LLM Benchmark Report

Generated by script

June 30, 2025

Summary Statistics

Model	Total	Correct	Partial	Incorrect	Errors	Avg. Score
o3	5	3	0	2	0	0.60

Problem Details

Problem 1 (Paper: 2506.17853v1)

Problem Statement

Background: The dynamics of a robotic mechanical metamaterial are described by a reaction-diffusion system for its dimensionless angular displacements, α and β:

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

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

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:

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

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

$\psi'(\alpha)$ is a non-monotonic, odd, piecewise continuous tri-linear function with roots at $\alpha = \{0, \pm\alpha^*\}$. Its derivative, $\psi''(\alpha)$, is:

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

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

Task: Derive the matrix $\mathbf{J}_\kappa$ for the linear stability analysis of this system. This matrix is defined as $\mathbf{J}_\kappa = \mathbf{J_R} - \kappa^2\mathbf{D}$, where $\mathbf{J_R}$ 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: 0.0)

Model Solution: $\mathrm{J}_{\boldsymbol{\kappa}} = \mathbf{J_R} - \kappa^2\mathbf{D} = \begin{pmatrix} -k_{\alpha\beta} - k_2 - \kappa^2 k_\alpha & k_{\alpha\beta} - \sigma \\ k_{\alpha\beta} & -k_{\alpha\beta} - \dfrac{\kappa^2 k_\beta}{\gamma} \end{pmatrix}$

Judge's Evaluation: The ground truth matrix is:

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

The model-generated matrix is:

$$\mathbf{J}_{\boldsymbol{\kappa}} = \begin{pmatrix} -k_{\alpha\beta} - k_2 - \kappa^2 k_\alpha & k_{\alpha\beta} - \sigma \\ k_{\alpha\beta} & -k_{\alpha\beta} - \frac{\kappa^2 k_\beta}{\gamma} \end{pmatrix}$$

Comparison:

1. Top-left element: - Ground truth: $-k_{\alpha\beta} - \psi''(0) - \kappa^2 k_\alpha$ - Model: $-k_{\alpha\beta} - k_2 - \kappa^2 k_\alpha$

Here, the model uses k_2 instead of $\psi''(0)$. If k_2 is meant to represent $\psi''(0)$, this is a variable name difference, which is acceptable. However, if k_2 is not equivalent to $\psi''(0)$, this is an error.

2. Top-right element: - Both: $k_{\alpha\beta} - \sigma$ — matches exactly.

3. Bottom-left element: - Ground truth: $\frac{k_{\alpha\beta}}{\gamma}$ - Model: $k_{\alpha\beta}$

The model is missing division by γ here, which is a significant error.

4. Bottom-right element: - Ground truth: $-\frac{k_{\alpha\beta}}{\gamma} - \frac{\kappa^2 k_\beta}{\gamma}$ - Model: $-k_{\alpha\beta} - \frac{\kappa^2 k_\beta}{\gamma}$

The model is missing division by γ in the first term.

Summary:

- The top-left element difference is likely a variable name difference (acceptable). - The bottom row elements are missing division by γ, which is a significant error. - The top-right element matches exactly.

Because there are more than one minor errors (two missing divisions by γ in the bottom row), the expression is not nearly correct.

Therefore, the score is 0.

Score: 0

Problem 2 (Paper: 2506.17853v1)

Problem Statement

Background: The dynamics of a robotic mechanical metamaterial with local rotational freedoms α and β are described by a set of dimensionless reaction-diffusion equations. The governing equations are:

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

where Δ is the Laplacian operator, t is dimensionless time, $\gamma = \eta_\beta/\eta_\alpha$, and $d = k_\beta/k_\alpha$. The reaction terms are:

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

where $k_{\alpha\beta}$ is a stiffness parameter, σ is a feedback parameter, and $\psi'(\alpha)$ is a non-monotonic function. At the quiescent, spatially uniform state $\mathbf{\Phi}_0 = (\alpha_0, \beta_0) = (0, 0)$, $\psi''(0) = k_2$. To analyze stability, introduce a small perturbation $\mathbf{\Phi}(x, t) = \mathbf{\Phi}_0 + \tilde{\mathbf{\Phi}}\mathrm{e}^{\mathrm{i}\kappa x + \Omega t}$, where $\tilde{\mathbf{\Phi}}^{\mathrm{T}} = [\tilde{\alpha}, \tilde{\beta}]$ is the perturbation amplitude, κ is the spatial frequency, and Ω is the perturbation growth rate. The system can be written 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 $\mathbf{J}_\kappa = \mathbf{J}_{\mathbf{R}} - \kappa^2\mathbf{D}$, where $\mathbf{J}_{\mathbf{R}}$ is the Jacobian matrix of $\bar{\mathbf{R}}(\alpha, \beta)$ 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

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

Model Outputs

Model: o3 (Score: 1.0)

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

Judge's Evaluation: The ground truth expression is:

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

The model-generated expression is:

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

—
Comparison:
1. **Form of the linear term (outside the square root):**
- Ground truth linear term inside parentheses is:

$$-\frac{k_{\alpha\beta}(\gamma+1) + k_2\gamma}{\gamma} - \frac{\gamma+d}{\gamma}\kappa^2$$

- Model linear term inside brackets is:

$$k_{\alpha\beta} + k_2 + \kappa^2 + \frac{k_{\alpha\beta} + d\kappa^2}{\gamma}$$

and the whole is multiplied by $-\frac{1}{2}$.

Rewrite the ground truth linear term:

$$-\frac{k_{\alpha\beta}(\gamma+1) + k_2\gamma}{\gamma} - \frac{\gamma+d}{\gamma}\kappa^2 = -\frac{k_{\alpha\beta}\gamma + k_{\alpha\beta} + k_2\gamma}{\gamma} - \frac{\gamma+d}{\gamma}\kappa^2$$
$$= -\left(k_{\alpha\beta} + \frac{k_{\alpha\beta}}{\gamma} + k_2\right) - \left(1 + \frac{d}{\gamma}\right)\kappa^2$$

But this is not exactly the same as the model's linear term:

$$k_{\alpha\beta} + k_2 + \kappa^2 + \frac{k_{\alpha\beta} + d\kappa^2}{\gamma} = \left(k_{\alpha\beta} + k_2 + \frac{k_{\alpha\beta}}{\gamma}\right) + \left(\kappa^2 + \frac{d}{\gamma}\kappa^2\right) = k_{\alpha\beta} + k_2 + \frac{k_{\alpha\beta}}{\gamma} + \kappa^2\left(1 + \frac{d}{\gamma}\right)$$

The model's linear term is positive inside the bracket, but the entire bracket is multiplied by $-\frac{1}{2}$, so the sign matches the ground truth.

So the linear terms are algebraically equivalent.

2. **Inside the square root:**

- Ground truth:

$$(\text{linear term})^2 - \frac{4}{\gamma}\left[k_{\alpha\beta}(k_2+\sigma) + (k_{\alpha\beta}(d+1) + k_2 d)\kappa^2 + d\kappa^4\right]$$

- Model:

$$(\text{linear term})^2 - \frac{4}{\gamma}\left[(k_{\alpha\beta} + k_2 + \kappa^2)(k_{\alpha\beta} + d\kappa^2) - k_{\alpha\beta}(k_{\alpha\beta} - \sigma)\right]$$

Let's expand the model's bracket:

$$(k_{\alpha\beta} + k_2 + \kappa^2)(k_{\alpha\beta} + d\kappa^2) - k_{\alpha\beta}(k_{\alpha\beta} - \sigma)$$

First, expand the product:

$$= k_{\alpha\beta}^2 + k_{\alpha\beta}k_2 + k_{\alpha\beta}\kappa^2 + dk_{\alpha\beta}\kappa^2 + dk_2\kappa^2 + d\kappa^4 - k_{\alpha\beta}^2 + k_{\alpha\beta}\sigma$$

Simplify:

$$= k_{\alpha\beta}k_2 + k_{\alpha\beta}\kappa^2 + dk_{\alpha\beta}\kappa^2 + dk_2\kappa^2 + d\kappa^4 + k_{\alpha\beta}\sigma$$

Group terms:

$$= k_{\alpha\beta}(k_2+\sigma) + \kappa^2(k_{\alpha\beta} + dk_{\alpha\beta} + dk_2) + d\kappa^4 = k_{\alpha\beta}(k_2+\sigma) + \kappa^2(k_{\alpha\beta}(d+1) + dk_2) + d\kappa^4$$

This matches exactly the ground truth bracket inside the square root.

—

Summary:

- The model's expression is algebraically equivalent to the ground truth expression. - The model uses a slightly different but equivalent form for the linear term and the term inside the square root. - No missing terms or incorrect signs. - The model expression includes the $\pm$ and the factor $\frac{1}{2}$ correctly. - Variable names match. - The model expression is mathematically equivalent to the ground truth.

—

Score: 1

The model-generated expression is completely correct and mathematically equivalent to the ground truth.

—

Score: 1

Problem 3 (Paper: 2506.18641v1)

Problem Statement

Background: Consider an undirected, unweighted network $G = (V, E)$ with N nodes, adjacency matrix $\mathbf{A}$, and diagonal degree matrix $\mathbf{D}$. The Laplacian matrix is $\mathbf{L} = \mathbf{D} - \mathbf{A}$, which has real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$.

The partition function Z_τ is defined as:

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

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

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

The spectral entropy S_τ is defined as:

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

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: $\mathrm{S}_\tau = -\sum_{i=1}^N \frac{e^{-\tau\lambda_i}}{\sum_j e^{-\tau\lambda_j}} \ln\left(\frac{e^{-\tau\lambda_i}}{\sum_k e^{-\tau\lambda_k}}\right) = \tau \frac{\sum_{i=1}^N \lambda_i e^{-\tau\lambda_i}}{\sum_{j=1}^N e^{-\tau\lambda_j}} + \ln\left(\sum_{i=1}^N e^{-\tau\lambda_i}\right)$

Judge's Evaluation: The model-generated expression first writes the entropy S_τ in the form of a sum over probabilities times the log of those probabilities, which is a standard definition of entropy. Then it simplifies this expression to the form:

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

This matches exactly the ground truth expression, except for the variable index in the logarithm sum (the ground truth uses j and the model uses i), which is irrelevant since the summation index is a dummy variable.

The model's expression is mathematically equivalent and correctly derived. There are no missing terms, no sign errors, and the coefficients and variables match perfectly.

Score: 1

Problem 4 (Paper: 2506.18641v1)

Problem Statement

Background: Consider an undirected, unweighted network $G = (V, E)$ with N nodes, adjacency matrix $\mathbf{A}$, and degree matrix $\mathbf{D}$. The Laplacian matrix is $\mathbf{L} = \mathbf{D} - \mathbf{A}$. $\mathbf{L}$ is real symmetric and can be diagonalized with eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_N$ and eigenvalues $\lambda_1, \ldots, \lambda_N$. The matrix exponential $e^{\mathbf{M}} = \sum_{k=0}^{\infty} \frac{\mathbf{M}^k}{k!}$. The trace of a square matrix $\mathbf{M}$, $\mathrm{Tr}(\mathbf{M})$, is the sum of its diagonal elements. The partition function Z_τ is defined as $Z_\tau = \mathrm{Tr}(e^{-\tau \mathbf{L}})$, where $\tau > 0$.

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

$Solution : Let \mathrm{L} be the Laplacian matrix with the eigenvalues \lambda_1, \ldots, \lambda_N$. Since $\mathbf{L}$ is a real symmetric matrix, it is diagonalizal

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 $Z_\tau = \sum_{i=1}^{N} e^{-\tau\lambda_i}$ exactly matches the final expression in the ground truth solution. It correctly represents the partition function as the sum over the exponentials of the eigenvalues multiplied by $-\tau$. There are no missing terms, incorrect coefficients, or sign errors. The expression is mathematically equivalent and fully correct.

Score: 1

Problem 5 (Paper: 2506.19617v1)

Problem Statement

Background: Consider two clusters of N Rössler oscillators. The dynamics of the i^{th} oscillator in the first cluster $(\mathbf{x}_i)$ and second cluster $(\mathbf{y}_i)$ are given by:

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

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

where a, b, c are Rössler parameters, ϵ is the intra-cluster coupling, and μ is the inter-cluster coupling. $\overline{x}^2$ and $\overline{y}^2$ are local centers of mass. The intra-cluster connectivity g_{ij}^k and inter-cluster connectivity D_{XY}, D_{YX} are defined as:

$$g_{ij}^k = \begin{cases} 1 \;\; if \;\; d_{ij}^k \leq d_0^k \\ 0 \;\; otherwise, \end{cases} \quad \text{and} \quad D_{XY} = D_{YX} = \begin{cases} 1 \;\; if \;\; s_{XY}, s_{YX} \leq s_0 \\ 0 \;\; otherwise \end{cases}$$

where d_{ij}^k and s_{XY} are Euclidean distances, and d_0^k and s_0 are distance thresholds.

To analyze stability, the error system is defined by $e_i^k = x_i^k - y_i^k$, with dynamics given by:

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

where $f_{1i} = \sum\limits_{j}^{N}\left(g_{ij}^1 - g_{ij}^2\right)\left(y_j^2 - y_i^2\right)$ and $f_{2i} = \frac{1}{m_1}\sum\limits_{j}^{m_1} g_{ij}^1 y_j^2 - \frac{1}{m_2}\sum\limits_{j}^{m_2} g_{ij}^2 y_j^2$. The non-linear term $e_i^1 e_i^3$ in $\dot{e}_i^3$ is disregarded.

A Lyapunov function candidate V_i and its time derivative $\dot{V}_i$ are given by:

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

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

Assume $y_m^1 = \max\left|y_i^1\right|$ and $y_m^3 = \max\left|y_i^3\right|$. Assume there exist positive constants Z_1, Z_2 such that $\sum\limits_{\substack{j=1,\\ j\neq i}}^{N} g_{ij}^1\left(e_j^2\right)^2 \leq Z_1\left(e_i^2\right)^2$ and $\sum\limits_{\substack{j=1,\\ j\neq i}}^{m_2} g_{ij}^2\left(e_j^2\right)^2 \leq Z_2\left(e_i^2\right)^2$, and 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 property $|A|\,|B| \leq \frac{1}{2}\left(A^2 + B^2\right)$ to all product terms involving error components. Then, substitute the definitions for γ_1 and γ_2, apply the Z bound, and 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: $\mathrm{f3}_{\mathrm{max}} = a + \epsilon\left(Z + \frac{1}{2}\right) + \frac{\mu}{2}$

Judge's Evaluation: The ground truth expression is:

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

The model-generated expression is:

$$f3_{\mathrm{max}} = a + \epsilon\left(Z + \frac{1}{2}\right) + \frac{\mu}{2}$$

Comparison:
- The model expression is missing the term $\frac{\mu D_{xy}}{m_2}$ inside the parentheses multiplied by Z. - The model expression does not include the $-\mu D_{xy}\left(\frac{1}{2} + \frac{1}{m_2}\right)$ term. - The model expression is missing the entire summation term $-\frac{\epsilon}{2}\sum_{j\neq i} g_{ij}^{1}$. - The model expression incorrectly groups ϵ with $Z + \frac{1}{2}$, which is not equivalent to the ground truth. - The model expression has a standalone $\frac{\mu}{2}$ term, which does not appear in the ground truth.

Overall, the model expression is missing multiple terms and has incorrect grouping of terms. It is not close to the ground truth expression.

Score: 0