Sensitivity Analysis for Neural Network Controllers

Kyle Morgenstein[1,2]
kjm3887

***Abstract*— Parameterized controllers like neural networks have rapidly become one of the most powerful techniques to control nonlinear systems. However, their black box nature has generally prevented them from taking advantage of control theoretic techniques such as perturbation and sensitivity analysis. In this work we derive a criterion for local exponential stability that does not require knowledge of the closed loop dynamics. We then apply that criterion to a tracking controller trained with reinforcement learning to assess its performance without ever evaluating trajectories from the policy.**

I. INTRODUCTION

Reinforcement learning (RL) has become the standard control technique in robotics due to its robustness, expressivity, 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 exploit the structure of the learning-based controller to assess its sensitivity to anomalous inputs to prevent catastrophic failure at runtime. We then approximate the closed loop dynamics via linearization, and derive a stability criterion for local exponential stability. We demonstrate the approach to assess the stability of a double integrator tracking controller learned via reinforcement learning.

[1]Aerospace Engineering, UT Austin `kylem@utexas.edu`
[2]Apptronik, Inc. `kylemorgenstein@apptronik.com`

II. SYSTEM MODELING

Consider a nonlinear, time varying system

$$\dot{x}_t = f(x_t) + g(x_t)u_t \tag{1}$$

with state $x_t \in \mathbb{R}^n$ and control signal $u_t \in \mathbb{R}^m$. We seek to understand the sensitivity of the closed loop dynamics without assuming knowledge of f or g. Let $u_t = \pi(z_t)$ be the output from a neural network controller tracking a reference signal such that the error dynamics

$$\begin{aligned} z_t &:= x_t - x_t^{\text{ref}} \\ \dot{z}_t &= F(z_t, \pi(z_t)) \end{aligned} \tag{2}$$

has an equilibrium at $F(0, \pi(0)) = 0 \forall t$. The linearized closed loop system Jacobian

$$J_{\text{cl}} = \left.\frac{\partial F}{\partial z_t}\right|_{z_t=0}. \tag{3}$$

III. SYSTEM ANALYSIS AND DESIGN

The closed loop Jacobian can be estimated as follows: For $k = 1, ..., N$ select perturbation direction $d^{(k)} \in \mathbb{R}^n$ with $||d^{(k)}|| = 1$ and radius $h \in \mathbb{R}$. Then define the forward difference

$$y^{(k)} := \frac{F(hd^{(k)}) - F(0)}{h} \tag{4}$$

with estimator $y^{(k)} = J_{\text{cl}} d^{(k)} + \frac{r^{(k)}}{h}$ and remainder $r^{(k)} = \frac{1}{2}(hd^{(k)})^T \mathcal{H}(hd^{(k)}) = \mathcal{O}(h^2)$. Let $Y = [y^{(1)}, ..., y^{(N)}] \in \mathbb{R}^{n \times N}$, $D = [d^{(1)}, ..., d^{(N)}] \in \mathbb{R}^{n \times N}$ and $R = [\frac{r^{(1)}}{h}, ..., \frac{r^{(N)}}{h}] \in \mathbb{R}^{n \times N}$. Then,

$$\begin{aligned} Y &= J_{\text{cl}} D + R \\ \hat{J}_{\text{cl}} &= Y D^T (DD^T)^{-1} \end{aligned} \tag{5}$$

if DD^T is invertible (i.e. $\{d^{(k)}\}$ spans $\mathbb{R}^n$). A valid choice of $\{d^{(k)}\}$ requires $N \geq n$.

To show convergence to the equilibrium, we must show that J_{cl} is Hurwitz. Define the estimation error

$$\Delta J_{\text{cl}} := \hat{J}_{\text{cl}} - J_{\text{cl}}. \tag{6}$$

From the estimator model we have

$$\Delta J_{\text{cl}} = R D^T (DD^T)^{-1} \tag{7}$$

with bound $||\Delta J_{\text{cl}}||_2 \leq ||R||_2 ||D^T(DD^T)^{-1}||_2$. Using the singular value decomposition $D = U\Sigma V^T$, the directional term on the right-hand side can be simplified $D^T(DD^T)^{-1} = V\Sigma^{-1}U^T$, yielding

$$||D^T(DD^T)^{-1}||_2 = \frac{1}{\sigma_{\min}(D)}. \tag{8}$$

Where $\{d^{(k)}\}$ is selected as N i.i.d unit norm isotropic random directions, $DD^T \approx \frac{N}{n}I$ results in $\sigma_{\min}(D) \approx \sqrt{\frac{N}{n}}$ with high probability for large N. Next, by assuming $F \in \mathcal{C}^2$, we can bound the second-order Taylor remainder as $||\frac{r^{(k)}}{h}||_2 \leq \frac{1}{2}L_2 h||d^{(k)}||_2$. The constant L_2 can be estimated via the second directional derivative

$$\begin{aligned} \hat{L}_2 &= \sup_{||d||=1} ||D^2 F(0)[d,d]||_2 \\ &\approx \max_k \left|\left| \frac{F(hd^{(k)}) - 2F(0) + F(-hd^{(k)})}{h^2} \right|\right| \end{aligned} \tag{9}$$

To account for numerical error, a small scale may be used $L_2 = \beta \hat{L}_2$, $\beta > 1$. Then,

$$\begin{aligned} ||R||_2 &\leq \max_k ||\frac{r^{(k)}}{h}||_2 \sqrt{\text{rank}(R)} \\ &\leq \max_k ||\frac{r^{(k)}}{h}||_2 \sqrt{\min(n, N)} \\ &\leq \frac{1}{2}L_2 h\sqrt{n} \end{aligned} \tag{10}$$

using the requirement that $\{d^{(k)}\}$ spans $\mathbb{R}^n$. Substituting bounds into Eq. 7, we find a computable bound on the estimation error

$$||\Delta J_{\text{cl}}||_2 \leq \frac{1}{2}L_2 h \frac{n}{\sqrt{N}} \tag{11}$$

From this upper bound we may now derive the conditions to certify that J_{cl} is Hurwitz given $\hat{J}_{\text{cl}}$. Assume J_{cl} is diagonalizable. Let $\hat{J}_{\text{cl}} = V\Lambda V^{-1}$ with eigenvalues $\{\hat{\lambda}_i\}$ and margin $\hat{\lambda} = -\max_i \operatorname{Re}\hat{\lambda}_i$. The Baur-Fike Theorem [1] gives a bound on the distance between an eigenvalue λ_i of $J_{\text{cl}} = \hat{J}_{\text{cl}} - \Delta J_{\text{cl}}$ and $\hat{\lambda}_i$:

$$\begin{aligned} |\lambda_i - \hat{\lambda}_i| &\leq \kappa(V)||\Delta J_{\text{cl}}||_2 \\ &\leq \kappa(V)\frac{1}{2}L_2 h \frac{n}{\sqrt{N}} \end{aligned} \tag{12}$$

with condition number $\kappa(V) = ||V||_2||V^{-1}||_2$ for the matrix of eigenvectors of $\hat{J}_{\text{cl}}$. Therefore, if

$$\kappa(V)\frac{1}{2}L_2 h \frac{n}{\sqrt{N}} \leq \hat{\lambda} \tag{13}$$

then

$$\begin{aligned} \operatorname{Re}\lambda_i &\leq \operatorname{Re}\hat{\lambda}_i + |\lambda_i - \hat{\lambda}_i| \\ &\leq -\hat{\lambda} + \kappa(V)\frac{1}{2}L_2 h \frac{n}{\sqrt{N}} \\ &< 0 \forall \lambda_i. \end{aligned} \tag{14}$$

Thus, Eq. 13 is sufficient to conclude that J_{cl} is Hurwitz. We have now certified that the closed loop dynamics are locally exponentially stable based on the sampled response within the perturbation radius h for each $t > t_0$. This proof holds despite only assuming J_{cl} is diagonalizable, $\{d^{(k)}\}$ is full rank, and $F \in \mathcal{C}^2$.

IV. SIMULATION RESULTS

The usefulness of the above procedure is demonstrated as follows: Let $\dot{x}_t = [v_t, \pi(z_t)]^T \in \mathbb{R}^4$ be a double integrator in the $X - Y$ plane controlled by an acceleration $a_t = \pi(z_t) \in \mathbb{R}^2$. The policy π_θ is represented as a multivariate Gaussian with mean $\mu_\theta(x_t)$ from a multi-layer perceptron (MLP) and covariance Σ_θ parameterized by $\theta = \left[\{W_i, b_i\}_0^{L-1}, \Sigma\right]$. An MLP is a nonlinear operator composition $g_i(s_i) = (\sigma_i \circ f_i)(s_i)$ of L affine maps $f_i(s_i) = W_i^T s_i + b_i$ and nonlinear differentiable activations $\sigma_i \in C^1$:

$$\begin{aligned} s_0 &= x_t \\ s_{i+1} &= g_i(s_i) = \sigma_i(W_i^T s_i + b_i) \\ \mu_\theta &= g_{L-1} \circ g_{L-2} \circ \cdots \circ g_0 \\ \pi_\theta(u_t|x_t) &= \mathcal{N}(\mu_\theta(x_t), \Sigma_\theta) \end{aligned} \tag{15}$$

After training, we treat the policy as deterministic $u_t = \pi_\theta(x_t) = \mu_\theta(x_t)$ for state observation x_t. We drop the weight parameterization θ for clarity. Without loss of generality we select $x_t^{\text{ref}} = 0 \forall t > t_0$ so $z_t = x_t$. The policy is trained via Proximal Policy Optimization [2] to minimize the cost

$$C(t) = ||x_t||_2^2 + 0.01||\dot{x}_t||_2^2 + 0.001||\ddot{x}_t||_2^2 \tag{16}$$

The policy is trained for 1000 iterations using the `IsaacLab` [3] simulation environment and algorithms adapted from `RSL-RL` [4]. To check for convergence, the value estimate from the critic is evaluated in Fig. 1 at 10,000 points uniformly sampled from the range $[-1.0, 1.0]$. While the critic nearly approximates the symmetric cost function, the level sets clearly show asymmetry and irregularity in the learned value estimate. The sensitivity can be quantified by inspecting the policy sensitivity matrix in Fig. 2.

$$\begin{aligned} J_\pi(x_t) &= \nabla_x u_t \\ \hat{C} &= \mathbb{E}_{x_t \sim \rho(x_0)}\left[J_\pi(x_t)^T J_\pi(x_t)\right] \end{aligned} \tag{17}$$

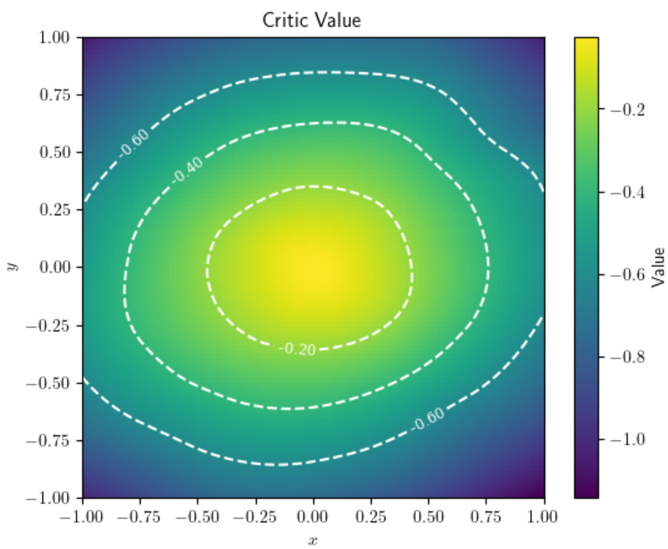

Fig. 1. Value estimates from the trained critic network.

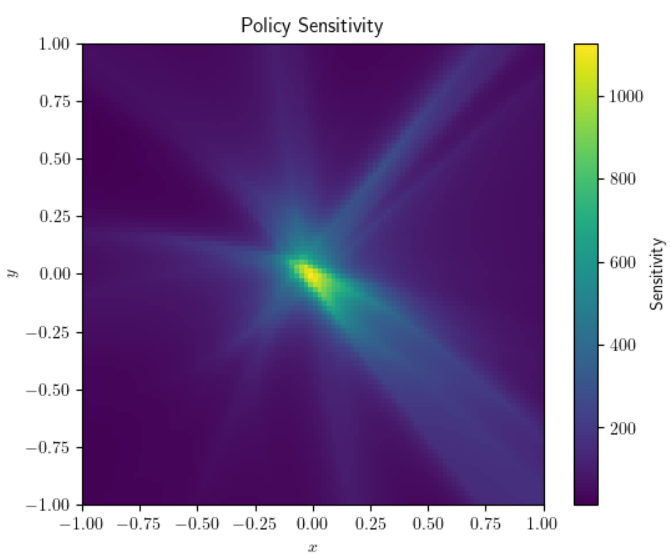

Fig. 2. Policy sensitivity over uniformly sampled states.

given uniform prior density ρ. Given the severe artifacts in the policy sensitivity, it is unlikely that the policy has converged to a locally exponentially stable solution. We compare the synthetic sensitivity to the sensitivity evaluated over real trajectories. We sample 8192 initial configurations in the unit disc and allow the policy to guide the trajectory over 100 steps. The resulting sensitivity is shown in Fig. 3 Following the finite differencing scheme, the eigenvalues of the linearized closed loop dynamics are shown in Fig. 4. The approximation was made over 5000 samples with radius $h = 0.01$. Because the eigenvalues have real part greater than

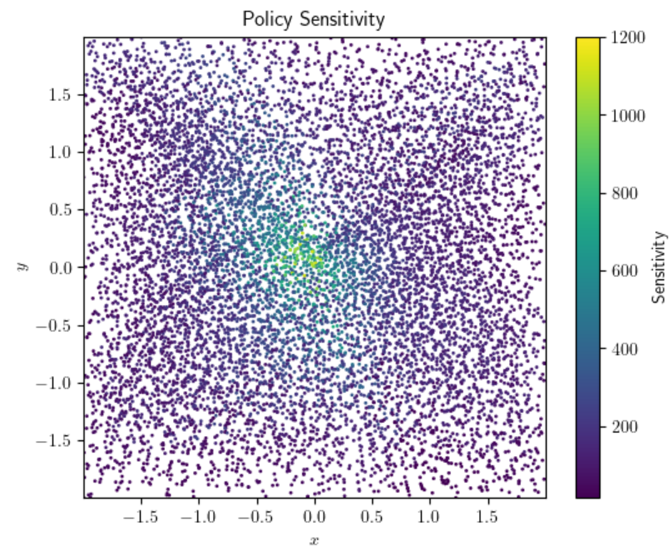

Fig. 3. Policy sensitivity over the policy state visitation distribution.

the derived stability condition, we can conclude that the closed loop dynamics are not locally exponentially stable. With this stability criterion, we can evaluate the behavior of the closed loop dynamics without ever having to run the policy. This is particularly valuable when it may be dangerous to run the policy due to uncertainty or cost.

V. CONCLUSIONS

In this work we derived a computable criterion for local exponential stability for the closed loop dynamics of a system controlled with a neural network. Additionally, we showed how using the Jacobian of the learned policy and critic modules could be used to analyze the performance of the controller without having to evaluate trajectories. In future work, we aim to extend this work to high degree of freedom systems such as a legged robot. When formulated as a tracking controller, the neural network has a natural equilibrium point at the origin at each time step in the reference, which provides a powerful starting point for analysis. To combat the curse of dimensionality, we also seek to use the eigenvectors of the sensitivity matrix as a search heuristic for sampling only the most sensitive directions for the policy. As robots rapidly enter human environments, it is vital that we ensure safety both for the robots and for the people they work with. Videos and additional results can be found at `https://github.com/KyleM73/nonlinear_ct/`.

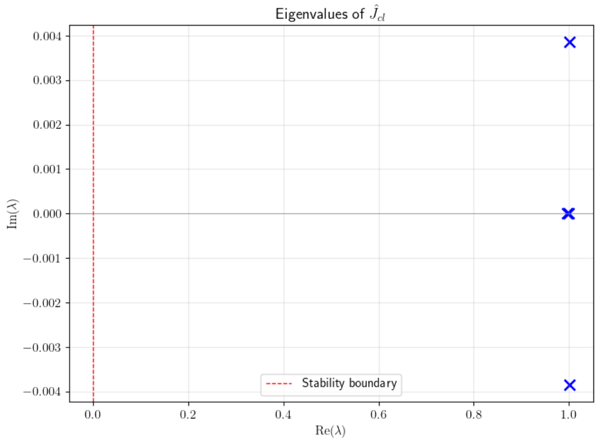

Fig. 4. Eigenvalues of the linearized closed loop dynamics.

References

[1] F. L. Bauer and C. T. Fike, "Norms and exclusion theorems," *Numerische mathematik*, vol. 2, no. 1, pp. 137–141, 1960.

[2] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *CoRR*, vol. abs/1707.06347, 2017. [Online]. Available: http://arxiv.org/abs/1707.06347

[3] M. Mittal, P. Roth, J. Tigue, A. Richard, O. Zhang, P. Du, A. Serrano-Muñoz, X. Yao, R. Zurbrügg, N. Rudin, L. Wawrzyniak, M. Rakhsha, A. Denzler, E. Heiden, A. Borovicka, O. Ahmed, I. Akinola, A. Anwar, M. T. Carlson, J. Y. Feng, A. Garg, R. Gasoto, L. Gulich, Y. Guo, M. Gussert, A. Hansen, M. Kulkarni, C. Li, W. Liu, V. Makoviychuk, G. Malczyk, H. Mazhar, M. Moghani, A. Murali, M. Noseworthy, A. Poddubny, N. Ratliff, W. Rehberg, C. Schwarke, R. Singh, J. L. Smith, B. Tang, R. Thaker, M. Trepte, K. V. Wyk, F. Yu, A. Millane, V. Ramasamy, R. Steiner, S. Subramanian, C. Volk, C. Chen, N. Jawale, A. V. Kuruttukulam, M. A. Lin, A. Mandlekar, K. Patzwaldt, J. Welsh, H. Zhao, F. Anes, J.-F. Lafleche, N. Moënne-Loccoz, S. Park, R. Stepinski, D. V. Gelder, C. Amevor, J. Carius, J. Chang, A. H. Chen, P. de Heras Ciechomski, G. Daviet, M. Mohajerani, J. von Muralt, V. Reutskyy, M. Sauter, S. Schirm, E. L. Shi, P. Terdiman, K. Vilella, T. Widmer, G. Yeoman, T. Chen, S. Grizan, C. Li, L. Li, C. Smith, R. Wiltz, K. Alexis, Y. Chang, D. Chu, L. J. Fan, F. Farshidian, A. Handa, S. Huang, M. Hutter, Y. Narang, S. Pouya, S. Sheng, Y. Zhu, M. Macklin, A. Moravanszky, P. Reist, Y. Guo, D. Hoeller, and G. State, "Isaac lab: A gpu-accelerated simulation framework for multi-modal robot learning," *arXiv preprint arXiv:2511.04831*, 2025. [Online]. Available: https://arxiv.org/abs/2511.04831

[4] C. Schwarke, M. Mittal, N. Rudin, D. Hoeller, and M. Hutter, "Rsl-rl: A learning library for robotics research," *arXiv preprint arXiv:2509.10771*, 2025.

[5] R. Hermann and A. Krener, "Nonlinear controllability and observability," *IEEE Transactions on Automatic Control*, vol. 22, no. 5, pp. 728–740, 1977.