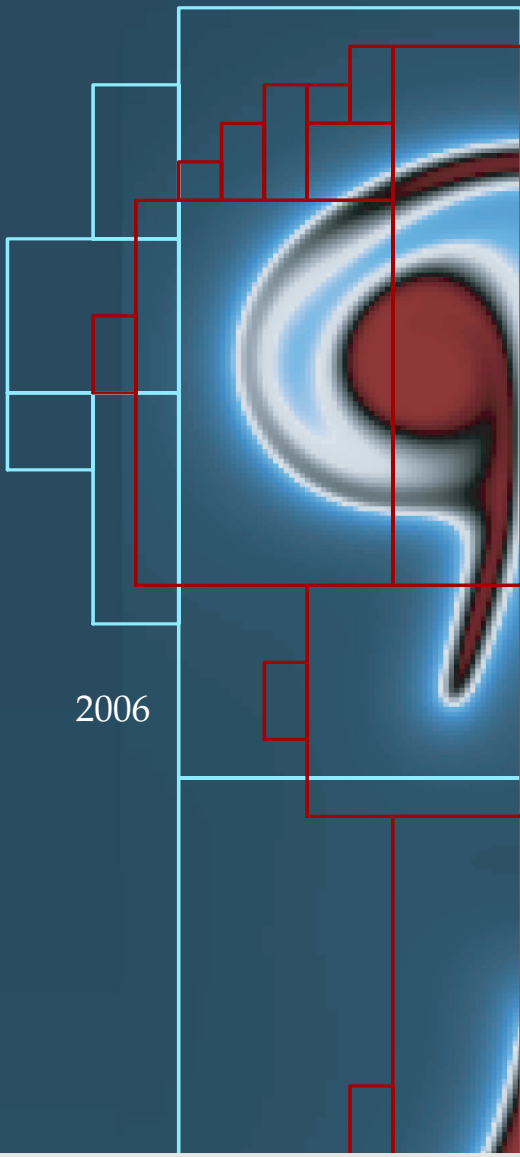
Communications in Applied Mathematics and Computational Science

Volume 1 No. 1 2006



PROBLEM REDUCTION, RENORMALIZATION, AND MEMORY

ALEXANDRE J. CHORIN AND PANAGIOTIS STINIS

mathematical sciences publishers

COMM. APP. MATH. AND COMP. SCI.
Vol. 1, No. 1, 2006

PROBLEM REDUCTION, RENORMALIZATION, AND MEMORY

ALEXANDRE J. CHORIN AND PANAGIOTIS STINIS

We present methods for the reduction of the complexity of computational problems, both time-dependent and stationary, together with connections to renormalization, scaling, and irreversible statistical mechanics. Most of the methods have been presented before; what is new here is the common framework which relates the several constructions to each other and to methods of theoretical physics, as well as the analysis of the approximate reductions for time-dependent problems. The key conclusions are: (i) in time dependent problems, it is not in general legitimate to average equations without taking into account memory effects and noise; (ii) mathematical tools developed in physics for carrying out renormalization group transformations yield effective block Monte Carlo methods; (iii) the Mori–Zwanzig formalism, which in principle yields exact reduction methods but is often hard to use, can be tamed by approximation; and (iv) more generally, problem reduction is a search for hidden similarities.

1. Introduction

There are many problems in science that are too complex for numerical solution as they stand. Examples include turbulence, molecular dynamics, and other problems where multiple scales must be taken into account. Such problems must be reduced to more amenable forms before one computes. In the present paper we would like to summarize some of the reduction methods that have been developed in recent years, together with an account of what was learned in the process. It is obvious that the problem has not been fully solved, but we think that the examples and the conclusions reached so far are useful.

In general terms, a reduction to a more amenable form is a renormalization group transformation, as in physics — a transformation of a problem into a more tractable form while keeping quantities of interest invariant. A renormalization group transformation involves an incomplete similarity transformation, and thus a

MSC2000: 65C20, 65Z05, 82B80, 76F30.

Keywords: problem reduction, renormalization, irreversible statistical mechanics, memory, Monte Carlo.

This work was supported in part by the National Science Foundation under Grant DMS 04-32710, and by the Director, Office of Science, Computational and Technology Research, U.S. Department of Energy under Contract No. DE-AC03-76SF000098.

reduction method is a search for hidden similarities. This is a general feature of reduction methods, and it will be illustrated in the examples. A successful problem reduction produces a new problem which must in some asymptotic sense be similar to the original problem. For general background on renormalization, see, e.g., [5; 20; 39].

In problems with strong time dependence, reduction methods resemble methods for the analysis of thermodynamic systems not in equilibrium; indeed, those aspects of the problem that are ignored in a reduced description conspire to destroy order and increase entropy. Problem reduction for time-dependent problems is basically renormalization group theory for non-equilibrium statistical mechanics. For background on such theory, see, e.g., [3; 22; 8; 44].

The content of the paper is as follows: In section 2 we consider Hamiltonian systems and their conditional expectations. In section 3 we narrow the discussion to statistically stationary Hamiltonian systems and recover Kadanoff real-space renormalization groups and an interesting block Monte Carlo method. In section 4 we display an example that exhibits and also extends the main features of this analysis in simple form.

In section 5 we explain the Mori–Zwanzig formalism for the reduction of statistically time-dependent problems. The analysis shows that averaging the equations is in general not enough; one must take into account noise and a temporal memory. The Mori–Zwanzig formalism is rather dense, and in the sections that follow we present various special cases in which it can be simplified, in particular when the memory is very short or very long. We wish to draw the reader's attention in particular to the "t-model", for which we present a new analysis; it seems to us that it represents a step forward in modeling for a relatively small price in added computational complexity.

One of our goals in exploring the connections between problem reduction and irreversible statistical mechanics is to point out some of the places where the knowledge acquired in statistical mechanics still awaits its proper integration into computational practice.

The paper [21] is a survey of reduction methods organized along different lines and can be profitably read in tandem with the present paper.

For the sake of readability, we remind the reader of the rudiments of similarity theory [3]. Suppose a variable a is a function of variables $a_1, a_2, \ldots, a_m, b_1, b_2, \ldots, b_k$, where $a_1, \ldots, a_m$ have independent units, for example units of length and mass, 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_{ij} are simple fractions, such that Π is a function of the Π_i:

$$\Pi = \Phi(\Pi_1, \ldots, \Pi_k). \tag{1}$$

This is just a consequence of the requirement that a physical relationship be independent of the size of the units of measurement. At this stage nothing can be said about the function Φ. Now suppose the variables Π_i are small or large, and assume that the function Φ has a non-zero finite limit as its arguments tend to zero or to infinity; then $\Pi \sim$ constant, and one finds a power monomial relation between a and the a_i. This is a complete similarity relation. If the function Φ does not have the assumed limit, it may happen that for Π_1 small or large, $\Phi(\Pi_1) = \Pi_1^{\alpha}\Phi_1(\Pi_1) + \cdots$, where the dots denote lower order terms, α is a constant, the other arguments of Φ have been omitted and Φ_1 has a finite non-zero limit. One can then obtain a scaling expression for a in terms of the a_i and b_i, with undetermined powers which must be found by means other than dimensional analysis. The resulting power relation is an incomplete similarity relation. Of course one may well have functions Φ with neither kind of similarity.

Incomplete similarity expresses what is invariant under a renormalization group; all renormalization group transformations involve incomplete similarity. The exponent α is called an anomalous exponent.

2. Averaging a Hamiltonian system

We begin by examining what happens when one tries to reduce the complexity of a Hamiltonian system by averaging (see also [15; 16; 38; 2]). This first section is partially historical – this is how our group in Berkeley started working on problem reduction; part of this development has been superseded by the theory in the section on the Mori–Zwanzig formalism below. It seems to us that this is still the right place to start, because the conclusions here explain the (less than intuitively obvious) need to go beyond averaging to a more complicated theory, and also because the theory in this section is the basis for the analysis of the stationary case in the two sections that follow.

Consider a system of nonlinear ordinary differential equations,

$$\begin{aligned} \frac{d}{dt}\varphi(t) &= R(\varphi(t)), \\ \varphi(0) &= x, \end{aligned} \tag{2}$$

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 (2) corresponds a trajectory $\varphi(t) = \varphi(x, t)$.

Suppose that we only want to find m of the n components of the solution vector $\varphi(t)$ without finding the $n - m$ others. One has to assume something about the variables that are not evaluated, and we assume that at time t=0 we have a joint probability density $f(x)$ for all the variables. The variables we keep will have definite initial values $x_1, x_2, \ldots, x_m$, and the rest of variables will then

have a conditional probability density $f_m = f(x_1, \dots, x_m, x_{m+1}, \dots)/Z_m$, where $Z_m = \int_{-\infty}^{+\infty} f(x_1, \dots, x_m, x_{m+1}, \dots) dx_{m+1} dx_{m+2} \cdots$ is a normalization constant. Without some assumption about the missing variables the problem is meaningless; this particular assumption is reasonable because in practice f can often be estimated from previous experience or from general considerations of statistical mechanics. The question is how to use this prior knowledge in the evaluation of $\varphi(t)$.

Partition the vector x so that $\hat{x} = (x_1, x_2, \dots, x_m)$, $\tilde{x} = (x_{m+1}, \dots, x_n)$ 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, $\hat{R} = \hat{R}(\varphi) = \hat{R}(\hat{\varphi}, \tilde{\varphi})$; if they do not we have a system of m equations in m variables and nothing further needs to be done. We want to calculate only the variables $\hat{\varphi}$; then $(d/dt)\hat{\varphi}(t) = \hat{R}(\varphi(t))$ where the right-hand side depends on the variables $\tilde{\varphi}$ which are unknown at time t. We shall call the variables $\hat{\varphi}$ the "resolved variables" and the remaining variables $\tilde{\varphi}$ the "unresolved variables".

Consider in particular a Hamiltonian system as in [15],[16]. There exists then by definition a Hamiltonian function $H = H(\varphi)$ such that for i odd R_i, the i-th component of the vector R in (2) 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 $f(\varphi) = Z^{-1} \exp(-H/T)$ where T is a parameter, known in physics as the "temperature", which will be set equal to one in much, but not all, of the discussion below. In physics this density appears naturally and is known as the "canonical" density; the normalizing constant $Z = Z(T)$ is the "partition function". This density f is invariant, i.e., sampling it and evolving the system in time commute.

A numerical analyst who wants to approximate the solution of an equation usually starts by approximating the equation. If one solves for the resolved variables one has values for the variables $\hat{\varphi}$ available at each instant t and the best approximation should be a function of these variables; it is natural to seek a best approximation in the mean square sense with respect to the invariant density f at each time; the best approximation in this sense is the conditional expectation $E[R(\varphi)|\hat{\varphi}] = \int R e^{-H} d\tilde{\varphi} / \int e^{-H} d\tilde{\varphi}$ (note that we set $T = 1$ for simplicity). This conditional expectation is the orthogonal projection of R onto the space of functions of $\hat{\varphi}$ with respect to the inner product $(u, v) = E[uv] = \int u(\varphi) v(\varphi) f(\varphi) d\varphi$, where $d\varphi$ denotes integration over all the components of φ. We then try to approximate the system (2) by:

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

It has been shown in [13; 15; 11] that: (i) the new system (3) is also Hamiltonian:

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

where $i \leq m =$ the dimension of $\hat{\varphi}$, and

$$\hat{H} = -log \int \exp(-H)d\tilde{\varphi} \tag{5}$$

is the new Hamiltonian.

(ii) the new canonical density $\hat{f} = Z^{-1}\exp(-\hat{H})$ 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.

Now the question is, what does the solution $\hat{\varphi}(t)$ of (3) represent? Having averaged the equations, one could hope that the result is an average of the solution, of course constrained by the initial data $\hat{x}$, i.e., that the solution of equations (3) is $E[\hat{\varphi}(t)|\hat{x}]$. This is the case for linear systems (where averaging and time integration commute), and is approximately the case for limited time in some other special situations – nearly linear systems and some systems where the "unresolved variables" are fast. However, in general this is not the case. On the other hand, the solution of equations (3) does not approximate the true values of $\hat{\varphi}(t)$ in the full system either – the latter depend strongly on the missing data $\tilde{x}$ while the former does not. We shall see below that a reduced description of the solution of nonlinear systems in time requires in general a "noise" (which describes the fluctuations in $\tilde{\varphi}(t)$) and a "memory" (which depends on the temporal fluctuations of the noise and on the history of the solutions).

The fact that the solution of the averaged equations is not the average of the solutions can be understood by the following physics argument. In physics, a system in which the values of all the variables are drawn from a canonical distribution is a system in thermal equilibrium. The assignment of definite values $\hat{x}$ to the variables $\hat{\varphi}$ at time $t = 0$ amounts 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. Thus the predictive value of the partial initial data $\hat{x}$ decreases in time; all averages of the $\hat{\varphi}$ approach equilibrium averages. However, the reduced system (3) is Hamiltonian, and the solutions it produces oscillate forever.

In Figure 1 we consider the Hald Hamiltonian system [13] with

$$H = \frac{1}{2}\left(\varphi_1^2 + \varphi_2^2 + \varphi_3^2 + \varphi_4^2 + \varphi_1^2\varphi_3^2\right) \tag{6}$$

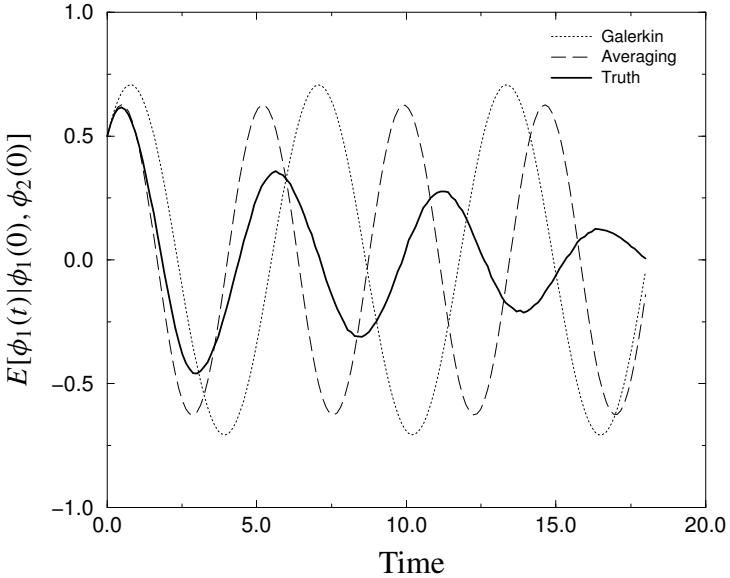


Figure 1. Comparison of the evolution of $E[\phi_1(t)|\phi_1(0), \phi_2(0)]$ (truth), to the prediction by the "Galerkin" approximation and the prediction by the averaging procedure described in the text.

(physically, two linear oscillators with a nonlinear coupling). We assume that $\varphi_1(0), \varphi_2(0)$ are given and sample the two other initial data from the canonical distribution with $T = 1$.

Figure 1 displays (1) the result for φ_1 of a "Galerkin" calculation in which the unresolved variables are set to zero (this is what is implicitly done in many unresolved computations); (2) the result of the averaging procedure just described, and (3) the true $E[\varphi_1(t)|\hat{x}]$, calculated by repeatedly sampling the initial data, solving the full system, and averaging. As one can see, averaging is initially better than the null "Galerkin" method, but in the long run the truth decays but the solution of the averaged system oscillates forever. For more detail, see [13].

Consider now the current practice of "large-eddy simulation" in hydrodynamics (see, e.g., [31]). One defines there as $\hat{\varphi}(t)$ "filtered" (i.e., locally averaged) variables and one finds for the time evolution of these variables equations obtained by relating various averaged terms in the Navier–Stokes equations to the filtered variables at one time. The result can be exactly equivalent to equations (3), as in [30], or indeed it could be an even worse approximation, because the conditional expectation of R is the best approximation of R by a function of the $\hat{\varphi}$. One should consider the possibility that some of the well-known difficulties of large-eddy simulation are

due to basic flaws in this procedure, and we will offer a possible alternative below. For a description of special cases, with small fluctuations and particular structures, where the use of equations (3) is legitimate, see [21].

3. Prediction with no data and block Monte Carlo

There is, however, a case where the construction of the preceding section can be very useful – when one tries to predict the future with no initial information. All the data are then sampled from the canonical density, which is invariant. If the system is ergodic, the solutions of equations (2) sample the space of solutions and their time average equals their average with respect to the canonical density. The system then simply samples the canonical density, and the reduction by conditional expectation of the previous section creates a smaller system whose variables have the same probability density after reduction as they had before reduction, and can be sampled at lower cost. This is the starting point for some interesting analysis as well as for block sampling methods (see [38; 2] for applications to molecular dynamics).

To see in detail what the reduction by conditional expectations of the previous section accomplishes under these circumstances, suppose the variables φ_i are associated with nodes on a regular lattice, for example, they may represent spins in a solid, or originate in the spatial discretization of a partial differential equation.

Divide the lattice into blocks of some fixed shape (for example, divide a regular one-dimensional lattice into groups of two contiguous nodes). We have not yet specified how the variables are to be divided into resolved and unresolved. Now decide to "resolve" one variable per block, and leave the others in the same block unresolved. The transformation between the old variables and the smaller set of resolved variables is a Kadanoff renormalization group transformation exactly as the latter are defined in [28] even if the steps which lead to it are presented differently; the Hamiltonian $\hat{H}$ defined above in equation (5) is the renormalized Hamiltonian in the sense of Kadanoff. This is an easy instance of our general claim that problem reduction is renormalization.

Suppose the system described by the Hamiltonian is translation invariant. The equations of motion at any one point, say at the location labeled by 1, have the same form as the equations of motion at any other point. The relation between the right-hand side of the reduced system and the right-hand side of the old system can be rewritten as:

$$\frac{\partial \hat{H}}{\partial \varphi_1} = E[\frac{\partial H}{\partial \varphi_1}|\hat{\varphi}], \tag{7}$$

where the expected value is with respect to the invariant density as before. This relation is the starting point for the evaluation of $\hat{H}$.

The key to success is to expand H and $\hat{H}$ in series, so that the calculation of the conditional expectations becomes easier for each term than it is for the Hamiltonians themselves. We use here a version of what is known in physics as an expansion in successive couplings (see [28]). The Hamiltonians are functions of the variables φ and can be expanded in the form:

$$H = \sum_j a_j \psi_j, \tag{8}$$

where the ψ_j are "elementary Hamiltonians". In a translation invariant system, where each equation has the same form as any other, the Hamiltonian is made up of sums over i of terms of the form $h(\varphi_i \varphi_{i+j})$ for various values of j, where h is some function; these terms represent "couplings" between variables j apart; one can then choose the elementary Hamiltonians to be polynomials in $x_i x_{i+j}$ with a fixed j in each ψ_j, i.e., one segregates the couplings between variables j apart into separate terms.

In a homogeneous system where there is only one variable per site, it is enough to satisfy (7) for one variable, say for φ_1. Define $\psi'_j = \frac{\partial}{\partial \varphi_1}\psi_j$, noting that though each ψ_j for a homogeneous system is necessarily a function with at least as many arguments as there are components on φ, ψ'_j can be sparse in the sense that it depends only on a few of the variables (for example, if $\psi_0 = \sum_i \varphi_i^2$, then $\psi'_0 = 2\varphi_1$). Equation (7) reduces to

$$\frac{\partial \hat{H}}{\partial \varphi_1} = \sum_j a_j P \psi'_j(\varphi) \tag{9}$$

with the projection P defined as before by $Pg(\varphi) = E[g|\hat{\varphi}]$ for any function g of φ. Now we're almost done. Pick a basis in $\hat{L}_2$, the subspace of square integrable functions that depend only on the variables $\hat{\varphi}$, made up of a subset of the set of functions ψ'_j. The right-hand side of equation (9) is then again a linear combination of ψ'_j; integration with respect to φ_1 requires only the erasure of the primes and yields a series for $\hat{H}$. The elements of $\tilde{\varphi}$ are now gone, and one can relabel the remaining variables $\hat{\varphi}$ so that the terms in the series have exactly the same form as before; the calculation can then be repeated, yielding a sequence of Hamiltonians with ever fewer variables: $H, H^{(1)} = \hat{H}, H^{(2)} = \hat{H}^{(1)}, \ldots$. The corresponding densities $f^{(n)} = Z^{-1} \exp(-H^{(n)}/T)$ can be sampled by any sampling scheme, for example, by Metropolis sampling (see, e.g., [10]).

At this point we have reduced the number of variables by a factor L equal to the number of variables in each block, but this may well seem to be a Pyrrhic victory. The Hamiltonians one usually encounters are simple in the sense that they involve few couplings – finite differences typically link a few neighboring variables, and so do the usual spin Hamiltonians in physics. As one reduces the number of variables,

the new Hamiltonians become more complex, with more terms in the series (8); the cost per time step of solving the equations in time or the cost per move in a Metropolis sampling typically increases quickly as well. To see what has been gained one must again turn to the physics literature (see, e.g., [28],[24]).

Consider the spatial correlation length ℓ which measures the range of values of $|j|$ over which the spatial covariances $E[\varphi_i \varphi_{i+j}]$ are non negligible, and the correlation time τ for which the temporal covariances $E[\varphi_i(t)\varphi_j(t+s)]$ are non-negligible. For very large and very small values of the temperature T (the variance parameter in the density f) both the correlation time and the correlation length are usually small (see [28],[17]); the properties of the system can then be found from calculations with a small number of variables and it is not urgent to reduce the number of variables. There is a range of intermediate values of T for which the correlation length and time are large and then the reduction is worthwhile. There often is a value T_c of T, the "critical value", for which $\ell = \infty$. Values of T around T_c are often of great interest.

Now we can see what the reduction can accomplish. If one tries to compute averages with T near T_c one finds that the cost of computation is proportional to τ and to some positive power of ℓ – one has to compute long enough to obtain independent samples of φ, and a new independent sample will not appear until a time $\sim \tau$ has passed. The reductions above produce a system with smaller ℓ and τ and therefore computation takes less time. Though we started with the declared goal of reducing the number of variables, what has been produced is more interesting: a new system with shorter correlations which is more amenable to computation. It is not the raw number of variables that matters. It is important to notice that what started as a scheme for winnowing out variables has ended up by producing a new system related to the original system by a scaling transformation.

The renormalization can be used with a multigrid scheme, in which one runs up and down on different levels of renormalization, on the finer ones to achieve accuracy and the cruder ones to move fast from one macroscopic configuration to another. It is well known that multigrid schemes require that one store conditional expectations (see, e.g., [7]), and the physicists' expansion in successive linkages provides an effective way to do so; for details see [10],[35].

An alternative method for obtaining the expansion coefficients for the renormalized Hamiltonians was proposed in [42]. The method is based on the maximization of the likelihood of the renormalized density. The maximization of the likelihood leads to a moment-matching problem. The moments in this case are the expectation values of the "elementary Hamiltonians" (see above) with respect to the renormalized density. The solution of the moment matching problem yields the expansion of the renormalized Hamiltonian.

The systematic development of the links of probability with renormalization began with Jona–Lasinio (see, e.g., [26]). The connection of renormalization with incomplete similarity is too well known (see [3; 28; 22]) to require further comment here. The analysis of this section provides a striking example of the benefits to be found in applying to computation ideas drawn from experience in statistical physics.

4. An example: The Korteveg–deVries–Burgers equation

As a further illustration of the ideas in the previous section, consider the equation

$$u_t + uu_x = \epsilon u_{xx} - \beta u_{xxx}, \tag{10}$$

with boundary conditions

$$u(-\infty) = u_0, \quad u(+\infty) = 0, \quad u_x(-\infty) = 0, \tag{11}$$

where the subscripts denote differentiation, x is the spatial variable, t is time, $\epsilon > 0$ is a diffusion coefficient, $\beta > 0$ is a dispersion coefficient, and $u_0 > 0$ is a given constant. The boundary conditions create a traveling wave solution moving to the right (towards $+\infty$) with velocity $u_0/2$ which becomes steady in a moving framework as $t \to \infty$. In nondimensional form the equation can be written as:

$$u_t + uu_x = \frac{1}{R} u_{xx} + u_{xxx}, \tag{12}$$

with $u_x(-\infty) = 0$, $u(+\infty) = 0$, $u(-\infty) = 1$; $R = \sqrt{\beta u_0}/\epsilon$ is a "Reynolds number". For $R \leq 1$ the traveling wave has a monotonic profile, while for $R > 1$ the profile is oscillatory, with oscillations whose wave length is of order 1 [6]. At zero diffusion ($R = \infty$) the stationary asymptotic wave train extends to infinity on the left. For finite R the wave train is damped and the solution tends to 1 as x decreases.

The steady wave profile can be found by noting that it satisfies an ordinary differential equation, whose solution connects a spiral singularity at $x = -\infty$ to a saddle point at $x = +\infty$. At the steady state we average the solution at each point x over the region $(x - \ell/2, x + \ell/2)$ and call the result $\bar{u}$. The task we set ourselves is to find an effective equation $g(v, v_x, v_{xx}, \ldots) = 0$ whose solution v approximates $\bar{u}$; v can be expected to be smoother than the solution of (12) and thus require fewer mesh points for an accurate numerical solution; this is analogous to finding a renormalized Hamiltonian further from the critical point so that the solution of the corresponding problem has lower fluctuations, as we did in the previous section; note that the problem of this present section is not Hamiltonian.

We now make an analogy between the conditional expectations which define the renormalized variables in the previous sections and an averaging in space which defines "renormalized" variables for solutions of the KdVB equations that are

stationary in a moving frame. 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 previous sections, 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. 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 theorems which guarantee that averaging the right-hand sides produces the correct statistics for the solutions.

We can look for an effective equation in the class of equations of the form

$$-cv_x + vv_x = \epsilon_{eff} v_{xx} + v_{xxx} + \beta |v_x|^{\alpha} v_{xx} + \cdots , \tag{13}$$

where $\epsilon \geq 0$, $\alpha \geq 0$, $\beta \geq 0$ are constants and $c = 1/2$ is the velocity of propagation of the steady wave (see also [4]). This expansion is analogous to the expansion in successive linkages (8) of the previous section; in a continuum limit, a series of partial Hamiltonians, whose derivatives have larger and larger "stencils" across which variables are connected, can be reorganized into an expansion in higher and higher derivatives of the unknown. One knows a priori that u and v propagate at the same velocity, which helps fix some of the parameters (i.e., expansion coefficients) at the outset. The problem is to find the values of the parameters in the effective equation which minimize

$$I = \int_{-\infty}^{+\infty} |\bar{u}(x) - v(x)|^2 dx. \tag{14}$$

One finds numerically that the last terms have little effect on the minimum of I when $\ell \geq 5$ (in physics terminology, they are "irrelevant"). The effective equation is thus a Burgers equation with a value of the dimensionless diffusion coefficient ϵ_{eff} different from $1/R$.

The minimization in (14) was carried out in [9], and it showed that the minimum was achieved when $\epsilon_{eff} = R^{\nu}\Phi(\ell)$, with the exponent $\nu \sim 0.75$. Note that when the diffusion coefficient $\epsilon \to 0$, then $\epsilon_{eff} \to \infty$! This is an incomplete similarity relation, as advertised, relating a "bare" Reynolds number R to a "dressed" Reynolds number ϵ_{eff}^{-1}. The form of the effective equation could conceivably have been found by

averaging the original equation, but the relation between the original ϵ and ϵ_{eff} requires some form of renormalization-like reasoning.

5. The Mori–Zwanzig formalism

We now return to the problem we started investigating in section 2: how to determine the evolution of a subset $\hat{\varphi}$ of components of a vector φ described by a nonlinear set of equations of the form (2). This is a nonlinear closure problem of a type much studied in physics, and a variety of formalisms is available for the job. We choose the Mori–Zwanzig formalism of irreversible statistical mechanics [19; 23; 33; 46; 34], because it homes in on the basic difficulty, which is the description of the memory in the system; the relation of this formalism to other nonlinear formalisms is described in [14]. That a reduced description of a nonlinear system involves a memory should be intuitively obvious: suppose you have $n > 3$ billiard balls moving about on top of a table and are trying to describe the motion of just three; the second ball may strike the seventh ball at a time t_1 and the seventh ball may then strike the third ball at a later time. The third ball then "remembers" the state of the system at time t_1, and if this memory is not encoded in the explicit knowledge of where the seventh ball is at all times, then it has to be encoded in some other way. We are no longer assuming that the system is Hamiltonian nor that we know an invariant density.

It is much easier to do theory for linear equations, and we start by finding a linear equation equivalent to (not approximating!) the system (2). Introduce the linear Liouville operator $L = \sum_{i=1}^{n} R_i(x)\frac{\partial}{\partial x_i}$, and the Liouville equation:

$$\begin{aligned} \frac{\partial}{\partial t}u(x,t) &= Lu(x,t) \\ u(x,0) &= g(x), \end{aligned} \tag{15}$$

with initial data $g(x)$. This is the partial differential equation for which (2) is the set of characteristic equations. One can verify that the solution of the Liouville equation is $u(x,t) = g(\varphi(x,t))$ (see, e.g., [11]). In particular, if $g(x) = x_i$, the solution is $u(x,t) = \varphi_i(x,t)$, the i-th component of the solution of (2). This linear partial differential equation is thus equivalent to the nonlinear system (2). The linearity of equation (15) greatly facilitates the analysis.

Introduce 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; therefore $e^{tL}g(x) = g(e^{tL}x)$, and one can also verify that $e^{tL}L = Le^{tL}$ (this can be seen to be a change of variables formula). Equation (15) becomes

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

We suppose that as before we are given the initial values of the m coordinates $\hat{x}$, and that the distribution of the remaining $n-m$ coordinates $\tilde{x}$ is the conditional density, f conditioned by $\hat{x}$, where f is initially given.

We define a projection operator P by $Pg = E[g|\hat{x}]$. The conditioning variables are the initial values of $\hat{\varphi}$; in section 2 the conditioning variables were the values of $\hat{\varphi}(t)$, which are unusable here when we do not know the probability density at time t. Quantities such as $P\hat{\varphi}(t) = E[\hat{\varphi}(t)|\hat{x}]$ are by definition the best estimates of the future values of the variables $\hat{\varphi}$ given the partial data $\hat{x}$ and are often the quantities of greatest interest.

Consider a resolved coordinate $\varphi_j(x,t) = e^{tL}x_j$ $(j \le m)$, and split its time derivative, $R_j(\varphi(x,t)) = e^{tL}Lx_j$ as follows:

$$\frac{\partial}{\partial t}e^{tL}x_j = e^{tL}Lx_j = e^{tL}PLx_j + e^{tL}QLx_j, \tag{16}$$

where $Q = I - P$. Define $\hat{R}_j(\hat{x}) = (PR_j)(\hat{x})$; the first term is $e^{tL}PLx_j = \hat{R}_j(\hat{\varphi}(x,t))$ and is a function of the resolved components only (but it is a function of the whole vector of initial data). Note that if Q were zero we would recover something that looks like the crude approximation of an earlier section; however the conditioning variables are not the same. We shall see that the term in Q is essential.

We further split the remaining term $e^{tL}QLx_j$. 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., [29]).

The evolution operators e^{tL} and e^{tQL} satisfy the Duhamel relation

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

Hence,

$$e^{tL}QLx_j = e^{tQL}QLx_j + \int_0^t e^{(t-s)L}PLe^{sQL}QLx_j\,ds. \tag{17}$$

Collecting terms, we find

$$\frac{\partial}{\partial t}e^{tL}x_j = e^{tL}PLx_j + \int_0^t e^{(t-s)L}PLe^{sQL}QLx_j\,ds + e^{tQL}QLx_j \tag{18}$$

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.

It is important to see that equation (18) is an identity. The memory and noise terms have not been added artificially, their presence is a direct consequence of the original equations of motion. However tempting it may be to average equations by taking one-time averages, the results will, in general, be wrong; one must add a memory and a noise as well. Note that the first term in equation (18) is, apart from the change of conditioning variables, the same as the right-hand side in equations (3).

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 L_2 to $\hat{\varphi}$ given the partial data $\hat{x}$), then one can premultiply equation (18) by P; the noise term then drops out and we find

$$\frac{\partial}{\partial t} P e^{tL} x_j = P e^{tL} P L x_j + P \int_0^t e^{(t-s)L} P L e^{sQL} Q L x_j \, ds. \tag{19}$$

Even if the system we start with is Hamiltonian, the Langevin equation (18) is not; the memory and the noise allow the system to forget its initial values and decay to "thermal equilibrium" as it should (see section 2).

Let $w(x,t) = e^{tQL} Q L x_j$; by definition w, the noise, is a solution of the initial value problem:

$$\begin{aligned} \frac{\partial}{\partial t} w(x,t) = QLw(x,t) \;&=\; Lw(x,t) - PLw(x,t) \\ w(x,0) &= QLx_j. \end{aligned} \tag{20}$$

If for some function $h(x)$, $Ph = 0$, then $Pe^{tQL}h = 0$ for all time t, i.e., e^{tQL} maps the null space of P into itself. The solution of the equations (20) defines the "orthogonal dynamics" for the system (2) with with data $\hat{x}$ and the given joint density for all the data at the initial time. The initial data for the orthogonal dynamics, $QLx_j = (I-P)R_j = R_j - E[R_j|\hat{x}]$ can be thought of as the fluctuations in the initial values of the R_j. The range of the projection P is everything that can be expressed as a function of $\hat{x}$, i.e., everything that can be predicted from the knowledge of $\hat{x}$; one can think of the range of P as the "resolved space". One can think of the range of Q as the "noise space". The orthogonal dynamics modify the temporal evolution that starts from QLx_j by continuously removing from the evolutes any component that can be resolved or predicted; the result always remains in the noise space.

We now show that the memory term is a functional of the temporal covariances of the noise (i.e., of covariances of stochastic processes confined to the noise space). To save on writing 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 canonical data, see [13],[18]; however, here the assumption of skew-symmetry is only an excuse

to reduce the number of symbols, not a return to the Hamiltonian case. Pick an orthonormal basis $\{h_k = h_k(\hat{x}), k = 1, \dots\}$ in the range of P, which is the space of functions of $\hat{x}$ (for example, the h_k could be Hermite polynomials in the variables $\hat{x}$). The projection of any function $\psi(x, t)$ can be written as $\psi = \sum_k (\psi(x, t), h_k) h_k(\hat{x})$, and in particular,

$$P(LQe^{sQL}QLx_j) = \sum_k (LQe^{sQL}QLx_j, h_k) h_k(\hat{x}), \tag{21}$$

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

$$\begin{aligned}\int_0^t e^{(t-s)L} PLe^{sQL}QLx_j ds &= \int_0^t \sum_k e^{(t-s)L}(LQe^{sQL}QLx_j, h_k) h_k(\hat{x}) ds \\ &= \sum_k \int_0^t (LQe^{sQL}QLx_j, h_k) h_k(\hat{\varphi}(t-s)) ds. \end{aligned} \tag{22}$$

In the last identity we used the fact that the inner product in parentheses is independent of time and therefore commutes with the time evolution operator e^{tQL}, and also the fact that $e^{(t-s)L} h_k(\hat{x}) = h_k(\hat{\varphi}(t-s))$. Now $(LQe^{sQL}QLx_j, h_k(\hat{x})) = -(e^{sQL}QLx_j, QLh_k(\hat{x}))$ by the symmetry of Q and the assumed skew-symmetry of L; each term on the right-hand side of equation (22) is the ensemble average of the product of the value of the stochastic process $e^{tQL}QLx_j$ at time $s = t$, with the value of the stochastic process $e^{tQL}QLh_k(\hat{x})$ evaluated at time $s = 0$, i.e., it is a temporal correlation. All these stochastic processes are in the range of Q for all t, and are therefore components of the noise. Remember that by definition $Lx_j = R_j$ (a right-hand side in equations (2)). PLx_j is then an average of the right-hand side of (2) and $QLx_j = R_j - E[R_j|\hat{x}]$ is the initial fluctuation in that right-hand side.

The first, "Markovian", term in equations (18) looks straightforward, but perils lurk there as well. In general R_j in equations (2) is nonlinear, and so is $PLx_j = E[R_j|\hat{x}]$. $e^{tL}PLx_j$ is a nonlinear function of the functions $\hat{\varphi}(t)$ that depends on all the components of x, not only on $\hat{x}$. Some way of approximating this function must be found. If one looks for conditional expectations, one must find a way to commute P with a nonlinear function; for a discussion, see [13]. This bullet was dodged in section 2 when the conditioning variables were chosen to be $\hat{\varphi}(t)$ which change in time, but it may be hard to dodge in general.

The task now at hand is to extract something usable from these rather cumbersome formulas. A very detailed presentation of the analysis in this section can be found in [17].

6. Fluctuation-dissipation theorems

We have established a relation between kernels in the memory term and the noise (the former is made up of covariances of the latter). This is the mathematical content of what are known as "fluctuation-dissipation theorems" in physics. A key difficulty is that the kernels in the memory term consist of covariances of the orthogonal dynamics, whose determination requires in principle the solution of the orthogonal dynamics equations (20), which can be very onerous. However, in the physics literature fluctuation-dissipation theorems are presented in a way that does not stress this difficulty, and we take a moment to explain how the usual physics versions of the theorems come about; they are worth understanding because even though they camouflage the orthogonal dynamics issue they contain significant additional insights.

In the physics literature one often takes a restricted basis in the range of P consisting of the coordinate functions $x_1, ..., x_m$ (the components of $\hat{x}$). The resulting projection is called the " linear projection" as if P as defined above were not linear. The use of this projection is appropriate when the amplitude of the functions $\hat{\phi}(t)$ is small. One then has $h_k(\hat{x}) = x_k$ for $k \leq m$. The covariances in equation (22) are then simply the temporal covariances of the fluctuations in the resolved variables only – all the other covariances have been set to zero. This is known as the fluctuation-dissipation theorem of the second kind. The fluctuations of course obey the orthogonal dynamics equation.

Specialize further to a situation where there is a single resolved variable, say ϕ_1, so that $m = 1$ and $\hat{\phi}$ has a single component. The Mori–Zwanzig equation becomes:

$$\frac{\partial}{\partial t} e^{tL} x_1 = e^{tL} PLx_1 + e^{tQL} QLx_1 + \int_0^t e^{(t-s)L} PLe^{sQL} QLx_1 ds,$$

or,

$$\begin{aligned}\frac{\partial}{\partial t}\phi_1(x,t) &= (Lx_1, x_1)\phi_1(x,t) + e^{tQL} QLx_1 \\ &\quad + \int_0^t (LQe^{sQL} QLx_1, x_1)\phi_1(x, t-s)ds \\ &= (Lx_1, x_1)\phi_1(x,t) + e^{tQL} QLx_1 - \int_0^t (e^{sQL} QLx_1, QLx_1)\phi_1(x, t-s)ds,\end{aligned} \tag{23}$$

where we have again inserted a harmless factor Q in front of e^{QL}, assumed that L was skew-symmetric as above, and for the sake of simplicity also assumed $(x_1, x_1) = 1$ (if the last statement is not true the formulas can be adjusted appropriately). Take

the inner product of equation (23) with x_1, you find:

$$\frac{\partial}{\partial t}(\phi_1(x,t),x_1) = (Lx_1,x_1)(\phi_1(x,t),x_1) + (e^{tQL}QLx_1,x_1) - \int_0^t (e^{sQL}QLx_1,\, QLx_1)\phi_1(x,t-s)ds$$
$$= (Lx_1,x_1)(\phi_1(x,t),x_1) - \int_0^t (e^{sQL}QLx_1,\, QLx_1)(\phi_1(x,t-s),x_1)ds, \quad (24)$$

because $Pe^{tQL}QLx_1 = (e^{tQL}QLx_1,x_1)x_1 = 0$ and hence $(e^{tQL}QLx_1,x_1) = 0$. Multiply equation (24) by x_1, and remember that $P\phi_1(x,t) = (\phi_1(x,t),x_1)x_1$. You find:

$$\frac{\partial}{\partial t}P\phi_1(x,t) = (Lx_1,x_1)P\phi_1(x,t) - \int_0^t (e^{sQL}QLx_1,\, QLx_1)P\phi_1(x,t-s)ds. \quad (25)$$

Observe that the covariance $(\phi_1(x,t),x_1)$ and the projection of ϕ_1 onto x_1 obey the same homogeneous linear integral equation. This is the fluctuation-dissipation theorem of the first kind, which embodies the Onsager principle, according to which spontaneous fluctuations in a system decay at the same rate as perturbations imposed by external means, when both are small (so that the linear projection is adequate). This reasoning can be extended to cases where there are multiple resolved variables, and this is usually done with the added simplifying assumption that $(x_i,x_j) = 0$ when $i \neq j$. We omit the details. Finally, if one makes short-memory approximations as in the next section, the issue of orthogonal dynamics disappears completely, as we shall now see.

7. Short-range memory

We have already pointed out that a salient difficulty in using the Mori–Zwanzig equations (18) is the need to solve the orthogonal dynamics equation. We wish now to examine what happens if one bypasses these equations by replacing the orthogonal dynamics by the real dynamics, i.e., if one sets:

$$e^{tQL} \cong e^{tL}. \quad (26)$$

We will show that this is a reasonable approximation under some important circumstances, and that the approximation leads to greatly simplified equations.

First, some heuristic comments. If the resolved dynamics (what happens in the range of P) have no effect on the noise, then the assumption (26) should be valid, for then the unresolved variables interact just with each other; the resulting noise remains unpredictable from the knowledge of $\hat{x}$ and thus remains in the noise space; e^{tQL} and e^{tL} acting on a vector in the noise space should be the same. The effect of

the resolved variables on the noise is small in particular if (i) the memory (i.e., the range of t's for which the covariances in the memory term is significant) is short, or (ii) the memory is long. The noise $e^{tQL}QLx_j$ starts out in the noise space by construction, and if the memory is short the operator e^{tL} can take the quantities QLx_j only a small distance out of the noise space before it becomes irrelevant for the evaluation of the covariances; in this short time $e^{tQL}QLx_j$ and $e^{tL}QLx_j$ are the same. If the memory is long, the noise goes on unaffected by the resolved variables. We therefore examine the approximation (26) in these two opposite cases.

In the present section we examine the case of short memory. The memory term in the Mori–Zwanzig equations (18) can be rewritten as

$$\int_0^t e^{(t-s)L}PLe^{sQL}QLx_j\,ds = \int_0^t e^{(t-s)L}PLQe^{sQL}QLx_j\,ds, \tag{27}$$

where the insertion of the extra Q is harmless. Adding and subtracting equal quantities, we find:

$$PLe^{sQL}QLx_j = PLQe^{sL}QLx_j + PLQ(e^{sQL} - e^{sL})QLx_j; \tag{28}$$

a Taylor series yields:

$$e^{sQL} - e^{sL} = I + sQL + \cdots - I - sL - \cdots = -sPL + O(s^2), \tag{29}$$

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

$$\int_0^t e^{(t-s)L}PLe^{sQL}QLx_j\,ds = \int_0^t e^{(t-s)L}PLQe^{sL}QLx_j\,ds + O(t^3). \tag{30}$$

If P is a finite rank projection then

$$PLe^{sQL}QLx_j = \sum_k (QLe^{sQL}QLx_j, h_k)h_k(\hat{x}), \tag{31}$$

where, as before, one can write $(QLe^{sQL}QLx_j, h_k)$ as $-(e^{sQL}QLx_j, QLh_k)$ when L is skew-symmetric. If the covariances $(e^{sQL}QLx_j, QLh_k)$ and also the covariances $(e^{sL}QLx_j, QLh_k)$ are significant only over short times t_0, the approximation (26) provides an approximation with an error $O(t_0^3)$ without requiring the solution of the orthogonal dynamics equation; this is still a short covariance time approximation but it can be preferable to a white noise approximation (see [41] for an application to the dimensional reduction of the Kuramoto–Sivashinsky equation and [2] for an application to molecular dynamics).

One important short-memory situation where the Mori–Zwanzig formalism simplifies even more is when the noise can be viewed as white noise. This is a valid approximation in a number of important cases, in particular when there is scale separation between the resolved and unresolved variables or when these variables

are weakly coupled (for recent reviews see, e.g., [21],[32], [40]). These situations are often encountered in applications, but we do not survey them here because their analysis does not require all of our machinery.

If the noise can indeed be viewed as white, one sets:

$$e^{tQL}QLx_j = A_j w_j'(t), \tag{32}$$

where the prime denotes a derivative, the $w_j(t)$ are independent unit Brownian motions so that that the w' are white noises, and the A_j are constants that must be derived from some prior knowledge. The covariances of the noise are then delta functions (thus the memory is vanishingly short). If one assumes further that the projection P is well represented by the physicists' "linear" projection, then the integral in the memory term can be easily seen to reduce to a constant times the unknowns, and equations (18) become stochastic ordinary differential equations of the usual kind. As usual (see, e.g., [27]), the corresponding probability densities can be found via Fokker–Planck formalisms (or Kolmogorov equations, in mathematicians' language).

It is important to note that the assumption of white noise does not require that the linear projection be used. More noise terms appear when one uses a more general linear projection, and one encounters situations where the additional noise terms can no longer be viewed as white and their uses detracts from the overall accuracy (see, e.g., [41; 42; 32]). These papers also include suggestions as to how to pick the best number of terms to use in the projections. Projections other than linear are important for mode-coupling theory in condensed matter physics, see, e.g., [45].

There is a comment to be added here. White noise and delta memory constitute an important special case. However, this is not the general case and maybe not even the usual case. It is rather surprising that 40 years after Alder and Wainwright [1] demonstrated the long-range memory in a common physical system, years during which physicists have learned how to model systems with arbitrary memory, most numerical treatments of dimensional reduction seem to assume that all memory is ultra-short. It is also surprising that most papers on dynamic renormalization (see, e.g., [24]) assume that the noise is white without comment, making it pointless to compare the schemes below with this dynamical renormalization literature.

Finally, it should be obvious that very short memory is very different from no memory, i.e., from situations where the memory term is absent altogether.

8. Long-range memory and the t-model

We examine now the validity of the ansatz $e^{tQL} \cong e^{tL}$ for cases with slowly decaying memory. Write the memory term in the Mori–Zwanzig equation (18) as

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

where we have used the commutation of L and QL with e^{tL} and e^{sQL}, respectively. At this point, make the approximation (26), which eliminates the s dependence of both integrands and we obtain:

$$\int_0^t e^{(t-s)L} PLe^{sQL} QLx_j ds \cong te^{tL} PLQLx_j. \tag{33}$$

All that remains of the integration in time is the coefficient t. To estimate the error, consider the difference between the full memory term and its approximation:

$$\int_0^t e^{(t-s)L} PLe^{sQL} QLx_j ds - te^{tL} PLQLx_j = \int_0^t [e^{(t-s)L} PLe^{sQL} - e^{tL} PL]QLx_j ds.$$

Adding and subtracting equal quantities, we find

$$e^{(t-s)L} PLe^{sQL} = e^{tL} PL + e^{tL}[e^{-sL} PLe^{sQL} - PL],$$

and a Taylor series around $s = 0$ gives

$$e^{-sL} PLe^{sQL} - PL = (I - sL + \ldots)PL(I + sQL + \ldots) - PL = O(s). \tag{34}$$

This implies

$$\int_0^t e^{(t-s)L} PLe^{sQL} QLx_j ds = te^{tL} PLQLx_j + O(t^2).$$

To understand this estimate, examine an alternate derivation of (33). Expand the integrand of the memory term of the Mori–Zwanzig equation around $s = 0$ and retain only the leading term, finding

$$\int_0^t e^{(t-s)L} PLe^{sQL} QLx_j ds = \int_0^t [e^{tL} PLQLx_j + O(s)]ds = te^{tL} PLQLx_j + O(t^2).$$

If we retain only the leading term, we do not keep any information about the time evolution of the integrand, which in turn means no information about the evolution of the resolved component and of the coupling to the orthogonal dynamics (through the term $(LQe^{sQL}QLx_j, h_k)$). Such a drastic approximation is expected to be appropriate in cases where the memory term integrand is slowly decaying, so that information about its initial value is sufficient to make predictions.

We have just seen that if the memory is long the ansatz $e^{tQL} \cong e^{tL}$ reduces the memory to a Markovian term with a time-dependent coefficient. Thus the assumption $e^{tQL} \cong e^{tL}$ greatly simplifies the equations, as expected. The resulting equations were introduced in [13] and are known as the "t-model".

As an example, consider again the Hald model whose Hamiltonian is

$$H(\phi) = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 + \phi_1^2\phi_3^2). \tag{35}$$

The resulting equations of motion are:

$$\begin{aligned}
\frac{d\phi_1}{dt} &= \phi_2 \\
\frac{d\phi_2}{dt} &= -\phi_1(1+\phi_3^2) \\
\frac{d\phi_3}{dt} &= \phi_4 \\
\frac{d\phi_4}{dt} &= -\phi_3(1+\phi_1^2).
\end{aligned}$$

Suppose one wants to solve only for $\hat{\phi} = (\phi_1, \phi_2)$, with initial data $\hat{x} = (x_1, x_2)$. Assume the initial data x_3, x_4 are sampled from a canonical density with temperature $T = 1$. A quick calculation yields $E[x_3^2|x_1, x_2] = 1/(1+x_1^2)$. The advance in time described by the multiplication by e^{tL} requires just the substitution $\hat{x} \to \hat{\phi}$. If one commutes the nonlinear function evaluation and the conditional averaging, i.e., writes $Pf(\hat{\phi}) = f(P\hat{\phi})$ (a "mean-field approximation"), and writes furthermore $\Phi(t) = P\hat{\phi} = E[\hat{\phi}|\hat{x}]$ one finds $Pe^{tL}PLx_1 = \Phi_2$, $Pe^{tL}PLx_2 = -\Phi_1(1 + 1/(1 + \Phi_2^2))$; one can calculate $Pe^{tL}LQLx_j$ for $j = 1, 2$ and finally one finds:

$$\begin{aligned}
\frac{d}{dt}\Phi_1 &= \Phi_2 \\
\frac{d}{dt}\Phi_2 &= -\Phi_1\left(1 + \frac{1}{1+\Phi_1^2}\right) - 2t\frac{\Phi_1^2\Phi_2}{(1+\Phi_1^2)^2}.
\end{aligned} \tag{36}$$

The last term represents the damping due to the loss of predictive power of partial data; the coefficient of the last term increases in time and one may worry that this last term eventually overpowers the equations and leads to some odd behavior. This

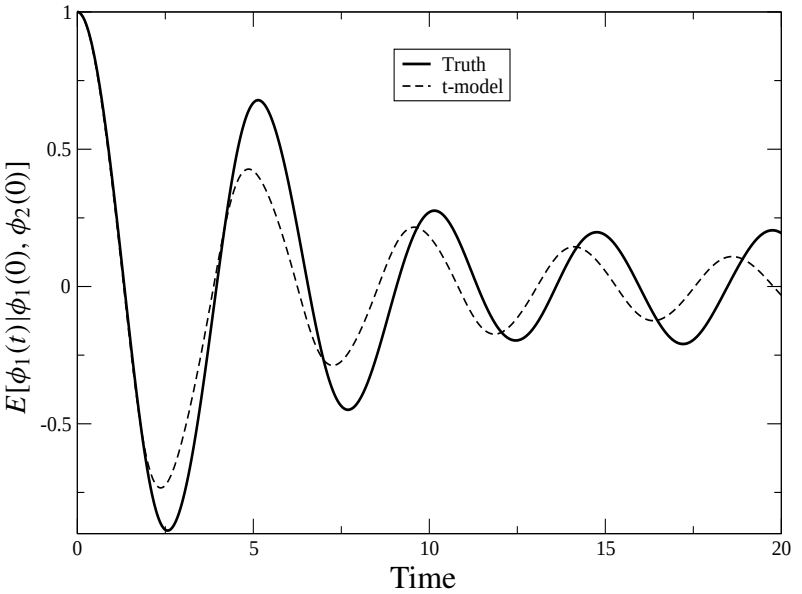


Figure 2. Comparison of the evolution of $E[\phi_1(t)|\phi_1(0), \phi_2(0)]$ (truth) with the prediction by the t-model; for comments, see the text.

is not the case. Indeed, one can prove the following general result: If the system one starts from, equation (2) is Hamiltonian with Hamiltonian H, and if the initial data are sampled from an initial canonical density conditioned by partial data $\hat{x}$, and if $\hat{H}$ is the renormalized Hamiltonian (in the sense of section 2), then $(d/dt)\hat{H} \leq 0$, showing that the components of $\hat{\phi}$ decay as they should. The proof requires a technical assumption (that the Hamiltonian H can be separated into the sum of a function of the momenta and a function of the position, a condition commonly satisfied) and we omit it (see [13]).

The solution of the t-model with the mean-field approximation for the Hald model is presented in Figure 2. The applicability of the approximation suffers from the fact that at the temperature $T = 1$ the Hald system is not ergodic. To see what has been gained, contrast this figure with Figure 1.

If the t-model is not sufficient for the approximation of a given problem, one can try to generalize it. Indeed, we have just seen that the t-model is the zero-th order term in a Taylor expansion (around $s = 0$) of the integrand of the memory term in (18). However, nothing prevents us from keeping more terms in this expansion. Let

$$K(\hat{\varphi}(t-s), s) = e^{(t-s)L} PLe^{sQL} QLx_j$$

and expand K around $s = 0$, i.e.,

$$K(\hat{\varphi}(t-s), s) = K(\hat{\varphi}(t), 0) + s\frac{\partial K}{\partial s}|_{s=0} + \frac{1}{2}s^2\frac{\partial^2 K}{\partial s^2}|_{s=0} + O(s^3).$$

In the case when P is the finite-rank projection and the density used to define the projection is invariant, the derivatives of K at $s = 0$ are equal-time (static) covariances. In mode-coupling theory, such expressions are known as "sum rules". One can assume a functional form for the memory term integrand around $s = 0$, e.g., a Gaussian ae^{-bs^2}, and use the derivatives of K at $s = 0$ to estimate a, b (see [37] for more on sum rules and mode-coupling theory). This is potentially another place where current ideas in physics can be helpful in numerical modeling.

The usefulness of the t-model depends on the range of the memory; this raises the question of what this range depends on and whether it can be modified. If the number of resolved variables is small, the range of the memory depends on the range of the memory in the full system (2)- indeed, if there are no resolved variables, as in section 3 above, the dynamics and the orthogonal dynamics are the same. However, in the general case, is it possible to have a reduced model with very short or very long memory, depending on how one coarse-grains a particular system at hand? In [41] evidence was presented that, for the Kuramoto–Sivashinsky equation, the range of the memory of a reduced model can vary dramatically, depending on whether all the unstable modes in the system are resolved or not. The construction of a reduced model corresponds to renormalization, and the two extreme cases can be interpreted as two fixed points of a renormalization scheme. In which one a reduced model will end up depends on how one renormalizes. How to formalize these remarks and put them to use remains a topic for further work.

Both the long memory approximation and the short memory approximation have been derived from the assumption $e^{tQL} \cong e^{tL}$, but this assumption has been used differently. In the short memory case one first makes this substitution in the memory term and then one performs the projection in that term; in the long memory case one performs these two operations in the reverse order. This leads to different results.

Finally, we go back to the remark at the end of section 2. We believe that the t-model is a sound basis for large eddy simulation in hydrodynamics; the equations are relatively simple and the memory is taken into account. We are acting on this basis and expect to publish results soon.

9. Intermediate-range memory

There are intermediate cases where the memory cannot be viewed as either short or long so that neither model above can be used. At present, it is not known how to deal effectively with such cases. In a series of papers [11]-[13] we presented special cases and their solutions. In particular in [13] we presented a detailed

analysis of the Hald system without the t-model assumptions. We showed that the memory decays roughly at the same rate as the solution itself (this is the general case in the absence of separation of scales). We expanded the various covariances at equilibrium (i.e., when there are no resolved variables) in Hermite polynomials, evaluated the coefficients in the expansions by Monte Carlo once and for all, and then obtained a system of integro-differential approximations to equations (18) which we then solved in various cases. This is a legitimate procedure which may be useful when the same system of equations has to be solved repeatedly. These calculations do exhibit a salient feature of model reduction in time-dependent problems, which is that its set-up costs are often very high. The future remedy, if there is one, will surely lie in a deeper understanding of dynamical renormalization and, in particular, of the way memory depends on scale.

10. Conclusions

We have made two sets of claims. First, theoretical claims: If one assumes that a probability density is initially available for all the degrees of freedom in a complex problem, then the problem of following the evolution of just a few degrees of freedom becomes a problem in statistical mechanics, of the equilibrium kind for problems with stationary densities, and of the non-equilibrium kind otherwise. Finding an equivalent problem with lesser complexity is equivalent to a renormalization, and a successful reduction in complexity corresponds to uncovering a similarity relation between the full problem and the reduced problem. Physics is often a good guide to what should be done.

On the practical side, reduction by conditional expectation is a powerful tool. In the stationary case we have used it to generate block Monte Carlo algorithms and effective equations for mean solutions. In the time dependent case it leads to the Mori–Zwanzig formalism, generalized Langevin equations, and promising approximation schemes. We have high hopes for the usefulness of one particular approximation scheme, the t-model, which yields good approximations in interesting cases with a relatively low overhead.

11. Acknowledgements

We would like to thank Prof. G.I. Barenblatt, Prof. O. Hald and Prof. R. Kupferman for all that we have learned from them; their ideas are present everywhere in this paper.

References

[1] B. Alder and T. Wainright, *Decay of the velocity correlation function*, Phys. Rev. A **1** (1970), 1–12.

[2] J. Barber, *Application of optimal prediction to molecular dynamics*, Ph.D. thesis, University of California, Berkeley, 2005.

[3] G. I. Barenblatt, *Scaling*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2003. MR 2005e:00011

[4] G. I. Barenblatt, M. Ya. Ivanov, and G. I. Shapiro, *On the structure of wave fronts in nonlinear dissipative media*, Arch. Rational Mech. Anal. **87** (1985), no. 4, 293–303. MR 86d:76024

[5] G. Benfatto and G. Gallavotti, *Renormalization group*, Physics Notes, vol. 1, Princeton University Press, Princeton, NJ, 1995. MR 97e:82001

[6] J. L. Bona and M. E. Schonbek, *Travelling-wave solutions to the Korteweg-de Vries-Burgers equation*, Proc. Roy. Soc. Edinburgh Sect. A **101** (1985), no. 3-4, 207–226. MR 87k:35208 Zbl 0594.76015

[7] A. Brandt and D. Ron, *Renormalization multigrid (rmg): Statistically optimal renormalization group flow and coarse-to-fine monte carlo acceleration*, J. Stat. Phys. **102** (2001), 231–257.

[8] L. Chen, P. Debenedetti, C. Gear, and I. Kevrekidis, *From molecular dynamics to coarse self-similar solutions: a simple example using equation-free computation*, J. Non-Newt. Fluid. Mech. **120** (2004), 215.

[9] A. J. Chorin, *Averaging and renormalization for the Korteveg-deVries-Burgers equation*, Proc. Natl. Acad. Sci. USA **100** (2003), no. 17, 9674–9679 (electr.). MR 2004f:35149

[10] ———, *Conditional expectations and renormalization*, Multiscale Model. Simul. **1** (2003), no. 1, 105–118 (electronic). MR 2004i:82043 Zbl 02139971

[11] A. J. Chorin, O. H. Hald, and R. Kupferman, *Optimal prediction and the Mori–Zwanzig representation of irreversible processes*, Proc. Natl. Acad. Sci. USA **97** (2000), no. 7, 2968–2973 (electronic). MR 2000m:82045 Zbl 0968.60036

[12] ———, *Non-Markovian optimal prediction*, Monte Carlo Methods Appl. **7** (2001), no. 1-2, 99–109. MR 2002a:65016 Zbl 0982.65011

[13] ———, *Optimal prediction with memory*, Physica D **166** (2002), no. 3-4, 239–257. Reviewed in: MR 2003e:62150 Zbl 1017.60046

[14] A. J. Chorin, O. H. Hald, and R. Kupferman, *Prediction from partial data, renormalization and averaging*, To appear in J. Sci. Comp., 2005.

[15] A. J. Chorin, A. P. Kast, and R. Kupferman, *Optimal prediction of underresolved dynamics*, Proc. Natl. Acad. Sci. USA **95** (1998), no. 8, 4094–4098 (electronic). MR 99a:65195 Zbl 0904.65117

[16] A. J. Chorin, R. Kupferman, and D. Levy, *Optimal prediction for Hamiltonian partial differential equations*, J. Comput. Phys. **162** (2000), no. 1, 267–297. MR 2001i:65137 Zbl 0960.65012

[17] A.J. Chorin and O. Hald, *Stochastic tools for mathematics and science*, Springer-Verlag, New York, 2005.

[18] D. Evans and G. Morriss, *Statistical mechanics of nonequilibrium liquids*, Academic Press, London, 1990.

[19] E. Fick and G. Sauermann, *The quantum statistics of dynamic processes*, Springer Series in Solid-State Sciences, vol. 86, Springer-Verlag, Berlin, 1990. MR 91m:82001

[20] M. E. Fisher, *Renormalization group theory: its basis and formulation in statistical physics*, Rev. Modern Phys. **70** (1998), no. 2, 653–681. MR 99e:82041

[21] D. Givon, R. Kupferman, and A. Stuart, *Extracting macroscopic dynamics: model problems and algorithms*, Nonlinearity **17** (2004), no. 6, R55–R127. MR 2097022

[22] N. Goldenfeld, *Lectures on phase transitions and the renormalization group*, Perseus Books, Reading, Mass., 1992.

[23] H. Grabert, *Projection operator techniques in nonequilibrium statistical mechanics*, Springer Tracts Modern Physics, vol. 95, Springer-Verlag, Berlin, 1982. MR 84k:82001

[24] P. Hohenberg and B. Halperin, *Theory of dynamical critical phenomena*, Rev. Mod. Phys. **49** (1977), 435–479.

[25] E. Ingerman, *Modeling the loss of information in optimal prediction*, Ph.D. thesis, University of California, Berkeley, 2003.

[26] G. Jona-Lasinio, *The renormalization group: a probabilistic view*, Nuovo Cimento B (11) **26** (1975), 99–119. MR 51 #4908

[27] W. Just, H. Kantz, C. Rödenbeck, and M. Helm, *Stochastic modelling: replacing fast degrees of freedom by noise*, J. Phys. A **34** (2001), no. 15, 3199–3213. MR 1836464

[28] L. P. Kadanoff, *Statistical physics*, World Scientific Publishing Co. Inc., River Edge, NJ, 2000. MR 2003f:82001

[29] L. D. Landau and E. M. Lifshitz, *Statistical physics, Part I*, Butterworth–Heinemann, 1980. MR 84m:82003a

[30] J. A. Langford and R. D. Moser, *Optimal LES formulations for isotropic turbulence*, J. Fluid Mech. **398** (1999), 321–346. MR 2000h:76103 Zbl 0983.76043

[31] M. Lesieur, P. Compte, and J. Zinn-Justin, *Mécanique des fluides numérique: Les houches, session lvix, 28 juin-30 juillet 1993*, North-Holland Publishing Co., Amsterdam, 1996. Reviewed in MR 97f:76001

[32] A. J. Majda, I. Timofeyev, and E. Vanden Eijnden, *A mathematical framework for stochastic climate models*, Comm. Pure and Applied Math. **54** (2001), no. 8, 891–974. MR 2002k:86012 Zbl 1017.86001

[33] H. Mori, *Transport, collective motion and brownian motion*, Prog. Theor. Phys. **33** (1965), 423–450.

[34] S. Nordholm, , and R. Zwanzig, *A systematic derivation of exact generalized Brownian motion theory*, J. Statist. Phys. **13** (1975), no. 4, 347–371. MR 56 #10694

[35] P. Okunev, *Application of optimal prediction to spin systems and to economics*, Ph.D. thesis, University of California, Berkeley, 2005.

[36] G. C. Papanicolaou, *Asymptotic analysis of stochastic equations*, Studies in probability theory, MAA Stud. Math., vol. 18, Math. Assoc. America, Washington, D.C., 1978, pp. 111–179. MR 80j:60095 Zbl 0443.60049

[37] Y. Pomeau and P. Resibois, *Time dependent correlation functions and mode-mode coupling theories*, Physics Reports C **2** (1975), 63–139.

[38] B. Seibold, *Optimal prediction in molecular dynamics*, Monte Carlo Methods Appl. **10** (2004), no. 1, 25–50. MR 2004m:81286 Zbl 02124459

[39] H. E. Stanley, *Scaling, universality and renormalization, three pillars of modern critical phenomena*, Rev. Mod. Phys. **71** (1999), S358–S366.

[40] P. Stinis, *A comparative study of two stochastic mode reduction methods*, To appear Physica D.

[41] ———, *Stochastic optimal prediction for the Kuramoto-Sivashinsky equation*, Multiscale Model. Simul. **2** (2004), no. 4, 580–612 (electronic). MR 2113171 Zbl 02148122

[42] ———, *A maximum likelihood algorithm for the estimation and renormalization of exponential densities*, J. Comp. Phys. **208** (2005), 691–703.

[43] R. Swendsen, *Monte-carlo renormalization group*, Phys. Rev. Lett. **42** (1979), 859–861.

[44] K. Theodoropoulos, Y.-H. Qian, and I.G. Kevrekidis, *"Coarse" stability and bifurcation analysis using timesteppers: a reaction diffusion example*, Proc. Natl. Acad. Sci. USA **97** (2000), 9840–9843 (electronic).

[45] R. van Zon and J. Schofield, *Mode-coupling theory for multiple-point and multiple-time correlation functions*, Phys. Rev. E (3) **65** (2002), no. 1, part 1, 011106, 17. MR 1892224

[46] R. Zwanzig, *Nonlinear generalized langevin equations*, J. Stat. Phys. **9** (1973), 215–220.

Received April 13, 2005.

ALEXANDRE J. CHORIN: chorin@math.berkeley.edu
Department of Mathematics, University of California, Berkeley CA 94720-3840, United States
http://math.berkeley.edu/∼chorin

PANAGIOTIS STINIS: pstinis@lbl.gov
Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Mail Stop 50A-1148, Berkeley, CA 94720-1148, United States