LEAST pTH POWER POLYNOMIALS ON A REAL FINITE POINT SET[1]

BY

T. S. MOTZKIN[2] AND J. L. WALSH[3]

It is our object in the present paper to study best approximation on a set E consisting of a finite number of distinct points $z_1, z_2, \cdots, z_m$, to a given function $f(z)$ defined on E, by polynomials $p_n(z)$ of given degree[4] $n(\leq m-2)$. As *deviation* or *measure of approximation* we use primarily the sum

$$\sum_{k=1}^{m} \mu_k \left| f(z_k) - p_n(z_k) \right|, \qquad \mu_k > 0, \tag{1}$$

where the μ_k are given, but also use various broad generalizations of (1). The analogous problem for approximation on a finite interval, where (1) is replaced by an integral, has been studied by Korkine and Zolotareff [6], and for approximation by linear families more general than polynomials by Laasonen [7].

An important special case of the problem of approximation is the choice $f(z) \equiv z^{n+1}$; here the difference $f(z)-p_n(z)$ is the *T-polynomial of degree* $n+1$, namely the polynomial $T_{n+1}(z) \equiv z^{n+1}+b_1z^n+ \cdots +b_n$ of best approximation on E to the function zero.

Our entire discussion deals largely with the separation of points of a real set E by zeros of $T_{n+1}(z)$ or by points of interpolation of $p_n(z)$ to $f(z)$.

Of especial interest is (1) as a measure of approximation on a real set in the case $n=m-2$, for then (as we shall prove) the polynomials of best approximation to $f(x)$ on E can be found by interpolation to $f(x)$ in certain preassigned points of E which are independent of $f(x)$. This is analogous to the results of Laasonen.

We emphasize approximation on a real set E, but some of the present results apply also to a non-real set, and a later paper will be devoted to the more general case. In §1 we mention briefly existence and nonuniqueness of extremal polynomials. In §2 we consider best approximation as determined by interpolation. In §3 we study separation by E of the zeros of T-polynomials for general deviation, and in §4 approximation by arbitrary families of func-

Presented to the Society, December 29, 1952; received by the editors August 28, 1953.

(1) This work was performed under a contract with the University of California for the National Bureau of Standards, and under a contract with Harvard University; it was sponsored (in part) by the Office of Naval Research, USN, and by the Office of Scientific Research, USAF.

(2) University of California and Harvard University.

(3) Harvard University and National Bureau of Standards.

(4) Any polynomial $a_0z^n+a_1z^{n-1}+ \cdots +a_n$ is said to be *of degree n*.

tions. In §5 we consider the totality of extremal polynomials as a convex set, in §6 approximation in the sense of least pth powers $(0<p<1)$, and in §7 we consider further the separation properties of zeros of T-polynomials. Then §8 is devoted to the specific determination of the polynomials $T_{m-1}(x)$, and finally §9 to the mutual separation of zeros of T-polynomials of various degrees.

1. **Existence and nonuniqueness of extremal polynomials.** The existence of an extremal polynomial follows (compare [12, §12.3]) with (1) as norm from the fact that if a sequence of polynomials of degree n is bounded on E, each coefficient is bounded and hence there exists some subsequence which converges on E.

But the extremal polynomial need not be unique. Even in the simple case $m=2$, $z_1=0$, $z_2=1$, $f(z)\equiv z$, $\mu_k=1$, every polynomial of degree zero of the form $p_0(z)=\mu$, $0\leqq\mu\leqq 1$, yields approximation with measure (1) equal to unity.

As another illustration of the nonuniqueness of the polynomial of best approximation, we mention $z_1=0$, $z_2=\delta$ $(0<\delta<1/2)$, $z_3=1-\delta$, $z_4=1$, $f(z_1)=f(z_4)=1$, $f(z_2)=f(z_3)=0$, $\mu_k=1$; any real linear function whose graph cuts the nonhorizontal sides of the trapezoid $(0, 1)$, $(\delta, 0)$, $(1-\delta, 0)$, $(1, 1)$ is a polynomial of degree unity of best approximation. Also in the important case of the T-polynomial, $T_{n+1}(z)$ may fail to be unique.

2. **Approximation determined by interpolation.** The case $m\leqq n+1$ is trivial, for in that case there exists an admissible polynomial $p_n(z)$ which coincides with $f(z)$ in all the points of E, and for which (1) vanishes. We shall consider in some detail the first nontrivial case, $m=n+2$.

THEOREM 1. *In the case $m=n+2$, let $P(z)$ be the unique polynomial of degree $n+1$ coinciding with $f(z)$ on E; then each extremal polynomial $p_n(z)$ with deviation* (1) *is a polynomial of degree n found by interpolation to $P(z)$ in the zeros of some $T_{n+1}(z)$; conversely, any polynomial of degree n found by interpolation to $P(z)$ in the zeros of a $T_{n+1}(z)$ is a polynomial of best approximation.*

In the notation $P(z)\equiv a_0z^{n+1}+a_1z^n+\cdots$, we have to study the minimum of

$$\sum_{k=1}^{n+2} \mu_k \left| P(z_k) - p_n(z_k) \right| = \sum_{k=1}^{n+2} \left| a_0 \right| \mu_k \left| z_k^{n+1} + \cdots \right|$$

and if $|a_0|\neq 0$ the solution of this minimum problem is

$$P(z) - p_n(z) \equiv a_0T_{n+1}(z).$$

If $p_n(z)$ is an extremal polynomial, then $p_n(z)$ coincides with $P(z)$ in the zeros of $T_{n+1}(z)$; if $p_n(z)$ coincides with $P(z)$ in the zeros of some $T_{n+1}(z)$, then the difference $P(z)-p_n(z)$ is divisible by $T_{n+1}(z)$, hence equal to $a_0T_{n+1}(z)$. The case $a_0=0$ is exceptional in this reasoning but trivial, for then $P(z)-p_n(z)\equiv 0$.

Theorem 1 is especially significant in the case that E is real, for then (§3, below) a suitable polynomial $T_{n+1}(z)$ can be chosen *to vanish only on E; the corresponding polynomial $p_n(z)$ interpolates to $P(z)$ and hence to $f(z)$ in $n+1$ points of E*; these $n+1$ points can be chosen to be independent of $f(z)$. However, we do not assert that all extremal polynomials $p_n(z)$ can be found in this way; the polynomial $T_{n+1}(z)$ is not necessarily unique, and under some conditions may be chosen to have not all its zeros on E; since $f(z)$ is assumed to be defined only on E, interpolation in points not on E to $f(z)$ has no significance. Even for families of functions more general than polynomials, interpolating extremal functions may exist; see §4.

Theorem 1 has been established for the case of (1) as a measure of approximation, but extends at once to various other measures of approximation on E, such as

$$\max\,[\mu_k\,|\,f(z_k) - p_n(z_k)\,|\,], \qquad \mu_k > 0; \tag{2}$$

$$\sum_{k=1}^{m} \mu_k\,|\,f(z_k) - p_n(z_k)\,|^p, \qquad p > 0,\ \mu_k > 0. \tag{3}$$

Indeed, Theorem 1 extends to an arbitrary measure of approximation which is homogeneous (not necessarily of degree unity) in the individual errors $|f(z_k)-p_n(z_k)|$.

For deviation (3) with $p>1$, the polynomial $p_n(z)$ of best approximation is unique; if two polynomials have the same deviation, half their sum has a smaller deviation.

It follows by reasoning due to Fejér [3; also in 10, §6] for deviations (1), (2), (3), or even for more general derivations and more general point sets, that the zeros of $T_{n+1}(z)$ lie in the convex hull of E. The zeros of the $T_{n+1}(z)$ have been studied in more detail by Fekete and von Neumann [5] with deviation (2) and by Fekete [4] with more general deviations.

3. **Zeros of T-polynomial for E real.** Methods developed by Korkine and Zolotareff [6] and by Achyeser [1] for approximation on an interval apply with some modification to the present study if E is real. Like Fejér, we use a deviation δ of considerable generality.

Fejér assumes merely that his norm (which he calls a *monotone deviation*) is defined for polynomials of given degree, rather than for arbitrary error functions of the form $f(z)-p_n(z)$, and assumes that his norm has a monotonic property for polynomials which may have zeros on E. He requires that the norm be greater for any polynomial $g(z)\equiv z^n+\cdots$ of degree n than for any *underpolynomial* $h(z)\equiv z^n+\cdots$ of the same degree; here $h(z)$ is defined to be an underpolynomial of $g(z)$ if $h(z)=g(z)$ on the subset of E on which $g(z)$ vanishes, and if $|h(z)|<|g(z)|$ on the remaining subset of E. For the case of polynomials Fejér's requirements are identical with ours, except that his set E may be infinite. His norm is more general than δ in the sense that because defined only for polynomials of given degree it may depend conceivably on special intrinsic properties of polynomials, such as coefficients or values in points not on E, hence may not be defined or not subject to monotonicity requirements for more general functions. Our δ is defined for functions not necessarily polynomials and depends only on the errors at the individual points of E. We state our results for an extremal polynomial(5), but we shall prove elsewhere that the set of extremal polynomials $x^n+\cdots$ for deviation (1) coincides (for real E) with the set of all polynomials $x^n+\cdots$ having no underpolynomials.

THEOREM 2. *Let E be a real point set $x_1, x_2, \cdots, x_m$ with $x_k<x_{k+1}$, and suppose $m\geqq n+2$. Let $\delta[\delta_1, \delta_2, \cdots, \delta_m]$ be a positive function of the non-negative variables δ_k when $\sum_k\delta_k>0$, which decreases whenever all the δ_k not zero decrease and the δ_k which are zero remain unchanged. Let $T_{n+1}(x)$ be the (or a) T-polynomial of degree $n+1$ for E with the deviation $\delta[|T_{n+1}(x_1)|, |T_{n+1}(x_2)|, \cdots, |T_{n+1}(x_m)|]$. If ξ is a zero of $T_{n+1}(x)$, then at least one point x_k at which $T_{n+1}(x)/(x-\xi)$ does not vanish lies in each of the intervals $-\infty<x\leqq\xi$ and $\xi\leqq x<\infty$. If ξ and $\eta(\geqq\xi)$ not necessarily distinct are two zeros of $T_{n+1}(x)$, then at least one point x_k at which $T_{n+1}(x)/(x-\xi)(x-\eta)$ does not vanish satisfies the inequalities $\xi\leqq x_k\leqq\eta$.*

We do not assume all the zeros of $T_{n+1}(x)$ real, but the reasoning of Fejér is valid, and hence all zeros lie in the convex hull of E, so these zeros are all real.

We set first $T_{n+1}(x)=(x-\xi)\phi(x)$, and introduce the auxiliary polynomial of degree $n+1$: $F(x)\equiv[(x-\xi)-\epsilon]\phi(x)$. If only points x_k at which $\phi(x)$ vanishes lie in the interval $\xi\leqq x<\infty$, we choose ϵ negative but so small that no point of E lies in the interval $\xi+\epsilon<x<\xi$. There exist points of E at which $\phi(x)$ fails to vanish, for we have $m=n+2$; all such points lie in the interval $-\infty<x<\xi+\epsilon$; at such a point x_j we have $|F(x_j)|<|T_{n+1}(x_j)|$ unless the latter number is zero; at each point x_j in the interval $\xi\leqq x<\infty$ we have $F(x_j)=T_{n+1}(x_j)=0$, so the deviation of $F(x)$ is less than that of $T_{n+1}(x)$ and the latter polynomial is not extremal. If only points x_k at which $\phi(x)$ vanishes lie in the interval $-\infty<x\leqq\xi$, a similar discussion with suitable choice of positive ϵ leads to a similar contradiction. The first part of Theorem 2 is established.

The second part of Theorem 2 is established similarly. We set $T_{n+1}(x)\equiv(x-\xi)(x-\eta)\phi(x)$, and introduce the auxiliary polynomial of degree $n+1$: $F(x)\equiv[(x-\xi)(x-\eta)-\epsilon]\phi(x)$, where $\epsilon(>0)$ is chosen so small that the zeros ξ_0 and $\eta_0(>\xi_0)$ of the square bracket are real, with no point of E in the open

(5) An extremal polynomial may fail to exist if the deviation (even though monotone) is not continuous. This is shown by the example $\delta(\delta_1, \delta_2)\equiv(1/2)\delta_1+\delta_2$ $(\delta_1+\delta_2<1;\ \delta_1+\delta_2=1,\ \delta_1<1/2)$, $\delta(\delta_1, \delta_2)\equiv\delta_1+\delta_2$ $(\delta_1+\delta_2>1;\ \delta_1+\delta_2=1,\ \delta_1\geqq 1/2)$, which is monotone for all non-negative δ_1 and δ_2. With the deviation $\delta(|T_1(0)|, |T_1(1)|)$ there exists no (extremal) T-polynomial of degree unity; in other words, the deviation $\delta(|a|, |1-a|)$ of the polynomial $z-a$ on the set E: (0, 1) has no minimum for variable a.

interval between ξ and ξ_0 or in the open interval between η and η_0. There exist points of E at which $\phi(x)$ is different from zero, for we have $m \geqq n+2$; if all such points x_j lie exterior to the interval $\xi \leqq x \leqq \eta$, we have $|F(x_j)| < |T_{n+1}(x_j)|$ unless the latter number is zero, so $T_{n+1}(x)$ is not an extremal polynomial. This completes the proof. It is also true, and may be proved by choosing ϵ negative, that at least one point x_k at which $\phi(x)$ does not vanish lies in the pair of intervals $-\infty < x_k \leqq \xi$, $\eta \leqq x_k < \infty$, but a stronger result has already been established. Theorem 2 shows the impossibility of various orderings of the points x_k and the zeros of $T_{n+1}(x)$. We mention, as a consequence of Theorem 2, some explicit situations which cannot occur; we do not formulate explicitly the situations obtained therefrom by reversal of order, which are of course likewise impossible. That Cases I and II are impossible follows from the first part of Theorem 2, and the remainder from the second part.

Case I. $T_{n+1}(\xi)=0$, $\xi \neq x_j$, all x_k(if any) with $\xi < x_k$ are zeros of $T_{n+1}(x)$. The impossibility here follows also from Fejér's results if $x_m < \xi$.

Case II. $T_{n+1}(x)$ has a multiple zero at $\xi = x_j$, all x_k (if any) with $\xi < x_k$ are zeros of $T_{n+1}(x)$.

Case III. $T_{n+1}(x)$ has a zero of multiplicity greater than one at $\xi \neq x_j$.

Case IV. $T_{n+1}(x)$ has a zero of multiplicity greater than two at $\xi = x_j$.

Case V. $T_{n+1}(x)$ has zeros at ξ and $\eta(>\xi)$ not in E; all points of E (if any) between ξ and η are zeros of $T_{n+1}(x)$.

Case VI. $T_{n+1}(x)$ has a multiple zero at ξ(in E) and a zero at η (not in E); all points of E (if any) between ξ and η are zeros of $T_{n+1}(x)$.

Case VII. $T_{n+1}(x)$ has multiple zeros at x_j and x_k, $j<k$; all points of E (if any) between x_j and x_k are zeros of $T_{n+1}(x)$.

Every multiple zero of $T_{n+1}(x)$ is of order two and lies on E.

It is a consequence of Case II that x_1 cannot be a double zero of $T_{n+1}(x)$, and a consequence of Case I that $T_{n+1}(x)$ can have no zero in the interval $x_1 < x < x_2$ if $T_{n+1}(x_1)=0$.

With the measure of approximation (1), further results are available, as we proceed to show.

THEOREM 3. *With the measure of approximation* (1), *whenever the extremal polynomial* $T_{n+1}(x)$ *has a zero* ξ *in the interval* $x_k < \xi < x_{k+1}$, *that zero can be chosen arbitrarily in the corresponding closed interval without modifying the extremal character of* $T_{n+1}(x)$.

As before, we set $T_{n+1}(x) \equiv (x-\xi)\phi(x)$, whence

$$\begin{aligned}\sum_j \mu_j |T_{n+1}(x_j)| &= \sum_{j\leqq k} \mu_j(\xi - x_j)|\phi(x_j)| + \sum_{j>k} \mu_j(x_j - \xi)|\phi(x_j)| \\ &= \xi\Big[\sum_{j\leqq k} \mu_j|\phi(x_j)| - \sum_{j>k} \mu_j|\phi(x_j)|\Big] \\ &\quad - \sum_{j\leqq k} \mu_j x_j|\phi(x_j)| + \sum_{>k} \mu_j x_j|\phi(x_j)| .\end{aligned}$$

The square bracket must be zero; otherwise the last member could be made smaller by suitable choice of ξ in the given open interval, contrary to the definition of $T_{n+1}(x)$. Since the square bracket is zero, the last member is independent of the value of ξ in the closed interval $x_k \leqq \xi \leqq x_{k+1}$.

COROLLARY 1. *If an extremal polynomial $T_{n+1}(x)$ with* (1) *as deviation has two coincident zeros at a point x_k of E, then one of these zeros can be displaced arbitrarily in the interval $x_{k-1} < x < x_{k+1}$ without altering the deviation of $T_{n+1}(x)$.*

The proof follows precisely that of Theorem 3.

It is a consequence of this corollary that with (1) as a measure of deviation *there exists some $T_{n+1}(x)$ whose zeros lie on E and are simple*; for one component of any multiple zero may be shifted continuously to a point of E not originally a zero of $T_{n+1}(x)$, without altering the extremal property of the polynomial $T_{n+1}(x)$; we use the impossibility of Case VII. This fact is of significance for Theorem 1; the zeros of $T_{n+1}(x)$, namely the points of interpolation of $p_n(x)$ to $f(x)$, may be chosen all distinct.

COROLLARY 2. *In the case $m = n+2 \geqq 3$ and with* (1) *as deviation, $\mu_j = 1$, there exists a $T_{n+1}(x)$ with zeros in both x_1 and x_m. Indeed, every $T_{n+1}(x)$ whose zeros lie on E and are simple vanishes in both x_1 and x_m.*

By the consequence of Corollary 1, at least one of the polynomials

$$F_k(x) \equiv \prod_{j \neq k} (x - x_j)$$

is extremal. The deviations of $F_1(x)$ and $F_2(x)$ are respectively

$$\left| (x_1 - x_2)(x_1 - x_3) \cdots (x_1 - x_m) \right|,$$
$$\left| (x_2 - x_1)(x_2 - x_3) \cdots (x_2 - x_m) \right|,$$

and it is clear that the latter is less than the former; a corresponding comparison applies to $F_{m-1}(x)$ and $F_m(x)$. Then $F_1(x)$ and $F_m(x)$ are not extremal, *so every extremal $T_{n+1}(x)$ whose zeros lie on E and are simple vanishes in both x_1 and x_m*. This conclusion is generalized below (§5).

Corollary 2 *does not extend to arbitrary m and n*, as we show by the example $m=6$, $x_1 = -2$, $x_2 = -1-\delta$, $x_3 = -1$, $x_4 = 1$, $x_5 = 1+\delta$, $x_6 = 2$, where δ is infinitesimal. For the polynomials x^2-4, $(x+2)(x+1)$, $(x+2)(x-1)$, x^2-1, the deviations are respectively (except for possible added infinitesimals) 12, 24, 8, 6, so for δ sufficiently small no extremal polynomial vanishes in x_1 or x_6.

Even in the case $m = n+2$, the deviation of $F_k(x)$ is not necessarily convex, considered as a function of x_k. We show this by the example $m=5$, $x_1 = -1-\delta$, $x_2 = -1$, $x_3 = 0$, $x_4 = 1$, $x_5 = 1+\delta$, $\mu_k = 1$. The corresponding deviations for $F_1(x)$, $F_2(x)$, $F_3(x)$ are respectively $\delta(1+\delta)(2+\delta)(2+2\delta) = 4\delta + 10\delta^2 + \cdots$, $2\delta(2+\delta) = 4\delta + 2\delta^2$, $(1+\delta)^2 = 1 + 2\delta + \delta^2$, and the nonconvexity for δ sufficiently small follows. This same example shows that $T_{n+1}(x)$ *may have a double zero on* E, for the two polynomials $F_2(x)=(x+1+\delta)x(x-1)(x-1-\delta)$ and $F_4(x)=(x+1+\delta)(x+1)x(x-1-\delta)$ are both extremal, hence half their sum $(x+1+\delta)x^2(x-1-\delta)$ is also extremal. Here with the special choice $\delta=2^{1/2}-1$, all three polynomials $F_2(x)$, $F_3(x)$, $F_4(x)$ have all their zeros on E and are all extremal.

4. **Approximation by arbitrary families of functions.** We investigate now, following suggestions made to the writers by Professor A. Ostrowski, best approximation to an arbitrary function $F(x)$ on E: $(x_1, x_2, \cdots, x_m)$ by linear combinations of the given functions $\psi_1(x), \psi_2(x), \cdots, \psi_{n+1}(x)$ defined on E, $n \leqq m-1$. The measure of approximation

$$\sum_{k=1}^{m} \mu_k \,|\, F(x_k) - P(x_k)\,|\,, \qquad P(x) = \sum_{\nu=1}^{n+1} \alpha_\nu \psi_\nu(x), \tag{4}$$

is to be minimized. It will be useful to have the functions $\psi_\nu(x)$ satisfy Condition A, namely, that *if any n points of E are given, there exists some linear combination of the $\psi_\nu(x)$ which vanishes in those points but is different from zero in at least one of the remaining points of E.* We prove

THEOREM 4. *Suppose $m \geqq n+1$, and the set $\psi_\nu(x)$ satisfies Condition* A. *Then there exists at least one extremal function $P(x)$ defined by* (4) *which coincides with $F(x)$ in at least $n+1$ points of E.*

Let $P(x)$ be the (or an) extremal function of the prescribed form which coincides with $F(x)$ in the maximum number ρ of points of E. If $\rho \geqq n+1$, the conclusion is satisfied. Otherwise let $\phi(x) \equiv \sum_1^{n+1} \beta_\nu \psi_\nu(x)$ vanish in this set E' of $\rho(\leqq n)$ points of E but be different from zero in at least one other point of E. For δ numerically small we consider

$$\begin{aligned}\sum_{k=1}^{m} \mu_k \,|\, (F(x_k) - P(x_k) - \delta\phi(x_k)\,| \\ &= \sum_{k=1}^{m} \mu_k \epsilon_k [F(x_k) - P(x_k) - \delta\phi(x_k)] \\ &= \sum_{k=1}^{m} \mu_k \,|\, F(x_k) - P(x_k)\,| - \delta \sum_{k=1}^{m} \mu_k \epsilon_k \phi(x_k);\end{aligned}$$

here ϵ_k is defined as plus or minus unity, according as its original factor is positive or negative. We choose δ so small that on $E-E'$ the functions $F(x_k)-P(x_k)$ and $F(x_k)-P(x_k)-\delta\phi(x_k)$ have the same algebraic sign. If the last coefficient of δ does not vanish, a suitable choice of positive or negative δ yields a smaller deviation for the function $P(x)-\delta\phi(x)$ of the linear family than for $P(x)$. On the other hand, if the last coefficient of δ vanishes, we can increase $|\delta|$ monotonically so that the deviation remains constant until

$F(x)-P(x)-\delta\phi(x)$ first vanishes in a point of $E-E'$; then an extremal function coincides with $F(x)$ in $\rho+1$ points of E, contrary to hypothesis. This contradiction completes the proof.

Of course for approximation by polynomials we have $\psi(x)=x^{\nu-1}$, $\nu=1, 2, \cdots, n+1$, and we need merely set $\phi(x)\equiv(x-x_1')(x-x_2')\cdots(x-x_n')$ to show that Condition A is satisfied, where the x_k' are the given n points of E.

Condition A need not be satisfied for an arbitrary set of functions $\psi_\nu(x)$; indeed we might have each $\psi_\nu(x)$ constant on E or on a subset of E. But if there always exists some linear combination of the $\psi_\nu(x)$ which assumes $n+1$ arbitrarily prescribed values respectively in $n+1$ arbitrary points of E (for which various sufficient conditions are well known), Condition A is satisfied. Adding to this "solvence" condition a similar one for "unisolvence," properties of approximating functions of a general, not necessarily linear, family are studied in [9]. Likewise, suppose whenever n columns are chosen from

$$\left\|\begin{matrix} \psi_1(x_1) & \psi_1(x_2) & \cdots & \psi_1(x_m) \\ \psi_2(x_1) & \psi_2(x_2) & \cdots & \psi_2(x_m) \\ \cdot & \cdot & \cdot & \cdot \\ \psi_{n+1}(x_1) & \psi_{n+1}(x_2) & \cdots & \psi_{n+1}(x_m) \end{matrix}\right\|$$

there exists some column not linearly dependent on them; then the functions $\psi_\nu(x)$ satisfy Condition A.

5. **Totality of extremal polynomials as a convex set.** It is clear from the triangle inequality that with (1) as deviation the totality of polynomials $p_n(x)$ of given degree of best approximation to an arbitrary $f(x)$ on an arbitrary set E form a convex set, in the sense that if $p_n^{(1)}(x)$ and $p_n^{(2)}(x)$ are two such polynomials, so also is $\lambda p_n^{(1)}(x)+(1-\lambda)p_n^{(2)}(x)$, $0<\lambda<1$. We proceed to investigate further the totality of such polynomials in the case that E is real, more especially when E consists of $m=n+2$ points, with $f(x)\equiv x^{n+1}$; thus we are studying the totality of the functions $T_{n+1}(x)$ of Theorem 1 in this case.

The points $y_1, y_2, \cdots, y_{n+1}$ are said to *separate* the points $x_1, x_2, \cdots, x_{n+2}$ (we suppose $x_k<x_{k+1}$) if and only if we have

$$x_1 \leqq y_1 \leqq x_2 \leqq y_2 \leqq \cdots \leqq y_{n+1} \leqq x_{n+2}, \tag{5}$$

and the separation is said to be *strong* if and only if this relation holds without the equality signs.

THEOREM 5. *If we set* $P_k(x)\equiv\prod_{j\neq k}(x-x_j)$, $k=1, 2, \cdots, n+2$, *then the polynomials* $P(x)=x^{n+1}+bx^n+\cdots$ *whose zeros separate the points* x_k *are precisely the polynomials*

$$P(x) \equiv \sum_{k=1}^{n+2} \lambda_k P_k(z), \qquad \lambda_k \geqq 0, \qquad \sum \lambda_k = 1. \tag{6}$$

Every polynomial $P(x)$ of form (6) with every $\lambda_k > 0$ has zeros strongly separating the x_k, for we have $P_{n+2}(x_{n+2}) > 0$, $P_{n+1}(x_{n+1}) < 0$, $P_n(x_n) > 0, \cdots$, whence $P(x_{n+2}) > 0$, $P(x_{n+1}) < 0$, $P(x_n) > 0, \cdots$. Consequently every polynomial which is a limit of polynomials $P(x)$ with $\lambda_k > 0$ also has zeros separating the x_k, so the sufficiency of (6) is established.

Conversely, let the zeros of $P(x) \equiv x^{n+1} + bx^n + \cdots$ separate strongly the points x_k. Each zero of $P(x)$ is simple, so we have $P(x_{n+2}) > 0$, $P(x_{n+1}) < 0$, $P(x_n) > 0, \cdots$. The linear independence of the polynomials $P_k(x)$ implies that we can write $P(x) \equiv \sum \lambda_k P_k(x)$, whence $\lambda_k = P(x_k)/P_k(x_k) > 0$. Moreover we have $\sum \lambda_k = 1$. Every polynomial $P(x) \equiv x^{n+1} + \cdots$ whose zeros separate the points x_k is the limit of a variable polynomial of the same form whose zeros separate strongly the x_k, so Theorem 5 is established.

In every case under Theorem 1, the totality of extremal polynomials $T_{n+1}(x)$ forms a convex set. By a *basic* extremal polynomial we understand an extremal polynomial which cannot be expressed as a linear combination (with positive coefficients) of two distinct extremal polynomials. Since the set of all extremal polynomials is convex, closed, and bounded, each extremal polynomial can be expressed [Minkowski, 8, vol. II, p. 160] as a linear combination of basic extremal polynomials. It follows from Theorem 3 and its Corollary 1 that if E is real, every basic extremal polynomial has its zeros simple and lying in points of E; every extremal polynomial can be expressed as a linear combination with positive coefficients of these, and these are finite in number.

We choose now $m = n+2$; let $r_\rho(x)$ denote a smallest basic set of extremal polynomials; the totality of extremal polynomials are then precisely the set

$$\sum \lambda_\rho r_\rho(x), \qquad \lambda_\rho \geqq 0, \qquad \sum \lambda_\rho = 1.$$

Let the points x_μ be the common zeros of all the $r_\rho(x)$, and the remaining points of E be the points x_ν. We set $\delta(x) = \prod_\mu (x - x_\mu)$, $r_\rho(x) = \delta(x) q_\rho(x)$. *Then the totality of extremal polynomials is precisely the set*

$$\delta(x) \sum \lambda_\rho q_\rho(x), \qquad \lambda_\rho \geqq 0, \qquad \sum \lambda_\rho = 1. \tag{7}$$

Except in the case $(m = n+2)$ that but one extremal polynomial exists, there are at least two basic extremal polynomials, the totality of zeros of those polynomials consists of E, and each polynomial $q_\rho(x)$ vanishes in all the points x_ν but one. Each x_ν is a nonzero of some $q_\rho(x)$. Thus the set $q_\rho(x)$ is precisely a set of the kind considered in Theorem 5, with a slight change of notation. The zeros of a polynomial $r_\rho(x)$ consist (by Theorem 5) of the x_μ and a set of points which separate the x_ν; the x_μ need not be distinct from the latter set. Conversely, any polynomial $T_{n+1}(x) = x^{n+1} + \cdots$ which vanishes in the x_μ and whose other zeros separate the x_ν is an extremal polynomial. Here we have a complete geometric characterization of the extremal polynomials. The zeros of every extremal polynomial separate the points of E, by Theorem 5.

It is a consequence of Corollary 2 to Theorem 3 that if $\mu_k=1$, $n\geqq 1$, $m=n+2$, every basic extremal polynomial vanishes in both x_1 and x_m, so there are at most n such polynomials. Every extremal polynomial vanishes in x_1 and x_m.

As an illustration we choose δ as in §3, $m=5$, $x_1=-1-\delta$, $x_2=-1$, $x_3=0$, $x_4=1$, $x_5=1+\delta$, $\delta=2^{1/2}-1$, $\mu_k=1$, and set $P_k(x)=\prod_{j\neq k}(x-x_j)$. The three polynomials P_2, P_3, P_4 are all basic, and an extremal polynomial with a double zero is $[P_2(x)+P_4(x)]/2$; here $x_3=0$ is both a point of E and a zero of some $q_\rho(x)$.

6. **Least pth power approximation.** The following theorem is based on a remark made to the writers by Professor A. Dvoretzky:

THEOREM 6. *Let E consist of the real points $x_1, x_2, \cdots, x_m$ $(m\geqq n+1)$, let $F(x)$ be defined on E, let p $(0<p<1)$ be given, and let the functions $\psi_1(x), \psi_2(x), \cdots, \psi_{n+1}(x)$ satisfy Condition* A (§4). *Then every function $P(x)\equiv\sum_1^{n+1}\alpha_\nu\psi_\nu(x)$ of best approximation measured by the deviation*

$$\sum_{k=1}^{m}\mu_k\left|F(x_k)-P(x_k)\right|^p \tag{8}$$

coincides with $F(x)$ in at least $n+1$ points of E.

As in the proof of Theorem 4, suppose $P(x)$ to be an extremal function which coincides with $F(x)$ in precisely $\rho(\geqq 0)$ points E' of E. We show that the assumption $\rho\leqq n$ leads to a contradiction. Let $\phi(x)\equiv\sum_1^{n+1}\beta_\nu\phi_\nu(x)$ vanish in E', but be different from zero in at least one other point of E. The function $P(x)+\delta\phi(x)$ belongs to the linear family under consideration.

For δ numerically small we consider

$$\sum_{k=1}^{m}\mu_k\left|F(x_k)-P(x_k)-\delta\phi(x_k)\right|^p, \tag{9}$$

a sum of which ρ terms vanish and of which other terms may conceivably be constant. Each of the remaining terms (at least one of which must exist, by the definition of $\phi(x)$) is of the form $|A+B\delta|^p$, where A and B are constants with $B\neq 0$. Unless the function $|A+B\delta|^p$ of δ with $B\neq 0$ is zero, its graph is locally concave downward, so a sum of such functions plus a constant cannot have a local minimum. Thus (9) is not a local minimum for $\delta=0$ unless at least one of the remaining terms vanishes for $\delta=0$; hence $P(x)$ coincides with $F(x)$ in at least $\rho+1$ points of E, contrary to the definition of ρ.

Theorem 6 implies that every extremal polynomial $P(x)$ is found by interpolation to $F(x)$ in $n+1$ points of E; there exist but a finite number of polynomials interpolating to $F(x)$ in $n+1$ points of E, so every extremal polynomial can be found merely by comparing their measures of approximation.

Of course Theorem 6 implies that every T-polynomial, namely a polynomial $T_{n+1}^{(p)}(x)=x^{n+1}+bx^n+\cdots$ of minimum norm $\sum_{k=1}^m \mu_k |T_{n+1}^{(p)}(x_k)|^p$, has all its zeros on E, whether or not we assume $m=n+2$.

7. **Zeros of T-polynomials, continued.** We return to the situation of §3, to study in more detail the location of the zeros of T-polynomials.

THEOREM 7. *With the hypothesis of Theorem* 2, *the ordered zeros* $y_1, y_2, \cdots, y_{n+1}$ *of every* $T_{n+1}(x)$ *separate a suitably chosen ordered subset* E': $x_1', x_2', \cdots, x_{n+2}'$ *of* E, *in the sense of* (5).

We establish this result by the use of the cases of impossibility enumerated in §3 as a consequence of Theorem 2, by examining the number $N_T(x)$ of zeros of $T_{n+1}(x)$ not greater than x and the number $N_{E'}(x)$ of points of E' not greater than x, as x increases monotonically. We show that for every x in the interval $x_1 \leqq x \leqq x_m$ we have

$$N_T(x) \leqq N_{E'}(x) \leqq N_T(x) + 1. \tag{10}$$

Always as x increases monotonically we adjoin each new x_k to E' if and only if after adjunction the relation (10) holds for $x=x_k$. In the proof the word *zero* refers to a zero of $T_{n+1}(x)$.

Each simple zero at a point of E shall be called a *zero of the first kind*; as x moves across such a zero y_k each of the numbers $N_T(x)$ and $N_{E'}(x)$ is increased by one unit, so that the difference is unchanged and (10) if originally valid persists with the same equality and the same inequality signs as before. Each other zero is called a *zero of the second kind*; in such a point y_k there is one more zero than points of E, so as x increases through y_k the difference $N_{E'}(x)-N_T(x)$ is decreased by one unit, and if (10) holds for $x<y_k$ with the respective signs $<$ and $=$ it becomes valid with the signs $=$ and $<$ for $x>y_k$. The precise content of the impossibility of Cases V, VI, and VII can be expressed: *between two successive zeros of the second kind must lie at least one point of E not a zero.*

The relation (10) holds for $x=x_1$; in every case we choose $x_1'=x_1$, a point which cannot (Case II) be a double zero. The relation (10) then holds for $x=y_1$ if y_1 is a zero of the first kind and also holds (Cases I and II) if y_1 is a zero of the second kind.

If $T_{n+1}(x)$ possesses only zeros of the first kind, the conclusion is immediate; the relation (10) holds in each x_k, and some x_k is $(m>n+1)$ not a zero, whence $N_T(x_m)=n+1<N_{E'}(x_m)$.

If $T_{n+1}(x)$ possesses zeros of the second kind, let y_k be the first (i.e., smallest) of them. By Cases I and II relation (10) holds for $x=y_k$. The next succeeding zero y_ρ of the second kind must be separated from y_k by at least one point of E not a zero, so (10) holds also for $x=y_\rho$. When we arrive at the last zero y_j of the second kind we have $N_T(y_j)\leqq N_{E'}(y_j)$. Cases I and II show that some $x_k>y_j$ is not a zero, so we have $N_T(x_m)=n+1<N_{E'}(x_m)$, which completes the proof.

Of course it is a consequence of (6) that the points y_k are the respective limits of points that strongly separate the points x_k' of E. It is to be noticed that Theorem 7 applies in particular to the deviation (3), for all $p>0$.

8. **Determination of the polynomials** $T_{m-1}(x)$. We have studied (Theorems 2 and 7) the separation of the points of E by the zeros of $T_{m-1}(x)$ especially by the method of adding positive or negative constants to various linear and quadratic factors of $T_{m-1}(x)$. It is conceivable that further information could be obtained by similar consideration of factors of $T_{m-1}(x)$ of degree higher than two. However, it turns out that the description of separation indicated in Theorem 7 is completely characteristic of the extremal polynomials with deviation (1), as we proceed to indicate for degree $m-1$ in the corollary to Theorem 8. In Theorem 8 the set E need not be real.

THEOREM 8. *The totality of T-polynomials $T_{m-1}(z)$ of degree $m-1$ for E: $(z_1, z_2, \cdots, z_m)$ with measure of approximation*

$$\mu(T_{m-1}) = \sum_1^m \mu_i \left| T_{m-1}(z_i) \right|, \qquad \mu_i > 0, \tag{11}$$

is found as follows. With the notation $\omega(z) \equiv \prod_1^m (z-z_i)$, arrange the numbers $\mu_i|\omega'(z_i)|$ in order of magnitude, and choose $T_{m-1}(z_i)=0$ except in the subset E': $(z_1', z_2', \cdots, z_p')$ of E on which $\mu_i|\omega'(z_i)|$ takes its smallest value; on E' choose $T_{m-1}(z_i)$ arbitrarily but so that $T_{m-1}(z_i')=\lambda_i\omega'(z_i')$, $\lambda_i \geqq 0$, $\sum_1^p \lambda_i = 1$; then we have

$$T_{m-1}(z) \equiv \omega(z) \sum_1^m \frac{\lambda_i}{z-z_i},$$

where we set $\lambda_i=0$ in the points of E not in E'.

Any polynomial $P(z)$ of degree $m-1$ is expressed by Lagrange's interpolation formula:

$$P(z) \equiv \sum_1^m \frac{P(z_i)}{\omega'(z_i)} \frac{\omega(z)}{z-z_i}, \tag{12}$$

and for $P(z) \equiv z^{m-1} + \cdots$ we have consequently

$$1 = \sum_1^m \frac{P(z_i)}{\omega'(z_i)}. \tag{13}$$

Thus $T_{m-1}(z)$ is the (or a) polynomial $P(z)$ defined by (12) where the arbitrary numbers $P(z_i)$ are chosen to satisfy (13) and to minimize

$$\mu(P) = \sum_1^m \mu_i \left| P(z_i) \right| = \sum_1^m \mu_i \left| \omega'(z_i) \right| \left| \frac{P(z_i)}{\omega'(z_i)} \right|.$$

It is now clear that the minimum of $\mu(P)$ is found by choosing

$$1 = \sum_{1}^{m} \left| \frac{P(z_i)}{\omega'(z_i)} \right|,$$

and further by choosing $P(z_i) = \lambda_i \omega'(z_i)$ in the manner described in the statement of Theorem 8. The polynomial $T_{m-1}(z)$ is unique when and only when E' consists of but one point.

In any case, it follows from Theorem 5 that the zeros of $T_{m-1}(x)$ separate the points of E, provided E is real.

Theorem 8 is somewhat similar to the determination of the T-polynomials using (2) as deviation, by Fekete and von Neumann [5], later studied more deeply by Fekete [4].

COROLLARY. *Under the conditions of Theorem 8 and with the choice* $\mu_i = \mu_0 / |\omega'(z_i)|$, $\mu_0 > 0$, *for every* i, *the set of* T-*polynomials* $T_{m-1}(z)$ *is precisely*

$$\sum_{1}^{m} \frac{\lambda_i \omega(z)}{z - z_i}, \qquad \lambda_i \geqq 0, \qquad \sum_{1}^{m} \lambda_i = 1.$$

In other words the totality of T-*polynomials is the convex set depending on the polynomials* $P_k(z) \equiv \omega(z)/(z - z_k)$, *and if* E *is real is precisely the set of polynomials whose zeros separate the points of* E.

The last statement follows from Theorem 5.

On another occasion the present writers expect to determine and characterize the T-polynomials defined by deviation (3), with especial reference to separation of points of E by zeros.

9. **Mutual separation of zeros of the** $T_n(x)$. Atkinson [2] has indicated that the classical properties [e.g., Szegö, 11, §3.3] of separation of zeros of polynomials orthogonal on a finite real interval are exhibited also by the zeros of the T-polynomials for that interval with deviation measured by the integral of the weighted pth power, $p > 1$; the classical case is precisely that of the T-polynomials for $p = 2$. The methods of Atkinson apply without essential change in our present case of deviation (3), $p > 1$, where E is real, as we now indicate. If $T_n(x)$ is the necessarily unique polynomial $T_n(x) \equiv x^n + bx^{n-1} + \cdots$ which minimizes

$$\sum_{k=1}^{m} \mu_k |T_n(x_k)|^p, \qquad p > 1, \tag{14}$$

and if $t(x)$ is an arbitrary polynomial of degree $n-1$, it is readily proved by variational methods that we have

$$\sum_{k=1}^{m} \mu_k |T_n(x_k)|^{p-1} sg[T_n(x_k)] t(x_k) = 0, \tag{15}$$

which represents a kind of orthogonality.

We omit the elementary proof of the

LEMMA. *If α and β are real, we have*

$$sg[\alpha\,|\,\alpha\,|^{p-2} + \beta\,|\,\beta\,|^{p-2}] = sg[\alpha + \beta];$$

if either of the numbers in square brackets vanishes, so does the other.

We suppose the points x_k of E so arranged that $x_1 < x_2 < \cdots < x_m$, and of course $m \geqq n+1$.

THEOREM 9. *If c is real and $0 < \nu \leqq n$, the polynomial $F(x) \equiv T_n(x) + cT_\nu(x) (\not\equiv 0)$ has at least ν sign-changes in I: $x_1 \leqq x \leqq x_m$.*

The function $F(x)$ cannot vanish at all points of E. If the theorem is false, there exists a polynomial $G(x)$ of degree less than ν which throughout I has the same sign as $F(x)$, and which by the lemma then has the same sign throughout I as $T_n(x)\,|\,T_n(x)\,|^{p-2} + cT_\nu(x)\,|\,T_\nu(x)\,|^{p-2}$. Thus we have

$$(16) \qquad \sum_{k=1}^{m} \mu_k \{\, |\, T_n(x_k)\,|^{p-1} sg[T_n(x_k)] + c\,|\, T_\nu(x_k)\,|^{p-1} sg[T_\nu(x_k)]\} \cdot G(x_k) > 0.$$

But (16) contradicts the relation (15) for orthogonality of $G(x)$ to $T_n(x)$ and $T_\nu(x)$, which completes the proof. We emphasize the fact that the polynomial $G(x)$ is subject to no condition at a point of I at which $F(x)$ vanishes but does not change sign, and is subject to no condition at either of the points x_1 or x_m if $F(x)$ vanishes there. In particular the choice $c=0$ shows that $T_n(x)$ itself has at least n sign changes interior to I, so *all the zeros of $T_n(x)$ are simple and lie in the open interval $x_1 < x < x_m$.*

It follows from Theorem 9 that the polynomial $H(x) \equiv T_n(x) + cT_{n-1}(x)$ has at least $n-1$ sign-changes in I, thus has at least $n-1$ zeros of odd order, so has its n zeros all real and distinct. Then the polynomials $T_n(x) + cT_{n-1}(x)$, $T_n'(x) + cT_{n-1}'(x)$ have no common zero, whence the polynomial

$$(17) \qquad T_n(x)T_{n-1}'(x) - T_n'(x)T_{n-1}(x)$$

has no real zeros. From the coefficients of x^n and x^{n-1} in $T_n(x)$ and $T_{n-1}(x)$ respectively it follows that the coefficient of x^{2n-1} in (17) is negative, so we have proved

THEOREM 10. *For all real x the polynomial* (17) *is negative.*

From Theorem 10 it follows that the polynomials $T_n(x)$ and $T_{n-1}(x)$ have no common zero, and also, by differentiation of the function $T_n(x)/T_{n-1}(x)$, that the latter increases monotonically wherever it is defined. If ξ and η are two successive zeros of $T_n(x)$, the function $T_n(x)/T_{n-1}(x)$ vanishes for $x=\xi$ and $x=\eta$, hence must have a discontinuity in that interval:

THEOREM 11. *The zeros of* $T_n(x)$ *are strongly separated by those of* $T_{n-1}(x)$, $n \geqq 2$.

Theorem 11 is the main result of §9.

We have already pointed out that every $T_n(x)$ has all its zeros *interior* to I. The monotonic character of $T_n(x)/T_{n-1}(x)$ can be written for $x > x_m$ in the form

$$\frac{T_n(x)}{T_{n-1}(x)} > \frac{T_n(x_m)}{T_{n-1}(x_m)},$$

so (as is likewise pointed out by Atkinson for the polynomials that he considers) the sequence

$$\frac{T_n(x)}{T_n(x_m)}, \qquad n = 0, 1, 2, \cdots,$$

increases monotonically with n.

BIBLIOGRAPHY

1. N. Achyeser, *Verallgemeinerung einer Korkine-Zolotareffschen Minimum-Aufgabe*, Comm. Inst. Sci. Math. Méch. Univ. Kharkoff (4) vol. 13 (1936) pp. 3–14.

2. F. V. Atkinson, *Über die Nullstellen gewisser extremaler Polynome*, Archiv der Mathematik vol. 3 (1952) pp. 83–86.

3. L. Fejér, *Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen*, Math. Ann. vol. 85 (1922) pp. 41–48.

4. M. Fekete, *On the structure of extremal polynomials*, Proc. Nat. Acad. Sci. U.S.A. vol. 37 (1951) pp. 95–103.

5. M. Fekete and J. von Neumann, *Über die Lage der Nullstellen gewisser Minimumpolynome*, Jber. Deutschen Math. Verein. vol. 31 (1922) pp. 125–138.

6. A. N. Korkine and G. Zolotareff, *Sur un certain minimum*, Nouvelles Annales vol. 12 (1873) pp. 337–355.

7. P. Laasonen, *Einige Sätze über Tschebyscheffsche Funktionensysteme*, Ann. Acad. Sci. Fennicae, Sec. A, Math. Phys. no. 52 (1949).

8. H. Minkowski, *Gesammelte Abhandlungen*, Leipzig, 1911.

9. T. S. Motzkin, *Approximation by curves of a unisolvent family*, Bull. Amer. Math. Soc. vol. 55 (1949) pp. 789–793.

10. T. S. Motzkin and J. L. Walsh, *On the derivative of a polynomial and Chebyshev approximation*, Proc. Amer. Math. Soc. vol. 4 (1953) pp. 76–87.

11. G. Szegö, *Orthogonal polynomials*, New York, 1939.

12. J. L. Walsh, *Interpolation and approximation*, New York, 1935.

UNIVERSITY OF CALIFORNIA,
LOS ANGELES, CALIF.
HARVARD UNIVERSITY,
CAMBRIDGE, MASS.
NATIONAL BUREAU OF STANDARDS,
LOS ANGELES, CALIF.