JOURNAL OF COMPUTATIONAL PHYSICS 8, 472–482 (1971)

Hermite Expansions in Monte-Carlo Computation*

ALEXANDRE JOEL CHORIN

Courant Institute of Mathematical Sciences, New York University, New York 10012

The expansion of estimators and estimands in Hermite (or Wiener–Hermite) series can substantially improve the accuracy and efficiency of Monte–Carlo algorithms.

INTRODUCTION

Monte–Carlo computations often yield numerical answers of limited accuracy, and are therefore employed as a last resort. It has been found, however, that some of the limitations of Monte–Carlo methods can be overcome through a judicious use of orthogonal expansions. When a numerical answer is obtained as the expected value of an estimator, expansion of that estimator in a series of orthogonal functions (or functionals) can reduce the variance of the estimate. Expansion of the estimand in orthogonal polynomials can increase accuracy and efficiency and simplify the solution of nonlinear problems. Hermite polynomials will be seen to play a particularly important role, and the purpose of this paper is to explain various aspects of their use through the solution of simple problems in one dimension. Generalization to multidimensional problems is immediate, and justifies the introduction of these methods.

The Hermite polynomials

$$H_n(u) = (-1)^n e^{u^2} \frac{d^n}{du^n} e^{-u^2}/c_n , \qquad n = 0, 1,..., c_n = \sqrt{2^n n!}$$

are orthonormal with respect to the weight $\pi^{-1/2}e^{-u^2}$, i.e.,

$$\pi^{-1/2} \int H_n(u)\, H_m(u)\, e^{-u^2} du = \delta_{n,m} ;$$

the set $\{H_n(u)\, e^{-u^2/2}\}$ is complete in $L_2(-\infty, +\infty)$ [5].

* Supported by the AEC Computing and Applied Mathematics Center, Courant Institute of Mathematical Sciences, New York University, under Contract AT(30-1)-1480 with the U.S. Atomic Energy Commission.

VARIANCE REDUCTION; SOLUTION OF THE HEAT EQUATION

Let $v(x, t)$ denote the solution of the heat equation

$$v_t = v_{xx}, \tag{1}$$

with initial data $v(x, 0) = f(x)$. We have

$$v(x, t) = (4\pi t)^{-1/2} \int f(x')\, e^{-(x-x)'^2/4t}\, dx';$$

or, writing

$$u = (x - x')/\sqrt{4t}$$

and

$$g(x, u) = f(x'),$$

we obtain

$$\begin{aligned} v(x, t) &= \pi^{-1/2} \int g(x, u)\, e^{-u^2}\, du, \\ &= E[g(x, \xi)], \end{aligned} \tag{2}$$

where ξ is a gaussianly distributed random variable, i.e., the probability that ξ lie between y and $y + dy$ is

$$P(y < \xi < y + dy) = \pi^{-1/2} e^{-y^2}\, dy,$$

and $E[g(x, \xi)]$ denotes the expected value of $g(x, \xi)$. ξ can be readily sampled, e.g., by an algorithm due to Paley and Wiener [7, p. 146]. $E[g]$ can be estimated by (see [4])

$$E[g] \cong N^{-1} \sum_{i=1}^{N} g(x, \xi_i),$$

where the ξ_i are drawn from the distribution above. The standard deviation SD of this estimate, which yields the order of magnitude of the error, is

$$N^{-1/2}\left(\pi^{-1/2} \int g^2 e^{-u^2}\, du - \left(\pi^{-1/2} \int g e^{-u^2}\, du\right)^2\right)^{1/2} \tag{3}$$

and may be inacceptably large for reasonable sample sizes N. Our goal is to obtain an estimate of v with smaller SD.

Assume $g(x, u)\, e^{-u^2/2}$ lies in $L_2(-\infty, +\infty)$ as a function of u, and expand $g(x, u)\, e^{-u^2/2}$ in Hermite series as a function of u:

$$g(x, u)\, e^{-u^2/2} = \sum_{n=1}^{\infty} a_n H_n(u)\, e^{-u^2/2},$$

or

$$g(x, u) = \sum_{n=1}^{\infty} a_n H_n(u),$$

where

$$a_n = \pi^{-1/2} \int H_n(u)\, g(x, u)\, e^{-u^2}\, du$$

$$= E[H_n(\xi)\, g(x, \xi)].$$

We note (i) $a_0 = v(x, t)$, (ii) $E[a_n H_n(\xi)] = 0$, $n \geqslant 1$, (iii) a_n can be evaluated by the same process, using the same samples ξ_i as $v(x, t) = a_0$. The method of reduction of the SD will be based on the following observations: For any $n + 1$ constants $b_0, b_1, \ldots, b_n$, we have

$$E[g(x, \xi)] = b_0 + E\left[g(x, \xi) - \sum_{n=0}^{m} b_n H_n(\xi)\right].$$

This identity does not imply that if the expectation on the left side and the expectation on the right side are estimated by a Monte–Carlo algorithm, the SD's will be equal. Thus we look for constants b_n which would lead to a SD as small as possible. If we could set $b_n = a_n$, the function whose expected value is being taken on the right side would be small, and so would be the SD of the estimate. Since the a_n are not known, we estimate them in the course of computation; these estimates will differ from the true a_n, but will still lead to a major reduction in variance if the resulting estimator is used rather than the original estimator; this will be shown below.

It is worth noting that for $m = 0$ (i.e., only the zero order term is used in the series), our method reduces to the standard variance reduction by subtraction of the estimate [4].

We thus proceed as follows: We generate N samples ξ_i, estimate a_n by

$$a_n^* = N^{-1} \sum_{i=1}^{N} g(x, \xi_i)\, H_n(\xi_i),\ n = 0, 1, \ldots, m,$$

and then pick N new samples ξ_i' and estimate $v(x, t)$ using

$$v(x, t) = a_0^* + E\left[g(x, \xi) - \sum_{n=0}^{m} a_n^* H_n(\xi)\right], \tag{4a}$$

i.e.,

$$v(x, t) \cong a_0^* + N^{-1} \sum_{i=1}^{N} \left(g(x, \xi_i') - \sum_{n=0}^{m} a_n^* H_n(\xi_i')\right). \tag{4b}$$

It should be noted that although the a_n^* are random variables, the expected value in (4a) is not, since this expected value is independent of the values of the coefficients of $H_n(\xi)$. Furthermore, the a_n^* are independent of the $H_n(\xi_i')$, since new samples ξ_i' have been generated. We have, in fact, a random estimator with a nonrandom expected value. However, the SD of the estimate (4b) will, of course, be random.

The SD of the estimate a_n^* of a_n is bounded by

$$N^{-1/2}\left(\pi^{-1/2} \int g^2 H_n^2 e^{-u^2}\, du\right)^{1/2};$$

Let $C = \limsup_{-\infty < x < \infty} |f(x)|$; the SD of a_n^* is bounded by

$$CN^{-1/2}\left(\int \pi^{-1/2} H_n^2 e^{-u^2}\, du\right)^{1/2} = CN^{-1/2};$$

thus, $a_n^* = a_n + \delta a_n$, $\delta a_n = 0(CN^{-1/2})$. Put

$$r_m = g(x, u) - \sum_{n=0}^{m} a_n H_n(u),$$

$$\| r_m \|^2 = \pi^{-1/2} \int r_m^2 e^{-u^2}\, du;$$

r_m is the remainder in the expansion of g. Thus

$$g(x, \xi) - \sum_{n=0}^{m} a_n^* H_n(\xi) = r_m - \sum_{n=0}^{m} \delta a_n H_n(\xi)$$

and the estimate (4b) of v has SD

$$N^{-1/2}\left\{\pi^{-1/2} \int \left(r_m - \sum_{n=0}^{m} \delta a_n H_n(u)\right)^2 e^{-u^2}\, du - \left(\pi^{-1/2} \int \left(r_m - \sum \delta a_n H_n(u)\right) e^{-u^2}\, du\right)^2\right\}^{1/2}$$

which is bounded by

$$N^{-1/2}\left\{\pi^{-1/2}\int\left(r_m-\sum_{n=0}^{m}\delta a_n H_n(u)\right)^2 e^{-u^2}\,du\right\}^{1/2},$$

which, in turn, is bounded by

$$N^{-1/2}\|r_m\|_\infty + N^{-1}0(Cm). \tag{5}$$

For fixed m, (5) is $0(N^{-1/2})$ for large N, but the constant may be small; thus we have an effective way for reducing variance. For f smooth enough and $m = 0(N^\epsilon)$; $\epsilon > 0$, (5) is $0(N^{-(1-\epsilon)})$. For each N, an optimum m could be determined by trial and error. For moderate m, the amount of labor required is not substantially larger than twice the amount required in a more standard computation; the recursion relations between Hermite polynomials can be put to good use. The increase in accuracy may be substantial. In Table I the results of a test computation are exhibited. The initial data are $f(x) = \sin \pi x$; $t = 0.049382$. The results of a nonaccelerated computation, a computation with $m = 4$, as well as their respective SD's, are compared with the exact answer. A method for reducing variance further with no increase in N will be presented in the next section.

TABLE I

Hermite Acceleration of a Monte-Carlo Computation
($t = 0.049382$, $N = 1000$)

x/π	No acceleration	$m = 4$	Exact answer
0.1	0.2070 ± 0.019	0.1902 ± 0.0013	0.1898
0.2	0.3654 ± 0.018	0.3627 ± 0.0015	0.3610
0.3	0.4945 ± 0.016	0.4963 ± 0.0019	0.4969
0.4	0.5806 ± 0.014	0.5841 ± 0.0007	0.5841
0.5	0.6041 ± 0.014	0.6146 ± 0.0004	0.6142

It should be noted that, in quadrature problems over bounded intervals, orthogonal polynomials other than Hermite polynomials may be used. In particular, the uniform boundedness of Legendre polynomials in the maximum norm makes possible an iterated use of the preceding procedure and further drastic reduction in variance.

The solution of the heat equation was reduced to a quadrature problem because the Green's function was known in advance. The acceleration method is in no way restricted to such special cases, and, in fact, may be used to generate Green's functions. In particular, the acceleration method is applicable whenever the solution

can be represented, explicitly or implicitly, as a function-space integral. It is obvious that $v(x, t)$ has such a representation: Let C be the space of real functions $z(t)$, continuous on the interval $0 \leqslant t \leqslant 1$, such that $z(0) = 0$; let dz denote the Wiener measure on C [7, 8]. Let $F[z]$ be the functional on C defined by

$$F[z] = g(x, z(1)).$$

Then

$$v(x, t) = \int F[z]\, dz.$$

As a further illustration, consider the solution $v_{a,b}(x, t)$ of the heat Eq. (1) on $a \leqslant x \leqslant b$, with $v_{a,b}(x, 0) = f(x)$, $v_{a,b}(a, t) = A$, $v(b, t) = B$. Put $(x - x')/\sqrt{4t} = u$, $f(x') = g(x, u)$, as above, and $a' = (x - a)/\sqrt{4t}$, $b' = (x - b)/\sqrt{4t}$.

Let $F_{a,b}[z]$ be the functional on C defined by $F_{a,b}[z] = g(x, z(1))$ if $z(t) \neq a'$, $z(t) \neq b'$, $0 \leqslant t \leqslant 1$, $F_{a,b}[z] = A$ if $z(t_1) = a'$, t_1 in $[0, 1]$, and $z(t) \neq b'$, $0 \leqslant t \leqslant t_1$ $F_{a,b}[z] = B$ otherwise.
Then

$$v_{a,b}(x, t) = \int F_{a,b}[z]\, dz.$$

the random paths $z(t)$ are readily constructed, and an estimator of $v_{a,b}(x, t)$ obtained. Our goal is to reduce the variance of the estimate. To achieve this reduction, we expand $F_{a,b}[z]$ in Wiener–Hermite functionals (for definitions and proof of completeness, see [1]). Wiener–Hermite functionals are Hermite functions of "linear functionals of the paths". If q is an integer such that in the time interval t/q a negligible fraction of paths $z(t)$ can reach the boundary $z = a'$ or $z = b'$, then at most q arguments need be used in the expansion. There is no crucial difference between the expansion of a functional of a one-dimensional path in Wiener–Hermite functionals in q arguments and the expansion of a functional of a q-dimensional path in q-dimensional Hermite polynomial series. The construction of expansion coefficients is obvious.

Scaling

The method of acceleration discussed above is of course most effective when the Hermite expansion of $g(x, u)$ is rapidly convergent. The rate of convergence of the expansion is to some extent under our control. For any $\alpha > 0$ we have

$$\begin{aligned} v(x, t) &= \pi^{-1/2} \int g(x, u)\, e^{-u^2}\, du = \pi^{-1/2} \int g_\alpha(x, u)\, e^{-\alpha u^2} \alpha^{1/2}\, du, \\ g_\alpha &= \alpha^{-1/2} g(x, u)\, e^{-(1-\alpha)u^2}, \end{aligned} \tag{6}$$

or

$$v(x, t) = E[g_\alpha(x, \xi_\alpha)],\ P(y < \xi_\alpha < y + dy) = (\pi/\alpha)^{-1/2}\, e^{-\alpha y^2}\, dy.$$

ξ_α is readily sampled. We now expand g_α in Hermite series with argument $\sqrt{\alpha}u$, which may converge more rapidly than the expansion of $g(x, u)$ in a series with argument u. No optimal determination of α has been made. The following heuristic consideration is helpful: α should be chosen so that g_α and $H_n(\sqrt{\alpha}u)$ have comparable characteristic lengths, i.e., they vary by comparable relative amounts over equal ranges of u. We have, of course,

$$a_n = E[g_\alpha(x, \xi_\alpha)\, H_n(\xi_\alpha)].$$

The scaling is essential for estimating $v(x, t)$ for large t. For $\alpha < 1$ we have the added advantage that g_α decays more rapidly as $u \to \pm\infty$ than g, thus reducing the variance of $a_n{}^*$ for n large enough.

TABLE II

Hermite Acceleration with Scaling
($t = 0.049382$, $\alpha = \frac{1}{2}$, $N = 1000$)

x/π	Scaled problem (no acceleration)	$m = 4$	Exact solution
0.1	0.2032 ± 0.020	0.1892 ± 0.0009	0.1898
0.2	0.3563 ± 0.019	0.3610 ± 0.0011	0.3610
0.3	0.4948 ± 0.019	0.4961 ± 0.0009	0.4969
0.4	0.5922 ± 0.018	0.5851 ± 0.0013	0.5841
0.5	0.6109 ± 0.018	0.6141 ± 0.0007	0.6142

TABLE III

Hermite Acceleration with Scaling
($t = 0.049382$, $\alpha = \frac{1}{2}$, $N = 10000$)

x/π	Scaled problem (no acceleration)	$m = 4$	Exact solution
0.1	0.1906 ± 0.0064	0.1893 ± 0.0003	0.1898
0.2	0.3713 ± 0.0063	0.3610 ± 0.0003	0.3610
0.3	0.4906 ± 0.0061	0.4969 ± 0.0002	0.4969
0.4	0.5841 ± 0.0060	0.5841 ± 0.0002	0.5841
0.5	0.6204 ± 0.0059	0.61421 ± 0.0001	0.61422

Tables II and III exhibit the result of a computation with $\alpha = \frac{1}{2}$ on the example of the preceding section. One can see that scaling of the unaccelerated computation is harmful, while its effect on the acceleration is mildly beneficial. Table III is

similar to Table II except for the increase in N. The increase in accuracy through acceleration is of order 30, which would have required a 900-fold increase in computational effort were it to be achieved by an increase in N. The SD's listed in the third column of Table III are as small as can be achieved without special precautions to limit the effect of round-off error. Qualitatively similar results were obtained in a number of examples, and in particular for all values of t.

Expansion of the Estimand; Solution of an Integral Equation

A Monte–Carlo computation typically yields estimates of a finite set of numbers. In the preceding sections it was shown how the expansion of the estimator can make these estimates more accurate. If the solution of the problem is a function, the question arises, which finite set of numbers should be evaluated to yield the most information about that function and to take advantage of the fact that Monte–Carlo computations are not tied to a finite grid. A reasonable answer can often be obtained by expending the estimated function in Hermite polynomials. For example, if the object of the calculations is to obtain a Green's function, one can readily construct an estimator for the n-th coefficient of its Hermite expansion, to which the acceleration technique can then be applied.

In kinetic theory one encounters problems whose solutions $f(x_1, x_2, \ldots)$ are close to Gaussian [a function $f(x)$ is said to be close to Gaussian if, for some β, $f(x)\, e^{\beta x^2}$ has a rapidly converging Hermite expansion], are functions of 2 or 3 variables, and can be expressed as integrals of higher multiplicity of an integrand which is not close to Gaussian. An appropriate procedure is to evaluate these functions f at the roots of a Hermite polynomial and then construct the Hermite expansions of f by Gauss–Hermite quadrature [5]. As an illustration of this technique, consider the integral equation

$$2 \int p(2u - v)\, p^*(v)\, dv = p(u), \tag{7}$$

with

$$p^*(v) = (\pi R T_0)^{-1/2}\, e^{-v^2/RT_0},\ R,\ T_0 \text{ constants},$$

$$p(u) \geqslant 0,\ \int p(u)\, du = 1.$$

$p(u)$ is the equilibrium velocity distribution in a one-dimensional phase space of particles which collide with target particles of another species. When a particle with velocity u collide with a target particle of velocity v, it acquires the velocity

$$u' = u - \theta(u - v),\ \theta = \tfrac{1}{2}\ (\text{"soft" collision}).$$

The solution of (7) is

$$p(u) = (\pi RT)^{-1/2} e^{-u^2/RT}, \qquad T = \frac{T_0}{3} = \int u^2 p(u)\, du.$$

Introduce an artificial time t and assume that during a time step each particle in phase space undergoes exactly one collision. Let $p_0(u)$ be an initial guess of $p(u)$; after the n-th time-step we have

$$p_n(u) = E[2p_{n-1}(2u - \xi)],$$

where

$$P(y < \xi < y + dy) = (\pi RT_0)^{-1/2} e^{-y^2/RT_0}\, dy.$$

Define $T_{n-1} = \int u^2 p_{n-1}(u)\, du$, and evaluate u at the l points $u_i = \sqrt{RT_{n-1}}\, x_i$, $i = 1,\ldots, l$. x_i the roots of a Hermite polynomial of degree l, using a Monte–Carlo method with acceleration. Assume $p_n(u)$ is in L_2, and expand $p_n(u)\, e^{u^2/2RT_{n-1}}$ in Hermite series:

$$p_n(u)\, e^{u^2/2RT_{n-1}} \cong \sum_{j=0}^{M} d_j H_j(u/\sqrt{RT_{n-1}})(\pi RT_{n-1})^{-1/2} e^{-u^2/2RT_{n-1}},$$

or

$$p_n(u) \cong \sum_{j=0}^{M} d_j H_j(u/\sqrt{RT_{n-1}})(\pi RT_{n-1})^{-1/2} e^{-u^2/RT_{n-1}}, \tag{8}$$

where

$$d_j = \int p_n(u)\, H_j(u/\sqrt{RT_{n-1}})\, du.$$

The integrals which yield d_j are performed by Gaussian quadrature. In particular, $\int p_n(u)\, du = 1$ (conservation of mass) implies $d_0 = 1$. Define $\int u^2 p_n(u)\, du = T_n$, and proceed to the next step.

In Table IV the results of such a computation are exhibited. n is the number of steps; $\int p_n(u)\, du = d_0$ is evaluated by Monte–Carlo and quadrature rather than set equal to 1, and affords a check on accuracy. $R = 0.1$, $T_0 = 1$; T_n should, and does, tend to $1/3$. The calculation uses a four-term acceleration without scaling, $(m = 4)$, $l = 7$, $M = 6$, $N = 10000$, $p_0 = (\pi RT_0)^{-1/2} e^{-u^2/\pi RT_0}$. After 5 steps, $|d_j| \leqslant 0.01$ for $j > 1$.

TABLE IV
Solution of an Integral Equation
($N = 10000$, $m = 4$, $l = 7$)

No. of steps	$\int p_n(u)\,du$	$T_n = \int u^2 p_n(u)\,du$
0	1	1
1	0.998	0.491
2	0.999	0.366
3	0.997	0.342
4	1.0009	0.337
5	1.0007	0.334

Several points remain to be made: M, the number of terms in (8), should be determined in the course of computation. Instead of (8) we could write, more generally,

$$p_n(u) \cong \sum_{j=0}^{M} d_j H_j(u/\sqrt{s})(\pi s)^{-1/2} e^{-u^2/s}, \tag{9}$$

where s is a scale factor, to be chosen for maximum computational convenience, and which may very in time (as above) and in multidimensional problems, from direction to direction. If M were predetermined and s fixed, (9) would be closely related to the expansion used by Grad [3].

Most importantly, the present technique can be used in nonlinear problems: Suppose the problem were to solve

$$2\int p(2u - v)\,p(v)\,dv = p(u), \tag{10}$$

i.e., the molecules collided with their own kind. If $p_n(u)$ has the expansion (9), then

$$p_{n+1}(u) = 2\sqrt{s}\,E\left[p_n(2u - \xi)\cdot\sum_{j=1}^{M} d_j H_j(\xi)\right],$$

$$P(y < \xi < y + dy) = \pi^{-1/2} e^{-y^2}\,dy.$$

For applications, see [2]. (10) has, of course, only the trivial solution for which $\int p(u)\,u^2\,du = 0$.

REFERENCES

1. R. H. CAMERON AND W. T. MARTIN, The orthogonal development of nonlinear functionals in series of Fourier–Hermite functionals, *Ann. of Math.* **48** (1967), 385.

2. A. J. CHORIN, Numerical Solution of the Boltzmann Equation, to appear in *Comm. Pure Appl. Math.*
3. H. GRAD, On the kinetic theory of rarefied gases, *Comm. Pure Appl. Math.* **2** (1949), 311.
4. J. M. HAMMERSLEY AND D. C. HANDSCOMB, Monte–Carlo methods, Methuen, London, 1965.
5. S. KACZMARZ AND H. STEINHAUS, Theorie der orthogonelreihen, Warsaw, 1935.
6. A. M. STROUD AND D. SECREST, "Gaussian Quadrature Formulas," Prentice Hall, Englewood Cliffs, N. J., 1966.
7. R. E. A. C. PALEY AND N. WIENER, "Fourier Transforms in the Complex Domain," Colloquium Publication Vol. 19, American Mathematical Society, Providence, R. I., 1934.
8. N. WIENER, "Nonlinear Problems in Random Theory," M.I.T. Press, Cambridge, Mass., 1958.