Lecture 10: Introduction to The Morse Functions

Topics in Computational Topology: An Algorithmic View

Scribed by: Brian Arand

3 May 2011

Today we will introduce Morse functions. Many specialized fields that have vested interest in the study of topology (computer graphics, scientific visualization, etc...) are often concerned with the study of functions defined on a topological space. Many times these functions are scalar. As we will see, Morse functions are a family of 'nice' functions that ease this study.

1 Introduction to Morse functions

1.1 The First Derivative

Before formally defining Morse functions, we first consider a simple example. Let $f\colon \mathbb{R} \to \mathbb{R}$. Such a function produces a graph like that of Figure 1.

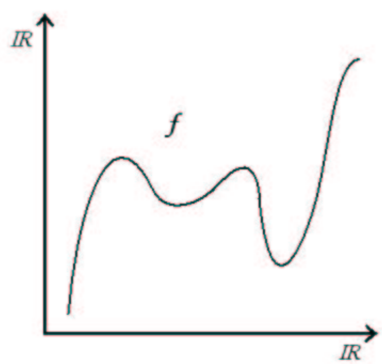

Figure 1: An arbitrary function $f : \mathbb{R} \to \mathbb{R}$ defined on the real line

Given such a function, an intuitive scalar property that one might want to measure would be its *derivative*. Recall from calculus:

Definition 1.1 (Derivative) *The derivative of a function is defined as*

$$\frac{d}{dx}f(x) = \lim_{t \to \infty} \frac{f(x+t) - f(x)}{t}$$

Essentially the derivative of a function at point x gives the rate of change of the value of $f(x)$ at x. This can be visualized as the slope of the tangent line of the function at x. We now consider the subset of x values where $\frac{d}{dx}f(x) = 0$. These x values are well known from our study of calculus to be the *critical points* of f. For a function defined on real line (see Figure 1), there are two types of critical points: minima and maxima. These points are considered to be 'critical' because they mark where the behavior of f changes. We suspend the formal definition of these points until later.

Now that we have a basis from which to build, we leave the one dimensional domain $\mathbb{R}$ in favor of a manifold M of arbitrarily dimension d. Let $f\colon \mathsf{M} \rightarrow \mathbb{R}$. With this new domain, we loose some of the simplicities of our earlier example. Derivatives, for instance, can no longer be represented as scalars, and the rate of change of the function is directionally dependent. However, at a base point $x \in \mathsf{M}$, we can still talk about the derivative of f in a certain direction $\vec{v}$. This is the so-called *directional derivative* $Df\vec{v}(x)$ along $\vec{v}$, and it measures the rate of change of f in the direction of $\vec{v}$. Now assume that there is a local coordinate frame $< x_1, x_2 ... x_d >$ around x. Then the information of all directional derivatives can be concisely represented by the *derivative of f at x*, denoted by $Df(x)$, defined as the directional derivative in each coordinate direction:

$$Df(x) = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} ... \frac{\partial f}{\partial x_d}]$$

The critical points of a function are those where every directional derivative vanishes. In other words, critical points are now the set of points $x \in \mathsf{M}$ such that:

$$Df(x) = [0, 0...0]$$

With this higher dimensional perspective of derivatives, we can now confidently define critical points as follows:

Definition 1.2 (Critical Point) *Given a smooth manifold* M *of dimension* d *and function* f *defined on* M, *the point* x *is a critical point iff*

$$\frac{\partial f(x)}{\partial x_i} = 0; \qquad 1 \leq i \leq d$$

where x_i *is the* i^{th} *component of the coordinate frame* $< x_1, x_2 ... x_d >$.

Another useful concept is that of a *critical value*.

Definition 1.3 (Critical Value) *The image* $f(x)$ *of a critical point* x.

If the input manifold M has a Riemannian structure on it (that is, there is a metric defined on M), then we can define the gradient of a function.

Definition 1.4 (Gradient) *A gradient* ∇f *is a vector field* $\nabla f\colon \mathsf{M} \rightarrow T\mathsf{M}$ *such that for any* V

$$< V(x), \nabla f(x) >= Df\vec{v}(x)$$

where M *is a manifold and* $T\mathsf{M}$ *is the tangent space of* M.

Intuitively, the direction of gradient $\nabla f(x)$ at a point x indicates the steepest descending direction of f (i.e, the direction where the function f decreases fastest), and its magnitude is the rate of change along this direction. The critical points of a function are basically where the gradient vector vanishes.

1.2 The Second Derivative

From the first derivative of a function we can determine critical points. But, by examining the second derivative we can classify a function's critical points as local minima, local maxima, saddle points, or degenerate critical points. Consider the Hessian Matrix.

Definition 1.5 (Hessian Matrix) *A Hessian Matrix of f at x is the matrix of second derivatives,*

$$Hessian(x) = \begin{vmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_2}2(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_d}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1}(x) & \frac{\partial^2 f}{\partial x_d \partial x_2}2(x) & \cdots & \frac{\partial^2 f}{\partial x_d \partial x_d}(x) \end{vmatrix}$$

The critical point x is *degenerate* if $Det(Hessian(x)) = 0$ ($Hessian(x)$ is not full rank). Otherwise, the critical point x is considered *non-degenerate*. For example,

$$Let f = x^2 + 1 \tag{1}$$
$$f'(x) = 2x \tag{2}$$
$$f'(0) = 0 \qquad (critical point) \tag{3}$$
$$f''(x) = \frac{\partial(2x)}{\partial x} \tag{4}$$
$$f''(0) = 2 \qquad (not degenerate) \tag{5}$$

$$Let f = x^3 \tag{6}$$
$$f'(x) = 3x^2 \tag{7}$$
$$f'(0) = 0 \qquad (critical point) \tag{8}$$
$$f''(x) = \frac{\partial(3x^2)}{\partial x} = 6x \tag{9}$$
$$f''(0) = 0 \qquad (degenerate) \tag{10}$$

Generally, the behavior around degenerate critical points are hard to manage. So, we'll only consider functions such that no critical points are degenerate. This brings us (finally) to the definition of Morse functions.

Definition 1.6 (Morse Function) *A function $f\colon \mathsf{M} \to \mathbb{R}$ is a Morse function iff the following conditions are met:*

1. *None of f's critical points are degenerate.*
2. *No two of f's critical points share the same function value.*

By limiting our study only to Morse functions, we're not actually loosing much because Morse functions are dense in the space of functions. That is, most functions can easily be 'tweaked' to achieve a Morse function without gravely changing the underlying behavior. At the same time, these restrictions of Morse functions result in a number of helpful characteristics like Morse Lemma.

Lemma 1.7 (Morse Lemma) *Given a Morse function $f\colon \mathsf{M} \to \mathbb{R}$, let p be a non-degenerate critical point of f, then there are local coordinates with $p = (0, 0, 0)$ such that, locally, the function f can be represented as*

$$f(x) = f(p) - x_1^2 - ...x_s^2 + x_s^2...x_d^2, \qquad for s \in [0, d]$$

for every point $x = (x_1, x_2...x_d)$ in a small neighborhood of p.

Consider a 2-manifold in $\mathbb{R}^3$, based on the Morse Lemma, there are three types of critical points that may arise:

Figure 2: The local neighborhoods of the possible critical points of a 2-manifold. Picture courtesy of [1].

1. Local Minima: $f(x) = f(p) + x_1^2 + x_2^2$, when $s = 0$.
2. Saddle Point: $f(x) = f(p) - x_1^2 + x_2^2$, when $s = 1$.
3. Local Maxima: $f(x) = f(p) - x_1^2 - x_2^2$, when $s = 2$.

Figure 1.2 (courtesy of [1]) helps us visualize this property of manifolds by shading in the areas in the local neighborhoods of each type of critical point. The darker areas correspond to regions with a smaller image value with respect to that of the critical value. The left image displays a local minima, the center is a saddle point, and the right displays a local maxima. In general, for a manifold of dimension d, there are $d + 1$ possible critical point types. Figure 3 below illustrates the critical point types ($s = 0, s = 1...s = 3$) of a 3-manifold.

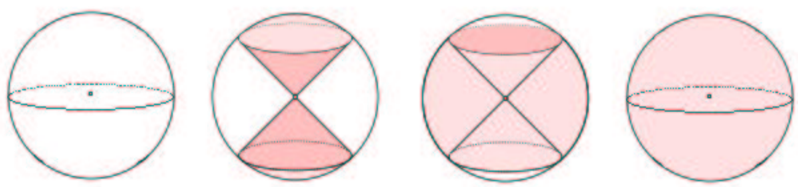

Figure 3: The local neighborhoods of the possible critical points of a 3-manifold. (Picture is courtesy of [1].)

2 Connection with Topology

First, a couple of definitions:

Definition 2.1 (Level Set) *Let $f: \mathsf{M} \to \mathbb{R}$. Then all real numbers a have a preimage, $f^-1(a)$, known as a level set.*

$$\mathsf{M}^a = f^-1(a) = \{x \in \mathsf{M} \mid f(x) = a\}$$

Informally, a level set is a set of all the points in M that result in the same function value.

Definition 2.2 (Interval-Level Set) *Let $f: \mathsf{M} \to \mathbb{R}$ and let $I \subseteq \mathbb{R}$. Then, the interval-level set is defined as:*

$$\mathsf{M}^I = f^-1(I) = \{x \in \mathsf{M} \mid f(x) \in I\}$$

An interval-level set is the union of all level sets for values from an interval I.

Definition 2.3 (Sublevel Set) *Let $f\colon \mathsf{M} \to \mathbb{R}$ Then, the sublevel set is defined as:*

$$\mathsf{M}^{(-\infty,a)} = \{x \in \mathsf{M} \mid f(x) \le a\}$$

And finally a sublevel set is a collection of all points with function values less than or equal to a certain value. It can be thought of as a sweeping through level sets.

By examining the behavior of a sublevel set's topology as we sweep through a function, we can see that the following theorem is true:

Theorem 2.4 *$H_k(M^{(\infty,a]})$ is isomorphic to $H_k(M^{(\infty,b]})$ unless the interval $[a,b]$ contains a critical value.*

An important note to take away from this is that critical points denote where the topology of a sweeping of the sublevel set of a function will change.

We end this recap with an example. Suppose we have a 2-manifold, in this case, a torus. And we define f to be the height function of each point on the torus as seen in Figure 4.

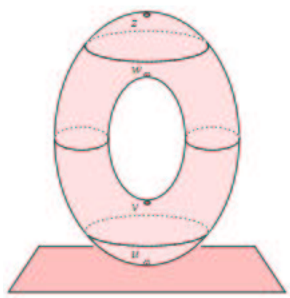

Figure 4: The height function defined on a torus with critical points u, v, w, and z (picture courtesy of [1]).

The level sets of this function would be the sets of points that share the same height or distance from the x-y plane. We can think of these sets as horizontal slices of the torus. And following this same scheme of thinking, a sublevel set of our height function would be sweeps of point from the base of the torus. A selected number of these sublevel sets are shown below in Figure 5.

Figure 5: A few important sublevel sets (picture courtesy of [1]).

From this visualization, it is easy to convince ourselves of the truth of Theorem 2.4. As we sweep through the sublevel sets of the torus and cross through the critical points u, v, w, and z we can consider the Betti numbers of the sublevel set. At u β_0 increases to 1. β_1 is incremented at both v and w. And β_2 is incremented at z. These increments illustrate a fundamental change in topology.

References

[1] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. 2009.