On Approximating the Smallest Enclosing Bregman Balls

Frank Nielsen
Sony Computer Science Laboratories Inc.
Tokyo, Japan
Frank.Nielsen@acm.org

Richard Nock
University of Antilles-Guyane
Martinique, France
Richard.Nock@martinique.univ-ag.fr

ABSTRACT

We present a generalization of Bădoiu and Clarkson's algorithm [3] for computing a $(1+\epsilon)$-approximation of the smallest enclosing ball of a point set equipped with a Bregman divergence as a distortion measure.

Categories and Subject Descriptors
I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—*Geometric algorithms, languages, and systems*

General Terms:
Algorithms, Theory

Keywords
Minimum enclosing ball, core-set, Bregman divergence

1. INTRODUCTION

Given a set $\mathcal{S} = \{\mathbf{s}_i\}_{i=1}^n$ of d-dimensional points, we are interested in finding a center point $\mathbf{c}^*$ of $\mathcal{S}$, where the "distance" measure $d(\cdot,\cdot)$ between any two points of $\mathcal{S}$, is a divergence (also called distortion[1]). Two optimization criteria may be considered for finding such a center: MinAvg which minimizes the *average divergence* or MinMax which minimizes the *maximal divergence* ($\mathbf{c}^* = \text{argmin}_{\mathbf{c}} d(\mathbf{c}, \mathcal{S})$, where $d(\mathbf{c}, \mathcal{S})$ denote the divergence from $\mathbf{c}$ to the furthest point of $\mathcal{S}$: $d(\mathbf{c}, \mathcal{S}) = \max_i d(\mathbf{c}, \mathbf{s}_i)$). These problems have been widely studied in computational geometry (1-center problem), computational statistics (1-point estimator), and machine learning (1-class classification). It is known that for the squared Euclidean distance (L_2^2) the *centroid* $\frac{1}{n}\sum_{i=1}^n \mathbf{s}_i$ is the MinAvg(L_2^2) center [2]. For the Euclidean distance L_2, the *circumcenter* of $\mathcal{S}$ is the MinMax(L_2) center, and the *Fermat-Weber point* is the MinAvg(L_2) center (see Figure 1). Finding the circumcenter of the unique smallest enclosing ball of $\mathcal{S}$ is weakly polynomial, and can be solved efficiently either numerically using second-order cone programming (SOCP), or combinatorially using a recent ball deflating heuristic, up to dimension 1000 and more [1].

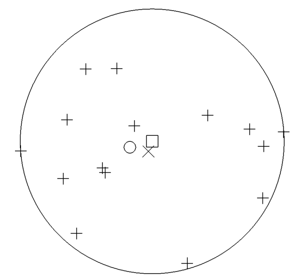

Figure 1: Three centers of a point set: centroid ($\times$), circumcenter ($\square$) and Fermat-Weber point ($\circ$).

2. BREGMAN DIVERGENCES

In computational machine learning, the L_2 geometric distance seldomly reflects the distance between two d-dimensional points. A more general distance framework, known as *Bregman divergences*, is rather used. Bregman divergences D_F are parameterized families of distortions defined on a convex domain $\mathcal{X} \subseteq \mathbb{R}^d$ for strictly convex and differentiable functions F on $\text{int}(\mathcal{X})$ (Figure 2) as $D_F(\mathbf{p}, \mathbf{q}) = F(\mathbf{p}) - F(\mathbf{q}) - \langle \mathbf{p} - \mathbf{q}, \boldsymbol{\nabla}_F(\mathbf{q})\rangle$, where $\boldsymbol{\nabla}_F$ denotes the gradient operator, and $\langle\cdot,\cdot\rangle$ the inner product (dot product). Informally speaking, a Bregman divergence D_F is the tail of a Taylor expansion of F. Bregman divergences include the squared Euclidean distance $D_F(\mathbf{p}, \mathbf{q}) = ||\mathbf{p} - \mathbf{q}||^2$ ($F(\mathbf{x}) = ||\mathbf{x}||^2$), the Kullback-Leibler divergence (also known as the Information divergence) $D_F(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^d p_i \log \frac{p_i}{q_i}$ ($F(\mathbf{x}) = \sum_{i=1}^d x_i \log x_i$ the negative entropy defined on the d-simplex), and the Itakura-Saito divergence $D_F(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^d (\frac{p_i}{q_i} - \log \frac{p_i}{q_i} - 1)$ ($F(\mathbf{x}) = -\sum_{i=1}^d \log x_i$ defined on $\mathbb{R}_+^d$). Bregman divergences define two families of Bregman balls: $\mathcal{B}_{\mathbf{c},r} = \{\mathbf{x} \in \mathcal{X} : D_F(\mathbf{c}, \mathbf{x}) \leq r\}$ and $\mathcal{B}'_{\mathbf{c},r} = \{\mathbf{x} \in \mathcal{X} : D_F(\mathbf{x}, \mathbf{c}) \leq r\}$, that are not necessarily convex nor identical (Figure 3). For a point set $\mathcal{S}$, the two smallest enclosing balls $\mathcal{B}^*(\mathcal{S})$ and $\mathcal{B}'^*(\mathcal{S})$ have been shown to be unique [4]. Bregman balls have many important applications in machine learning. For example, finding the minimum enclosing Itakura-Saito ball

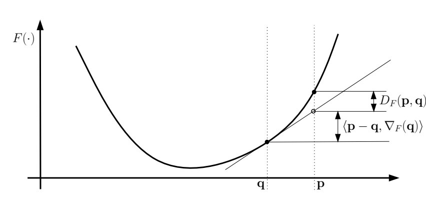

Figure 2: Visualizing the convex function F and its associated Bregman divergence $D_F(\cdot,\cdot)$.

*`http://www.csl.sony.co.jp/person/nielsen/BregmanBall`
[1] Symmetry and triangle inequality properties may not hold.

SCG'06, June 5–7, 2006, Sedona, Arizona, USA.
ACM 1-59593-340-9/06/0006.

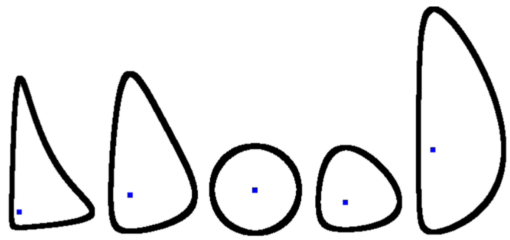

Figure 3: Examples of Itakura-Saito (×2), squared Euclidean (×1) and Kullback-Leibler (×2) balls.

is used to find the closest signal to a set of given signals in speech recognition [4].

3. APPROXIMATION ALGORITHMS

Bădoiu and Clarkson [3] introduced the notion of core-sets for balls. An ϵ-core-set $\mathcal{C}$ for the MinMax ball of $\mathcal{S}$ is a subset $\mathcal{C} \subseteq \mathcal{S}$ such that the circumcenter $\mathbf{c}$ of the MinMax ball of $\mathcal{C}$ is such that $d(\mathbf{c}, \mathcal{S}) \leq (1+\epsilon)r^*$, where r^* is the radius of the smallest enclosing ball of $\mathcal{S}$. They show that core-set sizes are *independent* of the dimension, and of size bounded by $\frac{2}{\epsilon}$. Further, they described a simple guaranteed $O(\frac{dn}{\epsilon^2})$-time $(1+\epsilon)$-approximation algorithm. Applying this algorithm to "skewed" divergences does not make sense and yield poor results [4]. In [4], we generalize their approximation algorithm to *arbitrary* Bregman divergences: `BBC`.

Choose at random $\mathbf{c} \in \mathcal{S}$
for $t = 1, 2, ..., T$ **do**
$\mathbf{s} \leftarrow \arg\max_{\mathbf{s}' \in \mathcal{S}} D_F(\mathbf{c}, \mathbf{s}')$
$\mathbf{c} \leftarrow \boldsymbol{\nabla}_F^{-1}\left(\frac{t}{t+1}\boldsymbol{\nabla}_F(\mathbf{c}) + \frac{1}{t+1}\boldsymbol{\nabla}_F(\mathbf{s})\right)$

4. EXPERIMENTS AND CONCLUSION

Figure 4 presents our experiments for Itakura-Saito and Kullback-Leibler divergences. Empirical studies suggest that the theoretical convergence is well below the upper bound $\frac{1}{T}$ of [3]. Figure 5 depicts the observed convergence rate for the Kullback-Leibler divergence. Banerjee et al. [2] proved that $\text{MinAvg}(D_F)$ is always the centroid whatever the Bregman divergence, and described a bijection between Bregman divergences and the exponential families in statistics. We exhibit yet another bijection between Bregman divergences and functional averages of core-sets [4].

5. REFERENCES

[1] F. Nielsen and R. Nock. Approximating smallest enclosing balls. In *ICCSA CGA*, pp. 147–157. LNCS #3045, Springer-Verlag, pp. 147–157, 2004.

[2] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh. Clustering with Bregman divergences. In *Journal of Machine Learning* (JMLR), 6:1705–1749, 2005.

[3] M. Bădoiu and K. L. Clarkson. Smaller core-sets for balls. In *ACM-SIAM Sympos. Discret Algorithms*, pp. 801–802, 2003. `http://cm.bell-labs.com/who/clarkson/`

[4] R. Nock and F. Nielsen. Fitting the smallest enclosing Bregman ball. In *European Conference on Machine Learning* (ECML), LNCS #3720, pp. 649–656, 2005.

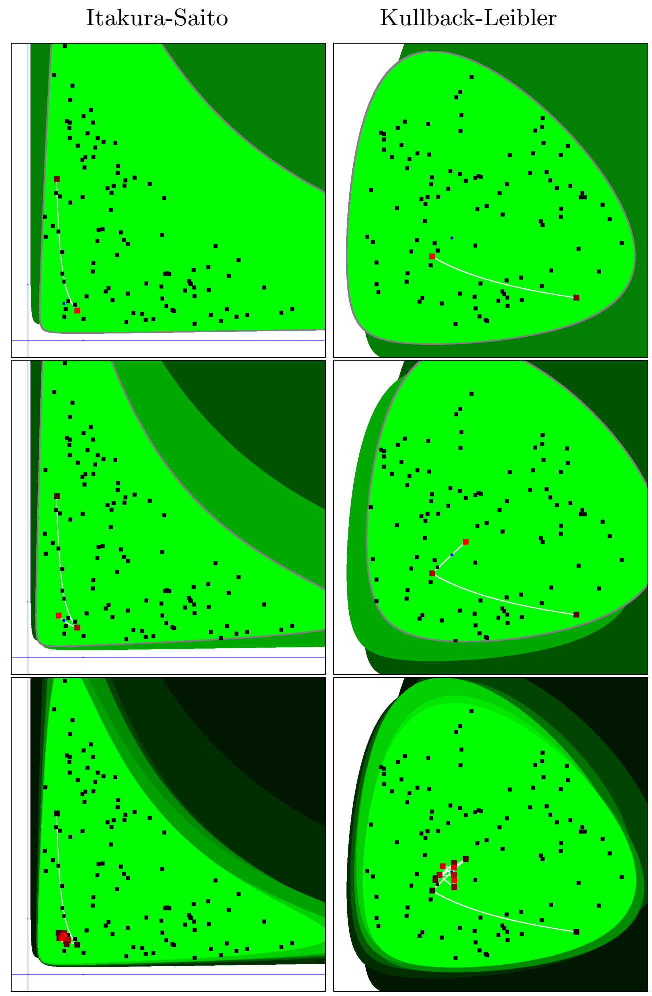

Figure 4: Algorithm `BBC`: Bregman enclosing balls after the first, second, and 10th iteration for the Itakura-Saito and Kullback-Leibler divergences. White lines depict the geodesics followed by c.

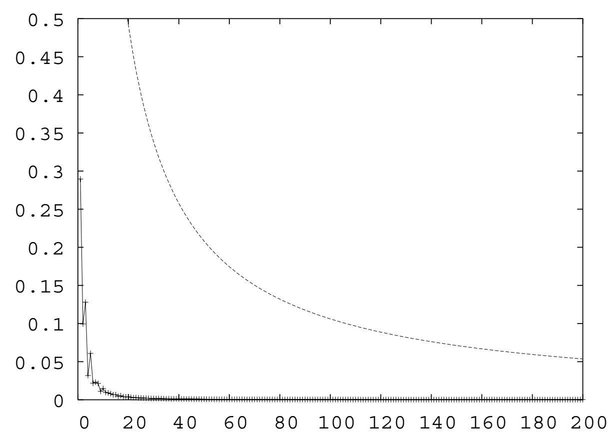

Figure 5: Convergence rate for the Kullback-Leibler divergence: $n = 1000$, $d = 2$ on 100 runs. Plain curve represent $\frac{D_F(\mathbf{c}^*, \mathbf{c}) + D_F(\mathbf{c}, \mathbf{c}^*)}{2}$, and dashed curve is the $\frac{1}{T}$ upperbound for L_2^2 [3].