<!-- Slide number: 1 -->
# Streaming Algorithms for Coresets
Pankaj K. Agarwal

![cs_logo](Picture4.jpg)
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 2 -->
# Geometric Summaries: Coresets
Tool for geometric approximation algorithms
S: Set of input points
m: Objective function (minimization)
C ⊆ S: coreset of S (wrt m)
 (1-ε) m(S) ≤ m(C) ≤ (1+e)m(S)
Example: Random sampling
X=(S,R), R ⊆ 2S: Set system
C ⊆S: e-approximation if ∀ r ∈R

Size: 1/e2
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 3 -->
# e-Approximation
Coreset in a combinatorial/statistical sense
E.g. approximate range searching
Approximates the distribution

Not a coreset in a metric/geometric sense
diam(C) doesn’t approximate diam(S)
Best-fit circle for C doesn’t approximate that of S

What about other sampling schemes?

![approx-examples.eps](Picture3.jpg)
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 4 -->
# ε-kernel
S: Set of points in Rd

Directional width: For u ∈ Sd−1,

ε-kernel: Q ⊆ S is an ε-kernel of S if

Coreset for many geometric optimization problems

![sphere-dir.eps](Picture8.jpg)

![pont-extent.eps](Picture7.jpg)
Size of e-kernel? How fast can they be computed
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 5 -->
# ε-kernel
Theorem A: [AHV, Ch, YAPV] S ⊆ Rd, ε > 0. An ε-kernel of S of size 1/ε(d−1)/2 can be computed in time n + 1/εd−3/2.

n + 1/εO(d)-time approximation algorithms for computing
 extent measures: diameter, width
smallest enclosing convex shapes
ball, ellipse
rectangle, simplex,…..
Coreset for mobile data

Coresets in streaming model?
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 6 -->
# Kernels in Streaming Model
Two crucial properties of e-kernels

R ⊆ Q :(e/2)-kernel of Q, Q ⊆ P: (e/2)-kernel of P
    R: e-kernel of P

C1 :e-kernel of A1, 	C2 :e-kernel of A2
         C1 U C2: e-kernel of A1 UA2

Merge and Reduce technique!

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 7 -->
# Streaming Algorithms for ε-kernel
S: Stream of points in R2; points arrive one-by-one
Maintain the ε-kernel using 1/εO(d) space                           [A., Har-Peled, Varadarajan; Chan; A., Yu; Zarrabi-Zadeh]
| d | Space | Update time | Ref |
| --- | --- | --- | --- |
| Rd | (1/ε(d-1)/2)logd n | 1/ε3(d-1)/2 | [A., Har-Peled,Varadarajan] |
| Rd | 1/εd-3/2logd (1/ε) | 1/εd-3/2 | [Chan] |
| R2 | 1/ε1/2 | log (1/ε) | [A., Yu] |
| Rd | 1/ε(d-1)/2 | 1/ε(d-3)/2 | [Zarrabi-Zadeh] |
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 8 -->
# Streaming Algorithm for ε-kernel
Problem is easy as long as S is fat (B ⊆ CH(S) ⊆aB)
B: box
Keep track of NN of each grid point
 Update time: log(1/ε)

![nn.eps](Picture4.jpg)

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 9 -->
# Streaming Algorithm for ε-kernel
Maintaining anchor points: epochs and subepochs

o: first point in the stream
xi: first point in the ith epoch
 yj : first point in the jth subepoch of the current epoch
 xi starts a new epoch if
 yj starts a new subepoch if

![overall.eps](Picture6.jpg)

![](Picture4.jpg)
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 10 -->
# Streaming Algorithm for ε-kernel
Maintain ε-kernels for log(1/ε) epochs
Maintain ε-kernels for log(1/ε) subepochs within each epoch
Points in earlier epochs are too close to o
 Points in earlier subepochs are too close to the line oxi
Size: (1/√ε) log2(1/ε)
Prune coresets from older epochs and subepochs
Size: (1/√ε)

Can be extended to higher dimensions
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 11 -->
# Streaming in High Dimensions
d is large and part of the input
 For ε = d1/3, size of ε-kernel is W(exp(d1/3))
 Coresets of size (d/ε)O(1) for some problems
Minimum enclosing ball
Minimum enclosing ellipsoid

Are there streaming algorithms that use (d/ε)O(1) space to maintain coresets?
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 12 -->
# Streaming in High Dimensions
Minimum enclosing ball (MEB) [Chan Zarrabi-Zadeh]
Maintains a single ball
Space Complexity: O(d), Update Time: O(d)
Quality: 1.5 approximation
 Any streaming algorithm that maintains exactly one ball and nothing more has an approximation ratio at least 1.203.

Diameter [Indyk]
c-approximation for c > √2
Size, update time:

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 13 -->
# Our Results
| Problem | Upper Bound | Lower Bound ( holds for any randomized streaming algorithm) |
| --- | --- | --- |
| α-MEB
 | α = (1+ √3)/2 +ε
Space: O(d/ε3 log 1/ε)
Update: O(d/ε5) | α < ((1+ √2)/2) (1-2/d1/3) requires Ω(exp(d)) space. |
| α-Coreset | α = √2+ε
Space: O(d/ε3 log 1/ε)
Update: O(d/ε2 log 1/ε) | α < √2 (1-2/d1/3) requires Ω(exp(d)) space.
 |
| α-Diameter | α = √2+ε
Space: O(d/ε3 log 1/ε)
Update: O(d/ε3 log 1/ε) | α < √2 (1-2/d1/3) requires Ω(exp(d)) space.
 |
| α-FN(x) | α = √2+ε
Space: O(d/ε3 log 1/ε)
Update: O(d/ε3 log 1/ε) | α < √2 (1-2/d1/3) requires Ω(exp(d)) space.
 |
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 14 -->
# Upper Bounds (Blurred ball cover)
S: set of points
K = {K1, . . . ,Ku}, Ki ⊆ S,
|Ki| ≈ 1/ε
Bi = MEB(Ki), ri = r(Bi)
K : ε-blurred ball cover if
 ri+1 ≥ (1 + ε2)ri
∀j ≤ i, Kj ⊆ Bi
S ⊂ U (1 + ε)Bi

B3

r3

r2
B2

r1

B1

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 15 -->
# Updating Blurred Ball Cover
p

B*
p
Update (p)
If p in U (1+ ε)Bi
Do nothing
Else
K*, B* = APPROX-MEB(p, U Ki)
Delete Ki from K if ri < ε r*/4
Add K* to K

ri+1 ≥ (1 + ε2)ri
∀j ≤ i, Kj ⊆ Bi
S ⊂ U (1 + ε)Bi
Trivially true!

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 16 -->
# Blurred Ball Cover
Space
ru ≤ 4.r1/ε ⇒ u ≈ 1/ε2, ∑i|Ki| ≈1/ε3
Update Time
dominated by computing APPROX-MEB [Badoiu Clarkson]
Single Point update: O(d/ε5)
Batched Update:  O(d/ε2 log 1/ε)

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 17 -->
# Minimum Enclosing Ball

Minimum Enclosing Ball
Return B = MEB(B1, . . . ,Bu)
S ⊂ (1 + ε/2)B
r(B)/r(MEB(S)) ≤(1+√3)/2 + ε

UKi  is a (√2+ ε )-coreset of MEB

![](Picture3.jpg)
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 18 -->
# Lower Bounds: INDEX problem
Alice has a string ∑ ∈{0,1}k
Bob has an index i ∈[1..k]
Alice send a message m to Bob.
Bob uses m to compute the ith bit of ∑ with probability 2/3
Communication complexity is |m| = Ω(k) bits.
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 19 -->
# Lower Bounds (Diameter)
Lemma: There is a centrally symmetric set K ⊂ Sd−1 s.t.
(i) |K| = exp(d1/3),
(ii) p, q ∈ K, p ≠ -q ⇒ ||pq|| ≈ √2
X = K ∩ {xd ≥ 0}
Define φ : [1 : k] → X, k ≈ exp(d1/3).
A: Randomized streaming algorithm for √2-Diameter.
Alice
∀j < k, If ∑j = 1, add φ(j) to A
Communicate workspace of A to Bob
Bob:  For index i, add -φ(i) to A
Communication Complexity = Size(A)
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 20 -->
# Lower Bounds (Diameter)
Diameter of points added by Alice is at most √2
If ∑i = 1
Alice adds φ(i)  and Bob adds - φ(i) ⇒ Diameter = 2.
If ∑i = 0
No pair of points are antipodal ⇒ Diameter ≈ √2
A can distinguish the two cases
Communication Compl. = Size (A) = Ω(k) = Ω(exp(d1/3))
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 21 -->
# Lower Bounds (Summary)
Ω(exp(d1/3))  bits required by any streaming algorithm that maintains
α-MEB for α < (1+ √2)/2 (1-2/d1/3)
α-Coreset for α < √2 (1-2/d1/3)
α-Diameter for α < √2 (1-2/d1/3)
α-Width for α < d1/3

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 22 -->
# Coresets for Dynamic Streams
insertions and deletions
Assume a point was inserted before being deleted

[Fahrling-Sohler’05]: coresets for k-medians
 set C of k points s.t. m(S,C) = ∑p ∈S d(p,C) is minimized
Q ⊆S, w: Q       R: e-coreset of S if
(1-e)m(S,C) ≤m(Q,C) ≤(1+e)m(Q,C)
Extends the approach by [Indyk 04]
k2poly(1/e, logn) space, kpoly(1/e, logn) update time
 size of the coreset: (k/ed)logn
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 23 -->
# Coresets vs Linear Sketches
Linear sketches
Handle insertion/deletions
Maintain norms, earth mover distance
Coresets
Subset of input points
Handle insertions
Maintain metric, extent properties
Relationship between the two is not clear
Linear sketch for the width of a point set?
Coreset for earth mover distance?

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 24 -->
# Open Problems
Coreset for
minimum enclosing ellipsoid in streaming model
minimum enclosing ball under a dynamic stream
preserving structural/topological properties

Coresets in distributed settings
Multiple streams
Minimize communication complexity

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 25 -->
# Problem Statement
S : set of n points
Minimum Enclosing Ball
α-MEB : r  ≤ α . rOPT

rOPT

r

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 26 -->
# Problem Statement

S : set of n points in Rd
Minimum Enclosing Ball
α-MEB : r  ≤ α . rOPT
α-Coreset: C ⊆ S,  S ⊆ α . MEB(C)

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 27 -->
# Problem Statement
S : set of n points in Rd
Diameter
 α-Diameter = (r,s)
||pq|| ≤ α. ||rs||
Farthest neighbor
α-FN(x) = p, ||xr|| ≤ α. ||xp||

p

r

s
q

x
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 28 -->
# Problem Statement
Streaming Algorithms for
α-MEB, α-Coreset and α-Diameter
Workspace: poly(d), sublinear(n)
Update Time: poly(d)
Insertion-only data structure to answer
α-FN (x) queries
Size:  poly(d), sublinear(n)
Query and update time: poly(d)
 α-Diameter trivially reduces to α-FN (x)
Lower Bound for α-Diameter
Upper Bound for α-FN (x)

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 29 -->
# Blurred Ball Cover: Applications
UKi  is a (√2+ ε )-coreset of MEB
  Bu = MEB(UKi) → ru < rOPT
Uki contains (√2+ ε )-FN(cu) → UKi  is a (√2+ ε )-coreset

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 30 -->
# MEB [Zarrabi-Zadeh Chan]
Maintains a ball B.
If p ∈ B, ignore p
Else B= MEB(p, B)
B is 1.5-MEB.
p

p

B
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 31 -->
# Diameter
Data Structure [Indyk’03]
Space Complexity:
Update Time:
Quality: c > √2

![](Picture4.jpg)

![](Picture2.jpg)
WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 32 -->
# Updating Blurred Ball Cover
Update (p)
If p in U (1+ ε)Bi
Do nothing
Else
K*, B* = APPROX-MEB(p, Ki)
Delete Ki from K if ri < ε r*/4
Add K* to K

p

B*
p

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 33 -->
# Blurred Ball Cover: Applications
(√2+ ε )-FN(x)
Return the farthest point to x in UKi
Key observation [Badoiu Har-Peled Indyk]
Any half-space that contains c(MEB(S)) also contains p ∈  ∂ MEB(S)

C

WAPMDS, IIT Kanpur
12/19/2009

<!-- Slide number: 34 -->
# Blurred Ball Cover: Applications
(√2+ ε )-FN(x)
Return the farthest point  to x in UKi
Proof:
∃Ki , r ∈ (1 + ε)Bi
||xr|| < ||xci|| + ||cir||
||xp|| > (||xci||2 + ||cip||2)1/2
||xr||/||xp|| < √2 + ε
p

r
Bi

x
WAPMDS, IIT Kanpur
12/19/2009