JOURNAL OF COMPUTATIONAL PHYSICS **11**, 146–151 (1973)

Note

An Economical Approximation for the Coefficients in the Development of a Function with Respect to an Orthogonal System

In applying Galerkin's method to partial differential equations, one frequently encounters the task of developing a function f in terms of a system of orthogonal functions g_k. The functions f are usually generated in the course of the computations. There exists some arbitrariness in the choice of the system of functions g_k. Often it is suggested by the operator governing the partial differential equation. The k-th coefficient in such a development is given by an integral, derived from the orthogonality relations, involving f and the function g_k. There are problems in which many such developments are required and then it is desirable to keep the number of values x small for which the function f must be evaluated. In this note we shall discuss a special method for approximating the development coefficients and present numerical data demonstrating its usefulness.

In all methods for the evaluation of the development coefficients the function f is computed at a finite number of points x_i of a certain interval. They will be called pivotal points. Let $f(x_i) = f_i$. We use m pivotal points x_i. (m must be at least as large as the number of functions required for an acceptable approximation of f, but usually larger values of m are used.) One can postulate that, except for rounding errors, the development coefficients are exact if f is given by any linear combination of the first m functions g_k. Such an approximation for f gives an expression which interpolates between the values of f at the pivotal points x_i using the functions g_k, $k = 1,..., m$.

We examine a special form of this approach suggested by problems in which the g_k are given by orthogonal polynomials—for instance, Legendre polynomials. (The example of Chebycheff polynomials may be misleading because of its simplicity.) One can start from the integral representation for the development coefficients—to be denoted by c_k—, and approximate these integrals by Gaussian quadrature based on the system of functions g_k. In this case the pivotal points for the integration are given by the zeros of g_{m+1}. Because of the properties of Gaussian integration this result is exact if the function f is a linear combination of the first m functions g_k. For a general choice of the g_k's a similar result is obtained by the following procedure. Let us denote by x_i the pivotal points of the interpolation scheme, not necessarily the zeros of g_{m+1}, by f^* a vector of dimension m whose i-th entry is given by f_i, by G a matrix of dimension m by infinity, where the element G_{ik} is given by $g_k(x_i)$ and by c a vector of dimension infinity, whose k-th entry is given by c_k. If f is given by a development

$$f(x) = \sum_{k=1}^{\infty} g_k(x)\, c_k\,,$$

then $f^* = Gc$.

Let $G^{(1)}$ be the matrix consisting of the first m columns of G, and let $\tilde{c}$ be the vector of dimension m, whose k-th entry is the approximation $\tilde{c}_k$ to the development coefficient c_k. The function $f(x)$ is now approximated by

$$\sum_{k=1}^{m} g_k(x)\, \tilde{c}_k\,.$$

The requirement that this approximation reproduces f exactly at the points x_i, $i = 1,..., m$ leads to the equation $G^{(1)}\tilde{c} = Gc$. Hence $\tilde{c} = Bc$, where

$$B = (G^{(1)})^{-1}G. \tag{1}$$

Independent of the choice of the pivotal points, B can be partitioned as follows:

$$B = [I_m \mid B^{(2)}], \tag{2}$$

where I_m is the identity matrix of dimension m. The elements of $B^{(2)}$ will retain the subscripts which they have as members of B. If the pivotal points x_i are the zeros of g_{m+1}, then it follows from the special properties of Gaussian quadrature that

$$B_{ij}^{(2)} = 0, \qquad 1 < i \leqslant m, m + 1 \leqslant j \leqslant 2m + 1 - i \tag{3}$$

and

$$\tilde{c}_k = c_k + \sum_{2m+2-k}^{\infty} B_{kl}^{(2)} c_l\,, \qquad k = 1,..., m. \tag{4}$$

In the method of interpolation described above, $\tilde{c}_k$ is used as an approximation to c_k, $k = 1,..., m$. The sum in (4) gives the error that occurs in this approximation. The special advantage of Gaussian quadrature is expressed by the lower limit of this sum. Without the special choice of the pivotal points the lower limit would be $m + 1$. Coefficients c_l for which l is smaller than this lower limit cannot falsify c_k. The elements of $B^{(2)}$ characterize the size of error in the approximations for the c_k's, provided that information is available about the behavior of these coefficients for larger values of k. The idea to approximate the coefficients in this manner is contained in a paper by Luke [1]. Luke's concern is primarily with the analytical aspects, in particular with means of computing the elements of $B^{(2)}$ from the recurrence relation for the orthogonal polynomials. The numerical procedure used here yields the same information. One should not expect that all of the elements of $B^{(2)}$ will be small.

If one uses approximations for the integral representations of the coefficients c_k, as given by sophisticated integration formula, then all values of c_l must be expected to contribute to the error terms. It is difficult to compare such a method with the present approach, for the characterization of the errors is different. Most integration formulae give error estimates in terms of some higher derivative of the integrand (here a product of f, g_k and the weight function), while at present we need information about the behavior of the c_l for large values of l. In many applications this information is not available for either kind of approximation. Some indication of the behavior of the c_l's can be obtained by inspection of the $\tilde{c}_l$'s, $l = 1,..., j < m$ as obtained by the computation. The representation of f by a truncated development in orthogonal functions is practical only if, from a certain l on, the values of c_l decrease at a significant rate. But then the present method of evaluating the coefficients will give good results. For poorly converging developments not only the method of determining the coefficients but also the form of the approximation for f is suspect.

We ask now to what extent the properties (2) and (3) of the matrix B will be carried over to a more general system of functions g_k—for instance, to a system suggested by the operator occurring in a partial differential equation. A problem of this kind is treated in [2], although with a different method of evaluating the coefficients. An approximation of the function f by an interpolation procedure involving the g_k guarantees that the decomposition of the matrix B shown in (2) is still valid. If one chooses the zeros of g_{m+1} as pivotal points, then one obtains, as before

$$B_{k,m+1} = 0, \qquad k = 1,..., m.$$

For orthogonal polynomials a number of other elements of B are zero [see Eq. (3)]. In general cases one cannot expect that the corresponding elements are also zero. A conjecture that these elements might be rather small is suggested by the observation that Chebycheff polynomials are closely related to trigonometric polynomials and that for trigonometric polynomials (3) holds again. The solutions of many second order Sturm–Liouville equations are asymptotically represented by trigonometric functions, and this fact is likely to be reflected in the properties of B. We have carried out numerical experiments for Bessel functions of different order. Bessel functions of order 1/2 are expressible by trigonometric functions; but as the order is increased, the applicability of asymptotic representations in terms of trigonometric functions is restricted to larger arguments; therefore, Bessel functions of higher order give a more severe test to our conjecture than the usual examples of Sturm–Liouville equations.

TABLE I

The first 12 Columns of $B^{(2)}$ for $m = 10$, $\nu = 5$ and Pivotal Points $x_i = \lambda_i/\lambda_{m+1}$.

-.146E-08	.933E-06	-.429E-05	.122E-04	-.302E-04	.734E-04	-.188E-03	.536E-03	-.189E-02	.102E-01	-.240E+00	.722E+00
-.673E-09	-.154E-05	.720E-05	-.212E-04	.554E-04	-.147E-03	.431E-03	-.158E-02	.934E-02	-.417E+00	.811E+00	-.972E-01
.214E-08	.241E-05	-.116E-04	.360E-04	-.103E-03	.315E-03	-.119E-02	.755E-02	-.563E+00	.733E+00	.239E+00	-.185E+00
.303E-09	-.381E-05	.192E-04	-.647E-04	.214E-03	-.851E-03	.570E-02	-.682E+00	.601E+00	.362E+00	.769E-01	-.162E+00
-.107E-08	.330E-05	-.340E-04	.132E-03	-.568E-03	.405E-02	-.776E+00	.461E+00	.365E+00	.210E+00	.351E-01	-.128E+00
-.409E-09	-.113E-04	.537E-04	-.344E-03	.267E-02	-.850E+00	.332E+00	.311E+00	.243E+00	.145E+00	.248E-01	-.953E-01
-.957E-09	.229E-04	-.176E-03	.158E-02	-.907E+00	.222E+00	.234E+00	.217E+00	.176E+00	.109E+00	.237E-01	-.672E-01
.112E-08	-.586E-04	.778E-03	-.949E+00	.133E+00	.155E+00	.162E+00	.154E+00	.127E+00	.821E-01	.226E-01	-.444E-01
-.271E-08	.247E-03	-.977E+00	.661E-01	.857E-01	.987E-01	.105E+00	.100E+00	.845E-01	.566E-01	.183E-01	-.264E-01
-.958E-09	-.994E+00	.220E-01	.329E-01	.419E-01	.488E-01	.519E-01	.502E-01	.427E-01	.293E-01	.103E-01	-.122E-01

TABLE II

The First 12 Columns of $B^{(2)}$ for $m = 10$, $\nu = 5$ and Pivotal Points $x_i = i/(m + 1)$.

-.103E-02	-.141E-02	-.337E-02	-.199E-02	-.858E-02	.311E-01	-.258E+00	.666E+00	-.696E+00	-.998E-01	.769E+00	-.614E+00
.187E-02	.248E-02	.693E-02	.350E-02	.639E-01	-.372E+00	.749E+00	-.211E+00	-.481E+00	-.387E-01	.478E+00	.382E+00
-.330E-02	-.328E-02	-.106E-01	.598E-01	-.484E+00	.692E+00	.278E-01	-.397E+00	-.346E+00	-.364E-01	.284E+00	.382E+00
.696E-02	.867E-02	.990E-01	-.509E+00	.684E+00	.219E+00	-.197E+00	-.302E+00	-.181E+00	.486E-01	.265E+00	.363E+00
-.130E-01	.671E-01	-.613E+00	.554E+00	.196E+00	-.198E+00	-.375E+00	-.369E+00	-.243E+00	-.720E-01	.827E-01	.171E+00
.113E+00	-.573E+00	.614E+00	.365E+00	.334E-01	-.143E+00	-.177E+00	-.107E+00	.153E-01	.149E+00	.261E+00	.322E+00
-.101E+01	-.239E+00	-.883E+00	-.162E+01	-.221E+01	-.262E+01	-.287E+01	-.299E+01	-.301E+01	-.296E+01	-.284E+01	-.268E+01
.340E+01	.627E+01	.932E+01	.124E+02	.153E+02	.179E+02	.200E+02	.216E+02	.226E+02	.230E+02	.228E+02	.220E+02
-.519E+01	-.119E+02	-.190E+02	-.258E+02	-.319E+02	-.370E+02	-.411E+02	-.440E+02	-.457E+02	-.463E+02	-.457E+02	-.439E+02
.369E+01	.741E+01	.115E+02	.156E+02	.193E+02	.225E+02	.251E+02	.270E+02	.282E+02	.286E+02	.283E+02	.272E+02

The computation is simply an evaluation of (1). This procedure can always be carried out if $G^{(1)}$ has an inverse. The approximations for the coefficient c_k, $k = 1,..., m$ are then given by

$$\tilde{c} = (G^{(1)})^{-1} f^*.$$

The procedures for computing $(G^{(1)})^{-1} f^*$ and for evaluating the coefficients from the orthogonality relations are very similar, they require about the same amount of work, provided that the number of pivotal points is the same. In our computation,

$$G_{ik} = g_k(x_i) = \sqrt{2}\, J_\nu(\lambda_k x_i)/|\, J_{\nu+1}(\lambda_k)|,$$

where λ_k is the k-th zero of the Bessel function J_ν. The computations have been carried out for $m = 10$ and $m = 15$, $\nu = 1, 2,..., 5$. The matrix elements of B are analytic functions of ν; fractional ν would give similar results. The pivotal points were chosen in accordance with our conjecture as the zeros of g_{m+1}, that is, $x_i = \lambda_i/\lambda_{m+1}$; alternately, we have used evenly spaced points

$$x_i = i/(m + 1).$$

Table I shows that first 12 columns of the matrix $B^{(2)}$ for $m = 10$, $\nu = 5$ and $x_i = \lambda_i/\lambda_{m+1}$. Elements of B which would be zero for orthogonal systems are very small. For smaller values of ν the results would be even more favorable. Our computations showed that the elements decrease somewhat with m. For orthogonal polynomials the first terms occurring in the sums of (4) have coefficients of order 1. The same behavior is observed here. Table II shows the first 12 columns of the matrix $B^{(2)}$ for the alternate set of pivotal points. A comparison of Tables I and II shows that the original choice is indeed advantageous.

The matrix $G^{(1)}$ is well conditioned; its inversion can be done routinely. If the matrix $G^{(1)}$ is large (that is, for large m), it may be practical to use an iterative procedure which starts with an approximate inverse obtained by inspection from the orthogonality relations. Let H be an approximation to $(G^{(1)})^{-1}$. In the present example we have chosen

$$H = (1/(m + 1))(G^{(1)})^T E,$$

where E is a diagonal matrix with $E_{ii} = \lambda_i/\lambda_{m+1}$. The matrix E accounts for the weight factor which is present in the orthogonality relations for Bessel functions. $HG^{(1)}$ is close to a diagonal matrix.

We write

$$HG^{(1)} = D + \Delta, \tag{5}$$

where D is a diagonal matrix and each diagonal entry of Δ is 0. Now one can compute

$$(HG^{(1)})^{-1} = (I - D^{-1}\Delta + (D^{-1}\Delta)^2 - (D^{-1}\Delta)^3 + \cdots)\, D^{-1} \tag{6}$$

and

$$(G^{(1)})^{-1} = (HG^{(1)})^{-1}H. \tag{7}$$

We have carried out this procedure for $m = 15$, $\nu = 4$. It converges rather well; one gains about one decimal digit in each iterative step. Moreover, the convergence can be improved by carrying out only one or two iterative steps in (6), and then by regarding (7) as a new approximation H of $(G^{(1)})^{-1}$. The procedure (5)–(7) is then repeated for this new H.

References

1. Yudell L. Luke, On the error in a certain interpolation formula and in the Gaussian integration formula, *J. Austral. Math. Soc.* to appear.
2. Karl G. Guderley and Chen-Chi Hsu, A special form of Galerkin's method applied to heat transfer in plane Couette–Poiseuille flows, *J. Comput. Phys.* to appear.

Received: April 19, 1972.

K. G. Guderley and R. H. Warren
Aerospace Research Laboratories,
Wright–Patterson Air Force Base, Ohio 45433

Printed in Belgium