Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 4774–4779, April 1999
Applied Mathematics

Optimal prediction and the Klein–Gordon equation

OLE H. HALD†

Department of Mathematics, University of California, Berkeley, CA 94720–3840

Communicated by Alexandre J. Chorin, University of California, Berkeley, CA, February 18, 1999 (received for review January 11, 1999)

ABSTRACT **The method of optimal prediction is applied to calculate the future means of solutions to the Klein–Gordon equation. It is shown that, in an appropriate probability space, the difference between the average of all solutions that satisfy certain constraints at time $t = 0$ and the average computed by an approximate method is small with high probability.**

1. Introduction

The method of optimal prediction was introduced by Chorin *et al.* (1–3) to study complicated flows, hopefully including turbulence at a future time. Instead of solving a particular initial value problem, we ask for 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 coefficients. Neither will determine the initial data uniquely. The idea then is to use statistical information to compensate for the incompleteness of the initial data. In its most elementary version, the method of optimal prediction is more expensive than solving the original initial value problem. The savings are achieved by finding an evolution equation for the constraints and, from this, determining the average of the solutions for $t > 0$. For nonlinear problems, this can only be done approximately. However, for linear problems, we can estimate the difference between the exact averages and the averages computed by the approximate method. We get the sharpest bound if the constraints are close to an invariant subspace for the adjoint of the differential equation. We apply the theory to the Klein–Gordon equation and prove that the difference between the exact mean at time t and the outcome of an approximate calculation is small with high probability. We also show that the exact averages converge with probability 1 as we increase the dimension of the trial space. This remains true even if the measure is carried by weak solutions that are difficult to obtain individually. We confine ourselves to a single case, but the arguments can be extended to the linear Schrödinger equation and to linear Korteveg de Vries equations.

2. Two Methods

In this section, we will present an exact and an approximate method for finding the average of the solutions to a differential equation. Let L be a real $m \times m$ matrix and let G be a real $m \times n$ matrix of rank $n < m$. We will look at the solutions $u(t)$ of

$$\dot{u}(t) = Lu(t) \qquad [1]$$

and assume that the initial conditions satisfy the constraint

$$G^T u(0) = v_0. \qquad [2]$$

If $S(t) = e^{tL}$ is our fundamental matrix, then $u(t) = S(t)u(0)$. To find the average of all u that satisfy **2**, we need a measure. Let A be a positive definite matrix of order n and define

$$P(u \in B) = \int_B Z^{-1} e^{-\frac{1}{2}u^T A u}\, du,$$

where Z is chosen so that $P(\mathbb{R}^m) = 1$. If $L^T A + AL = 0$, then P is an invariant measure, i.e., $P(B) = P(S(t)B)$ for all t. The matrix A may be chosen in many ways, but there is a natural choice if **1** is a Hamiltonian system. By restricting P to the set $G^T u = v_0$ and normalizing again, we get a measure P' that satisfies

$$\langle u \rangle = \int_{G^T u = v_0} u\, dP' = A^{-1} G M^{-1} v_0, \qquad [3]$$

where $M = G^T A^{-1} G$ (see refs. 1–3). Since $u(t) = S(t)u(0)$, we can determine the average of all solutions that satisfy $G^T u(0) = v_0$ and get

$$\langle u(t) \rangle_{exact} = S(t)\langle u(0) \rangle = S(t) A^{-1} G M^{-1} v_0. \qquad [4]$$

The approximate method is harder to motivate. We would not expect that $G^T u(t) = v_0$ for all $t > 0$; but there may exist a function $v(t)$ such that $G^T u(t) = v(t)$ for all $u(t)$ that satisfy $G^T u(0) = v_0$. The arguments for $t = 0$ are then applicable. After replacing v_0 in **3** by $v(t)$ we see that $\langle u(t) \rangle = A^{-1} G M^{-1} v(t)$. In addition, $v(t) = G^T \langle u(t) \rangle$, and it follows from **1** that $\dot{v}(t) = G^T L \langle u(t) \rangle$. We can now formulate the approximate method. Let $K = LA^{-1}$. Then

$$\langle u(t) \rangle_{approx} = A^{-1} G M^{-1} v(t) \qquad [5]$$

$$\dot{v}(t) = G^T K G M^{-1} v(t), \qquad v(0) = v_0. \qquad [6]$$

If $n \ll m$, it should be cheaper to find the approximate solution than the exact solution. The question is: "How good is the approximation?" To answer this question, we set

$$e(t) = \langle u(t) \rangle_{approx} - \langle u(t) \rangle_{exact}$$

$$E = L^T G + G M^{-1} G^T K G.$$

Suppose $L^T A + AL = 0$. Then $A^{-1} L^T + L A^{-1} = 0$, and it follows from **4**, **5**, and **6** that

$$\begin{aligned} \dot{e}(t) &= A^{-1} G M^{-1} \dot{v}(t) - \dot{S}(t) A^{-1} G M^{-1} v_0 \\ &= A^{-1} G M^{-1} G^T K G M^{-1} v(t) - L S(t) A^{-1} G M^{-1} v_0 \\ &= A^{-1} G M^{-1} G^T K G M^{-1} v(t) - L[A^{-1} G M^{-1} v(t) - e(t)] \\ &= L e(t) + A^{-1} [L^T G + G M^{-1} G^T K G] M^{-1} v(t). \end{aligned}$$

Using the explicit solution of inhomogeneous linear equations, (see ref. 4, p. 78), we obtain

$$e(t) = \int_0^t S(t-s) A^{-1} E M^{-1} v(s)\, ds. \qquad [7]$$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

†To whom reprint requests should be addressed. e-mail: hald@math.berkeley.edu.

LEMMA 1. *If $L^TA + AL = 0$, then*

$$|A^{1/2}e(t)| \le t\,|A^{-1/2}EM^{-1/2}|\,|M^{-1/2}v_0|.$$

Proof: To bound $e(t)$, we need two facts:

$$(A^{1/2}S(t)A^{-1/2})^T(A^{1/2}S(t)A^{-1/2}) = I \quad \textbf{[8]}$$

$$v^T(t)M^{-1}v(t) = v_0^TM^{-1}v_0. \quad \textbf{[9]}$$

Eq. **8** says that $A^{1/2}S(t)A^{-1/2}$ is orthonormal, while **9** corresponds to conservation of energy for **6**. Both are consequences of the assumption $L^TA+AL=0$. To prove **8**, we differentiate with respect to t, use $\dot S = LS$, and obtain

$$\frac{d}{dt}[A^{-1/2}S^T(t)AS(t)A^{-1/2}] = A^{-1/2}S^T(t)[L^TA + AL]S(t)A^{-1/2} = 0.$$

The matrix $A^{-1/2}S^T(t)AS(t)A^{-1/2}$ is, therefore, independent of time and equal to the identity when $t=0$. To prove **9**, we differentiate with respect to t, use **6** and $K^T+K=0$, and get

$$\frac{d}{dt}[v^T(t)M^{-1}v(t)] = v^T(t)M^{-1}G^T(K^T+K)GM^{-1}v(t) = 0.$$

This shows that $v^T(t)M^{-1}v(t)$ is independent of time. We can now complete the proof of *Lemma 1*. Multiplying both sides of **7** by $A^{1/2}$ and using **8**, **9** yields

$$\begin{aligned}|A^{1/2}e(t)| &\le \int_0^t |A^{1/2}S(t-s)A^{-1/2}| \\ &\quad\times |A^{-1/2}EM^{-1/2}|\,|M^{-1/2}v(s)|\,ds \\ &\le t\,|A^{-1/2}EM^{-1/2}|\,|M^{-1/2}v_0|.\end{aligned}$$

This completes the proof.

It follows from *Lemma 1* that $e(t)\equiv 0$ if $E=0$. This will occur if G is a left invariant subspace for L. To prove this, let $L^TG = GB$. Then $G^TA^{-1}L^TG = G^TA^{-1}GB$, and we see that $B = -M^{-1}G^TKG$ and $E = L^TG + GM^{-1}G^TKG = L^TG - GB = 0$.

3. Hamiltonian Systems

It is not true that, for every L, there is a positive definite matrix A such that $L^TA + AL = 0$. You need the eigenvalues of L to be purely imaginary and L must be diagonalizable. However, A exists for linear Hamiltonian systems. Lets look at $\ddot q(t) = -A_0^2q(t)$, where A_0 is positive definite. This equation describes small oscillations around equilibrium. Setting $\dot q(t) = p(t)$, we arrive at

$$\frac{d}{dt}\begin{bmatrix} q(t)\\ p(t)\end{bmatrix} = \begin{bmatrix} 0 & I\\ -A_0^2 & 0\end{bmatrix}\begin{bmatrix} q(t)\\ p(t)\end{bmatrix}.$$

The Hamiltonian for this system is $h = \frac12[p^Tp + q^TA_0^2q]$, i.e., $\dot q_i = \partial_{p_i}h$ and $\dot p_i = -\partial_{q_i}h$. It is natural to constrain p and q separately:

$$\begin{bmatrix} G_q^T & 0\\ 0 & G_p^T\end{bmatrix}\begin{bmatrix} q(0)\\ p(0)\end{bmatrix} = \begin{bmatrix} v_q(0)\\ v_p(0)\end{bmatrix}.$$

More complicated relations between $p(0)$ and $q(0)$ are possible and may be preferable in special cases. Letting $u(t) = \left[\begin{smallmatrix} q(t)\\ p(t)\end{smallmatrix}\right]$, we have $\dot u = Lu$, $G^Tu(0) = v_0$, and $h = \frac12 u^TAu$, as in **1** and **2**, where

$$L = \begin{bmatrix} & I\\ -A_0^2 & \end{bmatrix},\ G = \begin{bmatrix} G_q & \\ & G_p\end{bmatrix},\ A = \begin{bmatrix} A_0^2 & \\ & I\end{bmatrix}.$$

Set $|u|_A = |A^{1/2}u| = (2h)^{1/2}$. Since $M = G^TA^{-1}G$ and $K = LA^{-1}$, we obtain

$$M = \begin{bmatrix} G_q^TA_0^{-2}G_q & \\ & G_p^TG_p\end{bmatrix},$$

$$G^TKG = \begin{bmatrix} & G_q^TG_p\\ -G_p^TG_q & \end{bmatrix}.$$

Note that M is positive definite and that G^TKG is skew symmetric. To simplify the analysis, we assume that $G_p = G_q = G$ and hope that the double use of G will not cause confusion. The differential equation for the approximate method can then be written as

$$\frac{d}{dt}\begin{bmatrix} v_q(t)\\ v_p(t)\end{bmatrix} = \begin{bmatrix} & I\\ -(G^TG)(G^TA_0^{-2}G)^{-1} & \end{bmatrix}\begin{bmatrix} v_q(t)\\ v_p(t)\end{bmatrix} \quad \textbf{[10]}$$

(compare with **6**). If G consists of eigenvectors of A_0^2, then each eigenfrequency of **10** agrees with an eigenfrequency of the original problem and $e(t)=0$. To estimate the error in the approximate method, we must bound $|A^{-1/2}EM^{-1/2}|$ in *Lemma 1*. Since $E = L^TG + GM^{-1}G^TKG$, it follows that $A^{-1/2}EM^{-1/2} = \left[\begin{smallmatrix} 0 & F\\ 0 & 0\end{smallmatrix}\right]$, where

$$\begin{aligned}F &= -A_0G(G^TG)^{-1/2} \\ &\quad + A_0^{-1}G(G^TA_0^{-2}G)^{-1}(G^TG)^{1/2}.\end{aligned} \quad \textbf{[11]}$$

Thus, $|A^{-1/2}EM^{-1/2}| = |F|$, and it is enough to bound the 2-norm of

$$\begin{aligned}F^TF &= (G^TG)^{-1/2}(G^TA_0^2G)(G^TG)^{-1/2} \\ &\quad - (G^TG)^{1/2}(G^TA_0^{-2}G)^{-1}(G^TG)^{1/2}.\end{aligned} \quad \textbf{[12]}$$

To continue the analysis, we turn to a specific problem.

4. Klein–Gordon

In the papers by Chorin *et al.* (1–3), the method of optimal prediction was applied to linear and nonlinear Schrödinger equations. Here, we will study the Klein–Gordon equation

$$u_{tt} = u_{xx} - u \quad \textbf{[13]}$$

on the interval $0 \le x \le 2\pi$ with periodic boundary conditions. The equation describes dispersive waves on a string subject to a restoring force. A similar equation occurs in relativistic quantum field theory (5). The Hamiltonian for **13** is

$$h(t) = \frac12\int_0^{2\pi} (u_t)^2 + (u_x)^2 + (u)^2\,dx. \quad \textbf{[14]}$$

The corresponding Hamiltonian system is

$$\partial_t\begin{bmatrix} u(x,t)\\ \pi(x,t)\end{bmatrix} = \begin{bmatrix} 0 & I\\ \partial_x^2 - I & 0\end{bmatrix}\begin{bmatrix} u(x,t)\\ \pi(x,t)\end{bmatrix}, \quad \textbf{[15]}$$

where $\pi(x,t) = u_t(x,t)$. Note that $A_0^2 = -\partial_x^2 + I$ (for a derivation, see ref. 6). We constrain the initial data by prescribing local averages around the points $x_\alpha = 2\pi\alpha/(2n+1)$ for $\alpha = 0, 1, \dots, 2n$. Specifically,

$$\begin{aligned}\int_0^{2\pi} g(x-x_\alpha)u(x,0)\,dx &= v_{q,\alpha}(0),\\ \int_0^{2\pi} g(x-x_\alpha)\pi(x,0)\,dx &= v_{p,\alpha}(0).\end{aligned} \quad \textbf{[16]}$$

Let us imagine that $v_p(0)$, $v_q(0)$ are given and set $v_0 = \left[\begin{smallmatrix} v_q(0)\\ v_p(0)\end{smallmatrix}\right]$. Following Chorin *et al.* (1–3), we let

$$g(x) = \frac{1}{\sqrt{2\pi}}\sum_{k=-\infty}^{\infty} e^{-k^2\sigma^2/4}\frac{e^{ikx}}{\sqrt{2\pi}}. \quad \textbf{[17]}$$

The function g is positive and 2π periodic, has norm 1, and decreases away from the origin. As $\sigma \to 0$, g tends to a delta function. Since the measure P is finite dimensional, we assume that there is an integer $m \geq 0$ such that all $u(x,t)$ and $\pi(x,t)$ can be written as

$$\sum_{k=-m}^{m} c_k \frac{e^{ikx}}{\sqrt{2\pi}},$$

where $\bar{c}_k = c_{-k}$ and $m = n + r(2n+1)$. The complex notation is equivalent to

$$\frac{a_0}{\sqrt{2\pi}} + \sum_{k=1}^{m}\left(a_k \frac{\cos kx}{\sqrt{\pi}} + b_k \frac{\sin kx}{\sqrt{\pi}}\right)$$

when $c_0 = a_0$ and $c_k = (a_k - ib_k)/\sqrt{2}$ for $k = 1, 2, \ldots, m$. In the expansion of $g(x)$, we replace $\exp(-k^2\sigma^2/4)$ in **17** by 0 if $|k| > m$, thus obtaining $Proj_m\ g(x)$. Our basic variables are not the trigonometric functions but their Fourier coefficients. Let (a_i, b_i) be the Fourier coefficients for $u(x,t)$, and let (α_i, β_i) be the Fourier coefficients for $\pi(x,t)$. Set $q^T = (a_m, \ldots, a_0, b_1, \ldots, b_m)$ and $p^T = (\alpha_m, \ldots, \alpha_0, \beta_1, \ldots, \beta_m)$. We can then rewrite **15** as

$$\frac{d}{dt}\begin{bmatrix} q(t) \\ p(t) \end{bmatrix} = \begin{bmatrix} 0 & I \\ -\Lambda^2 & 0 \end{bmatrix}\begin{bmatrix} q(t) \\ p(t) \end{bmatrix}, \qquad \textbf{[18]}$$

where $\Lambda = \mathrm{diag}(\omega_m, \ldots, \omega_0, \ldots, \omega_m)$ and $\omega_k^2 = k^2 + 1$. Observe the shift in notation: the constants m and n from Section 2 have been replaced by $2(2m+1)$ and $2(2n+1)$. To find the analogue of **16**, we expand $u(x,t)$ in a complex Fourier series, use **17**, and get

$$v_{q,\alpha}(0) = \frac{1}{\sqrt{2\pi}} \sum_{\ell=-m}^{m} e^{-\ell^2\sigma^2/4} e^{i\ell x_\alpha}\ c_\ell.$$

Since the points x_α are equidistant, we have an aliasing effect. Let $\ell = k + j(2n+1)$ with $-n \leq k \leq n$ and $-r \leq j \leq r$. Then

$$\begin{aligned} v_{q,\alpha}(0) &= \sum_{k=-n}^{n} \frac{e^{ikx_\alpha}}{\sqrt{2n+1}} \cdot \sqrt{\frac{2n+1}{2\pi}} \\ &\quad \times \sum_{j=-r}^{r} e^{-[k+j(2n+1)]^2\sigma^2/4}\ c_{k+j(2n+1)} \\ &= \sum_{k=-n}^{n} U_{\alpha k} \cdot w_k. \end{aligned}$$

Note that $\bar{w}_k = w_{-k}$. The matrix U is the building block for the discrete Fourier transform and is unitary. Set

$$\Gamma = \sqrt{\frac{2n+1}{2\pi}}\ \mathrm{diag}\left(e^{-m^2\sigma^2/4}, \ldots, 1, \ldots, e^{-m^2\sigma^2/4}\right).$$

If $c^T = (c_{-m}, \ldots, c_0, \ldots, c_m)$, we can write **16** as $v_q(0) = U[I\ \cdots\ I]\Gamma c$ with $2r+1$ blocks of Is. To express the constraints as a product of real matrices, we let X and Y be of orders $2n+1$ and $2m+1$, respectively, and of the form

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & & & & & & & i \\ & 1 & & & & & i & \\ & & 1 & & & i & & \\ & & & \sqrt{2} & & & & \\ & & 1 & & -i & & & \\ & 1 & & & & -i & & \\ 1 & & & & & & -i & \end{bmatrix}.$$

Note that X, Y are unitary. The matrix $Q = UX$ is orthonormal and the αth row of Q is

$$\sqrt{\frac{2}{2n+1}}\left[\cos(nx_\alpha), \ldots, \cos(x_\alpha), \frac{1}{\sqrt{2}}, \sin(x_\alpha), \ldots, \sin(nx_\alpha)\right].$$

Since $c = Yq$ and $\Gamma Y = Y\Gamma$, we finally obtain $v_q(0) = UXX^*[I\ \cdots\ I]Y\Gamma q = QZ^T\Gamma q$. Because $v_q(0)$, Q, Γ, and q are real, Z must also be real. Let $G^T = QZ^T\Gamma = U[I\ \cdots\ I]Y\Gamma$. The analogue of **16** is then

$$\begin{bmatrix} G^T & \\ & G^T \end{bmatrix}\begin{bmatrix} q(0) \\ p(0) \end{bmatrix} = \begin{bmatrix} v_q(0) \\ v_p(0) \end{bmatrix}. \qquad \textbf{[19]}$$

We can now solve **18** and **19** by the exact method **4** and by the approximate method **5**. To estimate the difference, we use *Lemma 1* and need the following result.

LEMMA 2. *If* $n \geq 1$ *and* $(2n+1)\sigma^2 \geq 2$, *then*

$$|A^{-1/2}EM^{-1/2}| \leq (1.6)\,(2n+1)\,e^{-(2n+1)\sigma^2/4}.$$

Proof: To bound $|F|$, we will determine G^TG, $G^TA_0^2G$, $G^TA_0^{-2}G$ in **12** explicitly. Observe that $A_0 = \Lambda$. By using the complex representation of G, $Y\Gamma = \Gamma Y$, and $YY^* = I$, we see that $G^TG = UD_1U^*$, where

$$D_1 = \frac{2n+1}{2\pi}\ \mathrm{diag}_{-n\leq k\leq n}\left(\sum_{j=-r}^{r} e^{-[k+j(2n+1)]^2\sigma^2/2}\right).$$

Interchanging k and j with $-k$ and $-j$ shows that $(D_1)_{-k} = (D_1)_k$, which implies that $X^*D_1 = D_1X^*$. Since $U = QX^*$, we conclude that $G^TG = QD_1Q^T$. Similar arguments give $G^TA_0^2G = QD_2Q^T$ and $G^TA_0^{-2}G = QD_3Q^T$, where

$$\begin{aligned} D_2 &= \frac{2n+1}{2\pi}\ \mathrm{diag}_{-n\leq k\leq n} \\ &\quad \left(\sum_{j=-r}^{r} e^{-[k+j(2n+1)]^2\sigma^2/2}\{[k+j(2n+1)]^2+1\}\right) \\ D_3 &= \frac{2n+1}{2\pi}\ \mathrm{diag}_{-n\leq k\leq n} \\ &\quad \left(\sum_{j=-r}^{r} e^{-[k+j(2n+1)]^2\sigma^2/2}\{[k+j(2n+1)]^2+1\}^{-1}\right). \end{aligned}$$

We can now determine **12** explicitly. Since Q is orthonormal, it follows that

$$F^TF = Q[D_1^{-1}D_2 - D_1D_3^{-1}]Q^T.$$

If $r = 0$, then $E = F = 0$ and the approximate and exact method agree. Let $r \geq 1$ and suppose that the largest term in the diagonal matrix $D_1^{-1}D_2 - D_1D_3^{-1}$ occurs in the kth position. Set $d_j = \exp(-[k+j(2n+1)]^2\sigma^2/2)$ and $\lambda_j = [k+j(2n+1)]^2+1$. Extracting the leading order term in each sum, we get

$$\begin{aligned} &(D_1^{-1}D_2 - D_1D_3^{-1})_k \\ &\quad = \frac{\sum d_j\lambda_j}{\sum d_j} - \frac{\sum d_j}{\sum d_j\lambda_j^{-1}} \\ &\quad = \frac{(d_0\lambda_0 + a)(d_0\lambda_0^{-1} + b) - (d_0 + c)^2}{(d_0 + c)\ (d_0\lambda_0^{-1} + b)} \\ &\quad = \frac{a(d_0\lambda_0^{-1} + b) - d_0(c - \lambda_0 b) - c(d_0 + c)}{(d_0\lambda_0^{-1} + b)\ (d_0 + c)}. \end{aligned}$$

Since $c > \lambda_0 b$ and the 2-norm is invariant under orthonormal transformations, we see that

$$|F|^2 \le \frac{a}{d_0} = \sum_{|j|=1}^{r} e^{\{k^2-[k+j(2n+1)]^2\}\sigma^2/2} \times \{[k+j(2n+1)]^2+1\}. \quad [20]$$

To estimate the exponential, we observe that

$$\begin{aligned} k^2-[k+j(2n+1)]^2 &\le -j^2(2n+1)^2+2|k||j|(2n+1) \\ &= -(j^2-|j|)(2n+1)^2-|j|(2n+1) \\ &\quad -2|j|(n-|k|)(2n+1). \end{aligned} \quad [21]$$

Combining **20** and **21** with $|k| \le n$ results in

$$|F|^2 \le \sum_{j=1}^{r} e^{-(2n+1)\sigma^2/2} e^{-(j^2-j)(2n+1)^2\sigma^2/2} 2\,[j^2(2n+1)^2+k^2+1].$$

Since $k^2+1 \le (2n+1)^2/4$ when $|k| \le n$ and $n \ge 1$, we conclude that

$$|F|^2 \le (2n+1)^2 e^{-(2n+1)\sigma^2/2} \sum_{j=1}^{r} e^{-(j^2-j)(2n+1)^2\sigma^2/2}\, 2\,(j^2+1/4).$$

The last sum is less than 2.53 when $(2n+1)\sigma^2/2 \ge 1$ and $n \ge 1$. This completes the proof.

5. Stochastic Convergence

By combining *Lemma 1* and *Lemma 2*, we can bound the difference between the exact and the approximate method. Since $|M^{-1/2}v_0|$ depends on $2n+1$, σ, and v_0, we have not established convergence. Suppose $v_p(0)$ and $v_q(0)$ are generated by two particular random functions u and π, with $u(x,0)$ looking like Brownian motion and $\pi(x,0)$ resembling white noise. We can then show that the approximate method is close to the exact method if n is large. The rate of convergence is high if there is a substantial overlap of the kernels in the constraints. To measure the error we use the norm $|\cdot|_A$, the square of which equals twice the total energy.

THEOREM 1. *Let $n \ge 1$, and assume that $(2n+1)\sigma^2 \ge 6(\nu+1)\log(2n+1)$ with $\nu > 0$. Let p and q be picked at random with respect to P and set $v_p(0) = G^T p$, $v_q(0) = G^T q$. Consider all solutions of* **15** *that satisfy* **16**. *Then*

$$\left| \left\langle \begin{pmatrix} u(x,t) \\ \pi(x,t) \end{pmatrix} \right\rangle_{exact} - \left\langle \begin{pmatrix} u(x,t) \\ \pi(x,t) \end{pmatrix} \right\rangle_{approx} \right|_A \le \frac{2.3\ t}{(2n+1)^\nu}$$

with probability greater than $1-(2n+1)^{-\nu}$.

Proof: It follows from *Lemmas 1* and *2* that

$$|A^{1/2}e(t)| \le (1.6)\, t\,(2n+1)\, e^{-(2n+1)\sigma^2/4} |M^{-1/2}v_0|,$$

where $v_0 = \begin{bmatrix} v_q(0) \\ v_p(0) \end{bmatrix}$. To complete the proof, we use Chebyshev's inequality. Let E be the expected value corresponding to P. Since $2.3 > 1.6\sqrt{2}$ and σ is bounded below, we obtain

$$\begin{aligned} P\left(|A^{1/2}e(t)| > \frac{2.3\ t}{(2n+1)^\nu}\right) &\le P\left(|M^{-1/2}v_0| > \frac{\sqrt{2}e^{(2n+1)\sigma^2/4}}{(2n+1)^{\nu+1}}\right) \\ &\le \frac{E(|M^{-1/2}v_0|^2)}{2(2n+1)^{\nu+1}}. \end{aligned} \quad [22]$$

Using the definition of M from Section 3 in conjunction with $A_0 = \Lambda$ and **19**, we get

$$\begin{aligned} E(v_0^T M^{-1} v_0) &= E[(\Lambda q)^T \Lambda^{-1} G (G^T \Lambda^{-2} G)^{-1} G^T \Lambda^{-1} (\Lambda q)] \\ &\quad + E[p^T G (G^T G)^{-1} G^T p]. \end{aligned}$$

Since Λ is diagonal, the measure P is given by

$$dP = Z^{-1} e^{-\frac{1}{2}[a_0^2+\alpha_0^2+\sum_{k=1}^m \omega_k^2(a_k^2+b_k^2)+(\alpha_k^2+\beta_k^2)]}\, da_0 \cdots d\beta_m$$

$$Z = 2\pi \prod_{k=1}^{m} \left(\frac{2\pi}{\omega_k^2} \cdot 2\pi\right).$$

The components of Λq and p are, therefore, independent Gaussian random variables with mean 0 and variance 1, and it follows that

$$\begin{aligned} E(v_0^T M^{-1} v_0) &= \mathrm{tr}[\Lambda^{-1} G (G^T \Lambda^{-2} G)^{-1} G^T \Lambda^{-1}] \\ &\quad + \mathrm{tr}[G (G^T G)^{-1} G^T] \\ &= \mathrm{tr}[(G^T \Lambda^{-2} G)^{-1} (G^T \Lambda^{-2} G)] \\ &\quad + \mathrm{tr}[(G^T G)^{-1} (G^T G)] \\ &= 2(2n+1). \end{aligned}$$

Here, tr = trace, and we have used $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ if A is an $n \times m$ matrix and B is $m \times n$. Combining the last result with **22** and taking the complementary event finishes the proof.

We remark that the components of v_0 are strongly correlated. Indeed, it follows from **19** that

$$E(v_0 v_0^T) = \begin{bmatrix} G^T\Lambda^{-1} & \\ & G^T \end{bmatrix} E\begin{bmatrix} (\Lambda q)(\Lambda q)^T & \\ & pp^T \end{bmatrix} \times \begin{bmatrix} \Lambda^{-1} G & \\ & G \end{bmatrix} = M.$$

Using the spectral decomposition of $G^T G$, we can calculate the variances explicitly and get

$$\begin{aligned} \mathrm{var}[v_{p,\alpha}(0)] &= \sum_{k=-n}^{n} (Q_{\alpha,k})^2 (D_1)_k = \frac{1}{2\pi} \sum_{\ell=-m}^{m} e^{-\ell^2\sigma^2/2} \\ &= \int_0^{2\pi} [Proj_m\ g(x)]^2\, dx. \end{aligned}$$

The variance of $v_{p,\alpha}(0)$ is, therefore, of order $1/(\sqrt{2\pi}\sigma)$. For $v_{q,\alpha}(0)$ we get an additional factor of $\{\ell^2+1\}^{-1}$, and $1/(2\pi) < \mathrm{var}(v_{q,\alpha}(0)) < \coth(\pi)/2$.

Suppose the components of v_0 are chosen as independent, normally distributed random variables with mean 0 and variance 1. If $n \ge 4$ and $(2n+1)\sigma^2 \ge 6(\nu+1)\log(2n+1)$, we can show that any interval longer than 4 contains points t for which

$$E(|e(t)|_A^2) \ge (2n+1)^{(\nu+1)(2n+1)/4-1}.$$

The initial constraint v_0 must, therefore, be consistent with the mathematical model if we want convergence.

6. Convergence in L^2

In Section 5, we compared the outcome of two numerical methods. Both are defined on finite dimensional spaces and involve a finite number of Fourier coefficients. What happens if we fix the number of constraints but increase the dimension of the space? Each random choice of the Fourier coefficients $\{a_i, b_i, \alpha_i, \beta_i\}_{i=0}^{\infty}$ yields a sequence of constraint values. Such a sequence may or may not converge. We will show that the sequence of exact solutions generated by the constraints converges with probability 1. Note that we are not comparing results for different values of n. They differ by large amounts.

Let $m = n + r(2n+1)$. By solving **18** and **19** explicitly and using **4**, we find that the Fourier coefficients for the average of all solutions of **13** with the constraints **14** satisfy

$$\begin{bmatrix} \Lambda & \\ & I \end{bmatrix} \left\langle \begin{bmatrix} q(t) \\ p(t) \end{bmatrix} \right\rangle_{exact,r} = \begin{bmatrix} \cos \Lambda t & \sin \Lambda t \\ -\sin \Lambda t & \cos \Lambda t \end{bmatrix} \begin{bmatrix} \Lambda^{-1} & \\ & I \end{bmatrix} \times \begin{bmatrix} G & \\ & G \end{bmatrix} \begin{bmatrix} (G^T \Lambda^{-2} G)^{-1} & \\ & (G^T G)^{-1} \end{bmatrix} \begin{bmatrix} G^T q \\ G^T p \end{bmatrix}.$$

The index r reminds us of the dimension. Since $G^T = U[I \ \cdots \ I]\Gamma Y$ and $Y\Lambda = \Lambda Y$, it follows that

$$\langle \Lambda q(0) \rangle_{exact,r} = Y^* \begin{bmatrix} \Delta_{-r} \\ \vdots \\ \Delta_r \end{bmatrix} \left[\Delta_{-r}^2 + \cdots + \Delta_r^2\right]^{-1} \left[\Delta_{-r} \ \cdots \ \Delta_r\right] Y \Lambda q,$$

where $\Delta_j = \Gamma_j \Lambda_j^{-1}$. We get the formula for $\langle p(0) \rangle$ by replacing Λq by p and Δ_j by Γ_j. Next, let P_r be the probability measure from Section 5 on $\Omega_r = \mathbb{R}^{2(2m+1)}$. Since the random variables a_i, b_i, α_i, and β_i are independent, the measures P_r are consistent, and there is a probability space $(\Omega, \mathcal{F}, P)$ such that $P_r = P|\Omega_r$; see Billingsley (ref. 7, section 36). We can now formulate the following theorem.

THEOREM 2. *Let $n \geq 1$, and assume that $(2n+1)\sigma^2 \geq 6(\nu+1)\log(2n+1)$ with $\nu \geq 0$. Set $\epsilon_r = 4(2n+1)^{-(\nu+1)[1+r+r^2(2n+1)]}$. The limit of the exact method exists for almost all choices of the random Fourier coefficients, and*

$$\left| \left\langle \begin{pmatrix} u(x,t) \\ \pi(x,t) \end{pmatrix} \right\rangle_{exact,r} - \left\langle \begin{pmatrix} u(x,t) \\ \pi(x,t) \end{pmatrix} \right\rangle_{exact,\infty} \right|_A < \epsilon_r$$

with probability greater than $1 - \epsilon_r$.

Proof: Our proof is based on Borel-Cantelli (see ref. 7, p. 53). Here is an outline. Let $\psi_r(x,t,\omega) = \langle \binom{u(x,t)}{\pi(x,t)} \rangle_{exact,r}$, and define $\mathcal{A}_r = \{\omega\colon |\psi_r - \psi_{r+1}|_A \geq \epsilon_r\}$. Since $P(\mathcal{A}_r) < \epsilon_r$ and $\sum \epsilon_r < \infty$, it follows that $P(\bigcup_{s=0}^{\infty} \bigcap_{r=s}^{\infty} \mathcal{A}_r^c) = 1$. The sequence $\{\psi_r\}_{r=0}^{\infty}$ is, therefore, Cauchy for almost all $\omega \in \Omega$ and converges to an element in $H^1 \oplus H^0$. Instead of working with the random functions, we work with the Fourier coefficients and embed the smaller space into the larger space. Let $r < s$, and set

$$B_1^T = [\Delta_{-s} \cdots \Delta_s]$$
$$B_2^T = [\Delta_{-s} \cdots \Delta_{-r-1} \ 0 \cdots 0 \ \Delta_{r+1} \cdots \Delta_s]$$
$$B_3^T = [0 \cdots 0 \ \Delta_{-r} \cdots \Delta_r \ 0 \cdots 0]$$

Note that $B_i^T B_i$ are diagonal matrices of order $2n+1$. We can now write

$$\langle \Lambda q(0) \rangle_{exact,s} - \langle \Lambda q(0) \rangle_{exact,r} = b_1 + b_2 + b_3,$$

where

$$b_1 = Y^* B_2 (B_1^T B_1)^{-1} B_1^T Y \Lambda q$$
$$b_2 = Y^* B_3 (B_1^T B_1)^{-1} B_2^T Y \Lambda q$$
$$b_3 = -Y^* B_3 (B_1^T B_1)^{-1} B_2^T B_2 (B_3^T B_3)^{-1} B_3^T Y \Lambda q$$

Using Chebyshev's inequality and Cauchy-Schwarz, we see that

$$P\left(\left|\sum_{i=1}^{3} b_i\right| > \epsilon\right) \leq \epsilon^{-2} E \left|\sum_{i=1}^{3} b_i\right|^2 \leq 3\epsilon^{-2} \sum_{i=1}^{3} E|b_i|^2.$$

Since Λq are independent Gaussian random variables with mean 0 and variance 1 and $YY^* = I$, we get

$$E(b_1^T b_1) = \mathrm{tr}[Y^* B_1 (B_1^T B_1)^{-1} B_2^T Y Y^* B_2 (B_1^T B_1)^{-1} B_1^T Y] = \mathrm{tr}[(B_1^T B_1)^{-1} B_2^T B_2].$$

Now $\omega_k^2/\omega_{k+j(2n+1)}^2$ is less than 1 if $jk < 0$ and less than 0.2 if $jk \geq 0$ and $j \neq 0$. Combining $(B_1^T B_1)^{-1} < \Delta_0^{-2}$ with Eq. **21** and using $(2n+1)\sigma^2 \geq 6$ and $2n+1 \geq 3$, we obtain

$$E|b_1|^2 \leq \sum_{k=-n}^{n} \sum_{j=r+1}^{s} e^{-[(j^2-j)(2n+1)^2 - j(2n+1) - 2j(n-|k|)(2n+1)]\sigma^2/2} \quad (1.2)$$
$$\leq e^{-r^2(2n+1)^2\sigma^2/2} e^{-(r+1)(2n+1)\sigma^2/2} \sum_{\ell=1}^{\infty} (1.2)\, e^{-(\ell^2-\ell)3\cdot 3} \times \sum_{k=-n}^{n} e^{-(n-|k|)6}.$$

Since $(2n+1)\sigma^2 \geq 6(\nu+1)\log(2n+1)$ and the product of the two sums is less than 2.5, we conclude that

$$E|b_1|^2 \leq 2.5\,(2n+1)^{-3(\nu+1)[1+r+r^2(2n+1)]} = \epsilon'.$$

By almost the same arguments, we get $E|b_2|^2 \leq \epsilon'$ for the second term and for the third term, we find that $E|b_3|^2 \leq \mathrm{tr}[(\Delta_0^{-2} B_2^T B_2)^2]$. Since $\Delta_0^{-2} B_2^T B_2$ is diagonal with all terms less than 1, it follows that $E|b_3|^2 \leq \epsilon'$. The arguments for the p terms are similar and by combining all estimates, we obtain

$$P(|\psi_r - \psi_s|_A \geq (0.9)\epsilon_r) \leq (0.9\epsilon_r)^{-2}\ 3 \cdot 2 \cdot 3 \cdot \epsilon' < (0.9)\epsilon_r. \quad \textbf{[23]}$$

Thus $P(\mathcal{A}_r) < \epsilon_r$. Since $\sum \epsilon_r < \infty$, we conclude from Borel-Cantelli that $P(\bigcap_{s=0}^{\infty} \bigcup_{r=s}^{\infty} \mathcal{A}_r) = 0$. The sequence $\{\psi_r\}$ is, therefore, a Cauchy sequence with probability 1 and $\psi_r \to \psi_\infty$ in $H^1 \oplus H^0$. To estimate $\psi_r - \psi_\infty$, we set $\mathcal{B}_s = \bigcup_{r=s}^{\infty} \mathcal{A}_r$. Since $\mathcal{B}_1 \supset \mathcal{B}_2 \supset \cdots$, there exists an $s > r$ such that $P(\mathcal{B}_s) < (0.1)\epsilon_r$. Let $\mathcal{A}_{rs} = \{\omega\colon |\psi_r - \psi_s|_A < (0.9)\epsilon_r\}$. It follows from **23** that

$$1 - (0.9)\epsilon_r \leq P(\mathcal{A}_{rs} \cap \mathcal{B}_s^c) + P(\mathcal{A}_{rs} \cap \mathcal{B}_s).$$

The last term is less than $(0.1)\epsilon_r$, and for almost all $\omega \in \mathcal{A}_{rs} \cap \mathcal{B}_s^c$, we have

$$|\psi_r - \psi_\infty|_A \leq |\psi_r - \psi_s|_A + |\psi_s - \psi_{s+1}|_A + \cdots < (0.9)\epsilon_r + \sum_{j=s}^{\infty} \epsilon_j < \epsilon_r.$$

Since $|(\psi_r - \psi_\infty)(x,t,\omega)|_A$ does not depend on time, this completes the proof.

Suppose the constraints in **16** are generated by a smooth solution $\binom{u_0}{\pi_0}$ of **15**. If $n \geq 1$ and $(2n+1)\sigma^2 \geq 6(\nu+1)\log(2n+1)$ with $\nu \geq 0$, we can show that

$$\left| \left\langle \begin{pmatrix} u(x,t) \\ \pi(x,t) \end{pmatrix} \right\rangle_{exact,r} - \begin{pmatrix} u_0(x,t) \\ \pi_0(x,t) \end{pmatrix} \right|_A \leq \frac{3\sqrt{2.5}}{(2n+1)^{(3/2)(\nu+1)}} \left| \begin{pmatrix} u_0 \\ \pi_0 \end{pmatrix} \right|_A + \frac{1}{(n+1)^s} \left| \partial_x^s (I - Proj_n) \begin{pmatrix} u_0 \\ \pi_0 \end{pmatrix} \right|_A$$

The method of optimal prediction can therefore also be used, in principle, to solve the Klein–Gordon equation with smooth initial data.

I thank Bradford Chin, Alexandre Chorin, Craig Evans, William Kahan, and Nicolai Reshetikhin for helpful discussions. This work was supported in part by the National Science Foundation under Grant DMS-95-03483.

1. Chorin, A. J., Kast, A. P. & Kupferman, R. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 4094–4098.
2. Chorin, A. J., Kast, A. P. & Kupferman, R. (1999) *Comm. Pure Appl. Math.*, in press.
3. Chorin, A. J., Kast, A. P. & Kupferman, R. (1999) *Contemp. Math.*, in press.
4. Coddinton, E. A. & Levinson, N. (1955) *Theory of Ordinary Differential Equations* (McGraw-Hill, New York).
5. Ryder, L. H. (1996) *Quantum Field Theory* (Cambridge Univ. Press, Cambridge, U.K.), 2nd Ed.
6. Goldstein, H. (1950) *Classical Mechanics* (Addison-Wesley, Reading, MA).
7. Billingsley, P. (1986) *Probability and Measure* (Wiley, New York), 2nd Ed.