Stage-10 v2 (Geodesic Parser + Executor): Mathematical Specification

1 Setup and Notation

Let $x \in \mathbb{R}^F$ denote the latent state per time step. We perform one-time PCA whitening:

$$y = W\,(x - \mu) \in \mathbb{R}^d, \qquad W = \text{PCA whitener}, \quad d \ll F. \tag{1}$$

We are given K primitive prototypes $p_k \in \mathbb{R}^d$ (e.g., $K = 3$ for `flip_h`, `flip_v`, `rotate`), normalized and nearly orthogonal:

$$\|p_k\|_2 = 1, \qquad p_i^\top p_j \approx 0 \ (i \neq j). \tag{2}$$

Each primitive also has an *anchor* (center) $c_k \in \mathbb{R}^d$ estimated from data (e.g., per-class mean in the whitened space). A geodesic rollout produces a trajectory $t \mapsto y(t)$, $t = 0, \dots, T$.

2 Geodesic Mechanics

2.1 Action, Potentials, and Forces

In the whitened chart we take the metric $G = \mathbf{I}_d$ so free geodesics are straight lines. We *guide* the trajectory by a scalar potential that encodes active primitives. Let

$$U_k(y) \;=\; \tfrac{1}{2}\,\|\Pi_k\,(y - c_k)\|^2, \qquad \Pi_k \equiv \mathbf{I} - p_k p_k^\top, \tag{3}$$

so that U_k is minimized when y lies on the line $c_k + \alpha p_k$ (aligned with p_k and passing through c_k). Its gradient is

$$\nabla U_k(y) \;=\; \Pi_k\,(y - c_k), \tag{4}$$

which *removes* components orthogonal to p_k, pulling the state onto the p_k-axis.

Given (possibly time-varying) nonnegative weights $w_k(t)$ over primitives, define the composite potential

$$U(y,t) \;=\; \sum_{k=1}^{K} w_k(t)\, U_k(y). \tag{5}$$

The controlled Lagrangian is

$$\mathcal{L}(y, \dot{y}, t) \;=\; \tfrac{1}{2} m \|\dot{y}\|^2 - \lambda\, U(y,t), \tag{6}$$

with mass $m > 0$ and coupling $\lambda > 0$. Including Rayleigh dissipation $\mathcal{R} = \frac{1}{2}\gamma\|\dot{y}\|^2$ yields the Euler–Lagrange equations

$$m\,\ddot{y} \;=\; -\lambda\,\nabla U(y,t) \;-\; \gamma\,\dot{y} \;=\; -\lambda \sum_k w_k(t)\,\Pi_k\,(y - c_k) \;-\; \gamma\,\dot{y}. \tag{7}$$

Single-primitive runs. For the *oracle* or prototype runs used to estimate per-primitive energy traces, set $w_k(t) \equiv \mathbb{1}[k = k^\star]$ constant in time; Eq. (7) aligns the state with $p_{k^\star}$ (orthogonal components decay via the potential and damping).

Mixture control. For multi-task scenarios, $w_k(t)$ encodes a *schedule* over primitives (sequential or overlapping). Stage-10 v2 *parses* $w_k(t)$ from traces (Sec. 4) and then executes in that order (possibly with concurrency).

2.2 Discrete Integrator (CPU-friendly)

Let $\Delta t > 0$ be the step size. A stable semi-implicit Euler (or damped Verlet) discretization of (7) is:

$$v_{t+\frac{1}{2}} = \alpha\, v_t - \frac{\lambda\, \Delta t}{2m}\, \nabla U(y_t, t), \tag{8}$$

$$y_{t+1} = y_t + \Delta t\, v_{t+\frac{1}{2}}, \tag{9}$$

$$v_{t+1} = \alpha\, v_{t+\frac{1}{2}} - \frac{\lambda\, \Delta t}{2m}\, \nabla U(y_{t+1}, t + \Delta t), \tag{10}$$

with $\alpha = (1 - \frac{\gamma \Delta t}{2m})$. Parameters $\{m, \lambda, \gamma, \Delta t\}$ match the implementation knobs (`mass_scale`, `lambda`, `gamma`, `dt`).

3 Energies Along the Geodesic

Given a rollout $\{y(t)\}_{t=0}^{T}$, define (with c_k optionally set to 0 if data is zero-centered)

$$s_k(t) \;=\; \langle p_k |\, y(t) - c_k \,\rangle\,, \qquad \text{(signed alignment)} \tag{11}$$

$$E_{\parallel}^{(k)}(t) \;=\; s_k(t)^2 \;=\; \left(p_k^{\top}(y(t) - c_k)\right)^2, \tag{12}$$

$$E_{\perp}^{(k)}(t) \;=\; \|y(t) - c_k\|^2 - E_{\parallel}^{(k)}(t). \tag{13}$$

We then smooth each series with a short FIR/moving average to obtain $\widetilde{E_{\parallel}}^{(k)}$ and $\widetilde{E_{\perp}}^{(k)}$. Intuition: during a k-task window, $\widetilde{E_{\parallel}}^{(k)}$ peaks and $\widetilde{E_{\perp}}^{(k)}$ dips (relative to others).

3.1 Exclusive (Span-Complement) Residual

To suppress cross-talk between correlated prototypes, we use the *exclusive* residual for channel k:

$$r_k^{\text{ex}}(t) \;=\; \left(\mathbf{I} - Q_{-k} Q_{-k}^{\top}\right) z_k(t), \qquad z_k(t) = \text{zscore}\big(E_{\perp}^{(k)}(t)\big), \tag{14}$$

where Q_{-k} is an orthonormal basis (via QR) for the column space spanned by $\{z_j\}_{j \neq k}$. The parsing signal is then

$$\mathcal{E}^{(k)}(t) \;=\; \left[\, r_k^{\text{ex}}(t)\,\right]_+ * h, \tag{15}$$

i.e., positive part followed by a small smoothing kernel h; this is exactly what Stage-10 v2 integrates as area.

4 Parsing: Task Set, Order, Concurrency

4.1 Matched Filter and Peak Statistics

Let $q(\tau)$ be a fixed unimodal template ("bump"). The normalized cross-correlation for channel k is

$$C^{(k)}(t) \;=\; \frac{\sum_{\tau} \big(\mathcal{E}^{(k)}(t-\tau) - \overline{\mathcal{E}^{(k)}}\big)\big(q(\tau) - \bar{q}\big)}{\sqrt{\sum_{\tau}(\mathcal{E}^{(k)}(t-\tau) - \overline{\mathcal{E}^{(k)}})^2}\, \sqrt{\sum_{\tau}(q(\tau) - \bar{q})^2}}\,. \tag{16}$$

Let $t_k^\star = \arg\max_t C^{(k)}(t)$ and define a window $[a_k, b_k]$ around $t_k^\star$ (e.g., half-max width in $\mathcal{E}^{(k)}$). The area score is

$$A^{(k)} \;=\; \sum_{t=a_k}^{b_k} \mathcal{E}^{(k)}(t). \tag{17}$$

4.2 Decision Rules (Stage-10 v2)

1. **Presence:** include primitive k iff $A^{(k)} > \tau_{\text{area}}$ and $\max_t C^{(k)}(t) > \tau_{\text{corr}}$.
2. **Order:** sort included primitives by $t_k^\star$ (or by matched-filter lag; both are supported).
3. **Concurrency:** if windows overlap, $[a_i, b_i] \cap [a_j, b_j] \neq \emptyset$, add a concurrency edge (i, j).
4. **Confidence:** report $(A^{(k)}, \max C^{(k)})$ per task; overall confidence may be the harmonic mean of per-task normalized scores.

5 Executor (Primitive Geodesics)

For execution, we run (7) with a *single* active primitive at a time following the parsed order:

$$w_k(t) = \begin{cases} 1 & \text{if } k \text{ is the current primitive,} \\ 0 & \text{otherwise.} \end{cases}$$

At the end of each task window we update the grid/latent with the primitive's map (e.g., flip or rotation for ARC synthetics). Concurrency can be handled by either (i) short alternating micro-steps between active primitives or (ii) superposing small weights w_i, w_j during the overlap.

6 Parameters to Freeze (for Reproducibility)

Symbol/Flag	Typical Value	Meaning
d	19	PCA-whitened dimension
m	4.0	mass (`mass_scale`)
λ	0.35	potential coupling (`lambda`)
γ	0.04	damping/friction
Δt	0.02	integrator time step
T	600–720	rollout steps
h	window 7–15	smoother kernel (moving average)
q	width 50–80	matched-filter template
τ_{area}	$\approx$ 10 (synthetic)	area floor for presence
τ_{corr}	$\approx$ 0.7 (synthetic)	correlation floor for presence

7 Why It Works (Invariants)

- **Whitened orthogonality:** PCA + unit prototypes make directions comparable; with near-orthogonality, $\sum_k E_\parallel^{(k)} \approx \|y\|^2$.
- **Geodesic locality:** under U_k, the system damps components orthogonal to p_k and travels along the p_k-axis (the intended primitive).
- **Additivity:** sequential (or overlapping) $w_k(t)$ produce separated (or overlapping) lobes in $\mathcal{E}^{(k)}$.

- **Exclusive residual:** projecting onto the span-complement of $\{j \neq k\}$ eliminates cross-talk and sharpens detection.
- **Template robustness:** matched filtering provides a normalized peak statistic stable to scale/drift/timing jitter.

8 Unit Tests to Lock In

1. **Primitive oracle (single):** $\arg\min_k \sum_t E_\perp^{(k)}(t)$ equals ground truth (observed 100% on ARC-12 primitives).
2. **Stage-10 v2 synthetic:** parser+executor recovers task set and order end-to-end (25/25 observed).
3. **Stability:** $\pm 10\%$ timing jitter and amplitude noise do not change decisions.
4. **Ablations:** removing whitening hurts accuracy; lowering τ_{corr} adds spurious tasks; $\pm 25\%$ change in q leaves decisions unchanged.

9 Minimal Algorithms (Pseudocode)

Parser (from geodesic traces)

```
Inputs: y[0:T], {p_k, c_k}, template q, smoother h
for k in K:
    s_k[t]    = dot(p_k, y[t]-c_k)
    E_par[k]  = smooth(s_k^2, h)
    E_perp[k] = smooth(||y[t]-c_k||^2 - s_k^2, h)
# exclusive residual across channels
E_ex[k] = pos( (I - Q_{-k}Q_{-k}^T) zscore(E_perp[k]) )
# matched filter + decisions
C[k]    = nxcorr(E_ex[k], q); t*_k = argmax C[k]
A[k]    = sum_window(E_ex[k] around t*_k)
tasks   = { k : A[k] > tau_area and max C[k] > tau_corr }
order   = sort_by t*_k; concurrency = overlap(windows)
```

Executor (primitive geodesics)

```
for primitive k in order:
    set weights w_k(t)=1 (others 0)
    integrate m y'' = -lambda * grad U_k(y) - gamma y'
    apply primitive map to grid/latent at window end
# handle overlaps by alternating small steps or w_i,w_j >0 simultaneously
```

10 Empirical Summary

- **Stage-10 v2 (synthetic):** 25/25 correct (task set and order) using the parser + executor.
- **Oracle single-primitive:** 12/12 correct using the perpendicular-energy criterion.
- Visuals show clean, separable lobes in $\mathcal{E}^{(k)}$, high matched-filter peaks (~ 0.97) for active tasks, and stable ordering.