ and the functions Φ_i of $\mathcal{F}$ from a data set v. Therefore, we estimate at first $\mathcal{O}_i u$ from the data by finite differencing or alternative schemes. Especially in the presence of noise, Fourier methods or kernel estimation could be helpful. The result consists of K random variables $\hat{\mathcal{O}}_i \equiv \hat{\mathcal{O}}_i u$. If the underlying PDE is linear, i.e., the function $\mathcal{F}$ a multilinear one, we could solve the problem with linear regression methods. Second, to obtain a solution for the nonlinear problem, we solve

$$\inf_{\Phi \in S}\left\|\sum_{i=1}^{K}\Phi_i(\hat{\mathcal{O}}_i)\right\|=e^2 \tag{5}$$

in varying the functions Φ_i. The result are the estimates $\hat{\Phi}_i$. Up to now, we anticipated the number and type of operators $\mathcal{O}_i$. For unknown systems, which are the ones we treat, it would be necessary to extend the number of operators K to infinity. In practice one has to select a finite number. Redundant terms then deliver $\hat{\Phi}_i \approx 0$ as result. It is important to find a reasonable initial guess for the operators $\mathcal{O}_i$ of Eq. (4). If there already exists some description for the system in a special state, one starts with the operators that appear in the known equations and tries to determine additional terms that may appear when leaving that state. This situation appears, e.g., near some critical points, where one can start with already derived amplitude equations. If no known description exists, one should extract as much information as possible from other clues, e.g., symmetries, and then follow the usual schemes for deriving amplitude equations [1], where especially product terms are included in higher expansion terms.

In the case of data analysis, the optimization problem (5) becomes a multiple nonlinear regression problem that can be solved using the alternating conditional expectation algorithm (ACE) [8]. It has already been successfully applied in related fields of data analysis [6,9]. Since the algorithm is based on a statistical description, the norm used in Eqs. (3) and (5) is the L_2 norm. Using this algorithm, the functions Φ_i of Eq. (4) are seen as so-called *optimal transformations* [8], which are defined to solve regression problems of the form

$$E\left\{\left[\Phi_0(\hat{\mathcal{O}}_0)-\sum_{i=1}^{K}\Phi_i(\hat{\mathcal{O}}_i)\right]^2\right\}\overset{!}{=}\min. \tag{6}$$

The Φ_i are varied in the space S of the Borel-measurable functions with the additional requirement of zero expectation, $E[\Phi_i]=0$ $(i=0,\ldots,K)$, and $E[\Phi_0^2]=1$. The convergence of the ACE algorithm in the case of discrete samples rather than random variables has also been proved in [8]. The
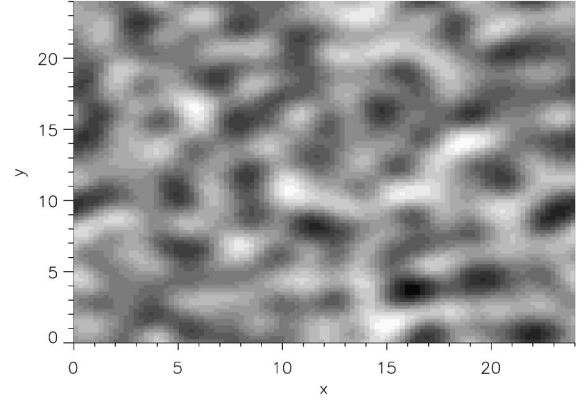

FIG. 1. The data sample $v(x,y,t_0)$ for the central time point t_0 encoded in gray values (small values dark). All quantities are plotted in dimensionless units.

minimization of the expectation value (6) is equivalent to the calculation of the *maximal correlation* Ψ [10]. Instead of e^2 we use the maximal correlation as a measure of the quality of the result: $\Psi \in [0,1]$ equals 1 for perfect estimation, the smaller it is the worse the estimation. The ACE algorithm achieves the maximization by iteratively transforming each input $\hat{\mathcal{O}}_i$ by suitable, generally *nonlinear*, transformations such as to yield a *linear* relationship between the output—the new random variables $\Phi_i(\hat{\mathcal{O}}_i)$.

As a first example, we analyze data $v(\vec{x},t)$ from the numerical integration of the Swift-Hohenberg equation [11]. The model is of the form

$$\partial_t u=[r-(\nabla^2+k^2)^2]u-u^3=(r-k^4)u-u^3$$
$$-2k^2(\partial_{xx}+\partial_{yy})u-(\partial_{xxxx}+\partial_{yyyy}+2\partial_{xxyy})u. \tag{7}$$

The global dynamics of the model can be derived from a potential, such that the asymptotic time dependence is trivial [2]. Therefore, we analyze a transient state to have a sufficient variation in the time derivative. For the identification procedure we use data produced by an explicit Euler integration scheme with a time step of 10^{-4}, a spatial discretization of $\Delta x=\Delta y=0.25$, and periodic boundary conditions. As initial conditions we choose uniformly distributed independent random numbers from the interval $[-10,10]$. The parameters are $r=0.1$ and $k=1$.

The differential operators are estimated by symmetric differencing schemes, e.g., $\hat{\partial}_t v(\vec{x},t)=[v(\vec{x},t+\Delta t)-v(\vec{x},t-\Delta t)]/2\Delta t$. Thus, to estimate the time derivatives of first order in each spatial data point, we need three consecutive ''pictures'' of data. The field size is 100×100 points, i.e., the data set $v(\vec{x},t)$ contains 3×10^4 values. The data for the central time point are shown in Fig. 1. In the following we can drop without ambiguity the hat $\hat{}$.

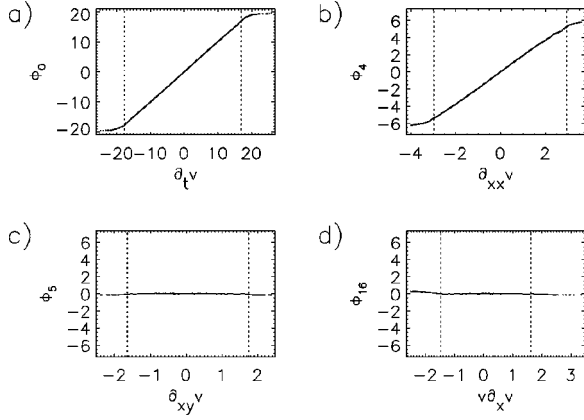To identify the unknown system, we use an ansatz with nonmixed terms (like $\partial_x v$) up to fourth order in the spatial

FIG. 2. Four exemplary estimates for the optimal transformations (Φ_0, Φ_4, Φ_5, Φ_{16}) are shown. The linear occurrence of $\partial_t u$ and $\partial_{xx} u$ is clearly recovered by Φ_0 and Φ_4, respectively. The redundant terms $\partial_{xy} u$ and $u \partial_x u$ vanish approximately. The dotted lines mark the interval on the abscissa in which 98% of the data points are located. Due to a very nonhomogeneous distribution of the data the optimal transformations outside the marked interval cannot be estimated reliably. The results for the optimal transformations not shown here deliver estimates of equal quality. As in Fig. 1, units are dimensionless.

derivatives. To show how to handle mixed terms we additionally include the terms $v\partial_x v$, $v\partial_y v$, $\partial_x v\partial_y v$:

$$\Phi_0(\partial_t v) = \Phi_1(v) + \Phi_2(\partial_x v) + \Phi_3(\partial_y v) + \Phi_4(\partial_{xx} v)$$
$$+ \Phi_5(\partial_{xy} v) + \Phi_6(\partial_{yy} v) + \Phi_7(\partial_{xxx} v) + \cdots$$
$$+ \Phi_{10}(\partial_{yyy} v) + \Phi_{11}(\partial_{xxxx} v) + \cdots + \Phi_{15}(\partial_{yyyy} v)$$
$$+ \Phi_{16}(v\partial_x v) + \Phi_{17}(v\partial_y v) + \Phi_{18}(\partial_x v\partial_y v), \quad (8)$$

with the nine redundant terms $\Phi_2(\partial_x v)$, $\Phi_3(\partial_y v)$, $\Phi_5(\partial_{xy} v)$, $\Phi_7(\partial_{xxx} v)$, $\Phi_8(\partial_{xxy} v)$, $\Phi_9(\partial_{xyy} v)$, $\Phi_{10}(\partial_{yyy} v)$, $\Phi_{12}(\partial_{xxxy} v)$, $\Phi_{14}(\partial_{xyyy} v)$, $\Phi_{16}(v\partial_x v)$, $\Phi_{17}(v\partial_y v)$, and $\Phi_{18}(\partial_x v\partial_y v)$. We choose this ansatz as a compromise between generality and computational effort. Comparing Eq. (7) with Eq. (8), one expects, in particular, the following for the solution of Eq. (6): Up to an arbitrary common factor, Φ_0 should be the identity, Φ_1 should be a polynomial of third order, and for $i = 4,6,11,13,15$ the Φ_i should be linear functions $\alpha_i \mathcal{O}_i$. All other estimates should vanish. Furthermore, we expect for the slopes of the linear functions, after division by the slope of the left-hand side to remove the arbitrary common factor, $\alpha_4 = \alpha_6 = \alpha_{13} = -2$, $\alpha_{11} = \alpha_{15} = -1$.

Performing the ACE algorithm, we find a maximal correlation of 0.9993 and optimal transformations as shown in Fig. 2. All functions approximate well the expected shape, and the terms that were expected to vanish are indeed very small compared to the other ones. Note that this does not mean that the values for the redundant terms are vanishing themselves but that these are independent of all the other terms involved. Comparison of the slope of the linear functions yields the possibility of estimating parameters; we obtain $\alpha_4 = \alpha_6 = -1.9$, $\alpha_{13} = -2.0$, and $\alpha_{11} = \alpha_{15} = -1.0$. The slopes of the terms that are expected to vanish are all smaller than 0.03 in absolute value. Since it is difficult to estimate
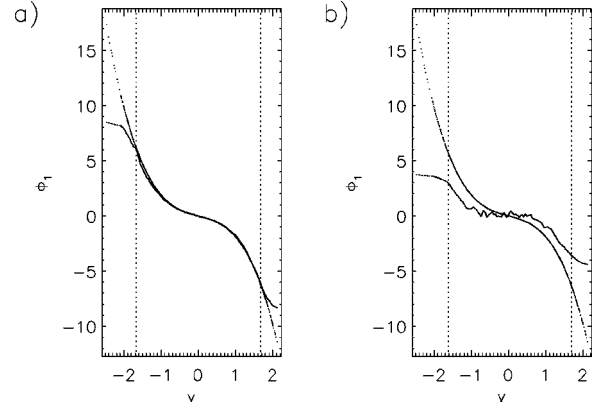

FIG. 3. The nonlinear term of the Swift-Hohenberg equation (7). (a) The estimated and the exact nonlinearity $-0.9u - u^3$. (b) The same with noise, as explained in the text. Units are dimensionless.

the errors of these values, it is advisable to check the range of validity of the reconstructed model by integration and comparison with the dynamics. In Fig. 3 the nonlinearity $-0.9u - u^3$ of Eq. (7) is compared explicitly with the estimate, taking now an ansatz with only the nonvanishing terms of the above result. Inside a range of 98% of the data values this term can be estimated with high accuracy.

From a practical point of view, it is essential to discuss the stability of the identification method against noise. The effect of noise is to increase the errors of the estimates for the partial differential operators applied to the data, and therefore also the value of the error estimate e^2. The identification detects the remaining correlations of the vector field and its derivatives via the function $\mathcal{F}$. In general, if the vector field and its derivatives are still strongly correlated for a reasonable low noise level of a noisy system, the minimum according to Eq. (5) should still be detectable.

To examine the stability against noise, we disturb each data point $v(\vec{x}, t)$ with additive Gaussian white noise with a standard deviation of 0.5% of the data standard deviation. The estimates of the partial derivatives are disturbed severely due to error propagation. In spite of this, using again Eq. (8), the overall shape of all functions can still be recovered satisfactorily, while being distorted (Fig. 3). The maximal correlation has decreased to 0.926.

Two other systems, which were analyzed with similar success as the example above, are (1) the Kuramoto-Sivashinsky equation (in the form given in Ref. [2]) in the fully chaotic regime and (2) a reaction-diffusion system with two components [12], where the nonlinearity in the inhibitor dynamics is a nonanalytic function. We reconstructed that function with high accuracy for different states, such as single rotating spiral and spiral-defect chaos.

Summarizing our experiences, we find the following data requirements. (1) In order to identify a system with N components it is sufficient to measure N independent variables. (2) The sampling of the vector fields in space as well as in time has to be appropriate to properly estimate the partial derivatives with respect to time and space. (3) In the preceding example we analyzed data at one time point but for an extended region in space. In other situations it may be useful to perform the analysis at a single spatial point but for a

longer interval in time, or even a combination of bothprocedures. However, as the example shows, even a small amount of data can be sufficient. (4) Since the reconstructed functions can only be determined at points that are attained by the data and the estimates for the differential operators, the data have to show a sufficient variability in space and time. (5) The inference from the data to a PDE is often not unique. But within the errors of the algorithm, our method reasonably reconstructs the dynamics on the trajectory given by the data. (6) We believe that due to the statistical nature of the method a higher noise level can typically be compensated by a larger data set.

We would like to stress the fact that the procedure presented above is essentially independent of the part of phase space the system moves on and the boundary and initial conditions. Thus we do not require the dynamics to be close to or away from an attractor. This kind of analysis is thus applicable in the case of high-dimensional chaotic motion, as exhibited by nonlinear spatially extended systems, as well as transient motion. We did not discuss nonhomogeneous or nonautonomous evolution laws, but, in principle, the arguments apply equally well, with some minor modifications, for those systems.

In conclusion, we presented a method for the identification of spatiotemporal systems by numerical reconstruction of the PDE which describes the system. The identification procedure consists in solving an optimization problem which results in a nonlinear regression problem. Using the ACE algorithm, we showed that this task can indeed be solved in the case of numerical examples. We consider the method a useful tool for the analysis of spatiotemporal systems and expect it to find a large variety of applications. The limits of this method and the requirements on the data have been discussed. Future work will concentrate on an extension of the ACE algorithm to the solution of multicomponent problems, and on the application to real data. Extensions to more general systems are envisaged.

[1] P. C. Cross and M. C. Hohenberg, Rev. Mod. Phys. **65**, 851 (1993).

[2] P. Manneville, *Dissipative Structures and Weak Turbulence* (Academic Press, New York 1990).

[3] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993); H. Kantz and T. Schreiber, *Nonlinear Times Series Analysis* (Cambridge University Press, Cambridge, 1997).

[4] C. Uhl, R. Friedrich, and H. Haken, Phys. Rev. E **51**, 3890 (1995); G. Berkooz, P. Holmes, and J. L. Lumley, Annu. Rev. Fluid Mech. **25**, 539 (1993); A. Armbruster, R. Heiland, and E. J. Kostelich, Chaos **4**, 421 (1994).

[5] M. J. Bünner, Th. Meyer, A. Kittel, and J. Parisi, Phys. Rev. E **54**, R3082 (1996): M. J. Bünner, Th. Meyer, A. Kittel, and J. Parisi, *ibid.* **56**, 5083 (1997).

[6] H. Voss and J. Kurths, Phys. Lett. A **234**, 336 (1997).

[7] E. Zeidler, *Nonlinear Functional Analysis and its Applications, Vol. III: Variational Methods and Optimization* (Springer, New York, 1984).

[8] L. Breiman and J. H. Friedman, J. Am. Stat. Assoc. **80**, 580 (1985).

[9] R. D. De Veaux and J. M. Steele, Technometrics **31**, 91 (1989).

[10] A. Rényi, *Probability Theory* (Akadémiai Kiadó, Budapest, 1970): For two random variables x_1 and x_2 Ψ is defined as $\Psi(x_1,x_2) := \sup_{\Phi_1,\Phi_2} |R(\Phi_1(x_1),\Phi_2(x_2))|$.

[11] J. Swift and P. Hohenberg, Phys. Rev. A **15**, 319 (1977).

[12] M. Bär and M. Eiswirth, Phys. Rev. E **48**, R1635 (1993); M. Hildebrand, M. Bär, and M. Eiswirth, Phys. Rev. Lett. **75**, 1503 (1995).