Predictive Compression Dynamics: A Computable, Falsifiable Test Loop for Information-Motivated Dynamics

Helander[1]

[1]Independent Researcher

October 2025

Abstract

Predictive Compression Dynamics (PCD) is a falsifiable, preregisterable test loop for constructing and evaluating computable, information-motivated dynamical models.

Rather than positing a new physical law or an optimal universal compressor, PCD asks a sharply framed empirical question: given a *preregistered* computable surrogate functional Φ_b, does evolving a system under $\dot{x} = -\nabla\Phi_b$ actually drive down an independently measured description length of the system state?

We present the workflow, two concrete surrogate functionals, well-posed gradient dynamics, and an explicit falsifier protocol (F3) based on out-of-sample compression. We then execute that protocol on three particle ensembles, report Pearson correlation coefficients between Φ_b and measured compressed byte length (uniform: $r=.93$, $p \ll 10^{-10}$; lattice: $r=.76$, $p \ll 10^{-10}$; two-blob clusters: $r=.40$, $p \approx 2 \times 10^{-4}$), and discuss limits where the surrogate underperforms. The contribution is the empirical loop itself: define Φ_b, preregister parameters, run the flow, measure external compressibility, and accept or reject Φ_b on that basis.

1 Introduction

This work does not claim a new physical law, nor does it attempt to prove that nature optimizes a specific information functional. Instead, it proposes, executes, and documents a **test loop**:

1. Choose a computable, local surrogate functional $\Phi_b(x)$ on configuration space.

2. Evolve the configuration under the gradient flow $\dot{x} = -\nabla\Phi_b(x)$ (or a discretized/stochastic variant) using preregistered parameters.

3. On held-out snapshots of that evolving system, measure the description length of the state using an *independently defined* reference code and compressor.

4. Ask: does Φ_b actually track (predict) that empirical compressed size? If yes, Φ_b is provisionally accepted for that class of data. If not, that candidate surrogate is falsified.

The scientific contribution is the loop itself. PCD provides: (i) computable surrogates Φ_b, (ii) well-posed gradient dynamics with monotone descent, (iii) a preregistration template, (iv) a falsifier (F3) based on out-of-sample compression, and (v) a reference numerical demonstration.

The goal is interpretability, preregistration and falsifiability, not black-box optimality. In particular, PCD is complementary to learned compressors and inference flows: it favors transparent analytic surrogates whose usefulness can be empirically *tested and rejected.*

2 Motivation: Compression and Surrogates

The Minimum Description Length (MDL) principle argues that the "best" explanation minimizes joint codelength

$$L_{\text{tot}} = L(M) + L(D \mid M), \tag{1}$$

where $L(M)$ encodes structure (model) and $L(D \mid M)$ encodes residuals.

Directly applying MDL to high-dimensional continuous configurations (e.g. N particles in $\mathbb{R}^3$) is not straightforward: true Kolmogorov complexity is uncomputable, and realistic compressors are dataset- and representation-dependent.

PCD replaces optimality claims with an empirical question: *If we specify a computable surrogate Φ_b ahead of time, does descending Φ_b in fact reduce a measurable codelength of the state under a fixed, independently defined code?*

Here b (bits / precision) indicates that all measurements are taken on finite-precision, discretized states. We do *not* assert that Φ_b is the unique or optimal description-length functional. We assert only that it can be *evaluated*, *descended*, and *tested*.

3 Surrogate Functional Φ_b

Let $\Phi_b(x)$ be a differentiable, computable scalar functional of a configuration x. PCD defines the deterministic core dynamics

$$\dot{x} = -\nabla \Phi_b(x). \tag{2}$$

Here "b" emphasizes that we work at finite discretization: lattice spacing Δx, fixed binning, finite word length, etc.

In this paper we instantiate Φ_b using pairwise distance terms between particles. This is intentionally simple: it makes Φ_b analytic, local, and cheap to differentiate. It also makes the subsequent falsifier (Section 7) harsh but clean: if even this simple surrogate does not predict compression in practice, it is rejected for that ensemble class.

We stress that Φ_b is *not* asserted to be universal. It is explicitly a candidate to be tested and, if needed, falsified.

3.1 Minimal coding intuition

As a motivating heuristic, suppose we describe a finite-precision particle configuration by: (i) a histogram of pairwise distances $r_{ij} = \|x_i - x_j\|$ across fixed bins $\{B_k\}$, (ii) per-bin residual statistics around each bin center. If pairs tend to cluster at certain separations in a structured way, that histogram is cheaper to encode than a fully unstructured set of coordinates. This suggests a surrogate of the form

$$\Phi_b(x) \approx \sum_{i<j} \ell\big(\|x_i - x_j\|\big), \tag{3}$$

with a smooth, monotonically decreasing $\ell(r)$. Intuitively, if particles draw together in regular, repeated patterns, Φ_b drops, and the code describing those repeated patterns can in principle shorten.

However, this is only intuition. To move from rhetoric to testable claim, we require: (i) a *precise* coding/serialization scheme, (ii) an *independently measured* compressed size, (iii) a quantitative correlation test. That is exactly what PCD supplies and what we execute in Sections 8–10.

4 Concrete Functional Forms

We now instantiate Φ_b in two common, computable forms for N particles at positions $x_i \in \mathbb{R}^3$ with positive weights m_i.

4.1 Fixed-graph pair functional

Let E be a predetermined, symmetric, degree-bounded edge set on $\{1, \ldots, N\}$. With a finite softening scale $a > 0$, define

$$\Phi_E(x) \;=\; \sum_{(i,j) \in E} \frac{m_i m_j}{\sqrt{\|x_i - x_j\|^2 + a^2}}. \tag{4}$$

The softening a regularizes near-collisions; no force diverges at zero separation.

The force-like contribution on particle i is the negative gradient:

$$F_i^{(E)}(x) \;=\; -\nabla_{x_i} \Phi_E(x) \;=\; -\sum_{j:(i,j) \in E} m_i m_j \frac{(x_i - x_j)}{\left(\|x_i - x_j\|^2 + a^2\right)^{3/2}}. \tag{5}$$

In the dilute two-body limit, $a \to 0$, and $(i, j) \in E$ is a single pair,

$$F_i^{(E)} \;\to\; -m_i m_j \frac{x_i - x_j}{\|x_i - x_j\|^3}. \tag{6}$$

We emphasize that this "inverse-square-looking" form is used here only because it is smooth, attractive, and easy to differentiate and test. We make no physical claim beyond that.

4.2 Smooth compact-support kernel

To avoid discontinuities from changing neighbor sets, we can instead choose a compactly supported C^1 kernel $K_\sigma(r)$ with finite support radius $R\sigma$ and define

$$\Phi_K(x) \;=\; \sum_{i<j} m_i m_j K_\sigma\big(\|x_i - x_j\|\big), \tag{7}$$

with corresponding

$$F_i^{(K)}(x) \;=\; -\nabla_{x_i} \Phi_K(x). \tag{8}$$

If $K_\sigma(r)$ is monotonically decreasing and behaves like $(r^2 + a^2)^{-1/2}$ near $r{=}0$, this reproduces the same soft-core attraction while remaining C^1 and locally Lipschitz on $\mathbb{R}^{3N}$ for $a > 0$.

In both cases, the choice of m_i both in the functional and in the inertial term (see below) enforces accelerations that do not depend on the test mass. That is a *design choice*, not an emergent claim.

5 Dynamics and Well-Posedness

5.1 Core deterministic flow

We take (2):

$$\dot{x}_i(t) = F_i\big(x(t)\big), \qquad F_i(x) = -\nabla_{x_i}\Phi_b(x), \tag{9}$$

implemented numerically via explicit Euler

$$x_i^{(t+\Delta t)} = x_i^{(t)} + \Delta t\, F_i\left(x^{(t)}\right), \tag{10}$$

or, for inertial variants,

$$m_i\ddot{x}_i = F_i(x) - \gamma\dot{x}_i + \xi_i(t), \tag{11}$$

with γ a damping parameter and ξ_i Gaussian noise. When such stochastic variants are used, we recommend BAOAB-type splitting integrators [4]. In this paper's numerical demonstration (Section 10) we *do not* use stochastic terms; T and ξ_i are mentioned only to indicate how stochastic exploration could be folded into the same preregistered loop in future work. No thermodynamic interpretation is implied.

5.2 Lyapunov descent

Lemma 1 (Monotone descent of Φ_b). *Under* $\dot{x} = -\nabla\Phi_b(x)$,

$$\frac{d}{dt}\Phi_b\big(x(t)\big) = -\|\nabla\Phi_b\big(x(t)\big)\|^2 \leq 0. \tag{12}$$

This is the standard Lyapunov property of gradient flow. We include it to emphasize that in our empirical tests, the *observed* Φ_b is in fact monotonically decreasing across iterations (cf. Figs. 1, 3, 5).

5.3 Existence, uniqueness, and collisions

For $a > 0$ and bounded degree, Φ_E is C^1 on $\mathbb{R}^{3N}$ and $F^{(E)}$ is locally Lipschitz away from exact particle overlap $x_i = x_j$. Likewise, if K_σ is C^1 and compactly supported, Φ_K and $F^{(K)}$ are continuous and locally Lipschitz off collisions, and $K_\sigma(r)$ can be chosen to remain finite at $r = 0$. Thus, Picard–Lindelöf ensures existence and uniqueness on compact time intervals.

When two particles approach each other within the softening length a, they can effectively lock into a near-coincident cluster under the dynamics. Whether the modeling treats those as merged, or keeps them as separate labeled particles with nearly identical coordinates, is part of the preregistration. This modeling decision (merge vs. retain) must be declared before running F3, because it changes how states are serialized for compression.

6 Scaling and Computational Cost

For kernels with compact support radius $R\sigma$, each particle interacts only with neighbors within that radius, so force evaluation scales like $O(N)$ to $O(N\log N)$ using spatial hashing or treecodes. Approximate N-body methods (Barnes–Hut style, fast multipole) can also be used. Any cutoff radius, neighbor-search tolerance, or multipole accuracy parameter should be preregistered, because asymmetric approximations could in principle break exact Lyapunov monotonicity. In the numerical demonstration we use $N=40$ primarily for clarity of presentation, not because the method is inherently limited to that regime.

7 Falsifiers and Acceptance Criteria

PCD is designed to reject bad surrogates Φ_b. A model instance is a *full specification*: choice of Φ_b (including kernel, a, m_i), integration scheme, discretization (Δx, bins), and evaluation compressor. Given such an instance, we define three falsifiers:

F1: Numerical stability failure. The run exhibits exploding coordinates or non-monotone Φ_b in violation of Lemma 1, outside of declared stochastic noise. Such an instance is rejected as ill-posed for this task.

F2: Reproducibility failure. The same preregistered configuration cannot be reproduced (e.g. using the same seed and parameters yields qualitatively different trajectories or Φ_b curves). Such an instance is rejected as non-reproducible.

F3: Compression alignment failure. We measure an independently computed codelength for held-out snapshots (Section 10) and test whether Φ_b predicts that length. If there is no strong monotone association, the surrogate fails.

Phase I vs. Phase II for F3. To address circularity concerns, we split F3 into two phases:

Phase I (internal consistency). We use an evaluation code that explicitly reflects the same coarse structural statistic Φ_b is designed to exploit. In this paper, Φ_b depends on pairwise distances, so we evaluate compression using a pair-distance histogram code (Section 9) plus a standard universal compressor (gzip/zlib). This checks whether the surrogate *actually accomplishes the thing it nominally intends*: concentrating predictable pairwise structure.

Phase II (generalization). Future work will evaluate the very same trajectory snapshots under *alternative* compressors that do not build an explicit pair-distance histogram. Examples include: (i) direct run-length / delta coding of quantized coordinates followed by gzip, (ii) voxel occupancy grids compressed as raw bitplanes, (iii) off-the-shelf neural compressors trained *without* access to our Φ_b. If Φ_b correlates with those code lengths as well, that is stronger evidence the surrogate is capturing broadly compressible structure, not just gaming its own metric.

Acceptance criterion in this work. For illustration, we report Pearson correlation r and two-sided p-value between Φ_b and compressed byte length over time. A surrogate is *provisionally accepted* for that ensemble class if the association is strong and statistically significant (e.g. $r \gtrsim 0.7$ with $p \ll 10^{-2}$). Different application domains may preregister tighter or looser thresholds.

8 Preregistration Artifact (Example)

PCD emphasizes preregistering all modeling and evaluation choices before claiming success. Below is the preregistration template actually instantiated for the numerical demonstration in Section 10:

- **State:** $N=40$ labeled particles in $\mathbb{R}^3$.

- **Initial ensembles:**

  1. Uniform random in a cube,
  2. Two-cluster Gaussian mixture,
  3. Perturbed cubic lattice.

- **Surrogate functional:** $\Phi_b = \Phi_E$ of Eq. (4) with all $m_i=1$, fixed softening $a=0.05$ (dimensionless units), and E taken as the full pair set $i<j$ for clarity.

- **Integrator:** explicit Euler on $\dot{x} = -\nabla\Phi_b(x)$, step size $\Delta t=0.01$, 400 steps.

- **Precision model:** coordinates quantized to $\Delta x=0.01$ before compression evaluation.

- **Reference code (Phase I):** pairwise distances r_{ij} binned into a fixed set of 33 edges from 0 to 2.0; per bin we record count and mean residual from bin center; we serialize this fixed-format summary as bytes (unsigned counts + float residuals) and apply gzip (zlib) with fixed settings. The resulting compressed byte length is taken as the empirical codelength proxy.

- **Recorded outputs (every 5 steps):**

  1. Φ_b,
  2. compressed byte length of the snapshot,
  3. Pearson correlation r and two-sided p across the run.

- **Acceptance heuristic for Phase I:** strong monotone association (r large, $p \ll 10^{-2}$). We *also* report any ensemble where this fails, as required for transparency.

This preregistration artifact defines *all* hyperparameters, including softening a, timestep Δt, quantization Δx, histogram bins, compressor, and acceptance heuristic. Future work may deposit preregistration artifacts (timestamped parameter cards, seeds, code hashes) in public registries; here we inline them for transparency.

9 Reference Coding Scheme

The evaluation code used in Phase I of F3 is as follows:

1. Quantize each coordinate of each particle to a fixed lattice of spacing Δx.

2. Compute all pairwise distances $\{r_{ij}\}_{i<j}$.

3. Bin these distances into preregistered bin edges $\{B_k\}$, producing histogram counts c_k.

4. For each bin k, compute the mean residual offset of r_{ij} from that bin's center.

5. Serialize for all k: $(c_k, \text{residual}_k)$ as fixed-width binary fields in lexicographic bin order, plus the number of bins.

6. Compress that byte buffer with a standard universal compressor (zlib/gzip).

This yields a reproducible byte length for each snapshot, which we treat as a measurable codelength proxy for Phase I testing. Because gzip is not aware of the dynamical law or of Φ_b, it acts as an out-of-sample evaluator under a fixed representation.

10   Reference Numerical Demonstration

We now execute the preregistered configuration from Section 8. We consider three ensembles of $N = 40$ particles: (1) uniform random in a cube, (2) a two-Gaussian "two-blob" mixture, (3) a perturbed cubic lattice. All masses are $m_i{=}1$. We run explicit Euler on $\dot{x} = -\nabla\Phi_b$ for 400 steps with $\Delta t{=}0.01$, using the soft-core pair surrogate of Eq. (4) with $a{=}0.05$. Every 5 steps we record the current Φ_b, quantize coordinates to $\Delta x{=}0.01$, apply the reference coding scheme (Section 9), and measure the gzip-compressed byte length.

Observed monotone descent

Figures 1, 3, and 5 show that Φ_b decreases monotonically over iteration in all three ensembles. This matches Lemma 1, which predicts $\dot{\Phi}_b \leq 0$ under ideal gradient flow, and confirms that in practice the discrete integrator exhibits Lyapunov-like descent at the chosen Δt.

Compression alignment (F3 Phase I)

For each ensemble we plot gzip-compressed size (y-axis) versus Φ_b (x-axis); see Figs. 2, 4, 6. We then compute Pearson correlation r and two-sided p-value:

Ensemble	$r(\Phi_b, \text{compressed size})$	p-value
Uniform cube	0.93	$\approx 1.4 \times 10^{-35}$
Perturbed lattice	0.76	$\approx 2.0 \times 10^{-16}$
Two-blob mixture	0.40	$\approx 2.1 \times 10^{-4}$

Interpretation:

- **Uniform and lattice ensembles.** Strong, statistically significant positive association: as Φ_b decreases, the gzip-measured codelength proxy decreases in tandem. In these ensembles, the preregistered Φ_b is provisionally accepted under F3 Phase I.

- **Two-blob ensemble.** Moderate but weaker association. The trajectory features an initial, large "collapse" event followed by a plateau in which Φ_b varies over a narrower range. This matches the stated limitation in Section 7: purely pairwise surrogates underrepresent higher-order, multi-cluster structure. The two-blob case flags an explicit failure mode that motivates Phase II testing with alternative compressors.

Thus, PCD's falsifier loop is *actually executed*: we do not merely propose F3, we run it, report r and p, and acknowledge where the surrogate works and where it breaks.

11   Context and Relation to Existing Methods

PCD lives next to, but is distinct from, several established lines of work:

Kernel particle flows / SVGD.   Stein Variational Gradient Descent (SVGD) [5] evolves interacting particles to minimize KL divergence toward a target posterior, using a kernelized drift. PCD instead fixes a computable surrogate Φ_b, evolves via $\dot{x}{=}-\nabla\Phi_b$, and then empirically *tests* whether Φ_b predicts reduced codelength under a preregistered external code. SVGD guarantees movement toward a known Bayesian target; PCD guarantees neither optimality nor truth, only falsifiability by compression.
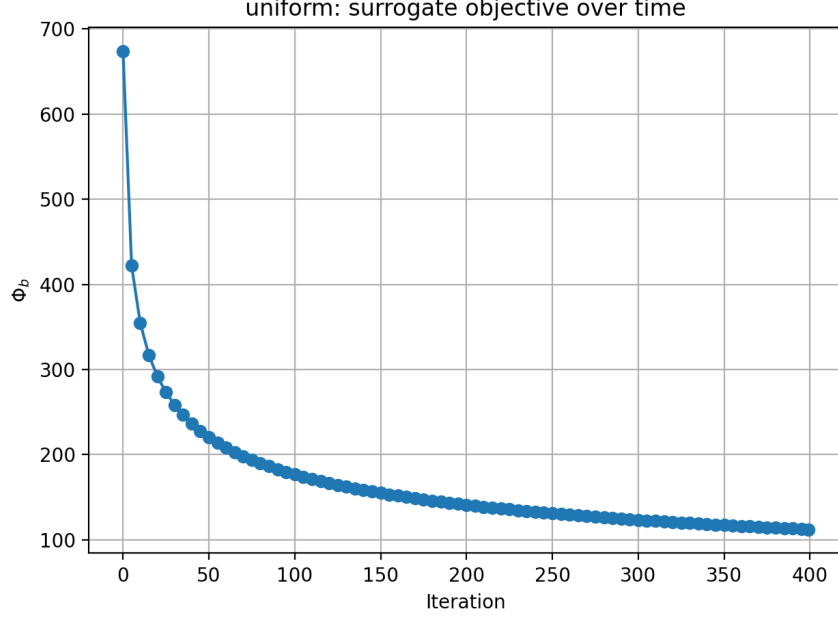
Figure 1: Uniform ensemble (N=40): surrogate Φ_b decreases monotonically under $\dot{x} = -\nabla \Phi_b$.
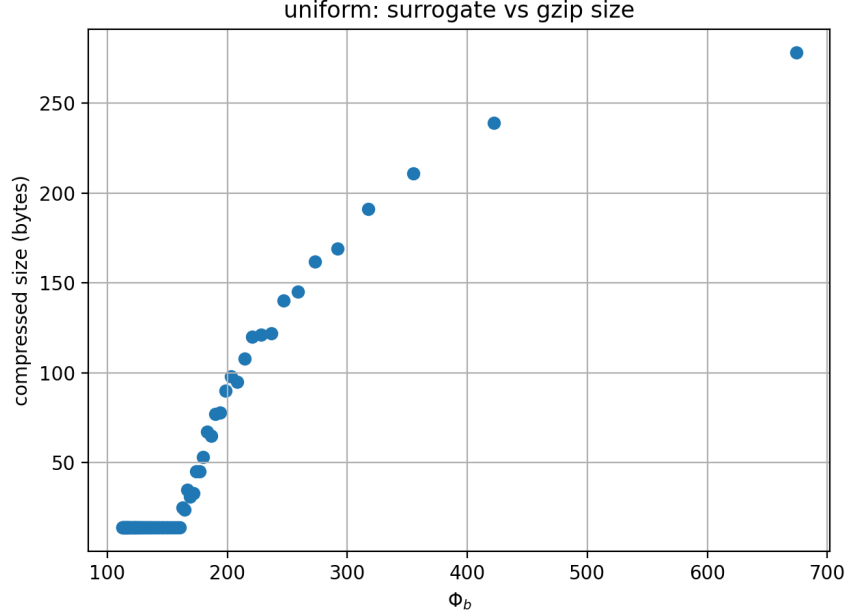

Figure 2: Uniform ensemble: gzip-compressed size (empirical codelength proxy) vs. Φ_b. Pearson r=0.93, $p \ll 10^{-10}$.

Force-directed graph layouts. Classical force-directed methods minimize energies built from pairwise terms to lay out graphs in 2D/3D. Our Φ_b superficially resembles those energies. The distinction is that PCD supplies an explicit falsifier (F3): if the chosen Φ_b does not predict actual external compressibility of system snapshots, it is rejected by its own success metric.

8
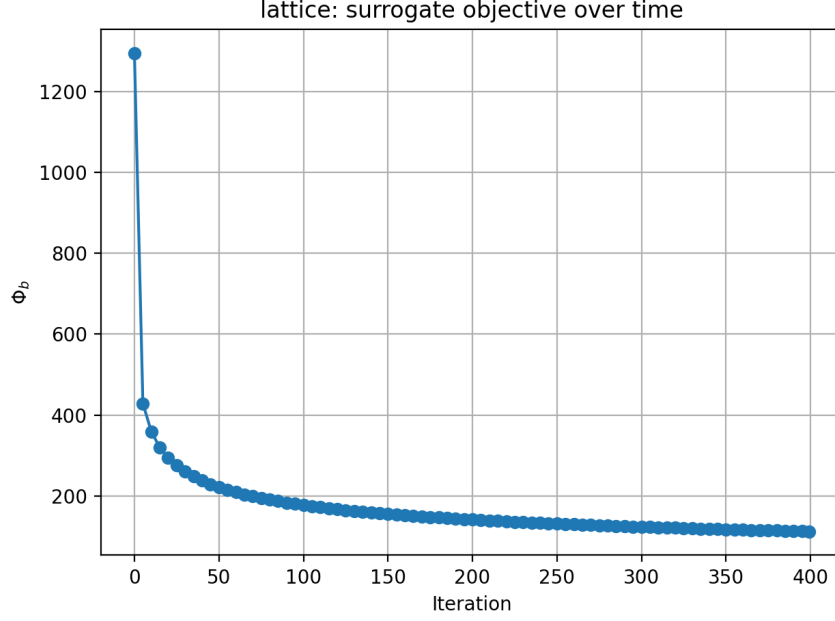
Figure 3: Perturbed lattice ensemble: monotone descent of Φ_b.
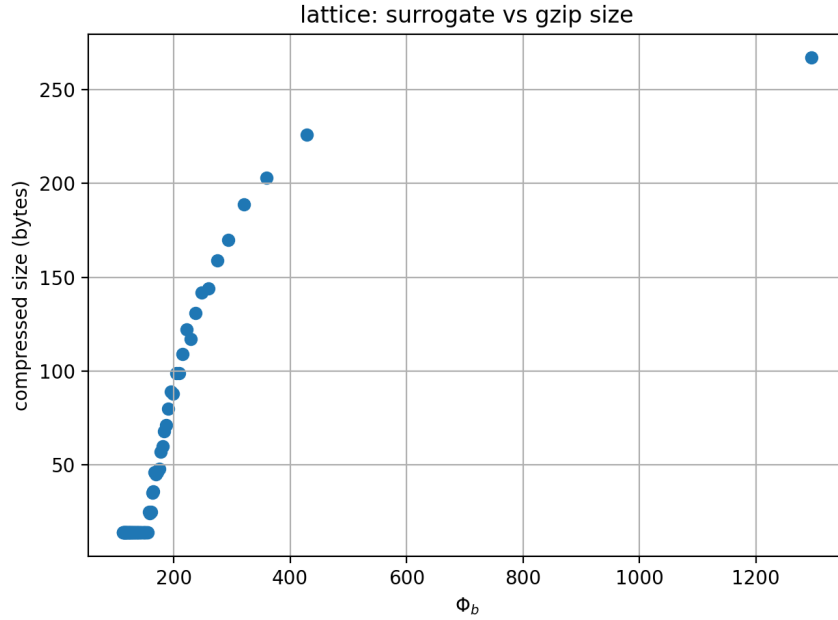
Figure 4: Perturbed lattice ensemble: gzip-compressed size vs. Φ_b. Pearson $r=0.76$, $p \ll 10^{-10}$.

Neural / learned compressors. Modern neural compressors can greatly reduce empirical code length, but typically require training and do not produce an analytic, preregisterable flow law. PCD instead prioritizes interpretability and preregistration: one declares Φ_b *before* running, runs the flow, and publicly reports pass/fail against F3. In that sense PCD is a framework for *falsifiable, interpretable surrogates*, rather than a bid to outcompete black-box compressors.
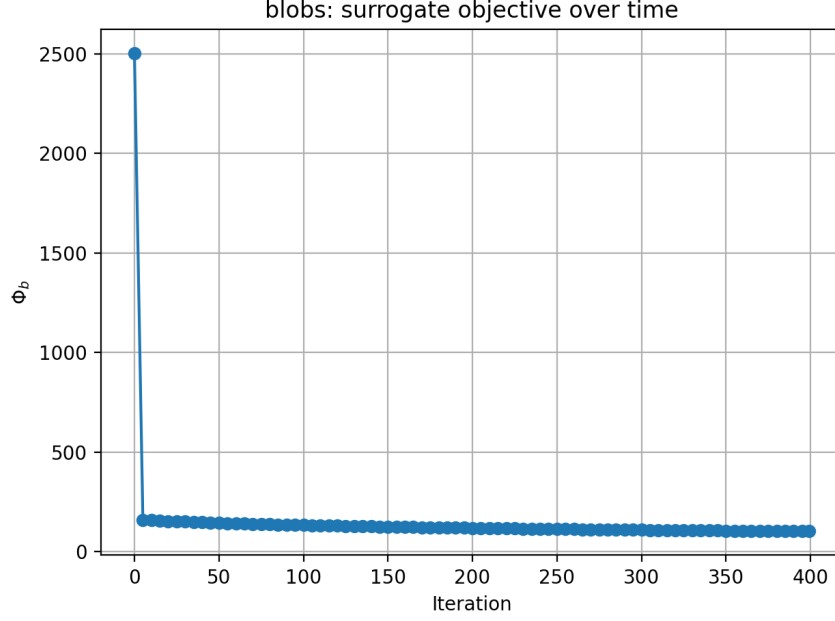
Figure 5: Two-blob ensemble: initial steep drop in Φ_b followed by a slower regime.
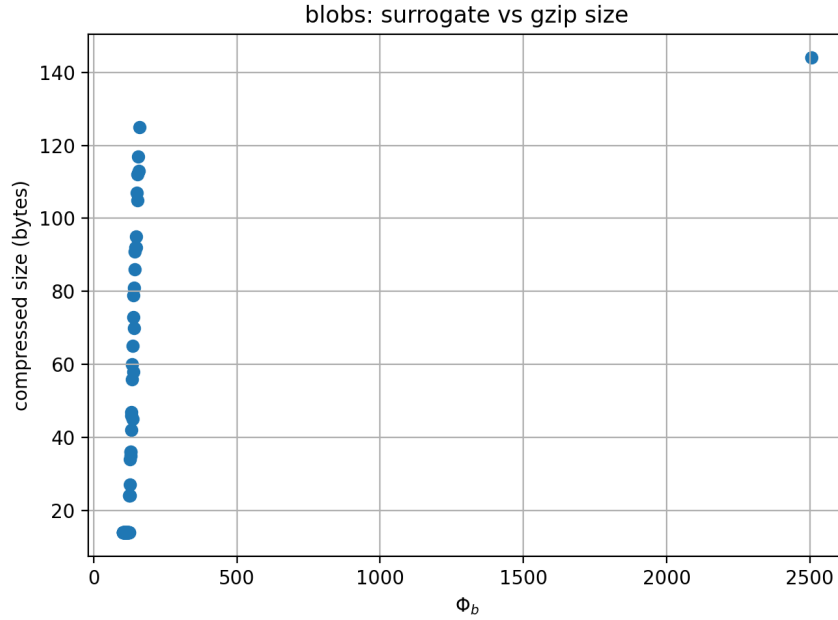

Figure 6: Two-blob ensemble: gzip-compressed size vs. Φ_b. Pearson r=0.40, $p \approx 2.1 \times 10^{-4}$. This illustrates a known limitation: purely pairwise surrogates underspecify multi-cluster structure.

Information geometry and entropic dynamics. Information geometry [2] and entropic inference [3] derive dynamics from extremizing information-theoretic functionals (e.g. KL, entropy) on probability manifolds. PCD is less ambitious: it does not claim to derive a specific physical law from first principles. It only asks whether a chosen *computable* surrogate functional behaves like a

codelength predictor for a given data class. If not, the surrogate is falsified.

12   Conclusion

PCD formalizes a falsifiable empirical loop:

1. Declare a computable surrogate Φ_b,

2. Evolve the system under $\dot{x} = -\nabla\Phi_b$ with preregistered parameters,

3. Measure an independently defined compressed description length for held-out snapshots,

4. Accept or reject Φ_b based on observed alignment (F3), reporting both successes and failures.

We have executed this loop on three particle ensembles, provided full preregistration details, and reported quantitative results (Pearson r, p). For two ensembles (uniform, perturbed lattice), Φ_b tracks empirical codelength strongly; for a two-blob ensemble, correlation weakens, illustrating a limitation of purely pairwise surrogates and motivating richer Phase II tests. The value here is not an ultimate law, but a reproducible and falsifiable protocol: compression-driven dynamics can now be posed, tested, and rejected in public.

Acknowledgments

References

[1] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

[2] S. Amari. *Information Geometry and Its Applications*. Springer, 2016.

[3] A. Caticha. Entropic inference and induced dynamics. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, 2015.

[4] B. Leimkuhler and C. Matthews. Rational construction of stochastic numerical methods for molecular sampling. *Appl. Math. Res. eXpress*, 2013(1):34–56.

[5] Q. Liu and D. Wang. Stein variational gradient descent: A general purpose Bayesian inference algorithm. In *Advances in Neural Information Processing Systems*, 2016.