Week 9

3.3 Polar coordinates

The algebraic expressions from the preceding section suggest a connection between the three types of curves. The first observation is that all the curves are described by quadratic expression, but looking closely we see that if we were working with complex numbers then (after substituting b with ib) ellipses and hyperbolas have the same expression.

In this section we will look at expressions for these curves in terms of focal polar coordinates, that is polar coordinates where the origin is a focus for the curve. This will highlight a deeper relationship between the curves and introduce the notion of eccentricity.

3.3.1 Ellipse or hyperbola

We begin with a treatment of the ellipse and hyperbola together. Let the origin $\mathbf{O} = \mathbf{F}_1$, be the first focus and let $\theta = 0$ indicate a direction towards $\mathbf{F}_2$. Let $\mathbf{P} = (r, \theta)$ be a point on the curve C, which is an ellipse or a hyperbola. For each θ there are two choices of r that give a point on the curve C. In the case of the ellipse we may always assume that _________, see Figure 12. In the case of the hyperbola things are slightly more complicated, here we say either: $\mathbf{P}$ is closer to $\mathbf{F}_2$ and _________; or instead $\mathbf{P}$ is closer to $\mathbf{F}_1$ and _________, see Figure 13. Therefore we have two cases for the hyperbola. Note that making the opposite choice would lead to a different, but perfectly good equation for the curve C.

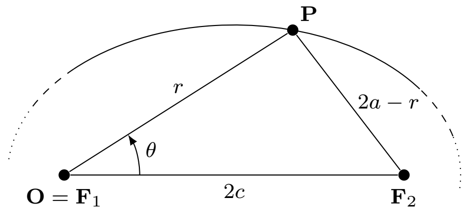


Figure 12: The radial distance r and angle θ for a point on an ellipse.

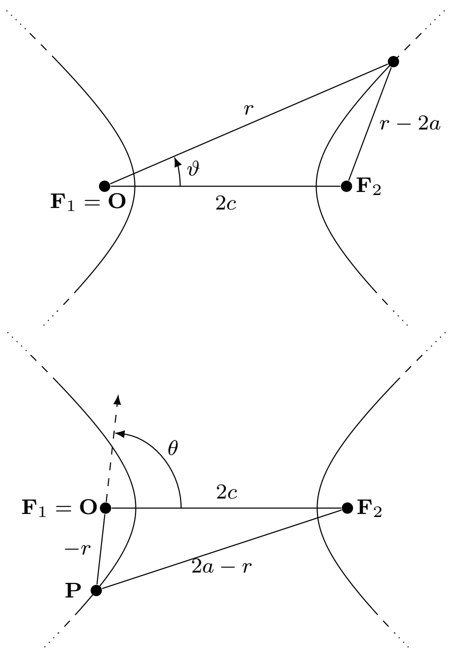


Figure 13: The radial distance r and angle θ for points on a hyperbola. In the left-hand diagram $r > 0$, whilst on the right $r < 0$.

Wikipedia: cosine rule

We can now apply the cosine rule to the triangle formed by the points $\mathbf{F}_1$, $\mathbf{F}_2$ and $\mathbf{P}$ to deduce that

$$(2a-r)^2 = r^2 + 4c^2 - 4cr\cos\theta \qquad \text{(ellipse or hyperbola case 1)}$$
$$(2a-r)^2 = r^2 + 4c^2 + 4cr\cos(\pi-\theta) \quad \text{(hyperbola case 2).}$$

But since $\cos(\pi - \theta) = -\cos\theta$ these are the same and we have

Letting $p = \frac{a^2-c^2}{a}$ and $e = \frac{c}{a}$ we say that the equation of the ellipse in polar coordinates is

$$r = \frac{p}{1-e\cos\theta}. \tag{3.7}$$

The parameter e is called the *eccentricity* of the curve and notice that for an ellipse $0 \le e < 1$, whilst for a hyperbola $e > 1$. In particular, if $e = 0$ then $c = 0$ and C is a circle.

Wikipedia: eccentricity

3.3.2 Parabola

The derivation of polar coordinates for a parabola is slightly simpler than for the other types of curve. In this situation we let the origin be the focus $\mathbf{O} = \mathbf{F}$, and say that the $\theta = 0$ direction is away from the directrix. As in the algebraic expression we say that the distance from focus $\mathbf{F}$ to directrix l is p.

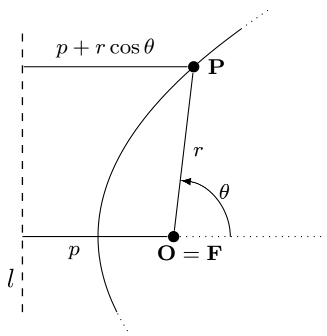


Figure 14: The radial distance r and angle θ for a point on an parabola.

Let $\mathbf{P} = (r, \theta)$ be a point on a parabola C. The distance from $\mathbf{P}$ to the directrix is $p + r\cos\theta$. The distance between $\mathbf{P}$ and $\mathbf{F}$ is simply r. Thus we have

$$r = \underline{\hspace{6cm}} \quad \Rightarrow$$

Notice that this matches equation (3.7) with $e = 1$, which is exactly the value of the eccentricity that we were missing.

Proposition 3.9 *Let C be an conic section. There exists a polar coordinate system (r, θ) of $\mathbb{A}^2$ with origin a focus point, and parameters $p, e \in \mathbb{R}$ with $e \geq 0$ such that*

$$C = \left\{ (r, \theta) \in \mathbb{A}^2 \mid r = \frac{p}{1 - e\cos\theta} \right\}. \tag{3.8}$$

Moreover, e is called the eccentricity (see Figure 15) and

$$\begin{aligned} e = 0 \quad &\Rightarrow \quad C \textit{ is } \underline{\hspace{4cm}}; \\ 0 < e < 1 \quad &\Rightarrow \quad C \textit{ is } \underline{\hspace{5cm}}; \\ e = 1 \quad &\Rightarrow \quad C \textit{ is } \underline{\hspace{5cm}}; \\ e > 1 \quad &\Rightarrow \quad C \textit{ is } \underline{\hspace{6cm}}. \end{aligned}$$

If C is a parabola then p is the distance between the focus and directrix. If C is an ellipse (including a circle) or a hyperbola then the distance between focus points is $|2c|$, and the defining distance of the curve is $|2a|$

where

$$a = \frac{p}{1-e^2} \qquad\qquad c = \frac{pe}{1-e^2}. \tag{3.9}$$

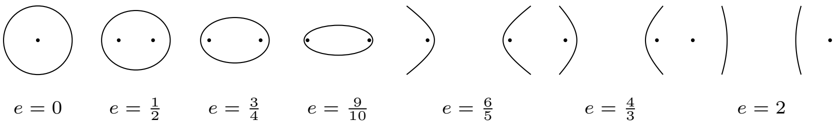


Figure 15: The conic sections shown with fixed distance $2a$, and in order of increasing eccentricity. The parabola (with $e = 1$) is not included as this has "infinite" distance.

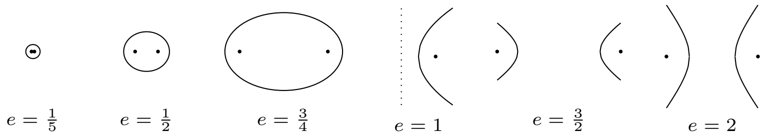


Figure 16: The conic sections shown with fixed focal parameter, p/e, and in order of increasing eccentricity. The circle (with e = 0) is not included as this has "infinite" focal parameter.

As a final remark to this section we address why ellipses, hyperbolas and parabolas are called conic sections. This is because all these curves are exactly given by the section of a cone passing through a plane. The type of curve, is dependent on the gradient of the plane with relative to the gradient of the cone. A plane with a smaller gradient intersects the cone in an ellipse; a plane with an equal gradient intersects the cone in a parabola; and a plane with a steeper gradient intersects the cone in a hyperbola. See Figure 17.

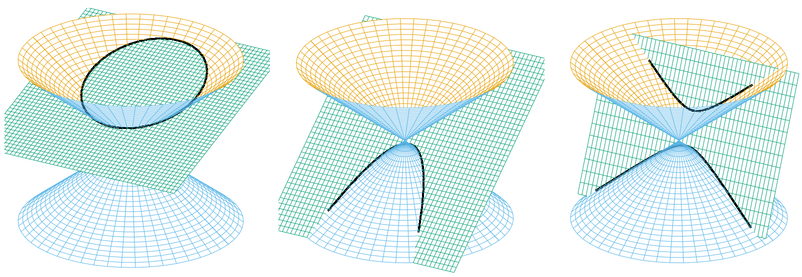


Figure 17: The curves formed by a plane intersecting a cone. On the left is a plane with a shallow gradient, giving an ellipse as its conic section. In the middle the plane has an equal gradient to the cone and intersects in a parabola. On the right the plane has a steeper gradient, producing a hyperbola.

4 Affine transformations

In this section we look at affine transformations; these are maps that act on affine spaces in a similar way to linear operators acting on vector spaces.

Wikipedia: affine transformation

Definition 4.1 (Affine transformation) Let $\mathbb{A}^n$ be an affine space. A function $f\colon \mathbb{A}^n \to \mathbb{A}^n$ is called an *affine transformation* if the following properties hold for all $\mathbf{P}_1, \mathbf{P}_2, \mathbf{Q}_1$ and $\mathbf{Q}_2$ in $\mathbb{A}^n$:

(1) the vectors $\mathbf{P}_2 - \mathbf{P}_1$ and $\mathbf{Q}_2 - \mathbf{Q}_1$ are ________________ if and only if the vectors $f(\mathbf{P}_2) - f(\mathbf{P}_1)$ and $f(\mathbf{Q}_2) - f(\mathbf{Q}_1)$ are ________________;

(2) if $(\mathbf{P}_2 - \mathbf{P}_1) = \lambda(\mathbf{Q}_2 - \mathbf{Q}_1)$______ for some value $\lambda \in \mathbb{R}$ then

$$f(\mathbf{P}_2) - f(\mathbf{P}_1) = \lambda\big(f(\mathbf{Q}_2) - f(\mathbf{Q}_1)\big)\_\_\_\_\_\_.$$

Remark 4.2 Two vectors are linearly dependent if and only if they are scalar multiples of one another, thus $P_2 - P_1$ and $Q_2 - Q_1$ are linearly dependent if and only if the line segment from P_1 to P_2 is parallel to the line segment from Q_1 to Q_2. Using this we can paraphrase the definition as:

(1) line segments are ____________ if and only if their images under f are ____________;

(2) the map f scales ____________ line segments by _______________ amount.

Recall that to each affine space $\mathbb{A}^n$, there is an associated vector space $\mathbb{E}^n$. Similarly, to each affine transformation of $\mathbb{A}^n$ there is an induced linear operator acting on $\mathbb{E}^n$.

Proposition 4.3 *Let $f\colon \mathbb{A}^n \to \mathbb{A}^n$ be an affine transformation and let $\mathbf{O}$ be a chosen origin in $\mathbb{A}^n$. The induced map $F\colon \mathbb{E}^n \to \mathbb{E}^n$ given by*

$$F(\mathbf{v}) = f(\mathbf{O} + \mathbf{v}) - f(\mathbf{O})$$

is an _______________ linear operator of $\mathbb{E}^n$, and is independent of the choice of __________.

Proof. First notice that for any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{E}^n$ and scalars $\lambda, \mu \in \mathbb{R}$ we have

$$(\mathbf{O} + \mu\mathbf{w}) - \mathbf{O} = \mu\mathbf{w} = (\mathbf{O} + \mu\mathbf{w} + \lambda\mathbf{v}) - (\mathbf{O} + \lambda\mathbf{v})$$

and so

$$f(\mathbf{O} + \mu\mathbf{w}) - f(\mathbf{O}) = f(\mathbf{O} + \mu\mathbf{w} + \lambda\mathbf{v}) - f(\mathbf{O} + \lambda\mathbf{v}).$$

Furthermore, by condition (2) in Definition 4.1, for any vector $\mathbf{v} \in \mathbb{E}^n$ and scalar $\lambda \in \mathbb{R}$ we have

$$\begin{aligned} (\mathbf{O} + \lambda\mathbf{v}) - \mathbf{O} &= \lambda\left((\mathbf{O} + \mathbf{v}) - \mathbf{O}\right) \\ \Rightarrow f(\mathbf{O} + \lambda\mathbf{v}) - f(\mathbf{O}) &= \lambda\left(f(\mathbf{O} + \mathbf{v}) - \mathbf{O}\right) \end{aligned}$$

We use this to show that the map is linear

$$\begin{aligned} F(\lambda\mathbf{v} + \mu\mathbf{w}) &= f(\mathbf{O} + (\lambda\mathbf{v} + \mu\mathbf{w})) - f(\mathbf{O}) \\ &= \Big(f(\mathbf{O} + \lambda\mathbf{v}) - f(\mathbf{O})\Big) + \Big(f(\mathbf{O} + \lambda\mathbf{v} + \mu\mathbf{w}) - f(\mathbf{O} + \lambda\mathbf{v})\Big) \\ &= \Big(f(\mathbf{O} + \lambda\mathbf{v}) - f(\mathbf{O})\Big) + \Big(f(\mathbf{O} + \mu\mathbf{w}) - f(\mathbf{O})\Big) \\ &= \lambda\Big(f(\mathbf{O} + \mathbf{v}) - f(\mathbf{O})\Big) + \mu\Big(f(\mathbf{O} + \mathbf{w}) - f(\mathbf{O})\Big) \\ &= \lambda F(\mathbf{v}) + \mu F(\mathbf{w}). \end{aligned}$$

To see that the map is independent of the choice of origin we note that for any $\mathbf{P} \in \mathbb{A}^n$

$$(\mathbf{P} + \mathbf{v}) - \mathbf{P} = \mathbf{v} = (\mathbf{O} + \mathbf{v}) - \mathbf{O}$$

and hence

$$f(\mathbf{P} + \mathbf{v}) - f(\mathbf{P}) = f(\mathbf{O} + \mathbf{v}) - f(\mathbf{O}) = F(\mathbf{v})$$

Finally, to see that the map is invertible we need only show $F(\mathbf{v}) = \mathbf{0}$ implies $\mathbf{v} = \mathbf{0}$, and then appeal to the rank-nullity theorem. This is clear from the Definition 4.1 (2) since

$$\mathbf{v} = (\mathbf{O} + \mathbf{v}) - \mathbf{O} \neq \mathbf{0} \qquad \text{and} \qquad F(\mathbf{v}) = f(\mathbf{O} + \mathbf{v}) - f(\mathbf{O}) = \mathbf{0}$$

is forbidden. □

The only additional information required to recover the affine transformation from its induced linear operator is the image of a single point, usually the origin: $\mathbf{O} \mapsto \mathbf{O} + \mathbf{r}$.

Proposition 4.4 *Let $f \colon \mathbb{A}^n \to \mathbb{A}^n$ be an affine transformation. Let F be the induced linear operator defined in Proposition 4.3. Fix any point $\mathbf{O} \in \mathbb{A}^n$ and let $\mathbf{O}' = f(\mathbf{O})$ be its image under f. The map f can be expressed in terms of F and $\mathbf{O}'$ by*

Proof. This is immediate from the definition of F since

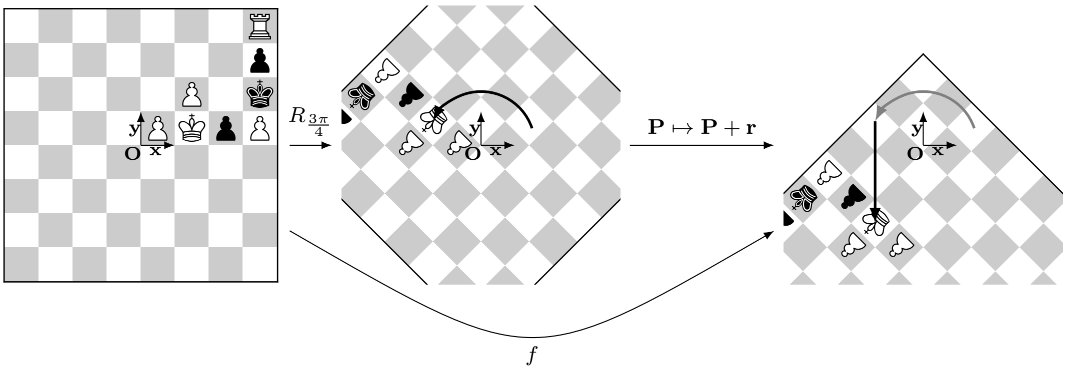
$$F(\mathbf{P} - \mathbf{O}) = f(\mathbf{P}) - f(\mathbf{O}) = f(\mathbf{P}) - \mathbf{O}'. \qquad \square$$

It is now clear that having fixed any origin $\mathbf{O}$ for affine space, an affine transformation can be expressed in terms of a linear operator together with a translation of the origin. Conversely, given any invertible linear operator and a translation we can construct an associated affine transformation. This leads to an alternative definition:

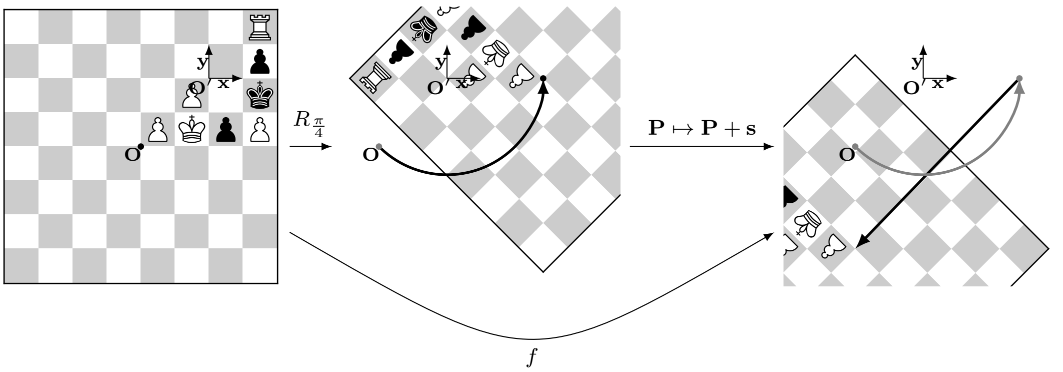
Alternative Definition 4.5 (Affine transformation) A map $f\colon \mathbb{A}^n \to \mathbb{A}^n$ is called an *affine transformation* if there is a point $\mathbf{O} \in \mathbb{A}^n$ (the origin), a translation vector $\mathbf{r}$ and an ________________ linear operator $F\colon \mathbb{E}^n \to \mathbb{E}^n$ such that for any point $\mathbf{P} \in \mathbb{A}^n$

$$f(\mathbf{P}) =$$

Example 4.6 The affine transformation f is made up of a rotation about $\mathbf{O}$ by $\frac{3\pi}{4}$ radians and a translation by $\mathbf{r} = -3\mathbf{y}$. The movement of the white king is indicated in the diagrams.



Example 4.7 If we express the map f from Example 4.6 with respect to a different origin, the linear map $R_{\frac{3\pi}{4}}$ remains the same but we must change the translation. For example, if $\mathbf{O}' = \mathbf{O} + 2\mathbf{x} + 2\mathbf{y}$ is the new origin then we must instead translate by the vector $\mathbf{s} = -2\sqrt{2}\mathbf{x} - 3\mathbf{y}$. The movement of the previous origin $\mathbf{O}$ is indicated in the diagrams.



Proposition 4.8 *An affine transformation $f\colon \mathbb{A}^2 \to \mathbb{A}^2$ is determined by the image of any ________ non-collinear points.*

Proof. Let $\mathbf{O}$, $\mathbf{P}_1$ and $\mathbf{P}_2$ be three points that do not lie on the same line. Let $\mathbf{e}_1 = \mathbf{P}_1 - \mathbf{O}$ and $\mathbf{e}_2 = \mathbf{P}_2 - \mathbf{O}$, so that $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2)$ is a basis for $\mathbb{E}^2$. Any linear operator is determined by the image of a basis, therefore the linear operator F, induced by f is determined by $F(\mathbf{e}_1) = f(\mathbf{P}_1) - f(\mathbf{O})$ and $F(\mathbf{e}_2) = f(\mathbf{P}_2) - f(\mathbf{O})$. Now we can use the fact that f is determined by F and a translation, which we can determine using the image of any of our three points. □

4.1 Matrix of an affine transformation

Just as we have matrix representations of a linear operator, which make calculations convenient, we can define a matrix representation of an affine transformation. Instead of the matrix being defined with respect to just a basis, we also need to pick an origin for the representation. The pair of a basis and an origin is called a *frame* for the affine space.

Definition 4.9 (Frame) Let $\mathbb{A}^n$ be affine space and let $\mathbb{E}^n$ be its associated Euclidean vector space. We call any pair $\mathcal{F} = (\mathcal{B}, \mathbf{O})$, consisting of a basis $\mathcal{B}$ of $\mathbb{E}^n$, together with a point $\mathbf{O} \in \mathbb{A}^n$ a *frame* for $\mathbb{A}^n$.

Wikipedia: frame

Example 4.10 A Cartesian coordinate system of $\mathbb{A}^n$ is a frame in which the basis is orthonormal.

A frame is convenient as it allows us to represent both points and vectors in column vector notation. We do this by adding an additional entry (row) to our vectors that is 0 in the case of vectors and 1 in the case of points.

Definition 4.11 (Vectors and points with respect to a frame) Let $\mathbb{A}^n$ be an affine space and $\mathbb{E}^n$ its associated vector space. Let $\mathcal{F} = \big((\mathbf{e}_1, \ldots, \mathbf{e}_n), \mathbf{O}\big)$ be a frame for $\mathbb{A}^n$. Let $\mathbf{v} = v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n$ be a vector in $\mathbb{E}^n$ and let $\mathbf{P} = \mathbf{O} + \mathbf{v}$ be a point of $\mathbb{A}^n$. Then we denote $\mathbf{v}$ and $\mathbf{O}$ *with respect to the frame* $\mathcal{F}$ by

$$[\mathbf{v}]_{\mathcal{F}} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \\ 0 \end{bmatrix} = \begin{bmatrix} [\mathbf{v}]_{\mathcal{B}} \\ 0 \end{bmatrix} \quad \text{and} \quad (\mathbf{P})_{\mathcal{F}} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \\ 1 \end{pmatrix} = \begin{pmatrix} [\mathbf{v}]_{\mathcal{B}} \\ 1 \end{pmatrix}. \tag{4.1}$$

Note that the choice of bracket is simply an additional aid to help distinguish points from vectors and is not of practical importance.

Definition 4.12 (Matrix of a affine transformation) Let $f\colon \mathbb{A}^n \to \mathbb{A}^n$ be an affine transformation and $\mathcal{F} = (\mathcal{B}, \mathbf{O})$ be a chosen frame for the affine space, with $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$. Let $F \colon \mathbb{E}^n \to \mathbb{E}^n$ be the induced linear operator of f and $F(\mathbf{e}_i) = a_{1,i}\mathbf{e}_1 + \cdots + a_{n,i}\mathbf{e}_n$ for each basis vector $\mathbf{e}_i$ in $\mathcal{B}$. Let $f(\mathbf{O}) = \mathbf{O} + \mathbf{b}$, where $\mathbf{b} = b_1\mathbf{e}_1 + \cdots + b_n\mathbf{e}_n$.

Wikipedia: (augmented) matrix representation of f with respect to $\mathcal{F}$

Then the *(augmented) matrix representation of f with respect to $\mathcal{F}$* is

$$[f]_{\mathcal{F}} = \left[\begin{array}{cccc|c} a_{1,1} & a_{1,2} & \ldots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \ldots & a_{n,n} & b_n \\ \hline 0 & 0 & \ldots & 0 & 1 \end{array}\right] = \begin{bmatrix} [F]_{\mathcal{B}} & [\mathbf{b}]_{\mathcal{B}} \\ 0\cdots 0 & 1 \end{bmatrix}. \tag{4.2}$$

Note that the dividing lines in the matrix are not strictly necessary and are included for illustrative purposes. These lines will be dropped in most situations.

Notice that, just as with the matrix of a linear operators, matrix multiplication takes the place of applying the either of the maps f or F.

Remark 4.13 Multiplication of the matrix representation of f with a vector is equivalent to applying the linear operator F to the vector: let $\mathbf{w} = \sum w_i\mathbf{e}_i \in \mathbb{E}^n$ be a general vector. Then

$$\begin{aligned} [f]_{\mathcal{F}}\,[\mathbf{w}]_{\mathcal{F}} &= \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \ldots & a_{n,n} & b_n \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} [F]_{\mathcal{B}}\,[\mathbf{w}]_{\mathcal{B}} \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} [F(\mathbf{w})]_{\mathcal{B}} \\ 0 \end{bmatrix} \\ &= [F(\mathbf{w})]_{\mathcal{F}}. \end{aligned}$$

Remark 4.14 Multiplication of the matrix representation of f with a point is equivalent to applying the affine transformation to the point:

let $\mathbf{P} = \mathbf{O} + \mathbf{w} = \mathbf{O} + \sum w_i \mathbf{e}_i \in \mathbb{A}^n$ be a general point. Then

$$\begin{aligned}
[f]_{\mathcal{F}} (\mathbf{P})_{\mathcal{F}} &= \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} & b_n \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} [F(\mathbf{w})]_{\mathcal{B}} + [\mathbf{b}]_{\mathcal{B}} \\ 1 \end{pmatrix} \\
&= (f(\mathbf{P}))_{\mathcal{F}}.
\end{aligned}$$

Example 4.15 The affine transformation f, of Example 4.6 was given with respect to a basis $\mathcal{B} = (\mathbf{x}, \mathbf{y})$ and an origin $\mathbf{O}$. Let $\mathcal{F} = (\mathcal{B}, \mathbf{O})$ be the frame determined by these data and recall that f was given by a rotation of $\frac{3\pi}{4}$ about $\mathbf{O}$ followed by a translation by $-3\mathbf{y}$. What is the matrix of f with respect tot he frame $\mathcal{F}$?