ELSEVIER

Physica D 166 (2002) 239–257

PHYSICA D

www.elsevier.com/locate/physd

Optimal prediction with memory

Alexandre J. Chorin [a], Ole H. Hald [a], Raz Kupferman [b,*]

[a] *Department of Mathematics, University of California, Berkeley, CA 94720, USA*
[b] *Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel*

Received 9 July 2001; accepted 21 February 2002
Communicated by R. Temam

Abstract

Optimal prediction methods estimate the solution of nonlinear time-dependent problems when that solution is too complex to be fully resolved or when data are missing. The initial conditions for the unresolved components of the solution are drawn from a probability distribution, and their effect on a small set of variables that are actually computed is evaluated via statistical projection. The formalism resembles the projection methods of irreversible statistical mechanics, supplemented by the systematic use of conditional expectations and new methods of solution for an auxiliary equation, the orthogonal dynamics equation, needed to evaluate a non-Markovian memory term. The result of the computations is close to the best possible estimate that can be obtained given the partial data. We present the constructions in detail together with several useful variants, provide simple examples, and point out the relation to the fluctuation–dissipation formulas of statistical physics.

PACS: 05.10.−a; 05.40.−a

Keywords: Optimal prediction; Memory; Langevin equations; Orthogonal dynamics; Underresolution; Hamiltonian systems; Hermite polynomials

1. Introduction

Many problems in science and engineering are described by nonlinear differential equations whose solutions are too complicated to be properly resolved and/or where needed data are unavailable. The problem of predicting the evolution of such systems has been addressed by the present authors and others in [1–11]. Nothing can be predicted without some knowledge about the unresolved ("subgrid") degrees of freedom. In the papers just cited it is assumed that one possesses, as one often does, prior statistical information about the system in the form of an initial probability distribution; what is sought is a mean solution with respect to this initial distribution, compatible with the partial information available initially as well as with the limitations on the computing power one can bring to bear. This mean solution is the conditional expectation of the solution given the partial initial data, and is the best available estimate of the solution of the full problem.

* Corresponding author.
E-mail addresses: chorin@math.berkeley.edu (A.J. Chorin), hald@math.berkeley.edu (O.H. Hald), raz@math.huji.ac.il (R. Kupferman).

0167-2789/02/$ – see front matter
PII: S0167-2789(02)00446-3

The simplest construction of this conditional expectation, first-order optimal prediction ([3–5], see also below), produces a small system of ordinary differential equations and works well for a time that depends on the degree of underresolution and on the uncertainty in the initial conditions. This approximation is optimal in the class of Markovian approximations [10], but eventually exhibits errors because the influence of partial initial data on the distribution of the solutions weakens in time, and this loss of information is not captured, see [6,7]. As shown in the present paper, an accurate estimate of a subset of variables requires the addition of a "memory" term, and the resulting prediction scheme becomes a generalized Langevin equation, similar to those in irreversible statistical mechanics [12–16].

Some of the relations between conditional expectations and irreversible statistical mechanics have been discussed in [6]. The memory depends on a solution of an auxiliary equation, the orthogonal dynamics equation, and in the present paper we also present algorithms for finding this solution. We also explain how the machinery can lead to novel ways of using prior measurements to predict the future behavior of complex systems. We apply our methods to a simple model problem. Related, partial and more heuristic, constructions have been presented in [1,17].

2. Projections of dynamical systems and Langevin equations

Consider a system of ordinary differential equations,

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi(t) = R(\varphi(t)), \qquad \varphi(0) = x, \tag{1}$$

where φ and x are n-dimensional vectors with components φ_i and x_i, and R a vector-valued function with components R_i; t is the time. We denote the vector space in which φ and x reside by Γ; in classical statistical physics this space is the $n = 6N$-dimensional space of coordinates and momenta (q_j, p_j), where N is the number of particles in the system. The case where Γ is infinite-dimensional and (1) is a partial differential equation can be analyzed by the methods of [18].

To each initial condition x in (1) corresponds a trajectory, $\varphi(t) = \varphi(x, t)$; the initial value x is emphasized by this notation in view of its key role in what follows. The map $x \mapsto \varphi(x, t)$ is the flow map. Our goal is to calculate average values of the first m components of φ, $m < n$, without necessarily calculating all the components; the average is over all the values that the remaining $n - m$ components may initially take. We assume that prior information allows us to make statistical statements about the missing initial data. To shorten notations, we denote by $\hat{x}$ the m-dimensional vector whose entries are the resolved components, $(x_1, \ldots, x_m)$, and by $\tilde{x}$ the $(n - m)$-dimensional vector of unresolved components, $(x_{m+1}, \ldots, x_n)$; thus, $x = (\hat{x}, \tilde{x})$. Similarly, $\hat{\varphi}(x, t) = (\varphi_1(x, t), \ldots, \varphi_m(x, t))$ denotes the m components of the solution that we focus on.

Let $L = \sum_{i=1}^{n} R_i(x)\partial_i$, $(\partial_i = \partial/\partial x_i)$, and consider the linear partial differential equation

$$\frac{\partial}{\partial t}u(x, t) = Lu(x, t), \qquad u(x, 0) = g(x) \tag{2}$$

for some function $g(x)$. This is the Liouville equation. One can verify that the solution of this equation is $u(x, t) = g(\varphi(x, t))$. In particular, if $g(x) = x_i$, the solution is $u(x, t) = \varphi_i(x, t)$, the ith component of the solution of (1).

We use the semigroup notation $u(x, t) = (\mathrm{e}^{tL} g)(x) = g(\varphi(x, t))$, where e^{tL} is the evolution operator associated with the Liouville equation (2) (see, e.g. [19]). A short calculation shows that $\mathrm{e}^{tL} L = L\,\mathrm{e}^{tL}$. Eq. (2) becomes

$$\frac{\partial}{\partial t}\mathrm{e}^{tL} g = L\,\mathrm{e}^{tL} g = \mathrm{e}^{tL} Lg.$$

Suppose that the initial conditions x are drawn from a probability distribution μ, where $\mu(\mathrm{d}x) = \varrho(x)\,\mathrm{d}x$, and $\varrho(x)$ a probability density function. Given the initial values of the m coordinates $\hat{x}$, the distribution of the remaining $n - m$ coordinates $\tilde{x}$ is given by the conditional measure, μ conditioned by $\hat{x}$. If the system (1) is Hamiltonian

with Hamiltonian H, one can use as initial distribution the canonical distribution with density $\varrho(x) = Z^{-1}\,\mathrm{e}^{-H(x)}$, where Z is a normalization constant. Hamiltonian systems are often of interest, and the canonical distribution is often natural for physical reasons. These choices simplify parts of the analysis.

Given μ, functions on Γ can be viewed as random variables, and one can use the terminology of probability theory. We define the expected value of g by

$$E[g] = \int_\Gamma g(x)\varrho(x)\,\mathrm{d}x.$$

We endow the space of functions on Γ with the inner product $(f, g) = E[fg]$, which makes it a Hilbert space $L^2(\Gamma, \mu)$ (L^2 for brevity). If (1) is a Hamiltonian system and the probability density is $\varrho(x) = Z^{-1}\exp(-H(x))$, then the operator L is skew-adjoint in this Hilbert space.

We now derive an equation, often referred to as the generalized Langevin equation, which is a reformulation of the equations of motion (1) for the resolved variables $\hat{\varphi}(x, t)$. The derivation uses projection operators: functions in L^2 are projected onto the space $\hat{L}^2$ of functions of the m-dimensional vector $\hat{x}$. Several different projections P are considered:

(1) Let $f \in L^2$, and consider the orthogonal projection of f onto the span of all functions of $\hat{x}$, given by

$$(Pf)(\hat{x}) = \frac{\int f(x)\varrho(x)\,\mathrm{d}\tilde{x}}{\int \varrho(x)\,\mathrm{d}\tilde{x}}, \quad \mathrm{d}\tilde{x} = \mathrm{d}x_{m+1}\cdots\mathrm{d}x_n.$$

In the language of probability, $(Pf)(\hat{x})$ is the conditional expectation of f given $\hat{x}$ and is denoted by $E[f|\hat{x}]$; see [20]. $E[f|\hat{x}]$ is the best approximation of f by a function of $\hat{x}$:

$$E[|f - E[f|\hat{x}]|^2] \le E[|f - h(\hat{x})|^2]$$

for all functions h in $\hat{L}^2$. P is the "nonlinear projection", used in [16] with a different interpretation, as well as in [3,4,6].

(2) Given $f \in L^2$, define

$$(Pf)(\hat{x}) = \sum_{i,j=1}^{m} a_{ij}^{-1}(f, x_i)x_j,$$

where a_{ij}^{-1} are the entries of a matrix whose inverse has entries $a_{ij} = (x_i, x_j)$. This is the linear projection widely used in irreversible statistical mechanics (see [12,13,21]).

(3) More generally, pick a set of functions of $\hat{x}$, say $h^\nu(\hat{x})$, $\nu = 1, \ldots, M$; for convenience, make them orthonormal: $(h^\nu, h^\mu) = \delta_{\mu\nu}$. Define a projection

$$(Pf)(\hat{x}) = \sum_{\nu=1}^{M}(f, h^\nu)h^\nu(\hat{x}).$$

If the h^ν span $\hat{L}^2$ as M increases, the result approximates the conditional expectation $E[f|\hat{x}]$. This is the finite-rank projection; it interpolates between the linear projection and the conditional expectation.

We now follow the Mori–Zwanzig procedure [12,14,15,21]. We consider the equation of motion for a resolved coordinate $\varphi_j(x, t) = \mathrm{e}^{tL}x_j$, and split the time derivative, $R_j(\varphi(x, t)) = \mathrm{e}^{tL}Lx_j$, into a term that depends only on $\hat{\varphi}(x, t)$ plus a remainder:

$$\frac{\partial}{\partial t}\mathrm{e}^{tL}x_j = \mathrm{e}^{tL}Lx_j = \mathrm{e}^{tL}PLx_j + \mathrm{e}^{tL}QLx_j, \tag{3}$$

where $Q = I - P$. Define $\mathrm{R}_j(\hat{x}) = (PR_j)(\hat{x})$; the first term is $\mathrm{e}^{tL} PLx_j = \mathrm{R}(\hat{\varphi}(x,t))$ and is a function of the resolved components of the solution only.

We further split the remaining term $\mathrm{e}^{tL} QLx_j$ as follows: Let $w(x,t)$ be a solution of the orthogonal dynamics equation:

$$\frac{\partial}{\partial t} w(x,t) = QLw(x,t) = Lw(x,t) - PLw(x,t), \qquad w(x,0) = QLx_j. \tag{4}$$

In semigroup notation, $w(x,t) = \mathrm{e}^{tQL} QLx_j$. An existence proof for Eq. (4) may be found in [22]. One verifies that if $Pf = 0$, then $P\,\mathrm{e}^{tQL} f = 0$ for all time t, i.e., e^{tQL} maps the null space of P into itself.

The evolution operators e^{tL} and e^{tQL} satisfy the Dyson formula [12]:

$$\mathrm{e}^{tL} = \mathrm{e}^{tQL} + \int_0^t \mathrm{e}^{(t-s)L} PL\,\mathrm{e}^{sQL}\,\mathrm{d}s,$$

which can be checked by differentiation. Hence,

$$\mathrm{e}^{tL} QLx_i = \mathrm{e}^{tQL} QLx_j + \int_0^t \mathrm{e}^{(t-s)L} PL\,\mathrm{e}^{sQL} QLx_j\,\mathrm{d}s. \tag{5}$$

Let

$$F_j(x,t) = \mathrm{e}^{tQL} QLx_j, \qquad K_j(\hat{x},t) = PLF_j(x,t).$$

Note that multiplication by P always yields a function of $\hat{x}$. Collecting terms, we obtain the generalized Langevin equation:

$$\frac{\partial}{\partial t}\mathrm{e}^{tL} x_j = \mathrm{e}^{tL}\,\mathrm{R}_j(\hat{x},t) + \int_0^t \mathrm{e}^{(t-s)L} K_j(\hat{x},s)\,\mathrm{d}s + F_j(x,t).$$

This is an identity, which in a more transparent form reads

$$\frac{\partial}{\partial t}\varphi_j(x,t) = \mathrm{R}_j(\hat{\varphi}(x,t)) + \int_0^t K_j(\hat{\varphi}(x,t-s),s)\,\mathrm{d}s + F_j(x,t). \tag{6}$$

The various terms in Eq. (6) have conventional interpretations. The first term on the right-hand side is the Markovian contribution to $\partial_t\varphi_j(x,t)$—it depends only on the instantaneous value of the resolved $\hat{\varphi}(x,t)$. The second term depends on x through the values of $\hat{\varphi}(x,s)$ at times s between 0 and t, and embodies a memory—a dependence on the past values of the resolved variables. Finally, the third term, which depends on full knowledge of the initial conditions x, lies in the null space of P and can be viewed as noise with statistics determined by the initial conditions. The fact that the memory depends on the noise is known in the physics literature as a fluctuation–dissipation theorem. The specific form of this relation given in physics books is obtained when P is the linear projection.

The last step is the multiplication of (6) by P:

$$\frac{\partial}{\partial t} P\varphi_j(x,t) = P\mathrm{R}_j(\hat{\varphi}(x,t)) + \int_0^t PK_j(\hat{\varphi}(x,t-s),s)\,\mathrm{d}s. \tag{7}$$

This identity involves only the known components $\hat{x}$ of the initial data. When P is the conditional expectation, $P\hat{\varphi}(x,t) = E[\hat{\varphi}(x,t)|\hat{x}]$, the right-hand side of Eq. (7) is exactly what we want: the derivative of the average of $\hat{\varphi}(x,t)$ conditioned by the initially given data.

Eqs. (6) and (7) are identities; their solution is exactly equivalent to the solution of the full problem (1) followed by averaging. In practice these equations have to be solved approximately; we shall show below how to perform the approximation term by term.

3. A model problem

We introduce a model problem to illustrate the formalism of the previous section and to test the accuracy of various approximations described below.

Consider the dynamical systems in $\Gamma = \Re^4$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_1 = \varphi_2, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\varphi_2 = -\varphi_1(1+\varphi_3^2), \qquad \frac{\mathrm{d}}{\mathrm{d}t}\varphi_3 = \varphi_4, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\varphi_4 = -\varphi_3(1+\varphi_1^2). \tag{8}$$

Eq. (8) are derived from the Hamiltonian

$$H(x) = \tfrac{1}{2}x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1^2x_3^2$$

with (x_1, x_2) and (x_3, x_4) canonical pairs of coordinates. This system describes two oscillators with a nonlinear coupling.

We take initial data randomly distributed with the canonical probability density $\varrho(x) = Z^{-1}\,\mathrm{e}^{-H(x)}$, thus endowing the space of functions on Γ with the inner product

$$(f, g) = \frac{\int f(x)g(x)\,\mathrm{e}^{-H(x)}\,\mathrm{d}x}{\int \mathrm{e}^{-H(x)}\,\mathrm{d}x}, \tag{9}$$

where $\mathrm{d}x = \mathrm{d}x_1 \cdots \mathrm{d}x_4$. We retain only two of the four variables, φ_1 and φ_2, thus $\hat{x} = (x_1, x_2)$ and $\tilde{x} = (x_3, x_4)$. The goal is to compute the average of $\varphi_1(x, t)$ and $\varphi_2(x, t)$ over all initial choices of x_3 and x_4.

In Fig. 1, we plot the time evolution of the mean values of $\varphi_1(x, t)$ and $\varphi_2(x, t)$, given $\varphi_1(x, 0) = 1$ and $\varphi_2(x, 0) = 0$. This graph was obtained by a Monte Carlo calculation: we generated a collection of 5×10^4 initial

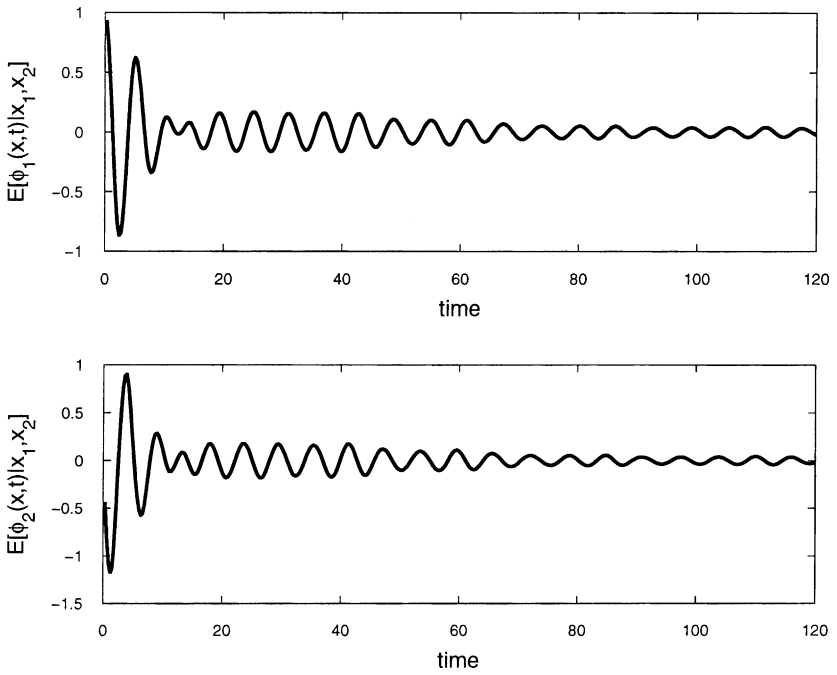

Fig. 1. Mean solutions $E[\varphi_1(x, t)|\hat{x}]$ (top) and $E[\varphi_2(x, t)|\hat{x}]$ (bottom) for the initial data $\hat{x} = (1, 0)$. The mean solution was computed by evolving in time a set of 5×10^4 solutions with initial conditions generated by Monte Carlo sampling.

conditions x by fixing $\hat{x} = (1, 0)$ and sampling x_3,x_4 from the canonical distribution. Each initial datum was evolved in time with an ODE solver. At each $t > 0$ the mean values of $\varphi_1(x, t)$ and $\varphi_2(x, t)$ were evaluated by averaging over the set of solutions. Note the decay of the mean solution towards its equilibrium value; this phenomenon has been explained in [6].

We now write down explicitly each of the projections defined in Section 2:

(1) The conditional expectation of a function $f(x)$ is

$$(Pf)(\hat{x}) = E[f|\hat{x}] = \frac{\int f(x)\,\mathrm{e}^{-H(x)}\,\mathrm{d}\tilde{x}}{\int \mathrm{e}^{-H(x)}\,\mathrm{d}\tilde{x}} \tag{10}$$

with $\mathrm{d}\tilde{x} = \mathrm{d}x_3\,\mathrm{d}x_4$. The density $\mathrm{e}^{-H(x)}$ is Gaussian when x_1, x_2 are fixed, thus the integrals in (10) can often be calculated explicitly. For example,

$$Px_3^{2n} = \binom{2n}{2}(1 + x_1^2)^{-n}, Px_4^{2n} = \binom{2n}{2}, \qquad n = 1, 2, \ldots .$$

(2) It is easy to verify that $(x_1, x_1) = c = 0.715$, $(x_2, x_2) = 1$, and $(x_1, x_2) = 0$, and so the linear projection is

$$(Pf)(\hat{x}) = c^{-1}(f, x_1)x_1 + (f, x_2)x_2.$$

(3) For functions f, g that depend only on $\hat{x}$, the inner product (9) takes the form:

$$(f, g) = \frac{Z^{-1}}{2\pi}\int \frac{\mathrm{e}^{-(1/2)x_1^2}\,\mathrm{e}^{-(1/2)x_2^2}}{\sqrt{1 + x_1^2}} f(\hat{x})g(\hat{x})\,\mathrm{d}\hat{x},$$

where

$$Z = \frac{1}{2\pi}\int \frac{\mathrm{e}^{-(1/2)x_1^2}\,\mathrm{e}^{-(1/2)x_2^2}}{\sqrt{1 + x_1^2}}\,\mathrm{d}\hat{x} = 0.78964.$$

Let $\alpha > -1/2$ be a parameter. For each fixed value of α the following family of functions constitutes an orthonormal basis in the space of square integrable functions of $\hat{x}$:

$$h^{\nu}(\hat{x}) = Z^{1/2}(1 + x_1^2)^{1/4}\tilde{H}_{\nu_1}(x_1)\tilde{H}_{\nu_2}(x_2), \tag{11}$$

where $\nu = (\nu_1, \nu_2)$, $\nu_{1,2} = 0, 1, \ldots$, and

$$\tilde{H}_k(z) = (1 + 2\alpha)^{1/4}H_k(\sqrt{1 + 2\alpha}z)\,\mathrm{e}^{-\alpha z^2/2}.$$

Here the $H_k(z)$ are Hermite polynomials satisfying the recursion relation

$$H_0(z) = 1, \qquad H_1(z) = z, \qquad H_k(z) = \frac{1}{\sqrt{k}}zH_{k-1}(z) - \sqrt{\frac{k-1}{k}}H_{k-2}(z).$$

For future use, we also note that

$$\frac{\mathrm{d}}{\mathrm{d}z}\tilde{H}_k(z) = \sqrt{1 + 2\alpha}\tilde{H}_{k-1}(z) - \alpha z\tilde{H}_k(z). \tag{12}$$

The span of a finite collection of these functions changes when α changes; we will use the freedom of choosing α to optimize the rate at which the finite-rank projection converges to the conditional expectation.

4. Conditional expectations calculated from previous measurements

Consider the problem (1) and suppose we know a large number of its solutions $\varphi(x,t)$ for various initial conditions x drawn from the distribution μ; these solutions may come from prior Monte Carlo computation or from experiment. The best L^2 estimate of the solution when $\hat{x}$ is given is $E[\varphi(x,t)|\hat{x}]$.

Let $h^\nu(\hat{x})$, ν in some finite index set I, be a set of orthonormal basis functions; we approximate the conditional expectation $E[\varphi_j(x,t)|\hat{x}]$ by a finite-rank projection

$$E[\varphi_j(x,t)|\hat{x}] = \sum_\nu (\varphi_j(t), h^\nu) h^\nu(\hat{x}) \approx \sum_{\nu\in I} c_j^\nu(t) h^\nu(\hat{x}), \tag{13}$$

where the inner products are integrations over Γ.

This approximation makes it possible to use information from collections of prior measurements to predict the behavior of a particular system with partially known initial conditions. If one has many functions $\varphi(x,t)$, one can evaluate the coefficients $c_j^\nu(t)$ and then make an optimal prediction for a specific case by substituting the known values of the initial data $\hat{x}$ into the right-hand side of (13).

Here we remind the reader of a basic fact of numerical analysis: Approximation by a finite set of orthogonal functions, especially on an infinite interval, may converge poorly (see, e.g. [23,24]); it is prudent to check the convergence, for example by checking the Bessel inequality. As an illustrative example, suppose we want to approximate the function $f(x) = \cos(xt)$ (t is a parameter) in the inner product space,

$$(f,g) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) g(x)\, \mathrm{e}^{-(1/2)x^2}\, \mathrm{d}x. \tag{14}$$

In this case, the functions $\tilde{H}_k(x)$ defined above form an orthonormal basis. If one approximates $f(x)$ by a finite sum of the form

$$\cos(xt) \approx \sum_{k=0}^{N} a_k(t) \tilde{H}_k(x), \tag{15}$$

where

$$a_k(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(xt) \tilde{H}_k(x)\, \mathrm{e}^{-(1/2)x^2}\, \mathrm{d}x,$$

the quality of the approximation can be assessed by the L^2 error:

$$\left(E[\cos^2(xt)] - \sum_{k=0}^{N} a_k^2(t) \right)^{1/2}.$$

In Fig. 2, we compare the variation in time of the L^2 norm of the function $\cos(xt)$ (solid line) with the L^2 norm of the finite sum (15) with $N = 7$ terms (dashed line). The dotted lines represent the contributions from $a_k^2(t)$ for $k = 1, 3, 5, 7$; by symmetry $a_{2k}(t) = 0$. The four graphs correspond to four different values of the parameter α. The finite sums approximate $\cos(xt)$ well for short times; the larger t, more modes are needed for an accurate approximation. These graphs demonstrate that a proper scaling of the basis is important for accuracy with an acceptable number of terms.

We now return to the model problem (8) and approximate the conditional expectation $E[\varphi_1(x,t)|\hat{x}]$ by a finite-rank projection of the form (13), with the $h^\nu(\hat{x})$ given by (11). The functions $c_1^\nu(t)$ are evaluated by averaging the products $\varphi_1(x,t)h^\nu(\hat{x})$ over a collection of 5×10^4 numerical solutions of (8) with initial conditions drawn from the canonical distribution.

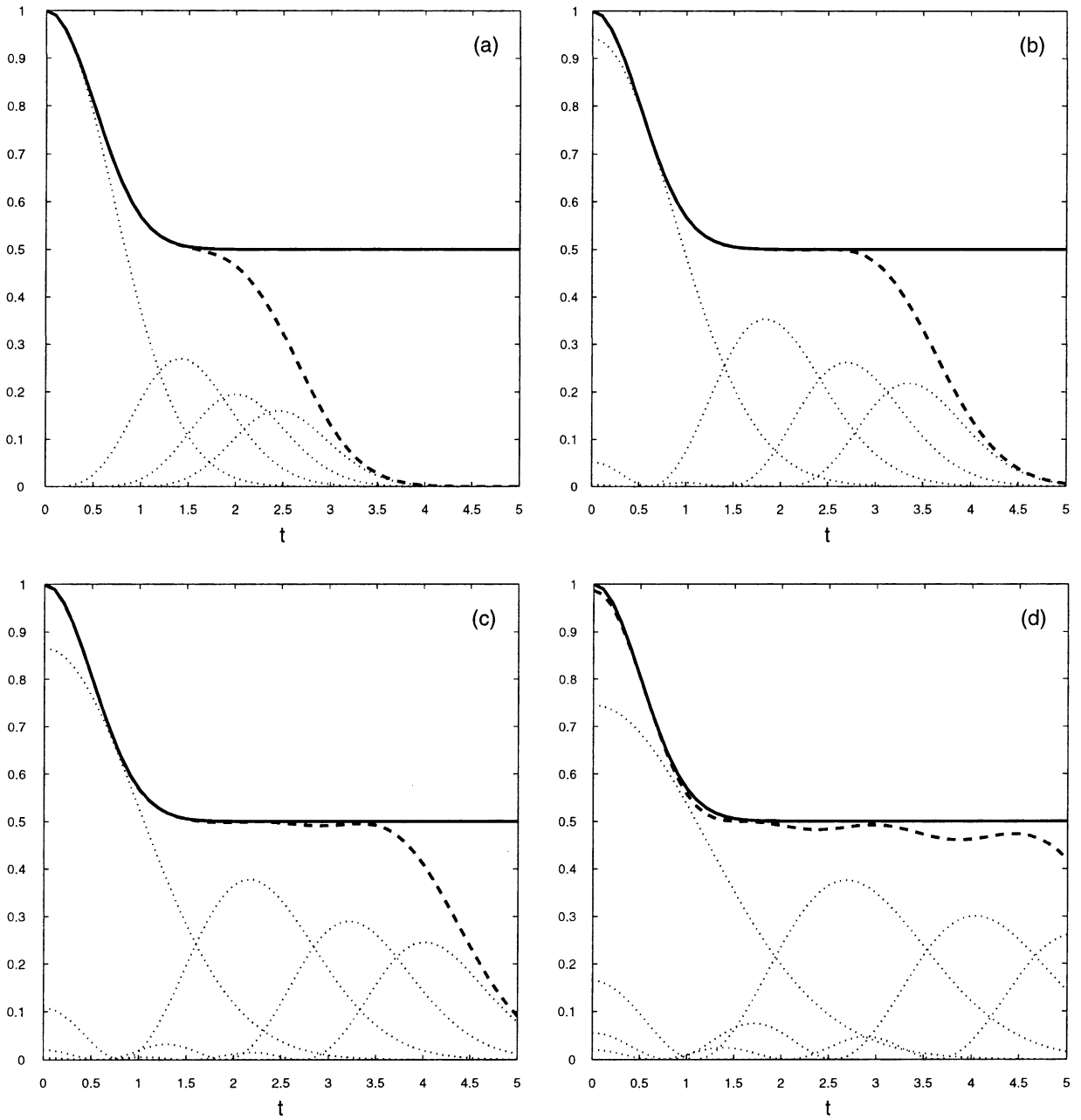

Fig. 2. Solid lines: the variation of the L^2 norm of $\cos(xt)$ in the inner product space (14). Dashed lines: the L^2 norm of the finite-rank approximation (15) with $N = 7$ terms. Dotted lines: $a_k^2(t)$ for $k = 1, 3, 5, 7$. The four plots correspond to: (a) $\alpha = 0$, (b) $\alpha = 0.5$, (c) $\alpha = 1$, and (d) $\alpha = 2$.

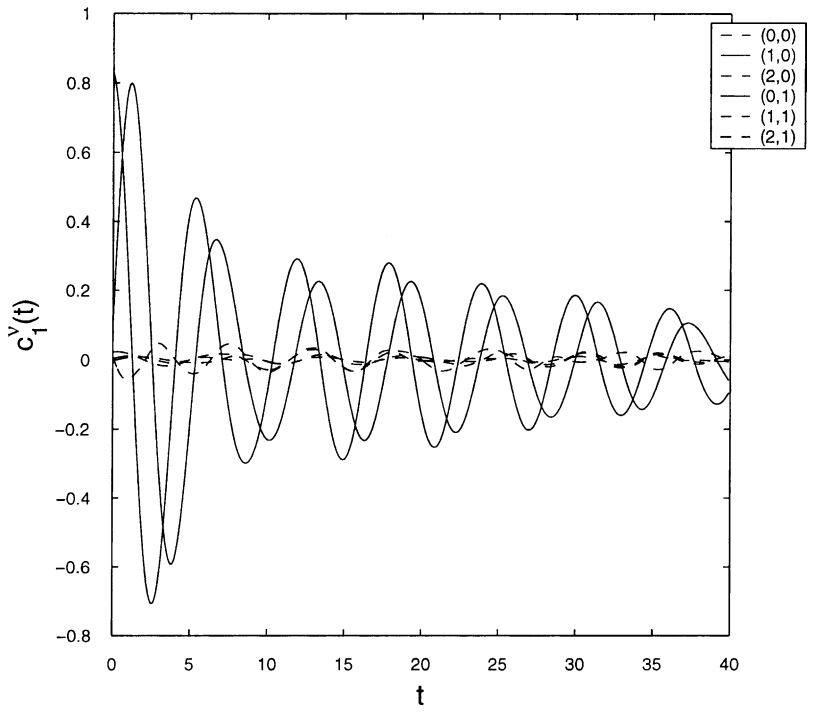

Fig. 3. $c_1^\nu(t)$ versus t for $\alpha = 0$ and various values of ν. The two dominant functions correspond to $\nu = (1, 0)$ and $\nu = (0, 1)$.

In Fig. 3, we plot several of the functions $c_1^\nu(t)$ with $\alpha = 0$, and observe that the two dominant contributions come from the components $\nu = (1, 0)$ and $\nu = (0, 1)$. This is consistent with the assumption often made in physics that the lower order terms are the most significant.

In Fig. 4, we compare the mean solution, $E[\varphi_1(x, t)|\hat{x}]$, $\hat{x} = (1, 0)$, generated by Monte Carlo sampling (dotted lines), with the finite-rank projection (13) (solid lines). The top two graphs correspond to $\alpha = 0$ and to (a) 3×3 and (b) 6×6 basis functions. In the first case, the finite-rank approximation deviates from the true solution already at short times, although the qualitative properties of the mean solution are well captured. With four times as many basis functions, better accuracy is preserved for a long time. In Fig. 4 (c) and (d), the value of α is modified, indicating how a significant reduction in computational effort can be obtained by scaling the basis functions.

Finally, note that the integrals that define the coefficients in the expansions are over all the components of x while the basis functions are functions only of $\hat{x}$; therefore the series expansion is an expansion, not of $\varphi(x, t)$ which may be very noisy, but of $E[\varphi(x, t)|\hat{x}]$ which is much smoother.

5. The Markovian term in the Langevin equation

We now examine the Markovian term $\mathrm{R}(\hat{\varphi}(x, t))$ in the Langevin equation (6). For the model problem (8) and P the conditional expectation, this term can be calculated explicitly:

$$\mathrm{R}(\hat{\varphi}(x, t)) = \begin{pmatrix} \varphi_2(x, t) \\ -\varphi_1(x, t)\left[1 + \dfrac{1}{1 + \varphi_1^2(x, t)}\right] \end{pmatrix}.$$

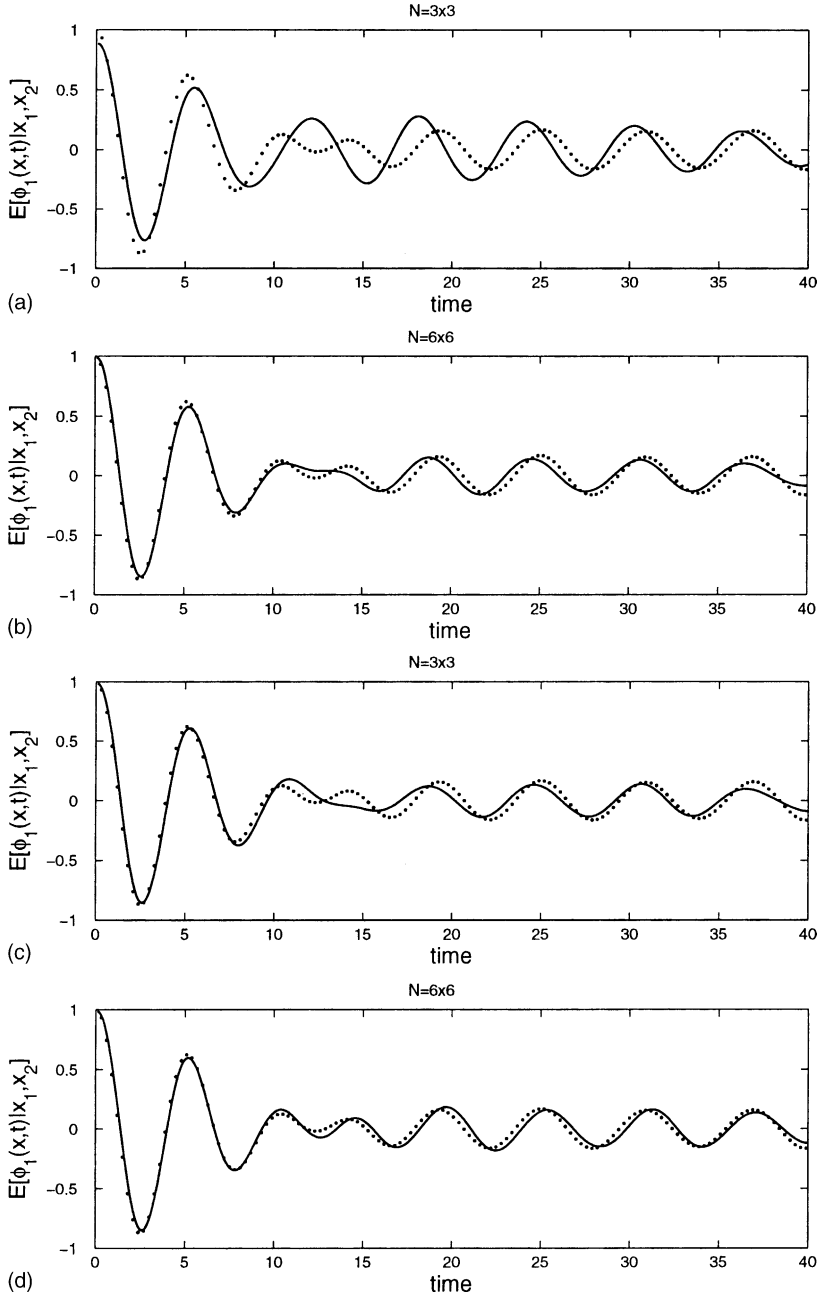

Fig. 4. Comparison of the mean solution $E[\varphi_1(x,t)|\hat{x}]$ (dotted lines) and the finite-rank approximation, $\sum_{\nu\in I} c_1^\nu(t)h^\nu(\hat{x})$ (solid lines) for the initial data $\hat{x} = (1, 0)$. The different graphs correspond to: (a) $I = \{0, 1, 2\}^2$ and $\alpha = 0$, (b) $I = \{0, \ldots, 5\}^2$ and $\alpha = 0$, (c) $I = \{0, 1, 2\}^2$ and $\alpha = 1$, (d) $I = \{0, \ldots, 5\}^2$ and $\alpha = 2$.

This expression is a function of all the components of x, not only of the $m = 2$ that we wish to work with. Next we apply the projection P and we do so now by interchanging the evaluations of R and P:

$$P\mathrm{R}(\hat{\varphi}(x,t) \approx \mathrm{R}(P\hat{\varphi}(x,t)).$$

The reader should not be horrified by the commutation of an average with a nonlinear function. The randomness in the problem is a reflection of the unresolved degrees of freedom in x. One alternative to our methodology is neglecting these degrees of freedom, which removes the randomness and makes the commutation perfectly legitimate. One is better off preserving these degrees of freedom and mistreating them slightly rather than omitting them. Another possible construction, more accurate but more expensive, consists of storing samples of $\mathrm{R}(\hat{\varphi}(x,t))$ for initial data drawn from the initial distribution and then projecting R just as we projected $\hat{\varphi}(x,t)$ in the previous section.

In what was called "first-order optimal prediction" in [3–5], the second and third terms in the generalized Langevin equation (6) are dropped; writing $\Phi(t) = P\hat{\varphi}(x,t)$ one obtains the approximate equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t) = \mathrm{R}(\Phi(t)), \qquad \Phi(0) = \hat{x}. \tag{16}$$

A convergence proof for this approximation applied to a nonlinear partial differential equation can be found in [10].

For the model problem (8), $\Phi(\hat{x},t) = (\Phi_1(\hat{x},t), \Phi_2(\hat{x},t))$, and the first-order optimal prediction equations are

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_1 = \Phi_2, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\Phi_2 = -\Phi_1\left(1 + \frac{1}{1+\Phi_1^2}\right). \tag{17}$$

As observed in [6], Eq. (16) are Hamilton's equations derived from the "effective" Hamiltonian

$$\mathrm{H}(\hat{x}) = -\log\int \mathrm{e}^{-H(x)}\,\mathrm{d}\tilde{x}$$

provided that if a variable is resolved so is its canonically conjugate variable.

In Fig. 5, we compare $\Phi_1(t)$, obtained by the integration of Eq. (17) with initial conditions $\Phi(0) = (1,0)$, to the function $E[\varphi_1(x,t)|\hat{x}]$, $\hat{x} = (1,0)$, resulting from the Monte Carlo sampling. The discrepancy between the two curves is due to the omission of the memory, see [7,6].

While first-order optimal prediction is accurate (in fact, optimal) only for short times, it may be exploited for longer times as a numerical "predictor" to improve the convergence rate of the finite-rank approximation (13). The system (16) is Hamiltonian and in particular time reversible, hence the flow map $\hat{x} \mapsto \Phi(\hat{x},t)$ is continuous and invertible (although it may be very complex). Thus, the mean solution $E[\hat{\varphi}(x,t)|\hat{x}]$ can be written as a function of $\Phi(\hat{x},t)$. Since $\Phi(\hat{x},t)$ approximates $E[\hat{\varphi}(x,t)|\hat{x}]$ well for short times, a finite-rank expansion of the latter in terms of the former may exhibit better convergence. The change of variables from $\hat{x}$ to $\Phi(\hat{x},t)$ is made easier for a Hamiltonian system by the following observation: If $h^\nu(\hat{x})$, $\nu \in I$, are orthonormal functions in the inner product space (9), and $\Phi(\hat{x},t)$ is the solution of (16), then the composite functions $h^\nu[\Phi(\hat{x},t)]$, t fixed, are orthonormal with respect to the same inner product. Indeed, let $\mu, \nu \in I$. Fix t, and consider

$$\begin{aligned}\ell^{\mu\nu} &= \int h^\mu[\Phi(\hat{x},t)]h^\nu[\Phi(\hat{x},t)]\,\mathrm{e}^{-H(x)}\,\mathrm{d}x = \int h^\mu[\Phi(\hat{x},t)]h^\nu[\Phi(\hat{x},t)]\left(\int \mathrm{e}^{-H(x)}\,\mathrm{d}\tilde{x}\right)\mathrm{d}\hat{x}\\ &= \int h^\mu[\Phi(\hat{x},t)]h^\nu[\Phi(\hat{x},t)]\,\mathrm{e}^{-\mathrm{H}(\hat{x})}\,\mathrm{d}\hat{x}.\end{aligned}$$

Change variables: $\hat{y} = \Phi(\hat{x},t)$. Since the map $\hat{x} \to \Phi(\hat{x},t)$ is Hamiltonian, it preserves the Lebesgue measure $\mathrm{d}\hat{x} = \mathrm{d}\hat{y}$ and the effective Hamiltonian $\mathrm{H}(\hat{x}) = \mathrm{H}(\Phi(\hat{x},t))$. Thus,

$$\ell^{\mu\nu} = \int h^\mu(\hat{y})h^\nu(\hat{y})\,\mathrm{e}^{-\mathrm{H}(\hat{y})}\,\mathrm{d}\hat{y} = \int h^\mu(\hat{y})h^\nu(\hat{y})\,\mathrm{e}^{-H(y)}\,\mathrm{d}y = \delta_{\mu\nu}.$$

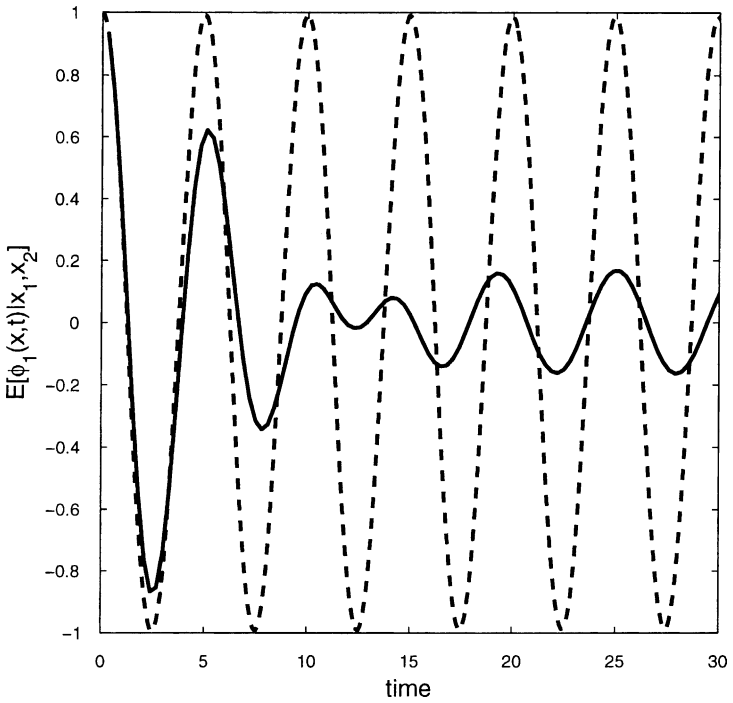

Fig. 5. Comparison of the mean solution $E[\varphi_1(x,t)|\hat{x}]$ (solid line) and the component $\Phi_1(t)$ of the solution of the optimal prediction equation (17) (dashed line) for the initial data $\hat{x} = (1, 0)$.

One can therefore replace the finite-rank projection (13) by

$$E[\varphi_j(x,t)|\hat{x}] \approx \sum_{\nu\in I}(\varphi_j(y,t), h^\nu[\Phi(\hat{y},t)])h^\nu[\Phi(\hat{x},t)].$$

6. Orthogonal dynamics and the memory kernel

We turn now to a general formalism for the evaluation of the noise $F(x,t)$ and the memory kernel $K(\hat{x},t)$. For each $j \le m$, the component $F_j(x,t)$ of the noise is the solution of the orthogonal dynamics equation

$$\frac{\partial}{\partial t}F_j(x,t) = QLF_j(x,t) = LF_j(x,t) - PLF_j(x,t), \tag{18}$$
$$F_j(x,0) = QLx_j = R_j(x) - \mathrm{R}_j(\hat{x}).$$

Each F_j is computed independently of the others. Eq. (18) is equivalent to the Dyson formula:

$$F_j(x,t) = \mathrm{e}^{tL}\,F_j(x,0) - \int_0^t \mathrm{e}^{(t-s)L}\,PLF_j(x,s)\,\mathrm{d}s. \tag{19}$$

This is a Volterra integral equation for $F_j(x,t)$, which we next decompose into its components in an orthonormal basis: Let P be a finite-rank projection. Then,

$$K_j(\hat{x},s) = PLF_j(x,s) = \sum_{\nu\in I} a_j^\nu(s)h^\nu(\hat{x}), \tag{20}$$

where

$$a_j^\nu(s) = (LF_j(s), h^\nu).$$

Consequently,

$$\mathrm{e}^{(t-s)L} PLF_j(x,s) = \sum_{\nu\in I} a_j^\nu(s) h^\nu(\hat{\varphi}(x, t-s)).$$

Substitute this equation into (19), multiply both sides by L, and take the inner product with h^μ; the result is

$$(LF_j(t), h^\mu) = (L\,\mathrm{e}^{tL} F_j(0), h^\mu) - \int_0^t \sum_{\nu\in I} a_j^\nu(s)(L\,\mathrm{e}^{(t-s)L} h^\nu, h^\mu\,\mathrm{d}s.$$

This is a Volterra integral equation for the functions $a_j^\nu(t)$, which can be rewritten as follows:

$$a_j^\mu(t) = f_j^\mu(t) - \int_0^t \sum_{\nu\in I} a_j^\nu(s) g^{\nu\mu}(t-s)\,\mathrm{d}s, \tag{21}$$

where

$$f_j^\mu(t) = (L\,\mathrm{e}^{tL} F_j(0), h^\mu), \qquad g^{\nu\mu}(t) = (L\,\mathrm{e}^{tL} h^\nu, h^\mu).$$

The functions $f_j^\nu(t)$, $g^{\mu\nu}(t)$ can be found by averaging over a collection of experiments or simulations, with initial conditions drawn from the initial distribution without reference to any specific realization; they are time-correlation functions (i.e., inner products of the form $(\mathrm{e}^{tL} g_1, g_2)$ for some functions (g_1, g_2) and not dependent on the orthogonal dynamics. The number of components a^ν depends only on the resolution needed in the space of functions of $\hat{x}$. Once the $a_j^\nu(t)$ have been calculated, the memory kernel is given by (20).

Finally, we perform the projection (7) of the Mori–Zwanzig equation. The finite-rank projection of the memory term

$$\int_0^t P\,\mathrm{e}^{(t-s)L} K_j(\hat{x}, s)\,\mathrm{d}s$$

is

$$\int_0^t \sum_{\nu,\mu\in I} a_j^\nu(s)\gamma^{\nu\mu}(t-s) h^\mu(\hat{x})\,\mathrm{d}s,$$

where

$$\gamma^{\nu\mu}(t) = (\mathrm{e}^{tL} h^\nu, h^\mu).$$

This implementation of P relies on an expansion of the several $h^\nu(\varphi(x,t))$ in scaled Hermite series and cannot be expected to be accurate with a moderate number of terms.

To formulate the algorithms compactly we introduce the matrices A, F, G, Γ: $A_{j\mu} = a_j^\mu$, $F_{j\mu} = f_j^\mu$, $G_{\nu\mu} = g^{\nu\mu}$, and $\Gamma_{\nu\mu} = \gamma^{\nu\mu}$, for $1 \le j \le m$ and $\nu, \mu \in I$. First let F, G be given and solve

$$A(t) = F(t) - \int_0^t A(s)G(t-s)\,\mathrm{d}s.$$

The function $P\hat{\varphi}(x,t)$ is approximated by $\Phi(\hat{x}, t)$ obtained by solving either

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t) = \mathrm{R}(\Phi(t)) + \int_0^t A(s)h(\Phi(t-s))\,\mathrm{d}s, \qquad \Phi(0) = \hat{x}, \tag{22}$$

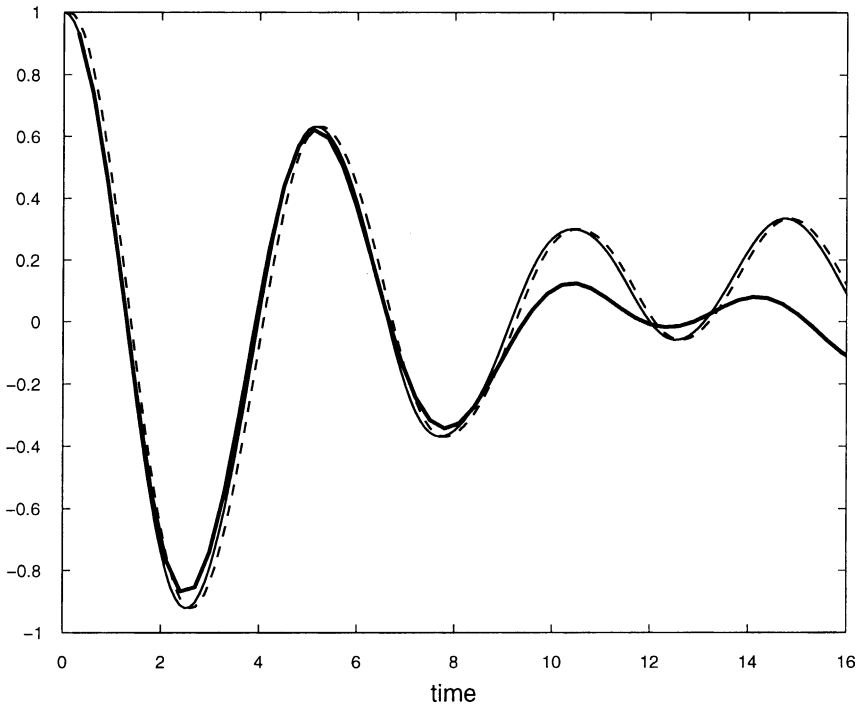

Fig. 6. Comparison of the mean solution $E[\varphi_1(x,t)|\hat{x}]$ (thick line) and the components $\Phi_1(t)$ obtained by solving Eq. (22) (thin line) and Eq. (23) (dashed line). The initial data are $\hat{x} = (1, 0)$.

or

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t) = \mathrm{R}(\Phi(t)) + \int_0^t A(s)\Gamma(t-s)h(\hat{x})\,\mathrm{d}s, \qquad \Phi(0) = \hat{x}. \tag{23}$$

Some results are shown in Fig. 6. The thick line represents the exact solution $E[\varphi_1(x,t)|\hat{x}]$, the thinner line is the function $\Phi_1(t)$ resulting from the integration of (22), whereas the dashed line results from the integration of (23). The quadratures in the Volterra equation were performed by the trapezoidal rule, 21 basis functions h^ν were used, the parameter α in the Hermite expansion was $\alpha = 7/6$, the ordinary differential equations were solved by a Runge–Kutta method, modified to take into account the integral term. The Monte Carlo summations used 10^4 sample solutions.

To improve the accuracy of these calculations one can: (1) increase the number of Monte Carlo samples, (2) decrease the time step, (3) allow the coefficient alpha to change with time, (4) truncate the integral kernel at an appropriate τ.

7. Short-memory approximations

In some situations of interest one can expect the support in time of the integrand in Eq. (7) to be small, and this can simplify the calculations. To analyze this situation, start from the Dyson formula (5) which can be the starting point of a perturbative evaluation of e^{tQL}. The zeroth-order approximation of this relation is

$$\mathrm{e}^{tQL} \cong \mathrm{e}^{tL}, \tag{24}$$

in which one replaces the flow in the orthogonal complement of the range of P by the real flow induced by L. Now consider the second term in Eq. (6):

$$\int_0^t \mathrm{e}^{(t-s)L} PL\,\mathrm{e}^{sQL} QLx_j\,\mathrm{d}s = \int_0^t \mathrm{e}^{(t-s)L} PLQ\,\mathrm{e}^{sQL} QLx_j\,\mathrm{d}s,$$

where the insertion of the extra Q is legitimate since e^{sQL} maps functions in the null space of P back into the same subspace. Adding and subtracting equal quantities, we find:

$$PL\,\mathrm{e}^{sQL} QLx_j = PLQ\,\mathrm{e}^{sL} QLx_j + PLQ(\mathrm{e}^{sQL} - \mathrm{e}^{sL})QLx_j,$$

a Taylor series yields:

$$\mathrm{e}^{sQL} - \mathrm{e}^{sL} = I + sQL + \cdots - I - sL - \cdots = -sPL + \mathrm{O}(s^2),$$

and therefore, using $QP = 0$, we find

$$\int_0^t \mathrm{e}^{(t-s)L} PL\,\mathrm{e}^{sQL} QLx_j\,\mathrm{d}s = \int_0^t \mathrm{e}^{(t-s)L} PLQ\,\mathrm{e}^{sL} QLx_j\,\mathrm{d}s + \mathrm{O}(t^3).$$

If P is the finite-rank projection then

$$PL\,\mathrm{e}^{sQL} QLx_j = \sum_{\nu\in I} (QL\,\mathrm{e}^{sQL} QLx_j, h^\nu) h^\nu(\hat{x}).$$

If the correlations $(L\,\mathrm{e}^{sQL} QLx_j, h^\nu)$ are significant only over short times, the approximation (24) provides an acceptable approximation without requiring the solution of the orthogonal dynamics equation. In statistical physics, one often makes an even more drastic approximation, in which it is assumed that the correlations vanish for $t \neq 0$ (see, e.g. the "high frequency approximation" in [12, p. 86]). Some applications of the short-memory approximation have been presented in [1].

8. The t-damping equation

A short-memory approximation of particular interest can be derived as follows: Write the memory term of the Langevin equation

$$\int_0^t K_j(\hat{\varphi}(x, t-s), s)\,\mathrm{d}s. \tag{25}$$

Expand the integrand in a Taylor series about $s = 0$, retaining only the leading term. The memory term reduces to

$$\int_0^t K_j(\hat{\varphi}(x,t), 0)\,\mathrm{d}s = tS_j(\hat{\varphi}(x,t)), \tag{26}$$

where

$$S_j(x) = PLQLx_j.$$

An alternative derivation uses the formalism of the previous section: write (25) as

$$\int_0^t \mathrm{e}^{(t-s)L} PL\,\mathrm{e}^{sQL} QR_j(x) = \int_0^t L\,\mathrm{e}^{(t-s)L}\,\mathrm{e}^{sQL} QR_j(x) - \int_0^t \mathrm{e}^{(t-s)L}\,\mathrm{e}^{sQL} QLQR_j(x),$$

where we have used the commutation of L and QL with e^{tL} and e^{tQL}, respectively. At this point, make the approximation (24) and replace the evolution operator of the orthogonal dynamics by the evolution operator of the Liouvillian flow, which eliminates the s dependence of both integrands, and (26) follows readily.

Writing $\Phi(t) \approx P\hat{\varphi}(x,t)$, the t-damping equations are

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t) = \mathrm{R}(\Phi(t)) + tS(\Phi(t)), \quad \Phi(0) = \hat{x}. \tag{27}$$

The form of this equation is surprising. All that remains of the memory is the coefficient t and one is left with a non-autonomous system of ordinary differential equations. No previously computed averages are needed.

For the model problem (8) Eq. (27) takes the explicit form

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_1 = \Phi_2, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\Phi_2 = -\Phi_1\left(1 + \frac{1}{1+\Phi_1^2}\right) - 2t\frac{\Phi_1^2\Phi_2}{(1+\Phi_1^2)^2}. \tag{28}$$

In Fig. 7, we present a comparison of the component $\Phi_1(t)$ of the solution of Eq. (28) with $E[\varphi_1(x,t)|\hat{x}]$. The new term, $tS(\Phi(t))$, leads to a damping of the solution.

We now show that the last term in Eq. (28), which approximates the memory term, leads to a decay just like the original term did. Set $q = (\hat{q}, \tilde{q})$, $p = (\hat{p}, \tilde{p})$, and write L, P, H as

$$L = \sum_{j=1}^{n}(H_{p_j}\partial_{q_j} - H_{q_j}\partial_{p_j}), \tag{29}$$

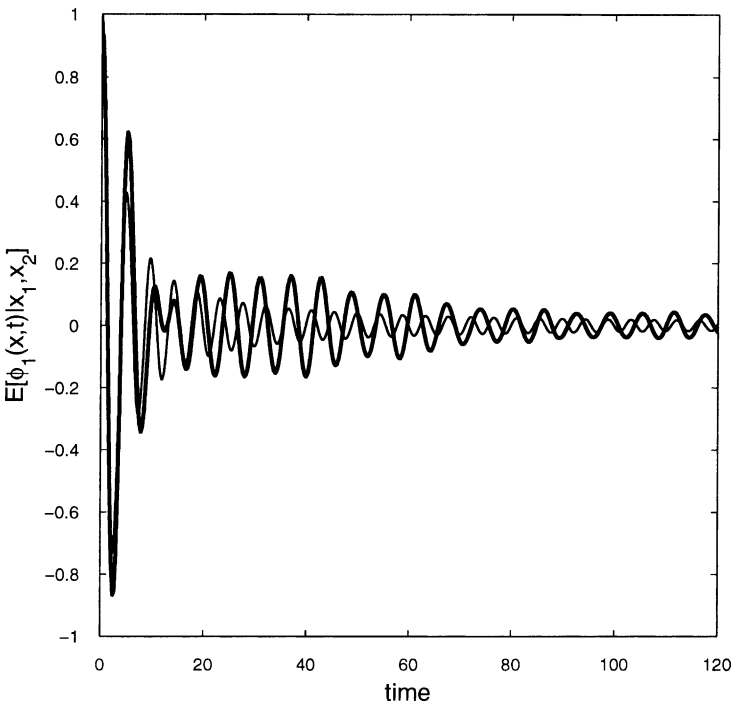

Fig. 7. Comparison of the mean solution $E[\varphi_1(x,t)|\hat{x}]$ (thick line) and the component $\Phi_1(t)$ of the solution of the t-damping equation (28) (thin line) for the initial data $\hat{x} = (1,0)$.

$$(Pf)(\hat{q},\hat{p}) = \frac{\int f(q,p)\,\mathrm{e}^{-H(q,p)}\,\mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p}}{\int \mathrm{e}^{-H(q,p)}\,\mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p}}, \tag{30}$$

and

$$\mathrm{H}(\hat{q},\hat{p}) = -\log\int \mathrm{e}^{-H(q,p)}\,\mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p}. \tag{31}$$

The subscripts in (29) represent differentiation.

Theorem 8.1. *Suppose $H(q,p) = T(p) + V(q)$. Let*

$$\dot{p}_i = (PL + tPLQL)p_i, \qquad \dot{q}_i = (PL + tPLQL)q_i \tag{32}$$

for $i = 1,2,\ldots,m$ with $(\hat{q}(0),\hat{p}(0))$ given. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{H}(\hat{q},\hat{p}) = -tP\left[\left|\sum_{j=1}^{m}(H_{p_j} - \mathrm{H}_{p_j})\mathrm{H}_{q_j}\right|^2 + \left|\sum_{j=1}^{m}(H_{q_j} - \mathrm{H}_{q_j})\mathrm{H}_{p_j}\right|^2\right].$$

Thus, if the Hamiltonian $H(q,p)$ is separable into a kinetic and a potential energy, then the effective Hamiltonian H for the solution of the t-damping equation (32) is a decreasing function of time. Similar results can be obtained for Fourier methods applied to Euler's equations in two- and three-dimensional flows.

Proof. We shall write Eq. (32) in terms of H and H. It follows from (29)–(31) that

$$Lp_i = -H_{q_i}, \qquad PLp_i = -PH_{q_i} = -\mathrm{H}_{q_i}, \qquad QLp_i = (I-P)Lp_i = -H_{q_i} + \mathrm{H}_{q_i}$$

for $i = 1,2,\ldots,m$. Since H_{q_i} and H_{q_i} do not depend on p_i (due to the assumed separability of H) Eqs. (29) and (30) imply

$$PLQLp_i = -\sum_{j=1}^{n} P[H_{p_j}(H_{q_i} - \mathrm{H}_{q_i})_{q_j}]. \tag{33}$$

To proceed we write $H(q,p) = T(p) + V(q)$ and get

$$P[H_{p_j}(H_{q_i} - H_{q_i})_{q_j}] = \frac{\int H_{p_j}\,\mathrm{e}^{-T(p)}\,\mathrm{d}\tilde{p}}{\int \mathrm{e}^{-T(p)}\,\mathrm{d}\tilde{p}} \cdot \frac{\int (H_{q_i} - \mathrm{H}_{q_i})_{q_j}\,\mathrm{e}^{-V(q)}\,\mathrm{d}\tilde{q}}{\int \mathrm{e}^{-V(q)}\,\mathrm{d}\tilde{q}} = PH_{p_j}\cdot P(H_{q_i} - \mathrm{H}_{q_i})_{q_j}. \tag{34}$$

For $j > m$ integration by parts gives

$$PH_{p_j} = -\frac{\int (\mathrm{e}^{-H})_{p_j}\,\mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p}}{\int \mathrm{e}^{-H}\,\mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p}} = 0. \tag{35}$$

Using (30) we see that

$$\partial_{q_j} P(H_{q_i} - \mathrm{H}_{q_i}) = P(H_{q_i} - \mathrm{H}_{q_i})_{q_j} + P[(H_{q_i} - \mathrm{H}_{q_i})(-H_{q_j})] + P(H_{q_i} - \mathrm{H}_{q_i})\cdot(-1)\cdot P(-H_{q_j}).$$

But H_{q_i} does not depend on $\tilde{q}$, $\tilde{p}$ and $P(H_{q_i} - \mathrm{H}_{q_i}) = 0$. Thus,

$$P(H_{q_i} - \mathrm{H}_{q_i})_{q_j} = P[(H_{q_i} - \mathrm{H}_{q_i})(H_{q_j} - \mathrm{H}_{q_j})]. \tag{36}$$

Combining (33)–(36) yields

$$PLQLp_i = -P\left[(H_{q_i} - \mathrm{H}_{q_i})\sum_{j=1}^{m}(H_{q_j} - \mathrm{H}_{q_j})\mathrm{H}_{p_j}\right].$$

The corresponding expression for $PLQLq_i$ is

$$PLQLq_i = -P\left[(H_{p_i} - \mathrm{H}_{p_i})\sum_{j=1}^{m}(H_{p_j} - \mathrm{H}_{p_j})\mathrm{H}_{q_j}\right].$$

We can now rewrite the system (32) as

$$\dot{p}_i = -\mathrm{H}_{q_i} + tPLQLp_i, \qquad \dot{q}_i = \mathrm{H}_{p_i} + tPLQLq_i.$$

Using the chain rule we find

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{H}(\hat{q},\hat{p}) &= \sum_{i=1}^{m}(\mathrm{H}_{q_i}\dot{q}_i + \mathrm{H}_{p_i}\dot{p}_i) = t\sum_{i=1}^{m}(\mathrm{H}_{q_i}PLQLq_i + \mathrm{H}_{p_i}PLQLp_i)\\ &= -tP\left[\sum_{i=1}^{m}(H_{p_i} - \mathrm{H}_{p_i})\mathrm{H}_{q_i}\sum_{j=1}^{m}(H_{p_j} - \mathrm{H}_{p_j})\mathrm{H}_{q_j}\right]\\ &\quad -tP\left[\sum_{i=1}^{m}(H_{q_i} - \mathrm{H}_{q_i})\mathrm{H}_{p_i}\sum_{j=1}^{m}(H_{q_j} - \mathrm{H}_{q_j})\mathrm{H}_{p_j}\right].\end{aligned}$$

This completes the proof. □

9. Conclusions

We have presented a general method for finding the best approximation of part of the solution of a partially resolved initial value problem, conditioned by partial initial data. We presented a formula, Eq. (6), which describes exactly the evolution in time of a few components of an underresolved problem. We have also introduced a collection of methods for approximating the general formula, in general but not always requiring prior information obtainable by Monte Carlo or by experiment but without reference to specific initial data. Some of the methods are expensive in practice but provide theoretical insight; some require additional information, for example about correlation times, which may be available; some are limited in accuracy by the properties of expansions in orthogonal functions. Many variants are possible. Theoretical issues remain to be discussed in greater depth. The existence of orthogonal dynamics is established in [22].

We applied the methods to a model problem which is easy to understand and analyze but is not typical of the problems one encounters in practice. We have not discussed properties of real problems such as separation of scales or convergence to a partial differential equation which help the application of our methods. We have not explored the effect of the choice of initial distribution, for example the possible advantages of a microcanonical distribution. Once the general framework has been established, further variants and assumptions are best discussed within the context of specific applications.

Acknowledgements

We would like to thank Prof. G.I. Barenblatt, Dr. E. Chorin, Mr. E. Ingerman, Dr. A. Kast, Mr. K. Lin, Mr. P. Okunev, Mr. B. Seibold, Mr. P. Stinis, Prof. J. Strain, and Prof. B. Turkington for helpful discussions and comments.

This work was supported in part by the Applied Mathematical Sciences subprogram of the Office of Energy Research of the US Department of Energy under Contract DE-AC03-76-SF00098, and in part by the National Science Foundation under Grant DMS98-14631. RK was supported by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities and by the Alon Fellowship.

References

[1] J. Bell, A. Chorin, W. Crutchfield, Stochastic optimal prediction with application to averaged Euler equations, in: C. Lin (Ed.), Proceedings of the Seventh National Conference on Comput. Fluid Mech., Pingtung, Taiwan, 2000, pp. 1–13.
[2] A. Chorin, Probability, Mechanics, and Irreversibility, Lecture Notes, Mathematics Department, University of California, Berkeley, CA, 2000.
[3] A. Chorin, A. Kast, R. Kupferman, Optimal prediction of underresolved dynamics, Proc. Natl. Acad. Sci. USA 95 (1998) 4094–4098.
[4] A. Chorin, A. Kast, R. Kupferman, On the prediction of large-scale dynamics using unresolved computations, AMS Contemporary Math. 238 (1999) 53–75.
[5] A. Chorin, A. Kast, R. Kupferman, Unresolved computation and optimal prediction, Comm. Pure Appl. Math. 52 (1999) 1231–1254.
[6] A. Chorin, O. Hald, R. Kupferman, Optimal prediction and the Mori–Zwanzig representation of irreversible processes, Proc. Natl. Acad. Sci. USA 97 (2000) 2968–2973.
[7] A. Chorin, R. Kupferman, D. Levy, Optimal prediction for Hamiltonian partial differential equations, J. Comp. Phys. 162 (2000) 267–297.
[8] N. Goldenfeld, A. Mckane, Q. Hou, Block spins for partial differential equations, J. Statist. Phys. 93 (1998) 699–714.
[9] O. Hald, Optimal prediction of the Klein–Gordon equation, Proc. Natl. Acad. Sci. USA 96 (1999) 4774–4779.
[10] O. Hald, R. Kupferman, Convergence of optimal prediction for nonlinear Hamiltonian systems, SIAM J. Numer. Anal. 39 (2001) 983–1000.
[11] A. Kast, Optimal prediction of stiff oscillatory mechanics, Proc. Natl. Acad. Sci. USA 97 (2000) 6253–6257.
[12] D. Evans, G. Morriss, Statistical Mechanics of Nonequilibrium Liquids, Academic Press, London, 1990.
[13] E. Fick, G. Sauerman, The Quantum Statistics of Dynamical Processes, Springer, Berlin, 1990.
[14] H. Mori, Transport, collective motion, and Brownian motion, Prog. Theoret. Phys. 33 (1965) 423–450.
[15] R. Zwanzig, Nonlinear generalized Langevin equations, J. Statist. Phys. 9 (1973) 215–220.
[16] R. Zwanzig, Problems in nonlinear transport theory, in: L. Garrido (Ed.), Systems far from equilibrium, Springer, New York, 1980, pp. 198–225.
[17] A. Chorin, O. Hald, R. Kupferman, Non-Markovian optimal prediction, Monte Carlo Meth. Appl. 7 (2001) 99–109.
[18] E. Hopf, Statistical hydromechanics and functional calculus, J. Rat. Mech. Anal. 1 (1952) 87–114.
[19] A. Pazy, Semigroups of Linear Operators and Applications to Partial Differential Equations, Springer, New York, 1983.
[20] K. Chung, A Course in Probability Theory, Academic Press, New York, 1974.
[21] D. Forster, Hydrodynamic Fluctuations, Broken Symmetry and Correlation Functions, Benjamin, Reading, MA, 1975.
[22] D. Givon, O. Hald, R. Kupferman, Existence of orthogonal dynamics, Preprint, 2001.
[23] A. Chorin, Hermite expansions in Monte-Carlo computations, J. Comp. Phys. 8 (1971) 472–482.
[24] D. Gottlieb, S. Orszag, Numerical Analysis of Spectral Methods, SIAM, Philadelphia, PA, 1977.