Application of Quantifier Elimination to Solotareff"s Approximation Problem

George E. Collins *
Research Institute for Symbolic Computation
Johannes Kepler University
A-4040 Linz, Austria
e-mail:
gcollins@risc.uni-linz.ac.at

March 10, 1997

Abstract

Solotareff's approximation problem is that of obtaining the best uniform approximation on the interval [-1,+1] of a real polynomial of degree n by one of degree $n-2$ or less. Without loss of generalitywe take the approximated polynomial to be $x^n + rx^{n-1}$, $r \geq 0$. We treat r as a parameter and seek to compute the coefficients of the best approximations, for small fixed values of n, as piecewise algebraic functions of r. We succeed only for $n \leq 4$, but also find that we can easily compute the coefficients for $n = 5$ for fixed values of r. The results serve to display the capabilities and limitations of our quantifier elimination program **qepcad**. We also prove that the coefficients of the best approximation, for any n, are continuous functions of r.

1 Introduction

In 1933 E. I. Solotareff (see [Achieser 56]) posed the problem of finding the best approximation, relative to the uniform norm on the interval [-1,+1], of a real polynomial of degree n by one of degree $n-2$ or less. This is easily seen to be equivalent to obtaining the best approximation to a polynomial $x^n + rx^{n-1}$, r a non-negative real number. This may be regarded as a generalization of the problem of obtaining the best approximation to a polynomial of degree n by one of degree $n-1$ or less, a problem solved by the Chebyshev polynomials. The existence of a quantifier elimination method for the theory of real-closed fields

*Supported by the Austrian Science Foundation (Grant P8572-PHY).

implies that the coefficients of the best approximation for Solotareff's problem are, for any fixed $n \geq 2$, piecewise algebraic functions of r. In the following we explore the extent to which our program, **qepcad**, for quantifier elimination based on partial cylindrical algebraic decomposition, is able to compute these functions.

The Solotareff problem is just one example of the many possible applications of an efficient quantifier elimination program. Our purpose in this paper, apart from the interest inherent in our results, is to indicate by example the current capabilities and limitations of **qepcad** and to explain some of its workings.

In Section 2 we use a characterization of the best approximation by Solotareff to obtain a desirable formulation of the problem as a quantified formula. In Section 3 we present complete solutions obtained for degrees 2, 3 and 4, and show that we can easily obtain solutions for particular values of r for degree 5. In Section 4 we give an introduction to the **qepcad** program and discuss the user interaction and program behavior for this application. We also prove that the coefficients of the best Solotareff approximations are continuous functions of r. This fact was used in obtaining the solution for $n = 4$, but does not appear to be a consequence of known theorems on approximation theory.

2 Formulation of the Problem

The definition of best uniform approximation by a polynomial of degree $n-2$ or less is a condition that can be expressed within the elementary theory of real-closed fields (ETRCF) for any fixed value of n since it expresses minimality among the class of all real polynomials of degree $n-2$ or less, and this class can be parameterized by the $n-1$ coefficients of any such polynomial. But we also must express the condition that one polynomial of degree n, the difference of $x^n + rx^{n-1}$ and a competing approximation, has a uniform norm no larger than that of another. This can also be done. Let $\| A \|$ denote the uniform norm of A on $[-1, +1]$. If $A(x)$ and $B(x)$ are two polynomials then $\| A \| \leq \| B \|$ is equivalent to

$$[-1 \leq y \leq +1 \wedge \mid A(x) \mid \leq \mid B(y) \mid]]. \tag{1}$$

The absolute value operator is not an accepted part of the ETRCF language but it can always be eliminated. $\mid x \mid \leq \mid y \mid$ is equivalent to $x^2 \leq y^2$ or to $[y \geq 0 \wedge -y \leq x \leq y] \vee [y \leq 0 \wedge y \leq x \leq -y]$.

To take the case $n = 3$ as an example, if we let $ax + b$ be the best approximation and $a'x + b'$ any competing approximation, we can express this by the quantified formula

$$\begin{aligned}(\forall a')(\forall b')(\forall x)[-1 \leq x \leq +1 \Rightarrow (\exists y)[-1 \leq y \leq +1 \wedge \\ \mid x^3 + rx^2 - ax - b \mid \leq \mid y^3 + ry^2 - a'y - b' \mid]].\end{aligned} \tag{2}$$

This formula has 3 free variables, r, a and b, and four bound variables, x, y, a' and b'. One can easily see that a formulation of this kind for arbitrary n will have n free variables and $n+1$ bound variables. To increase n by 1 we must add one free coefficient variable and one bound coefficient variable. However, we can reduce the number of variables, thereby enhancing the feasibility of eliminating the quantifiers, by utilizing some known theorems relating to the problem. Except where otherwise indicated these theorems can be found in [Achieser 56].

The best approximation is known to be unique, and it is characterized by the existence of n points, $-1 \leq u_1 < u_2 < \ldots u_n \leq +1$ such that, if $A_n(x)$ is the best approximation and $E_n(x)$ is the difference $x^n + rx^{n-1} - A_n(x)$, then $\| E_n \| = | E_n(u_1) |$ and $E_n(u_i) = (-1)^{i+1} E_n(u_1)$ for $2 \leq i \leq n$.

We will henceforth assume that $r \geq 0$, for if $r \leq 0$ and $A_n(x)$ is the best aproximation to $x^n - rx^{n-1}$ then $A_n(-x)$ is the best appproximation to $x^n + rx^{n-1}$. Achieser shows that

$$E_n(x) = (-1)^n 2^{-n+1} (1 + r/n)^n T_n((-x - r/n)/(1 + r/n)) \tag{3}$$

for $0 \leq r \leq S_n = n(tan(\pi/(2n)))^2$, where $T_n(x)$ is the Chebyshev polynomial of degree n, $T_n(x) = \cos(n \arccos x)$. This is an easily computable function, from which one can read off the coefficients of $A_n(x)$, and they will be polynomial functions of r of degree n or less. So quantifier elimination is needed only for $r > S_n$.

Each of the $n-2$ points $u_2, u_3, \ldots, u_{n-1}$ must clearly be a zero of of the derivative $E_n'(x)$. Elaboration of an argument in [Achieser 56] leads to the conclusion that $E_n(1) > 0$. These properties, in fact, characterize $E(x)$ and we arrive at the following quantified formula of ETRCF.

$$(\exists u_2) \ldots (\exists u_{n-1})[r > S_n \wedge -1 < u_2 < \ldots < u_{n-1} < 1 \wedge E_n(1) > 0$$
$$\wedge E_n(-1) = (-1)^{n+1} E_n(+1) \wedge \bigwedge_{i=2}^{n-1} [E_n(u_i) = (-1)^{n-i} E_n(1) \wedge E_n'(u_i) = 0]] \tag{4}$$

This formula has the same n free variables but only $n-2$ bound variables, a reduction of three. It has another very important advantage, namely the occurence in it of many equations. These equations are *equational constraints*, equations occurring in, and logically implied by, the quantifier-free part of the formula, to which the quantifiers are applied, which we call the *matrix*. We will explain in the following section how equational constraints are utilized in a recently completed version of our quantifier elimination program, which is called **qepcad**. Formula 4 contains $2n-3$ equational constraints, just one less than the number of variables. We might say that it is *almost* the problem of solving a system of polynomial equations; it fails being so only because of the imposed inequalities. We can simplify the formula slightly more by using the equation

$E_n(-1) = (-1)^{n+1} E_n(+1)$ to solve for one of the coefficients. Upon evaluation of $E_n(x)$ at $x = 1$ and $x = -1$ we obtain the linear equation $\sum_{i \text{ odd}} a_i = 1$, where $A_n(x) = \sum_{i=0}^{n-2} a_i x^i$. We use this equation to solve for a_0 if n is even, for a_1 if n is odd, in terms of the other a_i and substitute the result into the equations involving E_n. We are left with $2n-4$ equations in $2n-3$ variables.

3 Solutions

We would now like to solve the Solotareff problem for $n = 2, 3, 4, \ldots$. We take this to mean obtaining a constructive description of each coefficient of the best approximation, A_n, as a function of r in the form of a simple quantifier- free formula containing as free variables only r and the coefficient.

For $n = 2$, $S_2 = 2$ and Formula 3 produces

$$a = 1/2 + r/2 - r^2/8 \tag{5}$$

for $0 \le r \le 2$. When $n = 2$ there are no bound variables in Formula 4; it consists of only the equation $E_2(-1) = -E_2(1)$, which evaluates to $1-r-a = -1-r+a$. We conclude that $a = 1$ for $r > 2$. Of course we could also see this by simple reasoning.

For $n = 3$, let $A_3(x) = ax + b$. From the equation $E_3(1) = E_3(-1)$ we obtain $a = 1$. We substitute 1 for a, let $u_1 = u$, and compute $S_3 = 1$. Formula 4 becomes

$$(\exists u)[r > 1 \wedge -1 < u \wedge u < 1 \wedge 3u^2 + 2ru - 1 = 0 \wedge u^3 + ru^2 - u + r - 2b = 0] \tag{6}$$

qepcad produces the solution to this problem in 133 milliseconds. All times reported in this paper are based on the use of a Decstation 5000/240 having a 40 MHz R3400 risc processor. We find that that b is the lesser of the two real roots of the polynomial $P_3(r,b) = 27b^2 - 2r^3b - 36rb + r^4 + 11r^2 - 1$. Thus

$$[27b^2 - 2r^3b - 36rb + r^4 + 11r^2 - 1 = 0 \wedge 27b - r^3 - 18r < 0] \tag{7}$$

is a quantifier-free formula for the solution.

For $r \le 1$ Formula 3 yields the solutions

$$\begin{aligned} a = &\ 3/4 + 1/2\, r - 1/4\, r^2 \\ b = &\ 1/4\, r + 1/6\, r^2 - 1/108\, r^3. \end{aligned} \tag{8}$$

We can also solve for u as a function of r for $r > 1$. All we have to do is to existentially quantify b instead of u. From the input formula we see that u must be one of the two real roots of the equation $3u^2 + 2ru - 1 = 0$ and **qepcad** tells that u is the larger of the two roots. It is not difficult to see, or to prove by quantifier elimination, that as r increases from 1 to $+\infty$, $u = u_2$ decreases monotonically from $1/3$ to 0 and u_1 decreases monotonically from -1 to $-\infty$.

Furthermore we can compute the maximum error, $e_3 = \| E_3 \|$. $e_3 = E_3(1) = 1 + r - a - b = r - b$ so $b = r - e_3$ and $P_3(r, r - e_3) = 0$. By substitution $P_3(r, r - e_3) = 27e_3^2 + 2r^3 e_3 - 18re_3 - r^4 + 2r^2 - 1$. For $r > 1$ this polynomial has one positive root and one negative root; e_3 is the positive root. For $r \leq 1$, $e_3 = r - b = 3/4\, r - 1/6\, r + 1/108$.

Now consider the problem for $n = 4$. Let us first compute S_4. $S_4 = 4(\tan \pi/8)^2$. Let $a = \tan \pi/8$. Substituting $\pi/4$ for x in the double angle formula $\tan 2x = 2 \tan x/(1 - \tan^2 x)$, we obtain $1 = 2a/(1 - a^2)$, hence $a^2 + 2a - 1 = 0$. Let $b = S_4 = 4a^2$. Computing the resultant of $4a^2 - b$ and $a^2 + 2a - 1$ we deduce that $b^2 - 24b + 16 = 0$. Since $\tan \pi/8 < 1$, S_4 must be the lesser of the two real roots of this equation, $12 - 8\sqrt{2} = 0.6863-$. Consequently we can express $r > S_4$ by the formula $r^2 - 24r + 16 < 0 \vee [r > 1 \wedge r^2 - 24r + 16 \geq 0]$.

Let $A_4(x) = ax^2 + bx + c$. From $E_4(-1) = -E_4(+1)$ we obtain $c = 1 - a$. Let $u_1 = u$ and $u_2 = v$. $E_4(1) = 1 + r - a - b - (1 - a) = r - b$. Consequently the formula $E_4(u) = -E_4(1)$ simplifies to $u^4 + ru^3 - au^2 - bu - r + b + a - 1 = 0$. This factors into $(u^3 + ru^2 + u^2 - au + ru - b - a + r + 1)(u - 1) = 0$. Since $u < 1$ we can omit the factor $u - 1$. Similarly we can replace $E_4(v) = E_4(1)$ by $v^3 + rv^2 - v^2 - av - rv + v - b + a + r - 1 = 0$.

Putting this all together, we get the input formula

$$
\begin{aligned}
(\exists u)(\exists v)[[r^2 - 24r + 16 < 0 \vee [r > 1 \wedge r^2 - 24r + 16 >= 0]] \\
\wedge -1 < u \wedge u < v \wedge v < 1 \wedge r - b > 0 \\
\wedge u^3 + ru^2 + u^2 - au + ru + u - b - a + r + 1 = 0 \\
\wedge v^3 + rv^2 - v^2 - av - rv + v - b + a + r - 1 = 0 \\
\wedge 4u^3 + 3ru^2 - 2au - b = 0 \wedge 4v^3 + 3rv^2 - 2av - b = 0]
\end{aligned} \tag{9}
$$

We first obtain a as a function of r, $r > S_4$, by existentially quantifying Formula 9 with respect to b. **qepcad** produces the answer after 51.9 seconds of computation time. The polynomial

$$
\begin{aligned}
P_4(r, a) = 324a^4 + 324r^2a^3 - 2016a^3 + 108r^4a^2 - 1128r^2a^2 + 4576a^2 \\
+12r^6a - 224r^4a + 1392r^2a - 4480a - 15r^6 + 112r^4 - 608r^2 + 1600
\end{aligned} \tag{10}
$$

has two real roots for every $r > S_4$ and a is the larger of the two.

Quantifying a instead of b and applying **qepcad** again we obtain the solution for b as a function of r for $r > S_4$. This time the answer is a bit more complicated. The polynomial

$$
Q_4(r, b) = 78732b^4 + 8748r^3b^3 - 291600rb^3 + 324r^6b^2 - 18684r^4b^2
$$

$$+403272r^2b^2 - 1024b^2 + 4r^9b - 616r^7b + 12692r^5b$$
$$-246800r^3b + 2048rb - 3r^{10} + 280r^8 - 2708r^6 + 56332r^4 - 1024r^2 \quad (11)$$

has two real roots for every $r > S_4$ but the two roots coincide for $r = \alpha = 6.07286+$, the unique root of $243x^8 - 8532x^6 - 15920x^4 + 2624x^2 - 1024$ between 6 and 8. b is the smaller of the two roots for $S_4 < r < \alpha$, the larger root for $r > \alpha$. **qepcad** produces the answer in 53.8 seconds.

For $r \leq S_4$ application of 3 produces the formulas

$$\begin{aligned} a = & \ 1 + 1/2\, r - 5/16\, r^2 \\ b = & \ 1/2\, r + 1/4\, r^2 - 1/32\, r^3 \\ c = & \ -1/8 - 1/8\, r + 1/64\, r^2 + 3/128\, r^3 - 1/2048\, r^4 \end{aligned} \quad (12)$$

For $n = 5$ we first compute S_5. Applying the multiple-angle formula $\cos 5x = 16\cos^5 x - 20\cos^3 x + 5\cos x$ with $x = \pi/10$ we deduce that $\cos\pi/10$ is a positive real root of $16x^5 - 20x^3 + 5x = x(16x^4 - 20x^2 + 5)$. This polynomial has two positive real roots, 0.588- and 0.951+. Clearly $\cos\pi/10$ is the larger of the two. By solving the quantifier elimination problem $(\exists y)[16y^4 - 20y^2 + 5 = 0 \wedge y > 3/4 \wedge x^2 + y^2 = 1 \wedge x > 0]$ we find that $\sin\pi/10$ is the positive root of $4x^2 + 2x - 1$. Then the solution to the quantifier elimination problem $(\exists y)(\exists z)[y > 0 \wedge 4y^2 + 2y - 1 = 0 \wedge z > 3/4 \wedge 16z^4 - 20z^2 + 5 = 0 \wedge xz = y]$ reveals that $\tan\pi/10$ is the root of $5x^4 - 10x^2 + 1$ between 0 and 1. Finally the quantifier elimination problem $(\exists y)[5y^4 - 10y^2 + 1 = 0 \wedge y > 0 \wedge y < 1 \wedge x = 5y^2]$ yields the solution $S_5 = 0.5278640450+$, the unique root of $x^2 - 10x + 5$ between 1/2 and 1.

For $r < S_5$ Achieser's formula produces the results

$$\begin{aligned} a = & \ 5/4 + 1/2\, r - 7/20\, r^2 \\ b = & \ 3/4\, r + 3/10\, r^2 - 1/20\, r^3 \\ c = & \ -5/16 - 1/4\, r + 3/40\, r^2 + 1/20\, r^3 - 1/400\, r^4 \\ d = & \ -1/16\, r - 1/20\, r^2 - 1/200\, r^3 + 1/500\, r^4 - 1/50000\, r^5. \end{aligned} \quad (13)$$

The case $n = 5$ and $r > S_5$ proves to be too difficult for the current version of our program. Perhaps after further innovations we will be able to solve it. As we went from $n = 3$ to $n = 4$ the computation time jumped from 133 milliseconds to 51.9 seconds, a factor of about 400. So it is not surprising that $n = 5$ would be very difficult. However, we have been able to solve the particular case $r = 1$ for $n = 5$. Let $A_5(x) = ax^3 + bx^2 + cx + d$. The equation $E_5(-1) = E_5(+1)$ implies that $a + c = 1$, and we substitute $1 - a$ for c. For $r = 1$, a, b and d are uniquely determined real numbers, and then it follows that $u = u_1$, $v = u_2$ and $w = u_3$ are also uniquely determined. Consequently we can solve for $a, b, d, u, v,$ and w without quantifier elimination; we input the formula

$$
\begin{aligned}
[1-b-d>0 \wedge -1<u \wedge u<v \wedge v<w \wedge w<1 \\
\wedge 5u^4+4u^3-3au^2-2bu-(1-a)=0 \\
\wedge 5v^4+4v^3-3av^2-2bv-(1-a)=0 \\
\wedge 5w^4+4w^3-3aw^2-2bw-(1-a)=0 \\
\wedge u^5+u^4-au^3-bu^2-(1-a)u-d=-1+b+d \\
\wedge v^5+v^4-av^3-bv^2-(1-a)v-d=+1-b-d \\
\wedge w^5+w^4-aw^3-bw^2-(1-a)w-d=-1+b+d]
\end{aligned}
\tag{14}
$$

A solution is obtained in 21.1 seconds. Following are the values of the six variables and their minimal polynomials.

$$
\begin{aligned}
a &= 1.4581775889+, & & 675x^4-3240x^3+6048x^2-5120x+1600 \\
b &= 0.9557598789-, & & 2460375x^4-10060200x^3+15437952x^2 \\
& & & -10535680x+2697536 \\
d &= -0.1065834340+, & & 192216796875x^4+102160237500x^3+23548707018x^2 \\
& & & +1571039500x-117814 \\
u &= -0.5431806965+, & & 16875x^8+27000x^7-24300x-49880x^5 \\
& & & +10062x^4+30152x^3+108x^2-5832x-729 \\
v &= 0.1803173009+, & & 135x^4-18x^2+8x-1 \\
w &= 0.7730740989-, & & 16875x^8+27000x^7-24300x^6-49880x^5 \\
& & & +10062x^4+30152x^3+108x^2-5832x-729
\end{aligned}
\tag{15}
$$

4 Solution Methods

The results we have obtained for $n = 4$ and $n = 5$ were not obtained by straightforward application of **qepcad**. Although straightforward application would, in principle, suffice, computation times would be impractically large. **qepcad** is an interactive program whose behavior can be modified by various commands, given either in advance or during the course of the computation, that may be chosen by the user, on the basis of information that the program can be caused to display, and on the basis of his knowledge of the particular problem at hand. In the following we describe some of the features of the program that were used in an important way to obtain the solutions of the previous section.

A **qepcad** computation has four sequential phases:

1. Normalization

2. Projection

3. Partial CAD construction

4. Solution formula construction

In the normalization phase, the input formula is read and then rewritten in a standard form. The input formula must be in *prenex* form, consisting of a quantifier-free *matrix* to which a string of quantifiers is applied. In the normalization phase each atomic formula is rewritten so that the right-hand side of the relation symbol is zero. The resulting polynomials on the left-hand sides are then factored into irreducibles and the formula is rewritten using just these irreducible polynomials. A linear ordering of the variables in the formula must be specified by the user that is consistent with the order of the variables in the quantifier prenex.

The projection phase begins with the set of irreducible polynomials obtained from the normalization phase. Let r be the number of variables in the input formula. The r-variate irreducible polynomials are "projected", producing a set of polynomials in $r-1$ variables, from which the rth variable has been eliminated. The resulting projection polynomials are then factored into irreducible polynomials that are added to the existing set of irreducible polynomials. Then the $r-1$-variate polynomials are projected and factored. After $r-1$ projections and factorizations the projection phase is completed, having produced, for $1 \leq i \leq r$, a set of i-variate irreducible polynomials, called *projection factors*.

The next phase proceeds with constructing a cylindrical algebraic decomposition of r-dimensional real space consisting of a finite number of connected sets, called *cells*, in each of which the matrix of the input formula is invariant in truth value. Note that this can be ensured by making each projection factor invariant in sign in each cell, the key idea in the original CAD method for quantifier elimination [Collins 75]. But in [Collins–Hong 91] the method was improved by the introduction of two methods that usually avert construction of a full sign-invariant CAD. The construction begins with a decomposition of the real line into open intervals and one-point intervals in which the univariate projection factors are sign-invariant. The construction proceeds by choosing an existing cell in i-dimensional space and building a *stack* of cells in the cylinder having the chosen cell as base, in each of which the $(i+1)$-variate projection factors are sign-invariant. However this process is restricted in three important ways. First, *trial evaluation*, introduced in [Collins–Hong 91], may reveal that the matrix is true or false in the cell due to the presence in the matrix of some necessaary condition on only some of the variables. Clearly it is unnecessary to construct a stack over any cell whose truth is already known. Second, if f is the number of free variables and $i \geq f$, the truth value of the matrix in the cell may sometimes be determined by *truth propagation*, using the quantifiers in the prenex.. If, for example, it is established by trial evaluation that the matrix is true in a single cell of a stack corresponding to an existential quantifier of the prenex, then it follows that the input formula is true in the base cell of that stack. This truth value may be propagated further down by the application of other quantifiers. A third way, use of *equational constraints* was introduced more recently in [Collins95]. If any equation is implied by the matrix, this can be used to restrict the sets of projection polynomials. This has the further effect of reducing the number of projection polynomials and thereby the size of the sign-invariant CAD determined by the projection factors. Furthermore the equational constraints can be used during the partial CAD construction to immediately assign truth values to some of the cells in a newly constructed stack. Equational constraints are particularly effective for Solotareff approximation problems.

The solution formula construction phase attempts to construct a solution formula, that is, a quantifier-free formula equivalent to the quantified input formula, one that contains as its polynomials only the projection factors, and that is as simple as possible. This is not always possible; we are currently engaged in research on suitable ways to introduce additional polynomials when necessary.

Now let's see how the program was used, and what it did, in obtaining the solutions of the previous section. For $n = 3$ the variable ordering specified for the program was (r, b, u). u must be last because free variables must precede quantified variables. We put r before b because we want to obtain b as a function of r. The two equations in the input formula are both equational constraints. The program uses the fact that the resultant of two equational constraints is also an equational constraint, and such resultants are always included in the projection set. In this example the resultant of the two initial constraint polynomials is just four times the polynomial $B(r, b) = 27b^2 - 2r^3b - 36rb + r^4 + 11r^2 - 1$ occurring in the solution formula. It turns out that the only univariate projection factors in this problem are $r+1$, $r-1$ and r^2+3, arising as a factor of multiplicity three of the discriminant of $B(b, r)$. Therefore the CAD of one-dimensional space consists of just five cells, namely the one-point cells $\{-1\}$ and $\{+1\}$ and the three open intervals comprising the remainder of the real line. All of the cells except the open interval $(1, +\infty)$ are immediately marked false by trial evaluation by virtue of the condition $r > 1$ in the input matrix. Then a stack is constructed over the cell $(1, +\infty)$. This stack contains five cells, two of which are sections corresponding to the two roots of $B(r, b)$ and three are sectors. The sectors are marked false because $B(r, b)$ is an equational constraint. Finally stacks are constructed over the two sections, and application of the quantifier reveals that the lower section is true, the upper section is false. In this simple example the formula construction phase of **qepcad** is unsuccessful but it is obvious that using the first derivative of $B(r, b)$ with respect to b provides the stated solution formula.

Now let's consider the solution for a when $n = 4$. The ordering of variables chosen was (r, a, b, u, v). There are four equational constraints, two in the variable v and two in u. In the first projection the resultant of the two polynomials in v is computed, which is an equational constraint polynomial in r, a and b. In the second projection the resultant of the two u polynomials produces another equational constraint polynomial in r, a and b. Each of these projections also produces some other polynomials. In the third projection the resultant of the two equational constraint polynomials in r a and b, each of which has two irreducible factors, is computed, producing an equational constraint polynomial in r and a. This polynomial has seven irreducible factors, one of which is the polynomial $P_4(r, a)$ that provides the solution. Altogether there are 10 bivariate projection factors, which lead in turn to 37 univariate projection factors. There are 89 cells in the CAD of one-space, the first 48 of which correspond to values of r less than or equal to S_4 and are consequently marked false by trial evaluation. Next the program is instructed to construct stacks over only the remaining 41 cells that are open intervals. The reason for this is that we know that a as a function of r is continuous and we intend to use this to reduce the required computations. Next the program is instructed to construct stacks over any cell other than the one-point cells of one-space. This leads to truth values for all of the cells in two-space that were constructed. Finally the program is instructed to construct stacks over the one-point cells of one-space. Display of the resulting partial CAD then reveals that the only cells marked true are those that are the second (upper) of the two sections of $P_4(r, a)$. The solution formula construction phase of **qepcad** could not be used because cells remain (some sections in the stacks over the one-point cells of one-space) having undetermined truth values. The purpose of constructing the stacks over the one-point cells of one-space was to determine whether $P_4(r, a)$ always has two distinct roots when r has these values. This being indeed the case, we can conclude by continuity that, at these values of r also, a is the larger of the two roots.

The solution for b as a function of r for $r > S_5$ was obtained in a similar manner. For $n = 5$ and $r = 1$, after constructing the CAD of one-space a few of the one-point cells, which are values of b were displayed. One of these was guessed to be the most likely value for the correct value of b, a stack was constructed over this cell, and then stacks over cells in spaces of dimension greater than 1. This resulted in a value, confirming that the guess had been correct. Without this guidance the program would have proceeded to compute truth values for all of the 23 cells in one-space.

In our solutions for $n = 4$ we made vital use of continuity. Otherwise the program would have had to establish truth values for all of the one-point cells in two-space. Some of these have coordinates which are algebraic numbers of very high degree, resulting in enormous computational times. I have not been able to find in the literature any theorem which would imply the continuity of the coefficients of Solotareff best approximation polynomials so I have constructed the proof which follows. We first prove two lemmas.

Lemma 1 *Let $e_n(r)$ be the norm of the error polynomial $x^n + rx^{n-1} - A_n(r, x)$. e_n is a continuous function of r.*

Proof. Let $\delta > 0$ and assume that $\mid r_1 - r_2 \mid < \delta$. Define $P_n(r, x)$ as $x^n + rx^{n-1}$. Then $e_n(r_1) \leq \| P_n(r_1, x) - A_n(r_2, x) \| \leq \| P_n(r_1, x) - P_n(r_2, x) \| + \|$

$P_n(r_2, x) - A_n(r_2, x) \| \leq \delta + e_n(r_2)$. Symmetrically $e_n(r_2) \leq \delta + e_n(r_1)$. So $| e_n(r_1) - e_n(r_2) | \leq \delta$. ∎

Lemma 2 *Let $a_{n,i}(r)$ be the coefficient of x^i in $A_n(r, x)$. $| a_{n,i}(r) | \leq \binom{n}{i}(n+r)$ for all i.*

Proof. $E_n(r, x)$ has $n-1$ roots in $[-1, +1]$ and another root less than -1. The sum of the roots of $E_n(r, x)$ is $-r$ so the least root is greater than $-r-(n-1) > -r-n$. $a_{n,i}(r)$ is a symmetric polynomial in the roots of $E_n(r, x)$ that is the sum of $\binom{n}{i}$ terms, each of which is the product of i of the roots. Each such product has at most one factor, the least root of $E_n(r, x)$ that is not in $[-1, +1]$, and hence the absolute value of each term is less than $n + r$. ∎

Theorem 1 *The coefficients of $A_n(r, x)$ are continuous functions for $r > S_n$.*

Proof. Each coefficient of $A_n(r, x)$ is given by some sections in a CAD of two-dimensional space. There is one such section in each stack over a cell in the CAD of one- dimensional space. Each section over an open interval is a continuous function. Therefore each coefficient is continuous except perhaps at those isolated values of r corresponding to cells in the CAD of one- dimensional space consisting of a single point. Let $\overline{r}$ be any such point. We will show that each coefficient is also continuous at $\overline{r}$.

Let $a_{n,i}(r)$ be the coefficient of x^i in $A_n(r, x)$. $a_{n,i}(r)$ is defined and continuous in an open interval to the right of $\overline{r}$. Let $\overline{a}_i$ be the limit of $a_{n,i}(r)$ as r approaches $\overline{r}$ from the right. We must show that this limit exists and is finite. Existence follows from the fact that, as an algebraic function, $a_{n,i}(r)$ is monotone in some interval to the right of r. By Lemma 2 the limit must be finite.

By Lemma 1, $e(\overline{r}) = \lim_{r \leftarrow r+} e(r) = \lim_{r \leftarrow r+} \| (x^n + rx^{n-1}) - A_n(r, x) \| = \| \lim_{r \leftarrow r+}(x^n + rx^{n-1}) - \lim_{r \leftarrow r+} A_n(r, x) \| = \| (x^n + \overline{r}x^{n-1}) - \overline{A}_n(r, x) \|$. This proves that $\overline{A}_n(r, x)$ is the best approximation to $x^n + \overline{r}x^{n-1}$, from whioh the statement of the theorem follows. ∎

References

[Achieser 56] N. I. Achieser. *Theory of Approximation.* Frederick Ungar Publishing Co., New York, 1956.

[Collins 75] G. E. Collins. Quantifier elimination for the elementary theory of real closed fields by cylindrical algebraic decomposition. In *Lecture Notes In Computer Science, Vol. 33*, pages 134–183, Springer–Verlag, New York, 1975.

[Collins–Hong 91] G. E. Collins and H. Hong. Partial cylindrical algebraic decomposition for quantifier elimination. *Journal of Symbolic Computation*, 12(3): 299–328, 1991.

[Collins95] G. E. Collins. Quantifier elimination by cylindrical algebraic decomposition – twenty years of progress. In *Quantifier Elimination and Cylindrical Algebraic Decomposition*. Springer– Verlag, Vienna, 1995.