Non-Markovian Optimal Prediction

Alexandre J. Chorin and Ole H. Hald
Department of Mathematics
Lawrence Berkeley National Laboratory
1 Cyclotron Road
Berkeley CA 94720

Raz Kupferman
Institute of Mathematics
The Hebrew University
Jerusalem, Israel

Abstract

Optimal prediction methods compensate for a lack of resolution in the numerical solution of complex problems through the use of prior statistical information. We know from previous work that in the presence of strong underresolution a good approximation needs a non-Markovian "memory", determined by an equation for the "orthogonal", i.e., unresolved, dynamics. We present a simple approximation of the orthogonal dynamics, which involves an ansatz and a Monte-Carlo evaluation of autocorrelations. The analysis provides a new understanding of the fluctuation-dissipation formulas of statistical physics. A simple example is given.

1 Introduction

Many problems in science and engineering are described by nonlinear differential equations whose solutions are too complicated to be properly resolved. The problem of predicting the evolution of systems that are not well resolved has been addressed by the present authors and others in [?, ?, ?, ?, ?]. Nothing can be predicted without some knowledge about the unresolved ("subgrid") degrees of freedom. In the optimal prediction methods of [?, ?, ?, ?, ?] it is assumed that one possesses, as one often does, prior statistical information about the system in the form of an invariant measure; what is sought is a mean solution with respect to this prior measure, compatible with the information initially at hand as well as with the limitations on the computing power one can bring to bear.

The simplest version of this idea, Markovian optimal prediction, generates an approximating system of ordinary differential equations and works well for a time that depends on the degree of underresolution and on the uncertainty in the data. This version is optimal in the class of Markovian approximations, but it eventually exhibits errors, because the influence of partial initial data on the distribution of the solutions weakens in time if the system is ergodic, and this loss of information is not captured in full, see ([?, ?]). To obtain an accurate approximation of a subset of variables without solving the full problem requires the addition of a "memory" term, and the resulting prediction scheme becomes a generalized Langevin equation, similar to those in irreversible statistical mechanics ([]).

We present a general formalism for separating resolved and unresolved degrees of freedom, analogous to the nonlinear projection formalism of Zwanzig [] but using the language of probability theory. We find a solution of the equation for the orthogonal unresolved dynamics and find its statistics by Monte-Carlo integration; we use the results to construct a prediction scheme with memory. We apply the scheme to a simple model problem. In the conclusion we indicate how the construction is generalized to more complicated problems, and the new perspectives it opens for prediction as well as for irreversible statistical mechanics.

2 Linear and nonlinear projections of dynamical systems

Consider a problem of the form

$$\frac{d\varphi}{dt} = R(\varphi), \quad \varphi(0) = x, \tag{2.1}$$

where R and φ are n-dimensional vectors (n may be infinite), with components R_i and ϕ_i; t is time. When n is finite (2.1) is a system of ordinary differential equations. Our goal is to calculate the average values of m components of φ, $m < n$, , without calculating all the components; the average is over all the values that the missing, unresolved, components may assume; our prior information allows us to make statistical statements about these missing components.

We denote the phase space (the vector space in which φ resides) by $\mathbf{\Gamma}$; in classical statistical physics this phase space is the $n = 6\ell$ dimensional space of coordinates and momenta (q_i, p_i), where ℓ is the number of particles; the q_i, p_i at time t are then the entries of the vector φ. A solution of equation (2.1) is defined when an initial value $\varphi(t = 0) = x$ is given; to each initial condition $\hat{x}$ corresponds a trajectory, $\varphi(t) = \varphi(x, t)$; the initial value x is emphasized by this notation in view of its key role in what follows.

A phase variable u is a function on $\mathbf{\Gamma}$; u may be a vector, whose components are labeled as u_i. A phase variable varies when its argument varies in time, so that a phase variable whose value at $t = 0$ was $u(x)$ acquires at time t the value $u(\varphi(x, t))$. It will also be useful to examine the evolution of u in a more abstract setting: Introduce an evolution operator S^t for space variables by the relation

$$(S^t u)(x) = u(\varphi(x, t)). \tag{2.2}$$

Differentiation of (2.2) with respect to time yields

$$\frac{\partial}{\partial t}(S^t u)(x) = \sum_i R_i(\varphi(x, t)) \frac{\partial u}{\partial x_i}(\varphi(\hat{x}, t)) = LS^t u(\hat{x}), \tag{2.3}$$

where L, the Liouvillian, is the linear differential operator $L = \sum_i R_i(\varphi)\frac{\partial}{\partial \phi_i}$. Thus the phase variable $S^t u$ can be calculated in either of two ways: (i) for each x integrate to time t the equations of motion $\frac{d}{dt}\varphi(t) = R(\varphi(t))$ with initial conditions $\varphi(0) = x$ and evaluate the phase variable u at the point $\varphi(x,t)$; or (ii) solve the equation

$$\begin{cases} \dfrac{\partial}{\partial t} S^t u = L S^t u \\ S^0 u = u(x). \end{cases}$$

It is convenient to write $S^t = e^{tL}$; we do not inquire here as to the conditions under which this symbolic notation can be taken literally. The significant thing about equation (2.3) is that it is linear. For Hamiltonian systems L is antisymmetric.

One can check from the definitions that that $e^{tL}(uv) = (e^{tL}u)(e^{tL}v)$, $e^{tL}f(u) = f(e^{tL}u)$ for any function f, and $e^{tL}L = Le^{tL}$. In this notation, the symbol u refers to the data $u(x)$ at $t = 0$; the time dependence is described by the exponential.

Suppose that the initial data x are drawn from a probability distribution μ^0; each initial datum gives rise to a solution of equation (2.1) and the measure μ^0 evolves into a measure μ^t at time t. The evolution of μ^t is defined by the conditions

$$\int_{\mathbf{\Gamma}} u(\varphi(x,t))\,\mu^0(dx) = \int_{\mathbf{\Gamma}} u(x)\,\mu^t(dx)$$

for all sufficiently smooth phase variables u. We assume that the measure μ^0 is invariant under the flow (2.1): $\mu^t = \mu^0 = \mu$. Many systems have invariant measures, in particular, any Hamiltonian system with Hamiltonian H leaves invariant the canonical measure with density $Z^{-1}e^{-H(x)/T}$, where Z is a normalization constant and T is the variance of the samples, which in physics is the temperature. Given a phase variable u, we denote by $\mathbb{E}[u]$ the expected value of u with respect to the invariant measure μ,

$$\mathbb{E}[u] = \int_{\mathbf{\Gamma}} u(\hat{x})\,\mu(d\hat{x}).$$

We endow the space of phase variables with the inner product $(u, v) = \mathbb{E}[uv]$, which makes them elements of the Hilbert space $L_2[\mathbf{\Gamma}, \mu]$ (L_2 for brevity). The representations (**??**) and (**??**) are equivalent; in particular one can retrieve (**??**) from (**??**) by setting $u = \varphi_i$ (the i-th coordinate function).

We wish to follow a small number of variables $\varphi_i(x, t), i = 1, \ldots, m, m < n$ without calculating the rest. We write $\hat{\varphi} = (\phi_1, \ldots, \phi_m)$ for the vector whose entries are the variables we actually calculate. To find equations for $\hat{\varphi}$ we need projections of L_2 on subspaces of functions of $\hat{x} = (x_1, \ldots, x_m)$ or of $\hat{\varphi}(x, t) = (\varphi_1, \ldots, \varphi_m)$; these projections are not unique and we shall consider two different ones:

(i) Consider the conditional expectation $\mathbb{E}[v|u]$, where both u and v are phase variables; it satisfies:

1. $\mathbb{E}[v|u]$ is a function of u;

2. $\mathbb{E}[v|u]$ is linear in v:
$$\mathbb{E}[\alpha v_1 + \beta v_2 | u] = \alpha\, \mathbb{E}[v_1|u] + \beta\, \mathbb{E}[v_2|u].$$

3. $\mathbb{E}[v|u]$ is the best approximation of v by a function of u:
$$\mathbb{E}[|v - \mathbb{E}[v|u]|^2] \leq \mathbb{E}[|v - f(u)|^2] \tag{2.4}$$
for all functions f.

See Chung [**?**].

$\mathbb{E}[v|u]$ is the orthogonal projection of v on the space of functions of u, and we can write use $Pv = \mathbb{E}[v|u]$. This is the "nonlinear projection" used in ([**?**]) with a different interpretation, as well as in ([**?**, **?**, **?**, **?**]).

ii)For a function $w = w(x)$ in L_2, define
$$P'w = \Sigma_1^m \frac{(w, x_i)}{(x_i, x_i)} x_i / ||x_i||^2, \tag{2.5}$$

where $||x_i||^2 = \mathbb{E}[x_i^2]$; this is the "linear projection" used in most of irreversible statistical mechanics (see e.g. ([?])) . Note that in (2.5) the inner product is the expected value product defined above, and the final factor x_i is not a value of x_i at a point of $\mathbf{\Gamma}$ but a function in L_2, $q(x) = x_i$. A similar operation allows us to project onto the span of $\hat{\varphi}$.

Clearly the projection based on conditional expectations puts more information into the subspace of functions of the smaller set of variables and is thus preferable for our purposes, but we shall need both projections for technical reasons. There is a large set of projections intermediate between these two, obtained by spanning the space of functions of $\hat{x}$ by additional elements; we shall not discuss these intermediate projections here. One can also see that projecting the operator L in L_2 onto a space of functions of fewer variables $\hat{x}$ is equivalent to projecting x in $\mathbf{\Gamma}$ onto a subspace spanned by the components of $\hat{x}$.

We now follow the Mori-Zwanzig procedure ([?, ?, ?, ?]) and split the time derivative of $e^{tL}u$ into its projection on the functions of $\hat{\varphi}(t)$ plus a complement:

$$\frac{\partial}{\partial t}e^{tL}u = e^{tL}Lu = e^{tL}LP + e^{tL}QLu, \tag{2.6}$$

where P is either of the two projection just described and $Q = I - P$. If u is a scalar function of φ, it is projected in L_2 on a function of $\hat{\varphi}$; if u is a vector function of φ, each of its components is projected. The first term in (**??**) can be evaluated in terms of u and we denote it by $\mathfrak{R}(u)$:

$$e^{tL}\mathfrak{R}(u) = \mathfrak{R}(e^{tL}u).$$

To understand the second term, consider an evolution operator, e^{tQL}, defined for a phase variable v by the equation:

$$\frac{\partial}{\partial t}e^{tQL}v = QLe^{tQL}v = Le^{tQL}v - PLe^{tQL}v. \tag{2.7}$$

with $Pv(t=0)=0$. This is the orthogonal dynamics equation,which should be read as follows: Let $w=v$ at $t=0$; let $w(t)$ be a time-dependent phase variable that satisfies

$$\frac{\partial}{\partial t}w(t) = Lw(t) - PLw(t); \tag{2.8}$$

and set $w(t) = e^{tQL}v$, defining the orthogonal evolution operator e^{tQL}. Note that

$$\frac{\partial}{\partial t}Pe^{tQL}w = \frac{\partial}{\partial t}e^{tQL}Pw; \tag{2.9}$$

hence if v is orthogonal to the range of P then so is $e^{tQL}v$ for all times t. The operator e^{tQL} is the solution operator of the orthogonal dynamics.

The evolution operators e^{tL} and e^{tQL} satisfy the Dyson formula [**?**]:

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

which can be checked by differentiation. With the help of the Dyson formula the second term on the right hand side of (2.6) can be written as:

$$e^{tQL}v + \int_0^t e^{(t-s)L}PLe^{sQL}vds,$$

where

$$v = Lu - PLu = QLu.$$

Putting all the terms together, we obtain the generalized Langevin equation

$$\frac{\partial}{\partial t}e^{tL}u = \mathfrak{R}(e^{tL}u) + \int_0^t e^{(t-s)L}PLe^{sQL}vds\,ds + e^{tQL}v. \tag{2.10}$$

This is an identity between phase variables, the starting point for our approximations. If one projects this equation on the space spanned by the initial data for u, the last term drops out and the second term acquires a projection operator as prefactor.

The various terms in equation (2.10) have conventional interpretations. The first term on the right-hand side is a function only of $e^{tL}u$ and represents the self-interaction of the components of $e^{tL}u$; it is the Markovian contribution to $\frac{\partial}{\partial t}e^{tL}u$. The second term depends on $\boldsymbol{u}$ through the values of $e^{tL}u$ at all times s between 0 and t, and embodies a non-Markovian memory. Finally, the third term has no component in the range of P and if we know nothing outside this range it can be viewed as random, with statistics determined by the initial data.

In the special case of a linear projection P' the second term in equation (2.10) acquires a particularly simple form, well-known in statistical physics where it is the only form in general use. By definition, see above,

$$P'Le^{sQL}QLu = P'LQe^{sQL}QLu = \sum_{1}^{m}(LQe^{sQL}QLu, h_j(\hat{x}))h_j(\hat{x}), \qquad (2.11)$$

where $h_j(\hat{x}) = \hat{x}_j/||x_j||, ||x_j||^2 = E[x_j^2]$, are normalized coordinate functions. A short manipulation converts this sum into

$$-\sum K_j(s)h_j(\hat{x}), \qquad (2.12)$$

where $K_j(s) = (e^{sQL}QLu, QLh_j)$; the substitution $u = x_j$, used to go from L_2 to $\mathbf{\Gamma}$, yields $K_j(s) = (v_j(s), v_j(0))$, the autocorrelation of the solution of the orthogonal dynamics equation that starts from $v_j = QLx_j$. Substitution into the integral produces the term

$$-\int_0^t e^{t-s}\sum K_j(s)\varphi_j(\hat{x}, s). \qquad (2.13)$$

3 Implementation of the projections in a small dimensional space and first-order optimal prediction

The generalized Langevin equation (??) expresses the rate of change of the phase variable u as a sum of terms that depend on u and on the orthogonal dynamics. These expressions cannot be used directly for approximation when the equations are nonlinear. Indeed, the evaluation of a term such as $e^{tL}E[Lu|u] = E[Le^{tL}u|e^{tL}u]$ requires the evaluation of $e^{tL}u$, i.e., requires a solution of the full set of n equations. In terms of φ, this term becomes

$$\mathbb{E}[R(\varphi)|\hat{\varphi}(\hat{x},t)] = \mathbb{E}[R(\hat{y})|\hat{y} = \hat{\varphi}(\hat{x},t)], \tag{3.1}$$

where the last expression is the expected value of $R(\hat{y})$ given the value $\hat{\varphi}(\hat{x},t)$ of $\hat{y}$; in a more trasparent notation,

$$\mathbb{E}[R(\hat{y})|\hat{y} = \hat{\varphi}(\hat{x},t)] = \mathbb{E}[R(\hat{y})|\hat{y}]_{\hat{y}=\hat{\varphi}(\hat{x},t)]}. \tag{3.2}$$

(this is the expected value of R when we know that $\hat{y}$ has the value $\hat{\varphi}(\hat{x},t)$; we first find the best approximation of R by a function of $\hat{y}$, the substitute the value $\hat{y} = \hat{\varphi}(\hat{x},t)$ into the result). The vector $\hat{x}$ has components $\tilde{\hat{x}}$ not included in $\hat{x}$ and not known; $\hat{x} = (\hat{x},\tilde{\hat{x}})$. To obtain something that can be evaluated we move the expectation into the condition, i.e., we approximate $\mathbb{E}[R|\hat{y} = \varphi(\hat{x},\tilde{\hat{x}},t)]$ by $\mathbb{E}[R|\hat{y} = \mathbb{E}[\varphi(\hat{x},t)|\hat{x}] =$. We know that $\mathbb{E}[\varphi(\hat{x},t)|\hat{x}]$ is a function $\hat{y} = \hat{y}(\hat{x})$ of $\hat{x}$, but we do not know what function; it has to be determined so as to be consistent with the equations of motion.

Assume for a moment that the second and third terms in the Langevin equation (2.10) are negligible. Equation (2.10) becomes:

$$\frac{d}{dt}\hat{\varphi}(\hat{x},t) = \mathbb{E}[\hat{R}(\varphi(\hat{x},t)|\hat{\varphi}(\hat{x},t)]\mathbb{E}[R(\hat{y})|\hat{y}]_{\hat{y}=\hat{\varphi}(\hat{x},t)]}. \tag{3.3}$$

The conditioning data $\hat{\varphi}(\hat{x},t)$ depend not only on $\hat{x}$ but also on the other components of $\hat{x}$. The inner averaging just described makes the subsitution

$$\mathbb{E}[R(\hat{y})|\hat{y}] = \mathbb{E}[R(y)|\hat{y}]_{\hat{y}=\mathbb{E}[\hat{\varphi}(\hat{x},t)|\hat{x}]]}, \tag{3.4}$$

, where $\mathbb{E}[\hat{\varphi}(\hat{x},t)|\hat{x}]$ is a function of $\hat{x}$, $E[\hat{\varphi}(\hat{x},t)|\hat{x}] = \hat{y}(\hat{x})$.

Equation ([**?**]) becomes

$$\frac{d}{dt}\hat{\varphi}(\hat{x},t) = \mathbb{E}[\hat{R}(\hat{y})|\hat{y}]_{\hat{y}==\hat{y}(\hat{x})]}; \tag{3.5}$$

conditioning both sides of the equation by $\hat{x}$ yields

$$\frac{d}{dt}(\hat{y}) = \mathbb{E}[\hat{R}(\hat{y})|\hat{y}]_{\hat{y}=\hat{y}(\hat{x})]}. \tag{3.6}$$

The initial data $\hat{y}(0) = \hat{x}$ make $\hat{y}(t)$ a function of $\hat{x}$. Thus self-consistency and the equation of motion determine the averaged dependence of $\hat{y}$ on $\hat{x}$. Equation (3.6) is the first-order optimal prediction equation []. The presentations in these earlier references may bring out better the importance of self-consistency.

Let $\mathfrak{R}$ be the right hand side of equation (3.6); clearly $\mathbb{E}\hat{R}(\hat{y})|\hat{y}]_{\hat{y}=\hat{y}(\hat{x})]}$ is the best approximation of the mean of $\mathfrak{R}$ given the mean of $\hat{\varphi}$; first-order optimal prediction neglects however the effect of the fluctuations in $\mathfrak{R}$, as we already know from ([**?**]).

As an example, consider the model equations that arise from the Hamiltonian

$$H = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1^2 x_3^2), \tag{3.7}$$

(x_1, x_3 can be viewed as posotion variables q_1, q_2 and x_2, x_4 can be viewed as momenta p_1, p_2). The resulting equations of motion are:

$$\frac{d}{dt}\phi_1 = -\phi_2, \frac{d}{dt}\phi_2 = \phi_1 + \phi_1\phi_3^2, \frac{d}{dt}\phi_3 = -\phi_4 \frac{d}{dt}\phi_4 = \phi_3 + \phi_3\phi_1^2, \tag{3.8}$$

with $\varphi(0) = \hat{x}$. The probability density of x is $Z^{-1}e^{-H(x)/T}$, and that of φ is $Z^{-1}e^{-H(\varphi)/T}$. Keeping $m = 2$ out of the $n = 4$ equations, we find that the first term in the Langevin equation (2.10) becomes:

$$\frac{d}{dt}\phi_1 = -\phi_2, \frac{d}{dt}\phi_2 = \phi_1 + \frac{\phi_1}{1+\phi_1^2}, \tag{3.9}$$

, as can be deduced from the definition of conditional expectation (with $T = 1$ for simplicity):

$$\int_{-\infty}^{+\infty} x_2^2 e^{-H} dx_3 dx_4 / \int e^{-H} dx_3 dx_4 = x_1/(1 + x_1^2), \tag{3.10}$$

, and all the terms in (3.9) are evaluated along the

trajectories $\varphi = \varphi(x,t)$ of the full $4-$component system. The conditioning of $\hat{\varphi}(\hat{x},t)$ by $\hat{x}$ yields 2 $E[R_1(y)|\hat{y}]$ is the expected value of R

equations of the same form, but now the vector $\hat{\varphi}(x,t)$ is replaced by the vector $\hat{y} = (\hat{y}_1, \hat{y}_2)$ and the right-hand-side depends on $\hat{y}$ only. These are the first-order optimal prediction equations as presented in **??**.

The example allows us to clarify some of the notations. $E[R_1(y)|\hat{y}]$ is the expected value of R given $\hat{y}$- the best approximation of R_1 given $\hat{y}$- in the present case, $y_1/(1+y_3^2)$. $E[R_1(y)|\hat{y}]_{y1=sin(x)}$ is the previous expression in which the substitution indicated in the subscript has been made, $sinx/(1+sin^2x)$. If f is a one-to-one and onto mapping from the reals to the reals, then $E[R_1|\hat{y}]_{y1=f(x)}$ is $f(x)/(1+f^{2(x)}$. On the other hand, $E[R_1|f(x)]$, the best approximation of R_1 given $f(x)$, equals $E[R_1|x]$ and requires a careful and not obvious evaluation.

4 Zero-th order solution of the Langevin equation

Now consider the full Langevin equation 2.10. The evaluation of the memory term requires a solution of the orthogonal dynamics equation, which we shall present in detail elsewhere. In the present paper we shall be content with less, and the result is instructive.

With the appropriate substitutions, the Dyson formula **??** yields the identity

$$e^{tQL} = e^{tL} - \int_0^{te^{(t-s)QL}} XXe^{sL}ds, \tag{4.1}$$

,

which can be in principle be the starting point of a perturbative evaluation of e^{tQL}. The zero-th order approximation of this equation is:

$$e^{tQL} \cong e^{tQL}, \tag{4.2}$$

i.e., one replaces the true flow in the orthogonal complement of the range of P (or P') by the real flow induced by L (which of course has a component in the orthogonal complement). Suprisingly, this very crude approximation has some value. Part of the explanation may be the fact that we need the approximation only for a short time, as shown below. The general fact that the memory kernel is robust under approximation is discussed in [?].

At $t = 0$ the orthogonal dynamics start from $V(0) = R(x) - \mathbb{E}[R(x)|\hat{x}]$. At later time, in our poor approximation, $V(t) = e^{tQL}V(0) = R(\varphi(x,t)) - \mathbb{E}[R(y)|\hat{y}]_{\hat{y}=\hat{\varphi}_1(x,t)}$. As before, $\hat{\varphi}_1(x,t)$ is determined by the solution of the full system, and we approximate the value of $\hat{y}$ by a function of $\hat{x}$, say $\hat{y} = \mathbb{E}[\hat{\varphi}(x,t)|\hat{x}]$. In the example, $\mathbb{E}[R_1(y)|y_1] = y_1/(1+y_1^2)$; let $W(t) = \mathbb{E}[R(y)|\hat{y}]_{\hat{y}=\mathbb{E}[\hat{\varphi}(x,t)}|\hat{x}]$. Assume furthermore without loss of generality that $R(\phi) = 0$ for all t. The kernels K_j in 2.13 become

$$K_j(t-s) = K_j^0(t-s) - W_j(t)W_j(s). \tag{4.3}$$

where $K_j^0 = \mathbb{E}[V_j(t)V_j(s)]$ can be evaluated by Monte-Carlo: Sample initial data from the invariant distribution over and over, solve the resulting system of equations, whose solutions are of course statistically stationary in time, and average. Note that this calculation does not depend on the initial values $\hat{x}$, and thus can be done once and for all for a given equation and a given level of truncation. The time evolution of the $V_j(t)$ need be accurate only as long as $V_j(t)$ and $V_j(0)$ are correlated. The product of the W's depends only on $\hat{y}(\hat{x}) = \mathbb{E}[\hat{\varphi}(x,t)|\hat{x}]$.

Substitute this into the Langevin equation: there is no impediment to using the true e^{tQL} in the third term. Average with respect to x, keeping $\hat{x}$ fixed. Finally, note that the K_j are functions of time only and commute with a projection on the initial data, while the average of the third term is 0. Hence, a self-consistent equation for $\hat{y}(\hat{x})$ is defined by

$$\frac{d}{dt}\hat{y} = \hat{W} - \int_0^t (\sum_1^m (K_j^0(t-s) - W_j(t)W_j(s))y_j(s)ds, labeloops \tag{4.4}$$

. with $\hat{y}(\hat{x},0)=\hat{x}$.

Equation **??** has a simple intepretaion: Disregarding the inaccurate evaluation of e^{tQL}, one can view W as the mean solution and $V(t)$ as the fluctuation; equations oops states that the mean depends on the autocorrelations of the fluctuations. The mean is thus "renormalized" (see e.g [**?**]). The sources of error in **??** are: An inconsistent use of projections, an erroneous solution of the orthogonal dynamics, and possibly an inaccurate Monte-Carlo evaluation of the kernels K^0. All can be remedied.

In the example, **??** reduces to:

$$\frac{d}{dt}y_1 = E(t) - \int_0^t (K^0(t-s) - E(t)E(s))y_1(s)ds, \frac{d}{dt}y_2 = -y_1 \qquad (4.5)$$

, where $E(t) = y_2(t) + \frac{y_2}{1+y_2^2)}$, $\hat{y}(\hat{x},0) = \hat{x} = (x_1, x_2)$, and the function $\hat{y} = \mathbb{E}[\hat{\varphi}(x,t)|\hat{x}]$ is determined by the solution of the equation. The kernel K^0 is $K^0 = \mathbb{E}[R(\varphi(x,t)\varphi(x,s)]$, $R(\varphi) = \varphi_2 + \varphi_2\varphi_4^2$, and the expected value in K^0 is evaluated over all choices of x in $n=4$ dimensions drawn from the invariant distribution with density $Z^{-1}e^{-H/T}$.

In Figure 1 wedisplay some numerical results, for $y_1(t)$, with known data $x_1 = 1, x_2 = 0$. We shiow the "truth", the mean solution of equations **??** with these data obtained by Monte-Carlo, the Galerkin approximation that sets all unknown functions to zero, the first-order optimal prediction, and the solution of equation 4.5. The solutions are shown in a rather favorable case (they look less striking if one exchanges for example the values of x_1, x_2).