Chapter 1

Shadows from Higher Dimensions

A hemispherical bowl has a single, circular edge. So does a disc. If you sew the disc to the hemisphere you obtain a closed, two-sided surface; it has an inside and an outside. Topologists still call this a sphere, and it could be inflated back to a geometrical sphere if you like. Now take a long, narrow strip of paper, give it a half twist and glue the ends together. The resulting surface is not only one sided, it has but a single, closed edge. What happens if this edge is sewn to the rim of a disc?

Studying the variety of answers to this question, first asked by the astronomer August Möbius [1867], has been the gateway to topology for generations of students. Geometry is a “picture-friendly” branch of mathematics, or at least it was so a hundred years ago. Möbius’s contemporaries found a model for his surface among Steiner’s many brilliantly conceived but poorly “documented” geometrical constructions. Jakob Steiner, Pestalozzi student, tutor and “high school” teacher, colleague of Abel, Crelle and Jacobi, and ultimately holder of a special professorial chair in Berlin created by the brothers Humbolt, is the grandfather of synthetic geometry. This was a reaction to the analytical geometry that ruled the day. He left behind volumes of tantalizing theorems which were ultimately proved by Cremona and Veronese, using the very methods of analysis and algebraic geometry Steiner detested. Till the end of the century the best answer to Möbius’s question was the surface Steiner called his Roman surface in memory of a particularly productive sojourn in that romantic city.

The only Hilbert student who wrote a dissertation in geometry, Werner Boy [1901], found a “simpler” surface in 1900. It is an immersion of the projective plane in 3-space, but Boy could not find the algebraic equation for his surface. Only recently, François Apéry [1984] succeeded in doing that. Apéry is a student of that great topological visualizer Bernard Morin of Strasbourg, to whom descriptive topology owes the greatest debt. Morin’s vivid, pictorial description of his bold constructions has inspired their realization in many a drawing, model.
CHAPTER 1. SHADOWS FROM HIGHER DIMENSIONS

computer graphic and video. But he insists that ultimately, pictorial descriptions should also be clothed in the analytical garb of traditional mathematics. What follows is an introduction to Morin’s descriptive method.

This chapter is about curved surfaces in space; what they look like, how to draw them, and the algebra that is manifested in their construction. Most of them are best described in terms of more than three parameters. So the objects in 3-space pictured in this chapter are themselves shadows of objects spatially extended in higher dimensions. I also hope to teach you how to draw them on paper, blackboard and computer easels, as Adobe Illustrator etc. The use of 2-dimensional diagrams, composed of clearly labelled and textually interpreted plane arrangements of points and lines, is well established in mathematics and taught in school. It is another matter with 3-dimensional models. To be sure, cylinders, globes and cones, the Platonic solids and an occasional Möbius band are all part of our students’ mathematical furniture. These shapes can be reliably referred to by name, without further visual aid. This is not so for shapes the least bit more complicated. Here, the finished object, made with sticks and strings, or Plaster of Paris, or just its photographic image, is often all we show to the student. It is, however, in the making of the model, in the act of drawing a recognizable picture of it, or nowadays, of programming some interactive graphics on a microcomputer, that real spatial understanding comes about. It does this by showing how the model is generated by simpler, more familiar objects, for example, how curves generate surfaces.
**Crossing a Channel.**

This part of the story begins with a parabolic cylinder generated by a parabola inscribed in a rectangle. The parabola passes through two adjacent corners of the rectangle and is tangent to the opposite edge at its midpoint. Figure F1tl embodies a recipe for sketching this curve inside a given perspective rectangle. Draw the diagonals of the quadrilateral (rear panel) to find its center. Connect the center to the two vanishing points (arrows). This halves the edges, thus quartering the rectangle. Diagonalize again (middle panel) to find the centers of the lower quarters or guess their location. These points lie on the parabola, and they determine the quarter points on the base edge of the rectangle. The lines from these quarter points to the corners are tangent to the parabola. The arcs, shown on the front panel, are now easy to draw. If necessary, the foregoing procedure may be iterated in the new, smaller rectangle determined by the quarter points. It juts out of the front panel for emphasis.

**Figure 1.**

Now extrude this figure. That is, translate the $XZ$-rectangle, together with its inscribed parabola, orthogonally to itself, in the $Y$-direction. This forms the box which frames the surface. To draw F1te in a given box, construct parallel
parabolas on opposite faces and connect corresponding points by straight lines. If one parabola is reoriented, then the connecting lines cross, forming $F_{1_{tr}}$. This is a more general kind of extrusion.

The *keel* is the line joining the vertices of the front and back parabolas. Note that, in general position, the keel is *not* the contour of the surface. I have drawn the $XYZ$-coordinate frames away from the true positions in the pictures. In $F_{1_{tc}}$, the origin is located in the middle of the keel and not on the contour, as it is often drawn. In $F_{1_{tr}}$, the origin and the $Z$-axis is visible. Note how the design of the arrowheads on the coordinate frames helps tell the story. The frame on the right suggests that the surface is formed by extruding the $XY$-plane upward, making the surface the graph of a function of two variables, but with a serious singularity over the origin. More geometrically, imagine a double keel opening up with increasing angle as the horizontal plane rises, forming a second extrusion for the Whitney umbrella.

We can see how these extrusions produce the same surface by parametrizing the first thus.

$$
\begin{align*}
X &= -x & Y &= -1 & Z &= x^2 \\
X &= +x & Y &= +1 & Z &= x^2 \\
X &= xy & Y &= y & Z &= x^2 
\end{align*}
$$

Column three is the canonical parametrization of the *Whitney umbrella* we have met before in $C_{1F3}$. When sketching an umbrella remember that is it attached to its containing box along the keel. The keel splits into two crossing lines that become diagonals on the face opposite that of the keel $F_{1_{bl}}$. The line of double points, call it the *stalk*, is orthogonal to the keel at the pinch point. In a sense, the stalk continues past the pinch point to form the “handle of the umbrella,” also called its *whisker*. An algebraic geometer would eliminate the parameters, and consider the zero-set (variety) of $X^2 - Y^2 Z = 0$ over the complex numbers. The negative $Z$-axis in our “real” space is where the “imaginary”, and therefore invisible sheet of the surface crosses through itself. In $C_{2F11}$ we met a more complicated example of a whisker in connection with Morin’s pinch point cancellation move. It plays an important role also in Apéry’s story, told at the end of this chapter. Let us return to the drawing lesson.

The keel joins the vertices of the two parabolas. You can find the precise location of the contour by connecting corresponding points on the parabolas by straight construction lines and sketching their envelope. You could build a physical device by building a string model of this surface inside a transparent plastic box. Use this to cast instructive shadows. A few rules of thumb can help you guess a plausible placement of the contour in a picture. Curves on the surface cannot cross a contour because the surface bends away from view there. Thus the contour is tangent to the parabolas at opposite ends and to the keel at the pinch point. Hence it is transverse to the stalk. You should practice drawing Whitney umbrellas for various positions of the box. Note that for $F_{1_{bl}}$, the contour is obliged to form a cusp, in which case the pinch point becomes invisible, $F_{1_{br}}$.

You may have observed an *abus de dessin* in $C_{2F1_{bot}}$, where I used the same straight line for keel and contour on the upper umbrella and skipped the contour
entirely on the lower one. Such a graphical counterpart of Bourbaki’s *abus de langage* is harmless provided the “errors” can easily be corrected and they do not mislead the viewer. This is a drawing of two umbrellas joined crosswise to form a singular Möbius band which should, more properly, be called a *Möbius cap*. The surface is one-sided provided you consider the double line to be penetrable in some sense. My next figure explains how to do this.
to distinguish two generic cases depending how the surface crosses the principal
tangent plane at the pinch point. To model this, bend the keel of the canonical
umbrella either towards the stalk or towards the whisker by relacing $Z = x^2$
by $Z = x^2 \pm y^2$. Eliminating the parameters to obtain $X^2 - Y^2 Z \pm Y^4 = 0$, we
can study the cross-sections of the new shape by holding $Z$ constant. For the
negative sign the result looks like the shadow of a branch point, as in the center
detail of $C1F5$. Note how the contour develops two cusps near the pinch point.
The contour would close on a third cusp further up the stalk if the umbrella wer
to open so much wider that you could see the stalk from inside the channel.

When the sign is positive and the umbrella is entirely on one side of the
supporting tangent plane, the $Z$-sections are figure-8 shaped curves. This can
be seen by rewriting the equation: $X^2 = Y^2(Z - Y^2)$. Some conventional views
of the positively bent umbrella are $F_{2t}$ and $F_{2r}$, but despite appearances, $F_{2r}$
is not. Each figure has an *abus de dessin*. In $F_{2t}$ the pinch point is “correctly”
on the contour, but the cusp should not be on the stalk, which is a double
curve. While $F_{2t}$ shows a conical pinch point neighborhood very well, the
double curve should not end at a cusp of the contour. The Whitney umbrella is
a singular image of a disc, while $F_{2t}$ is a singular Möbius band. It is important
to note that the two “bent” umbrellas are part of an ambient isotopy through
the “straight” umbrella

$$
X(t) = X + tY^2 \\
Y(t) = Y \\
Z(t) = Z
$$

and thus do not violate the stability of the pinch point singularity.

We return to the canonical equation of the umbrella and “homogenize”
it to $WX^2 - Y^2 Z = 0$, which is the equation of a Plücker conoid. This de-
fines a surface in projective 3-space equipped with homogeneous coordinates
$W : X : Y : Z$. If you interpret Cartesian $XYZ$-space as the finite part ob-
tained by setting $W = 1$ you have what we started with.

Recall that a *projective space* of two, three or more dimensions may be
defined as the totality of all straight lines through the origin of a Euclidean
space of one dimension higher. To visualize what the projective umbrella looks
like at the “end” of the stalk and of the keel, take a different “picture” of
projective space. This time capture the lines through the origin in 4-space
on a 3-space cross section given by the equation $W + X + Y + Z = 1$. The
coordinate axes of 4-space cross this *hyperplane* at the vertices of a tetrahedron,
and we can interpret $W, X, Y, Z$ as barycentric coordinates in that 3-space. The
tetrahedron is drawn in the lower half of the figure, and the surface details
show what it looks like near the “four corners of the world”. This picture is a
*perspective chimera* in that it has several different triplets of vanishing points.
The special effect that can be achieved with multiple perspective is familiar to
fans of Escher’s drawings. Here I have drawn the umbrella at the lower corner
of the tetrahedron as if $W = 1$ held in an entire cubical neighborhood. So I used
the three other corners as the vanishing points of this detail; and similarly at
the other corners. Can you discover for yourself what’s in the opaque box?
Each line through the origin of a Euclidean space also pierces the unit hypersphere in two antipodal points. The sphere is thus a double cover of projective space. A hemisphere covers the projective space once, except along its equatorial border. If this border were smoothly sewn together so as to identify antipodes, a topological model of the projective space would be realized. In dimension two, such a surface was proposed by M"obius [1867] as an example of a closed, one-sided surface. He imagined it as a closed ribbon with an odd number of half twists sewn to a disc along its boundary. Of course, no such tailoring is possible in 3-space: we need an extra dimension.

Consider first a parametrization

\[
\begin{align*}
W &= z^2 \\
X &= xy \\
Y &= yz \\
Z &= x^2
\end{align*}
\]

of the conoid obtained by “homogenizing” the formulas for the umbrella.

Let us, for the sake of symmetry, append the remaining two quadratic monomials, \( U = y^2 \) and \( V = xz \), but consider \( U : V : W : X : Y : Z \) a point in real projective 5-space. This produces a parametrization of the Veronese surface. To see that the Veronese surface is an embedded projective plane, we check that \( \mathbb{P}^2 \to \mathbb{P}^5 \) is injective (one-to-one). Suppose that two values are proportional, \( \tilde{x}^2 = tx^2 \), \( \tilde{y}^2 = ty^2 \), \( \tilde{z}^2 = tz^2 \), \( \tilde{y}\tilde{z} = tyz \), \( \tilde{z}\tilde{x} = tzx \), \( \tilde{x}\tilde{y} = txy \),

then \( t \) must be positive and so have a square root. Without loss of generality we may set \( t = 1 \) in our equations because \( x : y : z = \sqrt{i}x : \sqrt{i}y : \sqrt{i}z \). Thus each coordinate with a tilde equals \pm the coordinate without a tilde, but we don’t know yet that all signs are matched. From the first, second and last equation we have \((\tilde{x} + i\tilde{y})^2 = (x + iy)^2\), and two of the three signs match. Cycling \( x, y, z \) makes all signs match, which was to be proved.

Thus, to visualize the M"obius surface we may look at lower dimensional projections of the Veronese surface. Pl"ucker’s conoid is such a shadow in 4-space, but it is a surface with singularities, as Jeff Weeks pointed out to me between the first and second printing of the first edition of this book. There are, however, projections of Veronese’s surface which are embeddings of M"obius’s surface in 4-space.
Here is another, more beautiful linear projection of Veronese's surface into 4-space which has an algebraically convenient parametrization, especially in cylindrical and spherical coordinates. The equatorial angle (azimuth) is $\theta$ and the elevation, $\alpha$, is the angle from the equator towards the north pole on the
unit sphere.

\[
\begin{align*}
W &= x^2 - y^2 &= r^2 \cos(2\theta) &= \cos^2(\alpha) \cos(2\theta) \\
X &= 2xz &= 2r \cos(\theta)z &= \sin(2\alpha) \cos(\theta) \\
Y &= 2yz &= 2r \sin(\theta)z &= \sin(2\alpha) \sin(\theta) \\
Z &= 2xy &= r^2 \sin(2\theta) &= \cos^2(\alpha) \sin(2\theta)
\end{align*}
\]

To see that this is also an embedded Möbius surface, proceed as before to show that two points on the unit sphere mapped to the same point in 4-space are antipodes. From the first and last line (in Cartesian coordinates) we have \((\hat{x} + i\hat{y})^2 = (x + iy)^2\), and \((\hat{x}, \hat{y}) = \pm (x + iy)\). From the second and third it follows that \((\hat{x} + i\hat{y})\hat{z} = (x + iy)z\). Thus \(\hat{z} = \pm z\), with the same \pm sign.

Let us first consider the \(XYZ\)-shadow of this surface and begin with the unit \(\theta\)-circle in the horizontal \(XY\)-plane. Hold \(r = 1\), \(z = 1/2\) constant and note how \(Z = \sin(2\theta)\) bobs up and down twice, which bends the circle into a wobbly hoop as in \(C_1F_2\) bot. Diametrically opposite points on the unit circle map to points on the wobbly hoop which have the same altitude. These four points determine a vertical rectangle, one for each pair \(\pm \theta\). As the rectangle turns inside the vertical cylinder, its height changes sinusoidally, \(F_3\) tc. Twice it degenerates into a diameter of the unit circle.

The mobile edge of this rectangle, connecting opposite points on the wobbly hoop, generates a ruled surface, \(F_3\) tl, consisting of two Whitney umbrellas joined as in \(C_2F_1\) br. This Möbius cap is a special projection of Pliker’s conoid; a calculation I leave to you. Use the cylindrical parametrization, third column, to generate this surface, holding \(r = 1\) constant and releasing \(z\) to go from \(+1/2\) to \(-1/2\). If, instead, you release \(r\) and hold \(z\) constant, you inscribe a flexing parabola in the rectangle, which generates the saddle surface, \(C_1F_2\) br. Glue these two surfaces together along the wobbly hoop and you get a topological model of the Roman surface. The algebraic surface is obtained by rounding the rotating rectangle into a moving ellipse, \(F_3\) tc. This bends the two umbrellas and flattens the parabolic saddle.

To verify, decompose the spherical parametrization (fourth column) thus:

\[
\begin{bmatrix}
\cos(\theta) \\
\sin(\theta) \\
0
\end{bmatrix}
\begin{bmatrix}
\sin(2\alpha) + \cos^2(\alpha) & 0 \\
0 & \sin(2\theta)
\end{bmatrix}
\].

The sliding coefficients, \(A = \sin(2\alpha)\) and \(B = \cos^2(\alpha)\), of this interpolation between two vectors satisfy the equation of a semi-ellipse inscribed in a unit square, \(A^2 = 4B(1 - B)\). To see this geometrically, note that \((1/2A)^2 = (1 - B)(B - 0)\), expresses Thales’ principle: as a point \(B\) moves across the diameter of a semi-circle, the altitude \(1/2A\) over \(B\) is the geometric mean of the distance from \(B\) to the ends of the diameter. Since circles go to ellipses under an affine transformation, such an interpolation between any two vectors produces the semi-ellipse inscribed in the associated parallelogram. This construction thus joins the more familiar trigonometric and linear interpolations, based on \(A^2 + B^2 = 1\).
and \( A + B = 1 \) respectively, which are so useful in spherical and projective coordinatizations.

The process also creates four new umbrellas, \( F_{3\text{tr}} \), whose stalks are the two degenerate ellipses. The three double lines cross at a triple point and end at six pinch points. Here are three ways to model the Roman surface as a 3D-sketch. In \( F_{3\text{nl}} \), pinch dimples on opposite sides of a clay ball, and imagine a place for each of the six pinch points. Progressively mold the pinch points in place as you rotate the model a third of a turn at a time. A piece-wise linear model, the heptadron (not shown), may be built out of three mutually orthogonal squares which cross each other on diagonals, plus three triangles. Imagine Cartesian coordinate planes inside an octahedron, with opposite octants capped by triangles or look at Figure 288 of Hilbert and Cohn-Vossen [1932]. A quick way to produce something “half way” between the heptahedron and the smooth surface from a ball of art-gum eraser is shown on the right, \( F_{3\text{br}} \). Take two quarter sections of the ball and cut them almost in half. Now turn the eighths against each other and stick the two pieces together. It looks like a ball with alternate octants removed.

When modelling or drawing the Roman surface, remember that it fits inside a tetrahedron, touching the faces along circles. The curvature of the surface changes at these four circles \( F_{3\text{pl}} \) and \( F_{3\text{hc}} \). Informally speaking, a surface bends to one side of its tangent plane in a neighborhood of a point of positive curvature and one says that the surface is supported by the plane. The surface crosses its tangent plane at a point of negative curvature.

To verify the first assertion algebraically, consider where the slant planes of the tetrahedron meets the surface. For example,

\[
1 + Z - X - Y = x^2 + y^2 + z^2 + 2xy - 2xz - 2yz = (x + y - z)^2 \geq 0.
\]

Since this is non-negative, the slant plane supports the surface on a curve which is the image of the circle on the source sphere cut by the plane \( z = x + y \).

To discover the shape of this curve consider the calculation

\[
1 = (X + Y - Z)^2 = X^2 + Y^2 + Z^2 + 8xyz(z - x - y)^2 = X^2 + Y^2 + Z^2.
\]

It shows that the distance from the origin in \( XYZ \)-space to this plane curve is constant; so it is a circle.

The curvature changes sign along what are ineptly called the parabolic curves on a surface. There is a practical way based on the Hessian principle of locating certain parabolic curves on a surface. I shall use it to establish the second assertion, that these circles of tangency to the tetrahedron are parabolic curves on the Roman surface. Let \( F : \mathbb{R}^2 \to \mathbb{R}^3 \) parametrize a surface in space. The cross-product \( N = F_x \times F_y \) of the two tangent vectors gives a vector normal to the surface. A change of coordinates multiplies \( N \) by the Jacobian determinant:

\[
F_u \times F_v = F_x \times F_y \frac{\partial (x, y)}{\partial (u, v)} \partial (u, v).
\]
Thus the *unit normal* $N/|N|$ is independent of the parametrization and defines a mapping from the surface to the unit sphere, called its *Gauss map*. For a non-orientable surface (like the Roman surface) we work on its (orientable) double cover. Thus the Roman surface has a pair of oppositely oriented unit normals at each point. The Jacobian of the Gauss map is called the *Gauss curvature function*. Here is less complicated way of thinking about this function. For each plane normal to a surface at a point, the curvature of the curve of intersection is a *sectional curvature* in that plane. The Gauss curvature at this point is the product of the maximum and the minimum sectional curvatures. (Their average is the *mean curvature* and this vanishes on surfaces minimizing area locally; soap films spanning wires, for example.)

For drawing or modelling it usually suffices to know where the curvature is positive or negative and to find the parabolic curves separating these regions. Suppose, we reparametrize the surface locally as the graph of a function $z = f(x, y)$. The surface bends to one side of its tangent plane where the Hessian determinant, $f_{xx}f_{yy} - (f_{xy})^2$, is positive. The surface crosses the tangent plane where the Hessian is negative. Though you can derive this from the second Taylor expansion, I prefer to argue geometrically as follows. The vector

$$N = \begin{bmatrix} -f_x \\ -f_y \\ 1 \end{bmatrix}$$

is normal to the surface and the Gauss map, $(x, y) \to N/|N|$, takes its values in the upper unit hemisphere. Since central projection preserves orientation, see $F_{4ul}$, you need only check the orientation of the planar map $\nabla f(x, y)$. The Jacobian of an orientation preserving (reversing) mapping is positive (negative). Thus, the orientation of the Gauss map matches the sign of the Hessian.

Let us apply this Hessian principle to a general model of a surface patch obtained by bending the keel of a U-shaped channel, $z = u(x), u'(0) = 0 < u''$, horizontally to $y = a(x), a'(0) = 0 < a''$, and perhaps also vertically to $z = b(x), b'(0) = 0$, see $F_{4fr}$. Compute:

$$z = f(x, y) = u(y - a(x)) + b(x)$$

$$\nabla f = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} b' - a'u' \\ u'' \end{bmatrix}$$

$$\begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = \begin{vmatrix} b'' - a''u' + (a')^2u'' - a'u''u'' \\ -a'u'' \end{vmatrix} = (b'' - a''u'u''),$$

whence the parabolic curves is given by $b'' - a''u' = 0$.

In the case of the Roman surface, the keel remains in a plane; so $b'' = 0$ in our local parametrization. Thus the curvature changes along $u' = 0$ which is exactly on the keel $y = a(x)$. In general, the parabolic curve shifts to one side of the keel according to the sign of $b''(x)$. For example, for the *Peano saddle* $F_{4fr}$, the keel is already in the region of negative curvature because $b'' < 0$. If
we parametrize this surface thus:

\[ f(x, y) = (y - x^2)(y - 3x^2) = (y - 2x^2)^2 - x^4, \]

then it changes curvature on the parabola \(2y = x^2\), on which \(\nabla f = (8x^3, -3x^2)\). In other words, the Gauss map develops a Whitney cusp, as discussed in Chapter 1. You will find a very informative treatment of these matters in the monograph on the cusps of Gauss maps by Banchoff-Gaffney-McCrory [1982]. It is challenging to draw a good picture of the Peano saddle, whether by hand or by computer. The one here is not a good one.

The \(YZW\)-shadow of the Veronese surface from 4- to 3-space is another important surface of Steiner, the cross cap surface, which factors thus:

\[
\begin{bmatrix}
\sin(\theta) & 0 & 0 \\
0 & \sin(2\alpha) + \cos^2(\alpha) & \sin(2\theta) \\
0 & \cos(2\theta) & 0
\end{bmatrix}.
\]

This time, the \(\theta\)-circle, \((\sin(\theta), \sin(2\theta), \cos(2\theta))\), wraps twice about the \(Y\)-axis, forming a double hoop, \(F_{4_{bc}}\), with the positive \(W\)-axis pointing to the node in the hoop. The double hoop is usually associated with the complex squaring function. You can use the Cartesian parametrization of the Veronese surface I gave earlier to visualize Riemann surface \(W + iZ = (X + iY)^2\) in \(XYWZ\)-space by holding \(z = 1/2\) constant and letting \(x + iy\) range over the entire complex plane.

As before, connect the \(\pm \theta\) places on the hoop by a line segment parallel to the \(Y\)-axis and complete the rectangle. This time the rectangle rotates about one of its sides, and degenerates once as it squeezes through the positive \(W\)-axis. Now release the \(\alpha\)-parameter, which, by Thales’ principle inscribes an ellipse in the rectangle, \(F_{4_{bc}}\). The completed surface looks like a Möbius band (with a window hiding the double curve) glued into a spherical shell, \(F_{4_{bl}}\), which is how Möbius defined his surface in 1858 [Scholz, p.147].

The cross cap is supported by the dihedral \(W \pm \sqrt{2}Y = 1\), which fits over it like a roof. The curvature changes at the two ellipses where the cross cap touches the roof, by the Hessian principle. You will find a more complete description of the cross cap, with many illustrations, in Chapter VI of Hilbert and Cohn-Vossen [1932]. The computer animation by Banchoff and Strauss [1977] shows the transition from the Roman surface to the cross cap, effected by a rotation in 4-space. This film dramatically reveals the many remarkable shadows this beautiful surface casts into our space.
Boy Surface.

During the latter half of the 19th century the theory of surfaces matured and flowered. Its influence on the topology of manifolds is nicely documented in Scholz’ history book [1980]. At the turn of the century, however, there remained the tantalizing question of what a projective plane might look like if it were immersed (no pinch points) in 3-space. A topological solution was found...
by Hilbert’s student Werner Boy [1901]; but he was unable to satisfy the conventional expectations of his day: to give analytic formulas for his surface. Only in the 1980s, Bernard Morin and his collaborators have succeeded doing this. But that is getting ahead of our story. You will find photographs of a wire mesh model of Boy’s surface in Hilbert and Cohn-Vossen [1932, Fig. 321a-d], along with an introduction to the topology of this surface, which I shall not repeat here.

It is natural to ask whether, like Steiner’s two surfaces, Boy’s surface is also the 3-dimensional shadow of a projective plane embedded in 4-space. The negative answer has an elegant demonstration, which I learned from Ben Halpern and Jeff Boyer. Let \( F : \mathbb{P}^2 \to \mathbb{R}^3 \) denote an immersion parametrizing the Boy surface. It has one double curve, \( K \), with one triple point. Therefore, it’s preimage, \( C = F^{-1}(K) \), is a single curve. Let \( f : S^1 \to \mathbb{P}^2 : f(S^1) = C \) denote a parametrization of this immersed circle for which \( F(f(\theta)) = F(f(\theta + \pi)) \). The difference function \( w(\theta) = W(f(\theta)) - W(f(\theta)) \), where \( W \) is the fourth coordinate of a presumed embedding of \( \mathbb{P}^2 \) in \( \mathbb{R}^4 \), is never zero. But if \( w(\theta) > 0 \) then \( w(\theta + \pi) = -w(\theta) < 0 \). By continuity it must vanish in between.

Since no suitable embedding exists, let us relax that condition, and see just how cleanly the projective plane can be placed in 4-space so that its shadow is Boy’s surface. For observing 4-dimensional fauna in the absence of convenient analytic formulas we need some descriptive tools. The 2-dimensional shadow-watchers in Plato’s cave could have benefited from perspective drawing to encode the 3-dimensional world outside. Let us place the colors of the rainbow at the disposal of 3-dimensional modellers so that they may paint the fourth coordinate onto their surfaces. Regrettably, book pages and blackboards are flat and most pictures drawn on them are monochromatic. So you will have to be content with windows and painting instructions.

To reduce the complexity of the line pattern, deform the conventional shape of Boy’s surface \( F_{5t} \) by rolling the cusp forward, as in \( F_{5tc} \). (We do this repeatedly in \( F_8 \).) This splits the contour of \( F_{5t} \) with its three cusps and three apparent crossings, see \( F_{8bot} \), into two curves, a deltoid inside an oval, \( F_{5tr} \). The double curve, dotted in \( F_{8t} \), a bouquet of three loops tied to the triple point, reappears in the same shape once the three windows in \( F_{5tr} \) are replaced. I have tilted the deltoid relative to the vertical direction so that the horizontal saddle is near one cusp. By rotating \( F_{5t} \) slightly to cancel an extra saddle against a maximum, this surface also supports a Morse height function with only three extrema, one maximum, one saddle and one minimum.

We map the fourth dimension into the chromatic scale, and paint \( F_{5tr} \) green for “ground” level; the surface is still in 3-space and singular. Now and reden the top window to eliminate the upper loop of the double curve in 4-space. Push both lower windows into the blue direction to clear the green surface. There are no self-intersections on the green level as we can see from the drawing. The red disc in \( F_{5mr} \) does not intersect the two blue discs, and we ignore it. Since the rims of the two blue discs link in 3-space, each of the two preimages of the vertical arc of intersection is an arc from the border to the interior of each disc. As the color of this arc in the right disc rises from deep violet to blue near the
The arc in the left disc falls from blue to violet. By continuity there is some point inbetween where both arcs have the same shade of indigo. This is where the node of the surface in 4-space occurs. Note that if, in 3-fold rotational symmetry, the top window had also been pushed into the blue direction, then two more indigo nodes would have been formed, one for each of Morin’s pinch point cancellations on the Roman surface raised to 4-space.

There is another way to build up dimensions by stacking consecutive 3-D sections. Slice $F_{5\text{tr}}$ by horizontal planes and reinterpret each plane as the horizontal shadow of a 3-dimensional slice of 4-space. Eight of these sections are stacked in a column, $F_{5\text{mt}}$ to $F_{5\text{bl}}$. I have used the convention of knot diagrams, interrupting a line as it passes under another. Note that “red” is “down” in each of these eight copies of 3-space. (I have omitted drawing the obvious point followed by an oval at the top, and the oval shrinking to a point at the bottom.) If you interpret the slicing plane of $F_{5\text{tr}}$ as falling in time, then the stack depicts eight instances, $\{T_1, T_2, ..., T_8\}$, of a curve moving in 3-space. It is a simple closed curve except at two instances. At $T_3$, a recombination (as in molecular genetics) signals a saddle orthogonal to the (time) × (color)-plane in 4-space. In my drawing, it has the shape of cubic surface, showing a cusp in the picture plane, $F_{5\text{center}}$. Later, at $T_6$, the curve crosses through itself, signifying the node on the surface in 4-space.

The shadow of this moving curve in the picture plane (close the gaps) depicts a regular homotopy of an immersed circle except at the recombination point. Stacking these temporal slices spatially, rebuilds Boy’s surface from the top down. This is, in fact the technique originally used by Boy. By $T_1$, a cusp has already occurred in the contour. A curve of double points is created between $T_1$ and $T_2$. After passing through the saddle, the curve has two opposing loops, $T_4$, which must be eliminated by the regular homotopy. A second double curve and a triple point is created as a segment sweeps across the left loop between $T_4$ and $T_5$. The last loop of the double curve disappears between $T_7$ and $T_8$, and a cusp allows the surface to close up. The stack from $T_4$ to $T_8$ is the crossed cusp shown in $F_{5\text{bc}}$ with and without windows, which form the triple point of Boy’s surface at $T_6$. A rearrangement of the stack produced $F_{5\text{br}}$ in Francis-Morin [1979] which inspired a beautiful sculpture by Benno Artman.
SLICE AND SHADOW.

Let me review this technique of visualizing painted surfaces in 4-space by applying it to the 3-dimensional surface primitives, the pinch point, \( \mathbf{F}_{\text{top}} \) and the double line, \( \mathbf{F}_{\text{bot}} \). Consider 4-space framed by an orthogonal triple, \( \mathbb{R}^2 \times \mathbb{R}^1 \times \mathbb{R}^1 \), consisting of the base plane, the temporal direction and the chromatic directions. Usually, the base plane is also the picture plane of the illustrations. For clarity, I have tilted the base plane. It is represented by the front

Figure 6.
face of the boxes framing all eight details.

The cubical boxes extrude the base plane into the chromatic direction, while the doubled boxes do so in the temporal direction. If you slice the surface in 4-space by moving the chromatic cube parallel to itself in the temporal direction, you will see a curve moving in space. At the top, the motion untwists an apparent loop. Below, the loop exchanges an over to an under crossing, thereby signalling a node, which is a one-point self intersection of the surface in 4-space. Projected to the base plane, the former motion looks like a loop being pulled taught, passing through a cusp singularity. The latter projection does not move.

Stacking the base curves into the temporal direction extrudes a Whitney umbrella (top) and a crossed tube (bottom). The pinch point in the umbrella corresponds to the cusp on the base plane shadow. On the other hand, a bit of calculus will convince you that the apparent cusp in the base projection of the loop indicates a tangency of the surface to the chromatic direction in 4-space. The node of the second surface is forced by the paint on linked of the windows used to remove the double locus. Let the front of the chromatic cubes be red, the rear blue, and the middle green. Tracing the loops top-down, the one in $F_{6_{bf}}$ turns green-blue-green-red-green, while $F_{6_{br}}$ dips into the red ink first and then into the blue. Therefore, in the temporal box, $F_{6_{bc}}$, the straight edge of each D-shaped windows is painted blue, and the round arc extends into the red range. The double line passes from red to blue on the rear window, as it goes from blue to red on the front one. At some place along the double curve both must be the same shade of green, which confirms a node.
Whitney Bottle.

The fact that the motion of the slicing curve for Boy’s surface, $F_{5_{14}}$, casts no cusped shadow in the (horizontal) base plane, says that the unit vector in the red direction is never tangent to the surface in 4-space. In other words, this 4-dimensional immersion of the projective plane supports a continuous field of normal vectors, which is not possible for an immersion of a closed non-orientable (1-sided) surface in 3-space. The embedding of such surfaces in 4-space which
admit a “furry pelt without bald spots” was investigated by Whitney [1941] in the course of developing the theory of characteristic classes. He proved that surfaces of odd Euler characteristic (such as Möbius’s surface, but not the Klein bottle) there are no furry embeddings. Thus the self-intersection cannot be removed by a smooth deformation through immersions (a regular homotopy) in 4-space.

Whitney’s conjecture for the case of even Euler characteristic, proved by Massey [1969], implies, in particular, that the Klein bottle has three inequivalent embeddings. One of these is furry without bald spots. You can construct the Whitney stack for this one by slicing the usual immersion of the Klein bottle in 3-space. To visualize the Klein bottle without normal field, described in Figure 4 of Whitney [1941], proceed as follows.

First note that if at time $T_4$ in Boy’s stack, $F_{5\text{bl}}$, we had simply untwisted the two little loops, an embedding of the projective plane would have resulted. The temporal shadow (project along the time-axis) of this projective plane would have a segment of double points ending in two pinch points. In other words, this is an isotope of Steiner’s cross cap. You should design the stack, whose shadow is the Roman surface, as an exercise. The temporal section of Whitney’s bottle, reading $F_{7\text{left}}$ from the top down, develops two loops that recombine across a saddle forming two linked ovals. Flip one of them over and reverse the process. A schematic assembly of details in the shadow of this surface, $F_{7\text{cent}}$, may be read from the Whitney stack. Note the singular Möbius band crossed by a plane with a triple point, in the center of the figure, corresponds to the flip of the oval. A complete assembly of the shadow is shown at the right, $F_7$ using both windows and zones.
The Romboy Deformation.  

There is a topological transition from Steiner's Roman surface to Boy's sur-
face based on the the cancellation of adjacent pinch points, see Figures C2F9-11. This deformation may well have been known to Boy and used by him to discover the immersion of the projective plane that bears his name. I learned it from Tim Poston in 1977, who learned it from Bernard Morin. For convenient comparison I have redrawn several views of the surfaces you have already met, together with their line patterns. I shall use row×column notation for the details of this 6×4 tableau. Begin with the conventional view of the Roman surface $F_{8,12}$. In its line pattern $F_{8,11}$ are two contours (solid lines), three double curves (dashed lines), one triple point (center), six pinch points and arrows indicating where to cut the surface. The section $F_{8,22}$ contains the southern and south-eastern pinch points, which are cancelled first, $F_{8,32}$. Globally, we have reached the line pattern $F_{8,41}$ from which you can draw the surface $F_{8,42}$. The western pinch points cancel next, $F_{8,51}$ and $F_{8,52}$.

You can also imagine this surface as the result of untwisting two little loops, one after the other, on the second curve from the top in the stacked cross-sections of Boy’s surface $F_{5,1}$. Once the last two pinch points cancel, $F_{8,62}$, you are looking at the backside of $F_{5,1}$, but in a mirror. Thus $F_{8,63}$ is a mirror image of $F_{8,62}$, and with three windows, $F_{8,64}$, it is the same view of Boy’s surface as $F_{5,1}$, but rotated 45°.

The third column depicts the analogous sequence taking the “flat” view of Boy’s surface $F_{5,1}$ to the conventional “round” view. The northern cusp rolls over to the contour, $F_{8,32}$ to $F_{8,33}$. This pair of cusps is an upside down view of the pair in $F_{5,1}$. Can you tell, by looking at $F_{8,44}$ and $F_{8,54}$ which cusp has to roll over to obtain view $F_{8,33}$? Cusp rolling is an isotopy in 3-space, but cancelling pinch points is not necessarily the shadow of an isotopy in 4-space. As we have seen, in contrast to the Roman surface, Boy’s surface cannot be the shadow of an embedded projective plane in 4-space.

The two center columns of this tableau also portray qualitatively the two methods used by François Apéry to obtain explicit parametrizations of Boy’s surface [1984]. For the first, recall that an ellipse generates Steiner’s Roman surface. The sculptor Max Sauze built such a model for Bernard Morin. He decided to use planar ovals also to construct a wire model of Boy’s surface. This inspired the physicist and mathematical artist, Jean-Pierre Petit and his collaborator, Souriau [1982], to program the surface on an Apple computer. Their ellipse changes its shape and plane but stays tangent to a base plane at the south pole of the surface. For the surfaces in the tableau the south pole lies on the window of $F_{8,62}$ and is furthest from you in $F_{8,63}$. It is your nearest point on $F_{8,12}$ and is visible through the three windows of $F_{8,13}$. Petit and Souriau empirically determined formulas for the the position of the ellipse in its plane and the tilt of the plane for each moment of the motion.

Apéry begins, instead, with an elliptical generation of the Roman surface. His is different from the one I used above. In Apéry’s version the ovals remain tangent to a supporting plane at the south pole. Consider a point moving on a horizontal circle of radius $r_1$ at height 1, going twice as fast and in the opposite direction as a point moving about a parallel circle of radius $r_0$ in the base plane.
22  

CHAPTER 1. SHADOWS FROM HIGHER DIMENSIONS

The south pole is the origin. 

\[
\begin{bmatrix}
  r_1 \cos(2\theta) \\
  r_1 \sin(2\theta) \\
  1 \\
\end{bmatrix} + \frac{1}{1+t^2} \begin{bmatrix}
  r_2 \cos(-\theta) \\
  r_2 \cos(-\theta) \\
  0 \\
\end{bmatrix}.
\]

Apéry takes the southern circle of radius \( r_0 = \frac{2}{3} \) and the northern circle at \( r_1 = \frac{1}{\sqrt{2}} \). The first vector in the formula points from the origin to the \( \theta \)-point on the upper circle, the second points to the lower \( \theta \)-point, and the sliding coefficients satisfy Thales principle. Thus Apéry’s surface is an ovaesque generated by ellipses, all tangent to the horizontal plane at the origin. The three components of this parametrization are quadratic polynomials in the Cartesian coordinates of the unit sphere. The parameter \( t \) is the tangent of the latitude on the sphere.

Apéry ingeniously perturbs the sliding coefficients by deforming their common denominator to \( 1 - 2\beta t + t^2 \) where \( \beta = \left(b/\sqrt{2}\right) \sin(3\theta) \). This has the remarkable effect of cancelling all six pinch points at \( b = 1/\sqrt{3} \). In an algebraic tour de force, he eliminates the parameters and obtains a polynomial equation of degree six for Boy’s surface.

For reasons derived in classical algebraic geometry, Boy’s surface must have degree at least six. In a review of Boy’s problem, Heinz Hopf [1955, 1983 p.104] conjectured that a polynomial parametrization would also be at least of degree six. Apéry, however, found a parametrization of Boy’s surface using homogeneous polynomials of degree four. Here, his approach is more closely related to Morin’s original way of finding parametrizations of sphere eversions [1978]. The visualization, both graphically and analytically, of these complicated surface homotopies is the subject of the next chapter. I shall conclude here with a thumbnail sketch of Apéry’s second surface.

Morin’s method is the analytical analogue to the graphical technique I described in Chapter 2. Starting from the contour, build up the surface from this 2-dimensional template. Apéry begins by stabilizing a singular map of the sphere into the plane so that the image of the critical points looks like \( F_{8_{14}} \), without the dotted curve. You may have noticed that the contour image of the cusp-rolling motion resembles the sequence of pictures associated with Thom’s hyperbolic umbilic catastrophe. More about this resemblance, which is not coincidental, is amply illustrated and patiently explained by Jim Callahan [1974, 1977].

Once the cusp-rolling procedure is captured algebraically and given a three-fold symmetry, Apéry has a motion from \( F_{8_{14}} \) to \( F_{8_{61}} \). The last step is to “inflate” this one parameter family of sphere-to-plane mappings. By chosing the third coordinates carefully so that the map from the projective plane into 3-space is an immersion at all stages he obtains an isotopy from \( F_{8_{13}} \) to \( F_{8_{63}} \).

The pleasant fact that all these surfaces and their deformations can now be coded in simple languages, like BASIC and viewed on inexpensive, easy to use graphics computers, like the Apple, places this medium on the same elementary level as the hand graphics in this picture book. That is why I have emphasized

\[\text{Revise out the anachronism here.}\]
the analytic counterparts of some of the drawing techniques. Nonetheless, the
limited resolution, slow speed and modest flexibility of computer graphics at a
less than heroic hardware level benefits greatly from the programmer’s ability
to sketch effectively those views of the objects the computer is to render.