The Isometric Log-Ratio Transform for Probabilistic Multi-Label Anatomical Shape Representation

Shawn Andrews, Neda Changizi, and Ghassan Hamarneh

Abstract—Sources of uncertainty in the boundaries of structures in medical images have motivated the use of probabilistic labels in segmentation applications. An important component in many medical image segmentation tasks is the use of a shape model, often generated by applying statistical techniques to training data. Standard statistical techniques (e.g. principal component analysis) often assume data lies in an unconstrained vector space, but probabilistic labels are constrained to the unit simplex. If these statistical techniques are used directly on probabilistic labels, relative uncertainty information can be sacrificed. A standard method for facilitating analysis of probabilistic labels is to map them to a vector space using the LogOdds transform. However, the LogOdds transform is asymmetric in one of the labels, which skews results in some applications. The isometric log-ratio (ILR) transform is a symmetrized version of the LogOdds transform, and is so named as it is an isometry between the Aitchison geometry, the inherent geometry of the simplex, and standard Euclidean geometry. We explore how to interpret the Aitchison geometry when applied to probabilistic labels in medical image segmentation applications. We demonstrate the differences when applying the LogOdds transform or the ILR transform to probabilistic labels prior to statistical analysis. Specifically, we show that statistical analysis of ILR transformed data better captures the variability of anatomical shapes in cases where multiple different foreground regions share boundaries (as opposed to foreground-background boundaries).

Index Terms—Uncertainty, Probabilistic Labels, Statistical Shape Analysis, Aitchison Geometry, Probabilistic Segmentation, Bayesian Inference, LogOdds, ILR

I. INTRODUCTION

In medical image segmentation, uncertainty in shape boundaries arises from numerous sources, including tissue heterogeneity [1], image acquisition artifacts, partial volume effects (PVE), segmentation by multiple-raters [2], and image segmentation algorithms intentionally designed to output results encoding uncertainty [3], [4]. These uncertainties should not be ignored in subsequent analyses [1], [2], [5], [6], and have been utilized in a variety of clinical applications [7]–[10]. As medical image analysis is often concerned with the segmentation of multiple structures, quantifying the relative uncertainty in the boundaries of neighboring structures is important.

In segmentation, every pixel (which we will use interchangeably with voxel) in an image is usually assigned a single label, but uncertainty can be encoded using a probabilistic segmentation, which assigns a probabilistic label to each pixel [1]–[4], [11]–[14]. A probabilistic label is a vector of non-negative fractions such that these fractions sum to one and represent our relative confidence in each label’s correctness. The space of such vectors of fractions is the unit simplex and can be interpreted as discrete distributions over the labels.

In medical images, different anatomical structures may be difficult to distinguish based on image features alone (e.g. due to similar intensity profiles, PVE obscuring boundaries). However, the relative locations and shapes of the anatomical structures are often consistent across different images, so prior shape knowledge can greatly assist in accurate segmentation. Prior shape knowledge is frequently obtained by applying statistical analysis techniques such as principal component analysis (PCA) to training segmentations (e.g. manually segmented images). The details of how PCA is applied depend on what representation is used for the training segmentations.

Cremers et al. perform PCA using a binary probabilistic label as their shape representation [15]. However, the space of probabilistic labels is constrained, and thus movement along the PCA modes can result in probabilistic labels outside of the simplex (e.g. negative fractions), which require projections back to the simplex. Projections may discard relative uncertainty information, particularly if a similar approach was applied to multi-label probabilistic segmentations, for example if many probabilities are projected back to 0.

A signed distance map (SDM) represents a shape using a real-valued function over the image domain that indicates the distance from the shape’s boundary to each pixel, with positive or negative signs indicating whether a pixel is interior or exterior to the shape, respectively. Multiple shapes can be represented by a vector of SDM values at each pixel. Many existing works perform statistical analysis on SDMs to create shape distributions [16]–[20]. One disadvantage of using SDMs for shape analysis is that they do not form a vector space (or even a convex space), which many common statistical techniques (e.g. PCA) assume. This is sometimes dealt with by repeated projections onto the manifold of valid SDMs [21].

Pohl et al. provide a connection between SDMs and probabilistic labels using the logarithm of odds (LogOdds) transform [22]. The LogOdds transform maps probabilistic labels to an unconstrained Euclidean vector space, and its inverse maps a vector of real values, e.g. the SDM values at a pixel, to a probabilistic label. When utilizing analysis techniques designed for unconstrained vector spaces (e.g. PCA) for shape analysis, the LogOdds representation requires no projections and ensures relative uncertainty values are preserved. Addition and
The isometric log-ratio (ILR) transform, proposed by Egozcue et al., is another log-ratio function that is symmetric between its components, and we explore its usefulness to image analysis. The ILR transform originates from the field of compositional data analysis. Compositional data is data that describes parts of a whole (e.g., the composition of a mineral), and so the absolute values of the components of a vector of compositional data are unimportant; only their relative values have meaning. Probabilistic labels, being constrained to the unit simplex, are a type of compositional data; the probability of an individual label is meaningful given the implicit understanding that the other label probabilities must sum to one. The work of Aitchison is central to compositional data analysis, and resulted in a Hilbert space structure for the simplex, based on the Aitchison distance and inner product \([23]–[25]\). The ILR transform is so named because it provides an isometry between the Aitchison distance and the Euclidean distance, preserving the Hilbert space structure of the simplex.

In this work, we extend our previous works that use the ILR transform to build probabilistic shape models \([26]–[28]\). Our main contribution is to demonstrate the impact that the choice of log-ratio transform (LogOdds or ILR) has on anatomical shape models generated using PCA. We show that the choice of log-ratio transform can indeed have a significant effect on the results of a task that utilizes these shape models (e.g., image segmentation). Specifically, PCA shape models built using the LogOdds transform focus on capturing variability in foreground-background boundaries, whereas shape models built using the ILR transform focus on capturing variability in all boundaries (foreground-foreground and foreground-background) equally. In images containing multiple relevant anatomical structures that share boundaries, we show the ILR transform is the more appropriate choice.

The remainder of the paper is laid out as follows: in Sec. II, we review and compare the LogOdds and ILR transforms and discuss the Aitchison geometry for the simplex; in Sec. III, we provide intuitive interpretations for the Aitchison distance and inner product in the context of probabilistic image segmentation; and in Sec. IV, we demonstrate the differences between performing PCA on LogOdds and ILR transformed probabilistic labels for synthetic and real medical data.

II. LOG-RATIO TRANSFORMS

A. LogOdds

We let \(L\) represent the number of labels (including background) in a given segmentation task. We define the unit simplex for \(L\) probabilistic labels as

\[
\mathbb{P}_L = \left\{ \mathbf{p} = [p_1, p_2, \ldots, p_L]^T \in \mathbb{R}^L : p_i > 0 ; 1 \leq i \leq L ; \sum_{i=1}^L p_i = 1 \right\}.
\]  

\(\mathbb{P}_L\) is defined such that each component is non-zero. Cromwell’s rule \([29]\) states that no probabilities should be exactly 0 in a Bayesian framework, as this does not allow for a change in one’s belief, regardless of the evidence observed. A probabilistic segmentation is a function \(\mathbf{q} : \Omega \to \mathbb{P}_L\), where \(\Omega\) is the discrete image domain, with \(n = |\Omega|\) pixels.

To facilitate the creation of statistical models over probabilistic segmentations, Pohl et al. \([22]\) proposed applying the LogOdds transform to probabilistic labels, mapping them to a vector space. The LogOdds transform is also referred to as the additive log-ratio (ALR) transform by Aitchison \([30]\), and for a probabilistic label \(\mathbf{p} \in \mathbb{P}_L\) it is defined as

\[
\text{ALR}(\mathbf{p}) = \left[ \ln \left( \frac{p_1}{p_L} \right), \ldots, \ln \left( \frac{p_{L-1}}{p_L} \right) \right]^T.
\]

\(\text{ALR}^{-1}(\mathbf{v}) = C \left( [\exp(v_1), \ldots, \exp(v_{L-1}), 1]^T \right)\).

We will adopt the convention of referring to a vector resulting from applying the ALR transform to a probabilistic label as an “ALR vector” and the space of all such vectors as “ALR space”. ALR provides a bijection between \(\mathbb{P}_L\) and \(\mathbb{R}^{L-1}\), and standard statistical techniques can be performed in ALR space without requiring any projections.

Pohl et al. construct a shape model by performing PCA on sets of \(N\) probabilistic training segmentations by first applying the LogOdds transform to probabilistic labels. With \(n\) pixels and \(L\) labels, probabilistic segmentations mapped by the ALR transform can be represented as vectors in \(\mathbb{R}^{n(L-1)}\). Performing PCA on the \(N\) training segmentations provides the mean segmentation and the \(K < N\) eigenmodes of greatest variance, which best capture how the training segmentations vary from the mean. The PCA mean and modes approximate the space of feasible segmentations, which can be used, for example, to guide the segmentation of a novel image.

Probabilistic training data is often not available, as is the case when training segmentations are constructed manually. Such “crisp” (non-probabilistic) segmentations could be represented by assigning a binary indicator vector of \((L-1)\) 0’s and a single 1 to each pixel, but vectors in \(\mathbb{P}_L\) cannot contain 0’s, and thus the ALR transform cannot be applied. There are several possibilities for converting a crisp segmentation to a probabilistic one.

Pohl et al. take the approach of constructing probabilistic labels by calculating the SDMs of each crisp label (other than the background) and treating the vector of SDMs at each pixel as an ALR vector which can be mapped back to a probabilistic label. This approach provides intuitive results because the ALR
The ILR transform has the convenient property that the $i^{th}$ component of $\text{ALR}(p)$ varies monotonically with $p_i$, so pixels closest to the center of a crisp shape are given probabilities for label $i$ close to 1 and pixels far away from the shape are given probabilities for label $i$ close to 0.

As mentioned, a potential drawback of the ALR transform is that it is symmetric in the first $L-1$ labels, but not in the $L^{th}$ label (usually the background in segmentation, see (2)). With $I$ as the $(L-1) \times (L-1)$ identity matrix, $1$ as the vector of $(L-1)$ 1’s, and using $[\ldots]$ to denote a block matrix, we can rewrite (2) as the matrix-vector product

$$\text{ALR}(p) = M^T \ln(p)$$

where

$$M = \begin{bmatrix} I & -I^T \end{bmatrix}.$$ (5)

We use the natural logarithm $\ln(\cdot)$ and its inverse $\exp(\cdot)$. In a slight abuse of notation, $\ln(\cdot)$ and $\exp(\cdot)$ are applied to vectors, denoting a component-wise logarithm or exponentiation.

The ALR transform is a linear combination of vectors in $\mathbb{R}^{L-1}$ with weights $\ln(p_i)$. The first $(L-1)$ vectors are the standard basis vectors of $\mathbb{R}^{L-1}$, $\{e_1, \ldots, e_{L-1}\}$, whereas the $L^{th}$ vector is $-1$. Since $1 = \|e_1\| < \| -1\| = \sqrt{L-1}$, for the binary case $L = 2$ the ALR transform is symmetric between the two labels, but for $L > 2$, changes in the last label’s probability are magnified by $\sqrt{L-1}$ when mapped to ALR space. Thus, when comparing pixels that have high probability for one label ($\ell_1$) to pixels that have high probability for another label ($\ell_2$), the magnitude of the difference in the ALR vectors at those pixels will be $\sqrt{L-1}$ times greater when either $\ell_1$ or $\ell_2$ correspond to the background compared to when both $\ell_1$ and $\ell_2$ are foreground labels. When PCA is used to construct shape models, this leads to a greater variance being detected around foreground-background boundaries than around foreground-foreground boundaries, as we demonstrate in Sec. IV.

### B. Isometric Log-Ratio Transform

Egozcue et al. proposed the isometric log-ratio transform for compositional data analysis [31], defined as

$$\text{ILR}(p) = U^T \ln(p)$$

and for a scalar $\alpha$, power transformation is defined as

$$\alpha \odot \mathbf{p} = \mathbf{C}([p_1^\alpha, \ldots, p_L^\alpha]^T).$$ (12)

Perturbation and power transformation satisfy the usual properties of vector addition and scalar multiplication [32]. Fig. 2 illustrates the effect of perturbation and power transformation on a set of compositional vectors. The analogue of subtraction is defined as

$$\mathbf{p} \ominus \mathbf{q} = \mathbf{p} \oplus ((-1) \odot \mathbf{q}).$$ (13)

Thus, the block matrix $[U \mathbf{1}/\sqrt{L}]^T$ is orthogonal, so

$$[U \mathbf{1}/\sqrt{L}]^T [U \mathbf{1}/\sqrt{L}] = U^T U + \mathbf{1}/L = I. \quad (9)$$

Similar to the ALR transform, we will refer to a vector resulting from applying the ILR transform to a probabilistic label as an “ILR vector”.

Like the ALR transform, the ILR transform is a linear combination of vectors in $\mathbb{R}^{L-1}$ with weights $\ln(p_i)$. In ILR, the vectors are the columns of $U^T$, which are all equidistant from each other and from the origin (forming the vertices of a generalized tetrahedron), making ILR symmetric in all $L$ labels. An illustration of the difference between ALR and ILR is seen in Fig. 1.

Many equally valid choices for $U$ exist, in this paper we use a basis suggested by Egozcue et al.:
It has been shown that both the ILR and ALR transforms map perturbation and power transformation to standard Euclidean vector addition and scalar multiplication [31]. Indeed, from (5) and (7) we see both log-ratio transforms are of the form $\text{LR}(p) = A^T \ln(p)$ where $A^T 1 = 0$. Thus, for normalization constant $Z$,

$$\text{LR}(p \oplus q) = A^T \ln \left(\left[p_1 q_1, \ldots, p_L q_L\right]^T / Z\right)$$

$$= A^T \ln(p) + A^T \ln(q) - A^T 1 / Z$$

$$= \text{LR}(p) + \text{LR}(q)$$

$$\text{LR}(\alpha \odot p) = A^T \ln \left(\left[p_1^\alpha, \ldots, p_L^\alpha\right]^T / Z\right)$$

$$= \alpha A^T \ln(p) - A^T 1 / Z$$

$$= \alpha \text{LR}(p).$$

Building on this simplicial geometry, Aitchison introduced the Aitchison inner product, norm, and distance, giving the mean of a vector. That is, for vectors $v, w \in \mathbb{R}^m$,

$$\langle v, w \rangle = \sum_{i=1}^{m} v_i w_i$$

$$\mu(v) = \frac{1}{m} \sum_{i=1}^{m} v_i.$$

Then, for $p, q \in \mathbb{P}_L$, the Aitchison inner product, norm, and distance are defined as

$$\langle p, q \rangle_a = \langle \ln(p), \ln(q) \rangle - L \cdot \mu(\ln(p)) \cdot \mu(\ln(q))$$

$$\|p\|_a^2 = \langle p, p \rangle_a$$

$$d_a(p, q) = \|p \oplus q\|_a.$$

The Aitchison inner product is defined so that it satisfies the standard properties of an inner product, specifically with respect to vector addition and scalar multiplication (i.e. perturbation and power transformation). The Aitchison distance and norm also satisfy the standard distance and norm properties.
We let \( T \) be a random variable for \( t \) and apply Bayes rules:

\[
\hat{q}_i(x) = \frac{\Pr(G = \ell_i | T = t)}{\Pr(T = t)} = \frac{\Pr(T=t|G=\ell_i) \Pr(G=\ell_i)}{\Pr(T=t)}.
\]

(32)

We will keep the exact form of \( t \) general, but as an example, the observation could be a label assigned to \( x \) by a human expert. We may then model \( \Pr(T = t|G = \ell_i) \) by assuming some small probability \( \epsilon \) that the label \( t \) is incorrect, giving

\[
\Pr(T = t|G = \ell_i) = \begin{cases} 
1 - \epsilon & \text{if } t = \ell_i \\
\epsilon / (L - 1) & \text{otherwise}.
\end{cases}
\]

(34)

Letting \( p_i(x, t) = \Pr(T = t|G = \ell_i) / \Pr(T = t) \) and \( p(x, t) = C \left( [p_1(x, t), \ldots, p_L(x, t)]^T \right) \), we rewrite (32) as

\[
\hat{q}_i(x) = p_i(x, t) q_i(x) \\
\hat{q}(x) = p(x, t) \oplus q(x).
\]

(35)

(36)

That is, calculating a posterior distribution at a pixel is equivalent to performing a perturbation between the prior distribution and the normalized likelihood function. We note that perturbation has built in renormalization (see (11)), so the posterior distribution is guaranteed to sum to 1.

This leads to the useful property, shown by Tunc et al. [34], that if we have two posterior distributions derived from the same prior but different observations, \( \hat{q}_1 = p(t_1) \oplus q \) and \( \hat{q}_2 = p(t_2) \oplus q \), the distance between their log-ratio transformations is independent of the prior used: \( \|\text{LR}(\hat{q}_1) - \text{LR}(\hat{q}_2)\| = \|\text{LR}(p(t_1)) - \text{LR}(p(t_2))\| \). Thus the Aitchison distance between posterior distributions is invariant to the choice of prior.

In practice one may wish to incorporate spatial dependencies between neighboring pixels when updating the label probabilities, instead of updating each pixel’s probabilities separately as is done in (32). The above framework is meant only as an example to provide an interpretation of perturbation.

If, \( m \) times consecutively, we make the same observation \( t \), we perform the perturbation in (36) \( m \) times, or equivalently, we perturb \( q(x) \) by \( (m \odot p(x, t)) \). Thus, the power transformation weights how much influence an observation has. For example, if \( t_1 \) is input provided by a reliable human expert and \( t_2 \) is similar input provided by an untrained human, we could use the power transformation to give \( t_1 \) greater influence.

A distance is in general a quantification of “how much” must be added to one point to get to another. The Aitchison distance between compositional vectors \( q(x) \) and \( \hat{q}(x) \) is a measure of how much \( q(x) \) must be perturbed to get \( \hat{q}(x) \). Equivalently, \( d_\alpha(q, q) \) measures how much evidence must be observed to convince ourselves of a posterior \( \hat{q}(x) \) starting from a prior \( q(x) \).

If we assume a likelihood where \( \|p(x, t)\|_\alpha \leq C \) for some constant \( C \) and for all \( t \) (e.g. using the likelihood in (34)), then we can make a more concrete statement about the Aitchison distance: to get from prior \( q(x) \) to posterior \( \hat{q}(x) \), we must observe corroborating evidence at least \( (d_\alpha(q(x), \hat{q}(x))/C) \) times. A graphic example of perturbing by vectors with different Aitchison norms is seen in Fig. 4.

An inner product provides a notion of orthogonality. For example, if we have a set of compositional vectors \( P = \{p_1, \ldots, p_k\} \), and another compositional vector \( r \) is orthogonal to each vector in \( P \), then no linear combination (with respect to perturbation and power transformation) of vectors in \( P \) can result in \( r \). If the vectors in \( P \) and \( r \) all correspond to likelihoods from different observations (see (36)), \( r \) provides information that is complimentary to the information provided by the vectors in \( P \): the evidence provided by \( r \) neither supports nor refutes the evidence provided by \( P \). A graphic example of how the Aitchison inner product can be used to identify parallel and orthogonal likelihood functions is seen in Fig. 4.

Since the ILR transform is an isometry between the Euclidean and Aitchison distances, the above interpretations apply to the standard Euclidean distance and inner product when used on ILR vectors. As discussed in Sec. II-C, the ALR transform is not an isometry, and thus distances between ALR vectors do not have such a straightforward interpretation.

In the next section, we provide examples for both synthetic data and real medical images, comparing statistical analysis done in both ALR and ILR space. These examples highlight how the symmetry of the ILR transform leads to more intuitive results in some applications.

**IV. RESULTS**

Principal component analysis (PCA) is a commonly used method for constructing a shape model from existing segmentations. In this section, we build shape models by applying the ALR and ILR transforms to rigidly aligned training segmen-
tions, using PCA to capture variability in the transformed segmentations, and then mapping the results back to the simplex of probabilistic labels. For brevity, we denote these shape models as ALR-PCA and ILR-PCA. For crisp training segmentations, we use the method outlined in Sec. II-A, where vectors of SDMs of the crisp shapes are treated as ALR vectors and mapped to the simplex to create probabilistic segmentations. A minimum probability of 0.01 is imposed.

To construct a PCA shape model, a set of ALR or ILR transformed training segmentations are cast to vectors in \( \mathbb{R}^{n(L−1)} \) and PCA is performed to find a mean segmentation \( \overline{v} \), a set of eigenmodes of greatest variance \( V = [v_1, \cdots, v_k] \), and the variances associated with the modes, \( \lambda_1 \geq \cdots \geq \lambda_k \). Since the segmentations are roughly aligned, the PCA modes \( V \) should capture how the boundaries of shapes tend to deform across the training data.

Since the mean can be calculated using only vector addition and scalar multiplication, \( \overline{v} \) is the same in ALR and ILR space. The modes, however, depend on distances, and thus will be different in the ALR- and ILR-PCA shape models. We note that if all of the modes are taken, the set of probabilistic segmentations in the span of the modes will be the same for ALR- and ILR-PCA, but the associated variances of the modes will differ. Since changes in the last component (the background) of a probabilistic label are magnified by a factor of approximately \( \sqrt{L−1} \) when mapped by the ALR transform (see Sec. II-A), these changes affect the variance much more than changes in the other components, and thus we expect the ALR-PCA shape model to focus more on modeling changes in the background.

A. Synthetic Data

We begin with a simple synthetic example to demonstrate how the asymmetry in ALR can affect PCA results. We uniformly sample (with respect to the Euclidean distance) 10,000 random probabilistic labels from \( \mathbb{P}_{L} \), for each \( L \in \{3, \cdots, 15\} \). We then take each sampled probabilistic label \( p \) and skew it away from the center of the simplex by pushing the 1\textsuperscript{st} label’s probability, \( p_1 \), away from 0.5 using a sigmoid function,

\[
p_1 \leftarrow (1 + \exp(-30(p_1 - 0.5)))^{-1},
\]

then renormalizing. An example of this skew for \( \mathbb{P}_3 \) is seen in Fig. 5a. We map the skewed labels using the ALR and ILR transforms and perform PCA on both to find \( k = L−1 \) modes. Since the probabilities are skewed along one label only, and are symmetric in the remaining labels, we expect that:

1) mode \( v_1 \) corresponds mostly to changes in the 1\textsuperscript{st} label,

2) variances \( \{\lambda_2, \cdots, \lambda_{L−1}\} \) are all similar.

We design a test to quantify both of these expectations, and repeat these tests 100 times on different randomly generated data, with the results summarized in Fig. 5.

To quantify the first expectation, we calculate how the label probabilities change when taking a small step along \( v_1 \) away from the center of the simplex (the origin in ALR and ILR spaces). We denote the change in probability \( p_i \) as \( \delta_i \). If \( v_1 \) does indeed correspond mostly just to changes in \( p_1 \), we expect the other label probabilities to change minimally, by about \( -\delta_1/(L−1) \) each, in order to maintain that the probabilities sum to one. That is, we expect

\[
\frac{\delta_i}{\delta_1} = -\frac{1}{L−1}.
\]

Thus we define the “error” in the change of \( p_i, i \geq 2 \), as

\[
E_i = \left| \frac{\delta_i}{\delta_1} - \frac{-1}{L−1} \right|.
\]

Fig. 5. (Color Figure) 5a demonstrates how vectors in \( \mathbb{P}_3 \) are stretched along the first label, 5b plots the mean error of the changes of the label probabilities (see (39)) when taking a small step along the eigenmode of greatest variance, \( v_1 \). 5c plots the CoV for \( \{\lambda_2, \cdots, \lambda_{L−1}\} \). See text for further details.

Fig. 6. (Color Figure) 6a shows an example slice from a thigh image, and 6b shows the corresponding 16 label segmentation.
We summarize these errors for a given test by taking the mean of \( \{E_2, \ldots, E_L\} \). Fig. 5b plots these mean errors. In ILR-PCA, the mean error is always close to 0, as expected. In ALR-PCA, the error increases for larger \( L \), since \( v_1 \) focuses more on capturing the increased variance in the last label.

To quantify the second expectation, we take a normalized measure of dispersion known as the coefficient of variation (CoV)\(^1\) of \( \{\lambda_2, \cdots, \lambda_{L-1}\} \). We expect the CoV to be close to 0, indicating these variances are close in value. Fig. 5c plots the CoV for \( \{\lambda_2, \cdots, \lambda_{L-1}\} \). In ILR-PCA, the CoV is around 0, as expected. In ALR-PCA, the CoV reaches a peak around \( L = 8 \), where the ALR space variance of the 1\(\text{st} \) and last components are about the same, and both the first and second modes are required to explain this variance.

While PCA in ILR space gives the expected results, PCA in ALR space does not, especially for larger \( L \). This results from changes in the last label being magnified in ALR space by \( \sqrt{L - 1} \) and the PCA modes capturing this variation.

### B. Medical Data - PCA Model Generalizability

In this section we provide an example illustrating the advantages of using the ILR transform to model anatomical shapes in real medical data. We build anatomical shape models using PCA in both ALR and ILR space, and then examine how accurately these shape models generalize to unseen shapes. We use a data set consisting of 39 MRIs of thighs with \( 300 \times 200 \times 30 \) voxels, each segmented into subcutaneous fat, intermuscular fat, cortical bone, bone marrow, 11 thigh muscles, and background [35], aligned as described by [28] (Fig. 6b). This data set is used because most of the boundaries are between adjacent anatomical regions, and not with the background. Since ALR-PCA gives extra weight to variations in the background, we expect the shapes of the muscle and bone labels to be modeled more accurately by the ILR-PCA shape model, since in ILR transformed data all label boundaries are treated equally. Conversely, we expect the shape of the background and the label it borders (the subcutaneous fat) to be more accurately modeled by the ALR-PCA shape model.

We first test the ability of the ALR- and ILR-PCA shape models to generalize to new images. We performed PCA on a training set of 20 of the images (chosen at random) to construct a shape model. \( k \) PCA modes were found, where \( k \) was chosen to account for 90\% of the variability in the training data (11 modes for both ALR- and ILR-PCA). The remaining 19 segmentations were used to test the generalizability of the shape models. Each testing segmentation was projected onto the span of the PCA modes. The projected segmentations represent the closest approximation to the original testing segmentations in the PCA shape space.

We evaluated the generalizability of the shape models by comparing the original segmentations to their projections. We compare them on a label by label basis using a fuzzy extension of the Dice similarity coefficient (DSC) (see appendix for further details). We use a fuzzy DSC instead of thresholding the segmentations so as not to lose uncertainty information. The DSCs for each label are summarized by their mean and standard deviation in Table I, along with the \( p \)-values generated from a paired \( t \)-test to test the null hypothesis that ALR- and ILR-PCA have the same mean DSC. We see the background label is more accurately modeled by ALR-PCA, but ILR-PCA models almost all of the foreground labels more accurately. The subcutaneous fat (label 15) is modeled equally well by both shape ALR- and ILR-PCA since half of its boundary is with the background and half with the other foreground labels.

### C. Medical Data - Segmentation Using PCA

In this section, we use the ALR- and ILR-PCA shape models constructed in Sec. IV-B in a fully automatic segmentation scheme in order to demonstrate the advantages of using ILR in a real application. For each of the 19 thigh images in the testing set, we construct a strictly convex energy function over the space of log-ratio (ALR or ILR) segmentations, and minimize that energy to produce a probabilistic segmentation.

---

**Table I**

The mean and standard deviation of the DSC values calculated between the 19 testing segmentations and their corresponding projections onto the ALR- and ILR-PCA shape spaces, built from the 20 training images. Also shown are the \( p \)-values generated from a paired \( t \)-test to determine if the differences in DSC values are statistically significant. Bolded \( p \)-values pass the Holm-Bonferroni test (used to correct for multiple comparisons) at a significance level of 0.05. We note that ILR-PCA performs better when modeling the labels that do not share a boundary with the background (\( L \in \{1, \ldots, 14\} \)), see Fig. 6.

<table>
<thead>
<tr>
<th>Type</th>
<th>( t )</th>
<th>ALR Proj. DSCs</th>
<th>ILR Proj. DSCs</th>
<th>( p )-Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muscle</td>
<td>1</td>
<td>0.63 ± 0.15</td>
<td>0.7 ± 0.15</td>
<td><strong>0.0019</strong></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.73 ± 0.12</td>
<td>0.73 ± 0.14</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.74 ± 0.1002</td>
<td>0.76 ± 0.14</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.73 ± 0.13</td>
<td>0.76 ± 0.12</td>
<td>3.7 \times 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.79 ± 0.061</td>
<td>0.81 ± 0.046</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.77 ± 0.11</td>
<td>0.80 ± 0.10</td>
<td><strong>0.00097</strong></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.84 ± 0.075</td>
<td>0.86 ± 0.064</td>
<td>0.0071</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.58 ± 0.19</td>
<td>0.64 ± 0.13</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.84 ± 0.056</td>
<td>0.86 ± 0.058</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.72 ± 0.10</td>
<td>0.74 ± 0.094</td>
<td>0.11</td>
</tr>
<tr>
<td>Bone</td>
<td>11</td>
<td>0.85 ± 0.030</td>
<td>0.88 ± 0.025</td>
<td>1.1 \times 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.56 ± 0.19</td>
<td>0.59 ± 0.15</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.50 ± 0.25</td>
<td>0.53 ± 0.24</td>
<td>0.48</td>
</tr>
<tr>
<td>Fat</td>
<td>14</td>
<td>0.38 ± 0.13</td>
<td>0.43 ± 0.11</td>
<td><strong>0.00062</strong></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.89 ± 0.11</td>
<td>0.79 ± 0.11</td>
<td>0.24</td>
</tr>
<tr>
<td>Background</td>
<td>16</td>
<td>0.9795 ± 0.0081</td>
<td>0.9702 ± 0.0005</td>
<td><strong>0.0013</strong></td>
</tr>
</tbody>
</table>

\( ^1 \)The CoV is the standard deviation divided by the mean.
for the image. We perform each segmentation twice, once using the ALR transform to represent the segmentation and once using the ILR transform instead.

A typical energy function used in segmentation might contain energy terms to
1) ensure regional intensity profiles match learned profiles
2) ensure regional boundaries match edges detected in the image, and
3) ensure regions’s shapes conforms to known shape priors.

Our energy consists of straightforward intensity-based, boundary-based, and shape-based terms, described below. LR(·) is used to denote either of the log-ratio transforms. We use $S : \Omega \to \mathbb{R}^{L-1}$ to represent a log-ratio space segmentation. Using $p : \Omega \to \mathbb{P}_L$ as an intensity based probabilistic segmentation (i.e. $p$ is based only on intensity values, and has no spatial regularization), we define the energy term:

$$E_{\text{intensity}}(S) = - \sum_{x \in \Omega} \langle S(x), \text{LR}(p(x)) \rangle$$, \hspace{1cm} (40)

which incentivizes the final segmentation to be similar to $p$. Using an “edgeness” function $h : \Omega \to [0,1]$, indicating the likely locations of region boundaries based on the image gradient, we define the weighted total variation term:

$$E_{\text{boundary}}(S) = \sum_{x \in \Omega} (1 - h(x)) \| \nabla S(x) \| \hspace{1cm} (41)$$

For the PCA mean $\bar{v}$, PCA modes $V = [v_1, \ldots, v_k]$, and the diagonal matrix $\Lambda$ with the variances $(\lambda_1, \ldots, \lambda_k)$ along the diagonal we define the terms:

$$E_{\text{shape}}(S) = (S - \bar{v})^T V \Lambda^{-1} V^T (S - \bar{v})$$ \hspace{1cm} (42)

$$E_{\text{proj}}(S) = \| (S - \bar{v}) - V V^T (S - \bar{v}) \|^2.$$ \hspace{1cm} (43)

In (42) and (43), $S$ is cast to an $n(L-1)$ length vector. $E_{\text{shape}}$ is a squared Mahalanobis distance penalty, encouraging $S$ to lie along the PCA modes of greatest variance. $E_{\text{proj}}$ penalizes the squared distance between $S$ and its projection onto the span of the PCA modes, encouraging the segmentation to not stray far from the PCA shape space. We note that if all the PCA modes are used, (42) is the same when using ALR- or ILR-PCA, though (43) will always be different.

Our automatic segmentation is then given by solving the minimization problem

$$S^* = \arg\min_S \left( w_1 E_{\text{intensity}} + w_2 E_{\text{boundary}} + w_3 E_{\text{shape}} + w_4 E_{\text{proj}} \right),$$ \hspace{1cm} (44)

where the $w_i$’s are scalar weights. $S^*$ can be calculated quickly and exactly since each of the terms are convex or strictly convex. To increase robustness to parameter tuning, each segmentation is performed 125 times with a different set of weights each time, with the best result taken. The 125 sets of weights were fixed for all images.

We chose a straightforward and general segmentation scheme since the goal of this section is not to construct a state-of-the-art segmentation method, but to demonstrate that the results in Sec. IV-A and Sec. IV-B extend to real applications. The simplicity of the scheme reduces the influences of extraneous factors (no user input, robust to parameter setting, initialization independent). We note that the problem of thigh muscle segmentation is difficult; even the DSC values for the initial segmentation shown here, can suffer significantly when working in ALR space. As seen in Sec. IV-B, the ALR space PCA

<table>
<thead>
<tr>
<th>Type</th>
<th>$t$</th>
<th>ALR-Seg DSCs</th>
<th>ILR-Seg DSCs</th>
<th>$p$-Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muscle</td>
<td>1</td>
<td>0.70 ± 0.19</td>
<td>0.74 ± 0.22</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.81 ± 0.13</td>
<td>0.82 ± 0.11</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.78 ± 0.14</td>
<td>0.85 ± 0.186</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.75 ± 0.12</td>
<td>0.77 ± 0.12</td>
<td>0.00018</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.79 ± 0.073</td>
<td>0.85 ± 0.078</td>
<td>0.00084</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.79 ± 0.15</td>
<td>0.80 ± 0.16</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.84 ± 0.13</td>
<td>0.87 ± 0.1</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.43 ± 0.24</td>
<td>0.46 ± 0.21</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.85 ± 0.081</td>
<td>0.86 ± 0.068</td>
<td>0.0114</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.75 ± 0.14</td>
<td>0.75 ± 0.13</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.89 ± 0.045</td>
<td>0.89 ± 0.043</td>
<td>0.32</td>
</tr>
<tr>
<td>Bone</td>
<td>12</td>
<td>0.62 ± 0.16</td>
<td>0.72 ± 0.13</td>
<td>4.4 \times 10^{-6}</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.61 ± 0.37</td>
<td>0.83 ± 0.2</td>
<td>0.0075</td>
</tr>
<tr>
<td>Fat</td>
<td>14</td>
<td>0.45 ± 0.14</td>
<td>0.59 ± 0.14</td>
<td>0.0061</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.72 ± 0.07</td>
<td>0.85 ± 0.075</td>
<td>8.9 \times 10^{-10}</td>
</tr>
<tr>
<td>Background</td>
<td>16</td>
<td>0.96 ± 0.033</td>
<td>0.98 ± 0.012</td>
<td>0.00029</td>
</tr>
</tbody>
</table>

**Table II**

<table>
<thead>
<tr>
<th>Type</th>
<th>$t$</th>
<th>ALR-Seg DSCs</th>
<th>ILR-Seg DSCs</th>
<th>$p$-Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muscle</td>
<td>1</td>
<td>0.70 ± 0.19</td>
<td>0.74 ± 0.22</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.81 ± 0.13</td>
<td>0.82 ± 0.11</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.78 ± 0.14</td>
<td>0.85 ± 0.186</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.75 ± 0.12</td>
<td>0.77 ± 0.12</td>
<td>0.00018</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.79 ± 0.073</td>
<td>0.85 ± 0.078</td>
<td>0.00084</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.79 ± 0.15</td>
<td>0.80 ± 0.16</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.84 ± 0.13</td>
<td>0.87 ± 0.1</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.43 ± 0.24</td>
<td>0.46 ± 0.21</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.85 ± 0.081</td>
<td>0.86 ± 0.068</td>
<td>0.0114</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.75 ± 0.14</td>
<td>0.75 ± 0.13</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.89 ± 0.045</td>
<td>0.89 ± 0.043</td>
<td>0.32</td>
</tr>
<tr>
<td>Bone</td>
<td>12</td>
<td>0.62 ± 0.16</td>
<td>0.72 ± 0.13</td>
<td>4.4 \times 10^{-6}</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.61 ± 0.37</td>
<td>0.83 ± 0.2</td>
<td>0.0075</td>
</tr>
<tr>
<td>Fat</td>
<td>14</td>
<td>0.45 ± 0.14</td>
<td>0.59 ± 0.14</td>
<td>0.0061</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.72 ± 0.07</td>
<td>0.85 ± 0.075</td>
<td>8.9 \times 10^{-10}</td>
</tr>
<tr>
<td>Background</td>
<td>16</td>
<td>0.96 ± 0.033</td>
<td>0.98 ± 0.012</td>
<td>0.00029</td>
</tr>
</tbody>
</table>

**Table II**

The mean DSC values comparing the automatically generated segmentations to the actual ground truth segmentations for the 19 volumetric thigh images used for testing. Also shown are the $p$-values generated from a paired $t$-test to determine if the differences in DSC values are statistically significant. Bolded $p$-values pass the Holm-Bonferroni test (used to correct for multiple comparisons) at a significance level of 0.05.
shape model does not capture non-background variability as well as the ILR shape model. We also note that the background was segmented more accurately by ILR-Seg, the opposite result of what we would expect from Sec. IV-B. We attribute this to the fact that in ILR space $E_{\text{boundary}}$ penalizes the boundary of the background more strongly than the boundaries of the foreground regions by a factor of approximately $\sqrt{L} - 1$. This results in either over-regularization of the background boundaries or under regularization of the background boundaries, depending on the choice of $w_2$.

V. DISCUSSION AND CONCLUSIONS

In this paper, we compared techniques for mapping probabilistic labels to a vector space in order to facilitate algebraic manipulation and statistical analysis. A well known method for performing this mapping is the LogOdds (ALR) transform, which has several useful properties. However, LogOdds is asymmetric when there are multiple foreground labels, and changes in the background probability cause a greater change in the ALR transformed data than equivalent changes in the foreground probabilities. Making use of established methods for compositional data analysis, we proposed using the symmetric ILR transform to map probabilistic label.

We demonstrated that the asymmetry of the ALR transform magnifies the importance of changes in the background label in statistical analysis techniques such as PCA, at the cost of other label boundaries, leading to shape models that capture foreground-background boundaries more accurately than foreground-foreground boundaries. Thus, the ILR transform is a more appropriate mapping in applications containing many foreground-foreground boundaries.

The choice of log-ratio transform (ALR, ILR, or other) affects the calculation of variance in statistical techniques, so while log-ratio transforms are a powerful tool for probabilistic shape analysis, it is important to use a transform appropriate for the application.

We note that an alternative approach to performing statistical analysis on probabilistic labels is to employ a non-Euclidean simplicial metric (e.g. the Fisher information metric or the Bhattacharyya metric) and use an extension of PCA to Riemannian manifolds (e.g. principal geodesic analysis [36]). In the future, we will explore the relationship between these metrics and the Aitchison metric in the context of shape analysis.

We note, however, several advantages the Aitchison metric provides that make it a strong choice when performing PCA on probabilistic segmentations:

1) It is isometric to the Euclidean metric through the ILR transform, simplifying the implementation of PCA.
2) The connections between the Aitchison metric and Bayesian inference, highlighted in Sec. III, motivate the use of the Aitchison geometry.
3) Many other simplicial metrics are compact, and would require projections or constraints to ensure probabilities stay in the simplex.

In the future, we will also explore how log-ratio transforms can be customized and extended for specific tasks. For example, if smaller regions were magnified in the log-ratio space (as the background is in ALR), shape models may encode these regions more accurately.

ACKNOWLEDGMENT

We thank Dr. W. D. Reid, Department of Physical Therapy, University of British Columbia for the thigh muscle MRI data and Dr. B. HajGhanbari for the corresponding manual segmentations. Partial funding for this project was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES


