A "FORCEFUL" CONSTRUCTION OF 1-1 COMPLEX POLYNOMIAL MAPPINGS

ALAN GLUCHOFF AND FREDERICK HARTMANN

1. Introduction

The polynomial $p_1(x) = x + (1/2)x^2$ is one-to-one on the interval $(-1, 1)$. We can show this algebraically: the equation $x_1 + (1/2)x_1^2 = x_2 + (1/2)x_2^2$ gives either $x_1 = x_2$ or $x_1 + x_2 = -2$. This yields a contradiction if $x_1 \neq x_2$ and both x_1 and x_2 are in $(-1, 1)$. An easier way is simply to note that $p_1'(x) = 1 + x > 0$ on the interval, so p_1 is increasing there. This also tells us about the behavior of p_1 on the interval.

What happens if we replace x by z in the complex plane $\mathbb{C}$ and $|x| < 1$ by $|z| < 1$? Does the polynomial $p_1(z) = z + (1/2)z^2$ remain one-to-one on the *unit disc* $D = \{z : |z| < 1\}$? Since the algebraic argument above carries through with complex numbers, p_1 is indeed one-to-one on D. This is a proof, but it is not very satisfying. Can we find a geometric reason for this behavior, something akin to the increasing of p_1 on $(-1, 1)$?

The simplest approach is to see how D is transformed by p_1; what is the image of D in $\mathbb{C}$ under p_1? In Figure 1a we take a polar co-ordinate grid of D and produce $p_1(D)$.

It appears that the disc has been stretched horizontally, folded and pinched to a cusp at $w = -1/2$ in the range. Writing $p_1(z) = -1/2 + (1/2)(z+1)^2$, we can verify that p_1 is one-to-one on D by using elementary mapping properties. The function first slides D one unit to the right. Then the squaring function folds the arc of the boundary touching the origin over the negative real axis. The remaining scaling and shifting give us the one-to-one image. We can also see that circles centered at the origin appear to go to heart-shaped curves in the range. More work shows that if the circles are traversed counterclockwise, then the range curves have increasing

Date: June 1, 2006.

1991 *Mathematics Subject Classification.* Primary: 30C45; Secondary: 30C50, 30C55.

Key words and phrases. Univalent functions, poynomials, starlike, convex, close-to-convex.

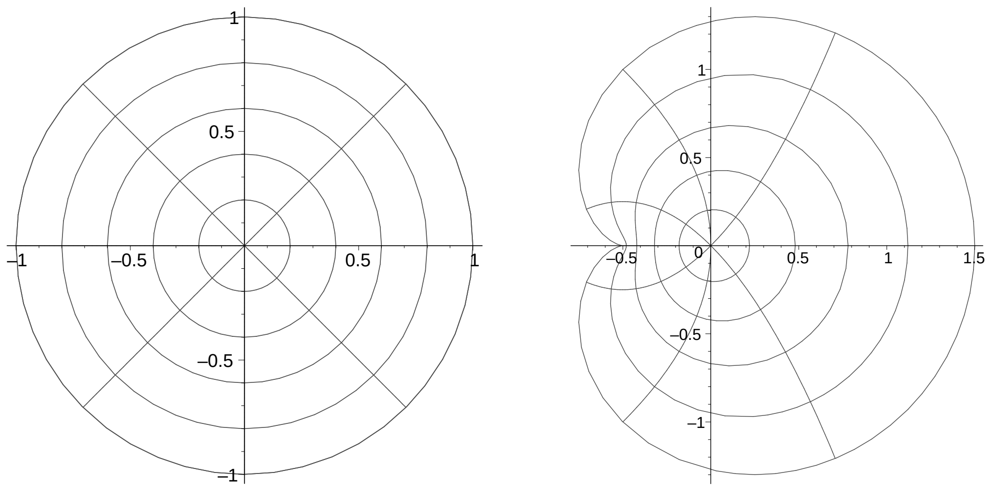


Fig. 1a. $p_1(z) = z + (1/2)z^2$

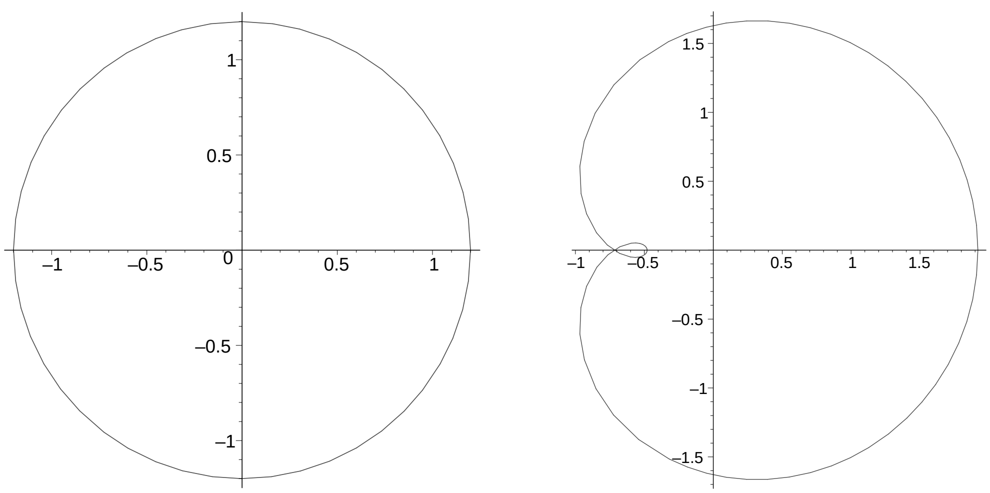


Fig. 1b. $p_1(z) = z + (1/2)z^2$

polar argument $\arg(p_1(\cos(t), \sin(t)))$ as t increases. There is no "looping" of these image curves over themselves. If there were, p_1 would clearly not be one-to-one. We do come close to this behavior at the range point $w = -1/2$, the image of the critical point $z = -1$. Here the cusp would yield to a small simple loop if a circle of radius $1 + \epsilon$ were input into p_1. See Figure 1b).

There is some interesting geometry here. The set $C = \{z : |z| = 1\}$ has been transformed into an epicycloid. This can be seen by substituting $z = \cos(t) + i\sin(t)$ into p_1 and using the DeMoivre relations to get

$$x = \cos(t) - (1/2)\cos(2t),\ y = \sin(t) - (1/2)\sin(2t),\ 0 \leq t \leq 2\pi.$$

If our goal was to produce an interesting complex polynomial which is one-to-one on D, we have succeeded, although only with a quadratic. But we were lucky. First, a polynomial which is monotonic on $(-1, 1)$ need not extend to one which is one-to-one on D. If $q_1(z) = (z + 9/8)^3$, then we have critical points at $z = -9/8$ only, and q_1 increases in $(-1, 1)$. But q_1 slides D into a disc centered at $z = 9/8$, then the complex cubing map twists that circle over itself on the negative real axis as shown in Figure 2a. Thus q_1 is not one-to-one. Second, even requiring $p'(z) \neq 0$ on all of D does not guarantee what we seek. Our q_1 also shows this.

On the other hand, a critical point anywhere in D ruins our chances. Our $p_1(z) = -1/2 + 1/2(z + 1)^2$ transforms any small disc centered at $z = -1$ into a self-overlapping image. If p_1 were changed slightly so that its critical point were in the interior, the new polynomial would fail to be one-to-one near that point. See Figure 2 b, where we have used $q_2(z) = z + (5/8)z^2$ with $q_2'(-4/5) = 0$. This kind of behavior occurs for any polynomial at a critical point, perhaps with a higher degree of looping. This can be seen by writing the polynomial in powers of ($z-$ critical point), as we have done for p_1.

Why is it so hard to produce the simplest (i.e., 1-1) mapping behavior on D for complex polynomials? What takes the place of "increasing or decreasing" in the complex plane? What is the "typical" one-to-one image of D under a polynomial? In this paper we explore these problems and give an approach that yields interesting examples of these "simplest" of functions. Our procedure uses some several variable calculus, brings in a force field first introduced by Gauss, and recalls polynomial mapping ideas which were perhaps more familiar in the first half of the 20th century but seem neglected now. On the way we will meet a complex version of Rolle's theorem. We will look at geometric properties of the images of D which were first considered by the early 20th century American topologist J. W. Alexander, perhaps better known for his knot polynomial and horned sphere (see [3]). We also study inverse images of the polar co-ordinate grid under polynomials. We hope that this approach serves as a good introduction to the geometry of complex polynomials.

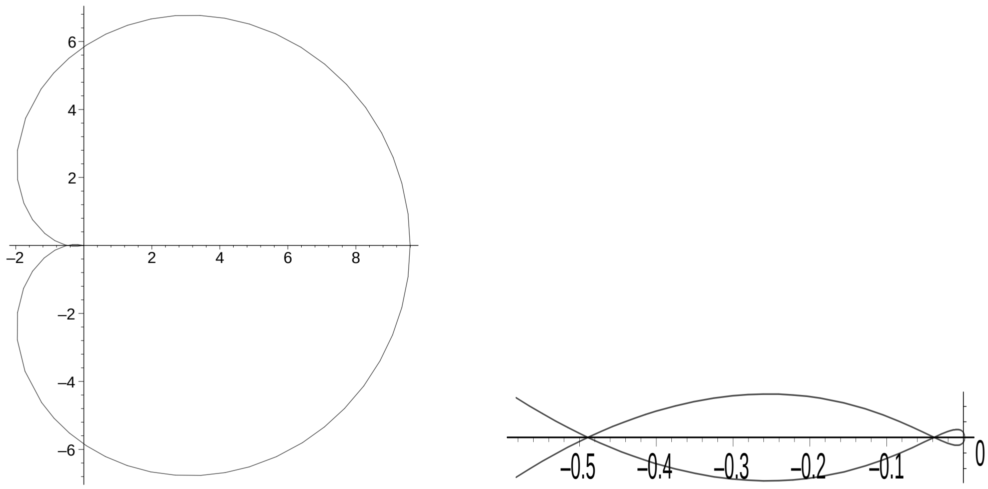


Fig. 2a. $q_1(z) = (z + 9/8)^3$

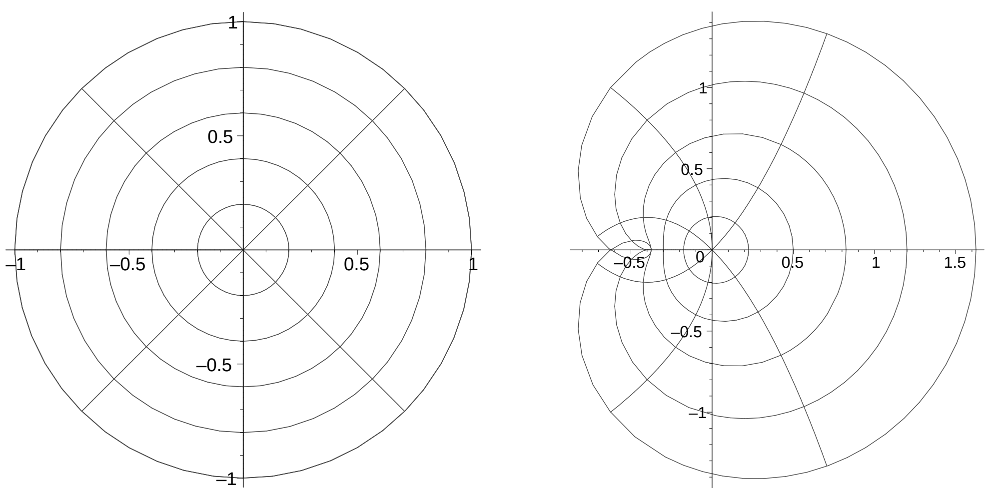


Fig. 2b. $q_1(z) = z + (5/8)z^2$

There are several alternative words for one-to-one complex functions which have become standard: *simple*, the German *schlicht* (= simple), and *univalent* (= one value) are all common. We shall use "univalent" as our synonym for one-to-one. All polynomials will be of the form $p(z) = z + a_2z^2 + \cdots + a_nz^n = z(1 - z/z_1)(1 - z/z_2)\ldots(1 - z/z_{n-1})$. Thus there will always be a single zero at the origin, and

$n-1$ "nontrivial" zeroes. *We assume for simplicity that the nontrivial zeroes are distinct.* This form causes no loss of generality, as we are off from an arbitrary polynomial only by a shift and a scaling. The identity function $I(z) = z$ is obviously univalent. It follows that if the remaining coefficients of $p(z)$ are small relative to 1 (or, equivalently, the nontrivial zeroes are placed far away from the origin) the resulting mapping is likely to remain one-to-one. But this is cheating! Such examples will only perturb D slightly. We wish to do better. Using D as a domain is somewhat arbitrary, but the work we are doing can be adjusted to any other domain. It will be necessary to have a disc domain for our examples in Sections 6 through 8, however.

2. The Gaussian Force Field

In this section we introduce a force field which is central to our work. Let's return to our first polynomial, $p_1(z) = z+(1/2)z^2$. We ask: what is the inverse image under p_1 of a polar co-ordinate grid in the range plane? We are looking for two sets of curves. The first is the set $\{z : |p_1(z)| = r\}$ for $r > 0$, the inverse image of a circle centered at 0 having radius r. These are known as the *level sets* of p_1. The second is the set $\{z : \arg[p_1(z)] = \theta\}$, the inverse image of a ray of polar argument θ through the origin. Figure 3 shows these sets for p_1 with several curves of the second set labelled with their repective θ values. The darkened lines are the preimage of the ray of argument $7\pi/8$.

We note that the two sets form orthogonal families. This follows from the fact that complex mappings are *conformal*, or angle and sense preserving, at all noncritical points. There is again some nice geometry here: the level sets are lemniscates of Bernoulli, first discovered by Jacques Bernoulli in 1694. The orthogonal family to the lemniscates may be shown to be hyperbolae.

We need to insert a technical point here. Suppose we write $p_1(x + iy) = (x + (1/2)x^2 - (1/2)y^2) + i(y + yx) = u(x, y) + iv(x, y)$. Our orthogonal families are produced by the computer algebra system Maple in the following way. We choose a point w in $\mathbb{C}$, let $upart = \Re(w)$, the real part of w, $vpart = \Im(w)$, the imaginary part of w. We would like to plot the curve $\dfrac{v(x,y)}{u(x,y)} = \dfrac{vpart}{upart}$. We use $v\cdot upart = u\cdot vpart$ to avoid vanishing denominators. Thus the curves seen are actually the inverse images of lines though the origin containing two rays in opposite directions. In addition the form used may introduce some extraneous curves. Deciding which ray yields which

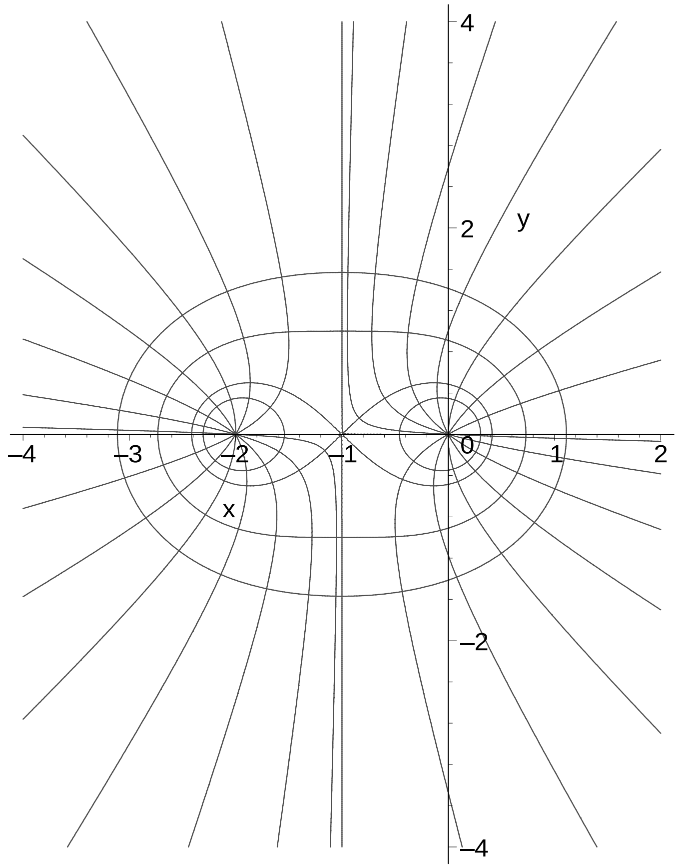


Fig. 3. $p_1(z) = z + (1/2)z^2$

portion of the orthogonal curves and recognizing the extraneous curves is usually easy.

The hyperbola family is suggestive. It reminds one of two fields of force, emanating from $z = 0$ and $z = -2$ respectively, combining to form a single field. A corresponding physical example would be the field between two repelling poles of point magnets. We note that these points are the zeroes of p_1. Can we find this force field explicitly?

Let's switch to several variable calculus mode, and make the correspondence $z = x + iy = (x, y) = x\overrightarrow{I} + y\overrightarrow{J}$. The leminscates are the level curves of $f(x, y) = |p_1(x+iy)|$ in the xy-plane. This brings to mind the gradient field of f, $\nabla(f)(x, y) = \partial/\partial x(f(x, y))\overrightarrow{I} + \partial/\partial y(f(x, y))\overrightarrow{J}$, since the gradient moves orthogonally to the level curves of $f(x, y)$. Our purpose will be better served by considering $\nabla(g)$, where $g(x, y) = (1/2)\ln|p_1(x + iy)|^2$. This gradient field will also follow the hyperbolae streamlines, since it too moves in the direction orthogonal to the level curves of f. A calculation shows that

$$\nabla(g) = [\frac{x}{x^2+y^2} + \frac{x+2}{(x+2)^2+y^2}]\overrightarrow{I} + [\frac{y}{x^2+y^2} + \frac{y}{(x+2)^2+y^2}]\overrightarrow{J}$$

.

But when this is written in complex variable notation, we get

$$\nabla(g)(x,y) = \frac{1}{\overline{z}} + \frac{1}{\overline{z}+2} = \frac{1}{\overline{z}} + \frac{1}{\overline{z}-\overline{-2}}$$

.

We have found that the field $F(x,y) = 1/\overline{z} + 1/(\overline{z}+2)$ moves along the hyperbola family. But there is more: F is a force field. If z_k is a zero of p_1, then $1/(\overline{z-z_k}) = ((z-z_k)/|z-z_k|)\cdot(1/|z-z_k|)$. This term can be viewed at a vector in the direction from z_k to z with magnitude inversely proportional to the distance from z_k to z. When placed at z it can be interpreted as a force vector from the "source" z_k. *Thus $F(x,y)$ is the sum of forces emanating from $z=0$ and $z=2$ acting at $z=(x,y)$.*

This force field was introduced by Gauss in 1816. His starting point was different: for $p(z) = z(z-z_1)\ldots(z-z_{n-1})$ he obtained by logarithmic differentiation that

$$\frac{p'(z)}{p(z)} = \frac{1}{z-0} + \frac{1}{z-z_1} + \ldots \frac{1}{z-z_{n-1}}.$$

Thus

$$\overline{p'(z)} = \overline{p(z)}[\frac{1}{(\overline{z}-\overline{0})} + \frac{1}{(\overline{z}-\overline{z_1})} + \cdots \frac{1}{(\overline{z}-\overline{z_{n-1}})}] = \overline{p(z)}F(x,y).$$

Gauss concluded that the critical points of p are the points of equilbrium of F, along with zeroes of higher multiplicity. We can see in our example that the critical point $z=-1$ lies at an equilibrium point on the real axis from the forces at 0 and -2.

This observation was extended by Lucas in 1874 to what is now called the Gauss-Lucas theorem: all the critical points of p lie in the smallest closed convex set in which the zeroes lie. The proof of this follows immediately from the fact if z is not in this set then the sum of the forces acting on z must point in a direction away from the set. Thus z cannot be an equilibrium point. The theorem is a kind of Rolle's theorem in $\mathbb{C}$. We will sharpen this result later in Section 5.

M. Marden's book [6], first published in 1949, has more details on this theorem. In another 1949 work [8], the American mathematician J. L. Walsh published the

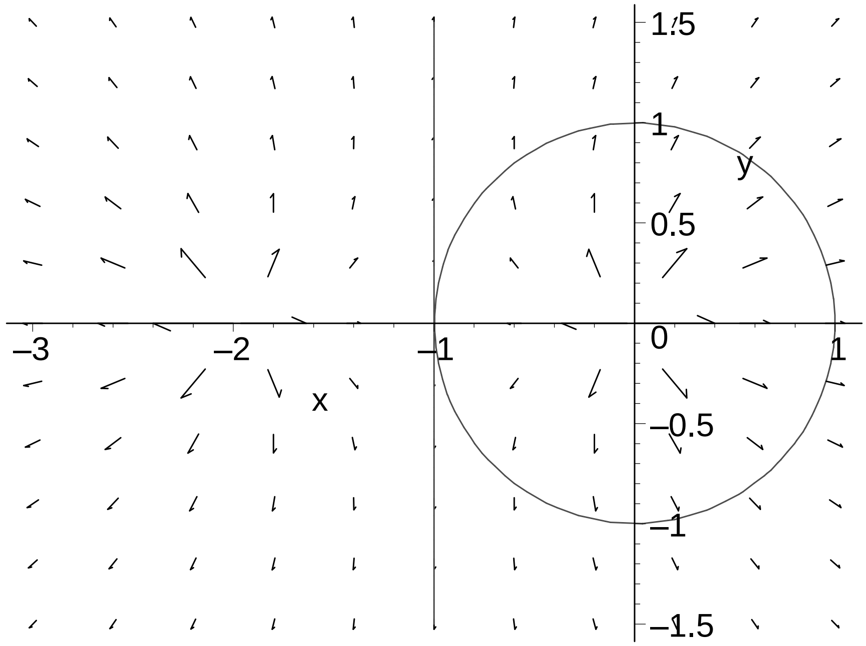


Fig. 4. Force field for $p_1(z) = z + (1/2)z^2$.

culmination of several decades of work on zeroes and critical points of polynomials using the force field as his main tool. It was occasionally referred to as the "mechanical analogy" [4]. We point out that, except at the zeroes of p, the $F(x, y)$ is sourceless (divergence-free) and irrotational (curl-free), in the language of vector calculus [2].

3. From The Force Field To Basins of Univalence

We again return to our initial example. The polynomial $p_1(z) = z + (1/2)z^2$ has zeroes at $z = 0$ and $z = -2$, and the Gaussian force field associated with these sources is illustrated in Figure 4.

We see that from each zero emanate lines of force dividing the complex plane into two regions: $B_0 = \{z : \Re(z) < -1\}$ and $B_1 = \{z : \Re(z) > -1\}$. The common boundary is the line $\{z : z = -1 + it, t \in \mathbb{R}\}$. We shall call these regions *basins.* The streamlines from this vectorfield follow the family of hyperbolae found previously. Furthermore, it is evident that each basin is mapped univalently by p_1. Each streamline must map into a ray which is strictly increasing as the streamline is traversed from the source out, since the streamline always travels in the direction of increasing $|p_1(z)|$. No two streamlines can be mapped to the same ray: following two such streams back to the origin produces two distinct segments emanating from zero. But $p_1(z)$ acts like the identity $I(z) = z$ near zero, so it can't send these segments to the same set. (An exception to this scenario occurs when a ray in the range contains the image of a critical point, a so-called *branch point*. We will deal with this case in the next section.) Since D is contained in B_0, it too must be mapped univalently.

This now suggests a plan to produce one-to-one polynomial mappings of D. We begin with a single zero at the origin, insuring that the polynomial begins as $z + \cdots$ We then place other zeroes outside D. We configure them so that the resulting Gaussian force field has D contained in the basin emanating from the origin. Let us call this basin B_0. The remaining zeroes cannot be too close to D, since, among other things, an equilibrium point might occur in D, spoiling our chances for univalence. But how far away and in what configurations can they lie to guarantee that D is covered by B_0? This question is not easy to answer, but experimenting with the process is a fascinating and frustrating activity. One needs to get a feeling for how the forces combine.

We illustrate the method with a second simple example. Place a zero at the origin and nontrivial zeroes balanced collinearly with 0 at $z_1 = i\sqrt{3}$, $z_2 = -i\sqrt{3}$. This gives $p_2(z) = z + (1/3)z^3$. We could show univalence of p_3 algebraically as we did for p_1, but again this does not illuminate the geometry. The force at 0 is opposed by the two forces at equal distances from the origin. The streamlines and the resulting vectorfield are shown in Figure 5, along with the basins. We can see that vectors from the origin are countered by those from the other sources until equilibrium lines are formed along two symmetric boundary curves. It appears that D is contained in B_0, so we tentatively conclude that p_2 maps D univalently. The image can be shown to be an epicycloid with two symmetric cusps.

We can generalize this by placing a zero at the origin and the remaining $n-1$ zeroes equiangularly on $\{z : |z| = \sqrt[n-1]{n}$. It can be shown that the resulting polynomial $z + z^n/n$ maps D univalently. The image is an epicycloid with $n-1$ equiangularly placed cusps. Since $\sqrt[n-1]{n} \to 1$ as $n \to \infty$, our point sources grow in number but paradoxically get closer to C! The cancellation of the forces due to their orientation accounts for this phenomenon.

We should note that general properties of the streamline sets can be found in [8] on p. 20. A similar analysis can be found in the first part of a long 1938 treatise by

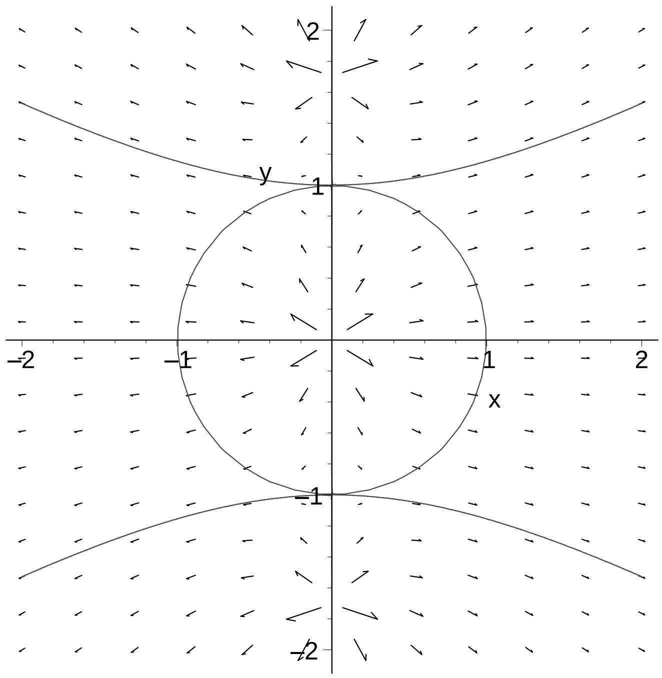


Fig. 5. $p_2(z) = z + (1/3)z^3$

the French mathematician L. Hibbert. This latter work includes the only diagrams of decompositions of the plane into basins known to the authors. Hibbert calls our basins "cellules d'univalence". We will not need these deeper analyses, due to the nature of our examples.

4. Using the Boundaries of the Basins

The success of this method hinges on being able to tell when D is in the basin emanating from the origin. This can be a difficult if we rely on images of the entire vectorfield or set of streamlines from each source. It might be hard to decide containment based on any single image, and more images woud be required. It is easier to find the boundary curves which separate the basins, and do analyses based on them. This is what we shall do, but the matter is a bit tricky for both mathematical and computational reasons. Let us see why.

The problem stems from the fact that the boundaries of our basins are related to the pre-images of rays through the origin which contain branch points. In Figure 4 we see that, for example, the inverse image under p_1 of a line through the origin of slope $\tan(7\pi/8)$ forms a hyperbola which is close to the real axis and the line

$\{z : \Re(z) = -1\}$. We can easily show that the preimage under p_1 of the negative real axis (which contains the branch point $w = -1/2$) consists of the line $\{z : \Re(z) = -1\}$ and the interval $[-2, 0]$. Note that this is the set toward which the highlighted portion of the hyperbola is leaning, and that the line $\{z : \Re(z) = -1\}$ is the boundary between B_0 and B_1. Thus this preimage gives us our boundary, and another extraneous piece, $[-2, 0]$. In addition, if we use the computational approach described earlier, the equation $v \cdot upart = u \cdot vpart$ becomes $y + xy = 0$. This gives the curves $x = -1$ and $y = 0$, in other words, our boundary together with the entire real axis!

In practice these extraneous pieces are easily identified. They often can be seen as dividing the flows into further compartments within a given basin. The real axis here separates the streamlines in the upper and lower half planes of $\mathbb{C}$. We will adopt the attitude that a little experimentation will be enough to decide which curves form the boundaries we need. A deeper study of these curves would allow us to understand the mapping behavior of polynomials on the entire complex plane, but that is not our goal here.

We can now return to $p_2(z) = z + z^3/3$. We find that there are critical points at $z = \pm i$, with branch points at $w = \pm(2/3)i$. With $u = x + (1/3)x^3 - xy^2$ and $v = y + x^2y - 1/3y^3$, our attempt to find the boundaries of the basins by our procedure yields $(2/3)x + (2/9)x^3 - (2/3)xy^2 = 0$. We have asked for the preimage of the imaginary axis to get this equation. The equation yields the hyperbola $y^2 - x^2/3 = 1$ and the imaginary axis $x = 0$. When we consider the three sources at 0, $\pm\sqrt{3}i$, it is clear that the hyperbola is the boundary of interest, and the imaginary axis is extraneous. Thus the three basins are determined by the three sections of the plane into which the hyperbola divides $\mathbb{C}$. Since D is contained in B_0, p_2 maps it univalently. The basins are illustrated in Figure 5.

5. More Examples

In this section we give several more examples of our approach. Bear in mind that these example are a result of some experimentation. They would not immediately suggest themselves. Rather they were found in some cases by trial and error and in other cases by looking at polynomials available in the literature.

Our first is a modification of q_2 in the introduction:

$$p_3(z) = \frac{(z + 2/\sqrt{3})^3 - (2/\sqrt{3})^3}{3(2/\sqrt{3})^2} = z + (\sqrt{3}/2)z^2 + z^3/4.$$

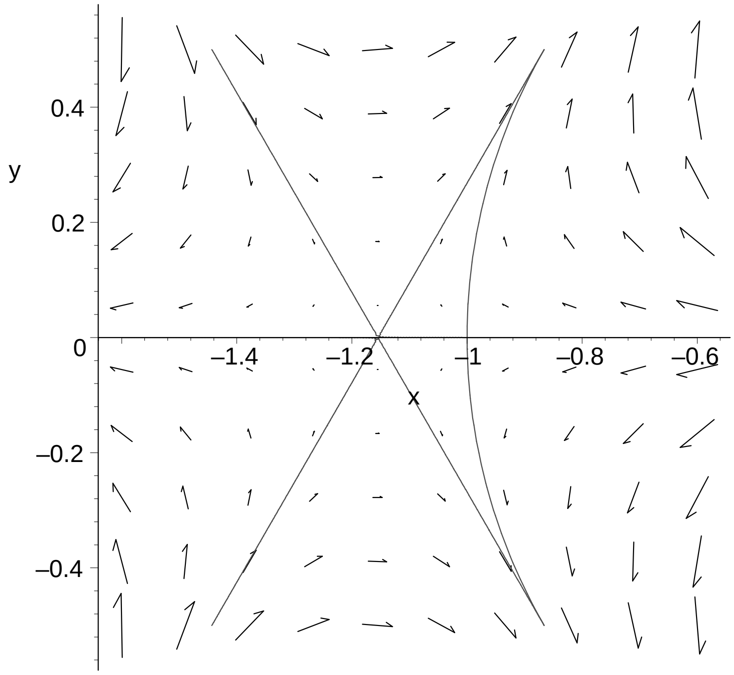


Fig. 6a. Force field for $z + (\sqrt{3}/2)z^2 + z^3/4$.

This cubic has nontrivial zeroes at $-\sqrt{3} \pm i$, a critical point of multiplicity 2 at $z = -2/\sqrt{3}$, and branch point $w = -2/\sqrt{9}$. The three zeroes lie equiangularly spaced on the circle $|z+2/\sqrt{3}| = 2/\sqrt{3}$, producing the single equilibrium point at the center of the equilateral triangle they form. We might anticipate that p_3 will twist D around the branch point. We find the basin boundaries by finding the preimage of the real axis. This gives the three lines $-2/\sqrt{3} + r[\cos(k\pi/3) + i\sin(k\pi/3)]$, for $k = 0, 1, 2,\ -\infty < r < \infty$. See Figure 6a.

B_0 is bounded by lines corresponding to $k = 1, 2$ above, and D is contained in B_0. Notice that some flowlines emanating from 0 exit D and then re-enter, and exit once again. One such is shown in Figure 6b and 6c. This implies that the corresponding image ray will exit $p_3(D)$, re-enter and exit as well. The image of D under p_3 is shown in Figure 7a and 7b, verifying our observations.

This polynomial is one-to-one on D, but only just so! It bends D around to meet itself on the negative real axis. This polynomial was introduced by the topologist J.W. Alexander in [1], the reading of which is an excellent exercise in understanding polynomial mappings.

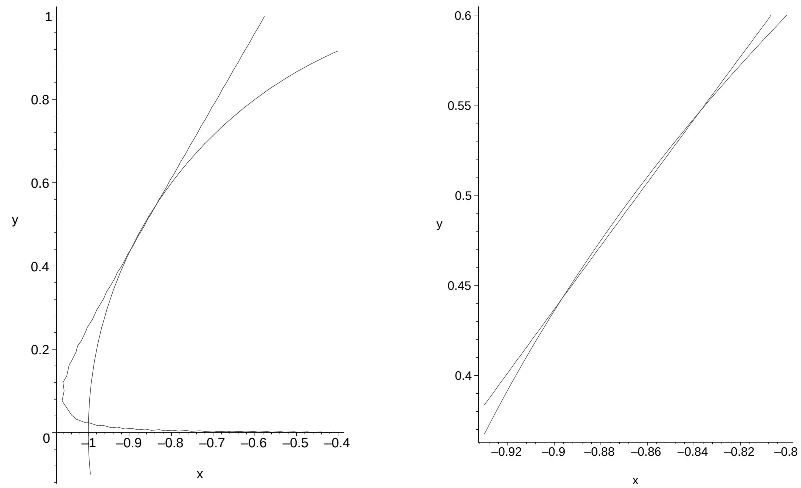


Fig. 6b, 6c. The unit disc and a streamline of p_3.

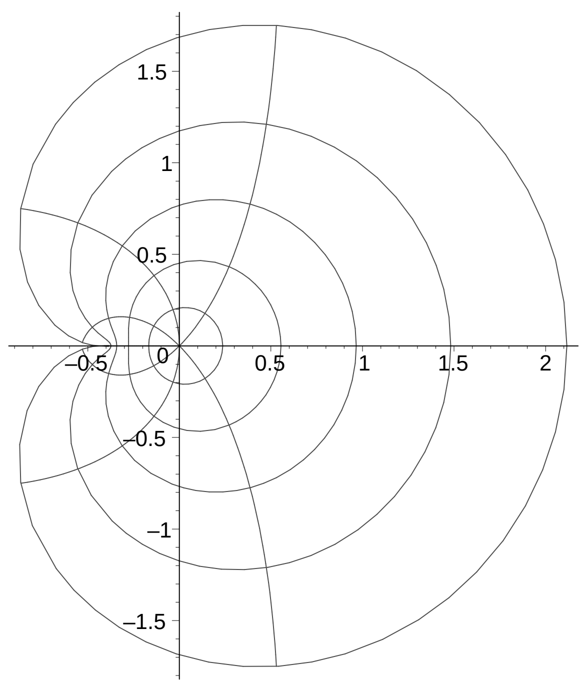


Fig. 7a. $p_3(z) = z + (\sqrt{3}/2)z^2 + z^3/4$

Our next example shows how complicated even cubics can be: let $p_4(z) = z + (7/8)z^2 + (7/25)z^3$. The roots, critical points and branch points are all conjugate pairs in the second and third quadrants. We omit the expressions since they are

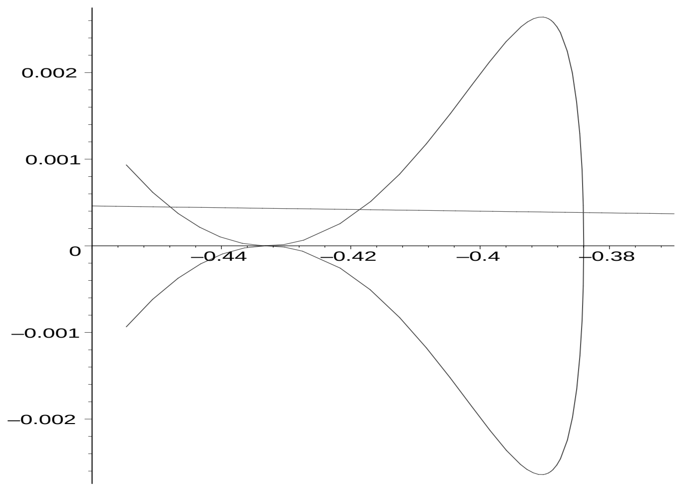


Fig. 7b. $p_3(z) = z + (\sqrt{3}/2)z^2 + z^3/4$, closeup of loop and ray.

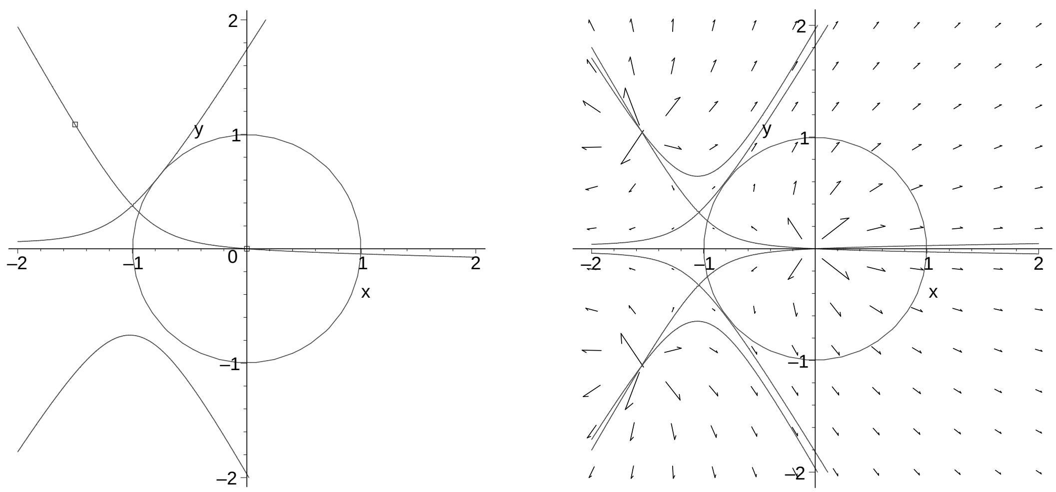


Fig. 8a, 8b. Basin boundaries and streamlines for $p_4(z) = z + (7/8)z^2 + (7/25)z^3$.

complicated and won't add to our analysis. Preimages of the lines through the branch points are real cubics of two variables known as cubic hyperbolae. One is shown in Figure 8a. A zero is marked by a square in the second quadrant. The intersection of two portions of the cubic hyberpola is a critical point.

There are three basins. The boundaries of the B_k are the portions of these curves which pass through no zero of p_4. They are darkened in Figure 8b. The zeroes are the centers of the three collections of longest arrows. It can be shown using MAPLE

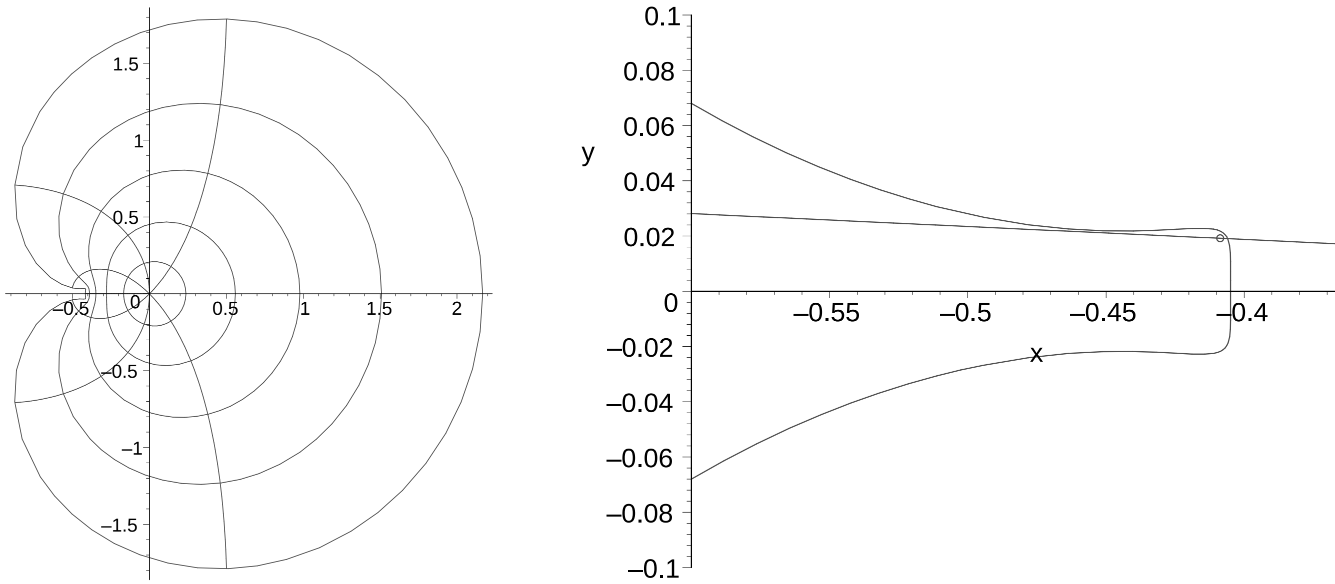


Fig. 9a,b. $p_4(z) = z + (7/8)z^2 + (7/24)z^3$

that D is contained in B_0. The image of D under p_4 is shown in Figure 9a. Figure 9b contains an enlargement of the part of the image near $w = -0.5$ along with a ray emanating from the origin.

The collection of curves in 8a illustrates a sharpening of the Gauss-Lucas theorem. Note that the critical point in the second quadrant lies at the end of a force line from 0 and one from the zero in the same quadrant. Thus if a streamline connects one zero of a polynomial to another, there must be a critical point on the line. This was proved by Hibbert. His proof did not use the force field, however. Also, one does not always have such a streamline for two arbitrary zeroes.

Moving to quartics, we let $p_5(z) = z + (7/6)z^2 + (4/6)z^3 + (1/6)z^4$. In this case we have zeroes at 0, -2, and $-1 \pm i\sqrt{2}$, with critical points at -1, $-1 \pm i\sqrt{2}/2$. We have built on p_1 by adding to $\{0, -2\}$ a conjugate pair of forces. These new forces preserve the critical point at $z = -1$ but are far enough away from C to not introduce any more such points in D. We find that there are four basins bounded by the line $\{z : \Re(z) = -1\}$ and the two branches of the hyperbola as shown in Figure 10.

Again it can be shown that D is contained in B_0, so p_5 maps D univalently. Note the healthy magnitudes coefficients of p_5!

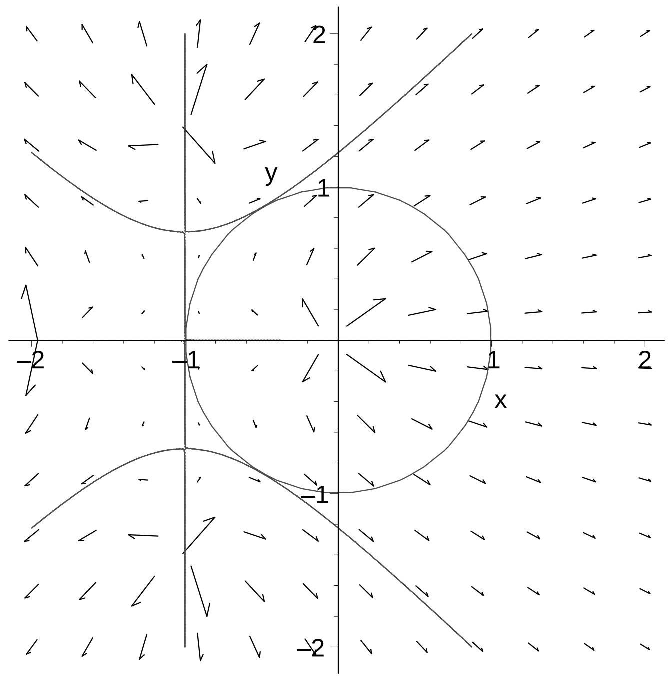


Fig. 10. Force field and basins for $p_5(z) = z + (7/6)z^2 + (4/6)z^3 + (1/6)z^4$.

We end this section with a higher order polynomial found by experimenting with the placements of the nontrivial zeroes. Set them around the circle $\{z : |z| = r\}$ at angles $\pm\pi/8$, $2\pi/3 \pm \pi/8$, and $4\pi/3 \pm \pi/8$. This choice yields $p_6(z) = z - \sqrt{2-\sqrt{2}}z^4/r^3 + z^7/r^6$. The choice of $r = 1.4$ gives the "septic" which sends D to the image shown in Figure 11. The basins are quite complicated in this case, but it is straighforward to keep D inside B_0. By allowing the zero pairs to get close together while pulling then further away from the origin (to weaken their acquired strength) we can produce many variations on this example.

6. Starlike Polynomials

Among the polynomial examples given so far we can detect one difference. For some, the force field streams emanating from 0 exit D and never re-enter. The polynomials p_1 and p_2 exhibit this. For others, given streamlines may exit, re-enter, and exit again, as in p_3. The first two are examples of *starlike* complex functions. Each streamline starting at 0 and ending on a point of C is sent to a ray in the range also starting at 0. The collection of all such rays is called a "star" centered at 0. The star as a set in the complex plane was introduced by the Swedish mathematician

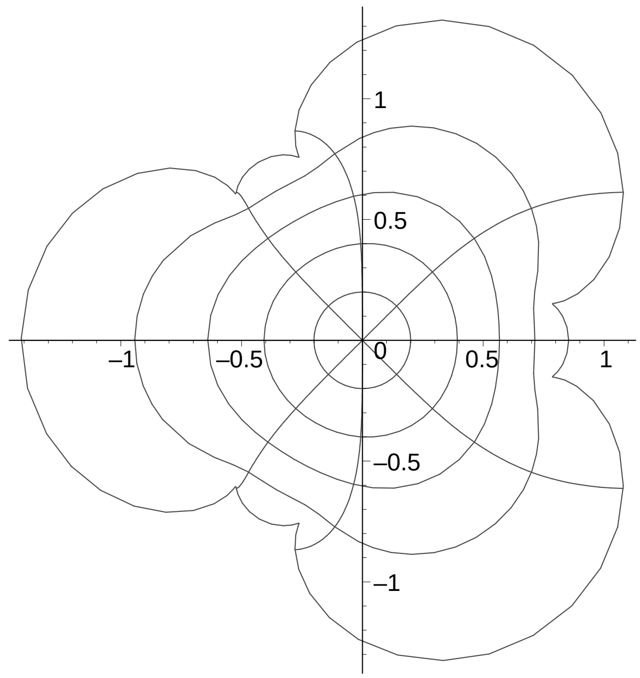


Fig. 11. $p_6(z) = z - \sqrt{2-\sqrt{2}}z^4/r^3 + z^7/r^6$ with $r = 1.4$.

G. Mittag-Leffler around the turn of the 20th century. Alexander defined it as "A region every point of which may be joined to a point 'a' [the center] by means of a linear segment consisting only of points of the region ... " The topologist looked at polynomials which mapped D onto star-shaped images centered at the origin by looking at the image of C as a curve. He noted that for these functions f the polar argument of $f(z)$ "...is a never decreasing function of $\theta = \arg(z)$ as z describes the unit circle [C] in the positive sense". In our example p_1 we saw this behavior. Hibbert noted that each of his "cellulues d'univalence" are mapped to "une étoile de Mittag-Leffler".

Our force-field interpretation is useful in dealing with starlike polynomials. The fields which yield star-shaped images are those where the force vectors from the origin have an "always exiting" behavior. More exactly, isolate a force vector at $z = r(\cos(t) + i\sin(t))$, where $0 < r < 1$. The tangent line to $\{z : |z| = r\}$ at z divides $\mathbb{C}$ into two half planes. The force vector will lie in the plane which does not contain $\{z : |z| = r\}$, thus pointing "outward" from the circle. This property does not allow a streamline to exit D and re-enter, as happens in p_3.

Using this idea, it is easy to construct a family of starlike univalent polynomials $p(z) = z + a_2 z^2 + \cdots + a_n z^n$ given by Alexander, and to prove one of his theorems related to them. Given a zero at 0, he found a "safety radius" R such that *any* placement of the $n-1$ remaining zeroes $z_1,\ z_2, \ldots z_{n-1}$ with $|z_k| \geq R$ yields a starlike univalent polynomial. For us this is asking for a value R with the following property: placing zeroes beyond or on $\{z : |z| = R\}$ insures that the field from the origin is "always exiting" concentric circles centered at 0 with radius less than 1. We claim that $R = n$ will do, and that it is the best that can be done. If we fix $z = r$ in D, then placing $z_1 = z_2 = \cdots = z_{n-1} = n$ produces a resultant force at z equal to $1/r - (n-1)/(n-r) = n(1-r)/r(n-r) > 0$, with equality if $z = 1$. It is clear that if the nontrivial zeroes are any complex numbers greater in magnitude than n, the "outward pointing" condition still holds at $z = r$, since the sum of the forces from $\{z_1,\ z_2, \ldots z_{n-1}\}$ is weakened. A similar argument applies to any z in D.

Thus for each integer $n > 1$ we have a starlike univalent polynomial $s_n(z) = z(1 - z/n)^{n-1}$, with all nontrivial sources concentrated at $z = n$. We also have a whole family of others with nontrivial zeroes larger in magnitude than n. We get a bonus: using $S_n(z) = [z(1 + z/n)^n]/(1 + z/n)$, and letting $n \to \infty$, we can conclude that ze^z is starlike and one-to-one on D, where $e^{x+iy} = e^x(\cos(y) + i\sin(y))$ is the complex exponential function. This requires knowing that the limit of univalent functions is univalent, a fact we borrow from complex variable theory. (Actually the convergence is required to be uniform on closed subdiscs of D.)

Finding such "safety radii", known as *radii of univalence* or *radii of starlikeness*, became of interest to some leading analysts of the early 20th century. Although hardly in the forefront of their research interests, the French mathematician J. Dieudonné and Hungarian mathematicians L. Féjèr and G. Szegö each visited this problem.

7. "Analyzing" Starlike Polynomials

We want to explore the work of starlike univalent polynomial construction from a different point of view and show how hard it can be. We will replace our force field condition with a compact one involving p and its derivative. Specifically, we will show that the "always exiting" property for p can be written as $\Re[zp'(z)/p(z)] > 0$ for z in D. The argument involves only elementary properties of complex numbers. To this end, we assume $p(z) = z(1 - z/z_1) \ldots (1 - z/z_{n-1})$ is starlike univalent. Let

$z = r(\cos(t) + i\sin(t))$ be in D, then the "always exiting" vector condition can be written as

$$t - \pi/2 < \text{polar argument}[\frac{1}{\overline{z} - \overline{0}} + \sum_{k=1}^{n-1} \frac{1}{\overline{z} - \overline{z_k}}] < t + \pi/2.$$

But remember that the polar argument for a complex number behaves logarithmically: $\arg(w_1 w_2) = \arg(w_1) + \arg(w_2)$. Thus we can subtract the t from each inequality and absorb it into the central expression, distributed on each term:

$$-\pi/2 < \arg[1/r + \sum_{k=1}^{n-1} \frac{\cos(-t) + i\sin(-t)}{\overline{z} - \overline{z_k}}] < \pi/2.$$

But the argument inequality remains the same if we multiply each term through by $r > 0$ and replace the polar complex number with a conjugate:

$$-\pi/2 < \arg[1 + \sum_{k=1}^{n-1} r\frac{\overline{\cos(t) + i\sin(t)}}{\overline{z} - \overline{z_k}}] < \pi/2.$$

The properties of complex conjugates allow us to pull the conjugates off of each term and replace them by a the conjugate of a single expression:

$$-\pi/2 < \arg\overline{[1 + \sum_{k=1}^{n-1} r\frac{\cos(t) + i\sin(t)}{z - z_k}]} < \pi/2.$$

Equivalently

$$-\pi/2 < \arg\overline{[1 + \sum_{k=1}^{n-1} \frac{z}{z - z_k}]} < \pi/2.$$

But this condition just says that the quantity inside the argument has positive real part:

$$\Re[\frac{z}{z - 0} + \sum_{k=1}^{n-1} \overline{\frac{z}{z - z_k}}] > 0.$$

Recalling Gauss' derivation of the force field idea we can remove the conjugates, since a number has positve real part if and only if its conjugate does:

$$\Re[\frac{z}{z - 0} + \sum_{k=1}^{n-1} \frac{z}{z - z_k}] > 0.$$

As in Gauss' equation we realize that this is the same as $\Re[zp'(z)/p(z)] > 0$ for z in D. These steps can be reversed. This is what we are after, an "analytic" equivalent condition for starlike univalence.

Before using this we note that the condition was introduced by the Finnish mathematician R. Nevanlinna in the 1920's to describe starlike behavior for any power series $f(z) = z + a_2z^2 + \cdots$ in D. Our derivation of it for polynomials appears to be novel. The exact relation for an arbitrary function $f(z) = z + a_2z^2 + \cdots$ complex differentiable in D is: f is univalent and starlike in D if and only if $\Re[zf'(z)/f(z)] > 0$ in D. For details on this statement, see [7], section 12.2. For example, $h(z) = z/(1-z)$ is a linear fractional transformation which maps D univalently onto $\{w : \Re(w) > -1/2\}$. It is then clearly starlike and univalent. An easy calculation shows that $zh'(z)/h(z) = 1/(1-z)$, which takes D onto $\{w : \Re(w) > 1/2\}$. This confirms the starlike univalence of h using the condition.

Another direct connection that can be made between the test quantity $\Re[zf'(z)/f(z)]$ and the geometry of an arbitrary function f is the relation $\Re[zf'(z)/f(z)] = \partial/\partial t[\arg\ f(r\cos(t) + ir\sin(t))]$, which can be shown by elementary properties of complex differentiation and the complex logarithm. Specifically, we first note that the complex logarithm can be defined as the inverse of the complex exponential mentioned earlier. A little consideration shows that this yields $\log(z) = \ln|z| + i\arg(z)$. For a complete justification of this definition, see [7], section 4.3. As long as we are working "locally", that is, in the neighborhood of a fixed $z \in D$, the principal value or any appropriate adjustment can be used for $\arg(z)$. By differentiating the relation $\exp(\log(z)) = z$ with respect to z we obtain $d/dz(\log(z)) = 1/z$, as in the real case. Next, use the Euler relation for the polar form of a complex number, $z = r(\cos(t) + i\sin(t)) = re^{it}$. Our connection follows:

$$\begin{aligned} d/dt[\arg\ f(re^{it})] &= d/dt\ \Im[\log(f(re^{it}))] \\ &= \Im[d/dt\ \log(f(re^{it}))] \\ &= \Im[ire^{it}f'(re^{it})/f(re^{it})] \\ &= \Re[zf'(z)/f(z)] \end{aligned}$$

where we have also used the chain rule. So our analytic condition $\Re[zf'(z)/f(z)] > 0$ is related to the increasing of the polar argument of $f(r\cos(t)+ir\sin(t))$ as described by Alexander in his definition of starlikeness. We also mentioned this in connection with P_1 in the Introduction.

As an application we analyze the cusp at the critical point $z = -1$ for p_1. If we look at Figure 4 we see that this critical point is the equilibrium for the forces from the origin and $z = -2$. If we replace the bit of C near this equilibrium with a vertical line through it, the force field acts on the line near $z = -1$ much as it does on the arc. But clearly the resultant forces on this line have zero horizontal component. Thus the resultant vectors on the arc have argument nearly $\pi/2$ above the real axis and nearly $-\pi/2$ below. Thus by our derivation $\Re[zf'(z)/f(z)]$ is nearly zero on this arc. But by the connection just stated, the image of this arc then has small argument change, a change nearing zero as we approach the equilibrium point. This is consistent with the cusp.

8. More On Starlike Polynomials

An analyst easily warms up to the condition in the previous section. To use it one usually writes

$$\frac{zp'(z)}{p(z)} = \frac{zp'(z)}{p(z)} \cdot \frac{\overline{p(z)}}{\overline{p(z)}} = \frac{zp'(z)\overline{p(z)}}{|p(z)|^2}.$$

Thus testing the real part of the numerator $N(z) = zp'(z)\overline{p(z)}$ for positive real part in D is all that is necessary. For our epicycloidal family $z + z^n/n$ we can calculate $\Re[N(r(\cos(t) + i\sin(t))] = \dfrac{r^{n-1}}{n}[1 + \cos(n-1)t] + [1 + r^{n-1}\cos(n-1)t] > 0$ for all r and t. This is a demonstration by analysis that the family is starlike univalent.

The condition is often difficult to apply. For example, for the arbitrary cubic $p(z) = z + a_2z^2 + a_3z^3$, even if we assume a_2 and a_3 are real, we have

$$N(r(\cos(t) + i\sin(t)) = r[1 + 3a_3^2r^4 + (4a_2 + 5a_2a_3r^2)r\cos(t) + 4a_3r^2\cos(2t)].$$

Thus checking in general for positivity amounts to choosing a_2 and a_3 so that this trig expression is positive for all r and t! Some relief is available. The function $\Re[zf'(z)/f(z)]$ for $f(z) \neq 0$ on D is *harmonic* ([7], Chapter 10). This property is physically exemplified by a steady-state temperature distribution $T(x, y)$ on D induced by heat sources outside of D. One of the many properties of harmonic functions is that positivity on C implies positivity throughout D. (The temperature on the interior of a circular plate will not drop below its temperature on the rim.) Thus we need only check for the positivity of $\Re[N]$ on C. This allows us to take $r = 1$ in our computations.

We can combine this test with the clever choice of point sources for the force field. For example, suppose we wish to place a conjugate pair $a \pm bi$ to induce a critical point at $w = 1$ but hope to preserve starlikeness. We may invoke the *center of gravity with respect to* $z = 1$. This is the point α at which the force at $z = 1$ due to the pair is equal to the force at $z = 1$ due to a double particle at α. If we set $\alpha = 3$ we can solve the force equation

$$\frac{1}{\overline{1} - \overline{a+bi}} + \frac{1}{\overline{1} - \overline{a-bi}} = \frac{2}{\overline{1} - \overline{3}}$$

to get $a + b^2/(a-1) = 3$, or, equivalently, $(a-2)^2 + b^2 = 1$. Thus any pair $a \pm bi$ of point sources on this circle will act at $z = 1$ like a double point source at $\alpha = 3$. Not every pair will give a starlike polynomial, however. We must choose and experiment. If we let $a \pm bi = 2 \pm i$, we obtain $p_7(z) = z - (4/5)z^2 + (1/5)z^3$. We then obtain $N(\cos(t) + i\sin(t)) = (8/5)(1 - \cos(t))^2 > 0$ for all t. Thus p_7 is starlike univalent in D. Often a computer algebra system is needed in more complicated examples to test positivity. But it is not hard to believe that in general we have a difficult analysis problem before us.

The authors have constructed a family of quartic starlike univalent polynomials in the following manner. Begin with the epicycloidal $z - z^4/4$, where the nontrivial zeroes are distributed at equal angles around $\{z : |z| = 4^{1/3}\}$. Then reduce the two angles between the three successive nontrivial zeroes, keeping them equal, while maintaining the zeroes on $\{z : |z| = 4^{1/3}\}$. This increases the total force at $z = 1$ from these zeroes. Now, keeping the magnitude of the zeroes equal, move them far enough away from the origin that the resulting polynomial still has the exiting force field condition on D. We allow a critical point at $z = 1$. The analytic test condition is used to determine how far back to pull the zeroes in order that the resulting the polynomials are starlike. The terminal polynomial in this family is the Alexander example $s_4(z) = z(1 - z/4)^3$ where the angle between any two nontrivial zeroes is 0. If we let a be our parameter, the family is given by

$$p(a, z) = z - \frac{a^3 - 4}{a^2(2a-3)}z^2 + \frac{a^3 - 4}{a^3(2a-3)}z^3 - \frac{1}{a^3}z^4, \text{ for } 4^{1/3} < a < 4.$$

As we allow a to vary we get a catalogue of images of D under the resulting polynomial mappings. Some of these are shown in Figure 12. When $a = 2$ we have $p_8(z) = z - z^2 + (1/2)z^3 - (1/8)z^4$, with zeroes at $z = 0,\ 1 \pm i\sqrt{3}$, and $z = 2$. The

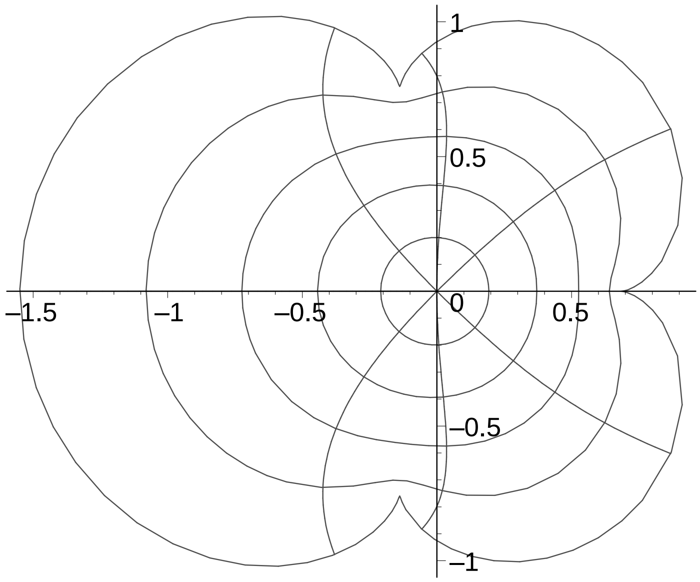


Fig. 12a. $p(a, z)$ for $a = 1.6$

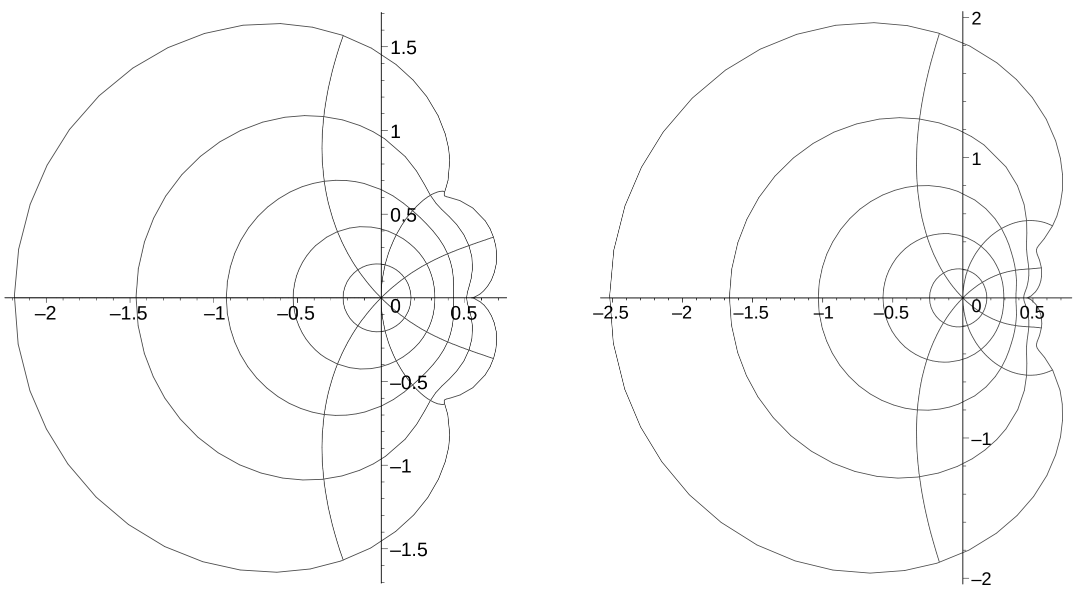


Fig. 12b, 12c. $p(a, z)$ for $a = 1.65, 1.72$.

nontrivial zeroes are placed equiangularly on $\{z : |z| = 2\}$. The critical point at $z = 1$ created by the zero set $\{ 0, 2 \}$ is maintained.

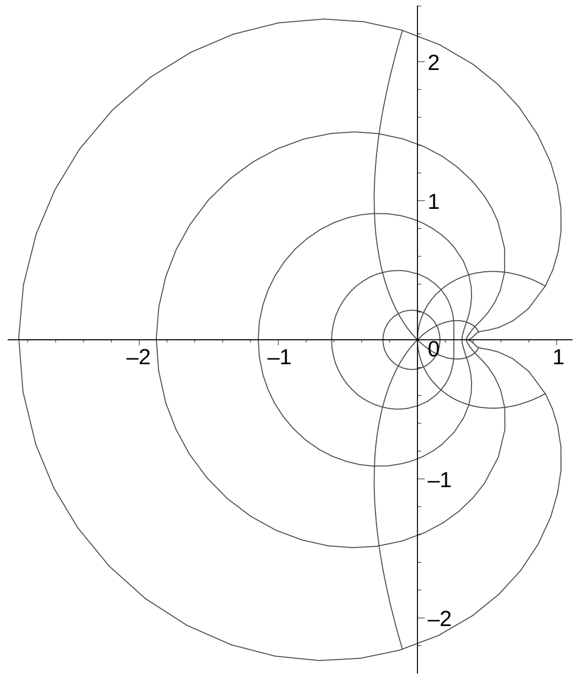
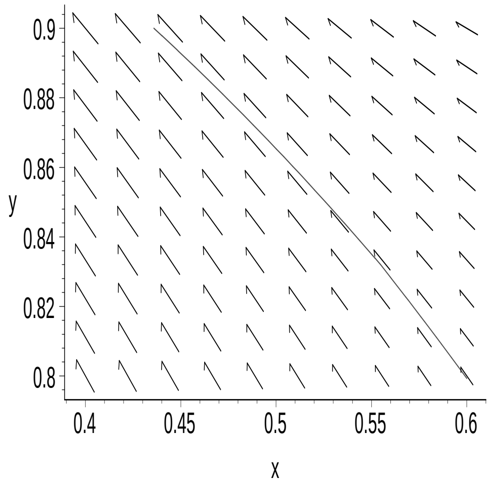


Fig. 13a. $p_9(z) = z - 1.085z^2 + .61344z^3 - .1662z^4$.

We conclude this section by modifying p_8. Maintain the arguments of the nontrivial zeroes but pull them closer to the origin. Specifically, replace $z_1 = 2$ by $z_1^* = 1.77$, and $z_2, z_3 = 1 \pm i\sqrt{3}$ with $z_2^*, z_3^* = 1.92(\cos(\pi/3)) \pm i\sin(\pi/3))$. One might think that this extinguishes univalence, but it doesn't. The critical point is moved to $z = 1.01\ldots$. Thus there is no cusp at $z = 1$, though it might appear so from Figure 13a. Moreover, near-tangential exiting of the flow field at two conjugate arcs on C occurs. See Figure 13b for one arc in the upper half plane. This is due to the confluence of streams from the origin and the remaining zeroes. A plot of the image of D under the resulting $p_9(z) = z - 1.085z^2 + .61344z^3 - .1662z^4$ is shown in Figure 13a. Note the enlarged "eyedropper" shape near $z = 1$ in Figure 13c. The tangent line in Figure 13c shows the image at a point where $\partial/\partial t(\arg(f(r(\cos(t) + i\sin(t))$ is small. This is the image of a point near $z = \cos(0.9) + i\sin(0.9)$, a point on the arc where the near-tangential exiting of the force field occurs. We have reached the limits of starlikeness by this near-tangential exiting in the domain. It is reflected by the argument of the image curve undergoing instantaneous almost-zero change.



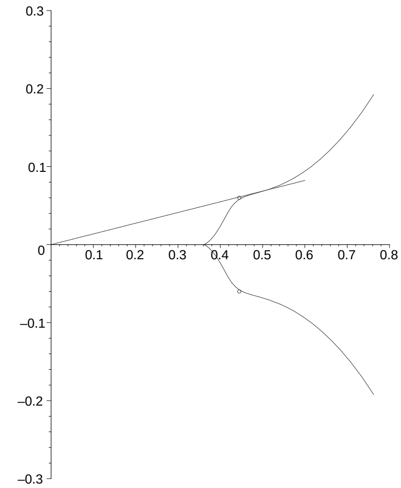


Fig. 13b, 13c. Force field for $p_9(z)$ and enlargement of image curve.

9. Conclusion

We have looked at the problem of finding one-to-one polynomial mappings of D from several points of view. It can be seen as a problem in algebra, though we didn't dwell on this aspect. Others have, and found it rough going. In a 1951 paper [**?**] in which the entire collection of cubics univalent on D is found algebraically the author remarks: "For polynomials of higher degree than three the discussion of the resultant is not an easy task because not only the resultant is of high degree but the explicit calculation of the functions of Sturm surpasses the possibilities of a normal man." (The resultant and Sturm functions are algebraic expressions.) Marden's book surveys the larger algebraic question of finding roots and critical points of polynomials. The geometry of complex mappings has played an important role in our work. We examined many curves, images and preimages under polynomials to help locate the domains of univalence. We needed to understand what these mappings do at critical points and "in the large". Complex and real analysis also played a role in considering starlike polynomials.

We hope that our use of the "mechanical analogy" provides a platform from which to explore these mappings. By studying the associated force field one can not only explore univalence but can to some extent predict mapping behavior, as we saw in the case of starlike polynomials. It is interesting to note that there are some univalent polynomials which don't arise from our approach. Let $p_{10}(z) = z+2\sqrt{2}/3z^2+(1/3)z^3$. Then p_{10} can be shown univalent on D, but the corresponding force field emanating from 0 does not cover D. As an exercise the reader may wish to analyze the field and basins for p_{10} and see how the image of D under p_{10} arises.

Our polynomials examples had real coefficients. This was the result of choosing nontrivial zeroes in conjugate pairs or on the real axis. But clearly the method applies to any nontrivial zero set. An easy example occurs by choosing $z_1 = 1.8$ and $z_2 = 1.825(1 + i)$. The reader can show that this polynomial is univalent on D by basin analysis.

The basins for the force fields provide a tool for decomposing the entire complex plane into tractably mapped regions, as Hibbert showed. The graph of a real polynomial can be understood by seeing where it is monotone. The study of the global behavior of a complex polynomial can be initiated by looking at its behavior on basins. But that is a much larger project. Our present investigation is ended.

References

[1] Alexander, J. W. II, *Functions which map the interior of the unit disk upon simple regions*, Ann. of Math. **17**(1915) 12-22.

[2] Braden, B., *Picturing functions of a complex variable*, The College Mathematics Journal **16, 1**(1985) 321-327.

[3] Cipra, B. "From Knot to Knot", in *What's Happening in the Mathematical Sciences*, Vol. 2, pp. 8-13, American Mathematical Society, 1994

[4] Curtiss, D. R., *A Mechanical Analogy In the Theory Of Equations*, Science, Vol. **55** February 24, 1922, pp. 189 - 194.

[5] Kössler, M., *Simple polynomials*, Czechoslovak Math. J., **1, 76**(1951) 5-15.

[6] Marden, M., *Geometry of Polynomials*, Amer. Math. Soc., Providence, R. I., 1966

[7] Silverman, H. *Complex Variables*, Houghton Mifflin, Boston, 1975

[8] Walsh, J. L., *The Location of the Critical Points of Analytical and Harmonic Functions*, AMS Colloquium Publication, vol. 34, AMS, Providence, RI, 1950

Department of Mathematical Sciences, Villanova University, Villanova, PA 19085

E-mail address: `frederick.hartmann@villanova.edu, alan.gluchoff@villanova.edu`