Natural Attractive Interactions and the Contraction of Information Radius

Tsutomu Mori*

Department of Human Lifesciences,
Fukushima Medical University School of Nursing, Fukushima, Japan.
*Correspondence: tmori@fmu.ac.jp

Abstract

Mutual information $MI(X_t; Y_t)$ quantifies how knowledge of one subsystem reduces uncertainty about another. We study the time evolution of the information radius

$$r(t) = H(X_t) + H(Y_t) - 2\,MI(X_t; Y_t)$$

when the joint law of (X_t, Y_t) evolves under a continuity equation with velocity fields (v_x, v_y). Under standard regularity and no-flux assumptions, we derive variational identities for the time derivatives of the marginal entropies and mutual information, and combine them into a master identity for $\dot{r}(t)$ in terms of the velocity fields and the pointwise mutual information (PMI) $\phi = \log p - \log p_X - \log p_Y$.

For canonical Gibbsian models with density proportional to $\exp\{-\beta(U(x) + U(y) + W(x, y))\}$, we show that the PMI gradients coincide with centered interaction forces and are $L^2(p)$-orthogonal to all marginal-only drifts. This yields a natural velocity decomposition $v = \gamma\nabla\phi + u$. When the dynamics is driven purely by PMI gradients and preserves the marginals, we obtain an exact contraction theorem: the information radius $r(t)$ decays at a rate proportional to a PMI "Fisher energy". When marginal drifts are present, we derive a quantitative inequality showing that $r(t)$ still decreases whenever the PMI-gradient energy dominates the marginal-entropy production. Gaussian and coupled Ornstein–Uhlenbeck examples illustrate these mechanisms and provide explicit parameter conditions for contraction.

Keywords: mutual information; information radius; pointwise mutual information; canonical Gibbs distributions; contraction; stochastic dynamics

1 Introduction

We investigate the time evolution of the *information radius*

$$r(t) = H(X_t) + H(Y_t) - 2\,MI(X_t;Y_t) = H(X_t \mid Y_t) + H(Y_t \mid X_t), \tag{1.1}$$

for a pair of random variables (X_t, Y_t). Throughout, logarithms are natural and entropies are expressed in nats. Throughout we write $MI(X;Y)$ for the mutual information between X and Y, reserving the symbol $I(\cdot)$ for the self-information. In discrete settings $r(t)$ is nonnegative and coincides with the classical variation of information, whereas in continuous settings the differential entropies $H(X_t)$ and $H(Y_t)$ may take negative values, so $r(t)$ may fail to be nonnegative.[1] Section 2 elaborates on this distinction. Our focus is therefore on the *monotonic behavior* of $r(t)$ under smooth evolutions of the joint density, rather than on its sign.

Recent studies have demonstrated that when the velocity fields governing the evolution of the joint density $p(x, y, t)$ align with the gradients of the pointwise mutual information (PMI; variational identities appear in Section 2), the information radius decreases monotonically. This provides an information-geometric interpretation of flows that strengthen statistical dependence. Motivated by physical models, we observe that many natural attractive interactions—especially under canonical approximations—possess PMI-gradient components after subtracting marginal mean forces. Natural attractive interactions considered here arise from potentials of the form $U(x) + U(y) + W(x, y)$, whose formal treatment appears in Section 3. This connection suggests a direct route from physical forces to contraction mechanisms previously established for idealized mutual-information gradient flows.

Contributions. We examine conditions under which natural attractive interactions contract the information radius and present two complementary results. First, for *marginal-preserving* dynamics that align exactly with PMI gradients, we establish a strict contraction principle: $r(t)$ decreases monotonically and remains constant only at statistical independence. Second, for general natural interactions that introduce small but nonzero marginal drift, we derive an *approximate contraction* inequality, providing a quantitative bound whose magnitude depends on the size of the marginal drift. These results are supported by a canonical representation of natural forces via PMI gradients, together with analytically tractable examples.

Scope and assumptions. We work with smooth, strictly positive densities on domains with no-flux boundaries so that integration by parts introduces no boundary terms. Our derivations rely on the continuity equation governing $p(x, y, t)$ and on variational identities for the time derivatives of H and I. The theorems apply to both discrete and continuous models under these regularity assumptions; for continuous variables, the conclusions concern the monotonicity of $r(t)$ rather than its absolute sign. Technical details of the variational identities are collected in Appendix A.

Related work. Information radius has long been used to quantify statistical separation, particularly within information theory and clustering. Contraction properties for gradient flows are well studied in information geometry and in Wasserstein settings; the present work connects these ideas to natural interactions by exhibiting PMI-gradient components of physical forces. From a statistical-mechanical perspective, our PMI representation parallels the classical decomposition of interaction forces into fluctuating and marginal mean components.

[1] In continuous settings, $r(t)$ is not a metric in the strict sense, as nonnegativity need not hold; we retain the term "information radius" by analogy with the discrete case.

Organization. Section 2 introduces foundational notions and variational identities used throughout the manuscript. Section 3 develops the representation of natural interactions through pointwise mutual-information gradients and formulates a projection identity. Section 4 states and proves the main contraction theorems: an exact result for marginal-preserving PMI-gradient flows and an approximate result with a quantitative drift-dependent bound. Section 5 offers illustrative examples. Section 6 discusses broader implications and connections. Section 7 concludes. Appendix A collects the variational identities used in Section 2. Appendix B contains the canonical PMI-gradient structure and the orthogonality result underlying Section 3. Appendix C presents the Gaussian Ornstein–Uhlenbeck example in detail, including the covariance dynamics and information-radius checks used in Section 5.

2 Preliminaries

This section fixes notation, states the regularity and boundary assumptions, and collects the variational identities used throughout the paper. We use natural logarithms and measure information in nats. Classical background on entropy, mutual information, and related quantities can be found in standard information-theory texts such as [2, 3]. Throughout we write $MI(X;Y)$ for the mutual information between X and Y, reserving the symbol $I(\cdot)$ for the self-information.

2.1 Notation and assumptions

Let $\Omega_x \subseteq \mathbb{R}^{d_x}$ and $\Omega_y \subseteq \mathbb{R}^{d_y}$ denote domains for X_t and Y_t. The joint density $p(x,y,t)$ is assumed to be strictly positive and smooth, with

$$\partial_t p + \nabla_x\cdot(p\,v_x) + \nabla_y\cdot(p\,v_y) = 0, \qquad (x,y,t) \in \Omega_x \times \Omega_y \times (0,T), \tag{2.1}$$

for some sufficiently regular velocity fields $v_x : \Omega_x\times\Omega_y\times(0,T) \to \mathbb{R}^{d_x}$ and $v_y : \Omega_x\times\Omega_y\times(0,T) \to \mathbb{R}^{d_y}$. The marginals are

$$p_X(x,t) = \int_{\Omega_y} p(x,y,t)\,dy, \qquad p_Y(y,t) = \int_{\Omega_x} p(x,y,t)\,dx,$$

and the corresponding (differential) entropies are

$$H(X_t) = -\int_{\Omega_x} p_X(x,t)\,\log p_X(x,t)\,dx, \qquad H(Y_t) = -\int_{\Omega_y} p_Y(y,t)\,\log p_Y(y,t)\,dy.$$

Assumption 2.1 (Regularity and boundary conditions). The fields p, $\partial_t p$, v_x, and v_y are sufficiently smooth to justify differentiation under the integral and integration by parts. Either (i) the domains are bounded with *no-flux* boundary conditions $p\,v_x\cdot n_x = 0$ on $\partial\Omega_x$ and $p\,v_y\cdot n_y = 0$ on $\partial\Omega_y$, or (ii) $\Omega_x = \mathbb{R}^{d_x}$ and $\Omega_y = \mathbb{R}^{d_y}$ and $p(x,y,t)$ and the fluxes $p(x,y,t)v_x(x,y,t)$, $p(x,y,t)v_y(x,y,t)$ decay faster than $\|(x,y)\|^{-1-\epsilon}$ (or exponentially) to ensure that score fields and all divergences remain integrable. Under such decay, all boundary terms vanish in the integration-by-parts identities used below.

Remark 2.2 (Discrete vs. continuous). For discrete variables,

$$r(t) = H(X_t) + H(Y_t) - 2\,MI(X_t;Y_t) = H(X_t \mid Y_t) + H(Y_t \mid X_t), \tag{2.2}$$

is nonnegative (the classical variation of information). For continuous variables, differential entropies may be negative, so $r(t)$ may fail to be nonnegative; our results concern the *monotonicity* of $r(t)$ under suitable flows, not its sign (cf. footnote in Section 1).

Remark 2.3 (Unbounded domains). If Ω_x or Ω_y is unbounded, it suffices that $p(x, y, t)$ and the fluxes $p(x, y, t)v_x(x, y, t)$, $p(x, y, t)v_y(x, y, t)$ decay faster than $\|(x, y)\|^{-1-\epsilon}$ (or exponentially) to ensure that score fields and all divergences remain integrable. Under such decay, all boundary terms vanish in the integration-by-parts identities used below.

2.2 Pointwise mutual information and score fields

The pointwise mutual information (PMI) is

$$\phi(x, y, t) = \log p(x, y, t) - \log p_X(x, t) - \log p_Y(y, t). \tag{2.3}$$

We define marginal score fields

$$s_X(x, t) = \nabla_x \log p_X(x, t), \qquad s_Y(y, t) = \nabla_y \log p_Y(y, t). \tag{2.4}$$

Conditional expectations of the velocities give the marginal drifts:

$$u_x(x, t) = \mathbb{E}[v_x(X_t, Y_t) \mid X_t = x], \qquad u_y(y, t) = \mathbb{E}[v_y(X_t, Y_t) \mid Y_t = y]. \tag{2.5}$$

In particular,

$$p_X(x, t)\, u_x(x, t) = \int_{\Omega_y} p(x, y, t)\, v_x(x, y, t)\, dy, \qquad p_Y(y, t)\, u_y(y, t) = \int_{\Omega_x} p(x, y, t)\, v_y(x, y, t)\, dx.$$

2.3 Variational identities

Derivations are provided in Appendix A.

Lemma 2.1 (Mutual-information derivative)**.** *Under Assumption 2.1,*

$$\frac{d}{dt} MI(X_t; Y_t) = \iint_{\Omega_x \times \Omega_y} p(x, y, t) \Big(v_x \cdot \nabla_x \phi + v_y \cdot \nabla_y \phi \Big)\, dx\, dy. \tag{2.6}$$

This sign convention is consistent with the transport identity $\frac{d}{dt}\mathbb{E}[f] = \mathbb{E}[\partial_t f + v \cdot \nabla f]$*, noting that* MI *is an integral functional rather than an expectation of a fixed function. This is the same identity as equation* (A.6) *proved in Appendix A.*

Lemma 2.2 (Marginal-entropy derivatives)**.** *Under Assumption 2.1,*

$$\frac{d}{dt} H(X_t) = -\iint p\, v_x \cdot s_X\, dx\, dy, \qquad \frac{d}{dt} H(Y_t) = -\iint p\, v_y \cdot s_Y\, dx\, dy. \tag{2.7}$$

Equivalently, using $p_X u_x = \int p\, v_x\, dy$*,*

$$\frac{d}{dt} H(X_t) = -\int_{\Omega_x} p_X\, u_x \cdot s_X\, dx, \qquad \frac{d}{dt} H(Y_t) = -\int_{\Omega_y} p_Y\, u_y \cdot s_Y\, dy. \tag{2.8}$$

These identities coincide with the marginal entropy derivatives (A.9)–(A.10) proved in Appendix A.

Corollary 2.3 (Information-radius identity)**.**

$$\frac{d}{dt} r(t) = \frac{d}{dt} H(X_t) + \frac{d}{dt} H(Y_t) - 2\, \frac{d}{dt} MI(X_t; Y_t). \tag{2.9}$$

The representation (2.9) is just the abstract identity obtained by combining the variational formulas in Appendix A, see in particular (A.12).

Remark 2.4 (Notation consistency). (X_t, Y_t) denote random variables; (x, y) denote coordinates. Spatial gradients ∇_x, ∇_y act on (x, y) with t fixed. Boldface is avoided to reduce clutter.

3 Natural Attractive Interactions as Mutual-Information Gradient Components

In many physical systems, interactions between two variables arise from an underlying potential energy landscape. When the joint distribution of (X_t, Y_t) is close to a canonical form, the structure of the interaction potential reveals a natural connection between physical forces and gradients of pointwise mutual information (PMI). This section establishes this connection in a form compatible with the notation and assumptions fixed in Section 2 and provides a projection identity that will be used in Section 4.

Assumption 3.1 (Canonical structure). There exist single-variable potentials $U : \Omega_x \to \mathbb{R}$, $U : \Omega_y \to \mathbb{R}$ and an interaction potential $W : \Omega_x \times \Omega_y \to \mathbb{R}$ such that, for each t,

$$p(x, y, t) = \frac{1}{Z(t)} \exp\Big(-\beta \, [\, U(x) + U(y) + W(x, y) \,] \Big), \tag{3.1}$$

with $\beta > 0$ and normalizer $Z(t) > 0$. We assume U and W are sufficiently smooth so that all derivatives below exist and are integrable under p. We note that the symbol W denotes the interaction potential; this differs from thermodynamic notation, where W often denotes mechanical work. Gibbsian forms of this type arise naturally in statistical mechanics and maximum-entropy modeling; see, for example, Jaynes' discussion of canonical ensembles [9].

3.1 PMI gradients under canonical structure

Let

$$\phi(x, y, t) = \log p(x, y, t) - \log p_X(x, t) - \log p_Y(y, t) \tag{3.2}$$

denote the pointwise mutual information. Differentiating (3.1) and separating marginal terms yields the identities (see Appendix B, equations (B.14)–(B.15), for a complete derivation)

$$\nabla_x \phi = -\beta\big(\nabla_x W(x, y) - \mathbb{E}[\nabla_x W \mid x]\big), \qquad \nabla_y \phi = -\beta\big(\nabla_y W(x, y) - \mathbb{E}[\nabla_y W \mid y]\big). \tag{3.3}$$

Thus the PMI gradients coincide with *centered* interaction forces. Writing $F_x(x, y) := -\nabla_x W(x, y)$ and $F_y(x, y) := -\nabla_y W(x, y)$, (3.3) becomes

$$\frac{1}{\beta} \nabla_x \phi = F_x - \mathbb{E}[F_x \mid x], \qquad \frac{1}{\beta} \nabla_y \phi = F_y - \mathbb{E}[F_y \mid y]. \tag{3.4}$$

Remark (Mean-field interpretation). The quantity $\mathbb{E}[\nabla_x W \mid x]$ may be interpreted as the *mean-field force* exerted by Y on X. The fluctuation term $\nabla_x W - \mathbb{E}[\nabla_x W \mid x]$ represents deviations from this mean field and corresponds exactly to the PMI gradient direction.

3.2 Projection identity and orthogonality

Proposition 3.1 (Force decomposition)**.** *Under Assumption 3.1, the interaction force decomposes into an information-gradient component and a marginal mean component:*

$$F_x = \frac{1}{\beta} \nabla_x \phi + \mathbb{E}[F_x \mid x], \qquad F_y = \frac{1}{\beta} \nabla_y \phi + \mathbb{E}[F_y \mid y]. \tag{3.5}$$

Moreover, the information-gradient component is orthogonal in $L^2(p)$ to all functions depending solely on the corresponding marginal coordinate: for every measurable $g : \Omega_x \to \mathbb{R}^{d_x}$ with $\int \|g(x)\|^2 p_X(x,t)\,dx < \infty$,

$$\iint p(x,y,t)\left(\tfrac{1}{\beta}\,\nabla_x \phi(x,y,t)\right)\cdot g(x)\,dx\,dy = 0, \tag{3.6}$$

and analogously in the y-variable.

This orthogonal decomposition is recorded in full generality as Proposition B.1 in Appendix B.

Sketch of proof. Using the tower property and (3.4),

$$\begin{aligned}\iint p\left(\tfrac{1}{\beta}\nabla_x\phi\right)\cdot g(x)\,dx\,dy &= \int p_X(x,t)\,\mathbb{E}\Big[\left(\tfrac{1}{\beta}\nabla_x\phi\right)\,\Big|\,x\Big]\cdot g(x)\,dx \\ &= \int p_X(x,t)\left(\mathbb{E}[F_x \mid x] - \mathbb{E}[F_x \mid x]\right)\cdot g(x)\,dx = 0.\end{aligned}$$

The y-case is identical. □

3.3 Interface with the dynamical formulation

Section 2 expresses the time evolution of the joint density via the continuity equation with velocity fields (v_x, v_y). In overdamped settings, it is natural to take velocities proportional to forces (after non-dimensionalization), so that the decomposition (3.5) induces a corresponding split of the velocity fields:

$$v_x = \gamma\,\nabla_x\phi + u_x, \qquad v_y = \gamma\,\nabla_y\phi + u_y, \tag{3.7}$$

where $\gamma > 0$ is a proportionality constant and $u_x = \mathbb{E}[v_x \mid x]$, $u_y = \mathbb{E}[v_y \mid y]$ are the marginal drifts introduced in Section 2. (In physical Langevin systems, γ corresponds to a mobility–temperature factor such as $\gamma \sim \mu k_B T$, though no specific physical scaling is assumed here.)

Equation (3.7) is the kinematic analogue of the force decomposition (3.5) and is the starting point for the contraction results in Section 4.

Remark 3.2 (Attractive interactions). We use the term "attractive" to reflect the qualitative behavior that the interaction potential W tends to increase statistical dependence between X_t and Y_t. The precise contraction statements are given in Section 4; no additional convexity or curvature condition on W is imposed here beyond Assumption 3.1.

Remark 3.3 (Consistency with Section 2). All gradients are taken with respect to spatial coordinates (x, y); conditional expectations such as $\mathbb{E}[\cdot \mid x]$ are defined with respect to the conditional law $p(y \mid x, t)$, consistent with Section 2. The use of the marginal drifts u_x, u_y in (3.7) matches the notation introduced in Lemma 2.2.

4 Main Results: Contraction Theorems

This section establishes contraction principles for the information radius under the assumptions and notation of Sections 2–3. We use the variational identities of Section 2 (see Lemma 2.1, Lemma 2.2, and Corollary 2.3 proved in Appendix A) and the velocity-field decomposition

$$v_x = \gamma\,\nabla_x\phi + u_x, \qquad v_y = \gamma\,\nabla_y\phi + u_y, \tag{4.1}$$

where $\gamma > 0$ and $u_x = \mathbb{E}[v_x \mid x]$, $u_y = \mathbb{E}[v_y \mid y]$ are the marginal drifts. Recall also that, by Proposition 3.1, the PMI gradients are orthogonal to all x-only and y-only functions in $L^2(p)$.

For convenience we write

$$\mathcal{I}_\phi(t) := \iint_{\Omega_x \times \Omega_y} p(x,y,t)\Big(\|\nabla_x \phi(x,y,t)\|^2 + \|\nabla_y \phi(x,y,t)\|^2\Big)\,dx\,dy \tag{4.2}$$

for the "PMI Fisher energy" at time t.

4.1 Exact contraction for marginal-preserving PMI-gradient flows

We first consider the case where the dynamics is driven purely by PMI gradients and preserves the marginals.

Theorem 4.1 (Exact contraction for PMI-gradient flows)**.** *Suppose Assumption 2.1 holds and that the velocity fields are* pure PMI-gradient *with preserved marginals, in the sense that*

$$v_x = \gamma\,\nabla_x \phi, \qquad v_y = \gamma\,\nabla_y \phi, \qquad \textit{and} \qquad u_x = u_y = 0. \tag{4.3}$$

Then the information radius $r(t)$ satisfies

$$\frac{d}{dt} r(t) = -2\gamma\,\mathcal{I}_\phi(t) = -2\gamma \iint p\Big(\|\nabla_x\phi\|^2 + \|\nabla_y \phi\|^2\Big)\,dx\,dy \;\le\; 0. \tag{4.4}$$

Moreover, if equality holds at some time t, then X_t and Y_t are statistically independent.

Proof. From Corollary 2.3 and Lemma 2.2,

$$\begin{aligned}\frac{d}{dt} r(t) = &-\iint p\,v_x \cdot s_X\,dx\,dy \;-\; \iint p\,v_y \cdot s_Y\,dx\,dy \\ &-\;2\iint p\Big(v_x \cdot \nabla_x \phi + v_y \cdot \nabla_y \phi\Big)\,dx\,dy.\end{aligned}$$

Substituting $v_x = \gamma \nabla_x \phi$, $v_y = \gamma \nabla_y \phi$ and using $\mathbb{E}[\nabla_x \phi \mid x] = \mathbb{E}[\nabla_y \phi \mid y] = 0$, we obtain

$$\begin{aligned}\iint p\,v_x \cdot s_X\,dx\,dy &= \gamma \int p_X(x,t)\,s_X(x,t) \cdot \mathbb{E}[\nabla_x \phi \mid x]\,dx = 0, \\ \iint p\,v_y \cdot s_Y\,dx\,dy &= \gamma \int p_Y(y,t)\,s_Y(y,t) \cdot \mathbb{E}[\nabla_y \phi \mid y]\,dy = 0,\end{aligned}$$

while

$$\iint p\Big(v_x \cdot \nabla_x \phi + v_y \cdot \nabla_y \phi\Big)\,dx\,dy = \gamma \iint p\Big(\|\nabla_x \phi\|^2 + \|\nabla_y \phi\|^2\Big)\,dx\,dy.$$

Equation (4.4) follows.

If equality holds at some time t, then $\nabla_x \phi = \nabla_y \phi = 0$ p-a.e., so ϕ is constant in (x,y). Since $p = e^{\phi} p_X p_Y$ and $\iint p\,dx\,dy = 1 = \iint p_X p_Y\,dx\,dy$, it follows that $e^{\phi} = 1$ and hence $p(x,y,t) = p_X(x,t)p_Y(y,t)$. Thus X_t and Y_t are independent. □

4.2 Approximate contraction with small marginal drifts

We now allow nonzero marginal drifts u_x, u_y and derive an estimate showing that contraction persists provided the PMI-gradient component dominates the drifts in a suitable sense.

Define the marginal score norms

$$C_X(t) := \|s_X(\cdot,t)\|_{L^2(p_X)} = \Big(\int_{\Omega_x} \|s_X(x,t)\|^2 p_X(x,t)\,dx\Big)^{1/2}, \tag{4.5}$$

and similarly

$$C_Y(t) := \|s_Y(\cdot,t)\|_{L^2(p_Y)} = \Big(\int_{\Omega_y} \|s_Y(y,t)\|^2 p_Y(y,t)\,dy\Big)^{1/2}, \tag{4.6}$$

and the marginal drift norms

$$\|u_x(\cdot,t)\|_{L^2(p_X)} = \Big(\int_{\Omega_x} \|u_x(x,t)\|^2 p_X(x,t)\,dx\Big)^{1/2}, \qquad \|u_y(\cdot,t)\|_{L^2(p_Y)} = \Big(\int_{\Omega_y} \|u_y(y,t)\|^2 p_Y(y,t)\,dy\Big)^{1/2}. \tag{4.7}$$

Theorem 4.2 (Approximate contraction with small drifts)**.** *Suppose Assumption 2.1 holds and that the velocity fields admit the decomposition* (4.1)*. Then, for each time t,*

$$\frac{d}{dt}r(t) \le -2\gamma \iint p\left(\|\nabla_x\phi\|^2 + \|\nabla_y\phi\|^2\right) dx\,dy \;+\; C_X(t)\,\|u_x(\cdot,t)\|_{L^2(p_X)} \;+\; C_Y(t)\,\|u_y(\cdot,t)\|_{L^2(p_Y)}. \tag{4.8}$$

In particular, if

$$C_X(t)\,\|u_x(\cdot,t)\|_{L^2(p_X)} + C_Y(t)\,\|u_y(\cdot,t)\|_{L^2(p_Y)} \;\le\; 2\gamma \iint p\left(\|\nabla_x\phi\|^2 + \|\nabla_y\phi\|^2\right) dx\,dy, \tag{4.9}$$

then $\frac{d}{dt}r(t) \le 0$ at that time.

Proof. Starting again from Corollary 2.3 and substituting (4.1) yields

$$\begin{aligned}\frac{d}{dt}r(t) = &-\iint p\,(\gamma\nabla_x\phi + u_x)\cdot s_X\,dx\,dy - \iint p\,(\gamma\nabla_y\phi + u_y)\cdot s_Y\,dx\,dy \\ &- 2\iint p\Big((\gamma\nabla_x\phi + u_x)\cdot\nabla_x\phi + (\gamma\nabla_y\phi + u_y)\cdot\nabla_y\phi\Big)\,dx\,dy.\end{aligned}$$

Using $\mathbb{E}[\nabla_x\phi \mid x] = \mathbb{E}[\nabla_y\phi \mid y] = 0$ and that u_x, u_y depend only on the marginal coordinates, we have

$$\begin{aligned}\iint p\,\gamma\nabla_x\phi\cdot s_X\,dx\,dy &= \gamma\int p_X(x,t)\,s_X(x,t)\cdot\mathbb{E}[\nabla_x\phi \mid x]\,dx = 0, \\ \iint p\,\gamma\nabla_y\phi\cdot s_Y\,dx\,dy &= \gamma\int p_Y(y,t)\,s_Y(y,t)\cdot\mathbb{E}[\nabla_y\phi \mid y]\,dy = 0.\end{aligned}$$

Thus

$$\begin{aligned}\frac{d}{dt}r(t) = &-\iint p\,u_x\cdot s_X\,dx\,dy - \iint p\,u_y\cdot s_Y\,dx\,dy \\ &- 2\iint p\Big(\gamma\|\nabla_x\phi\|^2 + \gamma\|\nabla_y\phi\|^2 + u_x\cdot\nabla_x\phi + u_y\cdot\nabla_y\phi\Big)\,dx\,dy.\end{aligned}$$

The first two terms are bounded by Cauchy–Schwarz:

$$\begin{aligned}\left|\iint p\,u_x\cdot s_X\,dx\,dy\right| &= \left|\int p_X(x,t)\,u_x(x,t)\cdot s_X(x,t)\,dx\right| \le C_X(t)\,\|u_x(\cdot,t)\|_{L^2(p_X)}, \\ \left|\iint p\,u_y\cdot s_Y\,dx\,dy\right| &\le C_Y(t)\,\|u_y(\cdot,t)\|_{L^2(p_Y)}.\end{aligned}$$

For the remaining terms, we use the inequality $-2u_x \cdot \nabla_x \phi \leq \|u_x\|^2 + \|\nabla_x \phi\|^2$, and similarly for u_y, to absorb the cross terms into the positive part of the drift contribution. Collecting these estimates gives

$$\frac{d}{dt} r(t) \leq -2\gamma \iint p \left(\|\nabla_x \phi\|^2 + \|\nabla_y \phi\|^2 \right) dx\, dy \\ + C_X(t) \, \|u_x(\cdot, t)\|_{L^2(p_X)} + C_Y(t) \, \|u_y(\cdot, t)\|_{L^2(p_Y)},$$

which is (4.8). The small-drift condition (4.9) then implies $\frac{d}{dt} r(t) \leq 0$. □

Remarks. (1) The score norms $C_X(t), C_Y(t)$ equal the square roots of the marginal Fisher informations and can be controlled in specific models via functional inequalities (e.g., Poincaré or logarithmic Sobolev), though such bounds are not needed for the statements above. (2) Under the marginal-preserving condition $u_x = u_y = 0$, Theorem 4.2 reduces to Theorem 4.1. (3) Integrating (4.4) or (4.8) in time yields cumulative contraction estimates over intervals. (4) Condition (4.9) indicates that contraction remains robust against small marginal drifts when the system exhibits strong interactions, i.e., when the PMI gradient $\nabla \phi$ is large.

5 Examples and Illustrations

This section provides analytically tractable examples that instantiate the contraction principles proved in Section 4. We first recall the velocity-field decomposition $v = \gamma \nabla \phi + u$ and then study a coupled Gaussian (Ornstein–Uhlenbeck) model, which serves as a canonical testbed connecting the PMI-gradient structure to concrete dynamics. Throughout, we maintain the notation of Sections 2–3.

5.1 A didactic check: exact PMI-gradient flow on a Gaussian family

Fix a bivariate Gaussian law with zero mean, common variance $\sigma^2(t)$, and correlation $\rho(t)$. If one drives the probability flow by the *pure PMI-gradient* velocity fields $v_x = \gamma \nabla_x \phi$ and $v_y = \gamma \nabla_y \phi$ (hence $u_x = u_y = 0$), Theorem 4.1 applies verbatim and yields the exact contraction

$$\frac{d}{dt} r(t) = -2\gamma \iint p \left(\|\nabla_x \phi\|^2 + \|\nabla_y \phi\|^2 \right) dx\, dy \leq 0, \tag{5.1}$$

with equality if and only if $\rho(t) = 0$ (statistical independence).

5.2 Coupled Gaussian Ornstein–Uhlenbeck system

Consider the linear SDE system

$$dX_t = \big(-a\, X_t - \kappa (X_t - Y_t) \big)\, dt + \sqrt{2D}\, dB_t^x, \tag{5.2}$$

$$dY_t = \big(-a\, Y_t - \kappa (Y_t - X_t) \big)\, dt + \sqrt{2D}\, dB_t^y, \tag{5.3}$$

with $a, \kappa, D > 0$ and independent Brownian motions. The joint law of (X_t, Y_t) stays Gaussian with zero mean and covariance $\Sigma_t = \begin{pmatrix} \sigma_t^2 & \rho_t \sigma_t^2 \\ \rho_t \sigma_t^2 & \sigma_t^2 \end{pmatrix}$.

This is the classical coupled Ornstein–Uhlenbeck model; see, for example, [10, 11, 12, 13]. A detailed analysis of the induced covariance and correlation dynamics, together with the behavior of the information radius $r(t)$ for this model, is carried out in Appendix C.

PMI gradients and marginal drifts. A direct computation (see Appendix C for details) gives

$$\iint p(\|\nabla_x \phi\|^2 + \|\nabla_y \phi\|^2)\, dx\, dy = \frac{2\rho_t^2}{\sigma_t^2(1-\rho_t^2)}.$$

The marginal drifts and scores satisfy

$$u_x = -(a + \kappa(1-\rho_t))x, \quad s_X = -x/\sigma_t^2, \quad C_X = 1/\sigma_t,$$

with analogous expressions for Y.

Approximate contraction bound. Substituting these expressions into Theorem 4.2 gives:

$$\frac{d}{dt} r(t) \leq -\frac{4\gamma\rho_t^2}{\sigma_t^2(1-\rho_t^2)} + 2|a + \kappa(1-\rho_t)|. \tag{5.4}$$

Remark (On the parameter γ in the OU example). In the physical OU system, the decomposition $v_x = \gamma \nabla_x \phi + u_x$ holds with a specific γ determined by the coupling constant κ and the Gaussian covariance; thus γ is not arbitrary in this model, although its exact value is irrelevant for the validity of the contraction inequality.

5.3 Summary

These Gaussian examples validate both the exact and approximate contraction mechanisms. They show that the PMI-gradient component of natural interactions governs the reduction of information radius, in line with Section 4.

6 Discussion

This section synthesizes the main conceptual messages of the paper, explains the scope and implications of the contraction principles proved in Section 4, and outlines limitations and avenues for future work. Throughout, we adhere to the notation fixed in Sections 2–3.

6.1 Conceptual synthesis

Our central observation is that, under canonical structure, the *pointwise mutual-information (PMI) gradients* isolate the *interaction component* of the physical force: cf. the centered representation

$$\frac{1}{\beta}\nabla_x \phi = F_x - \mathbb{E}[F_x \mid x], \qquad \frac{1}{\beta}\nabla_y \phi = F_y - \mathbb{E}[F_y \mid y], \tag{6.1}$$

which is a restatement of (3.4). The orthogonality structure (Proposition 3.1) implies that the PMI directions are $L^2(p)$-orthogonal to all marginal-only vector fields, hence they capture exactly the *pairwise* fluctuations responsible for changing statistical dependence. This observation ties the physics (through F_x, F_y) to information geometry (through ϕ) and enables the contraction results of Section 4. These conclusions rest on the variational identities collected in Appendix A and on the canonical PMI-gradient structure and orthogonality properties established in Appendix B.

6.2 Robustness and the small-drift condition

The exact contraction theorem (Theorem 4.1) shows that when the velocity fields coincide with PMI gradients and preserve marginals ($u_x = u_y = 0$), the information radius $r(t)$ obeys a dissipation law (4.4). In realistic models the marginals rarely remain strictly fixed; Theorem 4.2 quantifies precisely how much *marginal drift* can be tolerated while preserving monotone decrease:

$$\frac{d}{dt}r(t) \;\le\; -2\gamma \iint p\,\|\nabla\phi\|^2 + C_X(t)\,\|u_x\|_{L^2(p_X)} + C_Y(t)\,\|u_y\|_{L^2(p_Y)}. \tag{6.2}$$

The small-drift condition (4.9) expresses a simple balance: contraction persists when the *interaction strength* (PMI-gradient energy) dominates the *marginal-entropy production* induced by u_x, u_y. This matches the Gaussian OU analysis in Section 5, where the region of contraction can be read directly from explicit formulae.

6.3 Relations to thermodynamics and gradient-flow formalisms

The dissipation identity for $r(t)$ resembles entropy production laws in non-equilibrium thermodynamics and the "Fisher-information" dissipation encountered in Wasserstein gradient flows; see, for example, [14, 15, 16, 17, 18, 19]. Here the driving term is not a free-energy gradient but the PMI gradient, which measures pairwise dependence creation. The Fisher-like constants C_X, C_Y appearing in Theorem 4.2 underscore the role of marginal geometry and suggest that *functional inequalities* (Poincaré, log-Sobolev) could supply explicit contraction rates when u_x, u_y satisfy suitable regularity properties. Related information-geometric viewpoints on thermodynamic fluctuation theory and biological systems can be found in [20, 21].

6.4 Assumptions and limitations

Our analysis assumes smooth strictly positive densities, no-flux (or fast decay) boundary conditions, and—in the representation step—a canonical decomposition with sufficiently regular potentials. In continuous settings the quantity $r(t)$ is not a metric in the strict sense (nonnegativity can fail), and our conclusions concern monotonicity rather than absolute sign. The velocity decomposition uses a single scalar mobility γ; more general *anisotropic or state-dependent* mobilities could be treated with bookkeeping effort, but are beyond the scope of this paper. Finally, the PMI representation isolates *pairwise* interactions; higher-order interactions would require an extension of the formalism to multivariate PMI structures.

6.5 Extensions and open directions

- **Sharp constants.** Combining Theorem 4.2 with model-specific functional inequalities may yield explicit drift thresholds ensuring contraction, and possibly convergence rates.
- **Anisotropic mobilities.** Replacing γ by matrices or state-dependent mobilities should preserve the orthogonality arguments, but requires careful norm control.
- **Time-dependent or non-conservative interactions.** Extending Proposition 3.1 to time-dependent $W(x, y, t)$ (or beyond potential forces) is an open problem with practical relevance.
- **Higher-order structure.** Generalizing the decomposition to triplets (X, Y, Z) would connect PMI gradients to multi-information dynamics and networked interactions.

- **Applications.** The PMI-based viewpoint suggests connections to dependence shaping in statistical physics, causal inference under interventions, and representation learning where contrastive objectives implicitly target PMI gradients. Related ideas within the broader Influential Force Theory (IFT) framework, including applications to immuno-oncology, are developed in the author's earlier work [22].

7 Conclusion

We have investigated the time evolution of the information radius

$$r(t) = H(X_t) + H(Y_t) - 2\,MI(X_t; Y_t),$$

between two random variables X_t and Y_t evolving under a continuity equation with velocity fields (v_x, v_y). Our perspective combines variational identities, a canonical representation of interactions, and velocity-field decompositions to reveal how *pointwise mutual-information gradients* govern the contraction of $r(t)$.

We conclude by summarizing the main components of the analysis.

- **Variational identities (Section 2).** Under standard regularity and boundary assumptions, we expressed time derivatives of $H(X_t)$, $H(Y_t)$, and $MI(X_t; Y_t)$ in terms of the velocity fields (v_x, v_y) (Lemmas 2.1–2.2) and obtained the master identity (2.9), as proved in detail in Appendix A.
- **PMI representation (Section 3).** For canonical models we identified PMI gradients with centered interaction forces (Eq. (3.4)) and established an $L^2(p)$ orthogonality property (Proposition 3.1); these canonical PMI-gradient formulas and their orthogonality are developed systematically in Appendix B.
- **Contraction theorems (Section 4).** Using the velocity decomposition $v = \gamma\nabla\phi + u$, Theorem 4.1 shows exact contraction for marginal-preserving PMI-gradient flows, while Theorem 4.2 gives a quantitative bound that is robust to small marginal drifts.
- **Gaussian validation (Section 5).** Closed-form computations for Gaussian families and a coupled Ornstein–Uhlenbeck system verify the contraction mechanisms and provide interpretable thresholds in terms of correlation, variance, and drift; the detailed covariance dynamics and information-radius checks are collected in Appendix C.

Taken together, these elements show that under natural structural assumptions the PMI-gradient component of the dynamics provides a geometrically meaningful and physically interpretable mechanism for contracting the information radius between interacting subsystems. They also suggest a range of extensions, from sharper quantitative bounds based on functional inequalities to higher-order interaction structures and applications in statistical physics, information processing, and learning, as discussed in Section 6. A more detailed development of the mutual-information-gradient contraction mechanism, including further examples and applications, will be presented in a companion manuscript [23].

Appendix A: Variational Derivatives and Information Identities

In this appendix we collect the basic variational identities for the joint law (X_t, Y_t) that are used in Section 2. Throughout we assume the regularity and no–flux conditions stated in Assumption 2.1 of the main text: the joint density $p(x, y, t)$ is smooth in (x, y, t), decays sufficiently fast (or satisfies reflecting boundary conditions), and the probability current has vanishing normal component at the boundary so that integrations by parts produce no boundary terms.

Let $p(x, y, t)$ denote the joint density of (X_t, Y_t). The continuity equation reads

$$\partial_t p + \nabla_x \cdot (p\, v_x) + \nabla_y \cdot (p\, v_y) = 0, \tag{A.1}$$

with velocity fields $v_x : \Omega_x \times \Omega_y \times (0, T) \to \mathbb{R}^{d_x}$ and $v_y : \Omega_x \times \Omega_y \times (0, T) \to \mathbb{R}^{d_y}$. We write $p_X(x, t) = \int p(x, y, t)\, dy$ and $p_Y(y, t) = \int p(x, y, t)\, dx$ for the marginals, and

$$\phi(x, y, t) = \log p(x, y, t) - \log p_X(x, t) - \log p_Y(y, t)$$

for the pointwise mutual information.

A.1 Time derivative of mutual information

The mutual information $MI(X_t; Y_t)$ is

$$MI(X_t; Y_t) = \iint p(x, y, t)\, \phi(x, y, t)\, dx\, dy.$$

Differentiating under the integral sign yields

$$\frac{d}{dt} MI(X_t; Y_t) = \iint \partial_t p(x, y, t)\, \phi(x, y, t)\, dx\, dy + \iint p(x, y, t)\, \partial_t \phi(x, y, t)\, dx\, dy. \tag{A.2}$$

We first show that the second integral vanishes. Since

$$\phi = \log p - \log p_X - \log p_Y,$$

we have

$$\partial_t \phi = \partial_t \log p - \partial_t \log p_X - \partial_t \log p_Y.$$

We treat the three contributions separately.

First,

$$\iint p\, \partial_t \log p\, dx\, dy = \iint \partial_t p\, dx\, dy = \partial_t \iint p\, dx\, dy = \partial_t 1 = 0.$$

Next, using $p_X(x, t) = \int p(x, y, t)\, dy$,

$$\iint p\, \partial_t \log p_X\, dx\, dy = \int p_X(x, t)\, \partial_t \log p_X(x, t)\, dx = \int \partial_t p_X(x, t)\, dx = \partial_t \int p_X(x, t)\, dx = 0.$$

An analogous computation yields

$$\iint p\, \partial_t \log p_Y\, dx\, dy = 0.$$

Hence

$$\iint p(x, y, t)\, \partial_t \phi(x, y, t)\, dx\, dy = 0, \tag{A.3}$$

and (A.2) reduces to

$$\frac{d}{dt}MI(X_t;Y_t) = \iint \partial_t p(x,y,t)\,\phi(x,y,t)\,dx\,dy. \tag{A.4}$$

Using the continuity equation (A.1),

$$\begin{aligned}\iint \partial_t p\,\phi\,dx\,dy &= -\iint \nabla_x\cdot(p\,v_x)\,\phi\,dx\,dy - \iint \nabla_y\cdot(p\,v_y)\,\phi\,dx\,dy \\ &= \iint p(x,y,t)\,v_x(x,y,t)\cdot\nabla_x\phi(x,y,t)\,dx\,dy \\ &\quad + \iint p(x,y,t)\,v_y(x,y,t)\cdot\nabla_y\phi(x,y,t)\,dx\,dy,\end{aligned}$$

where we have integrated by parts and used the no-flux boundary conditions to discard boundary terms. Thus we obtain

$$\begin{aligned}\frac{d}{dt}MI(X_t;Y_t) = &\iint p(x,y,t)\,v_x(x,y,t)\cdot\nabla_x\phi(x,y,t)\,dx\,dy \\ &+ \iint p(x,y,t)\,v_y(x,y,t)\cdot\nabla_y\phi(x,y,t)\,dx\,dy.\end{aligned}$$

This can be written more compactly as

$$\frac{d}{dt}MI(X_t;Y_t) = \iint p\,v_x\cdot\nabla_x\phi\,dx\,dy + \iint p\,v_y\cdot\nabla_y\phi\,dx\,dy. \tag{A.5}$$

Thus we arrive at the basic variational identity

$$\frac{d}{dt}MI(X_t;Y_t) = \iint p\,v_x\cdot\nabla_x\phi\,dx\,dy + \iint p\,v_y\cdot\nabla_y\phi\,dx\,dy. \tag{A.6}$$

This identity coincides with the mutual information derivative used in Lemma 2.1 of Section 2.

A.2 Time derivative of the marginal entropies

We next derive variational formulas for the marginal entropies

$$H(X_t) = -\int p_X(x,t)\,\log p_X(x,t)\,dx, \qquad H(Y_t) = -\int p_Y(y,t)\,\log p_Y(y,t)\,dy.$$

We first consider $H(X_t)$. Differentiating under the integral sign yields

$$\frac{d}{dt}H(X_t) = -\int \partial_t p_X(x,t)\big(1+\log p_X(x,t)\big)\,dx. \tag{A.7}$$

Integrating the continuity equation (A.1) with respect to y we obtain

$$\partial_t p_X(x,t) + \nabla_x\cdot\Big(\int p(x,y,t)\,v_x(x,y,t)\,dy\Big) = 0.$$

It is convenient to write this in terms of a conditional expectation. We define

$$\mathbb{E}[v_x \mid x] = \frac{1}{p_X(x,t)}\int p(x,y,t)\,v_x(x,y,t)\,dy,$$

so that

$$\int p(x,y,t)\,v_x(x,y,t)\,dy = p_X(x,t)\,\mathbb{E}[v_x \mid x].$$

Substituting into the marginal continuity equation gives

$$\partial_t p_X(x,t) + \nabla_x \cdot \big(p_X(x,t)\, \mathbb{E}[v_x \mid x]\big) = 0. \tag{A.8}$$

Inserting (A.8) into (A.7) yields

$$\begin{aligned} \frac{d}{dt} H(X_t) &= \int \nabla_x \cdot \big(p_X\, \mathbb{E}[v_x \mid x]\big)\big(1 + \log p_X\big)\, dx \\ &= -\int p_X(x,t)\, \mathbb{E}[v_x \mid x] \cdot \nabla_x \log p_X(x,t)\, dx, \end{aligned}$$

where the last line follows from integration by parts and decay (or no-flux) assumptions. Using

$$\int p(x,y,t)\, v_x(x,y,t)\, dy = p_X(x,t)\, \mathbb{E}[v_x \mid x],$$

we can equivalently write this as

$$\frac{d}{dt} H(X_t) = -\iint p(x,y,t)\, v_x(x,y,t) \cdot \nabla_x \log p_X(x,t)\, dx\, dy. \tag{A.9}$$

An entirely analogous computation yields

$$\frac{d}{dt} H(Y_t) = -\iint p(x,y,t)\, v_y(x,y,t) \cdot \nabla_y \log p_Y(y,t)\, dx\, dy. \tag{A.10}$$

Taken together, the identities (A.9) and (A.10) coincide with the marginal entropy derivative formulas appearing as Lemma 2.2 in Section 2.

A.3 Derivative of the information radius

Finally we combine the preceding identities to obtain a variational formula for the information radius

$$r(t) = H(X_t) + H(Y_t) - 2\, MI(X_t; Y_t).$$

Differentiating,

$$\frac{d}{dt} r(t) = \frac{d}{dt} H(X_t) + \frac{d}{dt} H(Y_t) - 2\, \frac{d}{dt} MI(X_t; Y_t). \tag{A.11}$$

Substituting (A.6), (A.9), and (A.10), we obtain

$$\begin{aligned} \frac{d}{dt} r(t) &= \frac{d}{dt} H(X_t) + \frac{d}{dt} H(Y_t) - 2\, \frac{d}{dt} MI(X_t; Y_t) \\ &= -\iint p\, v_x \cdot \nabla_x \log p_X\, dx\, dy - \iint p\, v_y \cdot \nabla_y \log p_Y\, dx\, dy \\ &\quad - 2 \iint p\, v_x \cdot \nabla_x \phi\, dx\, dy - 2 \iint p\, v_y \cdot \nabla_y \phi\, dx\, dy. \end{aligned}$$

Equivalently,

$$\begin{aligned} \frac{d}{dt} r(t) = &-\iint p(x,y,t)\, v_x(x,y,t) \cdot \Big(\nabla_x \log p_X(x,t) + 2\, \nabla_x \phi(x,y,t)\Big)\, dx\, dy \\ &-\iint p(x,y,t)\, v_y(x,y,t) \cdot \Big(\nabla_y \log p_Y(y,t) + 2\, \nabla_y \phi(x,y,t)\Big)\, dx\, dy. \end{aligned} \tag{A.12}$$

This is exactly the time–derivative identity for the information radius used in Corollary 2.3 of Section 2.

Remarks on regularity and boundary conditions

For clarity we summarize the analytical assumptions that justify the above manipulations. We assume that $p(x, y, t)$ and the velocity fields $v_x(x, y, t)$, $v_y(x, y, t)$ are smooth and that p decays rapidly at infinity, or that the dynamics are confined to a bounded domain with reflecting boundary conditions. In either case the normal component of the probability current vanishes at the boundary so that all integrations by parts in x and y produce no boundary terms. These conditions are precisely those stated abstractly in Assumption 2.1 of the main text.

Appendix B: Canonical Distributions and PMI Gradient Identities

This appendix provides a complete derivation of the gradient identities used in Section 3 to connect natural attractive interactions with gradients of pointwise mutual information (PMI). Throughout we work under the canonical Gibbsian assumption stated in Assumption 3.1 of the main text: the joint density of (X, Y) is given by a Boltzmann–Gibbs distribution with interaction potential W.

B.1 Canonical density and notation

Let the joint density of (X, Y) be given by

$$p(x, y) = \frac{1}{Z} \exp\Big(-\beta \left[U(x) + U(y) + W(x, y) \right] \Big), \tag{B.1}$$

with $\beta > 0$ (inverse-temperature parameter), normalizing constant $Z > 0$, single-variable potentials $U : \Omega_x \to \mathbb{R}$ and $U : \Omega_y \to \mathbb{R}$, and interaction potential $W : \Omega_x \times \Omega_y \to \mathbb{R}$.

The marginals of p are

$$p_X(x) = \int p(x, y)\, dy, \qquad p_Y(y) = \int p(x, y)\, dx. \tag{B.2}$$

We denote by

$$\phi(x, y) = \log p(x, y) - \log p_X(x) - \log p_Y(y) \tag{B.3}$$

the pointwise mutual information.

B.2 Gradients of the canonical density

Differentiating $\log p(x, y)$ with respect to x yields

$$\nabla_x \log p(x, y) = -\beta \nabla_x U(x) - \beta \nabla_x W(x, y). \tag{B.4}$$

Similarly, the marginal density of X can be written as

$$\begin{aligned} p_X(x) &= \int p(x, y)\, dy & \text{(B.5)} \\ &= \frac{1}{Z} \exp\big(-\beta U(x) \big) \int \exp\big(-\beta [U(y) + W(x, y)] \big)\, dy. & \text{(B.6)} \end{aligned}$$

Taking logarithms,

$$\log p_X(x) = -\beta U(x) + \log \left(\int \exp\big(-\beta [U(y) + W(x, y)] \big) dy \right) - \log Z, \tag{B.7}$$

and therefore,

$$\begin{aligned} \nabla_x \log p_X(x) &= -\beta \nabla_x U(x) + \frac{\displaystyle\int \big(-\beta \nabla_x W(x, y) \big)\, \exp\big(-\beta [U(y) + W(x, y)] \big) dy}{\displaystyle\int \exp\big(-\beta [U(y) + W(x, y)] \big) dy} & \text{(B.8)} \\ &= -\beta \nabla_x U(x) - \beta\, \mathbb{E}[\nabla_x W(x, Y) \mid x]. & \text{(B.9)} \end{aligned}$$

Similarly,

$$\nabla_y \log p(x, y) = -\beta \nabla_y U(y) - \beta \nabla_y W(x, y), \tag{B.10}$$

and

$$\nabla_y \log p_Y(y) = -\beta \nabla_y U(y) - \beta\, \mathbb{E}[\nabla_y W(X, y) \mid y]. \tag{B.11}$$

B.3 PMI gradients in canonical form

By definition

$$\phi(x,y) = \log p(x,y) - \log p_X(x) - \log p_Y(y),$$

so differentiating with respect to x yields

$$\begin{aligned}
\nabla_x \phi(x,y) &= \nabla_x \log p(x,y) - \nabla_x \log p_X(x) && \text{(B.12)}\\
&= \big(- \beta \nabla_x U(x) - \beta \nabla_x W(x,y)\big) - \big(- \beta \nabla_x U(x) - \beta\, \mathbb{E}[\nabla_x W(x,Y) \mid x]\big) && \text{(B.13)}\\
&= -\beta\big(\nabla_x W(x,y) - \mathbb{E}[\nabla_x W(x,Y) \mid x]\big). && \text{(B.14)}
\end{aligned}$$

An entirely analogous computation with respect to y gives

$$\nabla_y \phi(x,y) = -\beta\big(\nabla_y W(x,y) - \mathbb{E}[\nabla_y W(X,y) \mid y]\big). \tag{B.15}$$

These formulas show that the PMI gradients coincide with *centered* interaction forces under the canonical Gibbsian form (B.1).

B.4 Orthogonality of the canonical decomposition

For later use in Section 3 it is convenient to record an orthogonality property of the canonical decomposition. Combining (B.14) with the definition of the conditional expectation, we can rewrite the interaction force as

$$\nabla_x W(x,y) = -\frac{1}{\beta}\, \nabla_x \phi(x,y) + \mathbb{E}[\nabla_x W(x,Y) \mid x]. \tag{B.16}$$

We view the two terms on the right-hand side as elements of the Hilbert space $L^2(p)$ with inner product

$$\langle f, g\rangle_{L^2(p)} = \iint f(x,y) \cdot g(x,y)\, p(x,y)\, dx\, dy.$$

Proposition B.1 (Orthogonality of the canonical PMI decomposition)**.** *Under the canonical assumption* (B.1)*, the decomposition* (B.16) *is orthogonal in* $L^2(p)$*, i.e.*

$$\big\langle - \tfrac{1}{\beta}\, \nabla_x \phi,\ \mathbb{E}[\nabla_x W(x,Y) \mid x]\big\rangle_{L^2(p)} = 0.$$

An analogous orthogonal decomposition holds for $\nabla_y W(x,y)$*.*

Proof. Using the definition of the inner product and the law of total expectation, we compute

$$\big\langle \nabla_x \phi,\ \mathbb{E}[\nabla_x W(x,Y) \mid x]\big\rangle_{L^2(p)} = \mathbb{E}\Big[\nabla_x \phi(X,Y) \cdot \mathbb{E}[\nabla_x W(X,Y) \mid X]\Big].$$

Conditioning on X and using the fact that

$$\nabla_x \phi(X,Y) = -\beta\big(\nabla_x W(X,Y) - \mathbb{E}[\nabla_x W(X,Y) \mid X]\big)$$

is centered in Y for each fixed X, we obtain

$$\mathbb{E}[\nabla_x \phi(X,Y) \mid X] = 0.$$

Hence

$$\big\langle \nabla_x \phi,\ \mathbb{E}[\nabla_x W(x,Y) \mid x]\big\rangle_{L^2(p)} = \mathbb{E}\Big[\mathbb{E}[\nabla_x \phi(X,Y) \mid X] \cdot \mathbb{E}[\nabla_x W(X,Y) \mid X]\Big] = 0,$$

which yields the claimed orthogonality. The statement for $\nabla_y W$ follows in exactly the same way from (B.15). □

This orthogonality property is the content of the projection identity stated as Proposition 3.1 in Section 3: the PMI gradient $-\beta^{-1}\nabla_x \phi$ captures the "fluctuating" part of the interaction force, while the conditional expectation $\mathbb{E}[\nabla_x W \mid x]$ represents the mean-field component.

B.5 Interpretation

Equations (B.14)–(B.15) establish the gradient identities used in Section 3, in particular the formulas (3.3)–(3.4) for the canonical PMI gradients.

$$\nabla_x \phi = -\beta(\nabla_x W - \mathbb{E}[\nabla_x W \mid x]), \qquad \nabla_y \phi = -\beta(\nabla_y W - \mathbb{E}[\nabla_y W \mid y]).$$

Thus the gradient of the pointwise mutual information corresponds to the *centered* interaction force, obtained by subtracting its conditional expectation. Together with Proposition B.1, this provides the canonical PMI decomposition and its orthogonality property used in the projection identity of Section 3 and in the contraction results of Section 4.

Appendix C: Gaussian Example: Detailed Calculations

This appendix provides the full derivations underlying the Gaussian example presented in Section 5. We consider a pair of coupled Ornstein–Uhlenbeck processes and compute explicitly the evolution of their covariance structure, the correlation coefficient, and the information radius.

C.1 Coupled Ornstein–Uhlenbeck dynamics

Let (X_t, Y_t) satisfy the linear stochastic differential equations

$$dX_t = -aX_t\,dt - \kappa\,(X_t - Y_t)\,dt + \sqrt{2D}\,dB_t^x, \tag{C.1}$$

$$dY_t = -aY_t\,dt - \kappa\,(Y_t - X_t)\,dt + \sqrt{2D}\,dB_t^y, \tag{C.2}$$

where $a, \kappa, D > 0$ and B_t^x, B_t^y are independent standard Brownian motions. The joint law of (X_t, Y_t) is Gaussian with zero mean and covariance matrix

$$\Sigma_t = \begin{pmatrix} \sigma_t^2 & \rho_t\sigma_t^2 \\ \rho_t\sigma_t^2 & \sigma_t^2 \end{pmatrix},$$

where $\sigma_t^2 = \mathbb{E}[X_t^2] = \mathbb{E}[Y_t^2]$ and $\rho_t = \mathbb{E}[X_tY_t]/\sigma_t^2$ is the correlation coefficient.

Applying Itô's lemma to X_t^2, Y_t^2, and X_tY_t yields ODEs for σ_t^2 and $c_t = \mathbb{E}[X_tY_t]$. A standard computation gives

$$\frac{d}{dt}\sigma_t^2 = -2(a+\kappa)\,\sigma_t^2 + 2\kappa\,c_t + 2D, \tag{C.3}$$

$$\frac{d}{dt}c_t = -2a\,c_t - 2\kappa\,(c_t - \sigma_t^2) + 2D. \tag{C.4}$$

In terms of (σ_t^2, ρ_t), with $c_t = \rho_t\sigma_t^2$, these become

$$\frac{d}{dt}\sigma_t^2 = -2(a+\kappa)\,\sigma_t^2 + 2\kappa\,\rho_t\sigma_t^2 + 2D, \tag{C.5}$$

$$\frac{d}{dt}(\rho_t\sigma_t^2) = -2a\,\rho_t\sigma_t^2 - 2\kappa\,(\rho_t\sigma_t^2 - \sigma_t^2) + 2D. \tag{C.6}$$

Expanding (C.6) and rearranging yields

$$\rho_t\,\frac{d}{dt}\sigma_t^2 + \sigma_t^2\,\frac{d}{dt}\rho_t = -2a\,\rho_t\sigma_t^2 - 2\kappa\,(\rho_t\sigma_t^2 - \sigma_t^2) + 2D. \tag{C.7}$$

Using (C.5) to replace $\frac{d}{dt}\sigma_t^2$ leads to an ODE for ρ_t:

$$\sigma_t^2\,\frac{d}{dt}\rho_t = -2D\,\rho_t + 2\kappa\,(1-\rho_t)\sigma_t^2. \tag{C.8}$$

Equivalently,

$$\frac{d}{dt}\rho_t = -2\,\frac{D}{\sigma_t^2}\,\rho_t + 2\kappa\,(1-\rho_t). \tag{C.9}$$

Thus the correlation dynamics are governed by a balance between a decay term proportional to D/σ_t^2 and an attractive term proportional to κ.

C.2 Quasi-stationary approximation

In many parameter regimes the variance σ_t^2 relaxes rapidly toward its stationary value, while the correlation ρ_t evolves on a slower time scale. In this regime it is natural to consider a *quasi-stationary* approximation in which σ_t^2 is replaced by a constant σ^2. Under this approximation, (C.5) suggests

$$0 \approx -2(a+\kappa)\,\sigma^2 + 2\kappa\,\rho_t\sigma^2 + 2D,$$

so that

$$D \approx (a + \kappa - \kappa\rho_t)\,\sigma^2.$$

Substituting into (C.9) gives the approximate ODE

$$\frac{d}{dt}\rho_t \approx -2(a+\kappa-\kappa\rho_t)\,\rho_t + 2\kappa\,(1-\rho_t) = -2a\,\rho_t + 2\kappa\,(1-\rho_t). \tag{C.10}$$

In particular, for $\kappa > 0$ and $a > 0$ we see that ρ_t increases monotonically toward

$$\rho_\infty = \frac{\kappa}{a+\kappa}.$$

C.3 Information radius and its derivative

For a bivariate Gaussian with covariance matrix Σ_t as above, the information radius is

$$r(t) = \log\bigl(2\pi e\,\sigma_t^2(1-\rho_t^2)\bigr). \tag{C.11}$$

Differentiating (C.11),

$$\frac{d}{dt}r(t) = \frac{d}{dt}\Bigl[\log\sigma_t^2 + \log(1-\rho_t^2)\Bigr] \tag{C.12}$$

$$= \frac{\dot{\sigma}_t^2}{\sigma_t^2} + \frac{-2\rho_t\,\dot{\rho}_t}{1-\rho_t^2}. \tag{C.13}$$

In the quasi-stationary regime we approximate $\dot{\sigma}_t^2 \approx 0$, so that

$$\frac{d}{dt}r(t) \approx \frac{-2\rho_t\,\dot{\rho}_t}{1-\rho_t^2}. \tag{C.14}$$

Using (C.10), this becomes

$$\frac{d}{dt}r(t) \approx \frac{-2\rho_t}{1-\rho_t^2}\,\bigl(-2a\,\rho_t + 2\kappa\,(1-\rho_t)\bigr). \tag{C.15}$$

For $\rho_t \in [0,1)$ and $\kappa > 0$, the factor

$$-2a\,\rho_t + 2\kappa\,(1-\rho_t) = 2\bigl(\kappa - a\rho_t - \kappa\rho_t\bigr)$$

is positive for a range of parameters, so the dominant prefactor $-2\rho_t/(1-\rho_t^2) \le 0$ drives $\frac{d}{dt}r(t)$ negative. More precisely, if

$$-2a\,\rho_t + 2\kappa\,(1-\rho_t) \ge 0$$

for ρ_t in the range of interest, then

$$\frac{d}{dt}r(t) \le 0 \qquad \text{whenever } \rho_t \ge 0.$$

Thus attractive coupling induces a reduction in the information radius.

C.4 Summary

The Gaussian example provides an explicit instance in which natural attractive interactions drive monotone changes in correlation and thereby reduce the information radius. It illustrates concretely how the contraction phenomena described in Section 4 manifest in a solvable dynamical system.

References

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27:379–423, 623–656, 1948. https://ieeexplore.ieee.org/document/6773024

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, 2006. https://onlinelibrary.wiley.com/doi/book/10.1002/047174882X

[3] R. M. Gray, *Entropy and Information Theory*, 2nd ed., Springer, 2011. https://link.springer.com/book/10.1007/978-1-4419-7970-4

[4] I. Csiszár and P. C. Shields, "Information Theory and Statistics: A Tutorial," *Foundations and Trends in Communications and Information Theory*, 1, 2004. https://www.nowpublishers.com/article/Details/CIT-004

[5] S. Amari and H. Nagaoka, *Methods of Information Geometry*, AMS, 2000. https://bookstore.ams.org/surv-191

[6] S. Amari, *Information Geometry and Its Applications*, Springer, 2016. https://link.springer.com/book/10.1007/978-4-431-55978-8

[7] C. Villani, *Topics in Optimal Transportation*, AMS, 2003. https://bookstore.ams.org/gsm-58

[8] P. Harremoës and F. Topsøe, "Inequalities between Entropy and Divergence," *IEEE Trans. Inf. Theory*, 47(4):1506–1516, 2001. https://doi.org/10.1109/18.923723

[9] E. T. Jaynes, "Information Theory and Statistical Mechanics," *Phys. Rev.*, 106:620–630, 1957. https://doi.org/10.1103/PhysRev.106.620

[10] G. E. Uhlenbeck and L. S. Ornstein, "On the Theory of the Brownian Motion," *Phys. Rev.*, 36:823–841, 1930. https://doi.org/10.1103/PhysRev.36.823

[11] J. L. Doob, "The Brownian Movement and Stochastic Equations," *Ann. Math.*, 43(2):351–369, 1942. https://doi.org/10.2307/1968873

[12] H. Risken, *The Fokker–Planck Equation: Methods of Solution and Applications*, 2nd ed., Springer, 1996. https://link.springer.com/book/10.1007/978-3-642-61544-3

[13] C. W. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed., Springer, 2009. https://link.springer.com/book/10.1007/978-3-540-70713-2

[14] C. Jarzynski, "Nonequilibrium Equality for Free Energy Differences," *Phys. Rev. Lett.*, 78:2690–2693, 1997. https://doi.org/10.1103/PhysRevLett.78.2690

[15] G. E. Crooks, "Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences," *Phys. Rev. E*, 60:2721–2726, 1999. https://doi.org/10.1103/PhysRevE.60.2721

[16] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, "Thermodynamics of Information," *Nature Phys.*, 11:131–139, 2015. https://doi.org/10.1038/nphys3230

[17] T. Sagawa and M. Ueda, "Second Law of Thermodynamics with Discrete Quantum Feedback Control," *Phys. Rev. Lett.*, 100:080403, 2008. https://doi.org/10.1103/PhysRevLett.100.080403

[18] J. M. Horowitz and M. Esposito, "Thermodynamics with Continuous Information Flow," *Phys. Rev. X*, 4:031015, 2014. https://doi.org/10.1103/PhysRevX.4.031015

[19] S. Still, D. A. Sivak, A. J. Bell, and G. E. Crooks, "Thermodynamics of Prediction," *Phys. Rev. Lett.*, 109:120604, 2012. https://doi.org/10.1103/PhysRevLett.109.120604

[20] G. Ruppeiner, "Riemannian Geometry in Thermodynamic Fluctuation Theory," *Rev. Mod. Phys.*, 67:605–659, 1995. https://doi.org/10.1103/RevModPhys.67.605

[21] W. Bialek, *Biophysics: Searching for Principles*, Princeton Univ. Press, 2012. https://press.princeton.edu/books/hardcover/9780691154561/biophysics

[22] T. Mori, "Influential Force: From Higgs to the Novel Immune Checkpoint KYNU," Jxiv preprint, 2022. https://doi.org/10.51094/jxiv.156

[23] T. Mori, "Mutual Information Gradient Induces Contraction of Information Radius," Jxiv preprint, submitted 2025.