

Treatise on Mathematical Models of Universal Stochastic Predictors (USP)

Extended and Unified Version

Adaptive Meta-Prediction Development Consortium

February 19, 2026

Contents

1 Theoretical Foundations and Architecture	2
1.1 Probability Spaces and Filtrations	2
1.2 The System Meta-State (Ξ_t)	2
1.3 Optimal Prediction Problem	2
1.4 Universal System Architecture	2
2 Phase 1: System Identification Engine (SIA)	4
2.1 Formalization of the Functional State Vector	4
2.2 Stationarity and Ergodicity Analysis	4
2.2.1 Strong Stationarity	4
2.2.2 Fractional Integration and Differentiation	4
2.3 Holder Regularity Analysis	4
2.3.1 Local Singularity Spectrum	4
2.4 Semimartingale Decomposition	5
2.4.1 Quadratic Variation	5
2.4.2 Bichteler-Dellacherie Theorem	5
2.5 Spectral Operators and Information Flow	5
2.5.1 Koopman Operator (\mathcal{K})	5
2.5.2 Filtration Enlargement (Grossissement de Filtration)	5
3 Phase 2: Formalization of the Prediction Kernels	6
3.1 Branch A: Projection in Hilbert Spaces	6
3.1.1 Orthogonality Principle and Wiener-Hopf	6
3.1.2 Paley-Wiener Condition	6
3.2 Malliavin Calculus and Stochastic Sensitivity	6
3.2.1 Occone-Haussmann Representation Theorem	6
3.3 Branch B: Evolution Equations and Viscosity	7
3.3.1 Infinitesimal Generator	7
3.3.2 Crandall-Lions Viscosity Solutions	7
3.3.3 Entropy Principle for Neural Approximation (DGM)	7
3.4 Malliavin Calculus on Poisson Spaces	8
3.4.1 Ito Formula for Semimartingales with Jumps	9
3.4.2 Temporal Discretization Schemes and Dynamic Transition	9
3.5 Branch D: Signature and Rough Paths (Topological Invariance)	11
3.5.1 Geometric Rough Paths Space	11
3.5.2 Signature (\mathcal{S}) and Hopf Algebra	11
3.5.3 Reparametrization Invariance Lemma	11
3.5.4 T-Linear Predictor	11

4 Phase 3: Adaptive Weighting Orchestrator	12
4.1 Optimal Transport Dynamics and Wasserstein Geometry	12
4.2 Large Deviations and Contraction Principle	12
4.3 Geometric Coupling and Fisher-Rao Metric	12
4.4 Global Lyapunov Functional	13
5 Phase 4: Convergence and Global Stability	14
5.1 Mixing Conditions	14
5.2 Sanov Theorem and Large Deviations	14
5.3 Mean L^p Stability and Lyapunov	14
5.4 Sequential Complexity and Generalization Bounds	14
5.5 Change-Point Detection and Stopping Times	15
5.5.1 Adaptive Threshold for Heavy Tails	15
6 Unified Operational Differential Equation	17
6.1 Meta-State Dynamics	17
6.2 Global Existence and Uniqueness Theorem	17
A Robustness Postulate for Singularities	18

Chapter 1

Theoretical Foundations and Architecture

This treatise formalizes the construction of a stochastic prediction system capable of operating on dynamic processes whose underlying probability law is unknown *a priori*.

1.1 Probability Spaces and Filtrations

We define a complete probability space (Ω, \mathcal{F}, P) . The evolution of information is modeled by a filtration $\{\mathcal{F}_t\}_{t \geq 0}$ that satisfies the usual conditions (Dellacherie-Meyer):

1. **Completeness:** \mathcal{F}_0 contains all P -null sets of \mathcal{F} .
2. **Right-continuity:** $\mathcal{F}_t = \bigcap_{s > t} \mathcal{F}_s$ for all $t \geq 0$.

This ensures that stopping times (such as those defined by the CUSUM algorithm) are measurable and the process admits cadlag modifications.

1.2 The System Meta-State (Ξ_t)

To unify control and prediction dynamics, we define the meta-state at time t as the functional triple in a Banach space:

$$\Xi_t = (V_s(t), w_t, \mathcal{P}_h^\cup)$$

Where V_s is the identification state, w_t the weight distribution on the statistical manifold, and \mathcal{P}_h^\cup the active prediction operator.

1.3 Optimal Prediction Problem

Definition 1.1 (Optimal Prediction Problem). *Given a stochastic process $X = \{X_t : t \in T\}$, we seek the operator \mathcal{P}_h such that $\hat{X}_{t+h} = \mathcal{P}_h(X_s, s \leq t)$ minimizes the norm in $L^2(\Omega, \mathcal{F}, P)$:*

$$\hat{X}_{t+h} = \underset{Z \in L^2(\mathcal{F}_t)}{\operatorname{argmin}} E [\|X_{t+h} - Z\|^2] = E[X_{t+h} | \mathcal{F}_t]$$

1.4 Universal System Architecture

The system is structured into three operational phases:

1. **Identification Engine (SIA):** Functional operator $\Psi(X) \rightarrow \mathcal{C}$. To ensure continuity of control operators on multifractal processes, the codomain \mathcal{C} is defined as the Besov space $B_{p,q}^s(\mathbb{R})$, which characterizes local singularities via wavelet decompositions.

2. **Prediction Kernels (\mathcal{P}_i):** Branches A (Hilbert), B (Markov/Fokker-Planck), C (Ito/Levy), D (Rough Paths/Signature).
3. **Adaptive Orchestrator (\mathcal{O}):** Optimal transport dynamics in the probability measure space $\mathcal{P}_2(\Omega)$ endowed with the Wasserstein metric.

Chapter 2

Phase 1: System Identification Engine (SIA)

The SIA characterizes the process topology via a functional state vector V_s .

2.1 Formalization of the Functional State Vector

The vector $V_s(t)$ consolidates structural metrics of the process:

$$V_s(t) = [d(t), \alpha(t), \sigma(\mathcal{K}), \mathcal{T}_{Y \rightarrow X}, [X]_t]^\top \in \mathcal{C}$$

2.2 Stationarity and Ergodicity Analysis

2.2.1 Strong Stationarity

The operator Ψ verifies invariance of the image measure under the time-translation group $\{\theta_\tau\}_{\tau \in \mathbb{R}}$:

$$P \circ \theta_\tau^{-1} = P \quad \forall \tau$$

2.2.2 Fractional Integration and Differentiation

For long-memory processes, we define the inverse of the unilateral Riesz kernel I^α :

$$Y_t = D^\alpha X_t = \frac{1}{\Gamma(-\alpha)} \int_{-\infty}^t (t-s)^{-\alpha-1} X_s ds$$

This generalizes the operator $(1 - L)^d$ to the continuous setting.

2.3 Holder Regularity Analysis

2.3.1 Local Singularity Spectrum

Local regularity is characterized by the pointwise Holder exponent $\alpha(t_0)$, defined as the supremum of α such that:

$$\limsup_{\epsilon \rightarrow 0} \frac{|X(t_0 + \epsilon) - X(t_0)|}{|\epsilon|^\alpha} < \infty$$

The function $\alpha(t)$ induces a stratification of the time domain $\bigcup_h \{t : \alpha(t) = h\}$.

2.4 Semimartingale Decomposition

2.4.1 Quadratic Variation

The quadratic variation process is defined as the uniform-in-probability limit:

$$[X]_t = P - \lim_{\|\Pi\| \rightarrow 0} \sum_i (X_{t_i} - X_{t_{i-1}})^2$$

2.4.2 Bichteler-Dellacherie Theorem

If X_t is an L^0 stochastic integrator, it admits the canonical decomposition:

$$X_t = X_0 + M_t + A_t$$

where M_t is a local martingale and A_t is a predictable finite-variation process.

2.5 Spectral Operators and Information Flow

2.5.1 Koopman Operator (\mathcal{K})

We define the composition operator on the observable space $L^\infty(\Omega)$:

$$\mathcal{K}^t g(\omega) = g(\theta_t \omega)$$

The point spectrum $\sigma_p(\mathcal{K})$ characterizes ergodic invariants of the dynamical system.

2.5.2 Filtration Enlargement (Grossissement de Filtration)

Let $\mathbb{G} = \{\mathcal{G}_t\}_{t \geq 0}$ be an enlargement of the original filtration \mathbb{F} such that $\mathcal{F}_t \subset \mathcal{G}_t$. By the Jeulin-Yor theorem, if hypothesis (H) fails, the \mathbb{F} -martingale M_t decomposes in \mathbb{G} as:

$$M_t = \tilde{M}_t + \int_0^t \alpha_s ds$$

where \tilde{M} is a \mathbb{G} -martingale and α_s is the information drift process. This formalizes the assimilation of exogenous latent variables.

Chapter 3

Phase 2: Formalization of the Prediction Kernels

3.1 Branch A: Projection in Hilbert Spaces

The predictor is defined in the space $\mathcal{H}_t = \overline{\text{span}}\{X_s : s \leq t\}$.

3.1.1 Orthogonality Principle and Wiener-Hopf

The prediction error must be orthogonal to the past history:

$$\langle X_{t+h} - \hat{X}_{t+h}, X_s \rangle = 0 \quad \forall s \leq t$$

This leads to the Wiener-Hopf integral equation for the optimal impulse response kernel $h(\tau)$:

$$\gamma(t + h - s) = \int_0^\infty h(\tau) \gamma(s - \tau) d\tau, \quad s \geq 0$$

3.1.2 Paley-Wiener Condition

To guarantee causality and the existence of spectral factorization $f(\lambda) = |\Psi(i\lambda)|^2$, we require:

$$\int_{-\infty}^\infty \frac{|\log f(\lambda)|}{1 + \lambda^2} d\lambda < \infty$$

3.2 Malliavin Calculus and Stochastic Sensitivity

We consider the canonical Wiener space (Ω, \mathcal{F}, P) and the Cameron-Martin subspace $H = L^2([0, T])$. We define the Malliavin derivative operator $D : \mathbb{D}^{1,2} \rightarrow L^2(\Omega; H)$ on cylindrical functionals $F = f(W(h_1), \dots, W(h_n))$ as:

$$D_t F = \sum_{i=1}^n \partial_i f(W(h_1), \dots, W(h_n)) h_i(t)$$

3.2.1 Ocone-Haussmann Representation Theorem

Every functional $F \in \mathbb{D}^{1,2}$ admits the integral representation:

$$F = E[F] + \int_0^T E[D_t F \mid \mathcal{F}_t] dW_t$$

This explicitly characterizes the integrand in the martingale decomposition of the optimal predictor as the conditional projection of the Malliavin derivative.

3.3 Branch B: Evolution Equations and Viscosity

3.3.1 Infinitesimal Generator

The evolution of the probability density $p(x, t)$ is governed by the adjoint operator \mathcal{L}^* . We consider the value function $V(t, x)$ associated with optimal stochastic control, which satisfies the Hamilton-Jacobi-Bellman (HJB) equation:

$$-\partial_t V + \inf_{u \in U} \{-\mathcal{L}^u V - f(x, u)\} = 0$$

3.3.2 Crandall-Lions Viscosity Solutions

Since V may be non-differentiable in C^2 , we define the solution in the viscosity sense. An upper semicontinuous function u is a viscosity subsolution of $F(x, u, Du, D^2u) = 0$ if for all $\phi \in C^2$ such that $u - \phi$ attains a local maximum at x_0 :

$$F(x_0, u(x_0), D\phi(x_0), D^2\phi(x_0)) \leq 0$$

This formulation guarantees existence and uniqueness for degenerate Hamiltonians typical in robust prediction.

3.3.3 Entropy Principle for Neural Approximation (DGM)

In the numerical implementation of Branch B via the Deep Galerkin Method (DGM), the neural network $V_\theta(t, x)$ approximates the value function. To ensure the neural solution is non-degenerate and captures the essential PDE structure, we define an entropy criterion.

Theorem 3.1 (Entropy Conservation Principle for Solution). *Let $V_\theta : [0, T] \times \Omega \rightarrow \mathbb{R}$ be the neural approximation of the value function satisfying the HJB equation, and let $g : \Omega \rightarrow \mathbb{R}$ be the terminal condition. Define the differential entropy of the solution at time t as:*

$$H_t[V_\theta] = - \int_{\Omega} p_t(v) \log p_t(v) dv$$

where $p_t(v)$ is the empirical probability density of values $\{V_\theta(t, x_i)\}_{i=1}^N$ on a grid $\{x_i\}$ that discretizes the domain Ω .

For the neural solution to be admissible, it must satisfy the entropy conservation criterion:

$$\frac{1}{T} \int_0^T H_t[V_\theta] dt \geq \gamma \cdot H[g]$$

where:

- $H[g] = - \int p_g(v) \log p_g(v) dv$ is the entropy of the terminal condition
- $\gamma \in [0.5, 1.0]$ is the entropy retention factor

This criterion prevents mode collapse, where the neural network converges to a constant or minimum-variance solution that trivially satisfies the PDE.

Proof. The value function $V(t, x)$ inherits informative structure from the terminal condition $g(x)$ through dynamic programming. Formally, for a finite-horizon optimal control problem with horizon T :

$$V(t, x) = \inf_{u \in \mathcal{U}} E \left[\int_t^T f(X_s^{t,x,u}, u_s) ds + g(X_T^{t,x,u}) \mid \mathcal{F}_t \right]$$

The conditional expectation is a contraction operator in L^2 , but it preserves information in the following sense: if g has diffuse support (high entropy), then $V(t, \cdot)$ for $t < T$ must also exhibit proportional spatial variability.

More precisely, consider the backward evolution of entropy. If $V(t, x)$ were constant in x for some $t < T$, then:

$$\nabla_x V(t, x) \equiv 0 \implies u^*(t, x) = \arg \max_u \{b(x, u) \cdot 0 + \dots\}$$

which implies the optimal policy u^* does not depend on the state x . This contradicts the non-triviality of $g(x)$ (assumed non-constant).

The entropy inequality follows by applying Jensen's inequality to the concave function $-x \log x$ combined with the comparison theorem for viscosity solutions, which guarantees that if V_θ approximates V well in L^∞ , then their induced distributions are close in the Kullback-Leibler sense, and thus their entropies are comparable.

Formally, under the Wasserstein metric between the pushforward measures:

$$W_2(V_\theta(\cdot, t)_\# \mu, V(\cdot, t)_\# \mu) < \epsilon \implies |H[V_\theta(\cdot, t)] - H[V(\cdot, t)]| < C\epsilon$$

where μ is the spatial measure on Ω and C is a domain-dependent constant. \square

Remark 3.1. *In practice, the entropy criterion is monitored during DGM training. If $H_t[V_\theta]$ falls persistently below the threshold $\gamma H[g]$, it is recommended to:*

1. Increase network capacity (more layers or neurons)
2. Adjust the learning rate to avoid premature convergence to trivial local minima
3. Modify the loss function to include an entropy regularization term:

$$\mathcal{L}_{total} = \mathcal{L}_{PDE} + \mathcal{L}_{BC} - \lambda H_t[V_\theta]$$

where $\lambda > 0$ penalizes low-entropy solutions

Corollary 3.1 (Relation to Solution Variance). *In the case of terminal conditions with approximately Gaussian distribution, the differential entropy relates to variance by:*

$$H[X] = \frac{1}{2} \log(2\pi e \sigma^2)$$

Therefore, the entropy conservation criterion implies variance conservation:

$$Var_x[V_\theta(t, x)] \geq \gamma' \cdot Var[g(x)]$$

with $\gamma' = \exp(2\gamma - 1)$. This is precisely the mode-collapse test described in the test document.

3.4 Malliavin Calculus on Poisson Spaces

For processes with jump components (Branch C), we extend the derivative operator D_t to the canonical Poisson space $(\Omega, \mathcal{F}, P, N)$. We define the difference operator $\mathcal{D}_{t,z}$ for functionals $F \in \mathbb{D}^{1,2}(\mu)$:

$$\mathcal{D}_{t,z} F(\omega) = F(\omega + \delta_{(t,z)}) - F(\omega)$$

The integrand in the predictable representation for the pure-jump martingale is given by the predictable projection of $\mathcal{D}_{t,z} F$.

3.4.1 Ito Formula for Semimartingales with Jumps

The process X_t decomposes according to the canonical Levy-Ito structure:

$$X_t = X_0 + \int_0^t b(X_{s-})ds + \int_0^t \sigma(X_{s-})dW_s + \int_0^t \int_{\mathbb{R}^n} z \tilde{N}(ds, dz)$$

The conditional expectation $u(t, x)$ satisfies the associated partial integro-differential equation (PIDE) for the generator \mathcal{L}^ν :

$$\mathcal{L}^\nu \phi(x) = \frac{1}{2}\sigma^2 \Delta \phi + b \cdot \nabla \phi + \int_{\mathbb{R}^d} [\phi(x+z) - \phi(x) - z \cdot \nabla \phi \mathbb{1}_{|z| \leq 1}] \nu(dz)$$

3.4.2 Temporal Discretization Schemes and Dynamic Transition

In the numerical implementation of Ito/Levy processes, the integration scheme selection is critical. Stochastic processes may exhibit variable stiffness, where temporal decompositions require implicit schemes for stability and convergence.

Stiffness Problem in Stochastic SDEs

We formally define stiffness of an SDE by the eigenvalue ratio of the diffusion operator:

$$\text{Stiffness Ratio} = \frac{\lambda_{\max}(J_\sigma)}{\lambda_{\min}(J_\sigma)}$$

where $J_\sigma = \frac{\partial}{\partial x}[\sigma(x)]$ is the Jacobian matrix of the diffusion coefficient. Processes with high ratios ($> 10^2$) exhibit multiscale dynamics where the explicit Euler-Maruyama scheme diverges.

Specifically, the explicit scheme:

$$X_{n+1} = X_n + b(X_n)\Delta t + \sigma(X_n)\Delta W_n$$

requires the time step Δt to satisfy the stability condition:

$$\Delta t < \frac{2}{\lambda_{\max}(J_b + J_\sigma^2)} \quad (\text{stochastic CFL condition})$$

where J_b, J_σ are the drift and diffusion Jacobians. In high-stiffness regimes, this bound is extremely restrictive, leading to prohibitive computational complexity.

Dynamic Transition Algorithm Between Schemes

We propose an adaptive algorithm that monitors stiffness in real time and transitions dynamically between the explicit scheme (fast, less accurate) and the implicit scheme (slow, robust):

- 1. Local Jacobian Estimation:** At each step n , we estimate Jacobians via finite differences:

$$\hat{J}_\sigma(X_n) \approx \frac{\sigma(X_n + \varepsilon e_i) - \sigma(X_n - \varepsilon e_i)}{2\varepsilon}$$

where $\varepsilon \sim 10^{-6}$ and e_i are basis vectors.

- 2. Stiffness Metric Computation:** We define the normalized stiffness metric:

$$S_t = \max \left(\text{Stiffness Ratio}, \left| \frac{d \log \sigma_t}{dt} \right| \cdot \Delta t \right)$$

The second term captures abrupt volatility changes.

3. **Scheme Decision Rule:** We establish adaptive thresholds:

$$\text{Scheme} = \begin{cases} \text{Explicit Euler} & \text{if } S_t < \theta_L \text{ (low stiffness)} \\ \text{Hybrid Transition} & \text{if } \theta_L \leq S_t < \theta_H \text{ (medium)} \\ \text{Implicit Euler} & \text{if } S_t \geq \theta_H \text{ (high stiffness)} \end{cases}$$

Typical thresholds are: $\theta_L = 100$, $\theta_H = 1000$.

4. **Hybrid Scheme in Transition:** In the intermediate region, we employ a convex mixture:

$$X_{n+1}^{(\lambda)} = (1 - \lambda)X_{n+1}^{(\text{exp})} + \lambda X_{n+1}^{(\text{imp})}$$

where $\lambda = \frac{S_t - \theta_L}{\theta_H - \theta_L} \in [0, 1]$ smoothly interpolates between schemes.

5. **Implicit Scheme for High Stiffness:** We use the Moulton-trapezoidal method:

$$X_{n+1} = X_n + \frac{\Delta t}{2} \left[b(X_n) + b(X_{n+1}^{(p)}) \right] + \sigma \Delta W_n$$

where $X_{n+1}^{(p)}$ is the explicit predictor. The implicit correction reduces truncation error and stabilizes diverging trajectories.

Strong Convergence Analysis

For the adaptive hybrid scheme, the global strong error is bounded by:

Theorem 3.2 (Adaptive Convergence Error). *Let $\{X_t\}$ satisfy the Ito SDE with Lipschitz coefficients and linear growth. For the dynamic transition scheme with steps $\Delta t_n = \Delta t / (1 + S_n)$ (where S_n is the stiffness metric at step n), the strong convergence error after N steps is:*

$$\mathbb{E}[|X_T - X_N|] \leq C \cdot \left(\sum_{n=0}^{N-1} (\Delta t_n)^{1.0} + \lambda_n \cdot (\Delta t_n)^{1.5} \right)$$

where λ_n is the adaptive coefficient (implicit fraction) at step n .

In low-stiffness regimes ($S_t < \theta_L$), the term $\lambda_n \approx 0$ recovers explicit Euler strong convergence of order 0.5 (weak order 1 with fixed Δt). In high-stiffness regimes ($S_t \geq \theta_H$), the dominant term is the implicit correction, ensuring stability without divergence.

Monitoring and Telemetry

During execution, the system records:

- **Scheme Frequency:** Proportion of steps with explicit vs implicit scheme. Expected: > 80% explicit in normal regime, < 50% in crisis.
- **Maximum Stiffness Metric:** $S_{\max}(t)$ over time. Alert threshold: $S_{\max} > 2000$ suggests processes with extreme fractal characteristics.
- **Number of Internal Iterations:** For iterative implicit methods (Newton), count correction iterations. Typical: 2–3 iterations; > 10 indicates anomalous behavior.
- **Implicit Residual Norm:** $\|X_{n+1}^{(k)} - X_{n+1}^{(k-1)}\|$ between iterations. Indicates convergence and potential numerical divergence.

Remark 3.2. In the context of the universal predictor, Branch C (Ito/Levy) alternates dynamically between schemes according to the local topology of the process X_t . This is especially important during regime transitions detected by CUSUM, where volatility multiplies by factors 3–10 in a few steps, inducing temporal stiffness that requires scheme changes to maintain numerical precision.

3.5 Branch D: Signature and Rough Paths (Topological Invariance)

For processes whose path roughness makes standard stochastic calculus infeasible (Holder exponent $H \leq 1/2$, variation $p \geq 2$), we operate within the Lyons rough paths framework.

3.5.1 Geometric Rough Paths Space

Let \mathbf{X} be a continuous process with values in the truncated tensor algebra $T^{(N)}(\mathbb{R}^d) = \bigoplus_{k=0}^N (\mathbb{R}^d)^{\otimes k}$. We define the space of geometric rough paths with finite p -variation $G\Omega_p(\mathbb{R}^d)$ as the closure of smooth paths under the p -variation metric:

$$d_p(\mathbf{X}, \mathbf{Y}) = \max_{k=1, \dots, [p]} \sup_{\mathcal{D}} \left(\sum_i |\mathbf{X}_{t_i, t_{i+1}}^k - \mathbf{Y}_{t_i, t_{i+1}}^k|^{p/k} \right)^{k/p}$$

3.5.2 Signature (\mathcal{S}) and Hopf Algebra

The signature map $\mathcal{S} : G\Omega_p([0, T], \mathbb{R}^d) \rightarrow T((\mathbb{R}^d))$ transforms the path into a formal series of non-commutative power series (Chen series):

$$\mathcal{S}(\mathbf{X})_{0,t} = 1 + \sum_{k=1}^{\infty} \int_{0 < u_1 < \dots < u_k < t} dX_{u_1} \otimes \dots \otimes dX_{u_k}$$

The image space is a Lie group under the \otimes operation, and its elements satisfy the shuffle product property for dual linear functionals $f, g \in T((\mathbb{R}^d))^*$:

$$\langle f, \mathbf{X} \rangle \langle g, \mathbf{X} \rangle = \langle f \amalg g, \mathbf{X} \rangle$$

This allows any continuous functional to be approximated by linear combinations of signature terms (Universal Approximation Theorem).

3.5.3 Reparametrization Invariance Lemma

Lemma 3.1. *The signature $\mathcal{S}(X)$ is invariant under any monotone time reparametrization $\psi(t)$:*

$$\mathcal{S}(X \circ \psi)_{0,T'} = \mathcal{S}(X)_{0,T}$$

This immunizes Branch D against irregular sampling noise, enabling a purely topological characterization.

3.5.4 T-Linear Predictor

The predictor is formalized as a linear functional in tensor space:

$$\hat{X}_{t+h} = \langle W, \mathbf{X}_{0,t} \rangle$$

Chapter 4

Phase 3: Adaptive Weighting Orchestrator

The Orchestrator \mathcal{O} manages the convex mixture $\hat{X}_{t+h}^{USP} = \sum w_i(t) \hat{X}_{t+h}^{(i)}$.

4.1 Optimal Transport Dynamics and Wasserstein Geometry

We consider the infinite-dimensional Riemannian manifold $\mathcal{M} = (\mathcal{P}_{ac}(\Delta^n), g_W)$ endowed with the metric structure W_2 . The free energy functional \mathcal{F} defines a gradient vector field $\nabla_{W_2}\mathcal{F}$. The evolution follows the JKO (Jordan-Kinderlehrer-Otto) flow, which is the limit as $\tau \rightarrow 0$ of the discrete variational scheme:

$$\rho_{k+1} \in \operatorname{argmin}_\rho \left\{ \frac{1}{2\tau} W_2^2(\rho, \rho_k) + \mathcal{F}(\rho) \right\}$$

This converges to the nonlinear Fokker-Planck equation:

$$\partial_t \rho = \nabla \cdot (\rho \nabla (\delta \mathcal{F} / \delta \rho))$$

4.2 Large Deviations and Contraction Principle

The convergence rate of the empirical measure L_n toward the invariant measure μ^* is governed by the action functional $I(\nu)$ (relative entropy or Kullback-Leibler divergence):

$$I(\nu) = \sup_{f \in C_b} \{ \langle f, \nu \rangle - \Lambda(f) \}$$

For dependent ϕ -mixing processes, the Large Deviations Principle (LDP) holds with a convex and lower semicontinuous rate function (good rate function).

4.3 Geometric Coupling and Fisher-Rao Metric

To incorporate sensitivity to the statistical manifold structure, we generalize the metric to a Hellinger-Kantorovich or Fisher-Rao structure deformed by the curvature tensor induced by the operator Ψ :

$$G(\rho) = e^{-\beta \|\nabla \Psi\|} G_{FR}(\rho)$$

where G_{FR} is the Fisher information metric. This defines an adaptive geodesic on the probability simplex.

4.4 Global Lyapunov Functional

The asymptotic stability of the orchestrator is established via the Lyapunov function based on relative entropy:

$$V(w) = \sum_{i \in \text{opt}} w_i^* \log \left(\frac{w_i^*}{w_i(t)} \right), \quad \frac{dV}{dt} \leq 0$$

Chapter 5

Phase 4: Convergence and Global Stability

5.1 Mixing Conditions

We assume β -mixing (absolute regularity) conditions with exponential decay:

$$\beta(\tau) = E \left[\sup_{B \in \mathcal{F}_{t+\tau}^\infty} |P(B|\mathcal{F}_{-\infty}^t) - P(B)| \right] \sim e^{-\lambda\tau}$$

5.2 Sanov Theorem and Large Deviations

The probability that the empirical error measure deviates from the optimal set decays exponentially:

$$P(\hat{L}_t \in \Gamma) \leq C \exp \left(-n \inf_{\nu \in \Gamma} I(\nu) \right)$$

5.3 Mean L^p Stability and Lyapunov

The exponential stability of the stochastic flow $\{\Xi_t\}$ is proven via the Foster-Lyapunov drift criterion for weakly continuous Markov generators. Let $V : \mathcal{H} \rightarrow \mathbb{R}_+$ be a compact Lyapunov function. If there exists a compact set (petite set) $K \subset \mathcal{H}$ and constants $\gamma > 0, b < \infty$ such that:

$$\mathcal{L}V(x) \leq -\gamma V(x) + b \mathbb{1}_K(x)$$

then the process is geometrically ergodic and admits a unique invariant measure π .

5.4 Sequential Complexity and Generalization Bounds

To bound excess risk in non-i.i.d. processes, we use conditional Rademacher complexity $\mathcal{R}_n(\mathcal{F}|\mathbf{x})$:

$$\mathcal{R}_n(\mathcal{F}|\mathbf{x}) = E_\sigma \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \sigma_i f(x_i) \right]$$

For β -mixing processes, the Bernstein blocking technique is applied to decompose temporal dependence and apply McDiarmid concentration inequalities.

5.5 Change-Point Detection and Stopping Times

We define the stopping time τ as the first barrier-crossing time of the generalized CUSUM process:

$$\tau = \inf\{t > 0 : \max_{0 \leq k \leq t} |S_t - S_k| \geq h(\Psi_t)\}$$

where S_t is the partial sum of standardized innovation residuals. Under the null hypothesis of stationarity, S_t converges weakly to a Brownian bridge. At time τ , the probability measure is reset to the uniform prior over the simplex: $\rho_\tau = \text{Unif}(\Delta^n)$ (maximum entropy).

5.5.1 Adaptive Threshold for Heavy Tails

In high-volatility regimes with non-Gaussian distributions (fat tails), a threshold based solely on variance can generate false positives. We formalize an adaptive threshold that incorporates the fourth moment.

Lemma 5.1 (Adaptive Threshold with Kurtosis). *Let $\{Z_t\}$ be the standardized residual process with finite fourth moment. We define the adaptive threshold:*

$$h_t = k \cdot \sigma_t \cdot \left(1 + \beta \cdot \frac{\kappa_t - 3}{\kappa_0}\right)$$

where:

- σ_t is the rolling standard deviation of residuals
- $\kappa_t = \frac{E[(Z_t - \mu_t)^4]}{\sigma_t^4}$ is the kurtosis (excess relative to the normal distribution)
- $k \in [3, 5]$ is the base sensitivity factor
- $\beta \in [0.1, 0.3]$ is the heavy-tail adjustment coefficient
- $\kappa_0 = 3$ is the Gaussian reference kurtosis

For heavy-tailed distributions ($\kappa_t > 3$), the threshold increases proportionally, reducing false alarms while preserving detection power for genuine structural changes.

Proof. Under the Lorden sequential detection framework, the false-alarm probability $P(FA)$ is bounded by:

$$P(FA) \leq e^{-\theta h}$$

where θ depends on the mean change rate. For sub-Gaussian distributions (exponentially bounded), this inequality holds with universal constants.

For distributions with heavier tails than Gaussian, variance no longer controls tail mass. The fourth moment (kurtosis) captures the frequency of extreme events. Formally:

$$P(|Z_t - \mu_t| > c\sigma_t) \leq \frac{\kappa_t}{c^4} \quad (\text{fourth-order Markov inequality})$$

By adjusting the threshold h_t via the factor $(1 + \beta(\kappa_t - 3)/\kappa_0)$, we guarantee that the conditional probability of exceeding the threshold under the null hypothesis remains uniformly bounded:

$$P\left(\max_{0 \leq k \leq t} |S_t - S_k| > h_t \mid H_0\right) \leq \alpha$$

where α is the desired significance level, independent of the kurtosis regime. \square

Remark 5.1. This adjustment is particularly relevant for financial processes that exhibit empirical kurtosis $\kappa \in [5, 20]$ (leptokurtic distributions). Without correction, the standard CUSUM detector generates excess signals during periods of high seasonal volatility without underlying structural drift changes.

Corollary 5.1 (Asymptotic Consistency). *For a sequence of kurtosis-calibrated thresholds $\{h_t\}$, the stopping time τ satisfies:*

$$\lim_{n \rightarrow \infty} P(\tau > n \mid H_1) = 0$$

where H_1 denotes the alternative hypothesis of regime change. That is, the detector retains full asymptotic power regardless of tail structure.

Chapter 6

Unified Operational Differential Equation

6.1 Meta-State Dynamics

The complete system is described by a nonlinear stochastic differential equation in the functional Hilbert space $H = \mathcal{C} \times L^2(\Delta^n) \times \mathcal{L}(\mathcal{H}, \mathcal{H})$ governing the meta-state Ξ_t :

$$d\Xi_t = (\Xi_t, X_t)dt + (\Xi_t, X_t)dW_t$$

The drift encapsulates:

1. The topological identification of the SIA operator (Ψ) .
2. The Wasserstein gradient flow $\text{grad}_{W_2}\mathcal{F}$ projected onto the tangent space of measures.
3. The evolution of local predictors.

6.2 Global Existence and Uniqueness Theorem

Theorem 6.1 (Weak Existence and Uniqueness). *Assuming the coefficients and are measurable and satisfy local Lipschitz and linear growth conditions (or that is Holder continuous and bounded, invoking the Yamada-Watanabe criterion in finite dimension), there exists a unique weak solution $(\Omega, \mathcal{F}, P, W, \Xi)$ to the operational stochastic differential equation such that:*

$$E \left[\sup_{0 \leq s \leq T} \|\Xi_s\|^2 \right] < C(T, \|\Xi_0\|)$$

Appendix A

Robustness Postulate for Singularities

Postulate A.1. *If the SIA detects a Hausdorff dimension $D > 1$ or $\alpha(t) \rightarrow 0$, the system prioritizes Branch D (Signature) and activates Huber regularization. This guarantees predictor operability in extreme roughness regimes where standard stochastic differential calculus fails.*