OCTA Research

Technical Textbook Series

Memory Arithmetic

Numbers, History, and the Mathematics of Irreversibility

Chapter 2 Standalone Packet

Forgetting, Metrics, and the Geometry of Identity

Expanded Math-Book Edition · Build v2.3

Standalone compile: title + notation + chapter + exercises + worked diagrams + solutions

---

Chapter 2 Thesis (Operational)

If memory numbers carry provenance, then *abstraction* cannot be informal. This chapter formalizes **forgetting families** (depth, budget, MDL, stochastic), builds **metrics** and **embedding cores** for histories, and defines **experiential sameness** as shared causal structure. A fundamental phenomenon appears immediately: at a fixed abstraction budget, outcomes need not be unique (**operational curvature**).

Contents

Preface (Chapter 2 Packet)

Chapter 1 introduced memory numbers $m = \langle v \mid h \rangle$ and the irreversibility discipline: composition grows provenance; value projection stays classical. Chapter 2 makes the theory operational:

- **Forget** becomes a family of operators with explicit constraints.
- **Compare** becomes a metric/embedding problem on histories.
- **Audit** becomes a core-extraction problem (shared causal substructure).
- **Geometry** emerges because compression is not unique at equal budgets.

Key Idea

Forget is not a narrative shortcut; it is a primitive map $\mathcal{F}_\theta : \mathbb{M} \to \mathbb{M}$ that can be analyzed (idempotence, monotonicity, stability, contraction, curvature).

Notation (Chapter 2)

- $\mathbb{V}$: value space (commutative monoid/semiring; typically $\mathbb{R}$).
- $\mathbb{H}$: history space (finite rooted ordered labeled trees or DAGs).
- $\mathbb{M} := \mathbb{V} \times \mathbb{H}$: memory numbers; $m = \langle v \mid h \rangle$.
- $\mathrm{val}(m) \in \mathbb{V}$: value projection; $\mathrm{hist}(m) \in \mathbb{H}$: provenance.
- $C(h)$: history cost (nodes, bytes, description length, etc.).
- $\mu(m)$: mass proxy; typically $\mu(m) = C(\mathrm{hist}(m))$.
- $\mathrm{Prune}_\theta : \mathbb{H} \to \mathbb{H}$: pruning/summarization map at parameter θ.
- $\mathcal{F}_\theta : \mathbb{M} \to \mathbb{M}$ induced by pruning:

$$\mathcal{F}_\theta(\langle v \mid h \rangle) = \langle v \mid \mathrm{Prune}_\theta(h) \rangle.$$

- $d_H(h_1, h_2)$: history pseudo-metric; $\mathrm{Edit}(h_1, h_2)$ placeholder.
- Weighted memory metric:

$$d(m_1, m_2) = |\,\mathrm{val}(m_1) - \mathrm{val}(m_2)| + \lambda\, d_H(\mathrm{hist}(m_1), \mathrm{hist}(m_2)).$$

- Embedding $h_0 \preceq h$: rooted ordered label-preserving embedding.
- $\mathrm{core}(h_1, h_2)$: maximal common embedded subhistory (canonical choice if non-unique).
- Similarity score:

$$\mathrm{sim}_e(h_1, h_2) = \frac{|\,\mathrm{core}(h_1, h_2)|}{\max(|h_1|, |h_2|)} \in [0, 1].$$

Chapter 2

Forgetting, Metrics, and the Geometry of Identity

2.1 Why forgetting must be an operator

If provenance is structural, then abstraction must also be structural. In Memory Arithmetic, "we ignore details" must be expressible as an explicit map.

Definition 2.1 (Forgetting operator (schema)). *A forgetting operator is a map $\mathcal{F}_\theta : \mathbb{M} \to \mathbb{M}$ such that:*

$$\mathrm{val}(\mathcal{F}_\theta(m)) = \mathrm{val}(m) \qquad \textit{for all } m \in \mathbb{M}.$$

The parameter θ encodes an abstraction regime: depth, budget, MDL cap, or stochastic temperature.

Warning / Pitfall

Forgetting is not "subtracting history." It is projecting history under a constraint. The value remains exact; identity becomes coarser.

2.2 Four primary forgetting families

2.2.1 Family A: depth truncation

Definition 2.2 (Depth truncation on histories). *For $k \in \mathbb{N}$, $\mathrm{Prune}_k^{\mathrm{depth}}(h)$ retains nodes up to depth k (root depth 0) and replaces each cut subtree by a single summary leaf $\bullet$.*

Definition 2.3 (Depth-forgetting on memory numbers).

$$\mathcal{F}_k^{\mathrm{depth}}(\langle v \mid h\rangle) = \langle v \mid \mathrm{Prune}_k^{\mathrm{depth}}(h)\rangle.$$

2.2.2 Family B: budget pruning (cost-constrained abstraction)

Definition 2.4 (History cost functional). *A cost functional is a map $C : \mathbb{H} \to \mathbb{R}_{\geq 0}$ measuring size (nodes, bytes, description length).*

Definition 2.5 (Budget-feasible prunings). *Fix $B \geq 0$. A history h' is B-feasible for h if h' can be obtained from h by summarizing subtrees and satisfies $C(h') \leq B$.*

Definition 2.6 (Set-valued budget pruning)**.** *Define the feasible pruning set:*

$$\mathcal{P}_B(h) := \{h' \in \mathbb{H} : \ h' \textit{ is } B\textit{-feasible for } h\}.$$

A deterministic budget pruning is any selector $\mathrm{Prune}_B^{\mathrm{budget}}(h) \in \mathcal{P}_B(h)$.

OCTA Research Note

Budget pruning is naturally multi-valued. Different selectors encode different policies (retain newest, rarest, causal, audit-relevant, highest MDL utility). This non-uniqueness is the source of operational curvature.

2.2.3 Family C: MDL pruning (minimum description length)

Definition 2.7 (Description length)**.** *Let* $\ell : \mathbb{H} \to \mathbb{R}_{\geq 0}$ *assign code length to histories (bits/symbols), induced by a prefix code or grammar for ordered labeled trees.*

Definition 2.8 (Information-loss functional)**.** *Let* $\mathcal{L} : \mathbb{H} \times \mathbb{H} \to \mathbb{R}_{\geq 0}$ *measure loss when replacing* h *by* h'*. Interpretation is domain-specific (audit loss, causal loss, predictive loss).*

Definition 2.9 (MDL pruning (optimization form))**.** *For cap* $L \geq 0$ *define:*

$$\mathrm{Prune}_L^{\mathrm{MDL}}(h) \in_{h' \in \mathbb{H}} \ \mathcal{L}(h, h') \quad \textit{subject to} \quad \ell(h') \leq L.$$

2.2.4 Family D: stochastic pruning (randomized abstraction)

Definition 2.10 (Stochastic kernel on histories)**.** *A Markov kernel is a map* $\mathcal{K}_\theta(\cdot \mid h)$ *assigning a distribution on* $\mathbb{H}$ *given input history* h*.*

Definition 2.11 (Stochastic forgetting on memory numbers)**.** *For* $m = \langle v \mid h \rangle$ *define:*

$$\mathcal{F}_\theta^{\mathrm{stoch}}(m) = \langle v \mid h' \sim \mathcal{K}_\theta(\cdot \mid h) \rangle.$$

Key Idea

Stochastic forgetting turns abstraction into probability and enables expected-mass and contraction-in-expectation theorems.

2.3 Universal axioms for admissible forgetting

Definition 2.12 (Admissible forgetting family)**.** *A family* $\{\mathcal{F}_\theta\}_{\theta \in \Theta}$ *is admissible if:*

(F1) ***Value preservation:*** $\mathrm{val}(\mathcal{F}_\theta(m)) = \mathrm{val}(m)$.

(F2) ***Idempotence:*** $\mathcal{F}_\theta(\mathcal{F}_\theta(m)) = \mathcal{F}_\theta(m)$.

(F3) ***Monotonicity:*** *if* $\theta \preceq \theta'$ *means "more forgetting," then* $\mathcal{F}_\theta(m)$ *is a pruning of* $\mathcal{F}_{\theta'}(m)$.

(F4) ***Cost control:*** $C(\mathrm{hist}(\mathcal{F}_\theta(m))) \leq \mathrm{budget}(\theta)$.

Proposition 2.13 (Depth truncation is admissible)**.** *The depth family* $\{\mathcal{F}_k^{\mathrm{depth}}\}_{k \in \mathbb{N}}$ *is admissible with budget interpreted as depth cap.*

2.4 Comparing histories: metrics and pseudo-metrics

Definition 2.14 (History pseudo-metric)**.** *A map* $d_H : \mathbb{H} \times \mathbb{H} \to \mathbb{R}_{\geq 0}$ *is a pseudo-metric if:*

(E1) $d_H(h, h) = 0$,

(E2) $d_H(h_1, h_2) = d_H(h_2, h_1)$,

(E3) $d_H(h_1, h_3) \leq d_H(h_1, h_2) + d_H(h_2, h_3)$.

Definition 2.15 (Weighted lift to memory numbers)**.** *Fix* $\lambda \geq 0$ *and define:*

$$d(m_1, m_2) = |\operatorname{val}(m_1) - \operatorname{val}(m_2)| + \lambda\, d_H(\operatorname{hist}(m_1), \operatorname{hist}(m_2)).$$

Theorem 2.16 (Metric lifting)**.** *If* d_H *is a metric on* $\mathbb{H}$ *then* d *is a metric on* $\mathbb{M}$ *(for* $\lambda > 0$*).*

2.5 Embeddings and experiential sameness

Definition 2.17 (Rooted ordered embedding)**.** *An embedding* $e : h_0 \hookrightarrow h$ *is an injective node map preserving root, ancestry, node labels, and left-to-right order of children. Write* $h_0 \preceq h$ *if such an embedding exists.*

Proposition 2.18. *The relation* $\preceq$ *is a preorder on* $\mathbb{H}$.

Definition 2.19 (Core operator)**.**

$$\operatorname{core}(h_1, h_2) \in \operatorname{argmax}\{|h_0| : \ h_0 \preceq h_1 \text{ and } h_0 \preceq h_2\}.$$

If multiple maximizers exist, choose a canonical representative (e.g. minimal signature).

Definition 2.20 (Experiential similarity score)**.**

$$\operatorname{sim}_e(h_1, h_2) = \frac{|\operatorname{core}(h_1, h_2)|}{\max(|h_1|, |h_2|)} \in [0, 1].$$

2.6 Operational curvature

Definition 2.21 (Curvature witness (budget form))**.** *Fix cost* C *and budget* B. *A history* h *exhibits curvature at* B *if there exist* $h_A, h_B \in \mathcal{P}_B(h)$ *with* $h_A \not\cong h_B$.

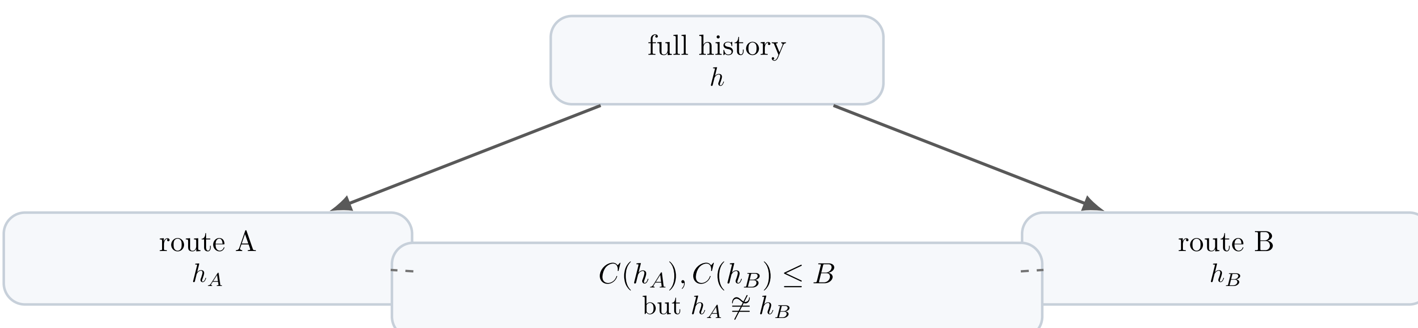


Figure 2.1: Operational curvature: equal budget does not imply unique abstraction outcome.

2.7 Exercises (each with worked diagrams and solutions)

Exercise

Exercise 1 (Depth-forgetting is admissible). Let $\mathcal{F}_k^{\mathrm{depth}}$ be induced by depth truncation on ordered labeled trees. Prove:

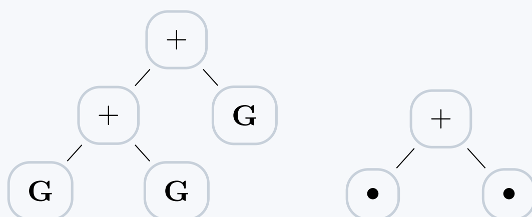
(a) $\mathrm{val}(\mathcal{F}_k^{\mathrm{depth}}(m)) = \mathrm{val}(m)$.

(b) $\mathcal{F}_k^{\mathrm{depth}}(\mathcal{F}_k^{\mathrm{depth}}(m)) = \mathcal{F}_k^{\mathrm{depth}}(m)$.

(c) If $k \le \ell$, then $\mathrm{Prune}_k^{\mathrm{depth}}(h)$ is a pruning of $\mathrm{Prune}_\ell^{\mathrm{depth}}(h)$.

Worked Example / Guided Work

Picture intuition. Truncation to depth k deletes everything below depth k. A second truncation to the same depth sees nothing new to delete.



Left: full. Right: depth-1 truncation.

Solution

Solution. (a) By definition, forgetting changes only hist, not val. (b) After truncation, all nodes below depth k are already replaced by leaves; truncating again has no effect. (c) If $k \le \ell$, then the set of nodes retained at depth k is a subset of those retained at depth ℓ; hence the k-truncation is a pruning of the ℓ-truncation.

Exercise

Exercise 2 (Metric lifting). Assume d_H is a metric on $\mathbb{H}$ and $|\cdot|$ is the usual metric on $\mathbb{R}$. Prove that

$$d(m_1, m_2) = |\,\mathrm{val}(m_1) - \mathrm{val}(m_2)| + \lambda\, d_H(\mathrm{hist}(m_1), \mathrm{hist}(m_2))$$

is a metric on $\mathbb{M}$ for $\lambda > 0$.

Worked Example / Guided Work

Triangle inequality blueprint. Use triangle inequality separately:

$$|\Delta v_{13}| \le |\Delta v_{12}| + |\Delta v_{23}|, \quad d_H(h_1, h_3) \le d_H(h_1, h_2) + d_H(h_2, h_3),$$

then add with weight λ.

Solution

Solution. Nonnegativity and symmetry are immediate. If $d(m_1, m_2) = 0$, then both summands are 0 so $\mathrm{val}(m_1) = \mathrm{val}(m_2)$ and $d_H(\mathrm{hist}(m_1), \mathrm{hist}(m_2)) = 0$. Since d_H is a metric, $\mathrm{hist}(m_1) = \mathrm{hist}(m_2)$, hence $m_1 = m_2$. Triangle inequality holds by summing triangle inequalities of each term.

Exercise

Exercise 3 (Compute a core in the associativity fracture). In the ordered-tree model, let

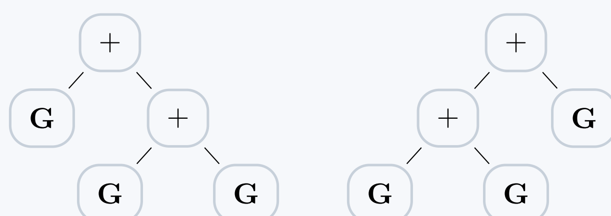
$$h_1 = +(\mathbf{G},\ +(\mathbf{G}, \mathbf{G})), \qquad h_2 = +(+(\mathbf{G}, \mathbf{G}),\ \mathbf{G}).$$

(a) Determine $\mathrm{core}(h_1, h_2)$ up to isomorphism (using ordered embeddings).

(b) Compute $\mathrm{sim}_e(h_1, h_2)$ when $|\cdot|$ counts internal nodes.

Worked Example / Guided Work

Draw both trees.



Because the internal + appears on opposite sides, the largest common ordered embedded structure is just the root + with children treated as leaves. So the core has one internal node.

Solution

Solution. (a) The maximal common ordered embedded subhistory is the single root node labeled + (with leaf children). Thus $|\mathrm{core}(h_1, h_2)| = 1$. (b) Each tree has two internal nodes, so

$$\mathrm{sim}_e(h_1, h_2) = \frac{1}{\max(2, 2)} = \frac{1}{2}.$$

Exercise

Exercise 4 (Curvature witness under equal budget). Let $h = +(A, B)$ be a history where $A \ncong B$ and child order matters. Define a budget rule: "retain exactly one subtree in full and summarize the other by $\bullet$." Show there exist two feasible prunings that are not isomorphic, hence curvature is witnessed.

Worked Example / Guided Work

Two routes.

$$
\begin{tikzcd}
& +(A,B) \arrow[dl] \arrow[dr] & \\
+(A,\bullet) & & +(\bullet,B)
\end{tikzcd}
$$

Order-preserving isomorphisms cannot swap left/right children; thus inequivalence is generic.

Solution

Solution. Both prunings satisfy the budget by construction. If $A \ncong B$ and order is preserved, then $+(A, \bullet) \ncong +(\bullet, B)$. Thus equal-budget abstraction is not unique.

Exercise

Exercise 5 (A clean 1-Lipschitz statement for truncation). Fix any $K \geq 0$ and define a capped-count pseudo-metric on histories:

$$d_H^K(h_1, h_2) := \left|\min(|h_1|, K) - \min(|h_2|, K)\right|.$$

Show that depth truncation is 1-Lipschitz with respect to d_H^K whenever $|\operatorname{Prune}_k^{\mathrm{depth}}(h)| \leq K$ for all h.

Worked Example / Guided Work

Key reduction. If truncation always lands inside the cap, then

$$\min(|\operatorname{Prune}_k(h)|, K) = |\operatorname{Prune}_k(h)|.$$

So you only compare output sizes, which truncation cannot increase.

Solution

Solution. Assume outputs satisfy $|\operatorname{Prune}_k(h)| \le K$. Then

$$d_H^K(\operatorname{Prune}_k(h_1), \operatorname{Prune}_k(h_2)) = \big|\,|\operatorname{Prune}_k(h_1)| - |\operatorname{Prune}_k(h_2)|\,\big|.$$

Truncation is deletion/summarization, hence it cannot increase size gaps under the capped regime, yielding

$$d_H^K(\operatorname{Prune}_k(h_1), \operatorname{Prune}_k(h_2)) \le d_H^K(h_1, h_2).$$

Exercise

Exercise 6 (Preorder property of embedding). Prove that $\preceq$ defined by rooted ordered embedding is reflexive and transitive. Give a condition under which it becomes antisymmetric (a partial order up to isomorphism).

Worked Example / Guided Work

Identity embedding gives reflexivity. Composition of embeddings gives transitivity. Antisymmetry holds (up to $\cong$) if embeddings are induced-subtree embeddings that preserve exact edges and forbid collapsing.

Solution

Solution. Reflexive via identity embedding. Transitive via composition. If embeddings are strict induced-subtree embeddings preserving exact parent-child relations (no collapsing), then $h_1 \preceq h_2$ and $h_2 \preceq h_1$ implies $h_1 \cong h_2$.

Exercise

Exercise 7 (Experiential overlap is not transitive). Fix $\alpha \in (0, 1)$. Construct h_A, h_B, h_C such that

$$\operatorname{sim}_e(h_A, h_B) \ge \alpha, \quad \operatorname{sim}_e(h_B, h_C) \ge \alpha, \quad \operatorname{sim}_e(h_A, h_C) < \alpha.$$

Worked Example / Guided Work

Let h_B contain two large disjoint modules X and Y. Let h_A share X with h_B and let h_C share Y with h_B. Then overlap occurs through B but not directly between A and C.

Solution

Solution. Take $h_B = +(X, Y)$ with $|X| = |Y|$ and $X \ncong Y$. Let $h_A = X$ and $h_C = Y$. Then $\mathrm{sim}_e(h_A, h_B) \approx 1/2$ and $\mathrm{sim}_e(h_B, h_C) \approx 1/2$. But $\mathrm{sim}_e(h_A, h_C)$ is small because X and Y share only trivial structure. Scaling sizes makes the strict inequality hold for a chosen α.

Exercise

Exercise 8 (MDL feasible set nesting). Show that if $L_1 \leq L_2$, then:

$$\min_{\ell(h') \leq L_2} \mathcal{L}(h, h') \;\leq\; \min_{\ell(h') \leq L_1} \mathcal{L}(h, h').$$

Worked Example / Guided Work

The feasible set $\{h' : \ell(h') \leq L_1\}$ is contained in $\{h' : \ell(h') \leq L_2\}$. Minimizing over a larger set cannot worsen the minimum.

Solution

Solution. If $L_1 \leq L_2$ then $\mathcal{F}(L_1) \subseteq \mathcal{F}(L_2)$ where $\mathcal{F}(L) = \{h' : \ell(h') \leq L\}$. Hence $\min_{\mathcal{F}(L_2)} \leq \min_{\mathcal{F}(L_1)}$.

Exercise

Exercise 9 (Depth profile: chain vs balanced). Let $R(h)$ count internal nodes retained after depth truncation. For $+$-only histories on n leaves, compare chain vs balanced at fixed small depth k and conclude chains are more fragile.

Worked Example / Guided Work

Chain retains at most about k internal nodes when $n \gg k$. Balanced retains about 2^k internal nodes (until saturating at n). So balanced keeps much more structure at the same k.

Solution

Solution. Chain: $R(\mathrm{Prune}_k(h_{\mathrm{chain}})) = O(k)$ for $n \gg k$. Balanced: $R(\mathrm{Prune}_k(h_{\mathrm{bal}})) = \Theta(2^k)$ until $2^k \approx n$. Thus at equal k and large n, balanced retains far more structure, so chains lose structure earlier and in sharper steps, increasing fragility.

Exercise

Exercise 10 (Expected retained nodes under Bernoulli retention). Assume each internal node is retained independently with probability p. If the original history has N internal nodes and X is the number retained, compute $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

Worked Example / Guided Work

Model $X = \sum_{i=1}^{N} I_i$ with i.i.d. Bernoulli(p) indicators.

Solution

Solution. $\mathbb{E}[X] = Np$ and $\mathrm{Var}(X) = Np(1-p)$ by independence and additivity of expectation/variance.

Chapter Summary

- Forgetting is a family: depth, budget, MDL, stochastic.
- Budget/MDL regimes are multi-valued; this produces operational curvature.
- History distances lift to memory distances by weighting value and identity.
- Experiential sameness is captured via maximal embedded cores and similarity thresholds.
- Forgetting profiles and fragility convert abstraction into geometry.

Roadmap (Where This Goes Next)

Chapter 3 (next) should:

- Prove irreversibility theorems formally (no cancellation, no cycles without forgetting).
- Build entropy/second-law statements with MDL semantics and Landauer-style bounds.
- Define decay operators (continuous families) and approximation regimes.
- Upgrade d_H to explicit edit scripts; discuss DAG relaxations.
- Add larger worked examples with distributivity and computation-graph witnesses.