NUMERICAL ANALYSIS PROJECT
MANUSCRIPT NA-~~90-15~~ 92-21

DECEMBER 1992

# On the Error Computation for Polynomial Based Iteration Methods

by

B. Fischer
G.H. Golub

NUMERICAL ANALYSIS PROJECT
COMPUTER SCIENCE DEPARTMENT
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

# On the Error Computation for Polynomial Based Iteration Methods

BERND FISCHER * and GENE H. GOLUB **

## Abstract

In this note we investigate the Chebyshev iteration and the conjugate gradient method applied to the system of linear equations $Ax = f$ where $A$ is a symmetric, positive definite matrix. For both methods we present algorithms which approximate during the iteration process the $k$th error $\varepsilon_k = \|x - x_k\|_A$. The algorithms are based on the theory of modified moments and Gaussian quadrature. The proposed schemes are also applicable for other polynomial iteration schemes. Several examples, illustrating the performance of the described methods, axe presented.

## 1. Introduction

Consider the system of linear equations

$$Ax = f, \tag{1.1}$$

where $A$ is a $N$ x N real symmetric, positive definite matrix and $f$ is a given vector. Let $x_0 \in \mathbb{R}^N$ be any initial guess for the solution of (1.1), and let $r_0 := f - Ax_0$, $\varepsilon_0$ := x-xc be the associated residual vector and error vector, respectively. Furthermore, let

$$\mathcal{P}_n := \{q(t) \equiv a_0 + a_1 t + \cdots + a_n t^n \mid a_0, a_1, \ldots, a_n \in \mathbb{R}\}$$

denote the set of all real polynomials of degree at most n.

A widely used class of iterative schemes for the solution of (1.1) are the so-called *polynomial iteration methods.* These methods generate iterates of the form

$$x_n = x_0 + q_{n-1}(A)r_0, \quad \text{where } q_{n-1} \in \mathcal{P}_{n-1}, \quad n = 1, 2, \ldots.$$

* Institute of Applied Mathematics, University of Hamburg, D-2000 Hamburg, F.R.G.

** Department of Computer Science, Stanford University, Stanford, CA 94305, U.S.A. The work of this author was in part supported by the National Science Foundation under Grant NSF CCR-8821078.

The corresponding residual vectors and error vectors are given in terms of the so-called residual polynomial $p_n$

$$\begin{aligned} r_n &= f - Ax_n = p_n(A)r_0, \\ \varepsilon_n &= 2 - x_n = p_n(A)\varepsilon_0, \end{aligned} \qquad \text{where } p_n(t) \equiv 1 - tq_{n-1}(t). \tag{1.2}$$

Notice that $\varepsilon_n = A^{-1}r_n$ and consequently

$$\|\varepsilon_n\|_A^2 := \varepsilon_n^T A\varepsilon_n = r_n^T A^{-1} r_n. \tag{1.3}$$

In this note we are concerned with the approximation of $\|\varepsilon_n\|_A$ as the iteration process goes along.

In §2 we show that the computation of $\|\varepsilon_n\|_A$ is equivalent to the evaluation of a certain integral. This integral will be (approximately) evaluated by means of Gaussian quadrature. The process is closely related to the computation of certain orthogonal polynomials. In §2.1 and §2.2 we show how to approximate $\|\varepsilon_n\|_A$ during the Chebyshev iteration and the conjugate gradient (CG) method, respectively. It turns out that the computation of the initial error $\|\varepsilon_0\|_A$ is special. §2.3 is devoted to this problem. Finally, in §3 we report on some numerical experiments.

## 2. Error computation

In this section we show that the A-norm of the error (1.3) can be expressed in terms of a certain Riemann-Stieltjes integral. To this end we expand the starting residual

$$r_0 = \sum_{k=1}^{N} \sigma_k v_k \tag{2.1}$$

into orthonormal eigenvectors $v_k$ of $A$. We denote by $\lambda_k$ the eigenvalues corresponding to $v_k$ and obtain, in view of (1.3) and (2.1),

$$\begin{aligned} \|\varepsilon_n\|_A^2 = (p_n(A)r_0)^T A^{-1} p_n(A)r_0 &= \sum_{k=1}^{N} \frac{\sigma_k^2}{\lambda_k} p_n^2(\lambda_k) \\ &= \int p_n^2(t)\, d\hat{\sigma}(t). \end{aligned} \tag{2.2}$$

The distribution function $\hat{\sigma}(t)$ in the above defined Riemann - Stieltjes integral is given by

$$\hat{\sigma}(t) \equiv \sum_{k=1}^{N} \frac{\sigma_k^2}{\lambda_k} H(t - \lambda_k), \quad \text{where } H(t) \equiv \begin{cases} 1 \text{ for } t \geq 0, \\ 0 \text{ otherwise.} \end{cases} \tag{2.3}$$

The problem of determining the error $\|\varepsilon_n\|_A$ is equivalent to the evaluation of an integral with distribution function $\hat{\sigma}(t)$. A standard tool for computing such an integral is the Gaussian quadrature . Let us assume (for a moment) that we know the orthonormal (*) polynomials $\hat{\psi}_k$, $k = 1, 2, \ldots, n$, relative to $\hat{\sigma}(t)$

$$\int \hat{\psi}_k(t)\hat{\psi}_m(t)d\hat{\sigma}(t) \begin{cases} = 1 \text{ if } k = m \\ = 0 \text{ if } k \neq m \end{cases}$$

in terms of their three-term recurrence coefficients $\hat{\alpha}_k$ and $\hat{\beta}_k$

$$t\hat{\psi}_k(t) = \hat{\beta}_k\hat{\psi}_{k+1}(t) + \hat{\alpha}_k\hat{\psi}_k(t) + \hat{\beta}_{k-1}\hat{\psi}_{k-1}(t).$$

We associate with $\hat{\sigma}(t)$ the tridiagonal matrix

$$\hat{J}_n = \begin{pmatrix} \hat{\alpha}_1 & \hat{\beta}_1 & & & \\ \hat{\beta}_1 & \hat{\alpha}_2 & \hat{\beta}_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \hat{\beta}_{n-1} \\ & & & \hat{\beta}_{n-1} & \hat{\alpha}_n \end{pmatrix} \tag{2.4}$$

Then, it is well known (see, e.g., Wilf [12]) that the Gaussian quadrature applied to (2.2) reads

$$\int p_n^2(t)d\hat{\sigma}(t) = \hat{\nu}_0 e_1^T p_n(\hat{J}_n)p_n(\hat{J}_n)e_1,$$

where $e_1 = (1, 0, \ldots, 0)^T$ denotes the first unit vector and

$$\hat{\nu}_0 = \int d\hat{\sigma}(t) = \|\varepsilon_0\|_A^2 \tag{2.5}$$

the zero-order moment, which turns out to be the square of the initial error. Hence, the evaluation of $\|\varepsilon_n\|_A$ "reduces" to the computation of the so-called Jacobi matrix $\hat{J}_n$ and the moment $\hat{\nu}_0$.

The computation of the orthogonal polynomials relative to $\hat{\sigma}(t)$ seems to be on the first glance a quite tricky problem. However, as we will see, it is not hard to compute the orthogonal polynomials

$$\psi_{k+1}(x) = (x - \alpha_k)\psi_k(x) - \beta_k\psi_{k-1}(x), \quad k = 0, 1, \ldots, \tag{2.6}$$

with respect to the distribution function (compare (2.3))

$$\sigma(t) \equiv \sum_{k=1}^{N} \sigma_k^2 H(t - \lambda_k). \tag{2.7}$$

(*) the assumption of orthonormality is more convenient but not necessary.

Again (cf. (3.4)). we associate with a(t) a Jacobi matrix $J_n$. For the CG-method the $\psi_k(t) \equiv p_k(t)$ are just the (explicitly known) residual polynomials (see 52.2). For general polynomial iteration methods the first $n$ polynomials $\psi_k$, $k = 1, 2, \ldots, n$, can be calculated via a modified moment *algorithm* (see, e.g., Sack, Donovan [10] and Wheeler [11]). This algorithm requires as input (only) the $2n + 1$ modified moments

$$\nu_j = \int p_j(t) d\sigma(t), \quad j = 0, 1, \ldots 2n. \tag{2.8}$$

These moments, however, are at hand during the iteration process. More precisely, we have by (1.2) and (2.1) (compare Dahlquist, Eisenstat and Golub [3]),

$$\begin{aligned} r_j^T r_0 = r_0^T p_j(A) r_0 &= \sum_{k=1}^{N} \sigma_k^2 p_j(\lambda_k) \\ &= \int p_j(t) d\sigma(t) = \nu_j. \end{aligned} \tag{2.9}$$

Thus after $2n$ steps of the polynomial iteration method we can calculate $\psi_1, \psi_2, \ldots, \psi_n$. Obviously, it would be advantageous to have the nth degree orthogonal polynomial after only $n$ iteration steps. In §2.2 we will see that this is precisely the case for the Chebyshev iteration.

Anyway, after having computed the $\psi_k$, the desired polynomials can be obtained from a "modification algorithm", by noting that, in view of (2.3) and (2.7), essentially $\hat{\sigma}(t) \equiv \sigma(t)/t$ holds. Such algorithms can be found in Gautschi [6] and Fischer and Golub [4]. Here, for the computation of the Jacobi matrix $\hat{J}_n$ (cf. *(2.4))* of order $n$, one needs to know beforehand the zero-order moment $\hat{\nu}_0$ (cf. (2.5)) and the Jacobi matrix $J_n$.

How to economically compute $J_n$ and $\hat{\nu}_0$, respectively, will be explained in the next subsections.

### 2.1 The Chebyshev iteration

The Chebyshev iteration method is a so-called parameter dependent method. The scheme needs estimates, a and *b,* for the smallest and largest eigenvalue of *A,* respectively,

$$\sigma(A) \subset [a, b], \quad \text{where} \quad 0 < a \leq \lambda_{min}(A) \leq X_{,,,}(A) \leq b. \tag{2.10}$$

The actual computation of these bounds will not be discussed in this paper. We refer the reader to Hageman and Young [9], Golub and Kent [7], Calvetti and Reichel [1], and references therein. The residual polynomial (cf. (1.2)) for the Chebyshev iteration

$$p_n(t) = \frac{T_n((a + b - 2t)/(a - b))}{T_n((a + b)/(a - b))} \tag{2.11}$$

is a suitable scaled and translated Chebyshev polynomial of the first kind. As is well-known, the Chebyshev polynomials fulfill the following identities

$$T_{2n-1}(t) = 2\,T_n(t)T_{n-1}(t) - t,$$
$$T_{2n}(t) = 2\,T_n(t) - 1.$$

This leads together with (2.9) and (2.11) to the modified moments (compare Golub and Iient [7] )

$$\nu_{2n-1} = r_n^T r_{n-1} + \frac{c}{T_{2n-1}(c)}(r_n^T r_{n-1} - \nu_1),$$
$$\nu_{2n} = r_n^T r_n + \frac{1}{T_{2n}(c)}(r_n^T r_n - \nu_0), \qquad \text{where } c = \frac{a+b}{a-b}.$$

Hence, after $n$ steps of the Chebyshev iteration, the modified moments $\nu_0, \nu_1, \ldots, \nu_{2n}$ are available. However, the price paid for the computation of the modified moments are 3 inner products per iteration. The same number of inner products is needed for one step of the conjugate gradient method.

In the next section we show (compare Dahlquist, Golub, and Nash [2]) how to compute the error $\|\varepsilon_n\|_A$ during the conjugate gradient iteration.

## 2.2 The conjugate gradient method

For the conjugate gradient method, there is, in contrast to general polynomial iteration methods, no additional work required for the computation of the orthogonal polynomials $\psi_k$. This is based on the equivalence of the CG method and the Lanczos algorithm and the fact that the Lanczos process directly computes the desired three-term recurrence coefficients (see, e.g., Golub and Van Loan [8]).

From the minimization property of the CG method one readily obtains [2]

$$\|\varepsilon_n\|_A^2 = \|\varepsilon_0\|_A^2 - \nu_0 e_1^T J_n^{-1} e_1. \tag{2.12}$$

In other words, $\|\varepsilon_n\|_A^2$ is the error obtained by approximating the initial error $\|\varepsilon_0\|_A^2 = \int d\hat{\sigma}(t)$ by the n-point Gaussian quadrature rule associated with the distribution function $\sigma(t)$.

However, the remaining (and main) problem is the computation of the initial error $\|\varepsilon_0\|_A$. The next section is devoted to this problem.

## 2.3. Computing the initial error

An effective scheme for the evaluation of

$$\|\varepsilon_0\|_A^2 = \hat{\nu}_0 = \int d\&(t)$$

is devised in Gautschi [5]. The scheme is based on the observation that $\hat{\nu}_0$ has a (convergent) continued fraction expansion in terms of the three-term recurrence coefficients of the orthogonal polynomial $\psi_k$ (cf. (2.6)):

$$\lim_{k\to\infty} C_k = -\|\varepsilon_0\|_A^2,$$

where

$$\begin{aligned} C_k &= (1 + \rho_k)C_{k-1} - \rho_k C_{k-2}, \\ \rho_{k+1} &= \frac{1}{1 - \beta_k(1 + \rho_k)/(\alpha_k\alpha_{k-1})} - 1, \end{aligned} \qquad k = 1, 2, \ldots . \tag{2.13}$$

As it is not surprising, the convergence rate of (2.13) depends on the condition number of $A$, which determines the (relative) distance between the singular point zero of (2.5) and the interval of integration.

A different approach for the approximation of the initial errror is provided by applying a Gauß-Radau rule onto (2.5) (for details see [2]). This leads to upper and lower bounds for $\|\varepsilon_n\|_A$ which can be used in conjunction with (2.12). As for the Chebyshev iteration (cf. (2.10)) this scheme requires the knowledge of upper and lower bounds for the spectrum of $A$.

## 3. Examples

In this section we present some numerical examples. The test problems are obtained by discretizing the Poisson equation

$$\mathrm{Au} = \boldsymbol{f}$$

on the unit square $\Omega = \{(x, y) | 0 \leq \mathrm{x}, \mathrm{y} \leq 0\}$ with Dirichlet boundary conditions $u(x, y) = 0$ on $\partial\Omega$. We approximate Au by the standard five-point stencil on a uniform n x n grid. This results in linear system $Ax = \boldsymbol{f}$ of order $N := n^2$, where the right hand side $\boldsymbol{f}$ is chosen such that the vector x = (1, 1, . . . , 1)$^T$ solves the system. In all experiments the initial vector $x_0$ is a random vector with elements uniformly distributed in [-1, 1]. All computations were carried out in MATLAB.

A crucial point for the correct error estimation is the convergence of the continued fraction (cf. (2.13)) to the initial error $\|\varepsilon_0\|_A$. In the following table we list the number $K$ of steps needed for the algorithm (2.13) to converge up to machine precision EPS ($\sim 2.22 * 10^{-16}$), i.e.,

$$\frac{|C_K - C_{K-1}|}{|C_K|} < \mathrm{EPS}, \tag{3.1}$$

In order to set the quantity $K$ into perspective we computed in addition the number $L$ of iteration steps required to reduce the (relative) A-norm of the error to less than $10^{-5}$, i.e.,

$$\frac{\|\varepsilon_L\|_A}{\|\varepsilon_0\|_A} \leq 10^{-5}. \tag{3.2}$$

In all examples described we chose for the Chebyshev iteration method the best possible interval [a = $\lambda_{min}$(A), *b = X,,,(A)]* (cf. (2.10)), in order to keep the issues of interest clear. Cond denotes the condition number of the respective systems.

| | | CG | | Cheb | |
|---|---|---|---|---|---|
| N | Cond | K | L | K | L |
| 400 | 178 | 62 | *62* | 64 | 123 |
| 900 | 388 | 91 | 89 | 89 | 187 |
| 2500 | 1053 | 141 | 145 | 142 | 309 |

**Table** 3.1. Steps needed for the continued fraction algorithm (2.13) to converge to the initial error $\|\varepsilon_0\|_A$

The convergence of algorithm (2.13) is quite rapidly. The next figure shows the typical convergence behaviour. Here, we ploted subsequent approximations $C_k$, $k = 1, 2, \ldots, 10$ (little circles) and the initial error $\|\varepsilon_0\|_A$ (dotted line) for the CG method and $N$ = 2500.

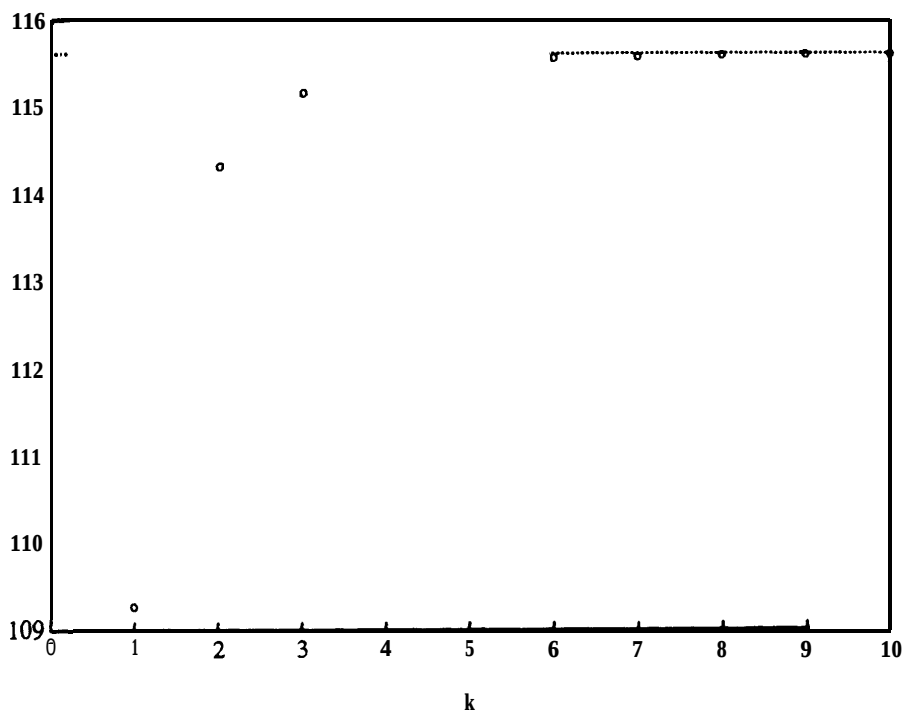


**Figure** 3.1. Approximation of $\|\varepsilon_0\|_A$ by the continued fractions $C_k$

Furthermore, it should be mentioned that in all our experiences the algorithm (2.13) turned out to be very stable.

In the nest table we compare the estimated error $\|\varepsilon_k^{est}\|_A$ with the true error $\|\varepsilon_k\|_A$ for the various examples and methods. We list the largest observed error (compare *(3.2)*)

$$\max_{k=1,2,\ldots,L} \left| \frac{\|\varepsilon_k^{est}\|_A - \|\varepsilon_k\|_A}{\|\varepsilon_0\|_A} \right|$$

In order to demonstrate the performance of the proposed schemes we first computed the initial error $\|\varepsilon_0\|_A$ up to machine precision (cf. (3.1)) and then used this value for the subsequent computations.

| N | CG | Cheb |
|---|---|---|
| 400 | $1.21 * 10^{-8}$ | $1.04 * 10^{-8}$ |
| 900 | $1.20 * 10^{-8}$ | $1.20 * 10^{-6}$ |
| 2500 | $1.15 * 10^{-8}$ | $2.47 * 10^{-5}$ |

**Table 3.2.** Maximal deviation of the estimated error from the true error

It is apparent that the error estimates for the Chebyshev iteration become worse with increasing order. This is due to instabilities in the modified moment algorithm. Note, that the computations for the case $N$ = 2500 involve orthogonal polynomials of degree 309. However, in practice the Chebyshev iteration is implemented as an adaptive scheme, which usually implies a restart after a moderate number of iterations. It should be mentioned that the error-estimation for the CG scheme may also suffer from instabilities. This is mainly due to the loss of orthogonality in the CG process and the possibility of cancelation in (2.12).

Finally, the upper bounds (dashed curve) and lower bounds (dotted curve) for $\|\varepsilon_k\|_A$ by Dahlquist, Golub and Nash [2] and $\|\varepsilon_k\|_A$ (continuous curve) are graphed in the next figure for the CG method and $N$ = 2500. Again, we used the best possible bounds $\lambda_{min}(A)$ and $\lambda_{max}(A)$ for the spectrum of $A$. Note that, according to Table 3.2., the curve ( $\|\varepsilon_k^{est}\|_A$ ) based on the use of continued fractions would not be distinguishable (for this scaling) from the continuous curve.

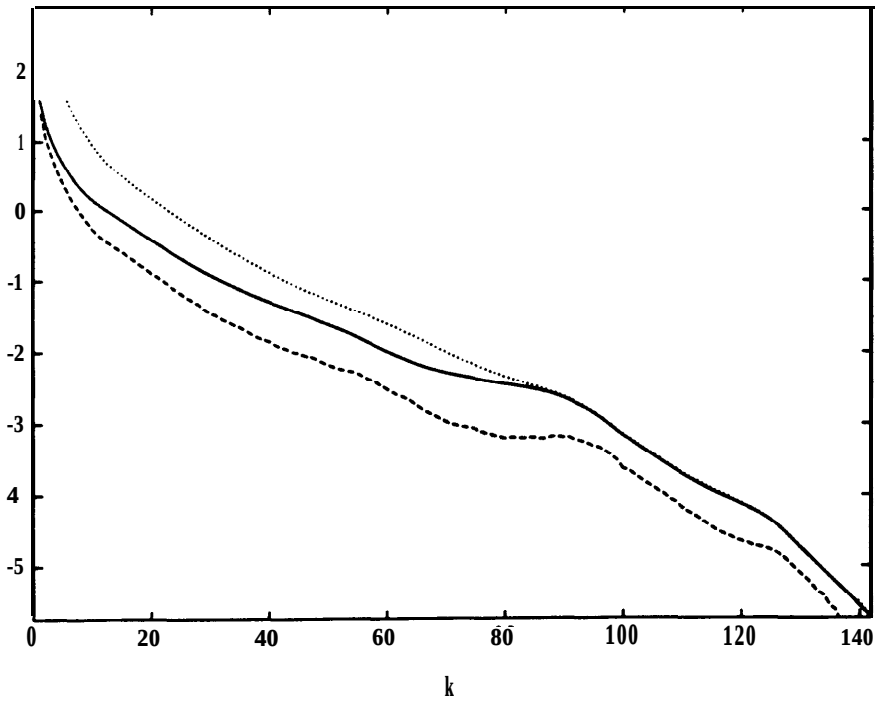


**Figure** 3.2. Estimates of $\|\varepsilon_k\|_A$


## Acknowledgement

The first author would like to thank Roland Freund and Henk van der Vorst for several helpful discussions.


## References


[1] D. **CALVETTI AND** L. **REICHEL,** *Adaptive Richardson iteration based on Leja points,* Tech. Rep. ICM-9210-42, Institute for Computational Mathematics, Kent State University, Kent, OH 44242, Oct. 1992.

[2] *G.* **DAHLQUIST,** G. H. **GOLUB, AND** *S.* **NASH,** *Bounds for the error in linear systems,* in Proceedings of the Workshop on Semi-Infinite Programming, R. Hettich, ed., Springer, 1978, pp. 154-172.

[3] G. G. **DAHLQUIST,** S. C. **EISENSTAT, AND** G. H. **GOLUB,** *Bounds for the error of linear systems of equations using the theory of* modified moments, J. Math. Anal. Appl., 37 (1972), pp. 151-166.

[4] B. **FISCHER AND** G. H. **GOLUB,** *How to generate unknown orthogonal polynomials out of known orthogonal polynomials,* Preprint 45, Institute of Applied

Mathematics, University of Hamburg, November 1991. To appear in the J. CAM.

[5] **W. GAUTSCHI,** *Computational aspects of three-term recurrence relations,* SIAM Rev., 9 (1967), pp. 24–82.

[6] ——, *An algorithmic implementation of the generalized Christoffel theorem,* in Numerical Integration, G. Hammerlin, ed., Basel, 1982, Birkhäuser, pp. 89-106. Internat. Ser. Numer. Math., v. 57.

[7] G. H. **GOLUB AND** M. D. **KENT,** *Estimates of eigenvalues for iterative methods,* Math. Comp., 53 (89), pp. 619-626.

[8] G. H. **GOLUB AND** *C.* F. **VAN LOAN,** *Matrix Computations,* The Johns Hopkins University Press, Baltimore, second ed., 1989.

[9] L. A. HAGEMAN AND D. M. **YOUNG,** *Applied iterative methods,* Academic Press, New York, 1981.

[10] R. A. **SACK AND** A. F. **DONOVAN,** *An algorithm for* Gaussian *quadrature given modified moments,* Numer. Math., 18 (1971/72), pp. 465478.

[11] J. C. **WHEELER,** *Modified moments and Gaussian quadrature,* Rocky Mt. J. Math., 4 (1974), pp. 287-296.

[12] H. **WILF,** *Mathematics for the Physical Sciences,* Wiley, New York, 1962.