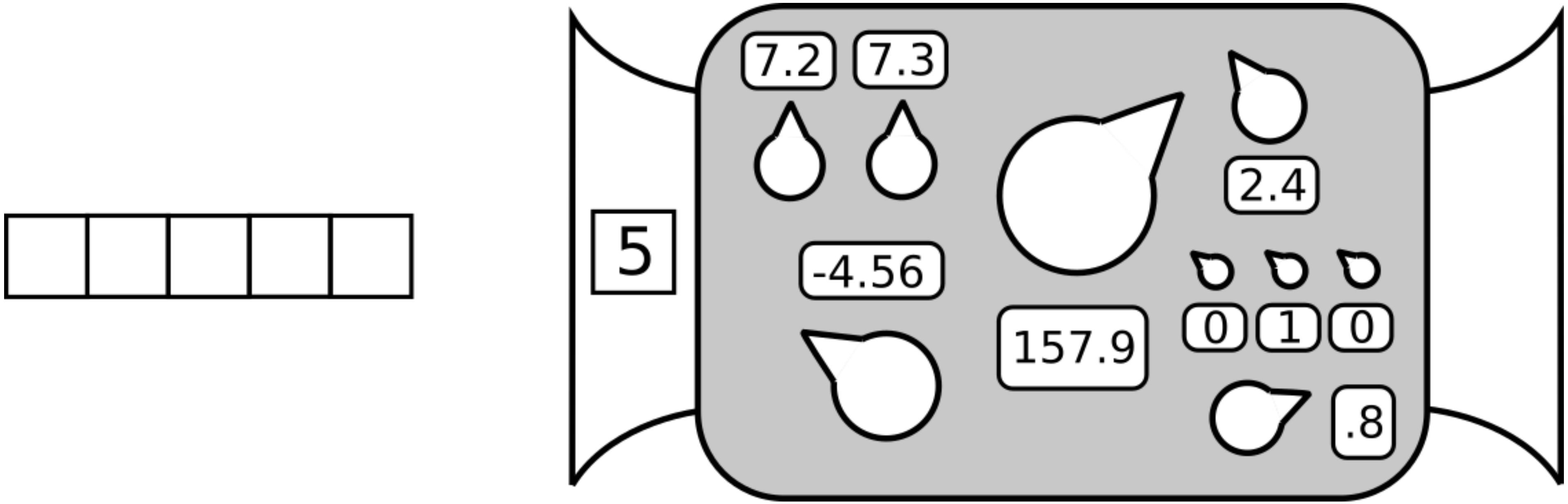
Cardinality-Agnostic Universal Approximation for Neural Networks on Point Clouds

Christian Bueno*

University of California, Santa Barbara

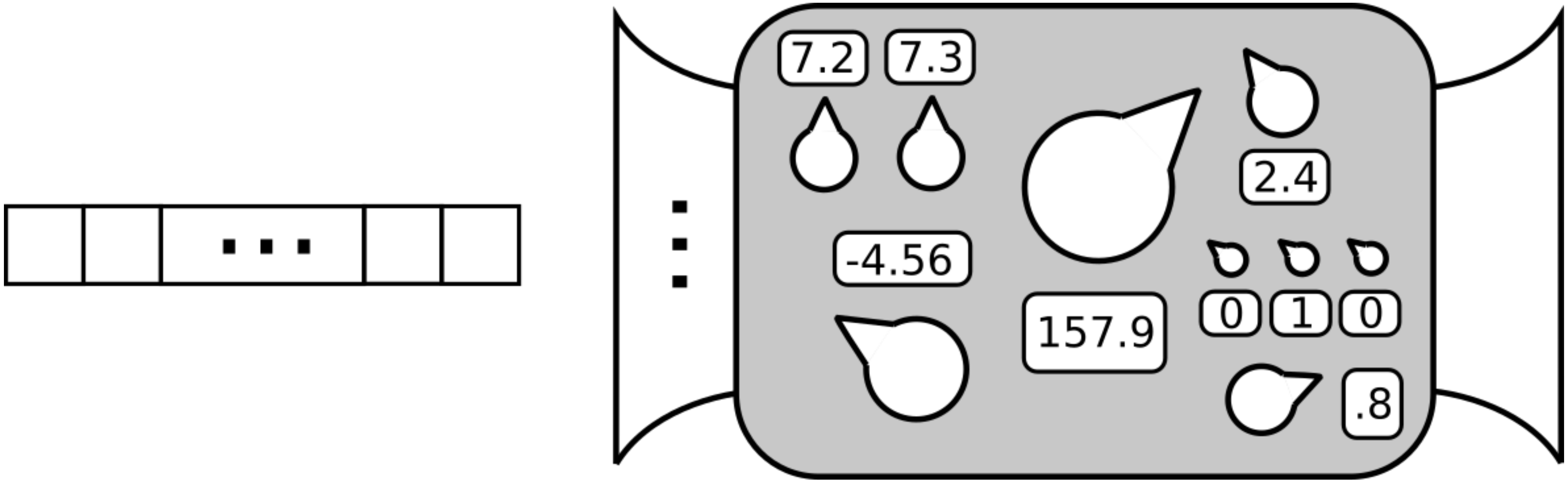
Alan G. Hylton

NASA Glenn Research Center



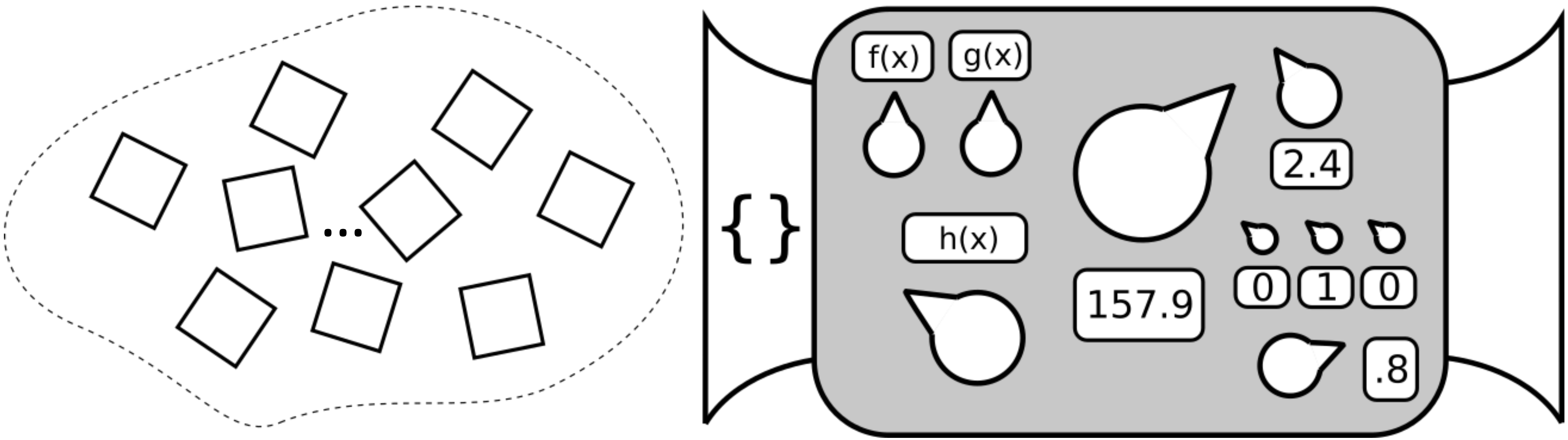
Feed-Forward Neural Networks consume fixed-size ordered data.

E.g. vectors



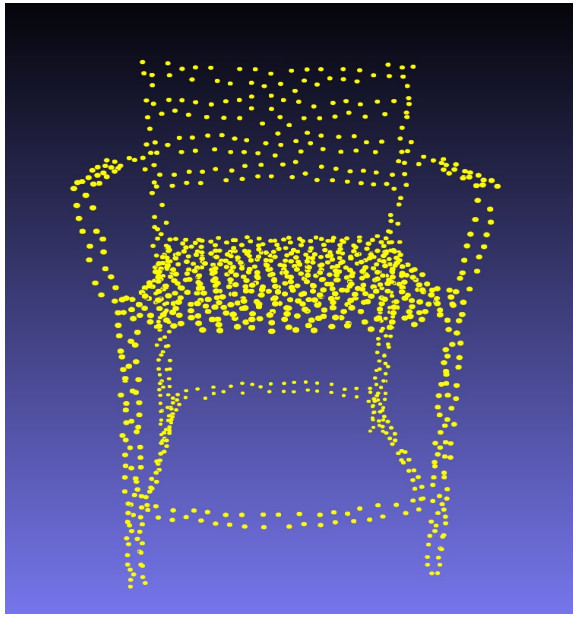
Recurrent Neural Networks consume arbitrary-size ordered data.

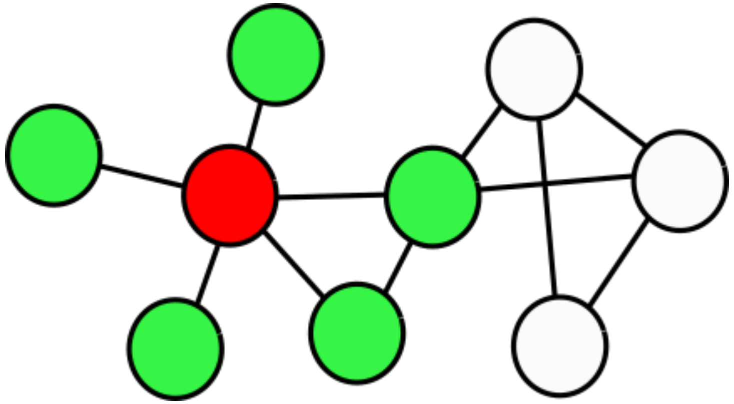
I.e. sequences

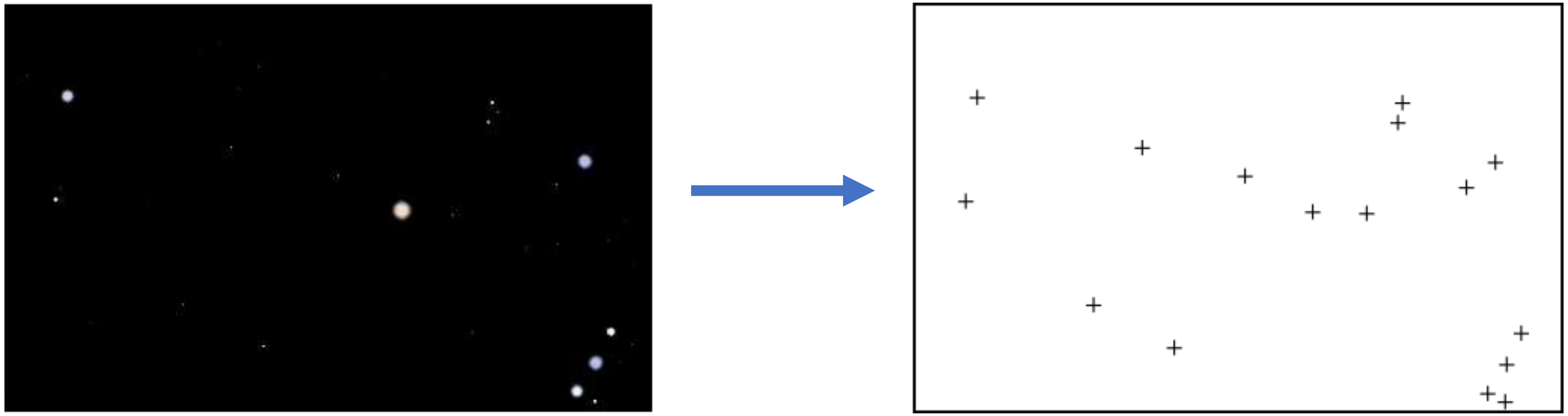


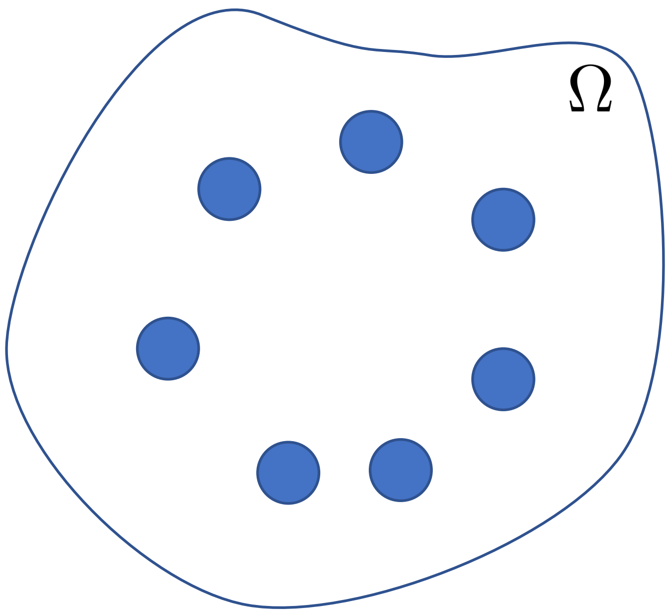
This Talk: Neural Networks that consume arbitrary-size un-ordered data.

I.e. sets



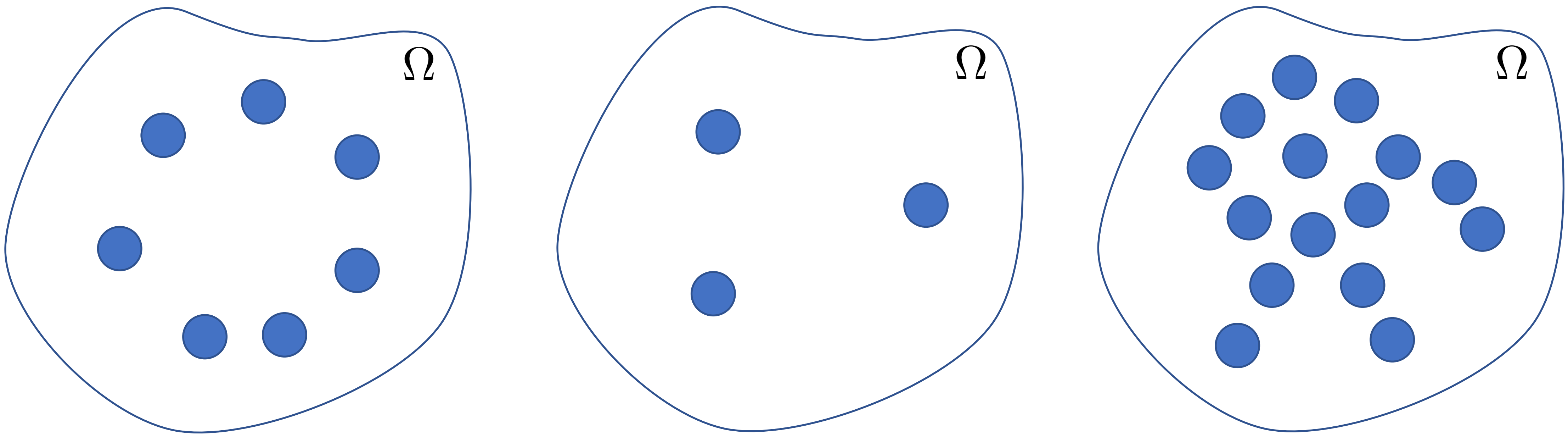






$$\mathrm{Fin}(\Omega)$$

Point Clouds

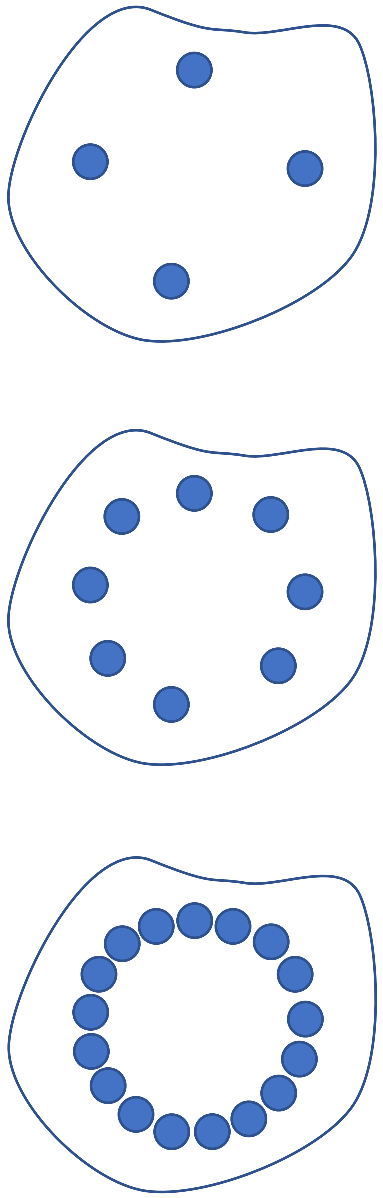


$$F : \mathrm{Fin}(\Omega) \longrightarrow \mathbb{R}^n$$

Permutation-Invariant

Cardinality-Agnostic

PointNet and DeepSets



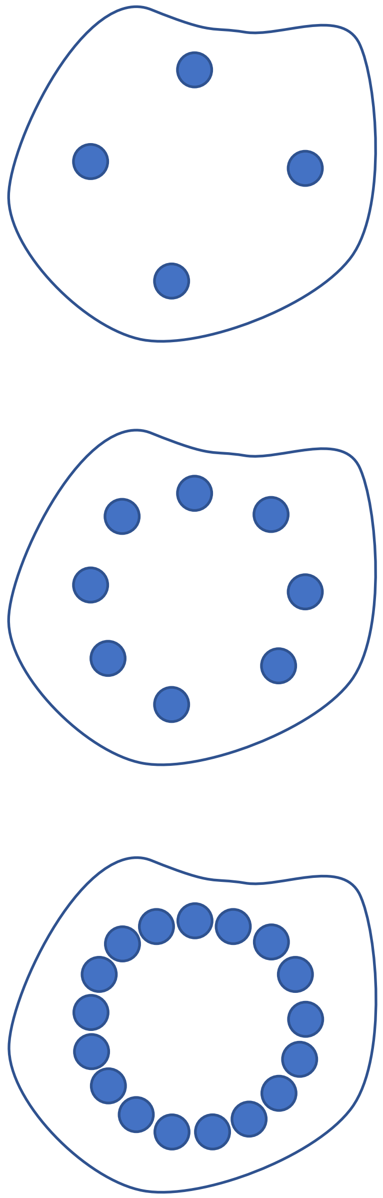
PointNet and DeepSets

$$F_{PN}(A) = \psi\left(\max_{a \in A} \varphi(a)\right)$$

(Qi et al. 2017)

$$F_{DS}(A) = \psi\left(\sum_{a \in A} \varphi(a)\right)$$

(Zaheer et al. 2017)



PointNet and DeepSets

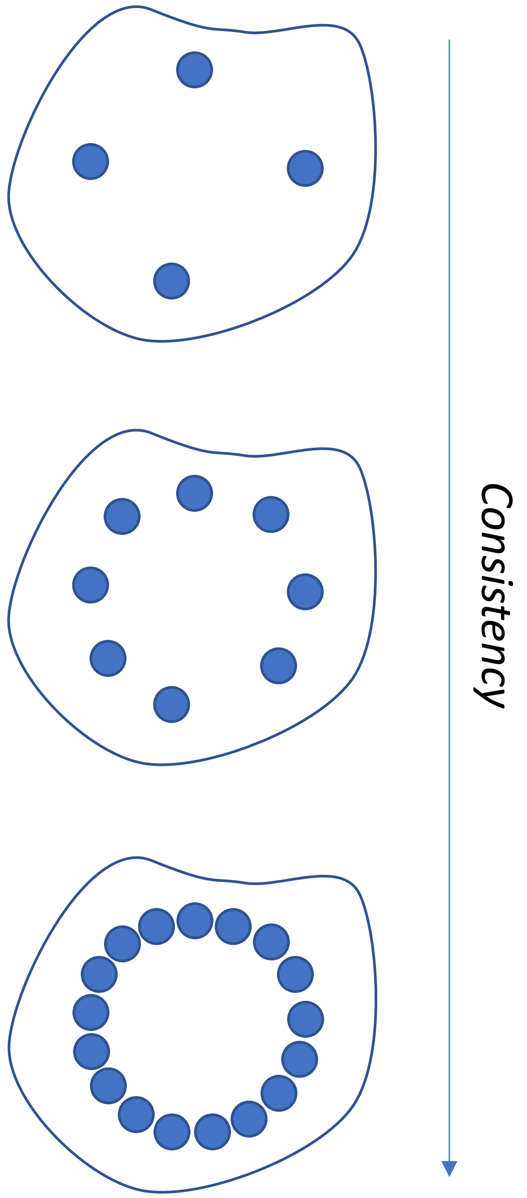
$$F_{PN}(A) = \psi\left(\max_{a \in A} \varphi(a)\right)$$

(Qi et al. 2017)

$$F_{DS}(A) = \psi\left(\sum_{a \in A} \varphi(a)\right)$$

(Zaheer et al. 2017)

$$F_{DS}(A) = \psi\left(\frac{1}{|A|}\sum_{a \in A} \varphi(a)\right)$$



Refactor

$$F_{PN} = \psi \circ \max_f$$

$$(\max_f)_i = \max_{f_i}$$

$$\max_{f_i} : \mathrm{Fin}(\Omega) \longrightarrow \mathbb{R}$$

$$F_{DS} = \psi \circ \mathrm{ave}_f$$

$$(\mathrm{ave}_f)_i = \mathrm{ave}_{f_i}$$

$$\mathrm{ave}_{f_i} : \mathrm{Fin}(\Omega) \longrightarrow \mathbb{R}$$

Continuity?

$$F_{PN} = \psi \circ \max_f \qquad F_{DS} = \psi \circ \mathrm{ave}_f$$

$$(\max_f)_i = \max_{f_i} \qquad (\mathrm{ave}_f)_i = \mathrm{ave}_{f_i}$$

$$\max_{f_i} : \mathrm{Fin}(\Omega) \longrightarrow \mathbb{R} \qquad \mathrm{ave}_{f_i} : \mathrm{Fin}(\Omega) \longrightarrow \mathbb{R}$$

What topologies yields continuity on $\mathrm{Fin}(\Omega)$?

Continuity?

$$F_{PN} = \psi \circ \max_f$$

$$F_{DS} = \psi \circ \mathrm{ave}_f$$

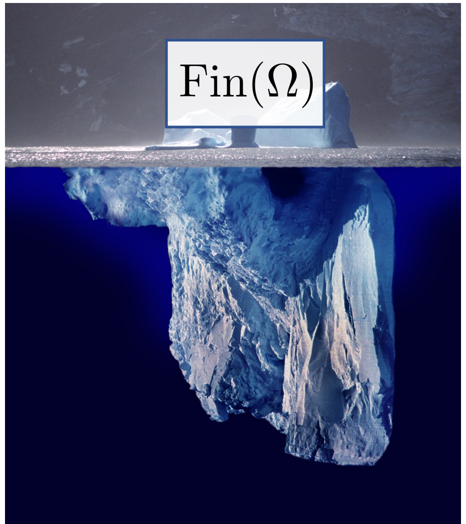
$$(\max_f)_i = \max_{f_i}$$

$$(\mathrm{ave}_f)_i = \mathrm{ave}_{f_i}$$

$$\max_{f_i} : \mathrm{Fin}(\Omega) \longrightarrow \mathbb{R}$$

$$\mathrm{ave}_{f_i} : \mathrm{Fin}(\Omega) \longrightarrow \mathbb{R}$$

What topologies yields continuity on $\mathrm{Fin}(\Omega)$?



Continuity?

$$F_{PN} = \psi \circ \max_f$$

$$F_{DS} = \psi \circ \text{ave}_f$$

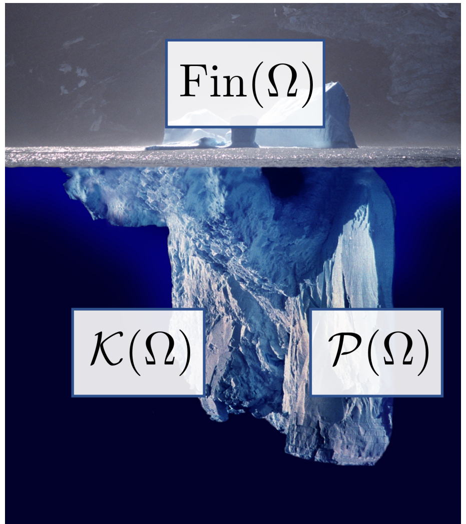
$$(\max_f)_i = \max_{f_i}$$

$$(\text{ave}_f)_i = \text{ave}_{f_i}$$

$$\max_{f_i} : \text{Fin}(\Omega) \longrightarrow \mathbb{R}$$

$$\text{ave}_{f_i} : \text{Fin}(\Omega) \longrightarrow \mathbb{R}$$

What topologies yields continuity on $\text{Fin}(\Omega)$?



$$(\mathcal{K}(\Omega), d_H)$$

Space of nonempty compact subsets with Hausdorff metric d_H

$$(\mathcal{P}(\Omega), d_W)$$

Space of Borel probability measures with Wasserstein metric d_W

Upgrade: Unique Continuous Extension

$$F_{PN} = \psi \circ \mathrm{Max}_f$$

$$\mathrm{Max}_{f_i}(A) = \max_{a \in A} f_i(a)$$

$$\mathrm{Max}_{f_i} : (\mathcal{K}(\Omega), d_H) \to \mathbb{R}$$

Space of nonempty compact subsets with Hausdorff metric d_H

$$F_{DS} = \psi \circ \mathrm{Ave}_f$$

$$\mathrm{Ave}_{f_i}(\mu) = \mathbb{E}_{x \sim \mu}[f_i(x)]$$

$$\mathrm{Ave}_{f_i} : (\mathcal{P}(\Omega), d_W) \to \mathbb{R}$$

Space of Borel probability measures with Wasserstein metric d_W

Upgrade: Unique Continuous Extension

$$F_{PN} = \psi \circ \mathrm{Max}_f$$

$$F_{DS} = \psi \circ \mathrm{Ave}_f$$

$$\mathrm{Max}_{f_i}(A) = \max_{a \in A} f_i(a)$$

$$\mathrm{Ave}_{f_i}(\mu) = \mathbb{E}_{x \sim \mu}[f_i(x)]$$

$$\mathrm{Max}_{f_i} : (\mathcal{K}(\Omega), d_H) \to \mathbb{R}$$

$$\mathrm{Ave}_{f_i} : (\mathcal{P}(\Omega), d_W) \to \mathbb{R}$$

Space of nonempty compact subsets with Hausdorff metric d_H

Space of Borel probability measures with Wasserstein metric d_W

Intuition

$$p : \mathbb{Q} \to \mathbb{Q}$$

polynomial

Upgrade: Unique Continuous Extension

$$F_{PN} = \psi \circ \mathrm{Max}_f$$

$$F_{DS} = \psi \circ \mathrm{Ave}_f$$

$$\mathrm{Max}_{f_i}(A) = \max_{a \in A} f_i(a)$$

$$\mathrm{Ave}_{f_i}(\mu) = \mathbb{E}_{x \sim \mu}[f_i(x)]$$

$$\mathrm{Max}_{f_i} : (\mathcal{K}(\Omega), d_H) \to \mathbb{R}$$

$$\mathrm{Ave}_{f_i} : (\mathcal{P}(\Omega), d_W) \to \mathbb{R}$$

Space of nonempty compact subsets with Hausdorff metric d_H

Space of Borel probability measures with Wasserstein metric d_W

Intuition

$$p : \mathbb{R} \to \mathbb{R}$$

polynomial

Upgrade: Unique Continuous Extension

$$F_{PN} = \psi \circ \mathrm{Max}_f$$

$$\mathrm{Max}_{f_i}(A) = \max_{a \in A} f_i(a)$$

$$\mathrm{Max}_{f_i} : (\mathcal{K}(\Omega), d_H) \to \mathbb{R}$$

Space of nonempty compact subsets with Hausdorff metric d_H

$$F_{DS} = \psi \circ \mathrm{Ave}_f$$

$$\mathrm{Ave}_{f_i}(\mu) = \mathbb{E}_{x \sim \mu}[f_i(x)]$$

$$\mathrm{Ave}_{f_i} : (\mathcal{P}(\Omega), d_W) \to \mathbb{R}$$

Space of Borel probability measures with Wasserstein metric d_W

Intuition

$$p : \mathbb{C} \to \mathbb{C}$$

polynomial

Upgrade: Unique Continuous Extension

$$F_{PN} = \psi \circ \mathrm{Max}_f$$

$$F_{DS} = \psi \circ \mathrm{Ave}_f$$

$$\mathrm{Max}_{f_i}(A) = \max_{a \in A} f_i(a)$$

$$\mathrm{Ave}_{f_i}(\mu) = \mathbb{E}_{x\sim\mu}[f_i(x)]$$

$$\mathrm{Max}_{f_i} : (\mathcal{K}(\Omega), d_H) \to \mathbb{R}$$

Space of nonempty compact subsets with Hausdorff metric d_H

$$\mathrm{Ave}_{f_i} : (\mathcal{P}(\Omega), d_W) \to \mathbb{R}$$

Space of Borel probability measures with Wasserstein metric d_W

$$(\Omega, d) \text{ compact} \implies (\mathcal{K}(\Omega), d_H) \text{ and } (\mathcal{P}(\Omega), d_W) \text{ compact}$$

Stability of Extension

Theorem. *Suppose* $\Omega \subseteq \mathbb{R}^N$ *is compact. Then every PointNet and normalized-DeepSet network with Lipschitz continuous activation functions is Lipschitz continuous on* $(\mathcal{K}(\Omega), d_H)$ *and* $(\mathcal{P}(\Omega), d_W)$ *respectively.*

$$\|F_{PN}(A) - F_{PN}(B)\| \leq K_{F_{PN}} d_H(A, B)$$

$$\|F_{DS}(\mu) - F_{DS}(\nu)\| \leq K_{F_{DS}} d_W(\mu, \nu)$$

Classical UAT → Topological UAT

Theorem. *Let X be compact Hausdorff and $\sigma \in C(\mathbb{R})$ not a polynomial. If $S \subseteq C(X)$ separates points and has a nonzero constant, then* $\operatorname{span}(\sigma \circ \operatorname{span} S)$ *is dense in $C(X)$. If S is a linear subspace, then* $\operatorname{span}(\sigma \circ S)$ *is dense in $C(X)$.*

Topological UAT → UAT for Extension

Theorem. *Let X be compact Hausdorff and $\sigma \in C(\mathbb{R})$ not a polynomial. If $S \subseteq C(X)$ separates points and has a nonzero constant, then* $\operatorname{span}(\sigma \circ \operatorname{span} S)$ *is dense in $C(X)$. If S is a linear subspace, then* $\operatorname{span}(\sigma \circ S)$ *is dense in $C(X)$.*

Letting $S_{PN} = \{\mathrm{Max}_f \mid f \in \mathcal{N}^\tau\}$ and $S_{DS} = \{\mathrm{Ave}_f \mid f \in \mathcal{N}^\tau\}$ works!

This yields a UAT for generalized PointNet and DeepSets on $\mathcal{K}(\Omega)$ and $\mathcal{P}(\Omega)$.

UAT for Extension → Point Cloud UAT

Theorem. *Let X be compact Hausdorff and $\sigma \in C(\mathbb{R})$ not a polynomial. If $S \subseteq C(X)$ separates points and has a nonzero constant, then* $\operatorname{span}(\sigma \circ \operatorname{span} S)$ *is dense in $C(X)$. If S is a linear subspace, then* $\operatorname{span}(\sigma \circ S)$ *is dense in $C(X)$.*

Letting $S_{PN} = \{\mathrm{Max}_f \mid f \in \mathcal{N}^\tau\}$ and $S_{DS} = \{\mathrm{Ave}_f \mid f \in \mathcal{N}^\tau\}$ works!

This yields a UAT for generalized PointNet and DeepSets on $\mathcal{K}(\Omega)$ and $\mathcal{P}(\Omega)$.

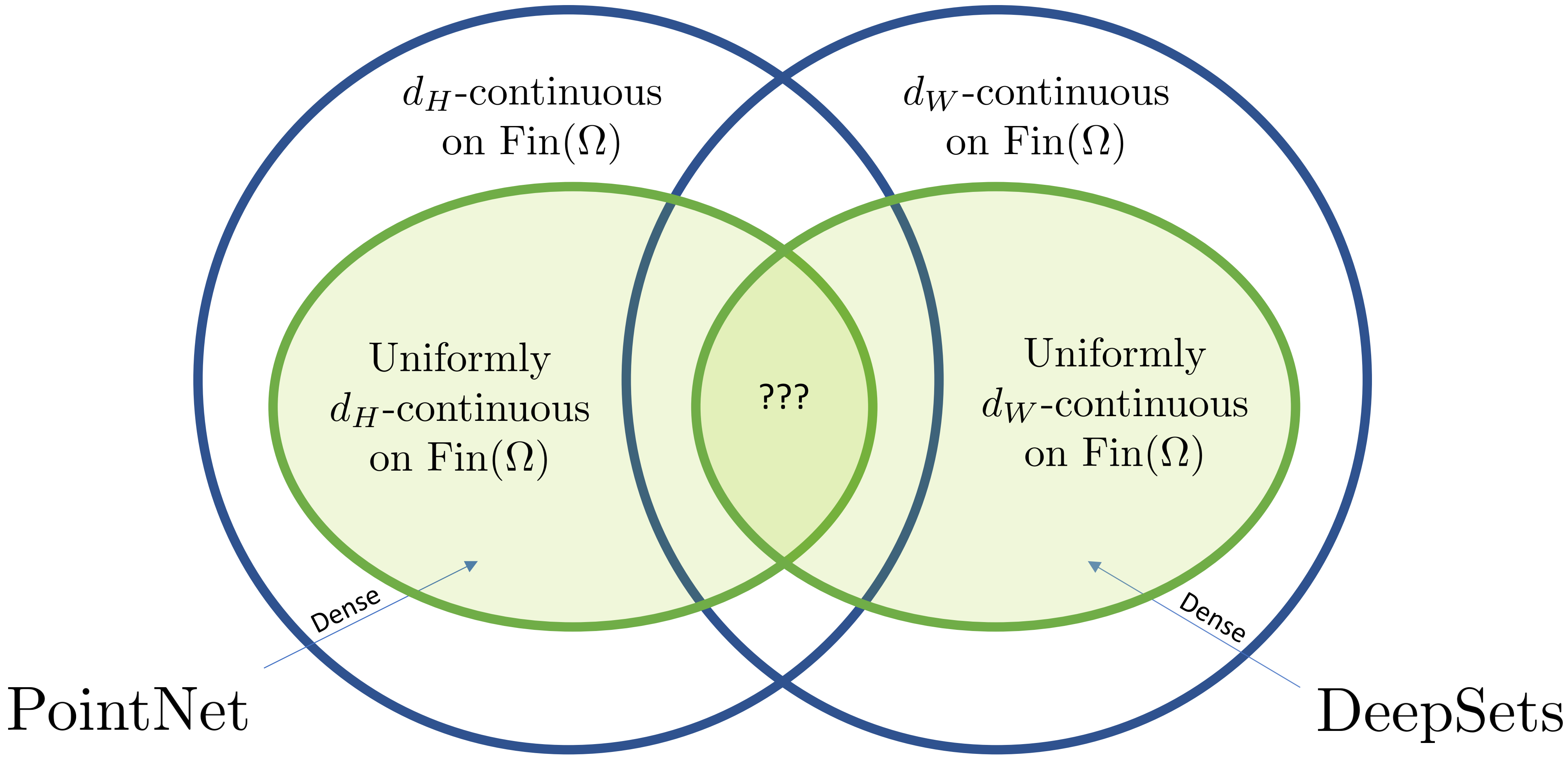
Theorem.

Restricting to $\mathrm{Fin}(\Omega)$

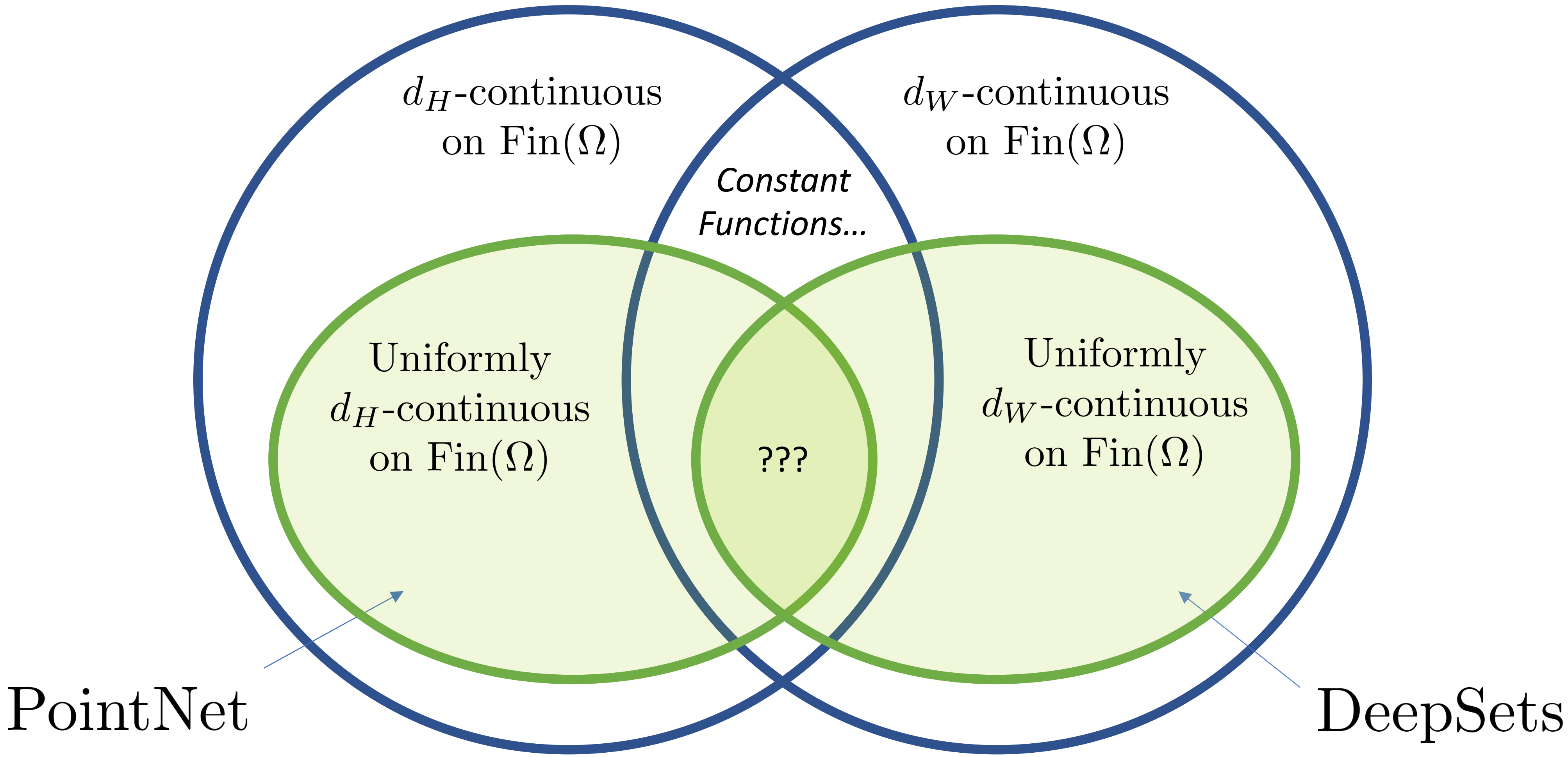
$\operatorname{span}(\sigma \circ \operatorname{span} S_{PN})$ → Uniform Closure → $\mathrm{UnifCont}(\mathrm{Fin}(\Omega), d_H)$

$\operatorname{span}(\sigma \circ S_{DS})$ → Uniform Closure → $\mathrm{UnifCont}(\mathrm{Fin}(\Omega), d_W)$

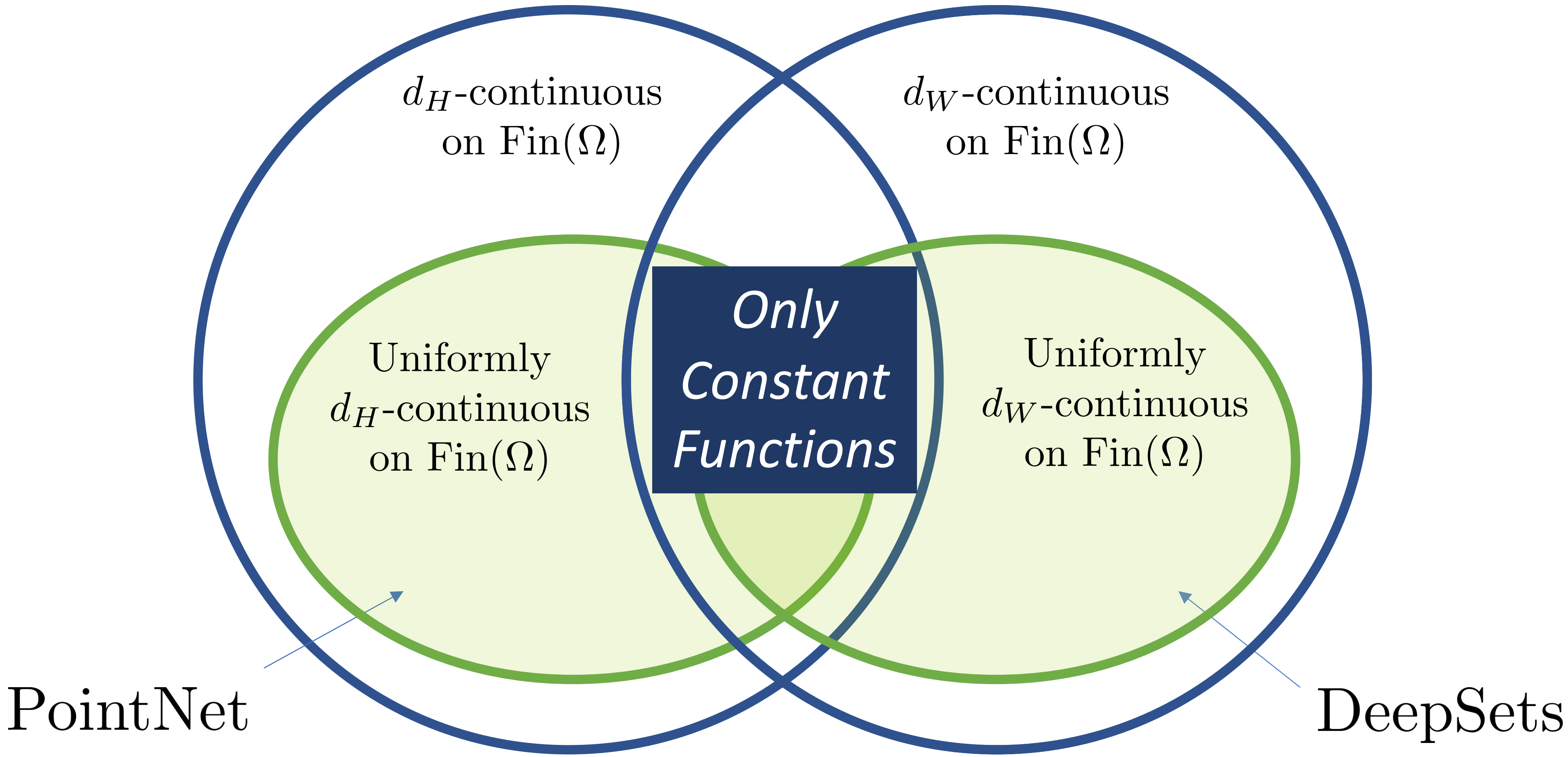
The Overlap and Limitations



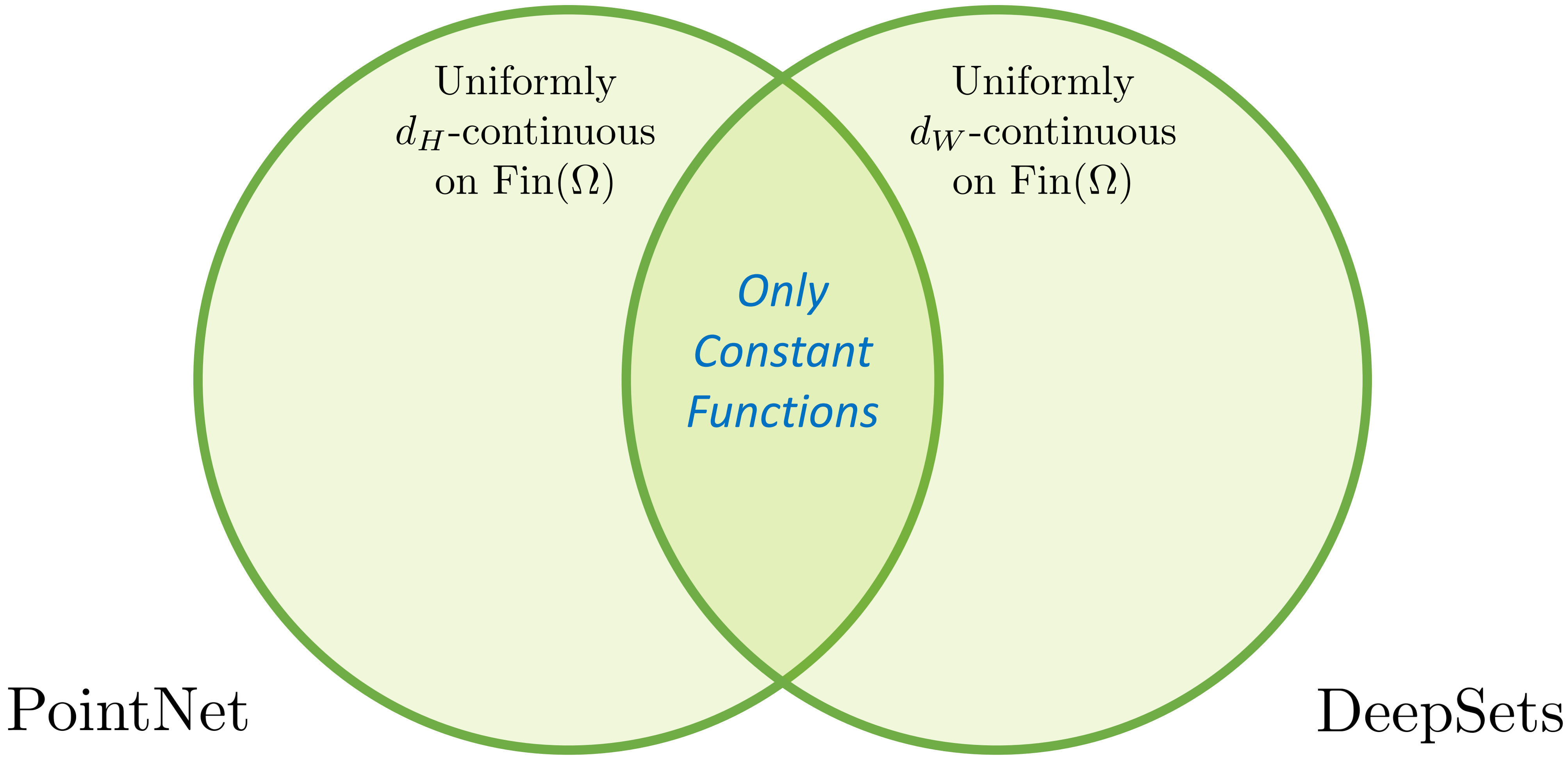
The Overlap and Limitations



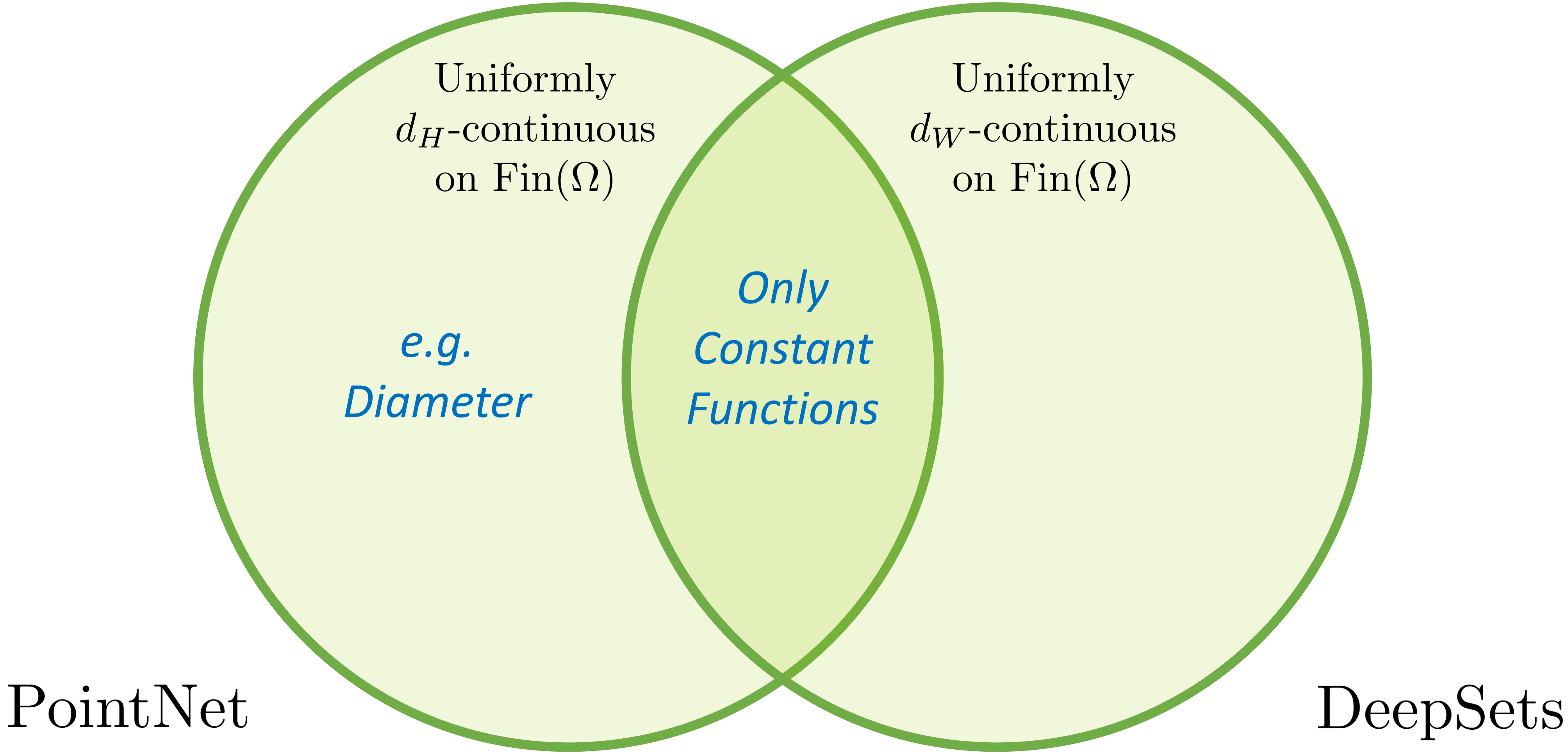
The Overlap and Limitations



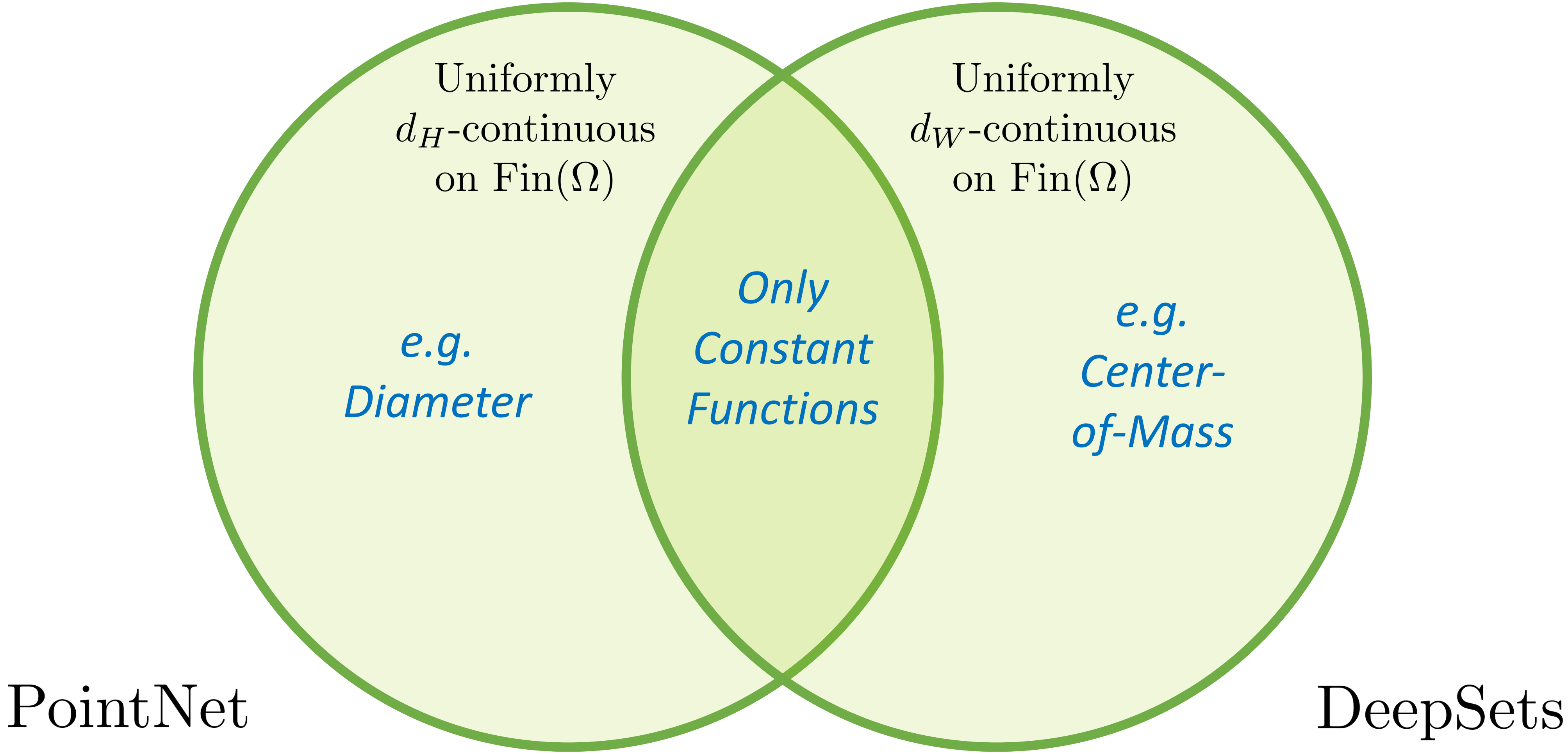
The Overlap and Limitations



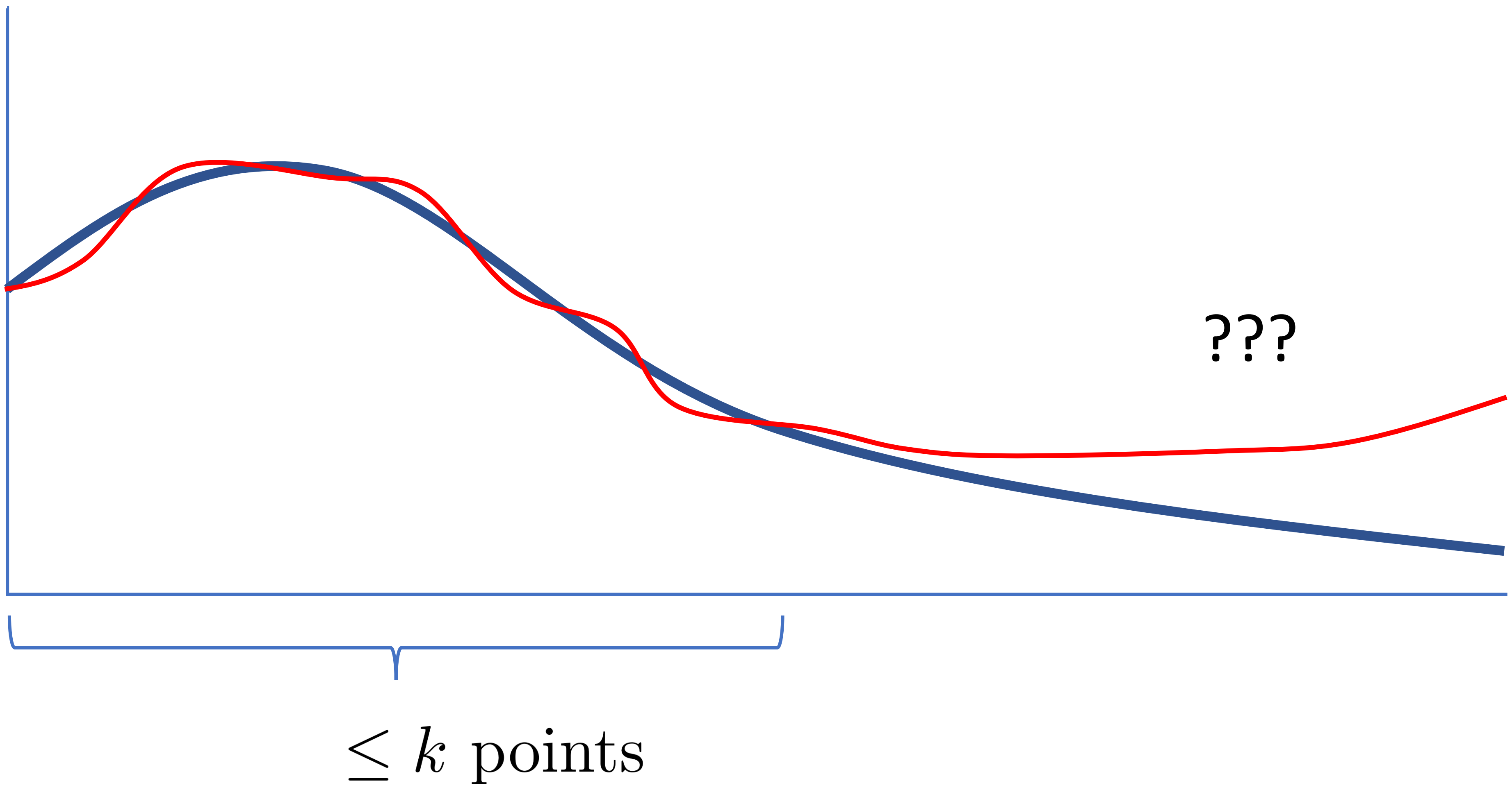
The Overlap and Limitations



The Overlap and Limitations

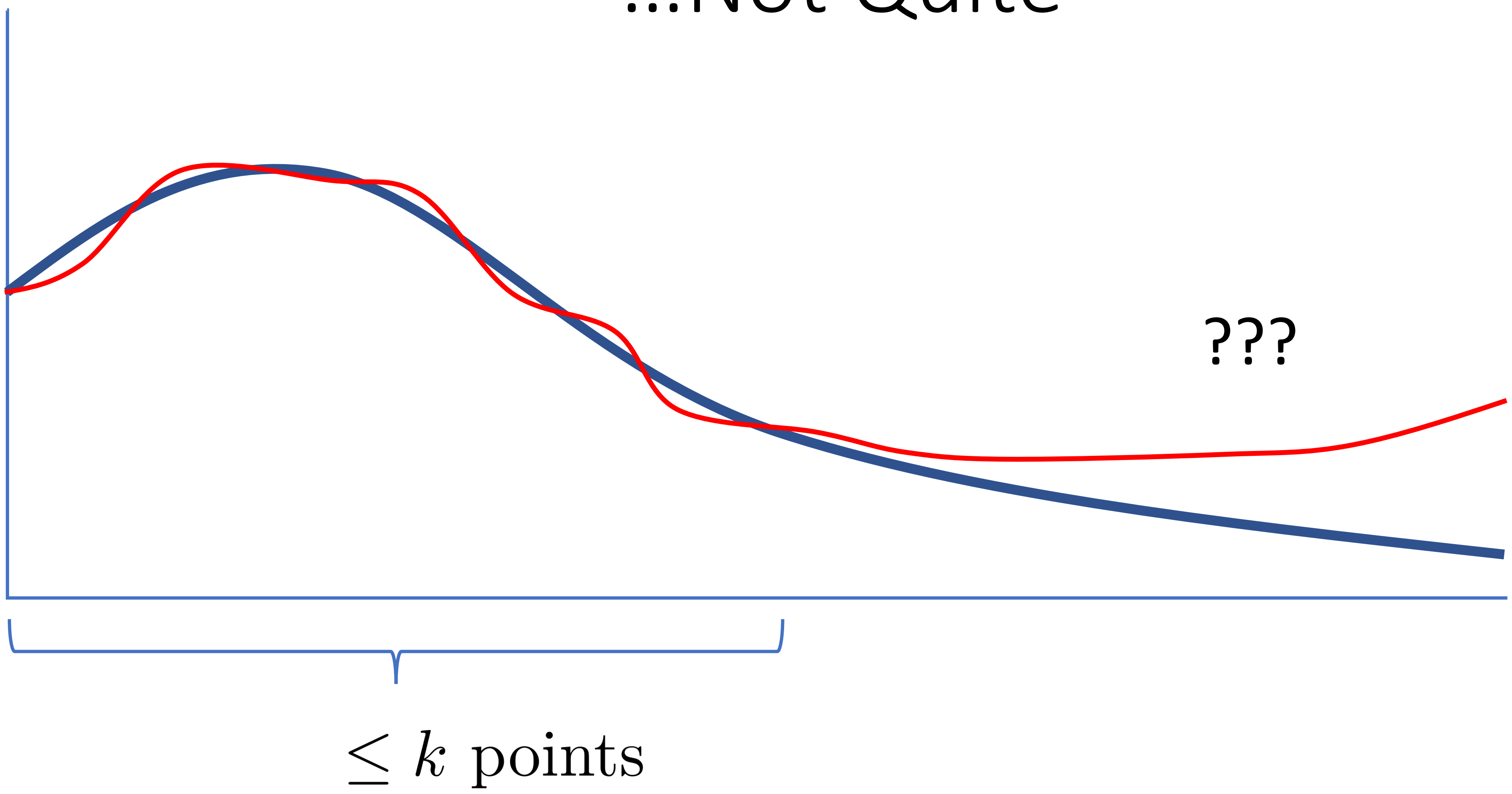


Is the Problem at Infinity?



Is the Problem at Infinity?

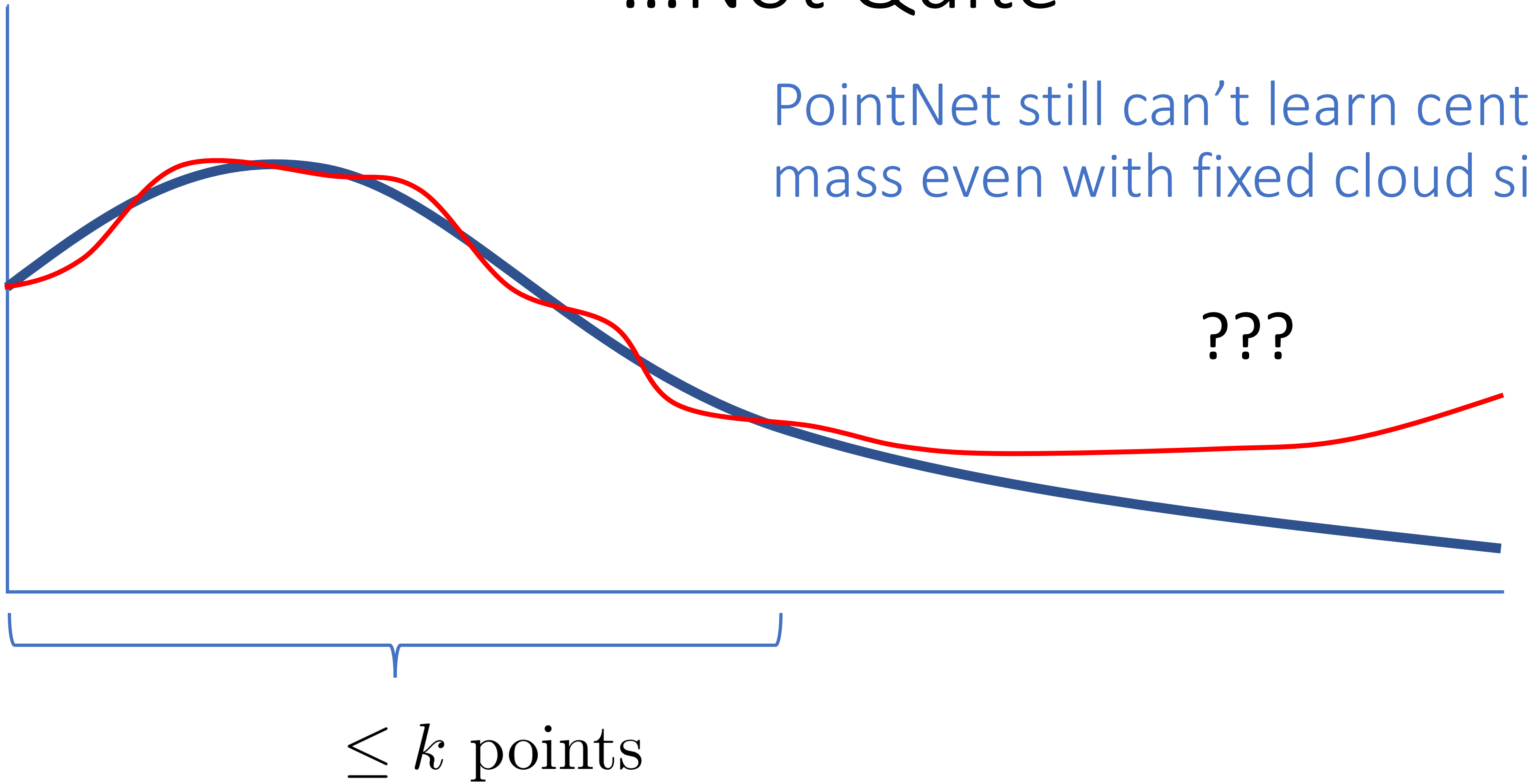
...Not Quite



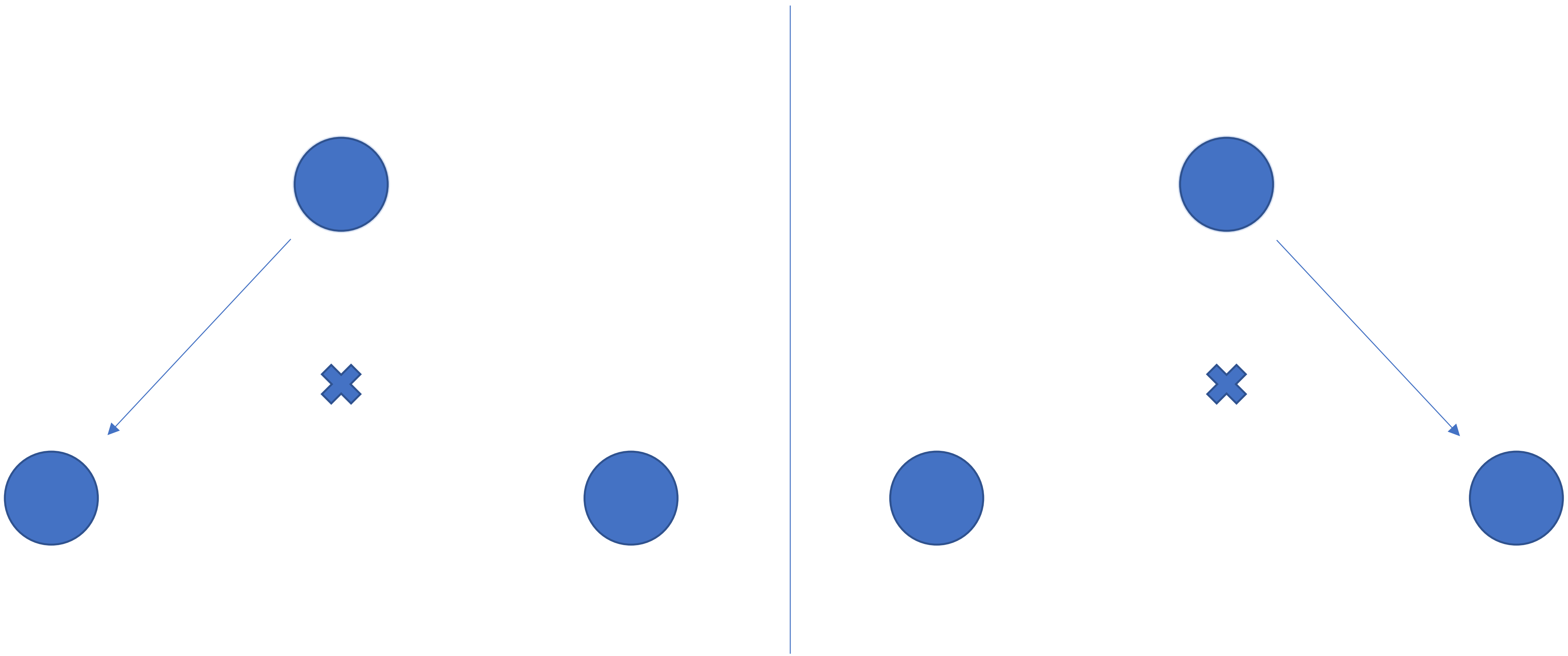
Is the Problem at Infinity?

...Not Quite

PointNet still can't learn center-of-mass even with fixed cloud size.



Center-of-Mass, PointNet, & Fixed Size Sets



Two d_H-continuous paths with same limit…
…But different limiting centers.

Error Lower Bound for $\mathbf{ave}_f$

Theorem. *Let* $\Omega \subseteq \mathbb{R}^n$ *be the unit ball,* $k \geq 3$*, and* $f : \Omega \to \mathbb{R}^n$ *continuous. Then for any distinct* $p, q \in \Omega$ *and* $0 < \tau < 1$ *there exists a* k*-point set* A *with* $p, q \in A \subseteq \Omega$ *so that*

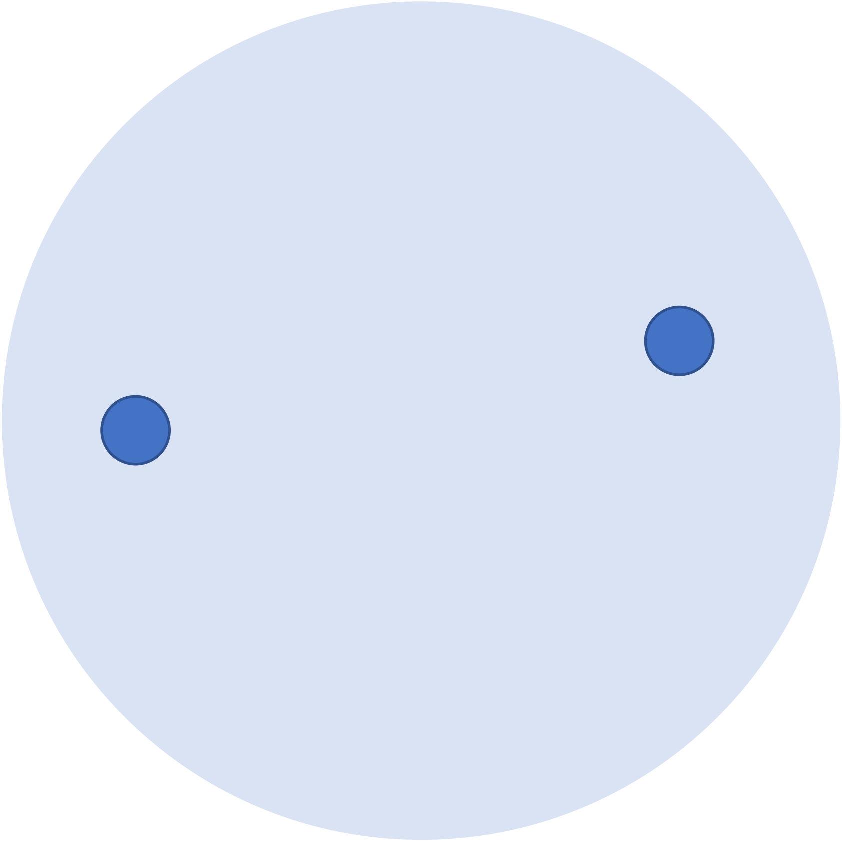
$$\|F_{PN}(A) - \mathrm{ave}_f(A)\| > (1 - \tau)\left(\frac{k-2}{2k}\right)\|f(p) - f(q)\|$$

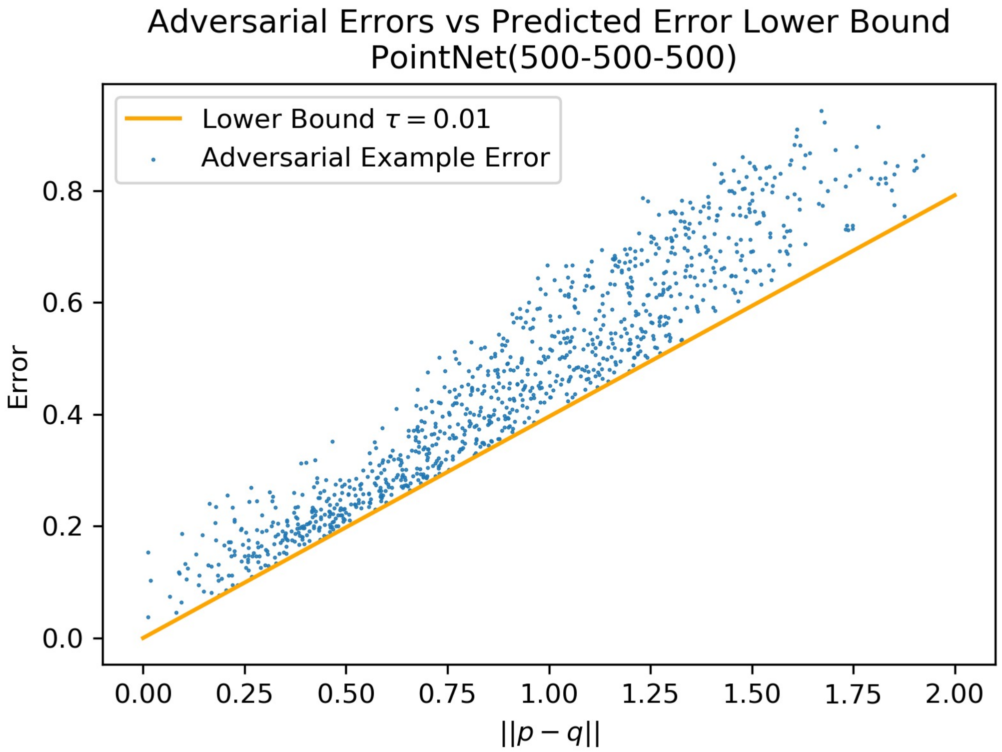
for any PointNet-type F_{PN}*, regardless of depth/width/training/etc. Thus,*

$$\|F_{PN} - \mathrm{ave}_f\|_{L^\infty(\mathrm{Fin}^k(\Omega))} \geq \left(\frac{k-2}{2k}\right)\mathrm{Diam}(f(\Omega))$$

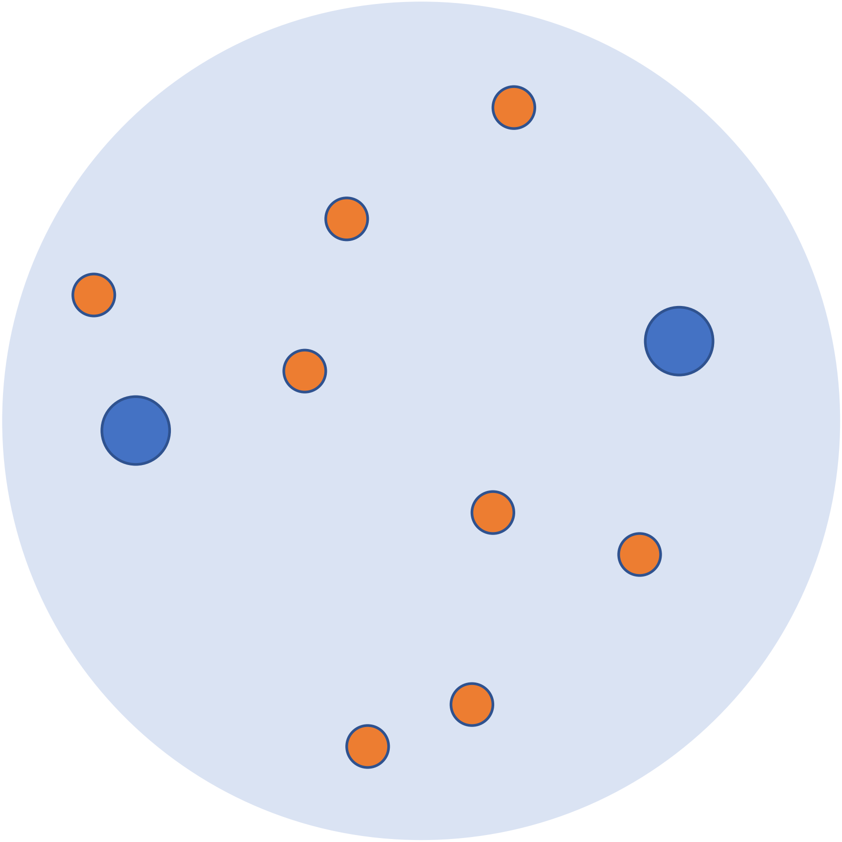
Moreover, we can construct such geometric "adversarial" examples

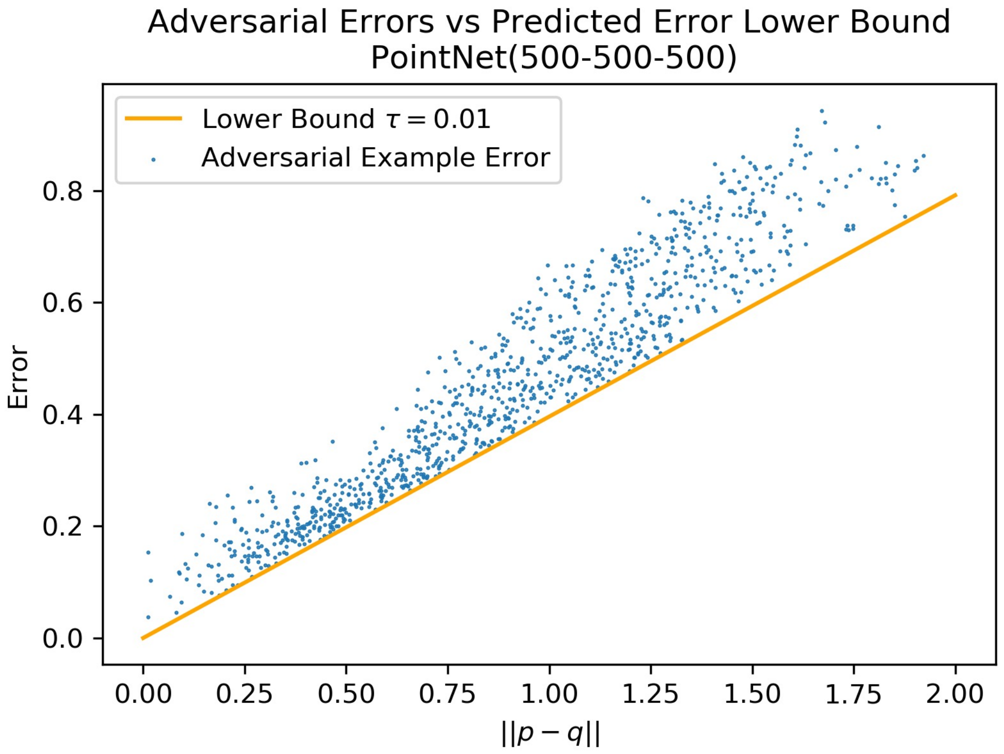
Test of Error Lower Bound (Center-of-Mass)



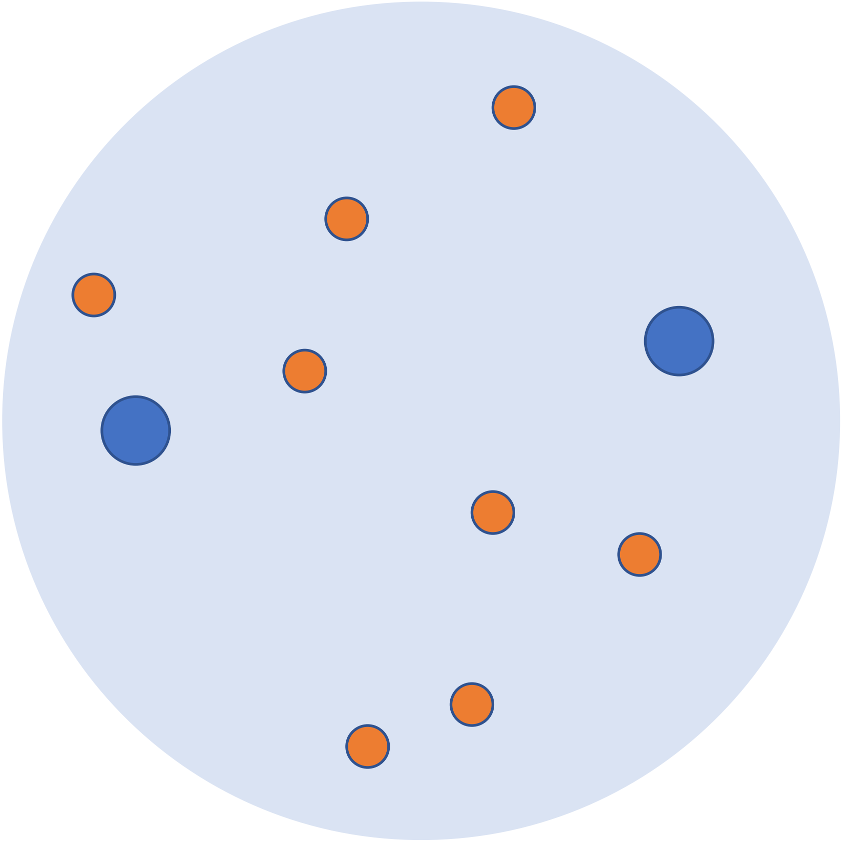


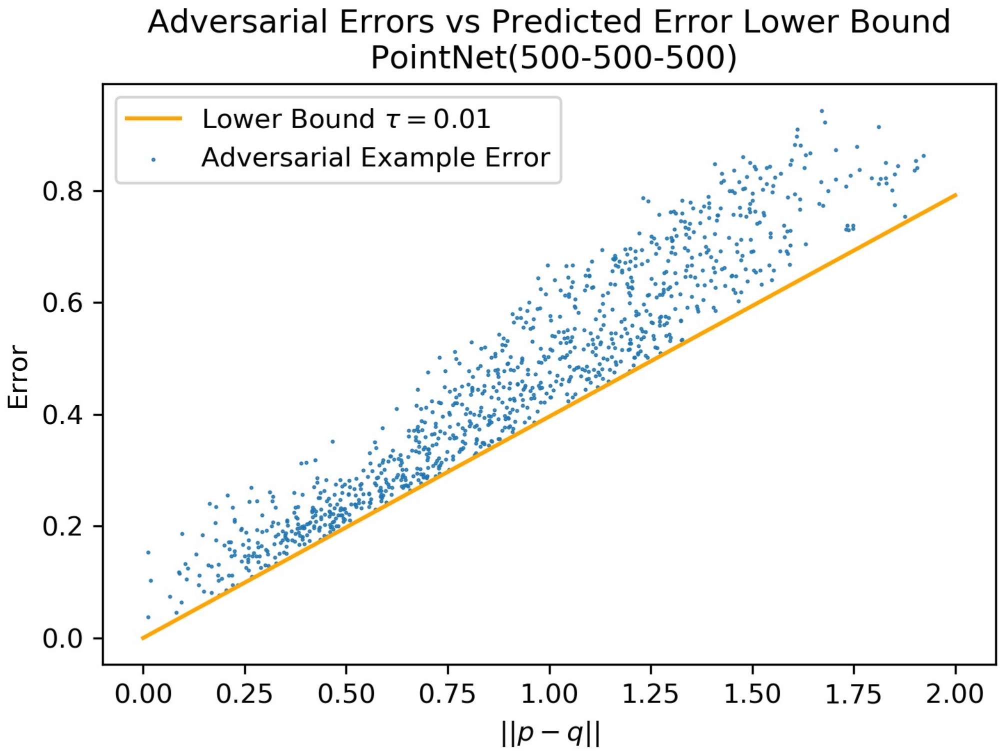
Test of Error Lower Bound (Center-of-Mass)



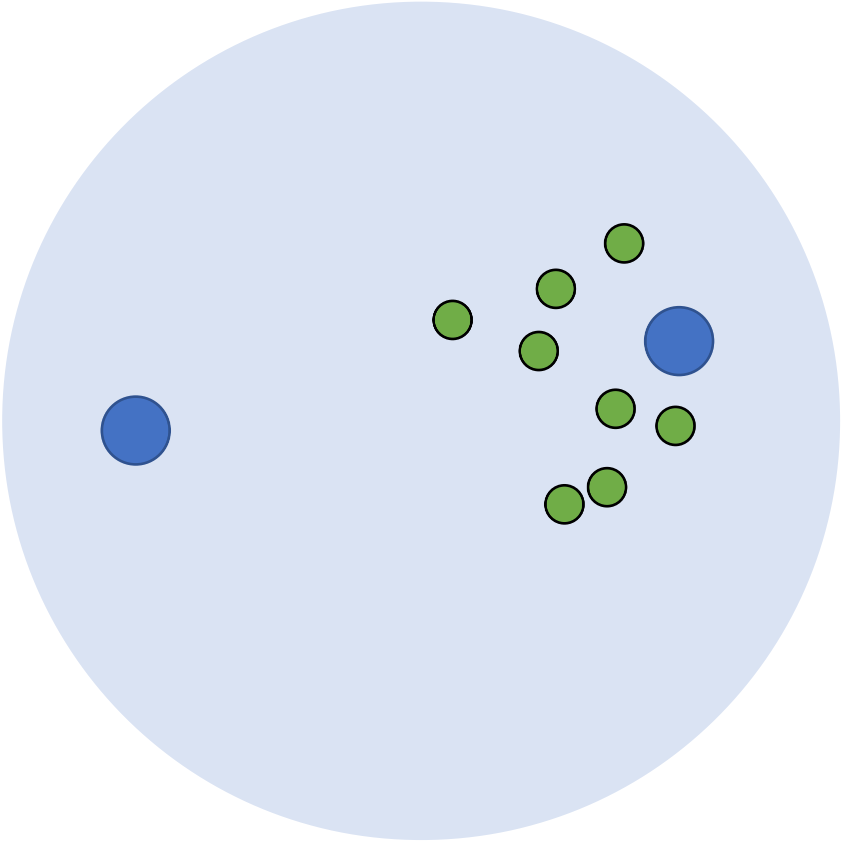


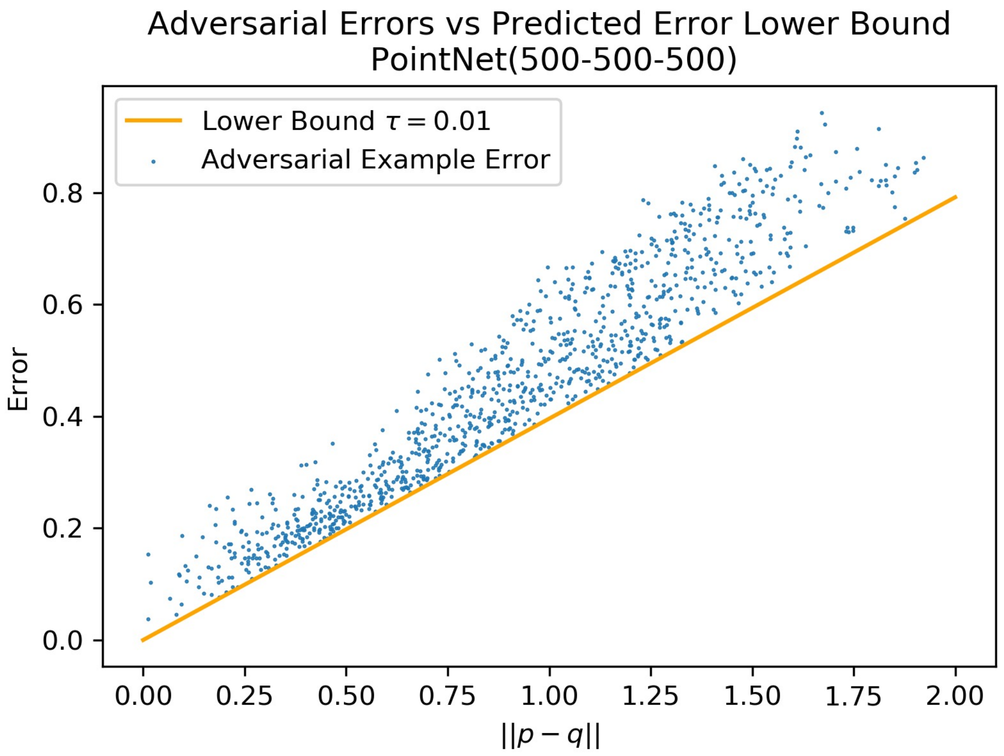
Test of Error Lower Bound (Center-of-Mass)





Test of Error Lower Bound (Center-of-Mass)





Summary

- PointNet & normalized-DeepSets uniquely continuously extend to $\mathcal{K}(\Omega)$ and $\mathcal{P}(\Omega)$ respectively.
- PointNet & normalized-DeepSets can uniformly approximate the uniformly continuous functions on $\text{Fin}(\Omega)$ w.r.t. d_H and d_W resp. They cannot uniformly approximate anything else.
- PointNet & normalized-DeepSets are Lipschitz if activations are Lipschitz
- Constants are only functions mutually approximable by PointNet and DeepSets on $\text{Fin}(\Omega)$.
- PointNet cannot uniformly approximate averages of continuous functions (*even for fixed cloud size*) and geometric adversarial examples are abundant and easily constructed.

References

Michael Fielding Barnsley. *Superfractals*. Cambridge University Press, 2006.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. *arXiv preprint arXiv:1806.07366*, 2018.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. *arXiv preprint arXiv:1806.07572*, 2018.

Moshe Leshno, Vladimir Ya Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks*, 6(6):861–867, 1993.

Ernest Michael. Topologies on spaces of subsets. *Transactions of the American Mathematical Society*, 71(1):152–182, 1951.

James Raymond. Munkres. *Topology*. Prentice Hall, 2000.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–660, 2017a.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5099–5108. Curran Associates, Inc., 2017b.

Walter Rudin. *Real and complex analysis*. Tata McGraw-hill education, 2006.

M.b. Stinchcombe. Neural network approximation of continuous functionals and continuous functions on compactifications. *Neural Networks*, 12(3):467–477, 1999. doi: 10.1016/s0893-6080(98)00108-7.

Cedric Villani. *Optimal transport: old and new*. Springer, 2009.

Edward Wagstaff, Fabian B Fuchs, Martin Engelcke, Ingmar Posner, and Michael Osborne. On the limitations of representing functions on sets. *arXiv preprint arXiv:1901.09006*, 2019.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.

Dmitry Yarotsky. Universal approximations of invariant maps by neural networks. *arXiv preprint arXiv:1804.10306*, 2018.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan R Salakhutdinov, and Alexander J Smola. Deep sets. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 3391–3401. Curran Associates, Inc., 2017. URL http://papers.nips.cc/paper/6931-deep-sets.pdf.

Thank You!