

# Generalized Rational Chebyshev Approximation

By Ichizo Ninomiya

**Abstract.** In this paper, a generalized rational Chebyshev approximation problem is considered.

The problem is this: To minimize the maximum absolute value of the "criterion function" of the error. By imposing a rather natural restriction on the criterion function, the problem is solved completely; the existence, the uniqueness and the characterization of the best approximation are clarified and interesting relationships between the best approximations corresponding to different criterion functions are found.

The theory is applied to the starting rational approximation for Newton iteration for  $x^{1/n}$ .

**1. Introduction.** A generalized rational Chebyshev approximation problem is determined by a septuple

$$[p, q; I, f(x), w(x); J, g(y)],$$

where  $p$  and  $q$  are nonnegative integers,  $I$  is a bounded closed interval,  $f$  is a continuous function on  $I$ ,  $w$  is a continuous and positive function on  $I$ ,  $J$  is an interval containing 0, and  $g$  is a continuous and strictly increasing function on  $J$  satisfying  $g(0) = 0$ .  $w$  is called a *weight function* and  $g$  a *criterion function*. No restrictions are imposed on the interval  $J$ , the domain of  $g$ .

For any continuous function  $F$  on  $I$ , the *error*  $e(F, x)$  is defined by

$$(1) \quad e(F, x) = (F(x) - f(x))/w(x)$$

and  $E(F)$  by

$$(2) \quad E(F) = \max [|g(e(F, x))|; x \in I].$$

A function is called *admissible* if it is a rational function of the form  $R(x) = P(x)/Q(x)$ , where  $P(x)$  and  $Q(x)$  are mutually prime polynomials of degrees not exceeding  $p$  and  $q$  respectively, and  $E(R)$  is finite. Evidently,  $R$  is admissible only if it is pole-free on  $I$ , and  $e(R, x) \in J$ . The set of admissible functions is denoted by  $\mathbf{R}$ .

The problem considered in this paper is the determination of an admissible function minimizing  $E(R)$ , i.e., an  $R^* \in \mathbf{R}$  such that

$$(3) \quad E(R^*) = \min [E(R); R \in \mathbf{R}].$$

Such an  $R^*$  is called a *best (rational Chebyshev) approximation to  $f$  with respect to the criterion  $g$* , or a *solution* of the problem  $[p, q; I, f, w; J, g]$ .

---

Received June 3, 1968, revised May 16, 1969.

*AMS Subject Classifications.* Primary 4117, 4140; Secondary 6520, 6525.

*Key Words and Phrases.* Rational approximation, criterion function, weight function, best approximation, existence of best approximation, uniqueness of best approximation, characterization of best approximation, relations between best approximations, starting approximation for square root.

The *standard Chebyshev approximation problem* is the special case,  $J = (-\infty, \infty)$ ,  $g(y) = y$  of this general problem.

Here a comment concerning the criterion function is appropriate. At first glance the condition imposed on  $g$  appears too restrictive to include a wide class of problems of practical interest in the present scope. However, every problem  $[p, q; I, f, w; J, g]$  for which  $|g(y)|$  is continuous, attains the minimum  $M$  at a point  $y^* \in J$  and is strictly monotone for  $y \leq y^*$  and  $y \geq y^*$ , can be reduced to an equivalent one  $[p, q; I, f^*, w; J^*, g^*]$  within the present scope by means of a simple transformation.

Such a transformation is given by

$$(4.1) \quad f^*(x) = f(x) + y^*w(x),$$

$$(4.2) \quad J^* = J - y^*,$$

$$(4.3) \quad g^*(y) = \operatorname{sgn}(y)(|g(y + y^*)| - M),$$

because  $g^*$  satisfies the condition of the criterion function and

$$(5) \quad E^*(F) = E(F) - M.$$

Recently, Dunham [2] and Moursund [3] treated analogous problems. In our notation, Dunham's definition of  $E(F)$  is

$$(6) \quad E(F) = \max [ |(g(F(x)) - f(x))/w(x)|; x \in I ],$$

and Moursund's is

$$(7) \quad E(F) = \max [ |W(x, F(x) - f(x))|; x \in I ],$$

where  $W(x, y)$  is his *generalized weight function*. Thus our problem is different from Dunham's and is included in Moursund's in its general form, although actually Moursund treated only polynomial approximations and imposed a stronger condition  $\lim_{|y| \rightarrow \infty} |W(x, y)| = \infty$  on  $W(x, y)$ .

The theory developed below gives a lucid insight into the nature of the approximation problem and leads to interesting relations between best approximations.

**2. Existence of Best Approximation.** As is well known, every standard problem always has a best approximation. The same is not the case, however, for a general problem. The following theorem gives a necessary and sufficient condition for the existence of a best approximation.

**THEOREM 1.** *A best approximation exists iff there are an  $S \in \mathbf{R}$  and a finite closed interval  $K \subset J$  such that*

$$(8) \quad [e(R, x); x \in I] \subset K$$

for every  $R \in \mathbf{R}$  satisfying  $E(R) \leq E(S)$ .

*Proof. Necessity:* Suppose a best approximation  $R^*$  exists, then

$$S = R^*, \quad K = \left[ \min_{x \in I} e(R^*, x), \max_{x \in I} e(R^*, x) \right]$$

satisfy the condition of the theorem, because  $R^*$  is unique.

*Remark.* Uniqueness will be proved in Theorem 4 below.

*Sufficiency.* It is readily seen that, for every  $R$  satisfying the condition of the

theorem, the set  $[R(x); x \in I]$  is uniformly bounded. The existence will now be shown by a straightforward extension of the arguments of the classical proof [1]. This completes the proof.

By virtue of Theorem 1, we can derive Theorem 2 which can be used to see whether every problem with a fixed pair  $[J, g]$  has a solution or not.

**THEOREM 2.** *Every problem  $[p, q; I, f, w; J, g]$  has a solution iff it belongs to one of the following cases.*

Case 1.  $J = (-\infty, \infty)$ .

Case 2.  $J = [c, \infty), c \leq 0$ .

Case 3.  $J = (-\infty, d], d \geq 0$ .

Case 4.  $J = (c, \infty), c < 0, G(c) \geq G(\infty)$ .

Case 5.  $J = (-\infty, d), d > 0, G(d) \geq G(-\infty)$ .

Remark.  $G(v) = \lim_{y \rightarrow v} |g(y)|$ .

*Proof.* It suffices to show that every problem has a solution for each of the above cases and there is a problem having no solution for every other case.

First, we shall list a choice of  $S$  and  $K$  satisfying the condition of Theorem 1 for each of the five cases. For shortness we shall use the abbreviations:

$$E = E(S), \quad L = [g^{-1}(-E), g^{-1}(E)],$$

$$F' = \min_{x \in I} f(x), \quad F'' = \max_{x \in I} f(x), \quad W' = \min_{x \in I} w(x).$$

Case 1. Subcase 1.1.  $G(-\infty) = G(\infty)$ .  $S = 0, K = L$ .

Subcase 1.2.  $G(-\infty) > G(\infty)$ .  $S = F'' + W'c', K = L$ . ( $G(c') = G(\infty)$ ).

Subcase 1.3.  $G(-\infty) < G(\infty)$ .  $S = F' + W'd'', K = L$ . ( $G(d'') = G(-\infty)$ ).

Case 2.  $S = F'' + W'c, K = [c, g^{-1}(E)]$ .

Case 3.  $S = F' + W'd, K = [g^{-1}(-)e, d]$ .

Case 4.  $S = F'' + W'c/2, K = L$ .

Case 5.  $S = F' + W'd/2, K = L$ .

Next, we shall show a problem having no solution for every other case.

Case 6.  $J = (c, \infty), c < 0, G(c) < G(\infty)$ .

Consider the problem  $[0, 0; [-1, 1], f, 1; J, g]$  with

$$f(x) = cx, \quad -1 \leq x \leq 0,$$

$$= -d''x, \quad 0 \leq x \leq 1,$$

where  $d'' > 0$  is such that  $G(c) = G(d'')$ . This problem has no solution, since, for any admissible function  $R$  which is a positive constant, there is a better approximation  $R/2$ .

Case 7.  $J = (-\infty, d), d > 0, G(d) < G(-\infty)$ .

Consider the problem  $[0, 0; [-1, 1], f, 1; J, g]$  with

$$f(x) = cx, \quad -1 \leq x \leq 0,$$

$$= -d'x, \quad 0 \leq x \leq 1,$$

where  $c' < 0$  is such that  $G(d) = G(c')$ . This problem has no solution, since, for any admissible function  $R$  which is a negative constant, there is a better approximation  $R/2$ .

Case 8.  $J = (c, d), [c, d), (c, d]$  or  $[c, d]$ .

Consider the problem  $[0, 0; [-1, 1], f, 1; J, g]$  with

$$f(x) = (d - c + 1)x.$$

This problem has no solution, since there is no admissible function for this problem.

**3. Characterization and Uniqueness of Best Approximation.** An *ordered point set of length  $n$*  is a point set  $[x_1, x_2, \dots, x_n]$  such that  $x_1 < x_2 < \dots < x_n$ . For any admissible function  $R = P/Q$ , its *degree*  $D(R)$  is defined by

$$(9) \quad \begin{aligned} D(R) &= p + q - \min [p - p^*, q - q^*], \quad \text{for } R \not\equiv 0, \\ &= p, \quad \text{for } R \equiv 0, \end{aligned}$$

where  $p^*$  and  $q^*$  are the exact degrees of  $P$  and  $Q$  respectively.

**THEOREM 3.** *If there is an  $R^* = P^*/Q^* \in \mathbf{R}$  such that  $e(R^*, x)$  alternates in sign on an ordered point set  $[x_1, x_2, \dots, x_n] \subset I$  of length  $n = D(R^*) + 2$ , then, for any  $R \in \mathbf{R}$ , we have*

$$(10) \quad E(R) \geq \min [|g(e(R^*, x_i))|; i = 1, 2, \dots, n].$$

*Proof.* Suppose the inequality (10) fails for an  $R = P/Q \in \mathbf{R}$ , then  $g(e(R, x)) - g(e(R^*, x))$  and consequently

$$e(R, x) - e(R^*, x) = (PQ^* - P^*Q)/wQQ^*$$

alternate in sign on the ordered point set. Since  $wQQ^*$  is of a constant sign, the polynomial  $PQ^* - P^*Q$  has at least  $n - 1$  zeros on  $I$ . From the definition of  $D(R^*)$ , however, its degree is at most  $D(R^*) = n - 2$ . This is a contradiction, proving the theorem.

Now the concept of the alternant will be introduced. An ordered point set is called an *alternant* of an  $R \in \mathbf{R}$ , if  $e(R, x)$  assumes its minimum and maximum on  $I$  alternately on the set and one of the following conditions holds.

$$(11.1) \quad E(R) = -g(e') = g(e''),$$

$$(11.2) \quad E(R) = g(e'') > -g(e'), e' = c,$$

$$(11.3) \quad E(R) = -g(e') > g(e''), e'' = d,$$

where  $e' = \min_{x \in I} e(R, x)$ ,  $e'' = \max_{x \in I} e(R, x)$  and  $c$  and  $d$  are the lower and upper closed ends of  $J$  respectively.

**THEOREM 4.** *An admissible function  $R^*$  is a best approximation iff it has an alternant of length  $D(R^*) + 2$ . If a best approximation exists, it is unique.*

*Proof.* Sufficiency for the case (11.1) follows immediately from Theorem 3. Sufficiency for other cases, uniqueness and necessity can be proved by a simple modification of Achieser's proof of the classical theorem [1].

**4. Relationship Between Best Approximations.** In this rather lengthy section, the relationship between best approximations will be investigated.

LEMMA 1. For any two criterion functions  $g$  and  $h$  defined on the same interval  $J$ , there is a continuous and strictly increasing function  $T$  such that  $g(y) = T(h(y))$  for any  $y \in J$ .

*Proof.* Consider the function  $T(u) = g(h^{-1}(u))$  on the interval with g.l.b.  $\inf_{y \in J} h(y)$  and l.u.b.  $\sup_{y \in J} h(y)$  which is closed, open or semi-open as the case may be. This is the desired function. Since  $g$  and  $h^{-1}$  are both continuous and strictly increasing,  $T$  is also.

Two criterion functions  $g$  and  $h$  are called *equivalent*, if any problems  $[p, q; I, f, w; J, g]$  and  $[p, q; I, f, w; J, h]$  have an identical solution or both have no solution.

THEOREM 5. Criterion functions  $g$  and  $h$  are equivalent iff

$$(12) \quad T(u) = -T(-u)$$

for any  $u \in (-M, M)$  or  $[-M, M]$  as the case may be, where

$$M = \min \left[ -\inf_{y \in J} h(y), \sup_{y \in J} h(y) \right].$$

*Proof.* Sufficiency is almost evident, since it is readily seen that a solution corresponding to  $h$  is also a solution corresponding to  $g$  and vice versa. In order to prove the necessity, we consider the problems such that

$$p = 0, \quad q = 0, \quad I = [-1, 1], \quad w = 1.$$

For any  $u$  satisfying the condition of the theorem, we put

$$\begin{aligned} f(x) &= h^{-1}(-u)x, & -1 \leq x \leq 0, \\ &= -h^{-1}(u)x, & 0 \leq x \leq 1. \end{aligned}$$

Clearly  $R^* = 0$  is the solution for  $h$ . By assumption, it is also the solution for  $g$ . It follows from (11.1) that

$$-g(h^{-1}(-u)) = g(h^{-1}(u)),$$

i.e.,  $-T(-u) = T(u)$ .

COROLLARY. If  $0$  is a closed end of  $J$ , then all the criterion functions defined on  $J$  are equivalent.

Now we shall examine the relationship between solutions corresponding to non-equivalent criterion functions. Although we can say hardly anything in general, strikingly simple relationships exist in the following two important cases. The one is the polynomial ( $q = 0$ ) or rational ( $p \geq q$ ) approximation problem with respect to *absolute error* ( $w = 1$ ), and the other is the rational approximation problem for a positive function  $f$  with respect to *relative error* ( $w = f$ ).

THEOREM 6. If  $p \geq q$  and  $w = 1$ , then the difference of solutions of any two problems differing only in criterion functions, when both solutions exist, is a constant.

*Proof.* There is no loss in generality in letting one problem be a standard problem. Therefore, let  $S$  be a solution of a standard problem,  $g$  be an arbitrary criterion function and  $J$  be its definition interval.

Consider the auxiliary function

$$(13) \quad H(z) = g(z - e) + g(z + e),$$

where

$$e = \max_{x \in I} e(S, x) = -\min_{x \in I} e(S, x).$$

*Case 1.*  $H(z)$  is not defined at all. This occurs when  $J$  is finite and

$$(14) \quad d - c \leq 2e,$$

where  $c$  and  $d$  are the g.l.b. and l.u.b. of  $J$  respectively. The equality in (14) may hold only if  $J$  is not closed. In this case there is no admissible function and a fortiori no solution for  $g$ . In order to see this, assume, on the contrary, that an  $R$  is admissible for  $g$ , then we have

$$(15) \quad c \leq e', \quad e'' \leq d,$$

where

$$e' = \min_{x \in I} e(R, x), \quad e'' = \max_{x \in I} e(R, x).$$

Define  $R^* = R - (e' + e'')/2$ , then  $R^*$  is admissible for the standard problem because  $p \geq q$ . Furthermore

$$\max_{x \in I} |e(R^*, x)| = (e'' - e')/2 < e$$

by virtue of (14), (15) and the fact that equalities in (14) and (15) never hold simultaneously. But this is a contradiction because  $S$  is the best approximation.

When  $H(z)$  is defined, it is defined on an interval  $K$  with g.l.b.  $c + e$  and l.u.b.  $d - e$ . Since  $H(z)$  is continuous and strictly increasing, it has at most one zero on  $K$ .

*Case 2.*  $H(z)$  has a zero  $z^*$  on  $K$ . Define  $R = S + z^*$ . Since  $p \geq q$  and

$$e(R, x) = e(S, x) + z^* \in [z^* - e, z^* + e] \subset J,$$

$R$  is admissible for  $g$ . Moreover, we have

$$E(R) = -g(e') = g(e'')$$

because  $e' = z^* - e$ ,  $e'' = z^* + e$ ,  $H(z^*) = g(e') + g(e'') = 0$ . Therefore, from (11.1),  $R$  is a solution for  $g$ .

*Case 3.*  $H(z)$  is positive on  $K$ . In this case  $c$  should be finite.

*Subcase 3.1.*  $c$  is the closed end of  $J$ . Define  $R = S + c + e$ . Since  $p \geq q$  and

$$e(R, x) = e(S, x) + c + e \in [c, c + 2e] \subset J,$$

$R$  is admissible for  $g$ . Moreover, we have

$$E(R) = g(e'') > -g(e')$$

because  $e' = c$ ,  $e'' = c + 2e$ ,

$$H(c + e) = g(c) + g(c + 2e) > 0.$$

Therefore, from (11.2),  $R$  is a solution for  $g$ .

*Subcase 3.2.*  $c$  is the open end of  $J$ . In this case no solution exists for  $g$ . In order to see this, assume that an  $R$  is a solution for  $g$ , then we have

$$(16) \quad c < e', \quad g(e') + g(e'') \leq 0$$

from (11.1) and (11.3). Since

$$H(c + e + 0) = g(c + 0) + g(c + 2e) \geq 0,$$

it follows from (16) that  $(e'' - e')/2 < e$ . From this, a contradiction can be reached by the same argument as was used in Case 1.

*Case 4.*  $H(z)$  is negative on  $K$ . This case can be treated quite similarly to Case 3. The l.u.b.  $d$  of  $J$  should be finite. If  $d$  is the closed end, then  $R = S + d - e$  is a solution for  $g$ . If, on the other hand,  $d$  is the open end, then there is no solution for  $g$ . This completes the proof.

LEMMA 2. *If  $w = f > 0$  on  $I$ , then every solution is positive on  $I$ .*

*Proof.* Assume, on the contrary, that there is a solution  $R^*$  such that  $\min_{x \in I} R^*(x) \leq 0$ . Then, since

$$\min_{x \in I} e(R^*, x) = \min_{x \in I} [R^*(x)/f(x) - 1] \leq -1,$$

we have  $E(R^*) \geq -g(-1)$ . Now consider  $R = F' = \min_{x \in I} f(x)$ . Then, since

$$0 \geq e(R, x) = F'/f(x) - 1 > -1,$$

$R$  is admissible and satisfies  $E(R) < -g(-1)$ . This means a contradiction  $E(R) < E(R^*)$ .

THEOREM 7. *If  $w = f > 0$  on  $I$ , then the ratio of solutions of any two problems differing only in criterion functions, when both solutions exist, is a positive constant.*

*Proof.* The proof is similar to that of Theorem 6, and we shall use the same symbols and abbreviations.

Defining  $S, g$  and  $J$  as before, we consider the function

$$(17) \quad H(z) = g((1 - e)z - 1) + g((1 + e)z - 1).$$

We first note that

$$(18) \quad 1 - e > 0$$

from Lemma 2.

*Case 1.*  $H(z)$  is not defined at all. This occurs when  $J$  is finite and

$$(19) \quad (1 + c)/(1 + d) \geq (1 - e)/(1 + e),$$

where the equality may hold only if  $J$  is not closed. In this case, there is no admissible function and a fortiori no solution for  $g$ . In order to see this, assume that an  $R$  is admissible for  $g$ , then we have  $c \leq e', e'' \leq d$  or

$$(20) \quad (1 + e')/(1 + e'') \geq (1 + c)/(1 + d).$$

Since the equalities in (19) and (20) never hold at the same time, we obtain

$$(21) \quad A = (1 + e')/(1 + e'') > (1 - e)/(1 + e) = B > 0.$$

Now define  $R^* = 2R/(2 + e' + e'')$ , then we have

$$\begin{aligned} \max_{x \in I} |e(R^*, x)| &= (e'' - e')/(2 + e' + e'') \\ &= (1 - A)/(1 + A) < (1 - B)/(1 + B) = e \end{aligned}$$

from (21). But this is a contradiction.

When  $H(z)$  is defined, it is defined on an interval  $K$  with

$$\text{g.l.b.} = \max [(1 + c)/(1 + e), (1 + c)/(1 - e)]$$

and

$$\text{l.u.b.} = (1 + d)/(1 + e).$$

Since  $H(z)$  is continuous and strictly increasing, it has at most one zero on  $K$ .

*Case 2.*  $H(z)$  has a zero  $z^*$  on  $K$ . Define  $R = z^*S$ . Since

$$0 < 1/(1 + e) < z^* < 1/(1 - e),$$

we have  $e(R, x) = (1 + e(S, x))z^* - 1 \in [(1 - e)z^* - 1, (1 + e)z^* - 1] \subset J$ . Hence  $R$  is admissible for  $g$ . Furthermore, we have

$$e' = (1 - e)z^* - 1, \quad e'' = (1 + e)z^* - 1,$$

$$H(z^*) = g(e') + g(e'') = 0.$$

Therefore  $E(R) = -g(e') = g(e'')$ . This implies from (11.1) that  $R$  is a solution for  $g$ .

*Case 3.*  $H(z)$  is positive on  $K$ . In this case  $1 + c > 0$ , since otherwise the g.l.b. of  $K$  would be nonpositive and  $H(z)$  would be negative in a neighborhood of  $z = 0$ .

*Subcase 3.1.*  $c$  is the closed end of  $J$ . Define  $R$  by

$$R = (1 + c)S/(1 - e),$$

then, since

$$e' = c, H((1 + c)/(1 - e)) = g(e') + g(e'') > 0,$$

we have  $E(R) = g(e'') > -g(e')$ . This implies from (11.2) that  $R$  is a solution for  $g$ .

*Subcase 3.2.*  $c$  is the open end of  $J$ . In this case no solution exists for  $g$ . In order to see this, assume that an  $R$  is a solution for  $g$ , then we have

$$(22) \quad c < e', g(e') + g(e'') \leq 0$$

from (11.1) and (11.3). On the other hand, we have

$$H((1 + c)/(1 - e) + 0) = g(c + 0) + g((1 + c)(1 + e)/(1 - e) - 1) \geq 0,$$

which, combined with (22), yields

$$(1 + e')/(1 + e'') > (1 - e)/(1 + e).$$

From here, a contradiction can be reached by the same argument as was used in Case 1.

*Case 4.*  $H(z)$  is negative on  $K$ . This case can be treated quite similarly as Case 3. The l.u.b.  $d$  of  $J$  should be finite. If  $d$  is the closed end, then  $R = (1 + d)S/(1 + e)$  is a solution for  $g$ . If, on the other hand,  $d$  is the open end, then there is no solution for  $g$ . This completes the proof.

**5. Applications.** In this section, two examples will be worked out to illustrate the applicability of the present theory.

*Example 1. One-sided polynomial approximation.* Consider the problem  $[n, 0;$



$I, f, 1; [0, \infty), g]$ . This problem always has a solution because it belongs to Case 2 of Theorem 2. For the purpose of obtaining a solution, we may assume  $g(y) = y$ , since all the criterion functions are equivalent in this case by the Corollary of Theorem 5. Let  $S$  denote the solution of the standard version of the problem. In accordance with Theorem 6, we consider the function

$$H(z) = g(z - e) + g(z + e) = 2z$$

which is defined for  $z \geq e$ . Since  $H(z)$  is positive and  $c = 0$  is the closed end, the solution for  $g$  is given by  $R = S + e$ . Clearly, we have  $E(R) = g(2e)$ .

Thus, for the special case of  $I = [-1, 1], f(x) = x^{n+1}$ , we obtain  $R = x^{n+1} - 2^{-n}(T_{n+1}(x) - 1), E(R) = g(2^{1-n})$ .

*Example 2. Starting rational approximation of Newton iteration for  $x^{1/n}$ .* The standard method for calculating  $x^{1/n}$  by a computer is to apply Newton iteration

$$(23) \quad R_{i+1} = ((n - 1)R_i + xR_i^{1-n})/n = N(R_i) = N^{i+1}(R_0)$$

to an appropriate starting approximation  $R_0$ . Let us assume as usual that  $x$  is confined in a positive finite closed interval  $I$ . A reasonable first choice for the starting approximation is to take  $R_0 = S$ , where  $S$  is the solution of the standard problem

$$(24) \quad [p, q; I, x^{1/n}, x^{1/m}; (-\infty, \infty), y].$$

There is, however, another more reasonable choice, i.e., Moursund's choice to take as  $R_0$  an admissible function  $R$  of the problem (24) so that  $R_i$  calculated from  $R_0 = R$  by (23) may be optimal in the sense that

$$(25) \quad \max_{x \in I} |e(R_i, x)| = \min \left[ \max_{x \in I} |e(R_i, x)|; R_i = N^i(R), R \in \mathbf{R} \right].$$

As Moursund has shown [4] in the case of  $x^{1/2}$ , the  $R$  satisfying (25) for  $i = 1$  satisfies (25) for every  $i$ .

Now it will be observed that the problem just posed can be formulated as a generalized problem

$$(26) \quad [p, q; I, x^{1/n}, x^{1/m}; (-1, \infty), g(y)]$$

where

$$(27) \quad g(y) = \operatorname{sgn}(y) \left( \frac{1}{n} \left( (n - 1)(1 + y) + \frac{1}{(1 + y)^{n-1}} \right) - 1 \right),$$

since we may assume  $R > 0$  or  $e(R, x) > -1$  by Lemma 2. The solution  $R$  of this problem is the desired starting approximation. Since we have

$$\lim_{y \rightarrow -1} |g(y)| = \lim_{y \rightarrow \infty} |g(y)| = \infty,$$

it belongs to Case 4 of Theorem 2, and therefore it always has a solution. From Theorem 7, the solution is given by  $R = z^*S$ , where  $z^*$  is a zero of the function

$$H(z) = g((1 - e)z - 1) + g((1 + e)z - 1).$$

Substituting (27), we find after an elementary calculation

$$(28) \quad z^* = \left( \frac{1}{2(n-1)e} \left( \frac{1}{(1-e)^{n-1}} - \frac{1}{(1+e)^{n-1}} \right) \right)^{1/n}.$$

In connection with the problem (26), we consider another problem

$$(29) \quad [p, q; I, x^{1/n}, x^{1/n}; (-1, \infty), h(y)],$$

with

$$(30) \quad y = h^{-1}(u) = \left( \frac{\sinh(n-1)u}{(n-1)\sinh u} \right)^{1/n} \exp u - 1$$

defined for  $u \in (-\infty, \infty)$ .

Substitution of (30) into (27) yields

$$T(u) = g(h^{-1}(u)) = \operatorname{sgn}(u) \left( \left( \frac{\sinh(n-1)u}{(n-1)\sinh u} \right)^{1/n} \frac{(n-1)\sinh u}{n \sinh(n-1)u} - 1 \right).$$

Since  $T(u)$  is an odd function, we conclude from Theorem 5 that  $g$  and  $h$  are equivalent, i.e., problems (26) and (29) have an identical solution.

*Case 1. Square root.* Putting  $n = 2$ , we obtain

$$z^* = (1 - e^2)^{-1/2},$$

$$h(y) = \log(1 + y).$$

The finding that the optimal starting approximation  $R$  in Moursund's sense is a simple multiple  $R = z^*S$  of the standard Chebyshev solution  $S$ , and can be obtained by solving (29) with  $h(y) = \log(1 + y)$ , is indeed the motive for the author to develop the present theory. It is interesting to notice that several authors including Chebyshev himself [5], [6] have obtained  $R$  correctly for some moderate values of  $p$  and  $q$  by solving not (24) but (29) for a technical reason to facilitate the analysis.

During the revision of the manuscript, the author was informed by the referee of the works of Sterbenz and Fike [7], and King and Phillips [8], where similar results are given.

*Case 2. Cubic root.* Putting  $n = 3$ , we obtain

$$z^* = (1 - e^2)^{-2/3},$$

$$h(y) = \frac{1}{2} \log \frac{(1 + 8(1 + y)^3)^{1/2} - 1}{2}.$$

The form of  $h(y)$  shows that the adoption of  $h(y) = \log(1 + y)$  is not successful in general.

**Acknowledgement.** The author is grateful to the referee for a number of helpful comments.

Faculty of Engineering  
Nagoya University  
Furo-cho, Chikusa-ku  
Nagoya, Japan

1. N. I. ACHESER, *Theory of Approximation*, OGIZ, Moscow, 1947; English transl., Ungar, New York, 1956. MR 10, 33; MR 20 #1872.
2. C. B. DUNHAM, "Transformed rational Chebyshev approximation," *Numer. Math.*, v. 10, 1967, pp. 147-152. MR 36 #585.
3. D. G. MOURSUND, "Chebyshev approximation using a generalized weight function," *SIAM J. Numer. Anal.*, v. 3, 1966, pp. 435-450. MR 34 #4771.
4. D. G. MOURSUND, "Optimal starting values for Newton-Raphson calculation of  $x^{1/2}$ ," *Comm. ACM*, v. 10, 1967, pp. 430-432.
5. W. J. CODY, "Double-precision square root for the CDC-3600," *Comm. ACM*, v. 7, 1964, pp. 715-718.
6. P. L. CHEBYSHEV, "Sur les expressions approchées de la racine carrée d'une variable par des fractions simples," in *Oeuvres*. Vol. 2, Chelsea, New York, pp. 542-558.
7. P. H. STERBENZ & C. T. FIKE, "Optimal starting approximations for Newton's method," *Math. Comp.*, v. 23, 1969, 313-318.
8. R. F. KING & D. L. PHILLIPS, "The logarithmic error and Newton's method for the square root," *Comm. ACM*, v. 12, 1969, pp. 87-88.