Conditional Expectations and Renormalization

Alexandre J. Chorin
Department of Mathematics
University of California
Berkeley, CA 94720

Abstract

Suppose one wants to approximate m components of an n-dimensional system of nonlinear differential equations ($m < n$) without solving the full system. In general a smaller system of m equations has right-hand-sides which depend on all of the n variables. The simplest approximation is the replacement of those right-hand-sides by their conditional expectations given the values of the m variables that are kept. It is assumed that an initial probability density of all the variables is known. This construction is first-order optimal prediction. We address here the problem of actually finding these conditional expectations in the Hamiltonian case.

We start from Hald's observation that if the full system is Hamiltonian then so is the reduced system whose right-hand-sides are conditional expectations. The relation between the Hamiltonians of the full system and of the reduced system is the same as the relation between a renormalized and a bare Hamiltonian in a renormalization group (RNG) transformation. This makes it possible to adapt a small-cell Monte-Carlo RNG method to the calculation of the conditional expectations. For the RNG the construction yields explicit forms of intermediate Hamiltonians in a sequence of renormalized Hamiltonians (the "parameter flow"). A spin system is used to illustrate the ideas.

Key words: Conditional expectations, optimal prediction, renormalization, parameter flow, critical exponents, spins, averaging.

1 Introduction

Consider a system of time-dependent differential equations (written as ordinary differential equations for simplicity) of the form

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

where ϕ, R and x are n-vectors with components ϕ_i, R_i, x_i, $i = 1, \ldots, n$ and t is the time. Suppose we can afford only to solve m equations (for example, because the system is too large to be readily solved or because we only have m initial data). the corresponding right-hand sides still depend on all the variables; in first-order optimal prediction (OP) methods one replaces these right-hand-sides by their best least squares approximations by functions of the m variables that are kept [7],[5],[8],[4]. The best least-square approximations are conditional expectations. In the present paper we address the often difficult problem of actually finding these best least-squares approximations in the case of Hamiltonian systems.

We note that first-order optimal prediction is a valid approximation only for short or moderate times; for longer times one has to take into account additional, usually non-Markovian, terms [9],[15],[6]. The computational issues relevant to these additional terms will be dealt with in subsequent papers.

Renormalization group (RNG) transformations [3],[12],[13] reduce the dimensionality of a system of equations by integrating out unwanted scales. When one performs this reduction one obtains a new Hamiltonian. That there is a qualitative resemblance between OP and RNG methods is quite clear, and has been pointed out in particular by Goldenfeld et al. [10],[11]. In the present paper we show that in the case of Hamiltonian systems the RNG transformations of the Hamiltonian can be obtained by integrating conditional expectations of the right-hand-sides of the equations of motion; loosely speaking, RNG transformations are integrals of OP reductions. This remark makes it possible to use simple Monte-Carlo RNG methods [3],[14] to evaluate both the coefficients of the new Hamiltonians in RNG transformations (the "RNG parameter flow") and the right-hand-sides of reduced systems of equations. To illustrate the construction, we apply it to spin systems and obtain explicitly the RNG parameter flows in addition to critical points, critical exponents, and order parameters. We exhibit in detail a particular implementation that is a little awkward if viewed as an instance of a RNG but is particularly convenient for OP.

In fact what is offered in the present paper is a numerical short-cut. Suppose $x = (x_1, x_2, \ldots)$ is a set of n random variables and let $m < n$; partition x so that $x = (\hat{x}, \tilde{x})$, $\hat{x} = (x_1, x_2, \ldots, x_m)$, $\tilde{x} = (x_{m+1}, x_{m+2}, \ldots)$. Let $p = p(x)$ be the joint probability density of all the variables, and consider the problem of finding a function $\hat{H} = \hat{H}(\hat{x})$ such that

$$\exp(-\hat{H}(\hat{x})) = \int p(x) d\tilde{x}, \tag{2}$$

where $d\tilde{x} = dx_{m+1} dx_{m+2} \cdots$. There is no question that $\hat{H}$ is well defined, at least when n is finite, but the obvious ways of finding it can be costly. We are offering effective ways to do so. There are other situations where one wants to integrate out unwanted variables inside nonlinear functions, and our short-cut may serve there as well; in subsequent papers we shall apply it to problems in irreversible statistical mechanics and, equivalently, to problems involving the full long-time OP equations.

The problem of evaluating the right-hand-sides in OP is an example of the "curse of circularity" in the formulation of simplified approximations of complex problems: One can find very good approximations if one knows the full solutions. The present paper is an attempt to lift part of the curse.

2 Conditional expectations and optimal prediction

Consider a set x of random variables $(x_1, x_2, \ldots, x_n)$ with a joint probability density of the form $Z^{-1} e^{-H(x)}$, $Z = \int e^{-H(x)} dx$, $dx = dx_1 dx_2 \ldots dx_n$. Consider the space L_2 of function $u(x), v(x), \ldots$, with the inner product $\langle u, v \rangle = E[uv] = \int u(x) v(x) Z^{-1} \exp(-H) dx$, where $E[\cdot]$ denotes an expected value.

Partition the variables into two groups as above, $x = (\hat{x}, \tilde{x})$, $\hat{x} = (x_1, \ldots, x_m)$, $m < n$. Given a function $f(x)$, its conditional expectation given $\hat{x}$ is

$$E[f(x)|\hat{x}] = \frac{\displaystyle\int f(x) e^{-H(x)} d\tilde{x}}{\displaystyle\int e^{-H(x)} d\tilde{x}}; \tag{3}$$

it is the average of f keeping $\hat{x}$ fixed. The conditional expectation is a function of $\hat{x}$ only, and it is the best approximation of f in the mean square sense by a function of $\hat{x}$:

$$E\left[(f(x) - E[f(x)|\hat{x}])^2\right] \le E\left[(f(x) - h(\hat{x}))^2\right] \tag{4}$$

for any function $h = h(\hat{x})$. $E[f|\hat{x}]$ is the orthogonal projection of f onto the subspace $\hat{L}_2$ of L that contains functions of $\hat{x}$ only. $E[f(x)|\hat{x}]$ can be approximated by expansion in a basis of $\hat{L}_2$; keeping only a suitable finite number ℓ of basis functions $\psi_1(\hat{x}), \psi_2(\hat{x}), \ldots, \psi_\ell(\hat{x})$, and minimizing the distance between f and the span of the $\psi_i(\hat{x})$, one finds

$$E[f|\hat{x}] = \sum_{i=1}^{\ell} c_i \psi_i(\hat{x}),$$

where the c_i satisfy the equation

$$\Phi c = r, \tag{5}$$

where Φ is the matrix with elements $\Phi_{ij} = \langle \psi_i, \psi_j \rangle$, $c = (c_1, \ldots, c_\ell)$, and $r = (\langle f, \psi_1 \rangle, \langle f, \psi_2 \rangle \ldots, \langle f, \psi_\ell \rangle)$. Usually the inner products can be calculated by Metropolis sampling, but note that they require integrations over the full space of n variables.

Suppose you want to find a function $\hat{H} = \hat{H}(\hat{x})$ such that

$$e^{-\hat{H}(\hat{x})} = \int e^{-H(\hat{x},\tilde{x})} d\tilde{x}, \tag{6}$$

i.e., write the marginal probability density of the variables $\hat{x}$ in exponential form. Then

$$\begin{aligned} E\left[\frac{\partial}{\partial x_i} H(x) | \hat{x}\right] &= \frac{\int \frac{\partial}{\partial x_i} H(x) e^{-H(x)} d\tilde{x}}{\int e^{-H(x)} d\tilde{x}} \\ &= \frac{\partial}{\partial x_i} \left(-\log \int e^{-H(x)} d\tilde{x}\right). \end{aligned} \tag{7}$$

An analogous relation between the derivative of a logarithm of a partially integrated density and a conditional expectation arises also in the context of expectation-maximization is statistics [1].

If one can find a basis for $\hat{L}_2$ consisting of functions of the form $\frac{\partial}{\partial x_1}\psi_j(\hat{x})$, $j = 1, \ldots,$ and provided the set of variables is homogenous so that for all $i \leq m$ the coefficients c_j in the expansions $\left[\frac{\partial}{\partial x_i} H|\hat{x}\right] = \sum_{j=1}^{\ell} c_j \frac{\partial}{\partial x_i}\psi_j(\hat{x})$ are independent of i, then the expansion

$$\hat{H}(\hat{x}) = \sum c_j \psi_j(\hat{x}). \tag{8}$$

follows immediately. This is our key observation. Note that $\hat{H}$ and its derivatives are calculated simultaneously and that we differentiated ψ_j with respect to only one variable x_1 because of the assumed homogeneity.

The relation in equation (6) arises in OP as well (this is Hald's theorem [5]). Consider again the system of differential equations (1):

$$\frac{d}{dt}\phi(t) = R(\phi(t)), \quad \phi(0) = x.$$

Suppose we partition as above $\phi = (\hat{\phi}, \tilde{\phi})$, $R = (\hat{R}, \tilde{R})$, where $\hat{\phi}$ contains the first m components of ϕ, etc. Suppose the system (8) is Hamiltonian, i.e., m, n are even, $R_i = \frac{\partial}{\partial x_{i-1}} H$ for i even, $R_i = -\frac{\partial}{\partial x_{i+1}} H$ for i odd; $H = H(x)$ is the Hamiltonian and $Z^{-1}e^{-H}$ is then an invariant probability density for the system.

Suppose we can afford to solve only $m < n$ of the equations in (1) or have only m data components $\hat{x}$ (for example, the variables ϕ may be Fourier components in the solution of a partial differential equation, and we may be able to keep only some of these components and have to discard the rest but also want to take into account the effect of the components that have been discarded on those that remain). We thus want to solve the equations for $\hat{\phi}$:

$$\frac{d\hat{\phi}}{dt} = \hat{R}(\phi), \quad \hat{\phi}(0) = \hat{x},$$

where $i \leq m$, but the argument of $\hat{R}$ is the whole vector ϕ. Assume that initially the variables that are not computed are sampled from the invariant distribution. It is natural to approximate $\hat{R}_i(\phi)$ by the closest function of $\hat{\phi}$ for each $i \leq m$, which we now know is the conditional expectation of ϕ given $\hat{\phi}$; this produces the approximating system:

$$\frac{d\hat{\phi}}{dt} = E[\hat{R}(\phi) \mid \hat{\phi}]. \tag{9}$$

The approximation (9) is valid only for a short time, as one can see from the full equation for the evolution of $\frac{d\hat{\phi}}{dt}$ in [5],[6], see also below. Hald's theorem asserts that the system (9) is Hamiltonian, with Hamiltonian $\hat{H} = \hat{H}(\hat{x}) = -\log \int e^{-H(x)} d\tilde{x}$, a relation identical to equation (6); equation (6) also contains the proof of Hald's theorem. The existence of $\hat{H}$ shows that the approximation (9) cannot be valid for long times: the predictive power of partial initial data decays at $t \to \infty$ for a nonlinear system, and the best estimate of $\hat{\phi}(t)$ should decay to unconditional mean of ϕ (which is usually zero). The existence of a reduced Hamiltonian shows that this decay can happen only to a limited extent and thus the approximation can in general be valid only for short times. Equations (9) constitute the short time, or "first-order", OP approximation.

Suppose however that instead of picking specific values for the initial data $\hat{x}$ one samples them from the invariant density $Z^{-1}e^{-\hat{H}(\hat{x})}$. The distribution of the $\hat{x}$'s is then invariant, and equal to their marginal distribution in the full system (8) when the data are sampled from the invariant distribution $Z^{-1}e^{-H(x)}$, as one can also see from the identities $\exp(-\hat{H}) = \exp\left(\log \int e^{-H} d\tilde{x}\right) = \int e^{-H} d\tilde{x}$. The system (9) then generates the marginal probability density of part of the variables of a system at equilibrium. Thus OP at equilibrium is a way of reducing the number of variables without affecting the statistics of the variables that remain. One can make short-time predictions about the future from the reduced system with coefficients computed at equilibrium because it is self-consistent to assume for short times that unresolved degrees of freedom are in thermal equilibrium, as is explained in the OP papers cited above.

3 Renormalization

For simplicity, we work here with real-space renormalization applied to variables associated with specific sites in a plane, $x^{(1)} = (x_{I_1}, x_{I_2}, \ldots)$, where $I_k = (i_k, j_k)$, i_k, j_k are integers, all the I_k are inside a square D of side N with N large, and the variables are listed in some convenient order. The Hamiltonian $H = H^{(1)}$ is a function of $x^{(1)}$, $H^{(1)} = H^{(1)}(x^{(1)})$. The need for the superscript (1) will appear shortly. We assume that the partition function $Z = \int e^{-H^{(1)}(x^{(1)})} dx^{(1)}$ is well defined, where $dx^{(1)} = dx^{(1)}_{I_1} dx^{(1)}_{I_2} \ldots$.

Suppose we group the variables $x_{I_1}, x_{I_2}, \ldots$ into groups of ℓ variables (for example, we could divide D into squares each containing 4 variables). The variables can be referred to as "spins"in conformity with common usage in physics. Associate with each group a new variable $x^{(2)}_{J_1}, x^{(2)}_{J_2}, \ldots$, where $J_1, J_2, \ldots$ is some ordering of the new variables and $x^{(2)}_{J_k}$ is a function (not necessarily invertible) of the $x^{(2)}_{I}$ in the group labeled by J_k, for example $x^{(2)}_{J_k} = g(x^{(1)}_{I_{m+1}}, x^{(1)}_{I_{m+2}}, \ldots, x^{(1)}_{I_{m+\ell}})$ for the appropriate m. The vector $x^{(2)}$ is $x^{(2)} = (x^{(2)}_{J_1}, x^{(2)}_{J_2} \ldots)$. We can write

$$\begin{aligned} Z &= \int e^{-H^{(1)}(x^{(1)})} dx^{(1)} \\ &= \int dx^{(2)} \int \delta\left(x^{(2)} - g(x^{(1)})\right) e^{-H^{(1)}(x^{(1)})} dx^{(1)} \end{aligned}.$$

where $dx^{(2)} = dx^{(2)}_{J_1} dx^{(2)}_{J_2} \cdots$, and the δ function is a product of delta functions, one per group. If one defines $H^{(2)}(x^{(2)})$ by the equation

$$e^{-H^{(2)}(x^{(2)})} = \int \delta\left(x^{(2)} - g(x^{(1)})\right) e^{-H(x^{(1)})} dx^{(1)}, \tag{10}$$

then $Z = \int e^{-H^{(2)}(x^{(2)})} dx^{(2)}$.

The mapping $x^{(1)} \to x^{(2)}$, followed by a change of numbering of the remaining variables so that $J_1, J_2 \ldots$ (the indices of the new variables $x^{(2)}$) enumerate the new variables by going through all integer pairs in a reduced domain of side $N/\sqrt{\ell}$, is a real-space renormalization group transformation; it produces a new set of variables which has less spatial detail than the previous set and such that distances between the remaining spins have been scaled down by $\sqrt{\ell}$. If the calculation is set up so that the mapping $x^{(1)} \to x^{(2)}$, $H^{(1)} \to H^{(2)}$ can be repeated, for example, if the range of the variables $x^{(1)}$ is invariant and the Hamiltonians $H^{(1)}, H^{(2)}$ can be represented in the same finite-dimensional basis, then one can produce in this way a sequence of Hamiltonians $H^{(1)}, H^{(2)}, H^{(3)}, \ldots$; the fixed points of the transformation $H^{(n)} \to H^{(n+1)}$ for a spin system of infinite spatial extent include the critical points of the system, see any discussion of the RNG, for example [12],[13].

Consider the special case where $x_J^{(2)}$ is one of the $x_I^{(1)}$ in its group–i.e., replace a block of spins by one of the spins in the block. More general and widely used assignments of block variables will not be needed in the present paper and will be discussed elsewhere (their OP analogues are problems where the variables one keeps are not necessarily the first entries in the list of components). We can identify the spins that remain with $\hat{x}$ of the preceding section and the spins that disappear with $\tilde{x}$. Equation (10) becomes a special case of equation (6), and can be solved for $H^{(2)}$ by taking conditional expectations of the derivatives of $H^{(1)}$.

Note that the usual RNG representation of a renormalized Hamiltonian by means of additional couplings [12] is interpreted here as an expansion of a conditional expectation in a convergent series. The new interpretation may be useful both in understanding what is happening on the computer and in deriving error estimates. The relation between the RNG and conditional expectations shows that the latter can be calculated recursively, as we show in the example below.

We have written the RNG transformation above in notation suitable for spins with a continuous range. The case of discrete (e.g., Ising) spins is automatically included, even though it may seem odd to differentiate functions with a discrete domain and range. Indeed, add to the Hamiltonian H a term of the form

$$\frac{1}{\varepsilon}\sum_i \left(\prod_j G(x_i - x_{0j})\right)$$

where ε is small, the sum is over all spins, the product is over a finite number of values x_{0j}, and $G \geq 0$ has a minimum at 0 and is positive elsewhere. For small ε such a term will constrain the x_i to take on the values x_{0j}, but since at the origin the derivative of G is zero the calculation of the conditional expectations is unaffected by this term and the limit $\varepsilon \to 0$ can be taken without actually doing anything on the computer. All one has to do is make sure that in the Monte-Carlo sampling only the values x_{0j} are sampled. Indeed, results below will be given for Ising spins which take on the values $+1$ and -1, with a "bare" (unrenormalized) Hamiltonian $H^{(1)} = \beta \sum x_I x_J$, with summation over locations

I, J that are neighbors on the lattice; $\beta = 1/T$, where T is the temperature.

4 A decimation RNG/OP scheme for a spin system

We consider in detail a RNG/recursive OP scheme where the number of variables is halved at each step. The spins are located on a square lattice with nodes $I_k = (i_k, j_k)$, i_k, j_k integers, and at each step of the recursion those for which $i_k + j_k$ is odd are eliminated while those for which $i_k + j_k$ is even are kept. The spins with $i_k + j_k$ even constitute $\hat{x}$ and the others $\tilde{x}$; the choice of which are even and which are odd is a matter of convention only (see Figure 1). The variables are labeled by $I_k : x_{I_1}, x_{I_2}, \ldots$.

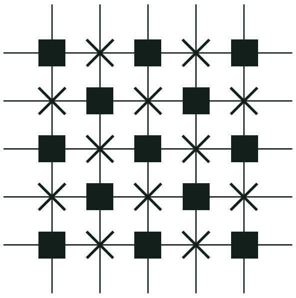

Figure 1: The decimation pattern

Given a location $I = (i, j)$, we group the other variables according to their distance from I: group 1 contains only x_I, the variable at I. Group 2 (relative to I) contains those variables whose distance from I is 1, group 3 contains those variables whose distance to I is $\sqrt{2}$, etc. We form the "collective" variables

$$X_{k,I} = \frac{1}{n_k} \sum_{\text{group } k} x_J$$

where n_k is the number of variables in the group (1 for group 1, 4 for group 2, etc.). From these variables one can form a variety of translation-invariant polynomials in x of various degrees: $\sum_I x_I X_{k,I} = \sum_I X_{1,I} X_{k,I}$, $\sum_I (X_{k,I})^2 (X_{k+1,I})^2$, $\sum_I (X_{k,I})^4, \ldots$. In practice the domain over which one sums must be finite, and it is natural to impose periodic boundary conditions at its edges to preserve the translation invariance. We wrote out explicitly only polynomials of even degrees because the Hamiltonians we consider are invariant under the transformation $x \to -x$. The translation-invariant polynomials built up from the $X_{k,I}$ can be labeled $\psi_1(x), \psi_2(x), \ldots$ in some order.

Expand the n-th renormalized Hamiltonian in a series and keep the first ℓ terms:

$$H^{(n)} = \sum_{k=1}^{\ell} \alpha_k^{(n)} \psi_k(x). \tag{11}$$

The derivative of this series at the spin x_I is

$$\frac{\partial}{\partial x_I} H^{(n)} = \sum_{1}^{\ell} \alpha_k^{(n)} \psi_k'(x), \qquad \psi_k' = \frac{\partial}{\partial x_I}\psi_k. \tag{12}$$

The functions ψ_k' are easily evaluated, for example:

$$\left(\sum_J x_J X_{k,J}\right)' = 2X_{k,I}$$

$$\left(\sum_J (X_{k,J})^4\right)' = 4\left(\sum_{\text{group } k} x_J^3\right)/n_k^2,$$

etc., where "group k" refers to distances from I, the variable with respect to which we are differentiating (see Figure 2). The use of these basis functions is equivalent to the physicists' expansion in successive couplings; One may think of looking for other bases that may lead to faster convergence of the conditional expectations, but the physicists' basis has an intriguing property: The expansions performed on the computer are in terms of the derivatives of the basis functions, and in the present case these derivatives are local, i.e., each derivative vanishes except on a small number of neighboring points. This may well lead to ways of breaking up the inner products into sums of inner products on spaces with smaller dimensions as will be explored in subsequent papers.

Pick a variable x_I in $\hat{x}$ (i.e., $I = (i,j)$, $i+j$ even in our conventions). Some of the functions ψ_k' in (12) will be functions of $\hat{x}$ only and some will be functions of both $\hat{x}$ and $\tilde{x}$ or of $\tilde{x}$ only. The task at hand is to project the latter on the former and then rearrange the series so as to shrink the scale of the physical domain. To explain the construction we consider a very special case.

Suppose one can write

$$H^{(n)}(x) = \alpha_2^{(n)}\psi_2 + \alpha_3^{(n)}\psi_3 + \alpha_4^{(n)}\psi_4, \tag{13}$$

where $\psi_k(x) = \sum x_I X_{k,I}$ for $k = 2, 3, 4, \ldots$. Note that $\psi_k' = \frac{\partial}{\partial x_I}\psi_k$ is a function only of $\hat{x}$ when $k = 3, 4$ (and when $k = 6$, as we shall need to know shortly) but not when $k = 2$ or 5 (see Figures 1, 2). We now calculate the conditional expectations of the derivatives of $H^{(n)}$ given $\hat{x}$ by projecting them on the space of functions of $\hat{x}$. First we project ψ_2' on the span of $\psi_3', \psi_4', \psi_6'$

(note that ψ_6' is not in the original expansion (13)). Form the matrix Φ with rows $\langle\psi_k', \psi_3'\rangle, \langle\psi_k', \psi_4'\rangle, \langle\psi_k', \psi_6'\rangle$ for $k = 3, 4, 6$, the primes once again denoting differentiation with respect to x_I. Form the vector r with component $(\langle\psi_2', \psi_3'\rangle, \langle\psi_2', \psi_4'\rangle, \langle\psi_2', \psi_6'\rangle)$. Let $c = (c_1, c_2, c_3)$ be the solution of $\Phi c = r$ (see equation (4)). The coefficients c are the coefficients of the orthogonal projection of ψ_2' onto the span of $\psi_3', \psi_4', \psi_6'$ which is contained in $\hat{L}_2$. After projection, the coefficients of ψ_3, ψ_4 in (13) become

$$\alpha_3^{\text{new}} = \alpha_3^{(n)} + \alpha_2^{(n)} c_1,$$

$$\alpha_4^{\text{new}} = \alpha_4^{(n)} + \alpha_2^{(n)} c_2,$$

and ψ_6 acquires the coefficient $\alpha_2^{(n)} c_3$.

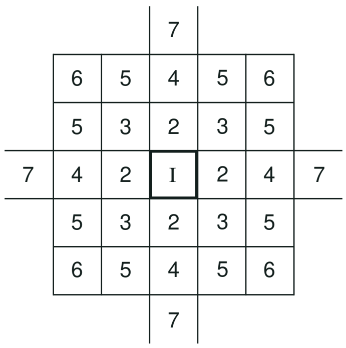

Figure 2: The collective variables

If one relabels the remaining spins so that they occupy the lattice previously occupied by all the spins, group 3 becomes group 2, group 4 becomes group 3, and group 6 becomes group 4 (see Figure 2). The new Hamiltonian $H^{(n+1)}$ now has the representation

$$H^{(n+1)} = \alpha_2^{(n+1)} \psi_2 + \alpha_3^{(n+1)} \psi_3 + \alpha_4^{(n+1)} \psi_4,$$

with

$$\begin{aligned} \alpha_2^{(n+1)} &= \alpha_3^{(n)} + \alpha_2^{(n)} c_1, \\ \alpha_3^{(n+1)} &= \alpha_4^{(n)} + \alpha_2^{(n)} c_2, \\ \alpha_4^{(n+1)} &= \alpha_2^{(n)} c_3. \end{aligned}$$

More generally, if $H^{(n)}$ is expressed as a truncated series, partition the terms in the series for $\frac{\partial}{\partial x_I} H^{(n)}$ into functions of $\hat{x}$ and functions of both $\hat{x}$ and $\tilde{x}$. Add to the terms which are functions of $\hat{x}$ additional terms which are functions of $\hat{x}$

and are chosen so that after relabelling they acquire the form of terms already in the series (just as above, terms that depend on $X_{3,J}$, for example, become terms that depend on $X_{2,J}$ after relabelling). Project the functions of x on the span of the expanded set of functions of $\hat{x}$, collect terms and relabel. This is a renormalization step, and it can be repeated.

Note that it if one wants to reduce the number of variables by a given factor, one can in principle use an analogous RNG/conditional expectation construction and get there in one iteration; the recursive construction is easier to do and the intermediate Hamiltonians, whose coefficients constitute the parameter flow in the renormalization, contain useful information.

The discussion so far may suggest that one sample the Hamiltonians recursively, i.e., start with $H^{(1)}$, find $H^{(2)}$, use Monte-Carlo to sample the density $Z^{-1}e^{-H^{(2)}}$ and find $H^{(3)}$ etc. The disadvantages of this approach are: (i) The sampling of the densities $Z^{-1}e^{-H^{(n)}}$ can be much more expensive for $n > 1$ than for $n = 1$ because each proposed Monte-Carlo move may require that the full series for $\frac{\partial}{\partial x_I}H^{(n)}$ be summed twice; and (ii) each evaluation of a new Hamiltonian is only approximate because the series are truncated, and, more important, the Monte-Carlo evaluation of the coefficients may have limited accuracy. These errors accumulate from step to step and may produce false fixed points and other artifacts.

The remedy lies in Swendsen's observation [3],[14] that the successive Hamiltonians can be sampled without being known explicitly. Sample the original Hamiltonian, remove the unwanted spins and relabel the remaining spins so as to cover the original lattice, as in the relabelling step in the renormalization; the probability density of the remaining spins is $Z^{-1}e^{-H^{(2)}}$; repeating n times yields samples of $Z^{-1}e^{-H^{(n+1)}}$. The price one pays is that to get an m by m sample of $Z^{-1}e^{-H^{(n)}}$ one has to start by sampling a $2^q m$ by $2^q m$ array of non-renormalized spins, where q is either $(n+1)/2$ or $n/2$ depending on the parity of n and on programming choices; the trade-off is in general very worthwhile. What has been added to Swendsen's calculation is an effective evaluation of the coefficients of the expansion of $H^{(n)}$ from the samples.

The programming here requires some care. With the decimation scheme as in Figure 1, after one removes the unwanted spins in $x^{(n)}$, the remaining spins, the variables $x^{(n+1)}$, live on a lattice with a mesh size $\sqrt{2}$ larger than before. To avoid sampling unnecessarily large arrays one wants to extract from this set of spins the largest square box that is inscribed in the previous array, and one finds by a simple picture that this largest box is half the size of the previous one and is rotated by an angle of $\pi/4$ from the previous box. At the next renormalization one needs $x^{(n+2)}$ which can be extracted from $x^{(n)}$ (not from $x^{(n+1)}$) by taking one spin in four and the resulting box size is the same as the size of the box of $x^{(n+1)}$ and is not rotated. One may worry a little about boundary conditions for $x^{(n+1)}$: At the edges of the boxes one normally imposes periodic boundary conditions; the rotated boxes are periodic but the periods are also rotated, thus the periodicity of $x^{(n)}$ is not the same as the periodicity one has to assume for $x^{(n+1)}$; the inconsistency leads to errors too small to be detected in our calculations.

5 Some numerical results

We now present some numerical results obtained with the RNG/conditional expectation scheme. The problem we apply the construction to is Ising spins; more interesting applications will be presented elsewhere. The point being made is that the construction can be effectively implemented. The results presented are for Ising spins.

In table I we list the coefficients $\alpha_k^{(n)}$ in the expansion of $H^{(n)}$ for $n = 1, \ldots 7$ and $T = 2.27$. The functions ψ_k are as follows:

$$\psi_k = \sum x_J X_{k,J} \quad \text{for } k = 1, 2, 3, 4, 5, 6$$

$$\psi_{6+k} = \sum (X_{k+1,J})^4, \quad \text{for } k = 1, 2, 3$$

$$\psi_{10} = \sum X_{2,J}^2 X_{3,J}^2.$$

Note that as a result of the numbering of the ψ's the last coefficient is not necessarily the smallest coefficient. This table represents the parameter flow and if the functions ψ_k are written in terms of the variables x_J the table defines the new system of equations for the reduced set of variables. Remember that in the projection on $\hat{L}_2$ additional functions are used so that after relabelling the series has the same terms , but maybe with different coefficients, as before the renormalization. In $H^{(1)}$, α_2 is the sole non-zero coefficient, and its value is determined by T and the definition of $X_{2,J}$, in particular the presence of the coefficient n_2 (see above).

It is instructive to use the parameter flow to identify the critical temperature T_c. For $T < T_c$ the renormalization couples ever more distant spins while for $T > T_c$ the spins become increasingly decoupled. One can measure the increasing or decreasing coupling by considering the quadratic terms in the Hamiltonian (the terms of the form $\sum x_J X_{k,J}$) and calculating the "second moments" M_2 of their coefficients $\alpha_k^{(n)}$:

$$M_2^{(n)} = \sum_{k=2}^{\ell} d_k^2 \alpha_k^{(n)}$$

where d_k is the distance from J of the spins in the group k (see the definition of $X_{k,J}$), $\alpha_k^{(n)}$ is the coefficient of $\sum x_J X_{k,J}$ in the expansion of $H^{(n)}$, and ℓ is the number of quadratic terms in this expansion. In Figure 3 we show the evolution of $M_2^{(n)}$ with n for various values of T (with $\ell = 5$ and 7 functions over-all in the expansion, including non-quadratic functions).

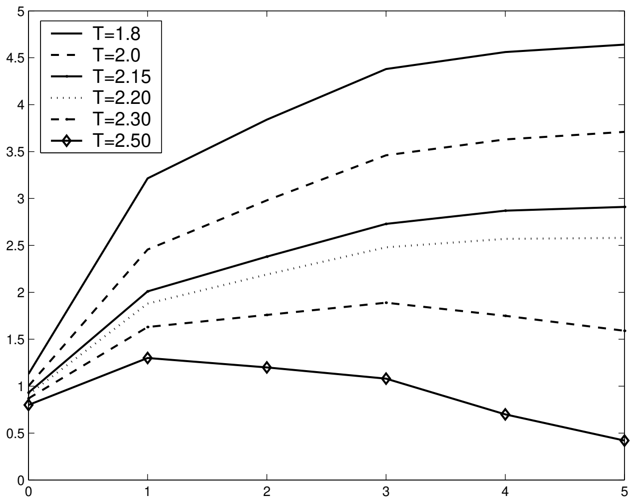

Figure 3: Second moments of the coefficients of the renormalized Hamiltonian for various values of T for successive iterations

In Figure 4 we show the evolution of $M_2^{(n)}$ near $T_c = 2.269\ldots$ with $\ell = 6$ and 10 terms in the expansion. The non-uniform behavior of M_2 is not a surprise (it is related to the non-uniform convergence of critical exponents already observed by Swendsen). Each step in the renormalization used 10^5 Monte-Carlo steps per spin. ¿From these graphs one would conclude that $T_c \sim 2.26$, an error of .5%. The accuracy depends on the number of terms in the expansion and on the choice of terms; with only 6 terms (4 quadratic and 2 quartic), the error in the location of T_c increases to about 3%. The point is not that this is a good way to find T_c but that it is a check on the accuracy of the parameter flow. From the Table one can see that the system first approaches the neighborhood of a fixed point and then diverges from it, as one should expect in a discrete sequence of transformations.

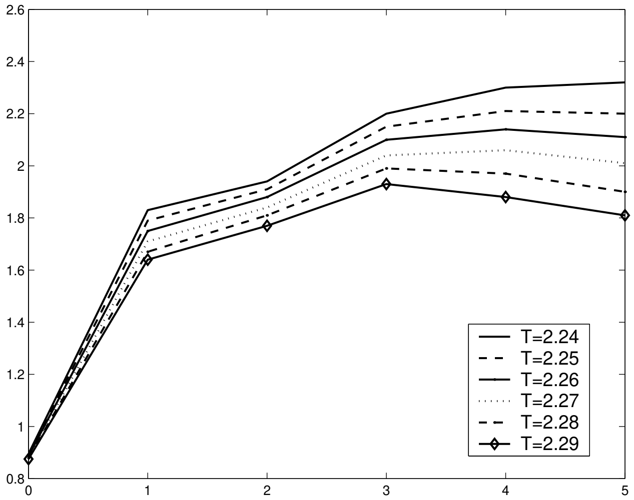

Figure 4: Second moments of the coefficients of the renormalized Hamiltonian near T_c

Table 1
Parameter flow for the Ising model $T = 2.26$, 10 basis functions

iteration	1	2	3	4	5	6	7
α_1	0	.26	.35	.44	.48	.52	.54
α_2	.893	.47	.47	.35	.30	.25	.21
α_3	0	.32	.20	.23	.21	.20	.18
α_4	0	.04	.08	.11	.12	.13	.13
α_5	0	.07	.11	.13	.13	.12	.12
α_6	0	$-.01$	.01	.01	.02	.03	.02
α_7	0	$-.08$	$-.07$	$-.10$	$-.09$	$-.09$	$-.08$
α_8	0	.04	.02	.02	.01	.00	$-.10$
α_9	0	$-.00$	$-.01$	$-.00$	$-.00$	.00	.00
α_{10}	0	$-.12$	$-.17$	$-.18$	$-.18$	$-.17$	$-.16$

We now use the renormalized system to calculate the magnetization $m = E[\sum x_I/n^2]$. To get the correct non-zero m for $T < T_c$ on a small lattice the symmetry must be broken, and we do this by imposing on all the arrays the boundary condition $x_{boundary} = 1$ rather than the periodic boundary conditions used elsewhere in this paper. In Figure 5 we display m computer with the bare (unrenormalized) Hamiltonian $H^{(1)}$ on 3 lattices: 20 by 20, 40 by 40, 60 by 60, as well as the results obtained on a 20 by 20 lattice by sampling the density defined by the renormalized Hamiltonian $H^{(5)}$ which corresponds in principle

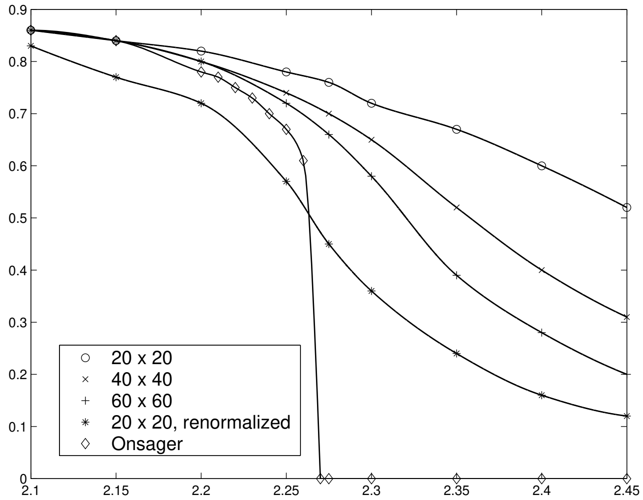

Figure 5: Bare and renormalized magnetization near T_c

to an 80 by 80 bare calculation. We also display the exact Onsager values of m. The calculations focus on values of T in the neighborhood of T_c where the size of the lattice matters; one cannot expect the results to agree perfectly with the Onsager results on a finite lattice with periodic boundary conditions for any n; all one can expect is to have the values of the small renormalized calculation be consistent with results of a larger bare calculation. We observe that they do, up to the shift in T_c already pointed out and due to the choice of basis functions.

The determination of the critical exponents for a spin model is independent of the determination of the coefficients in the expansion of $H^{(n)}$, and is mentioned here only because it does provide a sanity check on the constructions, in particular on the adequacy of the basis functions. For comparable earlier calculations, see in particular Swendsen's chapter in [3]. As is well known, if A is the matrix of derivatives $\partial \alpha_i^{(n+1)}/\partial \alpha_j^{(n)}$ at $T = T_c$, those of its eigenvalues that are larger than 1 are the critical exponents of the spin system [12]. The matrix A can be found from the chain rule [2],[3]

$$\frac{\partial}{\partial \alpha_j^{(n)}} E\left[\psi_k\left(x^{(n+1)}\right)\right] = \sum_i \frac{\partial \alpha_i^{(n+1)}}{\partial \alpha_j^{(n)}} \frac{\partial E[\psi_k(x^{(n+1)})]}{\partial \alpha_i^{(n+1)}}$$

and the sum is over all the coefficients that enter the expansion. The derivatives of the expectations are given by correlations as follows:

$$\frac{\partial E[\psi_k(x^{(n+1)})]}{\partial \alpha_j^{(n)}} = E\left[\psi_k(x^{(n+1)})\psi_j(x^{(n)})\right] - E[\psi_k(x^{(n+1)})]E[\psi_j(x^{(n)})],$$

$$\frac{\partial E[\psi_k(x^{(n+1)})]}{\partial \alpha_i^{(n+1)}} = E\left[\psi_k(x^{(n+1)})\psi_i(x^{(n+1)})\right] - E[\psi_k(x^{(n+1)})]E[\psi_i(x^{(n+1)})],$$

see [3]. In most of the literature on real-space renormalization for Ising spins the variables $x^{(n+1)}$ are obtained from $x^{(n)}$ by "majority rule", i.e., by assigning to the group that defines $x^{(n+1)}$ the value $+1$ if most of the members of the group are $+1$, the value -1 if most of the members of the group are -1, with ties resolved at random. For the decimation scheme described above our "pick one" rule ($x^{(n+1)}$ is one of the members of the group) is identical to the majority rule. There is an apparent difficulty in the decimation because at each recursion the number of terms in the summation that defines the basis functions is reduced by half while the square root of an integer is not in general an integer, so that one has to perform Swendsen sampling on rectangles so designed that the ratio of the areas of two successive rectangles is $1/2$. This has not turned out to be harmful, and the value of ν, the correlation exponent, was found to be 1 (the exact value) $\pm.01$ with 10^6 Monte-Carlo moves per spin, the error depending mainly on the number of Monte-Carlo moves which has to be very large, in line with previous experience [14]. We also checked that in a renormalization scheme where a 2×2 block of spins is replaced at each iteration by a single spin, the "majority rule" and our "pick one" rule for $x^{(n+1)}$ yield similar results. One needs fewer terms in the expansion of the Hamiltonian to get accurate values of the exponents than to get an accurate parameter flow, but a larger number of Monte-Carlo moves.

6 Conclusions

We have presented a simple relation between conditional expectations and the RNG, and used it to adapt Monte-Carlo RNG methods to the calculation of the conditional expectations that arise in the reduction of the dimensionality of systems of differential equations. We illustrated the methods on examples involving spins. In the case of spins, the numerical results emphasize the neighborhood of the critical point because this is where the variables are strongly coupled without separation of scales and a reduction in system size requires non-trivial tools. There may be a possibility to further simplify the calculations by exploiting the fact that the derivatives of the basis functions in the calculation of the conditional expectations are local, i.e., are non-zero only on a few neighboring points. We plan to apply these ideas to problems where a more accurate representation of the time dependence is needed, for example to underresolved finite-difference approximations of time-dependent partial differential equations, as was already suggested by Goldenfeld [10].

Acknowledgments. I would like to thank Prof. G.I. Barenblatt, Prof. N. Goldenfeld, Prof. O. Hald, Prof. R. Kupferman, Mr. K. Lin, and Mr. P. Stinis for very helpful discussions and comments, and the referees for useful remarks. This work was supported in part by the Office of Science, Office of Advanced Scientific Computing Research, Mathematical, Information, and Computational Sciences Division, Applied Mathematical Sciences Subprogram, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 and in part by the National Science Foundation under grant number DMS89-19074.

References

[1] P. Bickel and K. Doksum, Mathematical Statistics: Basic Ideas and Selected Topics, Prentice-Hall, New York, 2000, p. 133 and ff.

[2] J. Binney, N. Dowrick, A. Fisher, and M. Newman, The Theory of Critical Phenomena, The Clarendon Press, Oxford, 1992.

[3] T. Burkhardt and J. van Leeuwen, Real-Space Renormalization, Springer, Berlin, 1982.

[4] A. Chorin, Stochastic Methods in Applied Mathematics and Physics, Lecture notes, UC Berkeley Math. Dept., 2002.

[5] A. Chorin, O. Hald and R. Kupferman, Optimal prediction and the Mori-Zwanzig representation of irreversible processes. Proc. Nat. Acad. Sc. USA, 97, (2000), pp. 2968–2973.

[6] A. Chorin, O. Hald and R. Kupferman, Optimal prediction with memory, Physica D, 166, (2002), pp. 239-257.

[7] A. Chorin, A. Kast and R. Kupferman, Optimal prediction of underresolved dynamics, Proc. Nat. Acad. Sc. USA, 95 (1998), pp. 4094–4098.

[8] A. Chorin, R. Kupferman and D. Levy, Optimal prediction for Hamiltonian partial differential equations, J. Comput. Phys., 162, (2000), pp. 267–297.

[9] D. Evans and G. Morriss, Statistical Mechanics of Nonequilibrium Liquids, Academic, London, 1990.

[10] N. Goldenfeld, A. McKane and Q. Hou, Block spins for partial differential equations, J. Stat. Phys., 93, (1998), pp. 699–714.

[11] Q. Hou, N. Goldenfeld and A. McKane, Renormalization group and perfect operators for stochastic differential equations, Phys. Rev. E, 63 (2001), pp. 036125:1–22.

[12] L. Kadanoff, Statistical Physics: Statics, Dynamics, and Renormalization, World Scientific, Singapore, 2000.

[13] S.S. Ma, Modern Theory of Critical Phenomena, Benjamin, Reading, Mass, 1976.

[14] R. Swendsen, Monte-Carlo renormalization group, Phys. Rev. Lett. 42 (1979), pp. 859–861.

[15] R. Zwanzig, Nonlinear generalized Langevin equations, J. Stat. Phys., 9, (1973), pp. 215–220.