Singularities in Bairstow's method

A. Sofo* and A. Glasson* **

Abstract

It is shown that the nature and location of points at which Bairstow's method becomes undefined depend on elementary properties of the polynomial to which it is applied. Examples are given that illustrate the dynamics of Bairstow's method when a singularity occurs at a solution point, and a linear convergence rate is proved for polynomials with a repeated irreducible quadratic factor.

1. Introduction

If the polynomial $P(t)$ with real coefficients is divided by the quadratic $t^2 - xt - y$ with real coefficients x and y, then

$$P(t) = (t^2 - xt - y)Q(t) + tF(x, y) + G(x, y), \tag{1}$$

where the coefficients of the quotient polynomial $Q(t)$ also depend on x and y. If $(x, y) = (x^*, y^*)$ is a solution of the nonlinear equations

$$F(x, y) = 0, \quad G(x, y) = 0 \tag{2}$$

then $t^2 - x^*t - y^*$ is a quadratic factor of $P(t)$. Newton's method [3] for solving this system of two equations in two unknowns generates the sequence of points (x_n, y_n), starting with (x_0, y_0), from

$$x_{n+1} = f(x_n, y_n), \quad y_{n+1} = g(x_n, y_n), \quad \text{for } n = 0, 1, 2, 3, \dots,$$

where

$$\begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} \dfrac{\partial F(x,y)}{\partial x} & \dfrac{\partial F(x,y)}{\partial y} \\ \dfrac{\partial G(x,y)}{\partial x} & \dfrac{\partial G(x,y)}{\partial y} \end{pmatrix}^{-1} \begin{pmatrix} F(x,y) \\ G(x,y) \end{pmatrix}. \tag{3}$$

In this context the algorithm is referred to as Bairstow's method for finding a quadratic factor of a polynomial, which has proven to be an efficient way to numerically find all roots of a real polynomial working only in the reals. It necessarily exhibits the Newton method property of quadratic convergence rate in a

Received 10 March 2009; accepted for publication 21 November 2009.

*School of Engineering and Science, Victoria University, PO Box 14428 Melbourne, VIC 8001. E-mail: Anthony.Sofo@vu.edu.au

**E-mail: aglas1@bigpond.net.au

neighbourhood of (x^*, y^*) provided the Jacobian matrix in (3) is nonsingular when evaluated at this solution point. A quadratic convergence rate means the error at each stage of the iteration is proportional to the square of the previous error. This is more desirable in numerical calculations than a linear convergence rate when the error at each stage of the iteration is proportional to the previous error.

In this paper we consider (3) as the two-dimensional dynamical system

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} - \frac{1}{J(x,y)} \begin{pmatrix} J_1(x,y) \\ J_2(x,y) \end{pmatrix}, \tag{4}$$

where, omitting the dependence of the functions F and G on the variables x and y,

$$J(x,y) = \frac{\partial F}{\partial x}\frac{\partial G}{\partial y} - \frac{\partial F}{\partial y}\frac{\partial G}{\partial x}, \quad J_1(x,y) = F\frac{\partial G}{\partial y} - G\frac{\partial F}{\partial y},$$

$$J_2(x,y) = G\frac{\partial F}{\partial x} - F\frac{\partial G}{\partial x}$$

and focus on necessary and sufficient conditions for a singular Jacobian matrix, particularly when the singularity occurs at a solution point of (2) which corresponds to a quadratic factor of $P(t)$.

2. Conditions for a singular Jacobian

Using a prime to denote differentiation with respect to t we obtain from (1)

$$P'(t) = (t^2 - xt - y)Q'(t) + (2t - x)Q(t) + F. \tag{5}$$

If

$$t^2 - xt - y = (t - \alpha)(t - \beta) \quad \text{then } x = \alpha + \beta \text{ and } y = -\alpha\beta.$$

Substituting $t = \alpha$ and $t = \beta$ into (1) and (5) establishes the identities, for $\beta \neq \alpha$,

$$(\alpha - \beta)F = P(\alpha) - P(\beta), \quad (\alpha - \beta)G = \alpha P(\beta) - \beta P(\alpha)$$

$$(\alpha - \beta)Q(\alpha) = P'(\alpha) - F, \quad (\alpha - \beta)Q(\beta) = F - P'(\beta).$$

The Jacobian matrix in (3) is singular when $J(x,y) = 0$ where, for all α and β,

$$J(x,y) = Q(\alpha)Q(\beta)$$

is a consequence of matrix identities derived by Fiala and Krebsz [1] (this formula can also be derived for $\beta \neq \alpha$ using the identities in Section 4). They prove that the Jacobian is singular if and only if the quadratic divisor and quotient polynomial in (1) have a common linear or quadratic factor. The following theorems complement this result by considering the nature of the singularities in different regions of the plane, and where possible relate them to geometric properties of the graph of the polynomial.

Theorem 1. *When $x^2 + 4y > 0$, then $\beta \neq \alpha$ are both real, and there are points at which $J(x,y) = 0$ if and only if the graph of $P(t)$ has at least one inflection point.*

Proof. If $J(x,y)=0$ then

$$P'(\alpha)=\frac{P(\alpha)-P(\beta)}{\alpha-\beta} \quad \text{or} \quad P'(\beta)=\frac{P(\alpha)-P(\beta)}{\alpha-\beta}.$$

Geometrically this requires the chord joining two points, $(\alpha, P(\alpha))$ and $(\beta, P(\beta))$, on the graph of $P(t)$, be tangential to the graph at one, or both, of the points. This is impossible if the graph is concave or convex everywhere in the interval $\alpha \le t \le \beta$, so there must be at least one point on the graph where the curvature changes sign.

If the graph of $P(t)$ has an inflection point at $t=\gamma$ and it is the only one in the interval $t_1 < t < t_2$ then we can find an α^* satisfying $t_1 < \alpha^* < \gamma$ such that the tangent at $(\alpha^*, P(\alpha^*))$ intersects the graph at $(\beta^*, P(\beta^*))$ where $\gamma < \beta^* < t_2$ and so $J(\alpha^*+\beta^*, -\alpha^*\beta^*)=0$. All α satisfying $\alpha^* < \alpha < \gamma$ have the same property, with corresponding $\beta=\beta(\alpha)$ values satisfying $\gamma < \beta(\alpha) < \beta^*$ which gives $J(x,y)=0$ on a curve defined parametrically by $(x,y)=(\alpha+\beta(\alpha), -\alpha\beta(\alpha))$. The point $(2\gamma, -\gamma^2)$ lies on the curve which is continuous at this point. This does not preclude the tangent intersecting the graph at additional points and applies to each inflection point if there is more than one. □

Theorem 2. *When $x^2+4y<0$, then $\beta=\overline{\alpha}$, the complex conjugate of α, and $J(\alpha+\overline{\alpha}, -\alpha\overline{\alpha})=0$ if and only if $P(t)=(t^2-[\alpha+\overline{\alpha}]t+\alpha\overline{\alpha})^2R(t)+at+b$ for some real constants a and b. These points are isolated zeros of $J(x,y)$.*

Proof. In this case, if $J(x,y)=0$ then $Q(\alpha)=Q(\overline{\alpha})=0$. Hence

$$P'(\alpha)=P'(\overline{\alpha})=\overline{P'(\alpha)}=a$$

and

$$P(\alpha)-\alpha P'(\alpha)=P(\overline{\alpha})-\overline{\alpha}P'(\overline{\alpha})=\overline{P(\alpha)-\alpha P'(\alpha)}=b,$$

where a and b are real constants. Now

$$\begin{aligned} P(t) &= P(\alpha)+(t-\alpha)P'(\alpha)+(t-\alpha)^2\left[\frac{1}{2!}P''(\alpha)+\frac{(t-\alpha)}{3!}P'''(\alpha)+\cdots\right] \\ &= b+at+(t-\alpha)^2R_1(t) \end{aligned}$$

and similarly

$$P(t)=b+at+(t-\overline{\alpha})^2\overline{R_1}(t).$$

Combining the last two expressions for $P(t)$ we obtain

$$P(t)=b+at+[(t-\alpha)(t-\overline{\alpha})]^2R(t).$$

Conversely, if $P(t)$ has this latter form then $P'(\alpha)=P'(\overline{\alpha})$ and $P(\alpha)-\alpha P'(\alpha)=P(\overline{\alpha})-\overline{\alpha}P'(\overline{\alpha})$. It follows that $Q(\alpha)=0$ and $J(\alpha+\overline{\alpha}, -\alpha\overline{\alpha})=0$.

If

$$\begin{aligned} R_1(t) &= (t-\alpha)^{m-2}p(t-\alpha) \\ &= (t-\alpha)^{m-2}[p(0)+(t-\alpha)q(t-\alpha)] \quad \text{for } m\ge 2 \text{ and } p(0)\ne 0 \end{aligned}$$

then for complex $\epsilon \neq 0$,

$$
\begin{aligned}
&(\alpha - \overline{\alpha} + \epsilon - \overline{\epsilon})^2 Q(\alpha+\epsilon) \\
&\quad = \epsilon^{m-1}\Bigg\{ m(\alpha-\overline{\alpha})p(0) + (\epsilon - \overline{\epsilon})\left[mp(\epsilon) + \epsilon p'(\epsilon) - \epsilon p(\epsilon) + \overline{\epsilon}\left(\frac{\overline{\epsilon}}{\epsilon}\right)^{m-1}\overline{p(\epsilon)}\right] \\
&\qquad\qquad + \epsilon(\alpha-\overline{\alpha})\left[mq(\epsilon) + p'(\epsilon) - p(\epsilon) + \left(\frac{\overline{\epsilon}}{\epsilon}\right)^{m}\overline{p(\epsilon)}\right]\Bigg\} \\
&\quad = \epsilon^{m-1}[m(\alpha-\overline{\alpha})p(0) + \delta(\epsilon)],
\end{aligned}
$$

where $|\overline{\epsilon}/\epsilon| = 1$ and $\lim_{\epsilon\to 0}\delta(\epsilon) = 0$.

For sufficiently small $|\epsilon|$ we have $|Q(\alpha+\epsilon)| > 0$ and hence there are no other zeros of $J(x, y)$ in a neighbourhood of $(\alpha + \overline{\alpha}, -\alpha\overline{\alpha})$. □

Theorem 3. *When $x^2 + 4y = 0$, then $\beta = \alpha$ is real and $J(2\alpha, -\alpha^2) = 0$ if and only if the graph of $P(t)$ has zero curvature at $t = \alpha$.*

Proof. If $\beta = \alpha$ then

$$P(t) = (t-\alpha)^2 Q(t) + tF + G, \quad P''(t) = (t-\alpha)^2 Q''(t) + 4(t-\alpha)Q'(t) + 2Q(t),$$

and

$$J(2\alpha, -\alpha^2) = [Q(\alpha)]^2 = \tfrac{1}{4}[P''(\alpha)]^2$$

so that $J(2\alpha, -\alpha^2) = 0$ if and only if $P''(\alpha) = 0$ which is the definition of zero curvature of $P(t)$ at $t = \alpha$. It is a consequence of Theorems 1 and 2 that if the graph of $P(t)$ has a point of zero curvature at $(\alpha, P(\alpha))$ which is not an inflection point then $(2\alpha, -\alpha^2)$ is an isolated zero of $J(x, y)$. □

When the polynomial has a repeated linear factor, with multiplicity greater than 2, or a repeated quadratic factor then a singularity occurs at a solution point and Bairstow's method does not converge at a quadratic rate in a neighbourhood of the solution. The examples of the following section indicate alternatives to a quadratic convergence rate in these situations.

3. Examples

Numerical experiments with polynomials having a repeated real linear factor suggest the dynamics of Bairstow's method depends on the parity of the multiplicity of the factor. While it remains to be proved for the general case, Figures 1 and 2 illustrate this observation by considering the number of iterations required for convergence to t^2, the unique quadratic factor of $P(t) = t^3$ and $P(t) = t^4$ with initial x and y values in the interval $[-2, 2]$. Examples 1 and 2 analyse Bairstow's method when applied to these two polynomials, and a similar approach could be used on higher powers. The third example considers the simplest polynomial with a repeated irreducible quadratic factor and is a special case of the theorem in the next section.

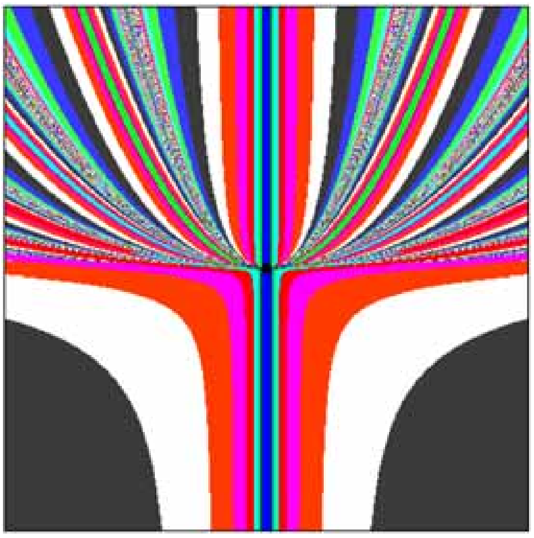

Figure 1. The number of iterations for convergence to $(0,0)$ when Bairstow's method is applied to $P(t) = t^3$ is complicated by the presence of a curve of singularities and points for which the algorithm diverges.

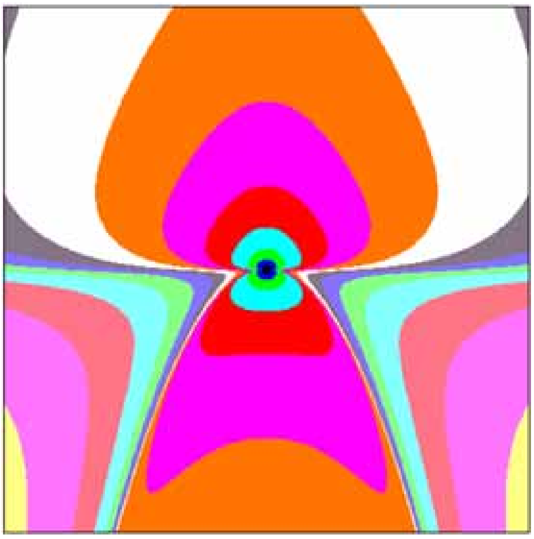

Figure 2. Convergence to $(0,0)$ when Bairstow's method is applied to $P(t) = t^4$ is a global property.

For a given polynomial $P(t) = \sum_{k=0}^{n} a_k t^k$ the two remainder functions $F(x,y)$ and $G(x,y)$ may be calculated, see [2], from the formulas

$$F(x,y) = \sum_{k=0}^{n} a_k F_k, \quad G(x,y) = a_0 + y \sum_{k=1}^{n} a_k F_{k-1},$$

where

$$F_0 = 0, \quad F_1 = 1 \quad \text{and} \quad F_k = xF_{k-1} + yF_{k-2}, \quad k = 2, 3, 4, \ldots$$

are the bivariate Fibonacci polynomials.

Example 1 ($P(t) = t^3$)**.** The graph of this polynomial, as for other odd powers, has an inflection point at $t = 0$. The Bairstow dynamical system (4) is

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{2x^2 - y} \begin{pmatrix} x^3 - xy \\ x^2 y \end{pmatrix}. \tag{6}$$

Consistent with Theorem 1, (6) has a curve of singular points passing through $(0, 0)$. It is convenient to define $B_3(x, y) = (x', y')$. Then $B_3(0, y) = (0, 0)$ for $y \neq 0$ so that convergence occurs in one step. When $x \neq 0$ then $B_3(x, 0) = (x/2, 0)$ indicating a linear convergence rate. For any other point where $y = zx^2$, (6) becomes $B_3(x, y) = (1/(2-z))(x(1-z), y)$ and if $y' = z'(x')^2$ then $z' = z(2-z)/(z-1)^2$. There are two invariant parabolas, obtained by solving $z' = z$, corresponding to $z_1^* = (1 - \sqrt{5})/2$ and $z_2^* = (1 + \sqrt{5})/2$. For the former $B_3(x, y) \approx (0.6x, 0.4y)$ so that convergence to the origin is linear, but on the latter $B_3(x, y) \approx (-1.6x, 2.6y)$ so that Bairstow's method diverges for an initial point on this parabola regardless of how close it is to the origin.

Example 2 ($P(t) = t^4$)**.** The graph of this polynomial, as for other even powers, has a point of zero curvature which is not an inflection point at $t = 0$. The Bairstow dynamical system (4) is

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{4y^2 + 4x^2 y + 3x^4} \begin{pmatrix} 2x(y^2 + x^2 y + x^4) \\ y(2y^2 + 3x^2 y + 2x^4) \end{pmatrix}. \tag{7}$$

Consistent with Theorem 3, (7) has an isolated singularity at $(0, 0)$. Defining $B_4(x, y) = (x', y')$ we have $B_4(0, y) = (0, y/2)$ and $B_4(x, 0) = (2x/3, 0)$. For points not on the axes $y = zx^2$, (7) becomes

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{4z^2 + 4z + 3} \begin{pmatrix} 2x(z^2 + z + 1) \\ y(2z^2 + 3z + 2) \end{pmatrix}.$$

Elementary methods can be used to prove

$$\left| \frac{x'}{x} \right| \leq \frac{3}{4} \quad \text{and} \quad \left| \frac{y'}{y} \right| \leq \frac{4 + \sqrt{2}}{8} < 1 \quad \text{for all } z,$$

and hence Bairstow's method converges to the only solution in a global sense. Further, there are two invariant parabolas corresponding to $z_1^* \approx -0.39$ and $z_2^* \approx -1.54$ on which convergence occurs at a linear rate.

Example 3 ($P(t) = (t^2 + 1)^2$)**.** It is convenient to define $\widetilde{y} = y + 1$ so that the quadratic factor corresponds to $(x, \widetilde{y}) = (0, 0)$ which, consistent with Theorem 2, is an isolated singularity of the transformed (4),

$$\begin{pmatrix} x' \\ \widetilde{y}' \end{pmatrix} = \frac{1}{2} \begin{pmatrix} x \\ \widetilde{y} \end{pmatrix} + \frac{1}{2(4x^2 + 4\widetilde{y}^2 + 4x^2\widetilde{y} + 3x^4)} \begin{pmatrix} x^5 \\ x^4\widetilde{y} + 2x^2\widetilde{y}^2 + 2x^4 \end{pmatrix}$$

and the convergence rate in a neighbourhood of the unique solution is again linear.

4. Linear convergence rate

Differentiating (1) with respect to x and y, substituting $t = \alpha$ and $t = \beta$ gives four equations which, for $\beta \neq \alpha$, solve to

$$(\alpha-\beta)\frac{\partial F}{\partial x} = \alpha Q(\alpha) - \beta Q(\beta), \quad (\alpha-\beta)\frac{\partial F}{\partial y} = Q(\alpha) - Q(\beta),$$
$$(\alpha-\beta)\frac{\partial G}{\partial x} = \alpha\beta[Q(\beta) - Q(\alpha)], \quad (\alpha-\beta)\frac{\partial G}{\partial y} = \alpha Q(\beta) - \beta Q(\alpha),$$

and then after some manipulation

$$(\alpha-\beta)J_1(x,y) = P(\alpha)Q(\beta) - P(\beta)Q(\alpha),$$
$$(\alpha-\beta)J_2(x,y) = \alpha P(\beta)Q(\alpha) - \beta P(\alpha)Q(\beta).$$

Theorem 4. *If the polynomial $P(t)$ has an irreducible quadratic factor $t^2 - x^*t - y^*$ of multiplicity $m \geq 2$ then in a neighbourhood of (x^*, y^*) Bairstow's method converges at a linear rate with the limiting ratio of consecutive error terms equal to $(1 - 1/m)$.*

Proof. Let $t^2 - x^*t - y^* = (t-\alpha)(t-\overline{\alpha})$ and let $P(t) = (t-\alpha)^m[p(0) + (t-\alpha)q(t-\alpha)]$ (using the same notation as in the proof of Theorem 2). Since $(x^*, y^*) = (\alpha + \overline{\alpha}, -\alpha\overline{\alpha})$ then for complex ϵ the point $(x, y) = (\alpha + \overline{\alpha} + \epsilon + \overline{\epsilon}, -(\alpha+\epsilon)(\overline{\alpha}+\overline{\epsilon}))$ is taken as a point arbitrarily close to the solution point. Then

$$\begin{aligned}
\frac{J_1(x,y)}{J(x,y)} &= \frac{P(\alpha+\epsilon)}{(\alpha-\overline{\alpha}+\epsilon-\overline{\epsilon})Q(\alpha+\epsilon)} + \frac{P(\overline{\alpha}+\overline{\epsilon})}{(\overline{\alpha}-\alpha+\overline{\epsilon}-\epsilon)Q(\overline{\alpha}+\overline{\epsilon})} \\
&= \frac{(\alpha-\overline{\alpha}+\epsilon-\overline{\epsilon})\epsilon[p(0)+\epsilon q(\epsilon)]}{m(\alpha-\overline{\alpha})p(0)+\delta(\epsilon)} + \frac{(\overline{\alpha}-\alpha+\overline{\epsilon}-\epsilon)\overline{\epsilon}[\overline{p(0)}+\overline{\epsilon q(\epsilon)}]}{m(\overline{\alpha}-\alpha)\overline{p(0)}+\overline{\delta(\epsilon)}} \\
&= \frac{\epsilon[1+\delta_1(\epsilon)]}{m} + \frac{\overline{\epsilon}[1+\overline{\delta_1(\epsilon)}]}{m},
\end{aligned}$$

where $\lim_{\epsilon\to 0}\delta_1(\epsilon) = 0$. Similarly

$$\begin{aligned}
\frac{J_2(x,y)}{J(x,y)} &= \frac{(\alpha+\epsilon)P(\overline{\alpha}+\overline{\epsilon})}{(\overline{\alpha}-\alpha+\overline{\epsilon}-\epsilon)Q(\overline{\alpha}+\overline{\epsilon})} + \frac{(\overline{\alpha}+\overline{\epsilon})P(\alpha+\epsilon)}{(\overline{\alpha}-\alpha+\overline{\epsilon}-\epsilon)Q(\alpha+\epsilon)} \\
&= \frac{(\alpha+\epsilon)(\alpha-\overline{\alpha}+\epsilon-\overline{\epsilon})\overline{\epsilon}[\overline{p(0)}+\overline{\epsilon q(\epsilon)}]}{m(\overline{\alpha}-\alpha)\overline{p(0)}+\overline{\delta(\epsilon)}} \\
&\quad + \frac{(\overline{\alpha}+\overline{\epsilon})(\overline{\alpha}-\alpha+\overline{\epsilon}-\epsilon)\epsilon[p(0)+\epsilon q(\epsilon)]}{m(\alpha-\overline{\alpha})p(0)+\delta(\epsilon)} \\
&= -\frac{\overline{\epsilon}[\alpha+\epsilon/2+\delta_2(\epsilon)]}{m} - \frac{\epsilon[\overline{\alpha}+\overline{\epsilon}/2+\overline{\delta_2(\epsilon)}]}{m},
\end{aligned}$$

where $\lim_{\epsilon \to 0} \delta_2(\epsilon) = 0$. Now, from (4), for small ϵ

$$x' - x^* \approx x - x^* - \frac{\epsilon + \bar{\epsilon}}{m} = \left(1 - \frac{1}{m}\right)(x - x^*)$$
$$y' - y^* \approx y - y^* + \frac{1}{m}(\epsilon\bar{\alpha} + \alpha\bar{\epsilon} + \epsilon\bar{\epsilon}) = \left(1 - \frac{1}{m}\right)(y - y^*).$$

This result contrasts with the situation where the repeated quadratic factor has real distinct linear factors when there is a curve of singular points which passes through, and is continuous at, the corresponding solution point. □

References

[1] Fiala, T. and Krebsz, A. (1987). On the convergence and divergence of Bairstow's method. *Numerische Mathematik* **50,** 477–482.
[2] Glasson, A.R. (1995). Remainder formulas involving generalized Fibonacci and Lucas polynomials. *The Fibonacci Quarterly* **33,** 268–272.
[3] Henrici, P. (1964). *Elements of Numerical Analysis.* John Wiley, New York.