1 Introduction

1.1 Morse functions

Let M be a smooth manifold. Given $f \in C^\infty(M, \mathbb{R})$, its differential df is a section of T^*M. The zeros of this section are called *critical points* of f, and form the subset $\mathrm{Crit}(f) \subset M$; the values of f on $\mathrm{Crit}(f)$ are called *critical values*. All other points and values are called *regular*. Always remember that if c is a critical value, not all points in $f^{-1}(c)$ need be critical.

We now define the Hessian of f, which is a symmetric bilinear form on the tangent space of a critical point:

Definition 1. The Hessian $D^2_p f \in \mathrm{Sym}^2 T^*_p M$ of a function f at a critical point $p \in M$ is defined via

$$D^2_p f(X, Y) = X(Y(f))|_p = Y(X(f))|_p, \tag{1}$$

for any vector fields X, Y. Note that $[X, Y](f)|p = df([X, Y])|_p = 0$ since p is a critical point, and also that $D^2_p f$ depends only on the values of X, Y at p, by the second and third terms, respectively, in (1).

Exercise 1. Show that if the first k terms (starting with the zeroth order term) of the coordinate Taylor expansion of a section of a vector bundle E vanish, then the k^{th} order term will be a well-defined (i.e. coordinate-independent) tensor in $\mathrm{Sym}^k T^*_p M \otimes E_p$. This is most easily seen through the properties of the so-called jet bundles $J^k E$.

Exercise 2. The tangent space H to the image of the section $df \in \Gamma(M, T^*M)$, at the point $(p, 0) \in T^*M$ (for $p \in \mathrm{Crit}(f)$), is a subspace of the tangent space $T_{(p,0)}T^*M$. Show that there is a natural isomorphism $T_{(p,0)}T^*M = T_pM \oplus T^*_pM$, and that the subspace H defines a map $T_pM \longrightarrow T^*_pM$ which coincides with the map $X \mapsto (Y \mapsto D^2_p f(X, Y))$.

Definition 2. $p \in \mathrm{Crit}(f)$ is *nondegenerate* when $D^2_p f$ is nondegenerate, i.e. the symmetric map $D^2_p f : T_pM \longrightarrow T^*_pM$ is an isomorphism. By Exercise 2, this is equivalent to the requirement that $df \in \Gamma(M, T^*M)$ intersect the zero section *transversally* at $(p, 0)$.

Remark 1. *Recall that transversality of subspaces U, V of the vector space W is the condition that $U + V = W$, without requiring that the sum be a direct sum.*

Definition 3. A Morse function $f \in C^\infty(M, \mathbb{R})$ is a function all of whose critical points are nondegenerate.

In view of the definition of nondegeneracy as a simple transversality, we see that this definition may be rephrased as saying f **is Morse iff** df **intersects the zero section transversally**. Since the transversal intersection of embedded submanifolds of codimension k, l yields an embedded submanifold of codimension $k + l$, we immediately obtain the following result:

Proposition 1.1. *The critical locus $\mathrm{Crit}(f)$ of a Morse function is an embedded 0-dimensional submanifold of M, so that the critical points are isolated. If M is compact, there must be finitely many critical points.*

Proof. If f is a Morse function on the n-manifold M, then the images of df and of the zero section are two embedded codimension n submanifolds of T^*M which intersect transversally. Hence they intersect in a zero-dimensional embedded submanifold. By the definition of embedded submanifolds, each point has a regular neighbourhood, and is hence isolated. □

Remark 2. *If M is compact, then not only must there be finitely many critical points, but there must be at least one (two if $\dim M > 0$), since the minimum and maximum value must be achieved, and these are automatically critical.*

Nondegenerate symmetric bilinear forms on a n-dimensional vector space V are classified by the signature (Sylvester's signature theorem), or equivalently we will classify them by the maximal dimension of a subspace $U \subset V$ on which the bilinear form is *negative-definite*. Using this, we obtain a numerical invariant at each critical point of a Morse function.

Definition 4. The *Morse index* λ_p of a critical point $p \in \mathrm{Crit}(f)$ is the maximal dimension of subspaces $U \subset T_pM$ on which $D_p^2 f$ is negative definite. It is an integer between 0 and $\dim M$ which indicates the number of directions in which f is decreasing. In still other words, it is the number of negative entries in the diagonalization of the bilinear form $D_p^2 f$.

It will be convenient for us to package the data of the Morse indices in a generating function, called the Morse polynomial:

$$\mathcal{M}_t(f) = \sum_{p \in \mathrm{Crit}(f)} t^{\lambda_p} = \sum_{\lambda} \mu_f(\lambda) t^{\lambda},$$

where the coefficients $\mu_f(\lambda)$ are simply the number of critical points with index λ – these are known as the Morse numbers of f. This generating function may even be used in cases with infinitely many critical points, as long as there are only finitely many critical points of a given index.

Example 1.2. *Let $f = x^0$ be the height function on the sphere $S^n = \{\sum_0^n (x^i)^2 = 1\} \subset \mathbb{R}^{n+1}$. Then $\mathrm{Crit}(f) = \{N, S\}$ consists of the North and South poles. Furthermore $\lambda_N = n$ while $\lambda_S = 0$, and*

$$\mathcal{M}_t(f) = 1 + t^n.$$

Example 1.3. *Let f and g be Morse functions on M, N respectively. Consider the product $M \times N$, with its projection maps p_M, p_N to M, N respectively. Then $p_M^* f + p_N^* g$ is a function on $M \times N$, whose derivative $p_M^* df + p_N^* dg$ vanishes on $\mathrm{Crit}(f) \times \mathrm{Crit}(g)$, and given such a critical point (p, q), we have the Morse index $\lambda_p + \lambda_q$. Hence we see that*

$$\mathcal{M}_t(p_M^* f + p_N^* g) = \mathcal{M}_t(f)\mathcal{M}_t(g).$$

In this way we obtain, for example, a Morse function on the n-torus T^n from that given on S^1 above.

Example 1.4. *The function $f = \sum_{i=0}^n \lambda_i |z_i|^2$, restricted to $S^{2n+1} \subset \mathbb{C}^{n+1}$, is invariant under $z \mapsto e^{i\theta} z$ and so descends to a smooth function on $\mathbb{C}P^n = S^{2n+1}/S^1$. By the Lagrange multiplier method, finding critical points of f subject to $g = \sum |z_i|^2 = 1$ is the same as finding critical points of $F = f - \lambda(g-1)$. This gives*

$$dF = \sum (\lambda_i - \lambda)(z_i d\bar{z}_i + \bar{z}_i dz_i) - (g-1)d\lambda,$$

so that critical points occur when $z_j = 0$ for all $j \neq i$, and $\lambda = \lambda_i$; that is, the critical points are the coordinate axes $L_i = [0 : \cdots : 1_i : \cdots : 0]$. Near such a point, we can use coordinates z_j and $\bar{z}_j$, $j \neq i$, and $r_i = |z_i|$ for the sphere S^{2n+1}, and in these coordinates we have

$$f = \sum_{j \neq i} (\lambda_j - \lambda_i) z_j \bar{z}_j + \lambda_i,$$

so that the Hessian has eigenvalues $\{\lambda_j - \lambda_i\}_{j \neq i}$, each with multiplicity two, and one zero with multiplicity 1 (corresponding to the S^1 symmetry direction). If we take λ_i real, and if we choose $\lambda_i < \lambda_{i+1}$, then the number of negative eigenvalues at the i^th axis is $2i$:

$$\mathcal{M}_t(f) = 1 + t^2 + \cdots + t^{2n}.$$

1.2 The Morse lemma

We will be analyzing the level set structure of a Morse function to understand the topology of the underlying manifold. The reason Morse functions are particularly good for this purpose is twofold: a) **they have a local normal form**, which means that they can be completely classified in small open sets, and b) **they exist in abundance**. Near points for which f is regular, by the constant rank theorem there exist coordinates $x^1, \dots x^n$ such that $f = x^1$. Near critical points, however, the following result provides a coordinate system in which the function is simply quadratic, encoding only the value of the Morse index.

Theorem 1.5. *Let f be a Morse function on a neighbourhood U of the origin in a vector space*[1]*, and let the origin be a critical point. Then there is another neighbourhood V of the origin and a diffeomorphism*

$$\phi : V \longrightarrow \phi(V) \subset U$$

with $\phi(0) = 0$, and such that

$$\phi^* f(x) = D_0^2 f(x, x).$$

If we diagonalize the bilinear form $D_0^2 f(x, x)$, then we obtain

$$\phi^* f = -x_1^2 - \cdots - x_{\lambda_0}^2 + x_{\lambda_0+1}^2 + \cdots + x_n^2,$$

where λ_0 is the Morse index of f at 0.

Proof. (Palais 1969) The strategy of this proof is one which appears many times in geometry, and is called Moser's trick. The idea is as follows: we need a diffeomorphism ϕ such that $\phi^* f$ is the quadratic function $Q : x \mapsto \frac{1}{2} D_0^2 f(x, x)$. We will construct a family of diffeomorphisms ϕ_t going from $\phi_0 = \mathrm{Id}$ to the desired $\phi_1 = \phi$, which has the special property that

$$\phi_t^* f = (1 - t) f + tQ. \tag{2}$$

That is, it accomplishes an interpolation from f to Q (in our case, this interpolation is a simple one: it is linear).

We will construct this family of diffeomorphisms by flowing along a time-dependent vector field for time t. This vector field will remain zero at 0, so that 0 is fixed by the resulting diffeomorphism.

The vector field X_t giving rise to a flow ϕ_t is generally defined by the equation

$$\tfrac{d}{dt}(\phi_t^* f) = X_t(\phi_t^* f).$$

In view of equation (2), this can be written (using X(f) = df(X))

$$Q - f = X_t((1 - t) f + tQ) = d((1 - t) f + tQ)(X_t).$$

This is an improvement because now this is a linear equation for X_t: we intend to solve for X_t in the above equation. We will first rewrite it as an equation $R(X_t, x) = S(x, x)$, and solve the vector equation $RX_t = Sx$ via $X_t = R^{-1} Sx$. This will then solve the above scalar equation.

In the same way that $Q(x) = \frac{1}{2} D_0^2 f(x, x)$ comes from a bilinear operator $D_0^2 f$, we can write f, using Taylor's theorem, as follows (assume $f(0) = 0$ for simplicity):

$$f(x) = \int_0^1 (1 - s) D_{sx}^2 f(x, x) ds.$$

This general formula is obtained by integration by parts, along the line from 0 to x. Then we define the family of bilinear forms

$$S|_x = \tfrac{1}{2} D_0^2 f - \int_0^1 [(1 - s) D_{sx}^2 f] ds,$$

[1] The proof of this theorem works for any $f \in C^{k+2}(U, \mathbb{R})$ and U a neighbourhood of the origin in a Banach space, yielding a diffeomorphism of class C^k.

so that $S|_x(x,x) = Q - f$. By Taylor's theorem again, we can write the derivative of f as an integral of the second derivative:

$$df|_x = \int_0^1 D^2_{sx} f(x,\cdot) ds.$$

Using this, we define the time-dependent family of operators

$$R|_{x,t} = (1-t)\int_0^1 D^2_{sx} f ds + tD^2_0 f,$$

so that $R|_{x,t}x = d((1-t)f + tQ)$. So, the equation we must solve is $R(x, X_t) = S(x,x)$. Instead, we will solve the vector equation $RX_t = Sx$.

Note that the operator R coincides with $D^2_0 f$ at $x = 0$ for all times $t \in [0,1]$. Since $D^2_0 f$ is nondegenerate, there exists a neighbourhood W of the origin where R is nondegenerate for all $t \in [0,1]$. So, in this neighbourhood we write

$$X_t|_x = R^{-1}Sx.$$

Clearly $X_t|_0 = 0$ for all t, so that 0 is a fixed point.

We have constructed a vector field on an open set W, so that by the Picard-Lindelöf theorem for ODEs, there is an open set surrounding $W \times \{0\} \subset W \times I$ where the flow φ_t is defined. However, since $0 \in W$ is a fixed point, we see that there must be a possibly smaller neighbourhood $V \subset W$ where the flow is defined for all $t \in [0,1]$, as required. $\square$

Exercise 3. Consider the standard Morse function $f = -x_1^2 - \cdots - x_{\lambda_0}^2 + x_{\lambda_0+1}^2 + \cdots + x_n^2$ on $\mathbb{R}^\lambda \times \mathbb{R}^{n-\lambda}$. Label the coordinates $(x,y) \in \mathbb{R}^\lambda \times \mathbb{R}^{n-\lambda}$ for convenience. Show that the -1 level set

$$-x^2 + y^2 = -1$$

is homotopic to $S^{\lambda-1}$ while the $+1$ level set is homotopic to $S^{n-\lambda-1}$. Show also that the zero level set is a cone on $S^{\lambda-1} \times S^{n-\lambda-1}$ and is homotopic to a point. Use $\cosh, \sinh$ to see the explicit homotopies.

As we pass through the zero level, the level sets can be viewed as undergoing a surgery and the sublevel sets can be seen to undergo a handle attachment. Investigate this process in low dimensions such as 1,2,3,4.

Remark 3. *There is a somewhat simpler proof of this result in Milnor's book, using a method which does not generalize to infinite dimensions. It may be helpful to see that proof as well.*