

A. Appendix

A.1. Value function hessian

$$\begin{aligned}
 \frac{d^2 V^{\bar{a}}(s_0)}{d\bar{a}^2} &= \frac{d}{d\bar{a}} \left[\frac{\partial V^{\bar{a}}}{\partial \bar{a}} + \frac{\partial V^{\bar{a}}}{\partial \mathbf{s}} \frac{d\mathbf{s}}{d\bar{a}} \right], \\
 &= \frac{d}{d\bar{a}} \left[\frac{\partial V^{\bar{a}}}{\partial \bar{a}} \right] + \frac{d}{d\bar{a}} \left[\frac{\partial V^{\bar{a}}}{\partial \mathbf{s}} \frac{d\mathbf{s}}{d\bar{a}} \right], \\
 &= \left[\frac{d\mathbf{s}^T}{d\bar{a}} \frac{\partial^2 V^{\bar{a}}}{\partial \mathbf{s} \partial \bar{a}} + \frac{\partial^2 V^{\bar{a}}}{\partial \bar{a}^2} \right] + \frac{d}{d\bar{a}} \left[\frac{\partial V^{\bar{a}}}{\partial \mathbf{s}} \right] \frac{d\mathbf{s}}{d\bar{a}} + \frac{\partial V^{\bar{a}}}{\partial \mathbf{s}} \frac{d}{d\bar{a}} \left[\frac{d\mathbf{s}}{d\bar{a}} \right], \\
 &= \left[\frac{d\mathbf{s}^T}{d\bar{a}} \frac{\partial^2 V^{\bar{a}}}{\partial \mathbf{s} \partial \bar{a}} + \frac{\partial^2 V^{\bar{a}}}{\partial \bar{a}^2} \right] + \left[\frac{d\mathbf{s}^T}{d\bar{a}} \frac{\partial V^{\bar{a}}}{\partial \mathbf{s}} + \frac{\partial^2 V^{\bar{a}}}{\partial \bar{a} \partial \mathbf{s}} \right] \frac{d\mathbf{s}}{d\bar{a}} + \frac{\partial V^{\bar{a}}}{\partial \mathbf{s}} \left[\frac{d\mathbf{s}^T}{d\bar{a}} \frac{\partial}{\partial \mathbf{s}} \frac{d\mathbf{s}}{d\bar{a}} + \frac{\partial}{\partial \bar{a}} \frac{d\mathbf{s}}{d\bar{a}} \right], \\
 &= \frac{\partial V^{\bar{a}}}{\partial \mathbf{s}} \left(\frac{d\mathbf{s}^T}{d\bar{a}} \frac{\partial}{\partial \mathbf{s}} \frac{d\mathbf{s}}{d\bar{a}} + \frac{\partial}{\partial \bar{a}} \frac{d\mathbf{s}}{d\bar{a}} \right) + \frac{d\mathbf{s}^T}{d\bar{a}} \left(\frac{\partial^2 V^{\bar{a}}}{\partial \mathbf{s}^2} \frac{d\mathbf{s}}{d\bar{a}} + 2 \frac{\partial^2 V^{\bar{a}}}{\partial \mathbf{s} \partial \bar{a}} \right) + \frac{\partial^2 V^{\bar{a}}}{\partial \bar{a}^2}.
 \end{aligned}$$

A.2. Detailed description of environments

Table 2: Summary of environments

Environment	State space	Action space	Total mass	Constraints	Additional Info
2D Simple Pendulum	5	1	50gr	Handle along horizontal axis	
3D Simple Pendulum	8	2	50gr	Handle on horizontal plane	
3D Double Pendulum	14	2	100gr	Handle on horizontal plane	
Cable driven payload	$2*4 + 2 = 10$	2	200gr	Handles along horizontal axis	Attachment to handle uses deformable cables
Rope in 3D	$3*5*2 + 2 = 34$	2	250gr	Handle on horizontal plane	All conections are deformable cables
Cloth	$3*25*2 + 2*3*2 = 162$	$2*3*2 = 12$	100gr	Handles move freely on 3D	Uses differentiable frictional contacts

For the following environments we used a fixed time horizon of 150 steps with a time step of 24ms, with the exception of the cloth where we used a time horizon of 60 steps with a time step of 16ms.

- **2D Simple Pendulum:** This system corresponds to a cable-driven pendulum in 2D (Figure 2 left). The handle of the pendulum is constrained to move only along the horizontal axis in order to test the degree to which a control policy can exploit the natural dynamics of the system.
- **3D Simple Pendulum:** For this system the pendulum is free to move in 3D, but the handle is restricted to moving along a horizontal plane.
- **3D Double Pendulum:** Extending the dynamical system above, the payload for this problem consists of two mass points that are coupled to each other via a stiff bilateral spring. The dimensionality of the state space doubles, and the system exhibits very rich and dynamic motions.
- **Cable driven payload 2D:** For this environment we have a densely connected network of 4 point masses and two handles that are constrained to move along the horizontal axis.
- **Rope in 3D:** For this environment we use 5 point masses to discretize a rope in 3D and one handle that is constrained to move on the horizontal plane.

- **Cloth:** For this environment we use 25 point masses to discretize a piece of cloth and both handles can move freely in 3D. Furthermore, frictional contact against a table is applied to each point mass.

A.3. Architecture of neural network policies

The neural networks representing the control policies for all our environments share the same architecture, 2 fully connected layers of 256 units each with ReLU activations and one output layer with Tanh activation, to ensure that the policy only outputs commands that are within the velocity limits.

A.4. Differentiable simulator

Following the approach in (Zimmermann et al., 2018), the sensitivity $\frac{ds}{da}$ has the structure of the figure below.

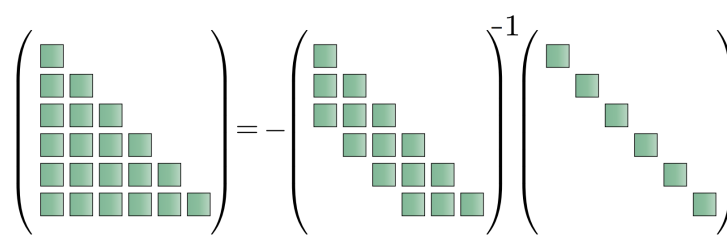
$$\frac{ds}{da} = - \left(\frac{\partial \mathbf{G}}{\partial s} \right)^{-1} \frac{d\mathbf{G}}{da}$$


Figure 7: Structure of sensitivity matrix $\frac{ds}{da}$ that encodes the dependency of a state on all the previous actions ¹

A.5. PODS: Additional figures

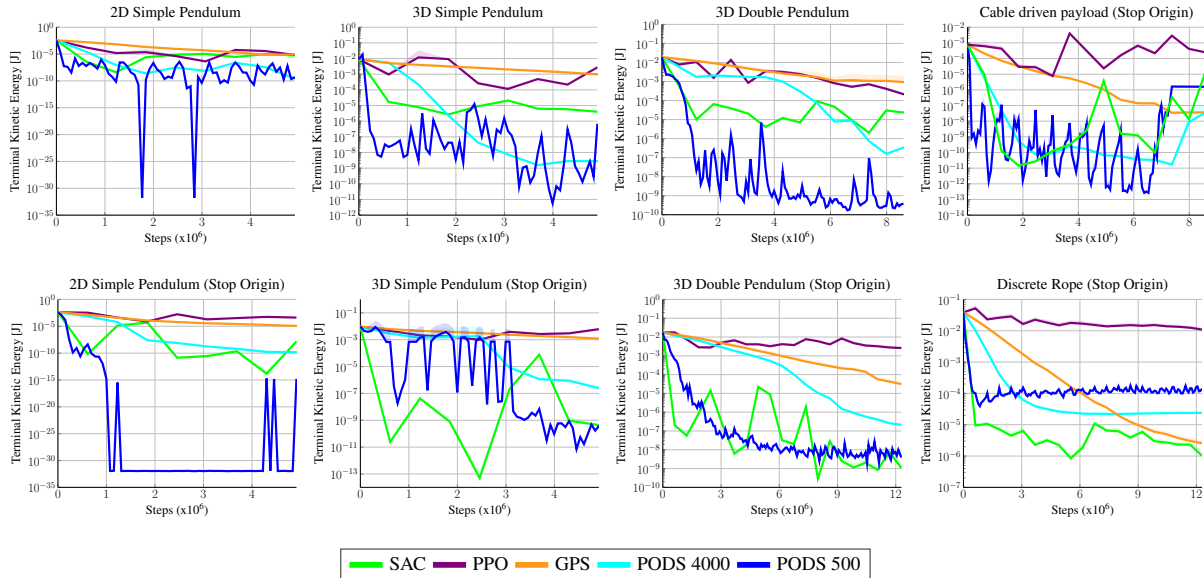


Figure 8: Final Kinetic Energy (averaged over a period of 10 time-steps after the policy is rolled out)

¹Figure reproduced with authorization of the authors (<http://arxiv.org/abs/1905.08534>)

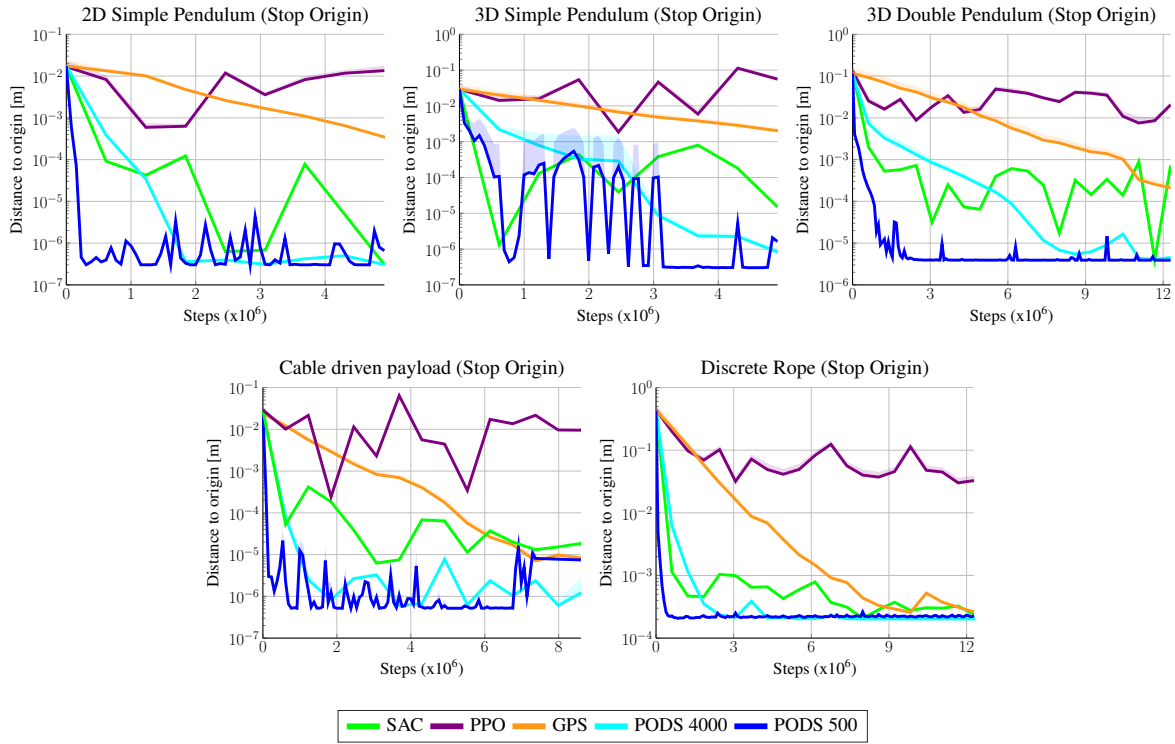


Figure 9: Final distance to Origin

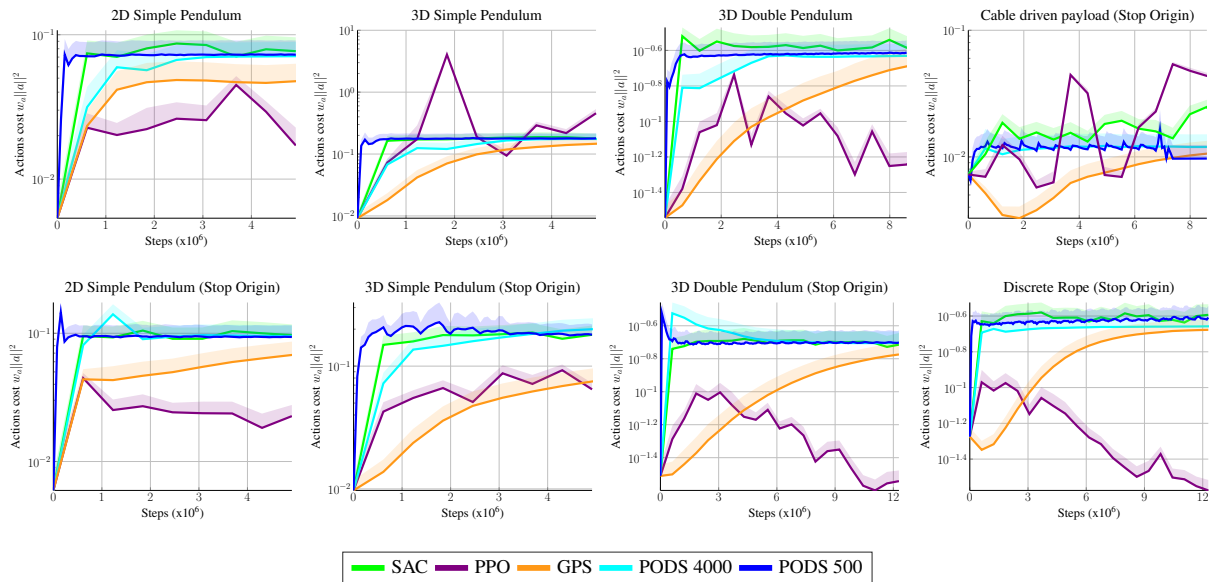


Figure 10: Average handle velocity (control effort)