Unresolved Computation and Optimal Predictions

ALEXANDRE J. CHORIN
ANTON P. KAST
AND
RAZ KUPFERMAN
Lawrence Berkeley National Laboratory

Abstract

We present methods for predicting the solution of time-dependent partial differential equations when that solution is so complex that it cannot be properly resolved numerically, but when prior statistical information can be found. The sparse numerical data are viewed as constraints on the solution, and the gist of our proposal is a set of methods for advancing the constraints in time so that regression methods can be used to reconstruct the mean future. For linear equations we offer general recipes for advancing the constraints; the methods are generalized to certain classes of nonlinear problems, and the conditions under which strongly nonlinear problems and partial statistical information can be handled are briefly discussed. Our methods are related to certain data acquisition schemes in oceanography and meteorology. © 1999 John Wiley & Sons, Inc.

1 Introduction

There are many problems in science whose solution is described in principle by a set of partial differential equations, but where the solutions of the equations are so complicated that in practice they cannot be found because they cannot be properly resolved; well-known examples include problems in statistical physics, such as many-body problems, as well as turbulence. There exists a large literature on methods for overcoming the lack of resolution by making additional, problem-specific assumptions.

The use of additional information cannot be avoided; our proposal in the present paper is to use prior statistical information and view the initial data that can be represented on the computer as constraints that limit the range of possible solutions. The resulting constrained statistics are then used to predict appropriate mean values and moments of future solutions. The availability of statistical information is not questioned here, and examples of what constitutes useful information are given. The greatest difficulty in the program just outlined is to find the future effect of initial constraints. For linear equations and in the presence of sufficient prior information, we solve this problem completely, at least in principle, and in certain other cases we solve it approximately. We also prepare the terrain for a discussion of nonlinear problems and of problems where the prior information is incomplete. Our approach provides predictions that, in a sense specified below, are the best information allowed by the lack of resolution.

Communications on Pure and Applied Mathematics, Vol. LII, 1231–1254 (1999)
 CCC 0010–3640/99/101231-24

We consider partial differential equations of the form

$$u_t = R(u, u_x, u_{xx}, \dots), \tag{1.1}$$

where the subscripts denote differentiation, x is a space variable, t is the time, and R is a (generally nonlinear) function of its arguments. We suppose that the solution of equation (1.1) is so complicated that it cannot be calculated accurately, for example, because it is chaotic, or nonsmooth, or because the number of spatial dimensions is too large. Assume for the time being that for every time t, we know a probability measure μ, the "prior measure," on the space of solutions of equation (1.1) viewed as functions of x. We suppose that we can only muster sufficient computational power to calculate N quantities, and that if these N quantities were Fourier coefficients of the solution or point values of the solution, the information that they contain would not be sufficient to calculate $R(u,\dots)$ and u_t accurately enough for a time integration starting from specific initial data. We assume that the N quantities we know initially are the values of N functionals of the solution, of the form

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

where the argument 0 in g refers to the initial time $t = 0$; we denote the value of the α^{th} functional by $\bar{u}_\alpha(0)$, so that

$$\bar{u}_\alpha(0) = (g_\alpha(0), u(0)). \tag{1.3}$$

If the "constraining kernels" g are local in space, one can think of the u_α as local averages of the solution; in the special case $g_\alpha = \delta(x - x_\alpha)$, $\bar{u}$ is a point value of the solution at $x = x_\alpha$. The set of functions u that satisfy the N initial constraints (1.3), the "initial constrained ensemble," is a subset of the set of functions that carry the measure μ at time $t = 0$ and inherits from μ a measure conditioned by the constraints; in other words, we pick initial data from the subset of functions on which the probability measure is defined that obey the initial constraints. If the problem is unresolved, this set of initial data contains functions that differ from each other significantly, and the prior measure allows us to assign probabilities to the occurrence of given functions in that initial ensemble. In physicists' notation, the constrained measure can be written as

$$\mu \prod_{\alpha=1}^{N} \delta[(g_\alpha(0), u) - \bar{u}_\alpha(0)] \tag{1.4}$$

where the normalization factor has been omitted.

If one uses each function in the initial constrained ensemble as an initial value for the partial differential equation (1.1), one obtains a set of solutions which, for each time t, inherits a measure from the initial conditions. Even when the prior measure μ does not change in time, the constrained measure does: For example, if one assumes that initially the functions assume certain values at N given points, there is no reason to believe that the solutions of the differential equation continue

to assume the very same values at these points later in time; indeed, if the prior measure μ is ergodic with respect to the equation, the constrained measure will converge in time to the prior measure.

To predict the future, we have to know how the constrained measure evolves in time. Our central hypothesis is that as the constrained measure evolves, it continues to be well approximated as the prior measure constrained by constraints of the form (1.3), but with kernels $g_\alpha(x,t)$ and constraint values $\bar{u}_\alpha(t)$ that may change in time. In physicists' notation, the time-dependent constrained measure can be written as

$$\mu \prod_{\alpha=1}^{N} \delta[(g_\alpha(t),u) - \bar{u}_\alpha(t)]\,. \tag{1.5}$$

If these kernels and values are known at a later time, one can use them to calculate averages with respect to the constrained measure and thus find the future averaged over the "invisible" degrees of freedom that cannot be represented in an unresolved calculation. In a sense explained below, our calculation minimizes the error on the average among all schemes that use the information available in the initial data. The central problem is to find the evolution of the constraining kernels $g_\alpha(x,t)$ and of the values $\bar{u}_\alpha(t)$ of the constraints.

We shall denote averages with respect to the prior measure μ by brackets $\langle\,\rangle$, and averages with respect to the constrained measure by subscripted brackets, $\langle\,\rangle_{\bar{u}}$.

We restrict ourselves in most of the present paper to problems where the prior measure is invariant in time. If one takes a subset of the functions that carry the measure initially (for example, those functions that at some fixed point x_0 take values between two given numbers a and b), then the differential equations may well take the corresponding solutions out of this set, but these solutions will be replaced by others so that the probability of finding a solution in the given subset remains fixed. Such measures are the analogues of equilibrium distributions in the kinetic theory of gases, where particles may collide and move but their distribution remains fixed.

Our approach shifts the burden of predicting the solution of unresolved problems from guessing relations between large-scale computable quantities and small scale, "invisible" quantities to determining appropriate prior distributions. In problems where invariant, i.e., equilibrium, measures serve as prior distributions, our approach is a numerical version of near-equilibrium statistical mechanics. Our use of prior and constrained "posterior" statistics is reminiscent of Bayesian methods, augmented here by the key device of time-dependent constraints determined by differential equations. Our construction is related to certain data acquisition schemes (see [5, 16, 19]) where one tries to make predictions on the basis of uncertain equations, partial and uncertain data, and prior information about the statistics of the errors. A simplified version of our methods has been briefly explained in [9]. 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., [14, 20]). Time dependence has similarly been approached through the estimation of time derivatives [1, 12]. Without a prior measure and a prescription for its constrained time evolution, however, the usefulness of such methods is necessarily limited.

The paper is organized as follows: In the next section we present for later use some lemmas on Gaussian distributions and related constrained expectations. In the third section we present a general discussion of the problem of advancing constraints; in particular, we show that in the linear case the assumption according to which the constrained measure remains the prior measure constrained by a fixed number of affine constraints is true, and that the exact evolution of the kernels is dual to the evolution of the solutions of the partial differential equation. We also discuss approximate evolutions of the constraints that may be easier to use in practice, and show that the constructions can be extended to certain nonlinear problems. In the three sections that follow, we apply our techniques to a linear problem, a nonlinear problem, and a problem with random forcing. Finally, we provide a short discussion of what one should do in problems that are strongly nonlinear and where the prior information is incomplete.

2 Gaussian Distributions and Constrained Expectations under Affine Constraints

We begin by presenting three elementary lemmas that relate Gaussian prior measures to the corresponding constrained measures. In all the examples we consider in the present paper, the measures either are Gaussian or can be viewed as perturbations of Gaussian measures. The general case of an arbitrary prior measure requires more sophisticated tools and will be briefly discussed in the final section. The Gaussian and near-Gaussian cases are sufficient to explain the ideas and are often encountered in practice.

Consider first a finite collection of random variables $u_1, \dots, u_n$. This collection has a Gaussian distribution when the values of the random variables have a probability distribution function of the form

$$\begin{aligned} &P(s_1 < u_1 \le s_1 + ds_1, \dots, s_n < u_n \le s_n + ds_n) \\ &\quad = f(\mathbf{s})\, ds_1 \cdots ds_N = Z^{-1} \exp\left(-\tfrac{1}{2} s_i a_{ij} s_j + b_i s_i\right) ds_1 \cdots ds_N , \end{aligned} \tag{2.1}$$

where a_{ij} are the entries of a symmetric positive definite matrix A, $i, j = 1, \dots, n$, the b_i are the entries of a vector $\mathbf{b}$, Z is a normalizing constant, and repeated indices imply summation. The means of the u_i are

$$\langle u_i \rangle = a^{-1}_{ij} b_j , \tag{2.2}$$

where the a^{-1}_{ij} are the entries of the matrix A^{-1}, and the covariances of the u_i are

$$\mathrm{Cov}[u_i, u_j] \equiv \langle u_i u_j \rangle - \langle u_i \rangle \langle u_j \rangle = a^{-1}_{ij} . \tag{2.3}$$

The expectations of higher moments are determined by Wick's theorem [13]:

$$\langle (u_{i_1} - \langle u_{i_1}\rangle)\dots(u_{i_l} - \langle u_{i_l}\rangle)\rangle = \begin{cases} 0, & l \text{ odd}, \\ \sum_{\substack{\text{all possible pairings} \\ \text{of}\{i_1\dots i_l\}}} \mathrm{Cov}[u_{i_{p_1}}, u_{i_{p_2}}]\dots\mathrm{Cov}[u_{i_{p_{l-1}}}, u_{i_{p_l}}], & l \text{ even}. \end{cases} \tag{2.4}$$

Suppose we draw $\mathbf{u} = u_1, \dots, u_n$ from a Gaussian distribution but accept only the samples that satisfy $N < n$ affine constraints,

$$\bar{u}_\alpha = g_{\alpha i} u_i\,, \quad \alpha = 1, \dots, N, \tag{2.5}$$

where the N numbers $\bar{u}_\alpha$ are given, the $g_{\alpha i}$ are the components of an $N \times n$ matrix G, of which the N column vectors indexed by α are linearly independent. Symbolically, we write that $\mathbf{u}$ is distributed according to the constrained probability density

$$f_{\bar{u}}(\mathbf{u}) = \frac{1}{Z'} f(\mathbf{u}) \prod_{\alpha=1}^{N} \delta[g_{\alpha i} u_i - \bar{u}_\alpha]\,, \tag{2.6}$$

where Z' is a normalization constant.

The mean of $\mathbf{u}$ with respect to the constrained probability density (2.6) is given explicitly in terms of A, $\mathbf{b}$, and the constraint data G and $\bar{u}$ by

$$\langle u_i\rangle_{\bar{u}} = a^{-1}_{ij} b_j + (a^{-1}_{ij} g_{\alpha j}) m^{-1}_{\alpha\beta} \left[\bar{u}_\beta - g_{\beta k} a^{-1}_{kl} b_l\right] \tag{2.7}$$

where, as before, the $a^{-1}_{\alpha\beta}$ are the entries of A^{-1}, and similarly for $m^{-1}_{\alpha\beta}$ and M^{-1}, where the matrix M has as entries the unconstrained covariances of the constrained quantities,

$$m_{\alpha\beta} = \left\langle (g_{\alpha i} u_i)\left(g_{\beta j} u_j\right)\right\rangle - \langle g_{\alpha i} u_i\rangle \left\langle g_{\beta i} u_i\right\rangle = g_{\alpha i} a^{-1}_{ij} g_{\beta j}\,. \tag{2.8}$$

The covariances of the $\mathbf{u}$ with respect to the constrained probability density (2.6) are

$$\mathrm{Cov}_{\bar{u}}[u_i, u_j] = a^{-1}_{ij} - \left(g_{\alpha k} a^{-1}_{ki}\right) m^{-1}_{\alpha\beta} \left(a^{-1}_{jl} g_{\beta l}\right). \tag{2.9}$$

Wick's theorem applies to these constrained expectations exactly as it applied to the unconstrained expectations:

$$\left\langle (u_{i_1} - \langle u_{i_1}\rangle_{\bar{u}})\dots(u_{i_l} - \langle u_{i_l}\rangle_{\bar{u}})\right\rangle_{\bar{u}} = \begin{cases} 0, & l \text{ odd}, \\ \sum_{\substack{\text{all possible pairings} \\ \text{of}\{i_1\dots i_l\}}} \mathrm{Cov}_{\bar{u}}[u_{i_{p_1}}, u_{i_{p_2}}]\dots\mathrm{Cov}_{\bar{u}}[u_{i_{p_{l-1}}} u_{i_{p_l}}], & l \text{ even}, \end{cases} \tag{2.10}$$

Equations (2.7) and (2.9) can be derived from standard linear regression theory (see, e.g., [17]). The derivation of the constrained version of Wick's theorem (2.10) proceeds by noting that a delta function can be represented as a limit of a Gaussian

variable with a small variance $\Delta^2/2$; therefore, the constrained probability density (2.6) can be written as

$$f_{\bar{u}}(\mathbf{u}) = \lim_{\Delta\to 0} \frac{1}{Z'} \exp\left(-\frac{1}{2}u_i a_{ij} u_j + b_i u_i\right) \times \prod_{\alpha=1}^{N} \frac{1}{\sqrt{\pi}\Delta} \exp\left[-\frac{1}{\Delta^2}\left(g_{\alpha i}u_i - \bar{u}_\alpha\right)^2\right]. \tag{2.11}$$

For any finite value of Δ, the constrained probability density is proportional to an exponential of a quadratic form of the vector $\mathbf{u}$; i.e., it is Gaussian, and therefore Wick's theorem applies. An appropriate limit can then be taken.

Equation (2.7) shows that constraints alter means in a way that is linear in the $\bar{u}_\alpha$ and independent of multiplicative factors in the covariance. Equation (2.9) shows that constrained covariances are determined by the matrix G alone, without reference to the $\bar{u}_\alpha$. Equation (2.10) shows that the constrained Gaussian distribution (2.6), while not satisfying the requirement that the covariance matrix be nonsingular, 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 and y in an appropriate range rather than only for integer values of i and j. The matrix A^{-1} becomes the integral operator whose kernel is the function a^{-1}. The kernel $a = a(x,y)$ of the operator A inverse to A^{-1} is defined by

$$\int dy\, a^{-1}(x,y)a(y,z) = \delta(x,z)\,. \tag{2.12}$$

The vectors with entries $g_{\alpha i}$ become functions $g_\alpha(x)$, and the constraint equations (10) become equations (2) of the introduction. The regression formula (2.7) becomes

$$\langle u(x)\rangle_{\bar{u}} = \langle u(x)\rangle + c_\beta(x)\left[\bar{u}_\beta - \left\langle \int g_\beta(y)u(y)\,dy \right\rangle\right] \tag{2.13}$$

where

$$c_\beta(x) = \left\{\int a^{-1}(x,y)g_\alpha(y)\,dy\right\} m^{-1}_{\alpha\beta}\,, \tag{2.14}$$

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

$$m_{\alpha\beta} = \iint g_\alpha(x)a^{-1}(x,y)g_\beta(y)dx\,dy\,. \tag{2.15}$$

The formula for the constrained covariance can be obtained from (2.9) 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 survives with the appropriate changes in notation. Note that the Greek indices that refer to the N initial data survive as integers.

3 Updating Constraints

We now turn to the central problem of updating the constraints (1.3) in time. Consider the special case of equation (1.1):

$$u_t = Lu\,, \tag{3.1}$$

where L is a linear operator. We are looking for the average (and possibly higher moments) of the solution at time t, knowing that at $t = 0$ the N equations (1.3) were satisfied, with given constraining kernels $g_\alpha = g_\alpha(0) = g_\alpha(x,0)$ and given values $\bar{u}_\alpha(0)$ of the constraints. The problem at hand is to find kernels $g_\alpha(t)$ and constraint values $\bar{u}_\alpha(t)$ that will encode the same information as the initial constraints.

A first construction proceeds as follows: Differentiate the α^{th} constraint with respect to t:

$$\frac{d}{dt}\bar{u}_\alpha(t) = ((g_\alpha)_t, u) + ((g_\alpha), u_t) = ((g_\alpha)_t + L^\dagger g_\alpha, u) \tag{3.2}$$

where $L^\dagger$ is the operator adjoint to L. If we choose $g_\alpha(t)$ so that it satisfies the equation

$$(g_\alpha)_t + L^\dagger g_\alpha = 0\,, \tag{3.3}$$

and if we set, in addition, $\bar{u}(t) = \bar{u}(0)$, the resulting constraint

$$(g_\alpha(t), u(t)) = \bar{u}_\alpha(0) \tag{3.4}$$

will be satisfied at time t by all solutions that satisfied the initial constraint. An analogous equation has been proposed for data acquisition [5]. The constrained probability distribution defined by these constraints is the probability distribution that develops from the initial constrained distribution; indeed, in physicists' notation, we write the constrained measure in the form

$$\mu_{\bar{u}}(t) = f_{\bar{u}}[du] = f(u)[du] \prod_{\alpha=1}^{N} \delta[(g_\alpha(t), u) - \bar{u}_\alpha(t)]\,, \tag{3.5}$$

where $[du]$ is a formal product of differentials, $f(u)$ is the prior probability density, and $f_{\bar{u}}(u)$ is the constrained probability density. The evolution of any probability density is given by the Liouville equation [18],

$$\frac{\partial f}{\partial t} + \int \frac{\delta f}{\delta u} Lu\, dx = 0\,, \tag{3.6}$$

where $\delta/\delta u$ denotes a functional derivative. One can readily verify that if $f(u)$, the prior density, is time-invariant ($\partial f/\partial t = 0$), then the constrained probability density $f_{\bar{u}}(u)$ satisfies (3.6) when the kernels obey (3.3). The constraints that satisfy (3.3) are therefore equivalent to the initial constraints and encode the same information. We call an evolution described by $-L^\dagger$ *dual* to the evolution described by L. Formula (3.3) solves the problem of determining the kernels g_α and determines

completely the evolution of all the moments of u. We have also proved in the linear case the validity of our assumption according to which the constrained measure remains for all time the prior measure subjected to N affine constraints.

It is natural to look for ways of updating constraints that are easier than solving equation (3.3), as well as easier to generalize to nonlinear problems. One starting point is the equations (2.7) and (2.9), which express the constrained means and variances in terms of the constraints. (Our discussion now will be restricted to the Gaussian case; for the more general situation, see the concluding section below.) By means of Wick's theorem, one can also express the higher-order moments at all points x in terms of the constraining kernels and the values of the constraints. The knowledge of the means, covariances, etc., at all x makes it possible to evaluate quantities such as $\langle u_t \rangle = \langle u \rangle_t$, $\langle (u^2)_t \rangle = \langle u^2 \rangle_t$, and the derivatives of various covariances and of higher moments, at the N points of the computational grid, and therefore calculate all these moments a short time Δt later. According to our main assumption, these later moments can be obtained by constraining the prior measure by a new set of N constraints; if we allow the functions g_α to depend on $\ell - 1$ parameters, and if one calculates ℓ moments per grid point, equations (2.7) and (2.9) and their higher-order analogues constitute ℓn nonlinear equations for the unknown parameters at time Δt and for the constraint values $\bar{u}(\Delta t)$.

Given the new constraining kernels and the new constraint values, the calculation can be repeated, and if it is repeated often and one makes $\Delta \to 0$, one obtains a set of of ordinary differential equations for the parameters and the constraints. The simplest approximation in this sequence of approximations is one in which only the mean is used to update the constraints; one can then choose as parameter the value $\bar{u}_\alpha$ of the constraint and keep the g_α constant in time, for example, equal to the bell-shaped function $g_\alpha = \exp[-(x - x_\alpha)^2/2\sigma]/\sqrt{2\pi\sigma}$, with width σ, centered at $x = x_\alpha$. Multiplication of equation (2.7) by $g_{\alpha i}$ and summation over i yields

$$\bar{u}_\alpha(t) = g_{\alpha i} \langle u_i \rangle_{\bar{u}}; \tag{3.7}$$

differentiation with respect to t gives

$$\frac{d\bar{u}_\alpha}{dt} = g_{\alpha i} \left\langle \frac{du_i}{dt} \right\rangle_{\bar{u}} \tag{3.8}$$

or, in the continuum version,

$$\frac{d\bar{u}_\alpha}{dt} = \left\langle \left(g_\alpha, \frac{du}{dt} \right) \right\rangle_{\bar{u}} = \langle (g_\alpha, Lu) \rangle_{\bar{u}} . \tag{3.9}$$

To understand what is being assumed here, consider the derivative of the constraint equation (3.2) when the kernel g_α is time invariant:

$$\frac{d}{dt} \bar{u}_\alpha(t) = (g_\alpha, u_t) = (g_\alpha, Lu) . \tag{3.10}$$

If the constraint $(g_\alpha, u(t)) = \bar{u}_\alpha(t)$ were indeed the evolute of the initial constraint $(g_\alpha, u(0)) = \bar{u}_\alpha(0)$, then the right-hand side of equation (3.10) would be a (nonrandom) function of t. If the constraint $(g_\alpha, u(t)) = \bar{u}_\alpha(t)$ is not the evolute of the initial constraint, then the right-hand side of equation (3.10) is a random variable; the approximation (3.8) replaces this random variable by its mean. If the standard deviation of (g_α, Lu) in the constrained ensemble were zero, equation (3.8) would be exact and not an approximation; if the constrained standard deviation of (g_α, Lu) is small, equation (3.8) is a good approximation. To make the constrained standard deviation of (g_α, Lu) small, g_α should have a support of nontrivial width so that (g_α, Lu) is a (locally) averaged quantity; the spatial derivatives of g_α should also be smooth, so that the derivatives in L are averaged. The approximate procedure is useful if (g_α, Lu) has a small constrained standard deviation with kernels g_α whose width (for example, σ in the Gaussian case) is comparable to the distance between points. Note the analogy with the conditions under which vortex methods converge [8]; they, too, rely on fixed kernels to approximate a varying field.

Given the kernels g_α, the information we have at any one time consists of the values of the N constraints. Suppose we use this information in some arbitrary way to construct an approximation u_t^* to $u_t = Lu$; u_t^* is a function of the $\bar{u}_\alpha, \alpha = 1, \ldots, N$, say $u_t^* = D(\bar{u}_1, \ldots, \bar{u}_N)$. Given a specific member $u = u(x)$ of the constrained ensemble, the error in the approximation is

$$E[u(\cdot)] = Lu - D(\bar{u}_1, \ldots, \bar{u}_N). \tag{3.11}$$

Linear regression, used to obtain the means $\langle u(x) \rangle_{\bar{u}}$ from the knowledge of the prior measure and the constraints, guarantees that the value of the time derivative we calculated in the preceding paragraph minimizes the mean square error

$$\left\langle [Lu - D(\bar{u}_1, \ldots)]^2 \right\rangle \tag{3.12}$$

over all approximations $D(\ldots)$. In this sense, our approximation of the time derivative is optimal.

The exact procedure for evolving g_α (equation (3.3)) requires the solution of the partial differential equation for the functions g_α; when the equation has constant coefficients, one may use as the g_α the spatial translates of a single function g and reduce the amount of work. Furthermore, in the exact construction, once the constraining kernels have been calculated, the moments of the solution can be found from the formulas in the previous section, and the equation $u_t = Lu$ is not used any further. In the second, approximate construction where the g_α are fixed and the $\bar{u}_\alpha$ vary, the latter have to be determined from the evolution of the mean solution in time so that the right-hand side of (3.8) can be evaluated, requiring the solution of N ordinary differential equations (see the examples in the following sections). These two procedures are merely the extreme members of a continuum of possibilities: One can solve equation (3.3) approximately, for example, by expanding g_α in a series and using as an approximate $g_\alpha(t)$ only the first few terms in the series that results from the application of (3.3): The approximate $g_\alpha(t)$ can then be inserted

into equation (3.2) to obtain an equation for $\bar{u}_\alpha(t)$ whose solution varies slowly and where the constrained variance of the right-hand side may be smaller than in the case of an invariant g_α.

Note that the approximation (3.8) is applicable as it stands to nonlinear equations; for nonlinear equations, one also has a whole range of intermediate possibilities. For example, one may write the nonlinear operator R of equation (1.1) as $R = L + Q$, where L is linear; then one finds

$$\frac{d\bar{u}_\alpha}{dt} = ((g_\alpha)_t, u) + (g_\alpha, Lu + Q(u)) = ((g_\alpha)_t + L^\dagger g_\alpha, u) + (g, Q(u)). \tag{3.13}$$

If one uses equation (3.3) to determine the $g_\alpha(t)$, one finds

$$\frac{d\bar{u}_\alpha}{dt} = \langle (g_\alpha, Q(u)) \rangle_{\bar{u}}. \tag{3.14}$$

The art is to find partitions $R = L + Q$ that minimize the constrained variance of (g_α, Q). Such partitions are reminiscent of the interaction picture in quantum mechanics.

4 Example 1: A Linear Schrödinger Equation

Consider the Schrödinger equation on the unit circle with a constant potential,

$$iu_t = -u_{xx} + m^2 u, \tag{4.1}$$

where u is a complex-valued function on the circle and m^2 is a constant. This equation is the Hamilton equation of motion for the Hamiltonian

$$H[u] = \int_0^{2\pi} \left[|u_x|^2 + m^2 |u|^2 \right] dx \tag{4.2}$$

(see [11]). Equation (4.1) preserves the density of the canonical ensemble,

$$f_0[u] = e^{-H[u]}, \tag{4.3}$$

where the temperature has been chosen equal to 1.

The measure defined by equation (4.3) is absolutely continuous with respect to a Wiener measure [15], and its samples are, with probability 1, almost nowhere differentiable. The corresponding solutions of the equations of motion are weak and difficult to approximate numerically.

By symmetry we see that the unconstrained means $\langle u(x) \rangle$ and $\langle u^*(x) \rangle$ are zero. To extract the covariance function A^{-1} from the probability density (4.3), we write $H[u]$ as a double integral,

$$H[u] = \int_0^{2\pi} \int_0^{2\pi} \left[u_x^*(x) \delta(x-y) u_x(y) + m^2 u^*(x) \delta(x-y) u(y) \right] dx\, dy, \tag{4.4}$$

where u^* is the complex conjugate of u. Integration by parts shows that the entries of the functional matrix A are

$$a(x,y) = -\delta''(x-y) + m^2 \delta(x-y), \tag{4.5}$$

where $\delta(x)$ is a δ-function and $\delta''(x)$ is the second derivative of a δ-function. The inverse of this function, $a^{-1}(x,y)$, is the function $K(x,y)$ that satisfies

$$\int_0^{2\pi} a(x,y)K(y,z)\,dy = \delta(x,z)\,. \tag{4.6}$$

Using the explicit form (4.5) of A, we see that $K(x,y)$ is translation invariant, i.e., a function only of $(x-y)$, and that it satisfies

$$-\frac{d^2K}{dx^2}(x-y) + m^2K(x-y) = \delta(x-y)\,. \tag{4.7}$$

This equation can easily be solved by Fourier series. The result is

$$K(x,y) = a^{-1}(x,y) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \frac{e^{ik(x-y)}}{k^2+m^2}\,. \tag{4.8}$$

We suppose that the initial data for equation (4.1) are drawn from the distribution (4.3) and that N initial observations have been made at the N grid points we can afford:

$$\bar{u_\alpha}(0) = (g_\alpha, u(0)) = \int_0^{2\pi} u(x,0)g_\alpha(x)dx\,, \tag{4.9}$$

where $\alpha = 1,\ldots,N$, so that the $x_\alpha = 2\pi\alpha/N$ form a regular mesh on the circle. We consider first an approximation where the functions g_α are fixed in time, and we furthermore take them to be translates of each other: $g_\alpha(x) = g(x-x_\alpha)$. We pick

$$g(x) = \frac{1}{\sqrt{\pi}\sigma}\sum_{\tau=-\infty}^{\infty} \exp\left[-\frac{(x-2\pi\tau)^2}{\sigma^2}\right], \tag{4.10}$$

i.e., a normalized Gaussian function whose width is σ, with suitable images to enforce periodicity; we will pick σ to be equal to the mesh spacing $2\pi/N$. The Fourier representation of $g(x)$ is

$$g(x) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ikx}e^{-\frac{1}{4}k^2\sigma^2}\,. \tag{4.11}$$

Given the constraints (4.9), we can find the mean initial condition everywhere by applying equation (2.7), which becomes

$$\langle u(x)\rangle_{\bar{u}} = c_\beta(x)\bar{u}_\beta \tag{4.12}$$

where

$$\begin{aligned} c_\beta(x) &= \left\{\int_0^{2\pi} a^{-1}(x,y)g(y-x_\alpha)\,dy\right\} m^{-1}_{\alpha\beta} \\ &= \left\{\frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ik(x-x_\alpha)}\frac{e^{-\frac{1}{4}k^2\sigma^2}}{k^2+m^2}\right\} m^{-1}_{\alpha\beta}\,, \end{aligned} \tag{4.13}$$

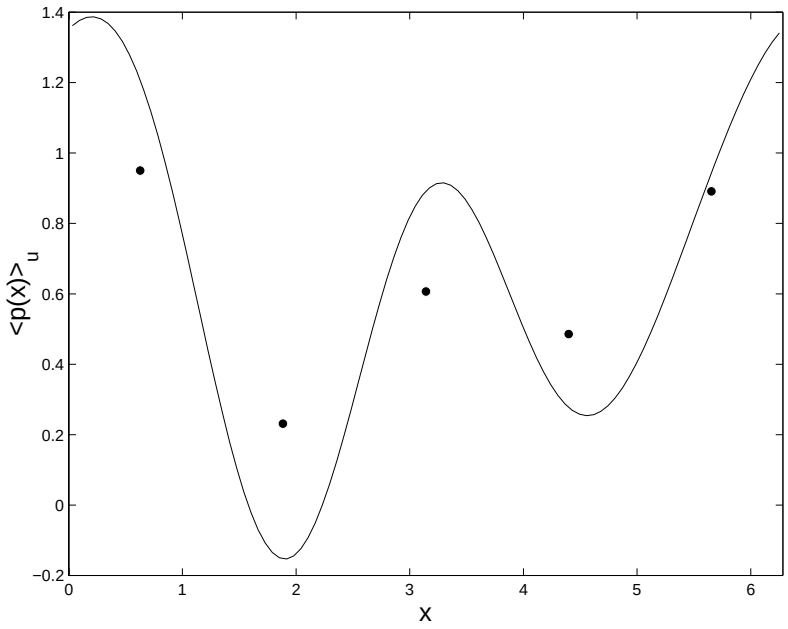


FIGURE 4.1. Example of a regression curve for the linear Schrödinger equation. Five constraint values (•) were chosen, representing local averages of $p(x) = \mathrm{Im}\, u(x)$ on a uniformly spaced grid. The constraining kernels are translates of each other and have a Gaussian profile of width $2\pi/5$ and are centered at the grid points. The solid line represents the constrained mean function $\langle p(x)\rangle_{\bar{u}}$ obtained by taking the imaginary part of equation (4.12).

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

$$
\begin{aligned}
m_{\alpha\beta} &= \int_0^{2\pi}\int_0^{2\pi} g(x-x_\alpha)a^{-1}(x,y)g(y-x_\beta)\,dx\,dy \\
&= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ik(x_\alpha - x_\beta)}\frac{e^{-\frac{1}{2}k^2\sigma^2}}{k^2+m^2}\,.
\end{aligned}
\tag{4.14}
$$

These Fourier series are easily evaluated numerically. An example of a regression for randomly chosen values of $\bar{u}_\alpha$ is shown in Figure 4.1.

We apply the approximate procedure in which the constraining kernels $g_\alpha(x)$ are fixed and only the values of the constraints $\bar{u}_\alpha$ change. The mean derivative of u is, from equation (4.1),

$$\begin{aligned}\langle u_t(x)\rangle_{\bar{u}} &= \langle iu_{xx}(x) - im^2u(x)\rangle_{\bar{u}} \\ &= i\frac{d^2}{dx^2}\langle u(x)\rangle_{\bar{u}} - im^2\langle u(x)\rangle_{\bar{u}} \\ &= \left\{i\frac{d^2c_\gamma}{dx^2}(x) - im^2c_\gamma(x)\right\}\bar{u}_\gamma.\end{aligned} \tag{4.15}$$

Substituting the Fourier representation of $c_\beta(x)$ from equation (4.13), we find

$$\langle u_t(x)\rangle_{\bar{u}} = -ig_\beta(x)m^{-1}_{\beta\gamma}\bar{u}_\gamma. \tag{4.16}$$

Multiplication of equation (4.16) by g_α and an integration yields, as in equation (3.8), the approximation

$$\frac{d\bar{u}_\alpha}{dt} = -\int_0^{2\pi} g_\alpha(y)\langle u_t(y)\rangle_{\bar{u}}\,dy = -i\left\{\int_0^{2\pi} g_\alpha(y)g_\beta(y)\,dy\right\}m^{-1}_{\beta\gamma}\bar{u}_\gamma, \tag{4.17}$$

where the integral in braces is another periodic Gaussian,

$$\int_0^{2\pi} g_\alpha(y)g_\beta(y)dy = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ik(x_\alpha - x_\beta)}e^{-\frac{1}{2}k^2\sigma^2}. \tag{4.18}$$

The important point about formula (4.17) is that it represents a closed system of ordinary differential equations for the $\bar{u}_\alpha$'s. The matrix elements in braces need be computed only once to define the scheme and are easily evaluated for all α and β.

We next calculate the exact mean future of the unresolved Schrödinger equation so that it can be compared with the result of the scheme we just presented. This could be done by using regression to find the mean solution at the initial time and then using the Green function to propagate this mean solution to a future time. Instead, we shall use the exact formalism (equation (3.3)) to find the time-dependent constraining kernels. A comment on these two ways of getting exact mean solutions in linear problems can be found below.

Starting from the initial kernels (4.11), the later kernels obey the dual equation, which is a time-reversed Schrödinger equation. Its solution is

$$g_\alpha(x,t) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ik(x-x_\alpha)}e^{i(k^2+m^2)t}e^{-\frac{1}{4}k^2\sigma^2}. \tag{4.19}$$

The mean solution at time t is constructed by regression from the probability density (4.3) and the constraining kernels (4.19). Repeating the calculation that led to (4.12), we find

$$\langle u(x,t)\rangle_{\bar{u}(t)} = c_\beta(x,t)\bar{u}_\beta \tag{4.20}$$

where

$$c_\beta(x,t) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ik(x-x_\alpha)}\frac{e^{-\frac{1}{4}k^2\sigma^2}}{k^2+m^2}e^{i(k^2+m^2)t}m^{-1}_{\alpha\beta}(t), \tag{4.21}$$

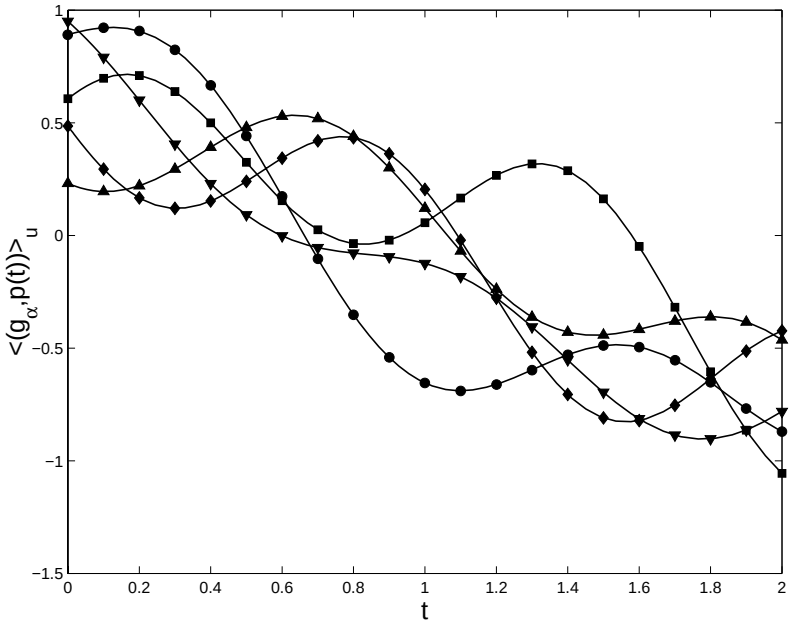


FIGURE 4.2. Evolution in time of the mean variables $\langle (g_\alpha, p(t)) \rangle_{\bar{u}}$, where $p(x,t) = \mathrm{Im}\, u(x,t)$, and g_α is a time-independent Gaussian constraint kernel as in the text. The symbols ($\alpha = 1\blacktriangledown$, $\alpha = 2\blacktriangle$, $\alpha = 3\blacksquare$, $\alpha = 4\blacklozenge$, $\alpha = 5\bullet$) are the exact values of these quantities as determined by equation (4.20). The solid lines are the approximations obtained by integrating the system of ordinary differential equations (4.17).

$$m_{\alpha\beta}(t) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{ik(x_\alpha - x_\beta)} \frac{e^{-\frac{1}{2}k^2\sigma^2}}{k^2+m^2} e^{2i(k^2+m^2)t}\,, \tag{4.22}$$

and as usual, the $m^{-1}_{\alpha\beta}$ are the entries of the matrix inverse to the matrix whose entries are $m_{\alpha\beta}$.

We can use this exact formula for mean futures (4.20) to test the accuracy of the approximation (4.17). Figure 4.2 compares the approximate prediction with the exact mean solution for initial values $\bar{u}(0)$ of the constraints picked at random. The approximation is not distinguishable from the exact solution on the scale of the plot for the duration of the calculation. Note that we have efficiently calculated the average behavior of an ensemble of solutions, the individual members of which are very difficult to evaluate.

As a final comment, we wish to point out that our techniques are useful even in linear problems where other ways of using regression to predict the mean future are available. First, we have seen and will see again below that our techniques are starting points for approximation. In addition, our methods apply for linear equations with nonconstant coefficients. Alternate techniques for advancing the mean

solution exactly in linear problems require the knowledge of a Green function, which is a function of $2d$ arguments, where d is the number of space dimensions. Our methods, even when an exact solution is wanted, require a knowledge of N functions of d arguments—a much smaller amount of information to store.

5 Example 2: A Nonlinear Hamiltonian System

We now consider a nonlinear generalization of the method demonstrated in the preceding section. We want to exhibit the power of our method by comparing the solutions that it yields with exact solutions; in the nonlinear case, exact solutions of problems with random data are hard to find, and we resort to a stratagem. Even though our method applies to the full nonlinear partial differential equation, we study a finite-dimensional system of n ordinary differential equations that is formally an approximation of a nonlinear Schrödinger equation:

$$i\frac{du_j}{dt} = -\frac{u_{j+1} - 2u_j + u_{j-1}}{\Delta^2} + \mathrm{Re}(u_j)^3 + i\,\mathrm{Im}(u_j)^3, \quad j = 1, \ldots, n, \tag{5.1}$$

where the n dependent variables u_j are complex-valued and $\Delta = \frac{1}{n}$. This system is nonintegrable for $n > 1$; it is the Hamilton equation of motion for the Hamiltonian

$$H(q_1, \ldots, q_n, p_1, \ldots, p_n) = \frac{1}{2}\sum_{j=1}^{n}\left[\left(\frac{q_{j+1} - q_j}{\Delta}\right)^2 + \left(\frac{p_{j+1} - p_j}{\Delta}\right)^2 + \frac{1}{2}\left(q_j^4 + p_j^4\right)\right], \tag{5.2}$$

where q_j and p_j are canonical variables related to u by $u_j = q_j + ip_j$. The differential equations preserve the canonical density,

$$F(q, p) = e^{-H(q,p)}. \tag{5.3}$$

The approximation is only formal because we shall be considering nonsmooth data which give rise to weak solutions that cannot be readily found by difference methods.

The probability density (5.3) is not Gaussian, raising a technical difficulty in computing constrained expectation values. We adopt an approximate procedure where the density (5.3) is approximated by a Gaussian density that yields the same means and covariances. The means are zero by symmetry,

$$\langle q_j \rangle = 0 = \langle p_j \rangle \tag{5.4}$$

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

$$\langle q_j p_k \rangle = 0, \tag{5.5}$$

since the density factors into a product of a density for the q's and a density for the p's. Thus $\langle q_j q_k \rangle = \langle p_j p_k \rangle$ are the only nontrivial covariances. Finally, since

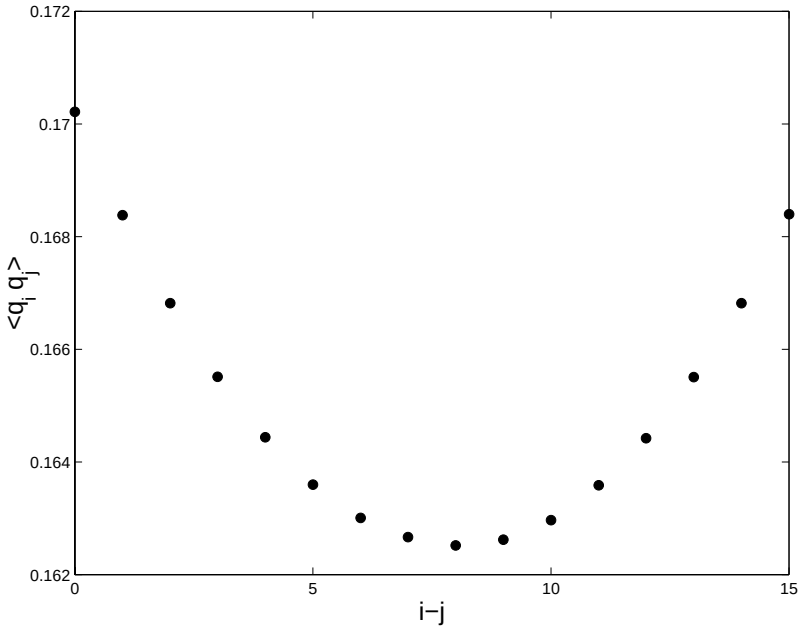


FIGURE 5.1. The covariance $\langle q_i q_j \rangle = \langle p_i p_j \rangle$ as a function of the separation $i-j$ for the probability density (5.3). These values were computed by a metropolis Monte Carlo algorithm with 10^5 samples. The statistical error is less than 1%.

the Hamiltonian is translation invariant, these covariances depend only on the separation of j and k and are symmetric in $j-k$. We thus have $\frac{n}{2}$ numbers to determine: $\langle q_1 q_1 \rangle$, $\langle q_1 q_2 \rangle$, ..., $\langle q_1 q_{n/2} \rangle$. We compute these numbers by a metropolis Monte Carlo algorithm [6]. The covariances obtained in this way are shown in Figure 5.1. With the unconstrained means and covariances in hand, the regression formula (2.7) can be used.

We assume that the number of degrees of freedom n is large; at time $t = 0$, we observe a small number N of quantities of the form

$$\bar{u_\alpha}(0) = \sum_{j=1}^{n} g_{\alpha j} u_j, \quad \alpha = 1, \ldots, N, \tag{5.6}$$

and try to predict the mean future of the right-hand sides of equation (5.6) when the u_j evolve according to the full system of equations (5.1); i.e., we try to predict the mean of a small number of quantities that depend on the full system of ordinary differential equations without solving the latter. In the calculations we exhibit, we chose $n = 16$ and $N = 2$ so that we reduced the number of degrees of freedom by a factor of 8; we are also seeking to save the effort of averaging the solution of (5.1) over many initial data drawn from the canonical distribution (5.3). We pick

as constraining kernels discretized Gaussian functions centered at $j = 1$ and $j = 9$,

$$g_{1\,j} = \frac{1}{Z} \exp\left\{ -\frac{d^2(j,1)}{n^2\sigma^2} \right\}, \quad g_{2\,j} = \frac{1}{Z} \exp\left\{ -\frac{d^2(j,9)}{n^2\sigma^2} \right\}, \tag{5.7}$$

where $\sigma = 0.25$, $d(j,k)$ is a distance function over the periodic index axis (i.e., it is the minimum of $|j-k|$, $|j-k+n|$, and $|j-k-n|$) and Z is a constant that normalizes the kernels so that $\frac{1}{n}\sum_{j=1}^{n} g_{\alpha j} = 1$.

Again, our first approximation is with fixed constraining kernels $g_{\alpha j}$, evolving the values $\bar{p}_\alpha$ and $\bar{q}_\alpha$ of the constraints. Since the equations of motion are nonlinear, the time derivatives no longer commute with the regressions. Instead, we evaluate the mean time derivatives in (5.1) by Wick's theorem (equation (2.10)), which in this case reduces to

$$\left\langle \left[p_j - \langle p_j \rangle_{\overline{qp}} \right]^3 \right\rangle_{\overline{qp}} = 0\,, \tag{5.8}$$

from which we may deduce that

$$\begin{aligned} \left\langle \frac{dq_j}{dt} \right\rangle_{\overline{qp}} &= -\left\langle \frac{p_{j+1} - 2p_j + p_{j-1}}{\Delta^2} \right\rangle_{\overline{qp}} + \langle p_j^3 \rangle_{\overline{qp}} \\ &= -\left\langle \frac{p_{j+1} - 2p_j + p_{j-1}}{\Delta^2} \right\rangle_{\overline{qp}} - 2\langle p_j \rangle_{\overline{qp}}^3 + 3\langle p_j^2 \rangle_{\overline{qp}} \langle p_j \rangle_{\overline{qp}}\,. \end{aligned} \tag{5.9}$$

Our lemmas allow us to evaluate this expression explicitly in terms of the $\bar{q}_\alpha$ and $\bar{p}_\alpha$. Specifically,

$$\langle p_j \rangle_{\overline{qp}} = \left[\langle p_j p_k \rangle g_{\alpha k} \right] m^{-1}_{\alpha\beta} \bar{p}_\beta\,, \tag{5.10}$$

with

$$m_{\alpha\beta} = g_{\alpha j} \langle p_j p_k \rangle g_{\beta k}\,, \tag{5.11}$$

and

$$\langle p_j^2 \rangle_{\overline{qp}} = \langle p_j \rangle_{\overline{qp}}^2 + \langle p_j^2 \rangle - \left[\langle p_j p_k \rangle g_{\alpha k} \right] m^{-1}_{\alpha\beta} \left[g_{\beta l} \langle p_l p_j \rangle \right] \quad \text{(no sum over } j\text{)}\,. \tag{5.12}$$

In the approximation (3.8), the equations for the values of the constraints are obtained by multiplying equation (5.9) by $g_{\alpha j}$ (with summation over the index j). Substituting the numerical values of the (unconstrained) covariance matrix, we obtain the following equation for $\bar{q}_1$:

$$\frac{d\bar{q}_1}{dt} = -19.5(\bar{p}_2 - \bar{p}_1) + \left[1.50\,\bar{p}_1^3 - 0.88\,\bar{p}_1^2\bar{p}_2 + 0.27\,\bar{p}_1\bar{p}_2^2 + 0.11\,\bar{p}_2^3 \right]. \tag{5.13}$$

The equation for $\bar{q}_2$ is obtained by substituting $1 \leftrightarrow 2$; the equations for $\bar{p}_1$ and $\bar{p}_2$ are obtained by the transformation $\bar{p} \to \bar{q}$ and $\bar{q} \to -\bar{p}$.

We compare the effective equations (5.13) with the exact mean solution of the underlying equations (5.1) as follows: First, we generate many initial conditions consistent with the constraints. Next, we evolve each initial condition using a

Runge-Kutta method. Finally, we average the results for the variables $\overline{p}(t)$ and $\overline{q}(t)$ and compare with the prediction from the effective equations (5.13). Figure 5.2 shows this comparison. Once again, the system of two equations (5.13) reproduces the average behavior of the resolved equations (5.1) to the resolution of the plot, but at a very much smaller computational cost.

6 Example 3: A Stochastically Driven Heat Equation

We now turn to a more difficult problem: a linear equation with random forcing. This problem poses a challenge to our methodology because at first sight the dual evolution is undefined. However, we shall see that the machinery above extends to this problem if one focuses on the right variables. We consider a one-dimensional bar located at $0 \le x \le 1$, heated at its center by a fluctuating heat source and cooled at its ends by heat baths. The temperature $u(x,t)$ of the bar is described by

$$u_t = u_{xx} + J(t)\,\delta\left(x - \tfrac{1}{2}\right), \tag{6.1}$$

where $J(t)$ is a random function of time, $\delta(x)$ is a δ-function, and $x = \frac{1}{2}$ is the location of the source. The boundary conditions are

$$u(0,t) = u(1,t) = 0\,. \tag{6.2}$$

We shall use an even number of computational points so that the heat source always falls between them; the heat source being singular, this is an unresolved problem: The key contribution to the dynamics occurs between computational points. Assume that $J(t)$ is white noise; i.e., $J(t)$ is a Gaussian random function of time that has as mean and covariance:

$$\langle J(t)\rangle = \overline{J}\,, \qquad \langle J(t_1)J(t_2)\rangle = \overline{J}^2 + \overline{J^2}\,\delta(t_1 - t_2)\,, \tag{6.3}$$

for some numbers $\overline{J}$ and $\overline{J^2}$.

The solution $u(x,t)$ of (6.1) depends on $J(t)$ and on some initial condition for $u(x,0)$. However, after long times it depends only on $J(t)$:

$$u(x,t) = \int_{-\infty}^{0} dt'\,K(x,-t')\,J(t+t')\,, \tag{6.4}$$

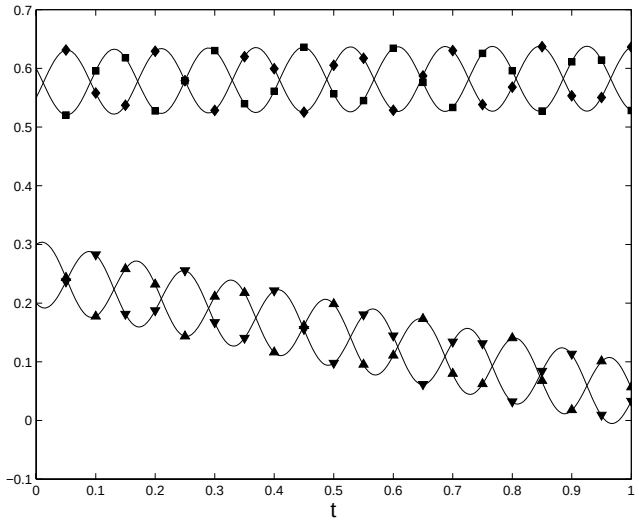
where $K(x,t)$ is the solution of

$$K_t = K_{xx}\,, \tag{6.5}$$

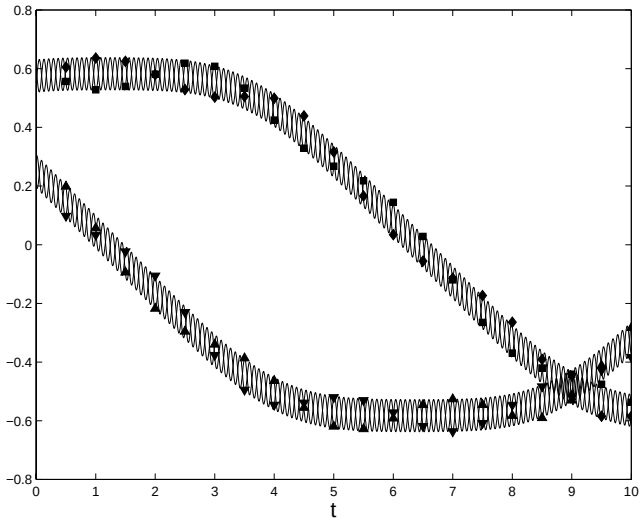
with initial condition $K(x,0) = \delta(x - \frac{1}{2})$ and boundary conditions

$$K(0,t) = K(1,t) = 0\,.$$

The heat kernel K is readily computed and so are, after suitably long times, average values of u and of functionals of u. Formula (6.3) defines the prior distribution of the solutions.



(a)



(b)

FIGURE 5.2. Evolution in time of the mean variables $\langle g_{\alpha i} p_i(t) \rangle_{\overline{pq}}$ for the nonlinear equations (5.1). The symbols represent the values of these quantities obtained by solving the 16 equations (5.1) for 10^4 initial conditions compatible with specific constraint values and averaging. The solid lines are the values of the four corresponding functions obtained by integrating equation (5.13). Figures (a) and (b) are for the time intervals $[0,1]$ and $[0,10]$, respectively.

In a Fourier representation, the kernel K can be written as

$$K(x,t) = \sum_{k=1}^{\infty} 2\sin\left(\frac{k\pi}{2}\right) e^{-k^2\pi^2 t}\sin(k\pi x)\,. \tag{6.6}$$

The unconstrained means are

$$\langle u(x)\rangle = \left\langle \int_0^\infty K(x,t)J(t)\,dt \right\rangle = \overline{J}\int_0^\infty K(x,t)\,dt = \frac{\overline{J}}{2}\begin{cases} x, & x < \frac{1}{2}\,, \\ 1-x, & x > \frac{1}{2}\,. \end{cases} \tag{6.7}$$

The unconstrained covariance is

$$\mathrm{Cov}[u(x),u(y)] \equiv \langle u(x)u(y)\rangle - \langle u(x)\rangle\langle u(y)\rangle = \overline{J^2}\int_0^\infty K(x,t)K(y,t)dt \tag{6.8}$$

(which we leave in integral form).

Suppose that we observe the function u at time $t = 0$ on a mesh of N points $x_\alpha = \alpha/N$. We denote the values of u at these mesh points by

$$\bar{u}_\alpha = u(x_\alpha)\,, \quad \alpha = 1,\ldots,N; \tag{6.9}$$

our constraints are of the standard form (1.3) with constraining kernels

$$g_\alpha(x) = \delta(x - x_\alpha)\,. \tag{6.10}$$

As in the previous example, we use a regression (2.7) to obtain the function $\langle u(x)\rangle_{\bar{u}}$ from the N constraints (6.9). The regression formula for this ensemble with δ-function constraining kernels is

$$\langle u(x)\rangle_{\bar{u}} = \langle u(x)\rangle + \mathrm{Cov}[u(x),u(x_\alpha)]m^{-1}_{\alpha\beta}\left[\bar{u}_\beta - \left\langle u(x_\beta)\right\rangle\right], \tag{6.11}$$

where the $m^{-1}_{\alpha\beta}$ are the entries of the matrix M^{-1} whose inverse is the matrix M with entries

$$m_{\alpha\beta} = \left\langle (g_\alpha,u)\left(g_\beta,u\right)\right\rangle - \left\langle (g_\alpha,u)\right\rangle\left\langle (g_\beta,u)\right\rangle\,. \tag{6.12}$$

Once again, this formula only involves readily evaluated Fourier series. An example of a regression with randomly chosen point values $\bar{u}(0)$ is shown in Figure 6.1.

To calculate the time evolution of the constrained measure, we proceed as follows: The values $u(x_\alpha,0)$ at time $t = 0$ constitute affine constraints on the stochastic source:

$$\bar{u}_\alpha(0) = \int_{-\infty}^0 K(x_\alpha,-t')J(t')dt'\,. \tag{6.13}$$

This constraint on J is of our standard form, with

$$g_\alpha(t',0) = K(x_\alpha,-t')\,. \tag{6.14}$$

The values of the source term at positive time do not affect the value of $u(x_\alpha,0)$, and thus as time advances the constraints recede into the past. Thus the constraint on J when the time has advanced to t is

$$g_\alpha(t',t) = K(x_\alpha,t-t')\,. \tag{6.15}$$

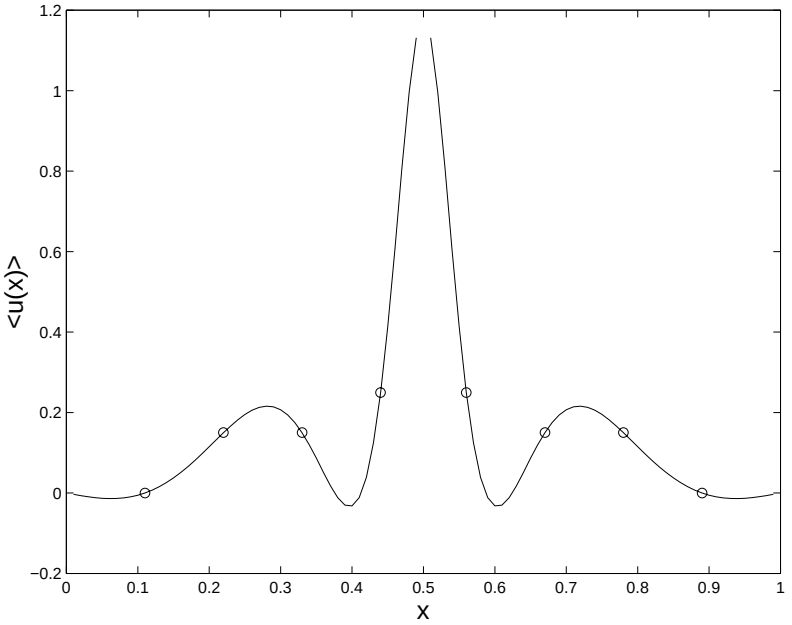


FIGURE 6.1. An example of a regression for the heat equation with $\overline{J} = 1$, $\overline{J^2} = 1$, and eight δ-function constraining kernels.

We apply our lemmas to obtain future constrained expectations of J and then compute the corresponding expectations of u through the relationship (6.4). The means and covariances of J without constraints were given in equation (6.3):

$$\langle J(t)\rangle = \overline{J}, \qquad a^{-1}(t_1,t_2) = \overline{J^2}\delta(t_1-t_2)\,. \tag{6.16}$$

The time-dependent constraints (6.15) allow us to define a time-dependent matrix M with entries

$$\begin{aligned} m_{\alpha\beta}(t) &= \big\langle (g_\alpha(t),J)(g_\beta(t),J)\big\rangle - \langle (g_\alpha(t),J)\rangle \big\langle (g_\beta(t),J)\big\rangle \\ &= \overline{J^2}\int_{-\infty}^{0} K(x_\alpha,t-t')K(x_\beta,t-t')dt'\,. \end{aligned} \tag{6.17}$$

The future constraints (6.15) imply that at time $t > 0$,

$$\begin{aligned} \big\langle J(t')\big\rangle_{\bar{u}} &= \big\langle J(t')\big\rangle + \left\{\int_{-\infty}^{0} g_\alpha(t'',t)a^{-1}(t'',t')dt''\right\} m^{-1}_{\alpha\beta}(t) \\ &\qquad \times [\bar{u}_\beta - \big\langle (g_\beta(t),J)\big\rangle] \\ &= \overline{J} + \overline{J^2}K(x_\alpha,t-t')\, m^{-1}_{\alpha\beta}(t)\, \left[\bar{u}_\beta - \overline{J}\int_{-\infty}^{0} K(x_\beta,t-t'')dt''\right]. \end{aligned} \tag{6.18}$$

Given the constrained J, we can calculate the future u: The mean of $u(x,t)$ given the N values $\bar{u}_\alpha(0) = u(x_\alpha, 0)$ is

$$\begin{aligned}\langle u(x,t)\rangle_{\bar{u}} &= \int_{-\infty}^{0} K(x,-t')\left\langle J(t')\right\rangle_{\bar{u}} dt' \\ &= \langle u(x)\rangle + \overline{J^2}\left\{\int_{-\infty}^{0} K(x,-t')K(x_\alpha,t-t')dt'\right\} m^{-1}_{\alpha\beta}(t) \\ &\qquad \times \left[\bar{u}_\beta - \left\langle u(x_\beta)\right\rangle\right].\end{aligned} \tag{6.19}$$

Future covariances can be obtained in a similar way. Note that the constraining kernels move backwards in time, a motion dual to the forward unfolding in time of the stationary stochastic process J. Our machinery thus applies provided one uses the appropriate variables. Note also that the prediction (6.19) decays to unconstrained averages after long times.

One may wonder whether it is possible to calculate the evolution of the solution u of the original equation (6.1) by evolving directly constraints on u without going back to J. The answer is apparently negative. No exact kernel evolution analogous to (3.3) can be defined, because no operator dual to L can be defined in the presence of the noise. One can also prove that the constant-kernel evolution approximation (3.8) is inaccurate in this case.

7 Conclusions and Further Work

We have presented methods for updating constraints in time that can be used to predict the future behavior of solutions of time-dependent partial differential equations on the basis of unresolved computations and partial data. We have concentrated on explaining the ideas, and the examples we have discussed have been relatively simple; in particular, they have explicitly known invariant measures that were either Gaussian or nearly Gaussian. For practical application, it is important to transcend these restrictive assumptions.

The assumption of near-Gaussianity is inessential. It can presumably be sometimes overcome by an assumption of local Gaussianity (i.e., the assumption that in a small neighborhood the probability densities are nearly Gaussian but with parameters that vary smoothly from neighborhood to neighborhood). A global alternative, when the prior measure is non-Gaussian, is to look for the measure that satisfies the constraints and is nearest to the prior measure; "nearest" should be interpreted in the sense of "having the smallest relative entropy with respect to the prior measure" [10]. Algorithms for finding such measures can be built in analogy to the constrained optimization methods of [7, 21, 22]. In strongly nonlinear problems, the simplified approximation (3.8) may be useful if it is supplemented by conditions on higher-order moments, as discussed at the end of Section 3.

A full knowledge of a prior measure is a luxury one cannot always expect. However, one can readily see from the discussion above that one can make do with the knowledge of covariances and maybe some higher-order moments and, in

addition, that this knowledge is needed only on scales comparable to the distance between computational points and smaller. Such knowledge is often available from asymptotics and scaling [2, 3, 4].

It is obvious that the constraints need not be all imposed at the initial time; subsequent information may sometimes be useful. Information about the right balance between increasing the number of data points and investing more effort into a careful updating of time-dependent constraining kernels awaits a broader experience with this type of prediction method.

Acknowledgment. We would like to thank Prof. G. I. Barenblatt, Mr. A. Gottlieb, Dr. D. Nolan, Prof. R. Miller, and Prof. B. Turkington for very helpful discussions, suggestions, and comments. This work was supported by the U.S. Department of Energy under contract DE-AC03-76-SF00098.

Bibliography

[1] Adrian, R. Stochastic estimation of sub-grid scale motions. *Appl. Mech. Rev.* **43** (1990), S214–S218.

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

[3] Barenblatt, G. I.; Chorin, A. J. New perspectives in turbulence: scaling laws, asymptotics, and intermittency. *SIAM Rev.* **40** (1998), no. 2, 265–291.

[4] Barenblatt, G. I.; Chorin, A. J. Scaling laws and vanishing viscosity limits in turbulence theory. *Recent Advances in Partial Differential Equations Venice 1996*, 1–25. Proceedings of Symposia in Applied Mathematics No. 54, American Mathematical Society, Providence, R.I., 1998.

[5] Bennett, A. F. *Inverse methods in physical oceanography*. Cambridge University Press, Cambridge, 1992.

[6] Binder, K.; Heermann, D. W. *Monte-Carlo simulation in statistical physics. An introduction.* 2nd ed., Springer Series in Solid-State Sciences, 80. Springer-Verlag, Berlin–New York, 1992.

[7] Boucher, C.; Turkington, B. Derivation of maximum entropy principles in two dimensional turbulence via large deviations. Preprint, University of Massachusetts, Amherst, 1998.

[8] Chorin, A. J. Vortex methods. *Mécanique des fluides numérique (Les Houches, 1993)*, 65–109. North-Holland, Amsterdam, 1996.

[9] Chorin, A. J.; Kast, A. P.; Kupferman, R. Optimal prediction of underresolved dynamics. *Proc. Natl. Acad. Sci. USA* **95** (1998), no. 8, 4094–4098.

[10] Dembo, A.; Zeitouni, O. *Large deviations techniques and applications*. Jones and Bartlett, Boston, 1993.

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

[12] Goldenfeld, N.; McKane, A.; Hou, Q. Block spins for partial differential equations. *J. Statist. Phys.* **93** (1998), no. 3–4, 699–714.

[13] Kleinert, H. *Gauge fields in condensed matter*. World Scientific, Singapore–Teaneck, N.J., 1989.

[14] McComb, W. D.; Watt, A. G. *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.

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

[16] Miller, R.; Carter, E.; Blue, S. Data assimilation into nonlinear stochastic models. *Tellus*, to appear.

[17] Papoulis, A. *Probability, random variables, and stochastic processes*. 3rd ed. McGraw-Hill, New York, 1991.

[18] Risken, H. *The Fokker-Planck equation. Methods of solution and applications*. Springer Series in Synergetics, 18. Springer-Verlag, Berlin–New York, 1984.

[19] Rozovskiĭ, B. L. *Stochastic evolution systems. Linear theory and applications to nonlinear filtering*. Translated from the Russian by A. Yarkho. Mathematics and Its Applications (Soviet Series), 35. Kluwer Academic, Dordrecht, 1990.

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

[21] Turkington, B. Statistical equilibrium measures and coherent states in two-dimensional turbulence. *Comm. Pure Appl. Math.* **52** (1999), no. 7, 781–809.

[22] Turkington, B.; Whitaker, N. Statistical equilibrium computations of coherent structures in turbulent shear layers. *SIAM J. Sci. Comput.* **17** (1996), no. 6, 1414–1433.

ALEXANDRE J. CHORIN
Lawrence Berkeley National Laboratory
Department of Mathematics
One Cyclotron Road
MS 50A-215
Berkeley, CA 94720
E-mail: chorin@math.berkeley.edu

ANTON P. KAST
Lawrence Berkeley National Laboratory
Department of Mathematics
One Cyclotron Road
MS 50A-215
Berkeley, CA 94720
E-mail: anton@math.lbl.gov

RAZ KUPFERMAN
Lawrence Berkeley National Laboratory
Department of Mathematics
One Cyclotron Road
MS 50A-215
Berkeley, CA 94720
E-mail: raz@snake.lbl.gov

Received May 1998.