in Theorem 3. It is of interest to note that Theorem 1 remains true if the norm $\|\cdot\|_\infty$ of (2) is replaced by the semi-norm $|\!|\cdot|\!|_{\infty,S}$ defined as

$$|\!|f(x)|\!|_{\infty,S} \equiv \sup_{x \in S} |f(x)|$$

where S is any subset of $[a, b]$ containing at least $n + 1$ points.

From the Weierstrass approximation theorem, it follows that

$$\lim_{n \to \infty} D(f, \hat{P}_n) = 0.$$

Furthermore, if $f(x)$ has r continuous derivatives in $[a, b]$, then by the convergence result for expansions in Chebyshev polynomials (see Theorem 9 in Subsection 3.4)

$$D(f, \hat{P}_n) = \mathcal{O}(n^{1-r}) \qquad \text{for } r \geq 2.$$

4.1. The Error in the Best Approximation

It is a relatively easy matter to obtain bounds on the deviation of the best approximation polynomial of degree n. Let us call this quantity

$$d_n(f) \equiv \min_{\{a_0, \ldots, a_n\}} d(a_0, \ldots, a_n) \equiv \min_{\{P_n(x)\}} D(f, P_n). \tag{4}$$

Then for any polynomial $P_n(x)$ we have the upper bound

$$d_n(f) \leq D(f, P_n).$$

Lowers bounds can be obtained by means of

THEOREM 2 (DE LA VALLÉE-POUSSIN). *Let an nth degree polynomial $P_n(x)$ have the deviations from $f(x)$*

$$f(x_j) - P_n(x_j) = (-1)^j e_j, \qquad j = 0, 1, \ldots, n + 1, \tag{5}$$

where $a \leq x_0 < x_1 < \cdots < x_{n+1} \leq b$ and all $e_j > 0$ or else all $e_j < 0$. Then

$$\min_j |e_j| \leq d_n(f). \tag{6}$$

Proof. Assume that for some polynomial $Q_n(x)$,

$$D(f, Q_n) < \min_j |e_j|. \tag{7}$$

Then the nth degree polynomial

$$Q_n(x) - P_n(x) = [f(x) - P_n(x)] - [f(x) - Q_n(x)]$$

has the same sign at the points x_j as does $f(x) - P_n(x)$. Thus, there are $n + 1$ sign changes and consequently, at least $n + 1$ zeros of this difference.

But then, this nth degree polynomial identically vanishes and so $P_n(x) \equiv Q_n(x)$, which from (7) and (2) is impossible. This contradiction arose from assuming (7); hence $D(f, Q_n) \geq \min_j |e_j|$ for every polynomial $Q_n(x)$, and (6) is established. ∎

To employ this theorem we need only construct a polynomial which oscillates about the function being approximated at least $n + 1$ times. This can usually be done by means of an interpolation polynomial of degree n.

A necessary and sufficient condition which characterizes a best approximation polynomial and establishes its uniqueness is contained in

THEOREM 3 (CHEBYSHEV). *A polynomial of degree at most n, $P_n(x)$, is a best approximation of degree at most n to $f(x)$ in $[a, b]$ if and only if $f(x) - P_n(x)$ assumes the values $\pm D(f, P_n)$, with alternate changes of sign, at least $n + 2$ times in $[a, b]$. This best approximation polynomial is unique.*

Proof. Suppose $P_n(x)$ has the indicated oscillation property. Then let x_j with $j = 0, 1, \ldots, n + 1$ be $n + 2$ points at which this maximum deviation is attained with alternate sign changes. Using these points in Theorem 2 we see that $|e_j| = D(f, P_n)$ and hence

$$d_n(f) \geq D(f, P_n).$$

From equation (4), the definition of $d_n(f)$, it follows that $D(f, P_n) = d_n(f)$ and the $P_n(x)$ in question is a best approximation polynomial. This shows the sufficiency of the *uniform oscillation property*.

To demonstrate the necessity, we will show that if $f(x) - P_n(x)$ attains the values $\pm D(f, P_n)$ with alternate sign changes at most k times where $2 \leq k \leq n + 1$, then $D(f, P_n) > d_n(f)$. Let us assume, with no loss in generality, that

$$f(x_j) - P_n(x_j) = (-1)^j D(f, P_n), \qquad j = 1, 2, \ldots, k,$$

where $a \leq x_1 < x_2 < \cdots < x_k \leq b$. Then, there exist points $\xi_1, \xi_2, \ldots, \xi_{k-1}$, separating the x_j, i.e.,

$$x_1 < \xi_1 < x_2 < \xi_2 < \cdots < \xi_{k-1} < x_k$$

and an $\epsilon > 0$ such that $|f(\xi_j) - P_n(\xi_j)| < D(f, P_n)$ and

$$-D(f, P_n) \leq f(x) - P_n(x) < D(f, P_n) - \epsilon,$$

for x in the "odd" intervals, $[a, \xi_1]$, $[\xi_2, \xi_3]$, $[\xi_4, \xi_5], \ldots$; while

$$-D(f, P_n) + \epsilon < f(x) - P_n(x) \leq D(f, P_n),$$

for x in the "even" intervals, $[\xi_1, \xi_2]$, $[\xi_3, \xi_4], \ldots$. For example, we may define $\xi_1 = \frac{1}{2}(\eta_1 + \zeta_1)$ where $\eta_1 = \text{g.l.b.}\{\eta\}$ for $a \le \eta \le x_2$ and $f(\eta) - P_n(\eta) = D(f, P_n)$; and similarly: $\zeta_1 = \text{l.u.b.}\{\zeta\}$ for $a \le \zeta \le x_2$ and $f(\zeta) - P_n(\zeta) = -D(f, P_n)$. Then $x_1 \le \zeta_1 < \eta_1 \le x_2$; otherwise, we may insert η_1 and ζ_1 in place of x_1 in the original sequence and find $k + 1$ alternations of sign. That is, alternately for each of the k intervals $[a, \xi_1], \ldots, [\xi_{k-1}, b]$, the deviation $f(x) - P_n(x)$ takes on only one of the extreme deviations $\pm D(f, P_n)$ and is bounded away from the extreme of opposite sign. The polynomial

$$r(x) = (x - \xi_1)(x - \xi_2)\cdots(x - \xi_{k-1})$$

has degree $k - 1$ and is of one sign throughout each of the k intervals in question. Let the maximum value of $|r(x)|$ in $[a, b]$ be M. Now define $q(x) \equiv (-1)^k r(x)/2M$ and consider the nth degree polynomial (since $k - 1 \le n$)

$$Q_n(x) = P_n(x) + \epsilon q(x),$$

for sufficiently small positive ϵ. We claim that $D(f, Q_n) < D(f, P_n)$, and so $P_n(x)$ could not be a best approximation. Indeed, in the interior of any of the "odd" intervals $(a, \xi_1), (\xi_2, \xi_3), \ldots$, we have that $-\frac{1}{2} \le q(x) < 0$ and conversely in the "even" intervals $(\xi_1, \xi_2), (\xi_3, \xi_4), \ldots$, we have that $0 < q(x) \le \frac{1}{2}$. However, recalling the above inequalities,

$$-D(f, P_n) - \epsilon q(x) \le f(x) - Q_n(x) \le D(f, P_n) - \epsilon[1 + q(x)], \quad x \text{ in odd intervals;}$$

$$-D(f, P_n) + \epsilon[1 - q(x)] \le f(x) - Q_n(x) \le D(f, P_n) - \epsilon q(x), \quad x \text{ in even intervals.}$$

From the signs and magnitude of $q(x)$ in each interval, it easily follows that $D(f, Q_n) < D(f, P_n)$ and the proof of necessity is completed.

To demonstrate uniqueness we assume that there are two best approximations say, $P_n(x)$ and $Q_n(x)$, both of degree at most n. Since by assumption $D(f, P_n) = D(f, Q_n) = d_n(f)$, we have in $[a, b]$,

$$\begin{aligned} |f(x) - \tfrac{1}{2}[P_n(x) + Q_n(x)]| &\le \tfrac{1}{2}|f(x) - P_n(x)| + \tfrac{1}{2}|f(x) - Q_n(x)| \\ &\le d_n(f). \end{aligned}$$

Thus, $\frac{1}{2}[P_n(x) + Q_n(x)]$ is another best approximation and we must have, by the first part of the theorem,

$$|f(x) - \tfrac{1}{2}[P_n(x) + Q_n(x)]| = d_n(f)$$

at $n + 2$ distinct points in $[a, b]$. From the inequality, it follows that at these points $f(x) - P_n(x) = f(x) - Q_n(x) = \pm d_n(f)$. Thus, the difference $[f(x) - P_n(x)] - [f(x) - Q_n(x)] = Q_n(x) - P_n(x)$ vanishes at $n + 2$ distinct points. Since this difference is an nth degree polynomial, it vanishes identically, i.e., $Q_n(x) \equiv P_n(x)$, and the proof is complete. ∎

This theorem can be used to recognize the best approximation polynomial. It is also the basis, along with Theorem 2, of various methods for approximating the best approximation polynomial. There is no finite procedure for constructing the best approximation polynomial for arbitrary continuous functions. However, the best approximation is known in some important special cases; see, for example, the next subsection and Problem 2.

As an obvious consequence of Theorem 3, it follows that the best approximation, $P_n(x)$, of degree at most n is *equal* to $f(x)$, the function it approximates, *at* $n + 1$ *distinct points*, say $x_0, x_1, \ldots, x_n$. Thus, $P_n(x)$ is *the* interpolation polynomial for $f(x)$ with respect to the points $\{x_i\}$ (since by Lemma 2.1 the interpolation polynomial of degree at most n is unique). Of course, for an arbitrary continuous function, $f(x)$, a corresponding set of interpolation points $\{x_i\}$ is not known a priori. Thus, this observation cannot, in general, be used to determine $P_n(x)$. However, if $f(x)$ has $n + 1$ continuous derivatives, Theorem 2.1 applies since $P_n(x)$ is an interpolation polynomial, and we have determined a form for the error in the best approximation of degree at most n. In summary, these observations can be stated as a

COROLLARY. *Let $f(x)$ have a continuous $(n + 1)$st derivative in $[a, b]$ and let $P_n(x)$ be the best polynomial approximation to $f(x)$ of degree at most n in this interval. Then, there exist $n + 1$ distinct points $x_0, x_1, \ldots, x_n$ in $a < x < b$ such that*

$$(8) \qquad R_n(x) \equiv f(x) - P_n(x) = \frac{(x - x_0)(x - x_1)\cdots(x - x_n)}{(n + 1)!} f^{(n+1)}(\xi),$$

where $\xi = \xi(x)$ is in the interval:

$$\min(x, x_0, \ldots, x_n) < \xi < \max(x, x_0, \ldots, x_n). \qquad \blacksquare$$

4.2. Chebyshev Polynomials

In the expression (8) for $R_n(x)$, the error of the best approximation, it will, in general, not be known at what point, $\xi = \xi(x)$, the derivative is to be evaluated. Hence, the value of the derivative is not known. An exception to this is the case when $f^{(n+1)}(x) =$ constant, which occurs if and only if $f(x)$ is a polynomial of degree at most $n + 1$. In this special case, the error (8) can be minimized by choosing the points $x_0, x_1, \ldots, x_n$ such that the polynomial

$$(9a) \qquad (x - x_0)(x - x_1)\cdots(x - x_n)$$

has the smallest possible maximum absolute value in the interval in question (say, $a \le x \le b$). In the general case, the choice of these same interpolation points may be expected to yield a reasonable approximation to the best polynomial, i.e., the smaller the *variation* of $f^{(n+1)}(x)$ in $[a, b]$, the better the approximation.

We are thus led to consider the following problem: *Among all polynomials of degree $n + 1$, with leading coefficient unity, find the polynomial which deviates least from zero in the interval* $[a, b]$. In other words, we are seeking the best approximation to the function $g(x) \equiv 0$ among polynomials of the form

$$x^{n+1} - P_n(x), \tag{9b}$$

where $P_n(x)$ is a polynomial of degree at most n. Alternatively, the problem can then be formulated as: *find the best polynomial approximation of degree at most n to the function x^{n+1}.*

For this latter problem, Theorem 1 is applicable and we conclude that such a polynomial exists and it is uniquely characterized in Theorem 3. Thus, we need only construct a polynomial of the form (9) whose maximum absolute value is attained at $n + 2$ points with alternate sign changes.

Consider, for the present, the interval $[a, b] \equiv [-1, 1]$. We introduce the change of variable

$$x = \cos\theta, \tag{10}$$

which takes on each value in $[-1, 1]$ once and only once when θ is restricted to the interval $[0, \pi]$. Furthermore, the function $\cos(n + 1)\theta$ attains its maximum absolute value, unity, at $n + 2$ successive points with alternate signs for

$$\theta = j\left(\frac{\pi}{n+1}\right), \qquad j = 0, 1, \ldots, n + 1.$$

Therefore, the function

$$T_{n+1}(x) = A_{n+1}\cos(n + 1)\theta = A_{n+1}\cos[(n + 1)\cos^{-1}x], \tag{11}$$

has the required properties as regards its extrema. To show that $T_{n+1}(x)$ is also a polynomial in x of degree $n + 1$ we consider the standard trigonometric addition formula

$$\cos(n + 1)\theta + \cos(n - 1)\theta = 2\cos\theta\cos n\theta, \qquad n = 0, 1, \ldots. \tag{12}$$

Let us define

$$t_n(x) \equiv \cos(n\cos^{-1}x), \qquad n = 0, 1, 2, \ldots, \tag{13a}$$

in terms of which (12) becomes

$$t_{n+1}(x) = 2t_1(x)t_n(x) - t_{n-1}(x), \qquad n = 1, 2, 3, \ldots. \tag{13b}$$

Clearly, from (13a), $t_0(x) = 1$, $t_1(x) = x$ and so, by induction, it follows from (13b) that $t_{n+1}(x)$ is a polynomial in x of degree $n + 1$. It also follows by induction that

$$t_{n+1}(x) = 2^n x^{n+1} + q_n(x), \qquad n = 0, 1, 2, \ldots, \tag{13c}$$

where $q_n(x)$ is a polynomial of degree at most n. Thus, with the choice $A_{n+1} = 2^{-n}$ in (11), these results imply

$$\begin{aligned} T_{n+1}(x) &= 2^{-n} \cos [(n + 1) \cos^{-1} x] = 2^{-n} t_{n+1}(x) \\ &= x^{n+1} + 2^{-n} q_n(x). \end{aligned} \tag{14}$$

At the $n + 2$ points

$$\xi_k = \cos \frac{k\pi}{n + 1}, \qquad k = 0, 1, \ldots, n + 1, \tag{15a}$$

which are in $[-1, 1]$, we have from (14)

$$T_{n+1}(\xi_k) = 2^{-n} \cos k\pi = 2^{-n}(-1)^k. \tag{15b}$$

Thus we have proven that $T_{n+1}(x)$ is the polynomial of form (9) which deviates least from zero in $[-1, 1]$; the maximum deviation is 2^{-n}.

The polynomials in (14) are called the Chebyshev polynomials (of the first kind—see Problem 9 of Section 3). If the zeros of the $(n + 1)$-st such polynomial are used to construct an interpolation polynomial of degree at most n, then for x in $[-1, 1]$ the coefficient of $f^{(n+1)}(\xi)$ in the error (8) of this approximation will have the least possible absolute maximum.

If the interval of approximation for the continuous function $g(y)$ is $a \le y \le b$, then the transformation

$$x = \frac{a - 2y + b}{a - b} \quad \text{or} \quad y = \tfrac{1}{2}[(b - a)x + (a + b)] \tag{16}$$

converts the problem of approximating $g(y)$ into that of approximating $f(x) \equiv g[y(x)]$ in the x-interval $[-1, 1]$. The zeros of $T_{n+1}(x)$ are at

$$x_k = \cos\left(\frac{2k + 1}{n + 1}\frac{\pi}{2}\right), \qquad k = 0, 1, \ldots, n; \tag{17a}$$

and the corresponding interpolation points in $[a, b]$ are then at

$$y_k = \tfrac{1}{2}[(b - a)x_k + (a + b)], \qquad k = 0, 1, \ldots, n. \tag{17b}$$

The value of the maximum deviation of $\prod_{j=0}^{n} (y - y_j)$ from zero in $[a, b]$ is then, using (16) and (17b):

$$\begin{aligned} \max_{a \le y \le b} \prod_{j=0}^{n} |y - y_j| &= \left|\frac{b - a}{2}\right|^{n+1} \cdot \max_{-1 \le x \le 1} \prod_{j=0}^{n} (x - x_k) \\ &= \frac{1}{2^n}\left|\frac{b - a}{2}\right|^{n+1}. \end{aligned} \tag{18}$$