KAGGLEYEs ACADEMY E-PRINT REGISTRY
ARCHIVAL PREPRINT / OPEN SCHOLARLY RECORD

KAGG RECORD
KAGG:2608.00004

Safe Velocity Interface Identification Under Input Saturation

Siyuan Cui • **Kaiwen Deng** • **Zhenghao Liu**

Autonomous Systems Center, National University of Defense Technology, Changsha, China

PUBLISHED	REGISTERED	RECORD TYPE
August 7, 2026	August 14, 2026	Preprint

ABSTRACT

Active calibration of opaque quadruped velocity interfaces, as in GAUGE, issues planner-conditioned probe commands and fits a command-to-motion model online. Informative probes can push the robot toward the saturation boundary of the interface, where the response becomes nonlinear and potentially unsafe on rough ground. This paper adds a probe safety layer: each probe is admitted only if it keeps the robot inside a verified reachable-and-recoverable region computed from a conservative inner model, namely a linear command-response model valid below a certified saturation bound with uncertainty from a recursive least-squares confidence set. Probes are selected to maximize planner-weighted information subject to saturation, excursion, and recovery-time constraints, and the safety margin is derived from the same confidence set as the identification estimate, so the safe region expands as confidence grows. Under bounded-disturbance and inner-linearity assumptions, the analysis shows that the certified identification residual and the safety margin evolve jointly: the margin is nonincreasing, safety holds with high probability, and both quantities decay at the same rate under persistent excitation. In 120 simulated identification sessions on a Unitree Go2 in a cluttered course with an 8-degree slope, the method reached a 10.8 percent RMS identification residual versus 11.3 percent for unconstrained GAUGE-style probes, reduced the mean maximum excursion outside the safe region from 5.4 cm to 4.7 cm, and reached navigation-grade accuracy in 52 s instead of 60 s with smaller probes. In ten hardware sessions, a runtime monitor truncated one probe after a 0.04 m/s violation of the certified saturation bound.

INDEX TERMS velocity interface identification, active calibration, input saturation, safe exploration, safety filter, quadruped robots, Unitree Go2

KAGG NOTE A probe is only as useful as the data it returns and only as safe as the model that certifies it; deriving the safety margin and the identification estimate from one confidence set lets the safe region grow exactly as fast as the evidence allows.

01 Introduction

Quadruped robots are increasingly deployed as velocity-commanded platforms: a navigation stack produces a planar body-velocity command, and a learned or model-based locomotion controller converts that command into joint torques [1, 2, 3]. From the planner's point of view the locomotion controller is an *opaque velocity interface*. Its command-to-motion response depends on the terrain, the payload, the gait, and the training distribution of the policy, and it is rarely documented beyond nominal command limits. A planner that assumes unit gain and instantaneous tracking will, on a slope or on loose ground, overshoot turns, under-deliver forward speed, and accumulate lateral drift.

GAUGE [4] addresses this mismatch through planner-conditioned active calibration: it issues probe commands chosen around the commands the planner is likely to request and fits a command-to-motion model online. Active probing is effective because it concentrates excitation where the downstream planner needs accuracy. It also introduces a risk that the calibration literature has not treated explicitly. An informative probe tends to be a large probe, and a large probe superimposed on a nominal command can push the commanded velocity toward the saturation boundary of the interface. Near that boundary the response is no longer linear, the controller's tracking authority is reduced, and on rough ground a probe that would be harmless on flat floor can produce a lateral excursion large enough to clip an obstacle or trigger a slip.

This paper studies how to keep active interface identification safe without giving up its efficiency. The key observation is that the quantity that makes probing informative (uncertainty about the interface) is the same quantity that determines how conservative a safety certificate must be. The approach therefore derives both the identification estimate and the safety margin from a single confidence set over a conservative inner model of the interface. Probes are admitted only if every interface consistent with that confidence set keeps the robot inside a region from which it can be steered back to the nominal trajectory within a bounded recovery time. As data accumulate, the confidence set contracts, the margin shrinks, the certified saturation bound grows, and the admissible probe set expands.

The contributions are as follows.

1. A conservative inner model of a saturating quadruped velocity interface, consisting of a linear autoregressive response valid inside a certified saturation box, together with a one-sided test that expands the box only when data support linearity.

2. A reachable-and-recoverable probe constraint, evaluated in closed form, that bounds the worst-case position excursion and recovery time of a probe over all interfaces in the current

KAGGLEYEs ACADEMY • E-PRINT REGISTRY • KAGG:2608.00004

confidence set, and a planner-weighted probe schedule that maximizes information subject to that constraint.

3. A joint-evolution result showing that, under stated assumptions, the certified identification residual and the safety margin are affine functions of the same nonincreasing quantity, so that the safe region is nested and expands at the rate permitted by the data, while safety holds with high probability.
4. An evaluation on a Unitree Go2 in 120 simulated identification sessions on a cluttered course with an 8-degree slope, with comparisons against unconstrained GAUGE-style probes, a probe-only calibration, and a fixed-margin inner-model baseline, an ablation of the three main components, and an analysis of recorded failure cases including one hardware runtime-monitor event.

02 Related Work

2.1 Calibration and adaptation of locomotion interfaces

GAUGE [4] is the work closest to ours. It treats the quadruped velocity controller as an opaque interface, issues probe commands conditioned on the planner's anticipated command distribution, and fits a calibrated command-to-motion model online so that the planner can compensate for gain, coupling, and lag. GAUGE optimizes where to probe for accuracy, but it does not constrain how far a probe may push the robot, and it does not model the saturation boundary as a region whose location is itself uncertain. The present work keeps the planner-conditioned objective and the opaque-interface view, and adds a safety layer that restricts probes to a certified inner region and adapts that region as identification confidence grows. The baselines in Section 5 include an unconstrained GAUGE-style prober so that the effect of the safety layer can be isolated.

A second line of work adapts the low-level policy rather than calibrating the interface seen by the planner. Universal policies with online system identification [5] estimate dynamics parameters from recent history and condition the controller on them. RMA [6] trains an adaptation module that infers an extrinsics latent from proprioception, enabling rapid adjustment to terrain and payload changes. DreamWaQ [7] learns an implicit terrain representation for robust blind locomotion. These methods improve tracking but leave a residual, terrain-dependent command-response map that a planner still has to model; they also do not expose an explicit saturation bound. Walk These Ways [8] exposes behavior parameters as part of the command interface, which enlarges the interface that a calibrator must characterize. Neural-Fly [9] learns a disturbance representation offline and adapts linear coefficients online for agile flight; its separation into a fixed basis and online-adapted linear weights parallels our linear-in-parameters inner model, but it targets tracking control rather than safe probe design.

2.2 Learned locomotion and navigation on rough terrain

Reinforcement-learned locomotion controllers trained in simulation with domain randomization [10, 11, 12] now produce robust gaits on challenging ground [1, 3], and massively parallel training in GPU simulators [13, 2, 14] makes such controllers easy to obtain. High-speed policies [15] demonstrate that the commanded-velocity envelope is often limited by the training curriculum rather than by hardware, which is one source of the saturation behavior studied here. End-to-end locomotion-navigation policies [16, 17] sidestep a separate velocity interface, whereas modular stacks with explicit traversability estimation [18] retain it and therefore benefit from calibration. Our method assumes the modular case: a planner issues body-velocity commands to a fixed, opaque locomotion policy.

2.3 Safe exploration and safe learning

Safe exploration seeks to gather information without leaving a safe set. Moldovan and Abbeel [19] formalize ergodicity-preserving exploration in MDPs. Turchetta et al. [20] expand a certified safe set using Gaussian-process confidence bounds, and Berkenkamp et al. [21] extend this idea to region-of-attraction estimates for continuous systems with Lyapunov guarantees. Koller et al. [22] propagate GP uncertainty through an MPC horizon to certify that exploratory trajectories can return to a safe set. Lew et al. [23] propose a sequential exploration-exploitation framework that learns dynamics with confidence bounds while maintaining safety, which is conceptually close to our joint treatment of identification and safety. Brunke et al. [24] survey these and related methods. Our formulation borrows the expanding-safe-set principle of [20, 21] and the return-to-safe-set requirement of [22], but specializes them to a low-dimensional linear-in-parameters interface model for which the reachable tube and the recovery time admit closed-form bounds that can be evaluated for dozens of candidate probes within a 20 ms control period. Unlike these works, the uncertain object includes the location of a saturation boundary, which is expanded by a separate one-sided test.

2.4 Safety filters, barrier functions, and reachability

Control barrier functions [25, 26] enforce forward invariance of a safe set through a quadratic program that minimally modifies a nominal input; learning-based variants correct model error in the barrier condition [27] or combine barriers with reinforcement learning [28]. Discrete-time barrier constraints within MPC [29] and multi-layered barrier-MPC designs for legged robots [30] show that such filters can operate at locomotion rates. Predictive safety filters [31, 32] certify a proposed input by verifying that a backup trajectory to a safe terminal set exists, and Hamilton-Jacobi reachability [33, 34] provides least-restrictive safety controllers for systems with bounded disturbance. Hsu et al. [35] unify these as safety filters, and shielding [36] provides the discrete analog for reinforcement learning. In legged robotics, Agile But Safe [37] switches between an agile policy and a learned recovery policy guided by a reach-avoid value, and Yang et al. [38] use a safety layer with a recovery policy during locomotion learning. Our probe constraint is a predictive safety filter whose backup policy is "remove the probe and resume nominal tracking"; the reachable tube is computed over the identification confidence set rather than a fixed disturbance set, which is what couples the filter's conservatism to identification progress. We use a box-and-tube overapproximation rather than a full Hamilton-Jacobi computation because the probe filter must evaluate many candidates per decision.

2.5 Sample complexity and active identification

Non-asymptotic analyses of linear system identification [39, 40] give self-normalized confidence bounds of the form used in Section 4. Active identification methods choose inputs to maximize information about linear [41] or nonlinear [42] systems with provable rates. These works assume either unconstrained inputs or a fixed input-energy budget; they do not restrict inputs

KAGGLEYEs ACADEMY • E-PRINT REGISTRY • KAGG:2608.00004

by a state-dependent safety region that itself depends on the estimate. Our analysis combines their confidence bounds with a safety constraint and shows that the two evolve together.

03 Problem Formulation

3.1 Velocity interface

Let $k \in \mathbb{N}$ index control steps of period Δt. The planner issues a body-frame velocity command $u_k = (u_k^x, u_k^y, u_k^\omega)^\top \in \mathcal{U} \subset \mathbb{R}^3$ (forward, lateral, and yaw rate), where $\mathcal{U} = \{u : |u^i| \le \bar{u}^i\}$ is the nominal command box published for the locomotion policy. The realized body velocity $v_k \in \mathbb{R}^3$ is estimated from proprioception and odometry. We model the interface as

$$v_{k+1} = F(v_k, u_k) + w_k, \tag{1}$$

where F is unknown and w_k is a disturbance due to terrain contact and estimation error. The interface saturates: there exists an unknown box $\mathcal{U}_{\text{lin}}(\rho^\star) = \{u : |u^i| \le \rho^{\star,i}\} \subseteq \mathcal{U}$ such that F is affine in (v, u) for $u \in \mathcal{U}_{\text{lin}}(\rho^\star)$ and is nonlinear outside it. On slopes and cluttered terrain, $\rho^\star$ can be markedly smaller than the nominal limits $\bar{u}$.

3.2 Inner model

Define the regressor $z_k = (v_k^\top, u_k^\top, 1)^\top \in \mathbb{R}^7$. Inside the linear region,

$$v_{k+1} = \Xi^\star z_k + w_k, \qquad \Xi^\star = [A^\star \;\; B^\star \;\; c^\star] \in \mathbb{R}^{3\times 7}, \tag{2}$$

where $A^\star$ captures the tracking lag and cross-coupling of the velocity channels, $B^\star$ the command gain and coupling, and $c^\star$ a constant bias such as yaw drift on a slope. The steady-state command-to-velocity map of the inner model is $h(u) = (I - A^\star)^{-1}(B^\star u + c^\star)$. A *conservative inner model* is a pair $(\mathcal{C}_k, \underline{\rho}_k)$ of a parameter set $\mathcal{C}_k \ni \Xi^\star$ and a certified saturation bound $\underline{\rho}_k \le \rho^\star$ (componentwise), both maintained online.

3.3 Probes and nominal tracking

At each step the commanded input is $u_k = u_k^{\text{nom}} + \delta_k$, where u_k^{nom} is the planner's tracking command and δ_k is a probe. Probes are organized in epochs of H_p steps; at the start of epoch n a probe shape $\delta^{(n)} = (\delta_0, \dots, \delta_{H_p-1})$ is chosen from a finite candidate set $\mathcal{D}$ that always contains the zero probe. Let $p_k \in \mathbb{R}^2$ and ψ_k denote the planar position and heading, and let p_k^{nom}, ψ_k^{nom} be the trajectory predicted under $\delta \equiv 0$. The excursion is $e_k = \|p_k - p_k^{\text{nom}}\|$, and the deviation in velocity and heading is $\tilde{v}_k = v_k - v_k^{\text{nom}}$, $\tilde{\psi}_k = \psi_k - \psi_k^{\text{nom}}$.

3.4 Reachable-and-recoverable region

Let d_k be the free clearance around the nominal trajectory, computed from the local occupancy and elevation map minus the body half-width. Given a margin $m_k \ge 0$, a recovery tolerance ε_v, and a maximum recovery time $\bar{T}_{\text{rec}}$, the *reachable-and-recoverable region* at step k is

$$\mathcal{R}_k = \Big\{(e, \tilde{v}, \tilde{\psi}) : \; e \le d_k - m_k, \; T_{\text{rec}}(\tilde{v}, \tilde{\psi}) \le \bar{T}_{\text{rec}}\Big\}, \tag{3}$$

where T_{rec} is the worst-case time, over all interfaces in $\mathcal{C}_k$, needed to bring $\|\tilde{v}\|_\infty$ below ε_v under $\delta \equiv 0$ while the excursion remains below $d_k - m_k$. The region is "verified" in the sense that membership is certified for every $\Xi \in \mathcal{C}_k$ and every admissible disturbance.

3.5 Problem statement

Find a probe policy and an update rule for $(\mathcal{C}_k, \underline{\rho}_k, m_k)$ such that (i) with high probability the robot state remains in $\mathcal{R}_k$ for all k, (ii) the identification residual on planner-relevant commands decreases as fast as possible, and (iii) the margin m_k is no more conservative than the current identification uncertainty requires.

04 Method

The method has five components, executed every control step ($\Delta t = 20$ ms) or every probe epoch: recursive identification with a confidence set, a certified saturation bound, a closed-form reachable-and-recoverable test, an adaptive margin, and a planner-weighted probe selector. A runtime monitor running at the 200 Hz state-estimation rate closes the loop. Table 1 lists the settings used in all experiments.

4.1 Recursive identification and confidence set

Only samples with $u_k \in \mathcal{U}_{\text{lin}}(\underline{\rho}_k)$ enter the regression. With regularization $\lambda > 0$, the Gram matrix and the regularized least-squares estimate are

$$V_k = \lambda I + \sum_{t<k} z_t z_t^\top, \qquad \hat{\Xi}_k = \Big(\sum_{t<k} v_{t+1} z_t^\top\Big) V_k^{-1}, \tag{4}$$

updated recursively in $O(49)$ operations per step. Assuming bounded, conditionally zero-mean disturbance (Assumption 2), the self-normalized martingale bound used in non-asymptotic system identification [39, 40] gives, for each row j and uniformly in k, with probability at least $1 - \delta$,

$$\begin{aligned} \|\Xi_j^\star - \hat{\Xi}_{k,j}\|_{V_k} &\le \beta_k, \\ \beta_k &= \bar{w}\sqrt{2\log\frac{3\det(V_k)^{1/2}}{\delta\,\lambda^{7/2}}} + \sqrt{\lambda}\,S, \end{aligned} \tag{5}$$

where $\bar{w}$ bounds $|w_k^j|$, S bounds $\|\Xi_j^\star\|_2$, and the factor 3 is a union bound over the velocity channels. The confidence set is the intersection of all past ellipsoids,

$$\mathcal{C}_k = \bigcap_{t \le k} \big\{\Xi : \|\Xi_j - \hat{\Xi}_{t,j}\|_{V_t} \le \beta_t, \; j = 1, 2, 3\big\}. \tag{6}$$

Because Eq. (5) holds uniformly in time, $\Xi^\star \in \mathcal{C}_k$ for all k on the same event, so the intersection is valid and $\mathcal{C}_{k+1} \subseteq \mathcal{C}_k$. In implementation $\mathcal{C}_k$ is represented by its support function along a fixed set of validation directions (below), which is all the subsequent tests require.

4.2 Certified saturation bound

The bound $\underline{\rho}_k$ starts at $\underline{\rho}_0 = 0.5\,\bar{u}$ and is expanded one channel at a time by a step $\Delta\rho$. For channel i, let $\mathcal{B}_k^i$ be the set of past samples whose command lies in the band $\underline{\rho}_k^i - \Delta\rho \le |u^i| \le \underline{\rho}_k^i$. The band is *certified linear* if it contains at least $n_{\min}$ samples and every sample satisfies

$$\big|v_{t+1}^j - \hat{\Xi}_{k,j} z_t\big| \le \beta_k \|z_t\|_{V_k^{-1}} + \bar{w} \quad \forall j, \; \forall t \in \mathcal{B}_k^i. \tag{7}$$

If the test passes, probes are allowed to reach $\underline{\rho}^i + \Delta\rho$; the new samples are admitted to the regression only after the next test on the new band passes. The test is one-sided: it can certify linearity in a band already visited but never extrapolates beyond it. Let δ_ρ denote the probability, over all tests in a session, that a band containing nonlinear response passes Eq. (7); this is controlled by $n_{\min}$ and by requiring the band to include samples near its outer edge.

KAGGLEYEs ACADEMY • E-PRINT REGISTRY • KAGG:2608.00004

4.3 Closed-form reachable tube

Consider a candidate probe $\delta^{(n)}$ starting at step k_0 from a measured deviation $(\tilde{v}_{k_0}, \tilde{\psi}_{k_0}, e_{k_0})$. For any $\Xi \in \mathcal{C}_{k_0}$, the velocity deviation obeys $\tilde{v}_{k+1} = A\tilde{v}_k + B\delta_k + (\Xi - \hat{\Xi})z_k^{\text{nom}} + w_k$, where the third term reflects that the nominal prediction uses $\hat{\Xi}$. We bound it by scalar recursions. Let $\hat{a} = \|\hat{A}\|_\infty$, let $\sigma_k = \lambda_{\min}(V_{k_0})^{-1/2}$, and define for $\ell = 0, 1, \ldots$

$$\bar{v}_{\ell+1} = (\hat{a} + \beta_{k_0}\sigma_k)\,\bar{v}_\ell + \|\hat{B}\delta_\ell\|_\infty + \beta_{k_0}\big(\|z_\ell^{\text{nom}}\|_{V_{k_0}^{-1}} + \sigma_k\|\delta_\ell\|_2\big) + \bar{w}, \tag{8}$$

$$\bar{\psi}_{\ell+1} = \bar{\psi}_\ell + \Delta t\,\bar{v}_\ell^{\omega}, \tag{9}$$

$$\bar{e}_{\ell+1} = \bar{e}_\ell + \Delta t\big(\bar{v}_\ell^{xy} + |u_\ell^{x,\text{nom}}|\,\bar{\psi}_\ell\big), \tag{10}$$

initialized at the measured deviation, with $\delta_\ell = 0$ for $\ell \geq H_p$ (the recovery phase). Here $\bar{v}^{xy}$ and $\bar{v}^{\omega}$ are the translational and yaw components of the bound obtained by applying Eq. (8) channelwise, and the last term of Eq. (10) accounts for lateral drift induced by heading error at forward speed. By Cauchy-Schwarz, $|(\Xi_j - \hat{\Xi}_j)z| \leq \beta\|z\|_{V^{-1}}$ for any $\Xi \in \mathcal{C}$, and $\|z\|_{V^{-1}} \leq \|z^{\text{nom}}\|_{V^{-1}} + \sigma\|z - z^{\text{nom}}\|_2$, so Eqs. (8)–(10) overapproximate every realization consistent with $\mathcal{C}_{k_0}$ and $|w| \leq \bar{w}$. The recovery time is

$$T_{\text{rec}}(\delta^{(n)}) = \Delta t \cdot \min\{\ell \geq H_p : \bar{v}_\ell \leq \varepsilon_v\}, \tag{11}$$

which is finite whenever the contraction factor $\hat{a} + \beta_{k_0}\sigma_k < 1$ and the steady-state uncertainty term is below ε_v; otherwise the candidate is rejected. The worst-case excursion is $E(\delta^{(n)}) = \max_{\ell \leq H_p + T_{\text{rec}}/\Delta t}\bar{e}_\ell$. The recursions cost $O(H_p + \bar{T}_{\text{rec}}/\Delta t)$ scalar operations per candidate, so the full candidate set is evaluated well within one control period.

4.4 Admissibility constraints

A candidate probe is admissible at epoch n if

$$\text{(C1)}\quad |u_\ell^{i,\text{nom}} + \delta_\ell^i| \leq \underline{\rho}_{k_0}^i \quad \forall \ell < H_p,\ i, \tag{12}$$

$$\text{(C2)}\quad E(\delta^{(n)}) \leq d_{\min,n} - m_{k_0}, \tag{13}$$

$$\text{(C3)}\quad T_{\text{rec}}(\delta^{(n)}) \leq \bar{T}_{\text{rec}}, \tag{14}$$

where $d_{\min,n}$ is the minimum clearance along the nominal trajectory over the probe and recovery window. Constraint (C1) keeps the probe inside the certified inner model, (C2) keeps the reachable tube inside the clearance corridor, and (C3) ensures the robot can return to nominal tracking before the planner's next replanning cycle so that consecutive probes do not compound. The nominal command is assumed to be applied with the planner's own feedback on e_k; the tube is therefore conservative, because it ignores the corrective action of that feedback. If u^{nom} itself lies outside $\mathcal{U}_{\text{lin}}(\underline{\rho})$, for example during an aggressive turn, probing is suspended for that epoch.

4.5 Adaptive margin and joint confidence evolution

Let $\mathcal{Z}_{\text{val}}$ be a finite set of regressors built from the planner's forecast command distribution, as in planner-conditioned calibration [4]. The *certified residual* is the half-width of $\mathcal{C}_k$ along these directions,

$$\varrho_k = \max_{z \in \mathcal{Z}_{\text{val}}} \max_j \tfrac{1}{2} \max_{\Xi, \Xi' \in \mathcal{C}_k} \big|(\Xi_j - \Xi'_j)z\big|, \tag{15}$$

and the margin is

$$m_k = m_0 + \gamma\,\varrho_k, \tag{16}$$

Table 1: Method settings used in all sessions.

Quantity	Value
Control period Δt	20 ms
State estimation and logging rate	200 Hz
Nominal limits $\bar{u}\ (x, y, \omega)$	1.5 m/s, 0.6 m/s, 1.5 rad/s
Initial saturation bound $\underline{\rho}_0$	$0.5\,\bar{u}$
Bound step $\Delta\rho$	0.1 m/s (0.1 rad/s)
Probe epoch H_p	25 steps (0.5 s)
Max. recovery time $\bar{T}_{\text{rec}}$	1.0 s
Recovery tolerance ε_v	0.05 m/s
Probe amplitudes	0.05–0.40 m/s, 6 levels
Confidence level $1 - \delta$	0.95
Regularization λ	1.0
Margin offset m_0, gain γ	0.03 m, 0.5 s

where m_0 absorbs state-estimation error and γ (units of seconds) converts a velocity uncertainty into a distance over the probe horizon. Because $\mathcal{C}_k$ is nested, ϱ_k and hence m_k are nonincreasing, and the region $\mathcal{R}_k$ in Eq. (3) grows for fixed clearance. A single ellipsoid gives the computable upper bound $\varrho_k \leq \beta_k \max_{z \in \mathcal{Z}_{\text{val}}}\|z\|_{V_k^{-1}}$, which we use when the intersection has not tightened it.

The margin, the tube in Eq. (8), the linearity test in Eq. (7), and the identification estimate all read the same pair (V_k, β_k). We refer to this coupling as *joint confidence evolution*: a probe that reduces uncertainty along planner-relevant directions immediately loosens the safety constraints for subsequent probes, and the saturation bound expands only once the same confidence set certifies the band below it.

4.6 Planner-weighted probe selection

Among admissible candidates, the selector maximizes

$$J(\delta^{(n)}) = \operatorname{tr}\big(WV_{k_0}^{-1}\big) - \operatorname{tr}\Big(W\big(V_{k_0} + \textstyle\sum_{\ell < H_p}\hat{z}_\ell\hat{z}_\ell^\top\big)^{-1}\Big) - \eta\|\delta^{(n)}\|_2^2 + \mu\,\mathbb{1}\big[\delta^{(n)} \text{ enters a band } \mathcal{B}^i\big], \tag{17}$$

where $W = |\mathcal{Z}_{\text{val}}|^{-1}\sum_{z \in \mathcal{Z}_{\text{val}}} zz^\top$ weights information by the planner's forecast commands, $\hat{z}_\ell$ are the regressors predicted under $\hat{\Xi}_{k_0}$, η penalizes probe energy, and μ rewards probes that sample the band just below the current saturation bound, which is the only way the certified bound can grow. The first two terms are the reduction in planner-weighted A-optimal uncertainty. The candidate set $\mathcal{D}$ contains steps and doublets along each of the three channels and along four mixed (x, ω) and (y, ω) directions, at the amplitudes listed in Table 1, plus the zero probe. Doublets are favored by Eq. (17) near obstacles because their tube in Eq. (10) integrates to a smaller excursion for the same information.

4.7 Runtime monitor

The tube certifies probes under Assumptions 1–3 below. To handle violations, a monitor compares each 200 Hz velocity estimate with the tube prediction. If $|v^j - \hat{v}^j| > \bar{v}^j + \varepsilon_{\text{mon}}$ for any channel, the monitor truncates the current probe ($\delta \equiv 0$ for the remainder of the epoch), retracts the saturation bound on the probed channel by $\Delta\rho$, excludes the epoch's samples from the regression, and logs the event. The monitor converts an assumption violation into a lost epoch rather than an unbounded excursion, although the excursion incurred before detection is not certified.

KAGGLEYEs ACADEMY • E-PRINT REGISTRY • KAGG:2608.00004

4.8 Analysis

Assumption 1 (Inner linearity)**.** *For all $u \in \mathcal{U}_{\text{lin}}(\rho^\star)$, Eq. (2) holds with $\|\Xi_j^\star\|_2 \le S$, and $\|A\|_\infty + \beta_0 \lambda^{-1/2} < 1$ for all $\Xi \in \mathcal{C}_0$.*

Assumption 2 (Disturbance)**.** *w_k is conditionally zero-mean given the past and $|w_k^j| \le \bar{w}$ for all k, j.*

Assumption 3 (Initialization and clearance)**.** *$\underline{\varrho}_0 \le \rho^\star$, the initial deviation is zero, and $d_k \ge m_0 + \gamma \varrho_0 + E(0)$ for all k, where $E(0)$ is the tube excursion of the zero probe.*

Theorem 1 (Joint evolution of residual and margin)**.** *Under Assumptions 1–3, with probability at least $1-\delta-\delta_\rho$, the following hold for all k.*

(a) Coupling. *For every $z \in \mathcal{Z}_{\text{val}}$ and channel j, $|(\hat{\Xi}_{k,j} - \Xi_j^\star)z| \le 2\varrho_k = 2(m_k - m_0)/\gamma$, and m_k is nonincreasing.*

(b) Safety and feasibility. *The zero probe is admissible at every epoch, and the realized deviation satisfies $e_k \le d_k - m_k$ and returns below ε_v within $\bar{T}_{\text{rec}}$ after each probe; that is, the state remains in $\mathcal{R}_k$.*

(c) Rate. *If there exist $M \in \mathbb{N}$ and $\kappa > 0$ such that the probes executed in every window of M consecutive epochs satisfy $\lambda_{\min}(\sum z_t z_t^\top) \ge \kappa$, then after N such windows $\varrho_k \le \beta_k z_{\max}/\sqrt{\lambda + \kappa N}$ with $z_{\max} = \max_{\mathcal{Z}_{\text{val}}} \|z\|_2$, so that both the certified residual and $m_k - m_0$ decay as $O(\sqrt{\log N / N})$.*

Proof sketch. Let $\mathcal{E}$ be the event on which Eq. (5) holds for all k and every passed linearity test is correct; $\Pr(\mathcal{E}) \ge 1-\delta-\delta_\rho$ by the uniform self-normalized bound and the definition of δ_ρ. On $\mathcal{E}$, $\Xi^\star \in \mathcal{C}_k$ for all k and $\underline{\rho}_k \le \rho^\star$. (a) Since $\hat{\Xi}_k$ and $\Xi^\star$ both lie in $\mathcal{C}_k$, their difference along z is at most the width $2\varrho_k$ of Eq. (15); the identity with m_k follows from Eq. (16), and monotonicity from $\mathcal{C}_{k+1} \subseteq \mathcal{C}_k$. (b) Under (C1) every executed command lies in $\mathcal{U}_{\text{lin}}(\rho^\star)$, so Eq. (2) governs the deviation. Each term of Eqs. (8)–(10) upper-bounds the corresponding term for $\Xi^\star$ and w, so by induction over ℓ the realized $(\|\tilde{v}\|_\infty, \tilde{\psi}, e)$ is bounded by $(\bar{v}, \bar{\psi}, \bar{e})$; (C2) and (C3) then give membership in $\mathcal{R}_k$. The zero probe is admissible because $\varrho_k \le \varrho_0$ and Assumption 3 reserves clearance for its tube, which yields recursive feasibility. (c) Under the excitation condition, $\lambda_{\min}(V_k) \ge \lambda + \kappa N$, and the single-ellipsoid bound on ϱ_k gives the rate, with β_k growing as $O(\sqrt{\log\det V_k}) = O(\sqrt{\log N})$. The margin inherits the rate through Eq. (16). □

Remark 1. *Part (a) is a statement about the certified residual, not the empirical one. The empirical residual of Section 5 also contains terrain-induced variability that is not explained by any Ξ and therefore has a floor. Part (c) requires excitation inside the admissible set; if clearance is persistently tight, the admissible probes may not excite all directions and the rate degrades, which is the expected price of safety. Part (b) holds only on the assumptions; the runtime monitor addresses violations of Assumption 1, and Section 5.6 reports one such event.*

05 Experiments

5.1 Platform and setup

All experiments use a Unitree Go2 quadruped driven by a fixed learned locomotion policy that accepts planar body-velocity commands. The identification sessions were run in simulation in Isaac Lab, the successor of Orbit [14] built on GPU-parallel physics [13], using a saturation-aware interface plant, on a cluttered lab course that includes an 8-degree slope. A total of 120 simulated identification sessions were run with the probe safety layers switched on or off according to the method under test. In addition, 10 real sessions were run on the Go2 hardware with the proposed method. Probe commands were logged at 200 Hz. Excursion is measured as the deviation of the robot outside the verified reachable-and-recoverable region of Eq. (3); for methods without a safety layer, the region is computed in shadow mode from the same logged data so that the metric is comparable across methods.

5.2 Compared methods

Unconstrained GAUGE-style probes. Planner-conditioned probes as in [4], with no constraint other than the nominal command box $\mathcal{U}$.

Probe-only calibration (no safety layer). The proposed probe objective of Eq. (17) and candidate set, with constraints (C1)–(C3) and the monitor removed.

Conservative inner-model probes (fixed margin). Constraints (C1)–(C3) with the saturation bound fixed at $\underline{\rho}_0$ and the margin fixed at its initial value $m_0 + \gamma\varrho_0$, so that the safe region never expands.

Ours. The full method of Section 4.

5.3 Metrics

The *identification residual RMS* is the root-mean-square open-loop prediction error of the identified model on held-out probe-free tracking segments, expressed as a percentage of the RMS measured velocity on those segments. The *maximum excursion outside the safe region* is the per-session maximum distance by which the robot left the region of Eq. (3), averaged over sessions. The *identification time to nav-grade* is the elapsed session time after which the residual stays below a navigation-grade threshold r_{nav} that is shared by all methods. The *probe command RMS* is the RMS of the probe component δ_k of the logged command.

5.4 Main comparison

Table 2 reports the main comparison. The proposed method attains the lowest identification residual (10.8%), the smallest maximum excursion outside the safe region (4.7 cm), the shortest time to nav-grade accuracy (52 s), and the smallest probe command RMS (14.9 cm/s). Relative to unconstrained GAUGE-style probes, this is a reduction of 0.5 percentage points in residual (about 4.4% relative), 0.7 cm in excursion (about 13%), 8 s in time to nav-grade (about 13%), and 1.5 cm/s in probe RMS (about 9%).

Three observations follow. First, the safety layer did not cost identification accuracy or speed in these sessions; the method is better than all three baselines on all four metrics. A plausible explanation, consistent with the smaller probe RMS, is that constraints (C1)–(C3) remove probes that drive the interface into its nonlinear region; such probes are large but yield samples that the linear inner model cannot explain and that inflate the residual. The probe-only baseline, which uses the same objective without the safety layer, has the highest residual (11.6%) and the largest probe RMS (16.8 cm/s), which supports this reading. Second, the fixed-margin baseline reduces excursion relative to the unconstrained probes (5.1 versus 5.4 cm) but does not improve the residual (11.4 versus 11.3%), in-

KAGGLEYEs ACADEMY • E-PRINT REGISTRY • KAGG:2608.00004

Table 2: Main comparison on the Go2 identification sessions. Lower is better in every column.

Method	Identification residual RMS (%)	Max excursion outside safe region (cm)	Identification time to nav-grade (s)	Probe command RMS (cm/s)
Unconstrained GAUGE-style probes	11.3	5.4	60	16.4
Probe-only calibration (no safety layer)	11.6	5.3	55	16.8
Conservative inner-model probes (fixed margin)	11.4	5.1	56	16.6
Ours	**10.8**	**4.7**	**52**	**14.9**

Table 3: Ablation of the safety-layered prober.

Variant	Identification residual RMS (%)
Ours (safety-layered probes)	10.8
w/o adaptive margin expansion	11.7
w/o recovery-time check	12.5
w/o joint confidence evolution	12.2

dicating that a safety layer alone is insufficient unless it adapts with confidence. Third, the excursion differences in session averages are modest: all methods have mean maximum excursions between 4.7 and 5.4 cm, and the full method does not drive the excursion to zero. A nonzero measured excursion is expected under the model, because the measurement includes state-estimation error and terrain disturbances that are not captured by the bounded-disturbance assumption, and because the region is evaluated with the online confidence set at each instant. The guarantee of Theorem 1(b) should therefore be read as a design target that shapes the probes, not as a measured property of every session.

5.5 Ablation

Table 3 removes one component at a time. Removing adaptive margin expansion raises the residual from 10.8% to 11.7% (+0.9 points), removing the recovery-time check (C3) raises it to 12.5% (+1.7 points), and removing joint confidence evolution (margin and saturation bound updated on a fixed schedule rather than from (V_k, β_k), and $\mu = 0$) raises it to 12.2% (+1.4 points).

All three ablated variants have higher residuals than every baseline in Table 2, so a partial safety layer is worse than none in this metric. The recovery-time check is the most important component. Our interpretation, which the logged metrics do not isolate, is that without (C3) consecutive probes start before the previous transient has decayed; the resulting samples are correlated with unmodeled transients, and more of them are flagged by the runtime monitor and discarded. Without joint confidence evolution, the margin and the saturation bound can be loosened before the identification data support it, or held tight after they do; either mismatch wastes probes. Removing adaptive expansion alone is the least harmful, consistent with the fixed-margin baseline being close to the unconstrained one.

5.6 Failure cases

Unconstrained probes on the slope. On the 8-degree slope, 6 of 20 sessions with unconstrained probes exceeded 21 cm of excursion outside the safe region, and one of these triggered a slip event. These excursions are roughly four times the session-averaged maximum of 5.4 cm in Table 2, which shows that the mean hides a heavy tail concentrated on sloped terrain, where the saturation level of the downhill-lateral and yaw channels is lowest. This is the regime the safety layer is designed for. The failure log does not list comparable slope excursions for the safety-layered prober, but it also does not report per-method tail statistics, so we do not claim a quantified reduction of the tail.

Fixed margin stays conservative. The fixed-margin baseline is recorded as taking 2.3 times longer to reach nav-grade accuracy, because it stays conservative past the point the data support. This ratio is larger than the ratio of the session means in Table 2 (56 s versus 52 s, about 1.08); we therefore read it as a failure observation in a subset of sessions rather than typical behavior. The mechanism is the one that adaptive expansion targets: with $\underline{\rho}$ and m frozen, probes cannot reach the commands near the saturation band that the planner later requests, so the residual on those commands decreases only slowly.

Hardware: inner-model violation. In one of the 10 real sessions, the inner model's saturation bound was violated by 0.04 m/s during a probe: the interface saturated below the certified bound $\underline{\rho}$, so Assumption 1 failed locally. The runtime monitor detected the departure from the tube, truncated the probe, retracted the bound, and logged the event. The event shows both that the certified bound can be wrong on hardware and that the monitor provides the intended fallback. It also shows that the guarantee of Theorem 1 is only as good as the linearity test of Eq. (7).

06 Limitations

The quantitative results come from simulation with a saturation-aware interface plant; the hardware evidence consists of 10 sessions and is reported qualitatively, so the size of the sim-to-real gap for these metrics is not established. The improvements in session averages are moderate (for example 0.5 percentage points in residual and 0.7 cm in excursion relative to unconstrained probes), and the results do not include per-session dispersion, so the statistical strength of individual differences cannot be assessed from these data. The theoretical guarantee relies on a bounded disturbance and on the correctness of the one-sided linearity test; the hardware violation of 0.04 m/s shows that the latter can fail, and the measured excursions show that the guarantee is not met exactly in practice. The inner model is affine and time-invariant within a session; gait transitions, payload changes, or terrain changes within a session would require forgetting or re-initialization of $\mathcal{C}_k$, which breaks the nesting used in Theorem 1. The tube is a box-type overapproximation that ignores the planner's corrective feedback and may be conservative in narrow corridors, where the excitation condition of Theorem 1(c) can fail. Finally, the navigation-grade threshold and the clearance computation are inherited from the planning stack and were not varied.

07 Conclusion

KAGGLEYEs ACADEMY • E-PRINT REGISTRY • KAGG:2608.00004

This paper added a safety layer to active identification of opaque quadruped velocity interfaces. Probes are restricted to a reachable-and-recoverable region computed from a conservative inner model, and the margin, the saturation bound, and the identification estimate are all derived from one confidence set, so that the safe region expands as identification confidence grows. Under stated assumptions, the certified residual and the margin evolve jointly and safety holds with high probability. On a simulated Go2 in a cluttered course with an 8-degree slope, the method reached a 10.8% identification residual in 52 s with a 4.7 cm mean maximum excursion and smaller probes than unconstrained GAUGE-style probing, a probe-only calibration, and a fixed-margin baseline; the ablation identified the recovery-time check as the most important component. Future work will quantify tail excursions on hardware, handle within-session interface changes, and replace the box tube with a tighter reachability computation.

References

[1] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter. Learning Quadrupedal Locomotion over Challenging Terrain. arXiv:2010.11251, 2020.

[2] N. Rudin, D. Hoeller, P. Reist, and M. Hutter. Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning. arXiv:2109.11978, 2021.

[3] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter. Learning Robust Perceptive Locomotion for Quadrupedal Robots in the Wild. arXiv:2201.08117, 2022.

[4] T. Zang, H. Huang, S. Cheng, and W. Zhang. GAUGE: Planner-Conditioned Active Calibration of Opaque Quadruped Velocity Interfaces. arXiv:2607.02841, 2026.

[5] W. Yu, J. Tan, C. K. Liu, and G. Turk. Preparing for the Unknown: Learning a Universal Policy with Online System Identification. arXiv:1702.02453, 2017.

[6] A. Kumar, Z. Fu, D. Pathak, and J. Malik. RMA: Rapid Motor Adaptation for Legged Robots. arXiv:2107.04034, 2021.

[7] I. M. A. Nahrendra, B. Yu, and H. Myung. DreamWaQ: Learning Robust Quadrupedal Locomotion With Implicit Terrain Imagination via Deep Reinforcement Learning. arXiv:2301.10602, 2023.

[8] G. B. Margolis and P. Agrawal. Walk These Ways: Tuning Robot Control for Generalization with Multiplicity of Behavior. arXiv:2212.03238, 2022.

[9] M. O'Connell, G. Shi, X. Shi, K. Azizzadenesheli, A. Anandkumar, Y. Yue, and S.-J. Chung. Neural-Fly Enables Rapid Learning for Agile Flight in Strong Winds. arXiv:2205.06908, 2022.

[10] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke. Sim-to-Real: Learning Agile Locomotion For Quadruped Robots. arXiv:1804.10332, 2018.

[11] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. arXiv:1710.06537, 2017.

[12] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter. Learning Agile and Dynamic Motor Skills for Legged Robots. arXiv:1901.08652, 2019.

[13] V. Makoviychuk et al.. Isaac Gym: High Performance GPU-Based Physics Simulation For Robot Learning. arXiv:2108.10470, 2021.

[14] M. Mittal et al.. Orbit: A Unified Simulation Framework for Interactive Robot Learning Environments. arXiv:2301.04195, 2023.

[15] G. B. Margolis, G. Yang, K. Paigwar, T. Chen, and P. Agrawal. Rapid Locomotion via Reinforcement Learning. arXiv:2205.02824, 2022.

[16] N. Rudin, D. Hoeller, M. Bjelonic, and M. Hutter. Advanced Skills by Learning Locomotion and Local Navigation End-to-End. arXiv:2209.12827, 2022.

[17] D. Hoeller, N. Rudin, D. Sako, and M. Hutter. ANYmal Parkour: Learning Agile Navigation for Quadrupedal Robots. arXiv:2306.14874, 2023.

[18] J. Frey, M. Mattamala, N. Chebrolu, C. Cadena, M. Fallon, and M. Hutter. Fast Traversability Estimation for Wild Visual Navigation. arXiv:2305.08510, 2023.

[19] T. M. Moldovan and P. Abbeel. Safe Exploration in Markov Decision Processes. arXiv:1205.4810, 2012.

[20] M. Turchetta, F. Berkenkamp, and A. Krause. Safe Exploration in Finite Markov Decision Processes with Gaussian Processes. arXiv:1606.04753, 2016.

[21] F. Berkenkamp, M. Turchetta, A. P. Schoellig, and A. Krause. Safe Model-based Reinforcement Learning with Stability Guarantees. arXiv:1705.08551, 2017.

[22] T. Koller, F. Berkenkamp, M. Turchetta, and A. Krause. Learning-based Model Predictive Control for Safe Exploration. arXiv:1803.08287, 2018.

[23] T. Lew, A. Sharma, J. Harrison, A. Bylard, and M. Pavone. Safe Active Dynamics Learning and Control: A Sequential Exploration-Exploitation Framework. arXiv:2008.11700, 2020.

[24] 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. arXiv:2108.06266, 2021.

[25] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada. Control Barrier Function Based Quadratic Programs for Safety Critical Systems. arXiv:1609.06408, 2016.

[26] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada. Control Barrier Functions: Theory and Applications. arXiv:1903.11199, 2019.

[27] A. J. Taylor, A. Singletary, Y. Yue, and A. D. Ames. Learning for Safety-Critical Control with Control Barrier Functions. arXiv:1912.10099, 2019.

[28] R. Cheng, G. Orosz, R. M. Murray, and J. W. Burdick. End-to-End Safe Reinforcement Learning through Barrier Functions for Safety-Critical Continuous Control Tasks. arXiv:1903.08792, 2019.

[29] J. Zeng, B. Zhang, and K. Sreenath. Safety-Critical Model Predictive Control with Discrete-Time Control Barrier Function. arXiv:2007.11718, 2020.

[30] R. Grandia, A. J. Taylor, A. D. Ames, and M. Hutter. Multi-Layered Safety for Legged Robots via Control Barrier Functions and Model Predictive Control. arXiv:2011.00032, 2020.

[31] K. P. Wabersich and M. N. Zeilinger. Linear Model Predictive Safety Certification for Learning-Based Control. arXiv:1803.08552, 2018.

[32] K. P. Wabersich and M. N. Zeilinger. A Predictive Safety Filter for Learning-Based Control of Constrained Nonlinear Dynamical Systems. arXiv:1812.05506, 2018.

[33] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin. Hamilton-Jacobi Reachability: A Brief Overview and Recent Advances. arXiv:1709.07523, 2017.

[34] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. Gillula, and C. J. Tomlin. A General Safety Framework for Learning-Based Control in Uncertain Robotic Systems. arXiv:1705.01292, 2017.

KAGGLEYEs ACADEMY • E-PRINT REGISTRY • KAGG:2608.00004

[35] K.-C. Hsu, H. Hu, and J. F. Fisac. The Safety Filter: A Unified View of Safety-Critical Control in Autonomous Systems. arXiv:2309.05837, 2023.

[36] M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu. Safe Reinforcement Learning via Shielding. arXiv:1708.08611, 2017.

[37] T. He, C. Zhang, W. Xiao, G. He, C. Liu, and G. Shi. Agile But Safe: Learning Collision-Free High-Speed Legged Locomotion. arXiv:2401.17583, 2024.

[38] T.-Y. Yang, T. Zhang, L. Luu, S. Ha, J. Tan, and W. Yu. Safe Reinforcement Learning for Legged Locomotion. arXiv:2203.02638, 2022.

[39] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. On the Sample Complexity of the Linear Quadratic Regulator. arXiv:1710.01688, 2017.

[40] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht. Learning Without Mixing: Towards a Sharp Analysis of Linear System Identification. arXiv:1802.08334, 2018.

[41] A. Wagenmaker and K. Jamieson. Active Learning for Identification of Linear Dynamical Systems. arXiv:2002.00495, 2020.

[42] H. Mania, M. I. Jordan, and B. Recht. Active Learning for Nonlinear System Identification with Guarantees. arXiv:2006.10277, 2020.

KAGGLEYEs ACADEMY • E-PRINT REGISTRY • KAGG:2608.00004