Linear algebra and the geometry of quadratic equations

Similarity transformations and orthogonal matrices

First, we need to recall some things from linear algebra. Two square matrices A and B are *similar* if there is an invertible matrix S such that $A = S^{-1}BS$. This is equivalent to $B = SAS^{-1}$. The expression SAS^{-1} is called a *similarity transformation* of the matrix A. A square matrix A is *diagonalizable* if it is similar to a diagonal matrix D. That is, A is diagonalizable if there is a diagonal matrix D and an invertible matrix S such that $D = SAS^{-1}$.

Similarity transformations can be thought of in terms of a change of basis (see Theorems CB, ICBM, and SCB of *A First Course in Linear Algebra*). Here, we'll limit our attention to the vector spaces $\mathbb{R}^n$. If $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a basis for $\mathbb{R}^n$, then $\{S\vec{v}_1, S\vec{v}_2, \ldots, S\vec{v}_n\}$ is another basis for $\mathbb{R}^n$. If $\vec{v}$ is in $\mathbb{R}^n$ and A is an $(n \times n)$ matrix, then we can rewrite the product $A\vec{v}$ as

$$A\vec{v} = (S^{-1}BS)\vec{v} = S^{-1}B(S\vec{v}).$$

We can read the last expression in the following way: Start with the vector $\vec{v}$, multiply by S to change bases, then multiply by B, and finally multiply by S^{-1} to return to the original basis. All of this is equivalent to multiplying by A in the original basis. So what's the point? If we choose S carefully, then multiplication by B is easier than multiplication by A. In particular, if A is diagonalizable, we get to multiply by a diagonal matrix which is particularly easy. This will be very useful in our application to quadratic equations below.

A similarity transformation is particularly nice if the matrix S is orthogonal. By definition, a (real) square matrix S is *orthogonal* if $S^T S = I$ where S^T is the transpose of S and I is the identity matrix (of the appropriate size). An orthogonal matrix is invertible and $S^{-1} = S^T$. It is a theorem (for example, see Theorem COMOS of FCLA) that a matrix $S = [\vec{S}_1, \vec{S}_2, \ldots, \vec{S}_n]$ is orthogonal if and only if the set of column vectors $\{\vec{S}_1, \vec{S}_2, \ldots, \vec{S}_n\}$ is an orthonormal set. Expressed in terms of inner product, this means $\langle \vec{S}_i, \vec{S}_j \rangle = 0$ for $i \neq j$ and $\langle \vec{S}_i, \vec{S}_i \rangle = 1$.

For the vector spaces $\mathbb{R}^2$ and $\mathbb{R}^3$, we can think of vectors both algebraically and geometrically as in multivariate calculus. For the most part, we'll focus on $\mathbb{R}^2$ so we can more easily draw pictures. The geometric interpretation of a vector is as a directed line segment, that is, an arrow. For our purposes here, it will be enough to focus on arrows with tail based at the origin of a chosen coordinate system. If the head of the arrow is at the point $P(x, y)$, we can make a correspondence with the column vector $\begin{bmatrix} x \\ y \end{bmatrix}$. Starting with the column vector $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ we can make a correspondence with the arrow having tail at the origin and head at the point $P(v_1, v_2)$. When thinking geometrically, we will denote the standard basis vectors by $\hat{i}$ and $\hat{j}$ (with $\hat{k}$ included when working in $\mathbb{R}^3$). We can then write

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = v_1 \hat{i} + v_2 \hat{j}.$$

So, when we are given a vector $\vec{v}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$, we can think of it as either an arrow vector or a column vector as suits our needs. Note a subtle point here: The correspondence we make depends on having picked a coordinate system for the geometric plane and a basis for the space of column vectors.

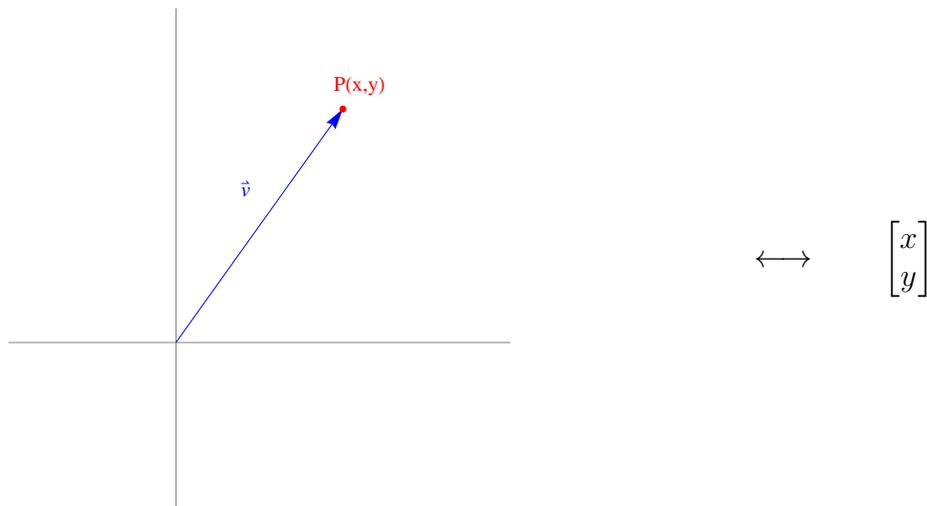$$\longleftrightarrow \qquad \begin{bmatrix} x \\ y \end{bmatrix}$$

Figure 1: Correspondence between arrows and column vectors

In the geometric view, we can think of the inner product of linear algebra as the dot product from multivariate calculus. That is, with

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = u_1\hat{\imath} + u_2\hat{\jmath} + u_3\hat{k}, \quad \text{and} \quad \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = v_1\hat{\imath} + v_2\hat{\jmath} + v_3\hat{k}$$

we have

$$\langle S\vec{u}, S\vec{v} \rangle = u_1 v_1 + u_2 v_2 + u_3 v_3 = \vec{u} \cdot \vec{v}.$$

Recall that $\|v\|^2 = \vec{v} \cdot \vec{v}$ gives the square of the *length* of the (arrow) vector $\vec{v}$. Angles enter through the result connecting algebra and geometry of dot products, namely

$$\vec{u} \cdot \vec{v} = \|\vec{u}\|\,\|\vec{v}\|\cos\theta$$

where θ is the angle between the (arrow) vectors $\vec{u}$ and $\vec{v}$. A pair of vectors that are orthogonal as column vectors ($\langle \vec{u}, \vec{v} \rangle = 0$) are perpendicular as arrow vectors ($\vec{u} \cdot \vec{v} = 0$).

Let's look at the geometry of multiplying a vector by an orthogonal matrix. We know that multiplication by an orthogonal matrix preserves the inner product (Theorem OMPIP) and hence the norm. That is, if S is orthogonal, then

$$\langle S\vec{u}, S\vec{v} \rangle = \langle \vec{u}, \vec{v} \rangle \tag{1}$$

for any pair of vectors $\vec{u}$ and $\vec{v}$ so

$$\|S\vec{w}\| = \|\vec{w}\|. \tag{2}$$

for any vector $\vec{w}$. To connect with a geometric interpretation in $\mathbb{R}^2$ and $\mathbb{R}^3$, we will think of the inner product as the dot product (from multivariate calculus). Geometrically, $\vec{w}$ and $S\vec{w}$ have the same length for any vector $\vec{w}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$. When we are

emphasizing geometric interpretations, we might write Display (1) in terms of the dot product as

$$(S\vec{u}) \cdot (S\vec{v}) = \vec{u} \cdot \vec{v} \tag{3}$$

Using the geometric interpretation of dot product, we have

$$\|S\vec{u}\| \, \|S\vec{v}\| \cos\phi = \|\vec{u}\| \, \|\vec{v}\| \cos\theta$$

where ϕ is the angle between $S\vec{u}$ and $S\vec{v}$ while θ is the angle between $\vec{u}$ and $\vec{v}$. Using the facts that $\|S\vec{u}\| = \|\vec{u}\|$ and $\|S\vec{v}\| = \|\vec{v}\|$, we see that $\cos\phi = \cos\theta$ so $\phi = \theta$. So, multiplication by an orthogonal matrix preserves lengths of vectors and angles between pairs of vectors. Note that the same holds true for the geometric transformation of *rotation through a specified angle*.

Exercise 1. Show that a (2×2) matrix S is orthogonal if and only if there is an angle θ such that

$$S = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{4}$$

or

$$S = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}. \tag{5}$$

Exercise 2. Convince yourself that multiplication by a matrix of the form Display (4) rotates a vector through the angle θ. Start by looking at the standard basis vectors $\hat{\imath}$ and $\hat{\jmath}$. You might also want to choose some specific angles θ with which to experiment. Try $\theta = 0$, $\theta = \pi/4$, $\theta = \pi/2$, and $\theta = \pi$.

Exercise 3. Determine the geometric interpretation of multiplication by a matrix of the form Display (5). Start by looking at the standard basis vectors $\hat{\imath}$ and $\hat{\jmath}$.

Quadratic equations and curves

Somewhere along the line, you learned that an ellipse can be described by an equation of the form is

$$\frac{x^2}{r_1^2} + \frac{y^2}{r_2^2} = 1. \tag{6}$$

If $r_1 > r_2$, then the major axis and minor axis of the ellipse lie along the x-axis and y-axis, respectively. The *semi-major axis length* is given by r_1 and the *semi-minor axis length* is given by r_2. You should remind yourself of the connection between the geometric definition of an ellipse and the analytic description given in Display (6).

Exercise 4. Here is a geometric definition of ellipse: Pick two points F_1 and F_2 and a distance d greater than $|F_1 F_2|$. An ellipse with foci F_1 and F_2 is the set of all points P such that $|PF_1| + |PF_2| = d$. Show the connection between this geometric definition and the analytic description given in Display (6). To do this, choose a coordinate system with origin at the midpoint of the segment $F_1 F_2$ and x-axis containing F_1 and F_2. Show P is on the ellipse if and only if the coordinates (x, y) of P satisfy Display (6).

You should also know that every quadratic equation in two variables corresponds to an ellipse, a hyperbola, or a parabola. Rather than look at all quadratic equations in two variables, we'll limit our attention to quadratic equations of the form

$$Ax^2 + 2Bxy + Cy^2 = 1. \tag{7}$$

(The factor of 2 in the cross term is for convenience.) Given an equation of this form, we want to know whether the equation corresponds to an ellipse, a hyperbola, or a parabola. We'll make nice use of some linear algebra to do this.

Start by defining

$$\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} A & B \\ B & C \end{bmatrix}.$$

You should check that

$$\vec{x}^T Q \vec{x} = Ax^2 + 2Bxy + Cy^2.$$

Note that Q is a real symmetric matrix so the eigenvalues of Q are real and the corresponding eigenvectors are orthogonal. (See Theorem HMRE and Theorem HMOE of FCLA. Note that a real symmetric matrix is Hermitian.) . Let α_1 and α_2 denote the eigenvalues of Q and let $\vec{u}_1$ and $\vec{u}_2$ denote corresponding eigenvectors. Choose $\vec{u}_1$ and $\vec{u}_2$ to be unit vectors.

Define the matrix $S = [\vec{u}_1, \vec{u}_2]$ and note that S is an orthogonal matrix since $\{\vec{u}_1, \vec{u}_2\}$ is an orthonormal set (Theorem COMOS of FCLA). We can also choose $\vec{u}_2$ so that S has the form given in Display (4). We can use S to diagonalize the matrix Q by a similarity transformation. That is,

$$SQS^T = D = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}. \tag{8}$$

Rewrite the similarity transformation in Display (8) to get $Q = S^T DS$. Using this, we have

$$\vec{x}^T Q \vec{x} = \vec{x}^T S^T DS \vec{x} = (S\vec{x})^T D(S\vec{x}) = \vec{X}^T D\vec{X}$$

where we have defined $\vec{X} = S\vec{x}$ to get the last expression. Introducing components of $\vec{X}$ as

$$\vec{X} = \begin{bmatrix} X \\ Y \end{bmatrix},$$

we can write

$$\vec{X}^T D\vec{X} = \begin{bmatrix} X & Y \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \alpha_1 X^2 + \alpha_2 Y^2.$$

Thus, our original equation

$$Ax^2 + 2Bxy + Cy^2 = 1$$

is equivalent to the new equation

$$\alpha_1 X^2 + \alpha_2 Y^2 = 1. \tag{9}$$

To see the utility of rewriting our original equation in this new form, let's look at the *geometric* relation between $\vec{x}$ and $\vec{X} = S\vec{x}$. Recall that S is orthogonal and of
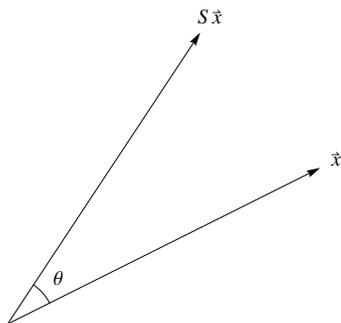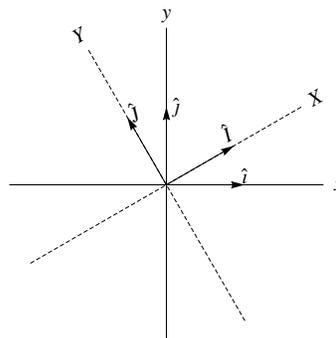
Figure 2: Multiplication by an orthogonal matrix

Figure 3: Rotation of coordinate axes

the form given in Display (4). The geometric effect of multiplication by S is rotation through the angle θ. That is, if we consider $\vec{x}$ as a geometric vector (i.e., a directed line segment of specific length and direction), then $S\vec{x}$ is the geometric vector obtained by rotating $\vec{x}$ as shown in Figure 2.

If we think of $\vec{x}$ as a position vector for a point with coordinates (x, y) in our original coordinate system, then $\vec{X}$ is a position vector for the point with coordinates (X, Y) with respect to a new coordinate system, one that has the same origin but with coordinate axes rotated by the angle θ with respect to the $x-$ and $y-$axes as shown in Figure 3. The new coordinate basis vectors $\hat{I}$ and $\hat{J}$ are rotations of $\hat{i}$ and $\hat{j}$. Note that since we built the roation (i.e., orthogonal) matrix S using eigenvectors $\vec{u}_1$ and $\vec{u}_2$ of the symmetric matrix Q, we have that $\hat{I} = S\hat{i}$ is the first column of S so $\hat{I} = \vec{u}_1$ and $\hat{J} = S\hat{j}$ is the second column of S so $\hat{J} = \vec{u}_2$. In other words, the X-axis and Y-axis lie along the eigenspaces of the symmetric matrix Q.

The equations in Display (7) and Display (9) describe the same curve in different coordinate systems. We can easily read off geometric information from Display (9). Recall that α_1 and α_2 are the eigenvalues of the matrix

$$Q = \begin{bmatrix} A & B \\ B & C \end{bmatrix}.$$

The characteristic polynomial of Q is

$$\lambda^2 - (A + C)\lambda + (AC - B^2) = (\lambda - \alpha_1)(\lambda - \alpha_2) = \lambda^2 - (\alpha_1 + \alpha_2)\lambda + \alpha_1\alpha_2.$$

By comparing the first and last expressions, we see that $AC - B^2 = \alpha_1\alpha_2$. (Note that $AC - B^2 = \det Q$ so this says that the determinant of Q is equal to the product of the eigenvalues. This statement is true for any square matrix.) We now consider cases.

1. $\det Q = AC - B^2 > 0$: In this case, the eigenvalues α_1 and α_2 are both nonzero and have the same sign. If both are negative, then Display (9) has no solutions. If both are positive, then Display (9) is the equation of an ellipse with major and minor axes along the X-axis and Y-axis.

2. $\det Q = AC - B^2 < 0$: In this case, the eigenvalues α_1 and α_2 are both nonzero and have opposite signs. Display (9) is the equation of a hyperbola with symmetry axes along the X-axis and Y-axis.

3. $\det Q = AC - B^2 = 0$: In this case, either $\alpha_1 = 0$ or $\alpha_2 = 0$. For $\alpha_1 = 0$, Display (9) reduces to $\alpha_2 Y^2 = 1$ so $Y = \pm 1/\sqrt{|\alpha_2|}$. You can think about this pair of lines as a "degenerate" hyperbola.

Example 1

Let's determine the geometry of the curve given by

$$73x^2 - 72xy + 52y^2 = 100.$$

We first divide both sides by 100 to get the standard form

$$\tfrac{73}{100}x^2 - \tfrac{72}{100}xy + \tfrac{52}{100}y^2 = 1$$

of Display (7). From this, we read off

$$Q = \begin{bmatrix} \frac{73}{100} & -\frac{36}{100} \\ -\frac{36}{100} & \frac{52}{100} \end{bmatrix} = \frac{1}{100}\begin{bmatrix} 73 & -36 \\ -36 & 52 \end{bmatrix}.$$

Using technology, we find the eigenvalues and eigenvectors of Q are

$$\alpha_1 = 1 \qquad \text{with} \qquad \vec{u}_1 = \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix}$$

and

$$\alpha_2 = \frac{1}{4} \qquad \text{with} \qquad \vec{u}_2 = \begin{bmatrix} \frac{4}{5} \\ -\frac{3}{5} \end{bmatrix}.$$

Since the eigenvalues are both positive, we have an ellipse. The major axis is in the direction $\vec{u}_2$ and the semi-major axis length is $1/\sqrt{1/4} = 2$. The minor axis is in the direction $\vec{u}_1$ and the semi-minor axis length is $1/\sqrt{1} = 1$. Note that the symmetry axes are perpendicular as we expect. The ellipse is shown in Figure 4.
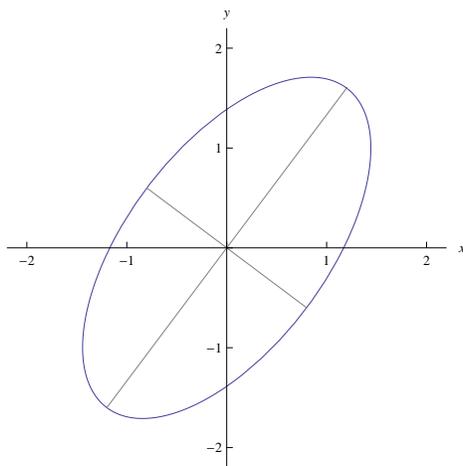

Figure 4: The ellipse of Example 1

In the context of solving linear systems of equations, we often arrive at a solution in the form

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a\cos t + b\sin t \\ c\cos t + d\sin t \end{bmatrix}. \tag{10}$$

We'll now show that this parametrizes an ellipse and we'll see how to determine the geometry of the ellipse using the ideas from above. We'll start by "deparametrizing" to get an equation satisfied by x and y. First, note that we can rewrite Display (10) as

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \cos t \\ \sin t \end{bmatrix}$$

If $\Delta = ad - bc \neq 0$, the matrix on the right side is invertible so we can write

$$\begin{bmatrix} \cos t \\ \sin t \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} dx - by \\ -cx + ay \end{bmatrix}.$$

This give us expressions for $\cos t$ and $\sin t$ in terms of x and y. Substituting into the Pythagorean identity gives us

$$1 = \cos^2 t + \sin^2 t = \frac{1}{\Delta^2}(dx - by)^2 + \frac{1}{\Delta^2}(-cx + ay)^2.$$

With some algebra, we can rewrite the right side to get

$$\frac{c^2 + d^2}{\Delta^2}x^2 - 2\frac{ac + bd}{\Delta^2}xy + \frac{a^2 + b^2}{\Delta^2}y^2 = 1.$$

This equation is in the form of Display (7) with

$$A = \frac{c^2 + d^2}{\Delta^2}, \qquad B = -\frac{ac + bd}{\Delta^2}, \qquad \text{and} \qquad C = \frac{a^2 + b^2}{\Delta^2}.$$

The geometry of the ellipse is determined by the eigenvalues and eigenvectors of

$$Q = \frac{1}{\Delta^2}\begin{bmatrix} c^2 + d^2 & -(ac + bd) \\ -(ac + bd) & a^2 + b^2 \end{bmatrix}. \tag{11}$$

Exercise 5. Compute $\det Q$ for the matrix in Display (11). Show that $\det Q > 0$ for $ad - bc \neq 0$. This confirms that Display (10) does indeed parametrize an ellipse.

Example 2

We will determine the geometry of the ellipse given parametrically by

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} \cos t - \sin t \\ -2\cos t \end{bmatrix}. \tag{12}$$

First, we can read off the values $a = 1$, $b = -1$, $c = -1$, and $d = 0$ by comparison with the general form given in Display (10). From these, we form the matrix

$$Q = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Using technology, we find the eigenvalues and eigenvectors of Q are

$$\alpha_1 = \frac{3+\sqrt{5}}{4} \qquad \text{with} \qquad \vec{u}_1 = \frac{1}{\sqrt{10+2\sqrt{5}}}\begin{bmatrix} 1+\sqrt{5} \\ 2 \end{bmatrix}$$

and

$$\alpha_2 = \frac{3-\sqrt{5}}{4} \qquad \text{with} \qquad \vec{u}_2 = \frac{1}{\sqrt{10-2\sqrt{5}}}\begin{bmatrix} 1-\sqrt{5} \\ 2 \end{bmatrix}$$

Note that we have chosen unit eigenvectors so that $S = [\vec{u}_1, \vec{u}_2]$ is orthogonal as required. Since the eigenvalues are both positive, we have an ellipse. The major axis is in the direction $\vec{u}_2$ and the semi-major axis length is $1/\sqrt{\alpha_1} = 2/\sqrt{3-\sqrt{5}}$. The minor axis is in the direction $\vec{u}_1$ and the semi-minor axis length is $1/\sqrt{\alpha_2} = 2/\sqrt{3+\sqrt{5}}$. The ellipse is shown in Figure 5.
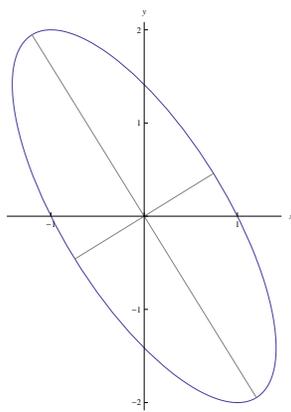

Figure 5: The ellipse of Example 2

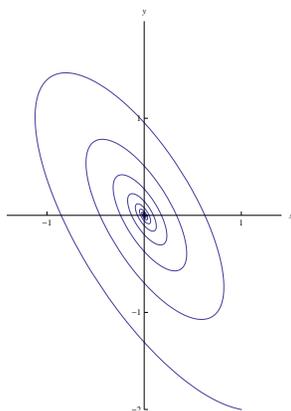If the parametrization in Display (12) included an decaying exponential factor, we would get an elliptic spiral as shown in Figure 6.

Figure 6: An elliptic spiral based on Example 2