Smallest Enclosing Spheres and Chernoff Points in Bregman Geometry*

Herbert Edelsbrunner, Žiga Virk, and Hubert Wagner

IST Austria (Institute of Science and Technology Austria), Am Campus 1, 3400 Klosterneuburg, Austria, {edels, ziga.virk, hwagner}@ist.ac.at

Abstract

Smallest enclosing spheres of finite point sets are central to methods in topological data analysis. Focusing on Bregman divergences to measure dissimilarity, we prove bounds on the location of the center of a smallest enclosing sphere that depend on the range of radii for which Bregman balls are convex.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Bregman divergence, smallest enclosing spheres, Chernoff points, convexity, barycenter polytopes.

1 Introduction

A standard step in data analysis interprets non-geometric data geometrically. Examples are abundant, including images [8], medical records [17], text documents [9], and speech samples [4]. The motivating reason for this reinterpretation of data is the availability of standard mathematical tools for multi-dimensional point sets, such as cluster analysis, nearest neighbor search, dimensionality reduction, data visualization etc. These tools rely on a notion of dissimilarity between data points, and the Euclidean distance is often not ideal. This is not surprising, since an object may be characterized by a discrete probability distribution, a text document may be represented by the histogram of its word frequencies, an image may be stored as a set of gradient-orientation histograms, and so on and so forth. A popular alternative to the Euclidean distance is the *Kullback–Leibler divergence*, also known as the *relative entropy* [12], which is built on information theoretic foundations and meaningfully compares probability distributions. There is experimental evidence for its efficacy, and this in spite of violating two of the three axioms we require from a metric; see [9] for a comparison of measures used to cluster text documents. The relative entropy belongs to the family of *Bregman divergences* [3]. Another member of this family is the Itakura-Saito divergence, which is classically used to compare power spectra of speech patterns [10]. The extension of topological data analysis methods from the Euclidean metric to Bregman divergences needs smallest enclosing spheres to turn data into Bregman–Čech complexes, and smallest circumspheres to turn data into Bregman–Delaunay complexes. We are therefore motivated to study these spheres in detail.

Notation and terminology. Before reviewing prior work and presenting our results, it is convenient to introduce the most important concepts. A function $F \colon \Omega \to \mathbb{R}$ on an open convex subset $\Omega \subseteq \mathbb{R}^d$ is of *Legendre type* if

- F is strictly convex,
- F is differentiable,
- the length of the gradient goes to infinity when we approach the boundary of Ω.

* This research is partially supported by the DFG Collaborative Research Center TRR 109, ‘Discretization in Geometry and Dynamics’, through grant no. I02979-N35 of the Austrian Science Fund (FWF).

Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

As proved in [5, Theorem 2.86], the combination of convexity and differentiability implies continuous differentiability. The somewhat technical third condition guarantees that the conjugate of F is also of Legendre type; see [18, page 259]. There will be no appearance of the conjugate in this paper, but we will make use of a consequence of the conjugate being of Legendre type proved in [7]. The *Bregman divergence* from x to y associated with F is the difference between F and the best linear approximation of F at y, both evaluated at x:

$$D_F(x\|y) = F(x) - [F(y) + \langle \nabla F(y), x - y \rangle]. \tag{1}$$

The divergence is not necessarily symmetric. We therefore define two balls with given center and radius, one by measuring the divergence *from* the center and the other *to* the center. Specifically, the *primal* and *dual Bregman balls* with center $x \in \Omega$ and radius $r \geq 0$ are

$$B_F(x;r) = \{y \in \Omega \mid D_F(x\|y) \leq r\}, \tag{2}$$

$$B_F^*(x;r) = \{y \in \Omega \mid D_F(y\|x) \leq r\}. \tag{3}$$

While the dual ball is necessarily convex, this is not true for the primal ball. Since we use F throughout this paper, we will feel free to drop it from the notation. An *enclosing sphere* of a set $X \subseteq \Omega$ is the boundary of a dual Bregman ball that contains all points of X. A *circumsphere* of X is an enclosing sphere that passes through all points of X.

Prior work and results. The family of Bregman divergences is named after Lev Bregman who studied convex programming problems in [3]. Each such divergence is based on a Legendre type function; see Rockafellar [18]. A prominent member of the family is the relative entropy, which is based on the Shannon entropy. Its introduction by Kullback and Leibler [12] predates the work of Bregman. Boissonnat, Nielsen, and Nock pioneered the study of Bregman divergences within computational geometry and information geometry. In [15, 16] they studied algorithms for fitting Bregman balls enclosing a set of points and in [1] they introduced Bregman–Voronoi diagrams. To get a useful dual structure, we need that the non-empty common intersections of primal Bregman balls with the corresponding Voronoi domains are contractible, a property proved in [7]. This opened the door to constructing filtrations of Čech and Delaunay complexes and to analyzing the data with persistent homology, which is one of the key tools in topological data analysis.

The bridge to the work in this paper is the observation that a collection of primal Bregman balls of radius r have a non-empty common intersection iff their centers are contained in a dual Bregman ball of the same radius r. In this paper, we study the location of the center of the smallest enclosing sphere of a finite set of points, and we follow [14] in calling this center the *Chernoff point* of the set. It is easy to show that the Chernoff point belongs to the convex hull of the finite set, which was first proven in [16]. For completeness, we present a proof of this observation based on the widely used Bregman–Pythagoras Theorem; see e.g. [1]. To improve on this insight, we distinguish between Bregman divergences with convex and with nonconvex balls. The original contributions presented in this paper are:

- for convex balls, we show that the Chernoff point of a simplex is contained in the convex hull of the Chernoff points of its facets, which is generally a much smaller space of possible locations,
- for nonconvex balls, we prove a weaker result with heavier machinery.

To provide context for these results, we mention that this paper follows [6, 7] as third in a series. The broader goal is to lay the theoretical foundations needed to expand the range of applications in which topological tools can be meaningfully used. This line of research established that Bregman divergences and the balls they induce are compatible with methods from computational topology, and in particular with persistent homology. The initial steps in this direction left however many important questions unanswered. In this undertaking, the location of the Chernoff point of a set plays a crucial role. A limiting factor was the general paucity of nontrivial bounds on its location. The new bounds are useful, for example, in pruning redundant computations when Chernoff points of many small and possibly overlapping point sets have to be computed — a scenario that is typical for topological methods. This contrasts the usual setting in which a single query for a large point set is considered.

In summary, we believe that a deeper understanding of the often counterintuitive behavior of Bregman divergences is needed to reach the full potential of the topological tools. This paper contributes by demonstrating that ideas form combinatorial topology are useful in the study of Chernoff points in particular and of Bregman divergences in general.

Outline. Section 2 proves basic properties of smallest enclosing spheres and smallest circumspheres. Section 3 introduces barycenter polytopes. Section 4 proves our main result in the easier convex case. Section 5 extends the main result to the more difficult nonconvex case. Section 6 shows that the nesting hierarchy proved in the nonconvex case is best possible. Section 7 concludes this paper.

2 Smallest Spheres

Growing primal Bregman balls from given points in Ω, we study the point at which these balls meet first. Equivalently, we study the Chernoff point, which is the center of the smallest enclosing sphere of the given points. In particular, we prove that the Chernoff point of a simplex lies in the simplex, and that the center of the smallest circumsphere lies in the affine hull of the simplex.

Bregman–Pythagoras. We use the following notation throughout this paper: letting $A \subseteq \Omega$ be a closed convex subset and $y \in \Omega$ a point, we write y_A for the point in A that minimizes the Bregman divergence to y: $y_A = \arg\min_{a \in A} D(a\|y)$. We will make use of an extension of Pythagoras' Theorem to Bregman divergences; see e.g. [1]. We give a proof for completeness.

▶ **Proposition 1** (Bregman–Pythagoras). *All points a of a closed convex set $A \subseteq \Omega$ satisfy $D(a\|y) \geq D(a\|y_A) + D(y_A\|y)$, with equality if A is an affine subspace.*

PROOF. We first assume that $F(y) = 0$ and f has its minimum at $y \in \Omega$. It follows that $\nabla F(y) = 0$ and $D(x\|y) = F(x) \geq 0$ for every $x \in \Omega$. The sets of constant distance r to y are therefore the level sets, $F^{-1}(r)$; see Figure 1. The point y_A is where the lowest level set touches A. The gradient

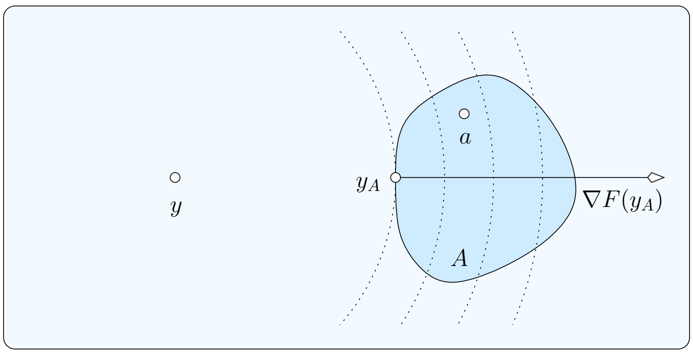

Figure 1: The gradient at a point of a level set of F forms a right angle with the level set.

of F at y_A is normal to this level set. Hence,

$$F(a) \geq F(a) - \langle \nabla F(y_A), a - y_A \rangle = D(a\|y_A) + F(y_A), \tag{4}$$

because the scalar product is necessarily non-negative. Rewriting the inequality, we get $D(a\|y) \geq D(a\|y_A) + D(y_A\|y)$, as claimed. If A is an affine subspace of $\mathbb{R}^d$, then the scalar product in (4) vanishes for all $a \in A$, which implies equality, again as claimed.

If F does not satisfy the simplifying assumption, then we construct $G\colon \Omega \to \mathbb{R}$ defined by $G(x) = F(x) - [F(y) + \langle \nabla F(y), x - y \rangle]$. It is clear that $G(y) = \nabla G(y) = 0$. For any two points $u, v \in \Omega$, we have $D_G(u\|v) = D_F(u\|v)$, so we get the claimed inequality from $D_G(a\|y) \geq D_G(a\|y_A) + D_G(y_A\|y)$, which is implied by the above argument. ◻

Smallest enclosing sphere. Recall that an enclosing sphere of a set $X \subseteq \Omega$ is the boundary of a dual Bregman ball that contains X. We are interested in the smallest such sphere in the case in which X is a set of $k+1 \leq d+1$ points. We refer to such a set X as an *(abstract) k-simplex*, and we write $\mathrm{conv}(X)$ for the corresponding geometric k-simplex.

▶ **Lemma 2** (Smallest Enclosing Sphere). *The Chernoff point of any k-simplex $X \subseteq \Omega$ is unique and contained in* $\mathrm{conv}(X)$*, for every $0 \leq k \leq d$.*

PROOF. Let $B^*(y; r)$ be a dual Bregman ball with smallest radius that contains all points of X. Writing $x_0, x_1, \ldots, x_k$ for the points in X, this implies $D(x_i\|y) \leq r$ for all i, with equality for at least one index i. To get a contradiction, we set $A = \mathrm{conv}(X)$ and assume $y \notin A$. Using Proposition 1, we get $D(x_i\|y) \geq D(x_i\|y_A) + D(y_A\|y)$ for all $0 \leq i \leq k$. Since $D(x_i\|y) \leq r$ and $D(y_A\|y) > 0$, by assumption of y not being in A, this implies $D(x_i\|y_A) < r$ for all $0 \leq i \leq k$. But this contradicts the minimality of $B^*(y; r)$, and we get $y \in A$ as desired. The uniqueness of y follows from the strict convexity of F. ◻

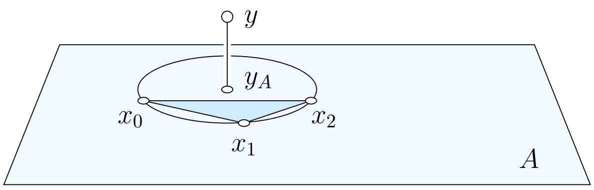

Figure 2: Assuming the center of the smallest circumsphere, y, does not lie in $A = \mathrm{aff}\, X$ leads to a contradiction.

Smallest circumsphere. Recall that a circumsphere of X is the boundary of a dual Bregman ball that passes through all points of X. There may or may not be any such ball whose center is contained in Ω. To simplify the discussion, we restrict ourselves to a case in which such centers are guaranteed to exist, namely when $\Omega = \mathbb{R}^d$ and X is a set of $k+1 \leq d+1$ points in general position in $\mathbb{R}^d$. Note that for $k+1 < d+1$, the circumsphere is not unique, and often the smallest one is of interest. Using Proposition 1, it is not difficult to prove that the center of the smallest circumsphere of X is contained in the affine hull of X.

▶ **Lemma 3** (Smallest Circumsphere). *The center of the smallest circumsphere of any k-simplex $X \subseteq \Omega$ is unique and contained in* $\mathrm{aff}\, X$*, for every $0 \leq k \leq d$.*

PROOF. Let $B^\circ(y; r)$ be the ball bounded by a smallest circumsphere of X. Writing $x_0, x_1, \ldots, x_k$ for the points in X, this implies $D(x_0\|y) = D(x_1\|y) = \ldots = D(x_k\|y) = r$, and that r is the smallest real number for which there is a point $y \in \Omega$ such that these equalities are satisfied; see Figure 2. To get a contradiction, we set $A = \mathrm{aff}\, X$ and assume $y \notin A$. Using Proposition 1 for affine subspaces, we get $D(x_i\|y) = D(x_i\|y_A) + D(y_A\|y)$ for all $0 \leq i \leq k$. Because $D(y_A\|y) > 0$, by assumption of y not being in A, this implies $D(x_0\|y_A) = D(x_1\|y_A) = \ldots = D(x_k\|y_A) < r$, which contradicts the minimality of $B^\circ(y;r)$. We get uniqueness because there is only one point in $A = \operatorname{aff} X$ equally far from all the x_i. ◻

3 Barycenter Polytopes

Given a simplex, we introduce the family of convex hulls of the face barycenters. The motivation for the study of these polytopes is the sharpening of Lemma 2.

Nested sequence of polytopes. Let $X = \{x_0, x_1, \ldots, x_k\}$ be a k-simplex in $\mathbb{R}^d$. For every subset $J \subseteq \{0, 1, \ldots, k\}$, we write $X_J \subseteq X$ for the corresponding face, $j = |J| - 1$ for the dimension of X_J, and $b_J = \frac{1}{j+1} \sum_{i \in J} x_i$ for the *barycenter* of X_J. For $0 \le j \le k$, the *j-th barycenter polytope* of X is

$$\Delta_j^k(X) = \operatorname{conv}\{b_J \mid |J| = j+1\}. \tag{5}$$

Note that $\Delta_0^k = \Delta_0^k(X) = \operatorname{conv}(X)$. In three dimensions, Δ_0^3 is a tetrahedron, Δ_1^3 is an octahedron, Δ_2^3 is again a tetrahedron, and Δ_3^3 is a point. It is not difficult to see that the barycenter polytopes are nested.

▶ **Lemma 4** (Nesting). *The barycenter polytopes of any k-simplex satisfy $\Delta_0^k \supseteq \Delta_1^k \supseteq \ldots \supseteq \Delta_k^k$.*

PROOF. Let $J \subseteq \{0, 1, \ldots, k\}$ of cardinality $j + 1 \ge 2$. We take the average of the barycenters that correspond to J with one index removed:

$$\frac{1}{j+1} \sum_{\ell \in J} b_{J \setminus \{\ell\}} = \frac{1}{j+1} \sum_{\ell \in J} \left[\frac{1}{j} \sum_{i \in J \setminus \{\ell\}} x_i \right] = \frac{1}{j(j+1)} \sum_{\ell \in J} \sum_{i \in J \setminus \{\ell\}} x_i. \tag{6}$$

Each point $x_i \in X_J$ appears j times in the double-sum, which implies that the above average is equal to b_J. We thus proved that every vertex of the j-th barycenter polytope is a convex combination of the vertices of the $(j-1)$-th barycenter polytope, for $1 \le j \le k$. Hence, $\Delta_j^k \subseteq \Delta_{j-1}^k$, as claimed. ◻

Face structure. It is instructive to take a closer look at Δ_1^3, which is the first barycenter polytope that is not a simplex. Being an octahedron, it has 8 faces of co-dimension one, which we refer to as *facets*. Four of the facets are the 1-st barycenter polytopes of the triangles bounding the tetrahedron, and the other four facets are homothetic copies of the original four triangle. More generally, most barycenter polytopes have twice as many facets as the defining simplex. Write $\#\mathrm{facets}(\Delta_j^k)$ for the number of facets of Δ_j^k.

▶ **Lemma 5** (Number of Facets). *Let $k \ge 1$. The number of facets of the j-th barycenter polytope of a k-simplex is*

$$\#\mathrm{facets}(\Delta_j^k) = \begin{cases} k+1 & \text{if } j = 0, k-1, \\ 2k+2 & \text{if } 1 \le j \le k-2. \end{cases} \tag{7}$$

PROOF. Index the coordinates of points in $\mathbb{R}^{k+1}$ from 0 to k, and let e_i be the unit vector in the i-th coordinate direction. Identifying the k-simplex with the endpoints of these vectors, we consider the $(k+1)$-dimensional cube spanned by e_0 to e_k. For $0 \le j \le k-1$, we define the *j-th slice* of this cube as the intersection with the k-plane of points $\sum_{i=0}^k \gamma_i e_i$ satisfying $\sum_{i=0}^k \gamma_i = j + 1$. It is the convex hull of the $(j+1)$-fold sums of the unit vectors. Scaling the j-th slice by a factor $\frac{1}{j+1}$, we get the j-th barycenter polytope of the k-simplex.

For $j = 0, k-1$, the k-plane intersects half the facets of the cube, and for $1 \leq j \leq k-2$, it intersects all facets of the cube. Each facet of the j-th slice, and after scaling of Δ_j^k, is the intersection of the k-plane with a facet of the cube, which implies the claimed number of facets of the barycenter polytope. ◻

It is easy to extend Lemma 5 and get explicit formulas for the number of i-dimensional faces of Δ_j^k, but this is not necessary for this paper. Observe that half the facets of the $(k+1)$-cube share the origin, and the other half share the point $(1, 1, \ldots, 1)$ as a vertex. We call the slice of a facet that shares the origin a *far facet* of the corresponding barycenter polytope, noting that it is a barycenter polytope of a facet of the given k-simplex. Similarly, we call the slice of a facet that shares $(1, 1, \ldots, 1)$ a *near facet* of the barycenter polytope, noting that it is the homothetic copy of a barycenter polytope of a facet of the given k-simplex.

Central projection. For the purpose of the proof of Theorem 8, we subdivide the boundary of Δ_j^k and re-associate the pieces to get the boundary complex of the k-simplex, at least topologically. Write $b(X)$ for the barycenter of the k-simplex, and introduce the central projection,

$$\pi_j^k \colon \partial \mathrm{conv}(X) \to \partial \Delta_j^k, \tag{8}$$

which we define by mapping $x \in \partial \mathrm{conv}(X)$ to the unique convex combination of x and $b(X)$ that belongs to the boundary of Δ_j^k. Figure 3 shows the picture of the two maps in the plane.

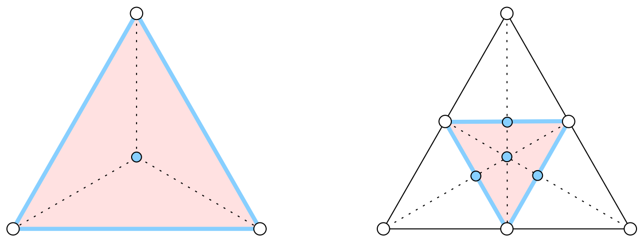

Figure 3: *Left:* the map π_0^2 is the identity on the boundary of the triangle. *Right:* the map π_1^2 projects the vertices of $\mathrm{conv}(X)$ to the midpoints of the edges of Δ_1^2. The structure of $\partial \mathrm{conv}(X)$ is recovered by gluing the half-edges in pairs at the shared endpoints.

In the general case, we subdivide $\partial \Delta_j^k$ along the image of the $(k-2)$-skeleton of $\partial \mathrm{conv}(X)$. To convince ourselves that this is well defined, we note that $\partial \Delta_j^k$ is a $(k-1)$-sphere for every $0 \leq j \leq k-1$. Similarly, $\partial \mathrm{conv}(X)$ is a $(k-2)$-sphere. The center of the projection, $b(X)$, lies in the interior of Δ_j^k and also in the interior of $\mathrm{conv}(X)$, which implies that π_j^k is a homeomorphism. We therefore reach our goal by first glueing the facets of Δ_j^k along their shared faces — which amounts to forgetting the decomposition of the $(k-2)$-sphere these facets imply — and second cutting the $(k-1)$-sphere along the image of the $(k-2)$-skeleton of $\partial \mathrm{conv}(X)$ — which effectively triangulates the $(k-1)$-sphere with $k+1$ $(k-1)$-simplices.

While not necessary, it is easy to understand the interaction between the two $(k-2)$-skeleta. Indeed, all far facets of Δ_j^k belong to facets of $\mathrm{conv}(X)$ and therefore do not get subdivided. Each near facet of Δ_j^k is subdivided into k cones that emanate from its center of mass, like the star of a vertex of $\partial \mathrm{conv}(X)$. The simplest example of this operation can be seen in Figure 3 on the right.

4 Theorem in Convex Case

The aim in this section is to sharpen Lemma 2 by further limiting the region in which the center of the smallest enclosing sphere can lie. Here we discuss the case in which all primal Bregman balls are convex.

Chernoff polytopes. As before, let X be a k-simplex in Ω. For each subset $J \subseteq \{0, 1, \ldots, k\}$, we recall that X_J is the corresponding face of X, and we let $B^*(d_J; R_J)$ be the smallest dual Bregman ball that contains X_J. Equivalently, $R_J = R(X_J)$ is the minimum radius r such that $\bigcap_{i \in J} B(x_i; r) \neq \emptyset$, and this intersection consists of a single point, namely the Chernoff point d_J of X_J. In analogy with the barycenter polytope of the previous section, we define the *j-th Chernoff polytope* of X as the convex hull of the Chernoff point of faces of dimension $j = |J| - 1$:

$$\Delta_j(X) = \text{conv}\,\{d_J \mid |J| = j + 1\} \tag{9}$$

for $0 \leq j \leq k$; see Figure 4 for an illustration. Note that $\Delta_0(X) = \text{conv}(X)$, but for positive indices j, $\Delta_j = \Delta_j(X)$ is not necessarily the j-th barycenter polytope because the vertices are not necessarily the barycenters of the faces.

Nesting. The Chernoff polytopes drawn in Figure 4 are nested, but that they retain this property of the barycenter polytopes in general needs a proof.

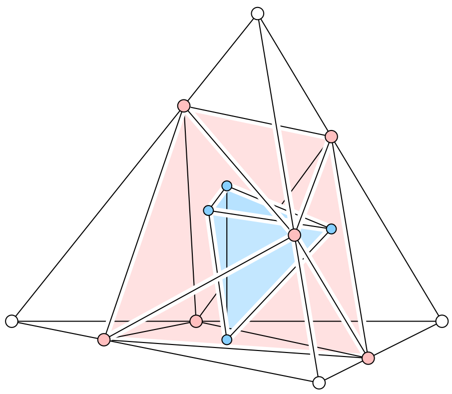

Figure 4: The 1-st Chernoff polytopes of the tetrahedron is the distorted octahedron (drawn with pink facets) whose vertices lie on the edges of the outer tetrahedron, and the 2-nd is the inner tetrahedron (drawn with blue facets) whose vertices lie on four of the triangles bounding the octahedron.

▶ **Theorem 6** (Nesting for Convex Balls)**.** *Let $F \colon \Omega \to \mathbb{R}$ be of Legendre type such that the primal Bregman balls are convex, and let $\Delta_0, \Delta_1, \ldots, \Delta_k$ be the Chernoff polytopes of a k-simplex $X \subseteq \Omega$. Then $\Delta_0 \supseteq \Delta_1 \supseteq \ldots \supseteq \Delta_k$.*

PROOF. We prove $\Delta_{j-1} \supseteq \Delta_j$ by showing that the Chernoff point of every j-face of $\text{conv}(X)$ is contained in the convex hull of the Chernoff points of the $(j-1)$-faces of this j-face. Indeed, this implies that the vertices of Δ_j lie in Δ_{j-1}, and the claimed inclusion follows. To formalize this idea, let $0 < j \leq k$, set $J = \{0, 1, \ldots, j\}$, and consider $J \setminus \{\ell\}$ for every $0 \leq \ell \leq j$. We prove that d_J belongs to the convex hull of the $d_{J \setminus \{\ell\}}$ in two steps.

For the first step, we write $\text{Sd}(X_J)$ for the *Chernoff subdivision* of $\text{conv}(X_J)$. We obtain it from the barycentric subdivision by moving each barycenter to the location of the corresponding Chernoff point. If all Chernoff points are different, the two subdivisions are isomorphic, but it is possible that two or more barycenters map to the same Chernoff point, in which case some of the simplices in the barycentric subdivision collapse to simplices of smaller dimension. Since $B(x_0; R_J)$ contains the point d_J, it also contains all points $d_{J'}$ with $0 \in J' \subseteq J$. Indeed, if the balls $B(x_i; R_J)$ with $i \in J$ have a point in common, then so do the balls with $i \in J'$. By convexity, $B(x_0; R_J)$ then contains all simplices in $\mathrm{Sd}(X_J)$ that share x_0. Similarly, $B(x_\ell; R_J)$ contains all simplices in $\mathrm{Sd}(X_J)$ that share x_ℓ, for each $\ell \in J$. It follows that the $j+1$ balls cover the entire Chernoff subdivision of $\mathrm{conv}(X)$ and thus the entire j-face:

$$\mathrm{conv}(X_J) \subseteq \bigcup_{\ell=0}^{j} B(x_\ell; R_J). \tag{10}$$

For the second step, we write Σ_J for the convex hull of the points $d_{J \setminus \{\ell\}}$, $\ell \in J$. It is the $(j-1)$-st Chernoff polytope of X_J and therefore also a j-simplex. Define $C_\ell = B(x_\ell; R_J) \cap \Sigma_J$, for each $0 \leq \ell \leq j$. By Lemma 2, the points $d_{J \setminus \{\ell\}}$ belong to $\mathrm{conv}(X_J)$, so $\Sigma_J \subseteq \mathrm{conv}(X_J)$, and (10) implies that the sets C_ℓ cover Σ_J. But this does not yet imply that the $(j+1)$-fold intersection of the balls, which is the point d_J, belongs to the C_ℓ and therefore to Σ_J. To prove this, we need the Nerve Theorem [2], which applies to the sets C_ℓ because they are convex. It implies that the nerve of the sets C_ℓ has the same homotopy type as Σ_J and is therefore contractible. The j-fold intersections are all non-empty, as witnessed by the vertices of Σ_J. Hence, the nerve contains the boundary complex of a j-simplex. Contractibility thus implies that the nerve also contains the j-simplex. In other words, $d_J \in C_\ell$ for $0 \leq \ell \leq j$ and therefore $d_J \in \Sigma_J$. ◻

We note that Theorem 6 tightens Lemma 2 in the case of convex Bregman balls: the center of the smallest enclosing sphere of a k-simplex is contained in the convex hull of the centers of the smallest enclosing spheres of all $(k-1)$-faces.

5 Theorem in Nonconvex Case

We will see shortly that the assumption of convex Bregman balls can be relaxed. The proof of the inclusions in the nonconvex case is the same as in the convex case, except that the individual steps are more complicated. We begin with an auxiliary result.

A fixed point lemma. To generalize Theorem 6 to the nonconvex case, we employ a classic result in topology proved in 1929 by Knaster, Kuratowski, and Mazurkiewicz [11]. It can be used to prove the Brouwer Fixed Point Theorem, which states that every continuous function from the n-dimensional closed ball to itself has a fixed point.

▶ **Proposition 7** (Fixed Point). *Let X be a k-simplex with vertices x_0 to x_k, and let C_0 to C_k be closed sets such that the union of any subcollection of the sets contains the face spanned the corresponding subcollection of vertices. Then $\bigcap_{i=0}^{k} C_i \neq \emptyset$.*

Take for example C_i equal to the closed star of vertex x_i in the barycentric subdivision of $\mathrm{conv}(X)$. The conditions in the proposition are satisfied, and the stars have indeed a non-empty common intersection, namely the barycenter of X. It is important to note that the lemma is topological and therefore also holds for homeomorphically deformed k-simplices.

Interrupted hierarchy. Now suppose that there is a threshold such that all balls with radius at most this threshold are convex, but this is not guaranteed for balls with radius larger than the threshold. An example is the Itakura–Saito divergence [10] defined on the standard simplex whose balls are convex provided the radius does not exceed $\ln 2 - \frac{1}{2} = 0.193\ldots$ [6]. How does this weaken the hierarchy in Theorem 6? To state our claim, we define $R_j = \max_{|J|=j+1} R_J$, noting that $r \geq R_j$ iff all $(j+1)$-fold intersections of the $B(x_i; r)$ are non-empty.

▶ **Theorem 8** (Nesting for Nonconvex Balls). *Let $F\colon \Omega \to \mathbb{R}$ be of Legendre type such that the primal Bregman balls are convex for radii $r \leq R_j$, and let $\Delta_0, \Delta_1, \ldots, \Delta_k$ be the Chernoff polytopes of a k-simplex $X \subseteq \Omega$. Then $\Delta_0 \supseteq \Delta_1 \supseteq \ldots \supseteq \Delta_j \supseteq \Delta_{j+1}, \Delta_{j+2}, \ldots, \Delta_k$.*

PROOF. To prove the inclusions $\Delta_0 \supseteq \Delta_1 \supseteq \ldots \supseteq \Delta_j$, it suffices to consider balls of radius R_j or smaller. These are convex, by assumption, so the inclusions are implied by Theorem 6. To prove $\Delta_j \supseteq \Delta_i$, for $j < i \leq k$, we show that for every i-face, the Chernoff point is contained in the j-th Chernoff polytope. It suffices to consider $i = k$. In the first step of the proof, we generalize (10) to

$$\mathrm{conv}(X_J) \subseteq \bigcup_{\ell \in J} B(x_\ell; R_J) \tag{11}$$

for every subset $J \subseteq \{0, 1, \ldots, k\}$ also in the nonconvex case. We use induction over the dimension. To simplify the notation, assume $J = \{0, 1, \ldots, j\}$ and let $H = H_J$ be the j-dimensional plane spanned by X_J. By inductive assumption, we have (11) for all $J \setminus \{\ell\}$, $0 \leq \ell \leq j$. By definition of R_J, the $j+1$ balls $B(x_\ell; R_J)$ have a non-empty common intersection, namely the point d_J. By Lemma 2, also the j-dimensional slices of the balls defined by H have the point d_J in common. It follows that the nerve of the sliced balls is a j-simplex, which is contractible. Since F is of Legendre type, so is its restriction to the j-plane, $F|_H$. As proved in [7], this implies that all intersections of the sliced balls are contractible. Hence, the Nerve Theorem applies, implying that also the union of the sliced balls is contractible. But this union covers the $(j-1)$-dimensional boundary of $\mathrm{conv}(X_J)$, so it must also cover $\mathrm{conv}(X_J)$ to be contractible. Hence (11) follows, and in particular $\mathrm{conv}(X) \subseteq \bigcup_{\ell=0}^{k} B(x_\ell; R_k)$.

For the second step, recall that $\Delta_j = \Delta_j(X)$ is the j-th Chernoff polytope of X. Define $C_\ell = B(x_\ell; R_k) \cap \Delta_j$, for $0 \leq \ell \leq k$. Since $\Delta_j \subseteq \mathrm{conv}(X)$, the balls $B(x_\ell; R_k)$ cover Δ_j. By the Nerve Theorem [2], the nerve of the sets C_ℓ has the same homotopy type as Δ_j and is therefore contractible. To apply Proposition 7 to Δ_j, we first interpret Δ_j as the homeomorphic image of a k-simplex. Specifically, we decompose its boundary by central projection of the boundary of $\mathrm{conv}(X)$, in which we use any point in the interior of Δ_j as center; see Figures 3 and 4 for illustrations. We denote this topological k-simplex by $\tilde{\Delta}_j$.

Recall that the facets of Δ_j are classified as near and far facets of the $k+1$ points in X. Each facet of $\tilde{\Delta}_j$ consists of a far facet and pieces of k near facets of Δ_j. To describe this in the necessary amount of detail, we denote the facets of $\tilde{\Delta}_j$ by $\Phi_0, \Phi_1, \ldots, \Phi_k$ and the facets of X by $X_\ell = X \setminus \{\ell\}$ for $0 \leq \ell \leq k$. The indexing is chosen so that Φ_ℓ consists of the far facet of x_ℓ — which is contained in $\mathrm{conv}(X_\ell)$ — together with pieces of the near facets of the vertices $x_i \in X_\ell$. By (11), the far facet of x_ℓ is covered by the balls $B(x_i; R_k)$, for $i \neq \ell$. Furthermore, the near facet of x_ℓ is covered by $B(x_\ell; R_k)$ simply because $B(x_\ell; R_j) \subseteq B(x_\ell; R_k)$, and the former ball is convex and contains the relevant vertices of Δ_j. It follows that Φ_ℓ is covered by the balls $B(x_i; R_k)$, for $i \neq \ell$. Hence, $\tilde{\Delta}_j$ and the sets C_ℓ satisfy the assumptions of Proposition 7. The proposition thus implies that the common intersection of the C_ℓ is non-empty. This intersection can only be the Chernoff point of X, which we therefore conclude lies inside Δ_j, as required. ◻

In particular, if the balls remain convex until radius R_{k-1}, then Theorem 6 still holds. Without any assumption on convexity, we do not claim anything beyond Δ_0 containing all points d_J, $J \subseteq \{0, 1, \ldots, k\}$, which is Lemma 2.

6 No Improvement

We finally show that Theorem 8 is best possible, in the sense that the hierarchy of inclusions cannot be extended beyond Δ_{j+1}. To this end, we construct a function of Legendre type, $F\colon \mathbb{R}^d \to \mathbb{R}$, and a d-simplex, $X \subseteq \mathbb{R}^d$, such that

- for radius $r \leq R_j$, the primal Bregman balls centered at the points of X are convex,
- there is at least one $(j+2)$-face of X whose Chernoff point is not contained in Δ_{j+1}.

It will suffice to consider the case $j+2 = d$. We begin with the construction in $d = 2$ dimensions, when $j = 0$ and $R_j = 0$, so the first condition is automatically satisfied. With an eye on the generalization to higher dimensions, we will nevertheless make sure that the balls with small but positive radius are convex. We simplify the task by requiring that F be convex but not necessarily strictly convex and not necessarily differentiable. Small bump functions can be used to eventually turn the convex function into a Legendre type function.

Two-dimensional construction. Let $\Delta_0^2 = \mathrm{conv}(X)$ with $X = \{x_0, x_1, x_2\}$ be an equilateral triangle with edges of length $\sqrt{2}$ and center at the origin in $\mathbb{R}^2$. The first barycenter polytope is $\Delta_1^2 = \mathrm{conv}(Y)$ with $Y = \{y_0, y_1, y_2\}$ and $y_i = \frac{1}{2}(x_{i+1} + x_{i+2})$, where we take indices modulo 3. Calculating the Euclidean inradii of Δ_1^2 and Δ_0^2, we choose a radius strictly between them, $1/\sqrt{24} < \varrho < 1/\sqrt{6}$, and we let $D(\varrho)$ be the (Euclidean) disk with this radius and center at the origin. As illustrated in Figure 5, $D(\varrho)$ is neither contained in Δ_1^2 nor does it contain Δ_1^2. We finally construct $F\colon \mathbb{R}^2 \to \mathbb{R}$ by mapping $a \in \mathbb{R}^2$ to $F(a) = \|a\|^2$ if $a \in \mathbb{R}^2 \setminus D(\varrho)$, and to $F(a) = \varrho^2$ if $a \in D(\varrho)$. The graph of F is a paraboloid with a flattened bottom. Recall that

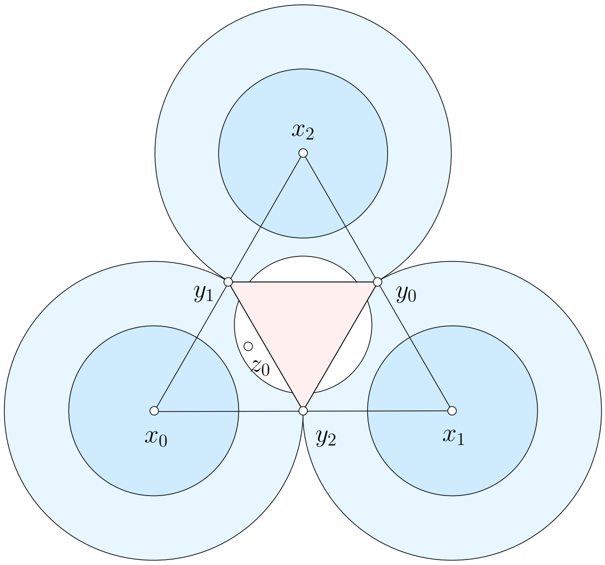

Figure 5: In the nonconvex case, we can arrange that the Chernoff point of X to lie outside the triangle spanned by the Chernoff points of the edges of X.

$B_F(x_i; r)$ can be constructed by vertically projecting all points of the graph that are visible from the point $(x_i, \|x_i\|^2 - r) \in \mathbb{R}^2 \times \mathbb{R}$. Writing $D_i(\sqrt{r})$ for the disk with center x_i and squared radius r, the primal Bregman ball satisfies

$$B_F(x_i; r) = \begin{cases} D_i(\sqrt{r}) & \text{if } r \leq (\sqrt{2/3} - \varrho)^2, \\ D_i(\sqrt{r}) \setminus D(\varrho) & \text{if } (\sqrt{2/3} - \varrho)^2 \leq r \leq 2/3 - \varrho^2, \\ D_i(\sqrt{r}) \cup D(\varrho) & \text{if } 2/3 - \varrho^2 < r. \end{cases} \quad (12)$$

The Bregman ball is convex in the first case, and it is nonconvex in the second case. To give the final touch, we observe that the gradient of F is bounded away from zero everywhere outside $D(\varrho)$. We can therefore change F so its graph over $D(\varrho)$ is an upside-down cone with apex $z_0 \in \mathrm{int}\, D(\varrho) \setminus \Delta_1^2$, and we can do this without violating convexity and without changing F outside this disk. With this modification, we get z_0 as the Chernoff point of X, which by construction lies outside Δ_1^2.

Higher dimensions. The 2-dimensional construction generalizes in a straightforward way to $d \geq 2$ dimensions. The only nontrivial step is to prove that $\varrho > 0$ can be chosen so that the Euclidean ball $D(\varrho)$ neither contains Δ^d_{d-1} nor is contained in it, and that a ball centered at x_i and touching $D(\varrho)$ in a single point contains the near facet of Δ^d_{d-2}. With such a ϱ, we can generalize the 2-dimensional construction so that the Chernoff point of X lies outside Δ^d_{d-1}. We now prove that such a ϱ exists. Let $\Delta^d_0 = \mathrm{conv}(X)$ be a regular d-simplex with edges of length $\sqrt{2}$ and center at the origin in $\mathbb{R}^d$. We need formulas for the Euclidean circumradius and height of Δ^d_0, and the Euclidean inradius of Δ^d_{d-1}. It is convenient to derive them for the standard d-simplex, which is the convex hull of the endpoints of the $d+1$ unit coordinate vectors of $\mathbb{R}^{d+1}$. We get the circumradius as the Euclidean distance between the vertices and the center at $(\frac{1}{d+1}, \frac{1}{d+1}, \ldots, \frac{1}{d+1})$:

$$R_d = \sqrt{\left(\tfrac{d}{d+1}\right)^2 + d\left(\tfrac{1}{d+1}\right)^2} = \sqrt{\tfrac{d(d+1)}{(d+1)^2}} = \sqrt{\tfrac{d}{d+1}}. \tag{13}$$

This radius is $d/(d+1)$ times the height of the standard simplex, which implies that the height is $H_d = (d+1)R_d/d = \sqrt{(d+1)/d}$. To compute the inradius of Δ^d_{d-1}, we observe that the Euclidean distance of the center of Δ^d_0 from a facet is $H_d/(d+1)$. Similarly, the Euclidean distance of a near facet of Δ^d_{d-1} from the parallel facet of Δ^d_0 is H_d/d. It follows that the inradius is

$$I_d = \left[\frac{1}{d} - \frac{1}{d+1}\right] H_d = \frac{1}{d(d+1)} H_d. \tag{14}$$

Consider the Euclidean ball with center x_i and radius $R_d - I_d$. By construction, it touches the $(d-1)$-st barycenter polytope of X at the center of one of its facets, which implies that it does not contain any of its vertices. Nevertheless, the ball contains the near facet of Δ^d_{d-2} in its interior, as we now prove.

▶ **Lemma 9.** $R_d - I_d > R_{d-2}$.

PROOF. Using (13) and (14), we simplify the expression for the difference on the left-hand side of the claimed inequality:

$$R_d - I_d = \sqrt{\tfrac{d}{d+1}} - \frac{1}{d(d+1)}\sqrt{\tfrac{d+1}{d}} = \frac{\sqrt{d+1}(d-1)}{\sqrt{d^3}}. \tag{15}$$

Dividing the claimed inequality by $R_{d-2} = \sqrt{(d-2)/(d-1)}$ and squaring, we get

$$\left[\frac{R_d - I_d}{R_{d-2}}\right]^2 = \left[\frac{\sqrt{d+1}(d-1)\sqrt{d-1}}{\sqrt{d^3}\sqrt{d-2}}\right]^2 = \frac{d^4 - 2d^3 + 2d - 1}{d^4 - 2d^3} > 1. \tag{16}$$

The claimed inequality follows. ◻

Finally note that the inradius of Δ^d_{d-1} is less than that of Δ^d_0: $I_d < J_d$. We can therefore choose $I_d < \varrho < \min\{J_d, R_d - R_{d-2}\}$, which is large enough so that $D(\varrho)$ is not contained in Δ^d_{d-1}, and it is small enough so that $D(\varrho)$ does not contain Δ^d_{d-1} and a touching Euclidean ball with center x_i contains the near facet of Δ^d_{d-2}.

7 Discussion

The contributions of this paper are geometric constraints on the location of the centers of smallest enclosing spheres for data in which dissimilarities are measured with Bregman divergences. The main tools used in their proofs are topological: the Nerve Theorem of Borsuk [2] and Leray [13] and the Fixed Point Lemma of Knaster, Kuratowski, and Mazurkiewicz [11]. Besides being of independent interest, the results are relevant to topological data analysis.

References

1 J.-D. Boissonnat, F. Nielsen and R. Nock. Bregman Voronoi diagrams. *Discrete Comput. Geom.* **44** (2010), 281–307.

2 K. Borsuk. On the imbedding of systems of compacta in simplicial complexes. *Fund. Math.* **35** (1948), 217–234.

3 L.M. Bregman. 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.* **7** (1967), 200–217.

4 G.N. Clements. The geometry of phonological features. *Phonology Yearbook* **2** (1985), 225–252.

5 A. Dhara and J. Dutta. *Optimality Conditions in Convex Optimization: a Finite-dimensional View.* CRC Press, Taylor & Francis Group, Boca Raton, Florida, 2012.

6 H. Edelsbrunner, Ž. Virk and H. Wagner. Topological data analysis in information space. Manuscript, IST Austria, Klosterneuburg, Austria, 2017.

7 H. Edelsbrunner and H. Wagner. Topological data analysis with Bregman divergences. *In* “Proc. 33rd Ann. Sympos. Comput. Geom., 2017”, 39:1–39:16.

8 J. Goldberger, S. Gordon and H. Greenspan. An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures. *In* “Proc. 9th IEEE Internat. Conf. Comput. Vision, 2003”, 487.

9 A. Huang. Similarity measures for text document clustering. *In* “Proc. 6th New Zealand Comput. Sci. Research Student Conf., 2008”, 49–56.

10 F. Itakura and S. Saito. An analysis-synthesis telephony based on the maximum likelihood method. *In* “Proc. 6th Internat. Congress Acoustics, 1968”, Tokyo, Japan, c17–c20.

11 B. Knaster, C. Kuratowski and S. Mazurkiewicz. Ein Beweis des Fixpunktsatzes für n-dimensionale Simplexe. *Fund. Math.* **14** (1929), 132–137.

12 S. Kullback and R.A. Leibler. On information and sufficiency. *Ann. Math. Stat.* **22** (1951), 79–86.

13 J. Leray. Sur la forme des espaces topologiques et sur les points fixes des représentations. *J. Math. Pure Appl.* **24** (1945), 95–167.

14 F. Nielsen. An information-geometric characterization of Chernoff information. *IEEE Signal Process. Lett.* **20** (2013), 269–272.

15 F. Nielsen and R. Nock. On the smallest enclosing information disk. *In* “Proc. 18th Canad. Conf. Comput. Geom., 2006”.

16 R. Nock, and F. Nielsen. Fitting the smallest enclosing Bregman ball. *In* “Proc. 16th European Conf. Machine Learning, 2005”, 649–656.

17 A. Rind, T.D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant and B. Shneiderman. Interactive information visualization to explore and query electronic health records. *Found. Trends in Human-Comput. Interaction* **5** (2011), 207–298.

18 R.T. Rockafellar. *Convex Analysis.* Princeton Univ. Press, Princeton, New Jersey, 1970.