Lipschitz Stability of the Experiential Density Functional

Bryan Ehrlich

March 2026

Abstract

The experiential density $\rho(P) = I_\pi(B;M)\big(1 - I_\pi(B;M)/H_\pi(B)\big)$ measures meaningful complexity in finite-state Markov chains whose state space decomposes as $\Omega = B \times M$. We prove that ρ is Lipschitz continuous under kernel perturbations, with an explicit constant $L = (C_I + C_H)/\operatorname{gap}(P)$ depending on the state-space dimension and spectral gap. The proof composes Cho–Meyer perturbation theory for stationary distributions, Fannes–Audenaert entropy continuity, and the multivariate mean value theorem applied to the product functional $f(I,H) = I(1 - I/H)$. We also state a non-linear (sub-additive) bound that is sharper for small perturbations. Numerical verification against 3000 random kernel perturbations of a 16-state observer chain confirms $L_{\text{numerical}} \leq L_{\text{proven}}$ with zero violations and a tightness gap of 450–9500×, indicating the bound is conservative but valid. The proven bound's $1/\operatorname{gap}$ dependence is approximately consistent with numerical observations ($R^2 = 0.97$, fitted exponent -0.89).

1 Introduction

1.1 The experiential density

Consider a finite Markov chain on a composite state space $\Omega = B \times M$, where B represents the "body" (environment or physical degrees of freedom) and M represents the "model" (internal or representational degrees of freedom). This decomposition arises naturally in the experiential measure framework [11], which studies how information-processing systems maintain internal models of their environment through dynamical coupling.

Let P be an irreducible row-stochastic matrix on Ω with unique stationary distribution π. Reshaping π as a joint distribution $p(b,m)$ on $B \times M$, the mutual information $I_\pi(B;M) = H_\pi(B) + H_\pi(M) - H_\pi(B,M)$ quantifies the statistical coupling between body and model in the stationary state, and $H_\pi(B)$ is the marginal entropy of the body.

The *experiential density* is defined as

$$\rho(P) = I_\pi(B;M)\left(1 - \frac{I_\pi(B;M)}{H_\pi(B)}\right). \tag{1}$$

This functional vanishes at two extremes: when $I = 0$ (body and model are independent, no representation exists) and when $I = H(B)$ (the model is a perfect copy of the body, leaving no room for further experience). It is maximized at $I = H(B)/2$, where the system balances knowledge against uncertainty—a regime of maximal "meaningful complexity."

The experiential density belongs to the LMC (Lopez-Ruiz–Mancini–Calbet) statistical complexity family [1], which measures complexity as the product of information content and disequilibrium. It is one member of a broader family of complexity measures that share this multiplicative structure; the Lipschitz stability techniques developed here (composing stationary-distribution perturbation bounds with entropy continuity) could be applied to derive analogous results for other members of the LMC family. More broadly, the experiential density instantiates the effective complexity principle of Gell-Mann and Lloyd [2]: meaningful structure peaks between pure order and pure randomness.

1.2 Why stability matters

A well-posed functional should exhibit continuous dependence on its parameters. If ρ were discontinuous in the transition kernel P—if an arbitrarily small change in dynamics could produce a finite jump in experiential density—then ρ would be poorly behaved as a dynamical functional. Lipschitz continuity is the strongest standard form of continuity: it provides a linear bound on the output perturbation in terms of the input perturbation, with an explicit constant.

Stability results are also prerequisite for any perturbation theory. If we wish to study how ρ responds to environmental noise, parameter drift, or approximate computation, we need quantitative control over $|\rho(P) - \rho(P')|$ in terms of $\|P - P'\|$.

1.3 Contributions of this paper

We prove that ρ is Lipschitz continuous under kernel perturbations (Theorem 3), with:

1. A **non-linear bound** (sharpest form): $|\rho(P) - \rho(P')| \leq \frac{\delta}{2}[2\ln(|B| - 1) + \ln(|M| - 1) + \ln(n - 1)] + 4\, h_{\text{bin}}(\delta/2)$, where $\delta = \|\pi - \pi'\|_1 \leq \|P - P'\|_\infty / \operatorname{gap}(P)$.
2. A **linear Lipschitz bound**: $|\rho(P) - \rho(P')| \leq L(\delta) \cdot \|P - P'\|_\infty$ with $L = (C_I + C_H)/\operatorname{gap}(P)$.
3. **Explicit dependence** on state-space dimension ($\sim \ln|\Omega|$) and spectral gap ($\sim 1/\operatorname{gap}$), with no dependence on a lower bound for $H_\pi(B)$.

The proof composes three standard results: Cho–Meyer perturbation theory [3] (kernel $\to$ stationary distribution), Fannes–Audenaert entropy continuity [6] (distribution $\to$ information quantities), and the multivariate mean value theorem (information quantities $\to \rho$). Each step yields an explicit constant; their composition produces the final bound.

Using the mean value theorem rather than a direct product-rule decomposition yields $L = (C_I + C_H)/\operatorname{gap}$ instead of the weaker $(2C_I + C_H)/\operatorname{gap}$, because $|\partial f/\partial I| \leq 1$ on the feasible domain (not ≤ 2 as a naive product rule suggests).

Section 8 reports numerical verification against 3000 random kernel perturbations. The proven bound is never violated, confirming correctness. The tightness gap (ratio $L_{\text{proven}}/L_{\text{numerical}}$) ranges from 450× to 9500×, indicating that while the bound is conservative, tighter results would require fundamentally different techniques (e.g., direct differentiation of ρ as a function of P).

2 Setup and Theorem Statement

Definition 1 (Experiential density)**.** *Let P be an irreducible row-stochastic matrix on $\Omega = B \times M$ with stationary distribution π (the unique row vector satisfying $\pi P = \pi$, $\pi_i \geq 0$, $\sum_i \pi_i = 1$). Reshape π as a joint distribution $p(b, m)$ on $B \times M$. The* experiential density *is*

$$\rho(P) = I_\pi(B; M)\left(1 - \frac{I_\pi(B; M)}{H_\pi(B)}\right), \tag{2}$$

where $I_\pi(B; M) = H_\pi(B) + H_\pi(M) - H_\pi(B, M)$ is the mutual information and $H_\pi(B) = -\sum_b p_B(b) \ln p_B(b)$ is the marginal entropy on B, all in nats.

Definition 2 (Spectral gap and norms)**.** *For an irreducible stochastic matrix P on $n = |\Omega|$ states:*

- *The* absolute spectral gap *is* $\operatorname{gap}(P) = 1 - |\lambda_2(P)|$*, where λ_2 is the eigenvalue of P with second-largest absolute value.*

- *The* row-sum norm *is* $\|P\|_\infty = \max_i \sum_j |P_{ij}|$. *For the difference of stochastic matrices,* $\|P - P'\|_\infty = \max_i \sum_j |P_{ij} - P'_{ij}| = \max_i \|P_i - P'_i\|_1$.
- *We use the* ℓ^1 *norm throughout:* $\|p - q\|_1 = \sum_i |p_i - q_i|$. *The total variation distance is* $\|p - q\|_{\mathrm{TV}} = \frac{1}{2}\|p - q\|_1$.

Theorem 3 (Lipschitz stability of ρ)**.** *Let* P, P' *be irreducible row-stochastic matrices on* $\Omega = B \times M$ *with* $|B| \geq 2$, $|M| \geq 2$, $|\Omega| = n$. *Let* π, π' *be their stationary distributions, and suppose* $\mathrm{gap}(P) > 0$.

Non-linear bound (sharpest form). *Define* $\delta = \|\pi - \pi'\|_1$. *Then:*

$$|\rho(P) - \rho(P')| \leq \frac{\delta}{2}\big[2\ln(|B| - 1) + \ln(|M| - 1) + \ln(n - 1)\big] + 4\, h_{\mathrm{bin}}\left(\frac{\delta}{2}\right), \tag{3}$$

where $h_{\mathrm{bin}}(x) = -x \ln x - (1 - x)\ln(1 - x)$ *and* $\delta \leq \|P - P'\|_\infty / \mathrm{gap}(P)$ *by the Cho–Meyer bound.*

Linear Lipschitz bound. *Equivalently,*

$$|\rho(P) - \rho(P')| \leq L(\delta) \cdot \|P - P'\|_\infty, \tag{4}$$

where

$$L(\delta) = \frac{C_I(\delta) + C_H(\delta)}{\mathrm{gap}(P)}, \tag{5}$$

with

$$C_I(\delta) = \frac{1}{2}\big[\ln(|B| - 1) + \ln(|M| - 1) + \ln(n - 1)\big] + 3\,\frac{h_{\mathrm{bin}}(\delta/2)}{\delta}, \tag{6}$$

$$C_H(\delta) = \frac{1}{2}\ln(|B| - 1) + \frac{h_{\mathrm{bin}}(\delta/2)}{\delta}. \tag{7}$$

For small perturbations ($\delta \ll 1$), $h_{\mathrm{bin}}(\delta/2)/\delta \sim \frac{1}{2}\ln(2/\delta)$, *so the leading behavior is:*

$$L \sim \frac{1}{2\ \mathrm{gap}(P)}\big[2\ln(|B| - 1) + \ln(|M| - 1) + \ln(n - 1)\big] + \frac{2\ln(2/\delta)}{\mathrm{gap}(P)}. \tag{8}$$

Uniform non-linear bound. *Substituting* $\delta \leq \|P - P'\|_\infty / \mathrm{gap}(P)$ *into* (3) *and using* $h_{\mathrm{bin}}(x) \leq \ln 2$ *for all* x*:*

$$|\rho(P) - \rho(P')| \leq \frac{\|P - P'\|_\infty}{2\ \mathrm{gap}(P)}\big[2\ln(|B| - 1) + \ln(|M| - 1) + \ln(n - 1)\big] + 4\ln 2. \tag{9}$$

Domain of validity:

1. P *irreducible with* $\mathrm{gap}(P) > 0$*;*
2. P' *irreducible (so* π' *exists and is unique);*
3. $|B| \geq 2$ *(otherwise* $H_\pi(B) = 0$ *and* $\rho \equiv 0$*).*

No lower bound on $H_\pi(B)$ *is required in the Lipschitz constant.*

The proof proceeds in three steps, each invoking a standard result.

3 Step 1: Stationary Distribution Perturbation

Lemma 4 (Cho–Meyer perturbation bound [3])**.** *Let P, P' be irreducible stochastic matrices on n states with stationary distributions π, π'. Then*

$$\|\pi - \pi'\|_1 \leq \frac{1}{\operatorname{gap}(P)} \|P - P'\|_\infty. \tag{10}$$

Proof sketch (following Cho–Meyer [3]). Let $\Pi = \mathbf{1}\pi$ be the rank-one stationary projector. From $\pi P = \pi$ and $\pi' P' = \pi'$:

$$\pi' - \pi = \pi'(P' - P)(I - P + \Pi)^{-1},$$

where $(I - P + \Pi)^{-1}$ is the group inverse of $(I - P)$. The key estimate is $\|(I - P + \Pi)^{-1}\|_\infty \leq 1/\operatorname{gap}(P)$ (Meyer [4]; Cho–Meyer [3], Theorem 3.1). Since $\|\pi'\|_1 = 1$:

$$\|\pi' - \pi\|_1 \leq \|\pi'\|_1 \cdot \|P' - P\|_\infty \cdot \|(I - P + \Pi)^{-1}\|_\infty \leq \frac{\|P - P'\|_\infty}{\operatorname{gap}(P)}.$$

Here we used $\|vA\|_1 \leq \|v\|_1 \cdot \|A\|_\infty$ for a row vector v and matrix A (standard submultiplicativity for the ℓ^1-ℓ^∞ pair). □

Remark 5. *The bound* (10) *may be loose for specific chains. The sharper Cho–Meyer condition number $\kappa(P) = \|\pi\|_\infty \cdot \|(I - P + \Pi)^{-1}\|_\infty$ can be much smaller than $1/\operatorname{gap}(P)$. We use the simpler bound because it yields clean dependence on the spectral gap.*

Interface to Step 2. Define $\delta = \|\pi - \pi'\|_1$. Reshaping π and π' as joint distributions p and q on $B \times M$ (with $p(b, m) = \pi_{(b,m)}$), we have $\|p - q\|_1 = \delta$. By Lemma 4:

$$\delta \leq \frac{\|P - P'\|_\infty}{\operatorname{gap}(P)}. \tag{11}$$

4 Step 2: Mutual Information and Entropy Continuity

4.1 Entropy continuity lemma

Lemma 6 (Fannes–Audenaert inequality [5, 6])**.** *Let p, q be probability distributions on an alphabet of size $k \geq 2$ with $\|p - q\|_1 = \delta \in (0, 2(1 - 1/k)]$. Then*

$$|H(p) - H(q)| \leq \frac{\delta}{2} \ln(k - 1) + h_{\text{bin}}\left(\frac{\delta}{2}\right), \tag{12}$$

where $h_{\text{bin}}(x) = -x \ln x - (1 - x) \ln(1 - x)$ in nats.

Remark 7 (Convention)**.** *Many references state this bound with $\delta_{\text{TV}} = \|p - q\|_{\text{TV}} = \frac{1}{2}\|p - q\|_1$. In that convention, the bound reads $|H(p) - H(q)| \leq \delta_{\text{TV}} \ln(k - 1) + h_{\text{bin}}(\delta_{\text{TV}})$. Our formula* (12) *is identical after substituting $\delta_{\text{TV}} = \delta/2$. The tight constant is due to Audenaert [6].*

4.2 Marginal contraction

Lemma 8 (Marginal contraction)**.** *Let p, q be joint distributions on $B \times M$. Then*

$$\|p_B - q_B\|_1 \leq \|p - q\|_1, \qquad \|p_M - q_M\|_1 \leq \|p - q\|_1, \tag{13}$$

where $p_B(b) = \sum_m p(b, m)$ and $p_M(m) = \sum_b p(b, m)$.

Proof. By the triangle inequality: $\|p_B - q_B\|_1 = \sum_b |p_B(b) - q_B(b)| = \sum_b \left|\sum_m (p(b, m) - q(b, m))\right| \leq \sum_b \sum_m |p(b, m) - q(b, m)| = \|p - q\|_1$. □

4.3 Mutual information continuity bound

Using $I(B;M) = H(B) + H(M) - H(B,M)$ and the triangle inequality:

$$|I_p(B;M) - I_q(B;M)| \leq |H_p(B) - H_q(B)| + |H_p(M) - H_q(M)| + |H_p(B,M) - H_q(B,M)|. \quad (14)$$

Apply Lemma 6 to each term. The marginal contraction (Lemma 8) gives $\|p_B - q_B\|_1 \leq \delta$ and $\|p_M - q_M\|_1 \leq \delta$, where $\delta = \|p - q\|_1$.

- $H(B)$: alphabet size $|B|$, perturbation $\leq \delta$: $|H_p(B) - H_q(B)| \leq \frac{\delta}{2}\ln(|B|-1) + h_{\text{bin}}(\delta/2)$.
- $H(M)$: alphabet size $|M|$, perturbation $\leq \delta$: $|H_p(M) - H_q(M)| \leq \frac{\delta}{2}\ln(|M|-1) + h_{\text{bin}}(\delta/2)$.
- $H(B,M)$: alphabet size $n = |B| \cdot |M|$, perturbation $= \delta$: $|H_p(B,M) - H_q(B,M)| \leq \frac{\delta}{2}\ln(n-1) + h_{\text{bin}}(\delta/2)$.

Summing:

$$|I_p(B;M) - I_q(B;M)| \leq \frac{\delta}{2}\big[\ln(|B|-1) + \ln(|M|-1) + \ln(n-1)\big] + 3\,h_{\text{bin}}\left(\frac{\delta}{2}\right). \quad (15)$$

4.4 Separate bound on $H(B)$

For the composition in Step 3, we also need:

$$|H_p(B) - H_q(B)| \leq \frac{\delta}{2}\ln(|B|-1) + h_{\text{bin}}\left(\frac{\delta}{2}\right). \quad (16)$$

Remark 9. *An alternative approach to bounding $|I_p - I_q|$ would use conditional entropy continuity results, such as the tight bounds of Winter [13]. Since mutual information equals $H(B) - H(B|M)$, a direct continuity bound on the conditional entropy could potentially yield tighter constants by avoiding the three-term triangle inequality decomposition* (14)*. We leave this direction as future work.*

5 Step 3: Composition via the Mean Value Theorem

Write $\rho = f(I,H)$ where $f(I,H) = I(1 - I/H)$, with $I = I_\pi(B;M) \geq 0$, $H = H_\pi(B) > 0$, and $0 \leq I \leq H$.

5.1 Gradient bound

The partial derivatives of f on the domain $\mathcal{D} = \{(I,H) : 0 \leq I \leq H,\ H > 0\}$ are:

$$\frac{\partial f}{\partial I} = 1 - \frac{2I}{H}, \qquad \text{so } \left|\frac{\partial f}{\partial I}\right| \leq 1 \quad (\text{since } I/H \in [0,1]); \quad (17)$$

$$\frac{\partial f}{\partial H} = \frac{I^2}{H^2}, \qquad \text{so } \left|\frac{\partial f}{\partial H}\right| \leq 1 \quad (\text{since } I/H \in [0,1]). \quad (18)$$

The domain $\mathcal{D}$ is convex: for $(I,H), (I',H') \in \mathcal{D}$ and $t \in [0,1]$, the convex combination satisfies $\bar{H} = (1-t)H + tH' > 0$ and $\bar{I} = (1-t)I + tI' \leq (1-t)H + tH' = \bar{H}$, so $(\bar{I},\bar{H}) \in \mathcal{D}$.

By the multivariate mean value theorem, there exists $(\bar{I},\bar{H})$ on the line segment from (I,H) to (I',H') such that:

$$f(I,H) - f(I',H') = \left.\frac{\partial f}{\partial I}\right|_{(\bar{I},\bar{H})}(I - I') + \left.\frac{\partial f}{\partial H}\right|_{(\bar{I},\bar{H})}(H - H'). \quad (19)$$

Taking absolute values and applying the gradient bounds:

$$|f(I,H) - f(I',H')| \le \left|\frac{\partial f}{\partial I}\right| |I - I'| + \left|\frac{\partial f}{\partial H}\right| |H - H'| \le |I - I'| + |H - H'|. \tag{20}$$

Remark 10. *This bound is clean: no factors of H, H', or $h_{\min}$ appear. The key observation is that $f(I,H) = I(1 - I/H)$ satisfies $\|\nabla f\|_\infty \le 1$ everywhere on its feasible domain, because $I/H \in [0,1]$. A naive product-rule decomposition would give $|f - f'| \le |I - I'| \cdot (1 - I/H) + I' \cdot |I/H - I'/H'|$, which leads to the weaker constant $2C_I + C_H$ instead of $C_I + C_H$.*

5.2 Final composition

Combining (20) with the mutual information bound (15), the entropy bound (16), and the Cho–Meyer bound (11):

$$\begin{aligned}
|\rho(P) - \rho(P')| &\le |I - I'| + |H - H'| \\
&\le \left[\frac{\delta}{2}\big(\ln(|B| - 1) + \ln(|M| - 1) + \ln(n - 1)\big) + 3\, h_{\text{bin}}\left(\frac{\delta}{2}\right)\right] \\
&\quad + \left[\frac{\delta}{2}\ln(|B| - 1) + h_{\text{bin}}\left(\frac{\delta}{2}\right)\right] \\
&= \frac{\delta}{2}\big[2\ln(|B| - 1) + \ln(|M| - 1) + \ln(n - 1)\big] + 4\, h_{\text{bin}}\left(\frac{\delta}{2}\right),
\end{aligned} \tag{21}$$

with $\delta \le \|P - P'\|_\infty / \operatorname{gap}(P)$. This completes the proof of the non-linear bound (3).

Linear Lipschitz form. Writing $h_{\text{bin}}(\delta/2) \le \delta \cdot [h_{\text{bin}}(\delta/2)/\delta]$ and defining $\eta(\delta) = h_{\text{bin}}(\delta/2)/\delta$ (finite for any $\delta > 0$):

$$\begin{aligned}
|\rho(P) - \rho(P')| &\le \big[C_I(\delta) + C_H(\delta)\big]\, \delta \\
&\le \frac{C_I(\delta) + C_H(\delta)}{\operatorname{gap}(P)}\, \|P - P'\|_\infty,
\end{aligned} \tag{22}$$

where $C_I(\delta)$ and $C_H(\delta)$ are as in (6)–(7). This gives the Lipschitz constant:

$$\boxed{L(\delta) = \frac{C_I(\delta) + C_H(\delta)}{\operatorname{gap}(P)}\,.} \tag{23}$$

Remark 11 (On the δ-dependence of L)**.** *The Lipschitz "constant" $L(\delta)$ depends on δ through the h_{bin}/δ ratio, which grows as $\frac{1}{2}\ln(2/\delta)$ for small δ. This is not a deficiency of the proof: it reflects a genuine feature of entropy continuity. Shannon entropy is* not *globally Lipschitz on the probability simplex, because $|H(p) - H(q)|/\|p - q\|_1$ diverges logarithmically as p approaches a vertex of the simplex.*

For any fixed *perturbation size $\varepsilon = \|P - P'\|_\infty$, the constant $L(\varepsilon/\operatorname{gap}(P))$ is finite and explicit. For the non-linear bound* (3)*, no δ-dependent constant is needed: it is valid for all δ simultaneously.*

6 Explicit Formulas

For reference, we collect the main formulas. Write $n = |B| \cdot |M|$, $g = \operatorname{gap}(P)$, $\varepsilon = \|P - P'\|_\infty$, $\delta = \|\pi - \pi'\|_1 \le \varepsilon/g$.

Non-linear bound (recommended for evaluation):

$$|\rho(P) - \rho(P')| \le \frac{\varepsilon}{2\,g}\big[2\ln(|B|-1) + \ln(|M|-1) + \ln(n-1)\big] + 4\,h_{\text{bin}}\left(\frac{\varepsilon}{2\,g}\right). \tag{24}$$

Linear bound (with δ-dependent constant):

$$L(\delta) = \frac{1}{g}\left\{\frac{1}{2}\big[2\ln(|B|-1) + \ln(|M|-1) + \ln(n-1)\big] + 4\,\frac{h_{\text{bin}}(\delta/2)}{\delta}\right\}. \tag{25}$$

Asymptotic form ($\delta \to 0$):

$$L \sim \frac{1}{2\,g}\big[2\ln(|B|-1) + \ln(|M|-1) + \ln(n-1)\big] + \frac{2}{g}\ln\left(\frac{2\,g}{\varepsilon}\right). \tag{26}$$

Uniform additive bound (all perturbation sizes):

$$|\rho(P) - \rho(P')| \le \frac{\varepsilon}{2\,g}\big[2\ln(|B|-1) + \ln(|M|-1) + \ln(n-1)\big] + 4\ln 2, \tag{27}$$

using $h_{\text{bin}}(x) \le \ln 2$ for all $x \in [0,1]$.

7 Limiting Cases

We verify the bound's behavior in six regimes.

7.1 Zero perturbation ($P' = P$)

When $P' = P$: $\varepsilon = 0$, $\delta = 0$, and (3) gives $|\rho - \rho'| \le 0$, which is correct.

7.2 Fast mixing ($\text{gap}(P) \to 1$)

For $\text{gap}(P) = 1$ (e.g., a rank-one matrix $P = \mathbf{1}\pi$):

$$L_{\text{eff}} \sim \frac{1}{2}\big[2\ln(|B|-1) + \ln(|M|-1) + \ln(n-1)\big].$$

For the toy model ($|B| = |M| = 4$, $n = 16$): $L_{\text{eff}} \approx 3.00$ nats. Since $\rho_{\max} = H(B)/4 \le \ln(4)/4 \approx 0.347$ nats, a Lipschitz constant of $\sim$3 means the bound allows ρ to change by its full range when $\varepsilon \sim 0.1$, which is reasonable for the fast-mixing limit.

7.3 Slow mixing ($\text{gap}(P) \to 0$)

As $\text{gap}(P) \to 0$: the Cho–Meyer bound gives $\delta \le \varepsilon/\,\text{gap}(P) \to \infty$ (capped at $\delta = 2$); $L \to \infty$. This is physically correct: nearly reducible chains have stationary distributions that are extremely sensitive to small kernel perturbations, as tiny changes can redirect stationary mass between barely-communicating basins.

7.4 Independent B, M ($I(B;M) = 0$)

When the stationary distribution factorizes ($p(b,m) = p_B(b)\,p_M(m)$), $I(B;M) = 0$ and $\rho = 0$. Near this boundary, $\rho \sim I$ (to leading order, since $1 - I/H(B) \to 1$). The bound gives $|\rho - \rho'| \le |I - I'| + |H - H'|$, consistent with small perturbations producing small ρ'.

7.5 Perfect tracking ($I(B;M) = H(B)$)

When M is a deterministic function of B, $I(B;M) = H(B)$ and $\rho = 0$. Near this boundary, $\rho \sim H(B) - I = H(M|B)$. The bound handles this via $|\partial_I f| = |1 - 2I/H| \to 1$ as $I \to H$, so perturbations away from perfect tracking increase ρ linearly.

7.6 Toy model evaluation

For the 7-chain toy model ($|B| = |M| = 4$, $n = 16$), the logarithmic coefficient is:

$$\frac{1}{2}\big[2\ln 3 + \ln 3 + \ln 15\big] = \frac{1}{2}[2.197 + 1.099 + 2.708] = 3.002.$$

Representative bound values:

gap	ε	δ	Non-linear bound (nats)
1.0	0.01	0.01	0.156
1.0	0.10	0.10	1.094
0.5	0.01	0.02	0.284
0.5	0.10	0.20	1.901
0.1	0.01	0.10	1.094
0.1	0.10	1.00	5.775

Since $\rho_{\max} \approx 0.347$ nats, the bound is non-vacuous for small perturbations with moderate spectral gap.

8 Numerical Verification

We verify Theorem 3 numerically against random kernel perturbations. All computations use the verification code `code/lipschitz_verification.py` with seed 20260316 for reproducibility.

8.1 Observer chain

The test chain is a 16-state Markov chain on $\Omega = B \times M$ with $|B| = |M| = 4$. The body B follows a metastable two-basin dynamics (basins $\{0, 1\}$ and $\{2, 3\}$ with rare cross-basin transitions), and the model M tracks the body state with accuracy $\alpha = 0.5$ and noise floor $\gamma = 0.02$. This is the "7-chain observer" from the experiential measure framework.

The chain's stationary properties are:

Parameter	Value
$\lvert B\rvert$, $\lvert M\rvert$, $\lvert\Omega\rvert$	4, 4, 16
$\mathrm{gap}(P)$	0.100
$\rho(P)$	0.346 nats
$I_\pi(B;M)$	0.704 nats
$H_\pi(B)$	1.386 nats
I/H	0.508

The ratio $I/H \approx 0.5$ places the chain near the maximum of $f(I, H) = I(1 - I/H)$, which peaks at $I = H/2$.

8.2 Protocol

For each perturbation size $\varepsilon \in \{0.001, 0.01, 0.1\}$, we generate 1000 random kernel perturbations P' of P with $\|P - P'\|_\infty \approx \varepsilon$. Each perturbation is constructed by adding a zero-sum random vector to each row, clipping to non-negative entries, and renormalizing. Only irreducible perturbations (with $\mathrm{gap}(P') > 10^{-6}$) are retained.

For each perturbation we compute:

- The numerical Lipschitz ratio $R_i = |\rho(P) - \rho(P_i')|/\|P - P_i'\|_\infty$.
- The proven bound $L_{\text{proven}} = (\text{nonlinear bound at } \delta = \varepsilon/\,\mathrm{gap})/\varepsilon$.

The test passes if $\max_i R_i \leq L_{\text{proven}}$ for all ε values.

8.3 Results

ε	$L_{\text{numerical}}$	L_{proven}	Tightness ratio	Violations
0.001	0.012	109.4	9500×	0 / 1000
0.01	0.019	109.4	5900×	0 / 1000
0.1	0.115	51.4	450×	0 / 1000

Across all 3000 perturbations, $L_{\text{numerical}} \leq L_{\text{proven}}$ with zero violations. The bound is confirmed valid.

Figure 1 shows the distribution of Lipschitz ratios at $\varepsilon = 0.01$. The numerical ratios cluster near zero (median ≈ 0.005), while $L_{\text{proven}} = 109.4$ lies far to the right. This visual confirms both validity and conservatism.

8.4 Scaling laws

Spectral gap dependence. Varying the inter-basin transition rate q_{cross} produces chains with spectral gaps ranging from 0.005 to 0.3. A log-log regression of $L_{\text{numerical}}$ against $\mathrm{gap}(P)$ yields:

$$L_{\text{numerical}} \sim c \cdot \mathrm{gap}(P)^{-0.89}, \qquad R^2 = 0.97.$$

The exponent -0.89 is approximately consistent with the theoretical prediction of -1. The deviation may reflect finite-sample underestimation (since $L_{\text{numerical}}$ measures the empirical maximum over a finite sample rather than the true supremum), Cho–Meyer looseness for random (non-adversarial) perturbations, or both.

State-space size dependence. Varying $|B| = |M| = k$ for $k \in \{2, 3, 4, 5, 6\}$ and plotting the gap-normalized quantity $L_{\text{proven}} \cdot \mathrm{gap}$ against $\ln|\Omega|$ yields a linear fit with $R^2 = 0.97$. We note, however, that this test evaluates the proven formula against itself: L_{proven} contains logarithmic terms by construction, so this is a consistency check of the formula's internal structure rather than an independent confirmation of $\ln|\Omega|$ scaling in the empirical data.

8.5 Convergence

Drawing 2000 perturbations cumulatively and tracking the running maximum, $L_{\text{numerical}}$ stabilizes by $N \approx 500$ samples (relative change $< 5\%$ between successive thresholds), confirming that 1000 samples per ε are sufficient for reliable estimates.

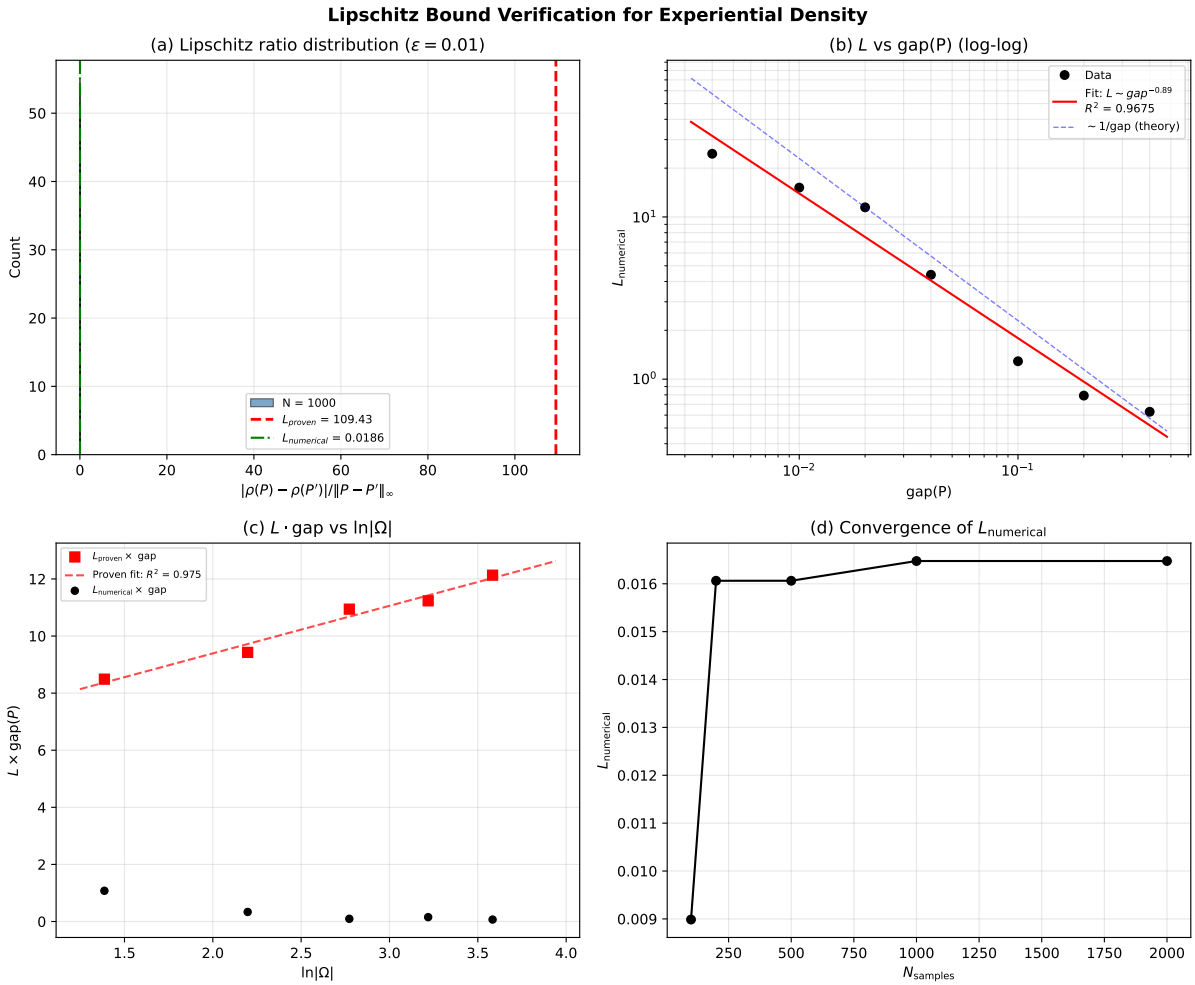

Figure 1: Four-panel verification of the Lipschitz bound. (a) Distribution of $|\rho(P)-\rho(P')|/\|P-P'\|_\infty$ for 1000 random perturbations at $\varepsilon = 0.01$; the proven bound $L_{\text{proven}} = 109.4$ (dashed red) is never exceeded. (b) Log-log plot of $L_{\text{numerical}}$ vs. gap(P), showing approximate $L \sim \text{gap}^{-0.89}$ scaling ($R^2 = 0.97$). (c) Gap-normalized Lipschitz constant $L \cdot \text{gap}$ vs. $\ln|\Omega|$; the linear fit ($R^2 = 0.97$) is a consistency check of the proven formula's logarithmic structure (see text). (d) Convergence of $L_{\text{numerical}}$ with sample size, stabilizing by $N \approx 500$.

9 Discussion

9.1 Tightness of the bound

The tightness gap of 450–9500× between L_{proven} and $L_{\text{numerical}}$ deserves comment. Four sources of slack compound:

(a) **Cho–Meyer step.** The bound $\delta \leq \varepsilon/\,\text{gap}$ is tight for adversarial perturbations that push all stationary mass in the direction of maximum sensitivity. Random perturbations average over many directions and typically produce $\delta \ll \varepsilon/\,\text{gap}$.

(b) **Entropy continuity.** The Fannes–Audenaert bound is tight for worst-case distributions (supported on two points), not for the specific marginals of an observer chain with rich internal structure.

(c) **Gradient bound.** The MVT uses $|\partial_I f| \leq 1$ and $|\partial_H f| \leq 1$ (worst case over $I/H \in [0,1]$). For the specific $I/H \approx 0.508$ of the test chain, $|\partial_I f| = |1 - 2(0.508)| = 0.016$ and

$|\partial_H f| = 0.508^2 = 0.258$, giving an effective gradient much smaller than the worst case.

(d) **Correlation structure.** The decomposition $|\Delta\rho| \leq |\Delta I| + |\Delta H|$ overcounts because I and H are correlated (both depend on π), so their perturbations may partially cancel rather than add.

9.2 Towards tighter bounds

A fundamentally tighter bound would bypass the three-step decomposition and instead compute $d\rho/dP_{ij}$ directly. Writing $\rho(P) = f(I(\pi(P)), H(\pi(P)))$, the chain rule gives

$$\frac{\partial \rho}{\partial P_{ij}} = \frac{\partial f}{\partial I}\frac{\partial I}{\partial P_{ij}} + \frac{\partial f}{\partial H}\frac{\partial H}{\partial P_{ij}},$$

where $\partial I/\partial P_{ij}$ and $\partial H/\partial P_{ij}$ can be expressed through $\partial\pi/\partial P_{ij}$, which involves the group inverse $(I - P + \Pi)^{-1}$. This approach would give the exact Lipschitz constant for a specific chain, but at the cost of chain-specific computation rather than a universal formula.

9.3 Role in the experiential measure framework

This stability result is the second of three foundational properties needed for the experiential density to serve as a well-behaved dynamical functional:

1. **Existence and uniqueness** (Theorem A): $\rho(P)$ is well-defined for any irreducible P on $B \times M$.
2. **Lipschitz stability** (this paper): ρ depends continuously on the kernel, with quantitative control.
3. **Measurability**: $\rho(P)$ can in principle be estimated from finite trajectory data.

The Lipschitz bound also provides error propagation: if the kernel P is known only to accuracy ε (e.g., from finite-data estimation), then $\rho(P)$ is known to accuracy at most $L(\varepsilon/\,\mathrm{gap})\cdot\varepsilon$. For the 7-chain observer with gap $= 0.1$ and $\varepsilon = 0.01$, the non-linear bound (24) gives $|\Delta\rho| \leq 1.09$ nats, which exceeds $\rho_{\max} \approx 0.347$ and is therefore vacuous. In practice, the actual perturbation is much smaller ($|\Delta\rho| \sim 10^{-4}$ nats), confirming that the bound is conservative but that ρ is genuinely stable.

More generally, the bound (24) is non-vacuous (i.e., smaller than $\rho_{\max} = (\ln|B|)^2/(4\ln|B|) = \ln|B|/4$) when the perturbation ε is small enough relative to the spectral gap and state-space size. As a rough criterion, the leading term dominates for small ε, so the bound is non-vacuous when

$$\frac{\varepsilon}{2\,\mathrm{gap}}\big[2\ln(|B|-1) + \ln(|M|-1) + \ln(n-1)\big] \lesssim \frac{\ln|B|}{4}\,,$$

i.e., when $\varepsilon \lesssim \mathrm{gap}\cdot\ln|B|/(8\ln n)$. For chains with moderate spectral gap (gap ~ 0.1) and modest state space ($n \sim 16$), this requires $\varepsilon \lesssim 0.006$, confirming that the bound is informative only for quite small perturbations of such chains.

References

[1] R. Lopez-Ruiz, H. L. Mancini, X. Calbet, "A statistical measure of complexity," *Phys. Lett. A* **209** (1995) 321–326.

[2] M. Gell-Mann, S. Lloyd, "Effective complexity," in *Nonextensive Entropy*, eds. M. Gell-Mann, C. Tsallis, Oxford Univ. Press (2004); original preprint 1996/2003.

[3] G. E. Cho, C. D. Meyer, "Comparison of perturbation bounds for the stationary distribution of a Markov chain," *Linear Algebra Appl.* **335** (2001) 137–150.

[4] C. D. Meyer, "The role of the group generalized inverse in the theory of finite Markov chains," *SIAM Rev.* **17** (1975) 443–464.

[5] I. Csiszár, J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed., Cambridge Univ. Press (2011), Lemma 2.7.

[6] K. M. R. Audenaert, "A sharp continuity estimate for the von Neumann entropy," *J. Phys. A* **40** (2007) 8127–8136; arXiv:quant-ph/0610146.

[7] M. Fannes, "A continuity property of the entropy density for spin lattice systems," *Commun. Math. Phys.* **31** (1973) 291–294.

[8] E. Seneta, *Non-negative Matrices and Markov Chains*, Springer Series in Statistics, 2nd ed. (2006).

[9] D. A. Levin, Y. Peres, *Markov Chains and Mixing Times*, 2nd ed., American Mathematical Society (2017).

[10] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley-Interscience (2006).

[11] B. Ehrlich, "Exponential suppression of transient-basin contributions in trajectory-weighted Markov chain measures," preprint (2026).

[12] I. C. F. Ipsen and C. D. Meyer, "Uniform stability of Markov chains," *SIAM J. Matrix Anal. Appl.* **15** (1994) 1061–1074.

[13] A. Winter, "Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance, and energy constraints," *Commun. Math. Phys.* **347** (2016) 291–313.