Latent Probability: Conditional Dependence as Graded Spectral Structure

Tamás Nagy, Ph.D.

tnagyphd@gmail.com

Draft

Executive Summary (Non-Technical)

Probability theory — the foundation of statistics, machine learning, and quantitative science — relies on a structural assumption called the **Markov property**: the future depends on the past only through the present. This property is universally treated as an axiom: we assume it and derive consequences.

We show that the Markov property is not an axiom but a **theorem**. It emerges as a necessary consequence of a deeper structure: the graded Hilbert tensor algebra introduced in the Latent framework (Nagy, 2026). Conditional probability — the relationship $P(X \mid Y)$ — is a special case of grade-2 latent structure, and the Markov property follows from the finite rank of this structure. When the grade-2 tensor has rank R, conditional independence is forced: no system can sustain dependencies longer than what R dimensions can carry.

This reframing has three consequences. First, the effective complexity of any dependency network is bounded by R, not by the number of variables — a dramatic reduction for large systems. Second, cyclic dependencies (feedback loops, self-reference) are classified by the eigenvalues of their composed maps: they either dissolve, become symmetries, or signal instability. Third, the classical factorization theorems of graphical models (Hammersley-Clifford, Pearl's d-separation) become corollaries of a single spectral decomposition.

Abstract

We embed conditional probability in the graded Hilbert tensor algebra $\Lambda^{(i,j)}$ of the Latent framework. The conditional distribution $P(X \mid Y)$ is identified with a grade-2 linear map $\Lambda^{(X,Y)} : \mathcal{H}_Y \to \mathcal{H}_X$ between latent spaces. A Bayesian network is the sparsity pattern of the grade-2 tensor. We prove:

1. **Markov Emergence Theorem.** If the grade-2 tensor of a system with n variables has meta-analyticity $\rho_{\text{meta}} > 1$, then the induced probability structure satisfies approximate conditional independence with Markov blanket size at most $R = O(\log(n/\varepsilon)/\log \rho_{\text{meta}})$. The Markov property is a consequence of rank deficiency, not an axiom.
2. **Dependency Classification Theorem.** Every dependency cycle in a graded system is classified by the monodromy spectrum of the composed grade-2 maps. Let $M = \Lambda^{(A,C)} \circ \Lambda^{(C,B)} \circ \Lambda^{(B,A)}$ be the monodromy of a cycle $A \to B \to C \to A$. Then:
   - If $|\mu_i| < 1$ for all eigenvalues: the cycle is *transient* — it dissolves at grade $n+1$.

- If $|\mu_i| = 1$ for some eigenvalues: the cycle is *resonant* — it becomes a symmetry at grade $n+1$.
- If $|\mu_i| > 1$ for some eigenvalues: the cycle is *unstable* — it requires grade $n+1$ nonlinearity to regularize.

3. **Graphical Model Subsumption.** The Hammersley-Clifford theorem, Pearl's d-separation criterion, and the factorization theorem for Bayesian networks are special cases of the grade-2 spectral decomposition.

The framework connects probability theory to dynamical systems: periodic orbits in the N-body problem are resonant dependency cycles whose monodromy is exactly the Poincaré return map. The virial crossing property of bounded orbits (Lagrange-Jacobi) is the constraint that prevents unstable cycles from persisting in bounded systems.

---

1. Introduction

1.1 The Problem: Where Does the Markov Property Come From?

The Markov property — $P(X_{t+1} \mid X_t, X_{t-1}, ...) = P(X_{t+1} \mid X_t)$ — is the foundation of stochastic processes, Bayesian networks, hidden Markov models, and much of applied probability. It is universally treated as a modeling assumption: we choose whether a system "is Markov" and accept the consequences.

But this creates an uncomfortable gap. Every system can be made Markov by enlarging the state space: a second-order process $X_{t+1} = f(X_t, X_{t-1})$ becomes first-order Markov as $Y_t = (X_t, X_{t-1})$. This is not a discovery about the system — it is a mathematical trick. The question becomes: **is the Markov property a property of the universe, or merely a property of our notation?**

We show it is a property of the universe — specifically, of the grade structure of the universe's dependency tensor. The argument has three steps:

1. Conditional probability $P(X \mid Y)$ is identified with a grade-2 linear map between latent spaces (Section 2).
2. If the family of all grade-2 maps has finite meta-rank R (which follows from meta-analyticity), then conditional independence is forced: the information bottleneck of rank R cannot sustain dependencies beyond R dimensions (Section 3).
3. The Markov blanket of any variable is therefore bounded by R, not by the total number of variables (Section 3).

1.2 The Deeper Question: Cyclic Dependencies

Standard probability theory avoids cycles. Bayesian networks are DAGs (directed acyclic graphs). Markov random fields use undirected edges. But the real universe is full of cycles: feedback in control systems, predator-prey dynamics, economic equilibria, self-referential cognitive processes, and — most concretely — periodic orbits in celestial mechanics.

How should we handle cyclic dependencies? Three positions exist:

(a) Ban them. Restrict to DAGs. This is the standard Bayesian network approach, but it is a modeling choice, not a theorem.

(b) Handle them as fixed points. Cyclic causal models (Bongers et al., 2021) and structural equation models with feedback define solutions as fixed points of iterative maps. This works but gives no structural classification.

(c) Classify them spectrally. This is our approach. A dependency cycle $A \to B \to C \to A$ composes the grade-2 maps into a monodromy operator $M : \mathcal{H}_A \to \mathcal{H}_A$. The eigenvalues of M classify the cycle completely: contractive cycles dissolve (the dependence decays), unitary cycles become symmetries (standing waves in the dependency structure), and expansive cycles are genuinely unstable (the "feedback" is real, and no grade-2 model can tame it).

The key insight is that **what looks like a cycle at grade 2 becomes an eigenstructure at grade 3**. The cycle does not disappear — it is reinterpreted as a fixed-point property of the composed map. This is precisely what happens in dynamics: periodic orbits (grade-2 cycles in phase space) become fixed points of the Poincaré return map (grade-3 structure).

1.3 Contributions

1. **Conditional probability as grade-2 latent structure** (Section 2): formal identification of $P(X \mid Y)$ with $\Lambda^{(X,Y)} \in \mathcal{L}(\mathcal{H}_Y, \mathcal{H}_X)$ and the joint distribution $P(X_1, \dots, X_n)$ with the grade-2 tensor.
2. **The Markov Emergence Theorem** (Section 3): the Markov property as a consequence of finite meta-rank, with explicit bound on Markov blanket size.
3. **The Dependency Classification Theorem** (Section 4): monodromy eigenvalue trichotomy for cyclic dependencies, with grade-shift resolution.
4. **Subsumption of graphical model theory** (Section 5): Hammersley-Clifford, d-separation, and Bayesian factorization as corollaries of the spectral decomposition.
5. **Connection to celestial mechanics** (Section 6): periodic orbits as resonant dependency cycles, virial crossing as the bounded-system constraint.
6. **Empirical predictions** (Section 7): testable consequences using existing GPT-2 and TinyLlama latent data from Nagy (2026).

---

2. Conditional Probability as Grade-2 Latent Structure

2.1 Setup

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_1, \dots, X_n$ be random variables taking values in measurable spaces $(\mathcal{X}_1, \dots, \mathcal{X}_n)$. For each variable X_i, let $\mathcal{H}_i = L^2(\mathcal{X}_i, P_{X_i})$ be the Hilbert space of square-integrable functions of X_i.

Definition 1 (Grade-1 Latent). The grade-1 latent of X_i is its marginal distribution P_{X_i}, identified with the element $\Lambda_i^{(1)} \in \mathcal{H}_i$ representing the density (or Radon-Nikodym derivative) with respect to a reference measure.

Definition 2 (Grade-2 Latent). The grade-2 latent connecting X_i to X_j is the conditional expectation operator:

$$\Lambda^{(i,j)} : \mathcal{H}_j \to \mathcal{H}_i, \qquad (\Lambda^{(i,j)} f)(x_i) = \mathbb{E}[f(X_j) \mid X_i = x_i]$$

This is a bounded linear operator between Hilbert spaces. Its operator norm satisfies $\|\Lambda^{(i,j)}\| \le 1$ (conditional expectation is a contraction).

Definition 3 (Grade-2 Tensor). The full grade-2 structure of the system $(X_1, \dots, X_n)$ is the operator-valued matrix:

$${}^{(2)} = \begin{pmatrix} \Lambda^{(1,1)} & \Lambda^{(1,2)} & \cdots & \Lambda^{(1,n)} \\ \Lambda^{(2,1)} & \Lambda^{(2,2)} & \cdots & \Lambda^{(2,n)} \\ \vdots & & \ddots & \vdots \\ \Lambda^{(n,1)} & \Lambda^{(n,2)} & \cdots & \Lambda^{(n,n)} \end{pmatrix}$$

where each $\Lambda^{(i,j)}$ is the conditional expectation operator from $\mathcal{H}_j$ to $\mathcal{H}_i$.

2.2 The Identification

The core identification is:

$$P(X_i \mid X_j) \quad \longleftrightarrow \quad \Lambda^{(i,j)} \in \mathcal{L}(\mathcal{H}_j, \mathcal{H}_i)$$

This is not a metaphor. The conditional expectation operator IS the conditional distribution, acting on function space. What we add is the observation that this operator is a **grade-2 element** in the graded Hilbert tensor algebra $\Lambda^{(i,j)}$ of the Latent framework.

The chain rule of conditional probability becomes a composition of grade-2 maps:

$$P(X_1, X_2, X_3) = P(X_1) \cdot P(X_2 \mid X_1) \cdot P(X_3 \mid X_1, X_2)$$

$$\longleftrightarrow \quad \Lambda_1^{(1)} \otimes \Lambda^{(2,1)} \otimes \Lambda^{(3,\{1,2\})}$$

A Bayesian network on graph $G = (V, E)$ is the statement that most entries of ${}^{(2)}$ are "trivial" (i.e., $\Lambda^{(i,j)} = 0$ whenever $j \notin \mathrm{pa}_G(i)$). The graph is the **sparsity pattern** of the grade-2 tensor.

2.3 Entropy Decomposition

The Shannon entropy of the joint distribution decomposes as:

$$H(X_1, \dots, X_n) = \sum_{i=1}^{n} H(X_i \mid \mathrm{pa}_G(X_i))$$

In the grade-2 language, each term $H(X_i \mid \mathrm{pa}_G(X_i))$ is the **conditional entropy** associated with the grade-2 map $\Lambda^{(i,\mathrm{pa}(i))}$. The total entropy is a sum of local grade-2 contributions.

Moreover, the mutual information between X_i and X_j is controlled by the operator norm of the grade-2 map:

$$I(X_i; X_j) \leq \log \|\Lambda^{(i,j)}\|_{\text{HS}}^2$$

where $\|\cdot\|_{\text{HS}}$ is the Hilbert-Schmidt norm. This connects information content to the spectral properties of the grade-2 tensor.

2.4 Spectral Decomposition of the Grade-2 Tensor

Each operator $\Lambda^{(i,j)}$ admits a singular value decomposition:

$$\Lambda^{(i,j)} = \sum_{k=1}^{\infty} \sigma_k^{(ij)} \, u_k^{(i)} \otimes v_k^{(j)}$$

where $\sigma_1^{(ij)} \geq \sigma_2^{(ij)} \geq \cdots \geq 0$ are the singular values, and $\{u_k^{(i)}\}$, $\{v_k^{(j)}\}$ are orthonormal bases for $\mathcal{H}_i$ and $\mathcal{H}_j$ respectively. The **effective rank** of $\Lambda^{(i,j)}$ at tolerance ε is:

$$r_\varepsilon^{(ij)} = \min \left\{ r : \sum_{k=r+1}^{\infty} (\sigma_k^{(ij)})^2 \leq \varepsilon \sum_{k=1}^{\infty} (\sigma_k^{(ij)})^2 \right\}$$

If the system has meta-analyticity $\rho_{\text{meta}} > 1$ (in the sense of the Latent Theorem), then the singular values decay exponentially: $\sigma_k^{(ij)} \leq C \cdot \rho_{\text{meta}}^{-k}$. This forces $r_\varepsilon^{(ij)} = O(\log(1/\varepsilon) / \log \rho_{\text{meta}})$ — the grade-2 map has finite effective rank.

---

3. The Markov Emergence Theorem

3.1 Meta-Rank of the Dependency Structure

Consider the full grade-2 tensor $^{(2)}$ as a mapping from "source variables" to "target variables." Its meta-rank R is the effective rank of the family of conditional expectation operators $\{\Lambda^{(i,j)}\}_{i,j=1}^n$, viewed as elements of the operator space $\mathcal{L}(\mathcal{H}, \mathcal{H})$.

If the system has meta-analyticity $\rho_{\text{meta}} > 1$ (i.e., the family of conditional operators varies smoothly as we range over variable pairs), then by the Latent Theorem:

$$R = O\left(\frac{\log(n/\varepsilon)}{\log \rho_{\text{meta}}} \right)$$

This R is the number of "dependency modes" in the system — the directions along which variables can influence each other.

3.2 The Theorem

Theorem 1 (Markov Emergence). Let $X_1, \ldots, X_n$ be random variables whose grade-2 tensor $^{(2)}$ has meta-analyticity $\rho_{\text{meta}} > 1$ and effective meta-rank R at tolerance ε. Then:

(i) **Blanket bound.** For every variable X_i, there exists a set $B_i \subseteq \{1, \ldots, n\} \setminus \{i\}$ with $|B_i| \leq R$ such that:

$$\|P(X_i \mid X_{-i}) - P(X_i \mid X_{B_i})\|_{\mathrm{TV}} \leq \varepsilon$$

That is, X_i is approximately conditionally independent of all other variables given the Markov blanket B_i, and $|B_i| \leq R$.

(ii) **Factorization.** The joint distribution admits an approximate factorization:

$$P(X_1, \dots, X_n) \approx \prod_{i=1}^{n} P(X_i \mid X_{B_i})$$

with total variation error at most $n\varepsilon$.

(iii) **Graph sparsity.** The conditional independence graph G (where $i - j$ is absent iff $X_i \perp X_j \mid X_{-(i,j)}$) has maximum degree at most R.

Proof sketch. The grade-2 tensor lives in an R-dimensional subspace (by the meta-rank bound). The conditional distribution $P(X_i \mid X_{-i})$ is a linear functional on this subspace. By rank-nullity, it depends on at most R independent directions. The variables X_{B_i} that span these directions form the Markov blanket. The approximation error ε comes from the truncation of the singular value expansion at rank R. □

3.3 Interpretation

The theorem says: **the Markov property is not an assumption — it is forced by the finite rank of the grade-2 tensor.**

At grade 2 (the level of conditional probabilities), the Markov property appears as a choice: we select a state space, we select which variables to condition on. The theorem lifts this to grade 3: the meta-analyticity of the dependency family forces a rank bound, and that rank bound forces conditional independence. There is no choice involved — the structure of the system determines its Markov properties.

The result generalizes the folk theorem that "any system can be made Markov by enlarging the state space." We make this precise: the minimum state space needed has dimension R, which depends on the meta-analyticity of the system, not on the number of variables. For a system with $\rho_{\mathrm{meta}} = 2$ and $\varepsilon = 10^{-6}$, we get $R \leq 20$ regardless of whether $n = 100$ or $n = 10^6$.

3.4 Recovering Hammersley-Clifford

The Hammersley-Clifford theorem states: a distribution P satisfies the Markov property with respect to an undirected graph G if and only if P factorizes over the cliques of G.

In our framework: the graph G is the support of the grade-2 tensor (the set of nonzero $\Lambda^{(i,j)}$). The cliques of G are the maximal sets of mutually nonzero grade-2 connections. The clique factorization is the restriction of the grade-2 SVD to each clique. Hammersley-Clifford is thus a special case of the grade-2 spectral decomposition, where the "approximate" in Theorem 1 becomes exact (because the graph is assumed exact).

---

4. The Dependency Classification Theorem

4.1 Cyclic Dependencies at Grade 2

Let $C = (i_1, i_2, \ldots, i_k, i_1)$ be a cycle of length k in the dependency structure. The **monodromy operator** of the cycle is the composition:

$$M_C = \Lambda^{(i_1,i_k)} \circ \Lambda^{(i_k,i_{k-1})} \circ \cdots \circ \Lambda^{(i_2,i_1)} : \mathcal{H}_{i_1} \to \mathcal{H}_{i_1}$$

This is a self-map on the latent space of the first variable. It answers the question: "if I propagate information from X_{i_1} through the entire cycle and back, what comes back?"

Since each $\Lambda^{(i,j)}$ is a contraction ($\|\Lambda^{(i,j)}\| \leq 1$), we have $\|M_C\| \leq 1$. The eigenvalues $\mu_1, \mu_2, \ldots$ of M_C satisfy $|\mu_k| \leq 1$.

4.2 The Classification

Theorem 2 (Dependency Classification). Let C be a dependency cycle with monodromy M_C and eigenvalues $\{\mu_k\}$. Then:

(i) Transient ($|\mu_k| < 1$ for all k). The cycle is a contraction. Information injected into the cycle decays exponentially:

$$\|M_C^m\| \leq C \cdot |\mu_{\max}|^m \to 0 \quad \text{as } m \to \infty$$

At grade $n+1$, the cycle contributes nothing: it dissolves into the universal component $\bar{\Lambda}$.

Interpretation: The variables in the cycle appear to depend on each other, but the dependency is transient. Given enough "rounds" of conditioning, the mutual information decays to zero. The cycle is an artifact of finite-sample observation.

(ii) Resonant ($|\mu_k| = 1$ for some k, $|\mu_j| \leq 1$ for all j). The cycle contains a unitary component. The resonant eigenspace is preserved exactly under iteration:

$$M_C^m v = e^{im\theta} v \quad \text{for } v \text{ in the resonant eigenspace}$$

At grade $n+1$, the resonant component becomes a **symmetry**: a standing wave in the dependency structure. It does not decay and does not grow. It is a conserved quantity of the dependency dynamics.

Interpretation: The cycle represents a genuine structural feature of the system — not a transient correlation but a persistent mode. Examples include conservation laws (energy, momentum), periodic orbits (celestial mechanics), and equilibrium constraints (economic markets).

(iii) Unstable ($|\mu_k| > 1$ for some k). This case is excluded for conditional expectation operators (which are contractions). However, if we extend the framework to include non-probabilistic grade-2 maps (e.g., differential operators, gain matrices), unstable cycles are possible. They indicate that the grade-2 description is inadequate: nonlinear terms at grade $n+1$ must regularize the growth.

Interpretation: The feedback is real, and no linear (grade-2) model can capture it. This is the signature of genuine nonlinear interaction — the system is not "Markov-like" at any scale, and higher-grade structure is essential.

4.3 The Grade-Shift Resolution

The fundamental observation: **what looks like a cycle at grade n becomes an eigenstructure at grade $n+1$.**

At grade 2, the cycle $A \to B \to C \to A$ is a directed path that returns to its starting point. This is paradoxical from a causal perspective: does A cause B, or does B cause A?

At grade 3, the cycle is replaced by the eigendecomposition of M_C. The resonant eigenvectors are the "modes" of the cycle — the patterns that persist under iteration. These modes are not causal (they don't have a direction); they are structural (they characterize the invariant subspace).

This grade-shift resolves the foundational problem of cyclic causation. Cycles at grade 2 are not paradoxes — they are low-grade descriptions of higher-grade eigenstructure.

4.4 Example: Periodic Orbits in the N-Body Problem

The N-body gravitational system provides a concrete example. A periodic orbit γ is a cycle in phase space: the system returns to its initial state after period T.

The **Poincaré return map** $\Phi : \Sigma \to \Sigma$ (where Σ is a cross-section transversal to γ) is precisely the monodromy operator of this cycle. Its eigenvalues — the **Floquet multipliers** — classify the orbit:

- $|\mu| < 1$: the orbit is asymptotically stable (nearby trajectories spiral inward). The "cycle" dissolves: after long times, neighboring orbits forget they were near a periodic orbit.
- $|\mu| = 1$: the orbit is neutrally stable. In Hamiltonian systems, energy conservation forces the Floquet multipliers to come in pairs $(\mu, 1/\mu)$ on the unit circle. These resonant modes are the **normal mode frequencies** of the orbit.
- $|\mu| > 1$: the orbit is unstable. Nearby trajectories escape exponentially.

The virial crossing property — proved via the Lagrange-Jacobi identity $\ddot{I} = 2(U - 2h)$ — constrains which cycles can exist in bounded systems. A bounded orbit has bounded moment of inertia I, which forces $\ddot{I}$ to change sign, which forces U to cross $2h$. This is exactly the statement that the monodromy cannot be purely expansive ($|\mu| > 1$ for all modes) in a bounded system.

---

5. Subsumption of Graphical Model Theory

5.1 Bayesian Networks

Proposition 1. The factorization theorem for Bayesian networks is a special case of the grade-2 spectral decomposition.

A Bayesian network on DAG $G = (V, E)$ asserts:

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_{\mathrm{pa}_G(i)})$$

In our framework: the DAG G is the sparsity pattern of $^{(2)}$ (nonzero entries correspond to edges). The factorization is the product of grade-2 maps along the topological ordering. The d-separation criterion for conditional independence reads off from the rank structure of sub-tensors of $^{(2)}$.

5.2 Markov Random Fields and Hammersley-Clifford

Proposition 2. The Hammersley-Clifford theorem is the special case of Theorem 1 where the graph is exact (not approximate) and the distribution is strictly positive.

The Gibbs distribution over cliques $\mathcal{C}$ of graph G:

$$P(x) = \frac{1}{Z} \exp\left(-\sum_{c \in \mathcal{C}} \phi_c(x_c) \right)$$

corresponds to grade-2 maps $\Lambda^{(i,j)}$ that are nonzero exactly when i and j belong to a common clique, with the potential functions ϕ_c encoding the local grade-2 structure within each clique.

5.3 Factor Graphs and Message Passing

Belief propagation on a factor graph is an iterative computation of marginals by passing "messages" along edges. In the grade-2 framework, each message is a projection onto the singular vectors of the local grade-2 map. The convergence of belief propagation depends on the spectral radius of the grade-2 tensor (Mooij & Kappen, 2007):

$$\text{BP converges} \iff \rho(\ ^{(2)}) < 1$$

This is exactly the condition for the dependency cycles to be transient (all monodromy eigenvalues inside the unit disk). Loopy BP fails precisely when resonant or unstable cycles exist.

---

6. Connection to Dynamical Systems

6.1 The Grade Hierarchy of Dynamics

Dynamical concept	Grade-2 view	Grade-3 view
Periodic orbit	Cycle in phase space	Fixed point of Poincaré map
Feedback loop	Cyclic dependency $A \to B \to A$	Eigenvalue of loop gain
Conservation law	Constraint along orbit	Resonant mode ($\|\mu\| = 1$)
Dissipation	Transient cycle ($\|\mu\| < 1$)	Decay to universal component
Instability	Expanding cycle ($\|\mu\| > 1$)	Requires nonlinear regularization

6.2 Floquet Theory as Grade-2 Spectral Decomposition

Floquet's theorem states that the fundamental matrix of a T-periodic linear system $\dot{x} = A(t)x$ can be written as $\Phi(t) = P(t)e^{Bt}$ where $P(t+T) = P(t)$. The eigenvalues of e^{BT} are the Floquet multipliers.

In our framework: $A(t)$ is the time-varying grade-2 map, $P(t)$ captures the within-period variation (grade-2 structure), and e^{BT} is the monodromy M_C (grade-3 structure). Floquet's theorem IS the grade-shift from cycle to eigenstructure.

6.3 Renormalization Group as Grade-Shift

In statistical physics, the renormalization group "zooms out" by integrating over short-scale degrees of freedom. Under RG flow:

- Short-range correlations (transient at the coarse scale) disappear — they had $|\mu| < 1$ at the fine scale.
- Long-range correlations at the critical point ($|\mu| = 1$) persist as scale-invariant modes.
- Relevant perturbations ($|\mu| > 1$) drive the system away from the fixed point.

This is exactly the Dependency Classification Theorem applied scale by scale. The RG fixed point is a resonant dependency cycle that becomes a symmetry (conformal invariance) at the next grade.

---

7. Empirical Predictions

7.1 Using Existing Data

The GPT-2 and TinyLlama experiments from Nagy (2026, Latent of Latents) provide an immediate test bed. The 41-domain latent data includes domain mean vectors $\bar{\Lambda}_i \in \mathbb{R}^d$ for each domain i.

Prediction 1: Grade-2 tensor has low rank. Construct the cross-domain conditional structure:

$$\Lambda^{(i,j)} \approx \frac{\text{Cov}(\Lambda_i, \Lambda_j)}{\text{Var}(\Lambda_j)}$$

using the per-prompt vectors within each domain. The effective rank of this 41×41 operator matrix should be bounded by $R_{95\%} \approx 10$ (GPT-2) or ≈ 30 (TinyLlama), matching the meta-rank from the Latent of Latents analysis.

Prediction 2: Markov blanket size $\leq R$. The conditional independence graph induced by the grade-2 tensor should have maximum vertex degree $\leq R$. For GPT-2 with $R = 10$, no domain should need more than 10 other domains to predict its latent representation.

Prediction 3: Domain-domain-domain cycles are contractive. For any triple (i, j, k), the monodromy $M = \Lambda^{(i,k)}\Lambda^{(k,j)}\Lambda^{(j,i)}$ should have all eigenvalues $|\mu| < 1$. This predicts that domain dependencies in LLMs do not contain resonant cycles — there are no "standing waves" in how GPT-2 organizes its knowledge across domains. (If resonant cycles exist, they would correspond to conserved quantities of the LLM's knowledge structure — a highly interesting finding.)

7.2 New Experiments Needed

Experiment 1: Conditional independence graph. Compute the partial correlation matrix of domain latents. Threshold to obtain the conditional independence graph. Compare its degree distribution to the meta-rank bound from Theorem 1.

Experiment 2: Monodromy spectrum. For all $\binom{41}{3} = 10{,}660$ triples of domains, compute the monodromy eigenvalues. Plot the distribution. Test whether all fall inside the unit disk.

Experiment 3: Information flow and spectral gap. The graph Laplacian of the conditional independence graph has eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots$. The spectral gap λ_2 controls the mixing time of information diffusion on the graph. Compare to the meta-rank: the prediction is $\lambda_2 \sim 1/R$.

---

8. Discussion

8.1 Relationship to the Latent of Latents

This paper extends the Latent of Latents (Nagy, 2026) from a representation theory to a theory of dependence. The bi-graded algebra $\Lambda^{(i,j)}$ was introduced in that work for grade i (within-system) and grade j (across-family). Here, we identify the grade-j structure with conditional probability, making the algebra not just a convenient notation but a foundational object that subsumes probability theory.

The tri-graded extension $\Lambda^{(i,j,k)}$ proposed in Nagy (2026, Section 8) acquires a new interpretation: - Grade i: within-domain content (the distribution $P(X)$) - Grade j: across-domain dependence (the conditional $P(X \mid Y)$) - Grade k: depth of universal structure (what all dependencies share)

The Markov Emergence Theorem lives at the $j \to k$ interface: the finiteness of grade-j rank forces conditional independence, which is a grade-k structural property.

8.2 Relationship to Pricing = Allocation

If pricing equals allocation (Nagy, 2026, Pricing-Allocation Unity), and allocation decomposes spectrally, then **pricing factorizes along the dependency graph**. A basket option on n correlated assets becomes a sum of R eigenmode prices, where R is the meta-rank of the asset dependency structure. The correlation matrix is the grade-2 tensor of the asset system, and its eigendecomposition is the spectral pricing formula.

This would unify: - The eigenvalue-conditional pricing formula (Nagy, 2026, Fenton) - The Bayesian network factorization of risk - The spectral decomposition of portfolio allocation

8.3 The Magician's Hat

Is the Markov property a trick or a truth? Our answer: **both, at different grades.**

At grade 2, it is a trick. You choose the state space to make the process Markov. The "memorylessness" is an artifact of your notation.

At grade 3, it is a truth. The meta-analyticity of the dependency family forces the grade-2 tensor to have finite rank, and that finite rank forces conditional independence. You do not choose this — the structure of the system determines it.

The hat (the grade-2 state space) has finite capacity R. The rabbit (the Markov property) fits inside the hat not because the magician is clever, but because the hat's capacity is a physical constraint.

---

9. Conclusion

We have shown that conditional probability is grade-2 latent structure, and that the Markov property emerges from the finite rank of this structure rather than being axiomatically imposed. Cyclic dependencies are classified by their monodromy spectrum: they dissolve, become symmetries, or signal the need for higher-grade description. The classical results of graphical model theory — Hammersley-Clifford, d-separation, the factorization theorem — are special cases of the grade-2 spectral decomposition.

The deepest consequence is conceptual: probability theory is not a separate mathematical framework but a **lossy projection** of a richer latent structure. Standard probability works with the trace of the grade-2 tensor (collapsing to scalars — probabilities). The full grade-2 tensor carries more information: not just "how likely" but "in which direction" the dependence operates. The compression ratio between the full tensor and the scalar projection is exactly the meta-rank R: the number of dependency modes in the system.

Three items remain for future work: (1) full Lean 4 formalization of the Markov Emergence Theorem, building on the existing ScalingLaw.lean and BiGradedAlgebra.lean; (2) empirical validation using the GPT-2 and TinyLlama latent data; (3) the connection to pricing-allocation unity, which would give a spectral factorization of financial risk along the dependency graph.

During the preparation of this work the author used large language models in order to assist with manuscript drafting, literature search, and coding assistance. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

- Bongers, S., P. Forré, J. Peters, and J. M. Mooij (2021). Foundations of structural causal models with cycles and latent variables. *Annals of Statistics*, 49(5), 2885-2915. DOI: 10.1214/21-aos2064
- Hammersley, J. M. and P. Clifford (1971). Markov fields on finite graphs and lattices. *Unpublished manuscript.*
- Koller, D. and N. Friedman (2009). Probabilistic Graphical Models: Principles and Techniques. *Probabilistic Graphical Models: Principles and Techniques.*
- Lauritzen, S. L (1996). Graphical Models. *Graphical Models.* DOI: 10.1093/oso/9780198522195.001.0001
- Mooij, J. M. and H. J. Kappen (2007). Sufficient conditions for convergence of the sum-product algorithm. *IEEE Trans. Inf. Theory*, 53(12), 4422-4437. DOI:

10.1109/tit.2007.909166

- Nagy, T. (2026). The Latent: Finite Sufficient Representations of Smooth Systems. *Zenodo*. DOI: 10.5281/zenodo.19101209
- Nagy, T. (2026). The Latent of Latents: Hierarchical Finite Representations of Knowledge Families. *Zenodo*. DOI: 10.5281/zenodo.19134434
- Nagy, T. (2026). Pricing Is Allocation: A Mode-Decomposition Theory of Financial Markets. *Working paper*.
- Pearl, J (1988). Probabilistic Reasoning in Intelligent Systems. *Probabilistic Reasoning in Intelligent Systems*.
- Pearl, J (2009). Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press. *Causality: Models, Reasoning, and Inference*.