

Learning Over Contracting and Lipschitz Closed-Loops for Partially-Observed Nonlinear Systems

Nicholas H. Barbara, Ruigang Wang, and Ian R. Manchester

Abstract—This paper presents a policy parameterization for learning-based control on nonlinear, partially-observed dynamical systems. The parameterization is based on a nonlinear version of the Youla parameterization and the recently proposed *Recurrent Equilibrium Network* (REN) class of models. We prove that the resulting Youla-REN parameterization automatically satisfies stability (contraction) and user-tunable robustness (Lipschitz) conditions on the closed-loop system. This means it can be used for safe learning-based control with no additional constraints or projections required to enforce stability or robustness. We test the new policy class in simulation on two reinforcement learning tasks: 1) magnetic suspension, and 2) inverting a rotary-arm pendulum. We find that the Youla-REN performs similarly to existing learning-based and optimal control methods while also ensuring stability and exhibiting improved robustness to adversarial disturbances.

I. INTRODUCTION

Deep reinforcement learning (RL) has been a driving force behind many recent successes in learning-based control, with applications ranging from discrete game-like problems [1], [2] to robotic locomotion [3]. As its popularity continues to grow, there is increasing need for a learning framework that offers the stability and robustness guarantees of classical control methods while still being fast and flexible for learning in complex environments.

One promising idea is to directly learn over a set of robustly stabilizing controllers for a given dynamical system. RL policies are then guaranteed to naturally satisfy robustness and stability requirements even during training. Parameterizing the space of all such controllers is well-studied for linear systems [4]. In fact, learning over this space results in policies that perform better and are more robust than those learned from typical RL frameworks which do not consider stability [5]. While there has been considerable work extending this parameterization to nonlinear systems with full state knowledge [6], [7] or specific structures [8], [9], the problem of general *partially-observed* nonlinear systems (full state information unavailable) is more challenging. In this paper, we present a parameterization of robust stabilizing controllers for partially-observed nonlinear systems that can be readily applied to learning-based control.

A. Previous work on linear systems

We recently proposed the Youla-REN policy class for learning over all stabilizing controllers for partially-observed

linear systems [10], [11]. It combines the classical Youla-Kucera parameterization with the *Recurrent Equilibrium Network* (REN) model architecture [12]. The Youla parameterization is an established idea in linear control theory that represents all stabilizing linear controllers for a given linear system [13]. One common construction augments an existing stabilizing “base” controller with a stable linear parameter Q , which is optimized to improve the closed-loop system performance [14], [15]. We extended this idea in [11], showing that if Q is a contracting and Lipschitz *nonlinear* system, the Youla parameterization represents all stabilizing *nonlinear* controllers for a given linear system.

A key feature of our work in [11] was using RENs for the Youla parameter Q . The *direct parameterization* presented in [12] allowed us to construct RENs that universally approximate all contracting and Lipschitz systems. This meant that we could use unconstrained optimization to train the Youla-REN. Training RENs in this way is less conservative than weight-rescaling methods such as [16], and significantly faster than solving large semi-definite programs during training [17] or projected gradient descent methods such as [18], [19]. Our aim is to retain this computational efficiency when extending our framework to learning stabilizing controllers for nonlinear systems.

B. Youla parameterization for nonlinear systems

Significant theoretical advances were made in the 1980s-90s to extend the Youla parameterization to partially-observed nonlinear systems. Early work by [20], [21] addressed the problem by using left coprime factorizations. However, state-space models of left coprime factors are limited to nonlinear systems with specific structures [8]. A more extensive framework was proposed by [22], [23], [24], [25] using kernel representations to parameterize all stabilizing controllers for nonlinear systems based on input-to-state stability. Despite providing a rather general result, kernel representations are often not intuitive to work with in practical learning-based control. Moreover, the focus of these works was on stabilizing a particular equilibrium state. Many applications require control systems that track a wide range of reference trajectories. This motivates the need for a framework with a sufficiently strong and flexible notion of stability that is also intuitive to implement in practice.

C. Contributions

In this paper, we extend the Youla-REN policy class proposed in [10], [11] to nonlinear systems and address the open questions outlined in Section I. In particular, we:

*This work was supported in part by the Australian Research Council, and the NSW Defence Innovation Network.

The authors are with the Australian Centre for Robotics (ACFR), and the School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, Sydney, NSW 2006, Australia nicholas.barbara@sydney.edu.au

- 1) Parameterize robust (Lipschitz) stabilizing feedback controllers for *partially-observed* nonlinear systems based on a strong stability notion — contraction.
- 2) Demonstrate that the Youla-REN can be applied to learning-based control by investigating its performance on two simulated RL tasks: 1) magnetic suspension and 2) inverting a rotary-arm pendulum.
- 3) Show through simulation that we can intuitively tune the trade-off between performance and robustness of the learned policy via the direct parameterization of RENs.

An extended version of this work (including all proofs and training details) can be found in [26].

D. Notation

Consider the set of sequences $\ell_{2e}^n = \{x \mid x : \mathbb{N} \rightarrow \mathbb{R}^n\}$, where n is omitted if it can be inferred from the context. For any $x \in \ell_{2e}^n$, write $x_t \in \mathbb{R}^n$ for the value of the sequence at time $t \in \mathbb{N}$. We define the subset $\ell_2 \subset \ell_{2e}^n$ as the set of all square-summable sequences such that $x \in \ell_2 \Leftrightarrow \|x\| := \sqrt{\sum_{t=0}^{\infty} |x_t|^2}$ is finite, where $|\cdot|$ denotes the Euclidean norm. We also define the norm of the truncation of $x \in \ell_2$ over $[0, T]$ as $\|x\|_T := \sqrt{\sum_{t=0}^T |x_t|^2}$ for all $T \in \mathbb{N}$ such that $x \in \ell_{2e}^n \iff \|x\|_T$ is finite. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is Lipschitz continuous if there exist a constant $\gamma \in \mathbb{R}^+$ such that $|f(a) - f(b)| \leq \gamma|a - b|, \forall a, b \in \mathbb{R}^n$.

E. Definitions

The main results in this paper concern the analysis of contracting (“stable”) and Lipschitz (“robust”) dynamical systems. Consider a system

$$\mathcal{T} : \begin{cases} x_{t+1} = f(x_t, u_t) \\ y_t = h(x_t, u_t) \end{cases} \quad (1)$$

with state $x_t \in \mathbb{R}^n$, inputs $u_t \in \mathbb{R}^m$, and outputs $y_t \in \mathbb{R}^p$.

Definition 1 (Contraction): \mathcal{T} is contracting with respect to a uniformly-bounded metric $V : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+$ if for any fixed input sequence $u \in \ell_{2e}^m$, the system satisfies

$$c_1 |x_t^1 - x_t^2|^2 \leq V(x_t^1, x_t^2) \leq c_2 |x_t^1 - x_t^2|^2 \quad (2)$$

$$V(x_{t+1}^1, x_{t+1}^2) \leq \alpha V(x_t^1, x_t^2) \quad (3)$$

where $c_2 \geq c_1 > 0$ and $\alpha \in [0, 1)$ is the contraction rate.

Intuitively, a contracting system is one that exponentially forgets its initial conditions. We introduce a slightly weaker notion of contraction based only on exponential convergence.

Definition 2 (Contraction with transients): \mathcal{T} is contracting with transients at rate $\alpha \in [0, 1)$ if for any two initial conditions $x_0^1, x_0^2 \in \mathbb{R}^n$ and for any fixed input sequence $u \in \ell_{2e}^m$, there exists $\beta(x_0^1, x_0^2) \geq 0$ such that

$$|x_t^1 - x_t^2| \leq \beta(x_0^1, x_0^2) \cdot \alpha^t \quad \forall t \in \mathbb{N}. \quad (4)$$

Note that the overshoot β is a function of initial conditions.

Definition 3 (Lipschitz with transients): A system \mathcal{T} is Lipschitz with transients (referred to simply as *Lipschitz* in this paper) if for any two input sequences $u^1, u^2 \in \ell_{2e}^m$ and initial conditions $x_0^1, x_0^2 \in \mathbb{R}^n$ we have

$$\|y^1 - y^2\|_T \leq \gamma \|u^1 - u^2\|_T + \kappa(x_0^1, x_0^2) \quad \forall T \in \mathbb{N} \quad (5)$$

where $\gamma \in \mathbb{R}^+$ is the Lipschitz constant and $\kappa(x_0^1, x_0^2) \geq 0$.

A system with a smaller Lipschitz bound is more robust to sudden changes in its inputs. Adding the condition $\kappa(a, a) = 0$ for any $a \in \mathbb{R}^n$ would recover the definition for a system with an incremental ℓ^2 gain bound of γ from [12], [27].

II. PROBLEM STATEMENT

Consider a nonlinear dynamical system \mathcal{G} described by

$$\mathcal{G} : \begin{cases} x_{t+1} = f(x_t, u_t + r_t) \\ y_t = c(x_t) \end{cases} \quad (6)$$

with internal states $x_t \in \mathbb{R}^n$, controlled inputs $u_t \in \mathbb{R}^m$, and measured outputs $y_t \in \mathbb{R}^p$. The control signal is perturbed by a known, bounded input $r_t \in \mathbb{R}^m$ (e.g., a reference signal). We denote the total inputs as $\tilde{u}_t = u_t + r_t$. The exogenous inputs and controlled outputs are r_t and $z_t = (x_t^\top, u_t^\top)^\top$, respectively.

Our aim is to learn feedback controllers of the form $u = \mathcal{K}_\theta(y)$ where θ is a learnable parameter. Controllers may be nonlinear and dynamical systems themselves. The closed-loop system of \mathcal{G} and \mathcal{K}_θ should satisfy the following stability, robustness, and performance criteria (respectively):

- 1) The closed-loop system is contracting (with transients) such that initial conditions are forgotten exponentially.
- 2) The closed-loop response to external inputs (the map $r \mapsto z$) is Lipschitz.
- 3) The controller \mathcal{K}_θ at least locally and approximately minimizes a cost function of the form

$$J_\theta = E \left[\sum_{t=0}^{T-1} g(x_t, u_t) + g_T(x_T) \right] \quad (7)$$

where the expectation is over x_0 and external inputs.

III. A NONLINEAR YOULA PARAMETERIZATION

A. The Youla architecture

Suppose \mathcal{G} is in feedback with a *base controller* \mathcal{K}_b consisting of an observer and state-feedback controller:

$$\mathcal{K}_b : \begin{cases} \hat{x}_{t+1} = f_o(\hat{x}_t, u_t + r_t, y_t) \\ \tilde{u}_t = k(\hat{x}_t) \end{cases} \quad (8)$$

where $\hat{x} = \mathcal{O}(\tilde{u}, y)$ is the observer, and its state vector $\hat{x}_t \in \mathbb{R}^n$ is the estimated value of the true state x_t . Denote the predicted measurements as $\hat{y}_t = c(\hat{x}_t)$ where $\hat{y} \in \mathbb{R}^p$. Our proposed controller parameterization (Fig. 1) augments the base controller with a (possibly nonlinear) system $\mathcal{Q} : (r, \tilde{y}) \mapsto \tilde{u}$, where $\tilde{y}_t = y_t - \hat{y}_t$ are the innovations. Specifically, the augmented controller is

$$\mathcal{K}_\mathcal{Q} : \begin{cases} \hat{x}_{t+1} = f_o(\hat{x}_t, u_t + r_t, y_t) \\ \tilde{u}_t = k(\hat{x}_t) + \tilde{u}_t \end{cases} \quad (9)$$

with

$$\mathcal{Q} : \begin{cases} q_{t+1} = f_q(q_t, r_t, \tilde{y}_t) \\ \tilde{u}_t = h_q(q_t, r_t, \tilde{y}_t), \end{cases} \quad q_t \in \mathbb{R}^q. \quad (10)$$

Here $\mathcal{K}_\mathcal{Q}$ is a nonlinear version of the Youla-Kucera parameterization, where \mathcal{Q} is the Youla parameter.

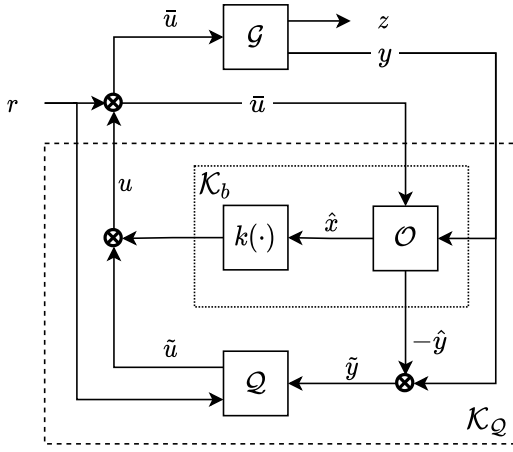


Fig. 1: A version of the Youla-Kucera parameterization, where the Youla parameter \mathcal{Q} augments an observer-based feedback controller \mathcal{K}_b .

B. Assumptions

We make the following assumptions, drawing inspiration from [24], [28]:

- A1) Robustly stabilizing base controller:** The closed-loop system $r \mapsto z$ composed of \mathcal{G} in feedback with \mathcal{K}_b is contracting and Lipschitz.
- A2) Observer correctness:** Given $\hat{x}_0 = x_0$, the observer can exactly replicate the plant dynamics. That is, $f(x_t, \bar{u}_t) = f_o(x_t, \bar{u}_t, c(x_t)) \forall t \in \mathbb{N}$.
- A3) Contracting & Lipschitz observer:** The observer \mathcal{O} is a contracting and Lipschitz system.
- A4) Lipschitz maps:** All functions are Lipschitz continuous.

C. Theoretical results

Our first main result is that augmenting a robustly stabilizing controller with a contracting and Lipschitz \mathcal{Q} ensures the closed-loop system will also be contracting and Lipschitz. This allows optimization of the closed-loop response via \mathcal{Q} while maintaining stability and robustness guarantees.

Theorem 1: The closed-loop system of \mathcal{G} in (6) and \mathcal{K}_Q in (8) – (10) is contracting *with transients* and Lipschitz from $r \mapsto z$ if the Youla parameter \mathcal{Q} is contracting and Lipschitz.

The proof is provided in [26] and can be summarized as follows. The observer error $\tilde{x}_t := x_t - \hat{x}_t$ exponentially converges to zero since the plant trajectory x is a particular solution of the observer, which is a contracting and Lipschitz system. This occurs independently of r_t and \tilde{u}_t , breaking the feedback loop through \mathcal{Q} . To prove contraction with transients, we show that the states of a contracting system exponentially converge given exponentially converging inputs, and apply this to \mathcal{Q} and the closed-loop system under \mathcal{K}_b (“base system”). We repeatedly apply Lipschitz properties of \mathcal{Q} and the base system to prove that the closed-loop system under \mathcal{K}_Q is also Lipschitz, completing the proof.

One interesting question is whether a contracting and Lipschitz closed loop can always be parameterized by a

contracting and Lipschitz \mathcal{Q} . Our second main result shows that this is true under additional assumptions. We consider a perturbed nonlinear system

$$\mathcal{G}_d : \begin{cases} x_{t+1} = f(x_t, u_t + r_t) + d_{x_t} \\ y_t = c(x_t) + d_{y_t}. \end{cases} \quad (11)$$

where $d_{x_t} \in \mathbb{R}^n$ and $d_{y_t} \in \mathbb{R}^p$ are additive process and measurement disturbances, respectively. The following is a stronger version of assumption **A1**.

- A5) Robustness to disturbances:** The closed-loop system $(r, d_x, d_y) \mapsto z$ composed of \mathcal{G}_d in feedback with \mathcal{K}_b is contracting and Lipschitz.

Theorem 2: Suppose that assumptions **A2** - **A5** hold. Then, any controller \mathcal{K} forming a contracting and Lipschitz closed-loop map $(r, d_x, d_y) \mapsto z$ with \mathcal{G}_d in (11) can be represented by (8) to (10) with contracting and Lipschitz \mathcal{Q} .

The proof follows by augmenting the robustly stabilizing controller \mathcal{K} with an observer \mathcal{O} to form a map $\mathcal{Q}_{\mathcal{K}} : (r, \tilde{y}) \mapsto \tilde{u}$, which is contracting and Lipschitz by comparison with the closed-loop system under \mathcal{K} (see [26]).

Remark 1: It is worth noting that the closed-loop system is not guaranteed to be contracting and Lipschitz under bounded but unknown additive disturbances (see Example 1). This is because the observer error does not converge to zero in the presence of unknown disturbances, which is the key property required in Theorem 1. However, the closed-loop states will still be bounded under bounded additive disturbances [29], [30].

Example 1: Consider the following scalar system

$$\mathcal{G} : x_{t+1} = 0.5 \sin(x_t) + \tilde{u}_t + d_t$$

$$\mathcal{K}_Q : \begin{cases} \hat{x}_{t+1} = 0.5 \sin(\hat{x}_t) + \tilde{u}_t \\ \tilde{u}_t = 10\tilde{y} = 10(x - \hat{x}_t) \end{cases}$$

which is contracting for the case $d_t \equiv 0$. When $d_t \equiv 1$, the system converges to multiple solutions, hence the closed-loop system is not contracting under non-zero disturbances.

IV. NUMERICAL EXPERIMENTS

We now examine the performance of the Youla-REN policy class on two RL problems: 1) magnetic suspension, and 2) inverting a rotary-arm pendulum. Each system is nonlinear and partially-observed, with different base controller designs to test the policy class under different architectures. We compare performance and robustness of three policy types:

- 1) *Youla-REN*: uses a contracting REN for the Youla parameter \mathcal{Q} (see (9) and (10)).
- 2) *Youla- γ REN*: uses a REN with a Lipschitz upper bound of γ (where $\gamma \rightarrow \infty$ recovers the contracting REN).
- 3) *Feedback-LSTM*: an LSTM network [31] augmenting the base controller via direct feedback of the measurement output, with $\tilde{u} = \mathcal{F}(y)$ for an LSTM system \mathcal{F} .

Since RENs are universal approximators of contracting and Lipschitz systems [11, Prop. 2], then by Theorem 1 the Youla- (γ) REN parameterizes a set of contracting and Lipschitz closed-loops for partially-observed nonlinear systems.

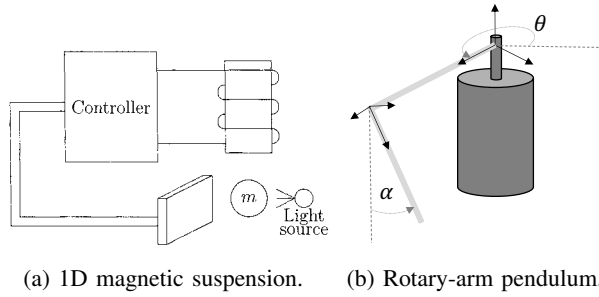


Fig. 2: The two RL environments examined in Sec. IV: (a) a magnetic ball moving vertically under the influence of an electromagnet, and (b) a rotary-arm pendulum driven by a single motor on the rotating arm. The light source in (a) measures the ball position. Fig. 2a from [33].

The Feedback-LSTM form is commonly used in deep RL (e.g: [3]) but provides no such stability or robustness guarantees. For more detail on contracting and Lipschitz parameterizations of RENs, see [12]. Our experiments were written in Julia using `RobustEquilibriumNetworks.jl` [32] and are available on [github](https://github.com/nic-barbara/CDC2023-YoulaREN)¹.

A. Problem setup

1) *Learning objective*: Let $z_t = p(x_t)$ be a performance variable to be tracked for some function p . We formulated the RL tasks as minimizing a quadratic cost on the differences $\Delta z_t = z_t - z_{\text{ref}}$, $\Delta u_t = u_t - u_{\text{ref}}$ between performance variables and controls, and their desired reference values (respectively). That is,

$$\min. E[J] \text{ s.t. } J(x_0) = \sum_{t=0}^{T-1} (|\Delta z_t|_Q^2 + |\Delta u_t|_R^2) \quad (12)$$

where the cost function J is weighted by matrices Q and R . The expectation is over all possible initial conditions and random disturbances. We used $T = 100$ time samples.

2) *Magnetic suspension*: Consider the one-dimensional magnetic suspension system presented in [33], illustrated in Fig. 2a. The system has three states (ball position, velocity, and coil current) and one input (coil voltage). Only the ball position and coil current are measured. We used the same nonlinear system model as in Exercise 13.27 of [33]. We added random noise to all states and measurements with standard deviations 5×10^{-4} and 10^{-3} (respectively).

The objective was to stabilize the ball at a height of 5 cm with minimal control effort. Our base controller consisted of a high-gain observer ([33, Sec. 14.5.2]) and a state-feedback controller designed with the backstepping and Lyapunov re-design methods outlined in [33, Sec. 14.2-14.3]. We encoded the learning objective in (12) with $z_t = x_t$, $Q = \text{diag}(1/0.025^2, 0, 0)$, $R = 1/50^2$.

3) *Rotary-arm pendulum*: Next we considered the rotary-arm pendulum system in Fig. 2b. The system has four states (rod angles and angular velocities) and one control input

(motor voltage). Only the angles are measured. The system dynamics are presented in [34, Eqn. 31]. We added noise with standard deviation 10^{-2} to all states and measurements.

The control objective was to stabilize the pendulum in its (unstable) upright equilibrium, again with minimal control effort. We designed a state-feedback policy consisting of an energy-pumping controller to swing the pendulum arm upwards ([35, Eqn. 8]) and a static linear quadratic regulator to balance the pendulum within 30° of the vertical. We completed the base controller with a high-gain observer. The learning objective was defined as per (12) with $z_t = (\cos \theta_t, \sin \theta_t, \cos \alpha_t, \sin \alpha_t)^\top$ for the arm and pendulum angles θ_t, α_t (respectively), and $Q = \text{diag}(5, 5, 10, 10)$, $R = 0.01$. For small deviations from vertical, this is approximately a quadratic cost on $\Delta \theta_t, \Delta \alpha_t$.

B. Results and discussion

1) *Learning performance*: Fig. 3 shows the mean test cost for each policy trained on the magnetic suspension and rotary-arm pendulum RL tasks. Policies were benchmarked against nonlinear model predictive controllers (NMPC), which are easy to design for low-dimensional problems. All of the learned policies show significant performance improvements over the base controller, and the best performing models reach the NMPC benchmarks. In particular, we see comparable performance between the Youla-REN and the Feedback-LSTM policy classes, with the Youla-REN achieving a lower cost on magnetic suspension, and the Feedback-LSTM performing slightly better on the rotary-arm pendulum. Together with the results of Sec. III, we therefore have a policy class that can perform as well as existing state-of-the-art methods on RL tasks for partially-observed nonlinear systems, while also providing stability guarantees for every policy trialled during training. Note that we have not compared the Youla-REN to an LSTM or REN in direct feedback without a base controller (the typical RL policy architecture) in Fig. 3. Training models in this configuration took significantly more epochs than the Youla and Feedback architectures, and achieved a worse final cost than even the base controllers on both tasks.

2) *Robustness*: One of the great advantages of the Youla-REN policy class is that we can control the performance-robustness trade-off by imposing a Lipschitz upper bound on the REN. The left panels in Figs. 4a and 4b show the effect of perturbing each trained policy with additive adversarial attacks of increasing size on the measurement signal y_t . The scatter plots to the right show the attack size required to meaningfully perturb each closed-loop system. We define a ‘‘critical’’ attack as one that shifts 1) the average ball position more than 1 cm from the target and 2) the average pendulum angle more than 30° from the vertical in the magnetic suspension and rotary-arm pendulum tasks, respectively. Adversarial attacks were computed with minibatch gradient descent over a receding horizon of 10 time samples.

Comparing Figs. 3 and 4 demonstrates the effect of the Lipschitz bound on the performance-robustness trade-off.

¹<https://github.com/nic-barbara/CDC2023-YoulaREN>

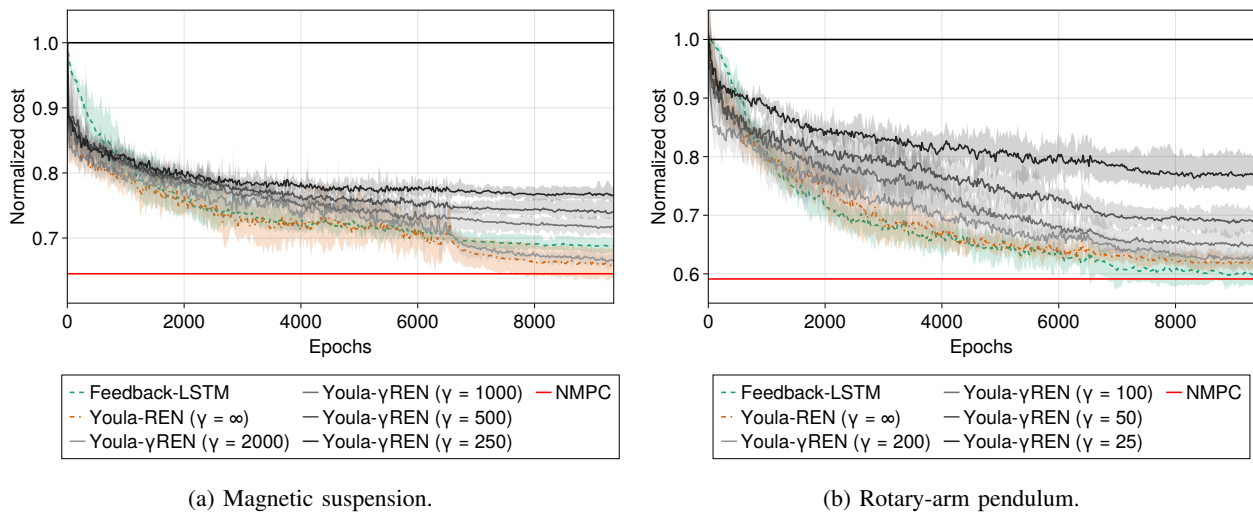


Fig. 3: Loss curves showing the mean test cost vs. training epochs for policies trained on (a) the magnetic suspension problem and (b) the rotary-arm pendulum problem. Colored bands show the range over the six random seeds for each model, lines show the mean. Test cost is normalized by the base controller cost, indicated by the solid black line at 1.0.

In Fig. 3, imposing a stronger Lipschitz bound (smaller γ) drives the Youla- γ REN policies to worse final costs. In Fig. 4, however, stronger Lipschitz bound can reliably lead to policies which are more robust to adversarial attacks, even if they perform worse in the unperturbed case. In particular, Fig. 4b shows that the base controller and Feedback-LSTM policies are highly sensitive to adversarial attacks in the rotary-arm pendulum environment. We suspect this is because the system has enough degrees of freedom to exhibit chaotic motion, and can be driven to extremely unstable closed-loop responses that are more difficult to recover from than in the magnetic suspension environment. Note that the relationship is not linear — imposing too strong a Lipschitz bound can lead to less robust policies (for example $\gamma = 25$ in Fig. 4b). In practice, careful tuning of the imposed Lipschitz upper bound is required for a given problem.

3) *Key results:* These results emphasize the strength of the Youla-REN policy class in learning-based control tasks. We can take an existing stabilizing controller for a (nonlinear) dynamical system and learn a robust stabilizing feedback controller that improves some user-defined performance metric. Moreover, we can search over a space of contracting and Lipschitz closed-loops, guaranteeing stability at every step of the training process, and still achieve similar performance to existing methods which provide no such guarantees. Our approach is intuitive in that we can balance the performance-robustness trade-off by tuning the Lipschitz bound of the REN. It is versatile since we do not require special solution methods or projections during training. We therefore expect the Youla-REN to be well-suited to learning-based control in safety-critical robotic systems where performance and robustness are crucial to successful operation.

REFERENCES

- [1] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovsky,
- S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, pp. 529–533, 2 2015.
- [2] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. V. D. Driessche, T. Graepel, and D. Hassabis, “Mastering the game of go without human knowledge,” *Nature*, vol. 550, pp. 354–359, 10 2017.
- [3] J. Siekmann, Y. Godse, A. Fern, and J. Hurst, “Sim-to-real learning of all common bipedal gaits via periodic reward composition,” 11 2021.
- [4] B. D. Anderson, “From youla-kucera to identification, adaptive and nonlinear control,” *Automatica*, vol. 34, pp. 1485–1506, 12 1998.
- [5] J. W. Roberts, I. R. Manchester, and R. Tedrake, “Feedback controller parameterizations for reinforcement learning,” pp. 310–317, 2011.
- [6] J. I. Imura and T. Yoshikawa, “Parametrization of all stabilizing controllers of nonlinear systems,” *Systems & Control Letters*, vol. 29, pp. 207–213, 1 1997.
- [7] L. Furieri, C. L. Galimberti, and G. Ferrari-Trecate, “Neural system level synthesis: Learning over all stabilizing policies for nonlinear systems,” pp. 2765–2770, 1 2023.
- [8] J. B. Moore and L. Irlicht, “Copriime factorization over a class of nonlinear systems,” *Proceedings of the American Control Conference*, vol. 4, pp. 3071–3075, 1992.
- [9] F. Blanchini, D. Casagrande, and S. Miani, “Parametrization of all stabilizing compensators for absorbable nonlinear systems,” *Proceedings of the IEEE Conference on Decision and Control*, pp. 5943–5948, 2010.
- [10] R. Wang and I. R. Manchester, “Youla-ren: Learning nonlinear feedback policies with robust stability guarantees,” *Proceedings of the American Control Conference*, vol. 2022-June, pp. 2116–2123, 2022.
- [11] R. Wang, N. H. Barbara, M. Revay, and I. R. Manchester, “Learning over all stabilizing nonlinear controllers for a partially-observed linear system,” *IEEE Control Systems Letters*, pp. 1–1, 2022.
- [12] M. Revay, R. Wang, and I. R. Manchester, “Recurrent equilibrium networks: Flexible dynamic models with guaranteed stability and robustness,” *IEEE Transactions on Automatic Control*, pp. 1–16, 2023.
- [13] D. C. Youla, J. J. Bongiorno, and H. A. Jabr, “Modern wiener-hopf design of optimal controllers — part ii: The multivariable case,” *IEEE Transactions on Automatic Control*, vol. 21, pp. 319–338, 1976.
- [14] F. Li, S. Yuan, F. Qian, Z. Wu, H. Pu, M. Wang, J. Ding, and Y. Sun, “Adaptive deterministic vibration control of a piezo-actuated active-passive isolation structure,” *Applied Sciences 2021, Vol. 11, Page 3338*, vol. 11, p. 3338, 4 2021.
- [15] I. Mahtout, F. Navas, V. Milanese, and F. Nashashibi, “Advances in youla-kucera parametrization: A review,” *Annual Reviews in Control*, vol. 49, pp. 81–94, 1 2020.

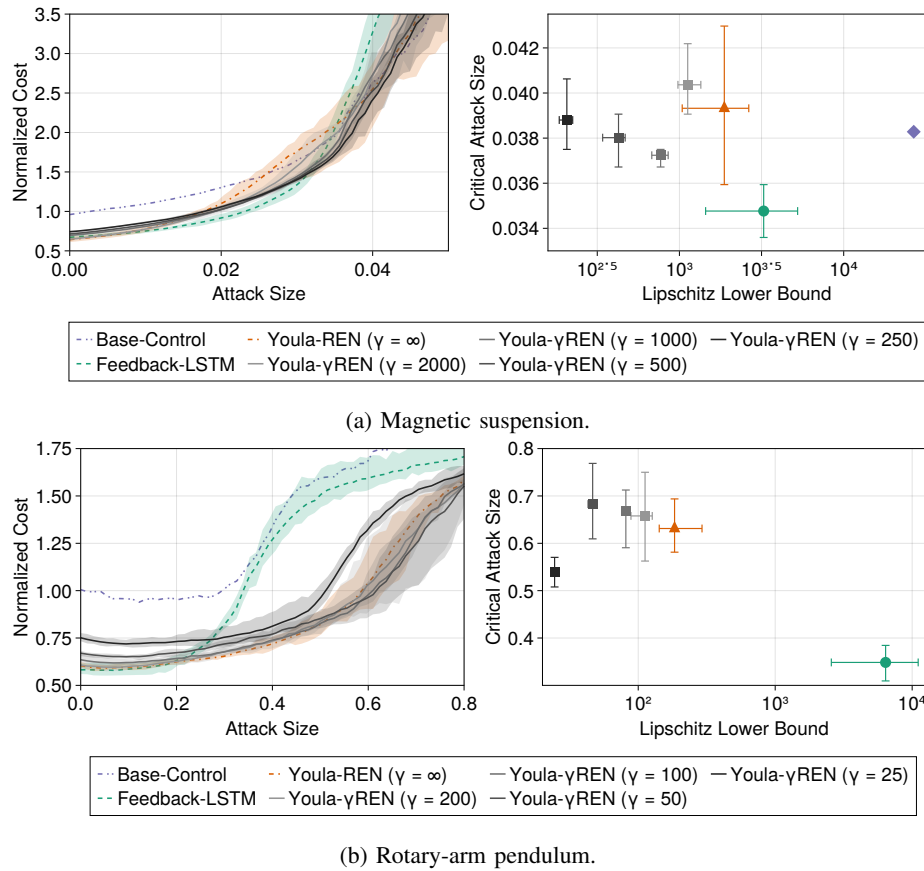


Fig. 4: Normalized test cost vs. adversarial attack size for each of the trained models in Fig. 3. Lines show the average cost over the six random seeds for each model, with bands showing the range. Test cost is normalized by the cost of the (unperturbed) base controller. Scatter plots show the “critical” attack size required to perturb (a) the mean ball position 1 cm from its target and (b) the mean pendulum angle to 30° from the vertical in each task, respectively. Error bars are for the six different random seeds for each model. The estimated Lipschitz lower bound of the base controller in (b) is extremely large and has been omitted from the scatter plot. Values of γ in the legend are Lipschitz upper bounds.

- [16] J. Miller and M. Hardt, “Stable recurrent models,” *7th International Conference on Learning Representations, ICLR 2019*, 5 2018.
- [17] P. Pauli, A. Koch, J. Berberich, P. Kohler, and F. Allgower, “Training robust neural networks using lipschitz bounds,” *IEEE Control Systems Letters*, vol. 6, pp. 121–126, 2022.
- [18] J. N. Knight and C. Anderson, “Stable reinforcement learning with recurrent neural networks,” *Journal of Control Theory and Applications*, vol. 9, pp. 410–420, 8 2011.
- [19] N. Junnarkar, H. Yin, F. Gu, M. Arcak, and P. Seiler, “Synthesis of stabilizing recurrent equilibrium network controllers,” pp. 7449–7454, 1 2023.
- [20] A. D. Paice and J. B. Moore, “On the youla-kucera parametrization for nonlinear systems,” *Systems & Control Letters*, vol. 14, pp. 121–129, 2 1990.
- [21] G. Chen and R. J. de Figueiredo, “Construction of the left coprime fractional representation for a class of nonlinear control systems,” *Systems & Control Letters*, vol. 14, pp. 353–361, 4 1990.
- [22] A. D. Paice and A. J. van der Schaft, “Stable kernel representations as nonlinear left coprime factorizations,” *Proceedings of the IEEE Conference on Decision and Control*, vol. 3, pp. 2786–2791, 1994.
- [23] A. D. Paice and A. J. V. D. Schaft, “The class of stabilizing nonlinear plant controller pairs,” *IEEE Transactions on Automatic Control*, vol. 41, pp. 634–645, 1996.
- [24] K. Fujimoto and T. Sugie, “Characterization of all nonlinear stabilizing controllers via observer-based kernel representations,” *Automatica*, vol. 36, pp. 1123–1135, 8 2000.
- [25] K. Fujimoto and T. Sugie, “State-space characterization of youla parametrization for nonlinear systems based on input-to-state stability,” *Proceedings of the IEEE Conference on Decision and Control*, vol. 3, pp. 2479–2484, 1998.
- [26] N. H. Barbara, R. Wang, and I. R. Manchester, “Learning over contracting and lipschitz closed-loops for partially-observed nonlinear systems (extended version),” *arXiv preprint arXiv:2304.06193*, 2023.
- [27] M. Revay, R. Wang, and I. R. Manchester, “A convex parameterization of robust recurrent neural networks,” *IEEE Control Systems Letters*, vol. 5, pp. 1363–1368, 10 2021.
- [28] B. Yi, R. Wang, and I. R. Manchester, “Reduced-order nonlinear observers via contraction analysis and convex optimization,” *IEEE Transactions on Automatic Control*, vol. 67, pp. 4045–4060, 8 2022.
- [29] W. Lohmiller and J. J. E. Slotine, “On contraction analysis for nonlinear systems,” *Automatica*, vol. 34, pp. 683–696, 6 1998.
- [30] F. Bullo, *Contraction Theory for Dynamical Systems*. Kindle Direct Publishing, 1.0 ed., 2022.
- [31] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.
- [32] N. H. Barbara, M. Revay, R. Wang, J. Cheng, and I. R. Manchester, “Robustneuralnetworks.jl: a package for machine learning and data-driven control with certified robustness,” *arXiv preprint arXiv:2306.12612*, 6 2023.
- [33] H. K. Khalil, *Nonlinear systems; 3rd ed.* Prentice-Hall, 2002. The book can be consulted by contacting: PH-AID: Wallet, Lionel.
- [34] B. S. Cazzolato and Z. Prime, “On the dynamics of the furuta pendulum,” *Journal of Control Science and Engineering*, vol. 2011, 2011.
- [35] K. J. Åström and K. Furuta, “Swinging up a pendulum by energy control,” *Automatica*, vol. 36, pp. 287–295, 2 2000.