

A NONLINEAR DIFFERENCE EQUATION WITH TWO PARAMETERS

A. BROWN¹

(Received 13 December 1983)

Abstract

The paper is mainly concerned with the difference equation

$$y_{n+1} = 2k / \{1 + (y_n - m)^2\},$$

where k and m are parameters, with $k > 0$. This equation arises from a method proposed for solving a cubic equation by iteration and represents a standardised form of the general problem. In using the above equation it is essential to know when the iteration process converges and this is discussed by means of the usual stability criterion. Critical values are obtained for the occurrence of solutions with period two and period three and the stability of these solutions is also examined. This was done by considering the changes as k increases, for a given value of m , which makes it effectively a one-parameter problem, and it turns out that the changes with k can differ strongly from the usual behaviour for a one-parameter difference equation. For $m = 2$, for example, it appears that the usual picture of stable 2-cycle solutions giving way to stable 4-cycle solutions is valid for smaller values of k but the situation is reversed for larger values of k where stable 4-cycle solutions precede stable 2-cycle solutions. Similar anomalies arise for the 3-cycle solutions.

1. Introduction

In some work on a stability problem in aerodynamics, Lin [8, 9] developed a method of finding factors of a polynomial by an iterative procedure. He was mainly concerned with quadratic factors, in order to determine complex roots of the polynomial, but in the second paper cited above he put forward a more

¹Department of Theoretical Physics, Research School of Physical Sciences, Australian National University, G.P.O. Box 4, Canberra, A. C. T. 2601.

© Copyright Australian Mathematical Society 1985, Serial-fee code 0334-2700/85

general scheme which could be used in approximating to a factor of any degree. For a linear factor, his scheme is equivalent to using the difference equation

$$x_{n+1} = \frac{-x_n G(0)}{G(x_n) - G(0)} \quad (1.1)$$

to give successive approximations to a root of the polynomial equation $G(x) = 0$. In this form it is discussed by Hildebrand [7], who includes an example where Lin's method breaks down. Lin was aware that his successive approximations might diverge rather than converge and he gives an example [8, page 239] where a shift of origin made the process convergent when it had previously been divergent.

Coppel [2] examined this problem for the case where the polynomial is a cubic

$$p(x) = x^3 + ax^2 + bx + c, \quad (1.2)$$

with c nonzero, and the iteration rule is

$$x_{n+1} = \frac{-c}{b + ax_n + x_n^2}. \quad (1.3)$$

He showed that $\{x_n\}$ converges monotonically to a root of $p(x) = 0$, provided the coefficients of $p(x)$ satisfy the conditions

$$c < 0, \quad b > \frac{1}{4}a^2, \quad c - \frac{ab}{2} + \frac{a^3}{8} > 0. \quad (1.4)$$

These conditions imply that $a < 0$ and $b < 0$, so at first sight they appear somewhat restrictive, but they can always be satisfied by shifting the origin far enough to the left [2].

It was suggested [3] that it would be of interest to look for periodic solutions of equation (1.3), in view of the increasing importance of nonlinear difference equations in a variety of problems [10, 5]. As can be seen from the monograph by Collet and Eckmann [1], most of the information available is for one-parameter problems whereas equation (1.3) involves two parameters. The work reported in this paper considers mainly solutions of period 2 and period 3 but it is enough to show conspicuous differences in behaviour compared with the results for one-parameter problems.

The paper starts by writing equation (1.3) in a standard form, which involves parameters m and k , with $k > 0$. After a discussion of the equilibrium solutions, expressions are obtained for the stability criterion, S_1 , and are used to see how S_1 varies for different values of m and k and for different solutions, where more than one solution is available. Coppel's conditions are considered in this context.

In Section 3, solutions with period two are discussed and it is shown that for a given value of m there is a minimum value of k at which C_2 solutions (solutions with a minimum period two) can occur. For $m < 1$ all the C_2 solutions are stable, whereas for a given value of m , with $m > 1$, there is a finite interval of stability

for smaller values of k and an infinite interval of stability for larger values of k . Between these there is a range of values of k for which the $C2$ solutions are unstable. Thus the usual arguments about bifurcation of $C2$ solutions into $C4$ solutions, and so on, must somehow be reversed in this case when k is large.

The $C3$ solutions are discussed in Sections 4 and 5 and their behaviour is also unorthodox. For $m < \sqrt{3}$, no $C3$ solutions occur. For m fixed, with $m > \sqrt{3}$, there is an interval $K_1 < k < K_2$ within which there are two $C3$ solutions for each value of k . K_1 and K_2 are found explicitly as functions of m and they are tabulated with other critical values of k in Table 2. One family of solutions is always unstable but the other family provides at least some stable $C3$ solutions. The general situation (for m sufficiently large) is that there are two intervals of k within which the solutions are stable, with unstable solutions occurring for the intervening k -values. Thus as k increases, with m fixed, it is possible to have stable $C2$ solutions preceding stable $C3$ solutions (for smaller values of k) and then to have another group of stable $C3$ solutions preceding stable $C2$ solutions (for larger values of k). It will be recognised that this is highly unorthodox [1].

Section 6 has some brief comments on superstable solutions, where the stability criterion is zero, and equations for locating these solutions are given in the $C3$ and $C4$ cases.

Section 7 goes back to the original problem of using equation (1.3) to obtain a root of the cubic $p(x) = 0$. To some extent this arose from the work done in Section 5, which involved solving about 100 cubic equations to obtain the elements of the $C3$ solutions for different values of m and k . This could be done by using tables, such as the tables published by Salzer, Richards and Arsham [11], but the availability of high speed computers makes it less convenient to look up tables. There was something of a challenge also in applying the information from the work on equation (1.3) to devise an efficient routine. This involved a discussion of the effect on S_1 of a change of origin and a look at cases where the denominator on the right-hand side of equation (1.3) is not positive definite. (The work in Sections 2 to 6 assumes that $b + ax + x^2$ is positive definite.)

2. Basic equation and equilibrium solutions

In using equation (1.3) there are obviously advantages in having $a^2 < 4b$, so that $x_n^2 + ax_n + b$ is always positive. As mentioned above, this condition can be met by a change of origin and we can then write $Q^2 = b - (1/4)a^2$, with $Q > 0$. If we put $a = -2mQ$, $c = -2kQ^3$ and $x_n = Qy_n$, equation (1.3) becomes

$$y_{n+1} = F(y_n) = \frac{2k}{(y_n - m)^2 + 1} \quad (2.1)$$

and we can treat this as a standard form, with m and k as the parameters. Since c is nonzero, k is also nonzero but otherwise there is no restriction on m and k . However, if we put $y_n = -u_n$ the equation becomes

$$u_{n+1} = \frac{-2k}{(u_n + m)^2 + 1}. \tag{2.2}$$

The convergence properties of the solutions will not be affected. Thus the behaviour for parameter values $(-m, -k)$ is the same as for $(+m, +k)$, so we can take k as positive in equation (2.1) and leave m unrestricted.

For any value of y_0 in equation (2.1) we will have $0 \leq y_1 \leq 2k$ and subsequent iteration will give $0 < y_n \leq 2k$ for $n > 1$. Thus any equilibrium solution must lie in $(0, 2k]$ and if we start with $0 \leq y_0 \leq 2k$ the iteration can be regarded as a mapping of the interval on itself. The equilibrium points and the elements of any periodic solution must be positive and they cannot be greater than $2k$. If k is small, any equilibrium solution must be close to zero and we can think of $y = 0$ as the limiting value of an equilibrium solution as k tends to zero.

If $G(y) = y^3 - 2my^2 + (1 + m^2)y$, the equilibrium solutions satisfy $G(y) = 2k$, with $y > 0$. For $y > 0$ and $m < \sqrt{3}$, $G(y)$ increases monotonically with y so there can only be one equilibrium solution. For $m > \sqrt{3}$, the graph of $G(y)$ has a maximum at $(Y_1, 2k_1)$ and a minimum at $(Y_3, 2k_3)$, where

$$Y_1 = \frac{1}{3} \{ 2m - \sqrt{m^2 - 3} \}, \tag{2.3}$$

$$k_1 = \frac{1}{27} \{ m^3 + 9m + (m^2 - 3)\sqrt{m^2 - 3} \}, \tag{2.4}$$

and Y_3, k_3 are obtained from Y_1, k_1 by changing the sign of the square root. Hence for $m > \sqrt{3}$ and $k_3 < k < k_1$, there are three equilibrium solutions which we can label as the left-hand solution ($0 < y < Y_1$), the intermediate solution ($Y_1 < y < Y_3$) and the right-hand solution ($y > Y_3$). For $k = k_1$, $y = Y_1$ is a double root of $G(y) = 2k$ and the remaining root occurs for $y > Y_3$. Similarly, $y = Y_3$ is a double root for $k = k_3$ and the remaining root occurs for $0 < y < Y_1$.

For $m = \sqrt{3}$, $Y_1 = Y_3 = (2/3)\sqrt{3}$ and $G(y) = 2k$ has a triple root for $k = (4/9)\sqrt{3}$. In this case there is only one equilibrium value of y for $k > 0$.

If Y is an equilibrium solution of equation (2.1), then the stability criterion at Y is $S_1 = F'(Y)$. For local stability, $|S_1|$ must be less than 1. We must have Y positive, with

$$G(Y) = Y\{1 + (Y - m)^2\} = 2k, \tag{2.5}$$

and we can make use of this equation to write

$$S_1 = \frac{-4k(Y - m)}{\{1 + (Y - m)^2\}^2} = \frac{-2Y(Y - m)}{1 + (Y - m)^2} = \frac{Y^2(m - Y)}{k}. \tag{2.6}$$

If we regard m as fixed and let k vary from 0 to ∞ , then $Y \sim 2k/(1 + m^2)$ for k small and $Y \sim (2k)^{1/3}$ for k large. It follows that S_1 is close to zero for k small, with $S_1 \rightarrow 0$ as $k \rightarrow 0$, and that $S_1 \rightarrow -2$ as k and Y tend to infinity. Also,

$$\frac{\partial S_1}{\partial Y} = \frac{2\{mY^2 - 2(1 + m^2)Y + m(1 + m^2)\}}{\{1 + (Y - m)^2\}^2}, \tag{2.7}$$

and this gives $(\partial S_1/\partial Y) < 0$ for $m \leq 0$. In this case, S_1 decreases from 0 to -2 as k and Y go from 0 to ∞ and the equilibrium is on the point of instability when $S_1 = -1$. From equation (2.6), this occurs when $2Y(Y - m) = 1 + (Y - m)^2$, which leads to

$$Y = Y^* = \sqrt{1 + m^2}, \quad k = k^* = (1 + m^2)\{-m + \sqrt{1 + m^2}\}. \tag{2.8}$$

Indeed, this result holds for all values of m , that is, $S_1 = -1$ when $Y = Y^*$ and $k = k^*$.

For $m > 0$, $(\partial S_1/\partial Y)$ is positive for small values of k and Y . As Y increases, S_1 attains a maximum when

$$Y = Y_2 = \frac{1}{m} \{(1 + m^2) - \sqrt{1 + m^2}\}. \tag{2.9}$$

At this maximum, $S_1 = -1 + \sqrt{1 + m^2}$, and for $0 < m < \sqrt{3}$ the maximum lies between 0 and 1. For $m = \sqrt{3}$, the maximum is 1 and it occurs for the triple root $Y = (2/3)\sqrt{3}$. For $m > \sqrt{3}$, the maximum is greater than 1 and there must be an intermediate value, with $0 < Y < Y_2$, at which $S_1 = 1$. It turns out that this occurs for $Y = Y_1$. For $Y > Y_2$, $\partial S_1/\partial Y$ is negative until $Y = Y_4$, where

$$Y_4 = \frac{1}{m} \{(1 + m^2) + \sqrt{1 + m^2}\}, \tag{2.10}$$

and $\partial S_1/\partial Y$ is positive again for $Y > Y_4$. When $Y = Y_4$, S_1 has its minimum value $-1 - \sqrt{1 + m^2}$ and it is clear that this minimum value is less than -2 (for any $m > 0$). For $Y > Y_4$, S_1 increases and tends to -2 as $Y \rightarrow \infty$.

This means that for $0 < m < \sqrt{3}$ the equilibrium solution is stable for $0 < k < k^*$. For $m > \sqrt{3}$, we still have $S_1 = -1$ when $k = k^*$ but we have an interval where $S_1 > 1$ and we need to know when $S_1 = 1$. From equation (2.6), $S_1 = 1$ when $-2Y(Y - m) = 1 + (Y - m)^2$ and this leads to the equation

$$3Y^2 - 4mY + (1 + m^2) = 0. \tag{2.11}$$

The left-hand side is simply $G'(Y)$, so $S_1 = +1$ when Y is a double root of $G(y) = 2k$, that is when $Y = Y_1$ and when $Y = Y_3$. This can easily be verified.

If we go back to the graph of $G(y)$, for $m > \sqrt{3}$ and $y > 0$, every point on this graph corresponds to an equilibrium solution for a suitably-chosen value of k . For points on the left-hand ascending branch, we have $0 < y < Y_1$ and $0 < S_1 < 1$, so these points correspond to stable equilibrium solutions. For the descending branch, we have $Y_1 < y < Y_3$ and $S_1 > 1$, so these points correspond to unstable

equilibrium solutions. For the right-hand ascending branch, we have $-1 < S_1 < 1$ for $Y_3 < y < Y^*$ and these points correspond to stable equilibrium solutions, but for $y > Y^*$ we have $S_1 < -1$ and the corresponding equilibrium solutions are unstable.

Although these results have been obtained for a standardised form of the difference equation, it is easy to see that a change of scale such as $x_n = Qy_n$ does not change the value of S_1 at an equilibrium point so the results for equation (1.3) can be discussed in the same way. With this comment, we can return to Coppel's conditions (1.4) and see how they apply to the cubic $p_1(y) = y^3 + a_1y^2 + b_1y + c_1$, where $a_1 = -2m$, $b_1 = 1 + m^2$, and $c_1 = -2k$. For $k > 0$ the first two conditions are satisfied and the third condition is also satisfied if

$$m - 2k > 0. \quad (2.12)$$

This rules out cases where $m \leq 0$ but this is understandable since $S_1 < 0$ for $m \leq 0$ and convergence to the equilibrium solution would not be monotonic. Because Coppel's conditions ensure monotonic convergence they should apply only for $0 < S_1 < 1$ or possibly for the limiting cases when $S_1 = 0$ or $S_1 = 1$. From equation (2.6), $S_1 = 0$ when $Y = Y_c = m$ and this requires $2k = G(m) = m$. Thus condition (2.12) excludes the case where $m = 2k$ and $S_1 = 0$. For $m > \sqrt{3}$, we have

$$Y_c = m > \frac{1}{3} \{2m + \sqrt{m^2 - 3}\} = Y_3, \quad (2.13)$$

which means that $y = Y_c$ gives a point on the right-hand ascending branch of the graph of $G(y)$ and condition (2.12) is valid for the points on the graph between $y = Y_3$ and $y = Y_c$. Thus we will have $0 < S_1 < 1$ for both the left-hand solution and the right-hand solution when there are three distinct solutions and condition (2.12) holds. (The intermediate solution is always unstable.)

A minor point is that condition (2.12) covers some cases where $S_1 = 1$, for example the "triple root" case when $m = \sqrt{3}$ and $k = (4/9)\sqrt{3}$ and the "double root" cases when $m > \sqrt{3}$ and $k = k_3$. In such cases, the stability is marginal and the convergence will be very slow.

3. Solutions of period two

If we look for solutions of equation (2.1) with minimum period 2 (C_2 solutions), then we want positive numbers b_1 and b_2 such that $b_1 \neq b_2$ and

$$2k = b_2 \{1 + (b_1 - m)^2\}, \quad (3.1)$$

$$2k = b_1 \{1 + (b_2 - m)^2\}. \quad (3.2)$$

These equations give

$$\beta = b_1 b_2 = 1 + m^2, \quad \alpha = b_1 + b_2 = 2m + \{2k/(1 + m^2)\}, \quad (3.3)$$

and the condition for real solutions for b_1 and b_2 can be written as $k > k^*$, with k^* as defined in equation (2.8). It can be shown that

$$2k^* - m = \frac{m + 2\sqrt{1 + m^2}}{\{m + \sqrt{1 + m^2}\}^2} > 0,$$

so $C2$ solutions are only possible when $2k > 2k^* > m$, that is, outside the range covered by equation (2.12).

The stability criterion for a $C2$ solution is $S_2 = F'(b_1)F'(b_2)$ and this leads to the expression

$$S_2 = (1 + m^2) \frac{(1 + m^2 - 2km)}{k^2} \tag{3.4}$$

and it can be checked that this gives $S_2 \rightarrow 1$ as $k \rightarrow k^*$ from above. Also,

$$\frac{\partial S_2}{\partial k} = 2(1 + m^2) \frac{km - (1 + m^2)}{k^3}. \tag{3.5}$$

The sign of this derivative depends on the factor $km - (1 + m^2)$. For $m \leq 0$ this factor is negative, so S_2 decreases as k increases (for m fixed). Also, S_2 is positive for $m \leq 0$ and tends to zero as $k \rightarrow \infty$, which means that all $C2$ solutions are stable, with $0 < S_2 < 1$ for $k > k^*$. This is in marked contrast to the usual behaviour for one-parameter mappings [1, pages 23–26].

For $m > 0$, we note that

$$k^*m - (1 + m^2) = -(1 + m^2)^{3/2} \{-m + \sqrt{1 + m^2}\} < 0,$$

so S_2 decreases at first as k increases above the critical value k^* . As before, we can keep m fixed and consider the changes in S_2 as k increases. S_2 becomes zero when $k = (1 + m^2)/(2m)$, reaches a minimum value when $k = (1 + m^2)/m$ and then increases again for larger values of k . At the minimum, $S_2 = -m^2$, and S_2 increases toward zero as $k \rightarrow \infty$. For $0 < m < 1$, S_2 lies between -1 and $+1$ for all admissible values of k and again all the $C2$ solutions are stable. For $m = 1$, the $C2$ solutions are almost all stable but with a solution on the margin for stability when $k = 2$. For $m > 1$, the limiting values of k for stability will occur when $S_2 = -1$ and from equation (3.4) the corresponding values of k are k_5 and k_6 , where

$$k_5, k_6 = (1 + m^2) \{m \pm \sqrt{m^2 - 1}\}. \tag{3.6}$$

If we take $k_5 < k_6$, then the $C2$ solutions are stable in the interval (k^*, k_5) , unstable in (k_5, k_6) and stable again for $k > k_6$. Once again, this is unusual behaviour compared with what happens for one-parameter mappings.

Another unorthodox feature is that it is possible to have a stable equilibrium solution and a stable $C2$ solution for the same parameter values m and k . As a numerical example we can take $m = 3$, $k = 5/3$. Then $p_1(y) = y^3 - 6y^2 + 10y - (10/3)$ and the equation $p_1(y) = 0$ has solutions 0.44184, 2.35587, and 3.20229.

The corresponding values for S_1 are 0.29965, 2.14501, and -1.24465 , so the solution at $y = 0.44184$ is stable. At the same time, there is a C_2 solution with $b_1 = 10/3, b_2 = 3$ and $S_2 = 0$. One would expect the two stable solutions to have different catchment areas, so to speak, with the intermediate solution as the watershed.

In the example above, the sum of the S_1 values is 1.20001 and this agrees with a general result, that the sum of the S_1 values is $(4m - 6k)/k$ when $p_1(y) = 0$ has three distinct real roots. This follows fairly simply from equation (2.6).

4. Equations for solutions of period three

For a C_3 solution (a solution with minimum period three) we need distinct positive numbers b_1, b_2, b_3 such that $b_2 = F(b_1), b_3 = F(b_2), b_1 = F(b_3)$. This gives a set of three equations of which

$$2k = (1 + m^2)b_2 - 2mb_1b_2 + b_1^2b_2 \tag{4.1}$$

is typical. Since the choice of b_1 is arbitrary, it is convenient to use symmetrical functions

$$\alpha = b_1 + b_2 + b_3, \quad \beta = b_1b_2 + b_2b_3 + b_3b_1, \quad \gamma = b_1b_2b_3,$$

and to look for equations which allow α, β, γ to be determined. Once they are known, b_1, b_2, b_3 are the roots of $h(x) = 0$, where $h(x) = x^3 - \alpha x^2 + \beta x - \gamma$.

Since the three equations represented by (4.1) have cyclic symmetry it is convenient to use Σ for cyclic summation over the three suffixes. For example,

$$\Sigma b_1^2b_2 = \Sigma b_3^2b_1 = b_1^2b_2 + b_2^2b_3 + b_3^2b_1.$$

Cyclic summation of equation (4.1) gives

$$6k = (1 + m^2)\alpha - 2m\beta + \Sigma b_1^2b_2, \tag{4.2}$$

and in the same way if we multiply equation (4.1) in turn by b_3, b_3^2, b_3b_1, b_2 and in each case use cyclic summation we get

$$2k\alpha = (1 + m^2)\beta - 6m\gamma + \alpha\gamma, \tag{4.3}$$

$$2k(\alpha^2 - 2\beta) = (1 + m^2)\Sigma b_1b_2^2 - 2m\gamma\alpha + \beta\gamma, \tag{4.4}$$

$$2k\beta = 3(1 + m^2)\gamma - 2m\alpha\gamma + \gamma(\alpha^2 - 2\beta), \tag{4.5}$$

$$2k\alpha = (1 + m^2)(\alpha^2 - 2\beta) - 2m\Sigma b_1b_2^2 + \beta^2 - 2\alpha\gamma. \tag{4.6}$$

In writing out these equations we have used identities such as $\Sigma b_1^2 = \alpha^2 - 2\beta$ and $\Sigma b_1^2b_2^2 = \beta^2 - 2\alpha\gamma$, and we can note also that

$$\alpha\beta = 3\gamma + \Sigma b_1^2b_2 + \Sigma b_1b_2^2. \tag{4.7}$$

If we use equations (4.2), (4.4), (4.5) and (4.7) to eliminate $\Sigma b_1^2 b_2$, $\Sigma b_1 b_2^2$ and $\beta\gamma$, we obtain

$$\begin{aligned} \beta \{ 6k + (2\alpha - 4m)(1 + m^2) \} + \gamma(\alpha^2 - 6m\alpha - 3 - 3m^2) \\ = 4k\alpha^2 + 12k(1 + m^2) - 2\alpha(1 + m^2)^2. \end{aligned} \tag{4.8}$$

Equations (4.3) and (4.8) are linear in β and γ and we can solve them to obtain

$$\begin{aligned} \beta D = 2\alpha(6m - \alpha)(1 + m^2)^2 + 18k(\alpha - 4m)(1 + m^2) \\ + 2k\alpha^2(\alpha - 6m), \end{aligned} \tag{4.9}$$

$$\gamma D = 2\alpha(1 + m^2)^3 - 12k(1 + m^2)^2 - 8km\alpha(1 + m^2) + 12k^2\alpha, \tag{4.10}$$

where

$$D = 3(1 + m^2)^2 + 6k(\alpha - 6m) + (\alpha - 4m)(\alpha - 6m)(1 + m^2). \tag{4.11}$$

For $D \neq 0$, these equations express β and γ as functions of α, k, m and we need an equation for α to complete the solution. If we substitute for β and γ in equation (4.5), the outcome is an equation of the form

$$0 = M_1(\alpha, m, k) = \sum_{j=0}^5 \alpha^j f_j(m, k), \tag{4.12}$$

where the coefficients $f_j(m, k)$ are polynomials in m and k . From other considerations, the equation for α should be of degree 2, so a factor of degree 3 has to be identified, explained and eliminated.

The clue to this process came from the case $m = 0$ where most of the terms disappear, leaving

$$\begin{aligned} M_1(\alpha, 0, k) &= (1 + 4k^2)\alpha^5 - 2k\alpha^4 + (10 + 36k^2)\alpha^3 \\ &\quad - (72k + 216k^3)\alpha^2 + (9 + 108k^2)\alpha - 54k \\ &= (\alpha^3 + 9\alpha - 54k) \{ (1 + 4k^2)\alpha^2 - 2k\alpha + 1 \}. \end{aligned}$$

Now although b_1, b_2 and b_3 should be distinct for a $C3$ solution we have not used this in forming the equations for α, β and γ . As a result, we can expect to have included the degenerate case where $b_1 = b_2 = b_3 = b$, with $G(b) = 2k$. For $m = 0, G(b) = b^3 + b$ and, since $\alpha = 3b$ in the degenerate case, the equilibrium equation is

$$(\alpha/3)^3 + (\alpha/3) = 2k$$

that is, $\alpha^3 + 9\alpha - 54k = 0$. This explains the cubic factor in $M_1(\alpha, 0, k)$ and equating the other factor to zero gives the appropriate values of α for a $C3$ solution. In fact, the second factor gives complex values for α so there are no real $C3$ solutions for $m = 0$. This is not surprising, since all $C2$ solutions are stable for $m < 1$.

For general values of m , the equilibrium solution $b_1 = b_2 = b_3 = b$ again gives $\alpha = 3b$ and in this case $G(b) = 2k$ leads to

$$E(\alpha, m, k) = \alpha^3 - 6m\alpha^2 + 9(1 + m^2)\alpha - 54k = 0.$$

It can be verified that $E(\alpha, m, k)$ is a factor of $M_1(\alpha, m, k)$ and the remaining factor gives

$$M_2(\alpha, m, k) = A\alpha^2 - 2B\alpha + C = 0, \quad (4.13)$$

where

$$A = (1 + m^2)^3 - 4km(1 + m^2) + 4k^2, \quad (4.14)$$

$$B = 3m(1 + m^2)^3 + k(1 + m^2)^2 - 12km^2(1 + m^2) + 12k^2m, \quad (4.15)$$

$$C = (1 + m^2)^4 + 8m^2(1 + m^2)^3 - 32km^3(1 + m^2) + 48k^2m^2. \quad (4.16)$$

For most values of k and m , the equations above suffice to determine α , β , and γ . A , B , and C can be calculated from equations (4.14), (4.15) and (4.16) and equation (4.13) then gives the appropriate values for α . If they are real, we can use equations (4.11), (4.9) and (4.10) to determine the corresponding values of β and γ .

Break-down sometimes occurred because D was zero (or close to zero). To deal with this case, β was eliminated between equations (4.3) and (4.5) and this gave a quadratic equation

$$2\gamma^2(\alpha - 6m) + \gamma\{3(1 + m^2)^2 + \alpha(\alpha - 2m)(1 + m^2) - 2k(\alpha + 6m)\} - 4k^2\alpha = 0, \quad (4.17)$$

which could be solved for γ . Equation (4.3) then gave β . In most cases, only one solution of the quadratic was required and the other gave a spurious solution which violated one or more of the check equations. (With a little bit of practice it was possible to pick out the appropriate value for γ and to incorporate this choice in the computing programme.)

One of the check equations came from eliminating $\Sigma b_1^2 b_2$ and $\Sigma b_1 b_2^2$ between equations (4.2), (4.6) and (4.7). This gave

$$\beta^2 + 2\beta(m^2 - 1 - m\alpha) - 2\gamma(\alpha - 3m) = 2k(\alpha - 6m) + \alpha(2m - \alpha)(1 + m^2) \quad (4.18)$$

Another came from multiplying equation (4.1) by b_3 and subtracting the corresponding equation (with the suffixes moved on by 1). This gave

$$2k(b_3 - b_1) = (b_1 - b_2)\{\gamma - (1 + m^2)b_3\} \quad (4.19)$$

and there are two corresponding equations, with $2k(b_1 - b_2)$ and $2k(b_2 - b_3)$ on the left-hand side. If we multiply the three equations together and cancel a common factor $(b_1 - b_2)(b_2 - b_3)(b_3 - b_1)$, assuming that b_1 , b_2 and b_3 are distinct, we get

$$\begin{aligned} 8k^3 &= \{\gamma - (1 + m^2)b_1\}\{\gamma - (1 + m^2)b_2\}\{\gamma - (1 + m^2)b_3\} \\ &= \gamma^3 - \alpha(1 + m^2)\gamma^2 + \beta(1 + m^2)^2\gamma - (1 + m^2)^3\gamma. \end{aligned} \quad (4.20)$$

The stability criterion for a $C3$ solution with elements (b_1, b_2, b_3) is $S_3 = F'(b_1)F'(b_2)F'(b_3)$ and we can write

$$F'(b_1) = \frac{-4k(b_1 - m)}{\{1 + (b_1 - m)^2\}^2} = \frac{-4k(b_1 - m)}{(2k/b_2)^2} = \frac{-b_2^2(b_1 - m)}{k}.$$

From this,

$$S_3 = (b_1 b_2 b_3)^2 (m - b_1)(m - b_2)(m - b_3)/k^3 = \gamma^2 h(m)/k^3. \tag{4.21}$$

Since $h(m) = m^3 - \alpha m^2 + \beta m - \gamma$, S_3 can be evaluated from k, m, α, β and γ without finding b_1, b_2 , and b_3 explicitly.

5. Critical values for solutions of period three

It was shown in Section 2 that all $C2$ solutions are stable for $m < 1$, so we shall assume that no $C3$ solutions occur unless $m \geq 1$. From equation (4.13) the values for α will not be real unless $B^2 \geq AC$ and this suggests that we examine $T(m, k) = B^2 - AC$, to find when it changes sign. In detail,

$$T(m, k) = -(1 + m^2)^6 + 2km(1 + m^2)^5 + (8km - 19k^2)(1 + m^2)^4 + (16k^2 + 56k^3m)(1 + m^2)^2 - (32k^3m + 48k^4)(1 + m^2) + 48k^4. \tag{5.1}$$

If we put $p = 4mk$ and $p_0 = (1 + m^2)^2$, we can write

$$16m^2 T(m, k) = -(p - p_0)^2 \{3p^2 - 8m^2(1 + m^2)p + 16m^2(1 + m^2)^2\}, \tag{5.2}$$

and the quadratic factor is zero when

$$p = (4/3)(1 + m^2)m \{m \pm \sqrt{m^2 - 3}\}. \tag{5.3}$$

This implies that the quadratic factor is positive definite for $0 < m < \sqrt{3}$ and in this case $T(m, k)$ is negative except when $p = p_0$. We shall look at this special case more carefully later.

For $m > \sqrt{3}$, $T(m, k) = 0$ when $p = p_0, p_1$ or p_2 , where p_1 and p_2 are given by equation (5.3). We can take $p_1 < p_2$ and it is easy to check that $p_1 < p_0 < p_2$. If we put $p_0 = 4mK_0, p_1 = 4mK_1$ and $p_2 = 4mK_2$, then

$$T(m, k) = 48m^2 (K_2 - k)(k - K_1)(k - K_0)^2. \tag{5.4}$$

which means that $T(m, k)$ is positive for $K_1 < k < K_0$ and $K_0 < k < K_2$, zero for $k = K_0, K_1, K_2$ and negative for $k < K_1$ or $k > K_2$. Thus for a given value of

m , with $m > \sqrt{3}$, $C3$ solutions cannot occur outside the interval $K_1 \leq k \leq K_2$. We note that $K_1 = (1/3)(m^2 + 1)\{m - \sqrt{m^2 - 3}\}$, $K_2 = (1/3)(m^2 + 1)\{m + \sqrt{m^2 - 3}\}$, $K_0 = (1 + m^2)^2/(4m)$.

For $m = \sqrt{3}$, $K_0 = K_1 = K_2 = (4/3)\sqrt{3}$ and $T(\sqrt{3}, k) = -144(k - K_0)^4$, so there cannot be a $C3$ solution when $k \neq K_0$.

When $B^2 = AC$, equation (4.13) has a double root and there is only one value of α . If we take $m > \sqrt{3}$ and examine the solution corresponding to $k = K_2$, then D is nonzero in general and we get a single solution for (α, β, γ) . If we write L for $\sqrt{m^2 - 3}$, then

$$\alpha = (3m^2 - 1) \frac{2m + L}{1 + m^2}, \tag{5.5}$$

$$\beta = \frac{(7m^4 - 6m^2 + 3) + 4m(m^2 - 1)L}{1 + m^2}, \tag{5.6}$$

$$\gamma = \frac{1}{3}(1 + m^2)(2m - L), \tag{5.7}$$

and it turns out that $A = \gamma^2$, $B = \gamma^2\alpha$, $C = \gamma^2\alpha^2$. A more important result is that $S_3 = 1$ for $k = K_2$. When $k = K_1$, L has to be replaced by $-L$ in equations (5.5), (5.6) and (5.7) but the solution follows the same lines, with a single solution for (α, β, γ) and with $S_3 = 1$.

When $k = K_0$, $A = 4K_0^2$, $B = 16mK_0^2$, $C = 64m^2K_0^2$ and $\alpha = 4m$. However, $D = 0$ and equations (4.9) and (4.10) for β and γ become indeterminate. If we turn to equation (4.17) and put $\gamma_1 = m\gamma/(1 + m^2)$, we get

$$4\gamma_1^2 - 2\gamma_1(3m^2 - 1) + (1 + m^2)^2 = 0, \tag{5.8}$$

an equation which has real, distinct roots only for $m^2 > 3$. This means that there are no $C3$ solutions for $m < \sqrt{3}$. For $m = \sqrt{3}$, equation (5.8) has a double root $\gamma_1 = 2$ and this leads to a unique solution for (α, β, γ) . The actual values are $\alpha = 4\sqrt{3}$, $\beta = 12$, $\gamma = (8/3)\sqrt{3}$ and they give a $C3$ solution $b_1 = 4.48$, $b_2 = 0.54$, $b_3 = 1.91$, with $S_3 = 1$. For $m > \sqrt{3}$, there are two solutions, with the same value for α but with different values for β and γ . For each solution the stability criterion can be expressed as a function of m . As $m \rightarrow \infty$, the stability criterion approaches $16 + 8\sqrt{5}$ in one case and $16 - 8\sqrt{5}$ in the other.

For $B^2 > AC$, equation (4.13) gives two real values for α and this leads to two distinct solutions for α, β and γ . The values for S_3 and for the solution elements (b_1, b_2, b_3) are also distinct. To illustrate what happens, Table 1 shows the solution details for $m = 2$, for a number of values of k . The entries have been rounded off to four decimal places to make the table more compact. It is clear that the solutions can be classed into two families from the values for S_3 .

For one family, labelled the unstable family in Table 1, $S_3 > 1$ for $K_1 < k < K_2$ and thus all of the solutions in this family are unstable. For the other family,

TABLE 1. C3 solutions for $m = 2$ ($K_1 = 5/3, K_0 = 25/8, K_2 = 5$)

k	"Stable" family						"Unstable" family					
	α	b_1	β	b_2	γ	S_3	α	b_1	β	b_2	γ	S_3
5/3	6.6	3.3195	13.4	1.2160	8.3333	+1.0	6.6	3.3195	13.4	1.2160	8.3333	1.0
1.75	6.5031	3.4810	12.6311	1.0960	7.3486	-0.9970	6.7980	3.3539	14.2797	1.2354	9.1514	3.3760
2.00	6.5641	3.8575	12.0657	0.8988	6.2679	-1.9306	7.0752	3.5626	15.2458	1.1622	9.7321	5.4306
2.50	7.0	4.6511	12.0	0.6228	5.0	-1.6	7.5	4.0	16.5	1.0	10.0	6.4
3.00	7.7603	5.5980	12.8496	0.4302	4.1716	-0.9755	7.8983	4.4584	17.5406	0.8518	9.8284	5.9384
3.125	8.0	5.8642	13.2087	0.3923	4.0109	-0.8400	8.0	4.5792	17.7913	0.8167	9.7391	5.7296
3.50	8.8270	6.7282	14.6599	0.2997	3.6277	-0.4959	8.3201	4.9649	18.5460	0.7150	9.3723	4.9974
4.00	10.0869	7.9479	17.4224	0.2199	3.3542	-0.1507	8.8096	5.5642	19.6120	0.5838	8.6458	3.3397
4.50	11.1991	8.9209	20.7092	0.1841	3.4384	+0.1535	9.4547	6.3611	20.8678	0.4496	7.5616	2.7326
4.75	11.4859	9.0874	22.2067	0.1854	3.7293	+0.3556	9.9200	6.9382	21.6646	0.3742	6.7707	2.0867
5.00	11.0	8.3022	23.0	0.2456	5.0	+1.0	11.0	8.3022	23.0	0.2456	5.0	1.0

labelled as the stable family in Table 1, S_3 decreases at first as k increases and there is a small interval in which the solutions are stable (until $S_3 = -1$). Beyond that, S_3 decreases to about -2 and then increases again until $S_3 = 1$ at $k = K_2$. From the table, the solutions in this family become stable again around $k = 3$ and remain stable until $k = 5 = K_2$. Thus the "stable" family has stable solutions in two intervals (along the k -axis) although outside these intervals the solutions are unstable. The boundaries for these stability intervals were determined more precisely by interpolation once the general character of the behaviour became clear.

A minor point about Table 1 is that although there is only one solution for $k = K_1$ and $k = K_2$, the solution has been entered under both families since it is a limiting case for both. Another point of detail is that for $K_1 < k < K_0$ the larger α -value occurs for the unstable family but there is a change-over at $k = K_0$ and the larger α -value occurs for the stable family for $K_0 < k < K_2$. Thus we cannot simply pick out the larger α -value as our criterion in classifying the solutions.

Similar surveys were made for $m = 1.75, 1.8, 1.9, 2.5, 3.0$ and 4.0 and in each case the values of S_3 for the unstable family of solutions were greater than 1 for $K_1 < k < K_2$. For the other family, S_3 decreased to a minimum and then increased again, although for $m = 1.75$ and $m = 1.8$ the minimum was greater than -1 . Thus for these values of m the stability interval is (K_1, K_2) . This suggested that there ought to be a value of m , say m_0 , for which the minimum value of S_3 is -1 and closer checking gave $m_0 = 1.8843457$, with the minimum occurring for $k = 2.1195885$. For $\sqrt{3} < m \leq m_0$, we can say that the stable family of solutions has $|S_3| \leq 1$ for $K_1 \leq k \leq K_2$, whereas for $m > m_0$ the stable family of solutions has $|S_3| \leq 1$ for $K_1 \leq k \leq K_1^*$ and for $K_2^* \leq k \leq K_2$, where $K_1^* < K_2^*$ and K_1^*, K_2^* are the values of k at which $S_3 = -1$. Table 2 shows the critical values k^*, k_5 and k_6 for C2 solutions and also K_1, K_1^*, K_0, K_2^* , and K_2 for different values of m .

Even for the limited range of values of m in Table 2 it is clear that the behaviour of the solutions is very different from that described by Collet and Eckmann [1] for one-parameter difference equations. The stability intervals (k^*, k_5) and (K_1, K_1^*) appear to be decreasing as m increases, and for larger values of k we have the reappearance of stable C2 and C3 solutions after an interval of instability.

To some extent, the appearance of the table can be explained by looking at the behaviour of k^*, k_5, k_6, K_1, K_0 and K_2 for large values of m . From the expressions for k^*, k_5 and K_1 , all three behave like $m/2$ for m large and

$$k_5 - k^* \sim K_1 - k_5 \sim \frac{1}{4m}.$$

TABLE 2. Critical values of k (for a given m).

m	C2 Solutions			C3 Solutions				
	k^*	k_5	k_6	K_1	K_1^*	K_0	K_2^*	K_2
1.0	0.8284	2.0000	2.0000	-	-	1.0000	-	-
1.5	0.9840	1.2414	8.5086	-	-	1.7604	-	-
$\sqrt{3}$	1.0718	1.2713	12.5851	2.3094	-	2.3094	-	2.3094
1.8	1.0987	1.2861	13.9779	1.8516	-	2.4969	-	3.2364
1.9	1.1391	1.3113	16.2067	1.7195	1.9405	2.7963	2.3626	4.1198
2.0	1.1803	1.3397	18.6603	1.6667	1.7503	3.1250	2.9784	5.0000
2.5	1.3962	1.5132	34.7368	1.6850	1.7014	5.2563	5.6889	10.3984
3.0	1.6228	1.7157	58.2843	1.8350	1.8432	8.3333	9.2799	18.1650
4.0	2.0928	2.1593	133.8407	2.2352	2.2391	18.0625	20.4424	43.0981

In the same way, the dominant terms for k_6 , K_0 and K_2 are $2m^3$, $(1/4)m^3$ and $(2/3)m^3$, respectively, so for large values of m , $k_6 - K_2$ and $K_2 - K_0$ are dominated by a term proportional to m^3 . It was noted earlier that the smaller S_3 value for $k = K_0$ tends to $16 - 8\sqrt{5} = -1.8885$ as $m \rightarrow \infty$, so we can expect to have K_2^* between K_0 and K_2 for larger values of m . From the table, this is already happening at $m = 2.5$.

6. Superstable solutions

The discussion of C2 and C3 solutions raises a number of questions concerning solutions with longer periods. For example, unstable C2 solutions only appear for $m > 1$ and unstable C3 solutions only appear for $m > 1.8843457$. Are there similar critical values for C4, C8, C16, ..., C5 solutions? One would guess that this is the case and that the critical m -values lie between 1 and 1.8843457.

If we take m as fixed, say $m = 2$, and let k vary then equation (2.1) can be regarded as a one-parameter first order difference equation, although the mapping is not on to a fixed interval but on to an interval $(0, 2k)$ which depends on the parameter k . It looks as if the usual arguments about bifurcations could be applied for the smaller critical values of k but what happens for the larger critical values? Do stable C3 solutions precede stable C5 solutions and so on, with eventually a coalescence of stable C4 solutions into C2 solutions when $k = k_6$? If so, does the same Feigenbaum ratio apply for k large as for k small?

One way of obtaining guidance on these questions might be to look at the superstable solutions, where the stability criterion is zero. From equation (2.6), S_1 is zero when $Y = m$ and this implies that $m = 2k$. From equation (3.4), S_2 is zero when $2k = (1 + m^2)/m$. Since $S_2 = F'(b_1)F'(b_2)$, one of these derivatives must be zero and this means that either $b_1 = m$ or $b_2 = m$. If we take $b_1 = m$, then equation (2.1) gives $b_2 = 2k$. For $m \neq 2k$, the superstable C2 solution is then

$(m, 2k)$ and the requirement that $m = b_1 = F(b_2)$ verifies the result obtained from equation (3.4). The example mentioned at the end of Section 3 is the particular case where $m = 3$.

In the same way, a $C3$ solution with $S_3 = 0$ must have m as one of its elements. We can take $b_1 = m$, $b_2 = 2k$, $b_3 = 2k/\{1 + (2k - m)^2\}$ and to complete the cycle we must have $b_1 = F(b_3)$. If we put $l = 2k - m$, the condition that $b_1 = F(b_3)$ leads to the equation

$$l(1 + l^2)^2 = l^2m(1 - lm)^2. \quad (6.1)$$

If $l = 0$, then $m = 2k$ and we have the equilibrium case $b_1 = b_2 = b_3 = m$. For a $C3$ solution we can omit the factor l in equation (3.1) and the result is

$$l^4 - m^3l^3 + 2l^2(1 + m^2) - lm + 1 = 0. \quad (6.2)$$

This equation has a double root when $m = 1.773471$ and the numerical evidence is that this is the smallest value of m for which $S_3 = 0$. (The corresponding value of k is 2.246115.) For $m = 2$, equation (6.2) gives $k = 1.6840$ and $k = 4.2534$, which agree with the critical values in Table 2. For $m = 3$, the corresponding values of k are 1.8371 and 14.6211, which again agree with Table 2. For $m = 3$ and $k = 1.8371$, equation (2.1) has a stable equilibrium point at $Y = 0.51047$ as well as the superstable $C3$ solution.

For a superstable $C4$ solution, we must again have $b_1 = m$, $b_2 = 2k$, $b_3 = 2k/\{1 + (2k - m)^2\}$, with $b_4 = F(b_3)$ and $b_5 = b_1 = m = F(b_4)$. The condition that $b_5 = m$ leads to an equation

$$l\{(1 + l^2)^2 + l^2(1 - lm)^2\}^2 = ml^2\{(1 + l^2)^2 - lm(1 - lm)^2\}^2, \quad (6.3)$$

where again $l = 2k - m$. However, equation (6.3) includes the solutions $l = 0$ and $l = 1/m$, which correspond to the equilibrium case and the $C2$ case. When we take out a factor $l(1 - lm)$, the condition for $C4$ solutions is an equation of degree 8 in l , namely

$$(1 + l^2)^4 + 2l^2(1 + m^2)(1 + l^2)^2(1 - lm) + (l^4 - l^3m^3)(1 - lm)^3 = 0. \quad (6.4)$$

For $m = 2$, this equation has roots at $l = 0.7314$ and $l = 24.5668$ and these correspond to $k = 1.3657$ and $k = 13.2834$. Thus there is one solution between k_5 and K_1 and another between K_2 and k_6 , which agrees with the idea that the stable $C4$ solutions occur between the stable $C2$ and $C3$ solutions at the lower end of the k -scale and between the stable $C3$ and $C2$ solutions at the upper end.

However, this tidy arrangement breaks down for larger values of m . For $m = 3$, there are superstable solutions at $k = 1.7285$, 1.9183, 5.8828 and 40.0081, which suggests that there are four intervals for k where the $C4$ solutions are stable. This behaviour occurs also for $m = 4$ and $m = 2.5$ and the change over to more than two superstable solutions appears to occur between $m = 2.3131$ and $m = 2.3132$.

7. Solving a cubic equation

Although the results for equation (2.1) are of interest because they are so different from the usual results for iterated mapping of an interval they also throw some light on the problem of solving a cubic equation by using equation (1.3). Coppel's conditions ensure monotonic convergence, which could be useful in some cases, but this restriction is not essential and an iteration process where $S_1 = -0.1$ would usually be preferable to one with $S_1 = 0.9$, even although the convergence is not monotonic when S_1 is negative. In the present investigation the problem of finding roots of a cubic came up in Sections 4 and 5, where b_1, b_2, b_3 had to be found from the values of α, β and γ . In some cases, b_1, b_2 and b_3 varied by a factor of 1000 or more and little was known about their location to begin with. The method that was used involved essentially transforming the cubic to a standard form and using a simple formula to give a first approximation to a root of the standard form. The iteration procedure then provided improved approximations.

As before, we can take the cubic as $p(x) = x^3 + ax^2 + bx + c$. If $u = x + (a/3)$, then

$$p(x) = p_2(u) = u^3 + b_2u + c_2, \quad (7.1)$$

with

$$b_2 = b - (a^2/3), \quad c_2 = c - \frac{ab}{3} + \frac{2a^3}{27}. \quad (7.2)$$

If $b_2 = 0$, we have $u = (-c_2)^{1/3}$ immediately (and this covers the case of a triple root). If $c_2 = 0$, then $u = 0$ is a solution and $u^2 + b_2 = 0$ gives the other roots. With these special cases out of the way we can take b_2 and c_2 as nonzero in the general discussion and the main classification is then (i) $b_2 > 0$, (ii) $b_2 < 0$.

If $b_2 = q^2 > 0$, with $q > 0$, we can put $u = qv$ and get $p_2(u) = q^3 p_3(v)$, where

$$p_3(v) = v^3 + v - 2K,$$

with $2K = -c_2/q^3$. This gives the equation to be solved in a standardised form

$$v^3 + v = 2K, \quad (7.3)$$

and we can take K as positive. Otherwise, replacing v by $-v$ gives the form we want. This equation has a single real root, say V , with $V > 0$, and it looks at first sight as if we could use equation (2.1) with $m = 0$ and k replaced by K . However, this would not work in all cases, since the equilibrium solution is unstable for $K > 1$, from equation (2.8). So we are forced to consider a change of origin as a way of keeping S_1 close to zero in the iteration.

From equation (2.12), $p_3(v)$ does not satisfy Coppel's conditions and it can be shown that the shift of the origin needed to make them apply must be at least $2V$ to the left. If we put $v = w - 2rV$, then we have a shift of origin to the left for $r > 0$ and to the right for $r < 0$. With this change of origin, the equation for w can be written as

$$w\{(w - 3rV)^2 + (1 + 3r^2V^2)\} = 2L_1, \tag{7.4}$$

where

$$2L_1 = V(1 + 2r) + V^3(1 + 8r^3). \tag{7.5}$$

This makes use of the equation $p_3(V) = 0$ to replace $2K$ by $V + V^3$. The real root of equation (7.4) is $V(1 + 2r)$ and the corresponding difference equation is

$$w_{n+1} = \frac{2L_1}{(1 + 3r^2V^2) + (w_n - 3rV)^2}. \tag{7.6}$$

This can be put into the standard form (equation (2.1)) by a change of scale and the stability criterion for the equilibrium solution works out as

$$S_1 = \frac{2(2r + 1)(r - 1)V^2}{1 + (4r^2 - 2r + 1)V^2}. \tag{7.7}$$

We note that $S_1 > 0$ for $r > 1$, with $S_1 \rightarrow 1$ as $r \rightarrow \infty$, and this is consistent with the point mentioned above, that Coppel's conditions are satisfied for $r > 1$. Since V is not known to begin with, any shift of origin to satisfy Coppel's conditions is liable to err on the safe side and this increases S_1 . For example, if $r = 2$ and $V = 1$, then $S_1 = 5/7$, which implies a comparatively slow convergence.

From equation (7.7), $S_1 = 0$ when $r = 1$ and this corresponds to the case where $Y = m = 2k$ in Section 2. We also have $S_1 = 0$ when $r = -1/2$ and this has a very simple explanation, that the origin has been shifted to the right to coincide with the real root we are looking for. As a check, we note that $L_1 = 0$ when $r = -1/2$ (from equation (7.5)) and then $w = 0$ satisfies equation (7.4). This suggests that we should look for an approximation to V , say V_0 , and shift the origin to the right by V_0 or to the left by $2V_0$, so that S_1 is close by zero. In practice, the origin was moved to the right, with $w = v - V_0$ and with

$$V_0 = \begin{cases} 2K, & \text{for } 0 \leq 2K < 0.45, \\ K + 0.16, & \text{for } 0.45 \leq 2K \leq 1.5, \\ (2K - 1)^{1/3}, & \text{for } 2K > 1.5. \end{cases} \tag{7.8}$$

Table 3 illustrates how this approximation works for a number of values of V . For each V , $2K$ was calculated as $V + V^3$ and V_0 was then obtained from K by using equation (7.8). Thus V_0 is our approximation to V and it will be seen that the approximation is fairly good. It does not matter that V_0 is not a continuous

function of K . All we want is a simple rule that can be fed to a computer. The iteration equation is then

$$w_{n+1} = \frac{2K - V_0 - V_0^3}{w_n^2 + 3V_0w_n + (1 + 3V_0^2)}, \tag{7.9}$$

with $w_0 = 0$.

In the case where $b_2 < 0$, we can write $b_2 = -q^2$ with $q > 0$. Then the substitution $u = qv$ gives $p_2(u) = q^3p_4(v)$, where

$$p_4(v) = v^3 - v - 2K. \tag{7.10}$$

As before, $2K = -c_2/q^3$ in equation (7.10) but we can work with the standardised equation

$$v^3 - v = 2K \tag{7.11}$$

and assume K is positive. With this assumption, $p_4(1) < 0$ and there is always a real root $V > 1$. For $K > 1/(3\sqrt{3})$, this is the only real root, but for $0 < K < 1/(3\sqrt{3})$ there are two more real roots between -1 and 1 . For $K = 1/(3\sqrt{3})$, there is a double root at $v = -1/\sqrt{3}$. We can argue that if one root of a cubic is known the problem reduces to solving a quadratic and this suggests that we should concentrate on the root V , which is always real and has the largest magnitude when there are three real roots. Also, we can be sure that it is not a double root.

To use the same strategy as before, we can put $w = v - V_1$, where V_1 is an approximation to the root V . The approximation used in this case was

$$V_1 = (2K + 1)^{1/3} \tag{7.12}$$

and Table 4 shows that this is a reasonably good approximation over a wide range of values of K . As in Table 3, $2K$ was calculated as $V^3 - V$ for each value

TABLE 3. Approximate solution of $v^3 + v = 2K$.

V = exact solution, V_0 = approximation from equation (7.8).

V	0.0	0.2	0.35	0.4	0.7	0.85	0.9	1.5	3.0	10.0	20.0
$2K$	0.0	0.204	0.393	0.464	1.043	1.464	1.629	4.875	30.0	1010.0	8020.0
V_0	0.0	0.204	0.393	0.392	0.682	0.892	0.857	1.571	3.072	10.03	20.02

TABLE 4. Approximate solution of $v^3 - v = 2K$.

V = exact solution, V_1 = approximation $(2K + 1)^{1/3}$.

V	1.0	1.1	1.2	1.4	1.6	1.8	2.0	4.0	10.0	20.0
$2K$	0.0	0.231	0.528	1.344	2.496	4.032	6.0	60.0	990.0	7980.0
V_1	1.0	1.072	1.152	1.328	1.518	1.714	1.913	3.936	9.97	19.98

of V and V_1 was then obtained from equation (7.12). One point which is of importance later is that V_1 is always an underestimate of V . From equation (7.10),

$$p_4(V_1) = 1 - (2K + 1)^{1/3} < 0$$

and this implies $V_1 < V$, since $p'_4(v) = 3v^2 - 1 \geq 2$, for $v \geq 1$.

After the change of origin, the iteration for w becomes

$$w_{n+1} = \frac{2K + V_1 - V_1^3}{(w_n + (3/2)V_1)^2 - 1 + (3/4)V_1^2}, \tag{7.13}$$

with $w_0 = 0$. If we put $2L_2 = 2K + V_1 - V_1^3$, then $2L_2 = -p_4(V_1) > 0$. Equation (7.13) can be put into the standard form (equation (2.1)) by a change of scale, provided $(3/4)V_1^2 - 1 > 0$, that is, provided $V_1 > 2/\sqrt{3} = 1.1547$. However, if $1 < V_1 < 1.1547$, we cannot get to the standard form by a change of scale. This difficulty arises because of an assumption made at the beginning of Section 2, that by a suitable change of origin we can have $a^2 < 4b$ in $p(x) = x^3 + ax^2 + bx + c$. This assumption is correct and we could satisfy it comfortably by using $w = v - V_2$, with

$$V_2 = \text{Max}\{V_1, (4/3)\}.$$

On the other hand, V_2 is not as good as V_1 as an approximation to V , so we can expect $|S_1|$ to be larger and the iteration longer in some cases if we use $w = v - V_2$.

A closer check shows that no difficulty arises if equation (7.13) is used, even when $(3/4)V_1^2 - 1 < 0$. If we rewrite the equation as

$$w_{n+1} = \frac{2L_2}{w_n^2 + 3V_1w_n + 3V_1^2 - 1}, \tag{7.14}$$

then for $w_n \geq 0$ and $V_1 > 1$

$$w_n^2 + 3V_1w_n + 3V_1^2 - 1 > 2, \tag{7.15}$$

and this ensures that $0 < w_{n+1} < L_2$. So if we start with $w_0 = 0$, then $w_n > 0$ for $n = 1, 2, 3, \dots$. Thus although the denominator in equation (7.14) is not positive definite we need not worry about it becoming zero or negative, since this cannot happen as long as $w_n \geq 0$. There will be an equilibrium value $W = W - V_1$ and from Table 4 it is clear that $0 < W < 0.1$. The stability criterion, S_1 , is the value of dw_{n+1}/dw_n when $w_n = W$ and this gives

$$S_1 = \frac{-W(2W + 3V_1)}{W^2 + 3V_1W + 3V_1^2 - 1}. \tag{7.16}$$

We note that $S_1 < 0$, so the convergence is not monotonic.

If we replace V_1 by $V - W$ in equation (7.16),

$$S_1 = \frac{-W(3V - W)}{W^2 - 3VW + 3V^2 - 1}$$

$$\doteq \frac{-3VW}{3V^2 - 1} \quad \text{for } W \text{ small.}$$

With this first order approximation,

$$|S_1| = 3W \frac{V}{3V^2 - 1} \quad (7.17)$$

and it is easy to show that $V/(3V^2 - 1)$ is monotonic decreasing for $V \geq 1$. Thus the first order term gives

$$|S_1| \leq \frac{3}{2} W < 0.15$$

and the available numerical evidence agrees with this.

It is of interest that the example cited by Hildebrand [7] to show how Lin's method could break down was $p(x) = x^3 - x - 1$. For this example, $p(x) = 0$ has a single real root at $x = 1.324718$, so Hildebrand started with $x_0 = 1.3$ and used the iteration formula

$$x_{n+1} = \frac{1}{x_n^2 - 1}, \quad (7.18)$$

which produces fluctuations of increasingly large amplitude. In contrast, equation (7.12) gives $X = 2^{1/3} \doteq 1.26$ as an approximation to the solution and the substitution $w = x - 1.26$ led to an iteration formula which gave $x = 1.324718$ after five iterations.

The method suggested above for solving a cubic is similar in principle to that given by Hartree [6] and by Fletcher and others [4], who noted that in general a cubic can be brought to the standard form $x^3 \pm x = c$ and then x can be found by interpolation in tables of $x^3 + x$ or $x^3 - x$. For the work in Section 5, that is determining (b_1, b_2, b_3) from $x^3 - \alpha x^2 + \beta x - \gamma = 0$, the time taken for each solution was about 9 milliseconds. This was using double precision and stopping when $|w_{n+1} - w_n|$ was less than 10^{-11} . Once one root had been found the corresponding linear factor was removed from the cubic so that the remaining roots could be determined by solving a quadratic. This procedure worked in almost all cases but it broke down when the remaining roots formed a double root. In such cases, rounding-off errors changed the double root to real roots with a small separation or to complex roots with a small imaginary part. This was covered in the computing programme by flagging cases where the roots were close together, so that these cases could be examined separately. (Apart from the cubics which arose in Section 5, a variety of examples was tried to test the procedure.)

Acknowledgements

As mentioned earlier, the suggestion that I look into this problem came from Mr. W. A. Coppel and I should like to thank him for this suggestion and for discussions while the work was in progress. The work was carried out as a Visiting Fellow in the Research School of Physical Sciences, Australian National University, and I am grateful to the Department of Theoretical Physics for the facilities it has provided.

References

- [1] P. Collet and J. P. Eckmann, *Iterated maps on the interval as dynamical systems* (Birkhäuser, Boston, 1980).
- [2] W. A. Coppel, "The solution of cubic equations by iteration", *Z. Angew. Math. Phys.* 9a (1958), 380–383.
- [3] W. A. Coppel, private communication.
- [4] A. Fletcher, J. C. P. Miller, L. Rosenhead and L. J. Comrie, *An index of mathematical tables* (Blackwell Scientific Publications, Oxford, 2nd edition, 1962), Section 5.64.
- [5] J. C. Frauenthal, *Introduction to population modeling* (Birkhäuser, Boston, 1980).
- [6] D. R. Hartree, *Numerical analysis* (Clarendon Press, Oxford, 1952), 200.
- [7] F. B. Hildebrand, *Introduction to numerical analysis* (McGraw-Hill, New York, 1956), 454–458.
- [8] S. N. Lin, "A method of successive approximations of evaluating the real and complex roots of cubic and higher-order equations", *J. Math. Phys.* 20 (1941), 231–242.
- [9] S. N. Lin, "A method for finding roots of algebraic equations", *J. Math. Phys.* 22 (1943), 60–77.
- [10] R. M. May, "Simple mathematical models with very complicated dynamics", *Nature* 261 (1976), 459–467.
- [11] H. E. Salzer, C. H. Richards and I. Arsham, *Table for the solution of cubic equations* (McGraw-Hill, New York, 1958).