Discrete Comput Geom
DOI 10.1007/s00454-010-9256-1

Bregman Voronoi Diagrams

Jean-Daniel Boissonnat · Frank Nielsen · Richard Nock

Received: 12 September 2008 / Revised: 17 February 2010 / Accepted: 17 February 2010

Abstract The Voronoi diagram of a finite set of objects is a fundamental geometric structure that subdivides the embedding space into regions, each region consisting of the points that are closer to a given object than to the others. We may define various variants of Voronoi diagrams depending on the class of objects, the distance function and the embedding space. In this paper, we investigate a framework for defining and building Voronoi diagrams for a broad class of distance functions called Bregman divergences. Bregman divergences include not only the traditional (squared) Euclidean distance but also various divergence measures based on entropic functions. Accordingly, Bregman Voronoi diagrams allow one to define information-theoretic Voronoi diagrams in statistical parametric spaces based on the relative entropy of distributions. We define several types of Bregman diagrams, establish correspondences between those diagrams (using the Legendre transformation), and show how to compute them efficiently. We also introduce extensions of these diagrams, e.g., k-order and k-bag Bregman Voronoi diagrams, and introduce Bregman triangulations of a set of points and their connection with Bregman Voronoi diagrams. We show that these triangulations capture many of the properties of the celebrated Delaunay triangulation.

A preliminary version appeared in the 18th ACM-SIAM Symposium on Discrete Algorithms, pp. 746–755, 2007. Related materials including demos and videos are available online at http://www.csl.sony.co.jp/person/nielsen/BregmanVoronoi/.

J.-D. Boissonnat (✉)
GEOMETRICA, INRIA Sophia-Antipolis, Sophia-Antipolis, France
e-mail: Jean-Daniel.Boissonnat@sophia.inria.fr

F. Nielsen
Fundamental Research Laboratory, Sony Computer Science Laboratories Inc., Tokyo, Japan

F. Nielsen
École Polytechnique, LIX, Palaiseau, France

R. Nock
CEREGMIA, Université Antilles-Guyane, Schoelcher, France

Published online: 02 April 2010

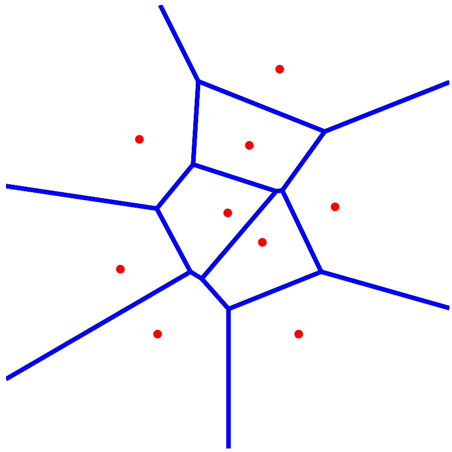


Fig. 1 Ordinary Euclidean Voronoi diagram of a given set $\mathcal{S}$ of nine sites

Keywords Computational Information Geometry · Voronoi diagram · Delaunay triangulation · Bregman divergence · Bregman ball · Legendre transformation

1 Introduction and Prior Work

The *Voronoi diagram* $\mathrm{vor}(\mathcal{S})$ of a set of n points $\mathcal{S} = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ of the d-dimensional space $\mathbb{R}^d$ is defined as the *cell complex* whose d-cells are the *Voronoi regions* $\{\mathrm{vor}(\mathbf{p}_i)\}_{i \in \{1,\ldots,n\}}$ where $\mathrm{vor}(\mathbf{p}_i)$ is the set of points of $\mathbb{R}^d$ closer to $\mathbf{p}_i$ than to any other point of $\mathcal{S}$ with respect to a *distance function* δ:

$$\mathrm{vor}(\mathbf{p}_i) \stackrel{\mathrm{def}}{=} \left\{\mathbf{x} \in \mathbb{R}^d \mid \delta(\mathbf{p}_i, \mathbf{x}) \le \delta(\mathbf{p}_j, \mathbf{x})\ \forall \mathbf{p}_j \in \mathcal{S}\right\}.$$

Points $\{\mathbf{p}_i\}_i$ are called the *Voronoi sites* or *Voronoi generators*. Since its inception in disguise by Descartes in the seventeenth century [20], the Voronoi diagram has found a broad spectrum of applications in science. Computational geometers have focused at first on *Euclidean* Voronoi diagrams [5] by considering the case where $\delta(\mathbf{x}, \mathbf{y})$ is the Euclidean distance $\|\mathbf{x}-\mathbf{y}\| = \sqrt{\sum_{i=1}^d (x_i - y_i)^2}$. Voronoi diagrams have been later on defined and studied for other distance functions, most notably the L_1 distance $\|\mathbf{x} - \mathbf{y}\|_1 = \sum_{i=1}^d |x_i - y_i|$ (Manhattan distance) and the L_∞ distance $\|\mathbf{x} - \mathbf{y}\|_\infty = \max_{i \in \{1,\ldots,d\}} |x_i - y_i|$ [5, 8]. Klein further presented an abstract framework for describing and computing the fundamental structures of abstract Voronoi diagrams [9, 24].

In Artificial Intelligence, Machine Learning techniques also rely on geometric concepts for building classifiers in supervised problems (e.g., linear separators, oblique decision trees, etc.) or *clustering data* in unsupervised settings (e.g., k-means, support vector clustering [7], etc.). However, the considered data sets $\mathcal{S}$ and their underlying spaces $\mathcal{X}$ are sometimes not metric spaces. The notion of distance between two elements of $\mathcal{X}$ needs to be replaced by a *pseudo-distance* that is not necessarily symmetric and may not satisfy the *triangle inequality*. Such a pseudo-distance is also referred to as a *distortion*, a (*dis*)*similarity* or a *divergence* in the literature. For example, in parametric statistical spaces, a vector point represents a distribution and its coordinates store the parameters of the associated distribution. A notion of "distance" between two such points is then needed to represent the divergence between the corresponding distributions.

Very few works have tackled an in-depth study of Voronoi diagrams and their applications for such a kind of statistical spaces. This is important even for ordinary Voronoi diagrams as Euclidean point location of sites are usually *observed* in *noisy* environments (e.g., imprecise point measures in computer vision experiments), and "noise" is often modeled by means of normal distributions (so-called "Gaussian noise"). To the best of our knowledge, statistical Voronoi diagrams have only been considered in a 4-page short paper of Onishi and Imai [33] which relies on Kullback–Leibler divergence of d-dimensional multivariate normal distributions to study combinatorics of their Voronoi diagrams, and subsequently in a 2-page video paper of Sadakane et al. [37] which defines the divergence implied by a convex function and its conjugate, and presents the Voronoi diagram via techniques of information geometry [1] (see also [34] and related short communications [22, 23]). Our study of Bregman Voronoi diagrams generalizes and subsumes these preliminary studies using an easier concept of divergence, namely the concept of Bregman divergences [6, 11] that does not rely *explicitly* on convex conjugates. Bregman divergences encapsulate the squared Euclidean distance and many widely used divergences, e.g., the Kullback–Leibler divergence. It should be noticed, however, that other statistical metric distances (called Rao's distances [2]) have been defined and studied in the context of Riemannian geometry [1]. Sacrificing some generality, while not very restrictive in practice, allows a much simpler treatment; in particular, our study of Bregman divergences is elementary and does not rely on Riemannian geometry.

In this paper, we give a thorough treatment of Bregman Voronoi diagrams which elegantly *unifies* the ordinary Euclidean Voronoi diagram and statistical Voronoi diagrams. Our contributions are summarized as follows:

- Since Bregman divergences are not symmetric, we define *two types* of Bregman Voronoi diagrams. One is an affine diagram with convex polyhedral cells, while the other one is curved. The cells of those two diagrams are in 1–1 correspondence through the Legendre transformation.
- We present a simple way to compute the Bregman Voronoi diagram of a set of points by lifting the points into a higher dimensional space. This mapping leads also to combinatorial bounds on the size of these diagrams. We also define weighted Bregman Voronoi diagrams and show that the class of these diagrams is identical to the class of affine (or power) diagrams. Special cases of weighted Bregman Voronoi diagrams are the k-order and k-bag Bregman Voronoi diagrams.
- We define the Bregman Delaunay triangulation of a set of points. This structure captures some of the most important properties of the well-known Delaunay triangulation. In particular, this triangulation is the geometric dual of the first-type Bregman Voronoi diagram of its vertices.

The outline of the paper is as follows: In Sect. 2, we define Bregman divergences and recall some of their basic properties. In Sect. 3, we study the geometry of Bregman spaces and characterize bisectors, balls and geodesics. Section 4 is devoted to

Bregman Voronoi diagrams and Sect. 5 to Bregman triangulations. Finally, Sect. 6 concludes the paper and mention further ongoing investigations.

Notations In the whole paper, $\mathcal{X}$ denotes an open convex domain of $\mathbb{R}^d$ and $F : \mathcal{X} \mapsto \mathbb{R}$ a strictly convex and differentiable function. $\mathcal{F}$ denotes the graph of F, i.e., the set of points $(\mathbf{x}, z) \in \mathcal{X} \times \mathbb{R}$ where $z = F(\mathbf{x})$. We write $\hat{\mathbf{x}}$ for the point $(\mathbf{x}, F(\mathbf{x})) \in \mathcal{F}$. ∇F, $\nabla^2 F$ and $(\nabla F)^{-1}$ denote respectively the gradient, the Hessian and the inverse gradient of F.

2 Bregman Divergences

In this section, we recall the definition of Bregman[1] divergences and some of their main properties (Sect. 2.1). We show that the notion of Bregman divergence encapsulates the squared Euclidean distance as well as several well-known information-theoretic divergences. We also introduce the notion of dual divergences (Sect. 2.2). Further results can be found in [6, 11, 18].

2.1 Definition and Basic Properties

Let $\mathcal{X}$ be an open convex subset of $\mathbb{R}^d$, and let F be a strictly convex and differentiable real-valued function defined on $\mathcal{X}$. For any two points $\mathbf{p} = (p_1, \dots, p_d)$ and $\mathbf{q} = (q_1, \dots, q_d)$ of $\mathcal{X}$, the Bregman divergence $D_F(\cdot \parallel \cdot)$ of $\mathbf{p}$ to $\mathbf{q}$ associated to F (which is called the *generator function* of the divergence) is defined [11, 13] as

$$D_F : \mathcal{X} \times \mathcal{X} \mapsto [0, +\infty)$$

$$D_F\big(\mathbf{p} \parallel \mathbf{q}\big) \stackrel{\text{def}}{=} F(\mathbf{p}) - F(\mathbf{q}) - \big\langle \mathbf{p} - \mathbf{q}, \nabla F(\mathbf{q})\big\rangle, \tag{1}$$

where $\nabla F = [\frac{\partial F}{\partial x_1} \ \cdots \ \frac{\partial F}{\partial x_d}]^T$ denotes the gradient operator, and $\langle \mathbf{p}, \mathbf{q}\rangle$ the inner (or dot) product, $\sum_{i=1}^d p_i q_i$. Informally speaking, Bregman divergence D_F is the *tail* of the Taylor expansion of F and has a nice geometric interpretation. Indeed, let $\mathcal{F} : z = F(\mathbf{x})$ be the graph of F and let $H_{\mathbf{q}}$ be the hyperplane tangent to $\mathcal{F}$ at point $\hat{\mathbf{q}} = (\mathbf{q}, F(\mathbf{q}))$. Since $H_{\mathbf{q}}$ is given by $z = H_{\mathbf{q}}(\mathbf{x}) = F(\mathbf{q}) + \langle \nabla F(\mathbf{q}), \mathbf{x} - \mathbf{q}\rangle$, we have $D_F(\mathbf{p} \parallel \mathbf{q}) = F(\mathbf{p}) - H_{\mathbf{q}}(\mathbf{p})$ (see Fig. 2).

Lemma 1 *The Bregman divergence $D_F(\mathbf{p} \parallel \mathbf{q})$ is geometrically measured as the vertical distance between $\hat{\mathbf{p}}$ and the hyperplane $H_{\mathbf{q}}$ tangent to $\mathcal{F}$ at point $\hat{\mathbf{q}}$.*

Observe that, for most functions F, the associated Bregman divergence is *not* symmetric, i.e., $D_F(\mathbf{p} \parallel \mathbf{q}) \neq D_F(\mathbf{q} \parallel \mathbf{p})$ (the symbol $\parallel$ is put to emphasize this point, as is standard in information theory).

We now recall some well-known properties of Bregman divergences.

[1] Lev M. Bregman historically pioneered this notion in the seminal work [11] on minimization of a convex objective function under linear constraints. See http://www.math.bgu.ac.il/serv/segel/bregman.html. The seminal paper is available online at http://www.lix.polytechnique.fr/Labo/Frank.Nielsen/.

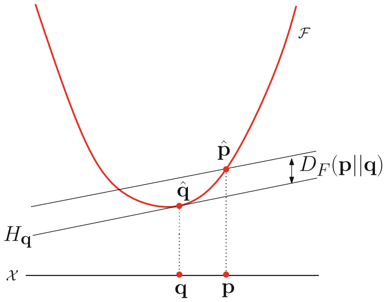


Fig. 2 Visualizing the Bregman divergence. $D_F(\cdot \parallel \mathbf{q})$ is the vertical distance between $\mathcal{F}$ and the hyperplane tangent to $\mathcal{F}$ at $\hat{\mathbf{q}}$

Property 1 (Non-negativity) *The strict convexity of generator function F implies that, for any $\mathbf{p}$ and $\mathbf{q}$ in $\mathcal{X}$, $D_F(\mathbf{p} \parallel \mathbf{q}) \geq 0$, with $D_F(\mathbf{p} \parallel \mathbf{q}) = 0$ if and only if $\mathbf{p} = \mathbf{q}$.*

Property 2 (Convexity) *Function $D_F(\mathbf{p} \parallel \mathbf{q})$ is convex in its first argument $\mathbf{p}$ but not necessarily in its second argument $\mathbf{q}$.*

Because positive linear combinations of strictly convex and differentiable functions are strictly convex and differentiable functions, new generator functions (and corresponding Bregman divergences) can also be built as positive linear combinations of elementary generator functions. The following property is important as it allows to handle mixed data sets of heterogenous types in a unified framework.

Property 3 (Linearity) *Bregman divergence is a linear operator, i.e., for any two strictly convex and differentiable functions F_1 and F_2 defined on $\mathcal{X}$ and for any $\lambda \geq 0$:*

$$D_{F_1+\lambda F_2}(\mathbf{p} \parallel \mathbf{q}) = D_{F_1}(\mathbf{p} \parallel \mathbf{q}) + \lambda D_{F_2}(\mathbf{p} \parallel \mathbf{q}).$$

Property 4 (Invariance under linear transforms) *$G(\mathbf{x}) = F(\mathbf{x}) + \langle \mathbf{a}, \mathbf{x} \rangle + b$, with $\mathbf{a} \in \mathbb{R}^d$ and $b \in \mathbb{R}$, is a strictly convex and differentiable function on $\mathcal{X}$, and $D_G(\mathbf{p} \parallel \mathbf{q}) = D_F(\mathbf{p} \parallel \mathbf{q})$.*

Examples of Bregman divergences are the squared Euclidean distance (obtained for $F(\mathbf{x}) = \|\mathbf{x}\|^2$) and the generalized quadratic pseudo distance function $D_F(\mathbf{p} \parallel \mathbf{q}) = (\mathbf{p} - \mathbf{q})^T \mathbf{Q}(\mathbf{p} - \mathbf{q})$ where $\mathbf{Q}$ is a positive definite symmetric matrix (obtained for $F(\mathbf{x}) = \mathbf{x}^T \mathbf{Q}\mathbf{x}$). When $\mathbf{Q}$ is taken to be the inverse of the variance–covariance matrix of some data set, D_F is the Mahalanobis distance, extensively used in Computer Vision and Data Mining. More importantly, the notion of Bregman divergence encapsulates various information measures based on entropic functions such as the Kullback–Leibler divergence based on the (unnormalized) Shannon entropy, or the Itakura–Saito divergence based on Burg entropy (commonly used in sound processing). Table 1 lists the main univariate Bregman divergences. Finally, we would like to point out that Banerjee et al. [6] have shown that there is a *bijection* between the regular exponential families in statistics [29] and a subset of the Bregman divergences called *regular Bregman divergences*.

Table 1 Some common univariate Bregman divergences D_F

Dom. $\mathcal{X}$	Function F	Gradient	Inv. grad.	Divergence $D_F(p \parallel q)$
$\mathbb{R}$	Squared function x^2	$2x$	$\frac{x}{2}$	Squared loss (norm) $(p-q)^2$
$\mathbb{R}^+$	Unnorm. Shannon entropy $x\log x - x$	$\log x$	$\exp x$	Kullback–Leibler div. (I-div.) $p\log\frac{p}{q} - p + q$
$\mathbb{R}$	Exponential $\exp x$	$\exp x$	$\log x$	Exponential loss $\exp p - (p-q+1)\exp q$
$\mathbb{R}^{+}*$	Burg entropy $-\log x$	$-\frac{1}{x}$	$-\frac{1}{x}$	Itakura–Saito divergence $\frac{p}{q} - \log\frac{p}{q} - 1$
[0, 1]	Bit entropy $x\log x + (1-x)\log(1-x)$	$\log\frac{x}{1-x}$	$\frac{\exp x}{1+\exp x}$	Logistic loss $p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$
$\mathbb{R}$	Dual bit entropy $\log(1+\exp x)$	$\frac{\exp x}{1+\exp x}$	$\log\frac{x}{1-x}$	Dual logistic loss $\log\frac{1+\exp p}{1+\exp q} - (p-q)\frac{\exp q}{1+\exp q}$
$[-1, 1]$	Hellinger-like $-\sqrt{1-x^2}$	$\frac{x}{\sqrt{1-x^2}}$	$\frac{x}{\sqrt{1+x^2}}$	Hellinger-like $\frac{1-pq}{\sqrt{1-q^2}} - \sqrt{1-p^2}$

2.2 Legendre Duality

We now turn to an essential notion of convex analysis, the Legendre transform. Legendre transform allows one to associate to any Bregman divergence a dual Bregman divergence.

Let F be a strictly convex and differentiable real-valued function on $\mathcal{X}$. The Legendre transformation associates to F a *convex conjugate* function $F^* : \mathbb{R}^d \mapsto \mathbb{R}$ given by [36]

$$F^*(\mathbf{x}') = \sup_{\mathbf{x}\in\mathcal{X}} \{\langle \mathbf{x}', \mathbf{x}\rangle - F(\mathbf{x})\}.$$

$\mathbf{x}'$ is called the dual variable. The supremum is reached at the *unique* point where the gradient of $G(\mathbf{x}) = \langle \mathbf{x}', \mathbf{x}\rangle - F(\mathbf{x})$ vanishes or, equivalently, when $\mathbf{x}' = \nabla F(\mathbf{x})$.

In the sequel, we will denote $\nabla F(\mathbf{x})$ by $\mathbf{x}'$, omitting the F in the notation as it should be clear from the context. Writing $\mathcal{X}'$ for the *gradient space* $\{\nabla F(\mathbf{x}) | \mathbf{x} \in \mathcal{X}\}$, the convex conjugate F^* of F is the real-valued function defined on $\mathcal{X}' \subset \mathbb{R}^d$

$$F^*(\mathbf{x}') = \langle \mathbf{x}, \mathbf{x}'\rangle - F(\mathbf{x}). \tag{2}$$

Figure 3 gives a geometric interpretation of the Legendre transformation. Consider the hyperplane $H_\mathbf{x}$ tangent to $\mathcal{F}$ at $\hat{\mathbf{x}}$. This hyperplane intersects the z axis at the point $(\mathbf{0}, -F^*(\mathbf{x}'))$. Indeed, the equation of $H_\mathbf{x}$ is $z(\mathbf{y}) = \langle \mathbf{x}', \mathbf{y} - \mathbf{x}\rangle + F(\mathbf{x}) = \langle \mathbf{x}', \mathbf{y}\rangle - F^*(\mathbf{x}')$. Hence, the z-intercept of $H_\mathbf{x}$ is equal to $-F^*(\mathbf{x}')$. Any hyperplane passing through another point of $\mathcal{F}$ and parallel to $H_\mathbf{x}$ necessarily intersects the z-axis above $-F^*(\mathbf{x}')$.

Since F is a strictly convex and differentiable real-valued function on $\mathcal{X}$, its gradient ∇F is well defined as well as its inverse $(\nabla F)^{-1}$, and $\nabla F \circ (\nabla F)^{-1} =$

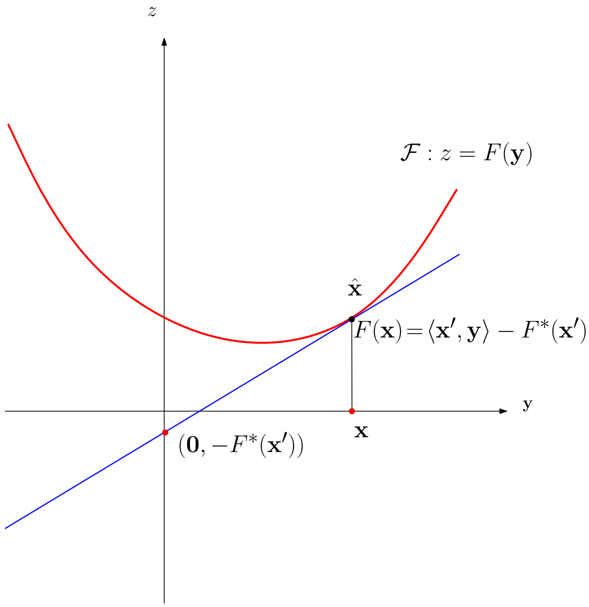


Fig. 3 The z-intercept $(\mathbf{0}, -F^*(\mathbf{x}'))$ of the tangent hyperplane $H_{\mathbf{x}}$ of $\mathcal{F}$ at $\hat{\mathbf{x}}$ defines the value of the Legendre transform F^* for the dual coordinate $\mathbf{x}' = \nabla F(\mathbf{x})$

$(\nabla F)^{-1} \circ \nabla F$ is the identity map. Taking the derivative of (2), we get

$$\langle \nabla F^*(\mathbf{x}'), \mathrm{d}\mathbf{x}' \rangle = \langle \mathbf{x}, \mathrm{d}\mathbf{x}' \rangle + \langle \mathbf{x}', \mathrm{d}\mathbf{x} \rangle - \langle \nabla F(\mathbf{x}), \mathrm{d}\mathbf{x} \rangle = \langle \mathbf{x}, \mathrm{d}\mathbf{x}' \rangle = \langle (\nabla F)^{-1}(\mathbf{x}'), \mathrm{d}\mathbf{x}' \rangle,$$

from which we deduce that $\nabla F^* = (\nabla F)^{-1}$.

The above discussion shows that D_{F^*} is a Bregman divergence, which we call the *Legendre dual divergence* of D_F. We have

Lemma 2 $D_F(\mathbf{p} \,\|\, \mathbf{q}) = F(\mathbf{p}) + F^*(\mathbf{q}') - \langle \mathbf{p}, \mathbf{q}' \rangle = D_{F^*}(\mathbf{q}' \,\|\, \mathbf{p}')$.

Proof By (1), $D_F(\mathbf{p} \,\|\, \mathbf{q}) = F(\mathbf{p}) - F(\mathbf{q}) - \langle \mathbf{p} - \mathbf{q}, \mathbf{q}' \rangle$, and, according to (2), we have $F(\mathbf{p}) = \langle \mathbf{p}', \mathbf{p} \rangle - F^*(\mathbf{p}')$ and $F(\mathbf{q}) = \langle \mathbf{q}', \mathbf{q} \rangle - F^*(\mathbf{q}')$. Hence, $D_F(\mathbf{p} \,\|\, \mathbf{q}) = \langle \mathbf{p}', \mathbf{p} \rangle - F^*(\mathbf{p}') - \langle \mathbf{p}, \mathbf{q}' \rangle + F^*(\mathbf{q}') = D_{F^*}(\mathbf{q}' \,\|\, \mathbf{p}')$ since $\mathbf{p} = \nabla F^{-1} \nabla F(\mathbf{p}) = \nabla F^*(\mathbf{p}')$. □

Observe that, when D_F is symmetric, D_{F^*} is also symmetric.

The Legendre transform of the quadratic form $F(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{Q} \mathbf{x}$, where $\mathbf{Q}$ is a symmetric invertible matrix, is $F^*(\mathbf{x}') = \frac{1}{2}\mathbf{x}'^T \mathbf{Q}^{-1} \mathbf{x}'$. Observe that the corresponding divergences D_F and D_{F^*} are both generalized quadratic distances.

To compute F^*, we use the fact that $\nabla F^* = (\nabla F)^{-1}$ and obtain F^* as $F^* = \int (\nabla F)^{-1}$. For example, the Hellinger-like measure is obtained by setting $F(x) = -\sqrt{1 - x^2}$ (see Table 1). The inverse gradient is $\frac{x}{\sqrt{1+x^2}}$ and the dual convex conjugate is $\int \frac{x \mathrm{d}x}{\sqrt{1+x^2}} = \sqrt{1 + x^2}$. Integrating functions symbolically may be difficult or even not possible, and, in some cases, it will be required to approximate numerically the inverse gradient $(\nabla F)^{-1}(\mathbf{x})$.

Let us consider the univariate generator functions defining the divergences of Table 1. Both the squared function $F(x) = \frac{1}{2}x^2$ and Burg entropy $F(x) = -\log x$ are *self-dual*, i.e., $F = F^*$. This is easily seen by noticing that the gradient and inverse gradient are identical.

For the exponential function $F(x) = \exp x$, we have $F^*(y) = y \log y - y$ (the unnormalized Shannon entropy) and for the dual bit entropy $F(x) = \log(1 + \exp x)$, we have $F^*(y) = y \log \frac{y}{1-y} + \log(1 - y)$, the bit entropy. Note that the bit entropy function is a particular Bregman generator satisfying $F(x) = F(1 - x)$.

3 Elements of Bregman Geometry

In this section, we discuss several basic geometric properties that will be useful when studying Bregman Voronoi diagrams. Since Bregman divergences are not symmetric, we describe two types of Bregman bisectors in Sect. 3.1. We subsequently characterize Bregman balls by using a lifting transform that extends a construction well-known in the Euclidean case (Sect. 3.2). Finally, we show an orthogonality property between bisectors and geodesics in Sect. 3.3.

3.1 Bregman Bisectors

Since Bregman divergences are not symmetric, we can define two types of bisectors. The Bregman bisector of the *first type* is defined as

$$BB_F(\mathbf{p}, \mathbf{q}) = \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{x} \,\|\, \mathbf{p}) = D_F(\mathbf{x} \,\|\, \mathbf{q})\}.$$

Similarly, we define the Bregman bisector of the *second type* as

$$BB'_F(\mathbf{p}, \mathbf{q}) = \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{p} \,\|\, \mathbf{x}) = D_F(\mathbf{q} \,\|\, \mathbf{x})\}.$$

These bisectors are identical when the divergence is symmetric. However, in general, they are distinct. As Lemma 3 below shows, the bisectors of the first type are hyperplanes while the bisectors of the second type are potentially curved (but always linear in the gradient space, hence the notation).

Lemma 3 *The Bregman bisector of the first type $BB_F(\mathbf{p}, \mathbf{q})$ is the hyperplane of equation*

$$BB_F(\mathbf{p}, \mathbf{q}, \mathbf{x}) = 0 \quad \text{where } BB_F(\mathbf{p}, \mathbf{q}, \mathbf{x}) = \langle \mathbf{x}, \mathbf{p}' - \mathbf{q}' \rangle + F(\mathbf{p}) - \langle \mathbf{p}, \mathbf{p}' \rangle - F(\mathbf{q}) + \langle \mathbf{q}, \mathbf{q}' \rangle.$$

The Bregman bisector of the second type $BB'_F(\mathbf{p}, \mathbf{q})$ is the hypersurface of equation

$$BB'_F(\mathbf{p}, \mathbf{q}, \mathbf{x}) = 0 \quad \text{where } BB'_F(\mathbf{p}, \mathbf{q}, \mathbf{x}) \langle \mathbf{x}', \mathbf{q} - \mathbf{p} \rangle + F(\mathbf{p}) - F(\mathbf{q}) = 0$$

(*a hyperplane in the gradient space $\mathcal{X}'$*).

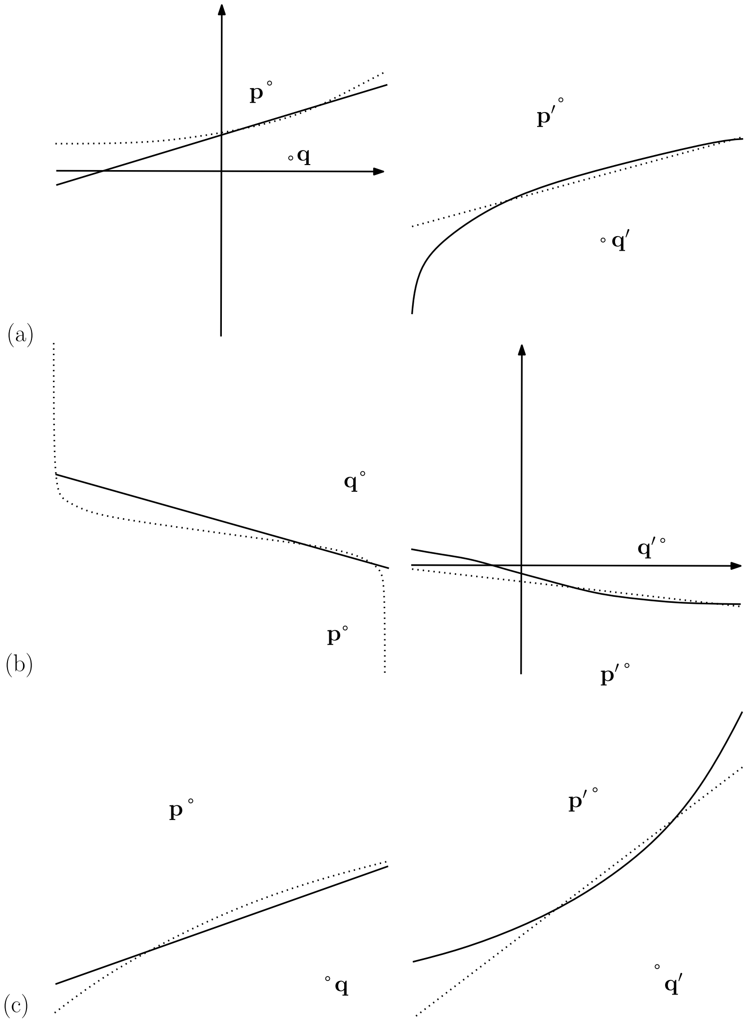


Fig. 4 Bregman bisectors. The first-type linear bisector and second-type curved bisector are displayed for pairs of primal/dual Bregman divergences: (**a**) exponential loss/Kullback–Leibler divergence, (**b**) logistic loss/dual logistic loss, and (**c**) self-dual Itakura–Saito divergence. (The scalings in $\mathcal{X}$ and $\mathcal{X}'$ do not correspond in order to improve readability.)

It should be noted that $\mathbf{p}$ and $\mathbf{q}$ lie necessarily on different sides of $BB_F(\mathbf{p},\mathbf{q})$ since $BB_F(\mathbf{p},\mathbf{q},\mathbf{p}) = D_F(\mathbf{p}\parallel\mathbf{q}) > 0$ and $BB_F(\mathbf{p},\mathbf{q},\mathbf{q}) = -D_F(\mathbf{q}\parallel\mathbf{p}) < 0$.

From Lemma 2, we know that $D_F(\mathbf{x}\parallel\mathbf{y}) = D_{F^*}(\mathbf{y}'\parallel\mathbf{x}')$ where F^* is the convex conjugate of F. We therefore have

$$BB_F(\mathbf{p},\mathbf{q}) = (\nabla F)^{-1}\big(BB'_{F^*}\big(\mathbf{q}',\mathbf{p}'\big)\big), \qquad BB'_F(\mathbf{p},\mathbf{q}) = (\nabla F)^{-1}\big(BB_{F^*}\big(\mathbf{q}',\mathbf{p}'\big)\big). \tag{3}$$

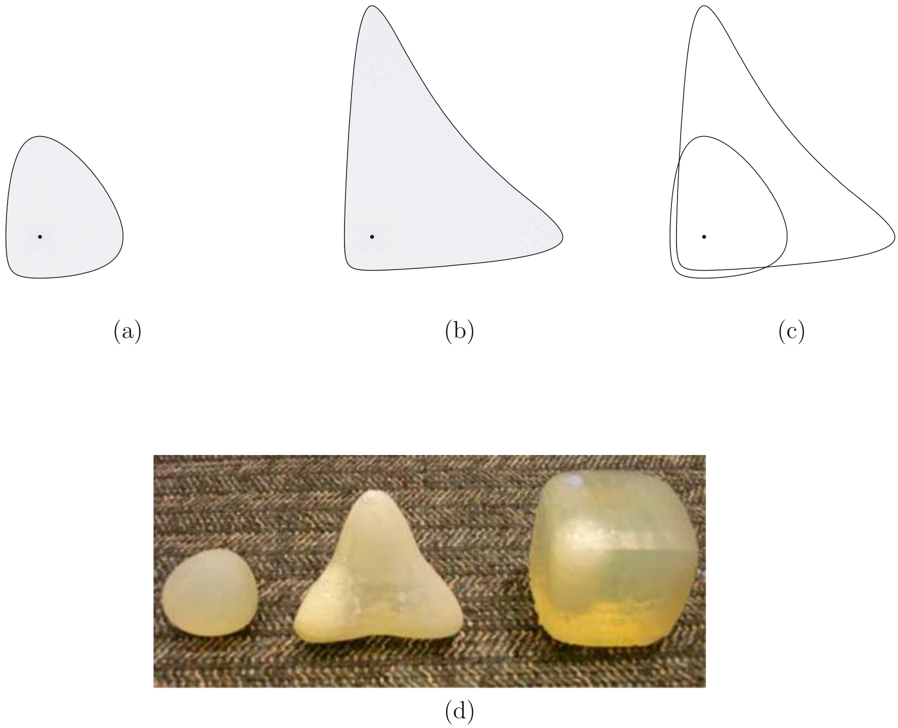


Fig. 5 Bregman balls for the Itakura–Saito divergence. The (convex) ball (**a**) of the first type $B_F(\mathbf{c}, r)$, (**b**) the ball of the second type $B'_F(\mathbf{c}, r)$ with the same center and radius, (**c**) superposition of the two corresponding bounding spheres; (**d**) shows 3D Bregman balls printed by a lithographic process (from *left* to *right*: Kullback–Leibler, Itakura–Saito and logistic balls)

Figure 4 depicts several first-type and second-type bisectors for various pairs of primal/dual Bregman divergences.

3.2 Bregman Spheres and the Lifting Map

We define the Bregman balls of, respectively, the first and the second types according to whether the center is taken as the first or the second argument of the Bregman divergence D_F:

$$B_F(\mathbf{c}, r) = \left\{\mathbf{x} \in \mathcal{X} \mid D_F\left(\mathbf{x} \parallel \mathbf{c}\right) \le r\right\} \quad \text{and} \quad B'_F(\mathbf{c}, r) = \left\{\mathbf{x} \in \mathcal{X} \mid D_F\left(\mathbf{c} \parallel \mathbf{x}\right) \le r\right\}.$$

The Bregman balls of the first type are convex, while this is not necessarily true for the balls of the second type as shown in Fig. 5 for the Itakura–Saito divergence (defined in Table 1). The associated bounding *Bregman spheres*[2] (i.e., $\partial B_F(\mathbf{c}, r)$ or $\partial B'_F(\mathbf{c}, r)$) are obtained by replacing the inequalities by equalities.

[2]For convenience, we simply say spheres instead of hyperspheres when there is no ambiguity.

From Lemma 2, we deduce that

$$B'_F(\mathbf{c}, r) = (\nabla F)^{-1}\big(B_{F^*}\big(\mathbf{c}', r\big)\big). \tag{4}$$

Let us now examine a few properties of Bregman spheres using a lifting transformation that generalizes a similar construct for Euclidean spheres (see [8, 38]).

Let us embed the domain $\mathcal{X}$ in $\hat{\mathcal{X}} = \mathcal{X} \times \mathbb{R} \subset \mathbb{R}^{d+1}$ using an *extra dimension* denoted by the Z-axis. For a point $\mathbf{x} \in \mathcal{X}$, recall that $\hat{\mathbf{x}} = (\mathbf{x}, F(\mathbf{x}))$ denotes the point obtained by lifting $\mathbf{x}$ onto the graph $\mathcal{F}$ of F (see Fig. 2). In addition, write $\mathrm{Proj}_{\mathcal{X}}(\mathbf{x}, z) = \mathbf{x}$ for the projection of a point of $\hat{\mathcal{X}}$ onto $\mathcal{X}$.

Let $\mathbf{p} \in \mathcal{X}$ and $H_{\mathbf{p}}$ be the hyperplane tangent to $\mathcal{F}$ at point $\hat{\mathbf{p}}$ of equation

$$z = H_{\mathbf{p}}(\mathbf{x}) = \big\langle \mathbf{x} - \mathbf{p}, \mathbf{p}' \big\rangle + F(\mathbf{p}), \tag{5}$$

and let $H_{\mathbf{p}}^{\uparrow}$ denote the halfspace above $H_{\mathbf{p}}$ consisting of the points $\mathbf{x} = [\mathbf{x}\ z]^T \in \hat{\mathcal{X}}$ such that $z > H_{\mathbf{p}}(\mathbf{x})$. Let $\sigma(\mathbf{c}, r)$ denote either the first-type or second-type Bregman sphere centered at $\mathbf{c}$ with radius r.

The lifted image $\hat{\sigma}$ of a Bregman sphere σ is $\hat{\sigma} = \{(\mathbf{x}, F(\mathbf{x})), \mathbf{x} \in \sigma\}$. We associate to a Bregman sphere $\sigma = \sigma(\mathbf{c}, r)$ of $\mathcal{X}$ the hyperplane

$$H_\sigma : z = \big\langle \mathbf{x} - \mathbf{c}, \mathbf{c}' \big\rangle + F(\mathbf{c}) + r, \tag{6}$$

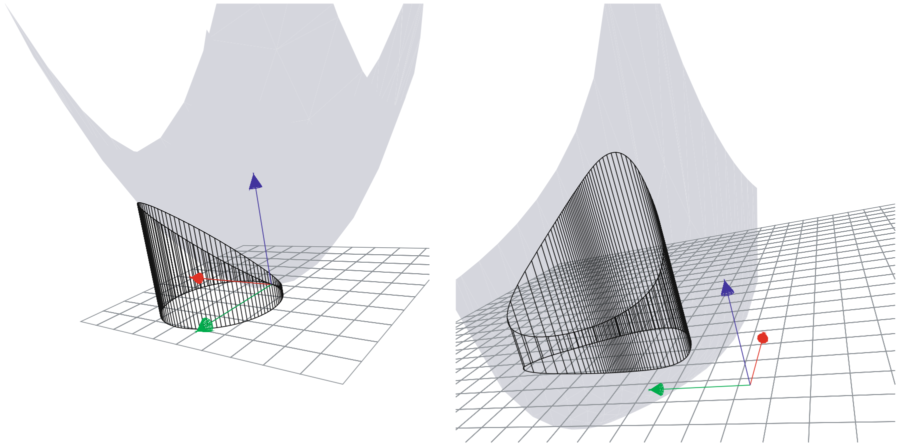
parallel to $H_{\mathbf{c}}$ and at vertical distance r from $H_{\mathbf{c}}$ (see Fig. 6). Observe that H_σ coincides with $H_{\mathbf{c}}$ when $r = 0$, i.e., when sphere σ is reduced to a single point.

Lemma 4 *$\hat{\sigma}$ is the intersection of $\mathcal{F}$ with H_σ. Conversely, the intersection of any hyperplane H with $\mathcal{F}$ projects onto $\mathcal{X}$ as a Bregman sphere. More precisely, if the equation of H is $z = \langle \mathbf{x}, \mathbf{a}\rangle + b$, the sphere of first type is centered at $\mathbf{c} = (\nabla F)^{-1}(\mathbf{a})$ and its radius is $\langle \mathbf{a}, \mathbf{c}\rangle - F(\mathbf{c}) + b$.*

Proof The first part of the lemma is a direct consequence of the fact that $D_F(\mathbf{x} \,\|\, \mathbf{y})$ is measured by the vertical distance from $\hat{\mathbf{x}}$ to $H_{\mathbf{y}}$ (see Lemma 1). For the second part, we consider the hyperplane $H^{\|}$ parallel to H and tangent to $\mathcal{F}$. From (5), we deduce $\mathbf{a} = \mathbf{c}'$. The equation of $H^{\|}$ is thus $z = \langle \mathbf{x} - (\nabla F)^{-1}(\mathbf{a}), \mathbf{a}\rangle + F((\nabla F)^{-1}(\mathbf{a}))$. It follows that the divergence from any point of σ to $\mathbf{c}$, which is equal to the vertical distance between H and $H^{\|}$, is $\langle (\nabla F)^{-1}(\mathbf{a}), \mathbf{a}\rangle - F((\nabla F)^{-1}(\mathbf{a})) + b = \langle \mathbf{a}, \mathbf{c}\rangle - F(\mathbf{c}) + b$. □

We have only considered so far Bregman spheres of codimension 1 of $\mathbb{R}^d$, i.e., hyperspheres. More generally, we can define the Bregman spheres of codimension $k+1$ of $\mathbb{R}^d$ as the Bregman (hyper)spheres of some affine space $\mathcal{Z} \subset \mathbb{R}^d$ of codimension k. The next lemma shows that Bregman spheres are stable under intersection.

Lemma 5 *The intersection of k Bregman spheres $\sigma_1, \dots, \sigma_k$ of the same type is a Bregman sphere σ of that type. If the σ_i pairwise intersect transversally, $\sigma = \bigcap_{i=1}^k \sigma_i$ is a Bregman sphere of dimension k.*



(a) Squared Euclidean distance (b) Itakura Saito divergence

Fig. 6 Two Bregman circles σ and the associated 3D curves $\hat{\sigma}$ obtained by lifting σ onto $\mathcal{F}$ (the plot of the function F is shown in *grey*). The closed curves $\hat{\sigma}$ are obtained as the intersection of the hyperplane H_σ with the convex hypersurface $\mathcal{F}$. 3D illustration with (**a**) the squared Euclidean distance, and (**b**) the Itakura–Saito divergence

Proof Consider first the case of Bregman spheres of the first type. The k hyperplanes H_{σ_i}, $i = 1, \dots, k$, intersect along an affine space H of codimension k of $\mathbb{R}^{d+1}$. Write G for the vertical projection of H onto $\mathbb{R}^d$, and $G^{\updownarrow} = G \times \mathbb{R}$ for the vertical flat of codimension $k-1$ that contains G (and H). Write further $\mathcal{F}_G = \mathcal{F} \cap G^{\updownarrow}$. Observing that $\mathcal{F}_G$ is the graph of the restriction of F to G and that H is a hyperplane of $G^{\updownarrow}$, we can apply Lemma 4 in $G^{\updownarrow}$, which proves the lemma for Bregman spheres of the first type.

The case of Bregman spheres of the second type follows from the duality of (4). □

Union and Intersection of Bregman Balls

Theorem 1 *Both the union and the intersection of n Bregman balls have combinatorial complexity* $\Theta(n^{\lceil \frac{d}{2} \rceil})$ *and can be computed in optimal time* $\Theta(n \log n + n^{\lceil \frac{d}{2} \rceil})$.

Proof Consider the case of a finite union $\mathcal{U}$ of balls and assume, without loss of generality, that the balls are in general position. To each ball, we can associate its bounding Bregman sphere σ_i which, by Lemma 4, is the projection by $\mathrm{Proj}_{\mathcal{X}}$ of the intersection of $\mathcal{F}$ with a hyperplane H_{σ_i}. The points of $\mathcal{F}$ that are below H_{σ_i} projects onto points that are inside the Bregman ball bounded by σ_i. Hence, the union of balls $\mathcal{U}$ is the projection by $\mathrm{Proj}_{\mathcal{X}}$ of the complement of $\mathcal{F} \cap \mathcal{H}^{\uparrow}$ where $\mathcal{H}^{\uparrow} = \bigcap_{i=1}^n H_{\sigma_i}^{\uparrow}$. $\mathcal{H}^{\uparrow}$ is a convex polytope defined as the intersection of n half-spaces in $\mathbb{R}^{d+1}$. The theorem follows from McMullen's theorem that bounds the number of faces of a polytope [27], and known optimal algorithms for computing convex hull/half-space intersection algorithm [15, 17]. Indeed, the number of vertices of $\mathcal{U}$ is at most twice the number of edges of $\mathcal{H}^{\uparrow}$ by convexity, and each vertex is incident to a bounded number of faces of $\mathcal{U}$ by the general position assumption. The result for the balls of the second type is deduced from the result for the balls of the first type and the duality of (4). The case of an intersection of balls is very similar (just replace $H^{\uparrow}_{\sigma_i}$ by the complementary halfspace $H^{\downarrow}_{\sigma_i}$). □

Note that output-sensitive algorithms may also be obtained following the guidelines in [14].

VC-dimension of Bregman Spheres

Theorem 2 *The VC-dimension of the class of all Bregman balls B_F of $\mathbb{R}^d$ (for any given strictly convex and differentiable function F) is $d+1$.*

Proof The result is known for Euclidean balls. Lemma 4 allows to extend the proof in [26] (Lemma 10.3.1) in a straightforward way to Bregman balls of the first type. The case of Bregman spheres of the second type follows from the duality of (4). □

Range spaces of finite VC-dimensions have found numerous applications in Combinatorial and Computational Geometry. We refer to Chazelle's book for an introduction to the subject and references wherein [16]. In particular, Brönnimann and Goodrich [12] have proposed an almost optimal solution to the disk cover algorithm, i.e., to finding a minimum number of disks in a given family that cover a given set of points. Theorem 2 allows one to extend this result to arbitrary Bregman ball cover (see also [21]).

Circumscribing Bregman Spheres There exists, in general, a unique Bregman sphere passing through $d+1$ points of $\mathbb{R}^d$. This is easily shown using the lifting map since, in general, there exists a unique hyperplane of $\mathbb{R}^{d+1}$ passing through $d+1$ points. The claim then follows from Lemma 4.

Deciding whether a point $\mathbf{x}$ falls *inside*, *on*, or *outside* a Bregman sphere $\sigma \in \mathbb{R}^d$ passing through $d+1$ points of $\mathbf{p}_0, \ldots, \mathbf{p}_d$ will be crucial for computing Bregman Voronoi diagrams and associated triangulations. The lifting map immediately implies that such a decision task reduces to determining the orientation of the simplex $(\hat{\mathbf{p}}_0, \ldots, \hat{\mathbf{p}}_d, \hat{\mathbf{x}})$ of $\mathbb{R}^{d+1}$, which in turn reduces to evaluating the sign of the determinant of the $(d+2)\times(d+2)$ matrix (see [28])

$$\texttt{InSphere}(\mathbf{x}; \mathbf{p}_0, \ldots, \mathbf{p}_d) = \begin{vmatrix} 1 & \cdots & 1 & 1 \\ \mathbf{p}_0 & \cdots & \mathbf{p}_d & \mathbf{x} \\ F(\mathbf{p}_0) & \cdots & F(\mathbf{p}_d) & F(\mathbf{x}) \end{vmatrix}$$

If one assumes that the determinant $\begin{vmatrix} 1 & \cdots & 1 \\ \mathbf{p}_0 & \cdots & \mathbf{p}_d \end{vmatrix}$ is positive, $\texttt{InSphere}(\mathbf{x}; \mathbf{p}_0, \ldots, \mathbf{p}_d)$ is negative, null or positive depending on whether $\mathbf{x}$ lies inside, on, or outside σ, respectively.

3.3 Projection and Orthogonality

We start with an easy property of Bregman divergences.

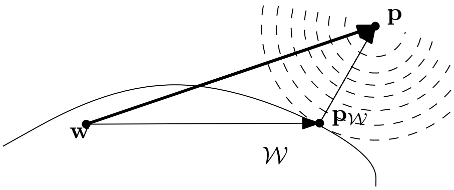


Fig. 7 The projection $\mathbf{p}_{\mathcal{W}}$ of point $\mathbf{p}$ to a convex subset $\mathcal{W} \subseteq \mathcal{X}$ and Bregman Pythagorean inequality

Property 5 (Three-point property) *For any triple* $\mathbf{p}$, $\mathbf{q}$ *and* $\mathbf{r}$ *of points of* $\mathcal{X}$, *we have* $D_F(\mathbf{p} \parallel \mathbf{q}) + D_F(\mathbf{q} \parallel \mathbf{r}) = D_F(\mathbf{p} \parallel \mathbf{r}) + \langle \mathbf{p} - \mathbf{q}, \mathbf{r}' - \mathbf{q}' \rangle$.

The following lemma characterizes the *Bregman projection* of a point onto a closed convex set $\mathcal{W} \subseteq \mathcal{X}$.

Lemma 6 (Bregman projection) *For any* $\mathbf{p}$ *in* $\mathcal{X}$, *there exists a unique point* $\mathbf{x} \in \mathcal{W}$ *that minimizes* $D_F(\mathbf{x} \parallel \mathbf{p})$. *We call this point the Bregman projection of* $\mathbf{p}$ *onto* $\mathcal{W}$ *and denote it* $\mathbf{p}_{\mathcal{W}}$ (*i.e.*, $\mathbf{p}_{\mathcal{W}} = \arg\min_{\mathbf{x} \in \mathcal{W}} D_F(\mathbf{x} \parallel \mathbf{p})$).

Proof Assume for a contradiction that there exists two points $\mathbf{x}$ and $\mathbf{y}$ of $\mathcal{W}$ that minimize the divergence to $\mathbf{p}$, and let $D_F(\mathbf{x} \parallel \mathbf{p}) = D_F(\mathbf{y} \parallel \mathbf{p}) = l$. Since $\mathcal{W}$ is convex, $(\mathbf{x}+\mathbf{y})/2 \in \mathcal{W}$ and, since D_F is strictly convex in its first argument (see Property 2 of Sect. 2.1), $D_F((\mathbf{x}+\mathbf{y})/2 \parallel \mathbf{p}) < D_F(\mathbf{x} \parallel \mathbf{p})/2 + D_F(\mathbf{y} \parallel \mathbf{p})/2 = l$, yielding a contradiction. □

We recall the following property already mentioned in [6] (see Fig. 7).

Property 6 (Bregman Pythagorean inequality) *Let* $\mathbf{p}_{\mathcal{W}}$ *denote the Bregman projection of point* $\mathbf{p}$ *to a convex subset* $\mathcal{W} \subseteq \mathcal{X}$. *For any* $\mathbf{w} \in \mathcal{W}$, *we have* $D_F(\mathbf{w} \parallel \mathbf{p}) \geq D_F(\mathbf{w} \parallel \mathbf{p}_{\mathcal{W}}) + D_F(\mathbf{p}_{\mathcal{W}} \parallel \mathbf{p})$, *with equality for and only for affine sets* $\mathcal{W}$.

Proof By the Three-point property, we have

$$D_F\big(\mathbf{w} \parallel \mathbf{p}_{\mathcal{W}}\big) + D_F\big(\mathbf{p}_{\mathcal{W}} \parallel \mathbf{p}\big) = D_F\big(\mathbf{w} \parallel \mathbf{p}\big) + \big\langle \mathbf{w} - \mathbf{p}_{\mathcal{W}}, \mathbf{p}' - \mathbf{p}'_{\mathcal{W}} \big\rangle.$$

From $\mathbf{p}_{\mathcal{W}} = \arg\min_{\mathbf{x} \in \mathcal{W}} D_F(\mathbf{x} \parallel \mathbf{p})$, we deduce that the inner product in the equality above is non positive, and zero if $\mathcal{W}$ is an affine set. □

We now introduce the notion of *Bregman orthogonality*. We say that the (ordered) triplet $(\mathbf{p}, \mathbf{q}, \mathbf{r})$ is *Bregman orthogonal* iff $D_F(\mathbf{p} \parallel \mathbf{q}) + D_F(\mathbf{q} \parallel \mathbf{r}) = D_F(\mathbf{p} \parallel \mathbf{r})$, or equivalently (by the three-point property), iff $\langle \mathbf{p} - \mathbf{q}, \mathbf{r}' - \mathbf{q}' \rangle = 0$. Observe the analogy with Pythagorean theorem in Euclidean space. It should be noted though that Bregman orthogonality depends on the order of the three points.

Notice that orthogonality is preserved (with reverse order) in the gradient space. Indeed, since $\langle \mathbf{p} - \mathbf{q}, \mathbf{r}' - \mathbf{q}' \rangle = \langle \mathbf{r}' - \mathbf{q}', \mathbf{p} - \mathbf{q} \rangle$, $(\mathbf{p}, \mathbf{q}, \mathbf{r})$ is Bregman orthogonal iff $(\mathbf{r}', \mathbf{q}', \mathbf{p}')$ is Bregman orthogonal.

More generally, we say that $I \subseteq \mathcal{X}$ is *Bregman orthogonal* to $J \subseteq \mathcal{X}$ ($I \cap J \neq \emptyset$) iff for any $\mathbf{p} \in I$ and $\mathbf{r} \in J$, there exists a $\mathbf{q} \in I \cap J$ such that $(\mathbf{p}, \mathbf{q}, \mathbf{r})$ is Bregman orthogonal.

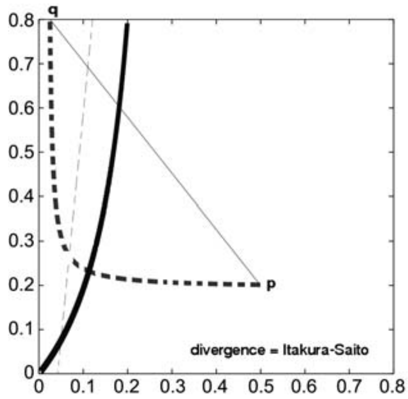


Fig. 8 Bregman bisectors $BB_F(\mathbf{p},\mathbf{q})$ (*thin dashed line segments*) and $BB_{F^*}(\mathbf{p},\mathbf{q})$ (*bold solid arcs*), and their relationships with respect to $\Lambda(\mathbf{p},\mathbf{q})$ (*thin solid line segments*) and $\Gamma_F(\mathbf{p},\mathbf{q})$ (*bold dashed arcs*), for the Itakura–Saito divergence (*left*) and Kullback–Leibler-divergence (*right*)

Let $\Gamma_F(\mathbf{p},\mathbf{q})$ be the image by $(\nabla F)^{-1}$ of the line segment $\mathbf{p}'\mathbf{q}'$, i.e.,

$$\Gamma_F(\mathbf{p},\mathbf{q}) = \{\mathbf{x}\in\mathcal{X} : \mathbf{x}' = (1-\lambda)\mathbf{p}' + \lambda\mathbf{q}', \lambda\in[0,1]\}.$$

We call $\Gamma_F(\mathbf{p},\mathbf{q})$ the *geodesic arc* joining $\mathbf{p}$ to $\mathbf{q}$. By analogy, we rename the line segment $\mathbf{pq}$ as

$$\Lambda(\mathbf{p},\mathbf{q}) = \{\mathbf{x}\in\mathcal{X} : \mathbf{x} = (1-\lambda)\mathbf{p} + \lambda\mathbf{q}, \lambda\in[0,1]\}.$$

In the Euclidean case ($F(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$), $\Gamma_F(\mathbf{p},\mathbf{q}) = \Lambda(\mathbf{p},\mathbf{q})$ is orthogonal to the bisector $BB_F(\mathbf{p},\mathbf{q})$. For general Bregman divergences, we have similar properties as shown next.

Lemma 7 *The Bregman bisector $BB_F(\mathbf{p},\mathbf{q})$ is Bregman orthogonal to $\Gamma_F(\mathbf{p},\mathbf{q})$ while $\Lambda(\mathbf{p},\mathbf{q})$ is Bregman orthogonal to $BB_{F^*}(\mathbf{p},\mathbf{q})$.*

Proof Since $\mathbf{p}$ and $\mathbf{q}$ lie on different sides of $BB_F(\mathbf{p},\mathbf{q})$, $\Gamma_F(\mathbf{p},\mathbf{q})$ must intersect $BB_F(\mathbf{p},\mathbf{q})$. Fix any distinct $\mathbf{x}\in\Gamma_F(\mathbf{p},\mathbf{q})$ and $\mathbf{y}\in BB_F(\mathbf{p},\mathbf{q})$, and let $\mathbf{t}\in\Gamma_F(\mathbf{p},\mathbf{q})\cap BB_F(\mathbf{p},\mathbf{q})$. To prove the first part of the lemma, we need to show that $\langle\mathbf{y}-\mathbf{t},\mathbf{x}'-\mathbf{t}'\rangle = 0$.

Since $\mathbf{t}$ and $\mathbf{x}$ both belong to $\Gamma_F(\mathbf{p},\mathbf{q})$, we have $\mathbf{t}'-\mathbf{x}' = \lambda(\mathbf{p}'-\mathbf{q}')$, for some $\lambda\in\mathbb{R}$, and, since $\mathbf{y}$ and $\mathbf{t}$ belong to $BB_F(\mathbf{p},\mathbf{q})$, we deduce from the equation of $BB_F(\mathbf{p},\mathbf{q})$ that $\langle\mathbf{y}-\mathbf{t},\mathbf{p}'-\mathbf{q}'\rangle = 0$. We conclude that $\langle\mathbf{y}-\mathbf{t},\mathbf{x}'-\mathbf{t}'\rangle = 0$, which proves that $BB_F(\mathbf{p},\mathbf{q})$ is indeed Bregman orthogonal to $\Gamma_F(\mathbf{p},\mathbf{q})$.

The second part of the lemma is easily proved by using the fact that orthogonality is preserved in the gradient space as noted above. □

Figure 8 shows Bregman bisectors and their relationships with respect to $\Lambda(\mathbf{p},\mathbf{q})$ and $\Gamma_F(\mathbf{p},\mathbf{q})$.

4 Bregman Voronoi Diagrams

Let $\mathcal{S} = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ be a finite point set in $\mathcal{X} \subset \mathbb{R}^d$. To each point $\mathbf{p}_i$ a d-variate continuous function D_i defined over $\mathcal{X}$ is attached. We define the *lower envelope* of the functions as the graph of $\min_{1\le i\le n} D_i$ and their *minimization diagram* as the subdivision of $\mathcal{X}$ into cells such that, in each cell, $\arg\min_i D_i$ is fixed.

The Euclidean Voronoi diagram is the minimization diagram for $D_i(\mathbf{x}) = \|\mathbf{x} - \mathbf{p}_i\|^2$. In this section, we introduce Bregman Voronoi diagrams as minimization diagrams of Bregman divergences (see Fig. 10).

We define two types of Bregman Voronoi diagrams in Sect. 4.1. We establish a correspondence between Bregman Voronoi diagrams, polytopes and power diagrams in Sect. 4.2. This correspondence leads to tight combinatorial bounds and efficient algorithms. Finally, in Sect. 4.3, we generalize Bregman Voronoi diagrams to k-order and k-bag diagrams.

Let $\mathcal{S}' = \{\nabla F(\mathbf{p}_i), i = 1, \ldots, n\}$ denote the *gradient point set* associated to $\mathcal{S}$.

4.1 Two Types of Diagrams

Because Bregman divergences are not necessarily symmetric, we associate to each site $\mathbf{p}_i$ *two types* of distance functions, namely $D_i(\mathbf{x}) = D_F(\mathbf{x} \,\|\, \mathbf{p}_i)$ and $D_i'(\mathbf{x}) = D_F(\mathbf{p}_i \,\|\, \mathbf{x})$. The *minimization diagram* of the D_i, $i = 1, \ldots, n$, is called the *first-type* Bregman Voronoi diagram of $\mathcal{S}$, which we denote by $\mathrm{vor}_F(\mathcal{S})$. The d-dimensional cells of this diagram are in *1–1 correspondence* with the sites $\mathbf{p}_i$ and the d-dimensional cell of $\mathbf{p}_i$ is defined as

$$\mathrm{vor}_F(\mathbf{p}_i) \overset{\text{def}}{=} \bigl\{\mathbf{x} \in \mathcal{X} \mid D_F\bigl(\mathbf{x} \,\|\, \mathbf{p}_i\bigr) \le D_F\bigl(\mathbf{x} \,\|\, \mathbf{p}_j\bigr)\ \forall \mathbf{p}_j \in \mathcal{S}\bigr\}.$$

Since the Bregman bisectors of the first-type are hyperplanes, the cells of any diagram of the first-type are convex polyhedra. Therefore, first-type Bregman Voronoi diagrams are *affine* diagrams [4, 5].

Similarly, the minimization diagram of the D_i', $i = 1, \ldots, n$, is called the *second-type* Bregman Voronoi diagram of $\mathcal{S}$, which we denote by $\mathrm{vor}'_F(\mathcal{S})$. A cell in $\mathrm{vor}'_F(\mathcal{S})$ is associated to each site $\mathbf{p}_i$ and is defined as above with permuted divergence arguments:

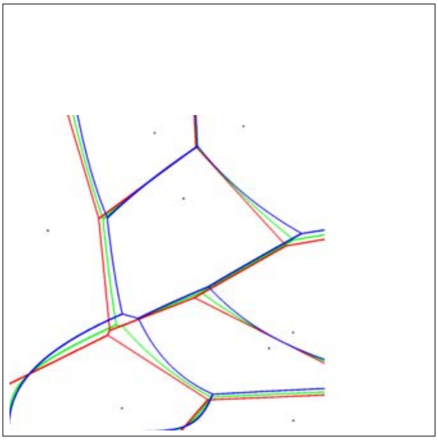
$$\mathrm{vor}'_F(\mathbf{p}_i) \overset{\text{def}}{=} \bigl\{\mathbf{x} \in \mathcal{X} \mid D_F\bigl(\mathbf{p}_i \,\|\, \mathbf{x}\bigr) \le D_F\bigl(\mathbf{p}_j \,\|\, \mathbf{x}\bigr)\ \forall \mathbf{p}_j \in \mathcal{S}\bigr\}.$$

In contrast with the diagrams of the first-type, the diagrams of the second type have, in general, curved faces.

Figure 9 illustrates these Bregman Voronoi diagrams for the Kullback–Leibler and the Itakura–Saito divergences. Note that the ordinary Euclidean Voronoi diagram is a Bregman Voronoi diagram since $\mathrm{vor}(\mathcal{S}) = \mathrm{vor}_F(\mathcal{S}) = \mathrm{vor}'_F(\mathcal{S})$ for $F(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$.

From the Legendre duality between divergences, we deduce correspondences between the diagrams of the first and the second types. As usual, F^* is the convex conjugate of F.

Lemma 8 $\mathrm{vor}'_F(\mathcal{S}) = (\nabla F)^{-1}(\mathrm{vor}_{F^*}(\mathcal{S}'))$ *and* $\mathrm{vor}_F(\mathcal{S}) = (\nabla F)^{-1}(\mathrm{vor}'_{F^*}(\mathcal{S}'))$.


(a)

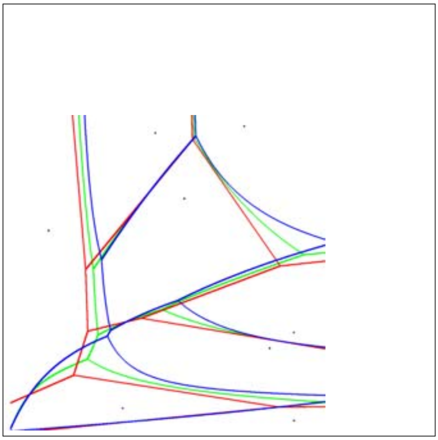

(b)

Fig. 9 Three types of Bregman Voronoi diagrams for (**a**) the Kullback–Leibler and (**b**) the Itakura–Saito divergences: the affine first-type Bregman Voronoi diagram, the associated curved second-type Bregman Voronoi diagram and, in between, the symmetrized Bregman Voronoi diagram associated to the distance functions $D''_i(\mathbf{x}) = \frac{1}{2}(D_i(\mathbf{x}) + D'_i(\mathbf{x}))$

Proof By Lemma 2, we have $D_F(\mathbf{x} \parallel \mathbf{y}) = D_{F^*}(\mathbf{y}' \parallel \mathbf{x}')$, which gives $\mathrm{vor}_F(\mathbf{p}_i) = \{\mathbf{x} \in \mathcal{X} \mid D_{F^*}(\mathbf{p}'_i \parallel \mathbf{x}') \leq D_{F^*}(\mathbf{p}'_j \parallel \mathbf{x}')\ \forall \mathbf{p}'_j \in \mathcal{S}'\} = (\nabla F)^{-1}(\mathrm{vor}'_{F^*}(\mathbf{p}'_i))$. This proves the second part of the lemma. The proof of the first part follows the same path. □

Hence, constructing the second-type curved diagram $\mathrm{vor}'_F(\mathcal{S})$ reduces to constructing an affine diagram in the gradient space $\mathcal{X}'$ (and mapping the cells by ∇F^{-1}).

4.2 Bregman Voronoi Diagrams, Polytopes and Power Diagrams

Let $H_{\mathbf{p}_i}$, $i = 1, \ldots, n$, denote the hyperplanes of $\hat{\mathcal{X}}$ defined in Sect. 3.2. For any $\mathbf{x} \in \mathcal{X}$, we deduce from Lemma 1

$$D_F(\mathbf{x} \parallel \mathbf{p}_i) \leq D_F(\mathbf{x} \parallel \mathbf{p}_j) \quad \Longleftrightarrow H_{\mathbf{p}_i}(\mathbf{x}) \geq H_{\mathbf{p}_j}(\mathbf{x}).$$

The first-type Bregman Voronoi diagram of $\mathcal{S}$ is therefore the maximization diagram of the n affine functions $H_{\mathbf{p}_i}(\mathbf{x})$ whose graphs are the hyperplanes $H_{\mathbf{p}_i}$ (see Fig. 10). Equivalently, the first-type Bregman Voronoi diagram $\mathrm{vor}_F(\mathcal{S})$ is obtained by projecting with $\mathrm{Proj}_{\mathcal{X}}$ the faces of the $(d+1)$-dimensional convex polyhedron $\mathcal{H} = \cap_i H^{\uparrow}_{\mathbf{p}_i}$ of $\hat{\mathcal{X}}$ onto $\mathcal{X}$.

Since the intersection of n half-spaces of $\mathbb{R}^d$ has complexity $\Theta(n^{\lfloor \frac{d}{2} \rfloor})$ and can be computed in optimal-time $\Theta(n \log n + n^{\lfloor \frac{d}{2} \rfloor})$ for any fixed dimension d [15, 27] and thanks to Lemma 8, we then deduce the following theorem.

Theorem 3 *The Bregman Voronoi diagrams of the first and the second types of a set of n d-dimensional points have complexity $\Theta(n^{\lceil \frac{d}{2} \rceil})$ and can be computed in optimal time $\Theta(n \log n + n^{\lceil \frac{d}{2} \rceil})$.*

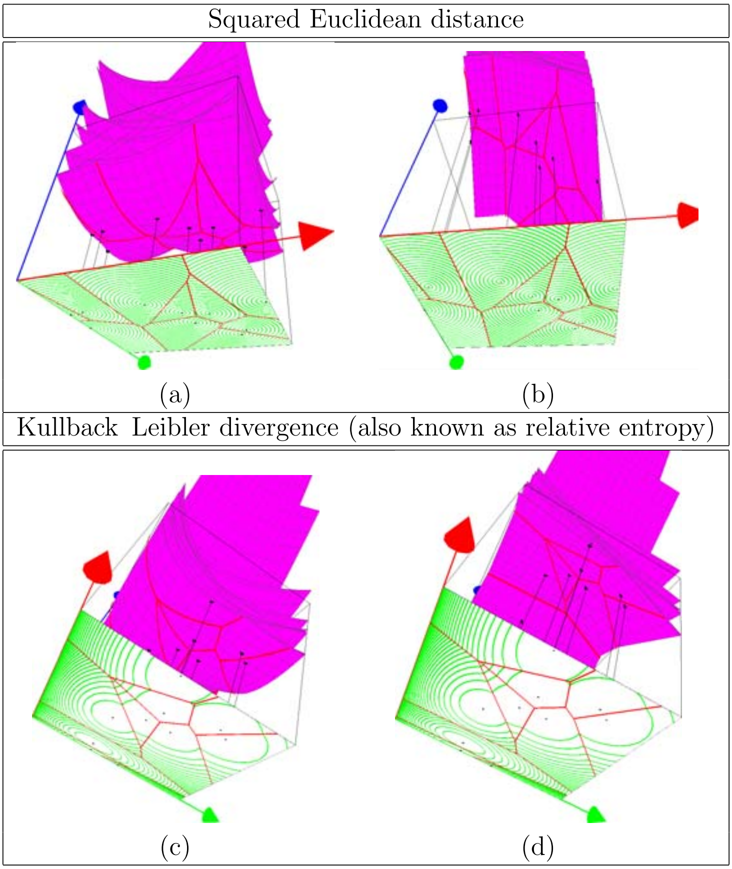


Fig. 10 Voronoi diagrams as minimization diagrams. The *first row* shows minimization diagrams for the Euclidean distance and the *second row* shows minimization diagrams for the Kullback–Leibler divergence. In the *first column*, the functions are the non-linear functions $D_i(\mathbf{x})$ and, in the *second column*, the functions are the linear functions $H_{\mathbf{p}_i}(\mathbf{x})$, both leading to the same minimization diagrams. Isolines are superimposed to the Voronoi diagrams

Since Bregman Voronoi diagrams of the first type are affine diagrams, Bregman Voronoi diagrams are power diagrams [3, 8] in disguise. The following theorem makes precise the correspondence between Bregman Voronoi diagrams and power diagrams (see Fig. 11).

Theorem 4 *The first-type Bregman Voronoi diagram of n sites is identical to the power diagram of the n Euclidean spheres of equations*

$$\langle \mathbf{x} - \mathbf{p}'_i, \mathbf{x} - \mathbf{p}'_i \rangle = \langle \mathbf{p}'_i, \mathbf{p}'_i \rangle + 2\big(F(\mathbf{p}_i) - \langle \mathbf{p}_i, \mathbf{p}'_i \rangle\big), \quad i = 1, \ldots, n.$$

Proof We easily have

$$\begin{aligned} &D_F(\mathbf{x} \,\|\, \mathbf{p}_i) \le D_F(\mathbf{x} \,\|\, \mathbf{p}_j) \\ &\quad\iff \quad -F(\mathbf{p}_i) - \langle \mathbf{x} - \mathbf{p}_i, \mathbf{p}'_i \rangle \le -F(\mathbf{p}_j) - \langle \mathbf{x} - \mathbf{p}_j, \mathbf{p}'_j \rangle \end{aligned}$$

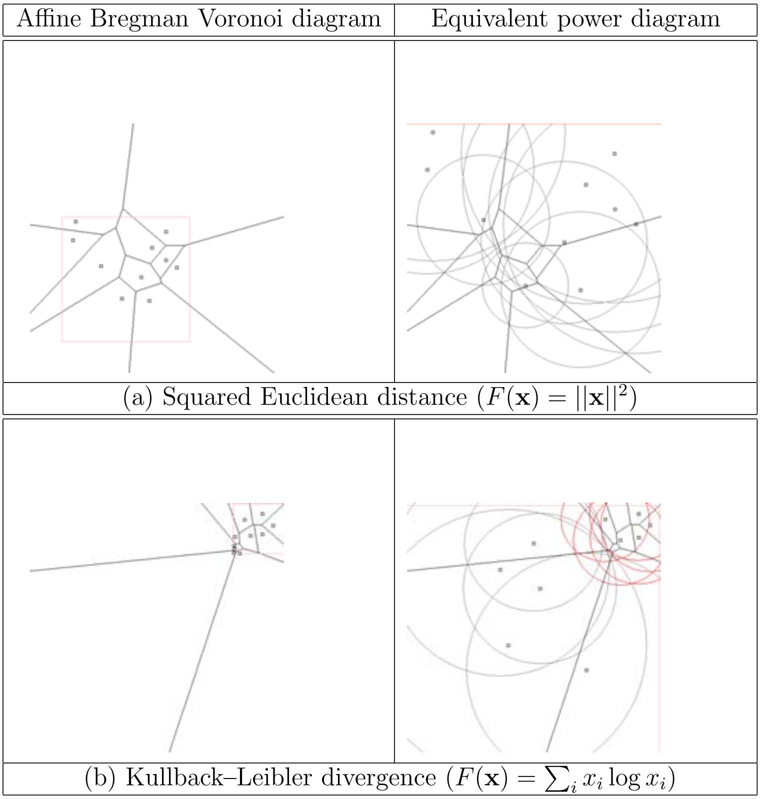


Fig. 11 Affine Bregman Voronoi diagrams (*left column*) can be computed as power diagrams (*right column*). Illustrations for the squared Euclidean distance (**a**), Kullback–Leibler divergence (**b**)

$$\begin{aligned}
\iff \quad & \langle \mathbf{x}, \mathbf{x}\rangle - 2\langle \mathbf{x}, \mathbf{p}_i'\rangle - 2F(\mathbf{p}_i) + 2\langle \mathbf{p}_i, \mathbf{p}_i'\rangle \\
& \quad \le \langle \mathbf{x}, \mathbf{x}\rangle - 2\langle \mathbf{x}, \mathbf{p}_j'\rangle - 2F(\mathbf{p}_j) + 2\langle \mathbf{p}_j, \mathbf{p}_j'\rangle \\
\iff \quad & \langle \mathbf{x} - \mathbf{p}_i', \mathbf{x} - \mathbf{p}_i'\rangle - r_i^2 \le \langle \mathbf{x} - \mathbf{p}_j', \mathbf{x} - \mathbf{p}_j'\rangle - r_j^2,
\end{aligned}$$

where $r_i^2 = \langle \mathbf{p}_i', \mathbf{p}_i'\rangle + 2(F(\mathbf{p}_i) - \langle \mathbf{p}_i, \mathbf{p}_i'\rangle)$ and $r_j^2 = \langle \mathbf{p}_j', \mathbf{p}_j'\rangle + 2(F(\mathbf{p}_j) - \langle \mathbf{p}_j, \mathbf{p}_j'\rangle)$. The last inequality means that the power of $\mathbf{x}$ with respect to the Euclidean (possibly imaginary) sphere $B(\mathbf{p}_i', r_i)$ is no more than the power of $\mathbf{x}$ with respect to the Euclidean (possibly imaginary) sphere $B(\mathbf{p}_j', r_j)$. □

For $F(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$, $\mathrm{vor}_F(\mathcal{S})$ is the Euclidean Voronoi diagram of $\mathcal{S}$. Accordingly, the theorem says that the centers of the spheres are the $\mathbf{p}_i$ and $r_i^2 = 0$ since $\mathbf{p}_i' = \mathbf{p}_i$.

Figure 11 displays affine Bregman Voronoi diagrams[3] and their equivalent power diagrams for the squared Euclidean, Kullback–Leibler and exponential divergences.

Since power diagrams are well defined over $\mathbb{R}^d$, this equivalence relationship provides a natural way to extend the scope of definition of Bregman Voronoi diagrams from $\mathcal{X} \subset \mathbb{R}^d$ to the full space $\mathbb{R}^d$. (The same observation holds for hyperbolic Voronoi diagrams [31] that are affine diagrams in disguise).

It is also to be observed that not all power diagrams are Bregman Voronoi diagrams. Indeed, in power diagrams, some spheres may have empty cells while each site has necessarily a nonempty cell in a Bregman Voronoi diagram (see Fig. 11 and Sect. 4.3 for a further discussion on this point).

4.3 Generalized Bregman Divergences and Their Voronoi Diagrams

Weighted Bregman Voronoi Diagrams

Let us associate to each site $\mathbf{p}_i$ a weight $w_i \in \mathbb{R}$. We define the *weighted divergence* between two weighted points as $WD_F(\mathbf{p}_i \| \mathbf{p}_j) \stackrel{\text{def}}{=} D_F(\mathbf{p}_i \| \mathbf{p}_j) - w_i + w_j$. We can define bisectors and weighted Bregman Voronoi diagrams in very much the same way as for non weighted divergences. The Bregman Voronoi region associated to the weighted point $(\mathbf{p}_i, w_i)$ is defined as

$$\mathrm{vor}_F(\mathbf{p}_i, w_i) = \left\{\mathbf{x} \in \mathcal{X} \mid D_F\left(\mathbf{x} \,\|\, \mathbf{p}_i\right) - w_i \le D_F\left(\mathbf{x} \,\|\, \mathbf{p}_j\right) - w_j \;\forall \mathbf{p}_j \in \mathcal{S}\right\}.$$

Observe that the bisectors of the first-type diagrams are still hyperplanes and that the diagram can be obtained as the projection of a convex polyhedron or as the power diagram of a finite set of spheres. The only difference with respect to the construction of Sect. 4.2 is the fact that now the hyperplanes $H_{\mathbf{p}_i}$ are no longer tangent to $\mathcal{F}$ since they are *shifted* by a z-displacement of length w_i. Hence Theorem 3 extends to weighted Bregman Voronoi diagrams.

k-order Bregman Voronoi Diagrams

We define the k-order Bregman Voronoi diagram of a finite point set $\mathcal{S}$ in $\mathcal{X}$ as follows. Let $\mathcal{T}$ be a subset of k sites of $\subset \mathcal{S}$. The cell of $\mathcal{T}$ in the k-order Bregman Voronoi diagram of $\mathcal{S}$ is defined as

$$\mathrm{vor}_F(\mathcal{T}) \stackrel{\text{def}}{=} \left\{\mathbf{x} \in \mathcal{X} \mid D_F\left(\mathbf{x} \,\|\, \mathbf{p}_i\right) \le D_F\left(\mathbf{x} \,\|\, \mathbf{p}_j\right) \;\forall \mathbf{p}_i \in \mathcal{T} \text{ and } \mathbf{p}_j \in \mathcal{S} \setminus \mathcal{T}\right\}.$$

The k-order Bregman Voronoi diagram of $\mathcal{S}$ of the first-type is then defined as the cell complex whose d-cells are the cells of all the subsets of k points of $\mathcal{S}$.

We can define in a similar way the k-order Bregman Voronoi diagram of $\mathcal{S}$ of the second-type.

Similarly to the case of higher-order Euclidean Voronoi diagrams, we have

[3] See Java™ applet at http://www.csl.sony.co.jp/person/nielsen/BVDapplet/.

Theorem 5 *The k-order Bregman Voronoi diagram of n d-dimensional points is a weighted Bregman Voronoi diagram.*

Proof Let $\mathcal{S}_1, \mathcal{S}_2, \ldots$ denote the subsets of k points of $\mathcal{S}$ and write

$$\begin{aligned} D_i(\mathbf{x}) &= \frac{1}{k}\sum_{\mathbf{p}_j\in\mathcal{S}_i} D_F\big(\mathbf{x}\,\|\,\mathbf{p}_j\big) \\ &= F(\mathbf{x}) - \frac{1}{k}\sum_{\mathbf{p}_j\in\mathcal{S}_i} F(\mathbf{p}_j) - \frac{1}{k}\sum_{\mathbf{p}_j\in\mathcal{S}_i}\big\langle \mathbf{x}-\mathbf{p}_j, \mathbf{p}'_j\big\rangle \\ &= F(\mathbf{x}) - F(\mathbf{c}_i) - \big\langle \mathbf{x}-\mathbf{c}_i, \mathbf{c}'_i\big\rangle - w_i \\ &= WD_F\big(\mathbf{x}\,\|\,\mathbf{c}_i\big), \end{aligned}$$

where $\mathbf{c}_i = (\nabla F)^{-1}(\frac{1}{k}\sum_{p_j\in S_i}\mathbf{p}'_j)$ and the weight associated to $\mathbf{c}_i$ is $w_i = -F(\mathbf{c}_i) + \langle \mathbf{c}_i, \mathbf{c}'_i\rangle + \frac{1}{k}\sum_{p_j\in S_i}(F(\mathbf{p}_j) - \langle \mathbf{p}_j, \mathbf{p}'_j\rangle) = -\frac{1}{k}\sum_{p_j\in S_i} F^*(\mathbf{p}'_j) + F^*(\mathbf{c}'_i)$.

Hence, $\mathcal{S}_i$ is the set of the k nearest neighbors of $\mathbf{x}$ iff $D_i(\mathbf{x}) \le D_j(\mathbf{x})$ for all j, or equivalently, iff $\mathbf{x}$ belongs to the cell of $\mathbf{c}_i$ in the weighted Bregman Voronoi diagram of the $\mathbf{c}_i$.

Constructing the k-order Bregman Voronoi diagram of $\mathcal{S}$ therefore reduces to constructing the power diagram of the weighted sites $(\mathbf{c}_i, w_i)$. □

k-bag Bregman Voronoi Diagrams

Let $F_1, \ldots, F_k$ be k strictly convex and differentiable functions, and $\boldsymbol{\alpha} = [\alpha_1 \ldots \alpha_k]^T \in \mathbb{R}^k_+$ a vector of positive weights. Consider the d-variate function $F_{\boldsymbol{\alpha}} = \sum_{l=1}^k \alpha_l F_l$. By virtue of the positive additivity property rule of Bregman generator functions (Property 3), $D_{F_{\boldsymbol{\alpha}}}$ is a Bregman divergence.

Now consider a set $\mathcal{S} = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ of n points of $\mathbb{R}^d$. To each site $\mathbf{p}_i$, we associate a weight vector $\boldsymbol{\alpha}_i = [\alpha_i^{(1)} \ldots \alpha_i^{(k)}]^T$ inducing a Bregman divergence $D_{F_{\boldsymbol{\alpha}_i}}(\mathbf{x}\,\|\,\mathbf{p}_i)$ anchored at that site. Let us consider the first-type of k-bag Bregman Voronoi diagram (k-bag BVD for short). The first-type bisector $K_F(\mathbf{p}_i, \mathbf{p}_j)$ of two weighted points $(\mathbf{p}_i, \boldsymbol{\alpha}_i)$ and $(\mathbf{p}_j, \boldsymbol{\alpha}_j)$ is the locus of points $\mathbf{x}$ at equidivergence to $\mathbf{p}_i$ and $\mathbf{p}_j$. That is, $K_F(\mathbf{p}_i, \mathbf{p}_j) = \{\mathbf{x} \in \mathcal{X} \mid D_{F_{\boldsymbol{\alpha}_i}}(\mathbf{x}\,\|\,\mathbf{p}_i) = D_{F_{\boldsymbol{\alpha}_j}}(\mathbf{x}\,\|\,\mathbf{p}_j)\}$. The equation of the bisector is simply obtained using the definition of Bregman divergences (1) as

$$F_{\boldsymbol{\alpha}_i}(\mathbf{x}) - F_{\boldsymbol{\alpha}_i}(\mathbf{p}_i) - \big\langle \mathbf{x}-\mathbf{p}_i, \nabla F_{\boldsymbol{\alpha}_i}(\mathbf{p}_i)\big\rangle = F_{\boldsymbol{\alpha}_j}(\mathbf{x}) - F_{\boldsymbol{\alpha}_j}(\mathbf{p}_j) - \big\langle \mathbf{x}-\mathbf{p}_j, \nabla F_{\boldsymbol{\alpha}_i}(\mathbf{p}_j)\big\rangle.$$

This yields the equation of the first-type bisector $K_F(\mathbf{p}_i, \mathbf{p}_j)$

$$\sum_{l=1}^{k}\big(\alpha_i^{(l)} - \alpha_j^{(l)}\big) F_l(\mathbf{x}) + \big\langle \mathbf{x}, \nabla F_{\boldsymbol{\alpha}_j}(\mathbf{p}_j) - \nabla F_{\boldsymbol{\alpha}_i}(\mathbf{p}_i)\big\rangle + c = 0, \tag{7}$$

where c is a *constant* depending on the weighted sites $(\mathbf{p}_i, \boldsymbol{\alpha}_i)$ and $(\mathbf{p}_j, \boldsymbol{\alpha}_j)$. Note that the equation of the first-type k-bag BVD bisector is linear if and only if $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_j$ (i.e., the case of standard BVDs).

Let us consider the linearization lifting $\mathbf{x} \mapsto \hat{\mathbf{x}} = [\mathbf{x}\ F_1(\mathbf{x})\ \ldots\ F_k(\mathbf{x})]^T$ that maps a point $\mathbf{x} \in \mathbb{R}^d$ to a point $\hat{\mathbf{x}}$ in $\mathbb{R}^{d+k}$. Then (7) becomes linear, namely $\langle \hat{\mathbf{x}}, \mathbf{a} \rangle + c = 0$ with

$$\mathbf{a} = \begin{bmatrix} \nabla F_{\boldsymbol{\alpha}_j}(\mathbf{p}_j) - \nabla F_{\boldsymbol{\alpha}_i}(\mathbf{p}_i) \\ \boldsymbol{\alpha}_i - \boldsymbol{\alpha}_j \end{bmatrix} \in \mathbb{R}^{d+k}.$$

That is, first-type bisectors of a k-bag BVD are associated to hyperplanes of $\mathbb{R}^{d+k}$. It follows that the k-bag Voronoi diagram is obtained by

- Computing the power diagram of a set of n spheres of $\mathbb{R}^{d+k}$.
- Computing the restriction of this diagram to the convex d-dimensional submanifold $\{\hat{\mathbf{x}} = [\mathbf{x}\ F_1(\mathbf{x})\ \ldots\ F_k(\mathbf{x})]^T \mid \mathbf{x} \in \mathbb{R}^d\}$.
- Projecting this restricted diagram onto $\mathbb{R}^d$.

The complexity of a k-bag Voronoi diagram is thus at most $O(n^{\lfloor \frac{d+k}{2} \rfloor})$.

Theorem 6 *The k-bag Voronoi diagram (for $k > 1$) on a bag of d-variate Bregman divergences of a set of n points of $\mathbb{R}^d$ has combinatorial complexity $O(n^{\lfloor \frac{k+d}{2} \rfloor})$ and can be computed within the same time bound.*

k-bag divergences and their Voronoi diagrams have been used implicitly in recent works on Bregman hard k-means clustering [32]. k-bag Bregman Voronoi diagrams are also related to the anisotropic Voronoi diagrams of Labelle and Shewchuk [25] where to *each* point $\mathbf{x} \in \mathcal{X}$ a metric tensor $\mathbf{M}_{\mathbf{x}}$ is associated, which tells how lengths and angles should be measured from the local perspective of $\mathbf{x}$.

5 Bregman Triangulations

Consider the Euclidean Voronoi diagram vor($\mathcal{S}$) of a finite set $\mathcal{S}$ of points of $\mathbb{R}^d$ (called sites). Let f be a face of vor($\mathcal{S}$), that is, the intersection of k d-cells of vor($\mathcal{S}$). We associate to f a dual face f^*, namely the convex hull of the sites associated to the subset of cells. If no subsets of $d+2$ sites lie on a same sphere, the set of dual faces (of dimensions 0 to d) constitutes a triangulation embedded in $\mathbb{R}^d$ whose vertices are the sites. This triangulation is called the *Delaunay triangulation* of $\mathcal{S}$, noted del($\mathcal{S}$). The correspondence defined above between the faces of vor($\mathcal{S}$) and those of del($\mathcal{S}$) is a bijection that satisfies $f \subset g \Rightarrow g^* \subset f^*$. We say that del($\mathcal{S}$) is the *geometric dual* of vor($\mathcal{S}$). See Fig. 12.

A similar construct is known also for power diagrams. Consider the power diagram of a finite set of spheres of $\mathbb{R}^d$. In the same way as for Euclidean Voronoi diagrams, we can associate a triangulation dual to the power diagram of the spheres. This triangulation is called the *regular triangulation* of the spheres. The vertices of this triangulation are the centers of the spheres whose cell is non empty.

We introduce Bregman Delaunay triangulations and show that they capture some important properties of Delaunay triangulations.

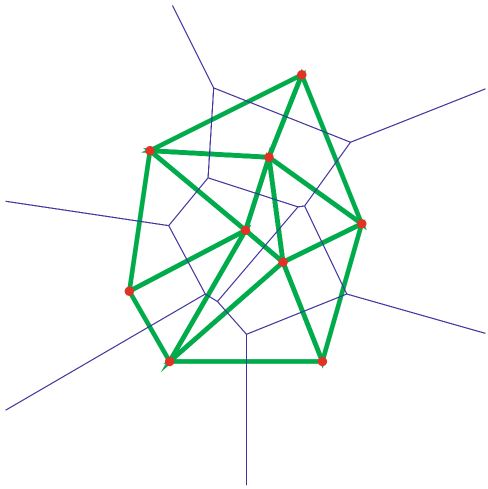


Fig. 12 Ordinary Voronoi diagram (*thin line*) and geometric dual Delaunay triangulation (*bold line*)

5.1 Bregman Delaunay Triangulations

Let $\hat{\mathcal{S}}$ be the lifted image of $\mathcal{S}$ and let $\mathcal{T}$ be the lower convex hull of $\hat{\mathcal{S}}$, i.e., the collection of facets of the convex hull of $\hat{\mathcal{S}}$ whose supporting hyperplanes are below $\hat{\mathcal{S}}$. We assume in this section that $\mathcal{S}$ is in *general position* if there is no subset of $d+2$ points lying on a same Bregman sphere. Equivalently (see Lemma 4), $\mathcal{S}$ is in general position if no subsets of $d+2$ points $\hat{\mathbf{p}}_i$ lie on the same hyperplane.

Under the general position assumption, each vertex of $\mathcal{H}=\bigcap_i H^{\uparrow}_{\mathbf{p}_i}$ is the intersection of exactly $d+1$ hyperplanes and the faces of $\mathcal{T}$ are all *simplices*. Moreover, the vertical projection $\mathrm{Proj}_{\mathcal{X}}(\mathcal{T})$ of $\mathcal{T}$ is a triangulation $\mathrm{del}_F(\mathcal{S})$ of $\mathcal{S}$ embedded in $\mathcal{X}\subseteq\mathbb{R}^d$ since the restriction of $\mathrm{Proj}_{\mathcal{X}}$ to $\mathcal{T}$ is bijective. Moreover, since F is convex, $\mathrm{del}_F(\mathcal{S})$ covers the convex hull of $\mathcal{S}$, and the set of vertices of $\mathcal{T}$ consists of all the $\hat{\mathbf{p}}_i$. Consequently, the set of vertices of $\mathrm{del}_F(\mathcal{S})$ is $\mathcal{S}$. We call $\mathrm{del}_F(\mathcal{S})$ the *Bregman Delaunay triangulation* of $\mathcal{S}$ (see Fig. 13). When $F(\mathbf{x})=\|\mathbf{x}\|^2$, $\mathrm{del}_F(\mathcal{S})$ is the Delaunay triangulation dual to the Euclidean Voronoi diagram. We will see (Theorem 11 below) that this duality property holds for general Bregman divergences.

We say that a Bregman sphere σ is *empty* if the open ball bounded by σ does not contain any point of $\mathcal{S}$. The following theorem extends a similar well-known property for Delaunay triangulations whose proof (see, for example, [8]) can be extended in a straightforward way to Bregman triangulations using the lifting map introduced in Sect. 3.2.

Theorem 7 *The first-type Bregman sphere circumscribing any simplex of* $\mathrm{del}_F(\mathcal{S})$ *is empty. If* $\mathcal{S}$ *is in general position*, $\mathrm{del}_F(\mathcal{S})$ *is the only triangulation of* $\mathcal{S}$ *with this property.*

Several other properties of Delaunay triangulations extend to Bregman triangulations. We list some of them.

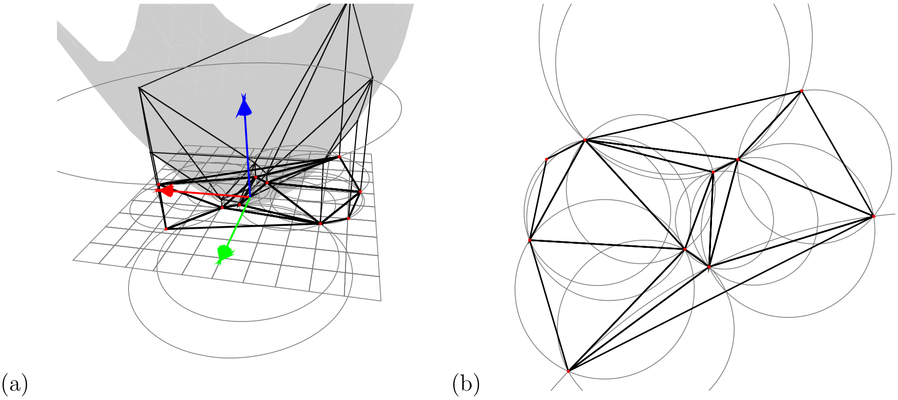


Fig. 13 Bregman Delaunay triangulation as the projection of the convex polyhedron $\mathcal{T}$. (**a**) The 3D convex polyhedron $\mathcal{T}$ of $\hat{\mathcal{X}}$ is shown in *thick lines* (wrt. the potential function F displayed in *grey*) and empty spheres are rasterized using *thin lines*. (**b**) The corresponding regular triangulation of $\mathcal{X}$

Theorem 8 (Empty ball) *Let $\mathcal{S} = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ be a set of n points in $\mathcal{X}$ in general position. If ν denotes a subset of at most $d+1$ indices in $\{1, \ldots, n\}$, the convex hull of the points $\mathbf{p}_i$, $i \in \nu$, is a simplex of the Bregman triangulation of $\mathcal{S}$ iff there exists an empty Bregman sphere σ passing through the $\mathbf{p}_i$, $i \in \nu$.*

The next property exhibits a local characterization of Bregman triangulations. Let $T(\mathcal{S})$ be a triangulation of $\mathcal{S}$. We say that a pair of adjacent facets $f_1 = (f, \mathbf{p}_1)$ and $f_2 = (f, \mathbf{p}_2)$ of $T(\mathcal{S})$ is regular iff $\mathbf{p}_1$ does not belong to the open Bregman ball circumscribing f_2 and $\mathbf{p}_2$ does not belong to the open Bregman ball circumscribing f_1 (the two statements are equivalent as is easily verified using the lifting map).

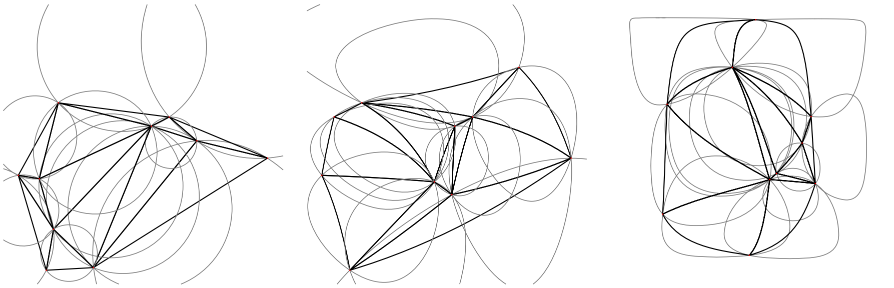
Theorem 9 (Locality) *Any triangulation of a given set of points $\mathcal{S}$ (in general position) whose pairs of facets are all regular is the Bregman triangulation of $\mathcal{S}$.*

Let $\mathcal{S}$ be a given finite set of points, $\mathrm{del}_F(\mathcal{S})$ its Bregman triangulation, and $\mathcal{T}(\mathcal{S})$ the set of all triangulations of $\mathcal{S}$. We define the min-containment Bregman radius of a d-simplex τ as the radius, denoted $r_{mc}(\tau)$, of the smallest Bregman ball containing τ. We further define the maximal min-containment Bregman radius of a triangulation $T \in \mathcal{T}(\mathcal{S})$ as $r_{mc}(T) = \max_{\tau \in T} r_{mc}(\tau)$. The following result is an extension of a result due to Rajan for Delaunay triangulations [35].

Theorem 10 (Max–min-containment) *For a given finite set of points $\mathcal{S}$, $r_{mc}(\mathrm{del}_F(\mathcal{S})) = \min_{T \in \mathcal{T}(\mathcal{S})} r_{mc}(T)$.*

The proof mimics Rajan's proof [35] for the case of Delaunay triangulations.

We will now show that $\mathrm{del}_F(\mathcal{S})$ is the geometric dual of $\mathrm{vor}_F(\mathcal{S})$. To this aim, we first introduce another (curved) triangulation of $\mathcal{S}$ that we call the Bregman geodesic triangulation of $\mathcal{S}$.



(a) Ordinary Delaunay (b) Exponential loss (c) Hellinger-like divergence

Fig. 14 An ordinary Delaunay triangulation (**a**) and two Bregman geodesic triangulations for the exponential loss (**b**) and for the Hellinger-like divergence (**c**). The Bregman balls circumscribing the simplices are shown in *light grey*

We have seen in Sect. 4.2 that the Bregman Voronoi diagram of a set of points $\mathcal{S}$ is the power diagram of a set of spheres $\mathcal{B}'$ centered at the points of $\mathcal{S}'$ (Theorem 4). Write $\mathrm{reg}_F(\mathcal{B}')$ for the regular triangulation dual to this power diagram. This triangulation[4] is embedded in $\mathcal{X}'$ and has the points of $\mathcal{S}'$ as its vertices. The image of this triangulation by $\nabla^{-1}F$ is a *curved triangulation*, denoted $\mathrm{del}'_F(\mathcal{S})$, whose vertices are the points of $\mathcal{S}$. The edges of $\mathrm{del}'_F(\mathcal{S})$ are curved arcs joining two sites. Since these arcs are geodesic arcs (see Sect. 3.3), we call $\mathrm{del}'_F(\mathcal{S})$ the *Bregman geodesic triangulation* of $\mathcal{S}$ (see Fig. 14).

Theorem 11 (Duality) *The Bregman Delaunay triangulation* $\mathrm{del}_F(\mathcal{S})$ *is the geometric dual of the* 1*st-type Bregman Voronoi diagram of* $\mathcal{S}$.

Proof We have, denoting $\overset{*}{\leftrightarrow}$ the dual mapping, and using Theorem 4

$$\mathrm{vor}_F(\mathcal{S}) = \mathrm{pow}\big(\mathcal{B}'\big) \overset{*}{\leftrightarrow} \mathrm{reg}\big(\mathcal{B}'\big) = \nabla F\big(\mathrm{del}'_F(\mathcal{S})\big). \tag{8}$$

It follows that $\mathrm{del}'_F(\mathcal{S})$ is a (curved) triangulation dual to $\mathrm{vor}_F(\mathcal{S})$.

We now show that $\mathrm{del}'_F(\mathcal{S})$ is isomorphic to $\mathrm{del}_F(\mathcal{S})$. Indeed, the two triangulations are embedded in $\mathbb{R}^d$, have the same vertices, and their d-simplices are in 1–1 correspondence. The last claim comes from the fact that the d-simplices of $\mathrm{del}'_F(\mathcal{S})$ are in 1–1 correspondence with the vertices of $\mathrm{vor}_F(\mathcal{S})$ by (8), and that the d-simplices of $\mathrm{del}_F(\mathcal{S})$ are in 1–1 correspondence with the centers of their circumscribing Bregman spheres, which are precisely the vertices of $\mathrm{vor}_F(\mathcal{S})$. □

6 Conclusion

We have defined the notion of Bregman Voronoi diagrams and showed how these geometric structures are a natural extension of ordinary Voronoi diagrams. Bregman

[4] Applet at http://www.csl.sony.co.jp/person/nielsen/BVDapplet/.

Voronoi diagrams share with their Euclidean analogs surprisingly similar combinatorial and geometric properties. In particular, we have shown how to define and build Bregman Voronoi diagrams using power diagrams and Legendre duality.

We hope that our results will make Voronoi diagrams and their relatives applicable in new application areas. In particular, Bregman Voronoi diagrams based on various entropic divergences are expected to find applications in information retrieval (IR), data mining, knowledge discovery in databases, image processing (e.g., see [22]). The study of Bregman Voronoi diagrams raises the question of revisiting computational geometry problems in this new light. This may also allow one to tackle uncertainty ('noise') in computational geometry for fundamental problems such as surface reconstruction or pattern matching. Bregman Voronoi diagrams can be extended using representational functions [30]. This allows one to compute other information-theoretic Voronoi diagrams for well-known divergences in information geometry: namely the α-divergences and the β-divergences.

A limitation of Voronoi diagrams and, in particular, of Bregman Voronoi diagrams is their combinatorial complexity that depends exponentially on the dimension (McMullen's upper bound theorem [27]). Since many applications are in high dimensional spaces, one may consider instead related but easier to compute data structures such as the witness complex [10, 19].

Acknowledgements Frédéric Chazal, David Cohen-Steiner and Mariette Yvinec are gratefully acknowledged for their comments on this paper. We thank the anonymous reviewers for their constructive comments. This research has been partially supported by the Agence Nationale de la Recherche (project GAIA 07-BLAN-0328-04) and DIGITEO (project GAS 2008-16D).

References

1. Amari, S., Nagaoka, H.: Methods of Information Geometry. Oxford University Press, London (2000). ISBN 13 9780821843024
2. Atkinson, C., Mitchell, F.S.: Rao's distance measure. Sankhya Indian J. Stat., Ser. A **43**, 345–365 (1981)
3. Aurenhammer, F.: Power diagrams: Properties, algorithms and applications. SIAM J. Comput. **16**(1), 78–96 (1987)
4. Aurenhammer, F., Imai, H.: Geometric relations among Voronoi diagrams. In: 4th Annual Symposium on Theoretical Aspects of Computer Sciences (STACS), pp. 53–65. Springer, London (1987)
5. Aurenhammer, F., Klein, R.: Voronoi diagrams. In: Sack, J., Urrutia, G. (eds.) Handbook of Computational Geometry, pp. 201–290. Elsevier, Amsterdam (2000). Chap. V
6. Banerjee, A., Merugu, S., Dhillon, I.S., Ghosh, J.: Clustering with Bregman divergences. J. Mach. Learn. Res. **6**, 1705–1749 (2005)
7. Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V.: Support vector clustering. J. Mach. Learn. Res. **2**, 125–137 (2002)
8. Boissonnat, J.-D., Yvinec, M.: Algorithmic Geometry. Cambridge University Press, New York (1998)
9. Boissonnat, J.-D., Wormser, C., Yvinec, M.: Curved Voronoi diagrams. In: Boissonnat, J.-D., Teillaud, M. (eds.) Effective Computational Geometry for Curves and Surfaces. Mathematics and Visualization, pp. 67–116. Springer, Berlin (2007)
10. Boissonnat, J.-D., Guibas, L.J., Oudot, S.Y.: Manifold reconstruction in arbitrary dimensions using witness complexes. Discrete Comput. Geom. **42**(1), 37–70 (2009)
11. Bregman, L.M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Comput. Math. Math. Phys. **7**, 200–217 (1967)
12. Brönnimann, H., Goodrich, M.T.: Optimal set covers in finite VC-dimension. Discrete Comput. Geom. **14**(4), 463–479 (1995)

13. Censor, Y.A., Zenios, S.A.: Parallel Optimization: Theory, Algorithms and Applications. Oxford University Press, London (1997)
14. Chan, T.M., Snoeyink, J., Yap, C.-K.: Output-sensitive construction of polytopes in four dimensions and clipped Voronoi diagrams in three. In: SODA '95: Proceedings of the Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 282–291. SIAM, Philadelphia (1995)
15. Chazelle, B.: An optimal convex hull algorithm in any fixed dimension. Discrete Comput. Geom. **10**, 377–409 (1993)
16. Chazelle, B.: The Discrepancy Method. Cambridge University Press, Cambridge (2000)
17. Clarkson, K., Shor, P.: Applications of random sampling in computational geometry, ii. Discrete Comput. Geom. **4**, 387–421 (1989)
18. Csiszár, I.: Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. Ann. Stat. **19**, 2032–2066 (1991)
19. de Silva, V.: A weak characterisation of the Delaunay triangulation. Geom. Dedic. **135**(1), 39–64 (2008)
20. Descartes, R.: Principia Philosophiae. Ludovicus Elzevirius, Amsterdam (1644)
21. Even, G., Rawitz, D., Shahar, S.: Hitting sets when the VC-dimension is small. Inf. Process. Lett. **95**(2), 358–362 (2005)
22. Inaba, M., Imai, H.: Geometric clustering models for multimedia databases. In: Proceedings of the 10th Canadian Conference on Computational Geometry (CCCG'98) (1998)
23. Inaba, M., Imai, H.: Geometric clustering for multiplicative mixtures of distributions in exponential families. In: Proceedings of the 12th Canadian Conference on Computational Geometry (CCCG'00) (2000)
24. Klein, R.: Concrete and Abstract Voronoi Diagrams. Lecture Notes in Computer Science, vol. 400. Springer, Berlin (1989). ISBN 3-540-52055-4
25. Labelle, F., Shewchuk, J.R.: Anisotropic Voronoi diagrams and guaranteed-quality anisotropic mesh generation. In: Proc. 19th Symposium on Computational Geometry (SoCG), pp. 191–200. ACM, New York (2003)
26. Matousek, J.: Lectures on Discrete Geometry. Springer, Berlin (2002)
27. McMullen, P.: The maximum numbers of faces of a convex polytope. J. Comb. Theory, Ser. B **10**, 179–184 (1971)
28. Nielsen, F.: Visual Computing: Geometry, Graphics, and Vision. Charles River Media/Thomson Delmar Learning (2005)
29. Nielsen, F., Garcia, V.: Statistical exponential families: A digest with flash cards. arXiv:0911.4863 (2009)
30. Nielsen, F., Nock, R.: The dual Voronoi diagrams with respect to representational Bregman divergences. In: International Symposium on Voronoi Diagrams (ISVD), DTU Lyngby, Denmark, June 2009. IEEE Press, New York (2009)
31. Nielsen, F., Nock, R.: Hyperbolic Voronoi diagrams made easy. arXiv:0903.3287 (March 2009)
32. Nock, R., Luosto, P., Kivinen, J.: Mixed Bregman clustering with approximation guarantees. In: ECML PKDD '08: Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases—Part II, pp. 154–169. Springer, Berlin (2008)
33. Onishi, K., Imai, H.: Voronoi diagram in statistical parametric space by Kullback–Leibler divergence. In: Proc. 13th Symposium on Computational Geometry (SoCG), pp. 463–465. ACM, New York (1997)
34. Onishi, K., Imai, H.: Voronoi diagrams for an exponential family of probability distributions in information geometry. In: Japan–Korea Joint Workshop on Algorithms and Computation, pp. 1–8 (1997)
35. Rajan, V.T.: Optimality of the Delaunay triangulation in $\mathbb{R}^d$. Discrete Comput. Geom. **12**, 189–202 (1994)
36. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
37. Sadakane, K., Imai, H., Onishi, K., Inaba, M., Takeuchi, F., Imai, K.: Voronoi diagrams by divergences with additive weights. In: Proc. 14th Symposium on Computational Geometry (SoCG), pp. 403–404. ACM Press, New York (1998)
38. Teillaud, M., Devillers, O., Meiser, S.: The space of spheres, a geometric tool to unify duality results on Voronoi diagrams. Technical Report Rapports de Recherche No. 1620, Institut National de Recherche en Informatique et en Automatique (1992)