Streaming Data

- Problems defined over points $P=\{p_1,\ldots,p_n\}$
- The algorithm sees $p_1$, then $p_2$, then $p_3$, …
- Key fact: it has limited storage
  - Can store only $s<<n$ points
  - Can store only $s<<n$ bits (need to assume finite precision)

$p_1\ldots p_2\ldots p_3\ldots p_4\ldots p_5\ldots p_6\ldots p_7\ldots$
Example - diameter
Problems

• Diameter
• Minimum enclosing ball
• $l_2$ norm of a high-dimensional vector
Diameter in $l_{\infty}^d$

- Assume we measure distances according to the $l_{\infty}$ norm
- What can we do?
Diameter in $l_\infty$, ctd.

• From previous lecture we know that
  \[ \text{Diam}_\infty(P) = \max_{i=1 \ldots d'} \left[ \max_{p \in P} p_i - \min_{p \in P} p_i \right] \]

• Can maintain max/min in constant space

• Total space = $O(d')$

• What about $l_1$?
Diameter in $l_1$

• Let $f : l_1^d \rightarrow l_\infty^{2^d}$ be an isometric embedding

• We will maintain $Diam_\infty(f(P))$
  – For each point $p$, we compute $f(p)$ and feed it to the previous algorithm
  – Return the pair $p,q$ that maximizes $||f(p)-f(q)||_\infty$

• This gives $O(2^d)$ space for $l_1^d$

• What about $l_2$?
$(1+\varepsilon)$-embedding of $l^d_2$ into $l^d_\infty$

- $(1+\varepsilon)$-embedding:
  - No expansion
  - Contraction by at most $1+\varepsilon$

- We will achieve $d'$ equal to $O(1/\varepsilon)^{(d-1)/2}$
Consider $d=2$

- For embedding into $l_1$ we used
  \[ f(x,y)=[x+y,x-y,-x+y,-x-y] \]
  - Since $f$ linear, we have $||f(p)-f(q)||=||f(p-q)||$
  - $||(x,y)||_1 = |x|+|y| = \max[ x+y, x-y, -x+y, -x-y ]$
Embedding of $l_2$ into $l_\infty$

- Again, use projections
  - Onto unit ($l_2$) vectors $v_1 \ldots v_k$
  - Requirement: vectors are “densely” spaced:
    - for any $u$ there is $v_i$ such that $u^*v_i \geq ||u||_2 / (1+\varepsilon)$
  - This implies $(1+\varepsilon)$ distortion
    - No expansion
    - Contraction by at most $1+\varepsilon$

- How big is $k$?
  - Can assume $||u||_2=1$
Lemma

- Consider two unit vectors \( u \) and \( v \), such that the \( \text{angle}(u,v) = \alpha \). Then \( u^*v \geq 1 - O(\alpha^2) \)
- Proof: \( u^*v = \cos(\alpha) = 1 - \Theta(\alpha^2) \)
- Therefore, suffices to use \( \frac{2\pi}{\varepsilon^{1/2}} \) vectors to get distortion \( 1 + O(\varepsilon) \)
Higher Dimensions

• For $d=2$ we get $d' = O(1/\varepsilon^{1/2})$
• For any $d$ we get $d' = O_d(1/\varepsilon)^{(d-1)/2}$
  – Can “cover” a unit sphere in $\mathbb{R}^d$ with $O_d(1/\alpha)^{d-1}$ vectors so that any $v$ has angle $< \alpha$ with at least one of the vectors
  – The remainder is the same
Covering vs Packing

• Assume we want to **cover** the sphere using disks of radius $\alpha$

• This can be achieved by **packing**, as many as possible, disks of radius $\alpha/2$

• How many disks can be pack?
  – Each disk has volume $\Theta_d(\alpha/2)^{d-1}$ times smaller than the volume of the sphere
  – Inverse of that gives the packing/covering bound
Diameter in $l_2$

- Let $f : l_2^d \rightarrow l_\infty^{d'}$, $d' = O(1/\varepsilon)^{(d-1)/2}$, be a $(1+\varepsilon)$-distortion embedding
- Apply the same algorithm as before
- Space: $O(1/\varepsilon)^{(d-1)/2}$
Minimum Enclosing Ball

• Problem: given $P=\{p_1, \ldots, p_n\}$, find center $o$ and radius $r>0$ such that
  – $P \subseteq B(o,r)$
  – $r$ is as small as possible

• Solve the problem in $l_\infty$

• Generalize to $l_1$ and $l_2$ via embeddings
MEB in $l_\infty$

- Let $C$ be the hyper-rectangle defined by max/min in every dimension
- Easy to see that min radius ball $B(o,r)$ is a min size hypercube that contains $C$
- Min radius = min hypercube side length/2
- How to solve it in $l_2$?
MEB in $l_2$

- Let $f: l_2^d \rightarrow l_\infty^{d'}$ be an “almost” isometric embedding

- Algorithm I:
  - Maintain $MEB_{\infty} B'(o',r)$ of $f(p_1)\ldots f(p_n)$
  - Compute $o$ such that $f(o) = o'$
  - Report $o$
Problem

• There might be NO $o$ such that $f(o)=o'$

• The problem is that $f$ is into, not onto
The Correct Version

• Algorithm II:
  – Maintain the min/max points \( f(p_1) \ldots f(p_{2d'}) \), two points per dimension
  – Compute MEB \( B(o,r) \) of \( p_1 \ldots p_{2d'} \)
  – Report \( o \)
Correctness

MEB radius for $P$

$= \min r \ s.t. \ \exists o \ P \subseteq B(o,r) \ (by \ definition)$

$\leq \min r' \ s.t. \ \exists f(P) \subseteq B(f(o),r') \ (no \ expansion \ of \ f)$

$= \min r' \ s.t. \ \exists \{f(p_1)\ldots f(p_{2d'})\} \subseteq B(f(o),r')$

(a set of points $f(P)$ is contained in a hypercube iff the extreme points of $f(P)$ are contained in that hypercube)

$\leq (1+\varepsilon) \min r \ s.t. \ \exists \{p_1\ldots p_{2d'}\} \subseteq B(o,r)$

(contraction by at most $1+\varepsilon$)

$= (1+\varepsilon) \ MEB \ radius \ for \ \{p_1\ldots p_{2d'}\}$
Digression: Core Sets

• In the previous slide we use the fact that in $l_\infty$, for any set $P$ of points, there is a subset $P'$ of $P$, $|P'|=2d'$, such that

$$\text{MEB}(P') = \text{MEB}(P)$$

• $P'$ is called a “core-set” for the MEB of $P$ in $l_\infty$

• For more on core-sets, see the web page by Sariel Har-Peled