Electronic Colloquium on Computational Complexity, Report No. 15 (2011)

Hardness and Non-Approximability of Bregman Clustering Problems*

Marcel R. Ackermann† Johannes Blömer† Christoph Scholz‡

December 8, 2010

Abstract

We prove the computational hardness of three k-clustering problems using an (almost) arbitrary Bregman divergence as dissimilarity measure: (a) The *Bregman k-center problem*, where the objective is to find a set of centers that minimizes the maximum dissimilarity of any input point towards its closest center, and (b) the *Bregman k-diameter problem*, where the objective is to minimize the maximum dissimilarity between pairs of points from the same cluster, and (c) the *Bregman k-median problem*, where the objective is to find a set of centers that minimizes the average dissimilarity of any input point towards its closest center. We show that solving these problems is $\mathcal{NP}$-hard, and that it is even $\mathcal{NP}$-hard to approximate a solution of (a) and (b) within a factor of (a) 3.32 and (b) 3.87, respectively. To obtain our results, we give a gap-preserving reduction from the Euclidean k-center (k-diameter, k-means) problem to the Bregman k-center (k-diameter, k-median) problem. This reduction combines the technique of Mahalanobis-similarity from Ackermann et al. (SODA '08) with a reduction already used by Chaudhuri and McGregor (COLT '08) to show the non-approximability of the Kullback-Leibler k-center problem, and a recent reduction given by Vattani to prove the $\mathcal{NP}$-hardness of the Euclidean k-means problem.

1 Introduction

Clustering is the problem of partitioning a set of data points into subsets (called clusters) such that points in a common cluster are similar. The quality of a given clustering is measured using a well defined objective function which may vary from application to application. Given an arbitrary dissimilarity function on the data points, there are three objective functions that have proven to be useful in practice: *k-center clustering*, where the objective is to find a set of centers that minimizes the maximum dissimilarity of any input point towards its closest center; *k-median clustering*, where the objective is to find a set of centers that minimizes the average dissimilarity of any input point towards its closest center; and *k-diameter clustering*, where the objective is to minimize the maximum dissimilarity between pairs of points from the same cluster.

The problem of clustering by these objective functions has received a lot of attention if the dissimilarity measure used is a metric (like the Euclidean distance) or at least the square of a metric. In particular, the computational hardness of finding an optimal k-center (k-median, k-diameter) clustering has been proven [10, 19, 21, 34, 15, 14, 26, 3, 31]. Additionally, in case of k-center or k-diameter clustering, it has also been shown that a clustering arbitrary close to the optimal solution can not be found in polynomial time if $k \geq 2$ is part of the input, unless $\mathcal{P} = \mathcal{NP}$ [35, 21, 23, 36, 16]. An overview of known hardness and non-approximability results for Euclidean k-clustering problems can be found in Table 1.

However, relatively little is known about the computational complexity of these clustering problems if the dissimilarity measure is neither a metric nor the square of a metric, or even asymmetric. Yet, there are a

*research supported by Deutsche Forschungsgemeinschaft (DFG), grant BL-314/6-1

†Department of Computer Science, University of Paderborn, {mra,bloemer}@uni-paderborn.de

‡Department of Electrical Engineering/Computer Science, University of Kassel, scholz@cs.uni-kassel.de

ISSN 1433-8092

	$k \geq 2$ *part of input,* d *constant*	$d \geq 2$ *part of input,* k *constant*
Euclidean k*-center*	$\mathcal{NP}$-hard for all $d \geq 2$ [19], no 1.82-approx. in poly(n,k) [36, 16], 2-approx. in poly(n,k) [21, 22, 16]	$\mathcal{NP}$-hard for all $k \geq 2$ [34] $(1+\epsilon)$-approx. in poly(n,d) [6, 5]
Euclidean k*-diameter*	$\mathcal{NP}$-hard for all $d \geq 2$ [21], no 1.96-approx. in poly(n,k) [16], 2-approx. in poly(n,k) [21, 23, 16]	$\mathcal{NP}$-hard for all $k \geq 3$ [10] 2-approx. in poly(n,d) [21, 23, 16]
Euclidean k*-means*	$\mathcal{NP}$-hard for all $d \geq 2$ [31, 43], $(9+\epsilon)$-approx. in poly(n,k) [27]	$\mathcal{NP}$-hard for all $k \geq 2$ [10, 15, 14, 26, 3] $(1+\epsilon)$-approx. in poly(n,d) [18, 29, 17, 12]
Kullback-Leibler k*-center*	$\mathcal{NP}$-hard for all $d \geq 3$ [11], no 3.32-approx. in poly(n,k) [11] $\mathcal{O}(\log n)$-approx. in poly(n,k) [11]	$\mathcal{O}(\max\{\log n, \log d\})$-approx. in poly$(n,d)$ [11]

Table 1: Overview of known hardness and (non-)approximability results.

large number of applications where k-clustering problems with respect to a non-metric dissimilarity measure are considered. For instance, in the spectral analysis of speech signals, k-median clustering by Itakura-Saito divergence is used to quantize speech signals [25], and in image retrieval, k-center clustering with respect to the Kullback-Leibler divergence (relative entropy) is used when indexing data bases [40]. Both of these dissimilarity measures are neither a metric, nor a symmetric distance function. These examples are instances of a broader class of dissimilarity measures that also includes well known symmetric distance functions such as the squared Euclidean distance and the Mahalanobis distances: the class of *Bregman divergences*.

Clustering with Bregman divergences is a problem that arises in many different disciplines of computer science, such as machine learning, data compression, data mining, speech processing, image analysis, or pattern recognition. However, the theoretical study and analysis of general Bregman k-clustering problems has only recently attained considerable attention in the theoretical computer science and machine learning community. A number of results have been achieved; in particular, Lloyds famous k-means heuristic [30] has been proven to be applicable to all Bregman k-median problems [7]. Furthermore, the use of adaptive seeding (i.e., kmeans++ seeding [4]) has been adopted to improve the performance of Lloyd's heuristic for Bregman divergences [38, 41, 1]. Additionally, a fast approximation algorithm for the Bregman 1-center problem has been developed [39]. The notion of Bregman Voronoi diagrams has been studied and it was shown how to compute them efficiently [37]. Recently, a first PTAS (for constant k) applicable to Bregman k-median problems has been given [2]. Also, the concept of coresets has been adopted to the Bregman k-median problem [1].

On the other hand, the computational hardness of these Bregman k-clustering problems has always been assumed. Yet, with the exception of the Euclidean case and a recent result considering the Kullback-Leibler divergence [11], no proofs have been known. In this paper, we resolve this open problem by showing hardness and non-approximability results for the Bregman k-center problem, the Bregman k-median problem, and the Bregman k-diameter problem when the number of clusters k is part of the input.

1.1 The k-center, the k-median, and the k-diameter problem

In this section, we introduce the formulation of the three k-clustering problems which we study in this paper. To this end, let $\mathbb{X} \subseteq \mathbb{R}^d$ denote an arbitrary ground set of possible data points. Furthermore, let $\mathrm{D}(\cdot,\cdot)$ denote an arbitrary dissimilarity measure on $\mathbb{X}$. Throughout this paper, let n denote the size of the input point set we wish to cluster.

In the *k-center problem* with respect to D, the objective is to minimize the maximum dissimilarity of any input point towards the center point of its cluster. That is, for finite point set $P \subseteq \mathbb{X}$ and any finite set of k

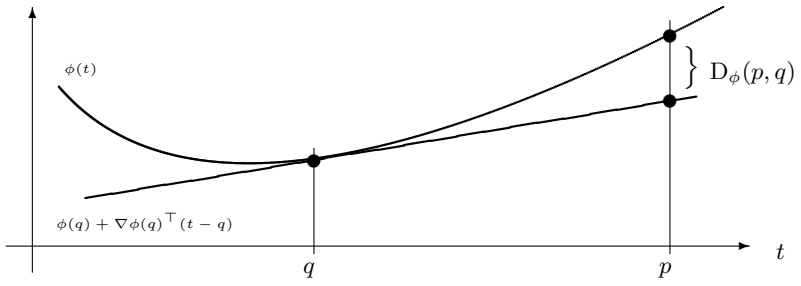

Figure 1: Geometric interpretation of a Bregman divergence.

centers $C \subseteq \mathbb{X}$, let

$$\mathrm{rad}_{\mathrm{D}}(P, C) = \max_{p \in P} \min_{c \in C} \mathrm{D}(p, c) \tag{1}$$

denote the k-center radius of P using center points C. Given a finite $P \subseteq \mathbb{R}^d$, the goal of the k-center problem is to find a set $C \subseteq \mathbb{X}$ of size k that minimizes $\mathrm{rad}_{\mathrm{D}}(P, C)$. We denote the radius of such an optimal solution by $\mathrm{opt}^{\mathrm{rad}}_{k,\mathrm{D}}(P)$.

In the *k-median problem* with respect to D, the objective is to minimize the average dissimilarity of any input point (or, equivalently, the total dissimilarity of all input points) towards the center point of its cluster. More precisely, for finite input set $P \subseteq \mathbb{X}$ and any finite set of k centers $C \subseteq \mathbb{X}$, let

$$\mathrm{cost}_{\mathrm{D}}(P, C) = \sum_{p \in P} \min_{c \in C} \mathrm{D}(p, c) \tag{2}$$

denote the k-median cost of P using center points C. Given a finite $P \subseteq \mathbb{R}^d$, the goal of the k-median problem is to find a set $C \subseteq \mathbb{X}$ of size k that minimizes $\mathrm{cost}_{\mathrm{D}}(P, C)$. We denote the cost of an optimal solution by $\mathrm{opt}^{\mathrm{cost}}_{k,\mathrm{D}}(P)$. Note that if dissimilarity measure D is the squared Euclidean distance, this problem is widely known as the Euclidean k-means problem.

In the *k-diameter problem*, the objective is to minimize the maximum dissimilarity between pairs of points from the same cluster. That is, let $P \subseteq \mathbb{X}$ be a finite point set and let $\mathfrak{P} = \{P_1, P_2, \ldots, P_k\}$ be a partition of P into k non-empty sets. Then, the k-diameter of P using partition $\mathfrak{P}$ is defined a

$$\mathrm{diam}_{\mathrm{D}}(\mathfrak{P}) = \max_{i=1,\ldots,k} \max_{p,q \in P_i} \mathrm{D}(p, q). \tag{3}$$

Given $P \subseteq \mathbb{R}^d$, the goal of the k-diameter problem is to find a partition $\mathfrak{P}$ of P into k subsets that minimizes $\mathrm{diam}_{\mathrm{D}}(\mathfrak{P})$. We denote the diameter of an optimal solution by $\mathrm{opt}^{\mathrm{diam}}_{k,\mathrm{D}}(P)$.

1.2 Bregman clustering

Throughout this paper, we consider an arbitrary but fixed dissimilarity measures that belongs to the class of Bregman divergences. In this case, we refer to the problems stated above as the Bregman k-center problem, the Bregman k-median problem, and the Bregman k-diameter problem, respectively. The dissimilarity measures known as Bregman divergences were introduced in 1967 by Lev M. Bregman [9]. Intuitively, a Bregman divergence can be seen as the error when approximating a strictly convex function by a tangent hyperplane (see Figure 1). Formally, let $\mathbb{X} \subseteq \mathbb{R}^d$ be a convex set and let $\mathrm{ri}(\mathbb{X})$ denote the relative interior of $\mathbb{X}$. For any strictly convex, differentiable function $\phi : \mathrm{ri}(\mathbb{X}) \to \mathbb{R}$ we define the *Bregman divergence* with respect to generating function ϕ as

$$D_\phi(p, q) = \phi(p) - \phi(q) - \nabla\phi(q)^\top (p - q) \tag{4}$$

for $p, q \in \mathrm{ri}(\mathbb{X})$. Here, $\nabla\phi(q)$ denotes the gradient column vector of ϕ at point q.

Bregman divergences include many prominent dissimilarity measures like the squared Euclidean distance $\mathrm{D}_{\ell_2^2}(p,q) = \|p-q\|_2^2$ (with $\phi_{\ell_2^2}(t) = \|t\|_2^2$ on $\mathbb{X}_{\ell_2^2} = \mathbb{R}^d$), the Kullback-Leibler divergence $\mathrm{D}_{\mathrm{KL}}(p,q) = \sum p_i \ln \frac{p_i}{q_i} - p_i + q_i$ (with $\phi_{\mathrm{KL}}(t) = \sum t_i \ln t_i - t_i$ on $\mathbb{X}_{\mathrm{KL}} = \mathbb{R}^d_{\geq 0}$), and the Itakura-Saito divergence $\mathrm{D}_{\mathrm{IS}}(p,q) = \sum \frac{p_i}{q_i} - \ln \frac{p_i}{q_i} - 1$ (with $\phi_{\mathrm{IS}}(t) = -\sum \ln t_i$ on $\mathbb{X}_{\mathrm{IS}} = \mathbb{R}^d_{\geq 0}$). We point out that, in general, Bregman divergences are asymmetric and do not satisfy the triangle inequality. Furthermore, D_ϕ may possess singularities, i.e., there may exist points $p, q \in \mathbb{X}$ such that $\mathrm{D}_\phi(p,q) = \infty$.

Clustering with Bregman divergences is a problem that arises in many different disciplines of computer science, such as machine learning, data compression, data mining, speech processing, image analysis, or pattern recognition. In fact, our work started with an industrial project on lossless compression of Java and C++ executable code on smartcards. This project included the design of a number of codebooks for a given instruction set in such a way that for an arbitrary executable source code file at least one these codebooks provides good compression. For any possible source code file, its optimal codebook is uniquely determined by the probability distribution given by the relative frequencies of the instructions from the instruction set. However, due to the strict space and memory limitations of a smartcard, only a small number of, say, k prototypical codebooks can actually be realized on a smartcard chip. These prototypical codebooks can be learned from a training set of n source code samples as follows. First, compute the optimal codebook and, hence, the corresponding probability distribution of each source code sample. Then, find the k prototypical codebooks that minimize the loss of compression when using the best fitting prototype instead of the optimal codebook. It is known from coding theory that this loss of compression is well approximated by the Kullback-Leibler divergence between the original and the prototypical probability distribution [13]. Hence, learning these prototypes immediately leads to a k-median clustering problem involving the Kullback-Leibler divergence.

1.3 Related work

It is known that the Euclidean k-center (k-diameter, k-means) problem can be solved in polynomial time if the dimension d and the number of clusters k are both constant [24]. Hence, there are usually two branches of hardness results that can be studied: First, the case when the number of clusters k is part of the input. In this case, the dimension d of the Euclidean space $\mathbb{R}^d$ may be fixed. Or, second, the case when k is a constant that is fixed for all input instances. In this case, the dimension d is part of the input and potentially unbounded. Hardness and and non-approximability results for these two settings are summarized in Table 1 on page 2.

In this paper, we concentrate on the case when k is part of the input. In this case, hardness and non-approximability results for the Euclidean k-clustering problems have been known for some time. In particular, Feder and Greene [16] gave a gap-introducing reduction from the degree-bounded planar vertex cover problem to show the non-approximability of the Euclidean k-center and the Euclidean k-diameter problem. Stated in detail, they show that the Euclidean k-center problem (and the Euclidean k-diameter problem) in $\mathbb{R}^d$ with $d \geq 2$ can not be approximated within factor $\alpha < 1.82$ (and $\alpha < 1.96$, respectively) in time polynomial in n and k, unless $\mathcal{P} = \mathcal{NP}$. To the best of our knowledge, these are still the best bounds on the non-approximability of the Euclidean k-center and the Euclidean k-diameter problem that are currently known. Also, note that there exist approximation algorithms with an approximation guarantee of $\alpha = 2$ and a running time polynomial in n and k for both the Euclidean k-center problem and the Euclidean k-diameter problem [21, 22, 23, 16]. Furthermore, it is an immediate consequence of [16] that the k-center and k-diameter problem with respect to the squared Euclidean distance can not be approximated within factor $\alpha < 1.82^2 \approx 3.32$ and factor $\alpha < 1.96^2 \approx 3.87$, respectively.

If k is part of the input, the Euclidean k-means problem (i.e., the k-median problem with respect to the squared Euclidean distance) in $\mathbb{R}^d$ with $d \geq 2$ is known to be $\mathcal{NP}$-hard [31, 43]. Although no non-approximability result is known, the best $\mathrm{poly}(n,k)$-time approximation algorithm currently known for the Euclidean k-means problem has an approximation guarantee of merely $9 + \epsilon$ [27]. It is an open question whether a better approximation can be achieved in time polynomial in n and k.

Recently, Chaudhuri and McGregor gave a gap-preserving reduction from the Euclidean k-center problem

to the Kullback-Leibler k-center problem [11]. Using this reduction and [16], they are able to show the non-approximability of the Kullback-Leibler k-center problem. The same reduction can be used to prove the non-approximability of the Kullback-Leibler k-diameter problem, although the authors did not consider this problem explicitly.

However, prior to our work, the special cases of the squared Euclidean distance and the Kullback-Leibler divergence were the only hardness results known for Bregman k-clustering problems.

A technique that has been used recently to obtain approximation algorithms for arbitrary Bregman divergences and that will also play an important role in our contribution is the technique of Mahalanobis-similarity [2, 1]. Generally speaking, the main observation of Mahalanobis-similarity is that all Bregman divergences behave locally just like a Mahalanobis divergence, up to some small multiplicative error. Here, the Mahalanobis distances are exactly the subclass of the Bregman divergences that are the square of a metric [37]. Mahalanobis-similarity has also been used implicitly in a number of different publications (e.g., [41, 8, 33]).

1.4 Our contribution

In all the recent work on Bregman clustering, the computational hardness of the Bregman k-clustering problems as given in Section 1.1 has always been assumed. However, beside the special case of the squared Euclidean distance and the recent result of Chaudhuri and McGregor considering the Kullback-Leibler divergence [11], no proof of hardness considering an arbitrary Bregman divergence has been known. We resolve this open problem in the case that the number of clusters k is part of the input. We prove that finding an optimal solution of the Bregman k-center problem, the Bregman k-median problem, and the Bregman k-diameter problem is $\mathcal{NP}$-hard. Furthermore, we show that it is even $\mathcal{NP}$-hard to approximate the Bregman k-center problem within a factor of 3.32, and the Bregman k-diameter problem within a factor of 3.87.

To achieve our result, we have to make two mild assumptions on Bregman divergence D_ϕ. First, we call a Bregman divergence D_ϕ *smooth* if its generating function ϕ is twice differentiable on $\mathrm{ri}(\mathbb{X})$ with continuous second-order partial derivatives. This assumption is not very strong since essentially all Bregman divergences that are used in practice are smooth. Second, we call a Bregman divergence D_ϕ *trivial* if its domain $\mathbb{X} \subseteq \mathbb{R}^d$ is completely contained in a 1-dimensional affine subspace of $\mathbb{R}^d$ (i.e, $\mathbb{X}$ is contained in a straight line). Otherwise, we call D_ϕ non-trivial. In the trivial case, there exists a simple $\mathrm{poly}(n, d, k)$-time algorithm that solves k-clustering problems optimally (see Section 4 for a discussion of this algorithm). Hence, the restriction to non-trivial Bregman divergences is a necessary assumption to show any hardness result.

Using the technique of Mahalanobis-similarity, we are able to generalize the approach of Chaudhuri and McGregor to all smooth and non-trivial Bregman divergences. Stated in detail, we prove the following theorems.

Theorem 1. *Let D_ϕ be a smooth, non-trivial Bregman divergence on domain $\mathbb{X} \subseteq \mathbb{R}^d$ and let $\alpha < 3.32$. There exists no α-approximation algorithm for the k-center problem with respect to D_ϕ with a running time polynomial in n and k, unless $\mathcal{P} = \mathcal{NP}$.*

Theorem 2. *Let D_ϕ be a smooth, non-trivial Bregman divergence on domain $\mathbb{X} \subseteq \mathbb{R}^d$ and let $\alpha < 3.87$. There exists no α-approximation algorithm for the k-diameter problem with respect to D_ϕ with a running time polynomial in n and k, unless $\mathcal{P} = \mathcal{NP}$.*

In the case of the k-median objective function, it turns out that we need a stronger reduction than the one suggested by Chaudhuri and McGregor. Instead, we use the Mahalanobis similarity to adopt a reduction that has been given recently by Vattani [43] to show the $\mathcal{NP}$-hardness of the Euclidean k-means problem. To this end, we have to strengthen our assumption of smoothness to what we call a *computationally smooth* Bregman divergence. We will state this notion in detail later in Section 3.3. Again, to the best of our knowledge, all Bregman divergences that are used in practice (such as the Kullback-Leibler divergence and the Itakura-Saito divergence) are computationally smooth.

Theorem 3. *Let D_ϕ be a computationally smooth, non-trivial Bregman divergence on domain $\mathbb{X} \subseteq \mathbb{R}^d$. There exists no algorithm solving the k-median problem with respect to D_ϕ optimally with a running time polynomial in n and k, unless $\mathcal{P} = \mathcal{NP}$.*

1.5 Organization

The rest of this paper is organized as follows. As a first step towards our results, in Section 2, we prove hardness and non-approximability results for Mahalanobis k-clustering problems. These results are an immediate consequence of the close relationship between Mahalanobis distances and the squared Euclidean distance. After that, in Section 3, we give a generalization of the reduction of Chaudhuri and McGregor [11] to prove Theorem 1 and Theorem 2. In addition to that, we combine the technique of Mahalanobis-similarity with a reduction given by Vattani [43] to show Theorem 3. In Section 4, we briefly discuss the simple polynomial time algorithm for trivial Bregman k-clustering problems. We end this paper with a discussion of open problems in Section 5.

2 Hardness of Mahalanobis k-clustering problems

Among the Bregman divergences, one particular subclass of dissimilarity measures plays an important role in our approach. For a symmetric positive definite matrix $A \in \mathbb{R}^{d\times d}$, the Mahalanobis distance with respect to A is defined as

$$\mathrm{D}_A(p,q) = (p-q)^\top A\,(p-q) \tag{5}$$

for all $p, q \in \mathbb{R}^d$. Mahalanobis distances were introduced in 1936 by P. C. Mahalanobis based on the inverse of the covariance matrix of two random variables [32]. It is easy to see that all Mahalanobis distances are Bregman divergences by means of the generating function $\phi_A(t) = t^\top A\,t$ on $\mathbb{R}^d$. In many ways, the class of Mahalanobis distances is a generalization of the squared Euclidean distance. This close relationship is formalized in the following lemma.

Lemma 4. *Let D_A be a Mahalanobis distance with respect to a symmetric positive definite matrix $A \in \mathbb{R}^{d\times d}$. There exists a non-singular matrix $U \in \mathbb{R}^{d\times d}$ such that for all $p, q \in \mathbb{R}^d$ we have*

$$\mathrm{D}_A(p,q) = \|Up - Uq\|_2^2. \tag{6}$$

Proof. Since A is a symmetric positive definite matrix, it is a well-known fact from linear algebra that there exists a non-singular matrix U with $A = U^\top U$. I.e., such a matrix U is given by the Cholesky decomposition of matrix A [42]. Hence, we obtain

$$\mathrm{D}_A(p,q) = (p-q)^\top U^\top U\,(p-q) = (Up-Uq)^\top (Up-Uq) = \|Up-Uq\|_2^2\ . \tag{7}$$

□

That is, a Mahalanobis distance is merely a squared Euclidean distance in a linearly transformed point space, where the linear transformation is given by the non-singular matrix U from Lemma 4.

2.1 $\mathcal{NP}$-hardness and non-approximability

Using Lemma 4, it is easy to show how the Euclidean k-center (k-diameter, k-means) problem can be reduced to a given Mahalanobis k-center (k-diameter, k-median) problem. We immediately obtain the following hardness results from the known results with respect to the squared Euclidean distance [16, 31, 43].

Corollary 5. *Let D_A be an arbitrary Mahalanobis distance on $\mathbb{R}^d$ with $d \geq 2$, and let $\alpha < 3.32$. Then there exists no α-approximation algorithm for the k-center problem with respect to D_A with a running time polynomial in n and k, unless $\mathcal{P} = \mathcal{NP}$.*

Corollary 6. *Let D_A be an arbitrary Mahalanobis distance on $\mathbb{R}^d$ with $d \geq 2$, and let $\alpha < 3.87$. Then there exists no α-approximation algorithm for the k-diameter problem with respect to D_A with a running time polynomial in n and k, unless $\mathcal{P} = \mathcal{NP}$.*

Corollary 7. *Let D_A be an arbitrary Mahalanobis distance on $\mathbb{R}^d$ with $d \geq 2$. Then there is no algorithm solving the k-median problem with respect to D_A optimally with a running time polynomial in n and k, unless $\mathcal{P} = \mathcal{NP}$.*

2.2 Mahalanobis k-clustering in bounded regions of $\mathbb{R}^d$

In addition to the results given above, we also find that the computational hardness of the Mahalanobis k-clustering problems remains unaltered if we restrict the problem to input points from some bounded (full-dimensional) region of $\mathbb{R}^d$. Let $B_r^d(z) \subseteq \mathbb{R}^d$ denote the Euclidean ball of radius $r > 0$ centered at point $z \in \mathbb{R}^d$, that is,

$$B_r^d(z) = \{x \in \mathbb{R}^d \,|\, \|x - z\|_2 \leq r\}. \tag{8}$$

In the lemma below, we show the computational hardness of Mahalanobis k-clustering problems restricted to input points from the unit ball $B_1^d(0)$. We will make use of Lemma 8 in the reduction in Section 3. Please note that the same result as given below applies for restrictions to any (full-dimensional) region of $\mathbb{R}^d$.

Lemma 8. *Let D_A be an arbitrary Mahalanobis distance on $\mathbb{R}^d$. If the Mahalanobis k-center (k-diameter, k-median) problem with respect to D_A and input domain $B_1^d(0)$ can be approximated within factor α in time polynomial in n and k, then the Mahalanobis k-center (k-diameter, k-median) problem with respect to D_A and input domain $\mathbb{R}^d$ can also be approximated within factor α in time in polynomial n and k.*

Proof. Assume that there exists an algorithm that computes an α-approximate solution to the Mahalanobis k-center (k-diameter, k-median) problem for arbitrary input sets from $B_1^d(0)$. Then we obtain an approximate solution to the Mahalanobis k-center (k-diameter, k-median) problem with respect to an arbitrary input instance $P \subseteq \mathbb{R}^2$ the following way. Let $\Delta = \max\{\|p\|_2 \,|\, p \in P\}$. Compute the scaled input set $\Delta^{-1}P = \{\Delta^{-1}p \,|\, p \in P\}$ and solve the k-center (k-diameter, k-median) problem for input instance $\Delta^{-1}P$.

Obviously, $\Delta^{-1}P \subseteq B_1^d(0)$. Note that by scaling the instance by factor Δ^{-1} the description size of the instance is increased by at most a constant. Also, we find that scaling does only change the numerical values of a clustering solution and not the clustering problem itself. That is, for all $C \subseteq P$ and all partitions $\mathfrak{P}$ of P we have

$$\begin{aligned}
\mathrm{rad}_{\mathrm{D}_A}(P, C) &= \Delta^2 \max_{p \in P} \min_{c \in C} (\Delta^{-1}p - \Delta^{-1}c)^\top A\, (\Delta^{-1}p - \Delta^{-1}c) \\
&= \Delta^2\, \mathrm{rad}_{\mathrm{D}_A}(\Delta^{-1}P, \Delta^{-1}C)\ , && (9) \\
\mathrm{diam}_{\mathrm{D}_A}(\mathfrak{P}) &= \Delta^2 \max_{i=1,\ldots,k} \max_{p,q \in P_i} (\Delta^{-1}p - \Delta^{-1}c)^\top A\, (\Delta^{-1}p - \Delta^{-1}c) \\
&= \Delta^2\, \mathrm{diam}_{\mathrm{D}_A}(\Delta^{-1}\mathfrak{P})\ , && (10) \\
\mathrm{cost}_{\mathrm{D}_A}(P, C) &= \Delta^2 \sum_{p \in P} \min_{c \in C} (\Delta^{-1}p - \Delta^{-1}c)^\top A\, (\Delta^{-1}p - \Delta^{-1}c) \\
&= \Delta^2\, \mathrm{cost}_{\mathrm{D}_A}(\Delta^{-1}P, \Delta^{-1}C)\ , && (11)
\end{aligned}$$

where $\Delta^{-1}\mathfrak{P} = \{\Delta^{-1}S \,|\, S \in \mathfrak{P}\}$. Hence, the lemma follows. □

3 Hardness of Bregman k-clustering problems

In this section, we prove the hardness and the non-approximability of Bregman k-clustering problems. We achieve our results by giving a polynomial time reduction from a Mahalanobis problem to the Bregman problem. This reduction is constructed in such a way that the Bregman k-center radius (k-diameter, k-median cost) of the reduced instance equals the Mahalanobis k-center radius (k-diameter, k-median cost) of the original instance within small multiplicative error.

3.1 Local Mahalanobis-similarity

We start by showing that the domain $\mathbb{X}$ of any smooth Bregman divergence D_ϕ contains a small region B such that we find that D_ϕ on B behaves just like a certain Mahalanobis distance, up to a small multiplicative error. To this end, we make the following elementary yet crucial observation: Since $\mathrm{D}_\phi(p,q)$ equals the remainder term of the first-order Taylor expansion of $\phi(p)$ at point q, the Bregman divergence D_ϕ can be expressed in terms of the Hessian matrix of ϕ.

Lemma 9. *Let D_ϕ be a smooth Bregman divergence on domain $\mathbb{X}$. Then for all $p, q \in \mathrm{ri}(\mathbb{X})$ there exists a point t on the line segment through p and q such that*

$$\mathrm{D}_\phi(p,q) = \frac{1}{2}(p-q)^\top \nabla^2\phi(t)\,(p-q)\ . \tag{12}$$

Here $\nabla^2\phi(t)$ denotes the Hessian matrix of ϕ at point t.

Proof. Since D_ϕ is smooth, the second-order partial derivatives of ϕ exist. Hence, consider the first-order Taylor expansion of $\phi(p)$ at point q, that is,

$$\phi(p) = \phi(q) + \nabla\phi(q)(p-q) + R_1(p)\ , \tag{13}$$

where $R_1(p)$ denotes the remainder term of the first-order Taylor expansion. Using the Lagrange form of the remainder term we obtain that there exists a point t on the line segment through p and q such that

$$\mathrm{D}_\phi(p,q) = R_1(p) = \frac{1}{2}(p-q)^\top \nabla^2\phi(t)\,(p-q)\ . \tag{14}$$

□

Lemma 9 enables us to show the local similarity of any Bregman divergence towards a Mahalanobis distance. Intuitively, this similarity follows from the fact that the right-hand side of (12) resembles the definition of a Mahalanobis distance. In fact, if the Hessian $\nabla^2\phi(t)$ is constant for all $t \in \mathrm{ri}(\mathbb{X})$, we obtain that D_ϕ is indeed a Mahalanobis distance. Also note that since the second partial derivatives of ϕ are continuous, we know that all Hessians $\nabla^2\phi(t)$ are symmetric.

In the sequel, let $z \in \mathrm{ri}(\mathbb{X})$ and $A = \nabla^2\phi(z)$ be such that A is positive definite, i.e., for all $x \in \mathbb{R}^d \setminus \{0\}$ we have $x^\top A\,x > 0$. Such points $z \in \mathrm{ri}(\mathbb{X})$ exist in abundance for the following reason. Since the generating function ϕ is strictly convex, we know that its Hessian matrix $\nabla^2\phi(t)$ is positive definite at almost all points t from $\mathrm{ri}(\mathbb{X})$, with the exception of a merely nowhere dense subset $Y \subsetneq \mathrm{ri}(\mathbb{X})$ (if such a subset Y exists at all). Hence, any point $z \in \mathrm{ri}(\mathbb{X}) \setminus Y$ will do.

Using these definitions and the smoothness of D_ϕ, we are able to show that there exists a small region $B^d_\delta(z)$ around center point z such that for all $t \in B^d_\delta(z) \cap \mathrm{ri}(\mathbb{X})$ the quadratic form $x^\top\nabla^2\phi(t)\,x$ is well approximated by $x^\top A\,x$.

Lemma 10. *Let z, A be as given above. For all $\epsilon > 0$ there exists a $\delta > 0$ and a δ-ball $B^d_\delta(z)$ centered at z such that for all $x \in \mathbb{R}^d$ and for all $t \in B^d_\delta(z) \cap \mathrm{ri}(\mathbb{X})$ we have*

$$(1-\epsilon)\,x^\top A\,x \le x^\top\nabla^2\phi(t)\,x \le (1+\epsilon)\,x^\top A\,x\ . \tag{15}$$

Proof. Let $x = (x_1, x_2, \ldots, x_d)^\top$. Note that for $x = 0$ the claim is trivially true. Hence, in the following, we may assume $x \in \mathbb{R}^d \setminus \{0\}$.

Recall that A is a symmetric positive definite matrix. Hence, all eigenvalues of A are positive reals. Let $\lambda_{\min} > 0$ denote the smallest eigenvalues of A. It is known from linear algebra that for all $x \in \mathbb{R}^d \setminus \{0\}$ the Rayleigh quotient of A and x is bounded from below by the smallest eigenvalue of A [42]. That is, we find

$$\frac{x^\top A\,x}{\|x\|_2^2} \ge \lambda_{\min}\ . \tag{16}$$

In the sequel, let $\nu(x) = 1/\|x\|_2^2$. Note that $\nu(x)$ is continuous in all points $x \in \mathbb{R}^d \setminus \{0\}$. Furthermore, for $t \in \mathrm{ri}(\mathbb{X})$ let $\phi_{ij}(t) = \frac{\partial^2}{\partial t_i \partial t_j}\phi(t)$ denote the second partial derivatives of ϕ. Since we assume Bregman divergence D_ϕ to be smooth, we know that for all i, j the functions $\phi_{ij}(t)$ are continuous on $\mathrm{ri}(\mathbb{X})$. Using the notation given above, it is easy to see that

$$h(t,x) = \frac{x^\top \nabla^2\phi(t)\, x}{\lambda_{\min}\|x\|_2^2} = \frac{1}{\lambda_{\min}}\,\nu(x)\sum_{i,j} x_i x_j \phi_{ij}(t) \tag{17}$$

is continuous for all $(t,x) \in \mathrm{ri}(\mathbb{X}) \times \big(\mathbb{R}^d \setminus \{0\}\big)$. Hence, for all $\epsilon > 0$ there exists a $\delta > 0$ and a Euclidean ball $B_\delta^d(z)$ of radius δ centered at z such that

$$\frac{1}{\lambda_{\min}\|x\|_2^2} \cdot |x^\top A\, x - x^\top \nabla^2\phi(t)\, x| = |h(z,x) - h(t,x)| \le \epsilon \tag{18}$$

for all $x \in \mathbb{R}^d \setminus \{0\}$ and for all $t \in B_\delta^d(z) \cap \mathrm{ri}(\mathbb{X})$. Using Inequality (16), we obtain

$$|x^\top A\, x - x^\top \nabla^2\phi(t)\, x| \le \epsilon\lambda_{\min}\|x\|_2^2 \le \epsilon\, x^\top A\, x\ . \tag{19}$$

Thus, the lemma follows. □

Using Lemma 9 and Lemma 10, we conclude that for all points from $B_\delta^d(z) \cap \mathrm{ri}(\mathbb{X})$, Bregman divergence D_ϕ is a $(1 \pm \epsilon)$-approximation of Mahalanobis distance $\mathrm{D}_{\frac{1}{2}A}$.

Lemma 11. *For ϵ, δ, z, A as given in Lemma 10 and for all $p, q \in B_\delta^d(z) \cap \mathrm{ri}(\mathbb{X})$ we have*

$$(1-\epsilon)\,\mathrm{D}_{\frac{1}{2}A}(p,q) \le \mathrm{D}_\phi(p,q) \le (1+\epsilon)\,\mathrm{D}_{\frac{1}{2}A}(p,q)\ . \tag{20}$$

Proof. Let $p, q \in B_\delta^d(z) \cap \mathrm{ri}(\mathbb{X})$. From Lemma 9 we know that there is a $\xi \in B_\delta^d(z) \cap \mathrm{ri}(\mathbb{X})$ such that $\mathrm{D}_\phi(p,q) = \frac{1}{2}(p-q)^\top \nabla^2\phi(\xi)\,(p-q)$. Hence, using Lemma 10 with $x = p - q$, we obtain

$$\mathrm{D}_\phi(p,q) = \frac{1}{2}(p-q)^\top \nabla^2\phi(\xi)\,(p-q) \le \frac{1+\epsilon}{2}(p-q)^\top A\,(p-q) = (1+\epsilon)\,\mathrm{D}_{\frac{1}{2}A}(p,q) \tag{21}$$

and

$$\mathrm{D}_\phi(p,q) = \frac{1}{2}(p-q)^\top \nabla^2\phi(\xi)\,(p-q) \ge \frac{1-\epsilon}{2}(p-q)^\top A\,(p-q) = (1-\epsilon)\,\mathrm{D}_{\frac{1}{2}A}(p,q)\ . \tag{22}$$

□

Lemma 11 can be used to relate the k-center radius (k-diameter, k-median cost) in terms of Bregman divergence D_ϕ to the k-center radius (k-diameter, k-median cost) of the same instance in terms of Mahalanobis distance $\mathrm{D}_{\frac{1}{2}A}$. We obtain the following corollary.

Corollary 12. *Let ϵ, δ, z, A be as given in Lemma 10. For all $P, C \subseteq B_\delta^d(z) \cap \mathrm{ri}(\mathbb{X})$ and for all partitions $\mathfrak{P}$ of P we have*

$$(1-\epsilon)\,\mathrm{rad}_{\mathrm{D}_{\frac{1}{2}A}}(P,C) \le \mathrm{rad}_{\mathrm{D}_\phi}(P,C) \le (1+\epsilon)\,\mathrm{rad}_{\mathrm{D}_{\frac{1}{2}A}}(P,C)\ , \tag{23}$$

$$(1-\epsilon)\,\mathrm{diam}_{\mathrm{D}_{\frac{1}{2}A}}(\mathfrak{P}) \le \mathrm{diam}_{\mathrm{D}_\phi}(\mathfrak{P}) \le (1+\epsilon)\,\mathrm{diam}_{\mathrm{D}_{\frac{1}{2}A}}(\mathfrak{P})\ , \tag{24}$$

$$(1-\epsilon)\,\mathrm{cost}_{\mathrm{D}_{\frac{1}{2}A}}(P,C) \le \mathrm{cost}\,\mathrm{D}_\phi(P,C) \le (1+\epsilon)\,\mathrm{cost}_{\mathrm{D}_{\frac{1}{2}A}}(P,C)\ . \tag{25}$$

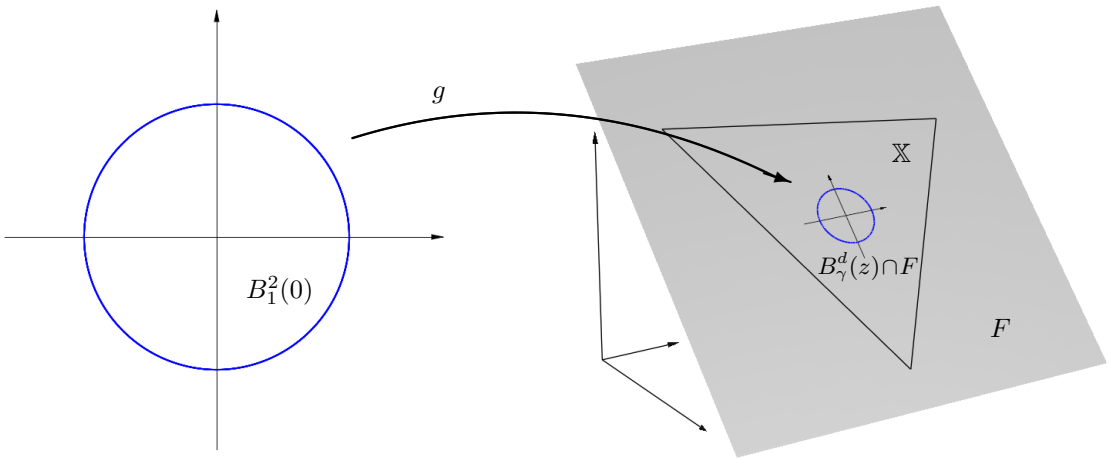

Figure 2: Sketch of mapping g.

3.2 Construction of a reduction function

In the following, we give the construction of a polynomial time computable reduction from a certain 2-dimensional Mahalanobis k-center (k-diameter, k-median) problem to the Bregman k-center (k-diameter, k-median) problem with respect to D_ϕ. This reduction is achieved by embedding a Mahalanobis k-clustering input instance $P \subseteq \mathbb{R}^2$ into a small region $B \subseteq \mathrm{ri}(\mathbb{X})$, where the local Mahalanobis-similarity results from Section 3.1 apply. This embedding is constructed in such a way that the Bregman k-center radius (k-diameter, k-median cost) of the embedded instance corresponds to the Mahalanobis k-center radius (k-diameter, k-median cost) of the original instance P within a multiplicative error of $1 \pm \epsilon$.

To this end, recall that we assume Bregman divergence D_ϕ to be non-trivial, i.e., its domain $\mathbb{X} \subseteq \mathbb{R}^d$ is not contained in a straight line. Thus, there exists a small 2-dimensional disc in $\mathbb{R}^d$ that is completely contained in $\mathrm{ri}(\mathbb{X})$. That is, there exist a 2-dimensional affine subspace $F \subseteq \mathbb{R}^d$, a constant $\beta > 0$, and a center point $z \in \mathrm{ri}(\mathbb{X}) \cap F$ such that $B_\beta^d(z) \cap F \subseteq \mathrm{ri}(\mathbb{X})$. Let $A = \nabla^2\phi(z)$. As has been discussed in Section 3.1, without loss of generality, we may assume that z is chosen in such a way that A is symmetric positive definite. Hence, for all $\epsilon > 0$ there exists a $\delta > 0$ such that the results from Section 3.1 apply for all points in $B_\delta^d(z) \cap \mathrm{ri}(\mathbb{X})$. Also note that once D_ϕ and ϵ are fixed, δ is a fixed constant. Furthermore, let $G \in \mathbb{R}^{d\times 2}$ be a matrix whose column vectors form an orthonormal basis of the 2-dimensional subspace $F - z = \{x - z \,|\, x \in F\}$.

Let $\gamma = \min\{\delta, \beta\} > 0$ be a constant that depends only on D_ϕ and ϵ. Obviously, we have $B_\gamma^d(z) \cap F \subseteq B_\delta^d(z) \cap \mathrm{ri}(\mathbb{X})$. We can define a mapping g between the 2-dimensional unit disc $B_1^2(0)$ and $B_\gamma^d(z) \cap F$ as follows. Let $g : \mathbb{R}^2 \to \mathbb{R}^d$ be given by

$$g(x) = \gamma G x + z \tag{26}$$

for all $x \in \mathbb{R}^2$. See Figure 2 for an illustration of this mapping. Also note that the inverse of g is given by the mapping $g^{-1}(x) = \frac{1}{\gamma} G^\top (x - z)$. Since G is a basis of $F - z$, we have $\gamma G x \in F - z$ for all $x \in \mathbb{R}^2$ and all $\gamma \in \mathbb{R}$. Hence, $g(x) = \gamma G x + z \in F$. Using the fact that the orthogonal mapping G is length-preserving, for all $x \in B_1^2(0)$ we find

$$\|g(x) - z\|_2 = \|\gamma G x\|_2 = \gamma \|x\|_2 \leq \gamma \tag{27}$$

and, thus, $g(x) \in B_\gamma^d(z)$. Therefore, $B_1^2(0)$ is indeed mapped into $B_\gamma^d(z) \cap F$.

In the sequel, let

$$A' = \frac{\gamma^2}{2} G^\top A\, G \in \mathbb{R}^{2\times 2}. \tag{28}$$

First, we argue that A' is indeed a symmetric positive definite matrix. Since A is symmetric positive definite we know that $x^\top A\, x > 0$ for all $x \in \mathbb{R}^d \setminus \{0\}$. Let $y \in \mathbb{R}^{d'} \setminus \{0\}$. Since $y \neq 0$ and G is non-singular, we have that $Gy \neq 0$. Hence,

$$y^\top A'\, y = \frac{\gamma^2}{2}\,(Gy)^\top A\,(Gy) > 0 \tag{29}$$

and A' is positive definite. Furthermore, since

$$A'^\top = (\frac{\gamma^2}{2}\,G^\top A\,G)^\top = \frac{\gamma^2}{2}\,G^\top A^\top G \tag{30}$$

we find that if A is symmetric, then so is A'. Hence, the Mahalanobis distance $\mathrm{D}_{A'}$ on domain $\mathbb{R}^2$ is well defined.

It is easy to show that using the mapping g, the Bregman divergence D_ϕ on $B^d_\gamma(z) \cap F$ approximates the 2-dimensional Mahalanobis distance $\mathrm{D}_{A'}$ on the unit disc $B^2_1(0)$.

Lemma 13. *For all $p, q \in B^2_1(0)$ we have*

$$(1-\epsilon)\,\mathrm{D}_{A'}(p,q) \leq \mathrm{D}_\phi\big(g(p), g(q)\big) \leq (1+\epsilon)\,\mathrm{D}_{A'}(p,q) \tag{31}$$

Proof. Let $p, q \in B^2_1(0)$ be arbitrary. We obtain

$$\begin{aligned}
\mathrm{D}_{A'}(p,q) &= \frac{\gamma^2}{2}(p-q)^\top G^\top A\,G\,(p-q) && (32)\\
&= \frac{1}{2}(\gamma Gp - \gamma Gq)^\top A\,(\gamma Gp - \gamma Gq) && (33)\\
&= \frac{1}{2}\big(g(p) - g(q)\big)^\top A\,(g(p) - g(q)) && (34)\\
&= \mathrm{D}_{\frac{1}{2}A}\big(g(p), g(q)\big) && (35)
\end{aligned}$$

Using Lemma 11, we have

$$(1-\epsilon)\,\mathrm{D}_{\frac{1}{2}A}\big(g(p), g(q)\big) \leq \mathrm{D}_\phi\big(g(p), g(q)\big) \leq (1+\epsilon)\,\mathrm{D}_{\frac{1}{2}A}\big(g(p), g(q)\big) \tag{36}$$

and the lemma follows. □

Corollary 14. *Let $P, C \subseteq B^2_1(0)$ and let $\mathfrak{P}$ be a partition of P. Furthermore, let $g(P) = \{g(p) \,|\, p \in P\}$ and $g(\mathfrak{P}) = \{g(S) \,|\, S \in \mathfrak{P}\}$. Then we have*

$$(1-\epsilon)\,\mathrm{rad}_{\mathrm{D}_{A'}}(P,C) \leq \mathrm{rad}_{\mathrm{D}_\phi}\big(g(P), g(C)\big) \leq (1+\epsilon)\,\mathrm{rad}_{\mathrm{D}_{A'}}(P,C)\ , \tag{37}$$

$$(1-\epsilon)\,\mathrm{diam}_{\mathrm{D}_{A'}}(\mathfrak{P}) \leq \mathrm{diam}_{\mathrm{D}_\phi}\big(g(\mathfrak{P})\big) \leq (1+\epsilon)\,\mathrm{diam}_{\mathrm{D}_{A'}}(\mathfrak{P})\ , \tag{38}$$

$$(1-\epsilon)\,\mathrm{cost}_{\mathrm{D}_{A'}}(P,C) \leq \mathrm{cost}\,\mathrm{D}_\phi\big(g(P), g(C)\big) \leq (1+\epsilon)\,\mathrm{cost}_{\mathrm{D}_{A'}}(P,C)\ . \tag{39}$$

Using Corollary 14, an α-approximation algorithm for a Bregman k-clustering problem can be used to approximate a solution to a Mahalanobis k-clustering problem with respect to A'. We obtain the following lemma.

Lemma 15. *Let D_ϕ be a smooth and non-trivial Bregman divergence, let $\epsilon > 0$ be arbitrary, and let A' be as given above. If the Bregman k-center (k-diameter, k-median) problem with respect to D_ϕ on domain $\mathbb{X}$ can be approximated within factor α in time polynomial in n and k, then the Mahalanobis k-center (k-diameter, k-median) problem with respect to $\mathrm{D}_{A'}$ on $\mathbb{R}^2$ can be approximated within factor $(1+\epsilon)\alpha$ in time polynomial in n and k.*

Proof of Lemma 15. Assume that there exists an algorithm that computes an α-approximate solution to the Bregman k-center (k-diameter, k-median) problem with respect to Bregman divergence D_ϕ in polynomial time. From Lemma 8, we know that it is sufficient to consider input instances of the Mahalanobis k-clustering problem from the unit disc $B_1^2(0)$. We obtain an approximate solution to the k-center (k-diameter, k-median) problem with respect to $\mathrm{D}_{A'}$ and arbitrary input instance $P \subseteq B_1^2(0)$ the following way. Apply mapping g to point set P to obtain point set $g(P) \subseteq \mathrm{ri}(\mathbb{X})$. This step can be done in time $\mathcal{O}(dn)$. Then, approximate the Bregman k-clustering problem on input instance $g(P)$ to obtain an α-approximate clustering in time $\mathrm{poly}(n,k)$. In case of center-based clustering (k-center, k-median), transform the approximate center set $C \subseteq \mathrm{ri}(\mathbb{X})$ into $g^{-1}(C) \subseteq B_1^2(0)$ using the inverse mapping of g. This step takes time $\mathcal{O}(dk)$.

According to Corollary 14, the clustering induced on the original input instance P by this clustering algorithm forms a $(1+\epsilon)\alpha$-approximate solution to the Mahalanobis k-clustering problem with respect to $\mathrm{D}_{A'}$. Hence, such an approximation is obtained in time polynomial in n and k. □

3.3 Proof of $\mathcal{NP}$-hardness and non-approximability

The non-approximability of the Bregman k-center and k-diameter problem (Theorems 1 and 2) is an immediate consequence of Lemma 15. This is due to the fact that for these objective functions, Corollaries 5 and 6, respectively, provide a constant non-approximability gap α for the Mahalanobis k-center (k-diameter) problem. Hence, an arbitrarily small, constant distortion of the k-center radius (k-diameter) as introduced by mapping g is negligible.

However, in case of a Bregman k-median problem, no such non-approximability gap is known. Thus, we have to show the hardness by other means. To this end, we make use of a recent reduction from the well known X3C decision problem to the Euclidean k-means problem [43]. Here, the X3C problem is defined as follows. Given a set U of size $|U| = 3n$ and a family $\mathcal{S} = \{S_i\}_{i=1,\dots,l}$ of subsets $S_i \subseteq U$ of size $|S_i| = 3$. Decide whether there is an index set $I \subseteq \{1,\dots,l\}$ such that $\{S_i\}_{i\in I}$ is an *exact covering* of U, that is, $U = \bigcup_{i\in I} S_i$ and $|I| = n$. It is known that X3C is $\mathcal{NP}$-complete [28, 20]. The following lemma is an immediate consequence of the construction of a reduction function as given by Vattani [43].

Lemma 16. *Let $(U,\mathcal{S})$ be an instance of the* X3C *problem with $|U| = 3n$ and $|\mathcal{S}| = l$. Then there exists a point set $P \subseteq \mathbb{R}^2$ of size $|P| = \mathrm{poly}(n,l)$, a cluster size $k = \Theta(|P|)$, and parameters $L = \mathrm{poly}(n,l)$ and $\alpha = 1/\mathrm{poly}(n,l)$ such that*

1. *If $(U,\mathcal{S}) \in$ X3C then $\mathrm{opt}^{\mathrm{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P) \leq L$.*
2. *If $(U,\mathcal{S}) \notin$ X3C then $\mathrm{opt}^{\mathrm{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P) \geq L + \alpha$.*

Furthermore, P, k, L, and α are computable in time polynomial in n and l.

Proof. Lemma 16 is an immediate consequence of the construction of the reduction instance as given by Vattani [43]. The reader is directed to Vattani's paper for any details on the construction.

Lemma 11 of [43] shows that instance P (named $G_{l,n} \cup X$ in [43]) has a k-means clustering of cost less or equal $L = \mathrm{poly}(n,l)$ if, and only if, the instance of the X3C problem is a yes-instance. Hence, part (1.) of Lemma 16 follows immediately. It remains to prove the second statement of the lemma.

To this end, note that Lemma 10 of [43] states that any optimal k-means clustering of P has cost exactly $L + (n-t)\alpha$ for some integer $t \leq n$ and an $\alpha = 1/\mathrm{poly}(n,l)$. That is, in case of a no-instance of the X3C problem, we obtain $n - t \geq 1$ and we have $\mathrm{opt}^{\mathrm{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P) = L + (n-t)\alpha \geq L + \alpha$. □

Corollary 17. *Lemma 16 also holds with respect to any Mahalanobis distance D_A.*

Note that the gap parameter α in Lemma 16 is non-constant yet computable in polynomial time. Our goal is to give a reduction function that allows to embed the instance P into a small region $B \subseteq \mathrm{ri}(\mathbb{X})$ such

that the cost of the optimal Bregman k-median solution of the embedded instance still decides the X3C problem, that is,

$$(1+\epsilon)L < (1-\epsilon)(L+\alpha) \tag{40}$$

for a sufficiently small ϵ. However, since α is not a constant, the mapping g from Section 3.2 with respect to a globally constant region $B^d_\delta(z) \cap \mathrm{ri}(\mathbb{X})$ will not suffice. Instead, for each instance of the X3C problem, we have to find a fitting δ in polynomial time. Then, a reduction from the X3C problem to a Bregman k-median problem can be obtained as follows. Given any X3C instance, construct point set $P \subseteq \mathbb{R}^2$ as well as k, L, and α as given in [43]. Let $\epsilon > 0$ be such that $(1+\epsilon)L < (1-\epsilon)(L+\alpha)$ or, equivalently,

$$\epsilon < \frac{\alpha}{2L+\alpha} = 1/\operatorname{poly}(n,l). \tag{41}$$

Compute parameter $\delta > 0$ such that the results from Section 3.1 apply with respect to the computed ϵ. Using this δ, we obtain mapping g as given in Section 3.2. The embedded instance is given by $g(P)$. By choice of parameter ϵ we find that $(U,\mathcal{S}) \in$ X3C if, and only if, $\mathrm{opt}^{\mathrm{cost}}_{k,\mathrm{D}_\phi}\big(g(P)\big) \leq (1+\epsilon)L$. Hence, if there exists a polynomial time algorithm solving the Bregman k-median problem optimally, then the X3C problem can be decided in polynomial time.

However, we have to assume that generating function ϕ allows us to compute a fitting δ efficiently. To this end, we call a smooth Bregman divergence *computationally smooth* if it satisfies the following condition: There exists an ϵ_0 such that for any $0 < \epsilon \leq \epsilon_0$, a parameter δ as given in Lemma 10 can be computed in time polynomial in $1/\epsilon$. Since in the case at hand $1/\epsilon = \operatorname{poly}(n,l)$, Theorem 3 follows for all computationally smooth Bregman divergences.

3.4 $\mathcal{NP}$-hardness of concrete Bregman k-median problems

To the best of our knowledge, all Bregman divergences that are used in practice are computationally smooth. An overview of the trade-off between ϵ and δ for a number of computationally smooth Bregman divergences can be found in Table 2.

As an example, in the following we give explicit proofs of the hardness of two of the most practically relevant Bregman divergences, namely the Kullback-Leibler divergence and the Itakura-Saito divergence.

3.4.1 Kullback-Leibler divergence

The (generalized) Kullback-Leibler divergence on domain $\mathbb{X}_{\mathrm{KL}} = \mathbb{R}^d_{\geq 0}$ is defined as

$$\mathrm{D}_{\mathrm{KL}}(p,q) = \sum_{i=1}^{d} \left(p_i \ln \frac{p_i}{q_i} - p_i + q_i \right) \tag{42}$$

for all $p, q \subseteq \mathbb{R}^d_{\geq 0}$. From Lemma 3.10 of [2] we know that when restricted to domain $[\lambda, \upsilon]^d$, the Kullback-Leibler divergence is similar to a certain Mahalanobis distance within a multiplicative error of $\frac{\lambda}{\upsilon}$. The following corollary is an immediate consequence of this result.

Corollary 18. *Let $\upsilon > \lambda > 0$. Then for all $P \subseteq [\lambda, \upsilon]^d$ we have*

$$\frac{1}{2\upsilon} \mathrm{opt}^{\mathrm{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P) \leq \mathrm{opt}^{\mathrm{cost}}_{k,\mathrm{D}_{KL}}(P) \leq \frac{1}{2\lambda} \mathrm{opt}^{\mathrm{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P). \tag{43}$$

Lemma 19. *Let $(U,\mathcal{S})$ be an instance of the* X3C *problem with $|U| = 3n$ and $|\mathcal{S}| = l$. Then there exist a point set $P' \subseteq \mathbb{R}^2_{\geq 0}$ of size $|P'| = \operatorname{poly}(n,l)$, a cluster size $k = \Theta(|P'|)$ and a constant $L' = \operatorname{poly}(n,l)$ such that*

$$(U,\mathcal{S}) \in \text{X3C} \iff \mathrm{opt}^{\mathrm{cost}}_{k,\mathrm{D}_{KL}}(P') \leq L'. \tag{44}$$

Furthermore, P', k, and L' are computable in time polynomial in n and l.

$\mathrm{D}_\phi(p,q)$	$\phi(t)$	z	$\nabla^2\phi(z)$	$B^d_\delta(z) \subseteq \mathbb{X} \subseteq \mathbb{R}^d$
squared Euclidean distance $\|p-q\|_2^2$	$\|t\|_2^2$	any	$2I_d$	any
Mahalanobis distance $(p-q)^\top A\,(p-q)$	$t^\top A\,t$	any	$2A$	any
Kullback-Leibler divergence $\sum p_i \ln(\frac{p_i}{q_i}) - p_i + q_i$	$\sum t_i \ln(t_i) - t_i$	$(1,\dots,1)^\top$	I_d	$\delta \leq \frac{\epsilon}{1+\epsilon}$
Itakura-Saito divergence $\sum \frac{p_i}{q_i} - \ln(\frac{p_i}{q_i}) - 1$	$-\sum \ln(t_i)$	$(1,\dots,1)^\top$	I_d	$\delta \leq \frac{\sqrt{1+\epsilon}-1}{\sqrt{1+\epsilon}}$
harmonic divergence ($a>0$) $\sum \frac{1}{p_i^a} - \frac{a+1}{q_i^a} + \frac{a p_i}{q_i^{a+1}}$	$\sum \frac{1}{t_i^a}$	$(1,\dots,1)^\top$	$(a^2+a)I_d$	$\delta \leq 1 - \sqrt[a+2]{\frac{1}{1+\epsilon}}$
norm-like divergence ($a>1$) $\sum p_i^a + (a-1)q_i^a - a p_i q_i^{a-1}$	$\sum t_i^a$	$(1,\dots,1)^\top$	$(a^2-a)I_d$	$\delta \leq \sqrt[a+2]{1+\epsilon} - 1$
exponential divergence $\sum \exp(p_i) - (p_i - q_i + 1)\exp(q_i)$	$\sum \exp(t_i)$	$(0,\dots,0)^\top$	I_d	$\delta \leq \ln(1+\epsilon)$
reciprocal exponential divergence $\sum \exp(-p_i) - (p_i - q_i + 1)\exp(-q_i)$	$\sum \exp(-t_i)$	$(0,\dots,0)^\top$	I_d	$\delta \leq \ln(1+\epsilon)$
logistic loss $\sum p_i \ln \frac{p_i}{q_i} + (1-p_i)\ln\frac{1-p_i}{1-q_i}$	$\sum t_i \ln t_i + (1-t_i)\ln(1-t_i)$	$(\frac{1}{2},\dots,\frac{1}{2})^\top$	$4I_d$	$\delta \leq \frac{1}{2}\sqrt{\frac{\epsilon}{1+\epsilon}}$
Hellinger-like divergence $\sum \frac{1-p_i q_i}{\sqrt{1-q_i^2}} - \sqrt{1-p_i^2}$	$-\sum \sqrt{1-t_i^2}$	$(0,\dots,0)^\top$	I_d	$\delta \leq \sqrt{1 - \frac{1}{(1+\epsilon)^{\frac{2}{3}}}}$

Table 2: Some computationally smooth Bregman divergences, with ϵ, δ, z as given in Lemma 10.

Proof. Let P, k, L, and α be given as in Lemma 16. Without loss of generality, we may assume $P \subseteq B_1^2(0)$.

The point set P can be mapped in to a small δ-region $B_\delta^2(z)$ around center point $z = (\frac{1}{2},\frac{1}{2})^\top \in \mathbb{R}^2_{\geq 0}$, using mapping $g(x) = \delta x + z$ as described in Section 3.2. Note that $g(P)$ is merely a scaled and translated version of P, that is,

$$\mathrm{D}_{\ell_2^2}\big(g(p), g(q)\big) = \|(\delta p + z) - (\delta q + z)\|^2 = \delta^2 \|p-q\|^2 = \delta^2\, \mathrm{D}_{\ell_2^2}(p,q) \tag{45}$$

for all $p, q \in P$. The intention of this mapping is that any Kullback-Leibler k-median clustering of $g(P)$ yields approximately the same cost as the Euclidean k-means clustering of $g(P)$, with only low (multiplicative) distortion. However, for each constant δ the distortion of this mapping is merely bounded by a constant. Since in the case at hand the gap parameter α is not a constant and depends reciprocally polynomially on n and l, no global constant δ will suffice. Instead, we have to show that for each instance of the X3C problem a fitting δ can be computed in time polynomial in n and l.

To this end, let $\delta = \frac{\alpha}{6L}$. Hence, $\delta = 1/\operatorname{poly}(n,l)$ and $g(P)$ can be computed in time polynomial in n and l. Using Corollary 18, for all $P' \subseteq B_\delta^2(z) \subseteq [\frac{1}{2}-\delta, \frac{1}{2}+\delta]^2$ we have

$$\frac{1}{1+2\delta}\operatorname{opt}^{\mathrm{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P') \leq \operatorname{opt}^{\mathrm{cost}}_{k,\mathrm{D}_{\mathrm{KL}}}(P') \leq \frac{1}{1-2\delta}\operatorname{opt}^{\mathrm{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P'). \tag{46}$$

Furthermore, note that by definition of δ we have

$$2\delta(2L+\alpha) < 6\delta L = \alpha \tag{47}$$

since $L > \alpha$. Hence, we find that

$$2\delta L < \alpha - 2\delta(L+\alpha) \tag{48}$$

which leads to

$$(1+2\delta)L = L + 2\delta L < L + \alpha - 2\delta(L+\alpha) = (1-2\delta)(L+\alpha). \tag{49}$$

Therefore, if $(U,\mathcal{S}) \in \text{X3C}$, using Lemma 16 and (46) we find

$$\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\mathrm{KL}}}\big(g(P)\big) \leq \frac{1}{1-2\delta}\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}\big(g(P)\big) = \frac{\delta^2}{1-2\delta}\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P) \leq \frac{\delta^2 L}{1-2\delta}. \tag{50}$$

On the other hand, if $(U,\mathcal{S}) \notin \text{X3C}$, using Lemma 16, (46), and (49) we conclude

$$\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\mathrm{KL}}}\big(g(P)\big) \geq \frac{1}{1+2\delta}\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}\big(g(P)\big) = \frac{\delta^2}{1+2\delta}\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P) \geq \frac{\delta^2(L+\alpha)}{1+2\delta} > \frac{\delta^2 L}{1-2\delta}. \tag{51}$$

Thus, the lemma follows by choice of $P' = g(P)$ and $L' = \frac{\delta^2 L}{1-2\delta} = \operatorname{poly}(n,l)$. □

Corollary 20. *There exists no algorithm solving the Kullback-Leibler k-median problem optimally with a running time polynomial in n and k, unless $\mathcal{P} = \mathcal{NP}$.*

3.4.2 Itakura-Saito divergence

The (discrete) Itakura-Saito divergence on domain $\mathbb{X}_{\mathrm{IS}} = \mathbb{R}^d_{>0}$ is defined as

$$\mathrm{D}_{\mathrm{IS}}(p,q) = \sum_{i=1}^{d}\left(\frac{p_i}{q_i} - \ln\frac{p_i}{q_i} - 1\right) \tag{52}$$

for all $p, q \subseteq \mathbb{R}^d_{>0}$. From Lemma 3.14 of [2] we know that when restricted to domain $[\lambda,\upsilon]^d$, the Itakura-Saito divergence is similar to a certain Mahalanobis distance within a multiplicative error of $\frac{\lambda^2}{\upsilon^2}$. The following corollary is an immediate consequence of this result.

Corollary 21. *Let $\upsilon > \lambda > 0$. Then for all $P \subseteq [\lambda,\upsilon]^d$ we have*

$$\frac{1}{2\upsilon^2}\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P) \leq \operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{IS}}(P) \leq \frac{1}{2\lambda^2}\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P). \tag{53}$$

Lemma 22. *Let $(U,\mathcal{S})$ be an instance of the* X3C *problem with $|U| = 3n$ and $|\mathcal{S}| = l$. Then there exist a point set $P' \subseteq \mathbb{R}^2_{>0}$ of size $|P'| = \operatorname{poly}(n,l)$, a cluster size $k = \Theta(|P'|)$ and a constant $L' = \operatorname{poly}(n,l)$ such that*

$$(U,\mathcal{S}) \in \text{X3C} \iff \operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{IS}}(P') \leq L'. \tag{54}$$

Furthermore, P', k, and L' are computable in time polynomial in n and l.

Proof. Let P, k, L, and α be given as in Lemma 16. Without loss of generality, we may assume $P \subseteq B^2_1(0)$. As in the proof of Lemma 19, the point set P can be mapped in to a small δ-region $B^2_\delta(z)$ around center point $z = (1,1)^\top \in \mathbb{R}^2_{>0}$ using mapping $g(x) = \delta x + z$, such that for all $p, q \in P$ we find

$$\mathrm{D}_{\ell_2^2}\big(g(p),g(q)\big) = \|(\delta p + z) - (\delta q + z)\|^2 = \delta^2\|p-q\|^2 = \delta^2\,\mathrm{D}_{\ell_2^2}(p,q). \tag{55}$$

We show that for each instance of the X3C problem a fitting δ can be computed in time polynomial in n and l such that any Itakura-Saito k-median clustering of $g(P)$ yields approximately the same cost as the Euclidean k-means clustering of $g(P)$.

To this end, let $\delta = \frac{\alpha}{7L}$. Hence, $\delta = 1/\operatorname{poly}(n,l)$ and $g(P)$ can be computed in time polynomial in n and l. Using Corollary 21, for all $P' \subseteq B^2_\delta(z) \subseteq [1-\delta, 1+\delta]^2$ we have

$$\frac{1}{2(1+\delta)^2}\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P') \leq \operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\mathrm{IS}}}(P') \leq \frac{1}{2(1-\delta)^2}\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P'). \tag{56}$$

Furthermore, note that by definition of δ we have

$$3\delta L + 2\delta(L+\alpha) = \delta(5L+2\alpha) < 7\delta L = \alpha \tag{57}$$

since $L > \alpha$. Hence, we find that

$$3\delta L < \alpha - 2\delta(L+\alpha) \tag{58}$$

which leads to

$$(1+3\delta)L = L + 3\delta L < L + \alpha - 2\delta(L+\alpha) = (1-2\delta)(L+\alpha). \tag{59}$$

Therefore, using $\delta < 1$ we find

$$(1+\delta)^2 L \leq (1+3\delta)L < (1-2\delta)(L+\alpha) \leq (1-\delta)^2(L+\alpha). \tag{60}$$

Hence, if $(U,\mathcal{S}) \in \text{X3C}$, using Lemma 16 and (56) we find

$$\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\text{IS}}}\big(g(P)\big) \leq \frac{1}{2(1-\delta)^2} \operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}\big(g(P)\big) = \frac{\delta^2}{2(1-\delta)^2} \operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P) \leq \frac{\delta^2 L}{2(1-\delta)^2}. \tag{61}$$

On the other hand, if $(U,\mathcal{S}) \notin \text{X3C}$, using Lemma 16, (56), and (60) we conclude

$$\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\text{IS}}}\big(g(P)\big) \geq \frac{1}{2(1+\delta)^2} \operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}\big(g(P)\big) = \frac{\delta^2}{2(1+\delta)^2} \operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_{\ell_2^2}}(P) \geq \frac{\delta^2(L+\alpha)}{2(1+\delta)^2} > \frac{\delta^2 L}{2(1-\delta)^2}. \tag{62}$$

Thus, the lemma follows by choice of $P' = g(P)$ and $L' = \frac{\delta^2 L}{2(1-\delta)^2} = \operatorname{poly}(n,l)$. □

Corollary 23. *There exists no algorithm solving the Itakura-Saito k-median problem optimally with a running time polynomial in n and k, unless $\mathcal{P} = \mathcal{NP}$.*

4 Polynomial time algorithms for trivial Bregman clustering problems

The tractability of a number of 1-dimensional Euclidean clustering problems has already been proven by Brucker in the late 1970s [10]. An adaptation of Brucker's dynamic programming scheme is also applicable to any trivial Bregman k-center (k-diameter, k-median) clustering problem. In a nutshell, the algorithm is based on two observations. First, it can be shown that for any input instance $P \subseteq \mathbb{X}$ there exists a partition of P into its optimal Bregman k-center (k-diameter, k-median) clusters such that any two distinct clusters are separated by a hyperplane. Such an optimal clustering is called linear separable. Second, note that if domain $\mathbb{X}$ is contained in a straight line, this means that there exists some ordering $p_1 \preceq p_2 \preceq \ldots \preceq p_n$ of $P = \{p_1, p_2, \ldots, p_n\}$ along this straight line.

An optimal solution is found as follows. Let P_1 denote the optimal linear separable cluster that contains p_1. A combination of both observations above implies that $P_1 = \{p_1, p_2, \ldots, p_i\}$ for some $1 \leq i \leq n-k$. Hence, we obtain

$$\operatorname{opt}^{\text{rad}}_{k,\mathrm{D}_\phi}(P) = \max\big(\operatorname{opt}^{\text{rad}}_{1,\mathrm{D}_\phi}(P_1), \operatorname{opt}^{\text{rad}}_{k-1,\mathrm{D}_\phi}(P \setminus P_1)\big)\ , \tag{63}$$

$$\operatorname{opt}^{\text{diam}}_{k,\mathrm{D}_\phi}(P) = \max\big(\operatorname{opt}^{\text{diam}}_{1,\mathrm{D}_\phi}(P_1), \operatorname{opt}^{\text{diam}}_{k-1,\mathrm{D}_\phi}(P \setminus P_1)\big)\ , \tag{64}$$

$$\operatorname{opt}^{\text{cost}}_{k,\mathrm{D}_\phi}(P) = \operatorname{opt}^{\text{cost}}_{1,\mathrm{D}_\phi}(P_1) + \operatorname{opt}^{\text{cost}}_{k-1,\mathrm{D}_\phi}(P \setminus P_1)\ , \tag{65}$$

where $\operatorname{opt}^{f}_{1,\mathrm{D}_\phi}(P_1)$ with $f \in \{\text{rad}, \text{diam}, \text{cost}\}$ can be found by enumerating and trying all subsets $\{p_1, p_2, \ldots, p_i\}$ with $1 \leq i \leq n-k$, and $\operatorname{opt}^{f}_{k-1,\mathrm{D}_\phi}(P \setminus P_1)$ can be found recursively. Therefore, an implementation of this approach using dynamic programming computes an optimal solution.

We obtain that any trivial Bregman k-center (k-diameter, k-median) problem can be solved optimally using at most $kn^2 T_f(n)$ arithmetic operations, including evaluations of Bregman divergence D_ϕ. Here, $T_f(n)$

is the number of operations necessary to compute the Bregman 1-center radius (1-diameter, 1-median cost) of a point set $\{p_i, \ldots, p_j\} \subseteq P$. That is, we have $T_{\text{rad}}(n) = \mathcal{O}(d)$ since the optimal 1-center can be found by computing the intersection of $\mathbb{X}$ with the Bregman bisector of p_i and p_j [37], $T_{\text{diam}}(n) = \mathcal{O}(1)$ for computing the maximum of $\mathrm{D}_\phi(p_i, p_j)$ and $\mathrm{D}_\phi(p_j, p_i)$, and $T_{\text{cost}}(n) = \mathcal{O}(dn)$ for finding the Bregman 1-median, which is given by the centroid of $\{p_i, \ldots, p_j\}$ [7]. Thus, the algorithm runs in time polynomial in n, d, and k.

5 Discussion

In this paper, we have shown the $\mathcal{NP}$-hardness of the Bregman k-center, the Bregman k-median, and the Bregman k-diameter problem if the number of clusters k is part of the input. Furthermore, we have shown that it is even $\mathcal{NP}$-hard to approximate the Bregman k-center problem within a factor of 3.32, and the Bregman k-diameter problem within a factor of 3.87. Additionally, we have shown that the Bregman k-median problem can not be approximated within factor $\alpha \geq 1$ in time polynomial in n and k, unless the Euclidean k-means problem can be approximated within a factor $\alpha + \epsilon$ in time polynomial in n and k. Hence, to prove non-approximability results for the Bregman k-median problem in general, it is sufficient to find non-approximability results for the Euclidean k-means problem. Unfortunately, it is an open question whether there holds any non-approximability result for the Euclidean k-means problem, or whether a $(1+\epsilon)$-approximation can be achieved in time polynomial in n and k.

In principle, an adaptation of our reduction function could be applied to the case when the number of clusters k is fixed but the dimension d is part of the input. The only modification we would have to make is to ensure that a d-dimensional unit ball can be embedded in the domain $\mathbb{X}$ of a given Bregman divergence. However, unlike the case we considered, there are no non-approximability results for Euclidean k-clustering problems if $k = \Theta(1)$. In particular, there exist $(1 + \epsilon)$-approximation algorithms with a running time polynomial in n and d (yet exponential in k) for both the Euclidean k-center [6, 5] and the Euclidean k-means problem [18, 29, 17, 12]. There even exists a PTAS for the Bregman k-median problem with constant k that is applicable to all instances generated by our reduction function [2, 1]. Hence, our approach can not be used to show the non-approximability of Bregman k-clustering problems if k is a constant, and it remains an open problem to show any hardness result in this case.

A variation of center based k-clustering problems (i.e., k-center, k-median) is defined as follows: Find a set of k centers that minimizes the objective function when the centers are restricted to be elements from the set of input points. We call these variations the *discrete k-center problem* and the *discrete k-median problem*. In case of the Kullback-Leibler divergence, Chaudhuri and McGregor [11] showed that there exists no constant factor approximation algorithm for the discrete Kullback-Leibler k-center problem and the discrete Kullback-Leibler k-median problem, unless $\mathcal{P} = \mathcal{NP}$. To this end, Chaudhuri and McGregor made use of the fact that there exist singularities on the domain $\mathbb{X}$ of the Kullback-Leibler divergence, that is, there are points q on the relative boundary of $\mathbb{X}$ and $\{q_i\}_{i \in \mathbb{N}} \subseteq \mathrm{ri}(\mathbb{X})$ with $q_i \to q$ such that for all $p \in \mathrm{ri}(\mathbb{X})$ we have $\mathrm{D}_\phi(p, q_i) \to \infty$. Many other Bregman divergences (such as the Itakura-Saito divergence) do exhibit singularities as well, and it is highly likely that the approach of Chaudhuri and McGregor can be extended to these Bregman divergences. On the other hand, in the case of the discrete Euclidean k-center and the discrete Euclidean k-means problem, there exist algorithms with an approximation guarantee of 2 [21, 22, 16] and $9 + \epsilon$ [27], respectively. Hence, there exist constant factor approximation algorithms for the case of the squared Euclidean distance (and, in fact, for the whole family of Mahalanobis distances). Essentially nothing is known on the (non-)approximability of these discrete k-clustering problems if Bregman divergence D_ϕ is not a Mahalanobis distance, but does not exhibit singularities, either. Therefore, it is an open problem to classify the (non-)approximability of discrete Bregman k-clustering problems in general.

References

[1] Marcel R. Ackermann and Johannes Blömer. Coresets and approximate clustering for Bregman divergences. In *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '09)*, pages 1088–1097. Society for Industrial and Applied Mathematics, 2009.

[2] Marcel R. Ackermann, Johannes Blömer, and Christian Sohler. Clustering for metric and non-metric distance measures. *ACM Transactions on Algorithms*, 6(4):59:1–26, August 2010.

[3] Daniel Aloise, Amit Deshpande, Pierre Hansen, and Preyas Popat. $\mathcal{NP}$-hardness of Euclidean sum-of-squares clustering. *Machine Learning*, 75(2):245–248, May 2009.

[4] David Arthur and Sergei Vassilvitskii. `k-means++`: the advantages of careful seeding. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '07)*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[5] Mihai Bădoiu and Kenneth L. Clarkson. Smaller core-sets for balls. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '03)*, pages 801–802. Society for Industrial and Applied Mathematics, 2003.

[6] Mihai Bădoiu, Sariel Har-Peled, and Piotr Indyk. Approximate clustering via core-sets. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC '02)*, pages 250–257. Association for Computing Machinery, 2002.

[7] Arindam Banerjee, Srujana Merugu, Inderjit S. Dhillon, and Joydeep Ghosh. Clustering with Bregman divergences. *Journal of Machine Learning Research*, 6:1705–1749, October 2005.

[8] Arnab Bhattacharya, Purushottam Kar, and Manjish Pal. On low distortion embeddings of statistical distance measures into low dimensional spaces. In *Proceedings of the 20th International Conference on Database and Expert Systems Applications (DEXA'09)*, pages 164–172. Springer, 2009.

[9] Lev M. Bregman. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.

[10] Peter Brucker. On the complexity of clustering problems. In *Optimization and Operations Research: Proceedings of a Workshop Held at the University of Bonn. Lecture Notes in Economics and Mathematical Systems 157*, pages 45–54. Springer, 1977.

[11] Kamalika Chaudhuri and Andrew McGregor. Finding metric structure in information theoretic clustering. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT '08)*, pages 391–402. Omnipress, 2008.

[12] Ke Chen. On coresets for k-median and k-means clustering in metric and Euclidean spaces and their applications. *SIAM Journal on Computing*, 39(3):923–947, August 2009.

[13] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, Hoboken, 2nd edition, 2006.

[14] Sanjoy Dasgupta. The hardness of k-means clustering. Technical Report CS2007-0890, University of California, San Diego, 2007.

[15] Petros Drineas, Alan M. Frieze, Ravi Kannan, Santosh Vempala, and V. Vinay. Clustering large graphs via the singular value decomposition. *Machine Learning*, 56(1–3):9–33, July 2004.

[16] Tomás Feder and Daniel H. Greene. Optimal algorithms for approximate clustering. In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing (STOC'88)*, pages 434–444. Association for Computing Machinery, 1988.

[17] Dan Feldman, Morteza Monemizadeh, and Christian Sohler. A PTAS for k-means clustering based on weak coresets. In *Proceedings of the 23rd ACM Symposium on Computational Geometry (SCG '07)*, pages 11–18. Association for Computing Machinery, 2007.

[18] Wenceslas Fernandez de la Vega, Marek Karpinski, Claire Kenyon, and Yuval Rabani. Approximation schemes for clustering problems. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC '03)*, pages 50–58. Association for Computing Machinery, 2003.

[19] Robert J. Fowler, Mike Paterson, and Steven L. Tanimoto. Optimal packing and covering in the plane are $\mathcal{NP}$-complete. *Information Processing Letters*, 12(3):133–137, June 1981.

[20] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. A Series of Books in the Mathematical Sciences. W. H. Freeman & Co., New York, 1979.

[21] Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, June 1985.

[22] Dorit S. Hochbaum and David B. Shmoys. A best possible heuristic for the k-center problem. *Mathematics of Operations Research*, 10(2):180–184, May 1985.

[23] Dorit S. Hochbaum and David B. Shmoys. A unified approach to approximation algorithms for bottleneck problems. *Journal of the ACM*, 33(3):533–550, July 1986.

[24] Mary Inaba, Naoki Katoh, and Hiroshi Imai. Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering. In *Proceedings of the 10th ACM Symposium on Computational Geometry (SCG '94)*, pages 332–339. Association for Computing Machinery, 1994.

[25] Fumitada Itakura and Shuzo Saito. Analysis synthesis telephony based on the maximum likelihood method. In *Reports of the 6th International Congress on Acoustics*, pages 17–20. Elsevier, 1968.

[26] Gaurav Kanade, Prajakta Nimbhorkar, and Kasturi R. Varadarajan. On the $\mathcal{NP}$-hardness of the 2-means problem. Manuscript, 2008.

[27] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu. A local search approximation algorithm for k-means clustering. *Computational Geometry*, 28(2-3):89–112, June 2004.

[28] Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, New York, 1972.

[29] Amit Kumar, Yogish Sabharwal, and Sandeep Sen. A simple linear time $(1+\epsilon)$-approximation algorithm for k-means clustering in any dimensions. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS '04)*, pages 454–462. IEEE Computer Society, 2004.

[30] Stuart P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982.

[31] Meena Mahajan, Prajakta Nimbhorkar, and Kasturi R. Varadarajan. The planar k-means problem is $\mathcal{NP}$-hard. In *Proceedings of the 3rd Annual Workshop on Algorithms and Computation (WALCOM '09)*, pages 274–285. Springer, 2009.

[32] Prasanta Chandra Mahalanobis. On the generalized distance in statistics. In *Proceedings of the National Institute of Sciences of India*, volume 2(1), pages 49–55. Indian National Science Academy, 1936.

[33] Bodo Manthey and Heiko Röglin. Worst-case and smoothed analysis of k-means clustering with Bregman divergences. In *Proceedings of the 20th International Symposium on Algorithms and Computation (ISAAC '09)*, volume 5878 of *Lecture Notes in Computer Science*, pages 1024–1033. Springer, 2009.

[34] Nimrod Megiddo. On the complexity of some geometric problems in unbounded dimension. *Journal of Symbolic Computation*, 10(3–4):327–334, 1990.

[35] Nimrod Megiddo and Kenneth J. Supowit. On the complexity of some common geometric location problems. *SIAM Journal on Computing*, 13(1):182–196, February 1984.

[36] Stuart G. Mentzer. Lower bounds on metric k-center problems. Manuscript, 1988.

[37] Frank Nielsen, Jean-Daniel Boissonnat, and Richard Nock. On Bregman Voronoi diagrams. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '07)*, pages 746–755. Society for Industrial and Applied Mathematics, 2007.

[38] Richard Nock, Panu Luosto, and Jyrki Kivinen. Mixed Bregman clustering with approximation guarantees. In *Proceedings of the 19th European Conference on Machine Learning (ECML '08)*, pages 154–169. Springer, 2008.

[39] Richard Nock and Frank Nielsen. Fitting the smallest enclosing Bregman ball. In *Proceedings of the 16th European Conference on Machine Learning (ECML '05)*, pages 649–656. Springer, 2005.

[40] Eric Spellman, Baba C. Vemuri, and Murali Rao. Using the KL-center for efficient and accurate retrieval of distributions arising from texture images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pages 111–116. IEEE Computer Society, 2005.

[41] Suvrit Sra, Stefanie Jegelka, and Arindam Banerjee. Approximation algorithms for Bregman clustering, co-clustering and tensor clustering. Technical Report MPIK-TR-177, Max Planck Institure for Biological Cybernetics, 2008.

[42] Gilbert Strang. *Linear Algebra and Its Applications*. Thomson Brooks/Cole, Belmont, 4th edition, 2006.

[43] Andrea Vattani. The hardness of k-means clustering in the plane. Unpublished manuscript, available at: `http://cseweb.ucsd.edu/users/avattani/`, 2010.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de