LBNL-44191
Preprint

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

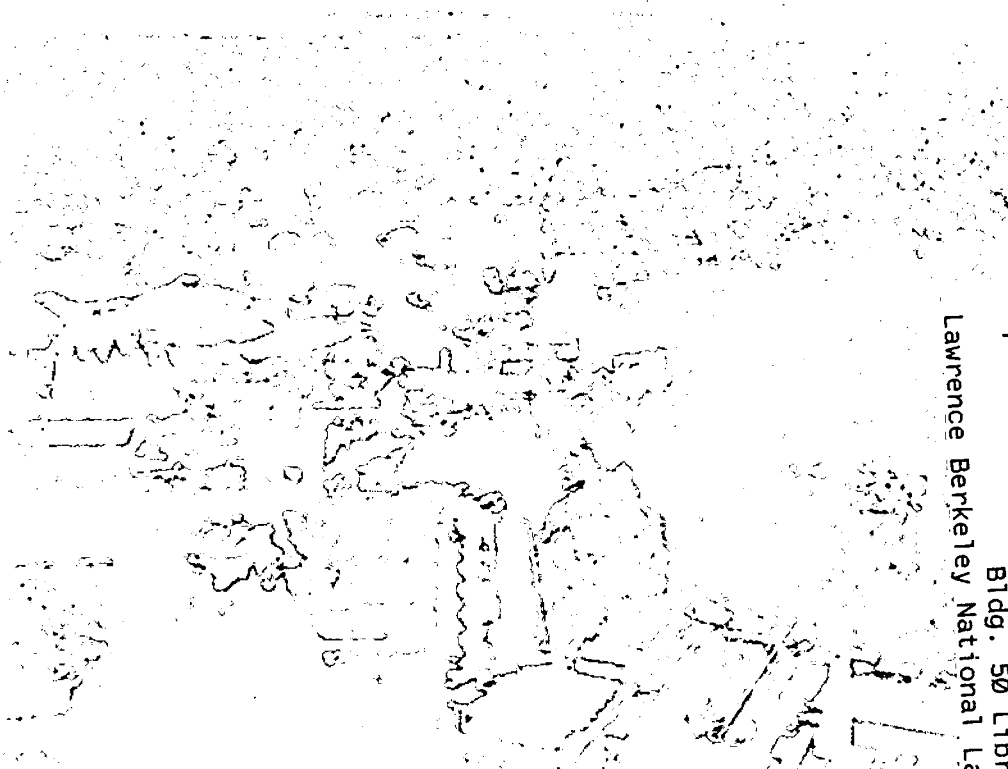
Experiments in First-Order Optimal Prediction

Thibaut Burin des Roziers and Alexander Gottlieb

Computing Sciences Directorate
Mathematics Department

August 1999

To be submitted
for publication



DISCLAIMER

LBNL-44191

EXPERIMENTS IN FIRST-ORDER OPTIMAL PREDICTION*

Thibaut Burin des Roziers and Alexander Gottlieb

Mathematics Department
Computing Sciences Directorate
Lawrence Berkeley National Laboratory
Berkeley, CA 94720, USA

August 1999

*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.

Experiments in First-Order Optimal Prediction

Thibaut Burin des Roziers and Alexander Gottlieb

Abstract

We review certain cases where first-order optimal prediction has been found to perform well. Numerical experiments show that the main reason for the success of optimal prediction in these cases is dynamical quasi-invariance of the collective variables.

1 First-order optimal prediction

A method of "optimal prediction of underresolved dynamics" has recently been proposed by Chorin, Kast, and Kupferman [1] for producing computationally feasible numerical schemes for nonlinear evolution problems, such as turbulent flow problems. They consider a space X of vectors or functions that is closed under the evolution

$$\frac{d}{dt}x = \mathcal{F}(x), \tag{1}$$

which may be either a system of ordinary differential equations or a partial differential equation. Given an ensemble of initial conditions for (1) distributed according to the probability measure P^0 on X , the ensemble of solution values at time t is distributed according to

$$P^t(A) = P^0\left(\{x : x(t) \in A \text{ when } x(0) = x\}\right).$$

Expectations with respect to the measure P^t satisfy the change of variables formula

$$\mathbb{E}_{P^t}[h(x)] = \mathbb{E}_{P^0}[h(x(t))], \tag{2}$$

where $x(t)$ is the solution at time t of (1) with initial data $x(0) = x$. The purpose of Optimal Prediction is to approximate the evolution $P^0 \to P^t$ by solving a system of k ordinary differential equations. The idea is to select a family $\{Q_\theta\}$ of probability measures on X parameterized by a subset of $\mathbb{R}^k$, and to derive differential equations $\frac{d}{dt}\theta = f(\theta)$ for the parameters from the original evolution equation (1), so that $Q_{\theta(t)} \approx P^t$ for as long as possible.

The papers [1, 3, 4, 5, 6, 7] develop a simplistic version of the method of optimal prediction, called first-order prediction. This method, as applied to large systems of ODEs, goes as follows:

Let

$$\frac{d}{dt}\mathbf{x} = F(\mathbf{x}) \tag{3}$$

be a system of n ODEs. (Imagine a Hamiltonian system with many degrees of freedom.) Let $g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_k(\mathbf{x})$ be k differentiable functions on $\mathbb{R}^n$, with $k << n$, and let Q be a measure on $\mathbb{R}^n$. (If the system of ODEs is Hamiltonian, the measure Q might be taken to be a canonical Gibbsian measure.) For each $\theta = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$ let Q_θ denote the conditional probability under Q given that $g_i(\mathbf{x}) = \theta_i$ for $i = 1, 2, \dots, k$. That is

$$Q_\theta(A) = \mathbb{P}_Q[A | g_1(\mathbf{x}) = \theta_1, \dots, g_k(\mathbf{x}) = \theta_k]. \tag{4}$$

Let $P^0 = Q_{\theta(0)}$ be the distribution of an ensemble of initial conditions for the ODEs (3). We wish to "predict" expectation values with respect to P^t.

Let us assume that the family $\{Q_\theta\}_{\theta \in \mathbb{R}^k}$ of measures is rich enough that the measures P^t can continually be approximated by measures from that family. This is called the *closure assumption* [6]. By the change of variables formula (2),

$$\begin{aligned} \frac{d}{dt}\mathbb{E}_{P^t}[g_i(\mathbf{x})] = \frac{d}{dt}\mathbb{E}_{P^0}[g_i(\mathbf{x}(t))] &= \mathbb{E}_{P^0}[\nabla g_i(\mathbf{x}(t)) \cdot F(\mathbf{x}(t))] \\ &= \mathbb{E}_{P^t}[\nabla g_i(\mathbf{x}) \cdot F(\mathbf{x})]. \end{aligned}$$

If the assumption that P^t can be approximated by a measure $Q_{\theta(t)}$ is valid, the preceding equation implies that $\frac{d}{dt}\mathbb{E}_{Q_{\theta(t)}}[g_i(\mathbf{x})] \approx \mathbb{E}_{Q_{\theta(t)}}[\nabla g_i(\mathbf{x}) \cdot F(\mathbf{x})]$. But $\mathbb{E}_{Q_{\theta(t)}}[g_i(\mathbf{x})] = \theta_i(t)$ by definition of the measure $Q_{\theta(t)}$, so the closure assumption suggests that the parameter $\theta(t)$ should obey the system of differential equations

$$\frac{d}{dt}\theta_i(t) = \mathbb{E}_{Q_{\theta(t)}}[\nabla g_i(\mathbf{x}) \cdot F(\mathbf{x})] \tag{5}$$

These are the *equations of first-order optimal prediction* for the evolution problem (3) with *prior measure* Q and *collective variables* $g_1(\mathbf{x}), \dots, g_k(\mathbf{x})$.

They may be called "first-order" because they do correctly predict $\mathbb{E}_{P^t}[g_i(\mathbf{x})]$ to first order for $0 \leq t << 1$: if $P^0 = Q_{\theta(0)}$ then

$$\frac{d}{dt}\mathbb{E}_{P^t}[g_i(\mathbf{x})]\Big|_{t=0} = \frac{d}{dt}\mathbb{E}_{Q_{\theta(t)}}[g_i(\mathbf{x})]\Big|_{t=0}.$$

2 When does first-order prediction work?

For a generic prediction problem, the closure assumption will not be valid and the first-order method is not expected to succeed. However, some numerical experiments have been performed which show that optimal prediction sometimes works insofar as it successfully predicts the expectations of the collective variables given their initial values. Speculation concerning circumstances which may contribute to the success of first-order optimal prediction has produced several hypotheses. Two of these hypotheses are mentioned in [6]: for PDEs, or for systems of ODEs derived from PDEs, first-order prediction is more apt to succeed if (i) the collective variables represent spatial averages of the primary variables which are not concentrated in space, and (ii) the collective variables are not funtions of *disjoint* groups of primary variables. Another hypothesis is that it is helpful to select collective variables that are functions of those primary variables that change more slowly under the dynamics. It also might seem favorable for optimal prediction that the prior measure Q be concentrated, i.e., that samples from Q have a small variance.

We have found that when first-order prediction succeeds, it is not primarily due to the preceding circumstances. Rather, the success of first-order optimal prediction depends mainly on the choice of collective variables: **first-order prediction works well if the set of functions of the collective variables is nearly invariant under the dynamics**. This hypothesis was suggested to us by the work of O. Hald on first-order prediction for linear Hamiltonian problems [2].

For special choices of the collective variables, first-order prediction correctly determines the expected values of the collective variables. The equations (5) of first-order optimal prediction for the system

$$\begin{aligned} \frac{d}{dt}u_i &= F_i(\mathbf{u})\ ; \quad \mathbf{u} = (u_1, u_2, \ldots, u_k) \\ \frac{d}{dt}w_i &= G_i(\mathbf{u}, \mathbf{w})\ ; \quad \mathbf{w} = (w_1, w_2, \ldots, w_{n-k}) \end{aligned} \tag{6}$$

with an *arbitrary* prior measure Q and collective variables

$$g_1(\mathbf{u},\mathbf{w}) = u_1,\ g_2(\mathbf{u},\mathbf{w}) = u_2,\ \ldots,\ g_k(\mathbf{u},\mathbf{w}) = u_k$$

are simply

$$\frac{d}{dt}\theta_i = F_i(\theta_1, \theta_2, \ldots, \theta_k)\ ; \quad i = 1, 2, \ldots, k.$$

If the ensemble of initial conditions for (6) is distributed according to $P^0 = Q_{\theta(0)}$ then by (2) and (4)

$$\mathbb{E}_{P^t}[g_i(\mathbf{u},\mathbf{w})] = \mathbb{E}_{P^t}[u_i] = \mathbb{E}_Q[u_i(t)|\mathbf{u}(0) = \theta(0)] = \theta(t),$$

and the first-order method correctly "predicts" the expected values of the collective variables at time t. This does not mean that $P^t \approx Q_{\theta(t)}$; first-order prediction correctly determines the expectations of functions $h(\mathbf{u})$ of the collective variables, but not the expectations of general functions $h(\mathbf{u},\mathbf{w})$.

Suppose, in particular, that A is an $n \times n$ matrix and that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ span a subspace of $\mathbb{R}^n$ that is invariant under A^{tr}. Then, the system

$$\dot{\mathbf{x}} = A\mathbf{x}$$

reduces to the form (6) just considered upon setting $u_i = \mathbf{x}\cdot\mathbf{v}_i$ and $w_j = \mathbf{x}\cdot\mathbf{w}_j$ for any vectors $\mathbf{w}_j$ which, together with the vectors $\mathbf{v}_i$, span $\mathbb{R}^n$. Thus, if $\theta(t)$ solves the equations of first-order optimal prediction for $\dot{\mathbf{x}} = A\mathbf{x}$ with collective variables $\mathbf{x}\cdot\mathbf{v}_1, \ldots, \mathbf{x}\cdot\mathbf{v}_k$ and *any* prior measure Q, then

$$\mathbb{E}_{P^t}[\mathbf{x}\cdot\mathbf{v}_i] = \theta_i(t)$$

if $P^0 = Q_{\theta(0)}$. First-order prediction exactly determines the expectations of the collective variables.

In light of the preceding, first-order optimal prediction may be expected to do well if the PDE or system of ODEs is a slight perturbation of a linear evolution equation for which the collective variables span an invariant subspace. We believe that every known success of first-order optimal prediction can be attributed to these circumstances, and that the other circumstances thought to favor the performance of optimal prediction matter much less. The numerical experiments reported in the next section support this opinion.

3 Tests of first-order prediction

In this section we review three of the cases where first-order optimal prediction has been found to work very well, and report the results of similar experiments we conducted ourselves. We interpret our results in Section 4.

3.1 Coupled harmonic oscillators

Kast [7] considers the following system of 1001 coupled oscillators:

$$\begin{aligned}
\dot{Q} &= P \\
\dot{q}_j &= j^2 p_j \; ; \quad j = 1, 2, \ldots, 1000 \\
\dot{P}_j &= -Q + \sum_{i=1}^{1000} (q_i - Q) \\
\dot{p}_j &= (Q - q_j) + \sum_{i=1}^{1000} (q_i - q_j) \; ; \quad j = 1, 2, \ldots, 1000.
\end{aligned} \tag{7}$$

The collective variables are taken to be $Q, P, q_1, p_1, \ldots, q_{10}, p_{10}$. The prior measure is the canonical measure $e^{-H} / \int e^{-H}$ determined by the Hamiltonian function H for which (7) are the canonical equations. Kast finds good agreement between

$$\mathbb{E}[P(t) | Q(0), P(0), q_1(0), \ldots, p_{10}(0)]$$

and the optimally predicted values until (at least) $t = 1$.

We tested a simpler version of Kast's system of coupled oscillators. We considered the system

$$\dot{q}_j = j^2 p_j \; ; \quad j = 1, 2, \ldots, 100 \tag{8}$$

$$\dot{p}_j = q_i + \sum_{i=1}^{100} (q_i - q_j) \; ; \quad j = 1, 2, \ldots, 100. \tag{9}$$

$$\tag{10}$$

As our prior measure Q we took

$$\frac{1}{Z} \exp\left(-\frac{1}{2} \sum_{j=1}^{100} j^2 p_j^2 + \frac{1}{4} \sum_{j,k=1}^{100} (q_j - q_k)^2 + \frac{1}{2} \sum_{j=1}^{100} q_j^2 \right) d\mathbf{q} d\mathbf{p}. \tag{11}$$

For several choices of the collective variables $g_i(\mathbf{q}, \mathbf{p})$, we took $P^0 = Q_{\theta(0)}$ as defined in formula (4). We compared the exact values of $\mathbb{E}_{P^t}[g_3(\mathbf{p})]$ to the estimates of those values $\mathbb{E}_{Q_{\theta(t)}}[g_3(\mathbf{p})]$ given by first-order optimal prediction. Here we report the results of seven experiments where twenty collective variables

$$g_1(\mathbf{q}), g_1(\mathbf{p}), \ldots, g_{10}(\mathbf{q}), g_{10}(\mathbf{q})$$

were taken to be:

(i) $q_1, p_1, \ldots, q_{10}, p_{10}$

(ii) $q_{11}, p_{11}, \ldots, q_{20}, p_{20}$

(iii) $q_{91}, p_{91}, \ldots, q_{100}, p_{100}$

(iv) $\sum_{j=1}^{10}(M_{1j}q_j), \sum_{j=1}^{10}(M_{2j}p_j), \ldots, \sum_{j=1}^{10}(M_{j,10}q_j), \sum_{j=1}^{10}(M_{j,10}p_j)$
where M was an 10×10 invertible matrix with integer entries

(v) $q_1 + 0.1q_{11}, p_1 + 0.1p_{11}, \ldots, q_{10} + 0.1q_{20}, p_{10} + 0.1p_{20}$

(vi) $q_1 + 0.2q_{11}, p_1 + 0.2p_{11}, \ldots, q_{10} + 0.2q_{20}, p_{10} + 0.2p_{20}$

(vii) $q_1 + 0.5q_{11}, p_1 + 0.5p_{11}, \ldots, q_{10} + 0.5q_{20}, p_{10} + 0.5p_{20}$

(viii) $q_1 + q_{11}, p_1 + p_{11}, \ldots, q_{10} + q_{20}, p_{10} + p_{20}$

The results of these experiments are shown in Figures 1 - 6. The solid curves in those figures graph $\mathbb{E}_{Q_{\theta(t)}}[g_3(\mathbf{p})]$ as a function of t, and the crosses mark the values of $\mathbb{E}_{P^t}[g_3(\mathbf{p})]$. The graph for experiment (iii) is not printed (the graph is too crowded) but first-order prediction was as successful for this case as it was for (i) and (ii).

We also performed two experiments with the collective variables

$$q_1, p_1, \ldots, q_{10}, p_{10},$$

but with different masses for the oscillators. In the first experiment, the masses were taken to be $1/j$ instead of $1/j^2$. Thus we replaced j^2 by j in the equation of motion (8) and in the fromula for the density of the prior measure (11). In the second experiment, the masses were all taken to be 1, so we replaced j^2 by 1 in (8) and (11). The results of these experiments are graphed in Figures 7 and 8.

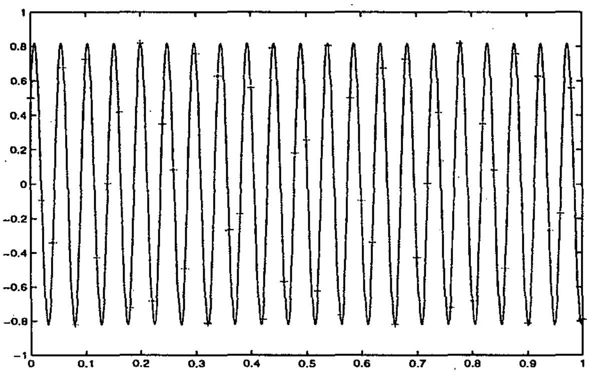


Figure 1: Coupled oscillators (ii)

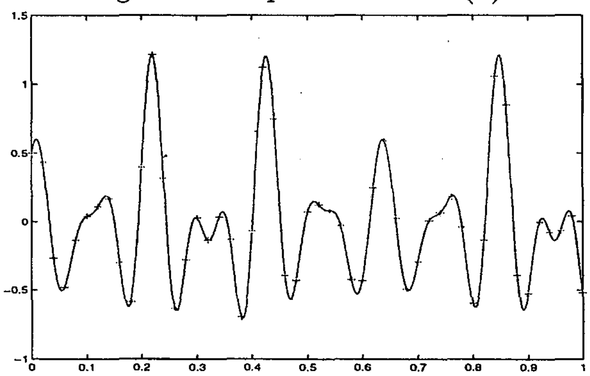


Figure 2: Coupled oscillators (iv)

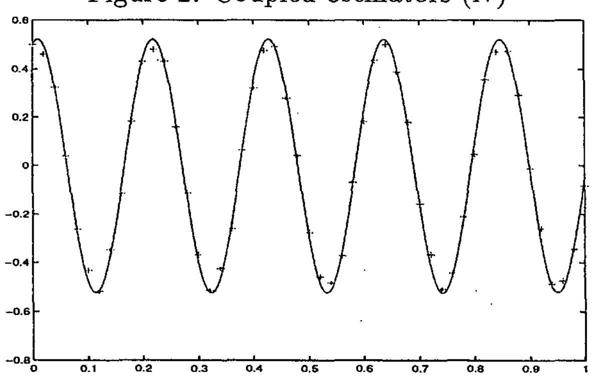


Figure 3: Coupled oscillators (v)

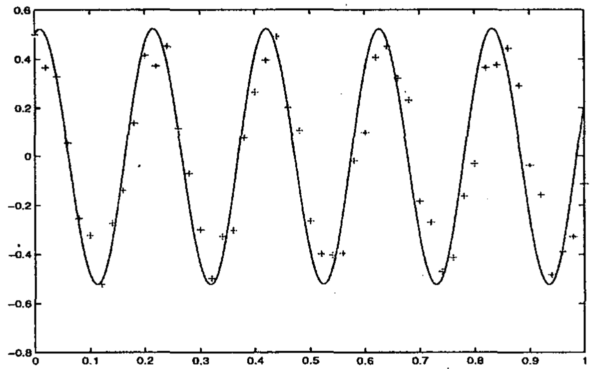


Figure 4: Coupled oscillators (vi)

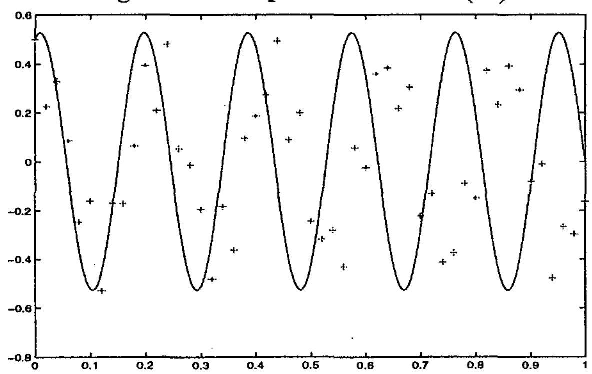


Figure 5: Coupled oscillators (vii)

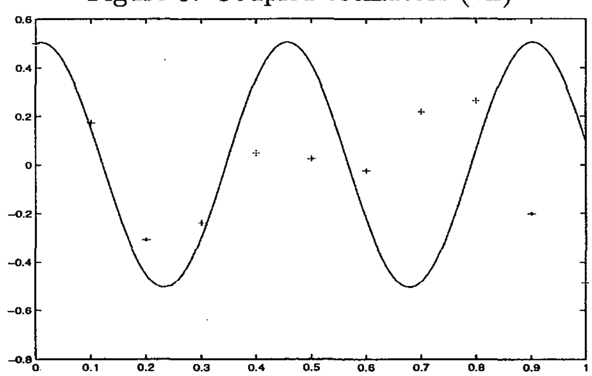


Figure 6: Coupled oscillators (viii)

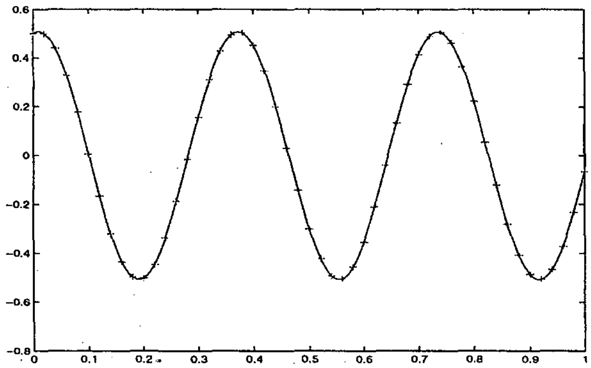


Figure 7: Oscillators with masses $1/j$

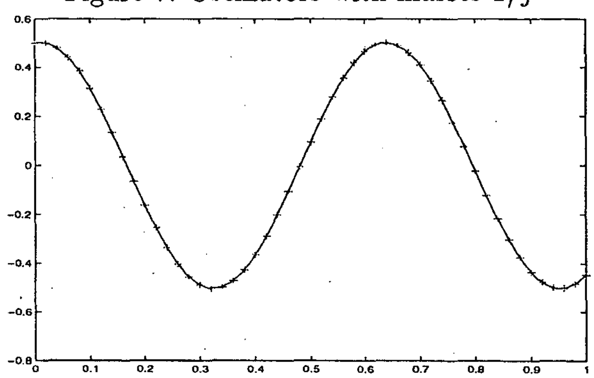


Figure 8: Oscillators with masses 1

3.2 A nonlinear system of ODEs

In [1, 4, 5], first-order optimal prediction is applied to the Hamiltonian system

$$\begin{aligned} \dot{q}_i &= -256(p_{i-1} - 2p_i + p_{i+1}) + p_i^3 \\ \dot{p}_i &= +256(q_{i-1} - 2q_i + q_{i+1}) - q_i^3 \; ; \quad i = 1, 2, \ldots, 16. \end{aligned} \tag{12}$$

Their prior measure Q is the canonical measure

$$\frac{1}{Z} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{16} \left(256(q_{i+1} - q_i)^2 + 256(p_{i+1} - p_i)^2 + \frac{1}{2}(q_i^4 + p_i^4) \right) \right\} d\mathbf{q} d\mathbf{p}. \tag{13}$$

Four collective variables

$$\begin{aligned} &\mathbf{g}_1 \cdot \mathbf{q},\ \mathbf{g}_2 \cdot \mathbf{q}, \qquad \mathbf{g}_1 \cdot \mathbf{p}, \mathbf{g}_2 \cdot \mathbf{p}; \\ &\mathbf{q} = (q_1, q_2, \ldots, q_{16}), \qquad \mathbf{p} = (p_1, p_2, \ldots, p_{16}) \end{aligned} \tag{14}$$

are selected, where

$$\begin{aligned} \mathbf{g}_1 &= (1, r, r^4, r^9, \ldots, r^{49}, r^{64}, r^{49}, \ldots, r^9, r^4, r)/z \\ \mathbf{g}_2 &= (r^{64}, r^{49}, r^{36}, \ldots, r^4, r, 1, r, r^4, \ldots, r^{36}, r^{49})/z \\ r &= e^{-1/16}, \quad z = 1 + r + r^2 + \cdots + r^{64} + \cdots + r. \end{aligned}$$

The conditional expectations that appear in the first-order equations (5) of optimal prediction are estimated by regression. The resulting equations are found to predict very well the conditional expectations

$$\mathbb{E}[\mathbf{g}_i \cdot \mathbf{q}(t) | q_1(0), \ldots, p_{16}(0)] \quad \text{and} \quad \mathbb{E}[\mathbf{g}_i \cdot \mathbf{q}(t) | q_1(0), \ldots, p_{16}(0)]$$

up until $t = 10$ (at least).

We repeated the preceding experiment, and obtained the same results as [1, 4, 5]. These results are presented in Figures 9. In this figure and the the rest of the figures of this section, the solid and dashed curves in those figures are the graphs of $\mathbb{E}_{Q_{\theta(t)}}[\mathbf{g}_i(\mathbf{q})]$ and $\mathbb{E}_{Q_{\theta(t)}}[\mathbf{g}_i(\mathbf{p})]$ for $i = 1, 2$, while the circles and crosses mark the values of $\mathbb{E}_{P^t}[\mathbf{g}_i(\mathbf{q})]$ and $\mathbb{E}_{P^t}[\mathbf{g}_i(\mathbf{p})]$. The solid and dashed curves are labelled $V1p, V2p, V1q, V2q$; the label $V2p$, for example, stands for $\mathbb{E}_{Q_{\theta(t)}}[\mathbf{g}_2(\mathbf{p})]$.

We also performed a variant of the original experiment wherein we used the same collective variables but changed the prior measure. We substituted 16 for 256 in formula (13) for the prior measure, which makes the prior measure about four times as spread out. The result of this experiment is presented in Figure 10.

To test the effect of the collective variables on the performance of first-order prediction, we used the same prior measure (13) but used different collective variables of the form (14), where the vectors $\mathbf{g}_1$ and $\mathbf{g}_2$ were taken to be

(i)

$$\begin{aligned} \mathbf{g}_1 &= (1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0) \\ \mathbf{g}_2 &= (0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1) \end{aligned}$$

(ii)

$$\begin{aligned} \mathbf{g}_1 &= (0,1,2,3,4,5,6,7,8,7,6,5,4,3,2,1)/8 \\ \mathbf{g}_2 &= (8,7,6,5,4,3,2,1,0,1,2,3,4,5,6,7)/8 \end{aligned}$$

(iii)

$$\begin{aligned} \mathbf{g}_1 &= (1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0) \\ \mathbf{g}_2 &= (0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1) \end{aligned}$$

(iv)

$$\begin{aligned} \mathbf{g}_1 &= (1,0,0,1,1,1,0,0,0,0,1,0,1,0,0,1) \\ \mathbf{g}_2 &= (0,1,0,0,0,1,1,1,1,1,0,0,0,1,0,1) \end{aligned}$$

(v)

$$\begin{aligned} \mathbf{g}_1(j) &= \cos(2\pi j/16), \quad j = 1,2,\ldots,16 \\ \mathbf{g}_2(j) &= \sin(2\pi j/16), \quad j = 1,2,\ldots,16 \end{aligned}$$

(vi)

$$\begin{aligned} \mathbf{g}_1(j) &= \cos(2\pi j/16) + \cos(4\pi j/16), \quad j = 1, 2, \ldots, 16 \\ \mathbf{g}_2(j) &= \sin(2\pi j/16) + \sin(4\pi j/16), \quad j = 1, 2, \ldots, 16 \end{aligned}$$

The results of these experiments are shown in Figures 11 - 15.

To compute expectations such as $\mathbb{E}_{P^t}[\mathbf{g}_i(\mathbf{q})]$, we sampled 100 initial conditions according to $P^0 = Q_{\theta(0)}$, solved the ODEs (12) for each initial condition, and averaged. We feel that 100 samples give us adequately accurate results because we redid a few of our experiments with as many as 10^4 samples and did not observe an appreciable change in the average values. The initial conditions were sampled using a Metropolis Markov chain Monte Carlo algorithm. We would allow 5×10^5 timesteps to pass before sampling the state of the Markov chain, then we let 10^4 timesteps pass between samples, for we estimated the decorrelation time to be between 10^3 and 10^4 timesteps. The coefficients of the equations of optimal prediction were also estimated exactly as in [1, 4, 5].

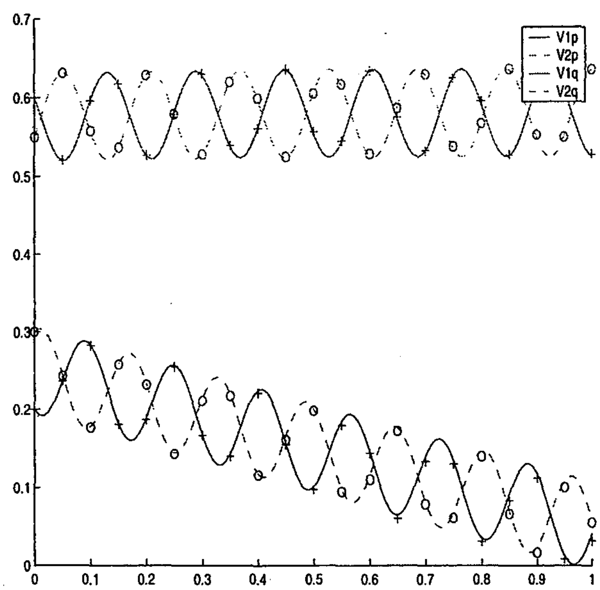


Figure 9: Nonlinear system: the original experiment

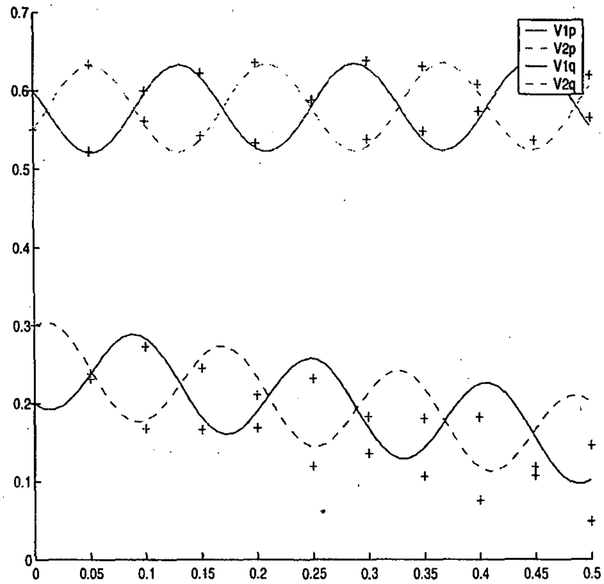


Figure 10: Nonlinear system with less concentrated prior

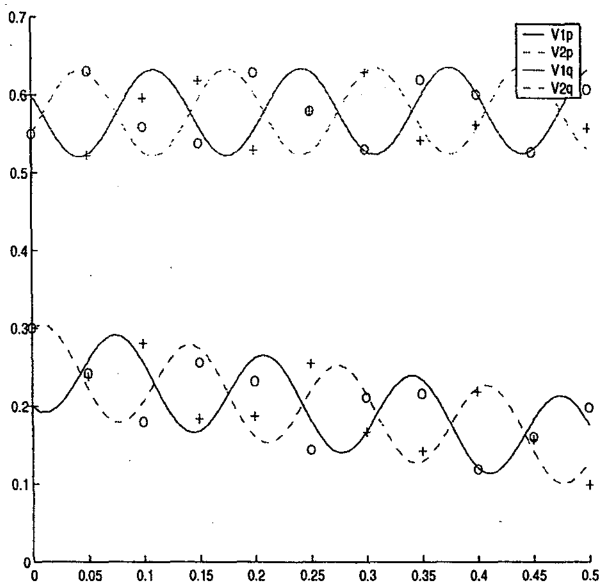


Figure 11: Nonlinear system (i)

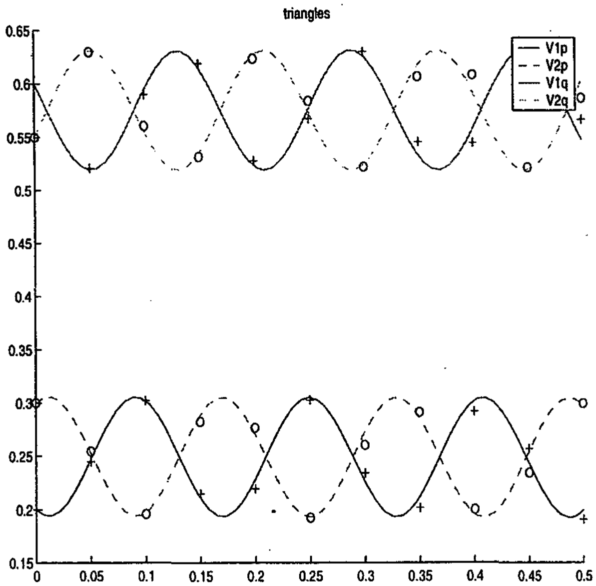


Figure 12: Nonlinear system (ii)

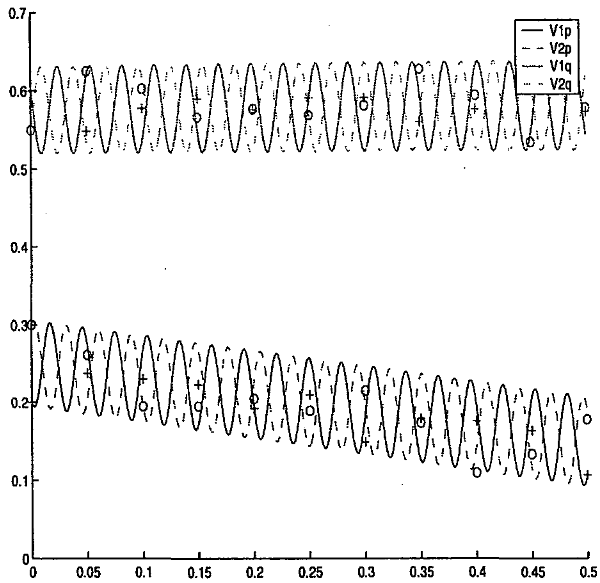


Figure 13: Nonlinear system (iv)

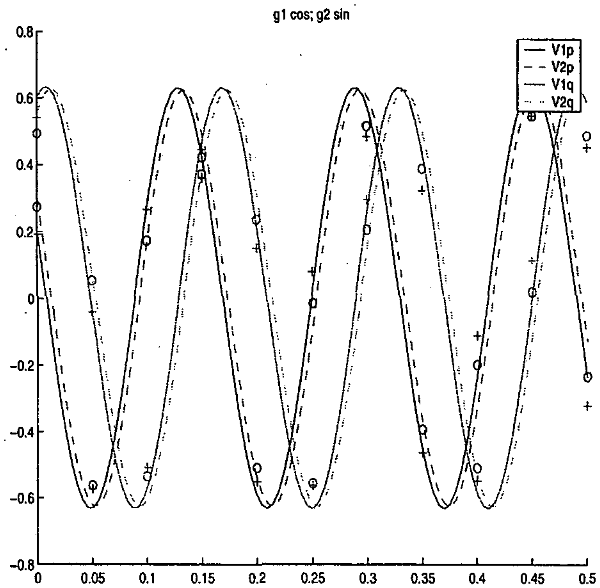


Figure 14: Nonlinear system (v)

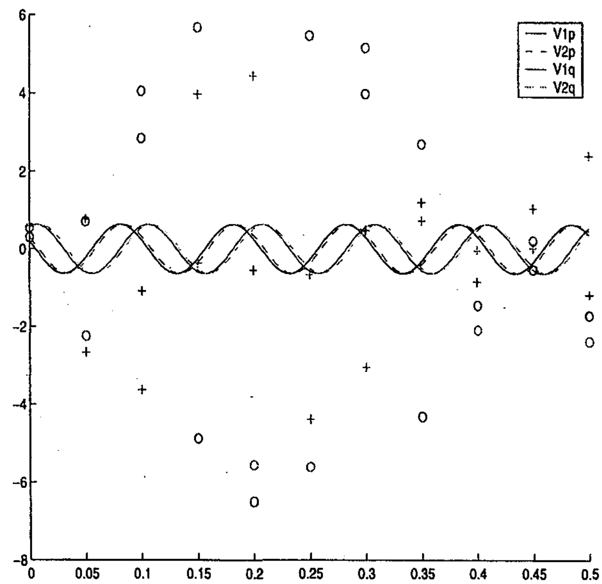


Figure 15: Nonlinear system (vi)

3.3 A linear PDE

In [1, 4, 5], first-order optimal prediction is applied to the partial differential equations

$$\begin{aligned} \frac{\partial}{\partial t} q(x,t) &= +\frac{\partial^2}{\partial x^2} p(x,t) - p(x,t) \\ \frac{\partial}{\partial t} p(x,t) &= -\frac{\partial^2}{\partial x^2} q(x,t) + q(x,t) \end{aligned} \tag{15}$$

on $[0, 2\pi]$ with periodic boundary conditions. Ten collective variables are selected, namely,

$$\begin{aligned} &\int q(x) g(x - 2\pi j/5) dx && \text{and} \\ &\int p(x) g(x - 2\pi j/5) dx, && j = 0, 1, 2, 3, 4; \end{aligned} \tag{16}$$

where

$$g(x) = \sum_{k \in \mathbb{Z}} e^{-\pi^2 k^2/25} e^{ikx} \qquad (i = \sqrt{-1}). \tag{17}$$

The prior measure selected is a certain Gaussian measure on $C([0, 2\pi])^2$ that resembles the law of circular Brownian motion. Because the prior measure is not supported on differentiable functions, (15) must be interpreted in the weak sense.

A numerical experiment shows that the the conditional expectations of the collective variables (16) at time t given their initial values are predicted well, at least up until $t = 2$.

4 Interpretation of the experiments

We performed many other experiments not reported here, whose results make it clear that first-order optimal prediction fails for "generic" collective variables. The results we presented in the preceding section were chosen to help us assess *why* first-order prediction works in the cases where it has been found to work.

4.1 The coupled harmonic oscillators

The span of a set of collective variables of the form $q_i, p_i, \ldots, q_{i+10}, p_{i+10}$ is nearly invariant under the dynamics (8) (9). Experiments (i) through (iv) show that first-order prediction works well for these collective variables. Experiment (iii) shows that "fast variables" can work as well as "slow variables." Experiments (v)-(viii) show how first-order prediction fails as the collective variables move farther and farther from invariance. The last two experiments also show that fast variables work as well as slow variables, as long as the collective variables are nearly invariant.

4.2 The nonlinear system of ODEs

Experiments (i),(iv) and (vi) show that first-order prediction does not work for some reasonable choices of the collective variables.

First-order prediction works well in experiments (iii) and (v) because the span of the collective variables is invariant under the linear part of (12), which dominates the dynamics. The linear part of the dynamics is diagonalized by functions of the form $\mathbf{f}_k \cdot (\mathbf{q} + i\mathbf{p})$ and $\mathbf{f}_k \cdot (\mathbf{q} - i\mathbf{p})$, where $\mathbf{f}_k$ is a discrete (sixteen point) Fourier vector of the form $\exp(-i2\pi jk/16)$; for $k = -7, -6, \ldots, 7, 8$. Since the eigenvalues of $\mathbf{f}_k \cdot (\mathbf{q} + i\mathbf{p})$ and $\mathbf{f}_k \cdot (\mathbf{q} - i\mathbf{p})$ are the same as the eigenvalues of $\mathbf{f}_{-k} \cdot (\mathbf{q} + i\mathbf{p})$ and $\mathbf{f}_{-k} \cdot (\mathbf{q} - i\mathbf{p})$, respectively, any function of the form $(c_1\mathbf{f}_k + c_2\mathbf{f}_{-k}) \cdot (\mathbf{q} + i\mathbf{p})$ or $(c_1\mathbf{f}_k + c_2\mathbf{f}_{-k}) \cdot (\mathbf{q} - i\mathbf{p})$ is invariant under the linear part of the dynamics. Thus, the collective variables of experiments (iii) and (v) span invariant subspaces for the linear part of (12).

Experiment (iii) indicates that collective variables can have disjoint supports and be fast variables; as long as they are invariant, first-order prediction can work.

Figure 10 shows that it helps somewhat to have a concentrated prior.

The success of first-order prediction in the original experiment and in experiment (ii) is a bit mysterious (see Figures 9 and 10) because the collective variables are not very close to spanning an invariant subspace for the linear part of (12), though they are still close. How close they come to invariance is indicated by the graphs of the discrete Fourier transforms of $\mathbf{g}_1 + \mathbf{g}_2$ and $\mathbf{g}_1 - \mathbf{g}_2$ in Figures 16 - 18. For example, Figures 17 and 18 show that $(\mathbf{g}_1 - \mathbf{g}_2) \cdot (\mathbf{q} + i\mathbf{p})$ is nearly invariant under the linear part of (12), but that $(\mathbf{g}_1 + \mathbf{g}_2) \cdot (\mathbf{q} + i\mathbf{p})$ is not as nearly invariant. (These Fourier transforms are real since $\mathbf{g}_1$ and $\mathbf{g}_2$ are even. The central bar represents the constant component; the outer bars represent higher frequency components.

The sum of $\mathbf{g}_1$ and $\mathbf{g}_2$ for experiment (iii) is constant, so the graph of its Fourier transform is omitted.)

4.3 The linear PDE

The span of the collective variables is very nearly invariant, for the sum (17) that defines the collective variables is very nearly equal to

$$\sum_{k=-2}^{2} e^{-\pi^2 k^2/25} e^{ikx} \qquad (i = \sqrt{-1}). \tag{18}$$

If the sum in (17) were *exactly* equal to (18), the span of the collective variables would be exactly invariant.

4.4 Yet another experiment

Another numerical test of optimal prediction is reported in the last section of [6]. First-order optimal prediction was applied to a nonlinear system of ODEs resembling (12), and several low wavenumber Fourier coefficients were used as collective variables. The first-order prediction of these Fourier coefficients agrees quite well with their conditional expectations given their initial values. Once again, the success of first-order prediction is attended by the circumstance of quasi-invariance of the collective variables.

It may seem that the success of first-order prediction in this case cannot be due to the quasi-invariance of the span of the Fourier coeficients, *for the approximate evolution of those Fourier coefficents given by the Galerkin scheme deviates substantially from the optimally predicted evolution.* We still maintain that first-order pediction works here because the collective variables are quasi-invariant, that is, because some change of variables that renames the collective variables $\mathbf{u}$ renders the dynamics are approximately of the form (6), so that approximately

$$\frac{d}{dt}\mathbf{u} = \mathbf{F}(\mathbf{u})$$

for some (unknown) function $\mathbf{F}$. The fact that the optimal predictions are not the predictions obtained from the Galerkin scheme shows only that $\mathbf{F}$ does not happen to be the one suggested by Galerkin projection. The fact that the collective variables are invariant under the dominating linear part of the dynamics suggests there might be *some* $\mathbf{F}$ that makes the decomposition (6) of the dynamics hold approximately. There is no reason $\mathbf{F}$ must have the form indicated by the Galerkin scheme, while the method of first-order prediction provides, in effect, a better choice of $\mathbf{F}$.

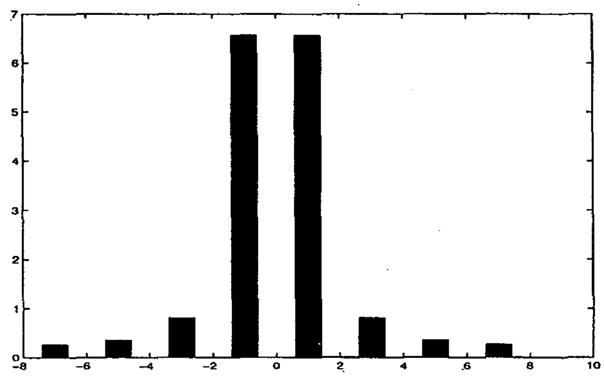


Figure 16: Fourier transform of $\mathbf{g}_1 - \mathbf{g}_2$ of experiment (ii)

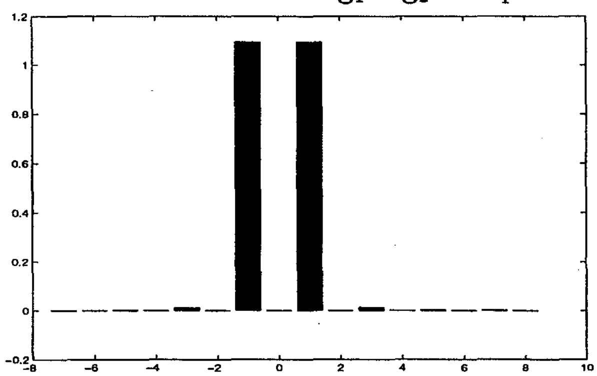


Figure 17: Fourier transform of $\mathbf{g}_1 - \mathbf{g}_2$ of original experiment

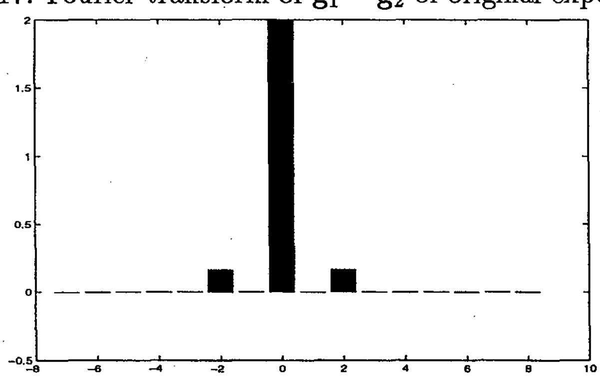


Figure 18: Fourier transform of $\mathbf{g}_1 + \mathbf{g}_2$ of original experiment

5 Conclusions

First-order optimal prediction generically fails because the closure assumption is rarely valid, but it can succeed when the set of functions of the collective variables is nearly invariant under the dynamics. Other circumstances thought to favor the success of the method seem to be much less important than the quasi-invariance of the collective variables.

6 Acknowledgments

The authors would like to thank Mr. C. Cameron, Prof. A.J. Chorin, Prof. O.H. Hald, Dr. A. Kast, Prof. D. Kessler, Dr. I. Kliakhandler, Dr. R. Kupferman, Dr. D. Levy, and Mr. K. Lin for helpful comments and suggestions.

References

[1] A. J. Chorin, A. P. Kast, and R. Kupferman. Optimal prediction of underresolved dynamics. *Proceedings of the National Academy of Sciences.* **95**: 4094 - 4098, 1998.

[2] O. H. Hald. Unpublished manuscript.

[3] O. H. Hald. Optimal prediction and the Klein-Gordon equation. *Proceedings of the National Academy of Sciences.* **96**: 4774 - 4779, 1999.

[4] A. J. Chorin, A. P. Kast, and R. Kupferman. Unresolved computation and optimal prediction. *Communications in Pure and Aplied Mathematics.* In press, 1999.

[5] A. J. Chorin, A. P. Kast, and R. Kupferman. On the prediction of large-scale dynamics using unresolved computations. *Contemporary Mathematics.* In press, 1999.

[6] A. J. Chorin, R. Kupferman, and D. Levy. Optimal prediction and perturbation theory. Preprint.

[7] A. P. Kast. Optimal prediction of stiff oscillatory systems. Preprint.

ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY
ONE CYCLOTRON ROAD | BERKELEY, CALIFORNIA 94720

Prepared for the U.S. Department of Energy under Contract No. DE-AC03-76SF00098