APPENDIX A

Projective Spaces and Transformations

Projective geometry is at the heart of computer graphics whichever view you take of it, practical or theoretical. The various transformations of real 3-space we learned to use for the purpose of animation in Chapter 4 and studied mathematically in Chapter 5 are, in fact, most naturally viewed as transformations of projective 3-space, following a so-called lifting of the scene from real to projective space. A consequence is that representing these transformations as projective is more efficient from a computational point of view, a fact that OpenGL takes constant advantage of in its design. Capturing the scene after a perspective projection on film – "shooting" as we imagine the OpenGL point camera to do – involves a projective transformation as well.

In fact, it's not an exaggeration to say that projective geometry is the mathematical foundation of modern-day CG, and that API's such as OpenGL "live" in projective 3-space. Unfortunately, though, because projective geometry works its magic deep inside the graphics pipeline, its importance often is not realized.

There are several books out there which discuss projective geometry – Coxeter [30], Henle [73], Jennings [78], Pedoe [111] and Samuel [124] come to mind – from mainly a geometer's point of view, as well as a few, such as Baer [5] and Kadison & Kromann [79], which take an algebraic standpoint. All these books, however, seem written primarily for a student of mathematics. There seems none yet dedicated to answering the computer scientist's (almost certainly a CG person) question of projective geometry, "What can you do for me?"

This appendix is a small attempt to fill this gap in the literature and introduce projective spaces and transformations from a CG point of view.

Projective spaces generalize real space. They are not difficult to understand, but geometric primitives, such as lines and planes, behave somewhat differently in a projective space than a real one. By applying a camera-view analogy from the outset, we try to convey a physical-based intuition for basic concepts, establishing at the same time connection with CG.

This appendix is long and the mathematics often admittedly abstract, but the payback for persevering through it comes in the form of a wealth of applications, including the projection transformation in the graphics pipeline, as well as the rational Bézier and all-important NURBS primitives, which are all topics of Chapter 20 on applications of projective spaces.

Logically, this appendix could as well have been a chapter of the book, just prior to Chapter 20. However, we decided against upsetting the fairly easy gradient of the book from the first chapter to the last with the insertion of a mathematical "hill".

In fact, Chapter 20 on applications has been written so that the reader reluctant to take on the venture into projective theory can still make her way through it with minimal loss. This is not in any way to diminish the importance of the material in this appendix, but merely recognition of the reality that there are numbers of people out there who would make fine CG professionals, but care little for abstract mathematics.

We begin in Section A.1 by invoking a camera's point of view to motivate the definition of the projective plane. The geometry of this plane, including its surprising point-line duality and coordinatization by means of the homogeneous coordinate system, is the topic of Sections A.2 and A.3. In Section A.4 we study the structure of the projective plane and learn that the real plane can be embedded in the projective, which in turn yields a classification of projective points into regular ones and those at infinity.

A particularly intuitive kind of projective transformation, the so-called snapshot transformation, comes next in Section A.5. Section A.6 covers a few applications of homogeneous polynomial equations, including an algebraic insight into the projective plane's point-line duality, and an algebraic method to compute the outcome of a snapshot transformation. Following a brief discussion of projective spaces of arbitrary dimension in Section A.7, we move on to projective transformations.

Projective transformations are first defined algebraically in Section A.8 and then understood geometrically in A.9. In Section A.10 we relate projective, snapshot and affine transformations, and see that projective transformations are more powerful than either of the other two. The process of determining the projective transformation to accomplish some particular mapping – often beyond the reach of an affine transformation – is the topic of Section A.11.

A.1 Motivation and Definition of the Projective Plane

Consider a viewer taking pictures with a point camera with a film in front of it. Light rays from objects in the scene travel toward the camera and their intersections with the film render the scene. See Figure A.1. Captured on film is the (perspective) projection of the objects. In the case of OpenGL, this is precisely the situation when the user defines a viewing frustum: the point camera is at the apex of the frustum, while the film lies along its front face.

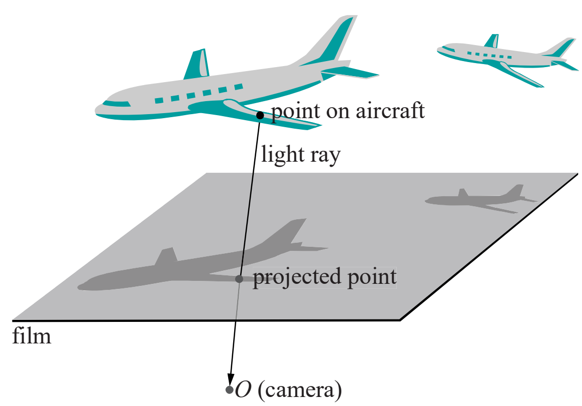

Figure A.1: Perceiving objects with a point camera and a plane film.

Clearly, points in the scene that lie on the same (straight) line through the camera cannot be distinguished by the viewer. In fact, all objects, e.g., points and line segments, lying on one line l through the camera cannot be distinguished by the viewer. They all project to and are perceived as a single point on the film. See Figure A.2(a). Assume for the moment that the film is two-sided and that objects behind project onto it as well (depicted is one such point). For now, ignore as well that lines through the camera parallel to the film, e.g., l', do not intersect the latter at all. This is owing to the alignment of the film, which can always be changed.

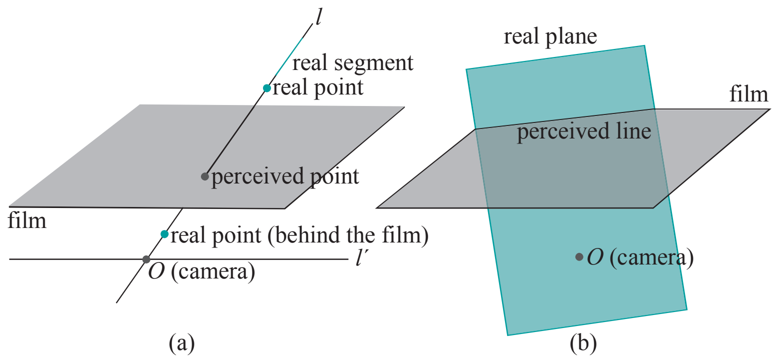

Figure A.2: Perceiving points, lines and planes by projection.

So, one can say that the viewer perceives *every* line through the camera as a point. What then does he perceive as a line? The likely answer is a plane. Indeed, any plane through the camera intersects the film in a line, though, again, the film may have to be re-aligned so as not to be parallel to the plane. See Figure A.2(b).

Lines are points, planes are lines, ... Let's take a moment to formalize, as a new space, the world as it is perceived through a point camera at the origin. Recall that a *radial* primitive is one which passes through the origin.

Definition A.1. A radial line in 3-space $\mathbb{R}^3$ is called a *projective point*. The set of all projective points lying on any one radial plane in $\mathbb{R}^3$ is called a *projective line*. (See Figure A.3.)

The set of all projective points is called 2-dimensional *projective space* and denoted $\mathbb{P}^2$. $\mathbb{P}^2$ is also called the *projective plane*.

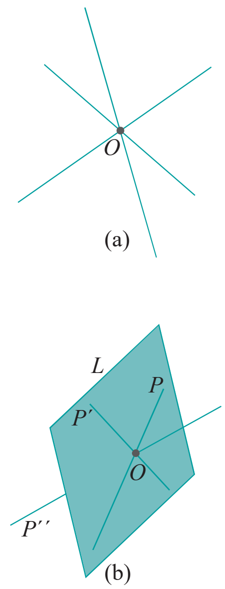

Figure A.3: (a) Projective points are radial lines (b) A projective line consists of all projective points on a radial plane: projective points P and P' belong to the projective line L, while P'' does not.

Remark A.1. We are taking a significant step up in abstraction in leaving $\mathbb{R}^2$ for $\mathbb{P}^2$. The real plane $\mathbb{R}^2$ is easy to visualize as, well, a real plane, e.g., a table top or a sheet of paper. Not so the projective plane. There is no real object to which it corresponds nicely.

Things such as a line, which is a set of points in one space, being just a point of another may seem a bit strange as well. It's mostly a matter of getting used to it though – like learning a foreign language. As with a new language, some words translate literally, but some don't simply because the concept isn't familiar (what's sandstorm in Eskimoan?).

It's recommended that the reader stick close to the real-based definitions at first. A thought process like "Hmm, the projective point P belongs to the projective line L. Well, then, this means that the real line which is P sits inside the real plane which is L" may seem cumbersome at first, but projective primitives will seem less and less strange as we go along.

The term "projective" arose because objects on the projective plane are perceived by projection onto a real one, which for us is the film. Observe that in Figure A.3(b) we denote by L both a radial plane (a primitive in $\mathbb{R}^3$), as well as the projective line (a primitive in $\mathbb{P}^2$) consisting of projective points that lie on that plane. There should be no cause for ambiguity as it'll be clear from the context which we mean.

Terminology: We'll generally use lower case letters to denote primitives in $\mathbb{R}^2$ and upper case for those in $\mathbb{P}^2$.

Remark A.2. The dimension of $\mathbb{P}^2$ is two (as indicated by the superscript). This is because, while points in $\mathbb{R}^3$ have three "degrees of freedom", radial lines in $\mathbb{R}^3$ have only two. We'll elaborate on the dimension of the projective plane in Section A.7.

A.2 Geometry on the Projective Plane and Point-Line Duality

We have, then, on the projective plane $\mathbb{P}^2$ projective points and projective lines, just as on the real plane $\mathbb{R}^2$ we have real points and lines. It's interesting to compare the relationship between points and lines in the two spaces.

Recall the following two facts from Euclidean geometry (geometry in real space is called Euclidean):

(a) There is a unique line containing two distinct points in $\mathbb{R}^2$.

(b) Two distinct lines in $\mathbb{R}^2$ intersect in a unique point, *except* if they are parallel, in which case they do not intersect at all.

What is the situation in projective geometry?

Two distinct projective points P and P' correspond to two distinct radial lines in $\mathbb{R}^3$, and, in fact, there is a unique radial plane L in $\mathbb{R}^3$ containing the latter two. See Figure A.4(a).

It follows that:

(A) There is a unique projective line containing two distinct projective points in $\mathbb{P}^2$.

How about two distinct projective lines? Observe that the corresponding two distinct radial planes, say, L and L' in $\mathbb{R}^3$, intersect in a unique radial line corresponding, in fact, to some projective point P. See Figure A.4(b). We have:

(B) Two distinct projective lines in $\mathbb{P}^2$ intersect in a unique projective point.

No exceptions! There's no such thing as parallelism in $\mathbb{P}^2$! Any two different lines always intersect in a point. Two points–one line, two lines–one point, *always*: $\mathbb{P}^2$ has better so-called *point-line duality* than $\mathbb{R}^2$. We'll have more to say about the point-line duality of $\mathbb{P}^2$ as we go along.

Exercise A.1. Consider three distinct projective lines L, L' and L''. We know that their pairwise intersections are three projective points, say, P, P' and P''. Give examples where (a) all three points are identical and (b) all three are distinct. Can only two of them be distinct? If all three are distinct can they be collinear, i.e., lie on one projective line?

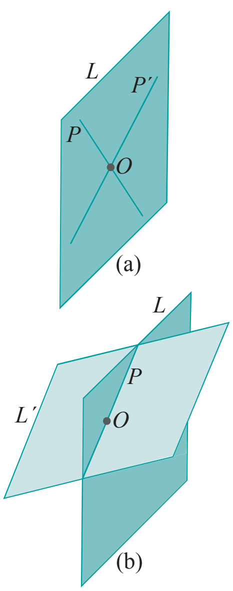

Figure A.4: (a) Radial lines corresponding to projective points P and P' are contained in a unique radial plane corresponding to the projective line L (b) Radial planes corresponding to projective lines L and L' intersect in a unique radial line corresponding to the projective point P.

A.3 Homogeneous Coordinates

We want to *coordinatize* $\mathbb{P}^2$, if possible, in a manner similar to that of $\mathbb{R}^2$ by Cartesian coordinates. This is important for the purpose of geometric calculations. For example, Cartesian coordinates on the real plane allow us to make a statement such as "The equation of the line through the $[-2 \ \ -5]^T$ and $[1 \ 1]^T$ is $y - 2x + 1 = 0$, which is satisfied as well by $[0 \ \ -1]^T$, so that all three points are collinear."

So how does one coordinatize $\mathbb{P}^2$? As follows:

Definition A.2. The *homogeneous coordinates* of a projective point are the Cartesian coordinates of any real point on it, *other than* the origin. (No, homogeneous coordinates are not unique, a projective point having many different homogeneous coordinates. This may seem strange at first but read on ....)

Example A.1. The projective point P corresponding to the radial line through $[1\ 3\ \ -2]^T$ has, as shown in Figure A.5, among others, homogeneous coordinates $[1\ 3\ \ -2]^T$, $[2\ 6\ \ -4]^T$, $[-1\ \ -3\ 2]^T$ and $[1.7\ 5.1\ \ -3.4]^T$. In fact, any tuple of the form $[c\ 3c\ \ -2c]^T$, where $c \neq 0$, can serve as homogeneous coordinates for P.

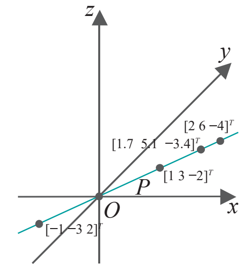

Figure A.5: The coordinates of any point on P, except the origin, can be used as its homogeneous coordinates – four possibilities are shown.

Terminology: To avoid clutter in diagrams, we'll often write homogeneous coordinates $[x\ y\ z]^T$ as (x, y, z).

That a projective point has infinitely many different homogeneous coordinates may seem odd, but it's not really a problem because two distinct projective points cannot share the same homogeneous coordinates. This is because two distinct radial lines do not share any point other than the origin. In other words, even though projective points have non-unique homogeneous coordinates, there is no risk of ambiguity. As an analogy, think of a roomful of people, each having multiple nicknames, but no two having a nickname in common – there is no danger of confusion then. As a non-zero tuple $[x\ y\ z]^T$ gives homogeneous coordinates of a unique projective point, we'll often refer to *the* projective point $[x\ y\ z]^T$ or write, say, the projective point $P = [x\ y\ z]^T$.

If you are wondering if $\mathbb{P}^2$ can at all be coordinatized in a unique manner, as is $\mathbb{R}^2$ by Cartesian coordinates, the answer is that there is no "natural" way to do this. Don't take our word for it, but give the question a bit of thought and you'll see the pitfalls. For example, a likely approach is to choose the coordinates of *one* real point from the radial line corresponding to each projective point. But then one has to come up with a *well-defined* way of choosing such a point; in other words, an *algorithm* that, given input a radial line, uniquely outputs a point from it. Try and devise such an algorithm! (The point on the line a unit distance from the origin? There are two such! The one in the positive direction? Be careful now: exactly which direction is this?)

Remark A.3. An important difference between the Cartesian and homogeneous coordinate systems is the lack of an origin in the latter. No matter how one sets up a Cartesian coordinate system in $\mathbb{R}^3$, i.e., no matter how one sets up the coordinate axes, the origin $(0, 0, \ldots, 0)$ is always distinguished as a special point. This is not the case for the homogeneous coordinate system in $\mathbb{P}^2$ – no projective point is special. It is truly homogeneous!

Example A.2. Find homogeneous coordinates of the projective point P of intersection of the projective lines L and L', corresponding, respectively, to the radial planes $2x + 2y - z = 0$ and $x - y + z = 0$.

Answer: Solving the simultaneous equations

$$\begin{aligned} 2x + 2y - z &= 0 \\ x - y + z &= 0 \end{aligned}$$

one finds that points on their intersecting line are of the form

$$y = -3x, \quad z = -4x$$

Therefore, homogeneous coordinates of P are (arbitrarily choosing $x = 1$)

$$[1\ \ -3\ \ -4]^T$$

Exercise A.2. Find homogeneous coordinates of the projective point P of intersection of the projective lines L and L' corresponding, respectively, to the radial planes $-x - y + z = 0$ and $3x + 2y = 0$.

Exercise A.3. Find the equation of the radial plane in $\mathbb{R}^3$ corresponding to the projective line L which intersects the two projective points $P = [1\ 2\ 3]^T$ and $P' = [2\ \ -1\ 0]^T$.

A.4 Structure of the Projective Plane

We're going to try and understand the structure of $\mathbb{P}^2$ by relating it to that of $\mathbb{R}^2$. In fact, we'll start off by using the homogeneous coordinate system of $\mathbb{P}^2$ to *embed* $\mathbb{R}^2$ inside $\mathbb{P}^2$.

A.4.1 Embedding the Real Plane in the Projective Plane

Associate a point $p = [x\ y]^T$ of $\mathbb{R}^2$ with the projective point $\phi(p) = [x\ y\ 1]^T$. The easiest way to picture this association is to first identify $\mathbb{R}^2$ with the plane $z = 1$; particularly, $[x\ y]^T$ of $\mathbb{R}^2$ is identified with $[x\ y\ 1]^T$ of $z = 1$. See Figure A.6. Following this, the association $p \mapsto \phi(p)$ is simply each real point with the radial line through it, in particular, the real point $[x\ y\ 1]^T$ (Cartesian coordinates) with the projective point $[x\ y\ 1]^T$ (homogeneous coordinates).

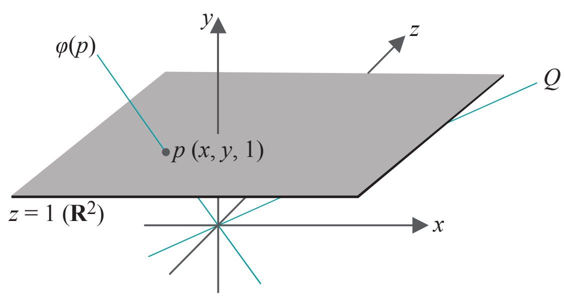

Figure A.6: Real point p on the plane $z = 1$ is associated with the projective point $\phi(p)$. Projective point Q, lying on the plane $z = 0$, is not associated with any real point.

The association $p \mapsto \phi(p)$ is clearly one-to-one as distinct points of $z = 1$ give rise to distinct radial lines through them. It's not onto as points of $\mathbb{P}^2$ that lie *on* the plane $z = 0$ or, equivalently, are parallel to $z = 1$, do not intersect $z = 1$ and, therefore, are not associated with any point of $\mathbb{R}^2$ (e.g., Q in the figure). Precisely, points of $\mathbb{P}^2$ with homogeneous coordinates of the form $[x\ y\ 0]^T$ are not associated with any point of $\mathbb{R}^2$.

$\mathbb{R}^2$, therefore, is embedded by ϕ as the *proper* subset of $\mathbb{P}^2$ consisting of radial lines intersecting $z = 1$. We're at the point now where we can try to understand how we ended up trading parallelism in $\mathbb{R}^2$ for perfect point-line duality in $\mathbb{P}^2$.

A.4.2 A Thought Experiment

Here's a thought experiment. Two parallel lines l and l' lie on $\mathbb{R}^2$, aka the plane $z = 1$ in $\mathbb{R}^3$, a distance of d apart. Points p and p' on l and l', respectively, start a distance d apart and begin to travel at the same speed and in the same direction on their individual lines. See Figure A.7. Evidently, they remain d apart no matter how far they go. Well, of course, as l and l' are parallel!

Consider next what happens to the projective points $\phi(p)$ and $\phi(p')$ associated with p and p', respectively. See again Figure A.7 to convince yourself that both $\phi(p)$ and $\phi(p')$ draw closer and closer to that particular radial line l'' on the plane $z = 0$ which is parallel to l and l'. As it lies on $z = 0$, l'' corresponds to a projective point P'' not associated with any real; in fact, P'''s homogeneous coordinates are of the form $[x\ y\ 0]^T$.

Observe that the projective point $\phi(p)$ itself travels along a projective line L – the one whose radial plane contains l. We'll call L the projective line corresponding to l. Likewise, the projective point $\phi(p')$ travels along the projective line L' corresponding to l'. Moreover, L and L' intersect in P''. See Figure A.8.

Let's take stock of the situation so far. The parallel lines l and l' on the real plane never meet, but the projective lines L and L' corresponding to them in $\mathbb{P}^2$ meet in P''.

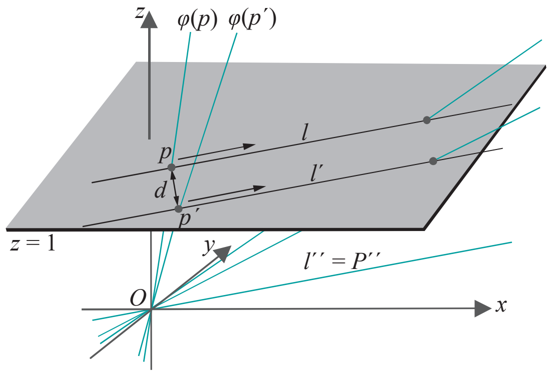

Figure A.7: The real points p and p' travel along parallel lines l and l'. Associated projective points $\phi(p)$ and $\phi(p')$ travel with p and p'.

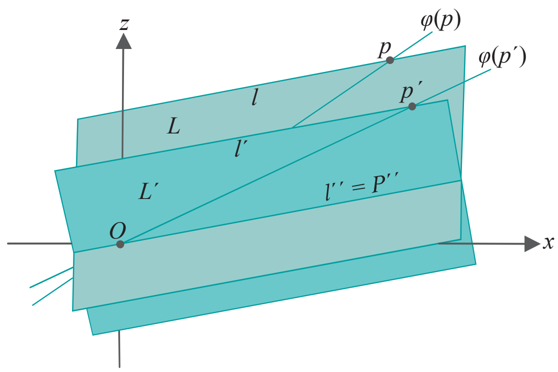

Figure A.8: $\phi(p)$ travels along L and $\phi(p')$ along L'. L and L' meet at P''.

Moreover, every point of L or L', *except* for P'', is associated by ϕ to a point of l or l', respectively. We can say then that the projective line L equals its real counterpart l *plus* the extra point P''; L', likewise, is its real counterpart l' *plus* P''. And, it's at this point P'', beyond the reals, that the two projective lines meet, while their real counterparts never do.

Example A.3. What if both points p and p', and together with them $\phi(p)$ and $\phi(p')$, travel along their respective lines in directions opposite to those indicated in Figure A.7? What if only one reversed its direction?

Answer: If both p and p' reversed directions, then again they would travel forever exactly d apart. If only one of the two reversed its direction, then, of course, the distance between them would continuously increase.

However, in either case, $\phi(p)$ and $\phi(p')$ draw closer, again both to P''. It seems that, whatever the sense of travel is of $\phi(p)$ and $\phi(p')$ along their respective projective lines L and L', they approach that one point of intersection of these two lines. Two points traveling in opposite directions along a real line ultimately grow farther and farther apart. A projective line, on the other hand, apparently behaves more like a circle.

A.4.3 Regular Points and Points at Infinity

Recall *equivalence relations* and *equivalence classes* from undergrad discrete math. In particular, recall that the lines of $\mathbb{R}^2$ can be split into equivalence classes by the equivalence relation of being parallel. Consider any equivalence class $\mathbf{l}$ of parallel lines of $\mathbb{R}^2$, the latter being identified with the plane $z = 1$ in $\mathbb{R}^3$ as before. There is a unique radial line l on the plane $z = 0$ parallel to the members of $\mathbf{l}$. See Figure A.9.

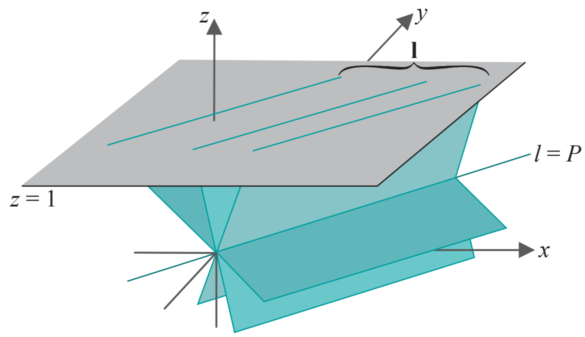

Figure A.9: The line l (= projective point P) is parallel to lines in $\mathbf{l}$. P is said to be the point at infinity along the equivalence class $\mathbf{l}$ of parallel lines.

Denote the projective point corresponding to l by P. Projective lines corresponding to lines in $\mathbf{l}$ all meet at P, because their radial planes each contain l. The point P, which is not associated with any real point by ϕ as it lies on $z = 0$, is called the *point at infinity* along $\mathbf{l}$ or, simply, the point at infinity along any one of the lines in $\mathbf{l}$. Conversely, any radial line l on the plane $z = 0$ is the point at infinity along the equivalence class of lines in $\mathbb{R}^2$ parallel to it. In other words, the correspondences

$$\begin{aligned} \text{equivalence class of parallel lines in } \mathbb{R}^2 &\leftrightarrow \text{radial line on } z = 0 \\ &\leftrightarrow \text{point at infinity of } \mathbb{P}^2 \end{aligned}$$

are both one-to-one. Note that points at infinity of $\mathbb{P}^2$ are precisely those with homogeneous coordinates of the form $[x\ y\ 0]^T$.

Returning to the thought experiment of Section A.4.2, one can imagine points at infinity plugging the "holes" along the "border" of $\mathbb{R}^2$ through which parallel lines "run off" without meeting, which explains why every pair of lines on the projective plane meets.

Projective points which are not points at infinity are called *regular points*. Regular points have homogeneous coordinates of the form $[x\ y\ z]^T$, where z is *not* zero. Moreover, regular points intersect $z = 1$, so are associated each by ϕ^{-1} with a point of $\mathbb{R}^2$ (remember ϕ takes a real point of $\mathbb{R}^2$, represented by the plane $z = 1$, to the projective point whose corresponding radial line passes through that point). Accordingly, one can write:

$$\mathbb{P}^2 = \mathbb{R}^2 \cup \{\text{points at infinity}\} = \{\text{regular points}\} \cup \{\text{points at infinity}\}$$

The union of all points at infinity, called the *line at infinity*, is the projective line whose radial plane is $z = 0$. Therefore, one can as well write:

$$\mathbb{P}^2 = \mathbb{R}^2 \cup \text{line at infinity} = \{\text{regular points}\} \cup \text{line at infinity}$$

Our embedding ϕ of $\mathbb{R}^2$ as a subset of $\mathbb{P}^2$ depends on the plane $z = 1$, particularly because we identify $z = 1$ with $\mathbb{R}^2$ and subsequently associate each point of $\mathbb{R}^2$ with the radial line in $\mathbb{R}^3$ through it. Is there anything special about the plane $z = 1$? Not at all. It just seemed convenient. In fact, we could have used any *any* non-radial plane p.

Exercise A.4. Why does p have to be non-radial?

Example A.4. Instead of $z = 1$, identify $\mathbb{R}^2$ with the plane $x = 2$ in $\mathbb{R}^3$. Accordingly, embed $\mathbb{R}^2$ into $\mathbb{P}^2$ by associating $[x\ y]^T$ with the radial line through $[2\ x\ y]^T$. Which now are the regular points and which are the points at infinity of $\mathbb{P}^2$?

Answer: The regular points of $\mathbb{P}^2$ are the radial lines in $\mathbb{R}^3$ which intersect the plane $x = 2$. These are precisely the radial lines which do not lie on the plane $x = 0$. The points at infinity are the radial lines which do lie on the plane $x = 0$. Equivalently, regular points have homogeneous coordinates of the form $[x\ y\ z]^T$, where $x \neq 0$, while points at infinity have homogeneous coordinates of the form $[0\ y\ z]^T$.

Exercise A.5. Identify $\mathbb{R}^2$ with the plane $x + y + z = 1$ in $\mathbb{R}^3$, embedding it into $\mathbb{P}^2$ by associating $[x\ y]^T$ with the radial line through $[x\ y\ \ 1 - x - y]^T$. Which now are the regular points and which the points at infinity of $\mathbb{P}^2$?

It may seem strange at first that the separation of $\mathbb{P}^2$ into regular points and points at infinity depends on the particular embedding of $\mathbb{R}^2$ in $\mathbb{P}^2$. However, this situation becomes clearer after a bit of thought. It's related, as a matter of fact, to the discussion at the beginning of the chapter, where we motivated projective spaces by observing that lines through a point camera are perceived as points on the plane film. Even though all lines through the camera do not intersect the film, we argued this to be merely an artifact of the alignment of the film, the latter being changeable. Therefore, we concluded that all radial lines should be taken as points in projective space.

We now come full circle back to this initially motivating scenario. Embedding $\mathbb{R}^2$ in $\mathbb{P}^2$ corresponds exactly to choosing an alignment of the film – the film is a copy of $\mathbb{R}^2$ and each point on it associated with the light ray (= radial line in $\mathbb{R}^3$ = point of $\mathbb{P}^2$) through that point to the camera. Light rays toward the camera which intersect the film are regular points of $\mathbb{P}^2$ and visible, while those which do not are points at infinity and invisible. Moreover, the line at infinity corresponds to the plane through the camera parallel to the film. And, of course, we are at perfect liberty to align the film, i.e., embed $\mathbb{R}^2$ in $\mathbb{P}^2$, as we like, different choices leading to different sets of visible and invisible light rays.

A.5 Snapshot Transformations

Here's another interesting thought experiment.

Example A.5. A point camera is at the origin with two power lines passing over it, both parallel to the x-axis. One lies along the line $y = 2, z = 2$ (i.e., the intersection of the planes $y = 2$ and $z = 2$) and the other along the line $y = -2, z = 2$.

Take "snapshots" of the power lines with the film aligned along (a) the plane $z = 1$ and (b) the plane $x = 1$. Sketch and compare the two snapshots.

Answer: This is one you might want to try yourself before reading on!

See Figure A.10. Figure A.10(a) shows the snapshot (or, projection) of the power lines (thin black lines) on the plane $z = 1$. These projections are the two *parallel* lines $y = 1$ and $y = -1$ (green). This is not hard to understand: by simple geometry, the line $y = 2, z = 2$ projects toward the origin (the camera) to the line $y = 1$ on the plane $z = 1$; likewise, $y = -2, z = 2$ projects to $y = -1$ on $z = 1$.

Figure A.10(b) shows the snapshot on the plane $x = 1$. It is the two *intersecting* lines $z = y$ and $z = -y$ making an X-shape. This requires explanation. The top of the X, above its center $[1\ 0\ 0]^T$, is formed from intersections with the film of light rays through points on the power lines with x-value greater than zero, while the bottom from rays through points with x-value less than zero. The rays from the points on either power line with x-value equal to zero do not strike the film.

The point $[1\ 0\ 0]^T$ at the center of the X is included in the snapshot, though no ray from either power line passes through it, because it's the intersection with the film of the "limit" of the rays from points on either power line as they run off to infinity. It's convenient to imagine the limits of visible rays as being visible as well and we ask the reader to accept this. In geometric drawing parlance $[1\ 0\ 0]^T$ is the *vanishing point* of the power lines – it's where they *seem* to meet on the film $x = 1$.

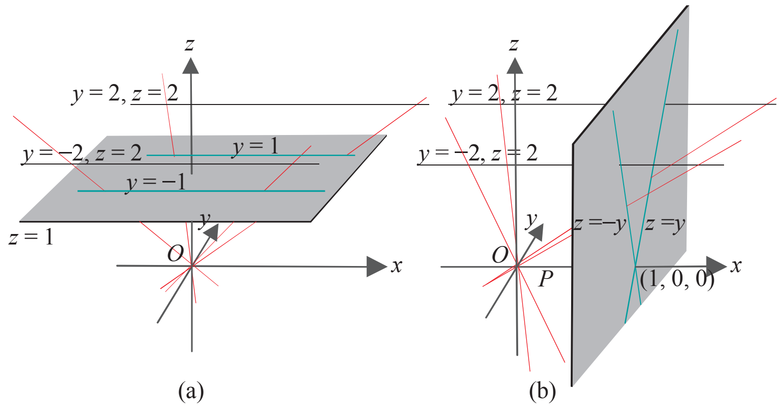

Figure A.10: Thin black power lines $y = \pm 2, z = 2$ projected onto the planes (a) $z = 1$ and (b) $x = 1$ as green lines. Red lines depict light rays. The x-axis corresponds to the projective point P.

Contemplate the situation from the point of view of projective geometry. The projective lines corresponding to the two power lines meet at the projective point P corresponding to the x-axis, as the radial planes through the power lines intersect in the x-axis. Now, P is a point at infinity with respect to the plane $z = 1$ (because the x-axis doesn't intersect this plane), while it's a regular point with respect to the plane $x = 1$ (because the x-axis intersects this plane at $[1\ 0\ 0]^T$). In terms of shooting pictures, then, the camera with its film along $z = 1$ cannot see where the two power lines meet, so they appear parallel. However, with its film along $x = 1$ the camera sees them meet at $[1\ 0\ 0]^T$.

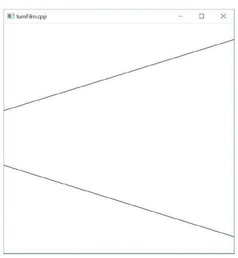

Figure A.11: Screenshot of `turnFilm.cpp`.

Experiment A.1. Run `turnFilm.cpp`, which animates the setting of the preceding exercise by means of a viewing transformation. Initially, the film lies along the $z = 1$ plane. Pressing the right arrow key rotates it toward the $x = 1$ plane, while pressing the left one reverses the rotation. Figure A.11 is a screenshot midway. You cannot, of course, see the film, only the view of the lines as captured on it.

The reason that the lower part of the X-shaped image of the power lines cannot be seen is that OpenGL film doesn't capture rays hitting it from behind, as the viewing plane is a clipping plane too. Moreover, if the lines seem to actually meet to make a V after the film turns a certain finite amount, that's because they are very long and your monitor has limited resolution!

This program itself is simple with the one statement of interest being `gluLookAt()`, which we ask the reader to examine next. **End**

Exercise A.6. (Programming) Verify that the `gluLookAt()` statement of `turnFilm.cpp` indeed simulates the film's rotation as claimed between the $z = 1$ and $x = 1$ planes.

Example A.6. Refer to Figure A.10(b). Suppose two power lines *actually* lie along the two intersecting lines $z = y$ and $z = -y$ on the plane $x = 1$, which is the snapshot on the plane $x = 1$ of the power lines of the preceding example. What would *their* snapshot look like on the films $z = 1$ and $x = 1$?

Answer: Exactly as in the preceding Example A.5, as depicted in Figures A.10(a) and (b)! It's not possible to distinguish between these two pairs of power lines – the pair in Example A.5 being "really" parallel and the current one "really" intersecting – with a point camera at the origin.

A somewhat whimsical take on all this is to imagine a Matrix-like world where one can never know reality. Perception is limited to whatever is captured on film. Therefore, one agent's intersecting power lines are just as real as the other's parallel ones!

It's useful to think of one snapshot of Example A.5 or A.6 as a *transformation* of the other. Keep in mind that if a snapshot appears as the two parallel lines $y = \pm 1$ on the film $z = 1$, then it always appears as the two intersecting lines $z = \pm y$ on the film $x = 1$, *regardless* of what the "real" objects are.

Convince yourself of this by mentally tilting one of the power lines in Figure A.10(a) on the radial plane (not drawn) through it, so that its projection on the $z = 1$ plane does not change. The power line's projection on the $x = 1$ plane remains unchanged, as well, because the set of light rays from it through the camera doesn't change. For this reason, it makes sense to talk of transforming one snapshot to another, without any reference to the real scene. We'll informally call such transformations *snapshot transformations*.

Remark A.4. Snapshot transformations as described are not really transformations in the mathematical sense, as they don't map some space to itself but, rather, one plane (film) to another. A rigorous formulation is possible, though likely not worth the effort, as we'll see soon that snapshot transformations are subsumed within the class of projective transformations, which we'll be studying in depth. Nevertheless, the notion of a snapshot transformation is geometrically intuitive and useful.

Here are more for you to ponder.

Exercise A.7. In each case below you are told what the snapshot looks like on the film $z = 1$, aka $\mathbb{R}^2$, and asked what is captured on the film $x = 1$. The $z = 1$ shots are drawn in Figure A.12, each labeled the same as the item to which it corresponds. You don't have to find equations for your answer for $x = 1$. Just a sketch or a verbal description is enough.

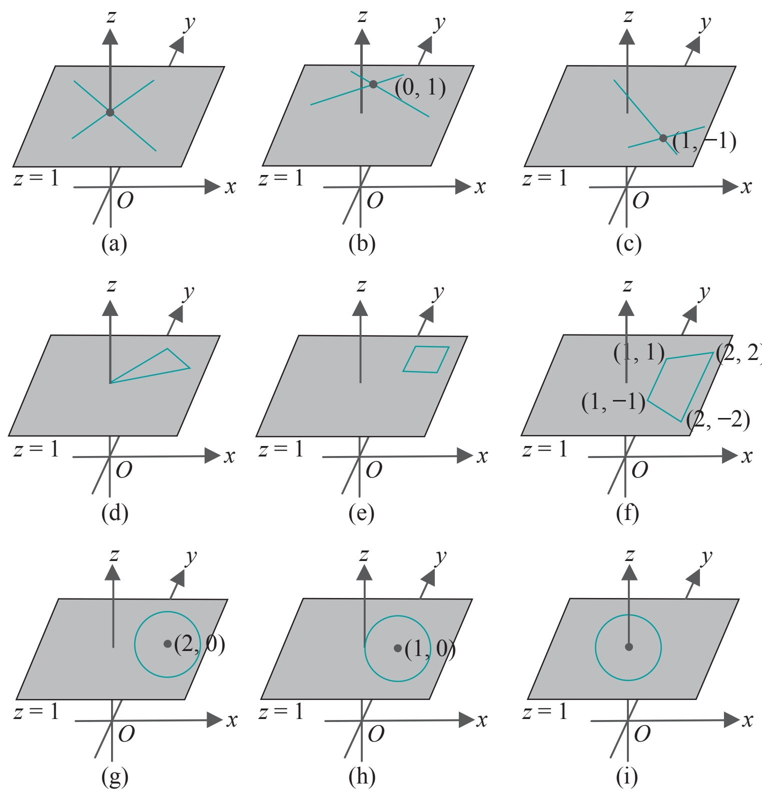

Figure A.12: Transform these snapshots on the plane $z = 1$ to the plane $x = 1$. Some points on the plane $z = 1$ are shown with their xy coordinates. Labels correspond to items of Exercise A.7.

Answers are in italics. Figure A.13 justifies the answer to (h).

(a) Two lines that intersect at the origin, neither being the y-axis (on $\mathbb{R}^2$).
Two parallel lines. Why the caveat? What happens if one is the y-axis?

(b) Two lines that intersect at the point $[0\ 1]^T$, neither being the y-axis.
Two parallel lines.

(c) Two lines that intersect at the point $[1\ \ -1]^T$.
Two intersecting lines.

(d) A triangle in the upper-right quadrant with one vertex at the origin but, otherwise, not touching the axes.
An infinitely long U-shape with straight sides.

(e) A square in the upper-right quadrant not touching any of the axes.
A quadrilateral with no two parallel sides.

(f) A trapezoid symmetric about the x-axis with vertices at $[1\ 1]^T$, $[1\ -1]^T$, $[2\ -2]^T$ and $[2\ 2]^T$.
A rectangle.

(g) A unit radius circle centered at $[2\ 0]^T$.
An ellipse.

(h) A unit radius circle centered at $[1\ 0]^T$.
A parabola – see Figure A.13.

(i) A unit radius circle centered at the origin.
A hyperbola.

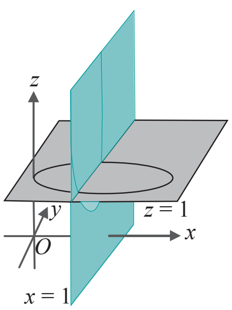

Figure A.13: Answer to Exercise A.7(h).

Remark A.5. Exercise A.7(f) seems innocuous enough, but it is very important. Its generalization to 3D will help convert viewing frustums to rectangular boxes in the graphics pipeline.

Exercise A.8. Refer to the geometric construction of conic sections in Section 10.1.5 as plane sections of a double cone, and show that any non-degenerate conic section can be snapshot transformed to another such.

Exercise A.9. (Programming) Write code similar to `turnFilm.cpp` to animate the snapshot transformation of Exercise A.7(h). Again, you'll see only part of the parabola because OpenGL cannot see behind its film.

It's not hard to see that none of the snapshot transformations of Exercise A.7, except for (c) and (g), can be accomplished using OpenGL modeling transformations. This is because they are not affine – recall from Section 5.4.5 that OpenGL implements only affine transformations.

Remark A.6. We just said that most of the snapshot transformations of Exercise A.7 are not affine and yet seem to be suggesting with the preceding Exercise A.9 that they may be implemented by means of an OpenGL viewing transformation. We know, however, that the latter is equivalent to a sequence of modeling transformations and, therefore, affine.

The apparent conundrum is not hard to resolve. The result of the viewing transformation of, e.g., `turnFilm.cpp`, is indeed a snapshot transformation in terms of what is *seen on the screen*. In other words, the transformation from the OpenGL window prior to applying the viewing transformation to that after is a snapshot transformation. However, the viewing transformation serves only to change the scene to one which OpenGL *projects* onto the window as the new one. A snapshot transformation, therefore, is more than a viewing transformation – it's a viewing transformation *plus* a projection.

Exercise A.10. By considering how to turn the film, i.e., viewing plane, show that implementing a snapshot transformation in OpenGL is equivalent to:

(a) setting the *centerx*, *centery*, *centerz*, *upx*, *upy* and *upz* parameters of the viewing transformation

```
gluLookAt(0, 0, 0, centerx, centery, centerz, upx, upy, upz)
```

and

(b) setting the *near* parameter of the perspective projection call

```
glFrustum( left, right, bottom, top, near, far)
```

where the other five parameters can be kept fixed at some initially chosen values.

A.6 Homogeneous Polynomial Equations

The only application we've made so far of homogeneous coordinates is to embed $\mathbb{R}^2$ in $\mathbb{P}^2$. We haven't used them yet to write equations of curves on the projective plane. Let's try now to do this.

We'll start with the simplest curve on the projective plane, in fact, a projective line. We want an equation – as for straight lines in real geometry – that will say if a projective point belongs to a projective line. For example, an equation such as $2x + y - 1 = 0$ for a straight line on the real plane gives the condition for a real point $[x\ y]^T$ to lie on that line.

Now, a projective point is a radial line and a projective line a radial plane. Moreover, a radial line lies on a radial plane if and only if any point of it, other than the origin, lies on that plane (the origin always does). See Figure A.14.

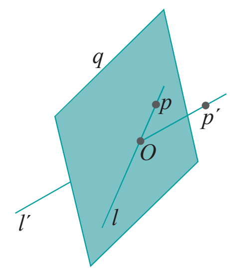

Figure A.14: Point p of radial line l lies on radial plane q, implying that l lies on q; point p' of l' doesn't lie on q, implying that no point of l', other than the origin, lies on q.

Therefore, a projective point $P = [x\ y\ z]^T$ belongs to a projective line L, whose radial plane has the equation $ax + by + cz = 0$, if and only if the real point $[x\ y\ z]^T$ lies on the real plane $ax + by + cz = 0$. It follows that the equation of L is identical to that of its radial plane:

$$ax + by + cz = 0 \tag{A.1}$$

Accordingly, a projective point $P = [x\ y\ z]^T$ belongs to L if it satisfies (A.1). Does it matter if we choose some other homogeneous coordinates for P? No, because

$$a(kx) + b(ky) + c(kz) = k(ax + by + cz) = 0$$

so any homogeneous coordinates $[kx\ ky\ kz]^T$ for P satisfy Equation (A.1).

Exercise A.11. Prove that if the projective line L is specified by the equation

$$ax + by + cz = 0$$

then it is specified by any equation of the form

$$(ma)x + (mb)y + (mc)z = 0$$

where $m \neq 0$, as well.

Exercise A.12. What is the equation of the projective line through the projective points $[2\ 1\ -1]^T$ and $[3\ 4\ 2]^T$?

Answer: Suppose that the line is L with equation

$$ax + by + cz = 0$$

Since $[2\ 1\ -1]^T$ and $[3\ 4\ 2]^T$ lie on L they must satisfy its equation, giving

$$\begin{aligned} 2a + b - c &= 0 \\ 3a + 4b + 2c &= 0 \end{aligned}$$

Any solution to these simultaneous equations, not all zero, then determines L. As there are more variables then equations, let's set one of them, say c, arbitrarily to 1, to get the equations

$$\begin{aligned} 2a + b - 1 &= 0 \\ 3a + 4b + 2 &= 0 \end{aligned}$$

These solve to give $a = 1.2$ and $b = -1.4$. The equation of the projective line L is, therefore,

$$1.2x - 1.4y + z = 0$$

(or, equivalently, $6x - 7y + 5z = 0$, from Exercise A.11.)

Exercise A.13. What is the projective point of intersection of the projective lines $3x + 2y - 4z = 0$ and $x - y + z = 0$?

Exercise A.14. When are three projective points $[x\ y\ z]^T$, $[x'\ y'\ z']^T$ and $[x''\ y''\ z'']^T$ collinear, i.e., when do they belong to the same projective line? Find a simple condition involving a determinant.

A.6.1 More About Point-Line Duality

In Section A.4.2 we tried to understand the point-line duality of the projective plane from a geometric point of view. We'll examine the phenomenon now from an algebraic standpoint.

The correspondence from the set of projective points to the set of projective lines given by

$$\text{projective point } [a\ b\ c]^T \quad \mapsto \quad \text{projective line } ax + by + cz = 0 \tag{A.2}$$

is well-defined as, whatever homogeneous coordinates we choose for a projective point, the image is the same projective line (by Exercise A.11). Moreover, the correspondence is easily seen to be one-to-one and onto.

Definition A.3. The projective line $ax + by + cz = 0$ is said to be the *dual* of the projective point $[a\ b\ c]^T$ and vice versa.

Exercise A.15. Prove that a projective point P belongs to a projective line L if and only if the dual of L belongs to the dual of P.

The preceding exercise implies that if some statement about the incidence of projective points and lines is true, then so is the dual statement, obtained by replacing "point" with "line" and "line" with "point".

Exercise A.16. What is the dual of the following statement? "There is a unique projective line incident to two distinct projective points."

From this last exercise one sees, then, the point-line duality of the projective plane as a consequence of the one-to-one correspondence (A.2) between projective points and lines. We ask the reader to contemplate if there exists a similar correspondence between real points and lines.

A.6.2 Lifting an Algebraic Curve from the Real to the Projective Plane

Let's see next projective curves more complex than a line. Consider, then, the curve Q' in $\mathbb{P}^2$ consisting of the projective points intersecting the parabola q

$$y - x^2 = 0 \tag{A.3}$$

on $\mathbb{R}^2$ (the plane $z = 1$). See Figure A.15.

The intersection of the projective point $P = [x\ y\ z]^T$ with the plane $z = 1$ is the real point $[x/z\ \ y/z\ 1]^T$, assuming $z \neq 0$, for, otherwise, there is no intersection. Now, $[x/z\ \ y/z\ 1]^T$ satisfies the equation of the parabola q if

$$y/z - (x/z)^2 = 0 \qquad \Longrightarrow \qquad yz - x^2 = 0$$

Accordingly, the curve consisting of projective points $[x\ y\ z]^T$ which satisfy

$$yz - x^2 = 0 \tag{A.4}$$

is called the *lifting* Q of q from the real to the projective plane. Q is sometimes simply called the lifting of q and also the *projectivization* of q. In this particular case, as a lifting of a parabola, Q is a *parabolic projective curve*.

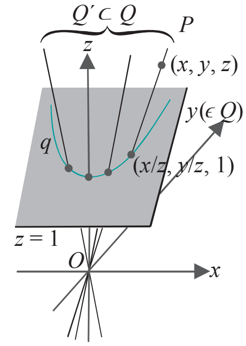

Figure A.15: Lifting a parabola drawn on the real plane $z = 1$ to the projective plane.

The camera analogy is that Q is the set of rays seen, by intersection with the film $z = 1$, as q. However, Q is actually one point *bigger* than Q', the set of projective points intersecting the parabola q on $\mathbb{R}^2$, as it includes the projective point $[0\ 1\ 0]^T$, the y-axis of 3-space, which satisfies (A.4), but does not intersect q. So, $Q = Q' \cup \{[0\ 1\ 0]^T\}$. We can justify the inclusion of this extra point, with the help of the proviso from Section A.5 that a limit of visible rays is visible, as follows.

From its equation $y - x^2 = 0$, a point of q is of the form $[x\ x^2\ 1]^T$, for any x. Therefore, the homogeneous coordinates of a projective point intersecting q are $[x\ x^2\ 1]^T$, for any x, as well. Rewriting these coordinates as $[\frac{1}{x}\ 1\ \frac{1}{x^2}]^T$ we see that its limit as $x \to \infty$ is indeed $[0\ 1\ 0]^T$. More intuitively, *a la* the thought experiment of Section A.4.2, as a point p travels off along either wing of the parabola, the projective point $\phi(p)$, corresponding to the line through p, approaches $[0\ 1\ 0]^T$, the projective point corresponding to the y-axis.

Definition A.4. A *homogeneous polynomial* is one whose terms each have the same degree, the degree of a term being the sum of the powers of the variables in the term. This common degree is called the degree of the homogeneous polynomial.

An equation with a homogeneous polynomial on the left and 0 on the right is called a homogeneous polynomial equation.

The equations $ax + by + cz = 0$ of a projective line and $yz - x^2 = 0$ of a parabolic projective curve are homogeneous polynomial equations of degree one and two, respectively. That they are both homogeneous is no accident, as we'll soon see.

Exercise A.17. Suppose that $p(x_1, x_2, \ldots, x_n)$ is a homogeneous polynomial in n variables. Then, if $[x_1\ x_2\ \ldots\ x_n]^T$ satisfies the equation $p(x_1, x_2, \ldots, x_n) = 0$, so does $[cx_1\ cx_2\ \ldots\ cx_n]^T$, for any scalar c.
Hint: Show, first, that, if $p(x_1, x_2, \ldots, x_n)$ is homogeneous of degree r, then

$$p(cx_1, cx_2, \ldots, cx_n) = c^r p(x_1, x_2, \ldots, x_n)$$

For example, for the homogeneous polynomial $yz - x^2$ of degree 2,

$$(cy)(cz) - (cx)^2 = c^2(yz - x^2)$$

So, in this case, if (x, y, z) satisfies $yz - x^2 = 0$, then so does (cx, cy, cz), because $(cy)(cz) - (cx)^2 = 0$ as well, by the equation just above.

The preceding exercise implies that a homogeneous polynomial equation of the form $p(x, y, z) = 0$ is legitimately an equation in $\mathbb{P}^2$, because a point of $\mathbb{P}^2$ can be tested if it satisfies $p(x, y, z) = 0$, independently of the homogeneous coordinates used to represent the point. Here are some definitions.

Definition A.5. An *algebraic curve* on the real plane consists of points satisfying an equation of the form

$$p(x, y) = 0$$

where p is a polynomial in the two variables x and y. The degree of the curve is the highest degree of a term belonging to $p(x, y)$.

Familiar algebraic curves of degree one include straight lines, e.g., $2x + y - 3 = 0$, while conic sections, e.g., the hyperbola $xy - 1 = 0$, are of degree two.

Definition A.6. A *projective algebraic curve* on the projective plane consists of points satisfying an equation of the form

$$p(x, y, z) = 0$$

where p is a homogeneous polynomial in the three variables x, y and z. The degree of the curve is the degree of $p(x, y, z)$.

Projective algebraic curves that we have already seen are the projective line $ax + by + cz = 0$ of degree one and the projective parabola $yz - x^2 = 0$ of degree two. Let's get a few more via lifting.

Example A.7. Lift the algebraic curve of degree 3

$$x^3 + 3x^2y + y^2 + x + 2 = 0$$

drawn on the plane $z = 1$, to $\mathbb{P}^2$.

Answer: The projective point $[x\ y\ z]^T$ intersects the plane $z = 1$ at the real point $[x/z\ y/z\ 1]^T$ (assuming $z \neq 0$). Accordingly, replace x by x/z and y by y/z in the given polynomial equation:

$$\begin{aligned} & (x/z)^3 + 3(x/z)^2(y/z) + (y/z)^2 + x/z + 2 = 0 \\ \Longrightarrow\quad & x^3/z^3 + 3x^2y/z^3 + y^2/z^2 + x/z + 2 = 0 \\ \Longrightarrow\quad & x^3 + 3x^2y + y^2z + xz^2 + 2z^3 = 0 \end{aligned}$$

defining the lifted curve, a projective algebraic curve of degree 3.

Exercise A.18. Lift the algebraic curve of degree 5

$$xy^4 - 2x^2y^2 + 3xy^2 + y^3 - xy + 2 = 0$$

drawn on the plane $z = 1$, to $\mathbb{P}^2$.

Exercise A.19. Show that the lifting of the straight line

$$ax + by + c = 0$$

drawn on the plane $z = 1$, to $\mathbb{P}^2$, in fact, is the projective line corresponding to it, as defined in Section A.4.2. Moreover, this line is a projective algebraic curve of degree 1.

It should be fairly clear at this point that the lifting of an algebraic curve $p(x, y) = 0$ is a projective algebraic curve $\overline{p}(x, y, z) = 0$ of the same degree. We leave a formal proof to the reader in the following exercise.

Exercise A.20. Show that the lifting of an algebraic curve $p(x, y) = 0$ of degree r is a projective algebraic curve $\overline{p}(x, y, z) = 0$ of degree r.

Definition A.7. The process of going from the equation of an algebraic curve on the real plane to the homogeneous polynomial equation of its lifting is called *homogenization*.

It's worth keeping mind that the process of homogenization depends on the particular plane on which the algebraic equation holds. E.g., in Example A.7 and Exercises A.18-A.19 the plane was $z = 1$. This need not always be the case as we see next.

Example A.8. Homogenize the polynomial equation

$$y^2 + z^2 + z = 0$$

drawn on the plane $x = 2$. (So, $x = 2$ is treated as a copy of the yz-plane.)

Answer: The projective point $[x\ y\ z]^T$ intersects the plane $x = 2$ at the real point $[2\ 2y/x\ 2z/x]^T$ (assuming $x \neq 0$, and multiplying $[x\ y\ z]^T$ by $2/x$). Accordingly, replace y by $2y/x$ and z by $2z/x$ in the given polynomial equation:

$$(2y/x)^2 + (2z/x)^2 + 2z/x = 0 \quad \Longrightarrow \quad 4y^2/x^2 + 4z^2/x^2 + 2z/x = 0$$

Multiplying throughout by x^2 one gets the homogenized polynomial equation

$$4y^2 + 4z^2 + 2xz = 0$$

Not surprisingly, giving the algebraic equation on different real planes corresponds, simply, to specifying the algebraic curve as seen by the viewer on differently aligned films. The lifting itself, of course, is the set of rays intersecting the film in the given curve, which does not change.

Exercise A.21. Homogenize the polynomial equation

$$3x^4 + 2x^2y + 2y^3 + 2x^2 + xy + x = 0$$

drawn on the plane $z = 4$.

Exercise A.22. Homogenize the polynomial equation

$$x^3 + 2xz - z^4$$

drawn on the plane $y = 2$.

Remark A.7. It's possible to define the homogenization of a polynomial in an abstract manner independent of reference to a particular plane. See Jennings [78].

One sees, then, that the algebraic analogue of lifting an algebraic curve from the real to the projective plane is homogenization. The reverse process of projecting a (projective algebraic) curve onto a real plane consists of taking the section of the projective points composing the curve with the given plane. Algebraically, this means simultaneously solving the equation of the curve and that of the plane – a process not surprisingly called *de-homogenization*.

Example A.9. Project the curve

$$yz - x^2 = 0$$

in $\mathbb{P}^2$ onto the real plane $z = 1$.

Answer: De-homogenize the equation of the curve by simultaneously solving

$$\begin{aligned} yz - x^2 &= 0 \\ z &= 1 \end{aligned}$$

to get

$$y - x^2 = 0$$

which is the equation of a parabola.

Exercise A.23. Project the curve of the preceding example onto the real plane $x = 1$.

Exercise A.24. Project the curve

$$4y^2 + 4z^2 + 2xz = 0$$

in $\mathbb{P}^2$ onto the real plane $y = -2$.

A.6.3 Snapshot Transformations Algebraically

It should make sense now that the snapshot transformation of an algebraic curve c from one real plane p to another p' can be determined by (a) first homogenizing the equation of c to lift it to the projective plane, and, then (b) de-homogenizing to project it back onto p'.

Example A.10. Let's solve the snapshot transformation problem of Exercise A.7(h) algebraically. The equation of the unit circle, centered at $[1\ 0]^T$ on the $z = 1$ plane, is

$$x^2 + y^2 - 2x = 0$$

Homogenizing, one gets

$$x^2 + y^2 - 2xz = 0$$

To project onto the plane $x = 1$, de-homogenize by simultaneously solving

$$\begin{aligned} x^2 + y^2 - 2xz &= 0 \\ x &= 1 \end{aligned}$$

to get

$$y^2 - 2z + 1 = 0 \quad \Longrightarrow \quad z = \frac{1}{2}y^2 + \frac{1}{2}$$

which indeed agrees with the sketch of a parabola in Figure A.13.

Exercise A.25. Solve Exercises A.7(g) and (i) algebraically.

A.7 The Dimension of the Projective Plane and Its Generalization to Higher Dimensions

Note: The next few paragraphs about $\mathbb{P}^2$ as a surface require recollecting some of the material from Section 10.2.12 on surface theory. If the reader is not inclined to do so, then she can safely skip ahead to Definition A.8. It won't affect her understanding of anything that follows.

Why do we say that the projective plane is a projective space of dimension 2? Because, as we'll see momentarily, $\mathbb{P}^2$ is a surface. In fact, it's a regular C^∞ surface, *except* that it is not a subset of $\mathbb{R}^3$: it's not possible to embed $\mathbb{P}^2$ in $\mathbb{R}^3$. One must go at least one dimension higher to $\mathbb{R}^4$.

Ignoring for now the question of the space in which it's embedded, it's not hard to find a coordinate patch containing any given point $P \in \mathbb{P}^2$. Suppose, for the moment, that P intersects the point p on the plane $z = 1$ (our favorite copy of $\mathbb{R}^2$). See Figure A.16. Let W be a closed rectangle containing p and B be the set of projective points intersecting W. The function

$$\text{point} \quad \mapsto \quad \text{the radial line through it}$$

from W to B is a one-to-one correspondence that makes B a coordinate patch.

And what if P doesn't intersect $z = 1$, i.e., if P is a point at infinity with respect to $z = 1$? Remember, there's nothing special about $z = 1$ – simply choose another non-radial plane with respect to which P is regular.

The reader has guessed by now that there exist projective spaces of various dimensions. True.

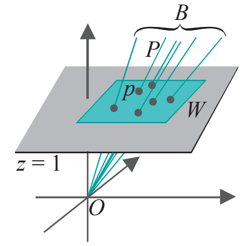

Figure A.16: The coordinate patch B containing P in $\mathbb{P}^2$ is in one-to-one correspondence with the rectangle W containing p in $\mathbb{R}^2$ (a few points in W and their corresponding projective points are shown).

Definition A.8. A radial line in $\mathbb{R}^{n+1}$ is said to be an *n-dimensional projective point*. The set of all n-dimensional projective points is *n-dimensional projective space*, denoted $\mathbb{P}^n$.

$\mathbb{P}^0$, not very interestingly, is a one-point space as there is only one line, radial or otherwise, in $\mathbb{R}^1$. We'll try to convince the reader next, without being mathematically precise, that $\mathbb{P}^1$ is a circle.

Let U be the upper-half of a circle centered at the origin of $\mathbb{R}^2$. Associate with each radial line in $\mathbb{R}^2$ its intersection(s) with U. See Figure A.17, where, e.g., the radial line P is associated with the point p. Each radial line in $\mathbb{R}^2$ is then associated with a unique point of U, except for the x-axis, which we denote Q; Q intersects U in two points q_1 and q_2. And, the other way around, every point of U is associated with a unique radial line, except only for q_1 and q_2, which are associated with the same one Q. It follows, then, that the set $\mathbb{P}^1$ of all radial lines in $\mathbb{R}^2$ is in one-to-one correspondence with the space obtained by "identifying" the two endpoints q_1 and q_2 of U as one. But this latter space is clearly a circle (imagine U as a length of string whose ends are brought together).

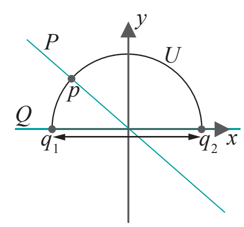

Figure A.17: Identifying $\mathbb{P}^1$ with a circle.

One can set up homogeneous coordinates for an arbitrary $\mathbb{P}^n$ in a manner similar to what we did for $\mathbb{P}^2$. For example, the homogeneous coordinates of a point $P \in \mathbb{P}^3$ are the coordinates of any point, other than the origin, on the radial line in $\mathbb{R}^4$ to which it corresponds. So the homogeneous coordinates of the point in $\mathbb{P}^3$ corresponding to the radial line through $[x\ y\ z\ w]^T$, where x, y, z and w are not all zero, is any tuple of the form $[cx\ cy\ cz\ cw]^T$, where $c \neq 0$.

It's hard to visualize $\mathbb{P}^3$ and higher-dimensional projective spaces for the same reason that it's hard to visualize $\mathbb{R}^4$ and higher-dimensional real spaces. The trick is to develop one's intuition in $\mathbb{P}^2$, as many of its properties do generalize.

A.8 Projective Transformations Defined

That the homogeneous coordinates of a point $P \in \mathbb{P}^2$ are of the form $[x\ y\ z]^T$ suggests defining transformations of $\mathbb{P}^2$ by mimicking the definition of a linear transformation of real 3-space. In particular, if

$$M = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

is a 3×3 matrix, then tentatively define a transformation of $\mathbb{P}^2$ by

$$[x\ y\ z]^T \mapsto M[x\ y\ z]^T \tag{A.5}$$

This definition has the virtue at least of being unambiguous because

$$[cx\ cy\ cz]^T \mapsto M[cx\ cy\ cz]^T = c(M[x\ y\ z]^T)$$

which represents the same point as $M[x\ y\ z]^T$, implying that the choice of any homogeneous coordinates for P gives the same image by the transformation.

The potential glitch to consider before putting (A.5) into production is if it maps a non-zero tuple to a zero tuple, for then it would map the homogeneous coordinates of a point $P \in \mathbb{P}^2$ to a value not even belonging to $\mathbb{P}^2$. However, we know from basic linear algebra that there is a non-zero tuple $[x\ y\ z]^T$ such that

$$M[x\ y\ z]^T = [0\ 0\ 0]^T$$

if and only if M is a singular matrix; otherwise, M maps non-zero tuples to non-zero tuples. We conclude that defining a transformation of $\mathbb{P}^2$ by (A.5) is indeed valid provided M is non-singular. Ergo:

Definition A.9. If M is a non-singular 3×3 matrix, then the transformation

$$[x\ y\ z]^T \mapsto M[x\ y\ z]^T$$

denoted h^M, is called a *projective transformation* of the projective plane. The transformation f^M of $\mathbb{R}^3$ – the linear transformation defined by M – is called a *related* linear transformation.

A simple relation between h^M and f^M is the following: if the radial line corresponding to a point P of $\mathbb{P}^2$ is l, then that corresponding to $h^M(P)$ is $f^M(l)$, the image of l by f^M.

Exercise A.26. Prove that if M is a non-singular 3×3 matrix and c is a scalar such that $c \neq 0$, then M and cM define the same projective transformation of $\mathbb{P}^2$, i.e., $h^M = h^{cM}$.

Remark A.8. The preceding exercise implies that actually there is not a unique linear transformation related to a projective transformation h^M, because f^{cM} is related to $h^{cM} = h^M$, for any non-zero c. However, when we do have a specific M that we are using to define h^M, then we'll often speak of *the* related linear transformation f^M.

Exercise A.27. Prove that a projective transformation h^M of $\mathbb{P}^2$ takes projective lines to projective lines.
Hint: The related (non-singular) linear transformation f^M takes radial planes in $\mathbb{R}^3$ to radial planes in $\mathbb{R}^3$.

Exercise A.28. Prove that the composition $h^M \circ h^N$ of two projective transformations of $\mathbb{P}^2$ is equal to the projective transformation h^{MN}.

A.9 Projective Transformations Geometrically

Our definition of projective transformations was purely algebraic. We would like to picture, if possible, how they transform primitives in $\mathbb{P}^2$. Now, projective primitives are "seen" by projection onto the real plane – by capture on a point camera's film as we've been putting it. Let's find out, then, what a projective transformation looks like through a point camera.

Here's what we plan to do. Start with a primitive s, on the plane $z = 1$, our favorite copy of $\mathbb{R}^2$, as the designated film. Suppose that the given projective transformation is h^M. Then we'll transform the lifting S of s by h^M to $h^M(S)$. Finally, we'll project $h^M(S)$ back to $z = 1$ to obtain a new primitive s'. It's precisely the change from s to s' which is seen as the transformation h^M by a point camera at the origin. For example, in Figure A.18, a boxy car is changed (fancifully) into a sleek convertible.

Back to reality, let's begin with a simple example. Consider a straight segment s joining two points p and q on $z = 1$. Given a projective transformation h^M, we want to determine s'. The lifting S of s, which is the set of all radial lines intersecting s, is

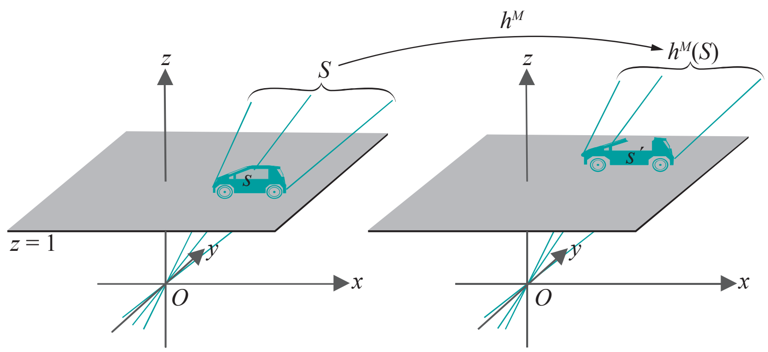

Figure A.18: Projective transformation of a car (purely conceptual!).

not hard to visualize: it forms an "infinite double triangle" which lies on the radial plane containing s and the origin. See Figure A.19(a). The radial lines through p and q are denoted P and Q, respectively.

The related linear transformation f^M transforms s to a segment $\overline{s} = \overline{p}\,\overline{q}$, where $f^M(p) = \overline{p}$ and $f^M(q) = \overline{q}$. See Figure A.19(b). Note that $\overline{s}$ can be anywhere in 3-space, depending on f^M, unlike s and s', which are both on $z = 1$.

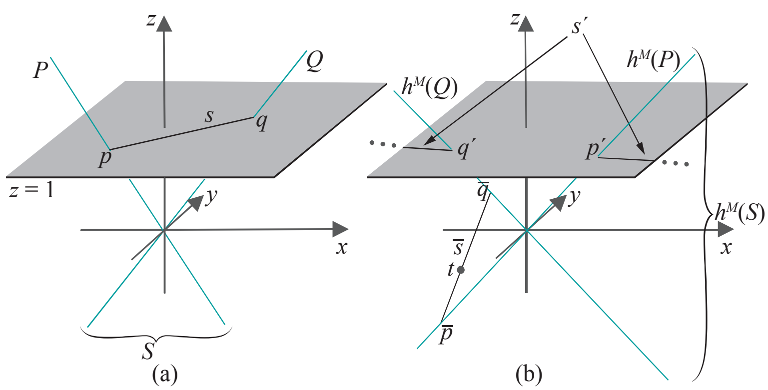

Figure A.19: (a) A segment s on $\mathbb{R}^2$ and its lifting S (b) f^M transforms s to $\overline{s}$ and S to $h^M(S)$, while s' is the intersection of $h^M(S)$ with $z = 1$.

Moreover, each radial line in S, the lifting of s, is transformed by f^M to a radial line in $h^M(S)$. Each radial line in $h^M(S)$, of course, intersects $\overline{s}$. A diagram depicting a particular disposition of $\overline{s}$, where it intersects the xy-plane in a single point t, is shown in Figure A.19(b).

The transformed primitive s' is the intersection of the radial lines in $h^M(S)$ with $z = 1$. At this time we ask the reader to complete the following exercise to find out for herself what it looks like, depending on the situation of $\overline{s}$.

Exercise A.29. Show that exactly one of (a)-(c) is true:

(a) $\overline{s}$ does not intersect the xy-plane, equivalently, every radial line in $h^M(S)$ is a regular point with respect to $z = 1$.

In this case, s' is the segment between the points p' and q' where $h^M(P)$ and $h^M(Q)$, respectively, intersect $z = 1$ (remember that P and Q are the radial lines through p and q, the endpoints of s, respectively). Sketch this case.

(b) $\overline{s}$ intersects the xy-plane at one point, equivalently, exactly one radial line in $h^M(S)$ is a point at infinity with respect to $z = 1$. Now, there are two subcases:

(b1) If the intersection point, call it t, is in the interior of $\overline{s}$, then s' consists of the *entire* infinite straight line through p' and q', where $h^M(P)$ and $h^M(Q)$, respectively, intersect $z = 1$, *minus* the finite open segment between p' and q'. This situation is sketched in Figure A.19(b).

(b2) If the intersection is an endpoint of $\overline{s}$, say $\overline{p}$, then s' is a straight line infinite in one direction and with an endpoint at q', where $h^M(Q)$ intersects $z = 1$, in the other. Sketch this case.

(c) $\overline{s}$ lies on the xy-plane, equivalently, every radial line in $h^M(S)$ is a point at infinity with respect to $z = 1$.

In this case, s' is empty.

The answer to the preceding exercise is not tidy, but in most practical situations it will be case (a), the most benign of the three, which applies.

So we know now what we set out to find: how the projective transformation of the lifting of a segment looks like on film. Generally, for any primitive s on the plane, if s' is the "film-capture" of the transformation by h^M of the lifting of s, we'll call s' the *projective transformation* of s by h^M, and denote it $h^M(s)$ – giving thus a geometric counterpart of the algebraic definition of a projective transformation in Section A.8. Although h^M is well-defined, it is not a transformation of $\mathbb{R}^2$ in general because $h^M(p)$ may not even exist for a point $p \in \mathbb{R}^2$, particularly if p's corresponding projective point is taken by h^M to a point at infinity (which has no film-capture).

In our usage, therefore, h^M can represent either a transformation of projective space (as defined in Section A.8) or a transformation of real primitives (as just defined above). There is no danger of ambiguity as the nature of the argument in $h^M(*)$ will make clear how it's being used.

Example A.11. The segment s joins $p = [1\ \ -1]^T$ and $q = [-2\ \ -2]^T$ on the plane $z = 1$, the latter identified with $\mathbb{R}^2$. The projective transformation $h^M : \mathbb{P}^2 \to \mathbb{P}^2$ is specified by

$$M = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

which is the matrix corresponding to a rotation f^M of $\mathbb{R}^3$ by $90°$ about the y-axis, clockwise when seen from the positive side of the y-axis. Determine $h^M(s)$.

Answer: f^M transforms s to the segment $\overline{s} = \overline{p}\,\overline{q}$, where $\overline{p}$ and $\overline{q}$ are the images by f^M of p and q, respectively. Multiplying p and q, written as points of $z = 1$, on the left by M we get:

$$\overline{p} = M[1\ \ -1\ \ 1]^T = [-1\ \ -1\ \ 1]^T$$

and

$$\overline{q} = M[-2\ \ -2\ \ 1]^T = [-1\ \ -2\ \ -2]^T$$

As the z-values of $\overline{p}$ and $\overline{q}$ are of different signs, an interior point of $\overline{s}$ lies on the xy-plane. Therefore, we are in case (b1) of Exercise A.29 above.

Let P and Q denote the radial lines through p and q, respectively. The radial line $h^M(P)$ through $\overline{p}$ meets $z = 1$ at $h^M(p) = [-1\ \ -1\ \ 1]^T$, which is $\overline{p}$ itself. The radial line $h^M(Q)$ through $\overline{q}$ meets $z = 1$ at $h^M(q) = [\frac{1}{2}\ \ 1\ \ 1]^T$ (multiplying the coordinate tuple of $\overline{q}$ by $-\frac{1}{2}$ to make its z-value equal to 1).

Applying Exercise A.29 case (b1), $h^M(s)$ is the entire straight line through the points $[-1\ \ -1]^T$ and $[\frac{1}{2}\ \ 1]^T$ minus the finite open segment joining $[-1\ \ -1]^T$ to $[\frac{1}{2}\ \ 1]^T$.

Example A.12. The rectangle r lies on the plane $z = 1$, the latter identified with $\mathbb{R}^2$. Its vertices are $p_1 = [0.5\ \ 1]^T$, $p_2 = [0.5\ \ -1]^T$, $p_3 = [1\ \ -1]^T$ and $p_4 = [1\ \ 1]^T$. See Figure A.20. Determine $h^M(r)$, where h^M is the same projective transformation as in the preceding example.

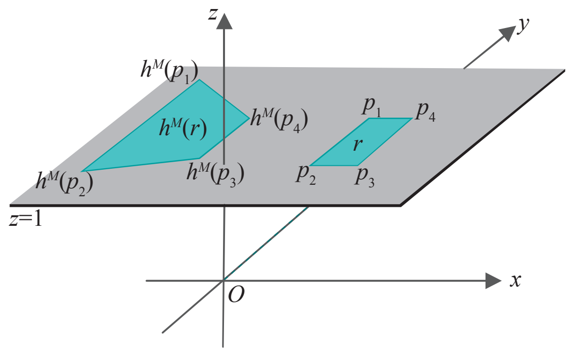

Figure A.20: Rectangle r is transformed to the trapezoid $h^M(r)$.

Answer: f^M transforms r to the rectangle $\overline{r}$ with vertices $\overline{p_i} = f^M(p_i)$, $1 \leq i \leq 4$. Multiplying each p_i, written as points of $z = 1$, on the left by M we get:

$$\begin{array}{rcl} \overline{p_1} & = & M[0.5\ 1\ 1]^T \ = \ [-1\ 1\ 0.5]^T \\ \overline{p_2} & = & M[0.5\ -1\ 1]^T \ = \ [-1\ -1\ 0.5]^T \\ \overline{p_3} & = & M[1\ -1\ 1]^T \ = \ [-1\ -1\ 1]^T \\ \overline{p_4} & = & M[1\ 1\ 1]^T \ = \ [-1\ 1\ 1]^T \end{array}$$

As the z-value of every $\overline{p_i}$, $1 \leq i \leq 4$, is greater than 0, none of the edges of $\overline{r}$ intersects the xy-plane. According to case (a) of Exercise A.29 then, $h^M(r)$ is the quadrilateral with vertices at the points $h^M(p_i)$, where the radial lines through $\overline{p_i}$, $1 \leq i \leq 4$, intersect $z = 1$. See Figure A.20. Multiply the coordinate tuple of each $\overline{p_i}$ by a scalar to make its z-value equal to 1, to find that

$$\begin{array}{rcl} h^M(p_1) & = & [-2\ 2\ 1]^T \\ h^M(p_2) & = & [-2\ -2\ 1]^T \\ h^M(p_3) & = & [-1\ -1\ 1]^T \\ h^M(p_4) & = & [-1\ 1\ 1]^T \end{array}$$

One sees, therefore, that $h^M(r)$ has vertices at $[-2\ 2]^T$, $[-2\ -2]^T$, $[-1\ -1]^T$ and $[-1\ 1]^T$, which makes it a trapezoid.

It's interesting to note that no affine transformation of $\mathbb{R}^2$ can map a rectangle to a trapezoid: as affine transformations preserve parallelism (see Proposition 5.1), at most they can transform a rectangle to a parallelogram.

Exercise A.30. Exercise A.7(f), where we snapshot transformed a trapezoid to a rectangle, evidently is related to the preceding example. Say how.

Clearly, with the help of Exercise A.29 we can determine the projective transformation of any shape specified by straight edges. More general shapes are curved and curves specified by equations. Let's see, for example, how a parabola is projectively transformed.

Example A.13. Determine how the parabola $y - x^2 = 0$ on $z = 1$, the latter identified with $\mathbb{R}^2$, is mapped by the same projective transformation h^M as in the previous example.

Answer: The point $[x\ y]^T$ on $z = 1$, which has coordinates $[x\ y\ 1]^T$ in $\mathbb{R}^3$, is transformed by f^M to the point $[\overline{x}\ \ \overline{y}\ \ \overline{z}]^T$, where

$$\begin{bmatrix} \overline{x} \\ \overline{y} \\ \overline{z} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ y \\ x \end{bmatrix}$$

which gives

$$\overline{x} = -1, \qquad \overline{y} = y, \qquad \overline{z} = x$$

The image $[x' \ y']^T$ of $[x \ y]^T$ by h^M, then, is the point $[\overline{x}/\overline{z} \ \ \overline{y}/\overline{z}]^T$, where the radial line through $[\overline{x} \ \overline{y} \ \overline{z}]^T$ intersects $z = 1$. Therefore:

$$x' = \overline{x}/\overline{z} = -1/x \quad \Longrightarrow \quad x = -1/x'$$

and

$$y' = \overline{y}/\overline{z} = y/x \quad \Longrightarrow \quad y = y'x = -y'/x' \ \ \text{(using } x = -1/x' \text{ from above)}$$

Plugging these expressions for x and y into the equation of the parabola $y - x^2 = 0$, we have the equation

$$-y'/x' - 1/x'^2 = 0, \text{ equivalently,} \quad x'y' + 1 = 0$$

of the transformed curve, which describes a hyperbola.

Here's another rather interesting example.

Example A.14. Determine how points of $\mathbb{R}^2$, identified with $z = 1$, are transformed by the projective transformation h^M of $\mathbb{P}^2$ specified by

$$M = \begin{bmatrix} 1 & 0 & 7 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Answer: The point $[x \ y]^T$ on $z = 1$, which has coordinates $[x \ y \ 1]^T$ in $\mathbb{R}^3$, is transformed by f^M to the point $[\overline{x} \ \overline{y} \ \overline{z}]^T$, where

$$\begin{bmatrix} \overline{x} \\ \overline{y} \\ \overline{z} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 7 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x+7 \\ y \\ 1 \end{bmatrix}$$

giving

$$\overline{x} = x + 7, \qquad \overline{y} = y, \qquad \overline{z} = 1$$

The image $[x' \ y']^T$ of $[x \ y]^T$ by h^M, then, is the point $[\overline{x}/\overline{z} \ \ \overline{y}/\overline{z} \ \ 1]^T$, where the radial line through $[\overline{x} \ \overline{y} \ \overline{z}]^T$ intersects $z = 1$. Therefore,

$$x' = \overline{x}/\overline{z} = x + 7 \quad \text{and} \quad y' = \overline{y}/\overline{z} = y$$

which is nothing but a *translation* by 7 units in the x-direction.

Incidentally, we did not pull the matrix M above out of a hat: it is the transformation matrix of a 3D shear whose plane is the xy-plane and line the x-axis (recall 3D shears from Section 5.4).

A projection transformation has just done something beyond the reach of linear transformations, for a linear transformation cannot translate. Translations, as we learned in Chapter 5, are in the domain of affine transformations. Further, in Example A.12, we saw a projective transformation convert a rectangle into a trapezoid, something beyond even affine transformations. For transformations inspired by and defined by matrix-vector multiplication, just like linear transformations, projective transformations certainly seem to carry plenty of additional firepower. It turns out that this makes them particularly worthy allies in the advancement of computer graphics.

Exercise A.31. Find a projective transformation to translate points of $\mathbb{R}^2$ 3 units in the x-direction and 2 in the y-direction, i.e., whose displacement vector is $[3 \ 2]^T$. *Hint*: Think another shear.

Exercise A.32. Determine how the segment s on $\mathbb{R}^2$, the latter identified with the plane $z = 1$, joining $p = [2 \ \ -2]^T$ and $q = [-2 \ 1]^T$, is mapped by the projective transformation h^M of $\mathbb{P}^2$ specified by

$$M = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Exercise A.33. Determine how the hyperbola $xy = 1$ on $z = 1$, the latter identified with $\mathbb{R}^2$, is mapped by the same projective transformation h^M as in the previous exercise.

Part answer: The problem is not hard but there is a fair amount of manipulation.

The point $[x \ y]^T$ on $z = 1$, which has coordinates $[x \ y \ 1]^T$ in $\mathbb{R}^3$, is transformed by f^M to the point $[\overline{x} \ \overline{y} \ \overline{z}]^T$, where

$$\begin{bmatrix} \overline{x} \\ \overline{y} \\ \overline{z} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Let's flip this equation over with the help of an inverse matrix:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \overline{x} \\ \overline{y} \\ \overline{z} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} \overline{x} \\ \overline{y} \\ \overline{z} \end{bmatrix}$$

which gives

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

Plugging these expressions into the equation of the hyperbola $xy = 1 = 1^2$ we get:

$$\frac{1}{4}(-\overline{x} + \overline{y} + \overline{z})(\overline{x} - \overline{y} + \overline{z}) = \frac{1}{4}(\overline{x} + \overline{y} - \overline{z})^2$$

Now, the image $[x' \ y']^T$ of $[x \ y]^T$ by h^M is the point $[\overline{x}/\overline{z} \ \ \overline{y}/\overline{z}]^T$, where the radial line through $[\overline{x} \ \overline{y} \ \overline{z}]^T$ intersects $z = 1$. We ask the reader to complete the exercise by dividing the preceding equation by $\overline{z}^2$ throughout to obtain an equation relating x' and y', and identifying the corresponding curve.

Exercise A.34. Determine how the straight line $x + y + 1 = 0$ on $z = 1$ is mapped by the same projective transformation h^M as in the previous exercise.

Exercise A.35. We saw in Example 5.4 that affine transformations preserve convex combinations and barycentric coordinates. Show that projective transformations in general do not.

Remark A.9. Projective transformations of $\mathbb{P}^2$ can be thought of as a powerful class of *pseudo-transformations* of $\mathbb{R}^2$ – pseudo because a projective transformation may map a regular point to a point at infinity, in which case the corresponding point of $\mathbb{R}^2$ has no valid image. If one is careful, however, to restrict its domain to a region of $\mathbb{R}^2$ where it *is* valid throughout, one may be able to exploit the ability of a projective transformation to do more than an affine one.

A.10 Relating Projective, Snapshot and Affine Transformations

We'll explore in this section the inter-relationships between projective, snapshot and affine transformations.

A.10.1 Snapshot Transformations via Projective Transformations

Snapshot transformations, being transformations of an object seen through a point camera as the film changes alignment, are geometrically intuitive. They are, in fact, a kind of projective transformation, as we'll now see.

Consider again Example A.12 for motivation. We saw that the rectangle r on the plane $z = 1$ (aka $\mathbb{R}^2$) with vertices at $p_1 = [0.5\ 1]^T$, $p_2 = [0.5\ \ -1]^T$, $p_3 = [1\ \ -1]^T$ and $p_4 = [1\ 1]^T$ is mapped to the trapezoid $r' = h^M(r)$ with vertices at $p'_1 = [-2\ 2]^T$, $p'_2 = [-2\ \ -2]^T$, $p'_3 = [-1\ \ -1]^T$ and $p'_4 = [-1\ 1]^T$, by the projective transformation h^M specified by

$$M = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

See Figure A.21(a).

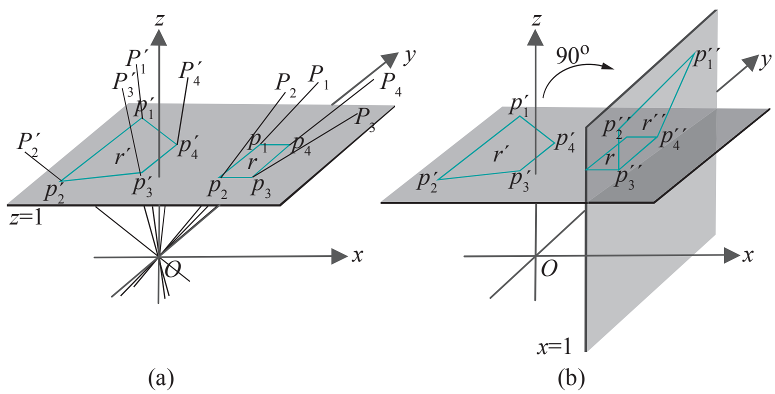

Figure A.21: (a) Projective transformation h^M maps rectangle r to trapezoid $r' = h^M(r)$ (b) r' is the "same" as r'', the picture of r captured on a film along $x = 1$.

We observed, as well, that the related linear transformation f^M is a rotation of $\mathbb{R}^3$ by 90° about the y-axis, which is clockwise when seen from the positive side of the y-axis.

Denote the radial line through p_i by P_i, $1 \leq i \leq 4$, and their respective images $h^M(P_i)$ by P'_i. Now rotate the radial lines P'_i, *as well as* the plane $z = 1$, an angle of 90° about the y-axis, this time counter-clockwise when seen from the positive side of the y-axis, in order to undo the effect of f^M; in other words, apply $f^{M^{-1}}$. We see the following:

(a) The radial line P'_i, of course, rotates back onto (its pre-image) the radial line P_i, $1 \leq i \leq 4$.

(b) The plane $z = 1$ is taken by the rotation onto the plane $x = 1$.

(c) The trapezoid r', as a consequence of (a) and (b), rotates onto a trapezoid r'' with vertices at the intersections p''_i of P_i with $x = 1$, for $1 \leq i \leq 4$. See Figure A.21(b) (note that the edge of r that happens to lie on the intersection of the planes $z = 1$ and $x = 1$ is shared with r'').

But r'' is precisely the snapshot transformation of r from the film along $z = 1$ to the one along $x = 1$! Here's what is happening. The image r' is obtained by applying the rotation f^M to the radials P_i and intersecting them with the plane $z = 1$, while r'' is obtained from r' by applying the reverse rotation $f^{M^{-1}}$, which takes the radials back to the where they were, and, at the same time, changes the intersecting plane from $z = 1$ to $x = 1$. Therefore, the transformation from r to r'' comes from a change in the plane (= film) intersecting the radials, which is precisely a snapshot transformation.

One sees, therefore, that, generally, a snapshot transformation in which the film is re-aligned by a rotation f about a radial axis is equivalent to a projective transformation whose related linear transformation is f^{-1}, in that the images are identical, though situated differently in space (precisely, the two images differ by a rigid transformation of $\mathbb{R}^3$). But, how about snapshot transformations where the new alignment of the film cannot be obtained from the original by mere rotation? To answer this question, we ask the reader, first, to prove the following, which says that an arbitrary snapshot transformation can be composed from two very simple ones.

Exercise A.36. Prove that any plane p in $\mathbb{R}^3$ can be aligned with any other p' by a translation parallel to itself followed by a rotation about a radial axis.

Therefore, any snapshot transformation is the composition of two: first, a snapshot transformation from one film to a parallely translated one and then another, where one film is obtained from the other by a rotation about a radial axis.
Hint: See Figure A.22.

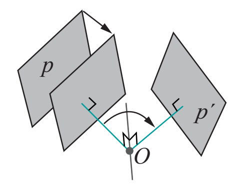

Figure A.22: Aligning plane p with p' by a parallel displacement, so that their respective distances from the origin are equal, followed by a rotation.

We have already seen how a snapshot transformation from one film to a rotated one is equivalent to a projective transformation. A snapshot transformation to a parallely translated one is also equivalent to a projective transformation, as the next exercise asks the reader to show.

Exercise A.37. Suppose that two parallel non-radial planes p and p' in $\mathbb{R}^3$ are at a distance of c and c' from the origin, respectively. Then the snapshot transformation from p to p' is equivalent to the projective transformation h^M, where

$$M = \begin{bmatrix} \frac{c'}{c} & 0 & 0 \\ 0 & \frac{c'}{c} & 0 \\ 0 & 0 & \frac{c'}{c} \end{bmatrix} = \frac{c'}{c} I$$

(i.e., a projective transformation whose related linear transformation is a *uniform* scaling of $\mathbb{R}^3$ by a factor of $\frac{c'}{c}$ in all directions).
Hint: See Figure A.23.

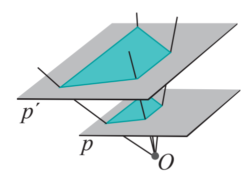

Figure A.23: A snapshot transformation to a parallel plane is equivalent to a scaling by a constant factor in all directions.

Putting the pieces together we have the following proposition:

Proposition A.1. *A snapshot transformation k from a non-radial plane p in $\mathbb{R}^3$ to another p' is equivalent to a projective transformation h^M of $\mathbb{P}^2$, in the sense that the images of primitives by k and h^M are identical modulo a rigid transformation of $\mathbb{R}^3$.*

In particular, k is equivalent to the projective transformation h^M which is the composition of a projective transformation h^{dI}, whose related linear transformation is a uniform scaling, with a projective transformation h^N, whose related linear transformation is a rotation of $\mathbb{R}^3$ about a radial axis.

In other words, k is equivalent to h^{dN}, where d is a scalar and N is the matrix of a rotation of $\mathbb{R}^3$ about a radial axis. □

Exercise A.38. Determine the projective transformation equivalent to the snapshot transformation from the plane $z = 1$ to the plane $x = 2$.

A.10.2 Affine Transformations via Projective Transformations

We begin by asking if there exist projective transformations of $\mathbb{P}^2$ that respect regular points, i.e., map regular points to regular points. Such a transformation could then be *entirely* captured on film because it takes no point of the film out of it, as would happen, say, if a regular point were mapped to one at infinity. Looking back at

Remark A.9, one could then say that such a projective transformation is no longer "pseudo", but a true transformation of $\mathbb{R}^2$.

So suppose the film lies along the plane (surprise) $z = 1$. What condition must a projective transformation h^M, where

$$M = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

satisfy in order to transform each point regular with respect to $z = 1$ to another such? Homogeneous coordinates of regular points are of the form $[x\ y\ 1]^T$. Now,

$$h^M([x\ y\ 1]^T) = M[x\ y\ 1]^T = \begin{bmatrix} a_{11}x + a_{12}y + a_{13} \\ a_{21}x + a_{22}y + a_{23} \\ a_{31}x + a_{32}y + a_{33} \end{bmatrix}$$

For this image point to be regular we must have

$$a_{31}x + a_{32}y + a_{33} \neq 0$$

However, if either one of a_{31} and a_{32} is non-zero, or if a_{33} is zero, then it's possible to find values of x and y such that $a_{31}x + a_{32}y + a_{33} = 0$. The conclusion then is that for h^M to transform all regular points to regular points, one must have both a_{31} and a_{32} equal to zero and a_{33} non-zero. Therefore, M must be of the form

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$$

with $a_{33} \neq 0$. By Exercise A.26, M can be multiplied by $1/a_{33}$ to still represent the same projective transformation, so one can assume $a_{33} = 1$, implying that the form of M is

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

In this case, h^M transforms $[x\ y\ 1]^T$ to

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11}x + a_{12}y + a_{13} \\ a_{21}x + a_{22}y + a_{23} \\ 1 \end{bmatrix}$$

Tossing the last coordinate, it transforms $[x\ y]^T \in \mathbb{R}^2$ to

$$\begin{bmatrix} a_{11}x + a_{12}y + a_{13} \\ a_{21}x + a_{22}y + a_{23} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}$$

which is precisely an affine transformation!

We conclude that a projective transformation of $\mathbb{P}^2$ that respects regular points gives nothing but an affine transformation of $\mathbb{R}^2$. Conversely, it's not hard to see that any affine transformation of $\mathbb{R}^2$ can be obtained as a projective transformation preserving regular points. We record these facts in the following proposition.

Proposition A.2. *An affine transformation of $\mathbb{R}^2$ is equivalent to a projective transformation of $\mathbb{P}^2$, in particular, one that respects regular points.*

Conversely, a projective transformation of $\mathbb{P}^2$ that respects regular points is equivalent to an affine transformation of $\mathbb{R}^2$.

Evidently, the constraint to respect regular points is a burden on projective transformations. It dumbs them down to affine and all the excitement of parallel lines turning into intersecting ones, rectangles into trapezoids, and circles into hyperbolas is lost!

However, one does see now a good reason for the use of homogeneous coordinates of real points in computing affine transformations. When first we did this in Section 5.2.3, it seemed merely a neat maneuver to obtain an affine transformation as a single matrix-vector multiplication. The bigger picture is that affine transformations are a subclass of the projective. Therefore, as the latter are obtained (by definition) from matrix-vector multiplication, so can the former, provided we relocate to projective space, in other words, use homogeneous coordinates.

A Roundup of the Three Kinds of Transformations

Snapshot and affine transformations are subclasses of the projective, as we have just seen. How about the relationship between these two subclasses themselves? Are snapshot transformations affine or affine transformations snapshot?

At the start of Section A.10.1 we saw a projective transformation, equivalent, in fact, to a snapshot transformation, map a rectangle to a trapezoid. This is not possible for an affine transformation to do, as it is obliged to preserve parallelism (Proposition 5.1). Therefore, snapshot transformations are certainly not all affine.

A shear on the plane, an affine transformation, can map a rectangle to a non-rectangular parallelogram. We leave the reader to convince herself that this is not possible for a snapshot transformation. So not all affine transformations are snapshot.

We see then that neither of the two subclasses, snapshot and affine, of projective transformations contains the other. However, what transformations, if any, do they have in common? We ask the reader herself to characterize the transformations at the intersection of affine and snapshot in the next exercise.

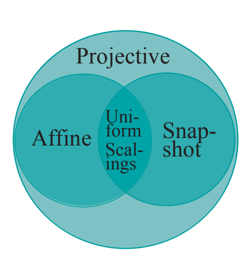

Figure A.24: Venn diagram of transformation classes of $\mathbb{R}^2$.

Exercise A.39. Prove that projective transformations which are both affine and snapshot are precisely those whose related linear transformation is a uniform scaling.

The final important question on the relationship between the three classes is if the union of snapshot and affine covers projective transformations or if the latter is strictly bigger. In Exercise A.42 in the next section we'll see an example of a projective transformation neither snapshot nor affine. Therefore, indeed, the class of projective transformations is strictly bigger than the union of snapshot and affine. Figure A.24 summarizes the relationship between the three classes.

A.11 Designer Projective Transformations

We know from elementary linear algebra that a linear transformation is uniquely specified by defining its values on a basis. Here's a like-minded claim for projective transformations of $\mathbb{P}^2$.

Proposition A.3. *If two sets* $\{P_1, P_2, P_3, P_4\}$ *and* $\{Q_1, Q_2, Q_3, Q_4\}$ *of four points each from* $\mathbb{P}^2$ *are such that no three in any one set are collinear, then there is a unique projective transformation of* $\mathbb{P}^2$ *that maps* P_i *to* Q_i*, for* $1 \le i \le 4$.

Proof. Choose non-zero vectors p_1, p_2, p_3 and p_4 from $\mathbb{R}^3$ lying on P_1, P_2, P_3 and P_4, respectively, and non-zero vectors q_1, q_2, q_3 and q_4 lying on Q_1, Q_2, Q_3 and Q_4, respectively.

Since P_1, P_2 and P_3 do not lie on one projective line, p_1, p_2 and p_3 do not lie on one radial plane. The latter three form, therefore, a basis of $\mathbb{R}^3$. Likewise, q_1, q_2 and q_3 form a basis of $\mathbb{R}^3$ as well.

Let c_1, c_2 and c_3 be arbitrary scalars, all three non-zero, whose values will be determined. As q_1, q_2 and q_3 form a basis of $\mathbb{R}^3$, so do c_1q_1, c_2q_2 and c_3q_3. Therefore, there is a unique non-singular linear transformation $f^M : \mathbb{R}^3 \to \mathbb{R}^3$ such that

$$f^M(p_i) = c_i q_i, \text{ for } 1 \le i \le 3$$

which, then, is related to a projective transformation $h^M : \mathbb{P}^2 \to \mathbb{P}^2$, such that

$$h^M(P_i) = Q_i, \text{ for } 1 \le i \le 3$$

It remains to make $h^M(P_4) = Q_4$.

As p_1, p_2 and p_3 form a basis of $\mathbb{R}^3$, there exist unique scalars α, β and γ such that

$$p_4 = \alpha p_1 + \beta p_2 + \gamma p_3$$

Now, α, β and γ are all three non-zero, for, otherwise, p_4 lies on the same radial plane as two of p_1, p_2 and p_3, which implies that P_4 lies on the same projective line as two of P_1, P_2 and P_3, contradicting an initial hypothesis. Likewise, there exist unique non-zero scalars, λ, μ and ν such that

$$q_4 = \lambda q_1 + \mu q_2 + \nu q_3$$

For

$$h^M(P_4) = Q_4$$

to hold, then, one requires a scalar $c_4 \neq 0$ such that

$$\begin{aligned} f^M(p_4) &= c_4 q_4 \\ &= c_4(\lambda q_1 + \mu q_2 + \nu q_3) \\ &= \lambda c_4 q_1 + \mu c_4 q_2 + \nu c_4 q_3 \end{aligned} \tag{A.6}$$

However,

$$\begin{aligned} f^M(p_4) &= f^M(\alpha p_1 + \beta p_2 + \gamma p_3) \\ &= \alpha f^M(p_1) + \beta f^M(p_2) + \gamma f^M(p_3) \\ &= \alpha c_1 q_1 + \beta c_2 q_2 + \gamma c_3 q_3 \end{aligned} \tag{A.7}$$

Combining (A.6) and (A.7) one has

$$\alpha c_1 q_1 + \beta c_2 q_2 + \gamma c_3 q_3 = \lambda c_4 q_1 + \mu c_4 q_2 + \nu c_4 q_3$$

As q_1, q_2 and q_3 is a basis of $\mathbb{R}^3$, it follows that

$$\alpha c_1 = \lambda c_4, \quad \beta c_2 = \mu c_4, \quad \gamma c_3 = \nu c_4$$

giving

$$c_1 = (\lambda/\alpha)c_4, \quad c_2 = (\mu/\beta)c_4, \quad c_3 = (\nu/\gamma)c_4$$

determining c_1, c_2, c_3 and c_4 uniquely, up to a constant of proportionality, so completing the proof. □

The following corollary, which is a straightforward application of the proposition, is particularly important.

Corollary A.1. *Any non-degenerate quadrilateral, i.e., one with no three collinear vertices, in* $\mathbb{R}^2$ *can be projectively transformed to any other such.* □

More than just theoretically, the proposition is important in that it suggests how to go about finding projective transformations specified at only a few points.

Example A.15. Determine the projective transformation h^M of $\mathbb{P}^2$ mapping the projective points

$$P_1 = [1\ 0\ 0]^T,\ P_2 = [0\ 1\ 0]^T,\ P_3 = [0\ 0\ 1]^T \text{ and } P_4 = [1\ 1\ 1]^T$$

to the respective images

$$Q_1 = [2\ 1\ 3]^T,\ Q_2 = [-1\ \ -1\ 1]^T,\ Q_3 = [0\ 1\ 1]^T \text{ and } Q_4 = [0\ 0\ 6]^T$$

Answer: Choose (not particularly imaginatively)

$$p_1 = [1\ 0\ 0]^T,\ p_2 = [0\ 1\ 0]^T,\ p_3 = [0\ 0\ 1]^T \text{ and } p_4 = [1\ 1\ 1]^T$$

in $\mathbb{R}^3$ lying on P_i, $1 \le i \le 4$, and

$$q_1 = [2\ 1\ 3]^T,\ q_2 = [-1\ \ -1\ 1]^T,\ q_3 = [0\ 1\ 1]^T \text{ and } q_4 = [0\ 0\ 6]^T$$

lying on Q_i, $1 \le i \le 4$.

The linear transformation $f^M : \mathbb{R}^3 \to \mathbb{R}^3$ such that $f^M(p_i) = c_i q_i$, for $1 \le i \le 3$, where c_1, c_2 and c_3 are non-zero scalars, is easily verified to be given by

$$M = \begin{bmatrix} 2c_1 & -c_2 & 0 \\ c_1 & -c_2 & c_3 \\ 3c_1 & c_2 & c_3 \end{bmatrix}$$

One can verify as well that

$$p_4 = p_1 + p_2 + p_3$$

and

$$q_4 = q_1 + 2q_2 + q_3$$

Therefore,

$$f^M(p_4) = f^M(p_1 + p_2 + p_3) = f^M(p_1) + f^M(p_2) + f^M(p_3) = c_1q_1 + c_2q_2 + c_3q_3$$

Accordingly, if $f^M(p_4) = c_4 q_4$, for some $c_4 \ne 0$, then

$$c_1q_1 + c_2q_2 + c_3q_3 = c_4(q_1 + 2q_2 + q_3) = c_4q_1 + 2c_4q_2 + c_4q_3$$

which implies that

$$c_1 = c_4, \quad c_2 = 2c_4, \quad c_3 = c_4$$

Setting $c_4 = 1$, one has: $c_1 = 1,\ c_2 = 2,\ c_3 = 1,\ c_4 = 1$. One concludes that the required projective transformation h^M is given by

$$M = \begin{bmatrix} 2 & -2 & 0 \\ 1 & -2 & 1 \\ 3 & 2 & 1 \end{bmatrix}$$

The following example will help in an application of projective transformations in the graphics pipeline.

Example A.16. Determine the projective transformation h^M of $\mathbb{P}^2$ that transforms the trapezoid q on the plane $z = 1$ (aka $\mathbb{R}^2$) with vertices at

$$p_1 = [-1\ 1]^T,\ p_2 = [1\ 1]^T,\ p_3 = [2\ 2]^T \text{ and } p_4 = [-2\ 2]^T$$

to the rectangle q' on the same plane with vertices at

$$p_1' = [-1\ 1]^T,\ p_2' = [1\ 1]^T,\ p_3' = [1\ 2]^T \text{ and } p_4' = [-1\ 2]^T$$

See Figure A.25.

Answer: Suppose that the required projective transformation h^M is defined by the matrix

$$M = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

We have to determine the a_{ij} up to a non-zero multiplicative constant.

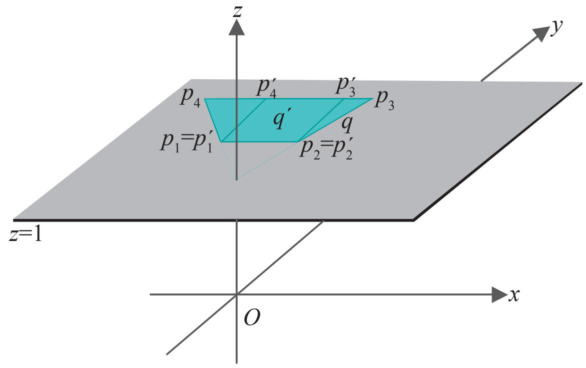

Figure A.25: Transforming the trapezoid q on $z = 1$ to the rectangle (bold) q'.

The two sides p_1p_4 and p_2p_3 of the trapezoid q meet at the regular point (with respect to $z = 1$) $[0\ 0\ 1]^T$, while the corresponding sides $p_1'p_4'$ and $p_2'p_3'$ of the rectangle q' are parallel and meet at the point at infinity $[0\ 1\ 0]^T$. The transformation must, therefore, map $[0\ 0\ 1]^T$ to $[0\ 1\ 0]^T$, yielding our first equation

$$h^M([0\ 0\ 1]^T) = [0\ 1\ 0]^T$$

(the RHS could be $c[0\ 1\ 0]^T$ for any non-zero scalar c, but there's no loss in assuming that $c = 1$) which translates to the matrix equation

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

giving

$$a_{13} = 0, \quad a_{23} = 1, \quad a_{33} = 0$$

So we write

$$M = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 1 \\ a_{31} & a_{32} & 0 \end{bmatrix}$$

That we have $h^M(p_1) = p_1'$ and $h^M(p_2) = p_2'$ gives two more matrix equations

$$\begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 1 \\ a_{31} & a_{32} & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -c \\ c \\ c \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 1 \\ a_{31} & a_{32} & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} d \\ d \\ d \end{bmatrix}$$

where c and d are arbitrary non-zero scalars, yielding the six equations

$$\begin{aligned} -a_{11} + a_{12} &= -c \\ -a_{21} + a_{22} + 1 &= c \\ -a_{31} + a_{32} &= c \\ a_{11} + a_{12} &= d \\ a_{21} + a_{22} + 1 &= d \\ a_{31} + a_{32} &= d \end{aligned} \tag{A.8}$$

Subtracting the first equation from the fourth, adding the second and fifth, and adding the third and sixth, one gets

$$a_{11} = \frac{c+d}{2}, \quad a_{22} = \frac{c+d}{2} - 1, \quad a_{32} = \frac{c+d}{2}$$

implying that

$$a_{22} = a_{11} - 1 \quad \text{and} \quad a_{32} = a_{11}$$

Likewise, adding the first and fourth equations, subtracting the second from the fifth, and subtracting the third from the sixth, one gets

$$a_{12} = a_{21} = a_{31}$$

We can now write

$$M = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{11}-1 & 1 \\ a_{12} & a_{11} & 0 \end{bmatrix}$$

That $h^M(p_3) = p'_3$ and $h_M(p_4) = p'_4$ give another two matrix equations

$$\begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{11}-1 & 1 \\ a_{12} & a_{11} & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} e \\ 2e \\ e \end{bmatrix} \quad \text{and}$$

$$\begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{11}-1 & 1 \\ a_{12} & a_{11} & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -f \\ 2f \\ f \end{bmatrix}$$

where e and f are arbitrary non-zero scalars. Again one obtains six equations, as in (A.8), which can be solved to find that

$$a_{11} = -1/2 \quad \text{and} \quad a_{12} = 0$$

We have, finally, that

$$M = \begin{bmatrix} -1/2 & 0 & 0 \\ 0 & -3/2 & 1 \\ 0 & -1/2 & 0 \end{bmatrix}$$

(or, a non-zero scalar multiple of the matrix on the RHS).

Exercise A.40. The projective transformation h^M of the preceding example mapped, by design, the regular point $[0\ 0\ 1]^T$ to the point at infinity $[0\ 1\ 0]^T$. What other regular points, if any, does it map to a point at infinity?

Exercise A.41. Determine the projective transformation h^M of $\mathbb{P}^2$ that transforms the rectangle q on the plane $z = 1$ with vertices at

$$p_1 = [0.5\ 1]^T,\ p_2 = [0.5\ -1]^T,\ p_3 = [1\ -1]^T \text{ and } p_4 = [1\ 1]^T$$

to the trapezoid q' on $z = 1$ with vertices at

$$p'_1 = [-2\ 2]^T,\ p'_2 = [-2\ -2]^T,\ p'_3 = [-1\ -1]^T \text{ and } p'_4 = [-1\ 1]^T$$

(see Example A.12 earlier for the solution).

Exercise A.42. Prove that there exist projective transformations which are neither affine nor snapshot.

Suggested approach: Corollary A.1 implies that a square can be projectively transformed to any non-degenerate quadrilateral. Non-degenerate quadrilaterals q' that can be obtained from a square q by a snapshot transformation are the intersections of a non-radial plane with the "cone" C through q (see Figure A.26). Those that can be obtained by an affine transformation, on the other hand, are parallelograms.

Therefore, if one can find a non-degenerate quadrilateral q'' which is neither a parallelogram nor the intersection of C with a plane, then one shows that there exists a projective transformation neither affine nor snapshot.

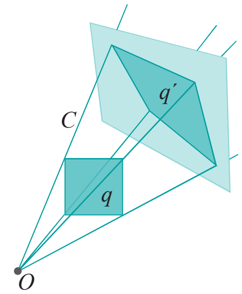

Figure A.26: The square q is mapped to the quadrilateral q' by a snapshot transformation.