Probabilistic clustering of high dimensional norms

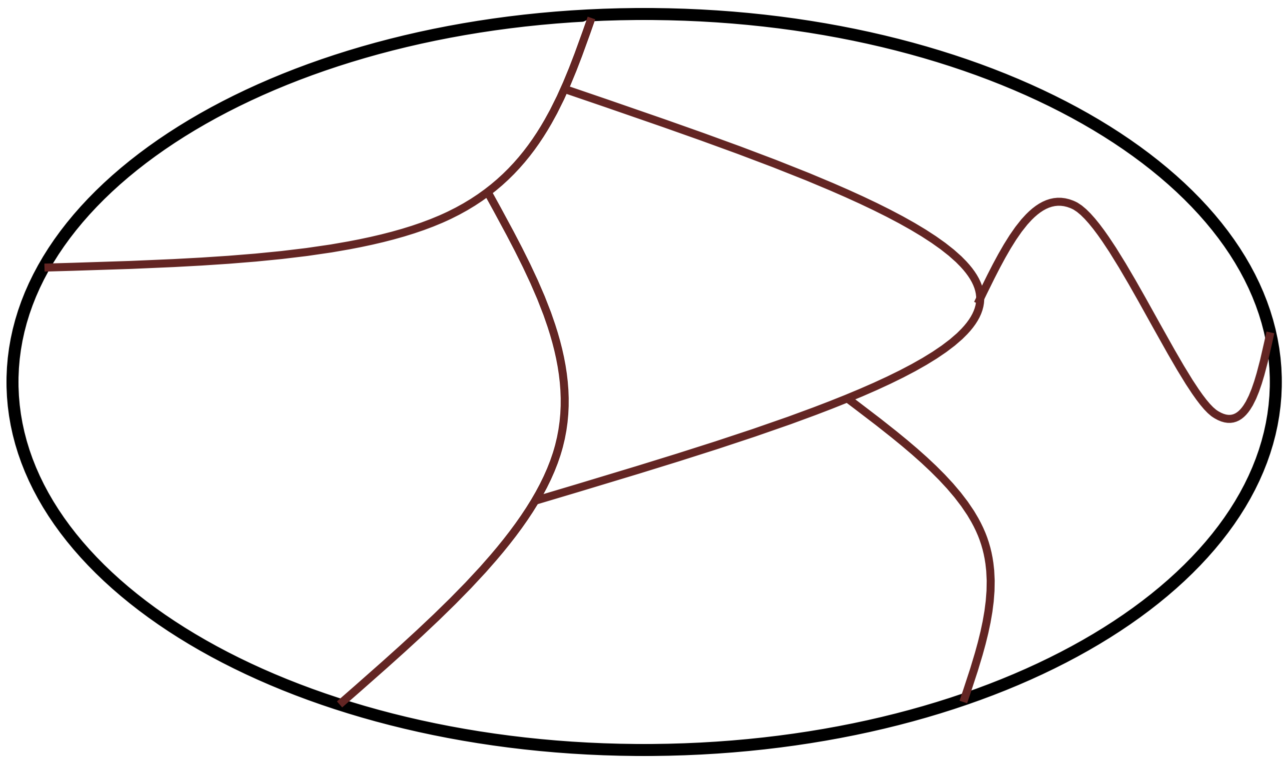
Assaf Naor

Princeton University

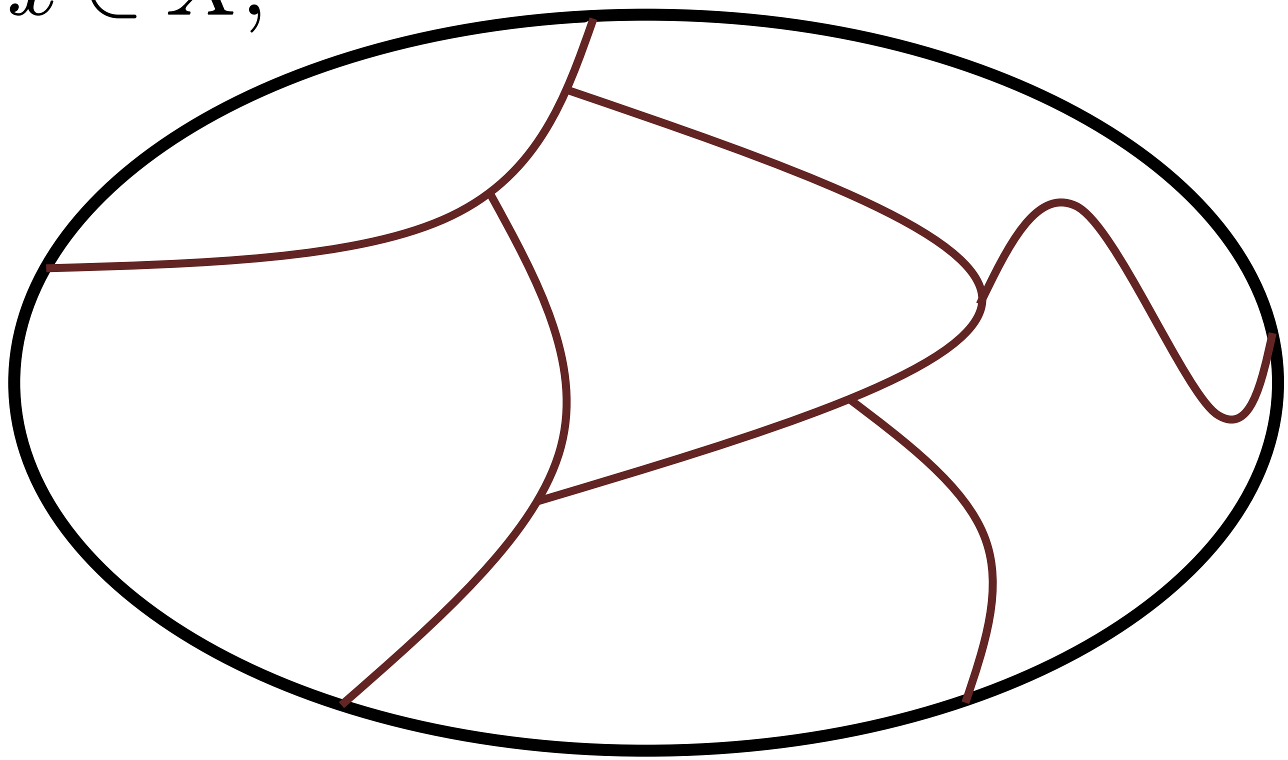
SODA'17

Partitions of metric spaces

Let (X, d_X) be a metric space and $\mathcal{P}$ a partition of *X*.

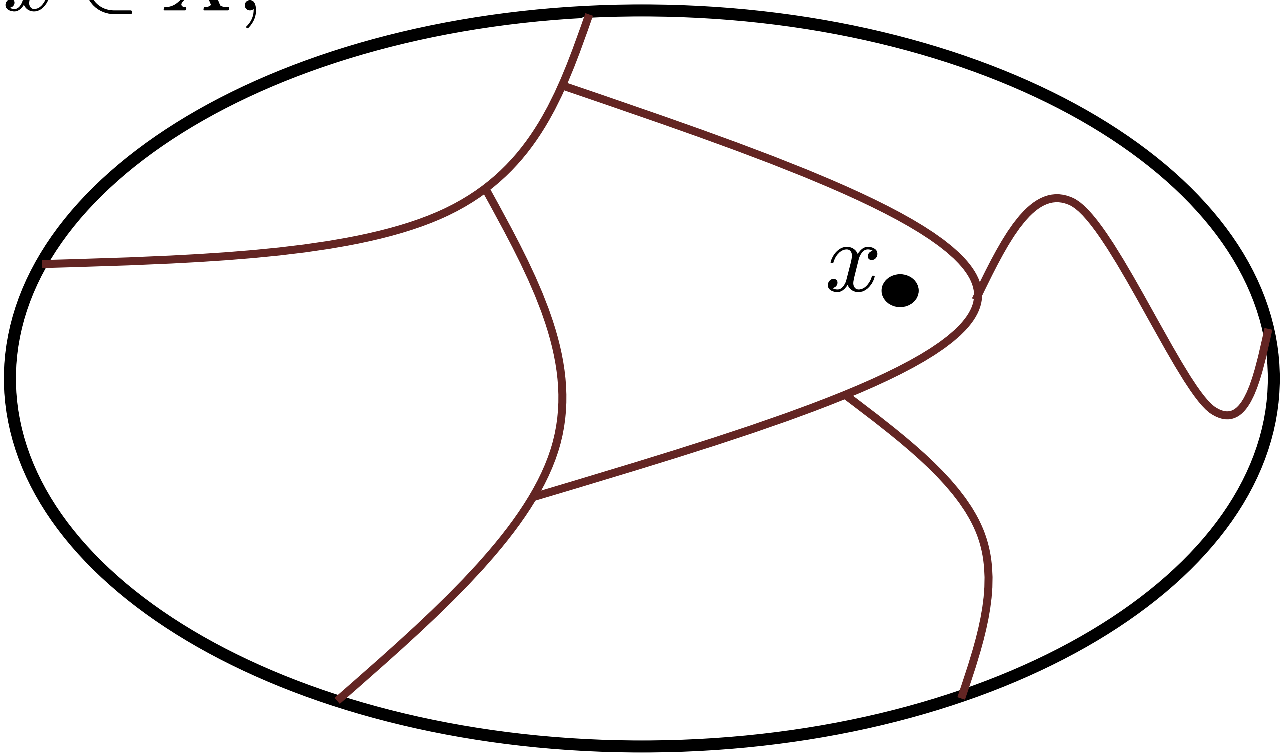


Given a point $x \in X$,



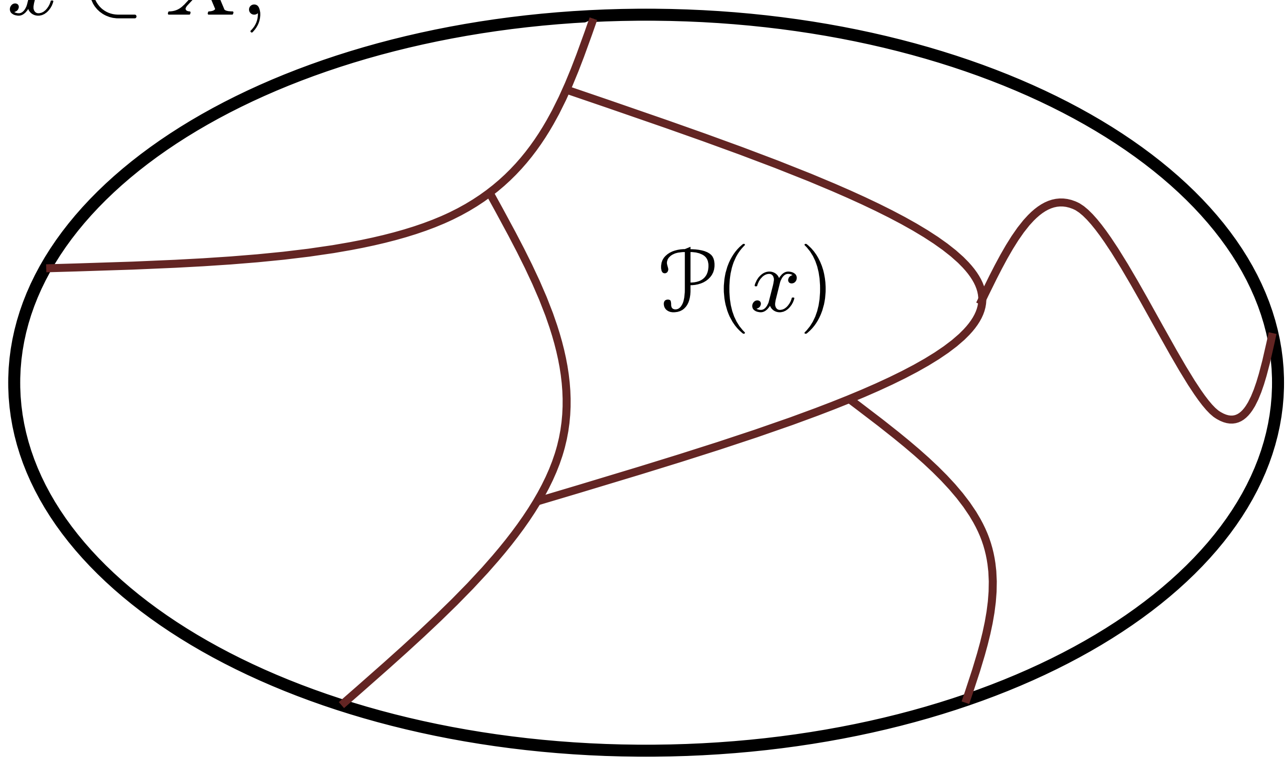
$\mathcal{P}(x)$ is the unique cluster in $\mathcal{P}$ which contains *x*.

Given a point $x \in X$,



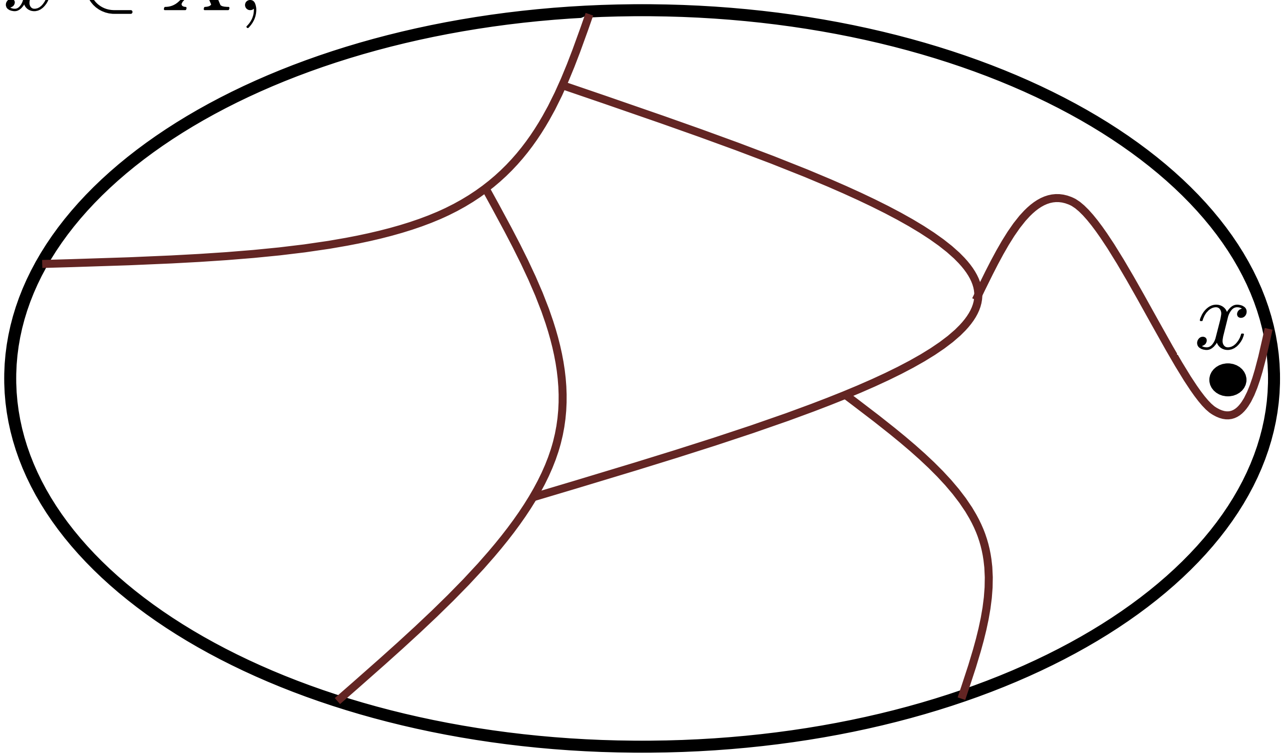
$\mathcal{P}(x)$ is the unique cluster in $\mathcal{P}$ which contains *x*.

Given a point $x \in X$,



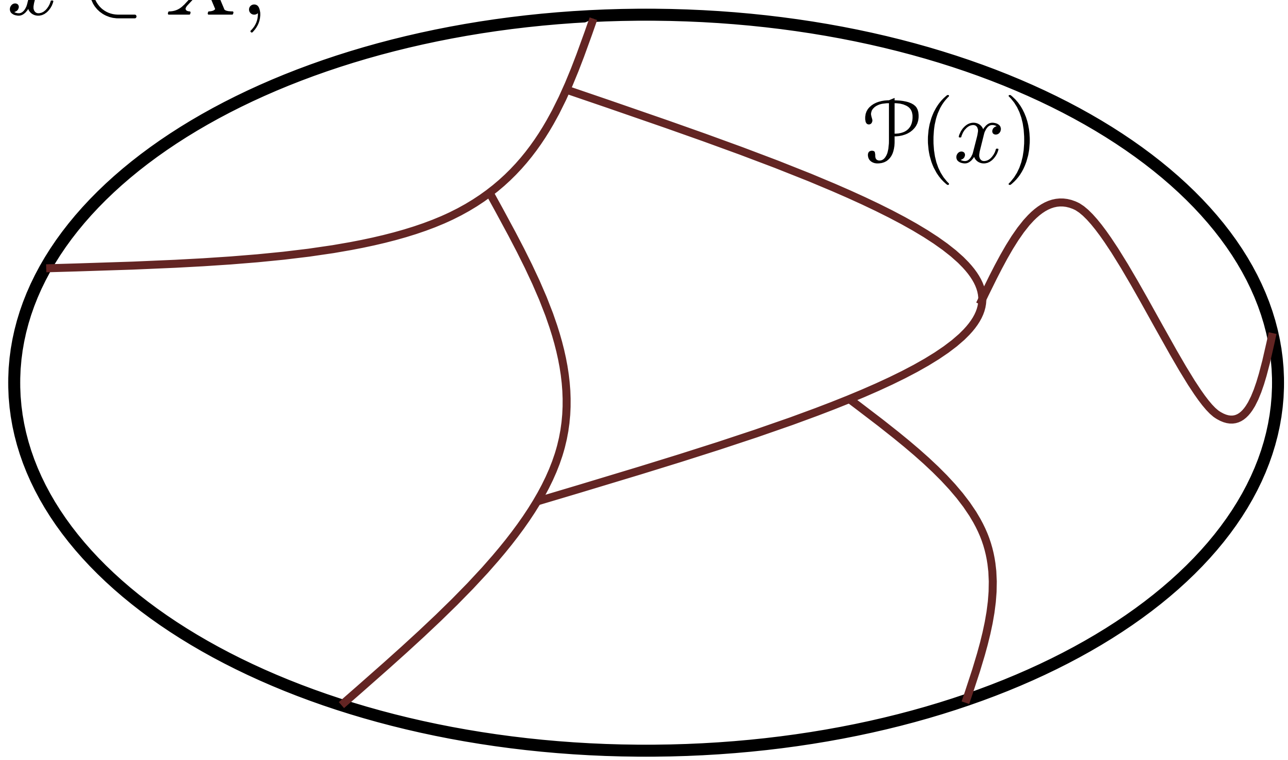
$\mathcal{P}(x)$ is the unique cluster in $\mathcal{P}$ which contains *x*.

Given a point $x \in X$,



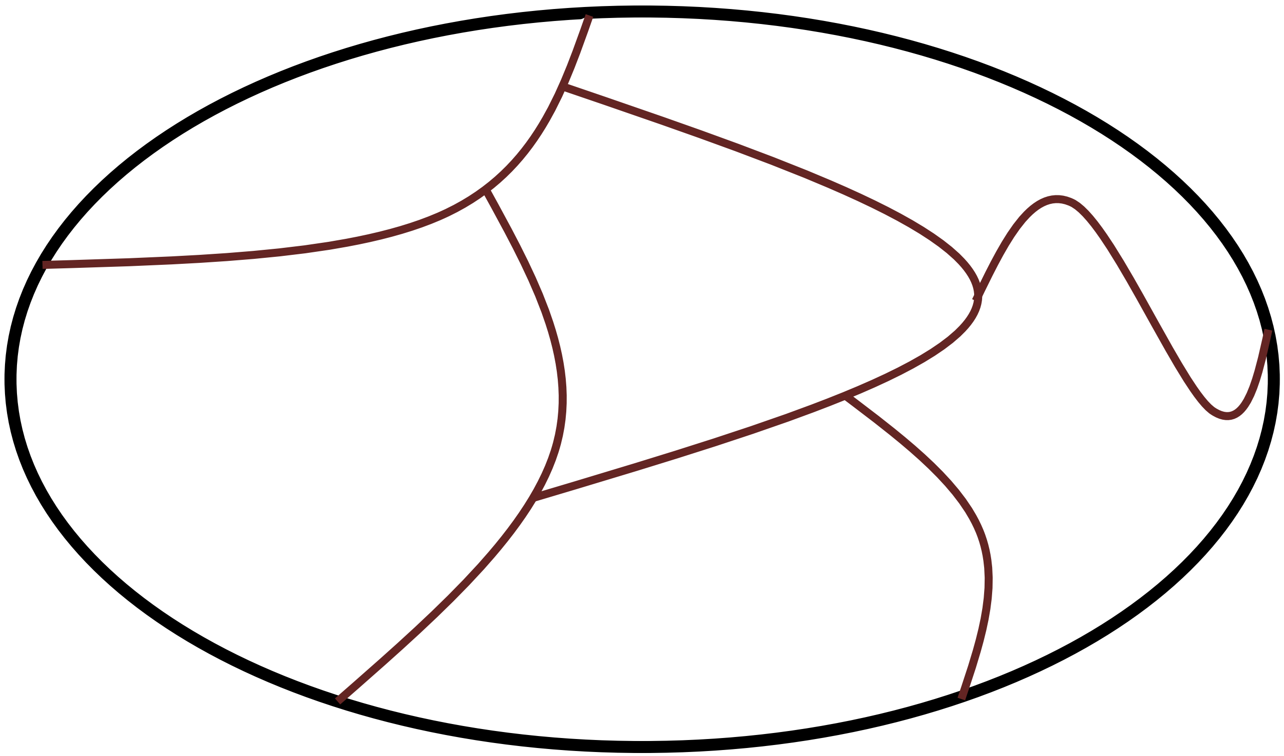
$\mathcal{P}(x)$ is the unique cluster in $\mathcal{P}$ which contains *x*.

Given a point $x \in X$,



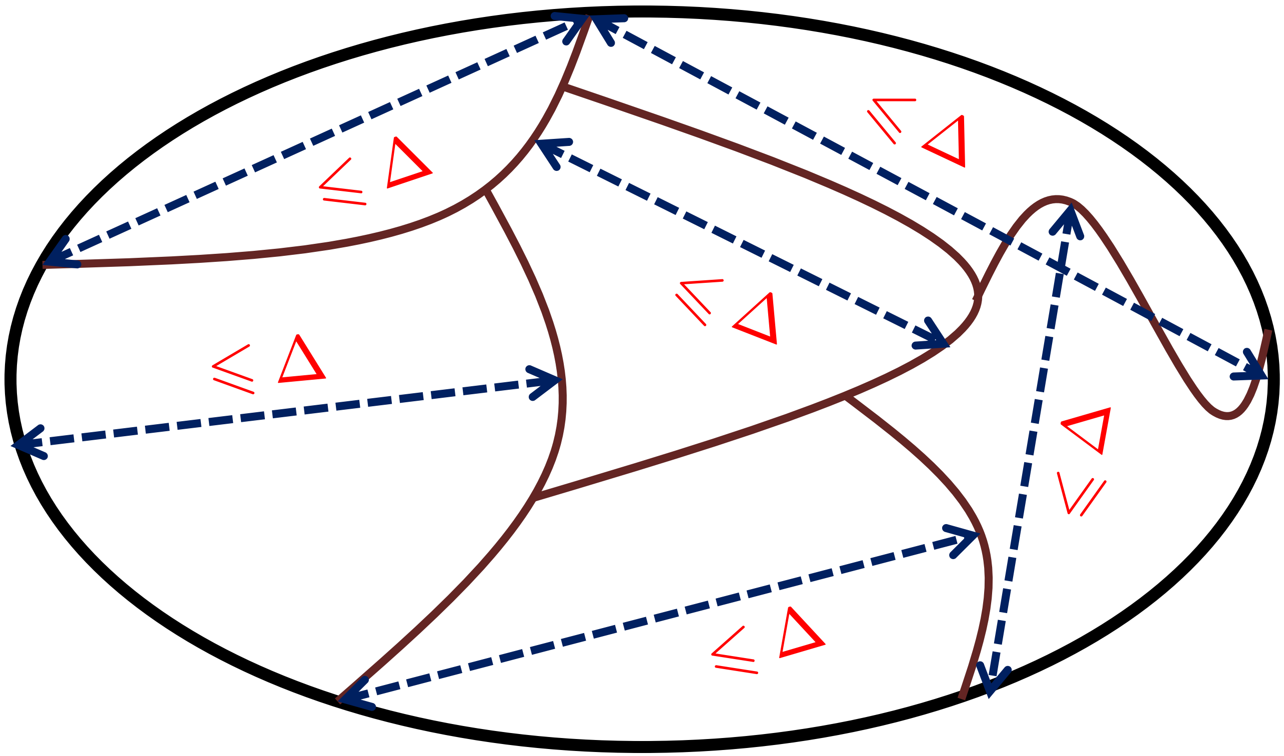
$\mathcal{P}(x)$ is the unique cluster in $\mathcal{P}$ which contains *x*.

Given $\Delta > 0$, the partition $\mathcal{P}$ is Δ–bounded if all the clusters of $\mathcal{P}$ have diameter at most Δ.



$$\forall\, x \in X, \quad \mathrm{diam}_X\big(\mathcal{P}(x)\big) := \max_{u,v\in\mathcal{P}(x)} d_X(u,v) \leqslant \Delta.$$

Given $\Delta > 0$, the partition $\mathcal{P}$ is $\Delta-$bounded if all the clusters of $\mathcal{P}$ have diameter at most Δ.



$$\forall\, x \in X, \quad \mathrm{diam}_X\big(\mathcal{P}(x)\big) := \max_{u,v\in\mathcal{P}(x)} d_X(u,v) \leqslant \Delta.$$

Random partitions

- Key goal in several areas of computer science and mathematics: use a $\Delta-$bounded partition to "simplify" the metric space.

Random partitions

- Key goal in several areas of computer science and mathematics: use a $\Delta -$ bounded partition to "simplify" the metric space.
- The partition should "mimic" the coarse geometric structure (at distance scale Δ) in some meaningful way.

Random partitions

- Key goal in several areas of computer science and mathematics: use a $\Delta-$bounded partition to "simplify" the metric space.

- The partition should "mimic" the coarse geometric structure (at distance scale Δ) in some meaningful way.

- Regions near boundaries should be "thin."

Random partitions

- Key goal in several areas of computer science and mathematics: use a $\Delta -$ bounded partition to "simplify" the metric space.
- The partition should "mimic" the coarse geometric structure (at distance scale Δ) in some meaningful way.
- Regions near boundaries should be "thin."
- Quite paradoxical, but randomness helps here…

Separating random partitions

<u>Definition (Bartal, 1996)</u>: Suppose that (X, d_X) is a metric space and $\sigma, \Delta > 0$.

A distribution $\mathcal{P}$ over $\Delta-$bounded random partitions of *X* is said to be $\sigma-$separating if

$$\forall\, x, y \in X, \qquad \mathbb{P}\big[\mathcal{P}(x) \neq \mathcal{P}(y)\big] \leqslant \frac{\sigma}{\Delta} d_X(x, y).$$

(Implicit in several early works, variety of applications: Leighton-Rao [1988], Auerbuch-Peleg [1990], Linial-Saks [1991], Alon-Karp-Peleg-West [1991], Klein-Plotkin-Rao [1993], Rao [1999].)

Modulus of separated decomposability

Denote by $\mathrm{SEP}(X)$ the minimum $\sigma > 0$ such that for every $\Delta > 0$ there is a $\sigma-$separating distribution over $\Delta-$bounded random partitions of (X, d_X).

Note: we are ignoring here technical measurability issues that are important for mathematical applications in the infinite setting. For TCS purposes, it suffices to deal with random partitions of finite subsets of *X*.

Theorem (Bartal, 1996): If $|X| = n$ then

$$\mathsf{SEP}(X) \lesssim \log n.$$

Goal of present work: to study $\mathsf{SEP}(X)$ for finite dimensional normed spaces *X* (and subsets thereof).

Originated in Peleg-Reshef [1998], followed by important work of Charikar-Chekuri-Goel-Guha-Plotkin [1998].

Sharp a priori bounds

<u>Theorem</u>: Suppose that *X* is an *n*-dimensional normed space. Then

$$\sqrt{n} \lesssim \mathsf{SEP}(X) \lesssim n.$$

The upper bound follows from [CCGGP98].

The lower bound hasn't been noticed before: it follows from a theorem of Bourgain-Szarek (1988) that is a consequence of the Bourgain-Tzafriri restricted invertibility principle (1987).

$$\sqrt{n} \lesssim \mathsf{SEP}(X) \lesssim n.$$

Both bounds are asymptotically sharp, as shown in [CCGGP98]. In fact, it is proved there that

$$\mathsf{SEP}(\ell_2^n) \asymp \sqrt{n} \qquad \text{and} \qquad \mathsf{SEP}(\ell_1^n) \asymp n.$$

For $p \in [1, \infty)$ and $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$\|x\|_{\ell_p^n} := \left(\sum_{j=1}^{n} |x_j|^p \right)^{\frac{1}{p}}$$

$$\|x\|_{\ell_\infty^n} := \max_{j \in \{1, \ldots, n\}} |x_j|.$$

In [CCGGP98], Charikar-Chekuri-Goel-Guha-Plotkin asserted that

$$\mathsf{SEP}(\ell_p^n) \asymp \begin{cases} n^{\frac{1}{p}} & \text{if } 1 \leqslant p \leqslant 2, \\ n^{1-\frac{1}{p}} & \text{if } 2 \leqslant p \leqslant \infty. \end{cases}$$

The upper bound on $\mathsf{SEP}(\ell_p^n)$ in the above equivalence is valid as stated for all $1 \leqslant p \leqslant \infty$,

In [CCGGP98], Charikar-Chekuri-Goel-Guha-Plotkin asserted that

$$\mathsf{SEP}(\ell_p^n) \asymp \begin{cases} n^{\frac{1}{p}} & \text{if } 1 \leqslant p \leqslant 2, \\ n^{1-\frac{1}{p}} & \text{if } 2 \leqslant p \leqslant \infty. \end{cases}$$

The upper bound on $\mathsf{SEP}(\ell_p^n)$ in the above equivalence is valid as stated for all $1 \leqslant p \leqslant \infty$, but we show here that the matching lower bound is incorrect when $2 < p \leqslant \infty$.

In [CCGGP98], Charikar-Chekuri-Goel-Guha-Plotkin asserted that

$$\mathsf{SEP}(\ell_p^n) \asymp \begin{cases} n^{\frac{1}{p}} & \text{if } 1 \leqslant p \leqslant 2, \\ n^{1-\frac{1}{p}} & \text{if } 2 \leqslant p \leqslant \infty. \end{cases}$$

The upper bound on $\mathsf{SEP}(\ell_p^n)$ in the above equivalence is valid as stated for all $1 \leqslant p \leqslant \infty$, but we show here that the matching lower bound is incorrect when $2 < p \leqslant \infty$. Thus, in particular, we obtain an asymptotically better probabilistic clustering of, say, ℓ_∞^n.

Theorem: For every $p \in [2,\infty]$,

$$\mathsf{SEP}(\ell_p^n) \lesssim \sqrt{n\min\{p,\log n\}}.$$

In particular, the previous best known bound when $p=\infty$ was $\mathsf{SEP}(\ell_\infty^n) \lesssim n$ (and this was asserted in [CCGGP98] to be sharp), but here we show that actually

$$\sqrt{n} \lesssim \mathsf{SEP}(\ell_\infty^n) \lesssim \sqrt{n\log n}.$$

The source of the error in [CCGGP98] was that it relied on unpublished work of Indyk (1998) that was not published since then; we confirmed with Indyk as well as with some of the authors of [CCGGP98] that there is indeed a flaw in the (unpublished) work of Indyk that was cited.

There is no flaw in the proof of [CCGGP98] in the range $p \in [1,2]$, i.e.,

$$p \in [1,2] \implies \mathsf{SEP}(\ell_p^n) \asymp n^{\frac{1}{p}}.$$

Refined probabilistic partitions for sparse or rapidly decaying vectors

For $n \in \mathbb{N}$ and $k \in \{1,\ldots,n\}$ denote by $(\ell_p^n)_{\leqslant k}$ the subset of $\mathbb{R}^n$ consisting of all of those vectors with at most *k* nonzero entries, equipped with the ℓ_p^n metric.

<u>Theorem</u>: For every $p \geqslant 1$ we have

$$\mathrm{SEP}\big((\ell_p^n)_{\leqslant k}\big) \lesssim k^{\max\left\{\frac{1}{p},\frac{1}{2}\right\}}\sqrt{\log\binom{n}{k} + \min\{p, \log n\}}.$$

The special case $p = 2$ becomes

$$\forall\, k \in \{1, \dots, n\}, \quad \mathsf{SEP}\big((\ell_2^n)_{\leqslant k}\big) \lesssim \sqrt{k \log\left(\frac{en}{k}\right)}.$$

A curious aspect of this bound is that despite the fact that it is a statement about Euclidean geometry, our proof involves non-Euclidean geometric considerations. Specifically, the ubiquitous "iterative ball partitioning method" is applied to balls in ℓ_p^n with $p = 1 + \log(n/k)$.

Mixed-metric random partitions

<u>Theorem</u>: For every $p \in [1, \infty]$ and $\Delta > 0$ there exists a distribution $\mathcal{P}$ over random partitions of $\mathbb{R}^n$ with the following properties.

1) $\forall x \in \mathbb{R}^n, \quad \mathrm{diam}_{\ell_p^n}\big(\mathcal{P}(x)\big) \leqslant \Delta.$
2) For every $x, y \in \mathbb{R}^n,$

$$\mathbb{P}\big[\mathcal{P}(x) \neq \mathcal{P}(y)\big] \lesssim \frac{n^{\frac{1}{p}}\sqrt{\min\{p, \log n\}}}{\Delta} \cdot \|x - y\|_{\ell_2^n}.$$

In particular, the special case $p = 2$ shows that one can obtain a random partition of $\mathbb{R}^n$ into clusters of ℓ_∞^n diameter at most Δ yet with the exponentially stronger Euclidean separation property

$$\forall\, x, y \in \mathbb{R}^n, \quad \mathbb{P}\big[\mathcal{P}(x) \neq \mathcal{P}(y)\big] \lesssim \frac{\sqrt{\log n}}{\Delta} \cdot \|x - y\|_{\ell_2^n}.$$

Iterative ball partitioning method

Karger-Motwani-Sudan (1998),

Charikar-Chekuri-Goel-Guha-Plotkin (1998),

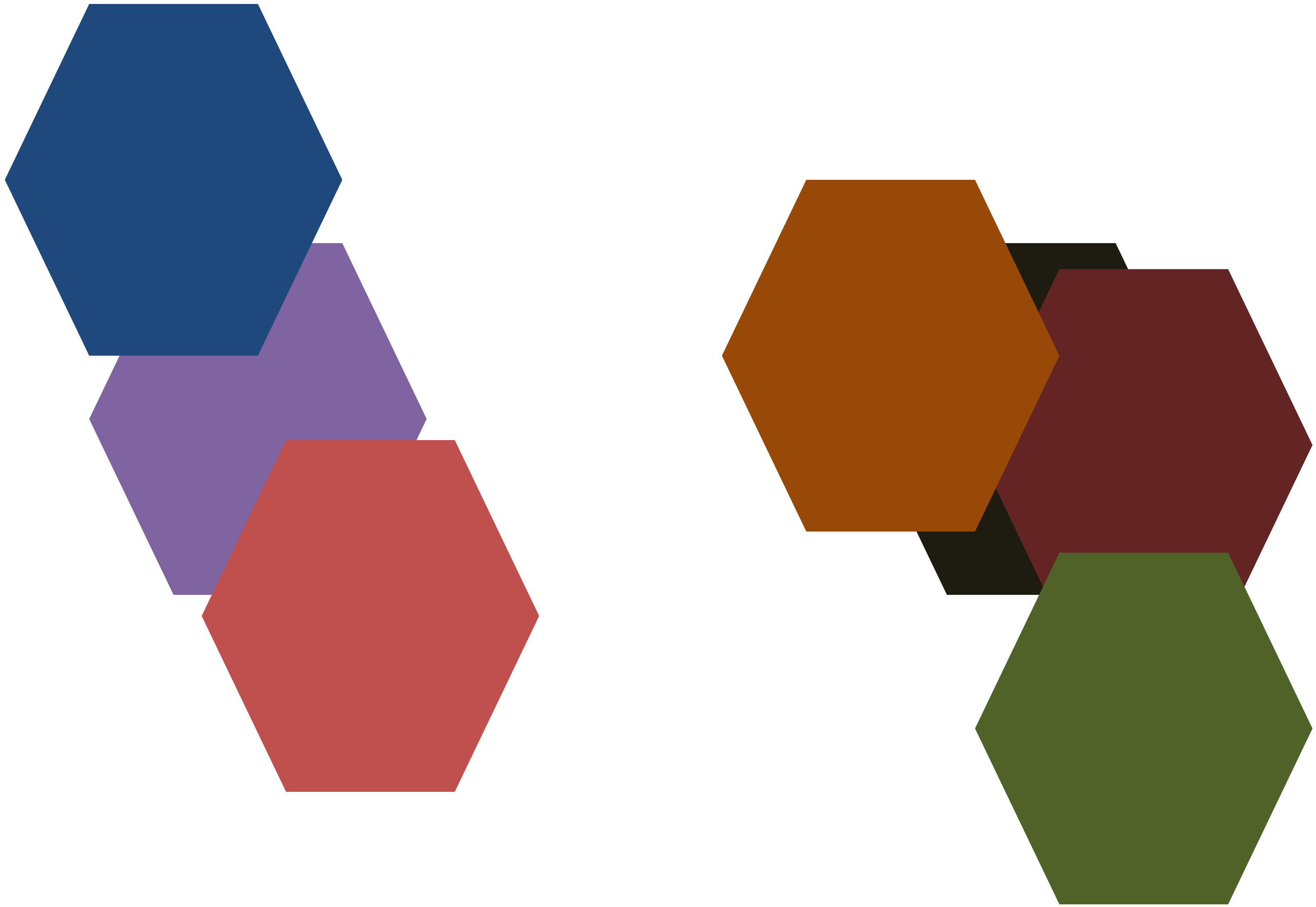
Calinescu-Karloff-Rabani (2001).

Iteratively remove balls of radius $\Delta/2$ centered at i.i.d. points in the normed space *X*.

$$B_X = \{x \in X : \ \|x\|_X \leqslant 1\}.$$

[redacted] $= x_1 + \frac{\Delta}{2} B_X.$

$$= x_2 + \frac{\Delta}{2} B_X .$$



Theorem: Let $\|\cdot\|_X$ be a norm on $\mathbb{R}^n$ and let $\mathcal{P}$ be the random partition that is obtained using iterative ball partitioning where the underlying balls are balls of radius $\Delta/2$ in the norm $\|\cdot\|_X$. Then (by design) $\mathrm{diam}_X\big(\mathcal{P}(x)\big) \leqslant \Delta$ for all $x \in \mathbb{R}^n$ and for every $x, y \in \mathbb{R}^n$ we have

$$\mathbb{P}\big[\mathcal{P}(x) \neq \mathcal{P}(y)\big]$$

$$\lesssim \frac{\mathrm{vol}_{n-1}\big(\mathsf{Proj}_{(x-y)^\perp}(B_X)\big)}{\Delta \mathrm{vol}_n(B_X)} \cdot \|x-y\|_{\ell_2^n}.$$

Sharp when the right hand side is < 1 (using Schmuckenschläger [1992]).

Extremal hyperplane projections

The previously stated theorems about random partitions of ℓ_p^n follow from this general theorem in combination with the evaluation of the extremal volumes of hyperplane projections of the unit ball of ℓ_p^n that were obtained by Barthe-N. (2002).

Extremal hyperplane projections

Theorem (Barthe-N., 2002): For every $a \in \mathbb{R}^n \setminus \{0\}$, the following function is increasing in *p*.

$$p \mapsto \frac{\mathrm{vol}_{n-1}\left(\mathsf{Proj}_{a^\perp}(B_{\ell_p^n})\right)}{\mathrm{vol}_{n-1}(B_{\ell_p^{n-1}})}.$$

When $p \geqslant 2$, the above ratio attains its maximum when $a = (1, 1, \ldots, 1)$.

The need to use an auxiliary metric

In the special case of ℓ_∞^n, if one applies iterative ball partitioning using balls in the intrinsic metric (which are in this case simply axis-parallel hypercubes $[-\Delta/2,\Delta/2]^n$), then one obtains a separation modulus of *n*.

In other words, one cannot obtain our better estimate $\mathrm{SEP}(\ell_\infty^n)\lesssim\sqrt{n\log n}$ using the intrinsic metric of the space that we wish to partition!

The need to use an auxiliary metric

Our bound follows by applying this procedure using balls in the metric that is induced from $\ell_{\log n}^n$.

The metrics on ℓ_∞^n and $\ell_{\log n}^n$ are *O(1)*-equivalent (the balls in $\ell_{\log n}^n$ are "rounded cubes"). But the corresponding volumes change drastically, which allows our theorem to yield a better (almost sharp) bound on $\mathsf{SEP}(\ell_\infty^n)$.

Further applications

- Solution of longstanding open problems on the extension of Lipschitz functions.
- Improved probabilistic partitions of the Schatten-von Neumann trace classes S_p^n and their subset consisting of all the matrices of rank at most *k* (N.-Schechtman, forthcoming); improved Lipschitz extension theorems for S_p^n.
- New volumetric stability theorems.
- Several additional results in full journal version.