

# Understanding Higher-Order Interactions in Information Space

Herbert Edelsbrunner<sup>1</sup>, Katharina Ölsböck<sup>1</sup> and Hubert Wagner<sup>2,\*</sup>

<sup>1</sup> ISTA (Institute of Science and Technology Austria), 3400 Klosterneuburg, Austria; edels@ist.ac.at (H.E.); katharina.oelsboeck@ist.ac.at (K.Ö.)

<sup>2</sup> Department of Mathematics, University of Florida, Gainesville, FL 32611, USA

\* Correspondence: hwagner@ufl.edu

**Abstract:** Methods used in topological data analysis naturally capture higher-order interactions in point cloud data embedded in a metric space. This methodology was recently extended to data living in an information space, by which we mean a space measured with an information theoretical distance. One such setting is a finite collection of discrete probability distributions embedded in the probability simplex measured with the relative entropy (Kullback–Leibler divergence). More generally, one can work with a Bregman divergence parameterized by a different notion of entropy. While theoretical algorithms exist for this setup, there is a paucity of implementations for exploring and comparing geometric-topological properties of various information spaces. The interest of this work is therefore twofold. First, we propose the first robust algorithms and software for geometric and topological data analysis in information space. Perhaps surprisingly, despite working with Bregman divergences, our design reuses robust libraries for the Euclidean case. Second, using the new software, we take the first steps towards understanding the geometric-topological structure of these spaces. In particular, we compare them with the more familiar spaces equipped with the Euclidean and Fisher metrics.

**Keywords:** higher-order interactions; topological data analysis; persistent homology; simplicial complex; alpha shape; wrap complex; information theory; Shannon entropy; relative entropy; Bregman divergence; non-Euclidean geometry; Bregman geometry



**Citation:** Edelsbrunner, H.; Ölsböck, K.; Wagner, H. Understanding Higher-Order Interactions in Information Space. *Entropy* **2024**, *26*, 637. <https://doi.org/10.3390/e26080637>

Academic Editors: Tiziano Squartini, Fernando Antônio Nóbrega Santos, Rossana Mastrandrea and Marco Piangerelli

Received: 19 June 2024

Revised: 19 July 2024

Accepted: 20 July 2024

Published: 27 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The motivation for the work reported in this paper is the deeper understanding of the effect of the ambient metric on the higher-order interactions arising in data, as well as its multi-scale geometric-topological properties. The first can be viewed via the Čech or Alpha complex, the second through their persistent homology. Specifically, we generalize geometric and topological data analysis methods from Euclidean geometry to Bregman geometries in which dissimilarity is measured with divergences that are generally non-symmetric and violate the triangle inequality. By necessity, these methods are sensitive to the dissimilarity defining the ambient geometry, and we exploit this sensitivity to quantify the difference between geometries.

As example geometries, we emphasize those related to information-theoretic concepts, such as the Shannon geometry and the Fisher geometry. What we call the Shannon geometry is induced by the relative entropy (Kullback–Leibler divergence) [1], which is based on Shannon’s entropy [2]. The Fisher geometry is induced by the Fisher distance (more technically the Fisher–Rao metric [3]). It was introduced by Rao [4], building on Fisher’s seminal work [5] in statistics.

These are examples of what we call information spaces [6]. Such spaces often exhibit unfamiliar behavior, different from the Euclidean space we are used to. Understanding such spaces is however important, as—plainly speaking—this is where many types of data live [7].

Importantly, these geometries are commonly used in data science. For example, relative entropy is a standard loss function used in dimensionality reduction [8,9] and deep learning [7,10]. (This connection may not be immediately obvious. Many models aim to explicitly minimize what is called cross-entropy. It has the same minimizer as the relative entropy, which can additionally be viewed as a distance, and therefore gives rise to a geometry). The Fisher geometry starts to be considered [11] as an alternative.

We therefore deem it important to shed light on the differences and commonalities between these geometries. Some pertinent questions are as follows: Is the simpler Fisher geometry a good approximation of the Shannon geometry? Can we see a significant difference between the Euclidean geometry and the non-Euclidean ones, as predicted by the discrepancy in their practical performances?

In the current work, we are particularly interested in the algorithms that underpin the data analysis methods, especially the topological ones. While the Fisher geometry can be handled with Euclidean tools [6], the Shannon geometry used to require customized ones [12]. We show that the Shannon geometry can also benefit from existing robust tools, although in this case the application is less direct. We also hope that this development opens new alleys for topological data analysis in information spaces.

**Prior work and results.** The research reported in this paper merges several lines of work. One is the study of Bregman divergences from the computational geometry point of view initiated in [13] and the extension of topological data analysis to Bregman and Fisher geometries started in [6,12]. Another is the study of higher-order interactions in high-dimensional probabilistic feature spaces [14,15], which generally lack information-theoretical interpretations.

Pivotal work at the intersection of computational geometry and Bregman geometry includes the extension of k-means clustering [16], proximity search data structures [17,18], and Voronoi tessellations [13] to the Bregman setting.

On a practical note, we mention a research direction focusing on detecting, loosely speaking, the holes in knowledge [14,15,19,20] understood as a collection of text documents. In the simplest case, each text document is represented as a discrete probability distribution (by counting characteristic keywords [21]). Interestingly, higher-order interactions among documents are necessary to capture the multi-scale topology of these data. Indeed, pairwise, triplewise, etc., intersections between balls (with respect to the appropriate metric or divergence) centered at the data points reveal the topological structure.

In this work, we will also use computational topology methods. One important topological tool is the nerve and the related Nerve Theorem. Intuitively, it allows one to capture the topology (more precisely, the homotopy type) of a union of (a finite collection of) objects using a finite simplicial complex. The theorem is usually attributed to Borsuk [22] or Leray [23], and is stated more explicitly by Weil [24]. See [25] for a comprehensive, modern treatment of this topic.

A particularly important concept in our investigations is the Bregman–Delaunay mosaic, which we formally define as the straight-line dual of the not necessarily straight-line Bregman–Voronoi tessellation obtained by measuring distance with the Bregman divergence from a data point. This mosaic was already defined in [13], and we explain how it can be computed as a weighted Euclidean Delaunay mosaic using standard geometric software. In Euclidean space, the Alpha shapes can be defined as the sublevel sets of the radius function on the Delaunay mosaic, which is a generalized discrete Morse function in the sense of Forman [26] and Freij [27]. The lower sets of the critical simplices of this function constitute the Wrap complex [28], which was introduced as a shape reconstruction tool in [29]. We extend this framework by introducing the rise function on a Bregman–Delaunay mosaic, which provides a convenient measure of the size of a Bregman ball. With these notions, we construct the shape of data in different geometries, and we use them to quantify the difference between the geometries.

We have implemented all the algorithms and used the resulting software to run experiments comparing the Euclidean, Shannon, and Fisher geometries for synthetic data.

We find that the Delaunay mosaics and their Alpha and Wrap complexes in these geometries show some, occasionally subtle, differences, which we quantify.

**Outline.** Section 2 provides the necessary background from discrete geometry and computational topology. Section 3 gives the details needed to compute Delaunay mosaics and their Alpha and Wrap complexes in Bregman and Fisher geometries using software for weighted Delaunay mosaics in Euclidean geometry. Section 4 presents computational experiments, and Section 5 discusses the quantification of the difference between Bregman and other geometries. Section 6 concludes the paper. In Appendices A–D, we provide supplementary experiments, results, and background information.

## 2. Background

We need some background on Bregman divergences, Delaunay mosaics, and discrete Morse functions. Indeed, this paper combines these concepts to obtain new insights into Bregman–Delaunay mosaics and their scale-dependent subcomplexes.

**Bregman divergence.** Given a suitable convex function on a convex domain, the best affine approximation at a point defines a dissimilarity measure on the domain (see [30]). We follow [31] in the details of this construction, requiring a technical third condition that guarantees a conjugate function of the same kind. Let  $\Omega \subseteq \mathbb{R}^d$  be an open and convex domain. A function  $F: \Omega \rightarrow \mathbb{R}$  is of Legendre type [32] if

- (i)  $F$  is differentiable;
- (ii)  $F$  is strictly convex;
- (iii)  $\nabla F$  diverges whenever we approach the boundary of  $\Omega$ .

If the boundary of the domain is empty, which is the case for  $\Omega = \mathbb{R}^d$ , then Condition (iii) is void. In other words,  $\|\nabla F(x)\|$  does not necessarily diverge when  $\|x\| \rightarrow \infty$ . Given points  $x, y \in \Omega$ , the Bregman divergence from  $x$  to  $y$  associated with  $F$  is the difference between  $F$  and the best affine approximation of  $F$  at  $y$ , both evaluated at  $x$ :

$$D_F(x||y) = F(x) - [F(y) + \langle \nabla F(y), x - y \rangle], \quad (1)$$

where  $\langle \cdot, \cdot \rangle$  denotes the standard dot product. Note that  $D_F(x||y) \geq 0$ , with equality iff  $x = y$ , resembling a metric. However, the other two axioms of a metric do not hold: the divergence is not necessarily symmetric, and it violates the triangle inequality in all non-trivial cases. In spite of these shortcomings, Bregman divergences are useful as measures of dissimilarity and are sufficient to define a geometry.

For a given  $h \geq 0$ , the primal ball with center  $x$  contains all points  $y$  such that the divergence from  $x$  to  $y$  is at most  $h$ , and the dual ball contains all points  $y$  such that the divergence from  $y$  to  $x$  is at most  $h$ :

$$B_F(x, h) = \{y \in \Omega \mid D_F(x||y) \leq h\}, \quad (2)$$

$$B_F^*(x, h) = \{y \in \Omega \mid D_F(y||x) \leq h\}. \quad (3)$$

The geometric intuition for (2) is to cast light onto the graph of  $F$  from a point vertically above  $x \in \mathbb{R}^d$  in  $\mathbb{R}^{d+1}$  and at distance  $h$  below the graph of  $F$ : the primal ball is the vertical projection of the lit up part of the graph onto  $\mathbb{R}^d$ . This ball is not necessarily convex. The geometric intuition for (3) is to intersect the graph of  $F$  with the tangent hyperplane at  $x$  shifted vertically upward by a distance  $h$ : the dual ball is the vertical projection of the part of the graph on or below this shifted hyperplane. This ball is necessarily convex.

The conjugate of  $F$  can be constructed with elementary geometric means. Specifically, we use the polarity transform that maps a point  $A = (a, a_{d+1}) \in \mathbb{R}^d \times \mathbb{R}$  to the affine map  $A^*: \mathbb{R}^d \rightarrow \mathbb{R}$  defined by  $A^*(x) = \langle a, x \rangle - a_{d+1}$ . Similarly, it maps  $A^*$  to  $A = (A^*)^*$ . The graph of  $F$  can be described as a set of points or a set of affine maps that touch the graph. The conjugate function,  $F^*: \Omega^* \rightarrow \mathbb{R}$ , is defined such that polarity maps the points of the graph of  $F$  to the tangent affine maps of the graph of  $F^*$ , and it maps the tangent affine

maps of the graph of  $F$  to the points of the graph of  $F^*$ . Since  $A$  and  $A^*$  switch position with gradient, so do  $F$  and  $F^*$ . More specifically,  $\Omega^* = \phi(\Omega)$  and  $F^* : \Omega^* \rightarrow \mathbb{R}$  are given by

$$\phi(x) = \nabla F(x), \tag{4}$$

$$F^*(\phi(x)) = \langle \nabla F(x), x \rangle - F(x), \tag{5}$$

$$\nabla F^*(\phi(x)) = x. \tag{6}$$

The convexity of  $\Omega$  and Conditions (i), (ii), (iii) imply that  $\Omega^*$  is convex and  $F^*$  satisfies (i), (ii), (iii). In other words, the conjugate of a Legendre type function is again a Legendre type function. Importantly, the Bregman divergences associated with  $F$  and with  $F^*$  are symmetric:  $D_F(x||y) = D_{F^*}(\phi(y)||\phi(x))$ . Hence,  $\phi$  maps primal balls to dual balls and it maps dual balls to primal balls:

$$B_{F^*}^*(\phi(x), h) = \phi(B_F(x, h)), \tag{7}$$

$$B_{F^*}(\phi(x), h) = \phi(B_F^*(x, h)). \tag{8}$$

Since all dual balls are convex, all primal balls are diffeomorphic images of convex sets. This implies that the common intersection of a collection of primal balls is either empty or contractible [12], so the Nerve Theorem applies.

**Examples.** An important example of a Legendre type function is  $\omega : \mathbb{R}^d \rightarrow \mathbb{R}$  defined by mapping  $x$  to half the square of its Euclidean norm:  $\omega(x) = \frac{1}{2}\|x\|^2$ . It is the only Legendre type function that is its own conjugate:  $\omega = \omega^*$ . The symmetry between the divergences of a Legendre type function and its conjugate thus implies  $D_\omega(x||y) = D_\omega(y||x)$  and  $B_\omega(x, h) = B_\omega^*(x, h)$ . Indeed, it is easy to see that the divergence is half the squared Euclidean distance,  $D_\omega(x||y) = \frac{1}{2}\|x - y\|^2$ , which is of course symmetric. This particular Legendre type function provides an anchor point for comparison.

The example that justifies the title of this paper is the (negative) Shannon entropy [2]. We add an extra term, resulting in  $E : \mathbb{R}_+^d \rightarrow \mathbb{R}$ , defined as  $E(x) = \sum_{i=1}^d [x_i \ln x_i - x_i]$ . Being linear, the additional term does not affect the resulting divergence, but simplifies certain computations later.

This function is of Legendre type and fundamental to information theory [2]. Its divergence,

$$D_E(x||y) = \sum_{i=1}^d [x_i \ln x_i - x_i \ln y_i - x_i + y_i], \tag{9}$$

is generally referred to as the relative entropy [10] or the Kullback–Leibler divergence [1] from  $x$  to  $y$ . We remark that the above derivation using the Bregman machinery results in a divergence which is valid on the entire positive orthant of  $\mathbb{R}^d$ , and agrees with the standard definition [1] on the standard simplex.

The gradient of the Shannon entropy at  $x$  is the vector  $\nabla E(x)$  with components  $\ln x_i$  for  $1 \leq i \leq d$ . Using Equation (5), one easily computes that the conjugate of  $E$  maps this vector to  $\sum_{i=1}^d x_i$ . Hence  $E^* : \mathbb{R}^d \rightarrow \mathbb{R}$  is defined by mapping  $y \in \mathbb{R}^d$  to  $E^*(y) = \sum_{i=1}^d e^{y_i}$ .

A case of special interest is the restriction of the Shannon entropy to the standard simplex, which is a subset of the positive orthant. Writing  $x = (x_1, x_2, \dots, x_d)$  for a point of  $\mathbb{R}_+^d$ , the *standard*  $(d - 1)$ -simplex, denoted  $\Delta^{d-1}$ , consists of all points  $x$  that satisfy  $x_1 + x_2 + \dots + x_d = 1$ . We use  $\Delta^{d-1}$  as the domain of a Legendre type function, which is the reason why we introduce  $\Delta^{d-1}$  as an open set. Finally, write  $E_\Delta : \Delta^{d-1} \rightarrow \mathbb{R}$  for the restriction of the Shannon entropy to the standard simplex. This setting is important because each  $x \in \Delta^{d-1}$  can be interpreted as a probability distribution on  $d$  disjoint events. Correspondingly,  $-E_\Delta(x) = -E(x)$  is the expected efficiency to optimally encode a sample from this distribution. Finally, the relative entropy from  $x$  to  $y$  is the expected loss in coding efficiency if we use the code optimized for  $y$  to encode a sample from  $x$ . Projecting the

gradient of the unrestricted Shannon entropy into the hyperplane of the simplex passing through the origin, we obtain the gradient of the restriction:

$$\nabla E_{\Delta}(x) = \begin{bmatrix} \ln x_1 \\ \ln x_2 \\ \vdots \\ \ln x_d \end{bmatrix} - \frac{1}{d} \sum_{i=1}^d \ln x_i \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{10}$$

Using (4) and (5), we compute the conjugate of  $E_{\Delta}$ , which we state in terms of the barycentric coordinates parametrizing  $\mathbb{R}^{d-1}$ . Specifically, we obtain  $\phi_{\Delta}(x) = \nabla E_{\Delta}(x)$  and

$$E_{\Delta}^*(\phi_{\Delta}(x)) = \langle \nabla E_{\Delta}(x), x \rangle - E_{\Delta}(x) \tag{11}$$

$$= 1 - \frac{1}{d} \sum_{i=1}^d \ln x_i \tag{12}$$

$$= 1 + \ln \sum_{i=1}^d e^{y_i}, \tag{13}$$

in which the  $y_i = \ln x_i - \frac{1}{d} \sum_{i=1}^d \ln x_i$  are the coordinates in conjugate space. Indeed, it is not difficult to verify (13) using  $\ln \sum_{i=1}^d x_i = 0$  for points in the standard simplex.

**Antonelli isometry.** A Bregman divergence gives rise to a path metric in which length is measured by integrating the square root of the divergence. As explained in [6], any divergence that decomposes into a term per coordinate implies an isometry between this path metric and the Euclidean metric. By (9), the relative entropy is an example of such a divergence, and the corresponding path metric is known as the Fisher metric, which plays an important role in statistics and information geometry [3]. Instead of formalizing the recipe for constructing the Fisher metric from the relative entropy, we note that the mapping

$$J(x) = (\sqrt{2x_1}, \sqrt{2x_2}, \dots, \sqrt{2x_d}) \tag{14}$$

is an isometry with Euclidean space, as first observed by Antonelli [33]. By virtue of being an isometry, the distance between points  $x, y \in \mathbb{R}_+^d$  under the Fisher metric satisfies  $\|x - y\|_{\text{Fsh}} = \|J(x) - J(y)\|$ . The path of this length from  $x$  to  $y$  is the preimage of the line segment from  $J(x)$  to  $J(y)$ , which is generally not straight.

Of special interest is the Fisher metric restricted to the standard  $(d - 1)$ -simplex, denoted  $\Delta^{d-1}$ . The mentioned isometry maps  $\Delta^{d-1}$  to  $J(\Delta^{d-1})$ , which is the positive orthant of the sphere with radius  $\sqrt{2}$  and center at the origin in  $\mathbb{R}^d$ . The shortest path between  $x, y \in \Delta^{d-1}$  is thus the preimage of the great-circle arc that connects  $J(x)$  and  $J(y)$  on the sphere. Since this arc is generally longer than the straight line segment connecting  $J(x)$  and  $J(y)$  in  $\mathbb{R}_+^d$ , the distance between  $x$  and  $y$  under the Fisher metric restricted to  $\Delta^{d-1}$  is generally larger than in the unrestricted case.

**Alpha shapes and Wrap complexes.** Two popular shape reconstruction methods based on Delaunay mosaics are the Alpha shapes introduced in [34] and the Wrap complexes first published in [29]. Both extend to generalized discrete Morse functions and therefore to Bregman–Delaunay mosaics and Bregman–Wrap complexes. While working with Bregman divergences, we can construct them using weighted Euclidean Delaunay mosaics, for which there is readily available fast software. For brevity, standard definitions and properties are available in Appendix B. Letting  $D$  be a simplicial complex and  $f: D \rightarrow \mathbb{R}$  a generalized discrete Morse function, the Alpha complex for  $h$  is the sublevel set,

$$\text{Alpha}_h(f) = f^{-1}(-\infty, h], \tag{15}$$

and the Alpha shape is the underlying space of the Alpha complex. In contrast to the Alpha shape, the assumption that  $f$  be a generalized discrete Morse function is essential in the definition of the Wrap complex. Recall that every step of a generalized discrete Morse function



is an interval of simplices in the Hasse diagram. We form the step graph,  $\mathcal{G} = \mathcal{G}_f$ , whose nodes are the steps and whose arcs connect step  $\varphi$  to step  $\psi$  if there are simplices  $P \in \varphi$  and  $Q \in \psi$  with an arc from  $P$  to  $Q$  in the Hasse diagram. By construction,  $f$  is strictly increasing along directed paths in the step graph, which implies that the graph is acyclic.

The lower set of a node  $v$  in  $\mathcal{G}$ , denoted  $\downarrow v$ , is the set of nodes  $\varphi$  for which there are directed paths from  $\varphi$  to  $v$ . Similarly, we write  $\downarrow N = \bigcup_{v \in N} \downarrow v$  for the lower set of a collection of nodes, and  $\bigcup \downarrow N$  for the corresponding collection of simplices. We are particularly interested in the set of singular intervals, and we recall that each such interval contains a critical simplex of  $f$ . We write  $\text{Sg}_f$  for the set of singular intervals, and  $\text{Sg}_f(h) \subseteq \text{Sg}_f$  for the subset whose simplices satisfy  $f(Q) \leq h$ . The Wrap complex for  $h$  is the union of steps in the lower sets of the singular intervals with value at most  $h$ :

$$\text{Wrap}_h(f) = \bigcup \downarrow \text{Sg}_f(h). \tag{16}$$

There are alternative constructions of the Wrap complex. Starting with the Alpha complex for  $h$ , we obtain the Wrap complex for the same value by collapsing all non-singular intervals that can be collapsed. The order of the collapses is not important as all orders produce the same result, namely  $\text{Wrap}_h(f)$ . Symmetrically, we may start with the critical simplices of value at most  $h$  and add the minimal collection of non-singular intervals needed to obtain a simplicial complex. The minimal collection is unique and so is the result,  $\text{Wrap}_h(f)$ . A proof of the equivalence of these three definitions of the Wrap complex is given in Appendix C.

### 3. Mosaics and Algorithms

In this section, we review Bregman–Delaunay and Fisher–Delaunay mosaics as well as their scale-dependent subcomplexes. All mosaics are constructed using software for weighted Delaunay mosaics in Euclidean geometry; all subcomplexes are computed by convex optimization, adapting the method from [12]. We begin with the mosaics in Bregman geometry.

**Bregman–Delaunay mosaics.** Let  $\Omega \subseteq \mathbb{R}^d$  be open and convex, consider a Legendre type function  $F: \Omega \rightarrow \mathbb{R}$ , and let  $U \subseteq \Omega$  be locally finite. Following [12,13], we define the Bregman–Voronoi domain of  $u \in U$ , denoted  $\text{dom}_F(u, \Omega)$ , as the points  $a \in \Omega$  that satisfies  $D_F(u||a) \leq D_F(v||a)$  for all  $v \in U$ . The Bregman–Voronoi tessellation is the collection of such domains, and the Bregman–Delaunay mosaic records all non-empty common intersections:

$$\text{Vor}_F(U, \Omega) = \{\text{dom}_F(u, \Omega) \mid u \in U\}, \tag{17}$$

$$\text{Del}_F(U, \Omega) = \{Q \subseteq U \mid \bigcap_{u \in Q} \text{dom}_F(u, \Omega) \neq \emptyset\}, \tag{18}$$

and we note that the mosaic is isomorphic to the nerve [35,36] of the tessellation. To develop geometric intuition, we observe that  $\text{Vor}_F(U, \Omega)$  can be obtained by growing primal Bregman balls with centers at the points  $u \in U$ . When two such balls meet, they freeze where they touch but keep growing everywhere else. Eventually, each ball covers exactly the corresponding domain. Since the primal balls are not necessarily convex, it is not surprising that the faces shared by the domains are not necessarily straight. Nevertheless, the Delaunay mosaic has a natural straight-line embedding as all its cells are vertical projections of the lower faces of the convex hull of the points  $(u, F(u)) \in \mathbb{R}^{d+1}$ . To see this, we note that each cell of the mosaic corresponds to a dual Bregman ball whose boundary passes through the vertices of the cell, and this ball is the vertical projection of the part of the graph of  $F$  on or below the graph of an affine function.

**Construction.** To construct the mosaic, we assume that  $U \subseteq \Omega$  is in general position, by which we mean that Conditions I and II from Appendix D are satisfied after transforming  $U \subseteq \Omega$  to  $X \subseteq \mathbb{R}^d \times \mathbb{R}$  such that  $\text{Del}_F(U, \Omega)$  is a subcomplex of the weighted Delaunay mosaic of  $X$ . Lifting the points from  $\mathbb{R}^d$  to  $\mathbb{R}^{d+1}$  and projecting the lower boundary of the convex hull back to  $\mathbb{R}^d$ , we obtain the mosaic. We remind the reader that relevant

background information can be found in Appendix B, and define  $\omega(a) = \frac{1}{2}\|a\|^2$ . We formalize this method while stating all steps in terms of weighted points in  $d$  dimensions:

STEP 1: Let  $X \subseteq \mathbb{R}^d \times \mathbb{R}$  be the set of weighted points  $x(u) = (u, 2\omega(u) - 2F(u))$ , with  $u \in U$ .

STEP 2: Compute the weighted Delaunay mosaic of  $X$  in Euclidean geometry, denoted  $\text{Del}(X)$ .

STEP 3: Select  $\text{Del}_F(U, \Omega)$  as the collection of simplices in  $\text{Del}(X)$  whose corresponding weighted Voronoi cells have a non-empty intersection with  $\Omega^*$ .

Indeed, the weighted Delaunay mosaic computed in Step 2 may contain simplices that do not belong to the Delaunay–Bregman mosaic of  $F$ . To implement Step 3, we note that  $\text{Del}_F(U, \Omega)$  is dual to  $\text{Vor}_F(U, \Omega)$ , which is isomorphic to  $\text{Vor}_{F^*}(\phi(U), \Omega^*)$ , and this Bregman–Voronoi tessellation is the weighted Voronoi tessellation of  $X$  restricted to  $\Omega^*$ . This tessellation has convex polyhedral cells and is readily available as the dual of  $\text{Del}(X)$ . Writing  $Y(Q) \subseteq X$  for the points  $x(u)$  with  $u \in Q \subseteq U$  and  $\text{dom}(Y)$  for the weighted Voronoi cell that corresponds to  $Y \in \text{Del}(X)$ , we have

$$\text{Del}_F(U, \Omega) = \{Q \subseteq U \mid \text{dom}(Y(Q)) \cap \Omega^* \neq \emptyset\}. \tag{19}$$

Instead of computing all these intersections, we can collapse  $\text{Del}(X)$  to the desired sub-complex and thus save time by looking only at a subset of the mosaic. We explain how the simplices can be organized to facilitate such a collapse. Recalling that  $\Omega^* \subseteq \mathbb{R}^d$  is open and convex, we introduce the signed distance function,  $\theta: \mathbb{R}^d \rightarrow \mathbb{R}$ , which maps every  $a \in \mathbb{R}^d$  to plus or minus  $r = r(a)$  such that the sphere with center  $a$  and radius  $r$  touches  $\partial\Omega^*$  but does not cross the boundary. Finally,  $\theta(a) = r(a)$  if  $a \notin \Omega^*$  and  $\theta(a) = -r(a)$  if  $a \in \Omega^*$ . Note that  $\Omega^* = \theta^{-1}[-\infty, 0)$  and that  $\Omega_t^* = \theta^{-1}[-\infty, t)$  is open and convex for every  $t$ . Now construct  $\vartheta: \text{Del}(X) \rightarrow \mathbb{R}$  by mapping  $Y \in \text{Del}(X)$  to the maximum  $t \in \mathbb{R}$  for which  $\text{dom}(Y) \cap \Omega_t^* = \emptyset$ . By (19), we obtain  $\text{Del}_F(U, \Omega)$  by removing all simplices  $Y$  with  $\vartheta(Y) \geq 0$ . The crucial observation is that for  $X$  in general position,  $\vartheta$  is a generalized discrete Morse function with a single critical vertex. To see this, we observe that  $\text{Vor}(X)$  decomposes  $\Omega_t^*$  into convex domains for every value  $t$ , which by the Nerve Theorem [22] implies that  $\vartheta^{-1}(-\infty, t)$  is contractible. Removing the simplices in a sequence of decreasing values of  $\vartheta$  thus translates into a sequence of collapses that preserve the homotopy type of the mosaic.

**Rise functions.** To introduce scale into the construction of Bregman–Delaunay mosaics, we generalize the radius function from Euclidean geometry to Bregman geometries, changing the name because size is more conveniently measured by height difference in the  $(d + 1)$ -st coordinate direction as opposed to the radius in  $\mathbb{R}^d$ . Let  $\hat{u} = (u, F(u))$  and  $\bar{u}: \mathbb{R}^d \rightarrow \mathbb{R}$  be the point and affine map that correspond to  $u \in \Omega$ , and let  $v: \mathbb{R}^d \rightarrow \mathbb{R}$  be the upper envelope of the  $\bar{u}$ ,  $u \in U$ . We introduce the rise function,  $q_F: \text{Del}_F(U, \Omega) \rightarrow \mathbb{R}$ , which maps each simplex,  $Q$ , to the minimum difference between  $F^*$  and  $v$  at points in the conjugate Voronoi cell:

$$q_F(Q) = \inf_{a \in \phi(\text{dom}(Q, \Omega))} [F^*(a) - v(a)]. \tag{20}$$

It is the infimum amount we have to lower the graph of  $F^*$  until it intersects the graph of  $v$  at a point vertically above the Voronoi cell in conjugate space. Without going to the conjugate, we can interpret  $q_F(Q)$  in terms of (primal) Voronoi domains and cones of light cast from the  $\hat{u}$  onto the graph, which we raise until the cones clipped to within their Voronoi domains have a point in common. This interpretation motivates the name of the function. Comparing (20) with (A8), we see that the two agree when  $F = \omega$  and  $\Omega = \mathbb{R}^d$ . Indeed, we obtain  $F^* = \omega$  and  $v = \zeta$ . Furthermore,  $\phi(\text{dom}(Q, \Omega)) = \text{dom}(Q, \Omega)$ , and taking the infimum is the same as taking the minimum.

For every  $h \in \mathbb{R}$ , we have a sublevel set,  $\text{Del}_{F,h}(U, \Omega) = \varrho_F^{-1}(-\infty, h]$ , which we refer to as the Bregman–Alpha complex of  $U$  and  $F$  for size  $h$ . For  $h < 0$ , this complex is empty, for  $h = 0$ , it is a set of vertices namely the points in  $U$ , and for sufficiently large positive  $h$ , this complex is  $\text{Del}_F(U, \Omega)$ .

**Information-theoretic interpretations.** We offer a brief information-theoretical interpretation for many of the geometric objects mentioned above in the case of relative entropy.

Let us start with a primal relative entropy ball of radius  $r$  around a finite probability distribution  $c$ . It contains all distributions that can be used to approximate  $c$ , incurring the efficiency loss of at most  $r$  bits. With this, it is easy to see that the Voronoi and Delaunay constructions have intuitive information-theoretical interpretations. For simplicity, let us consider the Čech rise (radius) function, which arises from the first point of a non-empty intersection between a number of primal balls [35,36]. This radius is therefore the smallest number of bits that need to be incurred when approximating the centers of these balls with a single distribution. This value can be interpreted as an information-theoretical quantification of the high-order interactions among these distributions. Of course, the intersection point is exactly the unique best choice for such a distribution.

The remaining objects can be interpreted in a similar way, which we leave for the reader. We stress that the birth and death values in the resulting persistence diagrams also express the loss of coding efficiency in bits—and stress that the diagrams are identical for the Delaunay and Čech case.

**Computation.** We compute the rise function following the intuition based on primal Voronoi domains explained below (20). Equivalently,  $\varrho_F(Q)$  is the minimum amount we have to raise the graph of  $F$  so it has a supporting hyperplane that passes through all points  $\dot{u}$ , with  $u \in Q$ , while all other point  $\dot{u}$ , with  $u \in U$ , lie on or above the hyperplane.

To turn this intuition into an algorithm, we consider the affine hull of  $Q$  and write  $\bar{v}: \text{aff } Q \rightarrow \mathbb{R}$  for the affine function that satisfies  $\bar{v}(u) = F(u)$  for all  $u \in Q$ . Let  $H: \text{aff } Q \cap \Omega \rightarrow \mathbb{R}$  measure the difference:  $H(a) = F(a) - \bar{v}(a)$ . Since  $F$  is of Legendre type, so is  $H$ . We are interested in the infimum of  $H$ , which either occurs at a point in  $\text{aff } Q \cap \Omega$  or at the limit of a divergent sequence. We therefore introduce a numerical routine that returns both, the infimum and the point where it occurs:

```

1  INFSIZE (function  $F$ , simplex  $Q$ ):
2       $(a_Q, h_Q) = (\text{arginf } H, \inf H)$ ;
3      return  $(a_Q, h_Q)$ .

```

Note that the dual Bregman ball centered at  $a_Q \in \text{aff } Q \cap \Omega$  of size  $h_Q$  contains  $Q$  in its boundary, and it may or may not contain points of  $U \setminus Q$  in its interior. If it does not, then  $\varrho_F(Q) = h_Q$ , otherwise,  $\varrho_F(Q)$  is the minimum function value of the proper cofaces of  $Q$ . To express this more formally, we write  $\text{coFacets}(Q)$  for the collection of simplices  $R \in \text{Del}(X)$  with  $Q \subseteq R$  and  $\#R = \#Q + 1$ . Since  $Q$  gets its value either directly or from a coface, it is opportune to compute the rise function in the order of decreasing dimension:

```

1  for  $p = d$  downto 0 do
2      forall  $p$ -simplices  $Q \in \text{Del}_F(U, \Omega)$  do
3           $(a_Q, h_Q) = \text{INFSIZE}(F, Q)$ ;
4          if  $B_F^*(a_Q, h_Q) \cap [U \setminus Q] = \emptyset$ 
5              then  $\varrho_F(Q) = h_Q$ 
6          else  $\varrho_F(Q) = \min_{R \in \text{coFacets}(Q)} \varrho_F(R)$ .

```

Note that this algorithm assigns a value to every simplex in  $\text{Del}_F(U, \Omega)$ . Indeed, the simplices in  $\text{Del}(X)$  that are not in  $\text{Del}_F(U, \Omega)$  have been culled in Step 3, as explained above.

**Fisher metric.** In addition to the Bregman divergences, we consider Delaunay mosaics under the Fisher metric. To construct them, we recall that the mapping  $j: \mathbb{R}_+^d \rightarrow \mathbb{R}_+^d$  defined by  $j(x) = (\sqrt{2x_1}, \sqrt{2x_2}, \dots, \sqrt{2x_d})$  is an isometry between the Fisher metric and the Euclidean metric. This suggests the following algorithm:



STEP 1: Compute the Delaunay mosaic of  $j(U)$  in Euclidean space.

STEP 2: Remove the simplices from  $\text{Del}(j(U))$  whose dual Voronoi cells have an empty intersection with  $\mathbb{R}_+^d$ .

STEP 3: Draw the resulting complex by mapping each point  $j(u)$  to the original point  $u \in U \subseteq \mathbb{R}_+^d$ .

The rise function in Euclidean geometry maps every simplex  $j(Q) \in \text{Del}(j(U))$  to the squared radius of the smallest empty circumsphere of  $j(Q)$ . By isometry, the preimage of this Euclidean sphere is the smallest empty circumsphere of  $Q$  under the Fisher metric, and the squared radius is the same. We thus obtain the rise function on the Fisher–Delaunay mosaic by copying the values of the rise function on the Delaunay mosaic in Euclidean geometry.

The construction of the mosaic for the Fisher metric restricted to the standard simplex,  $\Delta^{d-1}$ , is only slightly more complicated. As mentioned in Section 2, the isometry maps  $\Delta^{d-1}$  to  $\sqrt{2}\mathbb{S}_+^{d-1}$ , which is our notation for the positive orthant of the sphere with radius  $\sqrt{2}$  centered at the origin in  $\mathbb{R}^d$ . The distance between points  $u, v \in \Delta^{d-1}$  under the Fisher metric thus equals the Euclidean length of the great-circle arc connecting  $j(u), j(v) \in \sqrt{2}\mathbb{S}_+^{d-1}$ . The Delaunay mosaic of  $j(U)$  under the geodesic distance can be obtained by constructing the convex hull of  $j(U) \cup \{0\}$  in  $\mathbb{R}^d$  and centrally projecting all faces not incident to 0 onto the sphere. As before, we cull simplices whose dual Voronoi cells have an empty intersection with the positive orthant of the sphere, and we draw the mosaic in  $\Delta^{d-1}$  by mapping the vertices back to the original points. Furthermore, the rise functions of the mosaics in  $\sqrt{2}\mathbb{S}_+^{d-1}$  and in  $\Delta^{d-1}$  are the same. Note, however, that the geodesic radius is the arc-sine of and therefore slightly larger than the straight Euclidean radius in  $\mathbb{R}^d$ .

#### 4. Computational Experiments

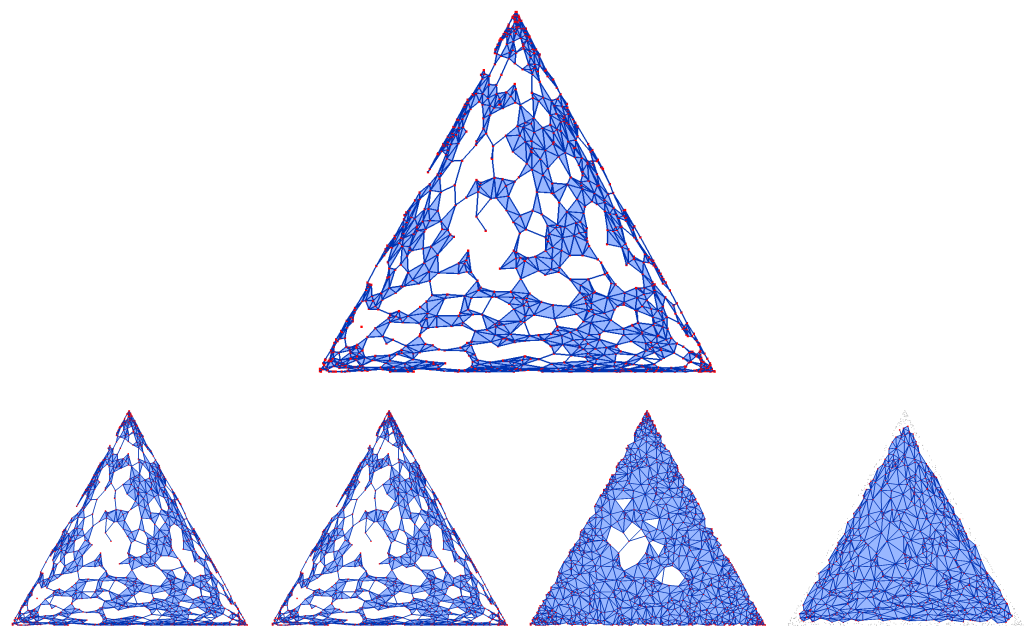
We illustrate the Bregman–Alpha and Bregman–Wrap complexes while comparing them to the conjugate, the Fisher, and the Euclidean constructions.

**Example in positive quadrant.** Let  $X$  be a set of 1000 points uniformly distributed according to the Fisher metric in  $(0, 2]^2 \subseteq \mathbb{R}_+^2$ . To sample  $X$ , we use the isometry,  $j: \mathbb{R}_+^2 \rightarrow \mathbb{R}_+^2$ , between the Fisher and the Euclidean metric mentioned in Section 2. Specifically, we sample 1000 points uniformly at random according to the Euclidean metric in  $(0, 2]^2$ , and we map each point with coordinates  $x_1, x_2$  to  $j^{-1}(x_1, x_2) = \frac{1}{2}(x_1^2, x_2^2)$ , which is again a point in  $(0, 2]^2$ . To compute the Delaunay mosaic in Fisher geometry, we construct the (Euclidean) Delaunay mosaic of  $j(X)$  and draw this mosaic with the vertices at the points in  $X$ . Recall however that the domain is  $\Omega = \mathbb{R}_+^d$  and not  $\mathbb{R}^d$ . A simplex whose corresponding Voronoi cell has an empty intersection with the positive orthant thus does not belong to the mosaic, which is restricted to  $\Omega$ . We identify these simplices and remove them from the Delaunay mosaic as described in Section 3.

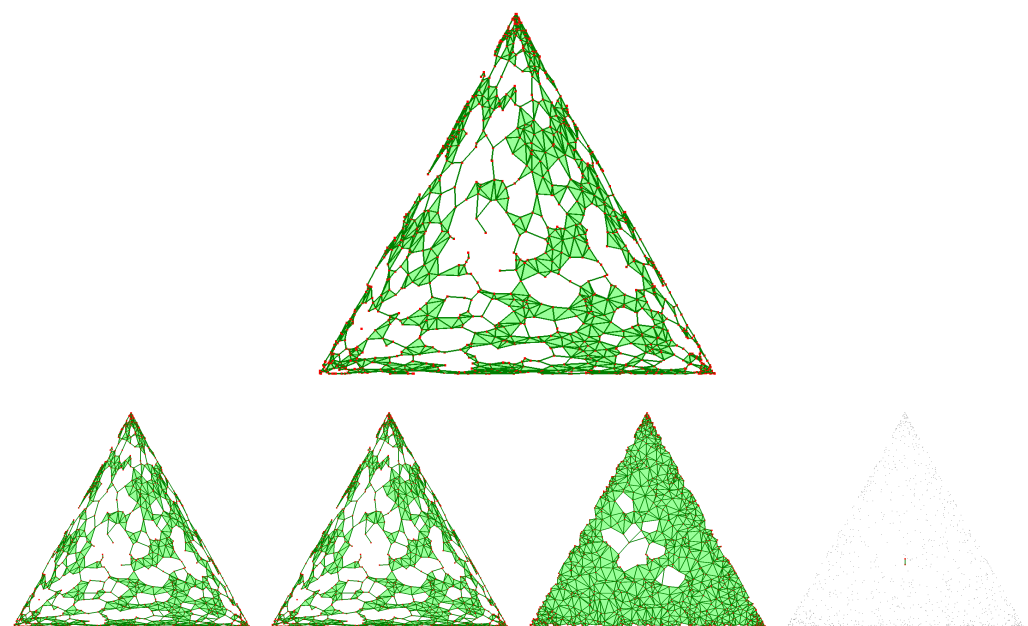
**Example in standard triangle.** Motivated by our interest in information-theoretic applications, we repeat the above experiment within the standard triangle,  $\Delta^2$ , which consists of all points  $(x_1, x_2, x_3) \in \mathbb{R}_+^3$  that satisfy  $x_1 + x_2 + x_3 = 1$ . Every point in  $\Delta^2$  can be interpreted as a probability distribution on three disjoint events, which is indeed the most relevant scenario for the application of the relative entropy. To sample a set  $Y$  of 1000 points uniformly at random according to the Fisher metric in  $\Delta^2$ , we use again  $j$ , now restricted to  $\Delta^2$ , whose image is the positive orthant of the sphere with radius  $\sqrt{2}$  centered at the origin of  $\mathbb{R}^3$ . Sampling 1000 points uniformly at random according to the geodesic distance on the sphere, we take the convex hull of  $j(Y) \cup \{0\}$  and obtain the mosaic by mapping the vertices to the points in  $Y = j^{-1}(j(Y))$ . Before drawing the faces in  $\Delta^2$ , we remove 0 and all incident faces, as well as all faces whose corresponding Voronoi cells have an empty intersection with  $\mathbb{R}_+^2$ .

Recall that the squared Fisher metric matches the relative entropy in the infinitesimal regime, which explains the random appearance of the reconstruction in Figure 1, for which we set the threshold to 0.0025. The reconstruction in Shannon geometry is similar to those in

conjugate Shannon geometry in Figure 1a and in Fisher geometry in Figure 1b. To interpret the reconstruction in Figure 1d, we observe that the difference between the Shannon entropy and the squared Euclidean norm has a minimum at the center and no other critical points in the interior of the triangular domain. Accordingly, the reconstruction removes simplices near the corners and the three sides first. More drastically, the Bregman–Wrap complex for the same data removes all simplices except for a single critical edge near the center (see Figure 2d).



(a) Conjugate Shannon. (b) Fisher. (c) Euclidean. (d) Weighted Euclidean.  
**Figure 1.** (Top) The Bregman–Alpha complex in Shannon geometry of a set  $Y$  of 1000 random points in  $\Delta^2$  with threshold  $h = 0.0025$ . (Bottom row) The reconstructions in four different geometries.



(a) Conjugate Shannon. (b) Fisher. (c) Euclidean. (d) Weighted Euclidean.  
**Figure 2.** (Top) The Bregman–Wrap complex in Shannon geometry of the same points and threshold as in Figure 1. (Bottom row) The reconstructions in four different geometries.

### 5. Quantification of Difference

We take a data-centric approach to quantifying the differences between the geometries. Given a common domain,  $\Omega$ , and a finite set of points,  $U \subseteq \Omega$ , we compare the corresponding mosaics and rise functions.

**Mosaics.** The Delaunay mosaics of  $U$  depend on the local shape of the balls defined by the metric or the divergence. Letting  $D$  and  $E$  be two Delaunay mosaics with vertex sets  $U$ , we compare them by counting the common cells:

$$J(D, E) = 1 - \frac{\#(D \cap E)}{\#D + \#E - \#(D \cap E)}, \tag{21}$$

which is sometimes referred to as the *Jaccard distance* between the two sets. It is normalized so that  $J = 0$  iff  $D = E$  and  $J = 1$  iff  $D$  and  $E$  share no cells at all. In our application, the two mosaics share all vertices, so  $J$  is necessarily strictly smaller than 1. To apply this measure to the constructions in Section 4, we write  $D_0, D_1, D_2, D_3, D_4$  for the mosaics in Figure A1, and we write  $E_0, E_1, E_2, E_3, E_4$  for the mosaics in Figure 3. All mosaics are different, except for  $D_0 = D_4$  and  $E_0 = E_4$ . The Jaccard distances are given in Table 1.

**Table 1.** The Jaccard distances between the Delaunay mosaics in Shannon, conjugate Shannon, Fisher, Euclidean, and weighted Euclidean geometries for points in the positive quadrant on the top and in the standard triangle on the bottom.

$J$	$D_0$	$D_1$	$D_2$	$D_3$	$D_4$
$D_0$	0.00	0.06	0.04	0.48	0.00
$D_1$		0.00	0.02	0.47	0.06
$D_2$			0.00	0.47	0.04
$D_3$				0.00	0.48
$D_4$					0.00
	$E_0$	$E_1$	$E_2$	$E_3$	$E_4$
$E_0$	0.00	0.10	0.06	0.52	0.00
$E_1$		0.00	0.04	0.51	0.10
$E_2$			0.00	0.51	0.06
$E_3$				0.00	0.52
$E_4$					0.00

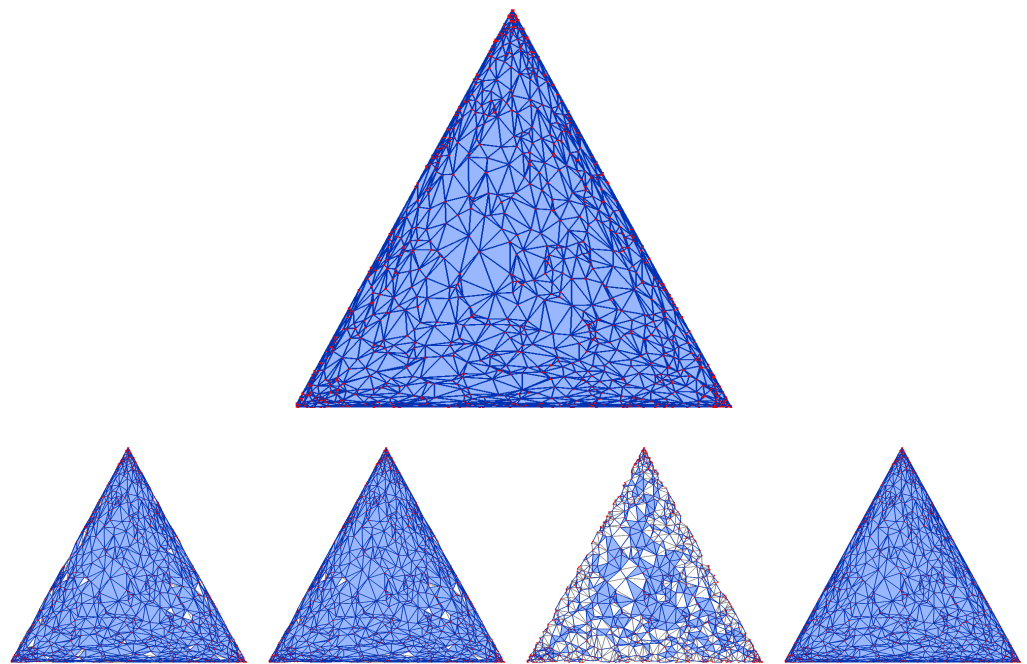
We see that the mosaics in conjugate Shannon geometry and in Fisher geometry are most similar to each other and less similar to the mosaic in Shannon geometry. The mosaic in Euclidean geometry is most dissimilar to the others (see Figure 3 and related Figure A1 in Appendix A for visual confirmation).

**Rise functions.** Different rise functions on the same mosaic can be compared by counting the inversions, which are the pairs of cells whose orderings are different under the two functions. Recall that  $D_0 = D_4$  and  $E_0 = E_4$ , let  $d_0: D_0 \rightarrow \mathbb{R}$  and  $e_0: E_0 \rightarrow \mathbb{R}$  be the rise functions in Shannon geometry, and let  $d_4: D_4 \rightarrow \mathbb{R}$  and  $e_4: E_4 \rightarrow \mathbb{R}$  be the rise functions in weighted Euclidean geometry. The normalized number of inversions are

$$I(d_0, d_4) = 0.476, \tag{22}$$

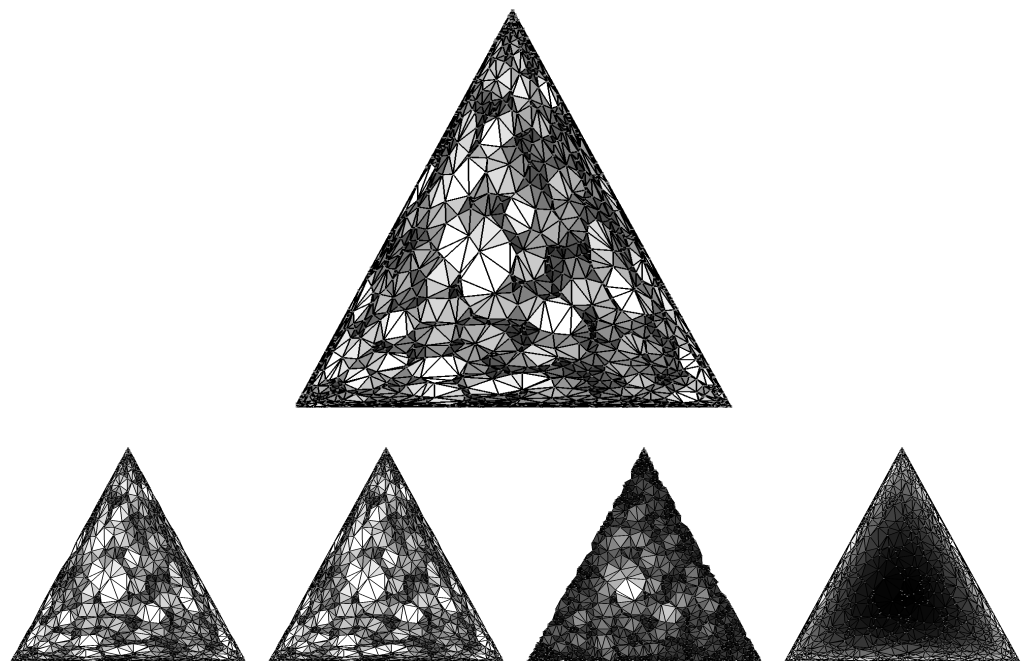
$$I(e_0, e_4) = 0.467. \tag{23}$$

In words, slightly fewer than half the pairs are inversions, both for  $d_0, d_4$  and for  $e_0, e_4$ . This is plausible because  $d_4$  orders the cells along the diagonal while  $d_0$  preserves the random character of the point sample (see Figure A2d). Similarly,  $e_4$  orders the cells radially, from the center of the standard triangle to its periphery, while  $e_0$  preserves again the random character of the sample (see Figure 4d).



(a) Conjugate Shannon. (b) Fisher. (c) Euclidean. (d) Weighted Euclidean.

**Figure 3.** (Top) The Bregman–Delaunay mosaic in Shannon geometry for the same set of points as used in Figures 1 and 2. (Bottom row) Four Delaunay mosaics whose triangles and edges are colored depending on whether or not they belong to the Shannon–Delaunay mosaic.



(a) Conjugate Shannon. (b) Fisher. (c) Euclidean. (d) Weighted Euclidean.

**Figure 4.** (Top) the color-coded Bregman–Delaunay mosaic in Shannon geometry of the same set of points as in Figure 3. (Bottom row) The color-coded Delaunay mosaics.

We can compare the rise functions also visually, by color-coding the 2-dimensional cells, and this works even if the mosaics are different. Specifically, we shade the triangles by mapping small to large rise function values onto dark to light colors. In Figure A2a,b, this leads to randomly mixed dark and light triangles, while in Figure A2c,d, there are clear but opposing gradients parallel to the diagonal. Similarly, in Figure 4c, we see the

rise function decreasing from the center to the boundary of the standard triangle, and in Figure 4d we see it increasing from the center to the boundary. In addition, we compare general rise functions by computing their persistence diagrams (see [36]). Writing  $\text{Dgm}(d)$  for the persistence diagram of function  $d$ , we quantify the difference with the bottleneck between the diagrams:

$$B(d, e) = W_\infty(\text{Dgm}(d), \text{Dgm}(e)). \tag{24}$$

As explained in [36], the bottleneck distance is 1-Lipschitz, that is:  $B(d, e) \leq \|d - e\|_\infty$ , but  $d \neq e$  does not necessarily imply  $B(d, e) \neq 0$ . The bottleneck distances between the  $d_i: D_i \rightarrow \mathbb{R}$  and the  $e_i: E_i \rightarrow \mathbb{R}$  are given in Table 2.

**Table 2.** The bottleneck distances between the persistence diagrams of the rise functions on the Delaunay mosaics in Shannon, conjugate Shannon, Fisher, Euclidean, and weighted Euclidean geometries for points in the positive orthant on the top and points in the standard triangle on the bottom.

$B$	$d_0$	$d_1$	$d_2$	$d_3$	$d_4$
$d_0$	0.0000	0.0028	0.0004	0.0126	0.0048
$d_1$		0.0000	0.0028	0.0126	0.0048
$d_2$			0.0000	0.0126	0.0048
$d_3$				0.0000	0.0126
$d_4$					0.0000
	$e_0$	$e_1$	$e_2$	$e_3$	$e_4$
$e_0$	0.0000	0.0006	0.0003	0.0031	0.0035
$e_1$		0.0000	0.0003	0.0030	0.0034
$e_2$			0.0000	0.0030	0.0034
$e_3$				0.0000	0.0023
$e_4$					0.0000

In part, this comparison agrees with the Jaccard distances between the mosaics given in Table 1. The most obvious disagreements are for  $d_0, d_4$  and for  $e_0, e_4$ , in which quite different functions are defined on identical mosaics.

### 6. Discussion

We extended two popular computational geometric-topological methods within the framework of discrete Morse functions and showed how this generalizes the methods to data in Bregman and Fisher geometries. Importantly, this allowed us to produce a robust implementation without the need to develop customized low-level software. We hope that this result will provide extra motivation for the development of geometric software for high-dimensional Euclidean space, as such software can be reused to handle data in information space. In particular, efficient computation of skeleta of weighted Delaunay mosaics would be relevant for topological analysis of high-dimensional data.

Turning the table, we use these generalized methods to compare different geometries experimentally. Our experimental approach to studying geometries is a first step towards an intuitive understanding of these often counter-intuitive geometries, as well as the high-order interactions occurring in them.

We reiterate one reason why understanding such geometries is important: they underpin many modern data science methods. It is, for example, a prerequisite for explaining the surprisingly good performance of many deep learning models (that output predictions as discrete probability distributions). More broadly, the space of discrete probability distributions is a basic object in mathematics, but is not well understood. We provide new tools for broadening the understanding of such spaces.

Perhaps the main observation coming from our experiments is the following. As we know, the pairwise distance in the Kullback–Leibler geometry and the Euclidean or Fisher one can be very different. Indeed, on the standard simplex, the former can approach infinity



while it is bounded in the latter. Still, our experiments show that the geometric-topological structures—at least in low dimensions—do not typically show large discrepancies between the geometries. Is this also true in higher dimensions? How does this generalize to other Bregman divergences and the corresponding generalized Fisher distances [6]? To what extent can the Fisher space replace the Shannon space in various applications?

Several open questions about experimental understanding of geometric spaces arise:

- How do these results generalize to other notions of entropy, such as the Burg entropy [37]? The resulting Bregman Itakura–Saito divergence [38] is used to compare speech samples, but little is known about the resulting geometry and how it compares with other geometries.
- Is there a deeper reason for the similarity of the studied geometries? For example: are there Pinsker-type inequalities [39,40] between the pairwise dissimilarities (as well as other Bregman divergences)?
- Can the sensitivity of Delaunay mosaics to the dissimilarity be quantified probabilistically, as the expected Jaccard distance for random point processes?
- Persistence has been used before to compare metric spaces [41], and it would be interesting to know whether there are deeper connections to our work.
- Can the Fisher metric be extended beyond the positive part of the sphere? What is the geometry of the preimage of the Antonelli map in this case?

We finish with a technical question concerning the Delaunay mosaics in Fisher geometry: is the drawing we obtain by mapping the vertices to the corresponding points and connecting them with straight edges, flat triangles, etc., necessarily a geometric realization of the mosaic?

**Author Contributions:** Conceptualization, H.E., K.Ö. and H.W.; methodology, H.E., K.Ö. and H.W.; software, K.Ö. and H.W.; writing, H.E., K.Ö. and H.W.; visualization, K.Ö.; supervision, H.E.; funding acquisition, H.E. and H.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme, grant no. 788183, from the Wittgenstein Prize, Austrian Science Fund (FWF), grant no. Z 342-N31, the DFG Collaborative Research Center TRR 109, ‘Discretization in Geometry and Dynamics’, Austrian Science Fund (FWF), grant no. I 02979-N35, the Austrian-Slovenian project on ‘Intrinsic Persistent Homology, Austrian Science Fund (FWF), grant no. I 4245-N35 and the 2022 Google Research Scholar Program grant ‘Algorithms for Topological Analysis of Neural Networks’.

**Data Availability Statement:** The software and data used for this study is available at: [https://git.ista.ac.at/katharina.oelsboeck/wrap\\_2\\_3-public/](https://git.ista.ac.at/katharina.oelsboeck/wrap_2_3-public/) (accessed on 19 July 2024).

**Acknowledgments:** We thank Anton Nikitenko for first observing that the Wrap complex can be characterized as stated in Claim (ii) of the Wrap Complex Lemma, and Ondrej Draganov for correcting a critical mistake in one of our formulas in Section 2.

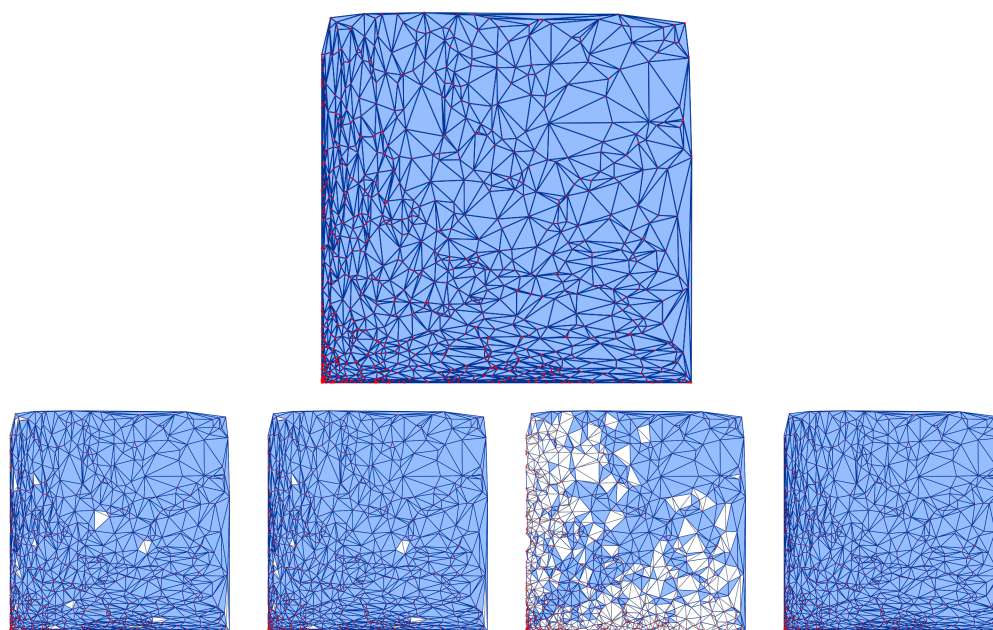
**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Extra Visualizations

In this section, we provide visualizations and discussion for the positive orthant case. We recall that the form of the Kullback–Leibler divergence we use works also in this case (and not only on the standard simplex). Figure A3 top displays the Bregman–Alpha complex in Shannon geometry for threshold 0.004. Infinitesimally, the relative entropy agrees with the squared Fisher metric, so the uniform distribution of the points translates into a fairly uniform arrangement of random holes in the complex. The closer we get to the left or the lower side of the square, the denser the points get and the more anisotropically aligned with the sides the edges and the triangles get.

For comparison, the bottom row in Figure A3 shows the Bregman–Alpha complex in conjugate Shannon geometry, in Fisher geometry, in Euclidean geometry, and in weighted

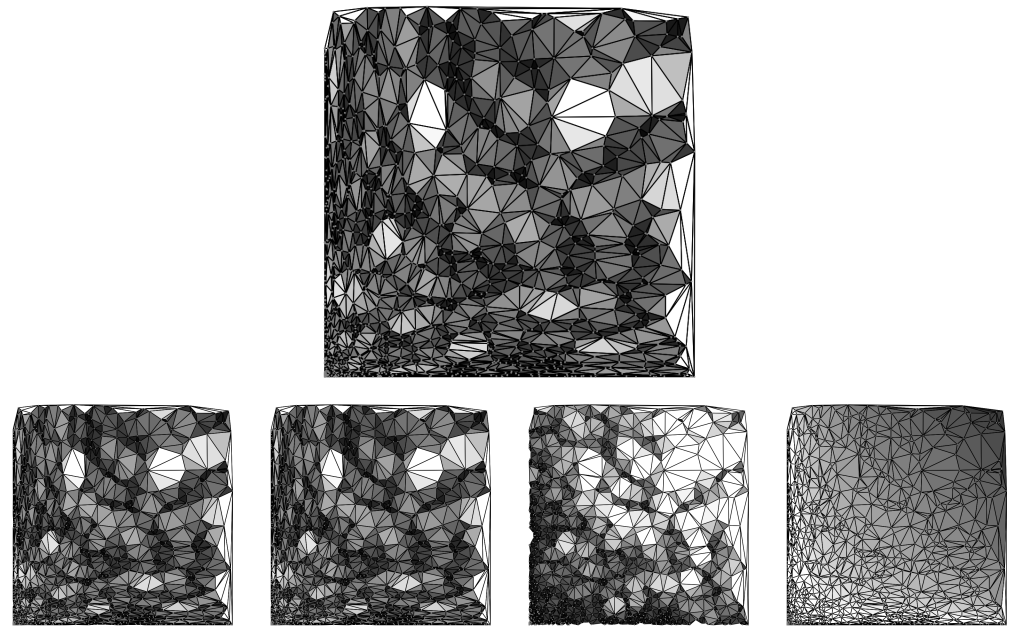
Euclidean geometry. The primal and the dual balls behave similarly, which explains the similarity of the complexes in Figure A3 top and (a). It should however be mentioned that the underlying triangulation in Figure A3a occasionally folds, which is caused by moving the vertices from the conjugate points (for which we have a straight-line embedding) to the original points. Not surprisingly, there is also a striking similarity to the reconstruction in Fisher geometry Figure A3b. The Bregman–Alpha complex in Euclidean geometry Figure A3c is just the usual Alpha complex of the points. It clearly shows that the density decreases along the diagonal. The complex in Figure A3d mixes aspects of Shannon and Euclidean geometry. In particular, it reuses the mosaic in Figure A3 top and assigns weights to the points such that this triangulation is the weighted Delaunay mosaic of the weighted points in Euclidean geometry. The corresponding rise function reflects the difference between the Shannon entropy and the squared Euclidean norm. Indeed, the rise function increases along the diagonal, which explains why the reconstructed complex is almost the entire mosaic, with cells along the left and bottom sides of the square domain missing.



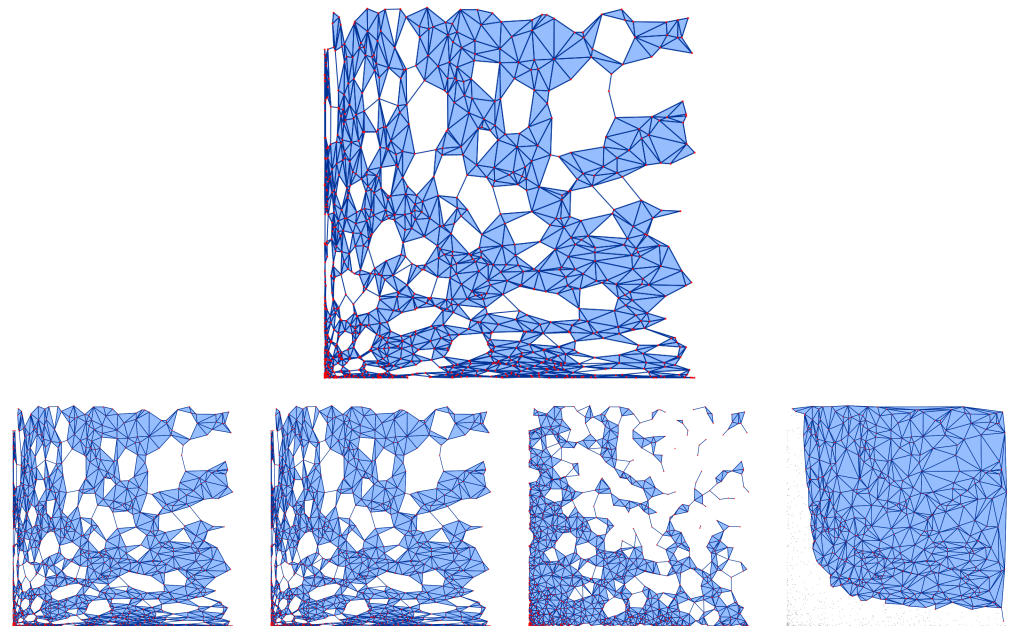
(a) Conjugate Shannon. (b) Fisher. (c) Euclidean. (d) Weighted Euclidean.

**Figure A1.** (Top) The Bregman–Delaunay mosaic in Shannon geometry for the same set of points as used in Figures A3 and A4. (Bottom row) Four Delaunay mosaics whose triangles and edges are colored depending on whether or not they belong to the Shannon–Delaunay mosaic.

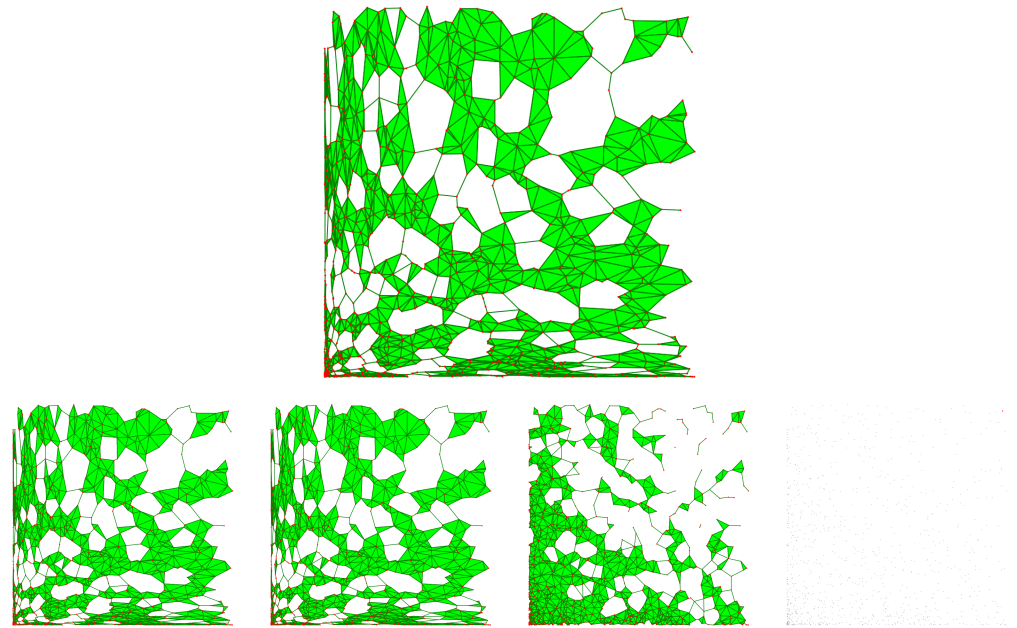
We see very similar reconstructions in Figure A4, which shows the Bregman–Wrap complexes for the same set of points and the same threshold. By construction, each Wrap complex is a homotopy equivalent subcomplex of the corresponding Alpha complex. The biggest difference occurs in weighted Euclidean geometry, in which we reuse the mosaic in Shannon geometry but filter with the rise function obtained from the squared Euclidean norm. The corresponding Bregman–Wrap complex consists of a single vertex near the upper right corner of the square domain (see Figure A4d). This reconstruction reflects the simple relation between the Shannon entropy and the halved squared Euclidean norm:  $\omega(x) - E(x)$  is monotonically increasing from left to right and from bottom to top. This translates into a discrete gradient that introduces a flow with a single critical cell, namely, the vertex near the upper right corner.



(a) Conjugate Shannon. (b) Fisher. (c) Euclidean. (d) Weighted Euclidean.  
**Figure A2.** (Top) The color-coded Bregman–Delaunay mosaic in Shannon geometry. The set of points is the same as in Figure A1. (Bottom row) The color-coded Delaunay mosaics for the same set  $X$ .



(a) Conjugate Shannon. (b) Fisher. (c) Euclidean. (d) Weighted Euclidean.  
**Figure A3.** (Top) The Bregman–Alpha complex in Shannon geometry of a set  $X$  of 1000 points uniformly distributed according to the Fisher metric in  $(0, 2]^2$  and a threshold  $h = 0.004$ . (Bottom row) The reconstructions in four different geometries.



(a) Conjugate Shannon. (b) Fisher. (c) Euclidean. (d) Weighted Euclidean.  
**Figure A4.** (Top) The Bregman–Wrap complex in Shannon geometry of the same points and the same threshold as in Figure A3. (Bottom row) The reconstructions in four different geometries.

### Appendix B. Background on Delaunay Mosaics and Related Topics

We recall standard definitions related to Delaunay mosaics, the corresponding liftings and projections, as well as discrete Morse theory.

**Delaunay mosaics.** In this paper, the ability to assign real weights to points is essential, so we go straight to the weighted generalizations of the Voronoi tessellation and the Delaunay mosaic. A weighted point is a pair  $x = (\text{pt}(x), \text{wt}(x)) \in \mathbb{R}^d \times \mathbb{R}$ , in which  $\text{pt}(x)$  is its *location* and  $\text{wt}(x)$  is its weight. The power distance of  $a \in \mathbb{R}^d$  from  $x$  is  $\pi_x(a) = \|\text{pt}(x) - a\|^2 - \text{wt}(x)$ . It is common to think of the weighted point as a ball with center  $\text{pt}(x)$  and squared radius  $\text{wt}(x)$ . With this interpretation,  $\pi_x(a)$  is negative inside, zero on the boundary, and positive outside the ball. Given a locally finite set of weighted points,  $X \subseteq \mathbb{R}^d \times \mathbb{R}$ , the (weighted) Voronoi domain of  $x \in X$  consists of all points  $a$  for which  $x$  minimizes the power distance, and the (weighted) Voronoi tessellation of  $X$  is the collection of such domains:

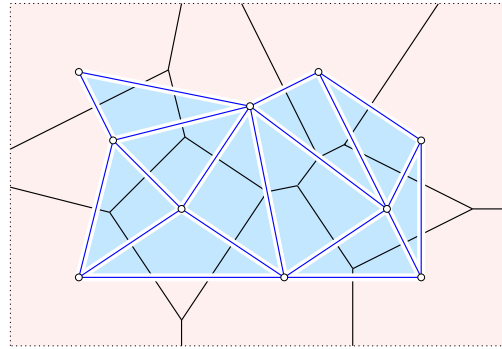
$$\text{dom}(x) = \{a \in \mathbb{R}^d \mid \pi_x(a) \leq \pi_y(a), \forall y \in X\}, \tag{A1}$$

$$\text{Vor}(X) = \{\text{dom}(x) \mid x \in X\}. \tag{A2}$$

A (weighted) Voronoi cell is the common intersection of Voronoi domains, and we write  $\text{dom}(Q) = \bigcap_{x \in Q} \text{dom}(x)$ . Note that the affine hull of  $\text{dom}(Q)$  contains a unique point, denoted  $a_Q$ , that minimizes the power distance to the weighted points in  $Q$ . Indeed,  $a_Q$  is at the intersection of the affine hull of  $\text{dom}(Q)$  and the affine hull of the locations  $\text{pt}(x)$ ,  $x \in Q$ . Let  $\#Q$  be the cardinality of  $Q$ . We are primarily interested in the generic case, when every non-empty Voronoi cell,  $\text{dom}(Q)$ , satisfies the following two general position conditions:

- I. the dimension of  $\text{dom}(Q)$  is  $d + 1 - \#Q$ ,
- II.  $a_Q$  does not belong to the boundary of  $\text{dom}(Q)$ .

By Condition I,  $\text{dom}(Q) = \emptyset$  whenever  $\#Q > d + 1$ . Condition I also implies that every non-empty Voronoi cell is the intersection of a unique collection of Voronoi domains.



**Figure A5.** The Voronoi tessellation restricted to the open rectangular region and its dual restricted Delaunay mosaic.

The (weighted) Delaunay mosaic is the collection of polytopes spanned by subsets of  $X$  that define non-empty Voronoi cells. It is convenient to identify such a subset,  $Q$ , with the polytope it spans, which is the convex hull of the locations of the weighted points in  $Q$ . In the assumed generic case, all polytopes are simplices, and the Delaunay mosaic is a simplicial complex geometrically realized in  $\mathbb{R}^d$ , which we denote as  $\text{Del}(X)$ . Most of the time, we restrict our attention to an open convex region,  $\Omega \subseteq \mathbb{R}^d$ , we assume  $X \subseteq \Omega \times \mathbb{R}$ , and we write  $\text{dom}(Q, \Omega) = \text{dom}(Q) \cap \Omega$ . Correspondingly, the restricted Voronoi tessellation and the restricted Delaunay mosaic are

$$\text{Vor}(X, \Omega) = \{\text{dom}(x, \Omega) \mid x \in X\}, \tag{A3}$$

$$\text{Del}(X, \Omega) = \{Q \subseteq X \mid \text{dom}(Q, \Omega) \neq \emptyset\} \tag{A4}$$

(see Figure A5).

**Lifting and projecting.** The Voronoi tessellation and the Delaunay mosaic can both be constructed as the projection of the boundary complexes of convex polyhedra in  $\mathbb{R}^{d+1}$ . To explain this, recall that  $\omega(a) = \frac{1}{2}\|a\|^2$  and map every weighted point,  $x$ , to the point  $\hat{x} \in \mathbb{R}^{d+1}$  and to the affine map  $\bar{x}: \mathbb{R}^d \rightarrow \mathbb{R}$  defined by

$$\hat{x} = (\text{pt}(x), \omega(\text{pt}(x)) - \frac{1}{2}\text{wt}(x)), \tag{A5}$$

$$\bar{x}(a) = \omega(\text{pt}(x)) + \langle \text{pt}(x), a - \text{pt}(x) \rangle + \frac{1}{2}\text{wt}(x) \tag{A6}$$

$$= \langle \text{pt}(x), a \rangle - \frac{1}{2}\|\text{pt}(x)\|^2 + \frac{1}{2}\text{wt}(x). \tag{A7}$$

The map is chosen so that the solution to  $\omega(a) - \bar{x}(a) = 0$  is the sphere with center  $\text{pt}(x)$  and squared radius  $\text{wt}(x)$ . The point is chosen so that a point on the graph of  $\omega$  lies on or below the graph of  $\bar{x}$  iff this point is visible from  $\hat{x}$ , by which we mean that the entire line segment connecting  $\hat{x}$  with this point lies below the graph of  $\omega$ .

Let  $\zeta: \mathbb{R}^d \rightarrow \mathbb{R}$  be the pointwise maximum of the affine maps,  $\zeta(a) = \max_{x \in X} \bar{x}(a)$ , and note that it is piecewise linear and convex. As observed already by Georges Voronoi [42], the vertical projection of its linear pieces gives the Voronoi tessellation of  $X$  in  $\mathbb{R}^d$ . To obtain a similar construction of the Delaunay mosaic, we take the convex hull of the points  $\hat{x} \in \mathbb{R}^{d+1}$ . We call a hyperplane that touches the polytope without intersecting its interior a support plane, and the intersection of the polytope with a support plane a face of the polytope. For points in general position, all faces are simplices. A lower face is the intersection of the polytope with a non-vertical support plane such that the polytope lies above the hyperplane. In analogy to the relation observed by Voronoi, the vertical projection of the lower faces of the convex hull gives the Delaunay mosaic of  $X$  in  $\mathbb{R}^d$ .

The interpretations of the Voronoi tessellation and the Delaunay mosaic as projections of boundary complexes of convex polyhedra provide geometrically intuitive interpretations of a function that plays a crucial role in this paper. Recall that each simplex,  $Q \in \text{Del}(X)$ , corresponds to a Voronoi cell,  $\text{dom}(Q)$ . The radius function, or more precisely the half-



squared radius,  $\varrho: \text{Del}(X) \rightarrow \mathbb{R}$ , maps  $Q$  to the minimum difference between  $\omega$  and  $\zeta$  at points in the Voronoi cell:

$$\varrho(Q) = \min_{a \in \text{dom}(Q)} [\omega(a) - \zeta(a)]. \tag{A8}$$

In words,  $\varrho(Q)$  is the amount we have to lower the graph of  $\omega$  until it intersects the graph of  $\zeta$  at a point vertically above  $\text{dom}(Q)$ . The function value is also the minimax difference between  $\omega$  and any affine map that satisfies  $\bar{y}(\text{pt}(x)) \leq \omega(\text{pt}(x)) - \frac{1}{2}\text{wt}(x)$  for all  $x \in X$  and with equality for all  $x \in Q$ . Specifically, we minimize the maximum  $\bar{y}(a) - \omega(a)$ , in which the maximization is over all  $a \in \mathbb{R}^d$ , and the minimization is over all affine maps,  $\bar{y}: \mathbb{R}^d \rightarrow \mathbb{R}$ , that satisfy the conditions stated above.

**Discrete Morse theory.** Assuming a general position, the radius function on the Delaunay mosaic enjoys structural properties, which we now formalize. Let  $K$  be a simplicial complex and  $P, Q \in K$  two simplices. For a monotonic function,  $f: K \rightarrow \mathbb{R}$ ,  $P \subseteq Q$  implies  $f(P) \leq f(Q)$ . The Hasse diagram of  $K$  is the directed graph whose nodes are the simplices and whose arcs are the codimension 1 face relations: every arc ends at a  $p$ -simplex and starts at a  $(p - 1)$ -dimensional face of this simplex. By construction, the values of a monotonic function are non-decreasing along directed paths in the Hasse diagram. A level set of  $f$  is a maximal collection of simplices with shared function value,  $f^{-1}(r) \subseteq K$ , and we call a maximal connected subset of a level set a step. For simplices  $P \subseteq R$  in  $K$ , call  $\psi = \{Q \in K \mid P \subseteq Q \subseteq R\}$  an interval,  $P = \text{lb}(\psi)$  its lower bound,  $R = \text{ub}(\psi)$  its upper bound, and note that  $\#\psi = 2^{\#R - \#P}$ . According to an inessential modification of the original formulation by Forman [26],  $f$  is a discrete Morse function if every step is an interval of size 1 or 2. A slightly weaker condition was introduced by Freij [27], calling  $f$  a generalized discrete Morse function if every step is an interval. The corresponding partition of  $K$  into intervals is called the generalized discrete gradient of  $f$ .

The singleton intervals are special, which is expressed by calling the simplices they contain and the corresponding values critical. To motivate this terminology, consider two contiguous values,  $r < s$ , and the corresponding sublevel sets,  $K_r = f^{-1}(-\infty, r]$  and  $K_s = f^{-1}(-\infty, s]$ . By assumption, no simplex maps to a value strictly between  $r$  and  $s$ , which implies that the difference between the two complexes is the level set at  $s$ . This level set is a disjoint union of steps, and because  $f$  is generalized discrete Morse, a disjoint union of mutually separated intervals. When we add the simplices of such an interval to  $K_r$ , then the homotopy type changes if the interval consists of a single, critical simplex, and it remains unchanged if the interval consists of two or more simplices. The operation of removing a non-singular interval is called a collapse. If all intervals in  $f^{-1}(s)$  are non-singular, then we write  $K_s \searrow K_r$  to express that  $K_r$  can be obtained from  $K_s$  by collapsing all intervals in the difference. More generally, if  $(r, t]$  contains no critical value of  $f$ , then  $K_t \searrow K_r$  (see Forman [26]).

### Appendix C. Equivalence of Definitions

This appendix proves that the three definitions of the Wrap complex offered in Section 2 are indeed equivalent. Given a generalized discrete Morse function  $f: D \rightarrow \mathbb{R}$ , we recall that  $\text{Wrap}_h(f) \subseteq \text{Alpha}_h(f)$  are the Wrap and the Alpha complex of  $f$  for  $h$ , and  $\text{Sg}_f(h)$  is the collection of singular steps whose critical simplices have function value at most  $h$ .

**Result 1 (Wrap Complex Lemma).** *Let  $f: D \rightarrow \mathbb{R}$  be a generalized discrete Morse function on a simplicial complex. Then*

- (i)  $\text{Wrap}_h(f)$  is the smallest complex  $K \subseteq D$  that satisfies  $\text{Alpha}_h(f) \searrow K$ , in which we restrict the collapses to intervals of  $f$ .
- (ii)  $\text{Wrap}_h(f)$  is the smallest subcomplex of  $D$  that contains  $\bigcup \text{Sg}_f(h)$  and is a union of intervals of  $f$ .

**Proof.** Consider two steps,  $\varphi$  and  $\psi$ , in the step graph  $\mathcal{G}$  of  $f$ . If there is an arc from  $\varphi$  to  $\psi$ , then  $\varphi$  contains a proper face of a simplex in  $\psi$ . This implies that if  $M$  is a collection of steps such that  $K = \bigcup M$  is a complex, then  $\psi \in M$  implies  $\varphi \in M$ . If both belong to  $M$ , then  $\varphi$  cannot be collapsed. On the other hand, if  $\varphi \in M$  and no successor of  $\varphi$  in  $\mathcal{G}$  belongs to  $M$ , then we can collapse  $\varphi$ ; that is:  $K \setminus \varphi$  is a complex. To prepare the proofs of (i) and (ii), we let  $M$  be a collection of steps such that

1.  $K = \bigcup M$  contains a critical simplex  $Q$  iff  $f(Q) \leq h$ ;
2.  $K$  is a complex;
3. there is no step  $\varphi \in M$  such that  $K \searrow K \setminus \varphi$ .

First, we claim that the three properties specify  $M$  uniquely. To prove this claim, let  $\varphi_0 \in M$  be non-singular and let  $\varphi_0, \varphi_1, \dots, \varphi_k$  be maximal such that  $\varphi_i \in M$  is a successor of  $\varphi_{i-1}$  in  $\mathcal{G}$  for  $1 \leq i \leq k$ . We note that  $k \geq 1$  because  $\varphi_0$  cannot be collapsed, and  $\varphi_k$  is singular because the sequence is maximal. To obtain a contradiction, we assume that  $N \neq M$  is another collection of steps that satisfies Properties 1, 2, and 3. Suppose first that  $N$  contains a step  $\mu_0 \notin M$ , and consider a maximal sequence  $\mu_0, \mu_1, \dots, \mu_\ell$  such that  $\mu_j \in N$  is a successor of  $\mu_{j-1}$  in  $\mathcal{G}$  for  $1 \leq j \leq \ell$ . Since  $\mu_0 \notin M$ , the step is necessarily non-singular, which implies  $\ell \geq 1$  and  $\mu_\ell$  singular. But then there is a first step along this sequence,  $\mu_j$ , that belongs to  $M$ . Since there is an arc from  $\mu_{j-1}$  to  $\mu_j$  and  $\mu_{j-1} \notin M$ , this contradicts that  $M$  is a complex. Suppose second that  $N$  contains no such step  $\mu_0$ , but  $M$  contains a step  $\varphi_0 \notin N$ . By the symmetric argument, this implies that  $N$  is not a complex, again a contradiction. We conclude that the collection  $M$  that satisfies Properties 1, 2, 3 is unique.

Second, we claim that the unique complex that satisfies Properties 1, 2, and 3 is  $\text{Wrap}_h(f)$ . By definition, the Wrap complex contains all critical simplices that satisfy  $f(Q) \leq h$ . The value of  $Q$  is the maximum of any step in the lower set of its singular interval, which implies that  $\text{Wrap}_h(f)$  contains no critical simplex with value larger than  $h$  and therefore satisfies Property 1. Property 2 is satisfied because all faces of a simplex in a step that are not in the step belong to predecessors of the step. Indeed, the directed path from a face to a simplex in the Hasse diagram maps to a possibly shorter directed path from the step of the face to the step of the simplex in  $\mathcal{G}$ . To see that Property 3 is satisfied as well, we note that every non-singular step  $\varphi_0 \subseteq \text{Wrap}_h(f)$  has a directed path to a singular interval and can therefore not be collapsed. We conclude that  $\text{Wrap}_h(f)$  is the only union of steps that satisfies Properties 1, 2, and 3.

To prove (i), we note that  $\text{Alpha}_h(f)$  satisfies Properties 1 and 2, so it cannot satisfy Property 3 unless it is equal to  $\text{Wrap}_h(f)$ . We can therefore collapse non-singular intervals. The process must halt, and the only way it can halt is when it reaches the unique union of steps that satisfies Properties 1, 2, 3, which is  $\text{Wrap}_h(f)$ .

To prove (ii), we observe that  $\text{Wrap}_h(f)$  contains  $\bigcup \text{Sg}_f(h)$  and is a union of steps. To see that it is the smallest such complex, suppose there is another complex,  $L = \bigcup N$ , that has this property and there exists a step  $\varphi \subseteq \text{Wrap}_h(f) \setminus L$ . As argued above, this contradicts that  $L$  is a complex, which implies (ii).  $\square$

#### Appendix D. Algorithm for Discrete Gradient

It is easy to see that the rise function defined in Section 3 is monotonic. As proved in [12], it also satisfies the more stringent requirements of a generalized discrete Morse function provided  $U \subseteq \Omega$  is in general position. The generalized discrete gradient of this function is a partition of the Delaunay mosaic into intervals, and this partition is instrumental in the construction of subcomplexes discussed in Section 4.

The construction of this partition is complicated by the impossibility of computing the rise function exactly, at least for general Legendre-type functions. Given a numerical approximation,  $g: K \rightarrow \mathbb{R}$ , our goal is therefore to first recover the generalized discrete Morse function that  $g$  approximates. Given a tolerance,  $\varepsilon \geq 0$ , we give an algorithm that computes such a function  $f: K \rightarrow \mathbb{R}$  with  $\|f - g\|_\infty \leq \varepsilon$  and such that the corresponding

partition is minimal in a restricted sense. To prepare the algorithm, we define the *gap* of a subset  $\varphi \subseteq K$  as the maximum difference of function values:

$$\text{gap } \varphi = \max_{P, Q \in \varphi, P \subseteq Q} [g(Q) - g(P)]. \quad (\text{A9})$$

If  $g$  is monotonic, then all gaps are non-negative. Otherwise, let  $-\varepsilon_0$  be the smallest (largest negative) gap between pairs  $P \subseteq Q$ , set  $g(P) = \min\{g(P), g(Q)\}$  whenever  $P \subseteq Q$ , and note that this makes  $g$  monotonic while changing the value of any simplex by at most  $\varepsilon_0$ . We will therefore assume that  $g$  is monotonic. Letting  $V$  be a partition of  $K$  into intervals, we call an interval  $\psi \notin V$  compatible with  $V$  if

- (i)  $\psi$  is the union of intervals in  $V$ ;
- (ii) every pair of simplices  $P \subseteq Q$  with  $P \in \psi$  and  $Q \notin \psi$  implies  $g(\text{ub}(\psi)) \leq g(Q)$ ,

in which  $\text{ub}(\psi)$  is the upper bound of the interval. The algorithm constructs the discrete gradient of  $f$  by adding compatible intervals to an initially trivial partition of  $K$ , namely the one in which every simplex belongs to its own set in the partition. The function itself is computed by spreading the function value of the upper bound to the other simplices in the interval. Let  $\psi_1, \psi_2, \dots, \psi_m$  be the collection of all intervals of  $K$ , sorted by gap, let  $\varepsilon \geq 0$  be a fixed threshold, and initialize  $i$  to 1 and  $V$  to the trivial partition of  $K$ .

```

1 while  $i \leq m$  and  $\text{gap } \psi_i \leq \varepsilon$  do
2   if  $\psi_i$  compatible with  $V$  then
3     remove all  $\varphi \in V$  with  $\varphi \subseteq \psi_i$  from  $V$ ;
4     add  $\psi_i$  to  $V$ ;
5     forall  $P \in \psi_i$  do set  $f(P) = g(\text{ub}(\psi_i))$ ;
6      $i = i + 1$ .
```

Condition (i) guarantees that the computed  $V$  is a partition of  $K$  into intervals. Condition (ii) makes sure that no relation is reversed, which implies that the computed function,  $f: K \rightarrow \mathbb{R}$ , is monotonic and that  $V$  is a refinement of its partition into steps. Finally,  $0 \leq f(P) - g(P) \leq \varepsilon$  for all simplices  $P \in K$ , as claimed. Without assuming that  $g$  is monotonic, the upper bound on the distance between the two functions is  $\varepsilon + \varepsilon_0$ .

A slight improvement of the algorithm takes into account that an interval can change from incompatible to compatible. By keeping track of this property throughout the algorithm, we can add an interval to the partition even after it was rejected earlier.

## References

- Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [\[CrossRef\]](#)
- Shannon, C. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [\[CrossRef\]](#)
- Amari, S.I.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Society: Providence, RI, USA, 2000; Volume 191.
- Rao, C. Information and Accuracy Attainable in Estimation of Statistical Parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–91.
- Fisher, R.A. On the Mathematical Foundations of Theoretical Statistics. *Philos. Trans. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character* **1922**, *222*, 309–368.
- Edelsbrunner, H.; Virk, Z.; Wagner, H. Topological data analysis in information space. In Proceedings of the 35th International Symposium on Computational Geometry, Portland, OR, USA, 18–21 June 2019; pp. 31:1–31:14.
- Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.
- van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
- McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3*, 861. [\[CrossRef\]](#)
- Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
- Scott, T.R.; Gallagher, A.C.; Mozer, M.C. von mises–fisher loss: An exploration of embedding geometries for supervised learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10612–10622.
- Edelsbrunner, H.; Wagner, H. Topological data analysis with Bregman divergences. In Proceedings of the 33rd International Symposium on Computational Geometry, Brisbane, Australia, 4–7 July 2017; pp. 39:1–39:16.
- Boissonnat, J.D.; Nielsen, F.; Nock, R. Bregman Voronoi diagrams. *Discret. Comput. Geom.* **2010**, *44*, 281–307. [\[CrossRef\]](#)

14. Wagner, H.; Dłotko, P. Towards topological analysis of high-dimensional feature spaces. *Comput. Vis. Image Underst.* **2014**, *121*, 21–26. [[CrossRef](#)]
15. Wagner, H.; Dłotko, P.; Mrozek, M. Computational Topology in Text Mining. In Proceedings of the Computational Topology in Image Context: 4th International Workshop, CTIC 2012, Bertinoro, Italy, 28–30 May 2012; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7309, p. 68.
16. Banerjee, A.; Merugu, S.; Dhillon, I.S.; Ghosh, J. Clustering with Bregman Divergences. *J. Mach. Learn. Res.* **2005**, *6*, 1705–1749.
17. Nielsen, F.; Piro, P.; Barlaud, M. Tailored Bregman ball trees for effective nearest neighbors. In Proceedings of the 25th European Workshop on Computational Geometry (EuroCG), Brussels, Belgium, 16–18 March 2009; pp. 29–32.
18. Nielsen, F.; Piro, P.; Barlaud, M. Bregman Vantage Point Trees for Efficient Nearest Neighbor Queries. In Proceedings of the 2009 IEEE International Conference on Multimedia and Expo, (ICME), New York, NY, USA, 28 June–3 July 2009; pp. 878–881. [[CrossRef](#)]
19. Gholizadeh, S.; Seyeditabari, A.; Zadrozny, W. Topological Signature of 19th Century Novelists: Persistent Homology in Text Mining. *Big Data Cogn. Comput.* **2018**, *2*, 33. [[CrossRef](#)]
20. Salnikov, V.; Cassese, D.; Lambiotte, R.; Jones, N.S. Co-occurrence simplicial complexes in mathematics: Identifying the holes of knowledge. *Appl. Netw. Sci.* **2018**, *3*, 1–23. [[CrossRef](#)]
21. Salton, G.; Wong, A.; Yang, C.S. A vector space model for automatic indexing. *Commun. ACM* **1975**, *18*, 613–620. [[CrossRef](#)]
22. Borsuk, K. On the imbedding of systems of compacta in simplicial complexes. *Fundam. Math.* **1948**, *35*, 217–234. [[CrossRef](#)]
23. Leray, J. L’anneau spectral et l’anneau filtré d’homologie d’un espace localement compact et d’une application continue. *J. Mathématiques Pures Appliquées* **1950**, *29*, 1–139.
24. Weil, A. Sur les théorèmes de de Rham. *Comment. Math. Helv.* **1952**, *26*, 119–145. [[CrossRef](#)]
25. Bauer, U.; Kerber, M.; Roll, F.; Rolle, A. A unified view on the functorial nerve theorem and its variations. *Expo. Math.* **2023**, *41*, 125503. [[CrossRef](#)]
26. Forman, R. Morse theory for cell complexes. *Adv. Math.* **1998**, *134*, 90–145. [[CrossRef](#)]
27. Freij, R. Equivariant discrete Morse theory. *Discret. Math.* **2009**, *309*, 3821–3829. [[CrossRef](#)]
28. Bauer, U.; Edelsbrunner, H. The Morse theory of Čech and Delaunay complexes. *Trans. Amer. Math. Soc.* **2017**, *369*, 3741–3762. [[CrossRef](#)]
29. Edelsbrunner, H. Surface reconstruction by wrapping finite point sets in space. In *Discrete and Computational Geometry. The Goodman–Pollack Festschrift*; Aronov, B., Basu, S., Pach, J., Sharir, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 379–404.
30. Bregman, L. The relaxation method of finding the common point of convex sets and its applications to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 200–217. [[CrossRef](#)]
31. Bauschke, H.; Borwein, J. Legendre functions and the method of random Bregman projections. *J. Convex Anal.* **1997**, *4*, 27–67.
32. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970. [[CrossRef](#)]
33. Antonelli, P. The geometry of random drift I. Stochastic distance and diffusion. *Adv. Appl. Prob.* **1977**, *9*, 238–249.
34. Edelsbrunner, H.; Kirkpatrick, D.; Seidel, R. On the shape of a set of points in the plane. *IEEE Trans. Inform. Theory* **1983**, *IT-29*, 551–559.
35. Alexandroff, P. Über kombinatorische Eigenschaften allgemeiner Kurven. *Math. Ann.* **1927**, *96*, 512–554. [[CrossRef](#)]
36. Edelsbrunner, H.; Harer, J. *Computational Topology. An Introduction*; American Mathematical Society: Providence, RI, USA, 2010.
37. Burg, J.P. *Maximum Entropy Spectral Analysis*; Stanford University: Stanford, CA, USA, 1975.
38. Itakura, F. Analysis synthesis telephony based on the maximum likelihood method. In Proceedings of the 6th International Congress on Acoustics, Tokyo, Japan, 21–28 August 1968; pp. 17–20.
39. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.
40. Rioul, O. A historical perspective on Schützenberger–Pinsker inequalities. In Proceedings of the International Conference on Geometric Science of Information, St. Malo, France, 30 August–1 September 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 291–306.
41. Chazal, F.; Cohen-Steiner, D.; Guibas, L.; Méholi, F.; Oudot, S. Gromov–Hausdorff stable signatures for shapes using persistence. *Comput. Graph. Forum* **2009**, *28*, 1393–1403. [[CrossRef](#)]
42. Voronoi, G. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième Mémoire: Recherches sur les paralléloèdres primitifs. *J. Reine Angew. Math.* **1908**, *134*, 198–287. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.