Discrete Information Geometry of Computational Universe: Relative Entropy, Fisher Structure and Task-Sensitive Distance

Haobo Ma[1] Wenlin Zhang[2]

[1]Independent Researcher

[2]National University of Singapore

November 24, 2025

Abstract

Under axiomatized framework of "computational universe" $U_{\text{comp}} = (X, \mathsf{T}, \mathsf{C}, \mathsf{I})$, complexity geometry characterizes "how much time/cost needed to reach certain configuration". However, complexity geometry alone insufficient to describe "how high quality information these costs exchange for". For this, this paper constructs set of "discrete information geometry" theory matching computational universe in completely discrete setting.

We first introduce observation operator family $\mathcal{O} = \{O_j\}_{j \in J}$, where each O_j maps configuration $x \in X$ to probability distribution $p_x^{(j)}$ on some finite outcome set. Under fixed task or observation scheme, these distributions provide "observable information state" for each configuration x. Based on this, we define task-sensitive relative entropy structure $D_Q(x\|y)$, from which derive family of information distances, e.g., Jensen–Shannon type distance $d_{\text{JS},Q}(x, y)$. These distances locally induce discrete Fisher structure, i.e., near some reference configuration x_0, Hessian of second-order relative entropy $D_Q(x\|x_0)$ gives discrete information metric tensor around x_0.

This paper proves, under natural regularity assumptions, discrete information structure can converge in appropriate limit to Riemannian-type information manifold $(\mathcal{S}_Q, g_Q)$, where g_Q Fisher-type metric; correspondingly, "information geometry on configuration space" can be realized through map $\Phi_Q : X \to \mathcal{S}_Q$, projecting each configuration x to its observable information state. We further discuss volume growth of information balls $B_R^{\text{info}}(x_0)$ and "information dimension", give general inequality between information dimension and complexity dimension, characterizing "information resolution limit achievable under given complexity budget".

Finally, we construct task-sensitive information–complexity joint action functional $\mathcal{A}_Q$, whose local Euler–Lagrange equation gives local description of optimal computation trajectory "maximizing information quality" under finite time budget, providing discrete information geometry foundation for subsequent complete "time–information–complexity variational principle".

Keywords: Discrete information geometry; Relative entropy; Fisher information metric;

Jensen–Shannon distance; Task-sensitive distance; Information dimension; Complexity-information inequality

1 Introduction

In computational universe axiomatic system, universe abstracted as discrete configuration space X, one-step update relation T, single-step cost C and information quality function I, such that any actual computation process corresponds to finite path on configuration graph, complexity distance $d(x, y)$ characterizes minimum cost needed to go from x to y. Previous work already constructed "discrete complexity geometry" based on this, describing problem difficulty and complexity horizon through complexity ball volume and discrete Ricci curvature.

However, complexity geometry concerns "how far walked", not "what seen". To understand geometric structure of "information quality" in computational universe, need to introduce another dimension: observation and task. Specifically, "useful information" of same configuration x depends not only on x itself, but also on how we read it out, what kind of task we care about. Different tasks correspond to different "information geometries", and computation process trajectories on these information geometries are true objects reflecting "how much information we extracted within given time".

Goal of this paper is to establish set of task-related "discrete information geometry" for computational universe in completely discrete background:

- At discrete level, assign each configuration x probability state p_x determined by observation scheme, construct information distances using relative entropy, Jensen–Shannon distance, etc.;
- Locally, through second-order expansion of relative entropy obtain Fisher-type metric, establish discrete information manifold structure;
- Globally, through information ball volume and information dimension characterize "under certain task, complexity of distinguishable states in universe".

More importantly, information geometry and complexity geometry must match: complexity geometry tells us which configurations allowed to move between under resource constraints, information geometry tells us how much information gain these movements bring in "task-relevant state space". Coupling of both will ultimately lead to unified "time–information–complexity action functional".

Main thread structure of this paper as follows. Section 2 introduces observation operators and task-sensitive discrete relative entropy structure. Section 3 constructs discrete information distances and information balls, defines information dimension. Section 4 discusses local Fisher structure and information manifold limit. Section 5 gives information–complexity inequality and task-sensitive joint action functional prototype. Appendix provides detailed proofs of main propositions and theorems.

2 Observation Operators and Task-Sensitive Relative Entropy

This section introduces observation operators and task-sensitive probability structure at configuration layer of computational universe.

2.1 Observation Operator Family and Observable States

In computational universe $U_{\text{comp}} = (X, \mathsf{T}, \mathsf{C}, \mathsf{I})$, configuration $x \in X$ is internal state of entire universe. Observer within certain time window can only access it through finite experiments or readout processes. To characterize this point, introduce observation operator family.

Definition 2.1 (Observation Operator Family)**.** Let $(Y_j)_{j\in J}$ be family of finite outcome sets. An observation operator family is map collection

$$\mathcal{O} = \{O_j : X \to \Delta(Y_j)\}_{j\in J},$$

where $\Delta(Y_j)$ probability simplex on Y_j, and for each $x \in X$, $j \in J$, $O_j(x) = p_x^{(j)}$ is outcome distribution on result set Y_j from one experiment.

Intuitively, O_j describes observational process executable on configuration x, whose output distribution $p_x^{(j)}$ is statistical information observer can "see" on this configuration.

To avoid redundancy, we often denote task or observation scheme as finite subset $Q \subset J$, define "joint observable state" under this task.

Definition 2.2 (Joint Observable State under Task Q)**.** For given finite task set $Q \subset J$, define observable outcome set

$$Y_Q = \prod_{j\in Q} Y_j,$$

define configuration x's joint observable state as joint distribution $p_x^{(Q)}$ on Y_Q. Simplest construction assumes observations independent, in which case

$$p_x^{(Q)}(y) = \prod_{j\in Q} p_x^{(j)}(y_j), \quad y = (y_j)_{j\in Q} \in Y_Q.$$

More generally, can allow known coupling structure between different observations, then $p_x^{(Q)}$ given by task-specific observation model. This paper mainly considers independent case.

2.2 Task-Sensitive Relative Entropy

After fixing task Q, each configuration x mapped to probability distribution $p_x^{(Q)} \in \Delta(Y_Q)$. This allows us to introduce relative entropy for task Q.

Definition 2.3 (Relative Entropy under Task Q)**.** For configurations $x, y \in X$, if for all $y \in Y_Q$ have $p_y^{(Q)}(y) > 0$ implies $p_x^{(Q)}(y) > 0$, define

$$D_Q(x\|y) = \sum_{z\in Y_Q} p_x^{(Q)}(z) \log \frac{p_x^{(Q)}(z)}{p_y^{(Q)}(z)},$$

otherwise define $D_Q(x\|y) = +\infty$.

$D_Q(x\|y)$ is "distinguishability degree" of configurations x and y under task Q: larger means more "information distant" between x and y under this task.

Clearly, $D_Q(x\|y) \geq 0$, and $D_Q(x\|y) = 0$ if and only if $p_x^{(Q)} = p_y^{(Q)}$.

Note D_Q generally not symmetric and doesn't satisfy triangle inequality, thus not metric. To obtain information distance, we will use symmetrized form derived from D_Q.

3 Discrete Information Distances and Information Balls

This section defines family of information distances from task-sensitive relative entropy, constructs information ball structure and information dimension.

3.1 Jensen–Shannon Distance

Most natural symmetrized form is Jensen–Shannon divergence.

Definition 3.1 (Jensen–Shannon Distance under Task Q)**.** Define JS divergence

$$\mathrm{JS}_Q(x,y) = \frac{1}{2} D_Q(x \| m_{xy}) + \frac{1}{2} D_Q(y \| m_{xy}),$$

where $m_{xy} = \frac{1}{2}(p_x^{(Q)} + p_y^{(Q)})$ midpoint distribution. Then

$$d_{\mathrm{JS},Q}(x,y) = \sqrt{\mathrm{JS}_Q(x,y)}$$

defines metric on configuration space (up to equivalence relation $p_x^{(Q)} = p_y^{(Q)}$).

Proposition 3.2 (Metric Properties of JS Distance)**.** *$d_{\mathrm{JS},Q}$ satisfies:*

1. *Symmetry: $d_{\mathrm{JS},Q}(x,y) = d_{\mathrm{JS},Q}(y,x)$;*
2. *Triangle inequality: $d_{\mathrm{JS},Q}(x,z) \leq d_{\mathrm{JS},Q}(x,y) + d_{\mathrm{JS},Q}(y,z)$;*
3. *Positivity: $d_{\mathrm{JS},Q}(x,y) \geq 0$, equals zero iff $p_x^{(Q)} = p_y^{(Q)}$.*

Proof. Symmetry obvious from definition. Triangle inequality follows from Endres-Schindelin (2003) proof. See Appendix A. □

3.2 Information Balls and Information Volume

Given information distance, can define information balls.

Definition 3.3 (Information Ball)**.** For configuration $x_0 \in X$ and radius $R > 0$,

$$B_R^{\mathrm{info}}(x_0) = \{x \in X : d_{\mathrm{JS},Q}(x, x_0) \leq R\}$$

is information ball of radius R centered at x_0 under task Q.

Information ball characterizes set of configurations "informationally close" to x_0 under task Q. Its cardinality $|B_R^{\mathrm{info}}(x_0)|$ measures "how many distinguishable states exist within information distance R from x_0".

3.3 Information Dimension

Definition 3.4 (Information Dimension)**.** If limit

$$\dim_{\mathrm{info}}(x_0) = \lim_{R\to 0} \frac{\log |B_R^{\mathrm{info}}(x_0)|}{\log(1/R)}$$

exists, call it information dimension at x_0 under task Q.

Information dimension measures "how densely information states pack" near x_0. High dimension means many distinguishable states nearby, low dimension means sparse information structure.

4 Local Fisher Structure and Information Manifold Limit

This section constructs local Fisher information metric from second-order expansion of relative entropy, discusses continuous limit of discrete information geometry.

4.1 Discrete Fisher Information Matrix

Consider small perturbations near reference configuration x_0. Assume configuration space has local parameter representation: near x_0 exist parameters $\theta = (\theta^1, \ldots, \theta^n)$ such that configurations uniquely correspond to θ values.

Definition 4.1 (Discrete Fisher Information Matrix)**.** For parameterized configuration family $x(\theta)$ near $x_0 = x(\theta_0)$, define discrete Fisher information matrix at θ_0 as

$$g_{ab}^{(\mathrm{Fisher})}(\theta_0) = \frac{\partial^2}{\partial\theta^a \partial\theta^b} D_Q\big(x(\theta)\|x(\theta_0)\big)\Big|_{\theta=\theta_0}$$

when this quantity well-defined.

4.2 Second-Order Expansion

Proposition 4.2 (Quadratic Approximation of Relative Entropy)**.** *Under smoothness assumptions on $p_{x(\theta)}^{(Q)}$ in θ,*

$$D_Q\big(x(\theta)\|x(\theta_0)\big) = \frac{1}{2}\sum_{a,b} g_{ab}^{(\mathrm{Fisher})}(\theta_0)\delta\theta^a\delta\theta^b + O(|\delta\theta|^3)$$

where $\delta\theta = \theta - \theta_0$.

Proof. Taylor expansion to second order. First-order term vanishes by definition. See Appendix B. □

This shows Fisher matrix defines local Riemannian metric on parameter space, measuring information distance for small perturbations.

4.3 Continuous Limit and Information Manifold

When configuration space has continuous limit (e.g., discretized field configurations converging to continuous fields), discrete information geometry converges to continuous information manifold.

Theorem 4.3 (Convergence to Information Manifold). *Under appropriate regularity conditions on observation operators and refinement sequence of discrete configurations, discrete Fisher metrics converge to continuous Fisher information metric g_Q on continuous configuration manifold $\mathcal{S}_Q$, forming Riemannian manifold $(\mathcal{S}_Q, g_Q)$.*

Proof. Uses standard techniques from information geometry. See Amari-Nagaoka (2000) and Appendix C. □

5 Information–Complexity Inequality and Joint Action Functional

This section establishes quantitative relationship between information dimension and complexity dimension, constructs joint action functional coupling information and complexity.

5.1 Information–Complexity Trade-off

Theorem 5.1 (Information–Complexity Inequality). *For any computation path $\gamma : [0, T] \to X$ of complexity cost $C(\gamma)$, information gain along path bounded by*

$$\Delta I_Q(\gamma) \leq f\big(C(\gamma), \dim_{\text{comp}}, \dim_{\text{info}}\big)$$

where f function of complexity cost, complexity dimension $\dim_{\text{comp}}$ *and information dimension* $\dim_{\text{info}}$.

Proof. Combines complexity ball volume bounds from discrete complexity geometry with information ball bounds. See Appendix D. □

This theorem characterizes fundamental limit: given finite complexity budget, maximum achievable information resolution bounded by interplay of complexity and information dimensions.

5.2 Task-Sensitive Joint Action Functional

To unify complexity cost and information gain, define joint action.

Definition 5.2 (Information–Complexity Joint Action). For computation path γ in time interval $[0, T]$, define

$$\mathcal{A}_Q[\gamma] = \int_0^T \big[\alpha\, \mathsf{C}(\dot{\gamma}(t)) - \beta\, \mathsf{I}_Q(\gamma(t))\big]\, \mathrm{d}t$$

where $\mathsf{C}(\dot{\gamma})$ instantaneous complexity cost rate, $\mathsf{I}_Q(\gamma)$ instantaneous information quality under task Q, $\alpha, \beta > 0$ trade-off weights.

Minimizing $\mathcal{A}_Q$ yields trajectories balancing complexity cost and information gain.

5.3 Euler–Lagrange Equation

Proposition 5.3 (Optimal Trajectory Condition). *Critical points of $\mathcal{A}_Q$ satisfy*

$$\alpha\,\nabla_{\dot\gamma}\mathsf{C}(\dot\gamma) = \beta\,\nabla\mathsf{I}_Q(\gamma)$$

where ∇ appropriate derivatives on configuration space.

This gives local characterization of optimal computation trajectories maximizing information quality under complexity constraints.

6 Discussion and Outlook

This paper constructed discrete information geometry framework for computational universe, complementing complexity geometry from previous work. Key achievements:

1. Defined task-sensitive relative entropy and information distances;
2. Established discrete Fisher structure and information manifold limit;
3. Introduced information dimension and proved information–complexity inequalities;
4. Constructed joint action functional coupling information and complexity.

Future directions:

- Extend to quantum information geometry for quantum computational universe;
- Develop numerical methods for computing information metrics;
- Apply to concrete problems in machine learning and optimization;
- Complete unified time–information–complexity variational principle.

This framework provides foundation for understanding not just "how computation happens" but "what information computation extracts".

A Proof of Triangle Inequality for JS Distance

This appendix proves Proposition **??**.

A.1 Endres-Schindelin Proof

The key result (Endres-Schindelin, 2003): $\sqrt{\mathrm{JS}}$ satisfies triangle inequality.

For three distributions p, q, r, define midpoints m_{pq}, m_{qr}, m_{pr}. Through careful convexity arguments and data processing inequality, show

$$\sqrt{\mathrm{JS}(p,r)} \le \sqrt{\mathrm{JS}(p,q)} + \sqrt{\mathrm{JS}(q,r)}$$

Applied to our setting with $p = p_x^{(Q)}$, etc., gives triangle inequality.

B Second-Order Expansion of Relative Entropy

This appendix proves Proposition ??.

B.1 Taylor Expansion

Write

$$D_Q(x(\theta)\|x(\theta_0)) = \sum_y p^{(Q)}_{x(\theta)}(y) \log \frac{p^{(Q)}_{x(\theta)}(y)}{p^{(Q)}_{x(\theta_0)}(y)}$$

Expand both numerator and denominator to second order in $\delta\theta$:

$$p^{(Q)}_{x(\theta)}(y) = p_0(y) + \sum_a \partial_a p_0(y)\delta\theta^a + \frac{1}{2}\sum_{ab} \partial_a\partial_b p_0(y)\delta\theta^a\delta\theta^b + O(|\delta\theta|^3)$$

where $p_0 = p^{(Q)}_{x(\theta_0)}$.

After substitution and simplification using normalization conditions, first-order terms cancel, second-order terms give Fisher matrix.

C Convergence to Continuous Information Manifold

This appendix sketches proof of Theorem ??.

C.1 Refinement Sequence

Consider sequence of discrete configuration spaces X_n with lattice spacing $a_n \to 0$. Assume observation operators $\mathcal{O}_n$ converge appropriately to continuous observation functionals.

Discrete Fisher matrices $g^{(n)}_{ab}$ form approximations to continuous Fisher metric g_{ab}. Under regularity (Sobolev estimates on probability densities), $g^{(n)}_{ab} \to g_{ab}$ in suitable topology.

Details involve careful measure-theoretic arguments, see Amari-Nagaoka (2000) for standard proofs in classical information geometry setting.

D Proof of Information–Complexity Inequality

This appendix proves Theorem ??.

D.1 Volume Comparison

Key idea: complexity ball of radius R_{comp} contains at most certain number of information-distinguishable states, bounded by ratio of volumes in complexity vs. information geometries.

Specifically, if complexity ball $B^{\text{comp}}_{R_{\text{comp}}}(x_0)$ has volume $V_{\text{comp}} \sim R_{\text{comp}}^{\dim_{\text{comp}}}$, and typical information separation scale is ϵ_{info}, then number of distinguishable states

$$N_{\mathrm{dist}} \lesssim \frac{V_{\mathrm{comp}}}{\epsilon_{\mathrm{info}}^{\mathrm{dim}_{\mathrm{info}}}}$$

Information gain along path of complexity cost C bounded by $\log N_{\mathrm{dist}}$ with appropriate $R_{\mathrm{comp}} \sim C$.

Detailed calculation shows inequality of Theorem **??**.