ON THE COST OF CERTIFYING NEURAL LYAPUNOV FUNCTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

A neural Lyapunov function certifies stability only where it is formally verified, and branch-and-bound verification can dominate the cost of current pipelines, varying by orders of magnitude between certificates of the same system. We show that this cost has two parts, both governed by the *certified radius*: the size of the largest box around a point that the verifier's bounding method can certify. The *frontier* part is geometric. Where the maximal level set touches the set on which the specification fails, the radius collapses: any verifier that checks the clauses separately on axis-aligned boxes needs $\Omega(\delta^{-(n-1-k)/2})$ boxes at slack δ, whatever its bounding method, with n the dimension and k the number of faces or kinks at the contact (for a nondegenerate contact and V not flat along any coordinate). The *bulk* part, paid everywhere, scales with the integral of the inverse n-th power of the radius, which a few thousand sampled points estimate without running branch-and-bound. On 25 certificates in two to six dimensions, the predicted exponents match at 22 contacts, and the estimate stays within a single-digit factor of the measured per-clause cost across five orders of magnitude. The decomposition turns the choice of the certified level into a computation, enlarging the certified region of a released quadrotor certificate 1.8-fold, and shows that certified training moves the frontier but not consistently the bulk.

1 INTRODUCTION

Neural Lyapunov certificates, learned jointly with a controller, have progressed from small examples to output-feedback systems with observers and to quadrotors (Chang et al., 2019; Dai et al., 2021; Wu et al., 2023; Yang et al., 2024). What makes such a certificate trustworthy is the formal check that follows training: a verifier proves that V decreases along the closed loop on a sublevel set $\Omega_c = \{x \in B : V(x) \le c\}$ and that trajectories stay in the box B, after which Ω_c is a certified region of attraction. In current pipelines this check is a branch-and-bound (BaB) search over the state space that bounds the network with linear relaxations (Zhang et al., 2018; Xu et al., 2021; Wang et al., 2021). It is the expensive step, taking hours for a six-dimensional system in Yang et al. (2024), and it has motivated both training methods that make certificates easier to verify (Xiao et al., 2019; Zhang et al., 2024; Shi et al., 2026) and verifiers that prune more aggressively (Zhou et al., 2025).

Two practical questions arise, and neither has a quantitative answer. The first concerns the level. Any c below the *maximal certifiable level* $c^\star$ is sound, and a larger c certifies a larger region; the released certificates we examine are verified 0.03% to 24% below $c^\star$. How fast does the cost grow as c approaches $c^\star$, and is the last part of the region worth its price? The second concerns the certificate: verification cost differs by orders of magnitude between certificates of one system. Which property of a certificate sets this cost, can it be measured before verifying, and does training change it?

Our answers rest on one quantity, the *certified radius*: the half side of the largest box around a point that the verifier's bounding method can certify. Up to constants, the number of boxes in a BaB proof is the integral of the inverse n-th power of this radius, and the integral splits into two parts with different causes. Where the level set $\{V = c^\star\}$ touches the violation set, the radius collapses for a geometric reason that no bounding method can repair: for a verifier that certifies the clauses separately, a box that meets the violation set can be accepted only by showing $V > c$ on all of it, so it must be short, relative to the slack $\delta = c^\star - c$, in every direction in which V changes. Counting such boxes gives a lower bound for every such verifier, with an exponent set by the shape of the contact.

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.

Away from the contact, the radius is a property of the certificate and the bounding method, and since it is defined point by point it can be sampled: its values at a few thousand points predict what a full BaB run will cost (Figure 1). We make four contributions.

- **A frontier law for the certified level** (Section 4.1). A verifier that certifies the clauses separately on boxes needs at least $\mathrm{vol}(S_\delta)\prod_i r_i/(2\delta)^n$ boxes at slack δ, with S_δ the δ-cap of the violation set and r_i descent rates of V (Theorem 1); at a regular contact this is $\Omega(\delta^{-(n-1-k)/2})$ (Proposition 2), a rate that contracted isotropic verifiers attain. Measured cap exponents match it at 22 of 25 contacts.
- **A cost estimator that does not run branch-and-bound** (Section 4.2). For contracted verifiers with monotone bounders, proof size is bounded above and below by multiples of the integral of the inverse n-th power of the certified radius (Theorem 4). Sampling the radius of the verifier's own bounder gives an estimate whose ratio to the measured bulk cost varies only 6.5-fold on 24 certificates spanning a $2 \cdot 10^5$-fold range of costs.
- **One-shot certification at the maximal level** (Section 5), which certifies a $1.8\times$ larger region of a released quadrotor certificate in 164 s on one GPU and a valid level for certified-trained ones.
- **What certified training changes** (Section 6.5). With paired seeds of CT-BaB (Shi et al., 2026) and a sampled-box variant, the training-time verified fraction does not predict cold certifiability, and training moves the contact, and with it the frontier of per-clause verifiers, but not consistently the bulk cost per certified volume.

2 Related work

Learning neural Lyapunov certificates. Counterexample-guided training with SMT or MIP verification established the pipeline (Chang et al., 2019; Dai et al., 2021; Edwards et al., 2024; Wu et al., 2023); Yang et al. (2024) scaled it to output feedback with α, β-CROWN and released certificates and verifier configurations, which we use unmodified. Sample-based certificates transfer a finite check to the region through Lipschitz constants (Boffi et al., 2020; Berkenkamp et al., 2017); our counting result covers such verifiers, since it only asks the bounder to be sound and monotone. Zubov-based continuous-time synthesis (Li et al., 2025) pushes the certified level toward its maximum by shrinking thresholds on counterexamples, for a different certificate class.

Training for cheaper verification. Inducing stable ReLUs (Xiao et al., 2019), synthesis with exact verification in the loop (Zhang et al., 2024), interval-bound certified training of classifiers (Gowal et al., 2018; Jovanović et al., 2022; Mao et al., 2024) and certified training with branch-and-bound for Lyapunov control (Shi et al., 2026, CT-BaB) optimize a bound rather than a sample loss. None identifies the property of a certificate that sets verification cost, or measures it under cold verification.

Verification cost. Continuous amortization bounds subdivision counts by an integral (Burr, 2016); two-sided characterizations exist for zeroth-order level-set estimation (Bachoc et al., 2021) and for subdivision in real algebraic geometry (Cucker et al., 2022). In neural network verification, branching heuristics are learned or engineered (Bunel et al., 2020; Lu & Kumar, 2020), preimage bounds handle control specifications (Kotha et al., 2023), verification can run in a latent space (Koller et al., 2026), hyperplane partitioning targets Lyapunov verification (Wayment et al., 2026), and Clip-and-Verify (Zhou et al., 2025) bounds one clause over the region cut out by the linear bound of another, the conjunction-aware mode we measure. Missing so far is a statement of how verification cost scales with the certified level and with the certificate, and a way to measure it before verifying.

3 Setting

States $x \in \mathbb{R}^n$ evolve as $x_{t+1} = F(x_t)$, with the learned controller absorbed into F. Verification takes place on a compact box $B \subset \mathbb{R}^n$ minus a small core box $B_0 \subset \operatorname{int} B$ around the equilibrium: $K = \overline{B \setminus B_0}$ is the *verification region*. A certificate is a continuous $V \geq 0$ with $V(0) = 0$ and $V > 0$ on K, with sublevel sets $\Omega_c = \{x \in B : V(x) \leq c\}$. Following Yang et al. (2024), the certificate is valid at level c if, for every $x \in K$,

$$\Pi_c(x): \qquad V(x) > c \quad \text{or} \quad \big(V(F(x)) \leq \lambda V(x) \ \text{ and } \ F(x) \in \operatorname{int} B\big), \tag{1}$$

with decay factor $\lambda \in (0,1)$: inside Ω_c, V must decrease and $F(x)$ stay in B. Let $D_q \subseteq K$ be the set where clause q fails (the decrease clause or a face of the containment clause) and $D = \bigcup_q D_q$ the *violation set*. Then Π_c holds on K exactly when $V > c$ on $K \cap D$, so every level below

$$c^\star = \inf_{x \in K \cap D} V(x) \tag{2}$$

is valid and no level above it is. We call $c^\star$ the *maximal certifiable level*; it is attained on the closure $\bar{D}$ at a *contact point* $x^\star$. We verify at $c = c^\star - \delta$ with *slack* $\delta > 0$, reported relative to $c^\star$ in the experiments. The violation set just above $c^\star$ is measured by the *cap* $S_\eta = \{x \in K \cap D : V(x) \le c^\star + \eta\}$ and the *cap exponent* $\gamma = \lim_{\eta \to 0} \log \mathrm{vol}\, S_\eta / \log \eta$, if the limit exists. If Π_c holds on K and the core is handled separately, Ω_c is forward invariant and every trajectory from it reaches B_0 in finitely many steps (Proposition 5 in Appendix B.1); this is what we mean by a certified region.

Verifiers. A box BaB verifier bounds the specification on each axis-aligned box with a *bounder* (CROWN and interval bounds in α, β-CROWN, or Lipschitz and Taylor bounds), accepts the box when the bounds prove Π_c on all of it, and otherwise splits it along one coordinate. Its cost N is the number of boxes bounded, which the verifier calls domains. Two acceptance rules matter. A *per-clause* verifier accepts a box for clause q when it proves, on the whole box, either clause q or $V > c$, so a box that meets D_q can be accepted only by proving $V > c$ on it. The released configuration of α, β-CROWN works this way (mode sb, splitting the coordinate with the largest bound sensitivity), as does the same verifier with longest-edge splitting (mode naive) and any interval, Lipschitz, Taylor or exact-range bounder. A *conjunction-aware* verifier may also accept a box after proving jointly that $K \cap D \cap \{V \le c\}$ misses it; Clip-and-Verify with complete clipping (mode clip) bounds one clause over the part of the box cut out by the linear bound of another.

Assumptions. Each result states which of the following it uses.

(A1) *Setting.* F is continuous on B and V on $\mathbb{R}^n$, K does not depend on the certificate, and D is not empty.

(A2) *Regular contact* (for the exponents). $V = c^\star$ at no other point of the closure $\bar{D}$ of D, only one clause has $x^\star$ in the closure of its violation set, and on each smooth piece of V and the margins near $x^\star$ (one piece if there are no kinks), V and the margins are C^2 and $x^\star$ is a nondegenerate KKT point of $\min V$ over $\bar{D}$: $k+1$ constraints with linearly independent normals are active, with strictly positive multipliers (the boundary of that clause and k linear pieces: faces of B or B_0, or kinks of V or of a clause margin), and the second-order sufficient condition holds on the tangent space (Appendix B.3).

(A3) *Verifier class.* The bounder is sound; for the counting results it is also inclusion monotone (its bound on a box is at least its bound on any sub-box), and the verifier is κ-contracted for some $\kappa < 1$: it accepts a box of radius w exactly when the bounder certifies the concentric box of radius w/κ. The frontier lower bound needs only per-clause acceptance on axis-aligned boxes.

4 Where the cost comes from

Figure 1(a) shows the two parts of the cost. We give the frontier lower bound, which needs no smoothness and no property of the bounder, the exponent it implies at regular contacts, and the counting theorem that turns the certified radius into a cost (proofs in Appendix B).

4.1 The frontier is geometric

The argument is elementary. A box B' of a per-clause proof at level $c = c^\star - \delta$ that meets the cap S_δ meets the violation set, so it must have been accepted by proving $\inf_{B'} V > c$. Any two points of $S_\delta \cap B'$ have $V \le c^\star + \delta$, and moving from one of them along coordinate i, in the direction in which V decreases, to the i-th coordinate of the other stays in the box. If V decreases at rate at least r_i in that direction, the move is shorter than $2\delta/r_i$, or V would reach c. So every coordinate projection of $S_\delta \cap B'$ is shorter than $2\delta/r_i$, and covering S_δ takes at least $\mathrm{vol}(S_\delta) \prod_i r_i/(2\delta)^n$ boxes.

Theorem 1 (Frontier lower bound)**.** *Let $\delta > 0$, $c = c^\star - \delta$, and let G be a set that contains the smallest box containing S_δ. Suppose there are signs $\sigma \in \{\pm 1\}^n$ and rates $r \in (0,\infty)^n$ such that $V(y - \sigma_i t\, e_i) \le V(y) - r_i t$ whenever $t \ge 0$ and both y and $y - \sigma_i t\, e_i$ lie in G. Then every per-clause box certificate at level c contains at least $\mathrm{vol}(S_\delta) \prod_{i=1}^n r_i/(2\delta)^n$ boxes that meet S_δ.*

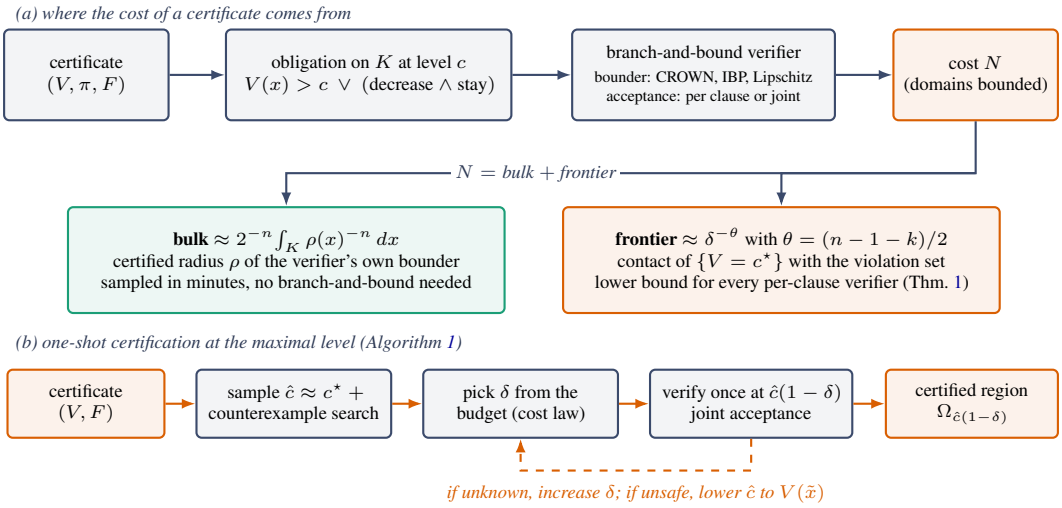


Figure 1: **Where the cost of certifying a neural Lyapunov function comes from.** (a) The cost of a branch-and-bound proof splits into a bulk term, the integral of the certified radius of the verifier's bounder, and a frontier term, set by the contact between $\{V = c^\star\}$ and the violation set. (b) The one-shot protocol certifies at $\hat{c}(1-\delta)$, with the slack δ chosen from the budget.

The bound holds for every bounder, including an exact range oracle, and for every splitting rule. At a C^1 contact where no partial derivative of V vanishes, the mean value theorem supplies the rates on a shrinking neighbourhood, and the bound becomes $N \gtrsim \mathrm{vol}(S_\delta)\, \delta^{-n}$ (Corollary 6). What remains is the volume of the cap, and this is where the shape of the contact enters.

Proposition 2 (Contact exponents)**.** *Under (A2) with k active linear pieces, $\mathrm{vol}\, S_\eta \asymp \eta^\gamma$ with $\gamma = (n+1+k)/2$. If, in addition, near $x^\star$ each partial derivative of V (one-sided at kinks) keeps a fixed sign and stays away from zero, every per-clause box verifier needs $N = \Omega(\delta^{-\theta})$ boxes at slack δ, where $\theta = n - \gamma = (n-1-k)/2$. A smooth interior contact ($k = 0$) gives $\theta = (n-1)/2$, a contact on a face of B gives $\theta = (n-2)/2$, and a vertex-type contact ($k = n-1$) gives $\theta = 0$.*

Under a transversality condition, a contracted isotropic verifier with Lipschitz bounds needs only $O(\log(1/\delta))$ boxes beyond a level-independent number N_{far} at a vertex-type contact and $N_{\mathrm{far}} + O(\delta^{-\theta})$ for $\theta > 0$, so the bound is attained (Theorem 10). Anisotropic splitting can do better only when the sign condition fails, typically because some $\partial_i V(x^\star)$ nearly vanishes when the binding clause barely depends on x_i (Remark 13), and Section 6.5 shows that training decides whether this happens. Conjunction-aware verifiers are outside the scope of Theorem 1: near a smooth interior contact, an idealized one with second-order affine minorants needs only $O(\log(1/\delta))$ boxes (Proposition 11, proof sketched), and Section 6.2 measures how far clip is from this ideal.

4.2 The bulk is the certified radius

The level-independent part of the cost, dominant in six dimensions, is set by the certified radius.

Definition 3 (Certified radius)**.** For an obligation $h(x) < 0$ and a bounder $U(x, w) \geq \sup_{\|z-x\|_\infty \leq w} h(z)$, the certified radius is $\rho(x) = \sup\{w \geq 0 : U(x, w) < 0\}$, with $\sup \emptyset = 0$. For the disjunction and the conjunction in (1), radii combine by $\max$ and $\min$.

For every sound and inclusion-monotone bounder, ρ is 1-Lipschitz in the ℓ_∞ norm (Theorem 7), and this makes it possible to count the boxes of a proof.

Theorem 4 (Proof size is the integral of the radius)**.** *Let the bounder be sound and inclusion monotone, with certified radius $\rho > 0$ on K. Let the verifier start from a cube $W \supseteq K$, split boxes dyadically, discard boxes that miss K, and accept a box of centre x and radius w exactly when $w \leq \kappa\rho(x)$, for*

Algorithm 1: One-shot certification at the maximal level

Input: certificate (V, F), region K, conjunction-aware verifier, domain budget M, slack δ_0
Output: a certified level c and region Ω_c

$\hat{c} \leftarrow$ minimum of V on $K \cap D$ from dense sampling, refinement and a gradient attack;
$\delta \leftarrow$ smallest slack on a log grid with predicted cost $N_\rho + A\delta^{-\theta} \le M$ (θ from the cap volumes at $\hat{c}$, A from one run at a larger slack; δ_0 if unavailable);
repeat
  run the verifier at $c = \hat{c}(1-\delta)$;
  if unsafe with counterexample $\tilde{x}$ **then** $\hat{c} \leftarrow \min(\hat{c}, V(\tilde{x}))$;
  if unknown (timeout) **then** $\delta \leftarrow 3\delta$;
until *the verifier returns* safe;
return c and Ω_c;

some $\kappa \in (0,1)$. Then its number of leaves N satisfies

$$a_{n,\kappa}\,\mathcal{P}(\rho) \;\le\; N \;\le\; \max\big\{1,\, A_{n,\kappa}\,\mathcal{P}(\rho)\big\}, \qquad \mathcal{P}(\rho) = \int_W \Big(\min_{y\in K}\big[\rho(y) + \|x-y\|_\infty\big]\Big)^{-n} dx,$$

with $a_{n,\kappa} = \frac{1}{2}(\bar{\kappa}/(56\kappa))^n$, $\bar{\kappa} = \min(\kappa, 1-\kappa)$, $A_{n,\kappa} = ((1+3\kappa)/\kappa)^n$; no dyadic proof with the same discharge and acceptance rules has fewer leaves.

The upper bound is an instance of continuous amortization (Burr, 2016); the lower bound is, to our knowledge, new, and it fails without contraction (Remark 9). Two consequences follow. First, tightness is amplified by the dimension: on K the integrand of $\mathcal{P}$ is ρ^{-n}, so enlarging the radius s-fold on a region divides its contribution by s^n (Proposition 15 links s to the relaxation error of a linear bounder). Second, cost can be estimated without running BaB, since ρ is defined pointwise: for each row of (1), bisection on the half width of the cube around x, clipped to B as the verifier's boxes are, gives $\rho_{\mathrm{row}}(x)$ for the verifier's own bounder. Disjunct q has radius $\rho_q = \max(\rho_V, \rho_{\mathrm{row}\,q})$, and

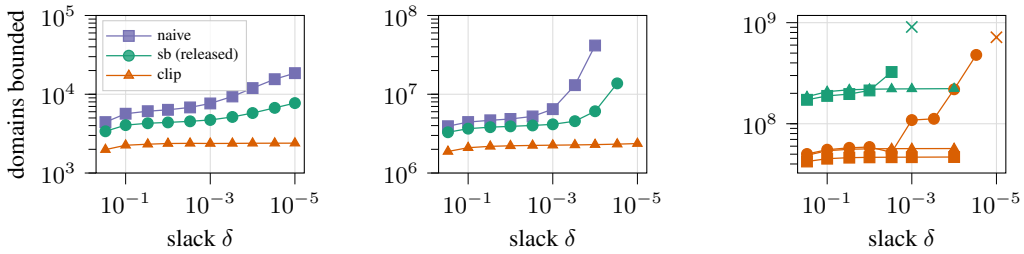
$$N_\rho \;=\; 2^{-n} \sum_q \int_K \min\{\rho_q(x), R\}^{-n}\, dx, \tag{3}$$

with R half the shortest side of B, is our estimate of the size of a per-clause proof with this bounder. Theorem 4 motivates it but does not cover it: the released verifier splits one coordinate at a time, is not contracted, and uses a bounder that is not monotone, so Section 6.3 tests the estimate empirically. The integrand is heavy-tailed near the equilibrium, where the decrease margin vanishes, so we estimate (3) by importance sampling, half uniform on K and half on logarithmic shells around the equilibrium (Appendix C.3); 8,000 points take 7 to 130 seconds on two CPU cores.

5 One-shot certification at the maximal level

The frontier law turns the choice of the level into a computation: verifying at the trained level leaves coverage unused, verifying at $c^\star$ is impossible because the frontier cost diverges as $c \to c^\star$, and verifying at $c^\star(1-\delta)$ costs a predictable amount once θ is known. Algorithm 1 puts this into practice. Its soundness does not depend on the estimate $\hat{c}$: the output is a level that the verifier has certified, and a wrong $\hat{c}$ produces an unsafe or unknown answer, never a false certificate. A counterexample $\tilde{x} \in K \cap D$ proves $c^\star \le V(\tilde{x})$, which the loop uses to tighten $\hat{c}$.

The first attempt succeeds whenever $\hat{c}$ is within δ of $c^\star$, which dense sampling achieves unless the violation set contains a thin sliver (Appendix D.7); hence the attack in step 1. The frontier cost grows as $\delta^{-\theta}$ while the coverage deficit is only of order δ (Appendix B.6), so the last fraction of a percent is expensive for per-clause verifiers at a smooth high-dimensional contact: on `pend-out-st`, isotropic splitting needs 6.5M domains at 0.1% below $c^\star$ and 41.6M at 0.01%.



(a) (b) (c)

Figure 2: **Cost of certifying $\Omega_{c^\star(1-\delta)}$ against the slack δ.** Domains bounded by α, β-CROWN in the modes naive, sb and clip; the slack decreases to the right. A cross marks a run stopped at its time limit without a verdict, drawn at the number of domains bounded until then, so the true cost is higher. (a) `pend-st`, a smooth contact in two dimensions. (b) `pend-out-st`, a smooth contact in four dimensions; sb at $\delta = 10^{-5}$ stalls at the float32 resolution of the bounds and is not shown (Appendix D.3). (c) Released quadrotor certificate (circles, clip) and CT-BaB certificates, seed 0, with the partition frozen after 5,000 steps (squares) or split to the end (triangles).

6 Experiments

6.1 Setup

Certificates. We use the released certificates of Yang et al. (2024): six with $n = 2$ (`pend`, `pend-quad`, `pend-st`, `path`, `path-quad`, `path-st`: pendulum and path tracking, neural or quadratic V, small-torque variants), three with output feedback and $n = 4$ (`pend-out`, `pend-out-nn`, `pend-out-st`) and a planar quadrotor with $n = 6$ (`quad`). We trained 17 more. With the released code of CT-BaB (Shi et al., 2026), which has no public checkpoints, we trained the pendulum, `pend-out` with three seeds, and the quadrotor with seeds 0 to 2 under two partition schedules plus one continued run; with a variant that trains on sampled boxes (Section 6.5), six `pend-out` certificates. Quadrotor seed 1 failed under both schedules ($c^\star$ of $1.9 \cdot 10^{-6}$ and $1.4 \cdot 10^{-7}$), leaving 15 trained certificates and 25 in total. We write c_0 for the released or trained level.

Verification. Every certificate is verified *cold*, from scratch, without a stored partition or an attack, using the released α, β-CROWN configurations in up to three modes: sb (released), naive (longest edge) and clip (Clip-and-Verify, complete clipping). Cost is the number of domains bounded, summed over the instances of a specification, on two CPU cores for the released two-dimensional certificates and one A100 otherwise, with per-instance timeouts of 300 to 5400 seconds. $c^\star$ is the minimum of V over $2 \cdot 10^6$ to $1.6 \cdot 10^7$ sampled points of $K \cap D$, refined in float64 and exact to within the finest slack that verified; cap exponents come from quasi-Monte Carlo cap volumes (Appendix C).

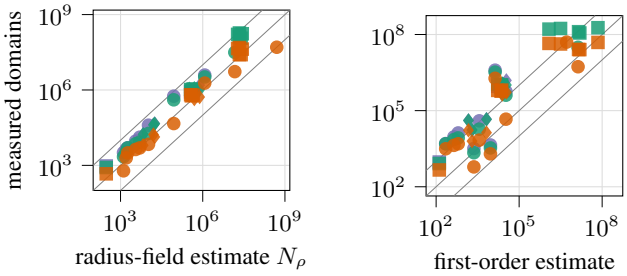
6.2 The frontier follows the contact geometry

Measured cap exponents match the predicted θ within 12% (within 0.05 where $\theta = 0$) at 22 of 25 contacts (Table 1), including a kink of V in all four coordinates (`pend-out-nn`, flat cost, with an uncertain cap fit) and the face and interior contacts of the certified-trained certificates. The exceptions are `pend-out` (a face with a kink of the decrease margin, 0.61, approaching 0.5 only over the finest levels), the frozen quadrotor at 15,000 steps (1.64, poorly sampled caps) and its continuation (1.53).

Per-clause verifiers pay the frontier at every contact with $\theta > 0$ and no degenerate axis (Figure 2). Isotropic splitting pays it even where an axis is nearly degenerate (Corollary 12): two `pend-out` certificates trained on sampled boxes bind at a face of the observer-error coordinate x_3 ($\theta = 1$), and naive grows from 1.9M and 1.6M domains at $\delta = 10^{-2}$ to 122M and 129M at 10^{-4}, local exponents 1.05 and 1.14, while the four that bind at the angle face x_0 grow only 1.7 to 2.5× and sb stays flat on all six (Figure 4). The contact sets the prefactor as well as the exponent. The lower bound's constant is loose (Appendix D.5), so we compare exponents, not counts. Below an absolute slack $c^\star \delta \approx 10^{-7}$ the verifier's float32 bounds can stall, which ends some sweeps.

Table 1: Contact geometry and frontier growth: θ predicted by Proposition 2 with the observed k, measured cap exponent, local cost exponent of the per-clause modes ($\log_{10}$ of the ratio of domains added over consecutive decades of slack), and growth of clip from $\delta = 10^{-2}$ to the finest verified slack. Released certificates above the rule, trained below (CT: CT-BaB, s-b: sampled boxes, $\times k$: k seeds). Coordinates are indexed from 0 as in the released code; for the quadrotor x is the horizontal position. *Stopped after $6 \cdot 10^9$ domains at the released level.

certificates	n	contact (active pieces k)	θ_{pred}	θ_{cap}	θ_{loc} naive; sb	clip growth
`pend-st`	2	smooth interior (0)	0.5	0.51	0.52; 0.51	1.01×
other five released; CT `pend`	2	vertex type (1)	0	0.00–0.01	flat	0.98–1.04×
`pend-out-st`	4	smooth interior (0)	1.5	1.50	1.34; 0.91	1.07×
`pend-out`	4	face + margin kink (2)	0.5	0.53–0.61	0.42; 0.41	1.02×
`pend-out-nn`	4	V kink, all axes (3)	0	0.04 ± 0.4	n/a; flat	1.01×
`quad`	6	smooth interior (0)	2.5	2.40	sb not run*	8.2×
CT `pend-out` (×3), s-b seed 1	4	face x_0, angle (1)	1.0	1.00–1.01	0.49–0.81; flat	0.99–1.01×
s-b `pend-out` seeds 0, 2	4	face x_3, observer error (1)	1.0	1.01–1.02	1.05, 1.14; flat	1.03, 0.97×
s-b + penalty (×3)	4	smooth interior (0)	1.5	1.41–1.50	float32 floor	1.13–1.25×
CT `quad` frozen (×2)	6	face $\dot{\vartheta}$ (1)	2.0	1.64, 1.97	sb timeout	1.01, 1.00×
CT `quad` split to end (×2)	6	face x, degenerate axis (1)	2.0	2.02, 2.22	sb flat	1.01×
CT `quad` frozen, 45k steps	6	face $\dot{\vartheta}$, decrease (1)	2.0	1.53	sb timeout	1.02×



(a) (b)

Figure 3: **Measured cost against estimates computed without branch-and-bound.** One point per certificate and mode at $\delta = 0.3$; colour gives the mode (naive, sb, clip), shape the method (circles: released, squares: CT-BaB, diamonds: sampled boxes); grey lines: equality and a factor 10. (a) Radius-field estimate N_ρ of (3). (b) First-order radius $m/\|g\|_1$ in place of ρ.

Table 2: **Measured over estimated cost across certificates.** At $\delta = 0.3$; spread: largest over smallest ratio. Leaf-width law: at $\delta = 10^{-2}$, certificates with recorded leaves (9, 14, 16).

	naive	sb	clip
certificates	18	24	25
cost range	$4.2{\cdot}10^3$	$2.3{\cdot}10^5$	$1.1{\cdot}10^5$
N/N_ρ	2.0–6.5	1.3–8.7	0.10–2.4
spread, N_ρ	3.2	6.5	24
spread, 1st order	600	660	640
spread, leaf law	26	$2.4{\cdot}10^4$	$3.6{\cdot}10^4$

The conjunction-aware mode behaves differently. It is flat, within 3% from $\delta = 10^{-2}$ to 10^{-4}, on every certified-trained certificate except the three with a looseness penalty (11 to 22% per decade), and in two dimensions it is 1.6 to 3.9× cheaper than the released configuration. On released certificates beyond two dimensions it grows, as $\delta^{-0.29}$ at the four-dimensional smooth contact and from 58.9M domains at $\delta = 10^{-2}$ to 480M at $3 \cdot 10^{-5}$ on the quadrotor: clipping removes the wedge where per-clause certification fails (Appendix B.7), not the relaxation error inside it.

6.3 The certified radius predicts the bulk cost

We computed (3) for every certificate at $\delta = 0.3$, where the frontier is negligible (the cost is 0.3 to 1.1 times that at $\delta = 10^{-2}$), with the bounder of the released verifier configurations and nothing taken from the BaB runs. For all 24 certificates with a completed per-clause run (18 of them in naive), the measured cost lies between 1.3 and 8.7 times N_ρ, over costs that span a factor $2 \cdot 10^5$ (Figure 3, Table 2), and two independent point sets agree within a median 5% and at most 37%. The ratio grows with the dimension (for sb: 1.7 to 2.7 in two dimensions, 1.3 to 4.6 in four, 4.5 to 8.7 in six), as the constants of Theorem 4 allow, and within one dimension each per-clause mode varies by at most 3.5× across certificates, training methods and cost levels. For clip the ratios are lower (0.36 to 2.4) because it accepts boxes that per-clause acceptance cannot; on the released quadrotor, where per-clause verification does not finish, clip needs 10× fewer domains than N_ρ.

Table 3: One-shot certification at $c^\star(1-\delta)$ against the released level c_0 ($\delta = 10^{-5}$ for $n = 2$, 10^{-3} for $n \geq 4$): certified-volume gain and increase in domains from c_0 to the target level (ranges over certificates; per certificate in Table 9). *Stopped after $6.3 \cdot 10^9$ domains, two of 12 instances.

certificates	$c_0/c^\star$	volume gain	N_{sb} increase	N_{clip} increase
six with $n = 2$	0.760–0.9997	1.000–1.191×	0.3–90%	0.1–7.4%
three with $n = 4$	0.978–0.994	1.009–1.092×	1.0–5.0%	1.1–1.8%
`quad`, $n = 6$	0.813	1.816×	stopped*	52.9M → 108.8M

Table 4: Certified training verified cold: CT-BaB's final partition and its own verified fraction (ver), trained (or released) over maximal level, binding contact, box fraction certified at $c^\star(1-10^{-3})$ (cert.), clip domains per unit certified fraction, and sb domains at $c^\star(1-10^{-3})$; values per seed (ranges for three). 45k row: frozen training partition; †training-batch estimate.

certificate	n	partition	ver	$c_0/c^\star$	contact	cert.	clip N/cert.	sb at 10^{-3}
released `pend-out`	4	–	–	0.982	face + kink	0.338	$1.7 \cdot 10^5$	0.50M
CT-BaB, 3 seeds	4	2.7–3.4M	0.990–0.997	0.979–0.998	face	0.603–0.615	$(0.95–1.2) \cdot 10^6$	1.0–1.2M
sampled boxes, 3 seeds	4	none	–	0.957–0.971	face	0.623–0.636	$(0.93–1.2) \cdot 10^6$	1.1–1.7M
+ looseness penalty	4	none	–	> 2000	interior	0	–	–
released `quad`	6	–	–	0.813	interior	0.0030	$3.6 \cdot 10^{10}$	not run
CT-BaB, frozen at 5k	6	32.3M, 29.0M	0.879, 0.858	1.48, 1.19	face $\dot\vartheta$	0.0180, 0.0275	$2.6 \cdot 10^9$, $1.0 \cdot 10^9$	timeout
CT-BaB, split to the end	6	65.0M, 57.5M	0.965, 0.964	1.08, 1.06	face x	0.0283, 0.0254	$2.0 \cdot 10^9$, $1.1 \cdot 10^9$	221M, 164M
CT-BaB, frozen, 45k steps	6	32.3M	0.886†	1.74	face $\dot\vartheta$	0.0145	$3.8 \cdot 10^9$	238M

Simpler predictors fail. First-order margin geometry, the same integral with $m/\|g\|_1$ in place of ρ (m and g the margin and gradient of a row), spreads by 600 to 660× per mode (790× pooled over per-clause modes) because it cannot see how loose the relaxation is. A leaf-width law with one relaxation constant per certificate fits the recorded leaves well, but across training methods its ratio to the measured cost spreads by $2.4 \cdot 10^4$ (sb) and $3.6 \cdot 10^4$ (clip), mostly through one containment row near a face of B, where its unclipped cubes must shrink and the verifier's clipped boxes need not (Appendix B.9). The radius field needs no fitted constant because it asks the bounder the question the verifier asks. Near the maximal level the estimate is less reliable: at $\delta = 10^{-2}$ per-clause ratios range from 0.78 to 9.1, and point sets disagree by 42 and 50× on two quadrotor certificates (Appendix C.3).

6.4 Gains from certifying at the maximal level

The released certificates sit 0.03% to 24% below $c^\star$ (Table 3). Certifying at $c^\star(1-\delta)$ instead enlarges the certified volume by up to 19% in two dimensions and 9% in four, for at most 7.4% more work in the conjunction-aware mode. On the quadrotor the certified volume at $c^\star(1-10^{-3})$ is 1.8× that at the released level, and clip proves it with 108.8M domains in 164 seconds on one A100, whereas the released per-clause configuration needed $5.0 \cdot 10^9$ domains and 23 minutes for the first of its 12 instances at the released level alone (Appendix D.2). For certified-trained quadrotor certificates the protocol is a necessity: all five valid ones were trained at a level $c_0 = 1$ above $c^\star$ (Table 4), and the one-shot level certifies each of them. We ran Algorithm 1 with a fixed slack and without the attack; the first attempt succeeded on every certificate.

6.5 Certified training under cold verification

CT-BaB optimizes the CROWN bound of (1) on an adaptively split partition of K, reuses that partition for fast verification, and by default freezes it after step 5,000. We retrained it with seeds 0 to 2 under this schedule and with splitting until the end, and verified every certificate cold (Table 4, Appendix D.9). Training is deterministic on our hardware, so paired runs are identical until step 5,000, where the schedules diverge (their logs first differ at steps 5,180 to 5,740).

The training-time verified fraction does not predict cold certifiability. Splitting to the end doubles the partition and raises CT-BaB's own verified fraction from 0.86–0.88 to 0.96–0.97 in both valid seeds, yet the certified fraction changes by ×1.57 and ×0.92, and the clip cost per certified fraction by ×0.77 and ×1.09, within the 2.6× spread between the two frozen seeds. Continuing the

frozen run for 30,000 more steps moves the two measures apart: the training-time fraction rises from 0.871 to 0.886 while the certified region shrinks by 19%. Seed 1 fails under both schedules, and its logs show it (2.9 attack counterexamples per batch against 0.25) before the schedules diverge.

Training moves the contact, and the contact decides which verifier pays the frontier. Both frozen certificates bind at the angular-velocity face $\dot{\vartheta} = 3$; both certificates split to the end bind at the face $x = 0.75$, where V barely depends on one coordinate (smallest normalized partial derivative $3 \cdot 10^{-4}$, against 0.015). As Remark 13 predicts, sensitivity splitting pays the frontier on the frozen certificates (no verdict after $9 \cdot 10^8$ domains at $\delta = 10^{-3}$ and $3 \cdot 10^{-3}$) and escapes it on the others (flat within 8% down to $\delta = 10^{-4}$), while clip is flat on all four (Figure 2c for seed 0, Table 12).

Neither intervention consistently lowers the bulk cost per certified volume. In six dimensions every valid CT-BaB quadrotor certificate is 9.5 to 36× cheaper per certified fraction than the released one, at 4.9 to 9.5× its certified fraction; in four dimensions the direction reverses (1.8 to 1.9× the fraction at 5.6 to 7.4× the cost per fraction). Training on boxes sampled afresh at every step (Appendix C.6) reproduces CT-BaB's certified region on `pend-out` (+1 to +5%) and its cold cost per certified fraction (clip ×0.98 to 1.10, sb ×1.00 to 1.51) in 18 to 22 instead of 28 to 35 minutes, but does not lower the cost; a penalty on the looseness of the bound collapses the certified region ($c^\star < 5 \cdot 10^{-4}$). The radius field agrees for the sampled boxes (N_ρ ×0.89 to 1.73) and falls by 14 to 21% under splitting to the end, within its band, while the measured cost does not. Neither training signal studied here targets the certified radius, the quantity that tracks cost across certificates.

7 Discussion

The cost of certifying a neural Lyapunov function has two parts: a frontier set by the contact of the level set with the violation set, and a bulk governed by the integral of the inverse n-th power of the bounder's certified radius. Both can be assessed in minutes before verifying, the frontier through its exponent and the bulk through N_ρ. They also show where verification-friendly training should act. The frontier is not the lever: conjunction-aware verification lowers its exponent, and training decides only as a side effect whether per-clause verifiers pay it. The bulk is, because its dependence on the certified radius is amplified by the dimension, and the training signals studied here do not target it.

Limitations. The lower bound's constant is loose; its matching upper bound covers only isotropic contracted verifiers. The radius field models per-clause acceptance on cubes, not the joint bound of clip or boxes elongated along degenerate axes, and its variance grows near $c^\star$. The certified-training study has two valid seeds per schedule in six dimensions and one system per dimension. Our protocol runs used a fixed slack and no attack, which sliver-shaped violation sets require.

AI use statement

In this work, we used generative AI tools to help develop the theoretical framework, formulate mathematical claims, provide ingredients for their proofs and write them, propose and refine hypotheses, design experiments, implement methods, and interpret results. We have not used generative AI tools to generate data sets, and translation, data cleaning and qualitative analysis are not applicable to this work. Additionally, we used generative AI tools for literature search, reference formatting, figure code, and drafting and editing text. We have reviewed all AI-assisted work, including every proof, table entry and reference. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

References

François Bachoc, Tommaso Cesari, and Sébastien Gerchinovitz. The sample complexity of level set approximation. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 130 of *Proceedings of Machine Learning Research*, 2021.

Felix Berkenkamp, Matteo Turchetta, Angela P. Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. URL https://arxiv.org/abs/1705.08551.

Nicholas M. Boffi, Stephen Tu, Nikolai Matni, Jean-Jacques E. Slotine, and Vikas Sindhwani. Learning stability certificates from data. In *Conference on Robot Learning (CoRL)*, volume 155 of *Proceedings of Machine Learning Research*, 2020. URL https://arxiv.org/abs/2008.05952.

Rudy Bunel, Jingyue Lu, Ilker Turkaslan, Philip H. S. Torr, Pushmeet Kohli, and M. Pawan Kumar. Branch and bound for piecewise linear neural network verification. *Journal of Machine Learning Research (JMLR)*, 21, 2020. URL https://arxiv.org/abs/1909.06588.

Michael A. Burr. Continuous amortization and extensions: With applications to bisection-based root isolation. *Journal of Symbolic Computation*, 77:78–126, 2016. URL https://arxiv.org/abs/1309.5991.

Ya-Chien Chang, Nima Roohi, and Sicun Gao. Neural Lyapunov control. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019. URL https://arxiv.org/abs/2005.00611.

Felipe Cucker, Alperen A. Ergür, and Josué Tonelli-Cueto. On the complexity of the Plantinga–Vegter algorithm. *Discrete & Computational Geometry*, 68(3):664–708, 2022.

Hongkai Dai, Benoit Landry, Lujie Yang, Marco Pavone, and Russ Tedrake. Lyapunov-stable neural-network control. In *Robotics: Science and Systems (RSS)*, 2021. URL https://arxiv.org/abs/2109.14152.

Alec Edwards, Andrea Peruffo, and Alessandro Abate. Fossil 2.0: Formal certificate synthesis for the verification and control of dynamical models. In *International Conference on Hybrid Systems: Computation and Control (HSCC)*, 2024. URL https://arxiv.org/abs/2311.09793.

Sven Gowal, Krishnamurthy Dvijotham, Robert Stanforth, Rudy Bunel, Chongli Qin, Jonathan Uesato, Relja Arandjelovic, Timothy A. Mann, and Pushmeet Kohli. On the effectiveness of interval bound propagation for training verifiably robust models. arXiv preprint arXiv:1810.12715, 2018. URL https://arxiv.org/abs/1810.12715.

Nikola Jovanović, Mislav Balunović, Maximilian Baader, and Martin Vechev. On the paradox of certified training. *Transactions on Machine Learning Research (TMLR)*, 2022. URL https://arxiv.org/abs/2102.06700.

Lukas Koller, Tobias Ladner, and Matthias Althoff. Out of the shadows: Exploring a latent space for neural network verification. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://arxiv.org/abs/2505.17854.

Suhas Kotha, Christopher Brix, J. Zico Kolter, Krishnamurthy Dvijotham, and Huan Zhang. Provably bounding neural network preimages. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. URL https://arxiv.org/abs/2302.01404.

Haoyu Li, Xiangru Zhong, Bin Hu, and Huan Zhang. Two-stage learning of stabilizing neural controllers via Zubov sampling and iterative domain expansion. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. URL https://arxiv.org/abs/2506.01356.

Jingyue Lu and M. Pawan Kumar. Neural network branching for neural network verification. In *International Conference on Learning Representations (ICLR)*, 2020. URL https://arxiv.org/abs/1912.01329.

Yuhao Mao, Mark Niklas Müller, Marc Fischer, and Martin Vechev. Understanding certified training with interval bound propagation. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://arxiv.org/abs/2306.10426.

Zhouxing Shi, Haoyu Li, Cho-Jui Hsieh, and Huan Zhang. Certified training with branch-and-bound for Lyapunov-stable neural control. In *Learning for Dynamics and Control Conference (L4DC)*, 2026. URL https://arxiv.org/abs/2411.18235.

Shiqi Wang, Huan Zhang, Kaidi Xu, Xue Lin, Suman Jana, Cho-Jui Hsieh, and J. Zico Kolter. Beta-CROWN: Efficient bound propagation with per-neuron split constraints for neural network robustness verification. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. URL https://arxiv.org/abs/2103.06624.

Jesse Wayment, Brian Yarbrough, Jingbo Wang, Shreyas Sundaram, and Philip E. Paré. HyParLyVe: Hyperplane partitioning for neural Lyapunov verification. arXiv preprint arXiv:2605.03992, 2026. URL https://arxiv.org/abs/2605.03992.

Junlin Wu, Andrew Clark, Yiannis Kantaros, and Yevgeniy Vorobeychik. Neural Lyapunov control for discrete-time systems. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. URL https://arxiv.org/abs/2305.06547.

Kai Y. Xiao, Vincent Tjeng, Nur Muhammad Shafiullah, and Aleksander Madry. Training for faster adversarial robustness verification via inducing ReLU stability. In *International Conference on Learning Representations (ICLR)*, 2019. URL https://arxiv.org/abs/1809.03008.

Kaidi Xu, Zhouxing Shi, Huan Zhang, Yihan Wang, Kai-Wei Chang, Minlie Huang, Bhavya Kailkhura, Xue Lin, and Cho-Jui Hsieh. Automatic perturbation analysis for scalable certified robustness and beyond. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. URL https://arxiv.org/abs/2002.12920.

Kaidi Xu, Huan Zhang, Shiqi Wang, Yihan Wang, Suman Jana, Xue Lin, and Cho-Jui Hsieh. Fast and complete: Enabling complete neural network verification with rapid and massively parallel incomplete verifiers. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://arxiv.org/abs/2011.13824.

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 for efficient synthesis and verification. In *International Conference on Machine Learning (ICML)*, 2024. URL https://arxiv.org/abs/2404.07956.

Hongchao Zhang, Sicun Gao, Zhizhen Qin, and Andrew Clark. SEEV: Synthesis with efficient exact verification for ReLU neural barrier functions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. URL https://arxiv.org/abs/2410.20326.

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 (NeurIPS)*, 2018. URL https://arxiv.org/abs/1811.00866.

Duo Zhou, Jorge Chavez, Hesun Chen, Grani A. Hanasusanto, and Huan Zhang. Clip-and-verify: Linear constraint-driven domain clipping for accelerating neural network verification. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. URL https://arxiv.org/abs/2512.11087.