Admissible Histories and the Structure of Physical Theory

Geometric Foundations of Dynamics, Irreversibility, and Observation

Flyxion
Independent Researcher

October 2026

Contents

Preface **v**

I Histories and Their Observable Representations **1**

1 Histories and Admissible Continuations **2**

2 Observation and the Sufficiency Question **4**

3 The Minimal Sufficient Observation **6**

4 Memory Depth and the Hierarchy of Observations **8**

5 Resolution, Refinement, and the Lattice of Observers **10**

6 Time Translation and Dynamics on the Minimal State **11**

II Projection and Composition **13**

7 Amplitudes, Probabilities, and the Failure of Composition **15**

8 Support Sufficiency and Probabilistic Sufficiency **18**

9 The Quantum State as a Minimal Probabilistic Quotient **20**

10 Sufficiency for Coherent Composition **22**

11 Context Graphs and the Gluing Obstruction **24**

12 Coarse-Graining and Effective Stochastic Dynamics **27**

13 Memory Kernels and the Projection Operator Method **30**

14 Consistent Histories and Decoherence Conditions **33**

15 Which Structures Survive Projection **36**

III Finite Models and Worked Examples **39**

16 Finite History Structures and Their Minimal Automata **41**

17 Two-Level Systems and Unistochastic Chains **43**

18 Hidden Markov Models and Insufficient Observations **45**

19 Lattice Gases and Local Admissibility **47**

20 Computing Minimal Sufficient Observations **49**

21 A Catalogue of Obstructions in Small Models **50**

IV Transport, Bundles, and Holonomy **52**

22 Families of Continuations as Bundles **53**

23 Distinction Transport Along Histories **55**

24 Connections on Continuation Spaces **57**

25 Holonomy of Closed Histories **59**

26 Curvature, Holonomy, and Irreversible Memory **60**

27 Gauge Structure From Transport **62**

28 Geometric Phase as Retained Information **64**

29 Gluing as Holonomy **66**

V Irreversible Dynamics and Memory **71**

30 Irreversibility as Loss of Sufficiency **73**

31 Entropy Production Under Projection **75**

32 Thermodynamic Descriptions as Observations **77**

33 Fluctuation Relations and Path Ensembles **79**

34 Non-Markovian Dynamics and Retained Variables **81**

35 Arrows of Time and Admissible Orderings **83**

VI Continuum Fields and the RSVP Construction **85**

36 From Histories to Fields **87**

37 Scalar Capacity, Directed Flow, and Entropy **89**

38 Closure of the Field Equations **91**

39 Conservation Laws and Stability **94**

40 Effective Metrics From Field Data **97**

41 Unistochastic Dynamics and the RSVP Bridge **99**

42 Gravitation and Effective Geometry **102**

VII Accessibility and Cosmology **104**

43 Physical Structure and Accessible Structure **106**

44 Accessibility Kernels **108**

45 Distinction Functionals **110**

46 Mixing, Attenuation, and Causal Separation **112**

47 What an Approach to Smoothness Establishes **114**

48 Turnover and Indistinguishability Results **116**

49 Expansion, Redshift, and Observational Comparison **119**

VIII Obstructions, Computation, and Open Problems 121

50 Four Obstructions and Their Common Form 123

51 Failure to Reconstruct a History 126

52 Failure to Compose Reduced Descriptions 128

53 Failure to Extend a Local Continuation Globally 130

54 Failure to Preserve Distinctions Under Observation 132

55 Computational Experiments 134

56 Open Problems 137

A Notation 140

B Divergences and Gaussian Channels 143

C The Lie–Poisson Structure of the Fluid Bracket 145

D The Geometry of the Probability Projection 148

E Growth and Horizons in the Flat ΛCDM Background 150

F The Experiment Script 152

Preface

This book asks how a physical world acquires a space of possible continuations, how those continuations compose, and how observers recover only part of that structure. The question is posed at a level of generality that allows quantum theory, thermodynamics, gravitation, and cosmology to be treated as instances of one mathematical problem, while allowing each instance to develop its own machinery and to be judged on its own terms.

The organizing distinction is between a physical history and its observable representation. A history carries everything that has happened and, with it, a set of admissible continuations. An observation carries less. Whether the less is enough, whether two observationally identical presents can be followed by different futures, is a precise question that has a definite answer in each model, and the answer determines what additional structure a theory must retain.

The development proceeds in a deliberate order. History, projection, and composition form the foundation, because they supply the definitions and the first theorems on which every later part depends. Transport and holonomy, irreversible memory, continuum field descriptions, and accessibility in cosmology are built on that foundation, and each is claimed only to the extent that its construction is carried out. Conjectural bridges are labelled as conjectures, and the assumptions that would be needed to close them are stated as part of the problem rather than hidden in the exposition.

The appendices collect material that supports several parts. Appendix A is a table of notation with a list of the symbols that carry more than one meaning. Appendices B to E give the proofs of the information-theoretic inequalities, the Lie–Poisson structure of the fluid bracket, the geometry of the probability projection, and the growth and horizon formulas of the cosmological background. Appendix F lists the script that reproduces every numerical and symbolic check.

Part I

Histories and Their Observable Representations

Chapter 1

Histories and Admissible Continuations

A physical theory, before it is a set of equations, is a specification of what can happen next given what has happened so far. The first task is to make that specification a mathematical object that does not presuppose a state space.

Definition 1.1 (History structure)**.** A *history structure* consists of a set $\mathcal{H}$ of histories, a set $\mathcal{F}$ of future segments, and a map $\mathcal{A}\colon \mathcal{H} \to \mathcal{P}(\mathcal{F})$ assigning to each history h its set $\mathcal{A}(h) \subseteq \mathcal{F}$ of *admissible continuations*. A history is *terminal* if $\mathcal{A}(h) = \varnothing$.

The set $\mathcal{F}$ is common to all histories. This is the only structural demand, and it encodes the idea that a continuation is described relative to the present moment and not relative to the past that produced it. A deterministic theory is one in which every $\mathcal{A}(h)$ is a singleton. A stochastic theory enriches $\mathcal{A}(h)$ with a measure, which is deferred until it is needed. The bare set-valued form is enough to formulate the questions of this part.

Example 1.2 (Free particle with hidden velocity)**.** Let $\mathcal{H}$ be the set of finite position records $h = (x_0, \dots, x_n)$ of a particle on the line moving at constant velocity $v \in \mathbb{R}$, sampled at unit time steps, with $n \geq 1$. Let $\mathcal{F}$ be the set of sequences $(y_1, y_2, \dots)$ of displacements from the present position. Then $\mathcal{A}(h)$ is the single sequence $y_k = kv$ with $v = x_n - x_{n-1}$. The history determines its continuation, but the present position x_n alone does not.

Example 1.3 (A bare branching rule)**.** Let $\mathcal{H}$ be the set of finite words over $\{a, b\}$ and let $\mathcal{F} = \{a, b\}^{\mathbb{N}}$ be the set of infinite words. Declare

$$\mathcal{A}(h) = \begin{cases} \{a, b\}^{\mathbb{N}}, & aa \text{ has not occurred in } h, \\ b\{a, b\}^{\mathbb{N}}, & aa \text{ has occurred in } h. \end{cases}$$

There are exactly two continuation classes, and membership in the second class is a global feature of the word that no bounded window of final letters can detect. This is only a bare history structure. Its continuation sets do not satisfy the extendability identity of Chapter 6, because they do not shrink consistently when a letter is appended.

Example 1.4 (A branching rule with sequential dynamics)**.** Let L be the prefix-closed language of finite words over $\{a, b\}$ in which no a occurs after the first occurrence of aa. Let $\mathcal{H} = L$, $\mathcal{F} = \{a, b\}^{\mathbb{N}}$, and let $\mathcal{A}(h)$ be the set of infinite words φ such that every finite prefix of $h\varphi$ lies in L. Appending a letter and removing it from the continuations then satisfies the extendability identity. There are three continuation classes: words in which aa has not occurred and which do not end in a (including the empty word), words in which aa has not occurred and which end in a, and words in which aa has occurred, for which $\mathcal{A}(h) = \{b^{\mathbb{N}}\}$.

The examples show the two ways a present can underdetermine its future. In the first, a single numerical datum is missing from the instantaneous description. In the second, the relevant information is a global feature of the past. The theory of the next chapters separates these cases and measures how much must be retained in each.

Chapter 2

Observation and the Sufficiency Question

An observation is a map out of the space of histories. It is the formal representative of what an observer, an instrument, or a coarse description retains.

Definition 2.1 (Observation)**.** An *observation* of a history structure $(\mathcal{H}, \mathcal{F}, \mathcal{A})$ is a surjective map $\pi\colon \mathcal{H} \to \mathcal{X}$. Two histories are *observationally equivalent* if $\pi(h_1) = \pi(h_2)$. The observation is *sufficient* if

$$\pi(h_1) = \pi(h_2) \implies \mathcal{A}(h_1) = \mathcal{A}(h_2),$$

and *insufficient* otherwise.

Insufficiency is exactly the phenomenon

$$\pi(h_1) = \pi(h_2) \qquad \text{while} \qquad \mathcal{A}(h_1) \neq \mathcal{A}(h_2),$$

that is, two observationally identical presents carrying different admissible futures. It is a property of the pair $(\pi, \mathcal{A})$ and not of the observation alone, so a given measurement can be sufficient for one theory and insufficient for another.

Proposition 2.2 (Factorization)**.** *An observation $\pi\colon \mathcal{H} \to \mathcal{X}$ is sufficient if and only if there is a map $f\colon \mathcal{X} \to \mathcal{P}(\mathcal{F})$ with $\mathcal{A} = f \circ \pi$.*

Proof. If π is sufficient, define $f(x) = \mathcal{A}(h)$ for any h with $\pi(h) = x$. Surjectivity of π makes f defined on all of $\mathcal{X}$, and sufficiency makes it well defined. Conversely, if $\mathcal{A} = f \circ \pi$, then $\pi(h_1) = \pi(h_2)$ gives $\mathcal{A}(h_1) = f(\pi(h_1)) = f(\pi(h_2)) = \mathcal{A}(h_2)$. □

Proposition 2.2 says that sufficiency is determination by the observed state: for a sufficient observation the set of admissible continuations is a function of the observed value, with no further dependence on the history. That a function exists is a set-theoretic fact. Whether it can be computed requires effective representations of observations and continuation sets, and an algorithm, which are not part of the definition.

Definition 2.3 (Sufficiency defect)**.** Let d be a metric on $\mathcal{P}(\mathcal{F})$ (or on the subfamily $\mathcal{A}(\mathcal{H})$). For an observation π, the *sufficiency defect* at $x \in \mathcal{X}$ is

$$\delta_\pi(x) = \sup\{\, d(\mathcal{A}(h_1), \mathcal{A}(h_2)) : \pi(h_1) = \pi(h_2) = x \,\}.$$

The defect vanishes identically if and only if π is sufficient. The Hausdorff distance is a metric on the nonempty compact subsets of a metric space $\mathcal{F}$, so it is available as a specialization only when the continuation sets are nonempty and compact. On wider classes it can be infinite or fail to separate distinct sets with the same closure, and terminal histories introduce the empty set. Under those assumptions a small defect says that the unobserved part of the history changes the admissible futures only slightly, which is the quantitative content of a statement such as “the state description is approximately closed.”

Chapter 3

The Minimal Sufficient Observation

The sufficiency question has a canonical answer, which is a close analogue of the Myhill–Nerode construction in the theory of automata [36] and of the causal states of computational mechanics [6, 39].

Definition 3.1. Define $\pi_{\mathcal{A}} \colon \mathcal{H} \to \mathcal{A}(\mathcal{H}) \subseteq \mathcal{P}(\mathcal{F})$ by $\pi_{\mathcal{A}}(h) = \mathcal{A}(h)$. The *continuation equivalence* on $\mathcal{H}$ is the relation $h_1 \sim_{\mathcal{A}} h_2$ if and only if $\mathcal{A}(h_1) = \mathcal{A}(h_2)$.

The comparison class for the following theorem is the class of deterministic observations $\pi \colon \mathcal{H} \to \mathcal{X}$ that are sufficient for the continuation sets $\mathcal{A}$, ordered by factorization as in Chapter 5.

Theorem 3.2 (Minimality)**.** *The observation $\pi_{\mathcal{A}}$ is sufficient, and it factors through every sufficient observation $\pi \colon \mathcal{H} \to \mathcal{X}$: there is a unique surjection $g \colon \mathcal{X} \to \mathcal{A}(\mathcal{H})$ with $\pi_{\mathcal{A}} = g \circ \pi$. Consequently the continuation equivalence is the coarsest equivalence relation on $\mathcal{H}$ whose quotient observation is sufficient.*

Proof. Sufficiency of $\pi_{\mathcal{A}}$ is immediate from the definition. Let π be sufficient. By Proposition 2.2 we have $\mathcal{A} = f \circ \pi$, so $g = f$ satisfies $\pi_{\mathcal{A}} = g \circ \pi$. Surjectivity of g follows from surjectivity of $\pi_{\mathcal{A}}$ onto its image, and uniqueness follows from surjectivity of π. For the last statement, if an equivalence relation $\approx$ has sufficient quotient map q, then $q(h_1) = q(h_2)$ implies $\mathcal{A}(h_1) = \mathcal{A}(h_2)$, so $\approx$ refines $\sim_{\mathcal{A}}$. □

The minimal sufficient observation is the answer to the question of what a theory must remember in order to determine the continuation sets $\mathcal{A}$. Everything finer than it is redundant for that purpose, and everything coarser loses information about it; redundancy is relative to the chosen continuation sets and not to every question one might ask of the

history. In the free particle of Chapter 1, with futures described as displacements from the present position, the minimal sufficient observation is the velocity, a datum not contained in the position. In Example 1.4 it takes three values, recording whether aa has occurred and, if it has not, whether the word ends in a. The bare structure of Example 1.3 has two classes.

Remark 3.3. The theorem does not claim that the minimal observation is physically accessible, finite, or small. Whether $\mathcal{A}(\mathcal{H})$ is a finite set, a manifold, or an object of no useful structure at all is a property of the theory, and the classification of those cases is the subject of later chapters.

Chapter 4

Memory Depth and the Hierarchy of Observations

A natural way to repair an insufficient observation is to remember more of the recent past. This chapter makes the repair precise and shows that it generates a strict hierarchy.

Suppose each history is a finite word $h = (o_1, \dots, o_n)$ over an alphabet Σ of instantaneous observations. For $k \geq 1$ define the depth-k observation

$$\pi_k(h) = (o_{n-k+1}, \dots, o_n),$$

with the convention that shorter histories are retained whole.

Proposition 4.1. *If π_k is sufficient then π_m is sufficient for every $m \geq k$.*

Proof. If $\pi_m(h_1) = \pi_m(h_2)$ with $m \geq k$, then the last k letters of the two histories agree, so $\pi_k(h_1) = \pi_k(h_2)$ and hence $\mathcal{A}(h_1) = \mathcal{A}(h_2)$. □

Definition 4.2. The *memory depth* of the history structure over Σ is the least k for which π_k is sufficient, or ∞ if no such k exists.

Proposition 4.3 (The hierarchy is strict)**.** *For every $k \geq 1$ there is a history structure of memory depth exactly k.*

Proof. For $k = 1$ the claim is that depth 1 is attained by a structure for which π_1 is sufficient, since the definition admits only $k \geq 1$; take $\Sigma = \{0, 1\}$ and $\mathcal{A}(h)$ depending on the last letter alone. For $k \geq 2$, take $\Sigma = \{0, 1\}$ and let $\mathcal{A}(h) = \{\mathbf{1}\}$ if h has length at least k and its last k letters are all 1, and $\mathcal{A}(h) = \{\mathbf{0}\}$ otherwise, where $\mathbf{1}$ and $\mathbf{0}$ are two distinct future segments. Then π_k determines $\mathcal{A}$. The histories 01^{k-1} and 1^k have the same last $k-1$ letters, namely 1^{k-1}, but different admissible sets, so π_{k-1} is insufficient. □

The history structure of Example 1.4 has memory depth ∞ over the alphabet $\{a, b\}$: for every k, the words aab^k and b^{k+2} have the same last k letters and different continuation classes. Its minimal sufficient observation is nonetheless finite, with three values, which separates two notions that are easily confused: the depth of memory needed to read the state off the recent record, and the cardinality of the minimal state itself. A history structure can need unbounded depth while its minimal sufficient observation takes only finitely many values, and the passage from the first description to the second is the replacement of a window by an internal variable.

Chapter 5

Resolution, Refinement, and the Lattice of Observers

Observations of a fixed history structure are ordered by how much they retain, and this order has a clean algebraic form.

Definition 5.1. For observations π_1, π_2 of $\mathcal{H}$ write $\pi_1 \preceq \pi_2$ if π_1 factors through π_2, that is, $\pi_1 = g \circ \pi_2$ for some g. Say that π_2 *refines* π_1.

Identifying observations that refine each other, the order $\preceq$ is the opposite of the inclusion order on the equivalence relations $\ker \pi \subseteq \mathcal{H} \times \mathcal{H}$, so the observations of $\mathcal{H}$ up to isomorphism form the lattice of partitions of $\mathcal{H}$, ordered by refinement.

Proposition 5.2. *The set of sufficient observations is an upper set in the refinement order, and by Theorem 3.2 it has the continuation partition as its least element.*

Proof. If π_1 is sufficient and $\pi_1 = g \circ \pi_2$, then $\pi_2(h_1) = \pi_2(h_2)$ forces $\pi_1(h_1) = \pi_1(h_2)$ and hence $\mathcal{A}(h_1) = \mathcal{A}(h_2)$. Minimality is Theorem 3.2. □

Two observers whose instruments are incomparable under refinement can each be insufficient while their joint observation, the meet of the kernels, is sufficient. In that case the missing information is distributed across the observers and no single one holds it. This is the first appearance in the book of a pattern that recurs: obstructions that are invisible to every single description can be removed by combining descriptions, and the combination has a definite mathematical form.

The lattice picture also fixes what is meant by coarse-graining in the later parts. A coarse-graining is a passage down the lattice, a loss of resolution, and the question of the next part is when a dynamics defined at one level of the lattice induces a dynamics at a lower one.

Chapter 6

Time Translation and Dynamics on the Minimal State

Up to now no dynamics has been introduced; $\mathcal{A}$ is only a set-valued map. A dynamics appears when histories can be extended and the extension is compatible with the continuation structure.

Definition 6.1. A history structure is *extendable* if $\mathcal{H}$ carries a partial operation $(h, s) \mapsto h \cdot s$ for s a one-step prefix of an element of $\mathcal{A}(h)$, and a shift $T\colon \mathcal{F} \to \mathcal{F}$ removing the first step, such that

$$\mathcal{A}(h \cdot s) = \{\, T(\phi) : \phi \in \mathcal{A}(h),\ \phi \text{ begins with } s \,\}.$$

This is the compatibility between the present and the future that makes the structure a dynamical system and not merely a list.

Theorem 6.2 (Dynamics on the minimal state)**.** *In an extendable history structure, the extension operation descends to the minimal sufficient observation: if $\mathcal{A}(h_1) = \mathcal{A}(h_2)$, then $\mathcal{A}(h_1 \cdot s) = \mathcal{A}(h_2 \cdot s)$ for every admissible one-step prefix s. Hence $\mathcal{A}(\mathcal{H})$ carries a well-defined partial transition $\mathcal{A}(h) \mapsto \mathcal{A}(h \cdot s)$.*

Proof. By the extendability identity, $\mathcal{A}(h \cdot s)$ is computed from $\mathcal{A}(h)$, s, and T alone. Equal sets $\mathcal{A}(h_1) = \mathcal{A}(h_2)$ therefore produce equal sets $\mathcal{A}(h_1 \cdot s) = \mathcal{A}(h_2 \cdot s)$. □

The theorem is the structural reason that a state space exists at all. A physical state is not assumed in advance; it is the quotient of histories by equality of admissible futures, and the dynamics on it is forced by extendability. The extendability identity is what supplies the compatibility $h \sim k \Rightarrow h \cdot s \sim k \cdot s$; for an arbitrary assignment of continuation sets this implication is an additional assumption, and it can fail.

A more familiar state space enters through an observation $p\colon \mathcal{H} \to P$, and the relation of P to the minimal state needs care. If p is sufficient, Theorem 3.2 gives a map $g\colon P \to \mathcal{A}(\mathcal{H})$, so the familiar state determines the minimal one. A map in the reverse direction exists only if the coordinates p are constant on continuation classes. The free particle shows the difference: velocity determines the future displacements, while position is needed in addition if the futures are to be described as absolute positions. Minimality is therefore relative to the selected description of the future, including its coordinate convention. The question whether a familiar description is closed is the question whether the corresponding observation is sufficient for the chosen continuation sets. The next part studies what happens when it is not, beginning with the case that motivates the book most directly.

Part II

Projection and Composition

The quantum case is the first place where the foundation of Part I meets a physical theory in which a natural observation is demonstrably insufficient. Passing from amplitudes to the probabilities of a fixed measurement forgets structure, and the forgotten structure returns the moment two processes are joined. This part treats the passage in three steps. First it shows exactly how the projected description fails to compose. Next it separates two grades of sufficiency, one for the support of future outcomes and one for their probabilities, and identifies the quantum state as the minimal quotient for the second grade. Finally it asks what a description must retain in order to compose, and finds that the answer is strictly larger than the collection of probabilities and strictly smaller than the full unitary. The constructions follow the author's reconstruction of quantum theory from admissible probabilistic composition [7], which is also the source of the context-graph results of the last chapter of this part.

Chapter 7

Amplitudes, Transition Probabilities, and the Failure of Composition

Let U be a unitary matrix acting on a finite-dimensional Hilbert space with a fixed orthonormal basis. Passing from amplitudes to transition probabilities replaces U by the matrix

$$T(U)_{ij} = |U_{ij}|^2,$$

which is doubly stochastic, and a stochastic matrix arising this way is called *unistochastic*. The map $U \mapsto T(U)$ forgets phases. Composing two evolutions gives U_2U_1, whereas composing the projected evolutions gives the matrix product $T(U_2)T(U_1)$, and the two need not agree.

Proposition 7.1 (The interference expansion)**.** *For unitary matrices U_1, U_2,*

$$T(U_2U_1)_{ik} = \sum_j T(U_2)_{ij}T(U_1)_{jk} \;+\; \sum_{j\neq l}(U_2)_{ij}\overline{(U_2)_{il}}\,(U_1)_{jk}\overline{(U_1)_{lk}}.$$

Proof. Expand $|\sum_j (U_2)_{ij}(U_1)_{jk}|^2$ as a double sum over j, l and separate the diagonal terms $j = l$, which are $|(U_2)_{ij}|^2|(U_1)_{jk}|^2$, from the off-diagonal terms. □

The first sum is the classical composition and the second collects the interference terms, which the product of projected matrices discards.

Proposition 7.2 (Failure of composition)**.** *There exist unitary matrices U_1, U_2 with $T(U_2U_1) \neq T(U_2)T(U_1)$.*

Proof. Let $H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and take $U_1 = U_2 = H$. Then $H^2 = I$, so $T(H^2) = I$. Every entry of H has squared modulus $\frac{1}{2}$, so $T(H) = J := \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, and $J^2 = J \neq I$. □

The same phenomenon occurs in every dimension $d \geq 2$ for which a Fourier matrix is available. For the 3×3 discrete Fourier matrix $F_{jk} = \omega^{jk}/\sqrt{3}$ with $\omega = e^{2\pi i/3}$, one has $(F^2)_{jk} = \delta_{j,-k}$, a permutation matrix, so $T(F^2) = F^2$, while $T(F)$ is the uniform matrix with every entry $\frac{1}{3}$ and $T(F)^2 = T(F)$. A deterministic outcome at the amplitude level becomes maximal uncertainty at the probability level.

The physical content of the calculation needs equal precision. The product $T(U_2)T(U_1)$ is not an arbitrary wrong answer; it is the correct description of a different experiment.

Proposition 7.3 (Probability-level composition is composition with an intervening measurement)**.** *Let Δ be the channel that measures in the computational basis and discards the outcome, $\Delta(\rho) = \sum_k |k\rangle\langle k|\rho|k\rangle\langle k|$. Then the probability of outcome i for the sequence "prepare $|j\rangle$, apply U_1, apply Δ, apply U_2, measure in the computational basis" is $\big(T(U_2)T(U_1)\big)_{ij}$.*

Proof. The state after U_1 is $U_1|j\rangle$, and Δ maps it to $\sum_k |(U_1)_{kj}|^2 \, |k\rangle\langle k|$. Applying U_2 and measuring gives $\sum_k |(U_2)_{ik}|^2 |(U_1)_{kj}|^2$. □

Coherent evolution through two Hadamard operations therefore differs from evolution with an intervening discarded measurement, and the matrix J describes the latter. The failure of Proposition 7.2 is the difference between two interventions on the same system.

Nonmultiplicativity of T shows that one proposed composition rule fails. A stronger statement is that the projected matrices cannot determine the composite probabilities by *any* rule, and this has its own counterexample.

Proposition 7.4 (No rule recovers composites from projections)**.** *Let $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Then $T(I) = T(Z) = I$, while $T(HIH) = I$ and*

$$T(HZH) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Proof. The first statement is immediate. Since $H^2 = I$, $T(HIH) = T(I) = I$. A direct computation gives $HZH = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, a permutation matrix. □

Two middle operations with identical projected description therefore produce different observable outcomes when placed between the same preparation and recombination operations. The phase difference between I and Z is invisible to any isolated measurement in the computational basis and becomes visible in a larger experimental context. This is the quantum instance of the insufficiency defined in Chapter 2, and the remaining chapters of this part make it into a structured theory.

Chapter 8

Support Sufficiency and Probabilistic Sufficiency

The history structures of Part I assigned each history a *set* of admissible continuations. That is an assertion about what can happen, and it ignores how likely each possibility is. Quantum theory, thermodynamics, and every stochastic model require the stronger notion.

Definition 8.1 (Probabilistic history structure)**.** A *probabilistic history structure* is a history structure $(\mathcal{H}, \mathcal{F}, \mathcal{A})$ together with a map $P\colon \mathcal{H} \to \mathrm{Prob}(\mathcal{F})$ assigning to each history a probability law on $\mathcal{F}$ whose support is contained in $\mathcal{A}(h)$. For a law on infinite words the support is taken to be the set of infinite words every finite prefix of which has positive probability, since an individual infinite word generally has probability zero. An observation π is *support sufficient* if it is sufficient in the sense of Part I, and *probabilistically sufficient* if

$$\pi(h_1) = \pi(h_2) \implies P(h_1) = P(h_2).$$

Proposition 8.2. *Every probabilistically sufficient observation is support sufficient when $\mathcal{A}(h)$ is taken to be the support of $P(h)$. The converse fails.*

Proof. Equal laws have equal supports. For the converse, let $\mathcal{H} = \{h_1, h_2\}$ and $\mathcal{F} = \{0, 1\}$, with $P(h_1)$ giving probability $\frac{1}{2}$ to each outcome and $P(h_2)$ giving probability $\frac{1}{4}$ and $\frac{3}{4}$. Both histories admit exactly the outcomes 0 and 1, so the trivial one-point observation is support sufficient but not probabilistically sufficient. □

The minimal construction of Chapter 3 repeats with P in place of $\mathcal{A}$. The equivalence $h_1 \approx h_2$ if and only if $P(h_1) = P(h_2)$ yields a minimal probabilistically sufficient observation, and Theorem 3.2 holds verbatim for it. Since equality of laws implies equality of supports, the probabilistic equivalence refines the support equivalence. The probabilistic quotient

is therefore at least as large as the support quotient, and the difference between them measures how much structure must be added to the state when one passes from asking what can happen to asking how often.

For dynamics, the compatibility needed in Theorem 6.2 acquires a probabilistic form. The identity $P(h \cdot s) = P(h)(\cdot \mid s)$, conditioning on the first step, is the statement that the law of the future after $h \cdot s$ is the conditional law of the future after h. Under that assumption the probabilistic quotient also carries a well-defined dynamics, now a stochastic transition rule on the quotient, which is the structure that computational mechanics calls the causal-state process [6, 39].

Chapter 9

The Quantum State as a Minimal Probabilistic Quotient

The quantum case admits an exact answer to the question of what the minimal probabilistically sufficient observation is. Fix a Hilbert space $\mathbb{C}^d$ with computational basis $\{e_i\}$ and an initial index j. Let histories be finite sequences of unitary gates, with $h = (U_1, \dots, U_n)$ denoting the product $U_h = U_n \cdots U_1$. Let a future consist of a further unitary V followed by a measurement in the computational basis, so that $P(h)$ is the family of outcome laws indexed by $V \in U(d)$, assigning to the outcome i after the gate V the probability

$$P(h)(V, i) = |(VU_h)_{ij}|^2.$$

Here the futures are allowed to use any gate before measuring, which models the freedom to choose the measurement context.

Theorem 9.1 (Quantum state as minimal quotient)**.** *Two histories h_1, h_2 satisfy $P(h_1) = P(h_2)$ if and only if the vectors $u_1 = U_{h_1} e_j$ and $u_2 = U_{h_2} e_j$ are equal up to a global phase. The minimal probabilistically sufficient observation is therefore the map $h \mapsto [U_h e_j]$ into the projective space $\mathbb{CP}^{d-1}$.*

Proof. If $u_2 = e^{i\varphi} u_1$ the two laws coincide, since probabilities depend on moduli squared. Conversely, suppose the laws coincide. Then $|\langle e_i | V u_1 \rangle|^2 = |\langle e_i | V u_2 \rangle|^2$ for all $V \in U(d)$ and all i. Writing $w = V^\dagger e_i$, which ranges over all unit vectors as V and i vary, this says $|\langle w | u_1 \rangle| = |\langle w | u_2 \rangle|$ for every unit vector w. Taking $w = u_1$ gives $|\langle u_1 | u_2 \rangle| = 1$, and equality in the Cauchy–Schwarz inequality forces u_2 to be a scalar multiple of u_1 of unit modulus. □

The theorem identifies the quantum state, as a ray, with the equivalence class of histories defined by the continuation structure alone, once the futures are rich enough to include

measurements in every basis. Within the assumed quantum model, which already supplies the Hilbert space, the unitary gates, and the Born probabilities, the state need not be posited separately: it is recovered as the quotient. The theorem does not derive those ingredients from anything more primitive, and a reconstruction of quantum theory from weaker assumptions is a separate task. The projected description, which retains only the vector of probabilities $p(h) = (|(U_h)_{ij}|^2)_i$ in the fixed basis, is a far coarser observation.

Proposition 9.2 (Fibres of the probability projection)**.** *The map $[u] \mapsto (|u_i|^2)_i$ from $\mathbb{CP}^{d-1}$ to the probability simplex is surjective. Over a point of the open simplex its fibre is the torus T^{d-1} of relative phases. Hence the projected description is insufficient for the futures of Theorem 9.1 for every $d \geq 2$, and restricting the futures to measurements in the computational basis alone, that is V a permutation matrix, makes it sufficient.*

Proof. Given probabilities p_i the vectors $u_i = \sqrt{p_i}e^{i\theta_i}$ realize them, and rays correspond to phase tuples (θ_i) modulo a common shift, which is a torus of dimension $d-1$ when all $p_i > 0$. Insufficiency follows from Theorem 9.1, because distinct rays are separated by some choice of V. If V ranges over permutations only, $|(VU_h)_{ij}|^2 = p_{\sigma(i)}$, which is determined by $p(h)$. □

The real case is smaller and instructive. If amplitudes are restricted to real orthogonal matrices, rays lie in $\mathbb{RP}^{d-1}$ and the fibre over an interior point of the simplex consists of the sign patterns modulo a global sign, a finite set of 2^{d-1} points. For $d = 2$ the projection forgets exactly one bit, the relative sign of the two amplitudes. This is the quantitative form of the statement that a real unistochastic description loses a discrete invariant where the complex one loses a continuous one, and it reappears in the gluing theorem of Chapter 11.

Chapter 10

Sufficiency for Coherent Composition

The previous chapter asked what must be remembered about a history in order to predict outcomes of later measurements. A different question concerns *operations*: what must be retained about a transformation U in order to know the description of any experiment that places U inside a larger one. The relevant sufficiency is a congruence property.

Definition 10.1 (Compositional sufficiency)**.** Let G be a group (or monoid) of transformations and $q\colon G \to Q$ a map. Then q is *compositionally sufficient* if $q(g) = q(g')$ implies $q(agb) = q(ag'b)$ for all $a, b \in G$, equivalently if the kernel of q is a two-sided congruence.

For $q = T$ on $G = U(d)$ this fails by Proposition 7.4. As in Theorem 3.2, there is a canonical remedy, the largest congruence contained in the kernel of T. It is the syntactic congruence: $U \equiv U'$ if and only if $T(AUB) = T(AU'B)$ for all $A, B \in U(d)$. It is the coarsest equivalence that remembers enough of U to determine the projected description of every experiment in which U is embedded.

Theorem 10.2 (The syntactic congruence of the probability projection)**.** *For $d \geq 2$, $U \equiv U'$ if and only if $U' = e^{i\varphi}U$ for some real φ. The minimal compositionally sufficient quotient of T on $U(d)$ is therefore the projective unitary group $PU(d) = U(d)/U(1)$.*

Proof. If $U' = e^{i\varphi}U$ then $T(AU'B) = T(AUB)$ because moduli are unchanged. Conversely, suppose $T(AUB) = T(AU'B)$ for all A, B. For a unit vector x choose B with $Be_1 = x$, and for a unit vector w choose A with $A^\dagger e_1 = w$. Then $T(AUB)_{11} = |\langle w|Ux\rangle|^2$, and the hypothesis gives $|\langle w|Ux\rangle| = |\langle w|U'x\rangle|$ for all unit vectors w and x. As in the proof of Theorem 9.1, this forces $U'x = c(x)Ux$ with $|c(x)| = 1$ for each x. Hence $W = U^\dagger U'$ maps every vector to a multiple of itself, so every nonzero vector is an eigenvector of W.

If $We_i = \lambda_i e_i$ and $We_j = \lambda_j e_j$, then applying the same property to $e_i + e_j$, which must also be an eigenvector, forces $\lambda_i = \lambda_j$. Hence W is a scalar, of unit modulus because W is unitary. □

The theorem corrects a loose reading of the slogan that composition requires "restoring the phases." A global phase is irrelevant to every sequential context of the form AUB considered here, and need not be restored; the theorem concerns such contexts and does not address a coherent controlled implementation of an unknown operation, where a global phase can become relative. The exact conclusion is that any two unitaries not related by a global phase are distinguishable in some allowed context. The retained structure is exactly $PU(d)$: strictly larger than the doubly stochastic matrix $T(U)$, which is not compositionally sufficient, and strictly smaller than U itself.

The orthostochastic case is analogous. For real orthogonal matrices the same argument gives the congruence $O \equiv \pm O$, and the minimal quotient is $PO(d) = O(d)/\{\pm I\}$.

Together with Chapter 9, this exhibits the two different senses in which T is insufficient. For histories, insufficiency is the failure to determine outcome laws of later measurements, and its repair is the ray. For operations, insufficiency is the failure to determine the description of composites, and its repair is the projective unitary. The two repairs are linked by the fact that a history is a product of operations, and the ray of a history is the image of the projective element under evaluation at the initial vector.

Chapter 11

Context Graphs and the Gluing Obstruction

Compositional sufficiency concerns a single chosen composition. A separate question concerns whether many individually admissible observations can be joined into one consistent description. This is the second obstruction in the author's reconstruction [7]: the two questions are logically independent, and the first is meaningful on an open path where no cycle occurs.

Let $G = (V, E)$ be a connected graph whose vertices are measurement contexts, that is orthonormal bases, and whose edges $e = (A, B)$ carry observed doubly stochastic matrices B_e of transition probabilities between the bases. The graph is *globally realizable* over a field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ if there is a single assignment of frames to vertices whose relative transition probabilities reproduce every B_e. This is a continuation problem in the sense of Part I: local data, each edge admissible by itself, are required to extend to a global object.

Proposition 11.1 (Trees carry no constraint)**.** *If each edge of a tree is individually admissible over $\mathbb{F}$, a global frame assignment reproducing every edge exists.*

Proof. Fix a root and propagate along the unique path to each vertex, composing the chosen lifts. Uniqueness of paths in a tree means no two propagations can disagree. □

Theorem 11.2 (Realizability criterion)**.** *A connected graph G with spanning tree T is globally realizable if and only if there is one choice of admissible lifts on the edges of T whose propagated frame assignment also reproduces every non-tree edge. There are $\beta_1(G) = |E| - |V| + 1$ non-tree-edge compatibility tests, all of which must hold for the same choice of tree lifts.*

The proof is the propagation argument of Proposition 11.1 together with the observation that any global assignment restricts to a choice of tree lifts. The theorem is existential and

not constructive. The cycle rank counts the tests in this construction; it does not by itself establish their functional independence or the codimension of the realizable data.

The smallest graph with a nontrivial constraint is the triangle on three contexts Z, X, Y, with $\beta_1 = 1$. Over $\mathbb{R}$ and for 2×2 matrices the constraint can be written explicitly, and here the labelling of outcomes matters. A real 2×2 bistochastic matrix has the form

$$B_{AB} = \begin{pmatrix} p_{AB} & 1 - p_{AB} \\ 1 - p_{AB} & p_{AB} \end{pmatrix}, \qquad \alpha_{AB} = \arccos \sqrt{p_{AB}} \in [0, \pi/2].$$

Realization is required to reproduce each B_{AB} exactly, entry by entry, with the first labelled vector of each basis (the first outcome) tracked consistently across all edges at that vertex.

Theorem 11.3 (Triangle gluing in $\mathbb{R}^2$, labelled form)**.** *Three real 2×2 bistochastic matrices B_{ZX}, B_{ZY}, B_{XY} are realized by a single assignment of real orthonormal bases to Z, X, Y if and only if*

$$p_{XY} = \cos^2\!\big(\varepsilon\, \alpha_{ZX} - \eta\, \alpha_{ZY}\big) \quad \textit{for some } \varepsilon, \eta \in \{-1, 1\}.$$

Proof. Necessity: place the first vector of Z at angle 0. The first vector of X then sits at an angle congruent to $\pm\alpha_{ZX}$ modulo π, since $\cos^2$ of that angle must equal p_{ZX}, and likewise the first vector of Y at an angle congruent to $\pm\alpha_{ZY}$. The entry p_{XY} is $\cos^2$ of the difference, which is unchanged by shifts of π, and so has the stated form. Sufficiency: given signs ε, η satisfying the equation, place the first vectors of Z, X, Y at angles 0, $\varepsilon\alpha_{ZX}$, $\eta\alpha_{ZY}$, and complete each to an orthonormal basis by the orthogonal partner. The entries p_{ZX}, p_{ZY} are reproduced by construction, p_{XY} by hypothesis, and the off-diagonal entries are then forced by bistochasticity. □

The criterion remains a theorem if one instead gives edge data only up to independent permutations of outcomes on each edge, in which case the relevant invariant is the canonical angle $\theta \in [0, \pi/4]$ and the condition is $\theta_{XY} \equiv \pm\theta_{ZX} \pm \theta_{ZY} \pmod{\pi/2}$. That is a different realization problem, because independent permutations on different edges need not come from a consistent relabelling at each vertex. For example, take $B_{ZX} = B_{ZY} = I$ and $B_{XY} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$: all three canonical angles vanish, yet the labelled problem has no solution, since the first two matrices force the labelled vectors of X and Y to coincide with those of Z and hence with each other.

The proof exhibits the fibre of the projection in the real case: the data p_{ZX}, p_{ZY} leave undetermined exactly one relative sign, and p_{XY} can take one of two values. This is the one bit of Chapter 9 for $d = 2$, now seen as a constraint on a cycle.

Corollary 11.4 (No three real mutually unbiased bases in the plane)**.** *There do not exist three pairwise mutually unbiased orthonormal bases of $\mathbb{R}^2$.*

Proof. Mutual unbiasedness requires $p_{ZX} = p_{ZY} = p_{XY} = \frac{1}{2}$, so $\alpha_{ZX} = \alpha_{ZY} = \pi/4$. The criterion would require $\frac{1}{2} = \cos^2(\pm\pi/4 \mp \pi/4)$ for some signs, but the right side takes only the values 1 and 0. The same argument applies to the canonical-angle formulation, since the unbiased matrix is invariant under outcome permutations. □

Within the real two-dimensional setting, for finite simple graphs whose edge parameters are sampled independently with densities, a cycle-closing edge is confined to finitely many values once the other edges of the cycle are fixed, so global realizability on a graph containing a cycle is a measure-zero condition. The statement does not extend automatically to complex frames or to higher dimensions. What is special about mutual unbiasedness is its symmetry and not its scarcity. Over $\mathbb{C}$ the triangle of mutually unbiased bases exists in dimension two, so the obstruction distinguishes the fields $\mathbb{R}$ and $\mathbb{C}$ at the level of consistent families of observations, a fact the monograph develops through the composite-system analysis that this book treats later.

The two obstructions of this part can now be compared. Compositional sufficiency asks whether projected data determine the projected data of a composite, and its answer is the projective unitary group. Global realizability asks whether independently admissible local data extend to a global assignment, and its answer is governed by the cycle rank. Both are failures of an observation to carry enough structure, and Part VIII returns to the question of whether they share a common mathematical form.

Chapter 12

Coarse-Graining and Effective Stochastic Dynamics

An observation that discards information can still support a dynamics of its own. Whether it does is a separate question from whether it is sufficient, and the two are easily conflated. This chapter separates them in the setting of Part I and then specializes to Markov chains, where the criterion becomes a linear-algebraic identity.

Recall that for an extendable history structure, an observation $\pi\colon \mathcal{H} \to \mathcal{X}$ is sufficient when it determines the continuation sets. It is *closed* if it is compatible with extension.

Definition 12.1 (Closed observation)**.** An observation π of an extendable history structure is *closed* if $\pi(h) = \pi(k)$, together with the existence of $h\cdot s$ and $k\cdot s$, implies $\pi(h\cdot s) = \pi(k\cdot s)$. A closed observation carries a transition relation $x \xrightarrow{s} x'$ on its values, which is the effective dynamics of the description.

Proposition 12.2 (Sufficiency and closure are independent)**.** *A sufficient observation need not be closed, and a closed observation need not be sufficient. The observation $\pi_{\mathcal{A}}$ is both, and it is the coarsest observation that is both.*

Proof. A constant observation is closed and, whenever $\mathcal{A}$ is not constant, not sufficient. For the other direction, let $\mathcal{H}$ be the set of all finite binary words, with $\mathcal{A}(h) = \mathcal{F}$ for every h, so that every observation is sufficient. Define $\pi(\varepsilon) = \pi(0) = 1$, $\pi(1) = 0$, $\pi(00) = 0$, and $\pi(h) = 0$ for all other h of length at least two; the value assigned to $\pi(1)$ is immaterial to the argument. Then $\pi(\varepsilon) = \pi(0)$ while $\pi(\varepsilon \cdot 0) = \pi(0) = 1$ and $\pi(0 \cdot 0) = \pi(00) = 0$, so π is sufficient but not closed. That $\pi_{\mathcal{A}}$ is closed and sufficient is Theorem 6.2, and that it is coarser than every sufficient observation is Theorem 3.2. □

The proposition says that an effective dynamics at the level of a given description exists exactly when that description is closed, and that being closed is not the same as being

predictive. Descriptions that are closed but insufficient give a self-contained dynamics that nonetheless fails to determine the continuations of the underlying system, and descriptions that are sufficient but not closed determine the continuations at each moment without supplying a rule for updating themselves. The minimal state is the only canonical object having both properties.

For Markov chains the criterion for closure takes a precise linear form. Let X be a finite set and M a column-stochastic matrix on $\mathbb{R}^X$, so that a distribution p evolves by $p \mapsto Mp$. Let $c\colon X \to Y$ be a surjection, and let C be the $|Y| \times |X|$ aggregation matrix with $C_{yx} = 1$ if $c(x) = y$ and 0 otherwise. Then Cp is the distribution of the coarse variable.

Theorem 12.3 (Lumpability)**.** *The following are equivalent. First, the coarse distribution one step later depends on the initial distribution only through the initial coarse distribution: $Cp = Cp'$ implies $CMp = CMp'$. Second, there is a matrix N with $CM = NC$. Third, for all blocks $A, B \subseteq X$ in the partition induced by c, the sum $\sum_{i \in B} M_{ij}$ takes the same value for every $j \in A$. In that case N is column stochastic with N_{BA} equal to this common value, and the same statement holds for every power of M.*

Proof. If $CM = NC$, then $Cp = Cp'$ gives $CMp = NCp = NCp' = CMp'$. Conversely, assume the first condition and let $v \in \ker C$. Then v has total sum zero, so $v = m(p - p')$ with $m \geq 0$ and p, p' distributions obtained from the positive and negative parts of v, and $Cp = Cp'$ because $Cv = 0$. The first condition gives $CMv = 0$. Hence $\ker C \subseteq \ker CM$, and since C is surjective there is a linear N with $CM = NC$. The third condition is the entrywise form of $CM = NC$ applied to the basis vectors: $(CM)_{Bj} = \sum_{i \in B} M_{ij}$ must equal $N_{B,c(j)}$. Stochasticity of N follows because CMp is a distribution for every distribution p. Finally $CM^n = N^n C$ by induction. □

Lumpability is the Markov-chain form of closure [34]: the aggregated process is again Markov, with transition matrix N, for every initial distribution.

Example 12.4 (A unistochastic chain that is not lumpable)**.** Let $\theta \in (0, \pi/2)$ and $U = R(\theta) \oplus 1$, the rotation of the first two basis vectors with the third left fixed, so that

$$T(U) = \begin{pmatrix} c & s & 0 \\ s & c & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad c = \cos^2\theta,\ s = \sin^2\theta.$$

Partition the three states into $A = \{0\}$ and $B = \{1, 2\}$. The sums of the columns of $T(U)$ over the rows in B are s for column 0, c for column 1, and 1 for column 2. The columns 1 and 2 both belong to block B but give different sums unless $c = 1$, so the partition is not lumpable.

On this example the coarse process (A versus B) cannot be written as a Markov chain on two states, because the probability of leaving B depends on whether the system sits in state 1 or state 2. A description that tracks only the block therefore has no dynamics of its own, and any effective description must either refine the partition or introduce memory. The next chapter constructs the memory explicitly.

A contrasting case shows the criterion can hold for a nontrivial unitary. Let F be the 3×3 Fourier matrix, for which $T(F)$ has every entry equal to $\frac{1}{3}$. For every partition, the sum over a block B of any column equals $|B|/3$, which does not depend on the column, so every partition is lumpable with $N_{BA} = |B|/3$. In that case the coarse process forgets its initial block entirely after one step.

Chapter 13

Memory Kernels and the Projection Operator Method

When a coarse description is not closed, the dynamics of the observed variables can still be written exactly, at the price of a memory term. This is the content of the projection operator method of Zwanzig and Mori [40, 35]. The finite-dimensional discrete-time form is elementary and suffices for the purposes of this book.

Let V be a finite-dimensional vector space with a linear map $M\colon V \to V$, and let $V = W \oplus W'$ be a decomposition with associated projections P onto W and $Q = 1 - P$ onto W'. Write M in block form with respect to this decomposition, with blocks $M_{PP}, M_{PQ}, M_{QP}, M_{QQ}$, and let $v_t = M^t v_0$, $x_t = P v_t$, $y_t = Q v_t$.

Theorem 13.1 (Discrete memory-kernel equation)**.** *For all $t \geq 0$,*

$$x_{t+1} = M_{PP}\, x_t + \sum_{s=0}^{t-1} K_{t-s}\, x_s + M_{PQ} M_{QQ}^{t}\, y_0, \qquad K_m = M_{PQ} M_{QQ}^{m-1} M_{QP} \quad (m \geq 1).$$

Proof. The block form gives $x_{t+1} = M_{PP}x_t + M_{PQ}y_t$ and $y_{t+1} = M_{QP}x_t + M_{QQ}y_t$. Solving the second recurrence gives $y_t = M_{QQ}^t y_0 + \sum_{s=0}^{t-1} M_{QQ}^{t-1-s} M_{QP} x_s$, which can be checked by induction on t. Substituting into the first recurrence and setting $m = t - s$ gives the stated equation. □

The theorem splits the observed dynamics into three parts: an instantaneous term $M_{PP}x_t$, a memory term whose kernel K_m records the effect of the observed state m steps ago passing through the unobserved subspace, and an inhomogeneous term depending on the unobserved part of the initial state. If the initial state lies in W, that is $y_0 = 0$, the last term vanishes.

Corollary 13.2 (Closure and memory)**.** *Suppose $y_0 = 0$. The observed dynamics is closed, in the sense that $x_{t+1} = M_{PP}x_t$ for all t and all $x_0 \in W$, if and only if $K_m = 0$ for all $m \geq 1$. If $M_{QQ}^r = 0$ for some r, then $K_m = 0$ for all $m \geq r+1$, so the recurrence for x_{t+1} involves x_t together with $x_{t-1}, \ldots, x_{t-r}$, a window of $r+1$ observed values, and for $y_0 \neq 0$ the inhomogeneous term $M_{PQ}M_{QQ}^t y_0$ vanishes for $t \geq r$. This is a* kernel lag *of r. It is not by itself the memory depth of Chapter 4, which concerns sufficiency for complete continuation sets, and any relation between the two would have to be established separately.*

Proof. If all K_m vanish the theorem gives closure. Conversely, closure for all $x_0 \in W$ means $x_{t+1} = M_{PP}x_t$, so the memory sum $\sum_{s=0}^{t-1} K_{t-s}x_s$ vanishes for every x_0 and every t. For $t = 1$ this reads $K_1 x_0 = 0$, so $K_1 = 0$. Suppose $K_1 = \cdots = K_{t-1} = 0$. The sum at time t then reduces to $K_t x_0$, which must vanish for all x_0, so $K_t = 0$. The finite-depth statement follows from $K_m = M_{PQ}M_{QQ}^{m-1}M_{QP}$ for $m - 1 \geq r$. □

The unitary dynamics of Chapter 7 fits this framework through the Liouville representation. Let V be the space of Hermitian matrices on $\mathbb{C}^d$ and $M = \mathcal{L}_U$ the map $\rho \mapsto U\rho U^\dagger$. Let $P = \Delta$ be the projection onto diagonal matrices, whose entries are the populations in the computational basis, and $Q = 1 - \Delta$ the projection onto coherences. For an initial state that is diagonal, $y_0 = 0$, and the first block is exactly the projected matrix:

$$M_{PP} = T(U).$$

This identification shows that the stochastic matrix $T(U)$ is the instantaneous part of the exact dynamics of the populations, and everything the probability description forgets is contained in the kernel.

Theorem 13.3 (First memory kernel of coherent evolution)**.** *For the Liouville dynamics $\rho \mapsto U\rho U^\dagger$ with diagonal initial state, the first kernel is*

$$K_1 = M_{PQ}M_{QP} = T(U^2) - T(U)^2.$$

Proof. The exact populations after two steps are $x_2 = T(U^2)x_0$. Theorem 13.1 with $y_0 = 0$ gives $x_1 = T(U)x_0$ and $x_2 = T(U)x_1 + K_1 x_0 = T(U)^2 x_0 + K_1 x_0$. Equating and using that x_0 is arbitrary gives the identity. □

In terms of Chapter 7, K_1 is the interference sum of Proposition 7.1 with $U_1 = U_2 = U$. A numerical check on a random 3×3 unitary reproduces $T(U^2) - T(U)^2$ from the blocks $M_{PQ}M_{QP}$ to machine precision, and $M_{PP} = T(U)$ to the same precision.

Example 13.4 (The Hadamard kernel)**.** For $U = H$ one has $T(H) = J$, $T(H^2) = I$, and therefore $K_1 = I - J = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Since $H^2 = I$, the populations satisfy $x_t = x_0$ for even t and $x_t = Jx_0$ for odd t. The kernel equation at $t = 2$ reads $x_3 = Jx_2 + K_1x_1 + K_2x_0$, that is $Jx_0 = Jx_0 + (I - J)Jx_0 + K_2x_0$, and $(I - J)J = 0$, so $K_2 = 0$. The block M_{QQ} of the Liouville map of H on the coherences is $-\frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$, which satisfies $M_{QQ}^2 = -M_{QQ}$, so $K_m = (-1)^m K_2 = 0$ for all $m \geq 2$. This is consistent with the periodic orbit: for even t one finds $Jx_t + K_1x_{t-1} = Jx_0 + (I - J)Jx_0 = Jx_0$, and for odd t one finds $Jx_t + K_1x_{t-1} = J^2x_0 + (I - J)x_0 = x_0$, because $J^2 = J$. A direct numerical computation of $M_{PQ}M_{QQ}^{m-1}M_{QP}$ for $m = 1, \ldots, 5$ agrees. The Hadamard evolution therefore has kernel lag one: for $t \geq 1$ the populations obey the second-order recurrence $x_{t+1} = Jx_t + (I - J)x_{t-1}$, which must be started from the explicit first step $x_1 = Jx_0$ with x_0 arbitrary, because for $t = 0$ there is no earlier population and only M_{PP} acts.

The example shows exactly what the projected description is missing. The matrix $T(H) = J$ would predict $x_2 = Jx_1 = Jx_0$, but the true value is x_0, and the difference is the single kernel term K_1x_0 acting on the population two steps back. The kernel lag, the number of earlier observed populations that enter the recurrence, is one here, while the minimal sufficient state of Chapter 9 is the ray. The two statements concern different things, since the kernel lag counts terms of a recurrence for the populations and the minimal state concerns sufficiency for complete continuation sets, and the kernel represents the effect of the unobserved coherences through the observed history.

Chapter 14

Consistent Histories and Decoherence Conditions

The consistent histories approach of Griffiths [29] and the logical analysis of Isham [30] ask the question of this part in a different vocabulary: when can probabilities be assigned to sequences of events so that they satisfy the sum rules of classical probability. The answer is a condition on the *interference between histories*, and it is the condition under which probability-level composition is legitimate.

Fix a basis and a sequence of unitaries $U_1, \dots, U_n$, and consider histories $\alpha = (k_1, \dots, k_n)$ of outcomes of measurements in the basis after each step, starting from the basis state $|j\rangle$. The amplitude of the history is

$$A_\alpha = (U_n)_{k_n k_{n-1}} \cdots (U_2)_{k_2 k_1} (U_1)_{k_1 j},$$

and the branch vector of a history $\alpha = (k_1, \dots, k_n)$ is $|\psi_\alpha\rangle = A_\alpha |k_n\rangle$. The decoherence functional for pure initial states and sharp rank-one projectors is therefore

$$D(\alpha, \beta) = \langle \psi_\beta | \psi_\alpha \rangle = A_\alpha \overline{A_\beta}\, \delta_{k_n, k'_n}.$$

Two histories with different final outcomes have orthogonal branch vectors, so $D(\alpha, \beta) = 0$ automatically when $k_n \neq k'_n$. A family is *weakly consistent* if $\operatorname{Re} D(\alpha, \beta) = 0$ for all $\alpha \neq \beta$, and *medium decoherent* if $D(\alpha, \beta) = 0$ for all $\alpha \neq \beta$.

Theorem 14.1 (Composition is a decoherence condition)**.** *For two steps and fixed initial index j and final index i,*

$$T(U_2 U_1)_{ij} - \big(T(U_2) T(U_1)\big)_{ij} = \sum_{k \neq l} D\big((k,i),(l,i)\big) = 2 \sum_{k<l} \operatorname{Re} D\big((k,i),(l,i)\big).$$

Hence probability-level composition holds at the entry (i, j) if and only if the real parts of the decoherence functional between the histories with these endpoints sum to zero. It holds for all entries if the family of two-step histories is weakly consistent, in particular if it is medium decoherent.

Proof. The amplitude for the composite transition is $\sum_k A_{(k,i)}$, so $T(U_2U_1)_{ij} = |\sum_k A_{(k,i)}|^2 = \sum_k |A_{(k,i)}|^2 + \sum_{k\neq l} A_{(k,i)}\overline{A_{(l,i)}}$. The first sum is $\sum_k T(U_2)_{ik}T(U_1)_{kj}$. The second sum is real, because interchanging k and l replaces each term by its complex conjugate, which also gives the second form of the right-hand side. □

The theorem identifies the interference term of Proposition 7.1 with the sum of off-diagonal entries of the decoherence functional. Composition at a given entry requires only that the real parts sum to zero, which is weaker than weak consistency of the whole family.

Example 14.2 (Weakly consistent but not medium decoherent)**.** Let $U_1 = H$ and $U_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$. This U_2 is unitary, and $U_2U_1 = \frac{1}{2}\begin{pmatrix} 1+i & 1-i \\ 1+i & i-1 \end{pmatrix}$, all of whose entries have squared modulus $\frac{1}{2}$. Hence $T(U_2U_1) = J = T(U_2)T(U_1)$, and composition holds, although neither matrix is a permutation. For $i = j = 0$ the two histories have amplitudes $A_0 = \frac{1}{2}$ and $A_1 = \frac{i}{2}$, so $D = A_0\overline{A_1} = -\frac{i}{4}$, which is nonzero but purely imaginary. The family is weakly consistent and not medium decoherent.

The failure of composition for the pair (H, H) in Proposition 7.2 and its success for the pair (H, U_2) in this example show that composition of probabilities is a property of the pair of operations together with the initial and final contexts, and not of the stochastic matrices alone. Proposition 7.4 established that no assignment of composites to projections can be correct for all pairs. The decoherence functional supplies, case by case, the missing condition.

Proposition 14.3 (Decoherence by an intervening channel with a record)**.** *Implement the dephasing channel Δ of Proposition 7.3 between consecutive steps by a unitary dilation that copies the outcome into an environment record, $|k\rangle|e_0\rangle \mapsto |k\rangle|e_k\rangle$ with orthonormal records e_k. In the dilated description of system and records, the probability of the history $(k_1, \ldots, k_n)$ is $\prod_t |(U_t)_{k_t k_{t-1}}|^2$ with $k_0 = j$, the family of histories is medium decoherent, and the probabilities of coarse-grained histories are obtained by summing, so that the dynamics of the populations is the Markov chain with transition matrices $T(U_t)$.*

Proof. The branch vector of the history α in the dilated description is $A_\alpha|k_n\rangle|e_{k_1}\cdots e_{k_{n-1}}\rangle$, with $A_\alpha = \prod_t (U_t)_{k_t k_{t-1}}$. Two distinct histories either differ in the final outcome, in which case the system vectors are orthogonal, or agree in the final outcome and differ in some

earlier outcome, in which case the record vectors are orthogonal. In both cases $D(\alpha, \beta) = 0$. The probability of α is $|A_\alpha|^2$, the product of the squared moduli. Summation over intermediate outcomes gives the product of the matrices $T(U_t)$. □

The record is needed for the medium-decoherence claim. Without it the system's final states alone need not make distinct histories orthogonal: two histories that agree in the final outcome and differ earlier have the same system vector, and their cross term is the interference term of Theorem 14.1. The reduced state of the system after the dilated evolution is the dephased state, so the populations are those of the channel Δ of Proposition 7.3.

The proposition completes the picture of Chapter 7. The classical composition $T(U_2)T(U_1)$ is the correct description of coherent evolution under exactly the conditions in which the decoherence functional has vanishing real part, and of evolution with an intervening dephasing channel in all cases. What the projection forgets is therefore the information needed to decide, for a given pair of operations and a given choice of contexts, which of these two situations holds.

Chapter 15

Which Structures Survive Projection

The preceding chapters show what projection loses. This chapter records what it keeps, so that the passage from amplitudes to probabilities can be assessed as an observation with a definite kernel and a definite image.

Proposition 15.1 (Gauge invariance of the projection)**.** *For diagonal unitary matrices D_1, D_2 and any unitary U, $T(D_1UD_2) = T(U)$. If all entries of U are nonzero, the orbit $\{D_1UD_2\}$ is a torus of dimension $2d-1$ contained in a single fibre of T.*

Proof. Multiplication by D_1 and D_2 multiplies the entry U_{ij} by a phase, which does not change its modulus. The torus $T^d \times T^d$ acts on $U(d)$ by $(D_1, D_2) \cdot U = D_1UD_2$, and $D_1UD_2 = U$ with all $U_{ij} \neq 0$ forces $(D_1)_{ii}(D_2)_{jj} = 1$ for all i, j, so $D_1 = cI$ and $D_2 = \bar{c}I$ for a scalar phase c. The stabilizer is thus a circle and the orbit has dimension $2d-1$. □

A count of the dimensions of domain and target does not by itself give the dimension of a fibre. At $U = I$ the fibre $T^{-1}(I)$ is the set of diagonal unitary matrices, which has dimension d, smaller than $2d-1$ for $d \geq 2$. The gauge-orbit statement is therefore restricted to matrices with all entries nonzero, where the orbit has dimension $2d-1$. If in addition U is a regular point, in the sense that the differential of T at U, as a map into the $(d-1)^2$-dimensional affine hull of the doubly stochastic matrices, has rank $(d-1)^2$, then by the regular value theorem the fibre near U is a submanifold of dimension $d^2-(d-1)^2 = 2d-1$, and the gauge orbit, which lies in the fibre and has the same dimension, coincides with the fibre near U. Full dimension of the image alone is not enough without this regularity of the map at U.

For $d = 2$ the conclusion holds globally for an interior matrix. If $T(U) = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$

with $0 < p < 1$, write the entries as $\sqrt{p}\,e^{i\varphi_{11}}$, $\sqrt{1-p}\,e^{i\varphi_{12}}$, $\sqrt{1-p}\,e^{i\varphi_{21}}$, $\sqrt{p}\,e^{i\varphi_{22}}$. Orthogonality of the columns requires $\varphi_{11} + \varphi_{22} - \varphi_{12} - \varphi_{21} \equiv \pi \pmod{2\pi}$, and this is the only constraint; the matrices $D_1 R(\theta) D_2$ with $\cos^2\theta = p$ have phases $\alpha_i + \beta_j$ plus the pattern $(0, \pi; 0, 0)$ of $R(\theta)$ and satisfy it, and they exhaust the solutions because three of the four phases can be matched by choosing $\alpha_1, \alpha_2, \beta_1, \beta_2$ and the fourth is then fixed by the constraint. The complex unitary fibre is therefore a single left-right diagonal-phase orbit; a random sample of 200 unitaries confirms the factorization to 10^{-15}. The finite sign ambiguity found in Chapter 9 belongs to the real state projection $[u] \mapsto (|u_i|^2)$ and not to this statement about the fibres of T on $U(2)$.

Proposition 15.2 (Zero patterns survive)**.** *$T(U)_{ij} = 0$ if and only if $U_{ij} = 0$. Moreover, if $B = T(U)$ for a unitary U then no two columns of B have supports that intersect in exactly one row.*

Proof. The first claim is immediate from $T(U)_{ij} = |U_{ij}|^2$. For the second, distinct columns u, u' of a unitary are orthogonal, so $\sum_i \overline{u_i} u'_i = 0$. If the supports meet in exactly one row i_0, the sum reduces to the single term $\overline{u_{i_0}} u'_{i_0}$, which is nonzero. □

Corollary 15.3 (Not every doubly stochastic matrix is unistochastic)**.** *For $n = 3$ the doubly stochastic matrix $B = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ is not of the form $T(U)$.*

Proof. The supports of the first and second columns are the rows $\{1, 3\}$ and $\{1, 2\}$, which intersect in the single row 1. By Proposition 15.2 no unitary has $T(U) = B$. □

For $n = 2$ every doubly stochastic matrix is unistochastic, as noted in Chapter 17, so the corollary marks the dimension at which the image of the projection becomes a proper subset of the doubly stochastic matrices. This is a constraint on the observation that cannot be seen from the observation alone: it expresses the fact that probabilities arising from unitary dynamics satisfy relations the abstract notion of a reversible stochastic process does not.

The levels of description studied in this part can now be compared. Each is an observation of the underlying unitary dynamics, and each is characterized by what it determines.

The table summarizes the main result of Part II. The projection to transition probabilities keeps gauge-invariant information and the zero pattern and loses relative phases. What a theory must retain depends on the use made of it. To predict the next measurement in a fixed basis, $T(U)$ is sufficient. To predict measurements in arbitrary bases the ray is the minimal state. To predict the behaviour of an operation inside larger experiments the projective unitary is the minimal operation. To assemble many partial descriptions into

Description	Retains	Determines
$T(U)$	moduli of entries; zero pattern; gauge-invariant data	single-step transition probabilities in the fixed basis
Ray $[Ue_j]$	state up to global phase	outcome laws of every later measurement and gate
$PU(d)$ element	operation up to global phase	the projected description of every sequential composite AUB
Context graph data	edge matrices on a graph	global frame assignment only when cycle compatibilities hold

Table 15.1: Levels of description of unitary dynamics and what each determines.

one the cycle compatibilities are the additional constraints. Part III confirmed these statements in small models, and Part IV asks whether the phases that composition retains can themselves be organized as a connection on a bundle of continuation spaces.

Part III

Finite Models and Worked Examples

The first two parts built definitions and theorems at a level of generality that can hide errors. This part tests them on models small enough that every quantity can be computed by hand or by a few lines of code. The models are chosen to exhibit, one at a time, the phenomena that the later parts treat in larger settings: a finite state hidden behind unbounded memory, a gap between support and probabilistic sufficiency, interference in a two-level system, local rules with global consequences, and the dependence of the minimal description on the class of questions asked. Each chapter ends with a quantity that can be checked independently, and the numerical values quoted were verified by direct computation.

Chapter 16

Finite History Structures and Their Minimal Automata

Let Q be a finite set of states, Σ a finite alphabet, $q_0 \in Q$, and $\delta\colon Q \times \Sigma \rightharpoonup Q$ a partial transition function. Call the automaton *trim* if every state reachable from q_0 admits an infinite run. Define a history structure by taking $\mathcal{H}$ to be the set of finite words h for which $\delta(q_0, h)$ is defined, $\mathcal{F} = \Sigma^{\mathbb{N}}$, and

$$\mathcal{A}(h) = \{\varphi \in \Sigma^{\mathbb{N}} : \delta(\delta(q_0, h), \varphi_1 \cdots \varphi_k) \text{ is defined for every } k\}.$$

This history structure is extendable in the sense of Chapter 6, and Example 1.4 is of this form.

Theorem 16.1 (Minimal sufficient observation of a finite automaton)**.** *Write $\mathcal{A}(q)$ for the set of infinite words admitted from the state q. Then $\mathcal{A}(h) = \mathcal{A}(\delta(q_0, h))$, so the minimal sufficient observation of the history structure is the quotient of the reachable states by the relation $q \sim q'$ if and only if $\mathcal{A}(q) = \mathcal{A}(q')$. In a trim automaton, $q \sim q'$ if and only if q and q' admit the same finite words, and $\sim$ is the limit of the relations $\sim_k$ defined by $q \sim_0 q'$ for all q, q' and*

$$q \sim_{k+1} q' \iff \text{the same letters are defined at } q \text{ and } q', \text{ and } \delta(q,a) \sim_k \delta(q',a) \text{ for every defined letter } a.$$

The sequence of relations stabilizes after at most $|Q| - 1$ steps.

Proof. The first statement is the definition of the run from q_0 after reading h. In a trim automaton, a finite word is admitted from q exactly when it is a prefix of an element of $\mathcal{A}(q)$: one direction is trivial, and for the other, extend the run to an infinite one using trimness.

Hence $\mathcal{A}(q) = \mathcal{A}(q')$ if and only if q and q' admit the same finite words, and $q \sim_k q'$ says that they admit the same words of length at most k. The relations $\sim_k$ form a decreasing chain of equivalence relations on a finite set. Once $\sim_{k+1}=\sim_k$ the chain is constant from that point on, because $\sim_{k+2}$ is computed from $\sim_{k+1}$ by the same rule. A strictly decreasing chain of equivalence relations on $|Q|$ points has at most $|Q| - 1$ strict steps, since each step increases the number of classes. □

This is the partition-refinement algorithm of Moore [11], here read as the computation of a minimal sufficient observation. Applied to Example 1.4, take as automaton the three-state machine with states N (no aa yet, last letter not a), A (no aa yet, last letter a) and D (dead zone, after aa). From N both letters are defined, with $\delta(N, a) = A$ and $\delta(N, b) = N$; from A both are defined with $\delta(A, a) = D$ and $\delta(A, b) = N$; from D only b is defined, with $\delta(D, b) = D$. Then $\sim_1$ separates D from N and A, and $\sim_2$ separates N from A because $\delta(N, a) = A$ admits b and a whereas $\delta(A, a) = D$ admits only b. The automaton is already minimal at three states. A five-state presentation that records the length of the current run of b would collapse onto the same three classes, which is the sense in which memory depth and minimal state differ.

Proposition 16.2 (Finite state versus finite depth)**.** *A structure of memory depth k over Σ has at most $|\Sigma|^k$ minimal states among histories of length at least k, together with at most $\sum_{m<k} |\Sigma|^m$ further classes from shorter histories. The converse fails: Example 1.4 has three minimal states and infinite memory depth.*

Proof. If π_k is sufficient then $\mathcal{A}$ factors through the map sending a history of length at least k to its last k letters, which takes at most $|\Sigma|^k$ values, and shorter histories are retained whole. The failure of the converse was shown in Chapter 4. □

Chapter 17

Two-Level Systems and Unistochastic Chains

The smallest quantum system is also the setting in which Part II is fully explicit. Every unitary $U \in U(2)$ satisfies $|U_{00}|^2 = |U_{11}|^2 = p$ and $|U_{01}|^2 = |U_{10}|^2 = 1 - p$, so

$$T(U) = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix},$$

and every such matrix arises: given $p \in [0,1]$ take θ with $\cos^2(\theta/2) = p$ and the real rotation $R_y(\theta) = \begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}$. Thus the unistochastic matrices coincide with the doubly stochastic ones for $n = 2$, as in the manuscript's degeneracy theorem [7].

Proposition 17.1 (Interference in the qubit rotation family)**.** *For rotations about a common axis, $R_y(\theta_2)R_y(\theta_1) = R_y(\theta_1 + \theta_2)$, and the probability of remaining in the initial state is*

$$\cos^2\!\Big(\frac{\theta_1 + \theta_2}{2}\Big) = \Big[\cos^2\tfrac{\theta_1}{2}\cos^2\tfrac{\theta_2}{2} + \sin^2\tfrac{\theta_1}{2}\sin^2\tfrac{\theta_2}{2}\Big] - \tfrac{1}{2}\sin\theta_1\sin\theta_2.$$

The bracket is the entry of the product $T(R_y(\theta_2))T(R_y(\theta_1))$, and the last term is the interference contribution of Proposition 7.1.

Proof. Write $a = \theta_1/2$ and $b = \theta_2/2$. Then $\cos^2(a+b) = (\cos a\cos b - \sin a\sin b)^2 = \cos^2 a\cos^2 b + \sin^2 a\sin^2 b - 2\sin a\cos a\sin b\cos b$, and $2\sin a\cos a\sin b\cos b = \tfrac{1}{2}\sin\theta_1\sin\theta_2$. □

For $\theta_1 = \theta_2 = \pi/2$ the bracket is $\tfrac{1}{2}$ and the interference term is $-\tfrac{1}{2}$, so the coherent probability vanishes, which is the Hadamard-type cancellation of Chapter 7 in a rotation

form. A numerical check at $\theta_1 = 0.7$, $\theta_2 = 1.9$ gives the same value, 0.3048, for the difference between the bracket and $\cos^2((\theta_1 + \theta_2)/2)$ and for $\frac{1}{2}\sin\theta_1 \sin\theta_2$.

Proposition 17.2 (Coherent evolution is not a Markov semigroup)**.** *Let $U(t) = R_y(2t)$ and $T(t) = T(U(t))$, so that $T(t)$ has stay-probability $p(t) = \cos^2 t$. The family $t \mapsto T(t)$ is not a semigroup: $T(s+t) \neq T(s)T(t)$ in general. For example $T(\pi/2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ whereas $T(\pi/4)^2 = T(\pi/4) = J$.*

Proof. A matrix of the symmetric form above has eigenvalue $2p - 1$ on the vector $(1, -1)^{\mathsf{T}}$, and the product of two such matrices has eigenvalue equal to the product of the eigenvalues. A semigroup would require $\cos 2(s+t) = \cos 2s \cos 2t$. At $s = t = \pi/4$ the left side is -1 and the right side is 0. □

The proposition separates two ways of passing to a classical chain. The chain obtained by measuring and discarding between steps, in the sense of Proposition 7.3, has n-step stay probability $\frac{1}{2}(1+\cos^n\theta)$ and relaxes to the uniform distribution. The coherent evolution has stay probability $\frac{1}{2}(1+\cos n\theta)$ and oscillates indefinitely. For histories consisting of rotations about the y-axis, with futures drawn from the same family followed by a computational-basis measurement, the ray of the state after n steps is fixed by $n\theta$ modulo 2π, so by Theorem 9.1 the minimal probabilistically sufficient state is a point on a circle, which no single stay probability determines, since $\cos n\theta$ fixes the angle only up to sign.

Chapter 18

Hidden Markov Models and Insufficient Observations

A hidden Markov model supplies observations through a state that the observer does not see. It is the standard setting in which a natural observation, here the most recent output symbol, fails to be sufficient.

Example 18.1 (The even process)**.** Take hidden states A, B, output alphabet $\{0, 1\}$, start in A. From A, emit 0 and stay in A with probability $\frac{1}{2}$, or emit 1 and move to B with probability $\frac{1}{2}$. From B, emit 1 and move to A with probability 1. Histories are the output words with positive probability, and $\mathcal{A}(h)$ is the set of infinite output words every finite prefix of which has positive conditional probability after h. The state reached after a word is A if the number of trailing 1s is even and B if it is odd, counting from the last 0 or from the start.

Proposition 18.2. *In the even process, the minimal sufficient observation has exactly two values, the parity of the current run of* 1*s, while the memory depth is infinite.*

Proof. After a word ending in an odd run the next symbol must be 1, so $\mathcal{A}$ contains only words beginning with 1, whereas after an even run both symbols are possible. The parity therefore separates two classes, and the transition structure shows that the classes form the two states A and B with no further distinction. For depth, fix k and compare 01^k and 01^{k+1}. They share their last k letters, both being 1^k, but have runs of lengths k and $k+1$ of opposite parity. Hence no π_k is sufficient. □

Each of the laws $P(h)$ is also determined by the parity, since the law of the future is determined by the hidden state reached and the hidden state is determined by the parity. The support quotient and the probabilistic quotient therefore coincide here. The next example shows that they need not.

Example 18.3 (A coin of unknown bias)**.** A hidden bit selects, with probability $\frac{1}{2}$ each, one of two coins: one gives 1 with probability p and the other with probability $1-p$, where $0 < p < 1$ and $p \neq \frac{1}{2}$. The process then emits independent flips of the chosen coin. Every finite word has positive probability, so every finite prefix of every infinite word has positive conditional probability after every history, and the support of every future law is all of $\{0,1\}^{\mathbb{N}}$.

Proposition 18.4. *In Example 18.3 the support quotient has a single class, while the probabilistic quotient is in bijection with $\mathbb{Z}$ via $h \mapsto n_1(h) - n_0(h)$, the excess of ones over zeros. The family of laws, however, spans a space of dimension two.*

Proof. Support: every future is admissible after every history, so $\mathcal{A}$ is constant. Probabilities: with $r = p/(1-p)$ the posterior probability of the first coin after a word with n_1 ones and n_0 zeros is

$$w = \frac{p^{n_1}(1-p)^{n_0}}{p^{n_1}(1-p)^{n_0} + (1-p)^{n_1}p^{n_0}} = \frac{1}{1 + r^{-(n_1 - n_0)}},$$

which is strictly increasing in $n_1 - n_0$ if $p > \frac{1}{2}$ (so $r > 1$) and strictly decreasing if $p < \frac{1}{2}$ (so $r < 1$); in either case it is injective in $n_1 - n_0$, which is all that is needed. The future law is the mixture $w\,\mathrm{Bern}(p)^{\otimes\mathbb{N}} + (1-w)\,\mathrm{Bern}(1-p)^{\otimes\mathbb{N}}$, and the probability that the next symbol is 1 equals $wp + (1-w)(1-p) = \frac{1}{2} + (w - \frac{1}{2})(2p-1)$, which is injective in w. Thus distinct values of $n_1 - n_0$ give distinct laws, and the quotient is indexed by $\mathbb{Z}$. Every future law lies in the span of the two component laws, so the span is two dimensional. □

The example shows that the minimal probabilistically sufficient observation can be an infinite set even though the laws it indexes lie in a two-dimensional linear space. The next chapter returns to this point, because it separates the cardinality of the minimal state from the dimension of the smallest linear representation of the process.

Chapter 19

Lattice Gases and Local Admissibility

Local rules constrain global configurations, and the interplay between the two is the finite-size version of the continuation problem of Part I. Two models make it concrete.

A hard-core constraint

Let the configurations be words over $\{0,1\}$ in which no two adjacent symbols are both 1. Read left to right as a history structure, $\mathcal{A}(h)$ is the set of admissible infinite continuations, and two histories are equivalent if they have the same last symbol being 1 or not.

Proposition 19.1. *The hard-core words have a minimal sufficient observation with two values, the indicator that the last letter is* 1*. The number of admissible words of length n is the Fibonacci number F_{n+2} (with $F_1 = F_2 = 1$), and their exponential growth rate is $\log \varphi$ with $\varphi = (1+\sqrt{5})/2$.*

Proof. After a word ending in 1 the next letter must be 0, and after a word ending in 0 both letters are possible, so these are the two classes and no finer distinction arises. Let a_n count admissible words of length n ending in 0 and b_n those ending in 1. Then $a_{n+1} = a_n + b_n$ and $b_{n+1} = a_n$, so the transfer matrix is $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ with largest eigenvalue φ. The totals $a_n + b_n$ for $n = 1, \ldots, 8$ are $2, 3, 5, 8, 13, 21, 34, 55$, which are F_{n+2}, and the growth rate is $\log \varphi$. $\square$

The model shows a local constraint, one forbidden pair, producing a global count governed by a two-state transfer matrix, with the minimal state equal to the number of rows of that matrix.

A conservative lattice gas

Let particles on a ring of N cells update synchronously by the rule that a particle advances one cell to the right if the cell ahead is empty and otherwise stays, which is the cellular automaton known as Rule 184 [12]. In formulas, $x'_i = x_{i-1}(1 - x_i) + x_i x_{i+1}$ with indices modulo N.

Proposition 19.2. *Rule 184 conserves the particle number $n = \sum_i x_i$. The observation $x \mapsto n$ is insufficient for the continuation sets.*

Proof. Conservation: summing the update over i gives $\left(n - \sum_i x_{i-1}x_i\right) + \sum_i x_i x_{i+1}$, and the two sums of products coincide after reindexing, so the total is n. Insufficiency: on a ring of four cells, the configurations 1100 and 1010 both have $n = 2$. The first update gives $1100 \mapsto 1010$ and $1010 \mapsto 0101$, which differ, so the one-step continuations are different. □

Here the dynamics is deterministic, so the continuation of a configuration is its sequence of successors. If the continuation excludes the present configuration, two configurations have the same continuation exactly when they have the same image under the update, since the first successor determines all later ones. The minimal state is then the image of the update and not the configuration, and the quotient can identify distinct configurations: on four cells both 1100 and 0101 map to 1010, so Rule 184 is not injective. If the present configuration is included in the continuation, the minimal state is the configuration itself. Particle number is a conserved coordinate on the state space, not a coordinate system for it. The distinction generalizes to every model in the book that has a conserved quantity: a conserved observable is sufficient for the question of whether it will remain the same, and insufficient for most other questions.

Chapter 20

Computing Minimal Sufficient Observations

Theorem 16.1 gives an algorithm for support sufficiency in finite automata, with at most $|Q| - 1$ refinement rounds, each of which inspects every state and letter once. For probabilistic sufficiency the right object is the Hankel matrix of the process.

Definition 20.1. For a process assigning probability $P(w)$ to each finite word w, the *Hankel matrix* has rows and columns indexed by finite words, with entry $H(u, v) = P(uv)$.

The rows of H are proportional to the future laws after the histories u, since $P(v \mid u) = P(uv)/P(u)$. Two histories have the same future law if and only if their rows are proportional, so the minimal probabilistically sufficient observation is the set of distinct normalized rows. The number of distinct rows can be infinite, as Proposition 18.4 shows, but the *rank* of H is a different measure of size.

Theorem 20.2 (Hankel rank; Fliess, Jaeger)**.** *A process has a finite-dimensional linear (observable-operator) representation if and only if its Hankel matrix has finite rank, and the minimal dimension of such a representation equals that rank [8, 10].*

For the coin of unknown bias, $P(w) = \frac{1}{2}a(w) + \frac{1}{2}b(w)$ with $a(w) = p^{n_1}(1-p)^{n_0}$ and $b(w) = (1-p)^{n_1}p^{n_0}$, both multiplicative under concatenation, so $H(u, v) = \frac{1}{2}a(u)a(v) + \frac{1}{2}b(u)b(v)$ has rank at most two. A direct computation on all words of length up to four with $p = 0.8$ returns rank exactly two. The same process thus has an infinite minimal probabilistic quotient and a linear representation of dimension two. The first measures how many distinct futures can occur and the second measures how many independent directions they span, and both are legitimate answers to the question of how much a theory must remember, depending on whether the remembered object is required to be a point or a vector.

Chapter 21

A Catalogue of Obstructions in Small Models

The models of this part and of Part II can be arranged by what kind of insufficiency each exhibits, which prepares the comparison of obstructions in Part VIII.

The even process of Chapter 18 exhibits support insufficiency of the instantaneous symbol, repaired by a two-valued internal state, with infinite memory depth. The coin of unknown bias exhibits the gap between the two grades of sufficiency: one support class, a countably infinite probabilistic quotient, and linear dimension two. The projection T on the qubit exhibits probabilistic insufficiency with a hidden phase relative to the Born probabilities, repaired by the ray, and compositional insufficiency, repaired by $PU(2)$. Rule 184 exhibits insufficiency of a conserved coordinate. The triangle of three real contexts exhibits a failure of local data to extend globally, with the single relative sign as the fibre.

Model	Support	Probabilistic	Linear dim.	Depth
Example 1.4	3	not defined	not defined	infinite
Even process	2	2	2	infinite
Unknown-bias coin	1	countably infinite	2	not applicable
Hard-core words	2	not defined	not defined	1
Qubit rotation chain	not defined	circle	not computed	not computed

Table 21.1: Sizes of the minimal description in the finite and low-dimensional models of this part. Entries marked "not defined" indicate that the model carries no probability law in the form used in the corresponding chapter.

Two patterns emerge from the table. First, finite state does not imply finite memory depth, and finite memory depth is a rare special case rather than the rule. Second, the three measures of the size of the description, number of classes, cardinality of the probabilistic quotient, and linear dimension, can take independent values. A theory that claims to have

found the state of a system has to say which of these it has computed. The remaining parts of the book return to this question in settings where the answer carries physical weight: the geometric structure of the quotient in Part IV, its behavior under coarse-graining in Part V, and its continuum limit in Part VI.

Part IV

Transport, Bundles, and Holonomy

Chapter 22

Families of Continuations as Bundles

The preceding parts construct states by identifying histories that support the same predictions. They also establish that a reduced description may fail to preserve composition. We now ask a further question: how can predictive or distinguishable structure at one state be compared with corresponding structure at another?

An assignment of continuation sets does not answer this question. Even when every state carries a vector space of the same dimension, the assignment alone specifies neither a topology on their union nor a rule for comparing their elements. Bundle structure and transport must therefore be constructed from additional data.

Throughout this part, let M be a smooth manifold representing a chosen family of continuation states. Its existence as a smooth manifold is an assumption about the model, rather than a consequence of the set-theoretic quotient construction of Part I.

Definition 22.1 (Distinction bundle)**.** A distinction bundle over M is a smooth real or complex vector bundle

$$p : E \longrightarrow M.$$

The fibre $E_x = p^{-1}(x)$ represents a specified vector space of distinction data at the state x.

The interpretation of the fibre must be supplied in each application. It may contain linearized perturbations, observable contrasts, amplitudes, or another explicitly constructed family of quantities. These interpretations are not interchangeable merely because they share a vector-space representation.

Proposition 22.2 (Constant-rank construction)**.** *Let W and Z be finite-dimensional vector spaces and let*

$$C : M \longrightarrow \mathrm{Hom}(W, Z)$$

be smooth with constant rank. Then

$$E = \{(x, w) \in M \times W : C(x)w = 0\}$$

is a smooth vector subbundle of $M \times W$*.*

Proof. Constant rank permits a nonvanishing maximal minor to be selected near each point. After a fixed reordering of coordinates, the corresponding equations express the pivot coordinates smoothly in terms of the free coordinates. These free coordinates provide a smooth local frame for the kernel. Its dimension is $\dim W - \operatorname{rank} C$, independently of the base point. □

This construction applies when admissible infinitesimal distinctions are the solutions of smooth linear constraints. If the rank changes, the fibre dimension changes and the result need not be a vector bundle. Singular families will require a stratified or otherwise generalized treatment.

Remark 22.3. Equal fibre dimensions alone do not specify the topology, local trivializations, or transition functions of a bundle. The kernel construction supplies these structures through its smooth constant-rank hypothesis.

Chapter 23

Distinction Transport Along Histories

A physical history can carry distinction data from one state to another. To formulate this without assuming reversibility, begin with a category of admissible paths.

Definition 23.1 (Admissible path category)**.** An admissible path category has states as objects and admissible directed paths as morphisms. It contains constant paths and is closed under concatenation, with paths identified under the reparametrizations needed to make concatenation associative.

Closure under reversal is not required. A history-dependent model may need an enlarged state space before admissibility is closed under concatenation: two paths that meet at the same reduced state need not be physically compatible if that state has discarded relevant history.

Definition 23.2 (Linear transport)**.** Linear transport assigns to each admissible path $\gamma : x \to y$ a linear map

$$P_\gamma : E_x \longrightarrow E_y$$

such that

$$P_{\mathrm{id}_x} = \mathrm{id}_{E_x}, \qquad P_{\gamma_2 \circ \gamma_1} = P_{\gamma_2} P_{\gamma_1}.$$

Thus transport is a functor from the admissible path category to the category of vector spaces. The maps may be invertible, contractive, or rank deficient, depending on the model.

Proposition 23.3 (Loss under composition)**.** *For composable transports,*

$$\operatorname{rank}(P_{\gamma_2} P_{\gamma_1}) \leq \min\{\operatorname{rank} P_{\gamma_1}, \operatorname{rank} P_{\gamma_2}\}.$$

Moreover,

$$\ker P_{\gamma_1} \subseteq \ker(P_{\gamma_2} P_{\gamma_1}).$$

Proof. The image of the composite is the image under P_{γ_2} of a subspace of $\operatorname{im} P_{\gamma_1}$, which gives both rank bounds. If $P_{\gamma_1} u = 0$, its image under P_{γ_2} is also zero. □

Once a distinction has been annihilated, subsequent transport cannot recover it from the transported vector alone. Recovery requires additional retained data or an enlarged description.

When all maps are invertible, two paths with common endpoints can be compared.

Definition 23.4 (Relative transport)**.** For $\gamma_1, \gamma_2 : x \to y$ with invertible transports, define

$$H(\gamma_1, \gamma_2) = P_{\gamma_2}^{-1} P_{\gamma_1} : E_x \to E_x.$$

The inverse in this expression is an algebraic comparison. It does not assert that the reverse of γ_2 is a physically admissible history. This distinction is essential in a theory whose history extensions are directed.

Chapter 24

Connections on Continuation Spaces

A smooth connection provides a local differential rule for invertible transport along smooth paths.

Definition 24.1 (Connection)**.** A connection on E is a linear operator

$$\nabla : \Gamma(E) \longrightarrow \Gamma(T^*M \otimes E)$$

satisfying

$$\nabla(fs) = df \otimes s + f\nabla s.$$

In a local frame, write

$$\nabla = d + A,$$

where A is a matrix-valued one-form. A section $u(t)$ along $\gamma(t)$ is parallel when

$$\frac{du}{dt} + A_{\gamma(t)}(\dot{\gamma}(t))u(t) = 0.$$

Proposition 24.2 (Parallel transport)**.** *A smooth connection determines invertible parallel transport along every piecewise smooth path. It satisfies concatenation and*

$$P_{\gamma^{-1}} = P_\gamma^{-1}.$$

Proof. In a local frame, the parallel equation is a linear ordinary differential equation with smooth coefficients. Existence and uniqueness give a linear solution operator. Solving along the reversed path yields its inverse. Uniqueness also identifies the solution along a concatenation with the composition of the solutions along its pieces. □

A connection can be restricted to physically admissible paths, but its mathematical reverse transport remains defined even when the reversed path is physically forbidden.

Connections therefore describe a specific class of transport models. They do not directly represent rank loss at finite time. For example, a smooth finite-dimensional linear evolution may contract vectors while its finite-time solution operator remains invertible. Exact erasure requires a singular limit, a projection, or a more general noninvertible transport rule.

Nor does every assignment of path maps arise from a connection. Smooth dependence, locality, and appropriate compatibility with path parametrization impose additional requirements. We assume a connection only where those requirements have been established.

Chapter 25

Holonomy of Closed Histories

Definition 25.1 (Holonomy)**.** For a piecewise smooth loop γ based at x, the parallel transport

$$P_\gamma : E_x \longrightarrow E_x$$

is its holonomy. The transports of all based loops form the holonomy group at x.

A loop in a reduced state space need not be a loop in the space of complete histories. The reduced state can return while the record of the journey grows. Holonomy therefore concerns return in a specified base description, not deletion of the intervening history.

Proposition 25.2 (Endpoint independence)**.** *Suppose M is path connected. Parallel transport depends only on its endpoints if and only if every loop has identity holonomy.*

Proof. Endpoint independence makes transport around every loop equal to transport along the constant path.

Conversely, let $\gamma_1, \gamma_2 : x \to y$. The loop obtained by following γ_1 and then γ_2^{-1} has transport $P_{\gamma_2}^{-1} P_{\gamma_1}$. If every loop has identity transport, then $P_{\gamma_1} = P_{\gamma_2}$. □

Trivial loop holonomy consequently permits a global identification of fibres by transport from a fixed reference state. Nontrivial holonomy obstructs that particular identification.

This gives a precise form of path dependence. The base endpoint alone does not determine the transported distinction. A sufficient description must retain enough information to distinguish the relevant transport classes, or retain the resulting fibre value.

Chapter 26

Curvature, Holonomy, and Irreversible Memory

The curvature of a connection is

$$F_\nabla = \nabla^2.$$

In a local frame,

$$F = dA + A \wedge A.$$

It measures the local failure of covariant derivatives to commute:

$$F(X,Y)s = \nabla_X \nabla_Y s - \nabla_Y \nabla_X s - \nabla_{[X,Y]} s.$$

Curvature and holonomy are related but distinct. Curvature is local differential data. Holonomy records transport around finite loops and can also detect global topology.

Example 26.1 (Flat connection with nontrivial holonomy)**.** On the trivial complex line bundle over the circle, let

$$\nabla = d + ia\, d\theta, \qquad a \in \mathbb{R},$$

where $d\theta$ is the globally defined angular one-form. The curvature vanishes, but transport once around the circle is

$$P_\gamma = e^{-2\pi i a}.$$

It is nontrivial whenever $a \notin \mathbb{Z}$.

Thus vanishing curvature does not imply globally path-independent transport. A flat connection may retain the homotopy class of a path.

Definition 26.2 (Transport record)**.** Given $u \in E_x$ and a loop γ at x, the transport record

of the loop relative to u is

$$m_\gamma(u) = P_\gamma u.$$

The record distinguishes two loops when their resulting vectors differ under the allowed observations.

This is a mathematical model of path-sensitive storage. Its interpretation as physical memory requires a preparation, a readable variable, and a mechanism that preserves the record.

Proposition 26.3 (Holonomy does not imply erasure)**.** *Parallel transport around a loop is invertible. Consequently nontrivial holonomy alone does not imply information loss.*

Proof. The inverse transport is $P_{\gamma^{-1}}$. Hence distinct input vectors remain distinct after transport. □

Neither nontrivial holonomy nor curvature alone establishes positive thermodynamic entropy production. Such a conclusion requires a physical dynamics, a thermodynamic state description, and an entropy balance. Geometric path dependence can occur without irreversible erasure.

This separates three questions that the later thermodynamic development must answer independently: whether transport depends on the path, whether the transported description loses distinctions, and whether the physical process produces entropy.

Chapter 27

Gauge Structure From Transport

Choosing a local frame gives numerical coordinates to fibre elements. A different frame changes those coordinates while leaving the represented element unchanged.

Adopt the convention

$$u' = g^{-1}u$$

for a smooth change of frame g. Covariance of the parallel equation requires

$$A' = g^{-1}Ag + g^{-1}dg.$$

Proposition 27.1 (Gauge covariance)**.** *Under a change of frame,*

$$F' = g^{-1}Fg, \qquad P'_\gamma = g(\gamma(1))^{-1}P_\gamma g(\gamma(0)).$$

For a loop based at x, its holonomy therefore changes by conjugation at x.

Proof. Substitution of $u = gu'$ into the parallel equation yields the connection transformation law and the endpoint transformation of its solution operator. Expanding $dA' + A' \wedge A'$ gives $g^{-1}Fg$. □

In finite dimensions, the trace, determinant, and spectrum of loop holonomy are consequently invariant under changes of frame. If the bundle carries a preserved inner product, unitary or orthogonal frame changes are natural, and corresponding invariant norms can also be used.

A proposed repair cost must respect this structure. For example, given a Hermitian metric,

$$L(\gamma, u) = \|P_\gamma u - u\|^2$$

for a based loop is independent of the chosen orthonormal frame when both the vector and

transport are transformed consistently.

The construction establishes gauge covariance as freedom in the description of transport. It does not by itself derive a physical gauge field. A physical theory must additionally specify the structure group, observable quantities, admissible couplings, and equations governing the connection. A field equation for A cannot be inferred from frame covariance alone.

Chapter 28

Geometric Phase as Retained Information

Quantum pure states provide a concrete bundle in which transport and observational projection interact.

The unit sphere in $\mathbb{C}^d$ projects onto projective space,

$$S^{2d-1} \longrightarrow \mathbb{CP}^{d-1}, \qquad \psi \longmapsto [\psi].$$

Its fibres are the phase orbits $\{e^{i\chi}\psi : \chi \in \mathbb{R}\}$.

For a normalized local representative ψ, define the real one-form

$$\mathcal{A} = -i\langle \psi, d\psi \rangle.$$

Under $\psi' = e^{i\chi}\psi$,

$$\mathcal{A}' = \mathcal{A} + d\chi.$$

Its curvature

$$\mathcal{F} = d\mathcal{A}$$

is therefore independent of the local phase convention.

For a closed ray path admitting a periodic normalized representative, define

$$\gamma_{\mathrm{g}} = -\oint \mathcal{A} \pmod{2\pi}.$$

Its exponential is independent of periodic changes of phase representative, since such changes alter the integral by an integer multiple of 2π.

Example 28.1 (A two-level loop)**.** Let

$$\psi(\vartheta, \varphi) = \begin{pmatrix} \cos(\vartheta/2) \\ e^{i\varphi} \sin(\vartheta/2) \end{pmatrix}.$$

Then

$$\mathcal{A} = \frac{1 - \cos\vartheta}{2}\, d\varphi, \qquad \mathcal{F} = \frac{1}{2} \sin\vartheta\, d\vartheta \wedge d\varphi.$$

For the loop at fixed ϑ with φ increasing from 0 to 2π,

$$\gamma_{\mathrm{g}} = -\pi(1 - \cos\vartheta) \pmod{2\pi}.$$

The ray returns to its starting point, while parallel transport retains a phase associated with the path. This phase is observable through comparison with a reference branch after dynamical contributions have been controlled or accounted for.

There is no conflict with the projective quotient of Part II. A common phase multiplying the complete state remains unobservable. In an interference experiment, the phase accumulated on one branch is relative to another branch, and the combined system contains that relative distinction.

Geometric phase therefore supplies an explicit instance of the book's organizing problem. An endpoint description can be sufficient for some predictions while failing to determine a comparison that depends on the route taken. The required repair is specified by transport and its holonomy.

The later field-theoretic construction must establish whether RSVP variables define an appropriate bundle, connection, and observable transport law. The present part supplies the mathematical conditions for that investigation. It does not identify an entropy field with curvature or infer a connection from a vector flow without an additional construction.

Chapter 29

Gluing as Holonomy

Part II treated the gluing of measurement contexts as a question about the existence of frames, and Part IV developed transport and holonomy as a theory of path dependence. This chapter connects the two. The connection is exact but narrower than a first reading suggests: the obstruction to gluing is not the nontriviality of a predetermined holonomy, but the impossibility of choosing the lifts of the edge data so that every cycle closes. The cases of real and complex frames, the relation to the projective unitary group, and the difference from Berry holonomy are treated in turn.

Labelled frames and edge lifts

Fix a field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and a dimension d. A *frame* is a matrix F_a in $O(d)$ or $U(d)$ whose columns are the labelled vectors of an orthonormal basis, the context a. For an oriented edge $a \to b$ of a connected graph G set

$$L_{ba} = F_b^\dagger F_a, \qquad B_{ba} = |L_{ba}|^{\circ 2}, \qquad L_{ab} = L_{ba}^{-1} = L_{ba}^\dagger,$$

where the modulus square is taken entrywise. The matrix B_{ba} is the transition-probability matrix from the labelled vectors of a to those of b, and $B_{ab} = B_{ba}^{\mathsf{T}}$. For a cycle $c = (v_0, v_1, \ldots, v_n = v_0)$ define its cycle product

$$H_c = L_{v_0 v_{n-1}} \cdots L_{v_2 v_1} L_{v_1 v_0}.$$

The observed data consist of the matrices B_e alone. A *lift* of an edge datum B_{ba} is a matrix $L_{ba} \in O(d)$ or $U(d)$ with $|L_{ba}|^{\circ 2} = B_{ba}$. A *lift assignment* chooses a lift for each oriented edge, consistently with $L_{ab} = L_{ba}^{-1}$. A lift assignment is a connection in the discrete sense of Part IV: it is a rule for transporting a frame along each edge, and H_c is its holonomy

around the cycle c.

Theorem 29.1 (Gluing is the existence of a flat lift assignment)**.** *The following are equivalent for edge data $\{B_e\}$ on a connected graph G with spanning tree T. First, the data are globally $\mathbb{F}$-realizable: there are frames F_a with $|F_b^\dagger F_a|^{\circ 2} = B_{ba}$ on every edge. Second, there is a lift assignment for which $H_c = I$ for every cycle c. Third, there is a lift assignment for which $H_c = I$ for each fundamental cycle of T.*

Proof. (First implies second.) Take $L_{ba} = F_b^\dagger F_a$. This is a lift of B_{ba}, consistent with inversion, and the product around any cycle telescopes: $L_{v_2v_1}L_{v_1v_0} = F_{v_2}^\dagger F_{v_1} F_{v_1}^\dagger F_{v_0} = F_{v_2}^\dagger F_{v_0}$, and in general $H_c = F_{v_0}^\dagger F_{v_0} = I$.

(Second implies third.) Fundamental cycles are cycles.

(Third implies first.) Fix a root r and set $F_r = I$. For a tree edge from a vertex a with F_a already defined to a child b, set $F_b = F_a L_{ab}$. This is orthogonal or unitary, and $F_b^\dagger F_a = L_{ab}^\dagger F_a^\dagger F_a = L_{ab}^\dagger = L_{ba}$, so every tree edge is reproduced and the propagated frames satisfy the telescoping identity along tree paths. Let $e = (a \to b)$ be a non-tree edge and let $a = u_0, u_1, \ldots, u_m = b$ be the tree path from a to b. By telescoping, $L_{u_m u_{m-1}} \cdots L_{u_1 u_0} = F_b^\dagger F_a$. The fundamental cycle of e is this path followed by the edge $b \to a$, and $H = I$ gives $L_{ab} F_b^\dagger F_a = I$, that is $F_b^\dagger F_a = L_{ab}^{-1} = L_{ba}$. Hence the edge is reproduced, and $|F_b^\dagger F_a|^{\circ 2} = B_{ba}$ on every edge. Identity holonomy is invariant under conjugation, so the choice of base point on each fundamental cycle is immaterial. □

Theorem 29.1 restates Theorem 11.2 in the language of Part IV: the tree criterion becomes the statement that the discrete connection is flat on the fundamental cycles, and the common choice of tree lifts becomes the requirement that a single lift assignment serve every cycle at once.

The obstruction is a nonexistence

The probabilities do not determine the lifts. An edge datum has in general a whole set of lifts, and the holonomy of a lift assignment is a property of the assignment and not of the data. A nonidentity cycle product for one chosen lift assignment therefore establishes nothing about realizability. The obstruction is that no lift assignment has trivial cycle products. This is a different kind of statement from the nontriviality of a given holonomy, and it is the reason the obstruction cannot generally be read as a $\mathbb{Z}_2$ or $U(1)$ invariant of the data.

Proposition 29.2 (The real mutually unbiased triangle)**.** *Let $B_{XZ} = B_{ZY} = B_{YX} = J$, where J is the 2×2 matrix with all entries $\frac{1}{2}$. Each edge has exactly eight lifts in $O(2)$,*

namely $R(\theta)$ and $R(\theta)\,\mathrm{diag}(1,-1)$ for $\theta \in \{\pi/4, 3\pi/4, 5\pi/4, 7\pi/4\}$. For every choice of lifts of the three edges, the cycle product is neither I nor $-I$. In particular, three lifts all equal to $R(\pi/4)$ have the cycle product $R(3\pi/4)$.

Proof. An orthogonal matrix with all entries of modulus $1/\sqrt{2}$ has first column $(\pm 1, \pm 1)/\sqrt{2}$, which determines it up to the sign of the second column, giving the eight matrices listed. Write each lift as $R(\theta_i)P^{k_i}$ with $P = \mathrm{diag}(1,-1)$, $k_i \in \{0,1\}$, and θ_i an odd multiple of $\pi/4$. The determinant of the product is $(-1)^{k_1+k_2+k_3}$, so a product equal to I or $-I$ requires $k_1 + k_2 + k_3 \in \{0, 2\}$. Using $PR(\theta)P = R(-\theta)$, the product with zero or two reflections is a rotation $R(\pm\theta_1 \pm \theta_2 \pm \theta_3)$. The angle $\pm\theta_1 \pm \theta_2 \pm \theta_3$ is a sum of three odd multiples of $\pi/4$, hence an odd multiple of $\pi/4$, and so is not a multiple of π. The product is therefore neither $R(0) = I$ nor $R(\pi) = -I$. □

An exhaustive check of all $8^3 = 512$ triples confirms that none has product $\pm I$. The proposition shows that the holonomy of a lift assignment can be a rotation through $3\pi/4$, an element of $SO(2)$ of order 8, so the obstruction cannot be recorded as a sign. What does reduce to a discrete choice is the selection among the lifts of each edge, the signs and π-shifts above. These discrete choices constrain a rotation-valued closure equation.

Proposition 29.3 (The real triangle as a closure equation)**.** *Let B_{ba} ($ba \in \{XZ, ZY, YX\}$) be real 2×2 bistochastic with parameters $p_e \in (0,1)$ and $\alpha_e = \arccos\sqrt{p_e} \in (0, \pi/2)$. Each edge has the eight lifts $R(\theta)P^k$ with $\theta \in \{\pm\alpha_e, \pm\alpha_e + \pi\}$, and there is a lift assignment with trivial cycle product if and only if*

$$\varepsilon_1\alpha_1 + \varepsilon_2\alpha_2 + \varepsilon_3\alpha_3 \equiv 0 \pmod{\pi} \qquad \text{for some signs } \varepsilon_i \in \{\pm 1\}.$$

This is equivalent to the criterion of Theorem 11.3.

Proof. As in the previous proof, a product of three lifts with zero or two reflections is a rotation by $\pm\theta_1 \pm \theta_2 \pm \theta_3$, and the product has determinant 1 only in these cases. Each θ_i ranges over $\pm\alpha_i + \{0, \pi\}$, and the signs of the three terms can be absorbed into the choice of θ_i. The rotation equals I if and only if the angle is a multiple of 2π, and because each θ_i can be shifted by π independently, this is possible if and only if $\sum \varepsilon_i\alpha_i \equiv 0 \pmod{\pi}$ for some signs. The equivalence with Theorem 11.3 follows from $p_{XY} = \cos^2\alpha_{XY}$, the evenness of $\cos^2$, and its period π. □

For the unbiased triangle all $\alpha_i = \pi/4$, and the sum $\pm\pi/4 \pm \pi/4 \pm \pi/4$ is an odd multiple of $\pi/4$, never a multiple of π, which is the content of Proposition 29.2 and of the no-three-real-MUBs corollary. The relative sign of Chapter 9 is the choice of the $\pm$ in the closure equation, and the equation itself is the constraint.

The projective unitary bridge

Suppose instead that the edge data are projective unitary operations $[W_e] \in PU(d)$ and that their cycle products are the identity in $PU(d)$. Choose unitary representatives. The kernel of $U(d) \to PU(d)$ is the central circle $\{\lambda I : |\lambda| = 1\}$, so each cycle product of representatives is a scalar $\lambda_c I$ with $|\lambda_c| = 1$. In the real projective case the kernel of $O(d) \to PO(d)$ is $\{\pm I\}$ and $\lambda_c \in \{\pm 1\}$. These scalars are well defined only under the hypothesis of projective closure, that the cycle products are trivial in the projective group, and they are not available for arbitrary probability-labelled context graphs, whose edge data are not projective operations to begin with.

Nor are the scalars gauge invariant without further structure. Replacing the representative W_e by $\mu_e W_e$ multiplies λ_c by the product of the μ_e along the cycle. If every phase is allowed on every edge, any scalar can be removed, and no invariant remains. An invariant exists only when the permitted changes of representative are restricted, for example by fixing the representatives on a spanning tree, by a determinant normalization, or by a specified transport structure. The scalar cycle phase is therefore a cocycle for a stated gauge group and not a property of the unlabelled data.

Frame gluing is not Berry holonomy

The cycle product of Theorem 29.1 is the identity whenever a global family of frames exists, because the relative transformations telescope. Parallel transport in a bundle of rays is a different structure. It is governed by a connection whose holonomy around a loop can be nontrivial even when every ray is described by a globally defined vector in some trivialization, as in the two-level loop of the preceding chapter on geometric phase. A global assignment of frames and a connection that transports along paths are different data, and neither implies the other.

The two structures meet in a gauge-invariant quantity of a closed cycle of frames. Choose from each context v_k one labelled vector $\psi_k = F_{v_k} e_1$. The *Bargmann invariant*

$$\Delta = \langle \psi_0 | \psi_1 \rangle \langle \psi_1 | \psi_2 \rangle \cdots \langle \psi_{n-1} | \psi_0 \rangle = \prod_k (L_{v_k v_{k+1}})_{11}$$

is unchanged by phase changes of each ψ_k separately, and for a closed polygon of rays it is the Pancharatnam phase that tends to the Berry phase as the polygon is refined. Its modulus is fixed by the probability data, $|\Delta|^2 = \prod_k (B)_{11}$ over the edges. Its argument is not determined by the edge probabilities alone in general. For three qubit states with Bloch

vectors n_1, n_2, n_3 one has

$$\Delta = \mathrm{Tr}(\rho_1\rho_2\rho_3) = \tfrac{1}{4}\big(1 + n_1\cdot n_2 + n_2\cdot n_3 + n_3\cdot n_1 + i\, n_1\cdot(n_2 \times n_3)\big).$$

The pairwise probabilities $|\langle\psi_i|\psi_j\rangle|^2 = \frac{1}{2}(1 + n_i \cdot n_j)$ fix the dot products and hence, by the Gram determinant, the magnitude of the triple product, but not its sign. The probability data therefore determine the Bargmann phase of a qubit triangle up to complex conjugation, which is the orientation of the triangle on the Bloch sphere. A numerical check on random triples reproduces the displayed formula.

The distinction clarifies the roles of the two chapters. The frame-gluing holonomy answers whether a consistent global assignment exists, and the answer is a yes-or-no existence statement. The Berry or Bargmann holonomy is a numerical invariant of a realized configuration, and the geometric phase of the preceding chapter is an instance. With a global frame assignment in hand the first is trivial by construction and the second is in general not, which is exactly why a description that retains only endpoint probabilities can be sufficient for single transitions and insufficient for the phase accumulated around a cycle.

The question that opens Part V follows from the same observation. Everything above concerns invertible transport. What changes when the transport ceases to be invertible, or when an observer projects away part of the transported information, is the subject of the next part.

Part V

Irreversible Dynamics and Memory

Part IV concerned invertible transport, and it showed that path dependence, in the form of holonomy, does not by itself imply any loss of information. This part begins where invertibility ends: what changes when transport ceases to be invertible, or when an observer projects away part of the transported information. The preceding parts show that a description can lose information in two distinct ways. It can fail to determine the continuations of the system, which is insufficiency, and it can fail to determine its own update, which is the lack of closure. This part asks what each loss means for irreversibility. Three questions are kept apart throughout, in the spirit of the caution recorded at the end of Part IV. The first is whether the induced dynamics of a description is logically reversible, that is, whether distinct descriptions can evolve into the same one. The second is whether the process, observed through the description, is statistically distinguishable from its time reversal. The third is whether a physical process produces entropy, which requires a thermodynamic state description and an entropy balance that no purely informational argument supplies. The results below concern the first two questions and the way both depend on the chosen observation. Statements about physical entropy production are made only where a thermodynamic setting is specified.

Chapter 30

Irreversibility as Loss of Sufficiency

For a deterministic dynamics $f\colon X \to X$ on a set of states, logical reversibility means that f is injective, so that the present state determines the past one. The question of this chapter is how reversibility of a dynamics relates to reversibility of its descriptions. Let $\pi\colon X \to Y$ be a surjective observation that is closed in the sense of Chapter 12, so that there is a map $g\colon Y \to Y$ with $\pi \circ f = g \circ \pi$.

Proposition 30.1 (Finite reversible dynamics has reversible closed descriptions)**.** *If X is finite and f is a bijection, then every closed surjective observation has an induced map g that is a bijection.*

Proof. Surjectivity of f and of π gives $g \circ \pi = \pi \circ f$ surjective onto Y, so g is surjective. A surjection of the finite set Y to itself is a bijection. □

The hypothesis of finiteness cannot be dropped.

Proposition 30.2 (Irreversibility from projection)**.** *Let $X = \mathbb{Z}$ and $f(n) = n+1$, a bijection. Define $\pi(n) = \min(n, 0)$ with values in $Y = \{0, -1, -2, \dots\}$. Then π is closed, with induced map $g(0) = 0$ and $g(-k) = -k + 1$ for $k \geq 1$, and g is not injective: $g(-1) = g(0) = 0$.*

Proof. If $\pi(n) = \pi(m) = 0$ then $n, m \geq 0$, so $n+1, m+1 \geq 1$ and $\pi(n+1) = \pi(m+1) = 0$. If $\pi(n) = \pi(m) < 0$ then $n = m$. Hence π is closed. The values of g follow from $g(\pi(n)) = \pi(n+1)$, and the states -1 and 0 of the description are both sent to 0. □

The mechanism is the identification of an infinite family of microstates into one macrostate, after which the description cannot recover which member it came from. Information about the past is not destroyed in the underlying dynamics; it is destroyed in the description. This is the sense in which irreversibility at the level of a description is a property of the pair of a dynamics and an observation.

The quantitative form of the loss is the elementary inequality for the entropy of a push-forward.

Proposition 30.3 (Entropy and non-injective maps)**.** *Let p be a probability distribution on a finite set and g a map. Then $H(g_*p) \le H(p)$, with equality if and only if g is injective on the support of p.*

Proof. Let $X \sim p$. The chain rule gives $H(X) = H(g(X)) + H(X \mid g(X))$, because $g(X)$ is a function of X. The conditional entropy is nonnegative and vanishes exactly when X is determined by $g(X)$ almost surely, which is injectivity on the support. □

The difference $H(p) - H(g_*p) = H(X \mid g(X))$ is the number of nats of information about the state that the step discards. This is the information-theoretic quantity that Landauer's principle [33] relates to a minimum heat cost when the discarding is carried out by a physical process; the relation requires the thermodynamic setting of Chapter 32 and is not assumed here.

For stochastic dynamics the analogue of reversibility is bistochasticity, which the manuscript identifies as the natural admissibility constraint. Closed descriptions of bistochastic dynamics need not inherit it.

Proposition 30.4. *Let $M = J_3$ be the 3×3 matrix with all entries $\frac{1}{3}$, and lump the states into the blocks $A = \{0\}$ and $B = \{1, 2\}$. The aggregated chain has column-stochastic matrix $N = \begin{pmatrix} 1/3 & 1/3 \\ 2/3 & 2/3 \end{pmatrix}$, which is not bistochastic, and its stationary distribution is $(1/3, 2/3)$, proportional to the block sizes.*

Proof. By Theorem 12.3, $N_{BA} = \sum_{i \in B} M_{ij}$ for $j \in A$, which gives the stated matrix. Its row sums are $\frac{2}{3}$ and $\frac{4}{3}$, and $N(1/3, 2/3)^{\mathsf{T}} = (1/3, 2/3)^{\mathsf{T}}$. □

A reversible fine dynamics therefore yields a coarse dynamics that is stochastic but has a nonuniform invariant measure, weighted by the number of microstates in each block. This is the point at which Boltzmann's counting entropy $\ln |B|$ enters, and Chapter 32 makes the role of that weighting precise.

Chapter 31

Entropy Production Under Projection

Statistical irreversibility is the asymmetry between a process and its time reversal. Let M be an irreducible column-stochastic matrix on a finite set X with stationary distribution π, and assume that $M_{yx} > 0$ implies $M_{xy} > 0$. The forward path measure on paths $\omega = (x_0, \dots, x_n)$ is

$$P_n(\omega) = \pi_{x_0} \prod_{k=0}^{n-1} M_{x_{k+1} x_k},$$

and the reversed measure is $R_n(\omega) = P_n(\omega^{\mathrm{r}})$, where $\omega^{\mathrm{r}} = (x_n, \dots, x_0)$. The *entropy production* of the process up to time n is the relative entropy

$$\Sigma_n = D(P_n \,\|\, R_n) = \sum_{\omega} P_n(\omega) \ln \frac{P_n(\omega)}{R_n(\omega)} \;\geq\; 0.$$

For a stationary chain $\Sigma_n = n\Sigma_1$ with

$$\Sigma_1 = \sum_{x,y} \pi_x M_{yx} \ln \frac{\pi_x M_{yx}}{\pi_y M_{xy}} = \sum_{x,y} \pi_x M_{yx} \ln \frac{M_{yx}}{M_{xy}},$$

the second form following from stationarity [37]. It vanishes exactly when detailed balance $\pi_x M_{yx} = \pi_y M_{xy}$ holds.

Given a map $c\colon X \to Y$ the observed process is the sequence $c(x_0), \dots, c(x_n)$. Its forward and reversed path measures are the push-forwards of P_n and R_n under the coordinatewise map c^{n+1}, because reversing a path commutes with applying c coordinatewise. The coarse entropy production is $\Sigma_n^c = D(c_*^{n+1} P_n \,\|\, c_*^{n+1} R_n)$. No Markov property of the coarse process is assumed.

Theorem 31.1 (Projection cannot create entropy production)**.** *For every n, $\Sigma_n^c \le \Sigma_n$. More generally, $D(\mu \circ \phi^{-1} \,\|\, \nu \circ \phi^{-1}) \le D(\mu \,\|\, \nu)$ for any measures μ, ν on a finite set and any map ϕ.*

Proof. For a block $B = \phi^{-1}(b)$ the log-sum inequality gives $\sum_{\omega \in B} \mu(\omega) \ln \frac{\mu(\omega)}{\nu(\omega)} \ge \mu(B) \ln \frac{\mu(B)}{\nu(B)}$. Summing over blocks gives the claim. Applying it with $\phi = c^{n+1}$, $\mu = P_n$, $\nu = R_n$ gives the first statement. □

The theorem is the relative-entropy data-processing inequality in path space. The coarse forward and reverse measures in it are pushforwards of P_n and R_n under the same map c^{n+1}, which is all the inequality uses. Reading $c_*^{n+1} R_n$ as the law of the time-reversed observed process is an additional step. In the present setting, where reversal is only the reversal of the order of indices, it holds because reversal commutes with the coordinatewise map. In a physical setting where time reversal acts on states by an involution θ, as it does on momenta, the reversed observed process is the pushforward under $c \circ \theta$, and it agrees with the pushforward under c only if the observation is compatible with time reversal in the sense $c \circ \theta = c$. Under that compatibility the theorem says that a coarser description exhibits no more irreversibility than a finer one. The following example shows the inequality can be strict, with the coarse description exhibiting none.

Example 31.2 (A biased cycle observed through its parity)**.** Let $X = \mathbb{Z}_4$ and let the chain step forward with probability $q \in (\frac{1}{2}, 1)$ and backward with probability $1 - q$. Then π is uniform and

$$\Sigma_1 = (2q - 1) \ln \frac{q}{1 - q} > 0.$$

Let $c(x) = x \bmod 2$. The parity alternates deterministically, since both forward and backward steps change it, so the observed process is the alternating sequence started in either phase with probability $\frac{1}{2}$. The reversal of an alternating sequence is again alternating, and the two phases have equal weight, so $c_*^{n+1} R_n = c_*^{n+1} P_n$ and $\Sigma_n^c = 0$. For $q = 0.8$ one has $\Sigma_1 \approx 0.8318$, and an exact enumeration of all paths of length 4 gives $\Sigma_4 = 3.3271 = 4\Sigma_1$ and $\Sigma_4^c = 0$.

The example shows that irreversibility is a property of a process together with an observation. The biased cycle has a strong direction, and an observer who records only parity cannot detect it. This does not contradict the second law, which concerns the fine dynamics; it records that a coarse description is blind to part of what the fine one carries.

Chapter 32

Thermodynamic Descriptions as Observations

A thermodynamic description summarizes a microscopic system by a small set of macroscopic variables. In the vocabulary of this book it is an observation, and the questions of sufficiency and closure apply to it. This chapter shows that the classical equilibrium description is a minimal sufficient observation in a precise sense and that a second law holds for descriptions that are closed.

Let X be a finite set of microstates and $f\colon X \to \mathbb{R}$ an observable such as the energy. For β in a set $B \subseteq \mathbb{R}$ with at least two elements, let $p_\beta(x) = e^{-\beta f(x)}/Z(\beta)$ be the Gibbs distribution. By Jaynes' principle [32], p_β is the maximum-entropy distribution with the given mean of f. Consider the family $\{p_\beta\}_{\beta\in B}$ in place of the family of futures of Part I: call two microstates *equivalent* if their likelihood functions $\beta \mapsto p_\beta(x)$ are proportional, which is the construction of Chapter 3 with the likelihood in place of the continuation set.

Proposition 32.1 (The energy is the minimal sufficient observation of the Gibbs family)**.** *Two microstates x, y are equivalent for the family $\{p_\beta\}_{\beta\in B}$ if and only if $f(x) = f(y)$. The observation f is sufficient in the sense that the conditional law of the microstate given $f(x) = m$ is uniform on the level set $f^{-1}(m)$ for every β, hence independent of β.*

Proof. The ratio $p_\beta(x)/p_\beta(y) = e^{-\beta(f(x)-f(y))}$ is independent of $\beta \in B$ exactly when $f(x) = f(y)$, because B has two distinct elements. For the second statement, $p_\beta(x \mid f = m) = e^{-\beta m}/\big(|f^{-1}(m)|e^{-\beta m}\big) = 1/|f^{-1}(m)|$. □

The proposition is the sufficiency principle of statistical inference in the language of Part I. Equilibrium thermodynamics works because a one-parameter family of ensembles is determined, on the level of inference about the microstate, by the value of the energy and nothing else. The phrase "nothing else" is the minimality statement.

Sufficiency does not provide dynamics. For the thermodynamic variables to evolve autonomously the description must also be closed. For doubly stochastic microscopic dynamics, which is the reversible stochastic case of Part II, a coarse second law follows from closure.

Theorem 32.2 (Second law for closed descriptions of bistochastic dynamics)**.** *Let M be a doubly stochastic matrix on a finite set X with $|X| = n$, and $c\colon X \to Y$ a surjection whose partition is lumpable, with aggregated matrix N and blocks $B_y = c^{-1}(y)$. For a distribution p on X with block distribution $P = Cp$ define*

$$S(P) = H(P) + \sum_y P_y \ln |B_y|,$$

the Shannon entropy of the block distribution plus the mean Boltzmann entropy of the blocks. Then $S(P) = \ln n - D(P \,\|\, u)$ with $u_y = |B_y|/n$, and $S(NP) \geq S(P)$.

Proof. Direct computation gives $D(P \,\|\, u) = \sum_y P_y \ln \frac{nP_y}{|B_y|} = \ln n - H(P) - \sum_y P_y \ln |B_y| = \ln n - S(P)$. Since M is doubly stochastic the uniform distribution on X is invariant, and its aggregate is u, so $Nu = u$ by $CM = NC$. Relative entropy to an invariant distribution does not increase under a stochastic matrix, hence $D(NP \,\|\, u) = D(NP \,\|\, Nu) \leq D(P \,\|\, u)$, which is the claim. □

The quantity S is the entropy of the uniform-within-block lift of P to X, and its nondecreasing behaviour under a closed coarse dynamics is the classical H-theorem in finite form. Three features should be noted. First, the monotonicity requires closure: for a non-lumpable partition the aggregated evolution depends on the within-block distribution and no function of P alone is guaranteed to increase. Second, when M is a permutation matrix and the partition is lumpable the blocks are mapped to blocks of equal size, and S is constant, so the monotone increase is a property of genuinely stochastic reversible dynamics. Third, the theorem is a statement about the information-theoretic entropy S of a stated description; identifying S with a physical entropy needs the thermodynamic setting of a heat bath and a Hamiltonian, which the theorem does not supply. In the example of Proposition 30.4, started at $P = (0.9, 0.1)$, one finds $S(P) \approx 0.394$ and $S(NP) = \ln 3 \approx 1.099$.

Chapter 33

Fluctuation Relations and Path Ensembles

The path measures of Chapter 31 give the entropy production a pathwise form, and the pathwise form has an exact symmetry [4, 38]. For a path ω with $P_n(\omega) > 0$ define

$$\sigma(\omega) = \ln \frac{P_n(\omega)}{R_n(\omega)} = \ln \frac{\pi_{x_0}}{\pi_{x_n}} + \sum_{k=0}^{n-1} \ln \frac{M_{x_{k+1}x_k}}{M_{x_k x_{k+1}}},$$

so that $\Sigma_n = \mathbb{E}_{P_n}[\sigma]$.

Theorem 33.1 (Fluctuation relations)**.** *Under the support-symmetry assumption, (i) the detailed relation $R_n(\omega) = P_n(\omega)e^{-\sigma(\omega)}$ holds, (ii) the integral relation $\mathbb{E}_{P_n}[e^{-\sigma}] = 1$ holds, and (iii) consequently $\Sigma_n = \mathbb{E}_{P_n}[\sigma] \geq 0$, with equality if and only if $\sigma \equiv 0$ on the support.*

Proof. Part (i) is the definition of σ. For (ii), $\mathbb{E}_{P_n}[e^{-\sigma}] = \sum_{\omega : P_n(\omega) > 0} R_n(\omega) = 1$, because the support of R_n lies within the support of P_n by symmetry of the transition support. For (iii), Jensen's inequality for the convex function $t \mapsto e^{-t}$ gives $1 = \mathbb{E}[e^{-\sigma}] \geq e^{-\mathbb{E}\sigma}$, hence $\mathbb{E}\sigma \geq 0$, with equality only if σ is almost surely constant, necessarily zero. □

Example 33.2 (The biased cycle)**.** In Example 31.2, with π uniform the boundary term vanishes and $\sigma(\omega) = (N_+ - N_-)\lambda$ with $\lambda = \ln \frac{q}{1-q}$, where $N_\pm$ count forward and backward steps. The integral relation reduces to the per-step identity $q \cdot \frac{1-q}{q} + (1-q) \cdot \frac{q}{1-q} = 1$, and the mean per step is $(q - (1-q))\lambda = (2q-1)\lambda = \Sigma_1$. Negative values of σ occur, on paths with more backward than forward steps, and they are exponentially rarer than the positive values of the same magnitude by the detailed relation.

The pathwise relation allows the effect of an observation to be stated at the level of individual coarse paths, which refines Theorem 31.1.

Theorem 33.3 (Coarse fluctuation relation)**.** *Let ω' be a coarse path with $P'_n(\omega') = c_*^{n+1} P_n(\omega') > 0$, and set $\sigma'(\omega') = \ln\bigl(P'_n(\omega')/R'_n(\omega')\bigr)$ with $R'_n = c_*^{n+1} R_n$. Then*

$$e^{-\sigma'(\omega')} = \mathbb{E}_{P_n}\bigl[e^{-\sigma} \,\big|\, \omega'\bigr], \qquad \textit{hence} \qquad \sigma'(\omega') \le \mathbb{E}_{P_n}\bigl[\sigma \,\big|\, \omega'\bigr].$$

The coarse process satisfies its own integral relation $\mathbb{E}_{P'_n}[e^{-\sigma'}] = 1$.

Proof. Write $R'_n(\omega') = \sum_{\omega \mapsto \omega'} R_n(\omega) = \sum_{\omega \mapsto \omega'} P_n(\omega) e^{-\sigma(\omega)}$. Dividing by $P'_n(\omega') = \sum_{\omega \mapsto \omega'} P_n(\omega)$ gives the conditional expectation of $e^{-\sigma}$. The inequality is Jensen's inequality for the conditional law. The coarse integral relation is Theorem 33.1(ii) applied to the push-forward measures. □

Averaging the pathwise inequality over P'_n recovers Theorem 31.1. The refinement shows where the apparent production is lost: on each coarse path it equals minus the logarithm of the average of $e^{-\sigma}$ over the fine paths compatible with it, which is smaller than their average of σ whenever those fine paths differ in σ. Heterogeneity among the hidden refinements of a coarse path is what the observer fails to see.

Chapter 34

Non-Markovian Dynamics and Retained Variables

An observed process that is not Markov can still be the projection of a Markov process on a larger state. The question is how much must be retained. Part I answers with the minimal sufficient observation, and the answer is a Markov process by construction.

Theorem 34.1 (Markov embedding by the minimal state)**.** *Let (X_t) be a stationary process over a finite alphabet, and let $s_t = \epsilon(X_{\le t})$ be its minimal probabilistically sufficient state in the sense of Chapter 8, so that s_t determines the law of the future and s_{t+1} is a function of s_t and X_{t+1}. Then the joint process (s_t, X_{t+1}) is Markov, and (X_t) is its projection onto the second component.*

Proof. By sufficiency, the conditional law of X_{t+1} given the entire past depends only on s_t. By the extension compatibility of Theorem 6.2 in its probabilistic form, $s_{t+1} = \Phi(s_t, X_{t+1})$ for a deterministic map Φ. Therefore the law of (s_{t+1}, X_{t+2}) given the whole past depends only on (s_t, X_{t+1}), which is the Markov property. □

The even process of Example 18.1 illustrates the situation completely. It has no finite Markov order, since no window of final symbols determines the next-symbol law, and yet it is the projection of a Markov chain on two retained states. In its stationary version the weights are found from $\pi_A = \frac{1}{2}\pi_A + \pi_B$ and $\pi_B = \frac{1}{2}\pi_A$, which give $\pi_A = \frac{2}{3}$ and $\pi_B = \frac{1}{3}$. The entropy of the retained state, the statistical complexity of computational mechanics [6, 39], is then

$$C = -\tfrac{2}{3}\ln\tfrac{2}{3} - \tfrac{1}{3}\ln\tfrac{1}{3} \approx 0.6365 \text{ nats},$$

while the entropy rate, the uncertainty per symbol remaining after the state is known, is $h = \frac{2}{3}\ln 2 \approx 0.4621$ nats, since only the state A has a random output. The retained state

therefore carries more information than the process generates per step, and the excess is the price of the long-range parity constraint.

The memory-kernel equation of Chapter 13 expresses the same fact in linear terms. The observed populations x_t do not form a closed system when the kernel is nonzero, while the pair (x_t, y_t) of populations and the retained coherences evolves under a single linear map. The retained variables of the Markov embedding are the unobserved block of the projection, and the kernel is what remains of that block after elimination. For the Hadamard evolution of Example 13.4 the kernel lag is one.

The retained variables are a mathematical device, and the cost of keeping them depends on their physical realization. In the biased cycle of Example 31.2 the parity-only description is closed and shows no irreversibility, but it does not retain the position, and the irreversible circulation is a property of the retained variable. The relation between the size of the retained state, the entropy production visible to an observer who holds it, and the work needed to maintain it is an area of active study in stochastic thermodynamics [38]; the present book records only the qualitative connection. Retaining the variables that make a description closed and sufficient restores the entropy production that projection hid, and it does so as a matter of mathematics, with no input from physical dynamics.

Chapter 35

Arrows of Time and Admissible Orderings

The history structures of Part I are directed from the start. The extension operation $h \mapsto h \cdot s$ is defined in one direction, and the set of histories is partially ordered by the prefix relation. Reversing the order requires a second structure, which for a stochastic process is the time-reversed path measure of Chapter 31. An *arrow of time* for an observation is the statistical asymmetry between its forward and reversed laws.

The asymmetry has an operational meaning. Suppose an observer is given a path and told only that it is a sample either of the forward measure P_n or of the reversed measure R_n, and is asked to decide which. The statement can be made precise under the hypotheses of the earlier sections: a finite state space, an irreducible stochastic matrix M with stationary distribution π, started in π, and the support condition that $M_{yx} > 0$ if and only if $M_{xy} > 0$. For a threshold $\epsilon \in (0,1)$ let $\beta_n(\epsilon) = \min\{R_n(A) : P_n(A) \geq 1 - \epsilon\}$, the smallest probability of declaring a reversed path forward-directed among tests that recognize a forward path with probability at least $1 - \epsilon$.

Theorem 35.1 (Stein exponent of the arrow of time)**.** *Under these hypotheses,* $-\frac{1}{n} \ln \beta_n(\epsilon) \to \Sigma_1$ *for every* $\epsilon \in (0,1)$.

Proof. The log-likelihood ratio is $\ln \frac{P_n}{R_n}(\omega) = \ln \frac{\pi_{x_0}}{\pi_{x_n}} + \sum_{k<n} \ln \frac{M_{x_{k+1}x_k}}{M_{x_k x_{k+1}}}$. The boundary term is bounded, and the sum is an additive functional of a finite irreducible chain, so by the ergodic theorem $\frac{1}{n} \ln \frac{P_n}{R_n} \to \Sigma_1$ in P_n-probability. For achievability, fix $\delta > 0$ and take $A_n = \{\frac{1}{n} \ln \frac{P_n}{R_n} \geq \Sigma_1 - \delta\}$. Then $P_n(A_n) \to 1$, and $R_n(A_n) \leq e^{-n(\Sigma_1 - \delta)} P_n(A_n) \leq e^{-n(\Sigma_1 - \delta)}$. For the converse, let $P_n(A) \geq 1 - \epsilon$ and put $G_n = \{\frac{1}{n} \ln \frac{P_n}{R_n} \leq \Sigma_1 + \delta\}$. Then $P_n(A \cap G_n) \geq 1 - \epsilon - o(1)$ and $R_n(A) \geq R_n(A \cap G_n) \geq e^{-n(\Sigma_1 + \delta)} P_n(A \cap G_n)$. Taking $n \to \infty$ and then $\delta \to 0$ gives the claim. $\square$

Thus Σ_1 is the rate at which the direction of time becomes detectable. An observer restricted to a coarse description cannot do better, because any test that uses only the coarse path is an event of the form $c^{-1}(A')$ for the fine path, and has the same probabilities under P_n and R_n as A' has under their pushforwards. The Stein exponent of the coarse observer, when it exists, is therefore at most Σ_1. This is consistent with Theorem 31.1, and it gives that theorem an operational reading.

Proposition 35.2 (The arrow is relative to the observation)**.** *Let Σ_1^c denote the rate of entropy production of an observation c when it exists. Then $0 \le \Sigma_1^c \le \Sigma_1$, and each bound can be attained. In Example 31.2, the identity observation attains $\Sigma_1 = (2q-1)\ln\frac{q}{1-q}$ and the parity observation attains 0.*

Proof. The inequalities are Theorem 31.1 applied to path measures of length n and divided by n. The two values for the example were computed in Example 31.2. □

Two consequences deserve emphasis. A process can have a pronounced arrow at one level of description and none at a coarser one, and it is meaningless to ask whether a process "has" an arrow without specifying what is observed. And the arrow in the sense of path-measure asymmetry must be kept apart from logical irreversibility, the non-injectivity of an induced map. A reversible fine dynamics can have a logically irreversible coarse description, as in Proposition 30.2, where two states of the description are sent to one: that is a loss of information about the past and not an asymmetry of probabilities. The two notions are independent, and Proposition 51.3 exhibits all four combinations. For a deterministic dynamics the path-measure notion is degenerate: the reversed path of a nonconstant orbit has probability zero under the forward measure, so the support condition $M_{yx} > 0 \Rightarrow M_{xy} > 0$ fails, the relative entropy Σ_n is infinite, and the direction is perfectly detectable from a single transition. The framework of Theorem 35.1 therefore concerns stochastic dynamics with symmetric transition support.

The forward and reverse descriptions of a process each have a minimal sufficient state, and these need not coincide. Ellison, Mahoney and Crutchfield [5] show that the forward and reverse causal states of a stationary process can differ in number and in the information they store, so the cost of predicting a process and the cost of retrodicting it are different quantities. In the language of Part I this is the statement that the forward and reversed history structures of one process have different minimal quotients. Whether such asymmetries are related to the entropy production Σ_1 is a question the framework poses and this book does not answer. It is recorded among the open problems of Part VIII together with the corresponding question for the geometric holonomy of Part IV, which has been shown above to be independent of entropy production unless a thermodynamic dynamics is specified.

Part VI

Continuum Fields and the RSVP Construction

This part constructs a continuum field description from the history framework and develops one model in full. The organizing decision is to separate three layers that are easily run together. The first layer is a definition of the fields from a family of probability measures on histories, with conservation treated as an assumption that has consequences and not as a consequence of a name. The second layer is a closed dynamics for those fields, written so that the reversible transport and the irreversible relaxation are generated by different structures and so that every property claimed for it can be verified. The third layer comprises the extensions: an effective propagation metric extracted from the equations, a bridge to unitary evolution, and a gravitational sector with its own metric and connection. Each extension is stated with the additional structure it needs, and any identification among the layers is labelled as a conjecture. The fields are the scalar capacity Φ, the directed flow $\mathbf{v}$ and the entropy S, in the notation of the author's RSVP programme. The results of this part concern the specific model constructed here and do not assert that any earlier formulation of that programme reduces to it.

Chapter 36

From Histories to Fields

Let $(\mathcal{H}, \mathcal{F}, \mathcal{A})$ be a probabilistic history structure as in Chapter 8, realized as a Markov chain or a continuous-time Markov process on a configuration space Ω of a spatially extended system, and let $\mathbb{P}$ be a family of path measures. A spatial observable is a collection of additive quantities $C_i(\eta)$ attached to sites i of a lattice, for example the number of particles at site i. Fix a coarse-graining scale ℓ, a partition of the lattice into blocks B_I of linear size ℓ, and set

$$\Phi_\ell(I,t) = \frac{1}{|B_I|}\,\mathbb{E}\Big[\sum_{i\in B_I} C_i(\eta_t)\Big].$$

This is the *capacity density* at scale ℓ. It is operationally defined by an additive observable and an expectation, and it carries no information beyond them.

Suppose the capacity is conserved microscopically, in the sense that over each time step the change at a site equals the net inflow across its two bonds:

$$C_i(\eta_{t+\tau}) - C_i(\eta_t) = J_{i-1/2} - J_{i+1/2},$$

where $J_{i+1/2}$ is the net transfer from site i to site $i+1$ during the step. This is an assumption about the model and not a consequence of any definition. Its consequence is exact and requires no limit.

Proposition 36.1 (Coarse-graining preserves the balance law)**.** *Under the microscopic balance above, the block content $\Phi_\ell(I,t)|B_I|$ satisfies*

$$\big(\Phi_\ell(I,t+\tau) - \Phi_\ell(I,t)\big)|B_I| = \mathbb{E}[J_{\partial^- I}] - \mathbb{E}[J_{\partial^+ I}],$$

where $\partial^\mp I$ are the left and right boundary bonds of the block. In particular the coarse capacity obeys a continuity equation with a flux that is the expected transfer across the block boundary.

Proof. Sum the microscopic balance over the sites of the block. The interior fluxes cancel in pairs, leaving the two boundary fluxes. Take expectations. □

The proposition states what conservation gives for free. It does not state that the flux is a function of the coarse capacities, and that is the closure problem treated in Chapter 38. The *directed flow* is defined from the flux by the operational rule

$$\mathbf{v}_\ell = \frac{\mathbf{j}_\ell}{\Phi_\ell}, \qquad \Phi_\ell > 0,$$

so that $\mathbf{v}_\ell$ is a transport velocity of the capacity and not an independent primitive. With this definition the continuity equation takes the conservative form $\partial_t \Phi + \nabla \cdot (\Phi \mathbf{v}) = 0$ in the continuum limit, whenever the limit exists. If the microscopic model permits creation, destruction, or exchange of the capacity, a source term must be retained.

Example 36.2 (Symmetric exclusion: a closed capacity field with a diffusive velocity)**.** Let particles on a ring of N sites obey the symmetric exclusion process in which each bond $\{i, i+1\}$ exchanges the contents of its two sites at rate 1. The generator applied to the occupation variable gives $L\eta_i = (\eta_{i+1} - \eta_i) + (\eta_{i-1} - \eta_i)$, so that

$$\frac{d}{dt}\,\mathbb{E}[\eta_i] = \mathbb{E}[\eta_{i+1}] + \mathbb{E}[\eta_{i-1}] - 2\mathbb{E}[\eta_i]$$

exactly, for every initial law (Liggett [23]). With $\Phi_i = \mathbb{E}[\eta_i]$ the flux across a bond is $J_{i+1/2} = \Phi_i - \Phi_{i+1}$, a function of the capacities, so the field is exactly closed at the level of the mean. The velocity defined by the rule above is $v_{i+1/2} = -(\Phi_{i+1} - \Phi_i)/\Phi_{i+1/2}$, a discrete form of $v = -\partial_x \ln \Phi$ for the interface density $\Phi_{i+1/2} = \frac{1}{2}(\Phi_i + \Phi_{i+1})$, and it is a diffusive velocity with no ballistic part. The hydrodynamic limit of the process, in which the empirical density converges to a solution of the heat equation, is classical [22].

The example shows that the operational velocity can be an entirely diffusive quantity, and that the closure of the mean field holds for this model for a special reason, the self-duality of the exclusion process. The correlation functions of each order also satisfy closed equations among correlation functions of the same order, but those of order two are not functions of the mean field alone, and for generic dynamics even the mean-field equation does not close. The general situation is the one described in Chapter 12: the coarse fields are closed only if the partition is lumpable, and otherwise the dynamics of the fields carries memory or requires additional variables.

Chapter 37

Scalar Capacity, Directed Flow, and Entropy

The three fields of the construction are the capacity density Φ, the momentum density $\mathbf{m} = \Phi \mathbf{v}$, and an entropy density S. The first two have been defined in the previous chapter from conservation and flux. The entropy needs a separate construction, because several quantities carry the name and are not interchangeable.

The first is the logarithm of the number of continuations compatible with a coarse description, a counting entropy of the kind used in Chapter 32. The second is a Shannon entropy of a law of futures, as in the statistical complexity of Chapter 34. The third is a thermodynamic entropy density defined through a local equilibrium family. These are different functionals with different monotonicity properties, and the arguments of this part do not transfer monotonicity from one to another.

For the first closed model the entropy is defined thermodynamically. Let $\varepsilon(\Phi, S)$ be an internal energy density, a function on the space of local equilibrium states, and define the temperature, the chemical potential and the pressure by

$$\vartheta = \varepsilon_S, \qquad \varepsilon_\Phi = \partial_\Phi \varepsilon, \qquad p = \Phi \varepsilon_\Phi + S\vartheta - \varepsilon.$$

The family of local equilibrium states is the family of Gibbs states of Chapter 32 indexed by (Φ, S) and the velocity. The continuation entropy of the earlier chapters remains available as a second functional, and its relation to S is a question to be investigated for each model; the dynamics below involves only S.

The earlier RSVP equation $\partial_t \Phi + \mathbf{v} \cdot \nabla \Phi = -\lambda \nabla \cdot \mathbf{v}$ [18] is kept here in its original form as a historical model, and the conservative equation is introduced separately. On a periodic

domain the historical model gives

$$\frac{d}{dt}\int \Phi \, dx = \int \Phi \, \nabla\cdot \mathbf{v} \, dx - \lambda \int \nabla\cdot \mathbf{v} \, dx,$$

because $-\int \mathbf{v}\cdot\nabla\Phi = \int \Phi\nabla\cdot\mathbf{v}$ by integration by parts. For a constant λ the last term vanishes, leaving $\int \Phi\nabla\cdot\mathbf{v}$, which is not zero in general, so the historical model does not conserve $\int\Phi$ in general; whether a particular model does depends on its complete form and its boundary conditions. The conservative equation $\partial_t\Phi + \nabla\cdot(\Phi\mathbf{v}) = 0$ has the advective form $\partial_t\Phi + \mathbf{v}\cdot\nabla\Phi = -\Phi\,\nabla\cdot\mathbf{v}$, which is the historical equation with the constant coefficient replaced by the field Φ. That replacement is a change of the dynamics and not a rewriting of it, and the two models should not be identified. The model below takes conservation as an explicit assumption. Whether other couplings have a useful physical or informational interpretation is a separate question that this part does not decide.

Chapter 38

Closure of the Field Equations

Closure is the question whether the selected fields determine their own future, and it is the content of Chapters 12 and 13 applied to the continuum. A local field equation can arise in one of two ways. It can be exact in a restricted model, as in Example 36.2, or it can be an approximation in which the dependence of the flux on the discarded degrees of freedom has been replaced by its value in the local equilibrium family. The second case is a named approximation: the first-order Chapman–Enskog expansion about local equilibrium produces viscous and heat-conduction terms proportional to gradients, and the discarded dependence is the memory term of Theorem 13.1, which is of higher order in the ratio of the microscopic relaxation time to the macroscopic time under the standard scaling assumptions of the expansion, assumptions that are not verified here.

The model constructed here is the one-dimensional compressible conducting fluid on a circle $\mathbb{S}^1$ of length ℓ_0, in the variables $z = (\Phi, m, S)$ with $m = \Phi v$ and $e = \frac{m^2}{2\Phi} + \varepsilon(\Phi, S)$ the total energy density. The equations are

$$\partial_t \Phi + \partial_x(\Phi v) = 0, \tag{38.1}$$

$$\partial_t m + \partial_x(mv + p) = \partial_x(\mu\, \partial_x v), \tag{38.2}$$

$$\partial_t S + \partial_x(Sv) = \frac{1}{\vartheta}\Big[\mu(\partial_x v)^2 + \partial_x(\kappa\, \partial_x \vartheta)\Big], \tag{38.3}$$

with viscosity $\mu \geq 0$ and thermal conductivity $\kappa \geq 0$, and with $\vartheta > 0$. The first two terms on the left of each equation are the reversible transport of Chapter 36; the right-hand sides are the first-order relaxation terms.

The structure of the model is displayed by writing it in the form

$$\partial_t z = \mathcal{L}(z)\, \frac{\delta E}{\delta z} + \mathcal{M}(z)\, \frac{\delta \mathcal{S}}{\delta z}, \qquad E = \int e\, dx, \quad \mathcal{S} = \int S\, dx, \tag{38.4}$$

which is the GENERIC structure of Grmela and Öttinger [19, 26]. The operators have to be supplied and their properties verified, which the rest of this chapter does.

Definition 38.1 (The reversible operator)**.** For functionals F, G of $z = (\Phi, m, S)$ define

$$\{F,G\} = -\int \Big[\Phi\big(F_m\,\partial_x G_\Phi - G_m\,\partial_x F_\Phi\big) + m\big(F_m\,\partial_x G_m - G_m\,\partial_x F_m\big) + S\big(F_m\,\partial_x G_S - G_m\,\partial_x F_S\big)\Big]\,dx,$$

and let $\mathcal{L}$ be the antisymmetric operator with $\langle \delta F, \mathcal{L}\,\delta G\rangle = \{F, G\}$.

Theorem 38.2 (The reversible part is a Lie–Poisson system)**.** *The bracket $\{\cdot,\cdot\}$ is antisymmetric and satisfies the Jacobi identity. It generates the reversible parts of Equations* (38.1)–(38.3)*: $\mathcal{L}\,\delta E/\delta z = \big(-\partial_x(\Phi v),\, -\partial_x(mv+p),\, -\partial_x(Sv)\big)$. The functionals $N = \int \Phi\,dx$ and $\mathcal{S} = \int S\,dx$ are Casimirs, so $\mathcal{L}\,\delta\mathcal{S}/\delta z = 0$.*

Proof. Antisymmetry is visible. For the Jacobi identity, let $\mathfrak{g} = \mathrm{Vect}(\mathbb{S}^1) \ltimes \big(C^\infty(\mathbb{S}^1) \oplus C^\infty(\mathbb{S}^1)\big)$ be the semidirect product of the vector fields on the circle with two copies of the functions on which they act by differentiation, with bracket

$$[(X,f,h),(Y,f',h')] = \big([X,Y],\, Xf' - Yf,\, Xh' - Yh\big), \qquad [X,Y] = XY' - YX'.$$

This is a Lie algebra, because the action of vector fields on functions is a representation of Vect on a vector space. The dual of $\mathfrak{g}$ is the space of triples (m, Φ, S) of a momentum density and two densities, and the Lie–Poisson bracket $\{F,G\} = -\langle z, [\delta F, \delta G]\rangle$ with $\delta F = (F_m, F_\Phi, F_S)$ expands to the displayed formula. The Lie–Poisson bracket of the dual of any Lie algebra satisfies the Jacobi identity, because it inherits it from the Jacobi identity of $\mathfrak{g}$ [21, 25]. For the generation of the equations, take $F = \int f\,\Phi$ with a test function f, and use $E_m = v$. The first component is $-\partial_x(\Phi E_m) = -\partial_x(\Phi v)$. For the momentum component take $F = \int f\,m$ and use $E_\Phi = -\frac{v^2}{2} + \varepsilon_\Phi$ and $E_S = \vartheta$:

$$-\Phi\partial_x E_\Phi - m\,\partial_x E_m - S\,\partial_x E_S - \partial_x(mE_m) = -\Phi\varepsilon_\Phi' - S\vartheta' - \partial_x(\Phi v^2) = -\partial_x p - \partial_x(\Phi v^2),$$

because $p_x = \Phi\varepsilon_{\Phi,x} + S\vartheta_x$ by $\varepsilon_x = \varepsilon_\Phi\Phi_x + \vartheta S_x$. The entropy component is $-\partial_x(SE_m) = -\partial_x(Sv)$. For the Casimirs, $\delta N = (0,1,0)$ and $\delta\mathcal{S} = (0,0,1)$ are constant, and each term of the bracket with $G = N$ or $\mathcal{S}$ contains either a derivative of a component of δG, which vanishes because that component is constant, or the component G_m, which is zero, so the bracket with them vanishes. □

Definition 38.3 (The dissipative operator)**.** Let $\mu, \kappa \geq 0$ and $\vartheta > 0$. Define $\mathcal{M}(z)$ by the symmetric bilinear form on covectors $a = (a_\Phi, a_m, a_S)$

$$\langle a, \mathcal{M}\,b\rangle = \int \Big[\mu\vartheta\,\chi(a)\chi(b) + \kappa\vartheta^2\,\zeta(a)\zeta(b)\Big]\,dx, \qquad \chi(a) = \partial_x a_m - \frac{\partial_x v}{\vartheta}\,a_S, \quad \zeta(a) = \partial_x\Big(\frac{a_S}{\vartheta}\Big).$$

Theorem 38.4 (The dissipative part)**.** *The operator $\mathcal{M}$ is symmetric and positive semidefinite, and it satisfies the degeneracy condition $\mathcal{M}\,\delta E/\delta z = 0$. It generates the relaxation terms of the model: $\mathcal{M}\,\delta\mathcal{S}/\delta z = \big(0,\ \partial_x(\mu\partial_x v),\ \frac{1}{\vartheta}[\mu(\partial_x v)^2 + \partial_x(\kappa\partial_x\vartheta)]\big)$. Moreover $\langle \delta\mathcal{S}, \mathcal{M}\,\delta\mathcal{S}\rangle = \int \sigma\,dx$ with $\sigma = \mu(\partial_x v)^2/\vartheta + \kappa(\partial_x\vartheta)^2/\vartheta^2 \geq 0$.*

Proof. Symmetry and semidefiniteness are immediate from the form of the bilinear expression. The energy covector is $\delta E = (E_\Phi, E_m, E_S) = (-\frac{v^2}{2} + \varepsilon_\Phi,\ v,\ \vartheta)$, so $\chi(\delta E) = \partial_x v - \frac{\partial_x v}{\vartheta}\vartheta = 0$ and $\zeta(\delta E) = \partial_x(1) = 0$, which is the degeneracy. For the entropy covector $\delta\mathcal{S} = (0, 0, 1)$ one has $\chi = -\partial_x v/\vartheta$ and $\zeta = -\partial_x\vartheta/\vartheta^2$. Then, for any test covector a,

$$\begin{aligned}\langle a, \mathcal{M}\,\delta\mathcal{S}\rangle &= \int \Big[-\mu\,\partial_x v\,\partial_x a_m + \frac{\mu(\partial_x v)^2}{\vartheta} a_S - \kappa\,\partial_x\vartheta\,\partial_x\Big(\frac{a_S}{\vartheta}\Big)\Big]\,dx \\ &= \int \Big[a_m\,\partial_x(\mu\partial_x v) + \frac{a_S}{\vartheta}\big(\mu(\partial_x v)^2 + \partial_x(\kappa\partial_x\vartheta)\big)\Big]\,dx\end{aligned}$$

where the second line follows by integration by parts, and it identifies $\mathcal{M}\,\delta\mathcal{S}$. Taking $a = \delta\mathcal{S}$ gives the entropy production. □

Remark 38.5. The dissipative operator was constructed by writing the standard Navier–Stokes–Fourier dissipation in the variables (Φ, m, e), where the entropy gradient is $(\,(v^2/2 - \varepsilon_\Phi)/\vartheta,\ -v/\vartheta,\ 1/\vartheta\,)$, and transforming the bilinear form to the variables (Φ, m, S) by the chain rule. The form is covariant under this change of variables, which is a point transformation without derivatives, and the same is true of the Poisson structure. The operator $\mathcal{M}$ is required to be symmetric and positive semidefinite and to satisfy the degeneracy condition; no Jacobi-type identity is required of it.

Chapter 39

Conservation Laws and Stability

The structure of the previous chapter implies the conservation and monotonicity properties of the model by short calculations.

Theorem 39.1 (Balance laws)**.** *For smooth solutions of* (38.1)–(38.3) *on the circle,* $N = \int \Phi\,dx$*, the momentum* $P = \int m\,dx$*, and the total energy* $E = \int e\,dx$ *are constant in time, and*

$$\frac{d}{dt}\int S\,dx = \int \sigma\,dx \geq 0.$$

Proof. Mass and momentum are conserved because their equations are divergences. For the energy, differentiate $e = \frac{m^2}{2\Phi} + \varepsilon$ along the equations to obtain $\partial_t e = v\,\partial_t m - \frac{v^2}{2}\partial_t\Phi + \varepsilon_\Phi \partial_t \Phi + \vartheta\,\partial_t S$. The terms coming from the reversible parts combine to $-\partial_x\big((e+p)v\big)$, which is an identity in $\varepsilon_x = \varepsilon_\Phi \Phi_x + \vartheta S_x$ and which was checked symbolically for an arbitrary smooth $\varepsilon(\Phi, S)$. The relaxation terms contribute $v\,\partial_x(\mu v_x) + \mu v_x^2 + \partial_x(\kappa\vartheta_x) = \partial_x(\mu v v_x) + \partial_x(\kappa\vartheta_x)$, a divergence. Hence $\partial_t e$ is a divergence and E is conserved. The entropy statement is the integral of Equation (38.3) rewritten as $\partial_t S + \partial_x(Sv - \kappa\vartheta_x/\vartheta) = \sigma$, using $\partial_x(\kappa\vartheta_x)/\vartheta = \partial_x(\kappa\vartheta_x/\vartheta) + \kappa\vartheta_x^2/\vartheta^2$. □

A numerical integration of the system with ideal-gas internal energy, a pseudo-spectral discretization with 128 modes and a fourth-order Runge–Kutta step, started from smooth nonconstant data, reproduces the theorem: over the interval $0 \leq t \leq 2$ the drifts of mass, momentum and energy are of the order of 10^{-15}, and the entropy, sampled at ten times, increases at every sample.

Linear stability about the uniform rest state follows from the structure and can be stated exactly. Let $(\Phi_0, 0, S_0)$ be a constant state with temperature ϑ_0, and let $H = \mathrm{Hess}_{(\Phi,S)}\,\varepsilon$ be the Hessian of the internal energy at that state.

Theorem 39.2 (Linear stability of the rest state)**.** *Assume that* H *is positive definite, which is the condition of thermodynamic stability of the local equilibrium family, and that* $\mu > 0$

and $\kappa > 0$. Let $(\delta\Phi, \delta v, \delta S)$ be a solution of the linearized system on the circle, and set $\delta g = \varepsilon_{\Phi\Phi}\delta\Phi + \varepsilon_{\Phi S}\delta S$, $\delta\vartheta = \varepsilon_{\Phi S}\delta\Phi + \varepsilon_{SS}\delta S$. Then

$$Q = \int \Big[\tfrac{1}{2}\Phi_0\, \delta v^2 + \tfrac{1}{2}\big(\delta\Phi, \delta S\big) H \big(\delta\Phi, \delta S\big)^{\mathsf{T}} \Big] dx$$

is nonnegative and satisfies

$$\frac{dQ}{dt} = -\int \Big[\mu\, (\partial_x \delta v)^2 + \frac{\kappa}{\vartheta_0} (\partial_x \delta\vartheta)^2 \Big] dx \le 0.$$

Moreover, for every wavenumber $q \neq 0$ the linearization at q is a 3×3 complex matrix all of whose eigenvalues satisfy $\operatorname{Re}\lambda < 0$.

Proof. The linearized equations are $\partial_t \delta\Phi = -\Phi_0 \partial_x \delta v$, $\Phi_0 \partial_t \delta v = -\partial_x(\Phi_0 \delta g + S_0 \delta\vartheta) + \mu \partial_x^2 \delta v$ and $\partial_t \delta S = -S_0 \partial_x \delta v + (\kappa/\vartheta_0)\partial_x^2 \delta\vartheta$, where the linearized pressure gradient is $\partial_x(\Phi_0 \delta g + S_0 \delta\vartheta)$. Differentiating Q and using $\frac{d}{dt}\frac{1}{2}(\delta\Phi, \delta S) H (\delta\Phi, \delta S)^{\mathsf{T}} = \delta g\, \partial_t \delta\Phi + \delta\vartheta\, \partial_t \delta S$ gives

$$\frac{dQ}{dt} = \int \Big[-\Phi_0 \big(\delta v\, \delta g_x + \delta g\, \delta v_x\big) - S_0 \big(\delta v\, \delta\vartheta_x + \delta\vartheta\, \delta v_x\big) + \mu\, \delta v\, \delta v_{xx} + \frac{\kappa}{\vartheta_0} \delta\vartheta\, \delta\vartheta_{xx} \Big] dx.$$

The first two groups are total derivatives and integrate to zero; the last two integrate by parts to the stated dissipation. The inequality alone does not give strict decay of every nonconstant mode, so the spectral argument is made separately. Fix $q \neq 0$, let u be an eigenvector of the linearization at q with eigenvalue λ, and let $Q(u)$ be the Hermitian form with the same expression as Q, which is positive for $u \neq 0$ since H and Φ_0 are positive. Along the solution $e^{\lambda t}u$ one has $Q = e^{2\operatorname{Re}\lambda\, t} Q(u)$, so $2 \operatorname{Re}\lambda\, Q(u)$ equals the negative of the dissipation of u, which is nonnegative, and hence $\operatorname{Re}\lambda \le 0$. Suppose $\operatorname{Re}\lambda = 0$. Then the dissipation of u vanishes; as $\mu > 0$ and $\kappa > 0$ this forces $q\, \delta v = 0$ and $q\, \delta\vartheta = 0$, hence $\delta v = 0$ and $\delta\vartheta = 0$ since $q \neq 0$. The momentum equation at wavenumber q then gives $iq\, \Phi_0\, \delta g = 0$, so $\delta g = 0$. Since H is nonsingular, $\delta g = \delta\vartheta = 0$ implies $\delta\Phi = \delta S = 0$, contradicting $u \neq 0$. Hence $\operatorname{Re}\lambda < 0$. □

Remark 39.3. The hypotheses are used. If $\kappa = 0$, the state with $\delta v = 0$ and $\Phi_0 \delta g + S_0 \delta\vartheta = 0$, a perturbation of the entropy at constant pressure, is a nonzero stationary solution at every wavenumber, an eigenvalue $\lambda = 0$ with $\operatorname{Re}\lambda = 0$. Positivity of H and a decreasing Q alone therefore do not exclude undamped modes; the exclusion needs the transport coefficients to be positive. The theorem concerns the linearization about the uniform rest state, and the numerical comparison below supports only the ideal-gas instance.

For the ideal gas $\varepsilon = \Phi c_v \vartheta$ with specific heats c_v, c_p per unit capacity and ϑ determined by the entropy through $S/\Phi = c_v \ln\vartheta - R \ln\Phi$ up to a constant, the eigenvalues at small wavenumber can be compared with the classical dispersion relation of a damped sound wave.

With $c_v = 2.5$, $c_p = 3.5$ ($R = 1$), $\Phi_0 = 1.3$, $S_0 = 2$, $\mu = 0.05$ and $\kappa = 0.08$, direct numerical diagonalization of the three-by-three matrix at $q = 0.01, 0.02, 0.04$ gives sound eigenvalues $\lambda = \pm i\, c\, q - \Gamma q^2$ with $c^2 = \gamma R \vartheta_0$ and

$$\Gamma = \frac{1}{2\Phi_0}\Big[\mu + \kappa\Big(\frac{1}{c_v} - \frac{1}{c_p}\Big)\Big],$$

with relative error of the eigenvalue 6×10^{-9} at the smallest wavenumber $q = 0.01$, growing as q^2 (sixteen times larger at $q = 0.04$) as expected for the neglected higher-order terms in the expansion. The decay of all non-constant modes was also confirmed numerically on the grid $0.05 \le q \le 5$. The nonlinear stability of the rest state and the long-time behaviour of the nonlinear system are not addressed here.

Chapter 40

Effective Metrics From Field Data

Two constructions that are often described by the same phrase should be kept apart. The first extracts from the field equations a metric governing the propagation of small disturbances. The second introduces a metric as an independent field of a gravitational theory. This chapter treats the first.

Write the nondissipative part of the model in the variables $u = (\Phi, v, S)$ as a first-order system $\partial_t u + A(u)\partial_x u = 0$. Direct computation gives

$$A = \begin{pmatrix} v & \Phi & 0 \\ p_\Phi/\Phi & v & p_S/\Phi \\ 0 & S & v \end{pmatrix}.$$

Proposition 40.1 (Characteristics and hyperbolicity)**.** *The eigenvalues of A are v and $v \pm c$, with $c^2 = p_\Phi + \frac{S}{\Phi}\, p_S = \frac{1}{\Phi}\, w^{\mathsf{T}} H w$ and $w = (\Phi, S)^{\mathsf{T}}$. If H is positive definite then $c^2 > 0$ and the system is strictly hyperbolic when $c \neq 0$.*

Proof. Subtract vI and write μ for the eigenvalue of the result. The characteristic polynomial is $-\mu\big(\mu^2 - p_\Phi - \frac{S}{\Phi} p_S\big)$. For the second form, $p_\Phi = \Phi\varepsilon_{\Phi\Phi} + S\varepsilon_{\Phi S}$ and $p_S = \Phi\varepsilon_{\Phi S} + S\varepsilon_{SS}$ so that $\Phi c^2 = \Phi^2\varepsilon_{\Phi\Phi} + 2\Phi S\varepsilon_{\Phi S} + S^2\varepsilon_{SS}$. □

For the ideal gas the formula gives $c^2 = \gamma R\vartheta$, and the numerical eigenvalues of A at a test state confirm the speeds $v - c, v, v + c$ both at rest and for a flow of speed 0.4. Three observations limit the use of this result.

First, in one space dimension the characteristic directions $dx/dt = v \pm c$ are the null directions of any two-dimensional Lorentzian metric in the conformal class of the *acoustic metric*

$$g_{\mathrm{ac}} = (dx - v\, dt)^2 - c^2\, dt^2.$$

The characteristics determine the metric only up to a conformal factor. A genuinely metrical

statement beyond the null cones, for example about lengths or curvature, needs more than the principal symbol.

Second, the entropy mode propagates along the flow lines with speed v, which is not null with respect to g_{ac}. The system has a sound sector with a Lorentzian cone and an entropy sector with a different characteristic, so no single metric governs all of its disturbances. In higher dimensions the same separation holds, and the familiar acoustic metric $g_{\mu\nu} \propto \big(-(c^2 - v^2),\, -v_j;\, -v_i,\, \delta_{ij}\big)$ is valid for irrotational, barotropic flow [27, 28].

Third, the dissipative terms make the full system hyperbolic–parabolic. The acoustic cone describes the high-frequency propagation of the ideal system at wavenumbers where the damping rate Γq^2 is small compared with cq, that is for $q \ll c/\Gamma$, and not the complete signal propagation of the model.

Nor does the acoustic metric satisfy gravitational field equations. Visser [28] emphasizes that it is determined by the fluid variables and does not obey the Einstein equations. The hyperbolicity calculation therefore provides a propagation geometry that can be used as an analogue, not a candidate gravitational metric.

Chapter 41

Unistochastic Dynamics and the RSVP Bridge

A classical continuum closure does not by itself establish unitary evolution. The bridge to the unistochastic projection of Part II has to supply a complex amplitude space, an inner product, and conditions ensuring unitary transport, and only then can $T_{ij} = |U_{ij}|^2$ be invoked. This chapter records the one exact construction of this kind that connects a conservative fluid to a unitary evolution, and states what it does and does not give.

Let $\psi : \mathbb{S}^1 \to \mathbb{C}$ be nowhere vanishing and write $\psi = \sqrt{\Phi}\, e^{i\theta}$ with $\Phi = |\psi|^2$. Define $v = (\hbar/m_0)\, \partial_x \theta$.

Proposition 41.1 (Madelung transformation)**.** *If ψ solves the Schrödinger equation $i\hbar\partial_t\psi = -\frac{\hbar^2}{2m_0}\partial_x^2\psi + V\psi$ with real V, then Φ and v satisfy the continuity equation $\partial_t\Phi + \partial_x(\Phi v) = 0$ and the Euler-type equation*

$$\partial_t v + v\,\partial_x v = -\frac{1}{m_0}\partial_x\big(V + Q\big), \qquad Q = -\frac{\hbar^2}{2m_0}\frac{\partial_x^2\sqrt{\Phi}}{\sqrt{\Phi}},$$

the second being the equation of an inviscid fluid with an additional pressure-like potential Q [24]. Conversely, a smooth solution of these two equations with irrotational $v = (\hbar/m_0)\partial_x\theta$ defines a solution of the Schrödinger equation, up to a time-dependent phase factor, since the Euler equation is the gradient of the Hamilton–Jacobi equation and fixes the phase only up to a function of time.

The proposition is classical, and its proof is a substitution of the polar form into the Schrödinger equation and separation of real and imaginary parts. Its content for the present programme is conditional. The conservative capacity equation holds for the fluid image of a unitary evolution, as it does in Part II for the populations of a fixed basis, but the

Euler equation carries the quantum potential Q. A classical closure of the form of Equations (38.1)–(38.3) with an equation of state $p(\Phi, S)$ contains no such term, and it does not produce unitary evolution unless Q is added as an assumption on the pressure.

The proposition clarifies the relation to Chapter 9, with care about what the velocity retains. A nowhere-vanishing ψ on a connected domain is determined up to a global phase by $\Phi = |\psi|^2$ and the phase θ, and θ is determined up to an additive constant by its gradient, hence by v. In the discretization where $\Phi_i = |\psi_i|^2$ are the populations of Part II, the projection $T(U)$ retains the populations and discards the relative phases. The velocity field retains the spatial gradient of the phase, which is the continuum counterpart of the phase differences between neighbouring sites, so it carries part of what the projection discards; it does not by itself carry every relative phase. Recovering ψ from (Φ, v) requires three further conditions. The circulation of v around every loop must satisfy the quantization condition below. The nodes of ψ, where $\Phi = 0$, θ is undefined and v is singular, must be excluded or treated separately. And the relative phase between disconnected components of the support of Φ is not determined by v at all, since the phase of each component is fixed only up to its own constant. When these conditions hold, (Φ, v) determines the ray, and Theorem 9.1 identifies the ray with the minimal probabilistically sufficient state for measurements in all bases; without them (Φ, v) determines less. The memory kernel of Chapter 13 is then what the retained phase gradient contributes to the dynamics of the density alone.

Single-valuedness of ψ on the circle gives a quantization of circulation:

$$\oint m_0 v \, dx = 2\pi\hbar \, n, \qquad n \in \mathbb{Z}.$$

This is the statement that the flat connection $\nabla = d - i\,(m_0 v/\hbar)\,dx$ on the trivial line bundle over the circle, whose parallel sections are $e^{i\theta}$, has holonomy $e^{i m_0 \oint v\, dx/\hbar}$ equal to the identity on the node-free domain, the condition that the connection define a global section. Trivial holonomy does not mean zero circulation: the integer n can be nonzero, and then v has a nonzero mean. The condition is a concrete place where the holonomy of Part IV enters the continuum description. It requires $\Phi > 0$ so that θ is defined; at zeros of ψ the velocity is singular.

The unistochastic bridge proposed in the earlier RSVP work, that coarse-grained field dynamics yields a transition matrix $T_{ij} = |U_{ij}|^2$ for a unitary U, is therefore a conjecture whose precise form is the following. A family of RSVP-type fluid closures whose pressure contains the quantum potential Q and whose velocity is irrotational is equivalent to a Schrödinger evolution, so that its coarse-grained populations are the diagonal projections of a unitary dynamics on a Hilbert space of amplitudes. Whether the closure of Chapter 38 can be modified to contain such a term from a derivation, rather than by postulate, is not

shown here.

Chapter 42

Gravitation and Effective Geometry

A gravitational sector requires an independent metric g and an affine connection Γ, a covariant action in which they appear, and field equations obtained by variation. This chapter states the minimal setup and records the places at which the RSVP fields can be attached.

In the Palatini formulation the action is

$$\mathcal{I}[g,\Gamma,\text{matter}] = \frac{1}{2\kappa_g}\int \sqrt{-g}\, g^{\mu\nu} R_{\mu\nu}(\Gamma)\, d^4x + \mathcal{I}_{\text{m}}[g,\text{matter}],$$

with $R_{\mu\nu}(\Gamma)$ the Ricci tensor of a torsion-free connection and the matter action independent of Γ. Variation with respect to Γ gives $\nabla_\lambda(\sqrt{-g}g^{\mu\nu}) = 0$, whose solution in dimension greater than two is the Levi-Civita connection of g up to a projective transformation, and variation with respect to g gives the Einstein equations with the stress–energy tensor of the matter. This is a standard result, and it identifies the connection with the Levi-Civita connection of the metric only as a consequence of the field equations under the stated coupling assumptions.

The RSVP fields enter as matter. The relativistic version of the capacity–flow–entropy system has the perfect-fluid stress tensor

$$T^{\mu\nu} = (\varepsilon + p)\, u^\mu u^\nu + p\, g^{\mu\nu},$$

with conserved capacity current $J^\mu = \Phi u^\mu$ and conservation of $T^{\mu\nu}$ following, on solutions of the matter equations, from diffeomorphism invariance of $\mathcal{I}_{\text{m}}$. Adding dissipation to a relativistic fluid is not a straightforward extension of Chapter 38: the first-order relativistic dissipative theories are known to have generic instabilities [20], and a causal and stable relativistic version requires a different structure, for example one with relaxation of the

fluxes. The dissipative sector of the nonrelativistic model above is not asserted to have a covariant extension in this book.

The distinction between the two geometries is the main point of the chapter. The acoustic metric of Chapter 40 governs the propagation of small disturbances in the fluid and is determined by the fluid variables. The gravitational metric g is a dynamical field with its own equations, which fluid matter sources. The two coincide only if a theorem or a postulate says so. An identification of the form $g = g_{\mathrm{ac}}$, up to a conformal factor, is therefore recorded as follows.

Conjecture 42.1 (Coincidence of propagation and gravitational geometry)**.** There exists a choice of the RSVP closure, of its coupling to the metric g, and of the conformal factor, for which the gravitational metric of the Palatini theory coincides with the acoustic metric of the fluid sector to the order at which the Einstein equations are satisfied.

The conjecture is not supported by the results of this part. Analogue-gravity arguments show that the acoustic metric reproduces kinematic aspects of curved spacetime, such as horizons and Hawking-type emission, for which the field equations of the metric are not required; they do not show that the acoustic metric obeys the Einstein equations [28].

One further remark concerns the nonlinearity of the connection in the metric. The Christoffel symbols are nonlinear functionals of g and its first derivatives, so the average of the connection over fluctuations of the metric does not in general equal the connection of the averaged metric. An elementary example in one dimension: take the metrics $g_\omega = (1 + \omega x)\, dx^2$ with $\omega = \pm\frac{1}{2}$ equally likely, so that $\langle g \rangle = dx^2$ has connection zero. The connection coefficient of g_ω is $\Gamma^x_{xx} = \omega/\big(2(1+\omega x)\big)$, whose average over ω at $x = 1$ is $\frac{1}{4}\big(\frac{1/2}{3/2} - \frac{1/2}{1/2}\big) = -\frac{1}{6} \neq 0$. This is an instance of the effect formalized by the affine quantum deformation principle of the author's earlier work [17], in which the averaged connection of a fluctuating geometry differs from the connection of the averaged metric by a correction tensor; the example here shows only the elementary nonlinearity and not the evaluation of that correction in any physical setting.

The ledger for this part is as follows. Established: the definition of the capacity density and its exact balance law under conservation; a closed model with an explicit Lie–Poisson reversible operator and an explicit dissipative operator, with the degeneracy conditions, energy conservation, entropy production and linear stability proved; the characteristic speeds and acoustic cone of the ideal sector; the Madelung correspondence between conservative irrotational flow with a quantum potential and unitary evolution. Not established: closure from microscopic dynamics beyond the cited exact example; nonlinear stability; a covariant dissipative theory; a derivation of the quantum potential; the identification of the propagation metric with a gravitational metric.

Part VII

Accessibility and Cosmology

This part makes precise a distinction that Parts I and V introduced in other settings: the structure a system has and the structure an observer can access are different quantities, and the second can decline while the first grows. The treatment begins with an operational definition, in which structure is the distinguishability of two candidate alternatives and access is mediated by a measurement channel, and it proves the exact mechanisms by which accessibility declines and the conditions under which it does not. A cosmological background enters only as an explicitly specified model, namely a spatially flat universe with matter and a cosmological constant, whose expansion history, linear growth factor and horizons are inputs. Nothing in the part derives expansion or redshift from declining accessibility, and the quantitative statements about the background are statements about that model and not about observation.

Chapter 43

Physical Structure and Accessible Structure

Let $\mathcal{X}$ be a measurable space of physical configurations and ρ_0, ρ_1 two probability laws on it, the *smooth reference* and the *structured alternative*. A *distinguishability measure* $\mathfrak{D}(\rho_0, \rho_1) \in [0, \infty]$ is a functional of the pair; the examples used here are the total variation distance, the relative entropy, and the other f-divergences, together with the Bayes error of the optimal test between the two. A *channel* is a Markov kernel C from $\mathcal{X}$ to a space $\mathcal{Y}$ of data, and it maps a law ρ to the law $C\rho$ of the data. The *physical structure* of the pair is $\mathfrak{D}(\rho_0, \rho_1)$, and the *accessible structure* for an observer with channel C is $\mathfrak{D}(C\rho_0, C\rho_1)$.

Proposition 43.1 (Accessible structure does not exceed physical structure)**.** *For every f-divergence $\mathfrak{D}$ and every channel C, $\mathfrak{D}(C\rho_0, C\rho_1) \le \mathfrak{D}(\rho_0, \rho_1)$. For the relative entropy on a finite space with ρ_0, ρ_1 of full support, equality holds if and only if the likelihood ratio ρ_1/ρ_0 is determined by the data, in the sense that C is sufficient for the pair $\{\rho_0, \rho_1\}$ in the sense of Part I.*

Proof. The inequality is the data-processing inequality for f-divergences, due to Csiszár [13]; for the total variation it follows from the triangle inequality and for the relative entropy from the log-sum inequality, as in the proof of Theorem 31.1. For the equality statement on a finite space, equality in the log-sum inequality holds for each data value y if and only if the ratio of the joint masses $\rho_1(x)C(y|x)/\rho_0(x)C(y|x)$ is constant in x over the support of $C(y|\cdot)$, that is, if $\rho_1(x)/\rho_0(x)$ is constant on the set of configurations compatible with each datum, which is the statement that the likelihood ratio factors through the data. □

The proposition says what the channel can lose, and it is the exact form of the claim in Part I that an insufficient observation discards predictive information about alternatives. It does not say that structure cannot grow. A Markov dynamics applied to both alternatives,

taken as laws, also contracts the divergence by the same inequality, so growth of structure cannot mean growth of the divergence between the laws of an ensemble evolving under a stochastic dynamics. It means growth of the separation of two particular configurations under a deterministic flow, which is a different quantity, and the observer's channel turns that separation into a distinguishability between laws of data.

Definition 43.2 (Structure under a flow)**.** Let $\varphi_t \colon \mathcal{X} \to \mathcal{X}$ be a flow, x_0 a structured configuration and μ_0 a smooth reference, and write $x_t = \varphi_t(x_0)$, $\mu_t = \varphi_t(\mu_0)$ and $\Delta_t = x_t - \mu_t$ in a vector-space chart. The *separation* of the pair is a norm $\|\Delta_t\|$, and the *accessible structure at time* t is $\mathfrak{D}(C_t\delta_{x_t}, C_t\delta_{\mu_t})$ for the channel C_t of the observer at time t.

The flow φ_t is a bijection, so no information about the initial pair is lost physically: (x_t, μ_t) determines (x_0, μ_0). The separation can nonetheless grow, as it does in gravitational instability, and the data of an observer whose channel degrades faster than the separation grows can become less informative at the same time. Chapter 45 gives this a closed form.

Chapter 44

Accessibility Kernels

Two objects are commonly called a kernel in this setting, and they support different arguments. An *instrument response kernel* is a linear operator W acting on a signal field δ, for example a spatial convolution with a window of finite width, possibly with a gain below or above one. A *normalized probabilistic channel* is a Markov kernel, as defined above. The first is not a channel and does not inherit the data-processing inequality.

Proposition 44.1 (A normalized channel contracts; a weighted integral need not)**.** *For a Markov kernel C and signed measures $\nu = \rho_1 - \rho_0$ of total mass zero, $\|C\nu\|_{\mathrm{TV}} \le \|\nu\|_{\mathrm{TV}}$. For a linear response $W\delta(y) = \int w(y,x)\delta(x)\,dx$ with nonnegative weights that are not normalized, $\|W\delta\|_1$ can exceed $\|\delta\|_1$, even for signals of total mass zero.*

Proof. The first statement is the contraction of total variation under a Markov kernel. For the second, take the response with gain two, $w(y,x) = 2\,\delta_{yx}$, which has nonnegative weights and is not normalized. Then $W\delta = 2\delta$ and $\|W\delta\|_1 = 2\|\delta\|_1 > \|\delta\|_1$ for every nonzero δ, in particular for those of total mass zero. □

A response kernel enters the data-processing argument only after it is combined with a noise model that makes the pair a channel. The standard case is the Gaussian channel.

Proposition 44.2 (Gaussian channel and signal-to-noise ratio)**.** *Let $Y = W\delta + \xi$ with $\xi \sim N(0,\Sigma)$. For two deterministic signals δ_0, δ_1 the laws of the data are $N(W\delta_0,\Sigma)$ and $N(W\delta_1,\Sigma)$, and*

$$\mathrm{KL}\big(N(W\delta_1,\Sigma)\,\|\,N(W\delta_0,\Sigma)\big) = \tfrac{1}{2}\,s, \qquad s = \big\|\Sigma^{-1/2}W(\delta_1-\delta_0)\big\|^2,$$

and the total variation distance between them is $2\Phi_{\mathrm{N}}(\sqrt{s}/2) - 1$, with Φ_{N} the standard normal distribution function. Both are increasing functions of the signal-to-noise ratio s.

Proof. Both are standard computations for Gaussian laws with common covariance: the relative entropy is half the squared Mahalanobis distance of the means, and the optimal test is the likelihood-ratio test on the projection of the data onto $\Sigma^{-1}W(\delta_1 - \delta_0)$, in which the two laws are unit-variance normals with means 0 and $\sqrt{s}$. □

In the Fourier-diagonal case, with independent modes k with response W_k and noise variance N_k, the ratio is $s = \sum_k |W_k|^2|\Delta_k|^2/N_k$, and each mode contributes its own s_k. The response enters only through $|W_k|^2/N_k$, the *effective gain-to-noise ratio* of the mode. A response kernel that attenuates the signal by a factor g lowers s by g^2 unless the noise variance falls as well, which is why attenuation and noise cannot be treated independently.

Proposition 44.3 (Cascade of Gaussian channels)**.** *Compose a first channel $Y_1 = g_1\delta + \xi_1$, $\xi_1 \sim N(0, N_1)$, with a second $Y_2 = g_2 Y_1 + \xi_2$, $\xi_2 \sim N(0, N_2)$, with independent noises. Then $Y_2 = g_1 g_2 \delta + \xi$ with $\xi \sim N(0, g_2^2 N_1 + N_2)$ and the signal-to-noise ratio of the cascade is*

$$s_2 = \frac{g_1^2 g_2^2\,|\Delta|^2}{g_2^2 N_1 + N_2} \le \frac{g_1^2 |\Delta|^2}{N_1} = s_1,$$

with equality if and only if $N_2 = 0$.

Proof. Substitute Y_1 into Y_2 and add the independent noise variances. The inequality is $g_2^2 N_1 + N_2 \ge g_2^2 N_1$, an equality only for $N_2 = 0$. It is the data-processing inequality for the particular cascade, since the second channel is a Markov kernel applied to the data of the first. □

Finally, a restricted causal domain is a kernel with support. If the response W_t is supported on a domain B_t, in the sense that $W_t\delta$ depends only on $\delta|_{B_t}$, then two signals that agree on B_t produce identical laws of data and $\mathfrak{D} = 0$ for every divergence, however large their separation outside B_t.

Chapter 45

Distinction Functionals

The scale-dependent distinction functional of a mode k at time t is the divergence between the laws of the data under the structured and the smooth alternative. Two forms are used. For a deterministic structured configuration against a smooth reference it is the Gaussian shift of Proposition 44.2, and for a random structured field against a smooth reference it is the divergence between zero-mean Gaussians of different variance.

Proposition 45.1 (The two forms depend on the signal-to-noise ratio only)**.** *Let a mode have signal variance P (for a random field) or squared amplitude $|\Delta|^2$ (for a configuration), effective response W and noise N, and put $s = |W|^2 P/N$ (respectively $|W|^2|\Delta|^2/N$). Then the relative entropy of the structured data law from the smooth one is $\frac{1}{2}s$ in the deterministic case and $g(s) = \frac{1}{2}\big(s - \ln(1+s)\big)$ in the random case with the structured law $N(0, |W|^2P + N)$ and the smooth law $N(0, N)$. Both are strictly increasing in s for $s > 0$.*

Proof. The deterministic case is Proposition 44.2. For the random case the relative entropy of $N(0,\sigma_1^2)$ from $N(0,\sigma_0^2)$ is $\frac{1}{2}\big(\sigma_1^2/\sigma_0^2 - 1 - \ln(\sigma_1^2/\sigma_0^2)\big)$, and $\sigma_1^2/\sigma_0^2 = 1 + s$. The derivative $g'(s) = \frac{1}{2}\big(1 - \frac{1}{1+s}\big)$ is positive. □

Because both forms are increasing functions of s_k, the sign of the time derivative of the distinction of a single mode is the sign of $\dot{s}_k$, and their turnover times coincide. The balance equation follows from writing s_k as a product.

Theorem 45.2 (Balance of the distinction of a mode)**.** *Write the signal variance of a mode as $P_k(t) = D_+(t)^2\, T_k(t)^2\, P_k^0$, where D_+ is the growth factor of the structure, $T_k \in (0,1]$ is a physical smoothing (transfer) factor, and P_k^0 is fixed, and write the effective gain-to-noise ratio as $\gamma_k(t) = |W_k(t)|^2/N_k(t)$. Then $s_k = \gamma_k D_+^2 T_k^2 P_k^0$ and, with the formation rate $F = 2\dot{D}_+/D_+$, the smoothing rate $\mathrm{Sm}_k = -2\, d\ln T_k/dt$ and the access-loss rate $A_k = -d\ln\gamma_k/dt$,*

$$\frac{d\ln s_k}{dt} = F - \mathrm{Sm}_k - A_k.$$

The distinction of the mode increases exactly when the formation rate exceeds the sum of the smoothing and access-loss rates.

Proof. Take the logarithm of $s_k = \gamma_k D_+^2 T_k^2 P_k^0$ and differentiate; the constant P_k^0 drops out, and the three remaining terms are $-A_k$, F and $-\mathrm{Sm}_k$ by the definitions. The smoothing and access rates are defined with the signs that make them nonnegative when the transfer factor decreases (physical smoothing) and when the effective gain-to-noise ratio decreases (loss of observational access). □

The three terms separate three mechanisms that are often conflated: structure formation, which increases the signal; physical smoothing, which decreases the signal itself; and loss of observational access, which decreases the signal that reaches the data without changing the signal. The decomposition is exact for the factorized form, and a model that does not factorize needs its own decomposition.

Theorem 45.3 (Turnover)**.** *Let $r(t) = F - \mathrm{Sm}_k - A_k$ be continuous on an interval. If $r(t_0) > 0$ and $r(t_1) < 0$ for some $t_0 < t_1$, then s_k attains a maximum over $[t_0, t_1]$ at an interior point t_* with $r(t_*) = 0$. If r has exactly one zero in (t_0, t_1), that zero is the unique point of maximum on the interval. A zero of r without a change of sign is a stationary point of $\ln s_k$ and need not be a maximum.*

Proof. The function $\ln s_k$ is continuously differentiable with derivative r. Since $r(t_0) > 0$ the function is increasing at t_0, and since $r(t_1) < 0$ it is decreasing at t_1, so the maximum of $\ln s_k$ over the compact interval is interior, and at an interior maximum the derivative vanishes. If r has one zero, then $r > 0$ before it and $r < 0$ after it by the intermediate value theorem, and the zero is the unique maximum. For the last statement, $\ln s = (t-1)^3$ has derivative $3(t-1)^2 \geq 0$, vanishing at $t = 1$, and is increasing through it. □

Turnover is therefore a testable property of the three rates, not a definition: it is the first-derivative test applied to an explicit balance. Different modes have different smoothing and access rates and generally peak at different times. The total distinction summed over modes has derivative $\sum_k g'(s_k)\, s_k\, r_k$ in the random form, a weighted sum of the individual rates, and a turnover of the total is not implied by turnovers of the modes without further assumptions.

Chapter 46

Mixing, Attenuation, and Causal Separation

Three mechanisms reduce what an observer can distinguish. They are different mathematical objects, and each supports its own exact statement.

The first is *mixing* of the physical state. If the dynamics is a doubly stochastic Markov chain and the description is lumpable, Theorem 32.2 shows that the relative entropy of the coarse state to the uniform reference does not increase: the state approaches the smooth reference. This is physical smoothing in the sense of the balance equation, a decrease of the signal itself, and it is a data-processing statement about the dynamics and not about the observer.

The second is *attenuation*, the lowering of the effective gain-to-noise ratio of a channel, treated in Proposition 44.3: any stage added to a channel can only lower the signal-to-noise ratio, and the attenuation of a stage reduces the ratio by the square of its gain when its noise is fixed. This is loss of access, a decrease of γ_k with the signal unchanged.

The third is *causal separation*. Let the observation at time t depend on the signal only on a domain B_t, as in the end of Chapter 44.

Theorem 46.1 (Nested domains and monotone accessibility)**.** *Suppose the data at time t are $Y_t = (\delta|_{B_t}, \xi_t)$ with independent noise, or any channel that factors through the restriction to B_t, and that $B_{t_2} \subseteq B_{t_1}$ for $t_2 \geq t_1$. Then the channel at t_2 factors through the channel at t_1, and for every divergence satisfying data processing,*

$$\mathfrak{D}\big(C_{t_2}\rho_0, C_{t_2}\rho_1\big) \leq \mathfrak{D}\big(C_{t_1}\rho_0, C_{t_1}\rho_1\big)$$

for any fixed pair of alternatives (ρ_0, ρ_1). In particular, if ρ_0 and ρ_1 coincide on B_t then the accessible structure vanishes.

Proof. A restriction from B_{t_1} to $B_{t_2} \subseteq B_{t_1}$ is a deterministic map, hence a Markov kernel, so $C_{t_2} = K \circ C_{t_1}$ for the kernel K that restricts the data. Proposition 43.1 then gives the inequality. For the last statement, equal restrictions give equal data laws. □

The hypothesis that the later channel factors through the earlier one is essential. It fails when new signals arrive, when the instrument changes, and when fresh structure forms inside B_t, because then the alternatives themselves change with time. In those cases accessibility need not decline, and nothing in the theorem implies that every evolving observer loses information.

A cosmological geometry supplies an exact example of nested shrinking domains. In a spatially flat Friedmann–Lemaître universe with scale factor a, Hubble rate $H(a)$, and comoving coordinates in which light travels along $d\chi = dt/a$, the *comoving event horizon* at scale factor a is

$$\chi_e(a) = \int_a^\infty \frac{da'}{a'^2 H(a')},$$

the comoving radius of the region from which a signal emitted now can still reach the observer, whenever the integral converges. In a matter-only universe with $H = H_0 a^{-3/2}$ the integrand is $a'^{-1/2}/H_0$ and the integral diverges, so there is no event horizon and no such shrinking. In a universe with a cosmological constant the integral converges and, because $H \to H_\Lambda$ as $a \to \infty$, $\chi_e(a) \sim 1/(aH_\Lambda)$, which tends to zero. For the background used below, matter fraction $\Omega_m = 0.3$ and cosmological-constant fraction 0.7, with $E(a) = H/H_0 = \sqrt{\Omega_m a^{-3} + \Omega_\Lambda}$, the event horizon in units of c/H_0 is 1.1407 at $a = 1$, 0.5937 at $a = 2$, 0.2389 at $a = 5$ and 0.1195 at $a = 10$, against the asymptote $1/(a\sqrt{\Omega_\Lambda})$ with values 1.1952, 0.5976, 0.2390 and 0.1195. The comoving particle horizon over the same range grows from 3.31 to 4.33 and saturates, which is the contrast: the region that has been seen saturates while the region that can still be influenced shrinks. These numbers are properties of the assumed model. An event horizon alone does not establish the nesting of Theorem 46.1: the domains B_t, their identification across times (for example by comoving coordinates), and the factorization of each later channel through the earlier one must be specified. Shrinking comoving reach describes the region that can still influence the observer, and it is a different object from the information that an observer has accumulated in a retained record, which behaves oppositely (Theorem 54.2).

Chapter 47

What an Approach to Smoothness Establishes

The Smoothness Paradox is the observation that a universe which continues to form structure can appear increasingly smooth to a bounded observer. The preceding chapters allow it to be stated as a precise claim with explicit conditions, and also to be bounded: a declining accessible distinction establishes less than it might seem to.

An *observer class* $\mathcal{O}$ is a set of channels, and a *tolerance* is a number $\epsilon \in (0,1)$. The structured alternative ρ_1 and the smooth reference ρ_0 are *ϵ-indistinguishable by $\mathcal{O}$* if $\sup_{C\in\mathcal{O}} \|C\rho_1 - C\rho_0\|_{\mathrm{TV}} \le \epsilon$, where the total variation distance is normalized so that it equals $2\Phi_{\mathrm{N}}(\sqrt{s}/2) - 1$ for the Gaussian case. By the standard relation between total variation and the error of the optimal test, the Bayes error of equally likely hypotheses is $\frac{1}{2}(1 - \mathrm{TV})$, so ϵ-indistinguishability bounds the advantage of every observer in $\mathcal{O}$ over guessing.

Theorem 47.1 (Exact criterion for the Gaussian channel)**.** *For the Gaussian channel of Proposition 44.2 the alternatives are ϵ-indistinguishable if and only if $s \le a_\epsilon^2$, where $a_\epsilon = 2\Phi_{\mathrm{N}}^{-1}\big(\frac{1+\epsilon}{2}\big)$. If the observer class is a nested family of Gaussian channels with signal-to-noise ratios s_C ordered by factorization, the supremum is attained at the finest channel.*

Proof. By Proposition 44.2, $\mathrm{TV} = 2\Phi_{\mathrm{N}}(\sqrt{s}/2) - 1 \le \epsilon$ is equivalent to $\sqrt{s} \le 2\Phi_{\mathrm{N}}^{-1}((1+\epsilon)/2)$. For a nested family, finer channels have larger s by Proposition 44.3 and the monotonicity of the total variation in s. □

The theorem makes indistinguishability a quantitative threshold on the signal-to-noise ratio. For $\epsilon = 0.1$ the threshold is $a_\epsilon \approx 0.2513$, and for $\epsilon = 0.5$ it is $a_\epsilon \approx 1.349$, so an observer needs a signal-to-noise ratio above a_ϵ^2 to reach advantage ϵ.

The next result limits what a decline of the accessible distinction can be taken to mean. It concerns a single mode with the factorization of Chapter 45.

Proposition 47.2 (A decline of accessibility does not imply a decline of structure)**.** *Let $s(t) = \gamma(t)P(t)$ with $\gamma > 0$. Then s can decrease while P increases, if γ decreases fast enough: for instance $P = t^{4/3}$ and $\gamma = e^{-t/\tau}$ give $\dot{s}/s = \frac{4}{3t} - \frac{1}{\tau} < 0$ for $t > \frac{4}{3}\tau$. Conversely, if $s(t_2) < s(t_1)$ and $\gamma(t_2) \geq \gamma(t_1)$ then $P(t_2) < P(t_1)$.*

Proof. The first statement is the computation of the logarithmic derivative. For the second, $P(t_2) = s(t_2)/\gamma(t_2) < s(t_1)/\gamma(t_2) \leq s(t_1)/\gamma(t_1) = P(t_1)$. □

The statement "the universe is approaching smoothness" is therefore, as a statement about data, a statement about the product of the signal and the channel. It supports a statement about the signal only under an additional assumption that the channel quality does not degrade.

Theorem 47.3 (Indistinguishability from a shrinking domain)**.** *Let the structured alternative differ from the smooth reference by a deterministic signal with squared-amplitude density bounded by $j(t) \leq \bar{j} < \infty$ per unit comoving volume, and let the observer's channel be the Gaussian channel restricted to a comoving domain B_t of volume $V(t)$, with unit effective gain-to-noise ratio inside the domain. Then the signal-to-noise ratio satisfies $s(t) \leq \bar{j}V(t)$. If $V(t) \to 0$ then for every $\epsilon > 0$ there is a time after which the alternatives are ϵ-indistinguishable, for this observer.*

Proof. By Proposition 44.2 with independent unit noise per mode inside B_t and no response outside, $s = \int_{B_t} |\Delta|^2 \, dV \leq \bar{j}V(t)$. By Theorem 47.1 the alternatives are ϵ-indistinguishable once $\bar{j}V(t) \leq a_\epsilon^2$, which holds for large t because $V \to 0$. □

In the background of the previous chapter the comoving event horizon tends to zero, so the volume of the ball of radius χ_e tends to zero like a^{-3}, and the theorem applies to an observer whose channel is limited to the region that can still be influenced, provided the domains and the factorization between channels are specified as in Theorem 46.1. Two features of the claim should be kept in view. It holds for an observer class defined by a causal restriction and a bounded signal density, and it says nothing about the existence of other observers, for example one who has recorded earlier data. And it establishes indistinguishability of a particular structured configuration from a smooth reference by a specified class, and does not establish that the configuration is smooth.

Chapter 48

Turnover and Indistinguishability Results

The balance equation, the turnover theorem and the indistinguishability theorem combine in a model that can be computed. Let the structure grow linearly with the growth factor $D_+(a)$ of the background, let the signal density per comoving volume at fixed mode content be $\bar{j}\, D_+(a)^2$ with $\bar{j}$ constant, and let the observer's channel be the Gaussian channel of Theorem 47.3 on the ball of comoving radius $\chi_e(a)$ with constant gain-to-noise ratio. Then the signal-to-noise ratio of the best test is

$$s(a) = \tfrac{4\pi}{3}\, \bar{j}\, D_+(a)^2\, \chi_e(a)^3.$$

This is a toy: the growth is linear, the spectrum is flat and bounded, and the instrument is idealized. It is useful because every ingredient is explicit and every conclusion can be checked.

The linear growth factor of a flat universe with matter and a cosmological constant is

$$D_+(a) \propto E(a) \int_0^a \frac{da'}{(a' E(a'))^3},$$

a result due to Heath [15] and derived in Appendix E, which solves the growth equation $D'' + (3/a + d\ln E/da) D' - \tfrac{3}{2}\Omega_m a^{-5} E^{-2} D = 0$ with primes denoting d/da. A numerical integration of this equation from $a = 10^{-3}$ with initial condition $D = a$ reproduces the integral formula to 10^{-11} at $a = 0.5$, where $D_+(0.5)/D_+(1) = 0.6118$.

Proposition 48.1 (Turnover condition in the model)**.** *Let $f = d\ln D_+ / d\ln a$. Then*

$$\frac{d\ln s}{d\ln a} = 2 f(a) - \frac{3}{a H(a) \chi_e(a)},$$

in units with $c = H_0 = 1$, so that $H = E$, which is the balance equation with formation rate $2f$, no physical smoothing, and a loss of access given by the shrinking of the event-horizon volume.

Proof. Differentiate the logarithm of s and use $d\chi_e/da = -1/(a^2 H)$, which gives $d\ln\chi_e/d\ln a = -1/(aH\chi_e)$. □

In a matter-only universe there is no event horizon, the second term is absent, the growth rate is $f = 1$, and s increases monotonically with a: there is no turnover. In the universe with $\Omega_m = 0.3$ the growth rate decreases (it is 0.5128 at $a = 1$, against $\Omega_m^{0.55} = 0.5157$, and 0.0093 at $a = 10$) while the second term tends to -3 because $aH\chi_e \to 1$. Numerically, the right-hand side is 1.472 at $a = 0.1$, 0.749 at $a = 0.3$, -0.044 at $a = 0.5$ and -1.604 at $a = 1$, with a single sign change at

$$a_* = 0.489 \quad (z_* = 1.04),$$

where s attains its maximum. A direct scan of $D_+^2\chi_e^3$ over $0.05 \le a \le 5$ places its maximum at $a = 0.486$, consistent with the root to the resolution of the scan. Relative to the peak, the signal-to-noise ratio is 0.224 at $a = 0.1$, 0.702 at $a = 0.25$, 0.567 at $a = 1$, 0.126 at $a = 2$, 0.0097 at $a = 5$ and 0.0013 at $a = 10$, and for large a it falls like a^{-3} because D_+ saturates and $\chi_e \propto a^{-1}$.

Proposition 48.2 (Computed turnover)**.** *For every flat background with $\Omega_m, \Omega_\Lambda > 0$ the right-hand side of Proposition 48.1 tends to 2 as $a \to 0$ and to -3 as $a \to \infty$, so it has at least one zero and $s(a)$ has at least one interior maximum (Corollary E.4). In the model with $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ the right-hand side changes sign exactly once on the grid examined, from positive to negative, at $a_* \approx 0.489$, so that by Theorem 45.3 the signal-to-noise ratio has its maximum there. The location and the uniqueness of the zero are numerical statements about this background.*

Theorem 48.3 (Eventual indistinguishability in the model)**.** *In the same model, D_+ is increasing and bounded, $\chi_e(a) \to 0$ as $a \to \infty$, and hence $s(a) \to 0$. For every tolerance $\epsilon > 0$ there is a scale factor beyond which the alternatives are ϵ-indistinguishable by the event-horizon-limited observer, and asymptotically s decays like a^{-3}. If $s(a_*) \le a_\epsilon^2$ the alternatives are never distinguishable at tolerance ϵ by this observer.*

Proof. The growth factor is increasing and bounded, and the event horizon satisfies $\chi_e \sim 1/(a\sqrt{\Omega_\Lambda}) \to 0$ (Propositions E.2 and E.3). Then $s = \frac{4\pi}{3}\bar{j}D_+^2\chi_e^3 \to 0$ like a^{-3}. The indistinguishability statements follow from Theorem 47.1, since $s \le a_\epsilon^2$ holds for all large a, and for all a if the maximum value $s(a_*)$ does. □

The location and uniqueness of the sign change are numerical observations for this background and not proved properties of every background. The qualitative conclusion, a turnover followed by indistinguishability, requires only that the growth rate fall to zero and that an event horizon exist and shrink, and it fails, for example, in the matter-only universe.

Chapter 49

Expansion, Redshift, and Observational Comparison

The results above use a background in which the expansion history, the growth factor and the horizons are given. This chapter states exactly what is assumed and what a comparison with observation would require.

The background is a spatially flat Friedmann–Lemaître universe with scale factor $a(t)$, Hubble rate $H = \dot{a}/a$ and $H^2 = H_0^2\big(\Omega_m a^{-3} + \Omega_\Lambda\big)$, with $\Omega_m + \Omega_\Lambda = 1$. The redshift of light emitted at scale factor a and observed today is $1 + z = 1/a$. The comoving particle horizon is $\chi_p(a) = \int_0^a da'/(a'^2 H)$ and the comoving event horizon is $\chi_e(a)$ above; the distinction between them is the standard one [14]. The linear growth of a pressureless density perturbation on this background is the growth factor D_+ of the previous chapter [15, 16]. Surface brightness of sources at redshift z is reduced by the factor $(1 + z)^{-4}$, the standard Tolman dimming, and enters a channel as a gain below one for the signal from such sources.

These are inputs. The results of Part VII are theorems about the interaction of a growing signal and a channel on a given background, and they transfer to any background for which D_+ and χ_e are specified. In particular:

First, nothing in this part derives the expansion, the redshift or the gravitational dynamics from declining accessibility. A decline of accessible structure is compatible with many backgrounds, including backgrounds with no event horizon, in which it does not occur (Proposition 48.1). Any proposal in which the redshift or the growth of structure is explained by the RSVP fields would have to supply its own field equations, a derivation of the corresponding background and growth factor, and a comparison of the result with data; the continuum model of Part VI does not contain a cosmological expansion, and the identification of its fields with a cosmological background is not made here.

Second, the toy channel of the previous chapter is not a forecast. A comparison with

observation would require specifying a survey: a selection function and sky coverage, a resolution and noise model for each tracer, the redshift range, the treatment of nonlinear structure and bias, and the cosmic variance of the reference. With those in place the distinction functional of Chapter 45 can be computed for each mode, its balance can be evaluated, and the turnover time can be compared across tracers. Only then does a numerical value of a_* acquire a meaning outside the model, and the value 0.489 found above should not be read as a prediction.

Third, the statement of the Smoothness Paradox established here is conditional on an observer class and a tolerance. It is true for observers restricted to the region that can still be influenced, and it is a statement about distinguishing a particular structured configuration from a smooth reference. It leaves open how a different observer, for example one who preserves earlier records, or a different notion of structure, for example nonlinear statistics, would fare. The nesting hypothesis of Theorem 46.1 marks where the argument stops.

The ledger for this part is as follows. Established: accessible structure never exceeds physical structure for any f-divergence, with the condition for equality; the separation of response kernels from normalized channels; the Gaussian signal-to-noise formulas and the cascade inequality; the exact balance equation for a mode and the first-derivative test for turnover; the monotone decline of accessibility under nested domains; an exact criterion for ϵ-indistinguishability; the proposition that a decline of accessible distinction does not imply a decline of structure unless channel quality is nondecreasing; and a computed example in the $\Omega_m = 0.3$ background with a single turnover at $a_* \approx 0.489$ followed by a proved eventual indistinguishability. Not established: the persistence of the result for nonlinear structure, for non-Gaussian channels, for other observer classes, and for backgrounds other than the one specified; any derivation of the background from the RSVP fields; any comparison with observation.

Part VIII

Obstructions, Computation, and Open Problems

This part draws together the constructions of the preceding parts. It does not announce a unification. It identifies four obstructions that the book has actually constructed, states the common setup and the differences of quantifier and remedy among them, records explicit maps between particular examples and the places where a proposed equivalence fails, collects the computations that were used and classifies them by what they establish, and lists the constructions that remain to be done before the physical programme can be said to follow from the mathematical one.

Chapter 50

Four Obstructions and Their Common Form

Each obstruction of the book has the same skeleton. There is a detailed object x ranging over a set $\mathcal{O}$, an observation map $\pi\colon \mathcal{O} \to \mathcal{Y}$ that forgets part of it, and a task that the observed description may or may not support.

Definition 50.1 (Task and support)**.** A *task* is a map $\tau\colon \mathcal{O} \to \mathcal{Z}$. The observation π *supports* τ if $\tau = f \circ \pi$ for some f.

Proposition 50.2 (Criterion and canonical remedy)**.** *An observation π supports τ if and only if* $\ker \pi \subseteq \ker \tau$*. The coarsest observation supporting τ is τ itself, up to relabelling of its values, and every supporting observation refines it.*

Proof. If $\tau = f \circ \pi$ then $\pi(x) = \pi(x')$ gives $\tau(x) = \tau(x')$. Conversely, if $\ker \pi \subseteq \ker \tau$ then $f(\pi(x)) = \tau(x)$ is well defined on the image of π. If π' supports τ then $\tau = f' \circ \pi'$, so τ is a coarsening of π'. □

This is Theorem 3.2 with the continuation sets replaced by an arbitrary task, and it is the common form of three of the four obstructions. The fourth has a different shape. The four are as follows.

Predictive sufficiency. The detailed objects are histories, and the task is the map $h \mapsto \mathcal{A}(h)$ or $h \mapsto P(h)$ to the support or the law of the future. The observation supports the task if the future laws factor through it. The canonical remedy is the continuation quotient of Chapter 3 and its probabilistic analogue of Chapter 8.

Compositional sufficiency. The detailed objects are operations, elements of a monoid G, and a family $\mathcal{C}$ of contexts, maps $c\colon G \to G$, such as $g \mapsto agb$. For an observation q of the operations the task is $\tau_{\mathcal{C}}(g) = (q(c(g)))_{c \in \mathcal{C}}$, the family of observed descriptions of g

embedded in every context. The observation is compositionally sufficient for $\mathcal{C}$ if q itself supports $\tau_{\mathcal{C}}$, that is, if $q(g) = q(g')$ implies $q(c(g)) = q(c(g'))$ for every $c \in \mathcal{C}$. The canonical remedy is the contextual equivalence of Chapter 52.

Observational discrimination. The detailed objects are the signals or states, two candidate laws ρ_0, ρ_1 are given, and the observation is a channel C. The exact form of the task is to preserve the likelihood ratio: the channel is sufficient for the pair if the ratio factors through the data, which by Proposition 43.1 is equivalent to preservation of the relative entropy, and then of every f-divergence. Preservation of the total variation alone is weaker: the channel that reports only which of ρ_1, ρ_0 is larger at each point preserves it without being sufficient. The approximate and asymptotic forms are quantitative and have no counterpart in the factorization framework.

Global realizability. This is not a factorization problem. The data are a family of local observations $(b_e)_e$, one for each piece of a cover, each with a set of lifts $\mathrm{Lift}_e(b_e) = \pi_e^{-1}(b_e)$, and the question is whether the local lifts can be chosen compatibly on overlaps, that is, whether the family extends to one global object. The criterion is existential: a single choice of lifts with trivial cycle products (Theorem 29.1).

Obstruction	Detailed object and task	Quantifier	Remedy
Predictive	histories; future law	for all pairs with equal observation, equal futures	quotient by equal futures
Compositional	operations; embedded description	for all allowed contexts	quotient by contextual equivalence
Gluing	local data on a cover	there exists one choice of lifts closing every cycle	none by refinement alone: change the data or the cover
Discrimination	pair of laws; separation after a channel	for the pair; at tolerance; in the limit	finer channel; larger observer class

Table 50.1: The four obstructions: object, quantifier, and remedy.

The common concern is retained structure, and the table shows that the quantifiers differ. The first and second ask that a relation hold for all pairs, and so are repaired by enlarging what is retained. The third asks for an existence, and refining the local data does not create existence. The fourth can be asked at three strengths. The following propositions record the explicit maps among the examples of the book and the places where a proposed equivalence fails.

Proposition 50.3 (Predictive and compositional sufficiency are different)**.** *The observation $R_j(U) = [Ue_j] \in \mathbb{CP}^{d-1}$ is predictively sufficient for all measurement futures after left multiplication by gates (Theorem 9.1) and is not compositionally sufficient for contexts that also multiply on the right.*

Proof. Take $d = 2$, $U = I$ and $U' = \mathrm{diag}(1, i)$. Then $R_1(U) = R_1(U') = [e_1]$. With the right context $B = H$, $R_1(UB) = [He_1] = [(1,1)^{\mathsf{T}}]$ and $R_1(U'B) = [(1,i)^{\mathsf{T}}]$, which differ. □

Proposition 50.4 (Compositional sufficiency does not give gluing)**.** *Let $G = PO(2)$ and let q be the identity, which is trivially compositionally and predictively sufficient. Edge data given by the classes $[R(\pi/4)]$ on each of the three oriented edges of a triangle cannot be realized by frames.*

Proof. A realization would give orthogonal frames with $F_b^{\mathsf{T}} F_a = \pm R(\pi/4)$ on each edge, and the cycle product would be $\pm I$. The product of the three rotations is $R(3\pi/4) \neq \pm I$. □

Proposition 50.5 (Gluing does not give sufficiency, and depends on the field)**.** *The mutually unbiased triangle Z, X, Y of a qubit is realizable over $\mathbb{C}$ and not over $\mathbb{R}$. Over $\mathbb{C}$ the frames $F_Z = I$, $F_X = H$, $F_Y = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$ give $|F_b^\dagger F_a|^{\circ 2} = J$ on every edge, and the Bargmann invariant of the first vectors is $(1+i)/4$. Over $\mathbb{R}$ no lift assignment closes (Proposition 29.2). In both cases the projection T is not compositionally sufficient (Proposition 7.4).*

Proof. The unitarity of F_Y and the entries of $F_X^\dagger F_Y = \frac{1}{2} \begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix}$ give the transitions. The overlaps of the first vectors are $\frac{1}{\sqrt{2}}$, $\frac{1+i}{2}$ and $\frac{1}{\sqrt{2}}$, with product $(1+i)/4$, of modulus squared $\frac{1}{8} = (\frac{1}{2})^3$. The real case is Proposition 29.2. □

Proposition 50.6 (Exact discrimination is predictive sufficiency for the pair)**.** *On a finite space with ρ_0, ρ_1 of full support, a channel C preserves the relative entropy between ρ_0 and ρ_1 if and only if it is sufficient for the family $\{\rho_0, \rho_1\}$ in the sense of Part I, with the likelihood function in place of the continuation set, and then it preserves every f-divergence; preservation of total variation alone is weaker. Approximate indistinguishability at a tolerance, and convergence of the observed distinguishability to zero along a family of channels, have no analogue in the factorization framework.*

The first sentence is Proposition 43.1, whose equality case is the factorization of the likelihood ratio, in the same spirit as Proposition 32.1; the second is a statement about quantifiers, and it is illustrated in Chapter 54. The propositions show that no two of the four are equivalent, that the first, second and fourth are repaired by refining the observation, and that the third constrains the data themselves.

Chapter 51

Failure to Reconstruct a History

Reconstruction of the past and prediction of the future are different tasks. Reconstruction asks for the task $\tau(h) = h$, or for a summary of the past, and an observation supports it only if the relevant part of the history can be recovered from the observed value.

Proposition 51.1 (Reconstruction implies sufficiency, and not conversely)**.** *If π is injective on histories then π supports every task, in particular the prediction of the future. The minimal continuation observation $\pi_{\mathcal{A}}$ is sufficient by construction and is not injective whenever two distinct histories have equal continuation sets.*

Proof. If $\ker \pi$ is trivial then $\ker \pi \subseteq \ker \tau$ for every τ, and Proposition 50.2 applies. For the second statement, the continuation equivalence identifies histories with equal continuation sets by definition. □

Compression of the past is exactly what the minimal state was constructed to perform. A finite-state structure retains at most $\log m$ nats about a past of unbounded length. The following gives the rate at which reconstruction fails for a stochastic process.

Proposition 51.2 (Reconstruction loss grows linearly)**.** *Let (X_t) be a stationary process with entropy rate h_μ and let S be a discrete random variable (such as the minimal probabilistic state) with entropy $H(S) = C$. Then the conditional entropy of a window of the past given S satisfies*

$$H(X_{-n+1}, \dots, X_0 \mid S) \;\geq\; n\, h_\mu - C.$$

Proof. For any random variables, $H(X \mid S) = H(X) - I(X;S) \geq H(X) - H(S)$. For a stationary process $H(X_{-n+1}, \dots, X_0) \geq n\, h_\mu$. □

For the even process the bound reads $0.4621\, n - 0.6365$ nats, positive for $n \geq 2$, so the retained two-valued state cannot reconstruct a window of the past beyond a bounded number of symbols, although it predicts the future exactly.

Failure of reconstruction is also neither evidence of predictive failure nor a physical arrow of time. The four combinations of lossless or lossy reconstruction with vanishing or positive entropy production all occur in one family of examples, the biased cycle $\mathbb{Z}_N$ of Chapter 31 with forward probability q.

Proposition 51.3 (Reconstruction and the arrow are independent)**.** *On the cycle $\mathbb{Z}_6$ with forward probability q, consider the identity observation and the observations $x \bmod 3$ and $x \bmod 2$. The identity observation reconstructs the state perfectly and has entropy production $\Sigma_1 = (2q-1)\ln\frac{q}{1-q}$, which is positive for $q \neq \frac{1}{2}$ and zero for $q = \frac{1}{2}$. The observation $x \bmod 3$ loses half the states, yet has the same entropy production rate. The observation $x \bmod 2$ is also lossy, and has none.*

Proof. The identity observation is injective. The map $x \bmod 3$ is lumpable, with coarse chain the biased cycle on $\mathbb{Z}_3$ with the same q, whose stationary distribution is uniform and whose production is the same expression; an exact enumeration of all paths of length 4 confirms $\Sigma_4^c = 4\Sigma_1$. The map $x \bmod 2$ is lumpable with deterministic alternation as coarse chain, and the argument of Example 31.2 gives $\Sigma_n^c = 0$. □

The proposition establishes that entropy production is not measured by how much of the past is lost, which in turn bears on the arrow of time: an arrow, in the sense of Chapter 35, is an asymmetry of path measures, and a lossy observation can retain it or erase it. The reversible bijection on the integers of Proposition 30.2 shows the converse direction: loss of reconstruction in a description can be produced by projection alone.

Chapter 52

Failure to Compose Reduced Descriptions

The contextual equivalence of Theorem 10.2 has a general form, in which the quotient depends on the allowed contexts in a monotone way.

Let G be a set, $q\colon G \to Q$ an observation, and $\mathcal{C}$ a family of maps $c\colon G \to G$. Define $g \equiv_{\mathcal{C}} g'$ if and only if $q(c(g)) = q(c(g'))$ for all $c \in \mathcal{C} \cup \{\mathrm{id}\}$.

Proposition 52.1 (Monotonicity and universality of contextual equivalence)**.** *If $\mathcal{C} \subseteq \mathcal{C}'$ then $\equiv_{\mathcal{C}'}$ refines $\equiv_{\mathcal{C}}$: enlarging the context family can only refine the quotient. If $\mathcal{C}$ is closed under composition then $\equiv_{\mathcal{C}}$ is invariant under every $c \in \mathcal{C}$, and it is the largest $\mathcal{C}$-invariant equivalence relation contained in* $\ker q$.

Proof. For monotonicity, equality of $q \circ c$ for all $c \in \mathcal{C}' \cup \{\mathrm{id}\}$ implies it for all $c \in \mathcal{C} \cup \{\mathrm{id}\}$. For invariance, suppose $g \equiv_{\mathcal{C}} g'$ and $c \in \mathcal{C}$; for $c' \in \mathcal{C} \cup \{\mathrm{id}\}$ the composite $c' \circ c$ lies in $\mathcal{C}$ by closure, so $q(c'(c(g))) = q(c'(c(g')))$, which is $c(g) \equiv_{\mathcal{C}} c(g')$. If $\sim$ is a $\mathcal{C}$-invariant equivalence contained in $\ker q$ and $g \sim g'$, then $c(g) \sim c(g')$ for each c, so $q(c(g)) = q(c(g'))$, hence $g \equiv_{\mathcal{C}} g'$. □

For the unitary group $G = U(d)$ and the projection $q = T$, three context families give three quotients.

Proposition 52.2 (A hierarchy of quotients for the probability projection)**.** *Let $\mathcal{C}_\emptyset = \varnothing$, let $\mathcal{C}_L = \{U \mapsto AU : A \in U(d)\}$ and let $\mathcal{C}_{LR} = \{U \mapsto AUB : A, B \in U(d)\}$. Then $U \equiv_\emptyset U'$ if and only if $T(U) = T(U')$; $U \equiv_L U'$ if and only if $U' = UD$ for a diagonal unitary D; and $U \equiv_{LR} U'$ if and only if $U' = e^{i\varphi}U$. The inclusions $\equiv_{LR} \subsetneq \equiv_L \subsetneq \equiv_\emptyset$ are strict for $d \geq 2$.*

Proof. The first is the definition. For the second, $T(AU) = T(AU')$ for all A says $|\langle w|Ue_j\rangle| = |\langle w|U'e_j\rangle|$ for all unit vectors w and all j, so each column $U'e_j$ is a unit-modulus multiple of Ue_j (as in the proof of Theorem 9.1), which is $U' = UD$; the converse is immediate. The third is Theorem 10.2. Strictness of the second inclusion: $D = \mathrm{diag}(1, i)$ gives $U' = D$ and $U = I$ with $U \equiv_L U'$ but not $U \equiv_{LR} U'$ since D is not scalar. Strictness of the first: $U = R(\pi/8)$ and $U' = R(-\pi/8)$ have equal T, while $U^{-1}U' = R(-\pi/4)$ is not diagonal, so $U \not\equiv_L U'$. □

The hierarchy identifies the remedy with the allowed contexts. If the only contexts are the measurements that follow the operation and the input is a fixed basis vector, the rays $[Ue_j]$ suffice, and the quotient is the classes UD. If the operation is embedded between arbitrary unitary operations on both sides, the quotient is $PU(d)$. The projection T with no contexts is a much coarser object, and its failure to compose, Proposition 7.2, is the failure of the relation $\equiv_\emptyset$ to be a congruence. A statement that "phases must be restored" therefore has no content until the contexts are named, and an enlargement of the contexts can only refine the quotient, strictly at each of the three stages recorded here.

Chapter 53

Failure to Extend a Local Continuation Globally

The worked result is the spanning-tree and cycle-product criterion of Theorem 29.1: a context graph with edge data B_e is globally realizable if and only if there is one lift assignment for which every fundamental cycle of a spanning tree has trivial product. Two features of the criterion deserve emphasis.

First, the probability data determine sets of lifts and not lifts. An edge datum has in general several lifts, eight for a real 2×2 matrix with entries in $(0, 1)$, and the holonomy of a lift assignment belongs to the assignment. A nonidentity cycle product for one chosen assignment proves nothing about realizability, and what realizability requires is a single common assignment satisfying all $\beta_1(G)$ cycle conditions.

Second, local admissibility is no evidence of global compatibility. Trees carry no constraint, by Proposition 11.1, so a path graph of individually admissible edges always glues. The first graph with a constraint is the triangle, and there the individually admissible data of the mutually unbiased triangle do not glue over $\mathbb{R}$. For the triangle with real 2×2 data the criterion is the closure equation of Proposition 29.3:

$$\varepsilon_1\alpha_1 + \varepsilon_2\alpha_2 + \varepsilon_3\alpha_3 \equiv 0 \pmod{\pi} \quad \text{for some signs } \varepsilon_i,$$

a single equation for a triple of angles, so that in the real two-dimensional setting, for independently and continuously distributed angles, realizability has probability zero. For the unbiased angles $\alpha_i = \pi/4$ the left side is an odd multiple of $\pi/4$ for every choice of signs, and an exhaustive enumeration of the $8^3 = 512$ triples of lifts finds none with product $\pm I$.

The comparison between fields is the clearest demonstration that gluing is a property of the pair of data and field. Over $\mathbb{C}$ the same data glue (Proposition 50.5) and the global frames give trivial cycle products by Theorem 29.1, while the Bargmann invariant of the

first vectors is $(1 + i)/4$, of argument $\pi/4$. The two numbers answer different questions. Trivial holonomy of the frame cycle is the existence of a realization; the Bargmann phase is a gauge-invariant functional of that realization, equal in this case to half the solid angle of the octant triangle on the Bloch sphere. Refining the edge data from probabilities to projective operations does not repair a failure of gluing, as Proposition 50.4 shows; it turns the existential question into a single check of cycle products. The converse implication holds: if projective data glue then so do the probabilities, by projecting.

Chapter 54

Failure to Preserve Distinctions Under Observation

Three statements about the distinguishability of two laws after observation are logically different, and the accessibility results of Part VII hold at different strengths. A pair (ρ_0, ρ_1) is *exactly identified* by a channel C if $C\rho_0 = C\rho_1$. It is *ϵ-indistinguishable by an observer class* if the total variation between the observed laws is at most ϵ for every channel in the class. And the observed distinguishability *converges to zero* if it tends to zero along a family of channels C_t without necessarily vanishing at any finite time. None of the three, alone, establishes that the underlying structure has homogenized.

Proposition 54.1 (The three notions and physical homogenization)**.** *In the Gaussian channel family $Y = W_t\Delta_t + \xi_t$ with $\xi_t \sim N(0, \Sigma_t)$ and $\Sigma_t \succeq \sigma^2 I$, $\sigma > 0$: (i) exact identification holds if and only if $W_t\Delta_t = 0$; (ii) ϵ-indistinguishability holds if and only if $\|\Sigma_t^{-1/2} W_t\Delta_t\| \le a_\epsilon$; (iii) convergence to zero holds if and only if $\|\Sigma_t^{-1/2} W_t\Delta_t\| \to 0$. Each can hold while $\|\Delta_t\|$ is arbitrarily large or growing. If $\|W_t\| \le w$ uniformly then physical homogenization $\|\Delta_t\| \to 0$ implies (iii), and the converse is false.*

Proof. The three criteria are Theorem 47.1 applied to $\mathrm{TV} = 2\Phi_{\mathrm{N}}(\|\Sigma_t^{-1/2} W_t\Delta_t\|/2) - 1$. For (i), take $\Delta \in \ker W$ nonzero and of any size. For (ii) take $W_t = \gamma\, I$ with $\gamma\|\Delta_t\| \le \sigma a_\epsilon$ and $\|\Delta_t\|$ large by taking γ small. For (iii), $\|\Sigma^{-1/2} W\Delta\| \le \sigma^{-1} w\|\Delta_t\| \to 0$ when $\|\Delta_t\| \to 0$. The converse fails by Proposition 47.2, where $\|\Delta_t\|$ increases while the observed distinguishability decreases. □

The qualification attached to the cosmological results should be stated again here. An event horizon alone does not establish a family of nested observation channels. What is needed is that the relevant domains be specified, that points of the domains be identified across times, for example by comoving coordinates, and that the channel at the later time

be shown to factor through the channel at the earlier time, as in Theorem 46.1. Shrinking comoving reach describes the region that can still influence the observer. It does not describe the information that the observer has accumulated, which is a different object, and the two have opposite monotonicity.

Theorem 54.2 (The cumulative record is monotone)**.** *Let the alternatives be laws ρ_0, ρ_1 on complete histories, fixed in time, and let the data available at time t be the record $Y_{\le t} = (Y_s)_{s \le t}$ of all the data collected up to t, each Y_s the output of a channel C_s applied to the history. Then for $t_1 \le t_2$ the channel producing $Y_{\le t_1}$ factors through the channel producing $Y_{\le t_2}$, and for every divergence satisfying data processing the accessible distinction of the cumulative record is nondecreasing in t.*

Proof. $Y_{\le t_1}$ is a coordinate projection of $Y_{\le t_2}$, a deterministic map and hence a Markov kernel. Proposition 43.1 applied to this kernel gives $\mathfrak{D}(C_{\le t_1}\rho_0, C_{\le t_1}\rho_1) \le \mathfrak{D}(C_{\le t_2}\rho_0, C_{\le t_2}\rho_1)$. □

The accessible distinction of the *current-reach channel* may decline, as Theorem 46.1 describes when its nested hypothesis holds, while that of the *retained record* cannot. An observer who preserves earlier data is therefore outside the observer class of Theorem 47.3, and the Smoothness Paradox, in the form established, is a statement about observers who do not. It follows that any inference from declining accessibility to declining physical structure needs two explicit ingredients: the factorization between structure strength and channel quality (Proposition 47.2) and the specification of which channel, current reach or retained record, is being described.

Chapter 55

Computational Experiments

The claims of the book rest on different kinds of evidence, and the computations are collected here by what each establishes. A single script, listed in Appendix F and executable as `python3 experiments.py`, reproduces all of them. It requires only numpy, scipy and sympy, runs in under ten seconds on a laptop, and prints one line per check with the measured value, the tolerance and a pass or fail verdict. All thirty-three checks pass. The categories are the following.

Category A, exact finite enumeration. The statement is proved by exhaustive inspection of a finite set, and the numerical output is an integer or an identity up to floating-point rounding of exact rationals. *Category B, symbolic verification.* A computer algebra system checks an identity for arbitrary smooth functions. *Category C, numerical corroboration of a proved result.* The result has a proof in the text and the computation tests the proof, the statement of the formula, or the parameters of an example; a pass is evidence of the absence of an algebraic slip and not an independent proof. *Category D, numerical exploration of an unproved claim.* The claim has no proof in the text, and the computation is the only evidence.

ID	Claim checked	Method	Result and tolerance
A1	eight lifts of J in $O(2)$; no triple has product $\pm I$	all $8^3 = 512$ triples	count $= 0$, exact
A2	Rule 184, $N = 4$: $1100 \to 1010$, $0101 \to 1010$ (update not injective), $1010 \to 0101$	direct update	exact
A3	hard-core word counts for $n = 1, \dots, 8$	enumeration	$2, 3, 5, 8, 13, 21, 34, 55$
A4	biased $\mathbb{Z}_4$ cycle, paths of length 4: Σ_4, fluctuation relation, parity-coarse Σ_4^c	all 4^5 paths	$\Sigma_4 - 4\Sigma_1 < 10^{-14}$; sum $= 1$; $\Sigma_4^c = 0$
A5	Hadamard, Z counterexample, weakly consistent pair, 3×3 Fourier	matrix algebra	$\leq 10^{-12}$
A6	doubly stochastic matrix with two columns sharing one row	inspection	exact
A7	$\mathbb{Z}_6$ observed mod 3: $\Sigma_4^c = 4\Sigma_1$	all 6^5 paths	$< 10^{-14}$

Table 55.1: Category A: exact finite enumeration.

ID	Claim checked	Method	Result
B1	energy-flux identity $\partial_t e = -\partial_x((e + p)v)$ + dissipative divergences for arbitrary $\varepsilon(\Phi, S)$	sympy, reversible part	residual 0
B2	Madelung: imaginary part is continuity, real part is Hamilton–Jacobi with Q	sympy, polar form	both identities hold

Table 55.2: Category B: symbolic verification.

The distinctions are meant to be assessed at a glance. The 512-triple enumeration (A1) is a proof of its statement for the stated lifts. The fluid checks (C9, C10) and the spectral computations (C12 to C15) test a proved theorem and a proved calculation, respectively, and do not substitute for the proofs. The cosmological checks of category C test the background formulas, and the turnover scale and its uniqueness (D1 to D3) are explorations whose uniqueness is claimed for a grid and for this background only. A reader who wishes to challenge a result should look first at its category.

ID	Claim checked	Method	Result and tolerance
C1	$K_1 = T(U^2) - T(U)^2$ and $M_{PP} = T(U)$	random 3×3 unitary, Liouville blocks	5×10^{-16}; tol 10^{-12}
C2	Hadamard kernel $K_1 = I - J$, $K_m = 0$ for $m = 2, \dots, 6$	Liouville blocks	2×10^{-16}
C3	Bargmann invariant equals $\mathrm{Tr}(\rho_1\rho_2\rho_3)$ formula	random qubit triple	2×10^{-16}
C4	qubit interference term $-\frac{1}{2}\sin\theta_1 \sin\theta_2$	direct evaluation	0
C5	geometric phase $-\pi(1 - \cos\vartheta)$	product of 2×10^4 overlaps	9×10^{-9}; tol 10^{-6}
C6	complex triangle Z, X, Y all transitions J; Bargmann $(1 + i)/4$	explicit frames	10^{-16}
C7	Hankel rank of the unknown-bias coin	words up to length 4, $p = 0.8$	rank 2
C8	even process: $C = 0.6365$, $h = 0.4621$ nats	closed form	agrees to 5×10^{-5}
C9	fluid: drift of mass, momentum, energy	RK4, 128 modes, $dt = 2 \times 10^{-4}$, $t \le 2$	1.8×10^{-15}
C10	fluid: entropy increases at ten samples	same run	minimum increment 7×10^{-4}
C11	$c^2 = p_\Phi + \frac{S}{\Phi} p_S = \gamma R \vartheta$	ideal gas	4×10^{-16}
C12	sound eigenvalue vs $icq - \Gamma q^2$	3×3 eigenproblem, $q = 0.01$	relative error 6×10^{-9}; at $q = 0.04$ the error is $16\times$ larger
C13	$\mathrm{Re}\,\lambda < 0$ for $0.05 \le q \le 5$	100 wavenumbers, $\mu, \kappa > 0$	all negative
C14	$\kappa = 0$ leaves an eigenvalue 0	$q = 0.7$	6×10^{-17}
C15	characteristic speeds $-c, 0, c$	eigenvalues of A	9×10^{-16}
C16	growth factor: integral formula vs ODE	$D(0.5)/D(1)$, tolerance 10^{-11}	5×10^{-12}
C17	event horizon asymptote $a\chi_e\sqrt{\Omega_\Lambda} \to 1$	quadrature at $a = 100$	5×10^{-8}
C18	$d = 2$: every unitary with moduli $\sqrt{p}, \sqrt{1-p}$ is $D_1 R(\theta) D_2$	200 random unitaries	2×10^{-15}

Table 55.3: Category C: numerical corroboration of proved results.

ID	Claim explored	Method	Result
D1	location of the turnover a_* of $D_+^2\chi_e^3$	root of $2f - 3/(aH\chi_e)$, $\Omega_m = 0.3$	$a_* = 0.4891$
D2	maximum of $D_+^2\chi_e^3$ on a grid	500 points, step 0.01	0.4865, within the grid step
D3	uniqueness of the sign change of the rate	400 points on $[0.05, 20]$	one sign change

Table 55.4: Category D: numerical exploration of unproved claims.

Chapter 56

Open Problems

The book establishes several precise ways in which reduced descriptions succeed or fail, develops an explicit continuum realization, and identifies the additional constructions that the proposed physical unification would require. The remaining obligations are the following, each stated with what would count as a solution.

The first concerns the geometry of the continuation quotient. Part I constructs the minimal state as a set, and Part IV assumes a smooth manifold and a bundle on it as additional data. The open problem is a theorem giving conditions on a history structure under which the continuation quotient is a smooth manifold, or a stratified space, and a family of continuation sets forms a vector bundle over it. The unitary example, whose quotient is complex projective space, is the model; the constant-rank construction of Part IV identifies one sufficient route. A solution would give intrinsic conditions that can be checked from the history structure alone.

The second is the controlled derivation of continuum closure. The exact closure of Example 36.2 rests on a special duality. The problem is to derive equations of the form of the model of Part VI from a microscopic stochastic or deterministic dynamics with an explicit error bound, for a family of microscopic models broad enough to include a nontrivial equation of state, and to identify the memory term that the first-order truncation discards. A solution is a theorem, with hypotheses, that bounds the difference between the coarse field and the solution of the closed equations.

The third is nonlinear stability. Theorem 39.2 concerns the linearization about a uniform rest state. The questions of global existence and of nonlinear stability for the model, with the constitutive assumptions of a general convex internal energy, are not addressed, and results in the literature on the one-dimensional compressible Navier–Stokes–Fourier system apply to restricted constitutive relations whose relation to the present model has not been checked. A solution is a statement of the conditions under which smooth solutions exist

globally and converge to a uniform state.

The fourth is a covariant dissipative theory. The dissipative operator of Part VI is constructed for a nonrelativistic model. The first-order relativistic dissipative theories are known to be unstable [20], and the second-order theories of Israel and Stewart [9] restore causality by adding relaxation variables, at the cost of enlarging the state. The problem is to formulate a relativistic extension that retains the structure of a reversible bracket and a symmetric dissipative operator with the degeneracy conditions, and to determine whether the additional variables are the memory variables of Chapter 34.

The fifth is the derivation of the quantum potential. Proposition 41.1 makes unitary evolution a conservative irrotational fluid with a pressure-like potential Q. The problem is to show that a microscopic closure of the capacity–flow–entropy fields yields that term, rather than postulating it, and to identify the amplitude space, inner product and unitarity conditions that the classical closure lacks. A solution would be a derivation, with the assumptions under which the Madelung structure appears.

The sixth is the relation between the acoustic and gravitational geometries. Conjecture 42.1 is unsupported. A solution is either a construction in which the gravitational metric of a Palatini theory with fluid matter coincides with the acoustic metric to a stated order, or a proof that no such choice exists for a stated class of closures, together with the corresponding statement about the conformal factor that the characteristics leave undetermined.

The seventh is the cosmological problem. The accessibility model of Part VII establishes consequences on a supplied background. The problem is to derive an expansion history and a growth factor from the field equations of the programme, to compute the accessibility quantities on that derived background, and to test the result against a specified survey with a channel model of the kind described in Chapter 49. Nothing in the book supplies this derivation, and the turnover scale computed there is a property of the assumed background only.

The remaining problems are internal to the framework. The sizes of the forward and reverse minimal states of a process can differ [5], and the book does not relate the difference to the entropy production Σ_1. The holonomy of Part IV has been shown to be independent of entropy production unless a thermodynamic dynamics is specified, and the conditions linking them have not been established. The accessibility results are proved for Gaussian channels and linear structure growth, and the extension to non-Gaussian channels, nonlinear statistics and other observer classes is open. The minimal-witness and extremal questions of the author's reconstruction of quantum theory [7] are outside this book, and the extension of the contextual hierarchy of Chapter 52 to composite systems and the tensor-product rule has not been carried out here.

The book therefore ends in the following position. It has fixed a definite object, the continuation quotient of a history structure, and a small set of tests under which an observation of it is adequate or inadequate. It has shown that the tests are distinct, supplied the canonical remedy for each where one exists, and recorded the examples that separate them. It has built a continuum model in which reversible transport, irreversible relaxation, conservation laws and stability can be proved, and it has shown how unitary evolution, an effective propagation geometry and a gravitational sector can be attached to it, each with the additional structure it requires. What it has not done is derive the physics from the mathematics, and the list above says what would have to be proved for that to be done.

Appendix A

Notation

The table lists the symbols used in more than one place, grouped by the part in which they are introduced. Several letters carry more than one meaning, because the book draws on literatures with conflicting conventions; in every case the meaning is fixed by the part and by the context, and the collisions are listed after the table.

Symbol	Meaning	Part
$\mathcal{H},\ \mathcal{F}$	set of histories; common set of future segments	I
$\mathcal{A}(h)$	admissible continuations of the history h, a subset of $\mathcal{F}$	I
$h \cdot s$	extension of a history by a one-step prefix s of a continuation	I
$\pi,\ \mathcal{X}$	an observation, a surjection $\mathcal{H} \to \mathcal{X}$; its set of values	I
$\pi_{\mathcal{A}},\ \sim_{\mathcal{A}}$	minimal sufficient observation; continuation equivalence	I
π_k	depth-k observation, the last k letters of a history	I
$\preceq$	refinement order on observations	I
δ_π	sufficiency defect of an observation	I
$P(h)$	law of the future after the history h	II
$T(U)$	transition-probability matrix $\lvert U_{ij}\rvert^2$ of a unitary or orthogonal U	II
J	the matrix with every entry $\frac{1}{2}$ (size 2)	II
Δ	dephasing channel, projection onto diagonal matrices	II
$PU(d),\ PO(d)$	projective unitary and orthogonal groups	II
$\equiv_{\mathcal{C}}$	contextual equivalence for a context family $\mathcal{C}$	II, VIII
$M_{PP}, M_{PQ}, M_{QP}, M_{QQ}$	blocks of a linear map for a splitting $V = W \oplus W'$	II

Symbol	Meaning	Part
K_m	memory kernel, $M_{PQ}M_{QQ}^{m-1}M_{QP}$	II
$D(\alpha,\beta)$	decoherence functional of two histories α,β	II
$B_e,\ \beta_1(G)$	edge transition matrix on a context graph; cycle rank	II
α_e	$\arccos\sqrt{p_e}$, the angle of a real 2×2 edge matrix	II, IV
$Q,\ \delta,\ \Sigma$ (automata)	state set, transition function, alphabet	III
$C,\ h_\mu$	statistical complexity and entropy rate of a process	III, V
$F_a,\ L_{ba}$	frame of a context a; relative transformation $F_b^\dagger F_a$	IV
H_c	cycle product of lifts around a cycle c	IV
$\nabla,\ A,\ F$	connection, connection form, curvature	IV
$P_n,\ R_n$	forward and reversed path measures of length n	V
$\Sigma_n,\ \Sigma_1$	entropy production of paths of length n; per step	V
$\sigma(\omega)$	pathwise log-ratio $\ln(P_n/R_n)$	V
KL, $D(\cdot\|\cdot)$	relative entropy	V, VII
$\Phi,\ \mathbf{m},\ \mathbf{v},\ S$	capacity density, momentum density, velocity, entropy density	VI
$\varepsilon(\Phi,S),\ \vartheta,\ p$	internal energy density; temperature ε_S; pressure	VI
$e,\ E,\ \mathcal{S}$	total energy density; energy functional; entropy functional $\int S\,dx$	VI
$\mathcal{L},\ \mathcal{M}$	reversible (Poisson) and dissipative operators	VI
$\mu,\ \kappa$	viscosity and thermal conductivity	VI
c	characteristic (sound) speed, $c^2 = p_\Phi + \frac{S}{\Phi}p_S$	VI
g_{ac}	acoustic metric	VI
$\mathfrak{D}$	distinguishability measure	VII
C_t	observer channel at time t (also the cumulative-record channel)	VII
$s,\ \gamma$	signal-to-noise ratio; effective gain-to-noise ratio $\lvert W\rvert^2/N$	VII
a_ϵ	$2\Phi_{\mathrm{N}}^{-1}((1+\epsilon)/2)$, the threshold for tolerance ϵ	VII
$D_+,\ f$	linear growth factor; growth rate $d\ln D_+/d\ln a$	VII
$E(a),\ \chi_e,\ \chi_p$	H/H_0; comoving event and particle horizons	VII
F, Sm, A	formation, smoothing and access-loss rates (Chapter 45)	VII
τ	a task, a map on detailed objects	VIII

Overloaded symbols. The letter T denotes the probability projection in Parts II and III and the transfer factor T_k of Part VII; the temperature is written ϑ to avoid a third use. The letter H denotes the Hadamard matrix in Parts II and III, the cycle product H_c in Part IV, the Hessian of the internal energy in Parts VI and VII, and the Hubble rate in Part VII; and $H(\cdot)$ with an argument is the Shannon entropy in Part V. The letter Q denotes the complementary projection in Chapter 13, the state set of an automaton in Part III, the Lyapunov functional in Chapter 39, and the quantum potential in Chapter 41. The letter D denotes the decoherence functional in Part II and the relative entropy in Parts V and VII, while the growth factor is written D_+. The letter Σ is an alphabet in Part I and Part III, a covariance in Parts VII and VIII, and the entropy production in Part V. The letter S is the entropy density in Parts VI and VII. The letter F is a frame in Part IV, the curvature in Part IV, and the formation rate in Part VII.

Appendix B

Divergences and Gaussian Channels

This appendix proves the information-theoretic facts used in Parts V, VII and VIII, on finite sets and for Gaussian laws. Throughout, μ, ν are probability distributions on a finite set X and C is a Markov kernel from X to a finite set Y, so that $(C\mu)(y) = \sum_x C(y|x)\mu(x)$.

Lemma B.1 (Log-sum inequality)**.** *For nonnegative numbers $a_1, \dots, a_n$ and $b_1, \dots, b_n$ with $b_i > 0$,*

$$\sum_i a_i \ln \frac{a_i}{b_i} \geq \Big(\sum_i a_i\Big) \ln \frac{\sum_i a_i}{\sum_i b_i},$$

with equality if and only if a_i/b_i is constant over $\{i : a_i > 0\}$.

Proof. The function $t \mapsto t \ln t$ is strictly convex on $[0, \infty)$. Put $B = \sum_i b_i$ and weights $w_i = b_i/B$. Jensen's inequality for the points $t_i = a_i/b_i$ gives $\sum_i w_i t_i \ln t_i \geq \big(\sum_i w_i t_i\big) \ln\big(\sum_i w_i t_i\big)$, and $\sum_i w_i t_i = \sum_i a_i/B$. Multiplying by B gives the claim. Equality in Jensen's inequality for a strictly convex function holds exactly when the t_i are equal over the points with positive weight. □

Theorem B.2 (Data processing for f-divergences)**.** *Let f be convex on $(0, \infty)$ with $f(1) = 0$, and $D_f(\mu\|\nu) = \sum_x \nu(x) f(\mu(x)/\nu(x))$ for ν of full support. Then $D_f(C\mu\|C\nu) \leq D_f(\mu\|\nu)$. In particular this holds for $f(t) = t \ln t$ (relative entropy) and $f(t) = \frac{1}{2}|t-1|$ (total variation).*

Proof. Assume $C\nu$ has full support. For each y, put $w_{x|y} = C(y|x)\nu(x)/(C\nu)(y)$, a probability distribution in x. Then $(C\mu)(y)/(C\nu)(y) = \sum_x w_{x|y}\, \mu(x)/\nu(x)$, so by Jensen's inequality $f\big((C\mu)(y)/(C\nu)(y)\big) \leq \sum_x w_{x|y} f\big(\mu(x)/\nu(x)\big)$. Multiplying by $(C\nu)(y)$ and summing over y gives $D_f(C\mu\|C\nu) \leq \sum_{x,y} C(y|x)\nu(x) f(\mu(x)/\nu(x)) = D_f(\mu\|\nu)$, because $\sum_y C(y|x) = 1$. □

For the relative entropy the proof by Lemma B.1 shows the equality condition used in Proposition 43.1: for each y equality holds if and only if $\mu(x)/\nu(x)$ is constant over the x with $C(y|x) > 0$ and $\mu(x) > 0$.

Proposition B.3 (Contraction of total variation)**.** *For signed measures $\theta = \mu - \nu$ of total mass zero, $\|C\theta\|_1 \le \|\theta\|_1$, and $\mathrm{TV}(C\mu, C\nu) \le \mathrm{TV}(\mu, \nu)$ with $\mathrm{TV} = \frac{1}{2}\|\cdot\|_1$.*

Proof. $\|C\theta\|_1 = \sum_y |\sum_x C(y|x)\theta(x)| \le \sum_{x,y} C(y|x)|\theta(x)| = \|\theta\|_1$. □

Proposition B.4 (Gaussian shift)**.** *Let $P_i = N(m_i, \Sigma)$ on $\mathbb{R}^n$ with $\Sigma \succ 0$, and put $\Delta = \Sigma^{-1/2}(m_1 - m_0)$ and $s = \|\Delta\|^2$. Then $\mathrm{KL}(P_1\|P_0) = \frac{1}{2}s$ and $\mathrm{TV}(P_0, P_1) = 2\Phi_{\mathrm{N}}(\sqrt{s}/2) - 1$.*

Proof. Both quantities are invariant under the invertible affine map $y \mapsto \Sigma^{-1/2}(y - m_0)$, which sends the laws to $N(0, I)$ and $N(\Delta, I)$. Rotating coordinates so that $\Delta = \sqrt{s}\, e_1$ leaves the remaining coordinates identically distributed under both laws, so only the first coordinate matters. The relative entropy of $N(\sqrt{s}, 1)$ from $N(0, 1)$ is $\mathbb{E}_{N(\sqrt{s},1)}\big[-\frac{1}{2}(y-\sqrt{s})^2 + \frac{1}{2}y^2\big] = \mathbb{E}[\sqrt{s}y - \frac{s}{2}] = \frac{s}{2}$. For the total variation, the densities cross at $y = \sqrt{s}/2$, the set where the second exceeds the first is $\{y > \sqrt{s}/2\}$, and $\mathrm{TV} = \int_{\sqrt{s}/2}^{\infty}\big(\varphi(y - \sqrt{s}) - \varphi(y)\big)dy = \Phi_{\mathrm{N}}(\sqrt{s}/2) - \Phi_{\mathrm{N}}(-\sqrt{s}/2) = 2\Phi_{\mathrm{N}}(\sqrt{s}/2) - 1$. □

Proposition B.5 (Gaussian variance shift)**.** *For scalars, $\mathrm{KL}\big(N(0, \sigma_1^2)\,\|\,N(0, \sigma_0^2)\big) = \frac{1}{2}\big(r - 1 - \ln r\big)$ with $r = \sigma_1^2/\sigma_0^2$. With $r = 1 + s$ this is $g(s) = \frac{1}{2}(s - \ln(1 + s))$, whose derivative $g'(s) = \frac{s}{2(1+s)}$ is positive for $s > 0$.*

Proof. $\mathrm{KL} = \mathbb{E}_1\big[-\frac{1}{2}\ln\sigma_1^2 - \frac{y^2}{2\sigma_1^2} + \frac{1}{2}\ln\sigma_0^2 + \frac{y^2}{2\sigma_0^2}\big] = -\frac{1}{2}\ln r - \frac{1}{2} + \frac{r}{2}$. □

Proposition B.6 (Cascade of Gaussian channels)**.** *If $Y_1 = g_1\delta + \xi_1$ and $Y_2 = g_2 Y_1 + \xi_2$ with independent $\xi_i \sim N(0, N_i)$, then $Y_2 = g_1 g_2 \delta + \xi$ with $\xi \sim N(0, g_2^2 N_1 + N_2)$, and the signal-to-noise ratio $s_2 = g_1^2 g_2^2 |\Delta|^2/(g_2^2 N_1 + N_2)$ does not exceed $s_1 = g_1^2|\Delta|^2/N_1$.*

Proof. Substitute, and add the variances of the independent noises. Then $s_2 \le s_1$ is $g_2^2 N_1 + N_2 \ge g_2^2 N_1$. □

Appendix C

The Lie–Poisson Structure of the Fluid Bracket

This appendix gives a direct proof of the Jacobi identity for the reversible bracket of Chapter 38, verifies the claims about Casimirs, and states how the structure behaves under a change of variables.

The Lie algebra

Let $\mathfrak{X} = \mathrm{Vect}(\mathbb{S}^1)$ be the vector fields on the circle with the bracket $[X, Y] = XY' - YX'$, and let $\mathcal{F} = C^\infty(\mathbb{S}^1)$, on which X acts by $u \mapsto Xu = X\,u'$. Put $\mathfrak{g} = \mathfrak{X} \ltimes (\mathcal{F} \oplus \mathcal{F})$ with the bracket

$$\big[(X_1, u_1, w_1), (X_2, u_2, w_2)\big] = \big([X_1, X_2],\ X_1u_2 - X_2u_1,\ X_1w_2 - X_2w_1\big).$$

Proposition C.1 ($\mathfrak{g}$ is a Lie algebra)**.** *The bracket is bilinear and antisymmetric and satisfies the Jacobi identity.*

Proof. Bilinearity and antisymmetry are immediate. The first component is the Lie bracket of vector fields, which satisfies the Jacobi identity. The second and third components have the same form, so it suffices to treat (X_i, u_i). The second component of $[[a_1, a_2], a_3]$ is $[X_1, X_2]u_3 - X_3(X_1u_2 - X_2u_1)$, that is, $X_1X_2u_3 - X_2X_1u_3 - X_3X_1u_2 + X_3X_2u_1$, using $[X, Y]u = X(Yu) - Y(Xu)$. Adding the two cyclic permutations gives twelve terms, which cancel in pairs:

$$(X_1X_2u_3 - X_2X_1u_3 - X_3X_1u_2 + X_3X_2u_1) + (X_2X_3u_1 - X_3X_2u_1 - X_1X_2u_3 + X_1X_3u_2) + (X_3X_1u_2 - X_1X_3u_2 - X_2X_3u$$

A symbolic computation with arbitrary functions X_i, u_i confirms the vanishing of both

components. □

The bracket on the dual

The dual of $\mathfrak{g}$ is identified, by the pairing $\langle (m,\Phi,S),(X,u,w)\rangle = \int (mX+\Phi u+Sw)\,dx$, with triples of a momentum density m and two densities Φ, S. For functionals F, G of $z=(m,\Phi,S)$ with gradients $\delta F = (F_m, F_\Phi, F_S) \in \mathfrak{g}$, the Lie–Poisson bracket is

$$\{F,G\}(z) = -\big\langle z, [\delta F, \delta G]\big\rangle = -\int \Big[m\big(F_m G'_m - G_m F'_m\big) + \Phi\big(F_m G'_\Phi - G_m F'_\Phi\big) + S\big(F_m G'_S - G_m F'_S\big)\Big]\,dx,$$

which is the bracket of Chapter 38.

Proposition C.2 (Jacobi identity for the Lie–Poisson bracket)**.** $\{\{F,G\},H\}+\{\{G,H\},F\}+\{\{H,F\},G\} = 0$.

Proof. For linear coordinate functionals $\ell_a(z) = \langle z, a\rangle$ with $a \in \mathfrak{g}$ one has $\delta \ell_a = a$ and $\{\ell_a, \ell_b\} = -\ell_{[a,b]}$. Hence the Jacobiator of three such functionals is $-\ell$ of the Jacobiator of the Lie algebra, which vanishes by Proposition C.1. The Jacobiator of a bracket satisfying the Leibniz rule in each argument is a trilinear derivation, so it is determined by its values on the coordinate functionals, and it vanishes for all functionals (see Holm, Marsden, Ratiu and Weinstein [21]). □

Casimirs and the generated flow

A Casimir is a functional C with $\{C,G\} = 0$ for all G. For $G = N = \int \Phi\,dx$ or $G = \mathcal{S} = \int S\,dx$, each term of the bracket contains either a derivative of a component of δG, which vanishes because $\delta N = (0,1,0)$ and $\delta \mathcal{S} = (0,0,1)$ are constant, or the component $G_m = 0$; hence $\{F,N\} = \{F,\mathcal{S}\} = 0$ for every F. These are the Casimirs asserted in Theorem 38.2. The flow of the energy $E = \int \big(m^2/2\Phi + \varepsilon(\Phi,S)\big)dx$ is computed from $E_m = v$, $E_\Phi = -v^2/2 + \varepsilon_\Phi$ and $E_S = \vartheta$ and is the reversible part of Equations (38.1)–(38.3), as shown in the proof of that theorem.

Change of variables and the degeneracy conditions

Let $z' = \psi(z)$ be a diffeomorphism acting pointwise, without derivatives, such as $(\Phi, m, S) \mapsto (\Phi, m, e)$. If $\mathcal{L}$ is a Poisson operator for z and $\mathcal{M}$ a symmetric operator, they transform as $\mathcal{L}' = J\mathcal{L}J^{\mathsf{T}}$ and $\mathcal{M}' = J\mathcal{M}J^{\mathsf{T}}$ with $J = \partial z'/\partial z$, gradients transform as $\delta F/\delta z = J^{\mathsf{T}}\,\delta F/\delta z'$, and the Jacobi identity, symmetry and positive semidefiniteness of the operators are preserved, as is the pairing $\langle \delta F, \mathcal{L}\,\delta G\rangle$. The degeneracy conditions $\mathcal{L}\,\delta\mathcal{S} = 0$ and $\mathcal{M}\,\delta E = 0$

are statements about the vectors $\mathcal{L}\,\delta\mathcal{S}$ and $\mathcal{M}\,\delta E$, which transform as J times the corresponding vectors in the other coordinates, so they hold in one set of variables if and only if they hold in the other. This is why the dissipative form was constructed in the variables (Φ, m, e), where the energy covector is constant, and transferred to (Φ, m, S).

Appendix D

The Geometry of the Probability Projection

Doubly stochastic matrices

Let $\mathrm{DS}(d)$ denote the set of $d \times d$ doubly stochastic matrices, those with nonnegative entries and all row and column sums equal to 1.

Theorem D.1 (Birkhoff–von Neumann)**.** $\mathrm{DS}(d)$ *is the convex hull of the* $d!$ *permutation matrices [1].*

Proposition D.2 (Dimension)**.** *The affine hull of* $\mathrm{DS}(d)$ *has dimension* $(d-1)^2$*.*

Proof. The $2d$ linear conditions on the d^2 entries have rank $2d-1$, since the sum of the row conditions equals the sum of the column conditions and no other dependence exists. The solutions of the homogeneous system are matrices with zero row and column sums, a space of dimension $d^2 - (2d-1) = (d-1)^2$, and the uniform matrix $\frac{1}{d}\mathbf{1}\mathbf{1}^{\mathsf{T}}$ is an interior point with positive entries, so the affine hull of $\mathrm{DS}(d)$ is this translate. □

Proposition D.3 (Entropy under doubly stochastic maps)**.** *For a probability vector* p *and* $B \in \mathrm{DS}(d)$*, the Shannon entropy satisfies* $H(Bp) \geq H(p)$*, with equality if* B *is a permutation matrix.*

Proof. By Theorem D.1, $B = \sum_\sigma \lambda_\sigma P_\sigma$ is a convex combination of permutation matrices. The entropy is concave and invariant under permutations of its argument, so $H(Bp) = H(\sum_\sigma \lambda_\sigma P_\sigma p) \geq \sum_\sigma \lambda_\sigma H(P_\sigma p) = H(p)$. □

By the theorem of Hardy, Littlewood and Pólya [2], $q = Bp$ for some doubly stochastic B if and only if q is majorized by p, and the entropy is Schur-concave, so Proposition D.3 is the statement that mixing moves a distribution down the majorization order.

Unistochastic matrices

Proposition D.4 (Two necessary conditions)**.** *If $B = T(U)$ for a unitary U, then $B \in \mathrm{DS}(d)$, and no two distinct columns of B have supports intersecting in exactly one row. For $d = 2$ every doubly stochastic matrix is of the form $T(R(\theta))$. For $d = 3$ the doubly stochastic matrix $\frac{1}{2}\left(\begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{smallmatrix}\right)$ is not of the form $T(U)$.*

Proof. Row and column sums of $|U_{ij}|^2$ are 1 by unitarity. If the supports of two distinct columns meet in exactly one row, their inner product is a single nonzero product, contradicting orthogonality. For $d = 2$ a doubly stochastic matrix is $\left(\begin{smallmatrix} p & 1-p \\ 1-p & p \end{smallmatrix}\right)$, and $R(\theta)$ with $\cos^2\theta = p$ realizes it. In the 3×3 example the first two columns have supports $\{1, 3\}$ and $\{1, 2\}$, which meet only in the first row. □

Gauge orbits and phase tori

The torus $\mathbb{T}^d \times \mathbb{T}^d$ acts on $U(d)$ by $(D_1, D_2) \cdot U = D_1 U D_2$ and preserves T. If every entry of U is nonzero, the stabilizer of U consists of the pairs $(cI, \bar{c}I)$ with $|c| = 1$, since $(D_1)_{ii} U_{ij} (D_2)_{jj} = U_{ij}$ forces $(D_1)_{ii}(D_2)_{jj} = 1$ for all i, j. The orbit is therefore a torus of dimension $2d - 1$ contained in a fibre of T. A count of dimensions of domain and target does not give the dimension of every fibre: at $U = I$ the fibre $T^{-1}(I)$ is the set of diagonal unitary matrices, of dimension d. If U has all entries nonzero and is a regular point of T, meaning that the differential of T at U, viewed as a map into the $(d-1)^2$-dimensional affine hull of $\mathrm{DS}(d)$, has rank $(d-1)^2$, then by the regular value theorem the fibre near U is a submanifold of dimension $2d - 1$, and it coincides near U with the gauge orbit, which has the same dimension and lies in it. For $d = 2$ and $0 < p < 1$ the complex unitary fibre of T over the interior matrix is a single orbit of the left-right diagonal phases, as shown in Chapter 15.

For states, the map $[u] \mapsto (|u_i|^2)_i$ from $\mathbb{CP}^{d-1}$ to the probability simplex has, over an interior point, the fibre $\{(\theta_i)\}/\text{common shift} \cong \mathbb{T}^{d-1}$, because a ray with prescribed moduli is determined by its phases up to a common phase. Over a boundary point where k coordinates vanish the fibre is a torus of dimension $d - 1 - k$, and over a vertex it is a point.

Appendix E

Growth and Horizons in the Flat ΛCDM Background

This appendix proves the facts about the background of Part VII that the text uses: the formula for the growth factor, the behaviour of its logarithmic derivative, and the asymptotics of the event horizon. Let $E(a) = \sqrt{\Omega_m a^{-3} + \Omega_\Lambda}$ with $\Omega_m, \Omega_\Lambda > 0$, so that $H = H_0 E$.

Proposition E.1 (Heath's formula)**.** *The growth equation*

$$D'' + \Big(\frac{3}{a} + \frac{E'}{E}\Big) D' - \frac{3}{2}\,\frac{\Omega_m}{a^5 E^2}\, D = 0, \qquad {}' = \frac{d}{da},$$

has the solutions $D_- = E$ and $D_+ = E(a)\int_0^a \frac{da'}{(a'E(a'))^3}$.

Proof. Differentiating $E^2 = \Omega_m a^{-3} + \Omega_\Lambda$ gives $E' = -\frac{3}{2}\Omega_m a^{-4}/E$, and a second differentiation gives $E'' = 6\Omega_m a^{-5}/E - \frac{9}{4}\Omega_m^2 a^{-8}/E^3$. Substituting $D = E$ into the left side, with $E'/E = -\frac{3}{2}\Omega_m a^{-4}/E^2$ and $E' \cdot \big(\frac{3}{a} + \frac{E'}{E}\big)$ expanded, the terms cancel identically; this was also confirmed symbolically. By reduction of order, a second solution is $D = E\int \frac{\exp(-\int p\, da)}{E^2} da$ with $p = \frac{3}{a} + \frac{E'}{E}$, and $\exp(-\int p) = 1/(a^3 E)$, so $D = E\int \frac{da}{a^3 E^3}$. The integral converges at 0 because $(aE)^{-3} \sim \Omega_m^{-3/2} a^{3/2}$. □

Proposition E.2 (Behaviour of the growth rate)**.** *Let $I(a) = \int_0^a (a'E)^{-3} da'$ and $f = d\ln D_+ / d\ln a$. Then*

$$f(a) = -\frac{3}{2}\frac{\Omega_m a^{-3}}{E^2} + \frac{1}{a^2 E^3 I(a)},$$

$f(a) \to 1$ as $a \to 0$, $f(a) \to 0$ as $a \to \infty$, and D_+ is increasing and bounded.

Proof. Differentiating $\ln D_+ = \ln E + \ln I$ and using $I' = (aE)^{-3}$ gives the formula, since $a\,E'/E = -\frac{3}{2}\Omega_m a^{-3}/E^2$. As $a \to 0$, $E^2 \sim \Omega_m a^{-3}$, so the first term tends to $-\frac{3}{2}$; and $I(a) \sim \Omega_m^{-3/2}\frac{2}{5}a^{5/2}$ while $a^2E^3 \sim \Omega_m^{3/2} a^{-5/2}$, so the second term tends to $\frac{5}{2}$, giving $f \to 1$.

As $a \to \infty$, $E \to \sqrt{\Omega_\Lambda}$, so the first term tends to 0, and $I(a) \to I_\infty = \int_0^\infty (a'E)^{-3} da' < \infty$, since $(a'E)^{-3} \le \Omega_\Lambda^{-3/2} a'^{-3}$ at large a'; the second term is $O(a^{-2})$, so $f \to 0$. Finally $D_+ = EI \to \sqrt{\Omega_\Lambda}\, I_\infty < \infty$. For monotonicity it suffices to show $f > 0$, that is $\frac{1}{a^2 E^3 I} > \frac{3}{2}\Omega_m a^{-3} E^{-2}$, which is equivalent to $I(a) < \dfrac{2a}{3\Omega_m E(a)}$. Put $G(a) = \dfrac{2a}{3\Omega_m E} - I(a)$. Then $G(0) = 0$ and, using $E' = -\frac{3}{2}\Omega_m a^{-4}/E$,

$$G'(a) = \frac{2}{3\Omega_m}\Big(\frac{1}{E} - \frac{aE'}{E^2}\Big) - \frac{1}{a^3E^3} = \frac{2}{3\Omega_m E} + \frac{1}{a^3E^3} - \frac{1}{a^3E^3} = \frac{2}{3\Omega_m E} > 0,$$

so $G > 0$ and $f > 0$. Hence D_+ is increasing and bounded. □

Proposition E.3 (The event horizon)**.** $\chi_e(a) = \int_a^\infty \frac{da'}{a'^2 E(a')}$ *is finite and satisfies*

$$\frac{1}{\sqrt{\Omega_m a^{-3} + \Omega_\Lambda}\, a} \;\le\; \chi_e(a) \;\le\; \frac{1}{\sqrt{\Omega_\Lambda}\, a},$$

so $a\chi_e(a)\sqrt{\Omega_\Lambda} \to 1$ *and* $\chi_e \to 0$ *as* $a \to \infty$*. If* $\Omega_\Lambda = 0$ *the integral diverges.*

Proof. For $a' \ge a$ one has $\sqrt{\Omega_\Lambda} \le E(a') \le E(a)$, and $\int_a^\infty a'^{-2} da' = 1/a$. Hence $1/(aE(a)) \le \chi_e(a) \le 1/(a\sqrt{\Omega_\Lambda})$. For $\Omega_\Lambda = 0$ the integrand is $a'^{-1/2}/\sqrt{\Omega_m}$, whose integral over $[a, \infty)$ diverges. □

Corollary E.4 (A turnover exists in the volume-limited model)**.** *Let* $r(a) = 2f(a) - \dfrac{3}{aE(a)\chi_e(a)}$ *be the logarithmic derivative of* $D_+^2\chi_e^3$ *with respect to* $\ln a$ *(up to the constant factor* H_0 *in* aH*). Then* $r(a) \to 2$ *as* $a \to 0$ *and* $r(a) \to -3$ *as* $a \to \infty$*. Hence* r *has at least one zero, and* $s = \frac{4\pi}{3}\bar{j} D_+^2 \chi_e^3$ *has at least one interior maximum.*

Proof. By Proposition E.2, $f \to 1$ at 0 and $f \to 0$ at ∞. By Proposition E.3, $aE(a)\chi_e(a) \ge 1$ and $aE(a)\chi_e(a) \le E(a)/\sqrt{\Omega_\Lambda} \to 1$, so $aE\chi_e \to 1$ as $a \to \infty$ and the second term of r tends to -3. As $a \to 0$, $\chi_e(a) \to \chi_e(0) = \int_0^\infty \frac{da'}{a'^2 E} < \infty$, because the integrand behaves like $a'^{-1/2}/\sqrt{\Omega_m}$ near 0, while $aE \sim \sqrt{\Omega_m}\, a^{-1/2} \to \infty$; hence $aE\chi_e \to \infty$, the second term tends to 0, and $r \to 2$. The intermediate value theorem gives a zero of r. If r changes sign from positive to negative at a zero, the function has a maximum there by Theorem 45.3; the existence of a sign change follows from the signs of the limits at the two ends, by applying the theorem on a compact interval whose end values have opposite signs. □

Corollary E.4 proves the existence of a turnover for every flat background with $\Omega_m, \Omega_\Lambda > 0$. The location $a_* \approx 0.489$ for $\Omega_m = 0.3$ and the uniqueness of the zero are numerical (checks D1 to D3 of Chapter 55).

Appendix F

The Experiment Script

The following script reproduces every check of Chapter 55. It prints one line per check with the identifier, the category, the claim, the measured value, the tolerance and a verdict, and ends with `ALL PASS` when every check passes. It was run with Python 3, numpy, scipy and sympy.

```
#!/usr/bin/env python3
"""Reproducible checks for the monograph.  Dependencies: numpy, scipy, sympy.

Each check returns (id, category, claim, value, tolerance, passed).
Categories:
  A  exact finite enumeration
  B  symbolic verification
  C  numerical corroboration of a proved result
  D  numerical exploration of an unproved claim
"""

import itertools
import sys
import numpy as np

RESULTS = []


def record(cid, cat, claim, value, tol, ok):
    RESULTS.append((cid, cat, claim, value, tol, bool(ok)))


def T(U):
    return np.abs(U) ** 2


H = np.array([[1, 1], [1, -1]]) / np.sqrt(2)
Zg = np.diag([1.0, -1.0])
J2 = np.full((2, 2), 0.5)
```

```
# ---------------------------------------------------------------- A
def a1_lifts():
    th = [np.pi / 4 + k * np.pi / 2 for k in range(4)]
    R = lambda t: np.array([[np.cos(t), -np.sin(t)], [np.sin(t), np.cos(t)]])
    P = np.diag([1, -1])
    L = [R(t) for t in th] + [R(t) @ P for t in th]
    ok_lifts = all(np.allclose(np.abs(l) ** 2, 0.5, atol=1e-14) for l in L)
    n = 0
    for a, b, c in itertools.product(L, repeat=3):
        M = a @ b @ c
        if np.allclose(M, np.eye(2), atol=1e-12) or np.allclose(
            M, -np.eye(2), atol=1e-12
        ):
            n += 1
    record(
        "A1",
        "A",
        "8 lifts of J in O(2); triples with product +-I",
        n,
        "exact",
        ok_lifts and n == 0,
    )


def a2_rule184():
    def step(x):
        n = len(x)
        return [x[i - 1] * (1 - x[i]) + x[i] * x[(i + 1) % n] for i in range(n)]

    ok = step([1, 1, 0, 0]) == [1, 0, 1, 0] and step([1, 0, 1, 0]) == [0, 1, 0, 1]
    ok = ok and step([0, 1, 0, 1]) == [1, 0, 1, 0]
    record(
        "A2",
        "A",
        "Rule 184 ring N=4: 1100->1010, 0101->1010 (update not injective), 1010->0101",
        1,
        "exact",
        ok,
    )


def a3_golden():
    def cnt(n):
        return sum(
            1
            for w in itertools.product([0, 1], repeat=n)
            if all(not (w[i] and w[i + 1]) for i in range(n - 1))
        )

    got = [cnt(n) for n in range(1, 9)]
    record(
        "A3",
```

```
        "A",
        "hard-core word counts n=1..8",
        got,
        "exact",
        got == [2, 3, 5, 8, 13, 21, 34, 55],
    )


def a4_cycle():
    q, n = 0.8, 4
    M = np.zeros((n, n))
    for x in range(n):
        M[(x + 1) % n, x] = q
        M[(x - 1) % n, x] = 1 - q
    pi = np.ones(n) / n
    P = lambda p: pi[p[0]] * np.prod([M[b, a] for a, b in zip(p[:-1], p[1:])])
    sig1 = (2 * q - 1) * np.log(q / (1 - q))
    L = 4
    tot, ift = 0.0, 0.0
    cf, cr = {}, {}
    for p in itertools.product(range(n), repeat=L + 1):
        pp, rr = P(p), P(p[::-1])
        if pp > 0:
            tot += pp * np.log(pp / rr)
            ift += rr
        c = tuple(x % 2 for x in p)
        cf[c] = cf.get(c, 0) + pp
        cr[c] = cr.get(c, 0) + rr
    sc = sum(v * np.log(v / cr[c]) for c, v in cf.items() if v > 0)
    record(
        "A4a",
        "A",
        "fine path entropy production Sigma_4 - 4 Sigma_1",
        tot - L * sig1,
        1e-12,
        abs(tot - L * sig1) < 1e-12,
    )
    record(
        "A4b",
        "A",
        "integral fluctuation relation sum of R over supp P",
        ift,
        1e-12,
        abs(ift - 1) < 1e-12,
    )
    record(
        "A4c",
        "A",
        "parity-coarse entropy production Sigma_4^c",
        sc,
        1e-12,
        abs(sc) < 1e-12,
    )
```

```
def a7_cycle6():
    q, n, L = 0.8, 6, 4
    M = np.zeros((n, n))
    for x in range(n):
        M[(x + 1) % n, x] = q
        M[(x - 1) % n, x] = 1 - q
    pi = np.ones(n) / n
    P = lambda p: pi[p[0]] * np.prod([M[b, a] for a, b in zip(p[:-1], p[1:])])
    sig1 = (2 * q - 1) * np.log(q / (1 - q))
    cf, cr = {}, {}
    for p in itertools.product(range(n), repeat=L + 1):
        pp, rr = P(p), P(p[::-1])
        c = tuple(x % 3 for x in p)
        cf[c] = cf.get(c, 0) + pp
        cr[c] = cr.get(c, 0) + rr
    sc = sum(v * np.log(v / cr[c]) for c, v in cf.items() if v > 0)
    record(
        "A7",
        "A",
        "Z6 cycle observed mod 3: Sigma_4^c - 4 Sigma_1 (lossy but arrow kept)",
        sc - L * sig1,
        1e-12,
        abs(sc - L * sig1) < 1e-12,
    )


def a5_hadamard():
    ok1 = np.allclose(T(H @ H), np.eye(2)) and np.allclose(T(H) @ T(H), J2)
    ok2 = np.allclose(T(H @ np.eye(2) @ H), np.eye(2)) and np.allclose(
        T(H @ Zg @ H), [[0, 1], [1, 0]]
    )
    U2 = np.array([[1, 1j], [1j, 1]]) / np.sqrt(2)
    ok3 = np.allclose(T(U2 @ H), T(U2) @ T(H)) and np.allclose(T(U2 @ H), J2)
    w = np.exp(2j * np.pi / 3)
    F = np.array([[w ** (j * k) for k in range(3)] for j in range(3)]) / np.sqrt(3)
    ok4 = np.allclose(
        T(F @ F), np.array([[1, 0, 0], [0, 0, 1], [0, 1, 0]])
    ) and np.allclose(T(F), 1 / 3)
    record(
        "A5",
        "A",
        "Hadamard, Z-counterexample, weakly consistent pair, 3x3 Fourier",
        1,
        "1e-12",
        ok1 and ok2 and ok3 and ok4,
    )


def a6_nonunistochastic():
    B = np.array([[1, 1, 0], [0, 1, 1], [1, 0, 1]]) / 2
    ds = np.allclose(B.sum(0), 1) and np.allclose(B.sum(1), 1)
```

```
    inter = np.sum((B[:, 0] > 0) & (B[:, 1] > 0))
    record(
        "A6",
        "A",
        "doubly stochastic; columns 1,2 share exactly one row",
        inter,
        "exact",
        ds and inter == 1,
    )


# ---------------------------------------------------------------- B
def b1_energy_flux():
    import sympy as sp

    x = sp.symbols("x")
    Ph, v, S = (sp.Function(n)(x) for n in ("Phi", "v", "S"))
    a, b = sp.symbols("a b")
    eps = sp.Function("eps")
    E = eps(a, b)
    sub = lambda f: f.subs({a: Ph, b: S}).doit()
    eP, th, e0 = sub(sp.diff(E, a)), sub(sp.diff(E, b)), sub(E)
    p = Ph * eP + S * th - e0
    m = Ph * v
    Phit, mt, St = -sp.diff(Ph * v, x), -sp.diff(m * v + p, x), -sp.diff(S * v, x)
    et = v * mt - (v**2 / 2) * Phit + eP * Phit + th * St
    rhs = -sp.diff((m**2 / (2 * Ph) + e0 + p) * v, x)
    res = sp.simplify(sp.expand(et - rhs))
    record(
        "B1",
        "B",
        "energy flux identity for arbitrary eps(Phi,S): residual",
        str(res),
        "symbolic",
        res == 0,
    )


def b2_madelung():
    import sympy as sp

    hb, m0 = sp.symbols("hbar m0", positive=True)
    Rs = sp.Symbol("R", positive=True)
    Rt, Rx, Rxx, tht, thx, thxx, Vs = [
        sp.Symbol(n, real=True) for n in "Rt Rx Rxx tht thx thxx V".split()
    ]
    # (i hbar d_t + hbar^2/2m d_x^2 - V) psi with psi = R exp(i theta), divided by exp(i theta)
    kin = Rxx + 2 * sp.I * Rx * thx + sp.I * Rs * thxx - Rs * thx**2
    res = sp.expand(sp.I * hb * (Rt + sp.I * Rs * tht) + hb**2 / (2 * m0) * kin - Vs * Rs)
    re, im = sp.re(res), sp.im(res)
    # continuity: d_t(R^2) + d_x(R^2 (hbar/m) theta_x) = 2 R im / hbar
    cont = 2 * Rs * Rt + (hb / m0) * (2 * Rs * Rx * thx + Rs**2 * thxx)
    ok_c = sp.simplify(sp.expand(2 * Rs * im / hb - cont)) == 0
```

```
    Q = -(hb**2) / (2 * m0) * Rxx / Rs
    hj = hb * tht + hb**2 / (2 * m0) * thx**2 + Vs + Q
    ok_h = sp.simplify(sp.expand(re / Rs + hj)) == 0
    record(
        "B2",
        "B",
        "Madelung: imaginary part = continuity; real part = Hamilton-Jacobi with Q",
        1,
        "symbolic",
        ok_c and ok_h,
    )


# ------------------------------------------------------------- C
def c_quantum():
    rng = np.random.default_rng(3)
    # K1 = T(U^2) - T(U)^2 from Liouville blocks, random 3x3 unitary
    A = rng.normal(size=(3, 3)) + 1j * rng.normal(size=(3, 3))
    U, _ = np.linalg.qr(A)
    S = np.kron(U.conj(), U)
    dg = [0, 4, 8]
    of = [i for i in range(9) if i not in dg]
    PP, PQ, QP = S[np.ix_(dg, dg)], S[np.ix_(dg, of)], S[np.ix_(of, dg)]
    e1 = np.abs(PQ @ QP - (T(U @ U) - T(U) @ T(U))).max()
    e2 = np.abs(PP - T(U)).max()
    record(
        "C1",
        "C",
        "K1 = T(U^2)-T(U)^2 and M_PP = T(U), random 3x3 unitary",
        max(e1, e2),
        1e-12,
        max(e1, e2) < 1e-12,
    )
    # Hadamard kernels
    SH = np.kron(H.conj(), H)
    PQh, QPh, QQh = (
        SH[np.ix_([0, 3], [1, 2])],
        SH[np.ix_([1, 2], [0, 3])],
        SH[np.ix_([1, 2], [1, 2])],
    )
    ks = [
        np.abs(PQh @ np.linalg.matrix_power(QQh, m - 1) @ QPh).max() for m in range(2, 7)
    ]
    k1 = np.abs(PQh @ QPh - (np.eye(2) - J2)).max()
    record(
        "C2",
        "C",
        "Hadamard kernel K1 = I-J, K_m = 0 for m=2..6",
        max(max(ks), k1),
        1e-12,
        max(max(ks), k1) < 1e-12,
    )
```

```
# Bargmann invariant for qubits
def ket(n):
    th = np.arccos(n[2])
    ph = np.arctan2(n[1], n[0])
    return np.array([np.cos(th / 2), np.exp(1j * ph) * np.sin(th / 2)])

ns = [v / np.linalg.norm(v) for v in rng.normal(size=(3, 3))]
ps = [ket(n) for n in ns]
D = np.vdot(ps[0], ps[1]) * np.vdot(ps[1], ps[2]) * np.vdot(ps[2], ps[0])
f = 0.25 * (
    1
    + ns[0] @ ns[1]
    + ns[1] @ ns[2]
    + ns[2] @ ns[0]
    + 1j * ns[0] @ np.cross(ns[1], ns[2])
)
record(
    "C3",
    "C",
    "Bargmann invariant = Tr(rho1 rho2 rho3) formula",
    abs(D - f),
    1e-12,
    abs(D - f) < 1e-12,
)
# qubit interference
t1, t2 = 0.7, 1.9
a, b = t1 / 2, t2 / 2
cl = np.cos(a) ** 2 * np.cos(b) ** 2 + np.sin(a) ** 2 * np.sin(b) ** 2
d = cl - np.cos((t1 + t2) / 2) ** 2 - 0.5 * np.sin(t1) * np.sin(t2)
record(
    "C4",
    "C",
    "qubit interference term -1/2 sin t1 sin t2",
    abs(d),
    1e-12,
    abs(d) < 1e-12,
)
# Berry phase of two-level loop by discrete product of overlaps
th = 1.1
N = 20000
ph = np.linspace(0, 2 * np.pi, N + 1)
psi = np.array([[np.cos(th / 2), np.exp(1j * p) * np.sin(th / 2)] for p in ph])
prod = np.prod([np.vdot(psi[k], psi[k + 1]) for k in range(N)])
gam = -np.angle(prod)
target = -np.pi * (1 - np.cos(th))
record(
    "C5",
    "C",
    "geometric phase -pi(1-cos theta), discrete product, N=20000",
    abs(gam - target),
    1e-6,
    abs(gam - target) < 1e-6,
)
```

```
    # complex MUB triangle: Z=I, X=H, Y=(1/sqrt2)[[1,1],[i,-i]]
    Y = np.array([[1, 1], [1j, -1j]]) / np.sqrt(2)
    ok = all(
        np.allclose(T(A.conj().T @ B), J2)
        for A, B in [(np.eye(2), H), (np.eye(2), Y), (H, Y)]
    )
    D3 = (1 / np.sqrt(2)) * ((1 + 1j) / 2) * (1 / np.sqrt(2))
    record(
        "C6",
        "C",
        "complex qubit triangle Z,X,Y has all transitions J; Bargmann (1+i)/4",
        abs(D3 - (1 + 1j) / 4),
        1e-14,
        ok and abs(D3 - (1 + 1j) / 4) < 1e-14,
    )


def c_fibre_d2():
    rng = np.random.default_rng(11)
    worst = 0.0
    for _ in range(200):
        A = rng.normal(size=(2, 2)) + 1j * rng.normal(size=(2, 2))
        U, _ = np.linalg.qr(A)
        p = abs(U[0, 0]) ** 2
        c, sn = np.sqrt(p), np.sqrt(1 - p)
        R = np.array([[c, -sn], [sn, c]])
        ph = np.angle(U) - np.angle(R + 0j)
        # phases of D1 R D2 are alpha_i + beta_j; solve for (alpha, beta) from three entries
        a1, a2, b1, b2 = 0.0, ph[1, 0] - ph[0, 0], ph[0, 0], ph[0, 1]
        V = np.diag(np.exp(1j * np.array([a1, a2]))) @ R @ np.diag(np.exp(1j * np.array([b1, b2])))
        worst = max(worst, np.abs(V - U).max())
    record(
        "C18",
        "C",
        "d=2: every unitary with moduli sqrt(p), sqrt(1-p) equals D1 R(theta) D2 (200 samples)",
        worst,
        1e-12,
        worst < 1e-12,
    )


def hankel_and_markov():
    p = 0.8

    def P(w):
        n1 = sum(w)
        n0 = len(w) - n1
        return 0.5 * p**n1 * (1 - p) ** n0 + 0.5 * (1 - p) ** n1 * p**n0

    words = [w for n in range(5) for w in itertools.product([0, 1], repeat=n)]
    Hk = np.array([[P(u + v) for v in words] for u in words])
    r = np.linalg.matrix_rank(Hk, tol=1e-10)
    record(
```

```
        "C7",
        "C",
        "Hankel rank of unknown-bias coin (words up to length 4)",
        r,
        "1e-10",
        r == 2,
    )
    C = -(2 / 3) * np.log(2 / 3) - (1 / 3) * np.log(1 / 3)
    h = (2 / 3) * np.log(2)
    record(
        "C8",
        "C",
        "even process: statistical complexity, entropy rate",
        (round(C, 4), round(h, 4)),
        "exact",
        abs(C - 0.6365) < 5e-5 and abs(h - 0.4621) < 5e-5,
    )


# fluid model (ideal gas) ------------------------------------------------
R_GAS, CV = 1.0, 2.5
GAM = (R_GAS + CV) / CV


def fluid_sim():
    N = 128
    L = 2 * np.pi
    x = np.arange(N) * L / N
    k = np.fft.fftfreq(N, L / N) * 2 * np.pi
    D = lambda f: np.real(np.fft.ifft(1j * k * np.fft.fft(f)))
    mu, kap = 0.02, 0.02

    def prim(rho, m, e):
        v = m / rho
        eps = e - 0.5 * m * v
        T_ = eps / (rho * CV)
        return v, T_, R_GAS * rho * T_

    def rhs(z):
        rho, m, e = z
        v, T_, p = prim(rho, m, e)
        vx = D(v)
        return np.array(
            [
                -D(m),
                -D(m * v + p) + D(mu * vx),
                -D((e + p) * v) + D(mu * v * vx + kap * D(T_)),
            ]
        )

    def ent(z):
        rho, m, e = z
        v, T_, p = prim(rho, m, e)
```

```
        return np.sum(rho * (CV * np.log(T_) - R_GAS * np.log(rho))) * L / N

    rho0 = 1 + 0.2 * np.sin(x)
    v0 = 0.3 * np.sin(x + 0.5)
    T0 = 1 + 0.1 * np.cos(2 * x)
    z = np.array([rho0, rho0 * v0, rho0 * CV * T0 + 0.5 * rho0 * v0**2])
    dt = 2e-4
    E0 = z[2].sum() * L / N
    M0 = z[0].sum() * L / N
    P0 = z[1].sum() * L / N
    S = [ent(z)]
    for n in range(1, 10001):
        k1 = rhs(z)
        k2 = rhs(z + dt / 2 * k1)
        k3 = rhs(z + dt / 2 * k2)
        k4 = rhs(z + dt * k3)
        z = z + dt / 6 * (k1 + 2 * k2 + 2 * k3 + k4)
        if n % 1000 == 0:
            S.append(ent(z))
    drift = max(
        abs(z[2].sum() * L / N - E0),
        abs(z[0].sum() * L / N - M0),
        abs(z[1].sum() * L / N - P0),
    )
    record(
        "C9",
        "C",
        "fluid: max drift of mass, momentum, energy over t in [0,2]",
        drift,
        1e-12,
        drift < 1e-12,
    )
    record(
        "C10",
        "C",
        "fluid: entropy increases at all 10 samples",
        float(np.diff(S).min()),
        "> 0",
        np.all(np.diff(S) > 0),
    )


def fluid_linear():
    # ideal gas eps = Phi*cv*T with T = exp((S/Phi + R ln Phi)/cv)
    rho0, s0 = 1.3, 2.0
    T0 = np.exp((s0 / rho0 + R_GAS * np.log(rho0)) / CV)
    cp = CV + R_GAS
    # derivatives of T(rho,s): dT/ds = T/(cv rho); dT/drho = T (R/rho - s/rho^2)/cv
    Ts = T0 / (CV * rho0)
    Tr = T0 * (R_GAS / rho0 - s0 / rho0**2) / CV
    # p = R rho T
    pr = R_GAS * T0 + R_GAS * rho0 * Tr
    ps = R_GAS * rho0 * Ts
```

```
mu, kap = 0.05, 0.08

def mat(q, mu=mu, kap=kap):
    I = 1j * q
    return np.array(
        [
            [0, -rho0 * I, 0],
            [-I * pr / rho0, -mu * q * q / rho0, -I * ps / rho0],
            [-(kap / T0) * q * q * Tr, -s0 * I, -(kap / T0) * q * q * Ts],
        ]
    )

c2 = pr + s0 * ps / rho0
record(
    "C11",
    "C",
    "c^2 = p_Phi + (S/Phi) p_S equals gamma R T",
    abs(c2 - GAM * R_GAS * T0),
    1e-12,
    abs(c2 - GAM * R_GAS * T0) < 1e-12,
)
Gam = (mu + kap * (1 / CV - 1 / cp)) / (2 * rho0)
errs = []
for q in (0.01, 0.02, 0.04):
    ev = np.linalg.eigvals(mat(q))
    lam = ev[np.argmax(ev.imag)]
    pred = 1j * np.sqrt(c2) * q - Gam * q * q
    errs.append(abs(lam - pred) / abs(pred.imag))
record(
    "C12",
    "C",
    "sound eigenvalue vs i c q - Gamma q^2, relative error at q=0.01",
    errs[0],
    1e-7,
    errs[0] < 1e-7,
)
record(
    "C12b",
    "C",
    "relative error at q=0.04 over error at q=0.01 (expect about 16, quadratic)",
    errs[2] / errs[0],
    "14 to 18",
    14 < errs[2] / errs[0] < 18,
)
allneg = all(
    np.linalg.eigvals(mat(q)).real.max() < 0 for q in np.linspace(0.05, 5, 100)
)
record(
    "C13",
    "C",
    "all eigenvalues Re<0 for 0.05<=q<=5 (mu,kappa>0)",
    int(allneg),
    "exact",
```

```
        allneg,
    )
    z = np.abs(np.linalg.eigvals(mat(0.7, kap=0.0))).min()
    record(
        "C14",
        "C",
        "kappa=0 leaves an eigenvalue 0 at q=0.7 (undamped entropy mode)",
        z,
        1e-12,
        z < 1e-12,
    )
    A = np.array([[0, rho0, 0], [pr / rho0, 0, ps / rho0], [0, s0, 0]])
    ev = np.sort(np.linalg.eigvals(A).real)
    c = np.sqrt(c2)
    record(
        "C15",
        "C",
        "characteristic speeds -c, 0, +c (v=0)",
        float(np.abs(ev - np.array([-c, 0, c])).max()),
        1e-12,
        np.abs(ev - np.array([-c, 0, c])).max() < 1e-12,
    )


def cosmology():
    from scipy.integrate import quad, solve_ivp
    from scipy.optimize import brentq

    Om, OL = 0.3, 0.7
    E = lambda a: np.sqrt(Om * a**-3 + OL)
    Dn = lambda a: E(a) * quad(lambda x: (x * E(x)) ** -3, 0, a, limit=200)[0]
    dlnE = lambda a: -1.5 * Om * a**-4 / E(a) ** 2
    rhs = lambda a, y: [
        y[1],
        -(3 / a + dlnE(a)) * y[1] + 1.5 * Om / (a**5 * E(a) ** 2) * y[0],
    ]
    sol = solve_ivp(
        rhs, [1e-3, 1.0], [1e-3, 1.0], rtol=1e-11, atol=1e-14, dense_output=True
    )
    ode = sol.sol(0.5)[0] / sol.y[0, -1]
    ratio = Dn(0.5) / Dn(1.0)
    record(
        "C16",
        "C",
        "growth factor: integral formula vs ODE, D(0.5)/D(1)",
        abs(ode - ratio),
        1e-9,
        abs(ode - ratio) < 1e-9,
    )
    chi_e = lambda a: quad(lambda x: 1 / (x * x * E(x)), a, np.inf, limit=200)[0]
    err = abs(chi_e(100.0) * 100 * np.sqrt(OL) - 1)
    record(
        "C17",
```

```
        "C",
        "event horizon a*chi_e*sqrt(OmegaL) -> 1 at a=100",
        err,
        1e-3,
        err < 1e-3,
    )

    def f(a):
        h = 1e-5 * a
        return (np.log(Dn(a + h)) - np.log(Dn(a - h))) / (np.log(a + h) - np.log(a - h))

    r = lambda a: 2 * f(a) - 3 / (a * E(a) * chi_e(a))
    astar = brentq(r, 0.3, 5.0)
    A = np.linspace(0.05, 5, 500)
    val = [2 * np.log(Dn(a)) + 3 * np.log(chi_e(a)) for a in A]
    amax = A[int(np.argmax(val))]
    grid = np.linspace(0.05, 20, 400)
    signs = np.sign([r(a) for a in grid])
    changes = int(np.sum(signs[1:] != signs[:-1]))
    record(
        "D1",
        "D",
        "turnover scale factor a_* (root of 2f - 3/(aH chi_e))",
        round(astar, 4),
        "root-finder xtol",
        abs(astar - 0.4891) < 1e-3,
    )
    record(
        "D2",
        "D",
        "argmax of D^2 chi_e^3 on grid (step 0.01) vs a_*",
        round(amax, 4),
        "grid step 0.01",
        abs(amax - astar) < 0.01,
    )
    record(
        "D3",
        "D",
        "number of sign changes of the rate on 0.05<=a<=20 (400 pts)",
        changes,
        "grid",
        changes == 1,
    )


def main():
    a1_lifts()
    a2_rule184()
    a3_golden()
    a4_cycle()
    a7_cycle6()
    a5_hadamard()
    a6_nonunistochastic()
```

```
    try:
        b1_energy_flux()
        b2_madelung()
    except ImportError:
        record("B", "B", "sympy not installed; symbolic checks skipped", 0, "-", False)
    c_quantum()
    c_fibre_d2()
    hankel_and_markov()
    fluid_sim()
    fluid_linear()
    cosmology()
    width = max(len(r[2]) for r in RESULTS)
    allok = True
    for cid, cat, claim, value, tol, ok in RESULTS:
        print(
            f"{cid:5s} {cat} {claim:{width}s}  value={value}  tol={tol}  {'PASS' if ok else 'FAIL'}"
        )
        allok = allok and ok
    print("ALL PASS" if allok else "SOME FAILED")
    return 0 if allok else 1


if __name__ == "__main__":
    sys.exit(main())
```

Bibliography

[1] G. Birkhoff, Tres Observaciones Sobre el Algebra Lineal, *Universidad Nacional de Tucumán, Revista, Serie A* **5** (1946), 147–151.

[2] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd ed., Cambridge University Press, 1952.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, 2006.

[4] G. E. Crooks, Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences, *Physical Review E* **60** (1999), 2721–2726.

[5] C. J. Ellison, J. R. Mahoney, and J. P. Crutchfield, Prediction, Retrodiction, and the Amount of Information Stored in the Present, *Journal of Statistical Physics* **136** (2009), 1005–1034, doi:10.1007/s10955-009-9808-z.

[6] J. P. Crutchfield and K. Young, Inferring Statistical Complexity, *Physical Review Letters* **63** (1989), 105–108.

[7] Flyxion, *Quantum Mechanics from Admissible Probabilistic Composition: A Reconstruction via Unistochastic Matrices*, unpublished manuscript, 2026.

[8] M. Fliess, Matrices de Hankel, *Journal de Mathématiques Pures et Appliquées* **53** (1974), 197–222.

[9] W. Israel and J. M. Stewart, Transient Relativistic Thermodynamics and Kinetic Theory, *Annals of Physics* **118** (1979), 341–372.

[10] H. Jaeger, Observable Operator Models for Discrete Stochastic Time Series, *Neural Computation* **12** (2000), 1371–1398.

[11] E. F. Moore, Gedanken-Experiments on Sequential Machines, in *Automata Studies*, Princeton University Press, 1956, 129–153.

[12] S. Wolfram, Statistical Mechanics of Cellular Automata, *Reviews of Modern Physics* **55** (1983), 601–644.

[13] I. Csiszár, Information-Type Measures of Difference of Probability Distributions and Indirect Observations, *Studia Scientiarum Mathematicarum Hungarica* **2** (1967), 299–318.

[14] T. M. Davis and C. H. Lineweaver, Expanding Confusion: Common Misconceptions of Cosmological Horizons and the Superluminal Expansion of the Universe, *Publications of the Astronomical Society of Australia* **21** (2004), 97–109.

[15] D. J. Heath, The Growth of Density Perturbations in Zero Pressure Friedmann–Lemaître Universes, *Monthly Notices of the Royal Astronomical Society* **179** (1977), 351–358.

[16] P. J. E. Peebles, *The Large-Scale Structure of the Universe*, Princeton University Press, 1980.

[17] Flyxion, Affine Quantum Deformation and the Geometry of Awareness: A Unified Variational Framework, unpublished manuscript, December 2025.

[18] Flyxion, Emergence of Quantum Unistochasticity from RSVP Fields, unpublished manuscript, July 2025.

[19] M. Grmela and H. C. Öttinger, Dynamics and Thermodynamics of Complex Fluids. I. Development of a General Formalism, *Physical Review E* **56** (1997), 6620–6632.

[20] W. A. Hiscock and L. Lindblom, Generic Instabilities in First-Order Dissipative Relativistic Fluid Theories, *Physical Review D* **31** (1985), 725–733.

[21] D. D. Holm, J. E. Marsden, T. Ratiu, and A. Weinstein, Nonlinear Stability of Fluid and Plasma Equilibria, *Physics Reports* **123** (1985), 1–116.

[22] C. Kipnis and C. Landim, *Scaling Limits of Interacting Particle Systems*, Springer, 1999.

[23] T. M. Liggett, *Interacting Particle Systems*, Springer, 1985.

[24] E. Madelung, Quantentheorie in Hydrodynamischer Form, *Zeitschrift für Physik* **40** (1927), 322–326.

[25] P. J. Morrison and J. M. Greene, Noncanonical Hamiltonian Density Formulation of Hydrodynamics and Ideal Magnetohydrodynamics, *Physical Review Letters* **45** (1980), 790–794.

[26] H. C. Öttinger and M. Grmela, Dynamics and Thermodynamics of Complex Fluids. II. Illustrations of a General Formalism, *Physical Review E* **56** (1997), 6633–6655.

[27] W. G. Unruh, Experimental Black-Hole Evaporation?, *Physical Review Letters* **46** (1981), 1351–1353.

[28] M. Visser, Acoustic Black Holes: Horizons, Ergospheres and Hawking Radiation, *Classical and Quantum Gravity* **15** (1998), 1767–1791.

[29] R. B. Griffiths, Consistent Histories and the Interpretation of Quantum Mechanics, *Journal of Statistical Physics* **36** (1984), 219–272.

[30] C. J. Isham, Quantum Logic and the Histories Approach to Quantum Theory, *Journal of Mathematical Physics* **35** (1994), 2157–2185.

[31] M. V. Berry, Quantal Phase Factors Accompanying Adiabatic Changes, *Proceedings of the Royal Society of London A* **392** (1984), 45–57.

[32] E. T. Jaynes, Information Theory and Statistical Mechanics, *Physical Review* **106** (1957), 620–630.

[33] R. Landauer, Irreversibility and Heat Generation in the Computing Process, *IBM Journal of Research and Development* **5** (1961), 183–191.

[34] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Van Nostrand, 1960.

[35] H. Mori, Transport, Collective Motion, and Brownian Motion, *Progress of Theoretical Physics* **33** (1965), 423–455.

[36] A. Nerode, Linear Automaton Transformations, *Proceedings of the American Mathematical Society* **9** (1958), 541–544.

[37] J. Schnakenberg, Network Theory of Microscopic and Macroscopic Behavior of Master Equation Systems, *Reviews of Modern Physics* **48** (1976), 571–585.

[38] U. Seifert, Stochastic Thermodynamics, Fluctuation Theorems and Molecular Machines, *Reports on Progress in Physics* **75** (2012), 126001.

[39] C. R. Shalizi and J. P. Crutchfield, Computational Mechanics: Pattern and Prediction, Structure and Simplicity, *Journal of Statistical Physics* **104** (2001), 817–879.

[40] R. Zwanzig, Memory Effects in Irreversible Thermodynamics, *Physical Review* **124** (1961), 983–992.