LLM Benchmark Report

Generated by script

June 27, 2025

Summary Statistics

Model	Total	Correct	Partial	Incorrect	Errors	Avg. Score
o3	49	22	7	20	0	0.52

Problem Details

Problem 1 (Paper: 2506.17853v1)

Problem Statement

Background: A robotic mechanical metamaterial is described by a continuum reaction-diffusion system for its dimensionless angular displacements α and β as functions of dimensionless spatial coordinate $\bar{x}$ and dimensionless time $\bar{t}$. The dynamics are governed by the following coupled partial differential equations:

$$\frac{\partial \alpha}{\partial \bar{t}} = k_\alpha \Delta \alpha + R_\alpha(\alpha, \beta)$$

$$\frac{\partial \beta}{\partial \bar{t}} = \frac{k_\beta}{\gamma} \Delta \beta + \frac{1}{\gamma} R_\beta(\alpha, \beta)$$

where Δ is the Laplacian operator, k_α and k_β are linearized torsional stiffnesses, and $\gamma = \eta_\beta / \eta_\alpha$ is the ratio of effective viscous coefficients. The non-conservative reaction functions are given by:

$$R_\alpha(\alpha, \beta) = k_{\alpha\beta}(\beta - \alpha) - \psi'(\alpha) - \sigma\beta$$

$$R_\beta(\alpha, \beta) = k_{\alpha\beta}(\alpha - \beta)$$

Here, $k_{\alpha\beta}$ and σ are positive constants. The term $\psi'(\alpha)$ is a non-monotonic, odd, piecewise continuous tri-linear function with three roots at $\alpha = \{0, \pm\alpha^*\}$. Its derivative, $\psi''(\alpha)$, is given by:

$$\psi''(\alpha) = \begin{cases} k_2, & |\alpha| \leq k_1 \alpha^* / (k_1 - k_2), \\ k_1, & |\alpha| > k_1 \alpha^* / (k_1 - k_2), \end{cases}$$

where k_1 and k_2 are constants. The system has a homogeneous equilibrium state at $(\alpha, \beta) = (0, 0)$. For linear stability analysis, a perturbation from this equilibrium is considered, which can be Fourier-expanded with spatial frequency $\kappa = 2\pi/\lambda$.

Task: Derive the matrix $\mathbf{J}_\kappa$ for the linear stability analysis of this system, which is defined as $\mathbf{J}_\kappa = \mathbf{J}_\mathbf{R} - \kappa^2 \mathbf{D}$, where $\mathbf{J}_\mathbf{R}$ is the Jacobian matrix of the reaction terms evaluated at the equilibrium state, and $\mathbf{D}$ is the diffusion matrix.

Ground Truth Solution

$$\boxed{\mathbf{J}_\kappa = \begin{bmatrix} -k_{\alpha\beta} - \psi''(0) - \kappa^2 k_\alpha & k_{\alpha\beta} - \sigma \\ k_{\alpha\beta}/\gamma & -k_{\alpha\beta}/\gamma - \kappa^2 k_\beta/\gamma \end{bmatrix}}$$

Model Outputs

Model: o3 (Score: 0.5)

Model Solution: $\mathrm{J}_\kappa = \begin{pmatrix} -\left(k_{\alpha\beta}+k_2\right)-k_\alpha\,\kappa^2 & k_{\alpha\beta}-\sigma \\ \dfrac{k_{\alpha\beta}}{\gamma} & -\dfrac{k_{\alpha\beta}}{\gamma}-\dfrac{k_\beta}{\gamma}\,\kappa^2 \end{pmatrix}$

Judge's Evaluation: The ground truth expression for the top-left element is:

$$-k_{\alpha\beta}-\psi''(0)-\kappa^2 k_\alpha$$

The model-generated expression for the top-left element is:

$$-\left(k_{\alpha\beta}+k_2\right)-k_\alpha\kappa^2$$

Here, the model uses k_2 instead of $\psi''(0)$. Since $\psi''(0)$ is a specific term (likely the second derivative of some function ψ evaluated at zero), replacing it with k_2 is not necessarily equivalent unless $k_2 = \psi''(0)$, which is not stated. This is a substitution of a variable that changes the meaning unless explicitly defined as equal.

The other elements match exactly in form and coefficients: - Top-right: $k_{\alpha\beta}-\sigma$ - Bottom-left: $k_{\alpha\beta}/\gamma$ - Bottom-right: $-k_{\alpha\beta}/\gamma-\kappa^2 k_\beta/\gamma$

Since the only difference is the replacement of $\psi''(0)$ by k_2, which is not justified or stated as equivalent, this counts as one minor error.

According to the rubric, one minor error (one wrong variable or missing minor constant term) warrants a partial credit score of 0.5.

Score: 0.5

Problem 2 (Paper: 2506.17853v1)

Problem Statement

Background: The dynamics of a robotic mechanical metamaterial, comprising two interacting subsystems with local rotational freedoms α and β, can be described by a set of dimensionless reaction-diffusion equations. For simplicity, overbars denoting dimensionless quantities are omitted. The governing equations are given by:

$$\alpha_{,t} = \Delta\alpha + R_\alpha(\alpha, \beta),$$
$$\gamma\beta_{,t} = d\Delta\beta + R_\beta(\alpha, \beta),$$

where Δ is the Laplacian operator, t is dimensionless time, $\gamma = \eta_\beta/\eta_\alpha$ is the ratio of temporal characteristics, and $d = k_\beta/k_\alpha$ is the ratio of diffusive characteristics. The reaction terms $R_\alpha(\alpha, \beta)$ and $R_\beta(\alpha, \beta)$ are defined as:

$$R_\alpha(\alpha, \beta) = k_{\alpha\beta}(\beta - \alpha) - \psi'(\alpha) - \sigma\beta,$$
$$R_\beta(\alpha, \beta) = k_{\alpha\beta}(\alpha - \beta),$$

where $k_{\alpha\beta}$ is a stiffness parameter for reciprocal interaction, σ is a feedback parameter for the non-reciprocal term, and $\psi'(\alpha)$ is a non-monotonic function. For the quiescent, spatially uniform state, $\mathbf{\Phi}_0 = (\alpha_0, \beta_0) = (0, 0)$, the derivative of $\psi'(\alpha)$ evaluated at $\alpha_0 = 0$ is $\psi''(0) = k_2$. To analyze the stability of this uniform state, a small perturbation of the form $\mathbf{\Phi}(\bar{x}, \bar{t}) = \mathbf{\Phi}_0 + \tilde{\mathbf{\Phi}}e^{i\kappa\bar{x}+\Omega\bar{t}}$ is introduced, where $\tilde{\mathbf{\Phi}}^{\mathrm{T}} = [\tilde{\alpha}, \tilde{\beta}]$ represents the perturbation amplitude, i is the imaginary unit, κ is the spatial frequency, and Ω is the perturbation growth rate. The system can be written compactly as $\mathbf{\Phi}_{,t} = \mathbf{D}\Delta\mathbf{\Phi} + \bar{\mathbf{R}}(\alpha, \beta)$, where $\mathbf{\Phi}^{\mathrm{T}} = [\alpha, \beta]$, $\mathbf{D} = \mathrm{diag}[1, d/\gamma]$, and $\bar{\mathbf{R}}^{\mathrm{T}}(\alpha, \beta) = [R_\alpha(\alpha, \beta), R_\beta(\alpha, \beta)/\gamma]$. The perturbation growth rates Ω are the eigenvalues of the matrix $\mathbf{J}_\kappa = \mathbf{J}_\mathbf{R} - \kappa^2\mathbf{D}$, where $\mathbf{J}_\mathbf{R}$ is the Jacobian matrix of $\bar{\mathbf{R}}(\alpha, \beta)$ evaluated at $\mathbf{\Phi}_0$.

Task: Derive the expressions for the perturbation growth rates, $\Omega_{1,2}(\kappa)$, in terms of the system parameters $k_{\alpha\beta}$, σ, k_2, d, γ, and κ.

Ground Truth Solution

$$\boxed{\Omega_{1,2}(\kappa) = \frac{1}{2}\left(\left(-\frac{k_{\alpha\beta}(\gamma+1) + k_2\gamma}{\gamma} - \frac{\gamma+d}{\gamma}\kappa^2\right) \pm \sqrt{\left(-\frac{k_{\alpha\beta}(\gamma+1) + k_2\gamma}{\gamma} - \frac{\gamma+d}{\gamma}\kappa^2\right)^2 - \frac{4}{\gamma}\left[k_{\alpha\beta}(k_2+\sigma) + (k_{\alpha\beta}(d+1)\right.}\right.}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: $\Omega_{1,2}(\kappa) = -\frac{1}{2}\left(a + \frac{b}{\gamma}\right) \pm \frac{1}{2}\sqrt{\left(a + \frac{b}{\gamma}\right)^2 - \frac{4}{\gamma}\Big[a\,b - k_{\alpha\beta}(k_{\alpha\beta} - \sigma)\Big]}$

Judge's Evaluation: The ground truth expression is a complicated formula for $\Omega_{1,2}(\kappa)$ involving terms with $k_{\alpha\beta}$, k_2, γ, d, σ, and powers of κ. It has the form:

$$\Omega_{1,2}(\kappa) = \frac{1}{2}\left(\left(-\frac{k_{\alpha\beta}(\gamma+1) + k_2\gamma}{\gamma} - \frac{\gamma+d}{\gamma}\kappa^2\right) \pm \sqrt{\left(-\frac{k_{\alpha\beta}(\gamma+1) + k_2\gamma}{\gamma} - \frac{\gamma+d}{\gamma}\kappa^2\right)^2 - \frac{4}{\gamma}\left[k_{\alpha\beta}(k_2+\sigma) + (k_{\alpha\beta}(d+1)\right.}\right.$$

The model-generated expression is:

$$\Omega_{1,2}(\kappa) = -\frac{1}{2}\left(a + \frac{b}{\gamma}\right) \pm \frac{1}{2}\sqrt{\left(a + \frac{b}{\gamma}\right)^2 - \frac{4}{\gamma}\left[ab - k_{\alpha\beta}(k_{\alpha\beta} - \sigma)\right]}$$

where a and b are not defined in the model expression, but presumably represent some combinations of parameters.

Key points of comparison:

1. **Form and structure:** - The ground truth has a factor of $\frac{1}{2}$ multiplying the entire expression inside the parentheses. - The model expression has $-\frac{1}{2}(a + \frac{b}{\gamma})$ plus/minus $\frac{1}{2}$ times the square root. - The ground truth expression has the first term inside the parentheses as a sum of two negative terms, while the model expression has a negative sign outside the first term.

2. **Variables and terms inside the square root:** - Ground truth: inside the square root is the square of the first term minus $\frac{4}{\gamma}$ times a complicated polynomial in κ. - Model: inside the square root is the square of $(a + \frac{b}{\gamma})$ minus $\frac{4}{\gamma}$ times $[ab - k_{\alpha\beta}(k_{\alpha\beta} - \sigma)]$.

3. **Dependence on κ:** - Ground truth explicitly includes κ^2 and κ^4 terms. - Model expression does not explicitly show κ dependence, and the terms inside the square root do not match the polynomial structure in the ground truth.

4. **Coefficients and signs:** - The model expression has a negative sign outside the first term, which is not consistent with the ground truth. - The terms inside the square root differ significantly in structure and content.

5. **Undefined variables:** - The model uses a and b without defining them, making it impossible to verify equivalence.

Conclusion:

The model-generated expression is structurally different, missing explicit κ dependence, has different terms inside the square root, and uses undefined variables a and b. It does not match the ground truth expression or its mathematical content.

Therefore, the model expression is **incorrect**.

Score: 0

Problem 3 (Paper: 2506.18641v1)

Problem Statement

Background: Consider an undirected and unweighted network $G = (V, E)$ with N nodes. Let $\mathbf{A}$ be its adjacency matrix, where $A_{ij} = 1$ if nodes i and j are connected, and $A_{ij} = 0$ otherwise. Let $\mathbf{D}$ be the diagonal degree matrix, where D_{ii} is the degree of node i. The Laplacian matrix of the network is defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$. The Laplacian matrix $\mathbf{L}$ is a real symmetric matrix, and thus it has N real eigenvalues, denoted as $\lambda_1, \lambda_2, \ldots, \lambda_N$.

The partition function Z_τ is a core concept for understanding information flow in networks, defined as:

$$Z_\tau = \mathrm{Tr}(e^{-\tau \mathbf{L}})$$

where τ is a positive diffusion scale parameter. The network density matrix $\boldsymbol{\rho}_\tau$ is defined as:

$$\boldsymbol{\rho}_\tau = \frac{e^{-\tau \mathbf{L}}}{Z_\tau}$$

The spectral entropy S_τ measures the diversity of information propagation paths and is defined as:

$$S_\tau = -\mathrm{Tr}(\boldsymbol{\rho}_\tau \log \boldsymbol{\rho}_\tau)$$

Task: Derive the expression for the spectral entropy S_τ in terms of the eigenvalues λ_i of the Laplacian matrix $\mathbf{L}$ and the diffusion scale parameter τ.

Ground Truth Solution

$$\boxed{S_\tau = \frac{\tau \sum_{i=1}^N \lambda_i e^{-\tau\lambda_i}}{\sum_{j=1}^N e^{-\tau\lambda_j}} + \log\left(\sum_{j=1}^N e^{-\tau\lambda_j}\right)}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\mathrm{S}_\tau = \ln\Big(\sum_{i=1}^N e^{-\tau\lambda_i}\Big) + \tau \dfrac{\sum_{i=1}^N \lambda_i\, e^{-\tau\lambda_i}}{\sum_{i=1}^N e^{-\tau\lambda_i}}$

Judge's Evaluation: The model-generated expression is:

$$S_\tau = \ln\left(\sum_{i=1}^N e^{-\tau\lambda_i}\right) + \tau \frac{\sum_{i=1}^N \lambda_i e^{-\tau\lambda_i}}{\sum_{i=1}^N e^{-\tau\lambda_i}}$$

The ground truth expression is:

$$S_\tau = \frac{\tau \sum_{i=1}^N \lambda_i e^{-\tau\lambda_i}}{\sum_{j=1}^N e^{-\tau\lambda_j}} + \log\left(\sum_{j=1}^N e^{-\tau\lambda_j}\right)$$

Comparing the two:

- The summation indices differ in the denominator of the fraction: the ground truth uses j while the model uses i. This is irrelevant since the index is a dummy variable. - The order of terms is reversed, but addition is commutative. - The logarithm is written as log in the ground truth and ln in the model. In mathematics and physics, log often denotes the natural logarithm unless otherwise specified, so this is acceptable. - The expressions are mathematically identical.

Therefore, the model-generated expression matches the ground truth exactly in mathematical content.

Score: 1

Problem 4 (Paper: 2506.18641v1)

Problem Statement

Background: Consider an undirected and unweighted network $G = (V, E)$ with N nodes. Let $\mathbf{A}$ be the adjacency matrix of the network, where $A_{ij} = 1$ if there is an edge between node i and node j, and $A_{ij} = 0$ otherwise. Let $\mathbf{D}$ be the diagonal degree matrix, where D_{ii} is the degree of node i. The Laplacian matrix of the network is defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$. Since $\mathbf{L}$ is a real symmetric matrix, it can be diagonalized, meaning there exists an orthonormal basis of eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_N$ corresponding to real eigenvalues $\lambda_1, \ldots, \lambda_N$. The matrix exponential $e^{\mathbf{M}}$ for any square matrix $\mathbf{M}$ is defined by its Taylor series expansion as $e^{\mathbf{M}} = \sum_{k=0}^{\infty} \frac{\mathbf{M}^k}{k!}$. The trace of a square matrix $\mathbf{M}$, denoted as $\mathrm{Tr}(\mathbf{M})$, is the sum of its diagonal elements. In the context of network information flow, the partition function Z_τ is defined as $Z_\tau = \mathrm{Tr}(e^{-\tau \mathbf{L}})$, where τ is a positive diffusion scale parameter.

Task: Derive the expression for the partition function Z_τ solely in terms of the eigenvalues of the Laplacian matrix $\mathbf{L}$.

Ground Truth Solution

$$\boxed{Z_\tau = \sum_{i=1}^{N} e^{-\tau \lambda_i}}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\mathrm{Z}_\tau = \sum_{i=1}^{N} e^{-\tau \lambda_i}$

Judge's Evaluation: The model-generated expression is exactly the same as the ground truth expression. Both have the same summation limits, the same exponential function, and the same variables and signs. The only difference is the presence of some extra spacing characters ( ) which do not affect the mathematical equivalence.

Score: 1

Problem 5 (Paper: 2506.19617v1)

Problem Statement

Background: Consider a network composed of two clusters, each containing N mobile Rössler chaotic oscillators. The internal dynamics of the i^{th} oscillator in the first cluster are described by:

$$\begin{cases} \dot{x}_i^1 = -x_i^2 - x_i^3 \\ \dot{x}_i^2 = x_i^1 + a x_i^2 + \epsilon \sum\limits_{j=1}^{N} g_{ij}^1(t)\left(x_j^2 - x_i^2\right) + \mu D_{xy}\left(\overline{y}^2 - x_i^2\right) \\ \dot{x}_i^3 = b + x_i^3\left(x_i^1 - c\right) \end{cases}$$

And for the i^{th} oscillator in the second cluster:

$$\begin{cases} \dot{y}_i^1 = -y_i^2 - y_i^3 \\ \dot{y}_i^2 = y_i^1 + a y_i^2 + \epsilon \sum\limits_{j=1}^{N} g_{ij}^2(t)\left(y_j^2 - y_i^2\right) + \mu D_{yx}\left(\overline{x}^2 - y_i^2\right) \\ \dot{y}_i^3 = b + y_i^3\left(y_i^1 - c\right) \end{cases}$$

Here, a, b, c are constant Rössler parameters, ϵ is the intra-cluster coupling coefficient, and μ is the inter-cluster coupling coefficient. $\overline{x}^2 = \frac{1}{m_1}\sum_{j=1}^{m_1} g_{ij}^1 x_j^2$ and $\overline{y}^2 = \frac{1}{m_2}\sum_{j=1}^{m_2} g_{ij}^2 y_j^2$ represent the local centers of mass for the first and second cluster, respectively, where m_1 and m_2 denote the number of nodes within the vision size in each cluster.

The intra-cluster connectivity $g_{ij}^k(t)$ between oscillators i and j in cluster k is defined as:

$$g_{ij}^k = \begin{cases} 1 \ if \ d_{ij}^k \le d_0^k \\ 0 \ otherwise, \end{cases}$$

where $d_{ij}^k(t) = \sqrt{\left(\eta_i^k(t) - \eta_j^k(t)\right)^2 + \left(\xi_i^k(t) - \xi_j^k(t)\right)^2}$ is the Euclidean distance between agents i and j in cluster k, and d_0^k is the agent vision size control parameter.

The inter-cluster connectivity functions D_{XY} and D_{YX} are defined as:

$$D_{XY} = D_{YX} = \begin{cases} 1 \ if \ s_{XY}, s_{YX} \le s_0 \\ 0 \ otherwise \end{cases}$$

where $s_{XY} = s_{YX} = \sqrt{(X^1(t) - X^2(t))^2 + (Y^1(t) - Y^2(t))^2}$ is the Euclidean distance between the lower-left corners of the clusters, and s_0 is the maximum distance threshold for inter-cluster connectivity.

To analyze the stability of synchronization between clusters, an error system is defined as $e_i^k = x_i^k - y_i^k$ for $k = 1, 2, 3$ and $i = 1, 2, ..., N$. The dynamics of the error system are given by:

$$\begin{cases} \dot{e}_i^1 = -e_i^2 - e_i^3 \\ \dot{e}_i^2 = e_i^1 + a e_i^2 + \epsilon \sum\limits_{j}^{N} g_{ij}^1\left(e_j^2 - e_i^2\right) + \epsilon f_{1i} + \\ \qquad \mu D_{xy}\left[-e_i^2 - \frac{1}{m_2}\sum\limits_{j}^{m_2} g_{ij}^2 e_j^2 + f_{2i}\right] \\ \dot{e}_i^3 = -c e_i^3 + e_i^1 y_i^3 + e_i^3 y_i^1 \end{cases}$$

where $f_{1i} = \sum\limits_{j}^{N}\left(g_{ij}^1 - g_{ij}^2\right)\left(y_j^2 - y_i^2\right)$ and $f_{2i} = \frac{1}{m_1}\sum\limits_{j}^{m_1} g_{ij}^1 y_j^2 - \frac{1}{m_2}\sum\limits_{j}^{m_2} g_{ij}^2 y_j^2$. For simplicity, the non-linear term $e_i^1 e_i^3$ in $\dot{e}_i^3$ is disregarded due to its negligible size as the system approaches synchronization.

A Lyapunov function candidate V_i is defined as:

$$V_i = \frac{1}{2}\left(\left(e_i^1\right)^2 + \left(e_i^2\right)^2 + \left(e_i^3\right)^2\right) + \int\limits_0^t \left(\gamma_1\left(e_i^1\right)^2 + \gamma_2\left(e_i^3\right)^2\right) dt$$

where γ_1 and γ_2 are parameters to be determined. The time derivative of this Lyapunov function candidate is:

$$\dot{V}_i = \left(-1 + y_i^3\right) e_i^1 e_i^3 + a\left(e_i^2\right)^2 - c\left(e_i^3\right)^2 + y_i^1\left(e_i^3\right)^2 + \gamma_1\left(e_i^1\right)^2$$
$$-\gamma_1\left(e_i^1(0)\right)^2 + \gamma_2\left(e_i^3\right)^2 - \gamma_2\left(e_i^3(0)\right)^2 + \epsilon \sum_{j\neq i}^{N} g_{ij}^1 e_j^2 e_i^2 -$$
$$\epsilon\left(e_i^2\right)^2 \sum_{j\neq i}^{N} g_{ij}^1 + e_i^2 \epsilon f_{1i} - \mu D_{xy}\left(1 + \tfrac{1}{m_2}\right)\left(e_i^2\right)^2 -$$
$$\tfrac{\mu D_{xy}}{2m_2} \sum_{j}^{m_2} g_{ij}^2 e_j^2 e_i^2 + e_i^2 \mu D_{xy} f_{2i}$$

Assume $y_m^1 = \max\left|y_i^1\right|$ and $y_m^3 = \max\left|y_i^3\right|$. Use the mathematical property $|A|\,|B| \leq \frac{1}{2}\left(A^2 + B^2\right)$ for real numbers A and B. Assume there exist real positive constants Z_1 and Z_2 such that $\sum_{\substack{j=1,\\ j\neq i}}^{N} g_{ij}^1\left(e_j^2\right)^2 \leq Z_1\left(e_i^2\right)^2$ and $\sum_{\substack{j=1,\\ j\neq i}}^{m_2} g_{ij}^2\left(e_j^2\right)^2 \leq Z_2\left(e_i^2\right)^2$. Let $Z = \max\left(Z_1, Z_2\right)$. Set $\gamma_1 = \frac{1}{2}\left(1 + y_m^3\right)$ and $\gamma_2 = y_m^1 - c + \frac{1}{2}\left(1 + y_m^3\right)$.

Task: Starting from the given expression for $\dot{V}_i$, apply the stated mathematical property $|A|\,|B| \leq \frac{1}{2}\left(A^2 + B^2\right)$ to all product terms involving error components (e.g., $e_i^1 e_i^3$, $e_j^2 e_i^2$, $e_i^2 f_{1i}$, $e_i^2 f_{2i}$). Then, substitute the definitions of γ_1 and γ_2, and use the Z bound to replace the terms involving $\sum g_{ij}^1 (e_j^2)^2$ and $\sum g_{ij}^2 (e_j^2)^2$. Finally, collect all terms proportional to $\left(e_i^2\right)^2$ to determine the expression for $f3_{max}$.

Ground Truth Solution

$$\boxed{f3_{max} = a + \frac{1}{2}\left(\epsilon + \frac{\mu D_{xy}}{m_2}\right) Z - \mu D_{xy}\left(\frac{1}{2} + \frac{1}{m_2}\right) - \frac{\epsilon}{2} \sum_{\substack{j=1,\\ j\neq i}}^{N} g_{ij}^1}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: $\mathrm{f3}_{\max} \;=\; a \;+\; \frac{\epsilon}{2}\,(1+Z) \;+\; \frac{\mu D_{xy}}{4\,m_2}\,(Z - m_2 - 4)$

Judge's Evaluation: The ground truth expression is:

$$f3_{max} = a + \frac{1}{2}\left(\epsilon + \frac{\mu D_{xy}}{m_2}\right) Z - \mu D_{xy}\left(\frac{1}{2} + \frac{1}{m_2}\right) - \frac{\epsilon}{2} \sum_{j=1, j\neq i}^{N} g_{ij}^1$$

The model-generated expression is:

$$f_{3_{\max}} = a + \frac{\epsilon}{2}(1 + Z) + \frac{\mu D_{xy}}{4m_2}(Z - m_2 - 4)$$

—

Comparison:

1. The ground truth has four terms: - a - $\frac{1}{2}\left(\epsilon + \frac{\mu D_{xy}}{m_2}\right) Z$ - $-\mu D_{xy}\left(\frac{1}{2} + \frac{1}{m_2}\right)$ - $-\frac{\epsilon}{2}\sum_{j\neq i} g_{ij}^1$

2. The model expression has three terms: - a - $\frac{\epsilon}{2}(1+Z)$ - $\frac{\mu D_{xy}}{4m_2}(Z - m_2 - 4)$

3. The model expression does not include the summation term $-\frac{\epsilon}{2}\sum_{j\neq i} g_{ij}^1$ at all, which is a significant omission.

4. The model's $\frac{\epsilon}{2}(1+Z)$ term does not match the ground truth's $\frac{1}{2}\epsilon Z$ term; the model adds an extra $\frac{\epsilon}{2}$ (the "1" inside the parentheses), which is not present in the ground truth.

5. The model's $\frac{\mu D_{xy}}{4m_2}(Z - m_2 - 4)$ term is not obviously equivalent to the ground truth's two separate terms involving μD_{xy}. The ground truth has $\frac{1}{2}\frac{\mu D_{xy}}{m_2} Z$ inside the first big parenthesis and a separate subtraction term $-\mu D_{xy}\left(\frac{1}{2} + \frac{1}{m_2}\right)$. The model's term is a different combination and includes constants like -4 which do not appear in the ground truth.

6. Overall, the model expression is structurally different, missing a key summation term, and has different coefficients and constants.

—

Conclusion:

The model expression is not mathematically equivalent to the ground truth. It misses the summation term entirely and has different coefficients and constants in the other terms. This is more than one minor error.

—

Score: 0

Problem 6 (Paper: 2506.19617v1)

Problem Statement

Background: Consider a system composed of two clusters of N mobile Rössler chaotic oscillators. The internal dynamics of the i^{th} oscillator in the first cluster are described by the state variables $\left(x_i^1(t), x_i^2(t), x_i^3(t)\right)$, and similarly, for the second cluster, the state variables are $\left(y_i^1(t), y_i^2(t), y_i^3(t)\right)$. The dynamics are given by: For the first cluster:

$$\begin{cases} \dot{x}_i^1 = -x_i^2 - x_i^3 \\ \dot{x}_i^2 = x_i^1 + a x_i^2 + \epsilon \sum\limits_{j=1}^{N} g_{ij}^1(t)\left(x_j^2 - x_i^2\right) + \mu D_{xy}\left(\overline{y}^2 - x_i^2\right) \\ \dot{x}_i^3 = b + x_i^3\left(x_i^1 - c\right) \end{cases}$$

For the second cluster:

$$\begin{cases} \dot{y}_i^1 = -y_i^2 - y_i^3 \\ \dot{y}_i^2 = y_i^1 + a y_i^2 + \epsilon \sum\limits_{j=1}^{N} g_{ij}^2(t)\left(y_j^2 - y_i^2\right) + \mu D_{yx}\left(\overline{x}^2 - y_i^2\right) \\ \dot{y}_i^3 = b + y_i^3\left(y_i^1 - c\right) \end{cases}$$

Here, ϵ is the intra-cluster coupling coefficient, and μ is the inter-cluster coupling coefficient. $\overline{x}^2 = \frac{1}{m_1}\sum_{j=1}^{m_1} g_{ij}^1 x_j^2$ and $\overline{y}^2 = \frac{1}{m_2}\sum_{j=1}^{m_2} g_{ij}^2 y_j^2$ represent the local centers of mass for the first and second cluster, respectively, where m_1 and m_2 are the number of nodes within the vision size in each cluster. The intra-cluster connectivity matrix g_{ij}^k is 1 if the Euclidean distance d_{ij}^k between agents i and j in cluster k is less than or equal to d_0^k, and 0 otherwise. The inter-cluster connectivity functions D_{XY} and D_{YX} are 1 if the Euclidean distance s_{XY} between the centers of mass of the two clusters is less than or equal to s_0, and 0 otherwise. For this problem, assume $D_{XY} = D_{YX} = 1$.

To analyze the stability of synchronization between the clusters, an error system $e_i^k = x_i^k - y_i^k$ is defined for $k = 1, 2, 3$ and $i = 1, 2, ..., N$. The error system equations are:

$$\begin{cases} \dot{e}_i^1 = -e_i^2 - e_i^3 \\ \dot{e}_i^2 = e_i^1 + a e_i^2 + \epsilon \sum\limits_{j}^{N} g_{ij}^1\left(e_j^2 - e_i^2\right) + \epsilon f_{1i} + \mu D_{xy}\left[-e_i^2 - \frac{1}{m_2}\sum\limits_{j}^{m_2} g_{ij}^2 e_j^2 + f_{2i}\right] \\ \dot{e}_i^3 = -c e_i^3 + e_i^1 y_i^3 + e_i^3 y_i^1 \end{cases}$$

where $f_{1i} = \sum\limits_{j}^{N}\left(g_{ij}^1 - g_{ij}^2\right)\left(y_j^2 - y_i^2\right)$ and $f_{2i} = \frac{1}{m_1}\sum\limits_{j}^{m_1} g_{ij}^1 y_j^2 - \frac{1}{m_2}\sum\limits_{j}^{m_2} g_{ij}^2 y_j^2$. For the purpose of this derivation, the non-linear term $e_i^1 e_i^3$ in $\dot{e}_i^3$ is disregarded due to its negligible size.

A Lyapunov function candidate V_i is defined as:

$$V_i = \frac{1}{2}\left(\left(e_i^1\right)^2 + \left(e_i^2\right)^2 + \left(e_i^3\right)^2\right) + \int\limits_0^t \left(\gamma_1\left(e_i^1\right)^2 + \gamma_2\left(e_i^3\right)^2\right) dt$$

The time derivative of this Lyapunov function candidate is given by:

$$\begin{aligned} \dot{V}_i = &\left(-1 + y_i^3\right) e_i^1 e_i^3 + a\left(e_i^2\right)^2 - c\left(e_i^3\right)^2 + y_i^1\left(e_i^3\right)^2 + \gamma_1\left(e_i^1\right)^2 \\ &-\gamma_1\left(e_i^1(0)\right)^2 + \gamma_2\left(e_i^3\right)^2 - \gamma_2\left(e_i^3(0)\right)^2 + \epsilon\sum\limits_{j\neq i}^{N} g_{ij}^1 e_j^2 e_i^2 - \\ &\epsilon\left(e_i^2\right)^2 \sum\limits_{j\neq i}^{N} g_{ij}^1 + e_i^2 \epsilon f_{1i} - \mu D_{xy}\left(1 + \frac{1}{m_2}\right)\left(e_i^2\right)^2 - \\ &\frac{\mu D_{xy}}{2m_2}\sum\limits_{j}^{m_2} g_{ij}^2 e_j^2 e_i^2 + e_i^2 \mu D_{xy} f_{2i} \end{aligned}$$

Let $y_m^1 = \max\left|y_i^1\right|$ and $y_m^3 = \max\left|y_i^3\right|$. Applying the inequality $|a|\,|b| \leq \frac{1}{2}\left(a^2 + b^2\right)$ to maximize the derivative yields:

$$\begin{aligned}&\dot{V}_i \le \tfrac{1}{2}\left(1+y_m^3\right)\left(e_i^1\right)^2+\left(y_m^1-c+\tfrac{1}{2}\left(1+y_m^3\right)\right)\left(e_i^3\right)^2+\\ &\left[a+\tfrac{\epsilon}{2}-\mu D_{xy}\left(\tfrac{1}{2}+\tfrac{1}{m_2}-\tfrac{1}{4m_2}\sum_{\substack{j=1,\\ j\neq i}}^{m_2} g_{ij}^2\right)-\tfrac{\epsilon}{2}\sum_{\substack{j=1,\\ j\neq i}}^{N} g_{ij}^1\right]\left(e_i^2\right)^2\\ &+\gamma_1\left(e_i^1\right)^2-\gamma_1\left(e_i^1\left(0\right)\right)^2+\gamma_2\left(e_i^3\right)^2-\gamma_2\left(e_i^3\left(0\right)\right)^2+\tfrac{\epsilon}{2}\left(f_{1i}\right)^2\\ &+\tfrac{\epsilon}{2}\sum_{\substack{j=1,\\ j\neq i}}^{N} g_{ij}^1\left(e_j^2\right)^2+\tfrac{\mu D_{xy}}{4m_2}\sum_{\substack{j=1,\\ j\neq i}}^{m_2} g_{ij}^2\left(e_j^2\right)^2+\tfrac{\mu D_{xy}}{2}\left(f_{2i}\right)^2\end{aligned}$$

Assume that all m_2 oscillators within the visual size are inherently connected to oscillator i, such that the degree of this node is naturally m_2, which implies $\frac{1}{4m_2}\sum_{\substack{j=1,\\ j\neq i}}^{m_2} g_{ij}^2=\frac{1}{4}$. Assume there exist real positive constants Z_1 and Z_2 such that $\sum_{\substack{j=1,\\ j\neq i}}^{N} g_{ij}^1\left(e_j^2\right)^2 \le Z_1\left(e_i^2\right)^2$ and $\sum_{\substack{j=1,\\ j\neq i}}^{N} g_{ij}^2\left(e_j^2\right)^2 \le Z_2\left(e_i^2\right)^2$. Let $Z=\max\left(Z_1,Z_2\right)$. This leads to the bound:

$$\left[\frac{\epsilon}{2}Z_1+\frac{\mu D_{xy}}{2m_2}Z_2\right]\left(e_i^2\right)^2 \le \left[\frac{\epsilon}{2}+\frac{\mu D_{xy}}{2m_2}\right]Z\left(e_i^2\right)^2.$$

Set the parameters $\gamma_1=\frac{1}{2}\left(1+y_m^3\right)$ and $\gamma_2=y_m^1-c+\frac{1}{2}\left(1+y_m^3\right)$. Let $\gamma=\max\left(\gamma_1,\,\gamma_2\right)$ and $\|e_i\|^2=(e_i^1)^2+(e_i^2)^2+(e_i^3)^2$.

Task: Using the provided expressions and assumptions, derive the upper bound for the time derivative of the Lyapunov function candidate, $\dot{V}_i$.

Ground Truth Solution

$$\boxed{\begin{aligned}&\dot{V}_i \le 2\gamma\|e_i\|^2-\gamma\|e_i\left(0\right)\|^2+\tfrac{\epsilon}{2}\left(f_{1i}\right)^2+\tfrac{\mu D_{xy}}{2}\left(f_{2i}\right)^2+\\ &\left[a+\tfrac{1}{2}\left(\epsilon+\tfrac{\mu D_{xy}}{m_2}\right)Z-\mu D_{xy}\left(\tfrac{1}{2}+\tfrac{1}{m_2}\right)-\tfrac{\epsilon}{2}\sum_{\substack{j=1,\\ j\neq i}}^{N} g_{ij}^1\right]\left(e_i^2\right)^2\end{aligned}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution:

$$\begin{aligned}\dot{V}_i &\le \gamma\big[(e_i^1)^2+(e_i^3)^2\big]\\ &\quad+\Big[a+\frac{\epsilon}{2}-\mu\Big(\frac{1}{2}+\frac{1}{m_2}-\frac{1}{4}\Big)-\frac{\epsilon}{2}\sum_{j\neq i} g_{ij}^1+\Big(\frac{\epsilon}{2}+\frac{\mu}{2m_2}\Big)Z\Big](e_i^2)^2\\ &\quad+\frac{\epsilon}{2}f_{1i}^2+\frac{\mu}{2}f_{2i}^2-\gamma_1\big(e_i^1(0)\big)^2-\gamma_2\big(e_i^3(0)\big)^2.\end{aligned}$$

Judge's Evaluation: Let's analyze the two expressions term-by-term.

—

Ground Truth Expression:

$$\begin{aligned}&\dot{V}_i \le 2\gamma\|e_i\|^2-\gamma\|e_i(0)\|^2+\tfrac{\epsilon}{2}(f_{1i})^2+\tfrac{\mu D_{xy}}{2}(f_{2i})^2+\\ &\left[a+\tfrac{1}{2}\left(\epsilon+\tfrac{\mu D_{xy}}{m_2}\right)Z-\mu D_{xy}\left(\tfrac{1}{2}+\tfrac{1}{m_2}\right)-\tfrac{\epsilon}{2}\sum_{j\neq i} g_{ij}^1\right](e_i^2)^2\end{aligned}$$

—

Model-Generated Expression:

$$
\begin{aligned}
\dot{V}_i \leq & \gamma \left[(e_i^1)^2 + (e_i^3)^2\right] \\
& + \left[a + \frac{\epsilon}{2} - \mu\left(\frac{1}{2} + \frac{1}{m_2} - \frac{1}{4}\right) - \frac{\epsilon}{2}\sum_{j\neq i} g_{ij}^1 + \left(\frac{\epsilon}{2} + \frac{\mu}{2m_2}\right) Z\right] (e_i^2)^2 \\
& + \frac{\epsilon}{2} f_{1i}^2 + \frac{\mu}{2} f_{2i}^2 - \gamma_1 (e_i^1(0))^2 - \gamma_2 (e_i^3(0))^2
\end{aligned}
$$

—

Detailed Comparison:

1. **Terms involving $\|e_i\|^2$ and $\|e_i(0)\|^2$:**

- Ground truth: $2\gamma\|e_i\|^2 - \gamma\|e_i(0)\|^2$.

- Model: $\gamma[(e_i^1)^2 + (e_i^3)^2] - \gamma_1(e_i^1(0))^2 - \gamma_2(e_i^3(0))^2$.

- The ground truth uses the norm squared $\|e_i\|^2$, presumably including components e_i^1, e_i^2, e_i^3, but the model only includes e_i^1 and e_i^3 in the first term and excludes e_i^2.

- The ground truth has a single γ coefficient for the initial error norm squared, while the model splits it into γ_1 and γ_2.

- The factor of 2 in front of $\gamma\|e_i\|^2$ in the ground truth is missing in the model (model has γ only).

- The model does not include the initial error term for $e_i^2(0)$.

2. **Terms involving f_{1i}^2 and f_{2i}^2:**

- Ground truth: $\frac{\epsilon}{2}(f_{1i})^2 + \frac{\mu D_{xy}}{2}(f_{2i})^2$.

- Model: $\frac{\epsilon}{2} f_{1i}^2 + \frac{\mu}{2} f_{2i}^2$.

- The model is missing the factor D_{xy} multiplying μ in the f_{2i}^2 term.

3. **Coefficient of $(e_i^2)^2$:**

- Ground truth:

$$
a + \frac{1}{2}\left(\epsilon + \frac{\mu D_{xy}}{m_2}\right) Z - \mu D_{xy}\left(\frac{1}{2} + \frac{1}{m_2}\right) - \frac{\epsilon}{2}\sum_{j\neq i} g_{ij}^1
$$

- Model:

$$
a + \frac{\epsilon}{2} - \mu\left(\frac{1}{2} + \frac{1}{m_2} - \frac{1}{4}\right) - \frac{\epsilon}{2}\sum_{j\neq i} g_{ij}^1 + \left(\frac{\epsilon}{2} + \frac{\mu}{2m_2}\right) Z
$$

- Differences:

- The model has an extra $\frac{\epsilon}{2}$ term outside the Z term, which is not present in the ground truth.

- The model's μ term inside the parentheses is $\frac{1}{2} + \frac{1}{m_2} - \frac{1}{4}$, whereas the ground truth has $\mu D_{xy}\left(\frac{1}{2} + \frac{1}{m_2}\right)$.

- The model is missing the factor D_{xy} multiplying μ in these terms.

- The Z term in the model is $\left(\frac{\epsilon}{2} + \frac{\mu}{2m_2}\right) Z$, while the ground truth has $\frac{1}{2}\left(\epsilon + \frac{\mu D_{xy}}{m_2}\right) Z$.

- Again, the model is missing D_{xy} in the μ term.

4. **Summary of discrepancies:**

- Missing factor D_{xy} in all μ-related terms in the model.

- The model splits γ into γ_1 and γ_2 and only includes e_i^1 and e_i^3 components, missing e_i^2.

- The factor 2 in front of $\gamma\|e_i\|^2$ is missing in the model.

- The model has an extra $\frac{\epsilon}{2}$ term in the coefficient of $(e_i^2)^2$ outside the Z term.

- The model has a $-\frac{1}{4}$ inside the μ parentheses, which is not in the ground truth.

—

Conclusion:

The model-generated expression has multiple minor but significant errors:

- Missing D_{xy} factors in μ-terms.

- Incorrect coefficients and extra terms in the $(e_i^2)^2$ coefficient.

- Missing factor 2 in the $\gamma\|e_i\|^2$ term.

- Splitting γ into γ_1 and γ_2 and missing e_i^2 terms.

These are more than one minor error, and the expression is not mathematically equivalent to the ground truth.

—

Score: 0

—

Justification:

The model expression is not equivalent to the ground truth. It misses key multiplicative factors (D_{xy}) in multiple terms, has incorrect coefficients, and omits some components of the error vector. These errors are more than minor and prevent the expression from being considered nearly correct.

Problem 7 (Paper: 2506.20163v1)

Problem Statement

Background: A simplicial complex is a mathematical structure built by attaching simplices (nodes, links, triangles, etc.) by their faces. A 0-simplex is a node, and a 1-simplex is a link. For a network consisting of N_n nodes and N_l links, the boundary operator $\mathbf{B}_1$ is an $N_n \times N_l$ matrix that maps 1-simplices to 0-simplices. Its elements are defined as $\mathbf{B}_1(i,j) = 1$ if the i-th 0-simplex is a face of the j-th 1-simplex with the same orientation, -1 if with opposite orientation, and 0 otherwise. Topological signals (or cochains) are functions defined on simplices. A 0-cochain $\vec{u}$ is a column vector of dimension N_n representing signals on nodes, and a 1-cochain $\vec{v}$ is a column vector of dimension N_l representing signals on links. The topological spinor is defined as $\vec{w} = \begin{bmatrix} \vec{u} \\ \vec{v} \end{bmatrix}$. The Dirac-Bianconi operator $\mathcal{D}$ couples cochains of adjacent dimensions and is given by $\mathcal{D} = \begin{pmatrix} 0 & \mathbf{B}_1 \\ \mathbf{B}_1^\top & 0 \end{pmatrix}$. The dynamics of a Dirac-Bianconi driven oscillator, where oscillations are induced by this coupling, are generally described by $\dot{\vec{w}} = \vec{F}(\vec{w}, \mathcal{D}\vec{w})$, which expands to $\begin{bmatrix} \dot{\vec{u}} \\ \dot{\vec{v}} \end{bmatrix} = \begin{bmatrix} \vec{f}(\vec{u}, \boldsymbol{B}_1\vec{v}) \\ \vec{g}(\vec{v}, \boldsymbol{B}_1^\top \vec{u}) \end{bmatrix}$. Consider a specific Dirac-Bianconi driven oscillator, inspired by the FitzHugh-Nagumo model, with $N_n = 3$ nodes and $N_l = 3$ links. The dynamics of its 0-cochains u_i and 1-cochains v_j are given by: $\dot{u}_i = u_i - u_i^3 - (\boldsymbol{B}_1\vec{v})_i + I_i$ $\dot{v}_j = \delta_j((\boldsymbol{B}_1^\top \vec{u})_j - b_j v_j + \alpha_j a_j)$ Here, I_i, δ_j, b_j, a_j are constant parameters, and α_j is a coefficient that accounts for the change of orientation of the j-th link. For this specific system, the boundary operator $\boldsymbol{B}_1$ is given by: $\boldsymbol{B}_1 = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$. Assume $\alpha_j = 1$ for all j.

Task: Derive the full system of six coupled ordinary differential equations that describe the dynamics of the Dirac-Bianconi driven oscillator for the given network structure and boundary operator.

Ground Truth Solution

$$\boxed{\begin{cases} \dot{u}_1 = u_1 - u_1^3 - (v_1 - v_3) + I_1 \\ \dot{u}_2 = u_2 - u_2^3 - (-v_1 + v_2) + I_2 \\ \dot{u}_3 = u_3 - u_3^3 - (-v_2 + v_3) + I_3 \\ \dot{v}_1 = \delta_1(u_1 - u_2 - b_1 v_1 + a_1) \\ \dot{v}_2 = \delta_2(u_2 - u_3 - b_2 v_2 + a_2) \\ \dot{v}_3 = \delta_3(-u_1 + u_3 - b_3 v_3 + a_3) \end{cases}}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution:

$$\begin{aligned} \dot{u}_1 &= u_1 - u_1^3 - \bigl(v_1 - v_3\bigr) + I_1 = u_1 - u_1^3 - v_1 + v_3 + I_1, \\ \dot{u}_2 &= u_2 - u_2^3 - \bigl(-v_1 + v_2\bigr) + I_2 = u_2 - u_2^3 + v_1 - v_2 + I_2, \\ \dot{u}_3 &= u_3 - u_3^3 - \bigl(-v_2 + v_3\bigr) + I_3 = u_3 - u_3^3 + v_2 - v_3 + I_3, \\ \dot{v}_1 &= \delta_1\Bigl[(u_1 - u_2) - b_1 v_1 + a_1\Bigr], \\ \dot{v}_2 &= \delta_2\Bigl[(u_2 - u_3) - b_2 v_2 + a_2\Bigr], \\ \dot{v}_3 &= \delta_3\Bigl[(-u_1 + u_3) - b_3 v_3 + a_3\Bigr]. \end{aligned}$$

Judge's Evaluation: The model-generated expression matches the ground truth exactly in all components. Each $\dot{u}_i$ and $\dot{v}_i$ term is identical, including signs, coefficients, and grouping. The model even expands the parentheses correctly for the $\dot{u}_i$ terms, showing equivalence explicitly. There are no missing terms, incorrect signs, or coefficient errors.

Score: 1

Problem 8 (Paper: 2506.20163v1)

Problem Statement

Background: Consider a Dirac-Bianconi driven oscillator, a system where dynamical variables are defined on nodes (0-cochains, denoted by $\vec{u} \in \mathbb{R}^{N_n}$) and links (1-cochains, denoted by $\vec{v} \in \mathbb{R}^{N_l}$) of a network. The state of the system is described by the topological spinor $\vec{w} = (\vec{u}^\top, \vec{v}^\top)^\top$, which has dimension $N_n + N_l$. The dynamics of this system, known as the Dirac-Bianconi FitzHugh-Nagumo (DBFHN) model, is given by:

$$\begin{cases} \dot{u}_i = u_i - u_i^3 - (\boldsymbol{B}_1 \vec{v})_i + I_i, \quad \text{for } i = 1, ..., N_n \\ \dot{v}_j = \delta_j((\boldsymbol{B}_1^\top \vec{u})_j - b_j v_j + \alpha_j a_j), \quad \text{for } j = 1, ..., N_l \end{cases}$$

Here, $\boldsymbol{B}_1$ is the $N_n \times N_l$ boundary operator matrix, and $\boldsymbol{B}_1^\top$ is its transpose (the coboundary operator). The terms $(\boldsymbol{B}_1 \vec{v})_i$ and $(\boldsymbol{B}_1^\top \vec{u})_j$ represent the i-th component of the vector $\boldsymbol{B}_1 \vec{v}$ and the j-th component of the vector $\boldsymbol{B}_1^\top \vec{u}$, respectively. The parameters $I_i, \delta_j, b_j, a_j, \alpha_j$ are constants. The dynamics can be compactly written as $\dot{\vec{w}} = \vec{F}(\vec{w})$. The Jacobian matrix $J(\vec{w})$ of the vector field $\vec{F}(\vec{w})$ is defined by its elements $J_{ab} = \frac{\partial F_a}{\partial w_b}$, where F_a is the a-th component of $\vec{F}(\vec{w})$ and w_b is the b-th component of $\vec{w}$. The vector $\vec{w}$ is ordered such that its first N_n components are $u_1, ..., u_{N_n}$ and its next N_l components are $v_1, ..., v_{N_l}$.

Task: Derive the full $(N_n + N_l) \times (N_n + N_l)$ Jacobian matrix $J(\vec{w})$ for the Dirac-Bianconi FitzHugh-Nagumo (DBFHN) model, expressing it in terms of $\vec{u}$, $\vec{v}$, the boundary operator $\boldsymbol{B}_1$, and the model parameters.

Ground Truth Solution

$$\boxed{J(\vec{w}) = \begin{pmatrix} \operatorname{diag}(1 - 3u_1^2, ..., 1 - 3u_{N_n}^2) & -\boldsymbol{B}_1 \\ \operatorname{diag}(\delta_1, ..., \delta_{N_l}) \boldsymbol{B}_1^\top & -\operatorname{diag}(\delta_1 b_1, ..., \delta_{N_l} b_{N_l}) \end{pmatrix}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: J()= $\begin{pmatrix} \boldsymbol{D}_u & -\boldsymbol{B}_1 \\ \boldsymbol{\Delta}\, \boldsymbol{B}_1^\top & -\boldsymbol{\Delta}\, \boldsymbol{B} \end{pmatrix}$

Judge's Evaluation: The ground truth expression is:

$$J(\vec{w}) = \begin{pmatrix} \operatorname{diag}(1 - 3u_1^2, ..., 1 - 3u_{N_n}^2) & -\boldsymbol{B}_1 \\ \operatorname{diag}(\delta_1, ..., \delta_{N_l}) \boldsymbol{B}_1^\top & -\operatorname{diag}(\delta_1 b_1, ..., \delta_{N_l} b_{N_l}) \end{pmatrix}$$

The model-generated expression is:

$$J(\vec{w}) = \begin{pmatrix} \boldsymbol{D}_u & -\boldsymbol{B}_1 \\ \boldsymbol{\Delta}\, \boldsymbol{B}_1^\top & -\boldsymbol{\Delta}\, \boldsymbol{B} \end{pmatrix}$$

Comparison:

- The top-left block in the ground truth is a diagonal matrix with entries $1 - 3u_i^2$. The model uses $\boldsymbol{D}_u$, presumably representing the same diagonal matrix. This is acceptable if $\boldsymbol{D}_u = \operatorname{diag}(1 - 3u_1^2, ..., 1 - 3u_{N_n}^2)$.
- The top-right block is $-\boldsymbol{B}_1$ in both expressions, matching exactly.
- The bottom-left block is $\operatorname{diag}(\delta_i) \boldsymbol{B}_1^\top$ in the ground truth, and $\boldsymbol{\Delta} \boldsymbol{B}_1^\top$ in the model. Assuming $\boldsymbol{\Delta} = \operatorname{diag}(\delta_1, ..., \delta_{N_l})$, this matches.
- The bottom-right block is $-\operatorname{diag}(\delta_i b_i)$ in the ground truth, but the model writes $-\boldsymbol{\Delta} \boldsymbol{B}$.

Here is the key discrepancy:

- The ground truth has a diagonal matrix with entries $\delta_i b_i$.
- The model uses $\boldsymbol{\Delta B}$, which is a product of a diagonal matrix $\boldsymbol{\Delta}$ and a matrix $\boldsymbol{B}$. This is not necessarily diagonal, and $\boldsymbol{B}$ is not specified as diagonal or equal to $\operatorname{diag}(b_i)$.
- Also, the ground truth uses $\boldsymbol{B}_1$ in the top-right and bottom-left blocks, but the model uses $\boldsymbol{B}$ in the bottom-right block, which is inconsistent.

Therefore, the bottom-right block in the model expression is not equivalent to the ground truth.

Since this is more than a minor error (it changes the structure from diagonal to a general matrix product and changes the matrix from $\boldsymbol{B}_1$ to $\boldsymbol{B}$), the model expression is not fully correct.

However, the rest of the expression is correct or matches the ground truth.

Given that there is one significant error (bottom-right block), the score should be 0.

Score: 0

Problem 9 (Paper: 2506.20572v1)

Problem Statement

Background: Consider a network of channels represented by an undirected graph where $(i, j) \in E$ denotes an edge connecting node i to node j. Fluid flows through these channels with a flux Q_{ij} along an edge of length L_{ij} and conductivity D_{ij}. The dissipated energy per unit time in this network, denoted by $\mathcal{P}$, is given by the sum over all edges:

$$\mathcal{P} = \sum_{(i,j)\in E} \frac{Q_{ij}^2 L_{ij}}{D_{ij}}$$

Each channel (i, j) has a volume $V_{ij} = \beta L_{ij}\sqrt{D_{ij}}$, where β is a constant related to the fluid properties. The total fluid volume in the network, $V = \sum_{(i,j)\in E} V_{ij}$, is conserved and remains constant. The function $g(Q_{ij})$, which describes the elasticity of the channels, is obtained by minimizing the dissipated energy $\mathcal{P}$ with respect to the conductivities D_{ij}, subject to the constraint that the total fluid volume V is constant.

Task: Derive the analytical expression for the function $g(Q_{ij})$.

Ground Truth Solution

$$\boxed{g(Q_{ij}) = Q_{ij}^{2/3}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: $\mathrm{g}(\mathrm{Q}_{ij}) = D_{ij} = \left(\frac{V}{\beta \sum_{(k,l)\in E} L_{kl} Q_{kl}^{2/3}}\right)^2 Q_{ij}^{4/3}$

Judge's Evaluation: The ground truth expression is simply $g(Q_{ij}) = Q_{ij}^{2/3}$.

The model-generated expression is:

$$g(Q_{ij}) = D_{ij} = \left(\frac{V}{\beta \sum_{(k,l)\in E} L_{kl} Q_{kl}^{2/3}}\right)^2 Q_{ij}^{4/3}$$

This is a much more complicated expression involving additional variables V, β, L_{kl}, and a summation over edges E. It also raises Q_{ij} to the power 4/3 instead of 2/3, and includes a squared prefactor.

Since the model-generated expression is not mathematically equivalent to the ground truth expression, and contains multiple differences (extra factors, different powers, summation terms), it cannot be considered partially correct.

Score: 0

Problem 10 (Paper: 2506.20572v1)

Problem Statement

Background: Consider a network of n sites arranged as the vertices of a regular square. Let the side length of this square be ℓ. The network forms the perimeter of this square, meaning the effectively conducting edges E' connect adjacent vertices along the perimeter. The average length ν between all pairs of sites in such a network is defined by the expression

$$\nu = \frac{1}{\binom{n}{2}} \sum_{(i,j)\in E'}^{\binom{n}{2}} L_{s_i s_j} = \frac{2}{n(n-1)} \sum_{(i,j)\in E'}^{\binom{n}{2}} L_{s_i s_j},$$

where $L_{s_i s_j}$ is the shortest path length between site i and site j along the conducting edges E'.

Task: Derive the average length ν for this square perimeter network, expressed in terms of ℓ.

Ground Truth Solution

$$\boxed{\nu = \frac{4}{3}\ell}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\nu = \dfrac{4}{3}\,\ell$

Judge's Evaluation: The model-generated expression $\nu = \frac{4}{3}\,\ell$ is exactly the same as the ground truth expression $\nu = \frac{4}{3}\ell$. The difference in spacing or formatting does not affect the mathematical equivalence. Both expressions represent the same quantity with the same variables and coefficients.

Score: 1

Problem 11 (Paper: 2506.21061v1)

Problem Statement

Background: Consider a quantum many-body system of $N = 16$ qubits arranged in a 4×4 square lattice. The system's dynamics are governed by a 2D spin-$\frac{1}{2}$ XY model Hamiltonian, given by $\hat{H} = \sum_{<i,j>} J_{ij}(\hat{\sigma}_i^+ \hat{\sigma}_j^- + \text{h.c.})$, where J_{ij} are nearest-neighbor coupling strengths, $\hat{\sigma}^\pm = (\hat{\sigma}^x \pm i\hat{\sigma}^y)/2$ are raising/lowering operators, and $\hat{\sigma}^x, \hat{\sigma}^y, \hat{\sigma}^z$ are Pauli matrices. This Hamiltonian conserves the total spin, meaning the operator $\hat{S}_z = \sum_j \hat{\sigma}_j^z$ commutes with $\hat{H}$. The system is initialized in a half-filling product state $\Psi_0 = 0101...01$, where the qubits are ordered such that the first qubit is in state 0, the second in 1, and so on. After evolving for a time t, the system is in the state $\Psi(t) = e^{-i\hat{H}t/\hbar}\Psi_0$. The system is partitioned into a local subsystem A consisting of two qubits, Q_5 and Q_6, and its complement B, which includes all other qubits. A projected ensemble is formed by performing projective measurements on subsystem B in the computational basis z_{B}, where z_{B} is a bit-string outcome. The quantum state of subsystem A, conditioned on the measurement outcome z_{B}, is denoted as $\Psi_{\mathrm{A}}(t, z_{\mathrm{B}})$. The half-filling sector of the projected ensemble is defined by outcomes z_{B} where the total number of excitations (number of '1's) in subsystem B, denoted $n(z_{\mathrm{B}})$, is 7. Recall that $\hat{\sigma}^z 0 = -0$ and $\hat{\sigma}^z 1 = 1$.

Task: Determine the general form of the projected state $\Psi_{\mathrm{A}}(t, z_{\mathrm{B}})$ for any z_{B} belonging to the half-filling sector of the projected ensemble.

Ground Truth Solution

$$\boxed{\Psi_{\mathrm{A}}(t, z_{\mathrm{B}}) = \alpha(t, z_{\mathrm{B}})01 + \beta(t, z_{\mathrm{B}})10}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $|\Psi_A(t, z_B)\rangle = \dfrac{\alpha(t, z_B)\,|01\rangle + \beta(t, z_B)\,|10\rangle}{\sqrt{|\alpha(t, z_B)|^2 + |\beta(t, z_B)|^2}}$

Judge's Evaluation: The ground truth expression is a linear combination of two basis states with coefficients $\alpha(t, z_B)$ and $\beta(t, z_B)$, without any normalization factor explicitly included.

The model-generated expression includes the same linear combination but divides by the normalization factor $\sqrt{|\alpha(t, z_B)|^2 + |\beta(t, z_B)|^2}$, effectively normalizing the state vector.

Since quantum states are typically normalized, the model's expression is mathematically equivalent to the ground truth expression up to a normalization factor. The ground truth expression may be implicitly normalized or just written in an unnormalized form.

Given that the model's expression is a normalized version of the ground truth, it is mathematically equivalent and correct, just more explicit.

Therefore, the model-generated expression matches the ground truth expression in form and meaning, with the only difference being the explicit normalization.

Score: 1

Problem 12 (Paper: 2506.21061v1)

Problem Statement

Background: Consider a quantum many-body system partitioned into a local subsystem A and its complement B. For a quantum state Ψ, the wave function can be expressed as $\Psi = \sum_{z_{\mathrm{B}}} \sqrt{p(z_{\mathrm{B}})}\Psi_{\mathrm{A}}(z_{\mathrm{B}}) \otimes z_{\mathrm{B}}$, where z_{B} is a measurement basis for B, and $p(z_{\mathrm{B}})$ is the probability of measuring the bit-string z_{B}. The state of subsystem A, conditioned on the measurement outcome z_{B} for subsystem B, is described by the density matrix $\hat{\rho}_{\mathrm{A}}(z_{\mathrm{B}})$. The ensemble-averaged entropy $\bar{E}_{\mathrm{A}}$ is defined as $\bar{E}_{\mathrm{A}} = -\sum_{z_{\mathrm{B}}} p(z_{\mathrm{B}}) \ln \mathrm{Tr}\hat{\rho}^2_{\mathrm{A}}(z_{\mathrm{B}})$. For a pure state $\hat{\rho}_{\mathrm{A}}(z_{\mathrm{B}})$, its purity $\mathrm{Tr}\hat{\rho}^2_{\mathrm{A}}(z_{\mathrm{B}}) = 1$, which implies $\bar{E}_{\mathrm{A}} = 0$ if all $\hat{\rho}_{\mathrm{A}}(z_{\mathrm{B}})$ are pure. When the system is coupled to the environment, the states $\hat{\rho}_{\mathrm{A}}(z_{\mathrm{B}})$ become mixed, and $\mathrm{Tr}\hat{\rho}^2_{\mathrm{A}}(z_{\mathrm{B}}) < 1$, leading to $\bar{E}_{\mathrm{A}} > 0$. Assume that for small information leakage, the state of the entire system at time t can be approximated as $\hat{\rho}(t) \approx \Psi(t)\Psi(t) + \delta\hat{\rho}(t)$, where $\delta\hat{\rho}(t)$ is a small perturbation due to noise. For small evolution times t, the entropy of the entire system, $S(t) = -\mathrm{Tr}(\hat{\rho}(t) \ln \hat{\rho}(t))$, can be approximated by $S(t) \approx t/\tau_S$, where τ_S is a characteristic time scale for entropy generation. Furthermore, assume that for weak and uniform noise across the system, the purity of the projected states, $\mathrm{Tr}\hat{\rho}^2_{\mathrm{A}}(z_{\mathrm{B}})$, is approximately independent of z_{B} and can be denoted by a time-dependent function $P(t)$. For small leakage, $P(t)$ can be expressed as $P(t) \approx 1 - \epsilon(t)$, where $\epsilon(t)$ is a small, positive quantity. The purity of the entire system, $\mathrm{Tr}(\hat{\rho}^2(t))$, is approximately $1 - S(t)$ for small leakage. Finally, assume that $\epsilon(t)$ is proportional to $S(t)$, and that $\bar{E}_{\mathrm{A}}$ can be expressed in the form $E_0 \frac{t}{\tau_{\mathrm{MB}}}$, where E_0 is a steady-state value of $\bar{E}_{\mathrm{A}}$ and τ_{MB} is a characteristic decoherence time for the many-body system.

Task: Derive the approximate expression for the ensemble-averaged entropy $\bar{E}_{\mathrm{A}}$ for small evolution times t, in terms of E_0, t, and τ_{MB}.

Ground Truth Solution

$$\boxed{\bar{E}_{\mathrm{A}} \approx E_0 \frac{t}{\tau_{\mathrm{MB}}}}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\quad {}_{\mathrm{A}}(t) \;\simeq\; E_0 \, \frac{t}{\tau_{\mathrm{MB}}}$

Judge's Evaluation: The model-generated expression is:

$$\bar{E}_{\mathrm{A}}(t) \simeq E_0 \frac{t}{\tau_{\mathrm{MB}}}$$

The ground truth expression is:

$$\bar{E}_{\mathrm{A}} \approx E_0 \frac{t}{\tau_{\mathrm{MB}}}$$

Comparison:

- Both expressions have the same variables and structure. - The model includes an explicit time dependence $\bar{E}_{\mathrm{A}}(t)$, which is consistent with the physical context and does not change the mathematical equivalence. - The approximation symbols $\simeq$ and $\approx$ are equivalent in meaning here. - The fraction and all variables match exactly. - The model expression includes some extra spacing and parentheses around t, which is stylistic and does not affect correctness.

Therefore, the model-generated expression is mathematically equivalent to the ground truth.

Score: 1

Problem 13 (Paper: 2506.21156v1)

Problem Statement

Background: Consider a quantum system of L qubits arranged in a chain with periodic boundary conditions, where $L = 2\ell$ for $\ell \in \mathbb{Z}^+$. The system's dynamics are governed by the PXP Hamiltonian, which in the qubit basis is given by $H_{\mathrm{PXP}} = \sum_{i=1}^{L} P_{(i-1) \bmod L} X_i P_{(i+1) \bmod L}$, where $P_i = 00$ is the projector to the 0 state, and $X_i = 01+10$ is the Pauli X operator acting on the i^{th} site. Due to the Rydberg blockade mechanism, adjacent atoms cannot simultaneously occupy 1 states, meaning states with consecutive $|1\rangle$s are forbidden.

To simplify the analysis, we can map two consecutive qubits to a single qutrit using the correspondence $|\iota_1\iota_2\rangle \equiv |\underline{2\iota_1 + \iota_2}\rangle$, where $\iota_1, \iota_2 \in \{0, 1\}$. The qutrit states are $|\underline{0}\rangle \equiv |00\rangle$, $|\underline{1}\rangle \equiv |01\rangle$, $|\underline{2}\rangle \equiv |10\rangle$, and $|\underline{3}\rangle \equiv |11\rangle$. Given the physical constraint, the $|\underline{3}\rangle$ state is forbidden. In this qutrit representation, the PXP Hamiltonian can be expressed in terms of Gell-Mann matrices, and further decomposed into a linear part, $H_{\mathrm{PXP,lin}}$, and a residual part, $H_{\mathrm{PXP,res}}$. For the purpose of analyzing the "arch" structure of the Lanczos coefficients, we consider only the linear part of the Hamiltonian, which is given by: $H_{\mathrm{PXP,lin}} = \sum_{i=1}^{\ell} \left[(E_\alpha + E_{-\alpha})_i + (E_\beta + E_{-\beta})_i \right]$, where $E_\alpha = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $E_{-\alpha} = E_\alpha^\dagger$, $E_\beta = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$, and $E_{-\beta} = E_\beta^\dagger$. These operators act on the qutrit states as defined by the mapping: $a|\underline{2}\rangle + b|\underline{0}\rangle + c|\underline{1}\rangle \doteq \begin{pmatrix} a \\ b \\ c \end{pmatrix}$.

We are interested in the time evolution of an initial state $|\Psi_0\rangle$ under $H_{\mathrm{PXP,lin}}$. The Lanczos algorithm constructs an orthonormal Krylov basis $|K_n\rangle$ from $|\Psi_0\rangle$ by iteratively applying the Hamiltonian: $|A_{n+1}\rangle = (H - a_n)|K_n\rangle - b_n|K_{n-1}\rangle$, with $|K_n\rangle = b_n^{-1}|A_n\rangle$, $a_n = \langle K_n|H|K_n\rangle$, and $b_n^2 = \langle A_n|A_n\rangle$. For the PXP Hamiltonian acting on product states, all diagonal Lanczos coefficients a_n are zero. The off-diagonal Lanczos coefficients b_m can be determined from the inner products of iterated Hamiltonian actions: $b_m = \left[\frac{\langle K_0|(H^-_{\mathrm{PXP,lin}})^m (H^+_{\mathrm{PXP,lin}})^m|K_0\rangle}{\langle K_0|(H^-_{\mathrm{PXP,lin}})^{m-1} (H^+_{\mathrm{PXP,lin}})^{m-1}|K_0\rangle} \right]^{1/2}$, where $H_{\mathrm{PXP,lin}} = H^+_{\mathrm{PXP,lin}} + H^-_{\mathrm{PXP,lin}}$.

Consider the initial state $|Z_k\rangle = |\ldots 1 \underbrace{0 \ldots 0}_{k-1} 1 \ldots\rangle$, which in the qutrit representation can be approximated as a product of two simpler product states: $|Z'_k\rangle \equiv |\underline{2}\rangle^{\otimes n_2} \otimes |\underline{0}\rangle^{\otimes n_0}$, where $n_2 = \ell \cdot \frac{2}{k}$ and $n_0 = \ell \cdot \frac{k-2}{k}$. For a state $|\underline{2}\rangle^{\otimes n}$, the relevant ladder operators are $H^{(2,n,+)}_{\mathrm{PXP,lin}} = \sum_{i=1}^{n} (E_{-\alpha} + E_{-\beta})_i = \frac{1}{\sqrt{2}} \sum_{i=1}^{n} (J_-)_i$, where $(J_-)_i$ are spin-1 lowering operators. The Lanczos coefficients for this case are $b_m^{(2,n)} = \sqrt{\frac{1}{2} m(2n - m + 1)}$. For a state $|\underline{0}\rangle^{\otimes n}$, the relevant ladder operators are $H^{(0,n,+)}_{\mathrm{PXP,lin}} = \sum_{i=1}^{n} (E_\alpha + E_{-\beta})_i$. The Lanczos coefficients for this case are $b_m^{(0,n)} = \sqrt{2m(n - m + 1)}$.

Task: Derive the expression for the Lanczos coefficients b_m for the initial state $|Z'_k\rangle$ under the evolution of the linear PXP Hamiltonian $H_{\mathrm{PXP,lin}}$.

Ground Truth Solution

$$\boxed{b_m = \sqrt{2}m \left[\frac{\sum_{l=0}^{m} \frac{2^{-2l}}{l!\left(\ell \cdot \frac{4}{k} - l\right)!(m-l)!\left(\ell \cdot \frac{k-2}{k} - m + l\right)!}}{\sum_{l=0}^{m-1} \frac{2^{-2l}}{l!\left(\ell \cdot \frac{4}{k} - l\right)!(m-l-1)!\left(\ell \cdot \frac{k-2}{k} - m + l + 1\right)!}} \right]^{1/2}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: $\mathrm{b}_m = \sqrt{\frac{2(k-1)}{k+2}\, m\left(\frac{(k+2)\,\ell}{k} - m + 1\right)}, \quad m = 1, 2, \ldots, \frac{(k+2)\,\ell}{k}$

Judge's Evaluation: The ground truth expression for b_m is a complicated ratio of sums involving factorials and powers of 2, with indices running from 0 to m and $m - 1$, respectively. It is a nested sum expression with factorial terms in the denominators and depends on parameters ℓ, k, and m.

The model-generated expression for b_m is a much simpler closed-form expression involving a square root of a product of terms, including m, ℓ, and k, but it does not resemble the sum-of-factorials form at all. It also includes a domain specification for m.

Since the model expression is not mathematically equivalent to the ground truth expression, and it does not match the structure or form of the ground truth, it cannot be considered correct or even partially correct. The model expression appears to be a different formula entirely.

Therefore, the model expression is incorrect.

Score: 0

Problem 14 (Paper: 2506.21156v1)

Problem Statement

Background: The PXP Hamiltonian describes a system of L qubits arranged in a chain with periodic boundary conditions, given by $H_{\mathrm{PXP}} = \sum_{i=1}^{L} P_{(i-1) \bmod L} X_i P_{(i+1) \bmod L}$, where $P_i = 00$ is the projector to the 0 state and $X_i = 01 + 10$ is the Pauli X operator acting on the i^{th} site. Due to the Rydberg blockade mechanism, adjacent qubits cannot simultaneously occupy the 1 state. This constraint implies that states like $|11\rangle$ are forbidden.

For an even lattice size $L = 2\ell$, the system can be effectively described using ℓ qutrits, where each qutrit corresponds to two consecutive qubits. The qutrit states are defined as $|\underline{0}\rangle \equiv |00\rangle$, $|\underline{1}\rangle \equiv |01\rangle$, and $|\underline{2}\rangle \equiv |10\rangle$. The forbidden $|11\rangle$ state corresponds to $|\underline{3}\rangle$.

The PXP Hamiltonian can be decomposed into a linear part, $H_{\mathrm{PXP,lin}}$, and a residual part, $H_{\mathrm{PXP,res}}$. The linear part is given by: $H_{\mathrm{PXP,lin}} = \sum_{i=1}^{\ell} \left[(E_\alpha + E_{-\alpha})_i + (E_\beta + E_{-\beta})_i \right]$, where $E_\alpha, E_{-\alpha}, E_\beta, E_{-\beta}$ are generators of the $\mathfrak{sl}_3(\mathbb{C})$ Lie algebra. Their matrix representations in the qutrit basis $\{|\underline{2}\rangle, |\underline{0}\rangle, |\underline{1}\rangle\}$ (corresponding to vector components a, b, c respectively, such that $a|\underline{2}\rangle + b|\underline{0}\rangle + c|\underline{1}\rangle \doteqdot \begin{pmatrix} a \\ b \\ c \end{pmatrix}$) are: $E_\alpha = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $E_{-\alpha} = E_\alpha^\dagger = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $E_\beta = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$, $E_{-\beta} = E_\beta^\dagger = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ The sum $(E_\alpha + E_{-\alpha})_i + (E_\beta + E_{-\beta})_i$ can be identified with $\sqrt{2}(J_x)_i$, where $(J_x)_i$ is the spin-1 matrix for the i^{th} qutrit.

The Lanczos algorithm is used to construct an orthonormal Krylov basis $\{|K_m\rangle\}$ from an initial state $|K_0\rangle = |\Psi_0\rangle$ by repeatedly applying the Hamiltonian H. The Hamiltonian takes a tri-diagonal form in this basis: $H|K_m\rangle = b_m|K_{m-1}\rangle + a_m|K_m\rangle + b_{m+1}|K_{m+1}\rangle$. For product initial states under the PXP Hamiltonian, all diagonal coefficients a_m are zero. The off-diagonal Lanczos coefficients b_m are given by $b_m = \left[\frac{\langle K_0|(H^-)^m (H^+)^m|K_0\rangle}{\langle K_0|(H^-)^{m-1}(H^+)^{m-1}|K_0\rangle} \right]^{1/2}$, where $H = H^+ + H^-$ is a decomposition into raising and lowering operators.

Consider initial states of the form $|Z_k\rangle = |\ldots 1 \underbrace{0 \ldots 0}_{k-1} 1 \ldots\rangle$, where $k - 1 \in \mathbb{Z}^+$. For analytical convenience, we consider a simplified initial state $|Z_k'\rangle \equiv |\underline{2}\rangle^{\otimes N_2} \otimes |\underline{0}\rangle^{\otimes N_0}$, where $N_2 = \ell \cdot \frac{2}{k}$ and $N_0 = \ell \cdot \frac{k-2}{k}$. The linear PXP Hamiltonian $H_{\mathrm{PXP,lin}}$ acts independently on each qutrit. The Lanczos coefficients generated by $H_{\mathrm{PXP,lin}}$ acting on a state of N_2 qutrits all in $|\underline{2}\rangle$ are $b_m^{(2,N_2)} = \sqrt{\frac{1}{2} m (2N_2 - m + 1)}$. The Lanczos coefficients generated by $H_{\mathrm{PXP,lin}}$ acting on a state of N_0 qutrits all in $|\underline{0}\rangle$ are $b_m^{(0,N_0)} = \sqrt{2m(N_0 - m + 1)}$.

Task: Derive the expression for the Lanczos coefficients b_m generated by the linear PXP Hamiltonian $H_{\mathrm{PXP,lin}}$ acting on the initial state $|Z_k'\rangle$.

Ground Truth Solution

$$\boxed{b_m = \sqrt{2} m \left[\frac{\sum_{l=0}^{m} \frac{2^{-2l}}{l! \left(\ell \cdot \frac{4}{k} - l\right)! (m-l)! \left(\ell \cdot \frac{k-2}{k} - m + l\right)!}}{\sum_{l=0}^{m-1} \frac{2^{-2l}}{l! \left(\ell \cdot \frac{4}{k} - l\right)! (m-l-1)! \left(\ell \cdot \frac{k-2}{k} - m + l + 1\right)!}} \right]^{1/2}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: $\mathrm{b}_m = \sqrt{\frac{1}{2} m \left(2N_2 - m + 1\right) + 2m \left(N_0 - m + 1\right)}$

Judge's Evaluation: The ground truth expression for b_m is a complicated ratio of sums involving factorials and powers of 2, all under a square root, multiplied by $\sqrt{2}m$. It explicitly depends on sums over l with factorial terms and parameters involving ℓ, k, and m.

The model-generated expression for b_m is a much simpler algebraic expression inside a square root, involving m, N_0, and N_2, with no sums, factorials, or the parameters ℓ and k. The structure and form are completely different.

There is no evident equivalence or even a close resemblance between the two expressions. The model expression does not match the ground truth expression in form or content, and it is not a minor variation or simplification of the ground truth. It appears to be a completely different formula.

Therefore, the model-generated expression is incorrect relative to the ground truth.

Score: 0

Problem 15 (Paper: 2506.21168v1)

Problem Statement

Background: Consider a quantum system evolving in a finite-dimensional Hilbert space $\mathcal{H}$. Let P be a projection operator onto a subspace $V \subset \mathcal{H}$, and $Q = 1 - P$ be the complementary projector onto $V^{\perp}$. The system undergoes unitary evolution described by $U(t) = e^{-iHt}$, where H is a time-independent Hamiltonian. We consider an initial quantum state ψ such that $\psi \in V$, which implies $P\psi = \psi$.

To monitor the quantum system, weak measurements are performed. These measurements are characterized by a coupling strength parameter $\eta \in (0, 1]$. The effect of these weak measurements on the system can be described by modified operators P_η and Q_η, which are related to the standard projection operators P and Q as follows: $P_\eta = \sqrt{\eta}P$ $Q_\eta = Q + \sqrt{1-\eta}P$

These operators satisfy a pseudo-completeness relation: $P_\eta + Q_\eta = 1 + f(\eta)P$, where $f(\eta) = \sqrt{\eta} + \sqrt{1-\eta} - 1$. Additionally, define $g(\eta) = 1 - \sqrt{1-\eta}$. Note that $P_\eta = \sqrt{\eta}P$ and $1 - Q_\eta = g(\eta)P$.

The monitoring process involves repeated applications of the unitary evolution followed by a weak measurement. The "weak survival operator" is defined as $\tilde{U}_\eta(t) = Q_\eta U(t)$. The "weak first return amplitude operator" $\hat{a}_n^\eta(t)$ describes the amplitude for the quantum walker to be detected for the first time at the n-th measurement step. It is defined as: $\hat{a}_n^\eta(t) = P_\eta U(t)\tilde{U}_\eta(t)^{n-1}P$ for $n \geq 1$.

For an initial state $\psi \in V$, the action of the weak first return amplitude operator can be expressed as: $\hat{a}_n^\eta(t)\psi = U(t)\tilde{U}_\eta(t)^{n-1}\psi - \tilde{U}_\eta(t)^n\psi + f(\eta)PU(t)\tilde{U}_\eta(t)^{n-1}\psi$.

The η-dependent first detection probability $p_n(\eta, t)$ is the probability of detecting the quantum walker at the n-th measurement step, given by $p_n(\eta, t) = \hat{a}_n^\eta(t)\psi^2$.

Task: Derive an analytical expression for the η-dependent first detection probability $p_n(\eta, t)$ in terms of the norms of the weak survival operator $\tilde{U}_\eta(t)$ acting on the initial state ψ.

Ground Truth Solution

$$\boxed{p_n(\eta, t) = \tilde{U}_\eta(t)^{n-1}\psi^2 - \tilde{U}_\eta(t)^n\psi^2}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\mathrm{p}_n(\eta, t) = \left\|\tilde{U}_\eta(t)^{\,n-1}\psi\right\|^2 - \left\|\tilde{U}_\eta(t)^{\,n}\psi\right\|^2$

Judge's Evaluation: The model-generated expression is:

$\mathrm{p}_n(\eta, t) = \|\tilde{U}_\eta(t)^{n-1}\psi\|^2 - \|\tilde{U}_\eta(t)^n\psi\|^2$

The ground truth expression is:

$\mathrm{p}_n(\eta, t) = \tilde{U}_\eta(t)^{n-1}\psi^2 - \tilde{U}_\eta(t)^n\psi^2$

Both expressions are mathematically identical. The only differences are stylistic: the model uses $\left\|...\right\|^2 insteadof ...^2, andthespacingisslightlydifferent.Thesearеequivalentnotationsforthenormsquared.Thevariablesa$

Therefore, the model-generated expression is completely correct.

Score: 1

Problem 16 (Paper: 2506.21168v1)

Problem Statement

Background: Consider a continuous-time quantum walk on a finite graph, where the quantum system evolves unitarily under a time-independent Hamiltonian H, such that the unitary evolution operator is $U(t) = e^{-iHt}$. The system is monitored stroboscopically at a constant rate, with a time interval t between consecutive measurements. The measurement process is implemented via weak measurements, characterized by a coupling parameter $\eta \in (0, 1]$.

Let $\mathcal{H}$ be the Hilbert space of the quantum system. We are interested in the first hitting time to a specific subspace $V \subset \mathcal{H}$. Let P be the projection operator onto V, and $Q = 1 - P$ be the complementary projector onto $V^{\perp}$. For weak measurements, the standard projection operators P and Q are effectively replaced by modified operators P_η and Q_η, which are defined as $P_\eta = \sqrt{\eta}P$ and $Q_\eta = Q + \sqrt{1-\eta}P$. The "survival operator" for the weakly monitored quantum walk is defined as $\tilde{U}_\eta(t) := Q_\eta U(t)$.

For an initial quantum state $\psi \in V$ (i.e., $P\psi = \psi$ and $\psi = 1$), the expected first hitting time after N weak measurements, denoted $\tau_N(\eta, t)$, is given by:

$$\tau_N(\eta, t) = \frac{\sum_{k=0}^{N-1} \tilde{U}_\eta(t)^k \psi^2 - N\tilde{U}_\eta(t)^N \psi^2}{1 - \tilde{U}_\eta(t)^N \psi^2}$$

Consider the "Zeno limit," where the time interval between measurements approaches zero, i.e., $t = 0$. In this limit, the unitary evolution operator becomes $U(0) = 1$.

Task: Derive the analytical expression for the expected first hitting time $\tau_N(\eta, 0)$ in the Zeno limit, for an initial state $\psi \in V$.

Ground Truth Solution

$$\boxed{\tau_N(\eta, 0) = \frac{1}{\eta} - \frac{N(1-\eta)^N}{1-(1-\eta)^N}}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\tau_N(\eta, 0) = \frac{1}{\eta} - \frac{N(1-\eta)^N}{1-(1-\eta)^N}$

Judge's Evaluation: The model-generated expression is:

$$\tau_N(\eta, 0) = \frac{1}{\eta} - \frac{N(1-\eta)^N}{1-(1-\eta)^N}$$

The ground truth expression is:

$$\tau_N(\eta, 0) = \frac{1}{\eta} - \frac{N(1-\eta)^N}{1-(1-\eta)^N}$$

Comparing the two, they are exactly the same mathematically. The only difference is in spacing, which is irrelevant. The variables, powers, coefficients, and signs all match perfectly.

Score: 1

Problem 17 (Paper: 2506.21173v1)

Problem Statement

Background: A quantum system is described by a Hilbert space $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$. The standard basis for this space is $\mathcal{B} = (00, 01, 10, 11)$. A time-evolving quantum state (trajectory) $\Psi(t)$ is given in this basis by:

$$\Psi(t) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ \cos(t) \\ \sin(t) \end{pmatrix}$$

for $t \in [0, \frac{\pi}{2}]$. A new tensor product structure (TPS) is defined by a basis $\mathcal{B}' = (++, +-, -+, --)$, where $+$ and $-$ are orthonormal basis states for each of the two constituent qubits. This new TPS is related to the standard one by a unitary transformation U, such that if a state is represented by a column vector v in basis $\mathcal{B}$, its representation in basis $\mathcal{B}'$ is Uv. The unitary matrix U is given by:

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & i & 0 & i \\ 0 & -i & 0 & i \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

A state is considered a product state in the new TPS if it can be written as $\Psi_1(t) \otimes \Psi_2(t)$, where $\Psi_1(t)$ is a state in the first qubit's Hilbert space and $\Psi_2(t)$ is a state in the second qubit's Hilbert space. In the basis $\mathcal{B}'$, a product state $(a + + b-) \otimes (c + + d-)$ corresponds to the column vector $\begin{pmatrix} ac \\ ad \\ bc \\ bd \end{pmatrix}$.

Task: Determine the explicit form of the transformed state $\Psi'(t)$ in the new basis $\mathcal{B}'$ and express it as a tensor product of two single-qubit states, $\Psi_1(t)$ and $\Psi_2(t)$, where $\Psi_1(t)$ and $\Psi_2(t)$ are linear combinations of $+$ and $-$.

Ground Truth Solution

$$\boxed{\frac{e^{-it}}{4} \left[(e^{it} - 1) + + (e^{it} + 1) -\right] \otimes \left[(e^{it} - 1) + + (e^{it} + 1) -\right]}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\Psi'(t) = \left(\sin \frac{t}{2} + - i \cos \frac{t}{2} -\right) \otimes \left(\sin \frac{t}{2} + - i \cos \frac{t}{2} -\right)$

Judge's Evaluation: Let's analyze the two expressions carefully.

Ground Truth Expression:

$$\frac{e^{-it}}{4} \left[(e^{it} - 1) + + (e^{it} + 1) -\right] \otimes \left[(e^{it} - 1) + + (e^{it} + 1) -\right]$$

Model-Generated Expression:

$$\Psi'(t) = \left(\sin \frac{t}{2} + - i \cos \frac{t}{2} -\right) \otimes \left(\sin \frac{t}{2} + - i \cos \frac{t}{2} -\right)$$

—

Step 1: Check if the two expressions are mathematically equivalent
Rewrite the ground truth expression inside the bracket:

$$(e^{it} - 1) + + (e^{it} + 1) -$$

Recall Euler's formula:

$$e^{it} = \cos t + i \sin t$$

Calculate each coefficient:
- $e^{it} - 1 = (\cos t - 1) + i \sin t$ - $e^{it} + 1 = (\cos t + 1) + i \sin t$
Now, factor out $e^{-it}/4$:

$$\frac{e^{-it}}{4}\left[(e^{it}-1)+ + (e^{it}+1)-\right] = \frac{1}{4}\left[(1-e^{-it})+ + (1+e^{-it})-\right]$$

Because $e^{-it}e^{it} = 1$, multiplying inside the bracket by e^{-it} shifts the exponentials.
Alternatively, let's try to rewrite the model expression in exponential form to see if it matches.
—
Step 2: Express the model expression in exponential form
Recall:

$$\sin\frac{t}{2} = \frac{e^{it/2} - e^{-it/2}}{2i}, \quad \cos\frac{t}{2} = \frac{e^{it/2} + e^{-it/2}}{2}$$

So,

$$\sin\frac{t}{2} + - i\cos\frac{t}{2} - = \left(\frac{e^{it/2} - e^{-it/2}}{2i}\right) + - i\left(\frac{e^{it/2} + e^{-it/2}}{2}\right) -$$

Multiply terms:

$$= \frac{e^{it/2} - e^{-it/2}}{2i} + - \frac{ie^{it/2} + ie^{-it/2}}{2} -$$

Note that $\frac{1}{i} = -i$, so:

$$\frac{e^{it/2} - e^{-it/2}}{2i} = -\frac{i}{2}(e^{it/2} - e^{-it/2}) = -\frac{ie^{it/2}}{2} + \frac{ie^{-it/2}}{2}$$

Therefore,

$$\sin\frac{t}{2} + - i\cos\frac{t}{2} - = \left(-\frac{ie^{it/2}}{2} + \frac{ie^{-it/2}}{2}\right) + - \frac{ie^{it/2} + ie^{-it/2}}{2} -$$

Group terms:

$$= \frac{i}{2}\left(e^{-it/2} - e^{it/2}\right) + - \frac{i}{2}\left(e^{it/2} + e^{-it/2}\right) -$$

Rewrite:

$$= \frac{i}{2}\left(e^{-it/2} - e^{it/2}\right) + - \frac{i}{2}\left(e^{it/2} + e^{-it/2}\right) -$$

Multiply numerator and denominator by $e^{it/2}$ to get terms in e^{it}:

$$= \frac{i}{2}e^{-it/2}(1 - e^{it}) + - \frac{i}{2}e^{-it/2}(e^{it} + 1) -$$

Factor out $\frac{i}{2}e^{-it/2}$:

$$= \frac{i}{2}e^{-it/2}\left[(1 - e^{it}) + - (e^{it} + 1) -\right]$$

—
Step 3: Compare with the ground truth
Ground truth inside the bracket is:

$$(e^{it} - 1) + + (e^{it} + 1) -$$

Model expression inside the bracket is:

$$(1 - e^{it}) + - (e^{it} + 1) - = -(e^{it} - 1) + - (e^{it} + 1) -$$

So the model expression is:

$$\frac{i}{2}e^{-it/2} \times \left[-(e^{it}-1)+-(e^{it}+1)-\right] = -\frac{i}{2}e^{-it/2}\left[(e^{it}-1)++(e^{it}+1)-\right]$$

The ground truth has a prefactor of $\frac{e^{-it}}{4}$, the model has $-\frac{i}{2}e^{-it/2}$.

—

Step 4: Check if the prefactors differ by a global phase and constant

Note that global phases and overall constants can be factored out in quantum states without changing the physical state.

Calculate the ratio of the prefactors:

$$\frac{-\frac{i}{2}e^{-it/2}}{\frac{e^{-it}}{4}} = -i \times 2 \times e^{it/2} = -2ie^{it/2}$$

This is a complex number with magnitude 2 and a phase factor.

Since quantum states are defined up to a global phase and normalization, the difference in prefactors is acceptable if the states are normalized.

—

Step 5: Conclusion

- The model expression matches the ground truth expression up to a global phase and normalization factor. - The bracketed terms are the same up to an overall minus sign, which can be absorbed into the global phase. - The model expression is a tensor product of the same single-qubit state, matching the ground truth's tensor product structure. - The difference in prefactors is a global phase and normalization factor, which is physically irrelevant.

Therefore, the model expression is **mathematically equivalent** to the ground truth expression.

—

Final assessment:

- The model expression is completely correct, modulo global phase and normalization. - It matches the ground truth expression exactly in form and content.

—

Score: 1

—

Score: 1

Problem 18 (Paper: 2506.21173v1)

Problem Statement

Background: Consider a complex Hilbert space $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$ with a standard basis $\mathcal{B} = (11, 12, 21, 22)$. Let $\Psi(t) = \begin{pmatrix} a(t) \\ b(t) \\ c(t) \\ d(t) \end{pmatrix}$ be a time-evolving quantum state, where $a(t), b(t), c(t), d(t)$ are continuous complex functions for $t \in [0, T]$. Assume that the set of all products of these functions, $\{a(t)^2, b(t)^2, c(t)^2, d(t)^2, a(t)b(t), a(t)c(t), a(t)$ forms a linearly independent set in the space of continuous functions $\mathcal{C}^0([0, T])$.

Suppose there exists a fixed tensor product structure (TPS) $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ that disentangles $\Psi(t)$. This TPS is defined by a new orthonormal basis $\mathcal{B}' = (1'1', 1'2', 2'1', 2'2')$. Let $U = (u_{ij})_{1 \leq i,j \leq 4}$ be the unitary matrix that maps the standard basis $\mathcal{B}$ to $\mathcal{B}'$, such that $Uj = j'$ for $j \in \{1, 2, 3, 4\}$. The state $\Psi(t)$ expressed in the new basis is $\Psi'(t) = U\Psi(t) = \begin{pmatrix} \Psi'_{11}(t) \\ \Psi'_{12}(t) \\ \Psi'_{21}(t) \\ \Psi'_{22}(t) \end{pmatrix}$. For $\Psi(t)$ to be a product state in this new TPS for all t, its components in $\mathcal{B}'$ must satisfy the condition $\Psi'_{11}(t)\Psi'_{22}(t) = \Psi'_{12}(t)\Psi'_{21}(t)$ for all $t \in [0, T]$.

Let $a_1(t) = a(t)$, $a_2(t) = b(t)$, $a_3(t) = c(t)$, $a_4(t) = d(t)$. The product state condition can be written as:

$$\left(\sum_{k=1}^{4} u_{1k} a_k(t)\right) \left(\sum_{l=1}^{4} u_{4l} a_l(t)\right) = \left(\sum_{k=1}^{4} u_{2k} a_k(t)\right) \left(\sum_{l=1}^{4} u_{3l} a_l(t)\right)$$

By the linear independence of the products $a_k(t)a_l(t)$, the coefficients of each product term must be equal on both sides of this equation. This implies two sets of conditions on the elements of U: 1. For any $j \in \{1, 2, 3, 4\}$: $u_{1j}u_{4j} = u_{2j}u_{3j}$. 2. For any distinct $j, k \in \{1, 2, 3, 4\}$: $u_{1j}u_{4k} + u_{1k}u_{4j} = u_{2j}u_{3k} + u_{2k}u_{3j}$.

Assume that there exists at least one index $j_0 \in \{1, 2, 3, 4\}$ such that $u_{1j_0}u_{4j_0} = u_{2j_0}u_{3j_0} \neq 0$. Without loss of generality, let this index be $j_0 = 1$. This implies that $u_{11}, u_{21}, u_{31}, u_{41}$ are all non-zero. Define new coefficients $v_{ij} = u_{ij}/u_{i1}$ for $i \in \{1, 2, 3, 4\}$ and $j \in \{1, 2, 3, 4\}$.

Task: Using the conditions on the coefficients of U derived from the product state requirement and the linear independence assumption, and the definition of v_{ij}, derive a simplified algebraic relationship that must hold for $v_{1j}, v_{2j}, v_{3j}, v_{4j}$ for any $j \in \{1, 2, 3, 4\}$.

Ground Truth Solution

$$\boxed{(v_{1j} - v_{2j})(v_{2j} - v_{4j}) = 0}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: $\mathrm{v}_{1j}\, v_{4j} = v_{2j}\, v_{3j}$

Judge's Evaluation: The ground truth expression is a product of two differences set equal to zero:

$$(v_{1j} - v_{2j})(v_{2j} - v_{4j}) = 0$$

This implies either $v_{1j} = v_{2j}$ or $v_{2j} = v_{4j}$.

The model-generated expression is:

$$v_{1j}v_{4j} = v_{2j}v_{3j}$$

This is a product equality involving four variables, including v_{3j}, which does not appear in the ground truth expression. The model's expression is not equivalent to the ground truth expression, nor is it a rearrangement or factorization of it.

Since the model's expression introduces a new variable and is not mathematically equivalent to the ground truth, it is incorrect.

Score: 0

Problem 19 (Paper: 2506.21226v1)

Problem Statement

Background: Consider a network automaton (NA) consisting of N nodes, where each node v_i can be in one of two states, $s(v_i,t) \in \{0,1\}$, at discrete time step t. The state of node v_i at time $t+1$ is determined by a uniform local update rule $\phi(s_i,\rho_i)$, where $s_i = s(v_i,t)$ is the current state of node v_i, and $\rho_i = \frac{1}{k_i}\sum_j A_{ij}s_j$ is the density of living nodes in v_i's neighborhood. Here, k_i is the degree of node v_i, and A_{ij} are elements of the adjacency matrix. The local update rule is defined as:

$$\phi(s_i,\rho_i) = \begin{cases} 1, & \text{if } s_i = 0 \text{ and } \rho_i \in \bigcup B, \\ 1, & \text{if } s_i = 1 \text{ and } \rho_i \in \bigcup S, \\ 0, & \text{else.} \end{cases}$$

Here, B and S are sets of subintervals of the unit interval, chosen from $R = \{R_j\}_{j=0}^{r-1}$, where r is an odd integer resolution parameter. The intervals R_j are defined as $R_j = \left[\frac{j}{r}, \frac{j+1}{r}\right[$ for $0 \le j < (r-1)/2$, $R_{(r-1)/2} = \left[\frac{(r-1)}{2r}, \frac{(r+1)}{2r}\right]$, and $R_j = \left]\frac{j}{r}, \frac{j+1}{r}\right]$ for $(r-1)/2 < j \le r-1$.

We are interested in how a defect propagates through the network. A defect is defined as the bitwise XOR operation on two nearly identical configurations, $s(v_i,t)$ and $s'(v_i,t)$. The normalized Hamming distance between these two configurations at time t is given by $\delta^t = \frac{1}{N}\sum_{i=1}^{N}(s(v_i,t) \oplus s'(v_i,t))$, where $\oplus$ is the XOR operator. We assume that the local topological structure can be ignored, meaning that probabilities of states and toggles follow simple binomial or hypergeometric distributions.

Specifically, for a given global state average $\rho^t = \frac{1}{N}\sum_{i=1}^{N} s(v_i,t)$ and normalized Hamming distance δ^t at time t, the probabilities are: - The probability of finding a central node in state s given ρ^t: $P(s|\rho^t) = (\rho^t)^s(1-\rho^t)^{1-s}$. - The probability of finding q living neighbors out of k for a node, given ρ^t: $P(q|k,\rho^t) = \binom{k}{q}(\rho^t)^q(1-\rho^t)^{k-q}$. - The probability that the central node v_i is toggled ($c=1$) or not ($c=0$), given δ^t: $P(c|\delta^t) = (\delta^t)^c(1-\delta^t)^{1-c}$. - The probability that d neighborhood states are toggled out of k neighbors, given δ^t: $P(d|k,\delta^t) = \binom{k}{d}(\delta^t)^d(1-\delta^t)^{k-d}$. - The probability that τ of the d toggles are 'killer toggles' (bit flips from 1 to 0), given k, d, and q (number of living neighbors in the original configuration): $P(\tau|k,d,q) = \binom{q}{\tau}\binom{k-q}{d-\tau}\Big/\binom{k}{d}$, where $\max(0,d+q-k) \le \tau \le \min(d,q)$.

The defect of node v_i at time $t+1$, denoted d_i^{t+1}, is given by $s(v_i,t+1) \oplus s'(v_i,t+1)$. This can be expressed using the local update rule as: $\mathcal{D}(k,s,q,c,d,\tau) = \phi\left(s, \frac{q}{k}\right) \oplus \phi\left(s \oplus c, \frac{q-2\tau+d}{k}\right)$. Here, s is the state of the central node in the original configuration, q is the number of living neighbors in the original configuration, c indicates if the central node itself is toggled, d is the number of toggled neighbors, and τ is the number of killer toggles among the d toggled neighbors. The term $q-2\tau+d$ represents the number of living neighbors in the perturbed configuration. The degree distribution of the network is $P(k)$.

Task: Derive the expected normalized Hamming distance at time $t+1$, denoted $\langle\delta^{t+1}\rangle$, as an analytical expression in terms of the given probabilities, the local update rule ϕ, and the defect function $\mathcal{D}$.

Ground Truth Solution

$$\boxed{\sum_k P(k) \sum_{s\in\{0,1\}} P(s|\rho^t) \sum_{q=0}^{k} P(q|k,\rho^t) \sum_{c\in\{0,1\}} P(c|\delta^t) \sum_{d=0}^{k} P(d|k,\delta^t) \sum_{\tau} P(\tau|k,d,q)\, \mathcal{D}(k,s,q,c,d,\tau)}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\left\langle \delta^{\,t+1} \right\rangle = \sum_{k=0}^{\infty} P(k) \sum_{s=0}^{1} (\rho^t)^s\,(1-\rho^t)^{1-s} \sum_{q=0}^{k} \binom{k}{q}(\rho^t)^q(1-\rho^t)^{k-q} \sum_{c=0}^{1}(\delta^t)^c(1-\delta^t)^{1-c} \sum_{d=0}^{k}\binom{k}{d}(\delta^t)^d(1-\delta^t)^{k-d} \sum_{\tau=\max(0,d+q-k)}^{\min(d,q)} \frac{\binom{q}{\tau}\binom{k-q}{d-\tau}}{\binom{k}{d}}\ \mathcal{D}(k,s,q,c,d,\tau)$

Judge's Evaluation: The ground truth expression is a general summation over probabilities $P(k)$, $P(s|\rho^t)$, $P(q|k,\rho^t)$, $P(c|\delta^t)$, $P(d|k,\delta^t)$, and $P(\tau|k,d,q)$, multiplied by $\mathcal{D}(k,s,q,c,d,\tau)$.

The model-generated expression explicitly writes out these probabilities assuming specific binomial or Bernoulli forms:

- $P(s|\rho^t) = (\rho^t)^s(1-\rho^t)^{1-s}$ for $s \in \{0, 1\}$, which is a Bernoulli distribution. - $P(q|k, \rho^t) = \binom{k}{q}(\rho^t)^q(1-\rho^t)^{k-q}$, a Binomial distribution. - $P(c|\delta^t) = (\delta^t)^c(1-\delta^t)^{1-c}$, Bernoulli. - $P(d|k, \delta^t) = \binom{k}{d}(\delta^t)^d(1-\delta^t)^{k-d}$, Binomial. - $P(\tau|k, d, q) = \frac{\binom{q}{\tau}\binom{k-q}{d-\tau}}{\binom{k}{d}}$, which is the hypergeometric distribution.

The summation limits also match the ground truth, with the τ sum running from $\max(0, d + q - k)$ to $\min(d, q)$, which is the correct support for the hypergeometric distribution.

The model-generated expression also includes the expectation notation $\langle \delta^{t+1} \rangle$, which is consistent with the sum over all these probabilities weighted by $\mathcal{D}$.

All components correspond exactly to the ground truth expression, just expanded explicitly in terms of binomial and Bernoulli probabilities.

Therefore, the model-generated expression is mathematically equivalent to the ground truth.

Score: 1

Problem 20 (Paper: 2506.21226v1)

Problem Statement

Background: A Life-like Network Automaton (LLNA) is a discrete dynamical system where N nodes, each with a binary state $s_i \in \{0,1\}$, evolve synchronously on a network. The state of a node v_i at time $t+1$, denoted $s(v_i, t+1)$, is determined by a uniform local outer-totalistic update rule $\phi(s_i, \rho_i)$. Here, $s_i = s(v_i, t)$ is the current state of node v_i, and $\rho_i = \frac{1}{k_i}\sum_j A_{ij}s_j$ is the density of living nodes in v_i's neighborhood, where k_i is the degree of node v_i and A_{ij} are elements of the adjacency matrix. The network has a degree distribution $P(k)$.

The global state average at time t is $\rho^t = \frac{1}{N}\sum_{i=1}^N s(v_i, t)$. To analyze the propagation of perturbations, we consider two configurations, $s(v_i, t)$ and $s'(v_i, t)$, which are initially nearly identical. The normalized Hamming distance between these configurations at time t is $\delta^t = \frac{1}{N}\sum_{i=1}^N \Big(s(v_i, t) \oplus s'(v_i, t)\Big)$, where $\oplus$ denotes the bitwise XOR operation. The Derrida plot describes how this distance evolves, mapping δ^t to the expected normalized Hamming distance at time $t+1$, $\langle \delta^{t+1} \rangle$.

To derive this expectation, we assume that local topological structure can be ignored, allowing us to use simple probabilistic distributions. The probability of a central node being in state s given the global state average ρ^t is $P(s|\rho^t) = (\rho^t)^s(1-\rho^t)^{1-s}$. The probability of q living neighbors out of k total neighbors, given ρ^t, is $P(q|k, \rho^t) = \binom{k}{q}(\rho^t)^q(1-\rho^t)^{k-q}$. When comparing two configurations, the probability that the central node v_i is toggled ($c = 1$) or not ($c = 0$) between the two configurations, given δ^t, is $P(c|\delta^t) = (\delta^t)^c(1-\delta^t)^{1-c}$. The probability that d out of k neighbors are toggled between the two configurations, given δ^t, is $P(d|k, \delta^t) = \binom{k}{d}(\delta^t)^d(1-\delta^t)^{k-d}$. Among the d toggled neighbors, let τ be the number of 'killer toggles' (bit flips from 1 to 0). The probability of τ killer toggles given k neighbors, d toggled neighbors, and q living neighbors in the first configuration, is $P(\tau|k, d, q) = \binom{q}{\tau}\binom{k-q}{d-\tau}\Big/\binom{k}{d}$, where $\max(0, d+q-k) \le \tau \le \min(d, q)$.

The defect of a single node v_i at time $t+1$, denoted d_i^{t+1}, is $s(v_i, t+1) \oplus s'(v_i, t+1)$. This can be expressed as $\mathcal{D}(k, s, q, c, d, \tau) = \phi\left(s, \frac{q}{k}\right) \oplus \phi\left(s \oplus c, \frac{q-2\tau+d}{k}\right)$, where s is the state of v_i in the first configuration, q is the number of living neighbors of v_i in the first configuration, $s \oplus c$ is the state of v_i in the second configuration, and $\frac{q-2\tau+d}{k}$ is the density of living neighbors of v_i in the second configuration.

Task: Derive the analytical expression for the expected normalized Hamming distance at time $t+1$, $\langle \delta^{t+1} \rangle$, by averaging over all nodes and considering the probabilities of different states and toggles.

Ground Truth Solution

$$\boxed{\langle \delta^{t+1} \rangle = \sum_k P(k) \sum_{s\in\{0,1\}} P(s|\rho^t) \sum_{q=0}^{k} P(q|k, \rho^t) \sum_{c\in\{0,1\}} P(c|\delta^t) \sum_{d=0}^{k} P(d|k, \delta^t) \sum_{\tau} P(\tau|k, d, q)\, \phi\left(s, \frac{q}{k}\right) \oplus \phi\left(s \oplus c, \frac{q-2\tau+d}{k}\right)}$$

Model Outputs

Model: o3 (Score: 0.5)

Model Solution:

$$\begin{aligned} \langle \delta^{t+1} \rangle(\rho^t, \delta^t) = \sum_{k=0}^{\infty} P(k) &\sum_{s=0}^{1}\sum_{q=0}^{k}\sum_{c=0}^{1}\sum_{d=0}^{k}\sum_{\tau=\tau_{\min}}^{\tau_{\max}} \mathcal{D}(k, s, q, c, d, \tau) \\ &\times \binom{k}{q}\binom{q}{\tau}\binom{k-q}{d-\tau} \left(\rho^t\right)^{s+q}\left(1-\rho^t\right)^{k+1-(s+q)} \\ &\times \left(\delta^t\right)^{c+d}\left(1-\delta^t\right)^{k+1-(c+d)} \end{aligned}$$

Judge's Evaluation: The ground truth expression is a nested sum over variables k, s, q, c, d, τ of the product of probabilities $P(k)$, $P(s|\rho^t)$, $P(q|k, \rho^t)$, $P(c|\delta^t)$, $P(d|k, \delta^t)$, $P(\tau|k, d, q)$ multiplied by a function involving ϕ and XOR operations.

The model-generated expression also sums over the same variables, but it explicitly writes out binomial coefficients and powers of ρ^t and δ^t that presumably correspond to the probabilities $P(s|\rho^t)$, $P(q|k, \rho^t)$, $P(c|\delta^t)$, and $P(d|k, \delta^t)$. It also includes a factor $\mathcal{D}(k, s, q, c, d, \tau)$ which is not defined here but presumably corresponds to the ϕ-based function in the ground truth.

Key points of comparison:

- The ground truth expression explicitly shows the XOR and ϕ functions, which are the core of the expected value calculation. - The model expression replaces the probability terms with explicit binomial distributions and powers, which is consistent with the probabilities in the ground truth if those are binomially distributed. - The model expression introduces $\mathcal{D}(k, s, q, c, d, \tau)$ which is not defined here, but presumably encodes the ϕ and XOR terms. - The model expression sums k from 0 to ∞, while the ground truth sums over k without explicit limits (likely over the support of $P(k)$). - The model expression sums s, c from 0 to 1, consistent with the ground truth. - The model expression sums q, d from 0 to k, consistent with the ground truth. - The model expression sums τ from τ_{min} to τ_{max}, consistent with the ground truth's sum over τ.

However, the model expression does not explicitly show the ϕ functions or the XOR operation, instead encapsulating them in $\mathcal{D}$. Since $\mathcal{D}$ is undefined, we cannot verify if it correctly represents the ϕ and XOR terms.

Because the model expression is structurally different and lacks explicit representation of the key functional terms, and because $\mathcal{D}$ is undefined, it is not possible to confirm full equivalence.

At the same time, the model expression is clearly attempting to represent the same summation with the same variables and probability distributions, just in a more expanded and factored form.

Given this, the model expression is very close but incomplete or at least not fully explicit in the key functional part.

Therefore, I assign partial credit.

Score: 0.5

Problem 21 (Paper: 2506.21243v1)

Problem Statement

Background: Let $J_m(x)$ denote the Bessel function of the first kind of order m. For $m \in \mathbb{N}_0 = \{0, 1, 2, ...\}$, let $j_{m,k}$ denote the k^{th} positive zero of J_m. Consider the domain $U = \{(\alpha, \kappa) \in \mathbb{R}^2 : \kappa^2 > \alpha^2\}$. For $m \in \mathbb{Z}$, define the function $F_m : U \to \mathbb{R}$ as:

$$F_m(\alpha, \kappa) = \frac{\kappa + \alpha}{\kappa - \alpha} J_{m-1}(\sqrt{\kappa^2 - \alpha^2}) + J_{m+1}(\sqrt{\kappa^2 - \alpha^2})$$

The Bessel functions satisfy the recurrence relations:

$$J'_m(x) = \frac{J_{m-1}(x) - J_{m+1}(x)}{2}, \qquad \frac{mJ_m(x)}{x} = \frac{J_{m-1}(x) + J_{m+1}(x)}{2}$$

For $m \in \mathbb{N}$, let j^*_m be defined as the infimum of positive κ values for which $F_m(\alpha, \kappa) = 0$ for some $\alpha \in \mathbb{R}$:

$$j^*_m = \inf\{\kappa > 0 : (\alpha, \kappa) \in F_m^{-1}(0) \text{ for some } \alpha \in \mathbb{R}\}$$

Task: Derive a lower bound for j^*_m in terms of $j_{m-1,1}$ for $m \in \mathbb{N}$.

Ground Truth Solution

$$\boxed{j^*_m > j_{m-1,1}}$$

Model Outputs

Model: o3 (Score: 0.5)

Model Solution: $j^*_m \geq j_{m-1,1}$

Judge's Evaluation: The ground truth expression is $j^*_m > j_{m-1,1}$, which uses a strict inequality "¿" (greater than).

The model-generated expression is $j^*_m \geq j_{m-1,1}$, which uses a non-strict inequality "" (greater than or equal to).

The expressions differ only in the inequality sign. The model's expression allows equality, while the ground truth does not.

This is a minor but significant difference in meaning. Since the model's expression is very close but not exactly the same, it should receive partial credit.

Score: 0.5

Problem 22 (Paper: 2506.21243v1)

Problem Statement

Background: Consider a Riemannian 3-manifold with boundary, $M_{a,b,L}$, defined as the Cartesian product $A_{a,b} \times /L$, where $A_{a,b} = \{(x, y) \in^2 : a^2 \leq x^2 + y^2 \leq b^2\}$ for $0 < a < b$, and $L > 0$. The manifold is endowed with the flat metric $g = dx^2 + dy^2 + ds^2$, where $(x, y) \in A_{a,b}$ and $s \in /L$ are Cartesian coordinates. We also consider two Killing vector fields: $T = \partial_s$ and $P = -y\partial_x + x\partial_y$.

Let $(M_{a,b,L},)$ denote the space of smooth complex-valued vector fields on $M_{a,b,L}$. We are interested in the space of complex-valued curl eigenfields $E_{,\lambda,\ell,m}(M_{a,b,L})$, defined as the set of vector fields $B \in (M_{a,b,L},)$ that satisfy the following conditions for given real numbers $\lambda \neq 0$, $\ell \in (2\pi/L)$, and $m \in$: 1. $B = \lambda B$ 2. $B \cdot n = 0$ on $\partial M_{a,b,L}$, where n is the outward unit normal. 3. $_T B = i\ell B$ 4. $_P B = imB$

For a non-trivial $B \in E_{,\lambda,\ell,m}(M_{a,b,L})$, it must be that $\lambda^2 > \ell^2$. Let $\mu\sqrt{\lambda^2 - \ell^2}$. A vector field $B \in E_{,\lambda,\ell,m}(M_{a,b,L})$ exists if and only if there exists a non-trivial scalar function $h \in C^\infty(M_{a,b,L},)$ such that: 1. $B = \lambda^2 hT + \lambda\nabla h \times T + i\ell\nabla h$. 2. h satisfies the partial differential equations: $\Delta h = -\lambda^2 h$, $_T h = i\ell h$, and $_P h = imh$. 3. h satisfies the following boundary conditions on $\partial M_{a,b,L}$: * $-\frac{\lambda m}{a} h + \ell \frac{\partial h}{\partial n} = 0$ on $\partial_1 M = \{(x, y) \in^2 : x^2 + y^2 = a^2\} \times /L$. * $\frac{\lambda m}{b} h + \ell \frac{\partial h}{\partial n} = 0$ on $\partial_2 M = \{(x, y) \in^2 : x^2 + y^2 = b^2\} \times /L$.

In cylindrical coordinates (r, θ, s), a function h satisfying the partial differential equations can be expressed as $h = (AJ_m(\mu r) + BY_m(\mu r))e^{im\theta}e^{i\ell s}$ for some constants $A, B \in \mathbb{C}$, where J_m and Y_m are Bessel functions of the first and second kind, respectively, of order m. Recall the Bessel function recurrence relations for $Z_m \in \{J_m, Y_m\}$: $Z'_m(x) = \frac{Z_{m-1}(x) - Z_{m+1}(x)}{2}$ $\frac{mZ_m(x)}{x} = \frac{Z_{m-1}(x) + Z_{m+1}(x)}{2}$

Task: Determine the condition on λ, ℓ, m, a, b that must be satisfied for a non-trivial $B \in E_{,\lambda,\ell,m}(M_{a,b,L})$ to exist. Express your answer as a determinant equation involving Bessel functions $J_k(\cdot)$ and $Y_k(\cdot)$ of orders $k \in \{m - 1, m + 1\}$.

Ground Truth Solution

$$\boxed{\begin{vmatrix} (\lambda + \ell)J_{m-1}(\mu a) + (\lambda - \ell)J_{m+1}(\mu a) & (\lambda + \ell)Y_{m-1}(\mu a) + (\lambda - \ell)Y_{m+1}(\mu a) \\ (\lambda + \ell)J_{m-1}(\mu b) + (\lambda - \ell)J_{m+1}(\mu b) & (\lambda + \ell)Y_{m-1}(\mu b) + (\lambda - \ell)Y_{m+1}(\mu b) \end{vmatrix} = 0}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\begin{vmatrix} (\lambda + \ell)J_{m-1}(\mu a) + (\lambda - \ell)J_{m+1}(\mu a) & (\lambda + \ell)Y_{m-1}(\mu a) + (\lambda - \ell)Y_{m+1}(\mu a) \\ (\lambda + \ell)J_{m-1}(\mu b) + (\lambda - \ell)J_{m+1}(\mu b) & (\lambda + \ell)Y_{m-1}(\mu b) + (\lambda - \ell)Y_{m+1}(\mu b) \end{vmatrix} = 0$

Judge's Evaluation: The model-generated expression matches the ground truth exactly. Both determinants have the same entries in the same positions, with identical coefficients and arguments for the Bessel functions $J_{m\pm1}$ and $Y_{m\pm1}$. The only difference is formatting (spacing and line breaks), which does not affect mathematical equivalence.

Score: 1

Problem 23 (Paper: 2506.21290v1)

Problem Statement

Background: The paper investigates the glass transition of molecular fluids using a novel Monte Carlo algorithm. To quantify the efficiency of their algorithm, the authors analyze the temperature dependence of the structural relaxation time, τ_α. For the physical molecular dynamics (MD) simulations, the relaxation time τ_α is found to follow an empirical parabolic law in Arrhenius coordinates:

$$\ln\left(\frac{\tau_\alpha}{\tau_o}\right) = J^2\left(\frac{1}{T} - \frac{1}{T_p}\right)^2$$

where T is the temperature, τ_o is a characteristic onset timescale, J is an energy parameter, and T_p is an alternative onset temperature. The experimental glass transition temperature, T_g, is conventionally defined as the temperature at which the relaxation time reaches a value 10^{12} times its onset timescale, i.e., $\tau_\alpha(T_g)/\tau_o = 10^{12}$.

Task: Derive the analytical expression for the experimental glass transition temperature T_g in terms of J and T_p.

Ground Truth Solution

$$\boxed{T_g = \left(\frac{1}{T_p} + \frac{\sqrt{12\ln(10)}}{J}\right)^{-1}}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\mathrm{T}_g = \dfrac{T_p}{1 + \left(\dfrac{T_p}{J}\right)\sqrt{12\ln 10}}$

Judge's Evaluation: Let's analyze the two expressions:

Ground Truth:

$$T_g = \left(\frac{1}{T_p} + \frac{\sqrt{12\ln(10)}}{J}\right)^{-1}$$

Model-Generated:

$$T_g = \frac{T_p}{1 + \left(\frac{T_p}{J}\right)\sqrt{12\ln 10}}$$

—

Step 1: Check equivalence

Rewrite the ground truth expression:

$$T_g = \frac{1}{\frac{1}{T_p} + \frac{\sqrt{12\ln(10)}}{J}} = \frac{1}{\frac{1}{T_p} + \frac{\sqrt{12\ln(10)}}{J}}$$

Find a common denominator inside the parentheses:

$$\frac{1}{T_p} + \frac{\sqrt{12\ln(10)}}{J} = \frac{J + T_p\sqrt{12\ln(10)}}{T_pJ}$$

Therefore,

$$T_g = \frac{1}{\frac{J+T_p\sqrt{12\ln(10)}}{T_pJ}} = \frac{T_pJ}{J + T_p\sqrt{12\ln(10)}}$$

—

Step 2: Compare with model expression

Model expression is:

$$T_g = \frac{T_p}{1 + \frac{T_p}{J}\sqrt{12\ln 10}} = \frac{T_p}{1 + \frac{T_p\sqrt{12\ln 10}}{J}} = \frac{T_p}{\frac{J + T_p\sqrt{12\ln 10}}{J}} = \frac{T_p J}{J + T_p\sqrt{12\ln 10}}$$

This matches exactly the simplified form of the ground truth.

—

Conclusion:

The model-generated expression is mathematically equivalent to the ground truth expression, just written in a different but equivalent form.

Score: 1

—

Final answer:

Score: 1

Problem 24 (Paper: 2506.21290v1)

Problem Statement

Background: In supercooled liquids, the structural relaxation time τ_α is intimately linked to the growth of dynamically correlated regions. The characteristic volume of these regions can be quantified by the peak value of the normalized four-point susceptibility, denoted as $(\chi_4/\chi_{\text{self}})^*$. Let $X = (\chi_4/\chi_{\text{self}})^*$. It has been empirically established that the relationship between τ_α and X can be described by a functional form that combines a power-law prefactor and an exponential term. The power-law prefactor is characterized by an integer exponent, and the argument of the exponential term is also characterized by an integer exponent, which is the same as the power-law prefactor's exponent. Furthermore, it is known that for this type of relationship, the relaxation time τ_α scales quadratically with X in its power-law prefactor. The exponential term's argument is also found to scale quadratically with a scaled X. The general form of this relationship is $\tau_\alpha = AX^\gamma \exp\left\{[BX]^\psi\right\}$, where A, B, γ, and ψ are empirical parameters.

Task: Determine the complete empirical functional form that describes the relationship between the structural relaxation time τ_α and the normalized peak four-point susceptibility $X = (\chi_4/\chi_{\text{self}})^*$, including the specific values of the empirical parameters A, B, γ, and ψ.

Ground Truth Solution

$$\boxed{\tau_\alpha \simeq 0.10 \left[\left(\frac{\chi_4}{\chi_{\text{self}}}\right)^*\right]^2 \exp\left\{\left[1.22 \times 10^{-3} \left(\frac{\chi_4}{\chi_{\text{self}}}\right)^*\right]^2\right\}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: $\tau_\alpha = X^2 \exp(X^2)$

Judge's Evaluation: The ground truth expression is:

$$\tau_\alpha \simeq 0.10 \left[\left(\frac{\chi_4}{\chi_{\text{self}}}\right)^*\right]^2 \exp\left\{\left[1.22 \times 10^{-3} \left(\frac{\chi_4}{\chi_{\text{self}}}\right)^*\right]^2\right\}$$

The model-generated expression is:

$$\tau_\alpha = X^2 \exp(X^2)$$

where presumably X is meant to represent $\left(\frac{\chi_4}{\chi_{\text{self}}}\right)^*$ or some related quantity.

Comparison:

- The model expression omits the prefactor 0.10. - The argument inside the exponential in the ground truth is $\left[1.22 \times 10^{-3} \left(\frac{\chi_4}{\chi_{\text{self}}}\right)^*\right]^2$, which is $(1.22 \times 10^{-3})^2 \times \left(\frac{\chi_4}{\chi_{\text{self}}}\right)^{*2}$. - The model expression uses simply X^2 inside the exponential, missing the factor 1.22×10^{-3}. - The model expression also uses an equals sign instead of $\simeq$, but this is minor. - The model expression is missing the scaling factor outside the exponential.

Because the model expression misses the numerical prefactor outside the exponential and the small coefficient inside the exponential, it is not mathematically equivalent to the ground truth expression.

This is more than one minor error: missing the prefactor and missing the coefficient inside the exponential.

Score: 0

—

Score: 0

Problem 25 (Paper: 2506.21318v1)

Problem Statement

Background: Consider a quantum system of N_S qubits coupled to a bath of N_B auxiliary qubits. The bath qubits are initially prepared in the product state $\hat{\Phi} = 00^{\otimes N_B}$. The system's density matrix $\hat{\rho}$ evolves under a repeated quantum channel, referred to as a "reset cycle," defined as $\mathcal{E}(\hat{\rho}) = \mathrm{tr}_B[\mathcal{Q}(\hat{\rho} \otimes \hat{\Phi})\mathcal{Q}^\dagger]$.

For this derivation, consider the unrandomized protocol, where the joint unitary $\mathcal{Q}$ is decomposed into $2M_T + 1$ unitary layers: $\mathcal{Q} = \mathcal{U}_{M_T} \dots \mathcal{U}_0 \dots \mathcal{U}_{-M_T}$. Each layer $\mathcal{U}_\tau$ is given by $\mathcal{U}_\tau = e^{-i\delta\theta\hat{V}_\tau} e^{-i\delta\hat{H}_B} e^{-i\delta\hat{H}_S}$, where θ is a small system-bath coupling parameter and δ is a small Trotter angle. $\hat{H}_S$ is the system Hamiltonian. The bath Hamiltonian is $\hat{H}_B = -\frac{h}{2}\sum_{\mu=1}^{N_B} \hat{Z}_\mu$, where h is a single bath energy scale and $\hat{Z}_\mu$ is the Pauli-Z matrix on the μ-th auxiliary qubit. The τ-dependent coupling operator is $\hat{V}_\tau = f_\tau \sum_{\mu=1}^{N_B} \hat{A}_\mu \hat{Y}_\mu$, where f_τ is a real-valued modulating function, $\hat{A}_\mu$ are Hermitian system operators, and $\hat{Y}_\mu$ is the Pauli-Y matrix on the μ-th auxiliary qubit.

The interaction picture is defined with respect to the free system-bath evolution as $\tilde{O}_\tau = U_0^{-\tau}(\hat{O}_\tau)U_0^\tau$, where $U_0 = e^{-i\delta\hat{H}_B} e^{-i\delta\hat{H}_S}$. In this picture, the interaction-picture map is $\tilde{\mathcal{E}}(\hat{\rho}) = \mathrm{tr}_B\left[\tilde{\mathcal{Q}}(\hat{\rho} \otimes \hat{\Phi})\tilde{Q}^\dagger\right]$. The bath correlation function is given by $\mathrm{tr}_B(\tilde{Y}_{\mu,\tau}\tilde{Y}_{\mu',\tau'}\hat{\Phi}) = \delta_{\mu,\mu'} e^{ih\delta(\tau'-\tau)}$.

Task: Derive the interaction-picture map $\tilde{\mathcal{E}}(\hat{\rho})$ to second order in the system-bath coupling parameter θ. Express the result in the Lindblad form: $\frac{\tilde{\mathcal{E}}(\hat{\rho})-\hat{\rho}}{\theta^2} = -i[\hat{G}^{\mathrm{LS}}, \hat{\rho}] + \{\hat{K}, \hat{\rho}\} + \sum_\mu \hat{L}_\mu \hat{\rho} \hat{L}_\mu^\dagger$. Specifically, derive the expressions for the jump operator $\hat{L}_\mu$, the Hermitian operator $\hat{K}$, and the Lamb-shift Hamiltonian $\hat{G}^{\mathrm{LS}}$ in terms of the given parameters and operators.

Ground Truth Solution

$$\boxed{\begin{aligned} \hat{L}_\mu &= \delta \sum_{\tau=-M_T}^{M_T} f_\tau e^{i\delta h\tau} \tilde{A}_{\mu\tau} \\ \hat{K} &= \frac{\delta^2}{2} \sum_\mu \sum_{\tau,\tau'=-M_T}^{M_T} e^{ih\delta(\tau'-\tau)} f_\tau f_{\tau'} \tilde{A}_{\mu,\tau} \tilde{A}_{\mu,\tau'} \\ \hat{G}^{\mathrm{LS}} &= \frac{\delta^2}{2i} \sum_{\mu,\tau,\tau'} \mathrm{sgn}(\tau-\tau') e^{i\delta h(\tau'-\tau)} f_\tau f_{\tau'}\ \hat{A}_{\mu\tau}^\dagger \hat{A}_{\mu\tau'} \end{aligned}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: To second order in the small system–bath coupling angle the joint unitary = exp(i V) may be expanded as

i V ² ²/2 , [V V] , (1)

with the discrete time–ordering operator and

V = f Â, Ŷ, , Â, U0 Â U0 , Ŷ, e+ih Ŷ . (2)

Tracing the state over the bath and using, for the initially polarised bath,

trB[Ŷ, Ŷ,] = , e$^{ih()}$, (3)

all terms that are linear in vanish, so that

() = + ² () + O(³) . (4)

Carrying out the second–order contraction with (3) one finds

()= ² , f f e$^{ih()}$(5)Ö[$\hat{A}, \hat{A},\cdot \hat{A}, \hat{A},$,].

To cast (5) into the standard Gorini–Kossakowski–Sudarshan–Lindblad (GKSL) form, separate the bath kernel into the real (even) and imaginary (odd) parts

C ² f f cos[h()], C ² f f sin[h()], (6)

so that e$^{ih()} = C + iC$.

(i) Jump (Lindblad) operators. Because the matrix C (C) is positive semidefinite it can be diagonalised, C = k k —vk)(vk—. For every bath qubit we therefore introduce

Lk k (vk) Â, . (7)

In the following we keep the index k implicit; the symbol already denotes the total sum over all jump operators.

(ii) Hermitian part coming from the symmetric kernel,

K = ½ , C Â, Â, = ½ L † L , (8)
(the second equality follows directly from (7)).
(iii) The Lamb–shift Hamiltonian, arising from the antisymmetric kernel,
G^{LS} = $\cdot\,, C\hat{A}, \hat{A}, .$(9)
Putting everything together, the interaction–picture map reads
[()] / ^{2} = i [G^{LS}$,] + K, + LL\check{D}$.(10)
Explicitly, with (6)–(9),

boxed $\dfrac{\tilde{\mathcal{E}}(\hat\rho)-\hat\rho}{\theta^2} = -\,i\left[\dfrac{1}{2}\sum_{\mu=1}^{N_B}\sum_{\tau,\tau'=-M_T}^{M_T}\delta^2 f_\tau f_{\tau'}\sin\bigl[h\delta(\tau'-\tau)\bigr]\,\tilde{A}_{\mu,\tau'}\tilde{A}_{\mu,\tau},\ \hat\rho\right] + \Biggl\{-\dfrac{1}{2}\sum_{\mu=1}^{N_B}\sum_{\tau,\tau'=-M_T}^{M_T}\delta^2 f_\tau f_{\tau'}\cos\bigl[h\delta(\tau'-\tau$

$\sum_{\mu=1}^{N_B}\hat{L}_\mu\hat\rho\hat{L}_\mu^\dagger,$

with the (time–coarse–grained) jump operators

$\hat{L}_\mu = \delta\sum_{\tau=-M_T}^{M_T} f_\tau\, e^{ih\delta\tau}\,\tilde{A}_{\mu,\tau}\ , \qquad (\mu = 1,\ldots,N_B),$

and where = -1$\overline{}$ $_{2\sum_\mu \hat{L}_\mu^\dagger \hat{L}_\mu, \hat{G}^{\mathrm{LS}} as given in the commutator term above.}$

Equations (7)–(10) constitute the desired second–order, interaction–picture Lindblad generator for the un-randomised reset cycle of the system–bath Trotter sequence.

Judge's Evaluation: Evaluation Error: No □ expression found in the model's output.

Problem 26 (Paper: 2506.21318v1)

Problem Statement

Background: Consider a quantum system with Hamiltonian $\hat{H}_S$, whose eigenstates are ϕ_a with corresponding eigenvalues ϵ_a. The transition frequencies are denoted by $\omega_{ab} = \epsilon_b - \epsilon_a$, and the average energies by $\bar{\epsilon}_{ab} = (\epsilon_a + \epsilon_b)/2$. The target thermal state is the Gibbs state, $\hat{\sigma}_\beta = \frac{1}{Z} e^{-\beta \hat{H}_S}$, where $Z = \mathrm{tr}(e^{-\beta \hat{H}_S})$ is the partition function and β is the inverse temperature.

The system evolves under a quantum channel $\mathcal{E}(\hat{\rho})$, which describes a reset cycle involving joint unitary evolution and bath qubit reset. At late times, the system converges to a fixed point $\hat{\sigma}$ of this channel. Under a weak system-bath coupling assumption, characterized by a small parameter θ, the steady state $\hat{\sigma}$ can be expanded as $\hat{\sigma} = \hat{\sigma}_0 + \theta^2 \hat{\zeta}$, where $\hat{\sigma}_0 = \hat{\sigma}_\beta$ represents the zeroth-order solution for the diagonal elements in the Hamiltonian eigenbasis.

The interaction-picture map $\tilde{\mathcal{E}}(\hat{\rho})$ is related to the Schrödinger-picture map $\mathcal{E}(\hat{\rho})$ by $\mathcal{E}(\hat{\rho}) = \mathcal{S}^{M_T+M_R} \circ \tilde{\mathcal{E}} \circ \mathcal{S}^{M_T+1}(\hat{\rho})$, where $\mathcal{S}(\hat{\rho}) = e^{-i\delta \hat{H}_S}(\hat{\rho}) e^{i\delta \hat{H}_S}$ is the system unitary evolution channel, δ is a small Trotter angle, and M_T is the number of unitary layers. An optional randomization step involves a random integer M_R, leading to an averaged dephasing channel $\mathcal{D}(\hat{\rho}) = \frac{1}{T_0} \int_0^\infty dT_R\, e^{-T_R/T0} e^{-i\hat{H}_S T_R}(\hat{\rho}) e^{i\hat{H}_S T_R} = \sum_{ab} \frac{\phi_a \phi_a \hat{\rho} \phi_b \phi_b}{1 - i\omega_{ab} T_0}$, where T_0 is a characteristic randomization time. The total reset time is $T = M_T \delta$.

The interaction-picture map $\tilde{\mathcal{E}}(\hat{\rho})$ can be expanded to second order in θ and expressed in a Lindblad-like form: $\frac{\tilde{\mathcal{E}}(\hat{\rho}) - \hat{\rho}}{\theta^2} = -i[\hat{G}^{\mathrm{LS}}, \hat{\rho}] + \{\hat{K}, \hat{\rho}\} + \sum_\mu \hat{L}_\mu \hat{\rho} \hat{L}_\mu^\dagger$. Here, $\hat{L}_\mu = \int_{-\infty}^\infty dt\, g(t) \tilde{A}_\mu(t)$ are the jump operators, where $g(t) = e^{iht} f_{t/\delta}$ and $f_\tau = \exp(-\frac{a^2\delta^2\tau^2}{2})$ is the filter function with $a = \sqrt{\frac{4h}{\beta}}$ and h being the bath energy scale. $\tilde{A}_\mu(t) = e^{i\hat{H}_S t} \hat{A}_\mu e^{-i\hat{H}_S t}$ are interaction-picture system operators, with $\hat{A}_\mu$ being local system operators. The operator $\hat{K} = \frac{1}{2}\sum_\mu \hat{L}_\mu^\dagger \hat{L}_\mu$. The Lamb shift Hamiltonian is $\hat{G}^{\mathrm{LS}} = \frac{1}{2i} \sum_\mu \iint_{-\infty}^\infty dt dt'\, \mathrm{sgn}(t-t') g^*(t) g(t') \tilde{A}_\mu^\dagger(t) \tilde{A}_\mu(t')$.

For the Gibbs state to be an exact fixed point, the Lindbladian must satisfy quantum detailed balance (QDB), which requires an ideal Lamb shift $\hat{G}^{\mathrm{DB}}$ such that $\hat{G}^{\mathrm{DB}}_{ab} = -i \tanh\left(\frac{\beta\omega_{ab}}{4}\right) \hat{K}_{ab}$. The actual protocol's Lamb shift $\hat{G}^{\mathrm{LS}}$ differs from this ideal form, leading to a deviation $\Delta\hat{G} = \hat{G}^{\mathrm{LS}} - \hat{G}^{\mathrm{DB}}$.

The leading-order correction to the coherences of the steady state, $\hat{\zeta}_{ab}$ (for $a \neq b$), is given by: $\hat{\zeta}_{ab} = -\frac{e^{-\beta\bar{\epsilon}_{ab}}}{Z} \frac{2i \sinh(\beta\omega_{ab}/2) e^{iT\omega_{ab}}}{[1 - e^{2iT\omega_{ab}} - i\omega_{ab} T_0]} (\Delta\hat{G})_{ab}$.

The steady state $\hat{\sigma}$ can also be viewed as approximately proportional to a Gibbs state defined with respect to a renormalized system Hamiltonian $\hat{H}'_S = \hat{H}_S + \theta^2 \hat{C}$, where $\hat{C}$ is a matrix representing the leading-order correction to $\hat{H}_S$ due to the bath interaction. The expansion of $e^{-\beta(\hat{H}_S + \theta^2 \hat{C})}$ using the Kubo formula is $e^{-\beta(\hat{H}_S + \theta^2\hat{C})} = e^{-\beta \hat{H}_S}\left(1 - \theta^2 \int_0^\beta d\beta' \hat{C}(\beta') + \mathcal{O}(\theta^4)\right)$, where $\hat{C}(\beta') = e^{\beta' \hat{H}_S} \hat{C} e^{-\beta' \hat{H}_S}$.

Task: Derive the matrix elements $\hat{C}_{ab}$ of the leading-order correction $\hat{C}$ to the system Hamiltonian $\hat{H}_S$, in the eigenbasis of $\hat{H}_S$, by equating the second-order steady-state coherence correction $\hat{\zeta}_{ab}$ with the corresponding term from the Kubo formula expansion of the renormalized Gibbs state $\hat{\sigma} \propto e^{-\beta \hat{H}'_S}$. Note that the diagonal part Λ_a remains undetermined in this second-order scheme.

Ground Truth Solution

$$\boxed{\hat{C}_{ab} = -\frac{i\omega_{ab} e^{i\omega_{ab} T}}{1 - i\omega_{ab} T_0 - e^{2i\omega_{ab} T}} (\Delta\hat{G})_{ab} + \delta_{ab} \Lambda_a}$$

Model Outputs

Model: o3 (Score: 0.5)

Model Solution: $$\mathrm{C}_{ab} = \begin{cases} \Lambda_a, & a = b, \\ \dfrac{i\,\omega_{ab}\, e^{iT\omega_{ab}}}{1 - e^{2iT\omega_{ab}} - i\omega_{ab} T_0} (\Delta G)_{ab}, & a \neq b, \end{cases}$$

Judge's Evaluation: The ground truth expression is a single formula for $\hat{C}_{ab}$ that includes both the diagonal and off-diagonal terms in one expression:

$$\hat{C}_{ab} = -\frac{i\omega_{ab}e^{i\omega_{ab}T}}{1 - i\omega_{ab}T_0 - e^{2i\omega_{ab}T}}(\Delta\hat{G})_{ab} + \delta_{ab}\Lambda_a$$

The model-generated expression splits the diagonal and off-diagonal terms explicitly:

$$C_{ab} = \begin{cases} \Lambda_a, & a = b, \\ \dfrac{i\omega_{ab}e^{iT\omega_{ab}}}{1 - e^{2iT\omega_{ab}} - i\omega_{ab}T_0}(\Delta G)_{ab}, & a \neq b, \end{cases}$$

Let's analyze the differences:

1. **Sign difference:** The ground truth has a negative sign in front of the fraction for the off-diagonal terms, while the model expression has a positive sign.

2. **Order of terms in denominator:** Ground truth denominator: $1 - i\omega_{ab}T_0 - e^{2i\omega_{ab}T}$ Model denominator: $1-e^{2iT\omega_{ab}}-i\omega_{ab}T_0$ The order of the last two terms is swapped, but addition is commutative, so this is equivalent.

3. **Variable notation:** The ground truth uses $\hat{C}_{ab}$ and $(\Delta\hat{G})_{ab}$, while the model uses C_{ab} and $(\Delta G)_{ab}$. This is acceptable as variable names can differ.

4. **Diagonal terms:** Ground truth uses $\delta_{ab}\Lambda_a$, which is equivalent to the model's piecewise definition Λ_a when $a = b$.

5. **Exponent order:** Both have $e^{i\omega_{ab}T}$ in the numerator, just written as $e^{iT\omega_{ab}}$ in the model, which is equivalent.

Summary: The only substantive difference is the missing negative sign in the off-diagonal term in the model expression. This is a significant error because it changes the sign of the entire off-diagonal term.

Since there is one major sign error (not a minor missing coefficient or constant), the model expression is not completely correct. However, it is very close to the ground truth, with only one sign error.

According to the rubric, one wrong sign is a minor error and warrants partial credit.

Score: 0.5

—

Score: 0.5

Problem 27 (Paper: 2506.21400v1)

Problem Statement

Background: Consider a two-dimensional Hermitian Hamiltonian h_0 given by $h_0 = p_x^2 - p_y^2 + \nu^2 x^2 + \Omega y^2 + gxy$, where x, y are position operators, p_x, p_y are momentum operators, and ν, Ω, g are real parameters. The ground state wavefunction of h_0 is $\phi_0(x,y) = C_0 e^{-\frac{\alpha x^2}{2} - \frac{\beta y^2}{2} + \gamma xy}$, where C_0 is a normalization constant and α, β, γ are real parameters.

We introduce a non-unitary similarity transformation using two operators, η_0 and η_1. The first operator is $\eta_0 = \exp\left(-\frac{\delta}{2}x^2 - \frac{\lambda}{2}y^2\right)$, where δ, λ are real parameters. Its adjoint actions on the momentum operators are $\eta_0 p_x \eta_0^{-1} = p_x - i\delta x$ and $\eta_0 p_y \eta_0^{-1} = p_y - i\lambda y$. The position operators transform as $\eta_0 x \eta_0^{-1} = x$ and $\eta_0 y \eta_0^{-1} = y$. The Hamiltonian h_0 is transformed into a non-Hermitian Hamiltonian $H_1 = \eta_0 h_0 \eta_0^{-1}$. The corresponding ground state wavefunction is $\psi_1(x,y) = \eta_0(x,y)\phi_0(x,y)$.

The second operator is $\eta_1 = \exp\left(\frac{\kappa}{2}p_x^2 + \frac{\xi}{2}p_y^2\right)$, where κ, ξ are real parameters. Its adjoint actions on the position operators are $\eta_1 x \eta_1^{-1} = x - i\kappa p_x$ and $\eta_1 y \eta_1^{-1} = y - i\xi p_y$. The momentum operators transform as $\eta_1 p_x \eta_1^{-1} = p_x$ and $\eta_1 p_y \eta_1^{-1} = p_y$. The Hamiltonian H_1 is further transformed into $H_2 = \eta_1 H_1 \eta_1^{-1}$. For this transformation, the parameters κ and ξ are chosen as $\kappa = \frac{\delta}{\delta^2 - \nu^2}$ and $\xi = \frac{\lambda}{\lambda^2 + \Omega}$. The ground state wavefunction $\psi_2(x,y)$ corresponding to H_2 is obtained by applying η_1 to $\psi_1(x,y)$ in momentum space: $\psi_2(x,y) = \mathcal{F}_{xy}^{-1}\left[\eta_1(p_x,p_y)\hat{\psi}_1(p_x,p_y)\right](x,y)$, where $\hat{\psi}_1(p_x,p_y)$ is the Fourier transform of $\psi_1(x,y)$. The Fourier transform of a 2D Gaussian function $f(\mathbf{x}) = e^{-\frac{1}{2}\mathbf{x}^T M \mathbf{x}}$ is $\hat{f}(\mathbf{p}) = \frac{(2\pi)^{n/2}}{\sqrt{\det M}} e^{-\frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p}}$. The inverse Fourier transform of a 2D Gaussian function $\hat{f}(\mathbf{p}) = e^{-\frac{1}{2}\mathbf{p}^T M_p \mathbf{p}}$ is $f(\mathbf{x}) = \frac{(2\pi)^{n/2}}{\sqrt{\det M_p}} e^{-\frac{1}{2}\mathbf{x}^T M_p^{-1}\mathbf{x}}$. The wavefunction $\psi_2(x,y)$ takes the form $C_2 e^{-\frac{\hat{\alpha}x^2}{2} - \frac{\hat{\beta}y^2}{2} + \hat{\gamma}xy}$, where C_2 is a normalization constant. You may ignore normalization constants C_0, C_1, C_2 in your derivation of the exponent parameters.

Task: Determine the expression for the parameter $\hat{\alpha}$.

Ground Truth Solution

$$\boxed{\hat{\alpha} = \frac{\left[(\alpha+\delta)(\beta\lambda - \Omega) - \gamma^2\lambda\right]\left(\delta^2 - \nu^2\right)}{\left(\alpha\delta + \nu^2\right)\left(\Omega - \beta\lambda\right) + \gamma^2\delta\lambda}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: $$\hat{\alpha} = \frac{\frac{\alpha+\delta}{(\alpha+\delta)(\beta+\lambda)-\gamma^2} - \frac{\lambda}{\lambda^2+\Omega}}{\left[\frac{\beta+\lambda}{(\alpha+\delta)(\beta+\lambda)-\gamma^2} - \frac{\delta}{\delta^2-\nu^2}\right]\left[\frac{\alpha+\delta}{(\alpha+\delta)(\beta+\lambda)-\gamma^2} - \frac{\lambda}{\lambda^2+\Omega}\right] - \left[\frac{\gamma}{(\alpha+\delta)(\beta+\lambda)-\gamma^2}\right]^2}$$

Judge's Evaluation: The ground truth expression is a single rational expression with a clear numerator and denominator:

$$\hat{\alpha} = \frac{\left[(\alpha+\delta)(\beta\lambda - \Omega) - \gamma^2\lambda\right]\left(\delta^2 - \nu^2\right)}{\left(\alpha\delta + \nu^2\right)\left(\Omega - \beta\lambda\right) + \gamma^2\delta\lambda}$$

The model-generated expression is a complicated nested fraction involving sums and differences of fractions, with denominators like $(\alpha+\delta)(\beta+\lambda) - \gamma^2$ and $\lambda^2 + \Omega$, and terms like $\frac{\alpha+\delta}{(\alpha+\delta)(\beta+\lambda)-\gamma^2}$, etc.

Key observations:

- The ground truth has terms $\beta\lambda - \Omega$ and $\Omega - \beta\lambda$, but the model uses $\beta + \lambda$ and $\lambda^2 + \Omega$, which are different. - The ground truth denominator is $\left(\alpha\delta + \nu^2\right)\left(\Omega - \beta\lambda\right) + \gamma^2\delta\lambda$, but the model denominator is a complicated expression involving sums and products of fractions, none of which match the ground truth denominator. - The model expression is structurally very different and does not simplify obviously to the ground truth. - The model expression contains multiple nested fractions and terms that do not appear in the ground truth. - The model expression has $\beta + \lambda$ where the ground truth has $\beta\lambda$, which is a significant difference. - The model expression has $\lambda^2 + \Omega$ where the ground truth has $\Omega - \beta\lambda$, again a significant difference.

Given these differences, the model expression is not mathematically equivalent to the ground truth. The differences are more than minor errors; they involve different operations (addition vs multiplication), different signs, and different terms.

Therefore, the model expression is incorrect.

Score: 0

Problem 28 (Paper: 2506.21400v1)

Problem Statement

Background: In quantum mechanics, canonical variables x, y, p_x, p_y satisfy the commutation relations $[x, p_x] = i$, $[y, p_y] = i$, and all other commutators are zero. Consider a non-unitary similarity transformation operator η_2 defined as $\eta_2 = \exp(\mu p_x p_y + \tau xy)$, where $\mu, \tau \in \mathbb{R}$. The parameter θ is defined as $\theta := \sqrt{\mu\tau}$. The adjoint action of an operator A on another operator B is given by the Baker-Campbell-Hausdorff formula:

$$e^A B e^{-A} = B + [A, B] + \frac{1}{2!}[A, [A, B]] + \frac{1}{3!}[A, [A, [A, B]]] + \cdots$$

Task: Derive the expression for the adjoint action of η_2 on the position operator x.

Ground Truth Solution

$$\boxed{\eta_2 x \eta_2^{-1} = \cosh(\theta) x - i\frac{\mu}{\theta}\sinh(\theta) p_y}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\eta_2\, x\, \eta_2^{-1} = x \cosh\!\big(\sqrt{\mu\tau}\big) \;-\; i\, \frac{\mu}{\sqrt{\mu\tau}} \sinh\!\big(\sqrt{\mu\tau}\big)\, p_y$

Judge's Evaluation: The ground truth expression is:

$$\eta_2 x \eta_2^{-1} = \cosh(\theta) x - i\frac{\mu}{\theta}\sinh(\theta) p_y$$

The model-generated expression is:

$$\eta_2 x \eta_2^{-1} = x\cosh(\sqrt{\mu\tau}) - i\frac{\mu}{\sqrt{\mu\tau}}\sinh(\sqrt{\mu\tau}) p_y$$

Comparison: - The ground truth uses the parameter θ. - The model uses $\sqrt{\mu\tau}$ in place of θ. - The structure of the expression is the same: a hyperbolic cosine term multiplying x, and a hyperbolic sine term multiplied by $-i\frac{\mu}{\theta}p_y$. - The model has the same form but replaces θ with $\sqrt{\mu\tau}$.

Since the problem states that differences in variable names are acceptable if the expressions are mathematically equivalent, the key question is whether $\theta = \sqrt{\mu\tau}$ or not.

If θ is defined as $\sqrt{\mu\tau}$, then the expressions are equivalent. However, the ground truth expression explicitly uses θ, and the model introduces $\sqrt{\mu\tau}$ without explanation.

Because the model replaces θ with $\sqrt{\mu\tau}$ without clarification, this could be a variable substitution or a redefinition. If θ is defined as $\sqrt{\mu\tau}$, then the expressions are equivalent.

Given the rubric allows differences in variable names, and the model's expression matches the ground truth in form and coefficients, the model's expression is mathematically equivalent assuming $\theta = \sqrt{\mu\tau}$.

Therefore, the model's expression is completely correct.

Score: 1

Problem 29 (Paper: 2506.21415v1)

Problem Statement

Background: The dynamics of a magnetized plasma can be described by an infinite-dimensional Hamiltonian system on the space $\mathfrak{s}^*$ of triples (n, P, ϱ), where $n(q)$ is the electron density, $P(q)$ is the electron momentum density, and $\varrho(q,\xi)$ is the centered electron velocity distribution. The spatial domain $q = (x, y)$ is a 2-torus $\mathbb{T}^2$ with period 2π in each direction, and the velocity domain $\xi = (\xi_x, \xi_y)$ is $\mathbb{R}^2$. The system is characterized by a dimensionless parameter ϵ and a magnetic field $B(q)$.

The following definitions and operators are used:

- The skew-symmetric matrix $\mathbb{J} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.
- The Laplace operator $\Delta = \partial_x^2 + \partial_y^2$.
- The inverse of Δ for zero-mean functions is denoted by $\mathcal{G}$.
- The L^2-orthogonal projection onto divergence-free vector fields is denoted by Π.
- For any function $Q(q,\xi)$, its average over velocity space is $\langle Q \rangle = \int Q\,\varrho\,d\xi$.
- The fluid vorticity is $\Omega = -\partial_q \cdot (\mathbb{J} n^{-1} P)$.
- The Lie bracket for vector fields is $[A, B] = A \cdot \partial_q B - B \cdot \partial_q A$.

The Poisson bracket $\{\cdot,\cdot\}_E$ between two functionals $F, G : \mathfrak{s}^* \to \mathbb{R}$ is given by:

$$
\begin{aligned}
\{F,G\}_E = &-\epsilon \int \left(\frac{\delta F}{\delta P} \cdot \left[\partial_q \frac{\delta G}{\delta n} - \partial_q \left\langle \frac{1}{n}\frac{\delta G}{\delta \varrho} \right\rangle + \left\langle \partial_q \frac{1}{n}\frac{\delta G}{\delta \varrho} \right\rangle \right] - \frac{\delta G}{\delta P} \cdot \left[\partial_q \frac{\delta F}{\delta n} - \partial_q \left\langle \frac{1}{n}\frac{\delta F}{\delta \varrho} \right\rangle + \left\langle \partial_q \frac{1}{n}\frac{\delta F}{\delta \varrho} \right\rangle \right] \right) n\,dq \\
&- \epsilon \int \left[\frac{\delta F}{\delta P}, \frac{\delta G}{\delta P} \right] \cdot P\,dq + \int B \frac{\delta F}{\delta P} \cdot \mathbb{J} \frac{\delta G}{\delta P}\, n\,dq \\
&+ \epsilon \int \left(\partial_q \frac{1}{n}\frac{\delta F}{\delta \varrho} \cdot \partial_\xi \frac{1}{n}\frac{\delta G}{\delta \varrho} - \partial_q \frac{1}{n}\frac{\delta G}{\delta \varrho} \cdot \partial_\xi \frac{1}{n}\frac{\delta F}{\delta \varrho} \right) n\,\varrho\,d\xi\,dq + \int (B - \epsilon\Omega) \left(\partial_\xi \frac{1}{n}\frac{\delta F}{\delta \varrho} \right) \cdot \mathbb{J} \left(\partial_\xi \frac{1}{n}\frac{\delta G}{\delta \varrho} \right) n\,\varrho\,d\xi\,dq \\
&- \epsilon \int \left(\left\langle \partial_q \frac{1}{n}\frac{\delta F}{\delta \varrho} \right\rangle \cdot \left\langle \partial_\xi \frac{1}{n}\frac{\delta G}{\delta \varrho} \right\rangle - \left\langle \partial_q \frac{1}{n}\frac{\delta G}{\delta \varrho} \right\rangle \cdot \left\langle \partial_\xi \frac{1}{n}\frac{\delta F}{\delta \varrho} \right\rangle \right) n\,dq - \int (B - \epsilon\Omega) \left\langle \partial_\xi \frac{1}{n}\frac{\delta F}{\delta \varrho} \right\rangle \cdot \mathbb{J} \left\langle \partial_\xi \frac{1}{n}\frac{\delta G}{\delta \varrho} \right\rangle n\,dq \\
&- \epsilon \int \left(\left[\frac{\delta F}{\delta P} - \left\langle \partial_\xi \frac{1}{n}\frac{\delta F}{\delta \varrho} \right\rangle \right] \cdot \partial_q \cdot \left[n \left\langle \partial_\xi \left(\frac{1}{n}\frac{\delta G}{\delta \varrho} \right) \xi \right\rangle \right] - \left[\frac{\delta G}{\delta P} - \left\langle \partial_\xi \frac{1}{n}\frac{\delta G}{\delta \varrho} \right\rangle \right] \cdot \partial_q \cdot \left[n \left\langle \partial_\xi \left(\frac{1}{n}\frac{\delta F}{\delta \varrho} \right) \xi \right\rangle \right] \right) dq.
\end{aligned}
$$

The quasineutral limit of this system corresponds to dynamics on a specific submanifold $\Sigma \subset \mathfrak{s}^*$, defined by the conditions $\partial_q n = 0$ and $\partial_q \cdot P = 0$. This implies that on Σ, n is a constant, denoted n_0, and P is a divergence-free vector field, denoted π. Thus, elements of Σ are of the form (n_0, π, ϱ).

For a point $\sigma = (n_0, \pi, \varrho) \in \Sigma$, the tangent space $T_\sigma\Sigma$ is given by $T_\sigma\Sigma = \{(\delta n_0, \delta\pi, \delta\varrho) \mid \delta n_0 \in \mathbb{R},\quad \delta\pi : \mathbb{T}^2 \to \mathbb{R}^2,\quad \delta\varrho : \mathbb{T}^2 \times \mathbb{R}^2 \to \mathbb{R},\quad \partial_q \cdot \delta\pi = 0\}$. The annihilator of $T_\sigma\Sigma$ in the dual tangent space $T^*_\sigma\mathfrak{s}^*$ is $T_\sigma\Sigma^\circ = \{(\delta n^*, \delta P^*, \delta\varrho^*) \in T^*_\sigma\mathfrak{s}^* \mid \exists \delta\Phi^* : \mathbb{T}^2 \to \mathbb{R},\quad \int \delta\Phi^*\,dq = 0,\quad \int \delta n^*\,dq = 0,\quad \delta P^* = \partial_q \delta\Phi^*,\quad \delta\varrho^* = 0\}$.

The submanifold Σ is a Poisson-Dirac submanifold, meaning it inherits a Poisson bracket $\{\cdot,\cdot\}_\Sigma$ from the ambient space $\mathfrak{s}^*$. For functionals $F, G : \Sigma \to \mathbb{R}$, the induced bracket $\{F,G\}_\Sigma(\sigma)$ is computed using the Poisson tensor π_E associated with $\{\cdot,\cdot\}_E$ and specific covector extensions $\widetilde{(dF_\sigma)}, \widetilde{(dG_\sigma)} \in T^*_\sigma\mathfrak{s}^*$. These extensions must satisfy two conditions: 1. They must agree with the differentials of F and G on $T_\sigma\Sigma$. For example, $\widetilde{(dG_\sigma)} \mid T_\sigma\Sigma = dG_\sigma$. 2. They must annihilate the image of $T_\sigma\Sigma^\circ$ under the Poisson tensor, i.e., $\widetilde{(dG_\sigma)} \mid \widehat{\pi}_E(T_\sigma\Sigma^\circ) = 0$. The map $\widehat{\pi}_E : T^*_\sigma\mathfrak{s}^* \to T_\sigma\mathfrak{s}^*$ is defined by the general Hamilton's equations, where $\partial_t n, \partial_t P, \partial_t(n\varrho)$ represent the components of $\widehat{\pi}_E$ when the functional derivatives $\frac{\delta G}{\delta n}, \frac{\delta G}{\delta P}, \frac{\delta G}{\delta \varrho}$ are replaced by the covector components $\delta n^*, \delta P^*, \delta\varrho^*$.

For a functional $G : \Sigma \to \mathbb{R}$, its differential dG_σ on $T_\sigma\Sigma$ is given by $dG_\sigma(\delta n_0, \delta\pi, \delta\varrho) = \int \frac{\delta G}{\delta n_0}\,\delta n_0\,dq + \int \frac{\delta G}{\delta \pi} \cdot \delta\pi\,dq + \int \frac{\delta G}{\delta \varrho}\,\delta\varrho\,d\xi\,dq$. The covector extension $\widetilde{(dG_\sigma)}$ can be written as $\int \widetilde{\frac{\delta G}{\delta n_0}}\,\delta n\,dq + \int \widetilde{\frac{\delta G}{\delta \pi}} \cdot \delta P\,dq + \int \widetilde{\frac{\delta G}{\delta \varrho}}\,\delta\varrho\,d\xi\,dq$. The conditions for $\widetilde{(dG_\sigma)}$ imply that the effective functional derivatives to be used in the ambient bracket $\{\cdot,\cdot\}_E$ (where n is replaced by n_0 and P by π) are:

$$\frac{\delta G}{\delta n} \to \frac{\delta G}{\delta n_0} - \frac{\pi}{n_0}\cdot\frac{\delta G}{\delta \pi} + \Delta^{-1}\left(\partial_q\cdot\left[\frac{1}{\epsilon}(B-\epsilon\,\Omega)\mathbb{J}\frac{\delta G}{\delta\pi} - \partial_q\cdot\left(\left\langle \partial_\xi\left(\frac{1}{n_0}\frac{\delta G}{\delta\varrho}\right)\xi\right\rangle\right) + \partial_q\left\langle\frac{1}{n_0}\frac{\delta G}{\delta\varrho}\right\rangle - \left\langle\partial_q\frac{1}{n_0}\frac{\delta G}{\delta\varrho}\right\rangle\right]\right)$$
$$\frac{\delta G}{\delta P} \to \frac{\delta G}{\delta \pi}$$
$$\frac{\delta G}{\delta \varrho} \to \frac{\delta G}{\delta \varrho}$$

Similar substitutions apply for the functional derivatives of F.

Task: Using the provided Poisson bracket $\{\cdot,\cdot\}_E$ and the effective functional derivative substitutions for F and G, derive the explicit formula for the induced Poisson bracket $\{\cdot,\cdot\}_\Sigma$ on the submanifold Σ.

Ground Truth Solution

$$\begin{aligned}
&\{F,G\}_\Sigma = \int (B-\epsilon\,\Omega)\,\frac{\delta F}{\delta\pi}\cdot\mathbb{J}\frac{\delta G}{\delta\pi}\,n_0\,dq \\
&-\epsilon\int\left(\frac{\delta F}{\delta\pi}\cdot\left[\left\langle\partial_q\frac{\delta G}{\delta\varrho}\right\rangle\right] - \frac{\delta G}{\delta\pi}\cdot\left[\left\langle\partial_q\frac{\delta F}{\delta\varrho}\right\rangle\right]\right)dq \\
&-\epsilon\int\left(\left[\frac{\delta F}{\delta\pi} - \left\langle\partial_\xi\frac{1}{n_0}\frac{\delta F}{\delta\varrho}\right\rangle\right]\cdot\partial_q\cdot\left[\left\langle\partial_\xi\left(\frac{\delta G}{\delta\varrho}\right)\xi\right\rangle\right] - \left[\frac{\delta G}{\delta\pi} - \left\langle\partial_\xi\frac{1}{n_0}\frac{\delta G}{\delta\varrho}\right\rangle\right]\cdot\partial_q\cdot\left[\left\langle\partial_\xi\left(\frac{\delta F}{\delta\varrho}\right)\xi\right\rangle\right]\right)dq \\
&+\epsilon\int\left(\partial_q\frac{1}{n_0}\frac{\delta F}{\delta\varrho}\cdot\partial_\xi\frac{1}{n_0}\frac{\delta G}{\delta\varrho} - \partial_q\frac{1}{n_0}\frac{\delta G}{\delta\varrho}\cdot\partial_\xi\frac{1}{n_0}\frac{\delta F}{\delta\varrho}\right)n_0\,\varrho\,d\xi\,dq + \int(B-\epsilon\,\Omega)\left(\partial_\xi\frac{1}{n_0}\frac{\delta F}{\delta\varrho}\right)\cdot\mathbb{J}\left(\partial_\xi\frac{1}{n_0}\frac{\delta G}{\delta\varrho}\right)n_0\,\varrho\,d\xi\,dq \\
&-\epsilon\int\left(\left\langle\partial_q\frac{1}{n_0}\frac{\delta F}{\delta\varrho}\right\rangle\cdot\left\langle\partial_\xi\frac{1}{n_0}\frac{\delta G}{\delta\varrho}\right\rangle - \left\langle\partial_q\frac{1}{n_0}\frac{\delta G}{\delta\varrho}\right\rangle\cdot\left\langle\partial_\xi\frac{1}{n_0}\frac{\delta F}{\delta\varrho}\right\rangle\right)n_0\,dq - \int(B-\epsilon\,\Omega)\left\langle\partial_\xi\frac{1}{n_0}\frac{\delta F}{\delta\varrho}\right\rangle\cdot\mathbb{J}\left\langle\partial_\xi\frac{1}{n_0}\frac{\delta G}{\delta\varrho}\right\rangle n_0\,dq.
\end{aligned}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: Introduce the short–hand notation

v$_F(q) := F/\pi(q), v_G(q) := G/\pi(q), (divergence–free vector fields)$

$_F := F/n_0,_G := G/n_0 (constants, because the density n_0 is spatially uniform),$

R$_F(q,) := F/\varrho(q,), R_G(q,) := G/\varrho(q,),$

(q) := - 1/n$_0$ $_q\cdot\big(\mathbb{J}\pi\big)(fluid vorticity).$

For every functional F define the scalar field

$_F(q) =_F -(1/n_0)\,\pi(q)\cdot v_F(q) +^{-1}\,\partial_q\cdot\Big[\frac{1}{\varepsilon}\,(B-\varepsilon)\,\mathbb{J}v_F - \frac{1}{n_0}\,\partial_q\cdot<{}_\xi R_F\,\xi> + \frac{1}{n_0}\,\partial_q< R_F > - \frac{1}{n_0}\,<{}_q R_F>\Big], (1)$

the last term being the inverse–Laplacian of a zero–mean function. With this notation the effective replacements dictated by the extension procedure are

F/n $\to {}_F, F/P\text{ß}v_F, F/\varrho\text{ß}R_F, and analogously for G.$

Inserting these expressions into the ambient bracket (keeping n=n$_0 = const. and P = \pi) gives, after the straightforwar$

0,

F,G$_\Sigma(n_0,\pi,\varrho) =$

- $\varepsilon n_0\int_{\mathbb{T}^2}\big(v_F\mathring{u}_{qG} - v_G\mathring{u}_{qF}\big)\,dq$

- $\varepsilon\int_{\mathbb{T}^2}[\,v_F, v_G\,]\mathring{u}\pi\,dq$

+ n$_0\int_{\mathbb{T}^2} B(q)\,v_F\mathring{u}\mathbb{J}v_G\,dq$

$+\,\varepsilon\frac{}{n_0\int_{\mathbb{T}^2}\int_{\mathbb{R}^2}\big({}_q R_F\mathring{u}_\xi R_G - {}_q R_G\mathring{u}_\xi R_F\big)\,\varrho\,d\xi\,dq}$

$+\,1\frac{}{n_0\int_{\mathbb{T}^2}\int_{\mathbb{R}^2}(B-\varepsilon)\,\big({}_\xi R_F\mathring{u}\mathbb{J}_\xi R_G\big)\,\varrho\,d\xi\,dq}$

- $\varepsilon\frac{}{n_0\int_{\mathbb{T}^2}\big(<{}_q R_F>\mathring{u}<{}_\xi R_G> - <{}_q R_G>\mathring{u}<{}_\xi R_F>\big)\,dq}$

- $1\frac{}{n_0\int_{\mathbb{T}^2}(B-\varepsilon)\,<{}_\xi R_F>\mathring{u}\mathbb{J}<{}_\xi R_G>\,dq}$

- $\varepsilon\int_{\mathbb{T}^2}\Big(\big[v_F - \frac{1}{n_0}<{}_\xi R_F>\big]\mathring{u}_q\cdot<{}_\xi R_G\,\xi> - \big[v_G - \frac{1}{n_0}<{}_\xi R_G>\big]\mathring{u}_q\cdot<{}_\xi R_F\,\xi>\Big)\,dq.(2)$

Equation (2) is the Poisson bracket that the quasineutral manifold inherits from the full extended-electron bracket. In (2)

$\langle\,\cdot>(q) =_{\mathbb{R}^2}(\cdot)\,\varrho(q,)\,d\xi,$

[· , ·] is the two–dimensional Lie bracket of vector fields, and $^{-1} istheinverseLaplacianactingonzero–meanfunctionso$
$(n_0, \pi, \varrho), withn_0constantand_q \cdot \pi = 0.$

The formula (2) together with the definition (1) therefore gives the explicit expression of the induced Poisson bracket

boxed{ F,G }$_\Sigma$

needed to describe the Hamiltonian dynamics in the quasineutral limit.

Judge's Evaluation: Evaluation Error: No □ expression found in the model's output.

Problem 30 (Paper: 2506.21424v1)

Problem Statement

Background: In the Nonideal Statistical Field Theory (NISFT), the critical properties of nonideal systems undergoing continuous phase transitions are described by critical exponents. These systems, unlike ideal ones, exhibit defects, inhomogeneities, and impurities, which are accounted for by a nonideal parameter a (where $0 < a < 2$). The theory generalizes the description of ideal systems, which are recovered in the limit $a \to 1$. Two fundamental nonideal critical exponents, η_a and ν_a, are given up to next-to-leading order (NLO) in an ϵ-expansion, where $\epsilon = 4 - d$ and d is the spatial dimension. For a system with N components, these exponents are:

$$\eta_a = \frac{(N+2)}{2a(N+8)^2}\epsilon^2 + \frac{(N+2)}{8a(N+8)^4}(-N^2 + 56N + 272)\epsilon^3$$

$$\nu_a = \frac{1}{2} + \frac{(N+2)}{4a(N+8)}\epsilon + \frac{(N+2)}{8a^2(N+8)^3}[N^2 + (19a+4)N + 92a - 32]\epsilon^2$$

The critical exponent γ_a, which describes the susceptibility scaling behavior, is related to η_a and ν_a through the nonideal scaling relation:

$$\gamma_a = \nu_a(2 - \eta_a)$$

Task: Using the provided expressions for η_a and ν_a, and the given scaling relation, determine the analytical expression for the critical exponent γ_a up to next-to-leading order, i.e., up to $\mathcal{O}(\epsilon^2)$.

Ground Truth Solution

$$\boxed{\gamma_a = 1 + \frac{(N+2)}{2a(N+8)}\epsilon + \frac{(N+2)}{4a^2(N+8)^3}[N^2 + (18a+4)N + 84a - 32]\epsilon^2}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\gamma_a = 1 + \dfrac{N+2}{2a\,(N+8)}\,\epsilon + \dfrac{(N+2)\left[N^2 + (18a+4)N + 84a - 32\right]}{4a^2(N+8)^3}\,\epsilon^2 + \mathcal{O}(\epsilon^3)$

Judge's Evaluation: The model-generated expression matches the ground truth exactly in all terms up to order ϵ^2. The coefficients, powers, and variables are all consistent. The only difference is that the model includes an additional $\mathcal{O}(\epsilon^3)$ term, which is a standard notation indicating higher-order terms and does not affect the correctness of the expression up to ϵ^2.

Since the model expression is mathematically equivalent and includes the same terms as the ground truth, it is completely correct.

Score: 1

Problem 31 (Paper: 2506.21424v1)

Problem Statement

Background: In the Nonideal Statistical Field Theory (NISFT), the critical properties of systems undergoing continuous phase transitions are described, accounting for defects, inhomogeneities, and impurities. This theory generalizes the standard O(N)-symmetric $\lambda\phi^4$ theory by introducing a nonideal distribution parameter a, where $0 < a < 2$. When $a \to 1$, the theory reduces to the ideal case. The system's order parameter has N components. Critical exponents, which characterize the behavior of physical quantities near the critical point, are calculated using renormalization group and ϵ-expansion techniques. The spatial dimension is d, and the expansion parameter is defined as $\epsilon = 4 - d$. One such critical exponent is δ_a, which describes the behavior of the order parameter at the critical point.

Task: Determine the analytical expression for the critical exponent δ_a up to next-to-leading order (NLO) in the ϵ-expansion, as derived within the Nonideal Statistical Field Theory framework.

Ground Truth Solution

$$\boxed{\delta_a = 3 + \epsilon + \frac{1}{2a(N+8)^2}[aN^2 + (16a-2)N + 64a - 4]\epsilon^2 + \frac{1}{4a(N+8)^4}[aN^4 + (29a+1)N^3(330a-54)N^2 + (1760a-3}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: $\delta_a = 3 + \epsilon + \left[\frac{1}{2} - \frac{a\,(N+2)}{(N+8)^2}\right]\epsilon^2 + \mathcal{O}(\epsilon^3)$

Judge's Evaluation: The ground truth expression for δ_a is a detailed expansion up to order ϵ^3, with explicit coefficients involving a, N, and powers of $(N+8)$. It includes terms up to ϵ^3 with complicated polynomial coefficients.

The model-generated expression only includes terms up to order ϵ^2, and the coefficient of ϵ^2 is given as $\frac{1}{2} - \frac{a(N+2)}{(N+8)^2}$, which is structurally different from the ground truth coefficient of ϵ^2, which is:

$$\frac{1}{2a(N+8)^2}\left[aN^2 + (16a-2)N + 64a - 4\right]$$

The model's coefficient does not match this expression, and the model does not include the ϵ^3 term at all, instead replacing it with $\mathcal{O}(\epsilon^3)$.

Therefore, the model's expression is not mathematically equivalent to the ground truth expression. It is missing the ϵ^3 term entirely and the ϵ^2 coefficient is different.

This is more than one minor error: the ϵ^2 coefficient is incorrect, and the ϵ^3 term is missing.

Hence, the score should be 0.

Score: 0

Problem 32 (Paper: 2506.21428v1)

Problem Statement

Background: Consider a quantum system described by a time-dependent Hamiltonian $H(t) = H_0 + e^{\epsilon t} f(t) B$, where H_0 is the unperturbed Hamiltonian, B is an operator representing the perturbation, $f(t)$ is a time-dependent external field, and $\epsilon \to 0^+$ is a positive infinitesimal ensuring the perturbation vanishes in the distant past. The system's density matrix $\rho(t)$ evolves according to the Schrödinger equation. The n-th order contribution to the density matrix, $\rho_n(t)$, satisfies the equation of motion $\dot{\rho_n}(t) = -i[H_0, \rho_n(t)] - ie^{\epsilon t} f(t)[B, \rho_{n-1}(t)]$. The average of an observable A at n-th order in perturbation theory, $\mathrm{tr}(A\rho_n(t))$, is defined in terms of the n-th order non-linear causal response function $\chi^{(n)}_{AB}(t - t_1, \ldots, t - t_n)$ as:

$$\mathrm{tr}(A\rho_n(t)) \equiv e^{n\epsilon t} \int \prod_{j=1}^{n} \mathrm{d}t_j f(t_j) e^{-\epsilon(t-t_j)} \chi^{(n)}_{AB}(t - t_1, \cdots, t - t_n)$$

The causal response function $\chi^{(n)}_{AB}$ vanishes if any of its arguments are negative. Its Fourier transform is denoted by $\chi^{(n)}_{AB}(\omega_1, \ldots, \omega_n)$. When performing Fourier transforms, the notation $\omega_j^+ = \omega_j + i\epsilon$ is used. The response function is symmetric under the exchange of any pair of frequency arguments, i.e., $\chi^{(n)}_{AB}(\ldots, \omega_j, \ldots, \omega_k, \ldots) = \chi^{(n)}_{AB}(\ldots, \omega_k, \ldots, \omega_j, \ldots)$. Assume that the Fourier transform of $f(t)$ is $f(\omega) = \int dt e^{i\omega t} f(t)$.

Task: Derive the recurrence relation that connects the n-th order causal response function $\chi^{(n)}_{AB}(\{\omega_j^+\}_{j=1}^n)$ to the $(n-1)$-th order causal response function.

Ground Truth Solution

$$\boxed{\Big(\sum_{j=1}^{n} \omega_j^+\Big) \chi^{(n)}_{AB}(\{\omega_j^+\}_{j=1}^n) = \chi^{(n)}_{[A,H_0]B}(\{\omega_j^+\}_{j=1}^n) + \frac{\chi^{(n-1)}_{[A,B]B}(\{\omega_j^+\}_{j=2}^n) + (\omega_1^+ \leftrightarrow \omega_{j\neq 1}^+)}{n}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: $\chi^{(n)}_{AB}(\omega_1^+, \ldots, \omega_n^+) = \frac{n}{\omega_1^+ + \cdots + \omega_n^+} \chi^{(n-1)}_{[A,B]\,B}(\omega_1^+, \ldots, \omega_{n-1}^+)$

Judge's Evaluation: The ground truth expression relates the sum of frequencies multiplied by the nth-order susceptibility $\chi^{(n)}_{AB}$ to a combination of a commutator susceptibility $\chi^{(n)}_{[A,H_0]B}$ and a symmetrized term involving $\chi^{(n-1)}_{[A,B]B}$ divided by n. It is an equation expressing a relationship between these susceptibilities with frequency arguments and symmetrization over frequencies.

The model-generated expression, however, expresses $\chi^{(n)}_{AB}$ directly as a fraction involving n divided by the sum of frequencies times $\chi^{(n-1)}_{[A,B]B}$ with one fewer frequency argument. It does not include the $\chi^{(n)}_{[A,H_0]B}$ term, nor the symmetrization indicated by $(\omega_1^+ \leftrightarrow \omega_{j\neq 1}^+)$. The structure and terms are quite different.

Specifically: - The ground truth has a sum over frequencies multiplying $\chi^{(n)}_{AB}$ on the left, while the model has $\chi^{(n)}_{AB}$ isolated on the left. - The ground truth includes a commutator susceptibility term $\chi^{(n)}_{[A,H_0]B}$ which is missing in the model. - The ground truth has a symmetrized term involving $\chi^{(n-1)}_{[A,B]B}$ divided by n, while the model has a factor of n in the numerator and no symmetrization. - The frequency arguments differ in indexing and symmetrization.

These differences are more than minor errors; the model expression is missing a key term and the symmetrization, and the structure is different.

Therefore, the model expression is incorrect relative to the ground truth.

Score: 0

Problem 33 (Paper: 2506.21428v1)

Problem Statement

Background: Consider a quantum system described by a time-independent Hamiltonian H_0, which is perturbed by an external field. The total Hamiltonian is given by $H(t) = H_0 + e^{\epsilon t} f(t) B$, where B is an operator, $f(t)$ is a time-dependent function, and $\epsilon \to 0^+$ ensures the perturbation vanishes in the distant past. The system is initially in thermal equilibrium, with the unperturbed density matrix $\rho_0 = \frac{1}{\mathcal{Z}} \exp(-\beta H_0)$, where $\mathcal{Z}$ is the partition function and $\beta = 1/(k_B T)$.

The n-th order causal response function, $\chi^{(n)}_{AB}(\{\omega_i\}_{i=1}^n)$, describes the system's response of an observable A to the perturbation. In the eigenbasis of H_0, with eigenvalues ϵ_k, the causal response function can be expressed in the Lehmann representation as:

$$\chi^{(n)}_{AB}(\{\omega_i\}_{i=1}^n) = \frac{1}{\mathcal{Z}} \sum_{i_1,\cdots,i_{n+1}} B^{i_1}_{i_2} \cdots B^{i_n}_{i_{n+1}} A^{i_{n+1}}_{i_1} f^{(n)}_{i_1,\cdots,i_{n+1}}(\{\omega_i\}_{i=1}^n)$$

where $A^k_l = \langle k|A|l\rangle$ and $B^k_l = \langle k|B|l\rangle$ are matrix elements, and $f^{(n)}_{i_1,\cdots,i_{n+1}}(\{\omega_i\}_{i=1}^n)$ is a function that captures the frequency dependence. The physical response function $\chi^{(n)}_{AB}$ is symmetric under the permutation of its frequency arguments $\{\omega_i\}_{i=1}^n$. This implies that $f^{(n)}_{i_1,\cdots,i_{n+1}}(\{\omega_i\}_{i=1}^n)$ must also be symmetrized over all $n!$ permutations of its frequency arguments.

The unsymmetrized function $f^{(n)}_{i_1,\cdots,i_{n+1}}(\omega_1,\cdots,\omega_n)$ satisfies the following recurrence relation:

$$f^{(n)}_{i_1,\cdots,i_{n+1}}(\omega_1,\cdots,\omega_n) = \frac{1}{\omega_t^+ + \epsilon_{i_{n+1}} - \epsilon_{i_1}} \left(f^{(n-1)}_{i_2,\cdots,i_{n+1}}(\omega_2,\cdots,\omega_n) - f^{(n-1)}_{i_1,\cdots,i_n}(\omega_1,\cdots,\omega_{n-1}) \right)$$

where $\omega_t^+ = \sum_{k=1}^n \omega_k^+$, $\omega_k^+ = \omega_k + i\epsilon$, and $\epsilon_{ij} = \epsilon_i - \epsilon_j$. The base case for this recurrence relation, for $n = 1$, is given by:

$$f^{(1)}_{i_1,i_2}(\omega) = \frac{e^{-\beta\epsilon_{i_2}} - e^{-\beta\epsilon_{i_1}}}{\omega^+ - \epsilon_{i_1 i_2}}$$

Let $\rho_a = e^{-\beta\epsilon_{i_a}}$ and define $q_a = \epsilon_{i_a} + \sum_{k=0}^{a-1} \omega_k^+$, where ω_0^+ is an auxiliary variable that cancels out from all calculations.

A useful identity for nested commutators is:

$$\underbrace{[\cdots[[A,B],B],\cdots,B]}_{n-\text{times}} = \sum_{m=0}^{n} \binom{n}{m} (-1)^{n-m} B^{n-m} A B^m$$

where $\langle \mathcal{O} \rangle_0 = \mathrm{tr}(\mathcal{O}\rho_0)$.

Task: First, show by induction that the unsymmetrized function $f^{(n)}_{i_1,\cdots,i_{n+1}}(\omega_1,\cdots,\omega_n)$ can be expressed as $\sum_{a=1}^{n+1} \rho_a \prod_{j\neq a} \frac{1}{q_a - q_j}$. Then, using this result and the provided commutator identity, derive the asymptotic expression for the n-th order causal response function in the limit where all frequencies are equal and tend to infinity, i.e., $\lim_{\omega\to\infty} \omega^n \chi^{(n)}_{AB}(\omega,\cdots,\omega)$.

Ground Truth Solution

$$\boxed{\lim_{\omega\to\infty} \omega^n \chi^{(n)}_{AB}(\omega,\cdots,\omega) = \frac{1}{n!} \langle \underbrace{[\cdots[[A,B],B],\cdots,B]}_{n-\text{times}} \rangle_0}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\lim_{\omega\to\infty} \omega^n \chi^{(n)}_{AB}(\omega,\ldots,\omega) = \frac{1}{n!} \left\langle \underbrace{[\,[\,[A,B],B],\ldots,B]}_{n\ \text{times}} \right\rangle_0$

Judge's Evaluation: The model-generated expression matches the ground truth very closely in structure and meaning. Both expressions have the same limit, the same factor of $\frac{1}{n!}$, and the same nested commutator structure. The only difference is in the annotation of the number of times the commutator with B is applied:

- Ground truth: n-times nested commutator is indicated as $n-$ times. - Model: n times.

This is a difference in notation only, not in mathematical content.

However, the ground truth explicitly states the nested commutator as $\underbrace{\cdots[[A,B],B],\cdots,B}_{n-\text{times}}$, which means the commutator with B is applied n times after the initial A. The model writes $\underbrace{[\,[\,[A,B],B],\ldots,B]}_{n\ \text{times}}$, which is the same meaning.

Therefore, the expressions are mathematically equivalent.

Score: 1

Problem 34 (Paper: 2506.21435v1)

Problem Statement

Background: In scalar field cosmology, the dynamics of a spatially flat Friedmann-Lemaître-Robertson-Walker (FLRW) universe with a scalar field $\phi(t)$ and a dust fluid can be analyzed using the Eisenhart-Duval lift. This technique transforms the original constrained Hamiltonian system into a set of null geodesic equations in an extended minisuperspace. For a scalar field potential $V(\phi) = \alpha(e^{\lambda\phi} + \beta)$, where α, β, and λ are constants, the lifted system admits additional conservation laws. These conservation laws, when projected back to the original system, manifest as nonlocal conservation laws.

For the phantom scalar field, corresponding to a scalar field parameter $\varepsilon = -1$, two such nonlocal conservation laws, I_1 and I_2, are given by:

$$I_1 = \frac{1}{3}ap_a + \sqrt{\frac{\rho_{m0}}{2}}t + 2\alpha \int \left(e^{\lambda\phi} + \beta\right) a^3 dt$$

$$I_2 = \frac{\lambda}{2}\left(\frac{a}{3}p_a + \sqrt{\frac{\rho_{m0}}{2}}t\right) - p_\phi + \lambda \int \left(\alpha\left(2e^{\lambda\phi} + \beta\right)\right) a^3 dt$$

where $a(t)$ is the scale factor, p_a and p_ϕ are conjugate momenta, ρ_{m0} is a constant related to the dust density, and t is the cosmic time.

Task: Determine the time-independent conserved quantity $\hat{I}$ by computing the specific linear combination $\hat{I} = I_1 - \frac{2}{\lambda}I_2$.

Ground Truth Solution

$$\boxed{\hat{I} = \frac{2}{\lambda}p_\phi - 2\alpha \int e^{\lambda\phi} a^3 dt}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $=2\frac{}{\lambda\, p_\phi - 2\alpha \int e^{\lambda\phi}\, a^3\, dt}$

Judge's Evaluation: The model-generated expression is exactly the same as the ground truth expression. Both have the same terms, coefficients, variables, and integral structure. The only difference is the presence of some spacing and formatting (e.g., spaces after commas), which does not affect mathematical equivalence.

Score: 1

Problem 35 (Paper: 2506.21435v1)

Problem Statement

Background: In scalar field cosmology, the evolution of the universe can be described by a system of coupled differential equations for the scale factor $a(t)$ and the scalar field $\phi(t)$. For a spatially flat Friedmann–Lemaître–Robertson–Walker (FLRW) geometry, a phantom scalar field (characterized by $\varepsilon = -1$) with a potential function $V(\phi) = \alpha\left(e^{\lambda\phi} + \beta\right)$, and a dust fluid source with density ρ_{m0}, the cosmological field equations can be reduced to a two-equation dynamical system. Let $x(t) = a(t)^{1/3}$ and $\Phi(t) = \dot{\phi}(t)$. The system is given by:

$$\frac{1}{3}\ddot{x} + \lambda\left(\dot{x}\Phi + x\dot{\Phi}\right) - \alpha\beta - \frac{1}{2}\rho_{m0} = 0 \tag{1}$$

$$\frac{1}{3}\left(\ddot{x} + 2\frac{\dot{x}^2}{x}\right) - \lambda\left(\dot{x}\Phi + x\dot{\Phi}\right) + x\left(\alpha\beta + \varepsilon\Phi^2\right) + \frac{3}{2}\rho_{m0} = 0 \tag{2}$$

For the phantom scalar field with $\lambda = 1$ and $\varepsilon = -1$, the solutions near a movable singularity t_0 can be expressed as Right Laurent series expansions in powers of $T = (t - t_0)$: $x(t) = x_0T^{-6} + x_1T^{-5} + x_2T^{-4} + x_3T^{-3} + x_4T^{-2} + x_5T^{-1} + x_6T^0 + x_7T^1 + x_8T^2 + \ldots$ $\Phi(t) = \Phi_0T^{-1} + \Phi_1T^0 + \Phi_2T^1 + \Phi_3T^2 + \Phi_4T^3 + \Phi_5T^4 + \Phi_6T^5 + \Phi_7T^6 + \Phi_8T^7 + \ldots$ Here, x_0 and t_0 are integration constants, and $\Phi_0 = 2$. By substituting these series into the differential equations and collecting terms by powers of T, the coefficients x_j and Φ_j can be determined. It is known that $x_1 = 0$, $x_3 = 0$, $x_5 = 0$, $\Phi_1 = 0$, $\Phi_3 = 0$, and $\Phi_5 = 0$. The coefficient x_7 is a third independent integration constant. The other coefficients are: $x_2 = -\frac{\alpha\beta x_0}{5}$ $x_4 = \frac{17x_0}{600}(\alpha\beta)^2$ $x_6 = -\frac{1049}{189000}x_0(\alpha\beta)^3$ $\Phi_2 = -\frac{\alpha\beta}{15}$ $\Phi_4 = \frac{7}{450}(\alpha\beta)^2$ $\Phi_6 = -\frac{289}{2365}(\alpha\beta)^3$ $\Phi_7 = \frac{7x_7}{x_0}$

Task: Determine the coefficient Φ_8 in terms of x_0, α, β, and ρ_{m0}.

Ground Truth Solution

$$\boxed{\Phi_8 = -\frac{2699(\alpha\beta)^4x_0 - 140000\rho_{m0}}{315000x_0}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: $\Phi_8 = \dfrac{7\,\rho_{m0}}{6\,x_0} - \dfrac{17}{720}\,(\alpha\beta)^3$

Judge's Evaluation: The ground truth expression is:

$$\Phi_8 = -\frac{2699(\alpha\beta)^4x_0 - 140000\rho_{m0}}{315000x_0}$$

The model-generated expression is:

$$\Phi_8 = \frac{7\rho_{m0}}{6x_0} - \frac{17}{720}(\alpha\beta)^3$$

Comparison: - The ground truth has a single fraction with a numerator involving $(\alpha\beta)^4x_0$ and ρ_{m0}, and denominator $315000x_0$. - The model expression is a sum of two terms, one proportional to ρ_{m0}/x_0 and the other proportional to $(\alpha\beta)^3$. - The powers of $(\alpha\beta)$ differ: ground truth has power 4, model has power 3. - The coefficients and denominators differ significantly. - The model expression is not a single fraction and does not match the structure or coefficients of the ground truth. - The model expression is not mathematically equivalent to the ground truth.

Given these multiple discrepancies (different powers, different coefficients, different structure), the model expression is incorrect.

Score: 0

Problem 36 (Paper: 2506.21447v1)

Problem Statement

Background: The double-scaled SYK (DSSYK) model with matter chords can be described by a chord Hilbert space. When specific constraints are imposed on this Hilbert space, the system can describe End-Of-The-World (ETW) branes in the bulk. These constraints lead to a Hamiltonian of the Al-Salam Chihara (ASC) form. Consider a system described by the Hamiltonian $_{\text{ASC}}$, which acts on a Hilbert space where $\hat{\ell}$ is a length-like operator and $\hat{P}$ is its conjugate momentum. The operators satisfy the canonical commutation relation $[\hat{\ell}, \hat{P}] = \lambda$. The Hamiltonian is given by:

$$_{\text{ASC}} = \frac{J}{\sqrt{\lambda(1-q)}}(^{-\hat{P}} + (X+Y)^{-\hat{\ell}} + (1 - XY^{-\hat{\ell}})^{\hat{P}}(1 -^{-\hat{\ell}}))$$

Here, J is a coupling constant, λ and q are fixed parameters of the theory, and X, Y are constant parameters. To study spread complexity, it is useful to express the Hamiltonian in a Hermitian form. This can be achieved by a transformation of the canonical momentum operators. The resulting Hermitian Hamiltonian $^{(\text{Her})}_{\text{ASC}}$ is given by:

$$^{(\text{Her})}_{\text{ASC}} = \frac{J}{\sqrt{\lambda(1-q)}}\left(^{\hat{P}}\sqrt{(1 - XYq^{\hat{n}-1})(1-q)} + \sqrt{(1 - XYq^{\hat{n}-1})(1-q)}^{-\hat{P}} + (X+Y)q^{\hat{n}}\right)$$

where $q^{\hat{n}} =^{-\hat{\ell}}$. The action of the momentum operator on the orthonormal basis states K_n (which are related to the chord number operator $\hat{n}$ by $\hat{n}K_n = nK_n$) is given by:

$$^{-\hat{P}}K_n = \sqrt{[n+1]_q}K_{n+1}$$

$$^{\hat{P}}K_n = \sqrt{[n]_q}K_{n-1}$$

where $[n]_q = \frac{1-q^n}{1-q}$ is the q-number. The Lanczos algorithm defines coefficients a_n and b_n for a Hamiltonian and an initial state K_0 (which is the state with zero chord number, $\hat{n}K_0 = 0$) via the recurrence relations:

$$K_n = a_n K_n + b_n K_{n-1} + b_{n+1}K_{n+1}$$

where $b_0 = 0$.

Task: Derive the expressions for the Lanczos coefficients a_n and b_n for the Hamiltonian $^{(\text{Her})}_{\text{ASC}}$ in terms of J, λ, q, X, Y, and n.

Ground Truth Solution

$$\boxed{a_n = \frac{J}{\sqrt{\lambda(1-q)}}(X+Y)q^n\ , \quad b_n = \frac{J}{\sqrt{\lambda}}\sqrt{(1 - XYq^{n-1})[n]_q}\ .}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution:

$$a_n = \frac{J\,(X+Y)\,q^{\,n}}{\sqrt{\lambda(1-q)}},$$

$$b_n = \frac{J}{\sqrt{\lambda}}\sqrt{\frac{1-q^{\,n}}{1-q}\,\big(1 - XY\,q^{\,n-1}\big)}, \qquad n \geq 1, \qquad b_0 = 0.$$

Judge's Evaluation: The ground truth expression for b_n is:

$$b_n = \frac{J}{\sqrt{\lambda}}\sqrt{(1 - XYq^{n-1})[n]_q}$$

where $[n]_q = \frac{1-q^n}{1-q}$.

The model-generated expression for b_n is:

$$b_n = \frac{J}{\sqrt{\lambda}} \sqrt{\frac{1-q^n}{1-q}(1 - XYq^{n-1})}$$

which is exactly the same as the ground truth, just written explicitly with the definition of $[n]_q$.

For a_n, the ground truth is:

$$a_n = \frac{J}{\sqrt{\lambda(1-q)}}(X+Y)q^n$$

and the model-generated expression is:

$$a_n = \frac{J(X+Y)q^n}{\sqrt{\lambda(1-q)}}$$

which is exactly the same, just rearranged.

The model also adds the conditions $n \geq 1$ and $b_0 = 0$, which are consistent with typical definitions of b_n in such contexts and do not contradict the ground truth.

Therefore, the model-generated expressions are mathematically equivalent to the ground truth expressions, with no errors or omissions.

Score: 1

Problem 37 (Paper: 2506.21447v1)

Problem Statement

Background: The doubled double-scaled SYK (DSSYK) model, relevant for de Sitter holography, can be understood as two decoupled DSSYK systems. Consider one such system, described by the Hamiltonian $= \frac{J}{\sqrt{\lambda(1-q)}}(^{-\hat{P}}+^{\hat{P}}(1-^{-\hat{\ell}}))$. The parameters J and λ are coupling constants of the theory, and $q \in [0,1]$ is a fixed parameter. The operators $^{-\hat{P}}$ and $^{\hat{P}}$ are related to creation and annihilation operators $\hat{a}^\dagger$ and $\hat{a}$ respectively, such that $\hat{a}^\dagger = \frac{1}{\sqrt{1-q}}\,^{-\hat{P}}$ and $\hat{a} = \sqrt{1-q}\,^{\hat{P}}$. The variable $\hat{\ell}$ is related to the chord number operator $\hat{n}$ by $q^{\hat{n}} =^{-\hat{\ell}}$. These operators satisfy the following commutation relations: $[\hat{n}, \hat{a}^\dagger] = \hat{a}^\dagger$, $[\hat{n}, \hat{a}] = -\hat{a}$, and the q-commutator $[\hat{a},\ \hat{a}^\dagger]_q = 1$, where $[\hat{A}, \hat{B}]_q = \hat{A}\hat{B} - q\hat{B}\hat{A}$. The Lanczos algorithm defines a Krylov basis K_n for a given Hamiltonian and initial state K_0 (where $\hat{n}K_0 = 0$) via the recurrence relations: $K_n = a_n K_n + b_n K_{n-1} + b_{n+1} K_{n+1}$, where a_n and b_n are the Lanczos coefficients.

Task: Derive the Lanczos coefficients a_n and b_n for the given Hamiltonian.

Ground Truth Solution

$$\boxed{a_n = 0\ , \quad b_n = \frac{J}{\sqrt{\lambda}}\sqrt{[n]_q}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution:
$$a_n = 0 \qquad (n = 0, 1, 2, \dots),$$
$$b_n = \frac{J}{\sqrt{\lambda\,(1-q)}}\,\sqrt{\frac{1-q^{\,2n}}{1-q^{\,2}}} \qquad (n = 1, 2, 3, \dots)$$

Judge's Evaluation: The ground truth expression is:

$$a_n = 0, \quad b_n = \frac{J}{\sqrt{\lambda}}\sqrt{[n]_q}$$

where $[n]_q$ is presumably the standard q-number defined as $[n]_q = \frac{1-q^n}{1-q}$.

The model-generated expression is:

$$a_n = 0 \quad (n = 0, 1, 2, \dots), \quad b_n = \frac{J}{\sqrt{\lambda(1-q)}}\sqrt{\frac{1-q^{2n}}{1-q^2}} \quad (n = 1, 2, 3, \dots)$$

Key points of comparison:
1. The ground truth uses $\sqrt{[n]_q}$ with the standard q-number $[n]_q = \frac{1-q^n}{1-q}$.
2. The model uses a different q-number-like expression:

$$\sqrt{\frac{1-q^{2n}}{1-q^2}}$$

and also has an extra factor of $\frac{1}{\sqrt{1-q}}$ in the denominator outside the square root.

3. The model's expression is not equivalent to the ground truth expression. The model's b_n can be rewritten as:

$$b_n = \frac{J}{\sqrt{\lambda}} \cdot \frac{1}{\sqrt{1-q}} \cdot \sqrt{\frac{1-q^{2n}}{1-q^2}} = \frac{J}{\sqrt{\lambda}}\sqrt{\frac{1-q^{2n}}{(1-q)(1-q^2)}}$$

which is not the same as $\frac{J}{\sqrt{\lambda}}\sqrt{\frac{1-q^n}{1-q}}$.

4. The model's expression uses $2n$ and q^2 in the q-number, which is a different q-number definition (a "q$^2 - number$"$ratherthana$"$q - number$").$Thisisasignificantdifference, notaminortypo.$

5. The domain of n is slightly different: ground truth does not specify, model specifies $n = 1, 2, 3, \ldots$ for b_n, but this is a minor detail.

Conclusion:

- The model's expression is not mathematically equivalent to the ground truth. - The difference is more than a minor coefficient or sign; it is a different q-number definition. - Therefore, the model's expression is incorrect.

Score: 0

—

Score: 0

Problem 38 (Paper: 2506.21457v1)

Problem Statement

Background: Consider a one-dimensional quantum particle with Hamiltonian h_x modeling two delta-interactions centered at points $-x/2$ and $x/2$. The operator h_x acts on functions $u(y)$ and is defined by $h_x u(y) = -u''(y)$ for almost every $y \in \backslash\{\pm x/2\}$. The domain of h_x is given by $D(h_x) := \{u \in H^2(\backslash\{\pm x/2\}) \cap H^1() : [u'](\pm x/2) = \alpha u(\pm x/2)\}$, where $[u'](y)$ denotes the jump of the derivative u' across y. The parameter α represents the strength of the interaction. The eigenvalues $z = -\lambda < 0$ of h_x are determined by the condition $\det(1 + \alpha_x(z)) = 0$, where $_x(z)$ is a 2×2 matrix given by

$$_x(z) = \frac{i}{2\sqrt{z}} \begin{bmatrix} 1 & e^{i\sqrt{z}\,|x|} \\ e^{i\sqrt{z}\,|x|} & 1 \end{bmatrix}, \quad \Im\sqrt{z} > 0.$$

This determinant condition is equivalent to the transcendental equation:

$$\left(\alpha + 2\sqrt{\lambda}\right)^2 = \alpha^2 e^{-2\sqrt{\lambda}\,|x|}.$$

Assume that $\alpha < 0$. For this case, the equation admits one or two solutions for $\lambda > 0$. Let $-\lambda_0(x)$ denote the lowest (most negative) eigenvalue. It is known that $\alpha^2/4 < \lambda_0(x) \le \alpha^2$. This implies that for $\lambda_0(x)$, the term $\alpha + 2\sqrt{\lambda_0(x)}$ is positive. Therefore, taking the positive square root of both sides of the transcendental equation yields:

$$\alpha + 2\sqrt{\lambda} = |\alpha| e^{-\sqrt{\lambda}|x|}.$$

The Lambert W-function, $W(z)$, is defined as the solution to the equation $W(z)e^{W(z)} = z$.

Task: Derive the explicit expression for $\lambda_0(x)$ in terms of the Lambert W-function.

Ground Truth Solution

$$\boxed{\lambda_0(x) = \left(\frac{W\left(\frac{|\alpha||x|}{2}\, e^{-\frac{|\alpha||x|}{2}}\right)}{|x|} + \frac{|\alpha|}{2}\right)^2}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\lambda_0(x) = \Big(\dfrac{|\alpha|}{2} + \dfrac{1}{|x|}\, W\Big(\dfrac{|\alpha|\,|x|}{2}\, e^{-|\alpha|\,|x|/2}\Big)\Big)^2$

Judge's Evaluation: The model-generated expression is:

$$\lambda_0(x) = \left(\frac{|\alpha|}{2} + \frac{1}{|x|} W\left(\frac{|\alpha||x|}{2} e^{-\frac{|\alpha||x|}{2}}\right)\right)^2$$

The ground truth expression is:

$$\lambda_0(x) = \left(\frac{W\left(\frac{|\alpha||x|}{2} e^{-\frac{|\alpha||x|}{2}}\right)}{|x|} + \frac{|\alpha|}{2}\right)^2$$

Comparing the two:

- The order of the two terms inside the parentheses is swapped, but addition is commutative, so this is equivalent. - The arguments inside the Lambert W function are identical. - The coefficients and denominators match exactly. - The entire expression is squared in both cases.

Therefore, the model-generated expression is mathematically equivalent to the ground truth expression.

Score: 1

Problem 39 (Paper: 2506.21457v1)

Problem Statement

Background: Consider a quantum system in one dimension consisting of a light particle interacting with two heavy particles. The dynamics of the heavy particles, after applying the Born-Oppenheimer approximation and shifting the energy by α^2, is described by an effective self-adjoint Hamiltonian acting on the Hilbert space $L^2()$. Here, $\ll 1$ is a small parameter proportional to the square root of the mass ratio, and $\alpha < 0$ is a fixed interaction parameter. The Hilbert space $L^2()$ is defined as $\{f \in L^2() : f(x) = f(-x)\}$, where $= +$ for bosonic heavy particles and $= -$ for fermionic heavy particles.

The effective Hamiltonian is given by:

$$= -^2 \frac{d^2}{dx^2} + V(x) +^2 R(x)$$

where $V(x)$ is a potential term derived from the lowest eigenvalue of the light particle's Hamiltonian, defined as $V(x) := -\lambda_0(x) + \alpha^2$. The function $\lambda_0(x)$ is given by:

$$\lambda_0(x) = \left(\frac{W\left(\frac{|\alpha||x|}{2} e^{-\frac{|\alpha||x|}{2}}\right)}{|x|} + \frac{|\alpha|}{2} \right)^2$$

Here, $W(z)$ is the Lambert W-function, defined as the solution to $W(z)e^{W(z)} = z$. For $|x| \ll 1$, the potential $V(x)$ has the asymptotic behavior $V(x) = |\alpha|^3|x| + O(x^2)$. The term $R(x)$ is given by $R(x) := \int_{|} \partial_x \psi^{BO}(x, y)|^2 dy$, where $\psi^{BO}(x, y)$ is the normalized eigenfunction corresponding to $-\lambda_0(x)$.

To analyze the eigenvalues of for small , we introduce the scaling parameter $\Lambda =^{-1}$. We define a scaled operator $\widetilde{}_\Lambda \Lambda^2{}_{1/\Lambda}$. For $\Lambda \gg 1$, the eigenvalues of $\widetilde{}_\Lambda$ are primarily determined by the dominant potential term $\Lambda^2 V(x)$, with $R(x)$ contributing to higher-order corrections.

Consider an auxiliary self-adjoint operator K^1 acting on $L^2()$, defined as:

$$\mathrm{K}^1 = -\frac{d^2}{dx^2} + |\alpha|^3|x|$$

The eigenvalues of K^1 are denoted by e_n, where $e_n = |\sigma_{2n}|\alpha^2$ for bosonic heavy particles and $e_n = |\sigma_{2n+1}|\alpha^2$ for fermionic heavy particles. The negative numbers σ_k are defined such that $'(\sigma_{2n}) = 0$ (extrema of the Airy function) and $(\sigma_{2n+1}) = 0$ (zeros of the Airy function), with $\cdots < \sigma_{2n+1} < \sigma_{2n} < \cdots < \sigma_1 < \sigma_0 < 0$.

Let U_Λ be a unitary operator in $L^2()$ defined by $(U_\Lambda f)(x) := \Lambda^{1/3} f(\Lambda^{2/3}x)$.

Task: Derive the asymptotic expansion for the n-th eigenvalue ${}_{,n}$ of for $\ll 1$.

Ground Truth Solution

$$\boxed{{}_{,n} = s_n \alpha^{22/3} + O()}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: ${}_{\varepsilon,n} = \varepsilon^{2/3} e_n + \mathcal{O}(\varepsilon^{4/3}) = \varepsilon^{2/3} |\alpha|^2 |\sigma_{2n+\delta}| + \mathcal{O}(\varepsilon^{4/3})$

Judge's Evaluation: The ground truth expression is:

$${}_{,n} = s_n \alpha^{22/3} + O()$$

The model-generated expression is:

$${}_{\varepsilon,n} = \varepsilon^{2/3} e_n + \mathcal{O}(\varepsilon^{4/3}) = \varepsilon^{2/3} |\alpha|^2 |\sigma_{2n+\delta}| + \mathcal{O}(\varepsilon^{4/3})$$

Comparison:

1. The leading order term in both expressions is proportional to $\varepsilon^{2/3}$. 2. The ground truth uses $s_n \alpha^2$ as the coefficient, while the model uses $e_n = |\alpha|^2 |\sigma_{2n+\delta}|$. 3. The model's coefficient is more explicit,

involving $|\alpha|^2$ and $|\sigma_{2n+\delta}|$, which presumably corresponds to s_n. 4. The ground truth error term is $O(\varepsilon)$, while the model's error term is $\mathcal{O}(\varepsilon^{4/3})$. Since $\varepsilon^{4/3} = \varepsilon^{1+1/3}$ is smaller order than ε, the model's error term is actually tighter (smaller) than the ground truth. 5. The model uses absolute values $|\alpha|^2$ instead of α^2. If α is real and positive, this is equivalent; if α is complex, the ground truth might be simplified or assume α^2 is real positive. This is a minor difference but could be significant depending on context. 6. The model introduces $|\sigma_{2n+\delta}|$ instead of s_n. This could be a difference in notation or a more explicit form. Without further context, this is acceptable as a variable name difference.

Overall, the model expression is mathematically equivalent in form, with a more explicit coefficient and a tighter error term. The difference in the error term order is not a mistake but an improvement. The use of absolute values is a minor difference in notation or assumption.

Therefore, the model expression matches the ground truth expression exactly in mathematical content, with only minor differences in notation and a better error term.

Score: 1

—

Score: 1

Problem 40 (Paper: 2506.21466v1)

Problem Statement

Background:

Consider the two-dimensional unit flat torus $\mathbb{T}^2 := (\mathbb{R}/\mathbb{Z})^2$. Let $S'(\mathbb{T}^2)$ denote the space of Schwartz distributions on $\mathbb{T}^2$. For $s \in \mathbb{R}$, $H^s(\mathbb{T}^2)$ denotes the usual fractional Sobolev space, and $\mathcal{C}^s(\mathbb{T}^2)$ denotes the Besov-Hölder space. Let $L^2(\mathbb{T}^2)$ be the usual Lebesgue space, with inner product $\langle \cdot, \cdot \rangle_{L^2(\mathbb{T}^2)}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a standard Wiener space, where the canonical coordinate process $X^0 = (X_t^0)_{t \geq 0}$ is an $L^2(\mathbb{T}^2)$-Brownian motion. For $T \geq 0$, define the mollifier ρ_T by its Fourier transform $\hat{\rho}_T(n) = \rho(\langle n \rangle / T)$ for $n \in \mathbb{Z}^2$, where $\rho \in C^\infty(\mathbb{R}; [0,1])$ is compactly supported on $B(0,2)$ and $\rho \equiv 1$ on $B(0,1)$, and $\langle n \rangle = \sqrt{4\pi^2 |n|^2 + m^2}$ for a fixed mass $m^2 > 0$. Define the Fourier multiplier J_t^0 by its symbol $\widehat{J_t^0}(n) = \frac{1}{\langle n \rangle} \sqrt{\partial_t \hat{\rho}_t^2(n)}$. The stochastic process $(W_T^0)_{T \geq 0}$ is defined by $W_T^0 := \int_0^T J_t^0 dX_t^0$. The law of $W_\infty^0 := \lim_{T \to \infty} W_T^0$ is denoted by μ^0, which is a centered Gaussian measure on $S'(\mathbb{T}^2)$ with covariance $(-\Delta_{\mathbb{T}^2} + m^2)^{-1/2}$.

Let $M_0 = [-1,1] \times \mathbb{T}^2$. For $\varphi^0 \in S'(\mathbb{T}^2)$, the m-harmonic extension $\overline{H}\varphi^0$ is the smooth function on M_0 that solves $(-\Delta + m^2)\overline{H}\varphi^0 = 0$ on $(-1,1) \times \mathbb{T}^2$ with boundary condition $\overline{H}\varphi^0|_{\{0\} \times \mathbb{T}^2} = \varphi^0$ and $\overline{H}\varphi^0|_{\{\pm 1\} \times \mathbb{T}^2} = 0$. For any $\varphi \in S'(M_0)$, $\varphi_T := \varphi * \rho_T$ denotes convolution in the periodic direction. The regularized covariance of the harmonic extension is $C_T^B(x,y) := \mathbb{E}_{\mu^0}[\overline{H}W_T^0(x)\overline{H}W_T^0(y)]$ for $x, y \in M_0$. The Wick powers of $\overline{H}W_T^0$ are defined as $\overline{H}W_T^0(x) = \overline{H}W_T^0(x)$, $(\overline{H}W_T^0)^2(x) = (\overline{H}W_T^0)^2(x) - C_T^B(x,x)$, $(\overline{H}W_T^0)^3(x) = (\overline{H}W_T^0)^3(x) - 3C_T^B(x,x)\overline{H}W_T^0(x)$, and $(\overline{H}W_T^0)^4(x) = (\overline{H}W_T^0)^4(x) - 6C_T^B(x,x)(\overline{H}W_T^0)^2(x) + 3C_T^B(x,x)^2$.

The boundary energy renormalization is $\delta_T^0 := -3 \cdot 4 \int_{M_0} \int_{M_0} C_T^B(x,y)^4 dx dy$. The regularized boundary potential $V_T^0(\varphi^0)$ is defined for $\varphi^0 \in S'(\mathbb{T}^2)$ by $V_T^0(\varphi^0) = \int_{M_0} (\overline{H}\varphi_T^0)^4 dx - \delta_T^0$. The unnormalized Laplace transform of $f \in C^\infty(\mathbb{T}^2)$ under the regularized interacting boundary measure is $\mathcal{Z}_T^0(f) := \mathbb{E}_{\mu^0}[\exp(\langle f, \varphi^0 \rangle_{L^2(\mathbb{T}^2)} - V_T^0(\varphi^0))]$.

Let $\mathbb{H}^0$ be the space of progressively measurable processes that are $\mathbb{P}$-almost surely in $L_t^2 L_z^2$. For $u \in \mathbb{H}^0$, define $Z_T^0(u) := \int_0^T J_t^0 u_t dt$. The initial variational representation of $-\log \mathcal{Z}_T^0(f)$ is given by:

$$-\log \mathcal{Z}_T^0(f) = \inf_{u \in \mathbb{H}^0} \mathbb{E}\left[\langle f, W_T^0 + Z_T^0(u) \rangle_{L^2(\mathbb{T}^2)} + \mathcal{V}_T^0(W_T^0, Z_T^0(u)) + \frac{1}{2}\|u\|_{L_t^2 L_z^2}^2\right]$$

where $\mathcal{V}_T^0(W_T^0, Z_T^0(u)) = \int_{M_0} 4(\overline{H}W_T^0)^3 \overline{H}Z_T^0(u) + 6(\overline{H}W_T^0)^2(\overline{H}Z_T^0(u))^2 + 4\overline{H}W_T^0(\overline{H}Z_T^0(u))^3 + (\overline{H}Z_T^0(u))^4 \, dx - \delta_T^0$.

Let $\Xi_T^0(X^0) := ((\overline{H}W_t^0)_{t \leq T}, ((\overline{H}W_t^0)^2)_{t \leq T}, ((\overline{H}W_t^0)^3)_{t \leq T})$ be a vector of stochastic processes, where $\Xi_T^{0,1}(X^0) = (\overline{H}W_t^0)_{t \leq T}$, $\Xi_T^{0,2}(X^0) = ((\overline{H}W_t^0)^2)_{t \leq T}$, and $\Xi_T^{0,3}(X^0) = ((\overline{H}W_t^0)^3)_{t \leq T}$. Let $\mathfrak{B}^0 := C_t L_\tau^4 \mathcal{C}_z^{-1/2-\kappa_0'} \times C_t L_\tau^8 \mathcal{C}_z^{-1/2-\kappa_0'} \times C_t L_\tau^1 \mathcal{C}_z^{-1/2-\kappa_0'}$ for some small $\kappa_0' > 0$. Let $\mathcal{L}^0 := L_t^2 W_z^{1/2-\kappa_0', 2+\kappa_0''}$ for some small $\kappa_0'' > 0$. Let $\mathbb{L}^0$ be the space of progressively measurable processes that are almost surely in $\mathcal{L}^0$. Define the boundary remainder map $\ell_T^0 : \mathfrak{B}^0 \times \mathcal{L}^0 \to \mathcal{L}^0$ by $\ell_T^0(\Xi^0, u)_t := 4J_t^0 \overline{H}^* \Xi_t^{0,3} 1_{t \leq T} + u_t$.

Task: Derive the expression for the renormalized boundary cost function $\mathbb{F}_T^{0,f}(u)$ such that $-\log \mathcal{Z}_T^0(f) = \inf_{u \in \mathbb{L}^0} \mathbb{F}_T^{0,f}(u)$.

Ground Truth Solution

$$\boxed{\mathbb{F}_T^{0,f}(u) = \mathbb{E}\left[\langle f, Z_T^0(u) \rangle_{L_z^2} + \int_{[-1,1] \times \mathbb{T}^2} 6\Xi_T^{0,2}(X^0)(\overline{H}Z_T^0(u))^2 + 4\Xi_T^{0,1}(X^0)(\overline{H}Z_T^0(u))^3 dx + \|\overline{H}Z_T^0(u)\|_{L^4([-1,1] \times \mathbb{T}^2)}^4 + \frac{1}{2}\|\ell_T^0}$$

Model Outputs

Model: o3 (Score: 0.5)

Model Solution: $\mathbb{F}_T^{0,f}(u) = \mathbb{E}\Big[\big\langle f,\, Z_T^0(u) \big\rangle_{L^2(\mathbb{T}^2)} + \int_{M_0} \Big(6\, \Xi_T^{0,2}(X^0)(x)\, [\overline{H}Z_T^0(u)(x)]^2 + 4\, \Xi_T^{0,1}(X^0)(x)\, [\overline{H}Z_T^0(u)(x)]^3 + [\overline{H}Z_T^0(u)(x)]^4 \Big)\, dx + \frac{1}{2}\|u\|_{L_t^2 L_z^2}^2 \Big]$

Judge's Evaluation: Let's compare the two expressions term-by-term:

1. **Inner product term:** - Ground truth: $\langle f, Z_T^0(u)\rangle_{L_z^2}$ - Model: $\langle f, Z_T^0(u)\rangle_{L^2(\mathbb{T}^2)}$

The ground truth specifies the inner product over L_z^2, while the model uses $L^2(\mathbb{T}^2)$. Since the domain in the integral is $[-1,1]\times\mathbb{T}^2$, the variable z presumably corresponds to $\mathbb{T}^2$. This is a minor difference in notation but not a substantive error.

2. **Integral term:** - Ground truth integral is over $[-1,1]\times\mathbb{T}^2$. - Model integral is over M_0, which is presumably the same domain (though not explicitly stated). - The integrand terms match exactly: - $6\Xi_T^{0,2}(X^0)(\overline{H}Z_T^0(u))^2$ - $4\Xi_T^{0,1}(X^0)(\overline{H}Z_T^0(u))^3$ - $(\overline{H}Z_T^0(u))^4$

The model writes the last term as $[\overline{H}Z_T^0(u)(x)]^4$ inside the integral, while the ground truth writes the L^4 norm of $\overline{H}Z_T^0(u)$ over the domain separately. This is a significant difference: the ground truth separates the L^4 norm term outside the integral, while the model includes the fourth power inside the integral.

However, the L^4 norm to the fourth power is exactly the integral of the fourth power over the domain, so these are mathematically equivalent.

3. **Norm term:** - Ground truth: $\frac{1}{2}\|\ell_T^0(\Xi_T^0(X^0),u)\|^2_{L_t^2L_z^2}$ - Model: $\frac{1}{2}\|u\|^2_{L_t^2L_z^2}$

This is a significant difference. The ground truth has the norm of $\ell_T^0(\Xi_T^0(X^0),u)$, which is a function of $\Xi_T^0(X^0)$ and u, while the model simply has the norm of u itself. This is not a minor difference; it changes the meaning of the term.

Summary:

- The inner product domain difference is minor and acceptable. - The integral domain notation is minor. - The placement of the fourth power term is mathematically equivalent. - The last norm term is significantly different: the model omits the operator $\ell_T^0(\Xi_T^0(X^0),\cdot)$ and replaces it with just u.

Because of this last difference, the model expression is not fully correct.

Score: 0.5

The model expression is very close but misses the key operator $\ell_T^0(\Xi_T^0(X^0),\cdot)$ in the last norm term, which is a nontrivial omission.

—

Score: 0.5

Problem 41 (Paper: 2506.21466v1)

Problem Statement

Background: Let $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$ denote the two-dimensional unit flat torus, and let $m^2 > 0$ be a fixed mass parameter. Let $S'(\mathbb{T}^2)$ be the space of Schwartz distributions on $\mathbb{T}^2$. Consider a probability measure ν^0 on $S'(\mathbb{T}^2)$. Define the Hilbert space $\mathcal{H} = L^2(\nu^0)$.

For any $\tau > 0$, let $\mathcal{A}_\tau : S'(\mathbb{T}^2) \times S'(\mathbb{T}^2) \to \mathbb{R}$ be a measurable map, referred to as the φ^4_3 amplitude. This amplitude satisfies the following properties:

1. Positivity: $\mathcal{A}_\tau(\varphi, \varphi') > 0$ for $\nu^0 \otimes \nu^0$-almost surely (φ, φ').
2. Gluing Property: For any $\tau_1, \tau_2 > 0$ and $\varphi_1, \varphi_3 \in S'(\mathbb{T}^2)$,
$$\mathcal{A}_{\tau_1+\tau_2}(\varphi_1, \varphi_3) = \mathbb{E}_{\nu^0}[\mathcal{A}_{\tau_1}(\varphi_1, \varphi_2)\mathcal{A}_{\tau_2}(\varphi_2, \varphi_3)]$$
where φ_2 is a random variable sampled according to ν^0.
3. Square Integrability: For any $\tau > 0$, the amplitude is square-integrable with respect to $\nu^0 \otimes \nu^0$, i.e.,
$$\int_{S'(\mathbb{T}^2)} \int_{S'(\mathbb{T}^2)} \mathcal{A}_\tau(\varphi, \varphi')^2 \nu^0(\varphi)\nu^0(\varphi') < \infty$$
Furthermore, there exists a constant $C > 0$ such that for any $\tau > 0$,
$$\int_{S'(\mathbb{T}^2)} \int_{S'(\mathbb{T}^2)} \mathcal{A}_\tau(\varphi, \varphi')^2 \nu^0(\varphi)\nu^0(\varphi') \leq e^{C\tau}$$

Define a family of linear operators $(\mathbf{P}_\tau)_{\tau \geq 0}$ on $\mathcal{H}$ as follows:

- For $\tau > 0$, $\mathbf{P}_\tau$ acts on $f \in \mathcal{H}$ by
$$\mathbf{P}_\tau f(\varphi) = \int_{S'(\mathbb{T}^2)} f(\varphi')\mathcal{A}_\tau(\varphi, \varphi')\nu^0(\varphi')$$
- For $\tau = 0$, $\mathbf{P}_0 = \mathrm{Id}_{\mathcal{H}}$ (the identity operator on $\mathcal{H}$).

It is known that $(\mathbf{P}_\tau)_{\tau \geq 0}$ forms a strongly continuous semigroup on $\mathcal{H}$. Let $\mathbf{H}$ be the infinitesimal generator of this semigroup, so $\mathbf{P}_\tau = e^{-\tau \mathbf{H}}$. The spectrum of $\mathbf{H}$, denoted $\sigma(\mathbf{H})$, is known to be bounded from below, i.e., there exists $E_0 \in \mathbb{R}$ such that $\sigma(\mathbf{H}) \subset [E_0, \infty)$. For any $\alpha \in \mathbb{C} \setminus [E_0, \infty)$, the resolvent operator is defined as $\mathbf{R}_\alpha = (\mathbf{H} - \alpha)^{-1}$.

Task: Derive the property of the spectrum of $\mathbf{H}$ that implies it consists solely of eigenvalues with finite multiplicity.

Ground Truth Solution

The spectrum of $\mathbf{H}$ is discrete and consists of eigenvalues of finite multiplicity with associated eigenvectors that form a

Model Outputs

Model: o3 (Score: 0.5)

Model Solution: H has compact resolvent (H-α)$^{-1}$ is compact; $\implies$ $\sigma(\mathbf{H}) = \{E_n\}_{n\in\mathbb{N}}$, $E_n \uparrow \infty$, each E_n an eigenvalue of finite multiplicity

Judge's Evaluation: The ground truth states that the spectrum of $\mathbf{H}$ is discrete, consists of eigenvalues of finite multiplicity, and the associated eigenvectors form an orthonormal basis of $\mathcal{H}$.

The model-generated expression states that $\mathbf{H}$ has compact resolvent, which implies the spectrum $\sigma(\mathbf{H}) = \{E_n\}_{n\in\mathbb{N}}$ is discrete, with eigenvalues E_n tending to infinity, each of finite multiplicity.

The model expression correctly captures the discreteness of the spectrum, the eigenvalues of finite multiplicity, and the fact that the eigenvalues tend to infinity. However, it does not explicitly mention that the eigenvectors form an orthonormal basis of $\mathcal{H}$, which is a crucial part of the spectral theorem for compact resolvent operators on Hilbert spaces.

Since the model expression is missing this important part about the eigenvectors forming an orthonormal basis, it is not fully correct. However, it is very close and captures the main spectral properties.

Therefore, the model expression deserves partial credit.

Score: 0.5

Problem 42 (Paper: 2506.21496v1)

Problem Statement

Background: The octonion algebra $\mathbb{O}$ is a real 8-dimensional nonassociative algebra with a unit element e_0 and seven imaginary basis elements e_i for $i = 1, \dots, 7$. The product of any two basis elements $e_I, e_J \in \{e_0, \dots, e_7\}$ is given by $e_I e_J = f_{IJ}^K e_K$, where f_{IJ}^K are the structure constants. The octonions are an alternative algebra, meaning the associator $[x, y, z] = (xy)z - x(yz)$ changes sign under the exchange of any two elements.

Consider a finite-dimensional, unital, alternative algebra $A = \bigoplus_{a=1}^n \mathbb{O}$, where each A_a is a copy of the octonion algebra $\mathbb{O}$. We denote the basis elements of A as $e^{(a)I} := (0, \dots, e^I, \dots, 0)$, where e^I is the I-th basis element of the a-th octonion factor.

Let $M = A \otimes A$ be a real vector space. M is equipped with left and right A-actions, denoted by $\cdot$, which are inherited from the octonion product. For $c_{(k)} \in A_k$ and $u \otimes^{(lm)} v \in A_l \otimes A_m$ (where $u \otimes^{(lm)} v$ denotes an element of the $A_l \otimes A_m$ component of M), these actions are defined as: $c_{(k)} \cdot (u \otimes^{(lm)} v) = \hat{\delta}_k^l (cu \otimes^{(lm)} v)$ $(u \otimes^{(lm)} v) \cdot c_{(k)} = \hat{\delta}_k^m (u \otimes^{(lm)} vc)$ where $\hat{\delta}_k^l$ is the Kronecker delta, and cu and vc denote the octonion product within the respective factors. This type of bimodule is called a split alternative bimodule.

A linear map $\Delta : A \to M$ is called a derivation if it satisfies the Leibniz rule: $\Delta[xy] = \Delta[x] \cdot y + x \cdot \Delta[y]$ for all $x, y \in A$. We are interested in derivations where the input elements x, y are imaginary octonion basis elements. A general linear operator $\Delta : A \to M$ can be expressed in terms of the basis elements as: $\Delta[e^{(a)i}] = \sum_{b,c=1}^n \sum_{J,K=0}^7 \Delta_{(bc)JK}^{(a)i} e^J \otimes^{bc} e^K$ where $\Delta_{(bc)JK}^{(a)i}$ are real coefficients, $i \in \{1, \dots, 7\}$ (imaginary input index), and $J, K \in \{0, \dots, 7\}$ (output indices, including the identity e_0).

Applying the Leibniz rule to $\Delta[e^{(a)i} e^{(b)j}]$ (where $i, j \in \{1, \dots, 7\}$ are imaginary indices) and equating the coefficients of $e^J \otimes^{cd} e^K$ on both sides yields the following condition on the coefficients $\Delta_{(bc)JK}^{(a)i}$: $\hat{\delta}^{ab} f_{ij}^L \Delta_{(cd)JK}^{(a)L} = \Delta_{(cd)JS}^{(a)i} \hat{\delta}_b^d f_{SK}^K + \Delta_{(cd)SJ}^{(b)j} \hat{\delta}_a^c f_{iS}^K$ where f_{ij}^L are the structure constants of the octonions, and Einstein's summation convention is used for repeated indices L, S (but not for bracketed indices like (ab) or (cd)). This equation must hold for all combinations of indices and factors.

Task: 1. Show that if $c \neq a = b \neq d$, the Leibniz rule implies that the coefficients $\Delta_{(cd)JK}^{(a)L}$ must be zero for all L, J, K. This means that $\Delta[e^{(a)i}]$ can only have non-zero components in $A_a \otimes A_c$ or $A_c \otimes A_a$ for some c. 2. Given that for the case where $a = b = c = d$, the non-zero coefficients are $\Delta_{(aa)i0}^{(a)i} = \kappa_{(aa)}$ and $\Delta_{(aa)0i}^{(a)i} = -\kappa_{(aa)}$, where $\kappa_{(aa)}$ is an arbitrary real constant. 3. Given that for the cases where $c = a \neq b = d$, $a = b = c \neq d$, and $a = b = d \neq c$, the non-zero coefficients are $\Delta_{(ab)i0}^{(c)i} = \kappa_{(ab)} \hat{\delta}^{ca}$ and $\Delta_{(ba)0i}^{(c)i} = -\kappa_{(ba)} \hat{\delta}^{bc}$, where $\kappa_{(ab)}$ and $\kappa_{(ba)}$ are arbitrary real constants. 4. Combine these results to express the general form of $\Delta[e^{(a)i}]$ for any $a \in \{1, \dots, n\}$ and imaginary index $i \in \{1, \dots, 7\}$.

Ground Truth Solution

$$\boxed{\Delta_\kappa[e^{(a)i}] = \sum_c (\kappa_{(ac)} e^i \otimes^{ac} e^0 - \kappa_{(ca)} e^0 \otimes^{ca} e^i)}$$

Model Outputs

Model: o3 (Score: 0.5)

Model Solution: $\Delta\left[e^{(a)i}\right] = \sum_{b=1}^n \Big(\kappa_{(ab)}\, e^{(a)i} \otimes^{(ab)} e_0 - \kappa_{(ba)}\, e_0 \otimes^{(ba)} e^{(a)i}\Big), \qquad i = 1, \dots, 7$

Judge's Evaluation: The ground truth expression is:

$$\Delta_\kappa[e^{(a)i}] = \sum_c \left(\kappa_{(ac)} e^i \otimes^{ac} e^0 - \kappa_{(ca)} e^0 \otimes^{ca} e^i\right)$$

The model-generated expression is:

$$\Delta\left[e^{(a)i}\right] = \sum_{b=1}^n \left(\kappa_{(ab)} e^{(a)i} \otimes^{(ab)} e_0 - \kappa_{(ba)} e_0 \otimes^{(ba)} e^{(a)i}\right), \quad i = 1, \dots, 7$$

Comparison:

1. **Summation index:** Ground truth sums over c, model sums over b. This is just a dummy index difference, acceptable.

2. **Coefficients:** Both have $\kappa_{(ac)}$ and $\kappa_{(ca)}$ (ground truth) vs $\kappa_{(ab)}$ and $\kappa_{(ba)}$ (model). This matches the dummy index change.

3. **Tensor factors:** - Ground truth has $e^i \otimes^{ac} e^0$ and $e^0 \otimes^{ca} e^i$. - Model has $e^{(a)i} \otimes^{(ab)} e_0$ and $e_0 \otimes^{(ba)} e^{(a)i}$.

Here, the model uses $e^{(a)i}$ instead of e^i, and e_0 instead of e^0.

4. **Index notation:** - The ground truth uses e^i and e^0 without the (a) superscript on e^i. - The model uses $e^{(a)i}$ consistently.

This is a difference in notation, but if $e^{(a)i}$ corresponds to e^i in the ground truth, this is acceptable.

5. **Subscript vs superscript on e_0 vs e^0:** - Ground truth uses e^0. - Model uses e_0.

This is a significant difference: e^0 and e_0 are generally not the same object (covariant vs contravariant components). This is a nontrivial difference.

6. **Tensor product superscripts:** - Ground truth uses $\otimes^{ac}$ and $\otimes^{ca}$. - Model uses $\otimes^{(ab)}$ and $\otimes^{(ba)}$.

This is consistent with the dummy index change.

7. **Extra index range $i = 1, \ldots, 7$ in model:** - Ground truth does not specify this, but this is likely context-dependent and not an error.

Summary:

- The main discrepancy is the use of e_0 instead of e^0. - Also, the model uses $e^{(a)i}$ instead of e^i, but this may be just notation. - The difference between e_0 and e^0 is more than a minor notation difference; it changes the nature of the vector.

Given that, the model expression has one significant error (covariant vs contravariant index on e^0).

No other errors are apparent.

According to the rubric:

- One minor error (one wrong sign, missing variable, or minor constant) gets 0.5. - More than one minor error or a significant error gets 0.

Is the difference between e_0 and e^0 a minor error? It is more than a minor error because it changes the object.

Therefore, the model expression is not completely correct, but it is very close.

I would assign 0.5 for partial correctness.

Score: 0.5

Problem 43 (Paper: 2506.21496v1)

Problem Statement

Background: Consider an internal spectral geometry defined by a triple $T_F = (A_F, H_F, D_F)$. The internal coordinate algebra is $A_F = \mathbb{O} \oplus \mathbb{O}$, where $\mathbb{O}$ denotes the octonion algebra. The Hilbert space is $H_F = A_F$, equipped with a natural inner product derived from the octonionic norm. The left and right actions of A_F on H_F are given by $\pi_L(a) = \begin{pmatrix} L_{a_1} & 0 \\ 0 & L_{a_2} \end{pmatrix}$ and $\pi_R(a) = \begin{pmatrix} R_{a_1} & 0 \\ 0 & R_{a_2} \end{pmatrix}$ for $a = (a_1, a_2) \in A_F$, where $L_x y = xy$ and $R_x y = yx$ for $x, y \in \mathbb{O}$. These actions satisfy the alternative identities:

$$\begin{aligned} \pi_L(ab) &= \pi_L(a)\pi_L(b) + [\pi_L(a), \pi_R(b)] \\ \pi_R(ab) &= \pi_R(b)\pi_R(a) + [\pi_R(b), \pi_L(a)] \\ [\pi_L(a), \pi_R(b)] &= [\pi_R(a), \pi_L(b)] \end{aligned}$$

for all $a, b \in A_F$. The octonion algebra is involutive, with $x^* = x_0 e_0 - \sum_{i=1}^7 x_i e_i$ for $x = x_0 e_0 + \sum_{i=1}^7 x_i e_i \in \mathbb{O}$. This involution extends to H_F via the operator $J_F = \begin{pmatrix} j & 0 \\ 0 & j \end{pmatrix}$, where $j(x) = x^*$. The right action $\pi_R(a)$ can be expressed in terms of the left action and J_F as $\pi_R(a) = J_F \pi_L(a)^* J_F$, where $\pi_L(a)^*$ denotes the adjoint (transpose) of the operator $\pi_L(a)$ on H_F.

The Dirac operator $D_F : H_F \to H_F$ is a Hermitian operator of the general form $D_F = M_{IJ} \begin{pmatrix} 0 & e_I \otimes e_J^* \\ e_J \otimes e_I^* & 0 \end{pmatrix}$, where e_I are octonion basis elements ($I = 0, ..., 7$) and $*$ denotes the dual with respect to the Hilbert space. The Dirac operator is required to satisfy the constraint $D_F J_F = \epsilon'_F J_F D_F$ for some $\epsilon'_F = \pm 1$. The external derivation $\Delta_D : A_F \to End(H_F)$ is defined as $\Delta_D(a) = [D_F, \pi_L(a)]$. This derivation must satisfy the Leibniz rule:

$$\Delta_D(ab) = \Delta_D(a) \cdot b + a \cdot \Delta_D(b)$$

for some left and right bimodule actions $\cdot : A_F \otimes \Omega^1_D A_F \to \Omega^1_D A_F$ and $\cdot : \Omega^1_D A_F \otimes A_F \to \Omega^1_D A_F$, where $\Omega^1_D A_F$ is the space of 1-forms generated by elements of the form $\Delta_D(a)$. Using the alternative identities, the expression for $\Delta_D(ab)$ can be expanded as:

$$\Delta_D(ab) = \Delta_D(a)\pi_L(b) + \pi_L(a)\Delta_D(b) + [D_F, (1-S)[\pi_L(a), \pi_R(b)] + S[\pi_R(a), \pi_L(b)]]$$

for some parameter $S \in \mathbb{R}$. The space of 1-forms $\Omega^1_D A_F$ is also equipped with an involution $\omega^* = \omega^T$ (transpose of the operator $\omega \in End(H_F)$). For the representation to be involutive, the left and right actions must be related by $\omega \cdot a = (a^* \cdot \omega^*)^*$.

Task: Derive the explicit expressions for the left and right bimodule actions, $a \cdot \omega$ and $\omega \cdot a$, for $\omega \in \Omega^1_D A_F$ and $a \in A_F$, such that the Leibniz rule for Δ_D is satisfied and the representation is involutive.

Ground Truth Solution

$$\boxed{\begin{aligned} a \cdot \omega &= \pi_L(a)\omega - \frac{\epsilon'_F}{2}[\pi_L(a), J_F \omega^* J_F] + \frac{1}{2}[J_F \pi_L(a)^* J_F, \omega] \\ \omega \cdot a &= \omega \pi_L(a) + \frac{1}{2}[\omega, J_F \pi_L(a)^* J_F] - \frac{\epsilon'_F}{2}[J_F \omega^* J_F, \pi_L(a)] \end{aligned}}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: Write every elementary 1–form as

x ,y := L (x) [D F ,L (y)], x ,y A F ,

and extend linearly so that $_D^1 A_F = Span x, y. For A_F define$

Left action a x ,y := L (a x) [D F ,L (y)] +(1-S) [D F ,[L (a),R (y)]], (1)

Right action x ,y a := L (x) [D F ,L (y)] L (a)+S [D F ,[R (y),L (a)]], (2)

where S is the same parameter that appears in $_D(ab) =_D (a)_L(b) +_L(a)_D(b) + [D_F, (1-S)[_L(a),_R(b)] + S[_R(a),_L(b)]]$.

Properties

1. Leibniz rule. For every a,bA$_F$

a$_D(b) +_D (a)b =_L (a)_D(b) + (1 - S)[D_F, [_L(a),_R (b)]] +_D (a)_L(b) + S[D_F, [_R(a),_L (b)]]$
$=_D(ab)$,
because the alternative identities give $[_L(a),_R (b)] = [_R(a),_L (b)]$.
2. Involution. With the involution $^*(operatortranspose)andusing_R(a) = J_{F\,L}(a)^* J_F onefindsfrom(1)that$
a=(a$^{**})^*$,
so the bimodule representation is involutive.
Thus (1) and (2) give the required explicit left and right A$_F$–$actionson^1_D A_F$.

Judge's Evaluation: Evaluation Error: No $\square$ expression found in the model's output.

Problem 44 (Paper: 2506.21498v1)

Problem Statement

Background: Consider a two-player, two-strategy symmetric game with strategies **C** (cooperation) and **D** (defection). The payoffs for Player 1 are given by the matrix:

	C	**D**
C	1	$-B$
D	$1-A$	0

where A and B are real-valued parameters. Player 1 chooses a row, and Player 2 chooses a column. The game is symmetric, meaning Player 2's payoffs are obtained by transposing this matrix.

Players update their strategies using introspection dynamics. At each discrete time step, one player is chosen uniformly at random (Player 1 or Player 2, each with probability 1/2) to revise their strategy. The chosen player compares their current payoff π with the payoff π' they would obtain by switching to the other strategy, assuming the co-player's strategy remains fixed. The player switches to the alternative strategy with a probability given by the Fermi function:

$$\phi(\Delta\pi) = \frac{1}{1 + \mathrm{e}^{-\beta_i \Delta\pi}}$$

where $\Delta\pi = \pi' - \pi$ is the payoff difference, and $\beta_i \geq 0$ is the selection strength of player i. Player 1 has selection strength β_1, and Player 2 has selection strength β_2.

The system can be described as a Markov chain with four states representing the joint strategies: **CC**, **CD**, **DC**, and **DD** (where the first letter denotes Player 1's strategy and the second denotes Player 2's strategy). This Markov chain has a unique stationary distribution $\mathbf{u}(A, B, \beta_1, \beta_2) = (u_{\mathbf{CC}}, u_{\mathbf{CD}}, u_{\mathbf{DC}}, u_{\mathbf{DD}})$, which represents the average frequency of each game outcome over many rounds.

Task: Derive the stationary distribution $\mathbf{u}(A, B, \beta_1, \beta_2)$ of the Markov chain in terms of A, B, β_1, and β_2.

Ground Truth Solution

$$\boxed{\mathbf{u}(A, B, \beta_1, \beta_2) = \frac{1}{e^{\beta_1 A} e^{\beta_2 A} + e^{\beta_1 B} + e^{\beta_2 B} + 1} \left(e^{\beta_1 A} e^{\beta_2 A}, e^{\beta_1 B}, e^{\beta_2 B}, 1\right)}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution:

$$u_{\mathbf{CC}} = \frac{1}{Z},$$

$$u_{\mathbf{CD}} = \frac{p_2 + K r_1}{(2 - r_1 - p_2)\, Z},$$

$$u_{\mathbf{DC}} = \frac{p_1 + K r_2}{(2 - p_1 - r_2)\, Z},$$

$$u_{\mathbf{DD}} = \frac{K}{Z},$$

Judge's Evaluation: The ground truth expression is a single vector-valued function $\mathbf{u}(A, B, \beta_1, \beta_2)$ explicitly given as a fraction with a denominator and a 4-component vector numerator involving exponentials of $\beta_1 A$, $\beta_2 A$, $\beta_1 B$, and $\beta_2 B$.

The model-generated expression, however, is a set of four scalar expressions $u_{\mathbf{CC}}, u_{\mathbf{CD}}, u_{\mathbf{DC}}, u_{\mathbf{DD}}$ defined in terms of variables p_1, p_2, r_1, r_2, K, Z. There is no direct correspondence or mapping shown between these variables and the parameters A, B, β_1, β_2 or the exponential terms in the ground truth. The model expression does not resemble the ground truth expression in form or content.

Since the model expression is completely different in structure and variables, and does not match or approximate the ground truth expression, it cannot be considered correct or partially correct.

Score: 0

Problem 45 (Paper: 2506.21498v1)

Problem Statement

Background: Consider a two-player, two-strategy symmetric game where players can choose between strategy **C** (cooperation) and **D** (defection). The payoffs for player 1 are given by the following matrix, where player 1 chooses a row and player 2 chooses a column:

	C	**D**
C	1	$-B$
D	$1-A$	0

Players adapt their strategies over time using introspection dynamics. In this process, a player compares their current payoff π with the payoff π' they would obtain by switching to the alternative strategy (while the co-player's strategy remains fixed). The probability of switching to the alternative strategy is given by the Fermi function:

$$\phi = \frac{1}{1 + \mathrm{e}^{-\beta_i\,(\pi' - \pi)}}.$$

Here, $\beta_i \geq 0$ represents player i's selection strength. This learning process is iterated over many time steps, leading to a stationary distribution of the four possible game outcomes: **CC**, **CD**, **DC**, and **DD** (where the first letter denotes player 1's strategy and the second denotes player 2's). Let $u_{\mathbf{CC}}$, $u_{\mathbf{CD}}$, $u_{\mathbf{DC}}$, and $u_{\mathbf{DD}}$ be the stationary frequencies of these outcomes.

When both players have the same selection strength, $\beta_1 = \beta_2 = \beta$, the stationary distribution simplifies to:

$$\mathbf{u}(A, B, \beta, \beta) = \frac{1}{2 + e^{A\beta} + e^{B\beta}}(e^{A\beta}, 1, 1, e^{B\beta}).$$

Player 1's average payoff, denoted $\Pi_{\beta_2}(\beta_1)$, is computed by multiplying the stationary frequencies with the respective stage game payoffs:

$$\Pi_{\beta_2}(\beta_1) = 1 \cdot u_{\mathbf{CC}} - B \cdot u_{\mathbf{CD}} + (1 - A) \cdot u_{\mathbf{DC}} + 0 \cdot u_{\mathbf{DD}}.$$

In the context of adaptive dynamics, the selection strength β is an evolving trait. We consider a monomorphic resident population where all individuals have a selection strength x. The average payoff of a resident player in this population is denoted $\Pi_x(x)$. This quantity is a crucial component in determining the fitness of mutant traits.

Task: Derive the analytical expression for player 1's average payoff $\Pi_x(x)$ in a monomorphic population where both players have the same selection strength x.

Ground Truth Solution

$$\boxed{\Pi_x(x) = \frac{e^{Ax} - A - B + 1}{2 + e^{Ax} + e^{Bx}}}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\Pi_x(x) = \dfrac{e^{Ax} + 1 - A - B}{2 + e^{Ax} + e^{Bx}}$

Judge's Evaluation: The ground truth numerator is $e^{Ax} - A - B + 1$, while the model's numerator is $e^{Ax} + 1 - A - B$.

Since addition is commutative, $e^{Ax} + 1 - A - B$ is mathematically equivalent to $e^{Ax} - A - B + 1$.

The denominator is exactly the same in both expressions.

Therefore, the model-generated expression is mathematically equivalent to the ground truth expression.

Score: 1

Problem 46 (Paper: 2506.21529v1)

Problem Statement

Background: The behavior of a Dirac electron in graphene under an external magnetic field is described by the Dirac Hamiltonian $H = v_F\alpha\cdot(\mathbf{p}+\mathbf{A})$, where v_F is the Fermi velocity, $\alpha = (\sigma_x, \sigma_y)$ are Pauli matrices, $\mathbf{p} = (p_x, p_y)$ is the momentum operator, and $\mathbf{A} = (A_x, A_y)$ is the vector potential. For a Dirac wavefunction $\Psi = \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}$, the equation $H\Psi = E\Psi$ leads to a second-order differential equation for the component $\psi_2(\mathbf{r})$. Defining $\mathcal{P} = p_x + ip_y$ and $\mathcal{A}(\mathbf{r}) = A_x(\mathbf{r}) + iA_y(\mathbf{r})$, this equation is $v_F^2\big[(\mathcal{P}+\mathcal{A}(\mathbf{r}))(\mathcal{P}^*+\mathcal{A}^*(\mathbf{r}))\big]\psi_2(\mathbf{r}) = E^2\psi_2(\mathbf{r})$. Explicitly, this can be written as: $\Bigg[p_x^2+p_y^2+A_x^2+A_y^2+2(A_xp_x+A_yp_y)+(p_xA_x)+(p_yA_y)+i(p_yA_x)-i(p_xA_y)\Bigg]\psi_2(x,y) = \left(\frac{E^2}{v_F^2}\right)\psi_2(x,y)$. Consider a magnetic field in the symmetric gauge, where the vector potential components are given by $A_x(x,y) = -Byf(r)$ and $A_y(x,y) = Bxf(r)$, with $r = \sqrt{x^2+y^2}$ being the radial variable and B a constant. The magnetic field $B_z(r)$ is defined as $B_z(r) = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}$. When transforming to polar coordinates (r,θ) and using the ansatz $\psi_2(r,\theta) = r^{-1/2}e^{im_l\theta}\chi(r)$, the equation for ψ_2 reduces to a one-dimensional Schrödinger-like equation for $\chi(r)$: $-\frac{d^2\chi(r)}{dr^2} + V_{m_l}(r)\chi(r) = \left(\frac{E^2}{v_F^2}\right)\chi(r)$, where the effective potential is $V_{m_l}(r) = B^2r^2f(r)^2 + \frac{m_l^2-\frac{1}{4}}{r^2} + 2B\left(f(r)m_l - f(r) - \frac{rf'(r)}{2}\right)$. For $m_l \le 0$, this equation can be factorized using the operator $\mathcal{A}_{m_l} = \frac{d}{dr} + Brf(r) - \frac{|m_l|+\frac{1}{2}}{r}$. The ground-state eigenfunction $\chi_{0,m_l}(r)$ satisfies $\mathcal{A}_{m_l}\chi_{0,m_l}(r) = 0$, which yields $\chi_{0,m_l}(r) = N_{m_l}e^{-\int^r Br'f(r')dr'}r^{|m_l|+\frac{1}{2}}$, where N_{m_l} is a normalization factor. The Abraham-Moses construction states that if $V(x)$ is a potential with ground-state wavefunction $\psi_0(x)$, then a new potential $\tilde{V}_\lambda(x) = V(x) - 2\frac{d^2}{dx^2}\ln[J(x)+\lambda]$ is strictly isospectral to $V(x)$, where $J(x) = \int_{-\infty}^x \psi_0(x')^2dx'$ and λ is a real constant. In the context of the effective potential $V_{m_l}(r)$ and its ground-state eigenfunction $\chi_{0,m_l}(r)$, this transformation implies a new function $f_{m_l,\lambda}(r)$ for the vector potential, related to the original $f(r)$ by: $Bf_{m_l,\lambda}(r) = Bf(r) + \frac{1}{r}\frac{d}{dr}\ln[J_{m_l}(r)+\lambda]$, where $J_{m_l}(r) = \int_0^r \chi_{0,m_l}(r')^2dr'$.

Task: Derive the expression for the non-uniform magnetic field $(B_z)_{m_l,\gamma}(r)$ that supports the Landau levels, given that the original magnetic field is uniform (i.e., $f(r) = 1$). Express your result in terms of B, r, m_l, and a real constant $\gamma = \lambda/N_{m_l}^2$.

Ground Truth Solution

$$\boxed{(B_z)_{m_l,\gamma}(r) = 2B + \frac{r^{2|m_l|}e^{-Br^2}}{\gamma+\int_0^r r'^{2|m_l|+1}e^{-Br'^2}dr'} + \frac{d}{dr}\left(\frac{r^{2|m_l|+1}e^{-Br^2}}{\gamma+\int_0^r r'^{2|m_l|+1}e^{-Br'^2}dr'}\right)}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: Introduce the abbreviation

$\mathrm{I}_{|m_l|}(r) = \int_0^r r'^{\,2|m_l|+1}\,e^{-Br'^2}\,dr'$.(1)

For the uniform field (f(r)=1) the Abraham–Moses prescription gives

$\mathrm{f}_{m_l,\lambda}(r) = 1 + \frac{1}{B\,r}\frac{d}{dr}\ln\big[I_{|m_l|}(r)+\gamma\big], \qquad \gamma = \frac{\lambda}{N_{m_l}^2}$.(2)

With $\mathrm{A}_x^{(\lambda)} = -B\,y\,f_{m_l,\lambda}(r), \qquad A_y^{(\lambda)} = B\,x\,f_{m_l,\lambda}(r)$, (3)

the corresponding magnetic field is

$(\mathrm{B}_z)_{m_l,\gamma}(r) = \partial_xA_y^{(\lambda)} - \partial_yA_x^{(\lambda)} = 2B\,f_{m_l,\lambda}(r) + B\,r\,\frac{df_{m_l,\lambda}}{dr}$.(4)

Using (2) one finds after a short calculation

$(\mathrm{B}_z)_{m_l,\gamma}(r) = 2B + \frac{2\,r^{2|m_l|}\,e^{-Br^2}\big[|m_l|+1-Br^2\big]}{I_{|m_l|}(r)+\gamma} - \frac{r^{4|m_l|+2}\,e^{-2Br^2}}{\big[I_{|m_l|}(r)+\gamma\big]^2}$.(5)

Expression (5) is the non-uniform magnetic field that still supports the full set of graphene Landau levels. In terms of the lower incomplete gamma–function (s,x)=$_0^x t^{s-1}e^{-t}dt, theintegral(1)canbewrittenas$

$\mathrm{I}_{|m_l|}(r) = \frac{1}{2}\,B^{-|m_l|-1}\,\gamma\big(|m_l|+1,\,Br^2\big)$, (6)

so that the field (5) may also be expressed completely with elementary factors and (—m$_l| + 1, Br$š).

Judge's Evaluation: Evaluation Error: No $\square$ expression found in the model's output.

Problem 47 (Paper: 2506.21529v1)

Problem Statement

Background: A Dirac electron in graphene is described by the two-dimensional Dirac Hamiltonian $H = v_F\alpha \cdot (\mathbf{p} + \mathbf{A})$, where v_F is the Fermi velocity, $\alpha = (\sigma_x, \sigma_y)$ are Pauli matrices, $\mathbf{p} = (p_x, p_y)$ is the momentum operator, and $\mathbf{A} = (A_x, A_y)$ is the vector potential. The electronic charge is taken as -1. Consider the asymmetric (Landau) gauge, where $A_x = 0$ and $A_y = A_y(x)$. The Dirac wavefunction is taken in the form $\Psi = e^{iky}\begin{pmatrix}\psi_1(x)\\ i\psi_2(x)\end{pmatrix}$, where k is a continuous momentum quantum number. The Dirac equation $H\Psi = E\Psi$ leads to a Schrödinger-like equation for $\psi_2(x)$ with an effective potential $V_k(x)$ given by: $V_k(x) = [k + A_y(x)]^2 - \frac{dA_y(x)}{dx}$. For a uniform magnetic field $B_z = B$, the vector potential is $A_y(x) = Bx$. In this case, the ground-state eigenfunction $\psi_{2,0,k}(x)$ (corresponding to $n = 0$) is given by $\psi_{2,0,k}(x) = N_k e^{-\frac{B(x+k/B)^2}{2}}$, where N_k is a normalization factor. The Abraham-Moses construction provides a family of strictly isospectral potentials $\tilde{V}_\lambda(x)$ related to an initial potential $V(x)$ by the formula: $\tilde{V}_\lambda(x) = V(x) - 2\frac{d^2}{dx^2}\ln[J(x) + \lambda]$, where $J(x) = \int_{-\infty}^{x} \psi_0(x')^2 dx'$ and $\psi_0(x)$ is the ground-state eigenfunction of the initial potential $V(x)$. Here, λ is a real constant.

Task: Derive the expression for the strictly isospectral effective potential $\tilde{V}_{k,\lambda}(x)$ for the Dirac electron in the asymmetric gauge, using the uniform magnetic field $B_z = B$ as the initial configuration. Assume the normalization factor N_k is absorbed into λ, so $J_k(x) = \int_{-\infty}^{x} e^{-B(x'+k/B)^2} dx'$.

Ground Truth Solution

$$\boxed{\tilde{V}_{k,\lambda}(x) = (k + Bx)^2 - B - 2\frac{d^2}{dx^2}\ln\left[\int_{-\infty}^{x} e^{-B(x'+k/B)^2} dx' + \lambda\right]}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution:

$$\begin{aligned}\tilde{V}_{k,\lambda}(x) &= V_k(x) - 2\,\frac{d^2}{dx^2}\ln\big[J_k(x) + \lambda\big]\\ &= (k + Bx)^2 - B + \frac{4B\,(x + k/B)\,e^{-B(x+k/B)^2}}{J_k(x) + \lambda} + \frac{2\,e^{-2B(x+k/B)^2}}{\big[J_k(x) + \lambda\big]^2},\end{aligned}$$

$$J_k(x) = \int_{-\infty}^{x} e^{-B(x'+k/B)^2}\,dx' = \frac{\sqrt{\pi}}{2\sqrt{B}}\Big[1 + \operatorname{erf}\big(\sqrt{B}(x + k/B)\big)\Big].$$

Judge's Evaluation: The ground truth expression is:

$$\tilde{V}_{k,\lambda}(x) = (k + Bx)^2 - B - 2\frac{d^2}{dx^2}\ln\left[\int_{-\infty}^{x} e^{-B(x'+k/B)^2} dx' + \lambda\right]$$

The model-generated expression is:

$$\tilde{V}_{k,\lambda}(x) = (k + Bx)^2 - B + \frac{4B(x + k/B)e^{-B(x+k/B)^2}}{J_k(x) + \lambda} + \frac{2e^{-2B(x+k/B)^2}}{(J_k(x) + \lambda)^2}$$

where

$$J_k(x) = \int_{-\infty}^{x} e^{-B(x'+k/B)^2} dx' = \frac{\sqrt{\pi}}{2\sqrt{B}}\left[1 + \operatorname{erf}(\sqrt{B}(x + k/B))\right].$$

—

Comparison and analysis:

1. The model starts from the ground truth expression and rewrites the second derivative of the logarithm explicitly using the known formula for the second derivative of a log function:

$$-2\frac{d^2}{dx^2}\ln[f(x)] = -2\left(\frac{f''}{f} - \left(\frac{f'}{f}\right)^2\right) = 2\left(\left(\frac{f'}{f}\right)^2 - \frac{f''}{f}\right).$$

2. The model's expression matches the ground truth in the first two terms: $(k + Bx)^2 - B$.

3. The integral inside the logarithm is correctly identified as $J_k(x)$.

4. The model expands the second derivative term explicitly, giving the additional terms involving $e^{-B(x+k/B)^2}$ and its powers divided by $J_k(x) + \lambda$.

5. The model's expression is mathematically equivalent to the ground truth expression, just written in a more explicit form.

6. The model also provides the closed form of the integral $J_k(x)$ in terms of the error function, which is consistent and correct.

7. There are no sign errors, missing terms, or incorrect coefficients.

—

Conclusion:

The model-generated expression is mathematically equivalent to the ground truth expression, just expanded explicitly. It matches perfectly in all aspects.

Score: 1

Problem 48 (Paper: 2506.21540v1)

Problem Statement

Background: Consider a one-dimensional quantum spin chain described by a continuum field theory in terms of a unit vector $\vec{n}(x,t) = (\sqrt{1-\psi(x,t)^2}\cos\theta(x,t), \sqrt{1-\psi(x,t)^2}\sin\theta(x,t), \psi(x,t))$, where x is the spatial coordinate and t is imaginary time. Here, $\theta(x,t)$ is the azimuthal angle in the xy-plane, and $\psi(x,t)$ is the z-component of the spin, n_z. The action for the isotropic Heisenberg ferromagnet with spin magnitude S and exchange coupling $\mathcal{J}$ is given by $\mathcal{S} = \int xt\,\mathcal{L}$, where the Lagrangian density is

$$\mathcal{L} = -iS(1-\psi)\partial_t\theta - \frac{\mathcal{J}S^2}{2}\left(\frac{(\nabla\psi)^2}{1-\psi^2} + (1-\psi^2)(\nabla\theta)^2\right)$$

The system is subject to U(1) spin rotation symmetry around the S^z axis, spatial translation and reflection symmetries, and a discrete symmetry that reverses the sign of S^z (e.g., time-reversal symmetry). Generic perturbations respecting these symmetries are added to the Hamiltonian.

Task: Derive the effective Lagrangian density $\mathcal{L}'$ for the critical theory by performing the following steps: (1) Expand the given Lagrangian density to quadratic order in the fields $\psi(x,t)$ and $\theta(x,t)$ and their derivatives, neglecting total derivative terms. (2) Identify the lowest-order potential terms in $\psi(x,t)$ that are allowed by the specified symmetries. (3) Combine these terms to form the full Lagrangian density. (4) Rescale the spatial coordinate x by $x_0 = \mathcal{J}^{1/2}S$ and the imaginary time t by $t_0 = S$, and absorb all resulting constant prefactors into the redefined coupling constants λ_2 and λ_4.

Ground Truth Solution

$$\boxed{\mathcal{L}' = i\,\psi\partial_t\theta + 12(\nabla\theta)^2 + 12(\nabla\psi)^2 + \lambda_2\psi^2 + \lambda_4\psi^4}$$

Model Outputs

Model: o3 (Score: 0.0)

Model Solution: L' $=$ i$\,\psi\,\partial_t\theta\ -\ \frac{1}{2}\left[(\partial_x\theta)^2 + (\partial_x\psi)^2\right]\ +\ \lambda_2\,\psi^2\ +\ \lambda_4\,\psi^4$

Judge's Evaluation: The ground truth expression is:

$$\mathcal{L}' = i\psi\partial_t\theta + \frac{1}{2}(\nabla\theta)^2 + \frac{1}{2}(\nabla\psi)^2 + \lambda_2\psi^2 + \lambda_4\psi^4$$

The model-generated expression is:

$$\mathcal{L}' = i\psi\partial_t\theta - \frac{1}{2}\left[(\partial_x\theta)^2 + (\partial_x\psi)^2\right] + \lambda_2\psi^2 + \lambda_4\psi^4$$

Key points of comparison:

1. **Sign in front of the gradient terms:** - Ground truth has **$+ (1/2)(\nabla\theta)^2 + (1/2)(\nabla\psi)^2 **- Model has ** - (1/2)[(\partial_x\theta)^2 + (\partial_x\psi)^2] **$

The sign is opposite, which is a significant error.

2. **Gradient operator:** - Ground truth uses ∇, which implies the full spatial gradient (all spatial dimensions). - Model uses only ∂_x, i.e., derivative with respect to one spatial coordinate only.

This is a simplification or omission of other spatial derivatives, which is a notable difference.

3. **Other terms:** - The $i\psi\partial_t\theta$ term matches exactly. - The $\lambda_2\psi^2$ and $\lambda_4\psi^4$ terms match exactly.

Overall, the model expression has two significant errors: the sign in front of the gradient terms is wrong, and the gradient is only one-dimensional rather than full spatial gradient. These are more than minor errors.

Therefore, the expression is incorrect.

Score: 0

Problem 49 (Paper: 2506.21540v1)

Problem Statement

Background: Consider a 1+d dimensional field theory described by the Euclidean action for an N-component real scalar field $\vec{\psi}$:

$$S = \int^{d} xt \left[\frac{1}{2} \vec{\psi} \cdot \left(\hat{\lambda}_2 - \nabla^2 + \partial_t^2 \nabla^{-2} \right) \vec{\psi} + \frac{\lambda_4}{N} (\vec{\psi}^2)^2 \right]$$

where $\hat{\lambda}_2$ and λ_4 are coupling constants, ∇^2 is the d-dimensional Laplacian, and ∇^{-2} is its inverse. The spatial dimension is $d = 2 - \epsilon$, where ϵ is a small positive parameter. A one-loop Wilsonian renormalization group (RG) analysis yields the following flow equation for the coupling λ_4:

$$\partial_\tau \lambda_4 = \epsilon \lambda_4 - (4N + 32) \lambda_4^2 I_2$$

where $\tau = \ln b$ is the RG flow parameter (with b being the rescaling factor). The loop integral I_2 is given by:

$$I_2 = \frac{K_d \Lambda}{4(\hat{\lambda}_2 + \Lambda^2)^{3/2}}$$

where $K_d = \frac{2^{1-d} \pi^{-d/2}}{\Gamma(d/2)}$ is a constant, and Λ is a momentum cutoff. Assume $\hat{\lambda}_2 \ll \Lambda^2$.

Task: Derive the simplified one-loop RG flow equation for λ_4 by evaluating the integral I_2 under the given approximation and substituting it into the flow equation. Then, perform a redefinition of λ_4 to absorb the constant prefactor of the λ_4^2 term, expressing the final RG equation in terms of the redefined λ_4.

Ground Truth Solution

$$\boxed{\partial_\tau \lambda_4 = \epsilon \lambda_4 - \lambda_4^2}$$

Model Outputs

Model: o3 (Score: 1.0)

Model Solution: $\partial_\tau g = \epsilon\, g - g^2$

Judge's Evaluation: The model-generated expression is:

$$\partial_\tau g = \epsilon g - g^2$$

The ground truth expression is:

$$\partial_\tau \lambda_4 = \epsilon \lambda_4 - \lambda_4^2$$

The only difference is the variable name: g instead of λ_4. Since the rubric states that differences in variable names do not affect correctness, the expressions are mathematically equivalent.

There are no missing terms, incorrect signs, or coefficients. The structure and form of the expression match exactly.

Score: 1