6 Primer in Morse Theory

In this section, we give a compact introduction to Morse theory. While it was originally developed within the calculus of variations for the infinite-dimensional case, we focus on the low-dimensional case and, in addition, sketch the recent discrete version of the theory.

6.1 Classical Morse Theory

Classical Morse theory is about smooth functions on manifolds. It is very intuitive but technically challenging in its relation to analysis and differential equations. We refer to [4J. Milnor, Morse Theory. Princeton Univ. Press, Princeton, New Jersey, 1963.] for the standard reference.

Generic smooth functions. Let $\mathbb{M}$ be a $d$-dimensional manifold (without boundary) and $f: \mathbb{M} \to \mathbb{R}$ a smooth function on the manifold. A critical point is a point $y \in \mathbb{M}$ such that the derivative of $f$ at $y$ is the zero map; see Figure 11. All other points are regular points of $f$. Similarly, a critical value of $f$ is the function value of a critical point, and all others are regular values.

A characteristic attitude in Morse theory is the restriction to generic smooth functions. In contrast to general smooth functions, they possess a regular local structure. To define what we mean by genericity, consider the Hessian of $f$ at a critical point $y \in \mathbb{M}$, which is the matrix of second derivatives:

$$H_f(y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_i \partial x_j}(y) \end{bmatrix} \quad (1 \leq i, j \leq d).$$

(27)

The critical point is non-degenerate if the Hessian is non-singular. Finally, we call $f$ a Morse function if

I. all critical points are non-degenerate;
II. the values of any two critical points are different.

Condition II is convenient but not always required. Importantly, the non-degeneracy condition implies that the critical points are isolated. Assuming a compact manifold, we therefore have at most a finite number of critical points. A significant insight is that the restriction to Morse functions is not a serious limitation. In other words, the Morse functions are dense in $C^\infty(\mathbb{M})$, the set of smooth real-valued functions on $\mathbb{M}$. Indeed, we can approximate any smooth function with a Morse function that has similar derivatives of all orders at all points.

Let now $y$ be a non-degenerate critical point of $f$. It is possible to re-parametrize the neighborhood of $y$, such that the function is locally quadratic.

**Morse Lemma.** Every non-degenerate critical point, $y$, of a smooth function, $f: \mathbb{M} \to \mathbb{R}$, has a local parametrization such that

$$f(x) = f(y) - x_1^2 - \ldots - x_p^2 + x_{p+1}^2 + \ldots + x_d^2$$

for every point $x = (x_1, x_2, \ldots, x_d)$ near $y$.

The number of negative signs, $p$, does not depend on the parametrization and is called the index of the critical point. For example, if $\mathbb{M}$ is a 2-manifold, then we have three types of critical points: minima of index 0, saddles of index 1, and maxima of index 2.

**Morse inequalities.** In Morse theory, the function is primarily a means to study the topology of the manifold on which it is defined. To this end, we look at sublevel sets, which are sets of the form $\mathbb{M}_a = f^{-1}(-\infty, a]$. As we increase $a$ from $-\infty$ to $+\infty$, it covers progressively more and eventually the entire manifold. As long as $t$ does not pass a critical value, the sublevel set remains topologically the same. At critical points, the topology changes, as we can see in Figure 11 for the height function of the torus. Specifically, we gain a component at the minimum,

![Figure 11: The height function on the torus has four critical points and, correspondingly, four critical values. For each interval of regular values, we sketch a representative sublevel set.](image)

we gain a 1-cycle each at the two saddles, and we gain a 2-cycle at the maximum. This suggests that the number of critical points of index $p$ is an upper bound for the

---

p-th Betti number, which is true. This connection is usually expressed in terms of a family of inequalities. Let $c_p$ be the number of critical points of index $p$, and write $\beta_p = \beta_p(M)$ for the $p$-th Betti number of the manifold.

**Morse Inequalities.** Let $M$ be a compact smooth $d$-manifold and $f : M \to \mathbb{R}$ a Morse function. Then

$$c_p - c_{p-1} + \ldots + c_0 \geq \beta_p - \beta_{p+1} - \ldots + \beta_0,$$

for every $0 \leq p \leq d$.

For $p = d$, the inequality is in fact an equality, stating that the Euler characteristic of the manifold is equal to the alternating sum of critical points. Starting with the stated inequalities, it is easy to prove the weak Morse inequalities: $c_p \geq \beta_p$, for every $0 \leq p \leq d$.

**Integral lines and stable manifolds.** Suppose now that we have a metric on $M$. Then we can construct the gradient of the function, $\nabla f$, which maps every point $x \in M$ to a vector in the tangent space at $x$. The result is a smooth vector field on $M$ that is zero at all critical points of $f$. An integral line is a curve $\gamma : \mathbb{R} \to M$ whose velocity vectors agree with the gradient of $f$ at all points $\gamma(t)$. Assuming the integral line is maximal, it necessarily goes from one critical point to another. More precisely, it approaches its origin and destination:

$$\text{org}(\gamma) = \lim_{t \to -\infty} \gamma(t), \quad (28)$$
$$\text{dest}(\gamma) = \lim_{t \to \infty} \gamma(t). \quad (29)$$

By the Fundamental Theorem on Differential Equations, two integral lines are either the same or they are disjoint. It follows that every regular point $x \in M$ belongs to a unique integral line, $\gamma_x$. The stable manifold of a critical point is the union of the images of all integral lines with that destination. Symmetrically, the unstable manifold is the union of images of all integral lines with that origin:

$$S(y) = \{ y \in M \mid y = \text{dest}(\gamma_x) \} \quad (30)$$
$$U(y) = \{ y \in M \mid y = \text{org}(\gamma_x) \}. \quad (31)$$

Consider a 2-manifold as an example. The stable manifold of a minimum is the minimum itself. The stable manifold of a saddle is an open curve obtained by taking the union of two integral lines and the saddle itself. The stable manifold of a maximum is the open region covered by the circle of integral lines approaching the maximum, together with the maximum. Assuming a Morse function on a $d$-manifold, the stable and unstable manifolds of an index $p$ critical point are two open topological balls, one of dimension $p$ and the other of dimension $d - p$.

With an additional genericity assumption, the stable manifolds form a complex, that is: the boundary of every stable manifold is a union of lower-dimensional stable manifolds. See Figure 12, where the boundary of the stable 2-manifold that belongs to the maximum is glued to a single point, namely the stable 0-manifold, which is the minimum itself. The additional assumption asserts that stable and unstable manifolds intersect transversally. A function $f : M \to \mathbb{R}$ that is Morse and satisfies this additional assumption is called a Morse-Smale function. Importantly, it is still true that this smaller set of functions is dense in $C^\infty(M)$. Assuming a Morse-Smale function, we can construct $M$ by adding stable manifolds, one at a time, in the order of non-decreasing dimension. The $p$-th Betti number can increase only when we add a stable $p$-manifold, which implies the weak Morse inequalities.

**6.2 Discrete Morse Theory**

Discrete Morse Theory is about matchings of simplices in a simplicial complex, called discrete gradient fields, and it has been developed relatively recently. In contrast to the smooth theory, it is more difficult to get a good intuition for this structure, but it has the advantage of being entirely combinatorial. We refer to for an introduction.

**Collapses.** A concept that predates discrete Morse theory is the deconstruction of a simplicial complex by elementary operations that remove simplices in pairs. Call a simplex $\sigma$ in a simplicial complex $K$ free if it is the face of a single other simplex, $\tau$, which necessarily satisfies

---

The operation that removes both simplices is called an elementary collapse, and we write $K \searrow K - \{\sigma, \tau\}$. For example, if $K$ consists of a single $d$-simplex and all its faces, we can perform $2^d - 1$ elementary collapses and be left with a single vertex; see Figure 13. The inverse operation, which adds simplices $\sigma$ and $\tau$ to the smaller complex, is referred to as an elementary anti-collapse.

![Figure 13: From left to right: a tetrahedron, the three triangles left after collapsing the tetrahedron, the three edges left after collapsing each of the three triangles, and the vertex left after collapsing each of the three edges.](image)

For each collapse, there is a (continuous) deformation retraction with the same effect on the underlying space, which proves that the collapse does not change the homotopy type of the complex. We call $K$ collapsible if there is a sequence of collapses that reduces $K$ to a single vertex. Correspondingly, the underlying space of $K$ is contractible or, equivalently, it has the homotopy type of a point.

**Discrete gradient fields.** In discrete Morse theory, we encode an elementary collapse as a pair of simplices. Importantly, we do not require that the lower-dimensional simplex in the pair is free. Letting $\sigma$ and $\tau$ be simplices in $K$, we write $\sigma \prec \tau$ if $\sigma$ is a face of $\tau$ and $\dim \sigma = \dim \tau - 1$. A discrete vector field on $K$ is a set of pairs $\alpha \prec \beta$ such that each simplex belongs to at most one pair. We refer to this set of pairs as $V$, suggesting that we think of it as a vector field. A path is a sequence of simplices of the form

$$\alpha_0 \prec \beta_0 \succ \alpha_1 \prec \beta_1 \succ \ldots \succ \alpha_n \prec \beta_n \succ \alpha_{n+1},$$

where $\alpha_i \prec \beta_i$ is in $V$, for $0 \leq i \leq n$, and $\alpha_i \neq \alpha_{i+1}$ for $0 \leq i \leq n$. The path is closed if $\alpha_0 = \alpha_{n+1}$.

**DEF.** A discrete gradient field is a discrete vector field without closed paths.

The intuition for this definition comes of course from classical Morse theory, in which the integral lines connect the critical points in an acyclic fashion.

**Discrete Morse inequalities.** A critical simplex of $K$ and $V$ is one that does not belong to any pair. As in smooth Morse theory, we can collect all simplices that belong to paths ending at a critical simplex. However, different from the integral lines, we do not get a partition. Nevertheless, if $K$ is a simplicial complex with a discrete gradient field, then it has the same homotopy type as a CW complex with one cell of dimension $p$ for each critical simplex of index $p$. For the analysis of the Vietoris-Rips complex in Section 5, a weaker statement about the Betti numbers suffices. Letting $s_p$ be the number of critical $p$-simplices, and writing $\beta_p$ for the $p$-th Betti number of $K$, we have a discrete analog of the Morse inequalities.

**Discrete Morse inequalities.** Let $K$ be a simplicial complex and $V$ a discrete vector field on $K$. Then

$$s_p - s_{p-1} + \ldots + \pm s_0 \geq \beta_p - \beta_{p-1} + \ldots \pm \beta_0,$$

for every $0 \leq p \leq d$.

Importantly, we have $s_p \geq \beta_p$, for every $p$, which is what we used in the analysis of the Vietoris-Rips complex in Section 5.