Journal of Machine Learning Research 24 (2023) 1-34 Submitted 11/21; Revised 8/23; Published 12/23

Reproducing Kernels and New Approaches in Compositional Data Analysis

Binglin Li BLI2@NCAT.EDU
Department of Mathematics and Statistics
North Carolina A&T State University
Greensboro, NC 27411, USA

Changwon Yoon CWYOON@KAIST.AC.KR
Jeongyoun Ahn* JYAHN@KAIST.AC.KR
Department of Industrial and Systems Engineering
Korea Advanced Institute of Science and Technology
Daejeon, 34141, Korea

Editor: Risi Kondor

Abstract

Compositional data, such as human gut microbiomes, consist of non-negative variables where only the relative values of these variables are available. Analyzing compositional data requires careful treatment of the geometry of the data. A common geometrical approach to understanding such data is through a regular simplex. The majority of existing approaches rely on log-ratio or power transformations to address the inherent simplicial geometry. In this work, based on the key observation that compositional data are *projective*, we reinterpret the compositional domain as a group quotient of a sphere, leveraging the intrinsic connection between projective and spherical geometry. This interpretation enables us to understand the function spaces on the compositional domain in terms of those on a sphere, and furthermore, to utilize spherical harmonics theory for constructing a *compositional Reproducing Kernel Hilbert Space (RKHS)*. The construction of RKHS for compositional data opens up new research avenues for future methodology developments, particularly introducing well-developed kernel methods to compositional data analysis. We demonstrate the wide applicability of the proposed theoretical framework with examples of nonparametric density estimation, kernel exponential family, and support vector machine for compositional data.

Keywords: Directional statistics, Group invariance, Homogeneous polynomials, Kernel methods, Spherical harmonics

1. Introduction

The recent popularity of human gut microbiome research has presented numerous data-analytic and statistical challenges (Calle, 2019). Among the many features of microbiomes and metagenomic data, we address their *compositional* nature in this work. Compositional data consist of n observations of $(d + 1)$ non-negative variables, where the values represent the relative proportions to other variables in the data. Compositional data are commonly observed in various scientific fields, including biochemistry, ecology, finance, and economics,

*. Corresponding author

among others. The most notable aspect of compositional data is the restriction on their domain, where the sum of the variables is fixed, lying within a simplex denoted as Δ^d:

$$\Delta^d = \left\{(x_1, \dots, x_{d+1}) \in \mathbb{R}^{d+1} \mid \sum_{i=1}^{d+1} x_i = 1,\ x_i \geq 0, \forall i\right\}, \tag{1}$$

referred to in this article as the "compositional simplex." The inclusion of zeros in (1) is crucial, as most microbiome data contain a substantial number of zeros.

The log-ratio transformation, which transforms an open simplex to a vector space, is arguably the most prominent approach to handling compositional data because it enables multivariate methods to be performed on a vector space (Filzmoser et al., 2009; Tomassi et al., 2019; Susin et al., 2020). As this transformation requires the data entirely in the open interior of Δ^d, denoted as $\mathcal{S}^d$, zeros in the data are typically treated by substituting them with a small positive number. However, it has been observed that the analysis results can significantly depend on how zeros are handled (Lubbe et al., 2021). Furthermore, transforming Δ^d into a vector space induces non-trivial topological issues, particularly when the data points are located at the boundary of Δ^d (Park et al., 2022).

According to Aitchison (1994), "any meaningful function of a composition must satisfy the requirement $f(ax) = f(x)$ for any $a \neq 0$." In the field of geometry and topology, a space that comprises such functions is referred to as a *projective space*, denoted by $\mathbb{P}^d$. Consequently, projective geometry emerges as the natural choice for modeling compositional data and is also in accordance with the original philosophy in Aitchison (1994). Since points within compositional domains cannot have negative values, it becomes evident that a compositional domain can be represented as a *positive cone* $\mathbb{P}^d_{\geq 0}$ within a complete projective space. A key property of projective spaces is that altering the length of a vector in $\mathbb{P}^d$ through stretching or shrinking does *not* change the corresponding point. Hence, it is possible to stretch a point within Δ^d to a point in the first orthant sphere by dividing it by its ℓ_2 norm. This is feasible because each point in Δ^d or $\mathbb{S}^d_{\geq 0}$ effectively represents a line passing through the origin in $\mathbb{R}^{d+1}$. Figure 1 visually demonstrates this idea, and it should be noted that "stretching" is not a transformation in the context of projective geometry.

In this article, the spherical geometric representation plays a particularly important role. Realizing $\mathbb{S}^d_{\geq 0}$ as the "universal cover" of the compositional domain, we show that it is the strict fundamental domain of the reflection group action $\Gamma \curvearrowright \mathbb{S}^d$, which yields $\mathbb{S}^d_{\geq 0} = \mathbb{S}^d/\Gamma$. Therefore, the fundamental geometric foundation of compositional data $\mathbb{P}^d_{\geq 0}$ can be represented in three equivalent forms, namely the positive orthant sphere $\mathbb{S}^d_{\geq 0}$, the spherical quotient $\mathbb{S}^d/\Gamma$, and the traditional compositional simplex Δ^d. We will rigorously show the equivalence of these representations, establishing $\mathbb{P}^d_{\geq 0} \cong \mathbb{S}^d_{\geq 0} \cong \mathbb{S}^d/\Gamma \cong \Delta^d$ in Section 2. While each of these equivalent representations provides a different perspective on the geometry of the data, we particularly focus on the spherical quotient $\mathbb{S}^d/\Gamma$ for developing a new framework for compositional data analysis. As an immediate utilization of the group quotient, we discuss a novel nonparametric compositional density estimation method. We also establish the asymptotic normality of the integral squared error of the estimator, which can be useful for a goodness-of-fit test.

Our ultimate goal in this work is to construct reproducing kernel structures specifically tailored for compositional data. The aforementioned geometric interpretation of the

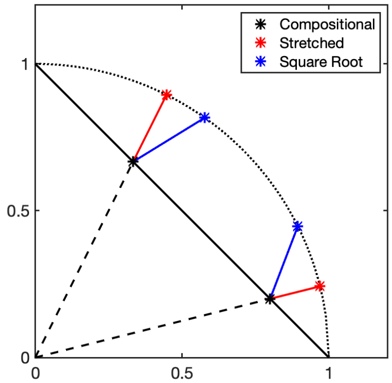


Figure 1: Illustration of the stretching action from Δ^1 to $\mathbb{S}^1$. It is observed that the stretching preserves the relative compositions, which is not the case for the square root transformation.

compositional domain as $\mathbb{S}^d/\Gamma$ establishes a novel connection between compositional data and *directional data*, which refer to data existing on spheres. This connection leads us to investigate how the functions defined on the compositional domain are related to those on the sphere. Specifically, we focus on square-integrable functions $L^2(\mathbb{S}^d)$ on the sphere, known as *spherical harmonics*. Notably, the theory of spherical harmonics reveals that each Laplacian eigenspace of $L^2(\mathbb{S}^d)$ is a reproducing kernel Hilbert space (RKHS). By leveraging the structure of RKHS on spheres and also "averaging out" the group action at the reproducing kernel level using orbital integrals, we construct a reproducing kernel structure on compositional domains.

The proposed compositional reproducing kernel introduces a novel paradigm for analyzing compositional data. It empowers us to utilize a diverse range of machine-learning techniques effectively with compositional data. Its immediate usefulness will be demonstrated through the application of representer theorems for the minimum norm optimization problem and subsequently with an application of kernel support vector machine.

In addition, we introduce a compositional kernel exponential family, which serves as a versatile distributional model suitable for representing compositional data. We observe that no distribution theory based on the log transformation provides a "natural" distribution on Δ^d with non-vanishing boundaries. Some previous works used the spherical Kent distribution to model compositional data that have been transformed by a square-root transformation (Scealy and Welsh, 2014; Paine et al., 2020). However, the compositional Kent distribution requires a large concentration parameter to ensure that most of the mass lies in the first orthant. Otherwise, a "density folding" adjustment is often necessary.

We note that there are two main aspects of our work that are shared with the current mainstream approaches to compositional data analysis. Firstly, the proposed framework acknowledges and embraces the inherent nonlinear nature of compositional domains Δ^d, which is also implied in the so-called Aitchison geometry induced by the log transformation. We find links between compositional data analysis and non-Euclidean geometry in the new compositional domain $\mathbb{S}^d_{\geq 0}$, which, equipped with non-linearity and compact topology, offers a more favorable environment for constructing meaningful distribution theories. Secondly, we establish a connection to a structure that exhibits linearity, namely, RKHS. One of the strong motivations of the log transformations is the ability to carry out "linear" multivariate analysis methods based on Euclidean geometry such as linear regression and principal component analysis. In this work, we utilize the projective geometry to "linearize" compositional data through kernel mean embedding (Muandet et al., 2017), which enables classical linear techniques to be effectively applied to compositional data.

1.1 Structure of the Article

We describe briefly the content of the main sections of this article:

In Section 2, we will establish a new geometric foundation for compositional domains using projective geometry and spherical geometry. We emphasize that the classical model based on the closed simplex Δ^d is topologically equivalent to this new foundation. The topological equivalence is represented as the following:

$$\Delta^d \cong \mathbb{P}^d_{\geq 0} \cong \mathbb{S}^d/\Gamma \cong \mathbb{S}^d_{\geq 0}, \tag{2}$$

where $\mathbb{S}^d_{\geq 0}$ is the first orthant sphere, which is also the fundamental domain of the group action $\Gamma \curvearrowright \mathbb{S}^d$. All of the four spaces in (2) will be referred to as "compositional domains". We present a direct application of our framework by proposing a compositional density estimation method based on the principles of spherical density estimation. We demonstrate that our compositional density estimator exhibits desirable properties, including its integral squared error that is asymptotically normally distributed.

Section 3 will be focused on the construction of compositional reproducing kernel Hilbert spaces (RKHS). Our approach relies on leveraging the reproducing kernel structures on spheres, which are derived from the rich theory of spherical harmonics. While Wahba (1981) previously constructed splines using reproducing kernel structures of spherical harmonics on the 2-dimensional sphere ($\mathbb{S}^2$), their work was limited to this specific case. In contrast, our theory encompasses the general d-dimensional case, necessitating a comprehensive understanding of spherical harmonics theory. To facilitate this, we will provide a review of spherical harmonics at the beginning of Section 3.

Section 4 will present two practical applications of the proposed compositional reproducing kernels. Firstly, we prove the RHKS representer theorem for the minimum norm regularization problem. Proof of the theorem requires a careful treatment of a notably distinctive feature of the compositional RKHS, which is that the space consists of degree $2m$ homogeneous polynomials, i.e., it is finite-dimensional with no transcendental functions. Thus, linear independence for distinct data points is not readily available. However, we show that with a sufficiently high degree m, linear independence still holds. An empirical example with support vector machines will also be presented to showcase its practical relevance in real-world problems. While our formulation of the representer theorem is not necessarily new from a pure RKHS-theoretical standpoint, its purpose is to show that intuitions derived from traditional statistical learning can still be employed in compositional data analysis. The second application is the construction of the compositional exponential family, which can be a useful model for the underlying distribution of compositional data.

This flexible distribution family is able to account for various dependence structures among compositional variables.

We conclude the paper in Section 5 with some discussions on relevant topics such diffusion kernels by Lafferty and Lebanon (2005) and a special orthogonal group $SO(d+1)$. We also discuss some computational issues regarding the implementation of the proposed approach.

2. New Geometric Foundation of Compositional Domains

In this section, we introduce a novel characterization of compositional domains as a cone $\mathbb{P}^d_{\geq 0}$ in a projective space. This representation allows us to further interpret compositional domains as spherical quotients obtained through reflection groups. We illustrate the practical benefits of this new approach with an application in compositional density estimation.

2.1 Projective Geometry and Spread-out Construction

Compositional data consist of the relative proportions of $d+1$ variables, implying that each observation can be associated with a point in a projective space. In particular, a d-dimensional projective space $\mathbb{P}^d$ comprises one-dimensional linear subspaces of $\mathbb{R}^{d+1}$. In projective geometry, points along a line passing through the origin are considered equivalent. Thus, instead of using classical linear coordinates $(x_1, \ldots, x_{d+1})$, a point in $\mathbb{P}^d$ can be expressed as a projective coordinate $(x_1 : \cdots : x_{d+1})$ possessing the following property:

$$(x_1 : x_2 : \cdots : x_{d+1}) = (\lambda x_1 : \lambda x_2 : \cdots : \lambda x_{d+1}), \quad \text{for any } \lambda \neq 0.$$

In compositional data analysis, it is natural to consider the *non-negative projective space* as the appropriate ambient space. The cone $\mathbb{P}^d_{\geq 0}$, defined as follows:

$$\mathbb{P}^d_{\geq 0} = \left\{ (x_1 : x_2 : \cdots : x_{d+1}) \in \mathbb{P}^d \, | (x_1 : x_2 : \cdots : x_{d+1}) = (|x_1| : |x_2| : \cdots : |x_{d+1}|) \right\},$$

is a new geometric foundation in compositional data analysis, and we refer to it as *the fundamental compositional domain.*

From the new geometric perspective, the compositional simplex Δ^d can be viewed as an equivalent representation of $\mathbb{P}^d_{\geq 0}$. This equivalence arises because each point in $\mathbb{P}^d_{\geq 0}$, which corresponds to a non-negative line passing through the origin in $\mathbb{R}^{d+1}$, intersects the compositional simplex Δ^d at a unique point. Conversely, each point within Δ^d determines a unique line passing through the origin in $\mathbb{R}^{d+1}$. This leads to the following equivalence:

$$\mathbb{P}^d_{\geq 0} \cong \Delta^d. \tag{3}$$

Another equivalent representation of the fundamental compositional domain is the first orthant sphere $\mathbb{S}^d_{\geq 0}$, which is the non-negative part of a d-dimensional unit sphere $\mathbb{S}^d$:

$$\mathbb{S}^d = \left\{ (x_1, x_2, \ldots, x_{d+1}) \in \mathbb{R}^{d+1} \mid \sum_{i=1}^{d+1} x_i^2 = 1 \right\}.$$

The following lemma states that $\mathbb{S}^d_{\geq 0}$ is another representation of $\mathbb{P}^d_{\geq 0}$.

Lemma 1. *There is a canonical identification of Δ^d with $\mathbb{S}^d_{\geq 0}$, namely,*

$$\begin{tikzcd} \Delta^d \arrow[r, shift left, "f"] & \mathbb{S}^d_{\geq 0} \arrow[l, shift left, "g"] \end{tikzcd} ,$$

where f is the inflation map and g is the contraction map, with both f and g being continuous and inverse to each other.

Proof It is straightforward to construct the inflation map f. For $v \in \Delta^d$, it is easy to see that $f(v) \in \mathbb{S}^d_{\geq 0}$ when $f(v) = v/\|v\|_2$, where $\|v\|_2$ is the ℓ_2 norm of v. Note that the inflation map ensures that $f(v)$ is in the same projective space as v. To construct the contraction map g, for $s \in \mathbb{S}^d_{\geq 0}$ we define $g(s) = s/\|s\|_1$, where $\|s\|_1$ is the ℓ_1 norm of s and see that $g(s) \in \Delta^d$. One can easily check that both f and g are continuous and inverse to each other. ■

Figure 1 illustrates the mapping f mentioned in Lemma 1. It visually demonstrates that the points x and $f(x)$ correspond to the same point within the projective space $\mathbb{P}^d$. Consequently, this identification maintains the projective characteristics of compositional domains, which sets it apart from the square-root transforms.

The most crucial representation of the fundamental compositional domain $\mathbb{P}^d_{\geq 0}$ in this work is $\mathbb{S}^d/\Gamma$, which denotes the quotient space of the group action $\Gamma \curvearrowright \mathbb{S}^d$. We define the reflection group Γ as follows:

Definition 2. *The reflection group Γ is a subgroup of general linear group $GL(d+1)$ and it is generated by $\{\gamma_i, i = 1, \ldots, d+1\}$. Given the elementary basis $\{e_1, \ldots, e_{d+1}\}$ for $\mathbb{R}^{d+1}$, the reflection γ_i is a linear map specified via:*

$$\gamma_i : \ (x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_{d+1}) \mapsto (x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_{d+1}).$$

Each reflection γ_i in the group Γ is an isometry of $\mathbb{S}^d$, which we denote by $\Gamma \curvearrowright \mathbb{S}^d$. Consequently, Γ can be regarded as a discrete subgroup of the isometry group of $\mathbb{S}^d$. In what follows, we establish that $\mathbb{S}^d_{\geq 0}$ serves as a fundamental domain for the group action $\Gamma \curvearrowright \mathbb{S}^d$ in a topological sense. It is important to note that there is no universally defined fundamental domain, but we will adopt the approach in Beardon (2012). To begin with, we introduce the concept of an *orbit*. For a given point $z \in \mathbb{S}^d$, the orbit associated with the group Γ is defined as follows:

$$\text{Orbit}^{\Gamma}_{z} = \{\gamma(z), \forall \gamma \in \Gamma\}. \tag{4}$$

Note that one can decompose $\mathbb{S}^d$ into a disjoint union of orbits. The size of an orbit does not necessarily match the size of the group $|\Gamma|$ due to the existence of a *stabilizer subgroup* of Γ, defined as

$$\Gamma_z = \{\gamma \in \Gamma, \gamma(z) = z\}. \tag{5}$$

Every element in $\text{Orbit}^{\Gamma}_{z}$ has isomorphic stabilizer subgroups. Consequently, the size of $\text{Orbit}^{\Gamma}_{z}$ is given by the quotient $|\Gamma|/|\Gamma_z|$, where $|\cdot|$ represents the cardinality of a set. Note that for the action $\Gamma \curvearrowright \mathbb{S}^d$, possible sizes of stabilizer subgroups are finite.

Definition 3. *Let G act properly and discontinuously on a d-dimensional sphere, with $d > 1$. A fundamental domain for the group action G is a closed subset F of the sphere such that every orbit of G intersects F at at least one point and if an orbit intersects with the interior of F, then it intersects F at only one point.*

A fundamental domain is *strict* if every orbit of G intersects F at exactly one point. The following proposition identifies $\mathbb{S}^d_{\geq 0}$ as the quotient topological space $\mathbb{S}^d/\Gamma$, i.e., $\mathbb{S}^d_{\geq 0} = \mathbb{S}^d/\Gamma$.

Proposition 4. *Let $\Gamma \rightsquigarrow \mathbb{S}^d$ be the group action described in Definition 2. Then $\mathbb{S}^d_{\geq 0}$ is a strict fundamental domain.*

In topology, there is a natural quotient map $\mathbb{S}^d \to \mathbb{S}^d/\Gamma$. We define a contraction mapping $c : \mathbb{S}^d \to \mathbb{S}^d_{\geq 0}$ that operates by taking the absolute values of each component, namely $(x_1, \dots, x_{d+1}) \mapsto (|x_1|, \dots, |x_{d+1}|)$. Then it is straightforward to see that c is indeed the topological quotient map $\mathbb{S}^d \to \mathbb{S}^d/\Gamma$, under the identification $\mathbb{S}^d_{\geq 0} = \mathbb{S}^d/\Gamma$.

Through the application of (3), Lemma 1, and Proposition 4, we have established the following equivalences:

$$\mathbb{P}^d_{\geq 0} \cong \Delta^d \cong \mathbb{S}^d_{\geq 0} \cong \mathbb{S}^d/\Gamma. \tag{6}$$

Remark 5. From this point onwards in the paper, we will freely interchange between these four characterizations of a compositional domain. However, the last representation of a compositional domain $\mathbb{S}^d/\Gamma$ (the spherical quotient representation) plays the most important role in our development of new framework for compostional data analysis:

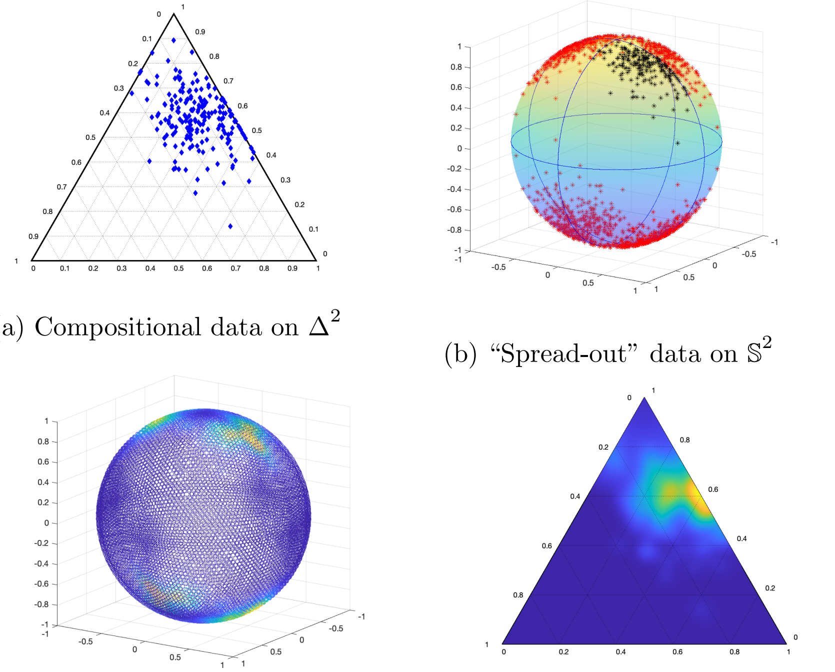
- ♠ The spherical quotient representation $\mathbb{S}^d/\Gamma$ reinterprets compositional data problems into Γ-invariant direction statistics with the help of spread-out construction (see (7)). This philosophy plays a key role in this paper, and two important applications of philosophy are compositional density estimation theory (Section 2.2) and constructing compositional reproducing kernels (Section 3).

- ♠ A longstanding challenge in compositional data analysis is due to the existence of the boundary in a compositional simplex. Recall that the fundamental geometric objects for directional statistics are spheres that are compact with no boundary. In many ways, the perspective of Γ-invariant directional statistics has essentially eliminated the boundary issues in compositional data analysis.

By utilizing the equivalence in (6), one can transform a compositional data problem into one on a sphere using the spread-out construction. The main concept behind this approach is to associate a compositional data point $z \in \Delta^d = \mathbb{S}^d_{\geq 0}$ with the Γ-orbit $\mathrm{Orbit}^{\Gamma}_z \subset \mathbb{S}^d$ as defined in (4). Formally, given a point $z \in \Delta^d$, we construct the following *dataset* (not necessarily a set due to possible repetitions):

$$c^{-1}(z) = \left\{ |\Gamma_{z'}| \text{ copies of } z', \text{ for } z' \in \mathrm{Orbit}^{\Gamma}_z \right\}, \tag{7}$$

where $\Gamma_{z'}$ represents the stabilizer subgroup of Γ with respect to z' as defined in (5). In general, if there are n observations in Δ^d, the spread-out construction will generate a dataset



(a) Compositional data on Δ^2

(b) “Spread-out” data on $\mathbb{S}^2$

(c) Density estimate on $\mathbb{S}^2$

(d) “Pulled-back” estimate on Δ^2

Figure 2: Toy compositional data on the simplex Δ^2 in (a) are spread out to a sphere $\mathbb{S}^2$ in (b). The density estimate on $\mathbb{S}^2$ in (c) are pulled back to Δ^2 in (d).

with $n2^{d+1}$ observations on $\mathbb{S}^d$, where observations with zero coordinates are replicated. We illustrate this concept in Figure 2 (a) and (b) with a toy data set with $d = 2$ and $n = 100$.

The spread-out construction presented in (7) essentially generates a directional dataset from the original compositional dataset. Leveraging this concept, it becomes quite feasible to adapt methods from directional statistics for practical application to compositional data. For instance, one can conduct independence or uniformity tests for a compositional dataset by simply performing an appropriate test for the "augmented" directional dataset (Jupp and Spurr, 1985; Jupp, 2008). Moreover, the spread-out approach could be used for obtaining density estimates for compositional data, which we will delve into in the following subsection.

2.2 Application to Compositional Density Estimation

Directional statistics has a rich history of exploring density estimation techniques for spherical data, tracing back to the late 1970s (Beran, 1979). Subsequently, Hall et al. (1987) and Bai et al. (1989) laid the groundwork for a comprehensive framework in spherical density estimation theory. In this section, we formulate nonparametric density estimation for compositional data, building upon the rich development in spherical density estimation. The main idea is that instead of directly estimating the density on Δ^d, we can obtain density estimates for the spread-out data on $\mathbb{S}^d$, which in turn allows us to derive compositional density estimates.

Consider a random vector Z distributed on $\mathbb{S}^d_{\geq 0}$, or equivalently on the compositional simplex Δ^d. Let $p(\cdot)$ denote its probability density function. The following proposition provides an expression for the density of the spread-out random vector $\Gamma(Z)$ defined on the entire sphere $\mathbb{S}^d$.

Proposition 6. *Let Z be a random variable on $\mathbb{S}^d_{\geq 0}$ with probability density $p(\cdot)$. Then the induced random variable $\Gamma(Z) = \{\gamma(Z)\}_{\gamma \in \Gamma}$, has the following density $\tilde{p}(\cdot)$ on $\mathbb{S}^d$:*

$$\tilde{p}(z) = \frac{|\Gamma_z|}{|\Gamma|} p(c(z)), \ z \in \mathbb{S}^d,$$

where $|\Gamma_z|$ is the cardinality of the stabilizer subgroup Γ_z of z.

Let c_* denote the corresponding operation for functions, analogous to the contraction map c applied to data points. It is evident that when given a probability density $\tilde{p}$ on $\mathbb{S}^d$, we can derive the original density on the compositional domain by utilizing the "pull back" operation c_*:

$$p(z) = c_*(\tilde{p})(z) = \sum_{x \in c^{-1}(z)} \tilde{p}(x), \quad z \in \mathbb{S}^d_{\geq 0}.$$

Now consider estimating the density on $\mathbb{S}^d$ with the spread-out data. For $x_1, \ldots, x_n \in \mathbb{S}^d$, a density estimate by Hall et al. (1987) is given by

$$\hat{f}_n(z) = \frac{c_h}{n} \sum_{i=1}^{n} K\left(\frac{1 - z^T x_i}{h}\right), \ z \in \mathbb{S}^d,$$

where K is a kernel function that satisfies common assumptions in Assumption 1, and c_h is a normalizing constant. Applying this to the spread-out data $\gamma(x_i)$, $i = 1, \ldots, n$, we have a density estimate of $\tilde{p}(\cdot)$ defined on $\mathbb{S}^d$:

$$\hat{f}_n^{\Gamma}(z) = \frac{c_h}{n|\Gamma|} \sum_{1 \le i \le n, \gamma \in \Gamma} K\left(\frac{1 - z^T \gamma(x_i)}{h}\right), \ z \in \mathbb{S}^d,$$

from which a density estimate on the compositional domain is obtained by applying c_*. That is,

$$\hat{p}_n(z) = c_* \hat{f}_n^{\Gamma}(z) = \sum_{x \in c^{-1}(z)} \hat{f}_n^{\Gamma}(x), \quad z \in \mathbb{S}^d_{\ge 0}. \tag{8}$$

Figure 2 (c) and (d) illustrate this process with toy data using the Gaussian kernel.

The consistency of the spherical density estimate $\hat{f}_n$ has been established by Zhao and Wu (2001) and García-Portugués et al. (2015), where it is shown that the integral squared error (ISE), $\int_{\mathbb{S}^d} (\hat{f}_n - f)^2 dz$, where the integration is with respect to the Lesbegue measure on $\mathbb{S}^d$, follows a normal distribution as the sample size increases. Similarly, we can show that the ISE of the proposed compositional density estimator $\hat{p}_n$ on the compositional domain also converges to a normal distribution through the central limit theorem (CLT). However, it is worth noting that the CLT for the ISE of spherical densities in Zhao and Wu (2001) assumes a finite support condition on the kernel function K. Although García-Portugués et al. (2015) relaxes this condition in their analysis of directional-linear data, their result does not directly apply to the pure directional context, thus their proof is not directly applicable. The following theorem establishes the asymptotic normality for ISE of our density estimator in (8). The proof of the theorem can be found in Appendix A.1.

Theorem 7. *Asymptotic normality for ISE holds for both directional and compositional data under the mild conditions specified in Appendix A.1, without finite support condition on the kernel function K.*

2.2.1 Comparison with Existing Approaches

We also give an empirical comparison between the proposed estimate in (8) and two existing methods for compositional density estimation from Aitchison and Lauder (1985). The existing methods are based on two kernels defined on open simplex $\mathcal{S}^d$: Dirichlet kernel and logistic-normal kernel. The Dirichlet kernel is based on the Dirichlet density function

$$D(x|\alpha) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_{d+1})}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_{d+1})} x_1^{\alpha_1 - 1} \ldots x_{d+1}^{\alpha_{d+1} - 1}, \quad x \in \mathcal{S}^d,$$

where $\alpha = (\alpha_1, \ldots, \alpha_{d+1}), \alpha_i > 0$ is the concentration parameter. Note that the mode of Dirichlet distribution is located at $(\alpha_i - 1) / \sum_j (\alpha_j - 1)$ for each x_i if every $\alpha_i > 1$, and the density becomes more concentrated at the mode as α_i increases. The Dirichlet kernel for density estimation is given by

$$K(x, x') = D(x | \mathbf{1} + x'/h), \quad x, x' \in \mathcal{S}^d,$$

where $\mathbf{1}$ is a $d + 1$ dimension vector of ones and h is a bandwidth parameter.

The logistic-normal kernel uses the additive log-ratio (alr) transformation which maps $x \in \mathcal{S}^d$ to $y \in \mathbb{R}^d$ by

$$y_i = \log(x_i/x_{d+1}), \quad i = 1, \ldots, d, \tag{9}$$

with which the following multivariate Gaussian kernel

$$\det(2\pi h\Sigma)^{-1/2}\exp\Big\{-\frac{1}{2}(y-y')^T(h\Sigma)^{-1}(y-y')\Big\}, \quad y, y' \in \mathbb{R}^d,$$

is used for estimating the density of y where h is again a bandwidth parameter. Here, Σ is set as a sample covariance matrix of the alr-transformed data. Note that (9) implies

$$x_i = \exp(y_i)/\{\exp(y_1) + \cdots + \exp(y_d) + 1\}, \quad i = 1, \ldots, d,$$

based on which a logistic-normal kernel $K(x, x')$ can be defined. Finally, we estimate the density at arbitrary point $x \in \mathcal{S}^d$ as

$$\hat{p}(x) = \frac{1}{n}\sum_{i=1}^{n} K(x, x_i). \tag{10}$$

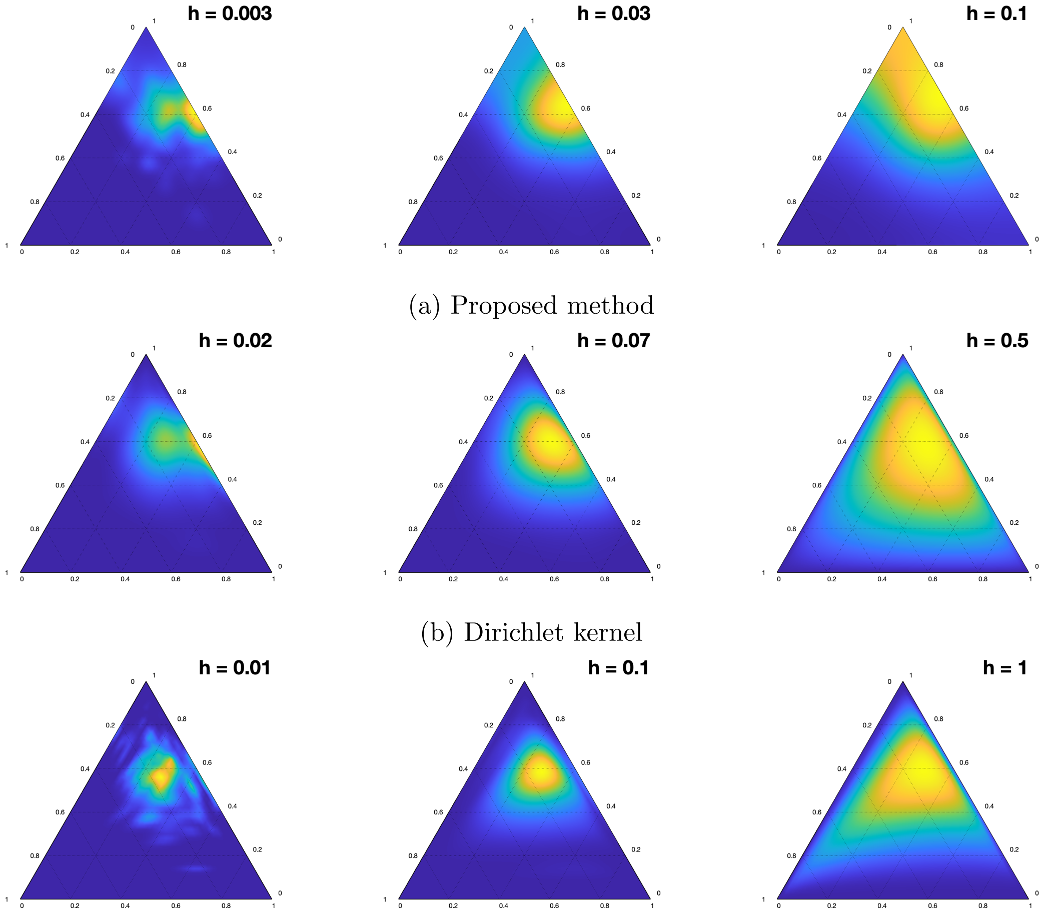
It is important to note that these two density estimates are defined only for the open simplex $\mathcal{S}^d$. Therefore, following the recommendation of Aitchison and Lauder (1985), we handle the zeros in each observation by replacing them with $\delta(c+1)(d+1-c)/(d+1)^2$. Here, c is the number of zero components in each observation, and δ is the maximum rounding-off error, which is set to 0.001.

Figure 3 compares the density estimate of the toy data shown in Figure 2 (a) based on the proposed method and the two existing methods based on the Dirichlet and logistic-normal kernels respectively. The results clearly demonstrate that the proposed method provides a more accurate representation of the data with minimal distortion. Both the Dirichlet and logistic-normal kernel density estimates exhibit distortions in the overall shapes of the density, particularly towards the center of the simplex and near the boundaries. Notably, the logistic-normal kernel density estimates in (c) show an undesirable concentration of mass towards the vertices. This distortion seems to be a consequence of the alr transformation, which is evident from the equivalent density estimation plots of the alr-transformed data in $\mathbb{R}^2$ shown in Figure 4.

3. Reproducing Kernels of Compositional Data

In this section, our goal is to construct a reproducing kernel Hilbert space (RKHS) for compositional data, utilizing the spherical quotient representation $\mathbb{S}^d/\Gamma$ discussed in Section 2. The key concept behind this construction is the quotient map $\pi : \mathbb{S}^d \to \mathbb{S}^d/\Gamma$, which allows us to establish a connection between function spaces defined on spheres and function spaces defined on compositional domains. By leveraging this relationship, we can construct reproducing kernels on the compositional domain $\mathbb{S}^d/\Gamma$ based on the ones defined on $\mathbb{S}^d$.

Reproducing kernels on $\mathbb{S}^d$ were actually discovered by Laplace and Legendre in the 19th century, though they were referred to as "zonal spherical functions" at that time.



(a) Proposed method

(b) Dirichlet kernel

(c) Logistic-normal kernel

Figure 3: Kernel density estimates of the toy compositional data from Figure 2, using (a) the proposed method in (8); the traditional estimates in (10) with (b) Dirichlet kernel and (c) logistic-normal kernel with varying bandwidths.

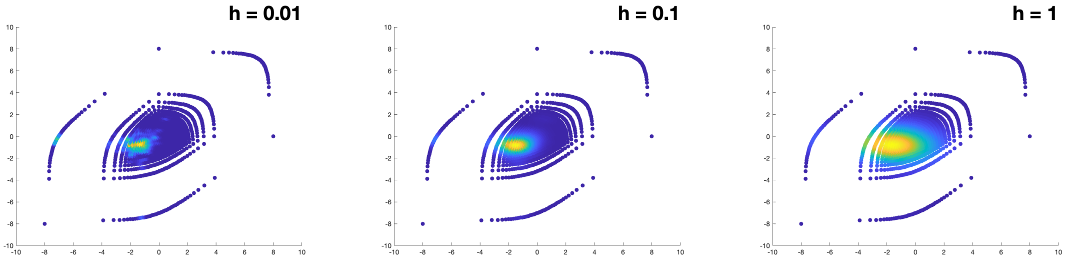


Figure 4: Kernel density estimates of the toy compositional data from Figure 2 in the alr-transformed space with Gaussian kernel. Each of the panels corresponds to the panels in Figure 3(c).

Both the theory of spherical harmonics and reproducing kernel Hilbert spaces (RKHS) have found applications in various fields, including functional analysis, representation theory of Lie groups, and quantum mechanics. In statistics, the successful application of RKHS in spline models by Wahba (1981) played a significant role in popularizing RKHS theory for $\mathbb{S}^d$. Their reproducing kernels were built using zonal spherical functions and motivated by spline models on the sphere $\mathbb{S}^2$. In contrast, our motivation for exploring reproducing structures on spheres is unrelated to spline models. However, we share a common foundation, namely the theory of spherical harmonics, as the building blocks for our approach.

Building upon the representation of a compositional domain $\mathbb{P}^d_{\geq 0}$ as $\mathbb{S}^d/\Gamma$, our aim is to construct reproducing kernels for compositions by leveraging reproducing kernel structures on spheres. Considering that spherical harmonics theory provides reproducing kernel structures on $\mathbb{S}^d$, and a compositional domain $\mathbb{S}^d/\Gamma$ is topologically covered by spheres with the group Γ leads to the two following questions, for both of which the answers are positive.

(i) Can function spaces on $\mathbb{S}^d/\Gamma$ be identified with Γ-invariant functions on $\mathbb{S}^d$?

(ii) Is it possible to construct Γ-invariant kernels using spherical reproducing kernels, with the hope that the Γ-invariant kernels can serve as reproducing kernels on $\mathbb{S}^d/\Gamma$?

Through the consideration of Γ-invariant objects in spherical function spaces, we will construct reproducing kernel structures for compositional domains, thereby establishing compositional RKHS. This opens up new opportunities to develop a framework for compositional data analysis, where we elevate compositional data points to functions via reproducing kernels, and classical statistical concepts such as means and variance-covariances are transformed into linear functionals and linear operators within the function space. This "kernel methods" framework provides fresh perspectives on various essential statistical topics, including dimension reduction, regression analysis, and many other inference problems.

3.1 Spherical Harmonics and Reproducing Kernels

This section provides a brief overview of the theory of spherical harmonics. For a more comprehensive introduction, we refer to Atkinson and Han (2012). Recall that a periodic function f in $\mathbb{R}$ with a period of one can be represented by Fourier expansions of the form:

$$f(x) \sim A_0 + \sum_{n=1}^{\infty} \big[a_n \cos(2\pi n x) + b_n \sin(2\pi n x)\big].$$

Noting that a function with periodicity one can be regarded as a function on the unit circle $\mathbb{S}^1$, we can view spherical harmonics theory as a generalization of the Fourier expansion, to a general sphere $\mathbb{S}^d$.

Let $L^2(\mathbb{V})$ denote the space of square-integrable functions on $\mathbb{V}$. For $f \in L^2(\mathbb{S}^d)$, consider the Laplacian

$$\Delta f = \sum_{i=1}^{d+1} \frac{\partial^2 f}{\partial x_i^2},$$

and let $\mathcal{H}_i$ denote the i-th eigenspace of the Laplacian operator. It is well known that the space $L^2(\mathbb{S}^d)$ can be orthogonally decomposed as follows:

$$L^2(\mathbb{S}^d) = \bigoplus_{i=1}^{\infty} \mathcal{H}_i, \tag{11}$$

where the orthogonality in this decomposition is defined with respect to the inner product in $L^2(\mathbb{S}^d)$, given by $\langle f, g\rangle = \int_{\mathbb{S}^d} f(t)\overline{g(t)}dt$.

Let $\mathcal{P}_i(d+1)$ be the space of homogeneous polynomials of degree i in $d+1$ coordinates on $\mathbb{S}^d$. A homogeneous polynomial is a polynomial whose terms are all monomials of the same degree, e.g., $\mathcal{P}_4(3)$ includes $xy^3 + x^2yz$. Further, let $H_i(d+1)$ be the space of homogeneous harmonic polynomials of degree i on $\mathbb{S}^d$, i.e.,

$$H_i(d+1) = \{P \in \mathcal{P}_i(d+1) \mid \Delta P = 0\}. \tag{12}$$

For example, $x^3y + xy^3 - 3xyz^2$ and $x^4 - 6x^2y^2 + y^4$ are members of $H_4(3)$.

Importantly, the theory of spherical harmonics establishes that each eigenspace $\mathcal{H}_i$ in the decomposition (11) is precisely the space $H_i(d+1)$. This crucial result implies that any function in $L^2(\mathbb{S}^d)$ can be approximated by a direct sum decomposition of orthogonal homogeneous harmonic polynomials. Moreover, it is known that the Laplacian constraint in (12) is not necessary, as stated in the following proposition:

Proposition 8. *Let $\mathcal{P}_m(d+1)$ be the space of degree m homogeneous polynomial of $d+1$ variables on $\mathbb{S}^d$ and $\mathcal{H}_i$ be the ith eigenspace of $L^2(\mathbb{S}^d)$. Then*

$$\mathcal{P}_m(d+1) = \bigoplus_{i=\lceil m/2\rceil - \lfloor m/2\rfloor}^{\lfloor m/2\rfloor} \mathcal{H}_{2i},$$

where $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ stand for ceiling and flooring integers, respectively.

Proposition 8 reveals that any L^2 function on $\mathbb{S}^d$ can be approximated by homogeneous polynomials. An important feature of spherical harmonics is that it provides reproducing structures on spheres. For the subsequent discussion, we will focus on a fixed Laplacian eigenspace $\mathcal{H}_i$ within $L^2(\mathbb{S}^d)$, which is a finite-dimensional Hilbert space on $\mathbb{S}^d$. It is important to note that the entire Hilbert space $L^2(\mathbb{S}^d)$ does not have a reproducing kernel given that the Delta functional on $L^2(\mathbb{S}^d)$ is *not* a bounded functional[1].

For each Laplacian eigenspace $\mathcal{H}_i$ within $L^2(\mathbb{S}^d)$, we can define a linear functional L_x on $\mathcal{H}_i$ such that, for any $Y \in \mathcal{H}_i$, $L_x(Y) = Y(x)$, where x is a fixed point in $\mathbb{S}^d$. It is known that there exists a function $k_i(x,t)$ such that:

$$L_x(Y) = Y(x) = \int_{\mathbb{S}^d} Y(t) k_i(x,t) dt, \ x \in \mathbb{S}^d.$$

The function $k_i(x,t)$, sometimes referred to as a zonal spherical function, serves as the representation of the functional $L_x(Y)$. Notably, these functions $k_i(x,t)$ can be considered as "reproducing kernels" in the sense of Aronszajn (1950). In other words, each Laplacian eigenspace $\mathcal{H}_i$ can be viewed as an RKHS on $\mathbb{S}^d$. The following proposition provides some fundamental properties of $k_i(x,t)$. The proofs of these properties can be found in various modern references on spherical harmonics, such as Stein and Weiss (1971).

Proposition 9. *The following properties hold for the function $k_i(x,t)$, which is also the reproducing kernel inside $\mathcal{H}_i \subset L^2(\mathbb{S}^d)$ with dimension a_i.*

(a) For any orthonormal basis $\{Y_1, \dots, Y_{a_i}\}$ in $\mathcal{H}_i$, we can express the kernel $k_i(x,t) = \sum_{i=1}^{a_i} \overline{Y_i(x)} Y_i(t)$, but $k_i(x,t)$ does not depend on the choice of basis.

(b) $k_i(x,t)$ is a real-valued function and symmetric, i.e., $k_i(x,t) = k_i(t,x)$.

(c) For any orthogonal matrix $R \in O(d+1)$, we have $k_i(x,t) = k_i(Rx, Rt)$.

(d) $k_i(x,x) = \dfrac{a_i}{\mathrm{vol}(\mathbb{S}^d)}$ for any $x \in \mathbb{S}^d$.

(e) $k_i(x,t) \leq \dfrac{a_i}{\mathrm{vol}(\mathbb{S}^d)}$ for any $x, \ t \in \mathbb{S}^d$.

Remark 10. The proposition mentioned above may appear obvious from a traditional perspective as if it could be found in any textbook. Readers with extensive experience in RKHS theory might consider it a trivial statement. However, we would like to emphasize two important points. First, function spaces over underlying spaces with different topological structures exhibit distinct behaviors. Spheres, being compact without boundaries, have function spaces with Laplacian operators whose eigenspaces are finite-dimensional and possess reproducing kernel structures. These remarkable properties are not necessarily expected to hold for other general topological spaces. Second, in comparison to more commonly used classical topological spaces, such as unit intervals or vector spaces, spheres possess a more "exotic" topological structure. Although spheres are simply connected, they

[1] At first glance, this may appear contradictory to the discussion on splines on $\mathbb{S}^2$ in Wahba (1981). However, upon closer examination, it is evident that a finiteness constraint was imposed in that context. It is important to note that $L^2(\mathbb{S}^2)$ itself is never claimed to be RKHS. Instead, the RKHS constructed on $\mathbb{S}^2$ is a subspace of $L^2(\mathbb{S}^2)$.

exhibit nontrivial higher homotopy groups. On the other hand, intervals or vector spaces are contractible with trivial homotopy groups. One way to appreciate the theory of spherical harmonics is by recognizing that classical "naïve" expectations can still be defined on spheres.

By utilizing the representation of a compositional domain as $\mathbb{S}^d/\Gamma$, the function space associated with compositional data can be identified as $L^2(\mathbb{S}^d/\Gamma)$. In other words, $L^2(\Delta^d) = L^2(\mathbb{P}^d_{\geq 0}) = L^2(\mathbb{S}^d/\Gamma)$. In the subsequent subsection, we describe the function space within the compositional domain in detail.

3.2 Function Spaces on Compositional Domains

Given the understanding of the function space $L^2(\mathbb{S}^d)$ through spherical harmonics theory, it is natural to establish a connection between $L^2(\mathbb{S}^d/\Gamma)$ and $L^2(\mathbb{S}^d)$. For a function $h \in L^2(\mathbb{S}^d/\Gamma)$, there exists an associated function $\pi^*(h) \in L^2(\mathbb{S}^d)$ obtained through the composition of maps:

$$\pi \circ h : \quad \mathbb{S}^d \xrightarrow{\pi} \mathbb{S}^d/\Gamma \xrightarrow{h} \mathbb{C} .$$

Consequently, the composition $\pi \circ h = \pi^*(h) \in L^2(\mathbb{S}^d)$ naturally leads to an embedding of the function space of compositional domains into that of a sphere. Thus, we have the mapping $\pi^* : L^2(\mathbb{S}^d/\Gamma) \to L^2(\mathbb{S}^d)$.

The embedding π^* identifies the Hilbert space of compositional domains as a subspace of the Hilbert space of spheres. A natural question is how to characterize the subspace in $L^2(\mathbb{S}^d)$ that corresponds to functions on compositional domains. The following proposition states that a function $f \in \mathrm{im}(\pi^*)$ if and only if f is constant on the fibers of the projection map $\pi : \mathbb{S}^d \to \mathbb{S}^d/\Gamma$, almost everywhere. In simpler terms, this means that f takes the same values on all points within each Γ orbit, i.e., on the points that are connected by "sign flipping".

Proposition 11. *The image of the embedding $\pi^* : L^2(\mathbb{S}^d/\Gamma) \to L^2(\mathbb{S}^d)$ consists of functions $f \in L^2(\mathbb{S}^d)$ such that up to a measure zero set, is constant on $\pi^{-1}(x)$ for every $x \in \mathbb{S}^d/\Gamma$, where π is the natural projection $\mathbb{S}^d \to \mathbb{S}^d/\Gamma$.*

We will refer to a function $f \in L^2(\mathbb{S}^d)$ that lies in the image of the embedding π^* as a Γ-*invariant function.* Now we construct the contraction map $\pi_* : L^2(\mathbb{S}^d) \to L^2(\mathbb{S}^d/\Gamma)$, which descends every function on spheres to a function on compositional domains. For $f \in L^2(\mathbb{S}^d)$, we define the associated Γ-invariant function f^Γ as follows:

Proposition 12. *Let f be a function in $L^2(\mathbb{S}^d)$. Then the following function*

$$f^\Gamma(z) = \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} f(\gamma z), \quad z \in \mathbb{S}^d, \tag{13}$$

is a Γ-invariant function.

Proof Each fiber of the projection map $\pi : \mathbb{S}^d \to \mathbb{S}^d/\Gamma$ is Orbit^Γ_z for some z in the fiber. For any other point y on the same fiber with z for the projection π, there exists a reflection

$\gamma \in \Gamma$ such that $y = \gamma z$. Then, this proposition follows from the identity $f^{\Gamma}(z) = f^{\Gamma}(\gamma z)$, which can be easily checked. ■

Note that the contraction map $f \mapsto f^{\Gamma}$ on spheres naturally yields the following map

$$\pi_* : \ L^2(\mathbb{S}^d) \to L^2(\mathbb{S}^d/\Gamma), \text{ with } f \mapsto f^{\Gamma}. \tag{14}$$

Furthermore, π_* has a section given by π^* and the composition $\pi_* \circ \pi^*$ induces the identity map from $L^2(\mathbb{S}^d/\Gamma)$ to itself. In particular, the contraction map π_* is a surjection. The relationship between $L^2(\mathbb{S}^d)$ and $L^2(\mathbb{S}^d/\Gamma)$ found in this section enables us to construct reproducing kernels on Δ^d using $L^2(\mathbb{S}^d)$, in Section 3.4.

3.3 Reduction to Even Degrees Homogeneous Polynomials

In this subsection, we provide a further simplification of the homogeneous polynomials in $\bigoplus_{i=0}^{m} \mathcal{H}_i$. According to Proposition 8, if m is even, then $\mathcal{P}_m(d+1) = \bigoplus_{i=0}^{m/2} \mathcal{H}_{2i}$, and if m is odd, then $\mathcal{P}_m(d+1) = \bigoplus_{i=0}^{(m-1)/2} \mathcal{H}_{2i+1}$, where $\mathcal{P}_m(d+1)$ represents the space of degree m homogeneous polynomials in $d+1$ variables. In either case (whether m is even or odd), the degree of the homogeneous polynomials m is the same as the max$\{2i, \lceil m/2 \rceil - \lfloor m/2 \rfloor \le i \le \lfloor m/2 \rfloor\}$. Consequently, we can decompose $\bigoplus_{i=0}^{m} \mathcal{H}_i$ into the direct sum decomposition of two spaces of homogeneous polynomials:

$$\bigoplus_{i=0}^{m} \mathcal{H}_i = \mathcal{P}_m(d+1) \bigoplus \mathcal{P}_{m-1}(d+1).$$

However, we note that not every function within $\bigoplus_{i=0}^{m} \mathcal{H}_i$ is meaningful for $\Delta^d = \mathbb{S}^d/\Gamma$. For example, it is possible to find a nonzero function $f \in \bigoplus_{i=0}^{m} \mathcal{H}_i$, but its Γ-invariant f^{Γ} is equal to zero. Notably, eigenspaces $\mathcal{H}_i$ with odd values of i do not contribute to $L^2(\mathbb{S}^d/\Gamma)$. In other words, the odd-numbered sum $\bigoplus_{i=0}^{m} \mathcal{H}_{2i+1}$ is "eliminated" and yields zero under the action of π_*. This simplification of the function space greatly streamlines computations. In the following lemma, we will show that any monomial term of odd degree will vanish when its Γ-invariant is taken.

Lemma 13. *For every monomial $\prod_{i=1}^{d+1} x_i^{\alpha_i}$ (each $\alpha_i \ge 0$), if there exits k with α_k being odd, then the monomial $\prod_{i=1}^{d+1} x_i^{\alpha_i}$ is a shadow function, that is, $(\prod_{i=1}^{d+1} x_i^{\alpha_i})^{\Gamma} = 0$.*

An important implication of this lemma is that all "odd" pieces in $L^2(\mathbb{S}^d) = \bigoplus_{i=0}^{\infty} \mathcal{H}_i$ do not contribute to $L^2(\mathbb{S}^d/\Gamma)$. Therefore, when employing spherical harmonics theory to study function spaces of compositional domains, it suffices to consider only even values of i for $\mathcal{H}_i$ within $L^2(\mathbb{S}^d)$. In summary, the function space on the compositional domain $\Delta^d = \mathbb{S}^d/\Gamma$ has the following eigenspace decomposition:

$$L^2(\mathbb{S}^d/\Gamma) = \bigoplus_{i=0}^{\infty} \mathcal{H}_{2i}^{\Gamma}, \tag{15}$$

where $\mathcal{H}_{2i}^{\Gamma} := \{h \in \mathcal{H}_{2i}, \ h = h^{\Gamma}\}$.

3.4 Reproducing Kernels on Compositional Domain

The main objective of this section is to establish reproducing kernels for compositional data. Within each Laplacian eigenspace $\mathcal{H}_i \subset L^2(\mathbb{S}^d)$, the Γ-invariant subspace $\mathcal{H}_i^\Gamma$, as defined in equation (15), can be regarded as a function space on $\Delta^d = \mathbb{S}^d/\Gamma$. To determine a potential candidate for a reproducing kernel within $\mathcal{H}_i^\Gamma$, we first identify the representing function for the linear functional L_z^Γ on $\mathcal{H}_i$ defined as follows: For any function $Y \in \mathcal{H}_i$,

$$L_z^\Gamma(Y) = Y^\Gamma(z) = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} Y(\gamma z),$$

where $z \in \mathbb{S}^d$ is a given point. It can be seen that L_z^Γ and L_z coincide on the subspace $\mathcal{H}_i^\Gamma$ within $\mathcal{H}_i$. Additionally, L_z^Γ can be viewed as a composite map $L_z^\Gamma = L_z \circ \pi_* : \mathcal{H}_i \to \mathcal{H}_i^\Gamma \to \mathbb{C}$. Note that although L_z^Γ is defined on $\mathcal{H}_i$, it can be regarded as a "Delta functional" on $\mathbb{S}^d/\Gamma = \Delta^d$.

In order to find the representing function for L_z^Γ, we make use of the reproducing kernels k_i in Section 3.1. We define the *compositional* kernel $k_i^\Gamma(\cdot,\cdot)$ be Γ-invariant version of k_i, as follows:

$$k_i^\Gamma(x,y) = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} k_i(\gamma x, y), \quad \forall x, y \in \mathbb{S}^d, \tag{16}$$

It can be easily verified that $k_i^\Gamma(z,\cdot)$ represents linear functionals of the form L_z^Γ, simply by following the definitions.

Remark 14. The definition of compositional kernels in (16) is not merely a technique to remove redundant points on spheres. Instead, it is motivated by the concept of *orbital integrals* in analysis and geometry. In our case, the "integral" takes on a discrete form since the compact subgroup is replaced by a finite discrete reflection group Γ. It is worth noting that this type of discrete orbital integral construction is not unique to our work. For instance, in statistical learning theory, Reisert and Burkhardt (2007) utilized a similar construction based on orbital integrals to investigate equivariant matrix-valued kernels.

At first glance, the compositional kernel may seem to lack symmetry since we are "averaging" only over the group orbit on the first variable of the function $k_i(x,y)$. However, due to the symmetric and orthogonally invariant properties of $k_i(x,y)$, as stated in Proposition 9, the compositional kernels surprisingly exhibit symmetry:

Proposition 15. *Compositional kernels are symmetric, namely* $k_i^\Gamma(x,y) = k_i^\Gamma(y,x)$.

Proof Recall that $k_i(x,y) = k_i(y,x)$ and that $k_i(Gx, Gy) = k_i(x,y)$ for any orthogonal matrix G. Note that every reflection $\gamma \in \Gamma$ can be realized as an orthogonal matrix.

Therefore,

$$\begin{aligned}
k_i^\Gamma(x,y) &= \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} k_i(\gamma x, y) \\
&= \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} k_i(y, \gamma x) = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} k_i(\gamma^{-1} y, \gamma^{-1}(\gamma x)) \\
&= \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} k_i(\gamma^{-1} y, x) \\
&= \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} k_i(\gamma y, x) \\
&= k_i^\Gamma(y,x)
\end{aligned}$$

■

Furthermore, it turns out that the compositional kernels are Γ-invariant on both arguments. In fact, they are reproducing kernels with respect to all Γ-invariant functions in $\mathcal{H}_i$. This result is stated in the following theorem:

Theorem 16. *Within $\mathcal{H}_i$, the compositional kernel $k_i^\Gamma(x,y)$ is Γ-invariant on both arguments x and y, and the compositional kernel is a reproducing kernel for $\mathcal{H}_i^\Gamma$.*

Proof The double Γ-invariance can be seen from the definition and Theorem 15. The reproducing property of $k_i^\Gamma(x,y)$ can be shown by observing that for any Γ-invariant function $f \in \mathcal{H}_i^\Gamma \subset \mathcal{H}_i$,

$$\begin{aligned}
< f(t), k_i^\Gamma(x,t) > &= < f(t), \sum_{\gamma\in\Gamma} \frac{1}{|\Gamma|} k_i(\gamma x, t) > \\
&= \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} < f(t), k_i(\gamma x, t) > \\
&= \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} f(\gamma x) = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} f(x) \quad (f \text{ is } \Gamma\text{-invariant}) \\
&= f(x)
\end{aligned}$$

■

Now, let us delve into the construction of an RKHS on Δ^d. Building upon the expression (15), our approach involves approximating the function space $L^2(\mathbb{S}^d/\Gamma)$ with the finite direct sum decomposition $\bigoplus_{i=0}^m \mathcal{H}_{2i}^\Gamma$. Note that this decomposition is orthogonal, and consequently reproducing kernels for $\bigoplus_{i=0}^m \mathcal{H}_{2i}^\Gamma$ can be represented by the sum $\sum_{i=0}^m k_{2i}^\Gamma(\cdot,\cdot)$.

Definition 17. *The degree m compositional reproducing kernel Hilbert space is defined to be the finite direct sum decomposition $\bigoplus_{i=0}^m \mathcal{H}_{2i}^\Gamma$, and its reproducing kernel is*

$$\omega_m(\cdot,\cdot) = \sum_{i=0}^m k_{2i}^\Gamma(\cdot,\cdot). \tag{17}$$

It is crucial to emphasize that our RKHS is *finite-dimensional*, as it comprises solely of degree $2m$ homogeneous polynomials. In other words, each function belonging to $\bigoplus_{i=0}^{m} \mathcal{H}_{2i}^{\Gamma}$ can be represented as a degree $2m$ homogeneous polynomial, including the reproducing kernel $\omega_m(\cdot,\cdot)$. Notably, for any point $(x_1, x_2, \ldots, x_{d+1}) \in \mathbb{S}^d$, the sum $\sum_{i=1}^{d+1} x_i^2$ is always equal to 1, enabling us to convert each element in $\bigoplus_{i=0}^{m} \mathcal{H}_{2i}^{\Gamma}$ into a homogeneous polynomial. To illustrate, let us consider the example of x^2+1. Although it is not initially a homogeneous polynomial, we can rewrite it as $x^2 + x^2 + y^2 + z^2 = 2x^2 + y^2 + z^2$, which is a homogeneous polynomial defined on the sphere $\mathbb{S}^2$. In fact, it is evident that $\bigoplus_{i=0}^{m} \mathcal{H}_{2i}^{\Gamma}$ comprises degree m homogeneous polynomials in terms of *squared* variables.

In what follows we provide details for computing (17). The Gegenbauer polynomial, which is instrumental in calculating the reproducing kernel $k_i(\cdot,\cdot)$ of the ith eigenspace of $L^2(\mathbb{S}^d)$, is defined recursively by

$$\begin{aligned} p_0(t) &= 1, \\ p_1(t) &= (d-1)t, \\ p_i(t) &= \frac{2t(i+(d-3)/2)}{i} p_{i-1}(t) - \frac{i+d-3}{i} p_{i-2}(t) \text{ for } i \geq 2. \end{aligned}$$

Then the kernel $k_i(\cdot,\cdot)$ is given by

$$k_i(v_1, v_2) = \frac{a_i}{\mathrm{vol}(\mathbb{S}^d)} p_i(\langle v_1 \cdot v_2 \rangle), \text{ for } v_1, v_2 \in \mathbb{S}^d,$$

where a_i is the dimension of Laplacian eigenspace $\mathcal{H}_i$:

$$a_i = \binom{d+i}{i} - \binom{d+i-2}{i-2}$$

for $i \geq 2$, and $a_1 = d+1$ and $a_0 = 1$. Also, $\mathrm{vol}(\mathbb{S}^d)$ is the volume of $\mathbb{S}^d$:

$$\mathrm{vol}(\mathbb{S}^d) = \frac{\pi^{(d+1)/2}}{G(1+(d+1)/2)},$$

where $G(\cdot)$ indicates the gamma function. Then, by plugging in $k_i(v_1, v_2)$ to (17), we obtain

$$\omega_m(v_1, v_2) = \sum_{i=0}^{m} \frac{a_{2i}}{|\Gamma| \mathrm{vol}(\mathbb{S}^d)} \sum_{\gamma \in \Gamma} p_{2i}(\langle \gamma(v_1) \cdot v_2 \rangle).$$

A simple numerical example of the calculation of $\omega_m(\cdot,\cdot)$ is provided in the Appendix A.6.

4. Applications of Compositional Reproducing Kernels

The introduction of compositional reproducing kernels brings forth numerous statistical and machine-learning techniques for compositional data analysis. In this context, we present two initial application scenarios to showcase the impact of RKHS theory on compositional data analysis. The first application pertains to the representer theorem, which draws motivation from the advancements in kernel-based machine learning. A compositional support vector machine (SVM) will be used to demonstrate the practical application of the representer theorem. The second example involves the construction of exponential families on compositional domains. Out exponential distribution models define explicit distributions on the compositional domain with *non-vanishing* densities on the boundary.

4.1 Compositional Representer Theorem

The representer theorem is instrumental in kernel-based learning. In this section, we focus specifically on minimal norm interpolations and least square regularizations, which play a crucial role in many data analytic scenarios, such as structured prediction, multi-task learning, and multi-label classification. An important assumption in the representer theorems is linear independence of the kernel evaluations (Micchelli and Pontil, 2005). As our compositional RKHS consists of finite-dimensional polynomials, the linear independence condition is not automatically guaranteed. However, the following theorem gives a positive answer:

Theorem 18. *Let $\{x_i\}_{i=1}^n$ be distinct data points on a compositional domain Δ^d. Then there exists a positive integer M, such that for any $m > M$, the set of functions $\{\omega_m(x_i, \cdot)\}_{i=1}^n$ is a linearly independent set in $\bigoplus_{i=0}^m \mathcal{H}_{2i}^{\Gamma}$.*

Proof The quotient map $c_\Delta : \mathbb{S}^d \to \Delta^d$ can factor through a projective space, i.e., $c_\Delta : \mathbb{S}^d \to \mathbb{P}^d \to \Delta^d$. The main idea is to prove a stronger statement, for which we show that distinct data points in $\mathbb{P}^d$ give linear independence of *projective kernels* for large enough m, where projective kernels are reproducing kernels in $\mathbb{P}^d$ whose definition is given in A.3. Then, we construct two vector subspaces V_1^m and V_2^m and a linear map g_m from V_1^m to V_2^m. The key trick is that the matrix representing the linear map g_m becomes diagonally dominant when m is large enough, which forces the spanning sets of both V_1^m and V_2^m to be linearly independent. More details of the proof are given in the Appendix A.3. ■

Theorem 18 indicates that when m is sufficiently large, $\omega_m(x_i, \cdot) \neq \omega_m(x_j, \cdot)$ holds for any distinct data points $x_i \neq x_j$. As m increases, not only do the reproducing kernels differentiate between points, but they also assign each data point its unique "dimension." A natural question would be regarding the necessary magnitude of m to guarantee linear independence in practical scenarios. In the most general case, the dimension of $\bigoplus_{i=0}^m \mathcal{H}_{2i}^{\Gamma}$ grows rapidly with m, indicating that a very large m may not be required. However, the situation where a large m becomes necessary can occur when there are many near-duplicates in the data. It is important to note that this issue primarily applies to *projective kernels.* For compositional kernels, which are linear combinations of projective kernels, in order to compromise linear independence, a significant number of data points would need to be not only extremely close to each other but also in close proximity to the boundary.

4.1.1 Minimal Norm Interpolation and Least Squares Regularization

Having established linear independence in Theorem 18, we can easily deduce the corresponding representer theorems for minimal norm interpolations and least square regularizations.

While these theorems do not introduce anything new from the perspective of general RKHS theory, we include them here for the sake of completeness.

The first representer theorem we present addresses the minimal norm interpolation problem. Given a fixed set of distinct points $\{x_i\}_{i=1}^n$ in Δ^d and a set of numbers $\{y_i\}_{i=1}^n$, let I_y^m denote the set of functions that interpolate the given data:

$$I_y^m = \{f \in \bigoplus_{i=0}^m \mathcal{H}_{2i}^\Gamma : \ f(x_i) = y_i\}.$$

Our objective is to find f_0 with the minimum ℓ_2 norm, expressed as:

$$||f_0|| = \inf\{||f||, f \in I_y^m\}.$$

Theorem 19. *For a set of distinct compositional data points* $\{x_i\}_{i=1}^n$*, choose* m *large enough so that the reproducing kernels* $\{\omega_m(x_i, t)\}_{i=1}^n$ *are linearly independent, then the unique solution to the minimal norm interpolation problem* $\min\{||f||, f \in \bigoplus_{i=0}^m \mathcal{H}_{2i}^\Gamma : \ f(x_i) = y_i\}$ *is given by the linear combination of the kernels:*

$$f_0(t) = \sum_{i=1}^n c_i \, \omega_m(x_i, t),$$

where $\{c_i\}_{i=1}^n$ *is the unique solution of the following system of linear equations:*

$$\sum_{j=1}^n \omega_m(x_i, x_j) c_j = y_i, \quad 1 \le i \le n.$$

Proof For any other f in I_y^m, define $g = f - f_0$. By considering the decomposition: $||f||^2 = ||g + f_0||^2 = ||g||^2 + 2 < f_0, g > + ||f_0||^2$, one can argue that the cross term $< f_0, g >= 0$. The detail can be found in the Appendix A.4. We point out that the linear independence of reproducing kernels guarantees the uniqueness and existence of f_0. ■

The second representer theorem is for a more realistic scenario with ℓ_2 regularization, which has the following objective:

$$\min_f \sum_{i=1}^n (f(x_i) - y_i)^2 + \mu||f||^2. \tag{18}$$

The goal is to find a Γ-invariant function $f_\mu \in \bigoplus_{i=0}^m \mathcal{H}_{2i}^\Gamma$ that minimizes (18). The solution to this problem is provided by the following representer theorem:

Theorem 20. *For a set of distinct compositional data points* $\{x_i\}_{i=1}^n$*, choose* m *large enough so that the reproducing kernels* $\{\omega_m(x_i, t)\}_{i=1}^n$ *are linearly independent. Then the solution to (18) is given by*

$$f_\mu(t) = \sum_{i=1}^n c_i \, \omega_m(x_i, t),$$

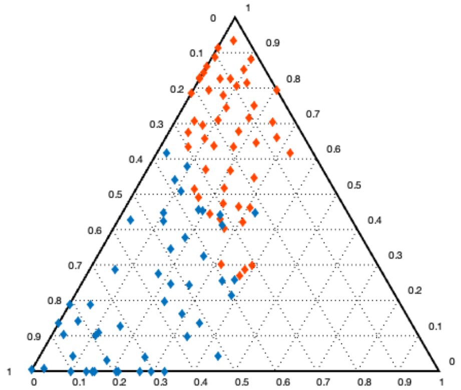


Figure 5: The toy data set used to train kernel SVM.

where $\{c_i\}_{i=1}^n$ *is the solution of the following system of linear equations:*

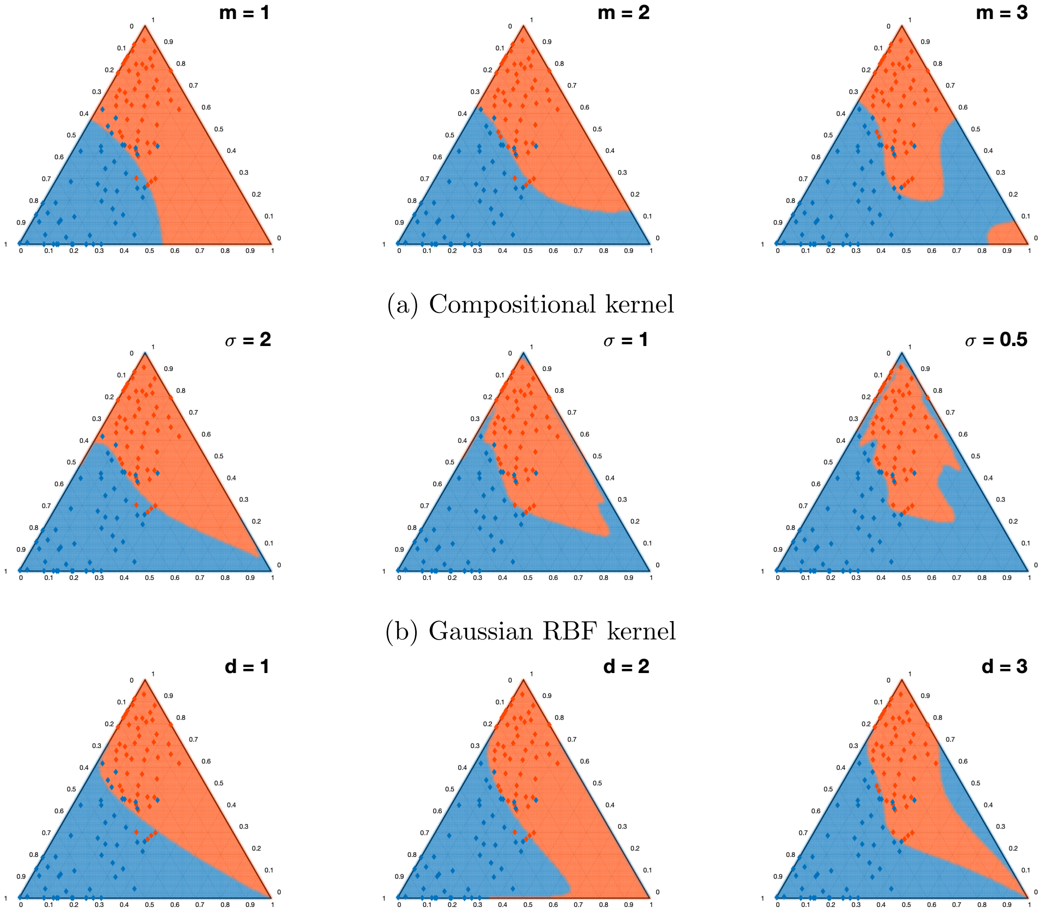
$$\mu c_i + \sum_{j=1}^{n} \omega_m(x_i, x_j) c_j = y_i, \quad 1 \le i \le n.$$

Proof The detail of this proof can be found in the Appendix A.5, but we point out how the linear independence condition plays a role here. In the middle of the proof we must show that $\mu f_\mu(t) = \sum_{i=1}^n \left[(y_i - f_\mu(x_i))\omega_m(x_i, t)\right]$, where $f_\mu(t) = \sum_{i=1}^n \omega_m(x_i, t)c_i$. We use the linear independence in Theorem 18 to establish the equivalence between this linear equation system of $\{c_i\}_{i=1}^n$ and that given in the theorem. ■

Remark 21. It is important to note that without the assumption of linear independence, the *uniqueness* of linear combinations of reproducing kernels may not hold, although a representer theorem can still be obtained. However, by assuming linear independence, we ensure the uniqueness of the solution for $\{c_i\}_{i=1}^n$ through a full rank system of linear equations, as well as canonical transitions between different systems of linear equations (as demonstrated in Theorem 20). This provides both computational convenience and theoretical elegance. Moreover, linear independence is also desired for higher-rank generalizations. In the context of learning vector-valued functions, Micchelli and Pontil (2005) and Muandet et al. (2017) assumed linear independence conditions. While the present work focuses on classical numerical valued functions, our approach suggests a direction for establishing such linear independence conditions even for more complex learning questions involving vector-valued functions.

4.1.2 Compositional Kernel SVM

In this section, we showcase the effectiveness of our compositional kernel in conjunction with a kernel support vector machine (SVM) using a toy dataset depicted in Figure 5. The toy dataset consists of 100 observations from two classes, exhibiting a certain degree of overlap. For comparison purposes, we also implement existing kernel SVM methods for



(a) Compositional kernel

(b) Gaussian RBF kernel

(c) Polynomial kernel

Figure 6: Classification boundaries of kernel SVM with the proposed compositional kernel and two existing kernels applied to alr transformed data.

compositional data, namely the Gaussian RBF kernel $K(x,x') = \exp(-\|x-x'\|^2/\sigma^2)$ and the polynomial kernel $K(x,x') = (1+x\cdot x')^d$ after the additive log ratio (alr) transformation. As similarly done in Section 2.2.1, we replace zeros with a small number before applying the alr transformation.

To implement the proposed compositional kernel, we utilized the `gegenbauerC` function in Matlab to compute $k_i(\cdot,\cdot)$ and the `fitcsvm` function for SVM. While we varied the hyperparameters of each kernel function, we kept the penalty parameter in SVM fixed at its default value. The degree of polynomials controls the complexity of both the compositional kernel and the alr-polynomial kernel, whereas the parameter σ in the alr-Gaussian kernel serves a similar purpose. The classification boundaries on the simplex for each method are depicted in Figure 6. As the hyperparameters change, all three methods exhibit noticeable changes in the classification boundary. Notably, in (b) and (c), the results are disrupted near the simplex edges, leading to unstable classification. In contrast, the compositional kernel does not display any peculiarities around the edges, indicating its robustness in handling such scenarios.

4.2 Compositional Exponential Family

Having established the RKHS, one can use it to define exponential families of probability distributions. Recall that for a function space $\mathcal{H}$ equipped with the inner product $\langle\cdot,\cdot\rangle$ on a general topological space $\mathcal{X}$, where $k(x,\cdot)$ denotes its reproducing kernel, the density of an exponential family $p(x,\theta)$ (Canu and Smola, 2006) with parameter $\theta\in\mathcal{H}$ is given by:

$$p(x,\theta) = \exp\{\langle\theta(\cdot),k(x,\cdot)\rangle - g(\theta)\},$$

where $g(\theta) = \log\int_{\mathcal{X}}\exp\big(\langle\theta(\cdot),k(x,\cdot)\rangle\big)dx$.

For compositional data, we define the density of the mth degree exponential family as:

$$p_m(x,\theta) = \exp\left\{\langle\theta(\cdot),\omega_m(x,\cdot)\rangle - g(\theta)\right\}, \quad x\in\mathbb{S}^d/\Gamma, \tag{19}$$

where $\theta\in\bigoplus_{i=0}^m \mathcal{H}_{2i}^\Gamma$ and

$$g(\theta) = \log\int_{\mathbb{S}^d/\Gamma}\exp\big(\langle\theta(\cdot),\omega_m(x,\cdot)\rangle\big)dx.$$

The density (19) can be expressed in a more explicit form by utilizing homogeneous polynomials. As mentioned earlier, any function within $\bigoplus_{i=0}^m \mathcal{H}_{2i}^\Gamma$ can be represented as a degree m homogeneous polynomial with squared variables, as shown in Lemma 13. Thus, the density in (19) can be written as follows for $x = (x_1,\ldots,x_{d+1})\in\mathbb{S}^d\geq 0$:

$$p_m(x,\theta) = \exp\{s_m(x_1^2,x_2^2,\ldots,x_{d+1}^2;\theta) - g(\theta)\}, \tag{20}$$

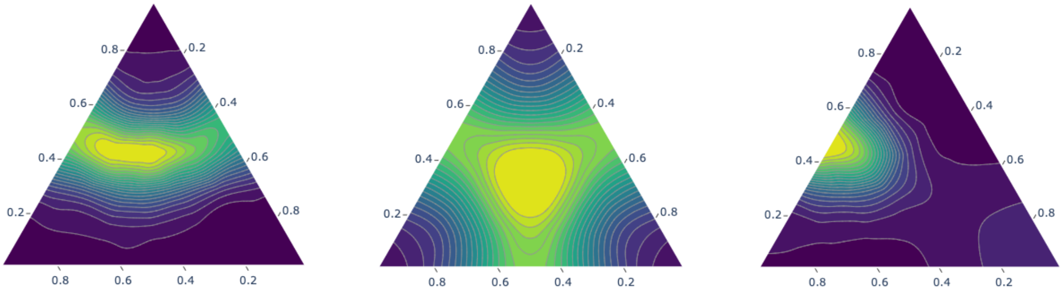
where s_m represents a polynomial on the squared variables x_i^2, with θ represented as the coefficients. The normalizing constant $g(\theta)$ can be computed by integrating over the entire sphere as follows:

$$g(\theta) = \int_{\mathbb{S}^d/\Gamma}\exp(s_m)dx = \frac{1}{|\Gamma|}\int_{\mathbb{S}^d}\exp(s_m)dx.$$

Figure 7 displays three examples of a compositional exponential distribution. The three densities respectively have the following functional s_m:

$$\begin{aligned} s_4(x,\theta_1) &= -2x_1^4 - 2x_2^4 - 3x_3^4 + 9x_1^2x_2^2 + 9x_1^2x_3^2 - 2x_2^2x_3^2, \\ s_4(x,\theta_2) &= -x_1^4 - x_2^4 - x_3^4 - x_1^2x_2^2 - x_1^2x_3^2 - x_2^2x_3^2, \\ s_4(x,\theta_3) &= -3x_1^4 - 2x_2^4 - x_3^4 + 9x_1^2x_2^2 - 5x_1^2x_3^2 - 5x_2^2x_3^2. \end{aligned}$$

The diversity of density contours shown in Figure 7 indicates that the proposed compositional exponential family is capable of modeling data with a broad spectrum of locations and correlation structures. Further exploration is warranted, especially concerning the estimation of parameters. Maximum likelihood estimation, coupled with regression-based approaches such as the one discussed by Beran (1979), serves as a natural starting point for tackling this task.



(a) $p_4(x,\theta_1)$

(b) $p_4(x,\theta_2)$

(c) $p_4(x,\theta_3)$

Figure 7: Three exemplary densities from the compositional exponential family. See text for specification of the parameters θ_1, θ_2, and θ_3.

5. Discussion

This research utilizes projective and spherical geometries to enhance the geometric understanding of compositional data. We propose a new nonparametric density estimation method and construct a reproducing kernel Hilbert space existing on the reinterpreted compositional domain. The proposed framework provides a novel approach to explore and analyze compositional data, eliminating the need for ad-hoc treatment of zeros on the boundary.

A practically important issue that needs to be addressed in future work is the computational complexity of the proposed RKHS, which may be significant if d is large. For instance, the kernel SVM experiment discussed in Section 4.1.2 took between 52.15 to 92.41 seconds to compute on an M1 MacBook Pro with 16 GB of memory, while other competing methods took much less computation times. To mitigate this computational burden, we plan to explore two potential approaches. The first approach involves subsampling the expanded data across the entire sphere, leveraging the symmetrical nature of Γ. By doing so, we can effectively reduce the number of terms involved in kernel evaluations, thereby enhancing computational efficiency. The second approach entails reducing the dimensionality of the data using techniques such as compositional principal component analysis.

As the editor has pointed out, a diffusion kernel proposed by Lafferty and Lebanon (2005), specifically the Fisher information kernel for multinomial distribution, which we call "Fisher kernel", could also be used for compositional data. The Fisher kernel is a solution to the heat equation that involves the Laplacian thus the application of this kernel could capture similar features of the data. However, there seem to be three important issues to address when one considers the Fisher kernel for compositional data. Firstly, the square-root transformation of the normalized multinomial counts may be necessary when the data are mapped into the sphere in order to preserve the Fisher information. This might yield a substantial difference in the results. Secondly, the Fisher kernel approach cannot naturally deal with the zeros in the simplex boundary, which was why a "rounding" procedure for the simplex boundary was suggested by Lafferty and Lebanon (2005) as a remedy. The third question regards the reproducing property and the characterization of the RKHS associated with the Fisher kernel. These considerations lead to an interesting direction of research on the Γ-*invariant Fisher kernel*.

It is also worth mentioning that even though not as intuitive as our approach, a similar result could be derived by harmonic analysis on special orthogonal group $SO(d+1)$, of which $\mathbb{S}^d$ is a homogeneous space, as the editor has pointed out. In fact, in directional statistics, using tensors to represent data points has been considered by Arnold et al. (2018) who explored statistical analysis of directional data in the coset space $SO(3)/K$, where K denotes a finite subgroup of $SO(3)$. Note that the compositional domain is given by $\mathbb{S}^d/\Gamma = O(d+1) \setminus O(d)/\Gamma$ which is a double coset space, and the reflection group Γ is not a subgroup of the *special* orthogonal group $SO(d+1)$.

Acknowledgments

Binglin Li wants to thank Thomas Nordahl (Professor Emeritus at UC Davis) for his complete financial support for a fall semester during his graduate study at the University of Georgia. This work was partially supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (2021R1A2C1093526, RS-2023-00218231, 2022M3J6A1063021).

Appendix A. Proofs and Example

A.1 Proof of Theorem 7

Assumption 1. *For all kernel density estimators and bandwidth parameters in this paper, we assume the following:*

- **H1** *The kernel function* $K : [0,\infty) \to [0,\infty)$ *is continuous such that both* $\lambda_d(K)$ *and* $\lambda_d(K^2)$ *are bounded for* $d \geq 1$*, where* $\lambda_d(K) = 2^{d/2-1}\mathrm{vol}(S^d)\int_0^\infty K(r)r^{d/2-1}dr$.
- **H2** *If a function* f *on* $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ *is extended to the entire* $\mathbb{R}^{d+1}\setminus\{0\}$ *via* $f(x) = f(x/\|x\|)$, *then the extended function* f *needs to have its first three derivatives bounded.*
- **H3** *The bandwidth parameter* $h_n \to 0$ *as* $nh_n^d \to \infty$.

Let f be the extended function from $\mathbb{S}^d$ to $\mathbb{R}^{d+1} \setminus \{0\}$ via $f(x/\|x\|)$, and let

$$\phi(f,x) = -x^T \nabla f(x) + d^{-1}(\nabla^2 f(x) - x^T(\mathcal{H}_x f)x) = d^{-1}\mathrm{tr}[\mathcal{H}_x f(x)],$$

where $\mathcal{H}_x f$ is the Hessian matrix of f at the point x.

Define the following:

$$b_d(K) = \frac{\int_0^\infty K(r)r^{d/2}dr}{\int_0^\infty K(r)r^{d/2-1}dr}$$

and

$$\phi(h_n) = \frac{4b_d(K)^2}{d^2}\sigma_x^2 h_n^4.$$

Proof The strategy in Zhao and Wu (2001) in the directional set-up follows that in Hall (1984), whose key idea is to give asymptotic bounds for degenerate U-statistics so that one can use Martingale theory to derive the central limit theorem. The step where the finite support condition was used in Zhao and Wu (2001), is when they were trying to prove the asymptotic bound:$E(G_n^2(X_1, X_2)) = O(h^{7d})$, where $G_n(x,y) = E[H_n(X, xH_n(X,y))]$ with $H_n = \int_{\mathbb{S}^d} K_n(z,x)K_n(z,y)dz$ and the centered kernel $K_n(x,y) = K[(1-x^Ty)/h^2] - E\{[K(1-x^TX)/h^2]\}$. During that proof, they were trying to show the following term

$$T_1 = \int_{\mathbb{S}^d} f(x)dx \int_{\mathbb{S}^d} f(y)dy \times \bigg\{ \int_{\mathbb{S}^d} f(z)dz \int_{\mathbb{S}^d} K[(1-u^Tx)/h^2]K[(1-u^Tz)/h^2]du \cdot \int_{\mathbb{S}^d} K[(1-u^Ty)/h^2]K[(1-u^Tz)/h^2]du\bigg\}^2$$

is $O(h^{7d})$, for which the finite support condition was substantially used.

The idea to avoid this assumption was based on the observation in García-Portugués et al. (2015) where they only concern the case of directional-linear CLT, whose result can not be directly used to the only directional case. Based on the method provided in Lemma 10 in García-Portugués et al. (2015), one can easily deduce the following asymptotic equivalence:

$$\int_{\mathbb{S}^d} K^j\Big(\frac{1-x^Ty}{h^2}\Big)\phi^i(y)dy \sim h^d\lambda_d(K^j)\phi^i(x),$$

where $\lambda_d(K^j) = 2^{d/2-1}\mathrm{vol}(\mathbb{S}^{d-1})\int_0^\infty K^j(r)r^{d/2-1}dr$. As a special case we have:

$$\int_{\mathbb{S}^d} K^2\Big(\frac{1-x^Ty}{h^2}\Big)dy \sim h^d\lambda_d(K^2)C, \text{ with } C \text{ being a positive constant.}$$

Now we will proceed with the proof without the finite support condition:

$$\begin{aligned}
T_1 &= \int_{\mathbb{S}^d} f(x)dx \int_{\mathbb{S}^d} f(y)dy \ \times \Big\{ \int_{\mathbb{S}^d} f(z)dz \int_{\mathbb{S}^d} K[(1-u^T x)/h^2]K[(1-u^T z)/h^2]du \\
&\qquad\qquad \cdot \int_{\mathbb{S}^d} K[(1-u^T y)/h^2]K[(1-u^T z)/h^2]du \Big\}^2 \\
&\sim \int_{\mathbb{S}^d} f(x)dx \int_{\mathbb{S}^d} f(y)dy \Big\{ \int_{\mathbb{S}^d} f(z)\Big[\lambda_d(K)h^d K\Big(\frac{1-x^T z}{h^2}\Big)\Big] \times \Big[\lambda_d(K)h^d K\Big(\frac{1-y^T z}{h^2}\Big)\Big]dz \Big\}^2 \\
&\sim \lambda_d(K)^4 h^{4d} \int_{\mathbb{S}^d} f(x)dx \int_{\mathbb{S}^d} f(y)\Big[\lambda_d(K)h^d K\Big(\frac{1-x^T y}{h^2}\Big)f(y)\Big]^2 dy \\
&= \lambda_d(K)^6 h^{6d} \int_{\mathbb{S}^d} \Big\{ \int_{\mathbb{S}^d} K^2\Big(\frac{1-x^T y}{h^2}\Big) f^3(y)dy \Big\} f(x)dx \\
&\sim \lambda_d(K)^6 h^{6d} \int_{\mathbb{S}^d} \lambda_d(K^2)h^d C \cdot f^3(x)f(x)dx \\
&= C\lambda_d(K)^6 \lambda_d(K^2) h^{7d} \int_{\mathbb{S}^d} f(x)dx = O(h^{7d}).
\end{aligned}$$

Thus we have proved $T_1 = O(h^{7d})$ without finite support assumption, then the rest of the proof will follow through as in Zhao and Wu (2001). Now combining this with the following equality

$$\int_{\mathbb{S}^d_{\geq 0}} (\hat{p}_n - p)^2 dx = |\Gamma| \int_{\mathbb{S}^d} (\hat{f}_n - \tilde{p})^2 dy$$

proves the CLT of compositional ISE.

■

A.2 Proof of Lemma 13

Proof A direct computation yields:

$$\begin{aligned}
(\textstyle\prod_{i=1}^{d+1} x_i^{\alpha_i})^\Gamma &= \frac{1}{|\Gamma|} \sum_{s_i \in \{\pm 1\}} \prod_{i=1}^{d+1} (s_i x_i)^{\alpha_i} \\
&= \frac{1}{|\Gamma|} \sum_{s_i \in \{\pm 1\}} \prod_{i \neq k} (s_i x_i)^{\alpha_i} x_k^{\alpha_k} + \sum_{s_i \in \{\pm 1\}} \prod_{i \neq k} (s_i x_i)^{\alpha_i} (-x_k)^{\alpha_k} \\
&= x_k^{\alpha_k} \frac{1}{|\Gamma|} \sum_{s_i \in \{\pm 1\}} \prod_{i \neq k} (s_i x_i)^{\alpha_i} - x_k^{\alpha_k} \frac{1}{|\Gamma|} \sum_{s_i \in \{\pm 1\}} \prod_{i \neq k} (s_i x_i)^{\alpha_i} \\
&= 0.
\end{aligned}$$

■

A.3 Proof of Theorem 18

We sketch a slightly more detailed (not complete) proof:

Proof This is the most technical lemma in this article. We will sketch the philosophy of the proof in here, which can be intuitively understood topologically.

Recall that we can produce a projective space $\mathbb{P}^d$ by identifying every pair of antipodal points of a sphere $\mathbb{S}^d$ (identify x with $-x$), in other words $\mathbb{P}^d = \mathbb{S}^d/\mathbb{Z}_2$ where $\mathbb{Z}_2 = \{0, 1\}$ is a cyclic group of order 2. Then we can define a projective kernel in $\mathcal{H}_i \subset L^2(\mathbb{S}^d)$ to be $k_i^p(x, \cdot) = [k_i(x, \cdot) + k_i(-x, \cdot)]/2$. We can also denote the projective kernel inside $\bigoplus_{i=0}^m \mathcal{H}_{2i}$ by $\underline{k}_m^p(x, \cdot) = \sum_{i=0}^m k_{2i}^p(x, \cdot)$.

Now we spread out the data set $\{x_i\}_{i=1}^n$ by "spread-out" construction in Section 2.1, and denote the spread-out data set as $\{\Gamma \cdot x_i\}_{i=1}^n = \{c_\Delta^{-1}(x_i)\}_{i=1}^n$ (a data set, not a set because of repetitions). A compositional reproducing kernel kernel is a summation of spherical reproducing kernels of on $c_\Delta^{-1}(x_i)$, divided by the number of elements in $c_\Delta^{-1}(x_i)$. This data set $c_\Delta^{-1}(x_i)$ has antipodal symmetry, then a compositional kernel is a linear combination of projective kernels. Notice that *different* compositional kernels are linear combinations of *different* projective kernels. It suffices to show the linear independence of projective kernels for distinct data points and large enough m, which implies the linear independence of compositional kernels $\{\omega_m(x_i, \cdot)\}_{i=1}^n$.

Now we are focusing on the linear independence of projective kernels. A projective kernel can be seen as a reproducing kernel for a point in $\mathbb{P}^d$. For a set of distinct points $\{y_i\}_{i=1}^l \subset \mathbb{P}^d$, we will show that the corresponding set of projective kernels $\{\underline{k}_m^p(y_i, \cdot)\}_{i=1}^l \subset \bigoplus_{i=0}^m \mathcal{H}_{2i}$ is linearly independent for an integer l and a large enough m.

Consider two vector subspace $V_1^m = \mathrm{span}\big[\{(y_i \cdot t)^{2m}\}_{i=1}^l\big]$ and $V_2^m = \mathrm{span}\big[\{\underline{k}_m^p(y_i, t)\}_{i=1}^l\big]$, both of which are inside $\bigoplus_{i=0}^m \mathcal{H}_{2i} \subset L^2(\mathbb{S}^d)$ (here we are implicitly using Proposition 8). Then we can define a linear map $h_m : V_1^m \to V_2^m$ by setting $h_m((y_i \cdot t)^{2m}) = \sum_{j=1}^l \langle (y_i \cdot t)^{2m}, \underline{k}_m^p(y_j, t)\rangle \underline{k}_m^p(y_j, t)$. The matrix representation of h_m with respect to these spanning sets is an $l \times l$ symmetric matrix whose diagonal elements are 1's, and off diagonal elements are $[(y_i \cdot y_j)]^{2m}$. Notice that $y_i \neq y_j$ in $\mathbb{P}^d$, which means that they are not antipodal to each other in $\mathbb{S}^d$, thus $|y_i \cdot y_j| < 1$. When m is large enough, all off-diagonal elements will go to zero while diagonal elements always stay constant, then the matrix representing h_m will become a *diagonally dominant* matrix, which is full rank. When the linear map h_m has full rank, both spanning sets $\{(y_i \cdot t)^{2m}\}_{i=1}^l$ and $\{\underline{k}_m^p(y_i, t)\}_{i=1}^l$ have to be a basis for V_1^m and V_2^m correspondingly, then the set of projective kernels $\{\underline{k}_m^p(y_i, t)\}_{i=1}^m$ have to be linearly independent when m is large enough.

■

A.4 Proof of Theorem 19

Proof Note that the set $I_y^m = \{f \in \bigoplus_{i=0}^m \mathcal{H}_{2i} : f(x_i) = y_i\}$ is non-empty, because the f_0 defined by the linear system of equation is naturally in I_y^m. Let f be any other element in I_y^m, define $g = f - f_0$, then we have:

$$\|f\|^2 = \|g + f_0\|^2 = \|g\|^2 + 2\langle f_0, g\rangle + \|f_0\|^2.$$

Notice that $g \in \bigoplus_{i=0}^{m} \mathcal{H}_{2i}$ and that $g(x_i) = 0$ for $1 \le i \le n$, we have:

$$\begin{aligned}\langle f_0, g\rangle &= \langle \textstyle\sum_{i=1}^{n} \omega_m(x_i,\cdot)c_i, g(\cdot)\rangle \\ &= \sum_{i=1}^{n} c_i \langle \omega_m(x_i,\cdot), g(\cdot)\rangle. \\ &= \sum_{i=1}^{n} c_i g(x_i) = 0.\end{aligned}$$

Thus $\|f\|^2 = \|g + f_0\|^2 = \|g\|^2 + \|f_0\|^2$, which implies that f_0 is the solution to the minimal norm interpolation problem. ■

A.5 Proof of Theorem 20

Proof First define the loss functional $E(f) = \sum_{i=1}^{n} |f(x_i) - y_i|^2 + \mu\|f\|^2$. For any Γ-invariant function $f = f^{\Gamma} \in \bigoplus_{i=0}^{m} \mathcal{H}_{2i}$, let $g = f - f_\mu$, then a simple computation yields:

$$E(f) = E(f_\mu) + \sum_{i=1}^{n} |g(x_i)|^2 - 2\sum_{i=1}^{n}(y_i - f_\mu(x_i))g(x_i) + 2\mu\langle f_\mu, g\rangle + \mu\|g\|^2.$$

I want to show $\sum_{i=1}^{n}(y_i - f_\mu(x_i))g(x_i) = \mu\langle f_\mu, g\rangle$, and an equivalent way of writing this equality is:

$$\sum_{i=1}^{n}\langle (y_i - f_\mu(x_i))\omega_m(x_i,t), g(t)\rangle = \mu\langle f_\mu, g\rangle.$$

Now I claim that $\mu f_\mu(t) = \sum_{i=1}^{n}\left[(y_i - f_\mu(x_i)) \cdot \omega_m(x_i,t)\right]$, which implies the above equality. To prove this claim, plug this linear combination $f_\mu = \sum_{i=1}^{n} c_i \cdot \omega_m(x_i,t)$ into the claim, then we get a system of linear equations in $\{c_i\}_{i=1}^{n}$, thus the proof of the claim breaks down to checking the system of linear equations in $\{c_i\}_{i=1}^{n}$, produced by the claim.

Note that $\{\omega_m(x_i,t)\}_{i=1}^{n}$ is a linearly independent set, so one can check that the system of linear equations in $\{c_i\}_{i=1}^{n}$ produced by the claim is true, if and only if $\{c_i\}_{i=1}^{n}$ satisfy $\mu c_k + \sum_{i=1}^{n} c_i \cdot \omega_m(x_i,x_k) = y_k$ for every k with $1 \le k \le n$, which is given by the condition of this theorem. The equivalence of these two systems of linear equations is given by the linear independence of the set $\{\omega_m(x_i,t)\}_{i=1}^{n}$. Therefore we conclude that the claim $\mu f_\mu(t) = \sum_{i=1}^{n}\left[(y_i - f_\mu(x_i)) \cdot \omega_m(x_i,t)\right]$ is true.

To finish the proof of this theorem, notice that

$$\begin{aligned}
E(f) &= E(f_\mu) + \textstyle\sum_{i=1}^n |g(x_i)|^2 - 2\sum_{i=1}^n (y_i - f_\mu(x_i))g(x_i) + 2\mu < f_\mu, g > +\mu||g||^2 \\
&= E(f_\mu) + \sum_{i=1}^n |g(x_i)|^2 + \mu||g||^2 + 2\big[\mu\langle f_\mu, g\rangle - \sum_{i=1}^n (y_i - f_\mu(x_i))g(x_i)\big] \\
&= E(f_\mu) + \sum_{i=1}^n |g(x_i)|^2 + \mu||g||^2 \\
&\quad +2\big[\mu\langle f_\mu, g\rangle - \textstyle\sum_{i=1}^n \langle (y_i - f_\mu(x_i)) \cdot \omega_m(x_i, t), g(t)\rangle\big] \\
&= E(f_\mu) + \sum_{i=1}^n |g(x_i)|^2 + \mu||g||^2 \\
&\quad +2\big[\langle \underbrace{\big(\mu f_\mu(t) - \textstyle\sum_{i=1}^n \big[(y_i - f_\mu(x_i)) \cdot \omega_m(x_i, t)\big]\big)}_{=0}, g(t)\rangle\big] \\
&= E(f_\mu) + \sum_{i=1}^n |g(x_i)|^2 + \mu||g||^2.
\end{aligned}$$

The term $\sum_{i=1}^n |g(x_i)|^2 + \mu||g||^2$ in the above equality is always non-negative, thus $E(f_\mu) \leq E(f)$, then the theorem follows. ■

A.6 A simple example of the compositional kernel evaluation

Suppose we have two composition points in Δ^2: $x = (0.2, 0.5, 0.3)'$ and $y = (0.1, 0.4, 0.5)'$. We first map them to $\mathbb{S}^2_{\geq 0}$: $x_s = (0.324, 0.811, 0.486)'$ and $y_s = (0.154, 0.617, 0.771)'$. We compute the degree m compositional kernel given x_s and y_s as

$$\omega_m(x_s, y_s) = \sum_{i=0}^m \frac{a_{2i}}{|\Gamma|\text{vol}(\mathbb{S}^2)} \sum_{\gamma \in \Gamma} p_{2i}(\langle \gamma(x_s) \cdot y_s \rangle),$$

where $\gamma(x_s)$ includes all eight sign-flipped versions of x_s. For instance with $m = 2$, $\omega_2(x_s, y_s)$ is

$$\frac{1}{|\Gamma|\text{vol}(\mathbb{S}^2)}\Big(a_0 \sum_{\gamma\in\Gamma} p_0(\langle \gamma(x_s) \cdot y_s\rangle) + a_2 \sum_{\gamma\in\Gamma} p_2(\langle \gamma(x_s) \cdot y_s\rangle) + a_4 \sum_{\gamma\in\Gamma} p_4(\langle \gamma(x_s) \cdot y_s\rangle)\Big),$$

which equals 0.803.

References

John Aitchison. Principles of compositional data analysis. *Lecture Notes-Monograph Series*, pages 73–81, 1994.

John Aitchison and Ian J. Lauder. Kernel density estimation for compositional data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 34:129–137, 1985.

Richard Arnold, Peter E. Jupp, and Helmut Schaeben. Statistics of ambiguous rotations. *Journal of Multivariate Analysis*, 165:73–85, 2018.

Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

Kendall Atkinson and Weimin Han. *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*. Springer, 2012.

Zhidong Bai, C Radhakrishna Rao, and Lei Zhao. Kernel estimators of density function of directional data. *Journal of Multivariate Analysis*, 27(1):24–39, 1989.

Alan Beardon. *The Geometry of Discrete Groups*, volume 91. Springer Science & Business Media, 2012.

Rudolf Beran. Exponential models for directional data. *Annals of Statistics*, 7(6):1162–1178, 1979.

M Luz Calle. Statistical analysis of metagenomics data. *Genomics & Informatics*, 17(1), 2019.

Stephane Canu and Alex Smola. Kernel methods and the exponential family. *Neurocomputing*, 69:714–720, 2006.

Peter Filzmoser, Kare Hron, Clemens Reimann, and Robert Garrett. Robust factor analysis for compositional data. *Computers & Geosciences*, 35:1854–1861, 2009.

Eduardo García-Portugués, Rosa M Crujeiras, and Wenceslao González-Manteiga. Central limit theorems for directional and linear random variables with applications. *Statistica Sinica*, 25:1207–1229, 2015.

Peter Hall. Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis*, 14:1–16, 1984.

Peter Hall, G. S. Watson, and Javier Cabrera. Kernel density estimation with spherical data. *Biometrika*, 74:751–762, 1987.

Peter E. Jupp. Data-driven Sobolev tests of uniformity on compact Riemannian manifolds. *Annals of Statistics*, 36(3):1246–1260, 2008.

Peter E. Jupp and B. D. Spurr. Sobolev tests for independence of directions. *Annals of Statistics*, 13(3):1140–1155, 1985.

John Lafferty and Guy Lebanon. Diffusion kernels on statistical manifolds. *Journal of Machine Learning Research*, 6:129–163, 2005.

Sugnet Lubbe, Peter Filzmoser, and Matthias Templ. Comparison of zero replacement strategies for compositional data with large numbers of zeros. *Chemometrics and Intelligent Laboratory Systems*, 210:104248, 2021.

Charles A. Micchelli and Massimiliano Pontil. On learning vector-valued functions. *Neural Computation*, 17:177–204, 2005.

Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, 10, 2017.

Phillip J Paine, S.P. Preston, M Tsagris, and Andrew Wood. Spherical regression models with general covariates and anisotropic errors. *Statistics and Computing*, 30:153–165, 2020.

Junyoung Park, Changwon Yoon, Cheolwoo Park, and Jeongyoun Ahn. Kernel methods for radial transformed compositional data with many zeros. In *International Conference on Machine Learning*, pages 17458–17472. PMLR, 2022.

Marco Reisert and Hans Burkhardt. Learning equivariant functions with matrix valued kernels. *Journal of Machine Learning Research*, 8(15):385–408, 2007.

J. L. Scealy and A. H. Welsh. Fitting kent models to compositional data with small concentration. *Statistics and Computing*, 24:153–165, 2014.

Elias M. Stein and Guido Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton Mathematical Series. Princeton University Press, 1971.

Antoni Susin, Yiwen Wang, Kim-Anh Lê Cao, and M. Luz Caller. Variable selection in microbiome compositional data analysis. *NAR Genomics and Bioinformatics*, 2, 2020.

Diego Tomassi, Liliana Forzani abd Sabrina Duarte, and Ruth M Pfeiffer. Sufficient dimension reduction for compositional data. *Biostatistics*, 22:687–705, 2019.

Grace Wahba. Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing*, 2(1):5–16, 1981.

Lincheng Zhao and Chengqing Wu. Central limit theorem for integrated square error of kernel estimators of spherical density. *Science in China Series A: Mathematics*, 44(4): 474–483, 2001.