A Geometric Interpretation of the Solution of the General Quartic Polynomial

William M. Faucette

Suppose we are given the general polynomial equation of degree n:

$$a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 = 0,$$

where each of the a_i's is a rational number and a_n is not zero. We might ask if the solutions of this equation can be expressed in terms of the coefficients $a_0, \ldots, a_n$ using only the operations of addition, subtraction, multiplication, division, and extraction of roots. One of the principal results of Galois Theory, Abel's theorem, states that such formulas exist for $n \leq 4$ and do not exist for $n \geq 5$. The reader can find a discussion of Abel's theorem in numerous sources, including **[A]**, **[F1]**, **[H1]**, and **[H2]**.

In this article we will first recall the explicit radical solution of cubic polynomials. We will then proceed to discuss the solution of the general quartic polynomial by reduction to an auxiliary cubic equation, the quartic's resolvent cubic. The algebraic solutions presented here appear in section 4.16 of the text **[E]**.

After defining algebraic plane curves and introducing a few facts about them, we will present an interesting algebro-geometric interpretation of the derivation of the resolvent cubic.

I would like to express thanks to Miles Reid for his text [**R**, pp. 22–24] which inspired this article. I would also like to thank Elsa Newman, Bill Rivera, David Smead, Chris Vaughn, and the referee, whose suggestions contributed to the readability of the finished product.

GALOIS SOLUTION OF THE GENERAL CUBIC POLYNOMIAL. Let $P(z) = z^3 + a_1 z^2 + a_2 z + a_3$ be a cubic polynomial with rational coefficients. To simplify the solution we eliminate the quadratic term by setting $z = x - \frac{1}{3}a_1$. Then $P(z)$ takes the form $\tilde{P}(x) = x^3 + px + q$, where p and q are polynomials in the coefficients of $P(z)$. Notice that solving $\tilde{P}(x)$ readily solves $P(z)$.

Let x_1, x_2, and x_3 be the roots of $\tilde{P}(x)$, which we assume to be distinct. Notice that since $\tilde{P}(x) = (x - x_1)(x - x_2)(x - x_3)$ has no quadratic term, the sum of the roots must be zero. Let ω be a primitive cube root of unity and define the *Lagrange resolvents*, $(1, x_1)$, (ω, x_1), and (ω^2, x_1), by

$$\begin{aligned}(1, x_1) &= x_1 + x_2 + x_3 = 0\\ (\omega, x_1) &= x_1 + \omega x_2 + \omega^2 x_3 \qquad (1)\\ (\omega^2, x_1) &= x_1 + \omega^2 x_2 + \omega x_3.\end{aligned}$$

Algebraic manipulation shows that the Lagrange resolvents can be computed in terms of the coefficients of $\tilde{P}(x)$ and the square root of the discriminant of $\tilde{P}(x)$. Solving equations (1) for x_1, x_2, and x_3 gives the roots of $\tilde{P}(x)$ in terms of the Lagrange resolvents. Substituting the value of the Lagrange resolvents into the solutions of (1) yields the zeroes of $\tilde{P}(x)$, from which the zeroes of $P(z)$ can be obtained.

GALOIS SOLUTION OF THE GENERAL QUARTIC POLYNOMIAL. Consider the general quartic with rational coefficients, given by $P(z) = z^4 + a_1 z^3 + a_2 z^2 + a_3 z + a_4$. As with the cubic, we first simplify the polynomial by the substitution $z = x - \frac{1}{4}a_1$, yielding

$$\tilde{P}(x) = x^4 + px^2 + qx + r \tag{2}$$

where p, q, and r are polynomials in the coefficients of $P(z)$.

Let x_1, x_2, x_3, and x_4 be the roots of $\tilde{P}(x)$. Since $\tilde{P}(x) = (x - x_1)(x - x_2)(x - x_3)(x - x_4)$ has no cubic term, the sum of the roots once again must be zero. Define $\alpha = (x_1 + x_2)(x_3 + x_4)$, $\beta = (x_1 + x_3)(x_2 + x_4)$, $\gamma = (x_1 + x_4)(x_2 + x_3)$. Let $h \in \mathbb{Q}[z]$ be the polynomial $h = (z - \alpha)(z - \beta)(z - \gamma)$, the *resolvent cubic of* $\tilde{P}(x)$. A little calculation shows that $h = z^3 - 2pz^2 + (p^2 - 4r)z + q^2$.

By solving this cubic equation using the method in the preceding section, one obtains α, β, and γ. Using

$$0 = (x_1 + x_2) + (x_3 + x_4) \text{ and } \alpha = (x_1 + x_2)(x_3 + x_4)$$
$$0 = (x_1 + x_3) + (x_2 + x_4) \text{ and } \beta = (x_1 + x_3)(x_2 + x_4)$$
$$0 = (x_1 + x_4) + (x_2 + x_3) \text{ and } \gamma = (x_1 + x_4)(x_2 + x_3),$$

one obtains roots of $\tilde{P}(x)$. The zeroes of the original quartic may then be easily obtained. For complete algebraic solutions of the general cubic and quartic polynomials, see [**E**, §4.16], [**W**, §64], and [**B**, 16.4.10 and 16.4.11.1].

EVERYTHING YOU NEED TO KNOW ABOUT ALGEBRAIC PLANE CURVES. To give an algebro-geometric interpretation of the resolvent cubic, we need to introduce a few basic facts about algebraic curves. For a complete introduction to algebraic plane curves, see the text [**F2**].

Let $\mathbb{C}$ denote the field of complex numbers and define the affine complex plane, $\mathbb{A}^2$, to be the set of all ordered pairs (a, b) where $a, b \in \mathbb{C}$. A complex affine plane curve is the locus of zeroes in $\mathbb{A}^2$ of a nonzero polynomial $f \in \mathbb{C}[X, Y]$. The complex projective plane, $\mathbb{P}^2$, is the set of all equivalence classes $[a, b, c]$ of ordered triples $(a, b, c) \in \mathbb{C}^3 \setminus (0, 0, 0)$ under the equivalence relation $(a, b, c) \sim (a', b', c')$ if $(a, b, c) = (\lambda a', \lambda b', \lambda c')$ for some nonzero complex number λ. Notice that if $c \neq 0$, we may divide the three coordinates by c and obtain coordinates $[a, b, 1]$. A complex projective plane curve is the locus of zeroes in $\mathbb{P}^2$ of a nonzero homogeneous polynomial $F \in \mathbb{C}[X, Y, Z]$. The degree of a plane curve is the degree of its defining polynomial. Curves of degrees one, two, three, and four are called lines, conics, cubics, and quartics, respectively.

The affine plane is contained in the projective plane by the inclusion $\mathbb{A}^2 \hookrightarrow \mathbb{P}^2$ given by $(x, y) \mapsto [x, y, 1]$, with the remainder of the projective plane forming the line at infinity, $L_\infty = \{[x, y, 0] \in \mathbb{P}^2\}$. If $f(X,Y)$ is an element of $\mathbb{C}[X,Y]$ of degree d, we can homogenize f by setting $F(X,Y,Z) = Z^d f(X/Z, Y/Z)$. F is then a homogeneous polynomial of degree d. If f defines an affine plane curve C, the projective plane curve defined by F is the *projective closure* of C.

A general conic in $\mathbb{P}^2$ is given as the set of zeroes of an equation

$$F(X,Y,Z) = aX^2 + bXY + cY^2 + dXZ + eYZ + fZ^2, \tag{3}$$

where at least one of these coefficients is nonzero, and this equation is unique up to multiplication by a nonzero constant. A conic with equation (3) is reducible if and only if the equations

$$F_X(X,Y,Z) = F_Y(X,Y,Z) = F_Z(X,Y,Z) = 0$$

have a common solution in $\mathbb{P}^2$, where we use the subscripts to denote partial derivatives. If we let $\mathbf{A}$ be the matrix associated with (3), then

$$\mathbf{A} = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix},$$

and we can rewrite equation (3) as

$$[X \quad Y \quad Z]\mathbf{A}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = 0.$$

The condition that the conic be reducible is equivalent to the condition that this associated matrix $\mathbf{A}$ is singular.

The set of all conics in $\mathbb{P}^2$ forms a five-dimensional projective space $\mathbb{P}^5$ in the following way. A general conic in $\mathbb{P}^2$ is given by an equation of the form (3), where at least one of these coefficients is nonzero, and this equation is unique up to multiplication by a nonzero constant. So, we may identify this conic with the point $[a,b,c,d,e,f] \in \mathbb{P}^5$. From this perspective, the conics in $\mathbb{P}^2$ passing through a given point P in $\mathbb{P}^2$ form a codimension one linear subspace in $\mathbb{P}^5$. That is, if $P = [u,v,w]$, then any conic through P must satisfy $F(u,v,w) = au^2 + buv + cv^2 + duw + evw + fw^2 = 0$, and this is a linear equation in a,b,c,d,e,f. Similarly the condition for a conic to contain points $P_1, P_2, P_3, P_4 \in \mathbb{P}^2$ is given by a system of four linear equations in a,b,c,d,e,f. From elementary linear algebra, the family of conics containing all four points will be a one-dimensional linear subspace of $\mathbb{P}^5$ exactly when these four conditions are linearly independent. We then have the following proposition:

Proposition. *The family of conics containing the distinct points P_1, P_2, P_3, and P_4 is (projective) one-dimensional if and only if P_1, P_2, P_3, and P_4 are noncollinear.*

Proof: Suppose the points are noncollinear. Without loss of generality, we may assume that P_1, P_2, and P_3 are noncollinear. It is sufficient to show there exists a conic containing $P_1, \ldots, P_t$ and not containing $P_{t+1}, \ldots, P_4$ for $t = 1,2,3$.

To produce a conic through P_1 and not through P_2, P_3, and P_4, choose two lines through P_1 and not containing P_2, P_3, or P_4. The union of these two lines is a reducible conic containing P_1 and not containing P_2, P_3, or P_4.

Let l be any line through P_1 not containing P_3 or P_4. Let l' be any line through P_2 not containing P_3 or P_4. The union of l and l' is a reducible conic containing P_1 and P_2, but not containing P_3 and P_4.

We divide the last part of the proof into two cases depending on the relative positions of P_1, P_2, and P_4. First, suppose P_1, P_2, and P_4 are noncollinear. Choose any line l' through P_3 not containing P_4. The union of l' and the line through P_1 and P_2 is then a reducible conic containing P_1, P_2, and P_3 and not P_4. On the other hand, suppose P_1, P_2, and P_4 are collinear. Let l be the line through P_1 and P_3 and let l' be the line through P_2 and P_3. Then the union of l and l' is a reducible conic containing P_1, P_2, and P_3, and not P_4. This shows the family of conics containing P_1, P_2, P_3, and P_4 has dimension one. In this context, a linear subspace of dimension one is called a *pencil*, so this family is a pencil of conics.

Conversely, if P_1, P_2, P_3, and P_4 are collinear, let l be the line containing these four points. Let l' be any line in $\mathbb{P}^2$. Then the union of l and l' is a reducible conic containing all four points. Since l' is an arbitrary line in $\mathbb{P}^2$, this family has dimension two. □

Now we wish to investigate briefly the number of points of intersection of two projective plane curves of various degrees. First, if we intersect a projective line with a conic, we always get two points if the points are counted properly. To see this, we can parametrize any line in the projective plane by

$$\begin{aligned} X &= a_1 s + b_1 t \\ Y &= a_2 s + b_2 t \\ Z &= a_3 s + b_3 t, \end{aligned} \tag{4}$$

where s and t cannot both be zero. Substituting these equations into the equation of a general conic gives a homogeneous quadratic polynomial in s and t. Setting this polynomial equal to zero and solving yields two points $[s, t]$ in the projective line $\mathbb{P}^1$. Substituting back into equations (4) yields the two points where the line meets the conic.

If we similarly investigate the intersection of two conics in the projective plane, we find that two conics always meet in four points if the points are counted properly. If one of the conics is reducible, this result follows from the previous paragraph, so we may assume the conics are nonsingular. Choose coordinates in the projective plane so that one conic has projective equation $XZ = Y^2$. We then parametrize this conic by the equations

$$\begin{aligned} X &= s^2 \\ Y &= st \\ Z &= t^2, \end{aligned} \tag{5}$$

where once again s and t cannot both be zero. Substituting these equations into the equation of a general conic gives a homogeneous quartic polynomial in s and t. Setting this polynomial equal to zero and solving yields four points $[s, t]$ in the projective line $\mathbb{P}^1$. Substituting back into equations (5) yields the four points where the two conics meet.

These two elementary computations are special cases of a more general result known as Bézout's Theorem, which says that projective algebraic curves of degrees m and n having no common component always meet in mn points if the points are counted properly. For our purposes, the two cases outlined above suffice.

A GEOMETRIC SOLUTION TO THE GENERAL QUARTIC. Let's go back to the reduced quartic polynomial given in equation (2):

$$x^4 + px^2 + qx + r = 0,$$

where $p, q, r \in \mathbb{Q}$. Considering these polynomials as having complex coefficients and setting $y = x^2$, we see that the solutions to equation (2) are the x-coordinates of the points of intersection of the conics with affine equations

$$\begin{aligned} y^2 + py + qx + r &= 0 \\ y - x^2 &= 0, \end{aligned}$$

in the affine plane $\mathbb{A}^2$. If we take the projective closure of these curves in $\mathbb{P}^2$, we get the projective curves C_1 and C_2 defined by polynomials

$$F_1(x, y, z) = y^2 + pyz + qxz + rz^2$$

$$F_2(x, y, z) = yz - x^2,$$

respectively. Using Bézout's Theorem, the curves C_1 and C_2 meet in the four points P_1, P_2, P_3, P_4, all of which lie in the finite plane and have affine coordinates $P_i = (x_i, x_i^2)$.

To see that the conditions imposed by P_1, P_2, P_3, P_4 are independent, we need only show that these points are noncollinear in $\mathbb{P}^2$. However, the four distinct points P_1, P_2, P_3, P_4 all lie on the irreducible conic $y = x^2$ in the affine plane, so they are not collinear, again by Bézout's theorem. It follows from the proposition that the set of conics in $\mathbb{P}^2$ containing P_1, P_2, P_3, P_4 forms a (projective) one-dimensional linear subspace Π of $\mathbb{P}^5$, so the conics C_1 and C_2 span Π. That is, any curve C in Π has equation $\lambda F_1 + \mu F_2 = 0$, where either λ or μ is not zero.

We now wish to find those conics C in the linear family Π that are reducible.

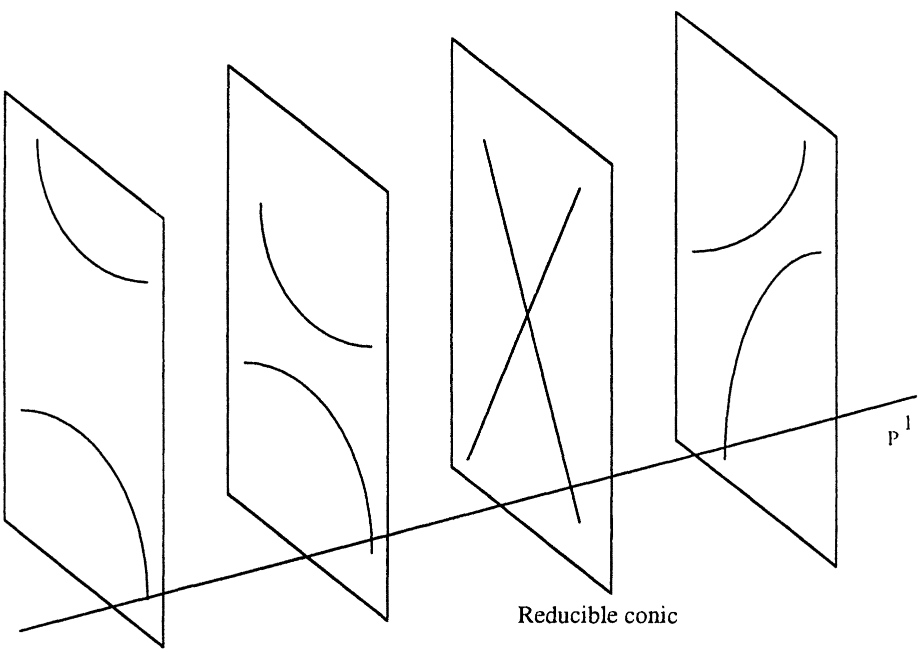

Figure 1. A Pencil of Conics Showing a Reducible Conic

The matrices $\mathbf{A}_i$ of conics C_i are given by

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & q/2 \\ 0 & 1 & p/2 \\ q/2 & p/2 & r \end{bmatrix} \qquad \mathbf{A}_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix},$$

so the matrix of the polynomial $F = \lambda F_1 + \mu F_2$ of any conic C in Π is given by the matrix

$$\begin{bmatrix} -\mu & 0 & \frac{1}{2}q\lambda \\ 0 & \lambda & \frac{1}{2}p\lambda + \frac{1}{2}\mu \\ \frac{1}{2}q\lambda & \frac{1}{2}p\lambda + \frac{1}{2}\mu & r\lambda \end{bmatrix},$$

and C is reducible precisely when this matrix is singular. The determinant of this matrix is

$$\tfrac{1}{4}\left[\mu^3 - q^2\lambda^3 + (p^2 - 4r)\lambda^2\mu + 2p\lambda\mu^2\right]. \tag{6}$$

As the reader can see, this equation is homogeneous in λ and μ of degree three, so the roots $[\lambda, \mu]$ of this equation correspond to three reducible conics in the family Π. Let L_{ij} be the line through P_i and P_j. Then L_{ij} has affine equation $Y = (x_i + x_j)X - x_i x_j$. One of the three reducible conics in the family Π is $L_{12} + L_{34}$, which satisfies the polynomial

$$\begin{aligned} &[Y - (x_1 + x_2)X + x_1x_2][Y - (x_3 + x_4)X + x_3x_4] \\ &\quad = Y^2 + (x_1x_2 + x_3x_4)Y + (x_1 + x_2)(x_3 + x_4)X^2 + qX + r \\ &\quad = F_1 - (x_1 + x_2)(x_3 + x_4)F_2, \end{aligned}$$

noting that, by assumption, the coefficient of the XY term is $-(x_1 + x_2 + x_3 + x_4) = 0$. Hence, one of the roots of polynomial (6) is $[1, -(x_1 + x_2)(x_3 + x_4)] = [1, -\alpha]$. Similarly, the remaining two roots of polynomial (6) are $[1, -\beta]$ and $[1, -\gamma]$, so that the solutions of the resolvent cubic correspond geometrically to finding the three reducible conics in the space of conics spanned by C_1 and C_2.

Since the reducible conics in Π are

$$\begin{aligned} Q_1 &= L_{12} + L_{34} \\ Q_2 &= L_{13} + L_{24} \\ Q_3 &= L_{14} + L_{23}, \end{aligned}$$

it is easy to see that the intersection of any two of these conics produces the desired points P_1, P_2, P_3, P_4.

Thus, if we interpret the roots of the general quartic as the first coordinates of points P_1, P_2, P_3, P_4 in the intersection of two conics in $\mathbb{P}^2$, we see that the resolvent cubic obtained from Galois Theory is, up to a nonzero constant multiple, just the determinant of the 3×3 matrix defining any conic in the family of conics containing the four points P_1, P_2, P_3, P_4. Solving the resolvent cubic corresponds geometrically to finding the reducible conics in this family. It is then a straightforward matter to solve the quartic equation geometrically.

REFERENCES

[A] E. Artin, *Galois Theory*, University of Notre Dame Press, Notre Dame, 1944.

[B] M. Berger, *Geometry*, Springer-Verlag, New York, 1987.

[E] G. Ehrlich, *Fundamental Concepts of Abstract Algebra*, PWS-Kent Publishing Company, Boston, 1991.

[F1] J. B. Fraleigh, *A First Course in Abstract Algebra*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1982.

[F2] W. Fulton, *Algebraic Curves*, Benjamin/Cummings Publishing Company, Reading, Massachusetts, 1981.

[H1] I. N. Herstein, *Topics in Algebra*, John Wiley & Sons, New York, 1975.
[H2] T. W. Hungerford, *Algebra*, Springer-Verlag, New York, 1974.
[R] M. Reid, *Algebraic Geometry*, Oxford Press, Oxford, 1970.
[W] B. L. van der Wärden, *Algebra*, Frederick Ungar Publishing Co., New York, 1970.

Department of Mathematics
West Georgia College
Carrollton, GA 30118
mfaucett@westga.edu

To learn Calculus without understanding what led to its development and how it was used by Newton and others, is like learning to play scales on the piano without being shown any compositions.

—*F. J. Swertz*

The incorporation of history in the teaching of mathematics is essential if the ideas of its purpose, its structure, its wonder, its creativeness are to be aroused in the child.

—*F. J. Swertz*

History is commonly taught in schools to initiate the young into a community—to give them an awareness of tradition, a feeling of belonging, and a sense of participation in an ongoing process or institution. Similar goals can be advocated for the teaching of the history of mathematics.

—*F. J. Swertz*