HOW MANY EQUILIBRIA ARE THERE?
AN INTRODUCTION TO MORSE THEORY

Robin Forman

Our goal in this lecture is to investigate a way of counting the equilibria of a dynamical system. Actually, we will not really count the equilibria, but rather we will relate the number of equilibria to the answer of a problem in topology that seems at first glance to have little to do with dynamical systems. The connection between these two subjects was discovered, at least in the form we will present, by Marston Morse, and many of the ideas we will discuss were first introduced in his fundamental works [Mo1] and [Mo2], but we will also refer to later insights. We warn the reader that this lecture is in the form of an informal discussion. We will make some rather vague comments along the way, with the goal of adding precision as we go along. However, some points will be remain imprecise throughout the entire lecture. We hope only to give the reader an appreciation for the wonderful subject which now goes under the name of "Morse Theory".

Section I: Dynamical Systems

The first topic on the agenda for today is *dynamical systems*. The sort of dynamical systems we have in mind are those in which a system is acting so as to minimize its energy. Such systems are found in abundance in nature. The equilibrium positions play an important role in our understanding of such a system, because they are the only states of the system that can be observed for more than a fleeting moment. Let us begin our discussion with a simple example.

When I was a student, I would often amuse myself, while listening to boring lectures, by trying to balancing the chair next to mine on its two back legs. (Many friends of mine would balance the chair they were sitting in on the back legs. It is significantly easier to balance a chair on its rear legs if you are sitting in it, because you can shift your own weight to help maintain balance. On the other hand there is much more at stake.) We all know that there is such an equilibrium position. Now I have a question for you. Why is it that when we enter a room filled with chairs, we never find that some of the chairs are balanced on their back legs. Most of the chairs are resting on all 4 legs, and perhaps we see a few chairs lying on their side, but I have never (yet!) come into a room to find a chair balancing on its back legs.

I'm sure you all know the answer to this mystery. The position of balancing on the two back legs is an *unstable* equilibrium. Even if we did manage to get a chair balancing on

Typeset by $\mathcal{AMS}$-TEX

its rear 2 legs, any movement forward or back, no matter how small, would send the chair forward to its standard position, or tipping completely over backwards.

Now let's go further. It seems pretty clear that there is an equilibrium position for the chair in which it is balanced on a single rear leg. Even my daredevil friends never tried that one. Why not? (I am not asking this question in order to encourage you to try!) The answer is that not only is it an unstable equilibrium, it is even more unstable than the equilibrium position of balancing on two legs.

What does it mean for one unstable equilibrium to be more unstable than another? How can we quantify levels of instability? Before reading further the reader should think about this point on his or her own.

.

.

.

Okay, perhaps that's long enough. I will now describe one way of measuring the instability of an equilibrium point. If we are balancing a chair on its rear two legs, there is only one component of motion, the forward-back component, that we need to worry about. If the chair is balancing on its rear two legs, unless there is a major disruption, it is unlikely to tip over on its side. We say that the forward-back component is an *unstable component* of direction. On the other hand, if we are balancing a chair on a single rear leg, we have to worry about two components of motion, the forward-back component, and the left-right component, since we can easily tip over in any direction. (Fortunately, with the combined help of gravity and the floor, we do not have to worry too much about the up-down component). In this case there are two unstable components of direction.

With all this in mind, let us define the *index* of an equilibrium position to be the number of unstable directional components. For example, the index of any stable equilibrium is 0. As we said earlier, our goal is to investigate the number of equilibria of a dynamical system. One of Morse's great insights is that the problem becomes easier if we keep track of the index of each of the equilibria.

Before moving on, let us spend a bit more time exploring the concept of the index of an equilibrium point. It is important that we come to grips with this idea, because it is central to the subject. Suppose there is a ball rolling around on the one-dimensional landscape shown in Figure 1. In this case, the ball is acting so as to minimize its height (or equivalently, its potential energy due to gravity). The equilibria are those points such that if we placed the ball there and let go (so that the ball starts with zero velocity) it will stay there. Another way to say this is that they are precisely the points such that the tangent line to the landscape is horizontal. (Throughout this discussion we will assume that the height function is differentiable, so that we may speak about tangent lines. Soon we will also assume that the second derivative of the height function exists.) There are three such points, which we have labeled A, B and C. The point A is a stable equilibrium, since if the ball begins at a point near A, it will roll towards A. Therefore, point A has index 0. The point B is clearly unstable, since if the ball begins near the point B, it will roll away from B. This equilibrium has index 1, since it is unstable in the left-right direction, and there are no other directions available to the ball. How about the point C? It is stable to the left (since if the ball starts slightly to the left of the point C it will roll towards C), but unstable to the right (since if a ball is placed slightly to the right of C it will roll away from C). What is the index of C? Should we say it is stable or unstable in the left-right component of the direction? There does not seem to be any correct answer to this question, so we will simply say that the index of the point C is undefined.

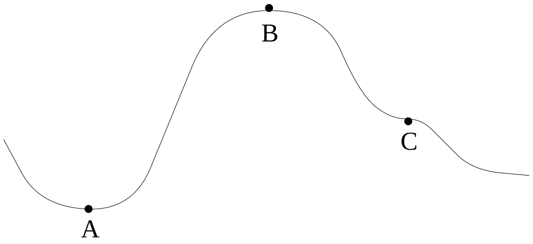

Figure 1. A one-dimensional landscape with three equilibria.

I'm sure that this last paragraph reminded you a bit of some discussions you had in your first calculus course. The point A is a local minimum of the energy (=height) function, the point B is a local maximum, and C is an inflection point. In fact, this entire discussion is probably best carried out in the language of calculus. The equilibria are those points where the tangent line to the graph of the height function is horizontal. Those are precisely the critical points of the height function. Now suppose that p is a critical point of the energy function E, so that $E'(p) = 0$. If, in addition, $E''(p) > 0$, then we know that p is a local minimum of E, so p is a stable equilibrium, and hence has index 0. If $E''(p) < 0$, then we know that p is a local maximum of E, so p is an unstable equilibrium of index 1. If $E''(p) = 0$ then the second derivative test does not tell us the index of p, or even if the index is well-defined. If $E''(p) = 0$ we say that p is a *degenerate critical point of* E. Conversely, if $E''(p) \neq 0$ we say p is a *nondegenerate critical point of* E. If all of the critical points of E are nondegenerate, so that, in particular, all of the critical points have a well-defined index that can be determined from the second derivative test, then we say that the energy function E is a *Morse function.*

In the example of a ball moving along the landscape, the set of possible positions of the ball can be identified with the points on the x-axis (i.e if you tell me the x-coordinate of the ball, I know immediately where it is). The set of possible positions of a dynamical system is called the *configuration space* of the system. In the previous example, the configuration space can be identified with a one-dimensional line. We will now consider dynamical systems with more interesting configuration spaces.

Example 1: Suppose we have a pendulum on a rigid rod. The configuration space for this system, that is the set of possible positions of the pendulum, can be identified with a circle (see Figure 2). The equilibria are the points labeled A and B. The point A is a stable equilibrium and has index 0, and the point B is an unstable equilibrium and has index 1. We will record this information as follows:

Configuration space = circle

Number of equilibria of index $0 = 1$

Number of equilibria of index $1 = 1$

Number of equilibria of index$\geq 2 = 0$.

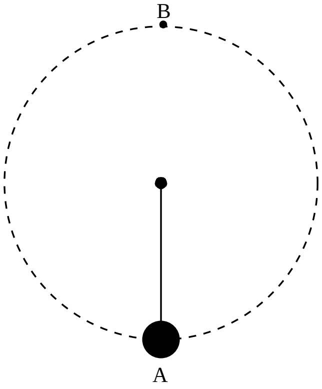

Figure 2. A pendulum on a rigid rod.

Our discussion can also be carried out in higher dimensions. Suppose we have a differentiable function $E : R^2 \to R$. We can graph this function in R^3 and think of a ball rolling around on this landscape. The equilibria are the points where the tangent plane is horizontal, and these are precisely the critical points of the function E. In our multi-variable calculus course, we are taught the second derivative test for such functions. If p is a critical point of a function E, we consider the Hessian of E at p. This is the 2×2 matrix of second derivatives

$$\begin{pmatrix} \frac{\partial^2 E(p)}{\partial^2 x} & \frac{\partial^2 E(p)}{\partial x \partial y} \\ \frac{\partial^2 E(p)}{\partial x \partial y} & \frac{\partial^2 E(p)}{\partial^2 y} \end{pmatrix}.$$

If both eigenvalues of this matrix are positive, for example if $E(x,y) = x^2 + y^2$, and p is the origin, then p is a local minimum of the energy function (Figure 3(i)). This implies that p is a stable equilibrium, and hence has index 0. If the Hessian has two negative eigenvalues, for example if $E(x,y) = -x^2 - y^2$, and again p is the origin, then p is a local maximum of the energy function (Figure 3(ii)). This implies that at p both the x and the y components of direction are unstable components, so p is an equilibrium of index 2. If the Hessian has one positive eigenvalue and one negative eigenvalue, for example if

$E(x,y) = -x^2 + y^2$, and again p is the origin, then p is a saddle point (Figure 3(iii)). In this case, the x component of direction (indicated by the dotted curve through p in Figure 3(iii)) is an unstable component, since if we place a ball at p and then move it slightly in the x direction and let go, it will roll away from p. On the other hand, p is stable in the y direction (the solid curve through the point p in Figure 3(iii)), since if we place a ball at p and then move it slightly in the y direction and let go, it will roll towards p. Therefore there is precisely one unstable component of direction at p, so p is an equilibrium of index 1.

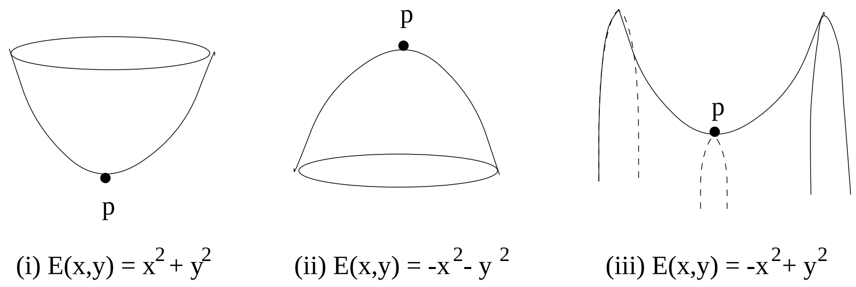
p
p
p
(i) E(x,y) = x^2 + y^2
(ii) E(x,y) = -x^2- y ^{2}
(iii) E(x,y) = -x^2+ y^2

Figure 3. Examples of two-dimensional equilibria.

If the Hessian has a zero eigenvalue, then the second derivative test does not tell us the index of p, or even whether p has a well-defined index. In this case, just as in the one-dimensional setting, we say that p is a degenerate critical point.

This discussion can be carried out in any dimension. Suppose that the configuration space is m-dimensional, and that near a critical point p the energy function E has the form

$$E(x_1, x_2, \ldots, x_m)) = -x_1^2 - x_2^2 - x_3^2 \cdots - x_i^2 + x_{i+1}^2 + \cdots + x_m^2, \tag{1}$$

so that there are exactly i unstable components of direction (i.e directions in which the energy is decreasing), then the critical point p has index i. Just as in the case of 2 dimensions, we can use the second derivative test to detect the index of a critical point.

The second derivative test: Suppose that p is a critical point of an energy function E, and the Hessian of E at p has no zero eigenvalues. Then the index of E at p is precisely the number of negative eigenvalues of the Hessian.

I have not labeled this statement a theorem because we do not yet have a precise definition for index. In fact, this statement is often used to define the index of an equilibrium point. An alternate point of view is that one can define a critical point p to have index i if near p the energy function E looks like the function shown in (1). If one takes this as the definition for index, then one then has to prove the above statement by showing that if the Hessian of E at p has i negative eigenvalues (and no zero eigenvalues) then E looks like the function (1). Here, we have been using a rather vague phrase "looks like" in reference to functions.

Unfortunately, I do not think that it would be worth the effort to make this phrase precise. For a precise statement of what we mean for a function to look like another in this context, and a proof of the second derivative test stated above, see the “The Morse Lemma” (the name given to this second derivative test), Lemma 2.2 in [Mi1].

Now let us apply these ideas to some 2-dimensional examples.

Example 2: Suppose we start with a round metal globe. The mathematical name for this shape is a *sphere*, or a *2-sphere* if we wish to emphasize that it is 2-dimensional, and is often denoted by the symbol S^2. Now suppose we place a marble on its surface, and that the marble is magnetized so that it stays on the globe as it rolls around. The equilibria are those points with the property that if the marble is placed there with zero velocity, then the marble will not move. These are the points where the tangent plane is horizontal. There are 2 equilibrium points on the sphere (Figure 4), which we have labeled A and B, and they clearly have indices 0 and 2, respectively.

Configuration space = sphere

Number of equilibria of index 0 = 1

Number of equilibria of index 1 = 0

Number of equilibria of index 2 = 1

Number of equilibria of index≥ 3 = 0.

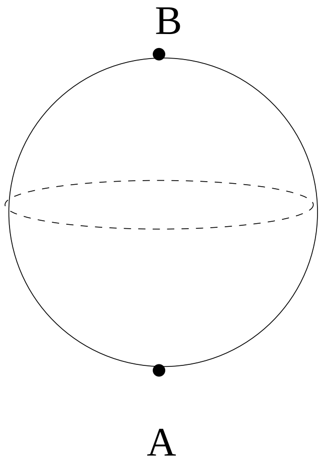

Figure 4. Equilibrium points for a magnetized marble rolling around on a metal globe.

Example 3: This time we start with a metal inner tube. The mathematical name for this shape is a *torus*. Now suppose we sit the metal torus on one end (see Figure 5) and place a magnetized marble on its surface. Here there are 4 equilibrium points, which we have labeled A, B, C and D. The point A is a stable equilibrium, and has index 0. The point D is a local maximum of the energy function and has index 2. The points B and C are a bit trickier. In fact, if one considers a small piece of the torus near the point B, it looks very much like the saddle point we drew in Figure 3(iii). The point B is stable in the left-right component of direction, and unstable in the forward-back component. Therefore B has index 1. The same holds true for the point C except that C is unstable in the left-right component of direction, and stable in the forward-back directions. Therefore, the point C also has index 1.

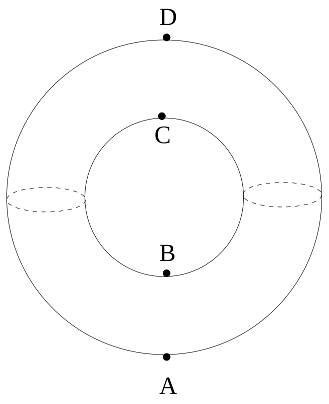

Figure 5. Equilibrium points for a magnetized marble rolling around on a metal torus.

Configuration space = torus

Number of equilibria of index 0 = 1

Number of equilibria of index 1 = 2

Number of equilibria of index 2 = 1

Number of equilibria of index≥ 3 = 0.

Section II: Topology

We will now briefly leave the subject of dynamical systems, and begin a discussion of some topics in topology. You probably already know something about topology. It is sometimes called “rubber geometry” because we will say that two shapes are topologically the same if one can be made into the other by stretching, pulling and twisting, and other similar operations. No cutting or pasting is allowed. If two shapes are the same in this way, we will say they are *topologically equivalent*, or (using the fancy word) *homeomorphic* (homeo=same, morph=shape). The sort of question we will be asking is typical in all branches of science (and, in fact, in many other areas of human endeavor). We will begin by describing a collection of simple objects, the building blocks of the theory. The main problem will then be to investigate how more complicated objects can be built from these building blocks.

The building blocks in the theory we wish to discuss are called *cells*, and there is one cell for each dimension.

The 0-dimensional cells are points (also called vertices).

The 1-dimensional cells are open intervals.

The 2-dimensional cells are discs (without their boundary). Since we are working in the world of topology, we can stretch these discs, and we can draw them as squares, or triangles, if we like. They are all 2-dimensional cells.

The 3-dimensional cells are the inside of a cube, or the inside of a sphere, or the inside of a pyramid, or ...

We illustrate these cells in Figure 6. The dotted lines in this figure are meant to indicate that the cells do not contain the points on their boundary.

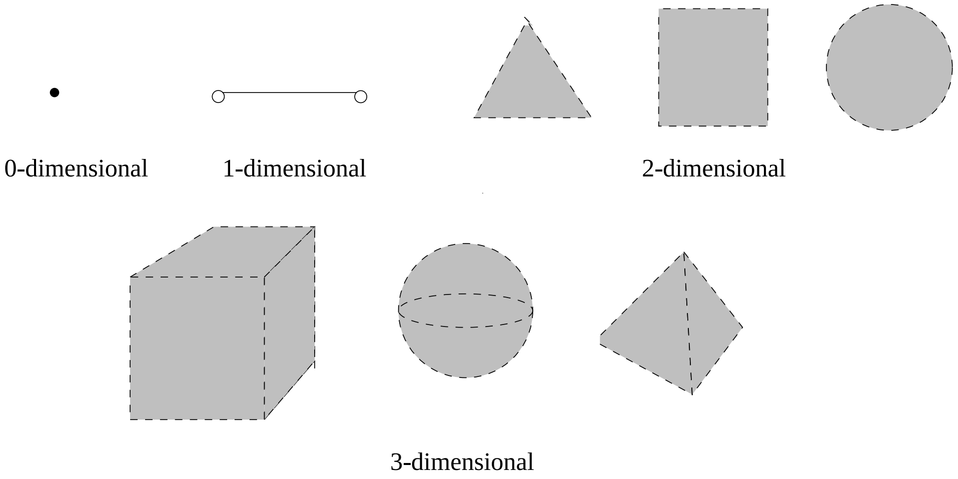

Figure 6. Topological cells

We can continue, and define cells of every dimension, but we will stop here. It is useful to introduce a bit of math shorthand. Instead of writing out “0-dimensional cell” each time, it is traditional to refer simply to a "0-cell". Similarly, we will sometimes refer to a 1-cell, or 2-cell.

Now let us return to the basic question. If we are given a shape, how can we build it out of these building blocks?

Example 1: Let's try this in the case of the circle. It is pretty easy to see that if I remove a single point (i.e a 0-dimensional cell) from the circle, then what remains is (topologically) an open interval (i.e. a 1-dimensional cell). Therefore, the circle can be built from one 0-dimensional cell and one 1-dimensional cell. Let us summarize this as follows.

Topological space = circle

Number of cells of dimension 0 = 1

Number of cells of dimension 1 = 1

Number of cells of dimension $\geq 2 = 0$.

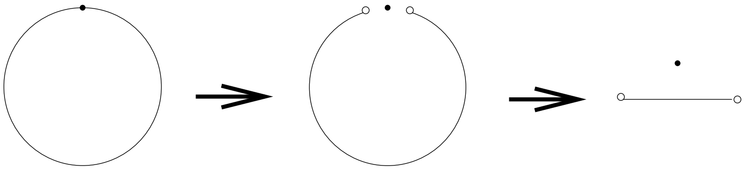

Figure 7. A decomposition of a circle into cells.

Example 2: For our next example, we consider the 2-dimensional sphere. If we remove a single point from the sphere, what remains can be stretched out flat to a round disc (i.e. a 2-dimensional cell). Therefore the 2-sphere can be built from one 0-dimensional cell and one 2-dimensional cell.

Topological space = sphere

Number of cells of dimension 0 = 1

Number of cells of dimension 1 = 0

Number of cells of dimension 2 = 1

Number of cells of dimension $\geq 3 = 0$.

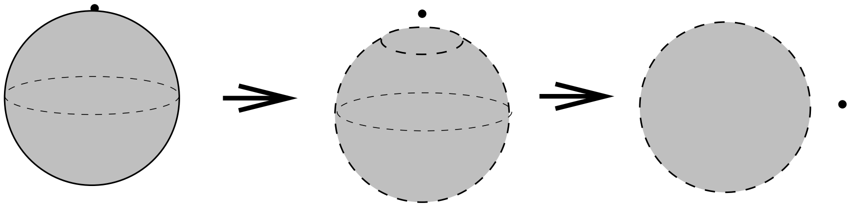

Figure 8. A decomposition of a 2-sphere into cells.

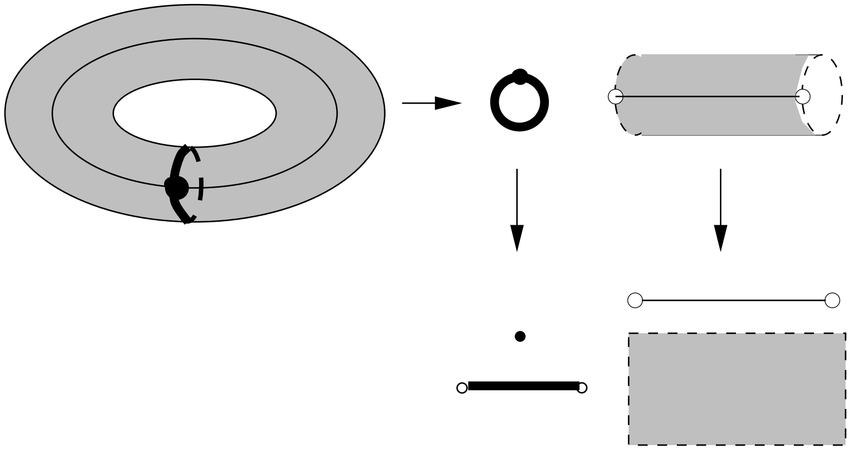

Figure 9. Decomposing the torus into cells.

Example 3: This time we start with a torus. We can build a torus as follows (see Figure 9). If we take a circle out of the torus, what remains is a cylinder which does not contain its boundary circles. We can build this circle from one 0-cell and one 1-cell. We can build the cylinder from one 1-cell and one 2-cell. Putting these constructions together, we have just shown how to build the torus from one 0-cell, two 1-cells, and one 2-cell.

Topological space = torus

Number of cells of dimension 0 = 1

Number of cells of dimension 1 = 2

Number of cells of dimension 2 = 1

Number of cells of dimension $\geq 3 = 0$.

Section III: Morse Theory

By now I hope the reader will have noticed the remarkable similarity between the examples in Section I and the examples in Section II. We first observe that the shapes considered in Section II are precisely the configuration spaces of the dynamical systems considered in Section I. Moreover, the number of cells required to build the configuration space seems to be the same as the number of equilibria of the dynamical system, with the dimension of the cell corresponding to the index of the equilibrium point. The main theorem of this lecture is that this is a general phenomenon. We are going to state the theorem using the precise mathematical terminology, all of which will be explained in the remainder of this section. What follows is the main theorem of Morse Theory.

Theorem: Let M be a closed, compact, smooth submanifold of Euclidean space (of any dimension). Let $E : M \to R$ be a smooth, real-valued function on M. Suppose that every critical point of E is nondegenerate. Then M can be built from a finite collection of cells, with exactly one cell of dimension i for each critical point of index i.

Some of the words in this theorem need to be explained. I will not give precise definitions, but only a rather vague description of what the hypotheses are requiring. First note that the theorem refers to *Euclidean space (of any dimension).* All of the examples we have considered take place in R^2 or R^3, but the phrase in the theorem means that everything can be placed in R^k for any k. A subset of a Euclidean space R^k is a *smooth submanifold* if it has the property that for each point p in the subset, the set of points in the subset which are near p looks just like the set of all points near the origin in some Euclidean space. In Figure 10 we show three subsets of R^2 which are *not* smooth submanifolds because in each case the point labeled A does not satisfy this condition. Note that in Figure 10(iii), near the point A the subset looks like the set of points near the origin in R^1 in a topological sense, since in topology we can always straighten out corners, but we actually need the space to look like Euclidean space in a stronger differentiable sense so that we can make sense of the idea of taking derivatives of functions on our space. In fact, we will need somewhat more because we will need to take the second derivative of our functions on M.

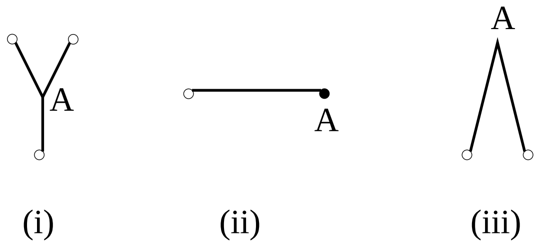

Figure 10. Subsets of R^2 which are not smooth submanifolds.

We must now explain the meaning of the word *compact.* A subset of Euclidean space is compact if and only if it satisfies two other properties. Namely, it must be *closed* and *bounded.* A subset M of R^k is closed if it has the following property. Suppose there is a point q in R^k such that we can get as close to q as we want while staying in M. Then q is required to be in M. For example, suppose we let M be the subset of R^1 consisting of all of R^1 except the number 0. Then M is not closed because you can get as close to 0 as you like while staying in M, but the point 0 itself is not in M. The same problem occurs if M is any open interval, for example, if $M = (0, 1)$. Then one can get as close as we like to the numbers 0 and 1 while staying in M, but the numbers 0 and 1 are not themselves in M. The condition that M is bounded means simply that it is possible to surround M by a (possibly very large) ball in R^k, i.e M does not go off to infinity in any directions. For example, the x-axis sitting in R^2 is closed but not compact.

Note that the three examples we examined in Sections 1 and 2 satisfied the hypotheses. It is a very good exercise for the readers to think about what can go wrong with the theorem if M is not required to be compact.

I will leave the reader with one last note concerning the statement of the theorem. We required that M be a subset of some Euclidean space. In fact, that is irrelevant for the theorem. We do need M to be a space on which it makes sense to take derivatives of functions. Such spaces are called *smooth manifolds.* I stated the theorem in this manner only because abstract manifolds, that is those which exist on their own, without reference to a surrounding Euclidean space, are somewhat harder to describe and to imagine.

Section IV: What the main theorem does not say

There is occasionally some confusion as to what the main theorem actually says. It does not say, for example, that all energy functions have the same number of critical points, or that the cell complex that you get in this manner is necessarily the "best possible" (i.e. has the fewest number of cells). Consider the example shown in Figure 11(i). In this case, our space is a topological circle. We see that there are four critical points, labeled A, B, C and D, with A and C having index 0, and B and D having index 1. Therefore, the main theorem implies that the circle can be built from two 0-cells and two 1-cells. This is shown in Figure 11(ii). The reader should compare this example with Example 1 of Sections I and II.

Section V: The idea of the proof

The proof we are going to describe is due to Smale [Sm1]. I apologize in advance that some details will be imprecise, and others will be precise, but will have unexplained terms. Still I hope that the reader will walk away with some understanding of what is going on.

Suppose p is a critical point of index i. Draw a small piece of the unstable directions near p. After adding the point p to this set, it will be a small i-cell. This is true because a critical point has index i precisely when there is an i-dimensional family of unstable directions. See Figure 12(i) where we have done this for each of the critical points of the function on the torus considered in Example 3 of Section I.

The next step is to draw a little arrow at every point on the manifold, indicating the direction in which the energy is decreasing the fastest (Figure 12(ii)). You may know that the *gradient* of the energy function points in the direction in which the energy is increasing the fastest, so the arrow we are drawing points precisely in the opposite direction of the

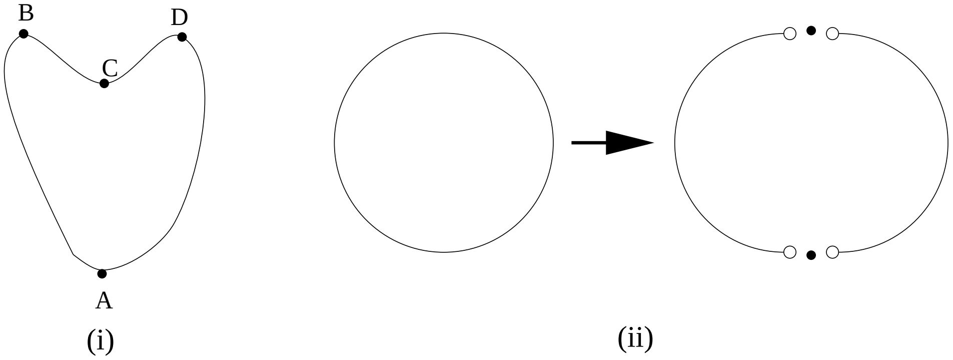

Figure 11. A configuration space which is a topological circle.

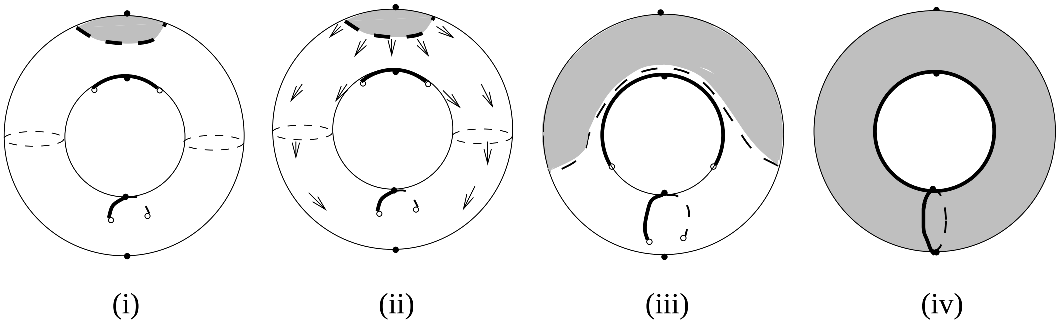

Figure 12. Growing a cell decomposition on the torus.

gradient. Now think of each point on the manifold flowing in the direction of the arrows we drew. If you are not sure what this means, think of putting a drop of molasses at a point on the manifold and watching it slowly make its way down. The drop will come very close to flowing in the direction of the arrows we have drawn.

Now watch what is happening to each of the cells we drew. As they flow they grow bigger and bigger, but they remain cells of the appropriate dimension (Figure 12 (iii)). The key point is that as we let more and more time go by, these cells fill up more and more of the manifold. If we let the amount of time go to infinity (!), the cells fill up the entire manifold (Figure 12 (iv)), and this shows that the manifold can be decomposed into the desired number of cells.

How do we know that the cells will fill up the entire manifold? Let q be a point in M. We need to see that one of the cells, which we are watching grow larger and larger, will eventually contain q. Which cell will it be? There is a fun way to tell. Let a drop of molasses begin at the point q, and this time let it flow **up** the manifold (or equivalently, turn the manifold upside down and let the drop flow down). If we let time approach infinity, the drop will get closer and closer to a critical point (here we are definitely using the compactness of M). The critical point the drop approaches is precisely the critical point whose cell will eventually contain q.

This point of view enables us to give a concise description of the resulting decomposition of M into cells. For each critical point p, Let $U(p)$ denote the set of all points in M which, when flowing **up** the manifold, flow towards p. We will include the point p in $U(p)$ even though it doesn't flow at all. The set $U(p)$ is called the *unstable cell* associated to p (or sometimes the *descending cell* associated to p). The crucial observations are:

1) $U(p)$ is a cell, and its dimension is the index of p.

2) If p and q are distinct critical points, then $U(p)$ and $U(q)$ are disjoint.

3)M is equal to the union of all the $U(p)$, where the union is taken over all critical points p.

Section VI: What does it really mean to build a shape from cells?

So far we have been a bit cavalier with our language when it comes to "building shapes out of cells." In fact, the sort of decomposition into cells that is provided by Morse Theory has some special properties which we will now discuss.

Suppose we already have a shape, and we wish to add a 1-cell to it. We know a 1-cell is an open interval, so that is what we are adding to our space. However, I will now add an important restrictions as to how that 1-cell can be added. From now on, we are not permitted to do this in any way we please. We must fill in the 2 endpoints of the interval with points that are already in our space. In other words, we begin with a closed interval, and glue the 2 endpoints to points in our space, so that the only points we are adding to the space are those in the open interval. It is important that we glue both points to our original space. For example, in Figure 12 we illustrate the case in which we begin with a 0-cell. In Figure 12(i) we show how to add a 1-cell to the 0-cell. The drawing in Figure 12(ii) is not permitted.

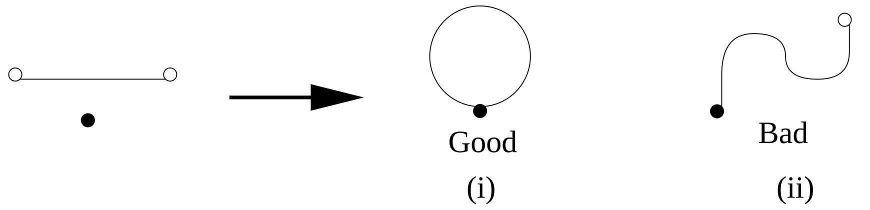

Figure 13. Adding a 1-cell to a point.

Adding a 2-dimensional cell is defined similarly. Instead of starting with a 2-dimensional disc without its boundary, we think of instead the 2-dimensional disc with its boundary.

The boundary is just a topological circle. To add a 2-cell to a space, we must glue every point on that circle to a point in our space. Moreover, we must do this in a continuous way.

Adding an i-cell, for $i > 2$ is defined similarly. A *cell complex* is a space that can be constructed by starting with a single point and adding one cell at a time in this manner.

The decomposition of M into cells which is provided by the Morse theorem has the important property of giving M the structure of a cell complex.

Note that to give a space the structure of a cell complex, one must start with a single point (which we count as a 0-cell) and then order the remaining cells, i.e. declare which one to add first, which one to add next, etc. How does this arise for M? Simply order the critical points of the energy function according to the value at the energy function at that point (ties can be broken arbitrarily). The first critical point in line will be the minimizer of the energy. This is always a stable critical point, and hence has index 0. This gives us our 0-cell to start with. We then add the cells one at a time, according to the value of the energy function at the critical point. In Figure 14 we illustrate this process for the example of the torus with the height function shown in Figure 4. In this case we add the cells corresponding to the critical points A, B, C, D in that order.

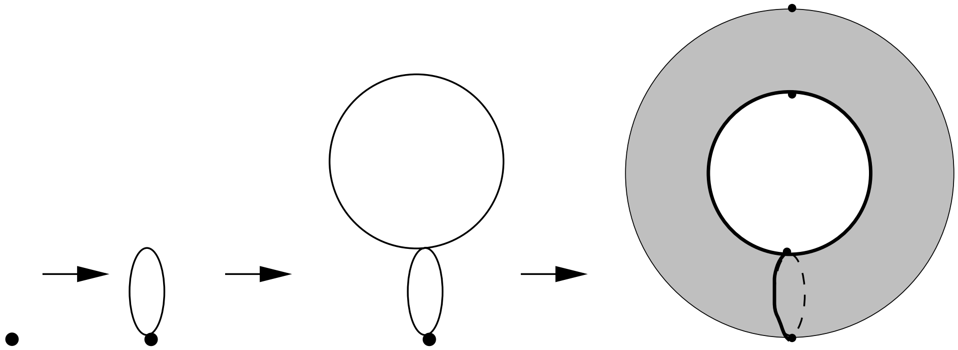

Figure 14. The torus as a cell-complex

We can now state a more precise version of the main theorem of Morse theory.

Theorem: Let M be a smooth compact submanifold of Euclidean space (of any dimension). Let $E : M \to R$ be a smooth, real-valued function on M. Suppose that every critical point of E is nondegenerate. Then M is homeomorphic (i.e. topologically equivalent) to a cell complex which has exactly one cell of dimension i for each critical point of index i.

There is one last subtlety to discuss. There is not complete uniformity in the literature as to the definition of a cell-complex. As it is often defined, a cell-complex is required to satisfy an additional condition. Namely, we required that when we add a cell to a space, the entire boundary of the cell must be glued to the space. There is often an additional requirement. Recall the space we are adding the cell to is itself the union of cells. The additional requirement is that the boundary of an i-cell can only be glued to cells of dimension less than i.

Take a look at the construction of the torus as a cell complex shown in Figure 14. This is not a cell-complex in this new, more restrictive, sense, because the boundary of the second 1-cell we add is glued to the first 1-cell, and this is not permitted. However, if we just tilt the torus slightly before beginning the growing process shown in Figure 12, then the second 1-cell would miss the first 1-cell entirely, and the result would be a cell-complex in our new sense. This always works. Given a Morse function E, if E does not give rise to a good cell-complex, then a generic small perturbation of E will do the trick. (The word "generic" means that those perturbations which will not work form a very small set.) Smale was the first to investigate this issue, and energy functions which give rise to good cell decompositions are now called Morse-Smale functions. (In fact somewhat more is required of a function before it is called Morse-Smale. Not only must the boundary of an i-cell only meet cells of lower dimension, but they must meet in a nice way. I will not say any more about this. See [Sm1] for details.)

Section VII: What now?

If all there were to Morse theory is the main theorem we have discussed, then I would still think it was a beautiful subject, but I would probably not be making such a fuss about it. In fact, Morse's work lead to a veritable revolution in the study of the topology of smooth manifolds. There is simply no time to give an overview of the hundreds of applications of Morse theory which have appeared in the literature. There are many examples of results that have be proved rather easily using Morse theory, and yet are quite difficult to prove by other means. The book [Mi1] describes some of these applications. Here, I will simply briefly describe two very striking (and historically significant) applications. I know that my brief remarks will be insufficient to give the reader a true understanding of these great works. I do hope that perhaps the reader will be sufficiently intrigued to study these topics further.

We recall that the main theorem of Morse Theory relates the number of equilibrium points of a dynamical system to the number of cells required to build the configuration space. One of the wonderful aspects of this theorem is that it has been very powerfully applied in each direction. That is, sometimes one wishes to understand the number of equilibria of a dynamical system, so one studies the topology of the configuration space and then applies the Morse theory. On the other hand, sometimes one wants to understand the topology of a space. One method is to put an energy function on the space. Information about the critical points of this function can then be translated into information about the original space.

One of the first major applications of Morse theory was Morse's investigation of the number of geodesics between two points in a manifold (see page 248 in [Mo2] and Part III of [Mi1]). Very roughly speaking, suppose we stretch a rubber string between the two points in a manifold. If, when we let go of the rubber string, it stays exactly as it is (we are assuming that it is restricted to stay on the manifold), then the path followed by the rubber string is a geodesic (see, for example, Part II of [Mi1] for a precise definition). The rubber string moves so as to minimize its energy, so this set-up is ripe for an application of Morse's theory. Namely, the configuration space of our dynamical system is the space of all possible ways for a rubber string to go from one point to another, i.e. the set of all paths from one point to another. The geodesics are precisely the equilibria of this dynamical system. Morse studied the number of geodesics by investigating the topology of the configuration space, and then applying his theory. For example, he was able to deduce that there are always infinitely many geodesics between any two points in a smooth compact manifold. (We have something more to say about this example a bit later.)

In Morse's work on geodesics, information about the topology of the configuration space was used to deduce information about the number of equilibria of the system. We now describe an example in which the flow of information goes the other way. In [Sm2] Smale used Morse theory to prove the higher dimensional Poincaré conjecture. The Poincaré conjecture states that any manifold which "looks like" a sphere, in some weak topological sense, is, in fact, topologically equivalent to a sphere (unfortunately, there is no time to explain the conjecture in any more detail than that). Smale's method was to put an energy function on the manifold and to then study the equilibria of the resulting dynamical system. He proved that if a manifold (of dimension ≥ 5) satisfies the hypothesis, then there is an energy function which has only two critical points, namely the maximum and the minimum. It then follows from a theorem of Reeb's that the manifold is a sphere (see Theorem 4.1 of [Mi1]). While Smale's proof used Morse theory quite extensively, he also used techniques from an approach to topology known as "handlebody theory". In [Mi2] Milnor presents a proof which is entirely in the language of Morse theory.

A great place to read about Morse theory, as well as some of the earlier, exciting, applications, is Milnor's wonderful book [Mi1]. The reader should be warned, however, that this book, like Morse's early writings, does not take the same point of view we have chosen, so some parts of the discussion may not look very familiar. Our discussion is more in line with the philosophy of [Sm1]. In addition to Morse, Milnor and Smale, Raoul Bott is one of the great practitioners of Morse theory, and the reader should certainly take a look at his beautiful survey article [Bo] for a look at some of the recent developments in the theory.

Morse theory has been extended and generalized far beyond what we presented in this lecture. Again, we will have to be content with just two examples. Some problems concerning the topology of spaces which are not smooth manifolds have been successfully solved by developing versions of Morse theory that can be applied to more general spaces (see, for example, [GM] and [Fo]). Perhaps the most exciting development has been the application of Morse theory to infinite-dimensional manifolds (!). It is interesting to note that one of Morse's first applications of his theory, which we described above, was to the study of an infinite dimensional manifold, the space of all paths connecting two points in a (finite-dimensional) manifold. However, he studied this space by considering finite-dimensional approximations to the space, and applying the finite-dimensional theory. Now we have the know-how to study the infinite-dimensional space of paths directly (see, for example, [Pa],[Sm3]). More recently, Floer's application of ideas from Morse theory to some infinite dimensional manifolds ([Fl]) has resulted in some extremely important advances in mathematical physics and topology.

I must warn you that this lecture is not sufficient preparation for reading most of the papers referenced in the previous paragraphs. However, I wanted to mention them so that you could see that Morse theory has been a growing and vibrant subject ever since its introduction almost 75 years ago. I think that it is one of the most beautiful mathematical developments of the century. The study of Morse theory has given me a lot of joy, and I am happy to have had this occasion to share the subject with you.

References

[Bo] R. Bott, *Morse Theory Indomitable*, Publ. Math. I.H.E.S. **68** (1988), 99–117.

[Fl] A. Floer, *An instanton-invariant for* 3*-manifolds.*, Comm. Math. Phys. **118** (1988), 215–240.

[Fo] R. Forman, *Morse Theory for Cell Complexes*, Adv. in Math . **134** (1998), 90–145.

[GM] M. Goresky and R. MacPherson, *Stratified Morse Theory*, in Singularities, Part I (Arcata, CA, 1981), Proc. Sympos. Pure Math., 40, Amer. Math. Soc., R.I., (1983), 517-533.

[Mi1] J. Milnor, *Morse Theory*, Annals of Mathematics Study No. 51, Princeton University Press, 1962.

[Mi2] ———, *Lectures on the h-Cobordism Theorem*, Princeton Mathematical Notes, Princeton University Press, 1965.

[Mo1] M. Morse, *Relations Between the Critical Points of a Real Function of n Independent Variables*, Trans. Amer. Math. Soc. **27** (1925), 345–396.

[Mo2] M. Morse, *The Calculus of Variations in the Large*, Amer. Math. Soc., Providence, R.I., **27** (1934).

[Pa] R. Palais, *Morse Theory on Hilbert Manifolds*, Topology **2** (1963), 299-340.

[Sm1] S. Smale, *On Gradient Dynamical Systems*, Annals of Math. **74** (1961), 199–206.

[Sm2] ———, *The Generalized Poincaré Conjecture in Dimensions Greater than Four*, Annals of Math. **74** (1961), 391–406.

[Sm3] ———, *Morse Theory and a Non-Linear Generalization of the Dirichlet Problem*, Annals of Math. **80** (1664), 382–396.