THE PRICE OF DIAGONAL MULTIPLIERS IN NEURAL NETWORK CERTIFICATES

Anonymous authors
Paper under double-blind review

ABSTRACT

Many semidefinite Lipschitz certificates for neural networks use diagonal multipliers on activation slope inequalities. We study their conservatism relative to a stated certificate class, using classical polytopic vertex-LMI ideas. For discrete-time neural dynamics, we derive necessary and sufficient vertex conditions for one-step quadratic incremental ℓ_2-gain certificates over the independent-slope abstraction (class-exact). We connect this characterization to robustness margins; its layerwise counterpart is NP-hard to compute. A valid block-multiplier hierarchy contains the diagonal level and reaches the reference under strict quadratic-stability assumptions. For equilibrium models, fixed-point certificates dominate storage-based ones, and their diagonal level contains LBEN. For stability, the stated Lur'e–Postnikov-type Lyapunov family is at least as powerful as quadratic functions. We characterize its full strength over sector-bounded or monotone classes by necessary and sufficient vertex conditions with a constructive, decrease-preserving reduction; slope restrictions can permit strict improvement. Exact witnesses enclose the maximum diagonal/reference class price over 72 feedforward networks in $[1.049355, 1.049357]$, and the maximum diagonal/full-block static class price over 828 equilibrium-model targets in $[1.067848, 1.122833]$. These are class-optimum enclosures; the separate 4.14% reduction from blocks of four compares returned feasible gains. Complete audits of twelve two-layer ReLU networks separate losses, with abstraction loss above 72% in one model. On two width-80 MNIST models, under a 60-second per-target budget, the original block policy loses to diagonal in certified accuracy at $\epsilon = 0.3$.

1 INTRODUCTION

Many robustness certificates for neural networks reduce a question about a nonlinear map to a semidefinite program (SDP). LipSDP bounds the Lipschitz constant of a feedforward network (Fazlyab et al., 2019); equilibrium and recurrent models are certified to be well posed and Lipschitz by linear matrix inequalities of the same type (Revay et al., 2020; 2024; Havens et al., 2023). The diagonal formulations compared here make two approximations. First, each activation is replaced by inequalities on its slope, which forgets which slopes can occur together (the *abstraction*). Second, these inequalities enter the SDP through a *diagonal* multiplier, one scalar per neuron (the *S-procedure*), which may lose information even relative to the abstraction. We measure these losses on small trained networks and random systems.

The coupled LipSDP-Network construction of Fazlyab et al. (2019) was proposed to reduce the second loss, but its cross-neuron constraints are invalid for increments (Pauli et al., 2022): two neurons with the same activation do not have the same *increments* (Revay et al., 2020). Several scalable incremental SDP certificates use diagonal multipliers (Pauli et al., 2024b; Xue et al., 2022). Classical vertex LMIs already characterize quadratic certificate classes (Boyd et al., 1994, Sec. 6.3.2); the question here is how to use an incremental neural-network specialization to quantify the loss.

We specialize this reference construction to discrete-time incremental neural dynamics, connect it to robustness margins, and develop a valid block hierarchy. The *price of diagonal multipliers* is the ratio of the gain certified with diagonal multipliers to the smallest gain that any certificate of the same type can certify (§3 defines it precisely); a price of one means that the diagonal S-procedure loses nothing. Because a single word "exact" would be misleading here, every claim carries one of

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.

Table 1: What is characterized, what is certified, and what remains restricted.

object	established result	scope or remaining restriction
Incremental gain	vertex characterization; conditional block-top equality	independent slopes, quadratic storage
Empirical class prices	certified optimum enclosures for 72 feedforward models and 828 DEQ targets	saved models; DEQ reference is static
Multilayer losses	complete separation on twelve two-layer models	abstraction and storage can both lose
Lur'e–Postnikov stability	sector/monotone characterization; strict slope-restricted example	stated function family; stability only

three labels: exact (it characterizes the property itself, e.g. the true Lipschitz constant), class-exact (it is the best possible within the stated certificate class and activation abstraction), or sound (it is valid but may be conservative).

Contributions and significance.

- *A reference value* (class-exact; §4). Specializing classical vertex LMIs, we characterize one-step quadratic incremental gain certificates over the independent-slope box (Theorem 2), with a gain-to-margin chain (Proposition 4). Computing the layerwise counterpart is NP-hard (Proposition 3).
- *A valid way beyond the diagonal* (sound at every level, class-exact at the top under a strict quadratic-stability condition; §5). Block multipliers with vertex conditions inside each block are valid, refine one another and interpolate between LipSDP-Neuron and the reference (Theorem 5).
- *Equilibrium models* (§6). Certificates on the fixed-point relation are never worse than storage-based ones at any level (Proposition 6); their diagonal level coincides with the condition of Lipschitz bounded equilibrium networks (LBEN; Revay et al. 2020), up to allowing $\lambda_j = 0$ (Remark 7), so our block levels tighten a certificate already in use.
- *Stability* (class-exact, stability only; §7). The stated Lur'e–Postnikov-type Lyapunov family contains quadratic functions and is therefore at least as powerful. Our necessary and sufficient vertex test captures its full strength over sector-bounded or monotone classes: an explicit quadratic reduction preserves the same uniform decrease constant (Theorem 8; App. C). Under slope restrictions, this family can be strictly stronger (Proposition 9).
- *Certified loss measurements* (§8). Maximum class prices lie in $[1.049355, 1.049357]$ over 72 feedforward models and $[1.067848, 1.122833]$ over 828 static DEQ targets. Complete audits of twelve two-layer models separate abstraction and storage losses, including positive storage loss in two MNIST models. Thus multiplier richness, realizable slopes and storage address distinct restrictions; the numerical 4.14% reduction compares returned DEQ gains.

2 Related work

SDP certificates with diagonal multipliers. LipSDP (Fazlyab et al., 2019) has been extended to new activation constraints (Pauli et al., 2024b) and general architectures (Pauli et al., 2024a; 2023) and scaled by decomposition (Xue et al., 2022; Wang et al., 2024); incremental dissipativity of layers (Aquino et al., 2022) and incremental quadratic constraints (Hashemi et al., 2021) also use diagonal multipliers. Wang & Manchester (2023, Remark 2.2) also point out that the claim that the coupled LipSDP-Network multipliers tighten the bound is false. Equilibrium models are certified by the LBEN condition (Revay et al., 2020; Havens et al., 2023), monotone operator theory (Winston & Kolter, 2020), semialgebraic models (Chen et al., 2021) and non-Euclidean contraction (Jafarpour et al., 2021; 2022).

How loose are relaxations? Input-specific attack relaxations have LP barriers (Salman et al., 2019) and SDP exactness conditions (Zhang, 2020; Raghunathan et al., 2018). We ask the analogous

question for global SDP certificates, with the multiplier class in the role of the relaxation. Kuang et al. (2026) studies reachability failures and reports that optimizing diagonal reparameterizations of product-of-norms bounds captures much of SDP tightness in examples. For one hidden ReLU layer, the vertex reference is no larger than this optimized product bound and can be strictly smaller (App. G). Exact Lipschitz computation is NP-hard (Scaman & Virmaux, 2018); mixed-integer methods handle small ReLU networks (Jordan & Dimakis, 2020).

Lyapunov functions in machine learning. At ML venues Lyapunov functions appear mainly as *learned* certificates: a network is trained to be a Lyapunov function for a given control system and then checked by a solver (Chang et al., 2019; Zhou et al., 2022; Yang et al., 2024; Dawson et al., 2023). Our use runs the other way: the network is the system, the Lyapunov function is the analysis object, and we ask which class suffices rather than how to learn one.

Control theory and other certificates. Vertex LMIs already optimize quadratic certificates for polytopic inclusions (Boyd et al., 1994, Sec. 6.3.2). Our reference specializes this framework to discrete-time incremental network models and their margin chain. Our strict dynamic full-block equivalence specializes Scherer (2001); we supply the neural-network construction and verified loss measurements. IQCs (Megretski & Rantzer, 1997) and Zames–Falb multipliers (Willems & Brockett, 1968; Carrasco et al., 2016; Ahmad et al., 2015) are classical; the last exploit temporal relations outside our one-step quadratic/static reference. Broader Lyapunov families admit sufficient tests (Drummond & Valmorbida, 2023); our stability theorem characterizes the stated restricted family over sector/monotone classes. Table 14 summarizes prior work; App. G discusses reported scales. Bound propagation (Wong & Kolter, 2018; Zhang et al., 2018), abstract interpretation (Gehr et al., 2018), complete verification (Katz et al., 2017) and randomized smoothing (Cohen et al., 2019) certify individual inputs; ours are global certificates.

3 Preliminaries and problem statement

Notation. $|\cdot|$ is the Euclidean norm and $\|\cdot\|_2$ the spectral norm; $|v|_P^2 = v^\top P v$; $\succ$ ($\succeq$) denotes positive (semi)definiteness; $\mathrm{diag}(k)$ is the diagonal matrix with diagonal k; e_j is the j-th unit vector; for a quantity q along two trajectories or at two inputs, $\Delta q = q - \tilde{q}$. Apps. B–C contain the proofs; App. A (Table 4) is the roadmap.

Three kinds of networks. A *feedforward network* with H hidden layers computes $x_0 = x$, $x_{\ell+1} = \phi_\ell(W_\ell x_\ell + b_\ell)$ for $\ell < H$ and $F(x) = W_H x_H + b_H$. A *deep equilibrium model* (DEQ) computes $z = \phi(Wz + Ux + b)$ and $F(x) = Vz + c$. A *recurrent update* $x_{k+1} = Ax_k + B\phi(C^\top x_k) + Ew_k$ with state $x_k \in \mathbb{R}^n$ reads an input sequence w_k. The activations ϕ act coordinatewise on m neurons.

One form for all three. All three are instances of the discrete-time Lur'e system

$$x_{k+1} = Ax_k + Bf_k(C^\top x_k) + Ew_k, \qquad y_k = C_y x_k + D_y f_k(C^\top x_k) + D_w w_k, \tag{1}$$

a linear system in feedback with the activations; Araujo et al. (2023) likewise view their unified Lipschitz-network layer structure as a Lur'e system. The recurrent update is (1) directly. A DEQ becomes (1) in the coordinates $\sigma = Wz + Ux + b$, with $A = 0$, $B = W$, $C = I$, $E = U$, $C_y = 0$, $D_y = V$, $D_w = 0$ and constant input $w = x$; its fixed points correspond one-to-one (Lemma 14). Each layer of a feedforward network is one step of (1) with $A = 0$ and $C = W_\ell^\top$, and a layer-dependent storage (§4.1). Constant biases are suppressed in (1); they cancel in incremental differences. This common form also lets us test the incremental theory on random systems.

Activations and the abstraction. For $\alpha_j \le \beta_j$ let $\mathfrak{B} = \{\mathrm{diag}(k) : k_j \in [\alpha_j, \beta_j]\}$ and let $\mathfrak{V}$ be its 2^m vertices. A scalar map φ is *slope-restricted* in $[\alpha, \beta]$ if $\alpha \le (\varphi(s) - \varphi(t))/(s - t) \le \beta$ for all $s \ne t$, and *sector-bounded* in $[\alpha, \beta]$ if $\varphi(0) = 0$ and $\alpha s^2 \le s\varphi(s) \le \beta s^2$. ReLU, $\tanh$ and hard-$\tanh$ are slope-restricted in $[0, 1]$, leaky ReLU in $[a, 1]$. We write $\mathcal{S}$ ($\mathcal{S}^{\mathrm{tv}}$) for decoupled maps $f = (\varphi_1, \ldots, \varphi_m)$ with slope-restricted, possibly different (and possibly time-varying) components. The abstraction behind every certificate in this paper is the following elementary fact.

Lemma 1 (Incremental secant). *If $f \in \mathcal{S}^{\mathrm{tv}}$, then for every time and all $\sigma, \tilde{\sigma}$ there is $K \in \mathfrak{B}$ with $f(\sigma) - f(\tilde{\sigma}) = K(\sigma - \tilde{\sigma})$. Conversely, $\sigma \mapsto K\sigma$ lies in $\mathcal{S}$ for every $K \in \mathfrak{B}$.*

The abstraction lets each neuron take any slope in its interval independently of the others, and of other times and pairs of points; this is what can make it loose (proof of Lemma 1 in App. B.1). A sector bound gives only $f(\sigma) = K\sigma$, a secant through the origin, so sector results give stability, and only slope restriction gives incremental (robustness) statements. We call a result a robustness certificate only when the whole chain is proved: slope model (Lemma 1) $\rightarrow$ incremental storage or fixed-point inequality $\rightarrow$ certified gain $\rightarrow$ margin (Proposition 4).

Problem statement. Fix a model of the form (1) and a certificate type: a quadratic storage (for DEQs, a quadratic inequality on the fixed-point relation) combined with static multipliers for the slope inequalities of Lemma 1. Let γ_{diag} be the infimum of the gains certifiable with diagonal multipliers and γ_{ref} the infimum over all certificates of the type. We ask whether γ_{ref} can be computed, and how large the *price of diagonal multipliers*, $\gamma_{\mathrm{diag}}/\gamma_{\mathrm{ref}} \geq 1$, is compared with the residual ratio $\gamma_{\mathrm{ref}}/\gamma_{\mathrm{true}}$, where γ_{true} is the true Lipschitz constant, when it can be computed. The reference is γ_{vx} of Theorem 2 for recurrent and random systems and its layerwise version for feedforward networks (§4.1); for DEQs we use the static full-block reference and certify bounds on its price without solving that reference SDP (§8). For stability the price is $\bar{\mu}_{\mathrm{vx}}/\bar{\mu}_{\mathrm{diag}}$, the ratio of the largest sector bounds certified at the vertex and diagonal levels. Ratios are reported for finite values with positive denominators.

4 The reference value

For two trajectories of (1) write $\Delta x = x - \tilde{x}$, etc. For $P = P^\top$ and $g \geq 0$ consider the one-step incremental supply inequality

$$\Delta x^{+\top} P \Delta x^{+} - \Delta x^\top P \Delta x \leq g|\Delta w|^2 - |\Delta y|^2. \tag{2}$$

With $A_K = A + BKC^\top$ and $C_{y,K} = C_y + D_y K C^\top$, let $\mathcal{N}_K(P, g)$ be the symmetric matrix of the quadratic form of (2) in $(\Delta x, \Delta w)$ for the linear map $f = K\sigma$:

$$\mathcal{N}_K(P,g) = \begin{bmatrix} A_K^\top P A_K - P + C_{y,K}^\top C_{y,K} & A_K^\top P E + C_{y,K}^\top D_w \\ E^\top P A_K + D_w^\top C_{y,K} & E^\top P E + D_w^\top D_w - gI \end{bmatrix}.$$

Theorem 2 (Vertex characterization of quadratic incremental dissipativity; class-exact). *Let $P \succeq 0$ and $g \geq 0$. The following are equivalent: (a) (2) holds for all $f \in \mathcal{S}^{\mathrm{tv}}$, all times and all pairs of states and inputs; (b) (2) holds for all linear $f = K\sigma$ with $K \in \mathfrak{B}$; (c) $\mathcal{N}_K(P, g) \preceq 0$ for all $K \in \mathfrak{V}$. Under (c), every pair of trajectories satisfies $\sum_{k<T} |\Delta y_k|^2 \leq g \sum_{k<T} |\Delta w_k|^2 + \Delta x_0^\top P \Delta x_0$ for all T. Consequently $g_{\mathrm{vx}}^\star := \inf\{g : \text{(c) holds for some } P \succeq 0\}$ is the smallest squared incremental ℓ_2 gain certifiable by a positive semidefinite quadratic storage through (2) over $\mathcal{S}$ or $\mathcal{S}^{\mathrm{tv}}$.*

The reference gain is $\gamma_{\mathrm{vx}} = (g_{\mathrm{vx}}^\star)^{1/2}$. *Proof idea.* By Lemma 1, $\Delta f = K_\star C^\top \Delta x$ with $K_\star \in \mathfrak{B}$, and the slack of (2) is a constant minus squared norms of affine functions of K, hence concave in K and minimized at a vertex (App. B.1). □

$g_{\mathrm{vx}}^\star$ is not the true incremental gain of a given network, because the abstraction lets slopes vary independently; nor is it the best gain certifiable with non-quadratic storage or dynamic multipliers. It is the best possible value *for the type of certificate that SDP methods use*: (b) and (c) are necessary as well as sufficient for (2), which makes it the right reference for the S-procedure: every static-multiplier certificate with quadratic storage certifies at least $g_{\mathrm{vx}}^\star$ (Theorem 5(i)), so the price is at least one. The continuous-time, non-incremental analogue is the textbook polytopic bounded-real lemma (Boyd et al., 1994, §6.3.2).

4.1 Feedforward networks, an exact case, and hardness

For a feedforward network with vertex sets $\mathfrak{V}_\ell$ per hidden layer and a weight $O \succeq 0$ on the output: if $g \geq 0$ and $P_1, \ldots, P_H \succeq 0$ satisfy, with $P_0 = gI$,

$$(KW_\ell)^\top P_{\ell+1} (KW_\ell) \preceq P_\ell \;\; (K \in \mathfrak{V}_\ell,\ \ell < H), \qquad W_H^\top O W_H \preceq P_H, \tag{3}$$

then $(F(x) - F(\tilde{x}))^\top O(F(x) - F(\tilde{x})) \le g|x - \tilde{x}|^2$, and the vertex conditions are necessary for the per-layer storage inequalities (Theorem 12; class-exact within layerwise quadratic storage). We write γ_{vx} also for $(g^\star)^{1/2}$ with $g^\star$ the optimal g here. For one hidden layer the optimum is $\max_{K \in \mathfrak{V}_0} \|O^{1/2} W_1 K W_0\|_2^2$, the derivative-box bound of Scaman & Virmaux (2018) and Latorre et al. (2020). For one hidden ReLU layer with W_0 of full row rank this value is exact: every activation pattern $K \in \{0,1\}^m$ is realized on an open set of inputs (solve $W_0 x + b_0 = s$ for any sign vector s), so $\max_K \|W_1 K W_0\|_2$ is attained by the Jacobian on a linear region and equals the Lipschitz constant. With W_0 rank-deficient (e.g. two-dimensional inputs) some patterns are not realizable and the gap can be large (§8).

Proposition 3 (NP-hardness of the layerwise class-exact value)**.** *Computing the optimal g in* (3) *is NP-hard, already for one hidden layer, scalar output and slopes in* $[0, 1]$.

The reduction is from MAX-CUT with $W_0 = [N^\top; -N^\top]$ for the signed incidence matrix N and $W_1 = \mathbf{1}^\top$, which gives $g^\star = 4\,\mathrm{MaxCut}$ (App. B.2). It concerns the certificate value, not the Lipschitz constant of the network, whose hardness is due to Scaman & Virmaux (2018).

4.2 Certified radii from class-pair margins

Proposition 4 (Margin certificate)**.** *Let $F : \mathbb{R}^d \to \mathbb{R}^{n_c}$ be a classifier with n_c classes and $|(F_y - F_j)(x) - (F_y - F_j)(\tilde{x})| \le \gamma_{yj}|x - \tilde{x}|_2$ for all $x, \tilde{x}$ and $j \ne y$, with $n_c \ge 2$ and $\gamma_{yj} > 0$. If $F_y(x) > F_j(x)$ for all $j \ne y$, then $\arg\max_i F_i(x + \delta) = \{y\}$ whenever $\|\delta\|_2 < r(x) = \min_{j \ne y}(F_y(x) - F_j(x))/\gamma_{yj}$. If F is evaluated with error, the margins must be reduced by a bound on that error (fixed-point error and floating-point evaluation); for a DEQ with $|\hat{z} - z^\star(x)| \le e_z$, the fixed-point part is $|(e_y - e_j)^\top V|\, e_z$. For a DEQ, F must be defined everywhere, i.e. a fixed point must exist for every input.*

Gains γ_{yj} follow from Theorem 2, Proposition 6 or (3) with the output selector $L = (e_y - e_j)^\top$ (proof in App. B.3). Since L and $-L$ give the same constraints, $n_c(n_c - 1)/2$ SDPs suffice (45 for MNIST).

5 A valid block-multiplier hierarchy

The reference costs 2^m LMIs. To measure how much of the gap between the diagonal certificate and the reference can be closed at lower cost, we need certificates in between that remain valid. Write $\xi = (\Delta x, \Delta f, \Delta w)$, $\Psi\xi = (\Delta f, C^\top \Delta x) \in \mathbb{R}^{2m}$, and let $\mathcal{F}(P, g)$ be the matrix of the form $\Delta x^{+\top} P \Delta x^+ - \Delta x^\top P \Delta x + |\Delta y|^2 - g|\Delta w|^2$ with $\Delta x^+ = A\Delta x + B\Delta f + E\Delta w$ and $\Delta y = C_y \Delta x + D_y \Delta f + D_w \Delta w$. For a partition $\mathcal{P}$ of the neurons, $\Pi(\mathcal{P})$ is the set of block-diagonal $\Pi = \sum_b S_b^\top \Pi_b S_b$, where S_b extracts $(\Delta f_b, \Delta \sigma_b)$ and each symmetric $\Pi_b \in \mathbb{R}^{2|b| \times 2|b|}$ satisfies

$$(\Pi_b)_{11} \preceq 0, \qquad [K_b; I]^\top \Pi_b [K_b; I] \succeq 0 \ \text{ at the } 2^{|b|} \text{ vertices } K_b \text{ of the block box.} \tag{4}$$

The diagonal family Π_{diag} uses $\lambda_j \begin{bmatrix} -1 & (\alpha_j + \beta_j)/2 \\ (\alpha_j + \beta_j)/2 & -\alpha_j \beta_j \end{bmatrix}$, $\lambda_j \ge 0$, per neuron (LipSDP-Neuron). Let $g^\star(\mathcal{P})$ and $g^\star_{\mathrm{diag}}$ be the infimum of g with $P \succeq 0$ and $\mathcal{F}(P, g) + \Psi^\top \Pi \Psi \preceq 0$. The conditions (4) couple neurons only through the vertices of their joint slope box, never through a repeated-nonlinearity assumption; this is what distinguishes them from the invalid LipSDP-Network multipliers. Strict top-level losslessness specializes Scherer (2001, App. B, Theorem 8); block sparsity and $(\Pi_b)_{11} \preceq 0$ enable finite vertex checks (App. G).

Theorem 5 (Hierarchy)**.** *(i) (Soundness, sound) Every feasible (P, g, Π) satisfies Theorem 2(a); hence $g^\star_{\mathrm{vx}} \le g^\star(\mathcal{P})$. (ii) (Monotonicity) If $\mathcal{P}'$ refines $\mathcal{P}$ then $\Pi(\mathcal{P}') \subseteq \Pi(\mathcal{P})$, so $g^\star(\mathcal{P}) \le g^\star(\mathcal{P}')$. (iii) (Diagonal level) $\Pi_{\mathrm{diag}} \subseteq \Pi(\text{singletons})$, and, when $\alpha_j < \beta_j$, the singleton family is exactly the diagonal multiplier plus the two linear constraints $\Delta f_j \Delta \sigma_j \ge \alpha_j \Delta \sigma_j^2$ and $\Delta f_j \Delta \sigma_j \le \beta_j \Delta \sigma_j^2$ (the inclusion holds for $\alpha_j = \beta_j$ as well, but the characterization fails there). (iv) (Top level, class-exact) If some $P_s \succ 0$ satisfies $P_s - A_K^\top P_s A_K \succ 0$ for all $K \in \mathfrak{V}$, then $g^\star(\{\{1, \dots, m\}\}) = g^\star_{\mathrm{vx}}$.*

Proofs are in App. B.4. Part (i) holds because the block quadratic form is concave in K_b when $(\Pi_b)_{11} \preceq 0$, so its minimum over the block box is attained at a vertex. For part (iv), strict vertex

feasibility extends over the box by convexity. Scherer's theorem on the graphs $\mathrm{im}[K; I]$ yields $\mathcal{F} + \Psi^\top \Pi \Psi \prec 0$; restricting to $\xi = (0, \Delta f, 0)$ gives $\Pi_{11} + B^\top P B + D_y^\top D_y \prec 0$, hence $\Pi_{11} \prec 0$. Every $g > g^\star_{\mathrm{vx}}$ is therefore one-block feasible, proving equality of infima without a rank assumption on C. The conditions (4) apply verbatim to stability certificates over the sector class, with $\sigma = C^\top x$ in place of $\Delta\sigma$; for feedforward networks the hierarchy is applied layer by layer (Proposition 13). A block of size b contributes 2^b LMIs of size b and a $2b \times 2b$ variable; we use contiguous blocks for synthetic systems and a greedy partition by $|W_{ij}| + |W_{ji}|$ for DEQs, with $b \in \{1, 2, 4\}$.

Verified computation. Interior-point solvers satisfy LMIs only up to a tolerance, so we never report a solver value as a certificate. Each SDP is solved with increasing strict margins $\varepsilon \in (10^{-7}, \ldots, 10^{-3})$ until the returned (P, Π, g), read as exact dyadic rationals with g inflated by $1 + 10^{-6}$, satisfies every constraint in exact integer arithmetic; positive definiteness is certified by a floating-point Cholesky factor whose exact residual is bounded below the shift, a self-contained variant of verified Cholesky tests (Rump, 2006) (Algorithm 1 in App. E).

6 Equilibrium models

For the DEQ in σ-coordinates, Theorem 2 gives a Lipschitz bound $|L(F(x) - F(\tilde{x}))| \le \sqrt{g}|x - \tilde{x}|$ by applying (2) at the fixed points, where $\Delta\sigma^+ = \Delta\sigma$ and the storage terms cancel (Theorem 15). This asks for more than necessary: a one-step decrease of the storage along *arbitrary* iterates, whereas the Lipschitz constant only concerns fixed points. With $\xi = (\Delta z, \Delta x)$, $\Psi_s \xi = (\Delta z, W\Delta z + U\Delta x)$ and $\mathcal{G}(g) = \mathrm{diag}(V^\top L^\top L V, -gI)$, we impose the constraint on the fixed-point relation only.

Proposition 6 (Static fixed-point certificates)**.** *Let Π be an admissible multiplier at any level of the hierarchy of §5 (block or diagonal). (i) (sound) If $\mathcal{G}(g) + \Psi_s^\top \Pi \Psi_s \preceq 0$, then $|L(F(x) - F(\tilde{x}))| \le \sqrt{g}\,|x - \tilde{x}|$ for all $x, \tilde{x}$ and all fixed points at x and $\tilde{x}$. (ii) If the inequality is strict, each input has at most one fixed point. (iii) The optimal values are monotone along the hierarchy, as in Theorem 5. (iv) If $g \ge 0$ and U has full row rank, $U\Delta x$ can be eliminated exactly, replacing gI by $g(UU^\top)^{-1}$. (v) (Dominance) Every storage certificate $(P \succeq 0, g, \Pi)$ of §5 for the σ-coordinate system yields a static certificate with the same g and Π, so static certificates are never worse at any level.*

Proofs are in App. B.5. Part (v) follows because storage terms cancel on $\Delta\sigma = W\Delta z + U\Delta x$. The DEQ reference is the static full-block optimum; its equality with the independent-slope fixed-point gain remains open. Part (ii) gives uniqueness, not existence. Existence for every input follows if $P \succ 0$ and $0 \le \rho < 1$ satisfy $(WK)^\top P (WK) \preceq \rho^2 P$ at every slope vertex: the σ-iteration is then a contraction. The monotone parameterization also ensures existence (Winston & Kolter, 2020).

Remark 7 (Relation to LBEN). For slopes in $[0, 1]$ and $\Lambda = \mathrm{diag}(\lambda) \succ 0$, the static diagonal certificate with gain γ is feasible if and only if $2\Lambda' - \Lambda' W - W^\top \Lambda' - \frac{1}{\gamma} V^\top L^\top L V - \frac{1}{\gamma} \Lambda' U U^\top \Lambda' \succeq 0$ with $\Lambda' = \Lambda/(2\gamma)$, which is the condition of Havens et al. (2023, Thm. 3.1) for the LBEN of Revay et al. (2020) (Schur complement, App. B.5); our diagonal level also admits $\lambda_j = 0$. The block levels therefore tighten that certificate without changing the certified object.

7 Stability: Lyapunov functions beyond quadratic storage

The dynamic certificates use quadratic storage. For stability, the larger *Lur'e–Postnikov* (Popov-type) class adds the terms $\sum_j \theta_j \int_0^{c_j^\top x} \varphi_j$, with c_j the columns of C, and exploits that the activation is time-invariant. Since $\theta = 0$ is admissible it certifies at least as much as a quadratic function; this section asks when it certifies strictly more. Consider (1) without input or output, $x^+ = Ax + Bf(C^\top x)$, A Schur, with f decoupled and sector-bounded in $[0, \mu_j]$ (class $\mathcal{A}$), monotone and sector-bounded ($\mathcal{M}$), or slope-restricted in $[0, \mu_j]$ with $\varphi_j(0) = 0$; here $\mathfrak{B}$ has vertices $\{0, \mu_1\} \times \cdots \times \{0, \mu_m\}$. A function V is a *strict certificate* over a class if $V(x^+) - V(x) \le -\varepsilon|x|^2$ for all x and all f in it, with one $\varepsilon > 0$ for the whole class.

Theorem 8 (Vertex characterization and collapse; class-exact, stability only)**.** *For $P \succ 0$, $x^\top P x$ is a strict certificate over $\mathcal{A} \iff$ over $\mathcal{M} \iff$ over the slope-restricted class $\iff$ $P - A_K^\top P A_K \succ 0$ for all $K \in \mathfrak{V}$. Moreover, for $\theta = \mathrm{diag}(\theta_j) \succeq 0$, if $V(x) = x^\top P x + \sum_j \theta_j \int_0^{c_j^\top x} \varphi_j$ is a strict certificate over $\mathcal{M}$ with constant ε, then $\widetilde{P} = P + \frac{1}{2} C \theta \mu C^\top$, $\mu = \mathrm{diag}(\mu_j)$, is a strict quadratic*

certificate over $\mathcal{A}$ with the same ε: a Lur'e–Postnikov certificate over $\mathcal{A}$ or $\mathcal{M}$ exists if and only if those same vertex LMIs are feasible.

Both parts are proved in App. C.1–C.2. Beyond the classical quadratic vertex test (Boyd et al., 1994, §5.1), the explicit $\widetilde{P}$ preserves the decrease constant and removes the integral terms. The vertex LMIs therefore decide whether *any* such certificate exists (the generalized Lur'e problem). They suffice for, but do not characterize, absolute stability: other analyses (Jury & Lee, 1964; Ahmad et al., 2015) can prove more under monotonicity or slope restriction. Narrowing the class matters:

Proposition 9 (Lur'e–Postnikov functions help over slope-restricted activations; stability only)**.** *There is a single-nonlinearity system ($n = 2$, $m = 1$) with a strict Lur'e–Postnikov certificate over the activations slope-restricted in $[0, \mu]$ with $\varphi(0) = 0$ for every $\mu \leq 1.6666$, while no (strict or non-strict) quadratic certificate exists for any $\mu \geq 1.228$. Both directions are verified in exact rational arithmetic, the second by a dual certificate of infeasibility (App. C.3).*

Diagonal multipliers are already strictly conservative for stability at $m = 2$, by printed rational certificates (App. C.4). On 120 random single-loop systems ($n \leq 4$), Lur'e–Postnikov functions raised the certified slope bound by more than 1% in 12.5% of instances and by up to 29%, never for monotone activations (App. F).

8 Experiments

The networks are small by design: the question is how much the multiplier relaxation loses, so we work at sizes where the reference, which costs 2^m LMIs, is numerically tractable. ReLU region enumeration is complete for the one-hidden-layer cases. We ask three questions. *Q1*: how large is the price on random systems, and how does it change with the number of neurons (E1)? *Q2*: on trained feedforward networks, how does the price compare with the abstraction gap (E2)? *Q3*: on trained equilibrium models, how large is the price, and does tightening the multipliers change certified accuracy (E3)? All experiments run on one CPU (Apple M1 Pro; Clarabel and SCS via CVXPY), and solve-plus-verify times are in Table 13. Algorithm 1 checks feasibility, not optimality. The original E1–E3 gain-ratio summaries use returned gains (solver values in E1, verified feasible values in E2–E3), not class infima. Their ratios alone do not certify class prices; the separate enclosures use verified upper and lower bounds (App. H). Apps. E–F give setup and full results. Failures are counted, and each experiment includes a falsifier.

E1: random systems (Q1). Nine cells of random instances of (1), 100 each: Gaussian systems with $(n, m) \in \{(4, 2), \ldots, (12, 8)\}$ and systems with $A = 0$, $C = I$ (the DEQ iteration, analysed with storage certificates) with $m \in \{2, 4, 6, 8\}$. We bisect the largest certified sector bound $\bar{\mu}$ at each level (stability only) and compute the incremental gain at slopes $[0, 0.8\bar{\mu}_{\mathrm{diag}}]$. The prices are $\gamma_{\mathrm{diag}}/\gamma_{\mathrm{vx}}$ (gain) and $\bar{\mu}_{\mathrm{vx}}/\bar{\mu}_{\mathrm{diag}}$ (stability). Falsifiers: a level certifying more than the vertex level, and non-monotone levels.

E1 records report all 900 instances verified exactly at every level, with no ordering-check violations at the stated tolerances (Table 6). The returned one-block and vertex values matched within numerical tolerances in every instance, as Theorem 5(iv) and its stability analogue predict. Across sampled cells, where n and m vary jointly, the fraction of Gaussian systems with a gain price above 1.005 rises from 0.39 at $m = 2$ to 0.90 at $m = 8$, and systems with $A = 0$ follow the same trend. The losses stay moderate in the median, 1.055 per cell at worst, but the largest returned-gain ratio is 1.765, on a two-neuron system (Table 6). An exact audit certifies $[1.764539, 1.764543]$ for a saved reconstruction of that seed, not the historical population maximum. Pair blocks close at least half of the gap in fewer sampled instances (0.42 at $m = 3$, 0.09 at $m = 8$); whether larger blocks resolve this trend remains a hypothesis.

E2: feedforward networks (Q2). Iris (Fisher, 1936), two moons, and PCA-50 MNIST (LeCun et al., 1998) and Fashion-MNIST (Xiao et al., 2017) (gains in PCA coordinates); widths (8), (16), $(8, 8)$; ReLU and $\tanh$; three seeds; 72 networks. Bounds: product of spectral norms, LipSDP-Layer and LipSDP-Neuron, the diagonal, singleton, pair and quadruple levels applied layer by layer (Proposition 13), and the vertex value (3); for ReLU networks also γ_{true} by enumerating linear regions, and for all networks a sampled lower bound (App. E). Falsifier: a lower bound above a verified upper bound.

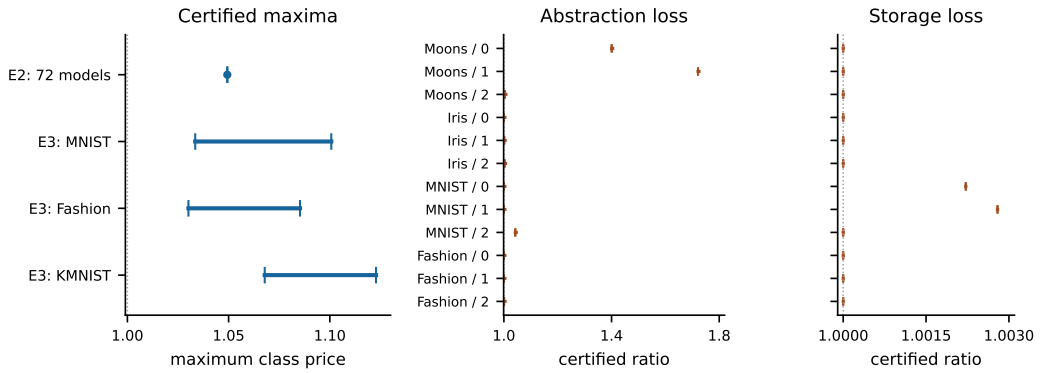

Figure 1: Certified evidence from saved models. *Left*: maximum diagonal/reference class-price enclosures for E2 (72 models, layerwise reference) and each E3 dataset (276 targets from six models, static full-block reference). *Middle/right*: isolated abstraction and storage losses for all twelve two-layer ReLU models; labels give dataset/seed. Intervals can be narrower than markers; horizontal scales differ. These are deterministic enclosures, not statistical confidence intervals (App. H).

Table 2: E2 by dataset and activation (9 networks each). *price*: returned $\gamma_{\mathrm{diag}}/\gamma_{\mathrm{vx}}$ with the layerwise vertex value; $\gamma_{\mathrm{vx}}/\gamma_{\mathrm{true}}$: vertex value over γ_{true} (ReLU only); *LipSDP-L/N*: median ratio of LipSDP-Layer to LipSDP-Neuron. Per-architecture bounds are in Table 7.

				$\gamma_{\mathrm{vx}}/\gamma_{\mathrm{true}}$		
data	act.	nets	price med./max	med.	max	LipSDP-L/N
Iris	relu	9	1.0000/1.0021	1.000	1.004	1.103
Iris	tanh	9	1.0000/1.0000	–	–	1.055
moons	relu	9	1.0000/1.0027	1.309	1.721	1.065
moons	tanh	9	1.0000/1.0494	–	–	1.083
MNIST	relu	9	1.0065/1.0235	1.000	1.042	1.070
MNIST	tanh	9	1.0047/1.0139	–	–	1.038
Fashion	relu	9	1.0005/1.0055	1.000	1.001	1.055
Fashion	tanh	9	1.0000/1.0006	–	–	1.024

All 360 certificates passed exact verification, and no lower bound exceeded a verified upper bound. Table 2 compares multiplier and residual losses. *The class price is small:* exact witnesses enclose its maximum over all 72 networks in $[1.049355, 1.049357]$; the returned-ratio median is 1.0000. Our LipSDP-Neuron raw solver outputs agree with the diagonal raw solver outputs, and LipSDP-Layer is looser (App. F). *The one-layer abstraction loss can be larger:* for one hidden ReLU layer on MNIST (W_0 of full row rank) the vertex value equals the true constant, as §4.1 predicts; on two-dimensional inputs it exceeds the exact constant by up to 43% with one hidden layer. *Certified loss decomposition.* Write $T \le B \le R \le D$ for the true gain, slope-box gain and layerwise-storage/diagonal class infima. The ratios D/R, B/T and R/B isolate multiplier, abstraction and storage losses. All twelve two-layer ReLU audits are complete: $B/T > 1.721$ in one moons model, while $R/B > 1.002$ in MNIST seeds 0/1 (Table 15). Thus richer multipliers alone cannot recover lost activation realizability or remove the layerwise-storage restriction. These examples establish distinct losses, not their relative sizes in all architectures.

E3: equilibrium models (Q3). Width-40 ReLU DEQs on raw MNIST, Fashion-MNIST and KMNIST (Clanuwat et al., 2018) pixels with a contractive ($\|W\|_2 \le 0.9$) or monotone ($I - (W + W^\top)/2 \succeq \omega I$, $\omega = 0.5$) parameterization, both with a unique fixed point for every input; three seeds, ten epochs. Static fixed-point certificates at the diagonal, singleton, pair and quadruple levels for all logits and the 45 class pairs; baselines $\|LV\|\|U\|/(1 - \|W\|)$ and the monotone-DEQ bound $\|LV\|\|U\|/\omega$ as stated by Havens et al. (2023). The one-block level has 2^{40} vertices and is not computed. We report the numerical ratios $\gamma_{\mathrm{diag}}/\gamma_{\mathrm{quads}}$ and $\gamma_{\mathrm{diag}}/\gamma_{\mathrm{low}}$, using returned feasible gains and a sampled lower estimate for the fixed-point abstraction. They suggest a bracket but are

Table 3: E3, means over 3 seeds. *baseline*: norm bound; *diag*: diagonal (contains LBEN); *quads*: blocks of four, verified accuracies. PGD: per-seed median finite ℓ_2 attack distance; failed attacks excluded, initially misclassified inputs assigned zero. More in Table 10.

				γ (all logits)			certified acc. at ϵ			
data	variant	S	acc.	baseline	diag	quads	0.1	0.3	0.5	PGD
MNIST	contractive	3	0.964	173.7	40.4	39.6	0.867	0.248	0.014	0.61
MNIST	monotone	3	0.965	41.2	26.2	25.8	0.910	0.566	0.123	0.74
Fashion	contractive	3	0.864	173.4	28.4	28.2	0.733	0.336	0.089	0.46
Fashion	monotone	3	0.860	32.6	19.7	19.4	0.744	0.419	0.136	0.50
KMNIST	contractive	3	0.845	126.9	26.4	25.8	0.697	0.297	0.073	0.57
KMNIST	monotone	3	0.843	27.4	17.0	16.7	0.739	0.451	0.185	0.67

not verified endpoints for the ratio of class infima. Certified accuracy is the fraction of all 10,000 test inputs whose ground-truth labels have verified ℓ_2 radii strictly above $\{0.1, 0.3, 0.5\}$. PGD uses 100 points per model. All datasets have complete gain-to-radius replays (App. H); Fashion/KMNIST use replacement training. Falsifier: a radius above a PGD distance.

All reported gains and radius chains replay; the numerical PGD checks find no violations (Table 3). Radius replay bounds fixed-point and logit-evaluation errors and rounds margins and radii downward, with gain bounds rounded upward. Quadruple blocks give paired within-model certified-accuracy gains over diagonal of 0.40–1.23 percentage points at $\epsilon = 0.3$ across all 18 models; 105/828 targets improve returned gains by at most 0.5% (Table 17). PGD-distance/radius diagnostics neither isolate multiplier headroom nor certify a population-wide gap.

Comparison with existing certificates. On the same width-40 models, SDP bounds improve on norm-based DEQ bounds by median factors up to 8.2; this changes the analysis, not just the multiplier class. Blocks of four reduce returned diagonal gains by up to 4.14% relative to that diagonal value. Three width-80 MNIST seeds per variant gain 0.11–0.22 (contractive) or 0.76–1.16 (monotone) certified-accuracy points at $\epsilon = 0.3$ retaining the better diagonal baseline (Table 18). At 60 seconds per target, the original block policy loses to diagonal at $\epsilon = 0.3$; an exploratory diagonal-first follow-up gains 0.21/0.84 points on the contractive/monotone seed-zero models (Tables 19–20). At the same radius, the policies tie at 30 seconds; at 15 seconds, monotone diagonal-first is worse. Figure 1 shows certified class-price enclosures: below 1.05 for the E2 maximum and $[1.067848, 1.122833]$ for the maximum over all 828 static E3 targets (Table 16).

How the class prices are certified. For diagonal and reference gain infima D, R, verified primal certificates give upper bounds U_D, U_R; exact moment and slope-response witnesses give lower bounds L_D, L_R. Thus $D/R \in [L_D/U_R, U_D/L_R]$ for positive lower bounds. Taking maxima of the endpoints bounds the maximum over audited targets. E3 uses the static full-block reference. Feasible-value ratios alone do not establish these enclosures.

9 Conclusion and Limitations

What this means in practice. On the audited feedforward models, diagonal gain penalties are below 5%; the E3 maximum exceeds 6.78%. Complete two-layer audits show abstraction loss above 72% in one moons model and storage loss above 0.2% in two MNIST models. Richer multipliers alone cannot remove either restriction; this is a finite-model finding, not a universal ordering. For stability, our necessary and sufficient test characterizes the stated Lur'e–Postnikov family over sector/monotone classes through a constructive quadratic reduction preserving the decrease constant (Theorem 8); slope restrictions can permit strict improvement (Proposition 9).

Limitations and future work. Gain references use the independent-slope abstraction and stated certificate classes. Broader Lyapunov families, dynamic multipliers and realizable-slope information may yield stronger certificates; extending these to certified robustness requires an incremental gain-to-margin chain (further open problems in App. F).

AI USE STATEMENT

Generative AI (Claude Code with Anthropic Claude Opus 5, and OpenAI Codex) was used, under the authors' guidance and verification, to help develop the conceptual framework, proofs, code and experiments, interpret results, and polish the writing; Codex also assisted with technical review and revisions. The authors review and verify the AI-assisted proofs and experiments and take full responsibility for the content of this paper.

ETHICS STATEMENT

This work develops and measures defensive tools: global certificates that bound how much the output of a neural network can change under input perturbations, and stability certificates. It involves no human subjects and no personal data; all datasets are public benchmarks (Iris, MNIST, Fashion-MNIST, KMNIST) or synthetic. The attacks used (PGD on our own small models) serve only to falsify the certificates. The main risk is overclaiming robustness; we mitigate it by scoping exactness claims, checking SDP feasibility in exact arithmetic, and distinguishing fully replayed gain-to-radius chains from historical numerical radius estimates (App. H), and counting failures instead of dropping them.

REPRODUCIBILITY STATEMENT

The theoretical results are proved in the appendix (Appendix A is a guide), and the certificates of Proposition 9 and of the two-neuron example are printed as rationals (App. C.3–C.4), so they can be checked without our code. Algorithm 1 and App. E give the verification, training and evaluation settings, and App. D (Table 5) describes the falsification tests. Experimental numbers are linked in a ledger to scripts, outputs and code versions; settings include fixed seeds and open-source solvers (Clarabel and SCS via CVXPY). The code (including the exact verifier), available saved models, training settings, ledger and Lean certificate checks will be released publicly upon acceptance. Missing historical checkpoints and the scope of independent replay are disclosed in App. H; seeds alone do not establish model identity.

REFERENCES

N. Syazreen Ahmad, J. Carrasco, and W. P. Heath. A less conservative LMI condition for stability of discrete-time systems with slope-restricted nonlinearities. *IEEE Transactions on Automatic Control*, 60(6):1692–1697, 2015. doi: 10.1109/TAC.2014.2361398.

Bernardo Aquino, Arash Rahnama, Peter Seiler, Lizhen Lin, and Vijay Gupta. Robustness against adversarial attacks in neural networks using incremental dissipativity. *IEEE Control Systems Letters*, 6:2341–2346, 2022. doi: 10.1109/LCSYS.2022.3150719.

Alexandre Araujo, Aaron Havens, Blaise Delattre, Alexandre Allauzen, and Bin Hu. A unified algebraic perspective on Lipschitz neural networks. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2303.03169.

Stephen Boyd, Laurent El Ghaoui, Eric Feron, and Venkataramanan Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics, 1994. doi: 10.1137/1.9781611970777.

Joaquin Carrasco, Matthew C. Turner, and William P. Heath. Zames–Falb multipliers for absolute stability: From O'Shea's contribution to convex searches. *European Journal of Control*, 28:1–19, 2016. doi: 10.1016/j.ejcon.2015.10.003.

Ya-Chien Chang, Nima Roohi, and Sicun Gao. Neural Lyapunov control. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://papers.nips.cc/paper_files/paper/2019/hash/2647c1dba23bc0e0f9cdf75339e120d2-Abstract.html.

Tong Chen, Jean-Bernard Lasserre, Victor Magron, and Edouard Pauwels. Semialgebraic representation of monotone deep equilibrium models and applications to certification. In *Advances in Neural Information Processing Systems*, volume 34, pp. 27146–27159,

2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/e3b21256183cf7c2c7a66be163579d37-Abstract.html.

Tarin Clanuwat, Mikel Bober-Irizar, Asanobu Kitamoto, Alex Lamb, Kazuaki Yamamoto, and David Ha. Deep learning for classical Japanese literature. *arXiv preprint arXiv:1812.01718*, 2018. URL https://arxiv.org/abs/1812.01718.

Jeremy Cohen, Elan Rosenfeld, and J. Zico Kolter. Certified adversarial robustness via randomized smoothing. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1310–1320, 2019. URL https://proceedings.mlr.press/v97/cohen19c.html.

Charles Dawson, Sicun Gao, and Chuchu Fan. Safe control with learned certificates: A survey of neural Lyapunov, barrier, and contraction methods for robotics and control. *IEEE Transactions on Robotics*, 39(3):1749–1767, 2023. doi: 10.1109/TRO.2022.3232542.

Ross Drummond and Giorgio Valmorbida. Generalized Lyapunov functions for discrete-time Lurie systems with slope-restricted nonlinearities. *IEEE Transactions on Automatic Control*, 68(10): 5966–5976, 2023. doi: 10.1109/TAC.2022.3233540.

Mahyar Fazlyab, Alexander Robey, Hamed Hassani, Manfred Morari, and George J. Pappas. Efficient and accurate estimation of Lipschitz constants for deep neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/95e1533eb1b20a97777749fb94fdb944-Abstract.html.

R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2): 179–188, 1936. doi: 10.1111/j.1469-1809.1936.tb02137.x.

Timon Gehr, Matthew Mirman, Dana Drachsler-Cohen, Petar Tsankov, Swarat Chaudhuri, and Martin Vechev. AI2: Safety and robustness certification of neural networks with abstract interpretation. In *2018 IEEE Symposium on Security and Privacy (SP)*, pp. 3–18, 2018. doi: 10.1109/SP.2018.00058.

Navid Hashemi, Justin Ruths, and Mahyar Fazlyab. Certifying incremental quadratic constraints for neural networks via convex optimization. In *Proceedings of the 3rd Conference on Learning for Dynamics and Control*, volume 144 of *Proceedings of Machine Learning Research*, pp. 842–853, 2021. URL https://proceedings.mlr.press/v144/hashemi21a.html.

Aaron J. Havens, Alexandre Araujo, Siddharth Garg, Farshad Khorrami, and Bin Hu. Exploiting connections between Lipschitz structures for certifiably robust deep equilibrium models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 21658–21674, 2023. doi: 10.52202/075280-0947.

Saber Jafarpour, Alexander Davydov, Anton V. Proskurnikov, and Francesco Bullo. Robust implicit networks via non-Euclidean contractions. In *Advances in Neural Information Processing Systems*, volume 34, pp. 9857–9868, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/51a6ce0252d8fa6e913524bdce8db490-Abstract.html.

Saber Jafarpour, Matthew Abate, Alexander Davydov, Francesco Bullo, and Samuel Coogan. Robustness certificates for implicit neural networks: A mixed monotone contractive approach. In *Learning for Dynamics and Control Conference*, volume 168 of *Proceedings of Machine Learning Research*, pp. 917–930, 2022. URL https://proceedings.mlr.press/v168/jafarpour22a.html.

Matt Jordan and Alexandros G. Dimakis. Exactly computing the local Lipschitz constant of ReLU networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7344–7353, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/5227fa9a19dce7ba113f50a405dcaf09-Abstract.html.

E. Jury and B. Lee. On the stability of a certain class of nonlinear sampled-data systems. *IEEE Transactions on Automatic Control*, 9(1):51–61, 1964. doi: 10.1109/TAC.1964.1105622.

Guy Katz, Clark Barrett, David L. Dill, Kyle Julian, and Mykel J. Kochenderfer. Reluplex: An efficient SMT solver for verifying deep neural networks. In *Computer Aided Verification (CAV)*, volume 10426 of *Lecture Notes in Computer Science*, pp. 97–117, 2017. doi: 10.1007/978-3-319-63387-9_5.

Simon Kuang, Yuezhu Xu, S. Sivaranjani, and Xinfan Lin. Demystifying Lipschitz verification: positive matrices, negative results. *arXiv preprint arXiv:2603.28113*, 2026. URL https://arxiv.org/abs/2603.28113. v2, 7 May 2026.

V. V. Kulkarni and M. G. Safonov. All multipliers for repeated monotone nonlinearities. *IEEE Transactions on Automatic Control*, 47(7):1209–1212, 2002. doi: 10.1109/TAC.2002.800642.

Fabian Latorre, Paul Rolland, and Volkan Cevher. Lipschitz constant estimation of neural networks via sparse polynomial optimization. In *International Conference on Learning Representations*, 2020. URL https://arxiv.org/abs/2004.08688.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

A. Megretski and A. Rantzer. System analysis via integral quadratic constraints. *IEEE Transactions on Automatic Control*, 42(6):819–830, 1997. doi: 10.1109/9.587335.

Sahel Vahedi Noori, Bin Hu, Geir Dullerud, and Peter Seiler. Stability and performance analysis of discrete-time ReLU recurrent neural networks. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, pp. 8626–8632, 2024. doi: 10.1109/CDC56724.2024.10886894.

Patricia Pauli, Anne Koch, Julian Berberich, Paul Kohler, and Frank Allgöwer. Training robust neural networks using Lipschitz bounds. *IEEE Control Systems Letters*, 6:121–126, 2022. doi: 10.1109/LCSYS.2021.3050444.

Patricia Pauli, Dennis Gramlich, and Frank Allgöwer. Lipschitz constant estimation for 1D convolutional neural networks. In *Learning for Dynamics and Control Conference*, volume 211 of *Proceedings of Machine Learning Research*, pp. 1321–1332, 2023. URL https://proceedings.mlr.press/v211/pauli23a.html.

Patricia Pauli, Dennis Gramlich, and Frank Allgöwer. Lipschitz constant estimation for general neural network architectures using control tools. *arXiv preprint arXiv:2405.01125*, 2024a. doi: 10.48550/arXiv.2405.01125.

Patricia Pauli, Aaron Havens, Alexandre Araujo, Siddharth Garg, Farshad Khorrami, Frank Allgöwer, and Bin Hu. Novel quadratic constraints for extending LipSDP beyond slope-restricted activations. In *International Conference on Learning Representations*, pp. 57238–57263, 2024b. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/fd8bec95da7283894f27576b5f2fb3fc-Abstract-Conference.html.

Aditi Raghunathan, Jacob Steinhardt, and Percy Liang. Semidefinite relaxations for certifying robustness to adversarial examples. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/29c0605a3bab4229e46723f89cf59d83-Abstract.html.

Max Revay, Ruigang Wang, and Ian R. Manchester. Lipschitz-bounded equilibrium networks. *arXiv preprint arXiv:2010.01732*, 2020. doi: 10.48550/arXiv.2010.01732.

Max Revay, Ruigang Wang, and Ian R. Manchester. Recurrent equilibrium networks: Flexible dynamic models with guaranteed stability and robustness. *IEEE Transactions on Automatic Control*, 69(5):2855–2870, 2024. doi: 10.1109/TAC.2023.3294101.

S. M. Rump. Verification of positive definiteness. *BIT Numerical Mathematics*, 46(2):433–452, 2006. doi: 10.1007/s10543-006-0056-1.

Hadi Salman, Greg Yang, Huan Zhang, Cho-Jui Hsieh, and Pengchuan Zhang. A convex relaxation barrier to tight robustness verification of neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/246a3c5544feb054f3ea718f61adfa16-Abstract.html.

Kevin Scaman and Aladin Virmaux. Lipschitz regularity of deep neural networks: analysis and efficient estimation. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/d54e99a6c03704e95e6965532dec148b-Abstract.html.

C. W. Scherer. LPV control and full block multipliers. *Automatica*, 37(3):361–375, 2001. doi: 10.1016/S0005-1098(00)00176-X.

Ruigang Wang and Ian R. Manchester. Direct parameterization of Lipschitz-bounded deep networks. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 36093–36110, 2023. URL https://proceedings.mlr.press/v202/wang23v.html.

Zi Wang, Bin Hu, Aaron J. Havens, Alexandre Araujo, Yang Zheng, Yudong Chen, and Somesh Jha. On the scalability and memory efficiency of semidefinite programs for Lipschitz constant estimation of neural networks: Scaling the computation for ImageNet. In *International Conference on Learning Representations*, pp. 19529–19555, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/54fa8255cdf30736ecad38e842725e7f-Abstract-Conference.html.

J. Willems and R. Brockett. Some new rearrangement inequalities having application in stability analysis. *IEEE Transactions on Automatic Control*, 13(5):539–549, 1968. doi: 10.1109/TAC.1968.1098999.

Ezra Winston and J. Zico Kolter. Monotone operator equilibrium networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 10718–10728, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/798d1c2813cbdf8bcdb388db0e32d496-Abstract.html.

Eric Wong and J. Zico Kolter. Provable defenses against adversarial examples via the convex outer adversarial polytope. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 5286–5295, 2018. URL https://proceedings.mlr.press/v80/wong18a.html.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. URL https://arxiv.org/abs/1708.07747.

Anton Xue, Lars Lindemann, Alexander Robey, Hamed Hassani, George J. Pappas, and Rajeev Alur. Chordal sparsity for Lipschitz constant estimation of deep neural networks. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 3389–3396, 2022. doi: 10.1109/CDC51059.2022.9993136.

Lujie Yang, Hongkai Dai, Zhouxing Shi, Cho-Jui Hsieh, Russ Tedrake, and Huan Zhang. Lyapunov-stable neural control for state and output feedback: A novel formulation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 56033–56046, 2024. URL https://proceedings.mlr.press/v235/yang24f.html.

Huan Zhang, Tsui-Wei Weng, Pin-Yu Chen, Cho-Jui Hsieh, and Luca Daniel. Efficient neural network robustness certification with general activation functions. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/d04863f100d59b3eb688a11f95b0ae60-Abstract.html.

Richard Y. Zhang. On the tightness of semidefinite relaxations for certifying robustness to adversarial examples. In *Advances in Neural Information Processing Systems*, volume 33, pp. 3808–3820, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/27b587bbe83aecf9a98c8fe6ab48cacc-Abstract.html.

Ruikun Zhou, Thanin Quartz, Hans De Sterck, and Jun Liu. Neural Lyapunov control of unknown nonlinear systems with stability guarantees. In *Advances in Neural Information Processing Systems*, volume 35, pp. 29113–29125, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/bba3160c42f4c1d23175b5abf04347f5-Abstract-Conference.html.

A GUIDE TO THE PROOFS

Appendix B proves the incremental, layered, hierarchy and equilibrium results of §§4–6; Appendix C proves the stability results of §7. Every result used in the paper is restated where it is proved. Table 4 lists where each result is stated and proved and what it relies on; one elementary tool, the vertex principle below, carries most of the arguments.

Table 4: Where each result is stated and proved, and what its proof uses.

result	stated in	proved in	proof uses
Lemma 10 (vertex principle)	App. A	App. A	convexity only
Lemma 1 (incremental secant)	§3	App. B.1	–
Theorem 2	§4	App. B.1	Lemmas 10, 1
Lemma 11 (strict feasibility)	App. B.1	App. B.1	Theorem 2
Theorem 12, Proposition 3	§4.1	App. B.2	Lemmas 10, 1
Proposition 13 (layerwise hierarchy)	App. B.2	App. B.2	Theorem 5, Lemma 1
Proposition 4	§4.2	App. B.3	–
Theorem 5 (i)–(iii)	§5	App. B.4	Lemmas 10, 1 Lemma 11; part (i);
Theorem 5 (iv)	§5	App. B.4	Scherer (2001, App. B, Thm. 8(a))
Lemma 14, Theorem 15	App. B.5	App. B.5	Lemmas 10, 1; Theorem 2
Proposition 6, Remark 7	§6	App. B.5	Theorem 5
Theorem 8, vertex part	§7	App. C.1	Lemmas 10, 16
Theorem 8, collapse part	§7	App. C.2	Lemmas 17, 18; vertex part
Proposition 9	§7	App. C.3	Lemmas 19, 20
System A gap	§7	App. C.4	Lemma 21; vertex part

Notation used throughout. $\mathfrak{B} = \{\mathrm{diag}(k) : k_j \in [\alpha_j, \beta_j]\}$ is the slope box and $\mathfrak{V}$ its 2^m vertices; $A_K = A + BKC^\top$ and $C_{y,K} = C_y + D_y KC^\top$; $|v|_P^2 = v^\top Pv$; $s_\pm = \max(\pm s, 0)$ for scalars; f denotes a value of the nonlinearity (to distinguish it from the map f). "Strict certificate" always means one $\varepsilon > 0$ for the whole class.

Lemma 10 (Vertex principle)**.** *The box $\mathfrak{B}$ is the convex hull of $\mathfrak{V}$. Consequently a function that is concave on $\mathfrak{B}$ attains its minimum over $\mathfrak{B}$ at a vertex, and a function that is convex on $\mathfrak{B}$ attains its maximum over $\mathfrak{B}$ at a vertex.*

Proof. The box is a product of intervals, each the convex hull of its two endpoints, and the convex hull of a product of sets is the product of their convex hulls. If q is concave and $K = \sum_i t_i K_i$ is a convex combination of vertices, then $q(K) \geq \sum_i t_i q(K_i) \geq \min_i q(K_i)$; the convex case is the same with the inequalities reversed. □

In every application below, $q(K)$ is a constant minus a sum of squared norms of vectors that are affine in K (hence concave), or a quadratic form $v(K)^\top M v(K)$ with $M \succeq 0$ and v affine in K (hence convex).

B PROOFS OF THE INCREMENTAL, HIERARCHY AND EQUILIBRIUM RESULTS

B.1 INCREMENTAL DISSIPATIVITY

Proof of Lemma 1 (incremental secant). Coordinatewise: if $\sigma_j \neq \tilde{\sigma}_j$, take k_j to be the difference quotient, which lies in $[\alpha_j, \beta_j]$ by slope restriction; if $\sigma_j = \tilde{\sigma}_j$, both sides vanish and any $k_j \in [\alpha_j, \beta_j]$ works. The converse is immediate: the difference quotient of $\sigma \mapsto k_j\sigma$ is k_j. □

Theorem (Theorem 2, restated)**.** *Let $P \succeq 0$ and $g \geq 0$. The following are equivalent: (a)* (2) *holds for all $f \in \mathcal{S}^{\mathrm{tv}}$, all times and all pairs of states and inputs; (b)* (2) *holds for all linear $f = K\sigma$ with $K \in \mathfrak{B}$; (c) $\mathcal{N}_K(P, g) \preceq 0$ for all $K \in \mathfrak{V}$. Under (c) every pair of trajectories satisfies $\sum_{k<T} |\Delta y_k|^2 \leq g \sum_{k<T} |\Delta w_k|^2 + \Delta x_0^\top P \Delta x_0$.*

Proof. (a)⇒(b). For $K \in \mathfrak{B}$ the linear map $\sigma \mapsto K\sigma$ lies in $\mathcal{S} \subseteq \mathcal{S}^{\mathrm{tv}}$ (Lemma 1).

(b)⇒(c). For $f = K\sigma$, $\Delta x^+ = A_K \Delta x + E\Delta w$ and $\Delta y = C_{y,K}\Delta x + D_w \Delta w$, so (2) reads $[\Delta x; \Delta w]^\top \mathcal{N}_K(P, g)[\Delta x; \Delta w] \leq 0$ for all $(\Delta x, \Delta w)$; take $K \in \mathfrak{V}$.

(c)⇒(a). Fix a time, a pair of states and inputs, and $f \in \mathcal{S}^{\mathrm{tv}}$. By Lemma 1, $\Delta f = K_\star C^\top \Delta x$ with $K_\star \in \mathfrak{B}$. For fixed $(\Delta x, \Delta w)$ define on $\mathfrak{B}$ the slack

$$s(K) = g|\Delta w|^2 + |\Delta x|_P^2 - \left|A_K \Delta x + E\Delta w\right|_P^2 - |C_{y,K}\Delta x + D_w \Delta w|^2.$$

Both vectors inside the norms are affine in K and $P \succeq 0$, so s is concave; by Lemma 10 its minimum over $\mathfrak{B}$ is attained at a vertex, where (c) gives $s \geq 0$. Hence $s(K_\star) \geq 0$, which is (2).

Horizon bound. Summing (2) over $k < T$ telescopes the storage terms, and $P \succeq 0$ lets us drop $\Delta x_T^\top P \Delta x_T \geq 0$. The characterization of $g_{\mathrm{vx}}^\star$ is the equivalence (a) $\iff$ (c). □

Lemma 11 (Strict feasibility)**.** *If some $P_s \succ 0$ satisfies $P_s - A_K^\top P_s A_K \succ 0$ for all $K \in \mathfrak{V}$, then for every $g > g_{\mathrm{vx}}^\star$ there is $P \succ 0$ with $\mathcal{N}_K(P, g) \prec 0$ for all $K \in \mathfrak{V}$.*

Proof. Step 1 (a strictly feasible point). For $c > 0$ large, the $(1,1)$ block of $\mathcal{N}_K(cP_s, g_1)$ is $-c(P_s - A_K^\top P_s A_K) + C_{y,K}^\top C_{y,K} \prec 0$; for such c and g_1 large, a Schur complement gives $\mathcal{N}_K(cP_s, g_1) \prec 0$, uniformly over the finitely many vertices. In particular the feasible set of Theorem 2(c) is nonempty.

Step 2 (a feasible point below g). Since $g > g_{\mathrm{vx}}^\star$, there is (P_0, g_0) feasible for Theorem 2(c) with $g_0 < g$.

Step 3 (interpolate). The LMIs are affine in (P, g), so $(1-t)(P_0, g_0) + t(cP_s, g_1)$ is strictly feasible for every $t \in (0, 1]$; for small t its g-value is below g, and $\mathcal{N}_K(P, \cdot)$ is nonincreasing in g. □

B.2 LAYERED NETWORKS AND HARDNESS

Theorem 12 (Layerwise vertex certificate; class-exact within layerwise quadratic storage)**.** *Under* (3)*, $(F(x) - F(\tilde{x}))^\top O (F(x) - F(\tilde{x})) \leq g|x - \tilde{x}|^2$. The vertex conditions are necessary for the per-layer inequalities $\Delta x_{\ell+1}^\top P_{\ell+1} \Delta x_{\ell+1} \leq \Delta x_\ell^\top P_\ell \Delta x_\ell$ to hold for all admissible activations. For $H = 1$ the optimal value is $\max_{K \in \mathfrak{V}_0} \lambda_{\max}(W_0^\top K W_1^\top O W_1 K W_0)$.*

Proof. Step 1 (one layer). By Lemma 1, $\Delta x_{\ell+1} = K_\star W_\ell \Delta x_\ell$ with $K_\star$ in the box. The map $K \mapsto \Delta x_\ell^\top W_\ell^\top K P_{\ell+1} K W_\ell \Delta x_\ell$ is convex because $P_{\ell+1} \succeq 0$, so by Lemma 10 its maximum over the box is attained at a vertex, where (3) bounds it by $\Delta x_\ell^\top P_\ell \Delta x_\ell$.

Step 2 (chain the layers). $\Delta F^\top O \Delta F \leq \Delta x_H^\top P_H \Delta x_H \leq \cdots \leq \Delta x_0^\top P_0 \Delta x_0 = g|\Delta x|^2$.

Step 3 (necessity). Linear maps with $K \in \mathfrak{V}_\ell$ are admissible activations.

Step 4 ($H = 1$). Enlarging P_1 only tightens the first set of constraints, so $P_1 = W_1^\top O W_1$ is optimal, and the remaining constraints give the stated maximum. □

The block levels used for feedforward networks in E2 are the hierarchy of Theorem 5 applied layer by layer.

Proposition 13 (Layerwise block-multiplier certificates)**.** *For each hidden layer $\ell < H$ let $\mathcal{P}_\ell$ be a partition of its neurons and $\Pi_\ell \in \Pi(\mathcal{P}_\ell)$ a block multiplier satisfying* (4) *for the slope box of layer ℓ (or a diagonal multiplier). If $g \geq 0$ and $P_1, \dots, P_H \succeq 0$ satisfy, with $P_0 = gI$, for every $\ell < H$ and all $(\Delta x_\ell, \Delta f_\ell)$,*

$$\Delta f_\ell^\top P_{\ell+1} \Delta f_\ell - \Delta x_\ell^\top P_\ell \Delta x_\ell + (\Delta f_\ell, W_\ell \Delta x_\ell)^\top \Pi_\ell \, (\Delta f_\ell, W_\ell \Delta x_\ell) \leq 0, \tag{5}$$

and $W_H^\top O W_H \preceq P_H$, then (i) $(F(x) - F(\tilde{x}))^\top O (F(x) - F(\tilde{x})) \leq g|x - \tilde{x}|^2$; (ii) the optimal g is monotone under refinement of the partitions and the diagonal family lies in the singleton level; and (iii) any feasible $(g, P_1, \dots, P_H)$ is feasible for the vertex conditions (3)*, so the vertex value is at most every block level.*

Proof. (i). Fix ℓ and a pair of inputs. By Lemma 1, $\Delta x_{\ell+1} = \Delta f_\ell = K_\star W_\ell \Delta x_\ell$ with $K_\star$ in the box. As in Theorem 5(i), the multiplier term in (5) is ≥ 0 at $(K_\star \sigma, \sigma)$, $\sigma = W_\ell \Delta x_\ell$ (concavity in K and Lemma 10), so $\Delta x_{\ell+1}^\top P_{\ell+1} \Delta x_{\ell+1} \leq \Delta x_\ell^\top P_\ell \Delta x_\ell$. Chaining over ℓ and using $W_H^\top O W_H \preceq P_H$ gives the bound, as in Step 2 of the proof of Theorem 12. *(ii).* Theorem 5(ii)–(iii) concern only the multiplier sets, which are used here layer by layer. *(iii).* At a vertex K of layer ℓ the multiplier term is ≥ 0 at $(K\sigma, \sigma)$ by (4), so evaluating (5) at $\Delta f_\ell = K W_\ell \Delta x_\ell$ gives $(K W_\ell)^\top P_{\ell+1} (K W_\ell) \preceq P_\ell$. □

Proof of Proposition 3 (NP-hardness of the class-exact *value). Step 1 (the gadget).* Let $G = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = p$ and signed incidence matrix $N \in \mathbb{R}^{|\mathcal{E}| \times p}$, so that $N^\top N = \mathcal{L}$ is the Laplacian and $\max_{s \in \{\pm 1\}^p} s^\top \mathcal{L} s = 4 \operatorname{MaxCut}(G)$. Take $W_0 = [N^\top; -N^\top]$ ($2p$ neurons, input dimension $|\mathcal{E}|$), $W_1 = \mathbf{1}^\top$, $O = 1$ and slopes $[0, 1]$.

Step 2 (evaluate at vertices). For $K = \operatorname{diag}(k^+, k^-)$ with $k^\pm \in \{0,1\}^p$, $W_1 K W_0 = (k^+ - k^-)^\top N^\top$, so by Theorem 12 $g^\star = \max_{s \in \{-1,0,1\}^p} s^\top \mathcal{L} s$.

Step 3 (remove the zeros). Since $\{\pm 1\}^p \subseteq \{-1, 0, 1\}^p \subseteq [-1, 1]^p$ and the convex function $s \mapsto s^\top \mathcal{L} s$ attains its maximum over the cube at a vertex (Lemma 10), $g^\star = 4 \operatorname{MaxCut}(G)$. The construction is polynomial in the size of G. □

B.3 Certified radii

Proof of Proposition 4 (margin certificate). For every $j \neq y$, $F_y(x + \delta) - F_j(x + \delta) \geq F_y(x) - F_j(x) - \gamma_{yj} \|\delta\|_2 > 0$ when $\|\delta\|_2 < r(x)$. With an evaluation error $|\hat{z} - z^\star| \leq e_z$, the margin changes by at most $|(e_y - e_j)^\top V (\hat{z} - z^\star)| \leq |(e_y - e_j)^\top V| \, e_z$. □

B.4 The block-multiplier hierarchy

Proof of Theorem 5 (i)–(iii). (i) Soundness. Let $\Delta f = K_\star \Delta \sigma$ with $K_\star \in \mathfrak{B}$ (Lemma 1). For each block b, the map $K_b \mapsto \Delta \sigma_b^\top [K_b; I]^\top \Pi_b [K_b; I] \Delta \sigma_b$ is concave because $(\Pi_b)_{11} \preceq 0$, so by Lemma 10 it is at least its minimum over the block vertices, which is ≥ 0 by (4). Summing over blocks, $(\Psi\xi)^\top \Pi \Psi \xi \geq 0$, and feasibility gives $\xi^\top \mathcal{F} \xi \leq -(\Psi\xi)^\top \Pi \Psi \xi \leq 0$, which is (2).

(ii) Monotonicity. If $b = b'_1 \cup \cdots \cup b'_q$ with $b'_i \in \mathcal{P}'$, embed $\Pi_b = \sum_i S_{b'_i \to b}^\top \Pi_{b'_i} S_{b'_i \to b}$. After a permutation $(\Pi_b)_{11} = \oplus_i (\Pi_{b'_i})_{11} \preceq 0$, and at a vertex $K_b = \oplus_i K_{b'_i}$, $[K_b; I]^\top \Pi_b [K_b; I] = \oplus_i [K_{b'_i}; I]^\top \Pi_{b'_i} [K_{b'_i}; I] \succeq 0$.

(iii) Diagonal level. With $c = (\alpha + \beta)/2$, $[k; 1]^\top \lambda \left[\begin{smallmatrix} -1 & c \\ c & -\alpha\beta \end{smallmatrix}\right] [k; 1] = \lambda (k - \alpha)(\beta - k)$ vanishes at $k \in \{\alpha, \beta\}$, so the diagonal multipliers are admissible singleton blocks. Conversely, a symmetric 2×2 Π_j is admissible iff $p(k) = \pi_{11} k^2 + 2\pi_{12} k + \pi_{22}$ has $\pi_{11} \leq 0$, $p(\alpha) \geq 0$ and $p(\beta) \geq 0$. With $\lambda = -\pi_{11}$, the polynomial $p(k) - \lambda(k - \alpha)(\beta - k)$ is affine and nonnegative at both endpoints, hence equals $a_1(k - \alpha) + a_2(\beta - k)$ with $a_1, a_2 \geq 0$ when $\alpha < \beta$; these are the two linear constraints. □

Proof of Theorem 5 (iv). Step 1 (strict vertex feasibility). Fix any $g > g^\star_{\mathrm{vx}}$. Lemma 11 supplies $P \succ 0$ with $\mathcal{N}_K(P, g) \prec 0$ at every vertex. For fixed $(\Delta x, \Delta w)$ the dissipation form after substituting $\Delta f = K C^\top \Delta x$ is convex in K, since $P \succeq 0$. Hence strict negativity extends to the entire slope box. For every $K \in \mathfrak{B}$, $\mathcal{F}(P, g)$ is therefore strictly negative on the nonzero vectors satisfying $\Delta f = K C^\top \Delta x$.

Step 2 (full-block S-procedure without a rank assumption). Apply Scherer (2001, App. B, Theorem 8(a)) with ambient subspace the full ξ-space, $U = I$, $Q = \mathcal{F}(P, g)$, $V = \Psi$, and $\mathcal{S}(K) = \operatorname{im} G_K$, where $G_K = [K; I]$. The finite slope box is compact and path-connected, including when some intervals are singletons. The graph subspaces vary continuously: their orthogonal projectors are $G_K(G_K^\top G_K)^{-1}G_K^\top$, and $G_K^\top G_K = I + K^\top K \succ 0$. Moreover, $\Psi\xi \in \mathcal{S}(K)$ is exactly $\Delta f = KC^\top \Delta x$. These facts require no rank condition on C or Ψ. The theorem yields a symmetric Π such that

$$G_K^\top \Pi G_K \succ 0 \quad (K \in \mathfrak{B}), \qquad \mathcal{F}(P, g) + \Psi^\top \Pi \Psi \prec 0.$$

Step 3 (the required leading-block sign). Restrict the second inequality to $\xi = (0, \Delta f, 0)$. Then $\Psi\xi = (\Delta f, 0)$, so

$$B^\top P B + D_y^\top D_y + \Pi_{11} \prec 0.$$

As $P \succeq 0$, this implies $\Pi_{11} \prec 0$. Together with the graph inequalities at the vertices, this makes Π admissible for the stated one-block cone (4).

Step 4 (equality of infima). Soundness gives $g^\star(\{\{1, \dots, m\}\}) \geq g_{\mathrm{vx}}^\star$. Steps 1–3 make every $g > g_{\mathrm{vx}}^\star$ feasible for the one-block family, proving the reverse inequality of infima. Neither optimum attainment nor feasibility at $g = g_{\mathrm{vx}}^\star$ is assumed. Strict quadratic stability is retained through Lemma 11; the argument concerns dynamic quadratic storage, not static fixed-point exactness. □

B.5 Deep equilibrium models

Lemma 14 (DEQ coordinates)**.** *For each x, $z^\star \mapsto \sigma^\star = Wz^\star + Ux + b$ is a bijection from fixed points of $z \mapsto \phi(Wz + Ux + b)$ onto fixed points of $T_x(\sigma) = W\phi(\sigma) + Ux + b$, with inverse $\sigma^\star \mapsto \phi(\sigma^\star)$.*

Proof. If $z^\star = \phi(Wz^\star + Ux + b)$ and $\sigma^\star = Wz^\star + Ux + b$, then $\phi(\sigma^\star) = z^\star$ and $T_x(\sigma^\star) = \sigma^\star$. Conversely, if $T_x(\sigma^\star) = \sigma^\star$ and $z^\star = \phi(\sigma^\star)$, then $Wz^\star + Ux + b = \sigma^\star$, so $\phi(Wz^\star + Ux + b) = z^\star$. The two maps are mutually inverse on these sets. □

Theorem 15 (DEQ well-posedness and storage-based Lipschitz bound)**.** *(i) (sound) If $P \succ 0$ and $0 \leq \rho < 1$ satisfy $(WK)^\top P (WK) \preceq \rho^2 P$ for all $K \in \mathfrak{V}$, then every T_x is a ρ-contraction in $|\cdot|_P$, the DEQ has a unique fixed point for every x, and the iteration converges from any initialization with $|\sigma_k - \sigma^\star|_P \leq \frac{\rho}{1-\rho}|\sigma_k - \sigma_{k-1}|_P$. (ii) (sound for the DEQ Lipschitz constant) If $(P \succeq 0, g)$ satisfies Theorem 2(c) for the σ-coordinate system with $D_y = LV$, then $|L(F(x) - F(\tilde{x}))| \leq \sqrt{g}|x - \tilde{x}|$ for all $x, \tilde{x}$ and all their fixed points.*

Proof. (i). By Lemma 1, $T_x(\sigma) - T_x(\tilde{\sigma}) = WK_\star(\sigma - \tilde{\sigma})$ with $K_\star \in \mathfrak{B}$. The map $K \mapsto |WK\Delta\sigma|_P^2$ is convex, so by Lemma 10 its maximum over $\mathfrak{B}$ is attained at a vertex, where it is at most $\rho^2|\Delta\sigma|_P^2$. Hence T_x is a ρ-contraction; Banach's fixed-point theorem, Lemma 14 and the standard a posteriori bound give the claims.

(ii). Apply Theorem 2(c)⇒(a), a pointwise one-step statement, with states $\sigma^\star, \tilde{\sigma}^\star$ and inputs $x, \tilde{x}$. The next states are again $\sigma^\star, \tilde{\sigma}^\star$, so the storage terms cancel and $|\Delta y|^2 \leq g|\Delta x|^2$ with $\Delta y = LV(\phi(\sigma^\star) - \phi(\tilde{\sigma}^\star)) = L(F(x) - F(\tilde{x}))$. □

Proof of Proposition 6 (static fixed-point certificates). Throughout, $\sigma^\star = Wz^\star + Ux + b$, $\tilde{\sigma}^\star$ likewise, and $\xi = (\Delta z, \Delta x)$.

(i) Soundness. By Lemma 1, $\Delta z = K_\star \Delta\sigma$ with $K_\star \in \mathfrak{B}$ and $\Delta\sigma = W\Delta z + U\Delta x$, i.e. $\Psi_s\xi = (K_\star\Delta\sigma, \Delta\sigma)$. As in Theorem 5(i), $(\Psi_s\xi)^\top \Pi \Psi_s \xi \geq 0$, so the certificate gives $|LV\Delta z|^2 - g|\Delta x|^2 \leq 0$, and $LV\Delta z = L(F(x) - F(\tilde{x}))$.

(ii) Uniqueness. Two fixed points $z^\star \neq \tilde{z}^\star$ for the same x give $\xi = (\Delta z, 0) \neq 0$ with $\xi^\top(\mathcal{G} + \Psi_s^\top \Pi \Psi_s)\xi \geq |LV\Delta z|^2 \geq 0$, contradicting strict negativity.

(iii) Hierarchy. This is Theorem 5(ii)–(iii), which concern only the multiplier sets.

(iv) Input elimination. Δx enters $\Psi_s\xi$ only through $v = U\Delta x$, and $\min\{|\Delta x|^2 : U\Delta x = v\} = v^\top(UU^\top)^{-1}v$ when U has full row rank.

(v) Static dominates storage-based. Restrict the form $\mathcal{F} + \Psi^\top \Pi \Psi$ (variables $(\Delta\sigma, \Delta f, \Delta w)$, with $A = 0, B = W, C = I, E = U, D_y = LV$) to the subspace $\Delta\sigma = W\Delta f + U\Delta w$. There $\Delta\sigma^+ = \Delta\sigma$, the storage terms cancel, and the form equals $|LV\Delta f|^2 - g|\Delta w|^2 + (\Delta f, \Delta\sigma)^\top \Pi (\Delta f, \Delta\sigma)$, which is the static form at $(\Delta z, \Delta x) = (\Delta f, \Delta w)$. □

Proof of Remark 7 (relation to LBEN). Step 1 (the static inequality as a matrix). For slopes in $[0, 1]$ the diagonal multiplier contributes $\sum_j \lambda_j(-\Delta z_j^2 + \Delta z_j \Delta\sigma_j)$ with $\Delta\sigma = W\Delta z + U\Delta x$. With $g = \gamma^2$ the static inequality is therefore

$$\begin{bmatrix} V^\top L^\top LV - \frac{1}{2}(2\Lambda - \Lambda W - W^\top \Lambda) & \frac{1}{2}\Lambda U \\ \frac{1}{2}U^\top \Lambda & -\gamma^2 I \end{bmatrix} \preceq 0.$$

Step 2 (Schur complement). For $\gamma > 0$ this is equivalent to $2\Lambda - \Lambda W - W^\top \Lambda - 2V^\top L^\top LV - \frac{1}{2\gamma^2}\Lambda U U^\top \Lambda \succeq 0$.

Step 3 (rescale). Substituting $\Lambda = 2\gamma\Lambda'$ and dividing by 2γ gives $2\Lambda' - \Lambda' W - W^\top \Lambda' - \frac{1}{\gamma}V^\top L^\top LV - \frac{1}{\gamma}\Lambda' U U^\top \Lambda' \succeq 0$, the condition of Havens et al. (2023, Thm. 3.1). □

C Proofs of the stability results

Every statement of §7 used in this paper is restated here with a complete proof, so the paper is self-contained. All results in this appendix are *stability* results.

Standing assumptions. $x^+ = Ax + Bf(C^\top x)$ with A Schur and f decoupled and time-invariant; the sector or slope bounds are $[0, \mu_j]$ with $\mu_j > 0$ and $\varphi_j(0) = 0$; $P \succ 0$; $\theta = \mathrm{diag}(\theta_j) \succeq 0$. Classes with a nonzero lower bound α_j reduce to this case by a loop transformation. The classes are nested: $\mathcal{A}$ (sector-bounded) $\supseteq \mathcal{M}$ (monotone and sector-bounded) $\supseteq$ the slope-restricted class.

Lemma 16 (Sector secant)**.** *If $f \in \mathcal{A}$, then for every x there is $K(x) \in \mathfrak{B}$ with $f(C^\top x) = K(x)\, C^\top x$.*

Proof. Coordinatewise with $\sigma = C^\top x$: if $\sigma_j \neq 0$, take $k_j = \varphi_j(\sigma_j)/\sigma_j \in [0, \mu_j]$ by the sector bound; if $\sigma_j = 0$, then $\varphi_j(\sigma_j) = \varphi_j(0) = 0$ and any $k_j \in [0, \mu_j]$ works. □

C.1 Theorem 8, vertex part

Theorem (Theorem 8, vertex part, restated)**.** *For $P \succ 0$, $x^\top Px$ is a strict certificate over $\mathcal{A} \iff$ over $\mathcal{M} \iff$ over the slope-restricted class $\iff$ $P - A_K^\top P A_K \succ 0$ for all $K \in \mathfrak{V}$.*

Proof. Step 1 (vertices ⇒ certificate over $\mathcal{A}$). Fix x and $f \in \mathcal{A}$. By Lemma 16, $x^+ = A_{K(x)}x$, so

$$V(x^+) - V(x) = -x^\top \big(P - A_{K(x)}^\top P A_{K(x)}\big)x.$$

The map $K \mapsto x^\top Px - |A_K x|_P^2$ is a constant minus the squared P-norm of a vector affine in K, hence concave; by Lemma 10 its minimum over $\mathfrak{B}$ is attained at a vertex. Therefore $V(x^+) - V(x) \leq -\varepsilon|x|^2$ with $\varepsilon = \min_{K\in\mathfrak{V}} \lambda_{\min}(P - A_K^\top P A_K) > 0$, uniformly in f.

Step 2 (the chain of classes). A certificate over $\mathcal{A}$ is one over the smaller classes $\mathcal{M}$ and the slope-restricted class.

Step 3 (certificate over the slope-restricted class ⇒ vertices). Each linear map $f(\sigma) = K\sigma$ with $K \in \mathfrak{V}$ is slope-restricted, and for it $V(x^+) - V(x) = -x^\top(P - A_K^\top P A_K)x \leq -\varepsilon|x|^2$ for all x, i.e. $P - A_K^\top P A_K \succeq \varepsilon I \succ 0$. □

C.2 Theorem 8, collapse part

Theorem (Theorem 8, collapse part, restated)**.** *If $V(x) = x^\top Px + \sum_j \theta_j \int_0^{c_j^\top x} \varphi_j$ is a strict certificate over $\mathcal{M}$ with constant ε, then $\widetilde{P} = P + \frac{1}{2}C\theta\mu C^\top$, $\mu = \mathrm{diag}(\mu_j)$, is a strict quadratic certificate over $\mathcal{A}$ with the same ε. Hence a Lur'e–Postnikov certificate over $\mathcal{A}$ or $\mathcal{M}$ exists if and only if the vertex LMIs are feasible.*

The idea: over a class as rich as $\mathcal{M}$, an adversary can choose the activation so that the integral term takes its largest possible value while the state moves as it would under a vertex gain. The certificate must survive that choice, and what it then says is exactly a vertex LMI for the enlarged matrix $\widetilde{P}$. Two lemmas make this precise.

Lemma 17 (A lower bound on the increment of squares)**.** *Let $\eta(s,t) = \max\{t_+^2 - s_+^2,\ t_-^2 - s_-^2\}$. Then $\eta(s,t) \geq t^2 - s^2$ for all s, t.*

Proof. The two numbers in the maximum sum to $(t_+^2 + t_-^2) - (s_+^2 + s_-^2) = t^2 - s^2$. At least one of them is ≤ 0: if $t \geq s$ then $t_- \leq s_-$, so the second is ≤ 0; if $t < s$ then $t_+ \leq s_+$, so the first is. The larger of two numbers whose sum is $t^2 - s^2$ and one of which is ≤ 0 is at least $t^2 - s^2$. □

Lemma 18 (Adversarial monotone activation)**.** *Let $\mu > 0$, $\sigma, \sigma^+ \in \mathbb{R}$, and f between 0 and $\mu\sigma$. For every $e > 0$ there is $\varphi \in \mathcal{M}$ with $\varphi(\sigma) = \mathrm{f}$ and $\int_\sigma^{\sigma^+} \varphi \geq \frac{\mu}{2}\eta(\sigma, \sigma^+) - e$.*

Proof. Case $\sigma < \sigma^+$. Let $k_0 = \mathrm{f}/\sigma$ ($k_0 = 0$ if $\sigma = 0$), so $k_0 \in [0, \mu]$. For small $\delta > 0$ (with $\delta \leq -\sigma$ if $\sigma < 0$) let φ_δ equal $k_0 s$ for $s \leq \sigma$, be affine on $[\sigma, \sigma + \delta]$ from f to $\mu(\sigma + \delta)_+$, and equal μs_+ for $s \geq \sigma + \delta$. It is continuous, nondecreasing, sector-bounded in $[0, \mu]$ and satisfies $\varphi_\delta(\sigma) = \mathrm{f}$, so $\varphi_\delta \in \mathcal{M}$; and

$$\int_\sigma^{\sigma^+} \varphi_\delta \ \longrightarrow\ \frac{\mu}{2}\big((\sigma^+)_+^2 - \sigma_+^2\big) = \frac{\mu}{2}\,\eta(\sigma, \sigma^+) \qquad (\delta \to 0).$$

Case $\sigma^+ < \sigma$. Apply the first case to the reflection $\psi(s) = -\varphi(-s)$, which maps $\mathcal{M}$ to itself. *Case $\sigma^+ = \sigma$.* Take $\varphi(s) = k_0 s$; both sides vanish. □

Proof of the collapse part. Step 1 (fix a vertex and a state). Fix x and $K \in \mathfrak{V}$, and put $\mathrm{f} = KC^\top x$, $\sigma = C^\top x$ and $\sigma^+ = C^\top A_K x$.

Step 2 (choose the adversary). Choosing each φ_j by Lemma 18 independently gives $f \in \mathcal{M}$ with $f(\sigma) = \mathrm{f}$, so the next state is $x^+ = A_K x$, and

$$V(x^+) - V(x)\ \geq\ -x^\top(P - A_K^\top P A_K)x + \frac{1}{2}\sum_j \theta_j\mu_j\,\eta(\sigma_j, \sigma_j^+) - e\sum_j \theta_j.$$

Step 3 (use the certificate and let $e \to 0$). Since $V(x^+) - V(x) \leq -\varepsilon|x|^2$ for this f,

$$x^\top(P - A_K^\top P A_K)x\ \geq\ \frac{1}{2}\sum_j \theta_j\mu_j\,\eta(\sigma_j, \sigma_j^+) + \varepsilon|x|^2.$$

Step 4 (remove η). By Lemma 17, $\eta(\sigma_j, \sigma_j^+) \geq (\sigma_j^+)^2 - \sigma_j^2$, and $\sum_j \theta_j\mu_j\big((\sigma_j^+)^2 - \sigma_j^2\big) = x^\top\big(A_K^\top C\theta\mu C^\top A_K - C\theta\mu C^\top\big)x$. Rearranging,

$$x^\top\big(\widetilde{P} - A_K^\top \widetilde{P} A_K\big)x\ \geq\ \varepsilon|x|^2 \qquad \text{for all } x \text{ and } K \in \mathfrak{V}.$$

Step 5 (conclude). $\widetilde{P} \succeq P \succ 0$ because $C\theta\mu C^\top \succeq 0$, so by the vertex part (Step 1 of its proof, which keeps ε) $x^\top\widetilde{P}x$ is a strict certificate over $\mathcal{A}$ with constant ε. For the converse direction of the "if and only if", take $\theta = 0$. □

C.3 Proposition 9: Lur'e–Postnikov functions help over slope-restricted activations

Over the slope-restricted class the adversary of Lemma 18 is no longer admissible, and the integral term can be bounded more tightly by co-coercivity.

Lemma 19 (Co-coercive integral bound)**.** *If φ is slope-restricted in $[0, \mu]$ with $\mu > 0$ (hence continuous), $\Phi(s) = \int_0^s \varphi$, $\mathrm{f} = \varphi(\sigma)$ and $\mathrm{f}^+ = \varphi(\sigma^+)$, then*

$$\int_\sigma^{\sigma^+} \varphi\ \leq\ \mathrm{f}^+(\sigma^+ - \sigma) - \frac{(\mathrm{f}^+ - \mathrm{f})^2}{2\mu}.$$

Proof. Step 1. The function $u(s) = \Phi(s) - \Phi(\sigma^+) - \mathrm{f}^+(s - \sigma^+)$ is C^1 and convex, with μ-Lipschitz derivative $u' = \varphi - \mathrm{f}^+$, and $u(\sigma^+) = u'(\sigma^+) = 0$; hence $u \geq 0$. *Step 2.* The descent inequality at $y = s - u'(s)/\mu$ gives $0 \leq u(y) \leq u(s) - u'(s)^2/(2\mu)$. *Step 3.* At $s = \sigma$, $u(\sigma) = \mathrm{f}^+(\sigma^+ - \sigma) - \int_\sigma^{\sigma^+} \varphi$ and $u'(\sigma) = \mathrm{f} - \mathrm{f}^+$; substituting gives the claim. □

Lemma 20 (Dual certificate for the vertex LMIs)**.** *Let $n \geq 1$. If $Z_K \succeq 0$ $(K \in \mathfrak{V})$ satisfy $S = \sum_{K\in\mathfrak{V}}(Z_K - A_K Z_K A_K^\top) \prec 0$, then no $P \succ 0$ satisfies $P - A_K^\top P A_K \succeq 0$ for all $K \in \mathfrak{V}$.*

Proof. For such a P, $0 \leq \sum_K \mathrm{tr}\left((P - A_K^\top P A_K) Z_K\right) = \mathrm{tr}(PS) < 0$, a contradiction. □

Proposition (Proposition 9, restated)**.** *There is a single-nonlinearity system ($n = 2$, $m = 1$) with a strict Lur'e–Postnikov certificate over the activations slope-restricted in $[0, \mu]$ with $\varphi(0) = 0$ for every $\mu \leq 1.6666$, while no (strict or non-strict) quadratic certificate exists for any $\mu \geq 1.228$.*

Proof. Take $A = \left[\begin{smallmatrix} 3/4 & 1/2 \\ -1/4 & 3/4 \end{smallmatrix}\right]$, $b = e_1$, $c = (-1/2, 1)$, $x^+ = Ax + b\varphi(c^\top x)$.

Step 1 (a sufficient LMI). For $\xi = (x, \mathrm{f}, \mathrm{f}^+)$, $\sigma = c^\top x$, $\sigma^+ = c^\top(Ax + b\mathrm{f})$ and $\Delta\sigma = \sigma^+ - \sigma$, let

$$\begin{aligned}\mathcal{R}(\xi) = |Ax + b\mathrm{f}|_P^2 - |x|_P^2 + \theta\Big(\mathrm{f}^+\Delta\sigma - \tfrac{(\mathrm{f}^+ - \mathrm{f})^2}{2\mu}\Big) \\ + 2\lambda\mathrm{f}\big(\sigma - \tfrac{\mathrm{f}}{\mu}\big) + 2\nu\mathrm{f}^+\big(\sigma^+ - \tfrac{\mathrm{f}^+}{\mu}\big) + 2\kappa(\mathrm{f}^+ - \mathrm{f})\big(\Delta\sigma - \tfrac{\mathrm{f}^+ - \mathrm{f}}{\mu}\big).\end{aligned}$$

Step 2 (soundness). For an admissible φ put $\mathrm{f} = \varphi(\sigma)$ and $\mathrm{f}^+ = \varphi(\sigma^+)$. The λ and ν terms are nonnegative by the sector bound at σ and at σ^+, and the κ term by the slope bound applied to (σ, σ^+). With $V(x) = x^\top P x + \theta\Phi(c^\top x)$ and Lemma 19,

$$V(x^+) - V(x) = |Ax + b\mathrm{f}|_P^2 - |x|_P^2 + \theta\int_\sigma^{\sigma^+} \varphi \;\leq\; \mathcal{R}(\xi) \qquad (\theta, \lambda, \nu, \kappa \geq 0).$$

Hence $\mathcal{R}(\xi) \leq -\varepsilon|x|^2$ for all ξ makes V a strict certificate, with $\lambda_{\min}(P)|x|^2 \leq V(x) \leq (\lambda_{\max}(P) + \theta\mu|c|^2/2)|x|^2$.

Step 3 (the certificate). At $\mu = 1.6666$ the rational values $P = \left[\begin{smallmatrix} 83286/975715 & -3049/489713 \\ -3049/489713 & 10012/24953 \end{smallmatrix}\right]$, $\theta = 437501/997424$, $\lambda = 45248/605215$, $\nu = 4/951427$, $\kappa = 8/986891$ and $\varepsilon = 1007069/997313191225$ give $P \succ 0$ and $-\mathcal{R} - \varepsilon|x|^2 \succ 0$ as a 4×4 rational matrix, verified by exact $LDL^\top$ factorization. Since the class for $\mu' \leq \mu$ is contained in the class for μ, the certificate covers all $\mu \leq 1.6666$.

Step 4 (no quadratic certificate). Linear maps $\varphi(s) = Ks$ with $K \in \{0, \mu\}$ are admissible, so a quadratic certificate requires $P \succ 0$ with $P - A_K^\top P A_K \succeq 0$ for both K, where $A_K = A + Kbc^\top$. At $\mu = 1.228$,

$$Z_0 = \begin{bmatrix} 667247963424311/777474308339432 & -16033/375514 \\ -16033/375514 & 869812397903/408029587715014 \end{bmatrix},$$
$$Z_\mu = \begin{bmatrix} 448798508890/76908678268867 & 20399/730427 \\ 20399/730427 & 9095705441815/67960484843327 \end{bmatrix}$$

satisfy $Z_0, Z_\mu \succ 0$ and $-S \succ 0$ in exact rational arithmetic, so Lemma 20 rules out a quadratic certificate at $\mu = 1.228$, and by nesting of the classes for every $\mu \geq 1.228$. □

C.4 Strict conservatism of diagonal multipliers (System A)

Lemma 21 (Dual certificate for the diagonal LMI)**.** *The diagonal certificate asks for $P \succ 0$ and $\lambda \geq 0$ with $F = [A\ B]^\top P [A\ B] - \mathrm{diag}(P, 0) + \sum_j \lambda_j S_j \prec 0$, where S_j is the matrix of the form $2\mathrm{f}_j(x_j - \mathrm{f}_j/\mu)$. If $Y \succeq 0$, $Y \neq 0$, satisfies $[A\ B]Y[A\ B]^\top - Y_{xx} \succeq 0$ and $\langle S_j, Y\rangle \geq 0$ for all j, no diagonal certificate exists.*

Proof. For such (P, λ), $\langle F, Y\rangle = \langle P, [A\ B]Y[A\ B]^\top - Y_{xx}\rangle + \sum_j \lambda_j\langle S_j, Y\rangle \geq 0$, while $F \prec 0$ and $Y \succeq 0$, $Y \neq 0$ give $\langle F, Y\rangle < 0$. □

Take $A = \left[\begin{smallmatrix} 3/4 & 3/4 \\ -3/4 & 1/4 \end{smallmatrix}\right]$, $B = \left[\begin{smallmatrix} 1 & -1/2 \\ 0 & 1/2 \end{smallmatrix}\right]$, $C = I$ and the sector $[0, \mu]$ on both channels. At $\mu = 0.26$,

$$Y = \begin{bmatrix} 20/41 & -11/50 & 8/95 & -8/93 \\ -11/50 & 43/91 & 1/75 & 5/72 \\ 8/95 & 1/75 & 1/46 & -1/94 \\ -8/93 & 5/72 & -1/94 & 1/56 \end{bmatrix}$$

satisfies $Y \succ 0$, $[A\ B]Y[A\ B]^\top - Y_{xx} \succ 0$, $\langle S_1, Y\rangle = 34/28405$ and $\langle S_2, Y\rangle = 5/3276$, so by Lemma 21 no diagonal-multiplier certificate exists. On the other hand $P = \left[\begin{smallmatrix} 38/73 & 9/43 \\ 9/43 & 35/73 \end{smallmatrix}\right]$ satisfies the vertex LMIs of Theorem 8 at $\mu = 0.28$, hence at $\mu = 0.26$. So no strict diagonal certificate exists at $\mu = 0.26$ while a quadratic one exists at $\mu = 0.28$, which is the claim of §7; a grid of such certificates brackets both thresholds, with the same exact verification.

D FALSIFICATION TESTS

Each new statement was tested for counterexamples before and after its proof was written. A violation is a sampled quantity exceeding a certified bound by more than the solver tolerance; the largest relative excess over all samples is reported (negative: no excess). Certificates in these tests are solver values, so relative excesses of order 10^{-6} reflect solver accuracy. Table 5 lists the tests.

Table 5: Falsification tests of the theoretical results (solver values; a violation is an excess above the solver tolerance).

test	instances	samples	violations	max. rel. excess
Thm. 2: dissipation along trajectories with random time-varying piecewise-linear slope-restricted activations	163	4 890 000	0	$-8.3 \cdot 10^{-5}$
Thm. 15: DEQ gain on sampled input pairs; uniqueness of fixed points	59	295 000	0	$8.9 \cdot 10^{-10}$
Prop. 4: certified radius vs. attack distance (tiny ReLU MLPs, 720 directions)	20	600	0	$-6.0 \cdot 10^{-6}$
Thm. 5: monotonicity along nested partitions; top level vs. vertex value	97	388	0	$6.7 \cdot 10^{-6}$
Prop. 3: $g^\star = 4\,\mathrm{MaxCut}$ on random graphs	26	26	0	$5.7 \cdot 10^{-9}$
Prop. 9: Lur'e–Postnikov decrease for random slope-restricted activations at $\mu = 1.6666$	60	–	0	–

Two further checks concern DEQs. On 120 random DEQs the static certificate at the one-block level matched a sampled lower bound on $\sup_{K \in \mathfrak{B}} \|LV(I - KW)^{-1}KU\|_2$ to within 0.1% (solver values), and on 150 further instances with U of full row rank to within 0.21%. We conjecture that the static one-block level is class-exact for the fixed-point abstraction when U has full row rank, but the proof of Theorem 5(iv) does not transfer: the reduced block $V^\top L^\top LV - gW^\top(UU^\top)^{-1}W$ was indefinite in all 150 instances. We do not claim this conjecture.

E EXPERIMENTAL DETAILS

Bounds compared in E2. LipSDP-Layer and LipSDP-Neuron are our own implementations of the two valid variants of Fazlyab et al. (2019). For ReLU networks, γ_{true} is obtained by enumerating linear regions; the original enumeration is exact for one hidden layer and a lower estimate for two. App. H independently verifies all two-layer ReLU regions. For every network we also report the sampled lower bound $\max_x \|J(x)\|_2$ over the evaluation inputs, where J is the Jacobian.

Software and verification. Python 3.12, CVXPY with Clarabel (SCS as fallback for stability SDPs), PyTorch for training, all on one CPU (Apple M1 Pro), one thread per solver process. Storage-based SDP proposals use Algorithm 1, increasing the strict margin when needed. Static DEQ verification omits the storage matrix; one-layer enumeration checks the resulting norm bounds directly. Appendix H specifies the additional exact primal/dual and radius audits. Gain candidates are inflated and rounded outward before exact acceptance.

Algorithm 1 Storage-based SDP proposals: certify, then verify exactly.

Input: model data, level (Π_{diag}, $\Pi(\mathcal{P})$ or vertex).

1. For $\varepsilon \in (10^{-7}, 10^{-6}, \ldots, 10^{-3})$: solve the SDP minimizing g with margins ε on every LMI ($P \succeq \varepsilon I$, main LMI $\preceq -\varepsilon I$, block constraints with margin ε) with Clarabel (SCS as fallback for stability SDPs).
2. Read the solution as IEEE doubles; every double is a dyadic rational, so model data and certificate are exact rationals. Inflate g to $g(1 + 10^{-6})$, clip diagonal multipliers at 0, and keep block multipliers as they are.
3. Check every constraint exactly: $P \succ 0$, the block conditions (4), and the main LMI (or all vertex LMIs).
4. Each positive-definiteness check of a matrix M: pick $\tau > 0$, compute a floating-point Cholesky factor R of $M - \tau I$, compute $\Delta = M - \tau I - R^\top R$ *exactly* in integer arithmetic, and accept if $\|\Delta\|_F^2 < \tau^2$ exactly. Then $M \succeq (\tau - \|\Delta\|_F) I \succ 0$. If input elimination is used with a floating-point approximation R_U of $(UU^\top)^{-1}$, also verify $R_U \preceq (UU^\top)^{-1}$ exactly via $UU^\top - UU^\top R_U UU^\top \succ 0$.
5. Return an outward upper bound on $\sqrt{g}$ as the certified gain at the first ε that passes; otherwise report failure.

E1. *Gaussian cells:* A Gaussian rescaled to spectral radius $\sim U(0.3, 0.9)$, B, C Gaussian with entries of variance $1/n$; *DEQ cells:* $A = 0$, $B = W$ Gaussian with variance $1/m$, $C = I$. Inputs $d = 2$, output $p = 1$, E Gaussian, C_y Gaussian with variance $1/n$, $D_y = 0$ (Gaussian cells) or Gaussian with variance $1/m$ (DEQ cells). Levels: diagonal, singletons, contiguous pairs, one block, vertex. Stability: bisection of $\bar{\mu}$ (sector $[0, \mu]$ on every channel; tolerance 10^{-3}) up to the smallest μ on a 0.01 grid at which some vertex gain matrix is not Schur, which bounds every level from above; the verified value is the largest of $\bar{\mu}, \bar{\mu} - 10^{-3}, \ldots, \bar{\mu} - 3 \cdot 10^{-3}$ that passes Algorithm 1. Block multipliers apply verbatim to the non-incremental sector constraints with σ in place of $\Delta\sigma$. Incremental gain: slopes in $[0, 0.8\bar{\mu}_{\mathrm{diag}}]$, all levels.

E2. Iris: a seed-zero NumPy permutation splits 150 points into 100 training and 50 test points; features use training-set means and standard deviations, with 400 epochs and batch size 32. Two moons: 2000 training and 1000 test points, noise 0.1, 200 epochs. MNIST and Fashion-MNIST: official training/test splits, pixels divided by 255, then 50 principal components without whitening, with centring and PCA fitted on training data only; 8 epochs and batch size 256. Two moons also uses batch size 256. All use Adam with learning rate $3 \cdot 10^{-3}$ and float64. Output $L = I$ (all logits). Levels: diagonal, singletons, contiguous pairs, contiguous blocks of 4, and the vertex value (enumeration of the 2^m patterns for one hidden layer, layerwise vertex SDP (3) for two). LipSDP-Neuron is solved separately from the formulation of Fazlyab et al. (2019). The ReLU Lipschitz constant γ_{true} enumerates realizable activation patterns by linear programming (a pattern counts if an interior point exists with margin above 10^{-9}, searched in $\|x\|_\infty \leq 10^6$), so it can only miss regions and is a lower bound on the true constant. For one hidden layer it is exact: the number of regions found equals the maximum $\sum_{i \leq d} \binom{m}{i}$ for m hyperplanes in $\mathbb{R}^d$ in all 24 one-hidden-layer networks, so no region was missed. We re-ran all 36 ReLU networks (24 one-layer and 12 two-layer) without the box, with normalized constraints and margin 10^{-12}: all 36 values and region counts were unchanged, and each maximizing region was confirmed by a forward pass at its witness point. The sampled lower bound is the largest $\|J(x)\|_2$ over 800 points (400 test points and Gaussian perturbations) and 30 Adam steps of gradient ascent from 40 starts.

E3. $z = \mathrm{ReLU}(Wz + Ux + b)$, $F = Vz + c$, width 40, inputs in $[0, 1]^{784}$. Contractive: $W = \widehat{W} \min(1, 0.9/\|\widehat{W}\|_2)$. Monotone: $W = (1 - \omega)I - A^\top A + B - B^\top$ with $\omega = 0.5$, so that $I - (W + W^\top)/2 \succeq \omega I$ (Winston & Kolter, 2020). Fixed points by forward–backward iteration, whose linear part has norm $q < 1$, with the a posteriori error bound $q/(1 - q)$ times the last step (used in Prop. 4); training by Adam (learning rate 10^{-3}, batch 128, 10 epochs) through one differentiable step at the fixed point. Blocks for pairs and quadruples are formed independently and greedily by coupling $|W_{ij}| + |W_{ji}|$; they need not be nested, so the refinement monotonicity theo-

rem does not order these two partitions. The input is eliminated exactly (Prop. 6(iv)); the inequality $R \preceq (UU^\top)^{-1}$ is verified in exact arithmetic. PGD: ℓ_2 steps within $[0,1]^{784}$, 40 steps, binary search over the radius (10 halvings up to 4). *What is and is not verified.* All reported SDP gain-to-radius chains use exact dyadic residuals, exact rational arithmetic and checked outward conversions (App. H). MNIST retains six saved models with replacement multiplier witnesses. Because the original Fashion-MNIST/KMNIST checkpoints were not located, we retrained their twelve models with the stated setup, saved the weights and verified new gain/radius packets. All 144 threshold counts and twelve clean accuracies match the historical records; this does not establish historical weight identity. E3 gain/accuracy tables and plots use the new released records; historical timings are labeled separately. Norm baselines, sampled lower bounds and PGD remain numerical diagnostics. The legacy radius routine used heuristic floating-point error estimates; an underflow counterexample motivated the exact replacement, whose proof and replay scope appear in Appendix H. The certified model is the real-arithmetic network with those stored parameters and stored input values; this does not by itself certify a deployed floating-point prediction routine. PGD distance medians include initially misclassified inputs as zero and exclude failed attacks on initially correct inputs. Reported distances average these seed-level medians. Distance/radius ratios use initially correct points and retain failed attacks as infinity.

F ADDITIONAL RESULTS

This appendix reports the full results behind the summary tables of §8. For E2, Table 7 gives the bounds per architecture, Table 8 the accuracy and the LipSDP bounds, and Table 9 the diagonal, pair, quadruple and vertex levels. Table 10 gives E3 on MNIST at all levels. Tables 11 and 12 give E1 per cell and level for stability and for the incremental gain. Table 13 reports solve-plus-verify times.

Open problems. Future directions include incorporating realizable activation patterns to reduce abstraction conservatism; improving neuron-block selection to obtain tighter certificates at a given computational budget; and extending the incremental gain-to-margin framework to richer storage functions and dynamic multipliers. Another open question is whether the static one-block DEQ certificate recovers the optimal gain bound over its independent-slope fixed-point abstraction.

Static versus storage certificates on random DEQs. On 120 random DEQs ($m \leq 6$), the static *diagonal* certificate was smaller than $\sqrt{g^\star_{\mathrm{vx}}}$ by more than 0.1% in 82% of instances, with ratio up to 7.9 (solver values).

Implementation agreement (E2). The maximum relative discrepancy between independently solved LipSDP-Neuron and diagonal raw solver outputs over all 72 networks is $3.8 \cdot 10^{-8}$, dividing by the diagonal raw output. This statistic precedes outward adjustment and verification; it does not compare the displayed verified bounds. LipSDP-Layer is looser by a median 6.3% and up to 22%.

Comparison with norm-based DEQ bounds (E3). The static SDP certificates, whose diagonal level contains LBEN, are tighter than the norm bounds used for contractive and monotone DEQs (Winston & Kolter, 2020; Havens et al., 2023) by median factors (over the 46 targets) of 5.6, 8.2 and 6.6 and of 1.8–1.9, respectively, on MNIST, Fashion-MNIST and KMNIST.

Lur'e–Postnikov versus quadratic certificates (§7). For random single-loop systems $x^+ = Ax + b\varphi(c^\top x)$ (A Schur, $n \in \{2,3,4\}$, 40 each) we bisect the largest certified μ for a quadratic certificate (vertex LMIs) and for $V(x) = x^\top P x + \theta \int_0^{c^\top x} \varphi$, over monotone and over slope-restricted activations (bisection tolerance 10^{-3}, solver values). Over monotone activations the two bounds agreed in all 120 instances, as Theorem 8 requires; over slope-restricted activations the Lur'e–Postnikov function improved the bound by more than 1% in 15 instances; over all 120 instances the median ratio was 1.000 and the maximum 1.285. Pure integral certificates ($P = 0$) are not considered: for $m < n$ they vanish on the kernel of $c^\top$ and cannot be positive definite. Table 6 summarizes E1 per cell.

Table 6: E1. G: Gaussian systems; D: systems with $A = 0$, $C = I$. *gap*: fraction of instances whose price exceeds one by more than 0.5%; *med.*, *max*: median and maximum price (solver values); $\bar{\mu}$ has a bisection tolerance of 10^{-3}; ordering checks allow 2×10^{-3} for adjacent stability levels, 10^{-3} against the Schur upper bound, and 10^{-4} relative gain undershoot against vertex; *verified*: instances in which every level passed Algorithm 1 for both problems.

		stability ($\bar{\mu}$)			gain (γ)			
cell	N	gap	med.	max	gap	med.	max	verified
G (4,2)	100	0.21	1.000	1.203	0.39	1.002	1.765	100/100
G (6,3)	100	0.43	1.004	1.105	0.65	1.017	1.252	100/100
G (8,4)	100	0.47	1.004	1.069	0.73	1.028	1.270	100/100
G (10,6)	100	0.58	1.009	1.120	0.81	1.041	1.461	100/100
G (12,8)	100	0.81	1.021	1.187	0.90	1.055	1.648	100/100
D (2,2)	100	0.00	1.000	1.002	0.21	1.000	1.068	100/100
D (4,4)	100	0.33	1.002	1.043	0.71	1.015	1.118	100/100
D (6,6)	100	0.51	1.005	1.045	0.80	1.023	1.117	100/100
D (8,8)	100	0.69	1.013	1.036	0.90	1.038	1.157	100/100

Table 7: E2 per architecture: ℓ_2 bounds (all logits; mean over 3 seeds). Only SDP certificates are exactly verified; norm and sampled bounds are numerical. *naive*: product of spectral norms; *diag*: diagonal multipliers (verified bounds; raw-solver agreement is described above); *pairs, quads*: block levels; *vertex*: returned feasible layerwise reference bound; *true*: numerical value from linear-region enumeration (ReLU only; App. H independently certifies complete region covers for all twelve two-layer models); *lower*: sampled $\max \|J(x)\|_2$.

data	width	act.	naive	diag	pairs	quads	vertex	true	lower
Iris	(8)	relu	13.76	10.39	10.39	10.39	10.39	10.39	9.42
Iris	(8)	tanh	15.17	12.85	12.85	12.85	12.85	–	11.58
Iris	(16)	relu	14.45	11.52	11.52	11.52	11.52	11.50	11.02
Iris	(16)	tanh	15.58	13.78	13.78	13.78	13.78	–	12.19
Iris	(8,8)	relu	25.12	17.16	17.14	17.14	17.14	17.12	13.89
Iris	(8,8)	tanh	27.60	23.89	23.89	23.89	23.89	–	17.77
moons	(8)	relu	17.00	11.78	11.77	11.77	11.77	11.32	8.70
moons	(8)	tanh	28.27	17.04	16.99	16.91	16.83	–	14.65
moons	(16)	relu	37.96	27.14	27.14	27.14	27.14	19.80	13.73
moons	(16)	tanh	34.32	21.09	21.09	21.09	21.09	–	18.63
moons	(8,8)	relu	66.37	39.31	39.31	39.31	39.31	27.11	24.78
moons	(8,8)	tanh	90.98	54.50	54.50	54.50	54.50	–	36.81
MNIST	(8)	relu	13.01	9.24	9.24	9.22	9.21	9.21	9.21
MNIST	(8)	tanh	12.02	10.45	10.45	10.45	10.45	–	10.24
MNIST	(16)	relu	14.34	9.87	9.83	9.80	9.79	9.79	9.56
MNIST	(16)	tanh	15.56	12.39	12.37	12.35	12.34	–	12.00
MNIST	(8,8)	relu	23.89	12.96	12.89	12.84	12.81	12.63	12.25
MNIST	(8,8)	tanh	34.52	19.59	19.50	19.44	19.41	–	17.51
Fashion	(8)	relu	13.41	10.19	10.19	10.19	10.19	10.19	10.19
Fashion	(8)	tanh	15.66	12.99	12.99	12.99	12.99	–	12.80
Fashion	(16)	relu	14.64	10.52	10.52	10.51	10.51	10.51	10.37
Fashion	(16)	tanh	15.97	13.99	13.99	13.99	13.99	–	13.63
Fashion	(8,8)	relu	29.07	18.71	18.70	18.69	18.68	18.68	18.43
Fashion	(8,8)	tanh	39.72	31.76	31.75	31.75	31.75	–	28.95

G EXTENDED RELATED WORK: THE CONTROL LITERATURE

This appendix expands the corresponding paragraph of §2. Here we record only what bears on the claims of this paper.

Table 8: E2, accuracy and published bounds, mean±std over 3 seeds. LipSDP-L and LipSDP-N: our CVXPY implementations of LipSDP-Layer and LipSDP-Neuron (raw solver values, not the adjusted verified bounds in Table 9).

data	width	act.	acc.	naive	LipSDP-L	LipSDP-N
Iris	(8)	relu	0.953	13.8±0.9	11.47	10.4±0.3
Iris	(8)	tanh	0.960	15.2±0.6	13.73	12.9±0.7
Iris	(16)	relu	0.953	14.5±0.3	12.44	11.5±1.3
Iris	(16)	tanh	0.960	15.6±0.3	14.55	13.8±0.4
Iris	(8,8)	relu	0.960	25.1±1.8	19.25	17.2±4.0
Iris	(8,8)	tanh	0.967	27.6±2.0	25.23	23.9±0.3
moons	(8)	relu	0.930	17.0±12.6	12.25	11.8±7.0
moons	(8)	tanh	0.994	28.3±5.2	18.89	17.0±3.4
moons	(16)	relu	0.998	38.0±0.7	28.42	27.1±1.0
moons	(16)	tanh	0.998	34.3±5.2	23.02	21.1±3.0
moons	(8,8)	relu	0.965	66.4±36.3	42.67	39.3±20.4
moons	(8,8)	tanh	0.999	91.0±10.4	59.56	54.5±3.4
MNIST	(8)	relu	0.908	13.0±0.4	9.65	9.2±0.1
MNIST	(8)	tanh	0.903	12.0±0.6	10.64	10.5±0.1
MNIST	(16)	relu	0.942	14.3±0.3	10.57	9.9±0.4
MNIST	(16)	tanh	0.932	15.6±0.1	12.88	12.4±0.1
MNIST	(8,8)	relu	0.913	23.9±2.2	14.98	13.0±1.1
MNIST	(8,8)	tanh	0.905	34.5±1.7	21.48	19.6±1.0
Fashion	(8)	relu	0.825	13.4±1.0	10.50	10.2±0.5
Fashion	(8)	tanh	0.820	15.7±0.6	13.23	13.0±0.1
Fashion	(16)	relu	0.846	14.6±0.6	11.06	10.5±0.5
Fashion	(16)	tanh	0.842	16.0±1.1	14.36	14.0±0.2
Fashion	(8,8)	relu	0.821	29.1±6.3	20.56	18.7±2.7
Fashion	(8,8)	tanh	0.825	39.7±1.3	33.53	31.8±1.9

Vertex LMIs and full-block multipliers. Boyd et al. (1994, Sec. 6.3.2, Eq. (6.53)) already optimize quadratic gain bounds for continuous-time polytopic inclusions. Scherer's full-block S-procedure (Scherer, 2001, App. B, Theorem 8) is an algebraic equivalence for strict quadratic inequalities over continuously parameterized subspaces on a compact path-connected set; it is not restricted to continuous-time dynamics. For our strict dissipation inequality, take his $U = I$, $Q = \mathcal{F}(P, g)$, $V = \Psi$, the ambient subspace to be the full space, and $\mathcal{S}(K) = \mathrm{im}[K; I]$ over the slope box. The theorem supplies a symmetric Π positive on every graph and satisfying $\mathcal{F} + \Psi^\top \Pi \Psi \prec 0$. Evaluating this inequality at $\xi = (0, \Delta f, 0)$ gives $\Pi_{11} + B^\top P B + D_y^\top D_y \prec 0$, hence $\Pi_{11} \prec 0$ because $P \succeq 0$. Thus strict top-level losslessness is a specialization of that theorem; equality of infima follows with Lemma 11. The proof in App. B.4 checks all hypotheses without requiring C to have full column rank. We do not claim a new general lossless S-procedure. Curvature restrictions also yield finite vertex tests in Scherer (2001, Sec. 4.1, Eq. (32)). His parameter-domain subdivision (Scherer, 2001, Sec. 4.2) differs from our grouping of neuron coordinates. Theorem 5 also records validity, refinement and the singleton characterization of the specified block cone; Theorem 2 and Proposition 4 give the incremental neural-network reference and margin chain.

Multipliers. Repeated-nonlinearity constraints (Kulkarni & Safonov, 2002; Noori et al., 2024) cannot in general be transferred unchanged to increments at different neuron operating points. Our hierarchy instead uses valid block slope constraints. Dynamic Zames–Falb multipliers (Willems & Brockett, 1968; Carrasco et al., 2016; Ahmad et al., 2015) exploit temporal relations of time-invariant activations absent from the one-step quadratic/static-multiplier reference studied here. Extending such analyses to neural robustness requires an incremental gain-to-margin chain. The static fixed-point and Lur'e–Postnikov characterizations retain their separately stated scopes.

Lyapunov functions for discrete-time Lur'e systems. Drummond & Valmorbida (2023, Sec. II–III) use generalized quadratic forms in state and activation coordinates, future-output integral terms, and relaxed parameter signs. Our $P \succ 0$, nonnegative-integral-weight family is a restricted in-

Table 9: E2, verified feasible bounds at the displayed levels, mean±std over 3 seeds. Independent solves and verification adjustments need not preserve the ordering of class optima; these entries are not optimum estimates.

data	width	act.	diag	pairs	quads	vertex
Iris	(8)	relu	10.4±0.3	10.4±0.3	10.4±0.3	10.4±0.3
Iris	(8)	tanh	12.9±0.7	12.9±0.7	12.9±0.7	12.9±0.7
Iris	(16)	relu	11.5±1.3	11.5±1.3	11.5±1.3	11.5±1.3
Iris	(16)	tanh	13.8±0.4	13.8±0.4	13.8±0.4	13.8±0.4
Iris	(8,8)	relu	17.2±4.0	17.1±4.0	17.1±4.0	17.1±4.0
Iris	(8,8)	tanh	23.9±0.3	23.9±0.3	23.9±0.3	23.9±0.3
moons	(8)	relu	11.8±7.0	11.8±7.0	11.8±7.0	11.8±7.0
moons	(8)	tanh	17.0±3.4	17.0±3.5	16.9±3.6	16.8±3.6
moons	(16)	relu	27.1±1.0	27.1±1.0	27.1±1.0	27.1±1.0
moons	(16)	tanh	21.1±3.0	21.1±3.0	21.1±3.0	21.1±3.0
moons	(8,8)	relu	39.3±20.4	39.3±20.4	39.3±20.4	39.3±20.4
moons	(8,8)	tanh	54.5±3.4	54.5±3.4	54.5±3.4	54.5±3.4
MNIST	(8)	relu	9.2±0.1	9.2±0.1	9.2±0.2	9.2±0.2
MNIST	(8)	tanh	10.5±0.1	10.5±0.1	10.5±0.1	10.5±0.1
MNIST	(16)	relu	9.9±0.4	9.8±0.4	9.8±0.3	9.8±0.3
MNIST	(16)	tanh	12.4±0.1	12.4±0.1	12.4±0.1	12.3±0.1
MNIST	(8,8)	relu	13.0±1.1	12.9±1.2	12.8±1.2	12.8±1.2
MNIST	(8,8)	tanh	19.6±1.0	19.5±1.0	19.4±1.0	19.4±1.0
Fashion	(8)	relu	10.2±0.5	10.2±0.5	10.2±0.5	10.2±0.5
Fashion	(8)	tanh	13.0±0.1	13.0±0.1	13.0±0.1	13.0±0.1
Fashion	(16)	relu	10.5±0.5	10.5±0.5	10.5±0.5	10.5±0.5
Fashion	(16)	tanh	14.0±0.2	14.0±0.2	14.0±0.2	14.0±0.2
Fashion	(8,8)	relu	18.7±2.7	18.7±2.7	18.7±2.7	18.7±2.7
Fashion	(8,8)	tanh	31.8±1.9	31.8±1.9	31.7±1.9	31.7±1.9

Table 10: E3 on MNIST, all levels, mean over 3 seeds: gain for all logits and median class-pair gain, estimated certified accuracy at ℓ_2 radii 0.1, 0.3, 0.5, and average certified radius. †A sampled value for the fixed-point abstraction, not a certificate; the accuracies and radii in that row are not guarantees. Norm-baseline rows are numerical estimates, not rounding-verified certificates. The exact checks cover SDP gain-to-radius chains; see App. H.

variant	level	γ all	γ pair	0.1	0.3	0.5	radius
contr.	baseline	173.7	182.6	0.002	0.000	0.000	0.039
	diag	40.4	32.3	0.866	0.239	0.013	0.225
	single	40.4	32.3	0.866	0.239	0.013	0.225
	pairs	40.0	32.2	0.866	0.242	0.013	0.226
	quads	39.6	32.0	0.867	0.248	0.014	0.228
	lower†	38.2	31.6	0.870	0.260	0.016	0.231
mon.	baseline	41.2	40.8	0.816	0.071	0.000	0.175
	diag	26.2	22.6	0.909	0.555	0.114	0.318
	single	26.2	22.6	0.909	0.555	0.114	0.318
	pairs	26.1	22.5	0.909	0.560	0.118	0.320
	quads	25.8	22.3	0.910	0.566	0.123	0.322
	lower†	24.4	21.5	0.913	0.589	0.142	0.332

stance, not their entire $N = 0$ family. Their Theorem 1 gives sufficient LMIs under sector and slope assumptions. Although presented for SISO systems, they explicitly discuss a MIMO extension; channel count is not the distinction. Our Theorem 8 instead characterizes the stated restricted family over sector/monotone classes. Their broader family and input–output gain results do not establish our restricted-family slope separation (Proposition 9) or the incremental pair-to-margin chain. Neither comparison implies dominance across different function and activation classes.

Table 11: E1, median returned sector bound $\bar{\mu}$ per cell and level (stability only; solver values, all verification records are stored separately after possible downward adjustment). "block" is the one-block level. The ordering check allows 2×10^{-3} between adjacent levels and 10^{-3} against the Schur upper bound; raw medians can invert within tolerance.

cell	N	diag	single	pairs	block	vertex
G (4,2)	100	1.108	1.108	1.108	1.108	1.109
G (6,3)	100	1.076	1.076	1.083	1.088	1.087
G (8,4)	100	0.909	0.909	0.912	0.912	0.912
G (10,6)	100	0.760	0.760	0.768	0.773	0.773
G (12,8)	100	0.671	0.672	0.682	0.695	0.695
D (2,2)	100	1.015	1.014	1.014	1.014	1.014
D (4,4)	100	0.910	0.910	0.910	0.910	0.910
D (6,6)	100	0.853	0.853	0.858	0.867	0.867
D (8,8)	100	0.827	0.827	0.829	0.833	0.833

Table 12: E1, median incremental gain γ at slopes $[0, 0.8\bar{\mu}_{\mathrm{diag}}]$ per cell and level (solver values, all verification records are stored separately after upward adjustment). The check allows 10^{-4} relative undershoot against the vertex value; numerical inversions do not establish an inversion of class optima.

cell	N	diag	single	pairs	block	vertex
G (4,2)	100	6.866	6.866	6.856	6.856	6.856
G (6,3)	100	5.914	5.914	5.858	5.745	5.745
G (8,4)	100	6.780	6.780	6.579	6.363	6.363
G (10,6)	100	6.237	6.237	6.216	6.006	6.006
G (12,8)	100	5.824	5.824	5.780	5.541	5.541
D (2,2)	100	4.469	4.469	4.435	4.435	4.435
D (4,4)	100	4.262	4.262	4.218	4.178	4.178
D (6,6)	100	4.729	4.729	4.602	4.519	4.519
D (8,8)	100	4.631	4.631	4.623	4.474	4.474

Comparison with optimized diagonal reparameterization. Consider a fixed network $f(x) = W_1 \,\mathrm{ReLU}(W_0 x + b)$ with a linear output and Euclidean norms. The positive diagonal reparameterization in Kuang et al. (2026, Sec. 2.3.1, Eq. (8)) preserves the function when the hidden bias is also scaled by D^{-1}. Define

$$B_{\mathrm{scale}} = \inf_{D \succ 0 \text{ diagonal}} \|W_1 D\|_2 \|D^{-1} W_0\|_2, \qquad \gamma_{\mathrm{vx}} = \max_{K \in \{\mathrm{diag}(k) : k \in \{0,1\}^m\}} \|W_1 K W_0\|_2.$$

The latter is the one-hidden-layer reference of §4.1, namely the established derivative-box bound (Scaman & Virmaux, 2018; Latorre et al., 2020). For every such D and K, commutation gives

$$\|W_1 K W_0\|_2 = \|(W_1 D) K (D^{-1} W_0)\|_2 \le \|W_1 D\|_2 \|D^{-1} W_0\|_2.$$

Maximizing over K and taking the infimum over D proves $\gamma_{\mathrm{vx}} \le B_{\mathrm{scale}}$. This comparison uses the global scaling infimum, not a particular optimizer's returned value.

For a strict example, take $f(x) = \mathrm{ReLU}(x) - \mathrm{ReLU}(x-1)$, so $W_0 = (1,1)^\top$, $W_1 = (1,-1)$ and $b = (0,-1)^\top$. The function is zero for $x \le 0$, equals x on $[0,1]$, and equals one for $x \ge 1$; hence its true Lipschitz constant is 1. Also $\gamma_{\mathrm{vx}} = \max_{k_1,k_2 \in \{0,1\}} |k_1 - k_2| = 1$. The storage $P_1 = W_1^\top W_1 \succeq 0$ and $g = 1$ satisfy (3) directly. For $D = \mathrm{diag}(d_1, d_2) \succ 0$,

$$\left(\|W_1 D\|_2 \|D^{-1} W_0\|_2\right)^2 = (d_1^2 + d_2^2)(d_1^{-2} + d_2^{-2}) = 4 + (d_1/d_2 - d_2/d_1)^2 \ge 4,$$

with equality at $d_1 = d_2$, so $B_{\mathrm{scale}} = 2$. Diagonal quadratic constraints already certify 1 here: for input increment u and activation increments q_i, the slope constraints give $q_i(u - q_i) \ge 0$, and

$$u^2 - (q_1 - q_2)^2 = (u - q_1 - q_2)^2 + 2q_1(u - q_1) + 2q_2(u - q_2) \ge 0.$$

Thus this elementary example separates the two bounds, not full-block from diagonal multipliers. It establishes neither a new priority claim nor a computational advantage, and does not compare against training or architecture changes proposed by Kuang et al. (2026).

Table 13: Median solve-plus-verify time per certificate (seconds, one CPU core). The reference and one-block levels both have exponentially many vertex constraints. For E2 the top-level column is the layerwise vertex value (enumeration at depth one); for E3 the reference is not computed at width 40. Each E3 row summarizes 276 targets from six historical models. These timings include solver retries and gain-feasibility verification, but exclude training and the later complete radius/dual audits. Fashion/KMNIST timings are not measurements of replacement runs; – denotes an unreported or uncomputed level.

problem	m	LMIs at top level	diag	pairs	quads	one block	vertex
E1 stability	2	2^2	0.19	0.17	–	0.16	0.16
E1 stability	3	2^3	0.29	0.29	–	0.34	0.43
E1 stability	4	2^4	0.50	0.45	–	0.68	1.41
E1 stability	6	2^6	0.81	0.98	–	4.48	11.26
E1 stability	8	2^8	1.36	1.29	–	21.24	72.09
E2 Lipschitz (8)	8	2^8 per layer	0.17	2.53	2.31	–	0.07
E2 Lipschitz (16)	16	2^{16} per layer	0.99	4.82	5.37	–	31.10
E2 Lipschitz (8,8)	16	2^8 per layer	0.30	2.46	2.55	–	7.77
E3 MNIST (historical)	40	2^{40}	10.13	35.34	41.90	–	–
E3 Fashion (historical)	40	2^{40}	8.64	30.30	37.47	–	–
E3 KMNIST (historical)	40	2^{40}	8.09	27.75	33.29	–	–

Table 14: How the results relate to prior work.

prior work	what it provides	what this paper adds
LipSDP-Neuron/-Layer (Fazlyab et al., 2019); LBEN, recurrent equilibrium networks (REN) (Revay et al., 2020; 2024; Havens et al., 2023)	valid certificates with diagonal multipliers	numerical comparisons and certified class-price enclosures (App. H); LBEN is contained in our diagonal fixed-point level (§6)
LipSDP-Network (Fazlyab et al., 2019)	coupled multipliers; invalid (Pauli et al., 2022; Wang & Manchester, 2023)	a valid block hierarchy between diagonal and reference (Theorem 5)
Optimal layer-wise LP relaxation (Salman et al., 2019)	the looseness of LP relaxations for input-specific verification, against their optimal member	the analogous question for global SDP certificates and their multipliers, with a class-exact reference
Reachability-based negative results (Kuang et al., 2026)	constructions on which SDP bounds fail through the abstraction	certified multiplier prices and separated abstraction/storage losses for one hidden layer and twelve two-layer models

Scale. Scalable SDP-based Lipschitz estimation has been demonstrated on CIFAR10 and ImageNet (Wang et al., 2024); other methods build Lipschitz constraints into architecture or training (Araujo et al., 2023; Wang & Manchester, 2023; Havens et al., 2023; Winston & Kolter, 2020). Our experiments address a different task: evaluating class references and auditing sources of conservatism where these comparisons can be checked. Small-block certification does not require full slope-box enumeration, but this study does not establish its performance across all large architectures.

H INDEPENDENT EXACT AUDITS

The audits replay saved weights, including the twelve newly trained Fashion/KMNIST replacement models and the four additional width-80 MNIST models described below; replay itself does not retrain them. They also include one explicitly saved reconstruction of an E1 seed. They supplement, rather than reinterpret as optimal, the returned feasible values in the original experiments. The exact witness packets and separate replay commands are retained with the accompanying code.

H.1 CLASS-OPTIMUM ENCLOSURES AND SEPARATED MULTILAYER LOSSES

Let T be the true global Euclidean Lipschitz constant of a ReLU map, B the maximum independent-slope product norm, R the infimum of its layerwise quadratic-storage vertex certificates, and D the diagonal-certificate infimum. Then $T \le B \le R \le D$. Convexity of the operator norm in each individual slope implies that B is attained at binary slopes, by replacing slopes successively with endpoints. This does not establish realizability or $B = R$.

Complete region verification. For each activation prefix, exact dyadic propagation gives signed affine constraints $a_i^\top x + b_i > 0$. A numerical LP only proposes either a strict witness or an exclusion. Each witness is checked against every exact constraint. An exclusion must supply rational $\lambda_i \ge 0$ with $\sum_i \lambda_i = 1$, $\sum_i \lambda_i a_i = 0$, and $\sum_i \lambda_i b_i \le 0$. A strict feasible point would make their weighted sum positive, a contradiction. The stored leaves form a prefix-free complete cover of all binary patterns; an independent checker verifies this cover and every witness without solving an LP. Failed rational reconstruction or an identically zero affine preactivation causes audit failure, not pruning; neither prevented these audits. For a continuous finite piecewise-affine map, the maximum norm on the resulting full-dimensional regions equals T: line segments split into affine pieces, and the reverse inequality holds within each open region.

For two-input Jacobians, exact squared comparisons enclose the analytic eigenvalues of their 2×2 Gram matrices. In higher dimensions, an exact Rayleigh quotient bounds the largest Gram eigenvalue below and an exactly checked $\beta I - JJ^\top \succ 0$ bounds it above. With wide full-row-rank W_0, region feasibility is checked in $\sigma = W_0 x + b_0$ coordinates after proving $W_0 W_0^\top \succ 0$. Spectral bounds still use the original Jacobian, so this change does not alter the Euclidean input metric. Maximization over the complete region cover encloses T; enumeration of all 2^{16} binary slope patterns encloses B for every audited 8–8 model.

Actual class infima. Exactly replayed primal witnesses give upper bounds U_R, U_D. With positive intervals $[T_L, T_U]$, $[B_L, B_U]$, and any verified reference lower bound $R_L \ge B_L$,

$$1 \le D/R \le U_D/R_L, \qquad \max(1, R_L/B_U) \le R/B \le \min(U_R, U_D)/B_L,$$
$$\max(1, B_L/T_U) \le B/T \le B_U/T_L.$$

All divisions are rational and rounded outward. Table 15 gives the isolated losses for all twelve saved two-hidden-layer ReLU models. For the three moons models, D/R and R/B are below 1.000001; the positive MNIST storage losses demonstrate why this cannot be assumed for other models.

Table 15: Complete exact audits of all twelve saved two-hidden-layer ReLU models. Regions are full-dimensional activation cells; endpoints are rounded outward. The region covers, all slope products, primal witnesses and available covariance-flow lower bounds are replayed.

Dataset	Seed	Regions	Abstraction B/T	Storage R/B
Moons	0	205	$[1.400531, 1.400532]$	$[1.000000, 1.000001]$
Moons	1	161	$[1.721271, 1.721272]$	$[1.000000, 1.000001]$
Moons	2	116	$[1.003706, 1.003707]$	$[1.000000, 1.000001]$
Iris	0	1888	$[1.000000, 1.000001]$	$[1.000000, 1.000001]$
Iris	1	1941	$[1.001263, 1.001264]$	$[1.000000, 1.000001]$
Iris	2	1570	$[1.002626, 1.002627]$	$[1.000000, 1.000001]$
MNIST	0	12658	$[1.000000, 1.000001]$	$[1.002217, 1.002221]$
MNIST	1	10027	$[1.000000, 1.000001]$	$[1.002792, 1.002795]$
MNIST	2	7475	$[1.042365, 1.042366]$	$[1.000000, 1.000001]$
Fashion	0	10549	$[1.000000, 1.000001]$	$[1.000000, 1.000001]$
Fashion	1	9122	$[1.000000, 1.000001]$	$[1.000000, 1.000001]$
Fashion	2	9643	$[1.000667, 1.000668]$	$[1.000000, 1.000001]$

Reference lower bounds from covariance flow. For $A_{0,k} = K_0 W_0$, $A_{1,k} = K_1 W_1$, choose PSD covariances $Y_{0,k}, Y_{1,k}$ by exact factorization $LL^\top$. Put $t = \sum_k \operatorname{tr} Y_{0,k} > 0$, $S_1 = \sum_k A_{0,k} Y_{0,k} A_{0,k}^\top$, $T_1 = \sum_k Y_{1,k}$ and $S_2 = \sum_k A_{1,k} Y_{1,k} A_{1,k}^\top$. If $S_1 \succeq T_1$, every feasible vertex certificate gives

$$gt \geq \operatorname{tr}(P_1 S_1) \geq \operatorname{tr}(P_1 T_1) \geq \operatorname{tr}(P_2 S_2) \geq \operatorname{tr}(W_2 S_2 W_2^\top).$$

Thus the downward square root of the last trace divided by t bounds R below. Numerical duals propose factors; PSD factors, flow slack and any repair covariance are checked exactly, including the repair's input-trace cost. Mapped factors are evaluated in the original input coordinates; numerical isometry is not assumed.

Tight price bounds for all E2 models. For every saved network, an exactly verified diagonal primal gives $g_U \geq D^2$. Any checked binary slope product J and nonzero input ray v give $q = \|Jv\|^2/\|v\|^2 \leq B^2 \leq R^2$. Hence $D/R \leq \sqrt{g_U/q}$. Neither global maximization of the sampled pattern nor its nonlinear realizability is required. All 72 checked upper endpoints are below 1.05, including tanh relative to its independent $[0, 1]$ slope abstraction.

To certify a lower endpoint for the maximum price, use a PSD moment matrix in (x, f) coordinates for the one-hidden-layer moons tanh seed 2. Write $X = \mathbb{E}[xx^\top]$, $F = \mathbb{E}[ff^\top]$ and $q_j = \mathbb{E}[f_j (W_0 x)_j - f_j^2] \geq 0$. Taking the trace of any diagonal dissipation inequality against this moment matrix proves $g \operatorname{tr} X \geq \operatorname{tr}(W_1 F W_1^\top)$. Exact factor and sector checks therefore give a lower bound for D. Independently enumerating all one-layer slopes bounds R above. Together with all 72 upper bounds, this encloses the actual maximum class price in $[1.049355, 1.049357]$ (endpoints rounded outward). Numerical dual stationarity is not a premise. A separate rational implementation rechecks all 72 primal/ray upper bounds and the lower-endpoint moment/vertex witnesses using exact LDL pivots, without importing the certificate-generation or original verification modules. This additional implementation covers the E2 enclosure, not the E3 radius chains.

A reconstructed E1 example. The retained peak seed is Gaussian $(n, m) = (4, 2)$, seed 6. Its original matrices were not archived; the new packet stores an explicit reconstruction and new primal and lower-bound witnesses. In coordinates (x, f, w), let $A_e = [A\ B\ E]$, $X_e = [I\ 0\ 0]$, $C_e = [C_y\ D_y\ D_w]$. For PSD Z, require $S = A_e Z A_e^\top - X_e Z X_e^\top \succeq 0$, $t = \mathbb{E}\|w\|^2 > 0$, and sector moments $\mathbb{E}[(f_j - \alpha_j \sigma_j)(\beta_j \sigma_j - f_j)] \geq 0$. Taking traces of the diagonal dissipation inequality gives $gt \geq \operatorname{tr}(C_e Z C_e^\top)$, since $P \succeq 0$ and diagonal weights are nonnegative. For the vertex reference, use sums of moments on exact graphs $f = KC^\top x$; the vertex inequalities imply their interior-slope counterparts by matrix convexity. The same flow argument yields a reference lower bound. Exact checks include all repair terms and their costs. The resulting gain-class price lies in $[1.764539, 1.764543]$. This certifies the saved reconstruction, not historical bitwise identity, the maximum across all 900 retained instances, or the sector-maximization ratios.

H.2 COMPLETE GAIN-TO-RADIUS VERIFICATION ON ALL THREE DATASETS

All finite stored binary64 parameters and inputs are interpreted as exact dyadic rationals. Matrix products, sums, ReLU signs and squared norms use arbitrary-precision integers; divisions use rational arithmetic. An outward float is accepted only after comparison with the exact rational target. Square-root candidates are scaled by even powers of two before conversion and accepted only after exact squared comparisons. Subnormals are handled by these same comparisons; nonfinite values, unrepresentable finite upper bounds or resource failure cause failure, not acceptance. This removes the legacy squared-norm underflow error (a scalar residual of 10^{-200} could previously be underestimated).

For positive α, exact verification of $q^2 I - M^\top M \succ 0$, $M = I - \alpha(I - W)$, with $0 \le q < 1$, proves that $\mathcal{T}(z) = \mathrm{ReLU}(z - \alpha(z - Wz - Ux - b))$ is a contraction. Its fixed points coincide with those of the ReLU DEQ by componentwise complementarity. At any finite candidate $\hat{z}$, compute the residual exactly and bound $r \ge \|\mathcal{T}(\hat{z}) - \hat{z}\|_2$ outward. Then $E_z \ge r/(1-q)$, using exact subtraction and division, bounds $\|\hat{z} - z^*\|_2$. For arbitrary finite proposed logits $\hat{F}$, exact evaluation of $V\hat{z} + c$ provides $e_i \ge |\hat{F}_i - (V\hat{z} + c)_i|$. With $N_{yj} \ge \|V_y - V_j\|_2$ checked outward, a downward evaluation gives

$$m_{yj} \le \hat{F}_y - \hat{F}_j - e_y - e_j - N_{yj} E_z \le F_y(x) - F_j(x).$$

Given a sound positive global gain Γ_{yj} for the exact output difference, a downward rational division yields $r_{yj} \le \max(m_{yj}, 0)/\Gamma_{yj}$. A radius $\epsilon < \min_{j \ne y} r_{yj}$ preserves strict positive margins; equality is not certified. Missing infinite gains yield zero. This certifies the real fixed-point classifier at the stored input, not a deployed floating-point prediction routine.

The MNIST audit archives and independently replays $1{,}104$ gain witnesses: six models, each with 45 class pairs and one all-logit target at four SDP levels. It checks exact LV subtraction, input elimination, every block QC, the full gain inequality, model hashes, partitions and complete target coverage. Historical multiplier arrays were not archived; these are new replacement witnesses, not recovered solver outputs. Pair and quadruple partitions remain the independently formed historical partitions. Some singleton proposals embed a diagonal QC with a small strict perturbation; the entire resulting inequality is reverified rather than assuming the perturbation is safe.

Using these new gains, the replay checks contraction anew and all 10,000 inputs for each saved MNIST model. All 72 integer accuracy counts at the three exact decimal thresholds match the retained results. The printed SDP gain and average-radius summaries also match, although the underlying new gain values are not byte-identical. A second CLI pass reads the complete gain and radius packets and rechecks every stage without fixed-point iteration or LP/SDP solves. Historical gains are comparison data only. The same exact-checker implementation is used in both passes; this is not a separately implemented verifier or formal verification of its runtime. For Fashion-MNIST and KMNIST, twelve replacement trainings supply saved weights, $2{,}208$ new gain witnesses and complete radius packets for all 10,000 test inputs per model. Separate disk replay verifies all gains and radius chains. All 144 threshold counts and twelve clean accuracies match the historical records; gain values can differ, and historical weight identity is not claimed. New PGD diagnostics use these replacement weights and are checked against their radius packets with no observed violation. The original MNIST PGD diagnostics are retained rather than rerun. Norm baselines, sampled lower bounds and PGD remain numerical. Separate static-class audits below cover all three datasets.

Static fixed-point optimum ratios. For a checked binary K, put $H = I - KW$. Exactly verifying $H^\top H - \eta^2 I \succ 0$ proves invertibility and $\|H^{-1}\| < 1/\eta$. For a nonzero input ray x and arbitrary candidate $\hat{z}$, compute $r = KUx - H\hat{z}$ exactly. Then $z = H^{-1}KUx$ satisfies

$$\|LVz\| \ge \|LV\hat{z}\| - \|LV\|_F \|r\|/\eta.$$

Outward norm bounds and rational division by $\|x\|$ yield a positive response lower bound ℓ. Every static certificate valid on the independent slope box bounds this response, so $F \ge \ell$ for the static one-block gain infimum F. Together with a new verified diagonal upper bound U_D, this proves $1 \le D/F \le U_D/\ell$. All 276 MNIST targets have a checked upper endpoint below 1.106. This does not assert nonlinear realizability, identify F with a dynamic reference, or certify the numerical lower endpoints obtained by dividing two primal feasible gains.

To obtain a certified positive lower price, put $G = UU^\top \succ 0$ and use a PSD moment matrix Z in (z, v) coordinates, where $v = Ux$. If $q_j = \mathbb{E}[z_j((Wz + v)_j - z_j)] \ge 0$, tracing any diagonal

static certificate gives $g\,\mathrm{tr}(G^{-1}Z_{vv}) \geq \mathrm{tr}(LVZ_{zz}V^\top L^\top)$. An exactly checked symmetric R_+ with $GR_+G - G \succ 0$ bounds G^{-1} above. Thus the downward square root of

$$\frac{\mathrm{tr}(LVZ_{zz}V^\top L^\top)}{\mathrm{tr}(R_+ Z_{vv})}$$

bounds D below when both traces are positive. We store $Z = AA^\top + H\,\mathrm{diag}(\delta)H^\top$, $H = [I; 2I - W]$, $\delta \geq 0$: each repair column adds one unit to its own sector moment and zero to the others. All products, sector checks and repair costs are evaluated exactly. Dividing this lower bound by a verified quadruple-block primal upper bound gives a valid lower bound on D/F. For the saved monotone MNIST seed 0, the resulting static class-price intervals are $[1.026324, 1.105171]$ for target 0–7 and $[1.033509, 1.070186]$ for target 5–9. Together with the all-target upper bounds, the initial maximum-price enclosure across the 276 saved MNIST targets is $[1.033509, 1.106]$. These intervals need not be tight and do not identify the dynamic storage reference.

Targeted response refinement. Further slope searches improve eleven exactly checked response lower bounds among twelve targets selected by their largest initial upper endpoints. For an old bound $U \geq U_D/\ell$ based on the diagonal primal U_D, increasing the response lower bound to ℓ' gives $\min\{U, U\ell/\ell'\}$, rounded outward. The original diagonal primal remains valid. Replaying these new witnesses and retaining every other target tightens the MNIST maximum-price upper endpoint to 1.100726; this changes only upper endpoints. Table 16 also includes the later block-eight lower refinement.

Extension to the replacement datasets. The same exact slope-response construction verifies lower bounds for all 552 Fashion-MNIST/KMNIST targets; dividing their verified diagonal upper gains by these lower bounds gives static class-price upper bounds. Exact diagonal moment witnesses additionally prove positive lower prices for Fashion-MNIST monotone seed 1, class pair 7–9, and KMNIST contractive seed 1, pair 5–6. Dividing each diagonal lower bound by a verified block upper gain bounds D/F below. Separate disk replay checks model identity, all primal inequalities and all new lower witnesses. Table 16 encloses the maximum over each dataset's 276 targets (six models, 46 targets each); together the 828-target maximum lies in $[1.067848, 1.122833]$. These statements concern static certificate-class infima over independent slopes, not actual nonlinear Lipschitz constants or dynamic storage references.

Table 16: Certified intervals for the maximum static diagonal/one-block class price per dataset; endpoints rounded outward.

Dataset	Targets	Maximum class-price interval
MNIST	276	$[1.033509, 1.100726]$
Fashion-MNIST	276	$[1.030148, 1.085279]$
KMNIST	276	$[1.067848, 1.122833]$

Paired seed-level benefits and small gains. Table 17 uses the new verified MNIST gain packets and released Fashion/KMNIST records. For each model we compute $100(1 - \gamma_{\mathrm{quads}}/\gamma_{\mathrm{diag}})$ over all 46 targets, then take its median. The table shows the range of these medians over the three seeds, not a confidence interval. Accuracy differences use strict radius-threshold counts on the same 10,000 test inputs. All 18 models improve certified accuracy at radius 0.3, but 105 of 828 targets have gain reductions at most 0.5%. One Fashion contractive seed-0 target has a slightly larger returned quadruple gain (less than 0.000025% excess); this is a feasible-solution comparison, not a reversal of class inclusion. The largest target-level returned-gain reduction, 4.14%, occurs in the KMNIST contractive seed-1 model. These results support modest, consistent model-level benefits without implying that every target benefits.

Table 17: Paired E3 benefits: minimum–maximum over three seeds per row. Gain reductions compare returned verified gains; accuracy gains are percentage points at radius 0.3. These are not class-optimum gaps.

data	variant	median gain reduction (%)	accuracy gain (points)
MNIST	contractive	0.74–1.44	0.64–1.13
MNIST	monotone	1.25–1.64	1.09–1.23
Fashion	contractive	0.51–0.62	0.40–0.46
Fashion	monotone	0.86–1.06	0.62–0.69
KMNIST	contractive	0.90–1.44	0.48–0.77
KMNIST	monotone	1.20–1.36	0.52–0.64

Additional width check. For two saved width-eight MNIST models (one contractive and one monotone, seed zero), we generate and replay 368 new gain certificates and 92 slope-response lower bounds. The resulting upper bounds on the maximum static diagonal/full-block price over each model's 46 targets are 1.017579 (contractive) and 1.012091 (monotone), rounded outward. These extend the checked widths but do not isolate a causal width effect: the trained weights differ, there is only one seed per variant, and upper bounds alone do not order the true prices. For six width-80 MNIST models (seeds zero, one and two for each parameterization), normalized SDP proposals supply checked diagonal and quadruple-class gains for all 46 targets per model. Seed zero uses the saved checkpoints; four additional models use the same ten-epoch training recipe, without seed selection. A separate integer implementation verifies all 552 gains and 60,000 test-input radius chains (1,080,000 label/level inequalities). Table 18 gives full-test accuracies. Quadruple-class gains include the better diagonal baseline; failed proposals and diagonal-embedding fallbacks are retained in the released records. The diagonal baseline is selected for 20, 21 and 19 contractive targets and 2, 4 and 3 monotone targets (seeds zero, one, two; 46 targets each). At radius 0.3, all three seeds improve by 0.11–0.22 percentage points (contractive) and 0.76–1.16 (monotone). These are feasible-certificate comparisons, not class-optimum gaps. The six models are not pooled into the original 828-target maximum. Three seeds at one additional width do not establish larger-vision or different-architecture-family generality, nor a causal effect of width.

Table 18: Additional width-80 MNIST models, three seeds per parameterization. All-logit gains are rounded upward; accuracies (percent) use all 10,000 test inputs. Quadruple gains include the diagonal baseline.

Model/seed	Class	Gain	$\epsilon = 0.1$	$\epsilon = 0.3$	$\epsilon = 0.5$
Contractive / 0	Diagonal	47.742	87.35	19.34	0.39
	Quadruple	47.295	87.36	19.45	0.41
Contractive / 1	Diagonal	55.183	86.26	14.67	0.35
	Quadruple	55.183	86.31	14.83	0.36
Contractive / 2	Diagonal	52.998	87.34	13.11	0.08
	Quadruple	52.467	87.40	13.33	0.08
Monotone / 0	Diagonal	29.106	91.69	53.45	10.09
	Quadruple	28.821	91.76	54.28	10.70
Monotone / 1	Diagonal	31.054	90.76	45.60	7.19
	Quadruple	30.695	90.83	46.36	7.46
Monotone / 2	Diagonal	29.960	91.89	49.13	8.16
	Quadruple	29.631	92.01	50.29	8.77

Controlled wall-budget comparison. On the two saved width-80 seed-zero models, we compare a diagonal-only policy with a policy that starts from the same spectral diagonal construction and then proposes a quadruple certificate. There is one timed run per target/policy pair. Each policy independently pays for its baseline; the latter does not receive the former's optimized certificate. All 46 targets use the same 60-second deadline, with incumbents recorded at 15 and 30 seconds along that trajectory. One normalized SCS proposal has a 40-second solve cap; model setup, proposal and exact acceptance count toward the deadline, while interpreter imports, common fixed-point preparation and subsequent independent replay do not. An unfinished worker is stopped at the deadline. Policies alternate order; one numerical worker runs at a time on an Apple M1 Pro (eight cores,

16 GiB). Accelerate's single-thread BLAS/LAPACK mode is set and checked on that worker's main thread; other host activity is not isolated. Table 19 reports complete-test accuracies from separately verified accepted gains and radius chains. A missing required pair gain gives radius zero. These are comparisons of two specified solver policies, not optimal allocation of compute or class-optimum estimates. Unlike Table 18, the quadruple policy does not freely inherit the separately optimized diagonal gain. At 60 seconds and radius 0.3, diagonal versus quadruple accuracy is 19.34% versus 10.09% (contractive) and 53.45% versus 53.15% (monotone). At radius 0.5, the monotone quadruple policy instead gives 10.47% versus 10.09%. Median completed target times are 18.93/44.65 seconds (contractive diagonal/quadruple) and 9.70/44.45 seconds (monotone). There are no missing certificates or deadline expirations; the quadruple policy retains only its spectral baseline on 12 contractive and one monotone target. Thus class inclusion and the seed-table improvements do not establish a uniform advantage under equal wall budgets.

Table 19: Equal per-target wall budgets on saved width-80 seed-zero models. Accuracies are percentages of 10,000 test inputs. Base counts targets whose best accepted witness is the spectral diagonal baseline; Missing counts targets with no accepted witness (out of 46, including the all-logit target).

Model	Policy	Budget (s)	$\epsilon = 0.1$	0.3	0.5	Base	Missing
Contractive	Diagonal	15	77.35	1.43	0.00	39	0
Contractive	Diagonal	30	87.35	19.34	0.39	0	0
Contractive	Diagonal	60	87.35	19.34	0.39	0	0
Contractive	Quadruple	15	75.72	1.01	0.00	46	0
Contractive	Quadruple	30	75.72	1.01	0.00	46	0
Contractive	Quadruple	60	83.59	10.09	0.12	12	0
Monotone	Diagonal	15	91.49	50.08	6.53	7	0
Monotone	Diagonal	30	91.69	53.45	10.09	0	0
Monotone	Diagonal	60	91.69	53.45	10.09	0	0
Monotone	Quadruple	15	89.35	30.14	1.92	46	0
Monotone	Quadruple	30	89.35	30.14	1.92	46	0
Monotone	Quadruple	60	91.67	53.15	10.47	1	0

Exploratory diagonal-first follow-up. After inspecting those models, we froze a second policy before any new timed run: compute its own spectral baseline and optimized diagonal certificate, retain the best accepted witness, then attempt one quadruple solve if more than six seconds remain. Both policies pay for fresh certificates; no separately computed diagonal result is supplied for free. The same 60-second target deadline, 15/30-second prefixes, hardware and thread controls apply. Each solve is capped at the smaller of 40 seconds and remaining time minus five seconds. Parent receipt times after durable packet publication determine eligibility. Model setup and exact acceptance are charged; imports, common fixed-point preparation and independent replay are excluded. Runs are serial with alternating policy order; other host activity is not isolated. This is an exploratory follow-up on already inspected models.

Table 20 reports all prefixes. At 60 seconds and $\epsilon = 0.3$, diagonal versus diagonal-first accuracy is 19.34% versus 19.55% (contractive) and 53.45% versus 54.29% (monotone): gains of 0.21 and 0.84 percentage points. The final best witnesses are quadruple on 33/42 targets and optimized diagonal on 13/4 targets, respectively. Two monotone runs reach the deadline and retain their diagonal witnesses; no target lacks a certificate. At 30 seconds the policies tie, while at 15 seconds monotone diagonal-first is worse (48.55% versus 49.72% at $\epsilon = 0.3$). Thus the observed benefit is specific to the tested 60-second policy and models, not uniform across budgets. Independently replaying all 259 distinct accepted gain witnesses covers 443 accepted events in 184 runs; separate radius replay checks 20,000 model/input records, 1,080,000 label/policy/budget inequalities and 36 strict threshold counts. The same 10,000 test images are evaluated by each model. These finite feasible outputs need not equal those in Table 18, and neither establish class optima nor optimal compute allocation.

Table 20: Exploratory diagonal-first follow-up with equal per-target deadlines. Accuracies are percentages of 10,000 test inputs per model. Base and Missing have the definitions in Table 19. Prefixes are observations along each 60-second run, not separately retuned runs.

Model	Policy	Budget (s)	$\epsilon = 0.1$	0.3	0.5	Base	Missing
Contractive	Diagonal	15	78.13	1.56	0.00	36	0
Contractive	Diagonal	30	87.35	19.34	0.39	0	0
Contractive	Diagonal	60	87.35	19.34	0.39	0	0
Contractive	Diagonal-first	15	78.13	1.56	0.00	36	0
Contractive	Diagonal-first	30	87.35	19.34	0.39	0	0
Contractive	Diagonal-first	60	87.38	19.55	0.42	0	0
Monotone	Diagonal	15	91.45	49.72	6.53	9	0
Monotone	Diagonal	30	91.69	53.45	10.09	0	0
Monotone	Diagonal	60	91.69	53.45	10.09	0	0
Monotone	Diagonal-first	15	91.36	48.55	6.46	8	0
Monotone	Diagonal-first	30	91.69	53.45	10.09	0	0
Monotone	Diagonal-first	60	91.75	54.29	10.68	0	0

Exploratory explanation and held-out check. On the six width-40 seed-zero models, we select among three coupling statistics by leave-one-model-out prediction error, retaining a parameterization intercept. The selected within-block coupling fraction fails to improve prediction of median returned-gain reductions on the twelve held-out seed-1/2 models: mean absolute error is 0.3183 versus 0.2835 percentage points for the parameterization-specific training-mean baseline. We therefore do not claim that this statistic explains when blocks help; this is a descriptive check within the same datasets, not a causal or architecture-general result. Without retuning, the same predictor also fails on the six width-80 models: mean absolute error is 2.3952 versus 0.5711 percentage points for the frozen baseline. This additional extrapolation check does not supply a validated explanation of block benefits.

A stronger lower endpoint for the E3 maximum. For the all-logit KMNIST contractive seed-1 target, an exact dual witness bounds the diagonal infimum below by 29.248719, while a verified block-eight primal bounds the full-block infimum above by 27.390322 (values rounded outward here). Their unrounded certified ratio exceeds 1.067848, raising the lower endpoint of the 828-target maximum to that value. A separate integer checker verifies the dual moments, input metrics, all block graph vertices and dissipation inequality. The near-matching diagonal primal shows that the remaining uncertainty is predominantly in the full-block reference. Additional moment relaxations and two alternative block-eight partitions do not improve the retained bounds. This is a certificate-class result, not a nonlinear Lipschitz ratio.

Direct primal refinement of the E3 bottleneck. A separate tighter-tolerance diagonal SDP solve for the all-logit KMNIST contractive seed-1 target, followed by exact feasibility verification, reduces its price upper bound from approximately 1.122832886 to 1.122832307. Thus the six-decimal overall upper endpoint 1.122833 is unchanged. This small improvement does not resolve the remaining class-price uncertainty or prove response-search optimality.

Separate implementation checks. An additional checker imports neither the certificate generator nor its original verification modules. It independently assembles all 3,312 E3 gain inequalities using integer dyadic arithmetic, checks the multiplier graph vertices and input elimination, and verifies positive definiteness by exact strict diagonal dominance after a proposed congruence. Numerical Cholesky only proposes that congruence; acceptance uses exact integer inequalities. All gains pass. A separate exact-residual and logit-margin calculation checks all 180,000 test inputs, every competing label and all four multiplier levels: 6,480,000 margin inequalities and all 216 reported strict radius-threshold counts pass. Thus both implementations check the complete saved gain-to-radius chain. This does not formally verify the Python runtime or floating-point deployment of the classifier.