Exploring the Use of Rough Path Signatures in Quantitative Finance Applications

1. Introduction

The increasing complexity and nonlinearity of financial time series necessitate advanced tools that can capture path-dependent structures beyond conventional statistical or machine learning approaches. One promising methodology from the field of stochastic analysis is the use of **rough path signatures**, which provide a systematic algebraic representation of sequential data.

This project aims to investigate the potential of signature methods in **portfolio clustering** and **financial time series generation**, two areas where capturing path-level features is critical. The study will compare performance against **non-signature-based methods** such as traditional econometric models, statistical feature extraction, and deep learning sequence models.

---

2. Objectives

1. **Portfolio Clustering**
    - Develop a framework for representing multi-asset time series using rough path signatures.
    - Apply unsupervised learning techniques (e.g., hierarchical clustering, spectral clustering, k-means) to group portfolios based on path signatures.
    - Compare clustering results with conventional distance measures (correlation-based distances, covariance matrices, dynamic time warping).
2. **Time Series Generation**
    - Explore the use of signature-based generative models for producing realistic synthetic financial time series.
    - Benchmark against standard generative models (ARIMA, GARCH, GANs, Variational Autoencoders).
    - Evaluate realism through distributional tests, autocorrelation structure, and stress scenarios.
3. **Comparative Analysis**
    - Quantitatively assess the advantages of signature methods in representing path-dependent features.

- Identify cases where non-signature methods perform equally well or better, thereby clarifying the value-added scope of rough paths in finance.

---

3. Literature Review

- **Rough Path Theory**: Lyons (1998), Friz & Victoir (2010) – foundational work in signature methods for stochastic processes.
- **Signatures in Finance**: Research into applications for volatility modeling, market microstructure, and derivative pricing (e.g., Gatheral et al., 2018; Arribas et al., 2020).
- **Portfolio Clustering**: Traditional correlation clustering (Mantegna 1999), hierarchical risk parity, network-based methods.
- **Time Series Generation**: Econometric approaches (Box & Jenkins), deep generative models in finance (Wiese et al., 2020 on GANs for financial data).

---

4. Methodology

4.1 Data

- Daily and intraday financial time series: equity indices, individual stocks, fixed income, FX.
- Use of both synthetic (simulated SDEs) and real-world datasets.

4.2 Signature Computation

- Compute truncated signature transforms for multivariate financial time series.
- Evaluate dimensionality trade-offs (depth vs computational cost).
- Explore log-signatures for stability and interpretability.

4.3 Portfolio Clustering

- Feature representation: signatures vs correlation-based statistics.
- Dimensionality reduction: PCA, diffusion maps, autoencoders.
- Clustering algorithms: hierarchical clustering, spectral clustering.
- Evaluation metrics: Silhouette score, cluster stability, interpretability in financial contexts.

4.4 Time Series Generation

- Develop generative models using signatures (e.g., signature-based GANs, controlled differential equations).
- Benchmark against: ARIMA, GARCH, LSTMs, GANs.
- Evaluation criteria: marginal distributions, autocorrelations, higher-order moments, tail behavior, backtest performance in trading strategies.

4.5 Comparative Analysis

- Statistical tests: Kolmogorov–Smirnov, Maximum Mean Discrepancy.
- Performance measures: reconstruction error, predictive performance in downstream tasks (e.g., risk forecasting).
- Computational efficiency analysis.