Archimedes the Numerical Analyst

Author(s): G. M. Phillips

Source: *The American Mathematical Monthly*, Mar., 1981, Vol. 88, No. 3 (Mar., 1981), pp. 165-169

Published by: Taylor & Francis, Ltd. on behalf of the Mathematical Association of America

Stable URL: https://www.jstor.org/stable/2320460

ARCHIMEDES THE NUMERICAL ANALYST

G. M. PHILLIPS
The Mathematical Institute, University of St. Andrews, St. Andrews, Scotland

1. Introduction. Let p_N and P_N denote half the lengths of the perimeters of the inscribed and circumscribed regular N-gons of the unit circle. Thus $p_3 = 3\sqrt{3}/2$, $P_3 = 3\sqrt{3}$, $p_4 = 2\sqrt{2}$, and $P_4 = 4$. It is geometrically obvious that the sequences $\{p_N\}$ and $\{P_N\}$ are respectively monotonic increasing and monotonic decreasing, with common limit π. This is the basis of Archimedes' method for approximating to π. (See, for example, Heath [2].) Using elementary geometrical reasoning, Archimedes obtained the following recurrence relation, in which the two sequences remain entwined:

$$1/P_{2N} = \tfrac{1}{2}(1/P_N + 1/p_N) \tag{1a}$$

$$p_{2N} = \sqrt{(P_{2N}p_N)}. \tag{1b}$$

We note that these involve the use of the harmonic and geometric means. Beginning with $N = 3$ and applying the recurrence formula five times, Archimedes established the inequalities

$$3\tfrac{10}{71} < p_{96} < \pi < P_{96} < 3\tfrac{1}{7}. \tag{2}$$

His skill in obtaining rational numbers $3\frac{10}{71}$ and (the very familiar) $3\frac{1}{7}$ so close to the irrational numbers p_{96} and P_{96} can be more readily appreciated if we display all four numbers to four decimal places:

$$p_{96} = 3.1410, \qquad 3\tfrac{10}{71} = 3.1408$$

$$P_{96} = 3.1427, \qquad 3\tfrac{1}{7} = 3.1429.$$

2. Stability of the Recurrence Relation. In any thorough study of a recurrence relation we need to consider the question of *numerical stability*, that is, whether rounding errors are magnified by the recurrence relation. As an example, consider the sequence $\{a_n\}$ defined by

$$a_n = \frac{2}{\pi}\int_0^{\pi} e^{\cos\theta}\cos n\theta \, d\theta. \tag{3}$$

(The a_n are the Chebyshev coefficients for e^x; see Clenshaw [1].) It is easily verified, on integrating (3) by parts, that this sequence satisfies the recurrence relation

$$a_{n+1} = a_{n-1} - 2na_n. \tag{4}$$

In principle, given a_0 and a_1, we may then use (4) to compute the value of any a_n. In practice, the recurrence relation (4) does not provide a satisfactory method of computing this sequence, because it is numerically unstable. To illustrate this, suppose we begin with $a_0 = 2.5321$ and $a_1 = 1.1303$, which are correct to 4 decimal places. Using (4) and rounding each a_n to 4 decimal places gives $a_2 = 0.2715$, $a_3 = 0.0443$, $a_4 = 0.0057$, $a_5 = -0.0013$, and $a_6 = 0.0187$. The true values, to 4 decimal places, are a_2 and a_3 as above and $a_4 = 0.0055$, $a_5 = 0.0005$, and $a_6 = 0.0000$. We can now see, on re-examining (4), that the error in a_{n+1} is approximately $(-2n)$ times the error in a_n, which shows why (4) is numerically unstable.

To examine the stability of (1) let us assume that, due to the effect of rounding errors, we actually compute numbers $\tilde{P}_{2N}$ and $\tilde{p}_N$ instead of P_{2N} and p_N, where

George M. Phillips did research in the theory of numbers under E. M. Wright at Aberdeen. He taught at Southampton (1963–1967) before moving to St. Andrews, where he is now Reader in Numerical Analysis. He has made research visits to the University of Texas at Austin and, on several occasions, to the University of Calgary. His main research interests are in numerical analysis and approximation theory. He is coauthor (with P. J. Taylor) of *Computers* (1969) and *Theory and Applications of Numerical Analysis* (1973).—*Editors*

$$\tilde{P}_{2N} = P_{2N}(1+\delta), \tag{5a}$$

$$\tilde{p}_N = p_N(1+\epsilon). \tag{5b}$$

We call δ and ϵ the relative errors in P_{2N} and p_N, respectively. To find the relative error in p_{2N}, we have

$$\tilde{p}_{2N} = \sqrt{(\tilde{P}_{2N}\tilde{p}_N)}. \tag{6}$$

Thus $\tilde{p}_{2N}$ (neglecting the rounding error incurred in evaluating the right side of (6)) is the number we would actually obtain, instead of p_{2N}. Substituting (5) into (6), we have

$$\frac{\tilde{p}_{2N}-p_{2N}}{p_{2N}} = (1+\delta)^{1/2}(1+\epsilon)^{1/2}-1 \tag{7}$$

as the relative error in p_{2N}. Using binomial expansions in (7) we see that, for small values of δ and ϵ,

$$\frac{\tilde{p}_{2N}-p_{2N}}{p_{2N}} \simeq \tfrac{1}{2}(\delta+\epsilon). \tag{8}$$

An analysis of (1a) produces a result similar to (8), showing that rounding errors are not magnified by the recurrence relation, which is thus stable.

3. Rate of Convergence. We have a great advantage over Archimedes in being able to express P_N and p_N in terms of circular functions. It is easily verified that

$$p_N = N\sin(\pi/N) \tag{9}$$

and

$$P_N = N\tan(\pi/N). \tag{10}$$

From (9) and (10) we can justify that (1a) and (1b) are indeed correct and, further, from our familiarity with the Maclaurin series for $\sin\theta$ and $\tan\theta$, we can establish the rate of convergence of the sequences $\{p_N\}$ and $\{P_N\}$. Considering p_N first, we have from (9)

$$p_N = N\left[\left(\frac{\pi}{N}\right) - \frac{1}{3!}\left(\frac{\pi}{N}\right)^3 + \frac{1}{5!}\left(\frac{\pi}{N}\right)^5 - \cdots\right] \tag{11}$$

so that, for large N,

$$\pi - p_N \simeq \frac{1}{6}\pi^3 \cdot \frac{1}{N^2}. \tag{12}$$

We could give a more precise form of (12) by writing down the first two terms of the series (11) plus a remainder term. We can now see from (8) that the error in p_{2N} is approximately one-quarter of the error in p_N. More precisely, we have

$$\lim_{N\to\infty} \frac{\pi - p_{2N}}{\pi - p_N} = \frac{1}{4}. \tag{13}$$

By considering the series for $\tan(\pi/N)$, we see that the errors in the sequence $\{P_N\}$ decrease at the same rate. An inspection of the values of p_N and P_N in Table 1 shows that one might guess this result. (An explanation of the last column of this table follows later.) Given the superb numerical skills of Archimedes, one is sorely tempted to conjecture that he must have been aware of the rate of convergence of his sequences.

4. "Faster" Convergence. We have just seen that the convergence of the sequences $\{P_N\}$ and $\{p_N\}$ is very slow, and it is interesting to consider how to improve on this. First we expand (10) in a Maclaurin series to give

$$P_N = N\left[\left(\frac{\pi}{N}\right) + \frac{1}{3}\left(\frac{\pi}{N}\right)^3 + \frac{2}{15}\left(\frac{\pi}{N}\right)^5 + \cdots\right]. \tag{14}$$

TABLE 1. The first few values of p_N, P_N, and u_N.

N	p_N	P_N	u_N
3	2.598076	5.196152	3.464102
6	3.000000	3.464102	3.154701
12	3.105829	3.215390	3.142349
24	3.132629	3.159660	3.141639
48	3.139350	3.146086	3.141596
96	3.141032	3.142715	3.141593
192	3.141452	3.141873	3.141593

We may now eliminate the terms in $1/N^2$ between (11) and (14) by writing

$$u_N = \frac{1}{3}(2p_N + P_N) = \pi + \frac{1}{20}\frac{\pi^5}{N^4} + \cdots, \tag{15}$$

so that

$$u_N - \pi \simeq \frac{1}{20}\frac{\pi^5}{N^4} \tag{16}$$

and u_N converges to π faster than p_N or P_N. The first few values of u_N are given in Table 1. If we re-calculate the numbers in Table 1 to greater accuracy, we find that u_{96} gives an approximation to π which is more accurate, by a factor greater than 1000, than either of Archimedes' approximations p_{96} and P_{96}.

The technique of eliminating the term in $1/N^2$ could also have been done between p_N and p_{2N} (or, equally, between P_N and P_{2N}). Thus, similarly to (16), we can show that, say,

$$v_N - \pi = \tfrac{1}{3}(4p_{2N} - p_N) - \pi$$

also behaves like a multiple of $1/N^4$ for large N. This process is called *extrapolation to the limit*. (See, for example, Phillips and Taylor [3].) This process can be repeated; that is, we can eliminate the term in $1/N^4$ between v_N and v_{2N}. In Table 2 we show the dramatic effect of repeated extrapolation to the limit. Note that the last two numbers in the final column of Table 2 give π correct to 9 decimal places, although it is only the effect of rounding error which has prevented us from achieving agreement to twice as many places of decimals. If we re-calculate the numbers p_N in Table 2 to 20 decimal places and carry out five extrapolations (rather than three given in the table), we obtain an approximation which differs from π by less than 10^{-18}. It is remarkable that such accuracy can be extracted from Archimedes' raw material.

TABLE 2. The effect of repeated extrapolation to the limit.

N	Extrapolated Values		Repeated Extrapolation	
	p_N	v_N		
3	2.598 076 211			
6	3.000 000 000	3.133 974 596		
12	3.105 828 541	3.141 104 721	3.141 580 063	
24	3.132 628 613	3.141 561 970	3.141 592 454	3.141 592 650
48	3.139 350 203	3.141 590 733	3.141 592 651	3.141 592 654
96	3.141 031 951	3.141 592 534	3.141 592 654	3.141 592 654

5. Analysis of Convergence. In this final section we analyze the behavior of the recurrence relation (1) with arbitrary positive starting values. In divorcing (1) from its geometrical context, we shall change the notation and rewrite (1) in the form

$$1/Q_{N+1} = \tfrac{1}{2}(1/Q_N + 1/q_N) \tag{17a}$$

$$q_{N+1} = \sqrt{(Q_{N+1}q_N)}, \tag{17b}$$

beginning with arbitrary q_0, $Q_0 > 0$. We examine separately the two cases $0 < q_0 < Q_0$ and $0 < Q_0 < q_0$.

Case 1. For $0 < q_0 < Q_0$ we shall write

$$\frac{q_0}{Q_0} = \cos\theta, \qquad \alpha = \frac{q_0 Q_0}{(Q_0^2 - q_0^2)^{1/2}}, \tag{18}$$

so that

$$Q_0 = \alpha \tan\theta, \qquad q_0 = \alpha \sin\theta. \tag{19}$$

Substituting (15) into (13), we easily obtain

$$Q_1 = 2\alpha \tan\tfrac{1}{2}\theta, \qquad q_1 = 2\alpha \sin\tfrac{1}{2}\theta. \tag{20}$$

It follows that

$$Q_N = 2^N \alpha \tan(\theta/2^N), \qquad q_N = 2^N \alpha \sin(\theta/2^N), \tag{21}$$

and hence the sequences $\{Q_N\}$ and $\{q_N\}$ converge to the common limit

$$\alpha\theta = \frac{q_0 Q_0}{(Q_0^2 - q_0^2)^{1/2}} \cos^{-1}(q_0/Q_0). \tag{22}$$

The "Archimedes case" corresponds to $q_0 = 3\sqrt{3}/2$, $Q_0 = 3\sqrt{3}$.

Case 2. For $0 < Q_0 < q_0$ we write

$$\frac{q_0}{Q_0} = \cosh\theta, \qquad \alpha = \frac{q_0 Q_0}{(q_0^2 - Q_0^2)^{1/2}}, \tag{23}$$

so that

$$Q_0 = \alpha \tanh\theta, \qquad q_0 = \alpha \sinh\theta. \tag{24}$$

Substituting (24) into (17), we obtain

$$Q_1 = 2\alpha \tanh\tfrac{1}{2}\theta, \qquad q_1 = 2\alpha \sinh\tfrac{1}{2}\theta.$$

It follows that

$$Q_N = 2^N \alpha \tanh(\theta/2^N), \qquad q_N = 2^N \alpha \sinh(\theta/2^N),$$

and hence the sequences $\{Q_N\}$ and $\{q_N\}$ again converge to a common limit which, in this case, is

$$\alpha\theta = \frac{q_0 Q_0}{(q_0^2 - Q_0^2)^{1/2}} \cosh^{-1}(q_0/Q_0). \tag{25}$$

As an amusing application of this last result, let us choose

$$Q_0 = 2t, \qquad q_0 = t^2 + 1$$

for any positive $t \neq 1$. Then from (25) the sequences $\{Q_N\}$ and $\{q_N\}$ have common limit

$$\frac{2t(t^2+1)}{(t^2-1)} \log t.$$

This gives a simple method for evaluating $\log t$ and repeated extrapolation may be used to accelerate convergence. However, this is not proposed as a practical algorithm for computing $\log t$.

Acknowledgments. I am indebted to my colleagues J. M. Howie and J. J. O'Connor for interesting discussions on this topic. I also wish to thank John W. Wrench, Jr., for his valuable comments and for the trouble he took to re-work independently the calculations embodied in Tables 1 and 2.

References

1. C. W. Clenshaw, Chebyshev Series for Mathematical Functions, Mathematical Tables, vol. 5, National Physical Laboratory, H.M.S.O., London, 1962.

2. T. L. Heath, The Works of Archimedes, Cambridge University Press, 1897.

3. G. M. Phillips and P. J. Taylor, Theory and Applications of Numerical Analysis, Academic Press, 1973.

A BRIEF HISTORY AND SURVEY OF THE CATENARY CHAIN CONJECTURES

L. J. RATLIFF, JR.
Department of Mathematics, University of California, Riverside, CA 92521

1. Introduction and Some Terminology. There is a collection of problems in commutative algebra known as the catenary chain conjectures. These conjectures, some of which have their origins in W. Krull's foundational work in 1937, are concerned with the extent to which certain useful properties hold in the integral closure of a Noetherian domain. The purpose of this article is to tell what the most important of these conjectures say, where they came from, and what their current status is.

A very brief summary of these conjectures is that they are concerned with maximal chains of prime ideals in integral extension domains of Noetherian domains. This summary will be made considerably more specific in Sections 2–7, but since some of the terminology in the preceding sentence may not be familiar to the reader, the remainder of this section will be devoted to explaining some of the relevant definitions and giving some examples to illustrate them. (Other definitions will be given when they are needed in later sections of the paper.)

The conjectures and related results are a small but well-defined and important area in the study of *Noetherian rings*—those rings R which are commutative, have an identity $1 \neq 0$, and for which every ideal is finitely generated or, equivalently, that satisfy the ascending chain condition (that is, every strictly ascending chain of ideals of R is finite). These rings are named after Emmy Noether, who, in 1921 in a very important paper **[27]**, was the first to recognize their importance. They have been extensively studied ever since, and many very important and interesting theorems concerning them have been discovered. They are now clearly one of the basic structures in all of mathematics.

Actually, most of the conjectures and problems in this area can be reduced to local considerations; that is, it is sufficient to restrict attention to Noetherian rings with a unique maximal ideal. Such rings are called *local rings*, and they arise naturally in algebraic geometry (in studying the geometry on an algebraic variety in the neighborhood of a point) and in algebraic number theory (in solving Diophantine problems). (The reduction from the global conjectures [for Noetherian rings in general] to their local versions is readily accomplished by localizing at maximal ideals of R. The method of reducing global problems to local ones is standard in commutative algebra and need not be considered here.) Local rings are topological rings, the topology being given by using the set of powers of the maximal ideal M as the

---

The author received his Ph.D. at the State University of Iowa under H. T. Muhly in 1961. He was a lecturer at Indiana University from 1961 to 1963 and has been at the University of California, Riverside, since then. He has written extensively on commutative algebra and has a monograph, *Chain Conjectures in Ring Theory*, in the Springer Lecture Notes in Mathematics. His main interests are in commutative algebra, Noetherian lattices, algebraic geometry, number theory, and teaching.—*Editors*