Lines: 80

Kd-trees work with separable Bregman divergences

Tuyen Pham ✉
University of Florida, Gainesville, US

Hubert Wagner ✉
University of Florida, Gainesville, US

Abstract

We report work in progress on generalizing Kd-trees to a broad family of Bregman divergences. Our focus is on separable Bregman divergences, which include practical divergences such as the Kullback–Leibler divergence (relative entropy) and the Itakura–Saito divergence.

2012 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems; Theory of computation → Computational geometry; Mathematics of computing → Combinatorial algorithms

Keywords and phrases Kd-tree, nearest neighbor search, Bregman divergence, separable Bregman divergence, KL-divergence, relative entropy

Funding 2022 Google Scholar Award in Algorithms and Optimization.

1 Introduction

Many techniques in computational geometry were developed with the Euclidean distance – or a more general metric distance – in mind. However, in many applied settings – particularly modern machine learning – non-metric distances play an important role. One example is the Kullback–Leibler divergence (KL), commonly used to compare discrete probability distributions [8]. In practice, it often serves as a loss to be minimized [15, 9, 10] – often under the name of relative entropy or the related cross entropy. KL is one member of the family of Bregman divergences – on which we center our attention.

A range of computational geometry techniques have been extended to the setting of Bregman divergences. This includes vantage-point trees [12], ball-trees [13], k-means clustering [1], Voronoi diagrams and Delaunay triangulations [3], Čech complexes [6]. We show that Kd-trees work for a class of Bregman divergences called separable Bregman divergences [11].

2 Background

that

We begin by setting up the definitions for Bregman divergences [4], which will serve as measures of distance. We note they are usually not symmetric and never satisfy the triangle inequality – and as such do not define a proper metric.

A *function of Legendre Type* [14] is a function $F : \Omega \to \mathbb{R}$ where $\Omega \subseteq \mathbb{R}^n$ is a convex set and F is differentiable, strictly convex, and satisfies $\lim_{x \to \partial \Omega} \|\nabla F(x)\| = \infty$ if $\partial \Omega$ is nonempty. Given a function of Legendre type, the Bregman divergence is

$$D_F(x\|y) = F(x) - (F(y) + \langle \nabla F(y), x - y \rangle).$$

} give intuition!

Separable Bregman divergences. In particular, a separable Bregman divergence can be decomposed into univariate divergences, i.e., $D_F(x\|y) = \sum_{i=1}^{n} D_{f_i}(x_i\|y_i)$ [11]. These divergences have seen many uses in machine learning. One popular example of separable Bregman divergences is the Kullback–Leibler (KL) divergence, $D_{KL}(x\|y) = \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{y_i}\right)$, \mathrm{KL} which is a standard distance between probability distributions [8] and is commonly used as a loss function in algorithms such as t-SNE [15], UMAP [9] and in deep learning [10]. We work

The 39th International Symposium on Computational Geometry.
Editors: Xavier Goaoc and Michael Kerber; Article No. 63; pp. 63:1–63:4
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

 ← Use full paper title.

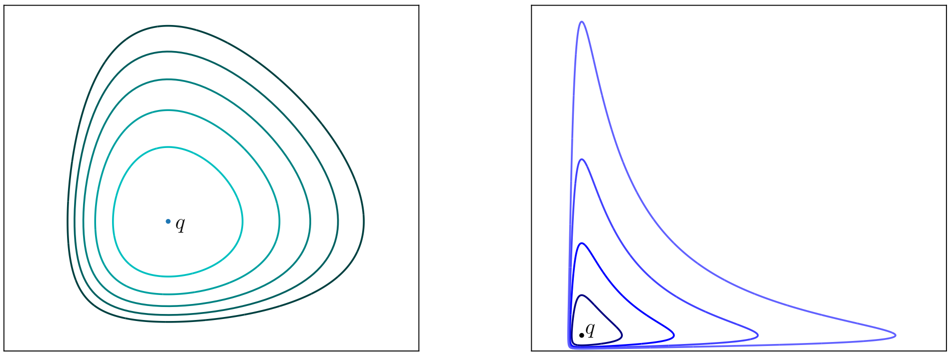

Figure 1 Left: primal generalized Kullback–Leibler balls. Right: Itakura–Saito balls

Don't use \displaystyle in the text.

with the generalized KL divergence $D_{GKL}(x\|y) = \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{y_i}\right) + \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i$ defined on $\mathbb{R}^n_+$ (which reduces to the usual KL for points on the open standard simplex). Another example is the Itakura–Saito (IS) divergence [7], $D_{IS}(x\|y) = \sum_{i=1}^{n} \left(\frac{x_i}{y_i} - \log\left(\frac{x_i}{y_i}\right) - 1\right)$ defined on $\mathbb{R}^n_+$, which is useful for speech and sound data [5].

and center q

We define the *primal Bregman ball* for radius $r \geq 0$ as $B_F(q;r) = \{y \in \Omega : D_F(q\|y) \leq r\}$ which contains all points y such that the divergence from the center q to y is at most r. See Figure 1 for an illustration.

Kd-trees. We briefly overview a simple version of the Kd-tree [2], which is a useful data-structure for nearest neighbor queries. We construct it as a binary tree encoding repeated partitions of Ω with axis-aligned hyperplanes. In finding the nearest neighbor of $q \in \Omega$ among $X \subset \Omega$, the crucial step is checking if the points on the other side of a hyperplane can be safely pruned. This reduces to checking if the hyperplane intersects the ball centered at q of radius equal to the distance to the current nearest neighbor candidate. With the Euclidean metric, this check is trivial. We now consider a version of this problem for separable Bregman divergences. We focus on finding $\operatorname{argmin}_{x \in X} D_F(q\|x)$; the other case is analogous.

3 Hyperplane Intersection Problem

Given an axis-aligned hyperplane $P \subset \mathbb{R}^n$ and a point $q \in \Omega \subset \mathbb{R}^n$, we check if the hyperplane intersects a Bregman ball of radius $r \geq 0$ centered at q. This can be solved by finding the Bregman projection of q onto P, namely the point $q_P = \operatorname{arginf}_{p \in P \cap \Omega} D_F(q\|p)$.

let

▶ **Lemma 1** (Hyperplane projection lemma). *Let $P \subset \mathbb{R}^n$ be an axis-aligned hyperplane, such that $P \cap \Omega \neq \emptyset$, and D_F be a separable Bregman divergence. Given $q \in \Omega$, the Bregman projection of q onto P is well-defined and coincides with the Euclidean projection of q onto P.*

Proof. Without loss of generality, we consider P with fixed $p_1 = c$ for each $p \in P$. Let $q = (q_1, q_2, \ldots, q_n)$ and $D_F(x\|y) = \sum_{i=1}^{n} D_{f_i}(x_i\|y_i)$. Since each D_{f_i} is a Bregman divergence, $D_{f_i}(x\|y) \geq 0$ with $D_{f_i}(x_i\|y_i) = 0$ if and only if $x_i = y_i$. As q is fixed and $p_1 = c$ for ~~all~~ each $p \in P$, $\min_{p \in P \cap \Omega} D_F(q\|p) \geq D_{f_1}(q_1\|p_1)$, with equality if and only if $p_i = q_i$ for $i = 2, 3, \ldots, n$. With

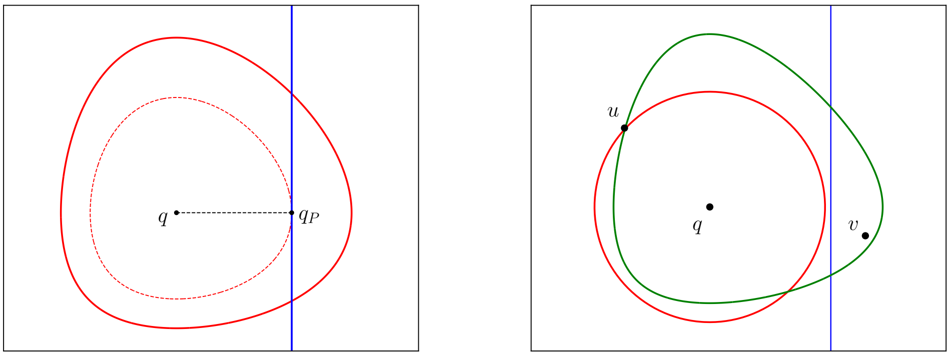

Figure 2 Left: A GKL ball intersects the hyperplane and contains the ball of radius $\inf_{p\in P} D_{GKL}(q\|p)$. Right: An IS ball intersects the hyperplane while the Euclidean ball does not, resulting in different nearest neighbors.

$f_i : \omega_i \to \mathbb{R}$ being a function of Legendre type, D_F is well-defined on Ω^2 where $\Omega = \prod_{i=1}^{n} \omega_i$ and so $q_P = (p_1, q_2, q_3, \cdots, q_n) \in \Omega$ and $D_F(q\|q_P)$ is well defined. \dots

Thus the Bregman projection of q onto an arbitrary axis-aligned hyperplane P with the i^{th} coordinate fixed at p_i is given by $q_P = (q_0, q_1, \ldots, q_{i-1}, p_i, q_{i+1}, \ldots, q_n)$. ◀

With this projection, we simply compare $D_F(q\|q_P)$ with a/the r to determine if the hyperplane intersects the Bregman ball $B_F(q;r)$. Despite the lack of the triangle inequality, this allows us to decide if the points on the ~~other~~ side of the hyperplane ~~than~~ opposite to q can be safely pruned. Indeed, these points are in the complement of the Bregman ball and thus have greater Bregman divergence. See Figure 2 for an illustration and Appendix A for an ~~extra~~ additional comment.

4 Summary

We showed that Kd-trees work with separable Bregman divergences. In particular, this result allows for efficient queries of probability distributions measured with the Kullback–Leibler divergence. We intend to sharpen this result for the important special case of points living on the simplex, which differs significantly from the Euclidean case.

We plan to test the efficiency of Kd-trees for various divergences and compare it to the alternatives mentioned in the introduction. We stress that, unlike most other methods, Kd-trees allow~~s~~ the divergence to be chosen *after* the data-structure is constructed. It makes the method an interesting choice for situations in which queries with respect to multiple divergences are performed or when the divergence changes over time.

References

1 Arindam Banerjee, Srujana Merugu, Inderjit S. Dhillon, and Joydeep Ghosh. Clustering with Bregman divergences. *Journal of Machine Learning Research*, 6(58):1705–1749, 2005. URL: http://jmlr.org/papers/v6/banerjee05b.html.

2 Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18:509–517, 1975.

3 Jean-Daniel Boissonnat, Frank Nielsen, and Richard Nock. Bregman Voronoi diagrams. *Discrete and Computational Geometry*, 44:281–307, 02 2010. doi:10.1007/s00454-010-9256-1.

4 L.M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967. ~~URL: https://www.sciencedirect.com/science/article/pii/0041555367900407~~, doi:https://doi.org/10.1016/0041-5553(67)90040-7.

5 Minh N Do and Martin Vetterli. Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance. *IEEE transactions on image processing*, 11(2):146–158, 2002.

6 Herbert Edelsbrunner and Hubert Wagner. Topological data analysis with Bregman divergences. In *33th International Symposium on Computational Geometry (SoCG ~~2017~~)*, 2017.

7 Fumitada Itakura. Analysis synthesis telephony based on the maximum likelihood method. *Reports of the* 6^{th} *Int. Cong. Acoust.*, ~~*1968*,~~ 1968.

8 Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

9 Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

10 Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

11 Frank Nielsen and Richard Nock. The bregman chord divergence. In *Geometric Science of Information: 4th International Conference, GSI ~~2019~~, ~~Toulouse, France, August 27–29, 2019, Proceedings 4~~,* pages 299–308. Springer, 2019.

12 Frank Nielsen, Paolo Piro, and Michel Barlaud. Bregman vantage point trees for efficient nearest neighbor queries. pages 878 – 881, 08 2009. doi:10.1109/ICME.2009.5202635. ?

13 Frank Nielsen, Paolo Piro, and Michel Barlaud. Tailored Bregman ball trees for effective nearest neighbors. In *Proceedings of the 25th European Workshop on Computational Geometry (EuroCG)*, pages 29–32, 2009.

14 Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, ~~Princeton~~, 1970. URL: ~~https://doi.org/10.1515/9781400873173, doi:~~doi:10.1515/9781400873173.

15 Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.
SNE

A Connectedness of Bregman balls.

We further comment on properties of Bregman balls to rule out potentially strange behavior, particularly lack of connectedness.

Primal Bregman balls may be nonconvex (when viewed as a subset of Euclidean space), but are connected (in the topology induced by the ambient Euclidean space). One way to see this is by considering the Legendre transform of a primal Bregman ball, in analogy to the proof of contractility of nonempty intersections of primal Bregman balls [6]. Its image is a dual Bregman ball (for some other Bregman divergence, potentially over a different convex domain) which is known to be convex hence connected [14]. Since the transform is a homeomorphism, a primal ball is connected as a homeomorphic image of a connected set.