Proceedings of the Royal Society of Edinburgh, **80A**, 45–56, 1978

The condition of gram matrices and related problems

J. M. Taylor

Mathematics Division, University of Sussex

(*Communicated by W. Ledermann*)

(MS received 3 June 1975. Revised MS received 21 September 1977. Read 31 October 1977)

Synopsis

It has been known for some time that certain least-squares problems are "ill-conditioned", and that it is therefore difficult to compute an accurate solution. The degree of ill-conditioning depends on the basis chosen for the subspace in which it is desired to find an approximation. This paper characterizes the degree of ill-conditioning, for a general inner-product space, in terms of the basis.

The results are applied to least-squares polynomial approximation. It is shown, for example, that the powers $\{1, z, z^2, \dots\}$ are a universally bad choice of basis. In this case, the condition numbers of the associated matrices of the normal equations grow at least as fast as 4^n, where n is the degree of the approximating polynomial.

Analogous results are given for the problem of finite interpolation, which is closely related to the least-squares problem.

Applications of the results are given to two algorithms—the *Method* of *Moments* for solving linear equations and *Krylov's Method* for computing the characteristic polynomial of a matrix.

1. Least Squares Problems and Gram Matrices

Let H be a (real or complex) inner-product space, with inner product $[.,.]$, and let $M \subset H$ be an n-dimensional subspace. For a given element $x \in H \backslash M$, the least squares approximation problem is to find the unique $u \in M$ such that

$$\|x-u\|^2 = [x-u, x-u] \tag{1.1}$$

is minimized.

If a basis $\{x_1, x_2, \dots, x_n\}$ is chosen for M, then

$$u = a_1x_1 + a_2x_2 + \cdots + a_nx_n, \tag{1.2}$$

where the coefficients $(a_1, a_2, \dots, a_n)$ are the solutions of the system of *normal equations*

$$\sum_{j=1}^{n} [x_j, x_i]a_j = [x, x_i], \quad i = 1, 2, \dots, n. \tag{1.3}$$

The matrix

$$A = ([x_j, x_i]) \tag{1.4}$$

is the *Gram matrix* for this basis, and is Hermitian positive definite.

If we denote by $\sigma(A)$ the spectrum (set of eigenvalues) of the matrix A and the *spectral radius* by

$$r(A) = \max\{|\lambda| : \lambda \in \sigma(A)\},$$

we can define the *P-condition number* of A by

$$P(A) = r(A)r(A^{-1}). \tag{1.5}$$

When $P(A)$ is large, the system of linear equations (1.3) is said to be *ill-conditioned* and it is difficult to obtain accurate solutions. Many authors have reported on this; for a convenient survey paper, see [**1**].

For a least squares problem, $P(A)$ depends on the basis chosen for the subspace M. For example, if an orthonormal basis is chosen, A is the identity matrix and $P(A) = 1$, which represents best conditioning. For another choice of basis, $P(A)$ could be quite large.

In section 2 a lower bound for $P(A)$ is obtained. This is the main result of this paper and characterizes the degree of ill-conditioning in terms of the basis chosen for the subspace M. Applications of this result, to least squares approximation by polynomials, are given in section 3. There are analogous results for the problem of finite interpolation (section 4). As a final application, results are obtained, in section 5, for the frequently occurring class of *moment matrices*.

2. The Basic Inequality

Given an Hermitian positive definite matrix A, we have

$$\lambda_{\min} \boldsymbol{a}^* \boldsymbol{a} \leq \boldsymbol{a}^* A \boldsymbol{a} \leq \lambda_{\max} \boldsymbol{a}^* \boldsymbol{a} \tag{2.1}$$

for any n-tuple $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^T$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of A, respectively; see [**2**, p. 110]. (T denotes transpose and * denotes conjugate transpose.)

Now $r(A) = \lambda_{\max}$ and $r(A^{-1}) = 1/\lambda_{\min}$, therefore, if we choose specific n-tuples in (2.1), it is possible to obtain a lower bound for $P(A)$.

For example, if $\boldsymbol{a} = \boldsymbol{e}_j$ in (2.1), where $\boldsymbol{e}_j$ is the jth column of the identity matrix, then

$$r(A) \geq \boldsymbol{e}_j^* A \boldsymbol{e}_j = [x_j, x_j] = \|x_j\|^2 \tag{2.2}$$

and

$$r(A^{-1}) \geq 1/\|x_j\|^2, \tag{2.3}$$

for $j = 1, 2, \ldots, n$.

Hence we have the quite useful inequality,

$$P(A) \geq \max_{1 \leq i \leq n} \|x_i\|^2 / \min_{1 \leq j \leq n} \|x_j\|^2. \tag{2.4}$$

Another result, also obtained from (2.1), is

Theorem 1. *If N_i denotes the subspace spanned by $\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$ then*

$$r(A^{-1}) \geq 1/\min_{u \in N_i} \|x_i - u\|^2, \tag{2.5}$$

for each $i = 1, 2, \ldots, n$.

Proof. Since A is Hermitian positive definite, so is A^{-1}. If we replace A by A^{-1} in (2.1), we obtain, with $\boldsymbol{a} = \boldsymbol{e}_i$,

$$\boldsymbol{e}_i^* A^{-1}\boldsymbol{e}_i \le r(A^{-1}) \cdot \boldsymbol{e}_i^*\boldsymbol{e}_i,$$

for each $i = 1, 2, \ldots, n$.

Thus, if $\boldsymbol{b}$ is the solution of

$$A\boldsymbol{b} = \boldsymbol{e}_i, \tag{2.6}$$

we have

$$\begin{aligned} r(A^{-1}) &\ge \boldsymbol{b}^* A\boldsymbol{b}, \\ &= \|v\|^2, \end{aligned} \tag{2.7}$$

where $v = b_1x_1 + b_2x_2 + \cdots + b_nx_n$.

From (2.6), we see that v is the unique element in M such that $[v, x_j] = \delta_{ij}$, for $j = 1, 2, \ldots, n$. Therefore, for any $u \in N_i$,

$$[v, x_i - u] = [v, x_i] = 1$$

and, from the Cauchy–Schwarz inequality, $1 \le \|v\|^2 \cdot \|x_i - u\|^2$. Thus $\|v\|^2 \ge 1/\min_{u\in N_i} \|x_i - u\|^2$ and the result follows from (2.7).

COROLLARY. *A lower bound for the P-condition number of the Gram matrix* (1.4) *is*

$$P(A) \ge 1/\min_{u\in N_i} \|\hat{x}_i - u\|^2, \tag{2.8}$$

for any $i = 1, 2, \ldots, n$, *where* $\hat{x}_i = \dfrac{1}{\|x_i\|} x_i$.

Proof. Now $P(A) = r(A)r(A^{-1}) \ge \|x_i\|^2/\min_{u\in N_i} \|x_i - \omega\|^2$, from (2.2) and (2.5). The result (2.8) follows immediately.

When any of the numbers

$$\delta_i = \min_{u\in N_i} \|\hat{x}_i - u\|^2, \quad \text{for } i = 1, 2, \ldots, n, \tag{2.9}$$

are small, we must expect computational difficulties in obtaining a solution of the linear equations (1.3). We may think of the δ_i as providing a measure of the deviation of the basis from linear dependence and so, in a sense, as its condition numbers. Clearly,

$$0 \le \delta_i \le 1, \quad \text{for } i = 1, 2, \ldots, n.$$

(Choose $u = 0$ in (2.9) to obtain the upper bound.) If the x_i's are linearly dependent, and therefore do not form a basis, we have $\delta_i = 0$. We can have $\delta_i = 1$, for all i, when, for example the x_i form an orthonormal system.

There is an alternative way of expressing the numbers (2.9) which is convenient in applications. Let $\{u_1, u_2, \ldots, u_n\}$ be the orthonormal system obtained by applying the Gram–Schmidt process to the basis $\{x_1, x_2, \ldots, x_n\}$. In this case, u_i is a linear combination of the first i elements of the basis and has an expression

$$u_i = k_{ii}x_i + \cdots + k_{i1}x_1, \tag{2.10}$$

for $i = 1, 2, \ldots, n$. The coefficients k_{ij} are easily calculated from the orthonormality conditions; see [2, p. 44]. Hence, from (2.10), we have

$$\min_{u \in N_n} \|x_n - u\|^2 = \min_{v} \left\| \frac{1}{k_{nn}} u_n - v \right\|^2,$$

where the minimum on the right is taken over the span of $\{u_1, u_2, \ldots, u_{n-1}\}$. Since the u's are an orthonormal system, the expression on the right is equal to $1/k_{nn}^2$. Therefore, from (2.5),

$$r(A^{-1}) \geq k_{nn}^2 \tag{2.11}$$

and

$$P(A) \geq \|x_n\|^2 k_{nn}^2. \tag{2.12}$$

The usefulness of the inequality (2.12) can be seen if we consider the problem of least squares polynomial approximation with the powers as a basis. In this case, as can be seen from (2.10), k_{nn} is the leading coefficient of the associated orthonormal polynomial. We now consider the application of (2.12) to this problem.

3. Least Squares Approximation by Polynomials

Let $q(t)$ be a positive, non-decreasing function of bounded variation on the interval $[a, b]$ which is not constant.

If $f(t), g(t)$ are real-valued Lebesgue–Stieltjes integrable functions with respect to $q(t)$ over $[a, b]$, we define the inner product,

$$[f, g] = \int_a^b f(t)g(t)dq(t),$$

in the usual way.

Let M be the $(n+1)$-dimensional subspace of polynomials of degree n or less and choose the powers $\{1, t, \ldots, t^n\}$ as a basis for M. The corresponding Gram matrices are

$$A_n = \left(\int_a^b t^{i+j-2} dq(t)\right), \tag{3.1}$$

for $i, j = 1, 2, \ldots, n$ and $n = 1, 2, \ldots$.

If the Gram–Schmidt process is used to construct an orthonormal system of polynomials, we obtain the classical system of polynomials, $p_n(t)$, corresponding to the distribution $dq(t)$; see [3, p. 25]. These polynomials satisfy a three-term recurrence relation of the form

$$p_n(t) = (a_n t + b_n)p_{n-1}(t) - c_n p_{n-2}(t), \quad n = 2, 3, \ldots; \tag{3.2}$$

see [3, p. 43].

If the highest coefficient of $p_n(t)$ is denoted by k_{nn}, then

$$a_n = k_{nn}/k_{n-1,n-1}. \tag{3.3}$$

Let

$$\mu_n = \int_a^b t^n dq(t), \quad n = 0, 1, \ldots. \tag{3.4}$$

From (2.12), we have

$$\begin{aligned} P(A_n) &\geq \int_a^b t^{2n} dq(t) \,.\, k_{nn}^2 \\ &= \mu_{2n} \,.\, k_{nn}^2. \end{aligned} \tag{3.5}$$

LEMMA. *For μ_n, defined as in* (3.4), *we have*

$$\lim_{n\to\infty} \mu_{2n}^{1/n} = \max(a^2, b^2),$$

if a and b are the smallest and largest points of increase of $q(t)$, respectively, in $[a, b]$.

Proof. A point in $[a, b]$ is said to be a *point of increase* of $q(t)$ if there does not exist a two-sided (one sided in the case of the end points a and b) interval which contains the point, and in which $q(t)$ is constant.

In this case,

$$\begin{aligned} \mu_{2n}^{1/n} &= \left\{\int_a^b t^{2n} dq(t)\right\}^{1/n} \\ &\leq \max(a^2, b^2)\left\{\int_a^b dq(t)\right\}^{1/n}. \end{aligned}$$

Hence, $\limsup_{n\to\infty} \mu_{2n}^{1/n} \leq \max(a^2, b^2)$.

However, if $K < \max(a^2, b^2)$, then

$$\mu_{2n}/K^n = \int_a^b (t^2/K)^n dq(t) \to \infty \quad \text{as } n \to \infty.$$

Therefore, $\liminf_{n\to\infty} \mu_{2n}^{1/n} \geq \max(a^2, b^2)$. The result follows immediately.

THEOREM 2. *If A_n is the matrix* (3.1) *then,*

$$\liminf_{n\to\infty} P(A_n)^{1/n} \geq \max(a^2, b^2) \liminf_{n\to\infty} a_n^2, \tag{3.6}$$

where a_n is defined in (3.2).

Proof. From (3.5), we have

$$\begin{aligned} \liminf_{n\to\infty} P(A_n)^{1/n} &\geq \liminf_{n\to\infty} \mu_{2n}^{1/n} \liminf_{n\to\infty} (k_{nn}^2)^{1/n}, \\ &= \max(a^2, b^2) \liminf_{n\to\infty} (k_{nn}^2)^{1/n}, \end{aligned}$$

on applying the lemma.

But the limit inferior of the nth roots of a positive sequence is not less than the limit inferior of the ratios of its terms; see [4, p. 277]. Hence,

$$\begin{aligned}\liminf_{n\to\infty} (k_{nn}^2)^{1/n} &\geq \liminf_{n\to\infty} (k_{nn}/k_{n-1,n-1})^2 \\ &= \liminf_{n\to\infty} a_n^2,\end{aligned}$$

from (3.3). Thus we have the result.

EXAMPLE. *Jacobi Polynomials.*

Corresponding to the weight function $dq/dt = w(t) = (1-t)^\alpha(1+t)^\beta$ in $[-1, 1]$, where $\alpha, \beta > -1$, we have the orthonormal system of polynomials

$$p_n(t) = \left\{\frac{2n+\alpha+\beta+1}{2^{\alpha+\beta+1}} \cdot \frac{(n+1)(n+\alpha+\beta+1)}{(n+\alpha+1)(n+\beta+1)}\right\}^{1/2} P_n^{(\alpha,\beta)}(t); \tag{3.7}$$

see Szegö [3, p. 68]. $P_n^{(\alpha,\beta)}(t)$ denotes the classical Jacobi polynomial.

From the three-term recurrence relation for these polynomials, we have

$$\begin{aligned}a_n &= \left\{\frac{(2n+\alpha+\beta+1)n(n+\alpha+\beta)}{(2n+\alpha+\beta-1)(n+\alpha)(n+\beta)}\right\}^{1/2} \frac{(2n+\alpha+\beta-1)(2n+\alpha+\beta)}{2n(n+\alpha+\beta)}, \\ &\to 2 \text{ as } n\to\infty.\end{aligned}$$

Hence, from (3.6), we see that $P(A_n)$ is diverging at least as rapidly as 4^n.

The classic example of ill-conditioning is that of the segments, $A_n = (1/(i+j-1))$, of the *Hilbert matrix*; see [5]. This corresponds to the matrix (3.1), with the weight function $w(t) = 1$, over the interval $[0, 1]$. The corresponding orthonormal system of polynomials is, from (3.7) with $\alpha = \beta = 0$,

$$2^{1/2} p_n(2t-1) = (2n+1)^{1/2} P_n^{(0,0)}(2t-1).$$

But the coefficient of the highest term t^n in $P_n^{(\alpha,\beta)}(t)$ is $2^{-n}\binom{2n+\alpha+\beta}{n}$; see [3, p. 63]. Hence, the highest coefficient of $2^{1/2}p_n(2t-1)$ is

$$k_{nn} = (2n+1)^{1/2}(2n!)/(n!)^2.$$

Thus, for the segments of the Hilbert matrix, we have from (3.5), since

$$\mu_{2n} = \int_0^1 t^{2n}\,dt = 1/(2n+1),$$

$$P(A_n) \geq (2n!)^2/(n!)^4 \sim 16^n/\pi n.$$

Further results can be obtained using the asymptotic properties of general orthogonal polynomials. For example, if $dq/dt = w(t)$ is a weight function on the interval $[-1, 1]$ such that $w(\cos\theta)|\sin\theta| = f(\theta)$ belongs to the class of functions $f(\theta) \geq 0$, defined and measurable in $[-\pi, \pi]$, for which the integrals $\int_{-\pi}^{\pi} f(\theta)d\theta$ and $\int_{-\pi}^{\pi} |\log(\theta)|d\theta$ exist with the first integral supposed positive, then we have the asymptotic formula for the leading coefficients of the associated orthonormal polynomials

$$k_{nn} \sim \pi^{-1/2} 2^n \exp\left\{-\frac{1}{2\pi}\int_{-1}^{1} \log w(t)/(1-t^2)^{1/2} dt\right\};$$

see [**3**, p. 309]. This indicates, on using (3.6), that for such a weight function the condition number is growing at least as rapidly as 4^n.

These examples illustrate the ill-conditioning which is associated with matrices of the type (3.1). We now prove that any matrix of the form (3.1), whatever the choice of function $q(t)$, is strongly ill-conditioned. The condition numbers $P(A_n)$, in every case, diverge at least as rapidly as 4^n.

THEOREM 3. *If A_n is the matrix* (3.1), *then*

$$r(A_n^{-1}) \geq [4/(b-a)]^{2n} \bigg/ \left[4\int_a^b dq(t)\right]. \tag{3.8}$$

Proof. We denote the Chebyshev polynomials on $[-1, 1]$ by

$$T_n(t) = \cos(n \text{ arc cos } t) = 2^{n-1}t^n + \text{terms of lower degree}.$$

Therefore,

$$\frac{(b-a)^n}{2^{2n-1}} T_n\left[\frac{2t-(b+a)}{(b-a)}\right]$$

is a monic polynomial, and

$$\min_p \int_a^b |t^n - p(t)|^2 dq(t) \leq [(b-a)^n/2^{2n-1}]^2 \cdot \int_a^b \left|T_n^2\left[\frac{2t-(b+a)}{(b-a)}\right]\right| dq(t),$$

where the minimum is taken over the class of polynomials of degree $n-1$ or less. Therefore, from (2.5), since $|T_n(t)| \leq 1$ on $[-1, 1]$,

$$r(A_n^{-1}) \geq [2^{2n-1}/(b-a)^n]^2 \bigg/ \int_a^b dq(t).$$

COROLLARY. *For any matrix A_n, as in* (3.1), *we have*

$$\liminf_{n\to\infty} P(A_n)^{1/n} \geq \max(a^2, b^2)[4/(b-a)]^2 \geq 4.$$

Proof. From (2.2) and (3.8),

$$P(A_n) \geq \int_a^b t^{2n} dq(t)[4/(b-a)]^{2n}/4\int_a^b dq(t).$$

Hence, we have

$$\liminf_{n\to\infty} P(A_n)^{1/n} \geq \max(a^2, b^2)[4/(b-a)]^2.$$

But $\max(a^2, b^2)/(b-a)^2 \geq \frac{1}{4}$, so we obtain the result.

The situation is even worse when the range of integration is infinite. From

(2.4), we have

$$P(A_n) \geq \max_{0 \leq i \leq n} \mu_{2i} / \min_{0 \leq i \leq n} \mu_{2i}.$$

Because of the infinite range of integration, this diverges more rapidly than any geometric sequence.

For example, if $w(t) = t^\alpha \exp(-t)$, where $\alpha > -1$, $a = 0$, $b = \infty$, we have

$$\mu_{2n} = \int_0^\infty t^{2n+\alpha} \exp(-t) dt = \Gamma(2n + \alpha + 1).$$

Therefore,

$$P(A_n) \geq \Gamma(2n + \alpha + 1)/\Gamma(\alpha + 1) = (2n + \alpha)(2n + \alpha - 1) \ldots (\alpha + 1).$$

4. Interpolation Problems

Closely related to the least squares problem is the problem of finite interpolation. Let X be a (real or complex) linear space and denote its (algebraic) dual by X^f.

Given an n-dimensional subspace $M \subset X$ and $x \in X \backslash M$, the problem of finite interpolation is to find $u \in M$ such that

$$f_i(u) = f_i(x), \quad i = 1, 2, \ldots, n, \tag{4.1}$$

where $f_i \in X^f$ for $i = 1, 2, \ldots, n$.

If a basis $\{x_1, x_2, \ldots, x_n\}$ is chosen for M, the solution of (4.1) is

$$u = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n, \tag{4.2}$$

where

$$G\boldsymbol{a} = \boldsymbol{b} \tag{4.3}$$

and $G = (f_i(x_j))$, $\boldsymbol{b} = (f_1(x), f_2(x), \ldots, f_n(x))^T$.

If we define the norm of the matrix G to be $\|G\| = r(G^*G)^{1/2}$, then we have the related condition number

$$k(G) = \|G\| \,.\, \|G^{-1}\|, \tag{4.4}$$

where G^{-1} exists if (4.1) has a solution.

It is easy to verify that the expression,

$$[x, y] = \sum_{i=1}^n f_i(x)\overline{f_i(y)}, \tag{4.5}$$

defines an inner product on M and a semi-inner product on X.

The solution to the problem

$$\min_{u \in M} \|x - u\|^2 = \min_{u \in M} \sum_{i=1}^n |f_i(x - u)|^2 \tag{4.6}$$

is obviously the solution of (4.1). The problem of finite interpolation can therefore be viewed as a least squares problem and has a solution (4.2) with $Aa = c$, where $A = G^*G$ and $c = G^*b$. We have $P(A) = k^2(G)$ and, from Theorem 1,

$$k^2(G) \geq \sum_{j=1}^{n} |f_j(x_i)|^2 / \min_{u \in N_i} \sum_{j=1}^{n} |f_j(x_i - u)|^2, \tag{4.7}$$

for any $i = 1, 2, \ldots, n$.

If $v \in N_i$ is the solution of the interpolation problem

$$f_j(v) = f_j(x_i) \quad \text{for} \quad j = 1, 2, \ldots, n, j \neq i, \tag{4.8}$$

for any $i = 1, 2, \ldots, n$, then

$$|f_i(x_i - v)|^2 = \sum_{j=1}^{n} |f_j(x_i - v)|^2 \geq \min_{u \in N_i} \sum_{j=1}^{n} |f_j(x_i - u)|^2$$

and

$$k^2(G) \geq \sum_{j=1}^{n} |f_i(x_i)|^2 / |f_i(x_i - v)|^2. \tag{4.9}$$

This is the corresponding result to Theorem 1, for the problem of finite interpolation.

EXAMPLE. *Vandermonde Matrices.*

Consider the Vandermonde matrices,

$$G_n = (z_r^s), \quad 1 \leq r \leq n, \quad 0 \leq s \leq n-1, \quad n = 1, 2, \ldots,$$

which arise in the well-known problem of polynomial interpolation.

In this case, the subspace M is the set of polynomials of degree less than n and the elements of the dual space are the functionals corresponding to evaluation at the points $z = z_i$, $i = 1, 2, \ldots, n$, i.e. $f_i(p) = p(z_i)$ for $i = 1, 2, \ldots, n$, if $p \in M$. The powers $\{1, z, z^2, \ldots, z^{n-1}\}$ form a basis for M.

Now, $v(z) = z^{n-1} - (z - z_1)(z - z_2) \ldots (z - z_{n-1})$ is the polynomial of degree $n - 2$ such that $v(z_j) = z_j^{n-1}$ for $j = 1, 2, \ldots, n-1$. Hence $v(z)$ is the solution, in this case, of the interpolation problem (4.8), with $i = n$. Therefore, from (4.9),

$$k^2(G_n) \geq \sum_{j=1}^{n} |z_j^{n-1}|^2 / |(z_n - z_1)(z_n - z_2) \ldots (z_n - z_{n-1})|^2, \tag{4.10}$$

for $n = 1, 2, \ldots$.

For example, if we choose the n equidistant points $z_j = (j-1)/(n-1)$, $j = 1, 2, \ldots, n$, from [0, 1], we have from (4.10),

$$\begin{aligned} k^2(G_n) &\geq \sum_{j=1}^{n} |(j-1)/(n-1)|^{2n-2} / \{(n-1)![1/(n-1)^{n-1}]\}^2 \\ &= \sum_{j=1}^{n} [(j-1)^{n-1}/(n-1)!]^2. \end{aligned}$$

Hence, $k(G_{n+1}) \geq n^n/n! \sim e^n/(2\pi n)^{1/2}$, by Stirling's formula. We therefore expect ill-conditioning for large n.

Consider the infinite triangle of points.

$$T: \begin{array}{lll} z_{11} & & \\ z_{21} & z_{22} & \\ z_{31} & z_{32} & z_{33}, \\ \vdots & & \end{array}$$

where we suppose that the points of T are chosen from some set E, in the complex plane, and are *dense in the limit*, i.e. for any neighbourhood of E, there exists an integer N such that for all $n > N$ every row of T contains points of the neighbourhood. This is the general situation for a sequence of interpolation problems. We denote by G_n the Vandermonde matrix associated with the nth row.

Let $T_n(z; E)$ denote the Chebyshev polynomial of degree n, with leading coefficient unity, associated with the set E; see [6, p. 265]. Then

$$\min \sum_{j=1}^{n} |z^{n-1} - p(z_{nj})|^2 \leq \sum_{j=1}^{n} |T_{n-1}(z_{nj}; E)|^2,$$

where the minimum is taken over the class of polynomials of degree $n-2$ or less. If $M_n(E) = \max\{|T_n(z; E)| : z \in E\}$, then we have, from (4.7),

$$k^2(G_n) \geq \sum_{j=1}^{n} |z_{nj}^{n-1}|^2 / nM_{n-1}^2(E).$$

It easy to prove, under the above assumptions, that

$$\lim_{n\to\infty} \sum_{j=1}^{n} |z_{nj}^{n-1}|^2 = \sup\{|z|^2 : z \in E\},$$

which is a similar result to the lemma in section 3. Hence,

$$\liminf_{n\to\infty} k^2(G_n)^{1/n} \geq \sup\{|z|^2 : z \in E\}/\rho^2(E),$$

where $\rho(E) = \lim_{n\to\infty} M_n(E)^{1/n}$ is the well-known *Chebyshev constant* associated with the set E; see [6, p. 266], For example, if $E = [a, b]$ then $\rho(E) = (b-a)/4$ and $\sup\{|z|^2 : z \in E\} = \max(a^2, b^2)$. Therefore, in this case,

$$\liminf_{n\to\infty} k^2(G_n)^{1/n} \geq 16 \,.\, \max(a^2, b^2)/(b-a)^2 \geq 4,$$

which is similar to the previous results in section 3.

5. Application to Moment Matrices

Let H be an inner product space, as before, and let $B : H \to H$ be a linear map. For a given $x \in H$, we have the *moments* of B,

$$x_i = Bx_{i-1} = B^{i-1}x, \quad i = 1, 2, \ldots, \tag{5.1}$$

if $x_1 = x$.

The matrix A, defined by (1.4) and (5.1), occurs in the *Method of Moments* for solving linear equations in Hilbert spaces; see [7, p. 18].

From (2.8) with $i = n$, we have

$$\begin{aligned} P(A) &\geq \|x_n\|^2/\min_{u\in N_n} \|u - x_n\|^2 \\ &= \|B^{n-1}x\|^2/ \min_{p(B)\in\Pi_{n-1}} \|p(B)x\|^2, \end{aligned} \tag{5.2}$$

where Π_n is the collection of monic polynomials in B of degree n or less. If H is a Hilbert space and B is a bounded, self-adjoint operator, then we have the integral representations

$$\|B^n x\|^2 = \int |z^{n2}| d\|E_z x\|^2$$

and

$$\|p(B)x\|^2 = \int |p(z)|^2 d\|E_z x\|^2,$$

where E_z denotes a family of projection operators associated with the mapping A; see [8, p. 351]. Hence we see that (5.2), in this case, reduces to the situation in section 3, with $q(t) = \|E_t x\|^2$. We therefore must expect moment matrices, formed from the moments of a bounded, self-adjoint operator, to be ill-conditioned when n is large. In fact the condition numbers are growing at a rate at least as fast as 4^n. This corroborates the experience of Vorobyev [7, p. 37].

Krylov's method, for calculating the characteristic polynomial of a finite dimensional operator B, is closely related to the previous problem. If B is n-dimensional and x is of grade n [see 9, p. 37], then there are coefficients $\{a_1, a_2, \ldots, a_n\}$ such that

$$x_n + a_n x_{n-1} + \cdots + a_1 x = 0.$$

This can be written as

$$Ga = -x_n, \tag{5.3}$$

where x_i, $i = 1, 2, \ldots, n$, are given by (5.1), and G is the matrix whose columns are the vectors x_i. Hence,

$$G^*G = (x_j^* x_i) = ([x_i, x_j]), \tag{5.4}$$

which is a matrix of type (1.4). Therefore, from (5.2),

$$k^2(G) = P(A) \geq [B^{n-1}x, B^{n-1}x]/ \min_{p(B)\in\Pi_{n-1}} [p(B)x, p(B)x]. \tag{5.5}$$

If B is *normal*, i.e. $B^*B = BB^*$, then

$$\begin{aligned} \min_{p(B)\in\Pi_{n-1}} [p(B)x, p(B)x] &\leq [T_{n-1}(B; E)x, T_{n-1}(B; E)x] \\ &\leq |r(T_{n-1}(B; E))|^2 x^* x, \end{aligned}$$

where E is any set containing the spectrum of B.

By the spectral mapping theorem, we have

$$r(T_{n-1}(B; E)) = \sup \{|T_{n-1}(z; E)|: z \in \sigma(B)\} \leq M_{n-1}(E).$$

Hence, from (5.5),

$$k(G)^{1/n} \geq \{[B^{n-1}x, B^{n-1}x]/x^*x\}^{1/2n}/M_{n-1}(E)^{1/n}$$
$$\sim r(B)/\rho(E), \tag{5.5}$$

since x is of grade n.

If, for example, the spectrum of B lies in the interval $E = [a, b]$, where a and b are eigenvalues of B, then

$$r(B)/\rho(E) = \max(|a|, |b|)/|\tfrac{1}{4}(b-a)| \geq 4.$$

This result is supported by computational experience with Krylov's method. It is very ill-conditioned when the matrix B has real eigenvalues; see [**9**, pp. 369–373].

References

1 J. Todd. On condition numbers. *Programmation en Mathématiques Numériques. Actes Colloq. Internat. C.N.R.S.* 165, *Besancon* (1966), 141–159. Paris: Editions C.N.R.S., 1968.
2 R. Bellman. Introduction to matrix analysis (New York: McGraw-Hill, 1960).
3 G. Szegö. *Orthogonal polynomials* (Providence, R.I.: Amer. Math. Soc., 1939).
4 K. Knopp. *Theory and application of infinite series* (Glasgow: Blackie, 1949).
5 J. Todd. The condition of the finite segments of the Hilbert matrix. *Nat. Bureau Standards Appl. Math. Ser.* **39** (1954), 109–116.
6 E. Hille. *Analytic function theory* II (Massachussets: Ginn, 1959).
7 U. V. Vorobyev. Moments method in applied mathematics. (Delhi: Hindustani Publishing, 1962).
8 A. E. Taylor. *Introduction to functional analysis* (New York: Wiley, 1958).
9 J. H. Wilkinson. *The algebraic eigenvalue problem* (Oxford: Clarendon Press, 1965).

(*Issued* 15 *September* 1978)