SHARP SAMPLE COMPLEXITY FOR LYAPUNOV-BASED CONTROL OF UNKNOWN SYSTEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning control from offline demonstrations requires transferring certificate guarantees to a replacement controller's trajectories. We study control-affine systems using one fixed batch of N independent expert episodes, including recorded controls. The method jointly learns a certificate and a drift predictor, then synthesizes constrained feedback through a soft control-Lyapunov quadratic program, without further plant interaction. A bounded, one-sided warning loss measures expert-margin deficits and harmful directional model errors. Under explicit encounter coverage, its population value bounds the probability of any certificate-decrease violation over a prescribed deployment horizon. Soundness requires no expert stability premise; useful guarantees additionally require an in-class decrease witness and directional approximation. We give same-batch bounds for parametric and smooth reproducing-kernel classes. For a fixed library of K certificates and an affine predictor with p coefficients, a convex scenario formulation yields a report of order $O\left(\frac{p+\log(K/\delta)}{\nu_{\mathrm{enc}} N}\right)$ at zero optimal warning slack. We quantify the probability of passing a requested risk tolerance. Separately, spatial logging coverage gives practical stability outside a ball whose radius shrinks with the loss bound. A realizable control family establishes matching sample, dimension, and coverage dependence within prescribed classes and synthesis architecture, despite nonvanishing minimax uniform drift-estimation error. Experiments illustrate directional learning and the distinct roles of source confidence and deployment coverage.

1 INTRODUCTION

Expert demonstrations offer a starting point for controlling an unknown system: learn a model and a certificate, then use them to synthesize feedback. The central difficulty is transferring evidence from those demonstrations to the replacement controller's own trajectories. Low training loss alone cannot establish this transfer, because the expert and learned controller may visit different states.

Learning certificates from trajectories provides a statistical framework for reasoning about unseen trajectories (Boffi et al., 2021; Zhang et al., 2022). Controller synthesis introduces an additional dependence: both the certificate being tested and the flow on which it is tested are learned from the data. Full-vector model accuracy can be more information than this task needs: the certificate derivative depends on model error only through its projection onto the certificate gradient, with a particular harmful sign.

We jointly learn a certificate V and drift predictor F, then convert them into feedback through a fixed control-Lyapunov-function quadratic program (CLF-QP). Every true decrease violation forces a scalar *warning* above a known threshold. The warning combines the expert's certificate-decrease deficit with underestimated drift along ∇V. Exact derivative and action records make it observable at logged states.

This objective ignores tangential and favorably signed errors and tolerates small harmful errors within a margin. To transfer logged evidence to the replacement flow, we impose *encounter coverage*: deployed warning encounters must be controlled by logged population loss, uniformly over eligible pairs. Its valid lower bound ν_{enc} enters reporting; fitting and feedback do not require it.

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.

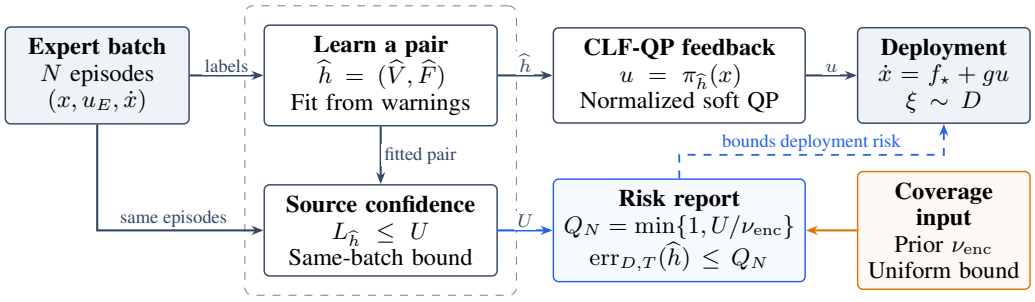

Figure 1: One-batch learning and deployment reporting. The learned certificate–predictor pair defines CLF-QP feedback. A source-loss bound U and justified coverage ν_{enc} give the trajectory-risk report Q_N. Deployment supplies no training data; the dashed link denotes the guarantee.

2 Related Work

Trajectory certificates and replacement control. Boffi et al. (2021) give fast realizable rates for learning parametric and kernel stability certificates from trajectories, with geometric conditions yielding regional stability. Zhang et al. (2022) extend this analysis to perturbed dynamics under incremental stability. Here the same batch learns both the certificate and a predictor whose synthesized controller changes the flow being certified.

Model-based and control-oriented learning. Joint certificate–controller synthesis is established through queryable demonstrators (Ravanbakhsh & Sankaranarayanan, 2019), neural learner–falsifier methods (Chang et al., 2019), mixed-integer verification (Dai et al., 2021), verified output feedback (Yang et al., 2024), and robust Lyapunov–barrier QPs (Dawson et al., 2022). For unknown plants, spatial model-error bounds transfer learned-model verification to true dynamics (Zhou et al., 2022). Offline precedents learn models and Lyapunov terminal costs (Mittal et al., 2020), policies and contraction metrics (Rezazadeh et al., 2022), or models, controllers, and Lyapunov functions jointly (Min et al., 2023). Control-oriented identification and certificate-directional uncertainty also predate this work (Singh et al., 2021; Dean et al., 2020; Taylor et al., 2019; 2021). Our statistical target is true replacement-flow risk from passive episodes using a learned, one-sided warning, without requiring uniform vector-model accuracy.

Demonstrations and deployment shift. DAgger queries the expert at learner-induced states (Ross et al., 2011). With fixed demonstrations, stability-constrained imitation and TaSIL control trajectory deviations through incremental gain or input-to-state stability (Tu et al., 2022; Pfrommer et al., 2022). We instead certify decrease, using encounter coverage for transfer and an in-class expert witness only for useful low-loss reports. PAC-Bayes approaches evaluate policies in sampled environments (Majumdar et al., 2021; Ren et al., 2021); our learner uses the supplied expert records alone.

Coverage and statistical tools. Concentrability and data-supported policy coverage govern offline transfer (Munos & Szepesvári, 2008; Chen & Jiang, 2019; Rashidinejad et al., 2021); function-class-restricted coefficients test relevant Bellman or model errors (Xie et al., 2021; Uehara & Sun, 2022). Our condition tests warnings generated by eligible learned pairs and their trajectory maxima, which occupancy averages alone do not control. Low-loss learning bounds (Srebro et al., 2010) and convex scenario theory (Calafiore & Campi, 2006; Campi & Garatti, 2008) supply the statistical tools. The control-specific contribution is the observable warning and its transfer to the synthesized controller's decrease event.

Contributions. We establish four linked results: (i) a CLF-QP comparison that turns decrease violations into observable directional warnings, with encounter coverage transferring source loss to replacement-flow risk (Theorem 1); (ii) same-batch parametric, RKHS, and convex scenario bounds, including numerical correction and probabilities of meeting a requested tolerance (Theorems 2, 3, 6); (iii) a source-density route to an invariant sublevel and a shrinking exceptional ball (Proposition 5); and (iv) matching minimax dependence on sample size, dimension, and inverse coverage within a

realizable control family, despite persistent uniform drift-estimation error (Theorem 4). Soundness uses regularity, containment, and the stated transfer premise. Useful reports additionally require an in-class expert decrease witness and directional approximation, without expert contraction or incremental stability.

3 LEARNING FRAMEWORK

Data and objective. We consider $\dot{x} = f_\star(x) + g(x)u$, where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^{m_u}$, the drift $f_\star$ is unknown, and the input map g is known. A fixed expert feedback u_E generates a batch $S = (E_1, \dots, E_N)$ of N independent episodes, starting from $\xi_i \sim D$. Each episode contains finitely many exact records $(x_{ij}, u_{ij}, \dot{x}_{ij})$ collected by a fixed episode-local rule. The learner reconstructs $y_{ij} = \dot{x}_{ij} - g(x_{ij})u_{ij} = f_\star(x_{ij})$ at these states. It receives no additional plant queries or off-record expert actions. Within-episode records may be dependent; throughout, N counts episodes.

A candidate is $h = (V, F) \in \mathcal{V} \times \mathcal{F}$, where V is a certificate and F predicts the drift; hats denote the fitted output $\widehat{h} = (\widehat{V}, \widehat{F})$. A prescribed synthesis map produces feedback π_h. Write $\phi_t^h(\xi)$ for the solution of the true deployed system $\dot{x} = f_\star(x) + g(x)\pi_h(x)$, and let $R_h(x) = \nabla V(x)^\top [f_\star(x) + g(x)\pi_h(x)] + \lambda V(x)$, with a prescribed decay rate $\lambda > 0$. Fix a finite deployment horizon $T > 0$. The learning target is

$$\mathrm{err}_{D,T}(h) = \mathbb{P}_{\xi \sim D}\left\{\exists t \in [0, T] : R_h(\phi_t^h(\xi)) > 0\right\}. \tag{1}$$

A useful report bounds this risk for the pair selected from the training batch. The probability in (1) is over a fresh initial state; confidence in that bound is over the training episodes.

Vector norms are Euclidean and matrix norms are induced unless marked $\|\cdot\|_F$ (Frobenius). Write $B_R = \{x : \|x\| < R\}$.

Assumption 1 (Dynamics)**.** On a ball $\Omega = B_R$ containing $\mathrm{supp}(D) \subseteq B_{R_0}$, the drift $f_\star$ and expert field $F_E = f_\star + gu_E$ are Lipschitz and vanish at the origin; g is bounded and Lipschitz. For fixed positive caps U_j, C_j, the expert satisfies $|u_{E,j}(x)| \le \min\{U_j, C_j\|x\|\}$.

Assumption 2 (Learned classes)**.** The classes are fixed before sampling and have uniform regularity bounds:

- Each $V \in \mathcal{V}$ has a Lipschitz Hessian, $V(0) = 0$, $\nabla V(0) = 0$, and $\alpha\|x\|^2 \le V(x) \le \bar{\alpha}\|x\|^2$ for fixed $0 < \alpha \le \bar{\alpha}$.
- Each $F \in \mathcal{F}$ has a bounded, Lipschitz derivative and satisfies $F(0) = 0$.

Positivity prevents the learner from reducing its loss by shrinking V to zero. The origin conditions make normalized directional errors well behaved near the equilibrium. The uniform derivative bounds used in the proofs are collected in Appendix A.

We treat bounded Lipschitz-parametric classes, including smooth neural classes with enforced positivity and origin constraints (Proposition 6), and smooth RKHS balls (Appendix E). A finite certificate library with affine predictors admits the convex specialization below.

Feedback synthesis. For $r = \|x\| > 0$, let $\mathcal{U}(x) = \{u : |u_j| \le \min(U_j, C_j r)\}$ and define $\pi_h(x)$ as the input component of

$$\begin{aligned} \underset{u \in \mathcal{U}(x),\ t \ge 0}{\text{minimize}} \quad & \frac{\|u\|^2}{2r^2} + \frac{\rho t^2}{2} \\ \text{subject to} \quad & \frac{\nabla V(x)^\top (F(x) + g(x)u) + \lambda V(x)}{r^2} + m \le t. \end{aligned} \tag{2}$$

Set $\pi_h(0) = 0$. Here $m > 0$ is a decrease margin and $\rho > 0$ penalizes slack. This is a normalized soft CLF-QP, using the standard optimization-based synthesis principle (Ames et al., 2014). It asks for modeled decrease with margin while respecting the actuator caps. The nonnegative slack t keeps the optimization feasible for every learned pair; the warning bound below determines its deployment guarantee.

Assumption 3 (Deployment)**.** Write $C = \|(C_1, \dots, C_{m_u})\|$ and choose $C/\sqrt{\rho} < m$. All expert and selectable replacement trajectories from $\mathrm{supp}(D)$ remain in the interior of Ω on $[0, T]$.

The design inequality is a convenient sufficient condition giving a detection threshold greater than $3m/4$. The containment requirement may be established by a uniform growth bound or a sharper system-specific bound. For example, $R_0 e^{(L_f+GC)T} < R$ suffices when L_f is a Lipschitz constant for $f_\star$ and G bounds $\|g\|$. Under these assumptions the QP has a unique optimizer and induces a Lipschitz feedback with a well-defined flow (Lemma 1). We next identify what this feedback needs the learned model to get right.

Directional warnings. Write $[z]_+ = \max\{z, 0\}$. For $x \neq 0$, define the expert score $q_E(V, x) = (\nabla V(x)^\top F_E(x) + \lambda V(x))/\|x\|^2$ and the directional model error $d_h(x) = \nabla V(x)^\top (f_\star(x) - F(x))/\|x\|^2$. The first measures how the expert changes the candidate certificate; the second is positive when the predictor underestimates the drift in a direction that increases it. Our warning is

$$s_h(x) = b_V(x) + [d_h(x)]_+, \qquad b_V(x) = [q_E(V, x) + m]_+. \tag{3}$$

All scores, including b_V and s_h, are set to zero at the origin. At a record, $F_E(x) = \dot{x}$ and $f_\star(x) = y$, so (3) is exactly computable for every candidate pair.

The CLF-QP makes this warning useful. With the design threshold $\theta = m - C^2/(4\rho m) > 0$, every state with $R_h(x) > 0$ must satisfy $s_h(x) > \theta$ (Lemma 3). This follows from comparing the QP solution with the expert's admissible action. Thus the warning need not measure the entire model error: its directional term ignores tangential errors and favorable overestimation. A certificate with a large expert deficit still pays for that deficit. Appendix B gives the complete comparison and detection proofs.

A bounded loss. Write $\mathrm{clip}(z, \ell, u) = \min\{u, \max\{\ell, z\}\}$ for clipping to an interval $[\ell, u]$. Choose thresholds $0 \le a < b < \theta$, fixed before sampling; $a = \theta/4$ and $b = \theta/2$ are convenient defaults. Define the logged-warning ramp $\psi_{\mathrm{lo}}(s) = \mathrm{clip}((s-a)/(b-a), 0, 1)^2$ and the deployed-encounter ramp $\psi_{\mathrm{hi}}(s) = \mathrm{clip}((s-b)/(\theta-b), 0, 1)^2$. An episode contributes its largest recorded warning, $H_h(E) = \max_j \psi_{\mathrm{lo}}(s_h(x_j))$. Its population and empirical losses are $L_h = \mathbb{E}_E H_h(E)$ and $\widehat{H}_h = N^{-1}\sum_{i=1}^N H_h(E_i)$. Empty episodes have loss zero. Boundedness controls the statistical scale; squaring the ramp makes $\sqrt{H_h}$ Lipschitz in the warning. The lower threshold detects logged warnings before deployed violations (Figure 3).

Encounter coverage. For a fixed pair let $S_h(\xi) = \sup_{t\in[0,T]} \psi_{\mathrm{hi}}(s_h(\phi_t^h(\xi)))$. This is a population quantity used in the analysis, not an extra observation available to the learner. We abbreviate this premise as *encounter coverage* (*EC*).

Assumption 4 (Coverage)**.** A supplied lower bound $\nu_{\mathrm{enc}} \in (0, 1]$ satisfies, for every pair in a predeclared eligible class,

$$\mathbb{E}_{\xi\sim D} S_h(\xi) \;\le\; \frac{L_h}{\nu_{\mathrm{enc}}}. \tag{4}$$

This function-class-restricted comparison covers both the changed flow and the logging schedule (Xie et al., 2021). A larger ν_{enc} means that records capture deployed warnings more effectively; expert stability and independent resets do not establish it. Eligibility is fixed before fitting, or certified through a simultaneous source bound (Proposition 3). Occupation coverage and invariant warning sets provide sufficient routes (Appendix C). Without justified coverage, the strict deployment report is $Q_N = 1$; spatial conclusions remain available under their separate premises.

4 Generalization Guarantees

Fix a confidence failure probability $\delta \in (0, 1)$. We first transfer population warning loss to deployment risk, then bound that loss when V and F are fitted to the same episodes.

Theorem 1 (Deployment guarantee)**.** *Under Assumptions 1–4, every eligible h satisfies* $\mathrm{err}_{D,T}(h) \le \min\{1, L_h/\nu_{\mathrm{enc}}\}$. *Suppose a source-loss bound U_h satisfies $L_h \le U_h$ simultaneously for all selectable outputs on an event of training probability at least $1-\delta$. Then any eligible pair $\widehat{h}$ selected from those outputs has the sound report $Q_N = \min\{1, U_{\widehat{h}}/\nu_{\mathrm{enc}}\}$ with $\mathbb{P}_{\mathrm{batch}}\{\mathrm{err}_{D,T}(\widehat{h}) > Q_N\} \le \delta$. On this confidence event, initial states with D-probability at least $1 - Q_N$ satisfy $\widehat{V}(\phi_t^{\widehat{h}}(\xi)) \le e^{-\lambda t}\widehat{V}(\xi)$ throughout $[0, T]$.*

Simultaneous validity permits fitting and reporting from the same batch. At a requested risk tolerance $\epsilon \in (0,1)$, the report passes when $Q_N \le \epsilon$; the chance of passing with true risk above ϵ is at most δ. The guarantee is over training batches, without conditioning confidence on the event of passing. Appendix C proves the theorem and the associated state-norm decay bound.

Parametric and kernel classes. For the parametric case, write $r_h(E) = \sqrt{H_h(E)}$. Suppose $h = h_\vartheta$, $\|\vartheta\| \le R_H$ in $\mathbb{R}^q$, and $\sup_E |r_{h_\vartheta}(E) - r_{h_{\vartheta'}}(E)| \le A_H \|\vartheta - \vartheta'\|$. The constant $B = 2A_H R_H$ groups the parameter radius and loss sensitivity; it does not depend on the batch.

Theorem 2 (Parametric learning)**.** *Under this root-loss modulus, define* $\Gamma_N = [8 + 4q\log(1 + B\sqrt{N}) + 4\log(1/\delta)]/N$. *With probability at least* $1-\delta$*, simultaneously over the parametric class,*

$$L_h \le U_h := \min\left\{1, \left(\sqrt{\widehat{H}_h} + \sqrt{\Gamma_N}\right)^2\right\}. \tag{5}$$

If the empirical optimization tolerance is $\tau \ge 0$*,* $\widehat{H}_{\widehat{h}} \le \inf_h \widehat{H}_h + \tau$ *and a fixed comparator has population loss* ℓ*, then, for* $\beta \in (0,1)$*, with probability at least* $1 - \delta - \beta$*, the deployment report is sound and obeys*

$$Q_N \;\le\; \min\left\{1, \frac{1}{\nu_{\mathrm{enc}}}\left[\ell_\tau + 8\sqrt{\frac{\ell_\tau \mathcal{C}_N}{N}} + \frac{16\mathcal{C}_N}{N}\right]\right\}, \quad \ell_\tau = \ell + \tau, \tag{6}$$

where $\mathcal{C}_N = 1 + q\log(1 + B\sqrt{N}) + \log(1/(\delta\beta))$.

With a zero-loss comparator and exact empirical fitting, the report scales as $\widetilde{O}(q/(\nu_{\mathrm{enc}}N))$, where $\widetilde{O}$ suppresses logarithmic factors. When $\ell_\tau > 0$, the statistical terms decrease toward an explicit loss floor. Covering the root episode loss permits parameter resolution $N^{-1/2}$ without counting correlated records as independent samples. Appendix D gives the exact KL inversion underlying (5), a sharper computable report, and the proof of (6).

For smooth RKHS classes, the same warning construction supports an optimistic bound in terms of norm and feature complexity (Proposition 7). For the episode maximum, the stated full-RKHS result requires a deterministic upper bound on records per episode; its complexity includes that dependence. A weighted-average episode loss gives an alternative without that record-count factor, with coverage stated for that loss. Finite-rank RKHS models may use Theorem 2 directly. The numerical value depends explicitly on these norm and feature constants; the convex specialization below uses additional optimization structure.

When useful reports are possible. A statistical guarantee can be valid even when the available class fits poorly. The following expert condition isolates the certificate part of a useful comparator.

Assumption 5 (Expert witness)**.** There exists $V_\star \in \mathcal{V}$ such that $q_E(V_\star, x) \le -m$ at every nonzero recorded state of a population episode, almost surely.

This witness is needed for a small comparator loss, not for validity of Theorem 1. It is an in-class decrease condition on the logged law. If it holds throughout a region containing the expert trajectories, it is the familiar regional exponential Lyapunov condition; incremental stability is not needed. For an admissible comparator predictor $F_c \in \mathcal{F}$, this assumption makes the expert deficit vanish, and comparator loss depends on F_c only through positive directional errors beyond a. In particular, $d_{(V_\star, F_c)} \le a$ at all recorded states almost surely gives $\ell = 0$, even with large full-vector model error. The model approximation condition is needed in addition to the expert witness. The resulting passing probabilities are given below and proved in Appendix D.4.

A convex specialization. Suppose $\mathcal{V} = \{V_1, \dots, V_K\}$ is a fixed certificate library and F_W is affine in p coefficients W, constrained by $\|W\| \le R_F$ with $R_F > 0$. The basis and library are fixed before sampling, and all possible fitted outputs must be eligible for the coverage premise. For each k, $G_k(W; E) = \max_j s_{(V_k, F_W)}(x_j)$ is convex in W (zero for an empty episode). Using lexmin to mean lexicographic minimization of the two displayed objectives, fit each candidate by

$$\operatorname*{lexmin}_{\|W\| \le R_F,\ \sigma \ge 0} (\sigma, \|W\|^2) \quad \text{subject to} \quad G_k(W; E_i) \le a + \sigma, \quad i = 1, \dots, N. \tag{7}$$

Algorithm 1 One-batch certificate learning, feedback synthesis, and reporting

Input: Expert episodes S, fixed $\{V_k\}_{k=1}^K$ and $\{F_W : \|W\| \le R_F\}$, known g, QP design (2), $0 \le a < b < \theta$, $\delta, \epsilon \in (0,1)$, and an optional justified ν_{enc}.

1. **Fit.** Reconstruct $y_{ij} = \dot{x}_{ij} - g(x_{ij})u_{ij}$. For each k, solve the two-stage program (7) exactly to obtain (σ_k, W_k).
2. **Bound.** Compute q_N by (8) if $N > p$, or set $q_N = 1$. For every k, set $c_k = \mathrm{clip}(\sigma_k/(b-a), 0, 1)^2$ and $U_k = q_N + (1 - q_N)c_k$.
3. **Select.** Choose $k_\star$ by lexicographic order on $(\sigma_k, \|W_k\|^2, k)$. Set $\widehat{h} = (V_{k_\star}, F_{W_{k_\star}})$, $U = U_{k_\star}$, and synthesize $\pi_{\widehat{h}}$ by (2), with $\pi_{\widehat{h}}(0) = 0$.
4. **Report.** If a justified ν_{enc} covers the selectable outputs, set $Q_N = \min\{1, U/\nu_{\mathrm{enc}}\}$; otherwise set $Q_N = 1$. Set *pass* $= \mathbf{1}\{Q_N \le \epsilon\}$.

Return on every batch: $(\widehat{h}, \pi_{\widehat{h}}, U, Q_N, \mathit{pass})$. Numerical fits use the correction or fallback below.

The first stage minimizes the common slack; the second selects its unique minimum-norm predictor. Write $b_{k,ij} = b_{V_k}(x_{ij})$ and $d_{k,ij}(W) = d_{(V_k,F_W)}(x_{ij})$. Each record constraint is equivalent to two affine inequalities: $b_{k,ij} \le a + \sigma$ and $d_{k,ij}(W) \le a + \sigma - b_{k,ij}$. Both are needed because model fitting cannot repair a fixed expert deficit. Sufficient slack ensures feasibility.

Theorem 3 (Scenario learning)**.** *Let (W_k, σ_k) be the exact solutions of (7). For $N > p$, define $q_N \in (0,1)$ by*

$$\sum_{j=0}^{p} \binom{N}{j} q_N^j (1 - q_N)^{N-j} = \delta/K, \qquad c_k = \mathrm{clip}\left(\frac{\sigma_k}{b-a}, 0, 1\right)^2. \tag{8}$$

For $N \le p$, set $q_N = 1$. With probability at least $1 - \delta$, all K fitted pairs satisfy $L_{(V_k,F_{W_k})} \le U_k := c_k + (1 - c_k)q_N$. Selecting k to minimize U_k is valid on the same event. With zero selected slack, the resulting report obeys $Q_N = \min\{1, q_N/\nu_{\mathrm{enc}}\} = O((p + \log(K/\delta))/(\nu_{\mathrm{enc}}N))$.

The scenario program has p coefficients and one slack; certificate choices cost $\log K$. This applies the convex scenario theorem of Campi & Garatti (2008); the warning constraint and its deployment transfer are control-specific. Positive slack contributes the explicit floor c_k. A uniformly low-warning comparator ensures low slack, whereas small mean loss alone does not. Appendix D proves the application and a numerical optimizer-gap extension. Arbitrary joint nonconvex training needs a different source bound.

Implementing the learner. The K programs are independent. The bound $\sigma_k^{\mathrm{lb}} = [\max_{i,j} b_{k,ij} - a]_+$, with $\max \varnothing = 0$, permits screening when it exceeds a feasible incumbent slack; tied certificates remain eligible, and confidence still uses the original K.

For a numerical predictor $\widetilde{W}_k$, let $\overline{\sigma}_k$ bound the exact first-stage optimum and let ζ_k certify a uniform warning difference from its exact minimum-norm predictor. Lemma 7 gives the corrected source report

$$\widetilde{c}_k = \min\left\{1, \left(\frac{\overline{\sigma}_k + \zeta_k}{b-a}\right)^2\right\}, \qquad \widetilde{U}_k = \widetilde{c}_k + (1 - \widetilde{c}_k)q_N. \tag{9}$$

The returned coefficients must be admissible, and solver termination alone does not certify ζ_k. Uncertified candidates receive $U_k = 1$; if all solves fail, return the prescribed admissible pair (V_1, F_0), where F_0 has zero coefficients. Either select corrected reports or apply the slack/norm selector and correct its output, as in the experiments. Tail inversion and reporting require outward upper bounds. General parametric/RKHS learners instead minimize empirical warning loss and use their simultaneous source bound; their optimization error affects usefulness.

Probability of a useful report. Let $r = \epsilon\nu_{\mathrm{enc}}$. For the parametric report, passing requires $\Gamma_N \le r$ and is equivalent to $\widehat{H}_{\widehat{h}} \le t_N = (\sqrt{r} - \sqrt{\Gamma_N})^2$. If $\widehat{H}_{\widehat{h}} \le \widehat{H}_{h_c} + \tau$ for a fixed comparator of population loss ℓ, then $t_N - \tau > \ell$ gives

$$\mathbb{P}_S\{\mathit{pass}\} \ge 1 - \exp[-N\,\mathrm{kl}(t_N - \tau \| \ell)],$$

where kl is binary relative entropy. For $\ell = 0$ and $\tau \le t_N$, passing is almost sure. A passing probability $1-\beta$ combines with soundness to give joint probability at least $1-\delta-\beta$ of passing with risk at most ϵ.

For exact scenario fitting, passing requires $q_N \le r$ and is equivalent to $\sigma_{k_\star} \le (b-a)\sqrt{(r-q_N)/(1-q_N)}$. A fixed comparator with $G_k(W;E) \le a$ almost surely ensures zero slack. Then $N \ge 2[p+\log(K/\delta)]/(\epsilon\nu_{\mathrm{enc}})$ suffices for almost-sure passing. Numerical outputs additionally require the certified effective slack $\overline{\sigma}_k + \zeta_k$. Appendix D.4 proves these statements, endpoint cases, and positive-loss floors; related probabilistic feasibility guarantees appear in Tu et al. (2022, Theorem 4.4).

A shrinking region around the equilibrium. A source-loss bound also gives a spatial guarantee without *EC*. Let μ_E be a fixed weighted law of logged states, so $\mathbb{E}_{\mu_E}\psi_{\mathrm{lo}}(s_h) \le L_h$. Suppose it has density at least $c_\mu > 0$ on a compact region $Y \subset \Omega$, and $R_{\widehat{h}}$ is L_R-Lipschitz there. For $0 < U < 1$ with $L_{\widehat{h}} \le U$, fix $\kappa > 0$, $\eta_g \in (0,1)$, and let v_n be the volume of the unit ball. Define

$$\varrho_U = \left(\frac{(1+\kappa)U}{c_\mu v_n}\right)^{1/n}, \qquad \Delta_U = L_R\varrho_U, \qquad r_{\mathrm{bad}} = \sqrt{\frac{\Delta_U}{\eta_g\lambda\alpha}}. \tag{10}$$

At every state whose radius-ϱ_U ball lies in Y, the residual is at most Δ_U: otherwise that whole ball would carry warning loss one and mass greater than U. If a compact full sublevel $K_c = \{x \in \Omega : \widehat{V}(x) \le c\}$ lies in this trimmed region and $c > \Delta_U/\lambda$, it is forward invariant, and for every $\xi \in K_c$ and $t \ge 0$,

$$\widehat{V}(\phi_t^{\widehat{h}}(\xi)) \le e^{-\lambda t}\widehat{V}(\xi) + \frac{\Delta_U}{\lambda}(1-e^{-\lambda t}). \tag{11}$$

Outside the radius-r_{bad} ball, $\dot{\widehat{V}} \le -(1-\eta_g)\lambda\widehat{V}$. With fixed constants, $r_{\mathrm{bad}} = O(U^{1/(2n)})$ and hence $\widetilde{O}((p/N)^{1/(2n)})$ at a realizable scenario source-loss rate. This is practical stability on a certified sublevel, not exclusion of near-origin violations from (1). The density, residual bound, and sublevel containment are additional requirements. Proposition 5 proves the statement and handles $U = 0$; Proposition 4 gives the complementary finite-horizon radius $O(Q_N^{1/(2n)})$ from a valid deployment-risk bound and flow sensitivity.

Sharp sample and coverage dependence. The upper bounds expose the cost of rare informative episodes. The next result shows that this cost corresponds to an actual control-learning obstruction, even with a stable expert and a realizable drift class.

Theorem 4 (Minimax rate)**.** *For every fixed $T > 0$, $d \ge 1$, and $\nu_0 \in (0, 1/2]$, there is a realizable family $\mathfrak{P}$ in dimension $n \ge 2d+1$ satisfying the preceding system assumptions and uniform coverage at least ν_0. It uses one terminal action-and-derivative record per episode and the declared d-parameter predictor class with the QP (2). For $N+1 > d/\nu_0$, universal constants $c_1, c_2 > 0$ satisfy*

$$c_1\frac{d}{\nu_0(N+1)} \le \inf_{\mathcal{A}}\sup_{P\in\mathfrak{P}}\mathbb{E}_{S,\mathcal{A}}\,\mathrm{err}_{D,T}(\mathcal{A}(S)) \le c_2\frac{d}{\nu_0(N+1)}. \tag{12}$$

The infimum ranges over randomized learners returning pairs in these classes through the prescribed QP. The same family retains a constant lower bound on the worst-case expected uniform drift-estimation error.

The matching rates in (12) concern this family and architecture. Rare informative records reveal hidden drift choices that affect deployment. Appendix F gives the construction and precise quantifiers, including a confidence–passing tradeoff. Logging and output restrictions are essential: initial expert actions reveal the hidden choices, and unrestricted feedback can solve this family without identifying them. The result does not assert optimality of every factor in the general RKHS or high-probability bounds.

5 Experiments

We separate three questions: whether the finite optimization supports its source report, whether the learned pair decreases along numerical deployment trajectories, and whether a continuous-state verifier certifies the frozen controller. The first study checks the scenario implementation; a

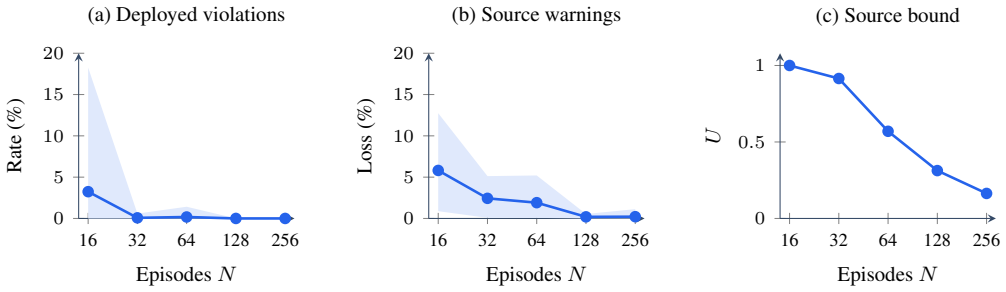

Figure 2: Independent scenario study: numerical violation rate, independently evaluated source warning loss, and 95% per-fit scenario source bound U, including certified optimizer correction. Curves and shading show the mean and range over eight batches. These source bounds are distinct from deployment confidence; the coverage-dependent offline report remains $Q_N = 1$ without a justified encounter coefficient.

Table 1: Detected decrease violations in the paired campaign, totaled over three fitted pairs and 128 starts per pair. The original release is the reporting run and is not pooled.

	Pendulum		Cart-pole	
Method	$N = 128$	$N = 1024$	$N = 128$	$N = 1024$
OURS-scenario	13/384	0/384	144/384	0/384
OURS-ERM	0/384	0/384	21/384	0/384
MSE-ID+QP	194/384	185/384	115/384	117/384
BC-ridge	0/384	0/384	0/384	0/384

second, paired two-plant campaign tests directional learning with stronger decay and conditioning requirements. Appendix G contains complete protocols, seed-level results, and verification records.

Independent scenario study. The pendulum of Chang et al. (2019) obeys $\dot{\varphi} = \omega$ and $\dot{\omega} = 19.62 \sin \varphi - \frac{8}{3}\omega + \frac{80}{3}u$, with expert $u_E = \mathrm{clip}(-23.1717\varphi - 6.7996\omega, -50, 50)$. Each episode records one reset drawn uniformly from the radius-2 disk. With $\lambda = 0.1$, $T = 0.5\,\mathrm{s}$, $K = 224$ fixed quadratic certificates and $p = 22$ affine coefficients, eight independent batches at each $N \in \{16, 32, 64, 128, 256\}$ produce 40 fitted pairs. The protocol precedes data generation; development records are excluded. Each output has 2,048 independent deployment tests and 16,384 independent source states. RK4 node tests use a 1 ms step and normalized residual tolerance 10^{-8}.

All 40 selected fits have certified zero optimal slack. Directed-rounding primal–dual checks transfer the exact-optimizer source report to the corrected coefficients actually deployed, with loss corrections below $1.51 \cdot 10^{-8}$. At $N = 128, 256$, every pair has 0/2,048 detected events; the per-pair two-sided 95% binomial upper endpoint is 0.180% for this numerical event. The mean source bounds at the five sample sizes are approximately $1, 0.915, 0.570, 0.313, 0.164$. At $N = 256$, the mean full-vector relative RMSE is 99.8%, while harmful directional RMS error is $8.61 \cdot 10^{-4}$. Most selected certificates have condition number near 1,000; the following campaign explicitly reduces this anisotropy.

Paired two-plant campaign. We use pendulum reset episodes and cart-pole episodes containing 51 records over 2 s. The independent sample count is the number of episodes, $N \in \{128, 1024\}$, with three training seeds and 128 paired test starts per seed. The smaller training sets are prefixes of the larger sets. Pendulum uses $(\lambda, T, K, p) = (3.2, 10, 641, 11)$; cart-pole uses $(0.46, 20, 320, 24)$. Both finite libraries impose condition number at most 100 up to roundoff, and all model-based learners receive the same exact kinematic rows. The scenario method jointly selects a certificate and warning-oriented predictor. MSE-ID+QP fits the same predictor class by squared error and selects its own certificate from the same library using logged warnings. R39-ERM uses a two-sided directional objective. BC-ridge learns the expert action from the same records and selects a diagnostic certificate. Additional neural and margin comparisons appear in Appendix G.9.

Directional accuracy and deployment. At $N = 1024$, the scenario method has zero detected violations on both plants, while MSE-ID+QP has 48.18% on pendulum and 30.47% on cart-pole (Table 1). Its lower vector error does not ensure the signed accuracy required by the certificate: scenario versus MSE relative error on the unknown drift rows is 41.89% versus 3.26% for pendulum and 89.59% versus 0.481% for cart-pole. Harmful directional RMS errors are respectively $9.71 \cdot 10^{-4}$ versus $2.14 \cdot 10^{-2}$ and $6.14 \cdot 10^{-5}$ versus $3.62 \cdot 10^{-3}$. The small-N results show batch sensitivity: scenario cart-pole violations fall from $144/384$ at $N = 128$ to zero at 1024, whereas R39-ERM has fewer events at 128. Behavior cloning also has zero events, consistent with the linear experts being well represented by its class. These comparisons support directional sufficiency, without establishing dominance over imitation or a general ordering of the two directional objectives.

Continuous-state verification of learned outputs. A separate post-fit interval calculation certifies all three $N = 1024$ campaign pendulum pairs. Their unchanged coefficients define a compact invariant sublevel $K = \{x : x^\top P x \leq 4.01\}$ containing the entire reset disk. Directed interval bounds on an enclosing rectangle establish $R_h(x) \leq -\|x\|^2/261063$. Consequently $V(x(t)) \leq e^{-3.2t} V(x(0))$ for every reset state and all $t \geq 0$, giving deterministic zero all-time decrease-violation risk for these specified benchmark controllers (Proposition 10). This model-based verification is performed after fitting and supplies no additional information to the learner.

Scope of the confidence reports. The campaign's floating-point optimizer diagnostics are not interval-certified scenario reports; its computed general source reports at $N = 1024$ are 0.1787 for pendulum and 0.3747 for cart-pole. No physical encounter coefficient is asserted, and the offline deployment report remains $Q = 1$. The preceding deterministic benchmark certificate is a separate guarantee. Independent development checks in Appendix G.6 illustrate why small reset loss alone does not establish encounter coverage; Proposition 1 supplies sufficient invariant-warning-set conditions for that transfer. Numerical event counts throughout concern the specified integrator, except where the separate continuum proof applies.

6 Discussion

The warning loss links offline learning to certificate decrease without full drift identification. The paired campaign exhibits this distinction under bounded certificate conditioning and long deployment horizons. Model-based interval verification additionally establishes all-time decrease for three frozen pendulum controllers.

Strict passive-data trajectory guarantees require justified encounter coverage; invariant warning sublevels provide sufficient transfer conditions. Spatial practical stability instead requires a lower sampling density, residual regularity, and certified sublevel containment. The benchmark interval calculation uses the specified plant after fitting and does not replace these information assumptions. Exact labels and continuous mathematical feedback remain the scope of the guarantees; sampled actuation and model uncertainty need additional analysis.

AI use statement. Generative AI assisted with writing, literature review, mathematical derivations and proof review, experiment design, and code. The authors take responsibility for all claims and results.

REFERENCES

Aaron D. Ames, Kevin Galloway, Jessy W. Grizzle, and Koushil Sreenath. Rapidly exponentially stabilizing control Lyapunov functions and hybrid zero dynamics. *IEEE Transactions on Automatic Control*, 59(4):876–891, 2014. doi: 10.1109/TAC.2014.2299335. URL https://doi.org/10.1109/TAC.2014.2299335.

Nicholas Boffi, Stephen Tu, Nikolai Matni, Jean-Jacques Slotine, and Vikas Sindhwani. Learning stability certificates from data. In *Proceedings of the 2020 Conference on Robot Learning*, volume 155 of *Proceedings of Machine Learning Research*, pp. 1341–1350. PMLR, 2021. URL https://proceedings.mlr.press/v155/boffi21a.html.

Giuseppe C. Calafiore and Marco C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006. doi: 10.1109/TAC.2006.875041. URL https://doi.org/10.1109/TAC.2006.875041.

Marco C. Campi and Simone Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008. doi: 10.1137/07069821X. URL https://epubs.siam.org/doi/10.1137/07069821X.

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

Jinglin Chen and Nan Jiang. Information-theoretic considerations in batch reinforcement learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1042–1051. PMLR, 2019. URL https://proceedings.mlr.press/v97/chen19e.html.

C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934. doi: 10.1093/biomet/26.4.404.

Hongkai Dai, Benoit Landry, Lujie Yang, Marco Pavone, and Russ Tedrake. Lyapunov-stable neural-network control. In *Proceedings of Robotics: Science and Systems*, 2021. doi: 10.15607/RSS.2021.XVII.063. URL https://www.roboticsproceedings.org/rss17/p063.html.

Charles Dawson, Zengyi Qin, Sicun Gao, and Chuchu Fan. Safe nonlinear control using robust neural Lyapunov-barrier functions. In *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, pp. 1724–1735. PMLR, 2022. URL https://proceedings.mlr.press/v164/dawson22a.html.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, 20(4):633–679, 2020. doi: 10.1007/s10208-019-09426-y. URL https://link.springer.com/article/10.1007/s10208-019-09426-y.

Anirudha Majumdar, Alec Farid, and Anoopkumar Sonar. PAC-bayes control: Learning policies that provably generalize to novel environments. *The International Journal of Robotics Research*, 40 (2–3):574–593, 2021. doi: 10.1177/0278364920959444. URL https://arxiv.org/abs/1806.04225.

Andreas Maurer. A vector-contraction inequality for Rademacher complexities. In *Algorithmic Learning Theory*, volume 9925 of *Lecture Notes in Computer Science*, pp. 3–17. Springer, 2016. doi: 10.1007/978-3-319-46379-7_1. URL https://link.springer.com/chapter/10.1007/978-3-319-46379-7_1.

Youngjae Min, Spencer M. Richards, and Navid Azizan. Data-driven control with inherent Lyapunov stability. In *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023. URL https://arxiv.org/abs/2303.03157.

Mayank Mittal, Marco Gallieri, Alessio Quaglino, Seyed Sina Mirrazavi Salehian, and Jan Koutník. Neural Lyapunov model predictive control: Learning safe global controllers from sub-optimal examples. *arXiv preprint arXiv:2002.10451*, 2020. URL https://arxiv.org/abs/2002.10451v2.

Rémi Munos and Csaba Szepesvári. Finite-time bounds for fitted value iteration. *Journal of Machine Learning Research*, 9(27):815–857, 2008. URL https://jmlr.org/papers/v9/munos08a.html.

Daniel Pfrommer, Thomas Zhang, Stephen Tu, and Nikolai Matni. TaSIL: Taylor series imitation learning. In *Advances in Neural Information Processing Systems*, volume 35, pp. 20162–20174. Curran Associates, Inc., 2022. doi: 10.52202/068431-1466. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/7f10c3d66c3b7863a9cda255dcac5bb7-Abstract-Conference.html.

Paria Rashidinejad, Banghua Zhu, Cong Ma, Jiantao Jiao, and Stuart Russell. Bridging offline reinforcement learning and imitation learning: A tale of pessimism. In *Advances in Neural Information Processing Systems*, volume 34, pp. 11702–11716, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/60ce36723c17bbac504f2ef4c8a46995-Abstract.html.

Hadi Ravanbakhsh and Sriram Sankaranarayanan. Learning control Lyapunov functions from counterexamples and demonstrations. *Autonomous Robots*, 43(2):275–307, 2019. doi: 10.1007/s10514-018-9791-9. URL https://doi.org/10.1007/s10514-018-9791-9.

Allen Ren, Sushant Veer, and Anirudha Majumdar. Generalization guarantees for imitation learning. In *Proceedings of the 2020 Conference on Robot Learning*, volume 155 of *Proceedings of Machine Learning Research*, pp. 1426–1442. PMLR, 2021. URL https://proceedings.mlr.press/v155/ren21a.html.

Navid Rezazadeh, Maxwell Kolarich, Solmaz S. Kia, and Negar Mehr. Learning contraction policies from offline data. *IEEE Robotics and Automation Letters*, 7(2):2905–2912, 2022. doi: 10.1109/LRA.2022.3145100. URL https://arxiv.org/abs/2112.05911.

Stéphane Ross, Geoffrey J. Gordon, and J. Andrew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pp. 627–635. PMLR, 2011. URL https://proceedings.mlr.press/v15/ross11a.html.

Sumeet Singh, Spencer M. Richards, Vikas Sindhwani, Jean-Jacques E. Slotine, and Marco Pavone. Learning stabilizable nonlinear dynamics with contraction-based regularization. *The International Journal of Robotics Research*, 40(10–11):1123–1150, 2021. doi: 10.1177/0278364920949931. URL https://journals.sagepub.com/doi/10.1177/0278364920949931.

Nathan Srebro, Karthik Sridharan, and Ambuj Tewari. Smoothness, low noise and fast rates. In *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010. URL https://proceedings.neurips.cc/paper/2010/hash/76cf99d3614e23eabab16fb27e944bf9-Abstract.html.

Andrew J. Taylor, Victor D. Dorobantu, Hoang M. Le, Yisong Yue, and Aaron D. Ames. Episodic learning with control Lyapunov functions for uncertain robotic systems. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6878–6884. IEEE, 2019. doi: 10.1109/IROS40897.2019.8967820. URL https://arxiv.org/abs/1903.01577.

Andrew J. Taylor, Victor D. Dorobantu, Sarah Dean, Benjamin Recht, Yisong Yue, and Aaron D. Ames. Towards robust data-driven control synthesis for nonlinear systems with actuation uncertainty. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 6469–6476. IEEE, 2021. doi: 10.1109/CDC45484.2021.9683511. URL https://doi.org/10.1109/CDC45484.2021.9683511.

Stephen Tu, Alexander Robey, Tingnan Zhang, and Nikolai Matni. On the sample complexity of stability constrained imitation learning. In *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*, volume 168 of *Proceedings of Machine Learning Research*, pp. 180–191. PMLR, 2022. URL https://proceedings.mlr.press/v168/tu22a.html.

Masatoshi Uehara and Wen Sun. Pessimistic model-based offline reinforcement learning under partial coverage. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=tyrJsbKAe6.

Tengyang Xie, Ching-An Cheng, Nan Jiang, Paul Mineiro, and Alekh Agarwal. Bellman-consistent pessimism for offline reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6683–6694, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/34f98c7c5d7063181da890ea8d25265a-Abstract.html.

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

Thomas Zhang, Stephen Tu, Nicholas Boffi, Jean-Jacques Slotine, and Nikolai Matni. Adversarially robust stability certificates can be sample-efficient. In *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*, volume 168 of *Proceedings of Machine Learning Research*, pp. 532–545. PMLR, 2022. URL https://proceedings.mlr.press/v168/zhang22a.html.

Ruikun Zhou, Thanin Quartz, Hans De Sterck, and Jun Liu. Neural Lyapunov control of unknown nonlinear systems with stability guarantees. In *Advances in Neural Information Processing Systems*, volume 35, pp. 29113–29125. Curran Associates, Inc., 2022. doi: 10.52202/068431-2111. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/bba3160c42f4c1d23175b5abf04347f5-Abstract-Conference.html.