ALTERNATIVE UNIVERSES AND COMPLEX ANALYSIS

RACHEL W. HALL* AND KREŠIMIR JOSIĆ†

Abstract. Trajectories of Hooke's Law in the complex plane, which are conic sections, are mapped onto trajectories of Newton's Law of Gravitation via the transformation $z \to z^2$. Newton's Law of Ellipses (objects attracted to a center by a force inversely proportional to the square of the distance travel in conic sections) follows from a geometric analysis of this map. An extension of this approach reveals a similar relation between more general pairs of power laws of centripetal attraction. The implications of these relations are discussed and a Matlab program is provided for their numerical study. This material is suitable for an undergraduate Complex Analysis class.

Key words. Two-body problem, Functions of a complex variable, Geometric function theory.

AMS subject classifications. 70F05, 30A99, 30C99.

1. Introduction. It is commonly known that paths of objects attracted to a center by a force inversely proportional to the square of the radius (Newton's Law of Gravitation) or by a force directly proportional to the radius (Hooke's Law) are conic sections. This striking coincidence can be explained by the fact that the two laws are related by the complex transformation $z \to z^2$ and can be called dual. In general, two laws of centripetal attraction are dual if the transformation $z \to z^\alpha$ takes a trajectory of one law to a trajectory of the other. Questions about duality date back to Newton, who was aware of several such pairs [4, pp. 114–125]. In the following discussion, complex analysis is used to demonstrate duality between power laws of centripetal attraction. It also gives some insight into the bizarre occurrences possible in a universe subject to a different law of gravity. The Appendix contains a Matlab program which plots dual trajectories. The body of this article is based on presentations by Arnold [3] and Saari [7].

2. Hooke's Law and Newton's Law. The following proof that the complex transformation $z \to z^2$ takes trajectories of Hooke's Law $w'' = -Cw$ to trajectories of Newton's Law $Z'' = -\tilde{C}\frac{Z}{|Z|^3}$ is attributed to Bohlin and Sundman.[1] Sundman tried to find a transformation that would remove the singularity present at the origin in $Z'' = -\tilde{C}\frac{Z}{|Z|^3}$. He noticed that nonsingular orbits coming close to the origin spin around it once and are ejected in a direction close to the incoming one. If one extends the singular orbits coming towards the origin so that they bounce off the origin and return back along their incoming paths, they correspond to images of lines passing through the origin under the map $z \to z^2$. Using this fact Sundman was able to find a series solution for the three-body problem [7]. Figure 2.1 shows an orbit of Hooke's Law which passes through the origin as a limit of orbits passing close to the origin, and its image under the squaring map, which is a path oscillating between the origin and another point.

THEOREM 2.1. *Suppose the motion of a point in the complex plane is given by Hooke's Law $w'' = -Cw$. Under the transformation $Z(\tau(t)) = [w(t)]^2$ for a suitable reparametrization of time $\tau(t)$ (depending on w) the motion of the point is described by:*

$$\frac{d^2 Z}{d\tau^2} = -\tilde{C}\frac{Z}{|Z|^3}.$$

where $\tilde{C} = 2(|w(0)'|^2 + C|w(0)|^2)$. That is, the point moves according to Newton's Law of Gravitation.

* Department of Mathematics, The Pennsylvania State University, University Park, PA 16802

† Department of Mathematics and Statistics, Boston University, 111 Cummington Street, Boston, MA 02215

[1] Arnold [3] attributes the idea of the proof to Bohlin and Saari [7] attributes it to Sundman.

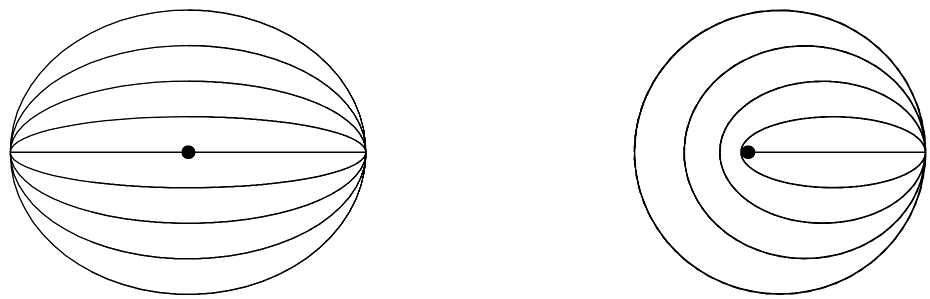

Fig. 2.1. *Trajectories for $w'' = -w$ and their images under the map $z \to z^2$. Note that the degenerate orbit is mapped to an orbit which "bounces off" the origin.*

Proof. First we reparametrize t. Newton showed that the law of areas, also known as Kepler's Second Law, must hold in any central field. This means that the line connecting an object to the center must sweep out equal areas in equal times. We can express Z and w in polar coordinates as $Z = (|Z|, \theta_Z)$ and $w = (|w|, \theta_w)$. Let A_1 and A_2 be the areas swept out by $w(t)$ and $Z(\tau)$, respectively. A calculus exercise shows that $w(t)$ satisfies the law of areas, and if $Z(\tau)$ is to satisfy Newton's Law, it too must satisfy the law of areas. Assuming that w, and hence Z, is nondegenerate (meaning, for our purposes, that it does not lie on a line through the origin),

$$\text{constant} = \frac{\frac{dA_1}{dt}}{\frac{dA_2}{d\tau}} = \frac{\frac{1}{2}|w|^2\frac{d\theta_w}{dt}}{\frac{1}{2}|Z|^2\frac{d\theta_Z}{d\tau}} = \frac{1}{2}\frac{d\tau}{dt}\frac{1}{|w|^2},$$

using the expression for area in polar coordinates and the identity $\theta_Z = 2\theta_w$. Therefore the reparametrization $\tau(t)$ must satisfy

$$\frac{d\tau}{dt} = k|w|^2$$

for some constant k. We will choose $k = 1$. If w lies on a line through the origin, we will consider w a limit of nondegenerate trajectories and use the same reparametrization.

The rest of the proof is a matter of using the chain rule

$$\begin{aligned}
\frac{d^2Z(\tau)}{d\tau^2} &= \frac{1}{|w(t)|^2}\frac{d}{dt}\left(\frac{1}{|w(t)|^2}\frac{dw^2}{dt}\right) \\
&= \frac{2}{w\overline{w}}\frac{d}{dt}\left(\frac{1}{\overline{w}}\frac{dw}{dt}\right) \\
&= -\frac{2}{w\overline{w}^3}\frac{dw}{dt}\frac{d\overline{w}}{dt} + \frac{2}{w\overline{w}^2}\frac{d^2w}{dt^2} \\
&= -\frac{2}{w\overline{w}}\left[\overline{w}^{-2}\frac{dw}{dt}\overline{\left[\frac{dw}{dt}\right]} + C\frac{w}{\overline{w}}\right] \\
&= -2w^{-1}(\overline{w})^{-3}[|w'|^2 + C|w|^2]
\end{aligned}$$

Now let $E_w = \frac{1}{2}(|w'|^2 + C|w|^2)$, and note that differentiation shows that E_w is constant on any trajectory. In physical terms, E_w is called the *energy* of the harmonic oscillator and is constant along the trajectories of motion because no friction is present to dissipate it. Then

$$\frac{d^2Z}{d\tau^2} = -4E_w w^{-1}\overline{w}^{-3} = -4E_w \frac{Z}{|Z|^3},$$

which is what we intended to show. □

So $z \to z^2$ maps parametric curves $w(t)$ which satisfy the equation $w'' = -Cw$ into parametric curves satisfying the equation $Z'' = -\tilde{C}\frac{Z}{|Z|^3}$ after a suitable reparametrization of time. Now we will show that the mapping is onto, at least for nonsingular orbits. A triple (w_0, w_0', C) consisting of initial conditions plus a choice of constant C for Hooke's Law uniquely determines a triple $(Z_0, Z_0', \tilde{C})$ giving initial conditions and the constant $\tilde{C}$ for the inverse square law by

$$(Z_0, Z_0', \tilde{C}) \quad = \quad \left(w_0^2, \ \frac{2w_0}{|w_0|^2} w_0', \ 2(|w_0'|^2 + C|w_0|^2) \right)$$

The mapping $(w_0, w_0', C) \to (Z_0, Z_0', \tilde{C})$ is onto and two-to-one on the initial conditions (except where $w_0 = 0$, where it is undefined) and onto the values for $\tilde{C}$.

Now we consider the effect of the squaring map on orbits. The differential equation $w'' = -Cw$ can be written as

$$\begin{aligned} x' &= y \\ y' &= -Cx \end{aligned}$$

and this system of equations is invariant under the map $(x, y) \to (-x, -y)$. With a little work, one can verify that trajectories in the complex plane satisfying Hooke's Law are ellipses in the case $C > 0$, lines when $C = 0$, and hyperbolas when $C < 0$ (more precisely, one branch of a hyperbola). An elliptical trajectory which satisfies the system and contains (w_0, w_0') will also contain $(-w_0, -w_0')$. Since (w_0, w_0') and $(-w_0, -w_0')$ are mapped to the same point, the squaring map is one-to-one on elliptical orbits. However, in the case of the non-closed orbits, (w_0, w_0') and $(-w_0, -w_0')$ lie in different trajectories (either in parallel lines or the different branches of a hyperbola), and the squaring map is two-to-one. In either case, any nonsingular orbit satisfying Newton's Law of Gravitation is an image of a curve satisfying Hooke's Law under the transformation $z \to z^2$.

The situation of singular orbits is more problematic. The map $z \to z^2$ takes differentiable orbits to nondifferentiable orbits. Any preimage of a singular orbit in the Z coordinates (that is, any orbit that passes through the origin in the w coordinates) can be brought to the real axis by rotation. Therefore, we will only discuss real solutions to $w'' = -Cw$. The solution will be either of the form

$$w(t) = Kt \tag{2.1}$$

for $C = 0$ or

$$w(t) = K\sin(\sqrt{C}t - \delta) \tag{2.2}$$

for $C > 0$, where K and δ are constants depending on the initial conditions. Under the map $Z = w^2$ these orbits are mapped to a portion of the positive real line. However, the parameterization is not smooth, since for real w,

$$\frac{dZ}{d\tau} = \frac{dt}{d\tau}\frac{d}{dt}w^2(t) = \frac{dt}{d\tau} \cdot 2w(t)w'(t) = \frac{2w'(t)}{w(t)}.$$

If w is given by either Equation 2.1 or 2.2, the speed diverges to infinity. This singularity is reached in finite time.

The natural convention to be adopted is that the object is reflected at the origin. This can be motivated in two ways: first, this is the limiting case as the orbits become more and more eccentric, and second, if we consider the images of the points of w under the squaring map regardless of the singularity, Z changes direction at the origin.[2]

3. The geometry of the map $z \to z^2$. The purpose of this section is to prove Newton's Law of Ellipses using duality and some geometrical properties of the map $z \to z^2$ acting on trajectories of Hooke's Law $w'' = -Cw$. Recall that trajectories in the complex plane satisfying Hooke's Law are ellipses in the case $C > 0$, lines when $C = 0$, and hyperbolas when $C < 0$ (more precisely, one branch of a hyperbola). The elliptic and hyperbolic trajectories are centered at the origin. Now we will investigate the effect that the transformation $z \to z^2$ has on conic sections in the complex plane. The so-called Zhukovskii ellipses and hyperbolas will be useful in our discussion. Their definition is contained in the following lemma:

LEMMA 3.1. *The transformation $T(z) = z + 1/z$ maps circles centered at the origin two-to-one and onto ellipses with foci at* ± 2 (Zhukovskii ellipses) *and lines through the origin two-to-one and onto hyperbolas with foci at* ± 2 (Zhukovskii hyperbolas).

Proof. Suppose $z = \rho e^{i\theta}$. Then

$$\mathrm{Re}(T(z)) = x = \left(\rho + \frac{1}{\rho}\right)\cos\theta$$
$$\mathrm{Im}(T(z)) = y = \left(\rho - \frac{1}{\rho}\right)\sin\theta$$

Eliminating θ by squaring both sides of these equations and adding them leads to

$$\frac{x^2}{\left(\rho + \frac{1}{\rho}\right)^2} + \frac{y^2}{\left(\rho - \frac{1}{\rho}\right)^2} = 1.$$

A circle of radius R centered at the origin is described in polar coordinates by fixing $\rho = R$. Therefore, its image is an ellipse with foci at $\pm\sqrt{\left(R + \frac{1}{R}\right)^2 - \left(R - \frac{1}{R}\right)^2} = \pm 2$, except in the degenerate case, $R = 1$, when the image is the line $[-2, 2]$ (see Figure 3.1). Notice that T takes circles of radius R and $1/R$ to the same ellipse, so the transformation is two-to-one.

Eliminating ρ in a similar way gives

$$\frac{x^2}{4\cos^2\theta} - \frac{y^2}{4\sin^2\theta} = 1.$$

A line through the origin has a fixed angle $\theta = \Theta$ (in polar coordinates). Its image under T is therefore one branch of a hyperbola with foci at ± 2. ∎

LEMMA 3.2. *The map $z \to z^2$ takes a Zhukovskii ellipse to an ellipse with one focus at the origin. It takes a Zhukovskii hyperbola to one branch of a hyperbola that has the origin as its focus, and a line to a parabola with focus at the origin.*

Proof. The map $z \to z^2$ takes circles of radius R centered at the origin to circles of radius R^2 centered at the origin, and lines through the origin to rays from the origin. Since $(z + 1/z)^2 = z^2 + (1/z)^2 + 2$ it follows that $(T(z))^2 = T(z^2) + 2$. Therefore $z \to z^2$ acts on Zhukovskii ellipses and hyperbolas as claimed.

[2] A different view of the problem is given by Milnor [6]. He claims that if we graph $Z(t)$ (which is a *real* variable here) as a function of time it satisfies the equation $(Z')^2 = 2E + C/r$ and is in fact a cycloid.

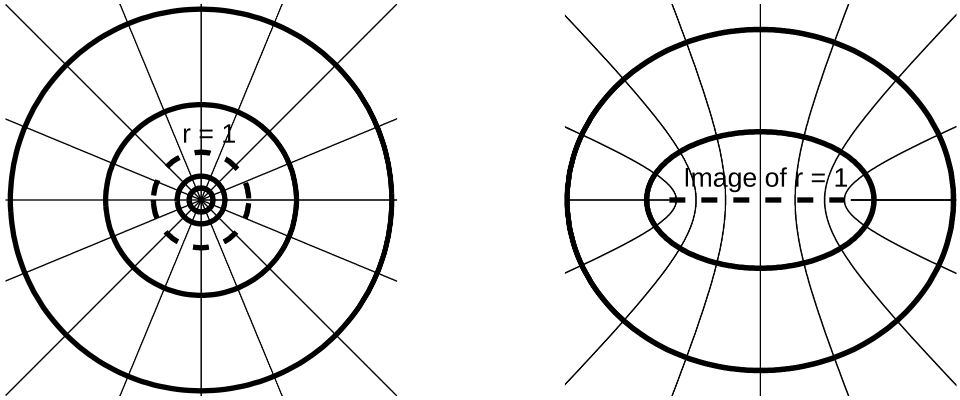

FIG. 3.1. *The transformation* $z \to z + \frac{1}{z}$

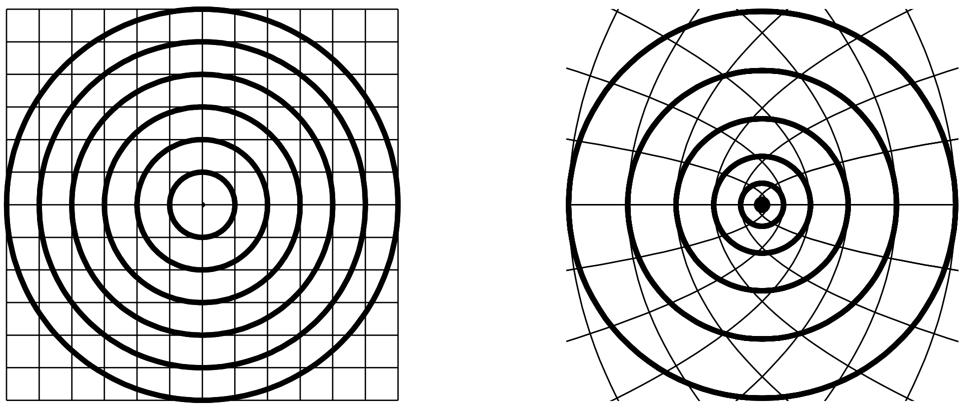

FIG. 3.2. *The transformation* $z \to z^2$ *acting on circles and lines*

The squaring map takes a line parallel to the real axis $z = x_0 + iy$ to a parabola in the complex plane with equation $x = x_0^2 - (y/2x_0)^2$, as shown in Figure 3.2. Any other line in the complex plane can now be obtained from a line parallel to the x-axis by a rotation. Since $[e^{i\theta}(x_0 + iy)]^2 = e^{i2\theta}[(x_0 + iy)]^2$ this completes the argument. $\square$

Since any ellipse or hyperbola with foci equidistant from the origin is obtained by dilation and rotation of a Zhukovskii ellipse or hyperbola, $z \to z^2$ maps all such ellipses to ellipses with one focus at the origin and all such hyperbolas to one branch of a hyperbola with one focus at the origin.

Recall that the type of conic section formed by orbits satisfying Hooke's Law is determined by the sign of C. Therefore, C also determines the type of conic section in the dual, Newton's Law. In the case of orbits satisfying Newton's Law, the energy of a unit mass is $E_Z = \frac{1}{2}|Z'|^2 - \frac{\tilde{C}}{|Z|}$. Using the initial condition for the dual equation and the fact that the energy E_Z is constant along any solution, we obtain

$$E_Z = 2\frac{|w'(0)|^2}{|w(0)|^2} - \frac{2|w'(0)|^2 + 2C|w(0)|^2}{|w(0)|^2} = -2C.$$

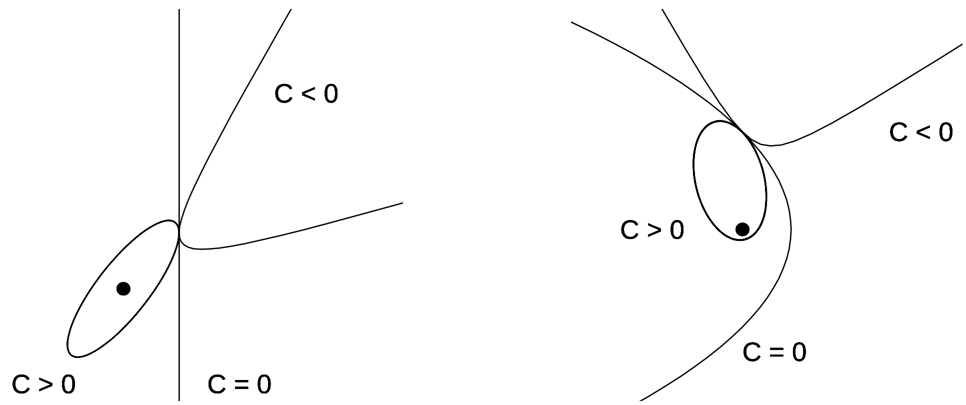

FIG. 3.3. *Trajectories of Hooke's Law and their images under $z \to z^2$*

In conjunction with Lemma 3.2 this observation proves Newton's Law of Ellipses:

THEOREM 3.3. *All trajectories of motion satisfying Newton's Law $Z'' = -\tilde{C}\frac{Z}{|Z|^3}$ are conic sections. In particular they are elliptical when $E_Z < 0$, hyperbolic when $E_Z > 0$ and parabolic when $E_Z = 0$.*

There is a slight subtlety in the case of the hyperbolic orbits. For any $\tilde{C}$, hyperbolic solutions to $Z'' = -\tilde{C}\frac{Z}{|Z|^3}$ consist of only one branch of a hyperbola. The choice of branch depends on whether the origin is attractive ($\tilde{C} > 0$) or repulsive ($\tilde{C} < 0$).

See Figure 3.3 for an illustration of the duality for different values of C. This figure was created using the Matlab routines given in the Appendix.

4. Duality of power laws. Hooke's Law and Newton's Law are only one example of a pair of dual laws. The following result is the general case:

THEOREM 4.1. *Trajectories of points in the complex plane under the centripetal attraction $w'' = -C\dfrac{w}{|w|^{1-a}}$ are mapped to the trajectories of $Z'' = -\tilde{C}\dfrac{Z}{|Z|^{1-A}}$ under the transformation $Z = w^\alpha$, where a, A and α satisfy*

$$(a+3)(A+3) = 4; \qquad \alpha = \frac{a+3}{2}.$$

Proof. The law of areas holds in any central field, and so we reparametrize time as

$$\frac{d\tau}{dt} = \frac{|Z(\tau(t))|^2}{|w(t)|^2} = |w(t)|^{2(\alpha-1)}.$$

The remaining calculations follow those of Theorem 2.1. $\square$

The duality of Newton's Law and Hooke's Law now becomes a special case of this observation, where $a = 1$, $A = -2$, and $\alpha = 2$.

5. The inverse fifth power law. From Theorem 4.1, the self-dual laws (that is, the laws of motion that are transformed into themselves under $z \to z^\alpha$) are those for which $a = -1$ or $a = -5$. If $a = -1$, then $\alpha = 1$ and we cannot get any information from the duality. However, in the case $a = -5$ we have $\alpha = -1$ and the trajectories of motion for the inverse fifth power law are transformed to trajectories for the same law by the inversion $z \to \frac{1}{z}$.

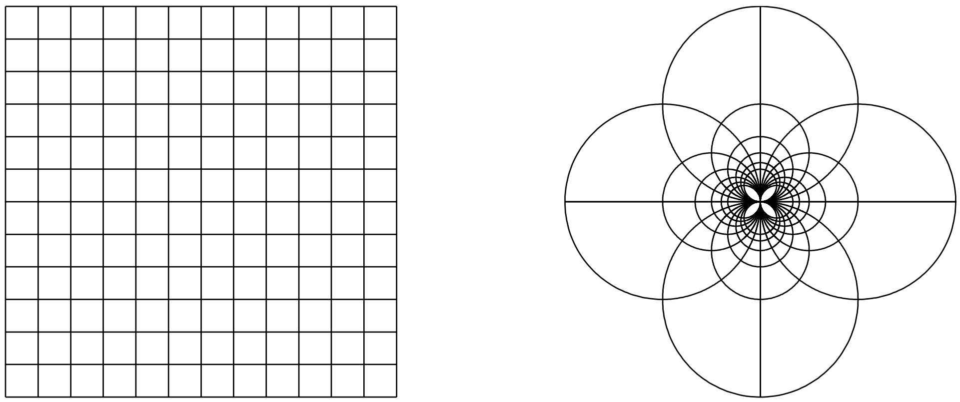

FIG. 5.1. *The transformation $z \to \frac{1}{z}$ acting on lines*

If we set $C = 0$, all trajectories of the inverse fifth power law $w'' = -C\frac{w}{|w|^6}$ are straight lines. Their images under the transformation $Z = w^{-1}$ are also trajectories for the same law, since it is self-dual, although the values of $\tilde{C}$ will vary depending on initial conditions. Figure 5.1 shows how inversion acts on horizontal and vertical lines in the complex plane.

COROLLARY 5.1. *In a field whose central attraction decreases as the fifth power of the distance, there exist circular trajectories that pass through the origin.*

Since we assumed that the sun is at the origin of the coordinate system, planets on such orbits would be on a collision course with the sun. This is an instance of a collision singularity, a phenomenon that has been much studied in celestial mechanics [5].

6. Applications. All solutions of the system $w'' = -Cw$ for $C > 0$ are ellipses, and if we change the initial values of a solution by a little, we obtain a different solution which remains close for all time to the original one. Such systems are called *stable*. An approximate solution to a stable system (found on a computer) will reflect the properties of the real solution. On the other hand, small differences in the initial conditions of *unstable* systems will typically lead to a divergence of orbits in time. Newton's Law describes an unstable system. Therefore the transformation described in Theorem 2.1 takes a stable system to an unstable system, allowing us to find solutions to the unstable system numerically.

Since we do not live in a two-dimensional universe, this approach is of limited practical value. However, if Newton's Law and Hooke's Law were dual in three dimensions, we could find solutions to problems which involve perturbation of the the three-dimensional two-body problem (for example, a satellite orbiting earth) computationally. It appears that the great Italian mathematician Levi-Civita unsuccessfully tried to find such an extension. The duality was shown by Kustaanheimo using *spinors* which are a natural generalization of complex variables [7].

Appendix. Illustrating the duality with Matlab.

The following Matlab programs will generate pictures of dual trajectories. To use it, download the files from *http://www.math.psu.edu/hall/newton/* or copy the program given here, save the first file as `duality.m`, and save the second as `orbit.m` in some directory. Open Matlab in that directory, and type `duality` at the `>>` prompt. This will give you a picture for the default values, which illustrate the solution to $w'' = -w$ with initial conditions $w_0 = 1+i,\ w_0' = -i$ for $0 \le t \le 50$ and the image of that trajectory under the map $z \to z^2$. To change the constants type `duality(`a`,`C`,`t`,`Y_0`)`,

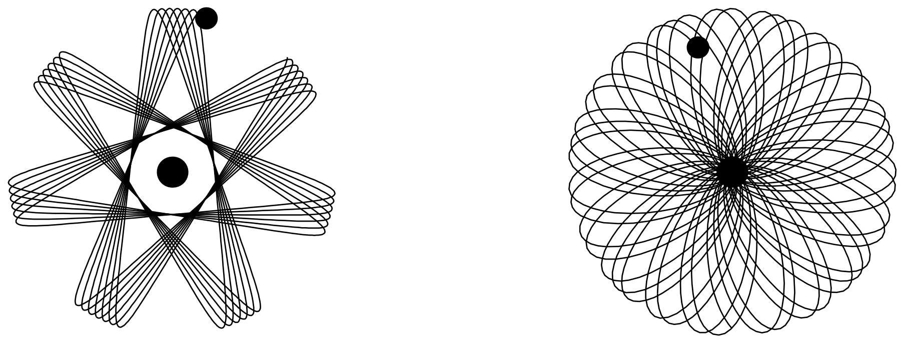

Fig. A.1. *A trajectory for the ninth-power law of attraction and its dual*

where a and C are as given in Theorem 4.1, t is the maximum time, and Y_0 is the vector [x_0 x'_0 y_0 y'_0]. For instance,
`duality(1, -1, 5, [1 0 -1 -1])`
gives the duality for Hooke's Law ($a = 1$) with a different value for C and set of initial conditions. You need not enter all four arguments; `duality(-3, -1)` will just give the inverse cube law with $C = -1$. Some problems you will notice in the program are due to the exponential function. Try `duality(-2)` and see what you get!

You will notice that some laws of attraction have strange, spirograph-type trajectories. For instance, Figure A.1 shows a trajectory for an ninth-power law with $C = 1$. For more information on laws that have only closed orbits and a more or less explicit solution to the laws of motion in a central field, see [2, pp. 33-42].

You can read more about this program at *http://www.math.psu.edu/hall/newton/* along with several articles about Newton's Laws and other topics from the vantage point of first-year Calculus.

Save this file as `duality.m`:

```
function duality(a, C, t, Y0)          % Y0 = [x0 x'0 y0 y'0]
global dualityfig

if nargin < 4, Y0 = [1 0 1 -1]; end
if nargin < 3, t = 50; end             % these are the
if nargin < 2, C = 1; end              % default values
if nargin < 1, a = 1; end

dualityfig = figure( ...
'Name','Dual Trajectories',...
'Userdata',[a C]);

options = odeset('RelTol',1e-4,'AbsTol',[1e-4 1e-4 1e-4 1e-4]);
       [T,Y] = ode45('orbit',[0 t],Y0,options);

X = Y(:,1);
Y = Y(:,3);
```

```
Z = X + i*Y;
x = real(Z.^((a+3)/2));
y = imag(Z.^((a+3)/2));

subplot(1,2,1)
plot(X,Y)
line('Marker', '.', 'Markersize', 35, 'xdata', 0, 'ydata', 0);
axis('square')

subplot(1,2,2)
plot(x,y)
line('Marker', '.', 'Markersize', 35, 'xdata', 0, 'ydata', 0);
axis('square')
```

Save this file as `orbit.m`:

```
function dy = orbit(t,y)
dualityfig = gcf;
A = get(dualityfig,'UserData');
a = A(1);  C = A(2);
dy = zeros(4,1);
dy(1) = y(2);                                        % y(1) = x
dy(2) = -C*y(1)/(y(1)^2 + y(3)^2)^((1-a)/2);    % y(2) = x'
dy(3) = y(4);                                        % y(3) = y
dy(4) = -C*y(3)/(y(1)^2 + y(3)^2)^((1-a)/2);    % y(4) = y'
```

REFERENCES

[1] L. V. Ahlfors, *Complex Analysis*, McGraw-Hill Book Company, New York, 1979.

[2] V. Arnold, *Mathematical Methods of Classical Mechanics*, Springer Verlag, New York, 1989.

[3] ———, *Huygens and Barrow, Newton and Hooke*, Birkhauser, Boston, 1990.

[4] S. Chandrasekhar, *Newton's* Principia *for the Common Reader*, Clarendon Press, Oxford, 1995.

[5] F. Diacu and P. J. Holmes, *Celestial Encounters: the Origin of Chaos and Stability*, Princeton University Press, Princeton, 1996.

[6] J. Milnor, *On the geometry of the Kepler problem*, American Mathematical Monthly, 90 (1983), pp. 353–365.

[7] D. G. Saari, *A visit to the Newtonian n-body problem via elementary complex variables*, American Mathematical Monthly, 97 (1990), pp. 105–119.