

A PROOF OF THE MORSE-BOTT LEMMA

AUGUSTIN BANYAGA AND DAVID E. HURTUBISE

ABSTRACT. It is often said that the Morse-Bott Lemma can be viewed as a “parameterized” Morse Lemma, and its proof should follow from the differentiability of the methods used to prove the Morse Lemma. The goal of this expository paper is to fill in the details. We present Palais’ proof of the Morse Lemma using Moser’s path method, which yields the necessary differentiability.

1. Introduction and Preliminaries

The goal of this expository paper is to give a complete proof of the Morse-Bott Lemma: a classical result, but one for which we do not know a complete proof in the literature. Usually authors just say that the Morse-Bott Lemma is a parameterized version of the Morse Lemma, and its proof follows from the differentiability of the methods used to prove the Morse Lemma [1], [2], [3]. This is actually true, but there are further subtleties involved. We take the opportunity in this exposition to give Palais’ proof of the Morse Lemma using Moser’s path method, which yields the differentiability needed to prove the Morse-Bott Lemma.

Let $f : M \rightarrow \mathbb{R}$ be a smooth function on a smooth manifold M . A critical point of f is a point $p \in M$ at which the differential $df_p : T_p M \rightarrow T_{f(p)} \mathbb{R} \approx \mathbb{R}$ is zero. Here $T_p M$ denotes the tangent space to M at p .

For each critical point p of f , there is a bilinear symmetric form $H_p(f)$ on $T_p M$, called the Hessian of f at p , defined as follows:

$$H_p(f)(V, W) = V \cdot (\tilde{W} \cdot f)(p)$$

where $V, W \in T_p M$ and \tilde{W} is an extension of W into a vector field defined near p [6]. If the Hessian $H_p(f)$ is a non-degenerate bilinear form, then we say that the critical point p is a non-degenerate critical point. The dimension of the subspace on which $H_p(f)$ is negative definite is called the index of p (sometimes called “the Morse index”).

A C^2 function $f : M \rightarrow \mathbb{R}$ is called a Morse function if all of its critical points are non-degenerate.

1991 *Mathematics Subject Classification.* 57R70; 53C15; 58E05.

Key words and phrases. critical points, hessian, nondegenerate critical points, Morse index, critical submanifold, Morse index of a critical submanifold, Morse function, Morse-Bott function.

Morse Lemma [6], [7]. *Let $p \in M$ be a nondegenerate critical point of a smooth function $f : M \rightarrow \mathbb{R}$. There exists a smooth chart $\phi : U \rightarrow \mathbb{R}^m$, where U is an open neighborhood of p with $\phi(p) = 0$, such that if $\phi(x) = (x_1, \dots, x_m)$ for $x \in U$, then*

$$(f \circ \phi^{-1})(x_1, \dots, x_m) = f(p) - x_1^2 - x_2^2 - \dots - x_k^2 + x_{k+1}^2 + x_{k+2}^2 + \dots + x_m^2$$

where k is the index of p .

It is easy to see that non-degenerate critical points are isolated. Hence, if f is a Morse function, then the set $Crit(f)$ of critical points is a 0 - dimensional manifold.

We now consider a function $f : M \rightarrow \mathbb{R}$ whose critical set is a disjoint union $\bigsqcup_i C_i$ of connected smooth submanifolds of dimensions $d_i \geq 0$, called critical submanifolds. Pick a Riemannian metric on M and use it to split $T_*M|_{C_i}$ as

$$T_*M|_{C_i} = T_*C_i \oplus \nu_*C_i$$

where T_*C_i is the tangent bundle of C_i and ν_*C_i is the normal bundle of C_i . Let $p \in C_i$, $V \in T_pC_i$, $W \in T_pM$, and let $H_p(f)$ be the Hessian of f at p . We have

$$H_p(f)(V, W) = V_p \cdot (\tilde{W} \cdot f) = 0$$

since $V_p \in T_pC_i$ and any extension \tilde{W} of W to a vector field satisfies $df(\tilde{W})|_{C_i} = 0$. Therefore, the Hessian $H_p(f)$ induces a bilinear symmetric form $\mathcal{H}_p(f)$ on $\nu_pC_i \approx T_pM/T_pC_i$.

A smooth function $f : M \rightarrow \mathbb{R}$ is called a Morse-Bott function if $Crit(f) = \bigsqcup_i C_i$ is a disjoint union of connected smooth submanifolds and for each C_i and $p \in C_i$ the Hessian $\mathcal{H}_p(f)$ is non-degenerate [2], [3].

One says that the Hessian is non-degenerate in the direction normal to the critical submanifolds.

Morse-Bott Lemma. *Let $f : M \rightarrow \mathbb{R}$ be a Morse-Bott function, C a connected component of $Crit(f)$ of dimension n , and $p \in C$. Then there exists an open neighborhood U of p and a smooth chart $\phi : U \rightarrow \mathbb{R}^n \times \mathbb{R}^{m-n}$, where $m = \dim(M)$, such that:*

a) $\phi(p) = 0$

b) $\phi(U \cap C) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^{m-n} | y = 0\}$ and

c) $(f \circ \phi^{-1})(x, y) = f(C) - y_1^2 - y_2^2 - \dots - y_k^2 + y_{k+1}^2 + \dots + y_{m-n}^2$

where $k \leq m - n$ is the index of $\mathcal{H}_p(f)$ and $f(C)$ is the common value of f on C .

This lemma implies that the index k of $\mathcal{H}_p(f)$ is locally constant, so it is the same for all points $p \in C$. We say that C is a critical manifold of index k .

We need first to recall the following Taylor formulas of order one and two:

Scholium. Let $f : U \rightarrow \mathbb{R}$ be a C^∞ function on a convex neighborhood U of $0 \in \mathbb{R}^m$. If $f(0) = 0$, then there exist C^∞ functions g_i with $g_i(0) = \frac{\partial f}{\partial x_i}(0)$ such that:

$$f(x_1, \dots, x_m) = \sum_{i=1}^m x_i g_i(x_1, \dots, x_m).$$

If $f(0) = \frac{\partial f}{\partial x_i}(0) = 0$ for $i = 1, \dots, m$ then there exist C^∞ functions h_{ij} with $h_{ij}(0) = \frac{\partial^2 f}{\partial x_i \partial x_j}(0)$ such that

$$f(x_1, \dots, x_m) = \sum_{ij} x_i x_j h_{ij}(x_1, \dots, x_m) = {}^t x S_x x$$

where x is the column vector transpose of ${}^t x = (x_1, \dots, x_m)$, and $S_x = (s_{ij}(x))$ is the symmetric matrix with entries

$$s_{ij}(x) = \frac{1}{2}(h_{ij}(x) + h_{ji}(x)).$$

Proof:

We have:

$$\frac{d}{dt} f(tx) = \sum_{i=1}^m x_i \frac{\partial f}{\partial x_i}(tx).$$

Hence, setting

$$g_i(x_1, \dots, x_m) = \int_0^1 \frac{\partial f}{\partial x_i}(tx) dt$$

we observe that

$$g_i(0) = \int_0^1 \frac{\partial f}{\partial x_i}(0) dt = \frac{\partial f}{\partial x_i}(0),$$

and we get:

$$\begin{aligned} f(x) &= f(x) - f(0) = \int_0^1 \frac{d}{dt} f(tx) dt \\ &= \int_0^1 \sum_{i=1}^m x_i \frac{\partial f}{\partial x_i}(tx) dt \\ &= \sum_{i=1}^m x_i g_i(x_1, \dots, x_m). \end{aligned}$$

This is a Taylor formula of order one.

Applying the preceding fact again to $g_i(x_1, \dots, x_m)$, and assuming $g_i(0) = 0$, we have

$$g_i(x_1, \dots, x_m) = \sum_{j=1}^m x_j h_{ij}(x_1, \dots, x_m)$$

where the h_{ij} are C^∞ functions with $h_{ij}(0) = \frac{\partial g_i}{\partial x_j}(0) = \frac{\partial^2 f}{\partial x_i \partial x_j}(0)$.

We have

$$\begin{aligned} f(x_1, \dots, x_m) &= \sum_{i=1}^m x_i \sum_{j=1}^m x_j h_{ij}(x_1, \dots, x_m) \\ &= \sum_{i,j} x_i x_j h_{ij}(x_1, \dots, x_m) \\ &= {}^t x S_x x. \end{aligned}$$

The expression $f(x) = {}^t x S_x x$ is a Taylor formula of order two. □

2. Palais' proof of the Morse Lemma [4], [9]

By replacing f by $f - f(p)$ and by choosing a suitable coordinate chart on M we may assume that the function f is defined on a convex neighborhood U_0 of $0 \in \mathbb{R}^m$ where $f(0) = 0$, $df(0) = 0$, and the matrix of the Hessian at $0 \in \mathbb{R}^m$,

$$M_0(f) = A = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(0) \right),$$

is a diagonal matrix with the first k diagonal entries equal to -1 and the rest equal to $+1$.

The matrix A induces a function $\tilde{A} : \mathbb{R}^m \rightarrow \mathbb{R}$ given by

$$\tilde{A}(x) = {}^t x A x = \langle Ax, x \rangle = \sum_{j=1}^m \delta_j x_j^2$$

where $\delta_j = \left. \frac{\partial^2 f}{\partial x_j^2} \right|_0 = \pm 1$ for all $j = 1, \dots, m$. We want to prove that there are neighborhoods U and U' of 0 with $U \subseteq U_0$ and a diffeomorphism $\varphi : U \rightarrow U'$ such that

$$f \circ \varphi = \tilde{A} \tag{1}$$

The idea of Moser's path method [8] is to interpolate f and \tilde{A} by a path such as,

$$f_t = \tilde{A} + t(f - \tilde{A}), \tag{2}$$

and to look for a smooth family φ_t of diffeomorphisms such that

$$f_t \circ \varphi_t = f_0 = \tilde{A}. \tag{3}$$

Then $\varphi = \varphi_1$ will satisfy $f \circ \varphi = \tilde{A}$.

We get φ_t as the solution of the differential equations

$$\frac{d\varphi_t}{dt}(x) = \xi_t(\varphi_t(x)); \quad \varphi_0(x) = x$$

where the smooth family ξ_t is the tangent along the curves $t \mapsto \varphi_t(x)$.

Taking the partial derivative with respect to t of both sides of (3) gives

$$(\dot{f}_t \circ \varphi_t + (\xi_t \cdot f_t) \circ \varphi_t)(x) = 0 \quad (4)$$

for all $x \in U$ where \dot{f}_t denotes $\frac{\partial f_t}{\partial t}$. Thus,

$$(\dot{f}_t + \xi_t \cdot f_t)(y) = 0 \quad (5)$$

for all $y \in U'$. But $\dot{f}_t = f - \tilde{A}$, and therefore (5) becomes

$$df_t(\xi_t) = g \quad (6)$$

where $g = \tilde{A} - f$.

Since $g(0) = \tilde{A}(0) - f(0) = 0 - 0 = 0$ and $dg(0) = d\tilde{A}(0) - df(0) = 0 - 0 = 0$, the Scholium gives:

$$g(x) = \langle G_x x, x \rangle$$

where G_x is a symmetric matrix depending smoothly on x .

The proof of the Scholium can be modified as follows to show that

$$df_t(x)(\xi_t) = \langle B_x^t \xi_t, x \rangle \quad (7)$$

where B_x^t is an $m \times m$ matrix with entries

$$(B_x^t)_{ij} = \int_0^1 \frac{\partial^2 f_t}{\partial x_i \partial x_j}(sx) ds.$$

We have,

$$\begin{aligned} \frac{d}{ds}(df_t(sx)(\xi_t)) &= \frac{d}{ds} \left(\sum_j \xi_t^j \frac{\partial f_t}{\partial x_j}(sx) \right) \\ &= \sum_{i,j} \xi_t^j \frac{\partial}{\partial x_i} \left(\frac{\partial f_t}{\partial x_j}(sx) \right) \cdot \frac{d}{ds} sx_i \\ &= \sum_{i,j} x_i \xi_t^j \frac{\partial^2 f_t}{\partial x_i \partial x_j}(sx). \end{aligned}$$

Hence, since $df_t(0) = 0$, we get

$$\begin{aligned} df_t(x)(\xi_t) &= df_t(x)(\xi_t) - df_t(0)(\xi_t) = \int_0^1 \frac{d}{ds} df_t(sx)(\xi_t) ds \\ &= \int_0^1 \sum_{i,j} x_i \xi_t^j \frac{\partial^2 f_t}{\partial x_i \partial x_j}(sx) ds \\ &= \sum_{i,j} x_i \xi_t^j (B_x^t)_{ij} \end{aligned}$$

and formula (7) follows.

Now observe that by (2)

$$B_0^t = \left(\frac{\partial^2 f_t}{\partial x_i \partial x_j} (0) \right)$$

is a diagonal matrix whose (j, j) -th entry is

$$2\delta_j + t \left(\frac{\partial^2 f}{\partial x_j^2} \Big|_0 - 2\delta_j \right) = (2-t)\delta_j$$

for all $j = 1, \dots, m$. Hence, B_0^t is non-degenerate for all $0 \leq t \leq 1$, and there exists a neighborhood \tilde{U} of $0 \in \mathbb{R}^m$ such that B_x^t is also non-degenerate for all t . For $x \in \tilde{U}$, we have a unique solution ξ_t of

$$\langle B_x^t \xi_t, x \rangle = \langle G_x x, x \rangle$$

This solution $\xi_t = (B_x^t)^{-1} G_x x$ of (5, 6), defined on \tilde{U} depends smoothly on both x and t . Clearly, $\xi_t(0) = 0$. Hence, by shrinking \tilde{U} we can integrate ξ_t and get a smooth family of diffeomorphisms φ_t from a smaller neighborhood U of 0 to another neighborhood U' of 0 which satisfies $f_t \circ \varphi_t = f_0 = \tilde{A}$.

□

3. Proof of the Morse-Bott Lemma

We will need the following lemma which is an important ingredient in Hirsch's proof of the Morse Lemma[5]:

Hirsch's Lemma. *Let $A = \text{diag}(a_1, a_2, \dots, a_m)$ be a diagonal matrix with diagonal entries $a_j = \pm 1$ for all $j = 1, \dots, m$. Then there is a neighborhood U of A in the vector space of symmetric matrices ($\approx \mathbb{R}^{m(m+1)/2}$) and a C^∞ map $P : U \rightarrow GL_m(\mathbb{R})$ such that: $P(A) = I_{m \times m}$ and for every $S \in U$, if $P(S) = Q$, then ${}^t Q S Q = A$.*

Proof:

We proceed by induction on the dimension m . First, suppose $m = 1$ and $A = (a)$ with $a = \pm 1$. If $S = (s)$ is any 1×1 matrix sufficiently close to A , then s will be non-zero with the same sign as a , and we define

$$P(S) = Q = \left(\frac{1}{\sqrt{|s|}} \right).$$

Now let $A = \text{diag}(a_1, a_2, \dots, a_m)$ where $a_j = \pm 1$ for all $j = 1, \dots, m$, and assume for the purpose of induction that there is a neighborhood U_1 of $A_1 = \text{diag}(a_2, \dots, a_m)$ in the vector space of $(m-1) \times (m-1)$ symmetric matrices such that for every $S_1 \in U_1$ there exists a smooth map $P_1 : U_1 \rightarrow GL_{m-1}(\mathbb{R})$ such that $P_1(A_1) = I_{(m-1) \times (m-1)}$ and ${}^t Q_1 S_1 Q_1 = A_1$ where $Q_1 = P_1(S_1) \in GL_{m-1}(\mathbb{R})$.

Let $S = (s_{ij})$ be a symmetric $m \times m$ symmetric matrix near enough to $A = \text{diag}(a_1, a_2, \dots, a_m)$ so that s_{11} is non-zero and has the same sign as a_1 . The symmetric matrix S determines a symmetric bilinear form $B : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ given by $B(x, y) = {}^t x S y$ for all $x, y \in \mathbb{R}^m$. Following the first step in the Gram-Schmidt orthogonalization process, we change the standard basis e_1, \dots, e_m of \mathbb{R}^m to a basis w_1, \dots, w_m where

$$w_1 = \frac{e_1}{\sqrt{|s_{11}|}}$$

and

$$w_j = e_j - B(w_1, w_1)B(w_1, e_j) w_1 = e_j - \frac{s_{1j}}{s_{11}} e_1$$

for all $j = 2, \dots, m$. The corresponding change of basis matrix $C \in GL_m(\mathbb{R})$ is given by

$$C = \begin{pmatrix} \frac{1}{\sqrt{|s_{11}|}} & -\frac{s_{12}}{s_{11}} & \dots & -\frac{s_{1m}}{s_{11}} \\ 0 & & & \\ \vdots & & I & \\ 0 & & & \end{pmatrix}$$

where I denotes the $(m-1) \times (m-1)$ identity matrix. The new basis satisfies $B(w_1, w_j) = 0$ for all $j = 2, \dots, m$, and thus it is easy to see that

$${}^t C S C = \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & S_1 & \\ 0 & & & \end{pmatrix}$$

where S_1 is an $(m-1) \times (m-1)$ symmetric matrix depending smoothly on S . If S is sufficiently close to A , then $S_1 \in U_1$, and we can apply the induction hypothesis to conclude that there exists some $Q_1 \in GL_{m-1}(\mathbb{R})$ depending smoothly on S_1 such that ${}^t Q_1 S_1 Q_1 = A_1 = \text{diag}(a_2, \dots, a_m)$. Define $P(S) = Q = CR$ where

$$R = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & Q_1 & \\ 0 & & & \end{pmatrix}$$

Then we have ${}^t Q S Q = A$ where $P(S) = Q \in GL_m(\mathbb{R})$ depends smoothly on S and $P(A) = I_{m \times m}$.

□

Let $C \subseteq M$ be an n -dimensional connected critical submanifold of f . By replacing f with $f - c$, where c is the common value of f on the critical submanifold C , we may assume that $f(p) = 0$ for all $p \in C$. Let $p \in C$, and choose a coordinate chart $\phi : U \rightarrow \mathbb{R}^n \times \mathbb{R}^{m-n}$ defined on an open neighborhood U of p such that $\phi(p) = (0, 0)$ and $\phi(U \cap C) = \mathbb{R}^n \times \{0\}$. By composing this chart with a diffeomorphism of $\mathbb{R}^n \times \mathbb{R}^{m-n}$ that is constant in the first component, we may assume that

the matrix of the Hessian in the direction normal to $\mathbb{R}^n \times \{0\}$ at $(0, 0) \in \mathbb{R}^n \times \mathbb{R}^{m-n}$ for the local expression $h(x, y) = (f \circ \phi^{-1})(x, y)$:

$$M_0^\nu(f) = \left(\frac{\partial^2 h}{\partial y_i \partial y_j} \Big|_{(0,0)} \right)$$

is a diagonal matrix with the first k diagonal entries equal to -1 and the rest equal to $+1$.

The assumption that f is Morse-Bott means that for every $x \in \mathbb{R}^n$ the quadratic form

$$q_x(y) = {}^t y \left(\frac{\partial^2 h}{\partial y_i \partial y_j} \Big|_{(x,0)} \right) y$$

is non-degenerate. If we fix $x \in \mathbb{R}^n$ and apply Palais' construction in the proof of the Morse Lemma to the quadratic form $q_x : \mathbb{R}^{m-n} \rightarrow \mathbb{R}$, then we get a family ψ_x of diffeomorphisms depending smoothly on x between neighborhoods of $0 \in \mathbb{R}^{m-n}$ such that

$$h(x, \psi_x(y)) = q_x(y).$$

Therefore, $\tilde{\phi}^{-1}(x, y) = \phi^{-1}(x, \psi_x(y)) : \mathbb{R}^n \times \mathbb{R}^{m-n} \rightarrow M$ is a chart such that

$$(f \circ \tilde{\phi}^{-1})(x, y) = q_x(y).$$

It's clear that $M_x^\nu(f)$ depends smoothly on x , and hence by Hirsch's Lemma, for every $x \in \mathbb{R}^n$ sufficiently close to 0 there exists a matrix $Q_x \in GL_{m-n}(\mathbb{R})$ depending smoothly on x such that

$${}^t Q_x \left(\frac{\partial^2 h}{\partial y_i \partial y_j} \Big|_{(x,0)} \right) Q_x = \left(\frac{\partial^2 h}{\partial y_i \partial y_j} \Big|_{(0,0)} \right).$$

Therefore, for all $x \in \mathbb{R}^n$ sufficiently close to 0 we have

$$\begin{aligned} (f \circ \tilde{\phi}^{-1})(x, Q_x y) &= q_x(Q_x y) \\ &= {}^t y {}^t Q_x \left(\frac{\partial^2 h}{\partial y_i \partial y_j} \Big|_{(x,0)} \right) Q_x y \\ &= {}^t y \left(\frac{\partial^2 h}{\partial y_i \partial y_j} \Big|_{(0,0)} \right) y \\ &= \sum_{j=1}^{m-n} \delta_j y_j^2 \end{aligned}$$

where $\delta_j = \frac{\partial^2 h}{\partial y_j^2}((0, 0)) = -1$ for all $j = 1, \dots, k$ and $\delta_j = \frac{\partial^2 h}{\partial y_j^2}((0, 0)) = +1$ for all $j = k + 1, \dots, m$. □

REFERENCES

- [1] M. Audin, *The topology of torus actions on symplectic manifolds*, Progress in Math. Vol. 93, Birkhauser, Basel, 1991.
- [2] R. Bott, *Nondegenerate critical manifolds*, Ann. Math. (2) 60 (1954) 248-261.
- [3] R. Bott, *Lectures on Morse theory, old and new*, Bull. Amer. Math. Soc (N. S.) 7 (1982) 331-358.
- [4] V. Guillemin and S. Sternberg, *Geometric asymptotics*, Amer. Math. Soc. Providence, R. I. (1977).
- [5] M. W. Hirsch, *Differential Topology*, Springer-Verlag, New York, 1994.
- [6] J. Milnor, *Morse theory*, Princeton University Press, N.J. 1963.
- [7] M. Morse, *Relations between the critical points of a real function of n independent variables*, Trans. Amer. Math. Soc. 27 (1925), 345-396.
- [8] J. Moser, *On the volume elements of a manifold*, Trans. Amer. Math. Soc. 120(1965) 286-294.
- [9] R. Palais, *The Morse lemma for Banach spaces*, Bull. Amer. Math. Soc. 75 (1969), 968-971.

Augustin Banyaga
Department of Mathematics
The Pennsylvania State University
University Park, PA 16802
banyaga@math.psu.edu

David E. Hurtubise
Department of Mathematics and Statistics
Penn State Altoona
Altoona, PA 16601-3760
hurtubis@math.psu.edu