SIAM J. NUMER. ANAL.
Vol. 39, No. 3, pp. 983–1000

CONVERGENCE OF OPTIMAL PREDICTION FOR NONLINEAR HAMILTONIAN SYSTEMS*

OLE H. HALD† AND RAZ KUPFERMAN‡

Abstract. Optimal prediction is a computational method for systems that cannot be properly resolved, in which the unresolved variables are viewed as random. This paper presents a first analysis of optimal prediction as a numerical method. We prove the convergence of the scheme for a class of equations of Schrödinger type and derive error bounds for the mean error between the optimal prediction solution and the set of exact solutions with random initial data. It is shown that optimal prediction is the scheme that minimizes the mean truncation error.

Key words. optimal prediction, nonlinear Schrödinger, statistical mechanics

AMS subject classifications. 65C20, 35Q55, 37K99

PII. S0036142900374482

1. Introduction.

The method of optimal prediction of Chorin, Kast, and Kupferman [6, 7, 8] is a computational approach to problems that are too complex to be properly resolved by standard numerical methods. Underresolution occurs when the number of variables used in a computation is insufficient to capture the full range of scales that occur in the solution. Such a situation is ubiquitous in complex nonlinear systems (e.g., turbulence, geophysical models). The traditional approach to such difficulty is "modeling"—formulating a modified set of equations that try to compensate for the lack of resolution. Optimal prediction is a new approach in which the unresolved variables are considered as random, and one tries to take advantage of their statistics to predict the *mean* value of the resolved variables given their value at the initial time.

The optimal prediction scheme was found to improve the accuracy of calculations for a certain class of problems (see Chorin, Kupferman, and Levy [9] and Kast [12]). In particular, equations of Schrödinger type in finite one-dimensional domains with rough initial data were studied; standard numerical schemes are known to be severely underresolved for such problems. Numerical experiments confirm that the optimal prediction scheme can capture the evolution of the mean value of a small number of Fourier modes over a short time interval better than a standard spectral scheme that uses truncation. In a subsequent publication (Chorin, Hald, and Kupferman [5]) the non-Markovian nature of underresolved dynamics was pointed out. A Langevin-type equation was derived and was found to yield reliable predictions for intermediate times.

*Received by the editors June 26, 2000; accepted for publication (in revised form) March 23, 2001; published electronically August 15, 2001. This work was supported in part by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, under contract DE-AC03-76-SF00098.

http://www.siam.org/journals/sinum/39-3/37448.html

†Department of Mathematics, University of California, Berkeley, CA, 94720 (hald@math.berkeley.edu).

‡Institute of Mathematics, The Hebrew University, Jerusalem, 91904, Israel (raz@math.huji.ac.il).
This author's work was supported by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities.

The development of optimal prediction theory has been motivated all along by the analogy between underresolved computations and statistical physics. The relation between ensemble averages and nonlinear dynamics of high-dimensional systems is one of the fundamental postulates of statistical mechanics. The numerical results reported above are encouraging and indicate that statistical physics may be a natural framework for considering underresolved computations.

From an analytical point of view, the theory of optimal prediction still has to be elaborated on. The only rigorous analysis at present is due to Hald [10] for a restricted set of linear problems. What is missing is an investigation of optimal prediction *as a numerical method*. Such an analysis, we hope, would clarify in what sense optimal prediction is really optimal and provide rigorous error bounds. It is the goal of this paper to fill this gap, at least in part.

In section 2 we present the general theory of optimal prediction. We introduce only the assumptions that are essential to the theory and deemphasize the physical context. Thus, some of the terminology and notations differ from the ones used in previous publications.

The fundamental idea in optimal prediction theory is to approximate the solution to a large system of n equations by the solution to a much smaller system of m equations. It is assumed that the two systems share at time $t = 0$ identical values for their common variables, whereas the remaining variables are random and drawn from a probability distribution that is invariant. The solution to the initial value problem in m variables is thus compared to a collection of initial value problems in n variables.

The main result in section 2 is that the mean deviation between the exact and approximate solutions is bounded by an expression that is proportional to the mean truncation error (Theorem 1). The surprising fact is that the mean truncation error does not depend on time. Optimal prediction is shown to be the scheme that yields the smallest possible truncation error. Furthermore, we obtain a lower bound on the mean deviation between the solutions to the large and small systems, which reflects the limits of predictability due to the randomness of the initial data.

In section 3 we investigate the defocusing nonlinear Schrödinger equation that was the object of numerical experiments in [6, 7, 8, 9]. To avoid the technical difficulties associated with weak solutions and measures on infinite-dimensional spaces [3], we first approximate the nonlinear Schrödinger equation by a large system of n equations that corresponds to spectral truncation. Then we try to approximate it by one that involves only a small number of Fourier modes. Our main result is that the conditions under which Theorem 1 holds are satisfied for this system (Theorem 2). Sections 4 and 5 contain the lemmas needed for the proof of Theorem 2.

Section 6 contains preliminary results concerning a novel idea, which is to use optimal prediction as a basis for a sampling procedure. Numerical experiments show that over long time intervals, optimal prediction cannot capture the decaying character of averages, which is due to the dispersion of solutions that start with different initial conditions. In [5] a non-Markovian version of optimal prediction was developed and was found to improve the intermediate-time prediction. An alternative approach is to solve the optimal prediction equations with m_2 variables many times and average over the last $m_2 - m_1$ components of the initial data. We show that such a sampling procedure may reduce the error. The establishment of stricter bounds is left for future investigation.

2. General theory. We consider a large system of ordinary differential equations,

$$\frac{d}{dt}u(t) = R(u(t)), \qquad u(0) = \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} = s, \tag{2.1}$$

where the first m coordinates, $s_1, \ldots, s_m$, are assumed to be given at time $t = 0$, whereas the remaining $n - m$ coordinates, $s_{m+1}, \ldots, s_n$, are random. Let the vector $Pu(t)$ contain the first m components of $u(t) = u(s,t)$. We seek an approximate method,

$$\frac{d}{dt}v(t) = \mathcal{R}(v(t)), \qquad v(0) = \begin{pmatrix} s_1 \\ \vdots \\ s_m \end{pmatrix} = Ps, \tag{2.2}$$

with $m < n$ and $v(t) = v(Ps,t)$ for which

$$|Pu(t) - v(t)|^2 = \sum_{j=1}^{m} |u_j(t) - v_j(t)|^2$$

is small in a sense to be specified. In the language of [6] the first m components of u represent a particular choice of collective variables; we focus on this case for simplicity, but our results can be extended to collective variables that have a general linear dependence on the components of u.

We make the following assumption regarding the dynamical systems (2.1), (2.2).

Assumption 1. (a) There exists a function H of n variables such that

$$\frac{d}{dt}H(u(t)) = 0,$$
$$Z = \int_{\mathbb{R}^n} e^{-\beta H(s)}\, ds_1 \cdots ds_n < \infty;$$

(b) the flow (2.1) preserves volume in phase-space, i.e.,

$$\sum_{j=1}^{n} \partial_{u_j} R_j(u) = 0;$$

(c) let $(v,w) = \sum_{j=1}^{m} v_j w_j$ be the discrete inner product; the function $\mathcal{R}$ satisfies the Lipschitz condition

$$(v - w, \mathcal{R}(v) - \mathcal{R}(w)) \le L\,|v - w|^2, \qquad v, w \in \mathbb{R}^m.$$

Condition (b) and the first part of (a) hold for all Hamiltonian systems; in this case the function H is the Hamiltonian and represents the total energy of the system. Condition (c) is quite restrictive by limiting the growth of the function $\mathcal{R}$.

Since the deviation $|Pu(s,t) - v(Ps,t)|$ depends on the random variables, $s_{m+1}, \ldots, s_n$, we cannot expect it to be small for all initial conditions s. Instead, our goal is to design a scheme such that the error be small *in the mean.* To define averages we introduce

$$\mathbb{E}\, f = \int_{\mathbb{R}^n} f(s) Z^{-1} e^{-\beta H(s)}\, ds_1 \cdots ds_n, \tag{2.3}$$

where $\beta > 0$. In statistical mechanics $1/\beta$ is proportional to the temperature and $Z^{-1}\exp(-\beta H(s))$ is called the canonical distribution. In probability language $Z^{-1}\exp(-\beta H(s))$ is a probability density, and the corresponding probability measure is $\mu_0(E) = \int_E Z^{-1}\exp(-\beta H(s))\,ds$. If Assumption 1 holds, then $\mu_0(\{s : u(s,t) \in E\})$ does not depend on time, and we say that the measure is *invariant.* In [6, 7, 8] this measure is called a prior measure because it embeds our belief about the distribution of the initial data, i.e., $\mathrm{Prob}(u(0) \in E) = \mu_0(E)$.

THEOREM 1. *Let Assumption* 1 *be satisfied. Then*

$$\left(\mathbb{E}\left|Pu(s,t) - v(Ps,t)\right|^2\right)^{1/2} \le \frac{e^{Lt}-1}{L}\left(\mathbb{E}\left|PR(s) - \mathcal{R}(Ps)\right|^2\right)^{1/2}.$$

Remark. This is a general result. It does not depend on $\mathcal{R}$ except via the Lipschitz constant L. It is possible that a local Lipschitz property may be sufficient in particular cases. To use the result we must select $\mathcal{R}$ appropriately and estimate the right-hand side. In this paper we consider two choices of $\mathcal{R}$. The first choice is similar to spectral truncation, and the second we call optimal prediction. In section 3 we use Theorem 1 to get error bounds for approximations of nonlinear Schrödinger equations. The detailed estimates are presented in sections 4 and 5.

Proof. The proof of Theorem 1 is similar to the convergence proof for Euler's method. The surprise is that the quantity that corresponds to the truncation error is independent of time. Let $\epsilon^2 = \mathbb{E}(Pu - v, Pu - v)$. Differentiating both sides with respect to t, writing $\dot\epsilon = \frac{d}{dt}\epsilon$, and using (2.1) and (2.2) yields

$$\begin{aligned}\epsilon\dot\epsilon &= \mathbb{E}(Pu - v, P\dot u - \dot v)\\ &= \mathbb{E}(Pu - v, PR(u) - \mathcal{R}(v))\\ &= \mathbb{E}(Pu - v, PR(u) - \mathcal{R}(Pu)) + \mathbb{E}(Pu - v, \mathcal{R}(Pu) - \mathcal{R}(v)).\end{aligned}$$

To estimate the first term we use the Cauchy–Schwarz inequality twice, and to estimate the last term we use the Lipschitz condition from Assumption 1(c):

$$\begin{aligned}\epsilon\dot\epsilon &\le \mathbb{E}\left|Pu - v\right|\left|PR(u) - \mathcal{R}(Pu)\right| + \mathbb{E}L\left|Pu - v\right|^2\\ &\le \left(\mathbb{E}\left|Pu - v\right|^2\right)^{1/2}\left(\mathbb{E}\left|PR(u) - \mathcal{R}(Pu)\right|^2\right)^{1/2} + L\,\mathbb{E}\left|Pu - v\right|^2\\ &\le \epsilon\tau + L\epsilon^2.\end{aligned} \tag{2.4}$$

This defines the truncation error, $\tau(t)$. We next show that $\tau(t)$ is independent of t. Because H is a constant of motion and the flow in $\mathbb{R}^n$ is measure preserving, it follows that

$$\begin{aligned}\tau^2(t) &= \mathbb{E}\left|PR(u(s,t)) - \mathcal{R}(Pu(s,t))\right|^2\\ &= \int_{\mathbb{R}^n}\left|PR(u(s,t)) - \mathcal{R}(Pu(s,t))\right|^2 Z^{-1}e^{-\beta H(s)}ds\\ &= \int_{\mathbb{R}^n}\left|PR(u(s,t)) - \mathcal{R}(Pu(s,t))\right|^2 Z^{-1}e^{-\beta H(u(s,t))}\left|\frac{\partial(u_1,\dots,u_n)}{\partial(s_1,\dots,s_n)}\right|ds.\end{aligned}$$

Changing variables $s' = u(s,t)$, we get

$$\tau^2(t) = \int_{\mathbb{R}^n}\left|PR(s') - \mathcal{R}(Ps')\right|^2 Z^{-1}e^{-\beta H(s')}\,ds' = \tau^2(0).$$

Finally, we solve the differential inequality (2.4) and obtain

$$\epsilon(t) \le \frac{e^{Lt}-1}{L}\tau(0)$$

as $Pu - v = 0$ at $t = 0$. This completes the proof. $\square$

To get a decent approximation to $Pu(s,t)$ we must select $\mathcal{R}$ appropriately. The simplest choice is to assume that $u_{m+1} = \cdots = u_n \equiv 0$ for all time and solve for the remaining variables (which we then denote by $v_1, \ldots, v_m$). This is a form of truncation and amounts to solving $\dot{v} = \widehat{\mathcal{R}}(v)$, where

$$\widehat{\mathcal{R}}(Ps) = PR(s), \tag{2.5}$$

with $s_{m+1} = \cdots s_n = 0$ and $s_1, \ldots, s_m$ are dummy variables.

A more sophisticated choice of $\mathcal{R}$ can be based on orthogonal projections. We start with the following question: how does one approximate a function of two variables, $f(x,y)$, by a function of one variable, $g(x)$, given a probability density $\rho(x,y)$ on $\mathbb{R}^2$? Consider the identity

$$\int_{x,y} (f-g)^2 \rho\, dx\, dy = \int_{x,y} \left(f - \frac{\int f\rho\, dy}{\int \rho\, dy} \right)^2 \rho\, dx\, dy + \int_{x,y} \left(\frac{\int f\rho\, dy}{\int \rho\, dy} - g \right)^2 \rho\, dx\, dy$$
$$+ 2 \int_x \left(\frac{\int f\rho\, dy}{\int \rho\, dy} - g \right) \int_y \left(f - \frac{\int f\rho\, dy}{\int \rho\, dy} \right) \rho\, dy\, dx.$$

The first integral does not depend on g, whereas the last integral vanishes. Hence, $g(x) = (\int \rho\, dy)^{-1} \int f\rho\, dy$ is the function that is closest to $f(x,y)$ in the mean sense. In probability theory this function is called the *conditional expectation* of f given x, and it is denoted by $\mathbb{E}[f|x]$ (see [1, section 34]). If we replace x, y, and ρ by $(s_1, \ldots, s_m)$, $(s_{m+1}, \ldots, s_n)$, and $Z^{-1}\exp(-\beta H(s))$, respectively, we obtain that

$$\mathcal{R}(Ps) = \mathbb{E}[PR|Ps] = \frac{\int PR(s) e^{-\beta H(s)}\, ds_{m+1} \cdots ds_n}{\int e^{-\beta H(s)}\, ds_{m+1} \cdots ds_n} \tag{2.6}$$

is the best approximation of $PR(s)$ among all functions of Ps. The system $\dot{v} = \mathcal{R}(v)$ is called optimal prediction and has the smallest truncation error of all approximation methods of order m. In particular,

$$\mathbb{E}\left| PR(s) - \mathcal{R}(Ps) \right|^2 \leq \mathbb{E}\left| PR(s) - \widehat{\mathcal{R}}(Ps) \right|^2. \tag{2.7}$$

Note that we use $\mathcal{R}$ in (2.6) rather than $\check{\mathcal{R}}$ or $\tilde{\mathcal{R}}$ because our focus is on optimal prediction, and the truncated scheme (2.5) is mainly used for comparison.

In Theorem 1 we compare the solution of (2.2) to solutions of (2.1) with the same values of $s_1, \ldots, s_m$ and $s_{m+1}, \ldots, s_n$ random. Instead, we would like to compare $v(Ps,t)$ to the average of all $Pu(s,t)$ with $s_1, \ldots, s_m$ given, i.e.,

$$\mathbb{E}\left[Pu(s,t) | Ps \right] = \frac{\int Pu(s,t)\, e^{-\beta H(s)} ds_{m+1} \cdots ds_n}{\int e^{-\beta H(s)} ds_{m+1} \cdots ds_n}.$$

Replacing $Pu(s,t)$ with $v(Ps,t)$ we see that $\mathbb{E}[v(Ps,t)|Ps] = v(Ps,t)$. Since $\mathbb{E}[\cdot|\cdot]$ is linear in the first variable and $\mathbb{E}[Pu - v|Ps]$ is the orthogonal projection of $Pu - v$ on the space of all functions of Ps, it follows from Pythagoras's law that

(2.8)
$$\begin{aligned}
&\mathbb{E}\left| Pu(s,t) - v(Ps,t) \right|^2 \\
&\quad = \mathbb{E}\left| Pu(s,t) - v(Ps,t) - \mathbb{E}[Pu(s,t) - v(Ps,t)|Ps] \right|^2 + \mathbb{E}\left| \mathbb{E}[Pu(s,t) - v(Ps,t)|Ps] \right|^2 \\
&\quad = \mathbb{E}\left| Pu(s,t) - \mathbb{E}[Pu(s,t)|Ps] \right|^2 + \mathbb{E}\left| \mathbb{E}[Pu(s,t)|Ps] - v(Ps,t) \right|^2.
\end{aligned}$$

Equation (2.8) bears an instructive interpretation. The mean deviation between Pu and v is the sum of two expressions. The first measures the mean deviation between Pu and its average over all solutions with the same initial Ps. This term makes no reference to the scheme $\dot{v} = \mathcal{R}(v)$. It is the intrinsic dispersion of solutions that have only partial initial data in common. The second expression measures the mean deviation between v and the average over all Pu having the same Ps. This is the term that one can hope to make small by a clever choice of the function $\mathcal{R}(v)$.

Combining Theorem 1 with (2.7) and (2.8) we obtain the following corollary.

COROLLARY 1. *If Assumption* 1 *holds and* v *satisfies* (2.2) *with a right-hand side given by* (2.5) *or* (2.6)*, then*

$$\left(\mathbb{E}\left|\mathbb{E}\left[Pu(s,t)|Ps\right] - v(Ps,t)\right|^2\right)^{1/2} \le \frac{e^{Lt}-1}{L}\left(\mathbb{E}\left|PR(s) - \widehat{\mathcal{R}}(Ps)\right|^2\right)^{1/2}.$$

In the terminology of classical numerical analysis, the left-hand side of (2.7) can be viewed as the *local truncation error*, whereas the left-hand side of (2.8) is the *global error*. The definitions of these errors differ from the standard ones due to the probabilistic interpretation of the underlying problem, which requires the consideration of mean errors. The goal of a computational scheme is to minimize the global error. In our case, this corresponds to the minimization of the second term on the right-hand side of (2.8). It is often possible to bound the global error by an expression that is proportional to the truncation error, and therefore the convergence of the former to zero guarantees the global convergence of the scheme. This is also the case for the optimal prediction scheme (Corollary 1).

There exists, however, an important distinction between classical numerical analysis and the analysis of underresolved computations. In the latter case, the mean truncation error is in general not small, and it is not clear how the global error builds up in time. Thus, a scheme that minimizes the mean truncation error is not guaranteed to minimize the global error. In particular, it is not unlikely that other approaches, for example, nonlinear Galerkin methods (see, e.g., [13]), could provide schemes of the form (2.2) that yield better approximations in the sense of (2.8). The optimal prediction scheme is optimal among all Markovian schemes in a "worst case" sense, when the only estimate available for the global error is the one based on the differential inequality (2.4).

Example of optimal prediction. To illustrate the difference between the truncated scheme and optimal prediction we consider a very simple system that consists of two harmonic oscillators coupled by a nonlinear spring. Let $H = \frac{1}{2}(p_1^2+q_1^2+p_2^2+kq_2^2+q_1^2q_2^2)$ with $k > 1$. Using Hamilton's equations, $\dot{p}_j = -\partial_{q_j}H$, $\dot{q}_j = \partial_{p_j}H$, we get

$$\begin{aligned} \dot{p}_1 &= -(1+q_2^2)q_1, & \dot{q}_1 &= p_1, \\ \dot{p}_2 &= -(k+q_1^2)q_2, & \dot{q}_2 &= p_2. \end{aligned}$$

If k is large, then the second particle will oscillate quickly. Setting $p_2 = q_2 = 0$ yields the truncated scheme

$$\dot{p}_1 = -q_1, \qquad \dot{q}_1 = p_1.$$

The equations for the optimal prediction are obtained from (2.6); evaluating the integrals we find that

$$\dot{p}_1 = -\left(1 + \frac{1}{\beta(k+q_1^2)}\right)q_1, \qquad \dot{q}_1 = p_1.$$

Thus, $q_1/[\beta(k+q_1^2)]$ is the function that best approximates $q_2^2 q_1$. If β is large or if k is large and β is of order 1, then the optimal prediction scheme is close to truncation. However, if β is small (high temperature) and k is moderate, then there is a big difference between the two methods. Because this is such a simple problem we can evaluate the truncation error for the two methods and conclude that

$$\left(\mathbb{E}\left|PR-\mathcal{R}(P)\right|^2\right)^{1/2}=\sqrt{\frac{2}{3}}\left(\mathbb{E}\left|PR-\widehat{\mathcal{R}}(P)\right|^2\right)^{1/2}.$$

From the point of view of error bounds, not much has been gained in this simple example.

3. The defocusing Schrödinger equation. In this section we consider a particular class of equations of the form (2.1), namely, finite-dimensional representations of equations of Schrödinger type. Let F be a function of two variables and consider the nonlinear Schrödinger equation

$$\begin{aligned} p_t &= +q_{xx}-(\partial_2 F)(p,q), \\ q_t &= -p_{xx}+(\partial_1 F)(p,q), \end{aligned} \tag{3.1}$$

where $p=p(x,t)$ and $q=q(x,t)$ are periodic functions with period 1; here subscripts denote differentiation and $(\partial_1 F)(x,y)=(\partial/\partial x)F(x,y)$. Associated with this system is a Hamiltonian (energy) function

$$H(p,q)=\int_0^1\left[\tfrac{1}{2}\left(p_x^2+q_x^2\right)+F(p,q)\right]\,dx. \tag{3.2}$$

Indeed, one readily checks that $\frac{d}{dt}H(p(t),q(t))=0$. In applications one frequently encounters the function

$$F(p,q)=\frac{\lambda}{4}(p^2+q^2)^2,$$

where $\lambda=\pm 1$. If $u=q+\imath p$, then (3.1) is equivalent to the cubic Schrödinger equation $\imath u_t+u_{xx}-\lambda\left|u\right|^2 u=0$. If $\lambda<0$, then random initial data lead to solutions that after a long time look like a soliton embedded in a sea of small scale fluctuation; see [4]. This will not happen if $\lambda>0$. We therefore call (3.1) focusing if $F<0$ and defocusing if $F>0$. In both cases the L^2 norm $\int_0^1(p^2+q^2)\,dx$ is an additional constant of motion and the equation is integrable; see [14, 2]. This creates a problem when one seeks equilibrium states using techniques from statistical mechanics; see Jordan, Turkington, and Zirbel [11] and Boucher, Ellis, and Turkington [2]. To circumvent this difficulty, Chorin, Kast, and Kupferman [6, 7, 8], Chorin, Kupferman, and Levy [9], and Chorin, Hald, and Kupferman [5] have used

$$F(p,q)=\frac{1}{4}\left(p^4+q^4\right).$$

To work in a finite-dimensional space, i.e., with systems of ordinary differential equations, we assume that p,q are periodic in $[0,1]$ and approximated by a finite Fourier series, i.e.,

$$p(x,t)=\sum_{j=0}^{2n'}p_j(t)\,e_j(x),\qquad q(x,t)=\sum_{j=0}^{2n'}q_j(t)\,e_j(x), \tag{3.3}$$

where $e_0(x) = 1$, $e_{2j-1}(x) = \sqrt{2}\cos(2\pi jx)$, and $e_{2j}(x) = \sqrt{2}\sin(2\pi jx)$. Note that the e_j are the normalized eigenfunctions of $y'' + \lambda y = 0$ with $\lambda_0 = 0$ and $\lambda_{2j-1} = \lambda_{2j} = (2\pi j)^2$. If we insert p, q from (3.3) into (3.2) we get

$$H = \sum_{j=0}^{2n'} \tfrac{1}{2}\lambda_j (p_j^2 + q_j^2) + \int_0^1 F(p,q)\,dx. \tag{3.4}$$

Hamilton's equations $\dot{p}_k = -\partial_{q_k} H$, $\dot{q}_k = \partial_{p_k} H$ then yield

$$\begin{aligned} \dot{p}_k &= -\lambda_k q_k - \int_0^1 (\partial_2 F)(p,q)\, e_k\,dx, \\ \dot{q}_k &= +\lambda_k p_k + \int_0^1 (\partial_1 F)(p,q)\, e_k\,dx \end{aligned} \tag{3.5}$$

for $k = 0, 1, \dots, 2n'$, that is, a system of $n = 2 + 4n'$ ordinary differential equations. To understand the connection to the original system (3.1) we multiply each equation by e_k and sum over k; this leads to

$$\begin{aligned} p_t &= +q_{xx} - \sum_{k=0}^{2n'} (\partial_2 F(p,q), e_k)\, e_k, \\ q_t &= -p_{xx} + \sum_{k=0}^{2n'} (\partial_1 F(p,q), e_k)\, e_k, \end{aligned}$$

where $(f,g) = \int_0^1 f(x)g(x)\,dx$. Therefore, we first expand p, q in a finite Fourier series, and then we truncate the Fourier series for $\partial_2 F(p,q)$ and $\partial_1 F(p,q)$. We can also write (3.5) as $\dot{u} = R(u)$, where $u = (p_0, q_0, \dots, p_{2n'}, q_{2n'})^T$.

The simplest approximation method, $\dot{v} = \widehat{\mathcal{R}}(v)$, is obtained via truncation; see (2.5). Let $m = 2 + 4m'$ with $m' < n'$; set $v = (p_0, q_0, \dots, p_{2m'}, q_{2m'})^T$ and solve

$$\begin{aligned} \dot{p}_k &= -\lambda_k q_k - (\partial_2 F(\hat{p}, \hat{q}), e_k)\,, \\ \dot{q}_k &= +\lambda_k p_k + (\partial_1 F(\hat{p}, \hat{q}), e_k) \end{aligned} \tag{3.6}$$

for $k = 0, \dots, 2m'$. Here, $\hat{p} = \sum_{j=0}^{2m'} p_j e_j$, and $\hat{q} = \sum_{j=0}^{2m'} q_j e_j$. Note that we use the same notation, p_j, q_j, for the coordinates in u and v.

To derive the method of optimal prediction, $\dot{v} = \mathcal{R}(v)$, we follow (2.6) and solve

$$\begin{aligned} \dot{p}_k &= -\lambda_k q_k - \frac{\int (\partial_2 F(p,q), e_k)\, e^{-\beta H(p,q)}\, dp_{2m'+1} \cdots dq_{2n'}}{\int e^{-\beta H(p,q)}\, dp_{2m'+1} \cdots dq_{2n'}}, \\ \dot{q}_k &= +\lambda_k p_k + \frac{\int (\partial_1 F(p,q), e_k)\, e^{-\beta H(p,q)}\, dp_{2m'+1} \cdots dq_{2n'}}{\int e^{-\beta H(p,q)}\, dp_{2m'+1} \cdots dq_{2n'}} \end{aligned} \tag{3.7}$$

for $k = 0, 1, \dots, 2m'$. Thus, we perform a partial averaging by integrating over the random coefficients for the high modes. To prove the convergence of optimal prediction we need restrictive assumptions about the function F.

Assumption 2. The function $F(x,y) = \frac{1}{2}c_0(x^2 + y^2) + G(x,y)$, with $c_0 > 0$, satisfies

(a) $F(0,0) = (\partial_1 F)(0,0) = (\partial_2 F)(0,0) = 0$;

(b) the matrix $A = \left(\begin{smallmatrix} \partial_{1,1}F & \partial_{1,2}F \\ \partial_{2,1}F & \partial_{2,2}F \end{smallmatrix}\right)$ is positive definite, and $|A| \le c_1$ for all (x,y);

(c) $|\nabla G(x,y)| \leq c_2$ for all x,y.

Here $\partial_{i,j} F = \partial_i \partial_j F$ and $|A|$ denotes the 2 norm. Condition (a) is satisfied for $F = \frac{1}{4}(p^2+q^2)^2$ and for $F = \frac{1}{4}(p^4+q^4)$, whereas (b) ensures that F is always positive (see Lemma 1 below). Equation (3.1) is therefore defocusing. Condition (c) implies that the normalization constant Z is finite (see below); it is needed because $\lambda_0 = 0$. The difficulty with Assumption 2 is that it restricts the growth of F for large (p,q) and excludes the standard choices of F. Instead, we may use the following functions:

$$
\begin{aligned}
F(p,q) &= \frac{1}{2\epsilon}\left[p^2+q^2-\frac{1}{\epsilon}\log\left(1+\epsilon(p^2+q^2)\right)\right],\\
F(p,q) &= \frac{p^4}{4+\epsilon p^2}+\frac{q^4}{4+\epsilon q^2}.
\end{aligned}
$$

These functions are quartic near the origin, grow quadratically at infinity, and satisfy Assumption 2.

THEOREM 2. *Let $\dot{u} = R(u)$ be the spectral method* (3.5) *for the Schrödinger equation* (3.1)*, and assume that $u \in \mathbb{R}^n$ for large $n = 2+4n'$. Let $\dot{v} = \mathcal{R}(v)$ be the optimal prediction scheme* (3.7) *with $m = 2+4m'$ and $m' < n'$. If Assumption* 2 *holds, then*

$$
\left(\mathbb{E}\,|\mathbb{E}[Pu(s,t)|Ps]-v(Ps,t)|^2\right)^{1/2} \leq \frac{e^{Lt}-1}{L}\,\frac{c_1\,e^{c_1/24}}{\pi\sqrt{\beta m'}},
$$

where the Lipschitz constant L is given by

$$
L = c_1\left[1+\frac{c_1}{\pi^2 m'}e^{2\beta c_2^2/c_0+2c_0/(\pi^2 m')}\right].
$$

Remark. Theorem 2 is also valid if we replace $\dot{v} = \mathcal{R}(v)$ by the truncated scheme $\dot{v} = \widehat{\mathcal{R}}(v)$, in which case we may use $L = c_1$.

Proof. Theorem 2 has the same structure as Corollary 1. It is therefore sufficient to check that Assumption 1 is satisfied. To do that we use Lemmas 1–4 below. Since (3.7) is a Hamiltonian system it follows that conditions (b) and the first part of (a) in Assumption 1 are automatically satisfied. For the second part of Assumption 1(a) we combine (3.4) with the first inequality of Lemma 1 and get

$$
\begin{aligned}
Z &= \int e^{-\beta H}\,dp_0\cdots dq_{2n'}\\
&\leq \int e^{\beta c_2^2/c_0}e^{-\beta\sum_j \frac{1}{2}(\lambda_j+c_0/2)(p_j^2+q_j^2)}\,dp_0\cdots dq_{2n'}\\
&= e^{\beta c_2^2/c_0}\prod_{j=0}^{2n'}\frac{1}{\beta(\lambda_j+c_0/2)};
\end{aligned}
$$

thus $Z < \infty$. Here it is important that $c_0 > 0$ as $\lambda_0 = 0$. The Lipschitz property in Assumption 1(c) is a consequence of Lemma 4. Finally, the estimate of the truncation error is given in Lemma 2. This completes the proof. ∎

4. Truncation errors. To prove Theorem 2 we must estimate the truncation error for $\dot{v} = \mathcal{R}(v)$ at $t = 0$. We begin with upper and lower bounds for the nonlinear terms $\int_0^1 F(p,q)\,dx$ in the Hamiltonian (3.4).

Lemma 1. *Let $p = \sum_j p_j e_j$, $q = \sum_j q_j e_j$, where $j = 0, 1, \ldots, 2n'$. If Assumption 2 holds, then $F(p,q) \geq 0$ and*

$$\int_0^1 F(p,q)\,dx \geq \frac{c_0}{4}\sum_{j=0}^{2n'}(p_j^2+q_j^2) - \frac{c_2^2}{c_0},$$
$$\int_0^1 F(p,q)\,dx \geq F(p_0,q_0),$$
$$\int_0^1 F(p,q)\,dx \leq F(p_0,q_0) + \frac{c_1}{2}\sum_{j=1}^{2n'}(p_j^2+q_j^2).$$

Proof. Since $F(0,0) = 0$ it follows from Assumption 2(c) and the mean value theorem that

$$\begin{aligned} F(p,q) &= \frac{c_0}{2}(p^2+q^2) + \int_0^1 \nabla G(\theta p, \theta q)\begin{pmatrix} p \\ q \end{pmatrix} d\theta \\ &\geq \frac{c_0}{2}(p^2+q^2) - \frac{1}{2}\cdot 2 \cdot \frac{c_2}{\sqrt{\frac{c_0}{2}}}\sqrt{\frac{c_0}{2}}\left|\begin{pmatrix} p \\ q \end{pmatrix}\right| \\ &\geq \frac{c_0}{2}(p^2+q^2) - \frac{c_2^2}{c_0} - \frac{c_0}{4}(p^2+q^2). \end{aligned}$$

To establish the first claim we integrate with respect to x and use the orthogonality of the $e_j(x)$. The proof of the remaining statements is based on the convexity of F. Let $g(x)$ be twice differentiable. Then, Taylor's theorem with an integral remainder yields

$$g(1) = g(0) + g'(0) + \int_0^1 (1-\theta) g''(\theta)\,d\theta. \tag{4.1}$$

If $g(\theta) = F(\theta p, \theta q)$ we conclude from Assumption 2(a), (b) that

$$F(p,q) = F(0,0) + \nabla F(0,0)\begin{pmatrix} p \\ q \end{pmatrix} + \int_0^1 (1-\theta)(p,q)\begin{pmatrix} \partial_{1,1}F & \partial_{1,2}F \\ \partial_{2,1}F & \partial_{2,2}F \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix} d\theta \geq 0.$$

Next let $g(\theta) = F(p_0 + \theta(p-p_0), q_0 + \theta(q-q_0))$. Since $\int_0^1 e_j\,dx = \delta_{0,j}$ it follows from (4.1) and Assumption 2(b) that

$$\begin{aligned} \int_{x=0}^1 F(p,q)\,dx &= F(p_0,q_0) + \nabla F(p_0,q_0)\int_{x=0}^1 \begin{pmatrix} p-p_0 \\ q-q_0 \end{pmatrix} dx \\ &\quad + \int_{x=0}^1\int_{\theta=0}^1 (1-\theta)(p-p_0, q-q_0)\begin{pmatrix} \partial_{1,1}F & \partial_{1,2}F \\ \partial_{2,1}F & \partial_{2,2}F \end{pmatrix}\begin{pmatrix} p-p_0 \\ q-q_0 \end{pmatrix} d\theta\,dx \\ &\geq F(p_0,q_0) + 0 + 0. \end{aligned}$$

The last inequality is also a consequence of Assumption 2(b). In other words,

$$\begin{aligned} \int_{x=0}^1 F(p,q)\,dx &\leq F(p_0,q_0) + 0 + \frac{c_1}{2}\int_{x=0}^1 \left|\begin{pmatrix} p-p_0 \\ q-q_0 \end{pmatrix}\right|^2 dx \\ &= F(p_0,q_0) + \frac{c_1}{2}\sum_{j=1}^{2n'}(p_j^2+q_j^2). \end{aligned}$$

This completes the proof. ∎

LEMMA 2. *If Assumption 2 holds, then*

$$\left(\mathbb{E}\left|PR(s) - \widehat{\mathcal{R}}(Ps)\right|^2\right)^{1/2} \le \frac{c_1\, e^{c_1/24}}{\pi\sqrt{\beta m'}}.$$

Proof. Let $\tau^2 = \mathbb{E}|PR - \widehat{\mathcal{R}}(P)|^2$. Remember that $p = \sum p_j e_j$, $q = \sum q_j e_j$, with $j = 0, 1, \ldots, 2n'$, and $\hat{p} = \sum p_j e_j$, $\hat{q} = \sum q_j e_j$, where $j = 0, 1, \ldots, 2m'$. Combining (3.5) and (3.6), canceling the linear terms, and using Bessel's inequality results in

$$\begin{aligned}\tau^2 &= \mathbb{E}\sum_{k=0}^{2m'} |(\partial_2 F(p,q) - \partial_2 F(\hat{p},\hat{q}), e_k)|^2 + \mathbb{E}\sum_{k=0}^{2m'} |(\partial_1 F(p,q) - \partial_1 F(\hat{p},\hat{q}), e_k)|^2 \\ &\le \mathbb{E}\,\|\partial_2 F(p,q) - \partial_2 F(\hat{p},\hat{q})\|_{L^2}^2 + \mathbb{E}\,\|\partial_1 F(p,q) - \partial_1 F(\hat{p},\hat{q})\|_{L^2}^2 .\end{aligned}$$

Using the mean value theorem, Assumption 2(b), and the orthogonality of the eigenfunctions $e_j(x)$ we obtain

$$\begin{aligned}\tau^2 &\le \mathbb{E}\left\|\int_{\theta=0}^{1} \begin{pmatrix} \partial_{1,1}F & \partial_{1,2}F \\ \partial_{2,1}F & \partial_{2,2}F \end{pmatrix} \begin{pmatrix} p-\hat{p} \\ q-\hat{q} \end{pmatrix} d\theta\right\|_{L^2}^2 \\ &\le c_1^2 \mathbb{E}\left\|\begin{pmatrix} p-\hat{p} \\ q-\hat{q} \end{pmatrix}\right\|_{L^2}^2 \\ &\le c_1^2 \mathbb{E}\sum_{k=2m'+1}^{2n'} (p_k^2 + q_k^2).\end{aligned}$$

The expected value (2.3) involves the Hamiltonian (3.4). Using Lemma 1 we get upper and lower bounds for the Hamiltonian. Thus,

$$\begin{aligned}\tau^2 &\le c_1^2 \frac{\int \sum_{k=2m'+1}^{2n'} (p_k^2 + q_k^2) e^{-\beta H}\, dp_0 \cdots dq_{2n'}}{\int e^{-\beta H}\, dp_0 \cdots dq_{2n'}} \\ &\le c_1^2 \frac{\int \sum_{k=2m'+1}^{2n'} (p_k^2 + q_k^2) e^{-\beta \sum_{j=1}^{2n'} \frac{1}{2}\lambda_j (p_j^2 + q_j^2)}\, dp_1 \cdots dq_{2n'}}{\int e^{-\beta\sum_{j=1}^{2n'} \frac{1}{2}(\lambda_j + c_1)(p_j^2+q_j^2)}\, dp_1 \cdots dq_{2n'}}.\end{aligned}$$

Here the cancellation of $\int \exp(-\beta F(p_0, q_0))\, dp_0 dq_0$ is justified by the first inequality in Lemma 1 with $(p, q) = (p_0, q_0)$. To evaluate the integrals we introduce

$$a_k = \frac{\int p_k^2 e^{-\frac{1}{2}\beta\lambda_k p_k^2}\, dp_k}{\int e^{-\frac{1}{2}\beta(\lambda_k + c_1)p_k^2}\, dp_k} = \frac{1}{\beta\lambda_k}\left(1 + \frac{c_1}{\lambda_k}\right)^{1/2}, \tag{4.2}$$

$$b_j = \frac{\int e^{-\frac{1}{2}\beta\lambda_j p_j^2}\, dp_j}{\int e^{-\frac{1}{2}\beta(\lambda_j + c_1)p_j^2}\, dp_j} = \left(1 + \frac{c_1}{\lambda_j}\right)^{1/2}. \tag{4.3}$$

The symmetry between p and q then implies that

$$\tau^2 \le 2c_1^2 \sum_{k=2m'+1}^{2n'} a_k \prod_{j=1, j\neq k}^{2n'} b_j \prod_{j=1}^{2n'} b_j = 2c_1^2 \sum_{k=2m'+1}^{2n'} \frac{1}{\beta\lambda_k} \prod_{j=1}^{2n'} \left(1 + \frac{c_1}{\lambda_j}\right).$$

Since $1 + c_1/\lambda \le e^{c_1/\lambda}$, $\sum_{k\ge 1} \lambda_k^{-1} \le (2\pi)^{-2} \cdot 2 \cdot (\pi^2/6)$, and $\sum_{k \ge m'+1} k^{-2} < 1/m'$ we conclude that

$$\tau^2 \le 2c_1^2 \frac{1}{\beta(2\pi)^2} \frac{2}{m'} e^{c_1/12}.$$

This completes the proof. ∎

5. The Lipschitz constant. In this section we show that the method of optimal prediction applied to the defocusing Schrödinger equation satisfies a Lipschitz condition, i.e., Assumption 1(c). We begin with the derivation of upper and lower bounds for the nonlinear term in the Hamiltonian (3.4).

LEMMA 3. *Let* $p = \sum_{j=0}^{2m'} p_j e_j$, $q = \sum_{j=0}^{2m'} q_j e_j$, $r = \sum_{j=2m'+1}^{2n'} r_j e_j$, *and* $s = \sum_{j=2m'+1}^{2n'} s_j e_j$. *If Assumption 2 holds, then*

$$\int_0^1 F(p+r,q+s)\,dx \geq \frac{c_0}{2}\sum_{j=0}^{2m'}(p_j^2+q_j^2) + \int_0^1 G(p,q)\,dx - \frac{c_2^2}{2c_0},$$
$$\int_0^1 F(p+r,q+s)\,dx \leq \frac{c_0}{2}\sum_{j=0}^{2m'}(p_j^2+q_j^2) + \int_0^1 G(p,q)\,dx + \frac{c_2^2}{2c_0} + c_0\sum_{j=2m'+1}^{2n'}(r_j^2+s_j^2).$$

Proof. Since $F(x,y) = \frac{1}{2}c_0(x^2+y^2) + G(x,y)$ and p,q are orthogonal to r,s, it follows that

$$\begin{aligned}
&\int_0^1 F(p+r,q+s)\,dx \\
&\quad = \int_0^1 \left\{ \tfrac{1}{2}c_0\left[(p+r)^2+(q+s)^2\right] + G(p,q) + G(p+r,q+s) - G(p,q)\right\}\,dx \\
&\quad = \tfrac{1}{2}c_0\sum_{j=0}^{2m'}(p_j^2+q_j^2) + \tfrac{1}{2}c_0\sum_{j=2m'+1}^{2n'}(r_j^2+s_j^2) + \int G(p,q)\,dx \\
&\qquad + \int_{x=0}^1\int_{\theta=0}^1 \nabla G(p+\theta r,q+\theta s)\begin{pmatrix} r\\ s\end{pmatrix}\,d\theta\,dx.
\end{aligned}\tag{5.1}$$

Denote the last term by I_1. Since $|\nabla G| \leq c_2$ and $2ab \leq a^2+b^2$ we get

$$\begin{aligned}
|I_1| &\leq \frac{1}{2}\int_{x=0}^1 2\frac{c_2}{\sqrt{c_0}}\sqrt{c_0}\left|\begin{pmatrix} r\\ s\end{pmatrix}\right|\,dx \\
&\leq \frac{1}{2}\left(\frac{c_2^2}{c_0} + c_0\int_0^1 (r^2+s^2)\,dx\right) \\
&= \frac{c_2^2}{2c_0} + \frac{c_0}{2}\sum_{j=2m'+1}^{2n'}(r_j^2+s_j^2).
\end{aligned}$$

Inserting in (5.1) completes the proof. ∎

LEMMA 4. *Let* $\dot v = \mathcal{R}(v)$ *be the optimal prediction scheme* (3.7) *for the Schrödinger equation* (3.1). *If Assumption 2 holds, then*

$$(v-w,\mathcal{R}(v)-\mathcal{R}(w)) \leq L\,|v-w|^2,$$

where

$$L = c_1\left[1 + \frac{c_1}{\pi^2 m'}e^{2\beta c_2^2/c_0 + 2c_0/(\pi^2 m')}\right].$$

Remark. Lemma 4 has been formulated for optimal prediction because that is what we are most interested in, but it is equally valid for the truncated scheme (3.6), in which case we may use $L = c_1$. The corresponding proof is straightforward and will not be presented.

Proof. The main difficulty in this proof is notational. Let $p = \sum p_j e_j$, $q = \sum q_j e_j$, $a = \sum a_j e_j$, and $b = \sum b_j e_j$, where $j = 0, 1, \dots, 2m'$. Set $v = \binom{p}{q}$, $\binom{p+a}{q+b}$, and rewrite (3.7) as

$$\begin{aligned} \dot{p}_k &= -\lambda_k q_k - g_k(v), \\ \dot{q}_k &= +\lambda_k p_k + f_k(v), \end{aligned}$$

where $f_k = (f, e_k)$ and $g_k = (g, e_k)$. We shall use v, w with two interpretations. In the first one, v is a vector with $2 + 4m'$ components p_j, q_j. In the second one, v is a function with two components $p(x), q(x)$. We can therefore write $|v| = \|v\|_{L^2}$. In the calculation of the inner product the linear terms cancel and we have

$$\begin{aligned} E &= (v - w, \mathcal{R}(v) - \mathcal{R}(w)) \\ &= \sum_{k=0}^{2m'} \left[a_k \left(g_k(v) - g_k(w)\right) - b_k \left(f_k(v) - f_k(w)\right)\right]. \end{aligned}$$

Using the Cauchy–Schwarz inequality, Bessel's inequality, and the mean value theorem gives

$$\begin{aligned} E &\le \left(\sum_{k=0}^{2m'} (a_k^2 + b_k^2)\right)^{1/2} \left(\sum_{k=0}^{2m'} \left[\left(g_k(w) - g_k(v)\right)^2 + \left(f_k(w) - f_k(v)\right)^2\right]\right)^{1/2} \\ &\le \left|\binom{a}{b}\right| \left(\int_{x=0}^{1} \left[|f(w) - f(v)|^2 + |g(w) - g(v)|^2\right] dx\right)^{1/2} \qquad (5.2) \\ &= |v - w| \left\| \int_{\theta=0}^{1} \frac{d}{d\theta} \binom{f}{g} (v + \theta(w - v))\, d\theta \right\|_{L^2}. \end{aligned}$$

To continue we need f, g explicitly. Here it is convenient to use a new notation for the high frequencies that are averaged out. Let

$$\begin{aligned} r &= \sum_{j=2m'+1}^{2n'} r_j e_j, & s &= \sum_{j=2m'+1}^{2n'} s_j e_j, \\ dr &= dr_{2m'+1} \cdots dr_{2n'}, & ds &= ds_{2m'+1} \cdots ds_{2n'}. \end{aligned}$$

Since $v = \binom{p}{q}$, $w - v = \binom{a}{b}$ it follows from (3.7) that

$$f\left(v + \theta(w - v)\right) = \frac{\int \partial_1 F e^{-\beta H}\, dr\, ds}{\int e^{-\beta H}\, dr\, ds},$$

where $\partial_1 F$ and H are evaluated at $(p + \theta a + r, q + \theta b + s)$. Canceling the common factor $\exp(-\beta \sum_j \frac{1}{2}\lambda_j[(p_j + \theta a_j)^2 + (q_j + \theta b_j)^2])$ in the numerator and denominator, differentiating with respect to θ, and finally restoring the common factor yield

$$\frac{df}{d\theta} = -\left[\int e^{-\beta H}\,dr\,ds\right]^{-2}\left[\int e^{-\beta H}(-\beta)\int_{x=0}^{1}\nabla F\cdot\begin{pmatrix}a\\b\end{pmatrix}dx\,dr\,ds\right]\left[\int \partial_1 F e^{-\beta H} dr\,ds\right]$$
$$+\left[\int e^{-\beta H} dr\,ds\right]^{-1}\left[\int \nabla\partial_1 F\cdot\begin{pmatrix}a\\b\end{pmatrix}e^{-\beta H} + \partial_1 F e^{-\beta H}(-\beta)\int_{x=0}^{1}\nabla F\cdot\begin{pmatrix}a\\b\end{pmatrix}dx\,dr\,ds\right].$$

To analyze this expression we need a more compact notation. Let

$$A = \partial_1 F, \qquad B = \nabla\partial_1 F\cdot\begin{pmatrix}a\\b\end{pmatrix},$$
$$\alpha = \int_{x=0}^{1}\nabla F\cdot\begin{pmatrix}a\\b\end{pmatrix}dx, \qquad d\mu = \frac{e^{-\beta H}\,dr\,ds}{\int e^{-\beta H}\,dr\,ds}.$$

Then

$$\frac{df}{d\theta} = \beta\int\alpha\,d\mu\int A\,d\mu + \int B\,d\mu - \beta\int A\alpha\,d\mu.$$

If we replace some of the r_j, s_j by r'_j, s'_j we can write the b terms as a double integral; indeed,

$$\frac{df}{d\theta} = \int B\,d\mu - \frac{\beta}{2}\int(A - A')(\alpha-\alpha')\,d\mu\,d\mu',$$

where A' and α' are evaluated at $(p+\theta a + r', q + \theta b + s')$. We use the mean value theorem to evaluate $\alpha - \alpha'$ and get

$$\alpha - \alpha' = \int_{x=0}^{1}\int_{\eta=0}^{1}(a,b)\begin{pmatrix}\partial_{1,1}F & \partial_{2,1}F\\ \partial_{1,2}F & \partial_{2,2}F\end{pmatrix}\begin{pmatrix}r-r'\\ s-s'\end{pmatrix}d\eta\,dx. \tag{5.3}$$

Here $\partial_i\partial_j F$ are evaluated at $(p+\theta a + r' + \eta(r-r'), q+\theta b + s' + \eta(s-s'))$. Similar arguments show that

$$A - A' = \int_{\eta=0}^{1}(\partial_{1,1}F, \partial_{1,2}F)\begin{pmatrix}r-r'\\ s-s'\end{pmatrix}d\eta.$$

If we replace $\partial_1 F$ by $\partial_2 F$ we obtain a similar expression for $dg/d\theta$ and conclude that

$$\begin{aligned}\frac{d}{d\theta}\begin{pmatrix}f\\g\end{pmatrix} &= \int\begin{pmatrix}\partial_{1,1}F & \partial_{2,1}F\\ \partial_{1,2}F & \partial_{2,2}F\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}d\mu\\ &\quad -\frac{\beta}{2}\int\int_{\eta=0}^{1}\begin{pmatrix}\partial_{1,1}F & \partial_{2,1}F\\ \partial_{1,2}F & \partial_{2,2}F\end{pmatrix}\begin{pmatrix}r-r'\\ s-s'\end{pmatrix}d\eta\,(\alpha-\alpha')\,d\mu\,d\mu'.\end{aligned}$$

The rest of the proof consists of a series of inequalities. We begin with $\alpha-\alpha'$ in (5.3). It follows from Assumption 2(b) and the Cauchy–Schwarz inequality that

$$\begin{aligned}|\alpha-\alpha'| &\le \int_{x=0}^{1}\left|\begin{pmatrix}a\\b\end{pmatrix}\right| c_1\left|\begin{pmatrix}r-r'\\ s-s'\end{pmatrix}\right|dx\\ &\le c_1\left(\int_{x=0}^{1}(a^2+b^2)\,dx\right)^{1/2}\left(\int_{x=0}^{1}\left[(r-r')^2+(s-s')^2\right]dx\right)^{1/2}\\ &= c_1\left(\sum_{j=0}^{2m'}(a_j^2+b_j^2)\right)^{1/2}\left(\sum_{j=2m'+1}^{2n'}\left[(r_j-r'_j)^2+(s_j-s'_j)^2\right]\right)^{1/2}.\end{aligned}$$

Since $a(x)$, $b(x)$ do not depend on r_j, s_j, and $\|\binom{a}{b}\|_{L^2} = |v-w|$ we find that

$$\begin{aligned}
&\left\|\frac{d}{d\theta}\begin{pmatrix} f \\ g\end{pmatrix}\right\|_{L^2} \le c_1 \left\|\begin{pmatrix} a \\ b\end{pmatrix}\right\|_{L^2} + \frac{\beta}{2}\int c_1 \left\|\begin{pmatrix} r-r' \\ s-s'\end{pmatrix}\right\|_{L^2} |\alpha-\alpha'|\; d\mu\, d\mu' \\
&\qquad = c_1\,|v-w| + \frac{\beta}{2}c_1^2\,|v-w| \int \sum_{k=2m'+1}^{2n'} \left[(r_k-r_k')^2 + (s_k-s_k')^2\right]\; d\mu\, d\mu'.
\end{aligned} \tag{5.4}$$

The treatment of the last integral, which we denote by I_2, follows the arguments in Lemma 2.

The measure $d\mu$ involves the Hamiltonian $H(p+\theta a+r, q+\theta b+s)$. From Lemma 3 we get upper and lower bounds for H, and canceling all terms involving $p_j+\theta a_j$,$q_j+\theta b_j$ we see that $\int f\,d\mu \le \int f\,d\nu$ if $f \ge 0$ and

$$d\nu = \frac{e^{\beta c_2^2/c_0} e^{-\beta\sum_{j=2m'+1}^{2n'} \frac{1}{2}\lambda_j(r_j^2+s_j^2)} dr\,ds}{\int e^{-\beta\sum_{j=2m'+1}^{2n'} \frac{1}{2}(\lambda_j+2c_0)(r_j^2+s_j^2)} dr\,ds}.$$

Since $\int r_k\,d\nu = \int s_k\,d\nu = 0$ and r, s occur symmetrically, we see that

$$\begin{aligned}
I_2 &\le 2\int \sum_{k=2m'+1}^{2n'} (r_k^2+s_k^2)\,d\nu \int d\nu' \\
&= 4\int \sum_{k=2m'+1}^{2n'} r_k^2\,d\nu \int d\nu'.
\end{aligned} \tag{5.5}$$

Since $1 + c_1/\lambda < e^{c_1/\lambda}$ and $\sum_{k=2m'+1}^{2n'} \lambda_k^{-1} < (2\pi)^{-2}\cdot 2/m'$ it follows from (4.3) with c_1 replaced by $2c_0$ that

$$\begin{aligned}
\int d\nu' &= e^{\beta c_2^2/c_0}\left(\prod_{j=2m'+1}^{2n'} b_j\right)^2 \\
&= e^{\beta c_2^2/c_0}\prod_{j=2m'+1}^{2n'}\left(1+\frac{2c_0}{\lambda_j}\right) \\
&\le e^{\beta c_2^2/c_0 + c_0/(\pi^2 m')}.
\end{aligned} \tag{5.6}$$

Using (4.2) together with (4.3) we obtain

$$\begin{aligned}
\sum_{k=2m'+1}^{2n'} \int r_k^2\,d\nu &= e^{\beta c_2^2/c_0} \sum_{k=2m'+1}^{2n'} a_k \prod_{j=2m'+1, j\neq k}^{2n'} b_j \prod_{j=2m'+1}^{2n'} b_j \\
&= e^{\beta c_2^2/c_0} \sum_{k=2m'+1}^{2n'} \frac{1}{\beta\lambda_k} \prod_{j=2m'+1}^{2n'}\left(1+\frac{2c_0}{\lambda_j}\right) \\
&\le \frac{2}{\beta(2\pi)^2 m'} e^{\beta c_2^2/c_0 + c_0/(\pi^2 m')}.
\end{aligned} \tag{5.7}$$

Combining (5.5), (5.6), and (5.7) gives

$$I_2 \leq \frac{2}{\beta\pi^2 m'} e^{2\beta c_2^2/c_0 + 2c_0/(\pi^2 m')}.$$

Finally we insert the bound for I in (5.4) and conclude from (5.2) that

$$E \leq |v-w|^2 c_1 \left[1 + \frac{c_1}{\pi^2 m'} e^{2\beta c_2^2/c_0 + 2c_0/(\pi^2 m')}\right].$$

This completes the proof. □

6. Optimal sampling. In Theorem 2 we have compared averages of the solutions of a large system (3.5) with a single solution of the optimal prediction equation (3.7). For nonlinear Schrödinger equations the numerical experiments by Chorin, Kupferman, and Levy [9] show that the averages will decay as time grows. If β is small the decay is rapid. This decay cannot be captured by the method of optimal prediction as (3.7) is a finite-dimensional Hamiltonian system. It is therefore natural to average the solutions of the optimal prediction equations to improve the performance of the method.

To formulate such an algorithm we modify the notation from section 2 slightly. We let $m_1 < m_2 < n$ and let $P_1 s$, $P_2 s$ consist of the first m_1, m_2 components of s. Here m_2 plays the same role as m in the previous sections. The idea is to average the optimal prediction solutions $v(P_2 s, t)$ over $s_{m_1+1}, \ldots, s_{m_2}$. Thus, we view $s_1, \ldots, s_{m_1}$ as given and $s_{m_1+1}, \ldots, s_{m_2}$ as random. To reveal the appropriate measure we observe that

$$\begin{aligned}
\mathbb{E}\left[v(P_2 s, t)|P_1 s\right] &= \frac{\int v(P_2 s,t) e^{-\beta H(s)} ds_{m_1+1}\cdots ds_n}{\int e^{-\beta H(s)} ds_{m_1+1}\cdots ds_n} \\
&= \frac{\int v(P_2 s,t) \int e^{-\beta H(s)} ds_{m_2+1}\cdots ds_n\, ds_{m_1+1}\cdots ds_{m_2}}{\int\int e^{-\beta H(s)} ds_{m_2+1}\cdots ds_n\, ds_{m_1+1}\cdots ds_{m_2}} \\
&= \frac{\int v(P_2 s,t) e^{-\beta H_0(P_2 s)} ds_{m_1+1}\cdots ds_{m_2}}{\int e^{-\beta H_0(P_2 s)} ds_{m_1+1}\cdots ds_{m_2}},
\end{aligned} \tag{6.1}$$

where $H_0 = -(1/\beta)\int \exp(-\beta H) ds_{m_2+1}\cdots ds_n$. In statistical mechanics H_0 is known as free energy and $Z_0^{-1}\exp(-\beta H_0)$ becomes the canonical distribution for $\dot{v} = \mathcal{R}(v)$ on the space $(s_1, \ldots, s_{m_2})$. It has been shown by Chorin, Hald, and Kupferman [5] that if (2.1) is a Hamiltonian system corresponding to H, then (2.2) is also Hamiltonian but corresponding to H_0.

Since $\mathbb{E}[\cdot|P_i s]$ is the orthogonal projection on the space of functions of $(s_1, \ldots, s_{m_i})$, it follows that

$$\begin{aligned}
\mathbb{E}\left|\mathbb{E}[P_2 u|P_2 s] - v\right|^2 &= \mathbb{E}\left|\mathbb{E}[P_2 u - v|P_2 s] - \mathbb{E}[\mathbb{E}[P_2 u - v|P_2 s]|P_1 s]\right|^2 \\
&\quad + \mathbb{E}\left|\mathbb{E}[\mathbb{E}[P_2 u - v|P_2 s]|P_1 s]\right|^2 \\
&= \mathbb{E}\left|\mathbb{E}[P_2 u|P_2 s] - \mathbb{E}[P_2 u|P_1 s] - (v - \mathbb{E}[v|P_1 s])\right|^2 \\
&\quad + \mathbb{E}\left|\mathbb{E}[P_2 u|P_1 s] - \mathbb{E}[v|P_1 s]\right|^2.
\end{aligned} \tag{6.2}$$

Here we have suppressed the dependence on s, t in u, v and used that $\mathbb{E}[\mathbb{E}[\cdot|P_2 s]|P_1 s] = \mathbb{E}[\cdot|P_1 s]$. This is a standard identity for conditional expectations, but in our setup it can be verified by a direct calculation. Combining (6.2) with Corollary 1 and the proof of Theorem 2 we get the following corollary.

COROLLARY 2. *Let $\dot{u} = R(u)$ be the spectral method* (3.5) *for the Schrödinger equation* (3.1) *and assume that $u \in \mathbb{R}^n$ for large $n = 2 + 4n'$. Let $\dot{v} = \mathcal{R}(v)$ be the optimal prediction scheme* (3.7) *with $m_2 = 2 + 4m_2'$ and $m_1' < m_2' < n'$. If Assumption* 2 *holds, then*

$$\left(\mathbb{E}\left|\mathbb{E}[P_2 u(s,t)|P_1 s] - \mathbb{E}[v(P_2 s,t)|P_1 s]\right|^2\right)^{1/2} \leq \frac{e^{Lt}-1}{L}\,\frac{c_1\, e^{c_1/24}}{\pi\sqrt{\beta m_2'}}.$$

Remark. Corollary 2 is also valid if we replace $\dot{v} = \mathcal{R}(v)$ by $\dot{v} = \widehat{\mathcal{R}}(v)$. The difficulty lies in the interpretation because the function H_0 in (6.1) is not the Hamiltonian for the truncated scheme (3.6). Thus we cannot interpret (6.1) as an average over the canonical measure.

7. Concluding remarks. The error bounds in this paper are unusual in several aspects. In numerical analysis we typically compare the computed solution with the exact solution, and the estimates involve bounds on the higher derivatives of the exact solution. In this paper we estimate the error in the mean, where the mean is taken over all initial conditions. This is similar to convergence proofs for stochastic differential equations, where the mean is taken over all random forcing functions. Thus, we cannot say that the error is small in any particular experiment, only that it is small with high probability.

Our convergence proof for the nonlinear Schrödinger equation exploits the fact that the equation can be regarded as a Hamiltonian system. In the limit $n \to \infty$ the initial data for the partial differential equation lie in H^α for $\alpha < 1/2$, but they are not smooth enough to be in $H^{1/2}$. Thus we approximate highly irregular functions and our rate of convergence is correspondingly low.

Acknowledgments. We are grateful to Profs. M. Ben-Artzi, A. Chorin, and B. Turkington for many fruitful discussions.

REFERENCES

[1] P. Billingsley, *Probability and Measure*, 3rd ed., John Wiley and Sons, New York, 1995.

[2] C. Boucher, R. Ellis, and B. Turkington, *Derivation of maximum entropy principles in two-dimensional turbulence via large deviations*, J. Statist. Phys., 98 (2000), pp. 1235–1278.

[3] J. Bourgain, *Global Solutions of Nonlinear Schrödinger Equations*, Amer. Math. Soc. Colloq. Publ. 46, AMS, Providence, RI, 1999.

[4] D. Cai and D. W. McLaughlin, *Chaotic and turbulent behavior of unstable* 1*-D nonlinear dispersive waves*. J. Math. Phys., 41 (2000), pp. 4125–4153.

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

[6] A. Chorin, A. Kast, and R. Kupferman, *Optimal prediction of underresolved dynamics*, Proc. Natl. Acad. Sci. USA, 95 (1998), pp. 4094–4098.

[7] A. Chorin, A. Kast, and R. Kupferman, *On the Prediction of Large-Scale Dynamics Using Unresolved Computations*, Contemp. Math. 238, AMS, Providence, RI, 1999.

[8] A. Chorin, A. Kast, and R. Kupferman, *Unresolved computation and optimal prediction*, Comm. Pure Appl. Math., 52 (1999), pp. 1231–1254.

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

[10] O. Hald, *Optimal prediction of the Klein-Gordon equation*, Proc. Natl. Acad. Sci. USA, 96 (1999), pp. 4774–4779.

[11] R. Jordan, B. Turkington, and C. Zirbel, *A mean-field statistical theory for the nonlinear Schrödinger equation*, Phys. D, 137 (1999), pp. 353–391.

[12] A. Kast, *Optimal prediction of stiff oscillatory mechanics*, Proc. Natl. Acad. Sci. USA, 97 (2000), pp. 6253–6257.

[13] M. Marion and R. Temam, *Nonlinear Galerkin methods*, SIAM J. Numer. Anal., 26 (1989), pp. 1139–1157.

[14] V. Zakharov and A. Shabat, *Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media*, Soviet Phys. JETP, 34 (1972), pp. 62–69.