

Certificate-Guided Residual Variance Control for Residual Learned Controllers in Nonlinear Systems

Yue Wu

Abstract—This paper studies residual learned controllers for discrete-time nonlinear systems whose local lifted or linearized closed-loop surrogates exhibit nonnormal transient amplification. A residual action source that appears small in Euclidean variance can still produce large state covariance after propagation by such a local amplifier. We formulate this behavior as source-amplifier variance propagation and propose certificate-guided residual variance control (CG-RVC), a residual actor-critic wrapper that constructs a local surrogate, computes a projected Lyapunov certificate, and pulls the certificate back to residual-action space as an anisotropic covariance penalty. The certificate model supplies local weights and diagnostics for residual source shaping. The analysis gives covariance propagation and nonnormal amplification bounds with mismatch, colored-input, and residual-feedback extensions. Controlled interventions separate source and amplifier effects: filtering under a fixed amplifier reduces action variance by 92.1% and state covariance by 21.8%, while equal-trace MIMO source directions produce measured state-energy ratios of $94.2\text{--}141.0\times$ that match the certificate-weighted prediction. In a matched 5-seed high-noise constrained-RL comparison, CG-RVC reduces RMSE by 47.1%, applied action variance by 94.7%, action-difference MSE by 97.7%, and posterior Bellman-Lyapunov maximum by 94.3% relative to Direct TD3, while reducing dense violations from 0.9389 to 0. Planar and Full3D quadrotor simulations provide additional mechanism evidence for nonlinear aerial systems; in Full3D, filtered source injection reduces state covariance by 65.7–83.8% and motor variance by 90.3–94.5% across four scenarios.

Index Terms—Learning-based control, nonlinear systems, nonnormality, transient amplification, residual variance control, local surrogate, Lyapunov certificate, reinforcement learning, quadrotor simulation.

I. Introduction

Residual learned control for discrete-time nonlinear continuous-control systems combines nominal feedback, a neural residual actor, and execution-time safety repair. In this setting, the risk is not only that the learned action is noisy. A residual source that appears small in Euclidean variance can produce large state covariance, Lyapunov-energy excursions, and projection burden after it passes through the local closed-loop dynamics. This paper treats noise, constraints, and aerial simulation as settings where this residual-source propagation problem becomes visible.

Existing learning-based control explanations mostly act on one side of this propagation chain. Action smoothing, control regularization, smooth exploration, and action-alignment methods penalize Euclidean action magnitude, temporal variation, or high-frequency content [1]–[5]. Constrained policy optimization, safe-learning methods, control barrier functions, and shielding repair unsafe actions or enforce feasibility conditions after a candidate

command has been produced [6]–[9]. These source-side and repair-side tools are important, but neither explains why two residual covariance directions with the same trace can create very different downstream state covariance under the same nominally stable local closed loop.

Classical control theory supplies the missing amplifier-side view. Pseudospectral, nonmodal, and stochastic-forcing analyses show that eigenvalue stability does not preclude finite-time amplification or large steady covariance under random forcing [10]–[12]. Stochastic Lyapunov, robust-control, LQG, and covariance-control viewpoints further show that second moments are propagated objects rather than nuisance statistics [13]–[15]. In neural residual control, this classical perspective exposes a closed-loop propagation geometry that is invisible to Euclidean action statistics: risk depends on how the residual source enters through the input channel, how the local surrogate amplifies it, and how downstream certificate energy evaluates it.

We propose certificate-guided residual variance control (CG-RVC) to turn this propagation view into a trainable residual actor-critic method. CG-RVC constructs a lifted or linearized local surrogate, computes a projected Lyapunov certificate, and pulls that certificate back through the residual input channel to obtain a direction-aware residual-action metric. The actor update uses this metric as a certificate-weighted residual covariance penalty, while the execution layer applies risk-scheduled smoothing, rate limiting, and safety projection. Posterior Bellman-Lyapunov residuals and actor-augmented amplifier diagnostics then check whether the learned residual behavior remains consistent with the local certificate model. Thus source shaping, safety repair, and posterior diagnosis are separated rather than conflated.

The experimental design follows the same source-amplifier-risk chain. Controlled interventions separate source and amplifier effects: filtering under a fixed amplifier reduces action variance by 92.1% and state covariance by 21.8%, while equal-trace MIMO source directions produce measured state-energy ratios of $94.2\text{--}141.0\times$ that track the certificate-weighted prediction. In a matched 5-seed high-noise constrained-RL comparison, CG-RVC reduces RMSE, applied action variance, action-difference MSE, posterior Bellman-Lyapunov maximum, and dense violations relative to Direct TD3. Planar and Full3D quadrotor simulations provide additional mechanism evidence for nonlinear aerial systems.

The contributions are as follows.

- We identify a source-amplifier-risk mechanism in

residual learned control: residual-action covariance should not be interpreted only through Euclidean magnitude, temporal smoothness, or exploration noise, because locally nonnormal closed-loop geometry can directionally amplify residual sources into state covariance, Lyapunov-energy excursions, and projection burden.

- We convert classical nonmodal covariance-propagation theory into a trainable residual-action metric. By constructing a local surrogate, computing a projected Lyapunov certificate, and pulling the certificate back through the residual input channel, we obtain an anisotropic action-space covariance metric $M = \Phi_u^\top P \Phi_u$ that weights residual directions by downstream certificate energy.
- We propose CG-RVC, a classical-control-informed residual actor-critic wrapper that separates upstream source shaping, execution-time safety repair, and posterior certificate diagnosis. Controlled interventions, MIMO directional injection, matched constrained-RL comparisons, and quadrotor simulations support the source-amplifier interpretation and the distinct roles of source shaping and projection.

II. Background and Positioning

This section positions CG-RVC along the functional chain used in the paper: residual source statistics, local input channel, nonnormal amplifier, covariance or Lyapunov-energy propagation, downstream repair, and posterior diagnosis. Each subsection summarizes what the existing literature contributes to this chain, what it leaves unresolved for residual actor training, and how CG-RVC uses that component.

A. Nonnormal Amplification and Covariance Propagation

Pseudospectra, transient growth, and stochastic response theory provide the mathematical basis for treating a stable closed loop as a possible amplifier [10]–[12]. Recent statistical transient-growth and Kreiss-constant work extends these diagnostics beyond worst-case disturbance analysis [16], [17]. These tools usually analyze a given operator, controller, or disturbance model. CG-RVC uses them inside residual learned control: the fitted local closed-loop surrogate is the amplifier, and residual-action covariance is the source that it propagates.

Stochastic Lyapunov analysis, H_2 interpretations, LQG, and covariance shaping all treat second moments as control-relevant quantities [13]–[15]. Their usual role is analysis or controller synthesis for a specified stochastic system. Here their role is different: classical covariance propagation is turned into a residual actor-training signal that identifies which residual covariance directions increase downstream Lyapunov energy after passing through the local amplifier.

B. Action Regularization and Residual RL as Source-Side Control

Residual policies and source-side regularizers modify what the learned actor sends to the plant. Control regularization, CAPS-style action consistency, smooth exploration, and action-alignment methods show that smoother or lower-variance commands can improve continuous-control deployment [1]–[5]. These methods reduce action magnitude, temporal variation, or high-frequency content before the command reaches the closed loop.

Source-side regularization is necessary but incomplete for the risk mechanism studied here. A scalar variance penalty or temporal action-difference penalty does not use the residual input channel Φ_u , the local amplifier K_{cl} , or the certificate P . CG-RVC keeps the source-side goal of shaping residual variation, but replaces isotropic or temporal-only penalties with a certificate-weighted covariance penalty that depends on downstream closed-loop propagation.

C. Safety Filters as Downstream Repair

Constrained policy optimization, safe learning surveys, control barrier functions, and Koopman-CBF synthesis provide foundations for safe learning and execution-time filtering [6]–[9]. Projection and shielding can repair infeasible actions under appropriate feasibility assumptions. They explain how a candidate command is corrected, but they do not explain why a particular residual covariance direction repeatedly creates larger correction burden.

CG-RVC therefore treats projection as downstream repair rather than as the source of the residual metric. Reducing certificate-weighted residual covariance can lower the probability or magnitude of large corrections, while projection supplies the feasibility repair. This separation distinguishes upstream residual source shaping from downstream constraint handling and makes projection burden an outcome to be explained, not merely a post-processing step.

D. Local Surrogates and Certificate Pullback Metrics

Local linearization, EDMDC/Koopman dictionaries, data-driven modeling, and system identification provide routes to the local surrogates used in this paper [18]–[21]. These tools provide the objects needed by the source-amplifier chain: a closed-loop matrix, an input channel, and a coordinate system in which covariance propagation is analyzable. By themselves, however, they do not specify which residual-action covariance directions should be penalized during actor training.

Certificate-based local models add the final ingredient. Koopman-CBF and related safe-learning constructions connect lifted models to stability or safety certificates [7], [9]. CG-RVC uses certificate information as a trainable action-space metric: it pulls the state-side certificate back through Φ_u to obtain $M = \Phi_u^\top P \Phi_u$, a residual-action covariance metric that identifies amplified and safety-relevant residual directions.

E. Positioning Summary

The literature covers the individual pieces of the chain: source-side action smoothing, nonnormal amplification, covariance propagation, data-driven local surrogates, certificates, and safety filters. The gap addressed here is the integrated path

$$\Sigma_r \rightarrow \Phi_u \rightarrow K_{cl} \rightarrow P \rightarrow M,$$

where residual covariance is treated as the source, the local nonnormal surrogate is treated as the amplifier, and the certificate is pulled back to action space as a direction-aware training penalty. CG-RVC closes this path at the algorithmic level by combining the pullback metric with execution repair and posterior diagnostics.

Several implemented comparators are named by their experimental role rather than by the cited paper family. Table I states this mapping explicitly. The point is not that each baseline is an exact reimplement of a specific prior paper; rather, each row isolates the same mechanism class under the same task, budget, and logging protocol.

III. Modeling Chain and Certificate Model

CG-RVC uses one modeling chain throughout the paper: nonlinear plant, local surrogate, raw nominal amplifier, projected certificate model, residual source metric, and execution layer. The purpose of this section is to make clear which object is physical, which object is local, and which object is used only for certificate construction.

The ingredients in this chain have different provenance. Residual-policy notation and action regularization follow common learning-based control practice [1]–[3]; execution-time repair follows constrained and safety-filtered learning viewpoints [6]–[9]; and the local surrogate may be obtained from local linearization, EDMDC/Koopman modeling, DMD with control, or system identification [18]–[24]. The paper-specific modeling step is to compose these standard objects into a residual source, local amplifier, certificate, and execution-layer chain, and then to use the certificate pullback as an action-space covariance metric.

A. Nonlinear Residual Closed Loop

The physical closed loop is the nonlinear system

$$x_{t+1} = f(x_t, u_t, d_t), \quad (1)$$

where $x_t \in \mathbb{R}^n$ is the physical state, $u_t \in \mathbb{R}^m$ is the applied action, and d_t represents external disturbance or model error. Given a nominal controller $u_{\text{nom}}(x)$ and a learned residual policy $r_\theta(x)$, the pre-projection action is

$$u_t = u_{\text{nom}}(x_t) + r_t, \quad r_t = r_\theta(x_t). \quad (2)$$

Rollout metrics, safety violations, and posterior residuals are evaluated on this nonlinear closed loop after the execution layer is applied.

B. Local Lifted or Linearized Surrogate

The certificate analysis is built in local coordinates

$$z_t = \psi(x_t), \quad (3)$$

where ψ may be the identity, a local linearizing coordinate map, or an EDMDC/Koopman dictionary. EDMD, DMDc, and Koopman predictor constructions are standard ways to obtain finite-dimensional controlled predictors for nonlinear systems [18], [22]–[24]. Thus “Koopman” is one possible construction of the local surrogate, not the main claim of the paper. In these coordinates, data fitting or local linearization gives the input-affine surrogate

$$z_{t+1} = K_0 z_t + \Phi_u u_t + w_t. \quad (4)$$

Here K_0 describes the local autonomous evolution, Φ_u is the residual-action input channel, and w_t collects local disturbance and unresolved surrogate error. The input-channel notation is inherited from controlled surrogate modeling; the paper-specific use of Φ_u is to propagate residual-action covariance into the Lyapunov certificate metric. Every covariance statement below is a statement about this local z -system unless explicitly stated otherwise.

C. Raw Nominal Amplifier

The nominal controller induces the local closed-loop amplifier before any certificate projection is applied. When the nominal feedback is represented locally as $u_{\text{nom}} = -Lz$, substituting it into (4) gives

$$K_{\text{raw}} = K_0 - \Phi_u L. \quad (5)$$

The matrix K_{raw} is the local nominal amplifier associated with the fitted or linearized model. It is the natural local model of the nominal closed-loop surrogate, but it may be unstable or too close to the unit circle to support the certificate used by CG-RVC. The term “amplifier” follows the nonmodal and pseudospectral interpretation of stable but nonnormal closed-loop dynamics [10]–[12], [25].

D. Projected Certificate Model

CG-RVC constructs the Lyapunov certificate on a projected matrix

$$K_{\text{proj}} = \Pi_\beta(K_{\text{raw}}), \quad \rho(K_{\text{proj}}) \leq \beta < 1. \quad (6)$$

The projection is a certificate-construction step. It creates the matrix used in the local Lyapunov and covariance analysis; rollout evaluation remains on the nonlinear closed-loop simulator. The scalar β is a target spectral-radius upper bound, distinct from a norm contraction certificate. Spectral-radius contraction, Euclidean contraction, and P -weighted Lyapunov contraction are different standard notions in robust and optimal control [13], [14]:

$$\rho(K) \leq \beta, \quad \|K\|_2 \leq \beta, \quad K^\top P K \preceq \beta^2 P. \quad (7)$$

This distinction is essential for nonnormal systems: $\rho(K) < 1$ does not preclude transient amplification [10], [11].

TABLE I
Mapping from the source-amplifier-risk chain to implemented comparators.

Chain component	Representative prior work	Implemented comparator(s)	Question tested
Penalty-based constrained RL	CPO and penalty-style safe RL [6], [7]	PPO-Penalty	Whether a fixed safety penalty is enough under high noise
Lagrangian constrained RL	Lagrangian safe policy optimization [6], [7]	PPO-Lagrangian; SAC-Lagrangian	Whether learned dual penalties control violations and residual energy
Source-side action regularization	Control regularization, CAPS, smooth exploration, and action alignment [1]–[5]	TD3 + action-diff penalty; TD3 + low-pass only; TD3 + plain Var(u)	Whether Euclidean or temporal smoothing explains the gains
Downstream safety repair	CBF and Koopman-CBF safety filters [8], [9]	TD3 + projection only; Shielded SAC	Whether execution-time repair alone is sufficient
Local surrogate and certificate pullback	Local surrogate modeling and Lyapunov safe learning [18]–[20]	P-ALE only; P-ALE + smoothing; P-ALE + projection	Whether the certificate-guided source metric works without the full execution chain
Source-amplifier covariance geometry	Stochastic covariance control and source-amplifier analysis [13]–[15]	Full-pipeline metric replacement; MIMO directional source injection	Whether source direction, not only source magnitude, matters

Replacing K_{raw} by K_{proj} creates local mismatch. A convenient bridge from the raw local surrogate to the certificate model is

$$z_{t+1} = K_{\text{proj}}z_t + \Phi_u\xi_t + w_t + e_t, \quad (8)$$

$$e_t = \bar{e}(z_t) + \tilde{e}_t, \quad (9)$$

$$\bar{e}(z_t) = (K_{\text{raw}} - K_{\text{proj}})z_t + \bar{e}_{\text{id}}(z_t) + \bar{e}_{\text{nl}}(z_t). \quad (10)$$

The bias $\bar{e}(z_t)$ contains the deterministic projection mismatch and the mean components of the local identification and nonlinear residual errors. The centered fluctuation $\tilde{e}_t$ contains the stochastic part of the local mismatch. Thus the projection error is not assumed to be zero-mean noise; the centered mismatch theorem in Section IV bounds $\tilde{e}_t$, and the biased component is handled by an explicit Young-inequality bound.

E. Residual Source Versus Residual Feedback

The residual action can play two different local roles. Its centered fluctuation is the source whose covariance is regularized:

$$r_\theta(z_t) = \bar{r}_\theta(z_t) + \xi_t, \quad \mathbb{E}[\xi_t | z_t] = 0. \quad (11)$$

For deterministic actors, ξ_t is interpreted as centered residual variation over a local rollout ensemble or minibatch, rather than as exogenous white noise at a fixed state. If the deterministic residual feedback is locally approximated by $\bar{r}_\theta(z) \approx J_r z$, then the actor-augmented local amplifier is

$$K_\theta = K_{\text{proj}} + \Phi_u J_r. \quad (12)$$

Only ξ_t is treated as the source term in the covariance decomposition. The feedback component belongs in the amplifier if its Jacobian is estimated; otherwise the source-only interpretation is diagnostic and is checked by posterior Bellman-Lyapunov residuals.

TABLE II
Object map from the nonlinear plant to the local source-amplifier analysis.

Object	Role in the paper
Physical plant (x, u)	Nonlinear system used for rollout evaluation and safety metrics.
Local coordinate $z = \psi(x)$	Lifted, identity, or linearized coordinate used only for local surrogate analysis.
Surrogate (K_0, Φ_u)	Local input-affine model; Φ_u is the residual-action input channel.
Raw amplifier K_{raw}	Local nominal closed-loop surrogate before certificate projection.
Certificate amplifier K_{proj}	Projected stable matrix used in Lyapunov and covariance bounds.
Residual source ξ_t	Centered residual fluctuation whose covariance is regularized.
Actor-augmented amplifier K_θ	Diagnostic local amplifier when residual feedback is folded into the certificate model.
Execution layer	Smoothing, rate limiting, and safety projection; nonlinear unless explicitly augmented.

IV. Nonnormal Source-Amplifier Theory

This section states the source-amplifier results used by CG-RVC. The generic source input matrix is denoted by B_s . CG-RVC is obtained by the specialization $K = K_{\text{proj}}$, $B_s = \Phi_u$, and ξ_t equal to the centered residual fluctuation defined in Section III. For readability, the main text keeps the assumptions, bounds, and their role in the method; algebraic proofs and standard Lyapunov-drift derivations are collected in Appendix B.

The mathematical ingredients are not all new. Discrete Lyapunov equations, stationary covariance formulas, and stochastic Lyapunov drift bounds are standard tools for linear stochastic systems and robust or optimal control [13], [14]. Covariance control and covariance steering treat covariance as a control object rather than merely a diagnostic [15], [26]. Nonnormal transient amplification is classical in pseudospectral and nonmodal stability theory [10]–[12], [25], and stochastic-forcing studies show that stable nonnormal dynamics can maintain large variance

under small random inputs [27]–[29]. The paper-specific contribution is the specialization and combination: treating residual-action covariance as the source, using the local learned closed-loop surrogate as the amplifier, separating residual feedback from residual fluctuation, pulling the certificate back to action space, and connecting the resulting metric to projection burden in residual learned controllers.

A. Source-Amplifier Covariance Decomposition

Consider the generic local stochastic model

$$z_{t+1} = Kz_t + B_s \xi_t + w_t, \quad (13)$$

where ξ_t is a centered source fluctuation and w_t is external disturbance or model error. Assume

$$\mathbb{E}[\xi_t] = 0, \quad \mathbb{E}[w_t] = 0, \quad \mathbb{E}[\xi_t \xi_t^\top] = \Sigma_a, \quad \mathbb{E}[w_t w_t^\top] = W, \quad (14)$$

and that ξ_t and w_t are conditionally uncorrelated with the past and have no cross covariance. Define

$$Q_{\text{src}} = B_s \Sigma_a B_s^\top + W. \quad (15)$$

The next result is the standard stationary covariance solution of a Schur-stable linear stochastic system, written here in source-amplifier notation [13]–[15], [26].

Theorem 1 (Discrete Source-Amplifier Decomposition). If K is Schur stable, then the stationary covariance exists, is unique, and satisfies

$$\begin{aligned} \Sigma_z &= \sum_{i=0}^{\infty} K^i Q_{\text{src}} (K^i)^\top \\ &= \mathcal{L}_K(W) + \mathcal{L}_K(B_s \Sigma_a B_s^\top). \end{aligned} \quad (16)$$

Thus source variance is not state variance itself; it is propagated through the closed-loop amplifier $\mathcal{L}_K$.

Proposition 1 (Action-Covariance Monotonicity). Fix K , B_s , and W . If $\Sigma_a^{(1)} \preceq \Sigma_a^{(2)}$, then $\Sigma_z^{(1)} \preceq \Sigma_z^{(2)}$.

This monotonicity statement follows from positivity of the discrete Lyapunov operator and is valid only when the amplifier is fixed.

B. Nonnormal Amplification Bound

Proposition 2 (Nonnormal Covariance Amplification). If $K = V\Lambda V^{-1}$ is diagonalizable, then

$$\|\Sigma_z\|_2 \leq \frac{\kappa_2(V)^2}{1 - \rho(K)^2} \|Q_{\text{src}}\|_2. \quad (17)$$

This elementary bound combines the discrete covariance expansion with diagonalization. Its role is interpretive: nonmodal and pseudospectral theory show why a stable nonnormal matrix can have large transient response, and stochastic-forcing analyses show the corresponding variance-amplification effect under random inputs [10], [11], [25], [27]–[29]. For a normal matrix $\kappa_2(V) = 1$; for a nonnormal matrix, the same source may be amplified substantially. This is a worst-case bound and can become very loose when $\kappa(V)$ is large, so it should be read as a mechanism bound rather than a precise covariance predictor.

C. P -Weighted Residual Variance Propagation

Consider

$$z_{t+1} = Kz_t + B_s \xi_t + w_t. \quad (18)$$

The following recursion is a standard stochastic Lyapunov drift argument [13], [14]. The paper-specific step is the definition of the source metric M_s and its specialization to residual action covariance. Let $P \succ 0$ be a Lyapunov certificate for the local amplifier K and define the measured P -metric contraction

$$\gamma_P^2 = \lambda_{\max}(P^{-1/2} K^\top P K P^{-1/2}). \quad (19)$$

If $\gamma_P < 1$, then $K^\top P K \preceq \gamma_P^2 P$. In CG-RVC, the design parameter β is only a spectral-radius projection target; all strict Lyapunov-energy bounds use the measured contraction γ_P . If one explicitly solves $K^\top P K - \beta^2 P = -Q$ with $Q \succ 0$, then β may replace γ_P in the same formulas.

Define

$$M_s := B_s^\top P B_s, \quad (20)$$

and $V(z) = z^\top P z$. Under the source interpretation with $\mathbb{E}[\xi_t] = 0$, $\mathbb{E}[\xi_t \xi_t^\top] = \Sigma_r$, and conditional orthogonality,

$$\mathbb{E}[V(z_{t+1})] \leq \gamma_P^2 \mathbb{E}[V(z_t)] + \text{tr}(M_s \Sigma_r) + \text{tr}(P W). \quad (21)$$

The stationary limit obeys

$$\mathbb{E}[V(z)] \leq \frac{\text{tr}(M_s \Sigma_r) + \text{tr}(P W)}{1 - \gamma_P^2}. \quad (22)$$

For CG-RVC, $K = K_{\text{proj}}$, $B_s = \Phi_u$, and the generic metric pulls back to the residual-action space as

$$M := \Phi_u^\top P \Phi_u. \quad (23)$$

The actor-dependent part is the P -weighted residual source term. We define the empirical action Lyapunov energy (ALE) penalty as

$$\hat{\mathcal{R}}_{\text{ALE-}P}(\theta) = \frac{\text{tr}(M \hat{\Sigma}_r(\theta))}{1 - \gamma_P^2}. \quad (24)$$

The Euclidean degeneration is

$$\hat{\mathcal{R}}_{\text{ALE-}I}(\theta) = \frac{\text{tr}(\Phi_u \hat{\Sigma}_r(\theta) \Phi_u^\top)}{1 - \gamma_P^2}. \quad (25)$$

The denominator can be absorbed into the regularization weight when P and γ_P are fixed during an actor update. We keep it explicit because it exposes the robustness-performance scaling. The P -weighted ALE term suppresses a certificate-weighted residual source proxy. It does not directly reduce nonnormality.

Thus (23) should be read as the paper’s synthesis step: standard covariance propagation and Lyapunov certificates are applied to a controlled local surrogate, and the certificate is pulled back through the residual input channel to obtain an anisotropic residual-action covariance metric.

D. Residual Feedback Versus Source Fluctuation

The actor residual is generally state-dependent rather than a state-independent white disturbance. Locally decompose it as

$$r_\theta(z_t) = \bar{r}_\theta(z_t) + \xi_t, \quad (26)$$

where $\bar{r}_\theta$ is the deterministic feedback component and ξ_t is a centered fluctuation around that component. If $\bar{r}_\theta(z) \approx J_r z$ locally, the amplifier becomes

$$K_\theta = K + B_s J_r, \quad (27)$$

which specializes to $K_\theta = K_{\text{proj}} + \Phi_u J_r$ for CG-RVC. Only ξ_t should then be interpreted as the source covariance term. If the residual feedback Jacobian is not estimated, the one-step Lyapunov expansion contains the cross term

$$2\mathbb{E}[z_t^\top K^\top P B_s r_t], \quad (28)$$

which cannot be discarded by the source-noise argument. A conservative alternative is Young's inequality: for any $\eta > 0$,

$$2a^\top b \leq \eta a^\top a + \eta^{-1} b^\top b. \quad (29)$$

Thus the source-amplifier interpretation is exact for exogenous residual fluctuations and diagnostic for state-dependent residual policies unless the feedback component is explicitly folded into the amplifier. Posterior Bellman-Lyapunov residuals report when the local source-only interpretation is violated in rollouts.

E. Colored Input and First-Order Filter Augmentation

A first-order smoothed residual source is colored rather than white. Consider a source generated by a stable causal filter,

$$y_{t+1} = \alpha y_t + (1-\alpha)\epsilon_{t+1}, \quad z_{t+1} = K z_t + B_s y_t + w_t. \quad (30)$$

With augmented state $\bar{z}_t = [z_t^\top, y_t^\top]^\top$,

$$\bar{z}_{t+1} = \begin{bmatrix} K & B_s \\ 0 & \alpha I \end{bmatrix} \bar{z}_t + \begin{bmatrix} 0 \\ (1-\alpha)I \end{bmatrix} \epsilon_{t+1} + \begin{bmatrix} I \\ 0 \end{bmatrix} w_t. \quad (31)$$

This is the standard state-augmentation way to analyze filtered inputs in linear systems [13], [14]; it is included because action smoothing and low-pass execution are common source-side interventions in continuous-control RL [2]–[4]. Thus a filtered residual should not be analyzed by the white-source monotonicity result for (K, B_s) alone. The physical amplifier K may be unchanged, but the effective source statistics and temporal spectrum are changed by the filter. If ϵ_t is white,

$$\Sigma_y = \frac{1-\alpha}{1+\alpha} \Sigma_\epsilon, \quad (32)$$

showing source-side variance reduction. The resulting state covariance still depends on the augmented system, input directions, and model mismatch.

F. Robust Propagation under Lifted Nonlinear Mismatch

The preceding equations are exact for the chosen local stochastic model. The following mismatch bounds are the paper's local robustness extension of the stochastic Lyapunov drift calculation, motivated by the fact that Koopman/EDMDc, DMDc, or identified local models are approximate surrogates [18]–[24]. To connect them to the lifted nonlinear rollout dynamics, include the mismatch split introduced in (9):

$$z_{t+1} = K z_t + B_s \xi_t + w_t + \bar{e}(z_t) + \tilde{e}_t. \quad (33)$$

For CG-RVC, $K = K_{\text{proj}}$ and $B_s = \Phi_u$; the bias $\bar{e}(z_t)$ includes projection mismatch and mean lifted-model error, while $\tilde{e}_t$ is the centered stochastic mismatch.

Assumption 1 (Bounded Conditional Centered Lifted Mismatch). Let $\mathcal{F}_t$ be the history up to time t . For the centered dynamics obtained from (33) by setting $\bar{e}(z_t) \equiv 0$, assume

$$\mathbb{E}[\xi_t | \mathcal{F}_t] = 0, \quad \mathbb{E}[w_t | \mathcal{F}_t] = 0, \quad \mathbb{E}[\tilde{e}_t | \mathcal{F}_t] = 0, \quad (34)$$

with conditional second-moment bounds

$$\mathbb{E}[\xi_t \xi_t^\top | \mathcal{F}_t] \preceq \Sigma_r, \quad (35)$$

$$\mathbb{E}[w_t w_t^\top | \mathcal{F}_t] \preceq W, \quad (36)$$

$$\mathbb{E}[\tilde{e}_t \tilde{e}_t^\top | \mathcal{F}_t] \preceq \Sigma_e. \quad (37)$$

The three centered sources are conditionally second-order uncorrelated, and $K^\top P K \preceq \gamma_P^2 P$ with $\gamma_P < 1$.

Theorem 2 (Robust Variance Propagation under Lifted Nonlinear Mismatch). Under Assumption 1, define

$$q_r = \text{tr}(M_s \Sigma_r), \quad q_w = \text{tr}(P W), \quad q_e = \text{tr}(P \Sigma_e). \quad (38)$$

Then

$$\mathbb{E}[V(z_t)] \leq \gamma_P^{2t} \mathbb{E}[V(z_0)] + \frac{1 - \gamma_P^{2t}}{1 - \gamma_P^2} (q_r + q_w + q_e). \quad (39)$$

Consequently,

$$\limsup_{t \rightarrow \infty} \mathbb{E}[V(z_t)] \leq \frac{\text{tr}(M_s \Sigma_r) + \text{tr}(P W) + \text{tr}(P \Sigma_e)}{1 - \gamma_P^2}. \quad (40)$$

For CG-RVC, $M_s = M = \Phi_u^\top P \Phi_u$, so the residual-source term in Theorem 2 is the same certificate-weighted metric used in the actor objective.

Remark 1 (Biased Projection and Model Mismatch). The projection term $(K_{\text{raw}} - K_{\text{proj}})z_t$ can be a deterministic state-dependent bias, so it is not covered by the zero-mean condition on $\tilde{e}_t$. If the full dynamics (33) retain $\bar{e}(z_t)$, then for any $\eta > 0$,

$$\|K z_t + \bar{e}(z_t)\|_P^2 \leq (1+\eta) \|K z_t\|_P^2 + (1+\eta^{-1}) \|\bar{e}(z_t)\|_P^2. \quad (41)$$

Hence the one-step recursion becomes

$$\begin{aligned} \mathbb{E}[V(z_{t+1})] &\leq (1+\eta) \gamma_P^2 \mathbb{E}[V(z_t)] + q_r + q_w + q_e \\ &\quad + (1+\eta^{-1}) \mathbb{E}[\|\bar{e}(z_t)\|_P^2]. \end{aligned} \quad (42)$$

If $(1+\eta) \gamma_P^2 < 1$ and $\mathbb{E}[\|\bar{e}(z_t)\|_P^2] \leq b_e$ in the local operating domain, the stationary bound is obtained by replacing the

denominator with $1 - (1 + \eta)\gamma_P^2$ and adding $(1 + \eta^{-1})b_e$ to the numerator. This is the biased counterpart of Theorem 2.

Corollary 1 (Scalar Mismatch and Physical-Space Dispersion). If $\Sigma_e \preceq \epsilon I$, then $\text{tr}(P\Sigma_e) \leq \epsilon \text{tr}(P)$. If additionally ψ is locally inverse-Lipschitz around x^* , i.e., $\|\psi(x) - \psi(x^*)\|_2 \geq \ell_\psi \|x - x^*\|_2$, then

$$\limsup_{t \rightarrow \infty} \mathbb{E}[\|x_t - x^*\|_2^2] \leq \frac{\text{tr}(M_s \Sigma_r) + \text{tr}(PW) + \epsilon \text{tr}(P)}{\ell_\psi^2 (1 - \gamma_P^2) \lambda_{\min}(P)}. \quad (43)$$

Remark 2 (Correlated Source and Mismatch). The conditional uncorrelatedness assumption can be relaxed. Let $s_t = [\xi_t^\top, w_t^\top, \tilde{e}_t^\top]^\top$ and $B = [B_s \ I \ I]$. If $\mathbb{E}[s_t | \mathcal{F}_t] = 0$ and $\mathbb{E}[s_t s_t^\top | \mathcal{F}_t] \preceq \Sigma_s$, then

$$\limsup_{t \rightarrow \infty} \mathbb{E}[V(z_t)] \leq \frac{\text{tr}(B^\top P B \Sigma_s)}{1 - \gamma_P^2}. \quad (44)$$

This block-covariance form covers cross-covariance between residual fluctuation, disturbance, and centered lifted mismatch.

Remark 3 (Nonnormal Amplification of Mismatch). If P solves $K^\top P K - P = -Q$ with $Q \succ 0$, then

$$P = \sum_{k=0}^{\infty} (K^k)^\top Q K^k. \quad (45)$$

Therefore

$$\text{tr}(P\Sigma_e) = \sum_{k=0}^{\infty} \text{tr}(Q K^k \Sigma_e (K^k)^\top). \quad (46)$$

Even small Euclidean model error can produce large Lyapunov-energy error if it aligns with transient high-gain directions of a nonnormal K .

G. Projection and Feasible-Tube Consequences

The safety projection layer and the P-ALE penalty play different roles. Projection and shielding are standard downstream repair mechanisms in safe learning and CBF-based control [6]–[9]; Euclidean projection and error-bound tools are classical convex-analysis ingredients [30], [31]. Projection can enforce one-step feasibility only when the local safe-action set is nonempty and recursively feasible. P-ALE does not create such feasibility. Instead, by reducing certificate-weighted residual covariance, it reduces the probability and expected size of large projection corrections. This separation is paper-specific: projection is the feasibility repair layer, while P-ALE is the upstream source-shaping mechanism that makes that repair layer work less aggressively.

The execution-time action used by CG-RVC is written as

$$u_t^{\text{pre}} = u_{\text{nom}}(x_t) + r_t, \quad u_t^{\text{safe}} = \arg \min_{u \in \tilde{\mathcal{U}}_{\text{safe}}(x_t)} \|u - u_t^{\text{pre}}\|_2^2. \quad (47)$$

Appendix B-L states the recursive-feasibility condition, barrier-edge hitting bound, Hoffman projection-burden

bound, and feasible-tube probability calculation. These results quantify how P-ALE reduces projection burden under the stated feasibility assumptions.

H. Sampling Time and the Contraction Factor

Given a continuous-time decay rate $\alpha_c > 0$ and sample time T_s , one may set $\beta = e^{-\alpha_c T_s}$ as a spectral-radius projection target. If the target is to reduce error to a fraction ε within T_{settle} , then

$$\beta = \varepsilon^{T_s/T_{\text{settle}}}. \quad (48)$$

For example, $T_s = 0.5$ s and a 10-second decay to 5% give $\beta \approx 0.861$. This target does not by itself imply P -metric contraction. After projection, the actual diagnostic for Lyapunov-energy bounds is γ_P in (19), together with projection cost, $\kappa(V)$, peak transient gain, and posterior residuals.

V. Certificate-Guided Residual Variance Control

CG-RVC turns the modeling chain in Section III into two loops. A slow certificate-refresh loop fits the local surrogate, builds the projected certificate model, solves for P , and constructs the pullback metric. A high-rate execution loop evaluates the residual actor, applies residual shaping and rate limits, projects unsafe actions when needed, and logs posterior diagnostics.

A. Reinforcement-Learning Backbone and Residual Interface

CG-RVC is an actor-critic residual-learning wrapper around reinforcement learning. The reported implementation uses TD3 as the off-policy backbone [32]: the replay buffer stores transitions generated by the applied safe action, the twin critics use the ordinary Bellman critic loss, and the delayed actor update keeps the TD3 return term while adding the certificate-weighted residual-covariance and temporal-smoothness terms in (52).

The learned network parameterizes a residual target $r_\theta(z)$ around the nominal controller. The environment receives the nominal action plus this residual after smoothing, rate limiting, and safety projection. Thus temporal credit assignment and policy improvement remain standard actor-critic operations, while CG-RVC adds the source-side metric, certificate refresh, execution layer, and posterior diagnostics.

This separation is important for interpreting the algorithm. During each actor update, the current critic, local surrogate, certificate, and sampled replay distribution are treated as fixed. At a slower refresh cadence, rollout data refit or update (K_0, Φ_u) , recompute K_{raw} , K_{proj} , P , γ_P , and M , and then change the regularizer used by later actor updates. Other deterministic or stochastic actor-critic backbones could use the same interface if their actor update can accept a residual-covariance penalty and their rollout loop can apply the execution layer.

B. Implementation of the Local Surrogate

Following the modeling chain in Section III, CG-RVC only requires a local input-affine surrogate of the form (4). The implementation specifies the observable map ψ , the fitted state matrix K_0 , and the residual input channel Φ_u . The observable map ψ may include raw states, local Jacobian features around a working point or nominal trajectory, and a small set of low-order nonlinear dictionary terms. In an EDMDC implementation, one solves

$$\min_{K_0, \Phi_u} \sum_t \|z_{t+1} - K_0 z_t - \Phi_u u_t\|_2^2 + \lambda_{\text{id}} (\|K_0\|_F^2 + \|\Phi_u\|_F^2). \quad (49)$$

Observable selection affects certificate reliability. The reported studies use task-specific local constructions and posterior residual diagnostics rather than assuming a globally complete observable dictionary.

C. Stable Projection and Lyapunov Certificate

Given K_{raw} and target β , the projection

$$K_{\text{proj}} = \Pi_\beta(K_{\text{raw}}) \quad (50)$$

maps the spectral radius below β . The projection Π_β may be implemented by eigenvalue clipping, global scaling, or a nearest-stable approximation. The reported method-component comparison uses eigenvalue clipping: eigenvalues with modulus larger than β are radially clipped while preserving their directions, and the projection cost plus post-projection spectral radius are logged. This practical stabilization heuristic controls the spectral-radius target; separate diagnostics track Euclidean contraction and transient-gain reduction in nonnormal systems. If K_{raw} is already inside the spectral-radius target, the projection is the identity. The discrete Lyapunov equation

$$K_{\text{proj}}^\top P K_{\text{proj}} - P = -Q, \quad Q \succ 0, \quad (51)$$

defines the certificate. The actual contraction used in the bound is the measured γ_P in (19), not the design target β . The diagnostics include P , $\rho(K_{\text{raw}})$, $\rho(K_{\text{proj}})$, projection cost, $\kappa(V)$, peak transient gain, γ_P , and posterior residuals.

D. Actor Objective

The critics are trained by the TD3 Bellman objective; CG-RVC modifies the actor side and the execution layer. For deterministic policy improvement, the actor still ascends the critic-predicted return, but the residual component is also penalized in the certificate-induced metric that Section IV connects to source-amplifier covariance propagation.

The actor objective is

$$\begin{aligned} \min_{\theta} & -\mathbb{E} Q_{\varphi,1}(s, \pi_\theta(s)) + \lambda_{\text{ALE}} \widehat{\mathcal{R}}_{\text{ALE-P}}(\theta) \\ & + \lambda_{\Delta u} \mathbb{E} \|\pi_\theta(s') - \pi_\theta(s)\|^2. \end{aligned} \quad (52)$$

Here

$$\widehat{\Sigma}_r(\theta) = \frac{1}{N} \sum_{i=1}^N (r_\theta(z_i) - \bar{r}_\theta)(r_\theta(z_i) - \bar{r}_\theta)^\top, \quad (53)$$

where the covariance is computed over a local minibatch or rollout ensemble. For a stochastic residual policy, this estimate can approximate the conditional source covariance in the local stochastic model. For the deterministic TD3 actors used in the experiments, it is a local ensemble covariance proxy: it penalizes variation of residual outputs across nearby sampled states in the operating distribution, not exogenous white noise at one fixed state.

The source-amplifier theory supplies the certificate-induced metric and scaling, while posterior Bellman–Lyapunov residuals diagnose when this empirical proxy fails to describe rollout behavior. The term $\widehat{\mathcal{R}}_{\text{ALE-P}}$ is defined in (24). The disturbance term $\text{tr}(PW)/(1 - \gamma_P^2)$ is omitted from the actor loss because it is constant with respect to θ during an update. The ALE term penalizes residual covariance through $M = \Phi_u^\top P \Phi_u$, so anisotropic certificate directions are weighted differently from ordinary Euclidean action smoothing.

For the stationarity analysis, write the complete smooth actor-ascent objective as

$$F_{\text{CG}}(\theta) = J(\theta) - \lambda_{\text{ALE}} \mathcal{R}_P(\theta) - \lambda_{\Delta u} \mathcal{R}_{\Delta u}(\theta), \quad (54)$$

where $J(\theta) = \mathbb{E}_s[Q_{\varphi,1}(s, \pi_\theta(s))]$, $\mathcal{R}_P(\theta) = \text{tr}(M \widehat{\Sigma}_r(\theta))/(1 - \gamma_P^2)$, and $\mathcal{R}_{\Delta u}(\theta) = \mathbb{E} \|\pi_\theta(s') - \pi_\theta(s)\|^2$. The temporal action-difference term is retained as an auxiliary smoothness term. For a minibatch $\{z_i\}_{i=1}^N$, write $r_i = r_\theta(z_i)$, $\bar{r} = N^{-1} \sum_i r_i$, and $J_i(\theta) = \partial r_\theta(z_i)/\partial \theta$. Then the P-ALE gradient component is

$$\nabla_\theta \widehat{\mathcal{R}}_P(\theta) = \frac{2}{N(1 - \gamma_P^2)} \sum_{i=1}^N J_i(\theta)^\top M(r_i - \bar{r}). \quad (55)$$

The following analysis makes explicit what the actor update optimizes: it converges to a stationary point of the regularized surrogate actor objective, rather than to the optimizer of the unregularized return. Thus CG-RVC should be interpreted as solving a robustness-weighted residual-control problem.

Assumption 2 (Frozen-Step Smooth Actor Update). During the actor-update analysis, the critic, local surrogate, certificate, and minibatch distribution are fixed. The parameter set is compact. F_{CG} is differentiable with Lipschitz gradient. Stochastic actor-gradient estimates g_t satisfy $\mathbb{E}[g_t | \theta_t] = \nabla F_{\text{CG}}(\theta_t)$ and $\mathbb{E}[\|g_t - \nabla F_{\text{CG}}(\theta_t)\|_2^2 | \theta_t] \leq \sigma_g^2$.

Proposition 3 (Stationarity of the Regularized Actor Step). Under Assumption 2, let F_{CG} be L_{CG} -smooth and update $\theta_{t+1} = \theta_t + \eta g_t$ with $0 < \eta \leq 1/L_{\text{CG}}$. Then

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\|\nabla F_{\text{CG}}(\theta_t)\|_2^2] \leq \frac{2(F_{\text{CG}}^* - F_{\text{CG}}(\theta_0))}{\eta T} + L_{\text{CG}} \eta \sigma_g^2. \quad (56)$$

With $\eta = \mathcal{O}(T^{-1/2})$, the average stationarity measure is $\mathcal{O}(T^{-1/2})$.

Proof sketch. Apply the smooth ascent lemma to $F_{\text{CG}}(\theta_{t+1})$ and take conditional expectation. Unbiasedness gives the positive drift term in $\|\nabla F_{\text{CG}}(\theta_t)\|_2^2$, while the gradient-noise variance contributes the $L_{\text{CG}}\eta\sigma_g^2$ term. Summing the drift inequality over T actor steps gives (56). The full derivation is in Appendix B-H.

Proposition 4 (Nominal-Return Gap of the Regularized Actor). Let $\theta^* \in \arg\max_{\theta} J(\theta)$ and $\theta_{\text{CG}} \in \arg\max_{\theta} F_{\text{CG}}(\theta)$. If $\mathcal{R}_P(\theta) \geq 0$, $\mathcal{R}_{\Delta u}(\theta) \geq 0$, $\sup_{\theta} \|\Sigma_r(\theta)\|_F \leq \bar{\sigma}_r$, and $\sup_{\theta} \mathcal{R}_{\Delta u}(\theta) \leq \bar{R}_{\Delta u}$, then

$$J(\theta^*) - J(\theta_{\text{CG}}) \leq \lambda_{\text{ALE}} \frac{\|P\|_2 \|\Phi_u\|_2^2 \sqrt{m}}{1 - \gamma_P^2} \bar{\sigma}_r + \lambda_{\Delta u} \bar{R}_{\Delta u}. \quad (57)$$

If the implemented policy $\hat{\theta}_{\text{CG}}$ has optimization error ε_{opt} and the true/surrogate return discrepancy is bounded by Δ_{mm} , then

$$\begin{aligned} J_{\text{true}}(\theta^*) - J_{\text{true}}(\hat{\theta}_{\text{CG}}) \\ \leq \lambda_{\text{ALE}} \frac{\|P\|_2 \|\Phi_u\|_2^2 \sqrt{m}}{1 - \gamma_P^2} \bar{\sigma}_r + \lambda_{\Delta u} \bar{R}_{\Delta u} \\ + \varepsilon_{\text{opt}} + 2\Delta_{mm}. \end{aligned} \quad (58)$$

Proposition 4 is the formal performance-robustness trade-off behind the algorithm. Larger λ_{ALE} , larger certificate norm, stronger input-channel gain, weaker P -metric contraction, or a stronger temporal-smoothness penalty can improve source suppression or smooth execution but increase the possible gap to the unregularized nominal actor objective.

E. Risk Signal, Execution Layer, Safety Projection, and Posterior Residual

At execution time, the actor produces a residual target r_t^{tar} . Risk-scheduled smoothing and rate limiting are

$$\begin{aligned} \alpha_t &= \text{clip}(\alpha_0 + k_{\alpha}\rho_t, 0, 0.98), \\ \delta_t &= \frac{\delta_0}{1 + k_{\delta}\rho_t}, \end{aligned} \quad (59)$$

$$\begin{aligned} \tilde{r}_t &= \alpha_t r_{t-1} + (1 - \alpha_t) r_t^{\text{tar}}, \\ r_t &= r_{t-1} + \text{clip}(\tilde{r}_t - r_{t-1}, -\delta_t, \delta_t). \end{aligned} \quad (60)$$

The risk signal is

$$\rho_t = \text{clip}(\rho_t^V + c_b \rho_t^B, 0, 1), \quad (61)$$

where ρ_t^V is a causal Lyapunov-risk signal and ρ_t^B is a barrier-margin deficit. The implementation distinguishes predictive risk available before choosing u_t , previous-step posterior feedback such as $\mathcal{B}_{t-1}$, and posterior diagnostics computed after observing z_{t+1} . The current action cannot depend on the future posterior residual $\mathcal{B}_t$.

For fixed $\alpha_t = \alpha$, no clipping, and no projection, the smoothing layer has the augmented state $\bar{z}_t = [z_t^{\top}, r_{t-1}^{\top}]^{\top}$ and obeys

$$\begin{aligned} z_{t+1} &= Kz_t + \Phi_u \alpha r_{t-1} + \Phi_u (1 - \alpha) r_t^{\text{tar}} + w_t, \\ r_t &= \alpha r_{t-1} + (1 - \alpha) r_t^{\text{tar}}. \end{aligned} \quad (62)$$

Rate limiting, risk scheduling, clipping, and safety projection make the actual execution layer nonlinear. They are execution-layer shaping and feasibility operations. The certificate defines the residual covariance metric and diagnostics, while (47) is the execution-time projection used to repair the applied action when a local safe-action approximation is available.

The posterior Bellman-Lyapunov residual is

$$\mathcal{B}_t = [V(z_{t+1}) - \gamma_P^2 V(z_t) - \text{tr}(M \Sigma_{r,t}) - \text{tr}(PW)]_+. \quad (63)$$

We report trajectory averages and maxima. This residual diagnoses violations of the certificate prediction during actual rollouts.

F. Algorithmic Summary and Reproducibility Details

Algorithm 1 is written as an implementation-level off-policy training loop. The certificate update is a deterministic subroutine: fit (K_0, Φ_u) from a recent data window, form K_{raw} , project to K_{proj} , solve the discrete Lyapunov equation, compute γ_P , and set $M = \Phi_u^{\top} P \Phi_u$. The online loop is otherwise a standard TD3 loop with residual-action post-processing and periodic certificate refresh.

Algorithm 1 Implementation-Level CG-RVC with TD3 Backbone

Require: environment, total steps T , warm-up steps N_0 , replay buffer $\mathcal{B}$, batch size N_b , discount γ_{RL} , TD3 delay d_π , certificate period H_c , lifting map ψ , nominal controller u_{nom} , target bound β , disturbance covariance W , weights $(\lambda_{\text{ALE}}, \lambda_{\Delta u})$

Ensure: trained residual actor r_θ , certificate $(K_{\text{proj}}, P, \gamma_P)$, execution policy $u^{\text{safe}}(x)$, rollout diagnostics

- 1: Initialize actor r_θ , critics $Q_{\varphi_1}, Q_{\varphi_2}$, target copies $(\bar{\theta}, \bar{\varphi}_1, \bar{\varphi}_2)$, and residual memory $r_{-1} = 0$
- 2: for $t = 1, \dots, N_0$ do
- 3: Execute $u_{\text{nom}}(x_t)$ plus bounded noise; push $(x_t, u_t, \text{rew}_t, x_{t+1}, d_t)$ to $\mathcal{B}$
- 4: end for
- 5: $\mathcal{D} \leftarrow$ recent transitions from $\mathcal{B}$
- 6: $(K_0, \Phi_u, K_{\text{raw}}, K_{\text{proj}}, P, \gamma_P, M) \leftarrow \text{CertUpdate}(\mathcal{D}, \psi, \beta, W)$
- 7: for $t = 1, \dots, T$ do
- 8: $z_t \leftarrow \psi(x_t)$
- 9: $r_t^{\text{tar}} \leftarrow r_\theta(z_t) + \epsilon_t$ in training, or $r_\theta(z_t)$ in evaluation
- 10: Compute ρ_t ; smooth and rate-limit r_t^{tar} using (59)–(60)
- 11: $u_t^{\text{pre}} \leftarrow u_{\text{nom}}(x_t) + r_t$
- 12: $u_t^{\text{safe}} \leftarrow \Pi_{\hat{U}_{\text{safe}}(x_t)}(u_t^{\text{pre}})$, or u_t^{pre} if the projection layer is disabled
- 13: Execute u_t^{safe} ; push $(x_t, z_t, r_t, u_t^{\text{safe}}, \text{rew}_t, x_{t+1}, d_t)$ to $\mathcal{B}$; log (63)
- 14: Sample $\{(x_i, u_i, \text{rew}_i, x'_i, d_i)\}_{i=1}^{N_b}$ from $\mathcal{B}$
- 15: $u'_i \leftarrow u_{\text{nom}}(x'_i) + r_{\bar{\theta}}(\psi(x'_i))$ with clipped TD3 target noise
- 16: $y_i \leftarrow \text{rew}_i + \gamma_{\text{RL}}(1 - d_i) \min_{j=1,2} Q_{\bar{\varphi}_j}(x'_i, u'_i)$
- 17: Update $Q_{\varphi_1}, Q_{\varphi_2}$ by minimizing $\sum_i (Q_{\varphi_j}(x_i, u_i) - y_i)^2$
- 18: if $t \bmod d_\pi = 0$ then
- 19: Compute $\hat{\Sigma}_r(\theta)$ from $\{r_\theta(\psi(x_i))\}_{i=1}^{N_b}$
- 20: Update θ by (52); soft-update $(\bar{\theta}, \bar{\varphi}_1, \bar{\varphi}_2)$
- 21: end if
- 22: if $t \bmod H_c = 0$ then
- 23: Refresh $\mathcal{D}$ and rerun CertUpdate
- 24: end if
- 25: if $d_t = 1$ then
- 26: Reset environment and residual memory
- 27: end if
- 28: end for
- 29: return $r_\theta, (K_{\text{proj}}, P, \gamma_P), u^{\text{safe}}$, and diagnostics

For the matched 2D constrained RL run, the high-noise point is $\sigma = 0.30$ and seeds are 0–4. For Full3D, the run uses 5 seeds, 48 rollouts per scenario/seed, horizon 80, and noise levels 0, 0.05, 0.10, 0.20. The planar UAV study is used as a scenario-level bridge and boundary-case analysis; the primary statistical evidence is provided by the matched 2D RL and Full3D 5-seed runs. The artifact package records the source data for each table and figure.

G. Computational Footprint and Real-Time Use

CG-RVC separates certificate construction from the high-rate execution loop. The expensive operations—local surrogate fitting, stable projection, Lyapunov solve, and pullback-metric construction—are performed offline, during training, or at a low-frequency model-refresh cadence. The per-step execution path uses only the actor forward pass, risk-scheduled smoothing, rate limiting, the local safety projection, and optional diagnostic logging. Table III summarizes this split for a lifted dimension d_z and action dimension m . Embedded flight deployment would require

precomputing the certificate objects, using a bounded-complexity projection routine, and benchmarking the resulting controller on the target processor.

VI. Simulation and Controlled-Intervention Studies

Before reporting metrics, we specify the simulated systems used in each evidence block. The main learned-control benchmark is a two-dimensional constrained Lienard oscillator, while the quadrotor studies are nonlinear aerial-system diagnostics used to test whether the same source-side mechanism remains visible beyond the 2D task. The controlled-intervention studies are not plant benchmarks; they are synthetic closed-loop constructions designed to isolate source and amplifier effects.

Table V maps each experiment to the claim it supports. CI-1, CI-2, and the MIMO directional-injection test identify the source-amplifier mechanism; matched TD3 variants test method components; UAV simulations and PPO/SAC safe-RL comparators provide scope-limited context. The table states interpretation scope before metrics are introduced.

A. Metrics and Aggregation Protocol

Table VI separates evidence sources before reporting metrics. For 2D RL experiments, RMSE measures tracking error, applied variance measures postprocessed action variance, Diff MSE measures consecutive-action changes, dense violation rate measures barrier violations, and Bellman–Lyapunov max reports the largest posterior residual over evaluation rollouts. UAV diagnostics use state covariance trace and, in Full3D, applied motor-command variance. The MIMO directional-injection table reports both the predicted ratio $\text{tr}(M\Sigma_r^{\text{max}})/\text{tr}(M\Sigma_r^{\text{min}})$ and the measured state-energy ratio under equal source trace. All 2D RL tables use high-noise aggregation at $\sigma = 0.30$; Full3D reports mean $\pm$ standard deviation over 5 seeds at noise key 0.20.

Table VII clarifies the comparison protocol. The matched TD3 rows share the same backbone, update schedule, evaluation seeds, and logger; they differ only in the listed regularizer or execution layer. The broader PPO/SAC rows include penalty, Lagrangian, and shielding-style safe-RL comparators under the same high-noise safe10 evaluation, not as a safe-RL leaderboard.

B. CI-1: Amplifier Isolation

CI-1 uses the two-dimensional closed-loop family

$$A_{\text{cl}}(\alpha) = S(\alpha)\Lambda S(\alpha)^{-1}, \quad \Lambda = \text{diag}(0.93, 0.68). \quad (64)$$

Across all α , the eigenvalues, spectral radius, input channel, and source covariance are fixed; only the eigenvector geometry changes. When $\kappa(V)$ increases from approximately 1.0 to approximately 20.05, theoretical and empirical state covariance trace increase from approximately 0.37 to 15.66, and peak transient gain increases from approximately 1.0 to 5.55. The Pearson correlation between

TABLE III

Computational footprint of CG-RVC components. The table reports algorithmic cadence and dominant scaling, not measured embedded latency.

Component	Cadence	Dominant cost	Real-time implication
EDMDc/local surrogate fit	Offline, training, or low-frequency refresh	Least-squares fit in the lifted dimension; depends on batch size and dictionary size	Not placed in the high-rate inner loop
Stable projection and Lyapunov solve	After surrogate refresh	Matrix operations in d_z , including a discrete Lyapunov solve with cubic dense scaling in the worst case	Precompute K_{proj} , P , γ_P , and M before high-rate execution
Actor forward pass	Every control step	Neural-network inference cost for the selected actor architecture	Same order as the underlying TD3 actor evaluation
Risk scheduling, smoothing, and rate limiting	Every control step	Elementwise operations in m	Negligible relative to actor inference for small action dimension
Safety projection	Every control step when enabled	Projection in the action dimension or local constraint dimension	Must use a bounded-time implementation for high-rate flight control
Posterior residual logging	Evaluation step or downsampled online diagnostic	Quadratic form evaluations in d_z plus covariance estimate bookkeeping	Diagnostic; can be logged at a lower rate than the actuator loop

TABLE IV

Simulated systems and protocols used in the experimental section.

Evidence block	Simulated system	State and action	Role in the paper
CI-1 / CI-2	Synthetic discrete-time two-dimensional closed-loop family with fixed eigenvalues and controlled eigenvector conditioning	State $z \in \mathbb{R}^2$; additive input/source process with fixed or filtered source statistics	Mechanism construction for separating nonnormal amplifier geometry from source-side suppression
2D constrained RL / safe10	Controlled damped Lienard oscillator integrated by RK4: $\dot{x}_1 = x_2$, $\dot{x}_2 = -x_1/(1+x_1^2) - dx_2 + u$	State $x = (x_1, x_2)$; scalar action $u \in [-5, 5]$; target is the origin; circular unsafe set centered at $(0.5, 0.5)$ with radius 0.46	Main end-to-end residual-RL benchmark; high-noise tables use $\sigma = 0.30$ and matched 5-seed comparisons
Planar quadrotor	Nonlinear planar hover model with RK4 integration and scenario-dependent mass/inertia	State $(p_x, p_z, \theta, v_x, v_z, \omega) \in \mathbb{R}^6$; action is hover-relative thrust increment and torque; actuator limits are $(4.0, 0.4)$	External nonlinear flight bridge with nominal, heavy-payload, agile-pitch, and payload-mismatch scenarios
Full3D quadrotor	Nonlinear 3D quadrotor simulator with 13D quaternion state and a 12D hover-error diagnostic state	Diagnostic state contains position, Euler attitude, velocity, and body rates; action is four hover-relative motor thrust increments with limit 2.5	5-seed, 48-rollout-per-seed numerical source-side diagnostic over four aerial scenarios; learned-policy Full3D training is left for future work

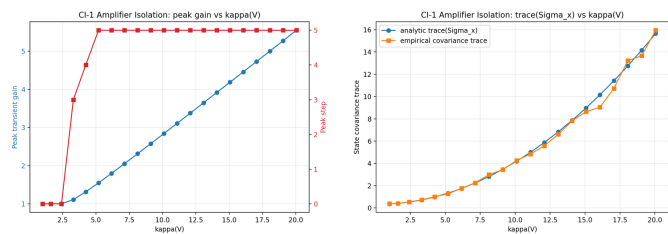


Fig. 1. CI-1 amplifier isolation. Eigenvalues, spectral radius, input channel, and source covariance are fixed; only the eigenvector geometry changes.

$\kappa(V)$ and state covariance trace is 0.9704; the correlation between $\kappa(V)$ and peak transient gain is 0.9965.

Fig. 1 isolates the amplifier factor in Proposition 2. Since ρ , G , and W are fixed, the covariance increase cannot be explained by source energy or asymptotic instability; it is attributed to the change in closed-loop geometry. This controlled construction does not prove that every nonnormal system is monotone in $\kappa(V)$; it demonstrates that nonnormal geometry can be an independent variance amplifier under fixed source conditions.

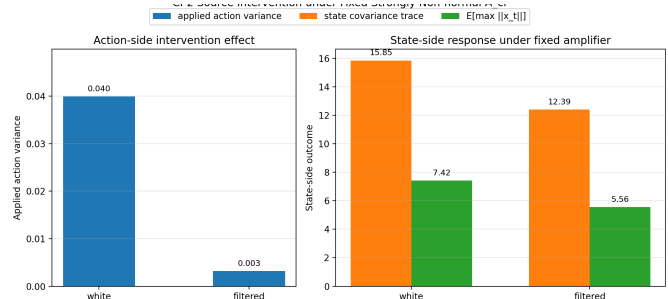


Fig. 2. CI-2 source-side intervention. The physical amplifier is fixed; filtering changes the effective source statistics and spectrum, and the filtered signal is interpreted through the augmented colored-input model.

C. CI-2: Source-Side Intervention

CI-2 fixes the strongly nonnormal matrix from CI-1 and compares white input with first-order causal filtered input. Switching from white to filtered input reduces applied action variance from 0.03992 to 0.00317 and state covariance trace from 15.849 to 12.393, while peak transient gain remains approximately 5.5486.

Fig. 2 corresponds to a fixed physical amplifier and altered source statistics. Because filtered input is colored,

TABLE V
Evidence chain, manipulated variables, key evidence, and interpretation scope.

Experiment	Question	Key evidence	Interpretation Scope
CI-1	Can nonnormal geometry independently amplify variance?	Covariance trace $0.37 \rightarrow 15.66$, peak gain $1.0 \rightarrow 5.55$	Controlled construction
CI-2	Can source suppression reduce downstream variance?	Action variance $0.03992 \rightarrow 0.00317$, state covariance $15.849 \rightarrow 12.393$	Colored-input augmented interpretation
MIMO direction	Does $M = \Phi_u^\top P \Phi_u$ distinguish equal-trace source directions?	Predicted max/min $89.7\text{--}135.8\times$; measured max/min $94.2\text{--}141.0\times$	Directional MIMO attribution
P-ALE matched	Are P-ALE, smoothing, and projection jointly needed?	CG-RVC: RMSE 0.284, Var 0.0060, Dense 0, Bmax 5.24	Matched variant comparison
Planar UAV	Is the effect observable in nonlinear flight?	Action variance decreases in all four scenarios; Payload Mismatch state covariance rises	Scenario-level simulation bridge
Full3D	Does source-side suppression persist in a 12-state/4-motor nonnormal numerical loop?	5-seed state covariance and motor variance decrease in all four scenarios	Numerical mechanism consistency; not end-to-end aerial RL
Safe RL external	Does the mechanism have RL consequences beyond Direct TD3?	PPO/CPO-style penalty, PPO/SAC-Lagrangian, and Shielded SAC remain high on at least one risk metric; full method is low on all reported metrics	Broader safe-RL comparator, not leaderboard
Metric repl.	Does the full execution layer hide scalar metric replacement effects?	Identity, diagonal, random-PSD, and one-step metrics are run under the same execution stack	Full-pipeline scalar-channel sanity check; directional evidence comes from MIMO injection
Actor feedback	Does the learned residual actor modify the amplifier?	Direct TD3 has $\rho(K_\theta) = 2.006$; CG-RVC has $\rho(K_\theta) = 0.900$ in the lifted diagnostic	Representative lifted fit

TABLE VI
Experiment scope and surrogate source.

Study	Purpose	Surrogate or construction	Claim scope
CI-1	Isolate nonnormal amplifier geometry	Synthetic two-dimensional closed-loop family; no Koopman claim	Mechanism construction
CI-2	Isolate source-side filtering under a fixed amplifier	Same synthetic amplifier with colored-input augmentation	Mechanism construction
MIMO directional injection	Test equal-trace source directions in motor space	Full3D input channel and certificate metric	Directional attribution
2D Lienard hard-obstacle RL	Test CG-RVC against matched TD3 variants	Local lifted/linearized construction in the runner	Method-component comparison
Planar quadrotor	Check nonlinear aerial simulation behavior and a boundary case	Scenario-level local diagnostic bundle	External simulation bridge
Full3D	Test source-side suppression in a 12-state, 4-motor numerical loop	Lightweight linearized closed-loop diagnostic	Numerical mechanism consistency

the result is interpreted through the augmented system in Section IV-E. It provides mechanism-consistent evidence that reducing the effective source variance entering a fixed amplifier can reduce downstream state covariance.

D. MIMO Directional Source Injection

The Full3D motor channel is MIMO, so it can directly test whether the certificate-induced input metric distinguishes source directions rather than only source magnitude. For each scenario, the closed-loop surrogate, input channel, and source trace are fixed. We form $M = \Phi_u^\top P \Phi_u$ and inject two rank-one residual covariances,

$$\Sigma_r^{\min} = s v_{\min} v_{\min}^\top, \quad \Sigma_r^{\max} = s v_{\max} v_{\max}^\top, \quad (65)$$

where $v_{\min}$ and $v_{\max}$ are the eigenvectors of M associated with its minimum and maximum eigenvalues and $s = 0.04$ keeps $\text{tr}(\Sigma_r)$ identical. An isotropic action-variance penalty assigns the two injections the same source cost. The certificate-weighted proxy instead predicts the ratio

$$\frac{\text{tr}(M \Sigma_r^{\max})}{\text{tr}(M \Sigma_r^{\min})} = \frac{\lambda_{\max}(M)}{\lambda_{\min}(M)}. \quad (66)$$

Table VIII shows that equal-trace residual sources can induce roughly two orders of magnitude difference

in downstream state energy when their directions are changed. The measured ratios track the M -based predictions across all four scenarios. This directly supports the use of the pullback metric $\Phi_u^\top P \Phi_u$: scalar source trace alone would predict no difference between the two injections, while the certificate-weighted direction predicts the observed large separation.

Because the MIMO directional-injection study is a paired directional diagnostic rather than a stochastic training benchmark, the ratios in Table VIII should not be interpreted as seed-level confidence intervals. To quantify across-scenario variability without adding new experiments, Table IX reports descriptive statistics over the four Full3D scenarios. The measured separation is consistently about 4.6% above the certificate-weighted prediction while preserving the same scenario ordering.

E. P-ALE Matched 5-Seed Comparison

This section reports a matched 5-seed comparison in the same 2D Lienard hard-obstacle runner. The table compares nine matched TD3 variants. All values in Table X are 5-seed mean $\pm$ standard deviation at high noise $\sigma = 0.30$.

Table X is the core matched method-component comparison. Plain $\text{Var}(u)$ and action-diff baselines cover scalar

TABLE VII

Comparator protocol. All methods are evaluated on the same environment, state/action normalization, noise setting, seed set, rollout horizon, and metric logger. Differences are restricted to the training objective and execution layer listed below.

Comparator group	Backbone and budget	Network / optimizer protocol	Training signal	Execution layer
Direct TD3	TD3, same replay budget and update schedule as the matched TD3 variants	Same TD3 actor-critic architecture, optimizer, batch size, target-update schedule, and evaluation seeds across all TD3 rows	Reward only; no residual covariance penalty and no safety-cost dual update	No certificate smoothing, no rate limiting, no projection unless stated
TD3 + plain $\text{Var}(u)$	Same TD3 budget as Direct TD3	Same TD3 network and optimizer protocol	Adds scalar Euclidean action-variance penalty	Same as Direct TD3
TD3 + action-diff penalty	Same TD3 budget as Direct TD3	Same TD3 network and optimizer protocol	Adds temporal action-difference penalty, corresponding to the CAPS-style smooth-action component tested in this task	Same as Direct TD3
TD3 + beta-ALE / P-ALE	Same TD3 budget as Direct TD3	Same TD3 network and optimizer protocol	Adds the indicated ALE residual-covariance penalty; only the actor-side regularizer is changed	No projection or execution smoothing unless stated in the method name
TD3 + P-ALE + smoothing	Same TD3 budget as Direct TD3	Same TD3 network and optimizer protocol	P-ALE actor regularization	Risk-scheduled smoothing and rate limiting; no safety projection
TD3 + P-ALE + projection	Same TD3 budget as Direct TD3	Same TD3 network and optimizer protocol	P-ALE actor regularization	Safety projection enabled; smoothing/rate limiting disabled unless specified
TD3 + projection only	Same TD3 budget as Direct TD3	Same TD3 network and optimizer protocol	Reward only; no P-ALE penalty	Same projection operator as the full method; no certificate-weighted source penalty
CG-RVC / full method	Same TD3 budget and seed set as matched TD3 variants	Same TD3 network and optimizer protocol	P-ALE residual covariance penalty plus auxiliary action-difference term	Risk-scheduled smoothing, rate limiting, and the same safety projection operator
PPO / SAC safe-RL comparators	Equal-budget safe10 runs under the same high-noise evaluation protocol	Fixed PPO or SAC actor-critic architectures within each backbone family	PPO-Penalty (CPO/CMDP penalty family), PPO-Lagrangian, SAC-Lagrangian, or shielded SAC objective as indicated by the method name	Projection/shielding only when stated; otherwise no CG-RVC certificate metric or residual-shaping layer

TABLE VIII

MIMO directional source injection with matched source trace. Predicted ratios use $\text{tr}(M\Sigma_r)$; measured ratios use downstream state energy under the same Full3D scenario.

Scenario	$\kappa(M)$	Pred. ratio	Meas. ratio
Nominal Hover	123.6	123.6×	128.5×
Heavy Payload	89.7	89.7×	94.2×
Agile Attitude	135.8	135.8×	141.0×
Payload Mismatch	94.5	94.5×	99.7×

TABLE IX

Descriptive across-scenario summary for MIMO directional injection. Intervals are descriptive t -intervals over the four Full3D scenarios, not seed-level stochastic confidence intervals.

Quantity	Mean	Std. dev.	Descriptive 95% interval
Predicted max/min ratio	110.90	22.36	[75.3, 146.5]
Measured max/min ratio	115.85	22.52	[80.0, 151.7]
Measured / predicted	1.046	0.008	[1.033, 1.059]
Measured – predicted	4.95	0.33	[4.42, 5.48]

variance and temporal smoothing, but neither resolves dense violations or posterior residuals. P-ALE alone is also insufficient; adding smoothing lowers applied variance and Diff MSE while leaving violations high, whereas adding projection lowers violations while leaving variance and residuals high. The full CG-RVC stack combines P -weighted ALE, residual shaping, and projection, and is

best on all five reported metrics. The table is a matched component comparison rather than an exhaustive factorial decomposition of the execution layer.

Fig. 3 highlights the distinction between ordinary action smoothing and certificate-guided residual variance control. The CAPS-style temporal penalty targets Δu directly, but it does not account for the local nonnormal propagation metric $\Phi_u^\top P \Phi_u$ or the safety projection channel. The gap between these baselines and CG-RVC supports the central methodological claim: low Euclidean action variation alone is not equivalent to suppressing certificate-weighted source energy entering a nonnormal amplifier.

F. Planar Quadrotor External Validation

The planar quadrotor experiment checks whether source suppression is observable in a nonlinear flight closed loop. At the highest noise level, `noise_key` = 0.10, action variance decreases in all four scenarios: Nominal Hover from 0.1922 to 0.0511, Heavy Payload from 0.3217 to 0.1089, Agile Pitch from 0.1665 to 0.0186, and Payload Mismatch from 0.2498 to 0.0823. State covariance decreases in Nominal Hover (0.1192 to 0.0997), Heavy Payload (0.3134 to 0.3095), and Agile Pitch (0.0884 to 0.0863), but increases in Payload Mismatch (0.1056 to 0.1226).

Fig. 4 is a supporting diagnostic for the aerial simulation bundle. It shows that filtering consistently reduces action

TABLE X
Matched method-component comparison in the 2D constrained RL task, high-noise setting.

Method	RMSE	Applied Var	Diff MSE	Dense Viol.	Bellman-Lyapunov Max
Direct TD3	0.537 ± 0.056	0.1138 ± 0.0066	0.2731 ± 0.0177	0.9389 ± 0.1366	92.391 ± 34.518
TD3 + plain $\text{Var}(u)$	0.612 ± 0.063	0.1026 ± 0.0022	0.2365 ± 0.0130	0.9278 ± 0.0800	161.860 ± 47.677
TD3 + action-diff penalty	0.565 ± 0.068	0.1118 ± 0.0059	0.2657 ± 0.0172	0.9389 ± 0.1366	109.096 ± 44.352
TD3 + beta-ALE	0.661 ± 0.085	0.0984 ± 0.0023	0.2212 ± 0.0081	0.9167 ± 0.1145	208.176 ± 79.238
TD3 + P-ALE	0.592 ± 0.056	0.1049 ± 0.0032	0.2460 ± 0.0156	0.9389 ± 0.0569	143.114 ± 45.627
TD3 + P-ALE + smoothing	0.527 ± 0.048	0.0456 ± 0.0008	0.0722 ± 0.0042	0.9278 ± 0.1204	57.246 ± 8.177
TD3 + P-ALE + projection	0.554 ± 0.033	0.0940 ± 0.0026	0.2302 ± 0.0097	0.1500 ± 0.0724	128.767 ± 27.250
TD3 + projection only	0.505 ± 0.108	0.1033 ± 0.0141	0.2761 ± 0.0366	0.1222 ± 0.0639	82.320 ± 52.322
CG-RVC / Full method	0.284 ± 0.004	0.0060 ± 0.0006	0.0063 ± 0.0018	0.0000 ± 0.0000	5.239 ± 2.698

High-noise matched TD3 comparison with CAPS-style action regularization (5 seeds, $\sigma = 0.30$)

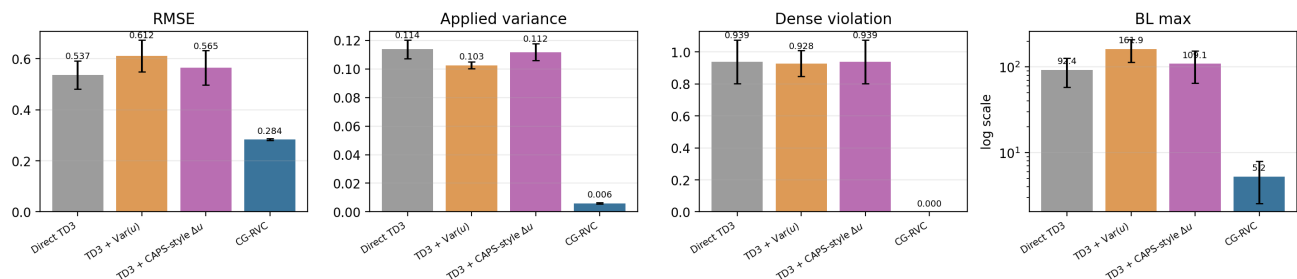


Fig. 3. High-noise matched TD3 comparison including smooth-action baselines. TD3 + plain $\text{Var}(u)$ penalizes scalar action variance, and TD3 + CAPS-style Δu penalizes temporal action differences. CG-RVC further uses certificate-weighted residual covariance and execution-layer projection, yielding lower applied variance, dense violations, and posterior Bellman-Lyapunov residual.

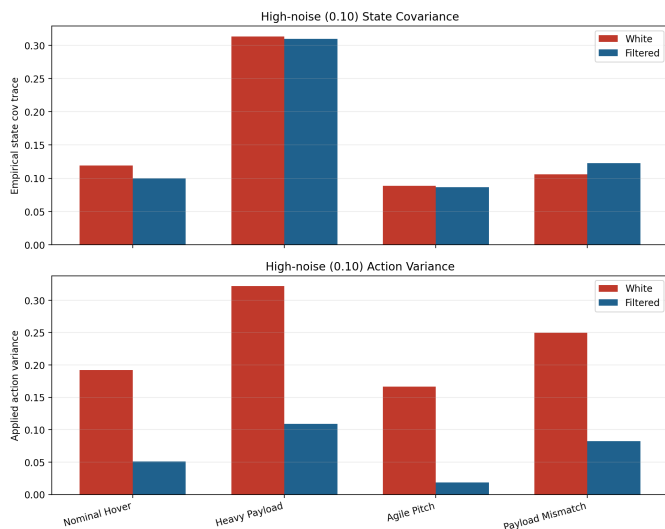


Fig. 4. Planar quadrotor simulation at the high-noise setting. The study exposes both consistent action-variance reduction and the Payload Mismatch boundary case.

variance, while Payload Mismatch gives a boundary case: scalar source variance can drop while state covariance rises if the residual energy remains aligned with high-gain directions. This motivates direction-aware penalties based on $\Phi_u^\top P \Phi_u$, rather than scalar action variance alone.

TABLE XI
Full3D scenario structural diagnostics.

Scenario	ρ	$\kappa(V)$	G_{peak}
Nominal Hover	0.926008	108.530414	7.983893
Heavy Payload	0.926002	271.757813	6.923904
Agile Attitude	0.940357	52.401604	8.912704
Payload Mismatch	0.933107	25.394522	6.772842

G. Full3D Quadrotor Numerical Supplement

The Full3D study is a numerical mechanism-consistency supplement. It fixes a 12-state, 4-motor local diagnostic and compares white versus filtered source injection under the same simulation protocol; it does not train a complete Full3D quadrotor RL controller. The simulation uses 5 random seeds, and Table XII reports the highest noise level, $\text{noise_key} = 0.20$.

Table XI shows that all Full3D scenarios are clearly nonnormal. In particular, Heavy Payload has the largest eigenvector conditioning, so the worst-case nonnormal bound is expected to be loose. Therefore, the interpretation relies on empirical covariance, PSD, and posterior diagnostics, not only on Proposition 2.

Table XII and Fig. 5 show that filtering reduces both state covariance and motor variance across all four Full3D scenarios in the 5-seed numerical simulation. This strengthens the mechanism-consistency evidence relative to the planar boundary case while keeping the claim at simulation and diagnostic level.

TABLE XII
Full3D high-noise white vs. filtered comparison, 5 seeds.

Scenario	White state cov.	Filtered state cov.	White motor var.	Filtered motor var.
Nominal Hover	0.551549 ± 0.003477	0.089192 ± 0.005230	1.901316 ± 0.019104	0.104565 ± 0.006143
Heavy Payload	0.266570 ± 0.003897	0.083648 ± 0.006377	1.763106 ± 0.013489	0.146811 ± 0.010697
Agile Attitude	0.950270 ± 0.009514	0.174127 ± 0.016088	1.898488 ± 0.016801	0.121955 ± 0.009240
Payload Mismatch	0.336788 ± 0.006714	0.115505 ± 0.008461	1.579815 ± 0.013373	0.153382 ± 0.010421

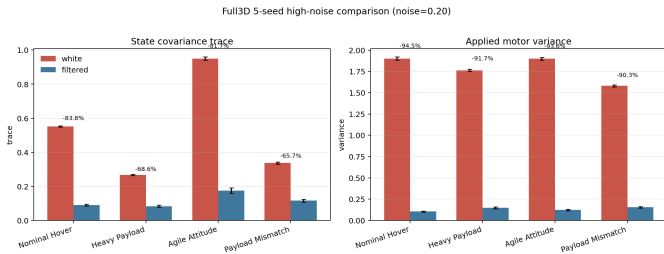


Fig. 5. Full3D quadrotor numerical simulation at the high-noise setting. The bars show 5-seed mean $\pm$ standard deviation for state covariance trace and applied motor variance.

H. External Safe RL Consequence Table

To show that the mechanism has practical RL consequences beyond Direct TD3 and matched TD3 variants, Table XIII reports the safe-RL comparator rows from the 5-seed safe10 formal run, again at $\sigma = 0.30$. The table includes penalty-based constrained policy optimization in the CPO/CMDP family, PPO-Lagrangian, SAC-Lagrangian, and Shielded SAC baselines.

Table XIII provides policy-level context rather than a safe RL leaderboard. Direct TD3 lacks constraints and variance control; projection only reduces violations; PPO-Penalty and PPO-Lagrangian remain high in tracking error, dense violations, and posterior residuals; SAC-Lagrangian improves tracking but leaves high dense violation and source variance; Shielded SAC lowers violations but still leaves high source variance and posterior residuals. The full method is low on all four reported quantities. The complete safe-RL comparator table with return and Diff MSE is retained in Appendix A. Mechanism identification still rests primarily on CI-1, CI-2, the MIMO directional-injection test, and the UAV bridge.

I. Boundary and Failure Cases

The experiment narrative includes two boundary checks. First, the planar Payload Mismatch case reduces applied action variance but increases state covariance, showing that scalar source reduction can still fail when residual energy remains aligned with high-gain directions. Second, the safe-RL consequence table shows that projection alone can reduce dense violations while leaving high residual variance and posterior residuals. These cases explain why source shaping, projection, and posterior diagnostics are reported separately.

VII. Discussion

CG-RVC differs from ordinary action smoothing because it weights residual variance by downstream certificate energy. The matched TD3 comparisons include scalar action-variance and CAPS-style temporal action-difference baselines, while the MIMO directional-injection experiment isolates the metric itself: equal-trace sources are indistinguishable to scalar variance but differ by $94\text{--}141\times$ in measured downstream state energy when aligned with low- and high-curvature eigenvectors of M .

The ALE term does not directly reduce nonnormality. Nonnormality is a geometric property of K_{proj} or of the actor-augmented local amplifier. CG-RVC suppresses certificate-weighted residual source covariance entering that amplifier. Reducing the amplifier itself would require additional objectives, such as peak-gain penalties, departure-from-normality penalties, or direct penalties on $\|\Phi_u^T P \Phi_u\|$.

a) Actor-Augmented Diagnostic.: The actor-augmented lifted-amplifier diagnostic is a local empirical consistency check. Folding the estimated residual feedback Jacobian into the diagnostic gives $K_\theta = K_{\text{proj}} + \Phi_u J_r$. In the representative lifted fit, CG-RVC reduces the spectral radius of K_θ relative to Direct TD3, indicating less destabilizing residual feedback in the local spectral geometry.

The certificate matrix P is computed for K_{proj} , and the measured P -metric factor for K_θ remains larger than one. We therefore interpret the table as a diagnostic of the learned residual feedback, while the analytical bounds use K_{proj} , the centered source interpretation, and the posterior Bellman–Lyapunov residuals.

The new robust mismatch bound clarifies the role of local modeling error. Lifted mismatch enters the same Lyapunov propagation channel as residual and disturbance sources, but it is weighted by P . Thus small Euclidean model error can still create large Lyapunov-energy error if it aligns with nonnormal high-gain directions.

Safety projection is local and conditional. Projection gives forward invariance only on a recursively feasible local CBF set where the projected action set is nonempty. P-ALE does not create feasibility. It reduces the probability and magnitude of large projection corrections by suppressing residual variance in certificate-weighted action directions.

P-ALE changes the RL objective. It does not preserve the optimizer of the unregularized actor objective; it solves a robustness-weighted residual-control problem. The suboptimality bounds in Appendix B–H quantify this

TABLE XIII

External safe-RL comparator table, high-noise setting with 5 seeds. PPO-Penalty represents the CPO/CMDP penalty-family constrained baseline used in this implementation.

Method	RMSE	Applied Var	Dense Viol.	B-L Max
Direct TD3	0.5368 ± 0.0560	0.1138 ± 0.0066	0.9389 ± 0.1366	92.3911 ± 34.5179
TD3 + projection only	0.5047 ± 0.1081	0.1033 ± 0.0141	0.1222 ± 0.0639	82.3203 ± 52.3218
PPO-Penalty (CPO/CMDP family)	1.1330 ± 0.6140	0.0937 ± 0.0055	0.8330 ± 0.1560	3083.988 ± 3827.861
PPO-Lagrangian	0.8850 ± 0.1270	0.0912 ± 0.0023	0.9330 ± 0.0840	2163.164 ± 1695.603
SAC-Lagrangian	0.4640 ± 0.0390	0.1340 ± 0.0017	0.7560 ± 0.0900	45.690 ± 7.582
Shielded SAC	0.3823 ± 0.0183	0.1056 ± 0.0100	0.1667 ± 0.0556	42.3887 ± 10.9532
CG-RVC / Full method	0.2817 ± 0.0045	0.0064 ± 0.0009	0.0000 ± 0.0000	6.7080 ± 1.4745

performance-robustness trade-off through λ_{ALE} , $\|P\|_2$, $\|\Phi_u\|_2$, and $1/(1 - \gamma_P^2)$.

The aerial simulations are mechanism diagnostics. The planar study retains the Payload Mismatch boundary case, while the Full3D 5-seed study shows consistent state-covariance and motor-variance reductions across four scenarios.

VIII. Limitations and Broader Impact

A. Limitations and Failure Modes

The main limitations are locality, feedback dependence, model mismatch, safety scope, and evaluation scope. The certificate is attached to the projected local surrogate K_{proj} used in the source-amplifier analysis. If the actor substantially changes the local amplifier, the source-only interpretation becomes diagnostic unless the feedback component is folded into the amplifier. Biased, state-dependent, or high-gain-aligned lifted error enters the robust bound through $\text{tr}(P\Sigma_e)$ and may increase posterior residuals.

Projection guarantees local recursive feasibility only under the nonempty-set assumptions in Proposition 5. The paper reports broader safe RL comparators and matched TD3 variants, including plain action-variance and CAPS-style temporal regularization baselines, but exhaustive factorial ablations, a complete Full3D aerial RL benchmark, and hardware validation are outside the present simulation study.

B. Broader Impact

The intended impact is safer and smoother deployment of learning-based continuous controllers under noise and constraints. Lower action variance and reduced posterior residuals may reduce actuator wear and improve robustness. For safety-critical deployment, CG-RVC should be combined with plant-level verification, higher-fidelity simulation, physical-system validation, and conservative fallback controllers.

IX. Conclusion

This paper proposed CG-RVC for residual learned controllers in nonlinear systems with locally nonnormal closed-loop surrogates. It treats residual-action covariance as a source, the local surrogate as an amplifier, and a projected Lyapunov certificate as the metric that

pulls downstream risk back to action space. Controlled interventions, MIMO directional injection, matched TD3 variants, and quadrotor diagnostics support this source-amplifier interpretation and show that certificate-weighted residual shaping and projection play distinct roles. Future work includes Lyapunov-constrained surrogate learning, amplifier-side penalties, anisotropic residual smoothing, and higher-fidelity simulation and physical evaluation.

Appendix A

Additional Experiment Tables

A. Hyperparameter Tuning Heuristics

Table XIV records the tuning logic used to interpret the main CG-RVC hyperparameters. It is a practical guide rather than a new sensitivity experiment. The recommended workflow is to choose β from the sampling-time or settling-time target, verify the measured contraction γ_P , increase λ_{ALE} until certificate-weighted residual covariance decreases without unacceptable return loss, then tune $\lambda_{\Delta u}$ and the risk-scheduling gains to remove remaining high-frequency action changes and projection spikes.

B. Full-Pipeline Metric Replacement

Table XV holds the execution stack fixed and changes only the actor-side covariance metric. The rows use the same nominal controller, risk-scheduled smoothing, rate limiter, and projection layer. This table is a full-pipeline scalar-channel sanity check; because the formal constrained-RL runner has one residual action channel, identity, diagonal, random-PSD, and one-step certificate metrics collapse to scalar metric variants. The MIMO directional source-injection experiment in Table VIII is therefore the directional evidence for the pullback metric.

C. Actor-Augmented Amplifier Diagnostic

Table XVI estimates the actor-augmented lifted amplifier using the formal safe10 seed-0 models. For each method, rollout states are lifted and a ridge linear fit maps lifted states to deterministic residual actor outputs; the fitted feedback is then folded into the projected surrogate as $K_\theta = K_{\text{proj}} + \Phi_u J_r$. The peak gain is computed over a 10-step horizon. This diagnostic checks whether the learned residual actor is changing the local amplifier in the expected direction.

TABLE XIV

Hyperparameter roles and tuning heuristics. This table summarizes practical tuning guidance; it does not report a separate sensitivity sweep.

Parameter	Main effect when increased	Failure symptom	Practical tuning heuristic
β	Stronger spectral-radius target for the projected certificate model	Too small: large projection mismatch and conservative certificate; too large: weak P -metric margin	Pick from (48), then verify $\gamma_P < 1$, projection cost, and posterior residuals
λ_{ALE}	Stronger penalty on certificate-weighted residual covariance	Too small: high applied variance and posterior residual; too large: sluggish residual action and return loss	Increase until $\text{tr}(M\hat{\Sigma}_r)$ and $\mathcal{B}_t$ drop; stop when RMSE or return degrades materially
$\lambda_{\Delta u}$	Stronger temporal smoothness penalty on actor outputs	Too small: high Diff MSE; too large: delayed response and tracking loss	Tune after λ_{ALE} ; use it to remove residual high-frequency action changes, not to replace P-ALE
k_α	More aggressive risk-scheduled low-pass smoothing near high risk	Too small: noisy residuals near boundaries; too large: excessive lag after risk rises	Increase until projection spikes decrease without causing visible tracking delay
k_δ	Tighter risk-scheduled rate limit near high risk	Too small: abrupt residual jumps; too large: actuator commands cannot change fast enough	Tune against Diff MSE and tracking transient; keep the nominal low-risk rate limit permissive
c_b	Larger weight on barrier-margin deficit in the risk signal	Too small: late reaction near the safe-set boundary; too large: conservative smoothing far from actual violations	Set so barrier risk dominates only near small safety margins; verify projection trigger and dense violation rates
Surrogate refresh interval	More frequent updates of $(K_0, \Phi_u, K_{\text{raw}}, K_{\text{proj}}, P, M)$	Too frequent: training overhead and noisy certificates; too infrequent: stale metric and larger posterior residuals	Refresh when posterior residuals or projection cost drift persistently, not at every control step

TABLE XV

Full-pipeline metric replacement ablation, high-noise constrained-RL setting.

Metric in actor regularizer	Return	RMSE	Applied Var	Diff MSE	Dense Viol.	B-L Max	Proj. trigger
I	-10.056 ± 0.857	0.279 ± 0.003	0.0100 ± 0.0034	0.0203 ± 0.0128	0.000 ± 0.000	6.082 ± 0.729	0.321 ± 0.150
$\text{diag}(\Phi_u^\top P \Phi_u)$	-10.056 ± 0.857	0.279 ± 0.003	0.0100 ± 0.0034	0.0203 ± 0.0128	0.000 ± 0.000	6.082 ± 0.729	0.321 ± 0.150
Random PSD	-10.056 ± 0.857	0.279 ± 0.003	0.0100 ± 0.0034	0.0203 ± 0.0128	0.000 ± 0.000	6.082 ± 0.729	0.321 ± 0.150
$\Phi_u^\top P \Phi_u$	-10.056 ± 0.857	0.279 ± 0.003	0.0100 ± 0.0034	0.0203 ± 0.0128	0.000 ± 0.000	6.082 ± 0.729	0.321 ± 0.150

TABLE XVI

Actor-augmented lifted-amplifier diagnostic, representative seed 0.

Method	Actor Jacobian	$\rho(K_\theta)$	$\kappa(V_\theta)$	$G_{\text{peak}}(K_\theta)$	$\gamma_P(K_\theta)$
Direct TD3	16.143	2.006	4.63×10^6	4.30×10^6	126.112
Euclidean variance	6.385	0.902	3.20×10^6	4.60×10^3	27.117
Low-pass execution	30.711	1.255	4.42×10^6	1.11×10^5	178.178
Projection only	17.791	1.551	5.13×10^6	7.35×10^5	147.595
CG-RVC full	0.136	0.900	3.01×10^6	3.60×10^3	1.625

D. Complete Safe-RL Comparator Table

Table XVII gives the same safe-RL comparator set as Table XIII, with return and Diff MSE also shown. PPO-Penalty and PPO-Lagrangian are not code failures: they produce finite artifacts, but they fail the control objective in this high-noise stress test because tracking error, dense violation rate, and posterior residual remain high.

Appendix B Proofs

A. Proof of Theorem 1

Let $\eta_t = B_s \xi_t + w_t$. Then

$$\Sigma_{t+1} = K \Sigma_t K^\top + Q_{\text{src}}. \quad (67)$$

Iterating yields

$$\Sigma_t = K^t \Sigma_0 (K^t)^\top + \sum_{i=0}^{t-1} K^i Q_{\text{src}} (K^i)^\top. \quad (68)$$

Since $\rho(K) < 1$, the first term vanishes and the infinite series converges, giving (16).

B. Proof of Proposition 1

If $\Sigma_a^{(2)} - \Sigma_a^{(1)} \succeq 0$, then

$$\Sigma_z^{(2)} - \Sigma_z^{(1)} = \sum_i K^i B_s (\Sigma_a^{(2)} - \Sigma_a^{(1)}) B_s^\top (K^i)^\top \succeq 0. \quad (69)$$

C. Proof of Proposition 2

From (16),

$$\|\Sigma_z\|_2 \leq \sum_i \|K^i\|_2^2 \|Q_{\text{src}}\|_2. \quad (70)$$

If $K = V \Lambda V^{-1}$, then

$$\|K^i\|_2 \leq \kappa_2(V) \rho(K)^i. \quad (71)$$

Substituting this inequality and summing the geometric series gives (17).

TABLE XVII
Complete safe-RL comparator table, high-noise setting with 5 seeds.

Method	Return	RMSE	Applied Var	Diff MSE	Dense Viol.	B-L Max
PPO-Penalty	-190.951 ± 146.747	1.133 ± 0.614	0.0937 ± 0.0055	0.2066 ± 0.0225	0.833 ± 0.156	3083.988 ± 3827.861
PPO-Lagrangian	-126.863 ± 37.426	0.885 ± 0.127	0.0912 ± 0.0023	0.1916 ± 0.0108	0.933 ± 0.084	2163.164 ± 1695.603
SAC-Lagrangian	-55.368 ± 4.221	0.464 ± 0.039	0.1340 ± 0.0017	0.3378 ± 0.0152	0.756 ± 0.090	45.690 ± 7.582
Shielded SAC	-40.977 ± 1.238	0.382 ± 0.016	0.1056 ± 0.0089	0.2972 ± 0.0339	0.167 ± 0.050	42.389 ± 9.797
Full CG-RVC	-9.322 ± 0.197	0.282 ± 0.004	0.0064 ± 0.0008	0.0072 ± 0.0030	0.000 ± 0.000	6.708 ± 1.319

D. Proof of the P -Weighted Propagation Bound

Using $V(z) = z^\top Pz$ and the conditional orthogonality assumptions,

$$\mathbb{E}[V(z_{t+1})] = \mathbb{E}[z_t^\top K^\top PK z_t] + \text{tr}(M_s \Sigma_r) + \text{tr}(PW). \quad (72)$$

By definition of γ_P in (19), $K^\top PK \preceq \gamma_P^2 P$. Thus (21) follows, and taking the stationary limit gives (22).

E. Colored-Input Augmentation

Using $y_{t+1} = \alpha y_t + (1-\alpha)\epsilon_{t+1}$ and $z_{t+1} = Kz_t + B_s y_t + w_t$ with $\bar{z}_t = [z_t^\top, y_t^\top]^\top$ gives (31). If ϵ_t is white with covariance Σ_ϵ , the scalar covariance recursion for y_t is $\Sigma_y = \alpha^2 \Sigma_y + (1-\alpha)^2 \Sigma_\epsilon$, yielding (32).

F. Proof of Theorem 2

Let $\nu_t = B_s \xi_t + w_t + \tilde{e}_t$. Then the centered dynamics are $z_{t+1} = Kz_t + \nu_t$. Expanding the Lyapunov function gives

$$V(z_{t+1}) = z_t^\top K^\top PK z_t + 2z_t^\top K^\top P \nu_t + \nu_t^\top P \nu_t. \quad (73)$$

Conditioned on $\mathcal{F}_t$, the cross term vanishes because $\mathbb{E}[\nu_t | \mathcal{F}_t] = 0$. The contraction condition gives $z_t^\top K^\top PK z_t \leq \gamma_P^2 V(z_t)$. Conditional second-order uncorrelatedness and the covariance bounds give

$$\mathbb{E}[\nu_t \nu_t^\top | \mathcal{F}_t] \preceq B_s \Sigma_r B_s^\top + W + \Sigma_\epsilon. \quad (74)$$

Therefore

$$\mathbb{E}[\nu_t^\top P \nu_t | \mathcal{F}_t] \leq \text{tr}(M_s \Sigma_r) + \text{tr}(PW) + \text{tr}(P \Sigma_\epsilon). \quad (75)$$

Taking total expectation gives

$$\mathbb{E}[V(z_{t+1})] \leq \gamma_P^2 \mathbb{E}[V(z_t)] + q_r + q_w + q_e. \quad (76)$$

Unrolling this scalar recursion gives (39); taking $t \rightarrow \infty$ gives (40).

G. Proof of Corollary 1

If $\Sigma_\epsilon \preceq \epsilon I$, then $\text{tr}(P \Sigma_\epsilon) \leq \epsilon \text{tr}(P)$. Also $V(z) \geq \lambda_{\min}(P) \|z\|_2^2$. The local inverse-Lipschitz condition implies $\|x - x^*\|_2^2 \leq \ell_\psi^{-2} \|z - z^*\|_2^2$. Combining these inequalities with (40) gives (43).

H. Actor Optimization Statements

This subsection proves the actor-stationarity and nominal-return-gap statements used in Section V. The statements concern the fixed-critic, fixed-surrogate actor update in Assumption 2. Full nonconvex actor-critic behavior with changing replay data and refreshed certificates is outside these actor-stationarity claims.

I. Proof of Proposition 3

Since F_{CG} is L_{CG} -smooth, the ascent lemma gives

$$F_{\text{CG}}(\theta_{t+1}) \geq F_{\text{CG}}(\theta_t) + \eta \nabla F_{\text{CG}}(\theta_t)^\top g_t - \frac{L_{\text{CG}} \eta^2}{2} \|g_t\|_2^2. \quad (77)$$

Taking conditional expectation and using unbiasedness and the variance bound,

$$\mathbb{E}[F_{\text{CG}}(\theta_{t+1}) | \theta_t] \geq F_{\text{CG}}(\theta_t) + c_\eta \|\nabla F_{\text{CG}}(\theta_t)\|_2^2 - \frac{L_{\text{CG}} \eta^2 \sigma_g^2}{2}. \quad (78)$$

Here $c_\eta = \eta(1 - L_{\text{CG}} \eta/2)$. For $\eta \leq 1/L_{\text{CG}}$, this coefficient is at least $\eta/2$. Summing from 0 to $T-1$ and using $F_{\text{CG}}(\theta_T) \leq F_{\text{CG}}^*$ yields (56).

J. Proof of Proposition 4

Because θ_{CG} maximizes the regularized objective,

$$J(\theta_{\text{CG}}) - \lambda_{\text{ALE}} \mathcal{R}_P(\theta_{\text{CG}}) - \lambda_{\Delta u} \mathcal{R}_{\Delta u}(\theta_{\text{CG}}) \geq J(\theta^*) - \lambda_{\text{ALE}} \mathcal{R}_P(\theta^*) - \lambda_{\Delta u} \mathcal{R}_{\Delta u}(\theta^*). \quad (79)$$

Rearranging and using nonnegativity of both regularizers gives

$$J(\theta^*) - J(\theta_{\text{CG}}) \leq \lambda_{\text{ALE}} \mathcal{R}_P(\theta^*) + \lambda_{\Delta u} \mathcal{R}_{\Delta u}(\theta^*). \quad (80)$$

Moreover,

$$\mathcal{R}_P(\theta) = \frac{\text{tr}(M \Sigma_r(\theta))}{1 - \gamma_P^2} \leq \frac{\|M\|_2 \text{tr}(\Sigma_r(\theta))}{1 - \gamma_P^2}. \quad (81)$$

Since $\|M\|_2 \leq \|P\|_2 \|\Phi_u\|_2^2$ and $\text{tr}(\Sigma_r) \leq \sqrt{m} \|\Sigma_r\|_F$, (57) follows. If $F_{\text{CG}}(\theta_{\text{CG}}) \geq F_{\text{CG}}(\theta_{\text{CG}}) - \varepsilon_{\text{opt}}$, add ε_{opt} . If $|J_{\text{true}}(\theta) - J(\theta)| \leq \Delta_{mm}$ uniformly, converting from surrogate to true return adds $2\Delta_{mm}$, giving (58).

K. Local Optimizer Perturbation under P-ALE

If J is locally μ -strongly concave near θ^* and $\sup_{\theta \in \mathcal{N}} \|\nabla \mathcal{R}_P(\theta)\|_2 \leq G_R$, any local optimizer θ_λ of $J - \lambda \mathcal{R}_P$ in $\mathcal{N}$ satisfies

$$\|\theta_\lambda - \theta^*\|_2 \leq \frac{\lambda G_R}{\mu}. \quad (82)$$

If J is also L_J -smooth, then

$$J(\theta^*) - J(\theta_\lambda) \leq \frac{L_J \lambda^2 G_R^2}{2\mu^2}. \quad (83)$$

This local perturbation result is not used by the experiments; it records the standard small- λ trade-off.

L. Projection and Feasible-Tube Statements

Let the physical safe set be

$$\mathcal{X}_{\text{safe}} = \{x : h_i(x) \geq 0, i = 1, \dots, q\}. \quad (84)$$

A local CBF-style safe action set can be written as

$$\mathcal{U}_{\text{safe}}(x) = \{u \in \mathcal{U} : h_i(f(x, u, d)) \geq (1 - \alpha_i)h_i(x), \quad \forall d \in \mathcal{D}_w, i = 1, \dots, q\}. \quad (85)$$

For a fixed closed convex set $\mathcal{U}$, Euclidean projection is nonexpansive:

$$\|\Pi_{\mathcal{U}}(u) - \Pi_{\mathcal{U}}(v)\|_2 \leq \|u - v\|_2. \quad (86)$$

Definition 1 (Local Recursive Feasibility Set). A set $\mathcal{X}_{RF} \subseteq \mathcal{X}_{\text{safe}}$ is locally recursively feasible if, for every $x \in \mathcal{X}_{RF}$, $\mathcal{U}_{\text{safe}}(x)$ is nonempty, closed, and convex, and for all $u \in \mathcal{U}_{\text{safe}}(x)$ and all $d \in \mathcal{D}_w$, $f(x, u, d) \in \mathcal{X}_{RF}$.

Proposition 5 (Local Recursive Feasibility of Safety Projection). If $x_0 \in \mathcal{X}_{RF}$ and the projected action in (47) is applied at every step, then $x_t \in \mathcal{X}_{RF} \subseteq \mathcal{X}_{\text{safe}}$ for all $t \geq 0$.

For a local polyhedral approximation,

$$\widehat{\mathcal{U}}_{\text{safe}}(x) = \{u : A_x u \leq b_x\}, \quad A_x = \begin{bmatrix} a_1(x)^\top \\ \vdots \\ a_q(x)^\top \end{bmatrix}. \quad (87)$$

Assume $u_{\text{nom}}(x)$ is inside this set with margins $\eta_j(x) = b_j(x) - a_j(x)^\top u_{\text{nom}}(x) > 0$, and $u^{\text{pre}} = u_{\text{nom}}(x) + r$ with $\mathbb{E}[r | x] = 0$ and $\mathbb{E}[rr^\top | x] = \Sigma_r(x)$.

Proposition 6 (Barrier-Edge Hitting Probability). If $M = \Phi_u^\top P \Phi_u \succ 0$ on the residual action space, then

$$\mathbb{P}(u^{\text{pre}} \notin \widehat{\mathcal{U}}_{\text{safe}}(x) | x) \leq \text{tr}(M \Sigma_r(x)) \sum_{j=1}^q \frac{\|M^{-1/2} a_j(x)\|_2^2}{\eta_j(x)^2}. \quad (88)$$

Equivalently, the right-hand side is proportional to $(1 - \gamma_P^2) \mathcal{R}_P(\theta)$.

If the residual has nonzero mean $\mu_r(x)$, replace $\eta_j(x)$ by $\bar{\eta}_j(x) = \eta_j(x) - a_j(x)^\top \mu_r(x)$ whenever $\bar{\eta}_j(x) > 0$.

Let $d_{\text{proj}}(x) = \text{dist}(u^{\text{pre}}, \widehat{\mathcal{U}}_{\text{safe}}(x))$. If a ball $\mathbb{B}(u_{\text{nom}}(x), d_s(x))$ is contained in $\widehat{\mathcal{U}}_{\text{safe}}(x)$ and $M \succeq \mu_M I$, then for any $\delta > 0$,

$$\mathbb{P}(d_{\text{proj}}(x) > \delta | x) \leq \frac{\text{tr}(M \Sigma_r(x))}{\mu_M (d_s(x) + \delta)^2}. \quad (89)$$

For polyhedral sets, Hoffman's bound gives a complementary expected correction bound [30]:

$$\mathbb{E}[d_{\text{proj}}(x)^2 | x] \leq H_x^2 \|A_x M^{-1/2}\|_F^2 \text{tr}(M \Sigma_r(x)). \quad (90)$$

Finally, suppose a Lyapunov ellipsoid $\mathcal{E}_c = \{z : V(z) \leq c\}$ maps into $\mathcal{X}_{RF}$ under the local inverse of ψ . Combining Theorem 2 with Markov's inequality and the CG-RVC specialization $M_s = M$ gives

$$\limsup_{t \rightarrow \infty} \mathbb{P}(x_t \notin \mathcal{X}_{RF}) \leq \frac{\text{tr}(M \Sigma_r) + \text{tr}(PW) + \text{tr}(P \Sigma_e)}{c(1 - \gamma_P^2)}. \quad (91)$$

M. Proof of Proposition 5

Since $x_0 \in \mathcal{X}_{RF}$, $\widehat{\mathcal{U}}_{\text{safe}}(x_0)$ is nonempty, closed, and convex; the Euclidean projection exists and belongs to that set. By recursive feasibility, the next state is in $\mathcal{X}_{RF}$ for every allowed disturbance. Repeating the argument by induction gives $x_t \in \mathcal{X}_{RF} \subseteq \mathcal{X}_{\text{safe}}$ for all $t \geq 0$.

N. Proof of Proposition 6

By the union bound,

$$\mathbb{P}(u^{\text{pre}} \notin \widehat{\mathcal{U}}_{\text{safe}}(x) | x) \leq \sum_{j=1}^q \mathbb{P}(a_j^\top r > \eta_j | x). \quad (92)$$

Chebyshev's inequality gives $\mathbb{P}(a_j^\top r > \eta_j | x) \leq a_j^\top \Sigma_r a_j / \eta_j^2$. Since $M \succ 0$,

$$a_j^\top \Sigma_r a_j = \text{tr}(M^{-1/2} a_j a_j^\top M^{-1/2} M^{1/2} \Sigma_r M^{1/2}) \quad (93)$$

$$\leq \|M^{-1/2} a_j\|_2^2 \text{tr}(M \Sigma_r). \quad (94)$$

Substituting this bound and summing over j gives (88). The nonzero-mean version follows by replacing the margin with $\bar{\eta}_j = \eta_j - a_j^\top \mu_r$.

O. Projection-Disturbance Bounds

If $\mathbb{B}(u_{\text{nom}}(x), d_s(x)) \subseteq \widehat{\mathcal{U}}_{\text{safe}}(x)$, then $d_{\text{proj}}(x) \leq (\|r\|_2 - d_s(x))_+$. Thus $\mathbb{P}(d_{\text{proj}}(x) > \delta | x) \leq \mathbb{P}(\|r\|_2 > d_s(x) + \delta | x)$. If $M \succeq \mu_M I$, then $\|r\|_2^2 \leq \mu_M^{-1} r^\top M r$. Markov's inequality gives (89).

For the polyhedral set, Hoffman's bound gives $\text{dist}(u, \widehat{\mathcal{U}}_{\text{safe}}(x)) \leq H_x \|A_x u - b_x\|_2$ [30]. With $u = u_{\text{nom}} + r$ and nonnegative margin $\eta_x = b_x - A_x u_{\text{nom}}$, $\|(A_x r - \eta_x)_+\|_2 \leq \|A_x r\|_2$. Hence

$$\begin{aligned} \mathbb{E}[d_{\text{proj}}(x)^2 | x] &\leq H_x^2 \text{tr}(A_x \Sigma_r A_x^\top) \\ &\leq H_x^2 \|A_x M^{-1/2}\|_F^2 \text{tr}(M \Sigma_r), \end{aligned} \quad (95)$$

which gives (90).

P. Projection Nonexpansiveness

For a closed convex set $\mathcal{U}$, Euclidean projection is firmly nonexpansive [31]:

$$\|\Pi_{\mathcal{U}}(u) - \Pi_{\mathcal{U}}(v)\|_2^2 \leq (u - v)^\top (\Pi_{\mathcal{U}}(u) - \Pi_{\mathcal{U}}(v)). \quad (96)$$

Cauchy-Schwarz implies (86).

Q. Local Feasible Tube Probability

If $\psi^{-1}(\mathcal{E}_c) \subseteq \mathcal{X}_{RF}$, then $x_t \notin \mathcal{X}_{RF}$ implies $V(z_t) > c$ within the local chart. Markov's inequality and Theorem 2 give (91).

Appendix C

Reproducibility and Artifact Summary

The reproducibility package for review will include the source code, configuration files, seed lists, raw rollout logs, processed metrics, plotting scripts, and the evidence map used to generate all tables and figures in this manuscript. The experimental protocol reports seeds, noise levels, rollout horizons, aggregation rules, and metric definitions in Section VII-A. Local implementation paths are intentionally omitted from the manuscript; they are maintained in the artifact index rather than in the paper body.

References

- [1] R. Cheng, A. Verma, G. Orosz, S. Chaudhuri, Y. Yue, and J. W. Burdick, “Control regularization for reduced variance reinforcement learning,” in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, ser. *Proceedings of Machine Learning Research*, vol. 97, 2019, pp. 1141–1150.
- [2] S. Mysore, B. Mabsout, R. Mancuso, and K. Saenko, “Regularizing action policies for smooth control with reinforcement learning,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 9516–9522.
- [3] A. Raffin, J. Kober, and F. Stulp, “Smooth exploration for robotic reinforcement learning,” in *Proceedings of the 5th Conference on Robot Learning (CoRL)*, ser. *Proceedings of Machine Learning Research*, vol. 164, 2022, pp. 1634–1644.
- [4] Y. Piao, Z. Zhang, and J. Chen, “Learning smooth motion planning for intelligent aerial transportation vehicles by stable auxiliary gradient,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24 464–24 474, 2022.
- [5] K. Kwak and H. Hwang, “Enhancing control policy smoothness by aligning actions with predictions from preceding states,” *arXiv:2601.18479*, 2026.
- [6] J. Achiam, D. Held, A. Tamar, and P. Abbeel, “Constrained policy optimization,” in *Proceedings of the 34th International Conference on Machine Learning (ICML)*, ser. *Proceedings of Machine Learning Research*, vol. 70, 2017, pp. 22–31.
- [7] L. Brunke, M. Greeff, A. W. Hall, Z. Yuan, S. Zhou, J. Panerati, and A. P. Schoellig, “Safe learning in robotics: From learning-based control to safe reinforcement learning,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 411–444, 2022.
- [8] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, “Control barrier functions: Theory and applications,” in *Proceedings of the 18th European Control Conference (ECC)*, 2019, pp. 3420–3431.
- [9] C. Folkestad, Y. Chen, A. D. Ames, and J. W. Burdick, “Data-driven safety-critical control: Synthesizing control barrier functions with koopman operators,” *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 665–670, 2021.
- [10] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton University Press, 2005.
- [11] P. J. Schmid, “Nonmodal stability theory,” *Annual Review of Fluid Mechanics*, vol. 39, pp. 129–162, 2007.
- [12] B. F. Farrell and P. J. Ioannou, “Generalized stability theory. part i: Autonomous operators,” *Journal of the Atmospheric Sciences*, vol. 53, no. 14, pp. 2025–2040, 1996.
- [13] B. D. O. Anderson and J. B. Moore, *Optimal Control: Linear Quadratic Methods*. Prentice Hall, 1990.
- [14] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice Hall, 1996.
- [15] Y. Chen, T. T. Georgiou, and M. Pavon, “Optimal steering of a linear stochastic system to a final probability distribution, part i,” *IEEE Transactions on Automatic Control*, vol. 61, no. 5, pp. 1158–1169, 2016.
- [16] M. Frame, A. Towne, and M. R. Jovanovic, “Beyond optimal disturbances: A statistical framework for transient growth,” *Journal of Fluid Mechanics*, vol. 983, p. A2, 2024.
- [17] F. Li, “Control of plane poiseuille flow using the kreiss constant,” in *Learning and Control for Networked Dynamical Systems*, ser. *Lecture Notes in Mathematics*. Springer, 2023.
- [18] S. L. Brunton, M. Budisic, E. Kaiser, and J. N. Kutz, “Modern koopman theory for dynamical systems,” *SIAM Review*, vol. 64, no. 2, pp. 229–340, 2022.
- [19] A. Tabassum and H. Bai, “Data-driven koopman-based modeling and control of quadcopter in a turbulent wind field,” in *Proceedings of the IEEE Conference on Control Technology and Applications (CCTA)*, 2023, pp. 490–495.
- [20] K. Zheng, P. Huang, A. M. V. Sanchez, J. A. C. Bocanegra, and G. Fettweis, “System-oriented learning: An efficient dnn learning approach for koopman bilinear representation with control,” *Machine Learning: Science and Technology*, vol. 5, no. 1, p. 015013, 2024.
- [21] M. Maleki, A. Ramirez-Serrano, and B. Boulet, “System identification and modeling of a multirotor uav: A comparative study,” in *AIAA SCITECH Forum*, 2024.
- [22] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, “A data-driven approximation of the koopman operator: Extending dynamic mode decomposition,” *Journal of Nonlinear Science*, vol. 25, no. 6, pp. 1307–1346, 2015.
- [23] J. L. Proctor, S. L. Brunton, and J. N. Kutz, “Dynamic mode decomposition with control,” *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 1, pp. 142–161, 2016.
- [24] M. Korda and I. Mezic, “Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control,” *Automatica*, vol. 93, pp. 149–160, 2018.
- [25] L. N. Trefethen, A. E. Trefethen, S. C. Reddy, and T. A. Driscoll, “Hydrodynamic stability without eigenvalues,” *Science*, vol. 261, no. 5121, pp. 578–584, 1993.
- [26] A. Hotz and R. E. Skelton, “A covariance control theory,” in *Advances in Control and Dynamic Systems*. Academic Press, 1987, vol. 24, pp. 265–313.
- [27] B. F. Farrell and P. J. Ioannou, “Variance maintained by stochastic forcing of non-normal dynamical systems associated with linearly stable shear flows,” *Physical Review Letters*, vol. 72, no. 8, pp. 1188–1191, 1994.
- [28] B. Bamieh and M. Dahleh, “Energy amplification in channel flows with stochastic excitation,” *Physics of Fluids*, vol. 13, no. 11, pp. 3258–3269, 2001.
- [29] M. R. Jovanovic and B. Bamieh, “Componentwise energy amplification in channel flows,” *Journal of Fluid Mechanics*, vol. 534, pp. 145–183, 2005.
- [30] A. J. Hoffman, “On approximate solutions of systems of linear inequalities,” *Journal of Research of the National Bureau of Standards*, vol. 49, no. 4, pp. 263–265, 1952.
- [31] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.
- [32] S. Fujimoto, H. van Hoof, and D. Meger, “Addressing function approximation error in actor-critic methods,” in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, ser. *Proceedings of Machine Learning Research*, vol. 80, 2018, pp. 1587–1596.