

Nonnormality in Reinforcement Learning Closed Loops: From Action Smoothing to Source-Amplifier Diagnostics

Bilingual independent draft package

Abstract

Continuous-control reinforcement learning often treats action jitter as a smoothness problem. This paper reframes the issue as a source-amplifier problem: residual actions, exploration noise, execution noise, and actuator disturbances become consequential when they enter a nonnormal closed-loop surrogate that can directionally amplify them into state covariance, safety-margin loss, projection burden, and posterior Bellman-Lyapunov residuals. We develop a minimal local model, a diagnostic table for source and amplifier effects, and a certificate-weighted response based on the action metric $M = \Phi_u^\top P \Phi_u$. The current evidence supports local-surrogate analysis and numerical simulation, not a global nonlinear safety guarantee. Controlled interventions show that, under fixed eigenvalues and source covariance, increasing eigenvector nonorthogonality can raise state covariance trace from about 0.37 to 15.66; under a fixed nonnormal amplifier, source filtering reduces action variance from 0.03992 to 0.00317 and state covariance from 15.849 to 12.393. A matched two-dimensional constrained-RL ablation further shows why scalar action variance and action-difference penalties are insufficient by themselves.

Keywords: reinforcement learning, nonnormality, transient amplification, action smoothing, Koopman surrogate, Lyapunov certificate, safe control

1 Introduction

High-return continuous-control policies can still produce jittery commands, dense constraint-margin losses, and frequent safety-layer corrections. The usual response is to penalize action variance, action differences, or actuator roughness. These penalties are useful, but they do not explain why the same magnitude of residual action noise can be benign in one closed loop and dangerous in another.

This paper argues that the central issue is not whether actions are smooth in Euclidean norm. The central issue is whether residual action covariance is injected into directions that a nonnormal closed-loop surrogate strongly amplifies. The action residual is a source; the closed-loop dynamics form an amplifier; state covariance, safety-margin loss, posterior Bellman-Lyapunov residuals, and projection burden are downstream consequences.

The goal is therefore a problem-first diagnostic/control framework rather than a broad safe-RL benchmark or a global nonlinear safety theorem. The concrete response

framework, certificate-guided residual variance control (CG-RVC), is used only as one implementation of this source-amplifier view. It combines local surrogate fitting, Lyapunov weighting, residual covariance control, smoothing, rate limiting, and safety projection.

2 Nonnormal Closed-Loop Amplifier

Consider a local lifted surrogate

$$z_{t+1} = Kz_t + \Phi_u r_t + w_t, \quad (1)$$

where $z_t = \psi(x_t)$ is a lifted state, r_t is a residual action or command fluctuation, Φ_u is the input channel, and w_t captures disturbances and model error. For white source covariance Σ_r and disturbance covariance W , the local stationary covariance satisfies

$$\Sigma_z = \sum_{i=0}^{\infty} K^i (\Phi_u \Sigma_r \Phi_u^\top + W) (K^i)^\top. \quad (2)$$

Equation (2) is the basic separation: action variance is not state variance itself; it becomes state variance only after propagation through the closed-loop amplifier.

Spectral-radius stability, $\rho(K) < 1$, is not enough to rule out short-horizon energy growth. If $K = V\Lambda V^{-1}$ is highly nonnormal, eigenvectors can be ill-conditioned even when eigenvalues lie inside the unit disk. Diagnostics such as $\kappa(V)$, peak transient gain $G_{\text{peak}} = \max_{t \leq H} \|K^t\|_2$, resolvent norms, and Kreiss-like proxies are therefore relevant to reinforcement-learning evaluation.

3 Why Ordinary Smoothing Is Not Enough

A scalar action-variance penalty,

$$R_{\text{var}} = \text{tr}(\Sigma_u), \quad (3)$$

measures total source energy but ignores input direction. A temporal roughness penalty,

$$R_{\Delta u} = \mathbb{E} \|u_t - u_{t-1}\|^2, \quad (4)$$

suppresses fast changes but does not determine whether the remaining residual covariance aligns with high-amplification directions.

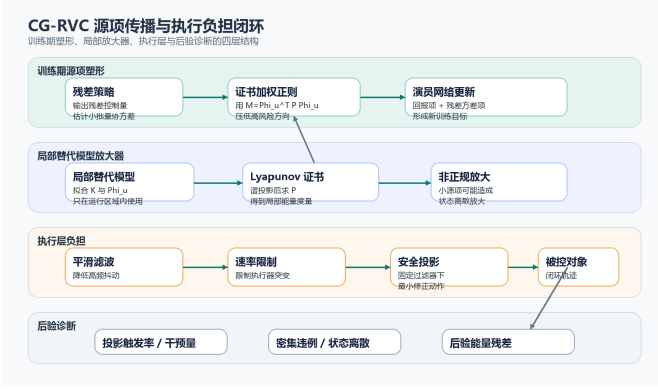


Figure 1: Source-amplifier view. Residual actions, exploration noise, execution noise, and disturbances enter the input channel Φ_u and are propagated by a nonnormal closed-loop amplifier K . The downstream consequences are state covariance, safety-margin loss, posterior Bellman-Lyapunov residuals, and projection burden.

Given a local Lyapunov metric P with measured contraction factor γ_P ,

$$K^T P K \preceq \gamma_P^2 P, \quad (5)$$

the residual input directions can be weighted by

$$M = \Phi_u^T P \Phi_u. \quad (6)$$

The corresponding P -weighted action Lyapunov energy (P-ALE) is

$$R_{P-ALE} = \frac{\text{tr}(M \Sigma_r)}{1 - \gamma_P^2}. \quad (7)$$

Smoothness is temporal regularity. P-ALE is amplifier-aware residual covariance control. They are related, but not equivalent.

4 Minimal Certificate-Guided Response

The response framework has five steps. First, fit a local surrogate $z_{t+1} = K_0 z_t + \Phi_u u_t + w_t$ using local linearization, EDMDC-style regression, or a replay-buffer local model. Second, construct a nominal or projected closed-loop matrix, for instance $K_{\text{raw}} = K_0 - \Phi_u L$ followed by $K_{\text{proj}} = \Pi_\beta(K_{\text{raw}})$. Third, solve a Lyapunov equation or inequality to obtain P . Fourth, add the P-ALE penalty in Eq. (7) to the actor objective. Fifth, apply execution-layer shaping through smoothing, rate limiting, and safety projection.

The spectral target β is not a P-metric contraction guarantee. Bounds and diagnostics should use the measured γ_P . Moreover, smoothing and projection make the execution layer nonlinear; the certificate is therefore a local design and diagnostic object rather than a global nonlinear plant certificate.

Table 1: Diagnostic quantities for a source-amplifier paper.

Layer	Quantity	Interpretation
Source	$\text{tr}(\Sigma_r)$, applied variance, diff MSE	How much residual or executed action jitter is injected.
Amplifier	$\rho(K)$, $\kappa(V)$, G_{peak} , resolvent norm	Whether stable-looking dynamics can transiently amplify sources.
Certificate	P , γ_P , $M = \Phi_u^T P \Phi_u$	Which action directions are high-energy under the local certificate.
Consequence	state covariance trace, dense violation, B-L residual	Whether source jitter becomes dispersion or safety-margin loss.
Execution	projection trigger rate, normalized correction	How much repair the safety layer must perform.

If the residual policy is state-dependent, the residual is not purely exogenous. A local diagnostic should estimate

$$K_\theta = K + \Phi_u J_r, \quad (8)$$

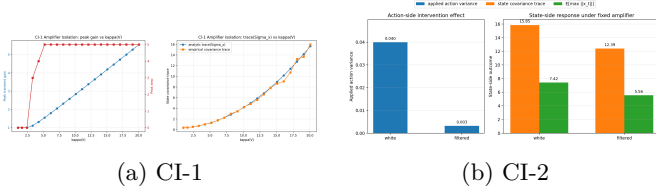
where J_r is a finite-difference actor-Jacobian proxy. When J_r is large, the feedback component must be folded into the amplifier.

5 Evidence

The evidence should separate source, amplifier, and safety repair. Figure 2 combines two controlled interventions. CI-1 fixes eigenvalues, spectral radius, input channel, and source covariance while changing eigenvector geometry. In the archived result, $\kappa(V)$ grows from about 1.0 to 20.05, state covariance trace grows from about 0.37 to 15.66, and peak transient gain grows from about 1.0 to 5.55. This supports the statement that nonnormal geometry can independently amplify a fixed source in a controlled matrix family.

CI-2 fixes the strongly nonnormal amplifier and changes the source. In the archived result, the filtered source lowers action variance from 0.03992 to 0.00317 and state covariance trace from 15.849 to 12.393, while structural peak gain stays fixed near 5.55. This is source-side evidence, but the filtered input is colored; it should be interpreted through an augmented colored-input model.

Table 2 shows why ordinary smoothing is incomplete in the high-noise constrained-RL task. Scalar action variance and action-difference penalties do not remove dense violations or posterior residuals. Projection alone lowers violations but leaves high applied variance and residuals. The full response combines source shaping and execution repair.



plains why ordinary smoothing can be insufficient and why certificate-weighted residual covariance is a natural diagnostic/control response for nonnormal RL closed loops.

Figure 2: Controlled interventions. CI-1 manipulates non-normal amplifier geometry while fixing source statistics. CI-2 manipulates source statistics while fixing the amplifier.

Table 2: Compressed high-noise 2D constrained-RL evidence. Values are mean summaries from the archived matched ablation.

Method	RMSE	Var.	Viol.	B-L max
Direct TD3	0.537	0.1138	0.9389	92.4
TD3 + $\text{Var}(u)$	0.612	0.1026	0.9278	161.9
TD3 + action diff	0.565	0.1118	0.9389	109.1
Projection only	0.505	0.1033	0.1222	82.3
Full response	0.284	0.0060	0.0000	5.24

Projection burden gives an operational downstream diagnostic. In the archived 2D high-noise run, normalized correction is about 0.09362 for projection-only, 0.05426 for P-ALE plus projection, and 0.01093 for the full method. A residual-feedback diagnostic gives another caveat: the finite-difference actor-Jacobian proxy is about 16.23 for Direct TD3 and 0.274 for the full method. This supports, but does not prove, the source-side interpretation.

6 Limitations

The claim level is deliberately narrow. First, all certificate claims are local-surrogate-level statements, not global nonlinear safety guarantees. Second, P-ALE reduces certificate-weighted residual covariance entering the amplifier; it does not directly reduce nonnormality of K . Third, if residual feedback is strong, the source-only interpretation becomes diagnostic unless the feedback is folded into Eq. (8). Fourth, projection depends on local feasibility and cannot create feasibility. Fifth, Full3D quadrotor and ControlGym results are simulation-level relevance checks, not hardware validation or broad safe-RL benchmark dominance.

7 Conclusion

Low-variance safe reinforcement learning should report both the source and the amplifier. The relevant object is not only $\text{Var}(u)$ or $\|u_t - u_{t-1}\|^2$, but the full chain from Σ_r through Φ_u , K , P , and M into state covariance, safety-margin loss, posterior Bellman-Lyapunov residuals, and projection burden. This source-amplifier framing ex-