LBNL-44105
Preprint

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

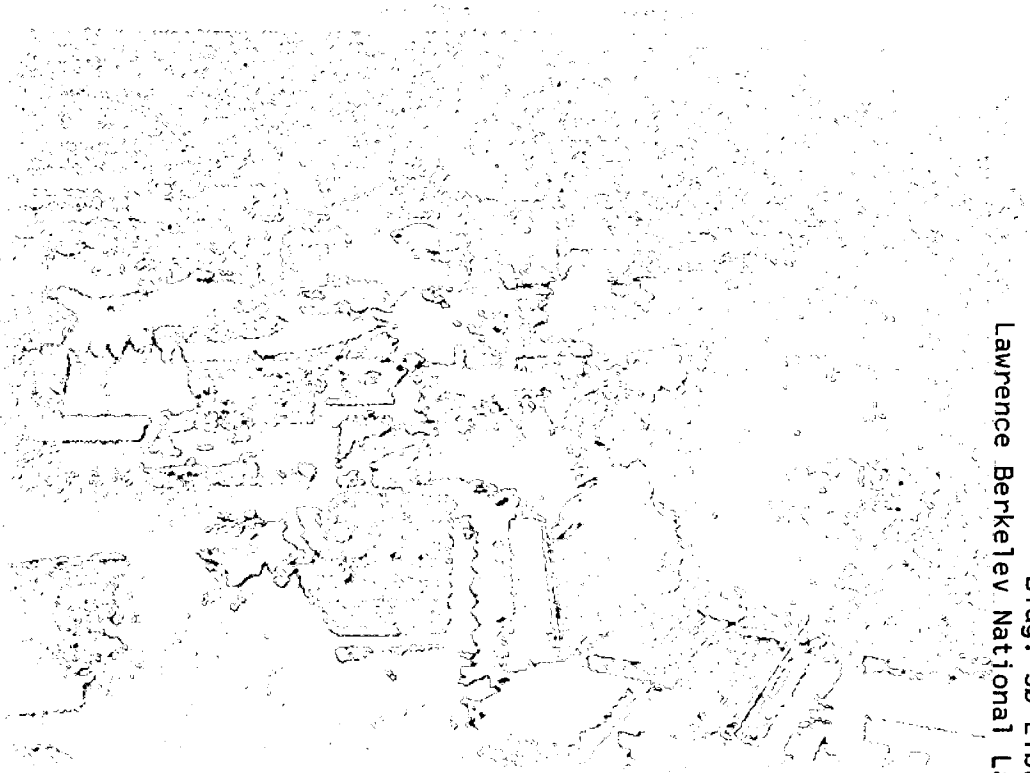
Optimal Prediction for Hamiltonian Partial Differential Equations

Alexandre J. Chorin, Raz Kupferman,
and Doron Levy

Computing Sciences Directorate
Mathematics Department

August 1999

To be submitted
for publication



DISCLAIMER

LBNL-44105

Optimal Prediction for Hamiltonian Partial Differential Equations

Alexandre J. Chorin, Raz Kupferman, and Doron Levy

Computing Sciences Directorate
Mathematical Department
Ernest Orlando Lawrence Berkeley National Laboratory
University of California
Berkeley, California 94720

August 1999

This work was supported in part by the Director, Office of Science, Office of Advanced Scientific Computing 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 the National Science Foundation under Contract No. DMS94-14631.

Optimal prediction for Hamiltonian partial differential equations

Alexandre J. Chorin[†] Raz Kupferman[‡] Doron Levy[§]

Abstract

Optimal prediction methods compensate for a lack of resolution in the numerical solution of time-dependent differential equations through the use of prior statistical information. We present a new derivation of the basic methodology, show that field-theoretical perturbation theory provides a useful device for dealing with quasi-linear problems, and provide a nonlinear example that illuminates the difference between a pseudo-spectral method and an optimal prediction method with Fourier kernels. Along the way, we explain the differences and similarities between optimal prediction, the representer method in data assimilation, and duality methods for finding weak solutions. We also discuss the conditions under which a simple implementation of the optimal prediction method can be expected to perform well.

Key words: Optimal prediction, underresolution, perturbation methods, regression, nonlinear Schrödinger, pseudo-spectral methods.

AMS(MOS) subject classification: Primary 65M99; secondary 81T15, 35Q55, 65M70.

1 Introduction

We consider a Cauchy problem

$$\begin{cases} u_t = R\left(x, u, u_x, u_{xx}, \dots\right), \\ u(x,0) = u_0(x), \end{cases} \tag{1.1}$$

where $R(x, u, u_x...)$ is a (generally nonlinear) function of $u = u(x,t)$, of its spatial derivatives, and of the independent variable x, in any number of dimensions; subscripts denote differentiation. We assume that we cannot afford to use enough computational

[†]Department of Mathematics, University of California, Berkeley, CA 94720, and Lawrence Berkeley National Laboratory; chorin@math.berkeley.edu

[‡]Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel; razk@math.huji.ac.il

[§]Department of Mathematics, University of California, Berkeley, CA 94720, and Lawrence Berkeley National Laboratory; dlevy@math.berkeley.edu

elements (for example, mesh points) to resolve the problem adequately, or that we do not wish to use many computational elements because we are only interested in some of the features of the solutions and it seems to be wasteful to compute all the details. We also assume that we have some prior statistical information about the distribution of possible solutions. The question we address is how the prior statistics can be used to obviate the need for resolution. In contrast to problems that arise in some applications, especially in geophysics [1], we assume that the equations we are solving are fully known, and that the data are knowable in principle, even if we may not be able or willing to store them all in a computer's memory.

In the present paper we assume that our statistical information consists of an invariant measure μ on the space of solutions, i.e., we assume that the initial data are sampled from a probability distribution on the space of data, and that, in principle, if one takes one instance of initial data after the other and computes the solutions produced by (1.1) at any later time t, then the set of solutions obtained at that later time t (viewed as functions of the spatial variables) has the same probability distribution as the set of initial data. We assume further that this probability distribution is explicitly known. In section 4 below we give an example where these assumptions are satisfied and provide more precise definitions. Our present assumptions may be unnecessarily strong for some practical problems, but they simplify the exposition and it is often easy to weaken them. We shall call the invariant measure the *prior measure*; thus the prior measure is the distribution of the data before anything has been specified about a particular problem. In Hamiltonian systems a natural prior measure is the canonical measure induced by the Hamiltonian $\mathcal{H}$, i.e., a measure defined by the probability density

$$f(u) = Z^{-1} e^{-\mathcal{H}(u)/T},$$

where Z is a normalization constant and the parameter T, which determines the variance of the density, is known as the "temperature". We do not assume that the differential equation (1.1) admits a unique invariant measure; in cases of non-uniqueness the right choice of measure is part of the formulation of the problem.

We further assume that all we know, or care to know, at time $t = 0$, is a small set of data. In the present paper we choose these data to be of the form

$$(g_\alpha, u) = \int g_\alpha(x) u(x,0)\, dx = V_\alpha, \qquad \alpha = 1, \ldots, N, \tag{1.2}$$

where the $g_\alpha = g_\alpha(x)$ are suitably chosen kernels (the choice of kernels g_α is at our disposal). The question we are asking can now be rephrased as follows: How do we best predict the future using only N variables defined as in (1.2), given the prior measure μ?

At time $t = 0$ one can, at least in principle, find the mean solution conditioned by the data (1.2), for every point x, i.e., at each x one can average over all those functions in the support of the invariant measure that also satisfy the conditions (1.2) and find the mean v of u given the values V_α, $\alpha = 1, ..., N$; symbolically,

$$v(x) = E\left[u(x) | V_1, ..., V_N\right], \tag{1.3}$$

where $E[\cdot]$ denotes an expectation value. This is a regression problem, which can be solved by standard tools ([21], see also below). If one is given only the information

contained in (1.2), this regression is the right substitute for a more detailed knowledge of the data. To perform the regression at a later time t we need appropriate conditions at that later time, which should encode the later effect of the initial data (1.2). The problem now at hand is how to find these conditions and how to do it efficiently, in linear and in nonlinear problems.

We have already explained our basic approach in [4, 5, 6]. In the present paper we explain it in a different way which we hope is more transparent; the suggestion that a weak formulation is the right starting point is due to Gottlieb [11]. The paper has several goals: to explain what constitutes the novelty of our approach, to show how it meshes with field-theoretical perturbation methods, to give a simple, explicit, nonlinear example of the ways in which the basic algorithm differs from, and is superior to, a numerical method that uses the information in the initial data without taking advantage of the prior statistics, and to discuss the domain of applicability of the current implementation of the basic ideas. Our examples are of Schrödinger type.

Other work along somewhat analogous lines includes Scotti and Meneveau [24], where a "fractal"interpolation can be viewed as an analog of sampling a measure, and a stochastic construction by Vaillant [25].

A comment on notations: The notation in equation (1.3) is clear but cumbersome. A shorter version is: $v = E[u|V]$. We shall also use the simpler but less transparent physicists' notations: $\langle u \rangle$ for $E[u]$ and $\langle u \rangle_V$ for $E[u|V]$. Both notations $\langle u \rangle_V$, $E[u|V]$ are of course shorthands for $\langle u \rangle_{V_\alpha, g_\alpha}$ etc., since the conditioning (1.2) depends on the kernels as well as on the right-hand sides.

The paper is organized as follows: In section 2 we present an overview of optimal prediction. In sections 3, 4, we present some needed background material on regression and invariant measures. In sections 5, 6 we explain how perturbation theory can be used to implement optimal prediction in nonlinear problems, and provide examples of success and failure in problems with sparse data. In section 7 we provide a detailed analysis of a problem in which perturbation theory is carried only to zero-th order (but with a ground state that our previous work shows to be optimal), and in which the kernels are trigonometric functions. These simplifications yield results that are particularly transparent. Conclusions are drawn in a final section.

2 Weak solutions, regression, and prediction

We start with well-known considerations about weak solutions of linear equations; they lead naturally to an algorithm for solving underresolved linear equations, which will be useful below, in particular because we shall be able to present our proposals by way of a contrast. Consider the equation:

$$u_t = Lu, \tag{2.1}$$

where L is a linear operator. Multiply (2.1) by a smooth test function g; for simplicity assume that the boundary conditions on u are periodic, and thus g can be also assumed

periodic. Integration over a periodic domain in x and between $t = 0$ and $t = \tau$, followed by an integration by parts, results in

$$\int_{t=0}^{\tau} \int_x \left(g_t + L^\dagger g\right) u \, dx \, dt + (u(\tau), g(\tau)) - (u(0), g(0)) = 0, \tag{2.2}$$

where $(u(t), g(t))$ denotes $\int u(x,t) g(x,t) \, dx$ and $L^\dagger$ is the adjoint of L. A weak solution of (2.1) is a function $u(x,t)$ that satisfies (2.2) for all test functions g (see e.g.[19]). In particular, if the test functions g_α satisfy the adjoint equation

$$\frac{\partial g_\alpha}{\partial t} + L^\dagger g_\alpha = 0, \qquad \alpha = 1, \ldots, N, \tag{2.3}$$

then u is a weak solution of (2.1) provided (u, g_α) is a constant independent of t for each α. This observation produces a possible numerical method for solving (2.1): One can construct a collection of functions g_α that satisfy the adjoint equation (2.3), find the numerical values V_α of the inner products $(u, g_\alpha), (g = g(x,0))$, u being the initial data, and finally reconstruct the weak solution at a later time τ from its inner products with the functions $g(x,\tau)$. These functions, $g(x,t)$, can be found at time $t = \tau$ if they are known at $t = 0$.

If this method is used with a small number of functions (i.e., small N), the solution at a time $t > 0$ will be underdetermined. In this case one can use the invariant measure μ to "fill in" the gaps through regression, i.e., replace the weak solution which is not completely known by its average (as determined by μ) over all solutions that satisfy the N conditions $(g_\alpha, u) = V_\alpha$. In other words, replace the function u which is not completely known by the regression $v = v(x,t) = \langle u \rangle_V$. Some technical background on regression follows in §3.

Note that so far, the construction resembles what is quite commonly done in underdetermined linear problems (for example, in the context of data assimilation [1]; the function g_α are analogous to the "representers" which are used there). The construction is computationally useful in certain linear problems even when alternate ways of finding future regressions are available, as happens whenever time evolution and averaging commute (see [5]). However, the construction just presented is restricted to linear problems, and the amount of work needed to evaluate and store the kernel functions g_α may not be trivial. It is also clear that not all of the available information has been used, as the evolution of the kernels g_α is independent of the invariant measure μ, and the measure used in the regression need not be connected with the differential equation. Indeed, in geophysical applications the measure is chosen according to considerations quite extraneous to the differential equations which may be only partly known.

We now wish to show how the machinery can be modified through the use of the invariant measure in the evolution equations for the conditions, so that it becomes more efficient as well as generalizable to nonlinear problems.

First, we must view the calculations differently. The measure μ induces a measure on the space of the data u that satisfy the initial conditions: the *conditional measure* μ_V. *The conditional measure is not invariant*; for example, if the initial conditions consist of N point-values of the functions $u(x,0)$, i.e., if we assume that at $t = 0$ the functions

satisfy $u(x_\alpha, t = 0) = V_\alpha$ (the functions g_α are then δ functions), there is no reason to believe that these conditions will be satisfied at all subsequent times with the same V_α by the solutions of the differential equations that arise out of these initial data.

Imagine first that one can sample the initial conditional measure μ_V; find the (in general, weak) solution of equation (2.1) that has this datum as initial condition, and perform this procedure repeatedly. At time t this produces an ensemble of functions $u(x,t)$ which inherits a measure from the initial data. In principle, this measure is well-determined; we wish to determine it in practice, and then to average with respect to this measure so as to obtain what we call an *optimal prediction.* Note that if the temporal evolution governed by (1.1) is ergodic with respect to the invariant measure, the conditional measure will eventually relax to the invariant measure and the initial conditions will be forgotten; we are in fact dealing with a computational analog of non-equilibrium statistical mechanics.

Given the initial conditional measure, we can find the statistics of Lu, or, in the general case (1.1), the statistics of $R(u)$ and consequently the statistics of u_t; thus, the evolution of the measure μ_V can be determined for a short time interval Δt. We cannot go beyond a short time interval because the measure μ_V at time Δt can no longer be described as the invariant measure conditioned by the conditions (1.2), at least not with the same functions g_α and the same V_α. This leads us to the closure assumption:

Assumption 2.1 (Closure) *The conditional probability measure at time t, $\mu_V(t)$, can be approximated by*

$$\mu_V(t) = \mu_{V(t)}, \tag{2.4}$$

where the left hand side $\mu_V(t)$ is the measure conditioned by the initial data (1.2), while the right hand side is the invariant measure conditioned by N affine conditions of the form

$$(g_\alpha(t), u(t)) = \int g_\alpha(x,t)u(x,t)\,dx = V_\alpha(t).$$

The kernels $g_\alpha(t)$ and the values $V_\alpha(t)$ of the inner products will generally be different at time t than at time $t = 0$, but it is assumed here that the affine form of the conditions and their number remain constant. We have already seen that in the linear case equation (2.4) is in fact a theorem (a different analysis of this fact was given in [5]).

It should be noted that even though this is the assumption that will be used in the present paper, it may be unduly restrictive in other situations; there is nothing magical about keeping the number of conditions fixed, and the conditions need not in general be affine.

What we wish to do is to advance the measure to time Δt and then to find the conditions that produce that new measure at the new time, when these conditions condition the invariant measure. In that way, the whole process can be taken a step further and then repeated as often as one may wish. In the linear case we have already produced a recipe for updating the conditions: there, we let g_α satisfy the adjoint equation and keep the numbers V_α fixed. We now propose different, *approximate*, ways

of finding conditions that describe the evolving measure. We are going to do so by matching moments; on one hand, we will calculate moments of the conditional measure by regression from the old moments, and on the other hand, we will produce conditions that produce the new moments from the invariant measure; this will produce equations of motion for the conditions.

More specifically, suppose that we compute Nq moments of u, at time Δt. For example, set $q = 2$, and compute the means and the variances with respect to the conditioned measure of the random variables $u(x, \Delta t)$ at each of the N points x_α. We can do this knowing the conditions at time $t = 0$ and the invariant measure. On the other hand, suppose we let the functions g_α depend on $q - 1$ parameters; for example, if $q = 2$, then we can pick $g_\alpha(x) = \exp((x - x_\alpha)^2/\sigma_\alpha)$, where the "centers" x_α are fixed and the numbers σ_α may be allowed to vary in time and will serve as our parameters. If we write down the requirement that the moments we calculated at time Δt match the moments produced by conditioning the invariant measure by affine conditions with unknown values of parameters such as the σ_α and unknown values of the right-hand sides V_α, we obtain Nq algebraic equations for the parameters, which we can try to solve. If we solve these equations successfully, we obtain a set of simultaneous *ordinary* differential equations for the parameters and the moments.

Before carrying out such a calculation, two remarks: There is no a priori guarantee that the algebraic equations we will obtain can be always solved: The basic assumption may fail, and the choice of parameters may be unsuitable. A good numerical program will inform us that a solution cannot be found. However, if a solution is found, the resulting moments are *realizable*. It is well-known that closures may well produce moments that not only fail to solve the problem at hand but do not solve any problem, because there is no stochastic process that admits the computed moments as its moments (see e.g.[22]). This is often a major difficulty in the formulation of mean equations, and it does not arise here.

We limit ourselves here to the simplest case with $q = 1$, i.e., for each $\alpha, \alpha = 1, ..., N$ we keep track of a single quantity, which we choose to be the mean value of the inner product of g_α and u, $\langle (g_\alpha, u) \rangle_V = (g_\alpha, \langle u \rangle_V)$, while at the same time we modify a single parameter in each condition; that single parameter is chosen to be the value V_α of the α-th product. Thus we must have for $\alpha = 1, ..., N$,

$$\frac{d}{dt} V_\alpha = \langle (g_\alpha, u_t) \rangle_V = \langle (g_\alpha, R(u)) \rangle_V = (g_\alpha, \langle R(u) \rangle_V),$$

where $R(u)$ is the right-hand side of equation (1.1). The final equation,

$$\frac{dV_\alpha}{dt} = (g_\alpha, \langle R(u) \rangle_V), \qquad \alpha = 1 \dots N, \tag{2.5}$$

is the main equation used in the present paper. The more transparent mathematical notation,

$$\frac{dV_\alpha}{dt} = E\left[(g_\alpha, R(u)) | V_1, \dots, V_N\right], \qquad \alpha = 1 \dots N,$$

makes explicit the fact that we are dealing with a system of N ordinary differential equations. Once the V_α are found at time t, a regression can be used to find the

average solution at any point x (see §3). We shall call an algorithm that uses, in the matching of moments and parameters, only means of the unknown solutions and no higher moments, a first-order prediction scheme; in the present paper this is the only prediction scheme we shall use. We hope to demonstrate that a first-order prediction is often an improvement over algorithms that take no cognizance of the invariant measure, in the sense that it requires less computational labor than the alternatives; higher-order and more sophisticated optimal prediction schemes will be described in subsequent work.

It is useful to contrast our algorithm with the one at the beginning of the section: We are keeping the kernels g_α fixed while changing the values V_α of the conditions, while the "natural", linear construction at the beginning of the section did the opposite.

When can we expect the first order optimal prediction scheme to be accurate? In a linear problem $u_t = Lu$, if the kernels g_α are eigenfunctions of the operator $L^\dagger$ adjoint to L, then one can readily see from the analysis at the beginning of the section that equations (2.5) are exact, and this remains true if the g_α span an invariant subspace of $u_t + L^\dagger = 0$. Lowest-order optimal prediction should be accurate as long as the g_α span a space that is approximately invariant under the flow induced by $L^\dagger$ (see [11],[12]). This remark also provides a recipe for choosing the kernels. Something similar remains true in nonlinear problems [7], [16],[26]: Define the space of functions spanned by the kernels g_α to be the resolved part of the solution; equations (2.5) should yield an accurate prediction of this resolved part (including a correct accounting for the effect of unresolved components on the resolved components) as long as there is no substantial transfer of information from the resolved part to the unresolved part and back; if such information transfer should occur, a correct description of the flow should include an additional, "memory", non-Markovian term. This remark also points out that equations (2.5) should not be accurate for very long times (because memory terms are important to the description of decay to equilibrium), and should be better at low temperatures T than at high temperatures (because the decay to equilibrium should be more rapid at high T).

There are other ways to ensure the accuracy of a first order optimal prediction scheme: The quantities $(g_\alpha, R(u))$ are, of course, random variables whose distribution depends on the measure μ_V. If the standard deviations of these variables are small, then equations (2.5) are be good approximations to the exact solution. Indeed, equation (2.5) merely equates these random variables to their means; a higher order approximation would take into account the variance of these variables as well. The smaller the variance of the variables $(g_\alpha, R(u))$, the smaller the error we expect; to some extent, we can control this variance by choosing the kernels appropriately. The larger the support of the kernels, the more these variables represent spatial averages, and the slower we may expect their variance to grow; thus if the scheme is to be accurate, it is most likely that the functions g_α should not be narrow, δ-function-like objects. Furthermore, the kernels should not have disjoint supports (for an analogous observation in the theory of vortex methods, see, e.g., [3]). Rigorous error bounds for the linear case can be found in [12]. These conditions allow us to call the variables (g_α, u) that appear in (1.2) *collective variables*; they are groupings of variables. Note that as the number N of collective variables increases, optimal prediction homes in an ever smaller set of initial data, and the variance of the variables $(g_\alpha, R(u))$ should decrease.

Finally, the presentation above started from a discussion of weak solutions, and indeed in all the examples below the solutions will be weak; why is that so? We shall present a detailed mathematical analysis elsewhere; here it should suffice to comment that if the solutions are not highly oscillatory on several levels, there is less interest in analyzing methods that fail to resolve them; solutions that do oscillate significantly on several scales appear, on the largest scale, as non-smooth and therefore weak.

3 Regression for Gaussian variables

It was mentioned in the previous section that once the data that condition the state of the system are found, or, at time $t = 0$ once the initial data have been chosen, the remainder of the solution can be replaced by a regression. Formula (2.5), the first order optimal prediction formula, is also a regression formula (an average conditioned by partial information). To illustrate regression, and more importantly, to remind the reader of formulas that will be used in the sequel, we collect in the present section some regression formulas for Gaussian measures, both for the discrete and the continuous case. More details can be found in standard books (e.g.[21] as well as in [6]).

We start by describing how to perform regressions on discrete sets of Gaussian (normal) variables. Let $\mathbf{u} = (u_1, \ldots, u_n)$ be a real vector of jointly normal random variables; it has a probability density $f(\mathbf{u})$ of the form,

$$P(s_1 < u_1 \leq s_1 + ds_1, \ldots, s_n < u_n \leq s_n + ds_n) = f(\mathbf{s})\, ds_1 \ldots ds_n = \\ = Z^{-1} \exp\left(-\frac{1}{2}(\mathbf{s}, A\mathbf{s}) + \mathbf{b}\cdot\mathbf{s}\right) ds_1 \ldots ds_n, \tag{3.1}$$

where Z is the appropriate normalization factor, $\mathbf{s} = (s_1, \ldots, s_n)$, and the $n \times n$ matrix A with entries a_{ij} is symmetric, positive definite, and has an inverse A^{-1} . The matrix A^{-1} is the pairwise covariance matrix with elements

$$a_{ij}^{-1} = \mathrm{Cov}\,\{u_i, u_j\} \equiv \langle u_i u_j \rangle - \langle u_i \rangle \langle u_j \rangle ,$$

where the brackets, $\langle\cdot\rangle$, denote averaging with respect to the probability density; the vector $\mathbf{b}$ with components b_i is related to the expectation values of $\mathbf{u}$, $\langle u \rangle = (\langle u_1 \rangle, \ldots, \langle u_N \rangle)$, by

$$A^{-1}\mathbf{b} = \langle u \rangle .$$

The distribution is fully determined by the n means and by the $\frac{1}{2}n(n+1)$ independent elements of the covariance matrix, and therefore all the expectation value of any observable can be expressed in terms of these parameters.

Next, we assume that the random vector $\mathbf{u}$ satisfies a set of conditions of the affine form,

$$g_{\alpha i} u_i = V_\alpha, \qquad \alpha = 1, \ldots, N < n, \tag{3.2}$$

where the index α enumerates the conditions and summation over repeated indices is assumed. Each equation in (3.2) is a discrete analog of one of the equations in

(1.2). The $N \times n$ matrix G, whose entries are $g_{\alpha i}$, determines the full set of conditions. To distinguish between the random variables $(u_1, \ldots, u_n)$, and the collective variables $(V_1, \ldots, V_N)$, we enumerate the former by Roman indices and the latter by Greek indices.

Our goal is to compute regressions, $E[\phi(u)|V]$, for various functions ϕ, i.e., conditional expectation values, or equivalently, averages over the functions that satisfy the conditions. We state three lemmas that will become handy below; for proofs, see [6].

Lemma 3.1 *The conditional expectation of the variables u_i is an affine function of the conditioning data V_α:*

$$\langle u_i \rangle_V = q_{i\alpha} V_\alpha + c_i, \tag{3.3}$$

where the $n \times N$ matrix Q whose entries are the $q_{i\alpha}$ and the n-vector $\mathbf{c}$ whose entries are the c_i are given by:

$$\begin{aligned} Q &= (A^{-1}G^\dagger)(GA^{-1}G^\dagger)^{-1}, \\ \mathbf{c} &= A^{-1}\mathbf{b} - (A^{-1}G^T)(GA^{-1}G^\dagger)^{-1}(GA^{-1}\mathbf{b}), \end{aligned} \tag{3.4}$$

where the dagger denotes a transpose.

Lemma 3.2 *The conditional covariance matrix has entries*

$$\begin{aligned} Cov\{u_i, u_j\}_V &= \langle u_i u_j \rangle_V - \langle u_i \rangle_V \langle u_j \rangle_V = \\ &= \left[A^{-1} - (A^{-1}G^\dagger)(GA^{-1}G^\dagger)^{-1}(GA^{-1})\right]_{ij}, \end{aligned} \tag{3.5}$$

where the subscript $[]_{ij}$ denotes the $\{ij\}$ component of a matrix.

Lemma 3.3 *Wick's theorem holds for constrained expectations, namely,*

$$\begin{aligned} &\left\langle \prod_{p=1}^{P} \left(u_{i_p} - \langle u_{i_p} \rangle_V\right) \right\rangle_V = \\ &= \begin{cases} 0 & P \text{ odd} \\ \sum_{perm} Cov\{u_{i_1}, u_{i_2}\}_V \cdots Cov\{u_{i_{P-1}}, u_{i_P}\}_V & P \text{ even} \end{cases} \end{aligned} \tag{3.6}$$

where the summation is over all possible pairings of the P coordinates that are in the list.

Equation (3.3) shows that conditioning data alter expectation values linearly in the V_α and independently of multiplicative factors in the covariances. Equation (3.5) shows that conditioned covariances are determined by the matrix G alone, without reference to the V_α. Equation (3.6) shows that the conditioned Gaussian distribution, while not satisfying the requirement that the covariance matrix be non-singular, retains a key property of Gaussian densities.

In the applications below we shall use Gaussian variables parameterized by a continuous variable x, i.e., Gaussian random functions $u = u(x)$. Their means, $\langle u(x) \rangle$, and covariances, $a^{-1}(x,y) = \langle u(x)u(y) \rangle - \langle u(x) \rangle \langle u(y) \rangle$, will be defined for all (x,y) in an appropriate range rather than only for integer values of (i,j). The matrix A^{-1} becomes

the integral operator whose kernel is a function a^{-1}. The kernel $a = a(x,y)$ of the operator A inverse to A^{-1} is defined by

$$\int a^{-1}(x,y)a(y,z)\,dy = \delta(x,z).$$

The vectors with entries $g_{\alpha i}$, become functions $g_\alpha(x)$, and the conditions (3.2) become equations (1.2). The regression formula, (3.3), then changes into

$$\langle u(x)\rangle_V = \langle u(x)\rangle + c_\beta(x)\left[V_\beta - \left\langle \int g_\beta(y)u(y)\,dy \right\rangle\right],$$

where

$$c_\beta(x) = \left\{\int a^{-1}(x,y)g_\alpha(y)\,dy\right\} m^{-1}_{\alpha\beta},$$

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

$$m_{\alpha\beta} = \int\int g_\alpha(x)a^{-1}(x,y)g_\beta(y)\,dx\,dy.$$

The formula for the constrained covariance can be obtained from (3.5) by replacing each i by an x, each j by a y, and each summation over a Latin index by the corresponding integration. Wick's theorem is still valid with the appropriate changes in notation; note that the Greek indices, which refer to the N initial data, remain integers.

4 An overview of invariant measures and Hamiltonian formalisms

Before proceeding with our numerical program, we summarize some material on Hamiltonian systems, invariant measures in finite and infinite dimensional systems, and the properties of certain measures. This material can be found in books on quantum field theory and related topics (see, e.g., [10, 18, 23]). More specific references will be given below; we take a very elementary point of view.

A Hamiltonian system is described in terms of n "position" variables, q_i, and their associated "momenta", p_i, $i = 1,\ldots,n$; a Hamiltonian function $\mathcal{H} = \mathcal{H}(q_i,p_i) = \mathcal{H}(q,p)$ is given, and the equations of motion are

$$\frac{dq_i}{dt} = \mathcal{H}_{p_i}, \quad \frac{dp_i}{dt} = -\mathcal{H}_{q_i}, \qquad i = 1,\ldots,n. \tag{4.1}$$

If the initial values of the $2n$ variables q,p are given, it is assumed that the system (4.1) has a global solution in time. Suppose the initial data are chosen at random in that $2n$ dimensional space, with a probability density $f(q,p,0)$; it is easy to check that the probability density of the q's and p's at time t satisfies the Liouville equation (see [10]):

$$f_t + \sum_i \left[\frac{dq_i}{dt} f_{q_i} + \frac{dp_i}{dt} f_{p_i} \right] = 0. \tag{4.2}$$

An invariant density is a probability density that does not depend on time, i.e., one that satisfies equation (4.2) with $f_t = 0$. Recall that by the definition of a probability density, $f \geq 0$, and $\int f \, dp \, dq = 1$ (with obvious notations). One can readily see that any function of $\mathcal{H}$ with these two properties is an invariant density; the one that is natural for physical reasons is $f = Z^{-1} \exp(-\mathcal{H}/T)$, where Z is a normalization constant and T is the "temperature". We set $T = 1$ unless specified otherwise. If the initial data are sampled from this initial distribution, and each of these samples is used as an initial datum for the equations of motion, then the probability distribution of the variables q and p at any later time t will be the same as it was initially. We now wish to generalize these notions of Hamiltonian systems and invariant distributions to the infinite-dimensional case, where the equations of motion will be partial differential equations and the invariant distributions will be called "invariant measures". We do so by way of an example that will be used in later sections.

Take the interval $[0, 2\pi]$ and divide it into n segments of length h. At each mesh point jh, $j = 1, \ldots, n$ define variables q_i, p_i ; introduce the "Hamiltonian"

$$\mathcal{H} = \sum_{i=1}^{n} h \left[\frac{(p_{i+1} - p_i)^2}{2h^2} + \frac{(q_{i+1} - q_i)^2}{2h^2} + F(p_i, q_i) \right], \tag{4.3}$$

where values of q, p outside the interval are provided by an assumption of periodicity, and the term in brackets is a Hamiltonian density. Consider the set of ordinary differential equations:

$$\frac{dp_i}{dt} = -\frac{1}{h} \mathcal{H}_{q_i} = \Delta_h q - F_{q_i}, \tag{4.4}$$

where $\Delta_h q = (q_{i+1} - 2q_i + q_{i-1})/h^2$, and similarly,

$$\frac{dq_i}{dt} = \frac{1}{h} \mathcal{H}_{p_i} = -\Delta_h p + F_{p_i}. \tag{4.5}$$

Note that the right hand side contains a factor h^{-1} that has no analog in the finite dimensional system above; its effect is to differentiate the Hamiltonian density rather than the Hamiltonian itself. This modification is needed to get self-consistent limits as $h \to 0$, a limit operation we shall now undertake (see [10, 18]). As $h \to 0$, these equations formally converge to:

$$\begin{cases} p_t = q_{xx} - F_q, \\ q_t = -p_{xx} + F_p, \end{cases} \tag{4.6}$$

or, writing $u = q + \imath p$ with imaginary $\imath$ and $F'(u) = F_q + \imath F_p$, we find an equation of Schrödinger type:

$$\imath u_t = -u_{xx} + F'(u). \tag{4.7}$$

The Hamiltonian, (4.3), converges formally to

$$\int_0^{2\pi} \left[\frac{1}{2}|u_x|^2 + F(u)\right] dx, \tag{4.8}$$

where the vertical lines denote a modulus. One has to examine what in these passages to a limit is justifiable, see e.g. [23] for a physics analysis, and [15] for a mathematical analysis.

For every finite h, the (finite dimensional) measure

$$Z^{-1} \exp\left[-h \sum \left(\frac{|u_{i+1} - u_i|^2}{2h^2} + F\right)\right], \tag{4.9}$$

is invariant for the system (4.4–4.5). Note that this is true both with and without the extra factor h^{-1} in equations (4.4–4.5). What is the limit of this measure as $h \to 0$? Set for a moment $F = 0$. The exponential in (4.3) factors into a product of terms each one of which is the exponential of a single difference quotient, of the form $\exp[-(p_{i+1} - p_i)^2/2h]$ or with q replacing p. Hence, the p variables are independent of the q variables. Furthermore, the "increments" $p_{i+1} - p_i$, $q_{i+1} - q_i$, are obviously Gaussianly distributed, have a variance proportional to the distance h, and are all independent of each other. Thus in the limit, the functions $q(x)$, (and similarly for $p(x)$), are made up of independent Gaussian increments. They differ from Brownian motion (see [9]) by being periodic rather than satisfying $q(0) = 0$ (they are "Brownian bridges" – which does not make a deep difference). Also, as long as $F = 0$, the common value of $q(0)$ and $q(2\pi)$ are undetermined because the exponent of the exponential is unchanged when one adds a constant to the q_i; one can remove this degeneracy by adding a term to the exponent that is sensitive to the value of $q(0)$. Thus, the limit of

$$Z^{-1} \exp\left\{-h \sum \left[\frac{(q_{i+1} - q_i)^2}{2h^2} + \frac{(p_{i+1} - p_i)^2}{2h^2} + F(q,p)\right]\right\} dq_1 \cdots dq_n dp_1 \cdots dp_n,$$

can be written as

$$dB_c \cdot \exp(-\int F dx), \tag{4.10}$$

where B_c is a suitably conditioned Brownian (Wiener) measure.

As is well-known (see, e.g., [9]), a sample Brownian path is, with probability one, nowhere differentiable. This fact is not changed by the factor $\exp(-\int F dx)$ in (4.10); thus if we sample initial data from (4.10) we obtain weak solutions of the equation of motion (4.7), as was indeed assumed in section 2. In addition, the integral $\int |u_x|^2 dx$ diverges, so that the limit in (4.3) is purely formal; its meaning is given by equation (4.10). An important consequence of these facts is that the exponential in (4.9) tends

to zero as $h \to 0$; this is indeed necessary if we are to have a reasonable function-space measure: As $h \to 0$, we have a measure on a space of increasing dimension, the density of functions that satisfy the set of inequalities $s_i < q_i \leq s_i + ds_i, i = 1, \ldots, n$, should decrease as n increases, thus $\exp(-\mathcal{H})$ should tend to zero and $\mathcal{H}$ should diverge.

Weak solutions that are spatially like Brownian motion are difficult to resolve; difference quotients do no converge, and the Fourier series expansions of these solutions converge slowly. We are thus consistent: our machinery applies in problems where it is indeed needed. Such problems are not exceptional; for example, in the vanishing-viscosity limit, the solutions of the Euler equations have a Hölder exponent of 1/3, i.e., they are even less smooth than Brownian motion (see [20]).

Finally, a computational comment that will be useful below: Brownian motions and Brownian bridges are easy to sample via interpolation formulas related to the regression formulas of the previous Section 3 (see e.g. [2],[17]); to modify the measures so as to take into account the factor $\exp(-\int F dx)$ requires simply that the samples be rejected or accepted with some form of a Metropolis algorithm.

5 Conditional expectations with a non-Gaussian prior

Our goal in this section is to introduce a systematic approach for solving the equations of optimal prediction (2.5) for nonlinear equations of the form (1.1). We assume that equation (1.1) has Hamiltonian form, i.e., we assume that there exists a Hamiltonian $\mathcal{H}$ such that (1.1) is the Hamilton equation of motion with this Hamiltonian. We can therefore assume the existence of a prior measure, $\mu(u)$ whose density f has the form

$$f(u) = Z^{-1} e^{-\mathcal{H}(u)}, \tag{5.1}$$

with Z a normalization constant.

In order to solve equation (2.5), we must compute the conditional expectations on its RHS,

$$(g, \langle R(u) \rangle_V), \tag{5.2}$$

with R being the RHS of equation (1.1). This task is relatively simple when the prior measure μ is Gaussian; we address here the problem of what to do when it is not. The method we present is based on perturbation theory, see e.g. [8, 13, 14]. The idea is to reduce the computation of (5.2) to the computation of regressions with respect to a Gaussian measure via a perturbation expansion.

We start by splitting the Hamiltonian into two parts,

$$\mathcal{H} = \mathcal{H}^0 + \mathcal{H}^1. \tag{5.3}$$

Here $\mathcal{H}^0$ is quadratic (producing a Gaussian measure μ^0) and $\mathcal{H}^1$ is a non-quadratic perturbation. A conditional expectation of a functional $\mathcal{F}(u)$ is defined as

$$\langle \mathcal{F} \rangle_V = Z^{-1} \int \mathcal{F}(u) d\mu_V, \tag{5.4}$$

with the normalization constant $Z = \int d\mu_V$. Averages with respect to Gaussian measures will be singled out by a superscript 0:

$$\langle \mathcal{F} \rangle_V^0 = Z_0^{-1} \int \mathcal{F}(u) d\mu_V^0, \tag{5.5}$$

with $Z_0 = \int d\mu_V^0$. Based on the Hamiltonian split, (5.3), we can carry out the expansion

$$e^{-\mathcal{H}} = e^{-\mathcal{H}^0} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (\mathcal{H}^1)^k, \tag{5.6}$$

which can be then utilized to write

$$\int \mathcal{F}(u) d\mu_V = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int \left[\mathcal{F}(u) (\mathcal{H}^1)^k \right] d\mu_V^0. \tag{5.7}$$

Hence, combining (5.4), (5.5) and (5.7) we find

$$\langle \mathcal{F} \rangle_V = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{Z_0}{Z} \left\langle (\mathcal{H}^1)^k \mathcal{F} \right\rangle_V^0.$$

The ratio Z_0/Z is:

$$\left(\frac{Z_0}{Z} \right)^{-1} = \frac{\int d\mu_V}{\int d\mu_V^0} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\int (\mathcal{H}^1)^k d\mu_V^0}{\int d\mu_V^0} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left\langle \mathcal{H}^1 \right\rangle_V^0,$$

and therefore:

$$\langle \mathcal{F} \rangle_V = \frac{\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left\langle (\mathcal{H}^1)^k \mathcal{F} \right\rangle_V^0}{\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left\langle (\mathcal{H}^1)^k \right\rangle_V^0}.$$

In particular, for $\alpha = 1, \ldots, N$,

$$(g_\alpha, \langle R(u) \rangle_V) = \frac{\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left(g_\alpha, \left\langle R(u) \left(\mathcal{H}^1 \right)^k \right\rangle_V^0 \right)}{\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left\langle (\mathcal{H}^1)^k \right\rangle_V^0}. \tag{5.8}$$

The RHS of (5.8) is now written in terms of expectation values that we already know how to compute since they are averages with respect to a Gaussian measure. Note that the division by Z/Z_0 can be avoided by removing certain terms from the numerator of (5.8); indeed, (5.8) is a conditioned expansion in Feynman diagrams and it can be normalized by removing unconnected graphs (see [8, 13, 14]). We choose not to explain this fact here. Note also that the leading term in the expansion, (corresponding to $k = 0$), where the measure μ_V is simply replaced by μ_V^0, already contains a contribution of the nonlinear terms in the equation of motion.

Computing a finite number of terms in the series expansion (5.8) can still be a relatively complicated task, and we demonstrate in §6 a step-by-step solution of a model

problem. The reader should not be unduly worried by the complexity of some of the expressions, because : (i) To make the exposition as clear as possible, no advantage is being taken here of various ways of simplifying the expressions, such as, e.g., using orthogonal functions for the g_α; (ii) much of the algebra can be automated (see below).

The partition of $\mathcal{H}$ into Gaussian and non-Gaussian parts is not unique, and in section 6 and 7 below we shall discuss some ways to optimize it in order to gain accuracy.

6 A perturbative treatment of a nonlinear Schrödinger equation

We now utilize the perturbation method to approximating solutions of the non-linear Schrödinger equation of the form (4.7),

$$\imath u_t = -u_{xx} + \frac{1}{4}\left[3\left|u\right|^2 u + u^{*3}\right], \quad u = q + \imath p, \tag{6.1}$$

in the interval $[0, 2\pi]$, with periodic boundary conditions. Equation (6.1) can also be written as the pair of equations,

$$\begin{cases} p_t = q_{xx} - q^3, \\ \\ q_t = -p_{xx} + p^3. \end{cases} \tag{6.2}$$

The corresponding Hamiltonian is

$$\mathcal{H}(p,q) = \frac{1}{2}\int \left(p_x^2(x) + q_x^2(x) + \frac{1}{2}\left[p^4(x) + q^4(x)\right]\right) dx, \tag{6.3}$$

and hence, equations (6.2) preserve the canonical density

$$f_0(p,q) = Z^{-1} e^{-\mathcal{H}(p,q)}.$$

(Note that the temperature T has been set equal to 1). To simplify the example, we follow [6] and use the same kernels in the definition of the collective variables for p and q, i.e., the collective variables are

$$\{U_\alpha^p, U_\alpha^q\} = \{(g_\alpha, p), (g_\alpha, q)\}, \qquad \alpha = 1, \ldots, N, \tag{6.4}$$

and their initial values, V_α^p, V_α^q, are given. The system of ordinary differential equations arising out of equation (2.5) for the V_α^p, V_α^q is:

$$\begin{cases} \dfrac{dV_\alpha^p}{dt} = +\left(g_\alpha, \dfrac{\partial^2}{\partial x^2}\langle q\rangle_V\right) - \left(g_\alpha, \langle q^3\rangle_V\right), \\ \\ \dfrac{dV_\alpha^q}{dt} = -\left(g_\alpha, \dfrac{\partial^2}{\partial x^2}\langle p\rangle_V\right) + \left(g_\alpha, \langle p^3\rangle_V\right). \end{cases} \qquad \alpha = 1, \ldots, N. \tag{6.5}$$

We therefore have to compute the four terms $(g_\alpha, \partial_{xx}\langle q\rangle_V)$, $(g_\alpha, \partial_{xx}\langle p\rangle_V)$, $(g_\alpha, \langle q^3\rangle_V)$ and $(g_\alpha, \langle p^3\rangle_V)$.

The first step is to split the Hamiltonian into two parts - a quadratic and a non-quadratic part, $\mathcal{H} = \mathcal{H}^0 + \mathcal{H}^1$; This will be done in the next section. We will just note at that point, that $\mathcal{H}^0$ will be of the form $\mathcal{H}^0 = \frac{1}{2}\int [p_x^2(x) + q_x^2(x) + m_0^2(p^2(x) + q^2(x))]\, dx$. Once this is done, The RHS of (6.5) can be computed following (5.8). For example, the term $(g_\alpha, \langle p^3\rangle_V)$ can be obtained by substituting p^3 for $R(u)$ in that equation. In particular, the zeroth, leading term is:

$$\left(g_\alpha, \langle p^3\rangle_V^0\right), \tag{6.6}$$

(note that this term already includes a nonlinear effect); a first-order (in the perturbation series) approximation will add the term

$$\frac{\left(g_\alpha, \langle p^3\mathcal{H}^1\rangle_V^0\right)}{\left(1 - \langle\mathcal{H}^1\rangle_V^0\right)}, \tag{6.7}$$

and so on, with similar expressions for the rest of the terms on the RHS of (6.5). Note that all these expansions are used to solve the equations of first-order prediction; in principle, we can perform higher-order predictions by using higher moments in setting up the matching of conditions before equation (2.5), and then improve the evaluation of the right-hand side of equation (2.5) be using more terms in the perturbation expansion. We consider here only the latter possibility.

The leading-order terms (6.6) can be computed using the results of §3. For the first-order term, (6.7), we have

$$\langle p^3\mathcal{H}^1\rangle_V^0 = \frac{1}{4}\left\langle p^3 \int \left[p^4 + q^4 - 2m_0^2(p^2 + q^2)\right] dz\right\rangle_V^0;$$

and we therefore have to compute

$$\left\langle p^3 \int p^s dz\right\rangle_V^0 = \int_z \langle p^3(x)p^s(z)\rangle_V^0\, dz, \qquad \left\langle p^3 \int q^s dz\right\rangle_V^0 = \int_z \langle p^3(x)q^s(z)\rangle_V^0\, dz$$

for $s = 2, 4$. To summarize, the RHS of (6.5), involves integration (with respect to z) and differentiation (with respect to x) of the following (with $j = 1, 3$ and $s = 2, 4$)

$$\langle p^j(x)p^s(z)\rangle_V^0, \quad \langle p^j(x)q^s(z)\rangle_V^0, \quad \langle q^j(x)p^s(z)\rangle_V^0, \quad \langle q^j(x)q^s(z)\rangle_V^0. \tag{6.8}$$

The terms in (6.8) can be computed by application of Wick's theorem (Lemma 3.3); the algebra can be performed with aid of a symbolic computer program (such as Math-

ematica). In particular, one obtains the identities:

$$\begin{aligned}
&\langle p(x)\, p^2(z)\rangle = -2\,\langle p(x)\rangle\,\langle p(z)\rangle^2 + 2\,\langle p(z)\rangle\,\langle p(x)p(z)\rangle + \langle p(x)\rangle\,\langle p^2(z)\rangle\,.\\
&\langle p(x)\, p^4(z)\rangle = 6\,\langle p(x)\rangle\,\langle p(z)\rangle^4 - 8\,\langle p(z)\rangle^3\,\langle p(x)p(z)\rangle -\\
&\quad -12\,\langle p(x)\rangle\,\langle p(z)\rangle^2\,\langle p^2(z)\rangle + 12\,\langle p(z)\rangle\,\langle p(x)p(z)\rangle\,\langle p^2(z)\rangle + 3\,\langle p(x)\rangle\,\langle p^2(z)\rangle^2\,.\\
&\langle p^3(x)\, p^2(z)\rangle = -12\,\langle p(x)\rangle^2\,\langle p(z)\rangle\,\langle p(x)p(z)\rangle + 6\,\langle p^2(x)\rangle\,\langle p(z)\rangle\,\langle p(x)p(z)\rangle +\\
&\quad + \langle p(x)\rangle^3\left(6\,\langle p(z)\rangle^2 - 2\,\langle p^2(z)\rangle\right) +\\
&\quad + \langle p(x)\rangle\left[6\,\langle p(x)p(z)\rangle^2 + \langle p^2(x)\rangle\left(-6\,\langle p(z)\rangle^2 + 3\,\langle p^2(z)\rangle\right)\right].\\
&\langle p^3(x)\, p^4(z)\rangle = 72\,\langle p(x)\rangle^2\,\langle p(z)\rangle\,\langle p(x)p(z)\rangle\left(\langle p(z)\rangle^2 - \langle p^2(z)\rangle\right) +\\
&\quad + \langle p(x)\rangle^3\left(-20\,\langle p(z)\rangle^4 + 36\,\langle p(z)\rangle^2\,\langle p^2(z)\rangle - 6\,\langle p^2(z)\rangle^2\right) +\\
&\quad +12\,\langle p(z)\rangle\,\langle p(x)p(z)\rangle\left[2\,\langle p(x)p(z)\rangle^2 + \langle p^2(x)\rangle\left(-2\,\langle p(z)\rangle^2 + 3\,\langle p^2(z)\rangle\right)\right].+\\
&\quad +9\,\langle p(x)\rangle\left[4\,\langle p(x)p(z)\rangle^2\left(-2\,\langle p(z)\rangle^2 + \langle p^2(z)\rangle\right) +\right.\\
&\quad \left. + \langle p^2(x)\rangle\left(2\,\langle p(z)\rangle^4 - 4\,\langle p(z)\rangle^2\,\langle p^2(z)\rangle + \langle p^2(z)\rangle^2\right)\right]. \qquad (6.9)
\end{aligned}$$

All the averages in equation (6.9) are of course conditioned by V; the constant repetition of the subscript V and the superscript 0 has been avoided for esthetic reasons.

We now pick the kernels $g_\alpha(x)$ to be translates of a fixed function $g(x)$, i.e., $g_\alpha(x) = g(x - x_\alpha)$, which is a normalized (not random!) Gaussian with periodic images and width σ:

$$g(x) = \frac{1}{\sqrt{\pi}\sigma}\sum_{\tau=-\infty}^{\infty}\exp\left[-\frac{(x-2\pi\tau)^2}{\sigma^2}\right]. \qquad (6.10)$$

We note that the Fourier representation of $g(x)$ is $g(x) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{\imath kx - \frac{1}{4}k^2\sigma^2}$. Given this choice of kernel functions we can write

$$\langle p(x)\rangle_V^0 = \sum_{\alpha=1}^{N} c_\alpha^{pp}(x) V_\alpha^p, \qquad (6.11)$$

where

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

and

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

This choice of kernels is the same as in previous work [5, 6]. It is far from optimal in the context of perturbation theory; in particular, orthogonal kernels such as the Fourier kernels used in section 7 below reduce the number of non-zero terms in the expansion. We thought that we should present at least once a perturbative calculation with a general kernel.

6.1 The partition of the Hamiltonian

We now turn to the question of how exactly the Hamiltonian should be divided into a sum of a quadratic part and a perturbation, $\mathcal{H} = \mathcal{H}^0 + \mathcal{H}^1$. Of course we wish $\mathcal{H}^1$ to be as small as possible, so as to have a perturbation series that behaves as well as possible; we therefore perform a partition with a few free parameters over which we shall minimize $\mathcal{H}^1$; we choose to write:

$$\begin{aligned} \mathcal{H}^0 &= \frac{1}{2}\int_0^{2\pi} \left(|u_x|^2 + m_0^2 |u|^2 + b\right) dx, \\ \mathcal{H}^1 &= \frac{1}{2}\int_0^{2\pi} \left(F - m_0^2 |u|^2 - b\right) dx, \end{aligned} \tag{6.12}$$

where, as before, $u = q + \imath p$, $F = \frac{1}{2}(p^4 + q^4)$. There are no odd powers in the partition because the measure is invariant under the reflection $u \leftrightarrow (-u)$; note that the term $m_0^2 |u|^2$ removes the indeterminacy in the Gaussian measure defined by $\mathcal{H}^0$. This is not the only partition that can be considered; one could for example add and subtract squares of fractional derivatives of u. The task at hand is to choose good values for the parameters m_0 and b. They cannot in general be chosen so as to make the perturbation series convergent; what one would really want is to make the measure defined by $\mathcal{H}$ and conditioned by the V_α be a small perturbation of the conditional measure defined by $\mathcal{H}^0$, and while this can in principle be done, and would presumably lead to time-dependent equations for m_0 and b, it is reasonable, as a first try, to choose m_0 and b so as to minimize $\langle \mathcal{H}^1 \rangle, \langle (\mathcal{H}^1)^2 \rangle$, where the averages are unconditional. Note also that the presence of the term b leaves all averages with respect to the measure defined by $\mathcal{H}^0$ unchanged—it gets absorbed into the normalization constant Z—but it does affect the expansion of $\exp(-\mathcal{H}^1)$.

A straightforward algebraic manipulation, simplified by the fact that p and q are uncorrelated and that $\langle p^s \rangle = \langle q^s \rangle$ for all s, leads to

$$\begin{aligned} \left\langle (\mathcal{H}^1)^2 \right\rangle^0 = \frac{1}{16}\int_{z=0}^{2\pi}\int_{x=0}^{2\pi} dx\, dz \Big[& 2\left\langle p^4(x)p^4(z)\right\rangle^0 + 2\left(\left\langle p^4(x)\right\rangle^0\right)^2 - \\ & -8m_0^2\left\langle p^4(x)p^2(z)\right\rangle^0 - 8m_0^2\left\langle p^4(x)\right\rangle^0\left\langle p^2(z)\right\rangle^0 + \\ & +8m_0^4\left(\left\langle p^2(x)\right\rangle^0\right)^2 + 8m_0^4\left\langle p^2(x)p^2(z)\right\rangle^0 \Big], \end{aligned} \tag{6.13}$$

where the $\langle\cdot\rangle^0$, denotes average with respect to the unconditional Gaussian measure. Using the expressions (6.9), one can numerically minimize (6.13) as a function of m_0. This minimum is obtained at $m_0 \approx 1.055$.

After setting m_0, we are free to pick the second constant, b, in (6.12). Changing b will not affect the variance which we just computed in (6.13). We choose to set the first moment of the perturbation to zero, i.e.,

$$\langle \mathcal{H}^1 \rangle^0 = 0. \tag{6.14}$$

Given the partition, (6.12), it is clear that (6.14) holds if

$$b\pi = \frac{1}{4} \left\langle \int \left[p^4 + q^4 - 2m_0^2(p^2 + q^2) \right] dx \right\rangle^0. \tag{6.15}$$

With the aid of Wick's theorem and equation (6.11), equation (6.15) can be rewritten as

$$b\pi = \left[\frac{3}{4\pi} \sum_{k=-\infty}^{\infty} \frac{1}{k^2 + m_0^2} - m_0^2 \right] \cdot \sum_{k=-\infty}^{\infty} \frac{1}{k^2 + m_0^2}. \tag{6.16}$$

If one chooses $m_0 = 1.055$ so as to minimize the second moment of $\langle \mathcal{H}^1 \rangle$, the RHS of (6.16) equals -1.197, which in turn determines $b = -0.381$ so that the first moment of the perturbation will vanish. Note that even the zero-th order expansion uses information about the higher-order terms, since the parameters that determine the partition and thus the zero-th order term depend on an analysis of the later terms.

One should also note that once the future conditions have been determined, one has to perform further regressions to obtain the mean solutions at various points in space; the machinery there is exactly analogous to what has just been done, and will not be spelled out here.

6.2 Numerical checks

We now check the perturbation series as well as the optimal prediction scheme by comparing the results they give with numerical results obtained at substantial expense by sampling the initial conditions, solving the differential equations over and over, and averaging. We concentrate in the present section on the case $N = 2$, i.e., a case where we have initially as data only four values of collective variables, and are trying to find the mean future conditioned by these four values. We display only the variation in time of these collective variables, $U_\alpha^p, U_\alpha^q, \alpha = 1, 2$, which were defined in equation (6.4); The kernels are taken as $g_\alpha(x) = g(x - x_\alpha)$ with $x_1 = \pi/8$ and $x_2 = 9\pi/8$. The parameter σ is set as π. We first use the perturbation series with the optimal values $m_0 = 1.055$ and $b = -0.381$ computed in the previous section 6.1. We then display results obtained with different values of m_0 and b. Specifically, we choose $m_0 = 0.9$ and $b = 0$, obtained by splitting the Hamiltonian into

$$\begin{aligned} \mathcal{H}^0 &= \frac{1}{2} \int_0^{2\pi} \left(|u_x|^2 + m_0^2 |u|^2 \right) dx, \\ \mathcal{H}^1 &= \frac{1}{2} \int_0^{2\pi} \left(F - m_0^2 |u|^2 \right) dx, \end{aligned}$$

and requiring only that the first moment of the perturbation vanish, i.e., $\langle \mathcal{H}^1 \rangle^0 = 0$ (compare with (6.12) and (6.14)).

We check these results by replacing the continuum equations (6.2), (6.5), by a formal finite difference approximation conditioned by discrete forms of these equations, and then display the convergence of the conditioned mean of many solutions of the difference equations to the optimal prediction obtained by the perturbation analysis. Specifically, we replace equation (6.2) by the following difference equations,

$$\begin{cases} \dfrac{dp(j)}{dt} = \dfrac{q(j-1) - 2q(j) + q(j+1)}{h^2} - q^3(j), \\ \dfrac{dq(j)}{dt} = -\dfrac{p(j-1) - 2p(j) + p(j+1)}{h^2} + p^3(j), \end{cases} \qquad j = 1, \dots, n, \tag{6.17}$$

where $h = 2\pi/n$ is the mesh size. The conditions (6.4) are replaced by the following discrete approximations:

$$U_\alpha^p = \sum_{j=1}^{n} h g_\alpha(j) p(j), \quad U_\alpha^q = \sum_{j=1}^{n} h g_\alpha(j) q(j), \qquad \alpha = 1, 2, \tag{6.18}$$

(A factor h has been introduced in the definition of the collective variables to allow them to converge to the continuum collective variables $(g_\alpha, u) = \int g_\alpha(x) u(x) dx$.)

The Hamiltonian (6.3) is replaced by the discrete Hamiltonian:

$$\mathcal{H}(p,q) = \tag{6.19}$$
$$= \frac{h}{2} \sum_{j=1}^{n} \left\{ \left[\frac{p(j+1) - p(j)}{h} \right]^2 + \left[\frac{q(j+1) - q(j)}{h} \right]^2 + \frac{1}{2} [p^4(j) + q^4(j)] \right\}$$

We present results obtained with two mesh sizes, corresponding to $n = 8$ and $n = 16$. We also checked that the results with $n = 32$ are very close to the results with $n = 16$. For each mesh size h we use a Metropolis Monte-Carlo algorithm to find 5000 initial data drawn from the distribution defined by (6.19) conditioned by the values of the collective variables, integrate the equations in time up to $t = 1$, and average the results at various fixed time intervals. This numerical calculation is very costly, even for moderate values of n, but it is independent both of the perturbative analysis and of the machinery of optimal prediction. It is important to note that as the mesh size h tends to zero, the results of a standard (i.e., non-averaged) finite-difference solution of equation (6.2) with data drawn from the distribution (6.19) diverge pointwise as $h \to 0$. Conditional averaging provides the only meaningful numerical solution of equations (6.17) for such initial data.

In Figure 6.1 we present the evolution in time of the four collective variables U_1^p, U_2^p, U_1^q and U_2^q with the optimal $m_0 = 1.055$, $b = -0.38$. Figure 6.2 presents the plots corresponding to the choice $m_0 = 0.9$ and $b = 0$. The zeroth-order solution is the optimal prediction solution obtained with only the zeroth, leading, term in the perturbation expansion (see, e.g., equation (6.6)). The first-order solution is the optimal prediction solution obtained after adding a first-order correction to the perturbation series. Note

however that the optimal choice of parameters uses information about terms of order one.

In both figures, the solid lines are the solutions obtained with the optimal prediction equations, while the dotted lines represent an average over over 5000 solutions that evolve from initial data sampled from the conditioned discrete Hamiltonian (6.19).

Clearly, there is an improvement when one uses additional terms in the perturbation series. Also, even though the individual numerical solutions converge only weakly to a continuum limit, the average over numerical solutions with data sampled from the discrete Hamiltonian on one hand and the solution of the optimal prediction equations on the other hand get close as n increases. The perturbation expansion converges rapidly; the key graph is the one on the bottom right of Figure 6.1: the comparison between the expansion up to first-order with the average numerical solution with $n = 16$. One conclusion we draw from these graphs and use in section 7 below is that with the optimal partition of the Hamiltonian one can obtain an accurate solution with only the zero-th term in the expansion; the computation of the optimal parameters in the expansion uses the first-order term.

The limitations of first-order optimal prediction are displayed in Figure 6.3, where the integrations are carried out to longer times. We are working with a temperature $T = 1$, i.e., the fluctuation in u are of order 1; by contrast, in the longer runs of [5],[6] we used a smaller temperature $T = \pi/15$; we also have only 4 collective variables, not enough at this temperature to keep the variances of the collective variables small. In Figure 6.3 the optimal prediction solution is based on the optimal choice of $m_0 = 1.055$, $b = -0.38$, and the discrete solution is presented for $n = 16$. Once again, the numerical solution is an average over 5000 individual solutions.

The effect of temperature is displayed in Figure 6.4, where we present the time evolution of the four collective variables, U_α^p, U_α^q up to time $t = 5$. Both graphs are for the same initial values of U_α^p and U_α^q but differ in the temperature which determines the distribution of initial data. The graph on the left corresponds to low temperature, $T = 0.2$, whereas the graph on the right corresponds to high temperature, $T = 4$. As expected, the higher the temperature, the faster is the decay towards equilibrium; the collective variables tend faster towards their equilibrium value of zero, and the first-order prediction scheme with a small, fixed number of conditions loses accuracy faster. There are two ways to improve the prediction: Go to more sophisticated prediction theory, as outlined in section 2, or increase the number of collective variables. The first alternative will be explored in later publications; the value of the second approach will be shown in the next section, with a choice of kernels that reduces the amount of labor and also makes possible an analytical estimate of the difference between optimal prediction and a simple scheme that makes no use of the prior measure. Note that the optimal prediction runs yield good results when the standard deviation of the values of the collective variables is as large as 50 percents of their mean (the standard deviation of the pointwise values of the solutions is much larger still).

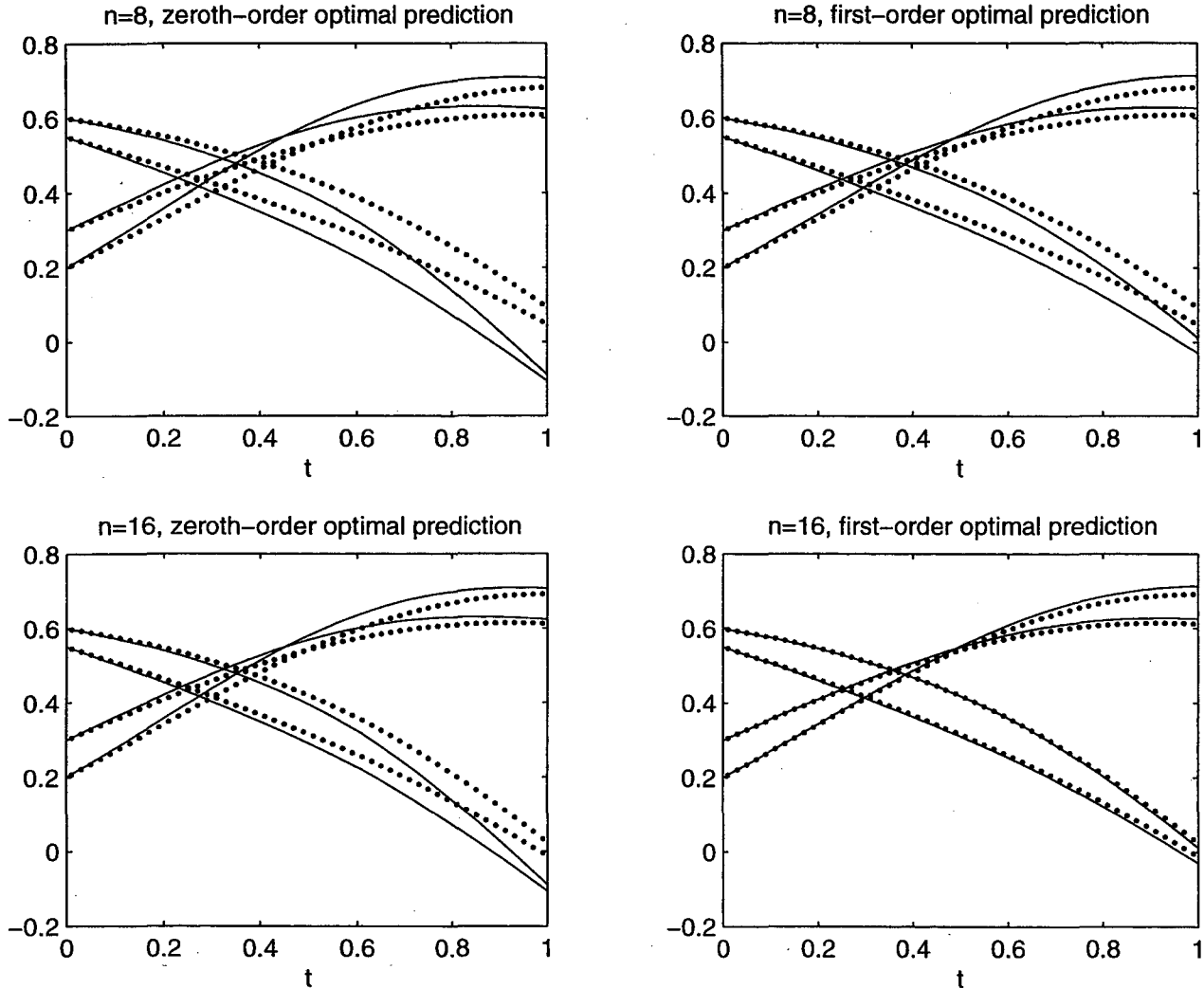


Figure 6.1: Time evolution of four collective variables U_1^p, U_2^p, U_1^q and U_2^q for the nonlinear Schrödinger equation (6.2),(6.17), with the optimal $m_0 = 1.055$, $b = -0.38$; Solid lines – optimal prediction equations. Dotted lines – average over 5000 solutions obtained from initial data sampled from the discrete Hamiltonian (6.19) with $n = 8$ and $n = 16$ points.

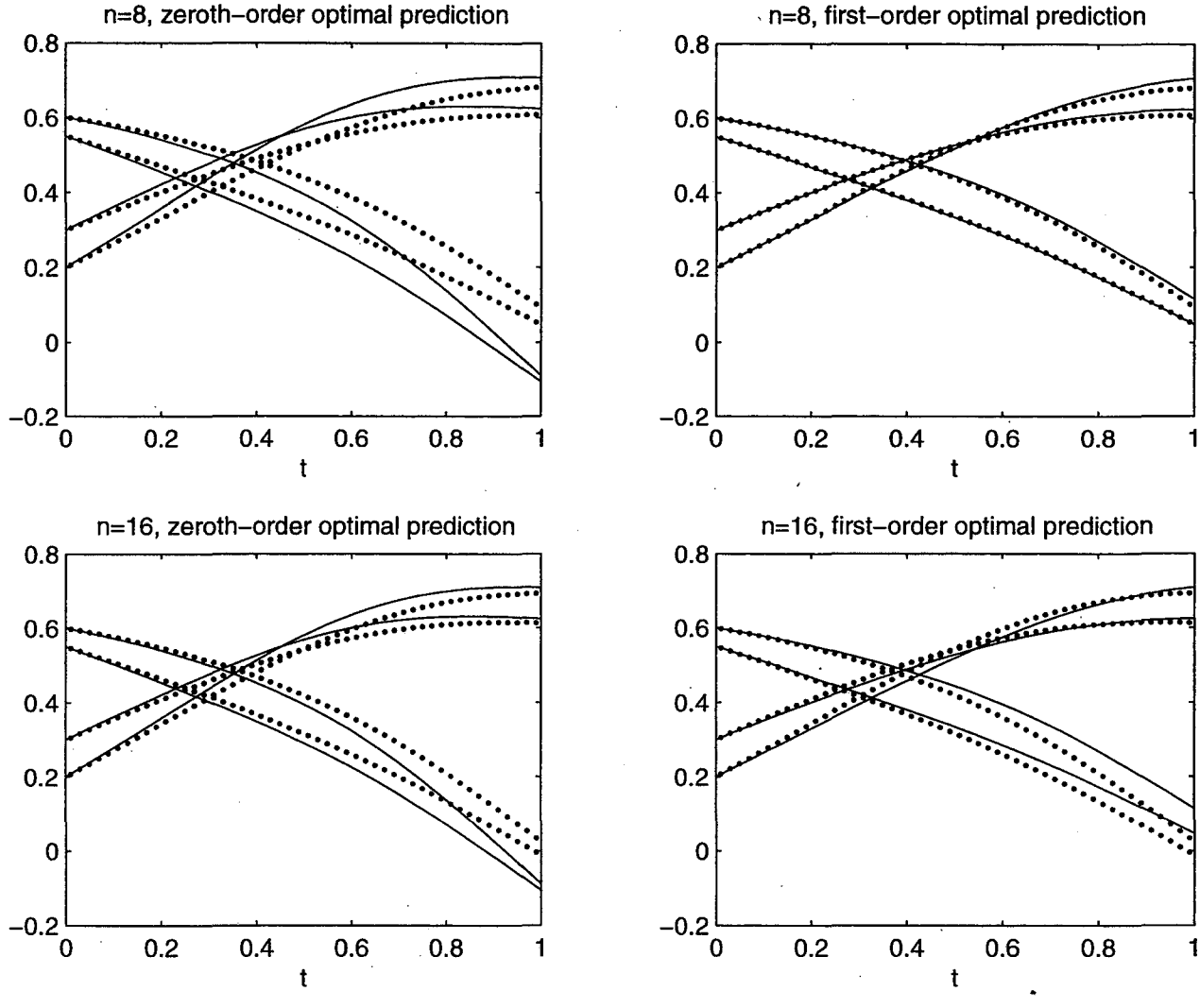


Figure 6.2: Time evolution of four collective variables U_1^p, U_2^p, U_1^q and U_2^q for the nonlinear Schrödinger equation (6.2),(6.17), with $m_0 = 0.9$, $b = 0$; Solid lines – optimal prediction equations. Dotted lines – average over 5000 solutions obtained from initial data sampled from the discrete Hamiltonian (6.19) with $n = 8$ and $n = 16$ points.

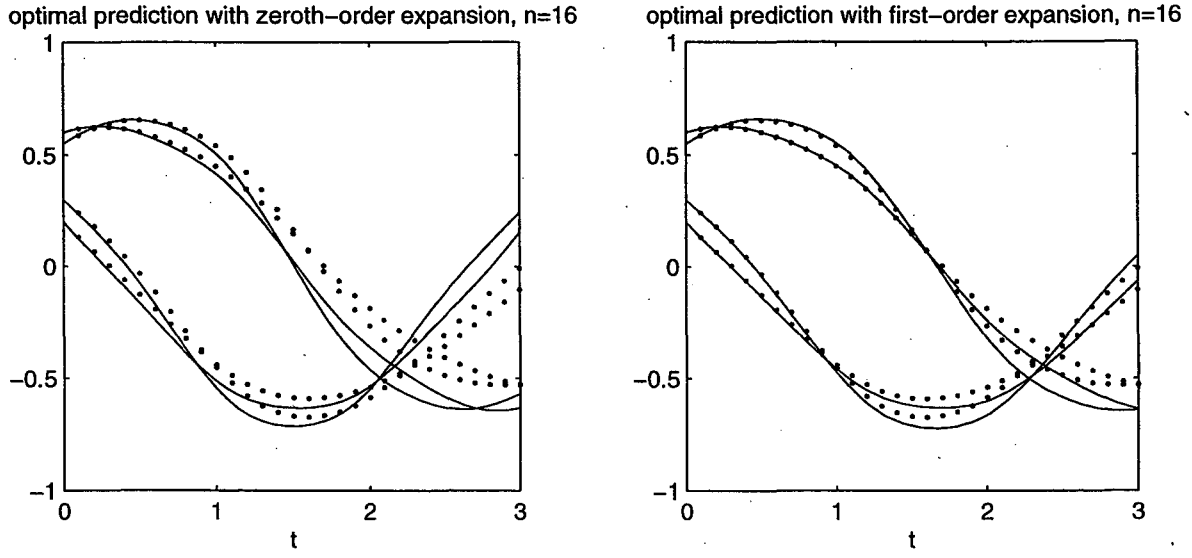


Figure 6.3: Longer time evolution of four collective variables U_1^p, U_2^p, U_1^q and U_2^q for the nonlinear Schrödinger equation (6.2),(6.17), with the optimal $m_0 = 1.055$, $b = -0.38$; Solid lines – optimal prediction equations. Dotted lines – average over 5000 solutions obtained from initial data sampled from the discrete Hamiltonian (6.19) with $n = 16$ points.

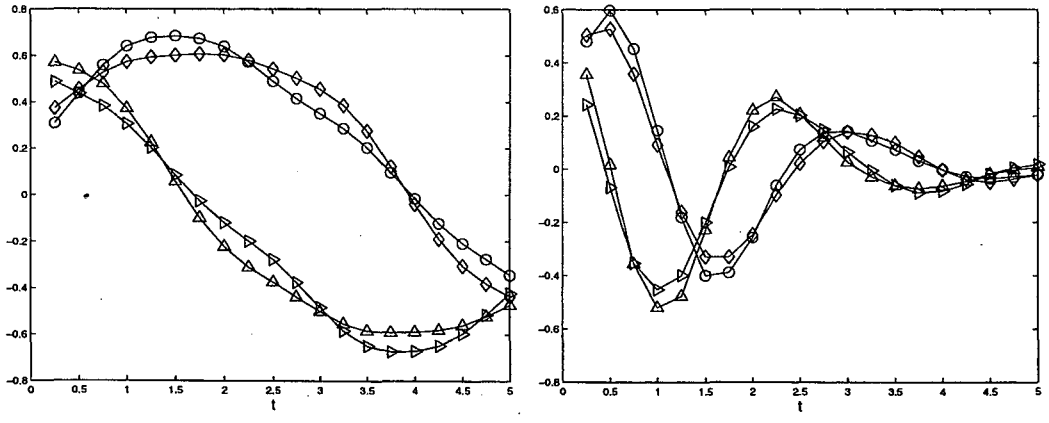


Figure 6.4: Long time evolution of the collective variables for initial distributions of the form $e^{-\mathcal{H}/T}$ with (a) $T = 0.2$ and (b) $T = 4$.

7 Pseudo-spectral optimal prediction for a model nonlinear problem

In the present section we consider a discrete version of the same Schrödinger equation as above; our goal is to show explicitly how the information in the prior measure improves the accuracy of an underresolved nonlinear calculation. One of the striking facts shown by the example is that first-order optimal prediction is useful in nonlinear problems. In view of the rapid convergence of the conditional expectations of discrete problems to their continuum limits, as displayed in the previous section, we shall be content with the discrete problem. Specifically, we shall contrast the solution of a discrete problem by a pseudo-spectral method that takes no cognizance of the prior measure with a closely related optimal prediction scheme with Fourier kernels, and show how the information in the prior measure improves the predictions. In the present section, a complex-function formalism turns out to be more transparent, and we therefore slightly change the notations. We consider a set of complex ordinary differential equations which is a formal discretization of our Schrödinger equation, (6.1),

$$\imath \frac{\partial u_j}{\partial t} = -\frac{u_{j-1} - 2u_j + u_{j+1}}{h^2} + \frac{1}{4}\left[3\left|u_j\right|^2 u_j + (u_j^*)^3\right], \tag{7.1}$$

with $j = 1, \ldots, n$ and $h = 2\pi/n$. Equations (7.1) are the Hamilton equations of motion derived from the following Hamiltonian,

$$\mathcal{H}(u) = \frac{1}{2}\sum_{j=1}^{n}\left\{\frac{\left|u_{j+1} - u_j\right|^2}{h} + \frac{h}{16}\left[u_j^4 + 6\left|u_j\right|^4 + (u_j^*)^4\right]\right\} \tag{7.2}$$

(see section 4 above). One can readily verify that this is the same discrete Hamiltonian as before. The prior measure is the canonical measure whose density is

$$f(u) = \frac{1}{Z}e^{-\mathcal{H}(u)}. \tag{7.3}$$

Note that we write u without boldface, as we did in the case of functions, but not as we did for vectors; this is done by analogy with the previous sections on the Schrödinger equation. (Those who look at our earlier papers [4],[5],[6] will notice that the measure here differs from the measure used there by a factor h in the exponent, and also that we used a different spatial period.) Equation (7.3) requires an explanation when the variables u_j are complex. A set of n complex variables can be treated as a set of $2n$ real variables. However, when the real and imaginary parts q_i, p_i of each variable u_i are independent and their means are zero, (as they are here), one can write:

$$\begin{aligned} &P(s_1 < q_1 \le s_1 + ds_1, \ldots, r_1 < p_1 \le r_1 + dr_1, \ldots, r_n < p_n \le r_n + dr_n) = \\ &\quad = F(\mathbf{z})\, dz_1 \ldots dz_n = Z^{-1} \exp\left(-\frac{1}{2}(\mathbf{z}, A\mathbf{z}^*)\right) dz_1 \ldots dz_N, \end{aligned} \tag{7.4}$$

where $z_i = s_i + \imath r_i, dz_i = ds_i dr_i$, and the matrix A is hermitian. When $\mathcal{H}$ is a quadratic function of the vector u, equation (7.3) defines A in the complex case.

The prior measure (7.3) being non-Gaussian, we proceed as above and partition the Hamiltonian into a quadratic part plus a perturbation, $\mathcal{H} = \mathcal{H}^0 + \mathcal{H}^1$. To make the example amenable to analysis we keep only the leading term in this expansion; we already pointed out that the leading term contains a contribution of the nonlinear terms in the equation and that the partition takes into account higher order terms in the expansion. As explained above, there is here only one relevant partitioning parameter, m_0. Thus we approximate the probability density by $f(u) = Z_0^{-1} e^{-\mathcal{H}^0(u)}$, where

$$\mathcal{H}^0(u) = \frac{1}{2} \sum_{j=1}^{n} \left\{ \frac{|u_{j+1} - u_j|^2}{h} + h m_0^2 |u_i|^2 \right\}. \tag{7.5}$$

We work here with $n = 32$. In this finite dimensional case, we rederived the optimal value of m_0 as follows: We calculated the two-point correlation function $\langle u_i u_j^* \rangle^0$ for the measure (7.3) by a Monte-Carlo method, and then found the value of m_0 that best reproduces this correlation function by minimizing the mean-square difference between this correlation function and the correlation function produced by (7.5). This yielded $m_0 = 1.055$, confirming the value obtained above. The procedure used here minimized the difference between the full measure and its quadratic piece while the more general procedure of the previous sections minimizes only the first few terms in an expansion; it is comforting that the results agree. For readers of our earlier papers [4, 5, 6], we point out that the present procedure differs from the "Gaussianization" proposed there by using an analytical expression for the approximate measure, whereas in the previous publications we needed to store the full covariance matrix. Furthermore, the present construction produces a first term in a systematic expansion.

In Figure 7.1 we compare the exact two-point correlation function $\langle u_i u_j^* \rangle$ obtained by a Monte Carlo sampling (open circles) to the Gaussian approximation (solid line),

$$C_{ij} = \langle u_i u_j^* \rangle^0 = \frac{1}{\pi} \sum_{k=1}^{n} \frac{e^{ik(x_i - x_j)}}{\frac{4}{h^2} \sin^2 \frac{hk}{2} + m_0^2}, \tag{7.6}$$

derived from (7.5). The discrepancy is negligible compared with the statistical uncertainty in the sampling procedure.

The formulas for calculating conditional expectations (3.3)–(3.5) can be generalized to deal with complex functions. Assume we have N conditions of the form,

$$g_{\alpha i} u_i = V_\alpha,$$

where the $g_{\alpha i}$ are the complex entries of the $N \times n$ kernel matrix G and the entries of the vector V_α are the values of the N collective variables defined by G. The formulas for the conditional means and variances of the complex vector u_i generalize equations (3.3) and (3.5); the conditional average of u_i is

$$\langle u_i \rangle_V^0 = q_{i\alpha} V_\alpha;$$

the $q_{i\alpha}$ are the entries of the matrix $Q = (CG^\dagger)(GCG^\dagger)^{-1}$, where C, whose elements are defined in (7.6), is the matrix that approximates the inverse of the matrix A defined in

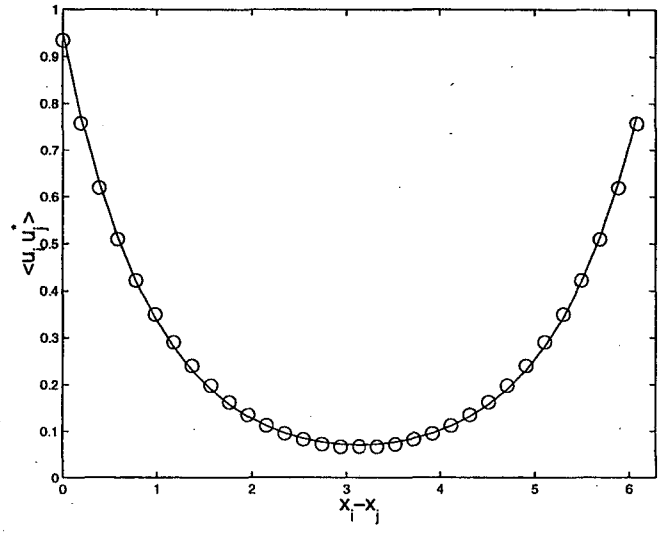


Figure 7.1: Comparison between the two-point correlation function $\langle u_i\, u_j^* \rangle$, as computed by a Monte Carlo sampling procedure, and its approximation (7.6) based on the Gaussian measure (7.5), with $m_0 = 1.05$. The number of points is $n = 32$.

(7.4). The dagger denotes an adjoint (hermitian transposed) matrix. The conditional covariance matrix has entries

$$\mathrm{Cov}\left\{u_i, u_j^*\right\}_V^0 = \left[C - (CG^\dagger)(GCG^\dagger)^{-1}(GC)\right]_{ij}.$$

We now make a special choice of kernels G: We pick them so that what is known at time $t = 0$ is a set of Fourier modes, fewer than are necessary to specify the solution completely. This makes the $g_{\alpha i}$ complex exponentials:

$$g_{\alpha i} = \frac{1}{n} \exp\left(-\imath K_\alpha x_i\right).$$

If the number of conditions N is even, the K_α take the values $-\frac{N}{2}+1, \ldots, \frac{N}{2}$. Note the following property of the resulting matrix G:

$$GG^\dagger = \frac{1}{n} I, \tag{7.7}$$

where I is the $N \times N$ identity. Spectral variables are particularly convenient here because, if u is expanded in Fourier series, and this series is substituted into the formula for the Hamiltonian $\mathcal{H}$, the result is a sum of squares of Fourier coefficients; the prior measure is "diagonal in Fourier space." This can also be deduced from the fact that the measure is Gaussian and translation-invariant in the variable x. In particular, after u is expanded in Fourier series, the matrices A, C are both diagonal. A short calculation yields:

$$Q = n\, G^\dagger,$$

from which follows

$$\langle u \rangle_V^0 = n\, G^\dagger V, \tag{7.8}$$

and

$$\mathrm{Cov}\left\{u_i, u_j^*\right\}_V^0 = \left[C - n\, G^\dagger G C\right]_{ij}. \tag{7.9}$$

Equation (7.8) is the interpolation formula for the first moment of u. Recall that $V = Gu$, hence $\langle u \rangle_V = n\, G^\dagger G u$. The operator $n\, G^\dagger G$ is a Galerkin projection operator which projects any vector onto the vector space spanned by the range of G. Thus, the mean of u produced by our regression formula equals the mean obtained from a simple Fourier series that uses the known coefficients V_α (we shall call this Fourier series the "Galerkin reconstruction"). This is as it should be: In a translation-invariant Gaussian measure the Fourier components are mutually independent; the knowledge of the first few does not condition the next ones, whose expected value is therefore zero. However, the measure does contain information about the higher moments of the higher Fourier coefficients, and this is important in a nonlinear problem.

The difference between the Galerkin reconstruction and the regression used in the optimal prediction of moments is demonstrated in Figure 7.2 . The three graphs depict the first three moments, calculated (i) by a Monte Carlo sampling of the measure (7.3) conditioned by $N = 4$ known collective variables (symbols), a Galerkin reconstruction (dashed lines), and our regression formulas (solid lines). We chose a system size of $n = 32$ with $N = 4$ resolved (i.e., known) Fourier modes. For the first moment, $\langle u_i \rangle_V$ (calculated by Monte-Carlo), the Galerkin reconstruction and the regression are close to each other. For the second and third moments the regression, which is the core of optimal prediction, is close to the truth, as revealed by the Monte-Carlo runs; the error is smaller than the statistical uncertainty in the sampling. The Galerkin reconstruction, on the other hand, deviates significantly from the truth. These graphs demonstrate the importance of the prior measure, which contains information about the mean squares of the unresolved Fourier components.

We next derive the optimal prediction scheme for our model problem with Fourier kernels. Given a set of Fourier modes, V_α, we replace the right-hand side of (7.1) by its conditional average, and multiply the result by the kernel matrix G . Thus,

$$\imath \frac{dV_\alpha}{dt} = g_{\alpha i} \left\langle -\frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} + \frac{1}{4}\left[3\left|u_i\right|^2 u_i + (u_i^*)^3\right] \right\rangle_V^0 .$$

Substituting the regression formulas (7.8) and (7.9) and using Wick's theorem and (7.7), we obtain the following set of N equations:

$$\imath \frac{dV_\alpha}{dt} = \frac{4}{h^2} \sin^2 \frac{K_\alpha h}{2} V_\alpha + \frac{1}{4} \sum_{\beta,\gamma,\epsilon} \left[3 V_\beta V_\gamma^* V_\epsilon\, \delta_{\alpha,\beta-\gamma+\epsilon} + V_\beta^* V_\gamma^* V_\epsilon^*\, \delta_{\alpha,-\beta-\gamma-\epsilon}\right] + \frac{6}{4} c\, V_\alpha, \tag{7.10}$$

where

$$c = \left[(I - n\, G^\dagger G) C\right]_{ii} \qquad \text{(no summation on } i\text{)}$$

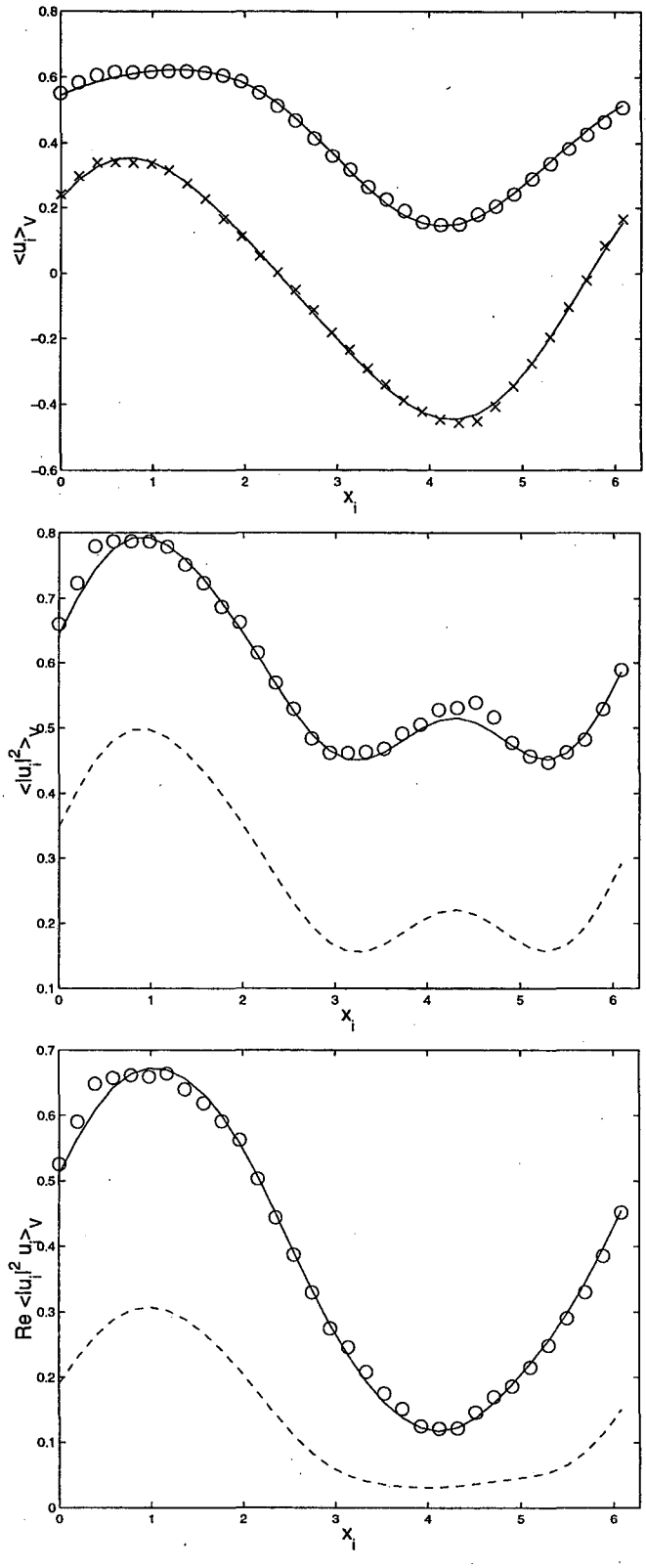


Figure 7.2: Conditional moments of u: comparison between true values (symbols), a Galerkin reconstruction (dashed lines), and regression (solid lines). (a) The conditional expectation $\langle u_i \rangle_V$; the circles represent the real part and the crosses represent the imaginary part. (b) The variance $\langle u_i^2 \rangle_V$. (c) The third moment Re $\langle |u_i|^2 u_i \rangle_V$. The number of points is $n = 32$, and the number of Fourier modes that are assumed known is $N = 4$.

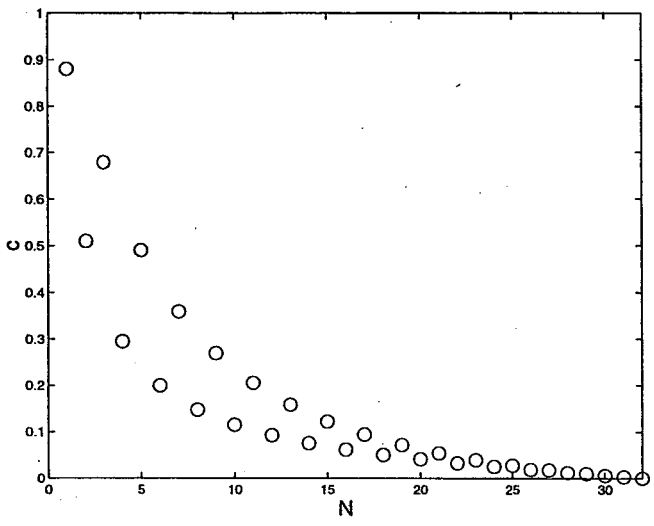


Figure 7.3: The parameter c as a function of the number N of resolved Fourier modes, for $n = 32$.

is a constant (the right-hand side is independent of i). Note that the last term comes from the evaluation of the nonlinear terms by Wick's theorem. The structure of equation (7.10) is enlightening: The first two terms on the right-hand side are precisely the Galerkin approximation for the evolution of a subset of Fourier modes; they constitute a pseudo-spectral approximation of the equations of motion. The third term, which is linear in V, represents information gleaned from the prior measure. The nice feature of this example is the sharp separation between the contribution from the resolved degrees of freedom and the contribution of the "subgrid" degrees of freedom, which happens to simply "renormalize" the linear part of the evolution operator.

One is of course interested in knowing how large is the extra term that makes up the entire difference between the optimal prediction scheme and a standard pseudo-spectral scheme; this difference is proportional to the coefficient c. In Figure 7.3 we plot the value of c as function of N for $n = 32$. As expected, c is larger when the number of resolved degrees of freedom is smaller , and vanishes when $N = n$, i.e., when the system is fully resolved. The oscillations in this graph result from the alternation between odd and even numbers of Fourier modes. Figure 7.3 demonstrates that optimal prediction is consistent: As the number of collective variables increases, its predictions converge to those of a resolved calculations (as indeed should be obvious from the derivation); when the number of collective variables is small, the corrections due to optimal prediction are substantial.

In Figure 7.4 we compare the time evolution of the first 4 Fourier coefficients predicted by our optimal prediction scheme (7.10) (solid lines), the time evolution of these modes predicted by a Galerkin scheme (dashed lines), and their exact mean evolution obtained by sampling 10^4 states from the conditional measure, evolving them in time, and averaging the first Fourier components (circles). We do this with $N = 4$ and $N = 8$; with $N = 8$ we display only 4 modes even though 8 are calculated, in the interest of clarity. To make the calculations with different values of N comparable, we pick the

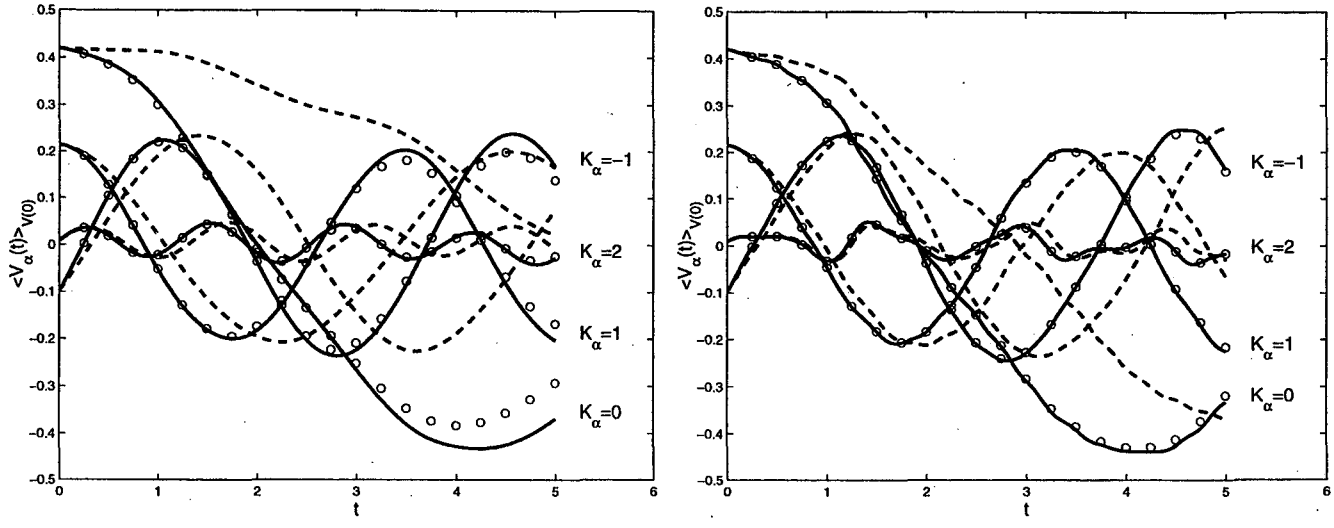


Figure 7.4: Time evolution of the real part of the 4 lowest Fourier modes for $n = 32$. The circles represent average values over an ensemble of 10^4 states; the dashed lines result from the Galerkin approximation; the solid lines result from our optimal prediction scheme (7.10). Results are presented for (a) $N = 4$ and (b) $N = 8$.

initial values of the collective variables in the way suggested by the analysis of Hald [12]: We sample an initial function u from the invariant measure, and then calculate values of the collective variables by performing the summations $g_{\alpha i}u_i$. The graph shows that in each case the simple Galerkin calculation deviates immediately from the true solution, while the optimal prediction remains accurate. The calculations show that optimal prediction improves the accuracy compared to a Galerkin calculation; the time during which the optimal prediction remains accurate increases with increasing N; we know from (7.10) that the cost of optimal prediction in this problem is small.

8 Conclusions

We have exhibited the value of the statistical information used in optimal prediction for the solution of an underresolved nonlinear problem, and we have shown that perturbation theory provides a ready-made machinery for applying the ideas of optimal prediction to problems where the invariant measure is non-Gaussian.

The first-order implementation of optimal prediction with a fixed, small, number of conditions breaks down after a finite time; the time for which it is valid increases as the temperature increases; the temperature determines the variance of the invariant measure and thus the uncertainty in the system. There are two ways to improve the prediction: Go to more sophisticated prediction theory, as outlined in section 2, or increase the number of collective variables. We have demonstrated the power of the second alternative; the first alternative will be explored in later publications.

Many aspects of the algorithms presented here require further work. The closure by means of a fixed number of affine conditions is only a first step; other closure schemes

will be investigated. In particular, optimal prediction fits within the framework of irreversible statistical mechanics, whose apparatus can be brought into use for finding closure schemes. More powerful versions of perturbation theory can be readily used. A careful perusal of our final example shows that there is a great advantage in using orthogonal kernels; in the interest of pedagogy, we have not used this possibility in the discussion of perturbation theory, and we have yet to explore orthogonal bases other than Fourier bases.

An inspection of the formulas derived by perturbation theory shows that though we assumed a knowledge of an invariant measure, all that is finally used is a set of moments; this is the opening for applying optimal prediction methods in problems where less than a full invariant measure is known.

We assume that all these issues will be handled as we progress to more complicated problems, in more dimensions, with dissipation (requiring a careful modeling of effective Hamiltonians), and with general boundary conditions. We shall explore problems with these additional features in future publications.

Acknowledgments

The authors would like to thank Prof. G.I. Barenblatt, Mr. T. Burin des Roziers, Dr. A. Gottlieb, Prof. O. Hald, Dr. A. Kast, Prof. D. Kessler, Dr. I. Kliakhandler, and Prof. P. Marcus for helpful discussions and comments. This work was supported in part by the Office of Science, Office of Advanced Scientific Computing 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 in part by the National Science Foundation under contract DMS94-14631.

References

[1] Bennett, A.F., *Inverse methods in physical oceanography*, Cambridge Monographs on Mechanics and Applied Mathematics, Cambridge, 1992

[2] Chorin, A.J., *Accurate evaluation of Wiener integrals*, Math. Comp. 27 (1973), 1-15.

[3] Chorin, A.J., *Vortex methods*, Les Houches Summer School of Theoretical Physics, 59, Elsevier (1995), 67-108.

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

[5] Chorin A.J., Kast A.P., Kupferman R., *Unresolved computation and optimal prediction*, Comm. Pure Appl. Math., in press, 1999.

[6] Chorin A.J., Kast A.P., Kupferman R., *On the prediction of large-scale dynamics using unresolved computations*, Contemp. Math., in press, 1999.

[7] Evans, D., Morriss, G., *Statistical Mechanics of Non-equilibrium Liquids*, Academic Press, NY, 1990.

[8] Fetter A.L., Walecka J.D., *Quantum theory of many particle systems*, McGraw-Hill, 1971.

[9] Freedman, F., *Brownian motion and diffusion*, Springer, NY, 1983

[10] Goldstein, H., *Classical mechanics*, Addison Wesley, Reading, Mass, 1980.

[11] Gottlieb, A.D., *First order optimal prediction*, in preparation.

[12] Hald O.H., *Optimal prediction and the Klein-Gordon equation*, Proc. Nat. Acad. Sci. USA 96 (1999), 4774-4779.

[13] Kleinert H., *Gauge fields in condensed matter*, Vol. I, World-Scientific, Singapore, 1989.

[14] Ma S.-K., *Modern theory of critical phenomena*, W.A. Benjamin, Massachusetts, 1976.

[15] MacKean, H., *Statistical mechanics of nonlinear wave equations IV: Cubic Schrödinger equations*, Comm. Math. Phys. 168, (1995), 479-491.

[16] Mori, H., *Transport, collective motion, and Brownian motion* Prog. Th. Phys., 33, (1965), 423-455.

[17] Morokoff, W., *Generating quasi-random paths for stochastic processes*, SIAM Rev., 40, (1998), 765-788.

[18] Mandl, F., *Introduction to quantum field theory*, Interscience, NY, 1959.

[19] Mock M., Lax P., *The computation of discontinuous solutions of linear hyperbolic equations*, Comm. Pure Appl. Math., 31, (1978), 423-430.

[20] Onsager, L., *Statistical hydrodynamics*, Nuovo Cimento, Suppl. to Vol. 6, (1949), 279-287.

[21] Papoulis, A., *Probability, random variables and stochastic processes*, McGraw-Hill, New York, 1991.

[22] Proudman, I., Reid, W., *On the decay of normally distributed and homogeneous turbulent velocity fields*, Phil. Trans. Roy. Soc. A247, (1954), 163-189.

[23] Ramond, P., *Field theory, a modern primer*, Benjamin, Reading, Mass, 1981

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

[25] Vaillant, O., PhD thesis, INRIA Sophia-Antipolis, 1999.

[26] Zwanzig, R., *Problems in nonlinear transport theory*, Systems Far From Equilibrium, L. Garrido (ed), Interscience, New York, 1961, 198-221.

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