Singularities Bifurcations and Catastrophes

James Montaldi
University of Manchester

Contents

1

What's it all about?

MATHEMATICAL MODELS OF NATURE almost always involve solving equations (often differential equations) and these models, and their equations, frequently depend on external parameters. Examples of parameters might be the temperature of a chemical reaction, or the load on a bridge, or the tension in a rope (will it snap?), or the temperature of the ocean for plankton populations. In such models, it is important to understand how phenomena associated to that model can change as the parameters are varied. Usually one finds that a small change in the value of the parameters produces a corresponding small change in the (set of) solutions of the equation. But occasionally, for particular values of the parameter there is a more radical change, and such changes are called *bifurcations*. Often these bifurcations involve simply a change in the number of solutions. This chapter illustrates these ideas with a few examples.

Bifurcation theory is the (mathematical) study of such qualitative changes arising as parameters are varied. In this book, we consider a subset of this very general theory, namely *local bifurcations*, which excludes for example, routes to chaos in dynamical systems and other global bifurcations: everything we study can be described by local questions and local changes.

The majority of applications of mathematics involve differential equations (ordinary or partial), and the theory of bifurcations can be applied to these in a straightforward manner, as we will see in the first example below. However, the ideas are more general, and can be applied to other systems that depend on parameters, not just differential equations.

The general approach is to consider an equation $g(x) = 0$, where g may have several components,

$$g(x) = (g_1(x),\ g_2(x), \dots,\ g_p(x)),$$

and indeed so may x, that is $x = (x_1, x_2, \dots, x_n)$. Then introduce a parameter λ (also possibly multi–dimensional), writing

$$g_\lambda(x) = 0,$$

or $G(x;\lambda) = 0$. This is called a ***family*** of equations, depending on the parameter λ. We shall always assume our families are smooth as functions of (x, λ) (i.e. of class C^∞). The basic question of bifurcation theory is, how do solutions in x of these equations change as λ varies? And a bifurcation occurs when the change is in some sense qualitative.

There are many applications where the equation is a so-called *variational problem*, which means that the equation $g(x) = 0$ is in fact of the form $\nabla V(x) = 0$ for some scalar function V, usually called the *potential*. Then solutions of the equations $g(x) = 0$ are critical points of the function V. Zeeman's catastrophe machine described in Section 1.3 is one such physical example. A more geometric example is described in Section 1.4.

1.1 The fold or saddle-node bifurcation

The simplest mathematical example exhibiting a bifurcation is provided by the ordinary differential equation (ODE),

$$\dot{x} = x^2 + \lambda. \tag{1.1}$$

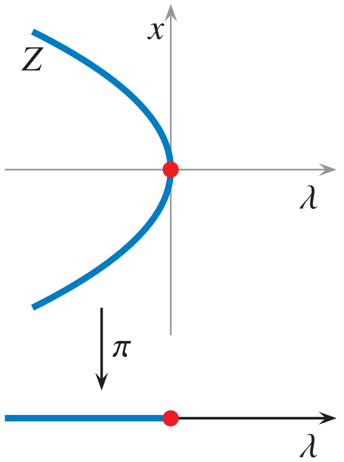
Here $\lambda \in \mathbb{R}$ is the parameter, and one often refers to $x \in \mathbb{R}$ as the ***state variable***. The dot over the x denotes the time-derivative, and a solution to the equation would be a function $x(t)$. Since this is a first order ODE, an equilibrium point occurs wherever the right-hand side vanishes. The equilibria therefore occur where

$$x^2 + \lambda = 0.$$

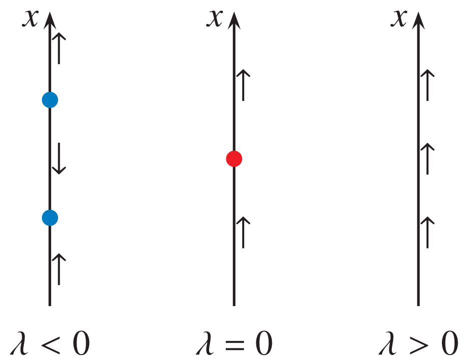
Define $g_\lambda(x) = x^2 + \lambda$. Then we are interested in solutions of $g_\lambda(x) = 0$, that is in the zeros of g. We call this set Z. Thus,

$$Z = \left\{(x, \lambda) \in \mathbb{R}^2 \mid x^2 + \lambda = 0\right\}.$$

The question we address is how the number of points in Z depends on λ. In this example, the curve Z is a parabola in the left half of the plane, as illustrated in Figure 1.1A. For $\lambda < 0$ there are two solutions (two equilibrium points), at $x = \pm\sqrt{-\lambda}$, and as λ increases to 0 these coalesce and then for $\lambda > 0$ there are no solutions (or they become complex, but we are just interested in real solutions). The transition, or *bifurcation*, occurs when $\lambda = 0$ (marked in red). The set of parameter values where such a bifurcation occurs is called the ***bifurcation set*** or ***discriminant***. The map π shown in the diagram is simply the projection taking $(x, \lambda) \in Z$ to the parameter value λ.



(A) The fold, or saddle-node, bifurcation diagram



(B) Phase diagram for the saddle-node bifurcation (1.1)

FIGURE 1.1 (A) shows the equilibrium points, forming a smooth curve in the (x, λ)-plane. (B) shows whether x is increasing or decreasing (the sign of $\dot{x}$) for different values of λ; the dots represent the equilibrium points and correspond to points on the curve in (A).

The behaviour of the differential equation is illustrated in Figure 1.1B. There are two equilibrium points when $\lambda < 0$ and none for $\lambda > 0$. In differential equations, this transition is often called a ***saddle-node bifurcation*** because in two dimensions, when $\lambda < 0$, one of the equilibria would be a saddle and the other a node. In singularity theory, where the specific application is not of concern, it is more generally called a ***fold bifurcation***, because of the shape of the curve Z folding over with respect to the parameter space (the λ-axis).

Remark In this simple example, the differential equation is a standard one and can be solved explicitly (by separation of variables, the type of solution depends on the sign of λ). However, more generally, bifurcation theory can be used to study equilibria (and neighbouring dynamics) of systems of ODEs where this is not the case, such as for example the ODE $\dot{x} = x^2 \mathrm{e}^x + \lambda$, which does not have a closed form solution but still exhibits a saddle-node bifurcation. ❞

The beauty of these ideas is that while the example above is so simple (g is quadratic), it contains essentially all that is expected to occur if there is just one parameter and no other restrictions. Imagine a small perturbation of the curve Z shown in Figure 1.1A; it seems reasonable to think that there will still be a single point where the curve 'folds over', with two solutions on the left and none on the

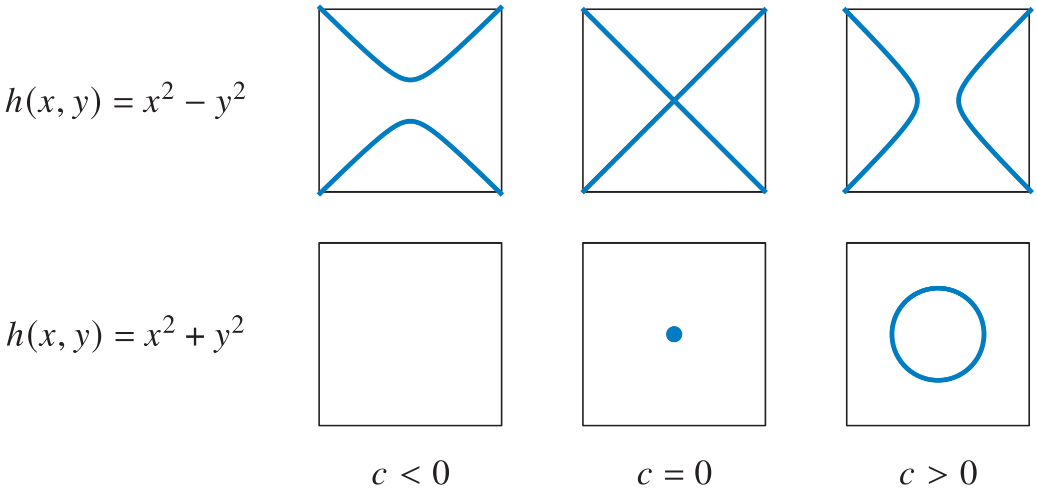


FIGURE 1.2 Two examples of bifurcation of contours $h(x, y) = c$ as c varies across the critical value of the function h.

right of this fold point (and one can prove this using the implicit function theorem; see Problem 1.6). This illustrates the *robustness* of the saddle-node bifurcation. In Section 1.5 below, we look briefly at an important bifurcation with one parameter, the pitchfork bifurcation, but where small perturbations do change its form. But first we look at two places bifurcations occur, the contours of a function as a parametrized set of equations, and a mechanical example with two parameters.

1.2 Bifurcations of contours

Landscape is determined in part by the height above sea level of each point of some region of the Earth. A ***contour*** is a curve on the landscape along which the height is constant; that is, for a given height the associated contour is the set of all points with that particular height. Let x, y be coordinates in the region in question, and $h(x, y)$ the height function. Then a contour at height c is the set of solutions of the equation

$$h(x, y) - c = 0.$$

Here we have a fixed function h and we can consider c as a parameter. Of course, height is only one example; another is the atmospheric pressure as a function on the surface, in which case the 'contours' are the familiar isobars from weather maps (although atmospheric pressure is best expressed as a function of three variables $P(x, y, z)$ as it varies with altitude z).

Consider a function $h(x, y)$ and the resulting equation $h(x, y) = c$. Most of the contours are curves, and a natural question to ask is, as c is varied, how can these curves change? The contours of a function are also called its ***level sets***.

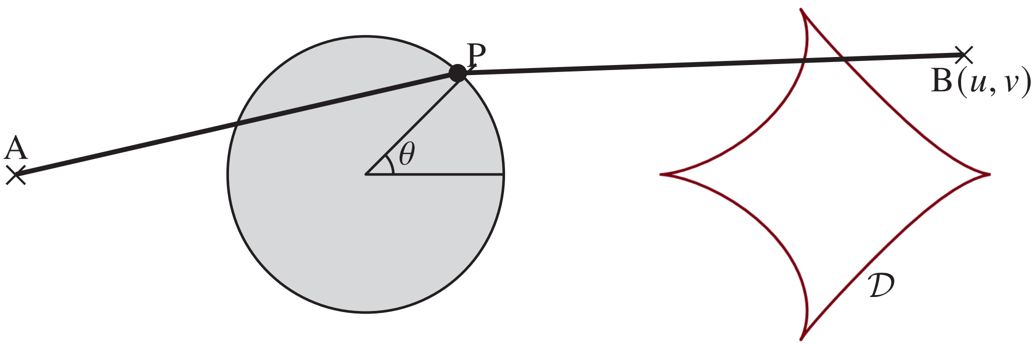


FIGURE 1.3 Schematic diagram of the Zeeman catastrophe machine. The red curve marked $\mathcal{D}$ is the discriminant or bifurcation set; notice the four cusps. Its precise size and position depend on the physical characteristics of the elastics and the position of the point A.

For example, suppose $h(x, y) = x^2 - y^2$. Then the contours are either hyperbolae or a pair of lines and the transition is depicted in the top row of Figure 1.2. For $h(x, y) = x^2 + y^2$ the contour is a circle for $c > 0$, a single point for $c = 0$ and is empty for $c < 0$. See the lower figures in Figure 1.2. In both cases a change occurs as one crosses the level $c = 0$, and one can show in general that qualitative changes only occur at *critical values* of the function; that is, the value the function takes at a critical point. We will study this in greater depth in later chapters.

A similar example in more variables is provided by $h(x, y, z) = x^2 + y^2 - z^2$. The zero level of this function is a circular cone in $\mathbb{R}^3$, while $h = 1$ is a one-sheeted hyperboloid and $h = -1$ is a two-sheeted hyperboloid.

1.3 Zeeman catastrophe machine

Conceived by Christopher Zeeman to illustrate the ideas of catastrophe theory, the Zeeman catastrophe machine consists of a wheel free to rotate about its centre, with a peg P attached at a point of its circumference. To the peg are attached two elastics: the other end of the first is pinned at a fixed point A in the plane of the wheel, while the other end of the second elastic is held by hand at a second point $B(u, v)$ in that plane. See Figure 1.3. The question is, how many equilibrium states are there of the wheel?

The answer will depend on where the end B is held; that is on the values of u and v, so these are the parameters. For each choice of point (u, v), the total elastic potential $V_{(u,v)}(\theta)$ is a function of θ, the position of the wheel (see Figure 1.4), and the equilibrium points are the points θ where V has a critical point: $\frac{\mathrm{d}}{\mathrm{d}\theta}V_{(u,v)}(\theta) = 0$.

The computation of the potential is straightforward but lengthy (and not relevant here), but the conclusion can be described simply. In the (u, v)-plane, there is a curve with four cusps, marked $\mathcal{D}$ in the figure. If the point B is within the curve, the wheel has four equilibrium points, two of which are stable (where $V'' > 0$) and two are unstable (where $V'' < 0$). On the other hand, if B lies outside this curve, then the wheel has only one stable and one unstable equilibrium point. The transition from four to two critical points happens when B approaches the curve $\mathcal{D}$ from the inside, and two of the critical points get closer and coalesce becoming degenerate in the process, and then disappearing; this curve $\mathcal{D}$ is therefore the discriminant of this family. This transition is the same as that in the fold bifurcation described above, although something more involved happens at the cusp points of the discriminant.

1.4 An example from geometry: the evolute

Consider a smooth simple closed curve C in the plane (e.g. an ellipse: a curve is said to be *smooth* if it has a parametrization whose derivative is nowhere zero). Let $P(u, v)$ be a point in the plane (possibly on C) with coordinates (u, v). The geometric question is, can you draw a perpendicular to the curve from the point P, and if so how many? (If P lies on the curve then we allow that the 'segment' (of zero length) from P to P is perpendicular to the curve.)

For example, if C is an ellipse, and P is at its centre, then it is not hard to see that there are 4 such perpendiculars – one to each of the points on the axes of the ellipse. What happens to those 4 points if P is perturbed? The feet of the perpendiculars will move, but can there be a different number of them? Imagine instead a point P' on the major axis of the ellipse, but outside the ellipse. It is easy to see that there are now only 2 perpendiculars from P' to C. See Figure 1.5. The bifurcation question is, how does 4 change to 2 as P is moved? And more completely, what is this number for all possible points P?

One observation is that for any P there are at least two such perpendiculars, and these arise at the nearest and furthest points of the curve to P as some thought should convince you (and which we prove below). This suggests defining the function on C which is the distance of each point of C to P. In fact we use the square of the distance which leads to simpler expressions after differentiating.

Let $\mathbf{r}(t)$ be a regular parametrization of the plane curve C, where 'regular' means that its derivative $\dot{\mathbf{r}}(t)$ is never zero, and for each point $\mathbf{c} = (u, v)$ in the

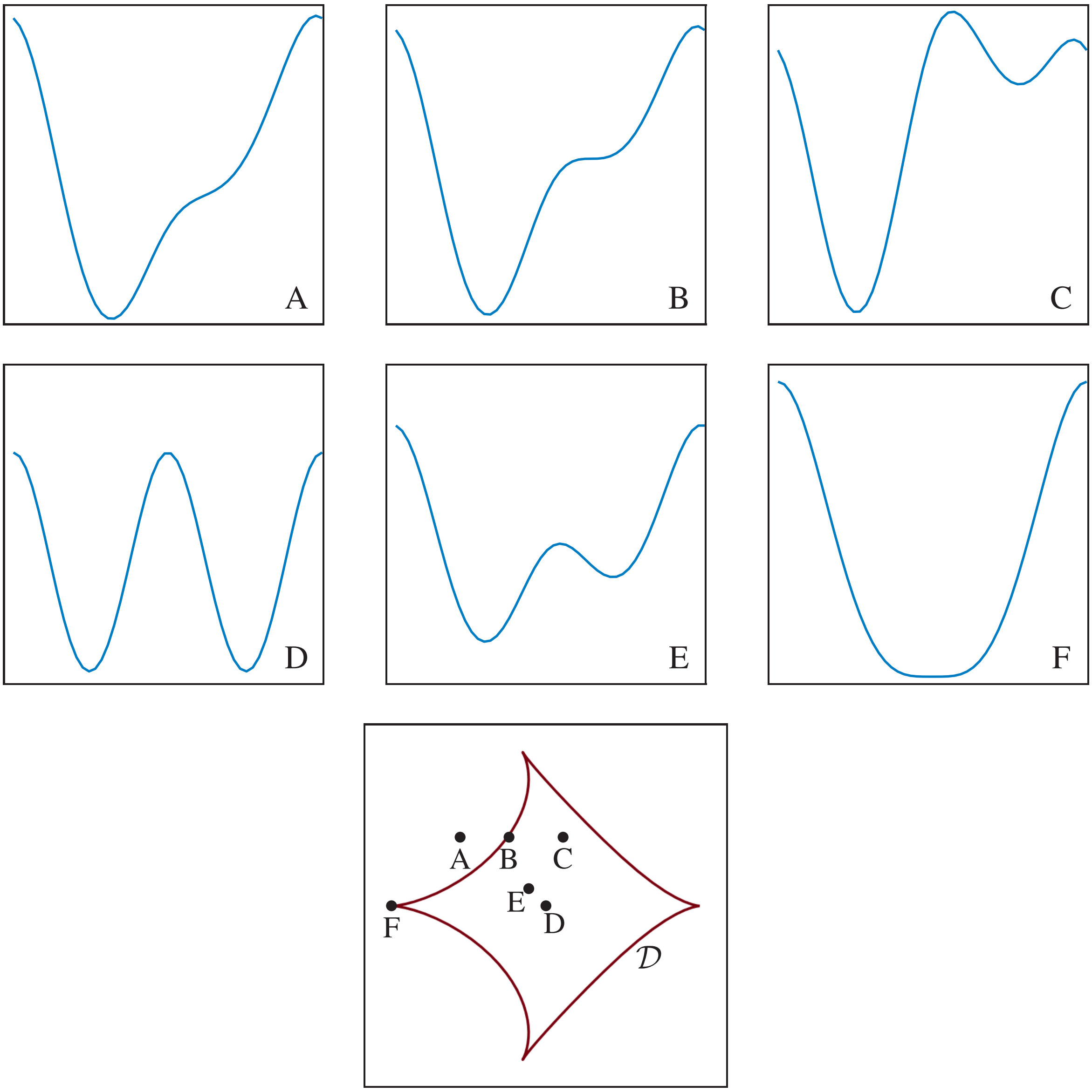


FIGURE 1.4 Graphs of the potential V in Zeeman's catastrophe machine for the six different parameter values shown in the bottom figure. Note that Figures B and F have degenerate critical points, and the corresponding points in the bottom diagram lie on the discriminant $\mathcal{D}$. The horizontal axis in diagrams A–F is $\theta \in [0, 2\pi]$.

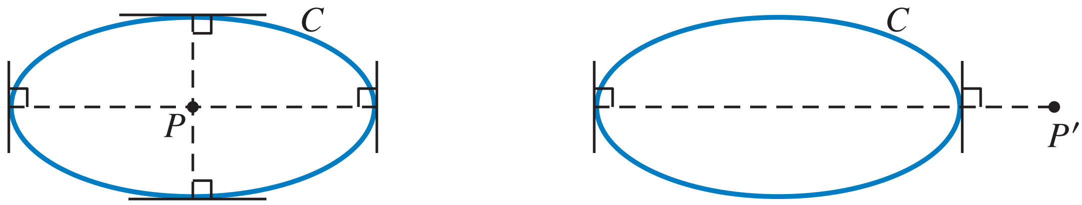


FIGURE 1.5 The dashed lines are perpendiculars from P and P' to the ellipse C.

plane, define the function

$$f_{\mathbf{c}}(t) = -\tfrac{1}{2}\,\|\mathbf{c} - \mathbf{r}(t)\|^2.$$

($\mathbf{c}$ is the position vector of the point P from the discussion above, and the factor $-\frac{1}{2}$ is for convenience.) This is a family of functions of t, with two parameters u and v. It measures the square of the distance from the point $\mathbf{c}$ to the point $\mathbf{r}(t)$; it's called the ***distance squared function***, or distance squared family.

Question: where does $f_{\mathbf{c}}$ have a critical point, and when is it degenerate?

First differentiate $f_{\mathbf{c}}$ (with respect to t),

$$f_{\mathbf{c}}'(t) = (\mathbf{c} - \mathbf{r}(t)) \cdot \dot{\mathbf{r}}(t). \tag{1.2}$$

Since $\dot{\mathbf{r}}(t)$ is the derivative of $\mathbf{r}(t)$, it represents a non-zero tangent vector to the curve. It follows that $f_{\mathbf{c}}'(t) = 0$, if $\mathbf{c}$ lies on the normal line to the curve at $\mathbf{r}(t)$. The set of critical points is therefore a very geometric object:

$$C = \{(t, u, v) \in \mathbb{R}^3 \mid (u, v) \text{ lies on the normal to the curve at } \mathbf{r}(t)\}. \tag{1.3}$$

Thus, for given P the original question is now, how many critical points does $f_{\mathbf{c}}$ have? In particular, the question of how many normals there are for a given point P is now cast as a variational problem.

Local changes in the number of critical points can only occur when a critical point is degenerate (as follows from the implicit function theorem). To see if the critical point is degenerate, we find the second derivative:

$$f_{\mathbf{c}}''(t) = (\mathbf{c} - \mathbf{r}(t)) \cdot \ddot{\mathbf{r}}(t) - \|\dot{\mathbf{r}}(t)\|^2. \tag{1.4}$$

Thus $f_{\mathbf{c}}$ has a degenerate critical point at t if both (1.2) and (1.4) are equal to zero. We can rewrite the two equations as,

$$\begin{cases} \dot{\mathbf{r}}(t) \cdot \mathbf{c} &= \mathbf{r}(t) \cdot \dot{\mathbf{r}}(t) \\ \ddot{\mathbf{r}}(t) \cdot \mathbf{c} &= \mathbf{r}(t) \cdot \ddot{\mathbf{r}}(t) + \|\dot{\mathbf{r}}(t)\|^2 \end{cases}$$

This is simply a pair of linear equations for $\mathbf{c}$, and if the coefficients $\ddot{\mathbf{r}}(t)$ and $\dot{\mathbf{r}}(t)$ are not parallel (they are both vectors), there is a unique solution $\mathbf{c}$, so giving a unique point[1] on that normal line. Call this point $\mathbf{e}(t)$: the resulting curve is called the ***evolute*** of the original curve. We have shown that the point $\mathbf{c} = \mathbf{e}(t)$ if and only if the function $f_{\mathbf{c}}$ has a degenerate critical point at t; the set of $\mathbf{e}(t)$ as t varies is therefore the discriminant of this family $f_{\mathbf{c}}$.

Example 1.1. As a specific example, consider the ellipse

$$\mathbf{r}(t) = (3\cos t,\ 2\sin t).$$

Then, with $\mathbf{c} = (u, v)$,

$$f_{\mathbf{c}}(t) = -\tfrac{1}{2}(u - 3\cos t)^2 - \tfrac{1}{2}(v - 2\sin t)^2. \tag{1.5}$$

The first two derivatives are $f_{\mathbf{c}}'(t) = -3u\sin t + 2v\cos t + 5\sin t\cos t$, and

$$f_{\mathbf{c}}''(t) = -3u\cos t - 2v\sin t + 10\cos^2 t - 5.$$

Solving $f'(t) = f''(t) = 0$ gives

$$u = \frac{5}{3}\cos^3 t, \quad v = -\frac{5}{2}\sin^3 t. \tag{1.6}$$

That is, $\mathbf{e}(t) = \left(\frac{5}{3}\cos^3 t,\ -\frac{5}{2}\sin^3 t\right)$; this curve is shown in Figure 1.6, together with the ellipse (notice that the ellipse is traversed anticlockwise, while the resulting parametrization of the evolute is clockwise). Note that this evolute or discriminant also has 4 cusps, like the ZCM above. We will see in later chapters that cusps occur very often on discriminants for 2 parameter families of functions, and using the theory of unfoldings we will explain why.

If $\mathbf{c}$ lies inside the evolute, the function $f_{\mathbf{c}}$ has 4 critical points, all nondegenerate, and if outside it has just 2. Indeed, using the symmetry of the ellipse if you take $\mathbf{c} = (0, 0)$ it is easy to see the 4 points of the curve for which the normal line passes through $\mathbf{c}$. If, on the other hand, $\mathbf{c}$ lies on the evolute but not at one of the cusps, then $f_{\mathbf{c}}$ has precisely 3 critical points, of which one is degenerate. Finally, if $\mathbf{c}$ lies at a cusp, $f_{\mathbf{c}}$ has a 'doubly' degenerate critical point and a nondegenerate one. As $\mathbf{c}$ varies from the interior of the evolute to the exterior, crossing at a regular point (ie, not at a cusp) then two of the critical points will coalesce and then disappear,

[1]it is in fact the *centre of curvature* of the curve at $\mathbf{r}(t)$; the evolute was originally defined by Huygens in the seventeenth century in his study of the pendulum, and it was later realised to be the locus of centres of curvature.

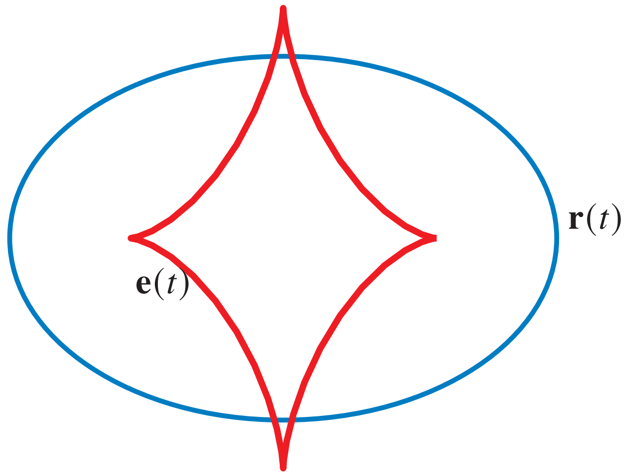


FIGURE 1.6 An ellipse with its evolute

just as they do for the fold family in Section 1.1 and the Zeeman Catastrophe Machine in Section 1.3. The 4 cusps are interesting geometrically: they are points on the evolute (centres of curvature) corresponding to points on the curve where the curvature has a local maximum or minimum.

If the major and minor axes of the ellipse were closer in value (here they are equal to 3 and 2 respectively), the evolute would be smaller, and in the limit as the ellipse tends to a circle, so the evolute tends to a single point: the centre of the circle. ✐

Applications of these ideas to the study of the geometry of curves and surfaces can be found in two books [18] and [61]; there is also a brief discussion in Chapter 15 in this book.

One question arising from the two very different examples, the evolute and Zeeman's catastrophe machine, is why do the bifurcation curves or discriminants have cusps? We will show in later chapters that this is very natural, given that we are studying a 2–parameter family of functions. The fact that in both cases there is only one variable θ or t turns out not to be important: it's the number of parameters that is central.

These two examples are both variational problems (arising from looking for critical points of functions), and such problems will be the study of the first part of this book. Later we will study more general (non-variational) bifurcation problems, but it will turn out that for two parameters, folds and cusps are still all that are to be expected.

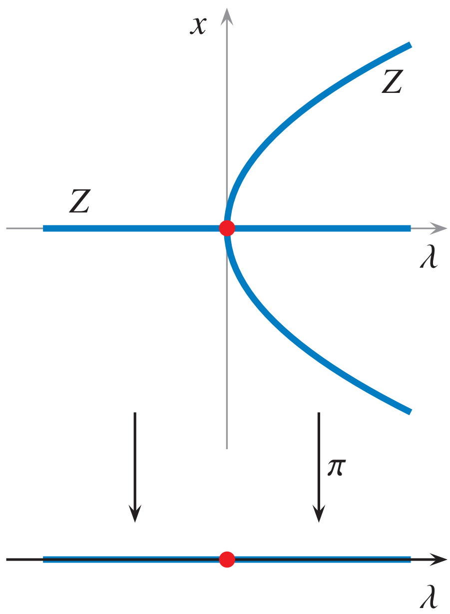


FIGURE 1.7 The pitchfork bifurcation: the zero-set Z consists of two intersecting curves, one of which is folded over relative to the projection.

1.5 Pitchfork bifurcation

Let us now look at a different 1–parameter example, namely the *pitchfork bifurcation*. Consider the family of ODEs,

$$\dot{x} = x^3 - \lambda x.$$

Again λ is the parameter.

This bifurcation often arises in problems where there is an assumed symmetry in the problem: notice that both sides of the equation are odd functions of x.

In this example $g_\lambda(x) = x^3 - \lambda x$, and the zero-set is

$$Z = \left\{(x, \lambda) \in \mathbb{R}^2 \mid x(x^2 - \lambda) = 0\right\}.$$

This is the set shown in the top diagram in Figure 1.7. For $\lambda < 0$ there is just one solution (namely $x = 0$), while for $\lambda > 0$ there are three. Again, the bifurcation point is at $\lambda = 0$ (marked in red), and the map $\pi : Z \longrightarrow \mathbb{R}$ is simply the projection $\pi(x, \lambda) = \lambda$.

In contrast to the fold bifurcation, this pitchfork bifurcation is not robust, or 'structurally stable', in the following sense. Consider the small perturbation given by $\dot{x} = x^3 - \lambda x + \varepsilon$ (where $\varepsilon \in \mathbb{R}$ is a small constant). The new zero-set is depicted in Figure 1.8, and looks very different (structurally different).

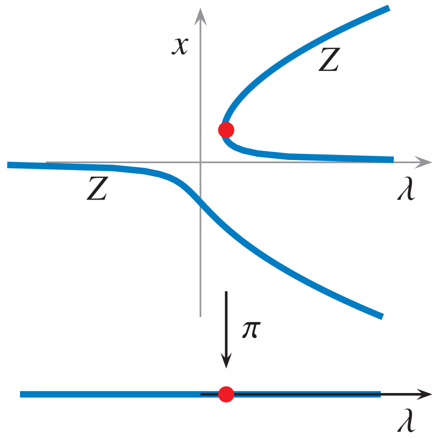


FIGURE 1.8 A perturbation of the pitchfork bifurcation

The symmetry of the pitchfork bifurcation was mentioned above, and the perturbation just given does not preserve this symmetry. It turns out that if only symmetric perturbations are allowed, then the pitchfork does not change qualitatively, and one says it is robust, or structurally stable, with respect to symmetric perturbations. This idea of symmetry breaking is important in applications: see for example the book [60] on *imperfect bifurcations*. These imperfect bifurcations arise by perturbing a 'perfect bifurcation': in this case the pitchfork would be the perfect bifurcation (with perfect symmetry), but in reality one expects systems not to have perfect symmetry and for the symmetry to be broken, leading to the 'imperfect bifurcation' shown in Figure 1.8.

An example of this scenario of pitchfork and perturbation can be seen within the Zeeman catastrophe machine described earlier. While the ZCM is a 2-parameter system, consider just the 1–parameter system where the point (u, v) lies on the line of symmetry $v = 0$, and let $(u_1, 0)$ and $(u_2, 0)$ be the two points where this line meets the bifurcation set $\mathcal{D}$ (the first is marked F in Figure 1.4). As u increases from $u = u_1$, a single solution splits into 3, giving a diagram similar to the one in Figure 1.7. A similar figure can be drawn for u varying through u_2, but with 3 solutions coalescing into 1, so the diagram should be reversed.

Now perturb this path a little, say by raising it (the line $v = 0.1$ say). The new bifurcation diagram near $u = u_1$ would be similar to the one in Figure 1.8, and near $u = u_2$ would be a reversed version, both with a single saddle-node bifurcation.

This idea of considering bifurcations as paths in some parameter space and their perturbations, was introduced in [48] and is central to the so-called *path approach* to bifurcation theory that we study in depth in Part III.

1.6 Conclusions

- A bifurcation problem is a family of equations, which we write $g_\lambda(x) = 0$ where λ is the parameter (in later chapters we will often use $u, v, \dots$ as parameters). This might arise as the equation for equilibria of the differential equation $\dot{x} = g_\lambda(x)$, but it may arise as an equation in its own right (such as for contours). One is interested in how the solutions of $g_\lambda(x) = 0$ change as the parameter is varied.

- A bifurcation is a *qualitative* change in the set of solutions to the bifurcation problem, as the parameter is varied. For example, if there are finitely many solutions, then a bifurcation occurs where the number of solutions changes, while if the solution set is a curve or surface then a bifurcation would occur when the topology of the set changes. The set of parameter values where a bifurcation occurs is called the *bifurcation set* or the *discriminant*.

- In many applications, a bifurcation problem arises as an equation for critical points of a function (of one or several variables). Such equations are called *variational problems*. So our $g_\lambda(x)$ would be equal to $\frac{\partial}{\partial x}V_\lambda(x)$ for some 'potential' function $V_\lambda(x)$. These variational problems form the part of bifurcation theory we call ***catastrophe theory***, and are the subject of the first part of this book. This is for both historical and technical reasons. Catastrophe theory was introduced by René Thom in the 1960s and the use of similar techniques in more general bifurcation theory was developed a decade or so later, and uses many of the ideas Thom introduced. Moreover, the techniques for variational problems are more straightforward.

The principal aim of this text is to show that there are only a few basic models for bifurcations, provided there are not too many parameters. For example, if there is just a single parameter, and no other restrictions, then the only bifurcation arising 'generically' is the saddle-node, or fold bifurcation described in Section 1.1, while if there are two parameters with no further restriction then there are two possibilities: the fold and the cusp (a further restriction might be something like the problem having some symmetry as described above). These ideas were first developed in the 1960s by the French mathematician and Fields medal winner René Thom.

A short account of early bifurcation theory can be found in the book by Drazin [33], who discusses how the first bifurcation problem to be studied was by Euler in the eighteenth century, who addressed what is now called the Euler beam problem, where a beam buckles under a load and as the load increases there is a (pitchfork) bifurcation.

Problems

1.1 Consider the following perturbation of the ODE given in Section 1.1, namely $\dot{x} = x^2 + \lambda + \varepsilon x$ where $\varepsilon \in \mathbb{R}$ is a small consant. Show that the bifurcation diagram is similar to the one in Figure 1.1A, but with discriminant equal to $\{\varepsilon^2/4\}$.

1.2 Investigate the bifurcations in the set of equilibria occurring as λ varies, in the family $\dot{x} = x^2 + \lambda^2 - 1$. In particular, show the bifurcation set consists of two points on the λ-axis.

1.3 Investigate the bifurcations in the set of equilibria occurring as λ varies, in the two pitchfork families of ODEs,

$$\dot{x} = \pm(x^3 - \lambda x),$$

showing the phase diagram analogous to Figure 1.1B. Notice that in both of these, the origin is attracting (stable) for λ on one side of the origin and and repelling (unstable) on the other. Correspondingly, the bifurcating equilibria are stable in one of these and unstable in the other. When they are stable it is called a *supercritical* pitchfork bifurcation, and when they are unstable it is a *subcritical* pitchfork bifurcation. (This is the only place in the book we consider dynamical properties such as stability and instability.) (†)

1.4 Investigate the contours of the height function $h(x, y) = 2x^4 + 4y^4 - x^2 + y^2$, and relate the transitions (bifurcations) to what was seen in Section 1.2, specifically in Figure 1.2. [*Hint*: *find the critical points, and hence critical values, and then use a graphing calculator, or Wolfram Alpha on the internet. For the latter you can enter an instruction such as,* `plot 2*x^4+4*y^4-x^2+y^2=0` *or* `contour plot 2*x^4+4*y^4-x^2+y^2`].

1.5 Investigate the level sets (contours) of the function $h_u(x) = x^3 - 3ux$ for different values of u. (Here the level sets are finite sets of points, so the question is, how many points in each level set? The answer depends on the value of c as well as u.)
[*Hint*: *begin by sketching the curve $y = h_u(x)$ for $u > 0, u = 0$ and $u < 0$. You will see that bifurcations occur at critical points of h_u (i.e., $\frac{\mathrm{d}}{\mathrm{d}x}(h_u) = 0$). Find the locus of points in the (u, c)-plane where these occur – that is the discriminant.*] (†)

1.6 Here we show the *robustness* of the saddle-node bifurcation. Consider the saddle-node bifurcation $g_\lambda(x) = x^2 + \lambda = 0$ shown in Figure 1.1A. Let $G(x, \lambda, u)$ be any smooth 'perturbation' of g; that is suppose G is a smooth function (with $u \in \mathbb{R}$) and that $G(x, \lambda, 0) = g_\lambda(x)$. Note that one cannot solve $g(x, \lambda) = 0$ for x as a function of λ, but one can instead solve for λ as a function of x, and this function has a nodegenerate extremum at the bifurcation point.

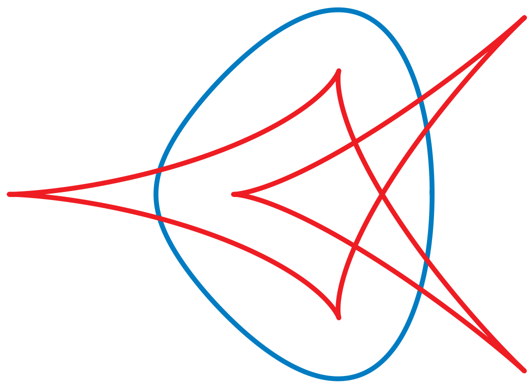
Use the implicit function theorem to show that for sufficiently small values of u, the perturbed bifurcation problem g_u given by $g_u(x, \lambda) = G(x, \lambda, u)$ (for u fixed) also has a saddle-node bifurcation, in the sense that one can solve (locally) for λ as a function of x (and u) and for each u this function λ has a nondegenerate extremum. (†)

1.7 In contrast to the saddle-node bifurcation the so-called *hysteresis bifurcation* is not robust. Consider $\dot{x} = x^3 + \lambda$, and its perturbation $\dot{x} = x^3 + \lambda + ux$. Sketch three phase portraits for this system similar to Figure 1.1B, one with $u > 0$, one with $u = 0$ and with with $u < 0$ (u fixed in each case). [See Figure 18.3C on p. 226 for the curves of equilibria.]

1.8 Investigate the change in contours for the family of surfaces $h(x, y, z) = x^2 + y^2 - z^2 = c$, distinguishing $c = 0$, $c > 0$ and $c < 0$. (†)

1.9 Find the evolute of the parabola $y = x^2$, and show it has a single cusp. On a sketch, show that from a point inside (or above) the evolute there are three lines perpendicular to the parabola, while from a point outside there is only one.

1.10 The diagram below shows a curve (in blue) and its evolute (in red). For a point P in each of the five components of the complement of the evolute find the number of perpendiculars to the curve from P.



[*Hint*: (1) *If P is far from the curve there are only two perpendiculars, and* (2) *as P crosses a smooth point of the evolute, two solutions are either created or destroyed (depending on the direction of crossing), like in saddle-node bifurcations.*] (†)

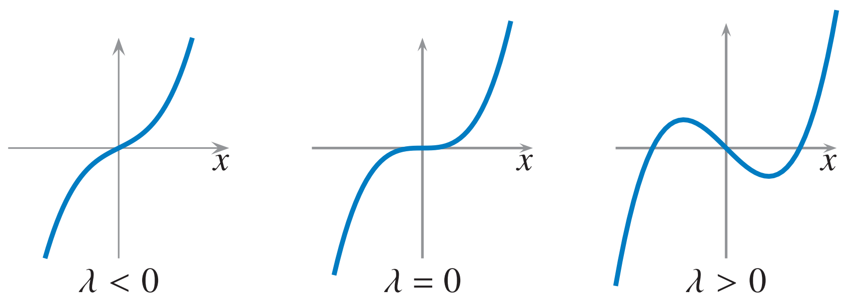
Part I

Catastrophe theory

2

Families of functions

CATASTROPHE THEORY is the study of families of functions, and in particular of their critical points, the principal motivation being to study bifurcations in variational problems.[1] For example the function considered in Problem 1.5, namely $f_\lambda(x) = x^3 - 3\lambda x$ has a single (degenerate) critical point when $\lambda = 0$, it has two critical points when $\lambda > 0$ and none at all when $\lambda < 0$ (or two complex ones if one prefers to include those). On the other hand if a family has only nondegenerate critical points when $\lambda = 0$, then for nearby values of λ it will still have nondegenerate critical points and they will be close to the original ones. We will prove this later in the book.



In most areas of mathematics the words 'function' and 'map' are more or less interchangeable; however it is traditional in this area to reserve the word 'function' to refer to scalar-valued functions (so $f: \mathbb{R}^n \to \mathbb{R}$ or $f: \mathbb{C}^n \to \mathbb{C}$), while 'map' refers to an $f: \mathbb{R}^n \to \mathbb{R}^p$ with $p > 1$, or the complex analogue.

2.1 Critical points

We consider smooth functions $f: \mathbb{R}^n \to \mathbb{R}$: here smooth means infinitely differentiable, although in practice, smooth just means 'as many times differentiable as is needed in the current context', so for example if we are talking about the second

[1]A *variational problem* is one involving finding critical points of functions. In applications these are often, but not always, required to be local maxima or minima.

derivative of f, we will need to assume it is twice differentiable, and the second derivatives are continuous (in short, f is of class C^2). However, we will make the blanket assumption that all functions we consider are infinitely differentiable.

In general, our function may not be defined on all of $\mathbb{R}^n$, and we adopt the notation $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}$ (notice the different arrow) to mean a function of n variables whose domain is some open subset of $\mathbb{R}^n$. This saves us writing things like, 'let $f\colon U \to \mathbb{R}$ be a smooth function, where U is an open subset of $\mathbb{R}^n$', and instead we write 'let $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}$ be a smooth function'. If we need to refer to the domain of such an f we will denote it $\mathrm{dom}(f)$.

Definition 2.1. A smooth function $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}$ has a ***critical point*** at x_0 if its differential there vanishes: $\mathsf{d}f_{x_0} = 0$. ✯

Here $\mathsf{d}f_{x_0}$ is the ***differential*** of f at x_0, so

$$\mathsf{d}f_{x_0} = \left(\frac{\partial f}{\partial x_1}(x_0),\ \frac{\partial f}{\partial x_2}(x_0),\ \ldots,\ \frac{\partial f}{\partial x_n}(x_0)\right).$$

Thus x_0 is a critical point of f if all n partial derivatives of f vanish at x_0:

$$\frac{\partial f}{\partial x_1}(x_0) = \frac{\partial f}{\partial x_2}(x_0) = \cdots = \frac{\partial f}{\partial x_n}(x_0) = 0.$$

Critical points are also known as *singular points* of the function.

Definition 2.2. If x_0 is a critical point of $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}$, then the ***Hessian*** matrix of f at x_0 is the symmetric $n \times n$ matrix of second partial derivatives,

$$H_f(x_0) = \mathsf{d}^2 f_{x_0} = (h_{ij})$$

where

$$h_{ij} = \frac{\partial^2 f}{\partial x_i\, \partial x_j}(x_0).$$

A critical point x_0 of f is ***nondegenerate*** if $\det H_f(x_0) \neq 0$: otherwise it is ***degenerate***. ✯

Example 2.3. Find the critical points of $f(x, y, z) = x^3 - xy^2 + 3x^2 + y^2 + z^2$, and determine whether each is nondegenerate.

Solution: Differentiating f with respect to each of the three variables gives the equations

$$3x^2 - y^2 + 6x = 0, \quad -2xy + 2y = 0, \quad 2z = 0.$$

The solutions, 4 in all, are found to be

$$(0,0,0), \quad (-2,0,0), \quad (1,3,0), \quad (1,-3,0).$$

The second differential (Hessian matrix) of f is

$$\mathsf{d}^2 f(x,y,z) = \begin{pmatrix} 6x+6 & -2y & 0 \\ -2y & -2x+2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

The Hessian at each of the critical points is,

$$\begin{pmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} -6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 12 & -6 & 0 \\ -6 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 12 & 6 & 0 \\ 6 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

A quick inspection shows that all four matrices are invertible, and hence all four critical points are nondegenerate. ✐

Recall that the eigenvalues of a symmetric matrix are all real. Moreover, at a critical point, the Hessian matrix is 'intrinsic' in the following sense. Suppose f has a critical point at the origin and let ϕ be a change of coordinates about the origin (i.e. such that $\phi(0) = 0$). Write H_f for the Hessian matrix of f at the origin in the original coordinates, and H'_f for the matrix in the transformed coordinates. Then by the chain rule for second derivatives (see Proposition A.10) $\mathsf{d}^2(f\circ\phi)_0(\mathbf{u}^2) = \mathsf{d}^2 f_0(\mathsf{d}\phi_0\mathbf{u})^2$ (using $\mathsf{d}f_0 = 0$), whence

$$\Phi^T H_f \Phi = H'_f, \tag{2.1}$$

where $\Phi = \mathsf{d}\phi_0$ (the Jacobian matrix of ϕ).

Definition 2.4. If $x_0 \in \mathbb{R}^n$ is a nondegenerate critical point of the smooth function f, then its ***index*** is the number of negative eigenvalues of the Hessian matrix, counting multiplicity. ✩

It is important to realize that while the eigenvalues of the Hessian do depend on the chosen basis (or coordinates), their signs do not; in particular the index of a critical point does not depend on the chosen coordinates, which follows from (2.1). For a symmetric $n \times n$ matrix, there can be anywhere between 0 and n negative eigenvalues, so the index of a nondegenerate critical point in n variables lies between 0 and n inclusive. The term 'counting multiplicity' means that for example if an eigenvalue is a double root of the characteristic polynomial of the matrix, then it should count twice; that way one always has exactly n eigenvalues 'counting multiplicity'. In particular, the identity matrix has eigenvalue 1 with multiplicity n, and the index of the critical point of the function $-x_1^2 - \cdots - x_n^2$ is n.

In Example 2.3 above, the origin is of index 0, while the other critical points are all of index 1. The index of a nondegenerate critical point is an important invariant and determines the geometry of the level sets of f near the critical point. For example, a point of index 0 is a local minimum of the function, and one of index n is a local maximum (if n is the number of variables). This follows from the following important result which we will return to later (Section 4.4).

Morse Lemma *Let $p \in \mathbb{R}^n$ be a nondegenerate critical point of f of index k. Then there is a change of coordinates $x = \phi(y)$ near p such that in these new coordinates y_i the function has the form*

$$f(y) = f(p) - y_1^2 - y_2^2 - \cdots - y_k^2 + y_{k+1}^2 + \cdots + y_n^2.$$

This is a particularly simple form of Taylor series for the function: since p is a critical point the linear terms in the Taylor series at p must all vanish, so this lemma is saying that if the critical point is nondegenerate then in some local coordinate system the Taylor series is purely quadratic.

Complex functions A complex analytic function $f\colon \mathbb{C}^n \to \mathbb{C}$ also has critical points, at points where all partial derivatives vanish, and one can form the Hessian matrix, but whose eigenvalues may now be complex. However, the notion of index is meaningless: firstly because the eigenvalues of the Hessian may not be real, and secondly if they are real, then a change of coordinates can change their sign: for example the function x^2 becomes $-y^2$ upon substituting $x = iy$.

2.2 Degeneracy in one variable

The story is fairly simple in one variable. Suppose $f\colon \mathbb{R} \rightarrowtail \mathbb{R}$ is a smooth function of a single variable. Then a point $x_0 \in \mathrm{dom}(f)$ is a critical point of f if $f'(x_0) = 0$. This critical point is nondegenerate if $f''(x_0) \neq 0$, otherwise it is degenerate. And one can continue looking at higher and higher derivatives as follows.

Definition 2.5. A critical point of a smooth function f of a single variable is of ***type*** A_k if the first k derivatives of f all vanish at that point, but the $(k+1)^{\text{th}}$ does not. ☆

The same definition is made for critical points of complex analytic functions $f\colon \mathbb{C} \to \mathbb{C}$. Thus in particular, a critical point (real or complex) of type A_1 is a nondegenerate critical point, and one sometimes says that a degenerate critical point is of type $\mathsf{A}_{\geq 2}$, or of type *at least* A_2.

Example 2.6. Consider the simple monomial $f(x) = x^{k+1}$, with $k \geq 1$ and $x \in \mathbb{R}$ or $\mathbb{C}$. Then

$$f'(0) = f''(0) = \cdots = f^{(k)}(0) = 0, \quad \text{but} \quad f^{(k+1)}(0) = (k+1)! \neq 0.$$

It follows that the function $x \mapsto x^{k+1}$ (whether real or complex) has a critical point of type A_k at $x = 0$. ✐

We will see in Chapter 4 that any function of one variable with a critical point of type A_k is *equivalent* to x^{k+1} (in a neighbourhood of the critical point), in the sense that there is a local change of coordinates that turns the given function into the appropriate monomial.

We will see later the corresponding definitions for functions of several variables; note for now that any nondegenerate critical point is said to be of type A_1, for any number of variables.

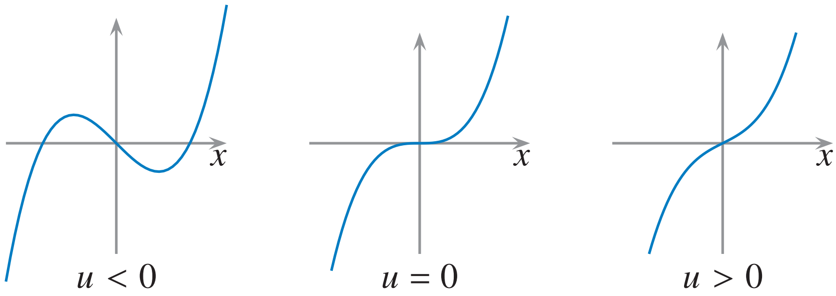
2.3 Families of functions

We begin with a simple example, demonstrating the issues of interest.

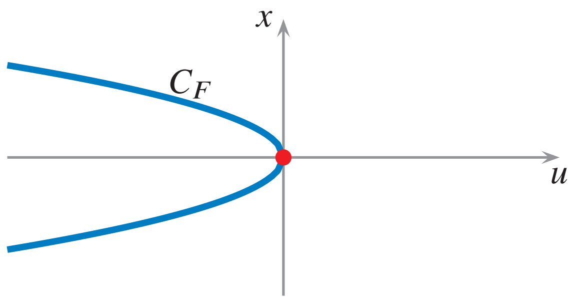
Example 2.7. To set the scene, consider the 1–parameter family of functions

$$F(x;u) \;=\; f_u(x) \;=\; x^3 + ux,$$

parametrized by $u \in \mathbb{R}$. This is called the ***fold family***, or ***fold catastrophe***. Note that when $u = 0$, the function $f_0(x) = x^3$ which has a critical point of type A_2 at the origin. The 3 figures below show graphs of the function f_u for $u < 0$, $u = 0$ and $u > 0$ respectively. Notice that there are 2, 1 and 0 critical points in the three figures, and as u increases from negative values to 0 so the 2 critical points coalesce and then disappear for $u > 0$. At the point of coalescing the critical point is degenerate.



To visualize this better, we plot below the set of critical points in (x, u)-space:



The curve $C_F = \{(x, u) \mid 3x^2 + u = 0\}$ is the set of critical points of f_u (since $(f_u)'(x) = 3x^2 + u$), and for each $u < 0$ there are two corresponding values of x (equal to $\pm\sqrt{-u/3}$), which coalesce when $u = 0$ and then for $u > 0$ there are no points on C_F (or at least, no real ones). The red dot is the point where the two critical points coalesce and in general is called the *singular set* of the family of functions, while C_F is called the *catastrophe set*. We will define these more generally below.

The reason this family is called the fold family is simply that C_F is folded over compared to the parameter space. This is the same as the saddle-node bifurcation described in Section 1.1. ✐

In general we consider a family of functions, $f_u(x)$, with $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^a$. This needs to be smooth in u as well as x, so altogether we require that the map (function),

$$f : \mathbb{R}^n \times \mathbb{R}^a \longrightarrow \mathbb{R}, \quad F(x, u) = f_u(x)$$

be smooth as a function of $n + a$ variables. We call these ***smooth families*** of functions.

Convention: As a rule, we will use x, y, z or $x_1, x_2, x_3, \ldots$ as variables, and u, v, w or $u_1, u_2, u_3, \ldots$ as parameters. A useful convention is to use a semicolon and write $F(x; u)$ instead of $F(x, u)$ to distinguish variables from parameters.

Definition 2.8. A smooth a–parameter ***family of functions*** on $\mathbb{R}^n$ is a smooth map

$$\begin{array}{rcl} f\colon \mathbb{R}^n \times \mathbb{R}^a & \longrightarrow & \mathbb{R} \\ (x,\, u) & \longmapsto & f_u(x). \end{array}$$

The notation $f_u(x)$ reflects the interpretation that u is a parameter and x the variable, so we are interested in the behaviour (especially critical points) of the function f_u for each u and how this varies as u changes. As already mentioned, we often write $F(x;u)$ rather than $F(x,u)$ to emphasize this distinction.

In applications x is often called the ***state variable***.

Since we are interested in the critical points of functions, it is natural to make the following definition. Let $f\colon \mathbb{R}^n \times \mathbb{R}^a \longrightarrow \mathbb{R}$ be a smooth family of functions, as above, and define the ***catastrophe set*** of F to be

$$C_F = \{(x,u) \in \mathbb{R}^n \times \mathbb{R}^a \mid \mathrm{d}(f_u)_x = 0\}\,.$$

In other words, it is the set of points $(x,u) \in \mathbb{R}^{n+a}$ such that f_u has a critical point at x:

$$\frac{\partial f_u}{\partial x_1}(x) = \cdots = \frac{\partial f_u}{\partial x_n}(x) = 0.$$

In many important cases, this will be a subset of dimension equal to a (the number of parameters), as we shall see.

An important subset of the catastrophe set is the ***singular set***, denoted Σ_F, which is the set of points in C_F where the critical point is *degenerate*:

$$\Sigma_F := \{(x,u) \in C_F \mid \det(H_{f_u}(x)) = 0\}.$$

Finally, we define the ***discriminant*** or ***bifurcation set*** Δ_F. This is a subset of the set of parameters, equal to the set of parameter values u for which f_u has a degenerate critical point. In other words, if we let π_F be the map projecting C_F onto the parameter space,

$$\begin{array}{rcl} \pi_F\colon C_F & \longrightarrow & \mathbb{R}^a \\ (x,u) & \longmapsto & u, \end{array}$$

then

$$\Delta_F = \pi_F(\Sigma_F) = \{u \in \mathbb{R}^a \mid \exists x \in \mathbb{R}^n,\ (x,u) \in \Sigma_F\}.$$

☆

The reason Δ_F is called the *bifurcation set* is that it is where transitions (or bifurcations) take place. The idea is that if $u_0 \notin \Delta_F$ then there is a neighbourhood U of u_0 such that $\forall v \in U$, the number of critical points of f_v is the same as for f_{u_0}. Thus Δ_F divides $\mathbb{R}^a$ into a number of connected components, and in each of these the number of critical points is constant. We will see this in examples.

In Example 2.7 we saw that the catastrophe set is the parabola

$$C_F = \{(x, u) \in \mathbb{R}^2 \mid 3x^2 + u = 0\}$$

and $\Sigma_F = \{(0, 0)\}$ so $\Delta_F = \{0\} \subset \mathbb{R}$ (ie, the point $u = 0$). The complement of Δ_F in $\mathbb{R}$ consists of two components $\{u < 0\}$ and $\{u > 0\}$, and the number of critical points is 2 and 0 respectively. Notice that C_F can be parametrized by x, with $u = -3x^2$; we will find that catastrophe sets often have good parametrizations.

2.4 Cusp catastrophe

The ***cusp family*** is the 2–parameter family given by

$$F(x; u, v) = \tfrac{1}{4}x^4 + \tfrac{1}{2}ux^2 + vx$$

(the coefficients of $1/4$ and $1/2$ are for convenience). With $u = v = 0$ the function $f_{(0,0)}$ has a critical point at the origin of type A_3 (see Definition 2.5). The catastrophe set C_F is given by the equation $f'_{(u,v)}(x) = 0$, so is given by

$$C_F = \left\{(x, u, v) \in \mathbb{R}^3 \,\middle|\, x^3 + ux + v = 0\right\}.$$

This can be parametrized by (x, u) with $v = -x^3 - ux$ and is the curved sheet in Figure 2.1. In this parametrization, the projection $\pi_f : C_F \longrightarrow B = \mathbb{R}^2$ is given by

$$\pi_F(x, u) = (u, v) = (u, -x^3 - ux). \qquad (2.2)$$

Now consider the Hessian matrix at the critical points (this is why it is useful to be able to parametrize C_F):

$$H_f = f''_{(u,v)}(x) = 3x^2 + u.$$

This means that the singular set Σ_F is

$$\Sigma_F = \left\{(x, u, v) \in C_F \,\middle|\, 3x^2 + u = 0\right\}.$$

So on Σ_F we have $u = -3x^2$. Given that $v = -x^3 - ux$ on C_F we have that

$$\Sigma_F = \left\{(x, u, v) \,\middle|\, u = -3x^2,\ v = 2x^3\right\}.$$

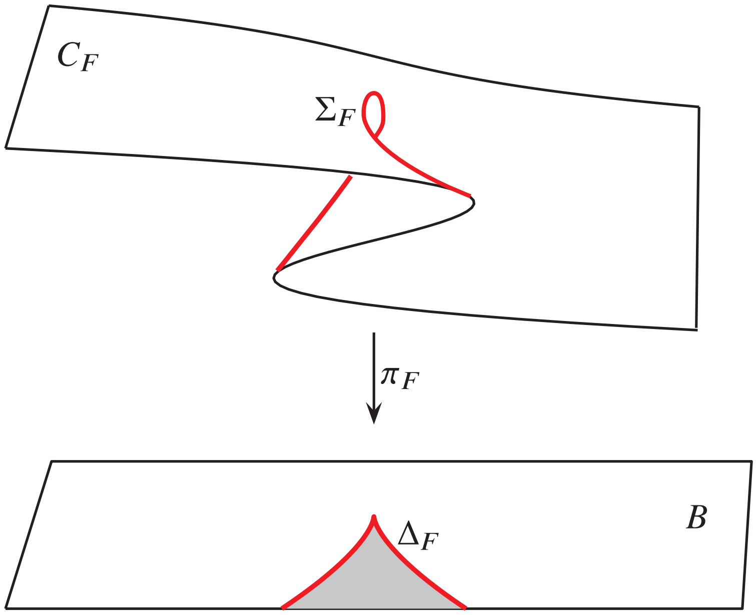


FIGURE 2.1 The cusp catastrophe: x is vertical.

This is a smooth curve in $\mathbb{R}^3$ (parametrized by x) – see the red curve in the top sheet of Figure 2.1.

The image of this singular set in $B = \mathbb{R}^2$ is,

$$\Delta_F = \left\{(u, v) \in \mathbb{R}^2 \,\middle|\, \exists x,\ u = -3x^2,\ v = 2x^3\right\}$$

which is the *discriminant* of F. Eliminating x gives a curve with equation

$$\left(\frac{u}{3}\right)^3 + \left(\frac{v}{2}\right)^2 = 0,$$

or $27v^2 + 4u^3 = 0$, which is the classical ***semicubical parabola*** or ***cusp***, and is why this is called the cusp family. See the red curve on the bottom sheet B in Figure 2.1.

Figure 2.2 shows the graph of $f_{(u,v)}$ for a few different points $(u, v) \in B$. Over points in the grey region in that and Figure 2.1, π_F has 3 preimage points, and so $f_{(u,v)}$ has 3 critical points, and over the other points it has 1, except along the discriminant, where it has precisely two (one of which is degenerate).

In both the Zeeman Catastrophe Machine and the evolute of the ellipse, described in Chapter 1, the catastrophe set has 4 cusp catastrophes occurring in different places, and this is a challenge to visualize globally.

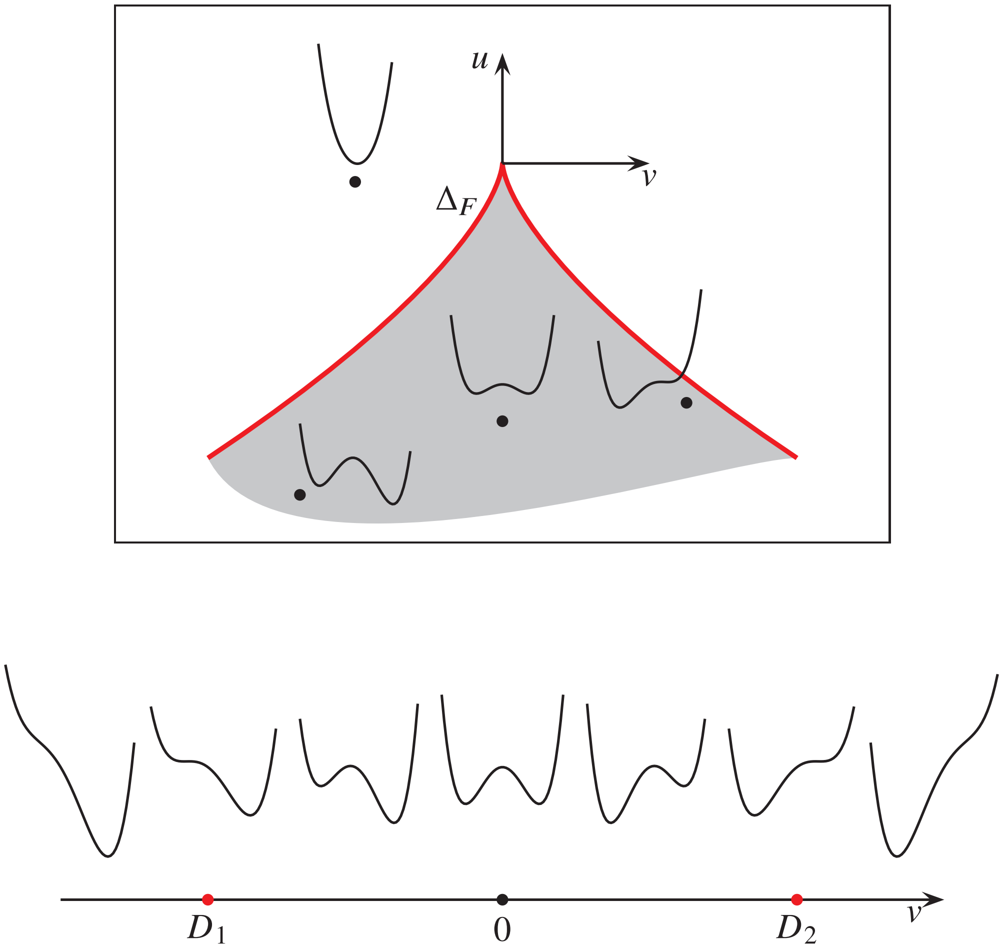


FIGURE 2.2 (TOP) The shape of the graph of $f_{(u,v)}$ in the cusp family, for different values of (u, v). (BOTTOM) The potential function $f_{u,v}(x)$ for $u = -1$ and different values of v. The points D_1 and D_2 represent the points where the line $u = -1$ crosses the discriminant.

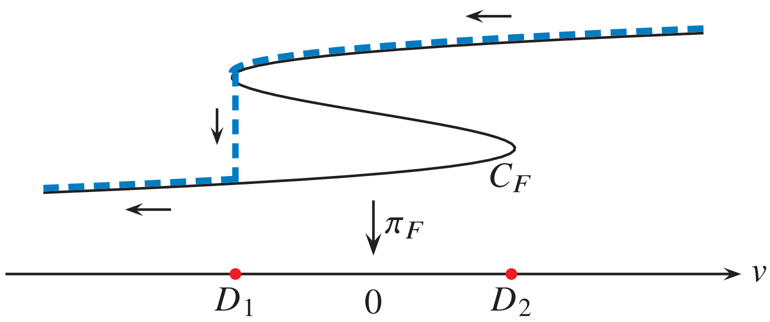


FIGURE 2.3 The effect of moving around the cusp point in the cusp catastrophe. See the text for discussion.

2.5 Why ‘catastrophes’

In his fundamental work [111], René Thom describes his approach to ‘qualitative’ modelling, which we discuss in the preface. He is particularly concerned with the origin of and changes in form or shape, or *morphogenesis*. He covers examples of forms that are physical, biological (at many different scales), or linguistic. A particular assumption in Thom’s work is that a form is selected by minimizing some potential function, which justifies the particular interest in critical points of functions.

Thom writes about *regimes of conflict*, which occur where there are coexisting local minima, and a ‘catastrophe’ occurs when the state of the system jumps from one local minimum to another. There are two types of conflict point, firstly where there are more than one critical point with the same critical value (the importance of these is attributed to Maxwell) and secondly where a local minimum disappears after collision with a saddle point and local maximum, as shown in the figure on p. 19, which is a (local) bifurcation point (a saddle-node or fold bifurcation). See also the changes that occur in Figure 2.2 as v crosses D_1 or D_2; the origin on that axis $v = 0$ corresponds to the Maxwell point.

The catastrophic jump resulting from the local bifurcation is illustrated in Figure 2.3 which represents a tour across (or around) the cusp catastrophe. Compare also with Figs 2.1 and 2.2. The curve in Figure 2.3 is a section through the catastrophe set C_F in Figure 2.1 with $u < 0$ constant (a similar picture is obtained from a path around the cusp point). If the parameters (u, v) are outside the shaded region within the discriminant (say on the side with $v > 0$) then there is only one possible state for the system (the unique minimum in the bottom right graph in Figure 2.2). Call this the top sheet. Now as the parameters vary to cross the discriminant (e.g.

v is decreased), then the state changes slowly as the parameter is varied, until the parameter passes the point D_1, where there is no more 'top sheet', and the system is forced to jump to the other equilibrium, the one on the 'bottom' sheet: this is a discontinuous, or 'catastrophic', change. Reversing the path in parameter space, the system would remain on the lower sheet until the parameter reaches the point D_2, when it would again jump to the upper sheet, giving a hysteresis effect.

Thom's remarkable work shows that for a small number of parameters (he considered up to four), there are only a few local models describing how these bifurcations occur. Without further constraints (such as symmetries, as in the pitchfork bifurcation discussed in Chapter 1), the only structurally stable bifurcations are named as follows:

- for 1 parameter, only the fold (saddle-node),
- for 2 parameters, the cusp,
- for 3 parameters, the swallowtail and the elliptic and hyperbolic umbilics,
- for 4 parameters, the butterfly and the parabolic umbilic.

These seven local models are called the *elementary catastrophes* by Thom, and we discuss these in more detail in Chapter 7 on 'unfoldings'.

Remark 2.9. The two types of conflict set described above give rise to two notions of bifurcation set. The one we have descibed is sometimes called the *local* bifurcation set, to distinguish from the *full* bifurcation set. The latter includes points in parameter space (the Maxwell set) where two or more critical *values* coincide, where a critical value is a value $f(x)$ where x is a critical point. Which type of bifurcation set is relevant to a particular problem depends on the context. Except for a brief metion in the final chapter of the book, we only consider the local bifurcation set as defined above. ❞

Problems

2.1 Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be the function $f(x,y) = x^4 - 2x^2 + y^2 + 11$. Find all the critical points of f, and determine whether each is nondegenerate, and if it is find its index.

2.2 Repeat the previous question for the function $f\colon \mathbb{R}^3 \to \mathbb{R}$ given by

$$f(x,y,z) = \tfrac{1}{4}x^4 + x^3 - x^2 + 2xy - y^2 - z^2.$$

2.3 The functions $f(x) = x^2\sin^5(x)$ and $g(x) = x^2\cos^5(x)$ have critical points at the origin. What are their types (ie A_k for which k)? (†)

2.4 Sketch the catastrophe set of the family $F(x;u) = x^3 - 3u^2x$. What is the singular set for this family? (†)

2.5 Let $F(x, y;\, u, v) = \frac{1}{3}x^3 - xy^2 - u(x^2 + y^2) - vx$. Find C_F, Σ_F and Δ_F. For each component of the complement of the discriminant, find the number of critical points of $f_{(u,v)}$. What are their indices?

2.6 Let $F(x;u) = \frac{1}{4}x^4 - \frac{3}{2}u^2x^2 + 2u^3x$. Sketch the catastrophe set C_F [*Hint*: *the expression for* C_F *factorizes*]. Find Σ_F, and show that $\Delta_F = \mathbb{R}$.

Part II

Singularity theory

Part III

Bifurcation theory

Part IV

Appendices

B

Local geometry of regular maps

THROUGHOUT THIS TEXT, WE consider maps from $\mathbb{R}^n$ to $\mathbb{R}^p$, and the purpose of this chapter is twofold. Firstly, in order to understand singularities one should first understand non–singularities – that is points where the rank of the Jacobian matrix of the map takes its maximal possible value. The local geometry in the neighbourhood of such a point is governed by the inverse function theorem and its brethren such as the local immersion and local submersion theorems. Secondly the inverse function theorem is fundamental to singularity theory through the resulting changes of coordinates. The proof we give (in Chapter 5) of the inverse function theorem involves existence and uniqueness theorems for ordinary differential equations, and the resulting idea of flows, all of which is covered in Appendix C. The principal reason for giving the proof is that one of the main theorems of the subject (the finite determinacy theorem of Chapter 5) has an analogous, if slightly more complex, proof.

We use the notation $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ to mean that f is a map whose domain is an open subset of $\mathbb{R}^n$.

Definition B.1. Let $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ be a smooth map. The ***rank of the map*** at a point x is defined to be the rank of its Jacobian matrix $\mathsf{d}f_x$ at that point, and denoted $\mathrm{rk}_x(f)$. ✰

The Jacobian matrix $\mathsf{d}f_x$ is a $p \times n$ matrix so $\mathrm{rk}_x(f) \leq \min\{n, p\}$. This chapter is principally about the structure of maps f near points x where $\mathrm{rk}_x(f) = \min\{n, p\}$.

B.1 Changes of coordinates and diffeomorphisms

In linear algebra, a change of basis is implemented by an invertible linear map. If $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for V, and $A \in \mathsf{GL}(n)$, then $\{\mathbf{e}'_1, \ldots, \mathbf{e}'_n\}$ with $\mathbf{e}'_i = A\mathbf{e}_i$ is another basis (because A is invertible). The coordinates of a point (vector) $\mathbf{v} \in V$ with respect to the basis $\{\mathbf{e}_i\}$ are the coefficients x_i appearing in the expression

$\mathbf{v} = \sum x_i \mathbf{e}_i$. In changing to the new basis the coordinates are also changed, to the coefficients x'_i in the expression $\mathbf{v} = \sum x'_i \mathbf{e}'_i$. And the two are related by $\mathbf{x} = A\mathbf{x}'$, that is $x_i = \sum_j a_{ij} x'_j$, so $x'_j = \sum_j b_{ij} x_j$, where $B = A^{-1}$.

On the other hand, when we are working with smooth maps and ignoring the linear structure on $\mathbb{R}^n$, we don't use bases so much as just the coordinates. This is because it is useful to consider nonlinear changes of coordinates which do not respect the nature of bases. Nonlinear changes of coordinates should be familiar from multiple integrals, for example changing from Cartesian coordinates to polar coordinates.

To define coordinates systems on $\mathbb{R}^n$, or on an open set $U \subset \mathbb{R}^n$, one can start with the usual Cartesian one on $\mathbb{R}^n$ and then say what constitutes a change of coordinates. The fundamental notion is the *diffeomorphism*.

Definition B.2. Let $U, V \subset \mathbb{R}^n$ be open sets. A smooth map $f\colon U \to V$ is a ***diffeomorphism*** if it has a smooth inverse; that is, if there is a smooth map $g\colon V \to U$ such that $f \circ g = \mathsf{Id}_V$ and $g \circ f = \mathsf{Id}_U$ (here Id_U is the identity map on U, and Id_V that on V). In this case one writes $g = f^{-1}$. ✯

Examples B.3.

(i). Let $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. This is not invertible so is certainly not a diffeomorphism.

(ii). Restrict this map to $f\colon (0,\infty) \to (0,\infty)$, with again $f(x) = x^2$. Now f *is* invertible, with $f^{-1}(y) = \sqrt{y}$, and both f and f^{-1} are smooth (note that $0 \notin \mathrm{dom}(f^{-1})$).

(iii). Let $g\colon \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = x^3$. This is invertible, with $g^{-1}(y) = y^{1/3}$, but g^{-1} is not differentiable at $y = 0$ so g is not a diffeomorphism. [This map $g(x) = x^3$ is an example of a *homeomorphism* which is not a diffeomorphism: it is a continuous and invertible map and its inverse is also continuous; however, its inverse is not differentiable.]

(iv). Let $U \subset \mathbb{R}^+ \times (-\pi,\pi)$ be any open set, where $\mathbb{R}^+$ is the set of strictly positive real numbers, and define

$$\phi(r,\theta) = (r\cos\theta,\ r\sin\theta).$$

Then ϕ is a diffeomorphism of U with $\phi(U)$ – the familiar change of coordinates between polar and Cartesian. Note that if one includes $r = 0$ in U then the map is defined but fails to be a diffeomorphism; indeed it fails to have an inverse as $\phi(0,\theta) = (0,0)$ for all θ.

(v). Any invertible linear map on $\mathbb{R}^n$ is a diffeomorphism, because both the map and its inverse are smooth. If we choose bases so the linear map is represented by a matrix, then the inverse map is represented by the inverse matrix. ✐

Returning to coordinate systems, suppose we start out with a given coordinate system on an open set $U \subset \mathbb{R}^n$ (possibly the usual Cartesian one), then a new coordinate system $\{x'_1, \ldots, x'_n\}$ is related to the original one by $x'_i = \phi_i(x_1, \ldots, x_n)$, for for some functions ϕ_i with $i = 1, \ldots, n$. These components define a map $\phi\colon U \to \mathbb{R}^n$ by

$$\phi(x_1, \ldots, x_n) = (\phi_1(x_1, \ldots, x_n), \ \ldots, \ \phi_n(x_1, \ldots, x_n)) \, .$$

What properties should this map ϕ have to be a change of coordinates? Firstly, it must be invertible, otherwise we cannot change back to the original coordinates. Moreover ϕ and ϕ^{-1} should both be differentiable (or better, smooth), otherwise a function which is differentiable (or smooth) in one coordinate system might not be in the other system (for an example, see Problem B.3). In short, ϕ must be a diffeomorphism. In Example B.3(iv) the diffeomorphism ϕ is the familiar change of coordinates from polar to Cartesian.

Thus a change of coordinates *is* a diffeomorphism, and vice versa: the new coordinates are the components of the diffeomorphism. That is, if we call the new coordinates $(y_1, \ldots, y_n)$ then

$$y_i = \phi_i(x_1, \ldots, x_n).$$

We will use the expression ***coordinates about a point*** q to mean coordinates defined on a neighbourhood of q, whose components are all equal to 0 at the point q.

One should think about coordinate systems and diffeomorphisms as follows, and this becomes particularly useful when we define submanifolds below.

Let $U \subset \mathbb{R}^n$ be an open set and $q \in U$. Now consider another copy of $\mathbb{R}^n$ with its usual Cartesian coordinates $(x_1, \ldots, x_n)$. A system of coordinates on U is then a map $\phi\colon U \to \mathbb{R}^n$ which is a diffeomorphism with its image $\phi(U)$, the coordinates being the components of ϕ; see Figure B.1. Thus a single coordinate, ϕ_1 say, is just a function on U, $\phi_1\colon U \to \mathbb{R}$, and is often written $x_1 \circ \phi$ since x_1 is a coordinate function on $\mathbb{R}^n$. Furthermore, ϕ defines a coordinate system *about* q if $\phi(q) = 0$.

In the next section we see how to determine whether a given smooth map is a diffeomorphism, at least in some neighbourhood of a given point. Before doing that, we give one final definition.

Definition B.4. Two subsets S_1, S_2 of $\mathbb{R}^n$ are ***diffeomorphic*** if there are neighbourhoods U_1 of S_1 and U_2 of S_2 and a diffeomorphism $\phi\colon U_1 \to U_2$ which maps S_1 to S_2. ✫

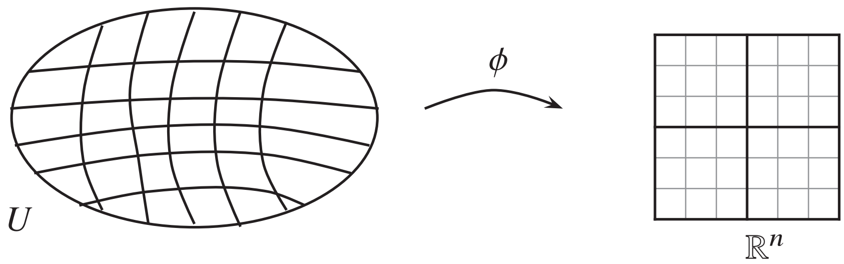


FIGURE B.1 A coordinate system on U

Maps and changes of coordinates Let $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ be a (smooth) map. The expression for the map will depend on the coordinates used on its domain. For example the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given in Cartesian coordinates by $f(x, y) = x^2 - y^2$, becomes in polar coordinates (see Example B.3(iv)) $f \circ \phi(r, \theta) = r^2 \sin 2\theta$. Here ϕ maps *from* $\mathbb{R}^+ \times (-\pi/\pi)$ *to* the domain of f.

We would often express this in reverse: if $U \subset \mathrm{dom}(f)$, then new coordinates on U would be defined by a diffeomorphism $\phi\colon U \to V$. Then the new expression for f would then be $f(x) = f \circ \phi^{-1}(y)$.

B.2 Inverse function theorem

Many of the methods of singularity theory involve calculations using infinitesimal data (differentials and vector fields) and making deductions about maps. The inverse function theorem is the archetype of such results, and not only is it a result of central importance, but the proof we give (in Chapter 5) provides a model for the proofs of several important theorems in this text and so is worthwhile understanding in detail.

Before stating the theorem, we show its converse: *if U and V are open sets in $\mathbb{R}^n$ and $f\colon U \to V$ is a diffeomorphism, then the differential $\mathsf{d}f_x$ is invertible for all $x \in U$.*

This is not surprising: if there is any justice in mathematics (and there is!), then the best linear approximation to an invertible map ought to be invertible. Indeed, since f is invertible, let $g = f^{-1}\colon V \to U$ be its inverse, so that $g \circ f\colon U \to U$ is the identity map on U. Applying the chain rule gives

$$\mathsf{d}g_y\, \mathsf{d}f_x = \mathsf{Id}_n,$$

where $y = f(x)$ and Id_n is the $n \times n$ identity matrix, so $[\mathsf{d}f_x]^{-1} = \mathsf{d}g_y$.

The inverse function is the converse of this observation. It says that if the best linear approximation (the differential) to a smooth map at a particular point is invertible, then so is the map itself – at least in some neighbourhood of that point. More formally:

Theorem B.5 (Inverse function theorem)**.** *Let* $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^n$ *be a smooth map with* $x_0 \in \mathrm{dom}(f)$*, and suppose* f *has rank* n *at* x_0*. Then there is a neighbourhood* U *of* x_0 *such that the restriction* $f_{|U}\colon U \to f(U)$ *is a diffeomorphism.*

The statement that f has rank n at x_0 is equivalent to saying that the Jacboian matrix $\mathsf{d}f_{x_0}$ is invertible. In most texts, this theorem is proved using the contraction mapping principle. We prove it using the 'homotopy method' in Chapter 5 (p. 72).

Remark B.6. Since diffeomorphisms are equivalent to smooth changes of coordinates, the statement of the theorem can also conclude that there is a change of coordinates in, say, the target, such that the map takes the form

$$f(x_1, \dots, x_n) = (x_1, \dots, x_n).$$

Namely, the change of coordinates is f^{-1}, for of course $f^{-1} \circ f = \mathsf{Id}$. Likewise, if one fixed the coordinates in the target, there is a change of coordinates in the source such that f takes the same form, since in this case $f \circ f^{-1} = \mathsf{Id}$.

Such choices of coordinates are a particular case of ***linearly adapted coordinates***; we will see more of this idea below. ❞

Example B.7. Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the smooth map

$$(u_1,\, u_2) = f(x_1,\, x_2) = (x_1 + x_2^2,\, x_2 + x_1^2).$$

Then at the origin $\mathsf{d}f_{(0,0)} = \mathsf{Id}$ which is invertible. Consequently there is a neighbouhood U of the origin in the source such that $f\big|_U\colon U \to f(U)$ is a diffeomorphism.

Moreover, if we let $y_1 = x_1 + x_2^2$ and $y_2 = x_2 + x_1^2$ then y_1, y_2 defines a coordinate system in a neighbourhood of $(0,0)$ in the source, and with these coordinates $f(y_1, y_2) = (y_1, y_2)$.

Likewise, if we let v_1, v_2 be such that $u_1 = v_1 + v_2^2$ and $u_2 = v_2 + v_1^2$ then v_1, v_2 defines a coordinate system in a neighbourhood of $(0,0)$ in the target, and with these coordinates $f(x_1, x_2) = (x_1, x_2)$.

Note that to find v_1, v_2 in terms of u_1, u_2, involves solving the equations which in general is not possible. ✐

B.3 Immersions & submersions

The inverse function theorem describes the local structure of a map $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ of maximal rank when $n = p$. There are similar descriptions for maps of maximal rank when $n \neq p$, and these are derived from the inverse function theorem. We turn to these now: first we consider $n < p$ (immersions) and then below $n > p$ (submersions).

B.3a Immersions

Definition B.8. A smooth map $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ is an ***immersion at*** x if f has rank n at that point. The map is an ***immersion*** if it is an immersion at every point in $\mathrm{dom}(f)$. ✩

Of course, this is only possible if $n \leq p$, and if $n = p$ then the property is precisely that of being a local diffeomorphism. See Figure B.2 for an example: notice that f is not 1–1 as there are two points mapping to the same point of intersection.

Example B.9. The graph of any smooth map $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ defines an immersion

$$g\colon\ \mathrm{dom}(f) \hookrightarrow \mathbb{R}^n \times \mathbb{R}^p,$$

given by $g(x) = (x, f(x))$. The proof is an exercise (see Problem B.13). ✐

In fact the graph is more than an immersion, it is an example of an ***embedding*** : an immersion $f\colon N \to P$ is an embedding if (i) it is 1–1, and (ii) for every open set U in N there is an open set $V \subset P$ for which $U = f^{-1}(V)$. Another way to say this second condition is (ii)′ if (x_j) is a sequence in N such that $f(x_j) \to f(q)$ for some $q \in N$ then $x_n \to q$, but we will not pursue this further here. The map depicted in Figure B.2 is an immersion but not an embedding, as it violates condition (i) (see Problem B.21 for an example of a 1-1 immersion which violates condition (ii)). As was mentioned earlier, being an immersion is a local property. On the other hand, being an embedding is definitely not local.

Theorem B.10 (Local immersion theorem)**.** *Let $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ be a smooth map with $n < p$. Suppose that f is an immersion at $x \in \mathrm{dom}(f)$. Then there is a neighbourhood U of x and local coordinates in the target about $f(x)$ such that*

$$f(x_1, \ldots, x_n) = (x_1, \ldots, x_n, 0, \ldots, 0).$$

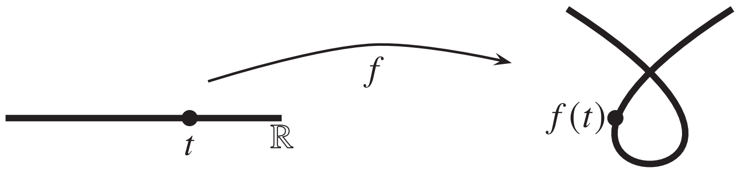


FIGURE B.2 An immersion: $f(t) = (t^3 - t,\, t^2)$.

PROOF: The proof is as follows, but the details are left to the reader. First permute the basis vectors of $\mathbb{R}^p$ so that

$$\mathsf{d}f_x = \begin{bmatrix} A \\ B \end{bmatrix},$$

where A is an invertible $n \times n$ matrix (this is done by permuting the rows of $\mathsf{d}f_x$ so that the first n rows are linearly independent). Next define $F\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^n$ by

$$F(x_1, \ldots, x_n, y_1, \ldots, y_{p-n}) = f(x_1, \ldots, x_n) + (0, \ldots, 0, y_1, \ldots, y_{p-n}).$$

It is easy to check that F satisfies the hypotheses of the inverse function theorem. Finally, one changes coordinates on $\mathbb{R}^p$ using F^{-1}. ✔

B.3B Submersions

The other major theorem on regular behaviour of maps is the submersion theorem, and the essentially equivalent implicit function theorem. We deduce this theorem from the inverse function theorem, although it can also be proved directly using the homotopy method.

Definition B.11. A smooth map $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ is a ***submersion at*** x if f has rank p at that point. It is a ***submersion*** if it is a submersion at every point in its domain dom(f). ☆

Example B.12. The projection $f\colon \mathbb{R}^p \times \mathbb{R}^k \to \mathbb{R}^p$ given by $f(x, y) = x$ is a submersion. ✐

In fact this example is the archetype of a submersion, at least locally, as the next theorem shows.

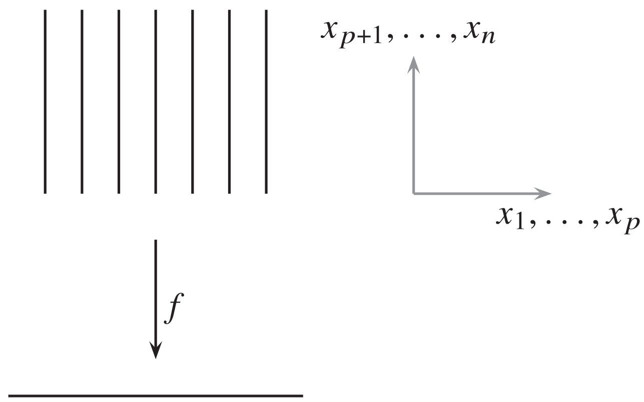


Figure B.3 A submersion

Theorem B.13 (Local submersion theorem)**.** *Let $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ be a smooth map with $n > p$, and suppose f is a submersion at $x_0 \in \mathrm{dom}(f)$, and $y_0 = f(x_0)$. Given any system of coordinates about y_0 in $\mathbb{R}^p$, there is a neighbourhood U of x_0 and coordinates about x_0 for which f takes the form*

$$f(x_1,\dots,x_p,x_{p+1},\dots,x_n) = (x_1,\dots,x_p).$$

That is, a submersion is locally a projection; see Figure B.3. Notice moreover that the image of any submersion is open. This theorem, like the next, follows from the extended implicit function theorem – the proof is given below.

In terms of applications, the implicit function theorem is the most important theorem of this appendix. A great deal of information about the theorem, its history and its applications can be found in [66].

Theorem B.14 (Implicit function theorem)**.** *Let $f\colon \mathbb{R}^p \times \mathbb{R}^k \rightarrowtail \mathbb{R}^p$ be a smooth map with $f(x_0, y_0) = z_0$, and suppose $\mathsf{d}f_{(x_0,y_0)}$ has the form*

$$\mathsf{d}f_{(x_0,y_0)} = [A \;\; B]\,, \tag{B.1}$$

where A is an invertible $p\times p$ matrix, and B is any $p \times k$ matrix. Then there are neighbourhoods U of x_0 and V of y_0 in $\mathbb{R}^p$ and $\mathbb{R}^k$ respectively, and a smooth map $h\colon V \to U$, with $h(y_0) = x_0$, such that, for $(x, y) \in U \times V$,

$$f(x,y) = z_0 \iff x = h(y).$$

In other words, the equation $f(x, y) = z_0$ can (in principle) be solved for x as a function of y. This function h is the eponymous ‘implicit function’. It is unusual to be able to find an explicit form for this implicit function, although its Taylor series to any order can be computed using implicit differentiation.

Example B.15. As a very simple example, consider the equation $x^3 - xy - y^4 = 5$. One sees that the point $(x, y) = (2, 1)$ satisfies the equation. Can the set of solutions near to this point be written as x being a function of y? Answer: yes! Because $\frac{\partial f}{\partial x}(2, 1) \neq 0$. And because $\frac{\partial f}{\partial y}(2, 1) \neq 0$, it can also be solved (locally) for y as a function of x. See Problem B.9. ✐

Both the submersion theorem and the implicit function theorem follow from the inverse function theorem, via the following result, which has a pleasing symmetry about its conclusion.

Theorem B.16 (Extended implicit function theorem)**.** *Assume f satisfies the hypothesis of the implicit function theorem above. Then there are neighbourhoods U, V, W of x_0, y_0 and z_0 in $\mathbb{R}^p$, $\mathbb{R}^k$ and $\mathbb{R}^p$ respectively and a smooth map $g\colon V \times W \to U$, satisfying for $(x, y, z) \in U \times V \times W$,*

$$f(x, y) = z \iff g(z, y) = x.$$

Proof: Define a map $F\colon \mathbb{R}^p \times \mathbb{R}^k \rightarrowtail \mathbb{R}^p \times \mathbb{R}^k$ by

$$F(x, y) = (f(x, y), y).$$

The Jacobian matrix of F at (x_0, y_0) is

$$\mathsf{d}F_{(x_0,y_0)} = \begin{bmatrix} A & B \\ 0 & I_k \end{bmatrix}.$$

This is invertible, with inverse $\begin{bmatrix} A^{-1} & -A^{-1}B \\ 0 & I_k \end{bmatrix}$ (as is readily checked), so by the inverse function theorem there is a neighbourhood U_1 of (x_0, y_0) (which we can take to be of the form $U \times V$) and a map $G\colon W \times V \to U$, where $W = f(U \times V)$ which is a neighbourhood of z_0, satisfying $F \circ G = \mathsf{Id}_V$. Moreover G has the form

$$G(z, y) = (g(z, y), y)$$

for some smooth map $g\colon V \to \mathbb{R}^p$ (exercise: prove this). Since $G = F^{-1}$ it follows that

$$F(x, y) = (z, y) \iff (x, y) = G(z, y),$$

and this is equivalent to $f(x, y) = z \iff x = g(z, y)$ as required. ✔

This extended implicit function theorem can be viewed as a parametrized version of the inverse function theorem: here y is the parameter, and for each value of y the maps f_y and g_y are mutually inverse. From this theorem we can derive the local submersion and the (ordinary) implicit function theorems as follows.

PROOF OF THEOREM B.13: As a first step, permute the columns of $\mathsf{d}f_{x_0}$ so that the first p columns are linearly independent; this amounts to permuting the basis vectors in $\mathbb{R}^n$. Then $\mathsf{d}f_{x_0}$ has the form (B.1), so we can apply the extended implicit function theorem, and write $x_0 = (u_0, y_0) \in \mathbb{R}^p \times \mathbb{R}^k$ where $k = n - p$.

The map $F(u, y) = (f(u, y), y)$ is therefore a diffeomorphism in a neighbourhood of x_0 (cf. the proof of Theorem B.16). Consider the change of coordinates F^{-1}. Then $f \circ F^{-1}$ has the required form. ✔

PROOF OF THEOREM B.14: Here one just defines $h(y) = g(z_0, y)$. ✔

Definition B.17. The k–dimensional ***suspension*** of a map $f : \mathbb{R}^n \to \mathbb{R}^p$ is the map

$$\begin{array}{ccc} \mathbb{R}^n \times \mathbb{R}^k & \longrightarrow & \mathbb{R}^p \times \mathbb{R}^k \\ (x, y) & \longmapsto & (f(x), y). \end{array}$$

☆

It is easy to see that if f is an immersion, submersion or diffeomorphism, then so, correspondingly, is any suspension of f.

B.4 Submanifolds and local straightening

The idea of submanifolds is a crucial concept in singularity theory, as well as in many other branches of mathematics such as topology, geometry, differential equations, classical mechanics and many more besides.

Definition B.18. A subset $M \subset \mathbb{R}^n$ is a ***submanifold*** if there is an integer d (called the ***dimension*** of M) such that for each point $q \in M$ there is a neighbourhood U of q in $\mathbb{R}^n$ and a *diffeomorphism* $\Phi \colon U \to V$, where $V \subset \mathbb{R}^n$ is an open set, such that

$$\Phi(M \cap U) = (\mathbb{R}^d \times \{0\}) \cap V \subset \mathbb{R}^d \times \mathbb{R}^{n-d}.$$

The diffeomorphism Φ is called a ***local straightening map*** for M at q (or in a neighbourhood of q). One also says that M is of ***codimension*** $(n - d)$ in $\mathbb{R}^n$. ☆

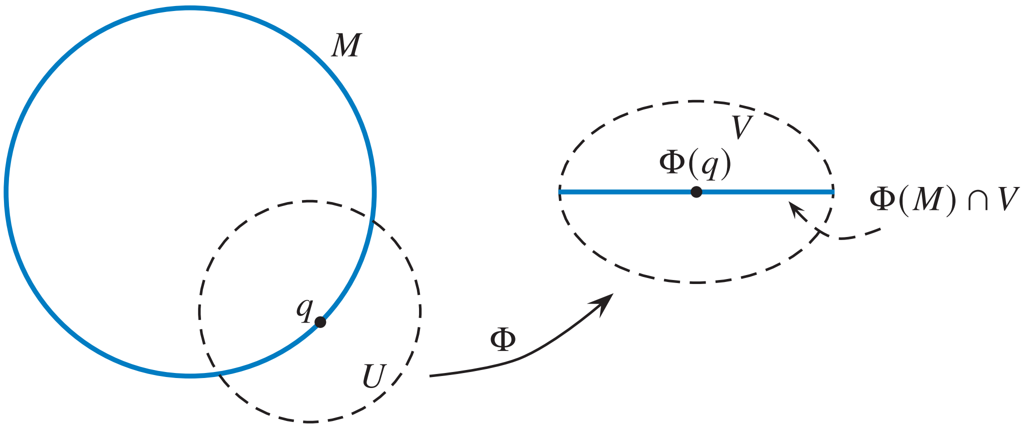


FIGURE B.4 A local straightening.

It is worth emphasizing that the straightening map Φ must be a diffeomorphism, and U and V are both open sets in $\mathbb{R}^n$. Such a submanifold is often referred to as an *embedded* submanifold, as distinct from an immersed submanifold, which we'll encounter later in this appendix.

Examples B.19 (of submanifolds)**.**

(i). Let $M = \{(x, y) \in \mathbb{R}^2 \mid y = x^2\}$ (the parabola in the plane). In this case we can define a single *global* straightening map $\Phi\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ by

$$\Phi(x, y) = (x, y - x^2).$$

Then $\Phi(M) = \mathbb{R} \times \{0\}$, showing that the parabola is a submanifold of dimension 1 (and codimension 1).

(ii). Consider the unit circle in the plane, centre the origin, usually denoted $S^1 \subset \mathbb{R}^2$. This needs more than one straightening map: for $q = (q_1, q_2) \in S^1$, define

- if $q_2 > 0$ put $\Phi(x, y) = (x,\ y - \sqrt{1 - x^2})$, which is a diffeomorphism for $U = \{(x, y) \in S^1 \mid y > 0\}$;
- if $q_2 < 0$ put $\Phi(x, y) = (x,\ y + \sqrt{1 - x^2})$, which is a diffeomorphism for $U = \{(x, y) \in S^1 \mid y < 0\}$;
- if $q_1 > 0$ put $\Phi(x, y) = (x - \sqrt{1 - y^2},\ y)$, which is a diffeomorphism for $U = \{(x, y) \in S^1 \mid x > 0\}$;
- if $q_1 < 0$ put $\Phi(x, y) = (x + \sqrt{1 - y^2},\ y)$, which is a diffeomorphism for $U = \{(x, y) \in S^1 \mid x < 0\}$.

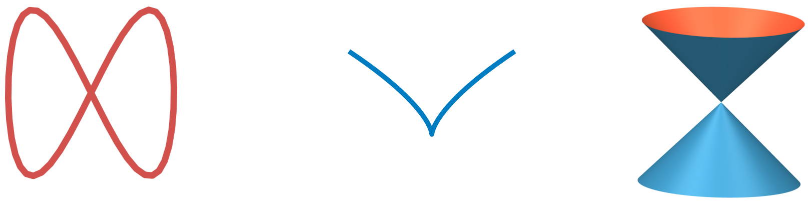


FIGURE B.5 None of these sets are submanifolds: the figure–8, the semicubical parabola (cusp) and the cone; each has one singular point.

These four straightening maps together cover the entire circle, which is enough to show it is a submanifold (of dimension 1). ✐

The following result is often used to show a subset is a submanifold. Recall that the graph of a map $f\colon X \longrightarrow Y$ is the set

$$\Gamma_f = \{(x, y) \in X \times Y \mid y = f(x)\}.$$

Proposition B.20. *Let $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^p$ be a smooth map. The graph Γ_f of f is a submanifold of $\mathbb{R}^n \times \mathbb{R}^p$ of dimension n.*

PROOF: Γ_f has a *global* straightening map, $\Phi\colon \mathbb{R}^n \times \mathbb{R}^p \longrightarrow \mathbb{R}^n \times \mathbb{R}^p$ defined simply by $\Phi(x, y) = (x, y - f(x))$. This is a diffeomorphism, as its inverse is

$$\Phi^{-1}(x, z) = (x, z + f(x)),$$

as can be readily checked, and both Φ and Φ^{-1} are smooth. Moreover,

$$\Phi(x,\ f(x)) = (x,\ f(x) - f(x)) = (x, 0),$$

so $\Phi(\Gamma_f) = \mathbb{R}^n \times \{0\}$, as required. ✔

Clearly in this setting, the codimension of the graph is $\operatorname{codim}(\Gamma_f) = p$.

Local parametrizations One of the main features of manifolds is they have (local) parametrizations, or coordinates. Let $M \subset \mathbb{R}^n$ be a submanifold of dimension d and let $p \in M$, and let $\Phi\colon U \longrightarrow \mathbb{R}^n$ be a local straightening map defined in a neighbourhood of p. Write $\Phi(x) = (\Psi(x), \chi(x))$, so $\Psi(x) \in \mathbb{R}^d$, and $\chi(x) = 0$ if and only if $x \in M \cap U$.

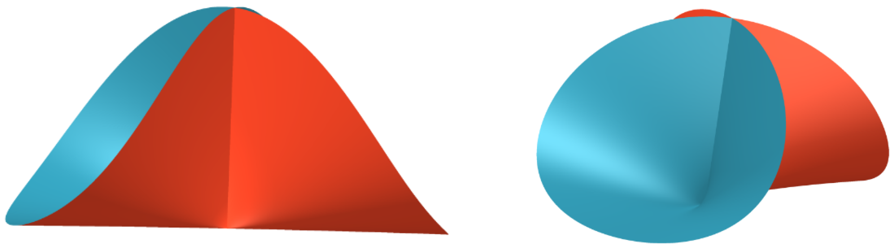


FIGURE B.6 Two views of the Cayley cross–cap – not a submanifold: it has a half–line of singular points. It is the image of the map $(x, y) \mapsto (x, xy, y^2)$ which fails to be an immersion only at $(0, 0)$.

If we restrict to $x \in M$, then the restriction $\Psi\colon M \cap U \longrightarrow \mathbb{R}^d$ defines a smooth map whose inverse is $\Psi^{-1}(y) = \Phi^{-1}(y, 0)$. Thus $\Psi^{-1}\colon U' \longrightarrow M$ is a smooth map of rank d whose image is a neighbourhood of x in M; here $U' = \Psi(M \cap U) \subset \mathbb{R}^d$. If we use coordinates $(x_1, \ldots, x_d)$ on U', then $\Psi^{-1}(x_1, \ldots, x_d)$ gives a smooth parametrization of a neighbourhood of x in M. Although (at this point) we have very little explicit information on what M can be like, this local parametrization allows us to define what is meant for a function on M to be *smooth*, namely $f\colon M \longrightarrow \mathbb{R}$ is ***smooth*** if and only if $f \circ \Psi^{-1}\colon U' \longrightarrow \mathbb{R}$ is smooth (see Problem B.16).

Example B.21. Continuing part (ii) of the previous example, with $M = S^1$ and $q = (0, 1)$, we have $\Psi(x, y) = x$, and since on M we have $x = \sqrt{1 - y^2}$, it follows that $\Psi^{-1}(x) = (x, \sqrt{1 - x^2})$, which is a parametrization of the upper half of the circle S^1 (i.e., of that part with $y > 0$). Moreover, the function $f(x, y) = \sin(y)$ is smooth on the portion of the circle parametrized by Ψ^{-1} because $f \circ \Psi^{-1}(x) = \sin(\sqrt{1 - x^2})$ which is smooth provided $|x| < 1$. ✐

Remark B.22. Readers familiar with the abstract definition of manifold will recognize the local parametrizations as the basis of that definition; indeed the map $\Psi|_M$ defines a local coordinate chart. The compatibility between different coordinate charts follows automatically here, since the composition of two diffeomorphisms is a diffeomorphism. ❞

Immersions and submersions If $f\colon \mathbb{R}^k \rightarrowtail \mathbb{R}^n$ is an immersion, the local immersion theorem tells us that locally (in $\mathbb{R}^k$) the image of f is a submanifold of $\mathbb{R}^n$ of dimension k. If on the other hand $f\colon \mathbb{R}^n \twoheadrightarrow \mathbb{R}^p$ is a submersion, then the local submersion theorem implies that all the level sets of f (the subsets of the form $f^{-1}(y)$) are submanifolds of dimension $n - p$; indeed, the reader can check that the change of coordinates in the definition provides the local straightening maps. It is useful to formalize this as the *Regular Value Theorem* – a regular value of a map f is a point y in the image for which f is a submersion at every $x \in f^{-1}(y)$.

Theorem B.23 (Regular value theorem)**.** *Let $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$, with $n > p$, be a smooth map, and $y \in \mathbb{R}^p$ a regular value of f. Then $f^{-1}(y)$ is a submanifold of $\mathbb{R}^n$ of codimension p, and hence dimension $n - p$.*

The proof follows from the local submersion theorem, and details are left to the reader.

Tangent spaces Let $X \subset \mathbb{R}^n$ be a submanifold with $x \in X$. Let $u\colon \mathbb{R} \rightarrowtail X$ be a smooth parametrized curve in X with $u(0) = x$. Then the vector $\dot{u}(0) = \frac{\mathsf{d}}{\mathsf{d}t}u(0)$ is called a ***tangent vector*** to X at x. The set of all such vectors is the ***tangent space*** of X at x written,

$$T_xX = \{\dot{u}(0) \mid u\colon \mathbb{R} \rightarrowtail X \text{ with } u(0) = x\}.$$

Proposition B.24. *Let X be a submanifold of $\mathbb{R}^n$ of dimension k and let $x \in X$.*

(i). If $g\colon \mathbb{R}^k \rightarrowtail \mathbb{R}^n$ is an immersion with image X then the tangent space

$$T_xX = \operatorname{im} \mathsf{d}g_q,$$

where $x = g(q)$.

(ii). If $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ is a submersion with $X = f^{-1}(0)$ (where $p = n - k$) then

$$T_xX = \ker \mathsf{d}f_x.$$

PROOF: (i) Let $v\colon \mathbb{R} \rightarrowtail \mathbb{R}^k$ be a path with $v(0) = q$. Then $u = g \circ v$ is a path in X, and its velocity vector is $\dot{u} = \mathsf{d}g_q\dot{v}$ so that $\dot{u} \in \operatorname{im} \mathsf{d}g_q$. Furthermore, any curve in X through x can be obtained in this way, so that indeed $T_xX = \operatorname{im} \mathsf{d}g_q$.
(ii) Let u be a path in X. Then $f \circ u \equiv 0$. Differentiating with respect to t gives $\mathsf{d}f_x\dot{u}(0) = 0$. Thus $T_xX \subset \ker \mathsf{d}f_x$. However, these two spaces have the same dimension, namely $k = n - p$, so they must be equal. ✔

B.5 Example: the set of matrices of given rank

Consider the vector space $\mathsf{Mat}(p, n)$ of $p \times n$ matrices (representing linear maps $\mathbb{R}^n \to \mathbb{R}^p$). Let $\Delta_r = \Delta_r(p, n) \subset \mathsf{Mat}(p, n)$ be the subset consisting of those matrices of rank equal to r.

Theorem B.25. *The subset $\Delta_r(p,n) \subset \mathsf{Mat}(p,n)$ is a submanifold of dimension $r(n+p-r)$.*

PROOF: Let $A \in \Delta_r$, and choose a basis so that A takes the form given in Proposition A.2. Now any matrix $A' \in \mathsf{Mat}(p,n)$ can be written in block form as

$$A' = \begin{bmatrix} I_r + B & C \\ D & E \end{bmatrix},$$

where I_r is the $r \times r$ identity matrix, $B \in \mathsf{Mat}(r,r)$, $C \in \mathsf{Mat}(r, n-r)$, $D \in \mathsf{Mat}(p-r, r)$ and $E \in \mathsf{Mat}(p-r, n-r)$. Choose a neighbourhood U of A by assuming that $B \in U$ is sufficiently small that $I_r + B$ is invertible. We now apply row operations on A', to show

$$\mathsf{A}' \sim \begin{bmatrix} I_r & (I_r+B)^{-1}C \\ D & E \end{bmatrix} \sim \begin{bmatrix} I_r & (I_r+B)^{-1}C \\ 0 & E - D(I_r+B)^{-1}C \end{bmatrix}.$$

This last matrix is of rank r if and only if $E = D(I_r+B)^{-1}C$. That is, in the neighbourhood U, the set Δ_r can be expressed as a graph of E as a function of B, C, D, and so (by Proposition B.20) Δ_r is indeed a submanifold of dimension $\dim \mathsf{Mat}(r,r) + \dim \mathsf{Mat}(r, n-r) + \dim \mathsf{Mat}(p-r, r)$, which is equal to $r(n+p-r)$. ✔

Notice in the proof that the rank r condition is written as a condition on E, which can be interpreted as saying that Δ_r has codimension $(n-r)(p-r)$ (equal to the number of entries in E). See Problems B.19 and B.20 for information about the tangent space to Δ_r.

B.6 Maps of constant rank

Let $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ be a smooth map. Since the entries of the matrix $\mathsf{d}f_x$ depend continuously on the point $x \in \mathrm{dom}(f)$, it follows that for each $r \in \mathbb{N}$ the set

$$S_r(f) := \{x \in \mathrm{dom}(f) \mid \mathrm{rk}_x(f) \leq r\}$$

is a *closed* subset of $\mathrm{dom}(f)$. Equivalently, the rank function $x \mapsto \mathrm{rk}_x(f)$ is *upper semicontinuous*, meaning that if x_j is a sequence of points in $\mathrm{dom}(f)$ converging to x_0, then

$$\lim_{j\to\infty} \mathrm{rk}_{x_j}(f) \geq \mathrm{rk}_{x_0}(f),$$

if the limit exists.

In all the theorems above, the rank of the map is maximal, and it follows therefore that it is constant in a neighbourhood of the point in question. There is a more general theorem which extends all the previous ones.

Theorem B.26 (Constant rank theorem)**.** *Let $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ be a smooth map of constant rank k, and let $x_0 \in \mathrm{dom}(f)$. Then there is a neighbourhood of x_0 and coordinates $x_1, \dots, x_n$ about x_0 and $y_1, \dots, y_p$ around $f(x_0)$ such that in these coordinates f is given by*

$$f(x_1, \dots, x_n) = (x_1, \dots, x_k, 0, \dots, 0).$$

Notice that one consequence is that image(f) is a k–dimensional immersed submanifold of $\mathbb{R}^p$ and each level set $f^{-1}(y)$ is a submanifold of $\mathbb{R}^n$ of dimension $n - k$.

This theorem contains the previous three as special cases (although its proof uses the implicit function theorem, which in turn relies on the inverse function theorem). We will prove this using 'linearly adapted coordinates' later in this chapter (see p. 354).

It is worth emphasizing one distinction between the four local theorems. In the inverse function theorem one can specify coordinates in the source or in the target, and deduce coordinates in the other so that the map has the special form (namely, the identity). In the local immersion theorem, one needs to choose (or change) the coordinates in the target, in the regular value theorem one needs to choose coordinates in the source, while in the constant rank theorem, one needs to choose coordinates in both in order to write f in the form given by the theorem.

B.7 Transversality

Two submanifolds X and Y of $\mathbb{R}^n$ are ***transverse*** at $x \in \mathbb{R}^n$, written $X \pitchfork_x Y$, if either $x \notin X \cap Y$ or their tangent spaces satisfy

$$T_x X + T_x Y = \mathbb{R}^n.$$

A map $f\colon \mathbb{R}^k \rightarrowtail \mathbb{R}^n$ is ***transverse*** to a submanifold $X \subset \mathbb{R}^n$ at $x \in X$, written $f \pitchfork_x X$ if

$$\mathrm{im}\, \mathsf{d}f_q + T_x X = \mathbb{R}^n,$$

for all $q \in f^{-1}(x)$. Two maps $f\colon \mathbb{R}^k \rightarrowtail \mathbb{R}^n$ and $g\colon \mathbb{R}^\ell \rightarrowtail \mathbb{R}^n$ are ***transverse*** at $x \in \mathbb{R}^n$ (written $f \pitchfork_x g$) if

$$\mathrm{im}\, \mathsf{d}f_q + \mathrm{im}\, \mathsf{d}g_p = \mathbb{R}^n,$$

for all q and p satisfying $f(q) = g(p) = x$. Finally, two such objects are simply ***transverse*** if they are transverse at x for all $x \in \mathbb{R}^n$.

The principal use of transversality is the following.

Theorem B.27. *Suppose X, Y are two submanifolds of $\mathbb{R}^n$.*

(i). If X and Y are transverse then their intersection $X \cap Y$ is also a submanifold. Moreover

$$\operatorname{codim}(X \cap Y) = \operatorname{codim}(X) + \operatorname{codim}(Y).$$

(ii). Let $f\colon \mathbb{R}^k \to \mathbb{R}^n$ be transverse to X. Then $f^{-1}(X)$ is a submanifold of $\mathbb{R}^k$, with

$$\operatorname{codim}(f^{-1}(X)) = \operatorname{codim}(X),$$

where the first codimension is in $\mathbb{R}^k$ and the second in $\mathbb{R}^n$.

Proof: We will prove (ii) and leave (i) to the reader. Suppose X is of dimension m. Let $q \in f^{-1}(X)$ and let U be a neighbourhood of $f(q)$ on which there is a straightening map for X: that is a diffeomorphism $\Phi\colon U \to \mathbb{R}^n$ with $\Phi(U \cap X)$ an open subset of $\mathbb{R}^m \times \{0\}$. We now proceed with the 'straightened version' of X; that is, we write

$$\mathbb{R}^n = \mathbb{R}^m \times \mathbb{R}^{n-m}$$

with $X \cap U \subset \mathbb{R}^m \times \{0\}$. Accordingly, write f as $f(x) = (f_1(x), f_2(x))$ with $f_1\colon \mathbb{R}^k \rightarrowtail \mathbb{R}^m$ and $f_2\colon \mathbb{R}^k \rightarrowtail \mathbb{R}^{n-m}$.

Now $T_{f(q)}X = \mathbb{R}^m \times \{0\}$, and hence f is transverse to X if and only if $\mathsf{d}f_2$ has rank $n - m$ at p; that is, f_2 is a submersion there. Moreover, for $x \in f^{-1}(U)$, $f(x) \in X$ if and only if $f_2(x) = 0$. Equivalently, $f^{-1}(U \cap X) = f_2^{-1}(0)$. The local result (in a neighbourhood of q) then follows from the local submersion theorem (Theorem B.13), including the statement about the codimension of $f^{-1}(U \cap X)$. Since this holds for all $q \in f^{-1}(X)$, the result follows. ✔

B.8 Linearly adapted coordinates

Recall that the rank of a map f at x is the rank of the Jacobian matrix $\mathsf{d}f_x$. Both words 'singular' and 'regular' have different meanings in mathematics depending on context. We use the following definition.

Definition B.28. A ***singular point*** or ***singularity*** of a smooth map $f: \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ is a point $x_0 \in \mathrm{dom}(f)$ where the rank $\mathrm{rk}_x(f) < p$. If $\mathrm{rk}_x(f) = p$ then x is a ***regular point*** of f. ✯

Note that if $p > n$ then every point in the source is singular. However, if $p > n$ and $\mathrm{rk}_(f) = n$ then one talks of f being a ***regular parametrization*** of its image (which by the local immersion theorem is a submanifold of $\mathbb{R}^p$, at least in a neighbourhood of x).

Note that if $p = 1$ (scalar-valued functions), a singular point is a point where all partial derivatives vanish, so a singular point is the same as a critical point.

A first step in the study of singularities is to introduce coordinates adapted to the situation. For non-singular points, the appropriate coordinates are the ones of the inverse, implicit, local immersion or constant rank theorems described above. Extending these to the case where the map is singular is the subject of the following theorem.

Theorem B.29 (Linearly adapted coordinates)**.** *Let $f: \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ be a smooth map with $f(0) = 0$, and $\mathrm{rk}_0(f) = k$. Then there is a neighbourhood U of 0 in $\mathbb{R}^n$ and coordinates $u_1, \dots, u_k, x_1, \dots, x_{n-k}$ on U and $(v_1, \dots, v_k, y_1, \dots, y_{p-k})$ in a neighbourhood of 0 in $\mathbb{R}^p$, and a smooth map $g: U \to \mathbb{R}^{p-k}$ with $g(u, 0) = 0$ and $\mathsf{d}g_0 = 0$, such that f takes the form,*

$$f(u, x) = (u, g(u, x)).$$

That is, in a neighbourhood of 0, f is given by

$$\begin{cases} v_i = u_i & (i = 1, \dots, k), \\ y_j = g_j(u, x) & (j = 1, \dots, p - k). \end{cases}$$

Proof: Since $\mathrm{rk}(\mathsf{d}f_0) = k$ we can choose bases in $\mathbb{R}^n$ and $\mathbb{R}^p$ so that

$$\mathsf{d}f_0 = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}. \tag{B.2}$$

Writing the target accordingly as $\mathbb{R}^p = \mathbb{R}^k \times \mathbb{R}^{p-k}$ we write

$$f(u, x) = (f_1(u, x), \ f_2(u, x)),$$

where $f_1: \mathbb{R}^n \to \mathbb{R}^k$. Now f_1 has rank k so it follows from the local submersion theorem that there are coordinates for which $f_1(u, x) = u$. In these coordinates, we have

$$f(u, x) = (u, f_2(u, x)).$$

Now make a further change of coordinates in the target by

$$(v, w) \longmapsto (v, w - f_2(v, 0)).$$

Then in these coordinates, f becomes

$$f(u, x) = (u, f_2(u, x) - f_2(u, 0)),$$

so putting $g(u, x) = f_2(u, x) - f_2(u, 0)$ we get the desired property that $g(u, 0) = 0$. ✔

Definition B.30. Let $f: \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ be a smooth map of rank k at the origin. Using linearly adapted coordinates we write $f(u, x) = (u, g(u, x))$ as in the theorem above. The map $g_0: \mathbb{R}^{n-k} \rightarrowtail \mathbb{R}^{p-k}$ defined by $g_0(x) = g(0, x)$ (which has rank 0 at the origin) is called the ***core*** of f (at the origin). ✯

The core of a map is important in the study of bifurcations: this is because the set of solutions of $f = 0$ is essentially the same as the set of solutions of $g_0 = 0$. As an expression in its coordinates, the core is not uniquely defined because linearly adapted coordinates are not unique. However, any two cores of a given map are 'contact equivalent' (in the sense of Chapter 11).

Example B.31. Consider the map $f: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ given by $f(x, y, z) = (x + y, x^2 + y^2 - z^2)$. The differential of f at the origin is

$$\mathsf{d}f_0 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This has rank 1, so we change basis so that $\mathsf{d}f$ is in the required form (B.2). So let $\mathbf{e}_1 = \frac{1}{2}(1, 1, 0)^T$ and $\mathbf{e}_2 = (1, -1, 0)^T$ and $\mathbf{e}_3 = (0, 0, 1)^T$ in the source, and leave the basis in the target as given. Note that $\{\mathbf{e}_2, \mathbf{e}_3\}$ spans $\ker \mathsf{d}f_0$. Then we have coordinates (u, v, w), with $(x, y, z) = u\mathbf{e}_1 + v\mathbf{e}_2 + w\mathbf{e}_3$ so that $x = \frac{1}{2}u + v$ and $y = \frac{1}{2}u - v$ and $z = w$. In these new coordinates, f takes the form

$$f(u, v, w) = \left(u, \ \tfrac{1}{2}u^2 + 2v^2 - w^2\right).$$

This is the expression of f in linearly adapted coordinates, with $g(u, v, w) = \frac{1}{2}u^2 + v^2 - w^2$, and the core is $g_0(v, w) = g(0, v, w) = v^2 - 2w^2$. Notice that $f(u, v, w) = 0$ if and only if $u = 0$ and $g_0(v, w) = 0$. ✐

The calculation in this example and the proof above, is formalized by a procedure called *Lyapunov–Schmidt reduction*; see the section below. But first we give a proof delayed from earlier.

Proof of the Constant Rank Theorem B.26: Let $q \in \mathrm{dom}(f)$, then $\mathrm{rk}\, \mathsf{d}f_q = k$. We can use linearly adapted coordinates in a neighbourhood U of q with $f(u,x) = (u, g(u,x))$. Then at any point $(u,x) \in U \subset \mathrm{dom}(f)$,

$$\mathsf{d}f_{(u,x)} = \begin{pmatrix} I_k & 0 \\ \mathsf{d}_u g & \mathsf{d}_x g \end{pmatrix},$$

where $\mathsf{d}_x g$ is the Jacobian matrix of the map $x \mapsto g(u,x)$ (consisting of partial derivatives with respect to the x variables), and $\mathsf{d}_u g$ analogously. This matrix has rank k if and only if $\mathsf{d}_x g(u,x) = 0$, and if this block of partial derivatives is identically zero, then $g(u,x)$ must be independent of x. That is,

$$f(u,x) = (u,\ g(u)).$$

The image of f is thus the graph of g, and we can change coordinates again on $\mathbb{R}^p$ by, $\psi(u,y) = (u, y - g(u))$. Then

$$\psi \circ f(u,x) = (u,0)$$

as required. ✔

The following 'reduction' procedure for cores is often useful.

Proposition B.32. *Let $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^a \times \mathbb{R}^b$ with $f(x) = (f_1(x), f_2(x))$ accordingly. Suppose $f(0) = (0,0)$ and that $f_1\colon \mathbb{R}^n \to \mathbb{R}^a$ is a submersion at* 0*. Then f and the restriction $f_2|_V$ have the same core at the origin, where $V = f_1^{-1}(0)$.*

Note that saying cores are 'the same', one means that we can choose coordinates so that they agree. One cannot expect more than that, given that – as mentioned above – the core is not unique.

Proof: Since f_1 is a submersion there is a neighbourhood of the origin in $\mathbb{R}^n$ on which there are coordinates for which $f_1(x,y) = x$, with $x \in \mathbb{R}^a$, $y \in \mathbb{R}^{n-a}$. In these coordinates, $V = \{(x,y) \mid x = 0\} = \mathbb{R}^{n-a}$. These are linearly adapted coordinates for f, since $f(x,y) = (x,\ f_2(x,y))$. The core of f is the same as the core of $f_2(0,y)$, and clearly $y \mapsto f_2(0,y)$ is the same as $f_2|_V$. ✔

B.9 Lyapunov–Schmidt reduction

Lyapunov–Schmidt reduction is a practical procedure for finding the core of a map, or linearly adapted coordinates, and is a straightforward application of the implicit function theorem. It is included here as it is often used as a first step in applications of bifurcation theory. This procedure is also valid in an infinite–dimensional context for bifurcation problems on Banach spaces (possibly arising from PDEs), although the required functional analysis would take us too far afield (see for example [49, 50] for details). See also [47] for further properties of this procedure.

Let $f:\mathbb{R}^n\times\mathbb{R}^k \rightarrowtail \mathbb{R}^p$ be a smooth map. We are interested in studying solutions of the 'bifurcation equation' $f(x;\lambda)=0$ (here we think of $\lambda\in\mathbb{R}^k$ as parameters, and $x\in\mathbb{R}^n$ as state variables) in a neighbourhood of a point (x_0,λ_0) which we take to be the origin.

There are two steps: (i) solve the largest non-degenerate part of f for the appropriate variables (using the implicit function theorem), and (ii) substitute those values into the remainder of f.

We write $f_0(x)=f(x,0)$. Then $\mathsf{d}(f_0)_0$ is the Jacobian matrix of f with respect to the x-variables (a $p\times n$ matrix) at the origin. Let $K=\ker \mathsf{d}f_0\subset\mathbb{R}^n$ be its kernel, $R\subset\mathbb{R}^p$ its image (range) and $Q=\mathbb{R}^p/R$ its cokernel. In practice one identifies Q with a subspace of $\mathbb{R}^p$ by writing $\mathbb{R}^p=R\oplus Q$. Let $\Pi\colon\mathbb{R}^p\to Q$ be the Cartesian projection with kernel R. Then $(I-\Pi)$ is the projection to R.

Write

$$f_1=(I-\Pi)\circ f,\quad \text{and}\quad f_2=\Pi\circ f. \tag{B.3}$$

Then $f(x,\lambda)=0$ if and only if $f_1(x,\lambda)=0$ and $f_2(x,\lambda)=0$.

Step (i): consider $f_1:\mathbb{R}^n\times\mathbb{R}^k\to R$. This is a submersion (in a neighbourhood of 0) and with $\mathbb{R}^n=Y\oplus K$ (for any choice of complementary subspace Y) one has $\mathsf{d}_x(f_1)_0=[A\ \ 0]$ with A invertible. Then by the implicit function theorem there is a neighbourhood of the origin and a map $h\colon K\times\mathbb{R}^k\rightarrowtail Y$ such that $f_1(x,\lambda)=0$ if and only if $x\in\Gamma_h$ (the graph of h). In other words, writing $(v,\lambda)\in K\times\mathbb{R}^k$, $f_1(x,\lambda)=0$ if and only if there is a $v\in K$ for which $(x,\lambda)=(h(v,\lambda),v,\lambda)\in Y\times K\times\mathbb{R}^k$ (all restricted to appropriate neighbourhoods of the origin).

Step (ii): define $g\colon K\times\mathbb{R}^k\rightarrowtail Q$ by

$$g(v,\lambda)=f_2(h(v,\lambda),v,\lambda).$$

Clearly, $f(x,\lambda)=0$ if and only if there exists $v\in K$ for which $(x,\lambda)=(h(v,\lambda),v,\lambda)$ and $g(v,\lambda)=0$. That is, the set of zeros of f forms a graph over the set of zeros of g. The map g is the result of Lyapunov–Schmidt reduction.

If $\mathsf{d}(f_0)_0$ has rank r, then $K \simeq \mathbb{R}^{n-r}$ and $Q \simeq \mathbb{R}^{p-r}$, so the reduction reduces the dimension by r in both source and target. Note also that $\mathsf{d}(g_0)_0 = 0$, so no further reduction is possible, and moreover that g_0 is the core of f_0.

See Problem B.18 for an example.

It should be pointed out that the resulting reduced map g depends on the choices of Q and Y (the splittings of $\mathbb{R}^n$ and $\mathbb{R}^p$). However, different choices lead to 'equivalent' maps (in particular, contact equivalent, or with the parameter, $\mathcal{K}_{\mathrm{un}}$-equivalent, as described in Chaps 11 and 14).

Problems

B.1 At which points is $f: \mathbb{R} \to \mathbb{R},\ f(x) = \cos x$ a local diffeomorphism? (†)

B.2 Let ϕ, ψ be local diffeomorphisms, ϕ defined in a neighbourhood of 0, and ψ in a neighbourhood of $\phi(0)$. Show that $\psi \circ \phi$ is a local diffeomorphism in a neighbourhood of 0.

B.3 Let x be the usual coordinate on $\mathbb{R}$, and let $x' = \phi(x) = x^3$. Consider the smooth function $f(x) = x^2$. Express f in terms of x', and show it is not differentiable at $x' = 0$. [*Note: this ϕ is a homeomorphism (i.e. ϕ is invertible and both ϕ and ϕ^{-1} are continuous). This example demonstrates why we need to consider changes of coordinates to be diffeomorphisms rather than just homeomorphisms.*]

B.4 Let $a \in \mathbb{R}$ and consider the map $f_a: \mathbb{R} \to \mathbb{R}$ defined by $f_a(x) = x + ax^2$. Use the inverse function theorem to show that f_a is a diffeomorphism on some neighbourhood of the origin. Now by explicitly solving $y = f_a(x)$ find the largest open interval U containing 0 such that $f: U \to f_a(U)$ is a diffeomorphism (the answer will depend on a). (†)

B.5 Consider the so-called ***folded handkerchief*** map $f: \mathbb{R}^2 \to \mathbb{R}^2,\ f(x, y) = (x^2,\ y^2)$. Find the singular points of f. Find also the image of f and the image of the singular points. (†)

B.6 Show that $f: \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = (\mathsf{e}^x \cos y,\ \mathsf{e}^x \sin y)$ is a local diffeomorphism at each point $(x, y) \in \mathbb{R}^2$, but that it is not a (global) diffeomorphism.

B.7 Define $f: \mathbb{R}^3 \to \mathbb{R}$ by $f(x, y, z) = x^2 + y^2 - z^2$.

(i). Show that 0 is the only singular value of f, (ii). Show that if a and b have the same sign, then $f^{-1}(a)$ and $f^{-1}(b)$ are diffeomorphic. [*Hint*: *consider scalar multiplication by a suitable constant on* $\mathbb{R}^3$.]

(iii). If a and b are of opposite signs, show that $f^{-1}(a)$ and $f^{-1}(b)$ are not diffeomorphic. [*Hint*: *show for example that one is connected and the other not.*]

B.8 Let $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^p$ be a submersion. Show it is an open map: that is, if $U \subset \mathrm{dom}(f)$ is open then $f(U)$ is open in $\mathbb{R}^p$. (†)

B.9 Consider the relation $x^2 + 1 - y^4 - y = 0$. Show that this can be solved for y as a function of x in a neighbourhood of the point $(1, 1)$. By using implicit differentiation, find the Taylor series of this function to order 3 about $x = 1$.

B.10 Consider the map $f\colon \mathbb{R}^3 \to \mathbb{R}^2$, $f(x, y, z) = (x + y^2 - xz, y - z^2)$. Define a new map $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ whose first two components are the same as f and whose third component is z (this exercise illustrates the proof of the submersion theorem).

(i). Show that f is a submersion at the origin in $\mathbb{R}^3$.

(ii). Show that F defines a local diffeomorphism at 0.

(iii). Show that $f \circ F^{-1}(X, Y, Z) = (X, Y)$ (as in the conclusion to the submersion theorem).

B.11 Find linearly adapted coordinates for the map $f(x, y) = (x, xy, y + x^2)$. (†)

B.12 Prove the Lagrange multiplier theorem: *Suppose* $f\colon \mathbb{R}^n \rightarrowtail \mathbb{R}$ *and* $g\colon \mathbb{R}^n \rightarrowtail \mathbb{R}^k$ *are smooth, and let* $p \in \mathrm{dom}(f) \cap \mathrm{dom}(g)$. *Suppose* $c = g(p)$ *is a regular value of* g *and let* $M = g^{-1}(c)$. *Then* p *is a critical point of the restriction* $f\big|_M$ *of* f *to* M *if and only if there are* $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ *such that*

$$\mathsf{d}f_p = \sum_{i=1}^{k} \lambda_i \, \mathsf{d}(g_i)_p.$$

[*Hint*: *start by applying the regular value theorem to* g.]

B.13 Prove the statement in Example B.9. Show moreover that the map defining the graph is an embedding.

B.14 Show that $S^n = \{x \in \mathbb{R}^{n+1} \mid \sum x_i^2 = 1\}$ is a submanifold of $\mathbb{R}^{n+1}$ of dimension n. It is called the n-sphere.

B.15 Let $X = \{(\mathbf{u},\mathbf{v}) \in \mathbb{R}^3 \times \mathbb{R}^3 \mid \|\mathbf{u}\|^2 = \|\mathbf{v}\|^2 = 1,\ \mathbf{u}\cdot\mathbf{v} = 0\}$. Show that X is a submanifold of $\mathbb{R}^6$ and determine its dimension. (†)

B.16 Let X be a subset of $\mathbb{R}^n$, and $f\colon X \to \mathbb{R}$ a function. One says f is ***smooth*** if for each $p \in X$ there is a neighbourhood U of p in $\mathbb{R}^n$ and a smooth function $\bar{f}\colon U \to \mathbb{R}$ such that $f = \bar{f}|_X$. Show that if X is a submanifold, this definition is equivalent to the one via straightening maps given on p. 347.

B.17 Consider the 3–dimensional vector space Sym(2) consisting of 2×2 symmetric matrices. Show that the subset consisting of symmetric matrices of rank r is a submanifold of Sym(2), for $r = 0, 1, 2$. What are their dimensions?

B.18 Apply Lyapunov–Schmidt reduction at the origin, to the bifurcation problem $f\colon \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^3$ given by

$$f(x, y, z, \lambda) = (z + x^2 - \lambda y,\ y + z - 2\lambda x, z - \lambda). \qquad (\dagger)$$

(This is a sufficiently simple example that everything can be done explicitly.)

B.19 Let $A \in \Delta_r$, the subset of $\mathsf{Mat}(p,n)$ consisting of those matrices of rank r (see Section B.5). Let $K \in \mathsf{Mat}(n,n)$ be any matrix of rank $n - r$ such that $AK = 0$, and let $Q \in \mathsf{Mat}(p,p)$ be any matrix of rank $p - r$ such that $QA = 0$. Show that the tangent space to Δ_r at A is given by

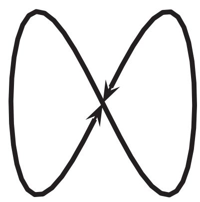
$$T_A\Delta_r = \{V \in \mathsf{Mat}(p,n) \mid QVK = 0\}.$$

[*Hint*: *choose bases so that A takes the normal form in Proposition A.2.*]

B.20 Let $A \in \Delta_r \subset \mathsf{Mat}(p,n)$. Use the previous problem to show that $PA \in T_A\Delta_r$ for all $P \in \mathsf{Mat}(p,p)$ and $AN \in T_A\Delta_r$ for all $N \in \mathsf{Mat}(n,n)$.

B.21 Show that the following 'figure-8 without double point' is an immersion but not an embedding. The map is $f\colon (0,\ 2\pi) \to \mathbb{R}^2$

$$f(t) = (\sin(t),\ \sin(2t)).$$



The diagram shows the figure-8 which is the image of f. Note that $f(\pi) = (0,0)$ so that the origin is in the image and there is no gap, and if either endpoint of the interval $[0,\ 2\pi]$ were included in the domain the map would no longer be injective. The arrows are there to emphasize the behaviour of f as $t \to 0$ and $t \to 2\pi$. [*Hint*: *to prove that f is not an embedding, consider the open set* $U = (\pi - \varepsilon,\ \pi + \varepsilon) \subset (0, 2\pi)$. *It is enough to show there is no neighbourhood V of* $(0,0)$ *in the target space* $\mathbb{R}^2$ *such that* $f^{-1}(V) = U$.]

Index

Figures are in **bold**. Individual singularity or bifurcation types are listed under 'singularity type' or 'bifurcation type' respectively