Coresets and Approximate Clustering for Bregman Divergences

Marcel R. Ackermann
University of Paderborn, Germany

Joint work with Johannes Blömer

ACM-SIAM Symposium on Discrete Algorithms (SODA)
January 6th, 2009
Introduction
- Bregman $k$-median clustering
- Bregman divergences
- Mahalanobis distances and $\mu$-similarity
- Our results

Weak coreset construction
- $\Gamma$-weak $(k, \epsilon)$-coresets
- Chen’s coreset construction for metrics
- Initial $O(\log k)$-approximation

$(1 + \epsilon)$-approximation algorithm
- Using weak coresets
- Size bound for $\Gamma$

Open problems
1 Introduction
- Bregman $k$-median clustering
- Bregman divergences
- Mahalanobis distances and $\mu$-similarity
- Our results

2 Weak coreset construction
- $\Gamma$-weak $(k, \epsilon)$-coresets
- Chen’s coreset construction for metrics
- Initial $O(\log k)$-approximation

3 $(1 + \epsilon)$-approximation algorithm
- Using weak coresets
- Size bound for $\Gamma$

4 Open problems
Bregman $k$-median clustering

Euclidean $k$-means clustering:

$n$ points in $\mathbb{R}^d$
Bregman $k$-median clustering

Euclidean $k$-means clustering:

$n$ points in $\mathbb{R}^d$
**Bregman k-median clustering**

Euclidean k-means clustering:

- n points in \( \mathbb{R}^d \)
- Find C of size k that minimizes \( \sum_{p \in P} \min_{c \in C} \| p - c \|^2 \)
Bregman $k$-median clustering

Euclidean $k$-means clustering:

- $n$ points in $\mathbb{R}^d$
- Find $C$ of size $k$ that minimizes $\sum_{p \in P} \min_{c \in C} \| p - c \|^2$

Bregman $k$-median clustering:

- Bregman divergence $D_\phi$
- Find $C$ of size $k$ that minimizes $\sum_{p \in P} \min_{c \in C} D_{\phi}(p, c)$
Bregman divergences

Some Bregman divergences:

- **Squared Euclidean distance** (geometric applications)
Coresets and Approximate Clustering for Bregman Divergences

Marcel R. Ackermann

Introduction

$k$-median clustering
Bregman divergences
$\mu$-similarity
Our results

Weak coresets

$(k, \epsilon)$-coresets
Chen's construction
Initial approximation

Algorithm

Using weak coresets
Size bound for $\Gamma$

Open problems

Bregman divergences

Some Bregman divergences:

- **Squared Euclidean distance** (geometric applications)
- **Kullback-Leibler divergence** (information theory, etc.)

Properties:

- centroid is the optimal 1-median of a given cluster
- optimal cluster partition is linear separable

In general:

- asymmetric, no triangle inequality
- may possess singularities on $\mathbb{R}^d$, i.e.:
  \[ D_\phi(p, q) = \infty \]
Bregman divergences

Some Bregman divergences:

- **Squared Euclidean distance** (geometric applications)
- **Kullback-Leibler divergence** (information theory, etc.)
- **Itakura-Saito divergence** (speech processing)
Bregman divergences

Some Bregman divergences:

- **Squared Euclidean distance** (geometric applications)
- **Kullback-Leibler divergence** (information theory, etc.)
- **Itakura-Saito divergence** (speech processing)
- **Mahalanobis distances** (statistical testing, etc.)
- many more …
Bregman divergences

Some Bregman divergences:
- **Squared Euclidean distance** (geometric applications)
- **Kullback-Leibler divergence** (information theory, etc.)
- **Itakura-Saito divergence** (speech processing)
- **Mahalanobis distances** (statistical testing, etc.)
- many more ...

Properties:
- **centroid** is the optimal 1-median of a given cluster
- optimal cluster partition is **linear separable**
Bregman divergences

Some Bregman divergences:
- **Squared Euclidean distance** (geometric applications)
- **Kullback-Leibler divergence** (information theory, etc.)
- **Itakura-Saito divergence** (speech processing)
- **Mahalanobis distances** (statistical testing, etc.)
- many more ...

Properties:
- **centroid** is the optimal 1-median of a given cluster
- optimal cluster partition is **linear separable**

In general:
- asymmetric, no triangle inequality
- may possess singularities on $\mathbb{R}^d$, i.e.: $D_\phi(p, q) = \infty$
### Recent related work

<table>
<thead>
<tr>
<th><strong>Squared Euclidean distance</strong> (a very small selection)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[Kumar et al., 2004]</td>
<td>$(1 + \epsilon) \quad \mathcal{O}(2^{(\frac{k}{\epsilon})^c} dn)$</td>
</tr>
<tr>
<td>[Chen, 2006]</td>
<td>$(1 + \epsilon) \quad \mathcal{O}(ndk + 2^{(\frac{k}{\epsilon})^c} d^2 n^\sigma)$</td>
</tr>
<tr>
<td>[Feldman et al., 2007]</td>
<td>$(1 + \epsilon) \quad \mathcal{O}(ndk + d^{(\frac{k}{\epsilon})^c} + 2^{\tilde{O}(\frac{k}{\epsilon})})$</td>
</tr>
<tr>
<td>[Arthur et al., 2007]</td>
<td>$\mathcal{O}(\log k) \quad \mathcal{O}(ndk)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Kullback-Leibler divergence</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>[Baker et al., 1998]</td>
</tr>
<tr>
<td>[Dhillon et al., 2003]</td>
</tr>
<tr>
<td>[Chaudhuri et al., 2008]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>General Bregman divergences</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>[Banerjee et al., 2005]</td>
</tr>
<tr>
<td>[Ackermann et al., 2008]*</td>
</tr>
<tr>
<td>[Nock et al., 2008]**</td>
</tr>
</tbody>
</table>

*) under assumption of $\mu$-similarity
**) under a similar assumption
Bregman divergences

Definition

Let $\phi : \mathcal{X} \rightarrow \mathbb{R}$ strictly convex, diff’able on convex $\mathcal{X} \subseteq \mathbb{R}^d$.

$$D_\phi(p, q) = \phi(p) - \phi(q) - \langle \nabla \phi(q), p - q \rangle$$

A Bregman divergence measures the error when approximating a convex function by a tangent hyperplane.
Bregman divergences

Definition

Let $\phi : \mathcal{X} \to \mathbb{R}$ strictly convex, diff’able on convex $\mathcal{X} \subseteq \mathbb{R}^d$.

$$D_{\phi}(p, q) = \phi(p) - \phi(q) - \langle \nabla \phi(q), p - q \rangle$$

A Bregman divergence measures the error when approximating a convex function by a tangent hyperplane.
Bregman divergences

**Definition**

Let \( \phi : \mathcal{X} \rightarrow \mathbb{R} \) strictly convex, diff’able on convex \( \mathcal{X} \subseteq \mathbb{R}^d \).

\[
D_{\phi}(p, q) = \phi(p) - \phi(q) - \langle \nabla \phi(q), p - q \rangle
\]

\( D_{\phi}(p, q) \) is **Lagrange remainder term** for some \( \xi \in [p, q] \):

\[
D_{\phi}(p, q) = \frac{1}{2} (p - q)^\top \nabla^2 \phi(\xi) (p - q)
\]
Bregman divergences

Definition

Let \( \phi : \mathcal{X} \to \mathbb{R} \) strictly convex, diff’able on convex \( \mathcal{X} \subseteq \mathbb{R}^d \).

\[
D_\phi(p, q) = \phi(p) - \phi(q) - \langle \nabla \phi(q), p - q \rangle
\]

\( D_\phi(p, q) \) is **Lagrange remainder term** for some \( \xi \in [p, q] \):

\[
D_\phi(p, q) = \frac{1}{2} (p - q)^\top \nabla^2 \phi(\xi) (p - q)
\]

**Mahalanobis distance** w.r.t. positive definite \( A \):

\[
D_A(p, q) = (p - q)^\top A (p - q)
\]
Mahalanobis distances and $\mu$-similarity

**Mahalanobis distance** w.r.t. positive definite $A$:

$$D_A(p, q) = (p - q)^\top A (p - q)$$

$D_A$ are "well-natured" Bregman divergences:
- symmetric
- double triangle inequality
Mahalanobis distances and $\mu$-similarity

**Mahalanobis distance** w.r.t. positive definite $A$:

$$D_A(p, q) = (p - q)^\top A (p - q)$$

$D_A$ are ”well-natured” Bregman divergences:
- symmetric
- double triangle inequality

**Definition**

$D_\phi$ on $\mathcal{X}$ is called $\mu$-**similar** iff there exists a Mahalanobis distance $D_A$ such that for all $p, q \in \mathcal{X}$:

$$\mu D_A(p, q) \leq D_\phi(p, q) \leq D_A(p, q)$$
Mahalanobis distance w.r.t. positive definite $A$:

$$D_A(p, q) = (p - q)\top A (p - q)$$

$D_A$ are "well-natured" Bregman divergences:
- symmetric
- double triangle inequality

**Definition**

$D_\phi$ on $X$ is called $\mu$-similar iff there exists a Mahalanobis distance $D_A$ such that for all $p, q \in X$:

$$\mu D_A(p, q) \leq D_\phi(p, q) \leq D_A(p, q)$$

E.g., choose $0 < \mu \leq \frac{\min_{\zeta \in X} (p - q)\top \nabla^2 \phi(\zeta) (p - q)}{\max_{\xi \in X} (p - q)\top \nabla^2 \phi(\xi) (p - q)} \quad \forall p, q \in X$
Our results

For all $\mu$-similar Bregman divergences $D_\phi$

- "strong coreset" construction from [Chen, 2006] gives "weak coresets" for the Bregman $k$-median problem
Our results

For all $\mu$-similar Bregman divergences $D_\phi$

- "strong coreset" construction from [Chen, 2006] gives "weak coresets" for the Bregman $k$-median problem
  
- using weak coresets and [A., Blömer, Sohler, 2008]:
  - $(1 + \epsilon)$-approximation algorithm
  - running time $O(ndk + d 2^{(k/\epsilon)c} \log^{k+2} n)$
    (previously: $O(d 2^{(k/\epsilon)c} n)$)
Coresets and Approximate Clustering for Bregman Divergences

Marcel R. Ackermann

Introduction
- $k$-median clustering
- Bregman divergences
- Mahalanobis distances and $\mu$-similarity
- Our results

Weak coreset construction
- $\Gamma$-weak $(k, \epsilon)$-coresets
- Chen’s coreset construction for metrics
- Initial $O(\log k)$-approximation

(1 + $\epsilon$)-approximation algorithm
- Using weak coresets
- Size bound for $\Gamma$

Open problems
"a small representation of the clustering behavior of $P$"
Using coresets

"a small representation of the clustering behavior of $P$"
Using coresets

"a small representation of the clustering behavior of $P$"
Using coresets

"a small representation of the clustering behavior of $P$"
Γ-weak \((k, \epsilon)\)-coresets

More formally:

A "strong" \((k, \epsilon)\)-coreset of \(P\) is a (weighted) set \(S \subseteq X\) such that for all \(C \subseteq X\) with \(|C| = k\):

\[
\text{cost}_w(S, C) = (1 \pm \epsilon) \text{cost}(P, C)
\]
\( \Gamma \)-weak \((k, \epsilon)\)-coresets

More formally:

A "strong" \((k, \epsilon)\)-coreset of \(P\) is a (weighted) set \(S \subseteq \mathcal{X}\) such that for all \(C \subseteq \mathcal{X}\) with \(|C| = k\):

\[
\text{cost}_w(S, C) = (1 \pm \epsilon) \text{ cost}(P, C)
\]

Relaxation: Consider only medians \(C\) that are relevant to \(P\)!
\( \Gamma \)-weak \((k, \epsilon)\)-coresets

More formally:

A "strong" \((k, \epsilon)\)-coreset of \( P \) is a (weighted) set \( S \subseteq \mathcal{X} \) such that for all \( C \subseteq \mathcal{X} \) with \(|C| = k\):

\[
\text{cost}_w(S, C) = (1 \pm \epsilon) \text{ cost}(P, C)
\]

Relaxation: Consider only medians \( C \) that are relevant to \( P \)!

**Definition**

For a **finite** \( \Gamma \subseteq \mathcal{X} \), a (weighted) set \( S \subseteq \mathcal{X} \) is called a \( \Gamma \)-weak \((k, \epsilon)\)-coreset of \( P \) iff for all \( C \subseteq \Gamma \) with \(|C| = k\):

\[
\text{cost}_w(S, C) = (1 \pm \epsilon) \text{ cost}(P, C)
\]
Chen’s coreset construction for metrics

Chen’s coreset construction for metrics

Introduction

- k-median clustering
- Bregman divergences
- \( \mu \)-similarity
- Our results

Weak coresets

- \((k, \epsilon)\)-coresets
- Chen’s construction
- Initial approximation

Algorithm

- Using weak coresets
- Size bound for \( \Gamma \)

Open problems
Chen’s coreset construction for metrics

1. Obtain initial $\alpha$-approximation $A = \{a_i\}$
Chen’s coreset construction for metrics

1. Obtain initial $\alpha$-approximation $A = \{a_i\}$
2. Partition $P$ into exponentially growing ring sets $P_{ij}$
1. Obtain initial $\alpha$-approximation $A = \{a_i\}$
2. Partition $P$ into exponentially growing ring sets $P_{ij}$
3. Choose $m$ points uniformly at random from each ring set and assign weight $\frac{1}{m}|P_{ij}|$
Existence of $\Gamma$-weak coresets for $D_\phi$

Straight-forward adaptation of [Chen, 2006] seems infeasible! (technical difficulties arise from asymmetry)
Existence of $\Gamma$-weak coresets for $D_\phi$

Straight-forward adaptation of [Chen, 2006] seems infeasible! (technical difficulties arise from asymmetry)

**Lemma**

Let $C \subseteq X$ be fixed. If $m = \Omega \left( \frac{\alpha^2}{\epsilon^2 \mu^2} \log(k \log(n)/\delta) \right)$ then with prob. $1 - \delta$ we have $\text{cost}_w(S, C) = (1 \pm \epsilon) \text{cost}(P, C)$.

(using adaptation of a proof from [Chen, 2006] and $\mu$-similarity)
Straight-forward adaptation of [Chen, 2006] seems infeasible! (technical difficulties arise from asymmetry)

**Lemma**

Let $C \subseteq \mathcal{X}$ be fixed. If $m = \Omega\left(\frac{\alpha^2}{\epsilon^2 \mu^2} \log(k \log(n)/\delta)\right)$ then with prob. $1 - \delta$ we have $\text{cost}_w(S, C) = (1 \pm \epsilon) \text{cost}(P, C)$.

(using adaptation of a proof from [Chen, 2006] and $\mu$-similarity)

**Theorem**

With high probability, Chen's construction gives a $\Gamma$-weak $(k, \epsilon)$-coreset of size $\Theta\left(\frac{\alpha^2 k}{\epsilon^2 \mu^2} \log(n) \log(|\Gamma|^k k \log n)\right)$.

(using Lemma, $\delta = \Theta(1/|\Gamma|^k)$, and union bound)
An initial $\mathcal{O}(\log k)$-approximation algorithm

Use $k$-means++ seeding from [Arthur, Vassilvitskii, 2007].

**Theorem**

Let $A$ be obtained by $D_\phi$-seeding. Then:

$$\mathbb{E}[\text{cost}(P, A)] \leq \frac{8}{\mu^2} (2 + \ln k) \text{opt}(P).$$
An initial $O(\log k)$-approximation algorithm

Use $k$-means++ seeding from [Arthur, Vassilvitskii, 2007].

**Theorem**

Let $A$ be obtained by $D_\phi$-seeding. Then:

$$
\mathbb{E}[\text{cost}(P, A)] \leq \frac{8}{\mu^2} (2 + \ln k) \text{opt}(P).
$$

Independently, (essentially) same generalization given by

- [Nock, Luosto, Kivinen, 2008]
- [Sra, Jegelka, Banerjee, 2008]
Coresets and Approximate Clustering for Bregman Divergences

Marcel R. Ackermann

Outline

1 Introduction
   - Bregman $k$-median clustering
   - Bregman divergences
   - Mahalanobis distances and $\mu$-similarity
   - Our results

2 Weak coreset construction
   - $\Gamma$-weak ($k, \epsilon$)-coresets
   - Chen’s coreset construction for metrics
   - Initial $O(\log k)$-approximation

3 $(1 + \epsilon)$-approximation algorithm
   - Using weak coresets
   - Size bound for $\Gamma$

4 Open problems
Using $\Gamma$-weak coresets

Algorithm $\text{CORECLUSTER}(P, k)$:

1. Construct $\Gamma$-weak $(k, \epsilon)$-coreset $(S, w)$ of $P$
2. Compute $(1 + \epsilon)$-approximation $\tilde{C}$ of $(S, w)$ using algorithm $\text{CLUSTER}$ from [A., Blömer, Sohler, 2008]

$\Gamma = \"\text{set of all medians that are relevant}\"$  size bound?
Using $\Gamma$-weak coresets

Algorithm $\text{CORECLUSTER}(P, k)$:

1. Construct $\Gamma$-weak $(k, \epsilon)$-coreset $(S, w)$ of $P$
2. Compute $(1 + \epsilon)$-approximation $\tilde{C}$ of $(S, w)$ using algorithm $\text{CLUSTER}$ from [A., Blömer, Sohler, 2008]

$$\Gamma = \ldots$$

$$\text{cost}(P, \tilde{C})$$
Using $\Gamma$-weak coresets

Algorithm $\text{CORECLUSTER}(P, k)$:

1. Construct $\Gamma$-weak $(k, \epsilon)$-coreset $(S, w)$ of $P$
2. Compute $(1 + \epsilon)$-approximation $\tilde{C}$ of $(S, w)$ using algorithm $\text{CLUSTER}$ from [A., Blömer, Sohler, 2008]

$$\Gamma = \Gamma_{\text{CLUSTER}} \cup \ldots$$

with: $\Gamma_{\text{CLUSTER}} \overset{\text{def}}{=} "\text{set of all outputs of algorithm } \text{CLUSTER"}"

$$\text{cost}(P, \tilde{C}) \leq (1 + \epsilon) \text{cost}_w(S, \tilde{C})$$
Using $\Gamma$-weak coresets

**Algorithm** \textsc{CoreCluster}($P, k$):

1. Construct $\Gamma$-weak $(k, \epsilon)$-coreset $(S, w)$ of $P$
2. Compute $(1 + \epsilon)$-approximation $\tilde{C}$ of $(S, w)$ using algorithm \textsc{Cluster} from [A., Blömer, Sohler, 2008]

\[
\Gamma = \Gamma_{\text{Cluster}} \cup \ldots
\]

with: $\Gamma_{\text{Cluster}}$ = ”set of all outputs of algorithm \textsc{Cluster}”

\[
\text{cost}(P, \tilde{C}) \leq (1 + \epsilon) \text{cost}_w(S, \tilde{C}) \leq (1 + \epsilon)^2 \text{opt}(S, w)
\]
Using $\Gamma$-weak coresets

Algorithm $\text{CORECLUSTER}(P, k)$:

1. Construct $\Gamma$-weak $(k, \epsilon)$-coreset $(S, w)$ of $P$
2. Compute $(1 + \epsilon)$-approximation $\tilde{C}$ of $(S, w)$ using algorithm $\text{CLUSTER}$ from [A., Blömer, Sohler, 2008]

$$\Gamma = \Gamma_{\text{CLUSTER}} \cup \ldots$$

with: $\Gamma_{\text{CLUSTER}} \triangleq ”$ set of all outputs of algorithm $\text{CLUSTER}”$

$C_{\text{opt}}(P) \triangleq ”$ optimal $k$-medians of $P”$

$$\text{cost}(P, \tilde{C}) \leq (1 + \epsilon) \text{cost}_w(S, \tilde{C}) \leq (1 + \epsilon)^2 \text{opt}(S, w) \leq (1 + \epsilon)^2 \text{cost}_w(S, C_{\text{opt}}(P))$$
Using $\Gamma$-weak coresets

Algorithm $\text{CoreCluster}(P, k)$:

1. Construct $\Gamma$-weak $(k, \epsilon)$-coreset $(S, w)$ of $P$
2. Compute $(1 + \epsilon)$-approximation $\tilde{C}$ of $(S, w)$ using algorithm $\text{Cluster}$ from [A., Blömer, Sohler, 2008]

\[
\Gamma = \Gamma_{\text{Cluster}} \cup C_{\text{opt}}(P)
\]

with: $\Gamma_{\text{Cluster}}$ = ”set of all outputs of algorithm $\text{Cluster}”$  
\[C_{\text{opt}}(P) = “\text{optimal } k\text{-medians of } P”\]

\[
\text{cost}(P, \tilde{C}) \leq (1 + \epsilon) \text{cost}_w(S, \tilde{C}) \leq (1 + \epsilon)^2 \text{opt}(S, w)
\]
\[
\leq (1 + \epsilon)^2 \text{cost}_w(S, C_{\text{opt}}(P)) \leq (1 + \epsilon)^3 \text{cost}(P, C_{\text{opt}}(P)) = \text{opt}(P)
\]
Size bound for $\Gamma$

$$\Gamma_{\text{\textsc{cluster}}} \doteq \"\text{set of all outputs}\ of\ \text{algorithm}\ \text{\textsc{cluster}}\"$$

Key features of algorithm $\text{\textsc{cluster}}$:

- each output median is computed as the (weighted) centroid of a poly($k/\epsilon$)-sized subset of $P$
Size bound for $\Gamma$

$\Gamma_{\text{CLUSTER}} \triangleq "\text{set of all outputs of algorithm } C_{\text{LUSTER}}"$

Key features of algorithm $C_{\text{LUSTER}}$:
- Each output median is computed as the (weighted) centroid of a $\text{poly}(k/\epsilon)$-sized subset of $P$

$\Rightarrow |\Gamma_{\text{CLUSTER}}| \leq n^{\text{poly}(k/\epsilon)}$
Size bound for $\Gamma$

$\Gamma_{\text{CLUSTER}} \triangleq \text{"set of all outputs of algorithm } C_{\text{LUSTER}}\text{"}$

Key features of algorithm $C_{\text{LUSTER}}$:

- each output median is computed as the (weighted) centroid of a $\text{poly}(k/\epsilon)$-sized subset of $P$

$\Rightarrow |\Gamma_{\text{CLUSTER}}| \leq n^{\text{poly}(k/\epsilon)}$

$C_{\text{opt}}(P) \triangleq \text{"optimal } k\text{-medians of } P\text{"}$

$\Rightarrow |C_{\text{opt}}(P)| = k$
Coresets and Approximate Clustering for Bregman Divergences

Marcel R. Ackermann

Introduction

$k$-median clustering
Bregman divergences
$\mu$-similarity
Our results

Weak coresets

$(k, \epsilon)$-coresets
Chen’s construction
Initial approximation

Algorithm

Using weak coresets
Size bound for $\Gamma$

Open problems

Size bound for $\Gamma$

\[ \Gamma_{\text{Cluster}} \triangleq \text{"set of all outputs of algorithm } C_{\text{Cluster}"} \]

Key features of algorithm $C_{\text{Cluster}}$:

- each output median is computed as the (weighted) centroid of a $\text{poly}(k/\epsilon)$-sized subset of $P$

\[ \Rightarrow |\Gamma_{\text{Cluster}}| \leq n^{\text{poly}(k/\epsilon)} \]

$C_{\text{opt}}(P) \triangleq \text{"optimal } k\text{-medians of } P"$

\[ \Rightarrow |C_{\text{opt}}(P)| = k \]

Lemma

\[ |\Gamma| \leq n^{\text{poly}(k/\epsilon)} + k \]
Results

Corollary

*W.h.p., using Chen’s construction we obtain a $\Gamma$-weak $(k, \epsilon)$-coreset $(S, w)$ of size $\Theta(\text{poly}(k/\epsilon) \log^2(n))$ in time $O(ndk + |S|)$.*
Results

Corollary

W.h.p., using Chen’s construction we obtain a $\Gamma$-weak $(k, \epsilon)$-coreset $(S, w)$ of size $\Theta(\text{poly}(k/\epsilon) \log^2(n))$ in time $O(ndk + |S|)$.

Theorem

W.c.p., algorithm CORECluster computes a $(1 + \epsilon)$-approximate solution to the $\mu$-similar Bregman $k$-median problem in time $O(ndk + d 2^{(k/\epsilon)c} \log^{k+2} n)$. 
1 Introduction
   - Bregman $k$-median clustering
   - Bregman divergences
   - Mahalanobis distances and $\mu$-similarity
   - Our results

2 Weak coreset construction
   - $\Gamma$-weak $(k, \epsilon)$-coresets
   - Chen's coreset construction for metrics
   - Initial $O(\log k)$-approximation

3 $(1 + \epsilon)$-approximation algorithm
   - Using weak coresets
   - Size bound for $\Gamma$

4 Open problems
Some open problems

- Are there ”strong” coresets for Bregman $k$-median problem?
Some open problems

- Are there "strong" coresets for Bregman $k$-median problem?
- Can we give coresets in the presence of singularities? (i.e., for non-similar Bregman divergences)
Some open problems

- Are there "strong" coresets for Bregman $k$-median problem?
- Can we give coresets in the presence of singularities? (i.e., for non-similar Bregman divergences)
- Are there coresets with size independent of $n$? (compare to [Har-Peled, Kushal, 2005])
Some open problems

- Are there "strong" coresets for Bregman $k$-median problem?
- Can we give coresets in the presence of singularities? (i.e., for non-similar Bregman divergences)
- Are there coresets with size independent of $n$? (compare to [Har-Peled, Kushal, 2005])
- Are there $(1 + \epsilon)$-approximation algorithms for Bregman divergences with singularities in their domain? (factor $O(\log n)$ from [Chaundhuri, McGregor, 2008] is tight!)

What further techniques from Euclidean geometry can be applied to Bregman divergences? (e.g., dimensionality reduction?)

What about non-Bregman, non-metric dissimilarity measures such as Pearson correlation, cosine similarity, etc.?
Some open problems

- Are there "strong" coresets for Bregman $k$-median problem?
- Can we give coresets in the presence of singularities? (i.e., for non-similar Bregman divergences)
- Are there coresets with size independent of $n$? (compare to [Har-Peled, Kushal, 2005])
- Are there $(1 + \epsilon)$-approximation algorithms for Bregman divergences with singularities in their domain? (factor $O(\log n)$ from [Chaundhuri, McGregor, 2008] is tight!)
- What further techniques from Euclidean geometry can be applied to Bregman divergences? (e.g., dimensionality reduction?)
Some open problems

- Are there “strong” coresets for Bregman $k$-median problem?
- Can we give coresets in the presence of singularities? (i.e., for non-similar Bregman divergences)
- Are there coresets with size independent of $n$? (compare to [Har-Peled, Kushal, 2005])
- Are there $(1 + \epsilon)$-approximation algorithms for Bregman divergences with singularities in their domain? (factor $\mathcal{O}(\log n)$ from [Chaundhuri, McGregor, 2008] is tight!)
- What further techniques from Euclidean geometry can be applied to Bregman divergences? (e.g., dimensionality reduction?)
- What about non-Bregman, non-metric dissimilarity measures such as Pearson correlation, cosine similarity, etc.?
Thanks for your attention!