Spatiotemporal system reconstruction using Fourier spectral operators and structure selection techniques

Daolin Xu and Omid Khanmohamadi

School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798

(Received 4 April 2008; accepted 28 October 2008; published online 9 December 2008)

A technique based on trigonometric spectral methods and structure selection is proposed for the reconstruction, from observed time series, of spatiotemporal systems governed by nonlinear partial differential equations of polynomial type with terms of arbitrary derivative order and nonlinearity degree. The system identification using Fourier spectral differentiation operators in conjunction with a structure selection procedure leads to a parsimonious model of the original system by detecting and eliminating the redundant parameters using orthogonal decomposition of the state data. Implementation of the technique is exemplified for a highly stiff reaction-diffusion system governed by the Kuramoto–Sivashinsky equation. Numerical experiments demonstrate the superior performance of the proposed technique in terms of accuracy as well as robustness, even with smaller sets of sampling data. © 2008 American Institute of Physics. [DOI: 10.1063/1.3030611]

Techniques of system reconstruction from time series are employed for data-driven modeling, adaptive control, monitoring and fault diagnosis, and so on. Reconstruction of temporal processes has been discussed extensively in the literature; reconstruction of spatiotemporal processes, however, is still an underdeveloped topic. In this work we present a method, based on trigonometric spectral methods and a structure selection procedure, for reconstructing spatiotemporal systems from time series. This method is capable of handling the challenging case of reconstructing highly nonlinear spatiotemporal systems with the general “structure” given by a Volterra expansion, leading to a nonlinear partial differential equation (PDE) of polynomial type with terms of arbitrary derivative order and nonlinearity degree. The Fourier spectral differentiation operators are used to obtain the regression form and a structure selection procedure in an orthogonal space detects and eliminates the redundant parameters based on their insufficient contribution to the reduction of the mean squared error (MSE). The system is identified from the resulting reduced dimension model based on the maximum likelihood principle for estimates with the assumption of the prediction errors being zero-mean white noise. We exemplify the entire method with numerical experiments for the reconstruction of the Kuramoto–Sivashinsky partial differential equation—a nonlinear, highly stiff, reaction-diffusion system. The effect of structure selection on the parameters estimated with spectral methods is compared with that of finite differences. Estimate errors are examined through the analysis of their magnitude and variation, with and without the application of the proposed structure selection.

I. INTRODUCTION

Obtaining a model capable of predicting a phenomenon has always been a problem of paramount importance in science. As it turns out, not always can we take a direct approach to that problem; constructing a model based on first principles and constitutive laws is usually not feasible for the complex systems encountered in the real world. We lack either the required principles/laws or the full set of conditions that should be imposed on them to yield the particular solution sought. For instance, for biological/biomedical systems, environmental systems, or process control, we lack first principles that lead to reasonable models. Furthermore, even when we do not lack those principles or conditions and a model can be obtained based on them, there is always a risk of unreasonable discrepancy between the behavior predicted by the model and the true behavior of the system, resulting from too much approximation in derivation of the model or the use of improper parameter values—the notorious simulation dilemma.

Various techniques for identifying physical systems from time series have been reported in the literature. Three main approximation techniques, namely, global, local, and radial basis functions, are usually employed to obtain the predictor model. Gouesbet et al., employ these standard functions and use the scalar observable together with its derivatives as embedding coordinates to reconstruct the true system. However, the quality of the reconstructed model may depend critically on the observable used. Cases for which there are unobserved variables are treated based on the concept of delay coordinates introduced by Packard et al., and phase space reconstruction is accomplished based on Takens’ theorem. A combination of parametric modeling, based on the methods of boundary value problems (mainly multiple shooting techniques), and nonparametric modeling, based on optimal transformations and nonlinear regressions, is another way to determine both the structure and parameters of the model. As for validation, synchronization may be used as a nontrivial test to determine the closeness of the reconstructed model to the true system and asymptotic scaling laws determine the uncertainty of the identified
parameters. Multistep integration schemes may be used to deal with low-resolution data.

Most of the aforementioned techniques deal with temporal systems—the underlying systems are modeled with ordinary differential equations (ODEs) that depend solely on a time variable. In this work we present a method for identifying spatiotemporal systems—the underlying systems are modeled with partial differential equations (PDEs) that depend on space as well as time variables. The difficulties in identifying PDEs in comparison with identifying ODEs are not only the computational cost, but also the algorithmic implementation, where boundary conditions, continuity constraints, and local minima problems need to be carefully treated during optimization procedures. The reports available on PDE identification problems may be classified as dynamical methods and regression methods. Dynamical methods are maximum likelihood estimators based on reducing subsets of piecewise continuous splines to best fit the full trajectory of experimental data. They are able to circumvent problems of local minima and to work with relatively large noise contaminations, but they exact high computational costs (maybe a thousand times that of the regression methods) and could be numerically sensitive to the initial parameter guess. Regression methods are of low computational cost and have straightforward implementations—centered forms of finite differences are traditionally used to approximate spatial and temporal derivatives from data to obtain the regression form—but they inflict relatively tight conditions on data resolution and clearness.

Most of the dynamical or regression methods tackle a parameter identification problem for a system of known governing equations with unknown parameters. However, in practice it is unlikely to know the exact governing equations. Here we reconstruct spatiotemporal systems without knowing the exact functional form of the systems. The model structure considered for the systems is a nonlinear PDE of polynomial type with an arbitrary nonlinearity degree and derivative order.

Our main goal is to introduce spectral differentiation operators into the realm of inverse problems (i.e., system identification) and to highlight the advantages. The key idea we develop is why and how to employ spectral differentiation to approximate the derivatives required in system identification. We show that this is more accurate than using finite differences especially when dealing with high order derivatives for which applying finite difference weights can be numerically ill conditioned, leading to severe cancellations and loss of significant digits in derivative approximations. Superior accuracy of spectral differentiation allows a coarser set of samples to be used in system identification, which is especially meaningful for reducing the cost in real experiments.

Our second goal is to introduce a customized structure selection procedure. For a model large enough to contain the true description of the system, when the number of parameters increases, so does the expected prediction error variance. Therefore, there is a strict penalty associated with the use of superfluous parameters. Hence, the model must be as parsimonious as possible. The proposed structure selection procedure uses an orthogonal decomposition and achieves parsimony by detecting and eliminating redundant terms from the model based on their insufficient contribution to the reduction of the mean squared error (MSE) decided by comparison with an error reduction ratio (ERR) threshold. This redundant-term elimination results in a reduced-order model which, compared with the full-order model, redeems two merits, in addition to parsimony: it has predefined fixed-point configuration, and its attractors are topologically closer to the original ones.

Finally, as a numerical experiment with the proposed “spectral identification” and “orthogonal structure selection,” the reconstruction of the Kuramoto–Sivashinsky (KS) PDE (a nonlinear highly stiff reaction-diffusion system) is presented and the quality of the model is examined through the analysis of estimate errors.

II. MODEL STRUCTURE

A model describes the relation between observations and allows for predictions. We assume a polynomial structure for our models that is general enough to be able to represent the true spatiotemporal system to be reconstructed. Specifically, we introduce the monomial \( x^{p_1} u^{p_2} (\partial u / \partial x)^{p_3} (\partial^2 u / \partial x^2)^{p_4} \cdots (\partial^{p_k} u / \partial x^{p_k})^{p_{k+2}} \) and take the model structure to be a Volterra expansion, using a linear combination of the aforementioned monomials:

\[
\frac{\partial^n u(t,x)}{\partial t^m} = \sum_{j=1}^{C} \alpha_j x^{p_1} u^{p_2} (\partial u / \partial x)^{p_3} (\partial^2 u / \partial x^2)^{p_4} \cdots (\partial^{p_k} u / \partial x^{p_k})^{p_{k+2}},
\]

for all \( p_j \in \mathbb{Z}^{\geq 0}, \ i \in \{1, \ldots, n+2\}, \) that \( \sum_{j} p_j = P \).

Equation (1) is of order \( m \) in time derivative and \( n \) in spatial derivatives. The monomials on its right-hand side are formed by taking the product of powers of the spatial variables \( x \), the dynamical variable \( u \), and the spatial derivatives of \( u \); the powers \( p_j \) are, by definition, non-negative integers and for each term they should add to at most the degree of nonlinearity \( P \) we assume for the model structure. The weights \( \alpha_j \) of this linear combination (the coefficients of the polynomial expansion) are the parameters to be estimated. Once we assume the highest order of spatial derivatives \( n \) and the degree of nonlinearity of the model structure \( P \), the number of terms \( C \) in Eq. (1), which is the dimension of the set of monomials used, is fixed, and equals \( (P+n+2)! / (P! (n+2)! \)).

To reconstruct a system with the structure of Eq. (1) from a time series, a key step is to approximate all the required derivatives from sampled data. Instead of traditional approximation methods based on finite differences, we propose spectral methods for this purpose which use spectral differentiation operators for approximating the derivatives. Spectral methods are based on special global interpolants, as opposed to finite difference/element methods which are based on local interpolants. These global interpolants lead to superior accuracy and less stringent requirements on the data resolution for spectral methods. Theoretical analysis shows that, in favorable conditions, spectral methods are of “infinite” accuracy and require much fewer samples to achieve a certain error tolerance than finite differences do. These fea-
We take grid points with grid function in and using a special nonuniform grid extension cannot help and using Fourier spectral methods. For nonperiodic domains, periodic differentiation is a linear operation in \( \mathcal{L}_N \); therefore, this operator has a corresponding matrix. To determine that matrix, we differentiate \( S_N(x) \),

\[
S_N'(x) = \frac{\cos(Nx/2)}{2\tan(x/2)} - \frac{\sin(Nx/2)}{2N\sin^2(x/2)},
\]

and evaluate \( S_N'(x) \) at the grid points:

\[
S_N'(x_j) = \begin{cases} 
0, & j = 0 \text{(mod} N) \\
\frac{1}{N}(1/2)(-1)^m \cot(jh/2), & j \neq 0 \text{(mod} N)
\end{cases}
\]

Therefore, the \((m, j)\) entry of \( D_N \) is

\[
(D_N)_{mj} = \begin{cases} 
0, & m = j \text{(mod} N) \\
\frac{1}{N}(1/2)(-1)^{m+j} \cot(x_m - x_j), & m \neq j \text{(mod} N)
\end{cases}
\]

Equation (8) shows that the operator \( D_N \) is a Toeplitz matrix depending only on \( m-j \) entries. In fact, it is more; it is circulant, since the entries depend on \((m-j)\text{mod} N\) only. A circulant matrix signals a convolution in a periodic domain, and is equivalent to a pointwise multiplication in the Fourier domain. Application of this special circulant matrix (Fourier spectral differentiation matrix) in the physical domain is equivalent to a multiplication by \( i k \) in the Fourier domain. Since the physical domain is periodic and discrete, the Fourier domain will be discrete and periodic; that is, the set of wavenumbers \( k \) will be discrete.

We take \( k=-N/2,-N/2+1,\ldots,N/2-1 \) as our fundamental set of wavenumbers. Using the connection between the two aforementioned spaces, the idea is to first transform the data \( v \) to the Fourier space by a discrete Fourier transform (DFT) to get \( \hat{v} \). Differentiation is done in that space and is simply a pointwise multiplication of \( \hat{v} \) by \( i k \). To have the derivative in the physical space, the result is then transformed back by an inverse DFT. Concisely,

\[
\hat{v}^{(n)} = F_N^{-1}((ik)^n \hat{v}) = \hat{v}^{(n)},
\]

where \( F_N \) and \( F_N^{-1} \) denote the DFT and inverse DFT matrices, respectively, and \( \hat{\times} \) and \( \times \) denote elementwise raising to power and multiplication. DFT and its inverse can be handled by fast Fourier transform algorithms efficiently. There is a subtle point here regarding the use of the fundamental wavenumbers for the differentiation in the Fourier domain. A special interpretation of the \( N/2 \) wavenumber is necessary when we deal with odd derivatives; specifically, it has to be set to 0, as at the grid points the modes \( \pm N/2 \) do not have any contributions to odd derivatives. Furthermore, the DFT matrix consists of the powers of the \( 1/N \)th root of unity: \( \exp(2\pi i jk/N) \). These exponentials, and therefore the DFT, are not changed if any integer multiple of \( N \) is added to \( k \). Thus, we may interpret \( \hat{v}_{-j} \) as \( \hat{v}_{j-1,2,\ldots,N/2-1} \) and write \( \hat{v}_{N/2} \) as \( \hat{v}_{-N/2} \), seeing it as coming from equal amounts...
of the modes $N/2$ and $-N/2$. Without this interpretation, the trigonometric polynomial takes complex values in between the real data points and its derivative becomes absurd.\textsuperscript{13} Putting these together, the elementwise multiplication should be done using

\begin{equation}
(f_k \cdot \mathring{\cdots} n_i = 0, \left(1, \ldots, \left(\frac{N}{2} - 1\right)\right) \cdot 0,
-((\frac{N}{2} - 1) \ldots, 1) \cdot n \quad \text{odd order deriv. } n),
\end{equation}

\begin{equation}
(f_k \cdot \mathring{\cdots} n_i = 0, \left(1, \ldots, \left(\frac{N}{2} - 1\right)\right) \cdot \frac{N}{2},
-((\frac{N}{2} - 1) \ldots, 1) \cdot n \quad \text{even order deriv. } n),
\end{equation}

where the operator $\cdot$ emphasizes that raising to power is implemented elementwise, the only way it makes sense for a vector.

It must be noted that the highest (in magnitude) mode attainable by the wavenumbers is

\begin{equation}
k_{\text{critical}} = (N/2) = \pi/h, \quad \text{or}
\end{equation}

\begin{equation}
f_{\text{critical}} = f_{\text{critical}}(2\pi) = 1/(2h) = (1/2)f_{\text{samples}},
\end{equation}

where $f_{\text{samples}} = 1/h$ and $f_{\text{critical}}$ is known as the Nyquist critical frequency. Equation (12) implies that the sampling frequency must be (at least) twice faster than the highest frequency present in the data; we must pick (at least) two samples in the smallest period present in the data. For a function that is not bandwidth limited to $|k| < k_{\text{critical}}$, any wavenumber component outside that band is, by the discrete Fourier transformation (DFT), indistinguishable from a wavenumber inside differing by an integer multiple of $N(=2\pi/h)$ and is, therefore, spuriously “folded over” into that band.\textsuperscript{37} This is known as the aliasing phenomenon and must be avoided by choosing $N$ large enough to give a large enough sampling frequency.

We now discuss the accuracy of the spectral differentiation and in particular see when and why it enjoys the so-called “infinite” accuracy. For a function with infinitely many continuous derivatives in $L^2$, Fourier spectral differentiation error is of order $O(N^{-m})$ as $N \to \infty$, for every $m \in \mathbb{R}_{\geq 0}$; as this holds for every $m$, it indicates that the error resulting from spectral differentiation, of a function with infinitely many continuous derivatives in $L^2$, dies faster than any fixed degree polynomial rate, hence the so-called “infinite” order of accuracy notion connected to spectral methods. On the other hand, using local methods (such as finite difference/element methods) the error would have a fixed degree polynomial decay rate. In other words, for local methods, the order of accuracy is some constant $m$ depending solely on the order of approximation and independent of the smoothness of the function whose derivatives are to be approximated. As spectral methods are global, they see “the bigger picture”; they look at the function as a whole and as a result their order of accuracy depends on the smoothness of the function they approximate.

If the function is further considered to be analytic in the complex strip $|\text{Im}(z)| \leq a$, and its $L^2$ norm along the horizontal line $\text{Im}(z) = y$ is uniformly bounded for all $y \in (-a,a)$, then the Fourier spectral differentiation error is of order $O(e^{-\pi(a-c)/h})$ as $h \to 0$, for every $\varepsilon > 0$ or $O(c^N)$ as $N \to \infty$, $(0 < c < 1)$; i.e., the error decays exponentially with the number of samples.\textsuperscript{41,44} We may justify these results, heuristically, by noting that a “smooth” function in the physical space decays rapidly in the frequency space because “smoothness” translates to slow changes in the former space and since large wavenumbers correspond to fast changes, such a function has “little” of them. As a result, for such a function, the discretization errors are “small” since they are caused by the aliasing of higher wavenumbers to lower ones.\textsuperscript{44} This may be further pinned down by noting that it is exactly the statement of the sampling theorem.\textsuperscript{37} A numerical experiment shows in Fig. 1 the considerable difference in accuracy, between finite differences and spectral methods, when approximating the first to fourth derivatives of $u(x) = \sin(x)\exp(\cos(x))$. The finite difference approximation employed in this experiment is centered and of fourth order.

**IV. LEAST SQUARES WITH QR DECOMPOSITION AND STRUCTURE SELECTION**

Now we can use spectral differentiation to approximate the spatial derivatives. Time derivatives can be approximated using finite differences; approximating time derivatives is not as critical as that of space derivatives since we are able to
make sensors sample data very densely in time. Dense spatial measurement, however, is usually infeasible for it requires a large number of sensors. Applying spectral differentiation in space and finite differences in time, and then picking the state of as many space points as required at one time point, Eq. (1) changes to a discrete regression form \( A_{N \times C} \hat{C}_{N \times 1} = b_{N \times 1} \), which is an overdetermined system of \( N \) linear equations in \( C \) unknowns, the parameters we seek. A vector of parameters \( \hat{\alpha} \) that minimizes the error \( \|e\|_2 = \|b - A\hat{\alpha}\|_2 \) is the optimal solution, since with proper statistical assumptions (zero mean white noise as the disturbance), it gives the maximum-likelihood estimate of the parameters. This is a least squares problem whose solution \( \hat{\alpha} \) is given by the normal equations \( A^T A \hat{\alpha} = A^T b \). Normal equations are always more badly conditioned than the original system since the condition number of \( A^T A \) is square of that of \( A \). Hence, we use an alternative approach and compute the parameter estimates by decomposing \( A \) into the product of a matrix \( Q \) with orthonormal columns and an upper triangular matrix \( R: A_{N \times C} = Q_{N \times N}(R_{C \times C}; 0) \), the so-called (full) QR decomposition.46

Orthogonal matrices are highly favorable in computations because of the crucial property that they do not change the norms (i.e., their condition number is 1) and hence they do not amplify errors. Triangular matrices are well known, among other things, for the straightforward solution of linear systems if they come up in the back substitution stage. Therefore, both factors of our decomposition have favorable properties and that justifies the computational cost we have to pay to get them. Using QR decomposition in linear least squares is due to Golub14 and the solution is obtained from

\[
Q\beta = b, \quad R\hat{\alpha} = \beta,
\]

in which the property \( Q^T Q = I \), due to the orthogonality of \( Q \), makes the first part already solved. It is noted that \( Q \) is not used in computing the parameter estimate and that using Householder reflections or Givens rotations, we are able to solve the least-squares problem using just the idea of QR, without actually computing \( Q \); this is in contrast to using Gram–Schmidt that explicitly computes \( Q \).31

The idea of structure selection (SS) finally chips in and is applied after the first part of Eq. (13). As \( Q \) has \( N \) orthonormal vectors, it forms a basis for \( \mathbb{R}^N \), so \( b \) is definitely in the column space of \( Q \); the first part has a unique solution. Since the basis vectors in \( Q \) are orthonormal, each component of the solution \( \beta_j \) is the contribution of the corresponding basis vector in a direction perpendicular to all others; no other \( \beta_{j \neq j} \) can contribute in that direction. Hence, the error reduction resulting from each parameter can be calculated separately.23 With no parameters, the maximum squared error would be incurred, equal to \( b^T b \). Inclusion of every \( \hat{\alpha}_j \) reduces this error by an amount \( \beta_{j1}^2 \). We define an error reduction ratio (ERR) equal to the ratio of the latter to the maximum:

\[
\text{ERR}_j = \frac{\beta_{j1}^2}{b^T b}.
\]

A threshold is then assumed for ERR\( _j \) below which the contribution of \( \hat{\alpha}_j \) to the error reduction is considered insignificant and allows us to omit the corresponding term. The estimates for this “reduced-dimension” regression are then recomputed, or, as is done in practice (for numerical efficiency), updated using the original estimates.46 This process is repeated until all remaining parameters give an ERR above the threshold. It should be pointed out that this structure selection procedure certainly yields a final parameter set with as many redundant parameters eliminated as possible but that set is not always “better” than a non-structure selected one in all senses. Eliminating \( \hat{\alpha}_j \)’s is equivalent to solving the original least-squares problem with some constraint equations. The original parameter estimates will be modified to take these constraints (eliminated parameters) into account. For finite differences, especially when the data have some error, the constrained parameter estimates (i.e., with structure selection) might have slightly more error than the unconstrained ones (i.e., with no structure selection). This happens because the remaining terms are not accurate enough to yield the least prediction error with parameter values close to the true ones; some undesired changes in the estimates with respect to the true parameter values is inevitable. An example of this phenomenon is discussed in more detail in the numerical experiment section.

V. NUMERICAL EXPERIMENTS ON IDENTIFICATION OF KS EQUATION

The KS equation was derived by Kuramoto and, independently, Sivashinsky; the former derived it in the context of angular-phase turbulence for a particular reaction-diffusion system24,25 and the latter in modeling small thermal diffusive instabilities in laminar flame fronts.68,39 It models pattern formations and is an instance of low-dimensional behavior in PDEs.21 It is characterized by the presence of coherent spatial structures and temporal chaos.34 We consider the identification of the following one-dimensional KS equation

\[
u_t = -u_{xx} + u - xu_{xxxx}, \quad (x,t) \in I \times \mathbb{R},
\]

\[
u(u,0) = u_0(x), \quad u(x + L, t) = u(x, t),
\]

where the subscripts denote partial differentiation, \( \varepsilon \) is a positive parameter and \( u_0 \) is of period \( L \) (the size of a typical pattern cell). In this particular example we pick \( \varepsilon = 1, I = [0,L], L = 32\pi, \) and \( u_0(x) = \cos(x/16)(1 + \sin(x/16)) \).

We compare the performance of system identification using finite differences and spectral methods. The required time derivative is approximated by centered differences of fourth order. The same approximation is applied to discretize the spatial derivatives in the finite differences case, too. For the spectral case, spatial derivatives are, of course, approximated using spectral differentiation. To obtain the regression form, the state of the system is then picked at \( N = 200 \) space points at a particular time point. We assume a degree of nonlinearity of \( P = 2 \) and a highest order of spatial derivative of \( n = 4 \). Therefore, there are a total number of \( C = 28 \) monomials in the model structure of Eq. (1) from which 25 monomials are redundant. All of the monomials, redundant or not, are listed in the first column of Table I. To generate the
Table I shows the comparison of the parameter estimates with and without structure selection with time steps of Δt=0.1 and Δt=0.01, using finite differences and spectral methods. Screening the results, we bring up three interesting points. (a) For both finite differences and spectral identifications, the technique of structure selection successfully eliminates most redundant monomials, making the model approach the true description of the system. More specifically, for the case of spectral identification, except for the term $u$, all redundant terms have been removed with the application of the 0.5% ERR threshold. (b) Whether employing the structure selection or not, all the test cases show that the spectral identifications outperform, significantly, finite difference identifications in terms of accuracy. (c) When the time step $\Delta t$ is chosen smaller, the parameter estimates generally become more accurate for the spectral method. However, for the finite differences case, a peculiar phenomenon was observed: the parameter estimates have no improvement for the case without structure selection (No SS), and become even worse after applying the structure selection (SS). For the case No SS, using finite differences, we believe estimate errors are induced mainly by rather inaccurate spatial derivative approximations, rather than time derivative approximations. For the case with SS, the deterioration, which we call “backfiring” phenomenon, is because the structure selection imposes additional constraints (namely, setting some of the parameters to zero) on the optimization of parameter estimates. These constraints leave fewer variables available for compensating the errors induced by spatial derivative approximations. This, in turn, results in some accuracy sacrifice in the true parameter estimates causing more errors in them than the unconstrained estimates (i.e., the case No SS), as mentioned before in the discussion of ERR threshold. Here, the key point is that finite differences cannot generate better spatial derivative approximations than spectral methods. The “backfiring” phenomenon in finite difference identifications persists in many of our test cases.

Another advantage of spectral methods, in inverse problems, is the less number of samples required for accurate identification. This is illustrated in Fig. 2(a) by the 2-norm of the estimate error for increasing number of samples whose states were picked at the 400th time point. Less error for the same number of samples and much faster decay of error are clearly seen for spectral methods. Also note that “paying” more samples to finite differences cannot “buy” the accuracy of spectral methods unless we are willing to pay a fortune.
Particularly in Fig. 2(a), the accuracy we get from spectral methods with 128 samples could not be caught up by finite differences even if we provided 200 samples. The choice of the time point at which to pick the state data can have a certain effect on the estimate error sense for both finite difference and spectral identifications, as illustrated in Fig. 2(a) state data picked at 400th time point and Fig. 2(b) state data picked at 200th time point. To find out if the time point choice could seriously affect the reconstruction results in the sense of the accuracy of the identified parameters, we investigated the variation of the accuracy of the identified parameters of the KS system with different time point choices via a systematic analysis for both methods. Figure 3 illustrates the effect of the time point choice on the relative error in the coefficients of the dominant terms, namely, $u_{xx}$, $u_{xxxx}$, and $uu_x$, of the reconstructed KS system. Figure 3(b) indicates that our proposed spectral identification is not “sensitive” to the time point choice as the relative error in the identified parameters does not exceed 0.05%. On the other hand, the finite difference identification is relatively sensitive to the time point choice, as the corresponding relative errors fluctuate as much as 26%, as shown in (b). However, finite difference identification is relatively sensitive to the time point choice, since the relative error in the identified parameters remains under 0.05%, as shown in (a).

In Fig. 4 we have highlighted the positive effect of the structure selection on the decrease of the estimate error as well as, again, the superiority of spectral identification. First of all, it should be noted that spectral identifications outdo finite difference identifications even if we ameliorate the latter with structure selection and do not apply that amelioration to the former. Second, when applied to spectral identifications, the structure selection decreases the estimate errors by several orders of magnitude. Lastly, a “backfiring” is noticed when the structure selection is applied to finite difference identifications. Structure-selected estimates coming from the finite difference case are actually worse than those with no structure selection because the terms surviving with structure selection and do not apply that amelioration to the former. Second, when applied to spectral identifications, the structure selection decreases the estimate errors by several orders of magnitude. Lastly, a “backfiring” is noticed when the structure selection is applied to finite difference identifications. Structure-selected estimates coming from the finite difference case are actually worse than those with no structure selection because the terms surviving with structure selection and do not apply that amelioration to the former. Second, when applied to spectral identifications, the structure selection decreases the estimate errors by several orders of magnitude. Lastly, a “backfiring” is noticed when the structure selection is applied to finite difference identifications. Structure-selected estimates coming from the finite difference case are actually worse than those with no structure selection because the terms surviving
In particular, we showed that, on periodic domains, Fourier

spectral methods yield a much more accurate solution for the parameter identification problem than finite differences do, even when we turn to higher order finite differences instead of the traditional centered second-order ones. To reach the same order of accuracy, spectral methods require much less number of samples than finite differences do. This advantage makes spectral methods of great practical significance as less number of sensors may be used with even higher identification accuracy. We also note that spectral methods cannot do magic: below a critical number of samples, depending on the highest frequency present in the system output, aliasing occurs and the identification would not be accurate.

In a robust test via a systematic study based on numerical experiments, the proposed spectral reconstruction was shown to be insensitive to the time point choice, while the finite differences could be relatively sensitive to that choice. Implementation of the proposed orthogonal structure selection technique allows us to reduce the order of the reconstructed model. Positive effects (as well as side effects, in the case of finite differences) of employing this structure selection were addressed. It was noted that as long as the related ERR threshold does not exceed a specific maximum value (contingent on the identification method and the system under reconstruction), most, if not all, of the redundant parameters can be eliminated from the model structure, resulting in a parsimonious model of the true system.

In this work we contributed a novel methodology for reconstructing spatiotemporal processes in order to boost this underdeveloped topic.19 The feasibility of the proposed spectral methods has been intensively verified in numerical experiments, but we note that the sample data used were “clean,” i.e., with no noise contamination. We fully understand the challenge of the global reconstructions from real experimental data reported in Refs. 27 and 28. Noise creates high frequency components in the data which might cause aliasing, leading to the erroneous identification of coefficients. Our preliminary studies based on numerical experiments (not reported in this manuscript) indicate that severe noise contamination can degrade the accuracy of spectral differentiation and cause relatively large deviation in the corresponding identification. We also remark that spectral methods are highly superior to finite differences in dealing with clean data but noise dilutes that superiority. A possible resolution in dealing with noise-contaminated data is to filter the unwanted high frequency components in data using, e.g., wavelet-based filters or other noise-filtering tools. The application of our proposed method to real experimental data is our next step in the extension of this research.

VI. CONCLUSION

In this work, we proposed Fourier spectral methods in conjunction with a structure selection procedure for reconstructing periodic spatiotemporal systems from time series. It was shown that the well-known accuracy of spectral methods for direct problems carries over to inverse problems as well. In particular, we showed that, on periodic domains, Fourier

FIG. 5. (Color online) Variation of estimate error with different ERR threshold selections for (a) finite difference and (b) spectral identifications (number of samples $N=200$, state data picked at 400th time point). The estimate error attains its minimum with ERR threshold in $[0.005, -0.01]$ for (a) and $[0.0035, -0.008]$ for (b). Reconstruction process may fail if the threshold is set in excess of a maximum threshold, 0.01 for (a) and 0.008 for (b). A logarithmic scale is used for the lower pane.