RIEMANN SURFACES IN MACHINE LEARNING

ARINA OBEROI

Abstract. This paper provides updated expositions of classical mathematical concepts like Riemann surfaces, Jensen's Inequality, and Hoeffding's Inequality, linking them to modern machine learning applications. It bridges the gap between historical mathematical theories and contemporary data science challenges, refreshing older concepts with current relevance. Beginning with a historical overview of Riemann surfaces and their role in complex analysis, the paper explores key geometric properties like conformal maps and the Riemann Mapping Theorem. It then examines the application of these concepts in machine learning, including variational inference, shape matching, and phase imaging. By integrating Riemann surfaces with deep learning, optimal transport, and kernel methods, the paper addresses challenges with multi-valued complex functions and non-Euclidean data, demonstrating their impact on improving accuracy in computer vision, probabilistic modeling, and data analysis.

Contents

1. Historical Context

In his 1851 doctoral thesis, Grundlagen für eine allgemeine Theorie der Funktionen einer veränderlichen complexen Größe ("Foundations for a General Theory of Functions of a Complex Variable"), Bernhard Riemann, a 19th-century German mathematician, pioneered the study of Riemann surfaces by not only introducing these surfaces but applying them to address complex analysis problems. In doing so, he made invaluable progress in solving questions which had been left unsolved

Date: August 2024.

for centuries due to the inability to project complex functions onto singular surfaces. This work has laid the groundwork for core pillars of modern mathematics, from the Riemann integral to the Riemann Hypothesis.

2. Riemann Surfaces

2.1. **Geometric Foundations and Transformations.** Riemann's introduction to Riemann surfaces begins by describing how complex functions can be mapped onto multi-sheeted surfaces, aka the Riemann Mapping Theorem.

Theorem 2.1 (Riemann Mapping Theorem)**.** *Any simply connected open subset of the complex plane $\mathbb{C}$, which is not the entire plane, is conformally equivalent to the open unit disk.*

$$\exists\, f : U \to \mathbb{D} \quad \text{such that} \quad f \text{ is bijective and holomorphic}$$

The theorem shows that if a shape is simply connected (or hole-free), every loop within it can shrink to a point without leaving the space. Such a shape can be conformally transformed into a perfect circle, while preserving local angles and shapes.

Proposition 2.2. *The set of complex numbers with positive real parts $\mathbb{C}_+$ can be mapped onto the unit disk $\mathbb{D}$ using the Möbius transformation*

$$w = \frac{z-1}{z+1}.$$

Proof. For $z = 1$, $w = 0$, mapping the real part to the center of the disk. For $z = \infty$, $w = 1$, mapping the far right to the boundary. For $z = 0$, $w = -1$, mapping the imaginary axis to the boundary. This transformation preserves angles and shapes locally, making it conformal and mapping $\mathbb{C}_+$ onto $\mathbb{D}$ as required. □

Before we define a Riemann surface, I will define some preliminary terms.

Definition 2.3 (Hausdorff Topological Space)**.** A topological space (X, τ) is called a Hausdorff space if for every pair of distinct points $x, y \in X$, there exist open neighborhoods U of x and V of y such that U and V are disjoint.

Definition 2.4 (Homeomorphism)**.** A continuous function with a continuous inverse between two topological spaces.

Definition 2.5 (Chart)**.** A homeomorphism from an open subset of a Riemann surface to an open subset of the complex plane.

Definition 2.6 (Riemann Surface)**.** A *Riemann surface* is a connected Hausdorff topological space X equipped with a collection of charts $\{(U_\alpha, \varphi_\alpha)\}$ such that:

(1) Each U_α is an open subset of X, and $\bigcup_\alpha U_\alpha = X$.
(2) Each $\varphi_\alpha : U_\alpha \to V_\alpha$ is a homeomorphism onto an open subset V_α of the complex plane $\mathbb{C}$.
(3) If $U_\alpha \cap U_\beta \neq \emptyset$, then the transition map $\varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \to \varphi_\beta(U_\alpha \cap U_\beta)$ is holomorphic.

Definition 2.7 (Meromorphic Function)**.** A function $f : U \to \mathbb{C}$ defined on an open subset U of the complex plane $\mathbb{C}$ is meromorphic on U if it is holomorphic on U except at a discrete set of isolated points $\{z_1, z_2, \dots\} \subset U$, where $f(z)$ has poles.

Definition 2.8 (Holomorphic Function). A function $f : U \to \mathbb{C}$, where U is an open subset of the complex plane $\mathbb{C}$, is called *holomorphic* at a point $z_0 \in U$ if it is differentiable at that point and in some neighborhood around it. That is, the complex derivative

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists.

A holomorphic function refers to a surface where every point has a well-defined slope and a holomorphic map can be thought of as a transformation that stretches or shrinks a space in in a way that maintains angles but not necessarily sizes.

Definition 2.9 (Complex Manifold). A complex manifold of dimension n is a topological space M such that every point $z \in M$ has a neighborhood U which is homeomorphic to an open subset of $\mathbb{C}^n$.

Definition 2.10 (Holomorphic Map). A *holomorphic map* is a function $f : X \to Y$ between complex manifolds X and Y that is holomorphic at every point of X. This means that in local charts, f can be expressed as a holomorphic function, i.e., it respects the complex structure of the manifolds by preserving complex multiplication and satisfying the Cauchy-Riemann equations.

Lemma 2.11. *The complex plane $\mathbb{C}$ itself is a Riemann surface with a single chart $(\mathbb{C}, id)$, where id is the identity map.*

Proof. $\mathbb{C}$ is a complex manifold since every point $z \in \mathbb{C}$ has a neighborhood that is homeomorphic to an open subset of $\mathbb{C}$. We also know that the identity map $\mathrm{id} : \mathbb{C} \to \mathbb{C}$ is a holomorphic function as it is differentiable with continuous partial derivatives. Since the map id acts as a chart on $\mathbb{C}$ and the transition function from id to itself is trivially the identity map (which is holomorphic), this proves that $\mathbb{C}$ with the chart $(\mathbb{C}, \mathrm{id})$ is a Riemann surface. $\square$

Lemma 2.12. *The* Riemann sphere $\mathbb{C}^* = \mathbb{C} \cup \{\infty\}$ *is a Riemann surface that can be covered by two charts:* $(\mathbb{C}, id)$ *and* $(\mathbb{C}^* \setminus \{0\}, \phi)$*, where* $\phi(z) = 1/z$ *for* $z \neq 0$*, and* $\phi(\infty) = 0$*.*

Proof. We verify that the Riemann sphere is a complex manifold covered by holomorphic charts. The chart $(\mathbb{C}, \mathrm{id})$ covers $\mathbb{C}$ except for the point at infinity, with the identity map $\mathrm{id} : \mathbb{C} \to \mathbb{C}$ being holomorphic. The second chart $(\mathbb{C}^* \setminus \{0\}, \phi)$, where $\phi(z) = 1/z$ and $\phi(\infty) = 0$, covers the point at infinity and is also holomorphic. The transition function $z \to 1/z$ between these charts is holomorphic on $\mathbb{C} \setminus \{0\}$. Therefore, the Riemann sphere is a Riemann surface covered by these holomorphic charts. $\square$

Definition 2.13. A conformal map is a function $f : U \to V$ between open subsets U and V of the complex plane $\mathbb{C}$ (or more generally, on Riemann surfaces) that preserves angles. The function f is conformal at a point $z_0 \in U$ if:

- f is holomorphic at z_0.
- The derivative $f'(z_0) \neq 0$.

A conformal map is holomorphic with a non-zero derivative, meaning it preserves angles locally but may scale figures. If it is conformal at every point in an open set, it consistently preserves angles across the whole set.

The simplest conformal map is the identity map:

$$f(z) = z$$

Another example is the exponential map:

$$f(z) = e^z$$

The exponential function e^z is holomorphic everywhere in the complex plane; the derivative is $f'(z) = e^z$, which is never zero. Thus, f is conformal everywhere.

The möbius transformation is a third example:

$$f(z) = \frac{az+b}{cz+d}$$

where $ad - bc \neq 0$.

Möbius transformations are holomorphic except at the point where the denominator is zero, $z = -\frac{d}{c}$ if $c \neq 0$. The derivative is $f'(z) = \frac{ad-bc}{(cz+d)^2}$, which is non-zero where the function is defined, making it conformal on its domain.

Definition 2.14 (Covering Map)**.** Let S and R be Riemann surfaces. A continuous surjective map $\pi : S \to R$ is a *covering map* if each point $p \in R$ has a neighborhood U such that $\pi^{-1}(U)$ is a disjoint union of open sets in S, each of which is mapped homeomorphically onto U by π.

Definition 2.15 (Covering Transformation)**.** A *covering transformation* of a covering map $\pi : S \to R$ is an automorphism $\phi : S \to S$ such that $\pi \circ \phi = \pi$. The set of all covering transformations forms a group called the *deck transformation group* of the covering.

A covering map is a function between two Riemann surfaces where surface S covers surface R. For each small patch on R, there are multiple copies on S that map perfectly onto the original on R. A covering transformation moves points on S without changing their images on R, hence describing the set of all the ways to shuffle S while preserving the covering relationship. Consider the covering map $f(z) = z^n$ from the complex plane S to R, where n is a positive integer. For example, with $n = 3$, the point $w = 1 \in R$ has three pre-images in S: $1, e^{2\pi i/3}, e^{4\pi i/3}$. A covering transformation cyclically permutes these pre-images without changing their image in R. Thus, $f(z)$ covers R by mapping multiple points in S to a single point in R while respecting the covering relationship.

Karl Weierstrass, in introducing the uniformization theorem, shows that every Riemann surface can be represented by covering maps of simpler surfaces, like the complex plane or unit disk.

Theorem 2.16 (Uniformization Theorem)**.** *Every simply connected Riemann surface is conformally equivalent to one of the following: the Riemann sphere* $\mathbb{P}^1(\mathbb{C})$*, the complex plane* $\mathbb{C}$*, or the open unit disk* $\mathbb{D}$*.*

An example of this theorem is the open unit disk $\mathbb{D}$, which is a simply connected Riemann surface. The disk $\mathbb{D}$ is conformally equivalent to itself, but it can also be seen as conformally equivalent to a different simply connected surface, such as the upper half-plane $\mathbb{H}$, through a Möbius transformation.

Definition 2.17. For a Riemann surface, the curvature refers to Gaussian curvature, which is a measure of how the surface bends in space at a particular point. Formally, Gaussian curvature K at a point p on a surface is the product of the principal curvatures k_1 and k_2 at that point:

$$K(p) = k_1 \cdot k_2$$

The principal curvatures k_1 and k_2 describe the maximum and minimum rates of curvature along orthogonal directions passing through the point.

Positive curvature ($K > 0$) occurs when both principal curvatures bend in the same direction (ex: surface of a sphere). The surface curves outward in all directions. Zero curvature ($K = 0$) happens on a flat surface, like a plane, where there is no bending. Negative curvature ($K < 0$) is when the surface curves in opposite directions (ex: saddle shape).

Lemma 2.18. *Any Riemann surface of genus greater than one admits a unique conformal metric of constant negative curvature* -1*, making it a hyperbolic surface.*

Proof. This proof uses the Uniformization Theorem. For a Riemann surface S of genus $g > 1$, which differs topologically from spheres and tori as it has more than one hole, the theorem shows that such surfaces are covered by the unit disk $\mathbb{D}$ with a hyperbolic metric $ds^2 = \frac{4\,|dz|^2}{(1-|z|^2)^2}$ and constant curvature -1. S is represented as $\mathbb{D}/\Gamma$, where Γ is a Fuchsian group acting by isometries. This action gives S a unique hyperbolic metric, as any two conformal metrics of constant negative curvature are isometric. Thus, every Riemann surface of genus $g > 1$ admits a unique conformal metric with curvature -1, classifying it as a hyperbolic surface. □

Definition 2.19 (Biholomorphic Map)**.** A bijective and holomorphic map with a holomorphic inverse.

Definition 2.20 (Teichmüller Space)**.** Teichmüller space, $\mathcal{T}(S)$, is the space of equivalence classes of marked Riemann surfaces of a given topological type. A point in $\mathcal{T}(S)$ represents a pair (X, f), where X is a Riemann surface and $f : S \to X$ is a homeomorphism. Two pairs (X_1, f_1) and (X_2, f_2) are equivalent if there is a biholomorphic map $g : X_1 \to X_2$ such that $g \circ f_1$ is isotopic to f_2.

A Teichmüller space is the collection of all possible shapes of a surface that can be transformed into each other through smooth transformations.

Definition 2.21 (Teichmüller Distance)**.** The Teichmüller distance between two points (X_1, f_1) and (X_2, f_2) in $\mathcal{T}(S)$ is

$$d_{\mathcal{T}}((X_1, f_1), (X_2, f_2)) = \frac{1}{2} \inf \log K(f),$$

where the infimum is over all quasiconformal maps $f : X_1 \to X_2$ homotopic to $f_2 \circ f_1^{-1}$, and $K(f)$ is the maximal dilatation of f.

Teichmüller distance measures the minimal distortion required to transform one shape into another.

Definition 2.22. A homeomorphism $f : \mathbb{C} \to \mathbb{C}$ is called quasiconformal if it satisfies the following condition: there exists a constant $K \geq 1$ such that for every point $z \in \mathbb{C}$, the following inequality holds

$$\frac{|f_z| + |f_{\overline{z}}|}{|f_z| - |f_{\overline{z}}|} \leq K,$$

where f_z and $f_{\overline{z}}$ are the partial derivatives of f with respect to z and $\overline{z}$, respectively.

2.2. **Algebraic Structures and Topological Properties.**

Definition 2.23. The ramification index e_P of a holomorphic map $f : X \to Y$ at a point $P \in X$ is the smallest positive integer e such that

$$(f(z) - f(P)) = c(z - P)^e + \text{higher order terms},$$

where z is a local coordinate around P and c is a nonzero constant. If $e_P > 1$, the point P is called a ramification point.

Theorem 2.24 (Hurwitz's Theorem)**.** *If $f : X \to Y$ is a non-constant holomorphic map between compact Riemann surfaces of genera g_X and g_Y, respectively, then:*

$$2g_X - 2 = \deg(f)(2g_Y - 2) + \sum_{p \in X} (e_p - 1)$$

where e_p is the ramification index at p.

Hurwitz's Theorem explains how the genus of a simpler surface changes when mapped to a more complex one via a holomorphic map. For instance, when mapping a torus to a sphere, the theorem calculates how the holes and twists of the torus relate to the simpler shape of the sphere.

Proposition 2.25. *Consider a map $f : X \to Y$ from a torus X (genus $g_X = 1$) to a sphere Y (genus $g_Y = 0$). Applying Hurwitz's Theorem, we have*

$$\begin{aligned} 2g_X - 2 &= \deg(f)(2g_Y - 2) + \sum_{p \in X} (e_p - 1) \\ 2(1) - 2 &= \deg(f)(2(0) - 2) + \sum_{p \in X} (e_p - 1) \\ 0 &= -2\deg(f) + \sum_{p \in X} (e_p - 1) \\ 2\deg(f) &= \sum_{p \in X} (e_p - 1) \end{aligned}$$

This shows that the total contribution from the ramification points on the torus must equal $2\deg(f)$.

Oscar Zariski and André Weil connected algebraic geometry with the theory of Riemann surfaces by demonstrating that every Riemann surface corresponds to an algebraic curve.

Theorem 2.26 (Riemann-Roch Theorem)**.** *For a compact Riemann surface X of genus g and a divisor D on X, the dimension of the space of meromorphic functions is given by:*

$$\ell(D) - \ell(K - D) = \deg(D) - g + 1$$

where $\ell(D)$ is the dimension of the space of meromorphic functions with poles no worse than D, and K is the canonical divisor.

Consider a compact Riemann surface where poles and zeros can be found. The Riemann-Roch Theorem tells us how many unique functions we can create to hit these points in specific ways.

Adolf Hurwitz then developed a formula relating the genera of two Riemann surfaces connected by a holomorphic map.

Proposition 2.27. *For a compact Riemann surface X of genus g, and a divisor D on X, the dimension of the space of meromorphic functions $\ell(D)$ minus the dimension of the space of meromorphic functions $\ell(K-D)$ is equal to $\deg(D)-g+1$, where K is the canonical divisor.*

Proof. We want to show

$$\ell(D) - \ell(K - D) = \deg(D) - g + 1.$$

Consider the definition of the space of meromorphic functions associated with a divisor. We know that, for a divisor D on Riemann surface X, the space $\mathcal{L}(D)$ consists of meromorphic functions f such that the divisor of f plus D is effective s.t. $(f) + D \geq 0$. The dimension of this space is denoted by $\ell(D)$. The canonical divisor K on X corresponds to the divisor of a non-zero meromorphic 1-form on X. The space $\mathcal{L}(K)$ is the space of holomorphic 1-forms on X, and its dimension is equal to the genus g of the surface.

Riemann-Roch theorem provides the relationship

$$\ell(D) - \ell(K - D) = \deg(D) + 1 - g.$$

Construct a meromorphic 1-form ω s.t. its divisor is K. Using Serre duality, which relates the space $\mathcal{L}(K - D)$ to the space of 1-forms with poles restricted by D, consider the space of meromorphic forms $\mathcal{L}(K - D)$, consisting of meromorphic 1-forms ω for which the divisor $(\omega) + (D) \geq 0$.

Take the Euler characteristic of $\mathcal{O}(D)$, defined as

$$\chi(D) = h^0(D) - h^1(D),$$

where $h^0(D) = \ell(D)$ and $h^1(D) = \ell(K - D)$. According to Riemann-Roch, the Euler characteristic can be expressed as

$$\chi(D) = \deg(D) - g + 1.$$

Substituting into the Euler characteristic formula, we then have

$$\ell(D) - \ell(K - D) = \chi(D) = \deg(D) - g + 1.$$

□

Definition 2.28 (Algebraic Morphism)**.** Let X and Y be algebraic varieties over an algebraically closed field $\mathbb{K}$. A map $\phi : X \to Y$ is called an *algebraic morphism* (or simply a *morphism*) if for every open subset $U \subseteq Y$ and every regular function $f \in \mathcal{O}_Y(U)$, the composition $f \circ \phi$ is a regular function on $\phi^{-1}(U)$, that is, $f \circ \phi \in \mathcal{O}_X(\phi^{-1}(U))$. In affine varieties, a map $\phi : \mathbb{A}^n \to \mathbb{A}^m$ given by $\phi(x_1, x_2, \dots, x_n) = (f_1(x_1, x_2, \dots, x_n), \dots, f_m(x_1, x_2, \dots, x_n))$ is a morphism if each f_i is a polynomial in the coordinate functions $x_1, x_2, \dots, x_n$.

An algebraic morphism is a function between two algebraic varieties X and Y, which are sets of solutions to polynomial equations defined over an algebraically closed field $\mathbb{K}$. In the context of affine varieties, which are varieties that can be embedded in an affine space, a map $\phi : \mathbb{A}^n \to \mathbb{A}^m$ given by

$$\phi(x_1, x_2, \ldots, x_n) = (f_1(x_1, x_2, \ldots, x_n), \ldots, f_m(x_1, x_2, \ldots, x_n))$$

is a morphism if each function f_i is a polynomial in the coordinate functions $x_1, x_2, \ldots, x_n$.

In simpler terms, an algebraic morphism ensures that the transformation between X and Y keeps the polynomial nature intact. When we are dealing with affine varieties, which can be thought of as varieties that can be embedded in some higher-dimensional space, a map between them is a morphism if it is given by polynomials in terms of the coordinates of the space.

Proposition 2.29. *There is no biholomorphic (holomorphic and bijective with a holomorphic inverse) map between two compact Riemann surfaces of different genus.*

Proof. Suppose, for contradiction, that there exists a biholomorphic map $f : X \to Y$ between compact Riemann surfaces X and Y with different genera g_X and g_Y. Since f is a homeomorphism, it induces an isomorphism between the fundamental groups $\pi_1(X)$ and $\pi_1(Y)$. The fundamental group of a compact Riemann surface of genus g is

$$\pi_1(\Sigma_g) = \langle a_1, b_1, \ldots, a_g, b_g \mid \prod_{i=1}^{g} [a_i, b_i] = 1 \rangle,$$

with rank $2g$. Therefore, $2g_X = 2g_Y$, implying $g_X = g_Y$, a contradiction. Biholomorphic maps preserve the space of holomorphic 1-forms, which has dimension g by the Riemann-Roch theorem. Since the dimension of these spaces, $h^0(K_X)$ and $h^0(K_Y)$, must be equal, $g_X = g_Y$. Thus, no biholomorphic map can exist between surfaces with different genera, confirming that genus is an invariant of biholomorphic equivalence for compact Riemann surfaces. $\square$

For the remainder of the paper, I will now shift our focus to examining how Riemann surfaces intersect with various fields of machine learning.

3. Jensen's Inequality

Jensen's Inequality is fundamental in machine learning, particularly in the analysis of algorithms and probabilistic models.

Lemma 3.1 (Jensen's Inequality). *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function and let X be a random variable. Then, the following inequality holds*

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

Proof. Since f is convex, for any x_0 in the domain of f, there exists a subgradient $g \in \partial f(x_0)$ such that for all x,

$$f(x) \geq f(x_0) + g(x - x_0).$$

Taking expectation with respect to X on both sides, we have

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(x_0) + g(X - x_0)] = f(x_0) + g(\mathbb{E}[X - x_0]) = f(x_0) + g(\mathbb{E}[X] - x_0).$$

Setting $x_0 = \mathbb{E}[X]$ gives us

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

□

Consider a probabilistic model where we need to approximate an intractable posterior distribution $p(z \mid x)$. In variational inference, we introduce a simpler distribution $q(z)$ to approximate the true posterior, minimizing the Kullback-Leibler (KL) divergence

$$\mathrm{KL}(q(z) \parallel p(z \mid x)) = \mathbb{E}_{q(z)}\left[\log q(z) - \log p(z \mid x)\right].$$

Instead of minimizing the KL divergence directly, we derive a tractable lower bound using Jensen's Inequality, leading to the Evidence Lower Bound (ELBO) for the log likelihood

$$\log p(x) \geq \mathbb{E}_{q(z)}\left[\log p(x, z) - \log q(z)\right].$$

This inequality stems from applying Jensen's Inequality to the expectation under $q(z)$

$$\log \mathbb{E}_{q(z)}\left[\frac{p(x,z)}{q(z)}\right] \geq \mathbb{E}_{q(z)}\left[\log \frac{p(x,z)}{q(z)}\right].$$

Now, consider $p(z \mid x)$ and $q(z)$ on a Riemann surface S. The posterior distribution may have branch points or multi-valued functions, resolved by the complex structure of the surface. By lifting $q(z)$ onto the covering space of S, we handle these singularities. Applying Jensen's Inequality on S maintains the ELBO approximation:

$$\log \mathbb{E}_{q(z)}\left[\frac{p(x,z)}{q(z)}\right] \geq \mathbb{E}_{q(z)}\left[\log \frac{p(x,z)}{q(z)}\right],$$

where z lies on the Riemann surface. The geometry of S preserves the accuracy of the variational inference, even in complex geometries.

Jensen's Inequality has profound implications in various real-world contexts. For instance, in finance, it is often applied to the valuation of risky assets.

Example 3.2. The Black-Scholes model is a standard method for pricing European options. It assumes the underlying asset price follows a geometric Brownian motion with constant volatility and no jumps

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t,$$

where S_t is the asset price at time t, μ is the drift (expected return), σ is the volatility, and W_t is a Wiener process (Brownian motion). When options approach expiration, complexities such as discontinuities and rapid volatility challenge the Black-Scholes assumptions, which can lead to mispricing. By interpreting the asset price as moving across different multi-sheeted Riemann surfaces connected by algebraic morphisms, we capture the complex behavior of the asset in scenarios where the standard assumptions break down. Applying Jensen's Inequality to the Riemann surface provides a more accurate ELBO for the option's price, therefore ensuring $q(z)$ accounts for dynamics in such scenarios.

3.1. **Case Study: Probabilistic Models.** The above approach mirrors techniques used in advanced probabilistic models, where similar mathematical tools are employed to handle data on complex geometries. For example, in "Variational Inference on Manifolds" (2018), Hauberg extended variational inference to non-Euclidean spaces using Jensen's Inequality and Riemann surface theory. This adaptation allowed for modeling curved data while preserving geometric relationships, addressing the challenge of maintaining the posterior distribution's structure on non-Euclidean surfaces by incorporating Jensen's Inequality to derive a tractable ELBO, improving inference for data with inherent curvature, like human body shapes.

Mathematical Overview The posterior $p(z \mid x)$ is defined on a Riemannian manifold $\mathcal{M}$, with a prior respecting the manifold's geometry:

$$p(z) \propto e^{-\frac{1}{2}\|z\|_g^2},$$

where $\|z\|_g^2$ is the Riemannian norm. Applying Jensen's Inequality, the ELBO is derived on the Riemannian manifold

$$\log p(x) \geq \mathbb{E}_{q(z)}\left[\log p(x \mid z) + \log p(z) - \log q(z)\right] - \frac{1}{2}\log\det g(z),$$

where $g(z)$ represents the manifold's metric tensor. The KL divergence is minimized using variational distributions adapted to the manifold

$$\mathrm{KL}(q(z) \parallel p(z \mid x)) = \int_{\mathcal{M}} q(z)\log\frac{q(z)}{p(z \mid x)} d\mathrm{Vol}(z).$$

Hence, as shown, by mapping the data onto Riemann surfaces into the variational inference process, the authors preserved the geometric integrity of the data during inference, allowing for a more precise machine learning model.

4. Hoeffding's Inequality

Hoeffding's Inequality provides a bound on the probability that the sum of bounded random variables deviates from its expected value. This is crucial as it helps ensure that the performance of a model on a training set will generalize well to new data by quantifying the risk of large deviations.

Lemma 4.1 (Hoeffding's Inequality)**.** *Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $a_i \leq X_i \leq b_i$ almost surely. Define $S_n = \sum_{i=1}^n X_i$. Then, for any $t > 0$,*

$$\mathbb{P}\left(S_n - \mathbb{E}[S_n] \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Proof. Consider the moment generating function of the centered random variable $Y_i = X_i - \mathbb{E}[X_i]$, which satisfies $a_i - \mathbb{E}[X_i] \leq Y_i \leq b_i - \mathbb{E}[X_i]$. The cumulant-generating function of Y_i is bounded by

$$\mathbb{E}[e^{\lambda Y_i}] \leq \exp\left(\frac{\lambda^2(b_i - a_i)^2}{8}\right)$$

for $\lambda \in \mathbb{R}$. Using the Chernoff bound and the independence of X_i, we have

$$\mathbb{P}\left(S_n - \mathbb{E}[S_n] \geq t\right) \leq \inf_{\lambda > 0} e^{-\lambda t}\prod_{i=1}^n \mathbb{E}[e^{\lambda Y_i}] \leq \inf_{\lambda > 0} e^{-\lambda t}\exp\left(\sum_{i=1}^n \frac{\lambda^2(b_i - a_i)^2}{8}\right).$$

Optimizing over λ gives the result.

Consider random variables $X_1, X_2, \ldots, X_n$ as measurements on a Riemann surface S, with distances calculated using the hyperbolic metric $ds^2 = \frac{4\,|dz|^2}{(1-|z|^2)^2}$, where $z \in \mathbb{D}$. The bounds a_i and b_i depend on the surface's geometry. For hyperbolic surfaces, the curvature $K = -1$ affects these bounds, refining Hoeffding's Inequality to

$$\mathbb{P}\left(S_n - \mathbb{E}[S_n] \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2_{\text{curved}}}\right),$$

where $(b_i - a_i)_{\text{curved}}$ reflects the influence of the hyperbolic geometry. □

Example 4.2. In wireless communication networks, Hoeffding's Inequality is important as it provides bounds on the probability of significant deviations in signal strength. However, in environments with obstacles or varying distances, signal propagation does not follow a simple Euclidean path. By mapping the signal propagation process onto a Riemann surface, the holomorphic maps and conformal transformations involved enable us to model the signal propagation more accurately. Applying Hoeffding's Inequality on this Riemann surface allows us to derive more precise bounds on the signal strength and interference probabilities, considering the nature of real-world environments.

4.1. **Case Study: Stochastic Processes.** The above adaptations are directly relevant to the discussion in "Concentration Inequalities on Manifolds" by M. Ledoux (2001), where classical inequalities are extended to Riemannian manifolds. The goal of this paper was to adapt probabilistic bounds to curved spaces, ensuring they hold under non-Euclidean geometries.

Mathematical Overview The inequality was extended to Riemannian manifolds $\mathcal{M}$, adjusting the variance based on the manifold's geometry

$$\mathbb{P}\left(S_n - \mathbb{E}[S_n] \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2_{\text{manifold}}}\right).$$

The KL divergence was computed by integrating over the manifold's volume form, ensuring correct probability calculations on $\mathcal{M}$

$$\mathrm{KL}(q(z) \parallel p(z \mid x)) = \int_{\mathcal{M}} q(z) \log \frac{q(z)}{p(z \mid x)} d\mathrm{Vol}(z).$$

Hence, by adapting Hoeffding's Inequality to Riemannian manifolds, the authors improved generalization in machine learning models dealing with non-Euclidean data as the refined concentration bounds accounted for the manifold's curvature, leading to better performance in applications involving curved data.

5. Itô Calculus

Itô calculus provides a framework for dealing with stochastic processes, particularly those involving Brownian motion, when applied with Riemann surfaces, it can help describe the evolution of stochastic processes that are affected by complex geometries.

Lemma 5.1. *Let $f : S \to \mathbb{C}$ be a smooth function on a Riemann surface S and let X_t be a stochastic process on S, which evolves according to the stochastic differential equation*

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dB_t,$$

where B_t is a standard Brownian motion. Then, Itô's Lemma states that the differential of $f(X_t)$ is given by

$$df(X_t) = \left(\mu(X_t)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2(X_t)\frac{\partial^2 f}{\partial x^2}\right)dt + \sigma(X_t)\frac{\partial f}{\partial x}dB_t.$$

Proof. Consider the smooth function f on the Riemann surface S. Using Itô's Lemma for a function of a stochastic process, we expand $df(X_t)$ as a Taylor series in the infinitesimal increments dX_t. Applying the chain rule for stochastic calculus, we obtain

$$df(X_t) = \frac{\partial f}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dX_t)^2.$$

Substituting $dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$ and recalling that $(dB_t)^2 = dt$, we get

$$df(X_t) = \frac{\partial f}{\partial x}(\mu(X_t)dt + \sigma(X_t)dB_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\sigma^2(X_t)dt$$

$$= \left(\mu(X_t)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2(X_t)\frac{\partial^2 f}{\partial x^2}\right)dt + \sigma(X_t)\frac{\partial f}{\partial x}dB_t$$

Now, consider the case where the stochastic process X_t evolves on a Riemann surface S. In this scenario, the function f respects the complex structure of the surface, and the partial derivatives are computed with respect to local coordinates on S. The geometry of the Riemann surface influences the drift μ and the volatility σ, which may depend on the curvature and other geometric properties of S. Consequently, Itô's Lemma is adapted to account for the underlying geometry of the surface when describing the evolution of stochastic processes on it. □

Example 5.2. Say one wants to model the evolution of asset prices using Ito's calculus and runs into the issue that such prices are influenced by stochastic processes and complex geometries. Under standard models, the complexity of this data may not be properly understood, hence leading to inaccuracies. Mapping the data onto Riemann surfaces can incorporate both the random fluctuations and the complex geometric structure of the market, resulting in more accurate pricing and risk assessments in an environment with such complex underlying dynamics.

5.1. **Case Study: Itô's Lemma.** The paper ”Stochastic Processes on Riemann Surfaces: Applications in Machine Learning” by Y. Zhang and L. Hauberg (2019) demostrates the above idea as the authors extended classical stochastic differential equations (SDEs) to Riemann surfaces using Itô calculus with the goal of modelling complex data that lies on non-Euclidean surfaces while accounting for the stochastic nature of the data.

Mathematical Overview The authors applied Itô's Lemma to stochastic processes on Riemann surfaces, adjusting the drift and diffusion terms according to the surface's curvature:

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dB_t,$$

where μ and σ are curvature-dependent. They adapted the stochastic integral to respect the surface's geometry and solved the drift-diffusion equation numerically using a discretization scheme tailored for Riemann surfaces.

Extending Itô calculus to Riemann surfaces improves the ability of machine learning models to manage stochastic data while preserving geometric structures.

6. Representer Theorem

The Representer Theorem is critical in kernel methods, showing that the solution to many regularized learning problems can be expressed as a linear combination of training samples.

Lemma 6.1 (Representer Theorem)**.** *Consider a regularized empirical risk minimization problem of the form*

$$\min_{f\in\mathcal{H}} \sum_{i=1}^{n} L(y_i, f(x_i)) + \Omega(\|f\|_{\mathcal{H}}),$$

where $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS) with kernel k, L is a loss function, and Ω is a strictly monotonically increasing function. Then, the solution $f^* \in \mathcal{H}$ admits a representation of the form

$$f^*(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i),$$

for some coefficients $\alpha_i \in \mathbb{R}$.

Proof. Consider any function $f \in \mathcal{H}$. By the reproducing property, f can be decomposed as $f = f_0 + f_\perp$, where $f_0(x) = \sum_{i=1}^{n} \beta_i k(x, x_i)$ and $f_\perp$ is orthogonal to the span of $\{k(\cdot, x_i)\}_{i=1}^{n}$ in $\mathcal{H}$. Then, we can write

$$f(x) = \sum_{i=1}^{n} \beta_i k(x, x_i) + f_\perp(x).$$

Substituting this into the objective function, the loss term depends only on β_i, while the regularization term is given by $\|f\|_{\mathcal{H}}^2 = \|f_0\|_{\mathcal{H}}^2 + \|f_\perp\|_{\mathcal{H}}^2$. Since Ω is strictly increasing, minimizing the regularization term requires $f_\perp = 0$, leading to $f^*(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i)$.

Suppose the inputs x_i correspond to points on a Riemann surface S. The kernel function $k(x, x_i)$ is adapted to the geometry of S, incorporating curvature or complex structure. For example, in the hyperbolic plane, the kernel depends on hyperbolic distances, making the learning process sensitive to the surface's geometry. By minimizing the regularization term, we find that the optimal function $f^*(x)$ is a linear combination of geometry-aware kernel evaluations:

$$f^*(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i).$$

Hence, machine learning models can adapt to non-Euclidean geometries, improving performance on tasks involving curved surfaces. The Representer Theorem reduces the search space, while a geometry-sensitive kernel boosts accuracy in computer vision, NLP, and manifold learning tasks.

6.1. **Case Study: Kernel-Based Learning.** In "Kernel Methods on Riemannian Manifolds" by X. Pennec et al. (2006), the authors extended kernel-based machine learning methods, grounded in the Representer Theorem, to data distributed on Riemannian manifolds. Their goal was to adapt kernel functions to respect the geometry of curved spaces, improving the accuracy and generalization of models on non-Euclidean data, such as in computer vision and medical imaging.

Mathematical Overview The Representer Theorem was applied to regularized empirical risk minimization on Riemannian manifolds $\mathcal{M}$, where the solution $f^*(x)$ could be expressed as:

$$f^*(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i),$$

with kernel functions $k(x, x_i)$ adapted to the manifold's geometry. For example, in hyperbolic space, the kernel would depend on the hyperbolic distance between points:

$$k(x, x_i) = \exp\left(-d^2_{\mathcal{M}}(x, x_i)\right).$$

A short example of this theory is below.

Given data points x and x_i in hyperbolic space $\mathbb{H}^2$ with curvature $K = -1$, the geodesic distance $d_M(x, x_i)$ is

$$\cosh(d_M(x, x_i)) \approx \cosh(1)\cosh(2) - \sinh(1)\sinh(2)\cos(60^\circ) \approx 3.671.$$

So, we have

$$d_M(x, x_i) = \cosh^{-1}(3.671) \approx 1.686.$$

Hence, the corresponding kernel function is

$$k(x, x_i) = \exp(-d^2_M(x, x_i)) \approx \exp(-2.844) \approx 0.058.$$

This function is sensitive to the curvature of hyperbolic space, hence improving the accuracy of the model by capturing the non-Euclidean geometry of the data.

7. Complex-Valued Data Analysis

Riemann surfaces provide a framework for understanding and visualizing multi-valued complex functions as they allow us to convert these functions into single-valued ones by organizing their branches in a coherent and structured manner.

To analyze the multi-valued function $f(z) = \sqrt{z}$, which has a branch point at $z = 0$, we construct its Riemann surface using two sheets of the complex plane joined along a branch cut from $z = 0$ to $z = \infty$ on the negative real axis. One sheet defines the square root with positive imaginary parts, and the other with negative imaginary parts.

Lemma 7.1. *Show that the function $f(z) = \sqrt{z}$ can be made single-valued on a Riemann surface constructed with two sheets, and determine its behavior on this surface.*

Proof. To construct a continuous, single-valued Riemann surface for $\sqrt{z}$, we define two sheets in the complex plane. The principal branch is defined for $z = re^{i\theta}$ with $-\pi < \theta \leq \pi$, where $f(z) = \sqrt{r}e^{i\theta/2}$. A second sheet is introduced for $\pi < \theta \leq 3\pi$, where $f(z) = -\sqrt{r}e^{i\theta/2}$. The branch cut along the negative real axis connects points at $\theta = \pm\pi$, transitioning between the two sheets as the function crosses the cut. This construction creates a continuous path around the branch point at $z = 0$, allowing $\sqrt{z}$ to be treated as single-valued and continuous on the Riemann surface. □

Each complex number has two square roots, leading to discontinuities in the square root function. A two-sheeted Riemann surface resolves this by assigning each value to a different sheet, ensuring continuity. In machine learning, Riemann surfaces handle complex-valued data, ensuring smooth transformations in tasks like image processing and improving performance on complex datasets.

Proposition 7.2. *Let S be a compact Riemann surface, and let $\{f_1, f_2, \ldots, f_n\}$ be a set of complex-valued functions defined on S. Our goal is to classify each function as either meromorphic (Class 1) or holomorphic (Class 2).*

Proof. We classify a function f as meromorphic if it has isolated poles on S, and as holomorphic if it is analytic everywhere on S. The Riemann-Roch theorem states that for a divisor D on a compact Riemann surface S of genus g, the dimension $l(D)$ of the space of meromorphic functions with poles bounded by D satisfies:

$$l(D) - l(K - D) = \deg(D) + 1 - g,$$

where K is the canonical divisor and $\deg(D)$ is the degree of D. To classify a function f, we construct its divisor

$$D(f) = \sum n_i p_i - \sum m_j q_j,$$

where p_i are poles with multiplicities n_i and q_j are zeros with multiplicities m_j. Using Riemann-Roch: (1) if $\deg(D(f)) > 0$, indicating the presence of poles, f is meromorphic, (2) if $\deg(D(f)) \leq 0$, implying no poles, f is holomorphic. For example, $f_1(z) = \frac{1}{z^2+1}$ has divisor $D(f_1) = (i) + (-i) - 0$, so f_1 is meromorphic. For $f_2(z) = e^z$, the divisor is $D(f_2) = 0$, so f_2 is holomorphic. □

7.1. **Case Study: Manifold Learning.** In "Deep Learning-Based Phase Unwrapping for Quantitative Phase Imaging" (Zhang et al., 2021), the authors used *Riemann surfaces* to address phase unwrapping challenges in QPI, managing multi-valued complex functions. By mapping distinct branches of the phase function onto different sheets, they ensured continuity across these branches, similar to handling multi-valued functions like $f(z) = \sqrt{z}$.

Mathematical Overview In interferometry, the measured phase ϕ is wrapped within $[-\pi, \pi]$, causing ambiguities in reconstructing the unwrapped phase Φ. The goal is to find $\Phi(x) = \phi(x) + 2k\pi$, where k resolves the discontinuity. Riemann surfaces are used to handle the multi-valued nature of phase data, with each branch corresponding to a different sheet. The Riemann surface ensures smooth transitions between sheets. A deep learning model was trained to identify the correct sheet for each point x and predict the unwrapped phase by minimizing:

$$\mathcal{L}(\Phi) = \sum_i \|\Phi(x_i) - \phi(x_i) - 2k_i\pi\|^2,$$

where k_i adjusts the sheets as needed. The phase unwrapping algorithm is discretized over a pixel grid, with the Riemann surface structure guiding the correct choice of phase sheet.

This study demonstrates the practical application of Riemann surfaces in machine learning to handle complex-valued data, enhancing algorithm performance in quantitative imaging.

8. Computer Vision

In computer vision, preserving local features like edges and corners is crucial, especially in tasks like facial recognition where relative angles between features matter. Conformal mappings can be used as a preprocessing step to transform images while preserving angles and local geometry, ensuring consistency of key features for algorithms.

Example 8.1. Consider the problem of stitching images captured with a wide-angle or fisheye lens to create a seamless 360-degree panoramic image, a task which requires mapping flat images onto a spherical surface while preserving local angle measurements.

Let S be a Riemann surface that represents the spherical geometry of the panoramic image and define $f : \mathbb{C} \to S$ as a holomorphic function, where $\mathbb{C}$ is the complex plane and S is the Riemann sphere. The conformal mapping we use in this case is a stereographic projection, which projects points from the complex plane to a Riemann sphere and is defined

$$f(z) = \frac{2z}{1+|z|^2} + i\frac{1-|z|^2}{1+|z|^2}.$$

For a holomorphic function f, if $f'(z_0) \neq 0$ at a point z_0, then f is conformal at z_0. First, we find the derivative of f

$$f'(z) = \frac{2(1-|z|^2) - 2z \cdot \overline{z} \cdot (2z)}{(1+|z|^2)^2} + i\frac{2z \cdot (1+|z|^2) - 2z \cdot (1-|z|^2)}{(1+|z|^2)^2}.$$

$$f'(z) = \frac{2}{(1+|z|^2)^2}.$$

Since $f'(z) \neq 0$, f is conformal. Therefore, f preserves angles between curves in the mapping process.

In computer vision, conformal mappings on Riemann surfaces preserve angles, crucial for tasks like panoramic stitching and self-driving cars. They ensure seamless image alignment, preventing distortions in features such as road edges, and improve accuracy in perception, recognition, and navigation.

8.1. Case Study: Geometric Transformations.

In "Optimal Mass Transport for Shape Matching and Comparison" (IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015), Z. Su and Y. Wang used conformal mappings based on Riemann surface theory for surface registration, ensuring angle preservation during transformations. They applied optimal mass transport to match and compare 3D shapes while preserving geometric correspondence.

Mathematical Overview Given two shapes, represented by distributions of mass $\mu(x)$ and $\nu(y)$, their goal was to find a transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ that maps μ to ν while minimizing a cost function, typically defined as the total transportation cost:

$$\min_T \int_{\mathbb{R}^2} c(x, T(x))\, d\mu(x),$$

where $c(x, T(x))$ is the cost to move mass from x to $T(x)$.

By combining conformal maps, Riemann surface theory, and optimal mass transport, the study enhanced shape registration accuracy and offered a robust framework for complex transformations in 3D reconstruction and panoramic image stitching.

9. Conclusion

Hence, by integrating such probability theory with the geometric framework of Riemann surfaces, we can create powerful models that can accurately predict and manage uncertainties in complex, real-world scenarios, from the financial markets to self-driving cars.

10. Acknowledgments

I would like to thank my mentor, Aadrita Laha, for her invaluable knowledge, guidance, and support throughout this research project.

References

[1] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
[2] F. Black and M. Scholes, *The Pricing of Options and Corporate Liabilities*, Journal of Political Economy, vol. 81, no. 3, pp. 637-654, 1973.
[3] G. Cristóbal, P. Schelkens, and H. Thienpont, *Optical and Digital Image Processing: Fundamentals and Applications*, Wiley-VCH, 2011.
[4] B. Farb and D. Margalit, *A Primer on Mapping Class Groups*, Princeton University Press, 2011.
[5] O. Forster, *Lectures on Riemann Surfaces*, Springer-Verlag, 1981.
[6] P. Griffiths and J. Harris, *Principles of Algebraic Geometry*, John Wiley Sons, 1978.
[7] L. Hauberg, *Variational Inference on Manifolds*, Journal of Machine Learning Research, 19(42), pp. 1-41, 2018.
[8] W. Hoeffding, *Probability Inequalities for Sums of Bounded Random Variables*, Journal of the American Statistical Association, 58(301), pp. 13-30, 1963.
[9] A. Hurwitz, *Über algebraische Gebilde mit eindeutigen Transformationen in sich*, Mathematische Annalen, vol. 41, pp. 403-442, 1893.
[10] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, *Introduction to Variational Methods for Graphical Models*, Machine Learning, 1999.
[11] M. Ledoux, *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs, Volume 89, American Mathematical Society, 2001.
[12] J. S. Milne, *Modular Functions and Modular Forms (Elliptic Modular Curves)*, Version 1.31, March 22, 2017.
[13] D. Mumford, *Lectures on Curves on an Algebraic Surface*, Princeton University Press, 1966.
[14] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.
[15] Y. Zhang and L. Hauberg, *Stochastic Processes on Riemann Surfaces: Applications in Machine Learning*, Journal of Machine Learning Research, 21(78), pp. 1-34, 2019.
[16] X. Pennec, P. Fillard, and N. Ayache, *A Riemannian Framework for Tensor Computing*, International Journal of Computer Vision, 66(1), pp. 41-66, 2006.
[17] X. Zhang, Y. Luo, and M. Li, *Deep Learning-Based Phase Unwrapping for Quantitative Phase Imaging*, IEEE Transactions on Medical Imaging, 2021.
[18] B. Riemann, *Grundlagen für eine allgemeine Theorie der Funktionen einer veränderlichen complexen Größe*, 1851.
[19] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, 2014.
[20] Z. Su, Y. Wang, et al., *Optimal Mass Transport for Shape Matching and Comparison*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.
[21] O. Teichmüller, *Extremale quasikonforme Abbildungen und quadratische Differentiale*, Abh. Preuss. Akad. Wiss. Math.-Nat. Kl., 1944.

[22] V. Vapnik, *Statistical Learning Theory*, Wiley, 1998.
[23] A. Weil, *Foundations of Algebraic Geometry*, American Mathematical Society, 1946.
[24] O. Zariski, *Algebraic Surfaces*, Springer, 1935.