MATHEMATICS OF COMPUTATION, VOLUME 33, NUMBER 146
APRIL 1979, PAGES 694–704

On Stable Calculation of Linear Functionals

By Sven-Åke Gustafson

Abstract. In this paper we discuss the recurrent task of evaluating a linear functional defined by (generally infinitely many) linear constraints. We develop a theory for the stability of this problem and suggest a regularization procedure, based on orthogonal expansions. Simple and efficient computational schemes for evaluating the functional numerically are given.

1. Calculation of Linear Functionals from Moment Conditions. We start by discussing the introductory

Example. $b \in C[0, 1]$, i.e. b is a continuous function over $[0, 1]$. Introduce the maximum norm on $C[0, 1]$. Let L be a bounded linear functional. Consider the problem

(1) Compute $L(b)$ when

(2) $$L(a_r) = c_r, \quad r = 1, 2, \ldots,$$

where $a_r(t) = t^{r-1}$.

As a particular instance of the problem (1) and (2) we take

(3) $$L(b) = \int_0^1 b(t)\ln(t^{-1})\,dt.$$

Then

$$c_r = L(a_r) = r^{-2}, \quad r = 1, 2, \ldots .$$

We shall call (2) *moment conditions.*

LEMMA 1. *Use the same notations and assumptions as in Example* 1. *Then* $L(b)$ *is uniquely determined by the sequence* $c_r = L(a_r)$, $r = 1, 2, \ldots$.

Proof. Let b_n be the polynomial of degree less than n which approximates b best in the maximum norm. b_n is uniquely determined and $\|b - b_n\| \longrightarrow 0$ when $n \longrightarrow \infty$. Hence $L(b) = \lim_{n\to\infty} L(b_n)$, and the conclusion follows. Q.E.D.

However, by practical calculations c_r are known only with a finite accuracy and only finitely many of the conditions (2) may be taken into account. Hence, a certain error is introduced in the calculated value of $L(b)$ which is determined by approximating b (directly or indirectly) with linear combinations of $a_1, a_2, \ldots, a_n$. The purpose of this paper is to extend and generalize the results in [6] and [7] as well as to de-

Received March 23, 1978; revised July 14, 1978

AMS (MOS) subject classifications (1970). Primary 65J05, 65F25; Secondary 65B10, 65D30.

Key words and phrases. Linear functionals, moment condition, linear space, error bound, duality lemma.

© 1979 American Mathematical Society
0025-5718/79/0000-0064/$03.75

scribe efficient computational schemes for evaluating $L(b)$ and assessing the associated error.

Definition 1. Let S be a fixed set, $F(S)$ the linear space of real-valued functions, defined on S. Thus, if $f, f_1 \in S$ and α is a real number, we define as usual $f + f_1$ and $\alpha \cdot f$ through

$$(f + f_1)(s) = f(s) + f_1(s); \quad (\alpha f)(s) = \alpha f(s), \qquad s \in S.$$

If $a_r \in F(s)$, $r = 1, 2, \ldots, n$, we denote by a the vector-valued function whose value $a(s)$, $s \in S$, is given by the column vector with components $a_r(s)$, $r = 1, 2, \ldots, n$. If L is a linear functional defined on $F(S)$, we write $L(a)$ for the column vector with components $L(a_r)$, $r = 1, 2, \ldots, n$. Thus, $L(a) \in R^n$, the n-dimensional Euclidean space.

LEMMA 2. *Let $a_1, a_2, \ldots, a_n$, b and L be as in Definition* 1 *and let b be a linear combination of $a_1, a_2, \ldots, a_n$,*

$$b(s) = y^T a(s), \qquad s \in S, \tag{4}$$

(*where $y \in R^n$ and superscript ^{T} as usual denotes transposition*). *Put*

$$L(a) = c. \tag{5}$$

Then

$$L(b) = c^T y.$$

Proof.

$$c^T y = y^T c = y^T(La) = L(y^T a) = L(b). \quad \text{Q.E.D.}$$

Remark. For examples and computational applications of Lemma 2 see [6] and [8]. In the special case when S is a finite set, $S = \{s_1, s_2, \ldots, s_N\}$, the functions a_r may be represented as vectors in R^N. Since $L(a_r)$ then may be represented in the form of scalar products, (5) takes the form of linear systems of equations.

We next establish a more general result.

THEOREM 1. *S, $a_1, a_2, \ldots, a_n$ and L are as in Definition* 1. *$b \in F(S)$ and $y \in R^n$. Introduce ϵ in $F(S)$ through*

$$y^T a = b + \epsilon. \tag{6}$$

Let further c, δ in R^n satisfy

$$L(a) = c + \delta. \tag{7}$$

Then

$$c^T y - L(b) = L(\epsilon) - \delta^T y. \tag{8}$$

Proof.

$$c^T y - L(b) = y^T c - L(y^T a - \epsilon) = y^T(c - L(a)) + L(\epsilon) = L(\epsilon) - \delta^T y. \quad \text{Q.E.D.}$$

The rest of the paper will be based on Theorem 1. We demonstrate first how it can be used to derive bounds on $L(b)$, provided ϵ and L meet certain further conditions.

Definition 2. L is called a *nonnegative* linear functional ($L \geqslant 0$), if $f(s) \geqslant 0$, $s \in S$, implies $L(f) \geqslant 0$.

COROLLARY 1. *Use the same notations as in Theorem* 1 *and require further that* $\delta = 0$ *and* $L \geqslant 0$. *Then*

$$y^{1T}a(s) \leqslant b(s) \leqslant y^{2T}a(s), \qquad s \in S,$$

implies

$$y^{1T}c \leqslant L(b) \leqslant y^{2T}c. \tag{9}$$

Further put $v_1 = \sup L(b)$ *over all* $L \geqslant 0$ *such that* $L(a) = c$ *and* $v_2 = \inf c^Ty$ *over all* $y \in R^n$ *such that* $y^Ta(s) \geqslant b(s)$, $s \in S$, *then*

$$v_1 \leqslant v_2. \tag{10}$$

Proof. With $\delta = 0$, (8) entails $L(b) = c^Ty - L(\epsilon)$. Since $L \geqslant 0$, $L(b) \leqslant c^Ty$, if $y^Ta(s) \geqslant b(s)$, $s \in S$, while $y^Ta(s) \leqslant b(s)$ implies $L(b) \geqslant c^Ty$. Thus (9) and (10) follow. Q.E.D.

Remark. (9) may be used to find bounds for $L(b)$. See [7]. (10) is a slight generalization of the duality lemma of semi-infinite programming, [4, Chapter II] and [10]. If S is a finite set, then (10) coincides with the duality inequality of linear programming. The conditions of Corollary 1 can generally not be met in practice, since the exact values of c_r cannot be represented in the computer. Thus, we arrive at the more general result.

COROLLARY 2. *Use the same notations as in Theorem* 1. *Put* $b_n = y^Ta$ *and let the linear functional* L *satisfy* $La = c$. *Let further* $u > 0$ *be a given* (*small*) *number and let* $\bar{c}_r$ *be such that*

$$|\bar{c}_r - c_r| \leqslant u|c_r|, \qquad r = 1, 2, \ldots,$$

then

$$|L(b_n) - y^T\bar{c}| \leqslant u \sum_{r=1}^{n} |y_rc_r| \leqslant u \max_{1 \leqslant r \leqslant n} |c_r| \sum_{r=1}^{n} |y_r|. \tag{11}$$

Proof. Put $\epsilon = 0$ and $|\delta_r| \leqslant u|c_r|$ in (8). Q.E.D.

Definition 3. Use the notations of Corollary 2 and put for $y^Tc \neq 0$,

$$\kappa_n = \sum_{r=1}^{n} |y_rc_r| / |y^Tc|.$$

We shall call κ_n the *condition number* of the problem to evaluate $L(b_n)$ from $L(a) = c$.

Remark. $\kappa_n u$ is an upper bound for the *relative* error in the value of $L(b_n)$ caused by a relative error in the components of c_r, bounded by u. However, we shall be interested in comparing various sequences $\{b_n\}^\infty$, of approximations to a fixed b. When $L(b)$, L given is sought, then it turns out to be simpler to use

(12) $$k_n = \max_{1 \leq r \leq n} |c_r| \sum_{r=1}^{n} |y_r|$$

as a measure of stability. We have namely $\kappa_n |y^T c| \leq k_n$; and hence, $k_n u$ is a bound for the *absolute* error in $L(b_n)$ caused by a relative error in the components c_r, bounded by u. When we want to determine $L(b)$ for a general b, we first approximate b by b_n. Combining (8) with Corollary 2 and setting $\epsilon = b_n - b$, we get

(13) $$|c^T y - L(b)| \leq |L(b_n - b)| + u \cdot k_n.$$

If we select a sequence of functions b_n, $n = 1, 2, \ldots$, such that $L(b_n - b) \longrightarrow 0$ when $n \longrightarrow \infty$, we achieve that the first term of the right-hand side of (13) decreases with n. However, k_n often increases with n and hence there is an optimal value, $n = n_0$, for which the right-hand side assumes its minimum value. n_0 depends on b, the approximating sequence $b_1, b_2, \ldots$ and u. We shall illustrate this fact on some simple but important problems in Section 2.

2. Stability of Some Convergence Acceleration Formulas. In this section we treat the problem: Let L be a bounded linear functional on $C[0, 1]$ when this space is equipped with the maximum norm. Put $c_r = L(a_r)$ for $a_r(s) = s^r$, $r = 0, 1, \ldots$. Compute for complex z

(14) $$F(z) = \sum_{r=0}^{\infty} c_r(-z)^r$$

using $c_0, c_1, \ldots$ as input data. As explained in [9] this problem can be cast into the form: Compute

(15) $$F(z) = L(f(z, \cdot))$$

for

(16) $$L(a_r) = c_r, \qquad r = 0, 1, \ldots, \qquad f(z, s) = (1 + sz)^{-1}.$$

Here (15) furnishes the analytic continuation of the function (14) to all z outside the set defined by z real and $z \leq -1$.

We investigate the stability of the convergence acceleration methods in Sections 2 and 3 of [9] and prove

THEOREM 2. *Let b_n be the polynomial of degree $n - 1$ obtained by developing $(1 + zs)^{-1}$ in a Taylor expansion around $s = t$ and retaining the first n terms. Put $B = |z|(1 + |t|)/|1 + zt|$. Then the condition number k_n of* (13) *has the properties*

k_n/B^n is bounded for $n > 1$, if $B > 1$,
k_n/n is bounded for $n > 1$, if $B = 1$,
k_n is bounded for $n > 1$, if $B < 1$.

Proof.

$$b_n(s) = \sum_{r=0}^{n-1} y_r s^r = (1 + tz)^{-1} \sum_{r=0}^{n-1} \left\{ \frac{z(t - s)}{1 + tz} \right\}^r.$$

Thus

$$\sum_{r=0}^{n-1} |y_r| \leqslant \frac{1}{|1+tz|} \sum_{r=0}^{n-1} B^r;$$

and hence,

$$k_n \leqslant \frac{\max_{0\leqslant r\leqslant n-1} |c_r|}{|1+tz|} (1+B+\cdots+B^{n-1}) \tag{17}$$

establishing the desired result, since c_r is bounded, $r \geqslant 1$. Q.E.D.

Remark. It may be of interest to compare the accuracy of the results reported in Table 2 of [9] with the estimates for k_n obtainable from Theorem 2. Thus $t = 1$, $z = 1$ gives $B = 1$, and by (17) $k_n \leqslant 0.5.n \max_{0\leqslant r\leqslant n-1} |c_r|$ in this case (Euler transformation), $t = 1, z = 10$ gives $B = 20/11$ (generalized Euler transformation), and $t = ½, z = 10$ gives $B = 5/2$. The bounds for k_n are sufficiently large to explain the observed loss in accuracy in our estimate of $F(z)$ for $z = 10$. For each n the bound for k_n is somewhat smaller for $t = 1$ than for $t = ½$. However, for $t = 1$ the optimal n-value is $n = 30$ but for $t = ½$ it is $n = 20$ and the corresponding values of k_n for optimal n turn out to be of the same magnitude.

We next study the stability of the Čebyšev acceleration, which is described in [9, Section 3]. For this purpose we need

LEMMA 3. *Let the function g satisfy* $(-1)^k g^{(k)}(s) > 0$, $k = 0, 1, \ldots$ *for* $s \in [0, 1]$ *and let* $Q(s) = \Sigma_{r=0}^{n-1} y_r s^r$, *interpolate g at* $t_1, t_2, \ldots, t_n$ *where* $0 \leqslant t_1 < t_2 < \cdots < t_n \leqslant 1$. *Then*

$$\sum_{r=0}^{n-1} |y_r| = Q(-1).$$

Further, if also $(-1)^k g^{(k)}(s) > 0$, $k = 0, 1, \ldots, s \in [-1, 1]$, *then* $Q(-1) \leqslant g(-1)$.

Remark. $g(s) = (1 + st)^{-1}$, $t > 0$, and $g(s) = e^{-s}$ are examples of functions satisfying the assumptions in Lemma 3.

Proof. According to Newton's formula with divided differences we can write

$$Q(s) = d_0 + (s - t_1)d_1 + (s - t_1)(s - t_2)d_2 + \cdots + (s - t_1)\cdots(s - t_{n-1})d_{n-1},$$

where $d_0 = g(t_1)$, $d_1 = g(t_1, t_2), \ldots, d_{n-1} = g(t_1, t_2, \ldots, t_n)$. As known, there is a $\xi_k \in (0, 1)$ such that $k!d_k = g^{(k)}(\xi_k)$. Hence $(-1)^k d_k > 0$, $k = 0, 1, \ldots, n - 1$. Rewriting Q in power form, we get an expression

$$Q(s) = \sum_{r=0}^{n-1} y_r s^r \quad \text{with } y_r = (-1)^r h_r,\ h_r > 0.$$

Thus

$$Q(-1) = \sum_{r=0}^{n-1} |y_r|,$$

proving the first assertion. Newton's interpolation formula with remainder gives

$$g(-1) = Q(-1) + \frac{g^{(n)}(\xi)}{n!} \prod_{i=1}^{n} (-1 - t_i),$$

where $\xi \in [-1, 1]$. Since $(-1)^n g^{(n)}(\xi) > 0$ and $t_i \in [0, 1]$, we get $Q(-1) \leqslant g(-1)$. Q.E.D.

LEMMA 4. *Put* $g(t) = (|1 + z| - |z| + t|z|)^{-1}$, $h(t) = (1 + zt)^{-1}$ *where* $\mathrm{Re}(z) > -½$, *and let* $0 \leqslant t_1 < t_2 < \cdots < t_n \leqslant 1$ *be fixed points. Then the divided differences satisfy*

$$(18) \quad |h(t_1, t_2, \ldots, t_k)| \leqslant (-1)^{k-1} g(t_1, t_2, \ldots, t_k), \qquad k = 1, 2, \ldots, n.$$

Further if

$$Q(s) = \sum_{r=0}^{n-1} y_r s^r, \quad R(s) = \sum_{r=0}^{n-1} w_r s^r$$

interpolate g and h respectively at $t_1, t_2, \ldots, t_n$, *then*

$$(19) \qquad \sum_{r=0}^{n-1} |w_r| \leqslant \sum_{r=0}^{n-1} |y_r|.$$

Proof. We have

$$h(t_1, t_2, \ldots, t_k) = (-z)^{k-1} \prod_{i=1}^{k} (1 + zt_i)^{-1},$$

$$(20) \qquad g(t_1, t_2, \ldots, t_k) = (-1)^{k-1} |z|^{k-1} \prod_{i=1}^{k} (|1 + z| - |z| + t_i|z|)^{-1}.$$

For $\mathrm{Re}(z) > -½$, $|1 + z| > |z|$. Hence by (20), $(-1)^{k-1} g(t_1, t_2, \ldots, t_k) > 0$. Since $|z|(|1 + z| - |z| + t_i|z|)^{-1} \geqslant |z(1 + zt_i)^{-1}|$, (18) follows. In order to verify (19) we express Q and R by means of Newton's formula with divided differences and evaluate y_r and w_r. Then the conclusion follows as in Lemma 3. Q.E.D.

We may now prove

THEOREM 3. *Let* $Q_n = \Sigma_{r=0}^{n-1} y_r t^r$ *be the polynomial of degree less than n which interpolates* $(1 + zt)^{-1}$ *at the zeros of* T_n^* *where* T_n^* *is the shifted Čebyšev polynomial of degree n defined by* $T_n^*(x) = T_n(2x - 1)$. *Here* T_n *is the usual Čebyšev polynomial of degree n. See* [2]. *Then*

$$(21) \qquad \sum_{r=1}^{n} |y_r| \leqslant (1 - z)^{-1}, \quad \text{if } z \text{ real and } z \in (0, 1),$$

$$(22) \qquad \sum_{r=1}^{n} |y_r| \leqslant (|1 + z| - 2|z|)^{-1}, \quad \text{if } |z - 1/3| < 2/3.$$

Thus, if z meets the requirements indicated in (21) *or* (22), *then the condition number* k_n *of the Čebyšev-acceleration in Section* 3 *of* [9] *remains bounded for all n.*

Proof. If $z \in [0, 1]$, then $(1 + zt)^{-1}$ meets the assumptions of Lemma 3 and (21) holds. (22) is established by Lemma 4, since if $|z - 1/3| < 2/3$, then $\mathrm{Re}(z) > -1/3$. Thus, we may replace $(1 + zt)^{-1}$ by $(|1 + z| - |z| + t|z|)^{-1}$ and apply Lemma 3 to that function, and the conclusion follows. Q.E.D.

Remark. If z real and $z > 1$, we may still apply Lemma 3 to find $\Sigma_{r=1}^n |y_r| = Q_n(-1)$, where Q_n is defined as in Theorem 3. Using the recurrence relation for T_n^* we establish as in Section 3 of [9] that $Q_n(-1)$ increases exponentially with n. If z is complex but does not meet the condition associated with (22), we get the inequality

$$\sum_{r=1}^{n} |y_r| \geqslant |Q_n(-1)|,$$

which gives lower bounds for the condition number. These are exponentially growing with n, if $\max(\lambda_1, \lambda_2) > 1$ where $\lambda_1 = |2z^{-1} + 1 + 2\sqrt{z^{-2} + z^{-1}}|$, $\lambda_2 = 1/\lambda_1$. Compare [9, Section 3]. Thus Theorems 2 and 3 are consistent with the numerical results reported in [9] which indicated the existence of an optimal n of modest magnitude giving the best estimate of $L(b)$ in (13).

3. Calculation of Linear Functionals by Means of Orthogonal Expansions.

Definition 4. Let $F(S)$ be as in Definition 1 and let $d\alpha$ be a nonnegative measure over S. We denote by $L_\alpha^2(S)$ the space of functions which are square-integrable over S with respect to $d\alpha$ and equipped with the scalar product

$$\langle f, g\rangle = \int_S f(s)g(s)\, d\alpha(s)$$

for f, g in $L_\alpha^2(S)$ and the norm given by $\|f\|_2^2 = \langle f, f\rangle$. The functions $a_1, a_2, \ldots, a_n$ are said to be linearly independent in $L_\alpha^2(S)$, if $\| \Sigma_{r=1}^n z_r a_r \|_2 = 0$ implies $z_1 = z_2 = \cdots = z_n = 0$.

We prove

LEMMA 5. *Let $a_1, a_2, \ldots, a_n$ be an orthonormal system in $L_\alpha^2(S)$ and put*

(23) $$b_n = y^T a \quad \textit{with } y_r = \langle b_n, a_r\rangle.$$

Then

(24) $$\sum_{r=1}^{n} |y_r| \leqslant \sqrt{n}\|b_n\|_2.$$

Proof. Since $a_1, a_2, \ldots, a_n$ are orthonormal, $y^Ty = \|b_n\|_2^2$. Next maximize $\Sigma_{r=1}^n |y_r|$ under the condition $y^Ty = \|b_n\|_2^2$ and (24) follows. Q.E.D.

Thus, k_n increases at most as $\sqrt{n}$, if $a_1, a_2, \ldots, a_n$ meet the conditions of Lemma 5. Consider again the problem:

Estimate $L(b)$ when $L(a_r) = c_r$, $r = 1, 2, \ldots, n$, are given numerically and $a_1, a_2, \ldots, a_n$ are linearly independent in the sense of Definition 4 for a certain measure $d\alpha$. In view of Lemma 5 it may be advantageous first to orthogonalize $a_1, a_2, \ldots, a_n$ in a preliminary step and then approximate $L(b)$ with $L(b_n)$. This type of stabilization of a problem is often advantageous. See [3, Chapter 1]. If this orthogonalization must be carried out numerically for a general system, then the *modified* Gram-Schmidt method [1] should be used in order to secure numerical stability of the transformation.

We describe now how to perform the transformation in the important case when S is a real interval and $a_r(s) = s^{r-1}$. Then we select $d\alpha$ such that the corresponding system $q_0, q_1, \ldots,$ of orthogonal polynomials has a three-term recurrence relation with coefficients u_r, v_r, which are known as analytic expressions,

$$q_0(s) = 1, \quad q_1(s) = s - u_1, \tag{25}$$

$$q_n(s) = (s - u_n)q_{n-1}(s) - v_n q_{n-2}(s), \qquad n = 2, 3, \ldots . \tag{26}$$

(The so-called classical orthogonal families: Jacobi, Hermite, Laguerre, etc., meet these requirements.) Afterwards, we normalize $q_0, q_1, \ldots$ by putting

$$e_r = w_r q_{r-1}, \tag{27}$$

where the constant w_r is determined (analytically) to render $\|e_r\| = 1$. There is a discrete measure $d\alpha_n$,

$$\int_S f(s)\, d\alpha_n(s) = \sum_{i=1}^{n} m_{ni} f(s_{ni}) \tag{28}$$

such that $m_{in} > 0$, $i = 1, 2, \ldots$, and

$$\int_S e_j(s) e_k(s)\, d\alpha_n(s) = \delta_{jk}, \qquad 1 \leqslant j \leqslant k < n. \tag{29}$$

m_{ni} and s_{ni} are weights and abscissae in the n-point Gaussian quadrature rule corresponding to the measure $d\alpha$ and are efficiently calculated using the computer codes in [5]. (29) follows from the relation

$$\delta_{jk} = \int_S e_j(s) e_k(s)\, d\alpha(s) = \sum_{i=1}^{n} m_{ni} e_j(s_{ni}) e_k(s_{ni}) = \int_S e_j(s) e_k(s)\, d\alpha_n(s),$$

which is valid since the n-point Gaussian rule gives an exact result for polynomials o degree $2n - 1$ or less.

We next determine Lq_r which is done by means of the recurrence relations (25), (26). Put $f_{kr}(s) = s^k q_r(s)$ $l_{kr} = L(f_{kr})$. We first calculate the auxiliary entities l_{kr} for $k + r < n$, $k = 0, 1, \ldots, n - 1$. (25), (26) give the relationships

$$l_{k,0} = c_{k+1}, \qquad k = 0, 1, \ldots, \tag{30}$$

$$l_{k,1} = c_{k+2} - u_1 c_{k+1}, \qquad k = 0, 1, \ldots, \tag{31}$$

$$l_{k,n} = l_{k+1,n-1} - u_n l_{k,n-1} - v_n l_{k,n-2}, \qquad n = 2, 3, \ldots . \tag{32}$$

We arrange the numbers $l_{k,n}$ in a triangular array

$$\begin{array}{ccccc}
l_{0,0} & & & & \\
 & l_{0,1} & & & \\
l_{1,0} & & l_{0,2} & & \\
 & l_{1,1} & & \vdots & \\
l_{2,0} & & l_{1,2} & & \\
 & l_{2,1} & \vdots & & \\
l_{3,0} & \vdots & & & \\
\vdots & & & &
\end{array} \tag{33}$$

By (27), $L(e_r) = w_r L(q_{r-1}) = w_r l_{0,r-1}$, $r = 1, 2, \ldots, n$. The desired elements $l_{0,r}$ appear in the top of the columns of the array (33). The numbers l_{kr} are calculated along ascending diagonals, i.e. in the order $l_{0,0}, l_{1,0}, l_{0,1}, l_{2,0}, l_{1,1}, l_{0,2}, \ldots$. Then, it follows from (30), (31), (32) that in order to calculate the elements in a certain ascending diagonal one needs access only to the two preceding diagonals. Hence, it is not necessary to store the entire array (33) simultaneously. In total about n^2 addition/subtractions and about n^2 multiplications are required for determining $L(e)$.

Next we put $b_n = y^T e$, where y is the optimal solution of

$$\min_{z \in R^n} \int_S [b(s) - z^T e(s)]^2 \, d\alpha_n(s). \tag{34}$$

If we take $z^T e$ as the unique polynomial of degree n which interpolates b at the zeros s_{ni} of e_n, then by (28) the integral (34) assumes the value 0; and hence, this $z^T e$ is optimal. But as known, the unique optimal solution of the least-squares problem (34) is given by $z = y$, where

$$y_r = \int_S b(s) e_r(s) \, d\alpha_n(s) = \sum_{i=1}^{n} m_{ni} b(s_{ni}) e_r(s_{ni}).$$

The calculation of y requires about n^2 multiplications and the same number of additions. Now to determine $L(b_n)$ only the computation of a further scalar product is called for.

Remark. If n is increased, $d\alpha_n$ is changed and most of the work including calculating functional values $b(s_{ni})$ must be redone from scratch.

We next discuss the case when S is a bounded interval, and we shall assume that by a linear transformation it has been transformed into the standard interval $[-1, 1]$. Then it is often suggested to take b_n as the polynomial of degree $< n$ which interpolates b at the zeros of the Čebyšev polynomial of degree n, $T_n(x)$. See, e.g., [3], [12] and [14]. Then

$$s_{ni} = \cos \frac{i - \frac{1}{2}}{n} \pi, \quad m_{ni} = \pi/n. \tag{35}$$

If the cost to evaluate $b(s_{in})$ is great in comparison to an arithmetic operation and it is not known which n-value is finally accepted, then one wants to avoid discarding previously calculated functional values. In this situation [11] suggests that one should start with an odd n-value, advance n according to

$$n_{\text{new}} = 2 \cdot n_{\text{old}} + 1 \tag{36}$$

and select s_{ni} as the zeros of U_n, the nth degree Čebyšev polynomial of the second kind. Thus,

$$s_{ni} = \cos \frac{\pi i}{n+1}, \quad m_{ni} = \frac{1}{n+1} \sin^2 \frac{\pi i}{n+1}. \tag{37}$$

From (37) we find that when n is advanced according to (36), all earlier computed functional values are retained.

It is known, that if b may be continued analytically to an ellipse in the complex plane with foci at $+1$ and -1, s_{ni} are given by (35) and b_n is the polynomial of degree $< n$ which interpolates b at s_{ni}, then

$$|b(s) - b_n(s)| \leqslant A\lambda^n, \tag{38}$$

where A and λ are constants and $\lambda < 1$. But if s_{ni} are given by (36), then we get a bound

$$|b(s) - b_n(s)| \leqslant nA_0\lambda^n, \tag{39}$$

where A_0 is another constant and λ has the same value as in (37). See [11]. A special case occurs in the Čebyšev acceleration scheme treated in Section 2 and in [9]. The bounds (38) and (39) are easily established by means of a straightforward application of Cauchy's integral formula.

Let Δ be a bound for the magnitude of the absolute error in the value of c_r. If $q_n(t) = z^T a(t)$ for some $z \in R^n$, then the resulting error in $L(q_n)$ is bounded by $\Delta \sum_{r=1}^n |z_r|$. We may derive simple expressions for this bound for $q_n = T_n$ and $q_n = U_n$ by means of:

LEMMA 6. *Let t_n and u_n denote the sums of the absolute values of the coefficients of T_n and U_n, the nth degree Čebyšev polynomials of first and second kind. Then*

$$t_n = \tfrac{1}{2}[(1 + \sqrt{2})^n + (1 - \sqrt{2})^n],$$

$$u_n = \frac{1}{2\sqrt{2}}[(1 + \sqrt{2})^{n+1} - (1 - \sqrt{2})^{n+1}].$$

Proof. It is well known that

$$T_n(z) = \tfrac{1}{2}\{(z + \sqrt{z^2 - 1})^n + (z - \sqrt{z^2 - 1})^n\},$$

$$U_n(z) = \frac{1}{2\sqrt{z^2 - 1}}\{(z + \sqrt{z^2 - 1})^{n+1} - (z - \sqrt{z^2 - 1})^{n+1}\}.$$

Since $t_n = |T_n(i)|$, $u_n = |U_n(i)|$ the stated result follows immediately. Q.E.D.

Remark. The result for t_n is also given in [13, p. 792, Eq. (14)]. We treat also the case $S = [0, 1]$. Then we use the shifted Čebyšev polynomials T_n^* and U_n^* defined by $T_n^*(x) = T_n(2x - 1)$ and $U_n^*(x) = U_n(2x - 1)$. See e.g. [9]. Lemma 6 is replaced by

LEMMA 7. *Let t_n^* and u_n^* denote the sums of the absolute values of the coefficients of T_n^* and U_n^*. Then*

$$t_n^* = \tfrac{1}{2}[(1 + \sqrt{2})^{2n} + (1 - \sqrt{2})^{2n}],$$

$$u_n^* = \frac{1}{4\sqrt{2}}\{(1 + \sqrt{2})^{2n+2} + (1 - \sqrt{2})^{2n+2}\}.$$

Proof. Use the fact that $T_n^*(x^2) = T_{2n}(x)$ and $2 \times U_n^*(x^2) = U_{2n+1}(x)$. Q.E.D.

Remark. It is apparent from the above that if $b_n = \Sigma_{r=1}^{n} y_r T_{r-1}$, the error in $L(b_n)$ computed with $L(a)$, $a_r(s) = s^{r-1}$ as input data depends on how rapidly y_r decreases with r. We have namely

$$L(b_n) = \sum_{r=1}^{n} y_r L(T_{r-1}),$$

and the errors in the calculated values of $L(T_{r-1})$ are bounded by a quantity of the form $A(1 + \sqrt{2})^r$ where A is a constant.

Also, if the relative error in the calculated value of $L(T_r)$ is greater than 100% we might as well put $L(T_r) = 0$ since this value is consistent with given data and associated error bounds. In this situation, the addition of more moments c_r does not improve upon our estimate of $L(b)$.

Computing Research Group
The Australian National University
P. O. Box 4
Canberra A.C.T. 2600, Australia

1. Å. BJÖRCK, "Solving linear least squares problems by Gram-Schmidt orthogonalization," *BIT*, v. 7, 1967, pp. 1–21.

2. C. W. CLENSHAW, "Chebyshev series for mathematical functions," *Mathematical Tables*, v. 5, National Physical Laboratory, HMSO, London, 1962.

3. G. DAHLQUIST, Å. BJÖRCK & N. ANDERSON, *Numerical Methods*, Prentice-Hall, Englewood Cliffs, N. J., 1974.

4. K. GLASHOFF & S.-Å. GUSTAFSON, *Einführung in die lineare Optimierung*, Wissenschaftliche Buchgesellschaft, Darmstadt, 1978.

5. G. H. GOLUB & J. H. WELSCH, "Calculation of Gauss quadrature rules," *Math. Comp.*, v. 23, 1969, pp. 221–230.

6. S.-Å. GUSTAFSON, "Control and estimation of computational errors in the evaluation of interpolation formulas and quadrature rules," *Math. Comp.*, v. 24, 1970, pp. 847–854.

7. S.-Å. GUSTAFSON, "On computational applications of the theory of moment problems," *Rocky Mountain J. Math.*, v. 4, 1974, pp. 227–240.

8. S.-Å. GUSTAFSON, "Some optimization problems in numerical analysis," *Methods of Operations Research*, v. 25, 1977, pp. 367–379.

9. S.-Å. GUSTAFSON, "Convergence acceleration on a general class of power series," *Computing*, v. 21, 1978, pp. 367–379.

10. S.-Å. GUSTAFSON & K. O. KORTANEK, "Numerical treatment of a class of semi-infinite programming problems," *Naval Res. Logist. Quart.*, v. 20, 1973, pp. 477–504.

11. S.-Å. GUSTAFSON & S. LINDAHL, "Numerical computation of an integral appearing in the Fröman-Fröman phase-integral formula for calculation of quantal matrix elements without the use of wave functions," *J. Computational Phys.*, v. 24, 1977, pp. 81–95.

12. I. MELINDER, "Accurate approximation in weighted maximum norm by interpolation," *J. Approximation Theory*, v. 22, 1978, pp. 33–45.

13. A.C. R. NEWBERY, "Error analysis for polynomial evaluation," *Math. Comp.*, v. 28, 1974, pp. 789–793.

14. M. J. D. POWELL, "On the maximum norm of polynomial approximation defined by interpolation and by least squares criteria," *Comput. J.*, v. 9, 1967, pp. 404–407.

15. T. RIVLIN, *The Chebyshev polynomials*, Wiley, New York, 1974.