Prediction from Partial Data, Renormalization, and Averaging

Alexandre J. Chorin,[1] **Ole H. Hald,**[1] **and Raz Kupferman**[2]

We summarize and compare our recent methods for reducing the complexity of computational problems, in particular dimensional reduction methods based on the Mori–Zwanzig formalism of statistical physics, block Monte-Carlo methods, and an averaging method for deriving an effective equation for a nonlinear wave propagation problem. We show that their common thread is scale change and renormalization.

KEY WORDS: Renormalization; problem reduction; averaging; scaling.

1. INTRODUCTION

We begin by summarizing the "optimal prediction" schemes [6–11] for the statistical estimation of the solutions of problems with too many variables by the solutions of problems with fewer variables, making use of prior statistical knowledge. These schemes have a natural connection with methods of irreversible statistical mechanics, as we explain.

How to reduce the number of variables in a problem while keeping salient statistical features unharmed is also considered in various renormalization group methods, and we show the identity between renormalization and prediction in the simple example of a thermodynamic system at equilibrium; we use this connection to derive a block Monte–Carlo algorithm as in [6].

The renormalization analysis here brings up an important observation. It is often implicitly assumed that to simplify a complex problem in order to make computation feasible means to reduce the number of variables. This is too narrow a view. Instead, one should look for the most

_____

[1] Department of Mathematics, UC, Berkeley, CA 94720, USA. E-mail: chorin@Math.Berkeley.EDU

[2] Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel.

245

amenable problem in a class of equivalent problems, where "equivalence" and "amenability" must be carefully defined in each case. In Sec. 6, a more general scheme for generating equivalent problems through averaging and scaling is used to find an effective equation for local means of the Korteveg-deVries–Burgers equation.

In the present paper the various methods are presented in a unified way designed to bring out their common features. In addition, the Mori–Zwangig formalism, the key to the analysis of time-dependent reduction methods, is presented as a special case of the moment-expansion methods common to all nonlinear statistical methods. Although the specific examples have appeared before, we hope that this presentation will contribute to answering the question: What do complexity reduction schemes have in common, and how does one design such a scheme for a particular given problem?.

2. PREDICTION: PRELIMINARIES

Consider a system of (in general nonlinear) ordinary differential equations,

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

where φ and x are n-dimensional vectors with components φ_i and x_i, and R is a vector-valued function with components R_i; t is time. To each initial value x in (1) corresponds a trajectory $\varphi(t) = \varphi(x, t)$. Such problems arise in particular after spectral expansion of solutions of initial value problems for partial differential equations.

Suppose that we are working with a computer on which only m of the n components of the solution vector $\varphi(t)$ can be effectively computed; the problem has to be simplified. Assume for the moment (this assumption will be modified as we go along) that a reasonable way to simplify the problem is to pick a certain number m of coordinates of $\varphi(t)$ and try to compute them without computing all the coordinates. The set of values of the uncomputed variables is to be represented by a probability density. Specifically, assume that at time $t = 0$ the probability density of all the components of the data, $F(x_1, x_2, \ldots, x_n)$, is known, from previous measurements or from general considerations of statistical mechanics:

$$F(y_1, \ldots, y_n)dy = Prob(y_1 < x_1 \leqslant y_1 + dy_1, \ldots), \tag{2}$$

where $dy = dy_1 \cdots dy_n$ and $Prob$ denotes "probability"; the question is how to use this prior knowledge in the evaluation of $\varphi(t)$.

Consider first the special case of a Hamiltonian system as in [10, 11]. There exists then a Hamiltonian function $H = H(\varphi)$ such that for i odd R_i, the ith component of the vector R in (1) satisfies $R_i = \partial H / \partial \varphi_{i+1}$ while for i even one has $R_i = -\partial H / \partial \varphi_{i-1}$, with n, the size of the system, even. Assume furthermore that F, the initial probability density, is $F(\varphi) = Z^{-1} \exp(-H)$ (in physics this density appears naturally and is known as the canonical density; the normalizing constant Z is the "partition function"). The measure defined by F is invariant, i.e. sampling the measure and evolving the system in time commute.

Partition the vector x so that $\hat{x} = (x_1, x_2, \ldots, x_m)$ (m must be even), $\tilde{x} = (x_{m+1}, \ldots, x_n)$ and $x = (\hat{x}, \tilde{x})$, and similarly $\varphi = (\hat{\varphi}, \tilde{\varphi})$, $R = (\hat{R}, \tilde{R})$. In general the first m components of R depend on all the components of φ, $\hat{R} = \hat{R}(\varphi) = \hat{R}(\hat{\varphi}, \tilde{\varphi})$, or else the first m equations of (1) are independent of the rest and the system is already reduced. Suppose we want to calculate only the variables $\hat{\varphi}$; we find that $(d/dt)\hat{\varphi}(t) = \hat{R}(\varphi(t))$ where the right-hand side depends on the variables $\tilde{\varphi}$ which are unknown. A numerical analyst who wants to approximate a solution usually starts by approximating the equation; a first way to do this is to replace the system (1) by:

$$\frac{d}{dt}\hat{\varphi}(t) = E[R(\varphi(t))|\hat{\varphi}(t)],$$
$$\hat{\varphi}(0) = \hat{x}, \tag{3}$$

where $E[R(\varphi)|\hat{\varphi}] = \int R e^{-H} d\tilde{\varphi} / \int e^{-H} d\tilde{\varphi}$ is the conditional expectation of R given $\hat{\varphi}(t)$. We remind the reader that the conditional expectation of R given $\hat{\varphi}$, the average of R over all values of $\tilde{\varphi}$, is the orthogonal projection of R onto the subspace $\hat{\Gamma}$ of functions of only $\hat{\varphi}$ in the Hilbert space Γ defined by the inner produce (u, v)=expected value of $uv = \int u(x)v(x)F(x)dx$. It is also the best approximation of R in Γ by a function of only $\hat{\varphi}$. The conditional expectation is well defined because the joint probability density of all the components of φ is known at each t (being the invariant density F).

We have shown in [9] that:

(i) The new system (3) is also Hamiltonian:

$$E\left[\frac{\partial H}{\partial \varphi_i}|\hat{\varphi}(t)\right] = \int \frac{\partial H}{\partial \varphi_i} \exp(-H)d\tilde{\varphi} / \int \exp(-H)d\tilde{\varphi} = \frac{\partial \hat{H}}{\partial \varphi_i}, \tag{4}$$

where $i \leqslant m =$ the dimension of $\hat{\varphi}$, and $\hat{H} = -\log \int \exp(-H)d\tilde{\varphi}$ is the new Hamiltonian.

(ii) The new canonical density $Z^{-1} \exp(-\hat{H})$ is invariant in the evolution of the new, reduced, system.

(iii)   When the data are sampled from the canonical distribution, the distribution of $\hat{\varphi}$ in the new system is its marginal distribution in the old system; equivalently, the partition function Z is the same for the old system and for the new system.

The approximation (3) is already useful (see, e.g., [11, 18, 24]). In physics a system in which the values of all the variables are drawn from a canonical distribution is a system in thermal equilibrium; then in Eq. (3) one has assumed that all the variables $\hat{\varphi}$ one is not resolving are in thermal equilibrium (which may be an acceptable assumption if the variables in $\tilde{\varphi}$ are, for example, varying very rapidly so that they approach equilibrium quickly). The assignment of definite values $\hat{x}$ to the variables $\hat{\varphi}$ at time $t = 0$ is tantamount to taking the system out of equilibrium at $t = 0$; if the system is ergodic it will then decay to equilibrium in time, so that all the variables become randomized and acquire the joint density F. This interpretation shows why the approximation cannot be accurate for long times: Suppose we are given partial data at $t = 0$; the predictive value of these data will decrease in time; all averages of the $\hat{\varphi}$ will approach equilibrium averages. However, the reduced system (3) being Hamiltonian, the averaged solutions it produces will oscillate forever.

We now turn to the derivation of more accurate prediction formulas, following [9].

3. THE MORI–ZWANZIG FORMALISM

Consider again the system (1). Let $L = \sum_{i=1}^{n} R_i(x)\partial_i$, $(\partial_i = \partial/\partial x_i)$, and consider the initial value problem for the linear partial differential equation

$$\frac{\partial}{\partial t} u(x, t) = Lu(x, t),$$
$$u(x, 0) = g(x) \tag{5}$$

for the initial 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). This linear partial differential equation is thus equivalent to the nonlinear system (1) (and similarly, one can derive a linear functional differential equivalent to a nonlinear partial differential equation, see [20]).

We use the semigroup notation $u(x, t) = (e^{tL}g)(x) = g(\varphi(x, t))$, where e^{tL} is the evolution operator associated with the operator L; One can verify that $e^{tL}L = Le^{tL}$. Equation (5) becomes

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

Suppose that the initial conditions x are sampled from a probability density F (note the we are no longer assuming that F is invariant nor that the system is Hamiltonian). Given the initial values of the m coordinates $\hat{x}$, the distribution of the remaining $n-m$ coordinates $\tilde{x}$ is the conditional density, F conditioned by $\hat{x}$. Given F, the functions $\varphi(t)$ can be viewed as stochastic processes on the space Γ with inner product given by the expectation of the product $(f, g) = E[fg]$.

We now transform the equations of motion into a form useful for the prediction problem. This transformation uses the linearity of the Liouville equation (5); this is why we introduced this equation. We define the projection operator P by $Pg = E[g|\hat{x}]$ (see above; note that the conditioning variables are now the initial values of $\hat{\varphi}$; we cannot find $E[g|\hat{\varphi}]$ without knowing the probability density of the φ). The construction is essentially the Mori–Zwanzig formalism of statistical physics [12, 17, 28, 32].

Consider a "resolved" coordinate $\varphi_j(x, t) = e^{tL} x_j$ ($j \leqslant m$), and split its time derivative, $R_j(\varphi(x, t)) = e^{tL} L x_j$, into a term that depends only on $\hat{\varphi}(x, t)$ plus a remainder:

$$\frac{\partial}{\partial t} e^{tL} x_j = e^{tL} L x_j = e^{tL} P L x_j + e^{tL} Q L x_j, \tag{6}$$

where $Q = I - P$. Define $\mathscr{R}_j(\hat{x}) = (P R_j)(\hat{x})$; the first term is $e^{tL} P L x_j = \mathscr{R}(\hat{\varphi}(x, t))$ and is a function of the resolved components only. Note that if Q were zero we would recover something that looks like the crude approximation of the previous section; however the conditioning variables are now the initial values and not the instantaneous values of the $\hat{\varphi}(t)$; in general we do not know without further work what the probability density of the $\hat{\varphi}(t)$ may be. Another key difference between the system (3) and the system obtained by setting $Q = 0$ in Eq. (6) is that the former is closed-its solutions are well-defined functions $\hat{\varphi} = \hat{\varphi}(\hat{x}, t)$ and live in the range of the operator P defined in the previous section, while here the situation is more complex.

We further split the remaining term $e^{tL} Q L x_j$. This splitting will bring it into a very useful form: a noise term, and a memory term whose kernel depends on the correlations of the noise term. The fact that such a splitting is possible is the essence of "fluctuation-dissipation" theorems (see, e.g., [23]). Let $w(x, t) = e^{tQL} Q L x_j$, i.e., $w(x, t)$ be a solution of the initial value problem:

$$\frac{\partial}{\partial t} w(x, t) = QL w(x, t) = L w(x, t) - P L w(x, t),$$
$$w(x, 0) = Q L x_j. \tag{7}$$

An existence proof for Eq. (7) may be found in [13]. If $Pf = 0$, then $Pe^{tQL} f = 0$ 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 Duhamel (also sometimes known as Dyson) formula [12]:

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

Hence,

$$e^{tL} QL x_j = e^{tQL} QL x_j + \int_0^t e^{(t-s)L} PL e^{sQL} QL x_j \, ds. \qquad (8)$$

Let:

$$J_j(x, t) = e^{tQL} QL x_j,$$
$$K_j(\hat{x}, t) = PL J_j(x, t).$$

Collecting terms, we obtain the generalized Langevin equation:

$$\frac{\partial}{\partial t} e^{tL} x_j = e^{tL} \mathscr{R}_j(\hat{x}, t) + \int_0^t e^{(t-s)L} K_j(\hat{x}, s) \, ds + J_j(x, t)$$

or equivalently,

$$\frac{\partial}{\partial t} \varphi_j(x, t) = \mathscr{R}_j(\hat{\varphi}(x, t)) + \int_0^t K_j(\hat{\varphi}(x, t-s), s) \, ds + J_j(x, t). \qquad (9)$$

The various terms in Eq. (9) have physical 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.

If what is desired is $P\hat{\varphi}(t)$, the conditional expectation of $\hat{\varphi}(t)$ given $\hat{x}$ (the best approximation in the sense of Γ to $\hat{\varphi}$ given the partial data $\hat{x}$), then one can multiply Eq. (9) by P to obtain:

$$\frac{\partial}{\partial t} P\varphi_j(x, t) = P\mathscr{R}_j(\hat{\varphi}(x, t)) + \int_0^t P K_j(\hat{\varphi}(x, t-s), s) \, ds. \qquad (10)$$

On the left-hand side of Eq. (10) is exactly what we want. How to evaluate the first term on the right-hand-side requires further thought in each particular problem.

Equations (9) and (10) 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; the only advantage of these equations is that they are a good starting point for approximations.

Note that even if the system we start with is Hamiltonian, the Langevin Eq. (9) is not; the memory and the "noise" allow the system to forget its initial values and decay to "thermal equilibrium".

4. CORRELATIONS AND CLOSURES, APPROXIMATIONS

We now show that what the memory term hides is higher order correlations of the stochastic processes $\varphi(t)$, specifically, that the memory term is a sum of convolutions of of the form $C * S$, where C is a correlation function of a "noise" (a stochastic process taking values in the null space of P) and S is a variable that describes the state of the system. We shall use this remark to imbed the Mori–Zwanzig formalism in the usual framework of nonlinear physics.

For simplicity, we restrict ourselves to cases where the operator L is skew-symmetric, i.e, $(Lu, v) = -(u, Lv)$, (remember $(u, v) = E[uv]$). The skew-symmetry holds in particular for Hamiltonian systems with F the canonical distribution, see [9, 12]). Pick an orthonormal basis $\{h_k = h_k(\hat{x}), k = 1, \ldots\}$ in the subspace of functions of only $\hat{x}$ (for example, the h_k could be Hermite polynomials). Then, for any function $\psi(x, t)$, one has $\psi = \sum_k (\psi(x, t), h_k)h_k$, and in particular,

$$P(LQe^{sQL}QLx_j) = \sum_k (LQe^{sQL}QLx_j, h_k)h_k(\hat{x}), \tag{11}$$

where a factor Q has been inserted before the exponentials, harmlessly because the operators that follow all live in the null space of P. The memory term now becomes

$$\int_0^t e^{(t-s)L} PLe^{sQL}QLx_j ds = \int_0^t \sum_k e^{(t-s)L}(LQe^{sQL}QLx_j, h_k)h_k(\hat{x})ds$$

$$= \sum_k \int_0^t (LQe^{sQL}QLx_j, h_k)h_k(\hat{\varphi}(t-s))ds. \tag{12}$$

Now $(LQe^{sQL}QLx_j, h_k) = -(e^{sQL}QLx_j, QLh_k)$ by the symmetry of Q and the assumed skew-symmetry of L; the right-hand side of Eq. (12) is the sum of correlations of stochastic processes with values in the null space of P, evaluated at time s, with functions h_k of the state of the

system at time $t - s$, as promised. The identity (12) reduces to more standard forms, for example to the Onsager relation between a response function and a correlation function, if one keeps only a small number of terms in the expansion of P.

These remarks allow one to look at the Mori–Zwanzig formalism in a broader perspective. The goal of the formalism is to calculate some variables [The variables $\hat{\varphi}(t)$] without calculating all the variables, on the basis of prior statistical information, just as in the kinetic theory of gases or in turbulence theory one wishes to find mean properties of a flow without calculating all the details of the flow. Equation (9) is thus analogous to the Reynolds averaged equations in turbulence: it is an identity, but it contains unknown quantities which must be found by other means. These quantities are correlation functions for the full solution. One can find an analogous equation for these correlation functions, but it depends then on higher-order correlations. This sets up a hierarchy of equations, analogous to the moment hierarchy in fluid mechanics or to the BBGKY hierarchy in kinetic theory. The point of the formalism is that the particular form of the Mori–Zwanzig equation makes it easier to guess suitable approximations.

A number of approximation schemes have been proposed for use in various settings (see [9, 16, 26, 27, 30]. The formalism is particularly useful when one can say something in advance about the characteristic time of the correlations in the memory term, i.e., about the length of time for which the correlations are significant. For the sake of completeness we summarize an example of such approximations (see [9]).

Suppose the characteristic time of the correlations is short. To analyze this situation, start from the Duhamel formula

$$e^{tL} = e^{tQL} + \int_0^t e^{(t-s)L} PLe^{sQL} \, ds,$$

The zeroth order approximation of this relation is:

$$e^{tQL} \cong e^{tL}, \tag{13}$$

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. (9):

$$\int_0^t e^{(t-s)L} PLe^{sQL} QLx_j \, ds = \int_0^t e^{(t-s)L} PLQe^{sQL} QLx_j \, ds,$$

where as above an extra Q has been inserted. Adding and subtracting equal quantities, we find:

$$PLe^{sQL}QLx_j = PLQe^{sL}QLx_j + PLQ(e^{sQL} - e^{sL})QLx_j$$

a Taylor series yields:

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

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

$$\int_0^t e^{(t-s)L}PLe^{sQL}QLx_j\,ds = \int_0^t e^{(t-s)L}PLQe^{sL}QLx_j\,ds + O(t^3).$$

If P is approximated by a finite number of terms in an orthonormal expansion, then

$$PLe^{sQL}QLx_j = \sum_k (Le^{sQL}QLx_j, h_k)h_k(\hat{x}).$$

If the correlations $(Le^{sQL}QLx_j, h_k)$ and $(Le^{sL}QLx_j, h_k)$ are significant only over times of order t_0 then one can check that the approximation is of order t_0^3 while it does not require the solution of the orthogonal dynamics equation. One often makes an even more drastic approximation, in which it is assumed that the correlations vanish for $t \neq 0$. Some applications of the particular short-memory approximation above have been presented in [30].

An interesting short-memory approximation can be derived as follows: write the memory term as:

$$\int_0^t e^{(t-s)L}PLs^{sQL}QLx_j\,ds = \int_0^t Le^{(t-s)L}e^{sQL}QLx_j\,ds$$
$$- \int_0^t e^{(t-s)L}e^{sQL}QLQLx_j\,ds, \qquad (14)$$

where the the commutativity of L and e^{tL}, and that of QL with e^{tQL}, have been used. Make the approximation $e^{tQL} \approx e^{tL}$ at this point, and you find that the memory term reduces to $tPLQLx_j$. This result is surprising at first glance: all that is left of the memory is the coefficient t. Partial differential equations with time-dependent coefficients appear elsewhere as well in theories of non-Markovian processes (see [22]).

5. CONNECTION WITH RENORMALIZATION

Consider again, as in Sec. 2, a Hamiltonian system with a canonical initial distribution. Suppose all the elements of the initial vector x are drawn from that canonical distribution, i.e., no initial values for the vector φ are known for sure even at time $t = 0$. By definition the system is now in equilibrium initially, and no memory and noise are needed to bring it to that equilibrium. The resulting problem is still interesting: if the system is at equilibrium the dynamics will sample the equilibrium distribution, and there are many problems in which such sampling is of interest. If we are interested in calculating the average of a function g of only the variables $\hat{\varphi}$, $g = g(\hat{\varphi})$, then one can sample the reduced system $\frac{d}{dt}\hat{\varphi} = E[\hat{R}|\hat{\varphi}]$ of Sec. 3, with the reduced Hamitonian $\hat{H}$, as we know that the probability density of the variables $\hat{\varphi}$ in the reduced system equals its marginal density in the full system.

Imagine now that the variables φ live on the nodes of a regular lattice. Divide the nodes into regular subsets (i.e., they all have the same shape and their union is the full lattice; for example, in the case of a one-dimensional lattice, the subsets could be adjacent pairs of points). In each subset pick one variable, in a fixed location within the shape of the subset, to be a member of $\hat{\varphi}$ and the remainder to be members of $\tilde{\varphi}$ (for more general choices see [29]) . Then write $H^{(0)} = H$ and $H^{(1)} = \hat{H}$, eliminate the members of $\tilde{\varphi}$ and renumber the members of $\hat{\varphi}$ so that they occupy the original lattice. We know (see Sec. 2) that the mapping from the old system to the new system preserves probability and equivalently, the partition function, and constitutes thus by definition a Kadanoff renormalization group transformation [21]. This transformation can be repeated and generates a sequence of Hamiltonians $H^{(0)}, H^{(1)}, H^{(2)}, \ldots$, and thus the conditional expectations that defines the reduced Hamiltonian can be evaluated recursively.

We know from Eq. (4) that for all i such that φ_i is in $\hat{\varphi}$, we have

$$\frac{\partial H^{(1)}}{\partial \varphi_i} = E\left[\frac{\partial H^{(0)}}{\partial \varphi_i}|\hat{\varphi}\right] \tag{15}$$

with similar identities connecting $H^{(n+1)}$ and $H^{(n)}$ for any $n \geqslant 0$. Consider a translation invariant system, in which one can satisfy this identity for all i by satisfying it for one value of i, say $i = 1$. Pick a basis, say $\{\psi_k, k = 1, 2, \ldots\}$ in the space of functions of $\hat{\varphi}$, expand $H^{(n)}$ in a series:

$$H^{(n)} = \sum_k a_k^{(n)} \psi_k(\varphi),$$

write $\psi_k' = \frac{\partial}{\partial \varphi_1} \psi_k$, and differentiate the $H^{(n)}$ with respect to φ_1; you find:

$$\frac{\partial}{\partial \varphi_1} H^{(n+1)} = \sum_k a_k^{(n)} P \psi_k'(\varphi),$$

where P is as before the projection operator defined by conditional expectation. A rearrangement of the terms and a trivial integration yield a series expansion for $H^{(n+1)}$ in terms of the ψ's. Details, a generalization to discrete variables, and examples, can be found in [6], with more coming in [29]; a Monte–Carlo sampling of $e^{-\hat{H}}$ is a block Monte–Carlo sampling (see, e.g., [5, 25, 31]).

This construction was brought up to make a point. Consider a set of variables in the plane, say they live on an $n \times n$ lattice with periodic boundary conditions. If the system is far from a critical point sampling with a small n will yield reasonable values for all quantities of interest. Near a critical point a large array size n may be needed, and a dimensional reduction like the one we just performed may be appropriate.

Suppose we carry out the construction just outlined and reduce n by a factor of, say, 2. We can now get statistical information about the equilibrium properties of the system of equations by working with $n/4$ variables. Note however that the amount of work per sweep of the array has not necessarily gone down. There are fewer variables, but the Hamiltonian has generally become substantially more complex as it was renormalized and more "couplings" (terms in the series expansion for H) have been added. However, the overall work has decreased, because, as is well-known, renormalization has taken us further from the critical point and the correlation time of the sampling has decreased [21].

A moment's reflection will show that what is important in the construction is not the reduction in the number of variables but the fact that the renormalization has picked out a more amenable problem in the class of problems with equal averages for selected variables. This conclusion has broad applicability. We now present an example where a generalization of our prediction procedure leads to more amenable problems without an explicit reduction in the number of variables.

6. SIMILARITY AND AVERAGING

We first briefly remind the reader of the fundamentals of similarity theory [2]. Suppose a variable a is a function of variables $a_1, a_2, \ldots, a_m$, $b_1, b_2, \ldots, b_k$, where in some system of units $a_1, \ldots, a_m$ have independent units while the units of $b_1, \ldots, b_k$, can be formed from the units

of $a_1, a_2, \ldots, a_m$. Then there exist dimensionless variables $\Pi = \frac{a}{a_1^{\alpha_1} \cdots a_m^{\alpha_m}}$, $\Pi_i = \frac{b_i}{a_1^{\alpha_{i1}} \cdots a_m^{\alpha_{im}}}$, $i = 1, \ldots, k$, where the α_i, α_{ik} are integers, such that Π is a function of the Π_i:

$$\Pi = \Phi(\Pi_1, \ldots, \Pi_k). \tag{16}$$

This is just a consequence of the requirement that a physical relationship be independent of the magnitude of the units of measurement. Suppose the variables Π_i are small, and suppose one assumes that the function Φ (about which we know nothing at this stage) has a non-zero finite limit as its arguments tend to zero; then $\Pi \sim$ constant, and one finds a power monomial relation between a and the a_i. The resulting relation is invariant under a group of scaling transformations generated by changes in the units. A similar argument works if the Π_i's are very large. If the function Φ does not have the assumed limit, it may happen that for Π_1 small, $\Phi(\Pi_1) = \Pi_1^q \Phi'(\Pi_1) + \cdots$, where the dots denote lower order terms, q is a constant, the other arguments of Φ have been omitted and Φ' has a finite non-zero limit. One can then obtain a power monomial expression for a in terms of the a_i and b_i, with undetermined powers which must be found by other means. The resulting power relation is an incomplete similarity relation, and the corresponding group of transformations under which the relationship is invariant is Barenblatt's renormalization group. A relation between Barenblatt's RNG and more standard definitions of the RNG is discussed in [14], see also the prescient paper [3].

Consider now the (non-dimensional) Korteveg-deVries–Burgers (KdVB) equation

$$u_t + u u_x = \frac{1}{R} u_{xx} + u_{xxx} \tag{17}$$

with $u(-\infty) = 1$, $u_x(-\infty) = 0$, $u(+\infty) = 0$; R is the "Reynolds number". The problem of deriving an effective equation for the mean solution of this equation was already considered in [1], where relations to Eddy modeling in turbulence were also explored. For long times the solution of this equation with these data becomes a traveling wave; for $R \leqslant 1$ the traveling wave has a monotonic profile, while for $R > 1$ the profile is oscillatory, with oscillations whose wave length is of order 1 [4]. At zero diffusion ($R = \infty$) the stationary asymptotic wave train extends to infinity on the left with a singularity near the location where there would be a shock in the absence of dispersion, and the left boundary conditions cannot be satisfied. For finite R the wave train is damped as is the singularity, and the boundary conditions can be satisfied. The solution of Eq. (3), shifted by $-t/2$, tends to a steady state.

Suppose we are interested only in local averages of u, e.g., in

$$\bar{u}(x) = \frac{1}{\ell} \int_{x-\ell/2}^{x+\ell/2} u(s)ds,$$

where ℓ is some predetermined length. Can we find an effective equation, easier to solve than Eq. (17), whose solution is $\bar{u}$?

We draw now an analogy between the conditional expectations which define the renormalized variables in the equilibrium Hamiltonian case and averaging in space. Averaging over an increasing length scale corresponds either to more renormalization steps or, equivalently, to renormalization with a greater number of variables grouped together. We pick a class of equations in which to seek the "effective" equation, the one whose solutions best approximate the averages of the true solution in the mean square sense; the choice of mean-square approximation in the KdVB case corresponds to the use of L_2 norms implied by the use of conditional expectations in the Hamiltonian case, and the choice of a class of equations in which to look for the effective equation is analogous to the choice of a basis for the representation of the Hamiltonian; the calculation of the best coefficients in the chosen class of "effective" equations corresponds to the evaluation of the coefficients in the series for the renormalized Hamiltonians. Note that in the Hamiltonian case we average the right-hand sides of the equations and in the analogous KdVB case we attempt to average the solutions; this must be so because in the KdVB case we do not have the theorems which guarantee that averaging the right-hand sides produces the correct statistics for the solutions. One expects that the effective equation would have smoother solutions than the original equation and would require fewer mesh points to be properly approximated; in this sense the number of variables is decreased, though we choose not to change the scale of the averaged solution and thus have no analogue of the relabeling of the variables in the Hamiltonian case.

We are looking here for an effective equation in the class of equations of the same form as the original KdVB equation but with a new, unknown coefficient ν replacing the coefficient $1/R$. This coefficient is the effective viscosity at scale ℓ. The equation we will be solving is thus

$$v_t + v v_x = \nu v_{xx} + v_{xxx},$$

where v obeys the same boundary conditions as u, and the problem is to find the coefficient ν which minimizes the integral

$$I = \int_{-\infty}^{+\infty} |\bar{u}(x) - v(x)|^2 dx.$$

The minimizing coefficient ν was determined numerically; the detailed procedure was presented in [7].

From scaling theory we know that

$$\nu = \Phi(R, \ell),$$

where Φ is an unknown dimensionless function, and we are interested in the case where both R and ℓ are large, We know a priori that we cannot have complete similarity, because if we did, ν would be independent of R for large R, an unlikely outcome. The numerical runs in [7] suggest

$$\Phi(R, \ell) = R^{3/4} \Phi(\ell)$$

an incomplete similarity relation where $\Phi(\ell)$ tends to a constant non-zero limit as $\ell \to \infty$. If this relation can be established with confidence, it could be used to calculate averaged solutions when R is so large that the full, unaveraged solution is out of numerical reach because the number of peaks in the full solution is too large. In Fig. 1 we display a full solution, an averaged solution and the solution of of the effective equation
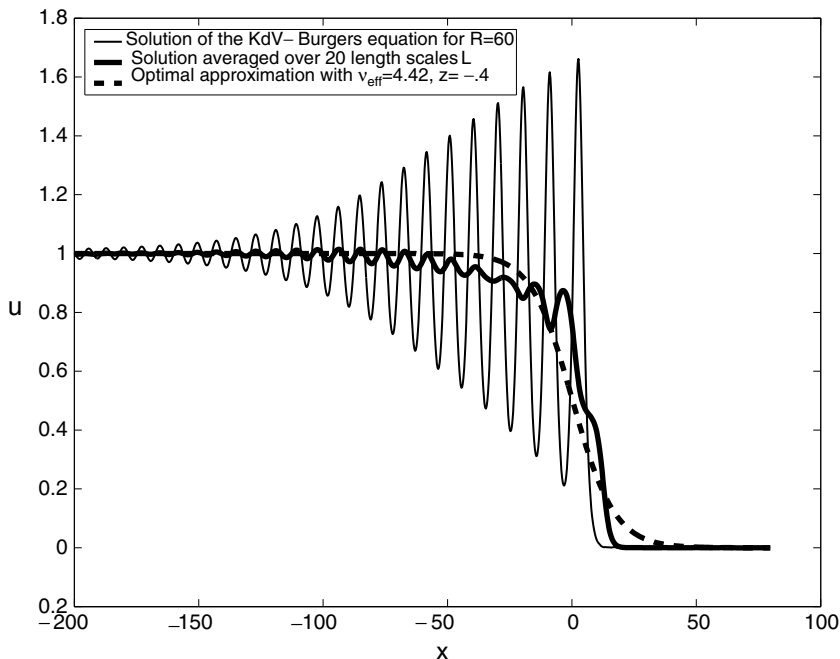

Fig. 1. Full solution, averaged solution and renormalized solution of the KdVB equation.

for a particular set of parameters. Note that the solutions of the effective equation are very smooth and thus easy to compute.

7. CONCLUSIONS

We have demonstrated the specific relationship between prediction methods and renormalization in a statistically stationary case and in a problem whose solution was stationary in a moving frame. When the time dependence is more complex one would want to compare the algorithms generated by the Mori–Zwanzig formalism to dynamic renormalization [15, 19]. However, dynamic renormalization has been derived only for problems where one can assume that the noise that represents the effect of one scale on another can be approximated by white noise- a special and generally unrealistic assumption. A proper dynamic renormalization would amount to a recursive implementation of the Mori–Zwanzig formalism, and it is not known at present how to do this and indeed whether it can be done. In this situation, the Mori–Zwanzig formalism is a good starting point for finding useful amenable problems in time-dependent cases, and current theories of dynamic renormalization must be viewed as mere heuristic approximation of that formalism.

The most interesting aspect of the methods just described is their usefulness in various concrete problems, as will be described in particular in our forthcoming publications.

ACKNOWLEDGMENTS

REFERENCES

1. Barenblatt, G. I., Ivanov, M., and Shapiro, G. I. (1985). On the structure of wave fronts in nonlinear dissipative media. *Arch. Rat. Mech. Anal.* **87**, 293–303.
2. Barenblatt, G. I. (2002). *Scaling*, Cambridge University Press, Cambridge.
3. Benettin, G., di Castro, C., Jona-Lasinio, G., Peliti, L., and Stella, A. (1976). On the equivalence of different renormalization groups. In M. Levy and P. Mitter (eds.), *New developments in quantum theory and statistical mechanics, Cargese Conf. Theor. Physics*, Springer, NewYork.

4. Bona, J. and Schonbek, M. (1985). Travelling-wave solutions to the Korteveg-deVries-Burgers equation. *Proc. R. Soc. Edinburgh* **101A**, 207–226.
5. Burkhardt, T. and van Leeuwen, J. (1982). *Real-Space Renormalization*, Springer, Berlin.
6. Chorin, A. J. (2003). Conditional expectations and renormalization. Multiscale Model. Simulation **1**, 105–118.
7. Chorin, A. J. (2003). Averaging and renormalization for the Korteveg-deVries-Burgers equation. *Proc. Nat. Acad. Sci.* **100**, 9674–9679.
8. Chorin, A. J., Hald, O., and Kupferman, R. (2000). Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *Proc. Nat. Acad. Sc. USA.* **97**, 2968–2973.
9. Chorin, A. J., Hald, O., and Kupferman, R. (2002). Optimal prediction with memory. *Phys. D* **166**, 239–257.
10. Chorin, A. J., Kast, A., and Kupferman, R. (1998). Optimal prediction of underresolved dynamics. *Proc. Nat. Acad. Sc. USA*, **95** 4094–4098.
11. Chorin, A. J., Kupferman, R., and Levy, D. (2000). Optimal prediction for Hamiltonian partial differential equations. *J. Comput. Phys.* **162**, 267–297.
12. Evans, D. and Morriss, G. (1990). *Statistical Mechanics of Nonequilibrium Liquids*. Academic, London.
13. Givon, D., Hald, O. and Kupferman, R. (2004). Existence proof for orthogonal dynamics in the Mori-Zwanzig formalism. *Israel J. Math.* **145**, 221–241.
14. Goldenfeld, N. (1992). *Lectures on Phase Transitions and the Renormalization Group*, Perseus Books, Reading, MA.
15. Goldenfeld, N., McKane, A., and Hou, Q. (1998). Block spins for partial differential equations. *J. Stat. Phys.* **93**, 699–714.
16. Gorban, A. (2001). Ehrenfest's argument extended to a formalism of nonequilibrium thermodynamics. *Phy. Rev. E* **63**, 0661241–0661246.
17. Grabert, H. (1982). *Projection Operator Techniques in Nonequilibrium Statistical Mechanics*, Springer, Berlin.
18. Hald, O., and Kupferman, R. (2001). Convergence of optimal prediction for nonlinear Hamiltonian systems. *SIAM J. Num. Anal.* **39**, 983–1000.
19. Hohenberg, P., and Halperin, B. (1977). Theory of dynamical critical phenomena. *Rev. Mod. Phys.* **49**, 435–479.
20. Hopf, E. (1952). Statistical hydromechanics and functional calculus. *J. Rat. Mech. Anal.* **1**, 87–141.
21. Kadanoff, L. (2000). *Statistical Physics: statics, Dynamics, and Renormalization*. World Scientific, Singapore.
22. Kupferman, R. (2004). Fractional kinetics in Kac-Zwanzig heat bath models. *J. Stat. Phys.* **114**, 291–326.
23. Landau, L., and Lifshitz, E.(1979). *Statistical Physics*. Pergamon, New York.
24. Langford, J., and Moser, R. (1999). Optimal LES formulations for isotropic turbulence. *J. Fluid Mech.* **398** 321–346.
25. Liu, J. (2001). *Monte-Carlo Strategies in Scientific Computing*. Springer-Verlag, New York.
26. Majda, A., Timofeyev, I., and Vanden Eijnden, E. (1999). Models for stochastic climate prediction. *Proc. Natl. Acad. Sci. USA* **96**, 14687–14691.
27. Majda, A., Timofeyev, I., and Vanden Eijnden, E., (2001). A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.* **54**, 891–974.
28. Mori, H. (1965). Transport, collective motion and Brownian motion. *Prog. Th. Phys.* **33**, 423–450.
29. Okunev, P. (2005). *Optimal prediction in physics and finance*, PhD thesis, math. dept.

30. Stinis, P. (2004). Optimal prediction for the Kuramoto-Sivashinski equation, *Multiscale Model. Simul.* **2**, 580–612.

31. Swendsen, R. (1979). Monte-Carlo renormalization group. *Phys. Rev. Lett.* **42**, 859–861.

32. Zwanzig, R. (1961). Memory effects in irreversible thermodynamics. *Phys. Rev.* **124**, 983–992.