COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 171–206 (1956)

On the Iteration of Closed Geodesics and the Sturm Intersection Theory*

RAOUL BOTT

The University of Michigan and the Institute for Advanced Study

Introduction

Although the original motivation of this paper is to be found in the Morse theory of closed geodesics, the results presented definitely belong into the realm of self-adjoint systems of ordinary differential equations. I have, therefore, written the main body of this paper purely in the context of this latter discipline, but would like to start here with a short account of the original problem as encountered in the Morse theory.

A closed geodesic g on a Riemannian manifold M is a map $g : R \to M$ of the reals into M, which satisfies the usual differential equations of a geodesic for all $x \in R$, and is periodic. The n-th iterate g^n of g is then defined in terms of g as the map

$$g^n(x) = g(nx), \qquad (n = \pm 1, \pm 2 \cdots), \ x \in R.$$

Thus g^1 represents the same geodesic as g; g^{-1} represents the oppositely oriented geodesic to g.

In [6] Morse assigns to each closed geodesic g two non-negative integers $\lambda(g)$ and $\nu(g)$; the index and nullity of the closed geodesic g respectively. ($\lambda(g)$ represents, roughly, the number of negative characteristic roots of the boundary value problem, associated to g, by Morse [6; p. 289]. $\nu(g)$, on the other hand, represents the multiplicity of the eigenvalue 0 of that same problem. In Section 1 these notions are precisely defined, for a general periodic operator L.) Our problem is to describe the nature of the sequences $\{\lambda(g^n); \nu(g^n)\} \quad (n = 1, 2, \cdots)$.

The sequences are of interest because of the following result due to Morse:

On a compact Riemannian manifold M the number of closed geodesics with prescribed index k is greater than or equal to a topological invariant

*The work on this paper was partially supported by a H. H. Rackham School Summer Faculty Research fellowship.

$R^k(M)$ of M (the k-th circular connectivity of M), provided the nullity of all closed geodesics on M is zero.

This theorem then furnishes a counting of the closed geodesics on M into which the nullity and index enter vitally. Since in this counting g^n, $n > 1$, counts as a distinct geodesic from g, the study of $\lambda(g^n)$, $\nu(g^n)$ is of some interest for purely computational reasons. On the other hand, if we consider g^n in some sense equivalent to g, which is usual when we speak of closed geodesics intuitively, the redundancy of the counting has to be adjusted, and this can only be done by setting bounds to the complexity of the sequence $\{\lambda(g^n)\}$.

The results of Section 1, when translated into this context can be summarized in the following theorems:

THEOREM A. *Every closed geodesic g determines non-negative integer-valued functions Λ_g and N_g on the unit circle, $|z| = 1$, such that*

$$\lambda(g^n) = \Sigma\, \Lambda_g(\omega),$$
$$\nu(g^n) = \Sigma\, N_g(\omega)$$

where ω ranges over the n-th roots of $+1$ or -1, depending on whether g is orientable or not.

Remark: g is orientable if the orientation of a coordinate system on M is not changed by parallel translation along a fundamental period of g. Otherwise g is non-orientable.

THEOREM B. *The function N_g $[\Lambda_g]$ of Theorem A is in turn completely determined (determined up to an additive constant) by a certain $2n \times 2n$ matrix P_g, the Poincaré matrix of g. Here $n + 1 = dim.\ M$.*

The manner in which P_g determines Λ_g is unfortunately rather complicated and will not be discussed in detail here (see Theorems III and IV, Section 1). We will only list a few of the properties of these functions which follow from these theorems.

THEOREM C. *The functions Λ_g* and *N_g have, among others, the following properties*:

1. $N_g(z) =$ *dimension of null-space of* $\{P_g - zI\}$ $(|z| = 1)$.
2. $\Lambda_g(z)$ *is constant at points at which* $N_g(z) = 0$, *and the jump of* Λ_g *at* z *is always bounded in absolute value by* $N_g(z)$.
3. $\Lambda_g(z^*) = \Lambda_g(z)$; $N_g(z^*) = N_g(z)$
   (z^* *denotes complex conjugate of* z.)

A real number ϱ, such that $N_g(e^{2\pi i\varrho}) \neq 0$ is called an imaginary (Poincaré) exponent of g. By part 1 of Theorem C, g has at most $2n$ distinct imaginary exponents mod 1. The following corollary to Theorems A and C

is now an immediate consequence of the definition of a Riemann integral.

COROLLARY I. $\lambda^*(g) = \lim\limits_{n\to\infty} \dfrac{\lambda(g^n)}{n}$ *is given by the formula*

$$\lambda^*(g) = \frac{1}{2\pi}\int_0^{2\pi} \Lambda_g(e^{i\theta})\, d\theta = a_0 + \Sigma a_i \varrho_i$$

where the a_i are integers, and $\{\varrho_i\}$ is a complete system of imaginary Poincaré exponents of g mod 1.

$$\nu^*(g) = \lim_{n\to\infty} \frac{\nu(g^n)}{n} = \frac{1}{2\pi}\int_0^{2\pi} N_g(e^{i\theta})\, d\theta = 0.$$

If the imaginary exponents of g are all irrational, then the formula of Theorem A can be given more explicitly, as can be easily verified and is expressed in

COROLLARY II. *If the imaginary exponents $\{\varrho_i\}$ of g are all irrational, then*

$$\lambda(g^n) = a_{-1} + a_0 n + \Sigma\, a_i[n\varrho_i]$$

where ϱ_i is again a complete system of imaginary exponents, and the a_{-1}, a_i $(i = 0, \cdots)$ are integers, while $[x]$ denotes the greatest integer $\leqq x$.

In this case $\nu(g^n)$ is clearly 0 for all n.

These results augment, partly overlap, and partly generalize results due to Hedlund [4], and announced results of Morse and Pitcher [7]. Hedlund restricts his discussion to dim. $M = 2$, and the "nondegenerate case" i.e., $\nu(g^n) = 0$ for all n. Our formula does not add to his rather complete discussion of this case, except for the following remark.

It follows at once from Corollary I, that since dim. $M = 2$,

$$\lambda^*(g) = a_0 + a_1\varrho + a_2(-\varrho) = a_0 + (a_1 - a_2)\varrho,$$

or

$$\lambda^*(g) = a_0$$

depending on whether g has the imaginary exponents ϱ and $-\varrho$, or has none at all. Hence if $\lambda^*(g)$ is rational, but not an integer, then g must have an imaginary exponent ϱ, which is also rational and not an integer, and $\nu(g^n)$ must be different from zero for some values of n. Hence the conditions of theorem IV, case 1 in [4] are incompatible.

The Proceedings note of Morse and Pitcher, treating the n-dimensional nondegenerate case, announces among other results the existence of the limit $\lambda^*(g)$ and that $\lambda(g^n)$ is determined by P_g and $\lambda(g)$. The formula of Corollary I is foreshadowed there by the proposition that if all the imaginary exponents of g are rational then $\lambda^*(g)$ is also.

That these authors restricted themselves exclusively to the non-degenerate case, $\nu(g^n) = 0$, is explained by the fact that in the classical Morse theory $\lambda(g)$ has "local topological significance" only if $\nu(g) = 0$. However, if g is contained in a compact manifold of closed geodesics, say σ, and if dim. $\sigma = \nu(g)$ for all $g \in \sigma$, then $\lambda(g)$ is again topologically significant, as is shown in [2]. I was therefore interested in freeing the discussion from the hypothesis of nondegeneracy.

Theorem A is essentially proved (modulo the translation into differential geometry) in Theorem I of Section 1. The main idea of this simple proof is the transition from the given problem which can be, and usually is, treated over the real numbers, to a Hermitian problem. Theorem B on the other hand I have found quite difficult, and it is to the proof of this theorem that the bulk of this paper is devoted. To solve it we introduce, what I will call, the "Sturm intersection theory". Briefly stated, its main idea is the following: From the problem $Ly = \lambda y$, where L is a self-adjoint differential operator of the second order, we go over to the standardly associated first order matrix equation

$$\frac{dX_\lambda}{dt} = A_\lambda X_\lambda .$$

The fundamental solution of this equation, normalized by the condition $X_\lambda(0) = I$ (all real λ), then defines a map X^L of the t, λ-plane into a certain subgroup $\mathfrak{H}$ of the full linear group. On the other hand, a self-adjoint boundary condition B is seen to define a cycle γ_B in $\mathfrak{H}$. It is then shown that the spectrum of $Ly = \lambda y$ subject to B at $t = 0$ and $t = a > 0$, corresponds precisely to the intersection (in a topological sense) of the curve $\lambda \to X^L(a, \lambda)$ with γ_B. Once this is established, standard deformation arguments yield Theorem B.

This "Sturm intersection theory" is, I hope, of some independent interest, and I have therefore developed it here for general self-adjoint (or Hermitian) boundary-conditions, rather than for the special case needed in Theorem B. Having done this I also indicate, in Sections 8 and 9, how it can be used to prove (a) the general existence theorem of an infinite but discrete spectrum for a regular problem, (b) the continuous dependence of the spectrum on the boundary conditions, (c) a generalization of the oscillation theorem of Sturm and the focal point theorem of Morse. All these theorems seem to be consequences of the fact that the intersection of two cycles, of which one is homologous to zero, is again homologous to zero.

The proof that the topological intersections of the curve: $\lambda \to X_\lambda(a)$, with γ_B correspond to the spectrum of $Ly = \lambda y$ is complicated by the fact that some points of the spectrum may have a multiplicity greater than one.

This phenomenon reflects itself in the carrier B^0 of γ_B. In general, this set is not a submanifold of $\mathfrak{H}$. To construct γ_B, we are therefore led to introduce the "resolution of B^0" in Section 4. This is a submanifold $B^{(1)}$ in the cartesian product of $\mathfrak{H}$ with a complex projective space $G^{(1)}$. Under the canonical projection $f^{(1)}$: $\mathfrak{H} \times G^{(1)} \to \mathfrak{H}$, $B^{(1)}$ then maps onto B^0, and it is the image of the fundamental class of $B^{(1)}$ under $f^{(1)}$ which is defined as γ_B.

1. Periodic Hermitian Systems

Let E_1 be the unitary space of complex n-tuples $x = \{x_1, \cdots, x_n\}$, equipped with the Hermitian inner product

$$(x, y) = \sum_1^n x_i y_i^*.$$

Here the star of a complex number denotes its complex conjugate. In general, if p: $E_1 \to E_1$ is a linear transformation p^* shall denote its adjoint. Thus

$$(px, y) = (x, p^*y) \text{ all } x, y \in E_1 .$$

In any unitary cordinate system on E_1, the matrix of p^* is therefore the complex conjugate transpose of the matrix of p. Throughout we will denote by $\nu(p)$ the dimension of the null space of a linear transformation p. S^1 shall stand for the circle $|z| = 1$, and R shall denote the real line.

Set $E = E_1 \oplus E_1$. The inner product in E is derived from the one in E_1 by the formula

$$(u, v) = (x_1, x_2) + (y_1, y_2)$$

if $u = \{x_1, y_1\}$, $v = \{x_2, y_2\}$, $x_i, y_i \in E_1$ $(i = 1, 2)$.

DEFINITION 1.1: *A Hermitian periodic operator L shall be defined as a second order differential operator on the vector functions* $t \to y(t) \in E_{(1)}$ $(t \in R)$ *of class* C'', *given by*

$$Ly \equiv -\{py' + qy\}' + q^*y' + ry. \tag{1.1}$$

Here, (as throughout the paper) the prime denotes differentiation with respect to t and $p(t)$, $q(t)$ *and* $r(t)$ *are continuously differentiable matrix functions defined on R, which satisfy the following conditions*:

(*a*) $p(t)$, $q(t)$ *and* $r(t)$ *are of period* 1,

(*b*) $p^*(t) = p(t)$; $r^*(t) = r(t)$ (*all t*),

(*c*) $(p(t)u, u) > 0$ *if* $u \neq 0$ (all *t*) (*i.e., p is positive definite — also denoted by* $p > 0$). *We reserve the symbol L for operators of the above type.*

Let

$$J\{x, y\} = \{-y, x\}, \qquad x, y \in E_1.$$

If y is a vector function from R to E_1 of class C'' and L is given, u_y shall always stand for the map of R into E defined in terms of the coefficients of L by

$$u_y(t) = \{y(t),\ p(t)y'(t) + q(t)y(t)\}. \tag{1.2}$$

The Hermitian character of L is then brought out by the following well-known identity, valid if $y(t)$, $x(t)$ are of class C'':

$$(Ly, x) - (y, Lx) = (Ju_y, u_x)'. \tag{1.3}$$

We will study the eigenvalue problem

$$Ly = \lambda y \tag{1.4}$$

subject to a family of boundary conditions $[n, z]$ ($n = \pm 1, \pm 2$ etc; $z \in S^1$.) Here a function $y(t) \in E_1$ of class C'' satisfies $[n, z]$ if and only if

$$u_y(t + n) = zu_y(t) \qquad \text{all } t \in R.$$

With (1.4) subject to any boundary condition B we associate the spectral multiplicity function Θ_B^L. Θ_B^L is defined on the complex λ-plane by the condition

$$\Theta_B^L(\lambda) = \text{number of linearly independent solutions of } Ly = \lambda y \text{ subject to } B. \tag{1.5}$$

The points at which $\Theta_B^L(\lambda) \neq 0$ are called the spectrum of (1.4) subject to B.

Since each condition B of the family $\{[n, z]\}$ described above is self-adjoint, or Hermitian as we will call it, (1.4) subject to $B \in \{[n, z]\}$ has the following well-known properties.

PROPOSITION 1.1. *If $B \in \{[n, z]\}$, then $\Theta_B^L(\lambda) = 0$ if λ is not real. On the real axis Θ_B^L is a finite valued function which is different from zero only on an infinite discrete subset of R which is bounded from below.*

The following continuity theorem also holds. It can be found in the literature (see for example, [6]; p. 91) but will also follow from subsequent arguments in this paper (see Section 7).

If τ is an interval on R which is bounded from above, we set

$$[\Theta_B^L : \tau] = \sum_{\lambda \in \tau} \Theta_B^L(\lambda). \tag{1.6}$$

By proposition (1.1) this number is always finite. If the end-points of τ do not lie on the spectrum of (1.4) subject to B, τ shall be called admissible with respect to Θ_B^L.

The continuity theorem which we are after can now be stated as follows:

PROPOSITION 1.2. *If τ is an admissible interval with respect to $\Theta_{[n,z]}^L$, then there exists a neighborhood U of z on S^1 such that*

(1) *τ is admissible with respect to $\Theta^L_{[n,z]}$ for all $z \in U$,*

(2) $[\Theta^L_{[n,z]}: \tau] = [\Theta^L_{[n,w]}: \tau]$, $\qquad z, w \in U$.

Having put these preliminaries down we can proceed to our first problem, which is to find the relations between the functions $\Theta^L_{[n,z]}$ when L is kept fixed and n, z are allowed to range over their respective domains. To simplify the notation we therefore drop the superscript L, and write Θ_z for $\Theta_{[1,z]}$. We then have the following Fourier Theorem, which is the basis of Theorem A of the introduction.

THEOREM I. *The spectral multiplicity functions $\Theta_{[n,z]}$ of (1.4) subject to $[n, z]$ are given by $\Theta_z = \Theta_{[1,z]}$, according to the formula*

$$\Theta[n, z] = \sum_{\omega} \Theta_\omega \tag{1.7}$$

where ω ranges over the n-th roots of z.

Proof: Every $y(t)$ subject to $[n, z]$ admits the unique Fourier expansion

$$y(t) = \sum \omega y_\omega(t)$$

where ω ranges over the n-th roots of z, and $y_\omega(t) = \frac{1}{n}\sum_0^{n-1} \omega^{-s} y(t+s-1)$. Now

$$y_\omega(t+1) = \frac{1}{n}\sum_0^{n-1} \omega^{-s} y(t+s). \tag{1.8}$$

If we set $k = s+1$, the right hand side of (1.8) becomes

$$\frac{\omega}{n}\left\{\sum_1^{n-1} \omega^{-k} y(t+k-1) + \omega^{-n} y(t+n-1)\right\},$$

$$\frac{\omega}{n}\left\{\sum_1^{n-1} \omega^{-k} y(t+k-1) + y(t-1)\right\}$$

since ω is an n-th root of z and $y(t+n) = zy(t)$. Hence $y_\omega(t+1) = \omega y_\omega(t)$; i.e., $y_\omega(t)$ satisfies $[1, \omega]$. Since L commutes with translations of length 1, Ly_ω again satisfies $[1, \omega]$. Hence y is a solution of (1.4) subject to $[n, z]$ if and only if each component $y_\omega(t)$ is subject to $[1, \omega]$. The theorem now follows.

Following Morse, the index and nullity (denoted by Λ^L_B and N^L_B, respectively) of the problem (1.4), subject to a Hermitian boundary condition B, is defined by

$$\Lambda^L_B = [\Theta^L_B: R^-], \qquad R^- = \{\lambda \mid \lambda < 0\}$$
$$N^L_B = \Theta^L_B(0).$$

Clearly Theorem I has the following

COROLLARY. *The index and nullity of $Ly = \lambda y$ subject to $[n, z]$ are given by*

$$\Lambda^L_{[n,z]} = \sum_\omega \Lambda^L_{[1,\omega]},$$

$$N^L_{[n,z]} = \sum_\omega N^L_{[1,\omega]}$$

where ω ranges over the n-th roots of z as before.

The sequence $\{\lambda(g^n)\}$ of the introduction is now seen easily to be identical with the sequence $\Lambda^L_{[n,1]}$ or $\Lambda^L_{[n,-1]}$ of the Hermitian operator L associated with g by Morse via the second variation. (The two cases correspond to the orientability of g.) We embark therefore on a detailed study of $\Lambda(z) = \Lambda^L_{[1,z]}$, $N(z) = N^L_{[1,z]}$ as functions on S^1.

PROPOSITION 1.3. *The functions $\Lambda(z)$, $N(z)$ defined above have the following properties:*

(1) *Both functions are non-negative integer valued functions on S^1.*

(2) *The inequality*

$$\lim_{z\to z_0^\pm} \Lambda(z) \geqq \Lambda(z_0) \tag{1.9}$$

holds.

(3) *$N(z) = 0$ except at at most $2n$ points $\{z_1, \cdots, z_r\} \in S^1$.*

(4) *The equality holds in (1.9) if z_0 is not one of these points $\{z_i\}$ (referred to as the Poincaré points of L). Furthermore, Λ is constant in the vicinity of points other than the Poincaré points and the jump of Λ at any point $z \in S^1$ is bounded in absolute value by $N(z)$.*

(5) *The sum $\sum N\{z_i\}$ extended over the Poincaré points of L does not exceed $2n$.*

(6) *If L is real, then $\Lambda(z) = \Lambda(z^*)$, $N(z) = N(z^*)$.*

Proof: (1) and (6) are clear from the definition of the functions. (2) is a consequence of the continuity theorem — Proposition 1.2. For, if 0 is not in the spectrum of (1.4) subject to $[1, z_0]$, then, by that proposition, Λ is constant in some vicinity of z_0. If 0 is in that spectrum, we can, because of the discreteness of the spectrum, define $\Lambda(z_0)$ by

$$\Lambda(z_0) = [\Theta^L_{[1,z_0]}: R^-_\varepsilon],\ R^-_\varepsilon = \{\lambda \mid \lambda \leqq -\varepsilon\},$$

for sufficiently small $\varepsilon > 0$. Let $\eta = \{\lambda \mid -\varepsilon \leqq \lambda < 0\}$. Then, since R^-_ε is admissible with respect to $\Theta^L_{[1,z_0]}$,

$$\Lambda(z) = \Lambda(z_0) + [\Theta^L_{[1,z]}: \eta]$$

in some vicinity of z_0. Since the last term on the right is only capable of positive values, property (2) follows. This argument clearly establishes (4) as well, since $N(z_0)$ can be defined by

$$N(z_0) = [\Theta^H_{[1,z_0]}: \delta] \text{ with } \delta = \{\lambda \mid -\varepsilon \leqq \lambda \leqq \varepsilon\}$$

for sufficiently small ε, and hence

$$N(z_0) \geqq [\Theta^L_{[1,z]}: \eta].$$

To establish (3), we make the usual transformation of the second order problem (1.4) over E_1 into a first order problem over E. As is easily checked, the equation

$$Ly = \lambda y$$

is quite equivalent to

$$u'_v(t) = A_\lambda(t) u_v(t) \tag{1.10}$$

where $A_\lambda(t)$ is the $2n \times 2n$ matrix constructed from the coefficients of L in the following manner:

$$A_\lambda(t) = \begin{cases} -p^{-1}(t)q(t), & p^{-1}(t) \\ r(t) - q^*(t)p(t)q(t) - \lambda, & q^*(t)p^{-1}(t). \end{cases} \tag{1.11}$$

From this point of view $\Lambda(z)$ and $N(z)$ are reinterpreted as:

$\Lambda(z)$ $[N(z)]$ $=$ number of linearly independent solutions of $u'(t) = A_\lambda(t)u(t)$ with $\lambda < 0$ $[\lambda = 0]$, subject to $u(t+1) = zu(t)$.

Let $X_\lambda(t)$ be the fundamental matrix solution of (1.10), normalized by the condition

$$X_\lambda(0) = I \text{ (identity on } E) \text{ for all } \lambda; \tag{1.12}$$

then every solution of (1.10) is of the form

$$u(t) = X_\lambda(t)v$$

where v is a constant vector of E. Hence, recalling our convention that $\nu\{\ \}$ denotes the complex dimension of the null space of the matrix $\{\ \}$, we obtain the formula

$$\Theta^L_{[1,z]} = \nu\{X_\lambda(1) - zI\} \tag{1.13}$$

and in particular

$$N(z) = \nu\{X_0(1) - zI\}. \tag{1.14}$$

However, this interpretation of $N(z)$ clearly establishes properties (3) and (5) of Proposition 1.2.

The matrix $X_0(1)$ was originally introduced by Poincaré in his study of periodic systems. We denote it by P^L or just P in this paper and will refer to it as the Poincaré matrix of L. Thus the Poincaré points of L are precisely those characteristic roots of P which are of absolute value 1.

By (1.14) P completely determines $N(z)$. In view of Proposition 1.2, it therefore seriously restricts the behavior of Λ on S^1. Actually, however, the following stronger proposition holds: