The zeros and the critical points of a polynomial: second moments and least squares fits of lines

G. KEADY

1 Introduction

There is a huge literature on the properties of zeros and critical points of univariate polynomials over the complex numbers. This note gives a generalization of results on cubic polynomials presented in [1]. See Figure 1, and Theorems 1 and 2 below, due to Siebeck and Coolidge respectively.

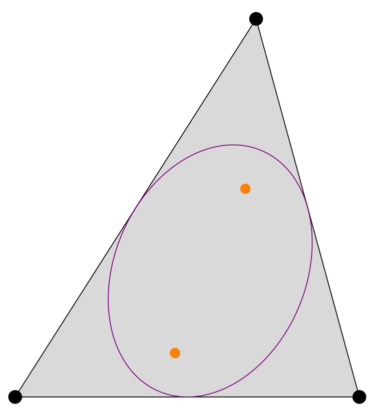

Figure 1: This figure is produced from the code of the “Marden Theorem” entry at the Wolfram demonstrations site. The triangle’s vertices – dark black dots – are to be regarded as the three zeros of a cubic polynomial, p. The pale dots are the zeros of its derivative p'. See Theorem 1.

The results in this note could have been found at any time over the last two centuries and have probably been rediscovered several times. Despite asking, for the last couple of years, both people and google, etc., I have not yet found papers with the results of my Theorem 3 or Corollary 7. Key ingredients, e.g.

equation (4), are presented in several papers. Perhaps readers of this paper will be able to report equivalent earlier work.

2 Triangles and cubic polynomials

Given a triangle T in the plane, there is an inscribed ellipse, tangent to each side at the midpoint of the side. This ellipse is usually called the *Steiner in-ellipse*. Its centre is at the centroid of the triangle, and it is the inscribed ellipse of largest area. The ellipse circumscribed about triangle T with least area is called the *Steiner circum-ellipse*. This also has its centre at the centroid of the triangle, has the same directions for its principal axes as the Steiner in-ellipse, and the lengths of the principal axes are a factor of 2 larger than those of the Steiner in-ellipse. There is a substantial popular literature on Steiner ellipses including books such as [2] (where Problem 98 is particularly relevant) and articles on the internet including in wikipedia and in MathWorld. Recent articles in publications by MAA, [3, 4, 1], and elsewhere, e.g.
`http://demonstrations.wolfram.com/MardensTheorem/`
have called attention to it, and some of its applications. (The Mathematica notebook on the Steiner CircumEllipse at MathWorld is comprehensive.) One of the results publicised in [3, 4, 1, 5] is the following.

Theorem 1 (Siebeck, $n = 3$.) *Suppose that z_1, z_2, z_3, are non-collinear points in the complex plane, and $p(z) = \prod_{j=1}^{3}(z - z_j)$. Then the roots of the derivative $p'(z)$ are the foci of the Steiner in-ellipse for the triangle with vertices at z_1, z_2, z_3.*

A related result, 2.4 in [1], concerns least squares fitting of lines with perpendicular offsets, called 'line of best fit' here:

Theorem 2 (Coolidge, $n = 3$.) *Suppose the triangle with vertices z_1, z_2, z_3 is nonequilateral. If $p(z) = \prod_{j=1}^{3}(z - z_j)$, then the line of best fit is the line through the roots of $p'(z)$.*

In this note we present generalizations of Coolidge's results to other $n > 3$. These are given in Theorems 4 and Theorem 5 and in the Corollary to Theorem 6 below.

3 Polynomials of degree n

3.1 Second moment matrices

Suppose that a set $\{z_j | 1 \le j \le n\}$ of $n \ge 3$ complex numbers is given. Define the first moment, the vertex average, $z_{\rm VA}$, of the set by $z_{\rm VA} = \frac{1}{n}\sum_{j=1}^{n} z_j$. Define also the second moment about a point z_* by

$$I_{\text{point}}(z_*) = \begin{bmatrix} \sum_{j=1}^n (x_j - x_*)^2 & \sum_{j=1}^n (x_j - x_*)(y_j - y_*) \\ \sum_{j=1}^n (x_j - x_*)(y_j - y_*) & \sum_{j=1}^n (y_j - y_*)^2 \end{bmatrix} \tag{1}$$

When no argument for z_* is given, I_{point} means that z_* has been taken to be z_{VA}.

3.2 Second moments for zeros and critical points of a polynomial

Consider next a polynomial of degree n, and, for the sake of definiteness, suppose that all the roots are distinct. Suppose also that the sum of the roots is zero, which is only supposed to save some writing in some future formulae:

$$p(z) = \prod_{j=1}^{n} (z - z_j), \qquad \sum_{j=1}^{n} z_j = 0. \tag{2}$$

We take the coordinates of the points to be represented as pairs of reals or, as $(z_j)_{j=1}^n$ complex numbers, choosing whichever representation is most convenient to the purpose at hand. The usual notation $z = x + iy$ is used. The derivative $p'(z)$ has $(n-1)$ its roots, sometimes called the *critical points* of p.

We began this study because of a connection with 'mechanics'. Suppose that each of these sets of roots is considered to be an equal point mass, pairwise joined by light rigid bars. This leads on to definitions of moments of inertia, as in equation (1). The connection with the Steiner in-ellipse is that if one repeatedly differentiates the polynomial until one reaches a cubic, then finds the Steiner in-ellipse associated with the triangle defined from the roots of the cubic, the principal axes of I_{point} are the axes of the Steiner in-ellipse. We will prove this and note that intermediate results on the way through to it are also memorable.

The derivative, of course is

$$p'(z) = n \prod_{k=1}^{n-1} (z - z'_k), \qquad \sum_{k=1}^{n-1} z'_k = 0. \tag{3}$$

The centroid of the $(n-1)$ zeros of p' and the n zeros of p coincide. There is a considerable literature concerning relationships between the zeros of the derivative and of the original polynomial. There is a nice picture, associated with illustrating the Gauss-Lucas Theorem, at the Mathematica demonstrations site,
`http://demonstrations.wolfram.com/LucasGaussTheorem/`
We will return to this later.

Result 1 *Denote the zeros of the polynomials p and p' as above, and suppose that their centroids are located at the origin. Then*

$$\sum (z'_k)^2 = \frac{(n-2)}{n} \sum z_j^2. \tag{4}$$

Proof. Considering the coefficient of z^{n-2} in p and of z^{n-3} in p', we have

$$n\sum_{k,l\ k\neq l} z'_k z'_l = (n-2)\sum_{j,l\ j\neq l} z_j z_l .$$

From this, on using the centroid conditions, first moment conditions, of equations (2,3) we obtain equation (4).

We are sure that this identity is "well known": see, for example, [6], equation (1.9). (There may be connections to Newton's identities, [7] equation (2), and it might be very old indeed.)

We have already noted that the centroids of the zeros and of the critical points of p coincide. This is a fact about the first moments. Consider next the second-moment matrices:

$$I_{\rm point} = \left[\begin{array}{cc} \sum x_j^2 & \sum x_j y_j \\ \sum x_j y_j & \sum y_j^2 \end{array}\right], \qquad I'_{\rm point} = \left[\begin{array}{cc} \sum (x'_k)^2 & \sum x'_k y'_k \\ \sum x'_k y'_k & \sum (y'_k)^2 \end{array}\right] .$$

Theorem 3 *$I_{\rm point}$ and $I'_{\rm point}$ have the same eigenvectors.*

Proof. We will establish this by showing that the matrices commute. (See [8], items 1.3.19 and 2.3.3. However, we only need the result that real symmetric, and hence diagonalizable, commuting matrices have the same eigenvectors, and the proof is easier then.)

The commutator $(II' - I'I)$ of two symmetric matrices I and I' is necessarily skew symmetric and it is the zero matrix iff

$$I'_{12}(I_{11} - I_{22}) - I_{12}(I'_{11} - I'_{22}) = 0.$$

Using the specific entries for our matrices

$$\begin{aligned} (II' - I'I) = 0 \quad &\Leftrightarrow \quad c := \sum x'_k y'_k \sum (x_j^2 - y_j^2) - \sum x_j y_j \sum ((x'_k)^2 - (y'_k)^2) = 0 \\ &\Leftrightarrow \quad c := \mathrm{Im}\left(\overline{\sum (z'_k)^2} \sum z_j^2\right) = 0. \end{aligned} \qquad (5)$$

Using equation (4) in the expression c of equation (5), we have that

$$\begin{aligned} c &= \frac{(n-2)}{n}\mathrm{Im}\left(|\sum z_j^2|^2\right) \\ &= 0. \end{aligned}$$

This establishes that the matrices $I_{\rm point}$ and $I'_{\rm point}$ commute and hence share eigenvectors.

Merely by way of discussion, see the picture associated with the Gauss-Lucas Theorem, at the Mathematica demonstrations site, the URL for which was given earlier. The picture shows, not just the (convex hulls of the) n-roots and the $(n-1)$-roots associated with p and p' respectively, but also the roots for the higher order derivatives. Eventually we get to a cubic polynomial, a triangle, and its Steiner in-ellipsoid. The axes of this ellipsoid are the principal axes for $I_{\rm point}$, and for $I_{\rm point}^{(k)}$ for any, e.g. k-th, derivative's set of points.

A couple of incidental remarks may be in order here.
Associated with the proof of Result 1 we noted reference [6]. As an aside, we note that the conjecture in [6], established since that paper (see [9, 10, 11, 14]), can be restated in terms of our matrices as follows.

$$(n-2)\ {\rm trace}(I_{\rm point}) \geq n\ {\rm trace}(I'_{\rm point})$$

with the equality sign if and only if the $(2n-1)$ points z_j and z'_k are on a straight line.
The paper [15] treats affinely-regular polygons, called there 'Steiner polygons' . Triangles are affinely regular. A quadrilateral is affinely regular if and only if it is a parallelogram. For any affinely regular polygon there are, with centres at $z_{\rm VA}$, circumscribing and inscribing ellipses which are similar. One generalization of Coolidge's result, Theorem 2, is as follows.

Theorem 4 *([15]) Let p be the vertex polynomial of an affinely-regular polygon. Then the roots of p' are collinear and $z_{\rm VA}$ lies on this line.*

Related material is given in [13]. There are various other places where affinely regular polygons arise: one which leads to nice animations is given in [12].

We will see another generalization of Coolidge's result later in this note.

3.3 Least squares fits of lines to sets of points

The fitting of lines here is done as described at
`http://mathworld.wolfram.com/LeastSquaresFittingPerpendicularOffsets.html`
The least-squares lines pass through the centroid of the set of points. From the derivation there, the slope of the line is given by one of the choices in

$$b = -B \pm \sqrt{B^2+1}$$

where, written in terms of the entries of the second moment matrix $I_{\rm point}$,

$$B = \frac{I_{11}-I_{22}}{2I_{12}}.$$

The $n=3$ case of the following is noted, together with the connections to the Steiner in-ellipse, in [1]. For general n, we have the following.

Theorem 5 *The line of least squares fit is in the direction of an eigenvector of $I_{\rm point}$.*

The proof is just routine algebra.

There is some sloppiness in the statement of the Theorem. It is clear that there need not be a unique line. As an example, let $n \geq 3$ be given. If the points are the set of all n-th roots of unity, any line through the origin is equally good for least-squares fitting with perpendicular offsets.

It is easy to improve on Theorem 5 using the following.

Theorem 6 *([1], 2.3.) Suppose z_j, $1 \leq j \leq n$, are complex numbers, $z_{\rm VA}$ is the centroid, and*

$$Z = \sum_{j=1}^{n} (z_j - z_{\rm VA})^2 = \left(\sum_{j=1}^{n} z_j^2 \right) - n {z_{\rm VA}}^2. \tag{6}$$

(a) If $Z = 0$, then every line through $z_{\rm VA}$ is a line of best fit for the points $z_1, \ldots, z_n$.
(b) If $Z \neq 0$, then the line through $z_{\rm VA}$ that is parallel to the vector from 0 to $\sqrt{Z}$ is the unique line of best fit for $z_1, \ldots, z_n$.

Corollary 7 *The line of least squares fit for the zeros z_j of p coincides with the line of least squares fit for the zeros z'_k of p'.*

The proof is to use equation (4) in the expressions, given in Theorem 6, for the Zs of equations (6) for first the zeros and next the critical points.

Acknowledgements

The author acknowledges input from Peter Scales, giving a result for the area moments of inertia matrix (about $z_{\rm VA}$) for a quadrilateral lamina, which ultimately led to this work on point moments of inertia.

References

[1] D. Minda and S. Phelps, Triangles, Ellipses and Cubic Polynomials *Amer. Math. Monthly* **115** (Oct 2008), 679-689.

[2] H. Dorrie, *100 Great Problems of Elementary Mathematics, Their History and Solution*, translated (from German to English) by D. Antin, Dover, New York, 1965.

[3] D. Kalman, The most marvelous theorem in mathematics, *Journal of Online Mathematics and its Applications*, available at
`http://www.JOMA.org` (for journal),
`http://mathdl.maa.org/mathDL/4/?pa=content&sa=viewDocument&nodeID=1663`
(direct to article).

[4] D. Kalman, An elementary proof of Marden's Theorem, *Amer. Math. Monthly* **115** (Apr 2008), 330-338.

[5] M. Marden, *The Geometry of the Zeros of a Polynomial in a Complex Variable*, American Math. Soc., New York, 1949.

[6] I. J. Schoenberg, A conjectured analogue of Rolle's Theorem for polynomials with real or complex coefficients, *Amer. Math. Monthly* **93** (Jan 1986), 8-13.

[7] S. Rosset, Normalized Symmetric Functions, Newton's Inequalities, and a New Set of Stronger Inequalities *Amer. Math. Monthly* **96** (Nov 1989), 815-819.

[8] R.A. Horn and C.A. Johnson, *Matrix Analysis*, Cambridge U.P., New York, 1985.

[9] W.S. Cheung and T.W. Ng, A companion matrix approach to the study of zeros and critical points of a polynomial, *J. Math. Analysis and Its Application*, **319** (2006), 690-707.

[10] S.M. Malamud, An Analog of the Poincaré Separation Theorem for Normal Matrices and the Gauss-Lucas Theorem, *Functional Analysis and Its Applications* **37** (2003), no.3, 232-235.

[11] S.M. Malamud, Inverse spectral problem for normal matrices and the Gauss-Lucas theorem, *Trans. Amer. Math. Soc.* **357** (2005), 4043-4064.

[12] B. Chow and D. Glickenstein, Semidiscrete geometric flows of polygons, *Amer. Math. Monthly* **114** (Apr 2007), 316-328.

[13] J. Clifford and M. Lachance, A Generalization of the Bôcher-Grace Theorem. arXiv:0910.2446v1, Oct 2009.

[14] R. Pereira, Differentiators and the geometry of polynomials. *J. Math. Anal. Appl.* **285** (2003), no. 1, 336-348.

[15] J. L. Parish, On the derivative of a vertex polynomial, *Forum Geometricorum*, **6** (2006), 285-288.

Grant Keady,
School of Mathematics and Statistics, University of Western Australia
email: keady@maths.uwa.edu.au