Approximate Nearest Neighbor Searching with Non-Euclidean and Weighted Distances

Ahmed Abdelkader[1], Sunil Arya[2], Guilherme D. da Fonseca[3], David M. Mount[1]

[1]University of Maryland, College Park

[2]The Hong Kong University of Science and Technology

[3]Université Clermont Auvergne, LIMOS, and INRIA Sophia Antipolis

SODA 2019
San Diego, CA
January 6th, 2019

Overview

Nearest-Neighbor Searching

Given n point sites P, preprocess P to answer nearest-neighbor queries:

Given a query point q, which site in P is closest?

Scope

- Fixed dimension: $P \subset \mathbb{R}^d$, where d is constant
- ε-approximate: Return any p', where $\text{dist}(q, p') \leq (1 + \varepsilon)\text{dist}(q, P)$

Performance Goals

- Query time: $O(\log(n/\varepsilon))$
- Storage: $O(n/\varepsilon^{O(d)})$

Overview

Nearest-Neighbor Searching

Given n point sites P, preprocess P to answer nearest-neighbor queries:

Given a query point q, which site in P is closest?

Scope

- Fixed dimension: $P \subset \mathbb{R}^d$, where d is constant
- ε-approximate: Return any p', where $\mathrm{dist}(q, p') \leq (1 + \varepsilon)\mathrm{dist}(q, P)$

Performance Goals

- Query time: $O(\log(n/\varepsilon))$
- Storage: $O(n/\varepsilon^{O(d)}) \longleftarrow$ Focus: Minimize ε-dependencies

Background

ε-ANN queries in logarithmic time $O(\log(n/\varepsilon))$

Euclidean metric:

- [Har-Peled 2001]: $\widetilde{O}(n/\varepsilon^d)$ storage by Approx. Voronoi Diagrams (AVD)
- [Arya et al., 2017b]: $O(n/\varepsilon^{d/2})$ storage by lifting + Macbeath regions

Non-Euclidean distances:

- [Har-Peled and Kumar, 2015]: $O(n/\varepsilon^{O(d^2)})$ storage by min. diagrams
- Various heuristic approaches, e.g., [Cayton, 2008], [Nielsen et al., 2009]

Question: ε-ANN for non-Euclidean distances with just $O(n/\varepsilon^{d/2})$ storage?

Background

ε-ANN queries in logarithmic time $O(\log(n/\varepsilon))$

Euclidean metric:

- [Har-Peled 2001]: $\widetilde{O}(n/\varepsilon^d)$ storage by Approx. Voronoi Diagrams (AVD)
- [Arya et al., 2017b]: $O(n/\varepsilon^{d/2})$ storage by lifting + Macbeath regions

Non-Euclidean distances:

- [Har-Peled and Kumar, 2015]: $O(n/\varepsilon^{O(d^2)})$ storage by min. diagrams
- Various heuristic approaches, e.g., [Cayton, 2008], [Nielsen et al., 2009]

Question: ε-ANN for non-Euclidean distances with just $O(n/\varepsilon^{d/2})$ storage?

Background

ε-ANN queries in logarithmic time $O(\log(n/\varepsilon))$

Euclidean metric:

- [Har-Peled 2001]: $\widetilde{O}(n/\varepsilon^d)$ storage by Approx. Voronoi Diagrams (AVD)
- [Arya et al., 2017b]: $O(n/\varepsilon^{d/2})$ storage by lifting + Macbeath regions

Non-Euclidean distances:

- [Har-Peled and Kumar, 2015]: $O(n/\varepsilon^{O(d^2)})$ storage by min. diagrams
- Various heuristic approaches, e.g., [Cayton, 2008], [Nielsen et al., 2009]

Question: ε-ANN for non-Euclidean distances with just $O(n/\varepsilon^{d/2})$ storage?

Main Results

Main Result

A general approach for answering ε-ANN queries with:

- Query time: $O(\log \frac{n}{\varepsilon})$
- Storage: $O((n/\varepsilon^{d/2}) \log \frac{1}{\varepsilon})$

Demonstrated on the following distance functions (defined later):

- Minkowski distance
- Minkowski distance & multiplicative weights
- Mahalanobis distance*
- Scaling distance functions*
- Bregman divergences*

* Additional admissibility conditions are needed

Main Results

Main Result

A general approach for answering ε-ANN queries with:

- Query time: $O(\log \frac{n}{\varepsilon})$
- Storage: $O((n/\varepsilon^{d/2}) \log \frac{1}{\varepsilon})$ $\longleftarrow$ Worse by a factor of only $\log \frac{1}{\varepsilon}$

Demonstrated on the following distance functions (defined later):

- Minkowski distance
- Minkowski distance & multiplicative weights
- Mahalanobis distance*
- Scaling distance functions*
- Bregman divergences*

* Additional admissibility conditions are needed

Main Results

Main Result

A general approach for answering ε-ANN queries with:

- Query time: $O(\log \frac{n}{\varepsilon})$
- Storage: $O((n/\varepsilon^{d/2}) \log \frac{1}{\varepsilon})$ $\longleftarrow$ Worse by a factor of only $\log \frac{1}{\varepsilon}$

Demonstrated on the following distance functions (defined later):

- Minkowski distance
- Minkowski distance & multiplicative weights
- Mahalanobis distance*
- Scaling distance functions*
- Bregman divergences*

* Additional admissibility conditions are needed

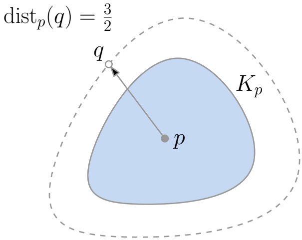
Distance Functions - Scaling Distances

- Minkowski Distance: $\|q-p\|_k = (\sum_{i=1}^{d} |p_i - q_i|^k)^{\frac{1}{k}}$.
  Our results apply for any real constant $k > 1$
- Multiplicative Weights: Each p has weight $w_p > 0$ and
  $\text{dist}_p(q) = w_p\|q-p\|_k$
- Mahalanobis Distance: Each p has a positive-definite matrix M_p and
  $\text{dist}_p(q) = \sqrt{(p-q)^\intercal M_p (p-q)}$
- Scaling Distance Functions: Each p has a closed convex body K_p containing the origin and $\text{dist}_p(q)$ is the smallest r such that
  $q \in p + r \cdot K_p$

Scaling distance functions generalize all the others

* Metric balls K_p are smooth and fat

** Triangle inequality may not hold



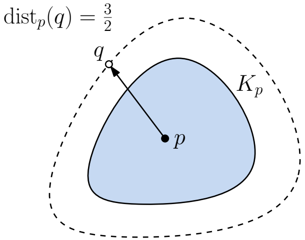
Distance Functions - Scaling Distances

- Minkowski Distance: $\|q-p\|_k = (\sum_{i=1}^{d} |p_i - q_i|^k)^{\frac{1}{k}}$.
  Our results apply for any real constant $k > 1$
- Multiplicative Weights: Each p has weight $w_p > 0$ and $\mathrm{dist}_p(q) = w_p\|q-p\|_k$
- Mahalanobis Distance: Each p has a positive-definite matrix M_p and $\mathrm{dist}_p(q) = \sqrt{(p-q)^\intercal M_p (p-q)}$
- Scaling Distance Functions: Each p has a closed convex body K_p containing the origin and $\mathrm{dist}_p(q)$ is the smallest r such that $q \in p + r \cdot K_p$

Scaling distance functions generalize all the others

* Metric balls K_p are smooth and fat

** Triangle inequality may not hold



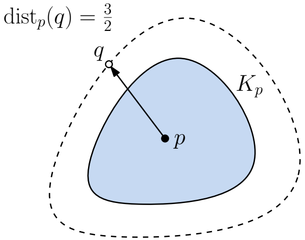
Distance Functions - Scaling Distances

- Minkowski Distance: $\|q-p\|_k = (\sum_{i=1}^{d} |p_i - q_i|^k)^{\frac{1}{k}}$.
  Our results apply for any real constant $k > 1$
- Multiplicative Weights: Each p has weight $w_p > 0$ and $\mathrm{dist}_p(q) = w_p\|q-p\|_k$
- Mahalanobis Distance: Each p has a positive-definite matrix M_p and $\mathrm{dist}_p(q) = \sqrt{(p-q)^\intercal M_p (p-q)}$
- Scaling Distance Functions: Each p has a closed convex body K_p containing the origin and $\mathrm{dist}_p(q)$ is the smallest r such that $q \in p + r \cdot K_p$

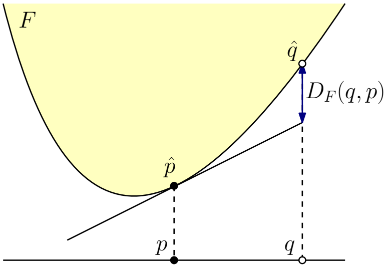
Scaling distance functions generalize all the others

* Metric balls K_p are smooth and fat

** Triangle inequality may not hold



Distance Functions - Bregman Divergences



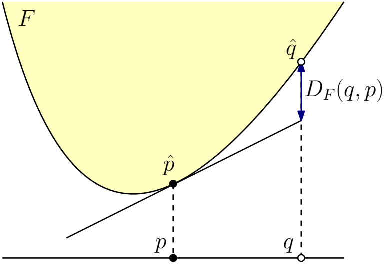
- Bregman Divergence: Given an open convex domain $\mathcal{X} \subseteq \mathbb{R}^d$, a strictly convex and differentiable real-valued function F on $\mathcal{X}$, and $q, p \in \mathcal{X}$, the *Bregman divergence* of q from p is

$$D_F(q,p) \;=\; F(q) - (F(p) + \nabla F(p) \cdot (q - p)).$$

where ∇F denotes the gradient of F and "$\cdot$" is the standard dot product.

* D_F is typically asymmetric

Distance Functions - Bregman Divergences



- Bregman Divergence: Given an open convex domain $\mathcal{X} \subseteq \mathbb{R}^d$, a strictly convex and differentiable real-valued function F on $\mathcal{X}$, and $q, p \in \mathcal{X}$, the *Bregman divergence* of q from p is

$$D_F(q,p) \;=\; F(q) - (F(p) + \nabla F(p) \cdot (q-p)).$$

where ∇F denotes the gradient of F and "$\cdot$" is the standard dot product.

* D_F is typically asymmetric

Distance Functions - Bregman Divergences

Worst-case bounds on ε-ANN for Bregman require additional assumptions.

- D_F is μ-asymmetric if for all $p, q \in \mathcal{X}$, $D_F(q, p) \leq \mu D_F(p, q)$
- D_F is μ-similar if for all $p, q \in \mathcal{X}$, $\|q - p\|^2 \leq D_F(q, p) \leq \mu \|q - p\|^2$

Prior results:

- [Abdullah et al., 2012]: Answer ε-ANN queries for decomposable Bregman divergences in spaces of constant dimension, with dependence on the defectiveness μ.
- [Abdullah and Venkatasubramanian, 2015]: cell probe lower bounds in terms of asymmetry

Our results hold under an intermediate condition, τ-admissibility, which states that the growth rate is polynomial w.r.t. Euclidean

Preliminaries - Minimization Diagrams

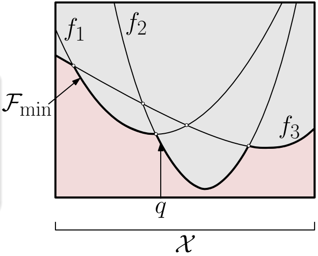
We apply a reduction from ε-ANN to approximate vertical ray shooting by [Har-Peled and Kumar, 2015]

Minimization Diagram

Given a convex domain $\mathcal{X} \subseteq \mathbb{R}^d$ and a set of functions $\mathcal{F} = \{f_1, \ldots, f_m\}$, where $f_i : \mathcal{X} \to \mathbb{R}^+$, let $\mathcal{F}_{\min}(x) = \min_{1 \le i \le m} f_i(x)$

Approximate Vertical Ray Shooting (ε-AVR)

Set f_i to the distance function to the ith site.
ε-ANN is equivalent to computing any index i such that $f_i(q) \le (1 + \varepsilon)\mathcal{F}_{\min}(q)$



Preliminaries - Minimization Diagrams

We apply a reduction from ε-ANN to approximate vertical ray shooting by [Har-Peled and Kumar, 2015]

Minimization Diagram

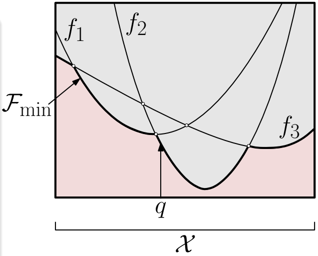
Given a convex domain $\mathcal{X} \subseteq \mathbb{R}^d$ and a set of functions $\mathcal{F} = \{f_1, \ldots, f_m\}$, where $f_i : \mathcal{X} \to \mathbb{R}^+$, let $\mathcal{F}_{\min}(x) = \min_{1 \le i \le m} f_i(x)$

Answering ε-AVR Queries [Arya et al., 2017a]

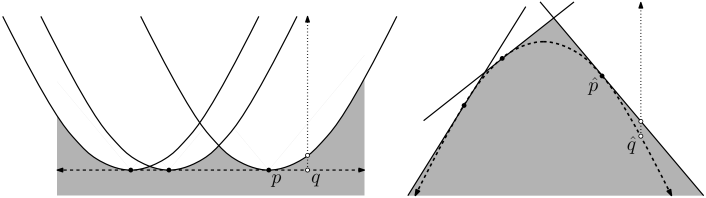
Query time: $O\left(\log \frac{1}{\varepsilon}\right)$ and Storage: $O\left(\frac{1}{\varepsilon^{d/2}}\right)$

When all f_i are concave and for all $x \in \mathcal{X}$

$$f_i(x) \in [0, 1] \quad \text{and} \quad \|\nabla f_i(x)\| \le 1$$



Preliminaries - Linearization (Lifting Transformation)



Consider ε-ANN in the squared Euclidean distance. The minimization diagram is the non-convex lower envelope of paraboloids (left figure)

These paraboloids share a common quadratic term $\sum_i x_i^2$

Subtract off this common term to linearize these paraboloids (right figure)
* Also called the lifting transformation

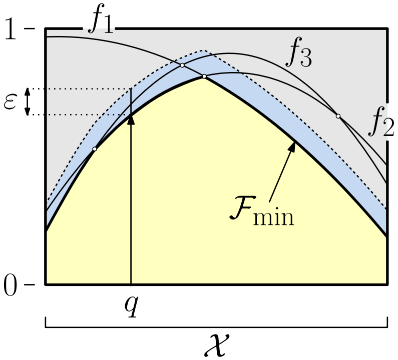
The minimization diagram maps to the lower envelope of linear functions, which is convex

Approximate ray-shooting in convex bodies can be applied [Arya et al., 2017a]

The Struggles and Hopes of Lifting

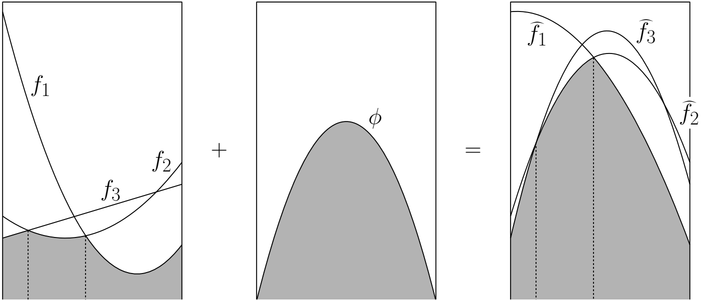
Unfortunately, linearization only works in this restricted context

But, [Arya et al., 2017a] do not require a piecewise linear envelope!
Only that it bounds a convex region



Convexification

Intuition: Subtract off a function with such a small second derivative (generally Hessian) that all the distance functions become concave.



The minimization diagram of concave functions bounds a convex set!

Convexification - The Details

- Distance functions: $\mathcal{F} = \{f_1, \ldots, f_m\}$
- Restricted domain: Euclidean ball B with center p and radius r
- Smoothness: Λ^+ is an upper bound on the largest eigenvalue of $\nabla^2 f_i(x)$ for any i and any $x \in B$
- Convexifying function:

$$\phi(x) = \frac{\Lambda^+}{2}(r^2 - \|x-p\|^2) = \frac{\Lambda^+}{2}\left(r^2 - \sum_{j=1}^{d}(x_j - p_j)^2\right)$$

- Convexified distance functions:

$$\widehat{f_i}(x) = f_i(x) + \phi(x), \quad \text{for } 1 \le i \le m, \text{ and}$$
$$\widehat{F}_{\min}(x) = \min_{1\le i\le m} \widehat{f_i}(x) = \mathcal{F}_{\min}(x) + \phi(x).$$

Convexification - Approximation

Convexification has been widely applied in the context of non-linear optimization, e.g., [Androulakis et al., 1995], [Bertsekas, 1979]

How to make it approximation-preserving?

- Smooth, centered, fat: Approximation errors grow with the eigenvalues of the Hessians of the distance functions

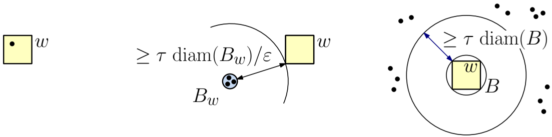
  Encapsulated as a property of distance functions, τ-admissibility

  E.g., Metric balls of scaling distance functions must be smooth, well-centered, and fat
- Separation and locality: Error distortion increases with variation of distance values within the ball B.
  We use a quadtree subdivision of space to establish separation properties and apply convexification to each leaf cell of the subdivision

Answering ε-ANN Queries

For τ-admissible Bregman divergences

Apply a quadtree subdivision so that for each leaf cell w either:



- NN is a single point within w: trivial
- Inner cluster: NN lies in a ball B_w and $\mathrm{dist}(B_w, w) = \Omega(\tau\,\mathrm{diam}(B_w)/\varepsilon)$
  - Any site within B_w can be used
- Outer cluster: NN is at distance $\Omega(\tau \mathrm{diam}(w))$ from w
  - Apply convexification to the distance functions of these sites.
  - Answer query by approximate vertical ray-shooting ε-AVR.

Concluding Remarks

- We have shown that best-known bounds on ε-ANN searching for the Euclidean metric can be almost* generalized to a variety of non-Euclidean distances: Minkowski metrics, multiplicatively weighted sites, Mahalanobis, convex scaling distances, Bregman divergences
- Our results arise from an application of convexification, which transforms the minimization diagram of smooth functions into convex form
- Open Questions:
  - Generalizations to non-smooth non-Euclidean distance functions? Target: $O(n/\varepsilon^d)$ storage
  - Is convexification applicable to other problems? (e.g., computing the diameter of a point set in different metrics)
  - *Eliminating the pesky $O(\log \frac{1}{\varepsilon})$ slop factor?

Concluding Remarks

- We have shown that best-known bounds on ε-ANN searching for the Euclidean metric can be almost* generalized to a variety of non-Euclidean distances: Minkowski metrics, multiplicatively weighted sites, Mahalanobis, convex scaling distances, Bregman divergences
- Our results arise from an application of convexification, which transforms the minimization diagram of smooth functions into convex form
- Open Questions:
  - Generalizations to non-smooth non-Euclidean distance functions? Target: $O(n/\varepsilon^d)$ storage
  - Is convexification applicable to other problems? (e.g., computing the diameter of a point set in different metrics)
  - *Eliminating the pesky $O(\log \frac{1}{\varepsilon})$ slop factor?

Thank you!

References I

A. Abdullah, J. Moeller, and S. Venkatasubramanian.
Approximate Bregman near neighbors in sublinear time: Beyond the triangle inequality.
In *Proc. 28th Annu. Sympos. Comput. Geom.*, pages 31–40, 2012.

A. Abdullah and S. Venkatasubramanian.
A directed isoperimetric inequality with application to Bregman near neighbor lower bounds.
In *Proc. 47th Annu. ACM Sympos. Theory Comput.*, pages 509–518, 2015.

I. P. Androulakis, C. D. Maranas, and C. A. Floudas.
αBB: A global optimization method for general constrained nonconvex problems.
J. Global Optim., 7:337–363, 1995.

References II

S. Arya, G. D. da Fonseca, and D. M. Mount.
Optimal approximate polytope membership.
In *Proc. 28th Annu. ACM-SIAM Sympos. Discrete Algorithms*, pages 270–288, 2017.

Arya, S., da Fonseca, G. D., and Mount, D. M. (2017).
Near-optimal ε-kernel construction and related problems.
In *Proc. 33rd Internat. Sympos. Comput. Geom.*, pages 10:1–15.

D. P. Bertsekas.
Convexification procedures and decomposition methods for nonconvex optimization problems 1.
J. Optim. Theory Appl., 29:169–197, 1979.

Cayton, L. (2008).
Fast nearest neighbor retrieval for Bregman divergences.
In *Proc. 25th Internat. Conf. Machine Learning*, pages 112–119.

References III

Har-Peled, S. (2001).
A replacement for Voronoi diagrams of near linear size.
In *Proc. 42nd Annu. IEEE Sympos. Found. Comput. Sci.*, pages 94–103.

Har-Peled, S. and Kumar, N. (2015).
Approximating minimization diagrams and generalized proximity search.
SIAM J. Comput., 44:944–974.

Nielsen, F., Piro, P., and Barlaud, M. (2009).
Bregman vantage point trees for efficient nearest neighbor queries.
2009 IEEE Internat. Conf. on Multimedia and Expo, pages 878–881.