Cut Locus of Submanifolds: A Geometric and Topological Viewpoint

A thesis submitted in partial fulfillment of the requirements for the award of the degree of

Doctor of Philosophy

Submitted by

Sachchidanand Prasad
(17RS038)

Under the supervision of

Dr. Somnath Basu

to the

Department of Mathematics and Statistics

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, KOLKATA

Declaration

I hereby declare that this thesis is my own work and, to the best of my knowledge, it contains no materials previously published or written by any other person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at IISER Kolkata or any other educational institution, except where due acknowledgement is made in the thesis.

December 24, 2022
IISER Kolkata

Sachchidanand Prasad
(17RS038)

Certificate

This is to certify that the thesis entitled *Cut Locus of Submanifolds: A Geometric and Topological Viewpoint* is a bona fide record of work done by Sachchidanand Prasad (17RS038), a student enrolled in the PhD Programme, under my supervision during August 2017 - October 2022, submitted in partial fulfillment of the requirements for the award of PhD Degree by the Department of Mathematics and Statistics (DMS), Indian Institute of Science Education and Research (IISER) Kolkata.

Supervisor
Dr. Somnath Basu
Associate Professor
DMS, IISER Kolkata

Acknowledgements

At first, I would like to thank my parents for their support and encouragement. Thanks to them for raising me to be what I am today, believing in me, giving me freedom and space to grow as I deemed fit. A special thanks go to my mother. Without her love and affection, it would have been impossible to finish my work. I also thank my brother and sisters, who have shown their belief in me which gave me constant confidence. Their constant support has often helped me sustain myself throughout. I also take this opportunity to thank one of my teachers Satyendra Singh who has always been helpful in many ways.

At this point, I must thank the faculties at the National Institute of Technology, Rourkela, who are responsible for the academic path that has led me here. I wish to thank Dr. Debajyoti Choudhuri for his unforgettable guidance during my stay at NIT Rourkela. Thanks to Dr. Bikash Sahoo for introducing me to Prof. Swadhin Pattanayak, who deserves a special mention for sharing his philosophy of mathematics and encouragement for the research.

I must also thank the IISER Kolkata for providing a friendly environment for my academic study. I also acknowledge the staff members and security persons of IISER Kolkata for helping me in numerous ways. The library resources have been tremendously helpful during my whole stay here. I must thank the library staff for this.

I am really grateful to the Mathematics Training and Talent Search (MTTS) program as they set the right base, which was helpful for me to build up on. A special thanks to Prof. Kumaresan and Prof. Bhabha Kumar Sharma for their wonderful teaching and ideas, which helped me to handle a research problem. The methodology given by Prof. Kumaresan was always giving me the strength to understand research articles.

This whole PhD would not have been possible without financial support. I must thank and acknowledge the CSIR-UGC Government of India for this support.

It is my privilege to express heartful thanks and sincere gratitude to the faculty members of the Department of Mathematics, IISER Kolkata. A special thanks to my RPC members Dr. Sushil Gorai and Dr. Subrata Shyam Roy, for listening to me each year and giving valuable comments. I want to extend my sincere thanks to Satyaki Sir and Shirshendu Sir for their guidance and for being a friend, which made me feel more comfortable. I am also grateful to the non-teaching staff of my department, Adrish Da and Rajesh Da, for their administrative and many other bits of help.

I would like to thank the thesis reviewers, Dr. Ritwik Mukherjee from NISER Bhubaneswar and Prof. Dr. Janko Latschev from University of Hamburg for their valuable suggestions. They have pointed out corrections and these have been invaluable in improving the mathematical exposition of the thesis.

I cannot begin to express my thanks to my friends, who played an important role during my stay. Prahllad Da's discussions were beneficial in the early days of my PhD. His experience and knowledge helped me a lot to understand basic tools in algebraic and differential topology. Many thanks to Sandip and Ramyak for the fruitful discussions. When I got stuck on some proofs, conversations with them often solved many problems. I gladly thank Golam, Subrata, Ashish Da and Mrinmoy Da for their availability for any help. Many thanks to Manish, Sanjoy, Jiten, Samiron, Avishek, Sugata. I thank Mukil, Saikat, Sandip, and Anant for their helpful discussions and make valuable comments after reading the thesis. I thank Saikat Panja from IISER Pune for many suggestions and daily discussions. Finally, I would also like to extend my deepest gratitude to Gaurav, Raksha and Sonu for their friendship. This journey would have been tough without you guys.

I thank Anada Marg and my father for teaching me meditation which always gave me positive energy. Thanks to Gurudev for always helping me, showing me the right path and for constant blessings.

I would also like to thank the creators of beautiful and open source software like LaTeX, INKSCAPE, GEOGEBRA and many more. Also, I should not forget to thank the webpages like Wikipedia, math stack exchange, math overflow and TeX stack exchange.

Finally, it comes to the director of the thesis, my supervisor Dr. Somnath Basu. He is the most important person without whom this was absolutely not possible. His way of doing mathematics is totally different. He has crystal clear concepts, geometric intuition and beautiful imagination. He has been very patient and attentive, simultaneously providing guidance and sharing his mathematical insights. I am thankful to him for his guidance and friendly encouragement throughout my work over the last five years. I feel lucky to have him as an advisor who has taken a keen interest in my progress. Apart from mathematics, his contagious enthusiasm and work ethic makes him a great teacher to work under. I could not imagine someone better to learn mathematics from! I also appreciate his patience and efficiency with regular and extensive online discussions during the lockdown period due to the COVID pandemic. I feel fortunate to know him; he will always be an inspiration to me. Thank you so much Sir, for providing a proper guidance, suggestions and feedback throughout my PhD tenure. It was impossible for me to complete this thesis without your support and supervision.

Finally, thanks to all who are not mentioned here but are associated with me.

To my family.

Publications related to the thesis

1. Basu, S. and Prasad, S. (2021) *A connection between cut locus, Thom space and Morse-Bott functions*, available at https://arxiv.org/abs/2011.02972, to appear in Algebraic & Geometric Topology.

Abstract

Associated to every closed, embedded submanifold N of a connected Riemannian manifold M, there is the distance function d_N which measures the distance of a point in M from N. We analyze the square of this function and show that it is Morse-Bott on the complement of the cut locus $\mathrm{Cu}(N)$ of N, provided M is complete. Moreover, the gradient flow lines provide a deformation retraction of $M - \mathrm{Cu}(N)$ to N. If M is a closed manifold, then we prove that the Thom space of the normal bundle of N is homeomorphic to $M/\mathrm{Cu}(N)$. We also discuss several interesting results which are either applications of these or related observations regarding the theory of cut locus. These results include, but are not limited to, a computation of the local homology of singular matrices, a classification of the homotopy type of the cut locus of a homology sphere inside a sphere, a deformation of the indefinite unitary group $U(p,q)$ to $U(p) \times U(q)$ and a geometric deformation of $GL(n,\mathbb{R})$ to $O(n,\mathbb{R})$ which is different from the Gram-Schmidt retraction.

If a compact Lie group G acts on a Riemannian manifold M freely then M/G is a manifold. In addition, if the action is isometric, then the metric of M induces a metric on M/G. We show that if N is a G-invariant submanifold of M, then the cut locus $\mathrm{Cu}(N)$ is G-invariant, and $\mathrm{Cu}(N)/G = \mathrm{Cu}(N/G)$ in M/G. An application of this result to complex projective hypersurfaces has been provided.

Contents

List of Figures

Notations

$A \cup B$:	union of sets A and B
$A \cap B$:	intersection of sets A and B
$A \times B$:	Cartesian product of A and B
$A \setminus B$:	set of elements in A but not in B
$A \subset B$:	A is a subset of B, not necessarily proper
$A \sqcup B$:	disjoint union of A and B
$A \oplus B$:	direct sum of A and B
$\mathbb{Z}$:	the set of all integers
$\mathbb{R}$:	the set of all real numbers
$\mathbb{C}$:	the set of all complex numbers
$\mathbb{Z}_d$:	the set of all integers modulo d, where d is a positive integer
$\mathbb{R}^n$:	the n-dimensional Euclidean plane, where n is a positive integer
$\mathbb{C}^n$:	the n-dimensional complex plane, where n is a positive integer
$\mathbb{S}^n$:	the unit sphere in $\mathbb{R}^{n+1}$
$\mathbb{D}^n$:	the unit disk in $\mathbb{R}^n$
$\bar{X}$:	the closure of the space X
$M(n, \mathbb{R})$:	the set of all $n \times n$ matrices
$GL(n, \mathbb{R})$:	the set of all $n \times n$ invertible matrices
$O(n, \mathbb{R})$:	the set of all $n \times n$ orthogonal matrices
$SO(n, \mathbb{R})$:	the set of all $n \times n$ orthogonal matrices with determinant 1
$U(n)$:	the set of all $n \times n$ unitary matrices
I_n :	identity matrix of size n
$\mathrm{tr}\,(A)$:	trace of a matrix A
A^T :	transpose of a matrix A
A^* :	conjugate transpose of a matrix A
T_pM :	tangent space of M at $p \in M$
$(T_pN)^\perp$:	orthogonal complement of T_pN, where N is a submanifold of M and $p \in N$
TM :	tangent bundle of M
ν :	normal bundle of N, where N is a submanifold of M
$S(\nu)$:	unit sphere bundle of ν

$D(\nu)$: unit disk bundle of ν

$\exp_p$: Riemannian exponential map at p

$\exp_\nu$: normal exponential map (see (2.1))

$d(p,q)$: the distance between points p and q

$d(A,B)$: the distance between sets A and B

$\nabla f(\mathbf{x})$: gradient of the function f at $\mathbf{x}$

$\pi_n(X)$: n^{th} homotopy group of the space X

$\pi_n(X,A)$: n^{th} relative homotopy group of the pair of spaces (X,A), where $A \subset X$

$H_n(X)$: n^{th} homology of the space X

$\tilde{H}_n(X)$: n^{th} reduced homology of the space X

$H_n(X,A)$: n^{th} relative homology of the pair of spaces (X,A), where $A \subset X$

$H^n(X)$: n^{th} cohomology of the space X

$\mathrm{Cu}(p)$: cut locus of the point p (see Definition 2.3.2)

$\mathrm{Cu}(N)$: cut locus of the set N (see Definition 3.1.2)

$\mathrm{Se}(p)$: separating set of a point p (see (2.5))

$\mathrm{Se}(N)$: separating set of the set N (see Definition 3.1.3)

$A \star B$: topological join of A and B

df_p : the derivative of f at p

$\mathrm{Hess}_p(f)$: the Hessian of f at p (see Definition 2.2.2)

Chapter 1

Introduction

On a Riemannian manifold M, the distance function $d_N(\cdot) := d(N, \cdot)$ from a closed subset N is fundamental in the study of variational problems. For instance, the viscosity solution of the Hamilton-Jacobi equation is given by the flow of the gradient vector of the distance function d_N, when N is the smooth boundary of a relatively compact domain in manifolds; see [Li and Nirenberg, 2005, Mantegazza and Mennucci, 2003]. Although the distance function d_N is not differentiable at N, squaring the function removes this issue. Associated to N and the distance function d_N is a set $\mathrm{Cu}(N)$, the cut locus of N in M. The cut locus of a point (submanifold) consists of all points such that a distance minimal geodesic (see Definitions 2.3.2 and 3.1.2) starting at the point (submanifold) fails its distance minimality property. The aim of the thesis is to explore the topological and geometric properties of cut locus of a submanifold.

1.1 A survey of the cut locus

This section is devoted to the literature survey and a discussion of some known results.

Cut locus of a point, a notion initiated by Henri Poincaré [Poincaré, 1905], has been extensively studied (see [Kobayashi, 1967] for a survey as well as [Buchner, 1977], [Myers, 1935], [Sakai, 1996], and [Wolter, 1979]). Prior to Poincaré it had appeared implicitly in a paper [von Mangoldt, 1881]. Other articles [Whitehead, 1935] and [Myers, 1935, Myers, 1936] describe topological behavior of the cut locus. Due to its topological properties, it became an important tool in the field of Riemannian geometry or Finsler geometry. We list a few references like [Klingenberg, 1959], [Rauch, 1959], and [Cheeger and Ebin, 1975, Chapter 5] for a detailed study of cut locus of a point. We also mention the work around the Blaschke conjecture which uses the geometry of the cut locus of a point, see [Besse, 1978, McKay, 2015]. A great source of reference for articles related to cut loci is [Sakai, 1984, §4]. Further, articles [Sakai, 1977, Sakai, 1978, Sakai, 1979] and [Takeuchi, 1978, Takeuchi, 1979] discussed cut loci in symmetric spaces. For questions on the triangulability of cut loci and differential topological aspects, see [Buchner, 1977, Singer and Gluck, 1976, Gluck and Singer, 1978, Wall, 1977].

Cut locus of submanifolds was first studied by René Thom [Thom, 1972]. We mention some references for cut locus of submanifolds where it has been analyzed via the Eikonal equations and Hamilton-Jacobi equation, for example, see [Angulo Ardoy and Guijarro, 2011, Mantegazza and Mennucci, 2003] as well as analyzed via topological methods, for example, see [Flaherty, 1965, Ozols, 1974, Singh, 1987a, Singh, 1987b, Singh, 1988].

1.2 Overview of results

Suitable simple examples indicate that $M - \mathrm{Cu}(N)$ topologically deforms to N. One of our main results is the following (Theorem 4.3.5).

> **Theorem A.** *Let N be a closed embedded submanifold of a complete Riemannian manifold M and $d : M \to \mathbb{R}$ denote the distance function with respect to N. If $f = d^2$, then its restriction to $M - \mathrm{Cu}(N)$ is a Morse-Bott function, with N as the critical submanifold. Moreover, $M - \mathrm{Cu}(N)$ deforms to N via the gradient flow of f.*

It is observed that this deformation takes infinite time. To obtain a strong deformation retract, one reparameterizes the flow lines to be defined over $[0, 1]$. It can be shown (Lemma 4.3.1) that the cut locus $\mathrm{Cu}(N)$ is a strong deformation retract of $M - N$. A primary motivation for Theorem A came from understanding the cut locus of $N = O(n, \mathbb{R})$ inside $M = M(n, \mathbb{R})$, equipped with the Euclidean metric. We show in Section 3.2 that the cut locus is the set Sing of singular matrices and the deformation of its complement is not the Gram-Schmidt deformation but rather the deformation obtained from the polar decomposition, i.e., $A \in GL(n, \mathbb{R})$ deforms to $A\left(\sqrt{A^T A}\right)^{-1}$. Combining with a result of J. J. Hebda [Hebda, 1983, Theorem 1.4] we are able to compute the local homology of Sing (cf Lemma 3.2.4 and Corollary 3.2.1).

> **Theorem B.** *For $A \in M(n, \mathbb{R})$*
>
> $$H_{n^2-1-i}(\mathrm{Sing}, \mathrm{Sing} - A; G) \cong \widetilde{H}^i(O(n-k, \mathbb{R}); G)$$
>
> *where $A \in \mathrm{Sing}$ has rank $k < n$ and G is any abelian group.*

When the cut locus is empty, we deduce that M is diffeomorphic to the normal bundle ν of N in M. In particular, M deforms to N. Among applications, we discuss two families of examples. We reprove the known fact that $GL(n, \mathbb{R})$ deforms to $O(n, \mathbb{R})$ for any choice of left-invariant metric on $GL(n, \mathbb{R})$ which is right-$O(n, \mathbb{R})$-invariant. However, this deformation is not obtained topologically but by Morse-Bott flows. For a natural choice of such a metric, this deformation (5.2) is not the Gram-Schmidt deformation, but one obtained from the polar decomposition. We also consider $U(p, q)$, the group preserving the indefinite form of signature (p, q) on $\mathbb{C}^n$. We show (Theorem 5.2.1) that $U(p, q)$ deforms to $U(p) \times U(q)$ for the left-invariant metric given by $\langle X, Y \rangle := \mathrm{tr}(X^* Y)$. In particular, we show that the exponential map is surjective for $U(p, q)$ (Corollary 5.2.1). To our knowledge,

this method is different from the standard proof.

For a Riemannian manifold we have the exponential map at $p \in M$, $\exp_p : T_pM \to M$. Let ν denote the normal bundle of N in M. We will modify the exponential map (see §4.3.2) to define the *rescaled exponential* $\widetilde{\exp} : D(\nu) \to M$, the domain of which is the unit disk bundle of ν. The main result (Theorem 4.3.2) here is the observation that there is a connection between the cut locus $\mathrm{Cu}(N)$ and Thom space $\mathrm{Th}(\nu) := D(\nu)/S(\nu)$ of ν.

Theorem C. *Let N be an embedded submanifold inside a closed, connected Riemannian manifold M. If ν denotes the normal bundle of N in M, then there is a homeomorphism*

$$\widetilde{\exp} : D(\nu)/S(\nu) \xrightarrow{\cong} M/\mathrm{Cu}(N).$$

This immediately leads to a long exact sequence in homology (see (4.8))

$$\cdots \to H_j(\mathrm{Cu}(N)) \xrightarrow{i_*} H_j(M) \xrightarrow{q} \widetilde{H}_j(\mathrm{Th}(\nu)) \xrightarrow{\partial} H_{j-1}(\mathrm{Cu}(N)) \to \cdots .$$

This is a useful tool in characterizing the homotopy type of the cut locus. We list a few applications and related results.

Theorem D. *Let N be a homology k-sphere embedded in a Riemannian manifold M^d homeomorphic to S^d.*

1. *If $d \geq k+3$, then $\mathrm{Cu}(N)$ is homotopy equivalent to S^{d-k-1}. Moreover, if M, N are real analytic and the embedding is real analytic, then $\mathrm{Cu}(N)$ is a simplicial complex of dimension at most $d-1$.*
2. *If $d = k+2$, then $\mathrm{Cu}(N)$ has the homology of S^1. There exists homology 3-spheres in S^5 for which $\mathrm{Cu}(N) \simeq S^1$. However, for non-trivial knots K in S^3, the cut locus is not homotopy equivalent to S^1.*

The above results are a combination of Theorem 4.3.3, Theorem 4.3.1 and Example 4.3.3. In general, the structure of the cut locus may be wild (see [Gluck and Singer, 1978], [Itoh and Sabau, 2016], and [Itoh and Vîlcu, 2015]). S. B. Myers [Myers, 1935] had shown that if M is a real analytic sphere, then $\mathrm{Cu}(p)$ is a finite tree each of whose edge is an analytic curve with finite length. Buchner [Buchner, 1977] later generalized this result to cut locus of a point in higher dimensional manifolds. Theorem 4.3.1, which states that the cut locus of an analytic submanifold (in an analytic manifold) is a simplicial complex, is a natural generalization of Buchner's result (and its proof). We attribute it to Buchner, although it is not present in the original paper. This analyticity assumption also helps us to compute the homotopy type of the cut locus of a finite set of points in any closed, orientable, real analytic surface of genus g (Theorem 4.3.4). In Example 4.3.3 we make some observations about the cut locus of embedded homology spheres of codimension 2. This includes the case of real analytic knots in the round sphere $\mathbb{S}^3$.

Let M be a closed Riemannian manifold and G be any compact Lie group

acting on M freely. Then it is known that M/G is a manifold. Further, if the action is isometric, then the metric on M induces a metric on M/G. If N is any G-invariant submanifold of M, then N/G is a submanifold of M/G. If the action is isometric, then we provide an equality between $\mathrm{Cu}(N)/G$ and $\mathrm{Cu}(N/G)$ (Theorem 6.1.1).

> **Theorem E.** Let M be a closed and connected Riemannian manifold and G be any compact Lie group which acts on M freely and isometrically. Let N be any G-invariant closed submanifold of M, then we have an equality
>
> $$\mathrm{Cu}(N)/G = \mathrm{Cu}(N/G).$$

1.3 Outline of Chapter 2

The majority of this chapter is an overview of recalling some basic results in Riemannian geometry and differential topology. This chapter also deals with some known results for cut locus of a point. Although this chapter may be interesting to read and help clarify the concepts, the experts can skip the details.

§2.1 Fermi coordinates

Fermi coordinates are important for studying the geometry of submanifolds. In this coordinate system the metric is rectangular and the derivative of metric vanishes at each point of a curve. It makes the calculations much simpler. This section is devoted to recalling the construction of Fermi coordinates in a tubular neighborhood of a submanifold of a Riemannian manifold. This requires us to define the exponential map restricted to the normal bundle. We have recollected some results which will be used to study the distance squared function from a submanifold. For example, it is shown that the distance squared function from a submanifold is sum of squares of Fermi coordinates in a tubular neighborhood of the submanifold.

§2.2 Morse-Bott theory

In order to study the space via critical points of some real valued function on that space, Morse theory plays an important role. If non-degenerate critical points are replaced by non-degenerate critical submanifolds (see Definition 2.2.5), then a generalization of Morse theory comes into the picture – Morse-Bott theory. In this section, we have recalled the definition of a Morse function and some examples of Morse functions. In §2.2.2 we have discussed Morse-Bott theory motivated by an example.

§2.3 Cut locus and conjugate locus

In a Riemannian manifold M a geodesic γ joining $p, q \in M$ is said to be *distance minimal* if $l(\gamma) = d(p, q)$, where d is the Riemannian distance. Cut locus

of a point captures all points in M beyond which geodesics fail to be distance minimal. In §2.3.1 we have discussed numerous example of cut locus of a point. Characterizations of cut locus has been discussed in terms of conjugate points (points p and q are said to be conjugate along a geodesic γ if there exists a non-vanishing Jacobi field vanishes at p and q) and number of geodesics joining the two points (Theorem 2.3.1). In particular, it says that a cut point is either the first conjugate point or there exists more than one geodesic joining the point and the cut point. We also have a characterization which shows the existence of a closed geodesic (Theorem 2.3.6). One of the result [Wolter, 1979, Theorem 1] is very important to find the cut points, which says that the cut locus of a point is the closure of points which can be joined by more than one geodesic (Theorem 2.3.3).

1.4 Outline of Chapter 3

This chapter serves as a motivation for the results of the subsequent chapters. It includes a detailed discussion of cut locus of submanifolds with numerous examples.

§3.1 Cut locus of submanifolds

To define cut locus of subset of a Riemannian manifold, one needs to define distance minimal geodesic starting from the subset. This section starts with defining the same (Definition 3.1.1) and then the cut locus of a subset is similarly. Example 3.1.9 shows that the cut locus need not be a manifold. Example 3.1.5 shows that the topological join of $\mathbb{S}^k$ and $\mathbb{S}^{n-k-1}$ is induced from cut locus by showing that $\mathrm{Cu}(\mathbb{S}^k_i) = \mathbb{S}^{n-k-1}_l$, where $\mathbb{S}^k_i \hookrightarrow \mathbb{S}^n$ denote the embedding of the k-sphere in the first $k+1$ coordinates and $\mathbb{S}^{n-k-1}_l$ denote the embedding of the $(n-k-1)$-sphere in the last $n-k$ coordinates. In §3.1.1 we have defined the separating set of a subset which consists of all points which have more than one distance minimal geodesic joining the subset. In Example 3.1.6 we have shown that the cut locus is strictly bigger than the separating set.

§3.2 An illuminating example

The main aim of this section is to find the cut locus of the set of all $n \times n$ orthogonal matrices. We have shown that the cut locus is the set of all singular matrices by showing that it is the separating set. We also analyzed the regularity of distance squared function on the singular set and outside the singular set, set of all invertible matrices. In fact, we have shown that the distance squared function is differentiable at A if and only if $A \in GL(n, \mathbb{R})$ In this section we have also shown that $GL(n, \mathbb{R})$ deforms to the set of all orthogonal matrices, but we noted that this deformation is different from one we obtained via Gram-Schmidt. We will also prove Theorem B.

1.5 Outline of Chapter 4

This chapter is based on joint work with Basu [Basu and Prasad, 2021]. Here we have explored some topological properties (relation with the Thom space (Theorem 4.3.2), homology and homotopy groups of cut locus) and geometric properties (regularity of the distance squared function §4.1, complement of cut locus deforms to the submanifold (Theorem 4.3.5)).

§4.1 Regularity of distance squared function

This section is motivated by the example of cut locus of $O(n,\mathbb{R})$ in $M(n,\mathbb{R})$ (§3.2). We proved that the distance squared function is not differentiable on the separating set (Lemma 4.1.1). We have also shown by an example that the distance squared function can be differentiable on points which are cut points but not separating points (Example 4.1.1).

§4.2 Characterizations of $\mathrm{Cu}(N)$

We have discussed two characterizations of cut locus. One in terms of first focal points (Definition 4.2.2) and number of geodesics joining the submanifold to the cut points (Theorem 4.2.1) and other is in terms of separating set (Theorem 4.2.2). The latter one is important for computation viewpoint. Let ν denotes the normal bundle of N and $S(\nu)$ be the unit sphere bundle. Consider a map

$$\rho : S(\nu) \to [0,\infty),$$
$$v \mapsto \sup\{t \in [0,\infty) : \gamma_v|_{[0,t]} \text{ is a distance minimal geodesic from } N\}$$

where γ_v means $\gamma'(0) = v$ (also see (4.4)).

Theorem. Let $u \in S(\nu)$. A positive real number T is $\rho(u)$ if and only if $\gamma_u : [0,T]$ is a distance minimal geodesic from N and at least one of the following holds:

(i) $\gamma_u(T)$ is the first focal point of N along γ_u,

(ii) there exists $v \in S(\nu)$ with $v \neq u$ such that $\gamma_v(T) = \gamma_u(T)$.

Theorem. Let $\mathrm{Cu}(N)$ be the cut locus of a compact submanifold N of a complete Riemannian manifold M. The subset $\mathrm{Se}(N)$, the set of all points in M which can be joined by at least two distance minimal geodesic starting from N, of $\mathrm{Cu}(N)$ is dense in $\mathrm{Cu}(N)$.

§4.3 Topological properties

In this section we start by showing that the cut locus is a simplicial complex for an analytic pair (following Buchner [Buchner, 1977]). In §4.3.2 we prove Theorem C and discuss some applications including Theorem D. We end this section by proving one of the main theorem Theorem A.

1.6 Outline of Chapter 5

We apply our study of gradient of distance squared function to two families of Lie groups - $GL(n,\mathbb{R})$ and $U(p,q)$. With a particular choice of left-invariant Riemannian metric which is right-invariant with respect to a maximally compact subgroup K, we analyze the geodesics and the cut locus of K. In both cases, we obtain that G deforms to K via Morse-Bott flow (Lemma 5.1.1 and Theorem 5.2.1). Although these results are deducible from classical results of Cartan and Iwasawa, our method is geometric and specific to suitable choices of Riemannian metrics. It also makes very little use of structure theory of Lie algebras.

1.7 Outline of Chapter 6

Consider a Riemannian manifold M on which a compact Lie group G acts freely. It is well known that the quotient M/G is a manifold. This chapter is devoted to the study of cut locus of a G-invariant submanifold N inside M. We will prove Theorem E. As an application of Theorem E, we have shown some examples of cut locus in orbit space. We also discuss an application to complex hypersurfaces. Let $\pi : \mathbb{S}^{2n+1} \to \mathbb{CP}^n$ be the quotient map. If

$$X(d) = \left\{ [z_0 : z_1 : \cdots : z_n] \in \mathbb{CP}^n : \sum_{i=0}^{n} z_i^d = 0 \right\}$$

and $\tilde{X}(d) := \pi^{-1}(X(d))$, then we make the following conjecture.

Conjecture. The cut locus of $\tilde{X}(d) \subseteq \mathbb{S}^{2n+1}$ is $\mathbb{Z}_d^{\star(n+1)} \times_{\mathbb{Z}_d} \mathbb{S}^1$, where $\times_{\mathbb{Z}_d}$ is the diagonal action of $\mathbb{Z}_d$ and $\star$ denotes the topological join of spaces.

We prove the above conjecture for two families: $d = 2, n$ arbitrary (Theorem 6.3.1) and $n = 1, d$ arbitrary (Theorem 6.3.2).

Chapter 2

Preliminaries

Contents

2.1 Fermi coordinates

In this section we give a brief overview of the Fermi coordinates which are generalizations of normal coordinates in Riemannian geometry. To study the distance squared function from a submanifold N of a Riemannian manifold M, it is essential to analyze the local geometry of M around N. For this the Fermi coordinates are the most convenient tool. In 1922, Enrico Fermi [Fermi, 1922] came up with a coordinate system in which the Christoffel symbols vanish along geodesics which makes the metric simpler. For an extensive reading we refer to the book [Gray, 2004, Chapter 2] and an article [Manasse and Misner, 1963].

2.1.1 Normal exponential map

Let N be an embedded submanifold of a Riemannian manifold M. We define the *normal bundle*, denoted by ν,

$$\nu := \left\{(p, v) : p \in N \text{ and } v \in (T_pN)^{\perp}\right\},$$

where $(T_pN)^\perp$ is the orthogonal complement of T_pN. Indeed, ν is a subbundle of the restriction of the tangent bundle TM to N. We can restrict the usual exponential map of the Riemannian manifold to the normal bundle to define the exponential map of the normal bundle. We define the *exponential map of the normal bundle* as follows:

$$\exp_\nu : \nu \to M,\ (p, v) \mapsto \exp_p(v), \tag{2.1}$$

where $\exp_p : T_pM \to M$ is the exponential map of M. We may write $\exp_\nu(v)$ in short and call this the *normal exponential map*. Note that we can identify N as the zero section of the normal bundle and hence N can be assumed to be submanifold of ν.

Lemma 2.1.1. [Gray, 2004, Lemma 2.3] Let M be a Riemannian manifold and N be any embedded submanifold. Then the normal exponential map $\exp_\nu : \nu \to M$ is a diffeomorphism from a neighbourhood of $N \subseteq \nu$ onto a neighbourhood of $N \subseteq M$.

Using the above lemma, let $\mathcal{U}_N$ be the largest open neighbourhood of $N \subseteq \nu$ for which $\exp_\nu$ is a diffeomorphism. We shall later be able to describe this neighbourhood in terms of a function ρ (4.4). We now ready to define the Fermi coordinates.

2.1.2 Fermi coordinate system

To define a system of Fermi coordinates, we need an arbitrary system of coordinates $(y_1, \cdots, y_k)$ defined in a neighborhood $\mathcal{O} \subseteq N$ of $p \in N$ together with orthogonal sections $\mathscr{E}_{k+1}, \cdots, \mathscr{E}_n$ of the restriction on ν to $\mathcal{O}$.

Definition 2.1.1 (Fermi coordinates). The *Fermi coordinates* $(x_1, \cdots, x_n)$ of $N \subseteq M$ centered at p (relative to a given coordinate $(y_1, \cdots, y_k)$ on N and given orthogonal sections $\mathscr{E}_{k+1}, \cdots, \mathscr{E}_n$ of ν) are defined by

$$x_l\left(\exp_\nu\left(\sum_{i=k+1}^{n} \tau_i \mathscr{E}_i\left(p'\right)\right)\right) = y_l\left(p'\right),\ l = 1, \cdots, k$$

$$x_m\left(\exp_\nu\left(\sum_{i=k+1}^{n} \tau_i \mathscr{E}_i\left(p'\right)\right)\right) = \tau_m,\ m = k+1, \cdots, n$$

for $p' \in \mathcal{O}$ provided the numbers $\tau_{k+1}, \cdots, \tau_n$ are small enough so that $\tau_{k+1}\mathscr{E}_{k+1}\left(p'\right) + \cdots + \tau_n \mathscr{E}_n\left(p'\right) \in \mathcal{U}_N$.

As the normal exponential map is a diffeomorphism on the set $\mathcal{U}_N$, $(x_1, \cdots, x_k, x_{k+1}, \cdots, x_n)$ defines a coordinate system near p. In fact, the restrictions to N of coordinate vector fields $\partial/\partial x_{k+1}, \ldots, \partial/\partial x_n$ are orthonormal.

Lemma 2.1.2. Let γ be a unit speed geodesic normal to N with $\gamma(0) = p \in N$. If $v = \gamma'(0)$, then there exists a system of Fermi coordinates $(x_1, \cdots, x_n)$ such that whenever $(p, tv) \in \mathcal{U}_N$, we have

$$\left.\frac{\partial}{\partial x_{k+1}}\right|_{\gamma(t)} = \gamma'(t),$$
$$\left.\frac{\partial}{\partial x_l}\right|_p \in T_pN, \text{ and } \left.\frac{\partial}{\partial x_i}\right|_p \in (T_pN)^\perp$$

for $1 \leq l \leq k$ and $k+1 \leq i \leq n$. Furthermore, for $1 \leq j \leq n$

$$(x_j \circ \gamma)(t) = t\delta_{j(k+1)}.$$

The following object will be useful while studying the distance squared function from a submanifold N.

Definition 2.1.2. Let N be a submanifold of a Riemannian manifold and let $(x_1, \cdots, x_n)$ be a system of Fermi coordinates for N. We define $\Delta(x_1, \cdots, x_n)$ to be the non-negative number satisfying

$$\Delta^2 = \sum_{i=k+1}^{n} x_i^2.$$

Lemma 2.1.3. Let $p \in N$. The Δ is independent of the choice of Fermi coordinates at p.

Proof. Let $(x_1', \cdots, x_n')$ be another system of Fermi coordinates at p, and let $\{\mathscr{E}_{k+1}', \cdots, \mathscr{E}_n'\}$ be the orthonormal sections of ν that give rise to it. We can write

$$\mathscr{E}_j' = \sum_{i=k+1}^{n} a_{ij}\mathscr{E}_i$$

where (a_{ij}) is a matrix of functions in the orthogonal group $O(n-k)$ with each a_{ji} being a smooth function on N. Now,

$$\begin{aligned} x_m\left(\exp_\nu\left(\sum_{j=k+1}^{n}\tau_j'\mathscr{E}_j'\right)\right) &= x_m\left(\exp_\nu\left(\sum_{j=k+1}^{n}\tau_j'\sum_{i=k+1}^{n}a_{ij}\mathscr{E}_i\right)\right) \\ &= x_m\left(\exp_\nu\left(\sum_{i=k+1}^{n}\left(\sum_{j=k+1}^{n}a_{ij}\tau_j'\right)\mathscr{E}_i\right)\right) \\ &= \sum_{l=k+1}^{n} a_{ml}\tau_l' \\ &= \sum_{l=k+1}^{n} a_{ml}x_l'\left(\exp_\nu\left(\sum_{j=k+1}^{n}\tau_j'\mathscr{E}_j'\right)\right). \end{aligned}$$

Therefore, we have

$$x_m = \sum_{l=k+1}^{n} a_{ml} x'_l, \ m = k+1, \cdots, n. \tag{2.2}$$

Now consider,

$$\begin{aligned}
\sum_{m=k+1}^{n} x_m^2 &= \sum_{m=k+1}^{n} \left(\sum_{l=k+1}^{n} a_{ml} x'_l \right)^2 \\
&= \sum_{m=k+1}^{n} \left(\sum_{l=k+1}^{n} \sum_{j=k+1}^{n} (a_{ml} x'_l)(a_{mj} x'_j) \right) \\
&= \sum_{l=k+1}^{n} \sum_{j=k+1}^{n} \left(\sum_{m=k+1}^{n} a_{ml} a_{mj} \right) x'_l x'_j \\
&= \sum_{l=k+1}^{n} \sum_{j=k+1}^{n} \delta_{lj} x'_l x'_j \\
&= \sum_{m=k+1}^{n} (x'_m)^2 .
\end{aligned}$$

□

2.2 Morse-Bott theory

This section will be devoted to a generalization of Morse function in which we study the space by looking at the critical points of a smooth real valued function. We will briefly recall Morse functions with a couple of examples, and then we will define Morse-Bott functions. The reference for this section will be the original article by Raoul Bott [Bott, 1954] and the book [Banyaga and Hurtubise, 2004, Section 3.5].

2.2.1 Morse functions

Broadly the "functions" and "spaces" are objects of study in analysis and geometry respectively. However, these two objects are related to each other. For example, on a line we can have functions like $f(x) = x$, $g(x) = x^2$ which takes arbitrarily large values, whereas on the circle there does not exist any function which takes arbitrarily large value. In this way, we are able to differentiate circles with lines by seeing functions on them. Morse theory studies relations between shape of space and function defined on this space. We study the critical points of a function defined on spaces to find out information on the space. More specifically, in Morse theory we study the topology of smooth manifolds by analyzing the critical point of a smooth real valued function. If $f : M \to \mathbb{R}$ is a smooth function on a smooth manifold M, then using Morse theory we can find a CW-complex which is homotopy equivalent to M and the CW-complex has one cell for each critical point

of f. For a detailed study of Morse theory we refer to the book [Milnor, 1963] by John Milnor.

Definition 2.2.1 (Critical Points). Let M and N be two smooth manifolds of dimension m and n respectively. A point $p \in M$ is said to be *critical point* of a smooth function $f : M \to N$ if the differential map

$$df_p : T_pM \to T_{f(p)}N$$

does not have full rank.

We confine our study to real-valued functions. In this case the above is equivalent to $df_p \equiv 0$. In a coordinate neighborhood $(\phi = (x_1, x_2, \ldots, x_n), U)$ around p, we have

$$\frac{\partial(f \circ \phi^{-1})}{\partial x_j}(\phi(p)) = 0,\ j = 1, \cdots, n. \tag{2.3}$$

At a critical point of $f : M \to \mathbb{R}$, we define the Hessian which is similar to the second derivative of the function.

Definition 2.2.2 (Hessian of f at p). Let $f : M \to \mathbb{R}$ be any smooth real valued function and p be any critical point of f. The *Hessian of f at p* is the map

$$\mathrm{Hess}_p(f) : T_pM \times T_pM \to \mathbb{R},\ \mathrm{Hess}_p(f)(V, W) = \tilde{V} \cdot \left(\tilde{W} \cdot f\right)(p), \tag{2.4}$$

where $\tilde{V}$ and $\tilde{W}$ are any extensions of V and W respectively.

Note that the Hessian is a bilinear form of V and W. Consider

$$\begin{aligned} V \cdot \left(\tilde{W} \cdot f\right)(p) - W \cdot \left(\tilde{V} \cdot f\right)(p) &= \left[\tilde{V}, \tilde{W}\right]_p \cdot f \\ &= df_p\left(\left[\tilde{V}, \tilde{W}\right]_p\right) \\ &= 0. \end{aligned}$$

Thus, Hessian is a symmetric bilinear form on $T_pM \times T_pM$. The above computation, in particular, also proves that the definition is well defined, that is, it is independent of the choice of extension.

Any critical point is categorized by looking at the value of Hessian at that point.

Definition 2.2.3. A critical point $p \in M$ of a smooth function $f : M \to \mathbb{R}$ is said to be *non-degenerate* if the Hessian is non-degenerate. Otherwise, we call p to be a *degenerate* critical point. The *index* of a non-degenerate critical point p is the dimension of the subspace of the maximum dimension on which $\mathrm{Hess}_p f$ is negative definite.

For example, the function $f : \mathbb{R} \to \mathbb{R}\ x \mapsto x^2$ has 0 a critical point which is non-degenerate but 0 is the degenerate critical point of the function $f(x) = x^3$.

Definition 2.2.4. A smooth function $f : M \to \mathbb{R}$ is said to be a *Morse function* if all its critical points are non-degenerate.

Example 2.2.1. The function

$$f : \mathbb{R}^2 \to \mathbb{R},\ (x, y) \mapsto x^2 - 3xy^2$$

is not a Morse function, as the critical point $(0, 0)$ is not non-degenerate.

Example 2.2.2 (Height function on sphere). The height function on the n-sphere is a Morse function with critical points $N = (0, 0, \cdots, 1)$ and $S = (0, 0, \cdots, -1)$. The index of N and S is n and 0 respectively.

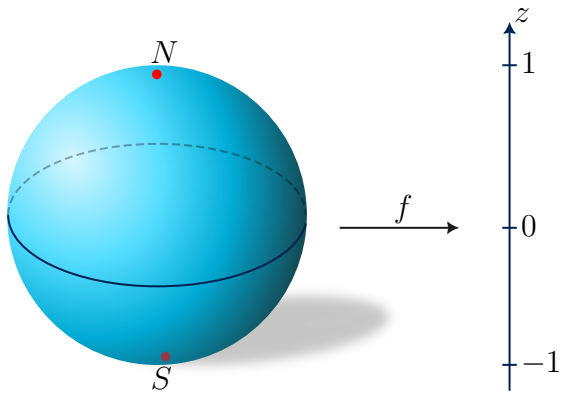


Figure 2.1: Height function on the 2-sphere is a Morse function with two non-degenerate critical points with index 2 and 0.

For this, let

$$\phi_1 : \mathbb{S}^n \setminus \{N\} \to \mathbb{R}^n,\ (x_1, \cdots, x_{n+1}) \mapsto \left(\frac{x_1}{1 - x_{n+1}}, \cdots, \frac{x_n}{1 - x_{n+1}} \right), \text{ and}$$

$$\phi_2 : \mathbb{S}^n \setminus \{S\} \to \mathbb{R}^n,\ (x_1, \cdots, x_{n+1}) \mapsto \left(\frac{x_1}{1 + x_{n+1}}, \cdots, \frac{x_n}{1 + x_{n+1}} \right)$$

be two charts of $\mathbb{S}^n$. The inverse is given by

$$\phi_1^{-1}(\mathbf{y}) = \left(\frac{2y_1}{\|\mathbf{y}\|^2 + 1}, \cdots, \frac{2y_n}{\|\mathbf{y}\|^2 + 1}, \frac{\|\mathbf{y}\|^2 - 1}{\|\mathbf{y}\|^2 + 1} \right)$$

$$\phi_1^{-1}(\mathbf{y}) = \left(\frac{2y_1}{\|\mathbf{y}\|^2 + 1}, \cdots, \frac{2y_n}{\|\mathbf{y}\|^2 + 1}, -\frac{\|\mathbf{y}\|^2 - 1}{\|\mathbf{y}\|^2 + 1} \right).$$

From equation (2.3), the critical points of f will be the critical points of $\psi_i = f \circ \phi_i^{-1} : \mathbb{R}^n \to \mathbb{R},\ i = 1, 2$. Note that

$$\psi_1(\mathbf{x}) = \frac{\|\mathbf{x}\|^2 - 1}{\|\mathbf{x}\|^2 + 1}, \text{ and } \psi_2(\mathbf{x}) = -\frac{\|\mathbf{x}\|^2 - 1}{\|\mathbf{x}\|^2 + 1},\ \mathbf{x} \in \mathbb{R}^n.$$

$$\implies (d\psi_1)_{\mathbf{x}} = \frac{4\mathbf{x}}{(\|\mathbf{x}\| + 1)^2}, \text{ and } (d\psi_2)_{\mathbf{x}} = -\frac{4\mathbf{x}}{(\|\mathbf{x}\| + 1)^2}.$$

Therefore, the critical points are $\phi_1^{-1}(\mathbf{0}) = S$ in $\mathbb{S}^n \setminus \{N\}$ and $\phi_2^{-1}(\mathbf{0}) = N$ in $\mathbb{S}^n \setminus \{N\}$. Note that

$$\mathrm{Hess}_S(f) = \left(\frac{\partial^2 \psi_1}{\partial x_i \partial x_j}(\mathbf{0})\right)_{1\le i,j\le n} = 4I_{n\times n}, \text{ and}$$

$$\mathrm{Hess}_N(f) = \left(\frac{\partial^2 \psi_1}{\partial x_i \partial x_j}(\mathbf{0})\right)_{1\le i,j\le n} = -4I_{n\times n}.$$

Hence, both critical points are non-degenerate and index of N and S is n and 0 respectively.

Example 2.2.3 (Height function on torus). If a and b be two positive real numbers with $0 < b < a$, then the torus is

$$\mathbb{T} =: \left\{(x, y, z) : x^2 + \left(\sqrt{y^2 + z^2} - a\right)^2 = b^2\right\}.$$

The function

$$f : \mathbb{T} \to \mathbb{R},\ (x, y, z) \mapsto z$$

is a Morse function with critical points $(0, 0, \pm(a + b))$ and $(0, 0, \pm(a - b))$.

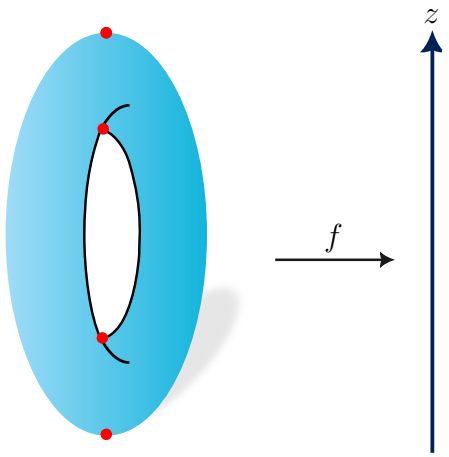


Figure 2.2: Height function on torus is a Morse function with four non-degenerate critical points

2.2.2 Morse-Bott functions

Morse-Bott functions are generalizations of Morse functions where we are allowed to have critical set need not be isolated but may form a submanifold. For example, let a torus be kept horizontally (a donut is kept in a plate). If f is the height function on the torus, then there are two critical submanifolds, the top and bottom circles.

Let M be a Riemannian manifold and f be any real valued smooth function on M. Let $\mathrm{Cr}(f)$ denotes the set of all critical points of f and N be any submanifold of M which is contained in $\mathrm{Cr}(f)$. For any point $p \in M$ we have the following decomposition:

$$T_pM = T_pN \oplus \nu_pN,$$

where $\nu_p N$ is the normal bundle at p. Note that if $p \in N$ then for any $V \in T_pN$ and $W \in T_pM$ the Hessian vanishes, i.e., $\mathrm{Hess}_p(f)(V, W) = 0$. Therefore, $\mathrm{Hess}_p(f)$ induces a symmetric bilinear form on $\nu_p N$. Now we can define non-degenerate critical submanifold similar to the non-degenerate critical points.

Definition 2.2.5 (Non-degenerate critical submanifold). Let $N \subset M$ be a submanifold of a Riemannian manifold M. Then N is said to be *non-degenerate critical submanifold* of f if $N \subseteq \mathrm{Cr}(f)$ and for any $p \in N$ the Hessian, $\mathrm{Hess}_p(f)$ is non-degenerate in the direction normal to N at p.

In the above definition, by $\mathrm{Hess}_p(f)$ is non-degenerate in the direction normal to N at p we mean that for any $V \in \nu_p N$ there exists $W \in \nu_p N$ such that $\mathrm{Hess}_p(f)(V, W) \neq 0$.

Definition 2.2.6 (Morse-Bott functions). The function $f : M \to \mathbb{R}$ is said to be *Morse-Bott* if the connected components of $\mathrm{Cr}(f)$ are non-degenerate critical submanifolds.

Example 2.2.4. Let $f : M \to \mathbb{R}$ be a Morse function. Then the critical submanifolds are zero-dimensional and hence the Hessian $\mathrm{Hess}_p(f)$ at any critical point p is non-degenerate in every direction as all the directions are normal. So f is Morse-Bott with critical submanifolds as critical points.

Example 2.2.5. Any constant function defined on a smooth manifold M is a Morse-Bott function with critical submanifold M.

Example 2.2.6. Let $M = \mathbb{R}^2$. Define

$$f : M \to \mathbb{R},\ (x, y) \mapsto x^4.$$

Then the derivative map

$$df_{(x,y)} = \left(4x^3, 0\right) = (0, 0) \implies x = 0.$$

Thus, the critical set is $\{(x, y) : x = 0\}$ which is y-axis. Now the Hessian at $(0, y)$ will be

$$\mathrm{Hess}_{(0,y)}(f) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

which is degenerate in every direction and hence it is not a Morse-Bott function.

Example 2.2.7. Let $M = \mathbb{R}^2$ with the Euclidean distance dist and $N = \{(x, x) : x \in \mathbb{R}\}$. Consider the function

$$f : M \to \mathbb{R},\ (x, y) \mapsto \mathrm{dist}^2((x, y), N) = \frac{(x - y)^2}{2}.$$

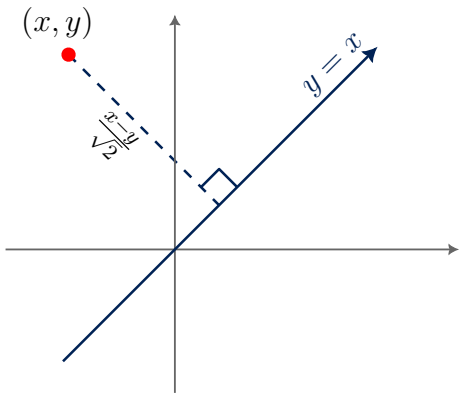


Figure 2.3: Distance of (x, y) from the line $y = x$.

So we have

$$df_{(x,y)} = (x - y, y - x) = 0 \implies x = y.$$

Thus, the critical submanifold is N. Now to see whether it is non-degenerate or not in the normal direction, we need to compute the Hessian. Let $(p, p) \in N$ be any critical point.

$$\mathrm{Hess}_{(p,p)}(f) = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Note that for any $\mathbf{v} = (a, -a) \in \left(T_{(p,p)}N\right)^{\perp}$ with $\mathbf{v} \neq 0$, we have

$$\mathrm{Hess}_{(p,p)}(f)(\mathbf{v}, \mathbf{v}) = \mathbf{v}^T \mathrm{Hess}_{(p,p)}(f)\mathbf{v} = 4a^2 \neq 0.$$

Thus, the given function is Morse-Bott.

Example 2.2.8. Let $M = \mathbb{R}^{n+1}$ with the Euclidean metric d. If $N = \mathbb{S}^n$ be the unit sphere, then the distance between a point $\mathbf{p} \in \mathbb{R}^{n+1}$ and N is given by

$$\mathrm{dist}(\mathbf{p}, N) := \inf_{\mathbf{q} \in N} \mathrm{dist}(\mathbf{p}, \mathbf{q}).$$

We shall denote by d^2 the square of the distance. Now consider the function

$$f : M \to \mathbb{R}, \quad \mathbf{x} \mapsto \mathrm{dist}^2(\mathbf{x}, N) = \left(\|\mathbf{x}\| - 1\right)^2.$$

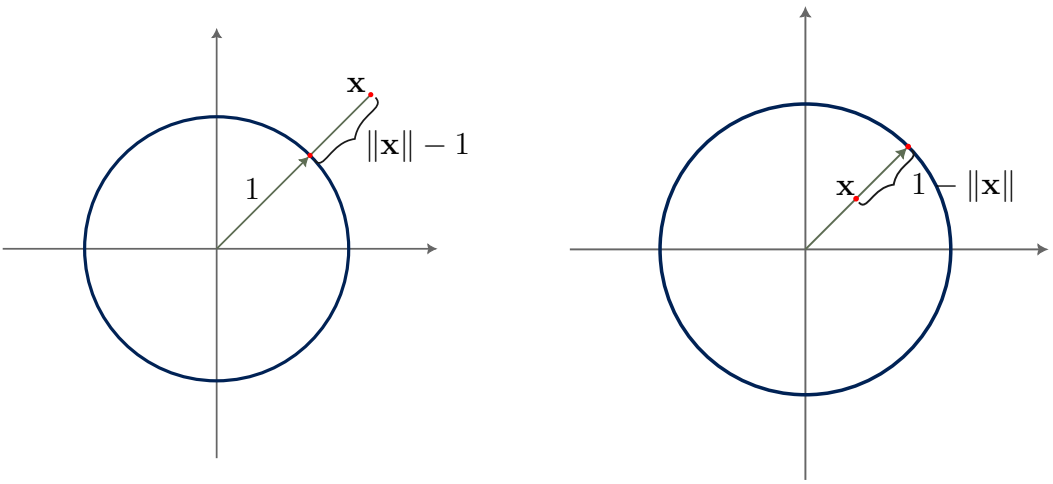


Figure 2.4: Distance of $\mathbf{x}$ from the unit circle.

The function $f : M - \{\mathbf{0}\}$ is a Morse-Bott function with $N = \mathbb{S}^n$ as the critical submanifold. We will see a general version of this example in Chapter 4.

Example 2.2.9. Consider the function

$$f : \mathbb{S}^2 \to \mathbb{R},\ (x, y, z) \mapsto z^2.$$

It is square of the height function discussed in the Example 2.2.8. We claim that f is a Morse-Bott function with critical set as $N = (0, 0, 1),\ S = (0, 0, -1)$ and the equator $E = \{(x, y, 0) : x^2 + y^2 = 1\}$.

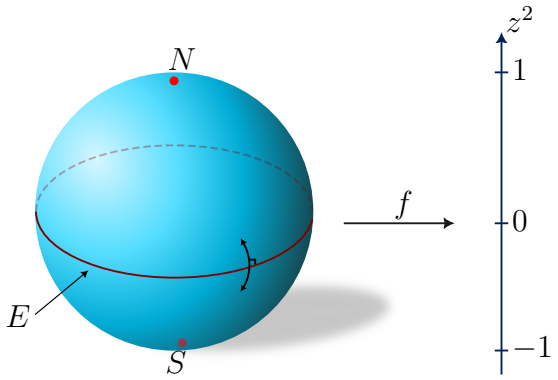


Figure 2.5: Square of the height function on sphere has three critical submanifolds; north pole, south pole and the equator circle.

We take the charts on $\mathbb{S}^2$ as given in Example 2.2.2. So we have

$$\psi_1 : \mathbb{R}^2 \to \mathbb{R},\ \mathbf{p} = (x, y) \mapsto \left(\frac{\|\mathbf{p}\|^2 - 1}{1 + \|\mathbf{p}\|^2}\right)^2, \text{ and}$$

$$\psi_2 : \mathbb{R}^2 \to \mathbb{R},\ \mathbf{p} = (x, y) \mapsto \left(\frac{1 - \|\mathbf{p}\|^2}{1 + \|\mathbf{p}\|^2}\right)^2.$$

The critical points are

$$d\psi_{1_\mathbf{p}} = (0, 0) \implies \frac{8\left(\|\mathbf{p}\|^2 - 1\right)}{\left(1 + \|\mathbf{p}\|^2\right)^3}\mathbf{p} = \mathbf{0}$$

$$\implies \|\mathbf{p}\| = 1 \text{ or } \mathbf{p} = 0.$$

Similarly, ψ_2 gives the same condition and hence the critical set will be N, S, and E (Figure 2.5). It is clear that the two submanifolds $\{N\}$ and $\{S\}$ are non-degenerate. To show that E is non-degenerate, we calculate the Hessian matrix at any point of E, say $\mathbf{p} = (x, y, 0)$. The Hessian with respect to the above charts is given by

$$\mathrm{Hess}_\mathbf{p}(f) = \begin{pmatrix} 2x^2 & 2xy \\ 2xy & 2y^2 \end{pmatrix}.$$

Note that for any $\mathbf{p} \in E$ the normal space $(T_\mathbf{p}E)^\perp$ is spanned by $(0, 0, 1)$. Since we have

$$T_\mathbf{p}\mathbb{S}^2 = T_\mathbf{p}E \oplus (T_\mathbf{p}E)^\perp = \mathrm{span}\{(-y, x, 0)\} \oplus \mathrm{span}\{(0, 0, 1)\}.$$

We need to show that for any $\mathbf{v}(0,0,\alpha)$ with $\alpha \neq 0$, there exists $\mathbf{w}(0,0,\beta)$ such that $\mathrm{Hess}_{\mathbf{p}}(f)(\mathbf{v},\mathbf{w}) \neq 0$. For that we will identify $T_{\mathbf{p}}\mathbb{S}^2$ with $\mathbb{R}^2$ Consider two curves γ and η passing through p.

$$\gamma(t) = \{(x\cos t - y\sin t, y\cos t + x\sin t, 0) : 0 \leq t \leq 2\pi\}$$
$$\eta(t) = \{(x\cos t, y\cos t, \sin t) : 0 \leq t \leq 2\pi\}.$$

Note that $\gamma'(0) = (-y, x, 0) = \mathbf{v}_1$ and $\eta'(0) = (0,0,1) = \mathbf{v}_2$. So we have

$$d\phi_{1_{\mathbf{p}}}(\mathbf{v}_1) = \frac{\mathrm{d}}{\mathrm{d}t}(\phi_1 \circ \gamma)(t)\big|_{t=0} = (-y, x)$$
$$d\phi_{1_{\mathbf{p}}}(\mathbf{v}_2) = \frac{\mathrm{d}}{\mathrm{d}t}(\phi_1 \circ \eta)(t)\big|_{t=0} = (-x, -y).$$

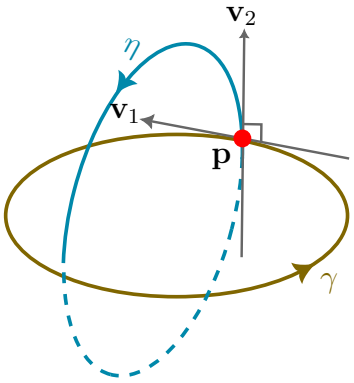


Figure 2.6: The curves γ and η passing through $\mathbf{p}$

So, we can define an isomorphism between

$$T_{\mathbf{p}}\mathbb{S}^2 \to \mathbb{R}^2,\ (-y,x,0) \mapsto (-y,x) \text{ and } (0,0,1) \mapsto (-x,-y).$$

Now note that

$$\begin{aligned}
\mathrm{Hess}_{\mathbf{p}}((-x,-y),(-x,-y)) &= (-x,-y)\begin{pmatrix} 2x^2 & 2xy \\ 2xy & 2y^2 \end{pmatrix}\begin{pmatrix} -x \\ -y \end{pmatrix} \\
&= (-x,-y)\begin{pmatrix} -2x^3 & -2xy^2 \\ -2x^2y & -2y^3 \end{pmatrix} \\
&= 2x^4 + 2x^2y^2 + 2x^2y^2 + 2y^4 \\
&= 2\left(x^2+y^2\right)^2 = 2.
\end{aligned}$$

Thus, Hessian is non-degenerate in the normal direction and hence it is a Morse-Bott functions.

Remark 2.2.1. The above example, in particular, shows that the critical submanifolds may have different dimensions.

Example 2.2.10. Let $f : M \to \mathbb{R}$ be a Morse-Bott function. If $\pi : X \to M$ is any smooth fiber bundle, then the composition $\pi \circ f : X \to M$ is a Morse-Bott function.

The trace function on $SO(n,\mathbb{R}), U(n,\mathbb{C})$ and $Sp(n,\mathbb{C})$ is a Morse-Bott function (cf [Banyaga and Hurtubise, 2004, page 90, Exercise 22]).

2.3 Cut locus and conjugate locus

Let M be a complete Riemannian manifold and $p \in M$. Let γ be a geodesic such that $\gamma(0) = p$. A *cut point* of p along the geodesic γ is the first point q on γ such that for any point $\tilde{q}$ on γ beyond q, there exists a geodesic $\tilde{\gamma}$ joining p to $\tilde{q}$ such that $l(\tilde{\gamma}) < l(\gamma)$, where $l(\gamma)$ is the length of γ. In simple words, q is the first point beyond which γ stops to minimize the distance. In this section we will recall the definition of cut locus of a point with some examples. We will also mention some important results which will be generalized in the upcoming chapters. The main references for this section are books [Sakai, 1996, Chapter 3, Section 4] and [Cheeger and Ebin, 1975, Chapter 5].

2.3.1 Cut locus of a point

Let M be a Riemannian manifold and $p, q \in M$ be two points. If there exists a piecewise differentiable curve joining them, then using the Riemannian metric we can measure the length of the curve. We now consider all possible curves joining these points. Then the distance between p and q is the infimum of the length of all (piecewise differentiable) curves joining p and q. This distance induces a metric. We call M to be *complete Riemannian manifold* if (M, d) is a complete metric space. From now onwards, we always consider M to be a complete Riemannian manifold. A geodesic $\gamma(t),\ t \in [a, b]$ is said to be *extendable* if it can be extended to a geodesic $\gamma(t),\ t \in [c, d] \supsetneq [a, b]$. A Riemannian manifold is said to be *geodesically complete* if any geodesic can be extendable for all $t \in \mathbb{R}$. Then the Hopf-Rinow Theorem [Hopf and Rinow, 1931] says that these two notions of completeness are equivalent. If a manifold is not complete, then we can not always extend a geodesic. For example, $\mathbb{R}^2 \setminus \{\mathbf{0}\}$ is not complete and the geodesic $\gamma(t) = t,\ t > 0$ is not extendable in the negative x-axis. This problem does not arise if the manifold is complete. The more is true which says that M is complete if and only if every geodesic can be extended for infinite time. The completeness of M also guarantees that any two points can be joined by a distance minimal geodesic which is defined as follows.

Definition 2.3.1 (Distance minimal geodesic). A geodesic joining p and q is said to be *distance minimal* if the length of the geodesic is equal to the distance between these points, i.e., $l(\gamma) = d(p, q)$.

We shall now define the cut locus, $\mathrm{Cu}(p)$ of a point p in a complete Riemannian manifold M. The notion of cut locus was first introduced for convex surfaces by Henri Poincaré [Poincaré, 1905] in 1905 under the name *la ligne de partage* meaning *the dividing line*.

Definition 2.3.2 (Cut locus of a point). Let M be a complete Riemannian manifold and $p \in M$. If $\mathrm{Cu}(p)$ denotes the *cut locus* of p, then a point $q \in \mathrm{Cu}(p)$ if there exists a minimal geodesic joining p to q any extension of which beyond q is not minimal.

Consider the set

$$S = \{s > 0 : \gamma(t),\ 0 \leq t \leq s \text{ is a distance minimal geodesic}\}.$$

If $S = (0, t_0)$, then $\gamma(t_0)$ is the cut point of p along γ, and if $S = (0, \infty)$, then the point p does not have a cut locus along $\gamma(t)$.

Note that if q_0 is a point on the geodesic $\gamma(t)$ which comes after the cut point, i.e., $q = \gamma(t_0)$ and $q_0 = \gamma(t),\ t > t_0$, then there is a geodesic $\eta(t)$ joining p to q_0 such that $l(\eta) < l(\gamma)$ (see Figure 2.7).

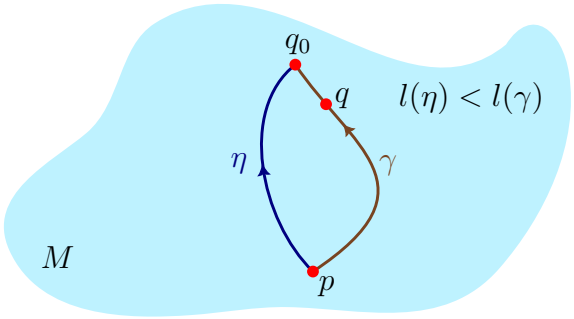


Figure 2.7: A point which is beyond cut point can be joined by a shorter geodesic

If q_0 comes before the cut point q, then we can not find any geodesic shorter than γ joining p to q_0. Moreover, we even can not find another geodesic η joining p to q_0 such that $l(\gamma) = l(\eta)$. So we can say that if q_0 is coming before cut point, then γ is the only minimal geodesic joining p to q_0. To prove this fact, we assume that if η is another geodesic joining p to q_0 such that $l(\gamma) = l(\eta)$ then

$$\delta(t) = \begin{cases} \eta(t),\ 0 \leq t \leq t_1 \\ \gamma(t),\ 0 \leq t_1 \leq t_0 \end{cases}$$

is a curve such that $l(\delta) = d(p, q) = l(\gamma)$. Choose two points q_1 and q_2 sufficiently close to q_0 as shown in Figure 2.8. Then α is a distance minimal geodesic joining points q_1 and q_2 and hence we got a curve ζ which is η from p to q_1, α from q_1 to q_2 and γ from q_2 to q. Note that $l(\zeta)$ is less than the distance between p and q, which is a contradiction.

We now will discuss some examples.

Example 2.3.1. Let $M = \mathbb{R}^n$ be the n-Euclidean plane equipped with the Euclidean metric. The cut locus of any point is a null set because any geodesic never fails to satisfy its distance minimizing property.

Example 2.3.2 (Cut locus of a point in n-sphere). Let $M = \mathbb{S}^n$ be the n-sphere with the round metric. The geodesics are great circles. The cut locus of the south pole is the north pole.

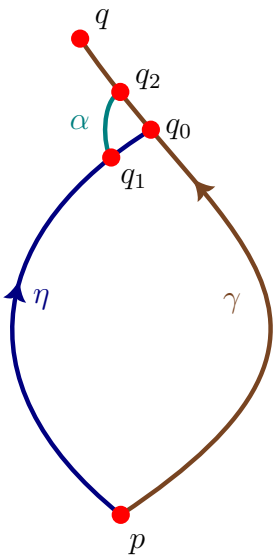


Figure 2.8: If a point appears before the cut point along the geodesic, then it can not be joined by two or more minimal geodesics

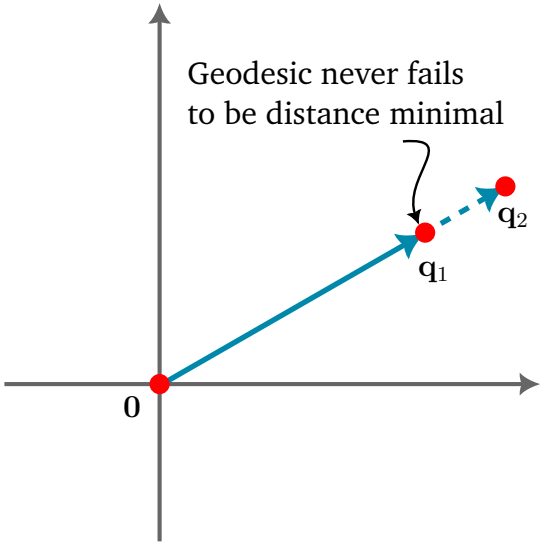
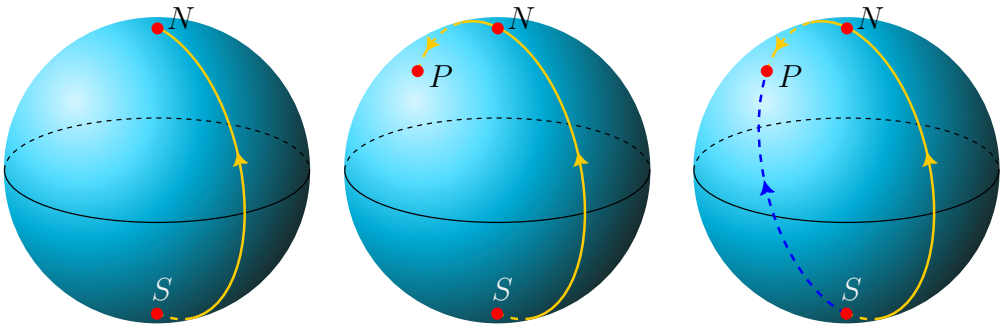


Figure 2.9: Cut locus of $(0, 0)$ in $\mathbb{R}^2$

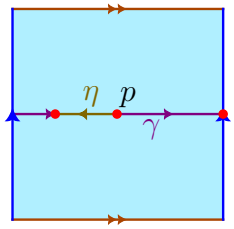


(a) (b) (c)

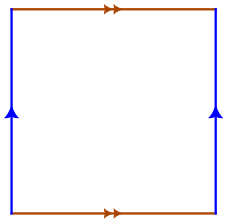
Figure 2.10: Cut locus of south pole in $\mathbb{S}^2$

In Figure 2.10 we have proven the claim. If γ is a geodesic from south pole S to north pole N, then the length of γ is π which is also the distance between these two points. Extending this geodesic beyond N (Figure 2.10b) makes its length more than π, whereas the distance between S to P is less than π (Figure 2.10c).

Example 2.3.3 (Flat torus). Consider $[0,1] \times [0,1] \subseteq \mathbb{R}^2$. We identify $(x,0)$ with $(x,1)$ and $(0,y)$ with $(1,y)$ where $x, y \in [0,1]$. The obtained quotient space is the flat torus. The metric is naturally induced from the Euclidean metric and hence the geodesics are straight lines. If p be the center $\left(\frac{1}{2}, \frac{1}{2}\right)$, then the cut locus is the wedge of two circles.



(a) Extending γ beyond the blue line fails to be distance minimal



(b) cut locus of p

Figure 2.11: Cut locus of a point in a flat torus

Example 2.3.4 (Real projective planes). We obtain the real projective plane $\mathbb{RP}^n$ by identifying the antipodal points of the round sphere $\mathbb{S}^n$. The metric on $\mathbb{RP}^n$ is induced from the metric on $\mathbb{S}^n$. If $\pi : \mathbb{S}^n \to \mathbb{RP}^n,\ p, -p \mapsto [p]$, then

$$\langle X, Y \rangle_{[p]} =: \left\langle (d\pi_p)^{-1}(X), (d\pi_p)^{-1}(Y) \right\rangle_p$$

is a metric on $\mathbb{RP}^n$. Since the antipodal map is an isometry of $\mathbb{S}^n$, the map π is a local isometry. Let $[p] \in \mathbb{RP}^n$ such that it is the image of north and south pole under the map π. Then the image of the equator of $\mathbb{S}^n$ under the quotient map π, $\mathbb{RP}^{n-1}$, is the cut locus of $[p]$. We will see a generalization of similar result in Chapter 6.

Example 2.3.5 (Cut locus a point in cylinder). Note that for a given point p if more than one distance minimal geodesic joining p and q exists, then q is a cut point. Using this, we observed that the cut locus of a point in cylinder is a line (shown in Figure 2.12). We also note that the point $-p$ is the closest point and there exists a closed geodesic passing through $-p$ starting and ending at p. This fact is more generally true (see Theorem 2.3.6). Generalizing this example, the cut locus of a point $(\mathbf{p}, \mathbf{v}) \in \mathbb{S}^n \times \mathbb{R}^m$ with the product metric is $\{-\mathbf{p}\} \times \mathbb{R}^m$.

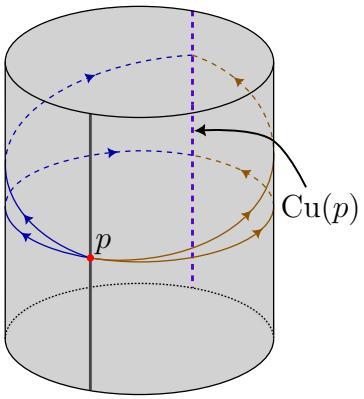


Figure 2.12: Cut locus of a point in a cylinder

2.3.2 Conjugate locus of a point

Let M be a Riemannian manifold and γ be any curve defined on $[a, b]$. A *variation of* γ is a function $\Gamma : [a, b] \times (-\varepsilon, \varepsilon) \to M$ such that $\Gamma(t, 0) = \gamma(t)$. So Γ is a one-parameter family of curves $\gamma_s(t) := \Gamma(t, s)$. If each of γ_s is a geodesic, then we call it is a *geodesic variation* . In this section by variation we mean the geodesic variation. The variation field $\dfrac{\partial \Gamma}{\partial s}(t, 0)$ is called a *Jacobi field* and we will denote it by $J(t)$.

Let γ be a geodesic. We say that a point $q \in M$ on γ is *conjugate to* $p \in M$ if we can find a variation γ_s of γ such that $\gamma_s(0) = p$ for $s \in (-\varepsilon, \varepsilon)$ and each of the geodesic γ_s meet *infinitesimally* at q. That is, if $\gamma(t_0) = q$, then

$$\left.\frac{\partial \gamma_s}{\partial t}\right|_{(t,s)=(0,0)} = 0 = \left.\frac{\partial \gamma_s}{\partial t}\right|_{(t,s)=(t_0,0)}.$$

The conjugate points can be defined in two more equivalent ways. One of them uses the Jacobi field and other uses the exponential map. Recall that a Jacobi field along a geodesic γ satisfies

$$\nabla_{\dot\gamma}\nabla_{\dot\gamma} J + R\left(\dot\gamma, J\right)\dot\gamma = 0,$$

where R is the Riemann curvature tensor.

Definition 2.3.3 (Conjugate points in terms of Jacobi fields). A point p is said to be *conjugate to q along a geodesic* γ if there exists a non-vanishing Jacobi field along γ which vanishes at p and q.

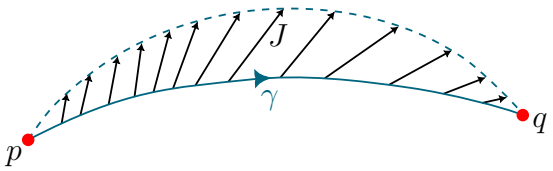


Figure 2.13: p and q are conjugate to each other along γ

Definition 2.3.4 (Conjugate points in terms of exponential map). For a point $p \in M$ we say that $\mathbf{v} \in T_pM$ is a *tangent conjugate point of* p if the derivative of the exponential map is singular at $\mathbf{v}$ i.e., $\det\left(d\left(\exp_p\right)_{\mathbf{v}}\right) = 0$. The point $q := \exp_p(\mathbf{v})$ is said to be *conjugate point of* p along the geodesic $\gamma(t) = \exp_p(t\mathbf{v})$.

For a proof of equivalence of these three definitions we refer the reader to books on Riemannian geometry, for example see [do Carmo, 1992].

The *multiplicity of a conjugate point* is defined to be the nullity of $d(\exp_p)_{\mathbf{v}}$. If nullity is one, then we say it is first conjugate point.

Example 2.3.6. Let $M = \mathbb{R}^n$ with the Euclidean metric, then there are no conjugate points along any geodesic.

Example 2.3.7. If $M = \mathbb{S}^n$ with the round metric, then any antipodal points are conjugate to each other along any great circles. In particular, south pole and north poles are conjugate to each other.

Example 2.3.8. If M is a flat torus, then there are no conjugate points along any geodesic. Recall that metric is flat if and only if the Riemann curvature tensor R vanishes, which implies the Jacobi field is affine and vanishes at two points forced it to be zero everywhere. This example, in particular, proves that if M is flat, then there are no conjugate points along any geodesic.

Example 2.3.9. Let M be the real projective plane, $\mathbb{RP}^n$, with the metric induced from $\mathbb{S}^n$. Here any point p is conjugate to itself along any geodesic.

2.3.3 Some results involving cut and conjugate locus

We will present some results related to the two concepts. As most of the results are standard, we will not provide proofs. Instead, we will mention references for each.

The following result is one of the most important characterization of cut locus in terms of first conjugate point.

Theorem 2.3.1. [Sakai, 1996, Chapter 3, Proposition 4.1] Let γ be a unit speed geodesic. Then $q = \gamma(t_0)$ is a cut point of $p = \gamma(0)$ along γ if either of the following holds.

(i) The point $q = \gamma(t_0)$ is the first conjugate point of p along γ.

(ii) There exists at least two distance minimal geodesic joining p to q.

The next result is also a relation between the two loci. In particular, it states that the cut point of p always comes before (if not the same) the conjugate point.

Theorem 2.3.2. [Kobayashi, 1967, Theorem 4.1] Let γ be a unit speed geodesic starting at p. Let $q = \gamma(t_0)$ be the first conjugate point along γ. Then γ is not a distance minimal geodesic beyond q.

There is one more characterization of the cut locus in terms of number of geodesics joining the point to the cut point. For $p \in M$ we define the set $\mathrm{Se}(p)$ as

$$\mathrm{Se}(p) := \left\{ q \in M \;\middle|\; \begin{array}{l} \text{there exists at least two distance} \\ \text{minimal geodesics joining } p \text{ to } q \end{array} \right\} \tag{2.5}$$

Note that $\mathrm{Se}(p) \subseteq \mathrm{Cu}(p)$. Franz-Erich Wolter in 1979 showed that the closure of $\mathrm{Se}(p)$ is the cut locus.

Theorem 2.3.3. [Wolter, 1979, Theorem 1] Let M be a complete Riemannian manifold and p be any point in M. Then

$$\overline{\mathrm{Se}(p)} = \mathrm{Cu}(p).$$

The above theorem, in particular, shows that the cut locus of a point is a closed set. He also proved that the distance squared function from the point p is not differentiable on the set Se.

Theorem 2.3.4. [Wolter, 1979, Lemma 1] Let $d^2(p,\cdot)$ denotes the square of the distance from the point p. Let $q \in \mathrm{Se}(p)$ and let γ_1 and γ_2 be two distance minimal geodesics joining p to q. Then the directional derivative of $d^2(p,\cdot)$ does not exist at q in the direction of $\gamma_i,\ i = 1, 2$.

For some special point of the cut locus of p we can improve Theorem 2.3.1.

Theorem 2.3.5. [Kobayashi, 1967, Theorem 4.4] Let q be a cut point of p and we assume that it is the closest point of p. Then q is either conjugate to p along a minimal geodesic joining these two points, or q is the mid-point of a closed geodesic starting and ending at p.

We can even make the above theorem sharper if we provide an additional condition.

Theorem 2.3.6. [Kobayashi, 1967, Theorem 4.5] Let p be any point in M such that $d(p, \mathrm{Cu}(p))$ is the smallest and q be any cut point closest to p. Then either q is conjugate to p with respect to a distance minimal geodesic joining p and q or q is the mid-point of a closed geodesic starting and ending smoothly at p.

Chapter 3

Cut locus of a submanifold

Contents

The cut locus of a point plays an important role in analyzing the local structure of a Riemannian manifold M. In the last chapter we had studied cut locus of a point, conjugate locus and some of their properties. Similarly, one can ask about the notion of cut locus for a non-empty subset of M. In order to give a similar definition, we need to first define distance minimal geodesics joining a point $p \in M$ to a subset $N \subseteq M$. If $q \in N$ such that the geodesic is distance minimal geodesic joining p and q, and it minimizes the distance between the set N and the point p then we call such a geodesic a distance minimal geodesic. Now the cut locus of N consists of all points $q \in M$ such that there exists a distance minimal geodesic which fails to be minimal beyond q. Conjugate locus is termed as focal locus if we replace point with submanifold. In this chapter we will study the cut locus of a subset, in particular, a submanifold. We will motivate the results based on some examples and the proofs will be discussed in the subsequent chapters.

3.1 Cut locus of submanifolds

In order to have a definition of the cut locus for a submanifold (or a subset), we need to generalize the notion of a minimal geodesic.

Definition 3.1.1. A geodesic γ is called a *distance minimal geodesic* joining N to p if there exists $q \in N$ such that γ is a minimal geodesic joining q to p and $l(\gamma) = d(p, N)$. We will refer to such geodesics as *N-geodesics*.

If N is an embedded submanifold, then an N-geodesic is necessarily orthogonal to N. This follows from the first variational principle. We are ready to define the cut locus for $N \subset M$.

Definition 3.1.2 (Cut locus of a subset). Let M be a Riemannian manifold and N be any non-empty subset of M. If $\mathrm{Cu}(N)$ denotes the *cut locus of* N, then we say that $q \in \mathrm{Cu}(N)$ if and only if there exists a distance minimal geodesic joining N to q such that any extension of it beyond q is not a distance minimal geodesic.

Example 3.1.1. Let $M = \mathbb{R}^2$ with the Euclidean metric and N be the x-axis. Then the cut locus of N will be empty. If we shoot any geodesic, which are straight lines, perpendicular to x-axis, these will never fail to be distance minimal and hence the cut locus will be empty.

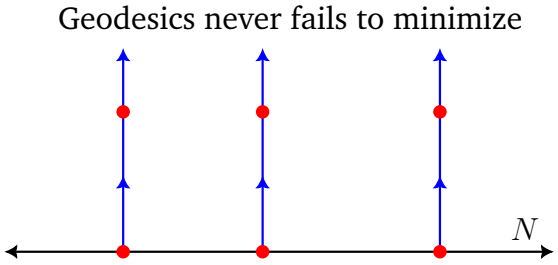


Figure 3.1: Cut locus of x-axis in $\mathbb{R}^2$

Example 3.1.2. Let $M = \mathbb{R}^{n+1}$ with the Euclidean metric and $N = \mathbb{S}^n$. Then the cut locus will be the center of the sphere. Note that if we shoot a minimal geodesic from any point of $\mathbb{S}^n$, then it fails its minimizing property beyond the origin (see Figure 3.2a). Hence, $\mathbf{0}$ is a cut point. To see this is the only cut point, we start with any point $\mathbf{a}$ other than origin. Consider a distance minimal geodesic γ starting at $\mathbb{S}^n$ to $\mathbf{a}$.

$$\gamma(t) = (1-t)\frac{\mathbf{a}}{\|\mathbf{a}\|} + t\mathbf{a},\ 0 \le t \le 1.$$

Note that

$$l(\gamma) = d(P,Q) = |1 - \|\mathbf{a}\||\,.$$

Let $\varepsilon = \frac{\|\mathbf{a}\|}{2}$. Consider the point $R = \gamma(1+\varepsilon)$. We have

$$l(\gamma)\big|_{[0,1+\varepsilon]} = d(P,R).$$

Now we will show that $d(\mathbb{S}^n, R)$ is same as the length of γ. Note that $R = (1+\varepsilon)\mathbf{a} - \frac{\varepsilon\mathbf{a}}{\|\mathbf{a}\|}$ which simplifies to $\frac{\mathbf{a}}{2}(1-\|\mathbf{a}\|)$. Note that

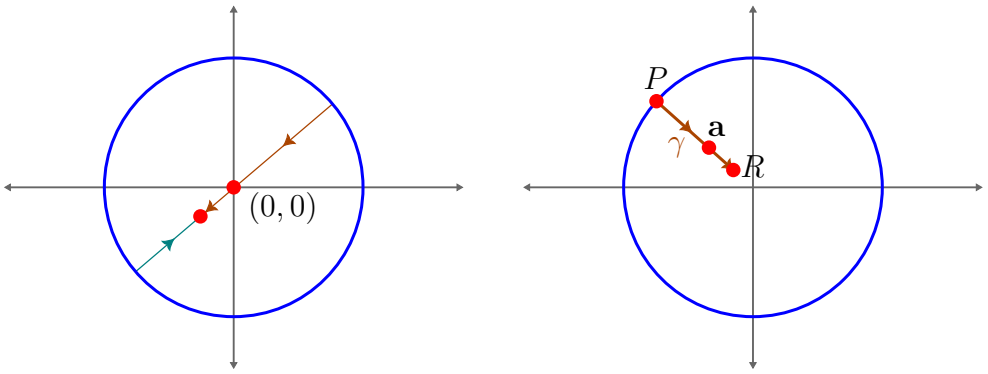
$$\begin{aligned} d^2(\mathbb{S}^n, R) &= \inf_{\mathbf{x}\in\mathbb{S}^n} d^2\left(\mathbf{x}, \frac{\mathbf{a}}{2}(1-\|\mathbf{a}\|)\right) \\ &= \min\left\{\|\mathbf{x}-\mathbf{v}\|^2 : \|\mathbf{x}\|^2 = 1\right\},\ \mathbf{v} = \frac{\mathbf{a}}{2}(1-\|\mathbf{a}\|). \end{aligned}$$

Set

$$f(\mathbf{x}) = \|\mathbf{x}-\mathbf{v}\|^2,\ g(\mathbf{x}) = \|\mathbf{x}\|^2 - 1.$$

So we want to minimize f such that $g(\mathbf{x}) = 0$.

$$\nabla f(\mathbf{x}) - \lambda\nabla g(\mathbf{x}) = 0 \implies (\mathbf{x}-\mathbf{v}) - \lambda\mathbf{x} = 0 \implies \mathbf{x} = \frac{\mathbf{v}}{1-\lambda}.$$



(a) $(0,0)$ is a cut point (b) No other points are cut points

Figure 3.2: Cut locus of $\mathbb{S}^1$ in $\mathbb{R}^2$

Note that the above quantity is well defined as $\mathbf{v} \neq 0$. Now we will use the given constrain,

$$\begin{aligned} \|\mathbf{x}\| = 1 &\implies \|\mathbf{v}\| = |1 - \lambda| \\ &\implies \lambda = 1 \pm \|\mathbf{v}\| \\ &\implies \mathbf{x} = \pm\frac{\mathbf{v}}{\|\mathbf{v}\|} = \pm\frac{\mathbf{a}}{\|\mathbf{a}\|}. \end{aligned}$$

The point $\frac{\mathbf{a}}{\|\mathbf{a}\|}$ corresponds to the minima and note that $P = \frac{\mathbf{a}}{\|\mathbf{a}\|}$. This proves the claim.

Example 3.1.3. Let $M = \mathbb{S}^2$ with the round metric and $N = \mathbb{S}^0 = \{\mathbf{e}_3, -\mathbf{e}_3\}$, where $\mathbf{e}_3 = (0,0,1)$. We claim that the cut locus is the equator, $\{(x,y,0) : x^2 + y^2 = 1\}$. Note that if γ is an N-geodesic (great circle) starting at the North Pole, then it remains distance minimal until it hits the equator circle. As soon as it goes beyond the circle, we can find another geodesic η from the South Pole which is shorter and hence γ is no longer distance minimal, see Figure 3.3.

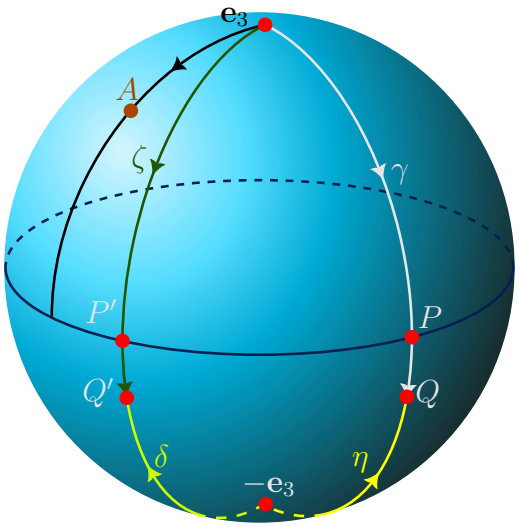


Figure 3.3: Cut locus of $\mathbb{S}^0$ in 2-sphere

Therefore, the equator circle is in the cut locus. We also note that any other point not on the equator is either on the top or bottom hemisphere. In either of the

cases, a minimal geodesic does not fail its distance minimal property beyond the point. So the cut locus is precisely the circle with $z = 0$. By the same argument it follows that the cut locus of $\mathbb{S}^0$ in $\mathbb{S}^n$ is

$$\mathbb{S}^{n-1} = \left\{(x_0, \cdots, x_{n-1}, 0) : x_0^2 + \cdots + x_{n-1}^2 = 1\right\}.$$

Example 3.1.4 (Cut locus of equator in 2-sphere). Let $M = \mathbb{S}^2$ with the round metric and $N = \mathbb{S}^1 = \{(x, y, 0) : x^2 + y^2 = 1\}$. The cut locus of N is $\mathbb{S}^0 = \{\mathbf{e}_3, -\mathbf{e}_3\}$. The argument is similar as above.

Example 3.1.5 (Join induced by cut locus). Let $\mathbb{S}_i^k \hookrightarrow \mathbb{S}^n$ denote the embedding of the k-sphere in the first $k+1$ coordinates while $\mathbb{S}_l^{n-k-1}$ denote the embedding of the $(n-k-1)$-sphere in the last $n-k$ coordinates. It can be seen that $\mathrm{Cu}(\mathbb{S}_i^k) = \mathbb{S}_l^{n-k-1}$. In fact, starting at a point $p \in \mathbb{S}_i^k$ and travelling along a unit speed geodesic in a direction normal to $T_p\mathbb{S}_i^k$, we obtain a cut point at a distance $\pi/2$ from $\mathbb{S}_i^k$.

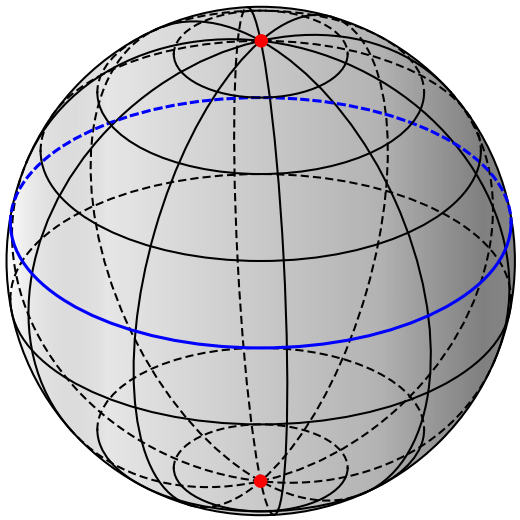


Figure 3.4: The cut locus of the equator in $\mathbb{S}^2$

Moreover, in this case $\mathrm{Cu}(\mathbb{S}_l^{n-k-1}) = \mathbb{S}_i^k$ and the n-sphere $\mathbb{S}^n$ can be expressed as the union of geodesic segments joining $\mathbb{S}_i^k$ to $\mathbb{S}_l^{n-k-1}$. This is a geometric variant of the fact that the n-sphere is the (topological) join of $\mathbb{S}^k$ and $\mathbb{S}^{n-k-1}$. We also observe that $\mathbb{S}^n - \mathbb{S}_l^{n-k-1}$ deforms to $\mathbb{S}_i^k$ while $\mathbb{S}^n - \mathbb{S}_i^k$ deforms to $\mathbb{S}_l^{n-k-1}$.

In our example, let ν_i^{n-k} and ν_l^{k+1} denote the normal bundles of $\mathbb{S}_i^k$ and $\mathbb{S}_l^{n-k-1}$ respectively. We may express $\mathbb{S}^n$ as the union of normal disk bundles $D(\nu_i)$ and $D(\nu_l)$. These disk bundles are trivial and are glued along their common boundary $\mathbb{S}_i^k \times \mathbb{S}_l^{n-k-1}$ to produce $\mathbb{S}^n$. Moreover, $\mathbb{S}_i^k$ is an analytic submanifold of the real analytic Riemannian manifold $\mathbb{S}^n$ with the round metric. There is a generalization of this phenomenon [Omori, 1968, Lemmas 1.3-1.5, Theorem 3.1].

Theorem 3.1.1 (Omori 1968). Let M be a compact, connected, real analytic Riemannian manifold which has an analytic submanifold N such that the cut point of N with respect to every geodesic, which starts from N and whose initial direction is orthogonal to N has a constant distance π from N. Then $N' = \mathrm{Cu}(N)$ is an analytic submanifold and M has a decomposition $M = DN \cup_\varphi DN'$, where DN, DN' are normal disk bundles of N, N' respectively.

3.1.1 Separating set

In all the examples in the previous section, we observed the cut locus of any submanifold N is same as the set of all points which has at least two minimal geodesics joining N to the point. This leads to the following definition.

Definition 3.1.3 (Separating set). Let N be a subset of a Riemannian manifold M. The set $\mathrm{Se}(N)$, called the *separating set*, consists of all points $q \in M$ such that at least two distance minimal geodesics from N to q exist.

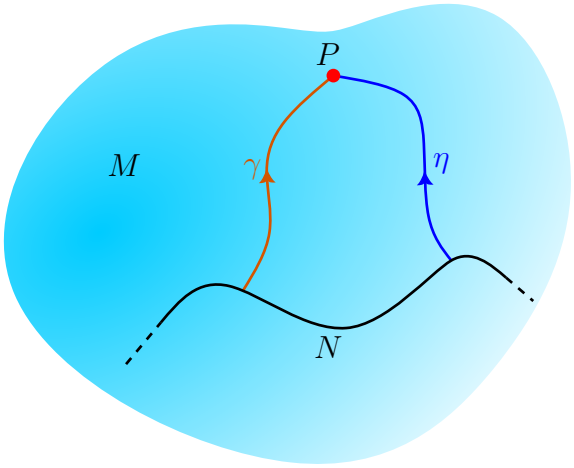


Figure 3.5: Separating set of N

The following example shows that for a given submanifold $N \subseteq M$, the separating set need not be same as the cut locus.

Example 3.1.6 (Cut locus of ellipse). Let $M = \mathbb{R}^2$ with the Euclidean metric and $N = \left\{(x, y) : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\right\}$ for some non-zero real numbers a and b with $a \neq b$. Let $A = (-a, 0)$ and $B = (a, 0)$ be two foci of the ellipse. Note that for any point

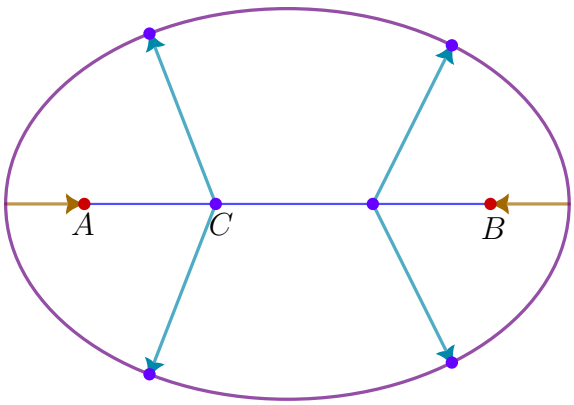


Figure 3.6: $\mathrm{Cu}(N)$ and $\mathrm{Se}(N)$

$C = (x, 0)$ with $x \in (-a, a)$, we have two N-geodesics joining N to C. Hence, all the points are separating point (see Figure 3.6). However, the two foci are not separating points, but they are in the cut locus. So $\mathrm{Se}(N) \neq \mathrm{Cu}(N)$.

Note that $\mathrm{Se}(N) \subset \mathrm{Cu}(N)$. Although the sets $\mathrm{Se}(N)$ and $\mathrm{Cu}(N)$ are not same, in general, we can ask whether including the limit points of $\mathrm{Se}(N)$ make them equal. In the next chapter, we will see that indeed this is the case, and we have $\overline{\mathrm{Se}(N)} = \mathrm{Cu}(N)$ (Theorem 4.2.2). This, in particular, proves that cut locus is a closed set. In general, the cut locus of a subset need not be closed, as illustrated by the following example [Sabau and Tanaka, 2016].

Example 3.1.7 (Sabau-Tanaka 2016). Consider $\mathbb{R}^2$ with the Euclidean inner product. Let $\{\theta_n\}$, with $\theta_1 \in (0, \pi)$, be a decreasing sequence converging to 0. Let $\overline{B(\mathbf{0}, 1)}$ be the closed unit ball centered at $(0, 0)$. Suppose $B_n := B(q_n, 1)$ is the open ball with radius 1 and centered at q_n. We have chosen q_n such that it does not belong to $\overline{B(\mathbf{0}, 1)}$ and denotes the center of the circle passing through $p_n = (\cos\theta_n, \sin\theta_n)$ and $p_{n+1} = (\cos\theta_{n+1}, \sin\theta_{n+1})$. Define $N \subset \mathbb{R}^2$ by

$$N := \overline{B(\mathbf{0}, 1)} \setminus \cup_{n=1}^{\infty} B(q_n, 1).$$

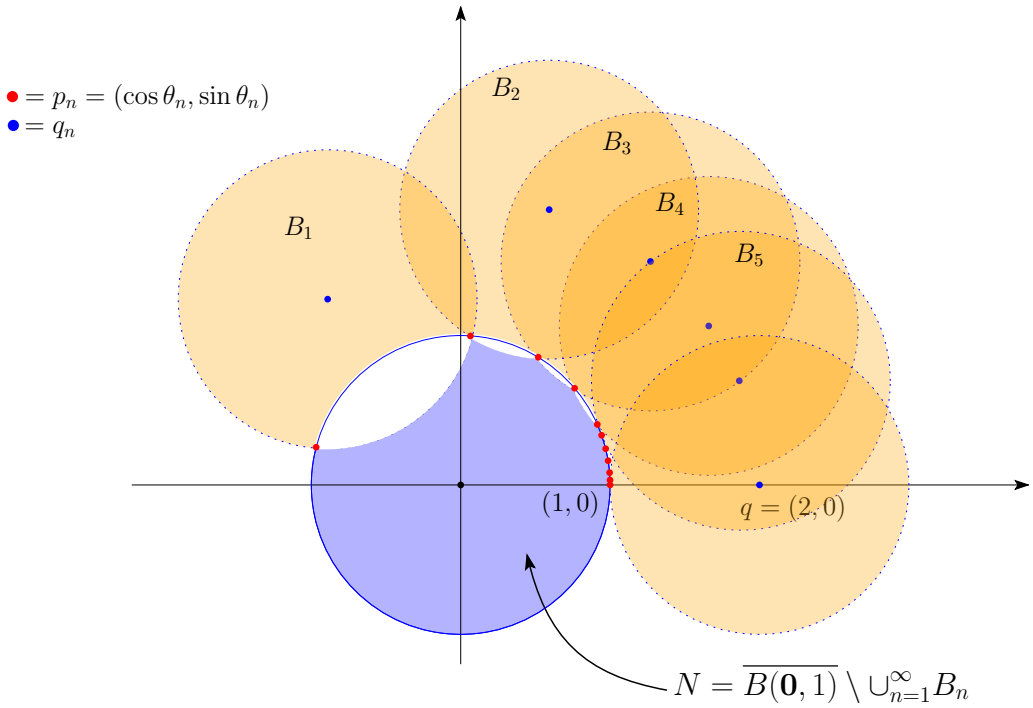


Figure 3.7: Cut locus need not be closed

Note that N is a closed set and the sequence $\{q_n\}$ of cut points of N converges to the point $(2, 0)$. However, $(2, 0)$ is not a cut point of N.

Using the characterization of cut locus in terms of the separating set, we will list some more examples. Most of the justification is provided by the help of pictures.

Example 3.1.8 (Cut locus of k points on the unit circle). Let $M = \mathbb{S}^1$ with the round metric and $N = \{A_1, A_2, A_3\}$. Then the cut locus will be $\{B_{12}, B_{23}, B_{31}\}$, see Figure 3.8. The above example can be generalized for any k-points on $\mathbb{S}^1$. The cut locus of $\{A_1, A_2, \cdots, A_k\}$ will be $\{B_{12}, B_{23}, \cdots, B_{k1}\}$ where B_{ii+1} is the mid-point of A_i and A_{i+1}.

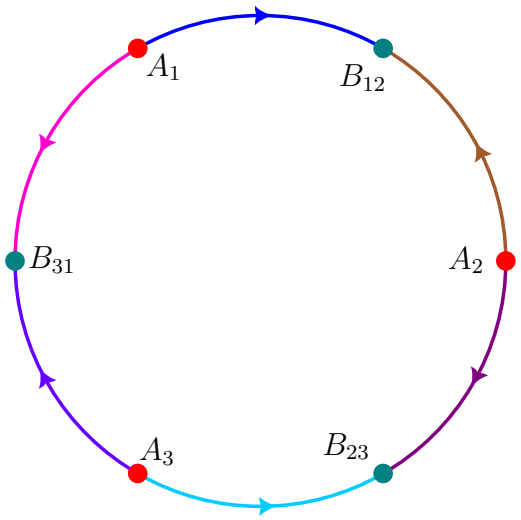
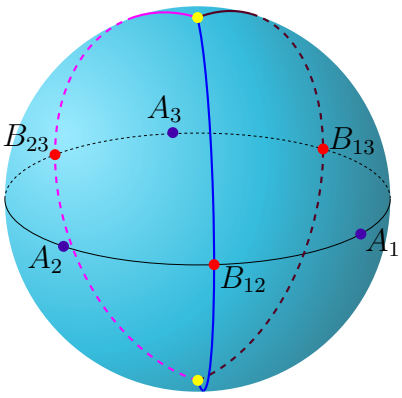
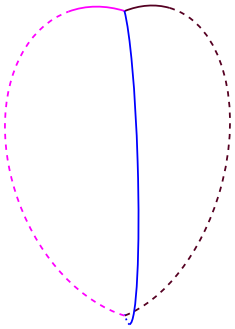


Figure 3.8: Cut locus of three points on unit circle

Example 3.1.9 (Cut locus of k points on $\mathbb{S}^2$). Let A_1, A_2 and A_3 be three points on the equator. The cut locus will be half great circles passing through the mid-points B_{12}, B_{23} and B_{31}, see Figure 3.9a. In fact, all these semicircles are the separating set of $\{A_1, A_2, A_3\}$, being closed the closure is itself. So, the cut locus is homotopic to wedge of two circles. The same can be generalized for k-points on the equator



(a) Cut locus of three points in $\mathbb{S}^2$



(b) Cut locus is homotopic equivalent to $\mathbb{S}^1 \vee \mathbb{S}^1$

Figure 3.9: Cut locus of three points in $\mathbb{S}^2$

of $\mathbb{S}^2$ to conclude that the cut locus is homotopic to $\vee_{k-1}\mathbb{S}^1$. Similarly, one can show that cut locus of k-points in $\mathbb{S}^n$ is homotopic to $\vee_{k-1}\mathbb{S}^{n-1}$. In this example also, the separating set is same as the cut locus as the separating set is closed.

The above example, in particular, shows that cut locus need not be a manifold.

Example 3.1.10. Let $M = \mathbb{R}^2$ with the Euclidean metric and N be the wedge of two circles. The cut locus of N consists of centers of these two circles and the y-axis with origin removed, see Figure 3.10. In fact, if we take any other point then that is not a cut point as any geodesic, a straight line, never fails its distance minimal property. This example, also shows that the cut locus of a subset need not be a closed set.

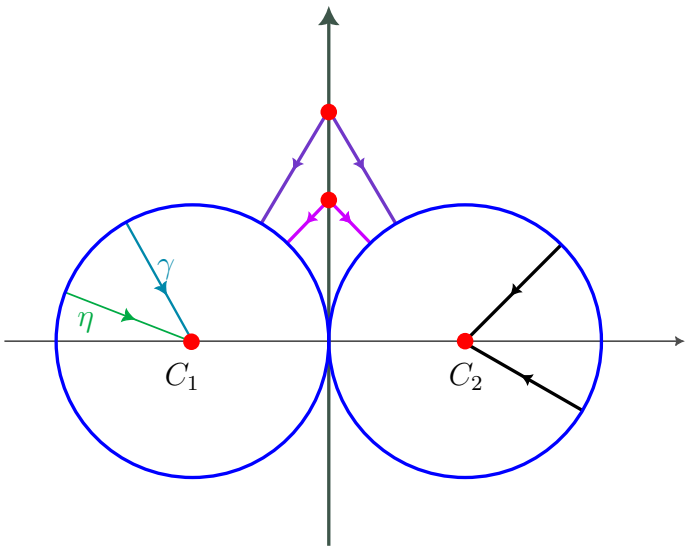


Figure 3.10: Cut locus of wedge of two circles in $\mathbb{R}^2$

Example 3.1.11. Let M be the cylinder $\mathbb{S}^1 \times \mathbb{R}$ with the product metric. Let $N = \{\mathbf{v}\} \times \mathbb{R}$ for some $\mathbf{v} \in \mathbb{S}^1$. Then cut locus of N is $\{-\mathbf{v}\} \times \mathbb{R}$, see Figure 3.11.

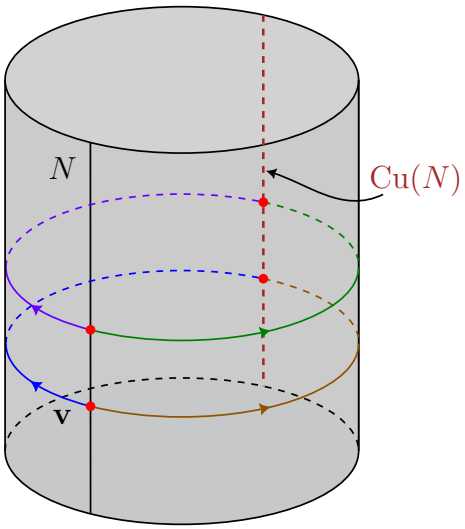


Figure 3.11: Cut locus of a line on cylinder

3.2 An illuminating example

Let $M = M(n, \mathbb{R})$ be the set of $n \times n$ matrices, and $N = O(n, \mathbb{R})$ be the set of all orthogonal $n \times n$ matrices. Let $A, B \in M(n, \mathbb{R})$. We fix the standard flat Euclidean metric on $M(n, \mathbb{R})$ by identifying it with $\mathbb{R}^{n^2}$. This induces a distance function given by

$$d(A, B) := \sqrt{\mathrm{tr}\left((A - B)^T (A - B)\right)}$$

Consider the distance squared function

$$f : GL(n, \mathbb{R}) \to \mathbb{R}, \quad A \mapsto d^2(A, O(n, \mathbb{R})).$$

In order to study this function, we want a closed formula for it.

Lemma 3.2.1. The function f can be explicitly expressed as

$$f(A) = n + \operatorname{tr}\left(A^T A\right) - 2\operatorname{tr}\left(\sqrt{A^T A}\right). \tag{3.1}$$

Proof. Let $A \in GL(n, \mathbb{R})$ be any invertible matrix. Then,

$$\begin{aligned}
f(A) &= \inf_{B \in O(n,\mathbb{R})} d^2(A, B) \\
&= \inf_{B \in O(n,\mathbb{R})} \|A - B\|^2 \\
&= \inf_{B \in O(n,\mathbb{R})} \operatorname{tr}\left((A-B)^T(A-B)\right) \\
&= \inf_{B \in O(n,\mathbb{R})} \operatorname{tr}\left(A^T A - A^T B - B^T A + B^T B\right) \\
&= \inf_{B \in O(n,\mathbb{R})} \left[\operatorname{tr}\left(A^T A\right) - \operatorname{tr}\left(A^T B\right) - \operatorname{tr}\left(B^T A\right) + \operatorname{tr}\left(B^T B\right)\right] \\
&= \operatorname{tr}\left(A^T A\right) + \inf_{B \in O(n,\mathbb{R})} \left[-2\operatorname{tr}\left(A^T B\right)\right] + n \\
&= \operatorname{tr}\left(A^T A\right) - 2 \sup_{B \in O(n,\mathbb{R})} \operatorname{tr}\left(A^T B\right) + n.
\end{aligned} \tag{3.2}$$

The problem of computing $f(A)$ is equivalent to maximizing the function

$$h_A : O(n, \mathbb{R}) \to \mathbb{R}, B \mapsto \operatorname{tr}\left(A^T B\right).$$

Case I: A is a diagonal matrix with positive entries. Then,

$$|h_A(B)| = \left|\operatorname{tr}\left(A^T B\right)\right| = \left|\sum_{i=1}^{n} a_{ii} b_{ii}\right| \leq \sum_{i=1}^{n} |a_{ii} b_{ii}| \leq \sum_{i=1}^{n} a_{ii} = \operatorname{tr}\left(A^T\right) = h_A(I).$$

Thus, one of the maximizer is $B = I$.

Case II: For any non-singular matrix A, we will use the *singular value decomposition* (SVD). Write $A = UDV^T$, where U and V are $n \times n$ orthogonal matrices and D is a diagonal matrix with positive entries. For any $B \in O(n, \mathbb{R})$ using the cyclic property of the trace we have

$$\operatorname{tr}\left(A^T B\right) = \operatorname{tr}\left(VDU^T B\right) = \operatorname{tr}\left(D(U^T B V)\right). \tag{3.3}$$

Since $U^T BV$ is an orthogonal matrix, maximizing over B reduces to the earlier observation that B will be a maximizer if $U^T BV = I$, which implies $B = UV^T$.

Since A is invertible, by the polar decomposition, there exists an orthogonal matrix Q and a symmetric positive definite matrix $S = \sqrt{A^T A}$ such that $A = QS$. As S is symmetric matrix we can diagonalize it, that is, $S = P\tilde{D}P^T$, where $P \in O(n, \mathbb{R})$ and $\tilde{D}$ is a diagonal matrix. Thus,

$$A = QS = QP\tilde{D}P^T.$$

Set $U = QP$, $V = P$ to obtain the SVD of A. In particular, the minimizer is given by

$$B = Q = A\left(\sqrt{A^T A}\right)^{-1}.$$

Therefore,

$$f(A) = n + \operatorname{tr}\left(A^T A\right) - 2\operatorname{tr}\left(\sqrt{A^T A}\right)$$

for invertible matrices.

To find out $f(A)$ for a non-invertible matrix A, we note that $GL(n,\mathbb{R})$ is dense in $M(n,\mathbb{R})$ and that $\sqrt{A^T A}$ is well-defined for $A \in M(n,\mathbb{R})$. The continuity of the map $A \mapsto \sqrt{A^T A}$ on $M(n,\mathbb{R})$ implies that the same formula (3.1) for f applies to A as well. □

In order to understand the differentiability of f, it suffices to analyze the function $A \mapsto \operatorname{tr}\left(\sqrt{A^T A}\right)$.

Lemma 3.2.2. The map $g : M(n,\mathbb{R}) \to \mathbb{R},\ A \mapsto \operatorname{tr}\left(\sqrt{A^T A}\right)$ is differentiable if and only if A is invertible.

Proof. Let A be an invertible matrix. We will prove that the function g is differentiable at A. Let $\mathscr{P}$ be the set of all positive definite matrices which is an open subset of the set of all symmetric matrices $\mathcal{S}$. We will prove that the map

$$r : \mathscr{P} \to \mathscr{P},\ A \mapsto \sqrt{A}$$

is differentiable. Define a function

$$s : \mathscr{P} \to \mathscr{P},\ A \mapsto A^2.$$

We will show that s is a diffeomorphism and from the inverse function theorem r will be differentiable. In order to show that s is a diffeomorphism, we claim that for $A \in \mathscr{P},\ ds_A : T_A\mathscr{P} \to T_{A^2}\mathscr{P}$ is injective. Note that $\mathscr{P}$ is an open subset of a vector space $\mathcal{S}$ and therefore, $T_A\mathscr{P} \cong \mathcal{S} \cong T_{A^2}\mathscr{P}$. So, take $B \in \mathcal{S}$ such that $ds_A(B) = 0$. We will show that $B = 0$. Recall that $ds_A(B) = AB + BA$. Now choose an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ of eigenspace of A and $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ ($\lambda_i > 0$). Then,

$$A(B\mathbf{v}_i) = -BA\mathbf{v}_i = -B\lambda_i\mathbf{v}_i = -\lambda i(B\mathbf{v}_i)$$

which implies Bv_i is also an eigenvector of A with eigenvalue $-\lambda_i < 0$. Hence, $Bv_i = 0$ which implies $B = 0$.
For the converse, we will show that if A is a singular matrix, then the map g is not directional differentiable. Let A be a singular matrix. Using the singular value decomposition, we write

$$A = U \begin{pmatrix} D & 0 \\ 0 & 0_k \end{pmatrix} V^T,$$

where D is a $(n-k) \times (n-k)$ diagonal matrix with positive entries. If

$$B = U \begin{pmatrix} 0_{n-k} & 0 \\ 0 & I_k \end{pmatrix}$$

then we claim that g is not differentiable in the direction of B. Since

$$\sqrt{(A+tB)^T(A+tB)} = V \begin{pmatrix} D & 0 \\ 0 & I_k|t| \end{pmatrix} V^T$$

the limit

$$\lim_{t\to 0}\frac{g(A+tB)-g(A)}{t} = \lim_{t\to 0}\frac{\operatorname{tr}\left(V\begin{pmatrix}D & 0\\ 0 & I_k|t|\end{pmatrix}V^T\right)-\operatorname{tr}\left(V\begin{pmatrix}D & 0\\ 0 & 0_k\end{pmatrix}V^T\right)}{t}$$
$$= k\lim_{t\to 0}\frac{|t|}{t}$$

does not exist and hence the function g is not differentiable. □

Proposition 3.2.1. If A is an invertible matrix, then

$$dg_A(H) = \left\langle A\left(\sqrt{A^TA}\right)^{-1}, H\right\rangle, \tag{3.4}$$

where H is a symmetric matrix of order n.

The following lemma along with chain rule will prove the above proposition.

Lemma 3.2.3. Let A be a positive definite matrix and $\psi : A \mapsto \sqrt{A}$. Then

$$d\psi_A(H) = \int_0^\infty e^{-t\sqrt{A}}He^{-t\sqrt{A}}\,dt,$$

for any symmetric matrix H.

Proof. As $\psi(A)\cdot\psi(A) = A$, differentiating at A we obtain

$$d\psi_A(H)\psi(A) + \psi(A)d\psi_A(H) = H. \tag{3.5}$$

We will show the following:

(i) For any positive definite matrix X and for any symmetric matrix Y the integral

$$\int_0^\infty e^{-tX}Ye^{-tX}\,\mathrm{d}t \tag{3.6}$$

converges. We note that the eigenvalues of e^{-tX} are $e^{-t\lambda_j}$, where λ_j are the eigenvalues of X. Since X is a positive definite matrix, each of the λ_j is positive. Without loss of generality, we assume that $\lambda = \lambda_1$ is the smallest eigenvalue of X. Then we have

$$e^{-t\lambda_j} \le e^{-t\lambda} \implies \left\|e^{-tX}\right\| = e^{-t\lambda}.$$

where $\|\cdot\|$ is the operator norm. Therefore, the operator norm of the integrand in (3.6) is bounded by $2e^{-t\lambda}\|Y\|$, which is an integrable function. Hence, the integral given by (3.6) converges.

(ii) The matrix $d\psi_A(H)$ satisfies (3.5). Observe that

$$\left(\int_0^\infty e^{-t\sqrt{A}}\cdot H\cdot e^{-t\sqrt{A}}\mathrm{d}t\right)\sqrt{A} + \sqrt{A}\left(\int_0^\infty e^{-t\sqrt{A}}\cdot H\cdot e^{-t\sqrt{A}}\,\mathrm{d}t\right)$$

$$= \int_0^\infty \left(e^{-t\sqrt{A}} \cdot H \cdot e^{-t\sqrt{A}} \sqrt{A} + \sqrt{A} e^{-t\sqrt{A}} \cdot H \cdot e^{-t\sqrt{A}} \right) \, \mathrm{d}t$$
$$= \int_0^\infty - \left(e^{-t\sqrt{A}} H e^{-t\sqrt{A}} \right)' \, \mathrm{d}t = H.$$

From (i), (ii) and the uniqueness of the derivative, the lemma is proved. □

We now give a proof of Proposition 3.2.1.

Proof of Proposition 3.2.1. Note that using Lemma 3.2.3, for any symmetric matrix H we have

$$dg_A(H) = \operatorname{tr} \left(\int_0^\infty e^{-t\sqrt{A^T A}} \left(A^T H + H^T A \right) e^{-t\sqrt{A^T A}} \, dt \right). \tag{3.7}$$

Let us simplify the above expression to get the desired result.

$$\begin{aligned}
dg_A(H) &= \int_0^\infty \left[\operatorname{tr} \left(e^{-2t\sqrt{A^T A}} H^T A \right) + \operatorname{tr} \left(e^{-2t\sqrt{A^T A}} A^T H \right) \right] \, dt \\
&= \operatorname{tr} \left(\int_0^\infty e^{-2t\sqrt{A^T A}} H^T A \, dt \right) + \operatorname{tr} \left(\int_0^\infty e^{-2t\sqrt{A^T A}} A^T H \, dt \right) \\
&= \operatorname{tr} \left(\left[\int_0^\infty e^{-2t\sqrt{A^T A}} \, dt \right] H^T A \right) + \operatorname{tr} \left(\left[\int_0^\infty e^{-2t\sqrt{A^T A}} \, dt \right] A^T H \right) \\
&= \operatorname{tr} \left(\left[-\frac{\left(\sqrt{A^T A}\right)^{-1}}{2} \int_0^\infty \frac{d}{dt} e^{-2t\sqrt{A^T A}} \, dt \right] H^T A \right) \\
&\qquad + \operatorname{tr} \left(\left[-\frac{\left(\sqrt{A^T A}\right)^{-1}}{2} \int_0^\infty \frac{d}{dt} e^{-2t\sqrt{A^T A}} \, dt \right] A^T H \right) \\
&= \operatorname{tr} \left(\frac{\left(\sqrt{A^T A}\right)^{-1}}{2} H^T A \right) + \operatorname{tr} \left(\frac{\left(\sqrt{A^T A}\right)^{-1}}{2} A^T H \right) \\
&= \frac{1}{2} \operatorname{tr} \left(\left(\sqrt{A^T A}\right)^{-1} A^T H \right) + \frac{1}{2} \operatorname{tr} \left(A^T H \left(\sqrt{A^T A}\right)^{-1} \right) \\
&= \operatorname{tr} \left(\left(\sqrt{A^T A}\right)^{-1} A^T H \right) = \left\langle A \left(\sqrt{A^T A}\right)^{-1}, H \right\rangle
\end{aligned}$$

Thus,

$$dg_A(H) = \left\langle A \left(\sqrt{A^T A}\right)^{-1}, H \right\rangle.$$

□

For any $A \in GL(n, \mathbb{R})$

$$df_A = 2A - 2A \left(\sqrt{A^T A}\right)^{-1} = -2A \left(\sqrt{A^T A}^{-1} - I \right).$$

Hence, the negative gradient of the function f, restricted to $GL(n,\mathbb{R})$ is given by

$$-\nabla f\big|_A = 2A\left(\sqrt{A^TA}^{-1} - I\right).$$

The critical points are orthogonal matrices. If $\gamma(t)$ is an integral curve of $-\nabla f$ initialized at A, then $\gamma(0) = A$ and

$$\frac{d\gamma}{dt} = -2\gamma(t) + 2\gamma(t)\left(\sqrt{\gamma(t)^T\gamma(t)}\right)^{-1} = -2\gamma(t) + 2\left(\gamma(t)^T\right)^{-1}\sqrt{\gamma(t)^T\gamma(t)}. \quad (3.8)$$

Take the test solution of (3.8) given by

$$\gamma(t) = Ae^{-2t} + (1-e^{-2t})\left(A^T\right)^{-1}\sqrt{A^TA} = Ae^{-2t} + (1-e^{-2t})A\left(\sqrt{A^TA}\right)^{-1}. \quad (3.9)$$

In order to show that $\gamma(t)$ satisfies (3.8), note that

$$\begin{aligned}
\gamma(t)^T\gamma(t) &= \left[Ae^{-2t} + (1-e^{-2t})\left(A^T\right)^{-1}\sqrt{A^TA}\right]^T \\
&\qquad \left[Ae^{-2t} + (1-e^{-2t})\left(A^T\right)^{-1}\sqrt{A^TA}\right] \\
&= \left[A^Te^{-2t} + (1-e^{-2t})\sqrt{A^TA}A^{-1}\right] \\
&\qquad \left[Ae^{-2t} + (1-e^{-2t})\left(A^T\right)^{-1}\sqrt{A^TA}\right] \\
&= A^TAe^{-4t} + 2e^{-2t}(1-e^{-2t})\sqrt{A^TA} \\
&\qquad + (1-e^{-2t})^2\left(\sqrt{A^TA}A^{-1}\left(A^T\right)^{-1}\sqrt{A^TA}\right) \\
&= A^TAe^{-4t} + 2e^{-2t}(1-e^{-2t})\sqrt{A^TA} \\
&\qquad + (1-e^{-2t})^2\left(\sqrt{A^TA}\left(A^TA\right)^{-1}\sqrt{A^TA}\right) \\
&= A^TAe^{-4t} + 2e^{-2t}(1-e^{-2t})\sqrt{A^TA} \\
&\qquad + (1-e^{-2t})^2\left(\sqrt{A^TA}\left(\sqrt{A^TA}\right)^{-1}\left(\sqrt{A^TA}\right)^{-1}\sqrt{A^TA}\right) \\
&= A^TAe^{-4t} + 2e^{-2t}(1-e^{-2t})\sqrt{A^TA} + \left(1-e^{-2t}\right)^2 I \\
&= \left(\sqrt{A^TA}\,e^{-2t} + \left(1-e^{-2t}\right)I\right)^2.
\end{aligned}$$

Thus,

$$\gamma(t)^T\gamma(t) = \left(\sqrt{A^TA}\,e^{-2t} + \left(1-e^{-2t}\right)I\right)^2$$

and hence

$$\left(\sqrt{\gamma(t)^T\gamma(t)}\right)^T = \left(\sqrt{A^TA}A^{-1}\gamma(t)\right)^T = \gamma(t)^T\left(A^T\right)^{-1}\sqrt{A^TA}$$

This implies that

$$\sqrt{\gamma(t)^T\gamma(t)} = \sqrt{A^TA}\left(e^{-2t}I + \left(\sqrt{A^TA}\right)^{-1}\left(1-e^{-2t}\right)\right)$$

$$\begin{aligned}
&\implies \sqrt{\gamma(t)^t\gamma(t)} = \sqrt{A^TA}A^{-1}\gamma(t) \\
&\implies \left(\sqrt{\gamma(t)^T\gamma(t)}\right)^T = \left(\sqrt{A^TA}A^{-1}\gamma(t)\right)^T = \gamma(t)^T\left(A^T\right)^{-1}\sqrt{A^TA} \\
&\implies \left(\gamma(t)^T\right)^{-1}\sqrt{\gamma(t)^T\gamma} = \left(A^T\right)^{-1}\sqrt{A^TA}.
\end{aligned}$$

The right hand side of (3.8), with the test solution, can be simplified to

$$-2Ae^{-2t} + 2e^{-2t}\left(A^T\right)^{-1}\sqrt{A^TA}$$

which is the derivative of γ. Thus, $\gamma(t)$, as defined in (3.9), is the required flow line which deforms $GL(n,\mathbb{R})$ to $O(n,\mathbb{R})$. In particular, $GL^+(n,\mathbb{R})$ deforms to $SO(n,\mathbb{R})$ and other component of $GL(n,\mathbb{R})$ deforms to $O(n,\mathbb{R}) \setminus SO(n,\mathbb{R})$. We note, however, that this deformation takes infinite time to perform the retraction.

Remark 3.2.1. A modified curve

$$\eta(t) = A(1-t) + tA\left(\sqrt{A^TA}\right)^{-1} \tag{3.10}$$

with the same image as γ, defines an actual deformation retraction of $GL(n,\mathbb{R})$ to $O(n,\mathbb{R})$. Apart from its origin via the distance function, this is a geometric deformation in the following sense. Given $A \in GL(n,\mathbb{R})$, consider its columns as an ordered basis. This deformation deforms the ordered basis according to the length of the basis vectors and mutual angles between pairs of basis vectors in a geometrically uniform manner. This is in sharp contrast with Gram-Schmidt orthogonalization, also a deformation of $GL(n,\mathbb{R})$ to $O(n,\mathbb{R})$, which is asymmetric as it never changes the direction of the first column, the modified second column only depends on the first two columns and so on.

We now show that f is Morse-Bott. The tangent space $T_IO(n,\mathbb{R})$ consists of skew-symmetric matrices while the normal vectors at I_n are the symmetric matrices. As left translation by an orthogonal matrix is an isometry of $M(n,\mathbb{R})$, normal vectors at $A \in O(n,\mathbb{R})$ are of the form AW for symmetric matrices W. Since

$$df_A(H) = 2\langle A, H\rangle - 2\left\langle A\big(\sqrt{A^TA}\big)^{-1}, H\right\rangle$$

the relevant Hessian is

$$\mathrm{Hess}(f)_A(H,H') = \lim_{t\to 0}\frac{df_{A+tH'}(H) - df_A(H)}{t}$$

with $H = AW, H' = AW'$ and symmetric matrices W, W'. Solving the Hessian expression, we have

$$\mathrm{Hess}(f)_A(H,H') = \lim_{t\to 0}\left(\frac{2\langle A+tH', H\rangle - 2\left\langle (A+tH')\sqrt{(A+tH')(A+tH')^T}^{-1}, H\right\rangle}{t} - \frac{2\langle A, H\rangle - 2\left\langle A\sqrt{A^TA}^{-1}, H\right\rangle}{t}\right)$$

$$= \lim_{t\to 0}\left(\frac{2\langle A,H\rangle + 2t\langle H',H\rangle - 2\left\langle A\sqrt{(A+tH')^T(A+tH')}^{-1},H\right\rangle}{t}\right.$$
$$\left.-\frac{2t\left\langle H'\sqrt{(A+tH')^T(A+tH')}^{-1},H\right\rangle + 2\langle A,H\rangle - 2\left\langle A\sqrt{A^TA}^{-1},H\right\rangle}{t}\right)$$

$$= \lim_{t\to 0}\left(\frac{2t\langle H',H\rangle - 2t\left\langle H'\sqrt{(A+tH')^T(A+tH')}^{-1},H\right\rangle}{t}\right.$$
$$\left.\frac{-2\left\langle A\sqrt{(A+tH')^T(A+tH')}^{-1},H\right\rangle + 2\left\langle A\sqrt{A^TA}^{-1},H\right\rangle}{t}\right)$$

$$= \lim_{t\to 0} 2\cancel{t}\left(\frac{\langle H',H\rangle - \left\langle H'\sqrt{(A+tH')^T(A+tH')}^{-1},H\right\rangle}{\cancel{t}}\right)$$
$$-2\lim_{t\to 0}\left(\frac{\left\langle A\sqrt{(A+tH')^T(A+tH')}^{-1},H\right\rangle - \left\langle A\sqrt{A^TA}^{-1},H\right\rangle}{t}\right)$$

$$= 2\langle H',H\rangle - 2\left\langle H'\cancel{\sqrt{A^TA}^{-1}}^{\,I},H\right\rangle - 2\left\langle A\left(Dg^{-1}\right)_A(H'),H\right\rangle$$

$$= -2\left\langle A\cdot\frac{A^TH'+H'^TA}{2},H\right\rangle = \langle H',H\rangle + \left\langle AH'^TA,H\right\rangle$$
$$= \langle H',H\rangle + \left\langle A(AW')^TA,AW\right\rangle = \langle H',H\rangle + \left\langle AW'^T,AW\right\rangle$$
$$= \operatorname{tr}\left(H'^TH\right) + \operatorname{tr}\left(W'^TW\right) = \operatorname{tr}\left(H'^TH\right) + \operatorname{tr}\left(H'^TH\right) = 2\operatorname{tr}\left(H'^TH\right).$$

Thus, the Hessian is,

$$\operatorname{Hess}(f)_A(H,H') = 2\operatorname{tr}\left(H^TH'\right) = 2\langle H,H'\rangle.$$

Therefore, the Hessian matrix restricted to $(T_AO(n,\mathbb{R}))^\perp$ is $2I_{\frac{n(n+1)}{2}}$. This is a recurring feature of distance squared functions associated to embedded submanifolds (see Proposition 4.1.1).

There is a relationship between the local homology of cut loci and the reduced Čech cohomology of the *link* of a point in the cut locus. This is due to Hebda [Hebda, 1983, Theorem 1.4 and the remark following it].

Definition 3.2.1. Let N be an embedded submanifold of a complete smooth Riemannian manifold M. For each $q \in \operatorname{Cu}(N)$, consider the set $\Lambda(q,N)$ of unit tangent vectors at q so that the associated geodesics realize the distance between q and N. This set is called the *link* of q with respect to N.

The set of points in N obtained by the end points of the geodesics asso-

ciated to $\Lambda(q, N)$ will be called the *equidistant set*, denoted by $\mathrm{Eq}(q, N)$, of q with respect to N.

Since the equidistant set $\mathrm{Eq}(q, N)$, consisting of points which realize the distance $d(q, N)$, is obtained by exponentiating the points in $\Lambda(q, N)$, there is a natural surjection map from $\Lambda(q, N)$ to $\mathrm{Eq}(q, N)$.

Theorem 3.2.1 (Hebda 1983). Let N be a properly embedded submanifold of a complete Riemannian manifold M of dimension n. If $q \in \mathrm{Cu}(N)$ and v is an element of $\Lambda := \Lambda(q, N)$, then for any abelian group G there is an isomorphism

$$\check{H}^i(\Lambda, v; G) \cong H_{n-1-i}(\mathrm{Cu}(N), \mathrm{Cu}(N) - q; G). \tag{3.11}$$

We are interested in computing $\Lambda(A, O(n, \mathbb{R}))$ for singular matrices A. Note that geodesics in $M(n, \mathbb{R})$, initialized at A, are straight lines and any two such geodesics can never meet other than at A. Therefore, there is a natural identification between the link and the equidistant set of A.

Lemma 3.2.4. [Basu and Prasad, 2021, Lemma 2.15] If $A \in M(n, \mathbb{R})$ is singular of rank k, then $\mathrm{Eq}(A, O(n, \mathbb{R}))$ is homeomorphic to $O(n - k, \mathbb{R})$.

Proof. Using the singular value decomposition, we write $A = UDV^T$, where $U, V \in O(n, \mathbb{R})$ and D is a diagonal matrix with entries the eigenvalues of $\sqrt{A^T A}$. If we specify that the diagonal entries of D are arranged in decreasing order, then D is unique. Moreover, as A has rank $k < n$, the first k diagonal entries of D are positive while the last $n - k$ diagonal entries are zero. In order to find the matrices in $O(n, \mathbb{R})$ which realize the distance $d(A, O(n, \mathbb{R}))$, by (3.2), it suffices to find $B \in O(n, \mathbb{R})$ such that

$$\sup_{B \in O(n,\mathbb{R})} \mathrm{tr}\left(A^T B\right) = \sup_{B \in O(n,\mathbb{R})} \mathrm{tr}\left(VDU^T B\right) = \sup_{B \in O(n,\mathbb{R})} \mathrm{tr}\left(DU^T BV\right)$$

is maximized. However, $U^T BV \in O(n, \mathbb{R})$ has orthonormal rows and the specific form of D implies that the maximum is attained if and only if $U^T BV$ has $e_1, \ldots, e_k$ as the first k rows, in order. Therefore, $U^T BV$ is a block orthogonal matrix, with blocks of I_k and $C \in O(n - k, \mathbb{R})$, i.e., $B \in U(I_k \times O(n - k, \mathbb{R}))V^T$. □

Corollary 3.2.1. Let Sing denote the space of singular matrices in $M(n, \mathbb{R})$. If $A \in \mathrm{Sing}$ is of rank $k < n$, then for any abelian group G there is an isomorphism

$$H_{n^2-1-i}(\mathrm{Sing}, \mathrm{Sing} - A; G) \cong \widetilde{H}^i(O(n - k, \mathbb{R}); G). \tag{3.12}$$

Proof. It follows from Lemma 3.2.4 that $\Lambda(A, O(n, \mathbb{R})) \cong O(n - k, \mathbb{R})$ if A has rank k. Since $O(n - k, \mathbb{R})$ is a manifold, Čech and singular cohomology groups are isomorphic. The space Sing is a star-convex set, whence all homotopy and homology groups are that of a point. Applying (3.11) in our case, we obtain an isomorphism

$$H_{n^2-1-i}(\mathrm{Sing}, \mathrm{Sing} - A) \cong \widetilde{H}^i(O(n - k, \mathbb{R}))$$

between reduced cohomology and local homology groups. In particular, the local homology of the cut locus at A detects the rank of A. □

Remark 3.2.2. Note that for a smooth manifold, the relative homology group $H_k(M, M - p)$ does not depend on the point p; it is in fact isomorphic to $H_k(\mathbb{R}^m, \mathbb{R}^m - \mathbf{0})$, where m is the dimension of M. However, the above result shows that $H_{n^2-1-i}(\mathrm{Sing}, \mathrm{Sing} - A; G)$ does depend on A (it depends on the rank of A). This is happening because Sing is not a smooth manifold. It is the zero set of the determinant map $\det : M(n, \mathbb{R}) \to \mathbb{R}$.

For a computation for $\tilde{H}^i(O(n-k, \mathbb{R}); \mathbb{Z})$, we refer the reader to [Hatcher, 2002, §3.D].

Similar computations hold for $U(n, \mathbb{C})$ and singular $n \times n$ complex matrices.

Theorem 3.2.2. Let $M(n, \mathbb{C})$ denotes the set of all $n \times n$ complex matrices and $U(n)$ denotes the set of all $n \times n$ unitary matrices. Then we have

(i) $\mathrm{Cu}(U(n)) = \mathrm{Sing} =$ set of all singular matrices in $M(n, \mathbb{C})$

(ii) If $A \in M(n, \mathbb{C})$ is singular of rank $k < n$, then $\mathrm{Eq}(A, U(n))$ is homeomorphic to $U(n-k)$.

(iii) If $A \in M(n, \mathbb{C})$ is singular of rank $k < n$, then for any abelian group G there is an isomorphism

$$H_{n^2-1-i}(\mathrm{Sing}, \mathrm{Sing} - A; G) \cong \widetilde{H}^i(U(n-k); G).$$

We end this chapter by mentioning some properties of the cut locus and separating set with the help of the listed examples. In the next chapter, we will prove these results.

(P1) For a submanifold N, the cut locus is the closure of the separating set.

(P2) The distance squared function $d^2(N, \cdot)$ from a submanifold N is not differentiable on the separating set $\mathrm{Se}(N)$.

(P3) The distance squared function from N is a Morse-Bott function with N as its critical submanifold.

(P4) The complement of $\mathrm{Cu}(N)$ deformation retracts to N. Also, the complement of N deforms to the cut locus of N.

Chapter 4

Geometric and topological nature of cut locus

Contents

The objective of this chapter is to analyze the geometric and topological properties of cut locus of submanifolds. In particular, we will study relations between the distance squared function from a submanifold, the cut locus of submanifold and Thom space of the normal bundle of the submanifold. We will also prove that the distance squared function is a Morse-Bott function. Results in this chapter are based on joint work with Basu [Basu and Prasad, 2021].

A result due to Wolter [Wolter, 1979, Lemma 1] may be generalized to prove (Lemma 4.1.1) that the distance squared function from a submanifold is not differentiable on the separating set. This result may be well known to experts, but we provide a proof, following Wolter, which is elementary.

4.1 Regularity of distance squared function

Recall Definition 2.1.2, where we have defined the Δ map. The following proposition describes Δ in terms of the distance function on the Riemannian manifold M from the submanifold N.

Proposition 4.1.1. Let U be a neighbourhood of N such that each point in U

admits a unique unit speed N-geodesic. If $p \in U$, then

$$\Delta(p) = \operatorname{dist}(N, p).$$

Proof. Since the expression of Δ is independent of the choice of the Fermi coordinates, we will make a special choice of the Fermi coordinates $(x_1, \cdots, x_n)$. For $p \in U$, choose the unique unit speed N-geodesic γ joining p to N. This geodesic meets N orthogonally at $\gamma(0) = p'$. Choose t_0 such that $\gamma(t_0) = p$. According to

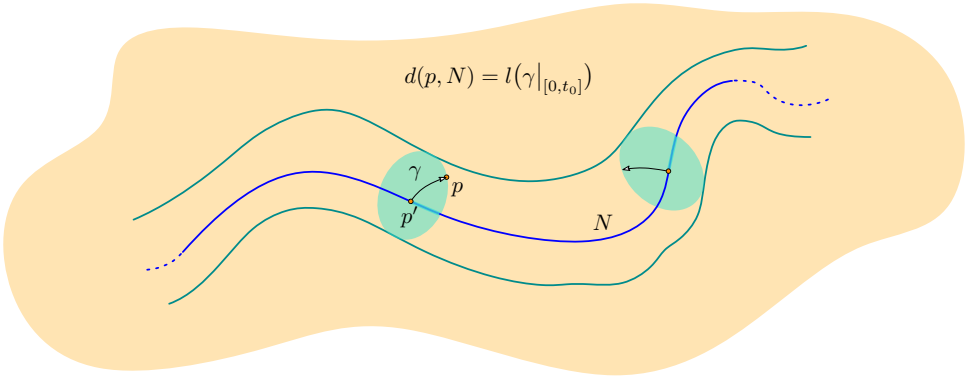


Figure 4.1: Distance via Fermi coordinates

Lemma 2.1.2, there is a system of Fermi coordinates $(x_1, \cdots, x_n)$ centered at p' such that $x_i(\gamma(t)) = t\delta_{i(k+1)}$. The sequence of equalities

$$\Delta(p) = x_{k+1}(\gamma(t_0)) = t_0 = \operatorname{dist}(p, N)$$

complete the proof. □

Corollary 4.1.1. Consider the distance squared function with respect to a submanifold N in M. The Hessian of the distance squared function at the critical submanifold N is non-degenerate in the normal direction.

Towards the regularity of distance squared function, the following observation will be useful. It is a routine generalization of [Wolter, 1979, Lemma 1].

Lemma 4.1.1. [Basu and Prasad, 2021, Lemma 3.7] Let M be a connected, complete Riemannian manifold and N be an embedded submanifold of M. Suppose two N-geodesics exist joining N to $q \in M$. Then $d^2(N, \cdot) : M \to \mathbb{R}$ has no directional derivative at q for vectors in direction of those two N-geodesics.

Proof. Let us assume that all the geodesics are arc-length parametrized. Let $\gamma_i : [0, \hat{t}] \to M, \quad i = 1, 2$, be two distinct geodesics with $\gamma_1(0), \gamma_2(0) \in N$ and $\gamma_1(l) = q = \gamma_2(l)$, where $l = d(N, q)$ and $0 < l < \hat{t}$. Suppose that the two geodesics start at p_1 and p_2 and so $d(p_1, q) = l = d(p_2, q)$. Note that the directional derivative of d^2 at q in the direction of $\gamma_i'(q)$ from the left is given by

$$(d^2)'_-(q) := \lim_{\varepsilon \to 0^+} \frac{(d(N, \gamma_i(l)))^2 - (d(N, \gamma_i(l - \varepsilon)))^2}{\varepsilon}$$

$$
\begin{aligned}
&= \lim_{\varepsilon \to 0} \frac{(d(p_i, \gamma_i(l)))^2 - (d(p_i, \gamma_i(l-\varepsilon)))^2}{\varepsilon} \\
&= \lim_{\varepsilon \to 0^+} \frac{l^2 - (l-\varepsilon)^2}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{l^2 - l^2 + 2l\varepsilon - \varepsilon^2}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{2l\varepsilon - \varepsilon^2}{\varepsilon} \\
&= 2l.
\end{aligned}
$$

Next, we claim that the derivative of the same function from the right is strictly bounded above by $2l$. Let $\omega \in (0, \pi]$ be the angle between the two geodesics γ_1 and γ_2 at q. Define the function,

$$u(\tau) := d(N, \gamma_1(l-\varepsilon)) + d(\gamma_1(l-\varepsilon), \gamma_2(\tau + l)).$$

By triangle inequality, we observe that

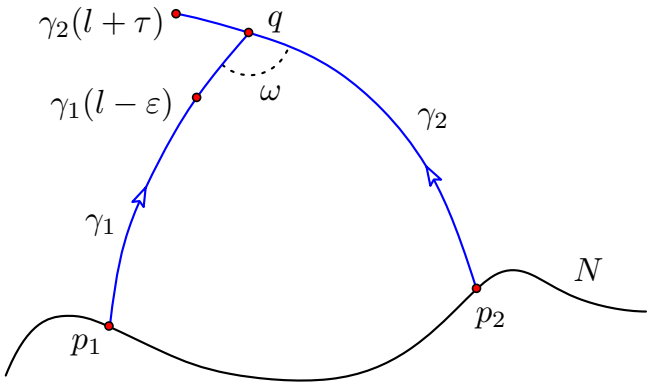


Figure 4.2: When two N-geodesics meet

$$f(\tau) := (u(\tau))^2 \geq d^2(p_1, \gamma_2(\tau + l)) \geq d^2(N, \gamma_2(\tau + l)),$$

and equality holds at $\tau = 0$ and $(u(0))^2 = d^2(N, q) = l^2$. Thus, in order to prove the claim, it suffices to show that the derivative of f from right, at $\tau = 0$, is bounded below by $2l$. We need to invoke a version of the cosine law for small geodesic triangles. Although this may be well-known to experts, we will use the version that appears in [Sharafutdinov, 2007] (also see [Daniilidis et al., 2018, Lemma 2.4] for a detailed proof). In our case, this means that

$$d^2(\gamma_1(l-\varepsilon), \gamma_2(\tau + l)) = \varepsilon^2 + \tau^2 + 2\varepsilon\tau\cos\omega + K(\tau)\varepsilon^2\tau^2$$

where $|K(\tau)|$ is bounded, and the side lengths are sufficiently small. Note that we are considering geodesic triangles with two vertices constant and the varying vertex being $\gamma_2(l + \tau)$. It follows from taking a square root and then expanding in powers of τ that

$$d(\gamma_1(l-\varepsilon), \gamma_2(\tau + l)) = \sqrt{\varepsilon^2 + \tau^2 + 2\varepsilon\tau\cos\omega}\,(1 + O(\tau^2)).$$

It follows that

$$u(\tau) = l - \varepsilon + \sqrt{\varepsilon^2 + \tau^2 + 2\varepsilon\tau\cos\omega}\,(1 + O(\tau^2)).$$

Therefore, $u'_+(0) = \cos\omega = d'_+(\gamma_1(l-\varepsilon), \gamma_2(l))$. Observe that

$$\begin{aligned} f'_+(\tau)\Big|_{\tau=0} &= 2d(N, \gamma_1(l-\varepsilon))d'_+(\gamma_1(l-\varepsilon), \gamma_2(l)) \\ &\qquad + 2d(\gamma_1(l-\varepsilon), \gamma_2(l))d'_+(\gamma_1(l-\varepsilon), \gamma_2(l)) \\ &= 2d(N, \gamma_1(l-\varepsilon))\cos\omega + 2d(\gamma_1(l-\varepsilon), \gamma_2(l))\cos\omega \\ &= 2\cos\omega\big[d(N, \gamma_1(l-\varepsilon)) + d(\gamma_1(l-\varepsilon), \gamma_2(l)))\big] \\ &= 2\cos\omega\big[d(N, \gamma_1(l-\varepsilon)) + d(\gamma_1(l-\varepsilon), \gamma_1(l)))\big] \\ &= 2d(N, \gamma_1(l))\cos\omega < 2l. \end{aligned}$$

Thus, we have proved the claim and subsequently the result. □

The above lemma shows that d^2 is smooth away from the cut locus. The following example suggests that d^2 can be differentiable at points in $\mathrm{Cu}(N) - \mathrm{Se}(N)$ (see Definition 3.1.3) but not twice differentiable.

Example 4.1.1 (Cut locus of an ellipse). We discuss the regularity of the distance squared function from an ellipse $x^2/a^2 + y^2/b^2 = 1$ (with $a > b > 0$) in $\mathbb{R}^2$. For a discussion of the cut locus for ellipses inside $\mathbb{S}^2$ and ellipsoids, see [Hebda, 1995, pages 90-91]. Let (x_0, y_0) be a point inside the ellipse lying in the first quadrant. The point closest to (x_0, y_0) and lying on the ellipse is given by

$$x = \frac{a^2 x_0}{t + a^2}, \;\; y = \frac{b^2 y_0}{t + b^2},$$

where t is the unique root of the quartic

$$\left(\frac{a x_0}{t + a^2}\right)^2 + \left(\frac{b y_0}{t + b^2}\right)^2 = 1$$

in the interval $(-b^2, \infty)$. Given (α, β) with $\beta > 0$, we set $P_\varepsilon(\alpha, \beta) = (\frac{a^2 - b^2}{a} + \varepsilon\alpha, \varepsilon\beta)$; this defines a straight line passing through $P_0(\alpha, \beta)$ in the direction of (α, β). For $\varepsilon > 0$, $P_\varepsilon(\alpha, \beta)$ lies in the first quadrant and we denote by $t = t(\varepsilon)$ be the unique relevant root of the quartic

$$\left(\frac{a(\frac{a^2 - b^2}{a} + \varepsilon\alpha)}{t + a^2}\right)^2 + \left(\frac{b\varepsilon\beta}{t + b^2}\right)^2 = 1.$$

Simplifying this after dividing by ε and taking a limit $\varepsilon \to 0^+$, we obtain

$$\frac{2a\alpha}{a^2 - b^2} = \lim_{\varepsilon \to 0^+} \left(\left(\frac{2}{a^2 - b^2}\right) \frac{t + b^2}{\varepsilon} - b^2\beta^2 \frac{\varepsilon}{(t + b^2)^2} \right).$$

On the other hand, the point $Q_\varepsilon(\alpha, \beta)$ on the ellipse closest to $P_\varepsilon(\alpha, \beta)$ is given by

$$x_\varepsilon = \frac{a^2(\frac{a^2 - b^2}{a} + \varepsilon\alpha)}{t + a^2}, \;\; y_\varepsilon = \frac{b^2\varepsilon\beta}{t + b^2}.$$

It follows that

$$d^2_\varepsilon(\alpha, \beta) := d^2(P_\varepsilon, Q_\varepsilon) = \frac{t^2}{a^2}\left(\frac{a^2 - b^2 + a\varepsilon\alpha}{t + a^2}\right)^2 + \frac{t^2}{b^2}\left(\frac{b\varepsilon\beta}{t + b^2}\right)^2 \tag{4.1}$$

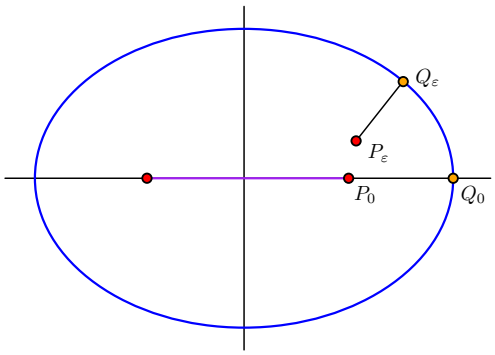


Figure 4.3: Cut locus of an ellipse

Using $t(0) = -b^2$ and simplifications lead us to the following

$$\begin{aligned}\lim_{\varepsilon\to 0^+}\frac{d_\varepsilon^2 - d_0^2}{\varepsilon} &= \frac{2ab^4\alpha}{a^2(a^2-b^2)} - \lim_{\varepsilon\to 0^+}\left(\frac{(t+b^2)(a^2b^2 - a^2t + 2b^2t)}{\varepsilon(t+a^2)^2} - \beta^2\frac{t^2\varepsilon}{(t+b^2)^2}\right)\\ &= \frac{2ab^4\alpha}{a^2(a^2-b^2)} - \frac{2b^2}{a^2-b^2}\lim_{\varepsilon\to 0}\frac{t+b^2}{\varepsilon} + \beta^2b^4\lim_{\varepsilon\to 0}\frac{\varepsilon}{(t+b^2)^2}\\ &= \frac{2ab^4\alpha}{a^2(a^2-b^2)} - \frac{2ab^2\alpha}{a^2-b^2} = -\frac{2b^2\alpha}{a}.\end{aligned}$$

On the other hand, for $\varepsilon < 0$, the point $P_\varepsilon(\alpha,\beta)$ lies in the fourth quadrant. By symmetry, the distance between $P_\varepsilon(\alpha,\beta)$ and $Q_\varepsilon(\alpha,\beta)$ is the same as that between $P_{-\varepsilon}(-\alpha,\beta)$ and $Q_{-\varepsilon}(-\alpha,\beta)$. However, it is seen that

$$d^2(P_{-\varepsilon}(-\alpha,\beta), Q_{-\varepsilon}(-\alpha,\beta)) = d_{-\varepsilon}^2(-\alpha,\beta)$$

as defined in (4.1). Therefore,

$$\begin{aligned}\lim_{\varepsilon\to 0^-}\frac{d^2(P_\varepsilon(\alpha,\beta),Q_\varepsilon(\alpha,\beta)) - d^2(P_0(\alpha,\beta),Q_0(\alpha,\beta))}{\varepsilon} &= \lim_{\varepsilon\to 0^-}\frac{d_{-\varepsilon}^2(-\alpha,\beta) - d_0^2(-\alpha,\beta)}{\varepsilon}\\ &= -\lim_{-\varepsilon\to 0^+}\frac{d_{-\varepsilon}^2(-\alpha,\beta) - d_0^2(-\alpha,\beta)}{-\varepsilon}\\ &= -\frac{2b^2\alpha}{a},\end{aligned}$$

where the last equality follows from the right hand derivative of d^2, as computed previously.

When $\beta = 0$ we would like to compute $d_\varepsilon^2(\alpha,0)$. If $\varepsilon > 0$, then

$$d_\varepsilon^2(\alpha,0) = (b^2/a - \varepsilon\alpha)^2 = \frac{b^4}{a^2} - \frac{2b^2\alpha\varepsilon}{a} + \alpha^2\varepsilon^2. \tag{4.2}$$

On the other hand, if $\varepsilon < 0$ is sufficiently small, then there are two points on the ellipse closest to $P_\varepsilon(\alpha,0) = (\frac{a^2-b^2}{a} + \varepsilon\alpha, 0)$, with exactly one on the first quadrant, say Q_ε. Since the segment $P_\varepsilon Q_\varepsilon$ must be orthogonal to the tangent to the ellipse at Q_ε, we obtain the coordinates for Q_ε:

$$x_\varepsilon = \frac{a^2(\frac{a^2-b^2}{a} + \varepsilon\alpha)}{a^2-b^2},\ y_\varepsilon^2 = b^2\left(1 - \frac{x_\varepsilon^2}{a^2}\right),\ y_\varepsilon > 0.$$

We may compute the distance

$$d_\varepsilon^2(\alpha,0) := (d(P_\varepsilon,Q_\varepsilon))^2 = \frac{b^4}{a^2} - \frac{2b^2\alpha\varepsilon}{a} - \frac{b^2\alpha^2\varepsilon^2}{a^2-b^2}, \tag{4.3}$$

where $\varepsilon < 0$. Combining (4.2) and (4.3) we conclude that d^2 is differentiable at $P_0 = (\frac{a^2-b^2}{a}, 0)$, a point in $\mathrm{Cu}(N)$ but not in $\mathrm{Se}(N)$. However, comparing the quadratic part of d^2 in (4.2),(4.3) we conclude that d^2 is not twice differentiable at P_0.

4.2 Characterizations of $\mathrm{Cu}(N)$

Let (M,g) be a complete Riemannian manifold with distance function d. The exponential map at $p \in M$

$$\exp_p : T_pM \to M$$

is defined on the tangent space of M. Moreover, there exists minimal geodesic joining any two points in M. However, not all geodesics are distance realizing. Given $v \in T_pM$ with $\|v\| = 1$, let γ_v be the geodesic initialized at p with velocity v. Let $S(TM)$ denote the unit tangent bundle and let $[0,\infty]$ be the one point compactification of $[0,\infty)$. Define

$$s : S(TM) \to [0,\infty],\ \ s(v) := \sup\{t \in [0,\infty) \,|\, \gamma_v|_{[0,t]} \text{ is minimal}\}.$$

Definition 4.2.1 (Cut Locus). Let M be a complete, connected Riemannian manifold. If $s(v) < \infty$ for some $v \in S(T_pM)$, then $\exp_p(s(v)v)$ is called a *cut point*. The collection of cut points is defined to be the cut locus of p.

As geodesics are locally distance realizing, $s(v) > 0$ for any $v \in S(TM)$. The following result [Sakai, 1996, Proposition 4.1] will be important for the underlying ideas in its proof.

Proposition 4.2.1. The map $s : S(TM) \to [0,\infty], u \mapsto s(u)$ is continuous.

The proof relies on a characterization of $s(v)$ provided $s(v) < \infty$, (Theorem 2.3.1). A positive real number T is $s(v)$ if and only if $\gamma_v : [0,T]$ is minimal and at least one of the following holds:

(i) $\gamma_v(T)$ is the first conjugate point of p along γ_v,
(ii) there exists $u \in S(T_pM), u \neq v$ such that $\gamma_u(T) = \gamma_v(T)$.

Remark 4.2.1. If M is compact, then it has bounded diameter, which implies that $s(v) < \infty$ for any $v \in S(TM)$. The converse is also true: if M is complete and connected with $s(v) < \infty$ for any $v \in S(TM)$, then M has bounded diameter, whence it is compact.

4.2.1 Characterization in terms of focal points

We shall be concerned with closed Riemannian manifolds in what follows. Let N be an embedded submanifold inside a closed, i.e., compact without boundary, manifold M. Let ν denote the normal bundle of N in M with $D(\nu)$ denoting the unit disk bundle. In the context of $S(\nu)$, the unit normal bundle and the cut locus of N, distance minimal geodesics or N-geodesics are relevant (see Definitions 3.1.1 and 3.1.2). We want to consider

$$\rho : S(\nu) \to [0,\infty),\ \ \rho(v) := \sup\{t \in [0,\infty) \,|\, \gamma_v|_{[0,t]} \text{ is an } N\text{-geodesic}\}. \tag{4.4}$$

Notice that $0 < \rho(v) \leq s(v)$ for any $v \in S(\nu)$. In the special case when $N = \{p\}$, ρ is simply the restriction of s to T_pM. The continuity of ρ requires a result similar to Theorem 2.3.1, which requires the definition of focal points.

Definition 4.2.2 (Focal point). Let $p \in N$ and $(p, v) \in S(\nu)$. We say that v is a *tangent focal point* of N if $d(\exp_\nu)_v$ is not of full rank. If γ is a geodesic from 0 to v in ν_p, then $\exp_\nu(v)$ is called a *focal point of* N *along* $\exp_\nu(\gamma)$.

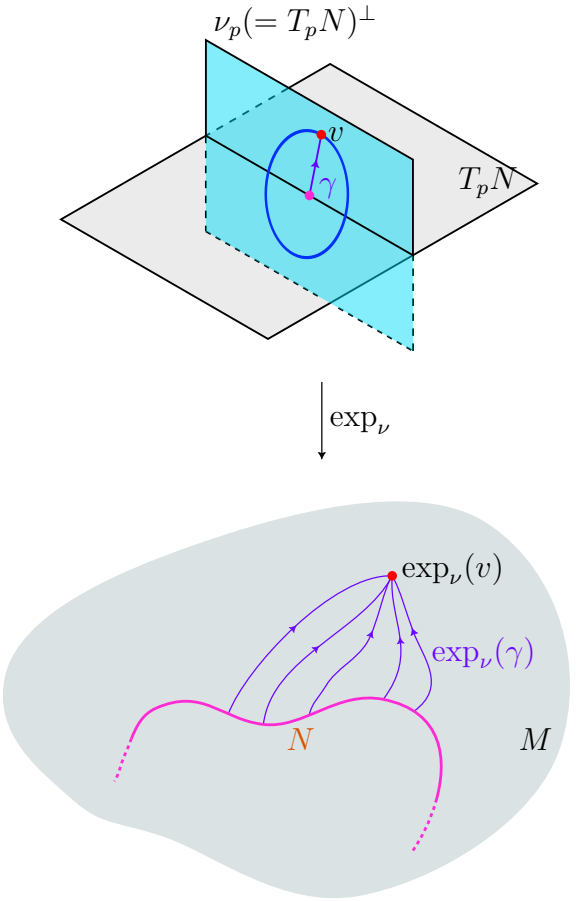


Figure 4.4: Focal points

The nullity of $d\exp_\nu$ at v is called the *multiplicity* of a focal point. If it is one, we say it the *first focal point*. Analogous to Theorem 2.3.1, we have the following result.

Theorem 4.2.1. Let $u \in S_p(\nu)$. A positive real number T is $\rho(u)$ if and only if $\gamma_u : [0, T]$ is an N-geodesic and at least one of the following holds:

(i) $\gamma_u(T)$ is the first focal point of N along γ_u,

(ii) there exists $v \in S(\nu)$ with $v \neq u$ such that $\gamma_v(T) = \gamma_u(T)$.

In order to prove the above theorem, we need the following observations.

Observation A [Sakai, 1996, Lemma 2.11, page 96] *Let N be a submanifold of a Riemannian manifold M and $\gamma : [t_0, \infty) \to M$ a geodesic emanating perpendicularly from N. If $\gamma(t_1)$ is the first focal point of N along γ, then for $t > t_1$, $\gamma|_{[t_0,t_1]}$ cannot be an N-geodesic, i.e., $L\left(\gamma|_{[t_0,t]}\right) > d(N, \gamma(t))$.*

Recall that a sequence $\{\gamma_n\}$ of geodesics, defined on closed intervals, is said to converge to a geodesic γ if $\gamma_n(0) \to \gamma(0)$ and $\gamma_n'(0) \to \gamma'(0)$. It follows from the continuity of the exponential map that if $t_n \to t$, then $\gamma_n(t_n) \to \gamma(t)$.

Observation B *Let γ_n be a sequence of unit speed N-geodesics joining $p_n = \gamma_n(0)$ to $q_n = \gamma_n(t_n)$. If γ_n converges to a geodesic γ and $t_n \to l$, then γ is a unit speed N-geodesic joining $p = \lim_n p_n$ to $q := \gamma(l) = \lim_n \gamma_n(t_n)$.*

Proof. The unit normal bundle $S(\nu)$ is closed. Since $\gamma_n'(0) \to \gamma'(0)$, it follows that $\gamma'(0) \in S(\nu)$. Note that

$$d(N, q) = \lim_{n\to\infty} d(N, q_n) = \lim_{n\to\infty} d(p_n, q_n) = \lim_{n\to\infty} t_n = l = L\left(\gamma|_{[0,l]}\right)$$

implies that γ is an N-geodesic. □

Proof of Theorem 4.2.1. If $\gamma_u(t)$ is the first focal point of N along γ_u, then Observation A implies that γ_u cannot be minimal beyond this value. If (ii) holds, then we need to show that for sufficiently small $\varepsilon > 0$ $\gamma_u|_{[0,T+\varepsilon]}$ is not minimal. Suppose, on the contrary, that γ_u is minimal beyond T. Take a minimal geodesic β joining $\gamma_v(T - \varepsilon)$ to $\gamma_u(T + \varepsilon)$. Observe that,

$$\begin{aligned} 2\varepsilon &= d\left(\gamma_u(T + \varepsilon), \gamma_u(T)\right) + d\left(\gamma_v(T), \gamma_v(T - \varepsilon)\right) \\ &> d\left(\gamma_u(T + \varepsilon), \gamma_v(T - \varepsilon)\right). \end{aligned}$$

If $p, q, r \in M$ such that $d(p, q) + d(q, r) = d(p, r)$ and there exist the shortest normal geodesic γ_1 and γ_2 joining p to q and q to r, respectively, then $\gamma_1 \cup \gamma_2$ is smooth at q and defines a shortest normal geodesic joining p to r. Therefore, we have

$$\begin{aligned} L(\gamma_v|_{[0,T-\varepsilon]} \cup \beta) &= T - \varepsilon + d(\gamma_v(T - \varepsilon), \gamma_u(T + \varepsilon)) \\ &< T + \varepsilon = L(\gamma_u|_{[0,T+\varepsilon]}). \end{aligned}$$

This contradiction establishes that $\gamma_u|_{[0,T+\varepsilon]}$ is not minimal.

For the converse, set $T = \rho(u)$ and observe that $\gamma_u|_{[0,T]}$ is an N-geodesic. Assuming that $q := \gamma_u(T)$ is not the first focal point of N along γ_u, we will prove that (ii) holds. Let $p = \gamma_u(0)$ and choose a neighbourhood $\tilde{U}$ of Tu in ν such that $\exp_\nu|_{\tilde{U}}$ is a diffeomorphism. For sufficiently large n, $q_n := \gamma_u\left(T + 1/n\right) \in \exp_\nu(\tilde{U})$. Take N-geodesics γ_n parametrized by arc-length joining p_n to q_n and set $u_n :=$

$\dot{\gamma}_n(0) \in S((T_{p_n}N)^\perp)$. Since $S((T_{p_n}N)^\perp)$ is compact, by passing to a subsequence, we may assume that u_n converges to $v \in S(N_p)$. By Observation B,

$$\gamma_v(T) = \lim_{n\to\infty} \gamma_{u_n}\left(T + \tfrac{1}{n}\right) = \gamma_u(T).$$

If $v = u$, then for sufficiently large n, $d(p, q_n)u_n \in \tilde{U}$, whence

$$\left(T + \tfrac{1}{n}\right) u = d(p, q_n)u_n.$$

Taking absolute values on both sides imply $T + 1/n > d(p, q_n)$. This contradiction implies $v \neq u$. □

We now will prove that the map ρ defined in 4.4 is a continuous function.

Proposition 4.2.2. The map $\rho : S(\nu) \to [0, \infty)$, as defined in (4.4), is continuous.

Proof. We will prove that $\rho(u_n) \to \rho(u)$ whenever $(p_n, u_n) \to (p, u)$ in the unit normal bundle $S(\nu)$. Let T be any accumulation point of the sequence $\{\rho(u_n)\}$ including ∞. By Observation B, $\gamma_u|_{[0,T]}$ is an N-geodesic and hence $T \leq \rho(u)$. If $T = +\infty$, we are done. So let us assume that $T < +\infty$. From Theorem 4.2.1, at least one of the following holds for infinitely many n.

(i) The sequence $\rho(u_n)$ is the first focal point to N along γ_{u_n}

(ii) there exist $v_n \in S(\nu)$, $v_n \neq u_n$ with $\gamma_{u_n}(\rho(u_n)) = \gamma_{v_n}(\rho(u_n))$.

If (i) is true for infinitely many n, then choose infinitely many unit vectors $\{w_n\}$ which belong to the kernel $\ker(D\exp_\nu(\rho(u_n)u_n))$ and are contained in a compact subset of $S(\nu)$. Choose a convergent subsequence whose limit w is contained in $\ker(D\exp_\nu(Tu))$. Since $w \neq 0$, the rank of $D\exp_\nu(Tu)$ is less than $\dim M$. Thus, $\gamma_u(T)$ is the first focal point of N along γ_u and $T = \rho(u)$.

If (ii) is true for infinitely many n, then we may assume that $v_n \to v \in S(\nu)$. If $v \neq u$, then Theorem 4.2.1 (ii) holds for T, whence $T = \rho(u)$. If $v = u$, we claim that $\gamma_u(T)$ is the first focal point of N along γ_u. If not, then the map $\exp_\nu$ is regular at $Tu \in \nu$ and hence the map

$$\Phi : \nu \to M \times M,\ (p, u) \mapsto (p, \exp_\nu(p, u))$$

is regular at Tu. Therefore, Φ is a diffeomorphism if restricted to an open neighbourhood $\tilde{U}$ of Tu in ν. Since $v = u$, which implies for sufficiently large n, $(p_n, \rho(u_n)u_n)$ and $(p_n, \rho(u_n)v_n)$ belong to $\tilde{U}$ and are different. On the other hand, by assumption $\Phi(\rho(u_n)u_n) = \Phi(\rho(u_n)v_n)$, which is a contradiction. Therefore, $\gamma_u(T)$ is the first focal point and $T = \rho(u)$. □

4.2.2 Characterization in terms of separating set

Recall that the separating set of N, $\mathrm{Se}(N)$, consists of all points $q \in M$ such that at least two distance minimal geodesics from N to q exist. If $q \in \mathrm{Se}(N)$ but $q \notin \mathrm{Cu}(N)$, then we have Figure 4.2, i.e., γ_1 is an N-geodesic beyond q while γ_2 is

another N-geodesic for q. The triangle inequality applied to $\gamma_1(0)$, $q = \gamma_1(l)$ and $\gamma_2(l+\tau)$ implies that

$$d(\gamma_2(l+\tau), N) < l + \tau$$

while for τ small enough $d(\gamma_2(l+\tau), N) = l + \tau$ as γ_2 is an N-geodesic beyond q. This contradiction establishes the well-known fact $\mathrm{Se}(N) \subseteq \mathrm{Cu}(N)$. In quite a few examples, these two sets are equal (see Example 3.1.6 where these two are not same). In the case of $M = \mathbb{S}^n$ with $N = \{p\}$, the set $\mathrm{Se}(N)$ consists of $-p$. There is an infinite family of minimal geodesics joining p to $-p$. An appropriate choice of a pair of such minimal geodesics would create a loop, which is permissible in the definition of $\mathrm{Se}(N)$. According to Theorem 4.2.1, a cut point is either a first focal point of N along a geodesic or it is a separating point. We will now prove our one of the observations in the last chapter, that cut locus of a submanifold N is the closure of separating set of N.

Theorem 4.2.2. Let $\mathrm{Cu}(N)$ be the cut locus of a compact submanifold N of a complete Riemannian manifold M. The subset $\mathrm{Se}(N)$ of $\mathrm{Cu}(N)$ is dense in $\mathrm{Cu}(N)$.

Proof. Let $q \in \mathrm{Cu}(N)$ but not in $\mathrm{Se}(N)$. Choose an N-geodesic γ joining N to q such that any extension of γ is not an N-geodesic. This geodesic γ is unique as $q \notin \mathrm{Se}(N)$. We may write $\gamma(t) = \exp_\nu(tx)$, where $\gamma(0) = p \in N$ and $\gamma'(0) = x_0 \in S(\nu_p)$. It follows from the definition of ρ that $q = \exp_\nu(\rho(x_0)x_0)$. We need to show that every neighborhood of q in $\mathrm{Cu}(N)$ must intersect $\mathrm{Se}(N)$. Suppose it is false. Let $\delta > 0$ and consider $\overline{B(x_0, \delta)}$, the closed ball with center x_0 and radius δ. Define the cone

$$\mathrm{Co}(x_0, \delta) := \left\{ tx : 0 \le t \le 1,\ x \in \overline{B(x_0, \delta)} \cap S(\nu) \right\}.$$

Since $\overline{B(x_0, \delta)} \cap S(\nu)$ is homeomorphic to a closed $(n-1)$-ball for sufficiently small δ, the cone will be homeomorphic to a closed Euclidean n-ball.

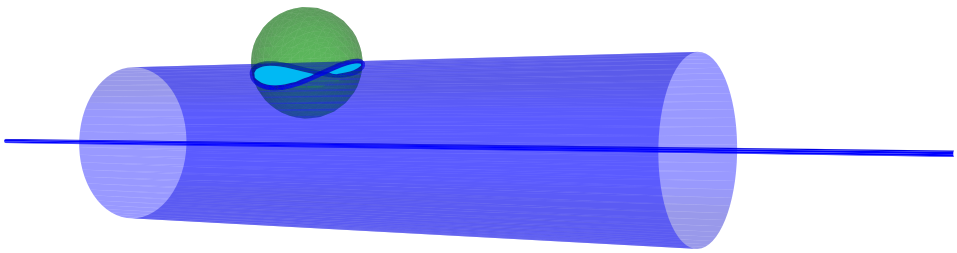


Figure 4.5: $\mathrm{Co}(x_0, \delta)$

Similarly, define another cone

$$\mathrm{Co}^\star(x_0, \delta) := \left\{ \rho\left(x/\left\|x\right\|\right) x \,|\, x \in \mathrm{Co}(x_0, \delta),\ x \neq 0 \right\} \cup \{0\}.$$

Note that $\rho(x_0)$ is finite. As ρ is continuous, due to Proposition 4.2.2, for sufficiently small δ the term $\rho\left(x/\left\|x\right\|\right)$ is still finite, whence $\mathrm{Co}^\star(x_0, \delta)$ is well defined. We claim that $\mathrm{Co}^\star(x_0, \delta)$ is also homeomorphic to a closed Euclidean $(n-k)$-ball. Indeed, a non-zero $x \in \mathrm{Co}(x_0, \delta)$ implies $x = \lambda\hat{x}$, for some $\lambda \in (0, 1]$ and

$\hat{x} \in \overline{B(x_0,\delta)} \cap S(\nu)$. Since $\rho(\hat{x})x = \lambda\rho(\hat{x})\hat{x}$, it follows that $\mathrm{Co}^\star(x_0,\delta)$ is the cone of the set

$$\{\rho(\hat{x})\hat{x} \,|\, \hat{x} \in \overline{B(x_0,\delta)} \cap S(\nu)\},$$

which is homeomorphic to $\overline{B(x_0,\delta)} \cap S(\nu)$. Now we have a dichotomy:

(a) For a fixed small $\delta > 0$, the restriction of $\exp_\nu$ to $\mathrm{Co}^\star(x_0,\delta)$ is a homeomorphism to its image because it is injective, or

(b) For any $\delta > 0$, the restriction of $\exp_\nu$ to $\mathrm{Co}^\star(x_0,\delta)$ is not injective.

If (b) holds, choose $v_n \neq w_n \in \mathrm{Co}^\star(x_0,\frac{1}{n})$ such that these map to q_n under $\exp_\nu$. Thus, $q_n \in \mathrm{Se}(N)$ and compactness of $S(\nu)$ ensures that q_n converges to q. If (a) holds, then let $B(q,\varepsilon)$ denote the open ball in M centered at q with radius $\varepsilon > 0$. We claim that it intersects the complement of $\exp_\nu(\mathrm{Co}^\star(x_0,\delta))$ in M. But it is true as $\rho(x_0)x_0$ lies on the boundary of $\mathrm{Co}^\star(x_0,\delta)$ and hence it has a neighborhood in $\mathrm{Co}^\star(x_0,\delta)$ which is homeomorphic to a closed n-dimensional Euclidean half plane. Since $\exp_\nu$ restricted to $\mathrm{Co}^\star(x_0,\delta)$ was a homeomorphism, the open ball $B(q,\varepsilon)$ must intersect the points outside the image of $\exp_\nu(\mathrm{Co}^\star(x_0,\delta))$.

Now take $\varepsilon = \frac{1}{n}$. For each n, there exists $q_n \in B\left(q,\frac{1}{n}\right)$ such that $q_n \notin \exp_\nu(\mathrm{Co}^\star(x_0,\delta))$. Since M is complete, for each point q_n let γ_n be a N-geodesics joining $p_n \in N$ to q_n. We may invoke the following result from Buseman's book [Busemann, 1955, Theorem 5.16, page 24]. Let $\{\gamma_n\}$ be a sequence of rectifiable curves in a finitely compact set X and the lengths $\ell(\gamma_n)$ are bounded. If the initial points p_n of γ_n forms a bounded set, then $\{\gamma_n\}$ contains a subsequence γ_{n_k} which converges uniformly to a rectifiable curve $\tilde{\gamma}$ in X and

$$\ell(\tilde{\gamma}) \leq \liminf \ell\left(\gamma_{n_k}\right)$$

Since $\{p_n\}$ lie in the compact set N, we obtain a rectifiable curve $\tilde{\gamma}$ such that

$$\ell(\tilde{\gamma}) \leq \liminf \ell\left(\gamma_{n_k}\right) = \lim_k \ell(\gamma_{n_k}) = \lim_k d(q_{n_k}, N) = d(q, N).$$

Thus, $\tilde{\gamma}$ is actually an N-geodesic joining $p' = \lim_k p_{n_k}$ to q and the unit tangent vectors $x_{n_k} = \gamma'_{n_k}(0)$ at p_{n_k} converges to the unit tangent vector $\tilde{x} = \tilde{\gamma}'(0)$ at p'. Since x_0 is an interior point of the set $\overline{B(x_0,\delta)} \cap S(\nu)$, any sequence in $S(\nu)$ converging to x_0 must eventually lie in $\mathrm{Co}(x_0,\delta)$. According to our choice, $q_{n_k} \notin \exp_\nu(\mathrm{Co}^\star(x_0,\delta))$ and x_{n_k} all lie outside of $\mathrm{Co}(x_0,\delta)$. Hence $x_0 \neq \tilde{x}$ and $\gamma \neq \tilde{\gamma}$. Thus, there are two distinct N-geodesics γ and $\tilde{\gamma}$ joining N to q, a contradiction to $q \notin \mathrm{Se}(N)$. This completes the proof. □

4.3 Topological properties

In this section we will study the structure of the cut locus and the relation of cut locus to the Thom space. We will also see various applications of this relation.

4.3.1 Structure of the cut locus

On a complete Riemannian manifold it is very difficult to analyze the structure of cut locus of a point or a submanifold. The main problem is that cut locus is not C^1-smooth. For example, the authors in [Gluck and Singer, 1978] showed that cut locus of a point is not always triangulable. Regarding the question of cut loci being triangulable, we recall the result [Buchner, 1977] that the cut locus (of a point) of a real analytic Riemannian manifold (of dimension d) is a simplicial complex of dimension at most $d-1$. It follows, without much changes, that the result holds for cut loci of submanifolds as well. Hence, we attribute the following result to Buchner.

Theorem 4.3.1 (Buchner 1977). Let N be an analytic submanifold of a real analytic manifold M. If M is of dimension d, then the cut locus $\mathrm{Cu}(N)$ is a simplicial complex of dimension at most $d-1$.

The obvious modifications to the proof by Buchner are the following:

(i) Choose ε to be such that there is a unique geodesic from p to q if $d(p,q) < \varepsilon$ and if $d(N,q) < \varepsilon$, then there is a unique N-geodesic to q;

(ii) Consider the set $\Omega_N(t_0, t_1, \ldots, t_k)$, the space of piecewise broken geodesics starting at N, and define $\Omega_N(t_0, t_1, \ldots, t_k)^s$ analogously;

(iii) The map

$$\Omega_N(t_0, t_1, \ldots, t_k)^s \to N \times M \times \cdots \times M,\ \omega \mapsto (\omega(t_0), \omega(t_1), \ldots, \omega(t_k))$$

determines an analytic structure on $\Omega_N(t_0, t_1, \ldots, t_k)^s$.

The remainder of the proof works essentially verbatim.

Remark 4.3.1. As we have seen in Example 3.1.5, the dimension of the cut locus of a k-dimensional submanifold can be of dimension $d-k-1$. However, generically, we may not expect this to be true. In fact, for real analytic knots (except the unknot) in $\mathbb{S}^3$, it is always the case that the cut locus cannot be homotopic to a (connected) 1-dimensional simplicial complex (cf Example 4.3.3).

4.3.2 Thom space via cut locus

Recall that the *Thom space* $\mathrm{Th}(E)$ of a real vector bundle $E \to B$ of rank k is $D(E)/S(E)$, where it is understood that we have chosen a Euclidean metric on E. If B is compact, then the Thom space $\mathrm{Th}(E)$ is the one-point compactification of E. In general, we compactify the fibres and then collapse the section at infinity to a point to obtain $\mathrm{Th}(E)$. Thus, Thom spaces obtained via two different metrics are homeomorphic. We start with a similar exponential map which is obtained by the ρ map, and we named it *rescaled exponential map*.

Definition 4.3.1 (Rescaled Exponential). The *rescaled exponential* or ρ-

exponential map is defined to be

$$\widetilde{\exp} : D(\nu) \to M,\ (p, v) \mapsto \begin{cases} \exp_p(\rho(\hat{v})v) & \text{if } v = \|v\|\hat{v} \neq 0 \\ p & \text{if } v = 0. \end{cases}$$

We are now ready to prove the main result of this section.

Theorem 4.3.2. Let N be an embedded submanifold inside a closed, connected Riemannian manifold M. If ν denotes the normal bundle of N in M, then there is a homeomorphism

$$\widetilde{\exp} : D(\nu)/S(\nu) \xrightarrow{\cong} M/\mathrm{Cu}(N).$$

Proof. It follows from Proposition 4.2.2 that the rescaled exponential is continuous. Moreover, $\widetilde{\exp}$ is surjective and $\widetilde{\exp}(S(\nu)) = \mathrm{Cu}(N)$. If there exists $(p, v) \neq (q, w) \in D(\nu)$ such that

$$\widetilde{\exp}(p, v) = \widetilde{\exp}(q, w) = p',$$

then $d(p', N)$ can be computed in two ways to obtain

$$d(p', N) = \rho(\hat{v})\|v\| = \rho(\hat{w})\|w\|.$$

Thus, $T = d(p', N)$ is a number such that $\gamma_v : [0, T]$ is an N-geodesic and $\gamma_v(T) = \gamma_w(T) = p'$. By Theorem 4.2.1, we conclude that $T = \rho(\hat{v}) = \rho(\hat{w})$, whence $\|v\| = \|w\| = 1$. Therefore, $\widetilde{\exp}$ is injective on the interior of $D(\nu)$.
As $\mathrm{Cu}(N)$ is closed and M is a compact metric space, the quotient space $M/\mathrm{Cu}(N)$ is Hausdorff. As the quotient $D(\nu)/S(\nu)$ is compact, standard topological arguments imply the map induced by the rescaled exponential is a homeomorphism. □

We will now revisit a basic property of Thom space via its connection to the cut locus. It can be seen that

$$\mathrm{Cu}(N_1 \times N_2) = (\mathrm{Cu}(N_1) \times M_2) \cup (M_1 \times \mathrm{Cu}(N_2)) \tag{4.5}$$

for an embedding $N_1 \times N_2$ inside $M_1 \times M_2$. If ν_j is the normal bundle of N_j inside M_j, then Theorem 4.3.2 along with (4.5) implies that

$$\begin{aligned} \mathrm{Th}(\nu_1 \oplus \nu_2) &\cong \frac{M_1 \times M_2}{(M_1 \times \mathrm{Cu}(N_2)) \cup (\mathrm{Cu}(N_1) \times M_2)} \\ &\cong \frac{M_1/\mathrm{Cu}(N_1) \times M_2/\mathrm{Cu}(N_2)}{M_1/\mathrm{Cu}(N_1) \vee M_2/\mathrm{Cu}(N_2)} \\ &\cong \mathrm{Th}(\nu_1) \wedge \mathrm{Th}(\nu_2). \end{aligned}$$

Let $N = N_1 \sqcup N_2$ be a disjoint union of connected manifolds of the same dimension. If $N \hookrightarrow M$, then let ν_j denote the normal bundle of N_j in M. If ν is the normal bundle of N in M, then

$$\mathrm{Th}(\nu) \cong \mathrm{Th}(\nu_1) \vee \mathrm{Th}(\nu_2). \tag{4.6}$$

This implies that

$$M/\mathrm{Cu}(N) \cong M/\mathrm{Cu}(N_1) \vee M/\mathrm{Cu}(N_2).$$

Example 4.3.1. Consider the two circles

$$N_1 = \{(\cos t, \sin t, 0, 0) \,|\, t \in \mathbb{R}\},\ \ N_2 = \{(0, 0, \cos t, \sin t) \,|\, t \in \mathbb{R}\}$$

in $\mathbb{S}^3$. The link $N := N_1 \sqcup N_2$ has linking number 1. We claim that the cut locus

$$\mathrm{Cu}(N) = \left\{ \frac{1}{\sqrt{2}}(\cos s, \sin s, \cos t, \sin t) \,|\, s, t \in \mathbb{R} \right\}$$

is a torus. We will prove the claim by showing that the above set is separating set of N. As it is closed, the claim will follow by using Theorem 4.2.2. Let $P = (a, b, 0, 0) \in N$ with $a^2 + b^2 = 1$. Note that

$$T_P\mathbb{S}^3 \cong T_PN \oplus (T_PN)^\perp \cong \mathrm{span}\{(-b, a, 0, 0)\} \oplus \mathrm{span}\{\mathbf{e}_3, \mathbf{e}_4\},$$

where $\mathbf{e}_3 = (0, 0, 1, 0)$ and $\mathbf{e}_4 = (0, 0, 0, 1)$. We take any unit vector at P which is perpendicular to T_PN, say $\mathbf{v} = (0, 0, \cos\theta, \sin\theta)$. An N-geodesic starting at P in the direction of $\mathbf{v}$ will be

$$\gamma(t) = P\cos t + \mathbf{v}\sin t, \ \ = (a\cos t, b\cos t, \cos\theta \sin t, \sin\theta \sin t), \ \ 0 \le t \le \pi.$$

We have

$$d(\gamma(t), N) = \inf_{X \in N} d(\gamma(t), X) = \min \left\{ \inf_{X \in N_1} d(\gamma(t), X), \inf_{X \in N_2} d(\gamma(t), X), \right\}.$$

Look at Figure 4.6 and note that

$$d(X, \gamma(t)) = \cos^{-1}(X \cdot \gamma(t)).$$

Therefore, the problem of finding the distance of N to $\gamma(t)$ is equivalent to maximizing the dot product $X \cdot \gamma(t)$. Let $X \in N_1$ and $X = (x, y, 0, 0),\ x^2 + y^2 = 1$. Then

$$X \cdot \gamma(t) = ax\cos t + by\cos t.$$

Maximizing the above such that $x^2 + y^2 = 1$ by the method of Lagrange multiplier, we have

$$x = \frac{a\cos t}{|\cos t|}, \text{ and } y = \frac{b\cos t}{|\cos t|}.$$

Note that the above expression is well-defined, as if $\cos t = 0$, then $X \cdot \gamma(t) = 0$. The maximum value of $X \cdot \gamma(t)$ will be $|\cos t|$ and this is achieved by only one point of N_1. Similarly, if we maximize the dot product over N_2, we get the maximum value $|\sin t|$, which is also obtained by a single maxima. Thus, $\gamma(t)$ will be a separating point if and only if

$$|\cos t| = |\sin t| \implies t = \frac{\pi}{4}, \frac{3\pi}{4}.$$

Therefore, the separating points will be

$$\left\{ \frac{1}{\sqrt{2}}(\cos s, \sin s, \cos\theta, \sin\theta) : s, \theta \in \mathbb{R} \right\}.$$

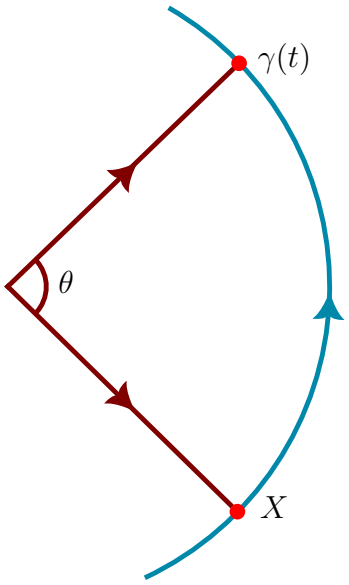


Figure 4.6: Distance of X to $\gamma(t)$

Note that $\mathrm{Cu}(N_1) = N_2$ and vice-versa as well as

$$\mathbb{S}^3/\mathrm{Cu}(N_j) \cong (S^1 \times S^2)/(S^1 \times \infty)$$

where $S^1 \times S^2$ is the fibrewise compatification of the normal bundle of N_j. We conclude that

$$\mathbb{S}^3/\mathrm{Cu}(N) \cong \left(\frac{S^1 \times S^2}{S^1 \times \infty}\right) \vee \left(\frac{S^1 \times S^2}{S^1 \times \infty}\right).$$

There are some topological similarities between $\mathrm{Cu}(N)$ and $M - N$. We recall that a topological pair (X, A) is called a *good pair* if A is closed in X and there is an open subset $U \subseteq X$ with $A \subseteq U$ such that A is a strong deformation retract in U.

Lemma 4.3.1. The cut locus $\mathrm{Cu}(N)$ is a strong deformation retract of $M - N$. In particular, $(M, \mathrm{Cu}(N))$ is a good pair and the number of path components of $\mathrm{Cu}(N)$ equals that of $M - N$.

Proof. Consider the map $H : (M - N) \times [0, 1] \to M - N$ defined via the normal exponential map

$$H(q,t) = \begin{cases} \exp_\nu \left[\left\{ t \cdot \rho\left(\frac{\exp_\nu^{-1}(q)}{\|\exp_\nu^{-1}(q)\|}\right) + (1-t)\,\|\exp_\nu^{-1}(q)\| \right\} \frac{\exp_\nu^{-1}(q)}{\|\exp_\nu^{-1}(q)\|} \right] & \text{if } q \notin \mathrm{Cu}(N) \\ q & \text{if } q \in \mathrm{Cu}(N). \end{cases}$$

If $q \in M - (\mathrm{Cu}(N) \cup N)$, then let γ be the unique N-geodesic joining N to q. The path $H(q,t)$ is the image of this geodesic from q to the first cut point along γ. The continuity of ρ implies that H is continuous. It also satisfies $H(q,0) = q$ and $H(q,1) \in \mathrm{Cu}(N)$. The claims about good pair and path components are clear. □

Remark 4.3.2. The above lemma, in particular, shows that the homotopy type of the cut locus of a submanifold is independent of the choice of the Riemannian metric.

Corollary 4.3.1. If two embeddings $f, g : N \to M$ are ambient isotopic, then $\mathrm{Cu}(f(N))$ and $\mathrm{Cu}(g(N))$ are homotopy equivalent.

Proof. The hypothesis implies that there is a diffeomorphism $\varphi : M \to M$ such that $\varphi(f(N)) = g(N)$. Thus, $M - \mathrm{Cu}(f(N))$ is homeomorphic to $M - \mathrm{Cu}(g(N))$ and the claim follows from the lemma above. Note that in the smooth category, the notion of isotopic and ambient isotopic are equivalent (refer to §8.1 of the book [Hirsch, 1976]). Thus, the same conclusion holds if we assume that the embeddings are isotopic. □

Remark 4.3.3. Without the assumption of M being closed, the above result fails to be true. One may consider $M = S^1 \times \mathbb{R}$ with the natural product metric and $N = S^1$. In fact, the universal cover of M is $\mathbb{R} \times \mathbb{R}$ while that of N is $\mathbb{R}$. If we choose a periodic curve in $\mathbb{R}^2$ which is isotopic to the x-axis and has non-empty cut locus in $\mathbb{R}^2$, then we may pass via the covering map to obtain an embedding g of N isotopic to the embedding f identifying N with $S^1 \times \{0\}$. For this pair, $\mathrm{Cu}(f(N)) = \varnothing$ while $\mathrm{Cu}(g(N)) \neq \varnothing$.

Several other identifications between topological invariants can be explored. For instance, if $\iota : N^k \hookrightarrow M^d$ is as before such that $M - N$ is path connected, then

$$\iota_* : \pi_j(\mathrm{Cu}(N)) \xrightarrow{\cong} \pi_j(M) \tag{4.7}$$

if $0 \leq j \leq d - k - 2$ while ι_* is a surjection for $j = d - k - 1$. The proof of this relies on a general position argument, i.e., being able to find a homotopy of the sphere that avoids N, followed by Lemma 4.3.1. Surjectivity of ι_* if $j \leq d - k - 1$ is imposed by the requirement that a sphere S^j in general position must not intersect N^k. Injectivity of ι for $j \leq d - k - 2$ is imposed by the condition that a homotopy $S^j \times [0, 1]$ in general position must not intersect N^k. This observation (4.7) generalizes a result in [Sakai, 1996, Proposition 4.5 (1)].

The inclusion $i : \mathrm{Cu}(N) \hookrightarrow M$ induces a long exact sequence in homology

$$\cdots \to H_j(\mathrm{Cu}(N)) \xrightarrow{i_*} H_j(M) \to H_j(M, \mathrm{Cu}(N)) \xrightarrow{\partial} H_{j-1}(\mathrm{Cu}(N)) \to \cdots$$

As $(M, \mathrm{Cu}(N))$ is a good pair (cf Lemma 4.3.1), we replace the relative homology of $(M, \mathrm{Cu}(N))$ with reduced homology of $M/\mathrm{Cu}(N) \cong \mathrm{Th}(\nu)$. This results in the following long exact sequence

$$\cdots \to H_j(\mathrm{Cu}(N)) \xrightarrow{i_*} H_j(M) \xrightarrow{q} \widetilde{H}_j(\mathrm{Th}(\nu)) \xrightarrow{\partial} H_{j-1}(\mathrm{Cu}(N)) \to \cdots \tag{4.8}$$

If $N = \{p\}$ is a point, then $\mathrm{Th}(\nu) = S^d$ and (4.8) imply isomorphisms

$$i_* : H_j(\mathrm{Cu}(p)) \xrightarrow{\cong} H_j(M), \; i^* : H^j(M) \xrightarrow{\cong} H^j(\mathrm{Cu}(p))$$

for $j \neq d, d-1$ (cf [Sakai, 1996, Proposition 4.5 (2)]).

Remark 4.3.4. The long exact sequence (4.8) can be interpreted as the dual to the long exact sequence in cohomology of the pair (M, N). If $N = N_1 \sqcup \cdots \sqcup N_l$ is a

disjoint union of submanifolds of dimension $k_1, \ldots, k_l$ respectively, then the Thom isomorphism implies that

$$\widetilde{H}_j(\mathrm{Th}(\nu)) \cong \widetilde{H}_j(\mathrm{Th}(\nu_1)) \oplus \cdots \oplus \widetilde{H}_j(\mathrm{Th}(\nu_l)) \cong H_{j-(d-k_1)}(N_1) \oplus \cdots \oplus H_{j-(d-k_l)}(N_l),$$

where ν_j is the normal bundle of N_j. Applying Poincaré duality to each N_j, we obtain isomorphisms

$$\widetilde{H}_j(\mathrm{Th}(\nu)) \cong \oplus_{i=1}^{l} H^{d-j}(N_i) = H^{d-j}(N).$$

Poincaré-Lefschetz duality applied to the pair (M, N) provides isomorphisms

$$\check{H}^j(M, N) \cong H_{d-j}(M - N). \tag{4.9}$$

As M and N are triangulable, Čech cohomology may be replaced by singular cohomology. Since $M - N$ deforms to $\mathrm{Cu}(N)$ by Lemma 4.3.1, we have isomorphisms

$$H^j(M, N) \cong H_{d-j}(\mathrm{Cu}(N)). \tag{4.10}$$

Combining all these isomorphisms, we obtain the long exact sequence in cohomology for (M, N) from (4.8).

Lemma 4.3.2. Let N be a closed submanifold of M with l components. If M has dimension d, then $H_{d-1}(\mathrm{Cu}(N))$ is free abelian of rank $l - 1$ and $H_{d-j}(\mathrm{Cu}(N)) \cong H^j(M)$ if $j - 2 \geq k$, where k is the maximum of the dimension of the components of N.

Proof. It follows from (4.9) that

$$H_{d-1}(\mathrm{Cu}(N)) \cong H^1(M, N).$$

Consider the long exact sequence associated to the pair (M, N)

$$0 \to H^0(M, N) \to H^0(M) \xrightarrow{i^*} H^0(N) \to H^1(M, N) \to H^1(M) \to H^1(N) \to \cdots$$

If N has l components, i.e., $N = N_1 \sqcup \cdots \sqcup N_l$ where N_j has dimension k_j, then $H^1(M, N)$ is torsion-free. This follows from the fact that $i^*(1) = (1, \ldots, 1)$ and $H^1(M)$ is free abelian.

Remark 4.3.5. In the above Lemma, if in particular, $H^1(M) = 0$, then $H_{d-1}(\mathrm{Cu}(N)) \cong \mathbb{Z}^{l-1}$.

The long exact sequence for the pair (M, N) imply that there are isomorphisms

$$H_{d-j}(\mathrm{Cu}(N)) \cong H^j(M, N) \xrightarrow{\cong} H^j(M) \tag{4.11}$$

if $j \geq k + 2$, where $k = \max\{k_1, \ldots, k_l\}$. □

Remark 4.3.6. Cut locus can be very hard to compute. For a general space, we have the notion of topological dimension. This notion coincides with the usual notion if the space is triangulable. However, in [Barratt and Milnor, 1962], the authors proved that the singular homology of a space may be non-zero beyond its topological dimension. Čech (co)homology is better equipped to detect topological dimension and is the reason why one may prefer it over singular homology due to the generic fractal like nature of cut loci (see the remarks following Theorem C in Chapter 1). Although the topological dimension of $\mathrm{Cu}(N)$ is at most $d-1$, it is not apparent that $H_{d-1}(\mathrm{Cu}(N))$ is a free abelian group.

There are several applications of this discussion.

Theorem 4.3.3. Let N be a smooth homology k-sphere embedded in a Riemannian manifold homeomorphic to S^d. If $d \geq k+3$, then the cut locus $\mathrm{Cu}(N)$ is homotopy equivalent to S^{d-k-1}.

Proof. As N has codimension at least 3, its complement is path-connected. It follows from (4.7) and Lemma 4.3.1 that $M - N$ is $(d-k-2)$-connected. In particular, $M-N$ is simply-connected and by Hurewicz isomorphism, $H_j(M-N) = 0$ if $j \leq d-k-2$. Note that $H_d(M-N) = 0$ as $M-N$ is a non-compact manifold of dimension d.

If $k > 0$, then by Lemma 4.3.2, $H_{d-1}(M-N) = 0$. Moreover, by Poincaré-Lefschetz duality (4.9), we infer that the only non-zero higher homology of $M-N$ is $H_{d-k-1}(M-N) \cong \mathbb{Z}$. By Hurewicz Theorem there is an isomorphism, $\pi_{d-k-1}(M-N) \cong \mathbb{Z}$. Let

$$\alpha : S^{d-k-1} \to M - N$$

be a generator. The map α_* induces an isomorphism on all homology groups between two simply-connected CW complexes. It follows from Whitehead's Theorem that α is a homotopy equivalence. Using Lemma 4.3.1, we obtain our homotopy equivalence $H_1 \circ \alpha : S^{d-k-1} \to \mathrm{Cu}(N)$.

If $k = 0$, then by Lemma 4.3.2, $H_{d-1}(M-N) \cong \mathbb{Z}$. Arguments similar to the $k > 0$ case now applies to obtain a homotopy equivalence with S^{d-1}. □

The above result is foreshadowed by Example 3.1.5 where we showed that the cut locus of $N = \mathbb{S}^k_i$ inside $M = \mathbb{S}^d$ is $\mathbb{S}^{d-k-1}_l$. It also differs from Poincaré-Lefschetz duality in that we are able to detect the exact homotopy type of the cut locus. In fact, when M and N are real analytic and the embedding is also real analytic, then by Theorem 4.3.1 we infer that $\mathrm{Cu}(N)$ is a simplicial complex of dimension at most $d-1$. Towards this direction, Theorem 4.3.3 can be pushed further.

Proposition 4.3.1. Let N be a real analytic homology k-sphere embedded in a real analytic homology d-sphere M. If $d \geq k+3$, then the cut locus $\mathrm{Cu}(N)$ is a simplicial complex of dimension at most $(d-1)$, having the homology of $(d-k-1)$-sphere with fundamental group isomorphic to that of M.

The proof of this is a combination of ideas used in the proof of Theorem 4.3.3. The homotopy type cannot be deduced here due to the presence of a non-trivial fundamental group. An intriguing example can be obtained by combining Proposition 4.3.1 and Poincaré homology sphere.

Example 4.3.2 (Cut locus of 0-sphere in Poincaré sphere). Let $\tilde{I}$ be the binary icosahedral group. It is a double cover of I, the icosahedral group, and can be realized a subgroup of $SU(2)$. It is known that $H_1(\tilde{I};\mathbb{Z}) = H_2(\tilde{I};\mathbb{Z}) = 0$, i.e., it is perfect and the second homology of the classifying space $B\tilde{I}$ is zero. A presentation of $\tilde{I}$ is given by

$$\tilde{I} = \langle s, t \,|\, (st)^2 = s^3 = t^5 \rangle.$$

In fact, if we construct a cell complex X of dimension 2 using the presentation above, then X has one 0-cell, two 1-cells and two 2-cells. The cellular chain complex, as computed from the presentation, is given by

$$0 \longrightarrow \mathbb{Z}^2 \xrightarrow{\begin{pmatrix} -1 & 2 \\ 3 & -5 \end{pmatrix}} \mathbb{Z}^2 \xrightarrow{0} \mathbb{Z} \longrightarrow 0$$

Therefore, $H_1(X) = H_2(X) = 0$ while $\pi_1(X) = \tilde{I}$.

In contrast, consider the cut locus C of the 0-sphere in $SU(2)/\tilde{I}$, the Poincaré homology sphere. As $SU(2)$ is real analytic, so is the homology sphere. By Proposition 4.3.1, C is a finite, connected simplicial complex of dimension 2 such that $\pi_1(C) \cong \tilde{I}$ and $H_\bullet(C;\mathbb{Z}) \cong H_\bullet(S^2;\mathbb{Z})$. The existence of this space is interesting for the following reason: although $X \vee S^2$ has the same topological invariants as C, we are unable to determine whether $X \vee S^2$ is homotopy equivalent to C.

In the codimension two case, we have two results.

Theorem 4.3.4. Let Σ be a closed, orientable, real analytic surface of genus g and N a non-empty, finite subset. Then $\mathrm{Cu}(N)$ is a connected graph, homotopy equivalent to a wedge product of $|N| + 2g - 1$ circles.

Proof. As $\Sigma - N$ is connected, Lemma 4.3.1 implies that $\mathrm{Cu}(N)$ is connected. It follows from Theorem 4.3.1 that $\mathrm{Cu}(N)$ is a finite 1-dimensional simplicial complex, i.e., a finite graph. In this case, $\mathrm{Th}(\nu)$ is a wedge product of $|N|$ copies of S^2 (cf (4.6)). We consider (4.8) with $j = 2$:

$$0 \xrightarrow{i_*} \mathbb{Z} \xrightarrow{q} \widetilde{H}_2(\vee_{|N|} S^2) \xrightarrow{\partial} H_1(\mathrm{Cu}(N)) \xrightarrow{i_*} H_1(\Sigma) \to 0$$

Note that $H_{d-1}(\Sigma)$ is torsion-free, whence all the groups appearing in the long exact sequence are free abelian groups. This implies that

$$\dim_{\mathbb{Z}} H_1(\mathrm{Cu}(N)) = 2g + |N| - 1.$$

As $\mathrm{Cu}(N)$ is connected finite graph, collapsing a maximal tree T results in a quotient space $\mathrm{Cu}(N)/T$ which is homotopic to $\mathrm{Cu}(N)$ as well being a wedge product of $|N| + 2g - 1$ circles. □

Remark 4.3.7. The authors in [Itoh and Vîlcu, 2015] proved that every finite, connected graph can be realized as the cut locus (of a point) of some surface. There

remains the question of orientability of the surface. As noted in the proof of Theorem 4.3.4, if the surface is orientable and $|N| = 1$, then the graph has an even number of generating cycles. If Σ is non-orientable, then $\Sigma \cong (\mathbb{RP}^2)^{\#k}$ has non-orientable genus k and the oriented double cover of Σ has genus $g = k - 1$. Recall that $H_1(\Sigma) \cong \mathbb{Z}^{k-1} \oplus \mathbb{Z}_2$ and $H_2(\Sigma) = 0$. Looking at (4.8) with $j = 2$ we obtain

$$0 \to \mathbb{Z} \to H_1(\mathrm{Cu}(p)) \to \mathbb{Z}^{k-1} \oplus \mathbb{Z}_2 \to 0.$$

Thus, $H_1(\mathrm{Cu}(p)) \cong \mathbb{Z}^k$ as homology of graphs are free abelian. Let $B_\varepsilon(\mathrm{Cu}(p))$ denote the ε-neighbourhood of $\mathrm{Cu}(p)$ in Σ. For ε sufficiently small, this is a surface such that $\overline{B_\varepsilon(\mathrm{Cu}(p))}$ has one boundary component. The compact surface $B_\varepsilon(\mathrm{Cu}(p))$ is reminiscent of ribbon graphs. The surface Σ can be obtained as the connect sum of a disk centered at p and the closure of $B_\varepsilon(\mathrm{Cu}(p))$. Therefore, non-orientability of Σ is equivalent to non-orientability of $B_\varepsilon(\mathrm{Cu}(p))$. A similar observation appears in the unpublished article [Itoh and Vîlcu, 2011, Theorem 3.7].

Example 4.3.3 (Homology spheres of codimension two). In continuation of Theorem 4.3.3, let $N \hookrightarrow S^{k+2}$ be a homology sphere of dimension $k \geq 1$. Since N has codimension two, $S^{k+2} - N$ is path connected and so is $\mathrm{Cu}(N)$. We are not assuming that the metric on S^{k+2} is real analytic. Using (4.10) and the long exact sequence in cohomology of (S^{k+2}, N), we infer that $H_1(\mathrm{Cu}(N)) \cong \mathbb{Z}$ and all higher homology groups vanish. However, the Hurewicz Theorem cannot be used here to establish that $\pi_1(\mathrm{Cu}(N)) \cong \mathbb{Z}$.

In particular cases, we may conclude that $\mathrm{Cu}(N)$ is homotopic to a circle. It was proved in [Plotnick, 1982] that certain homology 3-spheres N, obtained by a Dehn surgery of type $\frac{1}{2a}$ on a knot, smoothly embed in S^5 with complement a homotopy circle. Since $M - N$ deforms to $\mathrm{Cu}(N)$, it follows that there is a map $\alpha : S^1 \to \mathrm{Cu}(N)$ inducing isomorphisms on homotopy and homology groups.

If $k = 1$, then a homology 1-sphere is just a knot K in S^3. Since $S^3 - K$ deforms to $\mathrm{Cu}(K)$, the fundamental group of the cut locus is the knot group. Moreover, in the case of real analytic knots in $\mathbb{S}^3$, the cut locus is a finite simplicial complex of dimension at most 2 (cf Theorem 4.3.1). Except for the unknot, the knot group is never a free group while the fundamental group of a connected, finite graph is free. This observation establishes that $\mathrm{Cu}(K)$ is always a 2-dimensional simplicial complex, whenever K is a non-trivial (real analytic) knot in $\mathbb{S}^3$.

Finally, we will end this chapter by proving that the complement of cut locus deforms to the submanifold.

Theorem 4.3.5. Let N be a closed embedded submanifold of a complete Riemannian manifold M. Let $d : M \to \mathbb{R}$ be the distance function with respect to N. If $f = d^2$, then its restriction to $M - \mathrm{Cu}(N)$ is a Morse-Bott function, with N as the critical submanifold. Moreover, the gradient flow of f deforms $M - \mathrm{Cu}(N)$ to N.

Proof. It follows from Theorem 4.2.1 the map $\exp_\nu^{-1} : M - (\mathrm{Cu}(N) \cup N) \to \nu - \{0\}$ is an (into) diffeomorphism and $\mathrm{dist}(N, q) = \|\exp_\nu^{-1}(q)\|$ and hence the distance function is of class C^∞ at $q \in M - (\mathrm{Cu}(N) \cup N)$. Using Fermi coordinates (cf Proposition 4.1.1), we have seen that the distance squared function is smooth

around N and therefore it is smooth on $M - \mathrm{Cu}(N)$. By Corollary 4.1.1, the Hessian of this function at N is non-degenerate in the normal direction. It is well-known [Sakai, 1996, Proposition 4.8] that $\|\nabla d(q)\| = 1$ if d is differentiable at $q \in M$. Thus, for $q \in M - (\mathrm{Cu}(N) \cup N)$ we have

$$\|\nabla f(q)\| = 2d(q)\|\nabla d(q)\| = 2d(q). \tag{4.12}$$

Let γ be the unique unit speed N-geodesic that joins N to q, i.e.,

$$\gamma : [0, d(q)] \to M,\ \gamma(0) = p,\ \gamma(d(q)) = q,\ \|\gamma'\| = 1.$$

We may write $\nabla f(q) = \lambda\gamma'(d(q)) + w$, where w is orthogonal to $\gamma'(d(q))$. But

$$\left\langle \nabla f\big|_q, \gamma'(d(q)) \right\rangle = \frac{d}{dt} f(\gamma(d(q) + t))\Big|_{t=0} = \frac{d}{dt}(d(q)^2 + 2d(q)t + t^2)\Big|_{t=0} = 2d(q).$$

Thus, $\lambda = 2d(q)$ and combined with (4.12), we conclude that $\nabla f(q) = 2d(q)\gamma'(d(q))$. Therefore, the negative gradient flow line initialized at $q \in M - \mathrm{Cu}(N)$ is given by

$$\eta(t) = \gamma(d(q)e^{-2t}).$$

These flow lines define a flow which deform $M - \mathrm{Cu}(N)$ to N in infinite time. □

The reader may choose to revisit the example of $GL(n, \mathbb{R})$ discussed in Section 3.2 and treat it as a concrete illustration of the Theorem above.

Chapter 5

Application to Lie groups

Contents

Due to the classical results of Cartan, Iwasawa and others, we know that any connected Lie group G is diffeomorphic to the product of a maximally compact subgroup K and the Euclidean space. In particular, G deforms to K. For semisimple groups, this decomposition is stronger and is attributed to Iwasawa. The Killing form on the Lie algebra $\mathfrak{g}$ is non-degenerate and negative definite for compact semi-simple Lie algebras. For such a Lie group G, consider the Levi-Civita connection associated to the bi-invariant metric obtained from negative of the Killing form. This connection coincides with the Cartan connection.

We will consider two examples, both of which are non-compact and non semisimple. We prove that these Lie groups G deformation retract to maximally compact subgroups K via gradient flows of appropriate Morse-Bott functions. This requires a choice of a left-invariant metric which is right-K-invariant, and a careful analysis of the geodesics associated with the metric. In particular, we provide a possibly new proof of the surjectivity of the exponential map for $U(p,q)$. The results of this chapter is based on joint work with Basu [Basu and Prasad, 2021, §4].

5.1 Matrices with positive determinant

Let g be a left-invariant metric on $GL(n,\mathbb{R})$, the set of all invertible matrices. Recall that a left-invariant metric g on a Lie group is determined by its restriction at the identity. For $A \in GL(n,\mathbb{R})$, consider the left multiplication map $l_A : GL(n,\mathbb{R}) \to GL(n,\mathbb{R}),\ B \mapsto AB$. This extends to a linear isomorphism from $M(n,\mathbb{R})$ to itself. Thus, the differential $(Dl_A)_I : T_IGL(n,\mathbb{R}) \to T_AGL(n,\mathbb{R})$ is an isomorphism and given by l_A itself. For $X, Y \in T_IGL(n,\mathbb{R})$,

$$g_I(X,Y) = g_A((Dl_A)_IX, (Dl_A)_IY) = g_A(AX, AY).$$

We choose the left-invariant metric on $GL(n,\mathbb{R})$ generated by the Euclidean metric at I. Therefore,

$$g_{A^{-1}}(X,Y) = \langle AX, AY\rangle_I := \operatorname{tr}\left((AX)^T AY\right) = \operatorname{tr}\left(X^T A^T AY\right).$$

Note that this metric is right-$O(n,\mathbb{R})$-invariant. We are interested in the distance between an invertible matrix A (with $\det(A) > 0$) and $SO(n,\mathbb{R})$. Since $SO(n,\mathbb{R})$ is compact, there exists $B \in SO(n,\mathbb{R})$ such that $d(A,B) = \operatorname{dist}(A, SO(n,\mathbb{R}))$.

Lemma 5.1.1. If D is a diagonal matrix with positive diagonal entries $\lambda_1, \cdots, \lambda_n$, then

$$\operatorname{dist}(D, SO(n,\mathbb{R})) = d(D,I).$$

Moreover, I is the unique minimizer and the associated minimal geodesic is given by $\gamma(t) = e^{t\log D}$.

Proof. Let $B \in SO(n,\mathbb{R})$ satisfying $d(D,B) = \operatorname{dist}(D, SO(n,\mathbb{R}))$. Since with respect to the left-invariant metric $GL^+(n,\mathbb{R})$ is complete, there exists a minimal geodesic $\gamma : [0,1] \to GL^+(n,\mathbb{R})$ joining B to D, i.e.,

$$\gamma(0) = B, \quad \gamma(1) = D, \quad \text{and} \quad l(\gamma) = d(D,B).$$

The first variational principle implies that $\gamma'(0)$ is orthogonal to $T_B SO(n,\mathbb{R})$. It follows from [Martin and Neff, 2016, §2.1] that $\eta(t) = e^{tW}$ is a geodesic if W is a symmetric matrix. Moreover, $\eta'(0) = W$ is orthogonal to $T_I SO(n,\mathbb{R})$. As left translation is an isometry and isometry preserves geodesic, it follows that $\gamma(t) = Be^{tW}$ is a geodesic with $\gamma'(0)$ orthogonal to $T_B SO(n,\mathbb{R})$. By the defining properties of γ, $D = \gamma(1) = Be^W$. Since e^W is symmetric positive definite, we obtain two polar decompositions of D, i.e., $D = ID$ and $D = Be^W$. By the uniqueness of the polar decomposition for invertible matrices, $B = I$ and $D = e^W$.

In order to compute $d(I,D)$, note that

$$e^W = D = e^{\log D},$$

where $\log D$ denotes the diagonal matrix with entries $\log\lambda_1, \cdots, \log\lambda_n$. As W and $\log D$ are symmetric, and matrix exponential is injective on the space of symmetric matrices, we conclude that $W = \log D$. The geodesic is given by $\gamma(t) = e^{t\log D}$ and

$$\operatorname{dist}(D, SO(n,\mathbb{R})) = \|\gamma'(0)\|_I = \|\log D\|_I = \left(\sum_{i=1}^n (\log\lambda_i)^2\right)^{\frac{1}{2}}. \tag{5.1}$$

Thus, the distance squared function will be given by $\sum_{i=1}^n (\log\lambda_i)^2$. □

Now for any $A \in GL^+(n,\mathbb{R})$ we can apply the SVD decomposition, i.e., $A = UDV^T$ with $\sqrt{A^TA} = VDV^T$ and $\log\sqrt{A^TA} = V(\log D)V^T$. Note that $U, V \in SO(n,\mathbb{R})$ and D is a diagonal matrix with positive entries. The left-invariant metric is right-invariant with respect to orthogonal matrices. Thus,

$$\operatorname{dist}(A, SO(n,\mathbb{R})) = \operatorname{dist}(D, SO(n,\mathbb{R})) = \|\log D\|_I,$$

where the last equality follows from the lemma (see (5.1)). As

$$\|\log D\|_I = \left\|V(\log D)V^T\right\|_I = \left\|\log\sqrt{A^TA}\right\|_I,$$

it follows from the arguments of the lemma and the metric being bi-$O(n,\mathbb{R})$-invariant that

$$\gamma(t) = Ue^{t\log D}V^T$$

is a minimal geodesic joining UV^T to A, realizing $\mathrm{dist}(A, SO(n,\mathbb{R}))$. As the minimizer UV^T is unique, $\mathrm{Se}(SO(n,\mathbb{R}))$ is empty, implying that $\mathrm{Cu}(SO(n,\mathbb{R}))$ is empty as well. In fact, $UV^T = A\sqrt{A^TA}^{-1}$ and

$$\gamma(t) = Ue^{t\log D}V^T = UV^TVe^{t\log D}V^T = A\sqrt{A^TA}^{-1}e^{t\log\sqrt{A^TA}}. \tag{5.2}$$

If we compare (3.10), the deformation of $GL(n,\mathbb{R})$ to $O(n,\mathbb{R})$ inside $M(n,\mathbb{R})$, with (5.2), then in both of the cases, an invertible matrix A deforms to $A\sqrt{A^TA}^{-1}$. Finally, observe that the normal bundle of $SO(n,\mathbb{R})$ is diffeomorphic to $GL^+(n,\mathbb{R})$.

5.2 Indefinite unitary groups

Let n be a positive integer with $n = p+q$. Consider the inner product on $\mathbb{C}^n$ given by

$$\langle(w_1,\ldots,w_n),(z_1,\ldots,z_n)\rangle = z_1\overline{w_1}+\cdots+z_p\overline{w_p}-z_{p+1}\overline{w_{p+1}}-\cdots-z_n\overline{w_n}.$$

This is given by the matrix $I_{p,q}$ in the following way:

$$\langle\mathbf{w},\mathbf{z}\rangle = \overline{\mathbf{w}}^tI_{p,q}\mathbf{z} = \begin{pmatrix}\overline{w}_1 & \cdots & \overline{w}_n\end{pmatrix}\begin{pmatrix}I_p & 0\\ 0 & -I_q\end{pmatrix}\begin{pmatrix}z_1\\ \vdots\\ z_n\end{pmatrix}$$

Let $U(p,q)$ denote the subgroup of $GL(n,\mathbb{C})$ preserving this indefinite form, i.e., $\mathscr{A}\in U(p,q)$ if and only if $\mathscr{A}^*I_{p,q}\mathscr{A} = I_{p,q}$. In particular, $\det\mathscr{A}$ is a complex number of unit length. By convention, $I_{n,0} = I_n$ and $I_{0,n} = -I_n$, both of which corresponds to $U(n,0) = U(n) = U(0,n)$, the unitary group. In all other cases, the inner product is indefinite.

The group $U(1,1)$ is given by matrices of the form

$$\mathscr{A} = \begin{pmatrix}\alpha & \beta\\ \lambda\overline{\beta} & \lambda\overline{\alpha}\end{pmatrix},\ \lambda\in S^1,\ |\alpha|^2-|\beta|^2 = 1.$$

More generally, we shall use

$$\mathscr{A} = \begin{pmatrix}A & B\\ C & D\end{pmatrix}$$

to denote an element of $U(p,q)$. It follows from the definition that $\mathscr{A}\in U(p,q)$ if and only if

$$A^*A - C^*C = I_p$$

$$\begin{aligned} A^*B - C^*D &= 0_{p\times q} \\ B^*B - D^*D &= -I_q. \end{aligned}$$

Observe that if $Av = 0$, then

$$0 = A^*Av = C^*Cv + v,$$

which implies that C^*C, a positive semi-definite matrix, has -1 as an eigenvalue unless $v = 0$. Therefore, A is invertible, and the same argument works for D.

Lemma 5.2.1. The intersection of $U(p+q)$ with $U(p,q)$ is $U(p)\times U(q)$. Moreover, if $\mathscr{A} \in U(p,q)$, then $\mathscr{A}^*, \sqrt{\mathscr{A}^*\mathscr{A}} \in U(p,q)$.

Proof. If $\mathscr{A} \in U(p)\times U(q)$, then

$$\begin{aligned} A^*A + C^*C &= I_p \\ B^*B + D^*D &= I_q. \end{aligned}$$

This implies that both B and C are zero matrices. If $\mathscr{A} \in U(p,q)$, then $\mathscr{A}^* = I_{p,q}\mathscr{A}^{-1}I_{p,q}$ and

$$\begin{aligned} (\mathscr{A}^*\mathscr{A})^* I_{p,q}(\mathscr{A}^*\mathscr{A}) &= (\mathscr{A}^*\mathscr{A})I_{p,q}(\mathscr{A}^*\mathscr{A}) \\ &= I_{p,q}\mathscr{A}^{-1}I_{p,q}\mathscr{A}I_{p,q}I_{p,q}\mathscr{A}^{-1}I_{p,q}\mathscr{A} \\ &= I_{p,q} = \mathscr{A}^*I_{p,q}\mathscr{A}. \end{aligned}$$

This also implies that $\mathscr{A}I_{p,q}\mathscr{A}^* = I_{p,q}$.

All the eigenvalues of $\mathscr{A}^*\mathscr{A}$ are positive. Moreover, if λ is an eigenvalue of $\mathscr{A}^*\mathscr{A}$ with eigenvector $\mathbf{v} = (v_1,\ldots,v_p,v_{p+1},\ldots,v_n)$, then

$$I_{p,q}\mathbf{v} = \mathscr{A}^*\mathscr{A}\, I_{p,q}\mathscr{A}^*\mathscr{A}\mathbf{v} = \lambda(\mathscr{A}^*\mathscr{A}\, I_{p,q}\mathbf{v}),$$

which implies that λ^{-1} is also an eigenvalue with eigenvector $\mathbf{v}' = (v_1,\ldots,v_p,-v_{p+1},\ldots,-v_n)$. If $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ is an eigenbasis of $\mathscr{A}^*\mathscr{A}$ with (possibly repeated) eigenvalues $\lambda_1,\ldots,\lambda_n$, then

$$\sqrt{\mathscr{A}^*\mathscr{A}}\, I_{p,q}\sqrt{\mathscr{A}^*\mathscr{A}}\mathbf{v}_j = \sqrt{\mathscr{A}^*\mathscr{A}}\, I_{p,q}\sqrt{\lambda_j}\mathbf{v}_j = \sqrt{\lambda_j}\sqrt{\mathscr{A}^*\mathscr{A}}\mathbf{v}'_j = \mathbf{v}'_j = I_{p,q}\mathbf{v}_j.$$

Thus, $\sqrt{\mathscr{A}^*\mathscr{A}}$ satisfies the defining relation for a matrix to be in $U(p,q)$. □

We may use the polar decomposition (for matrices in $GL(n,\mathbb{C})$) to write

$$\mathscr{A} = U|\mathscr{A}|, \text{ where } U = \mathscr{A}\left(\sqrt{\mathscr{A}^*\mathscr{A}}\right)^{-1}, |\mathscr{A}| = \sqrt{\mathscr{A}^*\mathscr{A}},$$

where $U, |\mathscr{A}| \in U(p,q)$. For U$(1,1)$ this decomposition takes the form

$$\begin{pmatrix} \alpha & \beta \\ \lambda\overline{\beta} & \lambda\overline{\alpha} \end{pmatrix} = \begin{pmatrix} \frac{\alpha}{|\alpha|} & 0 \\ 0 & \lambda\frac{\overline{\alpha}}{|\alpha|} \end{pmatrix} \begin{pmatrix} |\alpha| & \frac{|\alpha|\beta}{\alpha} \\ \frac{|\alpha|\overline{\beta}}{\overline{\alpha}} & |\alpha| \end{pmatrix}$$

The Lie algebra $\mathfrak{u}_{p,q}$ is given by matrices $X \in M_n(\mathbb{C})$ such that

$$X^*I_{p,q} + I_{p,q}X = 0.$$

This is real Lie subalgebra of $M_{p+q}(\mathbb{C})$. It contains the subalgebras $\mathfrak{u}_p, \mathfrak{u}_q$ as Lie algebras of the subgroups $U(p)\times I_q$ and $I_p \times U(q)$. Consider the inner product

$$\langle\cdot,\cdot\rangle : \mathfrak{u}_{p,q}\times\mathfrak{u}_{p,q} \to \mathbb{R}, \;\; \langle X, Y\rangle := \text{trace}(X^*Y).$$

Lemma 5.2.2. The inner product is symmetric and positive-definite.

Proof. Note that

$$\langle X, Y\rangle = \text{trace}(-I_{p,q}XI_{p,q}Y) = \text{trace}(-I_{p,q}YI_{p,q}X) = \langle Y, X\rangle .$$

Since $\overline{\langle X, Y\rangle} = \langle Y, X\rangle$ due to the invariance of trace under transpose, we conclude that the inner product is real and symmetric. It is positive-definite as $\langle X, X\rangle = \text{trace}(X^*X) \geq 0$ and equality holds if and only if X is the zero matrix. □

The Riemannian metric obtained by left translations of $\langle \cdot, \cdot \rangle$ will also be denoted by $\langle \cdot, \cdot \rangle$. We shall analyze the geodesics for this metric. The Lie algebra $\mathfrak{u}_p \oplus \mathfrak{u}_q$ of $U(p) \times U(q)$ consists of

$$\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}, \; A + A^* = 0, \; D + D^* = 0.$$

Let $\mathfrak{n}$ denote the orthogonal complement of $\mathfrak{u}_p \oplus \mathfrak{u}_q$ inside $\mathfrak{u}_{p,q}$. As $\mathfrak{n}$ is of (complex) dimension pq, and

$$\left\{ \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix} \;\middle|\; B \in M_{p,q}(\mathbb{C}) \right\}$$

is contained in $\mathfrak{n}$, this is all of it. We may verify that

$$\begin{aligned} \left[\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}, \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix} \right] &= \begin{pmatrix} 0 & AB - BD \\ DB^* - B^*A & 0 \end{pmatrix} \in \mathfrak{n} \\ \left[\begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}, \begin{pmatrix} 0 & C \\ C^* & 0 \end{pmatrix} \right] &= \begin{pmatrix} BC^* - CB^* & 0 \\ 0A & B^*C - C^*B \end{pmatrix} \in \mathfrak{u}_p \oplus \mathfrak{u}_q. \end{aligned}$$

Lemma 5.2.3. Let γ be the integral curve, initialized at e, for a left-invariant vector field Y. This curve is a geodesic if $Y(e)$ either belongs to $\mathfrak{n}$ or to $\mathfrak{u}_p \oplus \mathfrak{u}_q$.

Proof. The Levi-Civita connection ∇ is given by the Koszul formula

$$2\langle X, \nabla_Z Y\rangle = Z\langle X, Y\rangle + Y\langle X, Z\rangle - X\langle Y, Z\rangle + \langle Z, [X, Y]\rangle + \langle Y, [X, Z]\rangle - \langle X, [Y, Z]\rangle .$$

Putting $Z = Y$ and X, two left-invariant vector fields, in the above, we obtain

$$\langle X, \nabla_Y Y\rangle = \langle Y, [X, Y]\rangle .$$

To prove our claim, it suffices to show that $\nabla_Y Y = 0$, ie, $\langle Y, [X, Y]\rangle = 0$ for any X. Let us assume that $Y(e) \in \mathfrak{n}$. If $X(e) \in \mathfrak{n}$, then $[X(e), Y(e)] \in \mathfrak{u}_p \oplus \mathfrak{u}_q$, which implies that $\langle Y(e), [X(e), Y(e)]\rangle = 0$. If $X(e) \in \mathfrak{u}_p \oplus \mathfrak{u}_q$, then

$$\begin{aligned} \langle Y, [X, Y]\rangle &= \left\langle \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}, \begin{pmatrix} 0 & AB - BD \\ DB^* - B^*A & 0 \end{pmatrix} \right\rangle \\ &= \text{trace} \begin{pmatrix} B(DB^* - B^*A) & 0 \\ 0 & B^*(AB - BD) \end{pmatrix} \\ &= \text{trace}(BDB^* - BB^*A) + \text{trace}(B^*AB - B^*BD) \\ &= 0 \end{aligned}$$

by the cyclic property of trace. Thus, $\nabla_Y Y = 0$ if $Y(e) \in \mathfrak{n}$; similar proof works if $Y(e) \in \mathfrak{u}_p \oplus \mathfrak{u}_q$. □

Remark 5.2.1. An integral curve of a left-invariant vector field (also called 1-parameter subgroups) need not be a geodesic in $U(p,q)$. For instance, if $X+Y$ is a left-invariant vector field given by $X(e) \in \mathfrak{u}_p \oplus \mathfrak{u}_q$ and $Y(e) \in \mathfrak{n}$, then $\nabla_{X+Y}(X+Y) = 0$ if and only if $\nabla_X Y = \frac{1}{2}[X,Y]$ and $\nabla_Y X = \frac{1}{2}[Y,X]$. This happens if and only if the metric is bi-invariant, i.e.,

$$\langle [X,Z],Y\rangle = \langle X,[Z,Y]\rangle\,.$$

This is not true in general; for instance, with $X(e) \in \mathfrak{u}_p \oplus \mathfrak{u}_q$ and linearly independent $Y(e), Z(e) \in \mathfrak{n}$, we get $\langle [X,Z],Y\rangle - \langle X,[Z,Y]\rangle \neq 0$.

Consider the matrix

$$Y = \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix} \in \mathfrak{n}.$$

Let $B = U\sqrt{B^*B}$ and $B^* = \sqrt{B^*B}\,U^*$ be the polar decompositions for the rectangular matrices. It follows from direct computation that

$$\begin{aligned} e^Y &= \begin{pmatrix} I_p + \frac{BB^*}{2!} + \frac{(BB^*)^2}{4!} + \cdots & \frac{B}{1!} + \frac{B(B^*B)}{3!} + \frac{B(B^*B)^2}{5!} + \cdots \\ \frac{B^*}{1!} + \frac{(B^*B)B^*}{3!} + \frac{(B^*B)^2B^*}{5!} + \cdots & I_q + \frac{B^*B}{2!} + \frac{(B^*B)^2}{4!} + \cdots \end{pmatrix} \\ &= \begin{pmatrix} \cosh(\sqrt{BB^*}) & U\sinh(\sqrt{B^*B}) \\ \sinh(\sqrt{B^*B})U^* & \cosh(\sqrt{B^*B}) \end{pmatrix}. \end{aligned}$$

It can be checked that

$$e^{\mathfrak{n}} \cap (U(p)\times U(q)) = \{I_n\}.$$

It is known that the non-zero eigenvalues of Y are the non-zero eigenvalues of $\sqrt{BB^*}$ and their negatives.

Theorem 5.2.1. For any element $\mathscr{A} \in U(p,q)$, the associated matrix $\sqrt{\mathscr{A}^*\mathscr{A}}$ can be expressed uniquely as e^Y for $Y \in \mathfrak{n}$. Moreover, there is a unique way to express $\mathscr{A}$ as a product of a unitary matrix and an element of $e^{\mathfrak{n}}$, and it is given by the polar decomposition.

In order to prove the result, we discuss some preliminaries on logarithm of complex matrices. In general, there is no unique logarithm. However, the Gregory series

$$\log A = -\sum_{m=0}^{\infty} \frac{2}{2m+1}\left[(I-A)(I+A)^{-1}\right]^{2m+1}$$

converges if all the eigenvalues of $A \in M_n(\mathbb{C})$ have positive real part, see [Higham, 2008, §11.3, page 273]. In particular, $\log A$ is well-defined for Hermitian positive-definite matrix. This is often called the *principal logarithm* of A. This logarithm satisfies $e^{\log A} = A$. There is an integral form of logarithm that applies to matrices without real or zero eigenvalues; it is given by

$$\log A = (A-I)\int_0^1 [s(A-I)+I]^{-1}\,ds.$$

Lemma 5.2.4. The inverse of $\mathscr{A}^*\mathscr{A} + I_n$ for $\mathscr{A} \in U(p,q)$ is given by

$$[\mathscr{A}^*\mathscr{A} + I_n]^{-1} = \frac{1}{2}\begin{pmatrix} I_p & -A^{-1}B \\ -B^*(A^*)^{-1} & I_q \end{pmatrix}.$$

Proof. Since $\mathscr{A}^*\mathscr{A}$ has only positive eigenvalues, $\mathscr{A}^*\mathscr{A} + I_n$ has no kernel. We note that

$$\mathscr{A}^*\mathscr{A} + I_n = \begin{pmatrix} 2C^*C + 2I_p & 2A^*B \\ 2B^*A & 2B^*B + 2I_q \end{pmatrix} = \begin{pmatrix} 2A^*A & 2A^*B \\ 2B^*A & 2D^*D \end{pmatrix}.$$

The inverse matrix satisfies

$$\begin{pmatrix} 2A^*A & 2A^*B \\ 2B^*A & 2D^*D \end{pmatrix}\begin{pmatrix} E & F \\ F^* & G \end{pmatrix} = \begin{pmatrix} I_p & 0 \\ 0 & I_q \end{pmatrix}.$$

As the matrices are Hermitian, the three constraints that E, F, G must satisfy (and are uniquely determined by) are

$$\begin{aligned} E &= \tfrac{1}{2}(A^*A)^{-1} - A^{-1}BF^* \\ G &= \tfrac{1}{2}(D^*D)^{-1} - D^{-1}CF \\ F &= -A^{-1}BG. \end{aligned}$$

We note that $E = \frac{1}{2}I_p$, $G = \frac{1}{2}I_q$ and $F = -\frac{1}{2}A^{-1}B$ satisfy the above equations. For instance,

$$\begin{aligned} \frac{1}{2}(A^*A)^{-1} - A^{-1}BF^* &= \frac{1}{2}(A^*A)^{-1} + \frac{1}{2}A^{-1}BB^*(A^*)^{-1} \\ &= \frac{1}{2}(A^*A)^{-1} + \frac{1}{2}A^{-1}(AA^* - I_p)(A^*)^{-1} = \frac{1}{2}I_p, \end{aligned}$$

where $BB^* = AA^* - I_p$ is a consequence of $\mathscr{A}^* \in U(p,q)$. Yet another consequence is $AC^* = BD^*$, which is equivalent to

$$A^{-1}B = (D^{-1}C)^*.$$

In a similar vein,

$$\begin{aligned} \frac{1}{2}(D^*D)^{-1} - D^{-1}CF &== \frac{1}{2}(D^*D)^{-1} + \frac{1}{2}D^{-1}CC^*(D^*)^{-1} \\ &= \frac{1}{2}(D^*D)^{-1} + \frac{1}{2}D^{-1}(DD^* - I_q)(D^*)^{-1} \\ &= \frac{1}{2}I_q, \end{aligned}$$

where $CC^* = DD^* - I_q$ is due to $\mathscr{A}^* \in U(p,q)$. □

Proof of Theorem 5.2.1. We use Gregory series expansion for computing the principal logarithm of $\mathscr{A}^*\mathscr{A}$ along with Lemma 5.2.4:

$$\log(\mathscr{A}^*\mathscr{A})$$

$$= \sum_{m=0}^{\infty} \frac{2}{2m+1} \left[2 \begin{pmatrix} A^*A - I_p & A^*B \\ B^*A & D^*D - I_q \end{pmatrix} \frac{1}{2} \begin{pmatrix} I_p & -A^{-1}B \\ -B^*(A^*)^{-1} & I_q \end{pmatrix} \right]^{2m+1}$$
$$= \sum_{m=0}^{\infty} \frac{2}{2m+1} \begin{pmatrix} 0 & A^{-1}B \\ B^*(A^*)^{-1} & 0 \end{pmatrix}^{2m+1}.$$

We set $Y = \frac{1}{2}\log(\mathscr{A}^*\mathscr{A})$. It is clear that $Y \in \mathfrak{n}$ and $e^Y = \sqrt{\mathscr{A}^*\mathscr{A}}$. It is known that the exponential map is injective on Hermitian matrices. This implies the uniqueness of Y.

If $U_1 e^{Y_1} = U_2 e^{Y_2}$ are two decompositions of $\mathscr{A} \in U(p,q)$ with $U_i \in \mathrm{U}(p) \times \mathrm{U}(q)$ and $Y_i \in \mathfrak{n}$, then

$$e^{2Y_1} = e^{Y_1} U_1^* U_1 e^{Y_1} = e^{Y_2} U_2^* U_2 e^{Y_2} = e^{2Y_2}.$$

By the injectivity of the exponential map (on Hermitian matrices), we obtain $Y_1 = Y_2$, which implies that $U_1 = U_2$. □

We infer the following (see [Yakubovich and Starzhinskii, 1975, Lemma 1, page 211] for a different proof) result.

Corollary 5.2.1. The exponential map $\exp : \mathfrak{u}_{p,q} \to U(p,q)$ is surjective.

Proof. Using the polar decomposition and Theorem 5.2.1,

$$\mathscr{A} = \mathscr{A}\left(\sqrt{\mathscr{A}^*\mathscr{A}}\right)^{-1} \sqrt{\mathscr{A}^*\mathscr{A}} = \mathscr{A}\left(\sqrt{\mathscr{A}^*\mathscr{A}}\right)^{-1} e^Y.$$

Since the matrix exponential is surjective for $U(p) \times U(q)$, choose $Z \in \mathfrak{u}_p \oplus \mathfrak{u}_q$ such that $e^Z = \mathscr{A}(\sqrt{\mathscr{A}^*\mathscr{A}})^{-1}$. By Baker-Campbell-Hausdorff formula, we may express $e^Z e^Y$ as exponential of an element in $\mathfrak{u}_{p,q}$. □

The distance from any matrix $\mathscr{A} \in U(p,q)$ to $U(p) \times U(q)$ is given by the length of the curve

$$\gamma(t) = \mathscr{A}\left(\sqrt{\mathscr{A}^*\mathscr{A}}\right)^{-1} e^{tY},$$

which can be computed (and simplified via left-invariance) as follows

$$\ell(\gamma) = \int_0^1 \|\gamma'(t)\|_{\gamma(t)}\, dt = \int_0^1 \|Y\|\, dt = \|Y\|.$$

Note that

$$\|Y\|^2 = \mathrm{trace}(Y^*Y) = \mathrm{trace}\left[\tfrac{1}{4}(\log(\mathscr{A}^*\mathscr{A}))^2\right].$$

Thus, the distance squared function is given by

$$d^2 : U(p,q) \to \mathbb{R},\ \mathscr{A} \mapsto \tfrac{1}{4}\mathrm{trace}\left[(\log(\mathscr{A}^*\mathscr{A}))^2\right].$$

Chapter 6

Equivariant cut locus

Contents

Let M be a smooth manifold on which a compact Lie group G acts freely. It is known that M/G is a smooth manifold. Moreover, if M has a Riemannian metric and G acts isometrically, then there is an induced Riemannian metric on M/G. Let N be a G-invariant submanifold of M. We want to find the cut locus of N/G in M/G. In this chapter we will discuss the equality between $\mathrm{Cu}(N)/G$ and $\mathrm{Cu}(N/G)$. We will start the chapter by motivating with an example, then state the theorem and then recall Riemannian submersion which will be useful for the proof of the theorem. At the end we will discuss an application of this result to complex projective hypersurfaces.

6.1 Statement of the theorem

Let us discuss an example which will be helpful to arrive at the statement of the main theorem. Let $\mathbb{RP}^n$ denotes the n-dimensional real projective space which is obtained from $\mathbb{S}^n$ by identifying points p and $-p$. Equivalently, this space can be obtained from the action of $\mathbb{Z}_2$ on $\mathbb{S}^n$. We know from Example 2.3.4 that for a point $p \in \mathbb{RP}^n$, the cut locus is $\mathbb{RP}^{n-1}$. In Example 3.1.5, we showed that $\mathrm{Cu}\left(\mathbb{S}^k_i\right) = \mathbb{S}^{n-k-1}_l$, where $\mathbb{S}^k_i \hookrightarrow \mathbb{S}^n$ denote the embedding of the k-sphere in the first $k+1$ coordinates while $\mathbb{S}^{n-k-1}_l$ denote the embedding of the $(n-k-1)$-sphere in the last $n-k$ coordinates. So we have

$$\begin{array}{ccc} \mathbb{S}^n \supset \{p,-p\} & \xrightarrow{\mathrm{Cu}} & \mathbb{S}^{n-1} \\ \downarrow \mathbb{Z}_2 & & \downarrow \mathbb{Z}_2 \\ \mathbb{RP}^n \supset \{p\} & \xrightarrow{\mathrm{Cu}} & \mathbb{RP}^{n-1} \end{array}$$

Similarly, one can see a similar diagram for the complex projective space $\mathbb{CP}^n$

which is obtained by taking $\mathbb{S}^1$ action on $\mathbb{S}^{2n+1}$. We have

$$\begin{array}{ccc} \mathbb{S}^{2n+1} \supset \mathbb{S}^1_i & \xrightarrow{\text{Cu}} & \mathbb{S}^{2n-1}_f \\ \downarrow \mathbb{S}^1 & & \downarrow \mathbb{S}^1 \\ \mathbb{CP}^n \supset \{p\} & \xrightarrow{\text{Cu}} & \mathbb{CP}^{n-1} \end{array}$$

Thus, it is natural to ask whether the following diagram commutes for a compact Lie group G.

$$\begin{array}{ccc} M \supset N & \xrightarrow{\text{Cu}} & \text{Cu}(N) \\ \downarrow G & & \downarrow G \\ M/G \supset N/G & \xrightarrow{\text{Cu}} & \text{Cu}(N)/G \end{array}$$

The above diagram make sense if N and $\text{Cu}(N)$ are G-invariant subsets. Note that if the action is free and isometric, i.e., length of the curves γ and $g \cdot \gamma$ are the same, then $\text{Se}(N)$ is G-invariant. Now to show that the cut locus of N, which

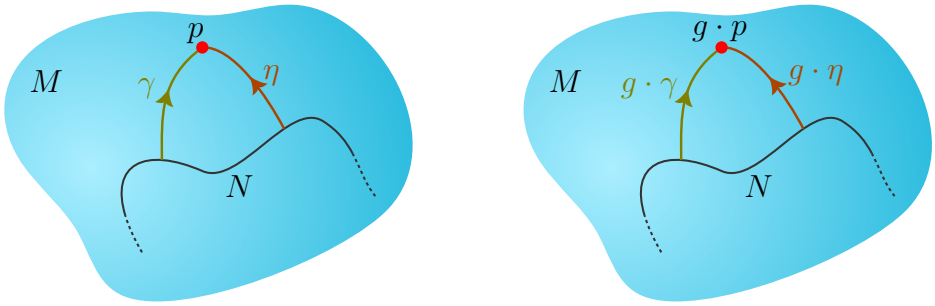


Figure 6.1: $\text{Se}(N)$ is G-invariant

is the closure of $\text{Se}(N)$, is G-invariant we take $x \in \overline{\text{Se}(N)}$. So there exists a sequence $(x_n) \subset \text{Se}(N)$ such that $x_n \to x$ (this convergence is with respect to the Riemannian metric). This implies $g \cdot x_n \to g \cdot x$ as the action is continuous. Hence, $g \cdot x \in \text{Se}(N)$.

The following theorem tells us that the above diagram commutes if G is a compact Lie group and the action is free and isometric.

Theorem 6.1.1 (Equivariant cut locus theorem). Let M be a closed and connected Riemannian manifold and G be any compact Lie group which acts on M freely and isometrically. Let N be any G-invariant closed submanifold of M, then we have an equality

$$\text{Cu}(N)/G = \text{Cu}(N/G).$$

Remark 6.1.1. (i) If the action of G is not isometric, then we can construct a G-invariant metric on M by averaging any metric on M over G. In fact, for any $p \in M$ and any vectors $\mathbf{v}_1, \mathbf{v}_2 \in T_pM$, we can define

$$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle := \int_G \langle \mathbf{v}_1, \mathbf{v}_2 \rangle_{g \cdot p} \, dg,$$

where the integral is taken with respect to the Haar measure. Then the theorem is valid with respect to the new metric.

(ii) Recall that, in Theorem 4.3.5, we proved that the gradient flow of distance squared function from the submanifold N deforms $M - \mathrm{Cu}(N)$ to N. By construction, the flow lines are G-invariant. So we also have that $M/G - \mathrm{Cu}(N)/G$ deforms to N/G.

To prove the theorem we have to encounter mainly two problems.

a. Whether a distance minimal geodesic in M projects down to a distance minimal geodesic in M/G?

b. Whether a distance minimal geodesic in M/G lifts to a distance minimal geodesic in M?

6.2 Proof of the equivariant cut locus theorem

In this section we will recall some results on Riemannian submersion and most of the results can be found in the book [Michor, 2008, Chapter V, section 26].

Definition 6.2.1. Let $\pi : E \to B$ be a smooth principal G-bundle. An *Ehresmann connection* on E is a smooth subbundle $\mathscr{H}$ of TE, called the *horizontal bundle* of the connection, such that $TE = \mathscr{H} \oplus \mathscr{V}$, where $\mathscr{V}_p = \ker\left(d\pi_p : T_pE \to T_{\pi(p)}B\right)$.

The bundle $\mathscr{V}$ is called the *vertical bundle*, and it is independent of the connection chosen. It follows from the definition that $\mathscr{H}_p$ depends smoothly on p and $\mathscr{H}_p \cap \mathscr{V}_p = \{0\}$. Moreover, the map $d\pi_p$ restricts to $\mathscr{H}_p$ is an isomorphism on $T_{\pi(p)}B$. If E is a Riemannian manifold with metric g, then by choosing a horizontal bundle $\mathscr{H}$ we have $\mathscr{H}_p = \mathscr{V}_p^{\perp}$.

Definition 6.2.2. A smooth submersion $\pi : (E, g) \to (B, g')$ is called a *Riemannian submersion* if the linear map $d\pi_p$ preserves the length of the horizontal vectors for each point $p \in E$. Equivalently, $d\pi_p$ is a linear isometry between $\mathscr{H}_p$ and $T_{\pi(p)}B$.

Using the above definitions we can define the following type of vectors.

Definition 6.2.3. Let $\pi : E \to B$ be a Riemannian submersion. A vector field $X \in \mathfrak{X}(E)$ is called

- *vertical* if for any $p \in E$, $X_p \in \mathscr{V}_p$, denoted by X^{ver}, and
- *horizontal* if for any $p \in E$, $X_p \in \mathscr{H}_p$, denoted by X^{hor}.

We can uniquely decompose any vector field $X \in \mathfrak{X}(E)$ as

$$X = X^{\text{ver}} + X^{\text{hor}}$$

into its horizontal and vertical components.

Once we have a connection, we now have a preferred way of lifting vectors from TB to TE. Recall that a vector $\tilde{X} \in T_eE$ is a *lift* of $X \in T_{\pi(e)B}$ if $T_e\pi(\tilde{X}) = X$. In absence of a connection, there are many different choices of lifts of a vector, and any two choices differ by a vertical vector. That is, if $\tilde{X}, \tilde{X}'$ are lifts of X, then $\tilde{X} - \tilde{X}'$ is vertical. Once we have a connection, we can define the *horizontal lift* (with respect to a connection $\mathscr{H}$) of X as the horizontal component of any lift of X. This definition is, of course, independent of the choice of lift, since any two differ by a vertical vector, whose horizontal component vanishes. Similarly, we can lift vector fields by lifting them in a pointwise fashion.

Definition 6.2.4 (Horizontal lift of vector fields). Let $X \in \mathfrak{X}(B)$ be a vector field and $\mathscr{H} \subset TE$ an Ehresmann connection on E. We define the *horizontal lift of* X as the vector field $\tilde{X} \in \mathfrak{X}(E)$ which satisfies $d\pi(\tilde{X}) = X$ and $\tilde{X}_e \in \mathscr{H}_e$ for all $e \in E$.

Suppose that we have a curve $\gamma : [0,1] \to B$. At each point over the curve, we have a vector $\gamma'(t) \in T_{\gamma(t)}B$, which we can lift to the fiber above $\gamma(t)$. So if we choose a starting point $e_0 \in \pi^{-1}(\gamma(0))$, we can find an integral curve along all these lifted vectors on the fibers over the curve γ. In the end we obtain a curve $\tilde{\gamma} : [0,1] \to E$ satisfying $\pi \circ \tilde{\gamma} = \gamma,\ \tilde{\gamma}(0) = e_0$, and $\tilde{\gamma}'(t) \in \mathscr{H}_{\tilde{\gamma}(t)}$ for all t. We call it a horizontal lift of γ.

Definition 6.2.5 (Horizontal lift of a curve). Let $\pi : E \to B$ be a fiber bundle with a connection $\mathscr{H}$. Let γ be a smooth curve in B through $\gamma(0) = b$. Let $e \in E$ be such that $\pi(e) = b$. A *horizontal lift* of γ through e is a curve $\tilde{\gamma}$ in E such that $\pi \circ \tilde{\gamma} = \gamma,\ \tilde{\gamma}(0) = e$, and $\tilde{\gamma}'(t) \in \mathscr{H}_{\tilde{\gamma}(t)}$.

For every point $t_0 \in (0,1)$, we can find $\varepsilon > 0$ such that the vector field γ' can be extended to a vector field over $\gamma\big|_{(t_0-\varepsilon,t_0+\varepsilon)}$. Then we look at the horizontal vector field $\tilde{X}$ defined on the bundle $E\big|_U$, where $U \supseteq \gamma((t_0-\varepsilon, t_0+\varepsilon))$. Then $\tilde{\gamma}$ is the integral curve of $\tilde{X}$ starting at the prescribed point $e_0 \in \pi^{-1}(\gamma(0))$. Since $[0,1]$ is compact, a usual gluing argument will help us to construct a horizontal lift of γ. Hence, we have the following proposition. For a detailed proof of the proposition we refer the reader to [Lang, 1999, Chapter XIV, Proposition 3.5(i)].

Proposition 6.2.1. Given a smooth path $\gamma : [0,1] \to B$ such that $\gamma(0) = b$ and $e_0 \in \pi^{-1}(b)$, there is a unique horizontal lift $\tilde{\gamma}$ of γ through $e_0 \in E$.

Recall the Quotient manifold theorem [Lee, 2013, Theorem 21.10].

Theorem 6.2.1 (Quotient Manifold Theorem). Suppose a Lie group G acting smoothly, freely, and properly on a smooth manifold M. Then the orbit space M/G is a topological manifold of dimension equal to $\dim M - \dim G$, and has a unique smooth structure with the property that the quotient map $\pi : M \to M/G$ is a smooth submersion.

Using the above theorem, we can define a unique metric on M/G such that π is a Riemannian submersion. In fact, for any $x \in M/G$ and $p \in \pi^{-1}(x)$, we take the vertical space $\mathscr{V}_p := \ker d\pi_p$ and the horizontal space $\mathscr{H}_p := \mathscr{V}_p^{\perp}$ so that $T_pM = \mathscr{V}_p \oplus \mathscr{H}_p$. Since $d\pi_p$ is surjective, then $d\pi_p\big|_{\mathscr{H}_p} : \mathscr{H}_p \to T_{\pi(p)}M/G$ is a bijection. Define

$$h_x(v, w) := \left\langle d\pi^{-1}\big|_{\mathscr{H}_p}(v), d\pi^{-1}\big|_{\mathscr{H}_p}(w) \right\rangle.$$

Since the action is isometric, the metric is independent of the choice of point p. Thus, h defines a well-defined metric on M/G and π is a Riemannian submersion.

The following is the key lemma for proving Theorem 6.1.1 and the proof of the same can be found in [Michor, 2008, Lemma 26.11].

Lemma 6.2.1. Let (E, g_E) and (B, g_B) be two Riemannian manifolds and $\pi : E \to B$ be a Riemannian submersion. Let γ be a geodesic in B and $\tilde{\gamma}$ be the horizontal lift of γ. Then we have:

1. The length of γ and $\tilde{\gamma}$ are same.
2. $\tilde{\gamma}'(t)$ is perpendicular to each fiber $E_{\tilde{\gamma}(t)}$.
3. $\tilde{\gamma}$ is a geodesic in E.

We also have the following result.

Theorem 6.2.2 (O'Neill). Let $\pi : E \to B$ be a Riemannian submersion. If $\tilde{\gamma}$ is a geodesic in E and $\tilde{\gamma}'(0) \in \mathscr{H}_{\tilde{\gamma}(0)}$, then $\tilde{\gamma}'(t) \in \mathscr{H}_{\tilde{\gamma}(t)}$ for all t. Moreover, $\pi \circ \tilde{\gamma}$ is a geodesic in B and the length is preserved.

The above theorem can also be proved using Lemma 6.2.1, see [Michor, 2008, Corollary 26.12].

Combining Lemma 6.2.1 and Theorem 6.2.2, we have the following correspondence.

Theorem 6.2.3. There is a one-to-one correspondence between the geodesics on M/G and geodesics on M which are horizontal.

We now ready to prove Theorem 6.1.1.

Proof of Theorem 6.1.1. Note that if $\tilde{\gamma}$ is an N-geodesic, then $\tilde{\gamma}'(1) \in \mathscr{H}_{\tilde{\gamma}(1)}$ and from Theorem 6.2.2 $\tilde{\gamma}'(t) \in \mathscr{H}_{\tilde{\gamma}(t)}$ and hence $\tilde{\gamma}$ is a horizontal geodesic which implies $\gamma = \pi \circ \tilde{\gamma}$ is a geodesic in M/G.

Let $\tilde{p} \in \mathrm{Se}(N)$. This implies there exists at least two N-geodesic, say $\tilde{\gamma}$ and $\tilde{\eta}$ such that $l(\tilde{\gamma}) = l(\tilde{\eta}) = d(\tilde{p}, N)$. Let us denote $\gamma = \pi \circ \tilde{\gamma}$ and $\eta = \pi \circ \tilde{\eta}$. Due to uniqueness of horizontal lift (Proposition 6.2.1), γ and η will be two different geodesics and the lengths will be same. Also, note that as $\tilde{\gamma}$ is an N-geodesic, γ will be an N/G geodesic. Otherwise, there exists an N/G-geodesic δ joining p to N/G which gives a horizontal lift $\tilde{\delta}$ whose length is strictly less than $\tilde{\gamma}$, a contradiction (see Figure 6.2). Hence, $\pi(\tilde{p}) \in \mathrm{Se}(N/G)$. On the other hand, if γ

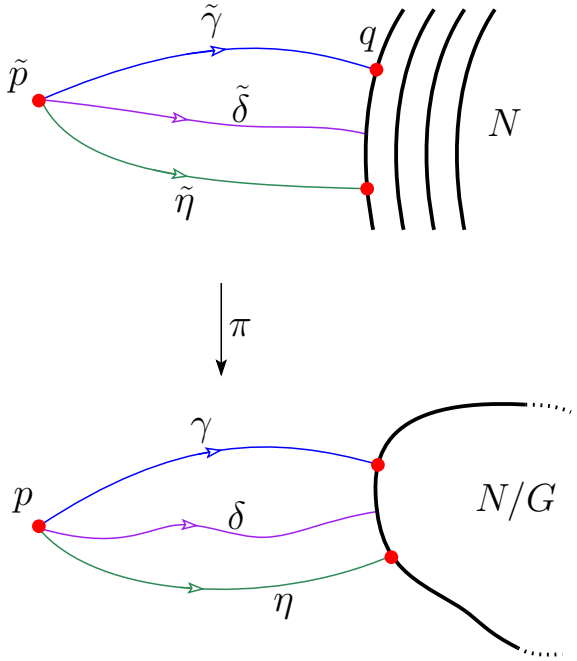


Figure 6.2: N-geodesics maps to N/G-geodesics

is an N/G geodesic starting from p, then its horizontal lift $\tilde{\gamma}$ will be a geodesic. In fact, it will be an N-geodesic. If not, let $\tilde{\eta}$ be such that $l(\tilde{\eta}) = d(\tilde{p}, N)$ which implies $\tilde{\eta}$ is horizontal. Hence, η will be a geodesic and

$$\begin{aligned} l(\gamma) &= d(p, N/G) = l(\eta) \\ &= l(\tilde{\eta}) < l(\tilde{\gamma}) = l(\gamma), \end{aligned}$$

a contradiction. Thus $\tilde{p} \in \mathrm{Se}(N)$. This proves that $\mathrm{Se}(N)/G = \mathrm{Se}(N/G)$. In order to prove the theorem, note that we have the following relation.

$$\mathrm{Se}(N/G) \subseteq \overline{\mathrm{Se}(N)}/G \subseteq \overline{\mathrm{Se}(N/G)}.$$

As $\overline{\mathrm{Se}(N)}$ is a closed set and G is a compact Lie group, so $\overline{\mathrm{Se}(N)}/G$ is closed. Thus, we have

$$\overline{\mathrm{Se}(N)}/G = \overline{\mathrm{Se}(N/G)} \implies \mathrm{Cu}(N)/G = \mathrm{Cu}(N/G).$$

□

We will discuss some examples based on the above result. Recall from Example 3.1.5,

$$\mathrm{Cu}\left(\mathbb{S}^k_i\right) = \mathbb{S}^{n-k-1}_l.$$

Example 6.2.1 (Real projective spaces). Take $M = \mathbb{S}^n,\ N = \mathbb{S}^k_i$ and $G = \mathbb{S}^0 \cong \mathbb{Z}_2$. Applying Theorem 6.1.1, we get

$$\mathrm{Cu}\left(\mathbb{RP}^k_i\right) \cong \mathbb{RP}^{n-k-1}_l.$$

Example 6.2.2. Take $M = \mathbb{S}^{2n+1},\ N = \mathbb{S}^{2k+1}_i$ and $G = \mathbb{S}^1$. Applying Theorem 6.1.1, we get

$$\mathrm{Cu}\left(\mathbb{CP}^k_i\right) \cong \mathbb{CP}^{n-k-1}_l.$$

Example 6.2.3 (Lens space). Consider $\mathbb{S}^{2n+1} \subset \mathbb{C}^{n+1}$. Let p be a prime number and $\xi = e^{\frac{2\iota\pi}{p}}$ be a primitive p^{th} root of unity and let $q_1, \cdots, q_{n+1}$ be integers coprime to p. Consider $\mathbb{Z}_p = \{1, \xi, \xi^2, \cdots, \xi^{p-1}\}$ and let it acts on $\mathbb{S}^{2n+1}$ by $\xi\,(z_1, \cdots, z_{n+1}) := (\xi^{q_1} z_1, \cdots, \xi^{q_{n+1}} z_{n+1})$. The orbit space is denoted by $L(p; q_1, \ldots, q_{n+1})$ and is called a *lens space*. Since $\mathrm{Cu}\,(\mathbb{S}^{2n-1} \subset \mathbb{S}^{2n+1}) = \mathbb{S}^1$, so taking the $\mathbb{Z}_p$ action, gives that the cut locus of $L\,(p; q_1, \ldots, q_n)$ in $L\,(p; q_1, \ldots, q_{n+1})$ is $\mathbb{S}^1$. In general, we have

$$\mathrm{Cu}\,(L\,(p; q_1, \ldots, q_{k+1})) \cong L\,(p; q_{k+2}, \ldots, q_{n+1})\,.$$

6.3 Cut locus of hypersurface in complex projective space

Let $\mathbf{z} = (z_0, \cdots, z_n) \in \mathbb{C}^{n+1}$ and $[\mathbf{z}] \in \mathbb{CP}^n$, then the Fermat hypersurface of degree d is given by the polynomial $f(\mathbf{z}) = z_0^d + \cdots + z_n^d$,

$$X(d) := \{[\mathbf{z}] : f(\mathbf{z}) = 0\}.$$

The homotopy type of the complement of the above hypersurface is well studied in the article [Kulkarni and Wood, 1980]. Since the partial derivatives $\frac{\partial f}{\partial z_j}$ do not vanish simultaneously on $\mathbb{C}^{n+1} - \{0\}$, the hypersurface is nonsingular. We wish to find its cut locus and want to compare our result with the result in [Kulkarni and Wood, 1980, Proposition 3.1]. In that paper, the authors found the homotopy type of the complement of the hypersurface.

Proposition 6.3.1 ([Kulkarni and Wood, 1980]). The complement of $X(d)$ is homotopic to the base space of the n universal principal $\mathbb{Z}_d$-bundle constructed by Milnor from the join of $n+1$ copies of $\mathbb{Z}_d$.

Using Theorem 4.3.5, the same can be studied by looking at the cut locus of $X(d)$ in $\mathbb{CP}^n$. We will use Theorem 6.1.1 to find the cut locus of $X(d)$. Recall that there is a principal $\mathbb{S}^1$-bundle given by

$$\pi : \mathbb{S}^{2n+1} \to \mathbb{CP}^n,\ (z_0, \cdots, z_n) \mapsto [z_0 : \cdots : z_n].$$

Therefore, using Theorem 6.1.1 it is enough to find the cut locus of $\pi^{-1}(X(d)) := \tilde{X}(d)$. We propose the following conjecture.

Conjecture. The cut locus of $\tilde{X}(d) \subseteq \mathbb{S}^{2n+1}$ is $\mathbb{Z}_d^{\star(n+1)} \times_{\mathbb{Z}_d} \mathbb{S}^1$, where $\times_{\mathbb{Z}_d}$ is the diagonal action of $\mathbb{Z}_d$ and $\star$ denotes the topological join of spaces.

Using Theorem 6.1.1, the cut locus of $X(d)$ will be $\mathbb{Z}_d^{\star(n+1)}$ and hence we are recovering the n^{th}-stage of Milnor join construction of classifying space for $\mathbb{Z}_d$. Note that we can see $\mathbb{Z}_d \subset \mathbb{S}^1$ and $\mathbb{S}^{2n+1}$ is the join of n circles. Hence, $\mathbb{Z}_d^{\star(n+1)} \subset \mathbb{S}^{2n+1}$ and also note that

$$\mathbb{Z}_d^{\star(n+1)} \times \mathbb{S}^1 \hookrightarrow \mathbb{S}^{2n+1}, \quad \left(\mathbf{v}, e^{\iota\theta}\right) \mapsto \mathbf{v}e^{-\iota\theta}$$

gives a well-defined map from $\mathbb{Z}_d^{\star(n+1)} \times_{\mathbb{Z}_d} \mathbb{S}^1$ to $\mathbb{S}^{2n+1}$. We will prove some particular cases of the above conjecture. More precisely, we will prove the above conjecture holds for $d = 2$ and arbitrary n (Theorem 6.3.1) and $n = 1$ and arbitrary d (Theorem 6.3.2). Let us denote

$$\widetilde{\mathrm{Cu}} = \mathrm{Cu}(\tilde{X}(d)), \text{ and } \widetilde{\mathrm{Se}} = \mathrm{Se}(\tilde{X}(d)).$$

Theorem 6.3.1 (Cut locus of $X(2)$). The cut locus of $\tilde{X}(2) \subseteq \mathbb{S}^{2n+1}$ is $\mathbb{S}^n \times_{\mathbb{Z}_2} \mathbb{S}^1 \cong \{(\mathbf{v}\cos\theta, \mathbf{v}\sin\theta) : \mathbf{v} \in \mathbb{S}^n, \theta \in [0, 2\pi]\}$. Hence, the cut locus of $X(2)$ in $\mathbb{CP}^n$ will be $\mathbb{RP}^{n+1}$.

Proof. We will show that $\{(\mathbf{v}\cos\theta, \mathbf{v}\sin\theta) : \mathbf{v} \in \mathbb{S}^n, \theta \in \mathbb{R}\} = \mathrm{Se}(\tilde{X}(2))$. Let $\mathbf{v} \in \mathbb{S}^n$ and $\theta \in \mathbb{R}$. Let us write $z_j = x_j + \iota y_j$. We can write $\tilde{X}(2)$ as

$$\begin{aligned}
\tilde{X}(2) &= \left\{(z_0, \cdots, z_n) \in \mathbb{C}^{n+1} : \sum_{i=0}^{n} z_i^2 = 0, \text{ and } \sum_{i=0}^{n} |z_i|^2 = 1\right\} \\
&= \left\{(x_0, y_0, \cdots, x_n, y_n) : \sum_{i=0}^{n} x_i^2 = \frac{1}{2} = \sum_{i=0}^{n} y_i^2, \text{ and } \sum_{i=0}^{n} x_i y_i = 0\right\} \\
&= \left\{(x_0, x_1, \cdots, x_n, y_0, y_1, \cdots, y_n) : \sum_{i=0}^{n} x_i^2 = \frac{1}{2} = \sum_{i=0}^{n} y_i^2, \text{ and } \sum_{i=0}^{n} x_i y_i = 0\right\}.
\end{aligned}$$

If $A \in O(n+1)$, then $\tilde{A} = \left(\begin{array}{c|c} A & \mathbf{0} \\ \hline \mathbf{0} & A \end{array}\right) \in SO(2n+2)$. Note that

i $\tilde{A} \in \mathrm{Iso}\left(\mathbb{S}^{2n+1}\right)$, where $\mathrm{Iso}(M)$ denotes the set of all isometries of M.

ii $\tilde{A}$ maps $\tilde{X}(2)$ to itself, and $\widetilde{\mathrm{Cu}}$ to itself.

Thus, if $p \in \widetilde{\mathrm{Se}} \subseteq \widetilde{\mathrm{Cu}}$ and let γ and η be two distance minimal geodesics joining p to $\tilde{X}(2)$, then $\tilde{A}\gamma$ and $\tilde{A}\eta$ will be two minimal geodesics joining $\tilde{A}p$ to $\tilde{X}(2)$. As the action of $O(n+1)$ on $\mathbb{S}^n$ is transitive, it suffices to check if $\mathbf{e}_1 \cos\theta + \mathbf{e}_{n+2}\sin\theta \in \widetilde{\mathrm{Se}}$.

We can further reduce our work by looking at the matrix

$$B = \left(\begin{array}{cccc|cccc} \cos\theta & 0 & \cdots & 0 & \sin\theta & 0 & \cdots & 0 \\ 0 & & & & 0 & & & \\ \vdots & & I_n & & \vdots & & \mathbf{0}_n & \\ 0 & & & & 0 & & & \\ \hline -\sin\theta & 0 & \cdots & 0 & \cos\theta & 0 & \cdots & 0 \\ 0 & & & & 0 & & & \\ \vdots & & \mathbf{0}_n & & \vdots & & I_n & \\ 0 & & & & 0 & & & \end{array}\right),$$

which is again an isometry of $\mathbb{S}^{2n+1}$ and sends $\mathbf{e}_1\cos\theta + \mathbf{e}_{n+2}\sin\theta$ to $\mathbf{e}_1$. Hence, it is enough to prove that $\mathbf{e}_1 \in \widetilde{\mathrm{Se}}$. Note that

$$\operatorname{dist}\left(\mathbf{e}_1, \tilde{X}(2)\right) = \frac{\pi}{4}.$$

Let

$$\mathbf{v}_1 = \mathbf{e}_{2n+2},\ \mathbf{v}_2 = -\mathbf{e}_{2n+2}.$$

Consider the geodesics

$$\gamma(t) = \mathbf{e}_1\cos t + \mathbf{e}_{2n+2}\sin t, \text{ and } \eta(t) = \mathbf{e}_1\cos t - \mathbf{e}_{2n+2}\sin t,\ t \in \mathbb{R}.$$

Note that γ and η intersect the set $\tilde{X(2)}$ at $t = \frac{\pi}{4}$, so their lengths are same, and it is equal to the distance between $\mathbf{e}_1$ and the submanifold. This proves that $\mathbf{e}_1 \in \widetilde{\mathrm{Se}}$.

Conversely, if $p = (\mathbf{v}, \mathbf{w}) \in \widetilde{\mathrm{Se}}$, then we will show that there exists $\mathbf{u} \in \mathbb{S}^n$ and $\theta \in \mathbb{R}$ such that $\mathbf{v} = \mathbf{u}\cos\theta$ and $\mathbf{w} = \mathbf{u}\sin\theta$. Note that if $\{\mathbf{v}, \mathbf{w}\}$ is linearly dependent, then there exists $\theta \in \mathbb{R}$ such that

$$\mathbf{v} = \hat{\mathbf{v}}\cos\theta,\ \ \mathbf{w} = \hat{\mathbf{v}}\sin\theta$$

and hence $p \in \widetilde{\mathrm{Se}}$. So we assume that $\mathbf{v}$ and $\mathbf{w}$ are linearly independent. Suppose $p \notin \mathbb{S}^n \times_{\mathbb{Z}_2} \mathbb{S}^1$. We need to show that $p \notin \widetilde{\mathrm{Se}}$. To the contrary, let us assume that $p \in \widetilde{\mathrm{Se}}$. Consider a unit speed geodesic $\gamma(t)$ in the direction of $\upsilon = (\mathbf{v}_1, \mathbf{v}_2) \in T_{(\mathbf{v},\mathbf{w})}\mathbb{S}^{2n+1}$ which implies

$$\|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 = 1 \tag{6.1}$$

$$\langle \mathbf{v}_1, \mathbf{v}\rangle + \langle \mathbf{v}_2, \mathbf{w}\rangle = 0 \tag{6.2}$$

Consider the curve $\gamma(t) = (\mathbf{v}\cos t + \mathbf{v}_1\sin t, \mathbf{w}\cos t + \mathbf{v}_2\sin t)$ for $t \in \mathbb{R}$. Note that $2 \le \operatorname{rank}[\mathbf{v}, \mathbf{w}, \mathbf{v}_1, \mathbf{v}_2] \le 4$. We will prove that none of the cases is possible. Since $p \in \widetilde{\mathrm{Se}}$, so $p \in \widetilde{\mathrm{Cu}}$ which implies there exists $t \in \mathbb{R}$ such that $\gamma(t) \in \tilde{X}(2), \gamma'(t) \in \left(T_{\gamma(t)}\tilde{X}(2)\right)^{\perp}$ and t will be minimum among all such values. So we have

$$\|\mathbf{v}\|^2\cos^2 t + \|\mathbf{v}_1\|^2\sin^2 t + \langle \mathbf{v}, \mathbf{v}_1\rangle \sin 2t = \frac{1}{2}, \tag{6.3}$$

$$\|\mathbf{w}\|^2\cos^2 t + \|\mathbf{v}_2\|^2\sin^2 t + \langle \mathbf{w}, \mathbf{v}_2\rangle \sin 2t = \frac{1}{2},$$

$$\langle \mathbf{v}, \mathbf{w}\rangle\cos^2 t + \langle \mathbf{v}_1, \mathbf{v}_2\rangle\sin^2 t + \frac{1}{2}\left(\langle \mathbf{v}, \mathbf{v}_2\rangle + \langle \mathbf{v}_1, \mathbf{w}\rangle\right)\sin 2t = 0. \tag{6.4}$$

Now we will make use of the other condition $\gamma'(t) \in \left(T_{\gamma(t)}\tilde{X}(2)\right)^{\perp}$. Consider the vector $u = (-\mathbf{w}\cos t + \mathbf{v}_2 \sin t, \mathbf{v}\cos t + \mathbf{v}_1 \sin t)$. We claim that $u \in T_{\gamma(t)}\tilde{X}(2)$. Note that $(\mathbf{v}, \mathbf{w}) \in T_{(\mathbf{p},\mathbf{q})}\tilde{X}(2)$ implies $\langle \mathbf{p}, \mathbf{v}\rangle = 0, \langle \mathbf{q}, \mathbf{w}\rangle = 0$ and $\langle \mathbf{p}, \mathbf{w}\rangle + \langle \mathbf{q}, \mathbf{v}\rangle = 0$.

$$\begin{aligned}\langle \mathbf{p}, \mathbf{v}\rangle &= \langle (\mathbf{v}\cos t + \mathbf{v}_1 \sin t), (-\mathbf{w}\cos t - \mathbf{v}_2 \sin t)\rangle \\ &= -\langle \mathbf{v}, \mathbf{w}\rangle \cos^2 t - \langle (\mathbf{v}, \mathbf{v}_2\rangle + \langle \mathbf{v}_1, \mathbf{w}\rangle) \cos t \sin t - \langle \mathbf{v}_1, \mathbf{v}_2\rangle \sin^2 t \\ &= 0 \qquad \text{(from (6.4))}.\end{aligned}$$

Similarly, $\langle \mathbf{q}, \mathbf{w}\rangle = 0$, and

$$\begin{aligned}&\langle \mathbf{p}, \mathbf{w}\rangle + \langle \mathbf{q}, \mathbf{v}\rangle \\ &= \langle \mathbf{v}\cos t + \mathbf{v}_1 \sin t, \mathbf{v}\cos t + \mathbf{v}_1 \sin t\rangle + \langle \mathbf{w}\cos t + \mathbf{v}_2 \sin t, -\mathbf{w}\cos t - \mathbf{v}_2 \sin t\rangle \\ &= \|\mathbf{v}\|^2 \cos^2 t + \|\mathbf{v}_1\|^2 \sin^2 t + 2\langle \mathbf{v}, \mathbf{v}_1\rangle \cos t \sin t - \|\mathbf{w}\|^2 \cos^2 t - \|\mathbf{v}_2\|^2 \sin^2 t \\ &\qquad - 2\langle \mathbf{w}, \mathbf{v}_2\rangle \cos t \sin t \\ &= \cos^2 t \left(\|\mathbf{v}\|^2 - \|\mathbf{w}\|^2\right) + \sin^2 t \left(\|\mathbf{v}_1\|^2 - \|\mathbf{v}_2\|^2\right) + \sin 2t \left(\langle \mathbf{v}, \mathbf{v}_1\rangle - \langle \mathbf{w}, \mathbf{v}_2\rangle\right) \\ &= 0 \qquad \text{(from (6.1) and (6.2))}.\end{aligned}$$

Therefore, $u \in T_{\gamma(t)}\tilde{X}(2)$ and hence $\langle u, \gamma'(t)\rangle = 0$ which implies

$$\langle \mathbf{v}, \mathbf{v}_2\rangle - \langle \mathbf{v}_1, \mathbf{w}\rangle = 0. \tag{6.5}$$

Define

$$\tilde{\mathbf{u}}_1 = \sqrt{2}\left(\mathbf{v}\cos t + \mathbf{v}_1 \sin t\right), \text{ and } \tilde{\mathbf{u}}_2 = \sqrt{2}\left(\mathbf{w}\cos t + \mathbf{v}_2 \sin t\right).$$

Note that $\tilde{\mathbf{u}}_1 \perp \tilde{\mathbf{u}}_2$ and both are vectors in $\mathbb{R}^{n+1}$. We extend $\{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2\}$ to an orthonormal basis of $\mathbb{R}^{n+1}$, say $\{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \tilde{\mathbf{u}}_3, \ldots, \tilde{\mathbf{u}}_{n+1}\}$. If $(\mathbf{w}_1, \mathbf{w}_2) \in T_{\gamma(t)}\tilde{X}(2)$, then $\langle \mathbf{w}_1, \tilde{\mathbf{u}}_1\rangle = 0$, $\langle \mathbf{w}_2, \tilde{\mathbf{u}}_2\rangle = 0$ and $\langle \mathbf{w}_1, \tilde{\mathbf{u}}_2\rangle + \langle \mathbf{w}_2, \tilde{\mathbf{u}}_1\rangle = 0$ as $\gamma(t) = \frac{1}{\sqrt{2}}(\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2)$. This implies $\mathbf{w}_1, \mathbf{w}_2 \in \text{Span}\{\tilde{\mathbf{u}}_3, \ldots, \tilde{\mathbf{u}}_{n+1}\}$ or

$$\mathbf{w}_1 = -\tilde{\mathbf{u}}_2 + \sum_{j\geq 3} c_j \tilde{\mathbf{u}}_j \text{ and } \mathbf{w}_2 = \tilde{\mathbf{u}}_1 + \sum_{j\geq 3} d_j \tilde{\mathbf{u}}_j.$$

Since $\mathbf{w}_i \in T_{\gamma(t)}\tilde{X}(2)$,

$$\langle \gamma'(t), (\mathbf{w}_1, \mathbf{w}_2)\rangle = 0 \implies \text{for } j \geq 3, \ \langle \gamma'(t), (\tilde{\mathbf{u}}_j, 0)\rangle = 0, \text{ and } \langle \gamma'(t), (0, \tilde{\mathbf{u}}_j)\rangle = 0.$$

The above implies

$$-\mathbf{v}\sin t + \mathbf{v}_1 \cos t, -\mathbf{w}\sin t + \mathbf{v}_2 \cos t \in \text{Span}\{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2\}. \tag{6.6}$$

Since $-\mathbf{v}\sin t + \mathbf{v}_1 \cos t \in \text{Span}\{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2\}$,

$$\begin{aligned}&-\mathbf{v}\sin t + \mathbf{v}_1 \cos t = \alpha\sqrt{2}\left(\mathbf{v}\cos t + \mathbf{v}_1 \sin t\right) + \beta\sqrt{2}\left(\mathbf{w}\cos t + \mathbf{v}_2 \sin t\right) \\ \implies &\mathbf{v}(\sin t - \alpha\sqrt{2}\cos t) + \mathbf{w}(-\beta\sqrt{2}\cos t) + \\ &\qquad \mathbf{v}_1(\cos t - \alpha\sqrt{2}\sin t) + \mathbf{v}_2(\beta\sqrt{2}\sin t) = 0.\end{aligned} \tag{6.7}$$

If $\text{rank}[\mathbf{v}, \mathbf{w}, \mathbf{v}_1, \mathbf{v}_2] = 4$, then from equation (6.7)

$$\sin t - \alpha\sqrt{2}\cos t = 0 = \cos t - \alpha\sqrt{2}\sin t, \text{ and } -\beta\cos t = 0 = \beta\sin t.$$

This implies $2\alpha^2 = -1$, which is absurd. Thus, all four vectors can not be linearly independent. Now we are remaining with two cases: $\operatorname{rank}[\mathbf{v}, \mathbf{w}, \mathbf{v}_1, \mathbf{v}_2] = 2$ or $\operatorname{rank}[\mathbf{v}, \mathbf{w}, \mathbf{v}_1, \mathbf{v}_2] = 3$.

Case 1: $\operatorname{rank}[\mathbf{v}, \mathbf{w}, \mathbf{v}_1, \mathbf{v}_2] = 2$: Let $\mathbf{v}_1 = \alpha\mathbf{v} + \beta\mathbf{w}$ and $\mathbf{v}_2 = \gamma\mathbf{v} + \delta\mathbf{w}$ for some $\alpha, \beta, \gamma, \delta \in \mathbb{R}$. Observe that

$$\begin{aligned}
(6.5) &\implies \gamma \|\mathbf{v}\|^2 - \beta \|\mathbf{w}\|^2 = (\alpha - \delta) \langle \mathbf{v}, \mathbf{w} \rangle \\
&\implies (\gamma + \beta) \|\mathbf{v}\|^2 - (\alpha - \delta) \langle \mathbf{v}, \mathbf{w} \rangle = \beta \qquad (6.8) \\
(6.2) &\implies \alpha \|\mathbf{v}\|^2 + \delta \|\mathbf{w}\|^2 = -(\beta + \gamma) \langle \mathbf{v}, \mathbf{w} \rangle \\
&\implies (\alpha - \delta) \|\mathbf{v}\|^2 + (\beta + \gamma) \langle \mathbf{v}, \mathbf{w} \rangle = -\delta. \qquad (6.9)
\end{aligned}$$

Using equations (6.8) and (6.9), we obtain

$$\|\mathbf{v}\|^2 \left((\alpha - \delta)^2 + (\beta + \gamma)^2 \right) = \beta\gamma + \beta^2 - \alpha\delta + \delta^2.$$

Note that $(\alpha - \delta)^2 + (\beta + \gamma)^2 = 0$ implies $\alpha = \delta$ and $\beta = -\gamma$. But from equations (6.8) and (6.9) implies that $\alpha = \beta = \delta = \gamma = 0$, which is not possible. Thus,

$$\begin{aligned}
\|\mathbf{v}\|^2 &= \frac{\beta\gamma + \beta^2 - \alpha\delta + \delta^2}{(\alpha - \delta)^2 + (\beta + \gamma)^2}, \\
\|\mathbf{w}\|^2 &= \frac{\beta\gamma + \delta^2 - \alpha\delta + \beta^2}{(\alpha - \delta)^2 + (\beta + \gamma)^2}, \text{ and} \\
\langle \mathbf{v}, \mathbf{w} \rangle &= \frac{-(\alpha\beta + \gamma\delta)}{(\alpha - \delta)^2 + (\beta + \gamma)^2}.
\end{aligned}$$

From equation (6.3), we have

$$\begin{aligned}
&\left(\alpha^2 + \gamma^2\right) \|\mathbf{v}\|^2 + 2 \left(\alpha\beta + \gamma\delta\right) \langle \mathbf{v}, \mathbf{w} \rangle + \left(\beta^2 + \delta^2\right) \|\mathbf{w}\|^2 = 1 \\
\implies &\left(\alpha^2 + \gamma^2\right) \left(\beta\gamma + \beta^2 - \alpha\delta + \delta^2\right) - 2(\alpha\beta + \gamma\delta)^2 \\
&\qquad + \left(\beta^2 + \delta^2\right) \left(\beta\gamma + \delta^2 - \alpha\delta + \beta^2\right) - (\alpha - \delta)^2 - (\beta + \gamma)^2 = 0 \\
\implies &(\beta\gamma - \alpha\delta) \left(\sum \alpha^2\right) + 2 \left(\alpha^2 + \gamma^2\right) \left(\beta^2 + \delta^2\right) - 2(\alpha\beta + \gamma\delta)^2 \\
&\qquad - \left(\sum \alpha^2\right) + 2(\alpha\delta - \beta\gamma) = 0 \\
\implies &(\beta\gamma - \alpha\delta - 1) \left(\sum \alpha^2\right) + 2(\beta\gamma - \alpha\delta)^2 - 2(\beta\gamma - \alpha\delta) = 0 \\
\implies &(\beta\gamma - \alpha\delta - 1) \left(\sum \alpha^2\right) + 2(\beta\gamma - \alpha\delta)(\beta\gamma - \alpha\delta - 1) = 0 \\
\implies &(\beta\gamma - \alpha\delta - 1) \left((\alpha - \delta)^2 + (\beta + \delta)^2 \right) = 0,
\end{aligned}$$

which implies

$$\beta\gamma - \alpha\delta = 1. \qquad (6.10)$$

Now we will use the condition that $\gamma(t) \in \tilde{X}(2)$ which was given by equations (6.3) and (6.4). From (6.3) we have

$$\begin{aligned}
&\|\mathbf{v}\|^2 \cos^2 + \left(\alpha^2 \|\mathbf{v}\|^2 + \beta^2 \|\mathbf{w}\|^2 + 2\alpha\beta \langle \mathbf{v}, \mathbf{w} \rangle\right) \sin^2 t \\
&\qquad + \left(\alpha \|\mathbf{v}\|^2 + \beta \langle \mathbf{v}, \mathbf{w} \rangle\right) \sin 2t = \frac{1}{2}
\end{aligned}$$

$$\begin{aligned}
\implies \sin^2 t\left(\alpha^2\left(1+\beta^2+\delta^2\right)+\beta^2\left(1+\alpha^2+\gamma^2\right)-2\alpha\beta(\alpha\beta+\gamma\delta)\right)& \\
+\cos^2 t\left(1+\beta^2+\delta^2\right)& \\
+\sin 2t\left(\alpha\left(1+\beta^2+\delta^2\right)-\beta(\alpha\beta+\gamma\delta)\right)=\frac{2+\sum\alpha^2}{2}& \\
\implies \left(\alpha^2+\beta^2+1\right)\sin^2 t+\left(\beta^2+\delta^2+1\right)\cos^2 t+(\alpha-\delta)\sin 2t=\frac{2+\sum\alpha^2}{2}& \\
\implies \alpha^2+\beta^2+\cos^2 t\left(\beta^2+\delta^2+1-\alpha^2-\beta^2+1\right)+(\alpha-\delta)\sin 2t=\frac{\sum\alpha^2}{2}& \\
\implies \left(\delta^2-\alpha^2\right)\cos^2 t+(\alpha-\delta)\sin 2t-\frac{\gamma^2+\delta^2-\alpha^2-b^2}{2}=0,&
\end{aligned}$$

which simplifies to

$$\left(\delta^2-\alpha^2\right)\cos 2t+2(\alpha-\delta)\sin 2t=\gamma^2-\beta^2.$$

Similarly, using (6.4) we have

$$-(\alpha+\delta)(\beta+\gamma)\cos 2t+2(\beta+\gamma)\sin 2t=(\alpha-\delta)(\beta-\gamma).$$

Writing the last two relation into matrix form we have

$$\begin{bmatrix}-(\alpha-\delta)(\alpha+\delta) & 2(\alpha-\delta)\\ -(\alpha+\delta)(\beta+\gamma) & 2(\beta+\gamma)\end{bmatrix}\begin{bmatrix}\cos 2t\\ \sin 2t\end{bmatrix}=\begin{bmatrix}\gamma^2-\beta^2\\ (\alpha-\delta)(\beta-\gamma)\end{bmatrix}. \tag{6.11}$$

Note that the rank of the coefficient matrix is 1 and this can occur if one row is linear multiple of the other.

- Let $\alpha=\delta$ and $\beta\neq-\gamma$. The system (6.11) has a solution, so $\beta-\gamma=0$. Note that

  $$(6.8)\implies \beta\left(\|\mathbf{v}\|^2-\|\mathbf{w}\|^2\right)=0\implies \beta=0 \text{ or } \|\mathbf{v}\|^2=\|\mathbf{w}\|^2.$$

  But, note that $\beta=0$ can not be possible because if $\beta=0=\gamma$, then using (6.10) $\alpha^2=-1$, a contradiction. Therefore, $\|\mathbf{v}\|^2=\frac{1}{2}=\|\mathbf{w}\|^2$. Again using (6.9),

  $$\langle\mathbf{v},\mathbf{w}\rangle=\frac{-\alpha}{2\beta}.$$

  Since

  $$\begin{aligned}\|\mathbf{v}_1\|^2+\|\mathbf{v}_2\|^2=1 &\implies \frac{(\alpha^2+\beta^2)}{2}+4\alpha\beta\left(\frac{-\alpha}{2\beta}\right)+\frac{(\alpha^2+\beta^2)}{2}=1\\ &\implies \beta^2-\alpha^2=1.\end{aligned}$$

  Consider

  $$\begin{aligned}\langle\mathbf{v},\mathbf{w}\rangle=-\frac{\alpha}{2\beta} &\implies 4\beta^2\langle\mathbf{v},\mathbf{w}\rangle^2=\alpha^2=\beta^2-1\\ &\implies \beta^2=\frac{1}{1-4\langle\mathbf{v},\mathbf{w}\rangle^2},\ \alpha^2=\frac{4\langle\mathbf{v},\mathbf{w}\rangle^2}{1-4\langle\mathbf{v},\mathbf{w}\rangle^2}.\end{aligned}$$

Note that the above expression is valid as $\mathbf{v}$ and $\mathbf{w}$ are linearly independent and each has norm $\frac{1}{\sqrt{2}}$. Thus, if $p \in \widetilde{\mathrm{Se}}$, then the only two possible directions are $\upsilon_1 = (\mathbf{v}_1, \mathbf{v}_2)$ and $-\upsilon = (-\mathbf{v}_1, -\mathbf{v}_2)$. If γ and η be two geodesics in υ_1 and $-\upsilon_1$ direction respectively, then they intersect $\tilde{X}(2)$ at t and $\pi - t$ respectively. As their lengths are same, and they are unit speed geodesics, so $t = \pi - t$ which implies $t = \frac{\pi}{2}$. This implies $(\mathbf{v}_1, \mathbf{v}_2) \in \tilde{X}(2)$ which means

$$\begin{aligned}\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0 &\implies \frac{\alpha\beta}{2} + \frac{\alpha\beta}{2} + 2\alpha\beta \langle \mathbf{v}, \mathbf{w} \rangle = 0 \\ &\implies \alpha\beta(2 + \langle \mathbf{v}, \mathbf{w} \rangle) = 0 \\ &\implies \alpha = 0.\end{aligned}$$

If $\alpha = 0 = \delta$, then $\beta^2 = 1$ which implies $\beta = \pm 1$, and hence $(\mathbf{v}, \mathbf{w}) \in \tilde{X}(2)$, which is a contradiction.

- Let $\beta + \gamma = 0$ or $\alpha - \delta \neq 0$. But $\alpha - \delta \neq 0$, so $\beta = \gamma = 0$ which implies $\mathbf{v}_1 = \alpha\mathbf{v}$ and $\mathbf{v}_2 = \delta\mathbf{w}$. Observe that

$$\begin{aligned}(6.8) &\implies \langle \mathbf{v}, \mathbf{w} \rangle = 0 \implies \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0. \\ (6.1) &\implies \alpha^2 \|\mathbf{v}\|^2 + \delta^2 \|\mathbf{w}\|^2 = 1 \\ &\implies \|\mathbf{w}\|^2 \left(\alpha^2 + \delta^2\right) = 1 - \alpha^2 \\ &\implies \|\mathbf{w}\|^2 = \frac{1 - \alpha^2}{\alpha^2 + \delta^2}, \text{ and } \|\mathbf{v}\|^2 = \frac{\delta^2 - 1}{\alpha^2 + \delta^2}.\end{aligned}$$

Now we use the condition that $\alpha\delta = -1$ to obtain,

$$\alpha^2 = \frac{\|\mathbf{w}\|^2}{\|\mathbf{v}\|^2}, \text{ and } \delta^2 = \frac{\|\mathbf{v}\|^2}{\|\mathbf{w}\|^2},$$

which is again fixed and there are only two possible directions υ_1 and $-\upsilon_1$ and hence this case is also not possible.

- Finally, both row are non-zero and let $\alpha + \delta = \lambda(\beta + \gamma)$. So (6.11) become

$$\begin{aligned}&\begin{bmatrix} -(\alpha+\delta)\lambda(\beta+\gamma) & 2\lambda(\beta+\gamma) \\ -(\alpha+\delta)(\beta+\gamma) & 2(\beta+\gamma) \end{bmatrix} \begin{bmatrix} \cos 2t \\ \sin 2t \end{bmatrix} = \begin{bmatrix} -(\beta^2 - \gamma^2) \\ \lambda(\beta^2 - \gamma^2) \end{bmatrix} \\ \implies &\begin{bmatrix} -(\alpha+\delta)\lambda & 2\lambda \\ -(\alpha+\delta) & 2 \end{bmatrix} \begin{bmatrix} \cos 2t \\ \sin 2t \end{bmatrix} = \begin{bmatrix} -(\beta+\gamma) \\ \lambda(\beta+\gamma) \end{bmatrix},\end{aligned}$$

which implies $\lambda^2 = -1$, a contradiction.

Thus, we have proved that $\mathrm{rank}[\mathbf{v}, \mathbf{w}, \mathbf{v}_1, \mathbf{v}_2] \neq 2$.

Case 2: $\mathrm{rank}[\mathbf{v}, \mathbf{w}, \mathbf{v}_1, \mathbf{v}_2] = 3$: Since rank is 3, without loss of generality we assume that $\mathbf{v}_2 \in \mathrm{Span}\{\mathbf{v}, \mathbf{w}, \mathbf{v}_1\}$. Let us write $\mathbf{v}_2 = a\mathbf{v} + b\mathbf{w} + c\mathbf{v}_1$ for some $a, b, c \in \mathbb{R}$. Since the rank is three we can assume that all the vectors are in $\mathbb{R}^3$ and let $\times$ denote the vector cross product. Let

$$\begin{aligned}\upsilon_1 &= (\mathbf{v}\cos t + \mathbf{v}_1 \sin t)) \times (\mathbf{w}\cos t + \mathbf{v}_2 \sin t) \\ &= (\mathbf{v} \times \mathbf{w})\cos^2 t + (\mathbf{v} \times \mathbf{v}_2)\cos t \sin t + (\mathbf{v}_1 \times \mathbf{w}\cos t \sin t) + (\mathbf{v}_1 \times \mathbf{v}_2)\sin^2 t.\end{aligned}$$

Using (6.6), we have

$$v_1 \cdot (-\mathbf{v}\sin t + \mathbf{v}_1 \cos t) = 0 \implies [\mathbf{v}, \mathbf{w}, \mathbf{v}_1]\cos t + [\mathbf{v}_1, \mathbf{v}, \mathbf{v}_2]\sin t = 0 \tag{6.12}$$

$$v_1 \cdot (-\mathbf{v}\sin t + \mathbf{v}_2 \cos t) = 0 \implies [\mathbf{v}, \mathbf{w}, \mathbf{v}_2]\cos t + [\mathbf{v}_1, \mathbf{w}, \mathbf{v}_2]\sin t = 0. \tag{6.13}$$

Note that

$$[\mathbf{v}_1, \mathbf{v}, \mathbf{v}_2] = [\mathbf{v}_1, \mathbf{v}, a\mathbf{v} + b\mathbf{w} + c\mathbf{v}_1] = b\,(\mathbf{v}, \mathbf{w}, \mathbf{v}_1)\,.$$

So from (6.12),

$$[\mathbf{v}, \mathbf{w}, \mathbf{v}_1]\cos t + b\,[\mathbf{v}, \mathbf{w}, \mathbf{v}_1]\sin t = 0 \implies \cos t + b\sin t = 0. \tag{6.14}$$

Above implies,

$$\cos t = \frac{-b}{1+b^2}, \text{ and } \sin t = \frac{1}{1+b^2}.$$

Similarly, using (6.13),

$$\begin{aligned}
&[\mathbf{v}, \mathbf{w}, a\mathbf{v} + b\mathbf{w} + c\mathbf{v}_1]\cos t + [\mathbf{v}_1, \mathbf{w}, a\mathbf{v} + b\mathbf{w} + c\mathbf{v}_1]\sin t = 0 \\
\implies & c\,(\mathbf{v}, \mathbf{w}, \mathbf{v}_1)\cos t + a\,[\mathbf{v}_1, \mathbf{w}, \mathbf{v}]\sin t = 0 \\
\implies & [\mathbf{v}, \mathbf{w}, \mathbf{v}_1]\,(c\cos t - a\sin t) = 0 \\
\implies & c\cos t = a\sin t.
\end{aligned} \tag{6.15}$$

Using (6.14) and (6.15) we obtain

$$a + bc = 0.$$

Now we collect some more conditions using previous conditions.

$$\begin{aligned}
(6.5) &\implies a\,\|\mathbf{v}\|^2 + b\,\langle \mathbf{v}, \mathbf{w}\rangle + c\,\langle \mathbf{v}, \mathbf{v}_1\rangle - \langle \mathbf{v}_1, \mathbf{w}\rangle = 0 \\
&\implies -bc\,\|\mathbf{v}\|^2 + b\,\langle \mathbf{v}, \mathbf{w}\rangle + c\,\langle \mathbf{v}_1, \mathbf{v}\rangle - \langle \mathbf{v}_1, \mathbf{w}\rangle = 0 \\
&\implies (\mathbf{v}_1 - b\mathbf{v}) \cdot (\mathbf{w} - c\mathbf{v}) = 0
\end{aligned} \tag{6.16}$$

$$\begin{aligned}
(6.2) &\implies \langle \mathbf{v}, \mathbf{v}_1\rangle + a\,\langle \mathbf{v}, \mathbf{w}\rangle + b\,\|\mathbf{w}\|^2 + c\,\langle \mathbf{v}_1, \mathbf{w}\rangle = 0 \\
&\implies \langle \mathbf{v}, \mathbf{v}_1\rangle - bc\,\langle \mathbf{v}, \mathbf{w}\rangle + b\,\|w\|^2 + c\,\langle \mathbf{v}, \mathbf{v}_1\rangle = 0 \\
&\implies \mathbf{v}_1 \cdot (\mathbf{v} - c\mathbf{w}) = b\mathbf{w} \cdot (c\mathbf{v} - \mathbf{w}).
\end{aligned} \tag{6.17}$$

Multiply (6.16) by c and add to (6.17) to obtain

$$\langle \mathbf{v}_1, \mathbf{v}\rangle = \frac{bc^2}{1+c^2}\,\|\mathbf{v}\|^2 - \frac{b}{1+c^2}\,\|\mathbf{w}\|^2]. \tag{6.18}$$

Similarly,

$$\langle \mathbf{v}_1, \mathbf{w}\rangle = \frac{-bc}{1+c^2} + b\,\langle \mathbf{v}, \mathbf{w}\rangle\,. \tag{6.19}$$

We use (6.3) and substitute the value of $\cos t, \sin t$ and use (6.19) and (6.18) to obtain

$$\|\mathbf{v}_1\|^2 = \frac{b^2\,(c^2-1)}{1+c^2}\,\|\mathbf{v}\|^2 - \frac{2b^2}{1+c^2}\,\|\mathbf{w}\|^2 + \frac{1+b^2}{2}.$$

Now use $\|\gamma(t)\|^2 = 1$,

$$\begin{aligned}
&\frac{b^2\left(c^2-1\right)}{1+c^2}\|\mathbf{v}\|^2 + \frac{b^2\left(c^2-1\right)}{1+c^2}\|\mathbf{w}\|^2 + \frac{b^2\left(c^2-1\right)^2+\left(c^2+1\right)^2}{2\left(1+c^2\right)} = 1\\
\implies &\frac{b^2\left(c^2-1\right)}{1+c^2}\left(\|\mathbf{v}\|^2+\|\mathbf{w}\|^2\right) + \frac{b^2\left(c^2-1\right)^2+\left(c^2+1\right)^2}{2\left(1+c^2\right)} = 1\\
\implies &2b^2\left(c^2-1\right)+b^2\left(c^2-1\right)^2 = 2\left(1+c^2\right)-\left(1+c^2\right)^2\\
\implies &\left(c^2-1\right)\left(2b^2+b^2c^2-b^2\right) = \left(1+c^2\right)\left(1-c^2\right)\\
\implies &\left(c^2-1\right)\left(b^2+b^2c^2+1+c^2\right) = 0\\
\implies &c = \pm 1.
\end{aligned}$$

We also have

$$\begin{aligned}
&(\mathbf{v}\cos t + \mathbf{v}_1 \sin t)\cdot(\mathbf{w}\cos t + (a\mathbf{v}+b\mathbf{w}+c\mathbf{v}_1)\sin t) = 0\\
\implies &(-b\mathbf{v}+\mathbf{v}_1)\cdot(-b\mathbf{w}+a\mathbf{v}+b\mathbf{w}+c\mathbf{v}_1) = 0\\
\implies &b^2c\|\mathbf{v}\|^2 - 2bc\langle\mathbf{v},\mathbf{v}_1\rangle + c\|\mathbf{v}_1\|^2 = 0\\
\implies &\frac{c\left(1+b^2\right)}{2} = 0\\
\implies &c = 0,
\end{aligned}$$

which is a contradiction. Hence, the rank can not be 3. Therefore, $\mathbf{v}$ and $\mathbf{w}$ are linearly dependent and hence, $(\mathbf{v},\mathbf{w}) \in \mathbb{S}^n \times_{\mathbb{Z}_2} \mathbb{S}^1$. □

We now prove that the cut locus of $\tilde{X}(d)$ when $n=1$ will be $(\mathbb{Z}_d \star \mathbb{Z}_d) \times_{\mathbb{Z}_d} \mathbb{S}^1$.

Note 6.3.1 (Cut locus of $X(2)$ for $n=1$)**.** For $n=1$, the cut locus of $\tilde{X}(2)$ is

$$\mathrm{Cu}(\tilde{X}(2)) = \left\{\frac{1}{2}(\cos s+\sin t, \sin s+\cos t, \sin s-\cos t, -\cos s+\sin t) : s,t\in\mathbb{R}\right\}.$$

Moreover, the cut locus of $X(2)$ will be $\mathbb{RP}^1$.

Proof. Let us write $z_j = x_j + \iota y_j$. Note that

$$\begin{aligned}
\tilde{X}(2) &= \left\{(z_0,z_1)\in\mathbb{C}^2 : z_0^2+z_1^2 = 0, \text{ and } |z_0|^2+|z_1|^1 = 1\right\}\\
&= \left\{(x_0,y_0,x_1,y_1) : x_0^2+x_1^2 = \frac{1}{2} = y_0^2+y_1^2, \text{ and } x_0y_0+x_1y_1 = 0\right\}.
\end{aligned}$$

Since (x_0,x_1) and (y_0,y_1) can not be zero vectors, so without loss of generality, we assume that $x_0y_1 \neq 0$. Since

$$x_0y_0+x_1y_1 = 0 \implies \frac{y_0}{y_1}+\frac{x_1}{x_0} = 0 \implies y_0 = -y_1\left(\frac{x_1}{x_0}\right).$$

Now as

$$y_0^2+y_1^2 = \frac{1}{2} \implies y_1^2\left(\frac{x_1}{x_0}\right)^2 + y_1^2 = \frac{1}{2} \implies y_1^2\left(\frac{x_1^2}{x_0^2}+1\right) = \frac{1}{2} \implies y_1 = \pm x_0.$$

Similarly, we have

$$x_1 = \pm y_0.$$

Therefore,

$$\begin{aligned}\tilde{X}(2) &= \left\{(x_0, y_0, y_0, -x_0) : x_0^2 + y_0^2 = \frac{1}{2}\right\} \sqcup \left\{(x_0, -y_0, y_0, x_0) : x_0^2 + y_0^2 = \frac{1}{2}\right\} \\ &= S_1^1 \sqcup S_2^1.\end{aligned}$$

Define a linear transformation

$$T : \mathbb{R}^4 \to \mathbb{R}^4,\ (a, b, c, d) \mapsto \frac{1}{\sqrt{2}}(a - d, b + c, b - c, a + d).$$

Note that T maps $\mathbb{S}^3$ onto $\mathbb{S}^3$ and is an isometry hence it preserves the cut locus. So

$$\begin{aligned}\mathrm{Cu}(T(\tilde{X}(2))) &= \mathrm{Cu}\left(T\left(S_1^1\right) \sqcup T\left(S_2^1\right)\right) \\ &= \mathrm{Cu}\left(\left\{(a, b, 0, 0) : a^2 + b^2 = 1\right\} \sqcup \left\{(0, 0, c, d) : c^2 + d^2 = 1\right\}\right) \\ &= \left\{\frac{1}{\sqrt{2}}(\cos s, \sin s, \cos t, \sin t) : s, t \in \mathbb{R}\right\}.\end{aligned}$$

Therefore, the cut locus of $X(2)$ can be found by the inverse transformation which is

$$T^{-1}(x, y, z, w) = \frac{1}{2}(x + 2, y + z, y - z, w - x).$$

Hence,

$$\begin{aligned}\mathrm{Cu}(\tilde{X}(2)) &= \left\{\frac{1}{2}(\cos s + \sin t, \sin s + \cos t, \sin s - \cos t, -\cos s + \sin t) : s, t \in \mathbb{R}\right\} \\ &\cong \mathbb{S}^1 \times \mathbb{S}^1.\end{aligned}$$

Quotient with $\mathbb{S}^1$ give the required cut locus. □

Theorem 6.3.2. For $n = 1$, we have

$$\widetilde{\mathrm{Cu}} = (\mathbb{Z}_d \star \mathbb{Z}_d) \times_{\mathbb{Z}_d} \mathbb{S}^1.$$

Proof. Note that

$$\mathbb{Z}_d \star \mathbb{Z}_d = \left\{\mathbf{v} = (\xi^k \cos\phi, \xi^l \sin\phi) : 0 \le k, l \le d - 1, \text{ and } 0 \le \phi \le \frac{\pi}{2}\right\},$$

where ξ^k is a d^{th} root of unity. Let $\mathbf{v} \in \mathbb{Z}_d \star \mathbb{Z}_d$ and $\theta \in (0, 2\pi)$. We will show that $\mathbf{v}e^{\iota\theta} \in \widetilde{\mathrm{Se}}$. Due to $\mathbb{S}^1$-symmetry, it is enough to show that $\mathbf{v} \in \widetilde{\mathrm{Se}}$. Now consider the matrix

$$A = \begin{pmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{pmatrix} \in U(2), \text{ such that } \lambda_1^d = 1 = \lambda_2^d.$$

We observe that A maps $\tilde{X}(d)$ to itself and hence it is enough to show that $p = (\cos\phi, 0, \sin\phi, 0) \in \widetilde{\mathrm{Se}}$. Note that

$$\tilde{X}(d) = \left\{(z_1, z_2) \in \mathbb{S}^3 : z_1^d + z_2^d = 0\right\}$$

$$
\begin{aligned}
&= \left\{(z_1, z_2) \in \mathbb{S}^3 : z_1^d = -z_2^d,\ |z_1| = \frac{1}{\sqrt{2}} = |z_2|\right\} \\
&= \left\{(z, \xi z) \in \mathbb{S}^3 : \xi^d = -1, \text{ and } |z| = \frac{1}{\sqrt{2}}\right\} \\
&= \bigsqcup_{k=0}^{d-1} \left\{(z, \xi z) : \xi = e^{\frac{(2k+1)\pi}{d}},\ |z| = \frac{1}{\sqrt{2}}\right\} = \bigsqcup_{k=0}^{d-1} X_k
\end{aligned}
$$

We now compute the distance of p from X_k and will show that the distance is same with at least two components, which will show that $p \in \widetilde{\mathrm{Se}}$. Note that for any point $Q = (x_1, y_1, x_2, y_2) \in \tilde{X}(d)$, the distance between p and Q is given by (look at the Figure 6.3)

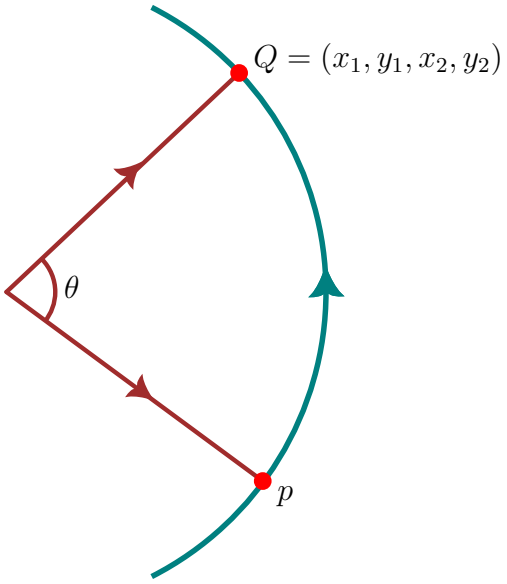


Figure 6.3: distance of p to $\tilde{X}(d)$

$$\mathrm{dist}(p, Q) = \cos^{-1}(p \cdot Q) = \cos^{-1}\left(x_1 \cos\phi + x_2 \sin\phi\right).$$

Therefore, the distance between p and the set $\tilde{X}(d)$ is given by

$$
\begin{aligned}
\mathrm{dist}(p, \tilde{X}(d)) &= \inf\left\{\mathrm{dist}(p, Q) : Q \in \tilde{X}(d)\right\} \\
&= \inf\left\{\cos^{-1}\left(x_1 \cos\phi + x_2 \sin\phi\right) : (x_1, y_1, x_2, y_2) \in \tilde{X}(d)\right\}.
\end{aligned}
$$

As $\cos^{-1}$ is a decreasing function, it is equivalent to maximize $x_1 \cos\phi + x_2 \sin\phi$ such that $x_2 = x_1 \cos\left(\frac{(2k+1)\pi}{d}\right) - y_1 \sin\left(\frac{(2k+1)\pi}{d}\right)$, and $x_1^2 + y_1^2 = \frac{1}{2}$. This maximum value will be

$$\sqrt{\frac{1 + \sin 2\phi \cos\left(\frac{(2k+1)\pi}{d}\right)}{2}}.$$

Therefore, the distance from X_k will be

$$\cos^{-1}\left(\sqrt{\frac{1 + \sin 2\phi \cos\left(\frac{(2k+1)\pi}{d}\right)}{2}}\right).$$

Note that if d is even, then the above distance is same from X_k and X_{d-1-k}, therefore, the point is a separating point and hence is a cut point. If d is odd, then the above still holds. The only thing to make sure that the distance from each component is smaller than the distance from $X_{\frac{d-1}{2}}$, but this is true as $0 \leq \phi \leq \frac{\pi}{2}$. Therefore, we proved that $p \in \widetilde{\mathrm{Se}}$ and hence it is in $\widetilde{\mathrm{Cu}}$. The inverse inclusion follows with a similar argument. In fact, if we take any other point, then there will be only one distance minimal geodesic, that will occur from one component. Let $p = \left(r_1 e^{\iota\theta_1}, r_2 e^{\iota\theta_2}\right)$ be any point in $\mathbb{S}^3$ not in the above form. Note that if $\theta_1 = \theta_2$, then due to $\mathbb{S}^1$-equivariant, $p \in \widetilde{\mathrm{Se}}$ if and only if $(r_1, 0, r_2, 0) \in \widetilde{\mathrm{Se}}$, which is a point in the above form. So we assume that $\theta_1 \neq \theta_2$ and $\theta_1, \theta_2 \in [0, 2\pi/d]$. Then we have to minimize the distance from p to $\tilde{X}(d)$, that is

$$\mathrm{dist}(p, \tilde{X}(d)) = \inf\left\{\cos^{-1}(p \cdot (z_1, z_2)) : (z_1, z_2) \in \tilde{X}(d)\right\}.$$

This is equivalent to maximizing the dot product $p \cdot (z_1, z_2)$, and this maximum will be achieved from one component, say X_l, of $\tilde{X}(d)$ as $0 < |\theta_1 - \theta_2| < 2\pi/d$. A similar computation shows that there is only one distance minimal geodesic from the component X_l, and this shows that $p \notin \widetilde{\mathrm{Se}}$. Hence, the theorem is proved. □

Remark 6.3.1. In [Audin, 2005, §2.2], the author has shown that the function

$$f : \mathbb{CP}^n \to \mathbb{R},\ [x + \iota y] \mapsto \frac{\|y\|^2}{\left(1 + \|y\|^2\right)^2}$$

is a Morse-Bott function with two critical submanifolds – $X(2)$ and $\mathbb{RP}^n$. We know that $\mathbb{CP}^n \setminus X(2)$ deformation retracts to $\mathbb{RP}^n$ via Morse-Bott flow, whereas Lemma 4.3.1 implies that $\mathbb{CP}^n \setminus X(2)$ deformation retracts to $\mathrm{Cu}(X(2))$. Hence, $\mathrm{Cu}(X(2))$ and $\mathbb{RP}^n$ have the same homotopy type. However, it's not a priori clear whether they are equal. Moreover, our calculation shows a computation of $\mathrm{Cu}(T_1\mathbb{S}^n) \subseteq \mathbb{S}^{2n+1}$, where $T_1\mathbb{S}^n$ is the unit tangent bundle of $\mathbb{S}^n$. We deduce that the cut locus is the unique non-trivial $\mathbb{S}^1$-bundle of $\mathbb{RP}^n$ and this is a new computation.

Bibliography

[Angulo Ardoy and Guijarro, 2011] Angulo Ardoy, P. and Guijarro, L. (2011). Balanced split sets and Hamilton-Jacobi equations. *Calc. Var. Partial Differential Equations*, 40(1-2):223–252.

[Audin, 2005] Audin, M. (2005). On the topology of Lagrangian submanifolds. Examples and counter-examples. *Port. Math. (N.S.)*, 62(4):375–419.

[Banyaga and Hurtubise, 2004] Banyaga, A. and Hurtubise, D. (2004). *Lectures on Morse homology*, volume 29 of *Kluwer Texts in the Mathematical Sciences*. Kluwer Academic Publishers Group, Dordrecht.

[Barratt and Milnor, 1962] Barratt, M. G. and Milnor, J. (1962). An example of anomalous singular homology. *Proc. Amer. Math. Soc.*, 13:293–297.

[Basu and Prasad, 2021] Basu, S. and Prasad, S. (2021). A connection between cut locus, Thom space and Morse-Bott functions. available at https://arxiv.org/abs/2011.02972, to appear in Algebraic & Geometric Topology.

[Besse, 1978] Besse, A. L. (1978). *Manifolds all of whose geodesics are closed*, volume 93 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin-New York. With appendices by D. B. A. Epstein, J.-P. Bourguignon, L. Bérard-Bergery, M. Berger and J. L. Kazdan.

[Bott, 1954] Bott, R. (1954). Nondegenerate critical manifolds. *Ann. of Math. (2)*, 60:248–261.

[Buchner, 1977] Buchner, M. A. (1977). Simplicial structure of the real analytic cut locus. *Proc. Amer. Math. Soc.*, 64(1):118–121.

[Busemann, 1955] Busemann, H. (1955). *The geometry of geodesics*. Academic Press Inc., New York, N. Y.

[Cheeger and Ebin, 1975] Cheeger, J. and Ebin, D. G. (1975). *Comparison theorems in Riemannian geometry*. North-Holland Mathematical Library, Vol. 9. North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York.

[Daniilidis et al., 2018] Daniilidis, A., Deville, R., Durand-Cartagena, E., and Rifford, L. (2018). Self-contracted curves in Riemannian manifolds. *J. Math. Anal. Appl.*, 457(2):1333–1352.

[do Carmo, 1992] do Carmo, M. P. a. (1992). *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA. Translated from the second Portuguese edition by Francis Flaherty.

[Fermi, 1922] Fermi, E. (1922). Sopra i fenomeni che avvengono in vicinanza di una linea oraria. *Rend. Lincei*, 31:21–23.

[Flaherty, 1965] Flaherty, F. J. (1965). *The cut locus of a submanifold*. PhD thesis. Thesis (Ph.D.)–University of California, Berkeley.

[Gluck and Singer, 1978] Gluck, H. and Singer, D. (1978). Scattering of geodesic fields. I. *Ann. of Math. (2)*, 108(2):347–372.

[Gray, 2004] Gray, A. (2004). *Tubes*, volume 221 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, second edition. With a preface by Vicente Miquel.

[Hatcher, 2002] Hatcher, A. (2002). *Algebraic topology*. Cambridge University Press, Cambridge.

[Hebda, 1983] Hebda, J. J. (1983). The local homology of cut loci in Riemannian manifolds. *Tohoku Math. J. (2)*, 35(1):45–52.

[Hebda, 1995] Hebda, J. J. (1995). Cut loci of submanifolds in space forms and in the geometries of Möbius and Lie. *Geom. Dedicata*, 55(1):75–93.

[Higham, 2008] Higham, N. J. (2008). *Functions of matrices*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Theory and computation.

[Hirsch, 1976] Hirsch, M. W. (1976). *Differential topology*. Graduate Texts in Mathematics, No. 33. Springer-Verlag, New York-Heidelberg.

[Hopf and Rinow, 1931] Hopf, H. and Rinow, W. (1931). Ueber den Begriff der vollständigen differentialgeometrischen Fläche. *Comment. Math. Helv.*, 3(1):209–225.

[Itoh and Sabau, 2016] Itoh, J.-i. and Sabau, S. V. (2016). Riemannian and Finslerian spheres with fractal cut loci. *Differential Geom. Appl.*, 49:43–64.

[Itoh and Vîlcu, 2011] Itoh, J.-i. and Vîlcu, C. (2011). Orientable cut locus structures on graphs. available at https://arxiv.org/abs/1103.3136.

[Itoh and Vîlcu, 2015] Itoh, J.-i. and Vîlcu, C. (2015). Every graph is a cut locus. *J. Math. Soc. Japan*, 67(3):1227–1238.

[Klingenberg, 1959] Klingenberg, W. (1959). Contributions to Riemannian geometry in the large. *Ann. of Math. (2)*, 69:654–666.

[Kobayashi, 1967] Kobayashi, S. (1967). On conjugate and cut loci. In *Studies in Global Geometry and Analysis*, pages 96–122. Math. Assoc. Amer. (distributed by Prentice-Hall, Englewood Cliffs, N.J.).

[Kulkarni and Wood, 1980] Kulkarni, R. S. and Wood, J. W. (1980). Topology of nonsingular complex hypersurfaces. *Adv. in Math.*, 35(3):239–263.

[Lang, 1999] Lang, S. (1999). *Fundamentals of differential geometry*, volume 191 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.

[Lee, 2013] Lee, J. M. (2013). *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, second edition.

[Li and Nirenberg, 2005] Li, Y. and Nirenberg, L. (2005). The distance function to the boundary, Finsler geometry, and the singular set of viscosity solutions of some Hamilton-Jacobi equations. *Comm. Pure Appl. Math.*, 58(1):85–146.

[Manasse and Misner, 1963] Manasse, F. K. and Misner, C. W. (1963). Fermi normal coordinates and some basic concepts in differential geometry. *J. Mathematical Phys.*, 4:735–745.

[Mantegazza and Mennucci, 2003] Mantegazza, C. and Mennucci, A. C. (2003). Hamilton-Jacobi equations and distance functions on Riemannian manifolds. *Appl. Math. Optim.*, 47(1):1–25.

[Martin and Neff, 2016] Martin, R. J. and Neff, P. (2016). Minimal geodesics on $\mathrm{GL}(n)$ for left-invariant, right-$\mathrm{O}(n)$-invariant Riemannian metrics. *J. Geom. Mech.*, 8(3):323–357.

[McKay, 2015] McKay, B. (2015). A summary of progress on the Blaschke conjecture. *ICCM Not.*, 3(2):33–45.

[Michor, 2008] Michor, P. W. (2008). *Topics in differential geometry*, volume 93 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI.

[Milnor, 1963] Milnor, J. (1963). *Morse theory*. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J. Based on lecture notes by M. Spivak and R. Wells.

[Myers, 1935] Myers, S. B. (1935). Connections between differential geometry and topology. I. Simply connected surfaces. *Duke Math. J.*, 1(3):376–391.

[Myers, 1936] Myers, S. B. (1936). Connections between differential geometry and topology II. Closed surfaces. *Duke Math. J.*, 2(1):95–102.

[Omori, 1968] Omori, H. (1968). A class of Riemannian metrics on a manifold. *J. Differential Geometry*, 2:233–252.

[Ozols, 1974] Ozols, V. (1974). Cut loci in Riemannian manifolds. *Tohoku Math. J. (2)*, 26:219–227.

[Plotnick, 1982] Plotnick, S. (1982). Embedding homology 3-spheres in S^5. *Pacific J. Math.*, 101(1):147–151.

[Poincaré, 1905] Poincaré, H. (1905). Sur les lignes géodésiques des surfaces convexes. *Trans. Amer. Math. Soc.*, 6(3):237–274.

[Rauch, 1959] Rauch, H. E. (1959). *Geodesics and curvature in differential geometry in the large*, volume 1. Yeshiva University, Graduate School of Mathematical Sciences.

[Sabau and Tanaka, 2016] Sabau, S. V. and Tanaka, M. (2016). The cut locus and distance function from a closed subset of a Finsler manifold. *Houston J. Math.*, 42(4):1157–1197.

[Sakai, 1977] Sakai, T. (1977). On cut loci of compact symmetric spaces. *Hokkaido Math. J.*, 6(1):136–161.

[Sakai, 1978] Sakai, T. (1978). On the structure of cut loci in compact Riemannian symmetric spaces. *Math. Ann.*, 235(2):129–148.

[Sakai, 1979] Sakai, T. (1979). Cut loci of compact symmetric spaces. In *Minimal submanifolds and geodesics (Proc. Japan-United States Sem., Tokyo, 1977)*, pages 193–207. North-Holland, Amsterdam-New York.

[Sakai, 1984] Sakai, T. (1984). Comparison and finiteness theorems in Riemannian geometry. In *Geometry of geodesics and related topics (Tokyo, 1982)*, volume 3 of *Adv. Stud. Pure Math.*, pages 125–181. North-Holland, Amsterdam.

[Sakai, 1996] Sakai, T. (1996). *Riemannian geometry*, volume 149 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI. Translated from the 1992 Japanese original by the author.

[Sharafutdinov, 2007] Sharafutdinov, V. A. (2007). Proof of soul theorem (translated with some modifications).

[Singer and Gluck, 1976] Singer, D. and Gluck, H. (1976). The existence of non-triangulable cut loci. *Bull. Amer. Math. Soc.*, 82(4):599–602.

[Singh, 1987a] Singh, H. (1987a). On the cut locus and the focal locus of a submanifold in a Riemannian manifold. *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.*, 30:139–144.

[Singh, 1987b] Singh, H. (1987b). On the cut locus and the focal locus of a submanifold in a Riemannian manifold. II. *Publ. Inst. Math. (Beograd) (N.S.)*, 41(55):119–124.

[Singh, 1988] Singh, H. (1988). Closest point of the cut locus to submanifold. *Proc. Indian Acad. Sci. Math. Sci.*, 98(2-3):179–186.

[Takeuchi, 1978] Takeuchi, M. (1978). On conjugate loci and cut loci of compact symmetric spaces. I. *Tsukuba J. Math.*, 2:35–68.

[Takeuchi, 1979] Takeuchi, M. (1979). On conjugate loci and cut loci of compact symmetric spaces. II. *Tsukuba J. Math.*, 3(1):1–29.

[Thom, 1972] Thom, R. (1972). Sur le cut-locus d'une variété plongée. *J. Differential Geometry*, 6:577–586.

[von Mangoldt, 1881] von Mangoldt, H. (1881). Ueber diejenigen Punkte auf positiv gekrümmten Flächen, welche die Eigenschaft haben, dass die von ihnen ausgehenden geodätischen Linien nie aufhören, kürzeste Linien zu sein. *J. Reine Angew. Math.*, 91:23–53.

[Wall, 1977] Wall, C. T. C. (1977). Geometric properties of generic differentiable manifolds. In *Geometry and topology (Proc. III Latin Amer. School of Math., Inst. Mat. Pura Aplicada CNPq, Rio de Janeiro, 1976)*, Lecture Notes in Math., Vol. 597, pages 707–774. Springer, Berlin.

[Whitehead, 1935] Whitehead, J. H. C. (1935). On the covering of a complete space by the geodesics through a point. *Ann. of Math. (2)*, 36(3):679–704.

[Wolter, 1979] Wolter, F.-E. (1979). Distance function and cut loci on a complete Riemannian manifold. *Arch. Math. (Basel)*, 32(1):92–96.

[Yakubovich and Starzhinskii, 1975] Yakubovich, V. A. and Starzhinskii, V. M. (1975). *Linear differential equations with periodic coefficients. 1, 2*. Halsted Press [John Wiley & Sons] New York-Toronto, Ont.,; Israel Program for Scientific Translations, Jerusalem-London. Translated from Russian by D. Louvish.

Index