Full-day mini-course on Path Signatures

Program:

09:00-10:45	**From Paths to Features: An Introduction to the Signature Transform**	**Fabian Harang**

Abstract: Many problems in mathematics, data science, economics, finance, and engineering involve sequential or streamed information: time series, trajectories, flows, or evolving systems. The path signature offers a principled way to represent such data by mapping a path to a sequence of iterated integrals that behave like coordinates for the path’s effects. This lecture gives a basic, researcher-friendly introduction to the signature transform: what it is, why it is natural, and how it can be used.

We will build intuition from simple examples (piecewise linear paths and low-order terms), explain how signatures encode order and interaction across time scales, and discuss how common preprocessing choices (time-augmentation, lead–lag, normalization, and truncation) affect what information is retained. Without emphasizing heavy algebra, we will highlight the key structural facts that make signatures useful in practice: stability under perturbations, principled feature construction for learning, and connections to controlled differential equations. We conclude with a high-level overview of application patterns—classification/regression on sequential data, model reduction and system identification, and examples from finance and engineering—preparing the ground for the subsequent lectures on algebraic structure and on path development and machine learning.

11:00-11:45	**Title: From Iterated Sums to Path Signatures: An Algebraic Perspective**	**Kurusch Ebrahimi-Fard**

Abstract: Iterated integrals and iterated sums provide a unifying language for encoding sequential and time-ordered data, ranging from classical control theory to rough path theory and modern machine learning. Central to this framework is the notion of a (discrete) signature: an infinite collection of iterated integrals (or sums) that captures the essential information of a path up to natural equivalences.
This lecture introduces the algebraic structures underlying iterated integrals and iterated sums, with an emphasis on tensor algebras, shuffle and quasi-shuffle products, and their Hopf-algebraic interpretation. These structures encode concatenation, time ordering, and reparameterization invariance in a transparent way, and explain the origin of Chen’s relations from purely algebraic principles. The discrete setting of iterated sums is treated in parallel with the continuous theory, highlighting both common features and structural differences.

The algebraic viewpoint is critical both for conceptual understanding and for computability and implementation: algebraic structures such as (quasi-)shuffle algebras and Hopf algebras make concatenation, truncation, and recursions explicit. The lecture aims to provide an accessible overview of the algebraic backbone of signature theory, with a view towards further developments in rough paths, numerical analysis, and (quantum) machine learning applications.

12:00-13:00	**Lunch**	
13:00-13:45	(Cont.) **From Iterated Sums to Path Signatures: An Algebraic Perspective**	**Kurusch Ebrahimi-Fard**
14:00-15:45	**Path Development Networks on Finite-Dimensional Lie Groups for Sequential Learning**	**Hao Ni**

Abstract: The path signature, originating from rough path theory, provides a mathematically principled and universal representation for sequential data and has found wide-ranging applications in modern machine learning. Despite its theoretical appeal, its practical use is often limited by the curse of dimensionality when dealing with high-dimensional paths. In this lecture, I will introduce path development, a trainable and dimension-efficient alternative to signature representations, which encodes sequential data through finite-dimensional Lie groups.

In the first half of the talk, I will present the theoretical foundations of path development and its key properties, elucidating its connections to path signatures, recurrent neural networks (RNNs), and neural controlled differential equations (neural CDEs). I will also explain the backpropagation-through-time algorithm that enables efficient optimization of model parameters via an optimization method on the manifold called "trivialization".

The second half of the lecture will focus on applications of path development to time-series learning. I will begin with a brief, hands-on tutorial demonstrating how to apply the path development layer to simple toy time-series datasets via python notebook. I will then present empirical results showing that path development networks significantly outperform signature-based methods. Moreover, the hybrid model by combining LSTM with the development network achieve state-of-the-art performance on several time series benchmarking datasets. The talk will conclude with diverse applications of the path development to complex real-world time series data, including skeleton-based action recognition, signature verification and medical time-series analysis, where the development network is carefully integrated into neural network designs to better capture domain knowledge and achieve superior empirical performance. This lecture is based on [1-4].

[1] Lou, H., Li, S. and Ni, H., Path Development Network with Finite-dimensional Lie Group. Transactions on Machine Learning Research. 2024
[2] Shi, Y., Jiang, L., Ni, H. and Jin, L., 2025, September. DevInSight: Weaving Path Development Into Online Signature Verification. In International Conference on Document Analysis and Recognition (pp. 398-415).
[3] Jiang, L., Yang, W., Zhang, X. and Ni, H., 2024. GCN-DevLSTM: Path development for skeleton-based action recognition. arXiv preprint arXiv:2403.15212.
[4] Feng, T., Li, Q., Zhang, Y., Liao, Y., Lu, D., Liping, W., Zhao, J., Jiang, L., Ni, H., Liu, H. and Deng, J., 2026. PathFusion-Net: A Rough Path Theory-Based Deep Learning Model for ECG Arrhythmia Classification. IEEE Journal of Biomedical and Health Informatics.