# Mathematical Frontiers for Next-Generation Autonomous Finance: A Strategic Research Report

## 1. The Convergence of Mathematical Rigor and Agentic Intelligence

The contemporary landscape of financial technology is undergoing a paradigmatic shift, transitioning from static descriptive analytics to autonomous, agentic reasoning. The vision for the SCALE application—an AI-powered personal finance platform capable of replacing human accountants—demands a technological foundation that transcends the capabilities of traditional linear regression and discrete time-series forecasting. While current industry standards rely on heuristic classification and established statistical methods like ARIMA or Prophet, these approaches fail to capture the high-dimensional, irregular, and topological complexity of human financial behavior. To construct a proprietary "moat" and realize the vision of an "AI Accountant" that not only tracks but actively manages wealth, the underlying prediction engines must integrate the most advanced mathematical breakthroughs of 2025 and 2026.

This research report synthesizes a comprehensive analysis of frontier mathematical domains—specifically Topological Data Analysis (TDA), Rough Path Theory, Hyperbolic Geometry, Optimal Transport, and Causal Inference—that have witnessed significant theoretical and applied advancements in the current research cycle. These fields offer novel mechanisms for modeling the "shape" of spending data, handling the irregularity of transaction streams, and distinguishing between correlation and causation in financial advice. By rigorously examining recent literature, including the resolution of the 3D Kakeya conjecture and advances in Hilbert's sixth problem, we establish a theoretical and practical blueprint for the SCALE architecture.

The central thesis of this analysis is that financial data should not be treated merely as a sequence of numbers, but as a complex physical system governed by geometric and topological laws. The integration of these advanced mathematical frameworks into a Python-based infrastructure, leveraging agents and neural differential equations, allows for the construction of prediction engines that operate with the nuance of a human CFO and the computational precision of a high-frequency trading algorithm. This report details the theoretical underpinnings, recent breakthroughs, and specific implementation strategies for each mathematical domain, providing a roadmap for engineering the next generation of financial intelligence.

# 2. Topological Data Analysis: The Geometry of Financial Stability

Traditional financial analysis treats transaction data as discrete points in time, often aggregating them into rigid temporal buckets such as daily or monthly totals. This reductionist approach discards the geometric "shape" of spending behavior—the loops, voids, and connected components that define the latent structure of a user's financial life. Topological Data Analysis (TDA) offers a mechanism to analyze this shape, providing a robust method for detecting structural changes in spending patterns that standard statistical methods miss.

## 2.1 Persistence Landscapes as Indicators of Financial Regime Shifts

The concept of "persistence" in TDA facilitates the separation of significant structural features from noise, a critical capability when dealing with the stochastic nature of personal financial data. Recent research has solidified the application of **Persistence Landscapes** as a stable, vector-based representation of these topological features, making them highly suitable for integration into machine learning pipelines.[1]

### 2.1.1 Theoretical Foundations and Stability

Persistent homology tracks the birth and death of topological features—such as connected components (0-dimensional holes), loops (1-dimensional holes), and voids (2-dimensional holes)—as a filtration parameter increases. A persistence diagram plots these birth-death pairs in a 2D plane. However, the space of persistence diagrams lacks the structure of a Hilbert space, making direct application in statistical learning challenging. Persistence landscapes address this by transforming the diagram into a sequence of continuous, piecewise linear functions, $\lambda_k(t)$, which reside in a Banach space.

The primary advantage of this transformation is stability. The Stability Theorem guarantees that small perturbations in the input data (noise) result in only bounded changes in the persistence landscape. In the context of the SCALE app, this means that minor fluctuations in transaction amounts or timing—common in financial data—will not trigger false alarms, whereas genuine structural shifts will be captured.[3]

### 2.1.2 2025 Breakthroughs in Financial Crisis Detection

Empirical studies conducted in 2025 have demonstrated that the $L^p$ norms (specifically $L^1$ and $L^2$) of persistence landscapes serve as powerful early warning signals (EWS) for critical transitions in financial markets. Research analyzing the S&P 500, NASDAQ, and other major indices has shown that spikes in the norm of the persistence landscape often precede systemic crises, such as the 2008 global financial crisis or the COVID-19 market crash, well before they are reflected in traditional volatility indices.[2]

These studies utilize a sliding window approach, mapping multivariate time series of returns into point clouds via Taken's embedding theorem. The resulting Vietoris–Rips filtration reveals the changing topology of the market. A sudden increase in the "life" of topological loops suggests a fragmentation of the market structure or the emergence of dangerous co-movement patterns indicative of a bubble or impending crash.[3]

#### 2.1.3 Implementation in the SCALE Anomaly Engine

For the SCALE app, implementing a Persistence Landscape layer within the Anomaly Detection Engine offers a significant upgrade over traditional methods like Isolation Forests or Z-scores.

- **Behavioral Topology:** Instead of flagging a single high-value transaction as an anomaly based on magnitude, TDA analyzes the *structure* of spending. For instance, a user may enter a cycle of taking out short-term loans to pay off other debts. In a high-dimensional feature space (Time, Amount, Merchant Category, Account Balance), this behavior manifests as a persistent loop. A linear model might see the inflows and outflows as cancelling each other out, but TDA identifies the stable cyclic structure, flagging it as a potential "debt spiral."
- **Regime Shift Detection:** Users transition through distinct financial "regimes" or life stages—such as shifting from "student" to "employed," or "single" to "parent." These regimes exhibit distinct topological signatures in their transaction graphs. By computing the persistence landscapes of a user's transaction history over sliding windows, the AI Accountant can detect a "regime shift" (e.g., a permanent, structurally integrated increase in recurring expenses) rather than misinterpreting it as a temporary spike.[5]
- **Algorithmic Strategy:** The implementation should leverage Python libraries such as gudhi or scikit-tda to compute Vietoris–Rips complexes and generate persistence landscapes. The "Topological Information Supervised (TIS)" framework, proposed in 2025 literature, introduces a neural network architecture (TIS-BiGRU) that learns to generate synthetic topological features. This approach significantly reduces the computational bottleneck associated with calculating persistent homology on the fly, allowing the SCALE app to provide real-time structural alerts without prohibitive latency.[1]

### 2.2 Topological Signal Processing for Sparse and Irregular Data

One of the primary challenges in personal finance modeling is data sparsity. Unlike high-frequency trading data, personal transactions occur irregularly; a user may have zero transactions for days, followed by a burst of activity. This sparsity renders standard time-series methods (like ARIMA or dense Neural Networks) unstable or reliant on heavy imputation. TDA excels in this environment because it focuses on connectivity and global shape rather than frequency or density.

Recent 2025 research into the **Topological Information Supervised (TIS)** prediction framework demonstrates that augmenting time-series models with a topological consistency loss function improves prediction accuracy on complex, irregular datasets.[1] This framework employs a Conditional Generative Adversarial Network (CGAN) to learn the distribution of topological features. By training the prediction engine to respect the topological structure of the spending data—ensuring that the forecasted spending trajectory preserves the user's historical spending "loops" and connected components—the model avoids generating physically impossible or structurally incoherent predictions.[8]

The integration of TDA into the SCALE prediction engine allows for a sophisticated "shape-based" forecasting capability. When the user asks, "Can I afford this vacation?", the AI does not just project a linear trend of bank balances. It analyzes the topological stability of the user's current spending regime. If the purchase would force a topological transition (e.g., breaking a stable "saving" loop or creating a "debt" loop), the AI can provide a mathematically grounded warning: "This purchase disrupts your stable spending geometry, likely leading to a new, less stable financial regime."

# 3. Rough Path Theory: The Calculus of Irregular Streams

To construct a truly proprietary prediction engine that serves as a robust "moat," the SCALE app must move beyond standard discrete time-series models. **Rough Path Theory**, a branch of stochastic analysis, has seen profound breakthroughs in 2025, particularly in its application to financial modeling via **Signatures** and **Neural Rough Differential Equations (Neural RDEs)**. These tools provide a rigorous mathematical framework for modeling continuous, irregular data streams without the need for artificial discretization or padding.

## 3.1 The Signature Transform as a Universal Feature Map

The core object of Rough Path Theory is the **Signature** of a path. For a stream of data (such as a sequence of transactions with timestamps, amounts, and merchant embeddings), the signature is an infinite collection of iterated integrals.

### 3.1.1 Mathematical Definition and Properties

Let $X_t$ be a path in $\mathbb{R}^d$ representing the financial state over time. The signature $S(X)_{a,b}$ over an interval $[a,b]$ is defined as the sequence of tensors:

$$S(X)_{a,b} = \left(1, \int_{a<t<b} dX_t, \int_{a<t_1<t_2<b} dX_{t_1} \otimes dX_{t_2}, \dots\right)$$

A fundamental theorem in the field states that the signature is a **universal nonlinearity**. This means that linear functions on the signature can approximate *any* continuous function of the path to arbitrary accuracy.[10] This property is analogous to the universality of polynomials (Stone-Weierstrass theorem) but for path-space.

The signature captures the geometric order of events. In financial terms, it distinguishes between "spending $100 then receiving $100" and "receiving $100 then spending $100," even if the net change is zero. Higher-order terms in the signature capture complex interactions, such as the acceleration of spending relative to income, or the volatility of merchant categories over time.[13]

### 3.1.2 2025 Advances: Neural RDEs and Log-Signatures

In 2025, the integration of **Neural Rough Differential Equations (Neural RDEs)** with truncated log-signature encoders emerged as a state-of-the-art (SOTA) technique for handling high-dimensional, path-dependent financial problems.[2]

- **Log-Signatures:** The signature contains redundant information due to algebraic relations (shuffle product). The **log-signature** provides a more concise, compressed representation of the path's information content, reducing the dimensionality of the feature space while retaining the universal approximation properties.
- **Neural RDEs:** Unlike standard Neural ODEs, which are driven by time ($dt$), Neural RDEs are driven by the data path itself ($dX_t$). The hidden state $H_t$ evolves according to:

  $$dH_t = f(H_t)dX_t$$

  This formulation allows the model to respond continuously to the incoming stream of transactions. The 2025 breakthroughs include the use of log-signature encoders to effectively summarize the "past" trajectory, addressing the information bottleneck that often plagues RNNs and LSTMs in long-sequence modeling.[3]

## 3.2 Application to the SCALE Prediction Engine

The implementation of a **Signature-based Prediction Layer** within the SCALE app addresses several critical requirements:

1. **Handling Irregularity:** Financial transactions are inherently irregular. A user might make three transactions in one hour and then none for two days. Standard RNNs/LSTMs struggle with this, often requiring binning (e.g., hourly sums) that destroys fine-grained temporal information. Neural RDEs naturally handle continuous-time inputs and irregular sampling, treating the latent state as a continuous trajectory driven by the transaction stream. This eliminates the need for potentially biasing interpolation or padding.[15]
2. **Proprietary Algorithmic Moat:** By utilizing signatures, the SCALE prediction engine can capture complex, non-linear dependencies that standard feature engineering misses. For example, the signature can encode the "area" generated by the path of (Income, Expenses), which corresponds to the accumulation of savings or debt. A prediction

model trained on signature features effectively "learns" the stochastic differential equations governing the user's financial life, providing a deeper, more physically grounded forecast than simple autoregressive models.[12]

3. **Computational Efficiency on Edge Devices:** Signatures compress the path history into a fixed-size vector. This is particularly advantageous for the "AI Accountant" vision, where the app may need to run inference on a mobile device. Instead of storing and processing the entire history of transactions for every query, the app can incrementally update the user's current "signature state." Python libraries like RoughPy and signatory, which have seen significant updates in 2025, enable highly efficient computation of signatures and log-signatures on both CPU and GPU backends.[16]
4. **Portfolio Optimization and Control:** Recent literature from 2025 highlights the use of **Signature Portfolios** for optimal control problems. The SCALE app can adapt this to "Personal Portfolio Optimization." By viewing the user's budget allocation as a control problem, the Neural RDE can suggest optimal "rebalancing" actions (e.g., "Move $500 from Entertainment to Savings") to maximize the probability of hitting financial goals, leveraging the "Physics-Informed" losses developed in recent rough path research.[15]

# 4. Hyperbolic Geometry: Hierarchical Taxonomy and Categorization

A central feature of the SCALE app is the "Intelligent Auto-Categorization Engine." The user explicitly aims to surpass standard industry tools like Plaid or basic keyword matching. The fundamental challenge in merchant categorization is that the data is inherently hierarchical. A transaction at "Starbucks" belongs to the category "Coffee Shop," which is a subset of "Food & Drink," which falls under "Discretionary Spending." Standard Euclidean embeddings (like those used in BERT or word2vec) struggle to represent these hierarchical relationships without significant distortion, as the volume of Euclidean space grows only polynomially.

## 4.1 The Power of Poincaré Embeddings

Hyperbolic space, specifically the Poincaré ball model, has a geometry where volume expands exponentially with radius. This property makes it mathematically isomorphic to tree-like structures. Embedding hierarchical data into hyperbolic space allows for the preservation of parent-child relationships and taxonomic distance with arbitrarily low distortion, even in low dimensions.

### 4.1.1 Hyperbolic Category Discovery (HypCD) Framework

A breakthrough paper presented at CVPR 2025, **"Hyperbolic Category Discovery" (HypCD)**, introduces a novel framework for Generalized Category Discovery (GCD) that operates entirely within hyperbolic space.[18] The HypCD framework addresses the limitations of Euclidean and spherical spaces by leveraging the exponential capacity of hyperbolic geometry to encode latent hierarchies.

The mechanics of the HypCD framework are directly applicable to the SCALE categorization engine:

- **Euclidean-to-Hyperbolic Mapping:** The process begins with a standard transformer backbone (like a fine-tuned BERT or RoBERTa on financial text) that produces Euclidean representations. These vectors are then mapped to the Poincaré ball ($H^n$) using an **exponential map** ($exp_o^c$). To prevent numerical instability and gradient vanishing near the boundary of the ball—a common issue in hyperbolic learning—the framework applies a feature clipping operation in the tangent space before mapping.[18]
- **Hyperbolic Contrastive Learning:** The core innovation is performing contrastive learning directly on the manifold. The loss function minimizes the **hyperbolic distance** between "positive" pairs (e.g., "Starbucks" and "Dunkin" as coffee shops) while maximizing the distance between "negative" pairs. Crucially, the hyperbolic metric naturally pushes generic, high-level categories (like "Food") towards the origin of the ball, while specific, granular merchants (like "Joe's Vegan Cheese Shop") are pushed towards the boundary.
- **Riemannian Optimization:** Optimization on the Poincaré ball requires specialized algorithms that respect the curvature of the space. The HypCD framework utilizes **Riemannian Adam (RAM)**, an extension of the Adam optimizer designed for non-Euclidean manifolds. This ensures that parameter updates follow geodesics rather than straight lines, preserving the geometric integrity of the embeddings.[20]

## 4.2 Implementation for the SCALE Categorizer

Integrating hyperbolic embeddings into the SCALE app creates a highly effective "proprietary moat" for the categorization engine.

- **Hierarchical Merchant Resolution:** By training a **Hyperbolic Merchant Encoder**, the app can construct a "Map of Commerce" within the Poincaré ball. When a new, unseen transaction string arrives (e.g., "SQ * ARTISAN ROAST"), the model projects it into the hyperbolic space. Even if the string is novel, its semantic features will land it in the "neighborhood" of other coffee shops. The nearest centroid in this curved space will likely be the "Coffee" category node, or potentially a new "Specialty Coffee" leaf node. This **Zero-Shot Categorization** capability is significantly more robust than flat classification models, which often fail on unseen labels.[22]
- **Drift-Resistant Taxonomy:** As new types of merchants emerge (e.g., new subscription services or crypto platforms), the hyperbolic space can accommodate them by expanding outward toward the boundary without disrupting the central organization of established categories. This provides a flexible, evolving taxonomy that grows with the market.
- **Integration with LLMs:** Recent 2025 research has also explored **Hyperbolic RAG**

**(Retrieval-Augmented Generation)**. By indexing the user's transaction history in a hyperbolic vector store (instead of a standard Euclidean one), the "AI Accountant" LLM can retrieve relevant context with much higher precision. A query like "How much do I spend on non-essential food?" implies a traversal of the hierarchy. Hyperbolic retrieval naturally understands that "Coffee" and "Fast Food" are children of "Non-essential Food" based on their geometric position, facilitating more accurate aggregation and reasoning.[22]

# 5. Optimal Transport: Detecting Behavioral Drift

A robust financial application must adapt to the evolving reality of its user. The "Concept Drift" problem—where a predictive model trained on past data becomes obsolete as user behavior changes (e.g., getting married, changing jobs, moving cities)—is a critical failure point for static models. **Optimal Transport (OT)** theory provides a rigorous, geometric framework for detecting and quantifying this drift with high sensitivity.

## 5.1 Wasserstein Distance for Distributional Shift

The **Wasserstein distance** (often referred to as the Earth Mover's Distance) quantifies the minimal "work" required to transform one probability distribution into another. Unlike information-theoretic measures like Kullback-Leibler (KL) divergence, the Wasserstein distance is a true metric that respects the geometry of the underlying space. It yields meaningful values even when the support of the two distributions does not overlap, making it ideal for comparing sparse or disjoint financial datasets.[25]

### 5.1.1 Optimal Transport-based Drift Detection (OTDD)

Research published in 2025 has introduced **Optimal Transport-based Drift Detection (OTDD)** algorithms specifically designed for streaming data environments.[28] The core methodology involves:

1. **Reference Window:** Defining a baseline distribution of spending behavior (e.g., "Spending patterns in Q1") represented as a probability measure $\mu_{ref}$ .
2. **Current Window:** Continuously computing the distribution of the most recent transactions, $\mu_{curr}$ .
3. **Drift Metric:** Calculating the Wasserstein distance $W_p(\mu_{ref}, \mu_{curr})$ . If this distance exceeds a statistically derived threshold—often determined via a method like **Data-Driven Threshold Estimation (DDTE)**—a concept drift is signaled.

This approach is far more sensitive than tracking simple averages. For instance, a user might maintain the same *total* monthly food spend (mean is stable), but switch from "many small grocery runs" to "few large luxury dinners." A mean-tracking model would miss this lifestyle inflation. The Wasserstein distance, however, would detect the change in the *shape* of the distribution (transporting mass from low-value/high-frequency to high-value/low-frequency), triggering an alert.[28]

## 5.2 Application to Robust Financial Forecasting

Beyond detection, Optimal Transport offers a framework for **Robust Forecasting**. The theory of **Martingale Optimal Transport (MOT)** allows for the pricing of risk and the bounding of uncertainty without assuming a specific parametric model for the underlying asset price dynamics.[30]

Adapted to the SCALE app, MOT can be used to generate "Robust Savings Forecasts." Instead of predicting a single "most likely" savings number, the engine can calculate the **Wasserstein Barycenter** of multiple predictive scenarios (e.g., "Recession," "Boom," "Status Quo"). This barycenter represents a geometric "average" distribution that minimizes the transport cost to all potential futures. The AI Accountant can then present a "Safe-to-Spend" limit that is robust to worst-case distributional shifts, providing users with a safety margin that is mathematically grounded in transport theory rather than arbitrary buffers.

Implementation of these OT methods is facilitated by the POT (Python Optimal Transport) library, which has seen significant performance optimizations (such as Sinkhorn iterations) in recent updates, making it feasible to compute Wasserstein distances on transaction batches in near real-time.[32]

# 6. The "AI Accountant" Brain: Causal Inference and Neuro-Symbolic Reasoning

The user's vision of an "AI Accountant" that replaces human professionals requires a system capable of reasoning about *cause and effect*. Current Large Language Models (LLMs) are fundamentally correlation engines; they can predict the next word, but they struggle to reliably infer causality (e.g., "Did buying this car *cause* my savings to dip, or was it the rent increase?"). To provide sound financial advice, the SCALE app must integrate **Causal Inference** frameworks.

## 6.1 The FinCARE Framework: Hybrid Causal Discovery

In late 2025, the **FinCARE (Financial Causal Analysis with Reasoning & Evidence)** framework was introduced, representing a hybrid approach that merges statistical causal discovery with the semantic reasoning capabilities of LLMs.[34] This framework addresses the limitations of pure statistical methods (which struggle with limited data) and pure LLMs (which hallucinate causality).

The FinCARE methodology consists of two synergistic components:

1. **LLM-Hypothesis Generation:** The LLM analyzes unstructured data (e.g., financial news,

user notes, transaction descriptions) to generate a set of potential causal edges (hypotheses). For example, it might hypothesize "High Uber usage $\rightarrow$ Credit Card Debt."

2. **Statistical Validation:** These hypotheses are then rigorously tested against the structured transaction data using causal discovery algorithms such as **PC (Peter-Clark)** or **GES (Greedy Equivalence Search)**. The algorithm retains only those edges that are supported by conditional independence tests in the data.

This hybrid approach allows the SCALE app to build a **Personalized Causal Graph** for each user. This graph is a Directed Acyclic Graph (DAG) where nodes are financial variables (Income, Category Spend, Savings, Debt) and edges represent confirmed causal links.

## 6.2 Individual Causal Inference (ICI) and Counterfactual Reasoning

For personal finance, population-level causality ("Buying cars generally reduces savings") is insufficient; the system needs **Individual Causal Inference (ICI)**. Recent 2025 research has formalized ICI using **Structural Causal Models (SCM)**, enabling rigorous counterfactual reasoning through a three-step process: **Abduction, Intervention, and Inference.**[38]

- **Abduction:** The model uses the user's observed history to infer the specific values of "exogenous noise" variables—hidden factors unique to the user, such as their latent risk tolerance or unobserved income sources.
- **Intervention ($do(X)$):** The AI simulates a hypothetical action by modifying the causal graph. For example, to answer "Can I afford this $50,000 car?", the Planner Agent performs a "graph surgery," severing the natural causes of the "Car Asset" node and forcing its value to $50,000.
- **Inference:** The model propagates this intervention through the SCM to predict the counterfactual distribution of the outcome variable (e.g., "Future Savings").

This framework enables the AI Accountant to give advice that is mathematically sound and personalized. instead of a generic "That's expensive," the AI can state: "Based on your causal graph, if you buy this car ($do(Car = 1)$), there is a 93% probability that your 'Emergency Fund' node will drop below the critical threshold within 6 months, caused primarily by the resulting increase in 'Insurance' and 'Maintenance' nodes."

## 6.3 Graph-Temporal Contrastive Learning for Fraud and Classification

To support this high-level causal reasoning, the underlying data representation must be robust. The **Graph-Temporal Contrastive Transformer (GTCT)** framework, detailed in 2025 literature for financial fraud detection, offers a powerful architecture for learning these representations.[39]

The GTCT model combines a **Graph Encoder** (modeling relationships between accounts/merchants) with a **Temporal Encoder** (modeling sequential patterns) and a

**Contrastive Learning Objective**. This contrastive loss ensures that the model learns to distinguish between similar but distinct transaction patterns (e.g., "Legitimate high spending" vs. "Fraudulent spike") even in the absence of large labeled datasets. For the SCALE app, deploying a GTCT-based encoder ensures that the node features fed into the Causal Discovery engine are rich, discriminative, and temporally aware, significantly improving the accuracy of the resulting causal graph.

# 7. Physics-Informed Financial Modeling: Fluid Dynamics and Sparse Signals

The user's request for "complex math" and "fluid dynamics" analogies is met by a serendipitous and historic breakthrough in 2025: the resolution of **Hilbert's Sixth Problem** regarding the derivation of fluid mechanics from particle dynamics.

## 7.1 The Navier-Stokes Analogy for Cash Flow

In 2025, mathematicians rigorously proved the derivation of macroscopic fluid equations (Navier-Stokes) from microscopic hard-sphere particle dynamics (Boltzmann equation).[40] This theoretical bridge provides a rigorous justification for modeling "macro" financial flows (monthly cash flow, savings rate) as the emergent result of "micro" particle interactions (individual transactions).

We can formalize a **Navier-Stokes Economic Model** for the SCALE app [45]:

- **Velocity Field ($\mathbf{u}$):** Represents the rate and direction of money transfer across the graph of accounts.
- **Pressure ($p$):** Analogous to financial demand or stress. High "pressure" in the "Bills" node (due to upcoming due dates) drives the flow of liquidity.
- **Viscosity ($\nu$):** Represents transaction friction—fees, transfer delays, or psychological resistance to spending.
- **Density ($\rho$):** The concentration of capital in specific categories or accounts.

By solving the discretized Navier-Stokes equations on the user's financial graph, the SCALE prediction engine can simulate the "flow" of money. This allows for the calculation of a **"Financial Reynolds Number" ($Re$)**. In fluid dynamics, a high $Re$ indicates turbulence. In the SCALE app, a high Financial Reynolds Number would indicate that the inertial forces of spending (volatility, impulse buys) are overwhelming the viscous forces (budgetary constraints, self-control), leading to "turbulent" and unpredictable financial instability. This provides a novel, physics-based metric for financial health monitoring.

## 7.2 Sparse Signal Processing: The Kakeya Conjecture

Financial data is notoriously sparse—most seconds of the day contain zero transactions. Efficiently processing this sparse signal requires advanced mathematical tools. The 2025 resolution of the **3D Kakeya Conjecture** by Wang and Zahl has introduced powerful new techniques in geometric measure theory and sparse signal analysis, specifically the **"Induction on Scales"** algorithm.[40]

The "Induction on Scales" method analyzes data at multiple resolutions simultaneously to identify structures (like lines or "tubes") that persist across scales.[52] Adapted for the SCALE app, this enables a **Multi-Scale Spending Analyzer**. Instead of analyzing spending only at the discrete "transaction" level or the aggregated "monthly" level, the algorithm processes the data stream at dyadic intervals ($2^k$ seconds). This allows the engine to detect "directional" features in high-dimensional sparse data—for example, identifying a trajectory of spending that is slowly accelerating towards a critical limit (bankruptcy or a savings goal) even when the individual data points are sparse and noisy.

Furthermore, to handle the "zeros" in transaction data effectively, the prediction engine should incorporate **Zero-Inflated Two-Stage Models**. As suggested in the sparse signal processing literature, this involves a "Gatekeeper" classifier (estimating $P(Transaction > 0)$) coupled with a magnitude regression model. This separation prevents the "averaging" error common in standard models, ensuring that the predicted spending stream retains the sparse, spiky nature of real financial reality.

# 8. Strategic Implementation and Architecture

To realize the vision of the SCALE app, a sophisticated "Hybrid Neuro-Symbolic Architecture" is required. This architecture integrates the deep mathematical engines described above into a cohesive, scalable system.

## 8.1 The "SCALE" Technical Stack

The proposed backend architecture is divided into three distinct layers, each utilizing specific Python libraries and technologies to implement the mathematical concepts.

### 8.1.1 Layer 1: The Deep Math & Feature Engineering Layer (Batch/Async)

This layer is responsible for heavy computational tasks that update the user's financial model periodically (e.g., nightly).

- **Topological Engine:** Computes Persistence Landscapes.
  - *Library:* gudhi or scikit-tda.
  - *Function:* Extracts topological feature vectors from the past 12 months of transaction history to detect regime shifts.

- **Causal Graph Builder:** Constructs the personalized Structural Causal Model.
  - *Library:* causal-learn (for PC/GES algorithms) and DoWhy (for causal reasoning).
  - *Function:* implementing the FinCARE framework to hybridize LLM hypotheses with statistical validation.
- **Infrastructure:** Python workers hosted on **Modal** or **AWS SageMaker** to handle the heavy compute loads of TDA and Causal Discovery without blocking the API.

### 8.1.2 Layer 2: The Fast Inference & Prediction Layer (Real-Time)

This layer serves immediate predictions to the frontend and the AI Accountant.

- **Neural RDE Predictor:** A Neural Rough Differential Equation model.
  - *Library:* torchcde (Torch CDE) and signatory (for path signature computation).
  - *Function:* Takes the live transaction stream, computes the log-signature, and evolves the hidden financial state to forecast future liquidity.
- **Hyperbolic Categorizer:** The HypCD implementation.
  - *Library:* geoopt (Riemannian optimization in PyTorch) and transformers.
  - *Function:* Maps incoming merchant strings into the Poincaré ball and finds the nearest category centroid using hyperbolic distance.
- **Infrastructure:** High-performance **FastAPI** services deployed on **Google Cloud Run** or **Railway**.

### 8.1.3 Layer 3: The Agentic Orchestration Layer ("The Brain")

This layer interfaces with the user, translating natural language into mathematical queries.

- **Orchestration Framework: LangGraph** (built on LangChain), as preferred by the user.
- **Mechanism:**
  - **Router Agent:** Classifies user intent (e.g., "Analysis," "Prediction," "Advice").
  - **Analyst Tool:** Invokes the Neural RDE Predictor for forecasts.
  - **Risk Tool:** Checks the Wasserstein Drift metric from Layer 1.
  - **Advisor Tool:** Performs "Intervention" simulations on the Causal Model (e.g., "What if I buy this?").
  - **Synthesis:** The LLM integrates these outputs into a coherent, empathetic response.

## 8.2 Data Infrastructure for Scale

To support scaling to "Big Data" levels (millions of transactions):

- **Data Lake: DuckDB** is the optimal choice for the embedded OLAP engine within the Python services. It handles vectorized operations efficiently and interacts seamlessly with Pandas/Arrow, capable of processing millions of rows on a single instance.
- **Vector Store:** For RAG (Retrieval Augmented Generation) and storing hyperbolic embeddings, **pgvector** within **Supabase** (which the user already employs) is recommended. It avoids the complexity of a separate vector database while supporting the necessary distance metrics.

### Table 1: Strategic Mathematical Stack for SCALE

| Feature | Current Industry Standard | Proposed Mathematical Innovation (2025/2026) | Theoretical Basis | Key Libraries/Tools |
|---|---|---|---|---|
| **Transaction Classification** | BERT / Fuzzy Matching (Euclidean) | **Hyperbolic Embeddings (HypCD)** | Hyperbolic Geometry, Poincaré Ball | geoopt, transformers |
| **Time-Series Prediction** | ARIMA / LSTM / Prophet | **Neural Rough Differential Equations (NRDE)** | Rough Path Theory, Signatures | signatory, torchcde |
| **Anomaly Detection** | Isolation Forest / Z-Score | **Persistence Landscapes ($L^p$ Norms)** | Topological Data Analysis (TDA) | gudhi, scikit-tda |
| **Drift / Habit Change** | KL Divergence | **Optimal Transport Drift Detection (OTDD)** | Wasserstein Distance, Optimal Transport | POT (Python Optimal Transport) |
| **Financial Advice** | Rule-based Heuristics | **Individual Causal Inference (ICI)** | Structural Causal Models (SCM), FinCARE | causal-learn, dowhy |
| **Cash Flow Modeling** | Linear Projections | **Navier-Stokes Fluid Analogy** | Fluid Dynamics, Hilbert's 6th Problem | Custom PINNs (Physics-Informed Neural Nets) |

# 9. Conclusion

The ambition to create an "AI Accountant" requires a fundamental re-engineering of the financial tech stack. By moving beyond linear approximations and adopting the frontiers of **Topological Data Analysis**, **Rough Path Theory**, and **Causal Inference**, the SCALE app can achieve a depth of understanding that mimics—and in some dimensions surpasses—human financial intuition.

The path forward is clear:

1. **Build the Moat:** Implement the **Hyperbolic Categorization Engine** immediately to differentiate the data layer.
2. **Ensure Robustness:** Deploy **Neural RDEs** for prediction to handle the irregularity of real-world spending.
3. **Enable Reasoning:** Construct the **Causal Graph** infrastructure to allow the AI to provide actionable, counterfactual advice ("If you do X, Y will happen").
4. **Monitor Geometry:** Use **Persistence Landscapes** and **Wasserstein Drift** to act as the "nervous system" of the app, detecting structural changes in financial health before they become crises.

This fusion of advanced mathematics and agentic AI positions the SCALE app not just as a tool for tracking, but as a true autonomous financial partner.

### Works cited


1. Time-Series Forecasting via Topological Information Supervised Framework with Efficient Topological Feature Learning - arXiv, accessed on February 11, 2026, https://arxiv.org/html/2503.23757v1
2. Topological Machine Learning for Financial Crisis Detection: Early Warning Signals from Persistent Homology - MDPI, accessed on February 11, 2026, https://www.mdpi.com/2073-431X/14/10/408
3. [2510.10728] Rough Path Signatures: Learning Neural RDEs for Portfolio Optimization, accessed on February 11, 2026, https://arxiv.org/abs/2510.10728
4. Persistence Landscapes—Implementing a Dataset Verification Method in Resource-Scarce Embedded Systems - MDPI, accessed on February 11, 2026, https://www.mdpi.com/2073-431X/12/6/110
5. Change Point Detection in Financial Market Using Topological Data Analysis - MDPI, accessed on February 11, 2026, https://www.mdpi.com/2079-8954/13/10/875
6. Class of topological portfolios: Are they better than classical portfolios? - arXiv, accessed on February 11, 2026, https://www.arxiv.org/pdf/2601.03974
7. the journal of - Portfolio Management Research, accessed on February 11, 2026, https://www.pm-research.com/content/iijjfds/7/1/local/complete-issue.pdf
8. [2503.23757] Time-Series Forecasting via Topological Information Supervised Framework with Efficient Topological Feature Learning - arXiv, accessed on February 11, 2026, https://arxiv.org/abs/2503.23757
9. Time-Series Forecasting via Topological Information Supervised Framework with Efficient Topological Feature Learning | Request PDF - ResearchGate, accessed

on February 11, 2026, https://www.researchgate.net/publication/396799095_Time-Series_Forecasting_via_Topological_Information_Supervised_Framework_with_Efficient_Topological_Feature_Learning
10. New Directions in the Applications of Rough Path Theory | Request PDF - ResearchGate, accessed on February 11, 2026, https://www.researchgate.net/publication/368433882_New_Directions_in_the_Applications_of_Rough_Path_Theory
11. Optimal Execution with Rough Path Signatures | Request PDF - ResearchGate, accessed on February 11, 2026, https://www.researchgate.net/publication/340960796_Optimal_Execution_with_Rough_Path_Signatures
12. Signature-Informed Transformer for Asset Allocation - arXiv, accessed on February 11, 2026, https://arxiv.org/html/2510.03129v3
13. Local Regression on Path Spaces with Signature Metrics - arXiv, accessed on February 11, 2026, https://arxiv.org/html/2510.16728v1
14. Signature-Based Models: Theory and Calibration | SIAM Journal on Financial Mathematics, accessed on February 11, 2026, https://epubs.siam.org/doi/10.1137/22M1512338
15. Rough Path Signatures: Learning Neural RDEs for Portfolio Optimization - arXiv, accessed on February 11, 2026, https://arxiv.org/html/2510.10728
16. RoughPy - SciPy Proceedings, accessed on February 11, 2026, https://proceedings.scipy.org/articles/DXWY3560
17. The Market Control Illusion: A Separated Control and Controllability Framework for Non-Monotonic Prediction–Profit Dynamics in Financial Markets - Preprints.org, accessed on February 11, 2026, https://www.preprints.org/manuscript/202512.2270/v1
18. CVPR Poster Hyperbolic Category Discovery, accessed on February 11, 2026, https://cvpr.thecvf.com/virtual/2025/poster/33798
19. Liu Hyperbolic Category Discovery CVPR 2025 Paper | PDF - Scribd, accessed on February 11, 2026, https://www.scribd.com/document/885996685/Liu-Hyperbolic-Category-Discovery-CVPR-2025-Paper
20. Hyperbolic Category Discovery, accessed on February 11, 2026, https://arxiv.org/abs/2504.06120
21. NeurIPS Poster HIDISC: A Hyperbolic Framework for Domain Generalization with Generalized Category Discovery, accessed on February 11, 2026, https://neurips.cc/virtual/2025/poster/119696
22. A Survey on Recent Advances in LLM-Based Multi-turn Dialogue Systems - ResearchGate, accessed on February 11, 2026, https://www.researchgate.net/publication/397078844_A_Survey_on_Recent_Advances_in_LLM-Based_Multi-turn_Dialogue_Systems
23. Exploring Hyperbolic Hierarchical Structure for Multimodal Rumor Detection - ACL Anthology, accessed on February 11, 2026, https://aclanthology.org/2025.findings-emnlp.8.pdf

24. OpenHype: Hyperbolic Embeddings for Hierarchical Open-Vocabulary Radiance Fields - arXiv, accessed on February 11, 2026, https://arxiv.org/html/2510.21441v1
25. Wasserstein Distance: A Comprehensive Guide for 2025 - Schattenkodierer - Shadecoder, accessed on February 11, 2026, https://www.shadecoder.com/de/topics/wasserstein-distance-a-comprehensive-guide-for-2025
26. A Survey on Optimal Transport for Machine Learning: Theory and Applications - IEEE Xplore, accessed on February 11, 2026, https://ieeexplore.ieee.org/iel8/6287639/10820123/10877795.pdf
27. Detection and Isolation of Incipiently Developing Fault Using Wasserstein Distance - MDPI, accessed on February 11, 2026, https://www.mdpi.com/2227-9717/10/6/1081
28. Optimal Transport based Drift Detection for Sensor ... - IIM Ahmedabad, accessed on February 11, 2026, https://www.iima.ac.in/sites/default/files/rnpfiles/16464727852021-09-01.pdf
29. Leveraging Optimal Transport for Distributed Two-Sample Testing - arXiv, accessed on February 11, 2026, https://arxiv.org/pdf/2506.16047
30. Optimal Transport in Math Finance | NYU Tandon School of Engineering, accessed on February 11, 2026, https://engineering.nyu.edu/events/2025/12/03/optimal-transport-math-finance
31. Specific Wasserstein Divergence Between Continuous Martingales | Mathematics of Operations Research - PubsOnLine, accessed on February 11, 2026, https://pubsonline.informs.org/doi/10.1287/moor.2025.0896
32. OptiMAG: Structure-Semantic Alignment via Unbalanced Optimal Transport - arXiv, accessed on February 11, 2026, https://arxiv.org/html/2601.22856v1
33. ICML 2025 Orals, accessed on February 11, 2026, https://icml.cc/virtual/2025/events/oral
34. Generative AI in Finance - NeurIPS, accessed on February 11, 2026, https://neurips.cc/virtual/2025/workshop/109564
35. Stefano Pasquali's research works | Gracie Square Hospital, New York, NY and other places, accessed on February 11, 2026, https://www.researchgate.net/scientific-contributions/Stefano-Pasquali-2225749182
36. FinCARE: Financial Causal Analysis with Reasoning and Evidence - arXiv, accessed on February 11, 2026, https://arxiv.org/html/2510.20221v1
37. FinCARE: Financial Causal Analysis with Reasoning and Evidence - OpenReview, accessed on February 11, 2026, https://openreview.net/pdf/cb7fe3187d61e796c5fccd42538d926f67a6e369.pdf
38. Individual Causal Inference with Structural Causal Model - arXiv, accessed on February 11, 2026, https://arxiv.org/abs/2506.17300
39. Graph-Temporal Contrastive Transformer for Financial Fraud Detection Using Transaction Behavior Modeling - MDPI, accessed on February 11, 2026, https://www.mdpi.com/1999-4893/18/12/770
40. 2025's Biggest Breakthroughs in Mathematics - YouTube, accessed on February 11, 2026, https://www.youtube.com/watch?v=hRpcWpAeWng

41. 2025 Simons Collaboration on Wave Turbulence Annual Meeting - Simons Foundation, accessed on February 11, 2026, https://www.simonsfoundation.org/event/simons-collaboration-on-wave-turbulence-annual-meeting-2025/
42. Mathematicians Crack 125-Year-Old Problem, Unite Three Physics Theories - Reddit, accessed on February 11, 2026, https://www.reddit.com/r/Physics/comments/1k2qtcz/mathematicians_crack_125yearold_problem_unite/
43. How a Breakthrough in Physics Can Revolutionise Climate Models, accessed on February 11, 2026, https://www.orfonline.org/expert-speak/how-a-breakthrough-in-physics-can-revolutionise-climate-models
44. [2503.01800] Hilbert's sixth problem: derivation of fluid equations via Boltzmann's kinetic theory - arXiv, accessed on February 11, 2026, https://arxiv.org/abs/2503.01800
45. Adapting the Navier-Stokes Equations to Economic Dynamics - SSRN, accessed on February 11, 2026, https://papers.ssrn.com/sol3/Delivery.cfm/5766223.pdf?abstractid=5766223&mirid=1
46. Modeling the Dynamics of Liquidity Flows and Systemic Risks Using Navier–Stokes Equations - arXiv, accessed on February 11, 2026, https://arxiv.org/pdf/2507.05287
47. A dynamic model for money velocity based on Navier-Stokes Equations of fluid motion, accessed on February 11, 2026, https://www.researchgate.net/publication/376956904_A_dynamic_model_for_money_velocity_based_on_Navier-Stokes_Equations_of_fluid_motion
48. AI-Assisted Mathematics And Recent Breakthroughs (2025) - IJCRT.org, accessed on February 11, 2026, https://mail.ijcrt.org/papers/IJCRT2601324.pdf
49. The Tiny Needle That Shapes AI: The Kakeya Conjecture and the Geometry of Understanding | by Aditya Inamdar | Medium, accessed on February 11, 2026, https://medium.com/@inamdaraditya98/the-tiny-needle-that-shapes-ai-the-kakeya-conjecture-and-the-geometry-of-understanding-63107056f1f0
50. A Once-in-a-Century Proof: The Kakeya Conjecture - YouTube, accessed on February 11, 2026, https://www.youtube.com/watch?v=5J3tYU_-IZI
51. 'Once in a Century' Proof Settles Math's Kakeya Conjecture - Institute for Advanced Study, accessed on February 11, 2026, https://www.ias.edu/news/once-century-proof-settles-maths-kakeya-conjecture
52. On the Erdos distinct distance problem in the plane - ResearchGate, accessed on February 11, 2026, https://www.researchgate.net/publication/47817494_On_the_Erdos_distinct_distance_problem_in_the_plane
53. NOTICES OF THE AMERICAN MATHEMATICAL SOCIETY, accessed on February 11, 2026, https://www.ams.org/journals/notices/200103/200103FullIssue.pdf?adat=March%202001&trk=&cat=none&type=.pdf