RESEARCH STATEMENT

EDOARDO LANARI

Overview

My main scientific interests are **Homotopy Theory** and **Higher Category Theory**, and, more recently, **Machine Learning** (in particular **Deep Learning**) and the contributions of **Topological Data Analysis** (TDA) to this field.
Higher categorical structures have proven to be the correct setting in which to develop homotopy-coherent mathematics. In the past two years, I have contributed significantly to the field, laying the foundations for the theory of $(\infty, 2)$-categories. This is an ongoing project, whose achievements and future goals I will describe in the following section, and which has resulted so far in [17, 16, 15, 18].

In addition to this, I am also involved in another project on TDA (whose first contribution is [27]), and it is my goal to further investigate how this new and exciting field in applied topology can become a fundamental part of the Machine Learning pipeline. For instance, this may come in the form of providing more complex architectures for neural networks, or by improving the explainability of real-world models.

Research[1]

Homotopy Theory/Higher Category Theory[2]. So far, the main focus of my research has been the development of the theory of $(\infty, 2)$-categories. Roughly speaking, (∞, n)-categories can be understood as weak categorical structures with cells in dimension k for $0 \leq k \leq \infty$, where all the k-cells with $k > n$ are invertible. In particular, $(\infty, 2)$-categories are the homotopy-coherent analogue of 2-categories, and as such are the correct framework where to express homotopy-coherent mathematics (*i.e.* ∞-category theory).

Motivations and Applications. The relevance of this work to other areas of Mathematics is well-established: most notably we can back this claim with two contributions from **(Derived) Algebraic Geometry** and **Goodwillie Calculus**. An example of the former is of course the work by Gaitsgory and Rozenblyum ([19]) aimed at proving the **geometric Langland conjecture**. This work relies heavily on the use of $(\infty, 2)$-categories to model **correspondences**, which are the main ingredient of their (arguably) most important results, *i.e.* the extension theorems of Chapter 8.

On the other hand, Lurie develops Goodwillie calculus in the context of $(\infty, 2)$-categories in [28], his seminal work that has been a great influence for me. There he uses $(\infty, 2)$-categories as the foundation for expressing the theory underlying Goodwillie Calculus, together with applications to first derivatives of functors.

Main results achieved thus far.

- the **equivalence** of all simplicial models for $(\infty, 2)$-categories. In [17] we solve the long-standing problem of whether the complicial model introduced by Verity in [43] and [42], and further investigated by Rovelli and Ozornova in [31] and [32], fits in the established web of equivalences of models for $(\infty, 2)$-categories (see Figure 1 of [17]).
- we adapt the **Gray tensor product** for complicial sets (see [43]) to the context of scaled simplicial sets in [16], and we introduce (normal) lax functors of $(\infty, 2)$-categories. We then establish a characterization of the Gray tensor product in terms of these maps, thus **proving a conjecture** by Gaitsgory and Rozenblyum formulated in [19].
- we introduced the notion of **fibrations** of $(\infty, 2)$-categories in [15], a fundamental piece of the theory that was not available in any of the previously known models. Leveraging on this notion we develop a theory that encompasses a full-fledged **(Un)Straightening theorem**, and we introduce **lax** and **weighted** (co)limits of $(\infty, 2)$-categories, generalizing the work of [40], [38] and [20]. Finally, we also develop a theory of **(co)final** functors that generalizes the work of [1] and part of [11].

[1] This is only a summary of the most recent work, for an exhaustive description please see my publications/preprints list.

[2] Collaborators for this project: A. Gagna and Y. Harpaz

- we show in [18] that, even though it was proven by [9] that **(lax, weighted) bilimits** cannot be expressed as "terminal cones" (in analogy with the ordinary 1-dimensional case), there is still a characterization in terms of "**final cones**". This makes use of the notion of **contraction**, borrowed from (strict) ∞-category theory (see [2]).

Future plans.

- Develop the theory of **2-fibrations**, **(formal) cofinality** and **flat 2-functors**[3] in the context of $(\infty, 2)$-categories. 2-fibrations over an $(\infty, 2)$-category $\mathcal{E}$ are (conjecturally) equivalent to maps of $(\infty, 2)$-categories of the form $\mathcal{E} \to \mathcal{C}at_{(\infty,2)}$, and constitute a fundamental building block of higher category theory. Formal cofinality refers to the optimal setting in which one can formalize the notion of cofinality, which significantly simplifies working with (co)limit-like constructions. Finally, flat 2-functors are developed for 2-categories in [11] and constitute the **points** of a **(Grothendieck)** $(\infty, 2)$**-topos** (see next bullet point). They are also crucial for defining a $2-\mathrm{Ind}$ **construction** for $(\infty, 2)$-categories, *i.e.* the free cocompletion under filtered colimits. This, in turn, has interesting geometrical applications and it is another direction of research that I intend to pursue.
- **(Elementary)** $(\infty, 2)$**-topoi**. This project can be thought of as the natural generalization of the work done in [37], [36] for the elementary case and in [29] and [41] for the Grothendieck case. This requires investigating the structure that ought to underlie an $(\infty, 2)$-topos, and (among other things) to define **universal fibrations** using results from [15]. Moreover, the plan is to generalize the theory of **localizations** of ordinary 2-categories (see [33], [34] and [35] for reference) to the context of $(\infty, 2)$-categories, in order to be able to define $(\infty, 2)$**-sheaves** (or $(\infty, 2)$**-stacks**). These have applications to the study of (derived) moduli spaces and descent theory, in analogy with what happens in the classical case.

Topological Data Analysis and Machine Learning. My interest in TDA comes from two main motivations: first, the interest in the field itself and my love for Algebraic Topology since my Bachelor years, and, more recently, its connections to Machine Learning (ML), in particular Deep Learning (DL). I will now outline the work I have done so far and future plans, highlighting this twofold source of motivation.

Motivations and Applications. TDA offers topological tools to investigate complex and high-dimensional datasets, highlighting the importance of the **shape** of data. For instance, persistent homology allows to grasp information about the distribution of data even when the high-dimensionality renders visualization and other techniques unfeasible. Classically, one of the shortcomings of statistical analysis is the **curse of dimensionality** (see for example Section 1.4 of [5]), and TDA addresses, among other things, possible solutions that can also scale up to the size of modern datasets.

All algebraic invariants studied in TDA ought to satisfy some **stability to perturbations** property, since real-world data is necessarily going to be noisy, and so the techniques employed must accomodate for this variance issue. Stability to noise is usually measured with respect to the **Gromov–Hausdorff** distance between compact metric spaces (see for example Section 1 of [6] for a relevant discussion). One direction I want to pursue for this ongoing project, which has already resulted in a paper ([27]), is to generalize our work on the homotopy interleaving distance of persistent spaces to sheaves and cosheaves, which are essential in studying local structures and properties. The **homotopy interleaving distance** is a natural metric to consider when dealing with stability of homotopy-invariant constructions, so this direction of work is inherently connected to noise-stability and also has practical applications.

Turning to ML, recently there has been a significant amount of work in developing and implementing **simplicial/cellular architectures for neural networks**, e.g. [7, 12, 21, 23] to name a few. The underlying idea is to model the **flow of information** on a combinatorial structure that is inherently more complex than the usual **hypergraph architecture** of a Deep Neural Network (DNN). Simplicial complexes can be employed to capture higher dimensional information beyond nodes and edges, and one of the problems addressed in the abovementioned papers is how to encode the flow of informations from simplices to their faces. Some further connections between TDA and DL have been investigated in [13] and [30].

One of the main motivations for my interest in DL, besides the firm belief that the next wave of enhancements in this field will primarily come from Mathematics, are its applications to the world of **Quantitative Finance** ([8, 10]). In this context, the gold standard for time series analysis is represented by Recurrent Neural Networks (RNN), and in particular the **Long Short-Term Memory** model (LSTM), see [14] and [26] for performance comparison with ordinary DNNs, Boosted-Trees and Random

[3] This third topic will be investigated with a fourth collaborator I. Di Liberti

Forests models. Recently, **Convolutional Neural Networks** (CNN) have also been proposed as models, see [39] for reference, with promising results.

Main results achieved thus far.

- In [27] we consider several distances introduced in [6] and [25] between (multi)persistent diagrams in a model category $\mathcal{M}$, with different degree of coherence. These are, respectively, the **homotopy interleaving distance** on $\mathcal{M}^{\mathbb{R}^m}$ (denoted by d_{HI}), the **interleaving distance in the homotopy category** on $\mathrm{Ho}(\mathcal{M}^{\mathbb{R}^m})$ (denoted by d_{IHC}) and the **homotopy commutative interleaving distance** on $\mathrm{Ho}(\mathcal{M})^{\mathbb{R}^m}$ (denoted by d_{HC}). One of the main results in the paper is that we have
$$3d_{HC} \geq d_{HI} \geq d_{IHC} \geq d_{HC}$$
and $d_{HC} \neq d_{HI}$ in the case of $\mathcal{M} = \mathbf{Top}$. This chain of inequalities shows that, for many purposes, these distances are **equivalent**.
- We also prove a **Persistent Whitehead Theorem**, conjectured by Blumberg and Lesnick in [6], which states that an interleaving between homotopy groups can be lifted to an interleaving in the homotopy category of persistent spaces. This is essentially the persistent counterpart of the classical Whitehead theorem in homotopy theory.

Future plans.

- Investigate **persistent sheaves**, defined over the site of open subsets of $\mathbb{R}$ (or, more generally $\mathbb{R}^m$ or even a nice enough metric space). This framework has been introduced in [25], where they also construct a (stable) distance between **persistent sheaves of chain complexes**. This distance is defined at the level of the **derived category**, and a non-trivial task is that of lifting this distance to a homotopy-interleaving distance d_{HI} on the actual category $\mathbf{Sh}(\mathbb{R}, \mathbf{Ch}_k)$ of persistent sheaves, where k is a given field. This could also be pushed further, and part of the project involves actually lifting this distance to d'_{HI}, a stable homotopy interleaving distance defined on **persistent sheaves of spaces**, so as to turn the chain complex functor $\mathbb{Z}[\bullet]\colon \mathbf{Sh}(\mathbb{R}, \mathbf{Top}) \to \mathbf{Sh}(\mathbb{R}, \mathbf{Ch}_k)$ into a contraction, i.e. a functor satisfying $d'_{HI}(X,Y) \leq d_{HI}(\mathbb{Z}[X], \mathbb{Z}[Y])$. In [4] the authors prove an isometry theorem by expressing the convolution distance of sheaves as a matching distance between combinatorial objects associated to them (called **graded barcodes**).

  Another part of the project is to consider **persistent cosheaves of sets** and their associated **Reeb Graphs** (see [3] for reference), which make use of ideas coming from Morse theory, to analyze filtered spaces presented by continuous maps $X \to \mathbb{R}$. Reeb graphs are combinatorial objects associated with such cosheaves, and in [3] the authors introduce stable distances in this framework. Again, the natural thing to do is to try to lift this distances to a (stable) homotopy-interleaving distance at the level of cosheaves of spaces, and then relate it (via a functorial contraction, as in the previous paragraph) to the Reeb distance. Finally, one can study how these different distances are related when applied to a map $p\colon X \to \mathbb{R}$, by looking at its sheaf of sections $(\mathbb{R} \supset U \mapsto \{f\colon U \to X \mid pf = \mathrm{Id}_{\mathbb{R}}\})$ and cosheaf of pre-images $(\mathbb{R} \supset U \mapsto p^{-1}(U))$.
- Analyze and push further the work done on simplicial architectures for DNNs and CNNs, and consider possible extensions and implementations in the context of RNNs architectures. Furthermore, in the context of time series, data comes naturally endowed with a notion of **orientation**, given by timestamps, and this consequently leads to consider more refined structures than bare simplicial complexes, in analogy with the refinement of ordinary graphs to directed ones. Since **simplicial complexes** admit embeddings in the category of **simplicial sets**, e.g. as in [24], we can look at endowing these simplicial sets with further structure. Among natural candidates feature **stratified sets**, both in the sense of [22] and [43] (note that these are conceptually different, but in some sense they both encode **directionality**).

References

1. Fernando Abellán García, *Marked cofinality for ∞-functors*, arXiv:2006.12416, 2020, preprint.
2. Dimitri Ara and Georges Maltsiniotis, *The homotopy type of the ∞-category associated to a simplicial complex*, arXiv:1503.02720, 2015, preprint.
3. U. Bauer, C. Landi, and F. Mémoli, *The reeb graph edit distance is universal*, Foundations of Computational Mathematics **270** (2020), no. 2, 654–669.
4. Nicolas Berkouk and Grégory Ginot, *A derived isometry theorem for constructible sheaves on* $\mathbb{R}$, arXiv1805.09694, 05 2018, preprint.
5. Christopher M. Bishop, *Pattern recognition and machine learning (information science and statistics)*, Springer-Verlag, Berlin, Heidelberg, 2006.
6. Andrew J. Blumberg and Michael Lesnick, *Universality of the homotopy interleaving distance*, 2017.
7. Eric Bunch, Qian You, Glenn Fung, and Vikas Singh, *Simplicial 2-complex convolutional neural networks*, arXiv2012.06010, 2020, preprint.

8. Ernie Chan, *Algorithmic trading: Winning strategies and their rationale*, 1st ed., Wiley Publishing, 2013.
9. Tslil Clingman and Lyne Moser, *2-limits and 2-terminal objects are too different*, arXiv:2004.01313, 2020, preprint.
10. Marcos Lopez de Prado, *Advances in financial machine learning*, 1st ed., Wiley Publishing, 2018.
11. Maria Descotte, Eduardo Dubuc, and Martín Szyld, *Sigma limits in 2-categories and flat pseudofunctors*, Advances in Mathematics **333** (2018), 266–313.
12. Stefania Ebli, Michael Defferard, and Gard Spreeman, *Simplicial neural networks*, arXiv2010.03633, 2020, preprint.
13. David Fernández, Asier Gutiérrez-Fandiño, Jordi Armengol-Estapé, and Marta Villegas, *Determining structural properties of artificial neural networks using algebraic topology*, 01 2021.
14. Thomas Fischer and Christopher Krauss, *Deep learning with long short-term memory networks for financial market predictions*, European Journal of Operational Research **270** (2018), no. 2, 654–669.
15. Andrea Gagna, Yonatan Harpaz, and Edoardo Lanari, *Fibrations and lax limits of $(\infty, 2)$-categories*, 2012.04537, 2020, preprint.
16. ______, *Gray tensor products and lax functors of $(\infty, 2)$-categories*, to appear in Advances in Mathematics (2020), available at arXiv:2006.14495.
17. ______, *On the equivalence of all models for $(\infty, 2)$-categories*, arXiv:1911.01905, 2020, preprint.
18. ______, *Bilimits are final object*, arXiv:1911.01905, 2021, preprint.
19. Dennis Gaitsgory and Nick Rozenblyum, *A study in derived algebraic geometry I & II*, Mathematical Surveys and Monographs, vol. 221, American Mathematical Soc., 2017.
20. David Gepner, Rune Haugseng, and Thomas Nikolaus, *Lax colimits and free fibrations in ∞-categories*, Doc. Math. **22** (2017), 1225–1266.
21. Celia Hacker, *k-simplex2vec: a simplicial extension of node2vec*, arXiv2010.05636, 2020, preprint.
22. Peter J. Haine, *On the homotopy theory of stratified spaces*, arXiv1811.01119, 11 2018, preprint.
23. Mustafa Hajij, Kyle Istvan, and Ghada Zamzmi, *Cell complex neural networks*, arXiv2010.00743, 2020, preprint.
24. JF Jardine, *Simplicial approximation*, Theory Appl. Categ **12** (2004), no. 2, 34–72.
25. Masaki Kashiwara and Pierre Schapira, *Persistent homology and microlocal sheaf theory*, Journal of Applied and Computational Topology **2** (2018), no. 1-2, 83–113.
26. Christopher Krauss, Xuan Anh Do, and Nicolas Huck, *Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500*, FAU Discussion Papers in Economics 03/2016, Nürnberg, 2016.
27. Edoardo Lanari and Luis Scoccola, *Rectification of interleavings and a persistent whitehead theorem*, arXiv:2010.05378, 2020, preprint.
28. Jacob Lurie, *$(\infty, 2)$-categories and the Goodwillie Calculus I*, arXiv:0905.0462, 2009, preprint.
29. ______, *Higher Topos Theory*, Annals of Mathematics Studies, vol. 170, Princeton University Press, Princeton, NJ, 2009.
30. Guido Montufar, Nina Otter, and Yu Guang Wang, *Can neural networks learn persistent homology features?*, arXiv2010.05636, 11 2020, preprint.
31. Victoriya Ozornova and Martina Rovelli, *Model structures for (∞, n)-categories on (pre)stratified simplicial sets and prestratified simplicial spaces*, Algebr. Geom. Topol. **20** (2020), no. 3, 1543–1600.
32. Viktoriya Ozornova, Martina Rovelli, and Dominic Verity, *Gray tensor product and saturated n-complicial sets*, arXiv:2007.01235, 2020, preprint.
33. Dorette Pronk, *Etendues and stacks as bicategories of fractions*, Compositio Mathematica **102** (1996).
34. Dorette Pronk and Laura Scull, *Bicategories of fractions revisited: towards small homs and canonical 2-cells*, 08 2019.
35. Dorette Pronk and Michael Warren, *Bicategorical fibration structures and stacks*, Theory and Applications of Categories **29** (2013).
36. Nima Rasekh, *Cartesian fibrations and representability*, arXiv1711.03670, 2017, preprint.
37. ______, *A theory of elementary higher toposes*, arXiv1805.03805, 2018, preprint.
38. Martina Rovelli, *Weighted limits in an $(\infty, 1)$-category*, arXiv:1902.00805, 2019, preprint.
39. Omer Sezer and Murat Ozbayoglu, *Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach*, Applied Soft Computing **70** (2018).
40. Ross Street, *Limits indexed by category-valued 2-functors*, Journal of Pure and Applied Algebra **8** (1976), no. 2, 149–181.
41. Ross Street, *Two-dimensional sheaf theory*, Journal of Pure and Applied Algebra **23** (1982), no. 3, 251–270 (English).
42. Dominic Verity, *Complicial sets characterising the simplicial nerves of strict ω-categories*, Mem. Amer. Math. Soc. **193** (2008), no. 905, xvi+184.
43. ______, *Weak complicial sets. I. Basic homotopy theory*, Adv. Math. **219** (2008), no. 4, 1081–1149.

Institute of Mathematics, Czech Academy of Sciences, Žitná 25, 115 67 Praha 1, Czech Republic

Email address: `edoardo.lanari.el@gmail.com`

URL: `https://edolana.github.io/`