

Article

Not peer-reviewed version

A Certificate Layer for Multi-Objective Training Across Refinement Scales

[Camilla Josephson](#)*

Posted Date: 22 December 2025

doi: 10.20944/preprints202512.1864.v1

Keywords: certificate layer; proof-carrying verification; refinement ladders; multi-objective learning; nonnegative ledgers; positive systems; Metzler matrices; copositive Lyapunov functions; small-gain; projective limits; dictionary transport



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Certificate Layer for Multi-Objective Training Across Refinement Scales

Camilla Josephson

Independent Researcher; Sweden; cmkjosephson@gmail.com

Abstract

Modern training pipelines are governed by multiple coupled nonnegative metrics (performance, constraints, robustness, calibration, and compute budgets) and are rerun across refinement ladders (width/depth scaling, discretization, basis growth, and data-fidelity upgrades). This paper develops a certificate layer for such settings. First, a Metzler comparison system for the ledger vector admits a one-clock reduction: if a Hurwitz witness exists, then a declared positive scalarization contracts exponentially, with explicit two-ledger small-gain formulas. Second, a Master Certificate upgrades four auditable programme lines verified in a single declared ruler—summable tail disturbances, a uniform contraction margin, projective Cauchy consistency, and a uniform dictionary for reported readouts—to existence and uniqueness of a refinement-limit learner on the certification window, together with rate inheritance and readout transport. The framework yields a proof-carrying stability grammar for learning under updates and refinement, intended to compose with external generalization and robustness modules.

Keywords: certificate layer; proof-carrying verification; refinement ladders; multi-objective learning; nonnegative ledgers; positive systems; Metzler matrices; copositive Lyapunov functions; small-gain; projective limits; dictionary transport

1. Introduction

Modern training pipelines are governed by *multiple* coupled nonnegative quantities: predictive performance, constraint and safety violations, robustness proxies, calibration errors, and compute/latency budgets. In multi-task and multi-constraint settings these ledgers can genuinely compete, so a single weighted loss is often only a proxy for what a pipeline is required to guarantee [1,2]. A second ubiquitous feature is *refinement*: the same pipeline is rerun across a ladder of resolutions (width/depth scaling [3], discretization refinement, feature-basis growth, data fidelity upgrades). In scientific computing, refinement limits are meaningful only once all levels are compared in a fixed energy norm and stable transfer maps are specified [4–6].

1.1. Problem and Certificate Goal

Problem (multi-ledger learning under refinement). Training is repeated across refinement levels, producing (i) learner trajectories/iterates and (ii) a vector of nonnegative ledgers tracking performance, constraints, robustness, calibration, and resource budgets. We seek *auditable* conditions under which:

- (i) a refinement-limit learner exists and is unique in a fixed, declared ruler, and
- (ii) the multi-ledger evolution admits a *single certified contraction clock* (a uniform exponential rate class) that does not degrade as the refinement index grows.

The formal ladder objects (state spaces, coarse maps/projections, and the declared ruler) are fixed once in Section 2. The two theorem engines that turn auditable inequalities into guarantees are proved in Sections 3 and 4, with instantiations in Section 5.

1.2. Context and Gap

Two adjacent traditions motivate the problem but do not, by themselves, provide a checker-facing interface.

Multiobjective learning (within a fixed representation).

Pareto structure and scalarization are classical in multiobjective optimization [1], and modern multi-task learning can be framed as multiobjective training with gradient-based trade-off schemes [2]. This line clarifies how to negotiate competing objectives at a given scale; it does not provide a refinement-consistent, verifier-friendly implication that survives changes of resolution.

Refinement limits (stability in a fixed norm).

In numerical analysis and multilevel methods, convergence across discretizations becomes meaningful only after fixing a single norm/ruler and stable transfer operators, as in stability–consistency paradigms and energy-norm FEM/multigrid constructions [4–7]. Semigroup approximation theory plays a related role for evolution families [8–10]. What is missing for modern training loops is a compact certificate interface that simultaneously (i) handles several interacting *nonnegative* quantities and (ii) remains coherent along a refinement ladder, returning a single auditable decay rate.

1.3. State of the Art, Approach, and Contributions

The paper sits at the intersection of three adjacent traditions.

State of the art (adjacent lines).

(i) *Multiobjective learning and scalarization*. The standard formal language for competing objectives is multiobjective optimization and Pareto trade-offs [1]; in modern ML, multi-task procedures can be interpreted as multiobjective updates with explicit trade-off mechanisms [2]. This line clarifies how to balance objectives at a fixed representation level, but it does not by itself provide a verifier-facing implication that survives changing resolution.

(ii) *Positive systems and comparison principles*. Metzler matrices, order-preserving comparison, and Perron–Frobenius structure form the classical backbone of positive linear systems and cooperative dynamics [11–13]. Copositive (linear) Lyapunov functions and robust/synthesis viewpoints for positive systems make the contraction certificates particularly witness-friendly [14]. What is missing for ML pipelines is an explicit “certificate grammar” that turns such comparison bounds into a shipped artifact and a small checker.

(iii) *Refinement ladders and stable limits*. In numerical analysis, meaningful convergence across discretizations requires a fixed measurement convention (energy norm) and stable transfer operators; this underlies stability–consistency paradigms and multilevel methods [4–7]. Abstract semigroup approximation theory provides a complementary operator-theoretic view of consistent approximations of evolution families [8–10]. These traditions explain why “one ruler” is non-negotiable, but they do not natively address coupled multi-ledger training objectives nor provide a checker interface for ML systems.

Approach (two theorem engines).

We separate the solution into two theorem engines: an *in-level* reduction that collapses many nonnegative ledgers to one certified clock, and a *cross-level* reduction that turns summable refinement discrepancies into a limit object in one ruler.

Engine A (positivity \Rightarrow one clock).

At each level K we target a cooperative comparison inequality of Metzler type for the ledger vector, possibly with a nonnegative remainder. In this setting, positive-systems theory supplies checker-friendly copositive Lyapunov witnesses and Perron–Frobenius structure [11–14]. The certificate layer

is organized so that a prover can ship a short witness (w, λ) and a verifier can check the one-clock claim by a finite set of inequalities (Section 3).

Engine B (one ruler + summable discrepancies \Rightarrow refinement limit).

Across levels, we realize $X^{(K)}$ inside a common ambient ruler space and compare levels only after applying the declared coarse projection. A summable tower defect (projective Cauchy condition) forces a Cauchy tower in the ambient ruler and therefore a unique refinement-limit learner on the declared window (Section 4). This is the refinement-ladder analogue of the basic “summable increments \Rightarrow convergence” mechanism in Banach spaces [15].

Contributions (with pointers).

- C1: One-clock reduction for nonnegative ledgers.** We develop a checker form of the Metzler/Hurwitz implication: a copositive witness yields exponential contraction of a declared scalar ledger, with an explicit two-ledger small-gain boundary and monotone design levers (Section 3).
- C2: Master Certificate across refinement ladders.** We prove that four auditable programme lines in one ruler (tail-robust envelope, uniform margin, projective Cauchy tower, uniform dictionary) imply existence/uniqueness of a refinement-limit learner on $[0, T]$ together with rate inheritance and readout transport (Section 4).
- C3: Instantiations.** We give (i) a fully checkable toy ladder with explicit constants and (ii) a width-ladder protocol sketch indicating how to populate the certificate from logged traces (Section 5).
- C4: Scope wall and composability.** The proved guarantees are training-time stability and refinement-limit existence in a declared ruler. Statistical generalization, out-of-distribution shift, and stochastic-optimizer noise are not asserted unless they are explicitly ledgerized or imported as separate modules (e.g. stability/PAC-Bayes and robustness/verification frameworks) [16–22].

Certificate interface.

The paper adopts a proof-carrying viewpoint: the producer supplies a certificate artifact (declared ruler, transfer maps, ledger definitions, budgets, and—when applicable—a Metzler witness), and a consumer validates it with a small checker, analogous in spirit to proof-carrying code [23].

2. Problem: Multi-Metric Learning on Refinement Ladders

This section fixes the formal objects referenced by the certificate obligations. We specify (i) a refinement-indexed ladder of state spaces realized inside a single declared ambient ruler, (ii) coherent coarse-graining/projection maps along the ladder, (iii) training dynamics (discrete updates or continuous flows) on each level, and (iv) ledger/readout maps that turn states into multiple nonnegative metrics. This setup is chosen so that the programme lines (O1)–(O4) in Section 4 can be stated unambiguously and checked mechanically.

2.1. Learner State, Dynamics, and a Single Ambient Ruler

2.1.1. Refinement-Indexed State Spaces and One Ruler

Fix a refinement index set $\mathcal{K} = \{K_0, K_0 + 1, \dots\}$. For each $K \in \mathcal{K}$, let $X^{(K)}$ be a metric space of learner states at resolution K (parameters, buffers, optimizer state, constraint multipliers, etc.). We assume the ladder embeds into an ambient normed space $(X^{(\infty)}, \|\cdot\|)$ with a *single ruler*:

$$X^{(K)} \subset X^{(\infty)} \quad \text{for all } K,$$

and we fix projections $P_K : X^{(\infty)} \rightarrow X^{(K)}$ such that

$$P_K P_L = P_K \quad (K \leq L), \quad \|P_K z\| \leq \|z\|, \quad P_K z \rightarrow z \text{ in } X^{(\infty)} \text{ as } K \rightarrow \infty. \quad (1)$$

The identities in (1) encode a coherent ladder geometry: coarse-graining is consistent along the ladder and contractive in the declared ruler.

2.1.2. Training Dynamics: Discrete Updates and Continuous Flows

At each K , training is described either by a discrete-time update map

$$x_{n+1}^{(K)} = \Phi^{(K)}(x_n^{(K)}, \xi_n), \quad n = 0, 1, 2, \dots, \quad (2)$$

or by a continuous-time (possibly differential-inclusion) flow

$$\dot{x}^{(K)}(t) = F^{(K)}(x^{(K)}(t), t), \quad t \in [0, T]. \quad (3)$$

Here ξ_n denotes exogenous randomness (mini-batch sampling, dropout masks, noise injections) when present. For (2) we assume $\Phi^{(K)}(\cdot, \xi)$ is measurable (often locally Lipschitz). For (3) we assume $F^{(K)}$ is measurable in time and locally Lipschitz in state, ensuring existence/uniqueness of trajectories on $[0, T]$ [24,25]. Stochasticity can be absorbed into the certificate layer as a nonnegative disturbance budget; see Grönwall-type propagation [26] and stochastic approximation frameworks [27–29].

2.1.3. State Augmentation

The state $x^{(K)}$ may include optimizer momentum, running moments (Adam/EMA), constraint multipliers, safety filters, or internal solver iterates. Analytically, this is harmless: auxiliary variables can be stacked into a product space and measured with a product norm, and many stability properties are naturally expressed on the augmented state [25].

2.2. Ledgers and Reported Readouts

2.2.1. Ledger Maps and Trajectories

Fix an integer $m \geq 1$. For each refinement level K , a *ledger family* is a vector of nonnegative maps

$$\mathcal{R}^{(K)} : X^{(K)} \rightarrow \mathbb{R}_{\geq 0}^m, \quad \mathcal{R}^{(K)}(x) = (R_1^{(K)}(x), \dots, R_m^{(K)}(x))^\top.$$

Each $R_i^{(K)}$ is a measurable functional representing a tracked metric (risk proxy, constraint residual, robustness proxy, calibration error, compute/latency deficit, etc.). Given a trajectory $\tau \mapsto x^{(K)}(\tau)$ (discrete $\tau = n$ or continuous $\tau = t$), we write

$$r^{(K)}(\tau) := \mathcal{R}^{(K)}(x^{(K)}(\tau)) \in \mathbb{R}_{\geq 0}^m.$$

2.2.2. Total Ledgers (Certificate-Compatible Scalarizations)

A *total ledger* is any positive weighting of the ledger vector:

$$R_{\text{tot}}^{(K)}(\tau) := w^\top r^{(K)}(\tau), \quad w \gg 0. \quad (4)$$

There are two complementary interpretations:

- (i) **Design aggregation:** w is chosen by a practitioner to reflect priorities (a scalarization choice) [1,2].
- (ii) **Certified aggregation:** w is produced by the one-clock reduction once a Metzler comparison inequality is established (a copositive Lyapunov weight), so that $R_{\text{tot}}^{(K)}$ decays with a certified exponent when the comparison matrix is Hurwitz [11,12,14].

2.2.3. Reported Readouts and Dictionary Conditioning

Refinement changes representation, so not all measured quantities are automatically comparable across K . We distinguish:

- **Geometric ledgers** defined by restriction of the ambient ruler (automatically consistent under refinement), as in stable discretization theory [4,6].
- **Reported/readout metrics** depending on a K -dependent apparatus (validation sets, simulator fidelity, feature maps). These require an explicit *dictionary* condition later (programme line (O4)) to prevent ill-conditioning from faking improvement.

2.2.4. A Minimal Regularity Interface

To transport convergence in the ruler to convergence of reported readouts, it is often enough to assume local Lipschitz control on the certification window: for each K and each i ,

$$|R_i^{(K)}(x) - R_i^{(K)}(y)| \leq L_i \|x - y\|, \quad (5)$$

with constants L_i independent of K on the relevant bounded set. Uniformity prevents refinement from silently changing the scale of what is being measured [8,10].

2.3. Cross-Level Maps and the “Same Task” Requirement

2.3.1. Projective Structure

Assume there exist coarse-graining/projection maps

$$\Pi_K^{K+1} : X^{(K+1)} \rightarrow X^{(K)},$$

representing restriction, pruning, averaging, or projection of refined states to coarser representations. We require *consistency* (a projective system):

$$\Pi_K^{K+2} = \Pi_K^{K+1} \circ \Pi_{K+1}^{K+2} \quad \text{for all } K \geq K_0. \quad (6)$$

This encodes that all levels represent the *same task object*, just at different resolution; see, e.g., FEM and multilevel treatments [4–6] and extensions beyond nested families [30].

2.4. What Will Be Certified Later

The remainder of the paper will not assume that ledgers decay automatically. Instead, it will require auditable inequalities of the form

$$\dot{r}^{(K)}(\tau) \leq M^{(K)} r^{(K)}(\tau) + d^{(K)}(\tau),$$

with $M^{(K)}$ Metzler and $d^{(K)} \geq 0$ a tail disturbance, and then impose uniformity/summability conditions across K (programme lines (O1)–(O4) in Section 4). Under those obligations, a one-clock total ledger and a refinement-limit learner follow from positive-systems theory and projective Cauchy arguments [11,12,14].

3. Certificate Layer: One-Clock Reduction for Nonnegative Ledgers

3.1. Why Positivity Is the Right Abstraction

3.1.1. Nonnegative Ledgers and Injections

A *ledger* is any measurable quantity $R_i(t) \geq 0$ that a practitioner wants to track (risk, constraint violation, robustness proxy, calibration error, solver residual, etc.). We collect them into a vector

$$r(t) := (R_1(t), \dots, R_m(t))^T \in \mathbb{R}_{\geq 0}^m.$$

Cross-effects appear as *injections*: if improving (or stabilizing) ledger j temporarily worsens ledger i , then \dot{R}_i may contain a nonnegative term proportional to R_j .

3.1.2. Why Positive Comparison Is Canonical

Nonnegativity induces the partial order $u \leq v$ iff $u_i \leq v_i$ for all i on \mathbb{R}^m . Many training-time inequalities have the funded-plus-injection form

$$\dot{R}_i(t) \leq -(\text{funded decay margin}) \cdot R_i(t) + \sum_{j \neq i} (\text{injection}_{i \leftarrow j}) \cdot R_j(t),$$

with all injection coefficients nonnegative. This is exactly the comparison form of a *positive linear system* (Metzler dynamics). It admits: (i) comparison principles, (ii) Perron–Frobenius structure for the dominant mode, and (iii) copositive Lyapunov functions $V(r) = w^\top r$ with $w \gg 0$; see, e.g., [11–14].

3.1.3. How Metzler Bounds Arise in Practice (Derivation Template)

The certificate layer does not assume access to model internals; it assumes that one can bound ledger evolution by auditable inequalities. A common derivation pattern is:

- (i) write a differential (or finite-difference) inequality for each nonnegative ledger R_i ;
- (ii) isolate a *self-decay* term $-2\lambda_i R_i$ (from dissipation, regularization, descent, contractive updates);
- (iii) bound cross-terms by *nonnegative injections* using operator norms or Lipschitz bounds and Young’s inequality, turning mixed products into sums of squares;
- (iv) collect coefficients into a Metzler matrix M with $M_{ii} = -2\lambda_i$ and $M_{ij} = \eta_{i \leftarrow j} \geq 0$.

Such reductions are standard in Lyapunov and comparison analyses; see, e.g., [25] for Lyapunov inequalities, [12,14] for positive-systems comparison, and [11] for the nonnegative-matrix structure behind Metzler semigroups. In discrete time, the same pattern yields a one-step domination

$$r_{n+1}^{(K)} \leq (I + hM) r_n^{(K)} + \zeta_n^{(K)}, \quad \zeta_n^{(K)} \geq 0,$$

and can be propagated by discrete Grönwall-type arguments for positive recursions [31,32].

3.2. Metzler Comparison Systems

3.2.1. Definition (Metzler Dominance)

Definition 1 (Metzler comparison inequality). *Let $r : [0, \infty) \rightarrow \mathbb{R}_{\geq 0}^m$ be absolutely continuous. A Metzler comparison system for r is an inequality*

$$\dot{r}(t) \leq M r(t) \quad \text{for a.e. } t \geq 0, \quad (7)$$

where $M \in \mathbb{R}^{m \times m}$ is Metzler, i.e. $M_{ij} \geq 0$ for all $i \neq j$.

3.2.2. Semantics: Funded Diagonals and Injections

Write

$$M_{ii} = -2\lambda_i, \quad \lambda_i \geq 0, \quad \text{and} \quad M_{ij} = \eta_{i \leftarrow j} \geq 0 \quad (i \neq j).$$

Then (7) reads componentwise as

$$\dot{R}_i(t) \leq -2\lambda_i R_i(t) + \sum_{j \neq i} \eta_{i \leftarrow j} R_j(t).$$

We interpret λ_i as a *funded decay margin* for ledger i and $\eta_{i \leftarrow j}$ as the *injection strength* from ledger j to ledger i .

3.2.3. Positivity of the Semigroup and Comparison

A basic fact is that Metzler matrices generate positive semigroups:

$$M \text{ Metzler} \implies e^{tM} \geq 0 \text{ entrywise for all } t \geq 0,$$

see, e.g., [11,12].

Lemma 1 (Linear comparison principle for Metzler dominance). *Assume (7) and $r(0) \in \mathbb{R}_{\geq 0}^m$. Let y solve the linear ODE $\dot{y} = My$ with $y(0) = r(0)$. Then for all $t \geq 0$,*

$$0 \leq r(t) \leq y(t) = e^{tM}r(0) \quad (\text{componentwise}). \quad (8)$$

Proof. Let $y(t) = e^{tM}r(0)$ solve $\dot{y} = My$ with $y(0) = r(0)$, and set $z(t) := y(t) - r(t)$. Then $z(0) = 0$ and for a.e. $t \geq 0$,

$$\dot{z}(t) = \dot{y}(t) - \dot{r}(t) \geq My(t) - Mr(t) = Mz(t).$$

Define the (componentwise) nonnegative forcing term

$$q(t) := \dot{z}(t) - Mz(t) \geq 0 \quad \text{for a.e. } t \geq 0.$$

Then z satisfies the Duhamel identity

$$z(t) = \int_0^t e^{(t-s)M} q(s) ds.$$

Because M is Metzler, $e^{(t-s)M} \geq 0$ entrywise for all $t \geq s$, and since $q(s) \geq 0$ componentwise, the integrand is componentwise nonnegative. Hence $z(t) \geq 0$ componentwise for all $t \geq 0$, i.e. $r(t) \leq y(t) = e^{tM}r(0)$ componentwise. \square

3.2.4. From Componentwise Bounds to Certified Scalar Ledgers

A comparison inequality of the form (7) controls the *vector* ledger $r(t)$. To obtain a single certified “clock” usable by a checker, we scalarize $r(t)$ by a positive weight. Copositive scalarizations are natural in positive systems: if $w \gg 0$, then $R_{\text{tot}}(t) := w^\top r(t)$ is itself a nonnegative ledger and preserves the cone order. This idea underlies linear copositive Lyapunov functions in positive systems [12,14] and is closely related to Perron–Frobenius theory for nonnegative matrices [11].

3.3. One-Clock Reduction

3.3.1. Spectral Abscissa and Effective Rate

Define the spectral abscissa

$$\mu(M) := \max\{\operatorname{Re} z : z \in \operatorname{spec}(M)\}.$$

For a Metzler matrix, $\mu(M)$ is real and governs the slowest exponential mode; see [11,12]. When $\mu(M) < 0$ we define the effective one-clock rate

$$\lambda_{\text{eff}} := -\mu(M) > 0.$$

3.3.2. Checker form: A Copositive Lyapunov Witness

Theorem 1 (One-clock reduction from a copositive Lyapunov witness). *Assume M is Metzler and (7) holds with $r(0) \in \mathbb{R}_{\geq 0}^m$. If there exist $w \gg 0$ and $\lambda > 0$ such that*

$$w^\top M \leq -\lambda w^\top \quad (\text{componentwise}), \quad (9)$$

then the scalar total ledger

$$R_{\text{tot}}(t) := w^\top r(t) \quad (10)$$

obeys

$$R_{\text{tot}}(t) \leq e^{-\lambda t} R_{\text{tot}}(0) \quad \forall t \geq 0. \quad (11)$$

In particular, M is Hurwitz and $\mu(M) \leq -\lambda$.

Proof. Since $r(t) \geq 0$ componentwise and $w \gg 0$, the scalar ledger $R_{\text{tot}}(t) = w^\top r(t)$ is nonnegative. Multiply (7) by w^\top and use (9): for a.e. t ,

$$\dot{R}_{\text{tot}}(t) = w^\top \dot{r}(t) \leq w^\top M r(t) \leq -\lambda w^\top r(t) = -\lambda R_{\text{tot}}(t).$$

Grönwall's inequality yields (11); see, e.g., [25,26]. The implication $\mu(M) \leq -\lambda$ (hence Hurwitz) from the existence of a strict copositive linear Lyapunov inequality is standard for Metzler matrices; see [12,14]. \square

3.3.3. From Hurwitzness to a Witness, and the Sharp Rate

The checker from Theorem 1 is the *verification* primitive: a certificate can literally ship (w, λ) and the checker verifies (9). The next result explains why such witnesses exist for Hurwitz Metzler matrices and how the sharp rate $\lambda_{\text{eff}} = -\mu(M)$ appears.

Proposition 1 (Hurwitz Metzler \Rightarrow existence of a strictly positive witness). *Let M be Metzler and assume $\mu(M) < 0$. Then there exists $w \gg 0$ such that*

$$w^\top M \leq \mu(M) w^\top \quad (\text{componentwise}). \quad (12)$$

Consequently, with $\lambda_{\text{eff}} := -\mu(M) > 0$, the same w satisfies $w^\top M \leq -\lambda_{\text{eff}} w^\top$ and is a witness in the sense of Theorem 1.

Proof. Choose $\beta > 0$ large enough that $A := M + \beta I$ is entrywise nonnegative (possible since M has nonnegative off-diagonals; take $\beta \geq -\min_i M_{ii}$). Then $A \geq 0$ and

$$\mu(M) = \mu(A - \beta I) = \rho(A) - \beta,$$

where $\rho(A)$ is the spectral radius of A (for nonnegative A , $\rho(A)$ is an eigenvalue by Perron–Frobenius; see [11]). Since $\mu(M) < 0$, we have $\rho(A) < \beta$.

For nonnegative A , the Collatz–Wielandt/Perron–Frobenius theory guarantees existence of a strictly positive left sub-eigenvector: there exists $w \gg 0$ with

$$w^\top A \leq \rho(A) w^\top \quad (\text{componentwise}),$$

see [11, Ch. 1] or [12]. Subtracting βw^\top gives

$$w^\top M = w^\top (A - \beta I) \leq (\rho(A) - \beta) w^\top = \mu(M) w^\top,$$

which is (12). Since $\mu(M) = -\lambda_{\text{eff}}$, we obtain $w^\top M \leq -\lambda_{\text{eff}} w^\top$. \square

Corollary 1 (Sharp one-clock envelope from a Perron–Frobenius witness). *Assume the hypotheses of Proposition 1 and let $w \gg 0$ satisfy (12). Then any r satisfying (7) obeys*

$$w^\top r(t) \leq e^{\mu(M)t} w^\top r(0) = e^{-\lambda_{\text{eff}} t} w^\top r(0), \quad \lambda_{\text{eff}} = -\mu(M) > 0. \quad (13)$$

Proof. Repeat the proof of Theorem 1 with (12) in place of (9):

$$\frac{d}{dt} (w^\top r(t)) = w^\top \dot{r}(t) \leq w^\top M r(t) \leq \mu(M) w^\top r(t),$$

and apply Grönwall [26]. \square

3.3.4. Two-Ledger Closed Form (Explicit Small-Gain Boundary)

The special case $m = 2$ admits explicit eigenvalues and hence an exact small-gain criterion and closed-form effective rate. This provides a particularly transparent diagnostic boundary and is the main “design lever” used in the toy instantiation.

Proposition 2 (Two-ledger small-gain criterion and explicit rate). *Assume $r(t) = (R_1(t), R_2(t))^T \in \mathbb{R}_{\geq 0}^2$ satisfies $\dot{r}(t) \leq Mr(t)$ with*

$$M = \begin{pmatrix} -2\lambda_1 & \eta_{1\leftarrow 2} \\ \eta_{2\leftarrow 1} & -2\lambda_2 \end{pmatrix}, \quad \lambda_1, \lambda_2 > 0, \quad \eta_{1\leftarrow 2}, \eta_{2\leftarrow 1} \geq 0.$$

Then M is Hurwitz if and only if

$$\eta_{1\leftarrow 2} \eta_{2\leftarrow 1} < 4\lambda_1 \lambda_2,$$

and in that case $\lambda_{\text{eff}} := -\mu(M) > 0$ with the explicit formula

$$\lambda_{\text{eff}} = (\lambda_1 + \lambda_2) - \sqrt{(\lambda_1 - \lambda_2)^2 + \eta_{1\leftarrow 2} \eta_{2\leftarrow 1}}.$$

Proof. For a real 2×2 matrix, Hurwitz stability is equivalent to $\text{tr}(M) < 0$ and $\det(M) > 0$. Here $\text{tr}(M) = -2(\lambda_1 + \lambda_2) < 0$ and $\det(M) = 4\lambda_1 \lambda_2 - \eta_{1\leftarrow 2} \eta_{2\leftarrow 1}$, giving the criterion. The eigenvalues are explicit and yield $\lambda_{\text{eff}} = -\mu(M)$ as stated. \square

3.3.5. Monotone Engineering Levers: Funded Diagonals vs. Injections

A distinctive advantage of Metzler comparison certificates is that they translate stability into *monotone* design parameters. Writing the comparison matrix as

$$M_{ii} = -2\lambda_i, \quad \lambda_i \geq 0, \quad \text{and} \quad M_{ij} = \eta_{i\leftarrow j} \geq 0 \quad (i \neq j),$$

makes the two control knobs explicit: *funded diagonals* (increase λ_i , making M_{ii} more negative) and *injections* (decrease $\eta_{i\leftarrow j}$, weakening cross-couplings).

Related “gain composition” ideas are classical in interconnected-system robustness [25,33,34]. In the Metzler setting, the ordering is literal: if $\tilde{M} \leq M$ entrywise, then $\mu(\tilde{M}) \leq \mu(M)$ and hence $\tilde{\lambda}_{\text{eff}} \geq \lambda_{\text{eff}}$, with generic strict improvement away from reducible/degenerate cases [11,12].

Corollary 2 (Funding vs. injection interventions (witness-preserving monotonicity)). *Assume (7). Suppose a witness (w, λ) satisfies $w^\top M \leq -\lambda w^\top$ with $w \gg 0$ and $\lambda > 0$. Let \tilde{M} be another Metzler matrix with $\tilde{M} \leq M$ entrywise (i.e. more negative diagonals and/or smaller off-diagonals). Then the same witness proves the same rate:*

$$w^\top \tilde{M} \leq w^\top M \leq -\lambda w^\top,$$

and hence, under $\dot{r} \leq \tilde{M}r$, one has $w^\top r(t) \leq e^{-\lambda t} w^\top r(0)$ for all $t \geq 0$.

Proof. Entrywise monotonicity gives $w^\top \tilde{M} \leq w^\top M$. Apply Theorem 1 to \tilde{M} . \square

3.3.6. A Checker View: What Must Be Provided and What Is Proved

For certificate checking, the key point is that Theorem 1 requires only an explicit witness: a positive weight vector $w \gg 0$ and a scalar $\lambda > 0$ such that $w^\top M \leq -\lambda w^\top$. A checker validates this inequality componentwise and then accepts the implication $w^\top r(t) \leq e^{-\lambda t} w^\top r(0)$, without spectral computations. This is the positive-systems analogue of “proof-carrying” evidence: a short witness certifies a global decay claim [12,14,23].

4. Master Certificate for Learning Under Refinement

We work on a user-declared certification window $[0, T]$ and in a single declared ruler $\|\cdot\|_W$. Training is run on each refinement level K and produces a trajectory $x^{(K)}(\cdot)$ together with a declared nonnegative scalar ledger $R^{(K)}(\cdot)$. The goal is to ensure that refinement does not change the meaning of “small” and that stability survives as $K \rightarrow \infty$.

Two quantitative inputs drive the refinement-limit guarantee.

- **One-ruler Cauchy geometry.** Cross-level comparisons are performed in the same instrument norm via the ambient realization (e.g. by comparing $z^{(K)} = i_K x^{(K)}$ to $P_K z^{(K+1)}$). If the adjacent discrepancies form a summable tail in K , then the tower is Cauchy in the ruler and determines a unique refinement-limit trajectory on $[0, T]$.
- **Tail-robust contraction on $[0, T]$.** Each level admits an inhomogeneous decay inequality for $R^{(K)}$ with a margin bounded below uniformly in K and a pollution budget whose integrated size is summable along the ladder. This yields a common exponential envelope on $[0, T]$, up to an explicit tail floor that vanishes with refinement.

Programme lines (O1)–(O4) record these requirements in auditable form (summable ladder budgets, a uniform margin, and a uniform dictionary linking $R^{(K)}$ to reported readouts). Under these lines, the Master Certificate theorem (Theorem 3) provides: a unique refinement-limit learner on $[0, T]$, a uniform exponent class for the ledger up to vanishing tails, and transport of the same rate to declared readouts.

4.1. Ladder Geometry and Refinement Limits (Analytic Core)

4.1.1. Ambient One-Ruler Structure and Projective Maps

We formalize the measurement contract by requiring that every refinement level is measured by restriction of a single ambient instrument. This is the analogue of fixing an energy norm across discretizations in stable refinement theory [4–6].

Definition 2 (One ruler (ambient instrument restriction)). *Fix a refinement ladder $\{X^{(K)}\}_{K \geq K_0}$. We say the ladder admits one ruler if there exist:*

- an ambient real Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$,
- linear realization maps $i_K : X^{(K)} \rightarrow \mathcal{H}$ for each K ,
- bounded projections $P_K : \mathcal{H} \rightarrow i_K(X^{(K)})$ for each K ,
- a bounded, self-adjoint, strictly positive operator $W : \mathcal{H} \rightarrow \mathcal{H}$ (the instrument),
- coarse-graining maps $\Pi_K^{K+1} : X^{(K+1)} \rightarrow X^{(K)}$ for each K ,

such that for all $K \geq K_0$:

- (i) **Compatibility (coarse-graining is realized by projection).** For all $x \in X^{(K+1)}$,

$$i_K(\Pi_K^{K+1}x) = P_K(i_{K+1}x). \quad (14)$$

- (ii) **Single instrument (restriction of one ambient ruler).** For $x \in X^{(K)}$ define

$$\|x\|_{W^{(K)}}^2 := \langle i_K x, W i_K x \rangle_{\mathcal{H}}. \quad (15)$$

Equivalently, $\|z\|_W^2 := \langle z, Wz \rangle_{\mathcal{H}}$ is the ambient ruler and $\|\cdot\|_{W^{(K)}}$ is its restriction to $i_K(X^{(K)})$.

We call (\mathcal{H}, W) the declared ruler and (i_K, P_K) the ambient realization of the ladder.

Assumption 1 (Instrument contractivity of coarse projections). *For each $K \geq K_0$, the projection P_K is contractive in the instrument norm on the next-level realization:*

$$\|P_K z\|_W \leq \|z\|_W \quad \text{for all } z \in i_{K+1}(X^{(K+1)}). \quad (16)$$

Lemma 2 (Non-expansiveness of coarse-graining in one ruler). *Assume Definition 2 and ?? 1. Then for all $K \geq K_0$ and $x \in X^{(K+1)}$,*

$$\|\Pi_K^{K+1}(x)\|_{W^{(K)}} \leq \|x\|_{W^{(K+1)}}. \quad (17)$$

Proof. Fix K and $x \in X^{(K+1)}$ and set $z := i_{K+1}x$. By (14), $i_K(\Pi_K^{K+1}x) = P_K z$. Hence

$$\|\Pi_K^{K+1}x\|_{W^{(K)}}^2 = \|P_K z\|_W^2 \leq \|z\|_W^2 = \|x\|_{W^{(K+1)}}^2,$$

where we used (16) and (15). Taking square roots gives (17). \square

4.1.2. Cross-Level Discrepancy and Telescoping

Given $x^{(K)} \in X^{(K)}$ and $x^{(L)} \in X^{(L)}$ with $L \geq K$, we compare them in the *same* ruler by projecting the fine object down:

$$d_{K \leftarrow L}(x^{(K)}, x^{(L)}) := \|i_K x^{(K)} - P_K(i_L x^{(L)})\|_W. \quad (18)$$

Lemma 3 (Projective telescoping in one ruler). *Assume Definition 2 and ?? 1. Fix $T > 0$ and let $x^{(K)} : [0, T] \rightarrow X^{(K)}$. Then for any integers $L > K \geq K_0$ and any $t \in [0, T]$,*

$$\|i_K x^{(K)}(t) - P_K(i_L x^{(L)}(t))\|_W \leq \sum_{j=K}^{L-1} \|i_j x^{(j)}(t) - P_j(i_{j+1} x^{(j+1)}(t))\|_W. \quad (19)$$

Proof. Fix $t \in [0, T]$ and write $z^{(j)} := i_j x^{(j)}(t) \in \mathcal{H}$. Using $P_K = P_K P_j$ for $j \geq K$ and contractivity of P_K in $\|\cdot\|_W$ (?? 1), we obtain

$$\|z^{(K)} - P_K z^{(L)}\|_W \leq \sum_{j=K}^{L-1} \|P_K(z^{(j)} - P_j z^{(j+1)})\|_W \leq \sum_{j=K}^{L-1} \|z^{(j)} - P_j z^{(j+1)}\|_W,$$

which is (19). \square

4.1.3. Existence and Uniqueness of a Refinement-Limit Object

Theorem 2 (Existence and uniqueness of a refinement-limit object). *Assume the ladder admits one ruler (Definition 2) and satisfies the projective consistency $\Pi_K^{K+2} = \Pi_K^{K+1} \Pi_{K+1}^{K+2}$. Let $\{x^{(K)}\}_{K \geq K_0}$ be a compatible tower, i.e. $x^{(K)} = \Pi_K^{K+1}(x^{(K+1)})$ for all K . Assume there exists a nonnegative sequence $(a_K)_{K \geq K_0}$ with $\sum_{K \geq K_0} a_K < \infty$ such that for all $K \geq K_0$ and all $t \in [0, T]$,*

$$\|(I - P_K) i_{K+1} x^{(K+1)}(t)\|_W + \|i_K x^{(K)}(t) - P_K i_{K+1} x^{(K+1)}(t)\|_W \leq a_K. \quad (20)$$

Then for every $t \in [0, T]$ the sequence $\{i_K x^{(K)}(t)\}_{K \geq K_0}$ is Cauchy in $(\mathcal{H}, \|\cdot\|_W)$ and converges to a unique limit $x^{(\infty)}(t) \in \mathcal{H}$:

$$x^{(\infty)}(t) := \lim_{K \rightarrow \infty} i_K x^{(K)}(t) \quad \text{in } (\mathcal{H}, \|\cdot\|_W).$$

Moreover, for every fixed $K \geq K_0$,

$$P_K x^{(\infty)}(t) = \lim_{L \rightarrow \infty} P_K i_L x^{(L)}(t) = i_K x^{(K)}(t) \quad \forall t \in [0, T]. \quad (21)$$

Proof. Fix $t \in [0, T]$ and let $L > K \geq K_0$. Write $z^{(j)} := i_j x^{(j)}(t) \in \mathcal{H}$. Then

$$\|z^{(L)} - z^{(K)}\|_W \leq \|z^{(L)} - P_K z^{(L)}\|_W + \|P_K z^{(L)} - z^{(K)}\|_W.$$

By telescoping and (20),

$$\|z^{(L)} - P_K z^{(L)}\|_W + \|P_K z^{(L)} - z^{(K)}\|_W \leq \sum_{j=K}^{L-1} a_j.$$

Since $\sum_{j \geq K_0} a_j < \infty$, the right-hand side tends to 0 as $K, L \rightarrow \infty$, so $\{i_K x^{(K)}(t)\}_K$ is Cauchy and converges in $(\mathcal{H}, \|\cdot\|_W)$. Projective consistency (21) follows by applying P_K to the tower identities and passing to the limit. \square

Remark 1 (Where this theorem is used). *The Master Certificate theorem (Theorem 3) uses Theorem 2 as the geometric engine that turns summable cross-level discrepancies into a refinement-limit object in the declared ruler. Concretely, in the Master Certificate proof one takes $a_K := \beta_K + \delta_K$ from (O3).*

4.2. Programme Lines (O1)–(O4) on $[0, T]$

The programme lines (O1)–(O4) are the paper's *checkable obligations* on a finite time window. They specify what a certificate must provide and what a verifier must validate: (i) a uniform-in-time contraction envelope for a declared total ledger up to a summable pollution budget, (ii) a K -uniform margin (one exponent class), (iii) a summable Cauchy tower in the declared ruler (no moving goalposts), and (iv) a uniform dictionary linking the declared ledger to reported readouts.

Certified time horizon.

Fix $T > 0$. All programme lines (O1)–(O4) are required on the window $[0, T]$.

4.2.1. Objects on the Window: Trajectories, Ruler, and a Declared Total Ledger

For each refinement level $K \geq K_0$, let $x^{(K)} : [0, T] \rightarrow X^{(K)}$ be a training trajectory. Assume the ladder admits one ruler in the sense of Definition 2, with embeddings $i_K : X^{(K)} \rightarrow \mathcal{H}$, projections $P_K : \mathcal{H} \rightarrow i_K(X^{(K)})$, and an instrument W inducing $\|z\|_W^2 = \langle z, Wz \rangle$. Write the realized trajectory in the ambient ruler as

$$z^{(K)}(t) := i_K x^{(K)}(t) \in \mathcal{H}.$$

Fix a nonnegative *total geometric ledger* $R^{(K)} : [0, T] \rightarrow \mathbb{R}_{\geq 0}$ (one scalar per level). This is the scalar object whose decay is certified; any additional structure is used only through inequalities.

(O1) Tail-robust contraction envelope (summable pollution budget)

There exist nonnegative functions $\tau_K \in L^1([0, T])$, numbers $\alpha_K \geq 0$ with $\sum_{K=K_0}^{\infty} \alpha_K < \infty$, and levelwise margins $\lambda_K > 0$ such that for a.e. $t \in [0, T]$,

$$\dot{R}^{(K)}(t) \leq -2\lambda_K R^{(K)}(t) + \tau_K(t), \quad \int_0^T \tau_K(s) ds \leq \alpha_K, \quad K \geq K_0. \quad (22)$$

(O2) Uniform margin (one exponent class across refinement)

There exists $\lambda > 0$ (independent of K) such that

$$\lambda_K \geq \lambda > 0 \quad \text{for all } K \geq K_0. \quad (23)$$

(O3) Projective Cauchy tower in one ruler (summable cross-level inconsistency)

There exist nonnegative sequences $(\beta_K)_{K \geq K_0}$ and $(\delta_K)_{K \geq K_0}$ such that

$$\sup_{t \in [0, T]} \left(\underbrace{\|(I - P_K) z^{(K+1)}(t)\|_W}_{\text{unresolved tail at level } K} + \underbrace{\|z^{(K)}(t) - P_K z^{(K+1)}(t)\|_W}_{\text{projected mismatch}} \right) \leq \beta_K + \delta_K, \quad \sum_{K=K_0}^{\infty} (\beta_K + \delta_K) < \infty. \quad (24)$$

(O4) Uniform dictionary / observability (reported readouts remain well-conditioned)

Let \mathcal{M} be a finite family of nonnegative reported readouts/metrics $M(\cdot) \geq 0$. Assume there exist constants $c_M, C_M > 0$, independent of K , such that for all $t \in [0, T]$ and all $K \geq K_0$,

$$c_M R^{(K)}(t) \leq M(x^{(K)}(t)) \leq C_M R^{(K)}(t), \quad \forall M \in \mathcal{M}. \quad (25)$$

4.2.2. Checker Summary (What the Certificate Must Provide)

A certificate on $[0, T]$ consists of: the declared ruler (\mathcal{H}, W) and projections (i_K, P_K) , the scalar ledger $R^{(K)}$, the budgets (α_K) from (O1) and (β_K, δ_K) from (O3), the uniform margin λ from (O2), and the dictionary constants (c_M, C_M) from (O4). Soundness (proved next) upgrades these inequalities to a unique refinement-limit learner and rate inheritance on $[0, T]$.

4.3. Master Certificate Theorem on $[0, T]$

We now state the soundness implication in its final form. Importantly, the theorem below does not restate the programme lines; it only invokes them by reference.

Theorem 3 (Master Certificate for Learning Under Refinement on $[0, T]$). *Assume the ladder admits one ruler (Definition 2) and let $x^{(K)} : [0, T] \rightarrow X^{(K)}$ be trajectories with ambient realizations $z^{(K)}(t) := i_K x^{(K)}(t) \in \mathcal{H}$. Let $R^{(K)} : [0, T] \rightarrow \mathbb{R}_{\geq 0}$ be a declared total geometric ledger. Assume the programme lines (O1)–(O4) from Section 4.2 hold on $[0, T]$. Then:*

(i) **Tail-robust levelwise envelope.** *With λ as in (O2) and α_K as in (O1), for all $K \geq K_0$ and all $t \in [0, T]$,*

$$R^{(K)}(t) \leq e^{-2\lambda t} R^{(K)}(0) + \alpha_K. \quad (26)$$

(ii) **Existence and uniqueness of a refinement-limit trajectory.** *Define the geometric tower budget*

$$a_K := \beta_K + \delta_K,$$

where (β_K, δ_K) are the sequences from (O3). Then $\sum_{K \geq K_0} a_K < \infty$, and there exists a unique trajectory $x^{(\infty)} : [0, T] \rightarrow \mathcal{H}$ such that $x^{(\infty)}(t) = \lim_{K \rightarrow \infty} z^{(K)}(t)$ in $(\mathcal{H}, \|\cdot\|_W)$, and

$$\sup_{t \in [0, T]} \|z^{(K)}(t) - P_K x^{(\infty)}(t)\|_W \leq \sum_{j=K}^{\infty} a_j \quad \forall K \geq K_0. \quad (27)$$

(iii) **Readout transport (rate inheritance at each level).** *For every $M \in \mathcal{M}$ (from (O4)), for all $K \geq K_0$ and all $t \in [0, T]$,*

$$M(x^{(K)}(t)) \leq C_M \left(e^{-2\lambda t} R^{(K)}(0) + \alpha_K \right). \quad (28)$$

Proof. Part (i) is the inhomogeneous Grönwall bound from (O1)–(O2). Part (ii) follows by applying Theorem 2 with $a_K := \beta_K + \delta_K$. Part (iii) is (O4) combined with (26). \square

5. Instantiations

This section explains how the abstract certificate obligations can be realized in concrete training pipelines. The goal is not to advocate a particular architecture, but to show (i) how to *choose* a ruler, a ladder, and a ledger vector; (ii) how to *derive* a Metzler comparison system with explicit margins and injections; and (iii) how to *audit* the refinement programme lines (O1)–(O4) from logs, bounds, or subproofs. We present two representative instantiations: a fully checkable toy ladder where all constants can be computed in closed form, and a practical “width/refinement” protocol that describes what evidence a producer must ship so that a consumer can validate the Master Certificate without re-running training.

At a high level, every instantiation follows the same recipe. First fix a single ambient ruler (a norm/energy that is independent of refinement) and specify the projections that implement “no moving goalposts.” Next define the nonnegative ledgers and establish an in-level comparison inequality $\dot{r}^{(K)} \leq M^{(K)}r^{(K)} + d^{(K)}$ with $M^{(K)}$ Metzler, extracting diagonal funding margins and off-diagonal injections via standard bounds (e.g. Lipschitz/operator-norm and Young inequalities). Finally, verify the cross-level conditions: tail terms are summable ((O1)), the exponent class is uniform ((O2)), projected states are Cauchy across levels ((O3)), and reported readouts remain uniformly comparable to the ruler ((O4)). The output is a proof-carrying artifact consisting of the matrices, constants, and tail budgets that a checker can validate, thereby upgrading verification to a refinement-limit guarantee with a certified one-clock rate.

5.1. Setup: an ℓ^2 Ladder with Diagonal Constraint Operator

Let ℓ^2 denote the real Hilbert space of square-summable sequences with inner product $\langle x, y \rangle = \sum_{i \geq 1} x_i y_i$ and norm $\|x\|_2^2 = \sum_{i \geq 1} x_i^2$. Define the ambient state space

$$X^{(\infty)} := \ell^2 \times \ell^2, \quad z = (x, \lambda) \in X^{(\infty)}.$$

Fix $p > 0$ and define the bounded diagonal operator $A : \ell^2 \rightarrow \ell^2$ by

$$(Ax)_i := a_i x_i, \quad a_i := i^{-p}. \quad (29)$$

Then $\|A\|_{\text{op}} = \sup_i |a_i| = 1$.

For each $K \geq 1$ define the refinement spaces

$$X^{(K)} := \mathbb{R}^K \times \mathbb{R}^K \subset \ell^2 \times \ell^2$$

embedded into $X^{(\infty)}$ by zero padding. Let P_K be orthogonal projection (truncate to first K coordinates), and set $A_K := P_K A P_K = \text{diag}(a_1, \dots, a_K)$.

5.2. Dynamics: Damped Primal–Dual Flow

Fix damping parameters $m > 0$ and $\alpha > 0$. For each K , consider the linear flow on $X^{(K)}$:

$$\begin{aligned} \dot{x}^{(K)}(t) &= -m x^{(K)}(t) - A_K^\top \lambda^{(K)}(t), \\ \dot{\lambda}^{(K)}(t) &= A_K x^{(K)}(t) - \alpha \lambda^{(K)}(t), \end{aligned} \quad t \geq 0. \quad (30)$$

5.3. Ledgers and Readouts

Define the two nonnegative ledgers and the ledger vector

$$R_1^{(K)}(t) := \|x^{(K)}(t)\|_2^2, \quad R_2^{(K)}(t) := \|\lambda^{(K)}(t)\|_2^2, \quad r^{(K)}(t) := (R_1^{(K)}(t), R_2^{(K)}(t))^\top. \quad (31)$$

As readout metrics, take $\mathcal{M} = \{M_1, M_2\}$ with $M_1(z) = \|x\|^2$ and $M_2(z) = \|\lambda\|^2$.

5.4. Metzler Comparison and a Checker-Friendly Hurwitz Witness

Lemma 4 (Metzler dominance (uniform in K)). *Assume $p > 0$ so that $\|A\|_{\text{op}} = 1$. For every K and all $t \geq 0$, the ledger vector satisfies*

$$\dot{r}^{(K)}(t) \leq M_{\text{Toy}} r^{(K)}(t),$$

where the same 2×2 Metzler matrix works for all K :

$$M_{\text{Toy}} := \begin{pmatrix} -(2m-1) & 1 \\ 1 & -(2\alpha-1) \end{pmatrix}. \quad (32)$$

Proof. Differentiate the squared norms using (30):

$$\dot{R}_1^{(K)}(t) = 2\langle x^{(K)}(t), \dot{x}^{(K)}(t) \rangle = -2m \|x^{(K)}(t)\|_2^2 - 2\langle x^{(K)}(t), A_K^\top \lambda^{(K)}(t) \rangle.$$

Using $\|A_K\|_{\text{op}} \leq \|A\|_{\text{op}} = 1$ and Young's inequality $2ab \leq a^2 + b^2$ gives

$$-2\langle x^{(K)}(t), A_K^\top \lambda^{(K)}(t) \rangle \leq 2 \|x^{(K)}(t)\|_2 \|A_K\|_{\text{op}} \|\lambda^{(K)}(t)\|_2 \leq 2 \|x^{(K)}(t)\|_2 \|\lambda^{(K)}(t)\|_2 \leq \|x^{(K)}(t)\|_2^2 + \|\lambda^{(K)}(t)\|_2^2.$$

Hence $\dot{R}_1^{(K)}(t) \leq -(2m - 1)R_1^{(K)}(t) + R_2^{(K)}(t)$.

Similarly,

$$\dot{R}_2^{(K)}(t) = 2\langle \lambda^{(K)}(t), \dot{\lambda}^{(K)}(t) \rangle = -2\alpha \|\lambda^{(K)}(t)\|_2^2 + 2\langle \lambda^{(K)}(t), A_K x^{(K)}(t) \rangle \leq -(2\alpha - 1)R_2^{(K)}(t) + R_1^{(K)}(t),$$

where we again used $\|A_K\|_{\text{op}} \leq 1$ and $2ab \leq a^2 + b^2$ on the cross term. Collecting the two inequalities yields $\dot{r}^{(K)}(t) \leq M_{\text{Toy}} r^{(K)}(t)$. \square

Proposition 3 (Hurwitz witness and certified one-clock rate). *Assume $m > 1$ and $\alpha > 1$. Let $w = (1, 1)^\top \gg 0$ and set*

$$\lambda_{\text{cert}} := 2 \min\{m - 1, \alpha - 1\} > 0. \quad (33)$$

Then $w^\top M_{\text{Toy}} \leq -\lambda_{\text{cert}} w^\top$ componentwise, and hence

$$R_{\text{tot}}^{(K)}(t) := w^\top r^{(K)}(t) = R_1^{(K)}(t) + R_2^{(K)}(t) \leq e^{-\lambda_{\text{cert}} t} R_{\text{tot}}^{(K)}(0) \quad \forall t \geq 0, \forall K.$$

Proof. Compute $w^\top M_{\text{Toy}} = (-2(m - 1), -2(\alpha - 1)) \leq -2 \min\{m - 1, \alpha - 1\} (1, 1)$. Apply Theorem 1 to Lemma 4.

\square

5.5. Programme Lines (O1)–(O4) (Vanishing Tails)

Lemma 5 (Exact projective consistency). *Let $L \geq K$. If $z^{(L)}(t)$ solves (30) at level L , then $P_K z^{(L)}(t)$ solves the level- K system with initial data $P_K z^{(L)}(0)$. In particular, for consistent initial data $z^{(K)}(0) = P_K z^{(K+1)}(0)$ we have $z^{(K)}(t) = P_K z^{(K+1)}(t)$ for all $t \geq 0$.*

Proof. For diagonal A , the first K coordinates of (30) depend only on the first K coordinates, and A_{K+1} restricted to the first K indices equals A_K . Uniqueness of ODE solutions yields the identity. \square

Proposition 4 (Toy ladder satisfies (O1)–(O4)). *Assume $m > 1$ and $\alpha > 1$. Then the toy ladder satisfies the Master Certificate programme lines on $[0, \infty)$:*

- (i) (O1) holds with $\tau_K \equiv 0$ and $\alpha_K \equiv 0$;
- (ii) (O2) holds with uniform margin $\lambda = \lambda_{\text{cert}}$ from (33);
- (iii) (O3) holds with $\beta_K \equiv 0$ and $\delta_K \equiv 0$ by Lemma 5;
- (iv) (O4) holds with $c_{M_1} = C_{M_1} = c_{M_2} = C_{M_2} = 1$ for $M_1(z) = \|x\|^2$ and $M_2(z) = \|\lambda\|^2$.

Proof. Items (i), (ii) follow from Lemma 4 and Proposition 3. Item (iii) is Lemma 5. Item (iv) holds because the readouts equal the ledgers. \square

5.6. Consequence: Refinement-Limit Existence and Inherited Clock

Corollary 3 (Toy model: unique refinement-limit trajectory and uniform rate inheritance). *Assume $m > 1$ and $\alpha > 1$ and consistent initial data $z^{(K)}(0) = P_K z^{(\infty)}(0)$ for some $z^{(\infty)}(0) \in \ell^2 \times \ell^2$. Then there*

K	$R_{\text{tot}}^{(K)}(0)$	$R_{\text{tot}}^{(K)}(T)$	$\max_{t \in [0, T]} \text{Rat}^{(K)}(t)$	$\text{Rat}^{(K)}(T)$
10	2.125857×10^{-1}	1.091714×10^{-10}	1.000000	1.131149×10^{-5}
20	2.370105×10^{-1}	1.247584×10^{-10}	1.000000	1.159436×10^{-5}
40	2.459643×10^{-1}	1.351165×10^{-10}	1.000000	1.209988×10^{-5}
80	2.486140×10^{-1}	1.378366×10^{-10}	1.000000	1.221192×10^{-5}

Table 1. Numerical sanity check of the toy certificate $R_{\text{tot}}^{(K)}(t) \leq e^{-\lambda_{\text{cert}} t} R_{\text{tot}}^{(K)}(0)$ with $\lambda_{\text{cert}} = 2$ for $m = 2.0$, $\alpha = 2.5$, $p = 1.0$, $T = 5.0$, $\Delta t = 10^{-3}$.

exists a unique refinement-limit trajectory $z^{(\infty)}(t) \in \ell^2 \times \ell^2$ with $z^{(K)}(t) = P_K z^{(\infty)}(t)$ for all K and $t \geq 0$, and the total ledger satisfies

$$R_{\text{tot}}^{(K)}(t) \leq e^{-\lambda_{\text{cert}} t} R_{\text{tot}}^{(K)}(0), \quad \lambda_{\text{cert}} = 2 \min\{m - 1, \alpha - 1\}.$$

Proof. Apply Theorem 2 using Proposition 4. \square

5.7. Numerical Sanity Check (Uniform in K)

Fix $m = 2.0$, $\alpha = 2.5$, $p = 1.0$, $T = 5.0$, and a forward Euler time step $\Delta t = 10^{-3}$, so that $\lambda_{\text{cert}} = 2$. Define the certificate ratio

$$\text{Rat}^{(K)}(t) := \frac{R_{\text{tot}}^{(K)}(t)}{e^{-\lambda_{\text{cert}} t} R_{\text{tot}}^{(K)}(0)}.$$

Table 1 reports $R_{\text{tot}}^{(K)}(0)$, $R_{\text{tot}}^{(K)}(T)$, $\max_{t \in [0, T]} \text{Rat}^{(K)}(t)$, and $\text{Rat}^{(K)}(T)$ for $K \in \{10, 20, 40, 80\}$.

Remark 2. The initial values increase mildly with K because $z^{(K)}(0)$ is obtained by truncating a fixed ambient ℓ^2 initial condition. The maximum ratio equals 1 (attained at $t = 0$), while $\text{Rat}^{(K)}(T) \sim 10^{-5}$, indicating the bound is conservative on this run.

5.8. Neural Instantiation (Protocol): A Width Ladder with Auditable Programme Lines

5.8.1. Setup: Width Ladder and a Declared Projection

Let K index width (or another scalable capacity parameter). For each K , let $X^{(K)}$ encode model parameters together with any augmented optimizer state (momenta, running averages, constraint multipliers). Fix an explicit coarse-graining map $\Pi_K^{K+1} : X^{(K+1)} \rightarrow X^{(K)}$.

5.8.2. Practical Ledgers (Loggable During Training)

Choose a finite nonnegative ledger vector $r^{(K)}(t) \in \mathbb{R}_{\geq 0}^m$ that is loggable during training, e.g.

$$r^{(K)}(t) = (R_{\mathcal{L}}^{(K)}(t), R_{\text{safe}}^{(K)}(t), R_{\text{rob}}^{(K)}(t), R_{\text{comp}}^{(K)}(t))^{\top}.$$

5.8.3. Estimating Metzler Coefficients from Traces

In discrete time, a checker-friendly target is the inequality

$$r_{n+1}^{(K)} \leq (I + hM^{(K)})r_n^{(K)} + \zeta_n^{(K)}, \quad \zeta_n^{(K)} \geq 0,$$

with $M^{(K)}$ Metzler. A Hurwitz witness can be supplied by checking $w^{\top} M^{(K)} \leq -\lambda w^{\top}$ (Proposition 1) or, in the two-ledger case, the small-gain product inequality (Proposition 2).

5.8.4. Empirical Checks for (O1)–(O4)

- **(O2) uniform margin:** fit envelope rates for $R_{\text{tot}}^{(K)}$ and test uniformity in K .
- **(O3) projective Cauchy:** measure $\|x^{(K)} - \Pi_K^{K+1}(x^{(K+1)})\|_{W^{(K)}}$ across checkpoints.
- **(O1) tail summability:** quantify unmodeled remainder budgets α_K and test $\sum_K \alpha_K < \infty$.

- **(O4) dictionary conditioning:** bound conditioning constants c_M, C_M uniformly in K .

5.8.5. Interpretation

The protocol certifies training-time contraction and refinement stability in a declared ruler when the programme lines hold. Generalization, domain shift, and stochastic optimization effects are separate modules unless explicitly ledgerized.

6. Concluding Discussion and Outlook

The paper's contribution is best read as a single verifier-facing implication. A producer supplies a finite certificate for a declared ladder and a declared ruler: (i) in-level one-clock evidence for the nonnegative ledger vector via a Metzler comparison system and a copositive witness, and (ii) cross-level programme-line budgets on the window $[0, T]$ ensuring summable refinement tails, a K -uniform decay margin, summable projective inconsistencies, and a uniform dictionary linking the declared geometric ledger to reported readouts. Soundness means the checker does not need to inspect the optimizer internals or re-run training: validating the ruler/transfer maps, the witness inequalities, and the ladder budgets is sufficient to conclude that refinement does not introduce drift, that the ladder admits a unique refinement-limit learner on $[0, T]$, and that the same exponent class is inherited up to the explicit vanishing tails. In particular, any readout covered by the dictionary line inherits the same rate class, so the certified clock is not confined to an internal ledger but transports to declared external metrics.

Where certificates fail is correspondingly concentrated. In applications one typically sees (a) the uniform margin degrade with refinement, (b) dictionary constants blow up (conditioning changes the meaning of the reported metric), or (c) the cross-level tower budgets cease to be summable. Each failure has a direct diagnostic: fit levelwise envelopes and monitor the inferred rate versus K ; track conditioning surrogates that upper bound C_M/c_M ; and measure projected mismatches $\sup_{t \in [0, T]} \|z^{(K)}(t) - P_K z^{(K+1)}(t)\|_W$ along checkpoints. When the bottleneck is the comparison matrix, the design lever is monotone: fund diagonals (increase self-decay margins) and/or reduce injections (weaken couplings), with an exact boundary in the two-ledger small-gain case.

The claim boundary is deliberate. The certificate is a stability-and-limit guarantee in a declared ruler on a declared window; it does not by itself imply statistical generalization or out-of-distribution robustness, and it does not control stochastic-gradient noise unless such effects are ledgerized or bounded by a separate probabilistic argument. Those properties remain composable add-ons: one may import generalization modules (e.g. stability or PAC-Bayes) or robustness modules (e.g. DRO/shift models) and then connect them to the certified ledgers through additional dictionary links.

Several extensions preserve the same checker philosophy. One can formulate stochastic certificates where the programme lines hold with high probability and the clock becomes probabilistic; allow switched or time-varying Metzler envelopes with common copositive weights; permit slowly varying rulers $W(t)$ under a separate "ruler-drift" budget; and replace scalar dictionary bounds by structured operator inequalities for richer readout transport. Finally, a useful empirical direction is a notion of certificate coverage: how much of an observed training trace is explained by the declared comparison model versus assigned to tail pollution, tracked as a function of refinement.

Notation

Symbol	Type / Domain	Meaning / Assumptions
Ambient ruler and refinement ladder		
\mathcal{H}	Hilbert space	Ambient realization space used to enforce a single ruler (Definition 2)
W	operator on \mathcal{H}	Bounded, self-adjoint, strictly positive instrument operator defining $\ z\ _W^2 = \langle z, Wz \rangle_{\mathcal{H}}$

Symbol	Type / Domain	Meaning / Assumptions
$\langle \cdot, \cdot \rangle$	inner product	Ambient Hilbert inner product on \mathcal{H}
$\ z\ _W$	norm	Instrument norm: $\ z\ _W := \sqrt{\langle z, Wz \rangle}$ on \mathcal{H}
$X^{(K)}$	set / space	State/parameter space at refinement level K (width, resolution, basis size, etc.)
K_0	integer	Minimal refinement index considered; ladder runs over $K \geq K_0$
i_K	map $X^{(K)} \rightarrow \mathcal{H}$	Realization/embedding of level- K states into the ambient ruler space
P_K	projection on \mathcal{H}	Orthogonal projection onto $i_K(X^{(K)}) \subset \mathcal{H}$
Π_K^{K+1}	map $X^{(K+1)} \rightarrow X^{(K)}$	Coarse-graining / restriction map between adjacent refinement levels
Π_K^L	map $X^{(L)} \rightarrow X^{(K)}$	Multi-step projection $\Pi_K^L := \Pi_K^{K+1} \circ \dots \circ \Pi_{L-1}^L$ (when well-defined)

Training dynamics (discrete and continuous)

$x^{(K)}(t)$	curve in $X^{(K)}$	Continuous-time training trajectory at refinement level K
$x_n^{(K)}$	sequence in $X^{(K)}$	Discrete-time iterates at refinement level K
$\Phi^{(K)}$	map on $X^{(K)}$	Discrete update map: $x_{n+1}^{(K)} = \Phi^{(K)}(x_n^{(K)})$
$\mathcal{V}^{(K)}$	vector field	Continuous-time flow: $\dot{x}^{(K)}(t) = \mathcal{V}^{(K)}(x^{(K)}(t))$
t	$\mathbb{R}_{\geq 0}$	Continuous time (or rescaled iteration-time)
n	\mathbb{N}	Discrete iteration index

Ledgers and scalarizations

$R_i(t)$	$\mathbb{R}_{\geq 0}$	Nonnegative ledger (risk, constraint violation, robustness proxy, etc.)
$r(t)$	$\mathbb{R}_{\geq 0}^m$	Ledger vector $r(t) = (R_1(t), \dots, R_m(t))^\top$
$r^{(K)}(t)$	$\mathbb{R}_{\geq 0}^m$	Level- K ledger vector along training at refinement K
$R_{\text{tot}}(t)$	$\mathbb{R}_{\geq 0}$	Declared total ledger (scalarization) used for the certificate
w	$\mathbb{R}_{> 0}^m$	Positive weight vector defining $R_{\text{tot}}(t) = w^\top r(t)$
W_{led}	SPD matrix	Optional quadratic ledger ruler: $R_{\text{tot}}(t) = \ r(t)\ _{W_{\text{led}}}^2 = r(t)^\top W_{\text{led}} r(t)$
\mathcal{M}	finite set	Family of reported readout metrics $M(\cdot) \geq 0$
c_M, C_M	scalars	Dictionary/observability constants in $c_M R^{(K)} \leq M(x^{(K)}) \leq C_M R^{(K)}$ (line (O4))

Metzler comparison system and one-clock quantities

M	matrix	Metzler comparison matrix: $M_{ij} \geq 0$ for $i \neq j$ (Definition 1)
$M^{(K)}$	matrix	Level- K comparison matrix in $\dot{r}^{(K)} \leq M^{(K)} r^{(K)}$
λ_i	scalar	Funded self-decay margin for ledger i when $M_{ii} = -2\lambda_i$
$\eta_{i \leftarrow j}$	scalar	Injection strength from ledger j to ledger i when $M_{ij} = \eta_{i \leftarrow j} \geq 0$
e^{tM}	matrix	Matrix exponential (positive for Metzler M)
$\text{spec}(M)$	set	Spectrum (eigenvalues) of M
$\mu(M)$	scalar	Spectral abscissa: $\mu(M) := \max\{\text{Re } z : z \in \text{spec}(M)\}$
Hurwitz	property	M Hurwitz $\Leftrightarrow \mu(M) < 0$

Symbol	Type / Domain	Meaning / Assumptions
λ_{eff}	scalar	Effective one-clock rate: $\lambda_{\text{eff}} := -\mu(M) > 0$ (when M is Hurwitz)
$\gamma_{1\leftarrow 2}$	scalar	Gain $\gamma_{1\leftarrow 2} := \eta_{1\leftarrow 2} / (2\lambda_1)$ in the 2×2 form
Refinement programme-line budgets (Master Certificate)		
T	$\mathbb{R}_{>0}$	Declared certification horizon: programme lines (O1)–(O4) are required on $[0, T]$
$\tau_K(t)$	$\mathbb{R}_{\geq 0}$	Tail disturbance in (O1) on $[0, T]$: $\dot{R}^{(K)} \leq -2\lambda_K R^{(K)} + \tau_K$
α_K	$\mathbb{R}_{\geq 0}$	Integrated pollution budget in (O1): $\int_0^T \tau_K(s) ds \leq \alpha_K$ with $\sum_K \alpha_K < \infty$
β_K	$\mathbb{R}_{\geq 0}$	Unresolved-tail budget in (O3): $\sup_{t \in [0, T]} \ (I - P_K)z^{(K+1)}(t)\ _W \leq \beta_K$ with $\sum_K \beta_K < \infty$
δ_K	$\mathbb{R}_{\geq 0}$	Projective mismatch budget in (O3): $\sup_{t \in [0, T]} \ z^{(K)}(t) - P_K z^{(K+1)}(t)\ _W \leq \delta_K$ with $\sum_K \delta_K < \infty$

Definitions

Entry	Definition / Formula	Role in the paper
Ladder geometry (“one ruler”)		
Projective consistency	$\Pi_K^{K+2} = \Pi_K^{K+1} \circ \Pi_{K+1}^{K+2}$	Encodes “same task” across refinement levels
One ruler (ambient instrument restriction)	There exist (\mathcal{H}, W) and realizations i_K, P_K such that $i_K(\Pi_K^{K+1}x) = P_K(i_{K+1}x)$ and $\ x\ _{W^{(K)}}^2 = \langle i_K x, W i_K x \rangle_{\mathcal{H}}$	Forbids moving goalposts (Definition 2)
Instrument norm and ruler square root	$\ z\ _W := \sqrt{\langle z, Wz \rangle_{\mathcal{H}}} = \ W^{1/2}z\ _{\mathcal{H}}$	Fixes the single measurement convention across all levels
Instrument contractivity (projection stability)	$\ P_K z\ _W \leq \ z\ _W$ on $i_{K+1}(X^{(K+1)})$	Ensures coarse projections do not inflate the ruler norm (Assumption 1)
Non-expansiveness across levels	$\ \Pi_K^{K+1}x\ _{W^{(K)}} \leq \ x\ _{W^{(K+1)}}$	Basic stability of coarse-graining (Lemma 2)
Cross-level discrepancy (ambient)	$d_{K\leftarrow L}(x^{(K)}, x^{(L)}) := \ i_K x^{(K)} - P_K(i_L x^{(L)})\ _W$	Canonical “apples-to-apples” cross-level distance
Projective telescoping bound	$\ i_K x^{(K)} - P_K(i_L x^{(L)})\ _W \leq \sum_{j=K}^{L-1} \ i_j x^{(j)} - P_j(i_{j+1} x^{(j+1)})\ _W$	Turns summable adjacent mismatches into a Cauchy tower (Lemma 3)
Refinement-limit learner (projective limit)	$x^{(\infty)}(t) \in \mathcal{H}$ with $P_K x^{(\infty)}(t) = i_K x^{(K)}(t)$ (compatible Cauchy tower)	Existence/uniqueness of a refinement-limit object (Theorem 2)
Ledgers, scalarizations, and readouts		
Ledger vector	$r(t) = (R_1(t), \dots, R_m(t))^{\top} \in \mathbb{R}_{\geq 0}^m$	Collects multiple nonnegative training-time quantities
Declared total ledger	$R_{\text{tot}}(t) = w^{\top} r(t)$ with $w \gg 0$	Single scalar clock target for contraction

Entry	Definition / Formula	Role in the paper
Quadratic ledger ruler (optional)	$R_{\text{tot}}(t) = \ r(t)\ _{W_{\text{led}}}^2 = r(t)^\top W_{\text{led}} r(t), W_{\text{led}} \succ 0$	Alternative scalarization when a quadratic contract is preferred
Reported readouts / metrics	$\mathcal{M} = \{M\}, \text{ with } M(\cdot) \geq 0$	External observables whose stability is transported
Dictionary (observability) line	$c_M R^{(K)}(t) \leq M(x^{(K)}(t)) \leq C_M R^{(K)}(t)$ uniformly in K	Transfers the certified clock to readouts ((O4))
Readout Lipschitz transport (optional)	$ M(u) - M(v) \leq L_M \ u - v\ _W$ on a declared bounded set	Converts ruler convergence to readout convergence (optional strengthening)
Metzler comparison and one-clock reduction		
Metzler matrix	$M \in \mathbb{R}^{m \times m}$ with $M_{ij} \geq 0$ for $i \neq j$	Positivity structure for ledger couplings
Metzler comparison inequality	$\dot{r}(t) \leq M r(t)$ componentwise	Auditable coupling model (Definition 1)
Semigroup positivity	M Metzler $\Rightarrow e^{tM} \geq 0$ entrywise for all $t \geq 0$	Enables order-preserving comparison
Comparison principle (Duhamel form)	If y solves $\dot{y} = My, y(0) = r(0)$ and $z := y - r$, then $z(t) = \int_0^t e^{(t-s)M} q(s) ds, q := \dot{z} - Mz \geq 0$	Correct proof mechanism for $r(t) \leq e^{tM} r(0)$ (Lemma 1)
Funding + injection parametrization	$M_{ii} = -2\lambda_i, M_{ij} = \eta_{i \leftarrow j} \geq 0$ ($i \neq j$)	Interpretable design levers (fund diagonals, reduce injections)
Spectral abscissa and effective rate	$\mu(M) := \max\{\text{Re } z : z \in \text{spec}(M)\}, \lambda_{\text{eff}} := -\mu(M)$ if $\mu(M) < 0$	Defines the certified one-clock exponent class
One-clock certificate (witness form)	$\exists w \gg 0, \lambda > 0$ with $w^\top M \leq -\lambda w^\top \Rightarrow w^\top r(t) \leq e^{-\lambda t} w^\top r(0)$	Core reduction theorem (Theorem 1)
Hurwitz \Rightarrow witness (sharp rate)	If $\mu(M) < 0$ and M is Metzler, $\exists w \gg 0$ with $w^\top M \leq \mu(M) w^\top$	Explains existence of witnesses / sharp clock (Proposition 1)
Two-ledger small-gain criterion	For 2×2 funded+injection M, M Hurwitz $\Leftrightarrow \eta_{1 \leftarrow 2} \eta_{2 \leftarrow 1} < 4\lambda_1 \lambda_2$	Exact design boundary (Proposition 2)
Master Certificate programme lines on $[0, T]$		
Certification horizon	$T > 0$ (declared)	All programme lines are audited on $[0, T]$
(O1) Tail-robust envelope	$\dot{R}^{(K)} \leq -2\lambda_K R^{(K)} + \tau_K, \int_0^T \tau_K \leq \alpha_K, \sum_K \alpha_K < \infty$	Controls time-direction pollution on $[0, T]$
(O2) Uniform margin	$\lambda_K \geq \lambda > 0$ independent of K	One exponent class across refinement
(O3) Geometric tower budget	$\sup_{t \in [0, T]} (\ (I - P_K)z^{(K+1)}(t)\ _W + \ z^{(K)}(t) - P_K z^{(K+1)}(t)\ _W) \leq \beta_K + \delta_K, \sum_K (\beta_K + \delta_K) < \infty$	Cauchy tower in one ruler (no drift) on $[0, T]$
(O4) Uniform dictionary on $[0, T]$	$c_M R^{(K)}(t) \leq M(x^{(K)}(t)) \leq C_M R^{(K)}(t)$ uniformly in K and $t \in [0, T]$	Transfers rates to reported metrics
Master Certificate (learning version)	(O1)–(O4) \Rightarrow refinement-limit trajectory + rate inheritance on $[0, T]$	Main soundness engine (Theorem 3)
Certificate artifact (what the prover ships)		

Entry	Definition / Formula	Role in the paper
Certificate artifact	Declared (\mathcal{H}, W) ; maps (i_K, P_K, Π_K^{K+1}) ; ledger definition $R^{(K)}$; budgets $(\alpha_K, \beta_K, \delta_K)$; margin λ ; dictionary constants (c_M, C_M) ; and (when applicable) a Metzler witness (w, λ)	Concrete verifier-facing interface (the “proof-carrying” object)
Witness-finding LP (optional recipe)	Find $w \in \mathbb{R}^m, \lambda \geq 0$ s.t. $w \geq \mathbf{1}, w^\top M \leq -\lambda w^\top, \sum_i w_i = 1$	Makes the one-clock witness actionable for ML/CS audiences

References

- Miettinen, K. *Nonlinear Multiobjective Optimization*; Vol. 12, *International Series in Operations Research & Management Science*, Kluwer Academic Publishers: Boston, MA, 1999. <https://doi.org/10.1007/978-1-4615-5563-6>.
- Sener, O.; Koltun, V. Multi-Task Learning as Multi-Objective Optimization. In Proceedings of the Advances in Neural Information Processing Systems, 2018, Vol. 31, pp. 525–536, [arXiv:cs.LG/1810.04650].
- Kaplan, J.; McCandlish, S.; Henighan, T.; Brown, T.B.; Chess, B.; Child, R.; Gray, S.; Radford, A.; Wu, J.; Amodei, D. Scaling Laws for Neural Language Models. *arXiv preprint* **2020**, [arXiv:cs.LG/2001.08361].
- Ciarlet, P.G. *The Finite Element Method for Elliptic Problems*; Vol. 4, *Studies in Mathematics and Its Applications*, North-Holland Publishing Company: Amsterdam, 1978.
- Hackbusch, W. *Multi-Grid Methods and Applications*; Vol. 4, *Springer Series in Computational Mathematics*, Springer-Verlag: Berlin, 1985. <https://doi.org/10.1007/978-3-662-02427-0>.
- Brenner, S.C.; Scott, L.R. *The Mathematical Theory of Finite Element Methods*, 3 ed.; Vol. 15, *Texts in Applied Mathematics*, Springer: New York, 2008. <https://doi.org/10.1007/978-0-387-75934-0>.
- Lax, P.D.; Richtmyer, R.D. Survey of the Stability of Linear Finite Difference Equations. *Communications on Pure and Applied Mathematics* **1956**, 9, 267–293. <https://doi.org/10.1002/cpa.3160090206>.
- Hille, E.; Phillips, R.S. *Functional Analysis and Semi-Groups*; Vol. 31, *American Mathematical Society Colloquium Publications*, American Mathematical Society: Providence, RI, 1957.
- Trotter, H.F. Approximation of Semi-Groups of Operators. *Pacific Journal of Mathematics* **1958**, 8, 887–919. <https://doi.org/10.2140/pjm.1958.8.887>.
- Pazy, A. *Semigroups of Linear Operators and Applications to Partial Differential Equations*; Vol. 44, *Applied Mathematical Sciences*, Springer: New York, 1983. <https://doi.org/10.1007/978-1-4612-5561-1>.
- Berman, A.; Plemmons, R.J. *Nonnegative Matrices in the Mathematical Sciences*; Academic Press, 1979. <https://doi.org/10.1016/C2013-0-10361-3>.
- Farina, L.; Rinaldi, S. *Positive Linear Systems: Theory and Applications*; Vol. 255, *Pure and Applied Mathematics*, Wiley–Interscience: New York, 2000.
- Smith, H.L. *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems*; Vol. 41, *Mathematical Surveys and Monographs*, American Mathematical Society: Providence, RI, 1995. <https://doi.org/10.1090/surv/041>.
- Briat, C. *Linear Parameter-Varying and Time-Delay Systems: Analysis, Observation, Filtering & Control*; Advances in Delays and Dynamics, Springer: Berlin, Heidelberg, 2015. <https://doi.org/10.1007/978-3-662-44050-6>.
- Kreyszig, E. *Introductory Functional Analysis with Applications*; Vol. 17, *Wiley Classics Library*, Wiley, 1989.
- Bousquet, O.; Elisseeff, A. Stability and Generalization. *Journal of Machine Learning Research* **2002**, 2, 499–526. <https://doi.org/10.1162/153244302760200704>.
- McAllester, D.A. PAC-Bayesian Model Averaging. In Proceedings of the Proceedings of the Twelfth Annual Conference on Computational Learning Theory (COLT '99), New York, NY, USA, 1999; pp. 164–170. <https://doi.org/10.1145/307400.307435>.
- Catoni, O. *PAC-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning*; Vol. 56, *Institute of Mathematical Statistics Lecture Notes–Monograph Series*, Institute of Mathematical Statistics: Beachwood, OH, 2007. <https://doi.org/10.1214/074921707000000391>.
- Duchi, J.C.; Namkoong, H. Learning Models with Uniform Performance via Distributionally Robust Optimization. *The Annals of Statistics* **2021**, 49, 1378–1406. <https://doi.org/10.1214/20-AOS2004>.
- Katz, G.; Barrett, C.; Dill, D.L.; Julian, K.; Kochenderfer, M.J. Reluplex: An Efficient SMT Solver for Verifying Deep Neural Networks. In Proceedings of the Computer Aided Verification (CAV 2017), Proceedings, Part

- I; Majumdar, R.; Kuncak, V., Eds., Cham, 2017; Vol. 10426, *Lecture Notes in Computer Science*, pp. 97–117. https://doi.org/10.1007/978-3-319-63387-9_5.
21. Gehr, T.; Mirman, M.; Drachler-Cohen, D.; Tsankov, P.; Chaudhuri, S.; Vechev, M. AI2: Safety and Robustness Certification of Neural Networks with Abstract Interpretation. In Proceedings of the 2018 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 2018; pp. 3–18. <https://doi.org/10.1109/SP.2018.00058>.
 22. Seshia, S.A.; Sadigh, D.; Sastry, S.S. Toward Verified Artificial Intelligence. *Communications of the ACM* **2022**, *65*, 46–55. An earlier technical version appeared as arXiv:1606.08514 (2016)., <https://doi.org/10.1145/3503914>.
 23. Necula, G.C. Proof-carrying code. In Proceedings of the Proceedings of the 24th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL '97), New York, NY, USA, 1997; pp. 106–119. <https://doi.org/10.1145/263699.263712>.
 24. Coddington, E.A.; Levinson, N. *Theory of Ordinary Differential Equations*; McGraw-Hill: New York, 1955.
 25. Khalil, H.K. *Nonlinear Systems*, 3 ed.; Prentice Hall: Upper Saddle River, NJ, 2002.
 26. Grönwall, T.H. Note on the Derivatives with Respect to a Parameter of the Solutions of a System of Differential Equations. *Annals of Mathematics (2)* **1919**, *20*, 292–296. <https://doi.org/10.2307/1967124>.
 27. Robbins, H.; Monro, S. A Stochastic Approximation Method. *The Annals of Mathematical Statistics* **1951**, *22*, 400–407. <https://doi.org/10.1214/aoms/1177729586>.
 28. Benaïm, M. Dynamics of Stochastic Approximation Algorithms. In *Séminaire de Probabilités XXXIII*; Azéma, J.; Émery, M.; Ledoux, M.; Yor, M., Eds.; Springer: Berlin, Heidelberg, 1999; Vol. 1709, *Lecture Notes in Mathematics*, pp. 1–68. <https://doi.org/10.1007/BFb0096509>.
 29. Kushner, H.J.; Yin, G.G. *Stochastic Approximation and Recursive Algorithms and Applications*, 2 ed.; Vol. 35, *Stochastic Modelling and Applied Probability*, Springer: New York, 2003. <https://doi.org/10.1007/978-1-4757-3104-9>.
 30. Bramble, J.H.; Pasciak, J.E.; Xu, J. Parallel Multilevel Preconditioners. *Mathematics of Computation* **1990**, *55*, 1–22. <https://doi.org/10.2307/2008686>.
 31. Thomée, V. *Galerkin Finite Element Methods for Parabolic Problems*, 2 ed.; Springer Series in Computational Mathematics, Springer: Berlin, Heidelberg, 2006. <https://doi.org/10.1007/3-540-33122-0>.
 32. Emmrich, E. Discrete Versions of Gronwall's Lemma and Their Application to the Numerical Analysis of Parabolic Problems. Preprint Reihe Mathematik 637, Technische Universität Berlin, 1999.
 33. Desoer, C.A.; Vidyasagar, M. *Feedback Systems: Input-Output Properties*; Academic Press: New York, 1975.
 34. Jiang, Z.P.; Teel, A.R.; Praly, L. Small-Gain Theorem for ISS Systems and Applications. *IEEE Transactions on Automatic Control* **1994**, *39*, 1609–1619. <https://doi.org/10.1109/9.310052>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.