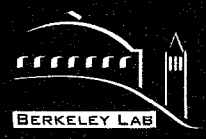
LBNL-44236



ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

The Optimal Prediction Method

Thibaut Burin des Roziers

Computing Sciences Directorate
Mathematics Department

August 1999

LBNL-44236

THE OPTIMAL PREDICTION METHOD*

Thibaut Burin des Roziers

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.

THE OPTIMAL PREDICTION METHOD

by Thibaut Burin des Roziers.
3eme annee Ensimag.
Option MCS.

Contents

1 Introduction

The prupose of the work is to test and show how well the numerical method called Optima Prediction works. This method is relatively new and only a few experiment have been made. We first did a series of simple tests to see how the method behaves. In order to have a better understanding of the method, we then reproduced one of the main experiment which was done about Optimal Prediction by Kupferman. Once we obtained the same results that kupferman had, we changed a few parameters to see how dependant the method was on this parameters. In this paper, I will present all the tests we made, the results we obtained and what we concluded about the method. Before talking about the experiments, I have to explain what is the Optimal Prediction method and how does it work. This will be done in the first section of this paper.

2 Presentation of the Optimal Prediction Method

2.1 Main Idea

The goal of this method is to be able to solve problems which can not be properly resolved because the solutions of the equations of the problem are too complicated. Optimal Prediction should also help to solve problem which are so large that they can not be solved with the actual technology. Here is the main idea. Instead of solving a particular initial value problem, we try to find the average of all solutions that satisfy certain constraints at time t=0.The constraints may be local averages of the initial data or a small number of Fourier coefficient. We then used some statistical information to compensate for the incompletness of the initial data. With this method, the number of variables to find is then a lot smaller than in the initial problem.

2.2 A Detailled Presentation

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 $g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_k(\mathbf{x})$ 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 $\theta = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k$ let Q_θ denote the conditional probability under Q given that $g_i(\mathbf{x}) = \theta_i$ for $i = 1, 2, \ldots, k$. That is

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

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* $g_1(\mathbf{x}), \ldots, g_k(\mathbf{x})$. They may be called "first-order" because they do correctly predict $\mathbb{E}_{P^t}[g_i(\mathbf{x})]$ for infinitesimally small $t > 0$ to first-order, that is, if $P^0 = Q_{\theta(0)}$ then

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

3 Experiments

3.1 Some Simple Experiments

The first tests we did, seemed to indicate that this method wasn't very good. It simply doesn't work in some very simple cases. These simple experiments were done mostly with the use of Matlab.

3.1.1 Experiment Number 1

Here are the parameters for this experiment. We look at N=12 particules which are the solutions of the following system :

$$\frac{d\theta_i}{dt} = \sum_{j=1}^{N} \frac{\cos(\theta_i - \theta_j)}{N} \quad \text{i=1..N}$$

There are NS=3 optimal prediction variables defined by

$$\Phi_i = \theta_k \quad for\ i = 1..NS\ and\ k \in \{1..N\}$$

and answering to the following equations:

$$\frac{d\Phi_i}{dt} = E[\frac{d\theta_k}{dt} | \Phi_l \text{ ,l=1..NS}] \quad for\ i = 1..NS\ and\ k \in \{1..N\}$$

This is equivalent to

$$\frac{d\Phi_i}{dt} = E[\sum_{j=1}^{N} \frac{\cos(\theta_k - \theta_j)}{N} | \Phi_l \text{,l=1..NS}] \quad for\ i = 1..NS\ and\ k \in \{1..N\}$$

$$\iff \frac{d\Phi_i}{dt} = \sum_{j=1}^{N} E[\frac{\cos(\theta_k - \theta_j)}{N}] | \Phi_l \text{,l=1..NS}] \quad for\ i = 1..NS\ and\ k \in \{1..N\}$$

$$\iff \frac{d\Phi_i}{dt} = \sum_{j=1}^{NS} \frac{\cos(\Phi_j - \Phi_i)}{N} + \sum_{j=NS+1}^{N} E[\frac{\cos(\theta_j - \Phi_i)}{N}]$$

$$\iff \frac{d\Phi_i}{dt} = \sum_{j=1}^{NS} \frac{\cos(\Phi_j - \Phi_i)}{N}$$

As we can see, the Optimal Prediction equations are exactly the same as the ones of the former system. In both cases, we used a fourth order Runge-Kutta algorithm to solve these systems. We had different kind of results. With N=12 and NS=3, the solutions of the Optimal Prediction system are similar to the solutions of the former system for about one second. This time increases if we increase N. However, it does not increase linearly. If NS = 1, then the methods of Optimal Prediction works perfectly: the solutions are exactly the same for both systems.

3.1.2 Experiment Number 2

In this experiment, we consider the same system as in experiment number 1. We just add the following term to the equations : $\sigma * dt^{(1/2)} * z_i$ where z_i is a random variable. We then have to solve a stochastic equation. The system becomes

$$\frac{d\theta_i}{dt} = \sum_{j=1}^{N} \frac{\cos(\theta_i - \theta_j)}{N} + \sigma dt^{1/2} z_i \quad \text{i=1..N}$$

However, the results were the same as in the first experiment.

3.1.3 Experiment number 3: coupled harmonic oscillator

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

$$\begin{aligned}
\dot{Q} &= P \\
\dot{q_j} &= j^2 p_j \ ; \quad j = 1, 2, \ldots, 1000 \\
\dot{P_j} &= -Q + \sum_{i=1}^{1000} (q_i - Q) \\
\dot{p_j} &= (Q - q_j) + \sum_{i=1}^{1000} (q_i - q_j) \ ; \quad j = 1, 2, \ldots, 1000.
\end{aligned} \tag{6}$$

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 (6) 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

$$\dot{q_j} = j^2 p_j \ ; \quad j = 1, 2, \ldots, 100 \tag{7}$$

$$\dot{p_j} = q_i + \sum_{i=1}^{100} (q_i - q_j) \ ; \quad j = 1, 2, \ldots, 100. \tag{8}$$

(9)

As our prior measure Q we took

$$\frac{1}{Z} \exp\left(-\frac{1}{2} \sum_{j=1}^{100} j^2 p_j^2 + \frac{1}{4} \sum_{j,k=1}^{100} (q_j - q_k)^2 + \frac{1}{2} \sum_{j=1}^{100} q_j^2 \right) d\mathbf{q} d\mathbf{p}. \tag{10}$$

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 - 5. 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 (7) and in the fromula for the density of the prior measure (10). In the second experiment, the masses were all taken to be 1, so we replaced j^2 by 1 in (7) and (10). The results of these experiments are graphed in Figures 6 and 7.

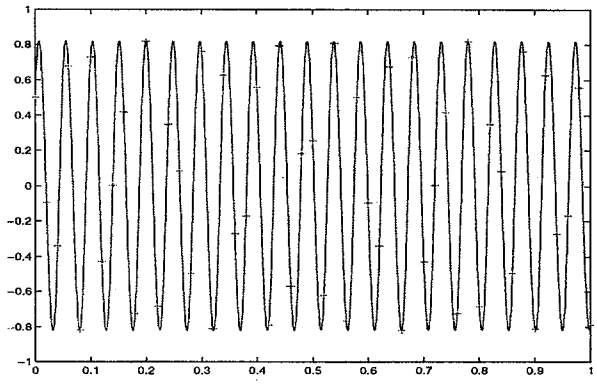


Figure 1: Coupled oscillators (ii)

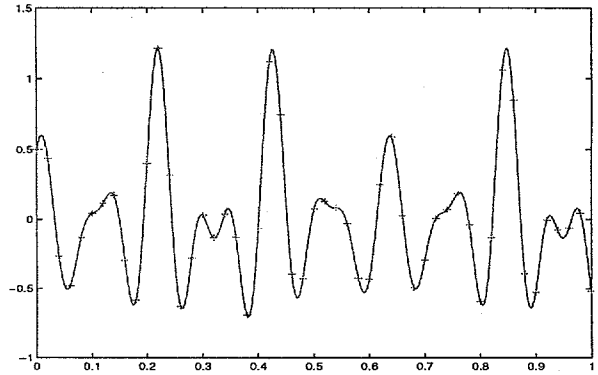


Figure 2: Coupled oscillators (iv)

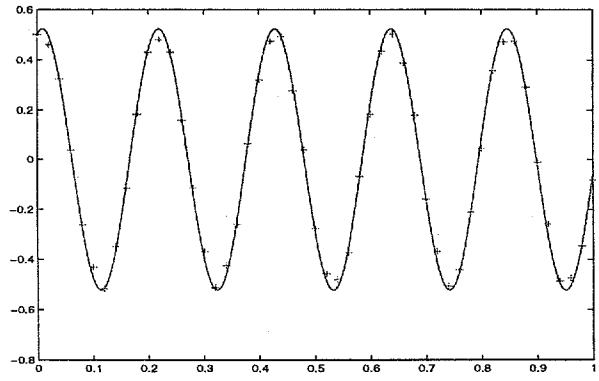


Figure 3: Coupled oscillators (v)

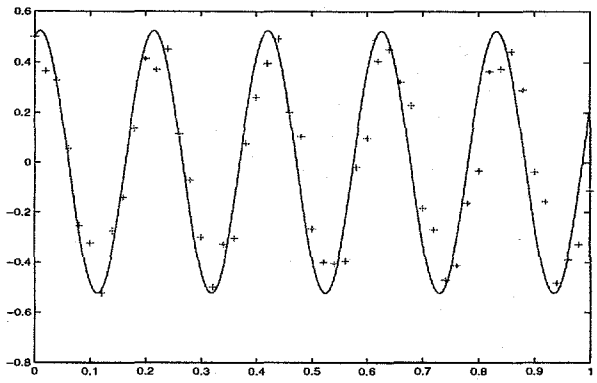


Figure 4: Coupled oscillators (vi)

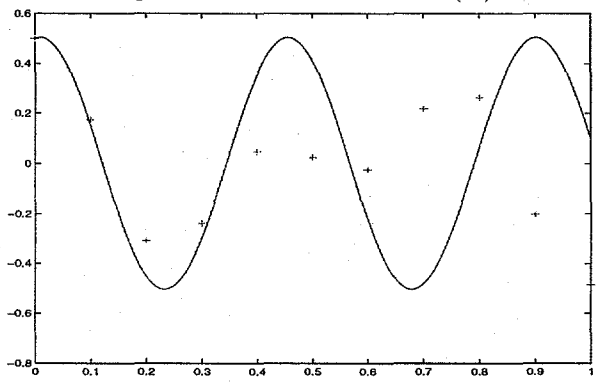


Figure 5: Coupled oscillators (viii)

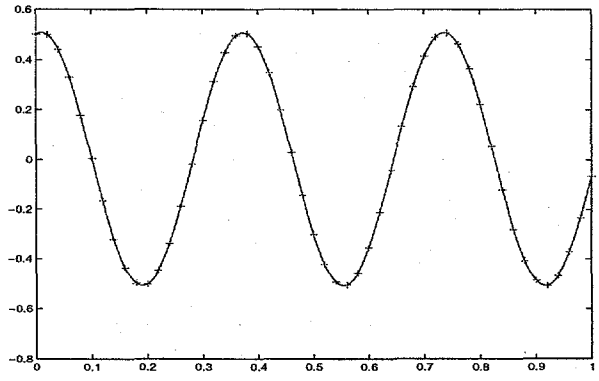


Figure 6: Oscillators with masses $1/j$

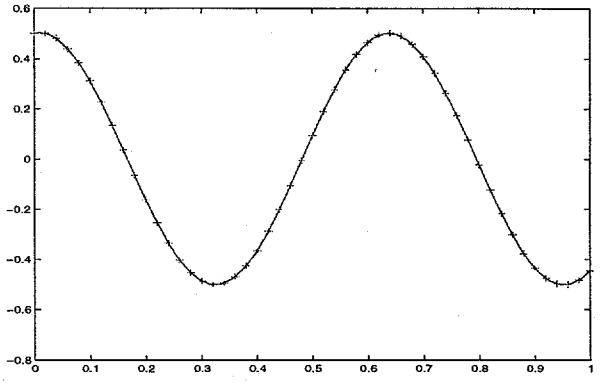


Figure 7: Oscillators with masses 1

3.2 A Nonlinear Hamiltonian System

3.2.1 Reproducing the Experiment

We then decided to do an experiment that was done a few months ago and which seems to validate the method. We wanted to understand why the method works for this test and doesn't work in simple cases. We first thought that by changing the value of some parameters, we could show that the method is not reliable.

We study a finite dimensional system of 2n (n=16) ordinary differential equations:

$$\begin{cases} \frac{dp(j)}{dt} = -\frac{q(j-1)-2q(j)+q(j+1)}{\Delta x^2} + q^3(j) \\ \frac{dq(j)}{dt} = +\frac{p(j-1)-2p(j)+p(j+1)}{\Delta x^2} - p^3(j) \end{cases}$$

for j=1,..,n and where $\Delta x = \frac{1}{n}$ We enforce periodicity with $p(0) \equiv p(n)$, $p(n+1) \equiv p(1)$, etc. This system admits the following Hamiltonian,

$$H[p,q] = \frac{1}{2}\sum_{i=1}^{n}\left\{\left[\frac{p(j+1)-p(j)}{\delta x}\right]^2 + \left[\frac{q(j+1)-q(j)}{\delta x}\right]^2 + \frac{1}{2}\left[p^4(j)+q^4(j)\right]\right\}$$

Therefore, we can defined the canonical density of the equations by

$$f_0(p,q) = \exp(-H(p,q))$$

which we assume to be the probability density.

The collective variables we want to study are of the form

$$U_\alpha^p[p,q] = (g_\alpha(.),p(.)) \equiv \sum_{j=1}^{n} g_\alpha(j)p(j)$$
$$U_\alpha^q[p,q] = (g_\alpha(.),q(.)) \equiv \sum_{j=1}^{n} g_\alpha(j)q(j)$$

with $\alpha = 1, .., N(N = 2)$ and where the g's represent the discrete kernels given by

$$g_1(j) = \frac{1}{Z} \exp(-\frac{d^2(1,j)}{n^2\sigma^2})$$
$$g_2(j) = \frac{1}{Z} \exp(-\frac{d^2(9,j)}{n^2\sigma^2})$$

where Z is a normalizing constant and $d(j_1, j_2)$ is a distance function which gives the minimun of $|j_1 - j_2|, |j_1 - j_2 - n|, |j_1 - j_2 + n|$.

Applying the optimal method to the above formulas, we find four differential equations for the four collective variables.

$$\frac{dV_1^p}{dt} = 19.5(V_2^q - V_1^q) - \left[1.50(V_1^q)^3 - 0.88(V_1^q)^2 V_2^q + 0.27 V_1^q (V_2^q)^2 + 0.11(V_2^q)^3\right]$$
$$\frac{dV_2^p}{dt} = 19.5(V_1^q - V_2^q) - \left[1.50(V_2^q)^3 - 0.88(V_2^q)^2 V_1^q + 0.27 V_2^q (V_1^q)^2 + 0.11(V_1^q)^3\right]$$
$$\frac{dV_1^q}{dt} = -19.5(V_2^p - V_1^p) + \left[1.50(V_1^p)^3 - 0.88(V_1^p)^2 V_2^p + 0.27 V_1^p (V_2^p)^2 + 0.11(V_2^p)^3\right]$$
$$\frac{dV_2^q}{dt} = -19.5(V_1^p - V_2^p) + \left[1.50(V_2^p)^3 - 0.88(V_2^p)^2 V_1^p + 0.27 V_2^p (V_1^p)^2 + 0.11(V_1^p)^3\right]$$

We used the same technics and the same algorithms that were formerly used. Once, we obtained exactly the same results (figure 8, 9, we were able to perform a few changes in order to see the limitation of the method.

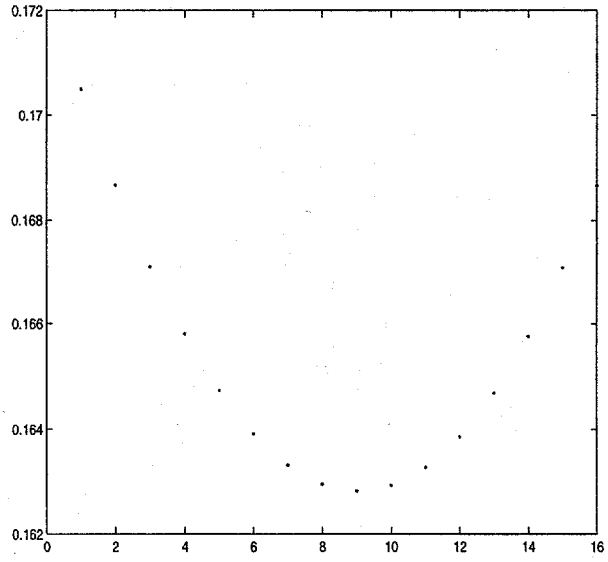


Figure 8: The covariances $< p(i)p(j) >=< q(i)q(j) >$ as a function of the grid separation i-j. these values were computed by a Monte Carlo algorithm

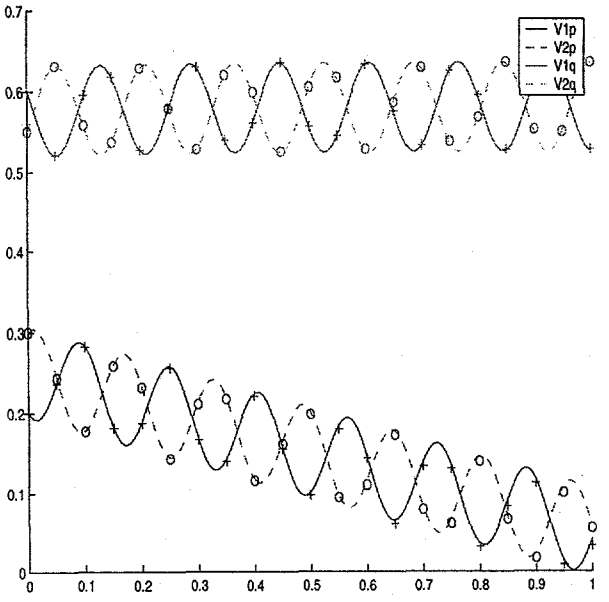


Figure 9: Evolution in time of the mean value of the four collectives variables: $V_1^p(+), V_2^p(o), V_1^q(+), V_2^q(o)$. These values were obtained by solving 32 equations for 500 initial conditions and averaging.

3.2.2 Changing some Parameters

There were several changes that we wanted to make in order to test the Optimal Prediction method. Most of these changes concerned the choice of the kernels that determinethe constraints. First we only changed the variance of the kernels. We increased it to 0.9 (instead of 0.25). We can see how the method reacts to that change on figure 10. The method works very well.

We then decided to change the hamiltonian. In fact, we only multiply the old one by Δx. This has the effect of making the prior measure about four times as spread out.

Then, we decided to see what was the influence of the kernels. First, we ran the experiment with the kernels defined below:

$$g_1(j) = \begin{cases} 1 & \text{if } 1 \leq j < 9 \\ 0 & \text{if } j \geq 9 \end{cases}$$
$$g_2(j) = \begin{cases} 0 & \text{if } 1 \leq j < 9 \\ 1 & \text{if } j \geq 9 \end{cases}$$

Figure 12 shows the results of this experiment. We can see in that case that the method doesn't work. We then tried a lot of others kernels. In most cases, we could see that the results of the Optimal prediction method did not correspond to the real solution. Here are some of the kernels we tried:

(i)
$$g_1(j) = \begin{cases} (j-1)/8 & \text{if } 1 \le j \le 9 \\ (9-j)/8 & \text{if } j \ge 9 \end{cases}$$
$$g_2(j) = \begin{cases} (17-j)/8 & \text{if } 1 \le j \le 9 \\ (j-9)/8 & \text{if } j \ge 9 \end{cases}$$

(ii)
$$g_1(j) = (j-1)/15 \quad for j \in \{1..16\}$$
$$g_2(j) = (16-j)/8 \quad for j \in \{1..16\}$$

(iii)
$$g_1(j) = \begin{cases} 1 & \text{if } j \bmod 2 = 0 \\ 0 & \text{if } j \bmod 2 \neq 0 \end{cases}$$
$$g_2(j) = \begin{cases} 0 & \text{if } j \bmod 2 = 0 \\ 1 & \text{if } j \bmod 2 \neq 0 \end{cases}$$

(iv)
$$g_1(j) = \begin{cases} 1 & \text{if } j \in \{1,4,5,6,7,11,,13,16\} \\ 0 & if not \end{cases}$$
$$g_2(j) = \begin{cases} 0 & \text{if } j \in \{2,6,7,8,9,10,14,16\}0 \\ 1 & if not \end{cases}$$

(v)
$$g_1(j) = \cos(2\pi j/N) + \cos(4\pi j/N) \quad \text{for } j = 1..16$$
$$g_2(j) = \sin(2\pi j/N) + \sin(4\pi j/N) \quad \text{for } j = 1..16$$

(vi)
$$g_1(j) = \cos(2\pi j/N) \quad \text{for } j = 1..16$$
$$g_2(j) = \sin(2\pi j/N) \quad \text{for } j = 1..16$$

Figures 12 through 18 show the results for these sets of kernels. In most cases, all the results seem to indicate the method doesn't work. we can see that the optimal prediction solution (plain lines) do not correspond with the real solution (cross and circle).

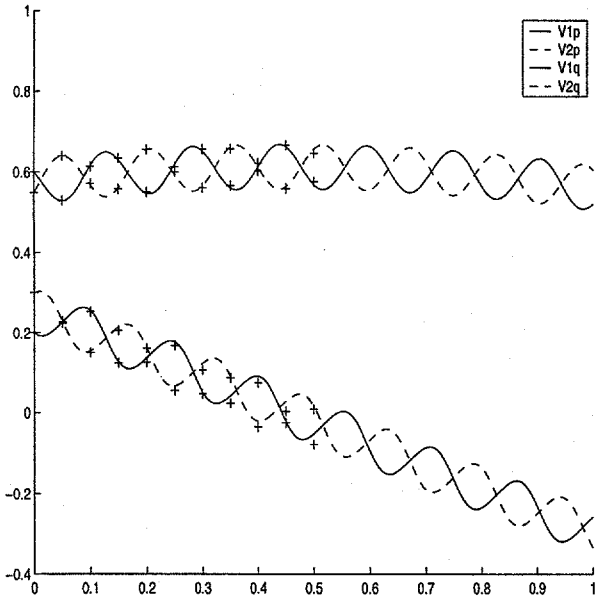


Figure 10: Evolution in time of the mean value of the four collectives variables: $V_1^p(+), V_2^p(o), V_1^q(+), V_2^q(o)$. These values were obtained by solving 32 equations for 100 initial conditions and averaging. Here, we change sigma to 0.9

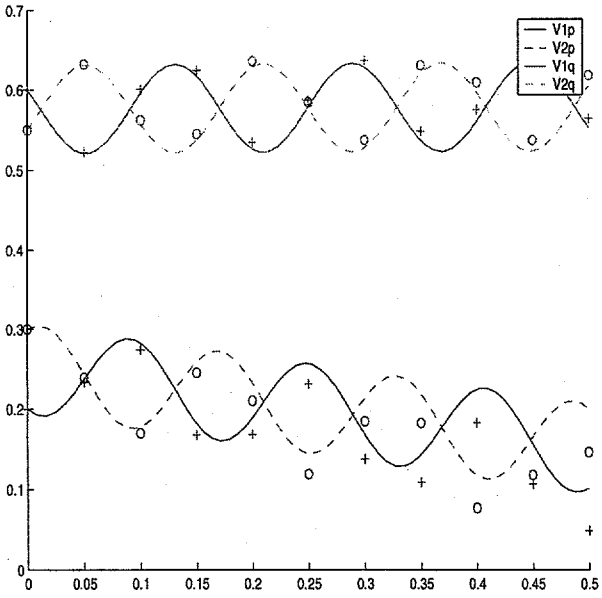


Figure 11: Evolution in time of the mean value of the four collectives variables: $V_1^p(+), V_2^p(o), V_1^q(+), V_2^q(o)$. These values were obtained by solving 32 equations for 500 initial conditions and averaging.We used for the hamiltonian $G = \Delta x * H$ where H is the former hamiltonian

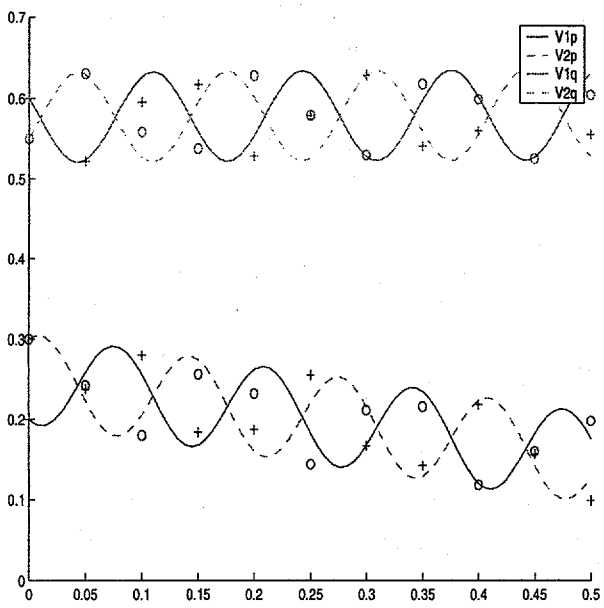


Figure 12: Evolution in time of the mean value of the four collectives variables: $V_1^p(+), V_2^p(o), V_1^q(+), V_2^q(o)$. The kernels were changed to "step kernels"

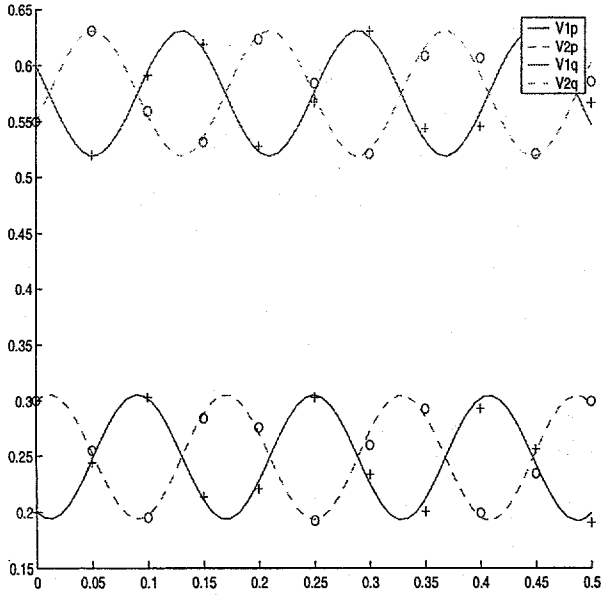


Figure 13: Evolution in time of the mean value of the four collectives variables: $V_1^p(+), V_2^p(o), V_1^q(+), V_2^q(o)$. The kernels were changed to "triangle kernels"

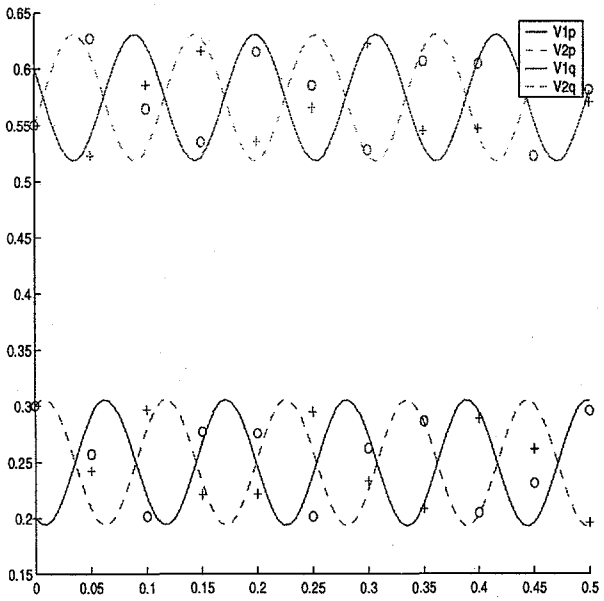


Figure 14: Evolution in time of the mean value of the four collectives variables: $V_1^p(+), V_2^p(o), V_1^q(+), V_2^q(o)$.

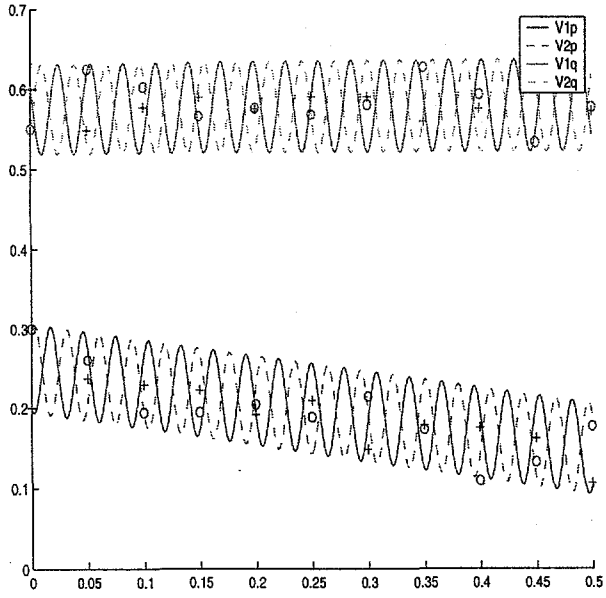


Figure 15: Evolution in time of the mean value of the four collectives variables: $V_1^p(+), V_2^p(o), V_1^q(+), V_2^q(o)$. The kernels were changed to "random kernels"

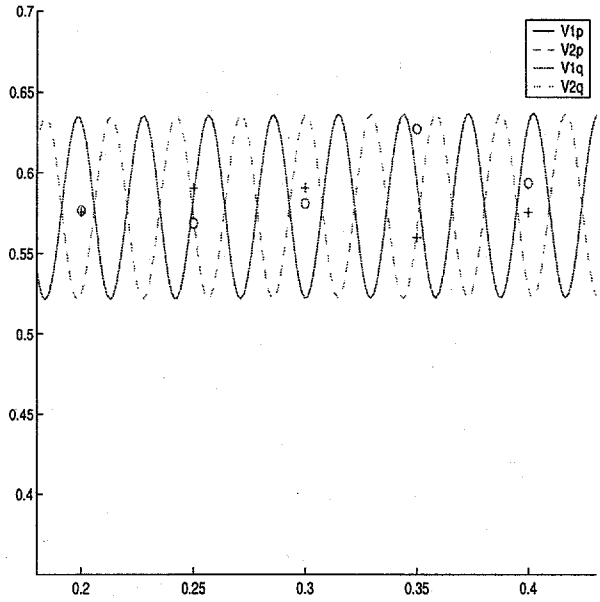


Figure 16: Evolution in time of the mean value of the four collectives variables: $V_1^p(+), V_2^p(o), V_1^q(+), V_2^q(o)$. The kernels were changed to "random kernels".This is a zoom of the precedent figure

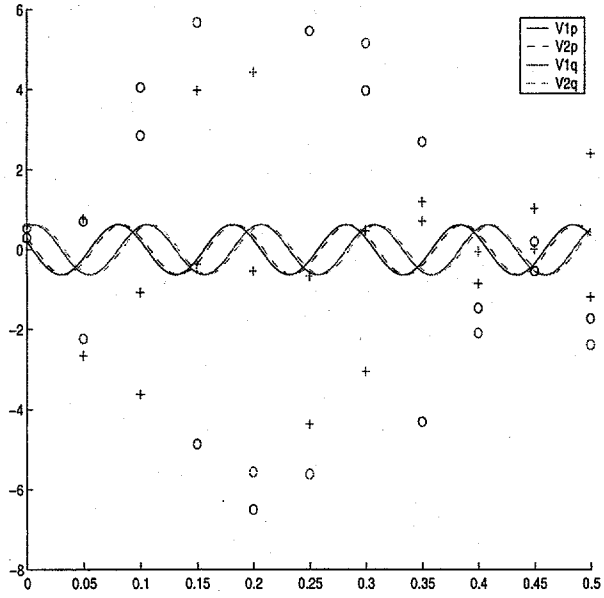


Figure 17: Evolution in time of the mean value of the four collectives variables: $V_1^p(+), V_2^p(o), V_1^q(+), V_2^q(o)$. The kernels were changed to "a sum of cosinus and sinus kernels"

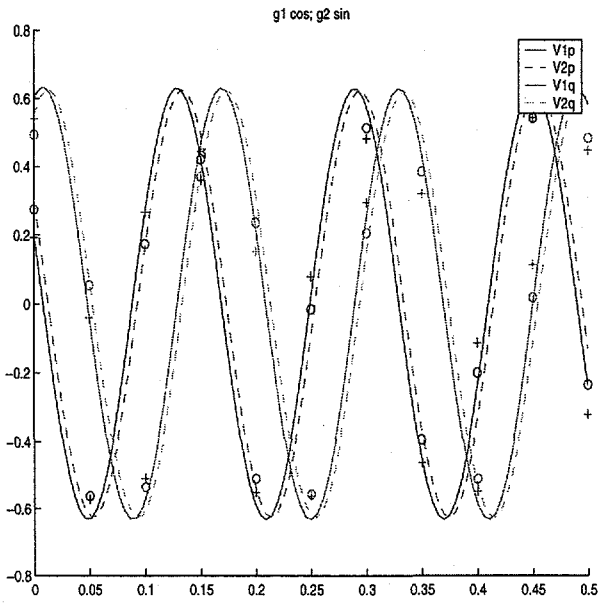


Figure 18: Evolution in time of the mean value of the four collectives variables: $V_1^p(+), V_2^p(o), V_1^q(+), V_2^q(o)$. The kernels were changed to "cosinus and sinus kernels"

4 Conclusion on the Experiments and on the method

All these experiments show that the method isn't working right now. A lot of work needs to be done. But why does Optimal Prediction give such good results with some of the experiments?

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

$$\begin{aligned} \frac{d}{dt}u_i &= F_i(\mathbf{u}) \; ; \quad \mathbf{u} = (u_1, u_2, \ldots, u_k) \\ \frac{d}{dt}w_i &= G_i(\mathbf{u}, \mathbf{w}) \; ; \quad \mathbf{w} = (w_1, w_2, \ldots, w_{n-k}) \end{aligned} \tag{11}$$

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 (11) 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 (11) 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.

4.0.3 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 (7) (8). 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) 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.

4.0.4 The Non Linear Hamiltonian

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 (3.2.1), which dominates the dynamics. The linear part of the dynamics is diagonalized by functions of the form $\mathbf{f} \cdot (\mathbf{q} + i\mathbf{p})$ and $\mathbf{f} \cdot (\mathbf{q} - i\mathbf{p})$, where $\mathbf{f}_k$ is a discrete (sixteen point) Fourier vector of the form $\exp(-i2\pi jk/16)$; for $k = -7, -6, \ldots, 7, 8$. Thus, the collective variables of experiments (iii) and (v) span invariant subspaces for the linear part of (3.2.1).

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 11 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 11) because the collective variables are not very close to spanning an invariant subspace for the linear part of (3.2.1), 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 19 - 21. (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.)

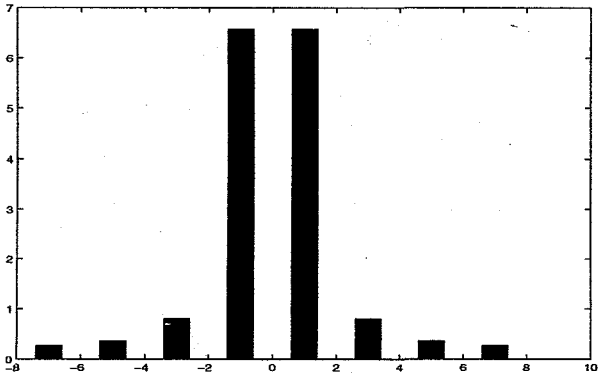


Figure 19: fourier transform of g2-g1 with the triangle kernels

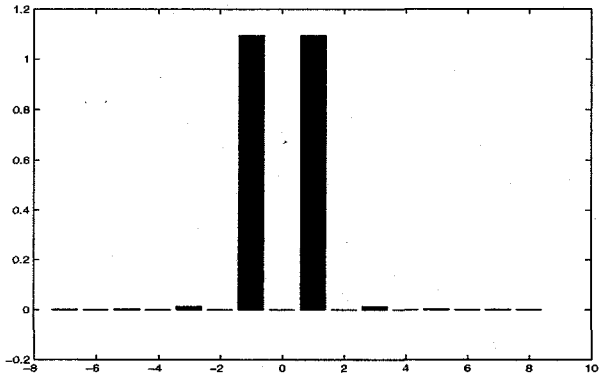


Figure 20: fourier transform of g1-g2 of the former system(ii)

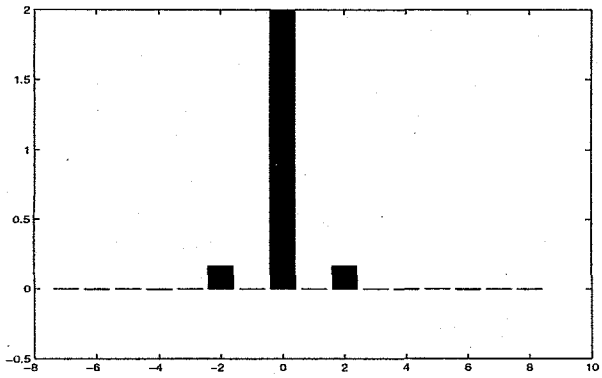


Figure 21: fourier transform of g1+g2 of the former system

5 Conclusion

This internship was very interesting for several reasons. I had the chance to work in both the mathematical field and the computing field. On the mathematical field, i was able to learn some new skill in the statistical domain and to increase my knowledge on how to solve Partial Differential Equations. As for the computation part, I now have a better grasp ont the C language. I was also initiated to C++, which could prove to be very helpfull for my first employement. I discovered some new software (Latex) and was able to perfect my knowledge of Matlab. But, what was the most interesting aspect of this internship, was the fact that I could see and help on the developemnt on a new numerical method. I could see all the different steps which are used in order to create such a method. Waiting for the results of an experiment was exciting since we didn't know what to expect exactly. Doing this internship abroad helped me to perfect my english and allowed to work in a different atmosphere than the one I was used to. There is no doubts that these two aspects will be very valuable for my first employement.

6 Acknowledgement

I'd like to thank all the people who work in the Mathematics Department at the Lawrence Berkeley Laboratory. I'm particurlarly gratefull toward Prof. Alexandre Chorin who gave me the opportunity to work in this departement. I also want to thank Alex Gottlieb for all the help he gave me during these six months. I'd like to mention Prof. Doron Levy, Raz kupferman, Andreas Wiegmann, Chris Cameron, Valery Heatlie and all the others members of the departement for their help and their friendship.

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.