PARAMETRIC NEURAL LYAPUNOV VERIFICATION WITH A SHARED PARTITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Independent verification characterizes the disturbance a neural Lyapunov certificate tolerates by rebuilding the branch-and-bound computation at every disturbance size and every candidate level. We show that on every state box the residual is affine in the disturbance size and in the decrease rate, with coefficients that depend on neither, so bounding those coefficients once gives a valid residual bound for every size and rate, and one adaptive partition of the state region serves the whole sweep. The same stored bounds, with validated points, bracket the smallest certificate level that survives a given disturbance size, from above and below and at every size in the interval rather than at the tested sizes alone, at a measured extra cost. On a power-system swing model and a Van der Pol oscillator the shared partition answers the same sequence of checks with 5.75 to 8.42 times fewer bound calls than independent verification. On seven released neural controllers for four plants, from two published pipelines, it uses 2.39 to 15.61 times fewer. In both comparisons it preserves every level the independent run certified. The same stored bounds answer decrease rates at or below the trained rate without a further bound call, and the advantage grew with the state dimension on a family of coupled oscillators. The partition is computation that branch-and-bound verification already performs, kept and reused instead of discarded with the answer.

1 INTRODUCTION

A neural certificate is a trained function whose stability conditions are checked over a region of state space by a verifier (Chang et al., 2019; Yang et al., 2024; Li et al., 2025b). Once such a certificate exists, the question asked of it is quantitative. An engineer wants to know how large a disturbance the closed loop absorbs and how close to the equilibrium the guarantee still holds, and the answer is not a single number. It is a profile: for each disturbance size, the smallest level of the certificate at which the verified decrease condition survives that size. Reading that profile is what turns a trained certificate into a statement about the plant.

A verifier is normally called on one check at a time: it builds a branch-and-bound tree over the state region for one disturbance size and one level, and discards the tree with the answer. A sweep over sizes with a level search inside it therefore pays for a fresh tree at every point, and the cost grows with the number of sizes times the depth of the level search. Monotonicity carries each certified level to the sizes below its grid point, so the sweep bounds the profile from above between the points, but it certifies no lower bound there, and how far its step curve sits above the profile is unknown. Both limitations come from the same source: the work a verifier does is thrown away with the tree that held it.

Control theory computes disturbance gains and parameter-range guarantees without a neural-network verifier, by linear programming over a simplicial grid (Li et al., 2015), by sum-of-squares programming for polynomial vector fields (Ichihara, 2012), and by certifying one inequality over a whole parameter range in reachability (Meyer, 2023; Althoff et al., 2008) and in parameter-dependent linear matrix inequalities (Gahinet et al., 1996). Each fixes the class of certificate in advance, and none reuses the work a bound-propagation verifier already does on a trained network.

The verification literature does reuse that work, in one direction. Incremental verification restarts from a stored branch-and-bound tree after the network weights change (Ugare et al., 2023), shared certificates reuse a propagated abstraction across specifications under a containment test (Fischer

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.

et al., 2022), learned conflicts reuse infeasible activation patterns when the second check refines the first (Elsaleh et al., 2026), and the two-stage pipeline of Li et al. (2025b) keeps verified subdomains while it moves the two levels of its own condition. The last three carry work from a check to a refinement of it, and a sweep over disturbance sizes moves the other way, because a larger size admits more disturbances and a lower level enlarges the region, so a retained positive answer does not transfer. A relaxation valid over the whole size interval follows from one pass with the size as an input coordinate (Xu et al., 2020); Section 5 measures it as a baseline. No method we have located performs one branch-and-bound partition of the state space and certifies a two-sided functional bound on a certificate-dependent threshold simultaneously over a continuum of an external parameter such as a disturbance magnitude, a perturbation radius, or a comparable specification scalar, though the same maximum-of-minima shape appears, solved by different means, in parametric flexibility analysis (Swaney & Grossmann, 1985).

This paper reuses the work in the direction the sweep needs, by keeping quantitative records instead of answers. For a fixed certificate the verified residual separates into a part that does not involve the disturbance and a part that scales linearly with its size, and the two parts are bounded independently on a box of states, so bounding a box once bounds it for every size in an interval, and the same stored numbers are re-read when the level moves. One adaptive partition of the state region then serves an entire sweep with the level searches inside it, and a check is settled on a leaf by re-evaluating three inequalities against the stored numbers rather than by calling the bound routine again.

The same records support a second guarantee over the interval of sizes: assembled over the retained leaves, they form an upper function of the size, validated points of the region form a lower function, and the threshold profile is enclosed between the two at every size including the sizes between the tested ones, at a cost measured against the sweep on the same instances.

Those records are the main contribution, and they have two uses. The first is as a verifier: under a strict budget cap the shared partition answers the recorded protocol of dissipation checks with 7.31 to 8.42 and 5.75 to 6.56 times fewer bound calls than independent verification on the two systems, with certified answers matching or stronger at every position of that protocol and no validated refutation lost, and it does the same on released closed loops of prior work. On a synthetic family of two to four states the advantage grew with the state dimension. The second is the enclosure of Theorem 1, assembled from the same records and validated points and certified over the whole interval to a width of at most 0.1628 and 0.0276 times the upper level, at a measured added cost of 22.3 and 19.7 percent of the bound calls used by the sweep when the sweep's levels serve as the target. The code, certificates and result files are released with the paper.

2 Setting and the parametric verification problem

The family of problems that verification work is reused across has to be written down first. This section fixes the plant class, the certificate class, and the one condition the verifier checks, says which of the four answers that condition can return constrain anything, and defines the threshold the disturbance sweep is searching for. That threshold depends on the trained certificate, the dynamics, the rate, and the region, and on nothing a particular search does, so it serves as an object two methods can be compared on.

Systems, certificates and the dissipation check. A neural certificate is a network trained to satisfy a Lyapunov-type condition on a region of state space, and a verifier checks that condition, for example by bound propagation inside a branch-and-bound loop (Chang et al., 2019; Zhang et al., 2018; Xu et al., 2020; 2021; Wang et al., 2021; Yang et al., 2024; Li et al., 2025b). We work with continuous-time input-affine systems subject to a bounded exogenous disturbance,

$$\dot{x}(t) \;=\; f_0(x(t)) + B\,w(t), \qquad x(t) \in \mathbb{R}^n,\; w(t) \in \mathbb{R}^m, \tag{1}$$

under Assumption 1 (Appendix B) and the standing conditions that the origin is an interior equilibrium, $f_0(0) = 0$ and $0 \in \operatorname{int} \mathcal{X}$, with $\mathcal{W}(\bar{w}) = \bar{w}\,\mathcal{W}_1$ for a fixed compact convex $\mathcal{W}_1$ containing the origin. Solutions are maximal Carathéodory solutions (Sontag, 1998). The construction of Section 3 rests on the residual being affine in the size, which the scaled family gives, as would a state-dependent $B(x)$ and several channels with separate sizes, whereas a family that is not a scaled copy of one set, or a parameter entering f_0, gives no such structure.

A candidate certificate is a continuously differentiable $V\colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ with $V(0) = 0$ that admits a quadratic lower bound on the domain: there is a symmetric positive definite P with

$$V(x) \;\geq\; x^\top P x \;\text{ on } \mathcal{X}, \qquad \lambda_P \coloneqq \lambda_{\min}(P) > 0. \tag{2}$$

The class used here adds to that quadratic form a one-layer input-convex residual of K softplus tangent-gap features with nonnegative weights (Appendix C), so that the quadratic lower bound holds by construction and ∇V is available in closed form. Fix a decrease rate $\kappa > 0$. For a disturbance size $\bar{w}$ the *worst-case residual* is

$$\rho_{\bar{w}}(x) \;\coloneqq\; \langle \nabla V(x), f_0(x) \rangle \;+\; \zeta_{\bar{w}}(x) \;+\; \kappa\, V(x), \qquad \zeta_{\bar{w}}(x) \;\coloneqq\; \sup_{w \in \mathcal{W}(\bar{w})} \langle \nabla V(x), B w \rangle, \tag{3}$$

where the disturbance term is the support function of $\mathcal{W}(\bar{w})$ at $B^\top \nabla V(x)$, with closed forms for the box, ball and polytope classes used here (Proposition 3). For $0 < a \leq b$ write $\mathcal{A}_a \coloneqq \{x \in \mathcal{X} : V(x) \leq a\}$ and $\mathcal{S}(a, b) \coloneqq \{x \in \mathcal{X} : a \leq V(x) \leq b\}$. The verifier checks the following condition, the *dissipation check*: given a size $\bar{w}$, a level c and an upper level $c_{\max}$,

$$\rho_{\bar{w}}(x) \;\leq\; \kappa\, c \qquad \text{for all } x \in \mathcal{S}(c, c_{\max}). \tag{4}$$

Our implementation evaluates the bounds behind (4) in float64 without directed rounding and subtracts a fixed margin of 10^{-9} from every clause, so the guarantee is the real-arithmetic statement up to a rounding error we bound empirically and do not prove.

Answers of the dissipation check. A check returns one of four answers: certified, when every leaf of a finite cover of $\mathcal{S}(c, c_{\max})$ is resolved, so (4) holds; counterexample, a validated refutation by a box whose enclosures place V at least $\varepsilon_0 > 0$ inside the shell and $\rho_{\bar{w}}$ above $\kappa c + \varepsilon_0$ at every point or by a point re-checked through the verifier's own enclosure; unresolved, when the budget or the width threshold was reached with boxes undecided and no refuting box; and error. Only the first two settle anything about a level, and an unresolved outcome is silent on both sides of the level (Lemma 1).

The certificate-dependent threshold. The level that separates the two settling answers is, once V, κ, and the region $\mathcal{A}_{c_{\max}}$ are fixed, a function of the disturbance size alone. Write

$$H(\bar{w}) \;\coloneqq\; \max_{x \in \mathcal{A}_{c_{\max}}} \min\big\{\, V(x),\; \rho_{\bar{w}}(x)/\kappa \,\big\}, \tag{5}$$

the certificate-dependent threshold of the fixed pair (V, κ) on the fixed region: every $c \in (H(\bar{w}), c_{\max})$ satisfies (4) on $\mathcal{S}(c, c_{\max})$ and every $c < H(\bar{w})$ in $(0, c_{\max})$ violates it (Lemma 3). A certified level $c^*(\bar{w})$ is therefore an upper bound of $H(\bar{w})$ reached at a finite budget, and the distance between the two is a property of the search. Since $\zeta_{\bar{w}} = \bar{w}\zeta_1$ with $\zeta_1 \geq 0$, the residual and H are nondecreasing in $\bar{w}$, which Theorem 1 uses to normalize its envelopes.

The threshold is worth certifying because a level converts into a trajectory bound. Under Assumptions 1–4, in particular the verified boundary separation $\inf_{\partial\mathcal{X}} V > c_{\max}$ and (4) at $(\bar{w}, c)$, Theorem 2 (Appendix A) states that every solution from $\mathcal{A}_{c_{\max}}$ stays there and satisfies $\limsup_{t\to\infty} \|x(t)\| \leq \sqrt{c/\lambda_P}$ for every measurable disturbance of size at most $\bar{w}$, which is the regional counterpart of the asymptotic gain of input-to-state stability, read off a verifier answer (Sontag, 1989; Sontag & Wang, 1996). Table 2 in Appendix C reports that reading on six certificates.

This paper is about producing the profile over sizes $\bar{w} \in [0, \bar{w}_{\text{cert}}]$, with $\bar{w}_{\text{cert}}$ the largest size of the sweep. The size, written ω in the tables, and the level c are two parameters over which a verifier is normally run one point at a time.

3 Parametric records and the shared partition

A sweep over disturbance sizes with a bisection over levels inside it runs the branch-and-bound check of Section 2 many times at different values of two parameters, and a fresh start discards the part of the proof independent of them. The residual depends affinely on the size on every box, so a bound on its two size-independent pieces holds for every size at once, and the level enters only through the clauses that settle a box, so one adaptive partition serves the whole sweep provided its records stay valid under refinement and as either parameter moves in either direction.

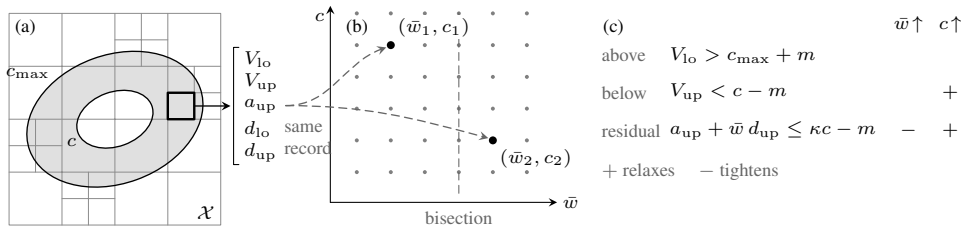


Figure 1: One partition serves every check of the sweep: (a) a leaf inside the shell between the level sets at c and $c_{\max}$ carries five stored bounds; (b) the same record settles checks at different sizes and levels, one of them inside a level bisection; (c) the three clauses that settle a leaf, with the direction in which each parameter moves them.

3.1 Records valid over an interval of sizes

Since $\mathcal{W}(\bar{w}) = \bar{w}\,\mathcal{W}_1$, the disturbance term of (3) is positively homogeneous in $\bar{w}$ and the residual splits as $\rho_{\bar{w}} = a + \bar{w}\, d$ with

$$a(x) \;:=\; \langle \nabla V(x), f_0(x) \rangle + \kappa V(x), \qquad d(x) \;:=\; \zeta_1(x) \;\ge\; 0, \tag{6}$$

both independent of $\bar{w}$. A leaf Q of the partition therefore carries a record of five numbers: interval bounds $V_{\rm lo}, V_{\rm up}$ on V, an upper bound $a_{\rm up}$ on a and bounds $d_{\rm lo}, d_{\rm up}$ on d over Q. One bound computation per leaf serves every size in $[0, \bar{w}_{\rm cert}]$, and since a is affine in κ as well, a record that bounds its two terms separately serves every decrease rate at once, which Section 5.2 uses.

A check at $(\bar{w}, c)$ is settled on a leaf by one of three clauses, each carrying the arithmetic margin m, set to 10^{-9}, in the conservative direction. The leaf lies above the region between the levels, with $V_{\rm lo} > c_{\max} + m$, or it lies below the region, with $V_{\rm up} < c - m$, or the reassembled residual bound clears the level, $a_{\rm up} + \bar{w}\, d_{\rm up} \le \kappa c - m$. Leaves that none of the three settles are split. Children start from their parent's bounds, which hold on any sub-box, and a parent never takes a child's. The two parameters move the clauses in opposite directions: raising $\bar{w}$ tightens the residual clause, while raising c relaxes it and at the same time removes part of the region through the second clause. The partition therefore retains the five numbers, and re-evaluates the three clauses from them as either parameter moves (Figure 1).

A check ends in a refutation when a point of the region is validated as a violation, evaluated as a degenerate box through the same verifier and margin. The residual at a fixed point is affine in the two parameters with coefficients that depend on neither, so a stored point's residual at any other size and rate is a closed-form evaluation. Validated points are kept beside the leaf records and revalidated at the start of later checks, so the same partition serves both answers.

A hash of the certificate, dynamics, domain, disturbance geometry, κ, $c_{\max}$, margin and bound-routine source accompanies every record, and a record with another hash is refused. A check returns one of the four answers of Section 2, and resetting the partition before every check reproduces the verifier of that section box for box, so the comparisons of Section 5 hold the branching and bounding fixed. The budget can be enforced strictly by slicing every call to the bound routine.

Anchored transport on an interval backend. A cached bound could be moved along the size axis at its cached sensitivity, or interpolated between two anchors, instead of being recomputed; on a backend that bounds a and d separately neither improves on the separable bound.

Proposition 1 (Interval identity)**.** *On a box Q with $0 \le d_{\rm lo} \le d \le d_{\rm up}$ and $a \le a_{\rm up}$, write $\mathrm{Sep}(\bar{w}) = a_{\rm up} + \bar{w} d_{\rm up}$ and take anchors $R_i = \mathrm{Sep}(\bar{w}_i)$, $R_j = \mathrm{Sep}(\bar{w}_j)$ at sizes $0 \le \bar{w}_i < \bar{w}_j$. The forward transport $R_i + d_{\rm up}(\bar{w} - \bar{w}_i)$ on $\bar{w} \ge \bar{w}_i$ and the chord through R_i and R_j on $[\bar{w}_i, \bar{w}_j]$ both equal $\mathrm{Sep}(\bar{w})$, and the backward transport $R_i - d_{\rm lo}(\bar{w}_i - \bar{w})$ exceeds $\mathrm{Sep}(\bar{w})$ on $[0, \bar{w}_i)$ unless $d_{\rm up} = d_{\rm lo}$, so none of them is tighter than $\mathrm{Sep}(\bar{w})$.*

The proof is a direct computation (Appendix D). The reassembled clause of Section 3.1 already equals what the forward transport and the chord return, so the shared partition needs no transport machinery. The proposition does not cover a backend that relaxes $a + \bar{w}d$ jointly, where the joint bound can be tighter than the separable one.

4 CERTIFIED ENCLOSURE OF THE THRESHOLD PROFILE

A sweep certifies levels at the sizes it visits, and monotonicity extends each level leftward into a step curve that bounds the threshold from above at every size up to the largest one tested. The sweep gives no lower bound, so how far that step curve sits above the threshold is unknown, and the curve is only as fine as the grid. This section states what the same partition and bound graph certify on both sides at every size of the interval. Each record bounds the residual for every size in the interval at once, so the largest level at which a box can still fail the check is bounded by a function of the size, and the maximum of these functions over the retained boxes is an upper envelope of the certificate-dependent threshold over the whole interval. Validated points of the region give a lower envelope through the same bound graph, so the pair brackets the threshold with a certified width.

Theorem 1 (Certified profile enclosure)**.** *Let finitely many boxes cover $\mathcal{A}_{c_{\max}}$, a box being discarded only when a validated lower bound on V proves it disjoint from $\mathcal{A}_{c_{\max}}$. For each retained box Q let $V_Q^+ \geq \sup_Q V$ and let $R_Q^+(\bar{w}) \geq \sup_{Q\cap\mathcal{A}_{c_{\max}}} \rho_{\bar{w}}$ for every $\bar{w} \in [0, \bar{w}_{\mathrm{cert}}]$, and let $x_1, \ldots, x_J \in \mathcal{A}_{c_{\max}}$ be validated points with validated lower values $v_j^- \leq V(x_j)$, $a_j^- \leq a(x_j)$ and $0 \leq d_j^- \leq d(x_j)$. Put*

$$U(\bar{w}) \coloneqq \max_Q \min\bigl\{c_{\max}, V_Q^+, R_Q^+(\bar{w})/\kappa\bigr\}, \quad L(\bar{w}) \coloneqq \max\Bigl\{0, \max_j \min\bigl\{v_j^-, (a_j^- + \bar{w}d_j^-)/\kappa\bigr\}\Bigr\}. \tag{7}$$

Then $L(\bar{w}) \leq H(\bar{w}) \leq U(\bar{w})$ at every $\bar{w} \in [0, \bar{w}_{\mathrm{cert}}]$; the normalizations $\tilde{U}(\bar{w}) = \inf_{t\geq\bar{w}} U(t)$ and $\tilde{L}(\bar{w}) = \sup_{t\leq\bar{w}} L(t)$ are nondecreasing and enclose H as well; and for any finite subdivision of $[0, \bar{w}_{\mathrm{cert}}]$, the largest over the subintervals I of an upper bound on $\sup_I U$ minus a lower bound on $\inf_I L$ is an upper bound of $\sup_{\bar{w}}(U(\bar{w}) - L(\bar{w}))$.

The proof is in Appendix D with the two conditions the implementation meets, clamping the lower sensitivity at zero and bounding the supremum of each candidate separately, since a maximum of minima of affine functions can peak inside a subinterval. The reported width therefore holds over the whole interval.

Returning a level from the upper envelope adds a margin τ (1.0×10^{-8} and 1.3×10^{-8} on the two instances), the enclosure's counterpart of m; the construction is anytime, and refinement splits one leaf at a time.

Which leaf to split is the main design choice, and three families of rules are compared (Appendix D): the max-gap rule splits the leaf contributing most to the widest subinterval; the curve-targeted rule splits where U most exceeds a supplied target, here the sweep's certified levels multiplied by $1+0.25$, so it measures the cost of certifying a supplied profile; and the oracle-free rules take no target and stop on a tolerance η fixed in advance. Section 5 reports two measures with their cost: closeness to the pointwise reference, defined as U at most $1 + 0.25$ times the sweep's level at every node, and the bound on the width of the enclosure (7) over the whole interval $[0, \bar{w}_{\mathrm{cert}}]$.

5 EXPERIMENTS

Two questions decide whether the construction is worth using: whether one shared partition answers a recorded list of checks for less work than independent verification with answers never weaker (Section 5.2), and what the enclosure of Section 4 costs (Section 5.3). Section 5.4 turns to larger state dimensions.

5.1 PROTOCOL

The study of the reuse rules uses four instances: a development and a held-out certificate for the single-machine infinite-bus (SMIB) swing dynamics (the primary system) and for the Van der Pol oscillator, with the held-out pair drawn from other seeds and left unused while the reuse rules were chosen. Each instance freezes the certificate, the dynamics, the domain box, the disturbance geometry, κ, the margin and the upper level. On the primary system the upper level is $c_{\max} = 1.3409$, which lies a margin below the certified boundary value 1.4115. The recorded list of checks is the node bisection of Appendix A, whose nodes are the sizes of the recorded grid and whose upper

Table 1: Cost and agreement of the shared partition against independent verification on the recorded protocol. Ratios are the row over its block's shared row. The interval rows are run under the strict budget and carry the ratio of median wall times with its range. The CROWN rows count relaxation calls against the implemented one-pass verifier replaying the independent run's per-check budgets, at its nominal budget, and at the nominal budget with its frontier grown to the shared partition's average batch, a variant that returns the same statuses at every position, each of the three variants carrying its paired wall-clock ratio (Tables 18 and 20).

instance, backend	method	calls	ratio	eq/str/weak	wall vs B1
SMIB-0, interval	B0 independent	3,390,698	8.42	91/0/0	4.67 [4.25, 6.63]
	B1 shared	402,547	1.00	72/19/0	
VDP-0, interval	B0 independent	1,449,794	5.75	91/0/0	7.25 [5.37, 8.54]
	B1 shared	252,161	1.00	91/0/0	
VDP-0, CROWN	one-pass independent (matched budgets)	1,133	3.79	77/0/0	3.50 [3.48, 3.83]
	one-pass independent (nominal budget)	1,162	3.89	77/0/0	3.81 [3.71, 4.05]
	one-pass independent, batched	378	1.26	77/0/0	1.33 [1.30, 1.33]
	B1 shared	299	1.00	77/0/0	

Node order; strict cap, 60,000 bound calls per check. Ratio, wall: row over B1, medians [range]; CROWN: paired, relaxation calls; batched: frontiers accumulated to the shared partition's boxes per call. Agreement: soft-cap run vs the independent run's trace. Appendix: fuller tables.

level the appendix tables write as b: 91 and 91 checks on the interval backend at $60{,}000$ bound calls each, and 77 checks at $2{,}000$ bound calls each on the linear-relaxation backend, replayed on the first backend in three orders (sizes increasing, sizes decreasing and one seeded permutation) and on the second in two, with three repetitions each. Bound calls, the number of boxes whose bounds are freshly computed, are the counted resource, with wall time beside them and the timing environment in Appendix D.2; the scan for a violating point is not charged to bound calls, so the methods are matched in bound calls alone.

5.2 Cost of answering the recorded checks

The methods below answer the same recorded list of checks, so a difference between them in bound calls is a difference in the reuse rule, with the branching and bounding held fixed and the comparison made within a backend. B0 resets the partition before every check and B1 shares it. Under the strict budget B1 also keeps an unaffordable frontier's parents as leaves and scans the whole residue for validated points, and B0 scans a capped subset. Rerun under those rules, B0 spends the same bound calls up to one call at a check that reaches its budget, and its answers change only where the full scan refutes a check it left unresolved. P0 adds the anchored residual records of the transport paragraph of Section 3.1. On the linear-relaxation backend, B1a takes one relaxation valid over the whole size interval by appending the size to the input box (Xu et al., 2020), and B1b carries the node sizes as separate output rows of one call, with a variant that first splits the size interval into K equal parts. B1c resets the partition at every change of size, reusing along the level axis alone, as a labelled adaptation of Li et al. (2025b), whose pipeline could not be run directly on these instances (Appendix D.8). One bound on the disturbance term over the whole region returns a level above $c_{\max}$ on every trained certificate of Table 2 and is not a baseline.

Table 1 gives the cost and the agreement counts: the cost in node order under the strict per-check budget and the counts from the soft-budget run. Over the three orders the bound-call ratios run 7.31 to 8.42 on SMIB and 5.75 to 6.56 on Van der Pol, and no check exceeded its budget in any block. Table 14 (Appendix D) reports the strict-budget run in full and Figure 8 its reduction against the number of sizes sharing the partition, and Tables 12 and 13 report every method and block of the soft-budget run, in which a check may overshoot its budget by one frontier. On these two instances the whole protocol runs in seconds either way, and on the released loops below in about a minute against a few seconds (Table 22, with the measurement conditions in Appendix D.2).

The shared partition's certified answers do not weaken at any position of the recorded protocol, and its 19 stronger answers in Table 1 occur where the independent run ended at its budget. It also loses no validated refutation at any of the $10{,}114$ positions of the 17 blocks measured under these rules, and at 14 of them a stored point turns an unresolved answer into a validated counterexample. No

reused answer certified a check the independent run refuted with a validated point, in any block, here or in the soft-budget replays and the batched pilot (0, 0, 0, 0 and 0 such answers).

The same protocol runs on released closed loops of prior work. We take the neural controllers released by Yang et al. (2024) for an inverted pendulum and by Li et al. (2025b) for the pendulum, path tracking, a double integrator and a Van der Pol oscillator, embed each controller in the bound-propagation graph unchanged, and train a certificate for the loop, which gives 7 instances on 4 plants on the linear-relaxation backend. On them the shared partition answers the 52 recorded checks with 2.39 to 15.61 times fewer bound calls than independent verification over the three orders, while issuing 8.99 to 21.63 times fewer relaxation calls. It weakens no level and leaves 0 soundness discrepancies over 1,092 positions (Table 21, Appendix D.5). Independent verification here is the scheduler's baseline, which relaxes a box in two passes (below). A one-pass verifier batched to the shared partition's frontier spends 1.657 to 9.204 times the shared partition's bound calls and is 1.525 to 6.219 times slower in wall time (Table 23).

Across the certificates of the seed sweep, 10 primary and 5 Van der Pol, the hard-cap ratio in node order lies between 4.344 and 12.958 (median 9.282) on the 7 primary certificates the verifier certifies and between 5.567 and 6.493 on Van der Pol, with 0 soundness discrepancies over 1365 positions, and both methods refuse the same 3 primary certificates at every check (Table 16, Appendix D). On those certificates the ratio rises to 90.846, 91.002 and 124.118, because every check on them ends in a validated counterexample, and a point validated once settles later checks against the stored records without refining anything, whereas the independent run spends its whole budget at every check before its residue scan exposes the violation, so these ratios scale with the budget. The other two orders give medians of 9.646 and 8.046 on the primary system and 6.512 and 6.447 on Van der Pol, and over the 45 runs of that cohort no certified level weakens, no validated refutation is lost and no soundness discrepancy occurs.

On the linear-relaxation backend the comparator is an implemented one-pass verifier, since the scheduler's own independent baseline issues a size-independent pass and then a joint pass on whatever the first leaves open, while a verifier written for one size issues one pass. On the evaluated Van der Pol checks, in paired within-session comparisons against that one-pass baseline at four threads, the shared partition cuts wall time by a factor of 3.503 in node order and 4.762 in the permuted order, with matching verification statuses at all 77 positions in both orders, and a second session at 8 threads agrees (Table 19). Each statistic is the median of three interleaved repetitions (Table 18). Here the one-pass baseline replays the independent run's per-check spending, and letting it spend the nominal budget instead gives wall-time factors of 3.813 and 4.940.

Most of that advantage is call structure. The boxes the two runs put through the relaxation stand in ratios of 1.009 and 1.000, while the relaxation calls that carry them stand in ratios of 3.789 and 5.294, because a verifier that restarts at the root bounds small frontiers. A one-pass verifier that grows its frontier, before each relaxation call, to the shared partition's average batch (148 boxes in node order and 209 in the permuted order) returns the same statuses at all 77 positions and is slower than the shared partition by the factors 1.330 and 1.646, since it bounds 1.595 and 1.823 times the boxes to reach large frontiers by pre-splitting (Table 20, Appendix D.4). That residual is the measured value of reuse over this batched one-pass verifier on this backend.

Appending the size to the input box and splitting the size interval before the joint bound put 23.0 and, at best, 12.7 percent more boxes through the relaxation than the shared partition, respectively (Table 27).

On SMIB and Van der Pol respectively, reuse along the level axis alone spends 1.34 and 1.39 times the shared partition's bound calls under the soft cap when the checks are grouped by size, and 5.00 and 4.34 times when they interleave sizes (Appendix D.8).

On the interval backend the stored partition can also be re-checked at other decrease rates. Re-read at 4 rates at or below the training rate with 0 and 0 further bound calls on the two systems, the partition built by the training-rate sweep certifies levels within 1.18 times those of independent sweeps on the primary system and equal to them on Van der Pol, with 0 soundness discrepancies over all 1092 positions of the six-rate record, using 27.0 and 20.7 times fewer bound calls in total than those sweeps. Letting the re-check refine the leaves that a stored record does not settle, within an allowance of 30,000 bound calls per check, brings the primary system to the independent sweeps' levels at every one of the 28 rate-node pairs and leaves 0 weaker positions out of 364, for 17.31

percent of those sweeps' bound calls in the re-check's level searches, or 29.27 percent also counting the same refinement run on every check of the independent sweeps' lists for the comparison by position. Above the training rate the certified level is looser than the independent level by factors of up to 1.71 on the primary system and 202.69 on Van der Pol, over the nodes that still certify a level (Table 24, Appendix D).

5.3 The profile enclosure and its price

Figure 10 in Appendix D plots the two envelopes beside the sweep's certified levels and $\tilde{U}$ over the sweep's level at each node against the line at $1 + 0.25$, and Table 29 prices every rule at two budgets. The curve-targeted rule brings $\tilde{U}$ to at most the target at every node on both instances, at $111{,}167$ bound calls on the primary system, where this study's sweep, run before the rules of Section 5.2 were adopted, spends $498{,}425$, and at $54{,}179$ on Van der Pol, where it spends $274{,}355$. That is 22.3 and 19.7 percent of the sweep's bound calls, so the sweep followed by the refinement costs 1.22 and 1.20 times the sweep alone in bound calls. The same runs put 2.32 and 3.77 times the sweep's boxes through the bounding backend (Table 29), and the enclosure remains slower than the sweep in wall time on this backend, where the bookkeeping around a cheap bound call dominates. The width guarantee the curve-targeted rule reaches at that budget is 0.1628 and 0.0276 times $c_{\max}$ over the whole interval. The max-gap rule at the same budget reaches 0.0556 and 0.0200 (Table 29) without bringing $\tilde{U}$ under the target at every node. Without a target the best oracle-free rule brings the largest node ratio $\tilde{U}/c^*$ to 1.2747 and 0.9983, under the target on Van der Pol alone, at 2.01 and 2.24 times the sweep's bound calls.

The finite-check baseline of Shubert (1972), with a validated modulus in place of an assumed Lipschitz constant, reaches a matched certified width of 0.05 times $c_{\max}$ at 15.75 and 8.20 times the enclosure's bound calls on the primary system and Van der Pol, respectively. Wall-time measurements differ from the call counts: on the interval backend the enclosure is slower, 23.1 against 11.9 seconds on the primary system (Table 26), and on the linear-relaxation backend the two are comparable at the reduced witness grid (Table 28). A sweep on a finer size grid with the same validated lower envelope is the second baseline: at 25 nodes it reaches a width of 0.0618 times $c_{\max}$ on the primary system at 4.8 times the max-gap enclosure's bound calls, wider than that enclosure's width, and 0.0104 on Van der Pol, narrower than any recorded enclosure width, which is bounded below there by its gap-certification tolerance (Table 25).

5.4 Larger state dimensions

The reuse mechanism has no step whose form depends on the state dimension. On a family of damped coupled oscillators with quadratic certificates (a family amended once, after its first version failed at four states), under the protocol of Section 5.2 with a hard budget of 60,000 bound calls per check at two and three states and 960,000 at four, the base check at zero disturbance and level $b/2$ certifies at two, three and four states. There the shared partition answers the 91 checks with 3.33, 8.06 and 11.98 times fewer bound calls than independent verification, with no soundness discrepancy and no weaker answer (Table 31, Appendix D.14). The ratio follows the independent run's budget saturation: that run ends at its budget on 4, 30 and 41 of those checks while the shared partition answers 53, 53 and 70 of them without a fresh bound call. The base verifier's covering count and the stored partition grew with the dimension on this family: at five states the check at zero disturbance is unresolved at 960,000 bound calls with no refutation, and the stored partition reaches 875.4 MiB at four states.

The released four-state cart-pole certificate of Li et al. (2025b) is checked under the protocol on a box of $1/32$ of the released half-widths, where the base check certifies under their rate-free condition and under ours at rate 0.05. Under ours the certified level rises with the disturbance size on the force channel, from 0.008172 at six newtons to 0.008686 at twelve, the level certified at six being refuted at twelve by validated points, and at eighteen newtons the verifier refutes the level 0.25 from stored points before any bound call, more than 28 times the level certified at twelve (Table 32, Appendix D.15). Over 15 checks at three sizes the shared partition uses 1.40 to 1.82 times fewer bound calls than independent verification under either condition, with no weaker answer. One check at zero disturbance, between the largest refuted and the smallest certified level, ends unresolved at the whole budget for both methods. The reduced box contains no sublevel set of the certificate above

the lower of the two released levels, so what is verified there is the decrease condition on the box, without an invariant set. On the released box the released pair of levels is refuted by validated points and the narrower pairs tested remain unresolved within the tested budget (Appendix D.13).

6 Related work

Incremental and shared-certificate verification. Within one check, branch and bound already passes bounds from parent to child, as the configuration released with Li et al. (2025b) does, and the reuse studied here is across checks at different parameter values. IVAN restarts from the branch-and-bound tree of a finished run after the weights change, with the property and the input region held fixed (Ugare et al., 2023, Def. 4.3, Sec. 4.2), and proof transfer carries intermediate-layer templates to quantized and pruned approximations (Ugare et al., 2022). Shared certificates hold the network fixed and vary the specification under a containment test between a newly propagated abstraction and the stored one (Fischer et al., 2022), with later work reporting subsumption rates that vary widely across scenarios (Das et al., 2026). Learned conflicts reuse infeasible combinations of activation phases when the second check refines the first (Elsaleh et al., 2026, Thm. 3.1), and Two-Stage moves the two levels of its condition during a run and keeps the subdomains already verified, which is sound because the move only removes counterexamples (Li et al., 2025b, Sec. 3.2). The specification-side arguments point from a check to a refinement of it, while a sweep over disturbance sizes with a level search inside it moves in both directions and needs records that stay valid either way. Section 5.2 measures the level-moving reuse of Li et al. (2025b) on these instances in a reimplementation.

Quantitative disturbance verification in control. Putting a disturbance bound inside a verified condition is established: the discrete-time Theorem 7 of Li et al. (2026, Sec. IV-D) does it at one bound, with no sweep across bounds in the section that states it, and Wang et al. (2020) certify constraint satisfaction for a neural closed loop under a perturbation bounded in ℓ_∞ norm, again at one prescribed bound. The neural pipelines we build on return a level set for an undisturbed or fixed-disturbance loop (Yang et al., 2024; Li et al., 2025b), certified training with branch and bound reuses its training-time subregions at test time (Shi et al., 2026), and discrete-time verified certificates cover incremental input-to-state stability (Basu et al., 2025), string stability (Zhou et al., 2025) and contraction metrics (Li et al., 2025a).

Certified envelopes over a scalar axis. Enclosing a scalar function between certified envelopes built from finitely many evaluations, and refining where the enclosure is widest, is the construction of Shubert (1972) for a known Lipschitz modulus, and it is the baseline of Section 5.3 with a validated modulus in place of an assumed one. Branch and bound maintaining two-sided bounds on one scalar functional is established for a fixed specification (Boetius et al., 2025). Parameterized abstract interpretation (Huang et al., 2026) parameterizes the relaxation itself, with the tightness parameters of affine bounds on attention products optimized per instance, whereas the parameter here belongs to the specification and the bound is certified as a function of it. The value function (5) has the shape of the flexibility index of Swaney & Grossmann (1985), which that literature solves by active-set and vertex arguments in parameter space rather than by a refined partition of the state space.

7 Conclusion

For a fixed neural certificate the residual splits into a part that carries the disturbance size linearly and a part that does not, so one adaptive partition of the state region answers a whole sweep with its level searches. On the recorded protocols that partition takes 7.31 to 8.42 and 5.75 to 6.56 times fewer bound calls than independent verification on the two systems, and 2.39 to 15.61 times fewer on released closed loops of prior work, weakening no certified level at the recorded positions of either protocol. The same records, with validated points and at additional cost, give a two-sided enclosure of the threshold, certified between the tested sizes as well.

Limitations. The enclosure is the more expensive output: without a supplied target it takes about twice the sweep's bound calls and comes within a factor $1 + 0.25$ of the sweep's levels on only one of the two instances. Beyond two states the evidence is a synthetic family with up to four states and the released cart-pole certificate on a box of $1/32$ of the released half-widths (Section 5.4).

AI USE STATEMENT

Large language model assistance was used throughout this work for coding and polishing under human direction. Claude Code (Anthropic; models Claude Fable 5.1 as the coordinating session and Claude Opus 5 / Sonnet 5 as task agents).

ETHICS STATEMENT

This work reads the outputs of a formal verifier as certified disturbance-to-state bounds for controlled dynamical systems with neural certificates, demonstrated on a power-system model, an oscillator, and released neural closed loops. Its intended use is to increase the safety and auditability of learned controllers. A certified gain holds only under the stated system model, disturbance class, region, and arithmetic assumptions, and a certificate read beyond them could give operators a false sense of security about a learned controller. In the other direction, a level whose check ended undecided is a statement about the search at its budget, and neither that answer nor a validated refutation is evidence that the controlled system is unsafe. The same guarantees would also strengthen controllers deployed for harmful ends.

REFERENCES

Matthias Althoff, Olaf Stursberg, and Martin Buss. Reachability analysis of nonlinear systems with uncertain parameters using conservative linearization. In *47th IEEE Conference on Decision and Control (CDC)*, pp. 4042–4048, 2008. doi: 10.1109/CDC.2008.4738704.

Ahan Basu, Bhabani Shankar Dey, and Pushpak Jagtap. Formally verified neural Lyapunov function for incremental input-to-state stability of unknown systems. *arXiv preprint arXiv:2501.05778*, 2025.

David Boetius, Stefan Leue, and Tobias Sutter. Solving probabilistic verification problems of neural networks using branch and bound. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*, pp. 4660–4699, 2025.

Ya-Chien Chang, Nima Roohi, and Sicun Gao. Neural Lyapunov control. In *Advances in Neural Information Processing Systems*, 2019.

Kanak Das, Shubham Ugare, Bor-Yuh Evan Chang, Sasa Misailovic, Gagandeep Singh, and Manu Sridharan. Uncovering the limits of proof sharing for neural networks. *arXiv preprint arXiv:2608.19351*, 2026.

Raya Elsaleh, Liam Davis, Haoze Wu, and Guy Katz. Incremental neural network verification via learned conflicts. *arXiv preprint arXiv:2603.12232*, 2026.

Marc Fischer, Christian Sprecher, Dimitar Iliev Dimitrov, Gagandeep Singh, and Martin Vechev. Shared certificates for neural network verification. In *Computer Aided Verification (CAV)*, volume 13371 of *Lecture Notes in Computer Science*, pp. 127–148, 2022. doi: 10.1007/978-3-031-13185-1_7.

Pascal Gahinet, Pierre Apkarian, and Mahmoud Chilali. Affine parameter-dependent Lyapunov functions and real parametric uncertainty. *IEEE Transactions on Automatic Control*, 41(3):436–442, 1996. doi: 10.1109/9.486646.

Sicun Gao, Jeremy Avigad, and Edmund M. Clarke. δ-complete decision procedures for satisfiability over the reals. In *International Joint Conference on Automated Reasoning*, pp. 286–300, 2012.

Pei Huang, Dennis Wei, Omri Isac, Haoze Wu, Min Wu, and Clark Barrett. Parameterized abstract interpretation for transformer verification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(42):35500–35508, 2026. doi: 10.1609/aaai.v40i42.40860.

Hiroyuki Ichihara. Sum of squares based input-to-state stability analysis of polynomial nonlinear systems. *SICE Journal of Control, Measurement, and System Integration*, 5(4):218–225, 2012. doi: 10.9746/jcmsi.5.218.

Hassan K. Khalil. *Nonlinear Systems*. Prentice Hall, Upper Saddle River, NJ, 3 edition, 2002.

Haoyu Li, Xiangru Zhong, Bin Hu, and Huan Zhang. Neural contraction metrics with formal guarantees for discrete-time nonlinear dynamical systems. In *Proceedings of the 7th Annual Learning for Dynamics and Control Conference (L4DC)*, volume 283 of *Proceedings of Machine Learning Research*, 2025a.

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)*, 2025b.

Haoyu Li, Xiangru Zhong, Hao Cheng, Bin Hu, and Huan Zhang. Bridging control with neural network verifier α, β-crown: A tutorial. In *American Control Conference (ACC), tutorial session*, 2026. arXiv:2605.26577.

Huijuan Li, Robert Baier, Lars Grüne, Sigurdur F. Hafstein, and Fabian R. Wirth. Computation of local ISS Lyapunov functions with low gains via linear programming. *Discrete and Continuous Dynamical Systems - B*, 20(8):2477–2495, 2015. doi: 10.3934/dcdsb.2015.20.2477.

Pierre-Jean Meyer. Reachability analysis of neural networks with uncertain parameters. *arXiv preprint arXiv:2303.07917*, 2023.

Ramon E. Moore, R. Baker Kearfott, and Michael J. Cloud. *Introduction to Interval Analysis*. SIAM, 2009.

R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

Zhouxing Shi, Haoyu Li, Cho-Jui Hsieh, and Huan Zhang. Certified training with branch-and-bound for Lyapunov-stable neural control. In *Proceedings of the Annual Learning for Dynamics and Control Conference (L4DC)*. PMLR, 2026.

Bruno O. Shubert. A sequential method seeking the global maximum of a function. *SIAM Journal on Numerical Analysis*, 9(3):379–388, 1972. doi: 10.1137/0709036.

Eduardo D. Sontag. Smooth stabilization implies coprime factorization. *IEEE Transactions on Automatic Control*, 34(4):435–443, 1989.

Eduardo D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Springer, 2nd edition, 1998.

Eduardo D. Sontag and Yuan Wang. New characterizations of input-to-state stability. *IEEE Transactions on Automatic Control*, 41(9):1283–1294, 1996.

Ross E. Swaney and Ignacio E. Grossmann. An index for operational flexibility in chemical process design. part I: Formulation and theory. *AIChE Journal*, 31(4):621–630, 1985. doi: 10.1002/aic.690310412.

Shubham Ugare, Gagandeep Singh, and Sasa Misailovic. Proof transfer for fast certification of multiple approximate neural networks. *Proceedings of the ACM on Programming Languages*, 6 (OOPSLA1):1–29, 2022. doi: 10.1145/3527319.

Shubham Ugare, Debangshu Banerjee, Sasa Misailovic, and Gagandeep Singh. Incremental verification of neural networks. *Proceedings of the ACM on Programming Languages*, 7(PLDI): 1920–1945, 2023. doi: 10.1145/3591299.

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 complete and incomplete neural network robustness verification. In *Advances in Neural Information Processing Systems*, 2021.

Yuh-Shyang Wang, Lily Weng, and Luca Daniel. Neural network control policy verification with persistent adversarial perturbation. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 10050–10059, 2020.

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*, 2020.

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*, 2021.

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 *International Conference on Machine Learning*, 2024.

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*, pp. 4944–4953, 2018.

Jingyuan Zhou, Haoze Wu, Haokun Yu, and Kaidi Yang. Scalable synthesis and verification of string stable neural certificates for interconnected systems. *arXiv preprint arXiv:2509.10118*, 2025.

A GUARANTEES FROM A CERTIFIED LEVEL

This appendix states and interprets the results that convert one certified level into a trajectory guarantee. Section 2 uses them; the proofs are in Appendix B.

A.1 THE LEVEL-PARAMETERIZED CHECK

The dissipation check of Section 2 takes a disturbance size $\bar{w}$ and a level c and asks whether the worst-case residual (3), in which the disturbance enters through the closed-form support function $\zeta_{\bar{w}}$, satisfies $\rho_{\bar{w}} \le \kappa c$ on the region between the levels c and $c_{\max}$. Near the equilibrium ∇V vanishes linearly, so $\zeta_{\bar{w}}$ is of first order in $\|x\|$ while the drift term and κV are of second order, and the residual is positive at points arbitrarily close to the equilibrium in every direction along which $\zeta_{\bar{w}}$ does not vanish. The check at level zero, decrease on the whole domain, therefore fails at every positive size. Above a level c the condition can hold, since the allowance κc on the right-hand side is available where $V \ge c$, and a positive answer at level c says that the excess of V over c decreases at rate κ on the region between the levels. The verifier bisects c at fixed $\bar{w}$ and $c_{\max}$, keeps the smallest tested level at which the check passes, and repeats this at each node of a grid of sizes. One boundary check, that V exceeds $c_{\max}$ on the boundary of the domain, makes the sublevel set at $c_{\max}$ the region from which the guarantees hold. The outputs of these finitely many checks form the object of Definition 1.

Definition 1 (Certified gain triple)**.** Let the system, domain, and disturbance class be as above and let V satisfy (2). A *certified gain triple* is a tuple $\mathcal{G} = (\gamma, \bar{w}_{\mathrm{cert}}, X_{\mathrm{cert}})$ in which $X_{\mathrm{cert}} = \mathcal{A}_{c_{\max}}$ for a level $c_{\max}$ satisfying the *verified boundary separation* $\inf_{\partial\mathcal{X}} V > c_{\max}$ (a verified computation); $\bar{w}_{\mathrm{cert}} = \bar{w}_N$ is the largest node of a grid $0 = \bar{w}_0 < \cdots < \bar{w}_N$ at which dissipation conditions sharing the level $c_{\max}$ have verified levels $c^*(\bar{w}_0) \le \cdots \le c^*(\bar{w}_N)$; and $\gamma\colon [0, \bar{w}_{\mathrm{cert}}] \to \mathbb{R}_{\ge 0}$ is the *gain function*, the left-continuous step interpolant through the node gains,

$$\gamma(\bar{w}) \;=\; \sqrt{c^*(\bar{w}_{i(\bar{w})})/\lambda_P}, \qquad i(\bar{w}) \coloneqq \min\{i : \bar{w}_i \ge \bar{w}\}, \tag{8}$$

where $\sqrt{c/\lambda_P} = \alpha_1^{-1}(c)$ for $\alpha_1(r) \coloneqq \lambda_P r^2$. We write $\rho(c)$ for the level-to-norm conversion, so that $\gamma(\bar{w}) = \rho(c^*(\bar{w}_{i(\bar{w})}))$; ρ is nondecreasing and $\mathcal{A}_c \subseteq \{\|x\| \le \rho(c)\}$ for every level c at which it is evaluated.

The gain gap. We report certificate quality as the *gain gap*

$$\mathrm{GG}(\bar{w}) \;\coloneqq\; \frac{\gamma(\bar{w})}{\gamma_{\mathrm{sim}}(\bar{w})}, \tag{9}$$

with γ_{sim} the largest ultimate deviation over a suite of worst-case disturbance simulations, a lower bound on the worst case, so that GG is a one-sided measure of the conservatism of the certificate and a value below one indicates an error in the simulation or in the certificate. Algorithm 1 (Appendix C.1) summarizes the pipeline.

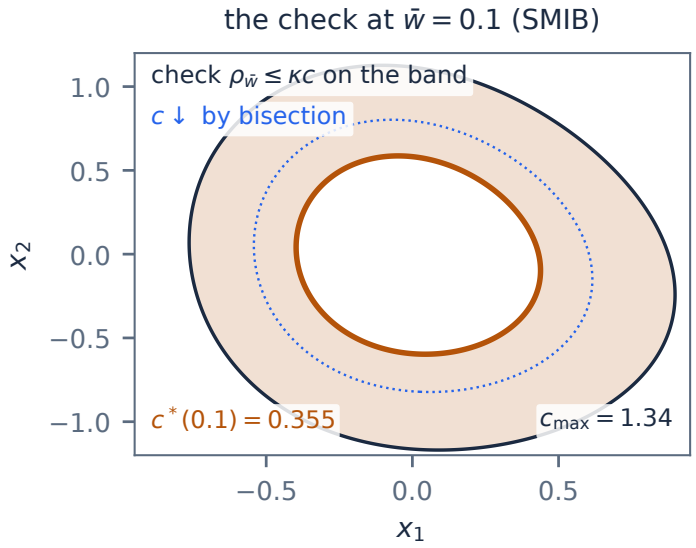


Figure 2: The check behind one point of the curve (SMIB, disturbance size $\bar{w} = 0.1$): the verifier checks $\rho_{\bar{w}} \le \kappa c$ on the region between the levels c and $c_{\max}$ (shaded), bisection lowers c until the check fails, and the last certified level $c^*(\bar{w})$ gives the point $\rho(c^*(\bar{w}))$ of the gain curve. The level reading of Section 5 repeats this check at every node of the size grid.

A.2 What a yes entitles you to

The standing assumptions (stated formally as Assumptions 1–4 in Appendix B) are: locally Lipschitz f_0, constant B, measurable disturbances valued in the compact convex $\mathcal{W}(\bar{w})$ almost everywhere; $V \in C^1$ with $V(0) = 0$ and the quadratic lower bound $V \ge x^\top P x$ on $\mathcal{X}$, with $\lambda_P = \lambda_{\min}(P) > 0$; the verified dissipation

$$\rho_{\bar{w}}(x) \;\le\; \kappa\, c \qquad \text{for all } x \in \mathcal{S}(c, c_{\max}); \tag{10}$$

and the verified strict boundary separation $\inf_{\partial\mathcal{X}} V > c_{\max}$.

Theorem 2 (Regional ISS with explicit constants)**.** *Let Assumptions 1–4 hold and let w be any admissible disturbance at level $\bar{w}$. Then for every $x_0 \in \mathcal{A}_{c_{\max}}$:*

1. ***(Existence, invariance)*** *the Carathéodory solution is unique, global, and stays in $\mathcal{A}_{c_{\max}}$ (never leaving $\mathcal{X}$);*
2. ***(Envelope, entry)*** *for all $t \ge 0$,*

$$V(x(t)) \le c + \big(V(x_0) - c\big)^+ e^{-\kappa t}, \tag{11}$$

every $\mathcal{A}_{c'}$, $c' \in [c, c_{\max}]$, is robustly invariant, and entry into $\mathcal{A}_{c'}$ occurs by

$$T(c') = \kappa^{-1} \ln\big((c_{\max} - c)/(c' - c)\big), \qquad c' \in (c, c_{\max}]; \tag{12}$$

3. ***(Ultimate bound)*** $\limsup_t V(x(t)) \le c$, *hence*

$$\limsup_{t\to\infty} \|x(t)\| \le \alpha_1^{-1}(c) = \sqrt{c/\lambda_P}. \tag{13}$$

4. ***(State-space estimate)*** *if in addition $V(x) \le \Lambda_V \|x\|^2$ on $\mathcal{X}$ for some $\Lambda_V \ge \lambda_P$, then for all $t \ge 0$,*

$$\|x(t)\| \;\le\; \sqrt{\Lambda_V/\lambda_P}\; e^{-\kappa t/2}\, \|x_0\| + \sqrt{c/\lambda_P}. \tag{14}$$

All bounds are uniform over admissible w and $x_0 \in \mathcal{A}_{c_{\max}}$.

Corollary 1 (Zero node and class-$\mathcal{K}$ gain)**.** *Let f_0 be C^2 near the origin with Hurwitz Jacobian A, let $S \succ 0$ solve $A^\top S + SA = -I$, and let $M_2(r)$, nondecreasing in r, satisfy $\|f_0(x) - Ax\| \le M_2(r)\|x\|^2$ for $\|x\| \le r$. Let $r_0 > 0$ satisfy $2\|S\|M_2(r_0)r_0 \le \frac{1}{2}$ and set $c_{\mathrm{loc}} = \lambda_P \lambda_{\min}(S) r_0^2 / \lambda_{\max}(S)$, with the disturbance class normalized so that $\sup_{v\in\mathcal{W}_1}\|v\| \le 1$. If the zero node is verified at a level $c_\varepsilon < c_{\mathrm{loc}}$, then under $w = 0$ every trajectory from X_{cert} converges exponentially to the origin, so the zero node gain may be replaced by $\gamma(0) = 0$. If in addition the first positive node $\bar{w}_1$ has certified level $c^*(\bar{w}_1) < c_{\mathrm{loc}}$, then $\limsup_t \|x(t)\| \le K\bar{w}$ for every trajectory from X_{cert}, every admissible disturbance of magnitude at most $\bar{w}$, and all $\bar{w} \le \bar{w}_{\mathrm{lin}} := \min\{\bar{w}_1, \bar{w}_{\mathrm{loc}}\}$, with K and $\bar{w}_{\mathrm{loc}}$ explicit in S, B, and r_0 (Appendix B); and if moreover $\bar{w}_{\mathrm{loc}} \ge \bar{w}_1$, so that $\bar{w}_{\mathrm{lin}} = \bar{w}_1$, and $\rho(c^*(\bar{w}_2)) \ge K\bar{w}_1$, then for every $\epsilon > 0$ the function $g(\bar{w}) + \epsilon\bar{w}$, with g piecewise linear through $(0,0)$, $(\bar{w}_1, \rho(c^*(\bar{w}_2)))$, and $(\bar{w}_i, \rho(c^*(\bar{w}_{i+1})))$ for $i \ge 2$, extended with constant value beyond its last node, is continuous, strictly increasing, zero at zero, and dominates the certified bound, as refined by $K\bar{w}$ on $[0, \bar{w}_{\mathrm{lin}}]$, at every $\bar{w} \in [0, \bar{w}_{\mathrm{cert}}]$: a class-$\mathcal{K}$ gain.*

In words, Theorem 2 says that a level certified at size $\bar{w}$ is an ultimate bound on V. From every start in X_{cert} and under every admissible disturbance of that size, the trajectory stays in X_{cert}, the excess of V over c decays at rate κ, and the distance to the equilibrium is bounded by $\sqrt{c/\lambda_P}$ in the limit, with the transient bounded by (14) when the upper quadratic bound is available. Corollary 1 says that when the zero node is verified below the level c_{loc} of a Lyapunov ball of the linearization, the undisturbed system converges to the equilibrium from all of X_{cert}, and that a certified segment to the first positive node gives a bound linear in $\bar{w}$ on the limiting distance at small sizes. Two curves follow from these statements.

The step curve γ_{raw} of Definition 1 takes the value $\rho(c^*(\bar{w}_i))$ on each interval $(\bar{w}_{i-1}, \bar{w}_i]$ of the disturbance magnitude grid and is positive at zero, since the zero node is verified at a positive level c_ε. The curve γ_{ISS} is a class-$\mathcal{K}$ majorant of Corollary 1: it is zero at zero, continuous and strictly increasing, dominates the certified bound, as refined by $K\bar{w}$ on $[0, \bar{w}_{\mathrm{lin}}]$, at every positive magnitude in the certified range, and is defined only when the hypotheses of the corollary hold. With γ_{ISS} and the upper quadratic bound, Theorem 2 is regional input-to-state stability in the $\beta + \gamma$ form of part 4 of that theorem at magnitudes at which γ_{ISS} dominates γ_{raw}, and in the ultimate bound of part 3 of that theorem on $[0, \bar{w}_{\mathrm{lin}}]$, and with γ_{raw} in place of γ_{ISS} it is regional practical input-to-state stability. Appendix C states for each table which of the two curves it shows.

Each check returns one of four answers, and only two of them say anything about a level. Certified means that every leaf of a cover of the region between the levels is resolved, either by an enclosure of the residual on it that stays below κc or by its provably containing no point of that region, so the dissipation hypothesis of Theorem 2 holds at that level. Counterexample means a validated refutation, either a box inside the region between the levels whose enclosure refutes the condition at every point of it or a point re-checked through the verifier's own enclosure on a degenerate box. Unresolved means the budget or the width threshold was reached with boxes left undecided and no such refuting box, which says nothing about the level. Error records a failure of the computation rather than a verification outcome. A bisection therefore returns three levels: the smallest level the run certified, $c^*(\bar{w})$, the largest refuted level $c_{\mathrm{rej}}(\bar{w})$, and the largest undecided level $c_{\mathrm{unres}}(\bar{w})$. Lemma 1 sorts these answers against the infimum $c_{\mathrm{inf}}(\bar{w})$ of Corollary 2: a certificate gives $c_{\mathrm{inf}}(\bar{w}) \le c^*(\bar{w})$, a validated refutation gives the strict bound $c_{\mathrm{rej}}(\bar{w}) < c_{\mathrm{inf}}(\bar{w})$, and an undecided level gives neither. The half-open interval $(\max\{c_{\mathrm{rej}}, c_{\mathrm{unres}}\}, c^*]$, reported in the released records and figures, is the interval a run leaves open, and its lower endpoint is labelled by its status: a refutation bounds c_{inf} from below, an exhausted budget marks where the search stopped.

A.3 What a no entitles you to

Branch-and-bound returns the unresolved subdomains (the *residue* $\overline{\mathcal{V}}$), an over-approximation of where the dissipation condition may fail. When the genuine violations lie above a separating level c_{f}, a dissipation check on the sub-region $\mathcal{S}(c_{\mathrm{low}}, c_{\mathrm{f}})$ together with boundary separation makes $\mathcal{A}_{c_{\mathrm{f}}}$ forward invariant, so trajectories from $\mathcal{A}_{c_{\mathrm{f}}}$ never meet the violation set $N(c_{\mathrm{low}})$. For a level c and upper level c_{top}, that set is

$$N(c) \;:=\; \big\{x \in \mathcal{S}(c, c_{\mathrm{top}}) : \rho_{\bar{w}}(x) > \kappa\, c\big\}. \tag{15}$$

Theorem 3 (Recovery below a separating level)**.** *Let Assumptions 1 and 2 hold, let* $0 < c_{\mathrm{low}} < c_{\mathrm{f}}$*, and suppose the verifier has established:*

1. ***(shell dissipation)*** $\rho_{\bar{w}}(x) \le \kappa\, c_{\mathrm{low}}$ *for all* $x \in \mathcal{S}(c_{\mathrm{low}}, c_{\mathrm{f}})$*;*
2. ***(boundary separation)*** $\inf_{x \in \partial\mathcal{X}} V(x) > c_{\mathrm{f}}$.

Fix any upper level $c_{\mathrm{top}} \ge c_{\mathrm{f}}$ *and define the violation set at the lower level,* $N(c_{\mathrm{low}}) \coloneqq \{x \in \mathcal{S}(c_{\mathrm{low}}, c_{\mathrm{top}}) : \rho_{\bar{w}}(x) > \kappa\, c_{\mathrm{low}}\}$*, as in* (15)*. Then for every admissible disturbance at level* $\bar{w}$ *and every* $x_0 \in \mathcal{A}_{c_{\mathrm{f}}}$*, the solution exists on* $[0, \infty)$*, stays in* $\mathcal{A}_{c_{\mathrm{f}}}$*, never intersects* $N(c_{\mathrm{low}})$*, and satisfies the full set of conclusions of Theorem 2 with* $(c, c_{\max})$ *replaced by* $(c_{\mathrm{low}}, c_{\mathrm{f}})$*; in particular*

$$\limsup_{t \to \infty} \|x(t)\| \;\le\; \sqrt{c_{\mathrm{low}}/\lambda_P}, \tag{16}$$

regardless of any violation of the dissipation condition on $\{V > c_{\mathrm{f}}\}$.

Proposition 2 (Impossibility of certification at an accumulating level)**.** *Fix* $\bar{w}$*, an upper level* c_{top}*, and a level* $c \in (0, c_{\mathrm{top}})$ *at which violations accumulate:* $\inf\{V(x) : x \in N(c)\} = c$*, with* $N(c)$ *the violation set* (15)*. Then every sub-region* $\mathcal{S}(c, c_{\mathrm{f}})$ *with* $c < c_{\mathrm{f}} \le c_{\mathrm{top}}$ *contains a point at which the dissipation condition at level* c *fails, so no sound verifier certifies any sub-region at level* c *at any budget. If this holds at every level in* $(0, c_{\mathrm{top}})$*, no regional guarantee with ultimate level below* c_{top} *is obtainable for this certificate through Theorem 2 or Theorem 3.*

A negative answer returns the residue at the level tested, and each box in it falls into one of two classes. A box is genuine when a floating-point evaluation of $\rho_{\bar{w}}$ at a point of the box within the region tested exceeds κc by more than the arithmetic margin. Such a point is evidence of a violation rather than a validated refutation, unless it is re-checked through the verifier's own enclosure on a degenerate box, and Section C.6 states which of the two the counts reported there rest on. A box is unknown when no such point is found and the enclosure of the residual over the box is too wide to decide whether $\rho_{\bar{w}} > \kappa c$. The two classes respond differently to the budget. Unknown boxes shrink as branch and bound splits them further, and a strictly positive margin of the condition on a sub-region is the standard hypothesis under which subdivision is known to terminate (Gao et al., 2012; Moore et al., 2009). Genuine boxes confirmed through the verifier's enclosure persist at every budget, since the verifier's enclosure on the degenerate box establishes $\rho_{\bar{w}} > \kappa c$ at that point. The minimum of V over the genuine boxes locates the evidence found. When it exceeds the lower level c_{low}, a separating level c_{f} is chosen between the two, the shell dissipation check of Theorem 3 is run on the region between c_{low} and c_{f} together with the boundary check at c_{f}, and a positive answer gives the conclusions of Theorem 2 for starts in $\mathcal{A}_{c_{\mathrm{f}}}$ with ultimate level c_{low}. When genuine boxes appear at every level tested down to c_{low}, no separating level is found among those tested. If violations accumulate at c_{low} itself, the hypothesis of Proposition 2 holds and no sound verifier certifies a sub-region at that level at any budget. Section C.6 shows both outcomes on the primary system.

B PROOFS

Assumption 1 (System regularity)**.** *f_0 is locally Lipschitz on an open set containing the compact domain $\mathcal{X}$; B is constant; disturbances are measurable with values in $\mathcal{W}(\bar{w})$ almost everywhere, $\mathcal{W}(\bar{w})$ compact and convex.*

Assumption 2 (Quadratic lower bound)**.** *$V \in C^1(\mathbb{R}^n, \mathbb{R}_{\ge 0})$, $V(0) = 0$, and $V(x) \ge x^\top P x$ on $\mathcal{X}$ with $P \succ 0$, $\lambda_P = \lambda_{\min}(P)$.*

Assumption 3 (Verified dissipation)**.** *For given levels $0 < c < c_{\max}$ and disturbance magnitude $\bar{w}$,*

$$\rho_{\bar{w}}(x) \;\le\; \kappa\, c \qquad \text{for all } x \in \mathcal{S}(c, c_{\max}) \quad \text{(equation 4)}.$$

Assumption 4 (Verified boundary separation)**.** $\inf_{x \in \partial\mathcal{X}} V(x) \;>\; c_{\max}$.

Theorem 2 (restated). Let Assumptions 1–4 hold and let w be any admissible disturbance at level $\bar{w}$. Then for every $x_0 \in \mathcal{A}_{c_{\max}}$:

1. **(Existence, invariance)** the Carathéodory solution is unique, global, and stays in $\mathcal{A}_{c_{\max}}$ (never leaving $\mathcal{X}$);

2. **(Envelope, entry)** for all $t \geq 0$,

$$V(x(t)) \leq c + \big(V(x_0) - c\big)^+ e^{-\kappa t},$$

every $\mathcal{A}_{c'}$, $c' \in [c, c_{\max}]$, is robustly invariant, and entry into $\mathcal{A}_{c'}$ occurs by

$$T(c') = \kappa^{-1} \ln\big((c_{\max} - c)/(c' - c)\big), \qquad c' \in (c, c_{\max}];$$

3. **(Ultimate bound)** $\limsup_t V(x(t)) \leq c$, hence

$$\limsup_{t\to\infty} \|x(t)\| \leq \alpha_1^{-1}(c) = \sqrt{c/\lambda_P}.$$

4. **(State-space estimate)** if in addition $V(x) \leq \Lambda_V \|x\|^2$ on $\mathcal{X}$ for some $\Lambda_V \geq \lambda_P$, then for all $t \geq 0$,

$$\|x(t)\| \;\leq\; \sqrt{\Lambda_V/\lambda_P}\, e^{-\kappa t/2}\, \|x_0\| + \sqrt{c/\lambda_P}.$$

All bounds are uniform over admissible w and $x_0 \in \mathcal{A}_{c_{\max}}$.

Proof of Theorem 2. Throughout this proof, write $A \coloneqq \mathcal{A}_{c_{\max}}$ and let

$$r(x) \;\coloneqq\; \langle \nabla V(x), f_0(x)\rangle + \zeta_{\bar{w}}(x),$$

so that $\rho_{\bar{w}}(x) = r(x) + \kappa V(x)$ and (4) reads $r(x) \leq \kappa(c - V(x))$ on $\mathcal{S}(c, c_{\max})$. Since ∇V and f_0 are continuous and $y \mapsto \sup_{w\in\mathcal{W}(\bar{w})}\langle y, Bw\rangle$ is convex and finite, hence continuous, r is continuous on $\mathcal{X}$.

Step 0 (solutions and the chain rule). The right-hand side $g(t, x) = f_0(x) + Bw(t)$ is measurable in t for fixed x, locally Lipschitz in x with a Lipschitz constant and a linear growth bound that are uniform in t because w is essentially bounded. By the Carathéodory existence and uniqueness theorem (Sontag, 1998, Appendix C) a unique maximal solution exists on some interval $[0, T_{\max})$, it is absolutely continuous, and $t \mapsto V(x(t))$ is absolutely continuous with

$$\tfrac{d}{dt} V(x(t)) \;=\; \langle \nabla V(x(t)), f_0(x(t)) + Bw(t)\rangle \;\leq\; r(x(t)) \qquad \text{for a.e. } t, \tag{17}$$

where the inequality inserts the pointwise bound $\langle \nabla V(x), Bw(t)\rangle \leq \zeta_{\bar{w}}(x)$, valid for a.e. t because $w(t) \in \mathcal{W}(\bar{w})$ a.e. (Proposition 3).

Step 1 (the trajectory cannot cross the level $c_{\max}$ while in $\mathcal{X}$). Let $J \coloneqq \{t \in [0, T_{\max}) : x(s) \in \mathcal{X} \text{ for all } s \in [0, t]\}$, an interval containing 0. We claim $V(x(t)) \leq c_{\max}$ for all $t \in J$. Suppose not, and set $t_1 \coloneqq \inf\{t \in J : V(x(t)) > c_{\max}\}$. By continuity $V(x(t_1)) = c_{\max}$ and $x(t_1) \in \mathcal{S}(c, c_{\max})$, so (4) gives $r(x(t_1)) \leq \kappa(c - c_{\max}) < 0$. By continuity of r and of $x(\cdot)$ there is $\delta > 0$ with $r(x(s)) \leq \frac{\kappa}{2}(c - c_{\max}) < 0$ for all $s \in [t_1, t_1 + \delta] \cap J$. Integrating (17),

$$V(x(t_1 + h)) \;\leq\; V(x(t_1)) + \int_{t_1}^{t_1+h} r(x(s))\, ds \;\leq\; c_{\max} + h\,\tfrac{\kappa}{2}(c - c_{\max}) \;<\; c_{\max}$$

for $0 < h \leq \delta$, contradicting the definition of t_1. Note that this argument uses continuity of r near the crossing point and the strict negativity $r \leq \kappa(c - c_{\max}) < 0$ on the level set $\{V = c_{\max}\} \cap \mathcal{X}$; no information about V above $c_{\max}$ is needed.

Step 2 (invariance of A, no escape from $\mathcal{X}$, global existence). By Assumption 4 the compact set $A = \{x \in \mathcal{X} : V(x) \leq c_{\max}\}$ satisfies $A \cap \partial\mathcal{X} = \emptyset$, so $\delta_0 \coloneqq \operatorname{dist}(A, \partial\mathcal{X}) > 0$. On J we have $x(t) \in \mathcal{X}$ and, by Step 1, $V(x(t)) \leq c_{\max}$, hence $x(t) \in A$ and $\operatorname{dist}(x(t), \partial\mathcal{X}) \geq \delta_0$. If $\sup J < T_{\max}$, continuity of $x(\cdot)$ would force $x(\sup J) \in \partial\mathcal{X}$ while also $\operatorname{dist}(x(\sup J), \partial\mathcal{X}) \geq \delta_0$, a contradiction; therefore $J = [0, T_{\max})$ and $x(t) \in A$ on the whole maximal interval. Since A is compact, the solution does not blow up and $T_{\max} = \infty$ (Sontag, 1998, Appendix C). This proves conclusion 1.

Step 3 (comparison). Define $v(t) \coloneqq \big(V(x(t)) - c\big)^+$, an absolutely continuous function. At a.e. t with $V(x(t)) > c$ we have $x(t) \in \mathcal{S}(c, c_{\max})$ by Steps 1–2, so (17) and (4) give $\dot{v}(t) = \dot{V}(x(t)) \leq \kappa(c - V(x(t))) = -\kappa v(t)$. At a.e. t in $\{s : V(x(s)) \leq c\}$ the a.e. derivative of v vanishes, because v is zero on that set and an absolutely continuous function has zero derivative a.e. on any of its level sets. Hence $\dot{v} \leq -\kappa v$ a.e. on $[0, \infty)$, and Grönwall's inequality for absolutely continuous functions

gives $v(t) \le v(0)e^{-\kappa t}$, which is (11). The entry time in conclusion 2 follows by reading (11) at the time $T(c')$ where the right-hand side equals c', and invariance of each $\mathcal{A}_{c'}$ follows by applying (11) with x_0 replaced by any point of $\mathcal{A}_{c'}$, since $v(0) \le c' - c$ implies $V(x(t)) \le c + (c' - c) = c'$ for all t. Conclusion 3 follows by letting $t \to \infty$ in (11) and then applying (2): $\lambda_P \|x(t)\|^2 \le V(x(t))$ forces $\limsup \|x(t)\| \le \sqrt{c/\lambda_P}$. Uniformity in w and x_0 holds because no step used anything about w beyond admissibility, and the envelope depends on x_0 only through $V(x_0) \le c_{\max}$.

Step 4 (state-space estimate). From (11) and $\lambda_P \|x\|^2 \le V(x) \le \Lambda_V \|x\|^2$ we obtain $\lambda_P \|x(t)\|^2 \le c + \Lambda_V \|x_0\|^2 e^{-\kappa t}$, and $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ gives (14). □

Corollary 2 (Monotone certified levels and gains)**.** *Fix $c_{\max}$ and suppose Assumption 3 holds at $(\bar{w}, c)$. Then it holds at $(\bar{w}', c)$ for every $0 \le \bar{w}' \le \bar{w}$, and also at $(\bar{w}, c'')$ for every $c'' \in [c, c_{\max})$. Consequently the infimum $c_{\inf}(\bar{w}) \coloneqq \inf\{c : (4)$ holds at $(\bar{w}, c)\}$ is nondecreasing in $\bar{w}$, every level returned by a level search is an upper bound of it, and the returned levels $c^*(\bar{w}_i)$ of Definition 1, once normalized by the running minimum from above, are certified and nondecreasing; hence $\gamma(\bar{w}) = \sqrt{c^*(\bar{w}_{i(\bar{w})})/\lambda_P}$ is nondecreasing and the step interpolant of Definition 1 is a sound upper bound for the gain between certified nodes.*

Proof of Corollary 2. The fold is positively homogeneous in the disturbance level: $\zeta_{\bar{w}}(x) = \bar{w}\,\zeta_1(x)$ with $\zeta_1 \ge 0$, since $\mathcal{W}(\bar{w}) = \bar{w}\mathcal{W}_1$ and $0 \in \mathcal{W}_1$. Hence $\rho_{\bar{w}'}(x) \le \rho_{\bar{w}}(x)$ pointwise for $\bar{w}' \le \bar{w}$, and the first claim follows because condition (4) has a right-hand side independent of $\bar{w}$. For the second claim, take $c'' \in [c, c_{\max})$ and $x \in \mathcal{S}(c'', c_{\max}) \subseteq \mathcal{S}(c, c_{\max})$; then $\rho_{\bar{w}}(x) \le \kappa c \le \kappa c''$. Monotonicity of $c_{\inf}(\bar{w})$ is immediate from the first claim; a level search that stops at a certified level returns a value at or above $c_{\inf}(\bar{w})$, the running-minimum normalization keeps each node at a certified level by the first claim (a level certified at a larger $\bar{w}$ is certified at every smaller one), and monotonicity of γ follows because α_1^{-1} is increasing. □

Corollary 3 (Verified sublevel radius)**.** *Let the hypotheses of Theorem 2 hold at $(\bar{w}, c)$, and let $c' > c$ and $R > 0$ be such that*

$$\{x \in \mathcal{X} : V(x) \le c'\} \;\subseteq\; \{x : \|x\| \le R\}. \tag{18}$$

Then $\limsup_{t\to\infty} \|x(t)\| \le R$ for every admissible disturbance at level $\bar{w}$ and every $x_0 \in \mathcal{A}_{c_{\max}}$. The containment (18) *is verified from a finite box partition of $\mathcal{X}$ on which every box whose lower bound of V is at most c' has an upper bound of $\|x\|$ at most R. At any node of Definition 1, $\min\{R, \sqrt{c/\lambda_P}\}$ may replace $\sqrt{c/\lambda_P}$, and radii verified at increasing levels are normalized across nodes by the running minimum from above.*

Proof of Corollary 3. By Theorem 2, $x(t) \in \mathcal{A}_{c_{\max}} \subseteq \mathcal{X}$ for all $t \ge 0$ and $\limsup_t V(x(t)) \le c < c'$, so there is T with $V(x(t)) \le c'$ for all $t \ge T$. By (18), $\|x(t)\| \le R$ for all $t \ge T$. For the verification, let $x \in \mathcal{X}$ with $V(x) \le c'$; x lies in some box Q of the partition, whose lower bound of V is at most $V(x) \le c'$, hence $\|x\| \le \sup_Q \|\cdot\| \le R$. The minimum with the closed form is sound because both bounds hold. For the normalization, a radius R_j verified at a level $c'_j \ge c'_i$ satisfies (18) at c'_i as well, since $\mathcal{A}_{c'_i} \subseteq \mathcal{A}_{c'_j}$ for $c'_i \le c'_j$, so the minimum over nodes at or above i is a valid radius at node i. □

Corollary 1 (restated). Under the hypotheses of Corollary 1: if the zero node is verified at a level $c_\varepsilon < c_{\mathrm{loc}}$, trajectories from X_{cert} under $w = 0$ converge exponentially to the origin, so the zero node gain may be replaced by $\gamma(0) = 0$; if in addition the first positive node $\bar{w}_1$ has certified level $c^*(\bar{w}_1) < c_{\mathrm{loc}}$, then $\limsup_t \|x(t)\| \le K\bar{w}$ for every trajectory from X_{cert}, every admissible disturbance of magnitude at most $\bar{w}$, and all $\bar{w} \le \bar{w}_{\mathrm{lin}} = \min\{\bar{w}_1, \bar{w}_{\mathrm{loc}}\}$, with $K = 4\|SB\|\sqrt{\lambda_{\max}(S)/\lambda_{\min}(S)}$ and $\bar{w}_{\mathrm{loc}} = r_0\sqrt{\lambda_{\min}(S)/\lambda_{\max}(S)}/(4\|SB\|)$; if moreover $\bar{w}_{\mathrm{loc}} \ge \bar{w}_1$, so that $\bar{w}_{\mathrm{lin}} = \bar{w}_1$, and $\rho(c^*(\bar{w}_2)) \ge K\bar{w}_1$, then for every $\epsilon > 0$ the function $g(\bar{w}) + \epsilon\bar{w}$, with g piecewise linear through $(0,0)$, $(\bar{w}_1, \rho(c^*(\bar{w}_2)))$, and $(\bar{w}_i, \rho(c^*(\bar{w}_{i+1})))$ for $i \ge 2$ and extended with constant value beyond its last node, is a class-$\mathcal{K}$ majorant of the certified bound, as refined by $K\bar{w}$ on $[0, \bar{w}_{\mathrm{lin}}]$, at every $\bar{w}$ in $[0, \bar{w}_{\mathrm{cert}}]$. The disturbance class is normalized so that $\sup_{v\in\mathcal{W}_1} \|v\| \le 1$ (the scalar box class used here satisfies this).

Proof of Corollary 1. Let $W(x) = x^\top S x$ and $h(x) = f_0(x) - Ax$. *Local decrease.* For $\|x\| \le r_0$ and $w = 0$,

$$\dot{W} = x^\top (A^\top S + SA)x + 2x^\top S h(x) \;\le\; -\|x\|^2 + 2\|S\| M_2(r_0)\|x\|^3 \;\le\; -\tfrac{1}{2}\|x\|^2,$$

using $2\|S\|M_2(r_0)r_0 \le \frac{1}{2}$. Since $W \le \lambda_{\max}(S)\|x\|^2$, this gives $\dot{W} \le -W/(2\lambda_{\max}(S))$ on the ball, the set $\{W \le c_W\}$ with $c_W = \lambda_{\min}(S) r_0^2$ lies inside the ball and is forward invariant, and W, hence $\|x\|$, decays exponentially from any point of it. *Transfer to* W. For $c' \le c_{\mathrm{loc}}$, (2) gives $\{V \le c'\} \subseteq \{\|x\|^2 \le c'/\lambda_P\} \subseteq \{W \le \lambda_{\max}(S)c'/\lambda_P\} \subseteq \{W \le c_W\}$. Theorem 2 at $(0, c_\varepsilon)$ gives entry of every trajectory from X_{cert} into $\{V \le c'\}$ for any $c' \in (c_\varepsilon, \min\{c_{\mathrm{loc}}, c_{\max}\}]$ within the time (12), after which the local decrease applies; so $x(t) \to 0$ exponentially, uniformly over X_{cert}, and the zero node gain may be replaced by $\gamma(0) = 0$. *Small disturbances.* On $\{W \le c_W\}$, for an admissible w at level $\bar{w}$, $\dot{W} \le -\frac{1}{2}\|x\|^2 + 2\|x\|\|SB\|\bar{w}$ (the disturbance class enters through $2x^\top SBw$ and $\|w\| \le \bar{w}$, which is the normalization $\sup_{v \in \mathcal{W}_1}\|v\| \le 1$ assumed in Corollary 1), which is negative for $\|x\| > \rho_{\mathrm{loc}} := 4\|SB\|\bar{w}$. When $\lambda_{\max}(S)\rho_{\mathrm{loc}}^2 \le c_W$, that is $\bar{w} \le r_0\sqrt{\lambda_{\min}(S)/\lambda_{\max}(S)}/(4\|SB\|) := \bar{w}_{\mathrm{loc}}$, the set $\{W \le \lambda_{\max}(S)\rho_{\mathrm{loc}}^2\}$ lies inside $\{W \le c_W\}$, is forward invariant, and is reached in finite time from every point of $\{W \le c_W\}$, so $\limsup_t \|x(t)\| \le \sqrt{\lambda_{\max}(S)/\lambda_{\min}(S)}\,\rho_{\mathrm{loc}} = K\bar{w}$ with $K := 4\|SB\|\sqrt{\lambda_{\max}(S)/\lambda_{\min}(S)}$ (Khalil, 2002, Theorem 4.18). Entry into $\{W \le c_W\}$ at a level $\bar{w} \le \bar{w}_1$: the dissipation condition certified at $(\bar{w}_1, c^*(\bar{w}_1))$ holds at $(\bar{w}, c^*(\bar{w}_1))$ by Corollary 2, so Theorem 2 gives entry into $\{V \le c'\}$ for any $c' \in (c^*(\bar{w}_1), \min\{c_{\mathrm{loc}}, c_{\max}\}]$, a set inside $\{W \le c_W\}$ by the inclusion of the transfer step since $c^*(\bar{w}_1) < c_{\mathrm{loc}}$. Both restrictions together give $\bar{w}_{\mathrm{lin}} = \min\{\bar{w}_1, \bar{w}_{\mathrm{loc}}\}$. *Majorant.* The certified bound on $(0, \bar{w}_{\mathrm{cert}}]$ is $\min\{K\bar{w}, \gamma(\bar{w})\}$ on $(0, \bar{w}_{\mathrm{lin}}]$ and $\gamma(\bar{w})$ beyond, with γ the step interpolant of Definition 1, whose value on $(\bar{w}_i, \bar{w}_{i+1}]$ is $\rho(c^*(\bar{w}_{i+1}))$. On $(0, \bar{w}_1]$ the linear piece of g has slope $\rho(c^*(\bar{w}_2))/\bar{w}_1 \ge K$, so $g \ge K\bar{w}$ there; on $(\bar{w}_i, \bar{w}_{i+1}]$ with $i \ge 1$, g interpolates between $\rho(c^*(\bar{w}_{i+1}))$ and $\rho(c^*(\bar{w}_{i+2}))$ (or stays at $\rho(c^*(\bar{w}_{\mathrm{cert}}))$ on the last piece), which are nondecreasing, so $g \ge \rho(c^*(\bar{w}_{i+1})) = \gamma(\bar{w})$ there. Hence g dominates the certified bound on $(0, \bar{w}_{\mathrm{lin}}]$ and on $(\bar{w}_1, \bar{w}_{\mathrm{cert}}]$, hence on all of $(0, \bar{w}_{\mathrm{cert}}]$ when $\bar{w}_{\mathrm{loc}} \ge \bar{w}_1$ so that $\bar{w}_{\mathrm{lin}} = \bar{w}_1$; it is continuous, nondecreasing, and zero at zero, and $g + \epsilon\bar{w}$ is in addition strictly increasing, hence of class $\mathcal{K}$ on $[0, \bar{w}_{\mathrm{cert}}]$. □

Remark 1 (Scope of the branch-and-bound certificate). The implementation verifies (4) through the disjunction that a box R in the subdivision is resolved if $\mathrm{UB}_R(\rho_{\bar{w}}) \le \kappa c - \varepsilon_0$, or $\mathrm{LB}_R(V) > c_{\max} + \varepsilon_0$, or $\mathrm{UB}_R(V) < c - \varepsilon_0$, where the bounds are sound interval or linear-relaxation enclosures and the two V-clauses are strict, so that a box meeting $\mathcal{S}(c, c_{\max})$ can never be discarded by them. The arithmetic margin ε_0 is the fixed positive constant of Section 2, of value 10^{-9} in our runs. The margin appears in the V-clauses as well as the residual clause, and in the conservative direction: resolving a box by a V-clause asserts that it contains no point of the region between the levels and so skips its residual check; an unjustified discard is the unsound direction. If every leaf of the subdivision is resolved and the leaves cover $\mathcal{S}(c, c_{\max})$, then (4) holds on $\mathcal{S}(c, c_{\max})$: a leaf resolved by a V-clause contains no point of the region between the levels ($\mathrm{LB}_R(V) > c_{\max} + \varepsilon_0$ excludes $V \le c_{\max}$ and $\mathrm{UB}_R(V) < c - \varepsilon_0$ excludes $V \ge c$), so every leaf containing a point of $\mathcal{S}(c, c_{\max})$ was resolved by the residual clause, whose upper bound then applies. Unresolved leaves are returned as the unverified subdomains (the residue) $\overline{\mathcal{V}}$, which Section A.3 discharges with a second certificate under the hypotheses stated there.

Lemma 1 (Bracket for the certified level)**.** *Fix a disturbance size $\bar{w}$ and an upper level $c_{\max}$, let $c_{\inf}(\bar{w})$ be the infimum of Corollary 2, and run the dissipation check* (4) *at finitely many levels $c \in (0, c_{\max})$ with the enclosures of Remark 1. Every answer falls into one of three cases, a computation that aborts being excluded because it asserts nothing about the level.*

1. ***(Certified)*** *Every leaf of a subdivision covering $\mathcal{S}(c, c_{\max})$ is resolved. Then $c_{\inf}(\bar{w}) \le c$.*
2. ***(Validated refutation)*** *A box $Q \subseteq \mathcal{X}$ is returned, of any width and in particular the degenerate box $\{x\}$ of a re-checked point, whose enclosures satisfy $\mathrm{LB}_Q(V) \ge c + \varepsilon_0$, $\mathrm{UB}_Q(V) \le c_{\max} - \varepsilon_0$, and $\mathrm{LB}_Q(\rho_{\bar{w}}) > \kappa c + \varepsilon_0$. Then, at every $x \in Q$,*

$$c_{\inf}(\bar{w}) \;\ge\; \min\{V(x),\, \rho_{\bar{w}}(x)/\kappa\} \;\ge\; c + \min\{1, \kappa^{-1}\}\,\varepsilon_0 \;>\; c. \tag{19}$$

3. ***(Undecided)*** *The subdivision stops at its box budget or its width threshold with unresolved leaves and no such box. Then neither $c_{\inf}(\bar{w}) \le c$ nor $c_{\inf}(\bar{w}) > c$ follows.*

Write L_+ for the levels answered as in case 1 and L_- for the levels answered as in case 2, both among the levels run. Then

$$\max L_- \;<\; c_{\inf}(\bar{w}) \;\le\; \min L_+, \tag{20}$$

each inequality holding when its set is nonempty. The lower endpoint may be replaced by the largest value of $\min\{\mathrm{LB}_{Q_j}(V), \mathrm{LB}_{Q_j}(\rho_{\bar{w}})/\kappa\}$ *over the boxes Q_j validated as in case 2 at any of the levels run. The largest undecided level is an endpoint of neither side of* (20).

Proof of Lemma 1. Step 1 (certified answers). By Remark 1, a subdivision whose leaves cover $\mathcal{S}(c, c_{\max})$ and are all resolved establishes (4) at $(\bar{w}, c)$. Hence c belongs to the set whose infimum defines $c_{\inf}(\bar{w})$ in Corollary 2, and $c_{\inf}(\bar{w}) \le c$.

Step 2 (validated refutations). Let Q be as in case 2, let $x \in Q$, and put $\mu := \min\{V(x), \rho_{\bar{w}}(x)/\kappa\}$. The enclosures are sound on Q and hold at each of its points, so $c + \varepsilon_0 \le V(x) \le c_{\max} - \varepsilon_0$ and $\rho_{\bar{w}}(x) > \kappa c + \varepsilon_0$, whence $\mu \ge \min\{c + \varepsilon_0, c + \varepsilon_0/\kappa\} = c + \min\{1, \kappa^{-1}\}\,\varepsilon_0$, the middle inequality of (19), and $\mu > c$. Let $0 < c' < \mu$. Then $V(x) \ge \mu > c'$ and $V(x) \le c_{\max}$, so $x \in \mathcal{S}(c', c_{\max})$, while $\rho_{\bar{w}}(x) \ge \kappa\mu > \kappa c'$, so (4) fails at $(\bar{w}, c')$. No level below μ is feasible, and $c_{\inf}(\bar{w}) \ge \mu$. Soundness of the enclosures on Q gives $\mu \ge \min\{\mathrm{LB}_Q(V), \mathrm{LB}_Q(\rho_{\bar{w}})/\kappa\}$ at every $x \in Q$, and the argument uses the box and not the level at which the box was found, so the largest of these enclosure values over the validated boxes is a lower bound as well.

Step 3 (undecided answers). An undecided answer records only that the subdivision stopped with unresolved leaves and no validated point, and both inequalities stay possible. A sound upper bound on $\rho_{\bar{w}} - \kappa c$ over a box of positive width can be positive even though the difference is negative on the whole box, so a level at which (4) holds can end undecided at a finite budget. A level at which (4) fails also ends undecided whenever the search validates no point through the enclosures of case 2. Neither inequality follows from the answer alone, so the largest undecided level is an endpoint of neither side. Collecting Steps 1 and 2 over the levels run gives (20). □

Proof of Theorem 3. Hypotheses 1 and 2 are Assumptions 3 and 4 with $(c, c_{\max}) = (c_{\mathrm{low}}, c_{\mathrm{f}})$, so Theorem 2 applies verbatim: solutions from $\mathcal{A}_{c_{\mathrm{f}}}$ are global, $\mathcal{A}_{c_{\mathrm{f}}}$ is robustly forward invariant, the envelope (11) holds with level c_{low}, and (16) follows from (13). For the discharge claim, hypothesis 1 forces $N(c_{\mathrm{low}}) \cap \mathcal{S}(c_{\mathrm{low}}, c_{\mathrm{f}}) = \emptyset$, and points of $N(c_{\mathrm{low}})$ have $V \ge c_{\mathrm{low}}$ by definition, so $N(c_{\mathrm{low}}) \subseteq \{x \in \mathcal{X} : V(x) > c_{\mathrm{f}}\}$, which is disjoint from the invariant set $\mathcal{A}_{c_{\mathrm{f}}}$ containing the trajectory. □

Lemma 2 (Existence of a separating level)**.** *Fix $0 < c < c_{\mathrm{top}}$ and let $N(c)$ be the violation set* (15)*. A level $c_{\mathrm{f}} \in (c, c_{\mathrm{top}}]$ with $\mathcal{S}(c, c_{\mathrm{f}}) \cap N(c) = \emptyset$ exists if and only if*

$$\inf\{V(x) : x \in N(c)\} \;>\; c \tag{21}$$

(with $\inf \emptyset = +\infty$), and then every $c_{\mathrm{f}} \in (c, c_{\mathrm{top}}]$ with $c_{\mathrm{f}} < \inf V(N(c))$ works.

Proof of Lemma 2. Assume (21) and pick $c_{\mathrm{f}} \in (c, c_{\mathrm{top}}]$ with $c_{\mathrm{f}} < \inf V(N(c))$ (possible since the infimum exceeds c, taking $c_{\mathrm{f}} = c_{\mathrm{top}}$ when $N(c)$ is empty). Any $x \in \mathcal{S}(c, c_{\mathrm{f}})$ has $V(x) \le c_{\mathrm{f}} < \inf V(N(c))$, so $x \notin N(c)$, and $\mathcal{S}(c, c_{\mathrm{f}}) \cap N(c) = \emptyset$. Conversely, assume $\mathcal{S}(c, c_{\mathrm{f}}) \cap N(c) = \emptyset$ for some $c_{\mathrm{f}} > c$. Any $x \in N(c)$ satisfies $c \le V(x) \le c_{\mathrm{top}}$ and, by the assumed disjointness, $x \notin \mathcal{S}(c, c_{\mathrm{f}})$, which forces $V(x) > c_{\mathrm{f}}$; taking the infimum over $N(c)$ gives $\inf V(N(c)) \ge c_{\mathrm{f}} > c$. □

Proof of Proposition 2. Fix $c < c_{\mathrm{f}} \le c_{\mathrm{top}}$. Since $\inf V(N(c)) = c < c_{\mathrm{f}}$, some $x \in N(c)$ has $V(x) < c_{\mathrm{f}}$; membership in $N(c)$ gives $V(x) \ge c$ and $\rho_{\bar{w}}(x) > \kappa c$, so x lies in the sub-region and the inequality $\rho_{\bar{w}} \le \kappa c$ is false there. A sound verifier certifies only true inequalities, so the check is refused at every budget. The cited theorems derive their conclusions from a certified sub-region, so at levels where the hypothesis holds neither applies. □

Proposition 3 (Disturbance folding)**.** *Let $\mathcal{W} \subset \mathbb{R}^m$ be compact and convex and let $y\colon \mathcal{X} \to \mathbb{R}^m$ be continuous (in our use, $y(x) = B^\top \nabla V(x)$). Define the support function $\zeta(x) = \sup_{w \in \mathcal{W}} \langle y(x), w \rangle$.*

1. ***(Pointwise validity along measurable signals)*** *For every measurable $w(\cdot)$ with $w(t) \in \mathcal{W}$ a.e. and every absolutely continuous trajectory $x(\cdot)$ in $\mathcal{X}$, $\langle y(x(t)), w(t) \rangle \le \zeta(x(t))$ for a.e. t.*

2. ***(Closed forms)*** *For the box $\mathcal{W} = \bar{w}\,[-1,1]^m$, $\zeta(x) = \bar{w}\|y(x)\|_1$; for the Euclidean ball of radius $\bar{w}$, $\zeta(x) = \bar{w}\|y(x)\|_2$; for a polytope $\mathcal{W} = \mathrm{conv}\{v_1,\dots,v_K\}$, $\zeta(x) = \max_k \langle y(x), v_k\rangle$.*

3. ***(Tightness)*** *For each of the three classes there is an admissible measurable signal attaining the bound along any given trajectory; for the box class it is $w^*(t) = \bar{w}\,\mathrm{sign}(y(x(t)))$ componentwise, with $\langle y(x(t)), w^*(t)\rangle = \zeta(x(t))$ for all t.*

Proof of Proposition 3. Claim 1 is immediate from the definition of the support function at each t where $w(t) \in \mathcal{W}$, which is a.e. by admissibility. Claim 2 is the standard computation of support functions (Rockafellar, 1970, Section 13): for the box, $\sup_{|w_j|\le\bar{w}} \sum_j y_j w_j = \bar{w}\sum_j |y_j|$, attained at $w_j = \bar{w}\,\mathrm{sign}(y_j)$; the ball and polytope cases are analogous. For claim 3, $t \mapsto y(x(t))$ is continuous, the componentwise sign map is Borel measurable, and a composition of a Borel map with a continuous map is measurable; the attainment identity is claim 2's attainment applied pointwise. For the ball class take $w^*(t) = \bar{w}\,y(x(t))/\|y(x(t))\|_2$ where $y(x(t)) \neq 0$ and $w^*(t) = 0$ otherwise; the quotient map is continuous off the closed zero set, so w^* is Borel, and on the zero set both sides of the attainment identity vanish. For a polytope take $w^*(t) = v_{k^*(t)}$ with $k^*(t)$ the least maximizing index of $k \mapsto \langle y(x(t)), v_k\rangle$; each set $\{t : k^*(t) = k\}$ is Borel as a finite intersection of closed conditions minus lower-index ones, so w^* is a measurable selection and attains by construction. Consequently the worst-case residual (3) is attained pointwise: at each state the bound inserted in (17) is met with equality by an admissible value of w, so folding introduces no conservatism at the level of the dissipation inequality. □

Theorem 4 (Two-sided quadratic sandwich)**.** *Let V be of the form* (23) *with weights (P, s_k, a_k, b_k). For every $x \in \mathbb{R}^n$,*

$$\lambda_{\min}(P)\,\|x\|^2 \;\le\; V(x) \;\le\; \Big(\lambda_{\max}(P) + \tfrac{1}{8}\sum_{k=1}^{K} s_k\|a_k\|^2\Big)\|x\|^2. \tag{22}$$

Both constants are closed-form in the weights, so $\alpha_1(r) = \lambda_{\min}(P)r^2$ and $\alpha_2(r) = (\lambda_{\max}(P) + \frac{1}{8}\sum_k s_k\|a_k\|^2)r^2$ are certified comparison functions at zero verification cost.

Proof of Theorem 4. The lower bound restates (2), valid globally here: each $\varphi_k \ge 0$ as a Bregman divergence of a convex function, each $s_k \ge 0$ by the softplus parametrization, and $x^\top P x \ge \lambda_{\min}(P)\|x\|^2$. For the upper bound, Taylor's theorem with integral remainder around b_k gives, with $z = a_k^\top x$,

$$\varphi_k(x) \;=\; \mathrm{sp}(b_k + z) - \mathrm{sp}(b_k) - \mathrm{sp}'(b_k)\,z \;=\; z^2\int_0^1 (1-\tau)\,\mathrm{sp}''(b_k + \tau z)\,d\tau \;\le\; \frac{z^2}{8},$$

because $\mathrm{sp}''(u) = \sigma(u)(1-\sigma(u)) \le \frac{1}{4}$ for all u and $\int_0^1 (1-\tau)\,d\tau = \frac{1}{2}$. By Cauchy–Schwarz, $z^2 \le \|a_k\|^2\|x\|^2$, and summing with weights s_k and adding $x^\top P x \le \lambda_{\max}(P)\|x\|^2$ proves (22). □

C Experimental details

C.1 The CertGain pipeline

Algorithm 1 lists the steps of the pipeline that produces the certified levels this appendix reports.

C.2 Release and checks

The code, the trained certificates, the seeds and configuration files of every training run, and the result files behind every table and figure are released. The constants of Theorem 2 and of the propositions, among them the envelope and entry time (11)–(12) and the sandwich constants of Theorem 4, were checked numerically against the trained certificates and the worst-case simulations. The certificate studies of this appendix predate the two asset generators, so the values quoted in its text are as recorded in the artifact, in the result files under `experiments/*/artifacts/` that each study writes, and not produced by a generator at build time.

Algorithm 1 CertGain: train, certify, and return a certified gain triple

Require: weights θ giving (V, P), rate κ, node grid $0 = \bar{w}_0 < \cdots < \bar{w}_N$, box budget, width threshold

train V of the form (23) by the sampled version of the hinge with the scale constraint and the boundary level, then the certified training phase
certify the upper level $c_{\max}$: verified boundary separation $\inf_{\partial\mathcal{X}} V > c_{\max}$
for each node $\bar{w}_i$ **do**
    bisect on c over dissipation checks (4) with the residual (3) to the certified level $c^*(\bar{w}_i)$ at the search tolerance, an upper bound of the infimum $c_{\inf}(\bar{w}_i)$ (Corollary 2)
end for
make the levels $c^*(\bar{w}_i)$ nondecreasing by the running minimum from above and form γ by (8) (Corollaries 2 and 3)
if a node check fails **then**
    diagnose, bisect upward for a separating level (Theorem 3); at an accumulating level (Proposition 2) no separating level exists
end if
return $\mathcal{G} = (\gamma, \bar{w}_{\mathrm{cert}}, X_{\mathrm{cert}})$ by Definition 1

The reported $c_{\max}$ on every system is the grid minimum of V over $\partial\mathcal{X}$ for the level $c_{\max}$ of Definition 1 (1.34 on the primary system) rather than a verified boundary separation. A separate certified boundary computation on the same certificate returns 1.4115, above the grid value, which discharges Assumption 4. A verified upper level with an explicit safety margin below this boundary value, 1.3409, is available from the same computation as the sound replacement.

C.3 Certificates, verifier and simulation protocol

The primary system is the single-machine infinite-bus (SMIB) swing dynamics with a scalar disturbance on the acceleration equation (Appendix C.7). The certificate is the class (23), trained by the sampled hinge under two scale constraints, the mean of V on a reference circle held at a fixed value and V held above a fixed level on the domain boundary, followed by the certified training phase. The verifier is an interval branch-and-bound with a budget of $60{,}000$ boxes per check, and the certified level at each node is the smallest level the bisection certifies at that budget. The node list has 7 nodes: zero, a small first positive node chosen so that its certified level lies below the level c_{loc} of Corollary 1, and equally spaced nodes up to $\bar{w} = 0.1$. The simulated deviation at each node is the largest late-time state norm over trajectories started at the origin and at points on the boundary of the certified region, under four fixed disturbance policies (the gradient sign, a constant of either sign, and a sign-switching signal) and an optimized piecewise-constant disturbance found by projected ascent on its switching values with random restarts. Each run records the peak, late-time maximum, and terminal norms, and a step-halving check confirms that the reported value moves by less than the configured tolerance. The certified curve is compared with the late-time maximum, and their ratio, the gain gap (9), appears in Table 2 under the heading certificate-to-simulation ratio, where a value of at least one is the expected ordering subject to simulation error.

Two simulation suites produced the floors reported in this paper, and the gain gap therefore appears at two values for the same certificate. The suite just described, which adds the optimized piecewise-constant disturbance to the four fixed policies, supplies the floors in Tables 2 and 3. An earlier suite, which ran the fixed policies alone, supplies the floors in Tables 6, 7, 9, 10 and 8. The earlier suite searches over a smaller set of disturbances, so its floors are no higher and the gaps computed against them are no smaller. The two are labelled where they are used and are not combined in any comparison.

C.4 Certified levels on six certificates

Table 2 reads one certificate and one verifier per row, all at the top disturbance size of that row's node list, and every row that certifies shows the step curve γ_{raw} converted to a distance by the verified radius. The two existing readings certify nothing at the sizes shown on any row: the check for decrease everywhere fails at zero disturbance and at the top size on every system at the stated

Table 2: Three readings of the same certificate and verifier at the top disturbance size of each row: the existing readings (whether the decrease condition verifies on the whole region at zero and at the top size; the level ν/κ obtained by bounding the disturbance term uniformly over the region, the continuous-time analogue of the ultimate level of the discrete-time Theorem 7 of Li et al. (2026), which was not proposed for this setting) and the level reading of this paper (certified level, its conversion to a distance by the closed form and by the verified radius, the certificate-to-simulation ratio against the late-time deviation), with the ratio after recovery where a check failed. Rows are certificates; "refused" marks a check the verifier did not certify at any node.

system	decrease everywhere ($\bar{w} = 0$ / top)	uniform-bound level ν/κ	certified level c^*	closed form / verified radius	certificate-to-simulation ratio (late)	after recovery
SMIB	no / no	$> c_{\max}$	0.3553	0.775 / 0.602	1.41	1.37
quadratic certificate	–	–	refused	refused	refused	–
Van der Pol	no / no	$> c_{\max}$	0.0524	0.840 / 0.214	1.55	–
Yang et al. loop	no / no	$> c_{\max}$	0.7876	3.799 / 1.072	14.75	2.75
Two-Stage, small torque	no / no	$> c_{\max}$	0.5785	0.866 / 0.844	48.54	–
Two-Stage, big torque	no / no	$> c_{\max}$	0.7196	1.076 / 0.274	58.71	–

budget, the zero-disturbance failure being $9{,}268$ undecided boxes at the tested level 10^{-4} rather than an exhibited violation, and the uniform-bound level exceeds $c_{\max}$ at the top size on every row of a trained certificate, so its sublevel set is not contained in the certified region.

The level reading returns a certified level on every row of a trained certificate, and the certificate-to-simulation ratio at the top size is 1.41 on SMIB and 1.55 on the Van der Pol oscillator through the unchanged pipeline. Over 10 training seeds, fixed before the sweep, 7 certify with ratios between 1.23 and 1.79, and the verifier refuses the rest at every node (Table 7).

The remaining three rows transfer the reading to closed loops released by two published certificate pipelines (Yang et al., 2024; Li et al., 2025b), with the certificate of this paper trained on those loops, and these rows are not a comparison of methods. On the loop of Yang et al. (2024) the ratio is 14.75 over the full region and 2.75 for starts in the sublevel set at the separating level, where Theorem 3 applies, and on the two Two-Stage loops, stiff closed loops on which this certificate class is loose, the ratios are 48.54 and 58.71. The quadratic certificate, the same pipeline with the tangent-gap weights fixed to zero, is refused at every node, so the trained quadratic factor alone yields no curve. The two distance columns convert the same certified level, and the verified radius tightens the conversion on SMIB from 0.775 to 0.602.

C.5 Coverage diagnostic: certified, refuted, and undecided levels

This diagnostic reports how far a search of a given budget got; it is not a property of the certificate. The search reports three levels at each node, and Figure 4 with Table 4 separates them on both systems: the certified level, which bounds the infimum $c_{\inf}$ from above, the largest level a validated refutation excluded, which bounds it from below, and the largest level the budget left undecided, which bounds it neither way. In a rerun of the protocol on both systems from the verified boundary level, the lower endpoint reported with each certified level agrees with the largest undecided level to the reported tolerance, with no level refuted, at 7 of 7 nodes on the primary system and 7 of 7 on the Van der Pol oscillator (largest relative disagreement 6.7×10^{-6} and 6.5×10^{-6}), while a validated refutation exists at only 4 and 2 of the nodes, each of those refuted levels lying strictly below the undecided one. The published endpoint therefore records where the search stopped. No refutation in this rerun excludes the level that endpoint reports. At the top node of the primary system, where nothing was refuted, the gap 3.3×10^{-4} to the certified level measures the resolution of the search at this budget. The remaining reference is the dense infimum computed on a grid of the domain, available at every node and uncertified. It lies 18 percent below the certified level at the top node (Figure 3, left).

Figure 3 (right) shows the step curve γ_{raw} against the late-time deviation, and at $\bar{w} = 0.1$ the certified gain is 0.602, the late-time deviation is 0.426, and their ratio 1.41 bounds the conservatism of the certificate from above, the simulated deviation being a lower bound on the worst case. The peak deviation at that node, 1.14 from starting points on the boundary of the certified region, is a transient

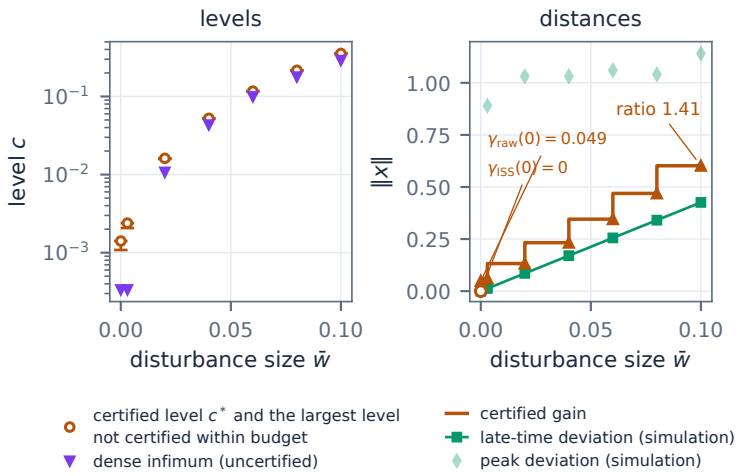


Figure 3: Certified levels and certified gain (SMIB). Left: the certified level at each disturbance size together with the largest level the budget left undecided, and the smallest level at which the condition holds on a dense grid (uncertified). Right: the certified gain against the late-time and peak deviations of the simulation protocol; the certificate-to-simulation ratio at the top node is 1.41, and the zero node gives $\gamma_{\mathrm{ISS}}(0) = 0$.

that the ultimate bound does not claim. At the zero node the certified level 0.0014 lies below the level $c_{\mathrm{loc}} = 0.0047$ of the Lyapunov ball of the linearization, so $\gamma_{\mathrm{ISS}}(0) = 0$ by Corollary 1, the first positive node certifies at 0.0024, also below c_{loc}, and the bound $\limsup_t \|x(t)\| \le K\bar{w}$ holds for every trajectory from X_{cert} up to $\bar{w}_{\mathrm{lin}} = 0.0030$. The uniform-bound level ν/κ of Table 2, with ν one interval bound of the folded disturbance term over the whole region, exceeds $c_{\max} = 1.34$ from the third positive node on and reaches 5.74 at the top node, a level whose sublevel set is not contained in the certified region.

C.6 Reading a failed check

Figure 5 and Table 5 show the recovery at $\bar{w} = 0.14$, beyond the node list, where the dissipation check at the lower level $c_{\mathrm{low}} = 0.65$ fails over the full region. At the budget of 60,000 boxes the residue contains 11,912 genuine and 30,240 unknown boxes (left), the genuine ones classified by floating-point evidence with none of them re-checked through the verifier's enclosure; the lower bound of V over every genuine box exceeds 0.857 (middle). The separating level $c_{\mathrm{sep}} = 0.717$ lies between c_{low} and that lower bound, and both the dissipation check on the region between c_{low} and c_{sep} and the boundary check at c_{sep} pass. Theorem 3 then gives the conclusions of Theorem 2 with ultimate level 0.65 for every start in the sublevel set at c_{sep}, which covers 53 percent of the certified region by grid area, an uncertified estimate. For these starts the largest late-time deviation the simulation found is 0.596 against the certified distance 0.816, a ratio of 1.37 (the after-recovery entry of Table 2). Raising the budget to 240,000 boxes reduces the volume of the unknown boxes from 0.0329 to 0.0146 while the volume of the genuine boxes stays at 0.0130 (right).

At $\bar{w} = 0.16$ the full region certifies only at $c^* = 0.944$, and genuine boxes appear at every level tested below it, so no separating level is found among the tested levels. Proposition 2 describes the limiting case of this pattern, in which violations accumulate at a level and no sound verifier certifies any sub-region at that level at any budget.

Seed robustness. The primary study uses seed 0, and a sweep reruns the entire primary pipeline for the other seeds under the identical recipe and budgets (Table 7). Of the 10 SMIB seeds, 7 certify the full node list end to end, all but one of them with a separating level found for the free-case discharge, and every certified seed reproduces the $\bar{w} = 0.16$ case with no separating level found. The seeds that do not certify converge to poor training minima whose candidate functions carry genuine pointwise violations at every level tested (for two of them, over 10^4 grid points with worst excess up to 0.42, unchanged under tripling the verification budget), and the verifier returns no certificate for them.

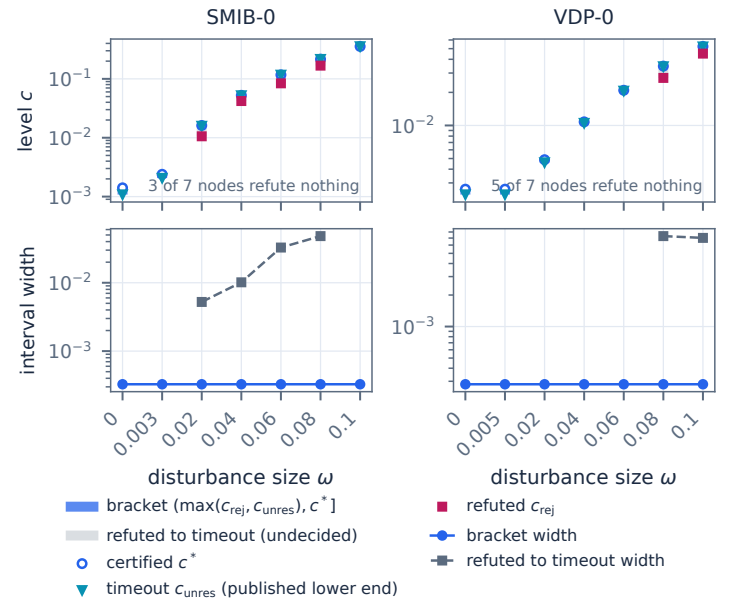


Figure 4: The levels a search returns at each node, on the primary system (left) and the Van der Pol oscillator (right), whose panel titles carry the instance names of the released records. Top: the certified level c^*, the largest refuted level c_{rej} where one exists, and the largest undecided level c_{unres}, on a logarithmic level axis; the dark interval above $\max\{c_{\mathrm{rej}}, c_{\mathrm{unres}}\}$ is the gap between where the search stopped and the certified level, and the light interval below it, between a refuted and an undecided level, is undecided rather than refuted. Bottom: the widths of those two intervals, since the first is thinner than a marker on the axis above. The lower endpoint published with each certified level agrees with the undecided level at 7 of 7 nodes on the left and 7 of 7 on the right.

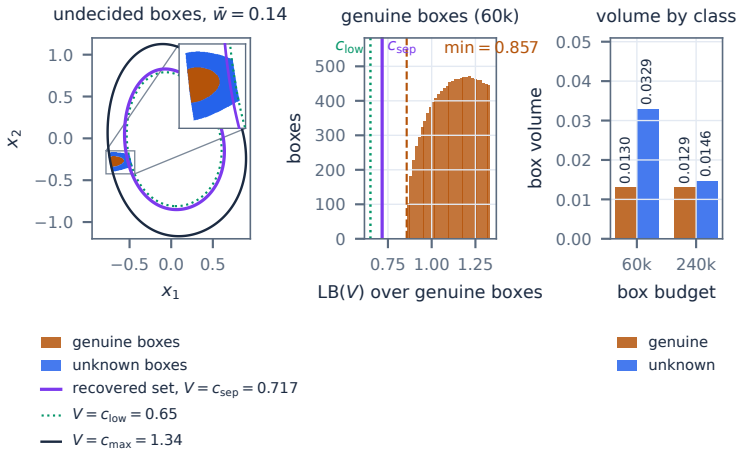


Figure 5: Content of a negative answer (SMIB, $\bar{w} = 0.14$, lower level 0.65). Left: the undecided boxes, genuine (classified by floating-point evidence at a point inside, none re-checked through the verifier's enclosure) and unknown (enclosure too wide), with the recovered set below the separating level 0.717. Middle: the lower V-bounds of the genuine boxes all lie above the separating level. Right: box volume by class at two budgets; the unknown volume falls, the genuine volume does not.

The separating level at $\bar{w} = 0.14$ is 0.717, 1.33, and 0.993 for seeds 0, 1, and 2. All 5 Van der Pol seeds certify the full curve under the selected configuration.

Table 3: SMIB levels and gains at the 7 nodes: the largest level the search did not certify within its budget, which is the undecided level of Table 4 at every node after the rescaling recorded there, the last certified level of the bisection, the dense infimum (uncertified, a grid reference), the uniform-bound level ν/κ, the step curve γ_{raw} as verified radius and as closed form, the class-$\mathcal{K}$ majorant γ_{ISS} of Corollary 1, the simulated deviation (peak, late-time, terminal), and the ratio of the certified gain to the late-time deviation.

$\bar{w}$	0	0.003	0.02	0.04	0.06	0.08	0.1
largest level not certified within budget	0.0011	0.0021	0.0158	0.0522	0.1166	0.2158	0.3550
certified level c^*	0.0014	0.0024	0.0161	0.0525	0.1170	0.2162	0.3553
dense infimum (uncertified)	0.0003	0.0003	0.0107	0.0432	0.0992	0.1780	0.2900
uniform-bound level ν/κ	0.000	0.172	1.148	2.295	3.443	4.590	5.738
γ_{raw} (verified radius)	0.049	0.061	0.132	0.233	0.346	0.469	0.602
closed form $\sqrt{c^*/\lambda_P}$	0.049	0.064	0.165	0.298	0.445	0.604	0.775
γ_{ISS} (class-$\mathcal{K}$ majorant)	0.000	0.165	0.298	0.445	0.604	0.775	0.775
simulated deviation, peak	0.000	0.890	1.032	1.032	1.060	1.041	1.141
simulated deviation, late-time	0.000	0.013	0.085	0.170	0.256	0.341	0.426
simulated deviation, terminal	0.000	0.013	0.081	0.169	0.192	0.313	0.407
ratio to simulation (late)	–	4.78	1.55	1.37	1.35	1.38	1.41

Table 4: The three levels of Section C.5 at every node of both systems, from a rerun of the protocol at the verified boundary level: the largest refuted level, the largest undecided level, and the certified level, beside the published certified level and the published lower endpoint, and the residual left after rescaling the published certified level by the ratio of the rerun's verified upper level b to the grid value b_{grid} the published run used. The rerun reproduces the published levels up to that rescaling. The box budget is a soft cap, tested before a frontier is bounded, so a check can overshoot it by one frontier.

instance	ω	refuted c_{rej}	timeout c_{unres}	certified c^*	published c^*	published lower end	$c^* - (b/b_{\mathrm{grid}})$ published c^*
SMIB-0	0	none	0.001082	0.001409	0.001410	0.001082	7.3×10^{-9}
	0.003	none	0.002064	0.002391	0.002392	0.002064	7.3×10^{-9}
	0.02	0.010575	0.015813	0.016140	0.016142	0.015814	7.2×10^{-9}
	0.04	0.042002	0.052150	0.052477	0.052481	0.052153	7.0×10^{-9}
	0.06	0.083903	0.116639	0.116966	0.116975	0.116647	6.6×10^{-9}
	0.08	0.167706	0.215828	0.216155	0.216171	0.215843	6.1×10^{-9}
	0.1	none	0.354954	0.355281	0.355307	0.354980	5.4×10^{-9}
VDP-0	0	none	0.002349	0.002630	0.002630	0.002349	1.5×10^{-8}
	0.005	none	0.002349	0.002630	0.002630	0.002349	1.5×10^{-8}
	0.02	none	0.004597	0.004878	0.004879	0.004598	1.5×10^{-8}
	0.04	none	0.010500	0.010781	0.010782	0.010501	1.5×10^{-8}
	0.06	none	0.020618	0.020899	0.020902	0.020621	1.5×10^{-8}
	0.08	0.027083	0.034391	0.034672	0.034677	0.034396	1.5×10^{-8}
	0.1	0.045071	0.052098	0.052379	0.052387	0.052106	1.5×10^{-8}

the published rejection endpoint coincides with the timeout level, not with a refuted level, at 7 of 7 SMIB nodes and 7 of 7 VdP nodes (largest relative disagreement 6.7×10^{-6} and 6.5×10^{-6}); a validated refutation exists at only 4 and 2 of them. soft cap: bound calls are tested before a frontier is bounded, so a check can overshoot the budget by one frontier

C.7 Primary system and protocol constants

The primary system is single-machine infinite-bus swing dynamics, $\dot{x}_1 = x_2$, $\dot{x}_2 = (P_m - b\sin(x_1 + \delta^*) - Dx_2 + w)/M$, with $M = 0.4$, $D = 0.3$, $b = 1$, $P_m = 0.5$, $\delta^* = \arcsin(P_m/b)$, on the domain $[-0.95, 0.95] \times [-1.2, 1.2]$, with decrease rate $\kappa = 0.15$. The certificate is the quadratic form plus a one-layer input-convex residual of softplus tangent-gap features,

$$V(x) = x^\top P x + \sum_{k=1}^{K} s_k \varphi_k(x), \quad \varphi_k(x) = \mathrm{sp}(a_k^\top x + b_k) - \mathrm{sp}(b_k) - \sigma(b_k) a_k^\top x, \tag{23}$$

Table 5: Recovery at $\bar{w} = 0.14$ with lower level $c_{\mathrm{low}} = 0.65$ at two box budgets: the undecided boxes by class and their volumes, the minimum of the lower bound of V over the genuine boxes, the separating level, the two checks of Theorem 3, the recovered fraction of X_{cert} (grid area, uncertified), and the ultimate level.

budget	60,000	240,000
undecided boxes	42,152	202,660
genuine boxes	11,912	95,279
unknown boxes	30,240	107,381
genuine volume	0.0130	0.0129
unknown volume	0.0329	0.0146
$\min \mathrm{LB}(V)$ over genuine boxes	0.8571	0.8600
separating level c_{sep}	0.7168	0.7168
band $[c_{\mathrm{low}}, c_{\mathrm{sep}}]$ certified	yes	yes
boundary check at c_{sep}	yes	yes
recovered fraction of X_{cert} (grid area)	0.529	0.529
ultimate level	0.65	0.65

Table 6: Certified gain curve for the time-reversed Van der Pol system ($\mu = 1$): same architecture, losses, budgets, and verification settings as the SMIB study, except for the rate κ. The certified gain is the verified sublevel radius of Corollary 3, capped by the closed form; the closed-form row is shown for comparison. Worst-case simulated lower bound from the earlier fixed-policy suite, Appendix C.3.

$\bar{w}$	0.00	0.02	0.04	0.06	0.08	0.10
certified level c^*	0.0026	0.0049	0.0108	0.0209	0.0347	0.0524
closed-form bound $\sqrt{c^*/\lambda_P}$	0.188	0.256	0.381	0.530	0.683	0.840
certified gain $\gamma(\bar{w})$ (verified radius)	0.064	0.077	0.104	0.139	0.175	0.214
worst simulated deviation	0.000	0.023	0.046	0.070	0.093	0.117
gain gap (certified / simulated)	–	3.31	2.24	1.99	1.88	1.84

with $P \succ 0$, $s_k \geq 0$, sp the softplus, $\sigma = \mathrm{sp}'$ and $K = 10$ features. Each dissipation check runs the interval branch and bound at a budget of 60,000 boxes with refinement threshold 5×10^{-4}, the computation of the verified sublevel radius has a budget of 200,000 boxes, and the level searches use 12 bisection rounds. The disturbance is the scalar load imbalance w with $|w| \leq \bar{w}$. Every constant is set in the experiment configuration file released with the code.

C.8 Full Van der Pol study

The time-reversed Van der Pol oscillator, $\dot{x}_1 = -x_2$, $\dot{x}_2 = x_1 - \mu(1 - x_1^2)x_2 + w$ with $\mu = 1$, has an exponentially stable origin and a polynomial nonlinearity that exercises the square and product rules instead of the sine rule. The pipeline runs unchanged, and the rate κ is the binding hyperparameter: at the SMIB rate $\kappa = 0.15$ the certificate leaves a nearly flat, $\bar{w}$-independent conservatism located in the vanishing-damping region near $|x_1| = 1$, with certified level $c^*(0) = 0.57$ against 0.0026 at $\kappa = 0.08$. Table 6 reports the resulting curve: near zero at zero, superlinear, dominating the worst simulated deviation at every tabulated node, with gain gap 1.55 at the top node. On this system the closed-form bound $\sqrt{c^*/\lambda_P}$ is loose: $\lambda_P = 0.074$, the sublevel sets are elongated, and the ball bound alone gives a gap of 6.1. The verified radius of Corollary 3, run at the same width threshold with 200,000 boxes per node and a total runtime under one second, gives the reported gap.

The Van der Pol certificate shows the same dichotomy as the SMIB discharge study with its own constants: it certifies on the full domain up to $\bar{w} = 0.38$, and at $\bar{w} = 0.42$ violations appear at every tested level and remain under further branch-and-bound, so no separating level was found there.

C.9 Baseline studies on shared plants

Baseline reproduction and comparison. We reproduced the primary baseline (Yang et al., 2024) end to end from its public code and checkpoint on the same machine as our studies: the training stage

Table 7: Per-seed replication of both studies under identical recipes and budgets. The gain gap is the ratio (9) of certified to simulated gain at the top node, with the verified sublevel radius for both systems, as in the respective studies. The three SMIB failures are diagnosed in the text. Worst-case simulated lower bound from the earlier fixed-policy suite, Appendix C.3.

system	seed	certified	c^*	γ	floor	gain gap	train (s)	note
SMIB	0	yes	0.3553	0.602	0.426	1.41	reused	fence found
SMIB	1	yes	0.3459	0.587	0.428	1.37	reused	fence found
SMIB	2	yes	0.3617	0.616	0.428	1.44	reused	fence found
SMIB	3	no	–	–	0.428	–	reused	refused from $\bar{w} = 0$; 49,704 genuine boxes
SMIB	4	no	–	–	0.377	–	reused	refused from $\bar{w} = 0$; 66,268 genuine boxes
SMIB	5	yes	0.5012	0.749	0.417	1.79	8	no fence
SMIB	6	yes	0.3923	0.635	0.429	1.48	10	fence found
SMIB	7	yes	0.2724	0.526	0.428	1.23	9	fence found
SMIB	8	yes	0.3566	0.604	0.428	1.41	7	fence found
SMIB	9	no	–	–	0.430	–	9	refused from $\bar{w} = 0$; 54,897 genuine boxes
SMIB summary: 7 of 10 seeds certified; gain gap in $[1.23, 1.79]$ over the certified seeds hyperparameters fixed before the sweep; no seed selected on its outcome								
Van der Pol	0	yes	0.0524	0.214	0.117	1.84	–	winner config, full sweep
Van der Pol	1	yes	0.0554	0.231	0.117	1.98	–	winner config, full sweep
Van der Pol	2	yes	0.0715	0.279	0.117	2.39	–	winner config, full sweep
Van der Pol	3	yes	0.0550	0.214	0.117	1.83	–	winner config, full sweep
Van der Pol	4	yes	0.0454	0.213	0.117	1.83	–	winner config, full sweep

(150 iterations, final falsifier pass with zero counterexamples, verified-region metric $\rho = 603.5$) and the formal verification stage (α, β-CROWN with input splits; all four specification instances verified safe, mean 10.0 seconds each).

Table 10 maps the two pipelines onto a common axis: the baseline certifies a qualitative object, a verified region-of-attraction level set for the undisturbed closed loop, by level-set bisection; ours certifies the quantitative gain curve over a disturbance range by ultimate-level bisection, through the same verifier family. The two bisections are the common interface, and we executed the shared-plant procedure: our pipeline ran on the baseline's own closed loop (their plant constants and their trained controller, loaded from their checkpoint with the equilibrium property verified), certifying a gain curve over torque disturbances up to 0.03 through the CROWN backend with their controller network embedded in the bound graph, the certified curve dominating the worst-case simulated deviation at every node, a consistency check rather than a soundness argument, in under a minute per configuration.

The comparison caveat is that their certificate addresses the discrete-time Euler map of the undisturbed loop while ours addresses the continuous-time disturbed flow. Our certificate on their plant carries a gain gap near $30\times$, caused by genuine violations confined to the region $V \geq 0.88$, where the CROWN bounds are close to the pointwise values and unchanged under further branch-and-bound from $3{,}000$ to $30{,}000$ boxes. Applying the free-case discharge of Theorem 3 to the same certificate finds a separating level at every disturbance level in the sweep; with the lower level c_{low} seeded at the pointwise-verified level, the regional certified level is ≈ 0.02 across the sweep, and the gap drops from $30\times$ to $7.3\times$ for initial conditions in the sublevel set at the separating level, without retraining. The remaining factor is a property of the trained certificate rather than of the verifier.

Training-side variants. Table 8 reports two retrained variants of the certificate on the same closed loop, a boundary-concentrated hinge and the same hinge with the quadratic factor frozen. With the quadratic factor frozen, a separating level exists at every disturbance level in the sweep, which indicates that the separating levels here follow the quadratic factor rather than the hinge placement.

Second baseline controller on the same plant. The Two-Stage pipeline of Li et al. (2025b) publishes finetuned controllers with its code, and its big-torque pendulum configuration matches exactly the plant above ($m = 0.15, l = 0.5, \beta = 0.1$, torque limit 6), so loading their shipped tanh controller (checked to preserve the equilibrium at load) gives a second independently trained closed loop on the same physical system, and we ran the same protocol: same domain, same CROWN backend with the controller embedded in the bound graph. On the full shared box the certificate fails with

Table 8: Training-side variants on the closed loop of Yang et al. (2024): gain gap at the top disturbance node, globally and for initial conditions in the sublevel set at the separating level (regional); simulated lower references from the earlier fixed-policy suite, Appendix C.3.

variant	global gap	regional gap
sampled hinge (Table 9)	$30\times$	$7.3\times$
boundary-concentrated hinge	$28\times$	$42\times$
boundary-concentrated hinge, quadratic factor frozen	$26\times$	$12\times$

Table 9: Three independently trained closed loops of one pendulum plant under the unchanged pipeline, with stiffness the measured $\max|f_2|$ on the shared box and gains at the top disturbance node. Worst-case simulated lower bound from the earlier fixed-policy suite, Appendix C.3.

closed loop	$\max\|f_2\|$	full box	γ / floor (top)	gap	discharge / mechanism
Yang et al. (ReLU, $\|u\| \le 6$)	56	certified	1.07 / 0.0353	$30\times$	separating levels at every $\bar{w}$; discharge gives regional gap $7.3\times$
Two-Stage big-torque (tanh, $\|u\| \le 6$)	160	fails (5/5 seeds)	0.27 / 0.0022 (near-origin box)	$123\times$	no separating level (limit branch); valley outside the class
Two-Stage small-torque (tanh, $\|u\| \le 0.75$)	34	certified	0.84 / 0.0108	$78\times$	near-origin gap $7.2\times$

genuine violations at every tested level down to $V = 0.02$, so no separating level exists at the tested levels and the discharge mechanism of Theorem 3 is inapplicable there, by the necessity direction of Lemma 2. The closed loop is two-time-scale, with $\max|f_2| = 160$ on a box where $|f_1| \le 1.2$, and the certificate it appears to require is a valley along the slow manifold, nonconvex across it, outside the class (23). Consistent with that reading, the same protocol certifies the full disturbance sweep on a near-origin box where the manifold is locally linear (half-width 0.15; certified γ from 0.17 to 0.27 across $\bar{w} \in [0, 0.03]$, certified above the worst-case simulated deviations) and fails at half-width 0.3. The certified curve on the small box is sound and loose (gap $123\times$ at the top node against simulated deviations near 0.002): high-gain feedback rejects disturbances by fast correction; the Lyapunov gain estimate pays the folded term $|\partial V/\partial x_2|\bar{w}/(ml^2)$, which the valley geometry forces to be large, so gain certificates in this class are conservative on this two-time-scale loop. Both findings hold across the five published controller seeds: the full-box certificate fails with genuine violations on every seed, and the near-origin certification passes the full sweep soundly on every seed. The release also includes a small-torque configuration of the same plant (saturation 0.75 in place of 6), whose stiffness on the shared box is $\max|f_2| = 34$ against 160; the same protocol certifies the full box (sound at every tested disturbance level) with gain gap $78\times$, and the near-origin configuration tightens to $7.2\times$, so certifiability in this class changes on this plant with controller-induced stiffness under everything else held fixed.

Table 9 reports the three closed loops in one view: the loops that certify and the one that does not are consistent with a threshold between the Yang loop's stiffness (56) and the big-torque loop's (160), and both certified loops admit a discharge or near-origin path to a substantially tighter gap. Worst-case simulations on these loops use RK4 with step 5×10^{-4}, at which the levels agree with step 5×10^{-5} to four decimals.

C.10 Training ablations

Table 11 removes the safeguards one at a time, and the measured effects differ in size. The ablation with the largest effect is dropping the global hinge from the certified training phase: box-only training no longer constrains the rest of the domain and every check fails, including at zero disturbance. The scale ablations order cleanly: 0.559 with the scale constraints, 0.420 without the circle-mean constraint, 0.299 with no scale constraint, with $c_{\max}$ tracking; the collapse direction exerts force only while the hinge binds, so the scale constraints remove a monotone drift toward the vacuous limit rather than a sudden collapse. Dropping the certified training phase at equal step budget is within two percent of the full method here. The value of that phase must be established on systems where sampled coverage genuinely fails.

Table 10: The reproduced baseline pipeline versus ours, on a common axis. Worst-case simulated lower bound from the earlier fixed-policy suite, Appendix C.3.

axis	Yang et al. (reproduced)	ours
deliverable	verified ROA level set $\{V \le \rho\}$	certified gain triple $(\gamma, \bar{w}_{\max}, X_{\text{cert}})$
disturbance model	none (undisturbed closed loop)	measurable $\lvert w\rvert \le \bar{w}$, curve in $\bar{w}$
verified quantity	$\rho = 603.5$; 4/4 specs safe	γ at 6 nodes, gap 1.84 at top (Van der Pol study)
level search	level-set bisection (theirs)	ultimate-level bisection $c^*(\bar{w})$ (ours)
verifier	α, β-CROWN (input split)	same family: auto_LiRPA CROWN backend
measured run	150 train iters, clean PGD; verify 10.0 s/spec	$c^*(0.10) = 0.0485$ at 2000 boxes, 3.2 s (Van der Pol, CROWN backend)

Table 11: Training ablations (identical seed, identical step budget). The circle-mean column is the mean of V on the reference circle; the full method holds it at 0.55.

variant	$c_{\max}$	circle mean V	$c^*(0.10)$
CertGain (full)	1.341	0.559	0.3553
no anchor (boundary term kept)	0.999	0.420	0.2614
anchor in phase 1 only	1.097	0.466	0.2856
phase 2 boxes only (no global hinge)	0.878	0.484	**fail**
sampled hinge only (no certified phase)	1.339	0.559	0.3502
no anchor, no boundary term	0.693	0.299	0.1594

D REUSE ACROSS SIZES: PROOFS, PROTOCOL AND FULL TABLES

This appendix supports Section 3. It proves the two statements the section defers, records the protocol under which the measurements were taken, and carries the tables and figures whose full column sets do not fit the main text. Every number quoted here is a generated constant produced from the released records by the asset script, as in the main text.

D.1 THE THRESHOLD AND THE INTERVAL IDENTITY

Lemma 3 (Threshold characterization)**.** *Let $\mathcal{X}$ be compact, $V \in C^1$, f_0 locally Lipschitz and $\mathcal{W}_1$ compact and convex, and let $H(\bar{w})$ be as in* (5)*. Then for every $c \in (0, c_{\max})$: if $c > H(\bar{w})$ the dissipation condition $\rho_{\bar{w}} \le \kappa c$ holds on $\mathcal{S}(c, c_{\max})$, and if $c < H(\bar{w})$ it fails at some point of $\mathcal{S}(c, c_{\max})$.*

Proof. The set $\mathcal{A}_{c_{\max}}$ is compact and $\rho_{\bar{w}}$ is continuous on it, the folded term being a support function of a compact convex set evaluated at a continuous argument, so the maximum in (5) is attained. Let $c > H(\bar{w})$ and take $x \in \mathcal{S}(c, c_{\max})$. Then $x \in \mathcal{A}_{c_{\max}}$ and $V(x) \ge c > H(\bar{w}) \ge \min\{V(x), \rho_{\bar{w}}(x)/\kappa\}$, so the minimum is not $V(x)$ and therefore $\rho_{\bar{w}}(x)/\kappa \le H(\bar{w}) < c$. Now let $c < H(\bar{w})$ and let x attain the maximum. Then $\min\{V(x), \rho_{\bar{w}}(x)/\kappa\} = H(\bar{w}) > c$, so $V(x) > c$ and $\rho_{\bar{w}}(x) > \kappa c$, while $V(x) \le c_{\max}$ by feasibility, so $x \in \mathcal{S}(c, c_{\max})$ and the condition fails there. □

At $c = H(\bar{w})$ the condition is not settled by this argument, and nothing in Section 3 relies on the equality case. A certified level $c^*(\bar{w})$ returned by the bisection satisfies $c^*(\bar{w}) \ge H(\bar{w})$, and the difference between the two is what the enclosures of Section 4 bound.

Proof of Proposition 1. Positive homogeneity of the support function in (3) gives $\rho_{\bar{w}} = a + \bar{w}d$ with a and d independent of $\bar{w}$, so $\mathrm{Sep}(\bar{w}) = a_{\mathrm{up}} + \bar{w}d_{\mathrm{up}}$ is an upper bound of $\rho_{\bar{w}}$ on Q for every $\bar{w} \ge 0$, using $a(x) \le a_{\mathrm{up}}$ and $d(x) \le d_{\mathrm{up}}$. For the forward transport, $R_i + d_{\mathrm{up}}(\bar{w} - \bar{w}_i) = a_{\mathrm{up}} + \bar{w}_i d_{\mathrm{up}} + d_{\mathrm{up}}\bar{w} - d_{\mathrm{up}}\bar{w}_i = \mathrm{Sep}(\bar{w})$. For the chord, write $\theta = (\bar{w} - \bar{w}_i)/(\bar{w}_j - \bar{w}_i) \in [0, 1]$; then $(1-\theta)R_i + \theta R_j = a_{\mathrm{up}} + ((1-\theta)\bar{w}_i + \theta\bar{w}_j)\, d_{\mathrm{up}} = \mathrm{Sep}(\bar{w})$, which is affine interpolation of an affine function. For the backward transport, $R_i - d_{\mathrm{lo}}(\bar{w}_i - \bar{w}) = \mathrm{Sep}(\bar{w}) + (d_{\mathrm{up}} - d_{\mathrm{lo}})(\bar{w}_i - \bar{w})$, and the

correction term is nonnegative on $[0, \bar{w}_i]$ and vanishes exactly when $d_{\mathrm{up}} = d_{\mathrm{lo}}$ or $\bar{w} = \bar{w}_i$. Every candidate is therefore at least $\mathrm{Sep}(\bar{w})$, so the pointwise minimum over the candidates is at least $\mathrm{Sep}(\bar{w})$, and it equals $\mathrm{Sep}(\bar{w})$ if $\mathrm{Sep}(\bar{w})$ is itself retained as a candidate, which the implementation does at every leaf. Without that, a leaf holding only an anchor above $\bar{w}$ is bounded by the backward transport, which exceeds $\mathrm{Sep}(\bar{w})$ when $d_{\mathrm{up}} > d_{\mathrm{lo}}$. □

Three hypotheses carry the argument and each fails for a reason worth naming. The bounds on a and on d must come from the same cached enclosure that the anchor is built from, since a cheaper and looser sensitivity makes the forward transport looser beyond the anchor size than Sep rather than equal to it. Both sizes must be nonnegative, which is what licenses $\bar{w}d \le \bar{w}d_{\mathrm{up}}$. And the arithmetic is exact in the statement; in floating point the two evaluation orders differ by a few units in the last place in either direction, so the identity is asserted in the tests up to a small multiple of machine epsilon and never bitwise. The identity also fails, in the useful direction, for a provider that relaxes $a + \bar{w}d$ jointly, because the maximizers of a and of d over Q need not coincide. That is the case measured on the linear-relaxation backend in Appendix D.4.

The measurement agrees with the proposition. On the interval backend the transports and chords resolve 0 leaves on the primary system and 0 on the Van der Pol instance, in all three orders and on the held-out certificates as well, at 12.4 to 19.5 and 21.8 to 43.8 percent more bound calls than the shared partition alone. A linear relaxation of $a + \bar{w}d$ need not be separable, and there the identity does not hold, the joint bound being strictly tighter than the reassembled one on at least 0.998 of sampled leaves at every sampled positive size, by a median relative amount of 0.0025 to 0.0822, which is smaller than the margins recorded for these checks. Transports and chords resolve 48 leaves there, 0.018 percent of the $271{,}868$ leaf resolutions of the same block, and the symbolic-size baseline resolves 0 leaves at 23.0 percent more boxes. A parameter-valid inequality over a fixed state subdomain is therefore available from a single pass, and on a separable backend it is the reassembled bound the shared partition already uses.

D.2 Instances, orders and budget contracts

Four instances carry the study: a development and a held-out certificate for the primary system and for the Van der Pol oscillator, the held-out pair trained from a different seed with the same configuration. Each instance freezes the certificate weights, the dynamics, the domain box, the disturbance geometry, the rate κ, the upper level $c_{\max}$ and the arithmetic margin, and the hash of that record accompanies every cached bound. The recorded list of checks is the node bisection of Section 5: 91 checks on the primary system and 91 on the Van der Pol instance on the interval backend at a budget of $60{,}000$ bound calls per check, 77 checks at a budget of $2{,}000$ on the linear-relaxation backend, and 416 checks in the held-out batches of 32 sizes. The interval replays run in three orders (sizes increasing, sizes decreasing, and one seeded permutation of the whole list) with three repetitions; the CROWN replay runs two orders with three repetitions, and the held-out batches two orders with two repetitions. Every deterministic counter was identical across repetitions in every block, so only wall time varies, and it is reported as the median with the observed minimum and maximum.

The replays answer a frozen list rather than running two live searches. Two live bisections diverge at the first level on which they disagree, after which a comparison by position compares different levels; the frozen list keeps the comparison matched. The one exception is the online search reported in Appendix D.9, which is compared per probe against a from-scratch search at that probe's own level.

Budget contracts. Under the soft contract the budget is tested before a frontier is bounded, so a check can overshoot by one frontier, and a shared frontier can be larger than the frontier an independent check reaches. On the pilots, the reuse methods spend 52.7 and 47.9 percent of their bound calls beyond the budget on the development primary instance and 57.3 and 60.5 percent on the held-out one, against 4.5 and 5.6 percent for independent checks, counted over the 1560 recorded checks each method answered on the development instance, pooling all orders and all four batch sizes. On the Van der Pol instances, where the partition saturates, no method exceeds 5.4 percent.

Under the strict contract every call to the bound routine is truncated at the budget, a frontier whose remainder the budget cannot afford keeps its parents as leaves rather than creating children that carry no bound of their own, and the bound recomputed after an interruption is dropped. The shared

partition also scans the whole residue for validated points, while the independent run scans a capped subset. When the independent run is rerun with the shared partition's frontier rule and full scan, over every hard-budget block of record and the fixed lists, it spends the same number of bound calls up to one call at a check that reaches its budget, and its answers change only at checks that the full scan, now run by the independent run too, refutes, which are the refutations Table 16 credits to the shared partition. Under that contract no check exceeded its budget in any block (0 and 0 over-budget checks).

The strict contract removes, in node order, 4.2 percent of the independent run's bound calls on the primary system and 27.4 percent of the shared partition's, and it costs the independent run 8 certified levels against 0 for the shared partition. Each of those levels was certified under the soft contract with bound calls beyond the budget. The independent run loses the same 8 certified levels in each of the other two orders, and the shared partition loses one certified level in the permuted order, at the opening position of that order, where its inherited partition is still empty and its search is the independent one. On the Van der Pol instances no method loses a level (0 and 0 certified levels lost).

The strict contract matches one resource, bound calls. The scan that looks for a violating point in the residue is capped by its own count of candidate centres and not by the budget, and that scan grows under the strict contract because the residue grows, so the comparison is matched in bound calls alone.

Timing environment. The wall times of the strict-budget replays of the two development instances (Table 1, Table 14) come from one serial pass over the final configuration at 4 threads: 5 sessions of 3 repetitions, one measuring process at a time, the one-minute load average between 1.26 and 1.60 before and after every block, with one other job on the machine throughout. The tables give the median over all 15 repetitions, and the range printed beside a ratio is the spread of the per-session ratios, which is wider than the spread within a session. Every row of that pass was compared with the record whose counts the paper reports, at 50,778 rows with 0 mismatches, so the wall times describe the same computation as the counts. Wall times on the other interval and CROWN replays were taken with the one-minute load average polled before the timing loop and the run held until it fell below a fixed threshold, with every reading stored in the record. Another process was active on the same machine during parts of those runs, so the spread over repetitions is the only evidence of variability offered and the separations quoted in Section 5.2 are those whose observed ranges do not overlap. The held-out pilot blocks carry their own timing label in the record, either multi-threaded or single-thread consistent, and no wall time is quoted from them anywhere in this paper. The whole-interval runs of Section 4 were run twice with identical deterministic counters, and the same computation was observed at different machine loads to move by a larger fraction than the differences between repetitions, so only orderings far outside that movement are stated.

Table 12: Cost of answering one recorded list of dissipation checks, compact form. B0 resets the partition before every check, B1 shares it, P0 adds the anchored residual records, B1a and B1b are the symbolic-size and batched relaxation baselines, and B1c is the adaptation of the retained partition of Li et al. (2025b). Bound calls count boxes whose bounds are freshly computed, and the last column counts the positions at which the method's answer equals, is stronger than, or is weaker than the independent run's, a stronger answer being one at a position the independent run left unresolved at its budget. The budget is soft here: the count is tested before a frontier is bounded, so a check can overshoot it by one batch of frontier boxes.

instance and backend	method	bound calls	vs B0	equal / stronger / weaker
SMIB-0, interval	B0 independent	3,538,039	1.00	91/0/0
	B1 shared	554,537	6.38	72/19/0
	B1c node reset (adaptation)	743,351	4.76	72/18/0
	B1c strict (adaptation)	2,407,046	1.47	80/11/0
	P0 transport	623,166	5.68	72/19/0
VDP-0, interval	B0 independent	1,477,707	1.00	91/0/0
	B1 shared	286,751	5.15	91/0/0
	B1c node reset (adaptation)	398,103	3.71	91/0/0
	B1c strict (adaptation)	1,089,801	1.36	91/0/0
	P0 transport	405,475	3.64	91/0/0
VDP-0, CROWN	B0 independent	68,000	1.00	77/0/0
	B1 shared	34,193	1.99	77/0/0
	P0 transport	34,156	1.99	77/0/0
	B1a symbolic	34,193	1.99	77/0/0
	B1b batched	34,193	1.99	77/0/0
SMIB-0 (dev), $K{=}32$	B0 independent	17,699,246	1.00	832/0/0
	B1 shared	5,325,513	3.32	618/214/0
	P0 transport	7,723,147	2.29	618/214/0
SMIB-1 (held out), $K{=}32$	B0 independent	18,998,312	1.00	832/0/0
	B1 shared	4,721,625	4.02	574/258/0
	P0 transport	8,681,781	2.19	574/258/0
VDP-0 (dev), $K{=}32$	B0 independent	7,793,612	1.00	832/0/0
	B1 shared	393,801	19.79	828/0/0
	P0 transport	956,509	8.15	828/0/0
VDP-1 (held out), $K{=}32$	B0 independent	8,427,960	1.00	832/0/0
	B1 shared	477,533	17.65	828/0/0
	P0 transport	1,185,270	7.11	828/0/0

soft cap: bound calls are tested before a frontier is bounded, so a check can overshoot the budget by one frontier. Measured price of that contract on the SMIB pilots: the reuse methods spend 47.9 to 60.5 percent of their bound calls beyond the cap, against 4.5 to 5.6 percent for B0. checks per block: SMIB-0, interval 91; VDP-0, interval 91; VDP-0, CROWN 77; SMIB-0 (dev), $K{=}32$ 416; SMIB-1 (held out), $K{=}32$ 416; VDP-0 (dev), $K{=}32$ 416; VDP-1 (held out), $K{=}32$ 416. Soundness-direction discrepancies are zero in every block (0, 0, 0, 0); wall times, provider boxes and the per-block soundness counters are in the appendix table

B1c is a labelled adaptation of Two-Stage's (NeurIPS 2025) dynamic (c1,c2) level update, run through this project's own SharedPartitionVerifier(mode='cached') restricted to the level axis by a per-node store reset; it is not a run of their public code. Appendix D.8 lists the four obstacles to running their pipeline directly on these instances.

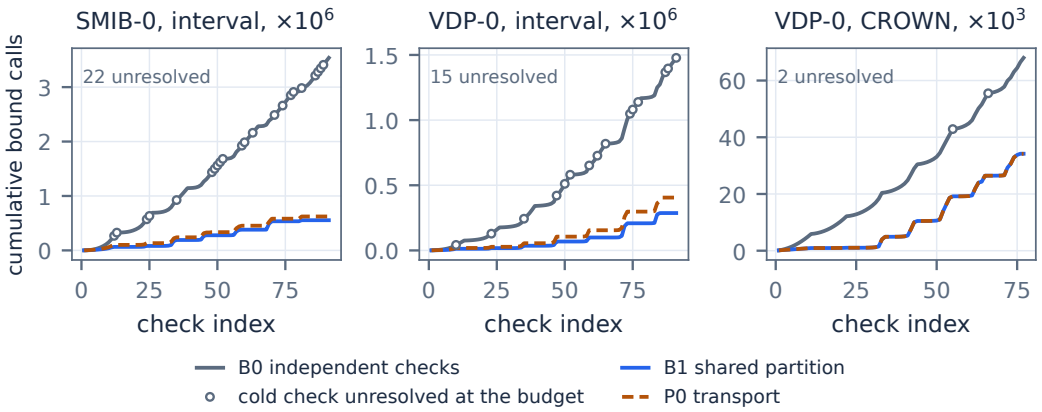


Figure 6: Cumulative bound calls against the position in the recorded list, sizes increasing, under the soft budget. Open markers are checks the independent run left unresolved at its budget.

Table 13: Cost of the recorded list of checks, all columns: bound calls, provider boxes (every box the provider was asked to bound, including the residue re-bound and the witness scan), wall time, the comparison of answers against the independent run and the soundness-direction counter, under the soft budget. B1c is the labelled adaptation of Li et al. (2025b).

D.3 Full cost and answer tables

instance and backend	method	queries	bound calls	provider boxes	vs B0
SMIB-0, interval	B0 independent	91	3,538,039	3,810,004	1.00
	B1 shared	91	554,537	685,823	6.38
	B1c node reset (adaptation)	91	743,351	879,227	4.76
	B1c strict (adaptation)	91	2,407,046	2,639,207	1.47
	P0 transport	91	623,166	773,604	5.68
VDP-0, interval	B0 independent	91	1,477,707	1,554,736	1.00
	B1 shared	91	286,751	338,090	5.15
	B1c node reset (adaptation)	91	398,103	473,268	3.71
	B1c strict (adaptation)	91	1,089,801	1,166,830	1.36
	P0 transport	91	405,475	464,824	3.64
VDP-0, CROWN	B0 independent	77	68,000	71,363	1.00
	B1 shared	77	34,193	44,125	1.99
	P0 transport	77	34,156	44,088	1.99
	B1a symbolic	77	34,193	54,282	1.99
	B1b batched	77	34,193	43,550	1.99
SMIB-0 (dev), $K=32$	B0 independent	416	17,699,246	19,376,414	1.00
	B1 shared	416	5,325,513	7,584,030	3.32
	P0 transport	416	7,723,147	11,008,392	2.29
SMIB-1 (held out), $K=32$	B0 independent	416	18,998,312	20,856,918	1.00
	B1 shared	416	4,721,625	6,612,407	4.02
	P0 transport	416	8,681,781	11,919,187	2.19
VDP-0 (dev), $K=32$	B0 independent	416	7,793,612	8,217,064	1.00
	B1 shared	416	393,801	534,883	19.79
	P0 transport	416	956,509	1,117,297	8.15
VDP-1 (held out), $K=32$	B0 independent	416	8,427,960	9,012,105	1.00
	B1 shared	416	477,533	617,766	17.65
	P0 transport	416	1,185,270	1,388,645	7.11

soft cap: bound calls are tested before a frontier is bounded, so a check can overshoot the budget by one frontier. Measured price of that contract on the SMIB pilots: the reuse methods spend 47.9 to 60.5 percent of their bound calls beyond the cap, against 4.5 to 5.6 percent for B0

instance and backend	method	wall median [min, max] s
SMIB-0, interval	B0 independent	3.31 [3.27, 3.42]
	B1 shared	0.92 [0.83, 0.98]
	B1c node reset (adaptation)	0.75 [0.74, 0.85]
	B1c strict (adaptation)	2.48 [2.44, 2.58]
	P0 transport	0.90 [0.88, 0.93]
VDP-0, interval	B0 independent	1.52 [1.50, 1.54]
	B1 shared	0.38 [0.37, 0.42]
	B1c node reset (adaptation)	0.48 [0.48, 0.51]
	B1c strict (adaptation)	1.23 [1.19, 1.30]
	P0 transport	0.51 [0.50, 0.52]
VDP-0, CROWN	B0 independent	28.06 [27.84, 28.45]
	B1 shared	5.29 [4.95, 5.56]
	P0 transport	4.68 [4.64, 4.95]
	B1a symbolic	7.99 [7.43, 10.91]
	B1b batched	5.11 [4.69, 5.38]
SMIB-0 (dev), $K=32$	B0 independent	not quotable‡
	B1 shared	not quotable‡
	P0 transport	not quotable‡
SMIB-1 (held out), $K=32$	B0 independent	not quotable‡
	B1 shared	not quotable‡
	P0 transport	not quotable‡
VDP-0 (dev), $K=32$	B0 independent	not quotable‡
	B1 shared	not quotable‡
	P0 transport	not quotable‡
VDP-1 (held out), $K=32$	B0 independent	not quotable‡
	B1 shared	not quotable‡
	P0 transport	not quotable‡

wall times are medians of the repetitions. ‡ the $K=32$ batch blocks carry the timing label multi-threaded and single thread consistent in their own record, so no wall time is quoted from them

instance and backend	method	equal / stronger / weaker	sound-ness
SMIB-0, interval	B0 independent	91/0/0	0
	B1 shared	72/19/0	0
	B1c node reset (adaptation)	72/18/0	0
	B1c strict (adaptation)	80/11/0	0
	P0 transport	72/19/0	0
VDP-0, interval	B0 independent	91/0/0	0
	B1 shared	91/0/0	0
	B1c node reset (adaptation)	91/0/0	0
	B1c strict (adaptation)	91/0/0	0
	P0 transport	91/0/0	0
VDP-0, CROWN	B0 independent	77/0/0	0
	B1 shared	77/0/0	0
	P0 transport	77/0/0	0
	B1a symbolic	77/0/0	0
	B1b batched	77/0/0	0
SMIB-0 (dev), $K=32$	B0 independent	832/0/0	0
	B1 shared	618/214/0	0
	P0 transport	618/214/0	0
SMIB-1 (held out), $K=32$	B0 independent	832/0/0	0
	B1 shared	574/258/0	0
	P0 transport	574/258/0	0
VDP-0 (dev), $K=32$	B0 independent	832/0/0	0
	B1 shared	828/0/0	0
	P0 transport	828/0/0	0
VDP-1 (held out), $K=32$	B0 independent	832/0/0	0
	B1 shared	828/0/0	0
	P0 transport	828/0/0	0

B1c is a labelled adaptation of Two-Stage's (NeurIPS 2025) dynamic (c1,c2) level update, run through this project's own SharedPartitionVerifier(mode='cached') restricted to the level axis by a per-node store reset; it is not a run of their public code. Appendix D.8 lists the four obstacles to running their pipeline directly on these instances.

Table 14: The same list of checks under the strict per-check budget, in which every provider call is sliced so that no check exceeds its budget: bound calls and wall time per method and check order on the two development instances, with the ratios over independent verification. The bound-call ratios of Section 5.2 are read from this table.

instance (hard cap)	method	natural bound calls	decreasing bound calls	shuffled bound calls	natural vs B0	decreasing vs B0	shuffled vs B0
SMIB-0 (dev)	B0 independent	3,390,698	3,390,698	3,390,698	1.00	1.00	1.00
	B1 shared	402,547	405,147	464,117	8.42	8.37	7.31
	P0 transport	524,485	521,667	613,141	6.46	6.50	5.53
VDP-0 (dev)	B0 independent	1,449,794	1,449,794	1,449,794	1.00	1.00	1.00
	B1 shared	252,161	220,839	232,537	5.75	6.56	6.23
	P0 transport	373,408	333,403	315,500	3.88	4.35	4.60

the hard-cap replay of the two development instances; hard cap: every provider call is sliced so bound calls never exceed the per-check budget, so these bound-call totals are the ones a budget statement can be read off

instance (hard cap)	method	natural wall median [min, max] s	decreasing wall median [min, max] s	shuffled wall median [min, max] s	natural wall vs B0	decreasing wall vs B0	shuffled wall vs B0
SMIB-0 (dev)	B0 independent	5.09 [4.03, 7.58]	5.35 [4.53, 6.04]	5.59 [4.85, 6.29]	1.00	1.00	1.00
	B1 shared	1.09 [0.56, 1.43]	1.17 [0.93, 1.30]	1.27 [0.86, 1.56]	4.67	4.59	4.38
	P0 transport	1.37 [0.87, 1.72]	1.22 [0.77, 1.62]	1.55 [1.35, 2.09]	3.73	4.39	3.60
VDP-0 (dev)	B0 independent	2.66 [2.34, 3.15]	2.54 [1.91, 3.20]	2.60 [1.80, 3.19]	1.00	1.00	1.00
	B1 shared	0.37 [0.31, 0.90]	0.57 [0.33, 0.93]	0.56 [0.33, 0.86]	7.25	4.49	4.65
	P0 transport	0.84 [0.49, 1.01]	0.79 [0.47, 1.05]	0.76 [0.42, 1.05]	3.17	3.23	3.41

wall seconds of the hard-cap protocol (a serial pass of the adopted configuration, four threads, five sessions of three repetitions pooled, one background GPU job present throughout with the 1-minute load recorded per block, 1.26 to 1.60): a cell is the block's median over its pooled repetitions with [min, max], labelled by method and order and not paired (this driver runs a block's repetitions consecutively); wall vs B0 is the ratio of the two blocks' medians

Table 15: Answers of every method against the independent run at the same position, by instance, backend and order, with the count of answers in the sound direction and the two refutation columns. A lost refutation is a check the independent run refuted with a validated point and the reuse method left unresolved, which is the conservative direction.

instance and backend	method / order	equal	stronger	weaker	soundness	ref. lost	ref. gained
SMIB-0, interval	B0/decreasing	91	0	0	0	0	0
	B0/natural	91	0	0	0	0	0
	B0/shuffled	91	0	0	0	0	0
	B1/decreasing	72	18	0	0	1	0
	B1/natural	72	19	0	0	0	0
	B1/shuffled	72	19	0	0	0	0
	P0/decreasing	72	18	0	0	1	0
	P0/natural	72	19	0	0	0	0
	P0/shuffled	70	19	2	0	0	0

instance and backend	method / order	equal	stronger	weaker	soundness	ref. lost	ref. gained
VDP-0, interval	B0/decreasing	91	0	0	0	0	0
	B0/natural	91	0	0	0	0	0
	B0/shuffled	91	0	0	0	0	0
	B1/decreasing	91	0	0	0	0	0
	B1/natural	91	0	0	0	0	0
	B1/shuffled	91	0	0	0	0	0
	P0/decreasing	91	0	0	0	0	0
	P0/natural	91	0	0	0	0	0
	P0/shuffled	91	0	0	0	0	0

instance and backend	method / order	equal	stron-ger	weak-er	sound-ness	ref. lost	ref. gained
VDP-0, CROWN	B0/natural	77	0	0	0	0	0
	B0/shuffled	77	0	0	0	0	0
	B1/natural	77	0	0	0	0	0
	B1/shuffled	77	0	0	0	0	0
	B1a/natural	77	0	0	0	0	0
	B1a/shuffled	77	0	0	0	0	0
	B1b/natural	77	0	0	0	0	0
	B1b/shuffled	77	0	0	0	0	0
	P0/natural	77	0	0	0	0	0
	P0/shuffled	77	0	0	0	0	0

instance and backend	method / order	equal	stron-ger	weak-er	sound-ness	ref. lost	ref. gained
SMIB-0 (dev), $K=32$	B0/natural	832	0	0	0	0	0
	B0/shuffled	832	0	0	0	0	0
	B1/natural	618	214	0	0	0	0
	B1/shuffled	618	212	2	0	0	0
	P0/natural	618	214	0	0	0	0
	P0/shuffled	614	214	4	0	0	0

instance and backend	method / order	equal	stron-ger	weak-er	sound-ness	ref. lost	ref. gained
SMIB-1 (held out), $K=32$	B0/natural	832	0	0	0	0	0
	B0/shuffled	832	0	0	0	0	0
	B1/natural	574	258	0	0	0	0
	B1/shuffled	576	256	0	0	0	0
	P0/natural	574	258	0	0	0	0
	P0/shuffled	574	256	2	0	0	0

instance and backend	method / order	equal	stron-ger	weak-er	sound-ness	ref. lost	ref. gained
VDP-0 (dev), $K=32$	B0/natural	832	0	0	0	0	0
	B0/shuffled	832	0	0	0	0	0
	B1/natural	828	0	0	0	0	4
	B1/shuffled	828	0	0	0	0	4
	P0/natural	828	0	0	0	0	4
	P0/shuffled	828	0	0	0	0	4

instance and backend	method / order	equal	stron-ger	weak-er	sound-ness	ref. lost	ref. gained
VDP-1 (held out), $K=32$	B0/natural	832	0	0	0	0	0
	B0/shuffled	832	0	0	0	0	0
	B1/natural	828	0	0	0	0	4
	B1/shuffled	828	0	0	0	0	4
	P0/natural	828	0	0	0	0	4
	P0/shuffled	828	0	0	0	0	4

instance and backend	method / order	equal	stron-ger	weak-er	sound-ness	ref. lost	ref. gained
SMIB-0 (dev) (adaptation)	B1c node reset/natural	72	18	0	0	1	0
	B1c node reset/decreasing	72	18	0	0	1	0
	B1c node reset/shuffled	88	3	0	0	0	0
	B1c strict/natural	80	11	0	0	0	0
	B1c strict/decreasing	80	11	0	0	0	0
	B1c strict/shuffled	91	0	0	0	0	0

instance and backend	method / order	equal	stron-ger	weak-er	sound-ness	ref. lost	ref. gained
VDP-0 (dev) (adaptation)	B1c node reset/natural	91	0	0	0	0	0
	B1c node reset/decreasing	91	0	0	0	0	0
	B1c node reset/shuffled	91	0	0	0	0	0
	B1c strict/natural	91	0	0	0	0	0
	B1c strict/decreasing	91	0	0	0	0	0
	B1c strict/shuffled	91	0	0	0	0	0

instance and backend	method / order	equal	stron-ger	weak-er	sound-ness	ref. lost	ref. gained
SMIB-0 (dev) (hard cap)	B0/decreasing	91	0	0	0	0	0
	B0/natural	91	0	0	0	0	0
	B0/shuffled	91	0	0	0	0	0
	B1/decreasing	65	26	0	0	0	0
	B1/natural	64	27	0	0	0	0
	B1/shuffled	65	26	0	0	0	0
	P0/decreasing	65	26	0	0	0	0
	P0/natural	65	26	0	0	0	0
	P0/shuffled	64	26	1	0	0	0

instance and backend	method / order	equal	stron-ger	weak-er	sound-ness	ref. lost	ref. gained
VDP-0 (dev) (hard cap)	B0/decreasing	91	0	0	0	0	0
	B0/natural	91	0	0	0	0	0
	B0/shuffled	91	0	0	0	0	0
	B1/decreasing	91	0	0	0	0	0
	B1/natural	91	0	0	0	0	0
	B1/shuffled	91	0	0	0	0	0
	P0/decreasing	91	0	0	0	0	0
	P0/natural	91	0	0	0	0	0
	P0/shuffled	91	0	0	0	0	0

the interval, CROWN, $K{=}32$ and (adaptation) rows are under the soft cap: bound calls are tested before a frontier is bounded, so a check can overshoot the budget by one frontier; the rows whose instance label carries the suffix (hard cap) are under the hard cap: every provider call is sliced so bound calls never exceed the per-check budget

Table 16: Seed robustness of the reuse ratio: the fixed protocol under the strictly sliced budget in node order on every trained certificate of both systems, with the seed-zero results reproduced exactly before the other seeds ran, and the ratio in the decreasing and permuted orders in the last two columns. For the certificates that fail verification the verifier returns a validated counterexample at every check under both methods.

system	seed	b	B0 calls	B1 calls	ratio	decr.	perm.	eq/str/weak	lost/gain	refused
SMIB	0	1.341	3,390,698	402,547	8.423	8.369	7.306	64/27/0	0/0	
	1	1.330	3,717,657	390,951	9.509	10.279	8.856	52/39/0	0/0	
	2	1.307	3,664,227	394,747	9.282	9.646	8.046	53/38/0	0/0	
	3	0.724	5,450,678	59,999	90.846	90.846	39.957	91/0/0	0/0	yes
	4	0.613	5,460,000	59,999	91.002	91.002	33.919	91/0/0	0/0	yes
	5	1.166	2,226,029	512,463	4.344	3.979	4.860	91/0/0	0/0	
	6	1.326	3,786,885	292,243	12.958	16.857	11.824	51/40/0	0/0	
	7	1.362	3,255,974	529,493	6.149	5.868	5.628	73/18/0	0/0	
	8	1.329	3,630,804	352,503	10.300	10.170	9.157	59/32/0	0/0	
	9	0.914	5,264,448	42,415	124.118	87.742	40.490	91/0/0	0/0	yes
VDP	0	1.151	1,449,794	252,161	5.749	6.565	6.235	91/0/0	0/0	
	1	1.026	1,566,771	244,569	6.406	6.136	6.447	91/0/0	0/0	
	2	0.863	1,403,378	252,087	5.567	6.512	5.107	91/0/0	0/0	
	3	1.022	1,445,596	222,629	6.493	7.148	8.186	91/0/0	0/0	
	4	0.879	1,988,456	344,987	5.764	6.450	6.783	89/0/0	0/2	

Interval backend; hard cap, 60,000 bound calls per check; natural node order; repetition 0 status counters, each row compared against that seed's own independent (cold) run. Reproduction check: seed 0 of both systems reproduces the hard-cap replay of the development instance exactly (bound calls and status counts) before the other seeds run. Refused seeds (marked yes) return a validated counterexample at every one of the 91 checks under both B0 and B1, not only at the zero node; the ratio column is still B0 calls over B1 there, just smaller. The decr. and perm. columns are the same ratio with the checks asked in the decreasing-node order and in the permuted order (the seed replay in the other orders, shuffle seed 2,026, the same certificates under the same cap); the independent run issues identical bound calls in every order on every seed, and the status columns are those of the natural order.

Tables 13 and 15 carry the columns dropped from Table 12, and Table 17 carries the transport comparison in full. The counts behind the soundness-direction comparison of Section 5.2 are 0 of 819 recorded positions on the primary system, 0 of 819 on the Van der Pol instance, 0 of 770 on the linear-relaxation backend, 0 of $37{,}440$ across the four held-out and development pilot instances, and 0 and 0 in the level-axis adaptation. Weaker answers than the independent run's occur in 17 records across the whole study, counting every pilot repetition, all of them in the permuted order: 15 from the

Table 17: Transport and chords against the shared partition alone, at identical positions. A positive count is work the transport method paid and the shared partition did not. On the linear-relaxation backend the column counts the marginal chord and transport resolutions, which is the transport method's total minus the same-size anchor reuse the shared partition already performs.

instance and backend	order	anchor-resolved leaves	extra bound calls	extra boxes	same statuses
SMIB-0, interval	natural	0	68,629	87,781	yes
	decreasing nodes	0	81,325	108,407	yes
	shuffled	0	121,244	188,623	no
VDP-0, interval	natural	0	118,724	126,734	yes
	decreasing nodes	0	63,238	52,134	yes
	shuffled	0	133,703	180,847	yes
VDP-0, CROWN	natural	48	−37	−37	yes
	shuffled	245	−214	−214	yes
SMIB-0 (dev), $K{=}32$	natural	0	2,397,634	3,424,362	yes
SMIB-1 (held out), $K{=}32$	natural	0	3,960,156	5,306,780	yes
VDP-0 (dev), $K{=}32$	natural	0	562,708	582,414	yes
VDP-1 (held out), $K{=}32$	natural	0	707,737	770,879	yes

soft cap: bound calls are tested before a frontier is bounded, so a check can overshoot the budget by one frontier

transport method and 2 from the shared partition. The shared partition's weaker records fall in the batched pilot on the primary development certificate at 32 sizes, where one check position returned unresolved in each of the pilot's two repetitions. Within the recorded protocol itself the shared partition is weaker nowhere, and the transport method is weaker at 2 of 819 primary positions. In this soft-budget replay, validated refutations are lost, in the conservative direction, at 1 of the primary positions for the shared partition and 2 over all methods on the primary system, when a differently shaped residue causes the capped witness scan to miss the center that a from-scratch scan happens to find.

D.4 The linear-relaxation backend

The scheduler's primitive is a size-independent pass over (V, a, d) followed, on whatever that pass leaves unresolved, by a joint pass over the residual at the requested size. Its independent-check baseline therefore issues two relaxation passes on every box the separable clause does not settle, whereas a one-pass verifier run from scratch on the same tree issues one. The record gives the split: 44,794 boxes went through the first pass and 26,569 through the second, in 2,208 relaxation calls altogether, of which 1,133 carried the first pass and 1,075 the second; with the joint boxes removed from the measured charge, a one-pass verifier would have been charged 41,431 bound calls against the 68,000 measured. Every ratio against that baseline is therefore reported both measured and re-estimated, and the three ratios that follow are node-order values. In bound calls the ratio is 1.99 measured against 1.21 estimated, in boxes carried 1.62 against 1.02, and in wall time 5.30 against 2.72, the last from an estimated one-pass block time of 14.4 seconds, all in the recorded pair run at 24 threads. In the permuted order the estimated wall ratio is 3.78 from an estimated 15.2 seconds. Each of the estimates rescales the measured block by the deterministic ratio of relaxation calls; none of them comes from a second run. In the timed sessions the one-pass tree equals the independent run's recorded tree at 73 of 77 positions in node order and 73 in the permuted order, and every exception is at a check where the independent run exhausted its budget (4 of 4 in node order).

Reusing the intermediate bounds of the first pass inside the second, a feature of the bounding library, reproduces the second pass's bounds bit for bit at 39.2 to 46.6 percent of its cost, an estimated 1.13 to 1.33 times on this block; it is orthogonal to partition sharing and was not used in any recorded run.

The implemented one-pass verifier of Section 5.2 supersedes them: within one session at four threads it reproduces the two recorded blocks, and against a freshly re-run independent baseline in that session the shared partition is 7.061 and 9.214 times faster in wall time. On this backend the

Table 18: The linear-relaxation backend. Top: the shared partition against the independent run as measured, and re-estimated against a one-pass cold verifier by removing the second pass from the same record, so that the estimated charged bound calls are the measured ones less the joint boxes, and the estimated provider boxes are the first-pass boxes. Middle: the joint bound of the residual against the separable reassembly on the leaves the replay itself built. Bottom: the implemented one-pass verifier, measured within one session at four threads over three interleaved repetitions, against the shared partition; the header of that block defines the two ratio statistics it carries and separates them.

quantity	B0 measured	B1 measured	ratio measured	B0 one-pass estimate	ratio estimated
charged bound calls	68,000	34,193	1.99	41,431	1.21
provider boxes	71,363	44,125	1.62	44,794	1.02
CROWN boxes	71,363	44,125	1.62	n/a	n/a
wall median, s	28.06	5.29	5.30	n/a	n/a

ω	joint strictly tighter (count)	median relative tightening	joint bound resolves only	symbolic bound resolves only	leaves sampled
0	0	0.0000	0	0	4,000
0.005	3,990	0.0025	1	0	4,000
0.02	3,990	0.0165	116	0	4,000
0.04	3,990	0.0576	216	0	4,000
0.06	3,990	0.0822	226	0	4,000
0.08	3,990	0.0613	236	0	4,000
0.1	3,990	0.0461	352	0	4,000

soft cap: bound calls are tested before a frontier is bounded, so a check can overshoot the budget by one frontier

engine, order	bound calls	compute_bounds calls	CROWN boxes	wall median, s	CPU, s	wall ratio vs B1, paired [min, max]
B1 shared, natural	34,193	299	44,125	3.973	11.304	1 (reference)
B1 shared, shuffled	35,975	214	44,537	2.946	10.058	1 (reference)
OP_matched, natural	41,431	1,133	44,516	14.666	32.945	3.503 [3.483, 3.826]
OP_matched, shuffled	41,431	1,133	44,516	15.011	33.486	4.762 [4.627, 5.296]
OP_budget, natural	53,113	1,162	62,905	15.947	35.658	3.813 [3.713, 4.048]
OP_budget, shuffled	53,113	1,162	62,905	15.650	35.442	4.940 [4.824, 5.366]

one-pass rows: the implemented one-pass $[V, \rho_\omega]$ cold engine on VdP-0, measured within one session at four torch threads, three repetitions per block, interleaved across methods and orders. OP_matched replays B0's per-check first-pass budget; OP_budget runs the nominal 2,000. The wall ratio column is paired: the median of the three same-repetition-index ratios against B1, bracketed by the minimum and maximum of those three, never a confidence interval; the WallRatio macros are instead a ratio of the two blocks' median wall times. B1 rows: the shared partition's own blocks of the same session, the reference of every ratio. Statuses equal the cold trace at all 77 check positions, soundness 0, in all eight blocks, and this supersedes the rescaling estimate above.

Table 19: The implemented one-pass verifier against the shared partition in a second session at eight threads, same protocol, same machine and the same three interleaved repetitions as the four-thread session of Table 18. The relaxation-call and box counts are identical to the first session's in every block, so among the measured quantities only the wall-clock and CPU seconds change between the sessions.

engine, order	bound calls	compute_bounds calls	CROWN boxes	wall median, s	CPU, s	wall ratio vs B1, paired [min, max]
B1 shared, natural	34,193	299	44,125	3.813	21.008	1 (reference)
B1 shared, shuffled	35,975	214	44,537	2.881	19.716	1 (reference)
OP_matched, natural	41,431	1,133	44,516	13.922	61.470	3.540 [3.214, 3.652]
OP_matched, shuffled	41,431	1,133	44,516	13.671	60.391	4.744 [4.732, 4.748]
OP_budget, natural	53,113	1,162	62,905	13.950	62.530	3.658 [3.431, 3.659]
OP_budget, shuffled	53,113	1,162	62,905	14.202	63.149	4.896 [4.217, 4.932]

one-pass rows at eight threads: the implemented one-pass $[V, \rho_\omega]$ cold engine on VdP-0 in one session at eight torch threads, a second session after the four-thread one, three repetitions per block, interleaved across methods and orders, OP_matched and OP_budget as above, B1 rows the shared partition's own blocks of this session. The wall ratio column is paired: the median of the three same-repetition-index ratios against B1, bracketed by the minimum and maximum of those, never a confidence interval. Statuses equal the cold trace at all 77 check positions in all eight blocks, soundness 0. Every counter of every block (calls, boxes, splits, residue) equals the four-thread session's, so the two sessions differ in wall and CPU seconds alone.

Table 20: The batching control: a third four-thread session with the same protocol, adding a one-pass verifier that grows its frontier to the shared partition's average batch size before each relaxation call, at the matched and at the nominal per-check budget. The nominal-budget batched verifier returns the shared partition's statuses at every position, so the ratio against it is the part of the advantage that batching to the same average batch size does not account for.

engine, order	bound calls	compute_bounds calls	CROWN boxes	wall median, s	CPU, s	wall ratio vs B1, paired [min, max]
B1 shared, natural	34,193	299	44,125	4.219	12.423	1 (reference)
B1 shared, shuffled	35,975	214	44,537	3.186	11.111	1 (reference)
B0 independent, natural	68,000	2,208	71,363	30.117	65.315	7.019 [6.916, 7.138]
B0 independent, shuffled	68,000	2,208	71,363	29.434	64.302	9.154 [8.844, 9.308]
OP_matched, natural	41,431	1,133	44,516	14.995	34.616	3.531 [3.425, 3.615]
OP_matched, shuffled	41,431	1,133	44,516	14.971	34.646	4.708 [4.699, 4.937]
OP_budget, natural	53,113	1,162	62,905	15.446	36.424	3.698 [3.661, 4.142]
OP_budget, shuffled	53,113	1,162	62,905	16.228	37.423	5.094 [4.876, 5.484]
OP_batched, natural	48,078	383	53,145	5.534	15.718	1.329 [1.306, 1.402]
OP_batched, shuffled	49,589	339	55,061	5.060	14.380	1.587 [1.574, 1.600]
OP_batched_budget, natural	60,926	378	70,358	5.613	16.769	1.330 [1.297, 1.333]
OP_batched_budget, shuffled	70,771	325	81,179	5.245	15.890	1.646 [1.620, 1.722]

the batching control session: the one-pass cold engine above, bounding at least B boxes per frontier call (a shorter frontier is split breadth-first without bounding), B the first session's B1 boxes per call rounded up (148 natural, 209 shuffled); VdP-0, a third session at four torch threads, three interleaved repetitions, the first session's four engines re-run beside it. OP_batched_budget (nominal 2,000) returns B1's statuses at all 77 check positions in both orders; OP_batched (matched budgets) is cut off at 39 and 57 checks and returns UNRESOLVED where B1 certifies at 25 and 43 positions; soundness 0 in all twelve blocks. Wall ratio: paired, the median of the three same-repetition-index ratios against B1, bracketed by their minimum and maximum, never a confidence interval. Batching alone (OP_batched_budget over OP_budget) is 0.360 and 0.323; $\ln 1.330/\ln 3.531 = 23$ and $\ln 1.646/\ln 4.708 = 32$ percent of the one-pass advantage survives.

cost of one relaxation call varies little with the number of boxes it carries, and the shared partition issues fewer and larger calls than the independent run.

The anchor mechanism is nonzero here because of the joint bound. On $4{,}000$ leaves sampled from the partition the replay built, the joint bound is strictly tighter than the separable reassembly on 0.998 of them at every positive size in the sweep, with median relative tightening from 0.0025 to 0.0822, and at size zero the two agree bitwise, as the record shows. That tightening resolves up to 352 of the sampled leaves that the separable bound does not. The shared partition already obtains it, because its cached mode recomputes the joint anchor at the size asked for, and what is left for transports and chords is the leaves that only a neighbouring size's anchor resolves: 48 leaves in the increasing order and 245 in the permuted one, saving 37 and 214 bound calls. In the increasing order those 48 leaves are 0.97 percent of the same block's $4{,}949$ same-size anchor resolutions by the shared partition and 0.018 percent of that block's $271{,}868$ leaf resolutions.

The two other baselines for this backend behave as the same arithmetic predicts. The symbolic-size baseline appends the disturbance size to the input box and takes one relaxation valid on the whole size interval, which agrees to machine precision with the library's own concretization, and it resolves 0 of the sampled leaves and 0 leaves in the replay, at $10{,}157$ boxes (23.0 percent) and 2.70 seconds more than the shared partition at the same positions. The reason is quantitative: one relaxation covering the whole interval exceeds the size-specific joint bound by the median fractions above, while the thresholds it has to clear are κc at the small levels this certificate certifies. A narrower size interval would shrink that excess and was not run.

The batched baseline carries the node sizes as separate output rows of one call, reproduces the per-size bounds to machine precision, and saves 575 boxes (1.30 percent) against the shared partition at the same positions, because branch and bound splits most anchored leaves before the next size reaches them; the two runs of the driver disagree on the sign of its wall-time difference, so no wall-time claim is made.

D.5 Released closed loops of prior work

The 7 instances of Section 5.2 each freeze a released controller and a certificate we train for that closed loop. Three are pendulum controllers: the released torque policies of Li et al. (2025b) at two saturation limits and the released policy of Yang et al. (2024), whose plant the former matches. The other 4 are controllers released by Li et al. (2025b) for path tracking at two torque limits, a double integrator and a Van der Pol oscillator, whose dynamics we read from the released source. Each controller enters the bound-propagation graph unchanged, so the released loop is checked against the condition of Section 2. For those 4 the certificates hold on boxes of 1 to 25 percent of the released half-widths, which bounds any statement about the region the controllers stabilise but not the comparison between the two methods, which is made on the same instance. The path-tracking certificates passed the pre-registered training sequence of 8 configurations, and the double-integrator and Van der Pol certificates were trained under an amendment of that sequence written after it had failed at every configuration on those two plants. The protocol is the one of Section 5.2 under the strict budget: 52 checks per instance at $2{,}000$ bound calls each, the independent run fixing the list, the shared partition replaying it in three orders with 2 repetitions. Table 21 gives every instance and check order.

The bound-call ratio varies more across these instances than across our own, from 2.39 to 15.61, and the relaxation-call ratio from 8.99 to 21.63.

In cached mode both methods recompute the joint bound of a stored leaf at every size a check asks for, including size zero, where it equals the separable bound in exact arithmetic: on $10{,}107$ stored leaves the two computed bounds differ by at most 4.5×10^{-13}. With that recomputation skipped at size zero by both methods and the check lists held fixed, the ratio runs from 2.34 to 15.13 and lies below the recorded ratio in 9 of the 21 blocks of instance and order, with no soundness discrepancy, no weaker answer and no lost refutation. The lists are held fixed because the independent run under that rule certifies 8 size-zero checks it had left unresolved at the budget, which would change the check lists of 3 instances.

Table 21: The shared partition against independent verification on the released closed loops, per instance and check order: bound calls, their ratio, the checks at which the independent run ended at its budget, the checks the shared partition answered without a fresh bound call, the agreement counts and the count of soundness violations. These runs shared the machine with other jobs, so their wall times are not reported, and Table 22 gives the wall times of a separate serial pass.

plant	controller	order	cap	B0 bound calls	B1 bound calls	ratio	B0 at cap	B1 at 0 calls	eq/str/ weak	sound- ness
pendulum	Yang	natural	2,000	59,247	12,370	4.79	11	36	52/0/0	0
		decreasing	2,000	59,247	15,654	3.78	11	38	52/0/0	0
		permuted	2,000	59,247	14,158	4.18	11	38	52/0/0	0
	Two-Stage small	natural	2,000	74,464	28,338	2.63	22	31	47/5/0	0
		decreasing	2,000	74,464	28,132	2.65	22	24	47/5/0	0
		permuted	2,000	74,464	31,178	2.39	22	30	47/5/0	0
	Two-Stage big	natural	2,000	92,685	7,092	13.07	23	41	35/17/0	0
		decreasing	2,000	92,685	7,553	12.27	23	44	35/17/0	0
		permuted	2,000	92,685	8,256	11.23	23	39	35/17/0	0
path tracking	Two-Stage big	natural	2,000	42,467	9,328	4.55	5	42	52/0/0	0
		decreasing	2,000	42,467	9,262	4.59	5	44	52/0/0	0
		permuted	2,000	42,467	9,514	4.46	5	35	52/0/0	0
	Two-Stage small	natural	2,000	35,214	6,439	5.47	4	38	52/0/0	0
		decreasing	2,000	35,214	7,587	4.64	4	43	52/0/0	0
		permuted	2,000	35,214	7,728	4.56	4	39	52/0/0	0
double integrator	Two-Stage	natural	2,000	48,418	13,998	3.46	8	35	52/0/0	0
		decreasing	2,000	48,418	17,024	2.84	8	40	52/0/0	0
		permuted	2,000	48,418	13,330	3.63	8	36	52/0/0	0
Van der Pol	Two-Stage	natural	2,000	82,777	5,862	14.12	20	29	33/19/0	0
		decreasing	2,000	82,777	5,303	15.61	20	37	33/19/0	0
		permuted	2,000	82,777	6,339	13.06	20	43	34/18/0	0
pendulum	Two-Stage big	natural	8,000	252,040	31,533	7.99	8	38	48/4/0	0

linear relaxation backend (CROWN, backward mode), CPU, 4 threads; hard cap: every provider call is sliced so bound calls never exceed the per-check budget (the cap column, 2,000 and 8,000 bound calls per check), exceeded by none. Plant and controller: the pendulum with the released controller of Yang et al. (Yang) and the small- and big-torque controllers of the Two-Stage pipeline (Two-Stage small and big, the big-torque loop on its near-origin box); path tracking with the big- and small-torque controllers of the Two-Stage pipeline, the double integrator with its Two-Stage controller and Van der Pol with its Two-Stage controller, each checked on a sub-box of its released verification box, 1 to 25 percent of the released half-widths. The double integrator and Van der Pol certificates were trained under an amendment of the certificate-training ladder dated 2026-09-19, written after the pre-registered ladder (8 rungs) failed at every rung on those plants; the path tracking certificates passed the pre-registered ladder. Each loop carries our certificate, with the disturbance sizes and the outer level of its certified sweep. 52 checks per loop (13 per size, 4 sizes); the shared partition in each order, repetition 0 of 2 interleaved repetitions, every counter bit-identical across repetitions. B0: independent verification, a fresh partition per check; B1: the shared partition; ratio: B0 over B1 bound calls. At cap: checks that ended at the budget; at 0 calls: B1 checks answered from the partition it already held. eq/str/weak: positions equal to, stronger than and weaker than B0 in that order; soundness: positions contradicting B0 (0 over every block and repetition). On every cap-2,000 row B1 follows the adopted rules: a parent is split only when both children can be bounded within the cap, so every leaf carries its own bound; the residue is scanned at every centre; and the counterexample points already validated at earlier checks are tried before any bound call. Every block ran concurrently with other work on the machine, so wall times are not of record and are not quoted. Rows at cap 8,000: the companion of the big-torque pendulum loop at the cap its first gate passed, natural order, 2 repetitions, a different fixed list, B1 under the no-spill rule alone, never pooled with the other rows.

Table 22: Wall time of the released closed loops in node order, from a serial pass at 4 threads with 2 repetitions, one measuring process at a time. A second job ran on the machine during this pass and the one-minute load read 1.85 to 2.13 before a block, which does not meet the condition of the pass behind Table 14, so these times are an indication of scale. Every row of the pass reproduces the counts of Table 21.

plant	controller	B0 wall median [min, max] s	B1 wall median [min, max] s	wall ratio	call ratio	relaxation-call ratio
pendulum	Yang	50.0 [49.3, 50.8]	2.9 [2.6, 3.1]	17.52	4.79	16.70
	Two-Stage small	53.3 [52.1, 54.5]	3.9 [3.7, 4.1]	13.64	2.63	16.38
	Two-Stage big	46.9 [46.5, 47.4]	2.3 [2.1, 2.5]	20.31	13.07	21.63
path tracking	Two-Stage big	68.4 [67.6, 69.2]	4.9 [4.3, 5.6]	13.87	4.55	19.12
	Two-Stage small	74.6 [74.1, 75.0]	5.9 [5.8, 6.0]	12.61	5.47	14.25
double integrator	Two-Stage	59.1 [57.4, 60.8]	4.5 [4.3, 4.8]	13.00	3.46	14.25
Van der Pol	Two-Stage	55.2 [55.1, 55.3]	7.1 [6.7, 7.5]	7.80	14.12	8.99

CROWN backward, CPU, four threads; natural order, 52 checks per loop, cap 2,000 bound calls per check. B0: independent verification; B1: the shared partition; a wall cell is the median over two consecutive repetitions with [min, max]; the three ratios are B0 over B1 in the same run (median wall seconds, bound calls, relaxation calls). Measurement condition: one loop per process and one measuring process at a time, under gate tier bg2 of the timing pass, which does not meet that pass's strict load line: the 1-minute load was 1.85 to 2.13 before a block and 1.88 to 4.35 after it (the reading after a block includes the measuring process's own threads), and the other processes on the machine held up to 219 percent of one core. The counters of every check equal, position by position, those of the runs the loop's bound calls are read from (1,456 rows compared, 0 mismatches).

Table 23: The batched one-pass verifier on the released closed loops, in bound calls, relaxation calls and wall time. Each ratio is the one-pass verifier's count over the shared partition's in the same record and check order. The three pendulum rows pair it with the shared partition as recorded in that run, the other rows with the rules of Section 5.2. Counts are identical across repetitions, and the wall ratios come from runs that shared the machine.

plant	controller	order	shared partition calls	one-pass control calls	bound call ratio	relaxation call ratio	wall ratio median [min, max]
pendulum	Yang	natural	26,864	64,914	2.416	2.477	2.568 [2.377, 2.749]
		permuted	26,976	61,102	2.265	2.257	2.760 [2.319, 2.986]
	Two-Stage small	natural	55,586	100,038	1.800	1.469	1.550 [1.339, 1.640]
		permuted	60,318	99,940	1.657	1.465	1.525 [1.427, 1.572]
	Two-Stage big	natural	11,259	69,822	6.201	5.366	6.219 [5.339, 6.969]
		permuted	12,468	66,114	5.303	4.082	4.492 [4.253, 4.546]
path tracking	Two-Stage big	natural	9,328	35,208	3.774	3.510	3.758 [3.648, 3.869]
		permuted	9,514	31,982	3.362	2.958	3.004 [2.983, 3.025]
	Two-Stage small	natural	6,439	25,142	3.905	3.346	3.476 [3.317, 3.636]
		permuted	7,728	24,819	3.212	2.671	2.689 [2.629, 2.749]
double integrator	Two-Stage	natural	13,998	42,660	3.048	2.675	3.063 [2.683, 3.444]
		permuted	13,330	40,011	3.002	2.326	2.750 [2.702, 2.799]
Van der Pol	Two-Stage	natural	5,862	53,956	9.204	3.190	3.407 [3.268, 3.547]
		permuted	6,339	55,048	8.684	4.772	4.796 [4.349, 5.244]

CROWN backward bounds, CPU, four threads, 52 checks per loop, cap 2,000 bound calls per check. Calls: bound calls, summed over the checks. One-pass control: every check starts from a fresh partition and bounds its frontier in batches of at least the shared partition's own boxes per relaxation call in the same record and order. The shared partition of a pendulum row runs the recorded scheduler of that run; that of the other rows runs the adopted rules (no-spill, full residue scan, witness points tried first). Every ratio is the control's count over the shared partition's in the same record and order, so no ratio mixes the two generations of the shared partition; a ratio above 1 means the control spends more. A relaxation call is one call to the bounding backend, which bounds a batch of boxes. Every bound-call and relaxation-call count is identical across the repetitions of its block (3 on the pendulum rows and 2 on the other rows). Wall ratio: the control's wall seconds over the shared partition's at the same repetition index, median [min, max]; every block ran concurrently with other work, so these wall seconds are not of record.

On these instances the independent run takes 46.9 to 74.6 seconds and the shared partition 2.3 to 7.1, a wall ratio of 7.80 to 20.31 that follows the relaxation-call ratio more closely than the bound-call ratio, since on this backend a relaxation call is the unit of cost.

D.6 The decrease rate as a second parameter

The residual $\rho_{\bar{w},\kappa} = \langle \nabla V, f_0 \rangle + \kappa V + \bar{w}\zeta_1$ is affine in the rate as well as in the size, so a record that stores an upper bound g_Q^+ on $\langle \nabla V, f_0 \rangle$ separately from V_Q^+ evaluates the residual clause at any pair $(\bar{w}, \kappa)$ with $\bar{w} \in [0, \bar{w}_{\mathrm{cert}}]$ and $\kappa > 0$, and the argument of Theorem 1 applies verbatim to the upper envelope $U(\bar{w}, \kappa) = \max_Q \min\{c_{\max}, V_Q^+, (g_Q^+ + \kappa V_Q^+ + \bar{w} d_Q^+)/\kappa\}$ and to the lower envelope from the same validated points. On the interval backend the stored a_{up} is already the sum of the two bounds, so the split changes no recorded decision: with the split records the B0, B1 and P0 blocks of the recorded protocol are reproduced bit for bit at the training rate on both instances. Table 24 and Figure 7 report the reading of the one partition at six rates, from a quarter of the training rate to twice it, against an independent verification at each rate.

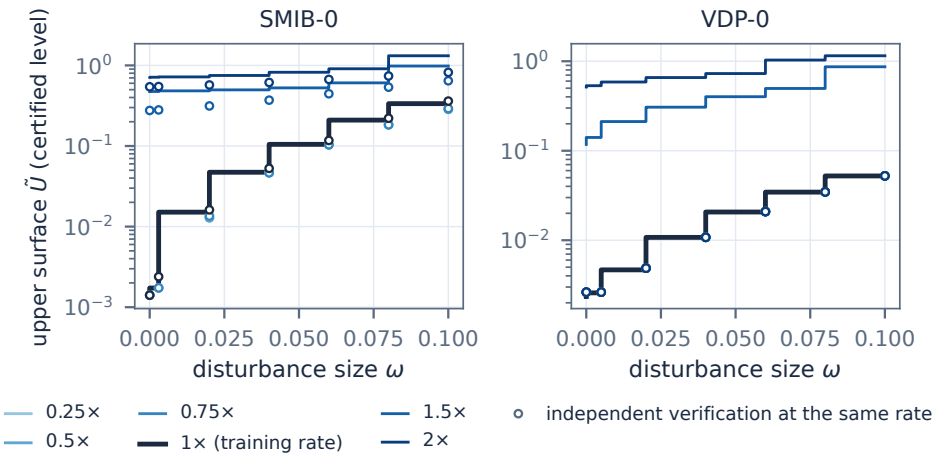
Table 24: One partition read at six decrease rates against independent verification at each rate, on the interval backend under the strict cap. The partition is the training-rate sweep's and the readings at the other rates spend no bound calls.

instance	rate factor	independent calls	eq/str/weak	max level ratio	node width$/b$
SMIB-0	0.25	2,082,399	73/0/10	1.18	0.0570
	0.5	2,465,284	73/0/10	1.18	0.0570
	0.75	2,976,506	73/0/10	1.15	0.0570
	1	3,359,122	73/15/0	1.00	0.0339
	1.5	3,347,069	48/0/33	1.71	0.3485
	2	2,519,319	43/0/38	1.61	0.4804
VDP-0	0.25	1,170,535	89/0/0	1.00	0.0042
	0.5	1,257,456	89/0/0	1.00	0.0042
	0.75	1,349,692	89/0/0	1.00	0.0042
	1	1,449,794	89/0/0	1.00	0.0042
	1.5	1,678,899	35/0/54	53.59	0.7121
	2	1,953,001	25/0/64	202.69†	0.9588

Interval backend; hard cap, 60,000 bound calls per check; the reading is the 13-check level search per node on the one partition, 91 positions per rate, compared position by position with an independent verification (a fresh partition per check) at that rate. Rate factor: the rate as a multiple of the training rate, 0.15 (SMIB-0) and 0.08 (VDP-0); the local rate ceilings are 0.312 (2.08 times the training rate) and 0.498 (6.23 times), so no rung was excluded. The shared partition's bound calls, 402,547 and 252,161, are the training-rate sweep's (the hard-cap replay's B1 block, cross-read), and the re-reads spent 0 bound calls beyond that sweep. The independent runs are native level searches at each rate (a fresh partition per check, bisection at that rate), not the replayed development record, so at the training rate the independent calls and the eq/str/weak counts can differ from the hard-cap replay's by the bisection path (SMIB-0: 3,359,122 against 3,390,698 bound calls, 73/15/0 against 64/27/0; VDP-0: 1,449,794 bound calls in both, 89/0/0 against 91/0/0), and the two are reported side by side, not pooled. Summed over the 6 rates the independent calls are 41.6 and 35.1 times the partition's; over the 4 rates at or below the training rate, 27.0 and 20.7 times. eq/str/weak: positions equal to, stronger than and weaker than the independent answer; the rest are point refutations the reading does not attempt (3 to 10 per rate on SMIB-0, 2 on VDP-0). Max level ratio: the reading's smallest certified level over the independent one, the maximum over the nodes; a dagger marks a rate at which some node certifies no level below b, and such nodes are left out of the maximum. Node width: the largest $U - L$ over the nodes, in units of b.

D.7 A finer size grid compared with the enclosure

A direct alternative to the enclosure is the recorded sweep on a finer size grid, whose certified levels form a step upper envelope by monotonicity, together with the same validated lower envelope the enclosure uses. Table 25 reports its cost on grids with 7, 13 and 25 nodes under the strict budget, with the shared partition, against the recorded enclosure widths in the same units. On the primary system the finest grid stays wider than the max-gap enclosure at 4.8 times that enclosure's bound calls. On Van der Pol it is narrower than every recorded enclosure width, and on that instance the recorded widths are bounded below by the enclosure's gap-certification tolerance of 0.023027 in



Interval backend; hard cap: every provider call is sliced so bound calls never exceed the per-check budget (60,000). Lines: the one partition built at the training rate, read at each rate at zero bound calls; steps, so nothing is claimed between nodes. A node with no certified level below b is drawn at b. At and below the training rate the certified levels are the same at every node and the drawn surfaces differ by at most 0.11 percent, so those lines overlap.

Figure 7: The upper envelope of the one partition against the disturbance size at each of the six rates, with the independent verification's smallest certified level at the same rate as hollow markers. At and below the training rate the two coincide on Van der Pol and lie within a factor of 1.18 on the primary system; above it the partition carries no slack and the reading loosens.

level units, so the comparison there is against the enclosure's recorded bound, and the enclosure's attainable width is not measured.

D.8 The level-axis adaptation

The retained-partition level search of Li et al. (2025b) reuses a verified partition across changes of level at a fixed disturbance, in the narrowing direction its own proof covers. Running their published pipeline directly on these instances was not possible for four independent reasons, recorded before the substitute was built: their verification driver is not part of the released material we obtained; their relaxation is documented for tanh and sigmoid multilayer networks rather than for the quadratic-plus-convex certificate used here; their condition carries no disturbance term and no rate constant, so these instances would have to be pushed through it by injecting both, which modifies their method rather than running it; and their outer level moves within a run while ours is frozen by the problem hash. The baseline is therefore a labelled adaptation, our own shared partition with the store reset at every change of size, which reproduces the mechanism (reuse along the level axis, none across sizes) without claiming to reproduce their implementation.

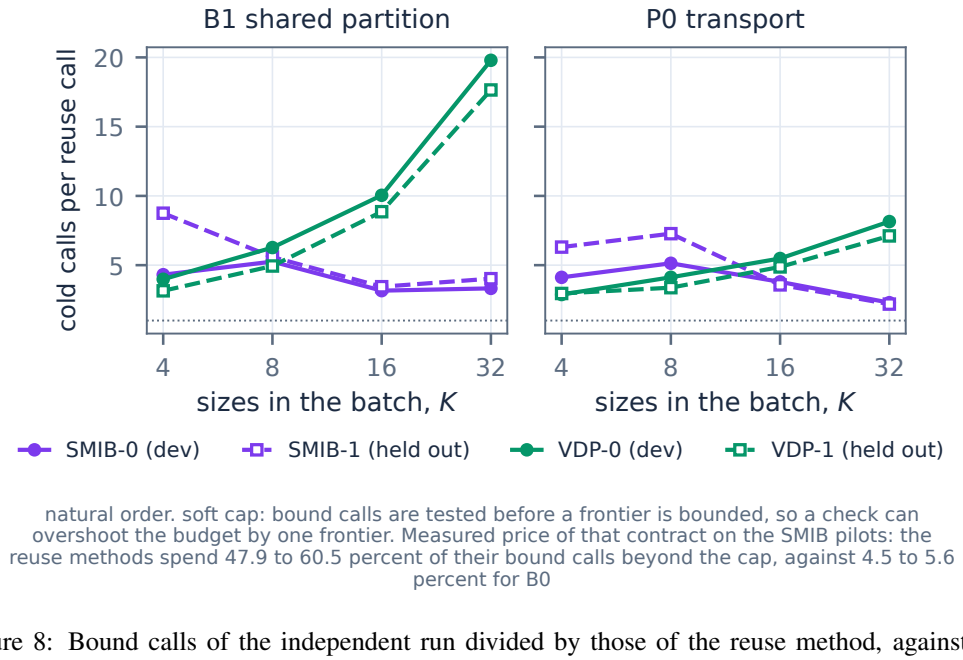
Its measured share of the shared partition's reduction depends on the order in which checks arrive. Grouped by size, the adaptation reaches 4.76 times fewer bound calls than independent checks on the primary system and 3.71 times fewer on the Van der Pol instance, which is 93.7 and 90.7 percent of the reduction the shared partition achieves, the remaining 188,814 and 111,352 bound calls being what crossing the size axis adds. In the permuted order the same quantities are 1.14 and 1.12, that is 14.6 and 13.0 percent, because the reset rule fires on most transitions. A stricter variant that also resets whenever the level drops below the running maximum at the current size was run as well and captures less in every order. Both orders are reported because neither is the workload: a level search at one fixed disturbance produces the grouped order, and a sweep that revisits sizes produces the interleaved one.

The measurement above isolates the direction of reuse that Section 6 distinguishes, measured on one instance family and one budget.

Table 25: A finer size grid against the enclosure: the recorded sweep on three grids with the shared partition, its step upper envelope and the recorded lower envelope, against the enclosure's recorded widths, in units of the upper level.

instance	nodes or rule	bound calls	width /b	calls over recorded sweep
SMIB-0	7 nodes (the recorded grid)	402,547	0.1166	1.00
	13 nodes	689,019	0.0816	1.71
	25 nodes	957,495	0.0618	2.38
	enclosure, max-gap rule, 200k budget	199,999	0.0556	0.50
	enclosure, best recorded width	999,999	0.0195	2.48
VDP-0	7 nodes (the recorded grid)	252,161	0.0193	1.00
	13 nodes	327,159	0.0140	1.30
	25 nodes	288,743	0.0104	1.15
	enclosure, max-gap rule, 200k budget	51,593	0.0200	0.20
	enclosure, best recorded width	613,325	0.0151	2.43

Interval backend; hard cap, 60,000 bound calls per check, exceeded by none; the shared partition in natural node order, one repetition. Grid 1 is the recorded grid; each finer grid adds the midpoints of the previous one (13 and 25 nodes); every node on every grid certifies a level. Upper envelope: the left-continuous step of the certified levels. Lower envelope: read from the enclosure records, no new witness, as the pointwise maximum of the max-gap run's subdivision step (347 witnesses over 5,940 intervals on SMIB-0, 538 over 13 on VDP-0) and the curve-targeted run's node values (444 and 653 witnesses). Width: the enclosure's own subinterval rule, exact for this step pair, in units of the upper level b (1.3409 and 1.1514). Calls over recorded sweep: bound calls over the recorded grid's, 402,547 and 252,161. The max-gap row is the band of record at its 200,000 budget (on VDP-0 it stopped on its own width target at 51,593 calls); the best recorded width is that of the oracle-free gap rule at the 1,000,000 budget, priced at its terminal bound calls, on VDP-0 first reached at 176,545 calls. Caveat on VDP-0: the enclosure's recorded widths at that level are floored by its own gap-certification tolerance (0.02 of the upper level, 0.023027 in level units), so what the VDP-0 comparison establishes is that the finer sweep beats the enclosure's recorded upper bounds on its width, not that it beats the enclosure's actual band. Midpoint excess: at the nodes a finer grid adds, the coarser grid's step envelope exceeds the finer sweep's own certified level by up to 196.3 percent (grid 1 to 2) and 49.5 percent (grid 2 to 3) on SMIB-0, 45.5 and 23.4 percent on VDP-0. No soundness-direction discrepancy in the 1,170 positions compared with the independent hard-cap verification.



natural order. soft cap: bound calls are tested before a frontier is bounded, so a check can overshoot the budget by one frontier. Measured price of that contract on the SMIB pilots: the reuse methods spend 47.9 to 60.5 percent of their bound calls beyond the cap, against 4.5 to 5.6 percent for B0

Figure 8: Bound calls of the independent run divided by those of the reuse method, against the number of sizes in the batch, on the development and held-out certificates of both systems. The quotient grows from 3.98 at four sizes to 19.79 at 32 on the Van der Pol development certificate and falls from 4.31 to 3.32 on the primary one, whose partition keeps growing rather than saturating.

Table 26: Top: the enclosure returned on the two instances, with the bound on the gap over the continuum and, beside it, the wider gap of the step enclosure through the same nodes. Middle: the finite-check baseline on the scalar reference, where the exact threshold is known, with the certified width and the width recomputed with the Lipschitz cones switched off. Bottom: cost to reach each target width, all five methods, in bound calls, wall seconds and estimated CPU seconds.

instance (certified profile band)	nodes	outer level b	certified gap bound	gap $/b$ (pct)	step-band gap	bound calls	wall s / CPU s
SMIB-0, partial band	7	1.3409	0.074609	5.56	0.163035	199,999	17.1 / 19.2
VDP-0, certified band	7	1.1514	0.023005	2.00	0.036343	51,593	7.0 / 7.7

gap bound: an upper bound of $\sup_\omega(U - L)$ over the whole size interval, not a node quantity. soft cap: bound calls are tested before a frontier is bounded, so a check can overshoot the budget by one frontier

κ	placement	witnesses	nodes	modulus K	certified band width	width with the cones off	largest node bracket η
0.5	adaptive	verifier-only	38	4.0020	0.0496	0.0561	0.0731
0.5	adaptive	grid	33	4.0020	0.0495	0.0495	0.0199
0.5	uniform	verifier-only	120	4.0020	0.0346	0.0542	0.0688
0.5	uniform	grid	95	4.0020	0.0364	0.0379	0.0199
1.5	adaptive	verifier-only	12	1.3340	0.0416	0.4458	0.0508
1.5	adaptive	grid	11	1.3340	0.0465	0.4388	0.0028
1.5	uniform	verifier-only	120	1.3340	0.0091	0.0186	0.0029
1.5	uniform	grid	33	1.3340	0.0293	0.0611	0.0028

These rows are the synthetic scalar reference of Appendix D.9, where $b = 1$ and H is known in closed form

instance	method	target width $/b$	achieved width $/b$	bound calls	wall s	CPU-s (est)	width contract
SMIB	B1	0.1		not reached			step band
	P2	0.1	0.0972	104,131	9.0	9.5	self-certified continuum
	P2-smallseed	0.1	0.0972	104,175	8.2	8.7	same contract as P2
	B3	0.1	0.0940	1,759,686	3.3	6.5	self-certified continuum
	B2	0.1		not reached			step band
	B1	0.05		not reached			step band
	P2	0.05	0.0494	282,607	23.1	24.5	self-certified continuum
	P2-smallseed	0.05	0.0492	283,751	23.3	24.7	same contract as P2
	B3	0.05	0.0438	4,451,274	11.9	23.2	self-certified continuum
	B2	0.05		not reached			step band
	B1	0.02		not reached			step band
	P2	0.02		not reached†			self-certified continuum
	P2-smallseed	0.02		not reached†			same contract as P2
	B3	0.02		not reached			self-certified continuum
	B2	0.02		not reached			step band
VDP	B1	0.1	0.0301	286,751	0.4	n/a	step band
	P2	0.1	0.0703	9,101	5.2	5.7	self-certified continuum
	P2-smallseed	0.1	0.0706	9,119	3.3	3.4	same contract as P2
	B3	0.1	0.0455	131,860	0.2	0.2	self-certified continuum
	B2	0.1	0.0316	104,043	4.3	4.8	step band
	B1	0.05	0.0301	286,751	0.4	n/a	step band
	P2	0.05	0.0398	16,087	6.0	6.5	self-certified continuum
	P2-smallseed	0.05	0.0399	16,137	3.9	4.1	same contract as P2
	B3	0.05	0.0455	131,860	0.2	0.2	self-certified continuum
	B2	0.05	0.0316	104,043	4.3	4.8	step band
	B1	0.02		not reached			step band
	P2	0.02	0.0200	51,593	8.0	8.6	self-certified continuum
	P2-smallseed	0.02	0.0200	51,807	5.8	6.0	same contract as P2
	B3	0.02	0.0198	300,076	0.4	0.4	self-certified continuum
	B2	0.02		not reached			step band

CPU-s (est) is the wall time times that method's own process/wall ratio in this record, not a second measurement; SMIB B3 ran multithreaded, so its wall column understates the work it did; † the SMIB P2 and SMIB P2-smallseed runs stopped on their budgets rather than on a target, so the widths those runs report and every target they never reached are censored by the budget and not by the method

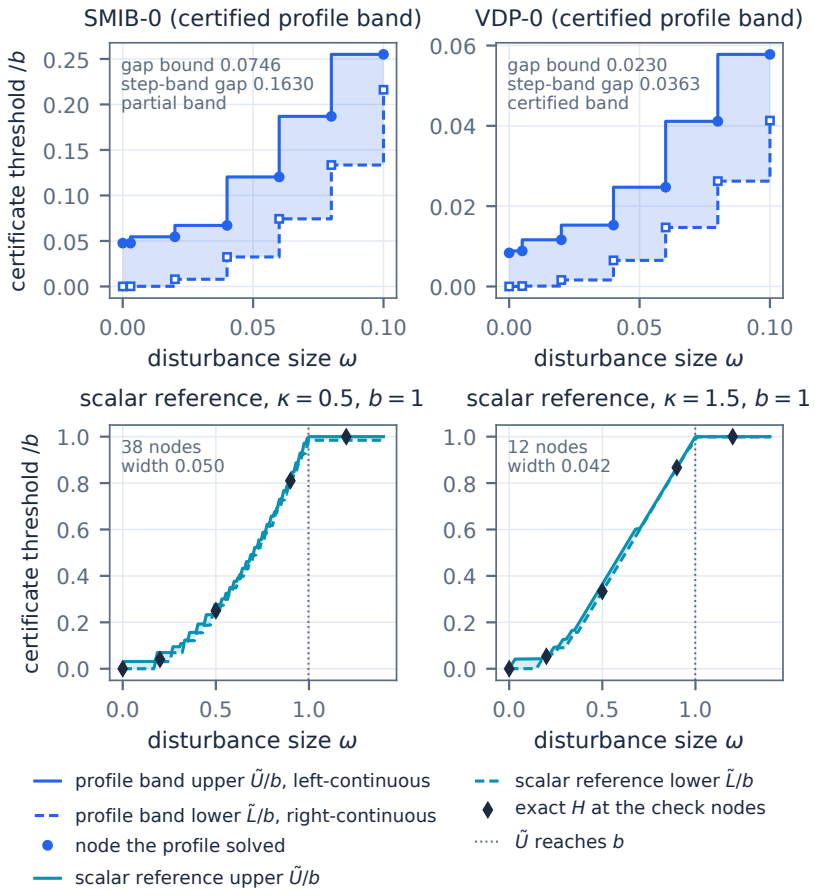


Figure 9: Two-sided enclosures of the threshold H over the size interval. Top: the two instances, upper values from the shared partition and lower values from validated points, drawn as the step interpolant the record certifies. Bottom: the finite-check baseline on a scalar reference whose H is known in closed form, the only panels in which the exact threshold can be drawn.

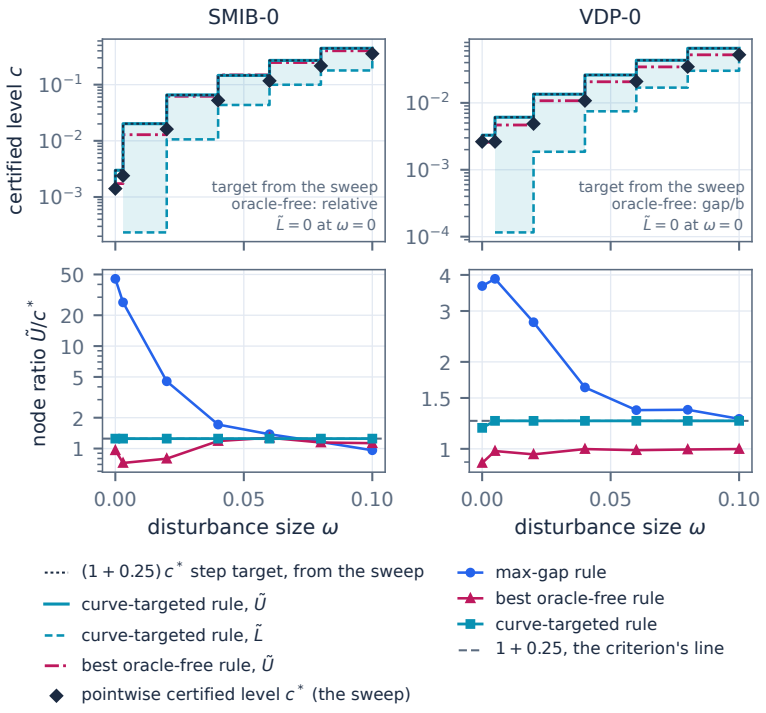


Figure 10: Enclosures of the certificate-dependent threshold on the two development instances, discussed in Section 5.3. Upper envelopes are left-continuous and lower envelopes right-continuous. Curves are step interpolants, so nothing is claimed between two nodes beyond the certified step. The top row draws the curve-targeted rule and the best oracle-free rule at 1,000,000 bound calls; the bottom row adds the max-gap rule's node ratio at its own 200,000. Under the soft cap, bound calls are tested before a frontier is bounded, so a check can overshoot the budget by one frontier.

D.9 THE WHOLE-INTERVAL ENCLOSURE AND ITS BASELINES

Table 26 gives the enclosure and the full cost-to-target comparison. Three points qualify the main-text summary. First, the returned object on the primary system is a partial enclosure that stopped on its budget of $200{,}000$ bound calls rather than on its target, so its width there is an upper bound on what more computation would buy, while the Van der Pol run reached its target within the same $200{,}000$-call budget and stopped there. Second, the step enclosure through the same nodes is a weaker object than the certified continuum gap, 0.163035 against 0.074609 on the primary system and 0.036343 against 0.023005 on the other, and any use of these numbers has to say which of the two it quotes. Third, refinement here is a deterministic heuristic that splits the leaf contributing most to the widest subinterval, and no convergence of the enclosure to the threshold is claimed; the upper values are loose at the small sizes, a consequence of a stopping rule driven by a single supremum over the whole interval.

Two further readings of the same runs belong here. At the node protocol's own cost the certified widths, in units of $c_{\max}$, are 0.0309 for the enclosure against 0.1250 for the step enclosure through the nodes on the primary system, and 0.0200 against 0.0301 on the Van der Pol instance, the first cut off by the $400{,}000$-call budget of that run rather than by the method.

The levels an online search reports also move under a shared partition: an online level search on it returns a certified level tighter than the independent search at 36 of the 42 combinations of node and block, equal at 6 and looser at 0, at the same budget per check.

The finite-check baseline places nodes where the enclosure is widest and joins them with cones of a certified Lipschitz modulus, which is Shubert's sequential sampling of a one-dimensional function (Shubert, 1972) with a verified modulus in place of an assumed one. On the two real instances the cones contribute nothing: the modulus is 41.16 on the primary system and 30.05 on the Van der Pol instance, so over a cell of the recorded spacing the cone allows more movement than the threshold makes over the whole interval. On the primary system this was measured directly, by recomputing the width with the cones switched off and obtaining the same number. On the Van der Pol instance the evidence is the baseline's own certified width, which equals the supremum of the step enclosure through its nodes. On these two instances the baseline is then monotone interpolation between certified nodes.

On the scalar reference of the specification, where the modulus is within a small factor of the true slope, the cones contribute: at $\kappa = 1.5$ the certified width is 0.0416 with the cones and 0.4458 without them, a factor of 10.72, while at $\kappa = 0.5$ the same comparison gives 1.13. On that reference adaptive placement needs 33 nodes against 60 for the smallest equally spaced design that certifies the same target (1.82) at $\kappa = 0.5$, and 11 against 18 (1.64) at $\kappa = 1.5$. The larger ratios 2.88 and 3.00 that follow from comparing against the a priori node count of the design formula carry the conservatism of that formula as well and are not the adaptivity advantage.

A consistency cross-check runs every method against every other at the recorded nodes: the largest validated lower value must sit below every method's upper value and the smallest certified upper value above every method's lower value. Four different methods supply the binding lower value at some node and four the binding upper value, and no pair violates the ordering on either instance. The lower side limits the width. A lower value needs a point, either a validated refutation or a validated witness, and where the search finds none the lower value falls back to zero and the node contributes only its upper end. The two endpoints a node bisection returns are asymmetric for the same reason, as Section C.5 reports per node.

Finally, we report the cost measures. The plan behind these experiments names time, in wall and CPU seconds, as the primary axis, and bound calls were not pre-registered as a sole measure, so Table 26 reports both and Section 5.3 quotes both. The bound-call result is a statement about a backend on which a bound call is expensive relative to the bookkeeping around it. On the interval backend of this comparison a bound call is cheap: 8.99×10^{-7} seconds for one interval bound call, against 2.00×10^{-4} seconds for one relaxation box. The same comparison on the linear-relaxation backend, on one instance, is reported in Table 28, where the two wall times are comparable at the reduced witness seed grid and the enclosure is slower at the grid of the reported runs.

D.10 Proof of the profile enclosure

Proof of Theorem 1. Write $h(x, \bar{w}) = \min\{V(x), \rho_{\bar{w}}(x)/\kappa\}$, so that $H(\bar{w}) = \max_{x \in \mathcal{A}_{c_{\max}}} h(x, \bar{w})$ by (5).

Step 1 (lower values). Each x_j lies in $\mathcal{A}_{c_{\max}}$, so $h(x_j, \bar{w}) \le H(\bar{w})$. Since $\bar{w} \ge 0$, $a_j^- \le a(x_j)$ and $0 \le d_j^- \le d(x_j)$, we have $a_j^- + \bar{w} d_j^- \le a(x_j) + \bar{w} d(x_j) = \rho_{\bar{w}}(x_j)$ by (6), and $v_j^- \le V(x_j)$, so $\min\{v_j^-, (a_j^- + \bar{w} d_j^-)/\kappa\} \le h(x_j, \bar{w})$. Taking the maximum over j preserves the inequality. For the outer maximum with zero, take $x = 0$, which lies in $\mathcal{A}_{c_{\max}}$ because $V(0) = 0$ (Section 2). There $f_0(0) = 0$ and $V(0) = 0$ reduce the residual (3) to the folded term, and $\zeta_{\bar{w}}(0) \ge 0$ because $0 \in \mathcal{W}(\bar{w})$, so $h(0, \bar{w}) = \min\{0, \zeta_{\bar{w}}(0)/\kappa\} = 0$ and $H(\bar{w}) \ge 0$.

Step 2 (upper values). Let $x \in \mathcal{A}_{c_{\max}}$. The boxes cover $\mathcal{A}_{c_{\max}}$ and a discarded box is disjoint from it, so x lies in some retained Q. There $V(x) \le \min\{c_{\max}, V_Q^+\}$ and $\rho_{\bar{w}}(x) \le R_Q^+(\bar{w})$, hence $h(x, \bar{w}) \le \min\{c_{\max}, V_Q^+, R_Q^+(\bar{w})/\kappa\} \le U(\bar{w})$. Maximizing over x gives $H(\bar{w}) \le U(\bar{w})$.

Step 3 (normalization). H is nondecreasing (pointwise monotonicity of $\rho_{\bar{w}}$ in $\bar{w}$, since $d \ge 0$ in (6); compare Corollary 2). For $t \ge \bar{w}$, $H(\bar{w}) \le H(t) \le U(t)$, and taking the infimum over such t gives $H(\bar{w}) \le \tilde{U}(\bar{w})$; the infimum is over a set that shrinks as $\bar{w}$ grows, so $\tilde{U}$ is nondecreasing. For $t \le \bar{w}$, $L(t) \le H(t) \le H(\bar{w})$, so $\tilde{L}(\bar{w}) \le H(\bar{w})$, and the supremum is over a set that grows with $\bar{w}$.

Step 4 (gap). Let the subintervals I cover $[0, \bar{w}_{\mathrm{cert}}]$, let $\overline{u}_I \ge \sup_I U$ and $\underline{\ell}_I \le \inf_I L$. For $\bar{w} \in I$, $U(\bar{w}) - L(\bar{w}) \le \overline{u}_I - \underline{\ell}_I$, and every $\bar{w}$ of the interval lies in some I, so the maximum of $\overline{u}_I - \underline{\ell}_I$ over the subdivision bounds $\sup_{\bar{w}}(U - L)$. □

Two conditions of the last step come from the implementation and are not hypotheses of the theorem. The shortcut that reads $\underline{\ell}_I$ as the value of L at the left endpoint of I needs every witness value nondecreasing in the size, which holds only with $d_j^- \ge 0$; a computed lower bound on d that comes out negative is replaced by zero, which nonnegativity of the folded term licenses, and without that replacement the reported width can fall below the true one while the pointwise inequalities still hold. Reading $\overline{u}_I$ as the value of U at an endpoint is likewise sound only while every candidate is nondecreasing in the size. A candidate that is not, such as a chord valid on a subinterval, can make the maximum peak inside I, so the implementation bounds the supremum of each candidate on I separately and takes the maximum of those bounds.

D.11 Size-interval splitting and the enclosure on the linear-relaxation backend

Table 27 carries the split study, and Table 28 the enclosure and its baselines on the linear-relaxation backend. Three qualifications belong with the second. The two enclosure rows differ only in the number of witness seed points, 1,000,000 against 10,201, and on this backend the larger grid adds cost without strengthening the enclosure, since the enclosure's cost there is almost entirely witness evaluation. Every method in the table ran multithreaded at 4 torch threads, so neither wall column is a single-core figure. The baseline B2 never reached the tighter of the two contracts, so its printed cost is the cost at which it stopped.

D.12 Cost of refinement, by rule and budget

Table 29 lists every rule at both budgets, and Section 5.3 quotes the rows that price the enclosure. Two of its rows coincide to every digit: the oracle-free rule that prioritizes by the gap in units of $c_{\max}$ is the max-gap rule with its width stopping criterion removed, so wherever the max-gap run also spends its whole budget, the two produce the same partition.

The fractions in the fourth column are readings in the size metric, the fraction of the size interval on which the stated accuracy holds, and they are not confidence statements.

Table 27: Splitting the size interval into K equal parts before the joint relaxation is taken, on the Van der Pol instance and the linear-relaxation backend. B1 is the shared partition run inside the same process, and K1 is the unsplit joint bound, which reproduces the symbolic-size baseline of Table 12 counter for counter. Splitting resolves leaves that the unsplit bound leaves open and costs more boxes through the relaxation. Wall times were taken on a shared machine and are labelled accordingly; the deterministic counters were identical across the three repetitions in all ten blocks.

order	block	bound calls	CROWN boxes	compute_bounds calls	symbolic resolved	statuses equal	sound-ness	wall median s (not of record)
natural	B1	34,193	44,125	299	0	77	0	9.03
	K1	34,193	54,282	447	0	77	0	11.51
	K2	31,662	49,742	441	656	77	0	15.11
	K4	32,950	51,395	433	1,155	77	0	16.85
	K7	32,997	51,259	428	1,522	77	0	16.40
shuffled	B1	35,975	44,537	214	0	77	0	6.53
	K1	35,975	57,241	314	0	77	0	12.97
	K2	36,132	56,790	313	710	77	0	11.82
	K4	36,175	56,529	313	1,206	77	0	15.56
	K7	36,261	56,424	313	1,583	77	0	14.83

K is the number of equal parts the size interval is split into before the joint symbolic bound is taken; B1 is the generic-caching control, re-run inside the same process at the same four torch threads, and K1 is the unsplit symbolic baseline, which reproduces the B1a block of the CROWN replay counter for counter. Wall times were taken on a shared machine and are not of record; the deterministic counters were bit identical across the three repetitions in all ten blocks. Statuses equal counts the replayed positions at which the method returned exactly the cold status of the frozen trace, out of 77 per block. soft cap: bound calls are tested before a frontier is bounded, so a check can overshoot the budget by one frontier

Table 28: Cost of reaching two certified widths on the linear-relaxation backend, Van der Pol instance. P2 is the enclosure of Section 4, P2-smallseed the same enclosure with a smaller witness seed grid, B3 the finite-check baseline and B2 the independent fixed-size solves. The two wall columns are repetition zero's crossing, read from its anytime curve, and the same crossing rescaled by the method's median-to-repetition-zero ratio.

method	width contract	bound calls	compute_bounds calls	wall s, rep 0	wall s, median scaled
P2	0.05 b	929	151	10.03	10.03
	0.02 b	1,977	165	10.79	10.79
P2-smallseed	0.05 b	1,151	32	2.67	2.63
	0.02 b	2,267	46	3.25	3.20
B3	0.05 b	72,184	173	2.44	2.50
	0.02 b	158,463	242	3.84	3.92
B2	0.05 b	5,951	1,071	41.46	41.02
	0.02 b	5,951†	1,071	41.46	n/a

P2 and P2-smallseed are the same profile band at two witness seed grids, 1,000,000 points and 10,201, and differ in nothing else; on this backend the seed grid is a cost, not a claim. Wall times are repetition 0's, since a crossing is read off that repetition's anytime curve, beside the same crossing rescaled by the method's median-to-repetition-0 ratio; every method here is multithreaded at four torch threads, so neither column is a single-core figure; † B2 at 0.02 b never reached the contract, so the cost printed is the cost at which the method stopped and the scaled wall column is empty: the record carries no crossing to rescale

Table 29: Cost of certifying a profile within 0.25 of the pointwise sweep at every node, for every refinement rule of Section 4 at two bound-call budgets. The rule cell says whether the rule reads the sweep's levels as its target, since a cost quoted against the sweep presupposes them. The last two columns carry the accuracy statement of Section 4 and its cost, the width guarantee on the threshold and the bound calls spent.

instance	refinement rule, sweep input	bound-call budget	max node $\tilde{U}/c^*$	fraction within 25 pct	gap bound $/b$	bound calls
SMIB-0	max-gap, no sweep input	200k	45.3227	0.5651	0.0556	199,999
	max-gap, no sweep input	1M	13.5776	0.8000	0.0200	703,509
	curve-targeted, target from sweep	200k	1.2500	1.0000	0.1628	111,167
	curve-targeted, target from sweep	1M	1.2500	1.0000	0.1628	111,167
	oracle-free gap$/b$, no sweep input	200k	45.3227	0.5651	0.0556	199,999
	oracle-free gap$/b$, no sweep input	1M	11.8490	0.9400	0.0195	999,999
	oracle-free radius, no sweep input	200k	2.7383	0.8772	0.0937	199,999
	oracle-free radius, no sweep input	1M	1.4584	0.9653	0.0668	999,999
	oracle-free relative, no sweep input	200k	2.8262	0.3995	0.5327	199,999
	oracle-free relative, no sweep input	1M	1.2747	0.9810	0.0830	999,999
VDP-0	max-gap, no sweep input	200k	3.8774	0.4704	0.0200	51,593
	max-gap, no sweep input	1M	3.8774	0.4704	0.0200	51,593
	curve-targeted, target from sweep	200k	1.2500	1.0000	0.0276	54,179
	curve-targeted, target from sweep	1M	1.2500	1.0000	0.0276	54,179
	oracle-free gap$/b$, no sweep input	200k	2.8869	0.7980	0.0151	199,999
	oracle-free gap$/b$, no sweep input	1M	0.9983	1.0000	0.0151	613,325
	oracle-free radius, no sweep input	200k	1.5291	0.8702	0.0259	199,999
	oracle-free radius, no sweep input	1M	0.9983	1.0000	0.0165	613,325
	oracle-free relative, no sweep input	200k	1.5291	0.8426	0.0251	199,999
	oracle-free relative, no sweep input	1M	0.9983	1.0000	0.0165	613,325

budgets are exact: 200k is 200,000 bound calls and 1M is 1,000,000. The pointwise sweep's own hard-cap totals, the cap the criterion compares against: SMIB 498,425, VdP 274,355 bound calls. These totals are the shared partition's first hard-cap sweep of the same 91 checks (natural order, median of three repetitions, the scheduler as first recorded, before the no-spill rule), which bounded SMIB 525,861, VdP 293,688 boxes; the curve-targeted runs at the cheapest budget within 25 pct bounded SMIB 2.32, VdP 3.77 times those boxes. Under the rules of the main comparison the same sweep spends SMIB 402,547, VdP 252,161 bound calls. A rule that reads the sweep's levels is marked in its row, since a cost quoted against the sweep presupposes them. The oracle-free gap$/b$ rule is the max-gap rule with its width stop removed, so the two coincide wherever the max-gap run also spent its whole budget. Bound calls are the budget currency; every run here was pinned to four torch threads and no wall time is printed, because the oracle-free rule runs ran on a contended machine and their timings are not of record.

D.13 Instances on which the pipeline returns no level

Six studies on larger systems returned no usable level, and the released reports carry their full records.

Released four-state and six-state closed loops. The 4-state cart-pole loop was run on the box of half-width 0.1 with upper level 2.01×10^{-3}, for 3,999 bound calls over 13 iterations, and the 6-state ducted-fan loop on the box of half-width 0.2 with upper level 1.19×10^{-2}, for 1,999 bound calls over 10 iterations. On both loops the enclosure of Section 4 certifies nothing stronger than the trivial statement that the threshold lies between zero and the upper level: the certified widths are 1.000001 and 1.000000 times that level, and the cart-pole record carries the status insufficient for dynamical guarantee and the ducted-fan record the status insufficient for dynamical guarantee. Two obstructions were separated in the diagnostics, and they are independent of one another. The dominant one sits in the released certificates: on the two loops, validated points of the fixed region attain threshold values of 0.999000 and 0.999000 times the upper level already at zero disturbance, so the dissipation condition of Section 2 fails there for the fixed pair of certificate and region, and no sound verifier can certify it. The second is the width of the relaxation on those graphs, which would keep the enclosure loose independently of the first obstruction: 0.7925 and 0.9470 of the leaves are blocked by it, carrying 0.1691 and 0.9482 of the retained volume. Neither obstruction is a statement about the plants.

The released cart-pole certificate under its own condition. The released verification of the cart-pole loop checks $\langle \nabla W, f(x, u(x)) \rangle < 0$ between two levels $c_1 \le W \le c_2$ of its certificate W, with no rate term and at zero disturbance, and the release states the pair $(0.01, 0.989)$ as the initial thresholds of an adaptive schedule rather than as a verified region between the levels. Independent verification under that condition on the released box, with the linear-relaxation backend and a budget of 2,000,000 bound calls per check, returns the status counterexample to condition at that pair: 34 of 5,000 residue centres and 2084 of 2084 ascent points validate as violations, the smallest validated at $W = 0.9608$, and all 5 released seeds violate at that pair on 3.5 to 6.4 percent of uniformly sampled points between the levels, reproduced through the authors' own code. At the narrower pair $(0.01, 0.95)$ of the release's fine-tuning default the check is unresolved at the same budget, with 1,189,343 residue boxes, none refuted and no validated counterexample, 88.9 percent of them straddling the upper level. The thinnest-margin points of that region between the levels resolve only at half-widths of 4.9×10^{-4} to 9.8×10^{-4} against a widest residue box of 0.225, and raising the lower level to 0.015 leaves the residue unchanged. In the seed sweep at that narrower pair, uniform sampling finds no violation on any seed and ascent finds one validated violation on one seed, just above the lower level. The result of the pre-registered sequence of attempts is NO-GO, and the six-state loops ship neither a pair of levels nor a box for this condition and were not run. Wall times of these checks were taken on a shared machine and are not of record.

Certificates trained under our condition on the cart-pole plant. 6 certificates were trained on the released four-state plant and its controller, on regions of half-width 0.05 to 0.1 with upper levels from 1.06×10^{-4} to 2.00×10^{-4}, and none was dropped. All 6 pass the first of the two pre-registered criteria, with 0 counterexamples found at zero disturbance and 0 at the largest magnitude over $4{,}096$ projected-gradient starts, and none passes the second: at each of the three levels probed, 0.5, 0.25 and 0.1 of the upper level, the check returns unresolved on every attempt, spending $407{,}465$ to $472{,}645$ bound calls against a box budget of $400{,}000$ and leaving a residue of $65{,}494$ to $110{,}880$ boxes that contains 0 refuted boxes. Training-time search finding no counterexample is evidence about the search and does not establish the condition. The measurements are consistent with the relaxation width as the obstruction, and shrinking the region as the protocol pre-registered left the relative crossing width unchanged at the sizes tried. A power law fitted to the relaxation width over the residue, with relative crossing widths of 9.77×10^{-4} at the median and 3.91×10^{-3} at the widest, projects a covering of 9.85×10^{7} to 2.78×10^{10} boxes and a tightening factor of 122.8 to 1538.8 before the budget would suffice. That projection assumes uniform refinement and does not bound what an adaptive or stronger method would need.

The optimized relaxation. A separate run replaced the frozen bound routine with an optimized relaxation on 200 of the widest residue boxes of the second attempt, the one trained at rate 0.5. At

0.5 of the upper level it raised the number of boxes the residual clause resolves from 33 to 74, that is, from a fraction 0.165 to 0.370, and lowered the median bound from 1.28×10^{-2} to 6.70×10^{-3}, a median tightening of 1.91 over the boxes, at 57.4 times the wall time. The fitted power law turns that into the tightening still required before the covering implied by the residue fits the gate's box budget, 57.5 on this attempt and 920.1 on the others, so under this fit no attempt comes inside the budget. The pre-registered rule stops there, full verification was not re-run, and the earlier negative result stands.

The same run diagnosed why the optimization had been inert before: the frozen bound routine evaluates with gradients disabled, which silently switches the optimization off. An audit of every relaxation record of this paper found that all of them requested the plain backward bound, which that switch leaves unchanged, so no reported number depended on the optimization, and the routine now either runs it or refuses the request.

Certified training and a quadratic certificate on the same plant. Two further studies changed the certificate and kept the released closed loop, its controller included, fixed. The first trained 6 certificates of the class above with the verifier's own bound inside the loss and a training-time branch and bound over the region, at 6 pre-registered settings of the rate, the region half-width, the capacity and the training budget, and put each to the two criteria of the previous paragraphs (Table 30). 4 of the 6 pass the first criterion, and on 2 of them the sampled threshold at zero disturbance lies below 10^{-6} of the upper level, so sampling and ascent find no point of the region above that level where the condition fails. The verifier still returns unresolved at all three levels on those, spending $805{,}849$ to $905{,}813$ bound calls per level against the box budget of $800{,}000$, and no certificate passes the second criterion. Training against the bound widened the relative crossing width from 9.77×10^{-4} to 3.91×10^{-3} on five of the six, a factor of 4.0, and the covering that width implies is 1.32×10^{9} boxes, 1.65×10^{3} times the budget.

A deeper run on one of the two certificates with a sampled threshold below 10^{-6} of the upper level, at 2.5 times the box budget, reached residue boxes 0.25 times the width the residual clause needs and remained unresolved with 0 refuted boxes. The surviving boxes straddle the outer level set, with upper bounds on V reaching 2.6 times the upper level, so they may contain points outside the region, where the condition asks nothing and the residual can be large. A box containing such a point is settled by the first clause of Section 3.1 and not by the residual clause. Settling it takes a partition that separates the level set, and a tighter bound leaves it in place. That mechanism is visible in the counts, since the boxes settled as lying above the region rose from $298{,}523$ to $502{,}139$ and the residue volume fell by a factor of 77.2 for the 2.5 times larger budget, and the run had not finished. Reaching the width the residual clause needs therefore did not suffice on this instance, and the covering that width implies understates the cost.

The second study replaced the trained certificate by the quadratic form $V = x^{\top} P x$ from the Lyapunov equation of the closed loop's linearization, a form that decreases at the rate 0.0454 near the equilibrium, on 6 pre-registered configurations over the rate and the region half-width. Every one of its 18 checks returns unresolved with 0 refuted boxes, and so does a re-run at $900{,}000$ boxes, while the sampled threshold at zero disturbance is at most 2.53×10^{-2} of the upper level and projected-gradient search finds 0 counterexamples. Its crossing width is 3.91×10^{-3}, the widest value among the trained certificates, and its usable disturbance size is at most 1.88×10^{-5} newtons, 16 times below the size pre-stated for the sweep. Across the three classes of certificate (sampled, trained against the bound, and quadratic), the residual bound on this closed loop clears the check only on boxes whose covering exceeds the budget, and the region clauses need a partition that separates the outer level set, so changing the certificate with the controller frozen moved neither obstruction far enough. The standard pipelines retrain the controller together with the certificate, which the frozen-loop protocol here did not permit.

D.14 Dimension scaling on a synthetic family

The family is a chain of $m = \lfloor n/2 \rfloor$ damped coupled oscillators with state $(q_1, v_1, \ldots, q_m, v_m)$ and, for odd n, one first-order state z driven by the last position,

$$\dot{q}_k = v_k, \qquad \dot{v}_k = -k_0 q_k - d_0 v_k + k_c(q_{k-1} - q_k) + k_c(q_{k+1} - q_k) - c_3 v_k^3 + w_k, \qquad \dot{z} = -\alpha z + \beta q_m, \tag{24}$$

Table 30: The twelve cart-pole certificates on which the verifier certifies no level at zero disturbance: six trained with the verifier's bound in the loss (A to F) and six quadratic forms from the Lyapunov equation of the loop (1 to 6), each under the criteria of the trained-certificate study. Sampled thresholds are lower estimates and certify nothing.

#	κ	h	capacity	b	sampled $H(0)/b$	PGD cx	status $0.5b$	status $0.25b$	status $0.1b$	f^*
A	1	0.1	16	1.78×10^{-4}	0.413136	0	unres.	refuted	refuted	3.91×10^{-3}
B	2	0.1	16	3.57×10^{-4}	0.163478	0	unres.	unres.	unres.	9.77×10^{-4}
C	1	0.1	16	2.47×10^{-4}	1	4096	refuted	refuted	refuted	3.91×10^{-3}
D	1	0.1	32	1.64×10^{-4}	0	0	unres.	unres.	unres.	3.91×10^{-3}
E	0.3	0.1	16	1.81×10^{-4}	1	4060	refuted	refuted	refuted	3.91×10^{-3}
F	1	0.05	16	1.91×10^{-4}	0	0	unres.	unres.	unres.	3.91×10^{-3}
1	2.27×10^{-2}	0.05	quadratic	3.72×10^{-6}	0	0	unres.	unres.	unres.	3.91×10^{-3}
2	2.27×10^{-2}	0.1	quadratic	1.49×10^{-5}	0.0253298	0	unres.	unres.	unres.	3.91×10^{-3}
3	1.13×10^{-2}	0.05	quadratic	3.72×10^{-6}	0	0	unres.	unres.	unres.	3.91×10^{-3}
4	1.13×10^{-2}	0.1	quadratic	1.49×10^{-5}	0	0	unres.	unres.	unres.	3.91×10^{-3}
5	4.54×10^{-3}	0.05	quadratic	3.72×10^{-6}	0	0	unres.	unres.	unres.	3.91×10^{-3}
6	4.54×10^{-3}	0.1	quadratic	1.49×10^{-5}	0	0	unres.	unres.	unres.	3.91×10^{-3}

Gate of the sampled-training probe on the released Two-Stage cart-pole closed loop: factored Lie-derivative graph, backward CROWN, $\omega = 0$. Rows A to F: certificates of our class trained with the verifier's own bound in the loss (pre-registered ladder, seed 0; box cap 800,000, 20,000 CROWN calls); capacity is the number of ICNN units. Rows 1 to 6: the quadratic certificate $V = x^\top P x$, P from $A^\top P + PA = -I$ at the closed loop's linearization, at κ equal to 0.5, 0.5, 0.25, 0.25, 0.1, 0.1 of its provable rate $\lambda_{\mathrm{dec}} = 0.0454$ (box cap 400,000). b from the certified boundary query. Sampled $H(0)/b$: 400,000 (rows A to F) or 200,000 (rows 1 to 6) samples plus projected ascent, a lower estimate that can refute a level and never certify one, printed as 0 when below 1.0×10^{-6} in magnitude. PGD cx: counterexamples at $\omega = 0$ out of 4,096 starts at level $0.5b$. unres.: UNRESOLVED within the box cap; refuted: COUNTEREXAMPLE_TO_CONDITION through validated witnesses. f^*: the largest probed relative box width at which the residual clause clears $\kappa\, 0.5b$ at twelve shell points, the width test of the sampled-training probe unchanged (its median 9.77×10^{-4}).

with the missing neighbours of the end oscillators dropped, $k_0 = 1$, $k_c = 0.5$, $c_3 = 0.1$, $\alpha = 1$, $\beta = 0.5$, and $d_0 = 1$ in family v2 or 0.4 in family v1, the first pre-registration. The region is $[-1, 1]^n$, the disturbance is a box of size $\bar{w}$ on every velocity, so the disturbance term is $\bar{w} \sum_k |\partial V / \partial v_k|$ and the residual is affine in the size, and the certificate is $V = x^\top P x$ with P from the Lyapunov equation $A^\top P + P A = -I$ of the linearization, $\kappa = 0.25 / \lambda_{\max}(P)$, and the upper level b equal to 0.95 times a certified boundary minimum of V. The protocol is that of Section 5.2 under the hard budget: 7 sizes, the largest from a validated scan, 13 checks per size fixed by the independent run and replayed by the shared partition in natural order, with the budget per check 60,000 at two and three states and 960,000 at four, the first of three increasing budgets at which the check at zero disturbance and level $b/2$ certifies. That gate is the pre-registered rule for running a dimension.

Family v1 passed the gate at two and three states, with ratios 3.40 and 14.49, and failed it at four, where the check is unresolved at 960,000 bound calls with 466,357 residue boxes and none refuted. The damping was then raised from 0.4 to 1 after a development scan of the gate check over damping and coupling values, the amendment was dated and its predicted outcome written into the configuration before the family v2 protocol ran, and the v1 records are kept beside the v2 records. Table 31 gives both.

Table 31: Dimension scaling of the shared partition on the oscillator family under the hard budget: bound calls of independent verification (B0) and of the shared partition (B1) and their ratio, the checks at which the independent run ended at its budget, the checks the shared partition answered without a fresh bound call, the agreement counts against the independent run, soundness discrepancies, and the store and peak memory of the shared partition. Rows marked v1 are the first pre-registered family.

n	cap	B0 bound calls	B1 bound calls	ratio	B0 at cap	B1 at 0 calls	eq/str/ weak	sound- ness	store used (MiB)	peak RSS B1 (MiB)
2	60,000	1,197,545	359,323	3.33	4	53	91/0/0	0	36.7	633
3	60,000	4,043,847	501,839	8.06	30	53	65/26/0	0	58.9	647
4	960,000	79,119,190	6,603,521	11.98	41	70	54/37/0	0	875.4	2,321
5	960,000	NO-GO at the gate								
2 (v1)	60,000	1,954,726	574,599	3.40	9	46	91/0/0	0	58.6	716
3 (v1)	60,000	4,648,214	320,837	14.49	41	69	52/39/0	0	37.6	653
4 (v1)	960,000	NO-GO at the gate								

Interval backend; hard cap: every provider call is sliced so bound calls never exceed the per-check budget (the cap column: 60,000 bound calls per check at $n = 2, 3$, and at $n = 4$ the first cap of the ladder 60,000, 240,000, 960,000 at which the gate check certified), exceeded by none; family v2 unless marked v1 (the two families differ in the damping alone, 1 against 0.4); the shared partition in natural node order, repetition 0 (1 and 2 repetitions, bit-identical counters where repeated). 91 checks per instance (13 per size, 7 sizes). B0: independent verification, a fresh partition per check; B1: the shared partition; ratio: B0 over B1 bound calls. At cap: checks that ended at the budget; at 0 calls: B1 checks answered from the partition it already held. eq/str/weak: positions equal to, stronger than and weaker than B0; refutations lost 0 and soundness discrepancies 0 over every measured block. Store used: rows times bytes per row, $75 + 16n$ bytes; peak RSS: the B1 process (2,321 MiB at $n = 4$ against 1,066 for the B0 process). $n = 5$ (v2) and $n = 4$ (v1): the gate check (zero disturbance, half the outer level) did not certify at any cap of the ladder, UNRESOLVED at 960,000 and 960,000 bound calls with 328,581 and 466,357 residue boxes, none refuted, widest residue box 0.125 and 0.0625, so no sweep ran and $n = 6$ was not run. The paper's two-state instance under this protocol gives 8.41 (its recorded trace replayed under the cap, 8.42). Wall times not of record.

On family v2 the base check certifies at two, three and four states, at 443, 6,007 and 325,315 bound calls, and is unresolved at five states at all three budgets, with 328,581 residue boxes at 960,000 bound calls, the widest of width 0.125, none refuted and no validated counterexample among 5,000 centres, so six states were not run. Over the three dimensions the shared partition uses 3.33, 8.06 and 11.98 times fewer bound calls than independent verification, with 0 soundness discrepancies over 2,002 records, no lost refutation, and stronger answers at 0, 26 and 37 positions, every one of them a check the independent run left unresolved at its budget, and no weaker answer at any dimension. The growth of the ratio follows the independent run's budget saturation, 4, 30 and 41 of 91 checks at the budget, and the two three-state families are consistent with this dependence at a fixed dimension: family v1, with 41 checks at the budget, gives 14.49, and family v2, with 30,

gives 8.06. On the same protocol the primary two-dimensional certificate gives 8.41, and replaying its recorded list under the hard budget reproduces the corresponding entry of Table 14 exactly, 8.42.

The partition's final leaf count equals the shared partition's bound calls, and each row costs $75+16n$ bytes. The store used is 36.7, 58.9 and 875.4 MiB, with a peak resident set for the replay process of 633, 647 and 2,321 MiB. The store grows as the leaf count times the row size and, once the base check certifies, is the memory resource we expect to bind first beyond four states on the machine used here. The enclosure of Section 4, run at four states with the max-gap rule and 200,000 bound calls, certifies a width of 1.00 times b, which is vacuous, and spends about two percent of the sweep's bound calls. The price at that dimension was not measured. Wall times of this study were taken on a shared machine and are not reported as results. Either a tighter bound on the residual or a splitting rule that follows the residual's gradient could let the base check certify at five states, and neither changes the numbers a leaf stores or the three inequalities read from them.

D.15 The released cart-pole certificate

The instance freezes the released four-state cart-pole certificate W and controller of Li et al. (2025b), seed 0, on the released box scaled by $1/32$, with half-widths $(0.1125, 0.075, 0.375, 0.375)$ in position, angle and their rates. On the released box the check between the released levels 0.01 and 0.95 stays unresolved within the tested budget (Section D.13), and the scaled box is the largest of the tested scales at which it certifies. The search differs from that of the earlier cart-pole studies in three ingredients read from the configuration released with that certificate: the split axis is chosen by the magnitude of its coefficient in the linear bound of the residual times the box width, a box is cut by the linear bounds before it is split (the cut keeps only points no clause settles), and a child keeps the tighter of its own bound and its parent's. The relaxation of the derivative of the activation is our own, since the library's lower relaxation of that operator is not a valid bound on some narrow intervals, which we confirmed against exact extrema and through the released pipeline; none of the graphs behind the other results of this paper uses that operator.

On that box the base check at zero disturbance certifies between the released levels under the released condition, which carries no rate term, in 25,817 bound calls, and under the condition of Section 2 at rate 0.05 in 21,071, both with an empty residue. The certified boundary minimum of W on the scaled box is 0.005665 against $W(0) = 0.005615$ and against 0.9267 on the released box, so the scaled box contains no sublevel set above the lower released level, and what is verified on it is the decrease condition on the box, without an invariant set.

The first bisections of this instance returned the same certified level at every size, and a sampled threshold showed why: the smallest level the bisection could reach, 0.01023, lay above the threshold at every size up to twelve newtons, where the sampled values run from 0.006951 to 0.008104. A level grid that does not resolve the threshold returns its own lowest level, which a reader cannot tell from a flat profile. With the window lowered to 0.002 (Table 32) the certified level is 0.008172 at zero and at six newtons and 0.008686 at twelve, and the level certified at six newtons is refuted at twelve by validated points, so the threshold separates those two sizes. At eighteen newtons a multi-start ascent inside the shell above the level 0.25 validates 986 of 4,087 candidates, and the verifier's own check at that level returns a counterexample from those stored points at 0 bound calls, the largest level any of them refutes being 0.3220. The threshold therefore rises by more than 28 times between twelve and eighteen newtons.

A multi-start ascent of the same kind, run before every bisection step, settles the refuted steps at six and twelve newtons without a bound call. The row at zero newtons was bisected without the ascent and cost 20,191,547 bound calls, against 479,974 and 455,281 for the two rows bisected with it.

Under the released condition, which carries no rate term, the same bisection on the same box returns the certified levels 0.007658, 0.008172 and 0.008686 at zero, six and twelve newtons. The level rises at every size, and each level certified at one size is refuted at the next by validated points. At eighteen newtons the verifier's check at the level 0.25 returns a counterexample from stored points at 0 bound calls, as under our condition. The refuted level is more than 28 times the level certified at twelve.

In the reuse protocol on this box (lower block of Table 32) both methods answer the same 15 checks at 3 sizes with at most 2,000,000 bound calls per check. The shared partition uses 1.41 to 1.82

Table 32: The threshold of the released cart-pole certificate on the scaled box, per size of a disturbance on the force channel: the smallest certified level of a bisection over the window $[0.002, 0.01023]$ at resolution 5.14×10^{-4}, the largest level refuted by validated points, a sampled threshold, which is not certified, the bound calls of the row and the answers of its steps (C certified, X validated counterexample, U unresolved). At the largest size only the refutation was run. The lower block is the reuse protocol on the same box: bound calls of independent verification (B0) and of the shared partition (B1) per condition and check order, their ratio, the ratio against the cold independent search, the agreement counts and the checks answered without a bound call.

condition	size (N)	certified level	verified lower bound	sampled threshold	bound calls	step codes
rate 0.05	0	0.008172	0.006966	0.006951	20,191,547	CXCUU
	6	0.008172	0.007743	0.007727	479,974	CXCXX
	12	0.008686	0.008211	0.008104	455,281	CXXCC
	18	–	0.25	0.3274	0	X
rate-free	0	0.007658	0.006398	0.006484	2,433,756	CXCUC
	6	0.008172	0.007743	0.007727	415,344	CXCXX
	12	0.008686	0.008205	0.008104	485,449	CXXCC
	18	–	0.25	0.3274	0	X

condition	order	B0 bound calls	B1 bound calls	ratio	cold ratio	eq/str/ weak	sound- ness	lost refut.	B0 at 0 calls	B1 at 0 calls asc./pool/part.
rate 0.05	natural	6,369,183	3,505,165	1.82	2.34	15/0/0	0	0	6	6/0/0
	decreasing	6,369,183	3,746,634	1.70	2.19	15/0/0	0	0	6	6/0/2
	permuted	6,369,183	4,507,894	1.41	1.82	15/0/0	0	0	6	6/0/3
rate-free	natural	5,359,950	3,033,877	1.77	2.36	15/0/0	0	0	6	6/0/0
	decreasing	5,359,950	3,218,151	1.67	2.22	15/0/0	0	0	6	6/0/2
	permuted	5,359,950	3,827,844	1.40	1.87	15/0/0	0	0	6	6/0/3

Released Two-Stage cart-pole certificate, seed 0, on the released box scaled by 1/32 (half-widths 0.1125, 0.075, 0.375, 0.375); band (0.01, 0.95). The base check at the band's inner level and zero disturbance certifies under our condition at rate 0.05 in 21,071 and under the rate-free condition in 25,817 bound calls. Per size: one probe at 0.999 of the window's top, then 4 bisection midpoints of the level window [0.002, 0.01023], resolution 5.14×10^{-4}, at most 20,000,000 bound calls and 1200 s per step. Before a step, a search for a violating point in the step's shell runs first, and a validated hit answers the step with no bound call (four steps of the 0 N row at rate 0.05 predate that search and ran under the earlier limit of 7200 s). Search: coefficient-driven branching, parent-slack order, domain clipping at the outer level, the inner level and the residual cut, clipped children, the no-spill rule, the witness pool offered before any bound call and the full residue scan. Certified level: the smallest level a completed check certified, an upper bound on the threshold. Verified lower bound: the largest level refuted by a point the sound bound computation validates. Sampled threshold: a lower estimate from sampling and ascent, which certifies nothing. Step codes, one per step: C certified, X refuted by a validated counterexample, U unresolved at the step's limit, which is a statement about the budget and not a refutation. 18 N, rate 0.05: no bisection; the level 0.25 was refuted by the verifier's own check at 0 bound calls from 986 validated points of 4,087 candidates, and the largest level any of them refutes is 0.3220. 18 N, rate-free: no bisection; the level 0.25 was refuted by the verifier's own check at 0 bound calls from 987 validated points of 4,090 candidates, and the largest level any of them refutes is 0.3258. Reuse protocol: 15 checks (3 sizes), at most 2,000,000 bound calls per check for both methods, and the identical search for a violating point before every check of both methods, charged to neither. Both methods clip at the outer level only, because the inner-level and residual cuts depend on the check's level and size, and a stored record serves checks at other levels and sizes. B0: independent verification in the cached mode of B1, with the partition and the witness pool emptied before every check, so that it recomputes the joint bound of a leaf at every positive size as B1 does; B1: the shared partition, which keeps both. B0 runs the recorded frontier rule, no recomputation of the joint bound at size zero, the witness pool offered before any bound call and the capped residue scan; B1 runs the no-spill rule, no recomputation of the joint bound at size zero, the witness pool offered before any bound call and the full residue scan. Ratio: B0 over B1 bound calls. Cold ratio: the bound calls of independent verification in cold mode over B1's; that run keeps nothing between checks, never recomputes the joint bound, and its answers fixed the list of checks. B0 returns the cold run's answer at every check. eq/str/weak: positions equal to, stronger than and weaker than the cold run's answer. At 0 calls: checks answered with no bound call; for B1 split by cause, that check's own search (asc.), the stored witness pool (pool), the stored partition (part.). At 0 N, rate 0.05, the independent run of the reuse protocol, which has no time limit per check, certified 0.007658, a level the profile above left unresolved at its limit. Wall times are not quoted.

times fewer bound calls than independent verification at rate 0.05 and 1.40 to 1.77 times fewer under the rate-free condition, and every answer equals the independent run's. The independent run here recomputes the joint bound at every positive size, as the shared partition does. An independent search that never recomputes it costs more, and against that search the ratios run from 1.82 to 2.36. Under both conditions one check at zero disturbance, at the level 0.007143 between the largest refuted and the smallest certified level, ends unresolved at the budget for both methods, as the bisection of the profile did at its time limit. That check accounts for 31.4 and 37.3 percent of the independent run's bound calls under the two conditions, and over the other checks the ratio runs from 1.74 to 2.90 at rate 0.05 and from 1.84 to 3.25 under the rate-free condition.