

# Nonlinear Sensitivity and Perturbation Analysis for Learning-based Legged Locomotion

Kyle Morgenstein

## I. MOTIVATION

Reinforcement learning (RL) has become the standard control technique in legged locomotion due to its robustness, expressiveness, and ease of deployment. Despite these attributes, RL control policies still fail catastrophically when evaluated on out of distribution (OOD) inputs. Due to the black box nature of RL policies, safety efforts largely focus on observing potentially destabilizing changes in the action space of the policy, and triggering safe modes when risk thresholds are reached (e.g. over-current protection). Efforts to quantify the distribution of valid inputs during training may result in more proactive runtime anomaly detection, but such efforts provide only weak guarantees of stability given the distribution shift between simulation-based training and hardware deployment. In this work we propose a more rigorous treatment of anomaly detection using tools from nonlinear sensitivity analysis. Treating the trained policy as an artifact, we aim to exploit the structure of the learning-based controller to provide stronger guarantees to prevent catastrophic failure at runtime.

## II. NONLINEAR SENSITIVITY FOR RL CONTROLLERS

### A. RL Preliminaries

We train a locomotion control policy in simulation to produce joint offset targets which are then actuated by a proportional-derivative (PD) control law

$$\begin{aligned} u_t &\sim \pi_\theta(u_t|x_t) \\ \tau_t &= K_p(q_{\text{default}} + K_u u_t - q_t) - K_d \dot{q}_t \end{aligned}$$

with gains  $K_p$ ,  $K_d$ , and  $K_u$ . The state of the system  $x_t = [v, \omega, g_{\text{proj}}, q, \dot{q}, u_{t-1}, c_t]$  contains the linear and angular velocity, the orientation of the body frame projected onto the gravity vector, the joint configuration, and the last policy output. The controller also receives a high-level command  $c_t = [v_{\text{SE2}}, f, h_{\text{body}}]$  containing the desired SE2 velocity, gait frequency, and body height. The policy  $\pi_\theta$  is represented as a multivariate Gaussian with mean  $\mu_\theta(x_t)$  from a multi-layer perceptron (MLP) and covariance  $\Sigma_\theta$  parameterized by  $\theta = [\{W_i, b_i\}_0^{L-1}, \Sigma]$ . An MLP is a nonlinear operator composition  $g_i(z_i) = (\sigma_i \circ f_i)(z_i)$  of  $L$  affine maps  $f_i(z_i) = W_i^T z_i + b_i$  and nonlinear differentiable activations  $\sigma_i \in C^1$ :

$$\begin{aligned} z_0 &= x_t \\ z_{i+1} &= g_i(z_i) = \sigma_i(W_i^T z_i + b_i) \\ \mu_\theta &= g_{L-1} \circ g_{L-2} \circ \dots \circ g_0 \\ \pi_\theta(u_t|x_t) &= \mathcal{N}(\mu_\theta(x_t), \Sigma_\theta) \end{aligned}$$

### B. Nonlinear Sensitivity Preliminaries

First define the Jacobian of the trained policy at state  $x_t$

$$J_\pi(x_t) = \left. \frac{\partial \pi_\theta(x)}{\partial x} \right|_{x_t}.$$

We may assume that the Jacobian is full rank. Then, the singular value decomposition (SVD) of the Jacobian

$$\begin{aligned} J_\pi(x_t) &= U \Sigma V^T \\ \Sigma &= \text{diag}(\sigma_1, \dots, \sigma_{\min}) \end{aligned}$$

with the singular values  $\Sigma$  in descending order. The columns of  $V$  define the principle directions of variation with corresponding singular values dictating the sensitivity of the action given perturbations in  $v_i$ . The Jacobian can be found numerically via auto-differentiation tools. A second order analysis can be performed by perturbing the input

$$\pi_\theta(x_t + \delta x) = \pi_\theta(x_t) + J_\pi(x_t) \delta x + \frac{1}{2} \delta x^T H_\pi \delta x + O(\|\delta x\|^3)$$

where  $H_\pi(x_t)$  is the Hessian along each output direction. We can find the direction of highest sensitivity via

$$\max_{\|\delta x\| < r} \|\pi_\theta(x_t + \delta x) - \pi_\theta(x_t)\|$$

within some  $\epsilon$ -ball of maximum radius  $r$ . The average active subspace of the inputs over the training state distribution can be found via the importance matrix

$$C = \mathbb{E}_{x_t \sim \mathcal{X}_{\text{train}}} [J_\pi(x_t)^T J_\pi(x_t)].$$

The sensitivity over time can be found by defining the  $k$ -step sensitivity matrix over a horizon  $H$

$$\begin{aligned} S_k(x_t) &= \left. \frac{\partial u_{t+k}}{\partial x} \right|_{x_t} \\ u_{t+k} &= \mathbb{E} \left[ \prod_{i=t}^{t+k-1} \pi_\theta(u_{i+1}|x_{i+1}) T(x_{i+1}|x_i, u_i) \right] \end{aligned}$$

with state transition dynamics  $T(x_{t+1}|x_t, u_t)$ . Then, the empirical observability Gramian over  $H$  can be found via

$$\mathcal{W} = \mathbb{E}_{x_t \sim \mathcal{X}_{\text{train}}} \left[ \sum_{k=0}^H S_k(x_t)^T S_k(x_t) \right].$$

## III. PLAN OF WORK

I will carry out the sensitivity analysis for a legged robot walking on flat ground. I aim to empirically compute the active subspace of the policy inputs to determine which inputs are most important for control. Finding the directions of maximum sensitivity to perturbation will allow for runtime switching to a safe control policy in the event of a large perturbation.