Universitext

Liviu Nicolaescu

An Invitation to Morse Theory

Springer

Liviu I. Nicolaescu

An Invitation to Morse Theory

Springer

Liviu Nicolaescu
Mathematics Department
University of Notre Dame
Notre Dame, IN 46556-4618
USA
nicolaescu.1@nd.edu

Mathematics Subject Classification (2000): MSC2000, 57R70, 57R80, 57R17, 57R91, 57T15, 58K05, 53D20, 14D05

Library of Congress Control Number: 2006938271

ISBN-10: 0-387-49509-6 e-ISBN-10: 0-387-49510-X
ISBN-13: 978-0-387-49509-5 e-ISBN-13: 978-0-387-49510-1

Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

springer.com

Universitext

Universitext

Editors (North America): S. Axler and K.A. Ribet

Aguilar/Gitler/Prieto: Algebraic Topology from a Homotopical Viewpoint
Aksoy/Khamsi: Nonstandard Methods in Fixed Point Theory
Andersson: Topics in Complex Analysis
Aupetit: A Primer on Spectral Theory
Bachman/Narici/Beckenstein: Fourier and Wavelet Analysis
Badescu: Algebraic Surfaces
Balakrishnan/Ranganathan: A Textbook of Graph Theory
Balser: Formal Power Series and Linear Systems of Meromorphic Ordinary Differential Equations
Bapat: Linear Algebra and Linear Models (2nd ed.)
Berberian: Fundamentals of Real Analysis
Blyth: Lattices and Ordered Algebraic Structures
Boltyanskii/Efremovich: Intuitive Combinatorial Topology (Shenitzer, trans.)
Booss/Bleecker: Topology and Analysis
Borkar: Probability Theory: An Advanced Course
Böttcher/Silbermann: Introduction to Large Truncated Toeplitz Matrices
Carleson/Gamelin: Complex Dynamics
Cecil: Lie Sphere Geometry: With Applications to Submanifolds
Chae: Lebesgue Integration (2nd ed.)
Charlap: Bieberbach Groups and Flat Manifolds
Chern: Complex Manifolds Without Potential Theory
Cohn: A Classical Invitation to Algebraic Numbers and Class Fields
Curtis: Abstract Linear Algebra
Curtis: Matrix Groups
Debarre: Higher-Dimensional Algebraic Geometry
Deitmar: A First Course in Harmonic Analysis (2nd ed.)
DiBenedetto: Degenerate Parabolic Equations
Dimca: Singularities and Topology of Hypersurfaces
Edwards: A Formal Background to Mathematics I a/b
Edwards: A Formal Background to Mathematics II a/b
Engel/Nagel: A Short Course on Operator Semigroups
Farenick: Algebras of Linear Transformations
Foulds: Graph Theory Applications
Friedman: Algebraic Surfaces and Holomorphic Vector Bundles
Fuhrmann: A Polynomial Approach to Linear Algebra
Gardiner: A First Course in Group Theory
Gårding/Tambour: Algebra for Computer Science
Goldblatt: Orthogonality and Spacetime Geometry
Gustafson/Rao: Numerical Range: The Field of Values of Linear Operators and Matrices
Hahn: Quadratic Algebras, Clifford Algebras, and Arithmetic Witt Groups
Heinonen: Lectures on Analysis on Metric Spaces
Holmgren: A First Course in Discrete Dynamical Systems
Howe/Tan: Non-Abelian Harmonic Analysis: Applications of $SL(2, R)$
Howes: Modern Analysis and Topology
Hsieh/Sibuya: Basic Theory of Ordinary Differential Equations
Humi/Miller: Second Course in Ordinary Differential Equations
Hurwitz/Kritikos: Lectures on Number Theory

(continued after index)

To my mother, with deepest gratitude

Preface

As the the title suggests, the goal of this book is to give the reader a taste of the "unreasonable effectiveness" of Morse theory. The main idea behind this technique can be easily visualized.

Suppose M is a smooth, compact manifold, which for simplicity we assume is embedded in a Euclidean space E. We would like to understand basic topological invariants of M such as its homology, and we attempt a "slicing" technique.

We fix a unit vector $\mathbf{u}$ in E and we start slicing M with the family of hyperplanes perpendicular to $\mathbf{u}$. Such a hyperplane will in general intersect M along a submanifold (slice). The manifold can be recovered by continuously stacking the slices on top of each other in the same order as they were cut out of M.

Think of the collection of slices as a deck of cards of various shapes. If we let these slices continuously pile up in the order they were produced, we notice an increasing stack of slices. As this stack grows, we observe that there are moments of time when its shape suffers a qualitative change. Morse theory is about extracting quantifiable information by studying the evolution of the shape of this growing stack of slices.

From a mathematical point of view we have a smooth function

$$h : M \to \mathbb{R}, \quad h(x) = \langle \mathbf{u}, x \rangle.$$

The above slices are the level sets of h,

$$\{ x \in M;\ \ h(x) = \text{const} \},$$

and the growing stack is the time dependent sublevel set

$$\{x \in M;\ \ h(x) \leq t\}, \quad t \in \mathbb{R}.$$

The moments of time when the pile changes its shape are called the *critical values* of h and correspond to moments of time t when the corresponding hyperplane $\{\langle \mathbf{u}, x\rangle = t\}$ intersects M tangentially. Morse theory explains how to describe the shape change in terms of *local* invariants of h.

A related slicing technique was employed in the study of the topology of algebraic manifolds called the *Picard–Lefschetz theory*. This theory is back in fashion due mainly to Donaldson's pioneering work on symplectic Lefschetz pencils.

The present book is divided into three conceptually distinct parts. In the first part we lay the foundations of Morse theory (over the reals). The second part consists of applications of Morse theory over the reals, while the last part describes the basics and some applications of complex Morse theory, a.k.a. Picard–Lefschetz theory. Here is a more detailed presentation of the contents.

In chapter 1 we introduce the basic notions of the theory and we describe the main properties of Morse functions: their rigid local structure (Morse lemma) and their abundance (Morse functions are generic). To aid the reader we have sprinkled the presentation with many examples and figures. One recurring simple example we use as a testing ground is that of a natural Morse function arising in the design of robot arms.

Chapter 2 is the technical core of the book. Here we prove the fundamental facts of Morse theory: crossing a critical level corresponds to attaching a handle and Morse inequalities. Inescapably, our approach was greatly influenced by classical sources on this subject, more precisely Milnor's beautiful books on Morse theory and h-cobordism [M3, M4].

The operation of handle addition is much more subtle than it first appears, and since it is *the* fundamental device for manifold (re)construction, we devoted an entire section to this operation and its relationship to cobordism and surgery. In particular, we discuss in some detail the topological effects of the operation of surgery on knots in S^3 and illustrate this in the case of the trefoil knot.

In chapter 2 we also discuss in some detail dynamical aspects of Morse theory. More precisely, we present the techniques of S. Smale about modifying a Morse function so that it is self-indexing and its stable/unstable manifolds intersect transversally. This allows us to give a very simple description of an isomorphism between the singular homology of a compact smooth manifold and the (finite dimensional) Morse–Floer homology determined by a Morse function, that is, the homology of a complex whose chains are formal linear combinations of critical points and whose boundary is described by the connecting trajectories of the gradient flow. We have also included a brief section on Morse–Bott theory, since it comes in handy in many concrete situations.

We conclude this chapter with a section of a slightly different flavor. Whereas Morse theory tries to extract topological information from information about critical points of a function, min-max theory tries to achieve the opposite goal, namely to transform topological knowledge into information about the critical points of a function. While on this topic, we did not want to avoid discussing the Lusternik–Schnirelmann category of a space.

Chapter 3 is devoted entirely to applications of Morse theory, and in writing it we were guided by the principle, few but juicy. We present relatively few examples, but we use them as pretexts for wandering in many parts of mathematics that are still active areas of research. More precisely, we start by presenting two classical applications to the cohomology of Grassmannians and the topology of Stein manifolds.

We use the Grassmannians as a pretext to discuss at length the Morse theory of moment maps of Hamiltonian torus actions. We prove that these moment maps are Morse–Bott functions. We then proceed to give a complete presentation of the equivariant localization theorem of Atiyah, Borel, and Bott (for S^1-actions only), and we use this theorem to prove a result of P. Conner [Co]: the sum of the Betti numbers of a compact, oriented smooth manifold is greater than the sum of the Betti numbers of the fixed point set of any smooth S^1-action. Conner's theorem implies among other things that the moment maps of Hamiltonian torus actions are *perfect* Morse–Bott function. The (complex) Grassmannians are coadjoint orbits of unitary groups, and as such they are equipped with many Hamiltonian torus actions leading to many choices of perfect Morse functions on Grassmannians.

We used the application to the topology of Stein manifolds as a pretext for the last chapter of the book on Picard–Lefschetz theory. The technique is similar. Given a complex submanifold M of a complex projective space, we start slicing it using a (complex) 1-dimensional family of projective hyperplanes. Most slices are smooth hypersurfaces of M, but a few of them are have mild singularities (nodes). Such a slicing can be encoded by a holomorphic Morse map $M \to \mathbb{CP}^1$.

There is one significant difference between the real and the complex situations. In the real case, the set of regular values is *disconnected*, while in the complex case this set is *connected*, since it is a punctured sphere. In the complex case we study not what happens as we cross a critical value, but what happens when we go once around it. This is the content of the Picard–Lefschetz theorem.

We give complete proofs of the local and global Picard–Lefschetz formulæ and we describe basic applications of these results to the topology of algebraic manifolds.

We conclude the book with a chapter containing a few exercises and solutions to (some of) them. Many of them are quite challenging and contain additional interesting information we did not include in the main body, since it have been distracting. However, we strongly recommend to the reader to try solving as many of them as possible, since this is the most efficient way of grasping the subtleties of the concepts discussed in the book. The complete solutions of these more challenging problems are contained in the last section of the book.

Penetrating the inherently eclectic subject of Morse theory requires quite a varied background. The present book is addressed to a reader familiar with the basics of algebraic topology (fundamental group, singular (co)homology, Poincaré duality, e.g., Chapters 0–3 of [Ha]) and the basics of differential geometry (vector fields and their flows, Lie and exterior derivative, integration on manifolds, basics of Lie groups and Riemannian geometry, e.g., Chapters 1–4 in [Ni]). In a very limited number of places we had to use less familiar technical facts, but we believe that the logic of the main arguments is not obscured by their presence.

Acknowledgments. This book grew out of notes I wrote for a one-semester graduate course in topology at the University of Notre Dame in the fall of 2005. I want to thank the attending students, Eduard Balreira, Daniel Cibotaru, Stacy Hoehn, Sasha Lyapina, for their comments questions and suggestions, which played an important role in smoothing out many rough patches in presentation. While working on these notes I had many enlightening conversations on Morse theory with my colleague Richard Hind. I want to thank him for calmly tolerating my frequent incursions into his office, and especially for the several of his comments and examples I have incorporated in the book.

Last, but not the least, I want thank my wife. Her support allowed me to ignore the "publish or perish" pressure of these fast times, and I could ruminate on the ideas in this book with joyous abandonment.

This work was partially supported by NSF grant DMS-0303601.

Notations and Conventions

- For every set A we denote by $\#A$ its cardinality.
- For $\mathbb{K} = \mathbb{R}, \mathbb{C}$, $r > 0$ and M a smooth manifold we denote by $\underline{\mathbb{K}}^r_M$ the trivial vector bundle $\mathbb{K}^r \times M \to M$.
- $\boldsymbol{i} := \sqrt{-1}$. **Re** denotes the real part, and **Im** denotes the imaginary part.
- For every smooth manifold M we denote by TM the tangent bundle, by T_xM the tangent space to M at $x \in M$ and by T^*_xM the cotangent space at x.
- For every smooth manifold and any smooth submanifold $S \hookrightarrow M$ we denote by T_SM the *normal bundle* of S in M defined as the quotient $T_SM := (TM)|_S/TS$. The *conormal bundle* of S in M is the bundle $T^*_SM \to S$ defined as the kernel of the restriction map $(T^*M)|_S \to T^*S$.
- $\mathrm{Vect}(M)$ denotes the space of smooth vector fields on M.
- $\Omega^p(M)$ denotes the space of smooth p-forms on M, while $\Omega^p_{cpt}(M)$ the space of compactly supported smooth p-forms.
- If $F : M \to N$ is a smooth map between smooth manifolds we will denote its differential by DF or F_*. DF_x will denote the differential of F at $x \in M$ which is a linear map $DF_x : T_xM \to T_xN$. $F^* : \Omega^p(N) \to \Omega^p(M)$ is the pullback by F.
- $\pitchfork :=$ transverse intersection.
- $\sqcup :=$ disjoint union.
- For every $X, Y \in \mathrm{Vect}(M)$ we denote by L_X the Lie derivative along X and by $[X, Y]$ the Lie bracket $[X, Y] = L_XY$. i_X or $X \lrcorner$ denotes the contraction by X.
- We will orient the manifolds with boundary using the outer-normal -first convention.
- The total space of a fiber bundle will be oriented using the fiber-first convention.
- $\underline{so}(n)$ denotes the Lie algebra of $SO(n)$, $\underline{u}(n)$ denotes the Lie algebra of $U(n)$ etc.
- $\mathrm{Diag}(c_1, \cdots, c_n)$ denotes the diagonal $n \times n$ matrix with entries $c_1, \dots, c_n$.

Contents

1

Morse Functions

In this first chapter we introduce the reader to the main characters of our story, namely the Morse functions, and we describe the properties which make them so useful. We describe their very special local structure (Morse lemma) and then we show that there are plenty of them around.

1.1 The Local Structure of Morse Functions

Suppose $F : M \to N$ is a smooth (i.e., C^∞) map between smooth manifolds. The differential of F defines for every $x \in M$ a linear map

$$DF_x : T_x M \to T_{F(x)} N.$$

Definition 1.1. *(a) The point $x \in M$ is called a* critical point of F *if*

$$\operatorname{rank} DF_x < \min(\dim M, \dim N).$$

A point $x \in M$ is called a regular point of F *if it is not a critical point. The collection of all critical points of F is called the* critical set of F *and is denoted by* $\mathbf{Cr}_F$.
(b) The point $y \in N$ is called a critical value of F *if the fiber $F^{-1}(y)$ contains a critical point of F. A point $y \in N$ is called a* regular value of F *if it is not a critical value. The collection of all critical values of F is called the* discriminant set of F *and is denoted by Δ_F.*
(c) A subset $S \subset N$ is said to be negligible *if for every smooth open embedding $\Phi : \mathbb{R}^n \to N$, $n = \dim N$, the preimage $\Phi^{-1}(S)$ has Lebesgue measure zero in* $\mathbb{R}^n$. □

Theorem 1.2 (Morse–Sard–Federer). *Suppose $F : M \to N$ is a smooth function. Then the Hausdorff dimension of the discriminant set Δ_F is at most $N-1$. In particular, the discriminant set is negligible in N. Moreover, if $F(M)$ has nonempty interior, then the set of regular values is dense in $F(M)$.* □

For a proof we refer to Federer [Fed, Theorem 3.4.3] or Milnor [M2].

Remark 1.3. (a) If M and N are real analytic manifolds and F is a proper real analytic map then we can be more precise. The discriminant set is a locally finite union of real analytic submanifolds of N of dimensions less than $\dim N$. Exercise 5.1 may perhaps explain why the set of critical values is called discriminant.

(b) The range of a smooth map $F : M \to N$ may have empty interior. For example, the range of the map $F : \mathbb{R}^3 \to \mathbb{R}^2$, $F(x, y, z) = (x, 0)$, is the x-axis of the Cartesian plane $\mathbb{R}^2$. The discriminant set of this map coincides with the range. □

Example 1.4. Suppose $f : M \to \mathbb{R}$ is a smooth function. Then $x_0 \in M$ is a critical point of f if and only if $df\,|_{x_0} = 0 \in T^*_{x_0}M$.

Suppose M is embedded in a Euclidean space E and $f : E \to \mathbb{R}$ is a smooth function. Denote by f_M the restriction of f to M. A point $x_0 \in M$ is a critical point of f_M if

$$\langle df, v\rangle = 0, \quad \forall v \in T_{x_0}M.$$

This happens if either x_0 is a critical point of f, or $df_{x_0} \neq 0$ and the tangent space to M at x_0 is contained in the tangent space at x_0 of the level set $\{f = f(x_0)\}$. If f happens to be a nonzero linear function, then all its level sets are hyperplanes perpendicular to a fixed vector $\mathbf{u}$, and $x_0 \in M$ is a critical point of f_M if and only if $\mathbf{u} \perp T_{x_0}M$, i.e., the hyperplane determined by f and passing through x_0 is tangent to M.

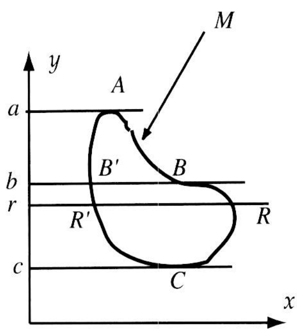

Fig. 1.1. *The height function on a smooth curve in the plane.*

In Figure 1.1 we have depicted a smooth curve $M \subset \mathbb{R}^2$. The points A, B, C are critical points of the linear function $f(x, y) = y$. The level sets of this function are horizontal lines and the critical points of its restriction to M are the points where the tangent space to the curve is horizontal. The points a, b, c on the vertical axis are critical values, while r is a regular value. □

Example 1.5 (*Robot arms: critical configurations*). We begin in this example the study of the critical points of a lsmooth function which arises in the design of robot arms. We will discuss only a special case of the problem when the motion of the arm is constrained to a plane. For slightly different presentations we refer to the papers [Hau, KM, SV], which served as our sources of inspiration. The paper [Hau] discusses the most general version of this problem, when the motion of the arm is not necessarily constrained to a plane.

Fix positive real numbers $r_1, \dots, r_n > 0$, $n \geq 2$. A (planar) *robot arm* (or *linkage*) with n segments is a continuous curve in the Euclidean plane consisting of n line segments

$$s_1 = [J_0 J_1], \quad s_2 = [J_1 J_2], \dots, \quad s_n = [J_{n-1} J_n]$$

of lengths

$$\text{dist}\,(J_i, J_{i-1}) = r_i, \quad i = 1, 2, \dots, n.$$

We will refer to the vertices J_i as the *joints* of the robot arm. We assume that J_0 is fixed at the origin of the plane, and all the segments of the arm are allowed to rotate about the joints. Additionally, we require that the last joint be constrained to slide along the positive real semiaxis (see Figure 1.2).

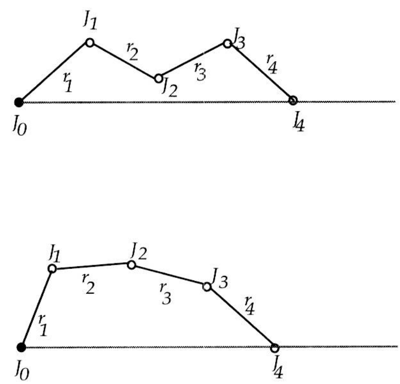

Fig. 1.2. *A robot arm with four segments.*

A (robot arm) configuration is a possible position of the robot arm subject to the above constraints. Mathematically a configuration is described by an n-uple

$$\mathbf{z} = (z_1, \dots, z_n) \in \mathbb{C}^n$$

constrained by

$$|z_k| = 1, \quad k = 1, 2, \dots, n, \quad \mathbf{Im} \sum_{k=1}^{n} r_k z_k = 0, \quad \mathbf{Re} \sum_{k=1}^{n} r_k z_k > 0.$$

Visually, if $z_k = e^{i\theta_k}$, then θ_k measures the inclination of the kth segment of the arm. The position of kth joint J_k is described by the complex number $r_1 z_1 + \cdots + r_k z_k$.

In Exercise 5.2 we ask the reader to verify that the space of configurations is a *smooth* hypersurface C of the n-dimensional manifold

$$M := \Big\{ (\theta_1, \dots, \theta_n) \in (S^1)^n;\ \sum_{k=1}^n r_k \cos\theta_k > 0 \Big\} \subset (S^1)^n,$$

described as the zero set of

$$\beta : M \to \mathbb{R},\ \ \beta(\theta_1, \dots, \theta_n) = \sum_{k=1}^n r_k \sin\theta_k = \mathbf{Im} \sum_{k=1}^n r_k z_k.$$

Consider the function $h : (S^1)^n \to \mathbb{R}$ defined by

$$h(\theta_1, \dots, \theta_n) = \sum_{k=1}^n r_k \cos\theta_k = \mathbf{Re} \sum_{k=1}^n r_k z_k.$$

Observe that for every configuration $\boldsymbol{\theta}$ the number $h(\boldsymbol{\theta})$ is the distance of the last joint from the origin. We would like to find the critical points of $h|_C$.

It is instructive to first visualize the level sets of h when $n = 2$ and $r_1 \neq r_2$, as it captures the general paradigm. For every configuration $\boldsymbol{\theta} = (\theta_1, \theta_2)$ we have

$$|r_1 - r_2| \leq h(\boldsymbol{\theta}) \leq r_1 + r_2.$$

For every $c \in (|r_1 - r_2|, r_1 + r_2)$, the level set $\{h = c\}$ consists of two configurations symmetric with respect to the x–axis. When $c = |r_1 \pm r_2|$ the level set consists of a single (critical) configuration. We deduce that the configuration space is a circle.

In general, a configuration $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in C$ is a critical point of the restriction of h to C if the differential of h at $\boldsymbol{\theta}$ is parallel to the differential at $\boldsymbol{\theta}$ of the constraint function β (which is the "normal" to this hypersurface). In other words, $\boldsymbol{\theta}$ is a critical point if and only if there exists a real scalar λ (Lagrange multiplier) such that

$$dh(\boldsymbol{\theta}) = \lambda d\beta(\boldsymbol{\theta}) \Longleftrightarrow -r_k \sin\theta_k = \lambda r_k \cos\theta_k,\ \ \forall k = 1, 2, \dots, n.$$

We discuss separately two cases.

A. $\lambda = 0$. In this case $\sin\theta_k = 0$, $\forall k$, that is, $\theta_k \in \{0, \pi\}$. If $z_k = e^{i\theta_k}$ we obtain the critical points

$$(z_1, \dots, z_n) = (\epsilon_1, \dots, \epsilon_n),\ \ \epsilon_k = \pm 1,\ \sum_{k=1}^n r_k \epsilon_k = \mathbf{Re} \sum_k r_k z k > 0.$$

B. $\lambda \neq 0$. We want to prove that this situation is impossible. We have