Predictive Compression Dynamics: A Methodological Framework for Computable Information-Motivated Modeling

Mats Helander[1] and Jeeves*[1]

[1]Independent Research

Dated: October 24, 2025

Abstract

We present *Predictive Compression Dynamics* (PCD), a methodological recipe for constructing *computable*, local functionals Φ_b and driving dynamics by gradient flow $\dot{x} = -\nabla\Phi_b(x)$ with preregistered parameters. Two concrete instances are given: (i) a fixed-graph pair functional and (ii) a smooth compact-support kernel; both yield stable, attractive gradient terms (after calibration) and admit Lyapunov descent. These models serve as methodological demonstrations of computable, information-motivated optimization. We make the MDL split $L_{\mathrm{tot}} = L(M) + L(D \mid M)$ explicit, give a minimal coding scheme linking Φ_b descent to achievable ΔL_{tot}, address well-posedness (smooth-kernel variant), recommend robust integrators (BAOAB for Langevin), and provide a preregistration/model-card template and falsifiers for a chosen model instance. The goal is a reproducible toolbox for compression-driven dynamics across domains.

1 Positioning and Commitments

A disciplined workflow to construct and test *computable* local functionals Φ_b whose gradients define algorithmic descent rules, with explicit preregistration (domain, discretization, kernels, parameters), numerical sanity checks, and internal falsifiers.

2 Operational Domain and Notation

We consider N point particles with positions $x_i \in \mathbb{R}^3$ and positive weights m_i. Computations use finite precision: lattice spacing a_{grid} and b bits/axis, stated *a priori.* The subscript b in Φ_b denotes dependence on numerical precision (bits of representation). A global calibration constant $G_{\mathrm{eff}} > 0$ maps dimensionless gradients to physical units if desired. Using identical m_i in both interaction and inertial terms is a modeling simplification that enforces equal accelerations by design. Vectors are written in plain type for brevity.

3 Model–Data Decomposition and Coding Link

Following MDL, we split description length as

$$L_{\mathrm{tot}} = L(M) + L(D \mid M), \tag{3.1}$$

*Jeeves is a pseudonym for an AI assistant contributing to analysis, code design, and manuscript preparation.

where $L(M)$ encodes modeled regularities and $L(D \mid M)$ encodes residuals given M. A decrease $\Delta L_{\text{tot}} < 0$ corresponds to realized compression. PCD treats a computable, local Φ_b as a proxy for the achievable total codelength L_{tot} itself; hence $\dot{x} = -\nabla\Phi_b$ implements a descent in surrogate description length under the chosen model family.

Minimal explicit coding scheme. Let (i, j) range over a symmetric set of "near" pairs. A two-part code describes (i) a shared pairwise template per distance bin and (ii) residual offsets:

- Partition distances into bins $\{B_k\}$ with centers r_k; encode the histogram counts using an arithmetic code with probability p_k proportional to frequency.
- For each pair (i, j) with $r_{ij} \in B_k$, encode a residual offset δr_{ij} relative to r_k using bounded precision.

The expected codelength per pair is

$$\ell(r_{ij}) = -\log p_k + H_{\text{res}}(\delta r \mid B_k),$$

and the overall code is prefix-free, satisfying Kraft's inequality. Then

$$L_{\text{tot}} \approx \text{const} + \sum_{(i,j)} \ell(r_{ij}). \tag{3.2}$$

Choosing

$$\Phi_b \ \propto \ \sum_{(i,j)} \ell(r_{ij})$$

makes $-\nabla\Phi_b$ a gradient descent in achievable total codelength. A convenient smooth surrogate is

$$\ell(r) \approx (r^2 + a^2)^{-1/2},$$

which provides bounded curvature at $r = 0$ and $1/r$ asymptotics. Other forms may be substituted without altering the workflow.

Proposition 3.1 (Surrogate MDL Descent). *Suppose $L_{tot} = \text{const} + \sum_{(i,j)} \ell(r_{ij})$ with $\ell'(r) \le 0$ and $\Phi_b = \kappa \sum_{(i,j)} \ell(r_{ij})$ for some $\kappa > 0$. Then along $\dot{x} = -\nabla\Phi_b(x)$ we have*

$$\frac{d}{dt}\Phi_b(x(t)) = -\|\nabla\Phi_b(x(t))\|^2 \le 0, \tag{3.3}$$

with equality iff $\nabla\Phi_b(x(t)) = 0$.

Interpretation. Φ_b is a computable surrogate for total codelength; its monotone decrease under $\dot{x} = -\nabla\Phi_b$ represents achievable compression within the chosen model family.

4 Information-Motivated Surrogates for Gradient Descent

4.1 Fixed-graph functional

Let $E \subset \{(i, j) : 1 \le i < j \le N\}$ be a symmetric, degree-bounded edge set. Define

$$\Phi_E(x) = \sum_{(i,j)\in E} \frac{m_i m_j}{\sqrt{\|x_i - x_j\|^2 + a^2}}, \qquad a > 0. \tag{4.1}$$

The gradient term for element i is

$$G_i^{(E)}(x) = -\nabla_{x_i}\Phi_E(x) = -\sum_{\substack{j:\\(i,j)\in E}} m_i m_j \frac{(x_i - x_j)}{\left(\|x_i - x_j\|^2 + a^2\right)^{3/2}}. \tag{4.2}$$

In the two-particle case with $(i, j) \in E$ and $a \to 0$,

$$G_i^{(E)} \to -\, m_i m_j \frac{x_i - x_j}{\|x_i - x_j\|^3}, \tag{4.3}$$

i.e. an attractive inverse-square form along the inter-particle direction.

4.2 Smooth-kernel functional

To avoid neighbor-set discontinuities, choose a compactly supported, C^1 radial kernel $K_\sigma : [0,\infty) \to \mathbb{R}_{\geq 0}$ with support $\subset [0, R\sigma]$. Define

$$\Phi_K(x) = \sum_{i<j} m_i m_j \, K_\sigma(\|x_i - x_j\|), \tag{4.4}$$

so $G_i^{(K)}(x) = -\nabla_{x_i}\Phi_K(x)$ is continuous and locally Lipschitz off collisions. If $K_\sigma(r) \sim (r^2 + a^2)^{-1/2}$ near $r = 0$, one recovers the regularized two-particle form (4.3).

5 Algorithmic Dynamics and Integrators

We preregister all parameters $(a_{\text{grid}}, b, a, \sigma, \Delta t, m_i, \gamma, T, \text{seeds})$.

Lemma 5.1 (Compression-Rate Identity). *Under $\dot{x} = -\nabla\Phi_b(x)$ the instantaneous surrogate codelength rate is*

$$\dot{\Phi}_b(t) = -\|\nabla\Phi_b(x(t))\|^2 \leq 0. \tag{5.1}$$

Hence Φ_b is a Lyapunov function and its monotone decrease represents achievable compression under the model.

Deterministic gradient flow. Explicit Euler:

$$x_i^{(t+\Delta t)} = x_i^{(t)} + \Delta t \, G_i(x^{(t)}), \quad G_i \in \{G_i^{(E)}, G_i^{(K)}\}. \tag{5.2}$$

For stability, use adaptive Δt or semi-implicit variants.

Stochastic descent (BAOAB recommended).

$$m_i \ddot{x}_i = G_i(x) - \gamma \, \dot{x}_i + \xi_i(t), \quad \langle \xi_i(t)\xi_j(t')\rangle = 2\gamma k_B T \, \delta_{ij}\delta(t-t'). \tag{5.3}$$

We recommend the BAOAB integrator with reported weak/strong orders.

6 Sanity Checks

With $a \to 0$ and a single pair, (4.3) holds (after calibration G_{eff}). For $r \gg a$,

$$\frac{r}{(r^2+a^2)^{3/2}} = \frac{1}{r^2}\left(1 - \frac{3a^2}{2r^2} + O\left(\frac{a^4}{r^4}\right)\right), \tag{6.1}$$

so

$$\|G_i^{(E)}\| = m_i m_j \frac{r}{(r^2+a^2)^{3/2}} \approx m_i m_j \frac{1}{r^2}\left(1 - \frac{3a^2}{2r^2}\right). \tag{6.2}$$

These expansions serve purely as numerical consistency checks.

7 Well-posedness

For $a > 0$ and bounded degree, $\Phi_E \in C^1(\mathbb{R}^{3N} \setminus \{x_i = x_j\})$ and $G^{(E)}$ is locally Lipschitz off collisions. For C^1 kernels with bounded K'_σ, $G^{(K)}$ is continuous and locally Lipschitz. Existence and uniqueness follow by Picard–Lindelöf on compact intervals. For dynamic kNN, gradients are piecewise smooth; employ hysteresis or prefer the smooth kernel.

8 Preregistered Model Card (example)

Domain. $a_{\text{grid}} = 10\,\mu\text{m}$, $b = 16$.
Functional. Φ_K with Wendland C^2 kernel ($\sigma = 0.5\,\text{mm}$); softening $a = 50\,\mu\text{m}$.
Dynamics. BAOAB stochastic descent with ($m_i \equiv 1, \gamma = 0.1, T = 300\,\text{K}$), $\Delta t = 1 \times 10^{-3}\,\text{s}$.
Calibration. Fit G_{eff} by least-squares on the slope of $\|G_i^{(E)}\|$ versus r^{-2} across sampled separations in a dilute two-point sandbox at $r \gg a$; hold fixed thereafter.
Sanity checks. Verify (4.3) and the far-field expansion; report seeds and residuals.

9 Falsifiers for a Chosen Instance

Given fixed (Φ_b, params), declare the instance falsified if:
(F1) Two-point trajectories disagree with the calibrated reference form beyond numerical error.
(F2) Smooth-kernel vs fixed-graph variants differ systematically at small r beyond topology effects.
(F3) The surrogate Φ_b correlates poorly with *out-of-sample* compression of generated data (e.g. compare Φ_b to actual Lempel–Ziv compression of held-out pair-distance histograms rather than the in-sample surrogate).

10 Discussion and Scope

PCD supplies a reproducible route from *computable* information-motivated functionals to concrete algorithmic descent schemes. That simple pairwise surrogates coincide with familiar inverse-square interactions is a feature for validation, not a claim of novelty. Future work will broaden Φ_b (e.g. learned local codes, graph Laplacians) under the same preregistration discipline.

Application domains. Although demonstrated on abstract particle configurations, the same workflow applies wherever local similarity drives redundancy reduction—particle-based learning objectives, swarm control, coarse-grained fluid solvers, or clustering under computational constraints. The physical units in examples (µm–mm) serve only as scale illustrations.

11 Context and Relation to Existing Frameworks

PCD complements algorithmic-thermodynamic and information-geometric programs by operating directly in finite-precision configuration space, with explicitly computable surrogates and preregistered parameters. It resembles force-directed graph energies and kernel particle flows such as Stein variational gradient descent (SVGD), but contributes (i) an explicit codelength linkage via Φ_b, (ii) a preregistered model card with declared parameters and calibration, and (iii) built-in falsifiers tied to out-of-sample compression. Unlike entropic-gravity or holographic approaches, PCD makes no physical claims beyond algorithmic optimization.

Acknowledgments

We thank colleagues for discussions on local estimators, kernels, integrators, and numerical reproducibility. Earlier drafts exploring broader interpretations are superseded by this methodological formulation.

References

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.* (1948).

[2] J. Rissanen, "Modeling by shortest data description," *Automatica* (1978).

[3] L. A. Levin, "On the notion of a random sequence," *Sov. Math. Dokl.* (1971).

[4] S. Amari, *Information Geometry and Its Applications*, Springer (2016).

[5] R. Jordan, D. Kinderlehrer, F. Otto, "The variational formulation of the Fokker–Planck equation," *SIAM J. Math. Anal.* **29** (1998).

[6] A. Caticha, *Entropic Inference and the Foundations of Physics* (2012).

[7] H. Wendland, "Piecewise polynomial, positive definite and compactly supported radial functions," *Adv. Comput. Math.* **4** (1995) 389–396.

[8] B. Leimkuhler, M. Matthews, "Rational construction of stochastic numerical methods for molecular sampling," *Appl. Math. Res. eXpress* (2013).

[9] B. Leimkuhler, C. Matthews, *Molecular Dynamics*, Springer (2016).

[10] R. C. Prim, "Shortest connection networks and some generalizations," *Bell Syst. Tech. J.* **36**, 1389–1401 (1957).

[11] L. Hernquist, "An analytical model for spherical galaxies and bulges," *ApJ* **356**, 359–364 (1990).

[12] H. C. Plummer, "On the problem of distribution in globular star clusters," *MNRAS* **71**, 460–470 (1911).