# CSC2414 - Metric Embeddings*
# Lecture 4: Big Core Theorem.

Notes taken by Hamed Hatami

**Summary:** In this tutorial we prove that a lower bound for the size of a core in a Negative type metric.

## 1 Introduction

In this tutorial we prove an asymptotically sharp lower bound for the size of a core in a Negative type metric. This result which improves the bound of [ARV04] is obtained by James Lee [Lee05]. As we saw in the lectures assuming the contrary of the structure theorem leads to the existence of certain structure in the metric, namely a core:

**Definition 1.1. Matching covers and cores:** For a finite set $X$ let $\mathcal{M}(X)$ denote the set of partial matchings on $X$. Given a subset $Y \subseteq X$ we say that $Y$ is $(\sigma, \delta, \ell)$-*matching covered* by $X$, if there exists a map

$$M : S^{d-1} \to \mathcal{M}(X)$$

such that

1. For every $u \in S^{d-1}$ and $(x, y) \in M(u)$, we have $\langle x - y, u \rangle \geq \sigma/\sqrt{d}$ and $\|x - y\| \leq \ell$.
2. For every $y \in Y$,
$$\Pr[\exists x \in X : (x, y) \in M(u)] \geq \delta.$$

$M$ is called the matching cover of $Y$. If $Y$ is $(\sigma, \delta, \ell)$-matching covers itself, we call it a $(\sigma, \delta, \ell)$-*core*.

Our goal now is to prove the main theorem of this tutorial.

**Theorem 1.2.** *Suppose $C \subseteq \mathbb{R}^d$ is a $(\sigma, \delta, \ell)$-core for some $\sigma, \delta \in (0, 1/2]$. Suppose furthermore that $d(x, y) = \|x - y\|^2$ is a metric on $C$. Then*

$$|C| \geq \exp\left(\Omega\left(\frac{\sigma^6}{\ell^4 \log^2(1/\delta)}\right)\right).$$

* Lecture Notes for a course given by Avner Magen, Dept. of Computer Sciecne, University of Toronto.

This theorem will finish the proof of the structure theorem as for the parameters that we get from there Theorem 1.2 implies that $|C| > n$, a contradiction.

**Definition 1.3.** We say that a point $x \in \mathbb{R}^d$ is $(\sigma, \delta, \ell)$-*covered* by a set $C \subseteq \mathbb{R}^d$ if the following condition is satisfied:

$$\Pr_{u \in S^{d-1}}[\exists y \in C \cap B(x, \ell) : \langle x - y, u \rangle \geq \frac{\sigma}{\sqrt{d}}] \geq \delta.$$

We also say that a set of points $S \subseteq \mathbb{R}^d$ is $(\sigma, \delta, \ell)$-*covered* by $C$ if every $x \in S$ is $(\sigma, \delta, \ell)$-covered by $C$.

The following lemma is a well-known fact.

**Lemma 1.4.** *If $z \in \mathbb{R}^d$, then*

$$\Pr\left[\langle z, u \rangle \geq \frac{\sigma}{\sqrt{d}}\right] \leq \exp\left(\frac{-\sigma^2}{2\|z\|^2}\right).$$

We can use this lemma to prove a lower bound for the size of a cover.

**Lemma 1.5.** *If $x$ is $(\sigma, \delta, \ell)$-covered by a set $C$, then*

$$|C| \geq \delta \exp\left(\frac{\sigma^2}{2\ell^2}\right).$$

**Exercise 1.6.** Prove Lemma 1.5.

A key step in the proof our main theorem is to attach a chain of covers together. The following lemma is the first step in this direction which shows that if $x$ is covered by $C$, then the cover can be extended to a nearby point $y$ with only a small loss in parameters.

**Lemma 1.7.** *Suppose that $x$ is $(\sigma, \delta, \ell)$-covered by $C$, and $z \in \mathbb{R}^d$. Then for every $t \geq 0$, $z$ is $(\sigma - t\|x - z\|, \delta - \exp(-t^2/2), \ell + \|x - z\|)$-covered by $C$.*

*Proof.* In order to have $\langle x - y, z \rangle \geq \sigma/\sqrt{d}$ for some $y \in C$, but

$$\langle z - y, u \rangle < \frac{\sigma - t\|x - y\|}{\sqrt{d}},$$

it must be the case that $\langle z - y, u \rangle \geq \frac{t\|x-y\|}{\sqrt{d}}$. But by Lemma 1.4, the probability of this over a random $u$ is at most $\exp(-t^2/2)$. In addition, clearly $\|y - z\| \leq \ell + \|x - z\|$ for every $y \in C$. $\square$

Now we can apply this lemma to sets and get the following corollary.

**Corollary 1.8.** *If $S \subseteq \mathbb{R}^d$ is $(\sigma, \delta, \ell)$-covered by $C$, then for every $t, \epsilon \geq 0$, the neighborhood $S_\epsilon = \{z \in \mathbb{R}^d : d(z, S) \leq \epsilon\}$ is $(\sigma - \epsilon t, \delta - \exp(-t^2/2), \ell + \epsilon)$-covered by $C$.*

The following lemma which can be proven by Levy's lemma shows that in a $(\sigma, \delta, \ell)$-cover by decreasing $\sigma$ slightly we can increase $\delta$ a lot.

**Lemma 1.9.** *Suppose that $x$ is $(\sigma, \delta, \ell)$-covered by $C$, then for every $\gamma > \sqrt{2\log(2/\delta)} + t$, $x$ is also $(\sigma - 2\ell\gamma, 1 - \exp(-t^2/2), \ell)$-covered by $C$.*

**Exercise 1.10.** Prove Lemma 1.9.

## 1.1 proof of Theorem 1.2

To prove Theorem 1.2 we will show that there exists a set $S_R \subseteq C$ of size

$$R = \left\lfloor \frac{\sigma^2}{2^{11}\ell^2 \log(8/\delta^2)} \right\rfloor$$

such that $S_R$ is $(\frac{\sigma R}{4}, 1 - \delta/2, 1)$-covered by $C$. Combining this with Lemma 1.5 completes the proof as

$$|C| \geq \exp(\Omega(\sigma R)^2) \geq \exp(\Omega(\sigma^6/\ell^4 \log^2(1/\delta))).$$

To prove that such a $S_R$ exists we start with $S_0 = C$ which is trivially $(0, 1, 0)$-covered by $C$. Then we "attach" matching edges to this cover inductively to obtain $S_R$. Lemma 1.11 below which is a major step towards the proof of Theorem 1.2 shows that how one can attach matching edges from a core to a cover to obtain a better cover. For subsets $S \subseteq Y \subseteq \mathbb{R}^d$, define

$$\Gamma_Y(S, r) = \{y \in Y : d(y, S) \leq r\}.$$

Additionally, for $k \in \mathbb{N}$, define $\Gamma_Y^k(S, r)$ inductively by

$$\Gamma_Y^k(S, r) = \Gamma_Y(\Gamma_Y^{k-1}(S, r), r),$$

with $\Gamma_Y^1(S, r) = \Gamma_Y(S, r)$.

**Lemma 1.11.** *Suppose that $C \subseteq \mathbb{R}^d$ is a $(\sigma_0, \delta_0, \ell_0)$-core. Additionally, suppose that $S \subseteq C$ is $(\sigma, 1 - \frac{\delta_0}{2}, \ell)$-covered by $C$. Let $\beta = \frac{|S|}{\Gamma_C(S, \ell_0)}$. Then there exists a subset $S' \subseteq \Gamma_C(S, \ell_0)$ with the following properties.*

- $|S'| \geq \frac{\delta_0 |S|}{4}$.
- *$S'$ is $(\sigma + \sigma_0, \frac{\delta_0 \beta}{4}, \ell + \ell_0)$-covered by $C$.*

*Proof.* Let $M : S^{d-1} \to \mathcal{M}(C)$ be the matching cover of $C$ by itself. Consider a point $x \in S$. Since $S$ is $(\sigma, 1 - \frac{\delta_0}{2}, \ell)$-covered by $C$, for a $1 - \delta_0/2$ fraction of directions $u \in S^{d-1}$, there exists some $y_u \in B_C(x, \ell)$ such that $\langle x - y_u, u \rangle \geq \frac{\sigma}{\sqrt{d}}$. In addition (since $C$ is a core), for a $\delta_0$ fraction of $u \in S^{d-1}$, there exists a point $z_u$ such that $(z_u, x) \in M(u)$, which implies that $\langle z_u - x, u \rangle \geq \frac{\sigma_0}{\sqrt{d}}$ and $z \in B_c(x, \ell_0)$ (in particular, $z \in \Gamma_C(S, \ell_0)$).

By a trivial intersection bound, for a $\delta_0/2$ fraction of $u \in S^{d-1}$, both events happen simultaneously, and we have $\langle z_u - y_u, u\rangle \geq \frac{\sigma+\sigma_0}{\sqrt{d}}$. In this case, we define $A(z_u, u) = y_u$. Observe that this is well-defined; since $M(u)$ is a matching, $A(z_u, u)$ is assigned at most once. Doing this for every $x \in S$, $u \in S^{d-1}$ defines a partial assignment $A : C \times S^{d-1} \to C$.

Define a measure $\mu_A$ on $C$ by

$$\mu_A(z) = \Pr_{u \in S^{d-1}}[A(z,u) \text{ is defined}].$$

First, we have $\mu_A(C) \geq \frac{\delta_0}{2}|S|$ by construction. Secondly, we have $\mu_A(z) > 0$ only if $z \in \Gamma_c(S, \ell_0)$, and trivially $\mu_A(z) \leq 1$ for every $z \in C$. define

$$S' = \left\{ z \in C : \mu_A(z) \geq \frac{\delta_0 \beta}{4} \right\},$$

and observe that

$$\frac{\delta_0}{2}|S| = \mu_A(C) \leq |\Gamma_C(S, \ell_0)| \frac{\delta_0\beta}{4} + |S'| = \frac{\delta_0}{4}|S| + |S'|.$$

We conclude that $|S'| \geq \frac{\delta_0}{4}|S|$. Additionally, every $z \in C$ is $(\sigma+\sigma_0, \mu_A(z), \ell+\ell_0)$-covered by the set $\{A(z,u) : A(z,u) \text{ is defined}\}$, so $S'$ itself is $(\sigma+\sigma_0, \frac{\delta_0\beta}{4}, \ell+\ell_0)$-covered by $C$. ☐

As we said above to prove Theorem 1.2 it is sufficient to show that there exists a set $S_R \subseteq C$ of size

$$R = \left\lfloor \frac{\sigma^2}{2^{11}\ell^2 \log(8/\delta^2)} \right\rfloor$$

such that $S_R$ is $(\frac{\sigma R}{4}, 1-\delta/2, 1)$-covered by $C$. We prove this by induction, where we show that:

For $0 \leq r \leq R$, there exists $S_r \subseteq C$ satisfying

1. $S_r$ is $(\frac{\sigma r}{4}, 1-\delta/2, 2\ell\sqrt{r})$-covered by $C$.
2. $|S_r| \geq (\frac{\delta}{4})^r |C|$.
3. $|S_r| \geq \delta |\Gamma_C(S_r, \ell)|$ (i.e. $\beta \geq \delta$ in Lemma 1.11).

**The base case:** Let $S_0 = C$. Then since $S_0$ is trivially $(0,1,0)$-covered by $C$, the inductive assumption is satisfied.

Now assume that $S_{r-1}$ satisfies the inductive assumption and that $r \leq R$. The construction of $S_r$ proceeds in three steps.

(S1) **Use the core to extend the set $S_{r-1}$ to $S'_r \subseteq \Gamma_C(S_{r-1}, \ell)$.**
We apply Lemma 1.11 to the set $S_{r-1}$ and the core $C$ to obtain $S'_r$. Observe that by property (3) of $S_{r-1}$, the value of $\beta$ in Lemma 1.11 is at least $\delta$. It follows that $S'_r$ is $(\frac{\sigma}{4}(r-1)+\sigma, \delta^2/4, \ell')$-covered by $C$ for some $\ell'$ (the value of which we address in step (S3)). Additionally, using property (1) of Lemma 1.11, $|S'_r| \geq (\delta/4)|S_{r-1}| \geq (\delta/4)^r|C|$.

(S2) **Grow $S'_r$ until it stops expanding.** The set $S'_r$ obtained above does not satisfy the property (3) of induction hypothesis. To fix this we do the following. Let $k \geq 0$ be the first value of which $|\Gamma_C^k(S'_r, \ell) \geq \delta|\Gamma_C^{k+1}(S'_r, \ell)|$. Let $S_r = \Gamma^k(S'_r, \ell)$. Notice that the neighborhood condition (3) is satisfied by construction. Condition (2) is satisfied since $S_r \supseteq S'_r$.

We claim that $S_r$ is $(\frac{\sigma}{4}(r-1)+\frac{\sigma}{2}, \delta^2/8, \ell'')$-covered by $C$ for some $\ell''$ addressed in (S3). First, since we had $|S'_r| \geq (\delta/4)^r|C|$ it follows that

$$k \leq \log_{1/\delta}\left(\frac{4}{\delta^2}\right)^r \leq 3r. \tag{1}$$

It follows that for every $a \in S_r$, there exists a $b \in S'_r$ and a sequence $a = a_0, \ldots, a_k = b$ of points in $C$ such that $\|a_i - a_{i+1}\| \leq \ell$ for $i = 0, \ldots, k-1$. Now use the fact that $d_C(x,y) = \|x-y\|^2$ is a metric on $C$ to conclude that

$$\|x-y\|^2 \leq \sum_{i=0}^{k-1} \|a_i - a_{i+1}\|^2 \leq 3r\ell^2 \tag{2}$$

i.e. $\|x-y\| \leq \ell\sqrt{3r}$. So $S_r \subseteq N_\epsilon(S'_r)$ for $\epsilon = \ell\sqrt{3r}$. Applying Corollary 1.8 to $S'_r$ with $t = \sigma/(2\epsilon)$, we conclude that $S_r$ is $(\frac{\sigma}{4}(r-1)+\sigma/2, \sigma^2/4 - \exp(-t^2/2), \ell')$-covered by $C$. This yields our desired conclusion as long as $\exp(-t^2/2) \leq \delta^2/8$. This is true as long as

$$r \leq \frac{\sigma^2}{24\ell^2\log(8/\delta^2)} \tag{3}$$

which holds true since $r \leq R$.

(S3) **Bounding $\ell''$ and boosting the cover to $1-\frac{\delta}{2}$.**

First we consider the size of $\ell''$. Observe that in (S1), in augmenting our cover with Lemma 1.11, we go at most "one step" (along some "edge" of the matching cover) when passing from $S_{r-1}$ to $S'_r$ (this corresponds to the fact that in property (2) of Lemma 1.11 the set $S'$ is covered by vectors of length at most $\ell+\ell_0$, where $\ell_0$ is the length of a vector in the matching cover). Additionally using the bound (1), we see that the total number of steps taken by (S2) is at most $3r$. using a similar calculation to the one in (2) we conclude that $\ell'' \leq 2\ell\sqrt{r}$.

Lemma 1.9 with $\gamma = \sigma/(8\ell'')$ and $t = \sqrt{2\log(2/\delta)}$ to conclude that $S_r$ is also $(\frac{\sigma}{4}r, 1-\frac{\delta}{2}, 2\ell\sqrt{r})$-covered by $C$. This is possible as long as

$$\gamma > \sqrt{2\log(2/\delta)} + t = 2\sqrt{2\log(2/\delta)},$$

which holds whenever

$$r < \frac{\sigma^2}{2^{11}\ell^2\log(2/\delta)},$$

which is true since $r \leq R$.

This completes the induction.

# References


[ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 222–231 (electronic), New York, 2004. ACM.

[Lee05] James R. Lee. Distance scales, embeddings, and metrics of negative type. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 92–101, Vancouer, BC, Canada, 2005.