LBNL-42283
Preprint

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

On the Prediction of Large-Scale Dynamics Using Unresolved Computations

Alexandre J. Chorin, Anton P. Kast,
and Raz Kupferman

Computing Sciences Directorate
Mathematics Department

September 1998
To be submitted
for publication

REFERENCE COPY
Does Not
Circulate
Bldg. 50 Library - Ref.
Lawrence Berkeley National Laboratory
Copy 1
LBNL-42283

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

LBNL-42283

ON THE PREDICTION OF LARGE-SCALE DYNAMICS USING UNRESOLVED COMPUTATIONS*

Alexandre J. Chorin, Anton P. Kast, and Raz Kupferman
Lawrence Berkeley National Laboratory
and
Department of Mathematics
University of California
Berkeley, CA 94720, USA

September 1998

*This work was supported in part by the Office of Energy Research, Office of Computational and Technology Research, Mathematical, Information, and Computational Sciences Division, Applied Mathematical Sciences Subprogram, of the U.S. Department of Energy, under Contract No. DE-AC03-76SF00098, and by the National Science Foundation under Grant DMS94-14631.

Contemporary Mathematics

On the Prediction of Large-Scale Dynamics using Unresolved Computations

Alexandre J. Chorin, Anton P. Kast, and Raz Kupferman

ABSTRACT. We present a theoretical framework and numerical methods for predicting the large-scale properties of solutions of partial differential equations that are too complex to be properly resolved. We assume that prior statistical information about the distribution of the solutions is available, as is often the case in practice. The quantities we can compute condition the prior information and allow us to calculate mean properties of solutions in the future. We derive approximate ways for computing the evolution of the probabilities conditioned by what we can compute, and obtain ordinary differential equations for the expected values of a set of large-scale variables. Our methods are demonstrated on two simple but instructive examples, where the prior information consists of invariant canonical distributions

1. Introduction

There are many problems in science that can be modeled by a set of differential equations, but where the solution of these equations is so complicated that it cannot be found in practice, either analytically or numerically. For a numerical computation to be accurate the problem must be well resolved, i.e, enough variables (or "degrees of freedom") must be represented in the calculation to capture all the relevant features of the solution; insufficient resolution yields sometimes disastrous results. A well-known example in which good resolution cannot be achieved is turbulent flow, where one has to resolve all scales ranging from the size of the system down to the dissipation scale—a prohibitively expensive requirement. One is then compelled to consider the question of how to predict complex behavior when the number of variables that can be used in the computation is significantly less than needed for full resolution. This is the question considered in the present paper; part of the theoretical framework and methods have already been briefly discussed in [**CKK98**].

Studies on underresolved problems exist in a wide range of different contexts, along with a large amount of literature that describes problem-specific methods. In

1991 *Mathematics Subject Classification.* Primary 65M99.

This work was supported in part by the US Department of Energy under contract DE-AC03-76-SF00098, and in part by the National Science Foundation under grant DMS94-14631.

©0000 (copyright holder)

turbulence, for example, there are various modeling methods for large eddy simulations. In all cases one needs to make additional assumptions about the relation between those degrees of freedom that are represented in the computation and the "hidden", or "invisible" degrees of freedom that are discarded from the computation. A number of interesting attempts have been made over the years to fill in data from coarse grids in difficult computations so as to enhance accuracy without refining the grid (see e.g. [**SM97, MW90**]). Indeed, nothing can be done without some information regarding the unresolved degrees of freedom. Such additional assumptions are usually motivated by intuitive reasoning and their validity is usually assessed by comparing the resulting predictions to experimental measurements.

In many problems the lack of resolution is due primarily to the insufficiency and sometimes also the inaccuracy of the measurements that provide initial conditions for the system of equations. This is the case for example in weather forecasting, where the initial information consists of local weather measurements collected at a relatively small number of meteorological stations. The problem of insufficient and sometimes noisy data is not considered in the present paper. We focus here on the case where underresolution is imposed by computational limitations. Initial data will be assumed to be available at will, and this assumption will be fully exploited by allowing us to select the set of degrees of freedom that are represented in the computation at our convenience. Another issue that often arises in the modeling of complex systems is uncertainty regarding the equations themselves. This important question is also beyond the scope of this paper; the adequacy of the system of equations to be solved is taken for granted.

We now define the problem and introduce some of the nomenclature: We consider a system described by a differential equation of the form

$$u_t = F(u), \tag{1.1}$$

where t is time, subscripts denote differentiation, $u(x,t)$ is the dependent variable, and $F(u) = F(u, u_x, u_{xx}, \ldots)$ is a (generally nonlinear) function of its arguments; the spatial coordinate x and the dependent variable u can be of arbitrary dimensionality.

To solve an equation of the form (1.1) on a computer one ordinarily discretizes the dependent variable $u(x,t)$ both in space and time and replaces the differential equation by an appropriate relation between the discrete variables. As described, the solution to the discrete system may approximate the solution of the differential equation well only if the discretization is sufficiently refined. It is our basic assumption that we cannot afford such a refined discretization, and must therefore be content with a much smaller number of variables. One still has the liberty to choose the degrees of freedom that are retained in the computation; those will be chosen, for convenience, to be linear functionals of the dependent variable $u(x,t)$:

$$U_\alpha[u(\cdot,t)] \equiv (g_\alpha(\cdot), u(\cdot,t)) \equiv \int g_\alpha(x) u(x,t)\, dx, \tag{1.2}$$

where α is an index that enumerates the selected degrees of freedom. Variables of the form (1.2) will be referred to as *collective variables*; every collective variable U_α is defined by a *kernel* $g_\alpha(x)$. Point values of $u(x)$ at a set of points x_α, and spectral components of $u(x)$ for a set of modes k_α are two special cases of collective variables; in the first case the corresponding kernels are delta functions, $g_\alpha(x) = \delta(x - x_\alpha)$, whereas in the second case the kernels are spectral basis functions, $\exp(ik_\alpha \cdot x)$. We

assume that our computational budget allows us to operate on a set of at most N collective variables, so that $\alpha = 1, \ldots, N$. The question is, what can be predicted about the state of the system at a future time t given the values of the collective variables U_α at an initial time $t = 0$?

Suppose that we know at time $t = 0$ that the collective variables U_α assume a set of values V_α. (We will denote by $U = (U_1, \ldots, U_N)^T$ and $V = (V_1, \ldots, V_N)^T$ the vectors whose entries are the collective variables and their initial values, respectively.) Our postulate that the number of collective variables N does not suffice to resolve the state of the system implies that the initial data, V, do not determine sharply enough the initial condition, $u(x, 0)$. A priori, every function $u(x, 0)$ that is compatible with the given values of the collective variables, that is, belongs to the set

$$\mathcal{M}(V) = \{v(x) : U_\alpha[v(\cdot)] = V_\alpha, \quad \alpha = 1, \ldots, N\}. \tag{1.3}$$

is a plausible initial condition. One could define underresolution in terms of the set of functions (1.3); the problem is underresolved if this set is non-trivial. Clearly, the state of the system at future times depends on the particular initial condition; in many cases it is even very sensitive to small variations in the initial condition. One wonders then in what sense the future can be predicted when the initial condition is not known with certainty.

The essence of our approach is the recognition that underresolution necessarily forces one to consider the evolution of a set, or ensemble, of solutions, rather than a single initial value problem. This requires the replacement of equation (1.1) by a corresponding equation for a probability measure defined on the space of the solutions of (1.1). The prediction of the future state of the system can then be reinterpreted as the prediction of most likely, or mean, properties of the system. Loosely stated, in cases where sufficient resolution cannot be achieved the original task of solving an initial value problem has to be replaced by a more modest one—the determination of "what is most likely to happen given what is initially known."

At first, there seems to be no practical progress in the above restatement of the problem. First, the statistical problem also requires initial conditions; a measure defined on the space of initial conditions $u(x, 0)$ must be provided for the statistical problem to be well-defined. Second, the high-dimensional Liouville equation that describes the flow induced by (1.1) is not easier to solve than the original initial value problem. It turns out that in many problems of interest there exists a natural measure μ that characterizes the statistical properties of the system; what is meant by "natural" has to be clarified; an important class of such measures are invariant ones. We are going to use this information to partially cure the two aforementioned difficulties: First, this measure will define the initial statistical state of the system by being interpreted as a "prior" measure—a quantification of our beliefs regarding the state of the system prior to the specification of any initial condition. The initial values of the collective variables are constraints on the set of initial states and induce on μ a conditional measure that constitutes an initial condition for the Liouville equation. Second, the existence of a distinguished statistical measure suggests a way to generate a hierarchy of approximations to the Liouville equation, examples of which will be described in the following sections.

The rest of this paper is organized as follows: In Section 2 we present our theory, and provide a recipe (2.11) for approximating the mean evolution of a set

of collective variables. In Section 3 we derive formulas for the calculation of conditional expectations in the case of Gaussian prior measures; these are necessary for the evaluation of the right-hand side of equation (2.11). In Sections 4 and 5 we demonstrate the power of our theory by considering two examples: a linear Schrödinger equation and a nonlinear Hamiltonian system. Conclusions are presented in Section 6.

2. Presentation of the theory

Our starting point is a general equation of motion of the form (1.1), and a set of collective variable U_α defined by (1.2) for a set of kernels $g_\alpha(x)$; the question of what constitutes a good choice of kernels will be discussed below.

In many problems of interest there exists a measure on the space of solutions of (1.1) that is invariant under the flow induced by (1.1); a measure that has this property is referred to as an *invariant measure*. Invariant measures are known to play a central role in many problems; macroscopic systems (that is, systems that have a very large number of degrees of freedom) whose macroscopic properties do no change in time, often exhibit an invariant statistical state. By that we mean the following: when the large scale observable properties of the system remain constant in time, the likelihood of the microscopic degrees of freedom to be in any particular state is distributed according to a measure that is invariant in time. We will assume that such an invariant measure μ_0 exists and that we know what it is. The measure μ_0 will then be postulated to be the *prior measure*, i.e, it describes the probability distribution of initial conditions before any measurement has been performed. We will denote averages with respect to the invariant measure μ_0 by angle brackets $\langle\cdot\rangle$; let $O[u(\cdot)]$ be a general functional of u, then

$$\langle O\rangle = \int O[u(\cdot)]\, d\mu_0, \tag{2.1}$$

where the integration is over an appropriate function space. We shall write formally,

$$d\mu_0 = f_0[u(\cdot)]\,[du], \tag{2.2}$$

as if the measure μ were absolutely continuous with respect to a Lebesgue measure, where $f_0[u]$ is the invariant probability density, and $[du]$ is a formal product of differentials.

We next assume that a set of measurements has been carried out and has revealed the values V_α of the collective variables U_α at time $t = 0$. This information can be viewed as a set of constraints on the set of initial conditions, which is now given by (1.3). Constraints on the set of functions $u(x)$ automatically induce on μ_0 a *conditional measure*, which we denote by μ_V. In a physicist's notation,

$$d\mu_V = f_V[u(\cdot)]\,[du] = c f_0[u(\cdot)]\,[du] \times \prod_{\alpha=1}^{N} \delta\left(U_\alpha[u(\cdot)] - V_\alpha\right), \tag{2.3}$$

where $f_V[u(\cdot)]$ is the conditional probability density, and c is an appropriate normalization factor. The conditional probability density is equal, up to a normalization, to the prior probability density projected on the space of functions $\mathcal{M}(V)$ that are compatible with the initial data. Note that the conditional measure μ_V is, in general, not invariant. Averages with respect to the conditional measure will be

denoted by angle brackets with a subscript that symbolizes the constraints imposed on the set of functions,

$$\langle O \rangle_V \equiv \int O[u(\cdot)] f_V[u(\cdot)]\,[du]. \tag{2.4}$$

The dynamics have not been taken into consideration so far, except for the fact that the measure μ_0 was postulated to be invariant. Let $f[u(\cdot),t]$ be the probability density of the solutions of (1.1) at time t, that is, the probability density that evolves from the initial probability density $f_V[u(\cdot)]$ under the flow induced by (1.1); it satisfies the Liouville equation [**Ris84**]

$$f_t + \left(\frac{\delta f}{\delta u}(\cdot), F(u(\cdot))\right) = 0, \tag{2.5}$$

where $\frac{\delta f}{\delta u}$ denotes a functional derivative. An equivalent statement is that if S_t denotes the time evolution operator induced by (1.1), i.e., $S_t : u(x,0) \longrightarrow u(x,t)$, then

$$f[u(\cdot),t] = f[S_t^{-1}u(\cdot),0] = f_V[S_t^{-1}u(\cdot)], \tag{2.6}$$

where S_t^{-1} is the operator inverse to S_t, which we assume to exist.

The objective that has been defined in the introductory section is to calculate the expectation value of observables $O[u(\cdot)]$ at time t, given the initial data V. In terms of the notations introduced above this is given by

$$\langle O[u(\cdot),t]\rangle_V = \langle O[S_t u(\cdot)]\rangle_V \tag{2.7}$$

(operators are generally treated as function of the dependent variable and time, $O[u(\cdot),t]$; when no reference to time is being made the expression refers to the initial time).

We next make the following observations: (i) The initial probability measure (2.3) is completely determined by the N numbers V_α. (ii) By the invariance of $f_0[u]$ and by equation (2.6), the probability density at later time t can still be represented as the invariant density projected on a set of N conditions; specifically,

$$f[u(\cdot),t] = c\, f_0[u(\cdot)] \prod_{\alpha=1}^{N} \delta\left[\left(g_\alpha(\cdot), S_t^{-1}u(\cdot)\right) - V_\alpha\right]. \tag{2.8}$$

Note however that the set of functions that support this measure at time t is generally not of the form (1.3), that is, the observable $\left(g_\alpha(\cdot), S_t^{-1}u(\cdot)\right)$ is not a linear functional of u.

These observations suggest an approximate procedure for solving the Liouville equation (2.5). We propose an ansatz in which the N conditions that are imposed on μ_0 remain for all times conditions on the values of the collective variables U; namely, the probability density is specified by a time-dependent vector of N numbers $V_\alpha(t)$, such that

$$f[u(\cdot),t] \approx c\; f_0[u(\cdot)] \prod_{\alpha=1}^{N} \delta\left[U_\alpha[u(\cdot)] - V_\alpha(t)\right]. \tag{2.9}$$

One has still to specify the time evolution of the vector $V(t)$. Suppose that the distribution of solutions is indeed given by (2.9) at time t, and consider a later time $t + \Delta t$. The value of the observable $U_\alpha[u(\cdot)]$ at the later time will, in general, not be uniform throughout the ensemble of solutions. The ansatz (2.9) projects

the distribution back onto a set of solutions $\mathcal{M}(V(t+\Delta t))$. A natural choice for $V_\alpha(t+\Delta t)$ is the expectation value of the collective variable $U_\alpha[u(\cdot)]$ given that the distribution at time t was (2.9):

$$\begin{aligned} V_\alpha(t+\Delta t) &\approx \langle U_\alpha[u(\cdot), t+\Delta t]\rangle_{V(t)} = \\ &= \langle U_\alpha[u(\cdot)]\rangle_{V(t)} + \Delta t\ \langle (g_\alpha(\cdot), F(u(\cdot)))\rangle_{V(t)} + O(\Delta t^2). \end{aligned} \tag{2.10}$$

Taking the limit $\Delta t \to 0$ we finally obtain,

$$\frac{dV_\alpha}{dt} = \langle (g_\alpha(\cdot), F(u(\cdot)))\rangle_{V(t)}\,. \tag{2.11}$$

Equation (2.11) is our main tool in the present paper and we will next discuss its implications:

- Equation (2.11) constitutes a closed set of N ordinary differential equations, which by our postulate is within the acceptable computational budget.
- The central hypothesis in the course of the derivation was that the distribution of solutions can be approximated by (2.9). This approximation assumes that for all times t the collective variable U_α has a uniform value V_α *for all the trajectories in the ensemble of solutions.* This assertion is initially correct (by construction) at time $t = 0$, but will generally not remain true for later times. The approximation is likely to be a good one as long as the above assertion is approximately true, that is, as long as the distribution of values assumed by the collective variables remains sufficiently narrow. In many cases it is possible to guarantee a small variance by a clever selection of collective variables (i.e., of kernels). Note furthermore that the smallness of the variance can be verified self-consistently from the knowledge of the probability density (2.9).
- Equation (2.11) still poses the technical problem of computing its right-hand side. This issue is the subject of the next section.
- The case where the equations of motion (1.1) are linear, i.e,

$$u_t = Lu, \tag{2.12}$$

with L being a linear operator, can be worked out in detail. Using the fact that $S_t = \exp(Lt)$, the solution to the Liouville equation (2.8) can be rearranged as

$$f[u(\cdot), t] = c f_0[u(\cdot)] \prod_{\alpha=1}^{N} \delta\left[\left(e^{-L^\dagger t} g_\alpha(\cdot), u(\cdot)\right) - V_\alpha\right], \tag{2.13}$$

where $L^\dagger$ is the linear operator adjoint to L. Thus, the probability density for all times is f_0 projected on the set of functions for which a set of N linear functionals of u have the values V; note that V here is not time dependent, but is the vector of initial values of the collective variables U. The kernels that define these functionals are time dependent, and evolve according to the dual equation

$$\frac{dg_\alpha}{dt} = -L^\dagger g_\alpha. \tag{2.14}$$

If the kernels g_α are furthermore eigenfunctions of the dual operator $L^\dagger$ with eigenvalues λ_α, the ansatz (2.9) is exact, with $V_\alpha(t) = V_\alpha(0)\, e^{\lambda_\alpha t}$. Hald [**Hal**] shows that by selecting kernels that are *approximate* eigenfunctions of $L^\dagger$,

one can bound the error introduced by the ansatz (2.9), while retaining the simplicity of the procedure.

- The two alternatives of evolving either the values V_α or the kernels $g_\alpha(x)$ are analogous to Eulerian versus Lagrangian approaches in fluid mechanics, or Schrödinger versus Heisenberg approaches in quantum mechanics. For nonlinear equations one has a whole range of intermediate possibilities; for example one may split the operator F in equation (1.1) as $F = L+Q$, where L is linear. The kernels can be evolved according to the linear operator, while the values of the collective variables can be updated by the remaining nonlinear operator. The art is to find partitions $F = L + Q$ that minimize the variance of the distribution of values assumed by the collective variables.
- Equation (2.11) should be viewed as a first approximation to the solution of the Liouville equation, where the only information that is updated in time is the mean value of a fixed set of collective variables. In principle, one could also update higher moments of those variables, and use this additional information to construct a better approximation. For example, equipped with the knowledge of means and covariances one could find new kernels and new values for the corresponding collective variables, such that the distribution obtained by conditioning the invariant distribution with those new constraints is compatible with the calculated means and covariances. Thus, one could imagine an entire hierarchy of schemes that take into account an increasing number of moments of the resolved variables.

3. Conditional expectation with Gaussian prior

Equation (2.11) is a closed set of equations for the vector $V(t)$, which requires the computation of a conditional average on its right-hand side. To have a fully constructive procedure, we need to evaluate conditional averages $\langle O[u(\cdot)]\rangle_V$, where O is an arbitrary observable, and V denotes as before the vector of values of a set of collective variables U of the form (1.2). In this section we present three lemmas that solve this problem for the case where the prior measure μ_0 is Gaussian. In the two examples below, the prior measure is either Gaussian or can be viewed as a perturbation of a Gaussian measure.

The random function $u(x)$ has a Gaussian distribution if its probability density is of the form

$$f_0[u(\cdot)] = Z^{-1} \exp\left(-\frac{1}{2}\iint u(x)a(x,y)u(y)\,dx\,dy + \int b(x)u(x)\,dx\right), \tag{3.1}$$

where $a(x,y)$ and $b(x)$ are (generalized) functions, and Z is a normalizing constant. The functions $a(x,y)$ and $b(x)$ are related to the mean and the covariance of $u(x)$ by

$$\langle u(x)\rangle = \left(a^{-1}(x,\cdot), b(\cdot)\right), \tag{3.2}$$

and

$$\mathrm{Cov}\,[u(x),u(y)] \equiv \langle u(x)u(y)\rangle - \langle u(x)\rangle\,\langle u(y)\rangle = a^{-1}(x,y), \tag{3.3}$$

where the generalized function $a^{-1}(x,y)$ is defined by the integral relation

$$\left(a(x,\cdot), a^{-1}(\cdot,y)\right) = \left(a^{-1}(x,\cdot), a(\cdot,y)\right) = \delta(x-y). \tag{3.4}$$

To compute the expectation value of higher moments of u one can use Wick's theorem [**Kle89**]:

$$(3.5)\quad \langle (u_{i_1} - \langle u_{i_1}\rangle)\cdots(u_{i_l} - \langle u_{i_l}\rangle)\rangle = \begin{cases} 0, & l \text{ odd} \\ \sum \mathrm{Cov}\left[u_{i_{p_1}}, u_{i_{p_2}}\right]\cdots \mathrm{Cov}\left[u_{i_{p_{l-1}}}, u_{i_{p_l}}\right], & l \text{ even} \end{cases},$$

with summation over all possible pairings of $\{i_1, \ldots, i_l\}$.

Next, suppose that the random function $u(x)$ is drawn from a Gaussian distribution, and a set of measurements reveal the vector of values V for a set of collective variables U of the form (1.2). This information changes the probability measure μ_0 into a conditional measure μ_V with density f_V given by (2.3). Conditional averages of operators $O[u(\cdot)]$ can be calculated by using the following three lemmas:

LEMMA 3.1. *The conditional expectation of the function $u(x)$ is a linear form in the conditioning data V:*

$$(3.6)\qquad \langle u(x)\rangle_V = \langle u(x)\rangle + \sum_{\alpha=1}^{N} c_\alpha(x)\left\{V_\alpha - \langle U_\alpha[u(\cdot)]\rangle\right\},$$

where the vector of functions $c_\alpha(x)$ is given by

$$(3.7)\qquad c_\alpha(x) = \sum_{\beta=1}^{N} \left(a^{-1}(x,\cdot), g_\beta(\cdot)\right) m^{-1}_{\beta\alpha},$$

and where the $m^{-1}_{\beta\alpha}$ are the entries of an $N\times N$ matrix M^{-1} whose inverse M has entries

$$(3.8)\qquad m_{\beta\alpha} = Cov[U_\beta[u(\cdot)], U_\alpha[u(\cdot)]] = \iint g_\beta(x) a^{-1}(x,y) g_\alpha(y)\, dx\, dy.$$

PROOF. Given the prior measure μ_0 and the values V of the collective variables U, we define a *regression function* (an approximant to $u(x)$) of the form

$$(3.9)\qquad R(x) = \sum_{\alpha=1}^{N} r_\alpha(x) V_\alpha + s(x),$$

where the functions $r_\alpha(x)$ and $s(x)$ are chosen such to minimize the mean square error,

$$(3.10)\qquad E(x) = \langle e^2(x)\rangle \equiv \left\langle \left[u(x) - \sum_{\alpha=1}^{N} r_\alpha(x) U_\alpha[u(\cdot)] - s(x)\right]^2 \right\rangle.$$

for all x. Note that this is an *unconditional* average with respect to μ_0.

Minimization with respect to $s(x)$ implies that

$$(3.11)\qquad \frac{\partial E(x)}{\partial s(x)} = \langle e(x)\rangle = \left\langle u(x) - \sum_{\alpha=1}^{N} r_\alpha(x) U_\alpha[u(\cdot)] - s(x)\right\rangle = 0,$$

which, combined with (3.9), yields

$$(3.12)\qquad R(x) = \langle u(x)\rangle + \sum_{\alpha=1}^{N} r_\alpha(x)\left\{\langle U_\alpha[u(\cdot)]\rangle - V_\alpha\right\}.$$

Minimization with respect to $r_\alpha(x)$ implies:

(3.13)
$$\frac{\partial E(x)}{\partial r_\alpha(x)} = \langle e(x)\, U_\alpha[u(\cdot)]\rangle = \left\langle \left[u(x) - \sum_{\beta=1}^{N} r_\beta(x) U_\beta[u(\cdot)] - s(x) \right] U_\alpha[u(\cdot)] \right\rangle = 0.$$

Equation (3.13) can be rearranged by substituting equations (3.3) and (3.11) into it, and using the fact that $U_\alpha[u(\cdot)] = (g_\alpha(\cdot), u(\cdot))$:

$$\sum_{\beta=1}^{N} \mathrm{Cov}\left[U_\alpha[u(\cdot)], U_\beta[u(\cdot)]\right] r_\beta(x) = \left(g_\alpha(\cdot), a^{-1}(x,.)\right). \tag{3.14}$$

One readily identifies the functions $r_\alpha(x)$ as satisfying the definition (3.7) of the functions $c_\alpha(x)$. Comparing (3.12) with (3.6), the regression function is nothing but the right-hand side of equation (3.6).

It remains to show that the regression curve equals also the left-hand side of (3.6). Consider equation (3.13): it asserts that the random variable $e(x)$ is statistically orthogonal to the random variables $U_\alpha[u(\cdot)]$. Note that both $e(x)$ and the collective variables U_α are linear functionals of the Gaussian function $u(x)$, and are therefore jointly Gaussian. Jointly Gaussian variables that are statistically orthogonal are independent, hence, the knowledge of the value assumed by the variables $U_\alpha[u(\cdot)]$ does not affect the expectation value of $e(x)$,

(3.15)
$$\left\langle u(x) - \sum_{\alpha=1}^{N} r_\alpha(x) U_\alpha[u(\cdot)] - s(x) \right\rangle_V = \left\langle u(x) - \sum_{\alpha=1}^{N} r_\alpha(x) U_\alpha[u(\cdot)] - s(x) \right\rangle.$$

The function $s(x)$ is not random and $\langle U_\alpha[u(\cdot)]\rangle_V = V_\alpha$, from which immediately follows that

$$\langle u(x)\rangle_V = \langle u(x)\rangle + \sum_{\alpha=1}^{N} r_\alpha(x) \left\{V_\alpha - \langle U_\alpha[u(\cdot)]\rangle\right\}, \tag{3.16}$$

This completes the proof. □

LEMMA 3.2. *The conditional covariance of the function $u(x)$ differs from the unconditional covariance by a function that depends on the kernels $g_\alpha(x)$, without reference to the conditioning data V:*

$$Cov[u(x), u(y)]_V = Cov[u(x), u(y)] - \sum_{\alpha=1}^{N} c_\alpha(x) \left(g_\alpha(\cdot), a^{-1}(\cdot, y)\right). \tag{3.17}$$

PROOF. The proof follows the same line as the second part of the proof of Lemma 3.1. Consider the following expression:

(3.18)
$$e(x)e(y) = \left[u(x) - \sum_{\alpha=1}^{N} r_\alpha(x) U_\alpha[u(\cdot)] - s(x) \right] \left[u(y) - \sum_{\beta=1}^{N} r_\beta(y) U_\beta[u(\cdot)] - s(y) \right].$$

Both $e(x)$ and $e(y)$ are independent of the collective variables U. It is always true that if A_1, A_2 and A_3 are random variables with A_3 being independent of A_1 and

A_2, then $\langle A_1 A_2 \rangle_{A_3} = \langle A_1 A_2 \rangle$. Hence,

$$\langle e(x)e(y)\rangle_V = \langle e(x)e(y)\rangle \,, \tag{3.19}$$

from which (3.17) follows after straightforward algebra. □

LEMMA 3.3. *Wick's theorem extends to conditional expectations:*

$$\left\langle (u_{i_1} - \langle u_{i_1}\rangle_V) \cdots (u_{i_l} - \langle u_{i_l}\rangle_V) \right\rangle_V = \begin{cases} 0, & l \text{ odd} \\ \sum Cov\left[u_{i_{p_1}}, u_{i_{p_2}}\right]_V \cdots Cov\left[u_{i_{p_{l-1}}}, u_{i_{p_l}}\right]_V, & l \text{ even} \end{cases}, \tag{3.20}$$

where again the summation is over all possible pairings of $\{i_1, \ldots, i_l\}$.

PROOF. Using the fact that a delta function can be represented as the limit of a narrow Gaussian function, the conditional expectation of any list of observables, $O_1[u(\cdot)], \ldots, O_p[u(\cdot)]$, can be expressed as

$$\langle O_1[u(\cdot)] \cdots O_p[u(\cdot)]\rangle_V = \lim_{\Delta \to 0} \int O_1[u(\cdot)] \cdots O_p[u(\cdot)]\, f_V^{\Delta}[u(\cdot)][du], \tag{3.21}$$

where

$$f_V^{\Delta}[u(\cdot)] = c_\Delta\, f_0[u(\cdot)] \prod_{\alpha=1}^{N} \frac{1}{\sqrt{\pi}\Delta} \exp\left[-\frac{(U_\alpha[u(\cdot)] - V_\alpha)^2}{\Delta^2}\right], \tag{3.22}$$

the coefficient c_Δ is a normalization, and the order of the limit $\Delta \to 0$ and the functional integration has been interchanged. Note that the exponential in (3.22) is quadratic in $u(x)$, hence the finite-Δ probability density $f_V^{\Delta}[u(\cdot)]$ is Gaussian, Wick's theorem applies, and the limit $\Delta \to 0$ can finally be taken. □

The conditional expectation of any observable $O[u(\cdot)]$ can be deduced, in principle, from a combination of Lemmas 3.1-3.3.

In the examples considered below, the dependent variable $u(x,t)$ is a vector; let $u^i(x,t)$ denote the i'th component of the d-dimensional vector $u(x,t)$. All the above relations are easily generalized to the vector case. To keep notations as clear as possible, we denote indices associated with the collective variables by Greek subscripts, and indices associated with the components of u by Roman superscripts. The probability density $f_0[u(\cdot)]$ is Gaussian if it is of the following form,

$$f_0[u(\cdot)] = \frac{1}{Z} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{d} \iint u^i(x) a^{ij}(x,y) u^j(y)\, dx\, dy + \sum_{i=1}^{d} \int b^i(x) u^i(x)\, dx\right), \tag{3.23}$$

where $a^{ij}(x,y)$ are now the entries of a $d \times d$ matrix of functions, and $b^i(x)$ are the entries of a vector of functions. These functions are related to the mean and the covariance of the vector $u(x)$ by

$$\langle u^i(x)\rangle = \sum_{j=1}^{d} \left([a^{-1}(x,\cdot)]^{ij}, b^j(\cdot)\right), \tag{3.24}$$

and

$$\mathrm{Cov}\left[u^i(x), u^j(y)\right] = [a^{-1}(x,y)]^{ij}, \tag{3.25}$$

where $[a^{-1}(x,y)]^{ij}$ is defined by

$$\sum_{j=1}^{d} \left([a^{-1}(x,\cdot)]^{ij}, a^{jk}(\cdot,y)\right) = \delta(x-y)\,\delta_{ik}. \tag{3.26}$$

Suppose now that a set of measurements reveals the values V_α^i of a matrix of collective variables of the form,

$$U_\alpha^i[u(\cdot)] = \left(g_\alpha(\cdot), u^i(\cdot)\right), \tag{3.27}$$

where $\alpha = 1,\ldots,N$ and $i = 1,\ldots,d$. The conditional expectation and covariance of $u^i(x)$ are given by straightforward generalizations of Lemmas 3.1 and 3.2:

$$\langle u^i(x)\rangle_V = \langle u^i(x)\rangle + \sum_{\alpha=1}^{N}\sum_{j=1}^{d} c_\alpha^{ij}(x)\left\{V_\alpha^j - \langle U_\alpha^j[u(\cdot)]\rangle\right\}, \tag{3.28}$$

and

(3.29)

$$\mathrm{Cov}\left[u^i(x), u^j(y)\right]_V = \mathrm{Cov}\left[u^i(x), u^j(y)\right] - \sum_{\alpha=1}^{N}\sum_{k=1}^{d} c_\alpha^{ik}(x)\left(g_\alpha(\cdot), [a^{-1}(\cdot,y)]^{kj}\right).$$

where

$$c_\alpha^{ij}(x) = \sum_{\beta=1}^{N}\sum_{k=1}^{d} \left([a^{-1}(x,\cdot)]^{ik}, g_\beta(\cdot)\right)[m^{-1}]_{\beta\alpha}^{kj}, \tag{3.30}$$

and where the $[m^{-1}]_{\beta\alpha}^{ij}$ are the entries of an $N\times N\times d\times d$ tensor M^{-1} whose inverse M has entries

$$m_{\beta\alpha}^{ij} = \iint g_\beta(x)[a^{-1}(x,y)]^{ij} g_\alpha(y)\,dx\,dy. \tag{3.31}$$

4. A linear Schrödinger equation

The equations of motion. The first example is a linear Schrödinger equation that we write as a pair of real equations:

$$\begin{aligned} p_t &= -q_{xx} + m_0^2 q \\ q_t &= +p_{xx} - m_0^2 p \end{aligned}, \tag{4.1}$$

where $p(x,t)$ and $q(x,t)$ are defined on the domain $(0,2\pi]$, m_0 is a constant, and periodic boundary conditions are assumed. Equations (4.1) are the Hamilton equations of motion for the Hamiltonian [**FH65**],

$$H[p(\cdot),q(\cdot)] = \frac{1}{2}\int_0^{2\pi} \left[(p_x)^2 + (q_x)^2 + m_0^2(p^2+q^2)\right]\,dx, \tag{4.2}$$

with $p(x)$ and $q(x)$ being the canonically conjugate variables.

The prior measure. Equation (4.1) preserves any density that is a function of the Hamiltonian. We will assume that the prior measure is given by the canonical density,

$$f_0[p(\cdot),q(\cdot)] = \exp\left\{-H[p(\cdot),q(\cdot)]\right\}, \tag{4.3}$$

where the temperature has been chosen equal to one.

The measure defined by equation (4.3) is absolutely continuous with respect to a Wiener measure [**McK95**], and its samples are, with probability one, almost

nowhere differentiable. The corresponding solutions of the equations of motion are weak and hard to approximate numerically.

The Hamiltonian (4.2) is quadratic in p and q, hence the probability density (4.3) is Gaussian. By symmetry we see that the unconstrained means $\langle p(x)\rangle$ and $\langle q(x)\rangle$ are zero. To extract the matrix of covariance functions A^{-1}, we write the Hamiltonian (4.2) as a double integral:

$$H[p(\cdot),q(\cdot)] = \iint \Big[p_x(x)\delta(x-y)p_x(y) + q_x(x)\delta(x-y)q_x(y) + \\ + m_0^2\, p(x)\delta(x-y)p(y) + m_0^2\, q(x)\delta(x-y)q(y)\Big]\, dx\, dy. \tag{4.4}$$

Integration by parts shows that the entries of the matrix of functions A are

$$a^{ij}(x,y) = \left[-\delta''(x-y) + m_0^2\,\delta(x-y)\right]\,\delta_{ij}, \tag{4.5}$$

where the indices i and j represent either p or q, and $\delta''(\cdot)$ is a second derivative of a delta function. The integral equation for the inverse operator A^{-1} can be solved by Fourier series. The result is a translation-invariant diagonal matrix

$$[a^{-1}(x,y)]^{ij} = \frac{1}{2\pi}\delta_{ij}\sum_{k=-\infty}^{\infty}\frac{e^{ik(x-y)}}{k^2+m_0^2}. \tag{4.6}$$

The collective variables. We assume that the initial data for equations (4.1) are drawn from the distribution (4.3), and that $2N$ measurements have revealed the values of the $2N$ collective variables,

$$\begin{aligned} U_\alpha^p[p(\cdot),q(\cdot)] &\equiv (g_\alpha(\cdot),p(\cdot)) = V_\alpha^p \\ U_\alpha^q[p(\cdot),q(\cdot)] &\equiv (g_\alpha(\cdot),q(\cdot)) = V_\alpha^q, \end{aligned} \tag{4.7}$$

for $\alpha = 1,\dots,N$. The kernels $g_\alpha(x)$ are translates of each other, $g_\alpha(x) = g(x-x_\alpha)$, and the points $x_\alpha = 2\pi\alpha/N$ form a regular mesh on the interval $(0,2\pi]$. We choose

$$g(x) = \frac{1}{\sqrt{\pi}\sigma}\sum_{\tau=-\infty}^{\infty}\exp\left[-\frac{(x-2\pi\tau)^2}{\sigma^2}\right], \tag{4.8}$$

i.e., the kernel is a normalized Gaussian whose width is σ, with suitable images to enforce periodicity. The Fourier representation of $g(x)$ is

$$g(x) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ikx}e^{-\frac{1}{4}k^2\sigma^2}. \tag{4.9}$$

We could have trivialized this example by choosing as kernels a set of trigonometric functions, which are eigenfunctions of the evolution operator. The goal here is to demonstrate what one could do when an exact representation of the eigenfunctions is not known.

Conditional expectation. We now demonstrate the application of the Lemmas derived in the previous section. Given the initial data, V^p and V^q, we may calculate the expectation of the functions $p(x)$ and $q(x)$; these conditional averages are given by equation (3.28). Because the unconditional averages of $p(x)$, $q(x)$, U_α^p and U_α^q all vanish, and the unconditional covariance $[a^{-1}(x,y)]^{ij}$ is diagonal with

respect to i and j (p and q are independent), equation (3.28) reduces to a simpler expression; the conditional average of $p(x)$, for example, is

$$\langle p(x)\rangle_V = \sum_{\alpha=1}^{N} c_\alpha^{pp}(x) V_\alpha^p, \tag{4.10}$$

where

$$c_\alpha^{pp}(x) = \sum_{\beta=1}^{N} \left([a^{-1}(x,\cdot)]^{pp}, g_\beta(\cdot)\right) [m^{-1}]_{\beta\alpha}^{pp} = c_\alpha^{qq}(x), \tag{4.11}$$

and $[m^{-1}]_{\beta\alpha}^{pp}$ are the entries of an $N \times N$ matrix M^{-1} (the upper indices p are considered as fixed) whose inverse M has entries

$$m_{\beta\alpha}^{pp} = \iint g_\beta(x)[a^{-1}(x,y)]^{pp} g_\alpha(y)\, dx\, dy = m_{\beta\alpha}^{qq}. \tag{4.12}$$

Substituting the Fourier representations of A^{-1} (4.6) and g (4.9), we obtain

$$c_\alpha^{pp}(x) = \frac{1}{2\pi} \sum_{\alpha=1}^{N} \sum_{k=-\infty}^{\infty} \frac{e^{-\frac{1}{4}k^2\sigma^2}}{k^2+m_0^2} \exp\left[ik(x-x_\beta)\right] [m^{-1}]_{\beta\alpha}^{pp}, \tag{4.13}$$

and

$$m_{\beta\alpha}^{pp} = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \frac{e^{-\frac{1}{2}k^2\sigma^2}}{k^2+m_0^2} \exp\left[ik(x_\alpha - x_\beta)\right]. \tag{4.14}$$

The regression function (4.10) can be viewed as an "optimal interpolant"; it is the expectation value of the function $p(x)$ given what is known. Examples of regression functions are plotted in Figure 1 for a mesh of $N = 5$ points. The open circles represent the values of the five collective variables V_α^p; the abscissa is the location of the point x_α around which the average is computed, and the ordinate is the value of the corresponding collective variable. The three curves represent the interpolating function (4.10) for three different values of the kernel width: $\sigma = \Delta x = 2\pi/N$ (solid line), $\sigma = 0.5\,\Delta x$ (dashed line), and $\sigma = 0.1\,\Delta x$ (dash-dot line). The parameter m_0 was taken to be one.

Time evolution. We next consider the time evolution of the mean value of the collective variables U^p and U^q, first based on the approximating scheme (2.11). The equation for V_α^p, for example, is

$$\begin{aligned} \frac{dV_\alpha^p}{dt} &= \left\langle \left(g_\alpha(\cdot), -q_{xx}(\cdot) + m_0^2 q(\cdot)\right)\right\rangle_V = \\ &= -\left(g_\alpha(\cdot), \frac{\partial^2}{\partial x^2}\langle q(\cdot)\rangle_V\right) + m_0^2\left(g_\alpha(\cdot), \langle q(\cdot)\rangle_V\right). \end{aligned} \tag{4.15}$$

Substituting the regression function (4.10) we find:

$$\frac{dV_\alpha^p}{dt} = \sum_{\gamma=1}^{N} \left\{ \sum_{\beta=1}^{N} \left(g_\alpha(\cdot), g_\beta(\cdot)\right) [m^{-1}]_{\beta\gamma}^{qq} \right\} V_\gamma^q. \tag{4.16}$$

A similar equation is obtained for V_α^q by the symmetry transformation $V_\alpha^p \to V_\alpha^q$ and $V_\alpha^q \to -V_\alpha^p$. Equation (4.16) represents a set of $2N$ ordinary differential equations that approximate the mean evolution of the collective variables. These equations are easy to solve with standard ODE solvers. Note that the matrix elements in braces need to be computed only once to define the scheme.

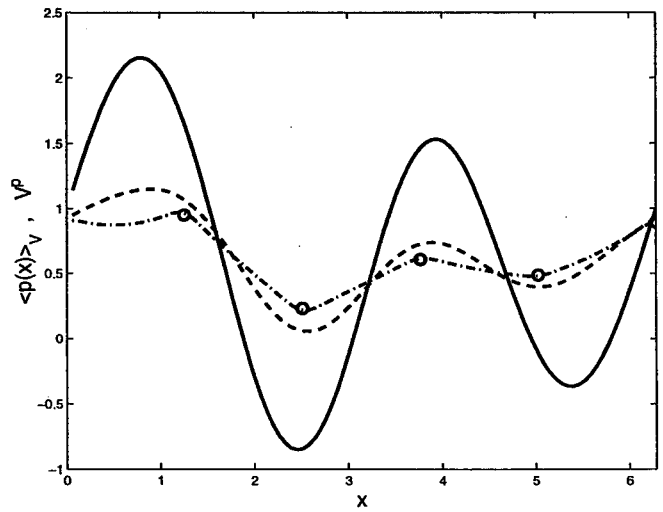

FIGURE 1. Example of regression functions for the linear Schrödinger equation. Values for five collective variables were chosen, representing local averages of $p(x)$ on a uniformly spaced grid. The kernels are translates of each other and have Gaussian profiles of width σ centered at the grid points. The lines represent the regression function, or optimal interpolant $\langle p(x)\rangle_V$ given by equation (4.10) for $\sigma = \Delta x$ (solid), $\sigma = 0.5\,\Delta x$ (dashed), and $\sigma = 0.1\,\Delta x$ (dash-dot).

We next calculate the *exact* mean value of the collective variables, U^p and U^q, at time t, conditioned by the initial data, V^p and V^q, at time $t = 0$, so that they can be compared with the result $V(t)$ of the scheme we just presented. We are able to do so in the present case because the equations are linear, and a simple representation of the evolution operator can be found.

The solution to the initial value problem (4.1) can be represented by Fourier series,

$$
\begin{aligned}
p(x,t) &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\int e^{ik(x-y)}\left[p(y)\,\cos\omega t + q(y)\,\sin\omega t\right]\,dy \\
q(x,t) &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\int e^{ik(x-y)}\left[q(y)\,\cos\omega t - p(y)\,\sin\omega t\right]\,dy
\end{aligned}
\tag{4.17}
$$

where $p(y)$ and $q(y)$ are the (random) initial conditions, and $\omega = k^2 + m_0^2$.

The expectation values of the collective variables U_α^p and U_α^q are obtained by averaging the scalar products $(p(\cdot,t), g_\alpha(\cdot))$ and $(q(\cdot,t), g_\alpha(\cdot))$ with respect to the initial distribution. Because equations (4.17) are linear in the random variables

$p(y)$ and $q(y)$ this gives

(4.18)

$$\langle U_\alpha^p[p(\cdot),q(\cdot),t]\rangle_V = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\int e^{ik(x_\alpha-y)-\frac{1}{4}k^2\sigma^2}\left[\langle p(y)\rangle_V\cos\omega t+\langle q(y)\rangle_V\sin\omega t\right]dy$$

$$\langle U_\alpha^q[p(\cdot),q(\cdot),t]\rangle_V = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\int e^{ik(x_\alpha-y)-\frac{1}{4}k^2\sigma^2}\left[\langle q(y)\rangle_V\cos\omega t-\langle p(y)\rangle_V\sin\omega t\right]dy$$

Note that in the linear case averaging and time evolution commute; equation (4.18) would have also been obtained if we first computed the mean initial state, $\langle p(y)\rangle_V$ and $\langle q(y)\rangle_V$, evolved it in time according to (4.17), and finally computed the collective variables by taking the appropriate scalar products.

To complete the calculation, we substitute the linear regression formula (4.10) for $\langle p(y)\rangle_V$ and $\langle q(y)\rangle_V$ and obtain:

$$\begin{aligned}\langle U_\alpha^p[p(\cdot),q(\cdot),t]\rangle_V &= \sum_{\beta,\gamma=1}^{N}\left\{c_{\alpha\beta}^C(t)[m^{-1}]_{\beta\gamma}^{pp}V_\gamma^p+c_{\alpha\beta}^S(t)[m^{-1}]_{\beta\gamma}^{qq}V_\gamma^q\right\}\\ \langle U_\alpha^q[p(\cdot),q(\cdot),t]\rangle_V &= \sum_{\beta,\gamma=1}^{N}\left\{c_{\alpha\beta}^C(t)[m^{-1}]_{\beta\gamma}^{pp}V_\gamma^q-c_{\alpha\beta}^S(t)[m^{-1}]_{\beta\gamma}^{qq}V_\gamma^p\right\}\end{aligned} \tag{4.19}$$

where

$$c_{\alpha\beta}^C(t) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\frac{\cos\omega t}{\omega}e^{ik(x_\alpha-x_\beta)}e^{-\frac{1}{2}k^2\sigma^2}, \tag{4.20}$$

and

$$c_{\alpha\beta}^S(t) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\frac{\sin\omega t}{\omega}e^{ik(x_\alpha-x_\beta)}e^{-\frac{1}{2}k^2\sigma^2}. \tag{4.21}$$

Results. We now compare the exact formula (4.19) for the future expectation value of the collective variables to the approximation (4.16). Figures 2*a*–2*c* compare between the two evolutions for $N=5$ and randomly selected initial data, V_α^p and V_α^q. The graphs show the mean time evolution of the collective variable $U_1^p[p(\cdot),q(\cdot)]$. The same set of initial values was used in the three plots; the difference is in the width σ of the kernels $g_\alpha(x)$: $\sigma=\Delta x$ (Figure 2*a*), $\sigma=0.5\,\Delta x$ (Figure 2*b*), and $\sigma=0.1\,\Delta x$ (Figure 2*c*). In the first case, in which the kernel width equals the grid spacing, the approximation is not distinguishable from the exact solution on the scale of the plot for the duration of the calculation. The two other cases show that the narrower the kernel is, the sooner the curve deviates from the exact solution.

5. A nonlinear Hamiltonian system

The equations of motion. The method demonstrated in the preceding section can be generalized to a nonlinear Schrödinger equation. However, we want to exhibit the power of our method by comparing the solutions that it yields to exact solutions; in the nonlinear case, exact solutions of problems with random initial conditions are hard to find, so we resort to a stratagem. Even though our method applies to nonlinear partial differential equations, we study instead a finite

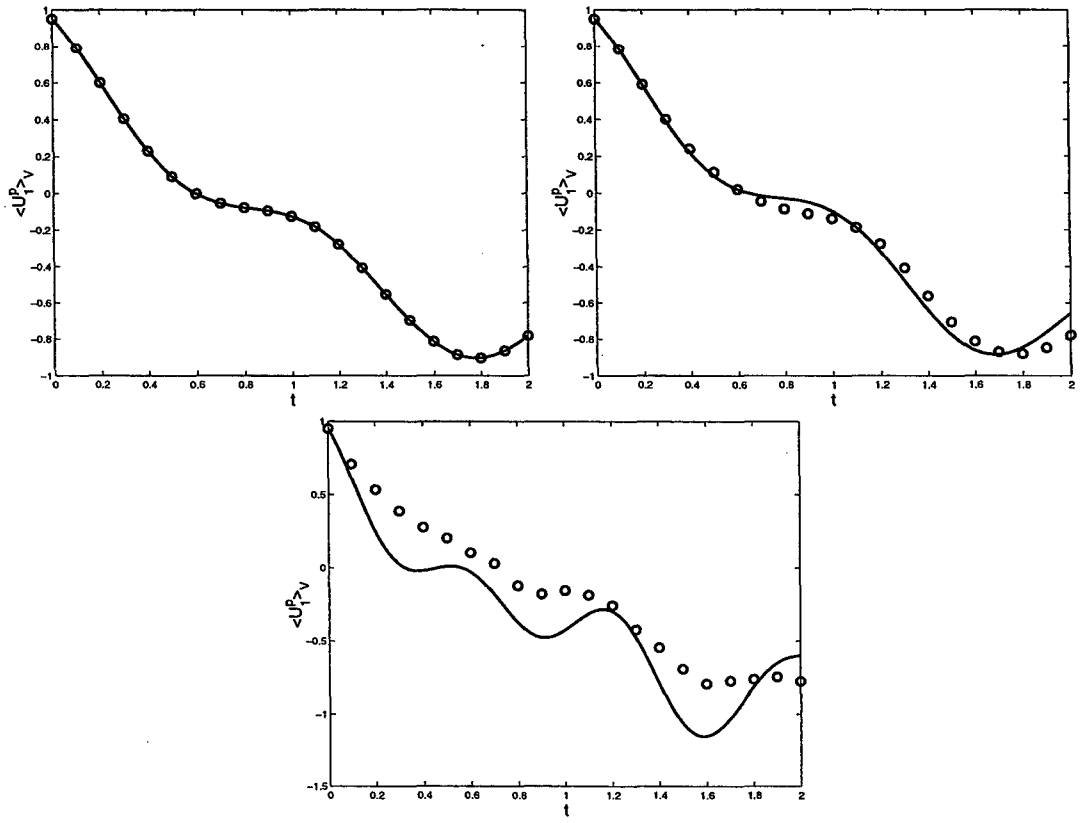

FIGURE 2. Mean evolution of the collective variable $U_1^p[p(\cdot), q(\cdot)]$ for $N = 5$, and a random choice of the initial data V^p and V^q. The open dots represent the exact solution (4.19), whereas the lines represent the approximate solution obtained by an integration of the set of 10 ordinary differential equations (4.16). The three graphs are for different values of the kernel width σ: (a) $\sigma = \Delta x$, (b) $\sigma = 0.5\,\Delta x$, and (c) $\sigma = 0.1\,\Delta x$.

dimensional system of $2n$ ordinary differential equations that is formally a finite difference approximation of a nonlinear Schrödinger equation:

$$\begin{aligned} \frac{dp(j)}{dt} &= -\frac{q(j-1) - 2q(j) + q(j+1)}{\Delta x^2} + q^3(j) \\ \frac{dq(j)}{dt} &= +\frac{p(j-1) - 2p(j) + p(j+1)}{\Delta x^2} - p^3(j) \end{aligned} \qquad j = 1, \ldots, n, \tag{5.1}$$

where $\Delta x = 1/n$ is the mesh spacing, and periodicity is enforced with $p(0) \equiv p(n)$, $p(n+1) \equiv p(1)$, etc; this system is non-integrable for $n > 1$. The approximation is only formal because we shall be considering non-smooth data which give rise to weak solutions that cannot be readily computed by difference methods.

We shall pretend that n is so large that the system (5.1) cannot be solved on a computer, and shall therefore seek an approximation that requires a computation

with fewer variables. In practice we shall pick an n small enough so that the results of the approximate procedure can be compared to an *ensemble* of exact solutions.

The prior measure. The system of equations (5.1) is the Hamilton equations of motion for the Hamiltonian

(5.2)
$$H[p,q] = \frac{1}{2}\sum_{j=1}^{n}\left\{\left[\frac{p(j+1)-p(j)}{\Delta x}\right]^2 + \left[\frac{q(j+1)-q(j)}{\Delta x}\right]^2 + \frac{1}{2}\left[p^4(j)+q^4(j)\right]\right\},$$

where $p \equiv (p(1),\ldots,p(n))$ and $q \equiv (q(1),\ldots,q(n))$. The differential equations (5.1) preserve the canonical density

$$f_0[p,q] = \exp\{-H[p,q]\}, \tag{5.3}$$

which we postulate, as before, to be the prior probability density.

The prior density (5.3) is not Gaussian, which raises a technical difficulty in computing expectation values. We adopt here an approximate procedure where the density (5.3) is approximated by a Gaussian density that yields the same first and second moments (means and covariances) of the vectors p and q. The means are zero by symmetry:

$$\langle p(j)\rangle = \langle q(j)\rangle = 0 \tag{5.4}$$

(positive and negative values of these have equal weight). Also all p's and q's are uncorrelated:

$$\langle p(j_1)q(j_2)\rangle = 0, \tag{5.5}$$

since the density factors into a product of a density for the p's and a density for the q's. Thus $\langle p(j_1)p(j_2)\rangle = \langle q(j_1)q(j_2)\rangle$ are the only non-trivial covariances. Finally, since the Hamiltonian is translation invariant, these covariances depend only on the separation between the indices j_1 and j_2, and are symmetric in $j_1 - j_2$.

To relate the present discrete problem to the continuous formalism used in the preceding section we write in analogy to (4.6)

$$\begin{aligned} \mathrm{Cov}\,[p(j_1),p(j_2)] &= [a^{-1}(j_1,j_2)]^{pp} = c(|j_1-j_2|) \\ \mathrm{Cov}\,[p(j_1),q(j_2)] &= [a^{-1}(j_1,j_2)]^{pq} = 0, \end{aligned} \tag{5.6}$$

with $j_1, j_2 = 1,\ldots,n$. We computed the numbers, $c(|j_1-j_2|)$, for $n = 16$ and $j_1 - j_2 = 0,\ldots,15$ by a Metropolis Monte-Carlo algorithm [**BH92**]; the covariances obtained this way are shown in Figure 3. Along with the zero means, the numbers represented in Figure 3 completely determine the *approximate* prior distribution.

The collective variables. We next define a set of $2N$ collective variables ($N < n$), whose values we assume to be given at the initial time. The class of collective variables that is the discrete analog of (4.7) is of the form

$$\begin{aligned} U_\alpha^p[p,q] &= (g_\alpha(\cdot),p(\cdot)) \equiv \sum_{j=1}^{n} g_\alpha(j)p(j) \\ U_\alpha^q[p,q] &= (g_\alpha(\cdot),q(\cdot)) \equiv \sum_{j=1}^{n} g_\alpha(j)q(j) \end{aligned} \qquad \alpha = 1,\ldots,N, \tag{5.7}$$

where the g's are discrete kernels. In the calculations we exhibit we chose $n = 16$ and $N = 2$ so that we aim to reduce the number of degrees of freedom by a factor

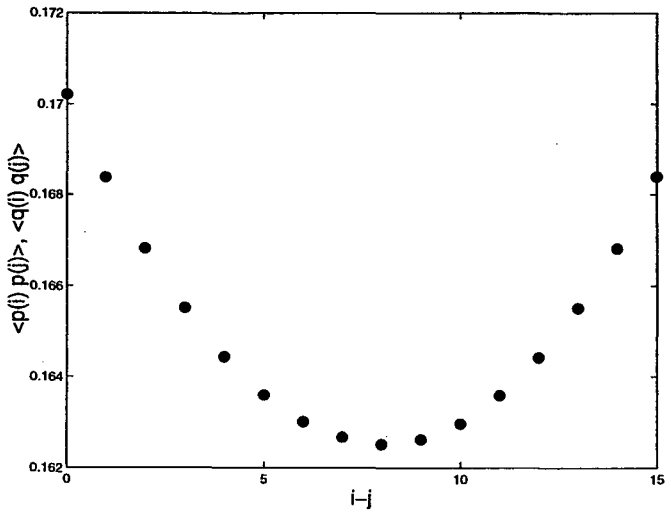

FIGURE 3. The covariance $\langle p(i)\,p(j)\rangle = \langle q(i)\,q(j)\rangle$ as function of the grid separation $i-j$ for the non-Gaussian probability distribution (5.3) with $n = 16$. These values were computed by a Metropolis Monte-Carlo simulation.

of 8. We pick as kernels discretized Gaussian functions centered at the grid points $j = 1$ and $j = 9$:

$$g_1(j) = \frac{1}{Z}\exp\left\{-\frac{d^2(1,j)}{n^2\sigma^2}\right\}$$
$$g_2(j) = \frac{1}{Z}\exp\left\{-\frac{d^2(9,j)}{n^2\sigma^2}\right\} \tag{5.8}$$

where Z is a normalizing constant, $\sigma = 0.25$, and $d(j_1, j_2)$ is a distance function over the periodic index axis, i.e., it is the minimum of $|j_1 - j_2|$, $|j_1 - j_2 - n|$, and $|j_1 - j_2 + n|$.

Conditional expectation. With the approximate measure defined by the covariances (5.6), and the collective variables (5.7), whose measured values are again denoted by V^p_α and V^q_α, we can approximate the conditional expectation of various observables $O[p,q]$. We shall need specifically the conditional expectation values of $p(j)$ and $p^3(j)$.

The approximate conditional expectation value of $p(j)$ is given by the discrete analog of equation (4.10), namely,

$$\langle p(j)\rangle_V = \sum_{\alpha=1}^{N} c^{pp}_\alpha(j) V^p_\alpha, \tag{5.9}$$

where

$$c^{pp}_\alpha(j) = \sum_{\beta=1}^{N} \left([a^{-1}(j,\cdot)]^{pp}, g_\beta(\cdot)\right) [m^{-1}]^{pp}_{\beta\alpha}, \tag{5.10}$$

and

$$m^{pp}_{\beta\alpha} = \sum_{j_1,j_2=1}^{n} g_\beta(j_1)[a^{-1}(j_1,j_2)]^{pp} g_\alpha(j_2). \tag{5.11}$$

(Again, the matrix inversion is only with respect to the lower indices α and β.)

To calculate the approximate conditional expectation value of $p^3(j)$ we first use Wick's theorem (Lemma 3.3):

$$\langle p^3(j)\rangle_V = 3\,\langle p^2(j)\rangle_V\,\langle p(j)\rangle_V - 2\,\langle p(j)\rangle_V^3\,, \tag{5.12}$$

and then calculate the conditional second moment by using the discrete analog of equation (3.17):

$$\langle p^2(j)\rangle_V = \langle p(j)\rangle_V^2 + [a^{-1}(j,j)]^{pp} - \sum_{\alpha=1}^{N} c^{pp}_\alpha(j)\left(g_\alpha(\cdot),[a^{-1}(\cdot,j)]^{pp}\right). \tag{5.13}$$

Time evolution. The approximating scheme for calculating the mean evolution of the $2N$ collective variables U^p and U^q is derived by substituting the kernels (5.8) and the equations of motion (5.1) in the approximation formula (2.11). The equation for V^p_α, for example, is

$$\begin{aligned}\frac{dV^p_\alpha}{dt} &= -\frac{1}{\Delta x^2}\sum_{j=1}^{n} g_\alpha(j)\left[\langle q(j-1)\rangle_V - 2\,\langle q(j)\rangle_V + \langle q(j+1)\rangle_V\right] + \\ &\quad + \sum_{j=1}^{n} g_\alpha(j)\,\langle q^3(j)\rangle_V\,.\end{aligned} \tag{5.14}$$

Substituting the expressions for the conditional expectations (5.9) and (5.12), and performing the summation, using the values of the covariances plotted in Figure 3, we explicitly obtain a closed set of 4 ordinary differential equations. The equation for V^p_1 is:

$$\begin{aligned}\frac{dV^p_1}{dt} &= -19.5\,(V^q_2 - V^q_1) + \\ &\quad + \left[1.50\,(V^q_1)^3 - 0.88\,(V^q_1)^2 V^q_2 + 0.27\,V^q_1 (V^q_2)^2 + 0.11\,(V^q_2)^3\right].\end{aligned} \tag{5.15}$$

The equation for V^p_2 is obtained by substituting $1 \leftrightarrow 2$; the equations for V^q_1 and V^q_2 are obtained by the transformation $p \to q$ and $q \to -p$.

Unlike in the linear case, we cannot calculate analytically the mean evolution of the collective variables. To assess the accuracy of the approximate equation (5.15) we must compare the solution it yields with an average over an ensemble of solutions of the "fine scale" problem (5.1). To this end, we generated a large number of initial conditions that are consistent with the given values, V^p and V^q, of the collective variables. The construction of this ensemble was done by a Metropolis Monte Carlo algorithm, where new states are generated randomly by incremental changes, and accepted or rejected with a probability that ensures that for large enough samples the distribution converges to the conditioned canonical distribution. We generated an ensemble of 10^4 initial conditions; each initial state was then evolved in time using a fourth-order Runge-Kutta method. Finally, for each time level we computed the distribution of collective variables, U^p and U^q; the average of this distribution should be compared with the prediction of equations (5.15).

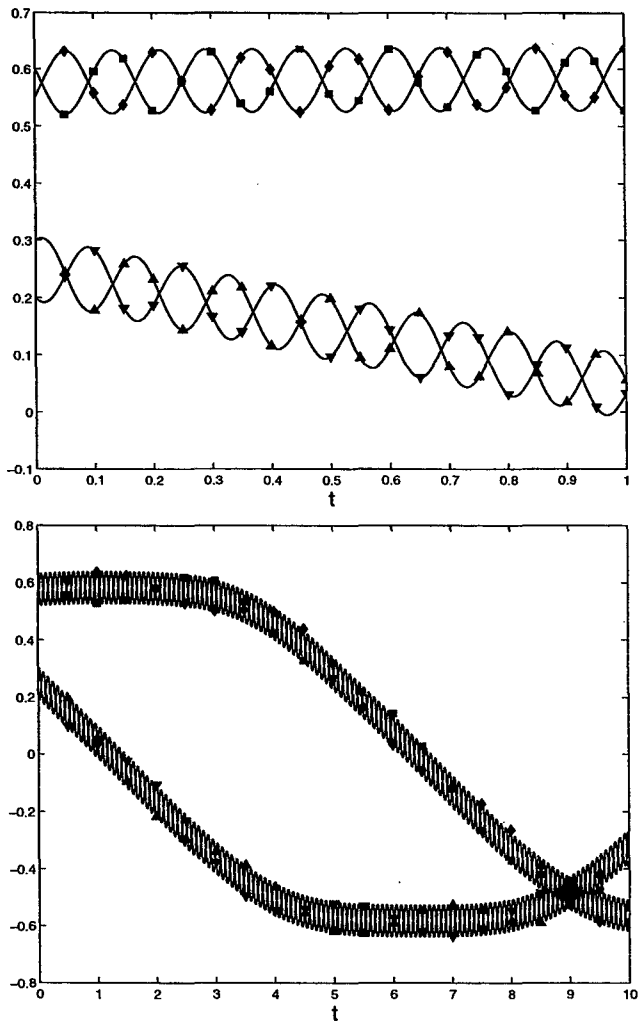

FIGURE 4. Evolution in time of the mean value of the four collective variable: V_1^p (▼), V_2^p (▲), V_1^q (■), and V_2^q (♦). The symbols represent the values of these quantities obtained by solving the 32 equations (5.1) for 10^4 initial conditions compatible with the initial data, and averaging. The solid lines are the values of the four corresponding functions obtained by integrating equation (5.15). Figures (a) and (b) are for the time intervals $[0, 1]$ and $[0, 10]$ respectively.

The comparison between the true and the approximate evolution is shown in Figure 4. Once again the reduced system of equations reproduces the average behavior of the collective variables with excellent accuracy, but at a very much smaller computational cost. Indeed, we compare one solution of 4 equations to 10^4 solutions of 32 equations.

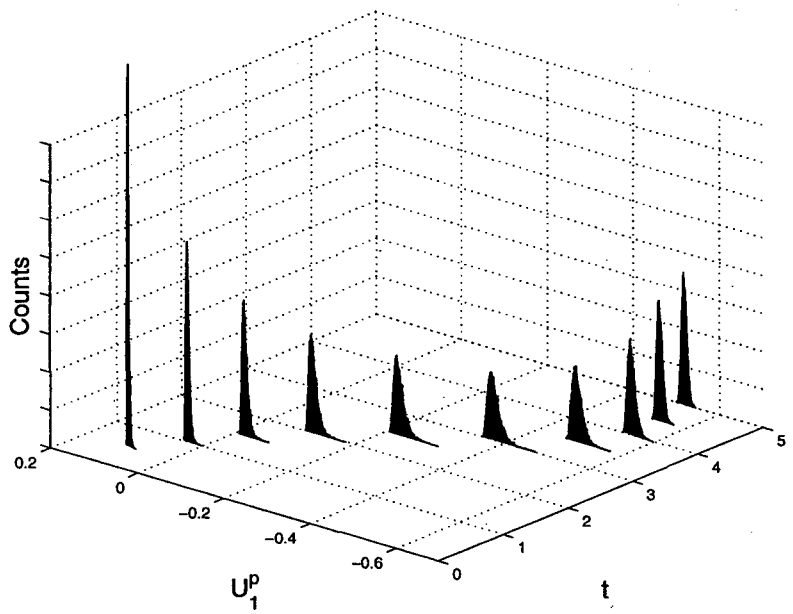

FIGURE 5. Evolution of the distribution of the collective variable U_1^p. The x-axis represents time, the y-axis represents the value of U_1^p, and the z-axis is proportional to the density of states that correspond to the same value of U_1^p at the given time.

In Figure 5 we show the evolution of the *distribution* of values assumed by the collective variable U_1^p; the data was extracted from the evolution of the ensemble. The distribution is initially sharply peaked, and spreads out as time evolves; yet, it remains sufficiently narrow throughout this computation, so that the approximation that projects that distribution back onto a sharp one is reasonable. This indicates that the choice of collective variables, or kernels, was appropriate. The use of narrow kernels, or even point values, would have yielded a distribution of value that spreads out almost instantaneously.

6. Conclusions

We have shown how to calculate efficiently, for a class of problems, the average behavior of an ensemble of solutions the individual members of which are very difficult to evaluate. The approach is reminiscent of statistical mechanics, where it is often easier to predict the evolution of a mole of particles than to predict the evolution of, say, a hundred particles, if one is content with the average behavior of a set of coarse variables (collective variables). The key step is the identification of a correspondence between underresolution and statistics; underresolved data define, together with prior statistical information, an ensemble of initial conditions, and the most one can aim for is to predict the expectation with respect to this ensemble of certain observables at future times. Our approach applies in those cases where prior statistical information is available, and is consistent with the differential equations; for example, it may consist of a measure invariant under the flow defined by the

differential equations. Fortunately, there are important classes of problems where we can find such information.

We proposed a scheme (2.11) that advances in time a set of variables that approximate the expectation values of a set of collective variables. As we explained, this scheme has to be viewed as a first approximation; more sophisticated schemes may be designed by allowing the kernels to vary in time and/or by keeping track of higher moments of the collective variables. Such refinements are the subject of ongoing research [**CKKT**].

One limitation of our present scheme can be perceived by considering the long time behavior of the nonlinear Hamiltonian system presented in Section 5. The flow induced by equations (5.1) is likely to be ergodic, hence the probability density function will approach, as $t \to \infty$, the invariant distribution. Indeed, the initial data have a decreasing influence on the statistics of the solutions as time progresses. This implies that the expectation values of the observables U^p and U^q will tend to their unconditional means, i.e., will decay to zero. On the other hand, no such decay occurs if one integrates the effective equations (5.15) for very long times. One must conclude that the present model is accurate for time intervals that are not longer than the time during which the initial data influence the outcome of the calculation.

The above discussion raises a number of questions interesting on their own: What is the range of influence, or the predictive power, of a given set of data? How much information is contained in partial data? These questions need to be formulated in a more quantitative way; they are intimately related to the question of how to choose appropriate collective variables, and their scope is beyond any particular method of solution.

Finally, a full knowledge of the prior measure is a luxury one cannot always expect. One needs to consider problems where the statistical information is only partial; for example, a number of moments may be known from asymptotics and scaling analyses (e.g., in turbulence theory [**Bar96, BC97, BC98**]). One can readily see from the nonlinear example that one can make do with the knowledge of means, covariances, and perhaps some higher-order moments. In addition, this knowledge is needed only on scales comparable with the widths of the kernels.

References

[Bar96] G.I. Barenblatt, *Scaling, self-similarity and intermediate asymptotics*, Cambridge University Press, Cambridge, 1996.

[BC97] G.I. Barenblatt and A.J. Chorin, *Scaling laws for fully developed turbulent flow in pipes*, Appl. Mech. Rev. **50** (1997), 413–429.

[BC98] G.I. Barenblatt and A.J. Chorin, *Scaling laws and vanishing viscosity limits in turbulence theory*, Proc. Symposia Appl. Math. AMS **54** (1998), 1–25.

[BH92] K. Binder and D. Heerman, *Monte-carlo simulation in statistical physics*, Springer, Berlin, 1992.

[CKK98] A.J. Chorin, A. Kast, and R. Kupferman, *Optimal prediction of underresolved dynamics*, Proc. Nat. Acad. Sci. USA **95** (1998), 4094–4098.

[CKKT] A.J. Chorin, A. Kast, R. Kupferman, and B. Turkington, *Optimal prediction of two-dimensional euler flow*, in preparation.

[FH65] R. Feynman and A. Hibbs, *Quantum mechanics and path integrals*, McGraw-Hill, New York, 1965.

[Hal] O.H. Hald, *Optimal prediction of linear systems*, in preparation.

[Kle89] H. Kleinert, *Gauge fields in condensed matter*, World Scientific, Singapore, 1989.

[McK95] H.P. McKean, *Statistical mechanics of nonlinear wave equations IV: Cubic Schrödinger equations*, Comm. Math. Phys. **168** (1995), 479–491.

[MW90] W.D. McComb and A.G. Watt, *Conditional averaging procedure for the elimination of the small-scale modes from incompressible fluid turbulence at high reynolds numbers*, Phys. Rev. Lett. **65** (1990), 3281–3284.

[Ris84] H. Risken, *The fokker-planck equation*, Springer, New York, 1984.

[SM97] A. Scotti and C. Meneveau, *Fractal model for coarse-grained partial differential equations*, Phys. Rev. Lett. **78** (1997), 867–870.

DEPARTMENT OF MATHEMATICS, LAWRENCE BERKELEY NATIONAL LABORATORY, MAIL STOP 50A-2152, 1 CYCLOTRON ROAD, BERKELEY, CA 94720
E-mail address: chorin@math.berkeley.edu
URL: http://math.berkeley.edu/~chorin

DEPARTMENT OF MATHEMATICS, LAWRENCE BERKELEY NATIONAL LABORATORY, MAIL STOP 50A-2152, 1 CYCLOTRON ROAD, BERKELEY, CA 94720
E-mail address: anton@math.lbl.gov
URL: http://www.lbl.gov/~anton

DEPARTMENT OF MATHEMATICS, LAWRENCE BERKELEY NATIONAL LABORATORY, MAIL STOP 50A-2152, 1 CYCLOTRON ROAD, BERKELEY, CA 94720
Current address: Institute of Mathematics, The Hebrew University, Jerusalem, 91904 Israel
E-mail address: raz@math.lbl.gov
URL: http://www.lbl.gov/~raz

Ernest Orlando Lawrence Berkeley National Laboratory
One Cyclotron Road | Berkeley, California 94720

Prepared for the U.S. Department of Energy under Contract No. DE-AC03-76SF00098