WHEN ARE ENERGY PROXIES RELIABLE? EXCITATION AND CURVATURE IN RELU EQUILIBRIA

Anonymous authors
Paper under double-blind review

ABSTRACT

An energy proxy replaces asymmetric recurrent interactions with a symmetric field, but a smaller field mismatch need not imply a more accurate answer. We construct a three-dimensional ReLU family whose original updates are uniformly contractive, yet the metric that minimizes rotational defect produces unbounded proxy error. This metric is fixed, has condition number below $3/2$, and performs worse than the bounded-error identity proxy. For fixed input distributions on a positive box, the energy penalty grows as ε^{-1} and the mean squared state error grows as ε^{-2}, where ε is the proxy curvature margin. We explain these rates with two-sided population bounds for strongly convex proxies, valid across activation changes and at any rotation strength. Inactive-unit slack can screen rotation even when the response second moment is nonsingular, so the lower bound measures excitation through jointly active responses. Complementarity supplies an energy-budget upper bound, and both bounds recover the growth rates on the separating family. Input-coverage and response-balance conditions specify when the lower bound is positive and quantitatively informative. Experiments on trained ReLU networks and a fixed-anchor recurrent Transformer support the directional mechanism: input-excited rotation predicts proxy errors better than rotation magnitude.

1 INTRODUCTION

Neural networks can compute by repeatedly updating an internal state with shared parameters instead of using separate parameters at each step. Equilibrium models define the resulting representation through a fixed point, while looped networks use repeated computation to increase effective depth (Bai et al., 2019; Giannou et al., 2023; Liang et al., 2025). An energy interpretation makes the state a minimizer of an input-dependent objective, enabling optimization tools to analyze and compute the answer. An objective used to explain or replace a trained recurrence should therefore preserve its computed state.

Symmetric recurrent networks provide classical examples of such interpretations (Hopfield, 1982; Hahnloser et al., 2003), and optimization perspectives have been developed for Transformer components and graph networks (Yang et al., 2022; 2025). For asymmetric interactions, however, the update field can contain a rotational component that a scalar gradient cannot generate in the chosen coordinates. A natural approximation is to remove this component and use the remaining symmetric field to define an energy proxy. Existing perturbation bounds compare contractive equilibria, including approximations based on the positive-semidefinite part of the symmetric weights (Tsuchida et al., 2022, Propositions 12–13). These bounds leave a selection question: when coordinate systems yield different proxies, does the one closest to a gradient field also preserve the answer?

We study this question in the threshold-linear recurrence

$$x^{t+1} = T_{W,b}(x^t) := [Wx^t + b]_+, \qquad W \in \mathbb{R}^{n\times n}, \quad b \in \mathbb{R}^n, \tag{1}$$

where x^t is the internal state, b is the fixed input, and ReLU replaces each negative coordinate by zero. This model retains both asymmetric interactions and changing activation patterns while admitting an explicit equilibrium analysis (Hadeler & Kuhn, 1987; Nozari & Cortés, 2018; Winston & Kolter, 2020). We compare the original equilibrium with the energy minimizer obtained after a positive diagonal change of coordinates and symmetrization, mapping both answers back to the original

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

state space. Positive diagonal scaling commutes with ReLU, so this comparison isolates the error introduced by the proxy from the coordinate transformation itself.

Our central finding is that the least rotational defect can select an arbitrarily inaccurate energy proxy, even when the original recurrence is uniformly contractive. In a three-dimensional construction, the defect-minimizing metric is fixed and has condition number below $3/2$, yet its answer error grows as ε^{-1} while the identity-metric proxy has bounded error (theorem 3). Here $\varepsilon > 0$ is the smallest curvature of the selected energy, which remains strongly convex for every member of the family. Thus the failure comes neither from unstable original dynamics nor from a singular coordinate transformation: making the update closer to a gradient field can flatten the energy in a direction that matters for the answer.

The mechanism involves three interacting quantities: the discarded asymmetry, the response excited by the input, and the energy curvature along the affected directions. If K denotes the skew-symmetric part of the transformed weights and z the proxy minimizer, then Kz is the discrepancy that the discarded interaction produces at that answer. Its effect depends on where it points, not just on the overall size of K. A flat direction can amplify this discrepancy, whereas an inactive ReLU unit can absorb it without changing the answer. Our analysis combines curvature with co-activation, which records the interacting units that respond together, to distinguish visible rotation from rotation absorbed by inactive-unit slack. The resulting bounds hold even when the original and proxy equilibria have different active sets.

This distinction between field approximation and answer fidelity leads to three contributions:

- **A failure theorem for energy-proxy selection.** We construct uniformly contractive ReLU networks for which minimizing rotational defect gives unbounded answer error despite a fixed, well-conditioned metric and a bounded-error identity proxy (theorem 3).
- **Input- and curvature-dependent reliability conditions.** We establish state and energy error bounds across activation changes, including a co-activation lower bound and an energy-budget upper bound with matching rates on the separating family (lemma 1 and theorem 2). Input coverage and directional tightness results specify when these certificates are informative (propositions 12, 13 and 15).
- **Controlled and trained-network evidence.** Analytic families test the predicted mechanisms; across twelve checkpoints of three trained ReLU classifiers, input-excited rotation ranks proxy errors better than rotational defect. A recurrent-Transformer intervention supports the same directional mechanism for local response errors (Section 6).

2 Related Work

Our question connects energy interpretations of neural computation with equilibrium sensitivity: an energy must both describe an update and preserve the state it computes. Symmetric recurrent-network energies (Hopfield, 1982; Hahnloser et al., 2003) connect fixed points to optimization. Optimization views of Transformers and graph networks extend this perspective to modern architectures (Yang et al., 2022; 2025; Gu et al., 2020). Vector-field equilibrium propagation and monotone-operator networks accommodate asymmetric dynamics (Scellier et al., 2018; Winston & Kolter, 2020; Baker et al., 2023). Tsuchida et al. (2022, Propositions 12–13) bound the difference between contractive equilibria and apply this bound to the positive-semidefinite part of the symmetric weights. Their approximation error depends on both negative symmetric eigenvalues and the skew norm. Our population lower bound identifies when the rotation discarded by the proxy must change the equilibrium, even when activation patterns change. The separation theorem addresses the coordinate choice itself: minimizing defect can reduce proxy curvature and cause unbounded error with uniformly contractive original dynamics.

Diagonal symmetrizability and detailed balance characterize exact potentials (Engel & Schneider, 1980; McKee & Smyth, 2020; Jiang et al., 2018). Frobenius balancing optimizes matrix norms under diagonal similarity (Osborne, 1960; Ostrovsky et al., 2017; Cohen et al., 2017), with a convex-flow formulation (Altschuler, 2022). Graph Hodge decomposition separates gradient and cyclic edge fields (Jiang et al., 2011). Together, these tools identify the coordinates that minimize rotational defect and the rotation that no coordinates can remove. This geometric choice depends on the weights alone; our separation theorem asks whether it also preserves an input-dependent equilibrium.

Related conservative-score regularization (Chao et al., 2023) and nearest reversible-chain optimization (Durastante et al., 2025) study conservative structure under their respective vector-field and stochastic constraints.

To connect the remaining rotation to answer error, we build on sensitivity analysis for strongly monotone variational inequalities (VIs) (Facchinei & Pang, 2003) and complementarity residual bounds for approximate solutions across active sets (Chen & Xiang, 2006). Fixed activation regions admit linear response calculations (Durstewitz, 2017). Duncan et al. (2016, Lemma 3 and equations (31)–(32)) express nonreversible sampling variance through the symmetric part of an inverse operator and a whitened skew perturbation. The same resolvent structure underlies our full-face refinement. The cross-face bound instead uses complementarity to relate signed forcing to a Gram matrix defined on active components. This relation gives an excitation condition even when the actual and proxy active sets differ.

3 Energy Proxies and Rotational Defect

Before comparing answers, we identify which coordinate changes preserve the recurrent computation and which asymmetric interactions remain. This separates an exact reparameterization from the approximation that creates an energy proxy.

3.1 Coordinates that preserve the recurrent update

The coordinate change must respect the activation, rather than replace it with a different projection. Positive diagonal scaling commutes with ReLU and defines exactly the constant metrics whose orthant projection is coordinatewise ReLU. Let $\mathcal{C} = \mathbb{R}^n_+$, let $\iota_{\mathcal{C}}$ be its convex indicator, and write $\|x\|_H^2 = x^\top Hx$ for $H \succ 0$. Keeping the state space, affine preactivation, and activation fixed, an exact projected-gradient representation exists precisely when some positive diagonal D satisfies

$$DW = W^\top D. \tag{2}$$

Appendix A.1 gives the formal definition and derivation. This identifies the coordinate class in which we compare approximate energy models.

3.2 Cycle certificates for unavoidable rotation

The weighted symmetry condition can fail because the required coordinate ratios disagree around a cycle. We quantify this obstruction and then identify the coordinates that minimize the remaining rotation. For this calculation, assume reciprocal same-sign support: every off-diagonal pair is either zero in both directions or satisfies $w_{ij}w_{ji} > 0$. Detailed balance then holds exactly when the edge log ratios have zero circulation on every cycle (Engel & Schneider, 1980; McKee & Smyth, 2020; Jiang et al., 2018). Choose one orientation $e = (i, j)$ of each support edge, let B have column $e_j - e_i$, and define

$$c_e = \log\left|\frac{w_{ij}}{w_{ji}}\right|, \qquad g_e = \sqrt{|w_{ij}w_{ji}|}, \qquad r(u) = c - \mathsf{B}^\top u. \tag{3}$$

For $D(u) = \mathrm{diag}(e^{u_1}, \ldots, e^{u_n})$, set

$$\widetilde{W}(u) = D(u)^{1/2} W D(u)^{-1/2}, \qquad K(u) = \frac{\widetilde{W}(u) - \widetilde{W}(u)^\top}{2}. \tag{4}$$

Then

$$\|K(u)\|_F^2 = J(u) := \sum_{e \in E} g_e^2 \big(\cosh r_e(u) - 1\big). \tag{5}$$

The corresponding Frobenius-norm identity is

$$C_W = \sum_i w_{ii}^2 + 2\sum_{e \in E} g_e^2, \qquad \left\|\widetilde{W}(u)\right\|_F^2 = C_W + 2J(u). \tag{6}$$

Frobenius matrix balancing therefore minimizes rotational defect (Osborne, 1960; Ostrovsky et al., 2017; Cohen et al., 2017). The minimizer is unique up to componentwise constants in u (Appendix A.3). The inequality $\cosh t - 1 \geq t^2/2$ also gives a coordinate-invariant lower bound:

$$J(u) \geq \mathcal{L}_{\circ}(W) := \frac{1}{2} \min_{v \in \mathbb{R}^n} \sum_{e \in E} g_e^2 \big(c_e - [\mathsf{B}^\top v]_e\big)^2. \tag{7}$$

This weighted Hodge residual (Jiang et al., 2011) is positive exactly when some cycle circulation is nonzero. Appendix A.4 gives its cycle-space formula and proves that the bound is sharp near balance. It measures the rotation that every compatible coordinate choice retains.

3.3 The energy proxy and its error

The geometric optimum minimizes a matrix mismatch, but our target is the resulting answer. We therefore define the proxy and compare its minimizer with the original equilibrium in the same state coordinates. Fix any positive diagonal $D = D(u)$, abbreviate $A = \widetilde{W}(u)$ and $c_D = D(u)^{1/2} b$, and decompose

$$A = S + K, \qquad S = S^\top, \qquad K = -K^\top. \tag{8}$$

In the scaled coordinates $y = D(u)^{1/2} x$, we compare the actual and energy-proxy maps

$$\mathcal{T}(y) = [Ay + c_D]_+, \qquad \mathcal{S}(y) = [Sy + c_D]_+. \tag{9}$$

The proxy objective is

$$\Psi_S(y) = \frac{1}{2} y^\top (I - S) y - c_D^\top y + \iota_{\mathcal{C}}(y). \tag{10}$$

Write $Q = I - S$, and assume $Q \succ 0$ when studying this proxy. Its minimizer z is unique. The actual equilibrium y also exists and is unique because $(Q - K)y - c_D$ is strongly monotone. We measure proxy fidelity by

$$\Delta_D(b) = \Psi_S(y) - \Psi_S(z), \qquad e_D(b) = \|x^\star - \widehat{x}_D\|_2, \qquad \widehat{x}_D = D^{-1/2} z, \tag{11}$$

where $y = D^{1/2} x^\star$. These are, respectively, the proxy energy penalty at the actual answer and the answer error in the original state coordinates. The equilibrium analysis applies to every symmetric–skew decomposition with $Q \succ 0$; reciprocal same-sign support supplies the additional cycle certificate.

4 When Rotational Defect Predicts Error

The proxy discards a matrix, but the input determines which part of that matrix affects the answer. We first bound this effect for one input, then identify the input populations that make rotation observable.

4.1 Input-dependent response

At a proxy minimizer, the discarded rotation acts as a forcing term; inactive-unit slack can absorb part of it. To separate forcing strength from activation effects, we scale the rotation by τ while fixing S and c_D; the original recurrence corresponds to $\tau = 1$.

Lemma 1 (Solver-independent equilibrium response)**.** *Suppose $Q = I - S \succ 0$, set $\mu = \lambda_{\min}(Q)$, $\kappa = \|K\|_2$, and $c_0 = \|[c_D]_+\|_2$. For every real τ, the map*

$$\mathcal{T}_\tau(y) = [(S + \tau K)y + c_D]_+ \tag{12}$$

has a unique fixed point $y_\tau^\star$. The point $z^\star = y_0^\star$ uniquely minimizes Ψ_S. Write $s^\star = Qz^\star - c_D$ and define

$$v_\tau = [\tau K z^\star - s^\star]_+, \qquad L_\tau = \|Q - \tau K\|_2, \qquad \ell_\tau = \lambda_{\max}(Q - \tau^2 K Q^{-1} K). \tag{13}$$

Then

$$\frac{\|v_\tau\|_2}{L_\tau} \leq \|y_\tau^\star - z^\star\|_2 \leq \min\left\{ \frac{|\tau|\,\|Kz^\star\|_2}{\mu}, \sqrt{\frac{(z^\star)^\top Q z^\star}{\mu}} \right\}, \tag{14}$$

$$\|y_\tau^\star - z^\star\|_2 \leq \frac{|\tau| \kappa c_0}{\mu^2}, \tag{15}$$

$$\frac{\|v_\tau\|_2^2}{2\ell_\tau} \leq \Psi_S(y_\tau^\star) - \Psi_S(z^\star) \leq \frac{1}{2} \min\big\{ \tau^2 (Kz^\star)^\top Q^{-1} (Kz^\star), (z^\star)^\top Q z^\star \big\}. \tag{16}$$

Moreover, $y_\tau^\star = z^\star$ if and only if $v_\tau = 0$. In particular, $Kz^\star = 0$ implies equality for every τ.

Let $\Delta_\tau = \Psi_S(y_\tau^\star) - \Psi_S(z^\star)$. Complementarity and the identity $y^\top K y = 0$ give

$$\Delta_\tau = \frac{1}{2} c_D^\top (z^\star - y_\tau^\star) = \frac{1}{2}\big((z^\star)^\top Q z^\star - (y_\tau^\star)^\top Q y_\tau^\star\big). \tag{17}$$

The proxy response therefore yields an upper bound independent of rotation strength. A comparison of the two μ-strongly monotone VIs gives the perturbative upper bounds (Facchinei & Pang, 2003). For the lower bounds, let $\delta = y_\tau^\star - z^\star$. Actual dual feasibility gives $\tau K z^\star - s^\star \le (Q - \tau K)\delta$ coordinatewise. The positive part has squared norm at most $\ell_\tau \delta^\top Q \delta$, and the proxy energy penalty is at least $\delta^\top Q \delta / 2$. Complementarity residual bounds give related state-error comparisons (Chen & Xiang, 2006); the energy bound retains the proxy curvature. Appendix A.5 proves the result and sharpness of the lower-bound constant across a face change.

The vector v_τ measures the part of the rotational forcing that violates dual feasibility at the proxy. On an inactive coordinate, the slack $s_i^\star$ can absorb that forcing. On an active coordinate the slack is zero. This distinction determines which input populations make rotational defect visible in the answer.

4.2 Population bounds across activation faces

To obtain an average-error guarantee, we need inputs that expose rotation on jointly active units, where slack cannot absorb it. The following quantity measures this exposure within each connected active component. Fix S, K, D and draw $c = c_D$ from ν. Write $z(c)$ and $y_\tau(c)$ for the proxy and actual equilibria, and $\Delta_\tau(c) = \Psi_{S,c}(y_\tau(c)) - \Psi_{S,c}(z(c))$. Let E contain the unordered off-diagonal support of K and let $\mathcal{K}_E$ be the space of skew matrices supported on E. Let $I(z) = \operatorname{supp}(z)$ and let $\mathcal{C}_E(z)$ be the collection of connected components of $E[I(z)]$ with at least two vertices. Write P_C for the coordinate projector onto C and z_C for the restriction of z to C. Define the directional excitation and its uniform coefficient by

$$\theta_C(z) = \frac{\min_{i \in C} z_i^2}{\|z_C\|_2^2}, \qquad g_\nu(H) = \mathbb{E}_\nu \left[\sum_{C \in \mathcal{C}_E(z)} \theta_C(z) \, \|P_C H z\|_2^2 \right], \tag{18}$$

$$\alpha_{\mathrm{co}} = \inf_{\substack{H \in \mathcal{K}_E \\ \|H\|_F = 1}} g_\nu(H), \qquad R_\nu = \mathbb{E}_\nu[z^\top Q z]. \tag{19}$$

The coefficient is the smallest eigenvalue of a Gram matrix that records proxy responses to edge-supported rotations (Appendix A.7). For $E = \varnothing$, set $\alpha_{\mathrm{co}} = 0$. Each active component has its own normalization, so an isolated unit leaves the Gram matrix unchanged. We use the support of K for E unless a larger edge space is specified.

A very weak edge can merge two active components and reduce their normalization factors. We can instead omit that edge from the component graph and explicitly bound its forcing. For a retained edge set $F \subseteq E$, let K_F agree with K on F and vanish elsewhere, and put $R_F = K - K_F$. Write $g_{\nu,F}$ for equation (18) computed on F, and define

$$\rho_F = \left(\mathbb{E}_\nu \|R_F z\|_2^2\right)^{1/2}, \qquad b_\nu(F) = \left(\sqrt{g_{\nu,F}(K_F)} - \rho_F\right)_+^2. \tag{20}$$

The next theorem combines these excitation quantities with the single-input bounds; the actual and proxy active sets may differ.

Theorem 2 (Population error across activation faces)**.** *Let $Q = I - S \succ 0$, $\mu = \lambda_{\min}(Q)$, $K \in \mathcal{K}_E$, and $\mathbb{E}_\nu[zz^\top] \preceq \beta I$ for some finite β. With ℓ_τ, L_τ from equation (13), every real τ satisfies*

$$\frac{\tau^2 g_\nu(K)}{2\ell_\tau} \le \mathbb{E}_\nu \Delta_\tau \le \frac{1}{2} \min\left\{ \frac{\tau^2 \beta \|K\|_F^2}{\mu}, R_\nu \right\}. \tag{21}$$

For $e_{\tau,D}(c) = \left\|D^{-1/2}(y_\tau(c) - z(c))\right\|_2$, $d_{\min} = \min_i D_{ii}$, and $d_{\max} = \max_i D_{ii}$,

$$\frac{\tau^2 g_\nu(K)}{d_{\max} L_\tau^2} \le \mathbb{E}_\nu e_{\tau,D}^2 \le \frac{1}{d_{\min}} \min\left\{ \frac{\tau^2 \beta \|K\|_F^2}{\mu^2}, \frac{R_\nu}{\mu} \right\}. \tag{22}$$

Both lower bounds also hold with $g_\nu(K)$ replaced by $b_\nu(F)$ for any $F \subseteq E$, with ℓ_τ, L_τ computed from the full K. Both lower bounds remain valid with $g_\nu(K)$ replaced by $\alpha_{\mathrm{co}} \|K\|_F^2$. In compatible coordinates with reciprocal same-sign support, this term can be further replaced by $\alpha_{\mathrm{co}} \mathcal{L}_\circ(W)$. In particular, nonzero cycle circulation and $\alpha_{\mathrm{co}} > 0$ force positive average error for every $\tau \neq 0$.

Within each active component, $(Kz)_C = K_{CC} z_C$, so $z_C^\top K_{CC} z_C = 0$. Weighted balance of the positive and negative forcing gives

$$\|v_\tau(c)\|_2^2 \geq \tau^2 \sum_{C \in \mathcal{C}_E(z)} \theta_C(z) \, \|P_C K z\|_2^2 . \tag{23}$$

We take expectations and apply lemma 1 to obtain the theorem. The factor θ_C is the largest scalar valid for every rotation on that connected component. The directional form retains which part of the edge space K occupies; α_{co} supplies a common bound over all rotations.

ReLU is nonexpansive, so replacing K by K_F changes the root mean squared dual violation by at most $|\tau| \rho_F$. This estimate proves the corrected bound. The choice $F = E$ recovers $g_\nu(K)$, and the largest bound over any collection of retained graphs is also valid. The correction obeys $\rho_F \leq \sqrt{\beta} \, \|R_F\|_F$. If $\rho_F \leq \eta \sqrt{g_{\nu,F}(K_F)}$ for $0 \leq \eta < 1$, then $b_\nu(F) \geq (1-\eta)^2 g_{\nu,F}(K_F)$. Appendix A.8 gives the corresponding curvature bounds and a weakly connected example.

4.3 From error bounds to reliability conditions

The bounds become useful when they can accept or reject a proxy at a chosen error tolerance. An upper bound below that tolerance certifies accuracy; a lower bound above it rules the proxy out. Both depend on the proxy response and known weights, without requiring the original equilibrium. We now ask when the lower bound is positive and when it captures a substantial fraction of the error.

Positive co-activation ensures that every allowed rotation is detected, whereas $g_\nu(K)$ concerns only the rotation actually discarded. The error bounds allow mixed-sign weights and dependent inputs; only the cycle comparison requires reciprocal same-sign support. Positive co-activation can be checked from the proxy Gram or from the input law: proposition 12 constructs input balls with isolated active-pair responses for every $Q \succ 0$. Positive probability on each ball gives $\alpha_{\mathrm{co}} > 0$, including for every nonsingular correlated Gaussian law. The constant is determined by these probabilities.

Sparse graphs permit a size-independent guarantee. Suppose $S \geq 0$, $S_{ii} = 0$, $S\mathbf{1} \leq r\mathbf{1}$ with $r < 1$, and the support lies in a degree-d graph E. Independent inputs bounded above by b_0, with probabilities at least p_+ on $[a, b_0]$ and p_- below $-M$, where $0 < a < b_0$ and $M > r b_0/(1-r)$, give

$$\alpha_{\mathrm{co}} \geq \frac{a^2}{2} p_+^2 p_-^{2d-2}. \tag{24}$$

The input law may also assign mass to intermediate values, so the result includes uniform inputs on intervals that span zero. For two-interval inputs with $p_+ = 1/d$, $d \geq 2$, proposition 13 gives $\alpha_{\mathrm{co}} = \Omega(d^{-2})$ and $O(d^2 \log(|E|/\delta))$ samples for a positive empirical coefficient. On rotating cycles, this yields matching $\Theta(n)$ energy-error bounds for fixed coupling and input parameters. In contrast, uniform one-hot responses have second moment I/n, but sufficient inactive slack makes every proxy answer exact.

A positive lower bound detects error, but its size can still be small if active responses are highly unequal. To quantify this loss, write $B_{\mathrm{co}} = \tau^2 g_\nu(K)/(2\ell_\tau)$. For $\tau \neq 0$ and $\mathbb{E}\|Kz\|_2^2 > 0$, if components with at most p active units and response ratio at most r carry a fraction $a > 0$ of the mean squared forcing, then proposition 15 gives

$$\frac{B_{\mathrm{co}}}{\mathbb{E}\Delta_\tau} \geq \frac{a\mu}{\ell_\tau p r^2}. \tag{25}$$

The remaining components may be arbitrarily imbalanced. Disjoint rotating pairs attain a constant fraction independent of network size through deactivation. On a persistent face, the resolvent refinement in Appendix A.10 separates the activation and spectral losses; an unexcited soft direction need not amplify error.

These conditions assess a fixed proxy. Comparing coordinate choices also requires a common scale: scalar rescaling of D leaves $\widehat{x}_D$ unchanged but rescales Δ_D. We fix the product of the diagonal

entries to one on each support component and keep the original input distribution fixed. Uniform metric conditioning and uniform constants in theorem 2 then turn equation (22) into a common answer-error certificate. The next result shows what can happen when the metric is selected by rotational defect alone.

5 Why the Least-Defect Proxy Can Fail

The bounds suggest why geometric fit alone can be misleading: a metric can reduce rotation while flattening the energy in an excited direction. We now show that this tradeoff can cause unbounded answer error even when the original recurrence remains uniformly contractive. For a fixed original bias b, abbreviate the error in equation (11) by $e(D) = e_D(b)$. For $W = W_\varepsilon$, use $e_\varepsilon(D)$ and write $J_\varepsilon(D)$ for equation (5) with $u_i = \log D_{ii}$.

Theorem 3 (Unbounded error of the least-defect metric)**.** *There exist a nonnegative three-dimensional family W_ε, $0 < \varepsilon < 1$, a fixed positive unit-norm bias b, a constant $q < 1$, and a fixed positive diagonal $D_\star$ with $\mathrm{cond}_2(D_\star) < 3/2$, such that $\|W_\varepsilon\|_2 \le q$ for every ε and $D_\star$ uniquely minimizes rotational defect up to scalar multiplication. Both the identity-metric and least-defect proxies are strongly convex and fully active, but*

$$J_\varepsilon(D_\star) < J_\varepsilon(I), \qquad \sup_{0<\varepsilon<1} e_\varepsilon(I) < \infty,$$

$$e_\varepsilon(D_\star) = \frac{C_\star}{\varepsilon} + O(1) \quad \text{as } \varepsilon \downarrow 0, \qquad C_\star > 0. \tag{26}$$

The least-defect proxy has curvature margin exactly ε; the identity proxy has margin at least $1 - q$. For every fixed input law supported on $[b_-, b_+]^3$, where $0 < b_- < b_+ < \infty$, the directional lower bounds and energy-budget upper bounds in theorem 2 give

$$\mathbb{E}\Delta_\varepsilon = \Theta(\varepsilon^{-1}), \qquad \mathbb{E}e_\varepsilon(D_\star)^2 = \Theta(\varepsilon^{-2}) \quad \text{as } \varepsilon \downarrow 0. \tag{27}$$

If, in addition, $\mathbb{E}[bb^\top] \succ 0$, then $\inf_{0<\varepsilon<1} \alpha_{\mathrm{co},\varepsilon} > 0$.

The construction starts from the Frobenius-balanced integer matrix

$$B = \begin{pmatrix} 0 & 7 & 1 \\ 5 & 0 & 7 \\ 5 & 5 & 0 \end{pmatrix}, \qquad \rho = \rho(B) < \lambda = \lambda_{\max}\left(\frac{B + B^\top}{2}\right). \tag{28}$$

For positive right and left Perron vectors r, l, set $H = \mathrm{diag}(l_i/r_i)$ and $P = H^{1/2} B H^{-1/2}$. The scaling gives $\|P\|_2 = \rho$ and can be normalized so that $\det H = 1$ and $\mathrm{cond}_2(H) < 3/2$. Take

$$W_\varepsilon = \frac{1 - \varepsilon}{\lambda} P, \qquad D_\star = H^{-1}, \qquad b = \frac{\mathbf{1}}{\sqrt{3}}. \tag{29}$$

The original recurrence has contraction factor at most $q = \rho/\lambda < 1$, and its identity-metric proxy inherits curvature at least $1 - q$. In the least-defect coordinates the matrix is $(1 - \varepsilon)B/\lambda$, whose symmetric proxy has curvature ε in its positive leading eigendirection. The fixed positive bias excites this direction, so the proxy answer grows as ε^{-1} while the actual answer stays bounded. Figure 1 illustrates this separation. Appendix A.11 proves the spectral inequalities, optimality, conditioning, and the leading constant.

Thus a least-defect choice admits no uniform additive or multiplicative answer-error guarantee relative to the identity proxy, even under fixed contraction and metric-conditioning bounds. The population statement holds for uniform box inputs and correlated continuous laws, with the input distribution fixed as ε decreases. The same error certificate explains this failure: $g_\nu(K_\varepsilon) = \Theta(\varepsilon^{-2})$, $\ell_1 = \Theta(\varepsilon^{-1})$, and $R_\nu = \Theta(\varepsilon^{-1})$. Its energy bounds have the same order, as do its squared-state bounds. The energy-budget upper bound is asymptotically exact, while the directional lower bound retains a fixed positive fraction of the penalty. For input laws with nonsingular second moments, the uniform coefficient stays bounded above and away from zero. A bound based on the least-excited direction therefore loses the observed error-growth rate. Appendix A.11 derives the constants.

Even a single fixed positive input reveals this distinction. Its response has a nonzero skew annihilator, so $\alpha_{\mathrm{co}} = 0$ (equation (192)). Yet the actual rotation has positive directional excitation for all sufficiently small ε, and the directional certificate recovers the divergent rates. Detecting the relevant rotation does not require detecting every possible rotation.

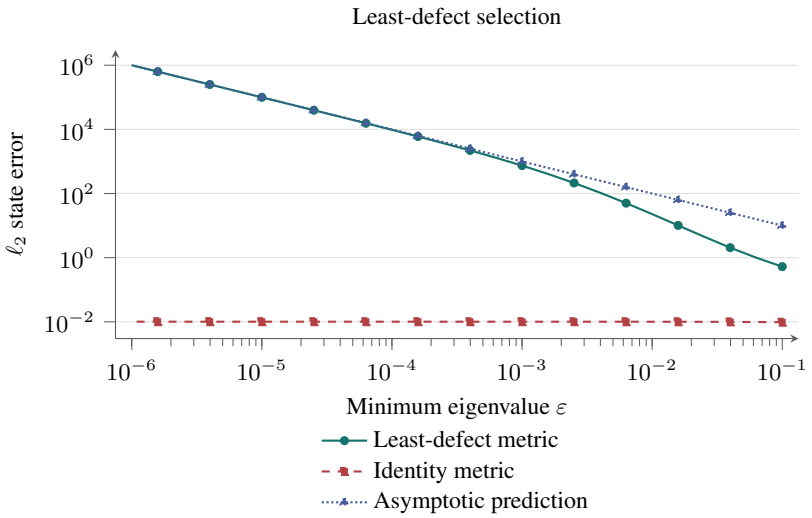


Figure 1: The least-defect proxy's ℓ_2 state error grows as ε^{-1}, where $\varepsilon = \lambda_{\min}(Q)$. The identity-proxy error remains bounded.

6 Experiments

The theory separates two questions: which quantities predict observed error, and which bounds certify its magnitude. Trained ReLU networks and a recurrent Transformer address the first; analytic families test the cross-face bounds and curvature-driven separation. Appendix B gives protocols, per-seed results, and numerical checks.

6.1 Trained ReLU equilibria

We first compare energy proxies for the same trained recurrence and inputs, so changes in answer error arise from the chosen metric. Three CIFAR-10 classifiers (Krizhevsky, 2009) use a convolutional encoder, a 256-dimensional ReLU equilibrium, and a linear head, with $\|W\|_2 \le 0.95$. Test accuracy at epoch 100 is 78.63–78.88%. At epochs 10, 30, 60, and 100, we compare 112 determinant-one metrics with condition number at most four, including identity and balancing. Sixteen additional metrics per checkpoint probe lower curvature. All 1,536 candidates have $Q \succ 0$ and $\mu \in [0.07046, 0.1511]$; Appendix B.1 specifies the candidate construction.

We hold the encoder inputs fixed across metrics and use 512 validation inputs for calibration and 1,024 test inputs for evaluation. We solve equilibria in 64-bit floating-point arithmetic (FP64). Normalized mean squared error (NMSE) is $\overline{e_D^2}/C_x$, where $C_x = \max\{\overline{\|z_I\|_2^2}, 10^{-12}\}$ uses the calibration identity proxy and is fixed across D. Appendix B.1.1 derives the original-coordinate interval $B_x \le e_D^2 \le U_x$ used in the predictor comparison.

At epoch 100, Spearman rank correlations between $\overline{\|Kz\|_2^2}$ and test mean squared error (MSE) are $0.987, 0.988, 0.992$, compared with $0.977, 0.954, 0.980$ for $\|K\|_F^2$. The improvement holds at all 12 checkpoints (Figure 2). Co-activation $\widehat{g}(K)$ gives $0.828, 0.833, 0.744$; the state upper bound gives $0.967, 0.953, 0.969$. Both predictors have lower correlations than forcing at all 12 checkpoints (Table 2).

6.2 Local response in a recurrent Transformer

We next test the directional mechanism in a recurrent Transformer, using a fixed-anchor intervention to obtain a positive-curvature setting for the local response comparison.

A public recurrent Transformer with 57.3 million parameters (Logan, 2026) converges on 126 of 128 contexts. Every converged context has a negative-curvature direction in $I - (J + J^\top)/2$, where

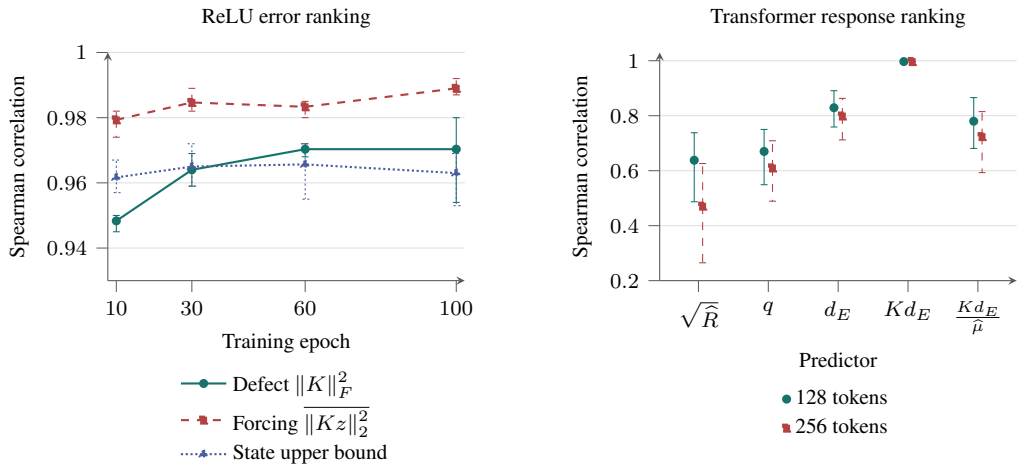


Figure 2: Input excitation and error prediction. Left: mean ReLU correlations across three training seeds; bars show their full range. The target is test MSE across 112 metrics. Right: recurrent-Transformer correlations with the L2 error of the symmetric response at perturbation size $a = 0.003$; bars are 95% bootstrap confidence intervals over documents. Vector labels denote Euclidean norms normalized by the square root of state dimension; $\widehat{R}$ estimates rotational defect.

$J = \partial_h T$. The fixed-anchor intervention $T_\gamma(h, e) = (1 - \gamma)h^* + \gamma T(h, e)$, with $\gamma = 1/4$, passes the convergence and estimated-curvature filters on 125 contexts. For an input perturbation δe, put $q = (\partial_e T)\delta e$, $S = (J + J^\top)/2$, $K = (J - J^\top)/2$, and $d_E = (I - \gamma S)^{-1}\gamma q$. The retained contexts have minimum eigenvalue estimates $\widehat{\mu} \geq 0.01$ for $I - \gamma S$. Excited rotation $\|Kd_E\|_2$ has Spearman correlation 0.997 with the L2 error of d_E at perturbation size $a = 0.003$, for both 128- and 256-token prefixes. Rotation magnitude gives 0.471–0.638, and symmetric response magnitude gives 0.799–0.829. Dividing excited rotation by the estimated curvature margin reduces correlation to 0.724–0.780. The full directional resolvent predicts this error vector with median relative L2 error 0.0130–0.0133% at $a = 0.003$, rising to 1.30–1.32% at 0.03 (Appendix B.2).

6.3 Analytic checks of the error bounds

The trained-model comparisons evaluate error ranking; the analytic families test the quantitative guarantees themselves. They verify cross-face bounds, inactive-unit slack, weak-edge corrections, and matching inverse-curvature rates (Figure 4).

The separation sweep uses 26 curvature values from 10^{-6} to 10^{-1} and a fixed metric with condition number 1.2767. At the smallest curvature, the least-defect state error is 9.996×10^5, while the identity proxy stays below 0.010076 throughout the sweep. For the six-input population, the energy-budget upper bound has relative error 0.034939% and the directional lower bound has relative error 70.648%. Both retain the predicted inverse-curvature rate, although their constants differ. The independent-pair experiment separates this curvature effect from graph size: component normalization keeps the same relative error as the number of pairs increases from one to 32 (Figure 5).

7 Conclusion

An energy proxy should preserve the answer, not merely approximate the update field. Our separation theorem shows that minimizing rotational defect can violate this requirement even for uniformly contractive dynamics and well-conditioned coordinates. The population bounds explain the failure through the rotation excited by the input and the curvature that converts it into error, with matching rates on the separating family. Co-activation identifies the forcing that survives inactive-unit slack, and the energy budget bounds its effect. The trained-network comparisons support the same directional distinction. Together, these results give a basis for evaluating energy proxies through the inputs they process and the answers they must preserve.

AI USE STATEMENT

We used generative AI tools to develop and check experimental code, improve the English presentation, prepare a Chinese reading version, audit mathematical derivations, and revise proof exposition. AI tools also assisted literature discovery and reference verification. The verification process combined explicit derivations, independent proof review, and tests of the analytic code against exact formulas and numerical bounds. These checks supplement the mathematical proofs rather than replace them. The authors take responsibility for the final text, mathematical claims, citations, code, and reported results.

REPRODUCIBILITY STATEMENT

Appendix A.1 through Appendix A.11 provide complete proofs. The anonymous code supplement generates the analytic experiments, including trial-level measurements, construction parameters, and summaries. Trained-network and Transformer results are retained as the reported aggregates in the manuscript; their full training artifacts and raw per-input records are not included in the code supplement. Appendix B gives protocols, dependencies, seeds, tolerances, and commands for reproducing the analytic experiments.

REFERENCES

Jason M. Altschuler. Flows, scaling, and entropy revisited: a unified perspective via optimizing joint distributions. *arXiv preprint arXiv:2210.16456*, 2022.

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Justin Baker, Qingsong Wang, Cory D. Hauck, and Bao Wang. Implicit graph neural networks: A monotone operator viewpoint. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 1521–1548, 2023.

Chen-Hao Chao, Wei-Fang Sun, Bo-Wun Cheng, and Chun-Yi Lee. On investigating the conservative property of score-based generative models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 4076–4095, 2023.

Xiaojun Chen and Shuhuang Xiang. Computation of error bounds for P-matrix linear complementarity problems. *Mathematical Programming*, 106(3):513–525, 2006.

Michael B. Cohen, Aleksander Madry, Dimitris Tsipras, and Adrian Vladu. Matrix scaling and balancing via box constrained Newton's method and interior point methods. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 902–913, 2017.

A. B. Duncan, T. Lelièvre, and G. A. Pavliotis. Variance reduction using nonreversible Langevin samplers. *Journal of Statistical Physics*, 163(3):457–491, 2016.

Fabio Durastante, Miryam Gnazzo, and Beatrice Meini. A Riemannian optimization approach for finding the nearest reversible Markov chain. *arXiv preprint arXiv:2505.16762*, 2025.

Daniel Durstewitz. A state space approach for piecewise-linear recurrent neural networks for identifying computational dynamics from neural measurements. *PLOS Computational Biology*, 13 (6):e1005542, 2017.

Gernot M. Engel and Hans Schneider. Matrices diagonally similar to a symmetric matrix. *Linear Algebra and its Applications*, 29:131–138, 1980.

Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*, volume I. Springer, 2003.

Angeliki Giannou, Shashank Rajput, Jy-Yong Sohn, Kangwook Lee, Jason D. Lee, and Dimitris Papailiopoulos. Looped transformers as programmable computers. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 11398–11442, 2023.

Fangda Gu, Heng Chang, Wenwu Zhu, Somayeh Sojoudi, and Laurent El Ghaoui. Implicit graph neural networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 11984–11995, 2020.

K. P. Hadeler and D. Kuhn. Stationary states of the Hartline–Ratliff model. *Biological Cybernetics*, 56(5–6):411–417, 1987.

Richard H. R. Hahnloser, H. Sebastian Seung, and Jean-Jacques Slotine. Permitted and forbidden sets in symmetric threshold-linear networks. *Neural Computation*, 15(3):621–638, 2003.

J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

Q. Jiang, M. Hlynka, P. H. Brill, and C. H. Cheung. Reversibility checking for markov chains. *arXiv preprint arXiv:1806.10154*, 2018.

Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial Hodge theory. *Mathematical Programming*, 127(1):203–244, 2011.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

Yingyu Liang, Zhizhou Sha, Zhenmei Shi, Zhao Song, and Yufa Zhou. Looped ReLU MLPs may be all you need as practical programmable computers. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 2647–2655, 2025.

Joe Logan. Per-token fixed-point convergence in depth-recurrent transformers. *arXiv preprint arXiv:2607.14427*, 2026.

Anton Lozhkov, Loubna Ben Allal, Leandro von Werra, and Thomas Wolf. FineWeb-Edu: the finest collection of educational content, 2024.

James McKee and Chris Smyth. Symmetrizable integer matrices having all their eigenvalues in the interval $[-2, 2]$. *Algebraic Combinatorics*, 3(3):775–789, 2020.

Erfan Nozari and Jorge Cortés. Stability analysis of complex networks with linear-threshold rate dynamics. In *2018 Annual American Control Conference (ACC)*, pp. 191–196, 2018.

E. E. Osborne. On pre-conditioning of matrices. *Journal of the ACM*, 7(4):338–345, 1960.

Rafail Ostrovsky, Yuval Rabani, and Arman Yousefi. Matrix balancing in L_p norms: Bounding the convergence rate of Osborne's iteration. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 154–169. Society for Industrial and Applied Mathematics, 2017.

Benjamin Scellier, Anirudh Goyal, Jonathan Binas, Thomas Mesnard, and Yoshua Bengio. Generalization of equilibrium propagation to vector field dynamics. *arXiv preprint arXiv:1808.04873*, 2018.

Russell Tsuchida, Suk Yee Yong, Mohammad Ali Armin, Lars Petersson, and Cheng Soon Ong. Declarative nets that are equilibrium models. In *International Conference on Learning Representations*, 2022.

Ezra Winston and J. Zico Kolter. Monotone operator equilibrium networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 10718–10728, 2020.

Yongyi Yang, Zengfeng Huang, and David P. Wipf. Transformers from an optimization perspective. In *Advances in Neural Information Processing Systems*, volume 35, pp. 36958–36971, 2022.

Yongyi Yang, Tang Liu, Yangkun Wang, Zengfeng Huang, and David Wipf. Implicit vs unfolded graph neural networks. *Journal of Machine Learning Research*, 26(82):1–46, 2025. arXiv:2111.06592.

A PROOFS OF THE ENERGY-PROXY RESULTS

The representation and balancing results identify the admissible coordinates and their unavoidable rotational defect. The proof of lemma 1 converts a dual-feasibility violation at the proxy into state and energy error. The component argument in theorem 2 then bounds this violation using proxy responses. The final construction applies the same population bounds to the metric selected by least rotational defect.

A.1 PROJECTION COMPATIBILITY AND FAITHFUL POTENTIALS

An exact optimization interpretation must reproduce both the affine preactivation and the ReLU projection. We impose these two requirements separately, which identifies the admissible metrics before determining the potential.

For $H \succ 0$, define the metric projection

$$P_{\mathcal{C}}^{H}(u) = \arg\min_{z \in \mathcal{C}} \frac{1}{2} \|z - u\|_H^2 . \tag{30}$$

Definition 4 (Generating potential)**.** A C^2 function f faithfully generates equation (1) in metric H if

$$x - H^{-1}\nabla f(x) = Wx + b, \qquad P_{\mathcal{C}}^{H}(u) = [u]_+ \quad \text{for all } x, u \in \mathbb{R}^n. \tag{31}$$

Then $T_{W,b}$ is exactly a unit metric projected-gradient step on $f + \iota_{\mathcal{C}}$.

For symmetric W, the generating objective is

$$\Phi(x) = \frac{1}{2} x^\top (I - W)x - b^\top x + \iota_{\mathcal{C}}(x). \tag{32}$$

Proposition 5 (Faithful metric representation)**.** *The recurrence $T_{W,b}$ has a faithful constant-metric representation if and only if a positive diagonal D satisfies $DW = W^\top D$. For each such D, the potential is unique up to an additive constant:*

$$f_D(x) = \frac{1}{2} x^\top D(I - W)x - b^\top Dx. \tag{33}$$

It is strongly convex if and only if $\lambda_{\max}(W) < 1$; all eigenvalues of W are real in this class. The stronger condition $\rho(W) < 1$ also makes the unit-step recurrence contractive in metric D.

Proposition 6 (Euclidean integrability)**.** *There exists $f \in C^2(\mathbb{R}^n)$ satisfying $x - \nabla f(x) = Wx + b$ for all x if and only if $W = W^\top$. The function is unique up to a constant and equals the smooth part of equation (32); it is convex exactly when $I - W \succeq 0$.*

Lemma 7 (ReLU-compatible metrics)**.** *For $H \succ 0$, the identity $P_{\mathcal{C}}^{H}(u) = [u]_+$ for every $u \in \mathbb{R}^n$ holds if and only if H is positive diagonal.*

Detailed proof of lemma 7. First suppose $H = \operatorname{diag}(h_1, \ldots, h_n)$ with every $h_i > 0$. The projection problem separates as

$$\min_{z_i \geq 0} \sum_{i=1}^{n} \frac{h_i}{2}(z_i - u_i)^2. \tag{34}$$

Each coordinate has the unique minimizer $z_i = \max\{u_i, 0\}$, so $P_{\mathcal{C}}^{H}(u) = [u]_+$.

Conversely, assume $P_{\mathcal{C}}^{H}(u) = [u]_+$ for every u. Fix distinct indices i and j, choose $u = e_i - e_j$, and set $z = [u]_+ = e_i$. The Karush–Kuhn–Tucker (KKT) optimality conditions for the strictly convex projection problem are

$$H(z - u) - \lambda = 0, \qquad z \geq 0, \quad \lambda \geq 0, \quad \lambda_k z_k = 0. \tag{35}$$

Because $z_i = 1 > 0$, complementary slackness gives $\lambda_i = 0$. Also $z - u = e_j$, so the ith stationarity coordinate gives $0 = [H(z - u)]_i = H_{ij}$. The ordered pair (i, j) was arbitrary; hence every off-diagonal entry vanishes. Positive definiteness then makes every diagonal entry positive. □

Proof of propositions 5 and 6. In Euclidean geometry, the preactivation identity forces $\nabla f(x) = (I - W)x - b$. Its constant Jacobian is $I - W$. Since the Hessian of a C^2 scalar function is symmetric, $W = W^\top$ is necessary. If $W = W^\top$, we can integrate the gradient to obtain $f(x) = \frac{1}{2}x^\top(I - W)x - b^\top x + C$, and its Hessian is positive semidefinite exactly when $I - W \succeq 0$.

For a general faithful constant metric H, lemma 7 first forces $H = D$ with D positive diagonal. The first identity in equation (31) is equivalent to

$$\nabla f(x) = D\big((I - W)x - b\big). \tag{36}$$

The Jacobian $D(I - W)$ is a Hessian exactly when $D(I - W) = (I - W)^\top D$, or equivalently $DW = W^\top D$. Under this equality, we integrate equation (36) to obtain equation (33). Its gradient verifies the converse. Any two such potentials have the same gradient and therefore differ by a constant.

Let $\widetilde{W} = D^{1/2} W D^{-1/2}$. From $DW = W^\top D$,

$$\widetilde{W}^\top = D^{-1/2} W^\top D^{1/2} = D^{1/2} W D^{-1/2} = \widetilde{W}. \tag{37}$$

Thus W is similar to a real symmetric matrix and has a real spectrum. The Hessian has the congruence factorization

$$D(I - W) = D^{1/2}(I - \widetilde{W})D^{1/2}. \tag{38}$$

Sylvester's law of inertia therefore gives $D(I - W) \succ 0$ if and only if every eigenvalue of $\widetilde{W}$ is smaller than one, i.e., if and only if $\lambda_{\max}(W) < 1$. If the stronger condition $\rho(W) < 1$ holds, then $\left\|\widetilde{W}\right\|_2 = \rho(W) < 1$. Positive diagonal scaling commutes with ReLU, so Euclidean nonexpansiveness of ReLU in the scaled coordinates proves

$$\|T_{W,b}(x) - T_{W,b}(y)\|_D \le \rho(W)\,\|x - y\|_D\,. \tag{39}$$

The unit-step estimate uses both ends of the spectrum through $\rho(W)$, whereas strong convexity depends on $\lambda_{\max}(W)$. □

A.2 Pairwise reciprocity and cycle balance

The representation theorem reduces the question to weighted symmetry. The next result shows when its pairwise scaling requirements fit together globally: the required ratios must agree along every cycle.

Theorem 8 (Cycle-balance characterization)**.** *A real matrix W admits a positive diagonal D satisfying equation (2) if and only if every off-diagonal pair has matching zero pattern and sign and, for every cycle $i_1, \dots, i_m, i_1$ in the bidirectional support graph,*

$$\prod_{k=1}^{m} w_{i_k i_{k+1}} = \prod_{k=1}^{m} w_{i_{k+1} i_k}, \qquad i_{m+1} = i_1. \tag{40}$$

Detailed proof of theorem 8. Suppose first that $D = \mathrm{diag}(d_i) > 0$ and $DW = W^\top D$. For $i \neq j$,

$$d_i w_{ij} = d_j w_{ji}. \tag{41}$$

Because both d_i and d_j are positive, one directed entry is zero exactly when its reverse is zero, and two nonzero reverse entries have the same sign. Dividing equation (41) gives $w_{ij}/w_{ji} = d_j/d_i$. The product of these ratios around a cycle equals one, which proves equation (40).

For the converse, form the undirected graph containing $\{i, j\}$ exactly when both reverse weights are nonzero. In each connected component choose a root r and set $d_r = 1$. For a path $r = i_0, i_1, \dots, i_k = j$, define

$$d_j = \prod_{\ell=0}^{k-1} \frac{w_{i_\ell i_{\ell+1}}}{w_{i_{\ell+1} i_\ell}}. \tag{42}$$

Every factor is positive by sign symmetry. If two paths lead from r to j, following the first and then the reverse of the second gives a closed walk. Any finite closed walk decomposes, after deleting

immediate backtracks and repeated-vertex segments, into simple cycles. The assumed cycle identity makes the product ratio on each simple cycle one, and a backtrack also contributes one. Hence the two products in equation (42) agree. The definition is therefore path-independent.

We extend a path to j across an edge j–!k to obtain $d_k/d_j = w_{jk}/w_{kj}$, hence $d_j w_{jk} = d_k w_{kj}$. Zero pairs satisfy the same equality trivially, diagonal entries impose no condition, and an isolated vertex may be assigned any positive d_i. Thus $DW = W^\top D$ on the whole matrix. □

A.3 Rotational defect, balancing, and duality

When cycle balance fails, no compatible metric removes all rotation. We now minimize the remaining defect and derive a dual formulation that certifies the geometric optimum.

Proposition 9 (Rotational defect and Frobenius balancing)**.** *Suppose W has reciprocal same-sign support. With the notation of equation (3), the defect and Frobenius norm satisfy equations (5) and (6). The Hessian of J is a positively weighted graph Laplacian. Consequently J has a unique minimizer modulo componentwise constants, and its minimum is zero exactly at cycle balance. Its dual and KKT conditions are equations (45) and (46).*

The dual follows the convex-flow formulation of matrix balancing (Altschuler, 2022). Direct differentiation and scalar conjugacy give

$$\nabla^2 J(u) = \mathsf{B}\,\mathrm{diag}\big(g_e^2 \cosh r_e(u)\big)\mathsf{B}^\top, \tag{43}$$

$$\phi_e^*(z) = z\ \mathrm{arsinh}\left(\frac{z}{g_e^2}\right) - \sqrt{g_e^4 + z^2} + g_e^2. \tag{44}$$

The exact paired-edge dual and its KKT equations are

$$\min_{u\in\mathbb{R}^n} J(u) = \max_{z:\,\mathsf{B}z=0} \left\{ c^\top z - \sum_{e\in E} \phi_e^*(z_e) \right\}, \tag{45}$$

$$z_e = g_e^2 \sinh r_e(u), \qquad \mathsf{B}z = 0. \tag{46}$$

Detailed proof of proposition 9. For an oriented edge $e = (i, j)$, write s_e for the common sign of w_{ij} and w_{ji}. The definitions of c_e and g_e give

$$w_{ij} = s_e g_e e^{c_e/2}, \qquad w_{ji} = s_e g_e e^{-c_e/2}. \tag{47}$$

Since $[\mathsf{B}^\top u]_e = u_j - u_i$, diagonal similarity yields

$$\widetilde{w}_{ij} = e^{(u_i-u_j)/2} w_{ij} = s_e g_e e^{r_e(u)/2}, \qquad \widetilde{w}_{ji} = s_e g_e e^{-r_e(u)/2}. \tag{48}$$

The two entries of K on this unordered pair are opposite and have magnitude $g_e|\sinh(r_e/2)|$. Their total squared contribution is therefore

$$2g_e^2 \sinh^2(r_e/2) = g_e^2(\cosh r_e - 1). \tag{49}$$

The sum over edges proves equation (5). Diagonal similarity preserves the diagonal entries, while the squares of the two off-diagonal entries in equation (48) sum to $2g_e^2 \cosh r_e$. This proves equation (6) and the exact equivalence with Frobenius matrix balancing.

Because $r = c - \mathsf{B}^\top u$, direct differentiation gives

$$\nabla J(u) = -\mathsf{B}\big(g_e^2 \sinh r_e(u)\big)_{e\in E}, \tag{50}$$

$$\nabla^2 J(u) = \mathsf{B}\,\mathrm{diag}\big(g_e^2 \cosh r_e(u)\big)\mathsf{B}^\top. \tag{51}$$

Every diagonal Hessian weight is strictly positive. Hence the Hessian kernel is exactly $\ker \mathsf{B}^\top$, the vectors constant on each connected component. Fix the gauge that u has zero mean on every component. On this subspace $\mathsf{B}^\top$ is injective, so finite-dimensional norm equivalence supplies $\sigma > 0$ with $\|\mathsf{B}^\top u\|_2 \geq \sigma \|u\|_2$. For $m = |E| > 0$,

$$\|c - \mathsf{B}^\top u\|_\infty \geq \frac{\sigma \|u\|_2 - \|c\|_2}{\sqrt{m}}. \tag{52}$$

Thus $\|r(u)\|_\infty \to \infty$ as the gauge-fixed norm of u tends to infinity. All g_e are positive, so $J(u) \to \infty$. Coercivity and strict convexity give existence and uniqueness on the gauge subspace. When E is empty, $J = 0$ and the gauge subspace contains only zero. Since every summand is nonnegative and vanishes only at zero, $J(u) = 0$ exactly when $c = \mathsf{B}^\top u$, which is the logarithmic cycle-balance condition.

It remains to derive the dual. Set $\phi_e(t) = g_e^2(\cosh t - 1)$. For fixed z, the first-order condition in the conjugate $\phi_e^*(z) = \sup_t\{zt - \phi_e(t)\}$ is $z = g_e^2 \sinh t$, so $t = \operatorname{arsinh}(z/g_e^2)$. We substitute this expression and use $\cosh(\operatorname{arsinh} a) = \sqrt{1 + a^2}$ to obtain equation (44).

Rewrite the primal as

$$\min_{r,u} \sum_e \phi_e(r_e) \quad \text{subject to} \quad c - r - \mathsf{B}^\top u = 0. \tag{53}$$

With multiplier z, its Lagrangian is

$$\sum_e \phi_e(r_e) + z^\top (c - r - \mathsf{B}^\top u). \tag{54}$$

The infimum over r is $-\sum_e \phi_e^*(z_e)$, and the infimum over u is finite exactly when $\mathsf{B}z = 0$. This gives the dual objective in equation (45). To verify attainment, take the primal minimizer $u_\star$ established above, set $r_\star = c - \mathsf{B}^\top u_\star$, and let $(z_\star)_e = g_e^2 \sinh((r_\star)_e)$. The gradient equation gives $\mathsf{B}z_\star = 0$. Equality in the scalar conjugacy relations yields

$$J(u_\star) = r_\star^\top z_\star - \sum_e \phi_e^*((z_\star)_e) = c^\top z_\star - \sum_e \phi_e^*((z_\star)_e). \tag{55}$$

Every feasible dual value is a lower bound by the same conjugacy inequality. Hence primal and dual optima are attained and equal. The equations in equation (46) are therefore necessary and sufficient for optimality. Since $\ker \mathsf{B}$ is the graph cycle space, the dual variable is a divergence-free cycle flow. □

A.4 The cycle lower bound

The nonlinear balancing problem identifies the best metric, but its minimum also admits a simpler lower bound. We express this bound directly through cycle circulations, which no diagonal scaling changes.

On one connected support component, write $G_E = \operatorname{diag}(g_e^2)$ and choose a full-row-rank cycle matrix $\mathsf{C}_\circ$ satisfying

$$\ker \mathsf{C}_\circ = \operatorname{im} \mathsf{B}^\top. \tag{56}$$

Its rows may be any basis of signed cycle vectors.

Lemma 10 (A coordinate-invariant cycle certificate)**.** *If the cycle space is nontrivial, the quadratic minimum in equation (7) equals*

$$\mathcal{L}_\circ(W) = \frac{1}{2}(\mathsf{C}_\circ c)^\top \left(\mathsf{C}_\circ G_E^{-1} \mathsf{C}_\circ^\top\right)^{-1} (\mathsf{C}_\circ c) \le J(u) \quad \textit{for every } u. \tag{57}$$

This value is invariant under positive diagonal similarity and the choice of cycle basis. On a tree it is zero, and $\min J = 0$*. For disconnected support, the certificates add over components.*

Detailed proof of lemma 10. Write $r = c - \mathsf{B}^\top u$ and $\Gamma = \mathsf{C}_\circ c$. Then $\mathsf{C}_\circ r = \Gamma$, while $\cosh t - 1 \ge t^2/2$ gives $J(u) \ge r^\top G_E r/2$. The Lagrangian of the quadratic minimization is $r^\top G_E r/2 - \lambda^\top(\mathsf{C}_\circ r - \Gamma)$. Stationarity gives $G_E r = \mathsf{C}_\circ^\top \lambda$; imposing the constraint then gives its unique minimizer

$$r = G_E^{-1} \mathsf{C}_\circ^\top (\mathsf{C}_\circ G_E^{-1} \mathsf{C}_\circ^\top)^{-1} \Gamma. \tag{58}$$

The inverse exists because $G_E \succ 0$ and $\mathsf{C}_\circ$ has full row rank. Substitution gives equation (57). Every feasible residual is attainable: $\mathsf{C}_\circ(c - r) = 0$ implies $c - r \in \operatorname{im} \mathsf{B}^\top$. The circulation Γ does not change when we add a gradient to c. The quadratic value also remains unchanged under a basis transformation $R\mathsf{C}_\circ$ with R invertible. On a tree, every edge field is a gradient, so $\min J = 0$. Disconnected components separate.

For local sharpness, fix g_e and perturb a balanced log field as $c_\varepsilon = \mathsf{B}^\top u_0 + \varepsilon h$. Let v minimize $\|h - \mathsf{B}^\top v\|_{G_E}^2$ and put $r_H = h - \mathsf{B}^\top v$. The cycle lower bound and evaluation at $u_0 + \varepsilon v$ give

$$\frac{\varepsilon^2}{2}\|r_H\|_{G_E}^2 \le \min_u J_\varepsilon(u) \le \sum_e g_e^2\big(\cosh(\varepsilon(r_H)_e) - 1\big) = \frac{\varepsilon^2}{2}\|r_H\|_{G_E}^2 + O(\varepsilon^4). \tag{59}$$

Thus the cycle certificate gives the leading defect near balance. □

A.5 Global equilibrium response

We now connect the discarded rotation to equilibrium error. Strong monotonicity first gives a unique equilibrium without assuming that the unit-step recurrence converges; complementarity then yields the state and energy bounds across activation changes.

Detailed proof of lemma 1. Write $F_\tau(y) = (Q - \tau K)y - c_D$. Since K is skew-symmetric,

$$\langle F_\tau(y) - F_\tau(v), y - v\rangle = (y - v)^\top Q(y - v) \ge \mu \|y - v\|_2^2 . \tag{60}$$

The operator is Lipschitz with constant $L_\tau = \|Q - \tau K\|_2$. For $0 < \eta < 2\mu/L_\tau^2$, nonexpansiveness of projection makes $P_{\mathcal{C}}(I - \eta F_\tau)$ a contraction, with squared factor at most $1 - 2\eta\mu + \eta^2 L_\tau^2 < 1$. Banach's theorem gives a unique fixed point on the closed orthant. Projection optimality identifies its fixed points, for every positive η, with the VI $\langle F_\tau(y), v - y\rangle \ge 0$ for all $v \in \mathcal{C}$. At $\eta = 1$, $y - F_\tau(y) = (S + \tau K)y + c_D$, so this VI is exactly the fixed-point equation for $\mathcal{T}_\tau$. At $\tau = 0$, the VI is optimality for the coercive strongly convex quadratic Ψ_S, so $z^\star$ is its unique minimizer.

Let $\delta = y_\tau^\star - z^\star$. The two VI comparison inequalities give

$$\langle F_\tau(y_\tau^\star), \delta\rangle \le 0, \qquad \langle F_0(z^\star), \delta\rangle \ge 0. \tag{61}$$

The difference between these inequalities gives $\delta^\top Q\delta \le \tau\langle Ky_\tau^\star, \delta\rangle$. Since $y_\tau^\star = z^\star + \delta$ and $\langle K\delta, \delta\rangle = 0$, this becomes

$$\mu \|\delta\|_2^2 \le \delta^\top Q\delta \le \tau \langle Kz^\star, \delta\rangle . \tag{62}$$

Cauchy–Schwarz proves the input-sensitive state upper bound, including $\delta = 0$. Complementarity at $z^\star$ also gives

$$\mu \|z^\star\|_2^2 \le (z^\star)^\top Qz^\star = c_D^\top z^\star \le c_0 \|z^\star\|_2 . \tag{63}$$

Thus $\|z^\star\|_2 \le c_0/\mu$, proving the second bound.

Both equilibria are feasible, so we expand the quadratic at $y_\tau^\star$ to obtain

$$\begin{aligned}\Psi_S(y_\tau^\star) - \Psi_S(z^\star) &= \langle F_0(y_\tau^\star), \delta\rangle - \frac{1}{2}\delta^\top Q\delta \\ &\le \tau\langle Ky_\tau^\star, \delta\rangle - \frac{1}{2}\delta^\top Q\delta = \tau\langle Kz^\star, \delta\rangle - \frac{1}{2}\delta^\top Q\delta \\ &\le \frac{\tau^2}{2}(Kz^\star)^\top Q^{-1}(Kz^\star).\end{aligned} \tag{64}$$

The last step maximizes the concave quadratic over all $\delta \in \mathbb{R}^n$. For the lower bounds, write $s^\star = Qz^\star - c_D$ and $t^\star = (Q - \tau K)y_\tau^\star - c_D$. Complementarity gives $s^\star, t^\star \ge 0$, $s^\star \perp z^\star$, and $t^\star \perp y_\tau^\star$. Therefore

$$(Q - \tau K)\delta = \tau Kz^\star - s^\star + t^\star, \qquad 0 \le v_\tau \le [(Q - \tau K)\delta]_+. \tag{65}$$

This proves $\|v_\tau\|_2 \le L_\tau \|\delta\|_2$. The energy expansion at $z^\star$ is

$$\Psi_S(y_\tau^\star) - \Psi_S(z^\star) = (s^\star)^\top y_\tau^\star + \frac{1}{2}\delta^\top Q\delta \ge \frac{1}{2}\delta^\top Q\delta. \tag{66}$$

Complementarity also gives $(z^\star)^\top Qz^\star = c_D^\top z^\star$ and $(y_\tau^\star)^\top Qy_\tau^\star = c_D^\top y_\tau^\star$, since $(y_\tau^\star)^\top Ky_\tau^\star = 0$. Thus the proxy energy at each equilibrium equals minus half its squared Q-norm, which proves equation (17). Positive definiteness gives $\Delta_\tau \le (z^\star)^\top Qz^\star/2$. Together with equation (66), this yields $\|\delta\|_2^2 \le (z^\star)^\top Qz^\star/\mu$. The minimum of these bounds and their perturbative counterparts gives both upper bounds in the lemma. To retain the curvature of the quadratic, observe that

$$(Q - \tau K)Q^{-1}(Q - \tau K)^\top = Q - \tau^2 KQ^{-1}K \succ 0. \tag{67}$$

With $A_\tau = (Q - \tau K)Q^{-1/2}$, the matrix on the left of equation (67) is $A_\tau A_\tau^\top$. Consequently $\|A_\tau\|_2^2 = \ell_\tau$ and

$$\|v_\tau\|_2^2 \le \left\|A_\tau Q^{1/2}\delta\right\|_2^2 \le \ell_\tau \delta^\top Q \delta \le 2\ell_\tau \Delta_\tau. \tag{68}$$

This gives the energy lower bound. Also $-KQ^{-1}K = KQ^{-1}K^\top \preceq (\|K\|_2^2/\mu)I$, proving $\ell_\tau \le \lambda_{\max}(Q) + \tau^2 \|K\|_2^2/\mu$.

If $v_\tau = 0$, then $(Q - \tau K)z^\star - c_D \ge 0$, and

$$(z^\star)^\top \big((Q - \tau K)z^\star - c_D\big) = (z^\star)^\top s^\star - \tau (z^\star)^\top K z^\star = 0. \tag{69}$$

Thus $z^\star$ also solves the actual complementarity problem. Uniqueness gives $y_\tau^\star = z^\star$. The converse follows from actual dual feasibility. When $Kz^\star = 0$, $v_\tau = 0$ for every τ. □

The energy lower-bound constant $1/2$ is sharp even after an active unit turns off. Take $Q = I_2$, $K = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, and $c_D = (1,1)^\top$. The proxy is $z^\star = (1,1)^\top$. For $\tau \ge 1$ the actual solution is $y_\tau^\star = (1,0)^\top$, so

$$\Delta_\tau = \frac{1}{2}, \qquad \|v_\tau\|_2^2 = \tau^2, \qquad \ell_\tau = L_\tau^2 = 1 + \tau^2. \tag{70}$$

Both the energy and squared-state lower bounds approach equality as $\tau \to \infty$.

A.6 Exact active-face rotational response

The preceding bounds allow the active set to change. When strict complementarity keeps that set fixed, we can solve for the error exactly and identify the interaction between rotation and directional curvature.

Proposition 11 (Exact active-face response to rotation)**.** *Assume $S = S^\top$, $Q = I - S \succ 0$, and $K = -K^\top$. Let $z^\star$ be the fixed point of $\mathcal{S}(y) = [Sy + c_D]_+$, set $I = \operatorname{supp}(z^\star)$ and $J = I^c$, and assume strict complementarity:*

$$\gamma = \min\left\{\min_{i\in I} z_i^\star, \min_{j\in J} -(Sz^\star + c_D)_j\right\} > 0, \tag{71}$$

where a minimum over an empty set is $+\infty$. The unique equilibrium of equation (12) has the same support I for all sufficiently small $|\tau|$. Define $Q_I = \mathbf{I}_{|I|} - S_{II} \succ 0$. Throughout any parameter range on which this support persists, with the natural empty-block convention,

$$(y_\tau^\star)_J = 0, \qquad (y_\tau^\star)_I = (Q_I - \tau K_{II})^{-1}(c_D)_I, \tag{72}$$

$$(y_\tau^\star - z^\star)_I = \tau (Q_I - \tau K_{II})^{-1} K_{II} z_I^\star, \tag{73}$$

$$\Psi_S(y_\tau^\star) - \Psi_S(z^\star) = \frac{1}{2}(y_\tau^\star - z^\star)_I^\top Q_I (y_\tau^\star - z^\star)_I. \tag{74}$$

Let $\overline{K}_I = Q_I^{-1/2} K_{II} Q_I^{-1/2}$ and $w_I = Q_I^{1/2} z_I^\star$. Then

$$\Psi_S(y_\tau^\star) - \Psi_S(z^\star) = \frac{\tau^2}{2} w_I^\top \overline{K}_I^\top (\mathbf{I} - \tau^2 \overline{K}_I^2)^{-1} \overline{K}_I w_I. \tag{75}$$

If $K_{II} z_I^\star = 0$, then $y_\tau^\star = z^\star$ throughout this parameter range. In particular, a skew perturbation is locally invisible to the equilibrium whenever at most one unit is active.

The exact identities imply, as $\tau \to 0$,

$$(y_\tau^\star - z^\star)_I = \tau Q_I^{-1} K_{II} z_I^\star + \tau^2 Q_I^{-1} K_{II} Q_I^{-1} K_{II} z_I^\star + O(\tau^3), \tag{76}$$

$$\Psi_S(y_\tau^\star) - \Psi_S(z^\star) = \frac{\tau^2}{2} \left\|\overline{K}_I w_I\right\|_2^2 - \frac{\tau^4}{2} \left\|\overline{K}_I^2 w_I\right\|_2^2 + O(\tau^6). \tag{77}$$

Detailed proof of proposition 11. Set $\mu = \lambda_{\min}(Q) > 0$. Because $Q = I - S \succeq \mu I$, its principal submatrix satisfies

$$Q_I = \mathbf{I}_{|I|} - S_{II} \succeq \mu \mathbf{I}_{|I|} \succ 0. \tag{78}$$

At the proxy fixed point, positivity on I makes ReLU the identity and zero coordinates on J remove all cross-face state terms. Thus

$$Q_I z_I^\star = (c_D)_I, \qquad z_J^\star = 0, \tag{79}$$

while strict complementarity says

$$S_{JI} z_I^\star + (c_D)_J < 0 \tag{80}$$

coordinatewise.

For every real τ, $Q_I - \tau K_{II}$ is invertible: a nonzero kernel vector v would give $0 = v^\top (Q_I - \tau K_{II}) v = v^\top Q_I v > 0$. Define the candidate

$$\widehat{y}_I(\tau) = (Q_I - \tau K_{II})^{-1} (c_D)_I, \qquad \widehat{y}_J(\tau) = 0. \tag{81}$$

This candidate is continuous at zero and $\widehat{y}_I(0) = z_I^\star > 0$. Therefore all its active coordinates remain positive for sufficiently small $|\tau|$. Its inactive preactivation is the continuous function

$$h_J(\tau) = (S_{JI} + \tau K_{JI}) \widehat{y}_I(\tau) + (c_D)_J. \tag{82}$$

Equation (80) gives $h_J(0) < 0$, so every inactive coordinate remains strictly negative after possibly shrinking the same neighborhood. Consequently

$$\widehat{y}(\tau) = [(S + \tau K)\widehat{y}(\tau) + c_D]_+. \tag{83}$$

Uniqueness from lemma 1 identifies this candidate with $y_\tau^\star$. More generally, the same identity holds for every real τ for which $\widehat{y}_I(\tau) > 0$ and $h_J(\tau) \le 0$: these explicit sign checks certify persistence of the equilibrium support. This proves equation (72). If I is empty, the same argument reduces to the strict inequality $c_D < 0$ and gives $y_\tau^\star = z^\star = 0$; all empty-block formulas below are then vacuous.

For nonempty I, combine equation (79) and equation (72):

$$\begin{aligned} (Q_I - \tau K_{II})\big((y_\tau^\star)_I - z_I^\star\big) &= Q_I z_I^\star - (Q_I - \tau K_{II}) z_I^\star \\ &= \tau K_{II} z_I^\star. \end{aligned} \tag{84}$$

Multiplication by the inverse gives equation (73).

It remains to prove the exact energy identity. Every vector v on this face has $v_J = 0$, so equation (10) restricts to

$$\Psi_S(v) = \frac{1}{2} v_I^\top Q_I v_I - (c_D)_I^\top v_I. \tag{85}$$

Set $\delta_I = (y_\tau^\star)_I - z_I^\star$ and use $(c_D)_I = Q_I z_I^\star$. Expanding equation (85) at $z_I^\star$ gives

$$\begin{aligned} \Psi_S(y_\tau^\star) - \Psi_S(z^\star) &= \frac{1}{2} \delta_I^\top Q_I \delta_I + (z_I^\star)^\top Q_I \delta_I - (c_D)_I^\top \delta_I \\ &= \frac{1}{2} \delta_I^\top Q_I \delta_I, \end{aligned} \tag{86}$$

which proves equation (74).

For the sharper energy representation, define $\overline{K}_I = Q_I^{-1/2} K_{II} Q_I^{-1/2}$ and $w_I = Q_I^{1/2} z_I^\star$. Congruence preserves skew-symmetry:

$$\overline{K}_I^\top = Q_I^{-1/2} K_{II}^\top Q_I^{-1/2} = -\overline{K}_I. \tag{87}$$

Moreover,

$$Q_I - \tau K_{II} = Q_I^{1/2} (\mathbf{I} - \tau \overline{K}_I) Q_I^{1/2}. \tag{88}$$

We apply this factorization to equation (73) and obtain

$$Q_I^{1/2} \delta_I = \tau (\mathbf{I} - \tau \overline{K}_I)^{-1} \overline{K}_I w_I. \tag{89}$$

The exact quadratic energy identity and $[(\mathbf{I} - \tau \overline{K}_I)^{-1}]^\top = (\mathbf{I} + \tau \overline{K}_I)^{-1}$ now imply

$$\begin{aligned} \Psi_S(y_\tau^\star) - \Psi_S(z^\star) &= \frac{\tau^2}{2} w_I^\top \overline{K}_I^\top (\mathbf{I} + \tau \overline{K}_I)^{-1} (\mathbf{I} - \tau \overline{K}_I)^{-1} \overline{K}_I w_I \\ &= \frac{\tau^2}{2} w_I^\top \overline{K}_I^\top (\mathbf{I} - \tau^2 \overline{K}_I^2)^{-1} \overline{K}_I w_I. \end{aligned} \tag{90}$$

The inverse factors are rational functions of the same matrix, so they commute. This proves equation (75). Also, $\mathbf{I} - \tau^2 \overline{K}_I^2 \succ 0$ for every real τ, because $-\overline{K}_I^2 \succeq 0$.

For the state expansion, factor the unwhitened resolvent as

$$(Q_I - \tau K_{II})^{-1} = (\mathbf{I} - \tau Q_I^{-1} K_{II})^{-1} Q_I^{-1}. \tag{91}$$

For $|\tau| \left\| Q_I^{-1} K_{II} \right\|_2 < 1$, its Neumann series gives

$$(Q_I - \tau K_{II})^{-1} = Q_I^{-1} + \tau Q_I^{-1} K_{II} Q_I^{-1} + O(\tau^2), \tag{92}$$

which proves equation (76). Finally,

$$(\mathbf{I} - \tau^2 \overline{K}_I^2)^{-1} = \mathbf{I} + \tau^2 \overline{K}_I^2 + O(\tau^4). \tag{93}$$

Skew-symmetry, $\overline{K}_I^\top = -\overline{K}_I$, gives

$$w_I^\top \overline{K}_I^\top \overline{K}_I w_I = \left\| \overline{K}_I w_I \right\|_2^2, \tag{94}$$

$$w_I^\top \overline{K}_I^\top \overline{K}_I^2 \overline{K}_I w_I = -\left\| \overline{K}_I^2 w_I \right\|_2^2. \tag{95}$$

Substitution proves equation (77), including the absence of odd powers and the $O(\tau^6)$ remainder. If $K_{II} z_I^\star = 0$, the right side of equation (84) is zero. Hence $y_\tau^\star = z^\star$ throughout every parameter range on which the support persists. For $|I| \le 1$, skew-symmetry forces $K_{II} = 0$, completing the proof. □

The exact law also yields a quantitative excitation criterion. On a persistent face, let $a_I = \left\| \overline{K}_I w_I \right\|_2$ and $k_I = \left\| \overline{K}_I \right\|_2$. Since $0 \preceq -\overline{K}_I^2 \preceq k_I^2 I$, spectral bounds in equation (75) give

$$\frac{\tau^2 a_I^2}{2(1 + \tau^2 k_I^2)} \le \Psi_S(y_\tau^\star) - \Psi_S(z^\star) \le \frac{\tau^2 a_I^2}{2}. \tag{96}$$

Thus nonzero excitation on the active set forces a positive penalty. These bounds agree to leading order as $\tau \to 0$. When every coordinate is active, this is the perturbative term in equation (16). If $Kz^\star \neq 0$, the ratio of actual energy penalty to that bound tends to one as $\tau \to 0$. Thus its leading constant $1/2$ cannot be uniformly reduced, even on strict full faces. Empty blocks have $a_I = k_I = 0$ by convention.

A.7 Co-activation and population bounds across faces

The global lower bound depends on the forcing that survives inactive slack. Within a jointly active component, skew-symmetry balances positive and negative forcing; we use this identity to obtain a population bound without requiring a common activation face.

Proof of theorem 2. All random variables below are functions of the bias c. The VI comparison for two biases gives $\|z(c) - z(c')\|_2 \le \|c - c'\|_2 / \mu$, so the proxy response is measurable. The same estimate holds for $y_\tau(c)$. Each component weight is continuous on its fixed active set, and there are finitely many such sets, so the weights are measurable as well. Let $\Xi = \mathbb{E}_\nu[zz^\top] \preceq \beta I$. Then $\mathbb{E}\|z\|_2^2 = \operatorname{tr} \Xi \le n\beta$ and $R_\nu \le \lambda_{\max}(Q) n\beta < \infty$.

Fix a proxy response z and a component $C \in \mathcal{C}_E(z)$. Edges from C to other active vertices are absent, and the response vanishes at inactive vertices. Hence $h = (Kz)_C = K_{CC} z_C$ and $z_C^\top h = 0$. If $h \neq 0$, its positive and negative parts h_+, h_- are both nonzero. Write $A = \operatorname{supp}(h_+)$, $B = \operatorname{supp}(h_-)$, and $m_B = \min_{i \in B} z_i > 0$. Weighted balance and Cauchy–Schwarz give

$$m_B \|h_-\|_2 \le z_C^\top h_- = z_C^\top h_+ \le \|z_A\|_2 \|h_+\|_2 . \tag{97}$$

The leftmost bound uses $z_C^\top h_- \ge m_B \|h_-\|_1 \ge m_B \|h_-\|_2$. Since A and B are disjoint, $\|z_A\|_2^2 \le \|z_C\|_2^2 - m_B^2$. It follows that

$$\|h\|_2^2 = \|h_+\|_2^2 + \|h_-\|_2^2 \le \frac{\|z_C\|_2^2}{m_B^2} \|h_+\|_2^2 \le \frac{\|h_+\|_2^2}{\theta_C(z)}. \tag{98}$$

The same argument applies to $-h$, and the inequality also holds for $h = 0$. Proxy complementarity gives $s_C = 0$. Apply the inequality to each C with either sign of τ and sum the squared norms over these disjoint components to obtain equation (23). If there are no nontrivial active components, its right side is zero.

If $K \neq 0$, use $H = K/\|K\|_F$ in equation (19) to obtain

$$\mathbb{E}_\nu \|v_\tau\|_2^2 \geq \tau^2 g_\nu(K) \geq \tau^2 \alpha_{\mathrm{co}} \|K\|_F^2 . \tag{99}$$

The energy and state lower bounds in lemma 1 now give the claimed lower bounds, using $\left\|D^{-1/2}(y_\tau - z)\right\|_2^2 \geq \|y_\tau - z\|_2^2 / d_{\max}$.

For the weak-edge correction, put $v_{\tau,F} = [\tau K_F z - s]_+$. Since $s \geq 0$, their norms are bounded by $|\tau| \|Kz\|_2$ and $|\tau| \|K_F z\|_2$, respectively. Thus both violations belong to $L^2(\nu; \mathbb{R}^n)$ under the same second-moment assumption. The component argument on F gives $\mathbb{E}\|v_{\tau,F}\|_2^2 \geq \tau^2 g_{\nu,F}(K_F)$. Nonexpansiveness gives $\|v_\tau - v_{\tau,F}\|_2 \leq |\tau| \|R_F z\|_2$ for each input. The reverse triangle inequality in $L^2(\nu; \mathbb{R}^n)$ therefore yields

$$\left(\mathbb{E}\|v_\tau\|_2^2\right)^{1/2} \geq \left(\left(\mathbb{E}\|v_{\tau,F}\|_2^2\right)^{1/2} - |\tau|\rho_F\right)_+ \geq |\tau|\sqrt{b_\nu(F)}. \tag{100}$$

Applying the same state and energy inequalities proves the corrected lower bounds. They hold simultaneously for every $F \subseteq E$.

For the upper bounds, $\Xi \preceq \beta I$ implies

$$\mathbb{E}_\nu (Kz)^\top Q^{-1}(Kz) = \mathrm{tr}(K^\top Q^{-1} K \Xi) \leq \frac{\beta}{\mu}\|K\|_F^2 , \qquad \mathbb{E}_\nu \|Kz\|_2^2 \leq \beta \|K\|_F^2 . \tag{101}$$

Apply the global upper bounds and $\left\|D^{-1/2}(y_\tau - z)\right\|_2^2 \leq \|y_\tau - z\|_2^2 / d_{\min}$. The energy identity also gives $\mathbb{E}\Delta_\tau \leq R_\nu/2$ and $\mathbb{E}e_{\tau,D}^2 \leq R_\nu/(\mu d_{\min})$. The smaller upper bound in each case gives the stated result. All expectations are finite by the second-moment assumption and the pointwise upper bounds. If $K = 0$, both equilibria coincide and every bound is zero. Finally, lemma 10 substitutes $\mathcal{L}_\circ(W)$ for $\|K\|_F^2 = J(u)$ in either lower bound. Here W is the base matrix with scaled decomposition $S + K$. □

To compute α_{co}, choose the Frobenius-orthonormal basis $H_{ij} = (e_i e_j^\top - e_j e_i^\top)/\sqrt{2}$ for $\{i, j\} \in E$, $i < j$. Let $\mathsf{A}_C(z)$ have columns $P_C H_{ij} z$. Then

$$G_\nu = \mathbb{E}_\nu \left[\sum_{C \in \mathcal{C}_E(z)} \theta_C(z) \mathsf{A}_C(z)^\top \mathsf{A}_C(z) \right], \qquad \alpha_{\mathrm{co}} = \lambda_{\min}(G_\nu). \tag{102}$$

Indeed, $H = \sum_e a_e H_e$ has $\|H\|_F^2 = \|a\|_2^2$, and the quadratic form in equation (19) is $a^\top G_\nu a$. This matrix depends only on the proxy responses and the chosen edge space.

A.8 Applicability and sharpness of co-activation

The population proof uses a separate normalization on each active component. We explain why this local choice matters, how to handle weak connections, and why its scalar weight cannot be uniformly improved.

The component normalization separates independent rotational subsystems. Adding an isolated vertex to E, with any nonnegative response, leaves every term of G_ν unchanged. More generally, the Gram matrix is block diagonal across the connected components of E. Its smallest eigenvalue is the minimum of the component eigenvalues, with no factor for the number of components.

For comparison, define the global weight $\theta_{\mathrm{g}}(z) = \min_{i \in I(z)} z_i^2 / \|z\|_2^2$ for $z \neq 0$, and set $\theta_{\mathrm{g}}(0) = 0$. Replacing every $\theta_C(z)$ by $\theta_{\mathrm{g}}(z)$ in equation (102) gives a Gram matrix G_{g}. Since $\theta_C(z) \geq \theta_{\mathrm{g}}(z)$ on every active component,

$$G_\nu \succeq G_{\mathrm{g}}, \qquad \alpha_{\mathrm{co}} \geq \lambda_{\min}(G_{\mathrm{g}}). \tag{103}$$

Global normalization can lose an arbitrarily large factor on a fixed rotational subsystem. Use the three-dimensional S, K in equation (132), append an isolated fourth coordinate with $Q_{44} = 1$, and give equal mass to $z_i = (\mathbf{1} + e_i/4, \eta)$, $i = 1, 2, 3$, with $c_i = Qz_i$. At $\tau = 1$, the fourth coordinate has zero displacement and

$$\mathbb{E}\Delta_1 = \frac{1}{1960}, \qquad \alpha_{\mathrm{co}} = \frac{1}{171}, \qquad \lambda_{\min}(G_{\mathrm{g}}) = \frac{\eta^2}{48(57/16 + \eta^2)} \quad (0 < \eta \le 1). \tag{104}$$

The local Gram matrix stays constant as η vanishes. Write $\bar{z}_i = \mathbf{1} + e_i/4$ for the three rotating coordinates. Then $\|\bar{z}_i\|_2^2 = 57/16$ and their second moment is $\mathbb{E}[\bar{z}\bar{z}^\top] = (7/6)\mathbf{1}\mathbf{1}^\top + I/48$. In dimension three, write $Hx = a \times x/\sqrt{2}$, so that $\|H\|_F = \|a\|_2$ and $\|Hx\|_2^2 = (\|a\|_2^2 \|x\|_2^2 - (a^\top x)^2)/2$. The expectation identifies the unweighted edge Gram, up to an orthogonal change of basis, as $(\operatorname{tr}\Xi_3 I - \Xi_3)/2$, where $\Xi_3 = \mathbb{E}[\bar{z}\bar{z}^\top]$. The eigenvalues of Ξ_3 are $169/48, 1/48, 1/48$, so the unweighted edge Gram eigenvalues are $1/48, 85/48, 85/48$. The local coefficient equals $16/57$ times the smallest of these eigenvalues; the global coefficient instead uses the factor $\eta^2/(57/16 + \eta^2)$. On $\mathbf{1}^\perp$, $Q = (6/5)I$ and $K^\top K = (3/100)I$, while $K\mathbf{1} = 0$. Therefore $\mathbb{E}\,\|K\bar{z}\|_2^2 = (3/100)(2/48) = 1/800$. The full-face energy formula gives

$$\mathbb{E}\Delta_1 = \frac{\mathbb{E}\,\|K\bar{z}\|_2^2}{2(6/5)(1 + 1/48)} = \frac{1}{1960}. \tag{105}$$

For weak edges between components, the remainder in equation (20) satisfies

$$\rho_F^2 = \operatorname{tr}(R_F \Xi R_F^\top) \le \beta \|R_F\|_F^2, \qquad \Xi = \mathbb{E}_\nu[zz^\top]. \tag{106}$$

The two lower-bound denominators also vary continuously with the omitted weights. Let $r_F = \|R_F\|_2$, and let $\ell_{\tau,F}, L_{\tau,F}$ be equation (13) evaluated at K_F. Expanding $K = K_F + R_F$ gives

$$-KQ^{-1}K + K_F Q^{-1} K_F = -K_F Q^{-1} R_F - R_F Q^{-1} K_F - R_F Q^{-1} R_F. \tag{107}$$

Its operator norm is at most $(2\|K_F\|_2 r_F + r_F^2)/\mu$. The eigenvalue perturbation inequality $|\lambda_{\max}(A) - \lambda_{\max}(B)| \le \|A - B\|_2$ for symmetric A, B, and the reverse triangle inequality for the operator norm, yield

$$|\ell_\tau - \ell_{\tau,F}| \le \frac{\tau^2}{\mu}(2\|K_F\|_2 r_F + r_F^2), \qquad |L_\tau - L_{\tau,F}| \le |\tau| r_F. \tag{108}$$

For fixed Q, ν, K_F and $g_{\nu,F}(K_F) > 0$, take $R_F = \delta R_0$ supported outside F. The proxy responses are fixed along this perturbation. If $\rho_0^2 = \mathbb{E}\,\|R_0 z\|_2^2$, then $\rho_F = |\delta|\rho_0$ and $b_\nu(F) = (\sqrt{g_{\nu,F}(K_F)} - |\delta|\rho_0)_+^2$. The corrected numerator thus converges to $g_{\nu,F}(K_F)$. Equation (108) gives convergence of both denominators, which are positive because $Q \succ 0$. The corrected lower bounds therefore converge to the original component bounds as $\delta \to 0$. The actual equilibria may change activation face along this limit.

The isolated-coordinate example gives an explicit comparison when the symmetric weights also change. Write S_0, K_0 for its four-dimensional matrices and retain the three triangle edges as F. For $\delta > 0$, set

$$S_\delta = S_0 + 2\delta(e_3 e_4^\top + e_4 e_3^\top), \qquad K_\delta = K_0 + \delta(e_3 e_4^\top - e_4 e_3^\top). \tag{109}$$

Keep the inputs of equation (104) fixed, with $0 < \eta < 1$ and $i = 1, 2, 3$. At $\tau = 1$, the new reciprocal weights are 3δ and δ. For $0 \le \delta < 1/5$, both S_δ and $S_\delta + K_\delta$ are nonnegative, with maximum row sums $2/5 + 2\delta$ and $2/5 + 3\delta$, respectively, both below one. The fixed biases are strictly positive. Their Neumann-series solutions are therefore fully active, and $Q_\delta = I - S_\delta \succeq (3/5 - 2\delta)I$. For each fixed bias, the resolvent identity gives

$$\|z_\delta - z_0\|_2 \le \frac{2\delta\|z_0\|_2}{3/5 - 2\delta}, \qquad \|y_\delta - y_0\|_2 \le \frac{3\delta\|y_0\|_2}{3/5 - 2\delta}. \tag{110}$$

Here the inverse norm of $Q_\delta - K_\delta$ is at most $(3/5 - 2\delta)^{-1}$ by strong monotonicity. The responses and their mean energy gap converge to the disconnected values. Also $\rho_F^2 = \delta^2\mathbb{E}[(z_\delta)_3^2 + (z_\delta)_4^2] \to 0$. Write Δ_δ for the energy gap of this family at $\tau = 1$. Let B_δ^{full} use the full-graph directional lower bound and B_δ^{corr} use $b_\nu(F)/(2\ell_1)$. Then

$$\lim_{\delta \downarrow 0} \frac{B_\delta^{\mathrm{full}}}{\mathbb{E}\Delta_\delta} = \frac{\eta^2}{57/16 + \eta^2}, \qquad \lim_{\delta \downarrow 0} \frac{B_\delta^{\mathrm{corr}}}{\mathbb{E}\Delta_\delta} = \frac{16}{57}. \tag{111}$$

Indeed, for every $\delta > 0$ the full active graph is connected, and its normalization tends to $\eta^2/(57/16 + \eta^2)$. The triangle normalization tends to $16/57$ and $\mathbb{E}\,\|K_0 z\|_2^2 = 1/800$ at $\delta = 0$. Since $\ell_1 \to 49/40$, substitution gives both limits.

The scalar $\theta_C(z)$ is optimal for the component norm used in equation (23). On a connected active component C, choose j with $z_j = \min_{i \in C} z_i$ and set $h_i = z_i$ for $i \neq j$ and $h_j = -\sum_{i \in C \setminus \{j\}} z_i^2 / z_j$. Then $z_C^\top h = 0$ and

$$\frac{\|h_+\|_2^2}{\|h\|_2^2} = \frac{z_j^2}{\|z_C\|_2^2} = \theta_C(z). \tag{112}$$

This forcing is attainable by a skew matrix on $E[C]$. Orient its edges and let B_C have column $e_j - e_i$ for the orientation $i \to j$. Connectivity gives $\operatorname{im} \mathsf{B}_C = \mathbf{1}^\perp$. The vector $w = \operatorname{diag}(z_C) h$ has zero sum, so choose an edge flow f with $\mathsf{B}_C f = w$. For each oriented edge, set

$$H_{ij} = -\frac{f_{ij}}{z_i z_j}, \qquad H_{ji} = -H_{ij}, \qquad H_{ij} = 0 \text{ off } E[C]. \tag{113}$$

At each vertex, $\operatorname{diag}(z_C) H_{CC} z_C = \mathsf{B}_C f = w$. Since every $z_i > 0$ on C, this proves $H_{CC} z_C = h$. Thus no larger scalar can bound both signs of the forcing uniformly on that component.

For $E \neq \varnothing$, positive co-activation has an equivalent rank condition: any $H \in \mathcal{K}_E$ with $P_{I(z)} H z = 0$ almost surely must be zero. Every θ_C is positive, and isolated active vertices have $(Hz)_i = 0$, so $a^\top G_\nu a = 0$ is exactly this kernel condition. For a finite population, stacking $\sqrt{p_t \theta_C(z_t)}\mathsf{A}_C(z_t)$ over inputs and components gives the rank test directly. Joint activity on every edge does not ensure full rank. For example, the constant response $z = \mathbf{1}$ on a triangle does not excite its circulating skew direction. The input model below provides positive edge coverage without prescribing the proxy responses.

A.9 Input coverage for general and sparse proxies

Co-activation is defined through proxy responses, whereas an experiment specifies an input law. We bridge these descriptions by constructing inputs that activate isolated pairs, then derive size-independent coverage conditions for bounded-degree graphs.

Proposition 12 (Input neighborhoods). *Let $Q \succ 0$, $\mu = \lambda_{\min}(Q)$, and let $E \neq \varnothing$ be any edge set. For $a, \zeta > 0$ and $e = \{i, j\} \in E$, set*

$$\begin{aligned} z^{(e)} &= a(e_i + e_j), \qquad s^{(e)} = \zeta(\mathbf{1} - e_i - e_j), \\ c^{(e)} &= Q z^{(e)} - s^{(e)}, \qquad \delta_0 = \min\left\{ \frac{\mu a}{2}, \frac{\mu \zeta}{2(\mu + \|Q\|_2)} \right\}. \end{aligned} \tag{114}$$

For an input law with $\mathbb{E}\,\|z\|_2^2 < \infty$, write $q_e = \mathbb{P}\{\|c - c^{(e)}\|_2 \le \delta_0\}$ and $q = \min_{e \in E} q_e$. Then

$$\alpha_{\mathrm{co}} \ge \frac{a^2 q}{8}. \tag{115}$$

For N independent draws from this law, let N_e count inputs in the ball about $c^{(e)}$. Then $\widehat{\alpha}_{\mathrm{co}} \ge (a^2/8) \min_e N_e/N$. If $q > 0$, $0 < \delta < 1$, and $N \ge (8/q)\log(|E|/\delta)$, this is at least $a^2 q/16$ with probability at least $1 - \delta$.

Proof. The pair $z^{(e)}, s^{(e)}$ satisfies proxy complementarity at $c^{(e)}$. Strong monotonicity gives $\|z(c) - z^{(e)}\|_2 \le \|c - c^{(e)}\|_2 / \mu$. Writing $s(c) = Q z(c) - c$ also gives

$$\left\| s(c) - s^{(e)} \right\|_2 \le \left(1 + \frac{\|Q\|_2}{\mu} \right) \left\| c - c^{(e)} \right\|_2. \tag{116}$$

Inside the prescribed ball, the two endpoint responses are at least $a/2$, and every other coordinate has slack at least $\zeta/2$. Complementarity therefore makes the active set exactly $\{i, j\}$. For any $H \in \mathcal{K}_E$, its Gram contribution is

$$\theta_{\{i,j\}}(z) \left\| P_{\{i,j\}} H z \right\|_2^2 = \min(z_i^2, z_j^2) H_{ij}^2 \ge \frac{a^2}{4} H_{ij}^2. \tag{117}$$

Balls for distinct edges are disjoint because their proxy active sets are different. All other inputs have nonnegative Gram contributions. The expected contribution is $g_\nu(H) \geq (a^2/4)\sum_e q_e H_{ij}^2 \geq (a^2 q/8)\|H\|_F^2$. The same argument with sample frequencies in place of probabilities proves the empirical bound. Each N_e is binomial; the multiplicative Chernoff bound and a union bound give $\mathbb{P}\{\min_e N_e < Nq/2\} \leq |E|\exp(-Nq/8)$. □

This condition allows mixed-sign weights and dependent input coordinates. Every nonsingular Gaussian input law gives $q > 0$ for every $Q \succ 0$ and edge set E. More quantitatively, if the input density is at least $f_0 > 0$ on the specified balls, then $q \geq f_0 \operatorname{vol}(B_2^n)\delta_0^n$, where B_2^n is the Euclidean unit ball. This certifies positivity; its size depends on the mass near the required activation patterns. The sparse input model below gives a lower bound independent of ambient dimension by imposing conditions only on an edge and its neighbors.

Proposition 13 (Local input coverage). *Let $E \neq \varnothing$ have m edges and maximum degree $d \geq 1$. Suppose $S = S^\top \geq 0$, $S_{ii} = 0$, its off-diagonal support lies in E, and $S\mathbf{1} \leq r\mathbf{1}$ for $0 \leq r < 1$. Let the input coordinates be independent, with $c_i \leq b_0$ almost surely, $\mathbb{P}(c_i \geq a) \geq p_+ > 0$, and $\mathbb{P}(c_i \leq -M) \geq p_- > 0$, where $0 < a \leq b_0$ and $M > rb_0/(1-r)$. The remaining mass may lie anywhere in $(-M, a)$. Set $q = p_+^2 p_-^{2d-2}$. Then*

$$\alpha_{\mathrm{co}} \geq \frac{a^2 q}{2}. \tag{118}$$

For N independent inputs, let N_e count the samples on which both endpoints of e have input at least a and every external neighbor of either endpoint has input at most $-M$. The empirical Gram satisfies

$$\widehat{\alpha}_{\mathrm{co}} \geq \frac{a^2}{2}\min_{e\in E}\frac{N_e}{N}. \tag{119}$$

For $0 < \delta < 1$, if $N \geq (8/q)\log(m/\delta)$, then $\widehat{\alpha}_{\mathrm{co}} \geq a^2 q/4$ with probability at least $1-\delta$.

Proof. Symmetry and the row-sum bound give $\lambda_{\max}(S) \leq \|S\|_\infty \leq r$, hence $Q = I - S \succeq (1-r)I \succ 0$. At the proxy equilibrium, $\|z\|_\infty \leq r\|z\|_\infty + b_0$, so $\|z\|_\infty \leq b_0/(1-r)$. If $c_i \leq -M$, then $(Sz+c)_i \leq rb_0/(1-r) - M < 0$, which forces $z_i = 0$. If $c_i \geq a$, nonnegativity of S gives $z_i \geq a$. The remaining coordinates may be activated by recurrent input.

For $e = \{i, j\}$, let d_e be the number of vertices adjacent to i or j, excluding the two endpoints. They form an isolated active pair whenever the input event defining N_e occurs. Independence makes its probability q_e at least $p_+^2 p_-^{d_e} \geq q$, since $d_e \leq 2d-2$. On this event, for any $H \in \mathcal{K}_E$,

$$\theta_{\{i,j\}}(z)\left\|P_{\{i,j\}}Hz\right\|_2^2 = \min(z_i^2, z_j^2)H_{ij}^2 \geq a^2 H_{ij}^2. \tag{120}$$

Events for edges sharing a vertex cannot occur together: their conditions would require one endpoint input to be both at least a and at most $-M$. Events for disjoint edges can occur together, and each then contributes a distinct active component. Thus for every sample the Gram sum dominates the sum of these event contributions, with no repeated component. The expected values therefore satisfy

$$\mathbb{E}\sum_C \theta_C(z)\|P_C Hz\|_2^2 \geq a^2 \sum_{\{i,j\}\in E} q_e H_{ij}^2 \geq \frac{a^2 q}{2}\|H\|_F^2. \tag{121}$$

The same argument with empirical frequencies proves equation (119). Across the N independent inputs, N_e is binomial with parameters N, q_e. The multiplicative Chernoff inequality at relative deviation $1/2$ gives $\mathbb{P}\{N_e < Nq_e/2\} \leq \exp(-Nq_e/8)$. A union bound over the m edges bounds the probability that any $N_e < Nq/2$ by $m\exp(-Nq/8) \leq \delta$. □

For independent uniform inputs on $[-B, b_0]$, choose $B > M$ and $a < b_0$. Then $p_+ = (b_0 - a)/(B + b_0)$ and $p_- = (B - M)/(B + b_0)$, giving

$$\alpha_{\mathrm{co}} \geq \frac{a^2}{2}\left(\frac{b_0 - a}{B + b_0}\right)^2\left(\frac{B - M}{B + b_0}\right)^{2d-2}. \tag{122}$$

The input density is positive near zero, and the proxy active set can differ from the positive-input set. For fixed degree, row-sum margin, and interval endpoints, this lower bound is independent of network

size. The argument also applies to dependent coordinates whenever the same local input events have probabilities at least q; independence supplies an explicit product bound for these probabilities.

The two-interval law with probabilities $p_+ = p$ and $p_- = 1 - p$ is a special case. Here the proxy active set is exactly the positive-input mask, and its active values lie in $[a, b_0/(1-r)]$. If $d \geq 2$ and $p = 1/d$, the inequality $\log(1 - 1/d) \geq -1/(d-1)$ gives $q \geq e^{-2}/d^2$. The population lower bound is then independent of n, and $8e^2 d^2 \log(m/\delta)$ samples suffice for the stated empirical guarantee. The expected number of active units is n/d. Bounded-degree networks can therefore have many simultaneously active units while retaining a positive, size-independent certificate.

This scaling has a direct consequence for the cycle lower bound. For an n-cycle, $n \geq 4$, take $S_{i,i+1} = S_{i+1,i} = s$ and $K_{i,i+1} = -K_{i+1,i} = k$ with cyclic indices, $0 < k < s < 1/2$, and use the two-interval input law with positive-input probability p and $M > 2sb_0/(1-2s)$. All other matrix entries are zero. Every edge has $g_e^2 = s^2 - k^2$ and oriented log ratio $\log((s+k)/(s-k))$. This constant cyclic field has zero divergence, so the Hodge formula gives

$$\mathcal{L}_\circ(S + K) = \frac{n}{2}(s^2 - k^2) \log^2 \frac{s+k}{s-k}. \tag{123}$$

Let $q = p^2(1-p)^2$ and $\bar{\ell}_\tau = 1 + 2s + 4\tau^2 k^2/(1-2s)$. The bounds $\mu \geq 1 - 2s$, $\|K\|_2 \leq 2k$, and $\ell_\tau \leq \bar{\ell}_\tau$, together with theorem 2, yield

$$\frac{n\tau^2 a^2 q (s^2 - k^2)}{8\bar{\ell}_\tau} \log^2 \frac{s+k}{s-k} \leq \mathbb{E}\Delta_\tau \leq \frac{2n\tau^2 k^2 p b_0^2}{(1-2s)^3}. \tag{124}$$

For the upper bound, apply equation (16) and $\mathbb{E}\,\|z\|_2^2 \leq npb_0^2/(1-2s)^2$, which follows from the Bernoulli active set and the response cap. For fixed parameters and $\tau \neq 0$, both bounds are proportional to n. For the general input law, including equation (122), replace q by $p_+^2 p_-^2$ and p in the upper bound by one. The response cap gives $\mathbb{E}\,\|z\|_2^2 \leq nb_0^2/(1-2s)^2$, so the same linear dependence on n holds.

Global normalization loses this size dependence. For any degree-d graph, only edges with two active endpoints contribute to the projected frame. Each such edge contributes $(z_i^2 + z_j^2)/2$, so

$$\sum_C \|\mathsf{A}_C(z)\|_F^2 = \frac{1}{2} \sum_{i \in I(z)} \deg_{E[I(z)]}(i) z_i^2 \leq \frac{d}{2} \|z\|_2^2. \tag{125}$$

The squared norm cancels when we multiply this inequality by $\theta_{\mathrm{g}}(z)$. If active responses are at most R, then $\operatorname{tr} G_{\mathrm{g}} \leq dR^2/2$ and

$$\lambda_{\min}(G_{\mathrm{g}}) \leq \frac{dR^2}{2|E|}. \tag{126}$$

For the cycle, $d = 2$, $|E| = n$, and $R = b_0/(1-2s)$. Its uniform lower bound based on G_{g} is therefore $O(1)$, while equation (124) gives mean penalty $\Theta(n)$.

For a uniform population $z = e_i + e_j$ indexed by the $m = |E|$ edges, $\theta_{\{i,j\}}(z) = 1/2$ and $\left\|P_{I(z)} H z\right\|_2^2 = 2H_{ij}^2$. Thus $G_\nu = I_m/(2m)$ and $\alpha_{\mathrm{co}} = 1/(2m)$. Any slack $s \geq 0$ supported outside $\{i, j\}$ defines a bias $c = Qz - s$ with exactly this proxy solution. These inputs can lie on different faces or at degenerate boundaries. Their actual faces are unrestricted.

Ordinary second moments do not provide the same guarantee. Take any $Q \succ 0$ and $K = -K^\top$, fix $T > 0$, and use the uniform responses $z_i = e_i$ with $s_i = a(\mathbf{1} - e_i)$, where $a \geq T \max_{j,k} |K_{jk}|$. With $c_i = Qe_i - s_i$,

$$\mathbb{E}[zz^\top] = I/n, \qquad \alpha_{\mathrm{co}} = 0, \qquad [\tau K e_i - s_i]_+ = 0 \quad (|\tau| \leq T). \tag{127}$$

Every proxy answer is therefore exact throughout this interval. The matrices S and K in equation (132) give a reciprocal same-sign example with nonzero cycle circulation. Input-wise activity of every unit and nonsingular second moments coexist with zero error.

For a size-dependent sharpness calculation, take m disjoint pairs, $Q = I_{2m}$, $c = z = a\mathbf{1}$ with $a > 0$, and

$$K = \operatorname{diag}(K_0, \ldots, K_0), \qquad K_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{128}$$

There is one edge per component, and its normalized forcing is $a^2 H_{ij}^2$. Thus $G_\nu = (a^2/2) I_m$, $G_{\mathrm{g}} = (a^2/(2m)) I_m$, and $\|K\|_F^2 = 2m$. For $0 \le \tau \le 1$, one pair has equilibrium $a(1+\tau, 1-\tau)/(1+\tau^2)$; for $\tau \ge 1$, it has equilibrium $(a, 0)$. Negative τ interchanges the coordinates. Consequently

$$\mathbb{E}\Delta_\tau = \begin{cases} ma^2\tau^2/(1+\tau^2), & |\tau| \le 1, \\ ma^2/2, & |\tau| \ge 1. \end{cases} \tag{129}$$

Since $\ell_\tau = L_\tau^2 = 1 + \tau^2$, the population lower bound is $B_{\mathrm{co}} = ma^2\tau^2/(2(1+\tau^2))$. Its ratio to the true penalty equals $1/2$ for $0 < |\tau| \le 1$ and $\tau^2/(1+\tau^2)$ for $|\tau| \ge 1$. The ratio is independent of m and tends to one after all pairs have changed face. Global normalization divides this ratio by m. The squared-state lower bound has the same ratio, since $Q = I$ and $\|y_\tau - z\|_2^2 = 2\Delta_\tau$.

For a general population, the slack and Gram data separate the sources of looseness. Write $k_E = (\sqrt{2} K_{ij})_{\{i,j\} \in E}$ and set

$$g_\nu(K) = k_E^\top G_\nu k_E, \qquad V_\tau = \mathbb{E}_\nu \|v_\tau\|_2^2. \tag{130}$$

For $\tau \neq 0$, the proof gives

$$\alpha_{\mathrm{co}} \|K\|_F^2 \le g_\nu(K) \le \mathbb{E}_\nu \left\| [\mathrm{sgn}(\tau) P_{I(z)} K z]_+ \right\|_2^2 \le \frac{V_\tau}{\tau^2} \le \frac{2\ell_\tau \mathbb{E}_\nu \Delta_\tau}{\tau^2}. \tag{131}$$

After multiplication by $\tau^2/(2\ell_\tau)$, these terms give the uniform Gram, directional Gram, active-coordinate, and dual-residual lower bounds, respectively. Each is computed from the proxy. The Gram bounds connect input coverage to geometric defect; the residual bound uses the signed forcing and slack of each input directly. A singular Gram matrix can still have $g_\nu(K) > 0$, so its directional certificate distinguishes an excited observed rotation from unexcited directions elsewhere in the edge space.

An explicit population crosses activation faces while keeping the biases fixed. Set

$$S = \frac{1}{5}(\mathbf{1}\mathbf{1}^\top - I), \qquad K = \frac{1}{10} \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}, \quad z_{i,\sigma} = \mathbf{1} - e_i, \quad c_{i,\sigma} = Q z_{i,\sigma} - \sigma e_i, \tag{132}$$

with $i \in \{1,2,3\}$ and $\sigma \in \{0, 1/200, 1/50, 1/20\}$, each of the twelve inputs equally likely. These responses have $\alpha_{\mathrm{co}} = 1/6$, $\beta = 4/3$, and $\mu = 3/5$. For $|\tau| \le 9/5$, $S + \tau K$ is nonnegative, and the two originally active coordinates have positive bias $4/5$. They remain active. For $z = (1,1,0)^\top$, the solution restricted to its two active coordinates is

$$y_1 = \frac{(4/5)(6/5 + \tau/10)}{24/25 + \tau^2/100}, \qquad y_2 = \frac{(4/5)(6/5 - \tau/10)}{24/25 + \tau^2/100}. \tag{133}$$

Substitution in the third preactivation gives $\sigma_c(\tau) - \sigma$, where

$$\sigma_c(\tau) = \frac{(6/5)(\tau/10)^2}{24/25 + (\tau/10)^2}. \tag{134}$$

Uniqueness and positivity of the other two coordinates show that the third coordinate is active exactly when $\sigma < \sigma_c(\tau)$. Cyclic permutation proves the same statement for the other inputs. Thus the population contains both persistent and changing faces as τ varies, including degenerate proxy boundaries at $\sigma = 0$. At each τ, applying the cycle certificate to the actual matrix $W_\tau = S + \tau K$ gives the lower bound $\alpha_{\mathrm{co}} \mathcal{L}_{\mathrm{o}}(W_\tau)/(2\ell_\tau)$.

A.10 Population refinements and certificate tightness

Input coverage makes a lower bound positive, but does not determine how much of the true error it captures. We first derive a sharper formula on a common full face, then separate the losses due to activation imbalance and spectral bounds.

Theorem 14 (Population bounds on a common full face). *Let $Q = I - S \succ 0$, $\mu = \lambda_{\min}(Q)$, $\Lambda = \lambda_{\max}(Q)$, and $K = -K^\top$. Suppose almost surely $\min_i z_i(c) \ge \gamma > 0$ and $\|z(c)\|_2 \le R$. Set $\Xi = \mathbb{E}_\nu[z(c) z(c)^\top]$, $k_Q = \left\| Q^{-1/2} K Q^{-1/2} \right\|_2$, and*

$$\tau_0 = \frac{\gamma\mu}{\|K\|_2 R}, \qquad a_\nu^2 = \mathrm{tr}(K^\top Q^{-1} K \Xi), \tag{135}$$

with $\tau_0 = +\infty$ if $K = 0$. For $|\tau| < \tau_0$, every equilibrium is fully active almost surely, and

$$\frac{\tau^2 a_\nu^2}{2(1+\tau^2 k_Q^2)} \le \mathbb{E}_\nu \Delta_\tau \le \frac{\tau^2 a_\nu^2}{2}. \tag{136}$$

If $0 < \alpha I \preceq \Xi \preceq \beta I$, then

$$\frac{\tau^2 \alpha \|K\|_F^2}{2\Lambda(1+\tau^2 k_Q^2)} \le \mathbb{E}_\nu \Delta_\tau \le \frac{\tau^2 \beta \|K\|_F^2}{2\mu}. \tag{137}$$

Under the reciprocal same-sign support condition, the lower bound also holds with $\|K\|_F^2$ replaced by $\mathcal{L}_\circ(W)$.

Detailed proof of theorem 14. Apply the global state bound in lemma 1 separately to every bias in the probability-one set specified by the assumptions:

$$\|y_\tau(c) - z(c)\|_\infty \le \|y_\tau(c) - z(c)\|_2 \le \frac{|\tau| \|K\|_2 R}{\mu} < \gamma. \tag{138}$$

Thus every coordinate of $y_\tau(c)$ is strictly positive for $|\tau| < \tau_0$, as is every coordinate of $z(c)$. This estimate certifies the common full face using only proxy responses. If $K = 0$, the two equilibria coincide for every τ and all conclusions hold with zero penalty.

For $K \neq 0$, let $\overline{K} = Q^{-1/2} K Q^{-1/2}$ and $v(c) = Q^{-1/2} K z(c)$. The full-face instance of equation (75) gives the exact identity

$$\Delta_\tau(c) = \frac{\tau^2}{2} v(c)^\top (I - \tau^2 \overline{K}^2)^{-1} v(c). \tag{139}$$

Since $0 \preceq -\overline{K}^2 \preceq k_Q^2 I$, the middle matrix is bounded between $(1+\tau^2 k_Q^2)^{-1} I$ and I. Also $\|z(c)\| \le R$ ensures integrability of every quadratic term. We take expectations and use

$$\mathbb{E}_\nu \|v(c)\|_2^2 = \operatorname{tr}(K^\top Q^{-1} K \Xi) = a_\nu^2 \tag{140}$$

to obtain equation (136). In particular, the exact population prediction from proxy second moments is

$$\mathbb{E}_\nu \Delta_\tau = \frac{\tau^2}{2} \operatorname{tr}\left[(I - \tau^2 \overline{K}^2)^{-1} Q^{-1/2} K \Xi K^\top Q^{-1/2} \right]. \tag{141}$$

Trace monotonicity for positive-semidefinite factors yields

$$a_\nu^2 \ge \alpha \operatorname{tr}(K^\top Q^{-1} K) \ge \frac{\alpha}{\Lambda} \|K\|_F^2, \tag{142}$$

$$a_\nu^2 \le \beta \operatorname{tr}(K^\top Q^{-1} K) \le \frac{\beta}{\mu} \|K\|_F^2. \tag{143}$$

Substitution proves equation (137). The geometric identity and global cycle bound give $\|K\|_F^2 = J(u) \ge \mathcal{L}_\circ(W)$ under their stated support conditions. Consequently a nonzero circulation and $\alpha > 0$ force strictly positive average penalty whenever $0 < |\tau| < \tau_0$. The cycle bound uses the support and defect of the fixed base matrix $A = S + K$. □

On the common full face, set $\delta_x = D^{-1/2}(y_\tau - z)$. The exact quadratic energy identity gives $2\Delta_\tau = \delta_x^\top D^{1/2} Q D^{1/2} \delta_x$, hence

$$\frac{\mu d_{\min}}{2} e_{\tau,D}(c)^2 \le \Delta_\tau(c) \le \frac{\Lambda d_{\max}}{2} e_{\tau,D}(c)^2. \tag{144}$$

Under $\alpha I \preceq \Xi \preceq \beta I$, this yields

$$\frac{\tau^2 \alpha \|K\|_F^2}{\Lambda^2 d_{\max}(1+\tau^2 k_Q^2)} \le \mathbb{E}_\nu e_{\tau,D}^2 \le \frac{\tau^2 \beta \|K\|_F^2}{\mu^2 d_{\min}}. \tag{145}$$

Write $m_\nu = \mathbb{E}_\nu z$ and $\Sigma_\nu = \mathrm{Cov}_\nu(z)$, so $\Xi = m_\nu m_\nu^\top + \Sigma_\nu$. Then

$$a_\nu^2 = \left\| Q^{-1/2} K m_\nu \right\|_2^2 + \mathrm{tr}(K^\top Q^{-1} K \Sigma_\nu). \tag{146}$$

If $Km_\nu = 0$, every occurrence of Ξ in equations (135) and (141) can be replaced by Σ_ν, and the spectral bounds in equation (137) use its eigenvalue bounds instead. For a general mean, the upper bound includes both terms in equation (146). At $K = 0$ all terms vanish. The forcing bounds also apply to singular response moments. For example, the matrices in equation (132) with the single bias $c = \mathbf{1}$ have $y = z = (5/3)\mathbf{1}$ and $Kz = 0$. This input has zero error and nonzero cycle circulation.

A finite collection of strictly positive proxy responses supplies explicit full-face and bounded-response constants. The construction below uses the same six biases across an entire circulation sweep and certifies their margin before computing actual equilibria. Response moments are computed in the chosen metric. For a fixed original input distribution and a family of metrics, uniform versions of these certificates, together with equation (144), give the state-error bounds in equation (145).

For a family attaining the lower bound, take the uniform distribution on the six fixed biases $c_{i,\pm} = \mathbf{1} \pm e_i/4$, $1 \le i \le 3$, and set

$$W_{\xi,d} = dI + \frac{1}{5} \begin{pmatrix} 0 & e^{\xi} & e^{-\xi} \\ e^{-\xi} & 0 & e^{\xi} \\ e^{\xi} & e^{-\xi} & 0 \end{pmatrix}, \qquad 0 \le \xi \le \frac{1}{2}, \quad d \in \{0, -2\}. \tag{147}$$

Use $D = I$, $S = (W_{\xi,d} + W_{\xi,d}^\top)/2$, $K = (W_{\xi,d} - W_{\xi,d}^\top)/2$, and $\tau = 1$. These matrices are already Frobenius-balanced, with

$$J = \frac{6}{25} \sinh^2 \xi, \qquad \mathcal{L}_\circ(W_{\xi,d}) = \frac{6}{25} \xi^2. \tag{148}$$

The eigenvalues of $Q = I - S$ are $\mu = 1 - d - (2/5)\cosh\xi$ on $\mathrm{span}\{\mathbf{1}\}$ and $\Lambda = 1 - d + (1/5)\cosh\xi$ on its orthogonal complement. Its inverse is entrywise nonnegative. Since $c_{i,\pm} \ge (3/4)\mathbf{1}$ and $\|c_{i,\pm}\| < 2$, all proxy responses obey $z \ge (3/(4\mu))\mathbf{1}$ and $\|z\| < 2/\mu$. On this parameter range $\mu > 0.54$ and $\|K\|_2 = (\sqrt{3}/5)\sinh\xi < 0.19$, giving the uniform sufficient radius $\tau_0 > 81/76 > 1$ (or $+\infty$ at $\xi = 0$). The population theorem therefore applies on the entire stated parameter range.

Here $m_\nu = \mu^{-1}\mathbf{1}$ is annihilated by K, while $\Sigma_\nu = Q^{-2}/48 \succ 0$. Rotation acts only on $\mathbf{1}^\perp$, where $Q = \Lambda I$ and $-K^2 = \|K\|_2^2 I$. Therefore the covariance lower bound in equation (137) is attained exactly:

$$\mathbb{E}_\nu \Delta_1 = \frac{J}{96\Lambda^3(1 + k_Q^2)}, \qquad k_Q = \frac{\sqrt{3}\sinh\xi}{5\Lambda}. \tag{149}$$

For $\xi > 0$, the ratio of the cycle-certified lower bound to this exact penalty is $(\xi/\sinh\xi)^2$, which tends to one as $\xi \downarrow 0$. The lower bound is attained for this fixed input distribution, and the cycle lower bound is asymptotically exact as $\xi \to 0$. The family with $d = -2$ also has $\|W_{\xi,-2}\|_2 > 1$.

Proposition 15 (Directional tightness of the co-activation bound)**.** *Suppose $Q \succ 0$, $K \in \mathcal{K}_E$, $\tau \ne 0$, and $\mathbb{E}_\nu \|z\|_2^2 < \infty$. Set $F_\nu = \mathbb{E}_\nu \|Kz\|_2^2$ and assume $F_\nu > 0$. Define*

$$A_\tau = Q - \tau^2 K Q^{-1} K, \qquad m_\tau = \lambda_{\min}(A_\tau), \qquad B_{\mathrm{co}} = \frac{\tau^2 g_\nu(K)}{2\ell_\tau}, \tag{150}$$

where $\ell_\tau = \lambda_{\max}(A_\tau)$ and $\mu = \lambda_{\min}(Q)$. For $p \ge 2$ and $r \ge 1$, let $\mathcal{C}_{p,r}(z)$ contain the components $C \in \mathcal{C}_E(z)$ with $|C| \le p$ and $\max_{i \in C} z_i / \min_{i \in C} z_i \le r$. If, for some $a > 0$,

$$\mathbb{E}_\nu \sum_{C \in \mathcal{C}_{p,r}(z)} \|P_C K z\|_2^2 \ge a F_\nu, \tag{151}$$

then the equilibria may have different active sets and still satisfy

$$\frac{a\mu}{\ell_\tau p r^2} \le \frac{B_{\mathrm{co}}}{\mathbb{E}_\nu \Delta_\tau} \le 1. \tag{152}$$

If both equilibria are fully active almost surely, the tighter comparison

$$\frac{m_\tau}{\ell_\tau} \frac{g_\nu(K)}{F_\nu} \le \frac{B_{\mathrm{co}}}{\mathbb{E}_\nu \Delta_\tau} \le \frac{g_\nu(K)}{F_\nu} \tag{153}$$

holds for every such full-face population. If every active component has size at most p and response ratio at most r almost surely, it gives

$$\frac{B_{\mathrm{co}}}{\mathbb{E}_\nu \Delta_\tau} \geq \frac{m_\tau}{\ell_\tau p r^2}. \tag{154}$$

On this full face, if E is connected on n vertices, the same ratio is at most $1/n$.

Proof. Every component in $\mathcal{C}_{p,r}(z)$ obeys

$$\theta_C(z) = \frac{\min_{i\in C} z_i^2}{\sum_{i\in C} z_i^2} \geq \frac{1}{|C| r^2} \geq \frac{1}{p r^2}. \tag{155}$$

Retaining just these nonnegative terms in the Gram sum gives

$$g_\nu(K) \geq \frac{1}{p r^2} \mathbb{E}_\nu \sum_{C\in\mathcal{C}_{p,r}(z)} \|P_C K z\|_2^2 \geq \frac{a F_\nu}{p r^2} > 0. \tag{156}$$

The other components and inputs are unrestricted. In particular, responses arbitrarily close to an activation boundary can occur outside the retained components. The global energy upper bound in lemma 1 gives $\mathbb{E}\Delta_\tau \leq \tau^2 F_\nu/(2\mu)$, while theorem 2 gives $\mathbb{E}\Delta_\tau \geq B_{\mathrm{co}} > 0$. Together, the two inequalities prove equation (152) without a restriction on the actual activation sets.

Now suppose both responses are fully active. Write $M_\tau = Q - \tau K$ and $w = Kz$. Full activity gives $M_\tau(y_\tau - z) = \tau w$ and zero proxy slack. Hence

$$2\Delta_\tau = \tau^2 w^\top M_\tau^{-\top} Q M_\tau^{-1} w = \tau^2 w^\top A_\tau^{-1} w, \tag{157}$$

because $M_\tau Q^{-1} M_\tau^\top = A_\tau$. The matrix $A_\tau \succeq Q \succ 0$, so its inverse lies between $\ell_\tau^{-1} I$ and $m_\tau^{-1} I$. In particular, $\mathbb{E}\Delta_\tau$ is positive and finite. The exact ratio factors as

$$\frac{B_{\mathrm{co}}}{\mathbb{E}\Delta_\tau} = \frac{g_\nu(K)}{F_\nu} \frac{F_\nu}{\ell_\tau \mathbb{E}[w^\top A_\tau^{-1} w]}. \tag{158}$$

The second factor lies in $[m_\tau/\ell_\tau, 1]$, proving equation (153). Full activity also gives $\sum_C \|P_C K z\|_2^2 = \|Kz\|_2^2$; isolated vertices contribute zero. The component assumptions therefore give $g_\nu(K) \geq F_\nu/(p r^2)$, which proves equation (154). For a connected graph, there is one active component and $\theta_C \leq 1/n$ pointwise. Therefore $g_\nu(K) \leq F_\nu/n$, proving the final assertion. If $F_\nu = 0$, the full-face equation instead gives zero error and zero lower bound; the ratios are then undefined. □

Response imbalance can change the ordering of errors even when the spectral factor in equation (158) is one. Take $D = Q = I$, $\tau = 1$, and the fixed contractive rotation

$$K = \frac{1}{2}\begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}, \qquad c_\eta = z_\eta = (\eta, 1, 1)^\top, \qquad 0 < \eta < \frac{1}{4}. \tag{159}$$

Here $S = 0$ and $\|K\|_2 = \sqrt{3}/2 < 1$. Since $K^2 = (\mathbf{1}\mathbf{1}^\top - 3I)/4$, the actual response is

$$y_\eta = \frac{1}{7}(2 + 5\eta, 8 - \eta, 4 + 3\eta)^\top, \qquad y_\eta - z_\eta = \frac{1-\eta}{7}(2, 1, -3)^\top. \tag{160}$$

All coordinates are positive, and substitution verifies $(I - K)y_\eta = c_\eta$. Consequently

$$\Delta_\eta = \frac{(1-\eta)^2}{7}, \qquad \|K z_\eta\|_2^2 = \frac{(1-\eta)^2}{2}, \qquad \frac{B_{\mathrm{co},\eta}}{\Delta_\eta} = \frac{\eta^2}{2 + \eta^2}. \tag{161}$$

The last equality follows from $\ell_1 = 7/4$ and the single-component coefficient $\theta = \eta^2/(2 + \eta^2)$. The true error decreases strictly as η increases, whereas

$$\frac{d}{d\eta} B_{\mathrm{co},\eta} = \frac{2\eta(1-\eta)(2 - 4\eta - \eta^3)}{7(2 + \eta^2)^2} > 0. \tag{162}$$

Thus the forcing gives the exact error ranking, whereas the co-activation lower bound reverses that ranking. All three edges have fixed magnitude $1/2$, the proxy curvature is one, and neither response changes face. This is an input comparison in a fixed network. It isolates the activation-normalization factor in equation (158).

The state response also admits a directional curvature refinement. For an input whose two equilibria are fully active, let $V \neq \{0\}$ be a linear subspace containing Kz and invariant under $M_\tau = Q - \tau K$. Define

$$\mu_V = \min_{v \in V \setminus \{0\}} \frac{v^\top Q v}{\|v\|_2^2}, \qquad L_V = \max_{v \in V \setminus \{0\}} \frac{\|M_\tau v\|_2}{\|v\|_2}. \tag{163}$$

Then

$$\frac{|\tau| \, \|Kz\|_2}{L_V} \leq \|y_\tau - z\|_2 \leq \frac{|\tau| \, \|Kz\|_2}{\mu_V}. \tag{164}$$

Indeed, M_τ is invertible by $Q \succ 0$. Its restriction to the finite-dimensional invariant subspace V is therefore onto V, so $\delta = y_\tau - z = \tau M_\tau^{-1} K z$ lies in V. The lower bound follows from $\|M_\tau \delta\|_2 \leq L_V \, \|\delta\|_2$. For the upper bound, skew-symmetry gives

$$\mu_V \, \|\delta\|_2^2 \leq \delta^\top Q \delta = \delta^\top M_\tau \delta = \tau \delta^\top K z \leq |\tau| \, \|\delta\|_2 \, \|Kz\|_2 . \tag{165}$$

Division proves the result when $\delta \neq 0$; otherwise both bounds are zero. The invariance condition ensures that the response remains in V; containing the initial forcing alone does not ensure this.

For a fixed-input example, set $D = I$, $\tau = 1$, $c = \mathbf{1}$, and

$$Q_\varepsilon = \operatorname{diag}(\varepsilon, 1, 1), \qquad K = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \qquad 0 < \varepsilon < 1. \tag{166}$$

The proxy and actual responses are $z_\varepsilon = (\varepsilon^{-1}, 1, 1)^\top$ and $y_\varepsilon = (\varepsilon^{-1}, 6/5, 2/5)^\top$. They are positive and solve their respective linear systems, giving

$$\|y_\varepsilon - z_\varepsilon\|_2^2 = \frac{2}{5}, \qquad U_x = \frac{2U_\Delta}{\lambda_{\min}(Q_\varepsilon)} = \frac{1}{2\varepsilon}. \tag{167}$$

Here $U_\Delta = 1/4$ is the perturbative branch of equation (16). The subspace $V = \operatorname{span}\{e_2, e_3\}$ is invariant and has $\mu_V = 1$, so equation (164) instead gives the uniform squared-error bound $1/2$. The restricted resolvent satisfies $\left\|(M_1|_V)^{-1}\right\|_2^2 = 4/5$ and recovers the exact error $2/5$. The eigenvalue ε lies in an unexcited direction; in theorem 3, the fixed biases excite the direction whose curvature vanishes.

A.11 An explicit unbounded separation for least-defect metrics

The construction must make the original recurrence uniformly stable while the least-defect proxy becomes nearly flat. We start with a balanced matrix, use a Perron-vector scaling to obtain the stable original coordinates, and then compare the two proxy responses.

Detailed proof of theorem 3. Start from the nonnegative irreducible integer matrix

$$B = \begin{pmatrix} 0 & 7 & 1 \\ 5 & 0 & 7 \\ 5 & 5 & 0 \end{pmatrix}, \qquad S_B = \frac{B + B^\top}{2} = \begin{pmatrix} 0 & 6 & 3 \\ 6 & 0 & 6 \\ 3 & 6 & 0 \end{pmatrix}. \tag{168}$$

For any matrix A in this similarity class, differentiation of equation (6) gives $\partial_i J = (\sum_j A_{ij}^2 - \sum_j A_{ji}^2)/2$. The row and column squared norms of B agree, both being $(50, 74, 50)$, so the gradient vanishes at B. Therefore B is already the unique Frobenius-balanced representative of its positive diagonal-similarity class up to the scalar gauge, by proposition 9. Its squared rotational defect is exactly 12. The vector $(1, 2, 1)^\top$ spans the kernel of $B - B^\top$, but $B(1, 2, 1)^\top = (15, 12, 15)^\top$ is not proportional to it. Thus B has no common positive left and right Perron vector.

Let $\rho = \rho(B)$ and $\lambda = \lambda_{\max}(S_B)$. The characteristic polynomial of B is $p(t) = t^3 - 75t - 270$, while

$$\lambda = \frac{3}{2}(1 + \sqrt{33}), \qquad p(10) < 0, \quad p(\lambda) = 6(\lambda - 9) > 0. \tag{169}$$

Since $p'(t) > 0$ for $t > 5$ and p has exactly one positive root,

$$10 < \rho < \lambda < 51/5. \tag{170}$$

In particular $q = \rho/\lambda < 1$ is a fixed constant.

Choose positive right and left Perron vectors r, l of B and define

$$H = \operatorname{diag}(l_i/r_i), \qquad P = H^{1/2} B H^{-1/2}. \tag{171}$$

The positive vector $v_i = \sqrt{l_i r_i}$ satisfies $Pv = P^\top v = \rho v$, hence $P^\top P v = \rho^2 v$. To identify the largest eigenvalue, let $V = \operatorname{diag}(v)$. The nonnegative matrix $V^{-1} P^\top P V$ has every row sum equal to ρ^2. Its spectral radius is at most its maximum row sum and at least its eigenvalue ρ^2. Similarity and positive semidefiniteness of $P^\top P$ therefore give

$$\|P\|_2 = \rho. \tag{172}$$

The common-Perron scaling gives an exact original-coordinate norm certificate for the construction.

The permutation that reverses the three coordinates conjugates B to $B^\top$, so we choose $l = (r_3, r_2, r_1)^\top$. Then $H = \operatorname{diag}(h, 1, h^{-1})$, where the first two Perron equations give

$$h = \frac{r_3}{r_1} = \frac{\rho^2 - 35}{\rho + 49}, \qquad 1 < h < \frac{(51/5)^2 - 35}{10 + 49} < \frac{6}{5}. \tag{173}$$

Thus $\det H = 1$ and $\operatorname{cond}_2(H) = h^2 < 36/25 < 3/2$.

For $0 < \varepsilon < 1$ set

$$t_\varepsilon = \frac{1 - \varepsilon}{\lambda}, \qquad W_\varepsilon = t_\varepsilon P, \qquad D_\star = H^{-1}, \qquad b = \frac{\mathbf{1}}{\sqrt{3}}. \tag{174}$$

Then $\|W_\varepsilon\|_2 = (1-\varepsilon)q \le q$. Its least-defect representative is $D_\star^{1/2} W_\varepsilon D_\star^{-1/2} = t_\varepsilon B$. Positive scalar multiplication does not change the balancing optimizer, so this $D_\star$ is fixed and optimal throughout the family. It differs from the identity because $h > 1$; uniqueness of balancing therefore gives the strict inequality

$$J_\varepsilon(D_\star) = 12 t_\varepsilon^2 < t_\varepsilon^2 \left\|(P - P^\top)/2\right\|_F^2 = J_\varepsilon(I). \tag{175}$$

Put $S_P = (P + P^\top)/2$. Since $S_P \ge 0$ and $\|S_P\|_2 \le \|P\|_2 = \rho$, both the actual equilibrium and identity proxy are strictly positive Neumann-series solutions:

$$x_\varepsilon = (I - t_\varepsilon P)^{-1} b, \qquad \widehat{x}_{\varepsilon, I} = (I - t_\varepsilon S_P)^{-1} b. \tag{176}$$

Their norms are at most $1/(1-q)$ and $I - t_\varepsilon S_P \succeq (1-q) I$. Hence $e_\varepsilon(I) \le 2/(1-q)$ uniformly.

For the least-defect proxy,

$$Q_\varepsilon = I - t_\varepsilon S_B \succ 0, \qquad \lambda_{\min}(Q_\varepsilon) = \varepsilon, \qquad \widehat{x}_{\varepsilon, D_\star} = H^{1/2} Q_\varepsilon^{-1} H^{-1/2} b. \tag{177}$$

This solution is also strictly positive, since $\rho(t_\varepsilon S_B) = 1 - \varepsilon < 1$. The eigenvalues of S_B are $\lambda = 3(1 + \sqrt{33})/2$, -3, and $\lambda_- = 3(1 - \sqrt{33})/2 < 0$. Let $p > 0$ be its unit eigenvector at λ and let Π_{-3}, Π_- be the orthogonal projectors onto the other two eigenspaces. The spectral decomposition gives the exact identity

$$Q_\varepsilon^{-1} = \varepsilon^{-1} p p^\top + U_\varepsilon, \qquad U_\varepsilon = \frac{\Pi_{-3}}{1 + 3t_\varepsilon} + \frac{\Pi_-}{1 - t_\varepsilon \lambda_-}. \tag{178}$$

In particular, $U_\varepsilon p = 0$ and $0 \preceq U_\varepsilon \preceq I$ for every $0 < \varepsilon < 1$. Thus the remainder is uniformly bounded in operator norm. The actual equilibrium remains bounded, whereas

$$\widehat{x}_{\varepsilon, D_\star} = \frac{H^{1/2} p \, (p^\top H^{-1/2} b)}{\varepsilon} + O(1). \tag{179}$$

The norm of the difference therefore satisfies equation (26) with

$$C_\star = \left\| H^{1/2} p \right\|_2 (p^\top H^{-1/2} b) > 0. \tag{180}$$

Every matrix, input, and metric apart from the scalar t_ε is fixed. The matrices of the identity and least-defect proxy maps have spectral norms at most q and $1-\varepsilon$, respectively, in their corresponding coordinates. Their iterations therefore converge to the respective equilibria for each $\varepsilon > 0$.

Keep the same matrices and fix any input law supported on $[b_-, b_+]^3$, where $0 < b_- < b_+ < \infty$. We first establish the uniform Gram guarantee under the additional assumption $\mathbb{E}[bb^\top] \succeq \sigma I$ for some $\sigma > 0$. In the least-defect coordinates, $z_\varepsilon(b) = Q_\varepsilon^{-1} H^{-1/2} b$. Since $\lambda_{\max}(Q_\varepsilon) \le 2$ and $D_\star = H^{-1}$,

$$\begin{aligned}\mathbb{E}[z_\varepsilon z_\varepsilon^\top] &= Q_\varepsilon^{-1} H^{-1/2} \mathbb{E}[bb^\top] H^{-1/2} Q_\varepsilon^{-1} \\ &\succeq \sigma Q_\varepsilon^{-1} D_\star Q_\varepsilon^{-1} \succeq \sigma d_{\min} Q_\varepsilon^{-2} \succeq \frac{\sigma d_{\min}}{4} I,\end{aligned} \tag{181}$$

where $d_{\min} = \min_i (D_\star)_{ii}$. Let $p > 0$ be the unit Perron eigenvector of S_B used above, and set $a_- = \min_{b,i} (H^{-1/2} b)_i / p_i$ and $a_+ = \max_{b,i} (H^{-1/2} b)_i / p_i$ over this box. Both are finite and positive. The inverse Q_ε^{-1} is entrywise nonnegative and $Q_\varepsilon^{-1} p = p/\varepsilon$, so the weight $\theta(z) = \min_i z_i^2 / \|z\|_2^2$ of the connected active graph satisfies

$$\frac{a_- p}{\varepsilon} \le z_\varepsilon(b) \le \frac{a_+ p}{\varepsilon}, \qquad \theta(z_\varepsilon(b)) \ge \theta_0 := \left(\frac{a_-}{a_+}\right)^2 \min_i p_i^2 > 0. \tag{182}$$

Thus for every skew matrix L,

$$\mathbb{E}[\theta(z_\varepsilon) \|L z_\varepsilon\|_2^2] \ge \frac{\theta_0 \sigma d_{\min}}{4} \|L\|_F^2, \qquad \alpha_{\mathrm{co},\varepsilon} \ge \frac{\theta_0 \sigma d_{\min}}{4} > 0. \tag{183}$$

The remaining error-rate argument uses only the fixed positive input box and applies even when $\mathbb{E}[bb^\top]$ is singular. The identity-proxy errors are at most $2\sqrt{3} b_+/(1-q)$ throughout the input box. The Perron expansion has a uniformly bounded remainder on this box, so

$$\mathbb{E} e_\varepsilon(D_\star)^2 = \frac{\mathbb{E} C_\star(b)^2}{\varepsilon^2} + O(\varepsilon^{-1}), \qquad C_\star(b) = \left\| H^{1/2} p \right\|_2 (p^\top H^{-1/2} b) > 0. \tag{184}$$

We now derive these rates from the population certificate itself. Put $K_0 = (B - B^\top)/(2\lambda)$, $K_\varepsilon = (1-\varepsilon) K_0$, and $M_0 = I - B/\lambda$. Define

$$a_b = p^\top H^{-1/2} b, \qquad A_\nu = \mathbb{E} a_b^2 > 0, \qquad \theta_* = \min_i p_i^2 > 0, \qquad u = K_0 p. \tag{185}$$

Here $u \neq 0$: otherwise $Bp = S_B p = \lambda p$, contradicting $\rho(B) < \lambda$. Write $c_b = H^{-1/2} b$. Equation (178) gives $z_\varepsilon(b) = a_b p/\varepsilon + U_\varepsilon c_b$ with a uniformly bounded remainder over the input box. The numbers a_b have a positive minimum, so $z_\varepsilon(b)/\|z_\varepsilon(b)\|_2 = p + O(\varepsilon)$ uniformly. The smallest squared coordinate of the normalized response then satisfies $\theta(z_\varepsilon(b)) = \theta_* + O(\varepsilon)$. Also $z_\varepsilon^\top Q_\varepsilon z_\varepsilon = c_b^\top Q_\varepsilon^{-1} c_b = a_b^2/\varepsilon + c_b^\top U_\varepsilon c_b$ and $K_\varepsilon z_\varepsilon = a_b u/\varepsilon + O(1)$. These pointwise expansions and their expectations give

$$z_\varepsilon(b) = a_b p/\varepsilon + O(1), \qquad \theta(z_\varepsilon(b)) = \theta_* + O(\varepsilon), \tag{186}$$

$$R_\nu = A_\nu/\varepsilon + O(1), \qquad g_\nu(K_\varepsilon) = \theta_* A_\nu \|u\|_2^2 / \varepsilon^2 + O(\varepsilon^{-1}). \tag{187}$$

In these expressions ν is the fixed bias population, mapped into the fixed coordinates $D_\star$; its responses depend on ε. For the curvature denominator, the same exact decomposition gives

$$Q_\varepsilon - K_\varepsilon Q_\varepsilon^{-1} K_\varepsilon = \frac{(1-\varepsilon)^2}{\varepsilon} u u^\top + Q_\varepsilon - K_\varepsilon U_\varepsilon K_\varepsilon. \tag{188}$$

The last two terms have uniformly bounded operator norm. Since $(1-\varepsilon)^2/\varepsilon = \varepsilon^{-1} + O(1)$, the largest-eigenvalue perturbation bound yields

$$Q_\varepsilon - K_\varepsilon Q_\varepsilon^{-1} K_\varepsilon = u u^\top/\varepsilon + O(1), \qquad \ell_1 = \|u\|_2^2/\varepsilon + O(1), \qquad L_1 \longrightarrow \|M_0\|_2 > 0. \tag{189}$$

The limit for L_1 also follows directly from $Q_\varepsilon - K_\varepsilon = I - t_\varepsilon B$. Here M_0 is invertible because $\rho(B)/\lambda < 1$. The actual responses stay bounded. Applying equation (17) therefore gives

$$\mathbb{E}\Delta_\varepsilon = \frac{A_\nu}{2\varepsilon} + O(1), \qquad \frac{g_\nu(K_\varepsilon)/(2\ell_1)}{\mathbb{E}\Delta_\varepsilon} \longrightarrow \theta_*, \qquad \frac{R_\nu/2}{\mathbb{E}\Delta_\varepsilon} \longrightarrow 1. \tag{190}$$

For squared state error, the directional lower bound has leading term $\theta_* A_\nu \|u\|_2^2 / (d_{\max} \|M_0\|_2^2 \varepsilon^2)$, and the energy-budget upper bound has leading term $A_\nu/(d_{\min}\varepsilon^2)$. These positive constants prove equation (27) directly from theorem 2, and agree in order with equation (184).

The distinction between directional and uniform excitation is necessary on this family. The explicit skew matrix

$$L = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -p_3 & p_2 \\ p_3 & 0 & -p_1 \\ -p_2 & p_1 & 0 \end{pmatrix} \tag{191}$$

has $Lp = 0$ and $\|L\|_F = \|p\|_2 = 1$. It follows that $Lz_\varepsilon(b) = LU_\varepsilon c_b = O(1)$. As $\theta(z) \le 1$, evaluating the Gram quadratic form at this fixed L gives $\alpha_{\mathrm{co},\varepsilon} = O(1)$. If $\mathbb{E}[bb^\top] \succ 0$, combining this upper bound with equation (183) gives $\alpha_{\mathrm{co},\varepsilon} = \Theta(1)$. For these input laws, the uniform energy lower bound has order ε, whereas retaining the observed rotation gives the correct order ε^{-1}. This completes both the single-input separation and the population statement of theorem 3. □

For a point-mass input, the directional rates still hold, but $\alpha_{\mathrm{co},\varepsilon} = 0$ at every ε. Indeed, for its positive response $z \in \mathbb{R}^3$, the skew matrix

$$L_z = \frac{1}{\sqrt{2}\,\|z\|_2} \begin{pmatrix} 0 & -z_3 & z_2 \\ z_3 & 0 & -z_1 \\ -z_2 & z_1 & 0 \end{pmatrix} \tag{192}$$

satisfies $\|L_z\|_F = 1$ and $L_z z = 0$. It belongs to the complete three-edge space of this construction, so $g_\nu(L_z) = 0$. The observed rotation has positive directional excitation for all sufficiently small ε and yields the same certified divergence rates.

The population statement therefore includes point masses, uniform box inputs, and correlated continuous laws on compact subsets of the positive orthant. Their input distributions remain fixed as the proxy curvature vanishes. For the deterministic numerical checks, we use the six biases

$$b_{i,\pm} = \frac{\mathbf{1}}{\sqrt{3}} \pm \frac{e_i}{4\sqrt{3}}, \qquad i \in \{1,2,3\}, \tag{193}$$

with equal probabilities. Their mean is $\mathbf{1}/\sqrt{3}$ and their covariance is $I/144$, so the proof applies with $\sigma = 1/144$.

B Experimental Details

The experiments separate comparisons on trained models from controlled tests of the error bounds. We give the data, metric choices, numerical checks, and evaluation protocols in that order, followed by commands for reproducing the analytic families.

B.1 Trained ReLU networks

To isolate proxy fidelity from training effects, we train each classifier once and hold its weights and encoder inputs fixed when comparing metrics. The following protocol specifies the shared data and candidate proxies used in each comparison.

We use a class-stratified split of the 50,000 CIFAR-10 training images into 45,000 training and 5,000 validation images. The calibration subset contains 512 validation images; the remaining 4,488 monitor training. A fixed class-stratified subset of 1,024 test images measures proxy error. We evaluate classifier accuracy on the full 10,000-image test set. Training seeds are 11, 22, and 33; split, calibration, and test-subset seeds are 20260909, 20260910, and 20260911. Training uses random crops with four-pixel padding and horizontal flips; evaluation uses neither. Channel means are $(0.4914, 0.4822, 0.4465)$ and standard deviations are $(0.2470, 0.2435, 0.2616)$.

The encoder has three 3×3 convolutions with 64, 128, and 256 output channels, strides 1, 2, and 2, and one-pixel padding. Group normalization with 8, 8, and 16 groups and ReLU follow the respective convolutions. Global average pooling and a 256×256 affine map produce b. The recurrent state satisfies $x = [Wx + b]_+$, and an affine head maps the 256-dimensional state to ten class logits. The

recurrent matrix starts at spectral norm 0.8 and is radially clipped to 0.95 after every optimizer step using its computed largest singular value. We impose no symmetry or excitation regularizer.

Training runs for 100 epochs with batch size 256 and AdamW. Learning rates are 10^{-3} for encoder and head and 3×10^{-4} for W, with five epochs of linear warmup and cosine decay to one percent of the initial rates. Matrix weight decay is 5×10^{-4} for encoder and head and zero for W, biases, and normalization parameters. The global gradient norm is clipped at one. Training uses 32-bit floating-point arithmetic (FP32), with automatic mixed precision (AMP) and TensorFloat-32 (TF32) disabled. Implicit differentiation solves $(I - W^\top P)\lambda = g$, where g is the loss gradient with respect to the equilibrium state and P is the diagonal active-set mask, to relative residual 10^{-5}. The resulting gradients are $P\lambda$ for b and $(P\lambda)x^\top$ for W. Small-dimensional finite-difference checks give maximum relative errors 2.2×10^{-9} for b and 7.3×10^{-8} for W.

Each checkpoint uses $D = \mathrm{diag}(e^u)$ with $\sum_i u_i = 0$ and $|u_i| \le \log 2$. The 112 main candidates are the identity, the box-constrained minimum of $\|K(u)\|_F^2$, 14 points $u = (j/15)u_{\mathrm{bal}}$ for $j = 1, \dots, 14$, and 96 random vectors. The latter use 24 centered Gaussian directions at each amplitude $0.1, 0.25, 0.5, \log 2$, normalized by their maximum absolute coordinate. Candidate seeds are derived from $(20260912, \text{training seed}, \text{epoch})$. For mixed-sign weights the balancing objective is evaluated as

$$\|K(u)\|_F^2 = \frac{1}{2}\sum_{i,j} w_{ij}^2 e^{u_i - u_j} - \frac{1}{2}\,\mathrm{tr}(W^2). \tag{194}$$

Sequential least-squares programming (SLSQP) uses the analytic gradient, a zero initialization, at most 2,000 iterations, and objective tolerance 10^{-12}. This is a convex objective on the prescribed box and gauge hyperplane. The extra 16 candidates are the endpoints reached along eight centered Gaussian rays and their negatives, stopping at the first crossing of $\mu = 0.005$ or at the box boundary. All 192 searches stop at the box boundary. Each checkpoint has 128 distinct valid candidates. Across the 1,536 records, margins lie in $[0.07046, 0.1511]$. Ranking comparisons use the 112 main candidates at each checkpoint.

We freeze the encoder, convert it to FP64, and export the same bias vectors for every metric. We compute the actual equilibrium once per input in the original coordinates. Proxy equilibria use accelerated projected gradient with step $1/\lambda_{\max}(Q)$, followed by an active-set linear solve and full primal, dual, and complementarity checks. There are 2,359,296 proxy-input solves and 18,432 actual solves. The maximum normalized KKT residual is 1.24×10^{-14}; the maximum actual relative fixed-point residual is 6.87×10^{-11}. Projected-gradient iterations have median 50 and maximum 100, followed by one active-set linear solve. Three expressions for the energy gap agree to relative error 3.18×10^{-11}. We compute co-activation from the active set of the quadratic-program solution. At epoch 100, 54.3–56.5% of units are active, and actual and proxy activation states differ on 3.17–5.26% of units.

B.1.1 State error bounds and numerical accuracy

We evaluate all candidates in the original state coordinates and check that solver error does not obscure the proxy error. This requires converting the theoretical bounds to those coordinates and monitoring the complementarity residuals of the computed solutions.

Table 1: Epoch-100 classifiers. Accuracy uses the full test set. The $W = 0$ ablation keeps the encoder and head fixed and reports the accuracy drop in percentage points on 512 of the 4,488 validation images used to monitor training. Iterations are the maximum forward count in the final training epoch.

Seed	Test accuracy (%)	$\|W\|_2$	Max. iterations	$W = 0$ drop (pp)
11	78.63	0.95	48	6.84
22	78.86	0.95	40	11.52
33	78.88	0.95	48	9.18

For each metric, abbreviate $c = c_D$, $z = z^\star$, and $v = v_1(c)$ at $\tau = 1$. Set

$$\mu_x = \lambda_{\min}(D^{1/2}QD^{1/2}), \qquad L_x = \left\|(Q-K)D^{1/2}\right\|_2,$$
$$U_{\Delta,\mathrm{pert}} = \frac{1}{2}(Kz)^\top Q^{-1}(Kz),$$
$$U_\Delta = \min\{U_{\Delta,\mathrm{pert}}, z^\top Qz/2\}, \qquad B_\Delta = \frac{\|v\|_2^2}{2\ell_1}. \tag{195}$$

The proof of lemma 1 gives

$$B_x := \frac{\|v\|_2^2}{L_x^2} \le e_D^2 \le U_x := \frac{2U_\Delta}{\mu_x}. \tag{196}$$

Indeed, write $d = x^\star - \widehat{x}_D$, so $y - z = D^{1/2}d$. Dual feasibility and the quadratic expansion at the proxy give

$$\|v\|_2 \le \left\|(Q-K)D^{1/2}d\right\|_2 \le L_x \|d\|_2,$$
$$\Delta \ge \frac{1}{2}d^\top D^{1/2}QD^{1/2}d \ge \frac{\mu_x}{2}\|d\|_2^2.$$

The inequality $\Delta \le U_\Delta$ then gives the upper bound. Since $\mu_x \ge d_{\min}\mu$ and $L_x^2 \le d_{\max}L_1^2$, these bounds retain the coordinate geometry more closely than the extreme-diagonal conversion in equation (22). Under $D \mapsto aD$ for $a > 0$, both sides of the state interval are unchanged, while the energy interval scales by a.

The energy-to-state conversion has an explicit excess term. Put $H_x = D^{1/2}QD^{1/2}$ and $s = Qz - c$. Since $s^\top z = 0$, the quadratic expansion is $2\Delta = d^\top H_x d + 2s^\top y$. Subtracting the squared state error gives

$$\frac{2\Delta}{\mu_x} - e_D^2 = \frac{d^\top(H_x - \mu_x I)d + 2s^\top y}{\mu_x}. \tag{197}$$

Both terms on the right are nonnegative. The first measures the cost of replacing curvature in the error direction by its minimum; the second comes from proxy slack at coordinates activated by the actual response. This excess vanishes exactly when $H_x d = \mu_x d$ and $s^\top y = 0$. Replacing Δ by U_Δ adds $2(U_\Delta - \Delta)/\mu_x$ to the right-hand side.

We account for proxy solver error before averaging the bounds. For a computed proxy $\widetilde{z}$, write $L_Q = \lambda_{\max}(Q)$, $\alpha = 1/L_Q$, and

$$r = \widetilde{z} - [\widetilde{z} - \alpha(Q\widetilde{z} - c)]_+, \qquad \eta_z = \frac{\|r\|_2}{\alpha\mu}. \tag{198}$$

The projected-gradient map has Lipschitz constant at most $1 - \alpha\mu$. The triangle inequality at its fixed point gives $\|\widetilde{z} - z\|_2 \le \|r\|_2 + (1 - \alpha\mu)\|\widetilde{z} - z\|_2$, hence $\|\widetilde{z} - z\|_2 \le \eta_z$. Put $\widetilde{v} = [(K - Q)\widetilde{z} + c]_+$ and define

$$v_- = (\|\widetilde{v}\|_2 - L_1\eta_z)_+, \qquad k_Q = \sqrt{\lambda_{\max}(K^\top Q^{-1}K)}, \tag{199}$$

$$a_+ = \sqrt{(K\widetilde{z})^\top Q^{-1}(K\widetilde{z})} + k_Q\eta_z, \qquad b_+ = \sqrt{\widetilde{z}^\top Q\widetilde{z}} + \sqrt{L_Q}\eta_z. \tag{200}$$

ReLU nonexpansiveness gives $\|v\|_2 \ge v_-$. Triangle inequalities in the two quadratic norms bound their exact values by a_+ and b_+, respectively. Thus the solver-adjusted intervals are

$$\frac{v_-^2}{2\ell_1} \le \Delta \le \frac{1}{2}\min(a_+^2, b_+^2), \qquad \frac{v_-^2}{L_x^2} \le e_D^2 \le \frac{\min(a_+^2, b_+^2)}{\mu_x}. \tag{201}$$

The norm error of a computed actual equilibrium $\widetilde{x}$ is at most $\eta_a = \|\widetilde{x} - T(\widetilde{x})\|_2/(1 - \|W\|_2)$. Consequently the measured state-error norm has uncertainty at most $\eta_a + \eta_z/\sqrt{d_{\min}}$. The squares of the nonnegative endpoints define the reference interval for the numerical checks. All reported bounds use these solver corrections and FP64 matrix factorizations and hold within the resulting numerical tolerances. The maximum η_a is 1.28×10^{-8}.

Table 2: Spearman correlations with test MSE across 112 metrics per checkpoint. Each cell lists seeds 11, 22, and 33 in that order. Overbars denote input averages.

Predictor	Epoch 10	Epoch 30	Epoch 60	Epoch 100
$\overline{\|K\|_F^2}$	0.950, 0.950, 0.945	0.969, 0.964, 0.959	0.972, 0.971, 0.968	0.977, 0.954, 0.980
$\overline{\|Kz\|_2^2}$	0.982, 0.974, 0.982	0.982, 0.983, 0.989	0.980, 0.985, 0.985	0.987, 0.988, 0.992
$\overline{\|v\|_2^2}$	0.976, 0.969, 0.970	0.977, 0.978, 0.976	0.978, 0.972, 0.983	0.975, 0.979, 0.982
$\widehat{g}(K)$	0.563, 0.855, 0.821	0.770, 0.854, 0.701	0.809, 0.647, 0.723	0.828, 0.833, 0.744
$\overline{B}_\Delta$	0.370, -0.032, 0.468	-0.183, -0.102, 0.437	0.199, 0.664, 0.071	0.400, 0.775, 0.423
$\overline{B}_x$	-0.243, -0.196, 0.060	-0.408, -0.170, -0.231	-0.310, 0.049, -0.100	-0.255, 0.211, -0.147
$\overline{U}_{\Delta,\mathrm{pert}}$	0.949, 0.948, 0.957	0.949, 0.980, 0.972	0.951, 0.925, 0.933	0.952, 0.933, 0.932
$\overline{U}_\Delta$	0.949, 0.948, 0.957	0.949, 0.980, 0.972	0.951, 0.925, 0.933	0.952, 0.933, 0.932
$\overline{U}_x/C_x$	0.961, 0.967, 0.957	0.959, 0.972, 0.964	0.972, 0.955, 0.970	0.967, 0.953, 0.969
$1/\mu_x$	0.931, 0.944, 0.894	0.902, 0.914, 0.918	0.957, 0.926, 0.933	0.953, 0.938, 0.942

Table 3: Spearman correlations with mean energy gap, with the same candidate sets and seed order as Table 2.

Predictor	Epoch 10	Epoch 30	Epoch 60	Epoch 100
$\overline{\|K\|_F^2}$	0.941, 0.960, 0.957	0.967, 0.968, 0.960	0.974, 0.965, 0.973	0.975, 0.955, 0.977
$\overline{\|Kz\|_2^2}$	0.989, 0.982, 0.994	0.986, 0.991, 0.988	0.986, 0.984, 0.990	0.992, 0.990, 0.991
$\overline{\|v\|_2^2}$	0.990, 0.987, 0.990	0.988, 0.987, 0.988	0.986, 0.979, 0.991	0.986, 0.987, 0.990
$\widehat{g}(K)$	0.568, 0.855, 0.821	0.766, 0.859, 0.715	0.808, 0.659, 0.720	0.840, 0.843, 0.760
$\overline{B}_\Delta$	0.414, 0.012, 0.510	-0.166, -0.098, 0.470	0.221, 0.701, 0.105	0.428, 0.801, 0.472
$\overline{B}_x$	-0.212, -0.188, 0.069	-0.398, -0.182, -0.222	-0.308, 0.072, -0.101	-0.246, 0.216, -0.125
$\overline{U}_{\Delta,\mathrm{pert}}$	0.968, 0.950, 0.970	0.957, 0.986, 0.976	0.958, 0.922, 0.930	0.950, 0.929, 0.931
$\overline{U}_\Delta$	0.968, 0.950, 0.970	0.957, 0.986, 0.976	0.958, 0.922, 0.930	0.950, 0.929, 0.931
$\overline{U}_x/C_x$	0.964, 0.961, 0.962	0.964, 0.978, 0.962	0.973, 0.942, 0.968	0.967, 0.948, 0.962
$1/\mu_x$	0.924, 0.935, 0.894	0.904, 0.917, 0.913	0.953, 0.900, 0.936	0.952, 0.933, 0.928

B.1.2 Error ranking

With the numerical accuracy checked, we compare how well each predictor orders the candidate proxies by observed error. We compute each ranking within a fixed checkpoint so that the network and inputs remain the same.

We compute Spearman correlations separately within each checkpoint over the 112 main candidates. Tables 2 and 3 report all three seeds at each evaluated epoch. Predictors use test-input averages. The positive normalization constant C_x does not change the MSE ranks within a checkpoint. The trained matrices have mixed-sign weights and correlated encoder outputs. We apply the general bounds in lemma 1 and the empirical directional excitation $\widehat{g}(K)$ from theorem 2. The cycle and independent-input corollaries are evaluated on the analytic families below.

The predictors in these tables use different parts of the response. The unweighted forcing Kz keeps the observed input direction; $v = [Kz - s]_+$ keeps only the part that violates dual feasibility. The co-activation lower bound replaces this signed violation by a guarantee over rotations on each active component. If a component has p active units, then

$$\theta_C(z) = \frac{\min_{i\in C} z_i^2}{\sum_{i\in C} z_i^2} \le \frac{1}{p}. \tag{202}$$

The coefficient equals $1/p$ for equal active responses and approaches zero when one response vanishes while another stays positive. Proposition 15 quantifies the resulting loss on a common full face. In equation (159), input variation changes the error ordering at fixed curvature. The trained-network comparison varies coordinates under a fixed input population.

At epoch 100, the inverse-curvature baseline has state-error correlations $0.953, 0.938, 0.942$, and the full state upper bound gives $0.967, 0.953, 0.969$, compared with $0.987, 0.988, 0.992$ for forcing. The

residual state lower bound gives correlations $-0.255, 0.211, -0.147$ at the same checkpoints. The two energy upper bounds have identical reported correlations.

The three seeds are independent training runs; Figure 2 shows their full range at each epoch.

B.2 Local response of a recurrent Transformer

The Transformer experiment tests local response errors rather than the global threshold-linear bounds. We specify the recurrent state, check the conditions for the anchored comparison, and evaluate the predicted responses against finite perturbations.

We use the public S1 checkpoint of Logan (2026), at step 49,209, with 57,301,056 unique parameters and 12,900,106,240 training tokens. The checkpoint and code revisions are `25c3a8b1bc0e` and `a3effdd6e4d1`, respectively; the supplied experiment record contains their full identifiers. The recurrence $T(h, e)$ includes the input adapter, all shared core blocks, and recurrence-boundary normalization. The prelude latent e is injected at every step. Adapter outputs match the official forward at depths 1, 2, and 8 with identical initialization. Attention masks and positions are fixed, and derivatives act on the entire token-by-hidden state, including cross-token dependencies. Linear solves use vectorized states, whose Euclidean norm equals the Frobenius norm of the token-by-hidden array.

From the first 20,000 FineWeb-Edu sample-10BT validation documents (Lozhkov et al., 2024), 16,947 have at least 257 tokens. We order their identifiers by a seeded SHA-256 key and select eight pilot documents and 64 main documents. Each main document contributes prefixes of 128 and 256 tokens. Dataset and SmolLM2-135M tokenizer revisions are `87f09149ef47` and `93efa2f097d58`, respectively. Analysis uses FP64 and the mathematical attention backend. The official root-mean-square normalization (RMSNorm) implementation casts its inputs to FP32. We replace this cast with precision-preserving evaluation in all 25 RMSNorm modules, without changing their FP32 outputs. In the pilot, the FP32 fixed-point residual plateaus near 1.7×10^{-6}, whereas dtype-preserving FP64 reaches 2.84×10^{-15} after 256 steps. Both the original and anchored recurrences use this FP64 implementation.

Starting from zero, we declare convergence after four successive relative residuals at most 10^{-6} and a further 16-step confirmation at 10^{-5}. We allow 128 iterations initially and extend to 256 if needed. Of 128 contexts, 126 converge, with median 57 steps and range 46–173. The two failures are both lengths of one document, with final residuals 0.0402 and 0.0286. At each converged basepoint, Jacobian–vector products (JVPs) and vector–Jacobian products (VJPs) define $J = \partial_h T$, $S = (J + J^\top)/2$, and $K = (J - J^\top)/2$. Lanczos uses 40 steps, two starts, and full reorthogonalization, with 80-step rechecks near the threshold. It finds a Rayleigh quotient for S above one on every converged context, with estimated largest eigenvalues approximately 2.5–4.0. These directions establish negative curvature of $I - S$.

The intervention keeps the original equilibrium h^* fixed:

$$T_\gamma(h, e') = (1 - \gamma)h^* + \gamma T(h, e'), \qquad Q_\gamma = I - \gamma S, \quad M_\gamma = I - \gamma J. \tag{203}$$

We test $\gamma = 0.75, 0.5, 0.25$ in descending order. At $\gamma = 0.75$ and 0.5, all 126 contexts have estimated minimum eigenvalues below 0.01. At $\gamma = 0.25$, 125 contexts have estimates in $[0.0194, 0.375]$, and one has an estimate below 0.01. The anchored recurrence also converges after a one-percent state perturbation on all 125 retained contexts. We write $\widehat{\mu} = 1 - \gamma\widehat{\lambda}_{\max}(S)$ for the Lanczos margin estimate. Ritz values estimate $\lambda_{\max}(S)$ from below, so $\widehat{\mu}$ estimates $\lambda_{\min}(Q_\gamma)$ from above. We retain contexts with $\widehat{\mu} \geq 0.01$ and use this estimate in the error-ranking comparison. Table 4 reports the context counts at each stage.

For each context, four orthogonal Gaussian latent directions are scaled to $\|\delta e\|_F = \|e\|_F$. Define

$$q = (\partial_e T)\delta e, \qquad d_E = Q_\gamma^{-1}\gamma q, \qquad d_A = M_\gamma^{-1}\gamma q, \qquad p = M_\gamma^{-1}\gamma K d_E. \tag{204}$$

These are the symmetric linear response, actual linear response, and predicted difference, all computed at the unperturbed state. Whenever the two linear systems are invertible, $M_\gamma d_E = \gamma q - \gamma K d_E$ implies $M_\gamma(d_A - d_E) = \gamma K d_E$, proving $d_A - d_E = p$. This is the same resolvent comparison as the fixed-face analysis. The conjugate-gradient method solves the symmetric system, and the restarted generalized minimal residual method (GMRES) solves the asymmetric systems. Their directly recomputed relative residuals are at most 9.99×10^{-7}. The relative error in $d_A - d_E = p$, normalized by $\|p\|_2$, is at most 7.16×10^{-6}.

Table 4: Context counts for the original and anchored recurrences. Anchoring uses the original equilibria and $\gamma = 0.25$.

Condition	Contexts
Document-prefix pairs	128
Original recurrence: converged	126
Original recurrence: $I - S \succ 0$	0
Anchored: converged with $\widehat{\mu} \geq 0.01$	125
Original recurrence: did not converge	2
Anchored: $\widehat{\mu} < 0.01$	1

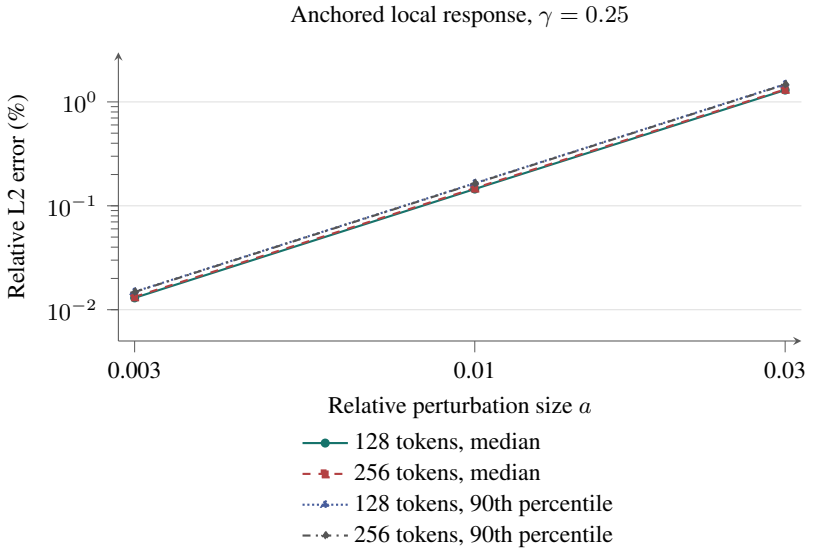


Figure 3: Relative L2 error between the nonlinear response error e_{obs} and the prediction p in equation (204), normalized by $\|p\|_2$. The 125 contexts contribute four directions at each perturbation size.

The response test solves the nonlinear equations independently:

$$h_{\pm}(a) = T_\gamma(h_{\pm}(a), e \pm a\delta e), \qquad e_{\text{obs}}(a) = \frac{h_+(a) - h_-(a)}{2a} - d_E, \quad a \in \{0.003, 0.01, 0.03\}. \tag{205}$$

Both solves start at h^* and hold its anchor fixed. The prediction p uses the Jacobian of the original recurrence at h^* and one asymmetric linear solve per direction. The relative L2 error is $\|e_{\text{obs}} - p\|_2 / \|p\|_2$, using vectorized states. Under a smooth local fixed-point branch, central differencing gives $e_{\text{obs}}(a) = p + O(a^2)$; Figure 3 shows the error as a function of perturbation size. All perturbed solves converge, giving 1,500 direction–amplitude pairs. Tightening tolerances tenfold on eight fixed contexts changes the measured responses by at most 9.89×10^{-7} in relative L2 norm. Table 5 reports relative L2 errors and cosine similarity.

Let $d_{\text{FD}} = (h_+ - h_-)/(2a)$. The computed vectors satisfy

$$e_{\text{obs}} - p = (d_{\text{FD}} - d_A) + (d_A - d_E - p). \tag{206}$$

The first term is the linearization error; the second is the numerical error in the resolvent identity. For exact solves on a smooth fixed-point branch, these terms are $O(a^2)$ and zero, respectively. Table 5 reports prediction error relative to $\|p\|_2$ and linearization error relative to $\|d_A\|_2$.

For scalar comparisons, m is the number of hidden-state entries, $m = 576$ times the prefix length. Thirty-two Rademacher random vectors estimate $\widehat{R} = (32m)^{-1} \sum_r \|Kr\|_F^2$. Table 6 compares the rotation norm with input forcing, symmetric response magnitude, excited rotation, curvature-adjusted

Table 5: Prediction accuracy for the anchored recurrence. Median and 90th percentile summarize $\|e_{\mathrm{obs}} - p\|_2 / \|p\|_2$. Cosine similarity is computed between e_{obs} and p and reported as a median. Linearization is the median of $\|d_{\mathrm{FD}} - d_A\|_2 / \|d_A\|_2$. The 128- and 256-token prefixes contribute 63 and 62 contexts, respectively, with four directions per context at each perturbation size.

Tokens	a	Median (%)	90th percentile (%)	Cosine similarity	Linearization (%)
128	0.003	0.0130	0.0149	1.000000	0.00328
128	0.01	0.1450	0.1650	0.999999	0.03640
128	0.03	1.3000	1.4800	0.999917	0.32700
256	0.003	0.0133	0.0148	1.000000	0.00333
256	0.01	0.1480	0.1640	0.999999	0.03700
256	0.03	1.3200	1.4700	0.999913	0.33200

Table 6: Spearman correlations with the L2 error $\|e_{\mathrm{obs}}\|_2$ of the symmetric response at $a = 0.003$ under anchoring. Brackets are 95% bootstrap confidence intervals over documents. The rotation and forcing predictors use the original K and q; the common factor $\gamma = 0.25$ leaves their ranks unchanged.

Predictor	128 tokens	256 tokens
$\sqrt{\widehat{R}}$	0.638 [0.487,0.738]	0.471 [0.265,0.626]
$\|q\|_2 / \sqrt{m}$	0.67 [0.549,0.750]	0.61 [0.489,0.709]
$\|d_E\|_2 / \sqrt{m}$	0.829 [0.759,0.891]	0.799 [0.712,0.863]
$\|Kd_E\|_2 / \sqrt{m}$	0.997 [0.995,0.998]	0.997 [0.994,0.998]
$\|Kd_E\|_2 / (\widehat{\mu}\sqrt{m})$	0.78 [0.681,0.866]	0.724 [0.593,0.815]
$\|p\|_2 / \sqrt{m}$	1 [1.000,1.000]	1 [1.000,1.000]

forcing, and the full prediction. At $a = 0.003$, excited rotation has correlation 0.997 at both lengths, whereas rotation magnitude gives 0.638 and 0.471. Dividing by the estimated minimum eigenvalue reduces correlation to 0.780 and 0.724; preserving the directional resolvent gives 1.000 to three decimal places. The ordering persists at both larger amplitudes. The symmetric-response baseline gives correlations 0.829 and 0.799, below those of excited rotation and above those of curvature-adjusted forcing. Thus the predictor $\|Kd_E\|_2$ improves on response magnitude alone, while the scalar curvature adjustment reduces ranking accuracy. The full prediction uses the directional action of M_γ^{-1} in equation (204). Bootstrap intervals resample documents as clusters, retaining all their lengths, directions, and amplitudes together, with 2,000 replicates and seed 20260919.

B.3 Analytic and random instances

The trained models combine several sources of error. The following small-dimensional families isolate activation changes, curvature, and graph structure so that each theoretical prediction can be checked against an accurately computed equilibrium.

The artifact uses NumPy, with base seed 20260820 and separate fixed seed offsets for each experiment family. The CSV files, JSON summary, plot data, and numerical macros are generated in one run. The reported errors are the energy gap Δ_τ, ℓ_2 state error $e_{\tau,D}$, and mean squared state error $\mathbb{E}e^2_{\tau,D}$. Averages are over inputs. Relative errors use the observed value as the denominator and are reported as percentages. Cases with zero observed error are reported separately.

For equation (132), the same twelve equally weighted biases are used at 37 equally spaced $\tau \in [-1.8, 1.8]$. All 444 input–parameter pairs satisfy the pointwise energy and state bounds, and the finite-population averages satisfy equations (21) and (22). The exact threshold equation (134) predicts all 216 changes of active set. For nonzero rotation, the maximum relative errors are 50.803% for the Gram lower bound and 74.988% for the cycle lower bound. The maximum fixed-point residual is 2.220×10^{-16}. Applying the fixed-active-set quadratic formula past a switch gives a maximum absolute prediction error of 8.258×10^{-4}. The one-hot control in equation (127) has second moment $I/3$, zero Gram matrix, and exact proxy answers at both endpoints of the same rotation range. The artifact records per-input solutions and bounds in `cross_face_inputs.csv`, and population averages,

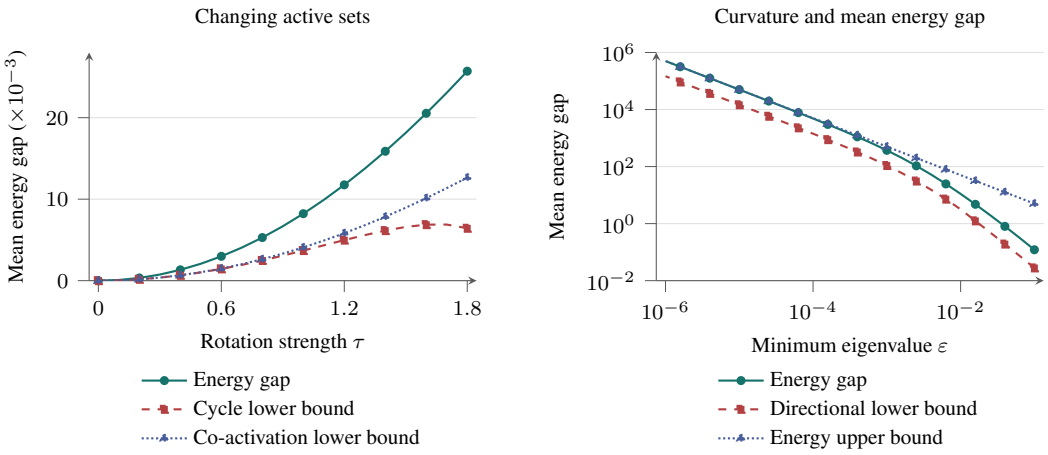


Figure 4: Mean energy gaps and theoretical bounds. Left: co-activation and cycle lower bounds remain positive as actual equilibria leave their proxy activation faces. Right: the directional lower bound and energy upper bound track the energy gap as it grows proportionally to ε^{-1}. Here $\varepsilon = \lambda_{\min}(Q)$, $\|W_\varepsilon\|_2 < 1$, and $\mathrm{cond}_2(D_\star)$ is fixed.

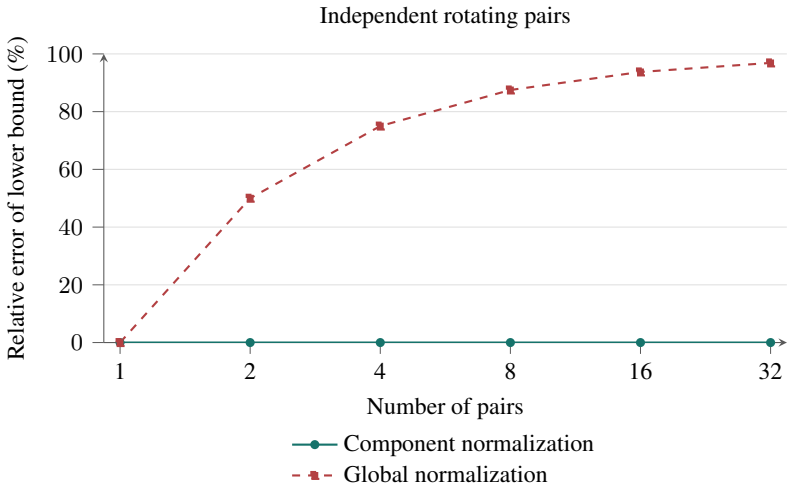


Figure 5: Relative error of the lower bounds for independent rotating pairs. The local bound has the same relative error for every network size. At $\tau = 10^3$, one unit in every pair has deactivated and the local lower bound is within 10^{-6} of the energy gap in relative terms.

Gram matrices, cycle certificates, and activation thresholds in `cross_face_population.csv`. Independent tests also verify the bounds on 800 inputs with noncommuting Q, K, including noncontractive maps and both signs of rotation.

To compare component and global normalization, we evaluate equation (104) at $\eta = 0$ and nine logarithmically spaced values from 10^{-8} to 1. The minimum eigenvalue of the local Gram matrix remains $1/171$. The local lower bound has relative error 71.93%. The global lower bound tends to zero as $\eta \to 0$, while the mean energy gap remains $1/1960$. For equation (128), we use $m \in \{1, 2, 4, 8, 16, 32\}$ and $\tau \in \{0.1, 1, 10, 10^3\}$. At $\tau = 10^3$, the relative error of the local lower bound is $1.000 \times 10^{-4}\%$ for every m; the global lower bound equals $1/m$ times the local bound. The files `coactivation_locality.csv` and `coactivation_modules.csv` record the solutions, eigenvalues, energy gaps, and lower bounds.

To check whether the corrected bound handles nearly disconnected components, use equation (109) with $\eta = 10^{-4}$ and $\tau = 1$. Use $\delta = 0$ and 21 logarithmically spaced bridge strengths from 10^{-12}

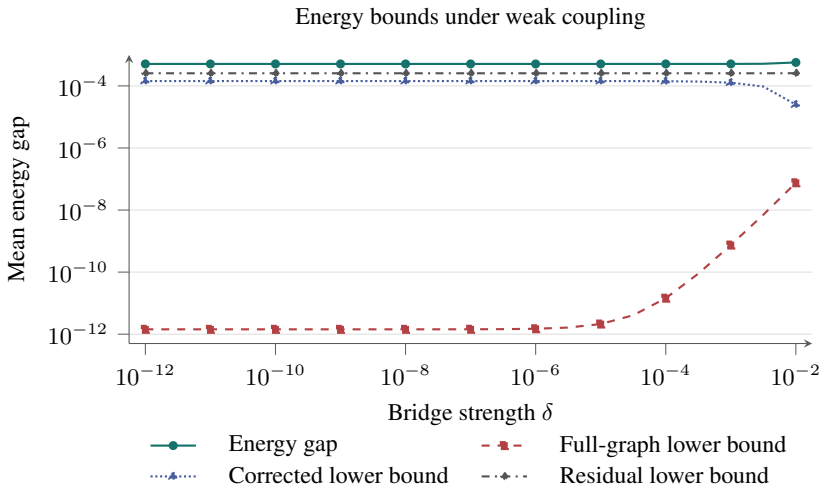


Figure 6: Fixed inputs under weak coupling. The corrected lower bound approaches the disconnected value as the bridge vanishes, while the full-graph bound remains small because its normalization includes the weakly active fourth coordinate.

to 10^{-2}, with the same three equally weighted inputs throughout. The retained graph consists of the triangle edges. All equilibria are fully active. Linear solves give the responses, mean energy gaps, and mean squared state errors at each bridge strength. The full-graph, corrected, and residual lower bounds are compared in Figure 6. At $\delta = 10^{-8}$ the mean gap is 5.102×10^{-4}, the full-graph bound is 1.433×10^{-12}, and the corrected bound is 1.432×10^{-4}. Both corrected energy and state bounds hold at all 22 bridge strengths. The maximum fixed-point residual is 4.441×10^{-16}. The file `coactivation_weak_coupling_population.csv` records the matrices, Grams, remainder term, errors, and bounds; `coactivation_weak_coupling_inputs.csv` records the 66 inputs and responses.

We next test the input-coverage guarantee on sparse graphs. On a cycle of size $n \in \{4, 6, 8\}$, set the two symmetric neighbor weights to 0.2 and the oriented skew weights to 0.15. Unit i receives its fixed positive amplitude, equally spaced between 1 and 1.5 over the coordinates, with probability p; otherwise it receives -1.5. The input coordinates are independent. Here $r = 0.4$, $a = 1$, $b_0 = 1.5$, and $M = 1.5 > rb_0/(1 - r) = 1$. We enumerate all 2^n input masks at $\tau \in \{1, 4\}$ and evaluate their exact probabilities at $p \in \{0.2, 0.5, 0.8\}$. Thus 672 input–parameter pairs give 18 input distributions and parameter settings, with 123 support changes among the distinct pairs. The minimum Gram eigenvalues satisfy the bound $p^2(1 - p)^2/2$. The maximum relative errors are 91.346% for the Gram lower bound and 93.629% for the cycle lower bound of the base matrix $S + K$. The files `coactivation_independent_inputs.csv` and `coactivation_independent_population.csv` record every mask, bias, equilibrium, Gram matrix, weighted mean, the linear-size bounds in equation (124), and the sample count from proposition 13 with $\delta = 0.05$.

To test inputs with mass near zero, use the same cycle matrices and draw 256 independent vectors from $\mathrm{Unif}[-3, 1.5]^n$ for each $n \in \{4, 6, 8\}$, with seed 20260908. The input-event thresholds are $a = 0.5$ and $M = 1.05 > 1$, so $p_+ = 1/4.5$ and $p_- = 1.95/4.5$. Equation (122) gives $\lambda_{\min}(G_\nu) \geq 1.159 \times 10^{-3}$ for this input law. We solve the proxy from each bias and the actual equilibrium at $\tau \in \{1, 4\}$, producing 1536 input–parameter pairs and 267 support changes. Across the distinct proxy inputs, recurrent coupling activates 126 coordinates with nonpositive bias. For each sampled population, the empirical Gram dominates the deterministic local-event count bound. Sample energy comparisons use the empirical Gram; the analytic input law bounds the population eigenvalue. The maximum relative error of the directional lower bound is 90.023%. The files `coactivation_continuous_inputs.csv` and `coactivation_continuous_population.csv` record all inputs, solutions, event counts, Grams, and upper and lower bounds.

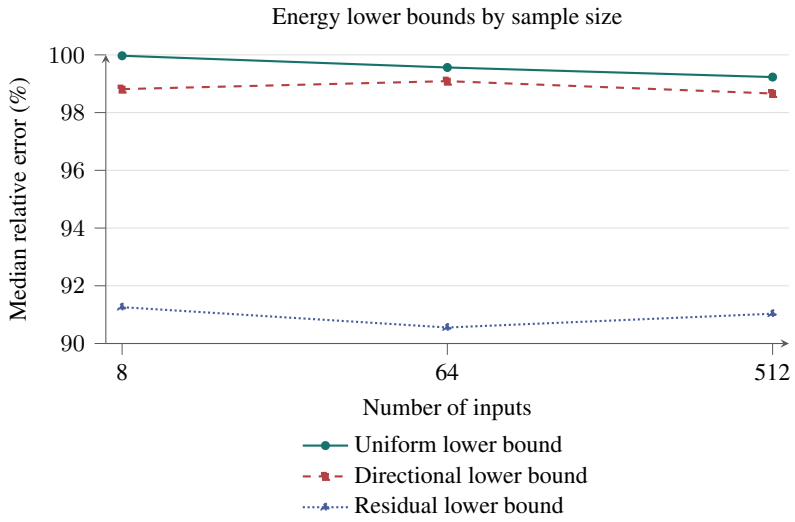


Figure 7: Relative error of the lower bounds at three sample sizes. Each point is the median over the same 20 matrices.

The sparse constructions isolate specific mechanisms; dense random instances test the bounds when several mechanisms interact. This experiment uses seed 20260907 and generates 100 populations in dimension four. In each trial, independently draw standard Gaussian matrices R, T and set $Q = R^\top R + 0.1I$, $K = T - T^\top$. Each of eight proxy responses has independent coordinates drawn uniformly from $[0.1, 2]$ with probability $1/2$ and zero otherwise. Inactive slacks are uniform on $[0, 0.5]$, active slacks are zero, and τ is uniform on $[-8, 8]$. The biases are $c = Qz - s$. All 800 inputs satisfy the pointwise bounds. Of the 100 Gram matrices, 65 are positive definite and 35 are singular. Median relative errors over all 100 trials, in the order of equation (131) are 99.985%, 99.068%, 96.07%, and 91.48%. The directional bound uses the observed rotation, and the residual bound also uses the signed forcing and slack of each input. The medians decrease along this sequence. The 100 trials are in `coactivation_stress_population.csv`; the 800 responses, slacks, biases, and solutions appear in `coactivation_stress_inputs.csv`.

The sample-size comparison fixes the first 20 of these matrices and rotation strengths. It keeps their original eight inputs, appends 504 independent draws from the same response–slack law using seed 20260909, and evaluates nested prefixes of sizes 8, 64, and 512. At 512 inputs, 20 of 20 Grams are full rank. Median relative errors over all 20 matrices are 99.23%, 98.661%, and 91.031% for the uniform, directional, and residual bounds. The residual lower bound has the smallest median relative error at each sample size. All 10,240 inputs are in `coactivation_sample_inputs.csv`; the 60 nested-population summaries are in `coactivation_sample_population.csv`.

To check the exact response formula without activation changes, we evaluate equation (147) at 21 equally spaced $\xi \in [0, 1/2]$ for each $d \in \{0, -2\}$, always using the same six equally weighted inputs $\mathbf{1} \pm e_i/4$ and $\tau = 1$. We compute the average over the complete finite population. All instances satisfy the active-set condition $1 < \tau_0$ in theorem 14. Across 42 instances, the maximum absolute difference between the mean energy gap and equation (141) is 9.324×10^{-18}. The covariance lower bound is attained, as predicted by equation (149). For nonzero circulation, the maximum relative error of the cycle lower bound is 7.9326%; it tends to zero as circulation vanishes. The $d = -2$ cases have $\|W_{\xi,-2}\|_2 > 1$. The CSV records the matrices, responses, eigenvalues, theoretical parameters, and bounds; the fixed biases are in `fixtures.json`. Figure 8 shows the $d = 0$ slice; both variants are in `population_excitation.csv`.

We evaluate the explicit family equation (174) at 26 logarithmically spaced values of $\varepsilon = \lambda_{\min}(Q) \in [10^{-6}, 10^{-1}]$. The bias has fixed unit norm, the condition number of $D_\star$ is 1.2767, and $\|W_\varepsilon\|_2 \le q < 1$ throughout, with $q \approx 0.99713$. The least-defect proxy error reaches 9.996×10^5 at $\varepsilon = 10^{-6}$, whereas identity-proxy error never exceeds 0.010076 on the sweep. At $\varepsilon = 10^{-6}$, the asymptotic prediction $C_\star/\varepsilon$, with $C_\star = 0.99996$, has relative error 0.034723%. All equilibria are

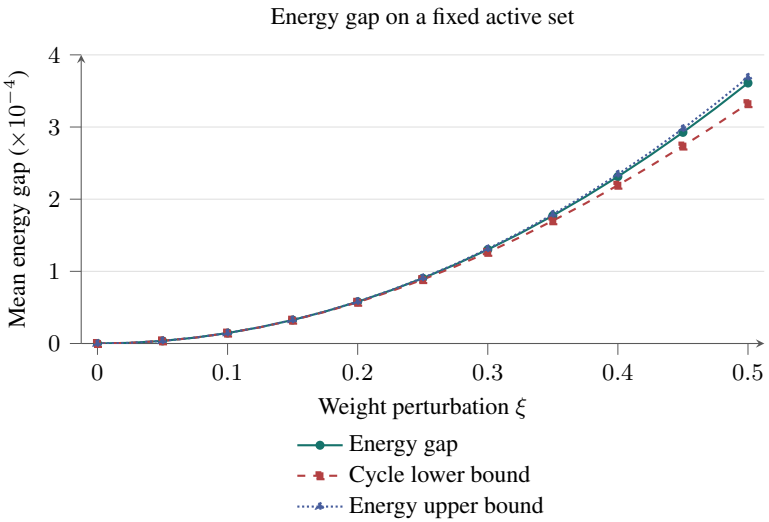


Figure 8: Mean energy gap and theoretical bounds for a fixed active set. The cycle lower bound approaches the mean gap as the perturbation vanishes.

fully active. We use dense linear solves on this active set, with maximum relative linear-system residual 7.063×10^{-11}. Raw values, both defects, minimum eigenvalues, and solutions are in `balancing_separation.csv`.

The six fixed inputs in equation (193) are evaluated at the same 26 values of ε. We compute the minimum Gram eigenvalue as the squared minimum singular value of the stacked feature matrix. The computed eigenvalues satisfy equation (183) throughout. At $\varepsilon = 10^{-6}$, the asymptotic predictions have relative errors 0.069458% for mean squared state error and 0.034937% for the mean energy gap. The directional energy lower bound has relative error 70.648%, and the energy upper bound has relative error 0.034939%. The responses, errors, bounds, and theoretical parameters are in `balancing_population.csv`. Figure 4 compares these bounds. The CSV records the asymptotic predictions and both upper bounds.

Two strict-complementarity families, one with a noncontractive proxy, check the active-set formulas and error bounds. The maximum absolute error in the exact energy formula is 2.638×10^{-16}. Empty and singleton supports are checked separately in `inactive_rotation.csv`.

For 250 four-dimensional trials, form $Q = R^\top R/4 + 0.2I$, $S = I - Q$, and $K = 2(Z - Z^\top)$ with independent standard-normal matrices R, Z, standard-normal bias, and $\tau \sim \mathrm{Unif}[-4, 4]$. We solve the linear complementarity problem (LCP) by enumerating all 16 activation faces. There are 249 trials with $\|S + \tau K\|_2 \geq 1$ and 178 with a different support from the proxy. Every state and energy bound holds; the maximum fixed-point residual is 2.887×10^{-15}. The complete instances and solutions are in `results/energy/noncontractive_proxy.csv`.

We independently perturb directed off-diagonal weights of a six-dimensional symmetrizable matrix by $\exp(\varepsilon Z_{ij})$, using 80 trials at each $\varepsilon \in \{0, 0.025, 0.05, 0.1, 0.2, 0.4\}$. Across 480 matrices, the maximum absolute error in equation (5) is 5.551×10^{-17}. The maximum constraint residual is 1.886×10^{-14}, and the maximum primal–dual gap is 8.327×10^{-17}. We evaluate the cycle lower bound using a cycle basis and weighted Hodge least squares. The two computations agree on every instance; the matrices, certificates, and nonlinear optima are recorded in `cycle_balance.csv`.

B.4 Reproduction details

The analytic experiments share a single entry point and save the quantities needed to check their bounds. The solver tolerances and verification commands below specify how to reproduce these checks independently of the manuscript build.

Run `python3 experiments/run_preliminary.py` from the paper directory. The script uses NumPy, Pillow, and the standard library. The JSON summary records dependency versions, random seeds, the experiment script's SHA-256 hash, and the generated files. The optional `--output` argument selects a separate artifact directory.

Small-dimensional equilibria are computed by active-set enumeration, with tolerance 10^{-10} for the scaled complementarity conditions. The fixed-point residual is $\|x - [Wx + b]_+\|_\infty$. Fully active separation instances use linear solves. For a system $Au = c$ with $c \neq 0$, its relative residual is $\|Au - c\|_2 \,/\, \|c\|_2$. For cycle geometry, Newton iteration with backtracking solves balancing in the zero-mean gauge; a singular value decomposition (SVD) of the incidence matrix gives a cycle basis.

The script checks geometric identities, equilibrium bounds, population certificates, and separation rates. Regression tests cover degenerate faces, singular excitation, energy–state conversion, and metric gauges. The command `python experiments/verify_analytic.py --artifact-root analytic_rerun` compares a separate rerun with the analytic plot sources, using the tolerances documented in the README. A standard pdfLaTeX–BibTeX build compiles the paper from the supplied PDF figures without running the experiments. The trained-network and recurrent-Transformer tables and plots retain the reported experimental aggregates.