Optimal Entropic Dimensionality: A Continuous Variational Principle for Geometric Equilibrium

Takeo Imaizumi*

Independent Researcher, Kyoto City, Japan

Abstract

Systems that encode information into high-dimensional degrees of freedom confront a geometric trade-off between dilution (expansion) and crowding (compression). We introduce *Optimal Entropic Dimensionality* (OED), a variational framework in which effective dimensionality is treated as a continuous state variable and an equilibrium is selected by minimizing geometric size. Under isotropic L_p geometry, this yields the scaling law $N^* = pH$, which balances geometric overhead against entropic density, with p determined by the underlying norm.

To ground this law in a canonical setting, we analyze the Euclidean case ($p = 2$) via information geometry. We model distinguishable configurations as small geodesic balls with respect to the Fisher information metric under Jeffreys measure. Using Gamma-function and Stirling-type volume asymptotics, we obtain a logarithmic expansion cost and derive $N^* = 2H$, up to constants independent of N.

We then extend the variational structure to anisotropic systems. Correlations consume a structural log-volume H_{struct}, leaving a free payload $H_{\text{free}} = H - H_{\text{struct}}$. This leads to the structure-corrected optimum $N^* \approx p(H - H_{\text{struct}})$, supported by both a geometric bulk hyper-rectangle model and an independent whitening-based entropy transformation. OED thus furnishes a geometric coordinate system (N, H, p) for complexity, organizing regimes from sparse ($p \approx 1$), through a log-base calibration point ($p = 1/\ln 2$), to robustness-oriented limits ($p \to \infty$) beyond the Euclidean benchmark.

Keywords: Effective dimension, Model selection, Scaling laws, Information geometry, Shannon entropy, Variational principle, L_p spaces, Fisher information metric, Complexity, Sparse representation

*Corresponding author

Email address: takeoimaizumi@song.ocn.ne.jp (Takeo Imaizumi)

Contents

1. Introduction: Dimensionality as a Continuous State Variable

1.1. The Fundamental Trade-off: Expansion vs. Compression

In the representation of complex systems, whether in statistical physics, model selection [1], or information theory [2], one encounters a fundamental geometric trade-off [3]. To encode a given informational payload H, a system must determine an appropriate effective number of degrees of freedom N, interpreted here as an effective geometric dimension. This choice is governed by two antagonistic tendencies:

- *Cost of Expansion:* Increasing the dimensionality N provides more "room" for the payload, reducing density. However, in high dimensions this expansion entails a *structural overhead*—a geometric penalty arising from the intrinsic vastness of the configuration space, which requires increasing informational resources to navigate.

- *Cost of Compression:* Conversely, reducing dimensionality forces the fixed payload H into a tighter configuration space. This creates an *entropic crowding* pressure that diverges as the system is squeezed into too few degrees of freedom, creating a resistance against over-compression.

Statistical and physical laws often manifest as equilibrium points between these two costs. While standard derivations explain these optimal dimensionalities using domain-specific constraints, OED investigates whether they share a common underlying variational structure governed by the geometry of the state space.

Remark 1.1 (Intuitive Analogy: The Tape Measure). To conceptualize dimensionality as a continuous state variable, consider a steel tape measure with a fixed length of 5 meters (representing the invariant payload H).

- High footprint (extended; $N = 1$): When fully extended, the tape acts as a one-dimensional object. While strictly linear, its representative geometric size is 5 meters in length, making it spatially unwieldy to transport.

- Lower footprint (coiled; $N = 2$): To minimize this geometric footprint for storage (i.e., "pocketability"), we coil the tape into a spiral. In this state, it effectively behaves as a two-dimensional disk, reducing its characteristic scale from 5 m in length to approx. 5 cm in diameter.

- Continuous transition ($1 < N < 2$): The act of winding the tape shifts the effective dimension continuously with the extended–coiled fraction.

From this perspective, the integer dimensions 1 and 2 are not rigid boundaries, but merely stable fixed points (extended vs. coiled) in a continuous variational landscape optimized for utility. OED formalizes this "winding" process as the minimization of geometric cost.

1.2. OED: A Geometric Coordinate System

This paper proposes *Optimal Entropic Dimensionality (OED)*, a variational framework that treats dimensionality not as a fixed background container, but as a *continuous state variable* optimized to balance these geometric costs. We formulate the problem as a geometric inverse problem: *Given a payload H, what is the optimal dimensionality N^* that minimizes the representative geometric size of the system?*

Central to this framework is the scaling law derived from L_p-norm geometry (formally derived in Section 2):

$$N^* = pH. \tag{1}$$

Within this framework, the Euclidean equilibrium ($N^* = 2H$) is identified as the specific balance point dictated by spaces with local quadratic (Gaussian) structure.

We emphasize that OED serves as a *geometric coordinate system* (N, H, p) that maps these phenomena onto a single variational landscape. In this view, p should be read as a continuous control parameter that indexes the effective geometry used to measure expansion, rather than as a fixed background constant. This perspective reveals that distinct equilibria can be interpreted as structural isomorphisms—manifestations of the same expansion–compression duality expressed through different metric choices, made precise at the level of the variational form in Section 3.

Notation. Throughout, H denotes the dimensionless payload (a log-volume constraint), $H(\cdot)$ denotes differential entropy when explicitly required, N denotes effective dimensionality, and $p \geq 1$ is the parameter of the underlying L_p metric used to measure geometric expansion. Unless stated otherwise we express H in natural units (nats); conversions to bits are obtained by an information-unit (log-base) gauge change $H_{\text{bits}} \equiv H/\ln 2$. These symbols remain fixed for the remainder of the paper.

1.3. Scope and Structure of this Paper

The remainder of this paper is organized as follows. Section 2 develops the general OED theory based on L_p geometry, establishing the cost trade-off and deriving the fundamental scaling law $N^* = pH$. Section 3 focuses on the Euclidean benchmark ($p = 2$), where the geometric law establishes a rigorous isomorphism with Gaussian thermodynamics ($N \leftrightarrow \sigma^2$). Section 4 bridges the theory to realistic, structured systems, deriving the structure-corrected law $N^* \approx p(H - H_{\text{struct}})$ via geometric and algebraic (whitening-based) approaches to quantifying effective degrees of freedom. Section 5 extends the framework to non-Euclidean $p \neq 2$ regimes, including sparsity ($p \approx 1$), an information-unit (log-base) gauge calibration point ($p = 1/\ln 2$), and robustness ($p \to \infty$). Section 6 discusses the outlook from a broader perspective: the structural overview of the OED framework, the model scope and limitations, and open questions across scales. Section 7 concludes the paper. Technical derivations and proofs are collected in the appendices.

2. Theoretical Foundation: The Geometry of Expansion and Compression

We complement domain-specific formulations with a derivation based on minimal geometric postulates. Throughout, we treat $N \in \mathbb{R}_{>0}$ as a continuous state variable in the OED variational landscape. Here "state variable" is used in a thermodynamic analogy, as a coarse-grained macroscopic coordinate selected by the variational principle rather than a literal microscopic count.

2.1. Minimal Postulates: Informational Payload and Geometric Scaling

Our framework rests on two primary postulates that define the basic "physics" of the geometric space.

Principle 1: Informational Invariance (Payload Conservation). We represent the information to be encoded by a single dimensionless scalar H, the *geometric payload*, defined as a log-volume (log distinguishability) in the chosen representation. Concretely, H may be read as the logarithm of the effective number of distinguishable configurations, and is treated as fixed prior to selecting its dimensional container. Standard notions such as Boltzmann or Shannon entropy correspond to particular choices of reference scale and logarithm base (an information-unit / log-base gauge) [2].

We impose the payload as a fixed log-volume budget: the configuration space must provide enough effective volume to realize the prescribed distinguishability, independently of the chosen dimensionality N:

$$\ln V(N) = H \quad \Longrightarrow \quad V(N) = e^H \tag{2}$$

Throughout the main text, V denotes a coordinate volume under a chosen gauge (measured relative to a fixed reference volume so that $\ln V$ is well-defined); the *representative size* of that volume is then evaluated using the L_p metric below.

Principle 2: Geometric Scaling (Homogeneity). We assume the system resides in a normed linear space with an L_p norm ($p \geq 1$). In an isotropic coordinate chart with characteristic scale a, the volume scales as $V = a^N$. Throughout, we take the *representative size* to be the L_p-typical radius of the volume-equivalent L_p ball, so that $L_p \propto N^{1/p} a$ up to N-subleading unit-ball constants [4]. Appendix A derives the same leading $\frac{1}{p} \ln N$ expansion term by writing the radius as $r = (V/V_{p,N})^{1/N}$ and using Gamma/Stirling asymptotics for $V_{p,N}$. Substituting the payload constraint $a = V^{1/N} = e^{H/N}$ yields

$$L_p(N) \propto N^{1/p} e^{H/N}. \tag{3}$$

2.2. The Variational Principle: Balancing Expansion and Compression

Principle 3: Dimensional Equilibrium (Variational Principle). We posit that, among admissible representations within the chosen coordinates and gauge, the system favors the most compact one, modeled as minimizing a representative geometric scale. The system therefore selects a dimensionality N^* that minimizes $L_p(N)$ (or equivalently its logarithm, the geometric cost $\mathcal{C}_p$):

$$\mathcal{C}_p(N, H) \equiv \ln L_p(N) = \underbrace{\tfrac{1}{p}\ln N}_{\text{Cost of Expansion}} + \underbrace{\tfrac{H}{N}}_{\text{Cost of Compression}} \tag{4}$$

Conceptually, this formulation shifts dimensionality from a rigid structural parameter to a *fluid capacity constraint*. By analogy with gas expansion, information spreads to occupy degrees of freedom; the OED equilibrium corresponds to the point where the geometric overhead of expansion is balanced by the crowding tendency.

Remark 2.1 (Microscopic Origin via Fisher Information). The logarithmic expansion cost is not an ad hoc postulate. As detailed in Appendix A, in the Riemannian ($p = 2$) case it follows from the Gamma-function scaling of Jeffreys volumes in high dimension, a manifestation of the concentration-of-measure phenomenon. For general L_p spaces ($p \neq 2$), this logarithmic form is adopted as a geometric generalization at the level of the variational form, motivated by the leading Gamma/Stirling scaling of L_p unit-ball volumes in high dimension.

This cost function encapsulates the competition between two opposing geometric effects:

- *Cost of Expansion* ($\frac{1}{p}\ln N$)*:* The penalty for increasing the number of degrees of freedom, arising from the $N^{1/p}$ scaling of the L_p norm. This term reflects concentration of measure: the logarithmic geometric work required to access the expanding volume shell in which representative states reside.
- *Cost of Compression* ($\frac{H}{N}$)*:* The penalty for forcing a fixed payload H into too few dimensions. This term reflects the exponential effort in required resolution needed to maintain a fixed informational payload within a restricted dimensional container.

For visualization, it is sometimes useful to consider the *informational compactness* $\rho(N)$, defined as the reciprocal of the representative size:

$$\rho(N) \equiv \frac{1}{L_p(N)} \propto \exp\bigl(-\mathcal{C}_p(N)\bigr) \tag{5}$$

Maximizing this compactness is equivalent to minimizing the cost, and identifies the dimensionality at which the system is most efficiently represented.

2.3. The General Law: $N^* = pH$

Minimizing Eq. (4) with respect to N (treating N as a continuous state variable) gives

$$\frac{d\mathcal{C}_p}{dN} = \frac{1}{pN} - \frac{H}{N^2} = 0$$

Physically (within this variational model), this equilibrium occurs when the marginal geometric work of expansion ($1/pN$) precisely balances the marginal reduction of crowding pressure (H/N^2). Solving for N yields the general scaling law:

$$N^* = pH \tag{6}$$

This relation states that the optimal dimensionality is proportional to the informational payload, with the proportionality constant determined entirely by the geometry (p) of the state space. Mathematical proofs of uniqueness and robustness with respect to geometric gauge choices are provided in Appendix B.

3. The Euclidean Benchmark ($p = 2$): Isomorphism with Gaussian Statistics

Having established the general scaling law $N^* = pH$ for isotropic systems, the Euclidean case ($p = 2$) yields the *Euclidean law*

$$N^* = 2H. \tag{7}$$

Specializing the OED cost to $p = 2$ gives the two-term variational form

$$\mathcal{C}_2(N, H) = \underbrace{\tfrac{1}{2}\ln N}_{\text{Expansion Cost}} + \underbrace{\tfrac{H}{N}}_{\text{Compression Cost}}. \tag{8}$$

Increasing N raises the logarithmic expansion term, while decreasing N raises the inverse compression term. The equilibrium $N^* = 2H$ is the unique point at which these opposing geometric pressures balance.

Figure 1 illustrates this landscape. As N grows, structural overhead increases while crowding pressure relaxes, and the representative size $L(N)$ attains its minimum precisely at $N^* = 2H$. The Euclidean case therefore supplies a clean geometric benchmark in which the trade-off between expansion and compression is expressed in its simplest form.

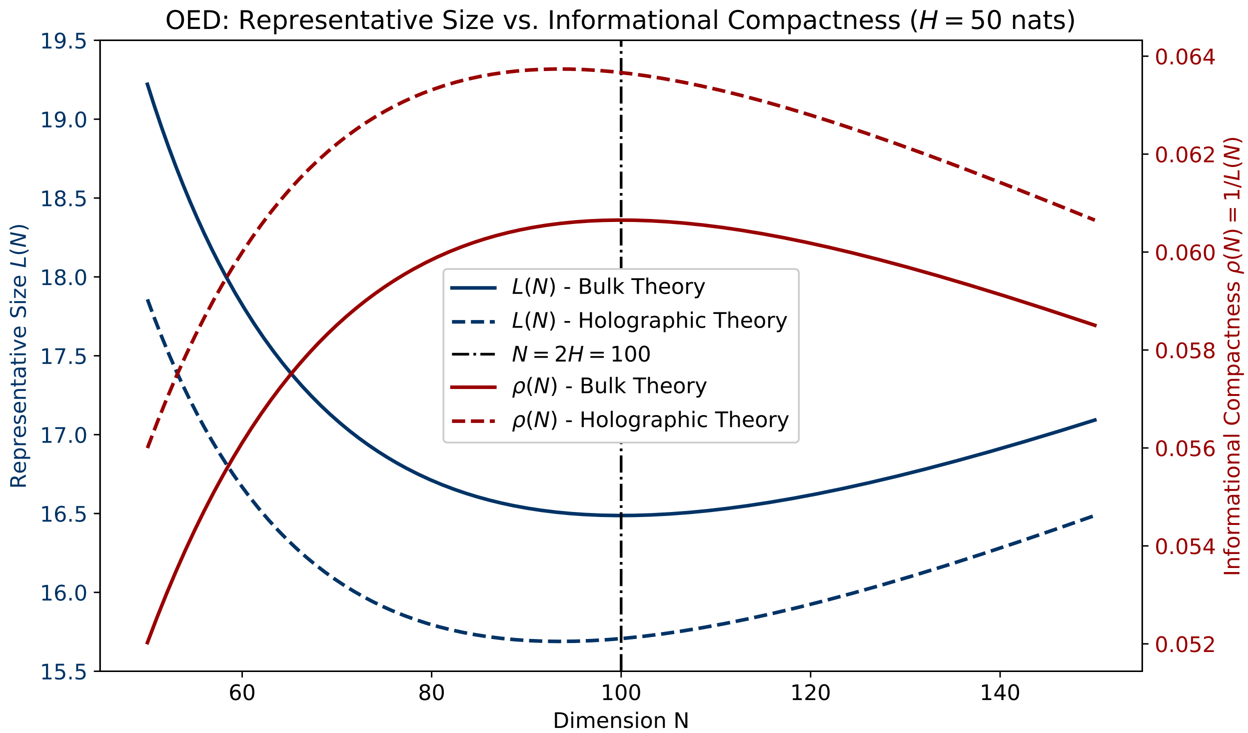

Figure 1: **Dimensional equilibrium in Euclidean geometry ($p = 2$).** Numerical optimization for a payload of $H = 50$. The representative size $L(N)$ reaches its minimum at $N^* = 100$, and the informational compactness $\rho(N) \propto 1/L(N)$ reaches its maximum at the same point.

We now show that this geometric equilibrium is variationally identical to the equilibrium of an isotropic Gaussian model. Consider a Gaussian distribution with variance σ^2 and residual sum of squares

$$S = \sum_{i=1}^{n} (x_i - \mu)^2, \quad E \equiv \frac{S}{2n}.$$

Normalizing per sample, the Gaussian free energy has the same two-term form:

$$F(\sigma^2) \propto \underbrace{\tfrac{1}{2}\ln \sigma^2}_{\text{Expansion Cost}} + \underbrace{\tfrac{E}{\sigma^2}}_{\text{Compression Cost}}. \tag{9}$$

Larger σ^2 expands the distribution and raises the logarithmic term; smaller σ^2 crowds the mass and raises the inverse term.

Differentiation gives the Gaussian optimum,

$$\sigma^{2*} = 2E,$$

which balances marginal expansion and compression costs in exact analogy with the OED optimum $N^* = 2H$.

The duality is now explicit. Both systems minimize a functional of the identical form

$$\text{log-expansion} \ + \ \text{inverse-compression}.$$

Under the identification

$$\sigma^2 \longleftrightarrow N, \qquad E \longleftrightarrow H,$$

the Gaussian free energy $F(\sigma^2)$ becomes the Euclidean OED cost $\mathcal{C}_2(N, H)$ up to affine rescaling, and the optima coincide:

$$\sigma^{2*} = 2E \quad \Longleftrightarrow \quad N^* = 2H.$$

Table 1 summarizes this dictionary. The Euclidean benchmark thus provides a geometric reinterpretation of Gaussianity: variance corresponds to dimensionality, quadratic energy to payload, and the familiar factor "two" emerges from the same quadratic equilibrium shared by both formulations.

Table 1: **Geometric isomorphism between Gaussian thermodynamics and OED geometry.**

Concept	Gaussian System	OED System
Diffusion-like scale	Variance σ^2	Dimensionality N
Fixed payload/invariant	Quadratic energy E	Geometric payload H
Variational cost	$\frac{1}{2}\ln\sigma^2 + \frac{E}{\sigma^2}$	$\frac{1}{2}\ln N + \frac{H}{N}$
Equilibrium condition	$\sigma^{2*} = 2E$	$N^* = 2H$

Within the OED coordinate system, the Euclidean case provides a uniquely transparent benchmark in which every structural element has a direct probabilistic analogue.

4. Structured Systems: Anisotropy and Dimensional Reduction

The general law $N^* = pH$ established in Section 2 assumes isotropy, i.e., that all dimensions are equivalent. Real-world data and systems, however, exhibit anisotropy: correlations, preferred directions, and heterogeneous variances are the norm rather than the exception. In this section we extend the OED framework to such structured settings by introducing a structural log-volume H_{struct} that quantifies how correlations consume degrees of freedom, and by showing that, under suitable conditions, the effective optimum follows the structure-corrected law $N^* \approx p\big(H - H_{\text{struct}}\big)$. This claim is supported by two independent constructions: a geometric bulk model and an algebraic derivation based on whitening and entropy transformations.

Throughout this section, we interpret N as an *effective (truncation) dimension*: the number of dominant principal axes retained in an ordered spectrum (e.g., $\sigma_1 \geq \sigma_2 \geq \cdots$). Accordingly, quantities such as $\sum_{i=1}^{N}$ are understood as being evaluated over the top-N principal directions. When we treat N as continuous, sums such as $\sum_{i=1}^{N}$ are understood in the usual truncated-spectrum sense (with the continuum approximation justified separately).

4.1. Geometric Derivation of the Structure-Corrected Law

To account for anisotropy, we generalize the geometric model from an isotropic hypercube to an *axis-aligned hyper-rectangle*. This subsection derives (i) the exact rectangle cost function under the bulk constraint and

(ii) its first-order minimizer in the smooth-spectrum, high-dimensional regime. Full algebraic details of the rectangle model are provided in Appendix C.

Let the edge lengths be proportional to the principal-axis standard deviations, $a_i = k\,\sigma_i$ with $k > 0$. We fix a reference scale σ_0 and interpret $\ln \sigma_i$ as $\ln(\sigma_i/\sigma_0)$, so that H_{struct} is dimensionless and the payload is defined up to an additive gauge. Under the bulk mapping, the volume constraint reads

$$V = \prod_{i=1}^{N} a_i = k^N \prod_{i=1}^{N} \sigma_i = e^H$$

We again measure representative size using an L_p norm,

$$L_p \;\propto\; \left(\sum_{i=1}^{N} a_i^p \right)^{1/p} = k \left(\sum_{i=1}^{N} \sigma_i^p \right)^{1/p}$$

Defining $\mathcal{C}_{\text{rect}} \coloneqq \ln L_p$ and eliminating k via the volume constraint yields the exact rectangle cost

$$\mathcal{C}_{\text{rect}}(N) = \frac{1}{p} \ln \left(\sum_{i=1}^{N} \sigma_i^p \right) + \frac{1}{N} \left(H - \sum_{i=1}^{N} \ln \sigma_i \right) \tag{10}$$

The structural term $H_{\text{struct}}(N) \coloneqq \sum_{i=1}^{N} \ln \sigma_i$ arises naturally (in the principal-axis gauge) from the determinant of the metric tensor (see Appendix A) and represents the log-volume consumed by fixed anisotropic correlations (evaluated in the principal-axis basis). The remaining quantity

$$H_{\text{free}}(N) \coloneqq H - H_{\text{struct}}(N)$$

may be interpreted as the "free" payload available for isotropic representation after accounting for structure.

Minimizing $\mathcal{C}_{\text{rect}}(N)$ with respect to N in full generality depends on the detailed spectral profile $\{\sigma_i\}$. However, in the high-dimensional regime with a sufficiently smooth spectrum (so that local spectral variations contribute only subdominantly at leading order), the leading-order stationarity balance matches the isotropic case, with H replaced by H_{free}, while spectrum-dependent corrections enter through subleading derivative terms. Assuming $H_{\text{free}} > 0$ for an interior optimum, this yields the first-order structure-corrected scaling law

$$N^* \approx p\big(H - H_{\text{struct}}\big) \tag{11}$$

where H_{struct} is evaluated at the effective dimension (i.e., $H_{\text{struct}} = H_{\text{struct}}(N^*)$). When $H_{\text{free}} \leq 0$, the interior optimum disappears in this first-order approximation and the optimum is pushed to the boundary (minimal effective dimension).

4.2. Algebraic Derivation via Data Whitening

The structure-corrected law admits an alternative algebraic derivation based on the transformation properties of differential entropy. This serves as an independent derivation of the geometric framework established in Section 4.1. In this subsection, $H(\cdot)$ denotes differential entropy. We interpret $H(X_N)$ as a log-volume–type payload defined relative to a fixed reference measure, so that it plays the same variational role as the abstract payload H introduced in Section 2.

Consider an anisotropic random vector X in a high-dimensional ambient space $\mathbb{R}^D$ characterized by a covariance matrix $\mathbf{\Sigma}$ with sorted eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D$. We seek to represent the system in an effective subspace of dimension $N \leq D$. Let $X_N \in \mathbb{R}^N$ be the projection of X onto the subspace spanned by the top-N principal components, with associated covariance $\Sigma_N = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$. Let

$$Y_N = \Sigma_N^{-1/2}(X_N - \mu_N)$$

be the *whitened representation* in this effective subspace, so that $\mathrm{Cov}(Y_N) = I_N$ (all directions have unit variance and no cross-covariances).

For an absolutely continuous random vector and an invertible linear transformation $Y = AX$, the differential entropy satisfies

$$H(Y) = H(X) + \ln|\det A|$$

Applying this change-of-variables relation within the N-dimensional active subspace (after projection) with $A = \Sigma_N^{-1/2}$ gives

$$H(Y_N) = H(X_N) + \ln\big|\det(\Sigma_N^{-1/2})\big| = H(X_N) - \frac{1}{2}\ln|\Sigma_N|$$

Using the relation $\ln|\Sigma_N| = \sum_{i=1}^{N}\ln(\sigma_i^2) = 2\sum_{i=1}^{N}\ln\sigma_i$, we have

$$H(Y_N) = H(X_N) - \sum_{i=1}^{N}\ln\sigma_i$$

Motivated by the geometric term introduced in Section 4.1, and using the same fixed reference scale, we define the structural entropy of the N-dimensional representation as

$$H_{\text{struct}}(N) \coloneqq \sum_{i=1}^{N}\ln\sigma_i$$

so that

$$H(Y_N) = H(X_N) - H_{\text{struct}}(N)$$

Since Y_N is isotropic, the isotropic OED law applies directly to Y_N, yielding

$$N^* = p\,H(Y_N) = p\big(H(X_N) - H_{\text{struct}}(N)\big)$$

Thus, at leading order, the effective entropic dimensionality is governed by the *free* information

$$H_{\text{free}}(N) \coloneqq H(X_N) - H_{\text{struct}}(N)$$

rather than by the raw entropy $H(X_N)$ alone.

Remark 4.1 (Note on Entropy Definitions). It is important to distinguish the definitions of H used here. In Section 2, H is a fixed "payload" parameter independent of dimensionality. In this algebraic derivation, $H(\cdot)$ refers to differential entropy, which inherently depends on coordinate scaling. The correspondence $H_{\text{struct}} = \sum\ln\sigma_i$ implicitly assumes a reference scale $\sigma_0 = 1$ (a dimensionless formulation). Despite these definitional nuances, the algebraic result coincides with the geometric derivation in the smooth-spectrum limit (see Appendix C), reinforcing that the structure-corrected form is not an artifact of a single derivation route.

5. Beyond the Euclidean/Hilbertian Regime ($p \neq 2$): A Phenomenological Hypothesis

The general law $N^* = pH$ suggests that the ubiquitous coefficient "two" is the signature of the Hilbertian (Euclidean) regime within OED. Among all L_p spaces, only the case $p = 2$ admits an inner product, which induces rotational symmetry of the norm and the quadratic structure that underlies Gaussianity. In OED, p is therefore promoted from a technical norm index to a geometric control parameter that labels regimes.

For $p \neq 2$, however, this promotion comes with a clear limitation: a microscopic derivation comparable in spirit to the Fisher–Jeffreys construction would likely require a genuinely non-Riemannian (Finsler-type) foundation, which lies beyond the scope of this paper. Accordingly, we treat the extension away from $p = 2$ as a *phenomenological hypothesis*: the two-term variational form of Section 2 is retained as a working model, and p is used as an organizing index for how a system's notion of "size" departs from Euclidean geometry.

From an applied viewpoint, p may be treated as an emergent descriptor of the effective geometry rather than a predefined constant. In principle, it can be estimated by fitting the observed expansion–compression trade-off to the OED cost form, i.e., by estimating the effective shape of distinguishable sets induced by the chosen coordinates.

5.1. Sparsity and Efficiency ($p = 1$)*: Geometric Adaptation in Biology*

A leading-order extension of the OED scaling to the L_1 regime (Manhattan geometry) gives $N^* = H$. While Fisher geometry itself is Riemannian ($p = 2$), this extension can be used as a working hypothesis: compared with the Euclidean optimum ($N^* = 2H$), an L_1-leaning regime is "frugal" in the sense that fewer degrees of freedom suffice to encode the same payload. This perspective is consistent with sparsity as an organizing principle in biological systems, where efficiency pressures favor sparse, modular representations in an appropriate basis (i.e., after a representation change that exposes sparsity) [5].

Quantitatively, this shift implies a substantial reduction in overhead: the system achieves "Laplacian" (i.e., L_1) efficiency, requiring only half the degrees of freedom at the OED optimum ($N^* = H$) compared with the Gaussian baseline ($N^* = 2H$) for the same payload definition. As a working hypothesis, this geometric frugality predicts a bias toward L_1-leaning architectures in resource-constrained environments.

5.2. Information-Unit Gauge and the Digital Calibration Point $\big(p = 1/\ln 2\big)$

The OED law $N^* = pH$ is stated in terms of the dimensionless payload H expressed using the natural logarithm (nats). Since many information-theoretic discussions use bits, it is useful to introduce an *information-unit (log-base) gauge* that re-expresses the same payload in base-2 units. Define

$$H_{\text{bits}} \;\coloneqq\; \frac{H}{\ln 2},$$

so that $H = (\ln 2)\, H_{\text{bits}}$ is merely a change of units. Substituting this into the general law isolates a dimensionless "redundancy factor" (degrees of freedom per bit) at equilibrium:

$$N^* = pH = (p \ln 2)\, H_{\text{bits}} \;\eqqcolon\; \alpha\, H_{\text{bits}}, \qquad \alpha \coloneqq p \ln 2.$$

The distinguished calibration $\alpha = 1$ occurs at

$$p = \frac{1}{\ln 2} \approx 1.44, \qquad \text{so that} \qquad N^* = H_{\text{bits}}.$$

We refer to this value as the *digital calibration point*: within the adopted variational form, the OED equilibrium assigns one effective degree of freedom per bit, yielding a one-to-one correspondence between optimal dimensionality and binary information content. This remark introduces no new microscopic assumptions beyond the retained $N^* = pH$ scaling; it simply highlights a particularly transparent normalization of that law.

5.3. Robustness and Control ($p \to \infty$)

At the opposite extreme are worst-case design objectives, where the relevant size is governed by the maximum component (the L_∞ norm). In this regime, the scaling $N^* = pH$ formally diverges as $p \to \infty$; in practice one may expect large but finite effective p when worst-case constraints dominate. This is consistent with the intuition that worst-case robustness often requires substantial redundancy: degrees of freedom are expanded to dilute the impact of coordinate-localized failures.

Taken together, sparsity ($p \approx 1$), the digital calibration point ($p = 1/\ln 2$), Gaussianity ($p = 2$), and worst-case robustness ($p \to \infty$) can be placed on a single Banach-geometric axis parameterized by p as phenomenological regimes of the same variational form. From this perspective, the question is not only "what is the effective dimension?", but also "what metric is the system implicitly choosing, and why?"

6. Discussion and Outlook

6.1. Structural Overview of the OED Framework

We summarize the conceptual structure of OED as developed throughout the preceding sections. The aim is not to supplant domain-specific models, but to provide a shared geometric coordinate system for comparison.

1. Dimensionality as a state variable; entropy as geometry. OED treats effective dimensionality N as a continuous state variable, rather than a fixed background parameter. The informational payload H is interpreted geometrically as a log-volume constraint—i.e., as the size (up to a logarithm) of the distinguishable region to be represented in a chosen coordinate space.

2. Expansion–compression trade-off. The OED cost $\mathcal{C}_p(N, H) = \frac{1}{p} \ln N + \frac{H}{N}$ decomposes into an expansion overhead and a compression (crowding) penalty. The equilibrium N^* is the point where these opposing tendencies balance, yielding the scaling law $N^* = pH$ in the isotropic case.

3. Beyond the Euclidean regime ($p \neq 2$). Expressing the cost in terms of L_p geometry makes the metric index p explicit. This highlights the special role of the Euclidean case $p = 2$, while also providing a systematic way to organize departures from Euclidean behavior within the same variational form.

4. The (N, H, p) coordinate system. The quantities N, H, and p form a compact coordinate system for the OED landscape: N encodes effective capacity (or diffusion-like spread), H encodes payload (log-volume), and p specifies the metric structure used to measure expansion.

5. Relation to classical information measures. OED does not modify the standard definitions or derivations of Shannon entropy or KL divergence. Its role is interpretive: it provides a geometric lens for relating familiar information measures to representational capacity and the choice of metric structure.

6.2. Model Scope and Limitations

The OED law $N^* = pH$ provides a compact description of optimal entropic dimensionality. Its use is subject to the following scope conditions and modeling limitations.

1. Geometric foundation and the L_p extension. The Euclidean benchmark ($p = 2$) is rigorously grounded in Fisher–Jeffreys (Riemannian) geometry via high-dimensional volume asymptotics. For $p \neq 2$, we retain the same variational form as a *phenomenological* extension: it organizes departures from Euclidean behavior, but a first-principles non-Riemannian (Finsler-type) derivation remains open.

2. Continuum approximation and low-payload regimes. OED treats N as a continuous state variable. This is justified when the payload is large ($H \gg 1$), where integer effects are negligible (see Appendix D). In low-payload or strongly quantized settings, discreteness (and, in quantum settings, non-commutativity) can invalidate the continuum picture.

3. Spectral smoothness in structured systems. The structure-corrected form $N^* \approx p(H - H_{\text{struct}})$ is a leading-order result under a sufficiently smooth eigenvalue spectrum, with subleading corrections analyzed in Appendix C.2. Strong spikes, low rank, or heavy tails can amplify higher-order terms, so the bulk formula should be read as a baseline estimator rather than an exact optimizer in such regimes.

4. Equilibrium principle vs. dynamics. OED is deliberately formulated as a static variational principle: it identifies the equilibrium N^* that minimizes $\mathcal{C}_p(N, H)$, but it does not derive the time evolution of $N(t)$. A dynamical completion would require additional modeling assumptions (e.g., dissipation, payload injection, or coupling), and we highlight this as an open direction.

6.3. Open Questions Across Scales

From interpretation to inquiry. A scientific theory derives value not only from prediction but also from its capacity to generate new questions and to reframe existing ones. Just as OED originated from a simple question about geometric stability, the framework invites reconsideration of fundamental problems in terms of expansion–compression equilibria. The items below summarize open questions and working hypotheses expressed in the (N, H, p) coordinate system.

1. Can artificial intelligence be compressed further? Large language models (LLMs) exhibit power-law scaling relationships that currently favor very large parameter counts [6]. Within OED, this trend can be interpreted as a *transient expansion regime* required to accommodate a rapidly growing effective payload. If, however, the free payload of knowledge (H_{free}) is ultimately finite, OED suggests a long-run equilibrium in which solutions become effectively confined to a substantially lower-dimensional manifold, with $N^* \approx 2H_{\text{free}}$ in the Euclidean case. Under this interpretation, one may anticipate a qualitative shift from massive over-parameterization toward more compact, structure-adapted geometric representations (speculatively reminiscent of biological efficiency). In OED terms, this corresponds to a transition from a high-N expansion regime toward an approximately Euclidean ($p \approx 2$) equilibrium near $N^* \simeq 2H_{\text{free}}$.

2. How did life acquire the capacity to manipulate dimensionality? The cell nucleus provides a physical instantiation of the "tape-measure" metaphor (Section 1) via a dynamic trade-off: transcription maps three-dimensional spatial organization onto a one-dimensional temporal sequence. *(1) Storage as spatial index ($N \to 3$):* folded chromatin can reduce geometric footprint and act as a topological search index, in which three-dimensional proximity influences regulatory access. *(2) Readout as temporal sequence ($N \to 1$):* transcription requires local dimensional reduction, locally unwinding the polymer so that relatively static spatial memory is converted into an ordered one-dimensional stream. In the OED picture, biological organization plausibly navigates the (N, H, p) landscape by alternating between high-N storage and low-N processing at approximately fixed payload H, treating dimensionality as a manipulable resource.

3. Why is space three-dimensional—and not ten-dimensional? Within OED, spatial dimensionality may be viewed as an equilibrium between free and structural entropy. If the universe entered a Euclidean-like regime ($p = 2$) and settled at $N = 3$, this would suggest that the accessible free entropy was extremely small, with the overwhelming majority absorbed into geometric structure—a perspective broadly consistent with Penrose's emphasis on the extraordinary smoothness of the early universe and its very low gravitational entropy [7]. In string-theoretic models with 10 or 11 spacetime dimensions, a related reading is that the additional correlational degrees of freedom need not "vanish"; rather, they may be compactified into extra dimensions whose curvature and topology act as a reservoir of structural entropy. On this view, only three spatial directions remain sufficiently unsaturated to unfold dynamically as macroscopic space, while the compact dimensions act as an entropic sink that helps stabilize the overall balance. In OED terms, one may say that H_{free} is expressed in the three extended directions, whereas H_{struct} is sequestered in the compact ones, with $H_{\text{free}} = H - H_{\text{struct}}$.

4. Can the OED geometry hint at a new class of dynamical laws? Since OED is posed at equilibrium, a natural next step is to ask what *minimal* dynamics could realize that equilibrium over time. One concrete route is a gradient-flow ansatz in which $\mathcal{C}_p$ acts as a Lyapunov functional and $N(t)$ relaxes according to

$$\dot{N}(t) \propto -\partial_N \mathcal{C}_p(N, H),$$

   optionally supplemented by terms encoding domain-specific effects such as payload injection ($\dot{H}$), dissipation, or constraints on admissible N (e.g., discreteness or feasible ranges). This viewpoint reframes the dynamical problem as an inverse-design question: which microscopic mechanisms (or control objectives) would coarse-grain to a flow whose stable fixed point is N^*? Beyond providing a heuristic evolution law, such a program would also clarify when the equilibrium picture is expected to hold and when it should fail.

7. Conclusion

This paper introduces the framework of *Optimal Entropic Dimensionality* (OED), casting effective degrees of freedom as the solution of a continuous variational principle. By treating dimensionality not as a fixed background parameter but as a continuous state variable, OED exposes the optimization structure underlying efficient information representation and, in the Euclidean case ($p = 2$), grounds it in the local geometry of probability space through the Fisher information metric.

A central result is a rigorous variational isomorphism with Gaussian thermodynamics, under which dimensionality plays a diffusion-like role that is mathematically equivalent to variance via an explicit identification. Within this correspondence, the ubiquitous coefficient "two" (i.e., $N^* = 2H$) is interpreted not as an *ad hoc* constant but as the characteristic signature of equilibrium in Euclidean (L_2) geometry. For structured systems, we further show that anisotropic correlations allocate part of the informational content to a structural log-volume H_{struct}, thereby reducing the free entropy available for representation. This yields the structure-corrected scaling law $N^* \approx p\,(H - H_{\text{struct}})$, which provides a geometric basis for dimensional reduction.

Extending the construction to a general (N, H, p) coordinate system organizes a range of scaling behaviors by the metric index p: sparsity-oriented efficiency ($p \approx 1$), a digital calibration point under an information-unit (log-base) gauge ($p = 1/\ln 2$), and robustness-oriented redundancy under worst-case constraints ($p \to \infty$). In this view, complexity resides not only in the payload H but also in the geometry used to carry it: the Euclidean case ($p = 2$) yields a stable equilibrium, whereas moving toward $p \approx 1$ favors more economical representations. The factor "two" thus emerges as an explicit marker of equilibrium in the familiar Euclidean setting.

Declarations

Funding: No funding was received to support this research.
Conflict of Interest: The author declares that there are no conflicts of interest.

References

[1] H. Akaike, A new look at the statistical model identification, IEEE Transactions on Automatic Control 19 (6) (1974) 716–723. doi:10.1109/TAC.1974.1100705.

[2] C. E. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (3–4) (1948) 379–423; 623–656, published in two parts (No. 3 and No. 4). doi:10.1002/j.1538-7305.1948.tb01338.x.

[3] S.-i. Amari, H. Nagaoka, Methods of Information Geometry, Vol. 191 of Translations of Mathematical Monographs, American Mathematical Society, Providence, RI, 2000.

[4] D. L. Donoho, High-dimensional data analysis: The curses and blessings of dimensionality, AMS Math Challenges Lecture, lecture notes / slides (commonly circulated). (2000).

[5] B. A. Olshausen, D. J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–609. doi:10.1038/381607a0.

[6] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, D. Amodei, Scaling laws for neural language models, arXiv preprint arXiv:2001.08361 (2020).
URL https://arxiv.org/abs/2001.08361

[7] R. Penrose, Singularities and time-asymmetry, in: S. W. Hawking, W. Israel (Eds.), General Relativity: An Einstein Centenary Survey, Cambridge University Press, Cambridge, 1979, pp. 581–638.

Appendix A. Microscopic Origin (p=2) and Coordinate-Space L_p Extension

We justify the OED expansion term $\frac{1}{p} \ln N$ by: (i) modeling the distinguishable region as a small geodesic ball in Fisher (Jeffreys) geometry, (ii) extracting the leading N-dependence of unit-ball volumes via Gamma/Stirling asymptotics, and (iii) showing how anisotropy induces the structural subtraction $H_{\text{struct}} = \sum_{i=1}^{N} \ln \sigma_i$.

Appendix A.1. Minimal definitions: Fisher information and Jeffreys volume

Let $\boldsymbol{\theta} \in \mathbb{R}^N$ parametrize a model $p(x \mid \boldsymbol{\theta})$. The Fisher information matrix $G(\boldsymbol{\theta}) = [g_{ij}(\boldsymbol{\theta})]$ defines a Riemannian metric:

$$g_{ij}(\boldsymbol{\theta}) = \mathbb{E}\left[\frac{\partial \ln p(X \mid \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln p(X \mid \boldsymbol{\theta})}{\partial \theta_j}\right] \tag{A.1}$$

The associated coordinate-invariant volume element (Jeffreys measure) is [3]

$$dV_J(\boldsymbol{\theta}) = \sqrt{\det G(\boldsymbol{\theta})}\, d\boldsymbol{\theta} \tag{A.2}$$

We encode distinguishability by the Jeffreys log-volume $H_J := \ln \mathrm{Vol}_J(\Omega)$ of a region Ω in parameter space.

Appendix A.2. Small-ball reduction and unit-ball constants

We relate the radius r of a distinguishable region to its log-volume. The only N-dependence we need comes from the unit-ball constant (Gamma-function scaling).

Geodesic ball approximation (Riemannian, $p = 2$). For sufficiently small radius, we approximate Ω by a geodesic ball B_r. In normal coordinates around $\boldsymbol{\theta}_0$, the metric is locally Euclidean, hence

$$\mathrm{Vol}_J(B_r) \;=\; V_{2,N}\, r^N\, (1 + O(r^2)) \tag{A.3}$$

where

$$V_{2,N} = \frac{\pi^{N/2}}{\Gamma(1 + N/2)} \tag{A.4}$$

is the Euclidean unit-ball volume in $\mathbb{R}^N$. (Curvature contributes only to the $O(r^2)$ factor and does not change the leading N-scaling used below.)

Coordinate-space L_p unit-ball volume (modeling generalization). For $p \geq 1$, the coordinate-space L_p unit-ball volume is

$$V_{p,N} = \frac{\big(2\Gamma(1 + 1/p)\big)^N}{\Gamma(1 + N/p)} \tag{A.5}$$

Remark Appendix A.1 (Scope of the $p \neq 2$ step)*.* The Fisher/Jeffreys construction is intrinsically Riemannian ($p = 2$). For $p \neq 2$ we keep the same *log-volume vs. radius* algebra but measure representative size using the coordinate-space L_p ball; thus the p-dependence is a controlled modeling generalization (same variational form), not an identity of Fisher geometry.

Appendix A.3. Extracting the expansion term via Gamma/Stirling asymptotics

Euclidean case ($p = 2$): where $\frac{1}{2} \ln N$ comes from. Imposing $\ln \mathrm{Vol}_J(B_r) = H_J$ gives

$$\ln V_{2,N} + N \ln r = H_J \quad \Rightarrow \quad \ln r = \frac{H_J}{N} - \frac{1}{N} \ln V_{2,N} \tag{A.6}$$

Using $\ln \Gamma(z) = z \ln z - z + O(\ln z)$ (Stirling) with $z = 1 + N/2$ yields

$$-\frac{1}{N} \ln V_{2,N} = \frac{1}{2} \ln N + O(1) \tag{A.7}$$

hence, up to N-independent constants,

$$\ln r = \frac{H_J}{N} + \frac{1}{2}\ln N + O(1) \tag{A.8}$$

This is the microscopic origin of the Euclidean expansion cost $\frac{1}{2}\ln N$.

Coordinate-space L_p case: where $\frac{1}{p}\ln N$ comes from. Repeating the same algebra with $V_{p,N}$ gives

$$\ln r = \frac{H_J}{N} - \frac{1}{N}\ln V_{p,N} \tag{A.9}$$

Applying Stirling to $\Gamma(1 + N/p)$ and keeping only the leading N-dependence yields

$$-\frac{1}{N}\ln V_{p,N} = \frac{1}{p}\ln N + O(1) \tag{A.10}$$

so

$$\ln r = \frac{H_J}{N} + \frac{1}{p}\ln N + O(1) \tag{A.11}$$

which reproduces the OED expansion term at leading order.

Appendix A.4. Anisotropy: why H_{struct} subtracts

We show that heterogeneous local scales rescale Jeffreys volume by $\sqrt{\det G}$, producing a log-determinant shift.

Assume locally (in an appropriate basis) the metric is approximately diagonal with heterogeneous scales:

$$G \approx \operatorname{diag}\left(\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_N^2}\right) \quad \Rightarrow \quad \sqrt{\det G} \approx \prod_{i=1}^{N} \sigma_i^{-1} \tag{A.12}$$

Then for a small region (same coordinates as above),

$$\operatorname{Vol}_J(B_r) \approx \left(\prod_{i=1}^{N} \sigma_i^{-1}\right) \operatorname{Vol}_{\text{coord}}(B_r) \approx \left(\prod_{i=1}^{N} \sigma_i^{-1}\right) V_{p,N}\, r^N \tag{A.13}$$

Taking logs and defining the coordinate log-volume $H := \ln \operatorname{Vol}_{\text{coord}}(B_r)$ gives

$$H_J = H - \sum_{i=1}^{N} \ln \sigma_i \;=:\; H - H_{\text{struct}}, \qquad H_{\text{struct}} := \sum_{i=1}^{N} \ln \sigma_i \tag{A.14}$$

(Here σ_i are understood as made dimensionless by a fixed reference scale, so $\ln \sigma_i$ is well-defined.) Substituting $H_J = H - H_{\text{struct}}$ into the leading radius relation yields

$$\ln r = \frac{H - H_{\text{struct}}}{N} + \frac{1}{p}\ln N + O(1) \tag{A.15}$$

Dropping the $O(1)$ constants, minimizing the leading cost $\mathcal{C}(N) = \frac{1}{p}\ln N + \frac{H - H_{\text{struct}}}{N}$ gives the structure-corrected optimum

$$N^* \approx p\,(H - H_{\text{struct}}) \tag{A.16}$$

consistent with the main text.

Appendix B. Uniqueness and Robustness of the Global Minimum

We show the OED cost has a unique global minimizer for $H > 0$, and that the leading-order scaling is stable under common geometric gauges (bulk vs. surface; cube vs. sphere) up to subleading corrections.

Appendix B.1. Uniqueness

Consider

$$\mathcal{C}_p(N) = \frac{1}{p}\ln N + \frac{H}{N}, \qquad N > 0,\ H > 0 \tag{B.1}$$

As $N \to 0^+$, $\mathcal{C}_p(N) \to +\infty$ due to H/N; as $N \to \infty$, $\mathcal{C}_p(N) \to +\infty$ due to $\ln N$. The derivative is

$$\frac{d\mathcal{C}_p}{dN} = \frac{1}{pN} - \frac{H}{N^2} = \frac{N - pH}{pN^2} \tag{B.2}$$

so it changes sign exactly once, at $N^* = pH$. Moreover,

$$\frac{d^2\mathcal{C}_p}{dN^2} = -\frac{1}{pN^2} + \frac{2H}{N^3} \quad \Rightarrow \quad \left.\frac{d^2\mathcal{C}_p}{dN^2}\right|_{N=pH} = \frac{1}{p^3H^2} > 0 \tag{B.3}$$

so $N^* = pH$ is the unique global minimum.

Appendix B.2. Robustness: surface vs. bulk (leading order)

We check that switching the constraint from bulk-like to surface-like only perturbs subleading terms in the optimum for large payload.

For an N-cube with side a, surface measure scales as $A \propto Na^{N-1}$. Imposing $A = e^H$ yields

$$\ln a = \frac{H - \ln(\mathrm{const}\cdot N)}{N-1} \tag{B.4}$$

Using $L_p \propto N^{1/p}a$ gives

$$\mathcal{C}_p^{(\mathrm{surf})}(N) = \frac{1}{p}\ln N + \frac{H - \ln(\mathrm{const}\cdot N)}{N-1} = \frac{1}{p}\ln N + \frac{H - \ln(\mathrm{const}\cdot N)}{N} + O\!\left(\frac{H + \ln N}{N^2}\right) \tag{B.5}$$

so the minimizer satisfies $N^* = pH + o(H)$ for $H \gg 1$ (the additional $-\ln N/N$ term produces only a subleading correction to N^* for $H \gg 1$).

Appendix B.3. Robustness: sphere vs. cube (Euclidean check)

For the Euclidean ball, $\mathrm{Vol} \propto V_{2,N}r^N$ implies

$$\ln r = \frac{H}{N} - \frac{1}{N}\ln V_{2,N} = \frac{H}{N} + \frac{1}{2}\ln N + O(1) \tag{B.6}$$

so minimizing $\ln r$ reproduces $N^* = 2H$. Thus the scaling comes from the same leading $\ln N$ vs. $1/N$ competition, not from the chosen shape.

Appendix C. Anisotropy: Rectangle Model and the Structure-Corrected Law

We (i) derive the exact anisotropic rectangle cost under a bulk constraint by eliminating a global scale factor, and (ii) show that under mild smooth-spectrum conditions the optimizer reduces to $N^* \approx p\,(H - H_{\mathrm{struct}})$ up to $O(1)$ corrections.

Appendix C.1. Exact rectangle cost (bulk constraint elimination)

Setup. Let principal-axis scales be $\sigma_1 \geq \sigma_2 \geq \cdots$ and set rectangle edges $a_i = k\sigma_i$ with $k > 0$. Impose the bulk constraint

$$V = \prod_{i=1}^{N} a_i = k^N \prod_{i=1}^{N} \sigma_i = e^H \quad \Rightarrow \quad \ln k = \frac{1}{N}\left(H - \sum_{i=1}^{N} \ln\sigma_i\right) \tag{C.1}$$

Measure representative size by an L_p norm:

$$L_p \propto \left(\sum_{i=1}^{N} a_i^p\right)^{1/p} = k\left(\sum_{i=1}^{N} \sigma_i^p\right)^{1/p} \tag{C.2}$$

Taking logs and substituting $\ln k$ gives the exact cost

$$\mathcal{C}_{\text{rect}}(N) \coloneqq \ln L_p = \frac{1}{p}\ln\left(\sum_{i=1}^{N}\sigma_i^p\right) + \frac{H - H_{\text{struct}}(N)}{N}, \qquad H_{\text{struct}}(N) \coloneqq \sum_{i=1}^{N}\ln\sigma_i \tag{C.3}$$

Appendix C.2. Leading-order optimizer (and what is being neglected)

We isolate the universal $\frac{1}{p}\ln N$ term and show the remaining N-dependences shift N^* only subleadingly when the spectrum varies slowly.

Write

$$\sum_{i=1}^{N}\sigma_i^p = N\,\overline{\sigma^p}(N), \qquad \overline{\sigma^p}(N) \coloneqq \frac{1}{N}\sum_{i=1}^{N}\sigma_i^p \tag{C.4}$$

so

$$\mathcal{C}_{\text{rect}}(N) = \frac{1}{p}\ln N + \frac{1}{p}\ln\overline{\sigma^p}(N) + \frac{H - H_{\text{struct}}(N)}{N} \tag{C.5}$$

Remark Appendix C.1 (Spectral smoothness and limitations). In the discrete setting, the increment $H_{\text{struct}}(N+1) - H_{\text{struct}}(N) = \ln\sigma_{N+1}$ is typically $O(1)$. In the continuum surrogate, the exact stationarity condition contains additional spectrum-derivative terms. The first-order form $N^* \approx p\,(H - H_{\text{struct}})$ is therefore valid only when those terms are negligible compared to the payload term at the optimum (e.g., in a high-payload regime with sufficiently mild anisotropy over the active band). For *singular*, *spiked*, or *heavy-tailed* spectra, these terms can be non-negligible and the full stationarity condition should be used.

The (continuum) stationarity condition for

$$\mathcal{C}_{\text{rect}}(N) = \frac{1}{p}\ln N + \frac{1}{p}\ln\overline{\sigma^p}(N) + \frac{H - H_{\text{struct}}(N)}{N}$$

is

$$0 = \frac{d\mathcal{C}_{\text{rect}}}{dN} = \frac{1}{pN} + \frac{1}{p}\frac{d}{dN}\ln\overline{\sigma^p}(N) - \frac{H - H_{\text{struct}}(N)}{N^2} - \frac{H'_{\text{struct}}(N)}{N}. \tag{C.6}$$

If the spectrum-derivative terms satisfy

$$\left|\frac{1}{p}\frac{d}{dN}\ln\overline{\sigma^p}(N)\right| \ll \frac{H - H_{\text{struct}}(N)}{N^2}, \qquad \left|\frac{H'_{\text{struct}}(N)}{N}\right| \ll \frac{H - H_{\text{struct}}(N)}{N^2},$$

then the leading balance reduces to $\frac{1}{pN} \approx \frac{H - H_{\text{struct}}(N)}{N^2}$ and hence

$$N^* \approx p\,(H - H_{\text{struct}}), \tag{C.7}$$

with H_{struct} evaluated at the effective dimension.

Appendix D. Stability and the Continuum Approximation

We quantify (i) local convexity at the optimum and (ii) the size of penalties due to (a) integer rounding of N and (b) payload estimation error. This justifies treating N as a continuum variable when H is large.

Consider

$$\mathcal{C}_p(N) = \frac{1}{p}\ln N + \frac{H}{N}, \qquad N^* = pH \tag{D.1}$$

The curvature at the optimum is

$$\left.\frac{d^2\mathcal{C}_p}{dN^2}\right|_{N=pH} = \frac{1}{p^3H^2} > 0 \tag{D.2}$$

so the minimum is locally convex.

Quadratic penalty near N^*. For a small perturbation δN,

$$\Delta\mathcal{C} \;\approx\; \frac{1}{2}\,\frac{d^2\mathcal{C}_p}{dN^2}\bigg|_{N^*}(\delta N)^2 = \frac{(\delta N)^2}{2p^3H^2} \tag{D.3}$$

Thus integer quantization (e.g. $|\delta N| \le 1/2$) incurs a vanishing $O(H^{-2})$ cost for $H \gg 1$.

Penalty from payload misspecification. If the payload is estimated as $H + \delta H$, the implied optimum shifts by $\delta N \approx p\,\delta H$. Substituting into the quadratic penalty yields

$$\Delta\mathcal{C} \;\approx\; \frac{(p\,\delta H)^2}{2p^3H^2} = \frac{1}{2p}\left(\frac{\delta H}{H}\right)^2 \tag{D.4}$$

Hence the relevant control parameter is the *relative* payload error $\delta H/H$, and the continuum treatment of N is justified in the high-payload regime.