Title
Comparative performance of three different algorithms for the non-Markovian optimal prediction applied to the Hald system

Permalink
https://escholarship.org/uc/item/25k4b77w

Author
Okunev, Pavel

Publication Date
2001-08-15
Comparative Performance of Three Different Algorithms for the Non-Markovian Optimal Prediction Applied to the Hald System

Pavel Okunev
Computing Sciences Directorate
Mathematics Department

August 2001
To be submitted for publication
DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.
Comparative Performance of Three Different Algorithms for the Non-Markovian Optimal Prediction Applied to the Hald System

Pavel Okunev

Computing Sciences Directorate
Ernest Orlando Lawrence Berkeley National Laboratory
University of California
Berkeley, California 94720

August 2001

This work was supported by the Director, Office of Science, Office of Advanced Scientific Computing Research, Mathematical, Information, and Computational Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Pavel Okunev
Department of Mathematics
LBNL
Berkeley, CA 94720
August 15, 2001

Abstract

We consider the Hald system of coupled oscillators and apply to it the optimal prediction algorithm with memory kernel. We consider three different constructions for the kernel and compare their relative performance. We also observe that the exact average solution to the Hald system depends on the initial conditions in a nearly linear fashion.

1 The Hald System and a Brief Overview of Optimal Prediction Theory.

1.1 The Hald System.

The Hald system is a Hamiltonian system of four differential equations defined below. It was introduced and considered in [1]. This system is a useful model because it is nonlinear and at the same time is simple enough for analysis. The Hald Hamiltonian is

\[ H(q_1, p_1, q_2, p_2) = \frac{1}{2}(q_1^2 + p_1^2 + q_2^2 + p_2^2 + q_1 q_2) \]

Here \( q_1, p_1 \) and \( q_2, p_2 \) represent the positions and momenta of two coupled oscillators. This leads to the following equations of motion:
\[
d\frac{dq_1}{dt} = p_1, \quad q_1(0) = q_{10} \tag{1}
\]
\[
d\frac{dp_1}{dt} = -q_1 - q_1 q_2^2, \quad p_1(0) = p_{10} \tag{2}
\]
\[
d\frac{dq_2}{dt} = p_2, \quad q_2(0) = q_{20} \tag{3}
\]
\[
d\frac{dp_2}{dt} = -q_2 - q_2 q_1^2, \quad p_2(0) = p_{20} \tag{4}
\]

In the rest of the article we assume that the initial conditions \(q_{10}, p_{10}\) for \(q_1(t)\), and \(p_1(t)\) are given real numbers, while the initial conditions \(q_{20}, p_{20}\) for \(q_2(t), p_2(t)\) are random. The joint probability distribution for \((q_{10}, p_{10}, q_{20}, p_{20})\) is assumed to have density

\[
\rho_{\text{canonical}}(q_{10}, p_{10}, q_{20}, p_{20}) = \frac{1}{Z} \exp(-H(q_{10}, p_{10}, q_{20}, p_{20})). \tag{5}
\]

Here \(Z\) is a normalization constant chosen to make the integral of the density function over the entire space \(R^4\) to be one. This measure on the space of initial conditions is known in physics as the \(\text{canonical measure}\). The distribution of the initial data \((q_{10}, p_{10}, q_{20}, p_{20})\) defined by this measure is known as the \(\text{canonical distribution}\). If \(q_{10}, p_{10}\) are fixed numbers then the probability distribution for \(q_{20}, p_{20}\) has the conditional density

\[
\rho_{\text{conditional}}(q_{20}, p_{20}) = \frac{1}{Z_m} \exp(-H(q_{10}, p_{10}, q_{20}, p_{20})) \tag{6}
\]

where \(Z_m = Z_m(q_{10}, p_{10})\) is a constant chosen to make the integral of the conditional density over \(q_{20}\) and \(p_{20}\) equal to one.

We are interested in computing the average time evolution of \(p_1(t), q_1(t)\) where the average is taken over all possible values of \(q_{20}, p_{20}\) with \(q_{10}, p_{10}\) fixed. In other words we want to compute

\[
y_1(t) = E[p_1(t) | p_1(0), q_1(0)]_{p_1(0)=p_{10}, q_1(0)=q_{10}} \tag{7}
\]

and

\[
y_2(t) = E[q_1(t) | p_1(0), q_1(0)]_{p_1(0)=p_{10}, q_1(0)=q_{10}}. \tag{8}
\]

The conditional expectation here is computed with respect to the canonical measure defined above. The definition of conditional expectation can be found in the appendix or in [2].
Informally, these quantities describe the result of the following experiment. First we fix values of \( q_{10}, p_{10} \). Second, we sample \( q_{20}, p_{20} \) from the distribution defined by the conditional density \( \rho_{\text{conditional}}(q_{20}, p_{20}) \). Third, now that we have the full set of initial conditions for the Hald system we compute time evolution of \( q_1(t) \) and \( p_1(t) \) up to some time \( T \). We repeat steps two and three many times. Finally for each time \( t \in [0, T] \) we compute the arithmetic average of the \( q_1(t), p_1(t) \) over all the samples we took. One approach to estimating the average time evolution is reviewed in the following section.

1.2 Non-Markovian Optimal Prediction Theory as it Applies to the Hald System.

In [1] it was shown through the use of the Mori-Zwanzig formalism and some approximations that under certain conditions the time evolution of \( y_1(t), y_2(t) \) can be approximated by the solutions of the following system of integrodifferential equations:

\[
\begin{align*}
\frac{dy_1}{dt} &= y_2, \quad y_1(0) = q_{10} \\
\frac{dy_2}{dt} &= -y_1 - \frac{y_1}{1 + y_1^2} - \int_0^t K(t, s)y_2(s)ds, \quad y_2(0) = p_{10}
\end{align*}
\]

Here \( K(t, s) \) is an appropriate memory kernel which remains to be defined. Observe that these equations are non-Markovian since the right hand side depends on the entire history of \( y_2(t) \).

The memory kernel \( K(t, s) \) is obtained as follows. Let \( R(q_1, p_1, p_2, q_2) = -q_1 - q_1q_2^2 \) be the right hand side of equation (2). Define the function

\[
w(q_1, p_1) = E[R(q_1, p_1, p_2, q_2)|q_1, p_1] = \\
\int_R \int_R R(p_1, q_1, p_2, q_2)\rho_{\text{canonical}}(p_1, q_1, p_2, q_2)dp_2dq_2 = -q_1 - \frac{q_1}{1 + q_1^2}
\]

Here the conditional expectation is taken with respect to the canonical measure. Observe that \( w(q_1, p_1) \) only depends on \( q_1 \), for this reason from now on we will refer to it as \( w(q_1) \).

Consider the set of solutions \( \{(q_1(t), p_1(t), p_2(t), q_2(t))\} \) of the Hald system for all possible initial conditions \( (q_1(0), p_1(0), p_2(0), q_2(0)) \). We assume that initial conditions are distributed in accordance with the canonical measure.
This set of all solutions can be viewed as a stochastic process \( \{S(t)\} \) which is completely defined by the canonical measure and the Hald system of differential equations. For convenience of notation we introduce three stochastic processes. The stochastic process \( \{A(t)\} \) is the right hand side of (2) when initial data are sampled from the canonical distribution:

\[
\{A(t)\} = \{R(q_1(t), p_1(t), p_2(t), q_2(t))\}.
\]

The stochastic process \( \{B(t)\} \) can be interpreted as an average of \( \{A(t)\} \):

\[
\{B(t)\} = \{w(q_1(t))\}.
\]

Finally, \( \{C(t)\} \) is the difference of \( \{A(t)\} \) and \( \{B(t)\} \):

\[
\{C(t)\} = \{A(t) - B(t)\} = \{R(q_1(t), p_1(t), p_2(t), q_2(t)) - w(q_1(t))\}.
\]

The memory kernel \( K(t, s) \) is chosen to be equal to the autocorrelation of the stochastic process \( C(t) \):

\[
K(t, s) = E[C(t)C(s)] = E[(A(t) - B(t))(A(s) - B(s))]
\]

The authors of [1] then go on to suggest the following approximation to the kernel:

\[
K(t, s) \approx K^1(t, s) = E[A(t)A(s)] - w(y_1(t))w(y_1(s)) =
E[A(t - s)A(0)] - w(y_1(t))w(y_1(s))
\]

This approximate kernel \( K^0(t, s) \) is to be inserted into the equation (9) in place of \( K(t, s) \).

This approximation is based on two observations. First, we note that if \( \{X(t)\} \) is a stochastic process then

\[
E[(X(t) - EX(t))(X(s) - EX(s))] =
E[X(t)X(s)] - E[X(t)]E[X(s)]
\]

From definitions (10)-(12) one can conclude that \( B(t) \) can be interpreted as an average of \( A(t) \). Thus by analogy with (16) we expect that:

\[
K(t, s) = E[(A(t) - B(t))(A(s) - B(s))] \approx E[(A(t)A(s)] - B(t)B(s)
\]
Unfortunately, $B(t)B(s)$ is a random variable for any fixed $t, s$ and we would need a scalar value of $K(t, s)$ so we additionally approximate $B(t)B(s) = w(q_1(t))w(q_1(s)) \approx w(y_1(t))w(y_1(s))$. This gives us

$$K(t, s) \approx E[A(t)A(s)] - w(y_1(t))w(y_1(s))$$

(18)

The second observation is that $A(t)$ is a stationary process, therefore $E[A(t)A(s)] = E[A(t - s)A(0)]$. Thus we arrive at (15). In the discussion of the numerical results given in the following section we will always refer to the approximation (15) as type 1 approximation.

Another approximation for the memory kernel $K(t, s)$ can be found in an unpublished FORTRAN program of Chorin. Let $R_2(q_1, p_1, p_2, q_2) = -q_1q_2$ and define

$$w_2(q_1, p_1) = E[R_2(q_1, p_1, p_2, q_2)|q_1, p_1] = \int_R \int_R R_2(p_1, q_1, p_2, q_2) \rho_{\text{canonical}}(p_1, q_1, p_2, q_2) dp_2 dq_2 = -\frac{q_1}{1 + q_1^2}$$

(19)

As before the conditional expectation is taken with respect to canonical distribution.

Additionally, we define two stochastic processes. The stochastic process $\{A_2(t)\}$ is equal to $R_2(q_1, p_1, p_2, q_2)$ when the initial data are sampled from the canonical distribution:

$$\{A_2(t)\} = \{R_2(q_1(t), p_1(t), p_2(t), q_2(t))\}.$$  \hspace{1cm} (20)

The stochastic process $\{B_2(t)\}$ can be interpreted as average of $\{A_2(t)\}$:

$$\{B_2(t)\} = \{w_2(q_1(t))\}. \hspace{1cm} (21)$$

Let stochastic process $\{C(t)\}$ be defined as before, observe that $C(t) = A_2(t) - B_2(t)$. By the same same argument as was used to arrive at approximation (15) we arrive at:

$$K(t, s) \approx K^2(t, s) = \frac{E[A_2(t)A_2(s)]}{E[q_1^2 + p_1^2]} - w_2(y_1(t))w_2(y_1(s)) = \frac{E[A_2(t-s)A_2(0)]}{E[q_1^2 + p_1^2]} - w_2(y_1(t))w_2(y_1(s))$$

(22)

The only new feature here is the the normalization factor $E[q_1^2 + p_1^2]$ which divides $E[A_2(t-s)A_2(0)]$. $E[q_1^2 + p_1^2]$ is computed with respect to the canonical
measure. The approximate kernel $K^2(t,s)$ is to be inserted in the right hand side of equation (9) in place of $K(t,s)$. We will from now on refer to this approximation as type 2 approximation. In the next section we discuss the comparative performance of the two approximation as well as the results obtained by using the memory kernel $K(t,s)$ directly without making further approximations.

2 Numerical Results and Interpretations.

2.1 Numerical Results Obtained for Type 1 Approximation.

As was discussed in the previous section type 1 approximation of the memory kernel leads to the following system of equations for the time evolution of $y_1(t), y_2(t)$:

\[
\begin{align*}
\frac{dy_1}{dt} &= y_2 \\
y_1(0) &= q_{10} \\
\frac{dy_2}{dt} &= -y_1 - \frac{y_1}{1 + y_1^2} - \int_0^t \left( K^0(t-s) - w(y_1(t))w(y_1(s)) \right) y_2(s) ds \\
y_2(0) &= p_{10}
\end{align*}
\]  

(23)

Here $K^0(t) = E[A(t)A(0)]$, with $\{A(t)\}$ defined by (11). We compute $K^0(t)$ using Markov chain Monte-Carlo sampling. In our computations we used 50,000 samples of initial conditions $(q_{10}, p_{10}, q_{20}, p_{20})$. For each particular set of initial conditions we then computed the time evolution of $(q_1(t), p_1(t), p_2(t), q_2(t))$ using the fourth order Runge-Kutta method. Finally for each $t$ we computed the value of $E[A(t)A(0)]$. The results are shown in Figure 1.

We want to compare the values of $y_1(t), y_2(t)$ given by (23) to the exact time evolution of $Y_1(t) = E[p_1(t)\mid p_1(0), q_1(0)]_{p_1(0)=p_{10}, q_1(0)=q_{10}}$ and $Y_2(t) = E[q_1(t)\mid p_1(0), q_1(0)]_{p_1(0)=p_{10}, q_1(0)=q_{10}}$. In order to do this we compute the exact averages $Y_1(t), Y_2(t)$ by Monte-Carlo sampling. For this computation we also used 50,000 samples. In Figure 2 we show the exact averages $Y_1(t), Y_2(t)$ and the averages given by type 1 approximation for initial data $q_{10} = 1, p_{10} = 0$. One can see that the approximations $y_1(t), y_2(t)$ follow the true evolution for a short time ($0 \leq t \leq 1$). For later times there is no agreement in
Kernel $K^0$

Approximation $Y_1(t)$
Exact evolution $Y_1(t)$

Figure 1

Approximation $Y_2(t)$
Exact evolution $Y_2(t)$

Figure 2
values between the approximations and the true evolution. Additionally, the approximations are out of phase with the exact evolution and even have different period. These results are very representative of the experiments we have done with different initial data.

Another interesting feature of type 1 approximation can be observed by looking at the behavior of the approximation for bigger initial data. Figure 3 shows the results for initial data $q_{10} = -10, p_{10} = -10$. Here we see that the approximations $y_1(t), y_2(t)$ blow up. On the other hand the exact averages $Y_1(t), Y_2(t)$ can be shown to tend to zero as $t \to \infty$. Thus the approximation gives bad results for these initial conditions. To understand the reasons for blow up we look more closely at the behavior of the solution for shorter times. Figure 4 shows the results for the same initial conditions $q_{10} = -10, p_{10} = -10$ but for $t \leq 0.2$. One observes that the first component of the approximation $y_1(t)$ is always decreasing while the second component $y_2(t)$ increases for very small times and decreases for all later times.

This behavior can be easily explained. Indeed, the derivative of $y_1(t)$ is $y_2(t)$. Since $y_2(t)$ is initially a negative number $y_1(t)$ will be decreasing as long as $y_2(t)$ remains negative. Thus $y_1(t)$ remains negative and its absolute value is increasing, given that we always have $y_2(t) \leq 0$. On the other hand the derivative of $y_2(t)$ is $-y_1 - \frac{y_1}{y_1 + y_2^2} - \int_0^t (K^0(t - s) - w(y_1(s))w(y_1(s)))y_2(s)ds$. Since $y_1(t)$ is initially large, we can approximate $-y_1 - \frac{y_1}{y_1 + y_2^2}$ by $-y_1$. Additionally we can approximate $w(y_1(t)) = -y_1(t) - \frac{y_1(t)}{1+y_1(t)^2}$ by $-y_1(t)$. Observe
that both $y_1(t)$ and $y_1(s)$, are big compared to values of $K^0$. Therefore
$K^0(t - s) - (-y_1(s))(-y_1(t)) \approx -y_1(s)y_1(t)$. Thus we have arrived at the
following approximation which is valid as long as $y_1(t)$ is bounded away from
zero by some number which is much bigger than 1 and much bigger than
maximum of $K^0(t)$:

$$\frac{dy_2}{dt} \approx -y_1 + \int_0^t y_1(t)y_1(s)y_2(s)ds \quad (24)$$

If $y_1(t), y_1(s)$ and $y_2(s)$ are large in absolute value and $t$ is not very small
then (24) can be additionally approximated by

$$\frac{dy_2}{dt} \approx \int_0^t y_1(t)y_1(s)y_2(s)ds \quad (25)$$

Thus the right hand side will be a negative number. Therefore $y_2(t)$ will
remain negative and will increase in absolute value. The only exception
occurs for very small times $t$. In this case the right hand side of (24) will be
dominated by $-y_1$ and therefore for small times $y_2(t)$ will be an increasing
function. However if $y_2(0)$ is big enough, $y_2(t)$ will not reach zero or become
too small in absolute value before the approximation (25) becomes valid.
The above analysis shows that if the initial conditions for $y_1(t), y_2(t)$ are
big negative numbers then $y_1(t), y_2(t)$ will be decreasing functions for all
sufficiently large times. Thus they cannot be decaying to zero, as the exact
average solutions $Y_1(t), Y_2(t)$ do.
2.2 Numerical Results Obtained For Type 2 Approximation.

In section 1.2 we discussed a type 2 approximation for the average time evolution of the Hald system.

\[
\begin{align*}
\frac{dy_1}{dt} &= y_2 \\
y_1(0) &= q_{10} \\
\frac{dy_2}{dt} &= -y_1 - \frac{y_1}{1 + y_1^2} - \int_0^t \left( K^0(t - s) - w_2(y_1(t))w_2(y_1(s)) \right) y_2(s) ds \\
y_2(0) &= p_{10}
\end{align*}
\]

Here \( K^0(t) = \frac{E[A_2(t-s)A_2(0)]}{E[q_1^2 + p_1^2]} \) and \( w_2(y_1(t)) = -\frac{y_1(t)}{1 + y_1(t)^2} \). As before we compute the memory kernel \( K^0(t) \) and the true average solutions \( Y_1(t), Y_2(t) \) by Markov chain Monte-Carlo sampling with 50,000 samples. Figure 5 shows the results of a computation with initial conditions \( q_{10} = 1, p_{10} = 0 \).

![Figure 5](image)

This represents a dramatic improvement compared to results obtained for type 1 approximation. The approximate averages \( y_1(t), y_2(t) \) follow the true average solution \( Y_1(t), Y_2(t) \) for much longer times, \( t \in [0, 5] \). Additionally for later times the approximation is roughly in phase with the correct average solution. Also \( y_1(t), y_2(t) \) decay at a rate approximating that of the
correct solution. We have done a number of experiments with various initial conditions and results were generally similar.

One of the features of type 2 approximation is that it involves a normalization factor $E[q_1^2 + p_1^2]$. If we do not normalize the memory kernel by this factor and choose $K^0(t) = E[A_2(t - s)A_2(0)]$ we will get, for the same initial conditions, different results shown in Figure 6. The approximations $y_1(t), y_2(t)$ obtained in this way do not follow exact solution as closely as in Figure 5. The approximations also fall out of phase after $t = 10$ and do not decay fast enough. Thus the normalization factor $E[q_1^2 + p_1^2]$ seems to improve the approximations significantly.

2.3 The Numerical Results Obtained By Using the Exact Kernel.

Finally we look at the results obtained by using the exact memory kernel $K(t, s)$ defined by (14). It is known that $\{C(t)\}$ is a stationary stochastic process. Therefore $K(t, s) = E[C(t)C(s)] = E[C(t - s)C(0)]$. We define

$$K^0(t) = E[C(t)C(0)]$$  \hspace{1cm} (27)

Then the system (9) becomes:

$$\frac{dy_1}{dt} = y_2$$
\begin{align*}
y_1(0) &= q_{10} \\
dy_2 \over dt &= -y_1 - {y_1 \over 1 + y_1^2} - \int_0^t K^0(t - s)y_2(s)ds \\
y_2(0) &= p_{10}
\end{align*}

(28)

Just as we did before we compute the memory kernel \(K^0(t)\) and the true average solutions \(Y_1(t), Y_2(t)\) using Markov chain Monte-Carlo sampling with 50,000 samples. Figure 7 shows the results obtained for initial conditions \(q_{10} = 6, p_{10} = 6\). We observe that up to time \(t \approx 9\) the results obtained for exact kernel agree very well with the results given by type 2 approximation. On this time interval both approximation also agree fairly well with the exact average solution. After \(t = 9\) both approximations fail to follow the correct solution. The exact kernel approximation decays too fast but follows the phase of the true solution better. The type 2 approximation falls out of phase and decays too slowly. However the rate of decay for type 2 approximation is closer to the rate of decay for true solution. This kind of behavior is typical for the case of relatively big initial data.

We now look at the results obtained for relatively small initial data \(q_{10} = 0.1, p_{10} = -0.1\) (Figure 8). For \(t \leq 5\) both approximation agree very closely with each other and do a good job of following the correct solution. After \(t \approx 5\) both exact kernel approximation and type 2 approximation fail completely to follow either the phase or the general shape of the true average solution.

Figure 7

Figure 8
This was typical for the experiments that we did with different small initial data. Experiment with mixed initial data (for example $q_{10}$ big, $p_{10}$ small) gave hybrid results, which do not add much to our understanding.

One advantage of using the exact kernel is that it does not require any normalization factors, like $E[q_1^2 + p_1^2]$. On the other hand type 2 approximation seems to predict the rate of decay of the true solution better, at least in the case of large initial conditions. This is surprising, since type 2 approximation is after all only an approximation to the procedure which uses the exact kernel (27).

2.4 Conclusions.

Our experiments have shown that type 1 approximation fails to predict correctly the behavior of the exact average solution. Type 2 approximation gives better results, but decays to slowly for large initial conditions. The procedure with the exact kernel performs better than type 1 approximation, but decays too fast, at least, in the case of large initial conditions. It is hard to say whether type 2 approximation is better than the procedure with the exact kernel or whether it is the other way around. Both approximations follow the exact solution well for roughly the same length of time. However, one of the advantages of the procedure with the exact kernel is that it does not require any normalization factors. Such normalization factors may be hard to find when we try to apply optimal prediction techniques to a different system.
of differential equations.

3 Acknowledgments.

We thank Professor Chorin and Professor Hald of UC Berkeley and LBNL for their advice and help in our work. We also wish to thank Mr. Benjamin Seibold for doing numerical experiments which provided independent verification for our initial observations of the facts described in section 2 and for numerous discussions which helped us formulate conclusions given in section 2 in the form in which they are now.

Appendix.

A Expectation and Conditional Expectation.

Let \( (p_1, q_1, p_2, q_2) \) denote a vector from four dimensional real space \( \mathbb{R}^4 \). Let a measure on this space be defined by its density \( \rho(p_1, q_1, p_2, q_2) \). Let \( X(p_1, q_1, p_2, q_2) \) be a random variable on this probability space.

Definition 1 The expectation of \( X(p_1, q_1, p_2, q_2) \) is defined by the formula:

\[
E[X(p_1, q_1, p_2, q_2)] = \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} X(p_1, q_1, p_2, q_2) \rho(p_1, q_1, p_2, q_2) dp_1 dq_1 dp_2 dq_2
\]

Definition 2 The conditional expectation of \( X(p_1, q_1, p_2, q_2) \) given \( p_1 \) and \( q_1 \) is defined by the formula:

\[
E[X(p_1, q_1, p_2, q_2)|p_1, q_1] = \frac{\int_{\mathbb{R}} \int_{\mathbb{R}} X(p_1, q_1, p_2, q_2) \rho(p_1, q_1, p_2, q_2) dp_2 dq_2}{\int_{\mathbb{R}} \int_{\mathbb{R}} \rho(p_1, q_1, p_2, q_2) dp_2 dq_2}
\]

More general definitions of the expectation and the conditional expectation can be found in [2].

References

