Polynomial Variables and the Jacobian Problem

Balwant Singh

Balwant Singh received his PhD from Bombay University in 1970. He works in commutative algebra and algebraic geometry, and is currently Professor and Dean, Mathematics Faculty, Tata Institute of Fundamental Research, Mumbai.

Most readers would be familiar with the technique of shifting the origin to simplify the equation of a locus. This process is the same as choosing new coordinate axes, parallel to the original ones. One can combine it with a rotation of the axes. If the angles of rotation of the two axes are different, we end up with a non-rectangular coordinate system.

It is clear that by transformations of the type described above, any two intersecting lines can be made to play the role of coordinate axes. While these transformations are adequate to deal with equations of low degree, it becomes necessary, for handling higher degree equations, to consider transformations which produce curvilinear coordinate axes, i.e. axes which may be curves rather than just straight lines.

As will become apparent soon, not every pair of intersecting curves can serve as a set of (curvilinear) coordinate axes. So we have our first question: How do we decide if a given pair of curves is admissible as a set of coordinate axes?

Suppose the curves are given by the equations $f = 0$ and $g = 0$, where $f = f(X, Y)$ and $g = g(X, Y)$ are polynomials in X, Y. We shall see below that for these two curves to be a set of coordinate axes it is necessary and sufficient that X and Y be expressible as polynomials in f, g. Moreover, for this to happen an obvious necessary condition is that the Jacobian of f and g be a nonzero constant.

Is this condition sufficient? Intriguing as it may sound, the answer to this simple and straightforward question, which has become famous as the *Jacobian Problem,* is not yet known in spite of the efforts of several mathematicians over the last few decades.

Our aim in this article is to give a brief description, at an elementary level, of some of the progress made on this problem.

Note that shifting the origin to the point $(-p, -q)$ amounts, algebraically, to making the *change of variables* $(X, Y) \mapsto (X+p, Y+q)$. Similarly, choosing the lines $aX+bY+p=0$ and $cX+dY+q=0$ as the new coordinate axes is the same as making the change of variables

$$(X, Y) \mapsto (aX+bY+p, cX+dY+q).$$

Such a change is called a *linear* change of variables because $aX+bY+p$ and $cX+dY+q$ are linear polynomials in X, Y. The two lines $aX+bY+p=0$ and $cX+dY+q=0$ will serve as coordinate axes if and only if the lines intersect, which condition is the same as $ad-bc \neq 0$.

Now, let $f = f(X,Y)$ and $g = g(X,Y)$ be polynomials of arbitrary positive degrees. Choosing the curves $f=0$ and $g=0$ as new coordinate axes is the same as making the change of variables $(X, Y) \mapsto (f, g)$.

A comparison with the linear case and a moment's thought will convince the reader that if a point has coordinates (α, β) in the old system (X, Y) then its coordinates in the new system (f, g) are $(f(\alpha,\beta), g(\alpha,\beta))$. So it is natural to say that the curves $f=0$ and $g=0$ are *admissible* as coordinate axes if (and only if) the map $(\alpha,\beta) \mapsto (f(\alpha,\beta), g(\alpha,\beta))$ is a bijection of the plane. In this case the corresponding change of variables may be called an *admissible change of variables.*

In making the above definition we have assumed tacitly that we are working in the plane $\mathbb{R}^2$ over the field $\mathbb{R}$ of real numbers. The definition is good so long as we work over an infinite field.

As an example, consider the change $(X, Y) \mapsto (X, Y+X^2)$. The map $(\alpha,\beta) \mapsto (\alpha, \beta+\alpha^2)$ is clearly a bijection. So this is an admissible change. Under this transformation the Y-axis remains unchanged while the parabola $Y+X^2=0$ becomes the other (curvilinear) coordinate axis.

On the other hand, the map $(\alpha,\beta) \mapsto (\alpha, \alpha+\beta^2)$ is not a bijection, so $(X, Y) \mapsto (X, X+Y^2)$ is not an admissible change.

A map $(\alpha,\beta) \mapsto (f(\alpha,\beta), g(\alpha,\beta))$, where f, g are polynomials, is called a *polynomial map* because the image of a point is given by a (fixed) pair of polynomials evaluated at that point.

It can be proved (assuming that we are working over an infinite field) that if a polynomial map is a bijection of the plane then its inverse is again a polynomial map.

The question posed above can be restated as follows: For given polynomials f and g how can one check if the map $(\alpha, \beta) \mapsto (f(\alpha, \beta), g(\alpha, \beta))$ is a bijection? If the polynomials are linear then we have seen an answer above. To confirm it in terms of the precise definition of admissibility, one checks easily that the map $(\alpha, \beta) \mapsto (a\alpha + b\beta + p, c\alpha + d\beta + q)$ is a bijection if and only if $ad - bc \neq 0$.

Before discussing the general case, note that the change

$$(X, Y) \mapsto (aX + bY + p, cX + dY + q)$$

can be made in two steps, namely $(X, Y) \mapsto (aX + bY, cX + dY)$ and

$$(aX + bY, cX + dY) \mapsto (aX + bY + p, cX + dY + q).$$

The second step is a shift of the origin which is always admissible. The admissibility of the first step is equivalent to the following condition: X and Y can be expressed as linear combinations of $aX + bY$, $cX + dY$.

This criterion extends to polynomials of higher degree as follows: The change $(X, Y) \mapsto (f, g)$ is admissible if and only if X and Y can be expressed as polynomials in f, g.

For a proof of this statement, see *Box 1*.

We introduce at this stage a convenient terminology. A pair (f, g) of polynomials is called a *pair of polynomial variables* if X and Y can be expressed as polynomials in f, g.

With this terminology the change of variables $(X, Y) \mapsto (f, g)$ is admissible if and only if (f, g) is a pair of polynomial variables. So the question raised above takes the following form: Given a pair of polynomials f, g how does one decide whether (f, g) is a pair of polynomial variables?

Box 1.

Suppose the change $(X, Y) \mapsto (f, g)$ is admissible. Then, as noted above, the inverse of the bijective map $(\alpha, \beta) \mapsto (f(\alpha,\beta), g(\alpha,\beta))$ is a polynomial map. So there exist polynomials φ, ψ such that

$$(\alpha,\beta) = (\varphi(f(\alpha,\beta), g(\alpha,\beta)), \psi(f(\alpha,\beta), g(\alpha,\beta)))$$

for all (α,β). Now, $\varphi(f,g)$ is a polynomial in f,g and a polynomial in X,Y. Regarding it as a polynomial in X,Y and evaluating it at (α,β) we get $\varphi(f,g)(\alpha,\beta) = \varphi(f(\alpha,\beta), g(\alpha,\beta)) = \alpha = X(\alpha,\beta)$. Thus the values of the two polynomials $\varphi(f,g)$ and X coincide at all points (α,β). Therefore (assuming that we are working over an infinite field) we get $X = \varphi(f,g)$, showing that X can be expressed as a polynomial in f,g. Similarly, $Y = \psi(f,g)$ is a polynomial in f,g.

Conversely, if it is given that X and Y are polynomials in f,g, say $X = \varphi(f,g)$ and $Y = \psi(f,g)$ then substituting (α,β) for (X,Y) we get $(\alpha,\beta) = (\varphi(f(\alpha,\beta), g(\alpha,\beta)), \psi(f(\alpha,\beta), g(\alpha,\beta)))$ for all (α,β). This shows that the map $(\alpha,\beta) \mapsto (f(\alpha,\beta), g(\alpha,\beta))$ is injective. To prove its surjectivity, consider the polynomial $h(X,Y) = X - f(\varphi,\psi)$. We have $h(f,g) = f - f(\varphi(f,g), \psi(f,g)) = f - f(X,Y) = 0$. Now, X and Y are independent variables. Therefore, since X and Y are polynomials in f,g, there cannot be a nontrivial polynomial relation between f and g. Consequently, since $h(f,g) = 0$, we must have $h = 0$, i.e. $X = f(\varphi,\psi)$. Similarly, $Y = g(\varphi,\psi)$. Now, if (γ,δ) is any point of the plane, we get $(\gamma,\delta) = (f(\varphi(\gamma,\delta), \psi(\gamma,\delta)), g(\varphi(\gamma,\delta), \psi(\gamma,\delta)))$, showing that (γ,δ) is in the image of the map $(\alpha,\beta) \mapsto (f(\alpha,\beta), g(\alpha,\beta))$. This proves the surjectivity of the map.

Suppose (f,g) is a pair of polynomial variables. Then there exist polynomials $\varphi = \varphi(X,Y)$ and $\psi = \psi(X,Y)$ such that $X = \varphi(f,g)$ and $Y = \psi(f,g)$. Taking partial derivatives with respect to X and Y and using the chain rule, we get

$$\begin{pmatrix} \frac{\partial\varphi}{\partial X}(f,g) & \frac{\partial\varphi}{\partial Y}(f,g) \\ \frac{\partial\psi}{\partial X}(f,g) & \frac{\partial\psi}{\partial Y}(f,g) \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial X} & \frac{\partial f}{\partial Y} \\ \frac{\partial g}{\partial X} & \frac{\partial g}{\partial Y} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

It follows that the product of the determinants of the two matrices appearing on the left side is 1. Now, these determinants are polynomials. Since the degree of the product of two polynomials is the sum of the degrees of the two polynomials, and 1 is a polynomial of degree zero, we conclude that each of these determinants is a polynomial of degree zero, hence a nonzero constant.

Thus we have shown that if (f,g) is a pair of polynomial variables then their *Jacobian* $\mathrm{J}(f,g)$ with respect to X, Y, which is defined by

$$\mathrm{J}(f,g) = \det \begin{pmatrix} \frac{\partial f}{\partial X} & \frac{\partial f}{\partial Y} \\ \frac{\partial g}{\partial X} & \frac{\partial g}{\partial Y} \end{pmatrix}$$

is a nonzero constant.

The famous Jacobian problem asks whether the converse is true:

Jacobian Problem. If f, g are polynomials in X, Y such that $\mathrm{J}(f,g)$ is a nonzero constant then is it true that (f,g) is a pair of polynomial variables?

The answer is trivially negative over fields of positive characteristic, as seen for example by considering the pair $(X + X^p, Y)$ over a field of characteristic $p > 0$.

Therefore we restrict ourselves to fields of characteristic zero, for example the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers.

We cannot draw a picture of the 'plane' $\mathbb{C}^2$ (which is 4-dimensional over $\mathbb{R}$). However, since our discussion is algebraic, and since algebra over $\mathbb{C}$ is simpler than algebra over $\mathbb{R}$, we shall work from now on over $\mathbb{C}$. This does not result in a loss of generality as far as the Jacobian problem is concerned.

For convenience, we state an affirmative answer to the Jacobian problem as the

Jacobian Conjecture. If f, g are polynomials in X, Y such that $\mathrm{J}(f,g)$ is a nonzero constant then (f,g) is a pair of polynomial variables.

The Jacobian problem seems to have appeared in print for the first time in 1939, formulated by O Keller. Several mathematicians have made attempts to solve the problem, particularly during the last three decades, but the only outcome of these efforts has been to obtain other equivalent formulations of the problem and some partial solutions in the form of an affirmative answer under additional hypotheses.

We aim to describe some of this progress. Our exposition is limited necessarily to those aspects which can be presented without getting into too many technicalities. It will be seen that in some formulations the gap between what has been proved and what is needed is tantalizingly small.

Birational Case. Suppose $J(f,g)$ is a nonzero constant. If X and Y can be expressed as rational functions (i.e. quotients of polynomials) in f,g then (f,g) is a pair of polynomial variables. This was proved already by Keller in 1939. The additional hypothesis means that the field of rational functions in X,Y is the same as the field of rational functions in f,g (hence the term 'birational').

Galois Case. Suppose $J(f,g)$ is a nonzero constant. If the field of rational functions in X,Y is a Galois extension of the field of rational functions in f,g then (f,g) is a pair of polynomial variables. This was proved in about 1973 independently by L A Campbell using topological methods and by S S Abhyankar using algebraic methods.

Injectivity is Enough. Suppose $J(f,g)$ is a nonzero constant. If the map $(\alpha,\beta) \mapsto (f(\alpha,\beta), g(\alpha,\beta))$ is injective then (f,g) is a pair of polynomial variables. This was proved in 1982 by H Bass, E H Connell and D Wright.

Points at Infinity. Let f be a nonzero polynomial in X,Y, and let $d = \deg(f)$. Then f can be written in the form $f = f_d + f_{d-1} + \cdots + f_1 + f_0$ with f_i homogeneous of degree i. Let Z be another variable and let $F = f_d + Zf_{d-1} + \cdots + Z^{d-1}f_1 + Z^d f_0$, so that F is a homogeneous polynomial in X,Y,Z of degree d, and defines the curve $F = 0$ in the projective plane. The points of intersection of this curve with the line at infinity (i.e. the line $Z = 0$) are called the *points at infinity* of the curve $f = 0$. These points are given in the projective plane by the equations $Z = 0$, $f_d = 0$. From this it is clear that the number of points at infinity of the curve $f = 0$ is the same as the number of mutually coprime factors of the homogeneous polynomial f_d. Recall here that, since we are working over $\mathbb{C}$, f_d factors into (possibly repeated) homogeneous linear factors.

Jacobian Condition Implies at most Two Points at Infinity. If $\mathrm{J}(f,g)$ is a nonzero constant then the curve $f = 0$ has at most two points at infinity.

Jacobian Conjecture in Terms of Number of Points at Infinity. The Jacobian conjecture is equivalent to the following statement: If $\mathrm{J}(f,g)$ is a nonzero constant then the curve $f = 0$ has only one point at infinity.

This is one formulation where, as mentioned above, the gap between what has been proved and what is needed appears to be very small.

Jacobian Conjecture in Terms of Degrees. The Jacobian conjecture is equivalent to the following statement: If $\mathrm{J}(f,g)$ is a nonzero constant, then of the two integers $\deg(f)$, and $\deg(g)$, one divides the other.

The above three results relating the Jacobian problem with the number of points at infinity and the degrees of the polynomials were proved in 1977 by Abhyankar.

Small Degrees are Okay. If $f = aX + bY + p$ and $g = cX + dY + q$ are linear polynomials then $J(f,g)$ is the constant $ad - bc$, and we have seen already that if $ad - bc \neq 0$ then (f,g) is a pair of polynomial variables. Thus the Jacobian conjecture holds for polynomials of degree one.

It was proved in 1983 by T T Moh that if the degrees of f and g are at most 100 and $\mathrm{J}(f,g)$ is a nonzero constant then (f,g) is a pair of polynomial variables. Thus, if the problem has a negative answer then at least one member of the pair of polynomials giving a counterexample would have to be of degree greater than 100.

It would be very instructive for the reader to play with a few polynomials and try to verify by direct computation that the Jacobian conjecture holds for polynomials of very small degree.

Jacobian Problem in Higher Dimensions. The problem generalizes clearly to the case of n polynomials in n variables: Let $f_1, f_2, \ldots, f_n$ be polynomials in n variables $X_1, X_2, \ldots, X_n$. Suppose $\mathrm{J}(f_1, f_2, \ldots, f_n)$, the Jacobian of $f_1, f_2, \ldots, f_n$ with respect to $X_1, X_2, \ldots, X_n$, is a nonzero constant. Is it then true that $X_1, X_2, \ldots, X_n$ can be expressed as polynomials in $f_1, f_2, \ldots, f_n$?

An affirmative answer to this problem may be called the Jacobian conjecture for general n.

As in the case $n = 2$, the Jacobian conjecture holds for general n in the Galois case (in particular, in the birational case). Similarly, the injectivity of the corresponding polynomial map of the affine n-space is enough.

A few more partial results and reductions in the general case are described below.

Reduction to Degree Three. The following two conditions are equivalent:

(1) The Jacobian conjecture holds for all n.

(2) The Jacobian conjecture holds whenever the degree of each f_i is at most three (and n is arbitrary).

This was proved by Bass–Connell–Wright in 1982.

It is important to emphasize here that the equivalence of the two conditions has been proved only by treating all n simultaneously. In fact, the reduction to degree three is achieved at the cost of a (large) increase in the number of variables.

The above reduction may be contrasted with the following positive result:

Degree Two is Okay. The Jacobian conjecture holds if the degree of each f_i is at most two.

This was proved by Wang in 1980.

So here again the gap between what is needed and what has been proved appears to be very small.

Polynomial Variable – Epimorphism Theorem We return to the case of two variables and discuss a related question. Call a polynomial f a *polynomial variable* if there exists a polynomial g such that (f, g) is a pair of polynomial variables. Equivalently, f is a polynomial variable if the curve $f = 0$ is a (curvilinear) coordinate axis under an (admissible) change of variables. The question is this: Given a polynomial f how does one decide whether f is a polynomial variable?

Let f and g be polynomials in X, Y and let C and D denote the curves $f = 0$ and $g = 0$, respectively. We say that the curve C is *isomorphic* to the curve D if there exist polynomial maps $C \to D$ and $D \to C$ which are inverses of each other.

For example, the parabola C defined by the equation $Y + X^2 = 0$ and the line D defined by the equation $Y = 0$ are isomorphic via the polynomial maps $\sigma : C \to D$ and $\tau : D \to C$ given by $\sigma(a, b) = (a, 0)$ and $\tau(a, 0) = (a, -a^2)$.

It is easy to see that the line given by a linear equation $aX + bY + c = 0$ is isomorphic to the line $Y = 0$. Let us call such a line a *straight line* (with respect to the coordinate system (X, Y)). Any two straight lines are isomorphic.

Call a curve $f = 0$ a *line* if it is isomorphic to a straight line, for example to the straight line $Y = 0$.

By this definition the parabola $Y + X^2 = 0$ is a line but not a straight line.

Now, suppose f is a polynomial variable. Let g be any polynomial such that (f, g) is a pair of polynomial variables. On making the change of variables $(X, Y) \mapsto (f, g)$ the curve $f = 0$ becomes one of the new (curvilinear) coordinate axes. This curve, though not a straight line in general, is nevertheless a line.

Let us verify this.

Since (f, g) is a pair of polynomial variables, the map $(\alpha, \beta) \mapsto (f(\alpha, \beta), g(\alpha, \beta))$ is a bijection of the plane, whence so also is the map $(\alpha, \beta) \mapsto (g(\alpha, \beta), f(\alpha, \beta))$. It is clear that under this bijection the curve $f = 0$ maps onto the line $Y = 0$. Further, the inverse map is again a polynomial map, as noted above. This proves that the curve $f = 0$ is isomorphic to the line $Y = 0$, hence is a line. (Note that, in fact, we have proved more, namely there exists not only an isomorphism between $f = 0$ and $Y = 0$, but an isomorphism of the entire plane onto itself which maps one curve to the other.)

Thus for a curve to be a (curvilinear) coordinate axis under a change of variables it is necessary that the curve be a line.

We can ask if the condition is sufficient: Is every line a (curvilinear) coordinate axis under some change of variables?

Each of the following two formulations is equivalent to the above question:

(1) If $f = 0$ is a line then is it true that f is a polynomial variable?

(2) Given an isomorphism of a curve with the line $Y = 0$, can it always be extended to an isomorphism of the entire plane onto itself?

For a straight line the answer is clearly affirmative. Such a line together with any non-parallel straight line would serve as a pair of coordinate axes.

For a more general line, the answer is still affirmative (remember that we are working over a field of characteristic zero) but the proof is surprisingly difficult. The first proof was provided in the well-known Abhyankar–Moh Epimorphism theorem which appeared in 1973.

The reader will notice a similarity between the above question (which has been solved) and the Jacobian problem (which is unsolved). It is very tempting to believe that the techniques which go into proving the Epimorphism theorem would produce something positive when applied to the Jacobian problem. An attempt in this direction has indeed been made. However, the result so far has been to generate other equivalent formulations of the Jacobian problem, leading to its proof in a few more special cases, but no real breakthrough yet for a complete solution of the problem.

Suggested Reading

[1] S S Abhyankar, Expansion techniques in algebraic geometry, *TIFR Lecture Notes*, 1977.

[2] H Bass, E H Connell and D Wright, The Jacobian Conjecture: Reduction of degree and formal expansion of the inverse, *Bull. AMS*, 7, 287–330, 1982.

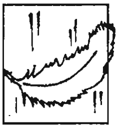
[3] L A Campbell, A condition for a polynomial map to be invertible, *Math. Ann.*, 205, 243–248, 1973.

[4] O H Keller, Ganze Cremona-Transformationen, *Monats. Math. Physik*, 47, 299–306, 1939

[5] T T Moh, On the Jacobian conjecture and the configuration of roots, *J. Reine Angew. Math.*, 340, 140–212, 1983.

[6] S Wang, A Jacobian criterion for separability, *J. Algebra*, 65, 453–494, 1980.

Address for Correspondence
Balwant Singh
School of Mathematics
Tata Institute of Fundamental Research
Homi Bhabha Road
Mumbai 400 005, India.
email: balwant@math.tifr.res.in



According to Bohr the notion of complementarity serves "to symbolize the fundamental limitation, met with in atomic physics, of our ingrained idea of phenomena as existing independently of the means by which they are observed"

Wolfgang Pauli