MORSE INFORMATION

JACOB CLEVELAND

But mathematics is the sister, as well as the servant, of the arts
and is touched by the same madness and genius.
- Marston Morse

Abstract. This paper carefully examines two important fundamental results about Morse functions on compact manifolds. The first result heavily restricts what Morse functions look like around a critical point, and allows one to instantly recognize certain smooth functions as not being Morse. The second result, somewhat contrary to the first result, shows that all smooth functions are arbitrarily close to some Morse function. Together, this means that although Morse functions are a heavily restricted class of functions, they are relevant to considering any smooth function on a compact manifold. These results form the foundation of Morse theory for compact manifolds of general dimension. We summarize the proofs from [Mat02] of these results and fill in some of the small gaps in the lemmas and supporting definitions.

Acknowledgements. The paper satisfies the written portion of the Qualifying Exam Part B requirements of the Mathematics PhD program at Colorado State University. This work is intended to compliment the associated presentation given to the part B committee consisting of Dr. Henry Adams and Dr. Amit Patel.

Acknowledgements are in order to Brian Collery for pointing out the first step of the proof of the Fundamental Theorem of Calculus below, Lemma 1.1, as well as Ignacio Rojas for the succinct picture explaining Lemma 1.2, also to Michael Moy for helping clarify some details in the initially ham-fisted proof of Lemma 2.4, and lastly to Yukio Matsumoto for writing such a succinct yet approachable introduction for Morse theory, without whom I would still be hopelessly flipping back and forth between Milnor and Wikipedia.

I dedicate this work to G. Griff Elder, whose careful mentorship inspired me to pursue math, and without whom I would not be on the path I am on now.

1. The Morse Lemma

In this section we state the Morse Lemma, sketching the proof and examining some explicit examples. We begin with an m-dimensional version of the Fundamental Theorem of Calculus, used in the proof of the Morse Lemma in [Mat02].

Lemma 1.1 (Fundamental Theorem of Calculus)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ has partial derivatives in all n coordinates and $f(0, \dots, 0) = 0$. Then*

$$f(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i \int_0^1 \frac{\partial f}{\partial x_i}(tx_1, \dots, tx_n)dt$$

Date: March 18, 2023.

Proof. Indeed, by the 1-dimensional Fundamental Theorem of Calculus,

$$\begin{aligned} f(x_1,\dots,x_n) &= f(tx_1,\dots,tx_n)\Big|_{t=0}^{1} \\ &= \int_0^1 \frac{df}{dt}(tx_1,\dots,tx_n)dt \\ &= \int_0^1 \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}(tx_1,\dots,tx_n)dt \\ &= \sum_{i=1}^n x_i \int_0^1 \frac{\partial f}{\partial x_i}(tx_1,\dots,tx_n)dt \end{aligned}$$

□

In fact, the following is also true using $g_i(t)$ where $g_i(0) = 0$, $g_i(1) = x_i$, and $g_i(t)$ is differentiable, so that

$$\begin{aligned} f(x_1,\dots,x_n) &= \int_0^1 \frac{df}{dt}(g_1(t),\dots,g_n(t))dt \\ &= \sum_{i=1}^n \int_0^1 \left(\frac{\partial g_i}{\partial t}(t)\right)\left(\frac{\partial f}{\partial x_i}(g_1(t),\dots,g_n(t))\right)dt \end{aligned}$$

Then Lemma 1.1 follows from using $g_i(t) = x_i t$.

The following lemma is used in the inductive step at the end of the proof of Theorem 1.3.

Lemma 1.2. *For $A_{ij} \in G$ with $(G,+)$ some abelian group,*

$$\sum_{i,j=1}^n A_{ij} = A_{11} + \sum_{j=2}^n A_{1j} + \sum_{i=2}^n A_{i1} + \sum_{i,j=2}^n A_{ij} \tag{$\flat$}$$

Proof. Let

$$A = \left[\begin{array}{c|ccc} A_{11} & A_{12} & \cdots & A_{1n} \\ \hline A_{21} & A_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ A_{n1} & \cdots & \cdots & A_{nn} \end{array}\right]$$

Then ($\flat$) is equivalent to the statement that summing up all entries in A is the same as adding A_{11}, the sum of the entries of the column vector $[A_{i1} : 2 \le i \le n]$, the entries of the row vector $[A_{1j} : 2 \le j \le n]$, and the entries of the matrix $[A_{ij} : 2 \le i, j \le n]$. □

Now, we state the Morse Lemma and sketch the proof seen in [Mat02].

Theorem 1.3 (Morse Lemma)**.** *Let $f : M \to \mathbb{R}$ be a smooth function with non-degenerate critical point p_0. Then one may choose a local coordinate system $(x_1, x_2, \dots, x_n)$ about p_0 such that f, when represented using these local coordinates, has the form*

$$f = -X_1^2 - \cdots - X_\lambda^2 + X_{\lambda+1}^2 + \cdots + X_n^2 + c \tag{$\spadesuit$}$$

where λ is the index of f at p_0.

Proof. Choose local coordinates $(x_1, \dots, x_n)$ at the critical point p_0 such that p_0 corresponds to the origin $(0, \dots, 0)$. Without loss of generality, $f(p_0) = 0$ since one may consider $f - f(p_0)$ instead of f. Therefore, $f(0, \dots, 0) = 0$ and we may apply Lemma 1.1 so that

$$(\clubsuit) \qquad f(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i \int_0^1 \frac{\partial f}{\partial x_i}(tx_1, \dots, tx_n)dt.$$

Let $g_i(x_1, \dots, x_n) = \int_0^1 \frac{\partial f}{\partial x_i}(tx_1, \dots, tx_n)dt$, defined in a neighborhood of the origin. Then $g_i(0, \dots, 0) = 0$ for all i because p_0 is a critical point of f. So we may apply Lemma 1.1 again to each g_i so that

$$(\diamond) \qquad g_i(x_1, \dots, x_n) = \sum_{j=1}^{n} x_j \int_0^1 \frac{\partial g_i}{\partial x_j}(tx_1, \dots, tx_n)dt.$$

Similarly, we let $H_{ij}(x_1, \dots, x_n) = \int_0^1 \frac{\partial g_i}{\partial x_j}(tx_1, \dots, tx_n)dt$ defined in a neighborhood of the origin. To see that $H_{ij} = H_{ji}$, observe[1] that

$$\begin{aligned}
\frac{\partial g_i}{\partial x_j} &= \frac{\partial}{\partial x_j} \int_0^1 \frac{\partial f}{\partial x_i}(x_1 t, \dots, x_n t)dt \\
&= \int_0^1 \frac{\partial^2 f}{\partial x_j \partial x_i}(x_1 t, \dots, x_n t)dt \\
&= \int_0^1 \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1 t, \dots, x_n t)dt \\
&= \frac{\partial}{\partial x_i} \int_0^1 \frac{\partial f}{\partial x_j}(x_1 t, \dots, x_n t)dt \\
&= \frac{\partial g_j}{\partial x_i}
\end{aligned}$$

Then integrating these as in Lemma 1.1, we see[2] that $H_{ij} = H_{ji}$. Combining equations $(\clubsuit)$ and $(\diamond)$ yields

$$(\heartsuit) \qquad f(x_1, \dots, x_n) = \sum_{i,j=1}^{n} x_i x_j H_{ij}(x_1, \dots, x_n)$$

Equation $(\spadesuit)$ is called the 'Morse form' (or standard form) of f, which is a special type of quadratic form for f. The goal at this point is to rewrite the representation given in $(\heartsuit)$ in standard form. This is accomplished inductively on the number of terms in $(\heartsuit)$.

Recall that p_0 is a non-degenerate critical point of f. This means that $\det(H_f(p_0)) \neq 0$ i.e. the Hessian matrix is non-singular, so after a suitable linear change of local coordinates, this implies that $\frac{\partial^2 f}{\partial x_1^2}(0, \dots, 0) \neq 0$. Note that by differentiating $(\heartsuit)$, one sees that $\frac{\partial^2 f}{\partial x_1^2}(0, \dots, 0) = 2H_{11}(0, \dots, 0)$ so $H_{11}(0, \dots, 0) \neq 0$. Moreover, H_{11} is continuous in a neighborhood of the origin so H_{11} is not zero on a neighborhood of the origin.

Now one can define a new system of coordinates, $(X_1, x_2, \dots, x_n)$ where

$$X_1 = \sqrt{|H_{11}|}\left(x_1 + \sum_{i=2}^{n} x_i \frac{H_{1i}}{H_{11}}\right)$$

[1]Being able to differentiate under the integral sign here follows from the fact that f is twice differentiable, hence g_i is continuous, so we are allowed to use the Leibniz Integral Rule.

[2]The text [Mat02] uses h_{ij} instead and defines $H_{ij} = (h_{ij} + h_{ji})/2$ to get $H_{ij} = H_{ji}$ without much extra work, but I think it's more beneficial to understand why $H_{ij} = H_{ji}$ as we have defined it.

which means

$$X_1^2 = |H_{11}|\left(x_1 + \sum_{i=2}^{n} x_i \frac{H_{1i}}{H_{11}}\right)^2$$
$$= \begin{cases} H_{11}x_1^2 + 2\sum_{i=2}^{n} x_1 x_i H_{1i} + (\sum_{i=2}^{n} x_i H_1 i)^2/H_{11} & (H_{11} > 0) \\ -H_{11}x_1^2 - 2\sum_{i=2}^{n} x_1 x_i H_{1i} - (\sum_{i=2}^{n} x_i H_1 i)^2/H_{11} & (H_{11} < 0) \end{cases}$$

Alternatively,

$$H_{11}x_1^2 + 2\sum_{i=2}^{n} x_1 x_i H_{1i} + \left(\sum_{i=2}^{n} x_i H_1 i\right)^2 / H_{11} = \begin{cases} X_1^2 & (H_{11} > 0) \\ -X_1^2 & (H_{11} < 0) \end{cases}$$

Observe that

$$(\mho) \qquad \pm X_1^2 + \sum_{i,j=2}^{n} x_i x_j H_{ij} - (\sum_{i=2}^{n} x_i H_{1i})^2/H_{11} = H_{11}x_1^2 + \sum_{i=2}^{n} x_1 x_i H_{1i} + \sum_{i=2}^{n} x_i x_1 H_{i1} + \sum_{i,j=2}^{n} x_i x_j H_{ij}$$

By Lemma 1.2 above with $A_{ij} = x_i x_j H_{ij}$, the right hand side of ($\mho$) is equal to ($\heartsuit$), so that

$$f(X_1, x_2, \dots, x_n) = \pm X_1^2 + \sum_{i,j=2}^{n} x_i x_j H_{ij} - (\sum_{i=2}^{n} x_i H_{1i})^2/H_{11}$$

depending if H_{11} is positive or negative. The remaining sum $f - (\pm X_1^2)$ is a function with one fewer term which can be rewritten using the same process. Hence proceeding by induction, such an f can always be represented in Morse form. □

Corollary 1.4. *Non-degenerate critical points are isolated.*

If a non-degenerate critical point could be not isolated on the manifold M, say critical point p_0, then every neighborhood of p_0 would contain a critical point. But then there would be no coordinate neighborhood of p_0 where f could be represented as in (♠), because this only has one critical point. In other words, (♠) forces critical points to be 'spread out' in some sense.

Corollary 1.5. *A Morse function on a compact manifold admits only finitely many critical points.*

Suppose M is compact, and has a Morse function f with infinitely many critical points p_i. Then there exists a compact subset K of a coordinate neighborhood U with K containing infinitely many of the p_i. Since K is homeomorphic to a compact subset of $\mathbb{R}^n$, there exists a convergent subsequence of the infinitely many p_i in K. However, the limit of this subsequence would not be isolated, which violates Corollary 1.4.

Remark 1.6. *Theorem 1.3 only applies to non-degenerate critical points, because among other things, one needs the index of the critical point, which might not be stable under small pertubations if the critical point is degenerate.*

2. Morse Approximations of Smooth Functions

In this section, we summarize the fact that loosely, Morse functions are abundant in the space of smooth functions $f : M \to \mathbb{R}$.

Lemma 2.1 (Linear Morsification). *For $U \subseteq \mathbb{R}^m$ open, $f : U \to \mathbb{R}$ smooth, there exist $a_1, \ldots, a_m \in \mathbb{R}$ such that $f(x_1, \ldots, x_m) - (a_1x_1 + \cdots + a_mx_m)$ is Morse.*

This lemma intuitively states that one can always subtract a suitably 'flat' hyperplane from the function to eliminate all degenerate critical points of f. Note that in the process, one may also gain non-degenerate critical points.

Definition 2.2. *Suppose $f, g : M \to \mathbb{R}$, $U \subseteq M$ a coordinate neighborhood of M with coordinates $(x_1, \ldots, x_n)$, and K a compact subset of U. Given $\varepsilon > 0$, f is said to be a (C^2, ε) approximation of g in K if the following inequalities hold for all $p \in K$:*

$$\begin{cases} |f(p) - g(p)| < \varepsilon & & (C^0 \text{ closeness}) \\ |\frac{\partial f}{\partial x_i} - \frac{\partial g}{\partial x_i}| < \varepsilon & i = 1, \ldots, n & (C^1 \text{ closeness}) \\ |\frac{\partial^2 f}{\partial x_i \partial x_j} - \frac{\partial^2 g}{\partial x_i \partial x_j}| < \varepsilon & i, j = 1, \ldots, n & (C^2 \text{ closeness}) \end{cases}$$

Loosely speaking, this definition of (C^2, ε) is saying that f and g are C^2 close whenever their function values, first partials, and second partials are close.

To define a (C^2, ε) approximation on a compact manifold M, pick compact subsets K_i of M such that $M = \cup_i K_i$, and associate to this covering by compact sets the covering by coordinate neighborhoods $M = \cup_i U_i$ where $K_i \subseteq U_i$. We know such a covering by compact subsets exists by the following: Fix compact manifold M. Consider all possible coordinate neighborhoods U of M, and for each U, consider all possible m-disks D. Note that since the U range over all points in M, the interiors $\text{int}(D)$, which are open, forms an open cover of M. Hence there is finite subcover $\text{int}(D_i)$ of M. Note that D_i, which are compact, must then also cover M. So M can be covered by finitely many compact subsets D_i contained inside finitely many coordinate neighborhoods U_i.

Definition 2.3. *Given $M = \cup_i K_i$ a compact covering, a function $f : M \to \mathbb{R}$ is a (C^2, ε)-**approximation** of function $g : M \to \mathbb{R}$ if f is a (C^2, ε) approximation of g on K_i for each $i = 1, \ldots, n$.*

The next lemma loosely establishes that non-degeneracy of critical points is preserved when moving from a function to a (C^2, ε) approximation.

Lemma 2.4. *Let C be a compact subset of an m-dimensional manifold M. Suppose that $g : M \to \mathbb{R}$ has no degenerate critical point in C. Then for a sufficiently small $\varepsilon > 0$, any (C^2, ε)-approximation f of g has no degenerate critical point in C.*

Proof. Fix coordinate neighborhood U_i with coordinates $(x_1, \ldots, x_n)$ and associated compact K_i. Also fix a smooth function $g : M \to \mathbb{R}$. Denote

$$H_g = \left[\frac{\partial^2 g}{\partial x_i \partial x_j} \right]$$

the Hessian of g with respect to these coordinates.

Indeed, there are no non-degenerate critical points of g in $C \cap K_i$ iff the condition

$$(\sharp) \qquad \left| \frac{\partial g}{\partial x_1} \right| + \cdots + \left| \frac{\partial g}{\partial x_n} \right| + \left| \det \left(\frac{\partial^2 g}{\partial x_i \partial x_j} \right) \right| > 0$$

holds on $C \cap K_i$. This can be seen by the following argument.

Suppose p_0 is a critical point of g. Then evaluated at p_0, each first derivative $\frac{\partial g}{\partial x_i} = 0$. Then p_0 by definition is non-degenerate iff $\det \left(\frac{\partial^2 g}{\partial x_i \partial x_j} \right) \neq 0$ iff $\left| \det \left(\frac{\partial^2 g}{\partial x_i \partial x_j} \right) \right| > 0$.

Now suppose p_0 is not a critical point of g. Then $\frac{\partial g}{\partial x_i} \neq 0$ at p_0 for some index i, and so condition (♯) is satisfied, even if $\det\left(\frac{\partial^2 g}{\partial x_i \partial x_j}\right) = 0$.

Next observe that a version of condition (♯) is also true of any (C^2, ε) approximation f of g for ε small enough. To see this, suppose g has no non-degenerate critical points in $C \cap K_i$. Then condition (♯) holds for g. Since det() is a continuous function of matrices in the sense that two matrices A and B are close if pairwise, their entries are at most ε apart, then there exists ε such that[3]

$$\left|\frac{\partial^2 f}{\partial x_i \partial x_j}(p_0) - \frac{\partial^2 g}{\partial x_i \partial x_j}(p_0)\right| < \varepsilon$$

implies that

$$\left|\det(H_f(p_0)) - \det(H_g(p_0))\right| < |\det(H_g(p_0))|$$

for all p_0 in $C \cap K_i$, and f some (C^2, ε) approximation of g with ε given above.

This forces $H_f(p_0) \neq 0$ so any critical point of f is non-degenerate on $C \cap K_i$. Note that if p_0 is not a critical point of f, even if it is a critical point of g, it doesn't matter the values of the second partials of f at p_0. Thus we have shown that for any (C^2, ε) approximation f of g, f also has no degenerate critical points on $C \cap K_i$, or equivalently, (♯) is true for f. Repeating this process for all K_i, $i = 1, \ldots, n$, yields n values for ε, so pick the smallest to yield the intended result. □

Lemma 2.5 (Continuous Functions Send Compacts to Compacts). *If $f : K \to \mathbb{R}$ is continuous on K compact, then $f(K)$ is compact as well.*

Proof. Let $C = f(K)$. Then C has a (possibly infinite) open covering $C = \cup_i U_i$. Since f is continuous, $f^{-1}(U_i)$ is open in K. Since f is onto C its image, every point in C comes from a point in K, so the preimage of the open covering is a covering of K, i.e. $K = \cup_i f^{-1}(U_i)$. Then since K is compact, there are finitely many U_i such that $K = \cup_{i=1}^n f^{-1}(U_i)$. Then C is covered by finitely many open sets, $C = \cup_{i=1}^n f(f^{-1}(U_i)) = \cup_{i=1}^n U_i$. □

In particular, f achieves a maximum and minimum value. So, if f is smooth, then each partial derivative of each order also achieves maximum and minimum values, and hence are each bounded.

Lemma 2.6 (Existence of Step Functions). *For subsets of a manifold $K \subset V \subset L \subset U$, where K and L are compact, V is open, and U is a coordinate neighborhood (i.e. open), then there exists a smooth function $h : U \to \mathbb{R}$ where*

(i) $0 \leq h \leq 1$
(ii) $h(V) = 1$
(iii) $h(U - L) = 0$

Theorem 2.7 (Existence of Morse Functions). *For M a closed m-manifold, $g : M \to \mathbb{R}$ a smooth function defined on M, there exists a Morse function $f : M \to \mathbb{R}$ where f is a (C^2, ε)-approximation of g.*

Proof. Choose a covering of M by compact sets K_l each contained in coordinate neighborhood U_l, i.e.

$$\begin{aligned} M &= K_1 \cup \cdots \cup K_n \\ &= U_1 \cup \cdots \cup U_n \end{aligned}$$

The proof inductively constructs functions f_l such that f_l has no degenerate critical points in $C_l := K_1 \cup \cdots \cup K_l$. Take $g : M \to \mathbb{R}$ smooth as f_0, and set $C_0 = \emptyset$. As the inductive hypothesis, suppose we have $f_{l-1} : M \to \mathbb{R}$ with no degenerate critical points in C_{l-1}.

We proceed by looking at U_l with coordinates $(x_1, \ldots, x_m)$. By Lemma 2.1, there exist $a_1, \ldots, a_n$ small enough such that

$$f_{l-1}(x_1, \ldots, x_n) - (a_1 x_1 + \cdots + a_m x_m) \tag{♮}$$

[3]Note that this first inequality is equivalent to saying that the Hessian matrices are close in the sense above.

is Morse on U_l. Moreover, applying Lemma 2.6 to $K_l \subseteq U_l$, one gets a smooth step function $h_l : U_l \to \mathbb{R}$. Let L_l be the compact subset between K_l and U_l as in Lemma 2.6. Then define

$$f_l = \begin{cases} f_{l-1}(x_1, \ldots, x_m) - (a_1x_1 + \cdots + a_mx_m)h_l(x_1, \ldots, x_m) & \text{(in } U_l\text{)} \\ f_{l-1}(x_1, \ldots, x_m) & \text{(outside } L_l\text{)} \end{cases}$$

which is well defined, because $h_l = 0$ in $U_l - L_l$. Moreover, function (♮) and f_{l-1} agree on some open neighborhood containing the compact K_l, since $h_l = 1$ in K_l. Thus since (♮) has no degenerate critical points on K_l, f_l also has no degenerate critical points on K_l, and is a smooth function on U_l.

Now we wish to determine constraints on the a_i so that f_l is a (C^2, ε) approximation of f_{l-1}. First restricting $p = (x_1, \ldots, x_m)$ to U_l, we apply Definition 2.2 to obtain the following:

$$(\circledast) \qquad \begin{cases} |f_{l-1}(p) - f_l(p)| = |(a_1x_1 + \cdots + a_mx_m)|h_l(p) \\ \left|\frac{\partial f_{l-1}}{\partial x_i} - \frac{\partial f_l}{\partial x_i}\right| = |a_ih_l(p) + (a_1x_1 + \cdots + x_mx_m)\frac{\partial h_l}{\partial x_i}(p)|, & i = 1, 2, \ldots, m \\ \left|\frac{\partial^2 f_{l-1}}{\partial x_i\partial x_j} - \frac{\partial^2 f_l}{\partial x_i\partial x_j}\right| = |a_i\frac{\partial h_l(p)}{\partial x_j} + a_j\frac{\partial h_l(p)}{\partial x_i} + (a_1x_1 + \cdots + x_mx_m)\frac{\partial^2 h_l}{\partial x_i\partial x_j}(p)|, & i, j = 1, 2, \ldots, m \end{cases}$$

So, by Lemma 2.5, the partial derivatives of h_l are all bounded, so one can choose the a_i small enough to make the right hand side of ($\circledast$) arbitrarily small. This means f_l can be made a (C^2, ε) approximation of f_{l-1} in the compact set K_l.

Now to constrain the a_i to get a (C^2, ε) approximation of f_{l-1} in the other compact sets K_j, consider the coordinate neighborhood U_j containing K_j with coordinates $(y_1, \ldots, y_m)$. Recall that outside L_l, $f_l = f_{l-1}$ so we should restrict our attention to $K_j \cap L_l \subseteq U_j \cap U_l$. We can rewrite ($\circledast$) using the appropriate coordinate transformation from $(x_1, \ldots, x_m)$ to $(y_1, \ldots, y_m)$. Furthermore, the components of the Jacobian of this coordinate transformation are each continuous and hence are bounded, so we may again choose the a_i small enough to make the right hand side of the transformed version of ($\circledast$) arbitrarily small. Repeating this process for each K_j means we can define f_l as a (C^2, ε) approximation on M for any value of ε by taking the smallest of each a_i for each K_j.

Moreover, since f_{l-1} has no degenerate critical points on C_{l-1}, one can pick ε small enough such that f_l also has no degenerate critical points on C_{l-1} by Lemma 2.4. Lastly, since f_l was constructed to not have degenerate critical points on K_l, f_l has no degenerate critical points on $C_l = C_{l-1} \cup K_l$. Proceeding by induction, one obtains $f = f_k$, the (C^2, ε) approximation of g with no degenerate critical points on M. □

References

[Mat02] Yukio Matsumoto. *An Introduction to Morse Theory*, volume 208 of *Translations of Mathematical Monographs*. American Mathematical Society, 1 edition, 2002.