Smallest Enclosing Spheres and Chernoff Points in Bregman Geometry

Herbert Edelsbrunner
IST Austria (Institute of Science and Technology Austria)
Am Campus 1, 3400 Klosterneuburg, Austria
edels@ist.ac.at

Žiga Virk
Faculty of Mathematics and Physics, University of Ljubljana
Jadranska ulica 19, 1000 Ljubljana, Slovenia
ziga.virk@fmf.uni-lj.si

Hubert Wagner
IST Austria (Institute of Science and Technology Austria)
Am Campus 1, 3400 Klosterneuburg, Austria
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. These bounds depend on the range of radii for which Bregman balls are convex.

2012 ACM Subject Classification Theory of computation → Computational geometry, Mathematics of computing → Combinatoric problems, Mathematics of computing → Algebraic topology

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

Digital Object Identifier 10.4230/LIPIcs.SoCG.2018.35

Funding This research is partially supported by the Office of Naval Research, through grant no. N62909-18-1-2038, and the DFG Collaborative Research Center TRR 109, ‘Discretization in Geometry and Dynamics’, through grant no. I02979-N35 of the Austrian Science Fund (FWF).

Acknowledgements The authors thank Arseniy Akopyan for fruitful discussions and an anonymous reviewer for his or her comments on the proof of Theorem 8 and the construction in Section 6.

1 Introduction
Interpreting non-geometric data geometrically is a standard step in data analysis. 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, dimension reduction, data visualization etc. These tools rely on a notion of dissimilarity between data points, and the Euclidean distance is often not ideal. Keeping in mind that a point often represents a histogram describing the corresponding non-geometric object, this is not surprising. 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.** We introduce the most important concepts studied in this paper before reviewing prior work and presenting our results. A function $F : \Omega \to \mathbb{R}$ 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 $\Omega$.

The combination of convexity and differentiability implies continuous differentiability; see [5, Theorem 2.86]. 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 $x$:

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

$$B_F(x; r) = \{ y \in \Omega \mid D_F(x\|y) \leq r \} ,$$  \hspace{1cm} (2)

$$B_F^*(x; r) = \{ y \in \Omega \mid D_F(y\|x) \leq r \} .$$  \hspace{1cm} (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 the fields of computational 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 the non-empty common intersections of primal Bregman balls with the corresponding Voronoi domains be contractible, a property proved in [7]. This opened the door to constructing filtrations of Bregman–Čech and Bregman–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 if 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 constructions. This contrasts the usual setting in which the computation of the Chernoff point for a single and possibly 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 from 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 $\Omega$, 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 $y$: $y_A = \arg\min_{a \in A} D(a \parallel y)$. We will make use of an extension of Pythagoras’ Theorem to Bregman divergences; see e.g. [1]. We give a proof for completeness.
Smallest Enclosing Spheres and Chernoff Points in Bregman Geometry

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

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

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

because the scalar product is necessarily non-negative. Substituting $F(a) = D(a\|y)$ and $F(y_A) = D(y_A\|y)$, 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 $G: \Omega \to \mathbb{R}$ 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 $\text{conv}(X)$ 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 $\text{conv}(X)$, 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 $A = \text{conv}(X)$ 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 $F$.  

\[\Box\]
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 $\Omega$. 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 $X \subseteq \mathbb{R}^d$ is unique and contained in aff $X$, for every $0 \leq k \leq d$.

Proof. Let $B^0(y;r)$ 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 $y \in \mathbb{R}^d$ such that these equalities are satisfied; see Figure 2. To get a contradiction, we set $A = \text{aff } X$ and assume $y \not\in A$. 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 $B^0(y;r)$. We get uniqueness because there is only one point in $A = \text{aff } X$ 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 $0 \leq j \leq k$, the $j$-th barycenter polytope of $X$ is

$$\Delta^j(X) = \text{conv}\{b_J \mid |J| = j + 1\}. \quad (5)$$

Note that $\Delta^0(X) = \text{conv}(X)$. In three dimensions, $\Delta^0$ is a tetrahedron, $\Delta^1$ is an octahedron, $\Delta^2$ is again a tetrahedron, and $\Delta^3$ 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 $\Delta_0^k \supseteq \Delta_1^k \supseteq \ldots \supseteq \Delta_k^k$.

![Figure 2](image-url) Assuming the center of the smallest circumsphere, $y$, does not lie in $A = \text{aff } X$ leads to a contradiction.
Numerical expressions:

\[
0 = \frac{1}{j+1} \sum_{\ell \in J} b_{J \setminus \{\ell\}} = \frac{1}{j+1} \sum_{\ell \in J} \left[ \frac{1}{j} \sum_{i \in J \setminus \{\ell\}} x_i \right] = \frac{1}{j+1} \sum_{\ell \in J} \sum_{i \in J \setminus \{\ell\}} x_i.
\]

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. Hence, \(\Delta^k_j \subseteq \Delta^k_{j-1}\) for \(1 \leq j \leq k\), as claimed.

**Face structure.** It is instructive to take a closer look at \(\Delta^3_1\), 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 \#facets(\(\Delta^k_j\)) for the number of facets of \(\Delta^k_j\).

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

\[
\#\text{facets}(\Delta^k_j) = \begin{cases} 
  k + 1 & \text{if } j = 0, k - 1, \\
  2k + 2 & \text{if } 1 \leq j \leq k - 2.
\end{cases}
\]

**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 \(e_k\) and note that this cube has \(2k + 2\) facets. For \(0 \leq j \leq k - 1\), 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 \(\Delta^k_j\), is the intersection of the \(k\)-plane with a facet of the cube, which implies the claimed number of facets of the barycenter polytope.

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. After scaling, 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 after scaling, 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 \(\Delta^k_j\) 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,

\[
\pi^k_j: \partial\text{conv}(X) \to \partial\Delta^k_j,
\]
The map $\pi^2_0$ is the identity on the boundary of the triangle. The map $\pi^2_1$ projects the vertices of $\text{conv}(X)$ to the midpoints of the edges of $\Delta^2_1$. The structure of $\partial\text{conv}(X)$ is recovered by gluing the half-edges in pairs at the shared endpoints.

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

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

### 4 Theorem in convex case

This section sharpens 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 $\Omega$. For each subset $J \subseteq \{0, 1, \ldots, k\}$, we recall that $X_J$ is the corresponding face of $X$, and we let $B^r(d_J; R_J)$ 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 points of faces of dimension $j = |J| - 1$:

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

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.

**Theorem 6 (Nesting for Convex Balls).** Let $F : \Omega \to \mathbb{R}$ be of Legendre type such that all 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 $\Delta_0 \supseteq \Delta_1 \supseteq \ldots \supseteq \Delta_k$. 

**Figure 3** Left: the map $\pi^2_0$ is the identity on the boundary of the triangle. Right: the map $\pi^2_1$ projects the vertices of $\text{conv}(X)$ to the midpoints of the edges of $\Delta^2_1$. The structure of $\partial\text{conv}(X)$ is recovered by gluing the half-edges in pairs at the shared endpoints.
To formalize this idea, let whose vertices lie on four of the triangles bounding the octahedron. and therefore centers of the smallest enclosing spheres of all center of the smallest enclosing sphere of a . It implies that the nerve also contains the . The implies that the nerve of the sets is contained in the convex hull of the Chernoff subdivision of the tetrahedron is the pink octahedron whose vertices are . If all Chernoff points are different, the two subdivisions are.

**Proof.** We prove by showing that the Chernoff point of every -face of is contained in the convex hull of the Chernoff points of the -faces of this -face. Indeed, this implies that the vertices of lie in , and the claimed inclusion follows. To formalize this idea, let , set , and consider for every . We prove that belongs to the convex hull of in two steps.

For the first step, we write for the Chernoff subdivision of . 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 contains the point , it also contains all points with . Indeed, if the balls with have a point in common, then so do the balls with . By convexity, contains all simplices in that share . Similarly, contains all simplices in that share , for each . It follows that the balls cover the entire Chernoff subdivision of and thus the entire -face:

\[
\text{conv}(X_J) \subseteq \bigcup_{\ell=0}^j B(x_{\ell}; R_J).
\]  

(10)

For the second step, we write for the convex hull of the points . It is the -st Chernoff polytope of and necessarily contractible. Define , for each . By Lemma 2, the points belong to , so and (10) implies that the sets have the same homotopy type as and is therefore contractible. The -fold intersections are all non-empty, as witnessed by the vertices of . Hence, the nerve contains the boundary complex of a -simplex. Contractibility thus implies that the nerve also contains the -simplex. In other words, for and therefore .

We note that Theorem 6 tightens Lemma 2 in the case of convex Bregman balls: the center of the smallest enclosing sphere of a -simplex is contained in the convex hull of the centers of the smallest enclosing spheres of all -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 $\text{conv}(X)$. 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 primal Bregman 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 $F: \Omega \to \mathbb{R}$ be of Legendre type such that all primal Bregman balls of radius $r \leq R_j$ are convex, and let $\Delta_0, \Delta_1, \ldots, \Delta_k$ be the Chernoff polytopes of a $k$-simplex $X \subseteq \Omega$. Then $\Delta_0 \supseteq \Delta_1 \supseteq \cdots \supseteq \Delta_j \supseteq \Delta_{j+1}, \Delta_{j+2}, \ldots, \Delta_k$.

Proof. To prove the inclusions $\Delta_0 \supseteq \Delta_1 \supseteq \cdots \supseteq \Delta_j$, 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

$$\text{conv}(X_J) \subseteq \bigcup_{t \in J} B(x_t; R_J)$$

(11)

for every $J \subseteq \{0, 1, \ldots, k\}$ 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 $J \setminus \{i\}$, $0 \leq j < i$. By definition of $R_J$, the $j + 1$ balls $B(x_t; R_J)$ 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. By the Nerve Theorem [2, 13], the union of the sliced balls is contractible as well. But this union covers the $(j - 1)$-dimensional boundary of $\text{conv}(X_J)$, so it must also cover $\text{conv}(X_J)$ to be contractible. Hence (11) follows, and in particular $\text{conv}(X) \subseteq \bigcup_{t=0}^k B(x_t; R_k)$.
For the second step, recall that \( \Delta_j = \Delta_j(X) \) is the \( j \)-th Chernoff polytope of \( X \). Define \( C_\ell = B(x_\ell; R_\ell) \cap \Delta_j \), for \( 0 \leq \ell \leq k \). Since \( \Delta_j \subseteq \text{conv}(X) \), the balls \( B(x_\ell; R_\ell) \) cover \( \Delta_j \). By the Nerve Theorem, the nerve of the sets \( C_\ell \) has the same homotopy type as \( \Delta_j \) and is therefore contractible. To apply Proposition 7 to \( \Delta_j \), we first interpret \( \Delta_j \) as the homeomorphic image of a \( k \)-simplex. Assuming \( \Delta_j \) is \( k \)-dimensional, we decompose its boundary by central projection of the boundary of \( \text{conv}(X) \), in which we use any point in the interior of \( \Delta_j \) as center; see Figures 3 and 4 for illustrations. It is possible that the dimension of \( \Delta_j \) is less than \( k \), namely when some \( j \)-faces of \( \text{conv}(X) \) have coincident Chernoff points. We can perturb the coincident Chernoff points slightly and continue the proof with the perturbed \( \Delta_j \), which is now \( k \)-dimensional. In either case, we denote the topological \( k \)-simplex by \( \tilde{\Delta}_j \).

Recall that the facets of \( \Delta_j \) 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 \( \Delta_j \). 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 \( \Phi_\ell \) consists of the far facet of \( x_\ell \) — which is contained in \( \text{conv}(X_\ell) \) — together with pieces of the near facets of the vertices \( x_\ell \in X_\ell \). The far facet of \( x_\ell \) is covered by the balls \( B(x_\ell; R_\ell) \), for \( i \neq \ell \), as a consequence of (11). Furthermore, the near facet of \( x_\ell \) is covered by \( B(x_\ell; R_\ell) \) simply because \( B(x_\ell; R_\ell) \subseteq B(x_\ell; R_k) \), and the former ball is convex and contains the relevant vertices of \( \Delta_j \). It follows that the \( \Phi_\ell \) is covered by the balls \( B(x_\ell; R_\ell) \), for \( i \neq \ell \). Hence, \( \tilde{\Delta}_j \) and the sets \( C_\ell \) satisfy the assumptions of Proposition 7. The proposition thus implies that the common intersection of the \( C_\ell \) is non-empty. This intersection can only be the Chernoff point of \( X \), which we therefore conclude lies inside \( \Delta_j \), 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 \( \Delta_0 \) 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 \( \Delta_{j+1} \). To this end, we construct a function of Legendre type, \( F : \mathbb{R}^d \to \mathbb{R} \), 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 \( \Delta_{j+1} \).

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 first simplify the task by requiring that \( F \) be convex but not necessarily differentiable and not necessarily strictly convex. Appropriate small bump functions are used to eventually turn the convex function into a Legendre type function.

#### Two-dimensional construction.
Let \( \Delta_0^2 = \text{conv}(X) \) 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 \( \Delta_1^2 = \text{conv}(Y) \) 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 \( \Delta_1^2 \) and \( \Delta_0^2 \), we choose a radius strictly between them, \( 1/\sqrt{24} < \rho < 1/\sqrt{6} \), and we let \( D(\rho) \) be the (Euclidean) disk with this radius and center at the origin. As illustrated in Figure 5, \( D(\rho) \) is neither contained in \( \Delta_1^2 \) nor
Figure 5 In the nonconvex case, we can arrange that the Chernoff point of $X$ lies outside the triangle spanned by the Chernoff points of the edges of $X$.

does it contain $\Delta^2_{1}$. We finally construct $F : \mathbb{R}^2 \to \mathbb{R}$ 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 $B_F(x_i; r)$ 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

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

(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 $z_0 \in \text{int} D(\varrho) \setminus \Delta^2_{1}$, and we can do this without violating convexity and without changing $F$ outside this disk. We can turn $F$ into a differentiable and strictly convex function by substituting slightly curved arcs for the generating lines of the cone and by rounding off the sharp corners at the apex and the circle at which the cone meets the paraboloid. With these modifications, we get $z_0$ as the Chernoff point of $X$, which by construction lies outside $\Delta^2_{1}$.

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 $\Delta^d_{d-1}$ 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 $\Delta^d_{d-2}$. With such a $\varrho$, we can generalize the 2-dimensional construction so that the Chernoff point of $X$ lies outside $\Delta^d_{d-1}$. We now prove that such a $\varrho$ exists. Let $\Delta^d = \text{conv}(X)$ 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 $\Delta^d$, and the Euclidean inradius of $\Delta^d_{d-1}$. It is convenient to derive them for the standard $d$-simplex, which is the convex hull of the endpoints of the $d + 1$
Smallest Enclosing Spheres and Chernoff Points in Bregman Geometry

unit coordinate vectors of $\mathbb{R}^{d+1}$. We get the circumradius as the Euclidean distance between the vertices and the center at $(\frac{1}{d+1}, \frac{1}{d+1}, \ldots, \frac{1}{d+1})$:

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

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 $\Delta_{d-1}^d$, we observe that the Euclidean distance of the center of $\Delta_{d-1}^d$ from a facet is $H_d/d$. Similarly, the Euclidean distance between parallel facets of $\Delta_{d-1}^d$ and $\Delta_0^d$ is $H_d/d$. It follows that the inradius is

$$ I_d = \frac{1}{d - \frac{1}{d+1}} H_d = \frac{1}{d(d+1)} H_d. \quad (14) $$

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 barycenters of the $(d-2)$-faces of $\Delta_{d-1}^d$ incident to $x_i$ and therefore the entire near facet of $\Delta_{d-2}^d$, as we now prove. Since $\Delta_{d-2}^d$ is a regular simplex, the distance between its barycenter and its vertices is $R_{d-2}$.

**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:

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

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

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

The claimed inequality follows. \hfill \blacksquare

Finally note that the inradius of $\Delta_{d-1}^d$ is less than that of $\Delta_0^d$: $I_d < J_d$. We can therefore choose $I_d < \rho < \min\{J_d, R_d - R_{d-2}\}$, which is large enough so that $D(\rho)$ is not contained in $\Delta_{d-1}^d$, and it is small enough so that $D(\rho)$ does not contain $\Delta_{d-1}^d$ and a touching Euclidean ball with center $x_i$ contains the near facet of $\Delta_{d-2}^d$.

# 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