Learning Robotic Interaction Tasks with Stability Guarantees

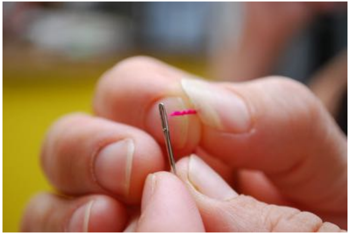
Shahbaz Abdul Khader, Hang Yin, Pietro Falco, Danica Kragic

IEEE International Conference on Multimedia and Expo

6-10 July 2020, London, UK

Multi-Modal Fusion and Learning for Robotics

Interaction Tasks (Contact-Rich Manipulation)



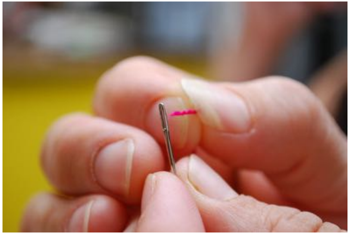
Contact-free manipulation: threading a needle[1]

Contact-rich manipulation: inserting a key[2]

[1] Cervantes Photography. March 2009. Online image. Flickr. June 2020. https://www.flickr.com/photos/g1gant3/3511501043/

[2] multifacetedgirl. Online image. Needpix. June 2020. https://www.needpix.com/photo/download/1119484/key-door-opening-door-lock-security-house-open-home-secure

Interaction Tasks (Contact-Rich Manipulation)



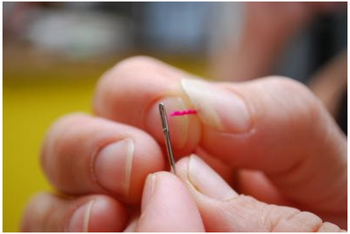
Contact-free manipulation: threading a needle

Contact-rich manipulation: inserting a key

- No exchange of forces and energy
- Purely kinematic
- Control = trajectory planning + feedback control

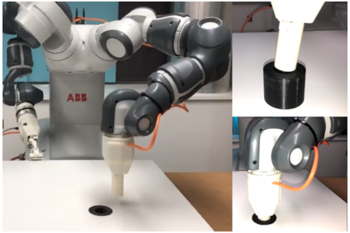
- Exchange of forces and energy
- Dynamics is complex and unknown
- Control ?
- Majority of human manipulation tasks

Interaction Tasks (Contact-Rich Manipulation)



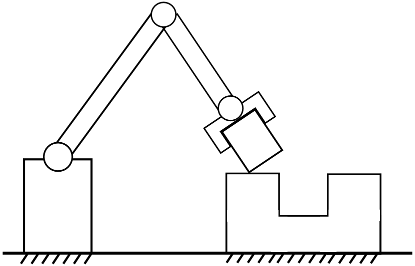
Contact-free manipulation: threading a needle

Contact-rich manipulation: inserting a key

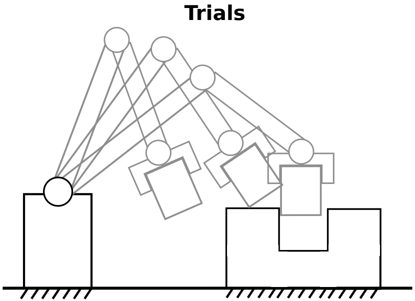


Peg-in-hole: benchmark for robots

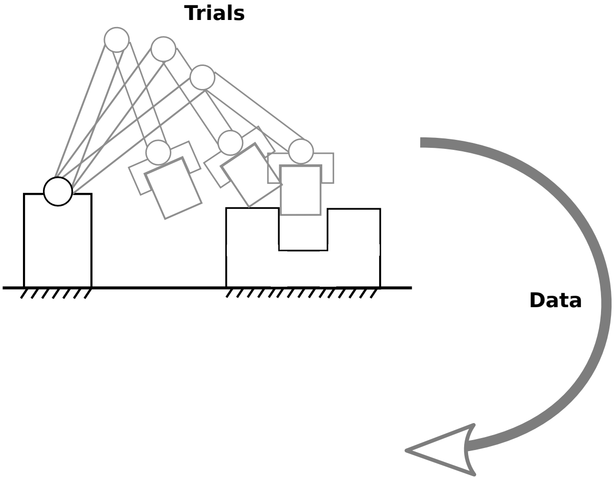
Reinforcement Learning (RL) of Interaction Tasks



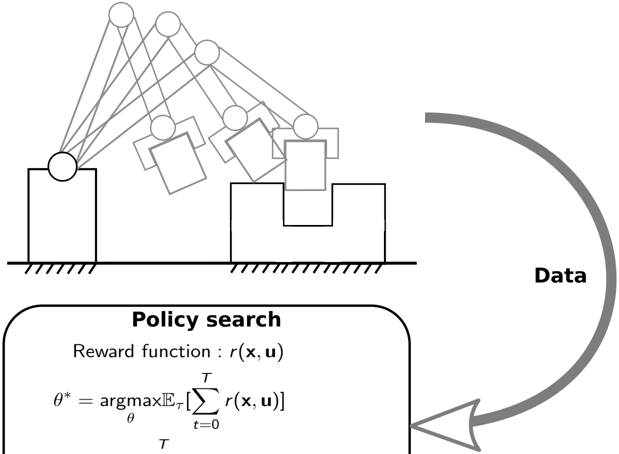
Reinforcement Learning (RL) of Interaction Tasks



Reinforcement Learning (RL) of Interaction Tasks



Reinforcement Learning (RL) of Interaction Tasks



Policy search

Reward function : $r(\mathbf{x}, \mathbf{u})$

$$\theta^* = \underset{\theta}{\text{argmax}} \mathbb{E}_\tau [\sum_{t=0}^{T} r(\mathbf{x}, \mathbf{u})]$$

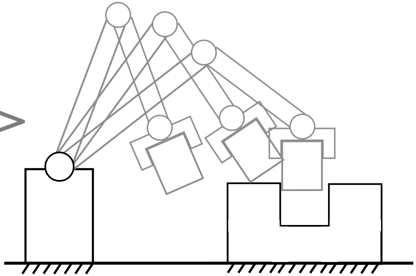
$$p(\tau) = p(\mathbf{x}_1) \prod_{t=1}^{T} \underbrace{\pi_\theta(\mathbf{u}_t|\mathbf{x}_t)}_{\text{policy}} \underbrace{p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t)}_{\text{dynamics}}$$

$$\tau : \{\mathbf{x}_1, \mathbf{u}_1, ..., \mathbf{x}_T, \mathbf{u}_T\}$$

Reinforcement Learning (RL) of Interaction Tasks

Model-free RL

Trials



Policy

Data

Policy search

Reward function : $r(\mathbf{x}, \mathbf{u})$

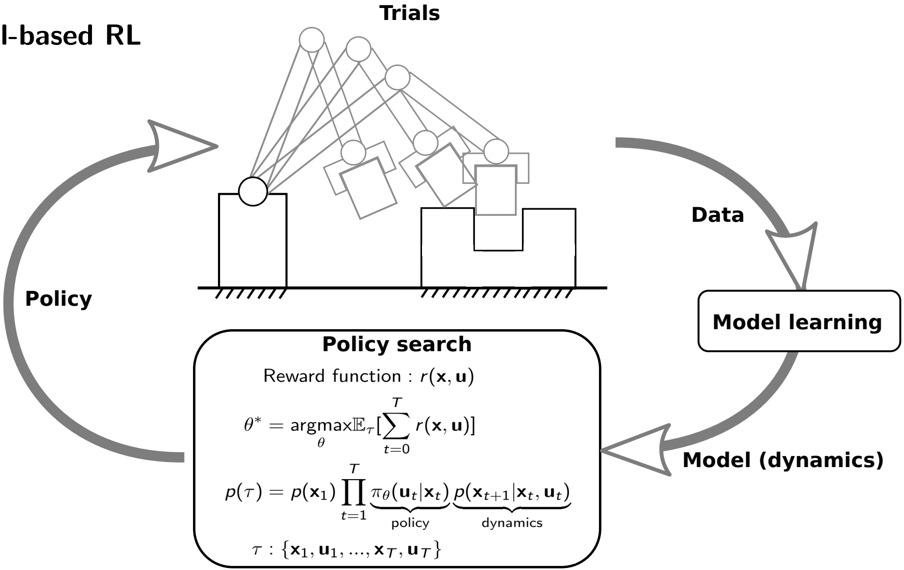
$$\theta^* = \underset{\theta}{\operatorname{argmax}} \mathbb{E}_\tau [\sum_{t=0}^{T} r(\mathbf{x}, \mathbf{u})]$$

$$p(\tau) = p(\mathbf{x}_1) \prod_{t=1}^{T} \underbrace{\pi_\theta(\mathbf{u}_t | \mathbf{x}_t)}_{\text{policy}} \underbrace{p(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{u}_t)}_{\text{dynamics}}$$

$\tau : \{\mathbf{x}_1, \mathbf{u}_1, ..., \mathbf{x}_T, \mathbf{u}_T\}$

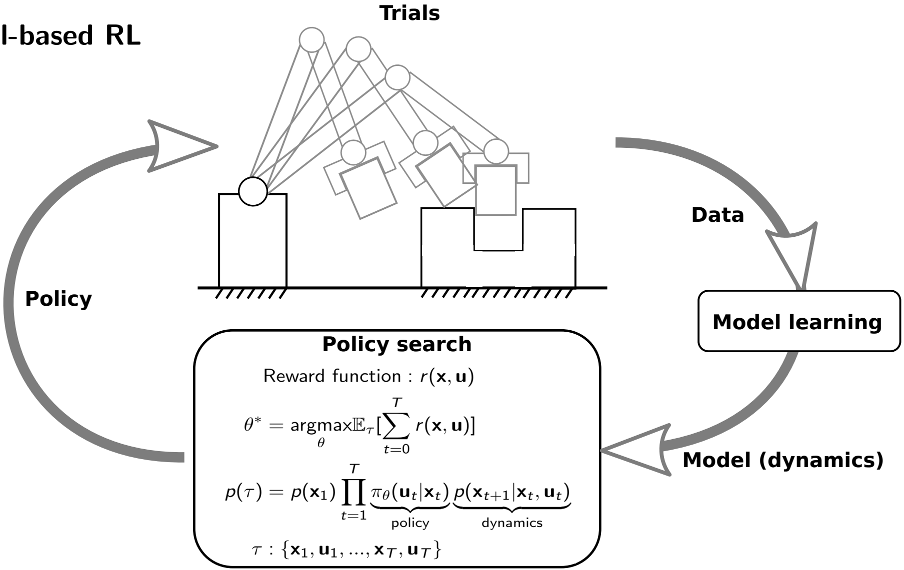
Reinforcement Learning (RL) of Interaction Tasks

Model-based RL



Reinforcement Learning (RL) of Interaction Tasks

Model-based RL



- **Specify only reward function! No trajectory planning + feedback control, no dynamics modeling**

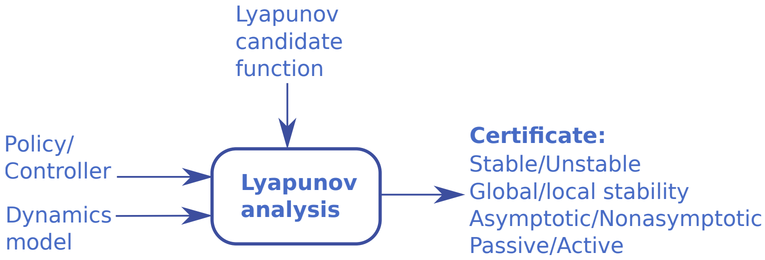
Unstable Motions in RL

Why stability?

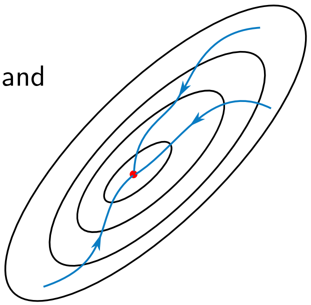
- A stability certificate is indispensable for real-world deployment
- It is the main means for safety

time-stability!

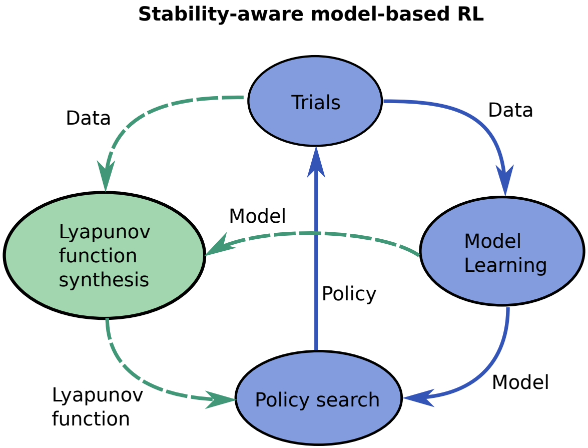
Lyapunov Stability



- Usually require analytic forms for policy, dynamics and Lyapunov function
- Lyapunov analysis is usually done manually



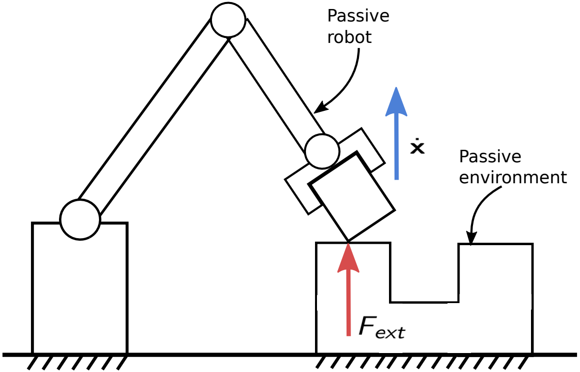
A Straightforward Approach for *all-the-time-stability*



- Model-learning of contact dynamics is hard
- How to cope with nonanalytic forms
- How to automate Lyapunov analysis?

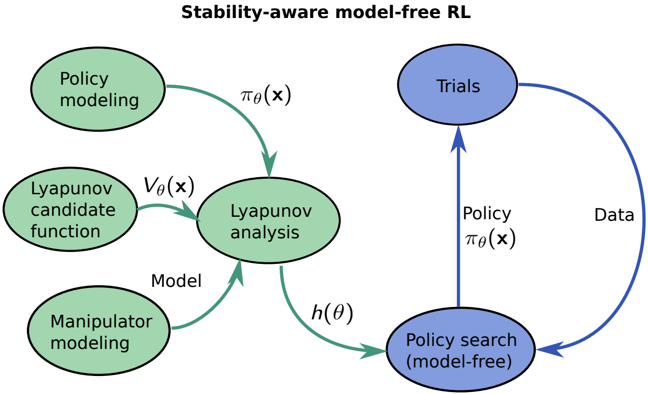
Stability of Robot-Environment Interaction

- Stability is retained when a *passive* manipulator interacts with a *passive* (unknown) environment[1].
- *Passivity*: can only dissipate or store energy but not create it.
- *Passive* manipulator $\implies$ a *passive* map from F_{ext} to $\dot{\mathbf{x}}$ w.r.t V, or $\dot{V} \leq F_{ext}^T \dot{\mathbf{x}}$.



[1] Colgate, J.E. and Hogan, N. (1988). Robust control of dynamically interacting systems, Int. J. Control, 48(1):65–88

A Smarter Approach

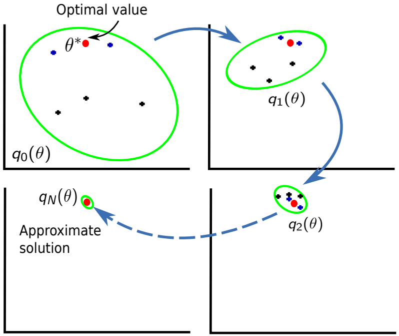


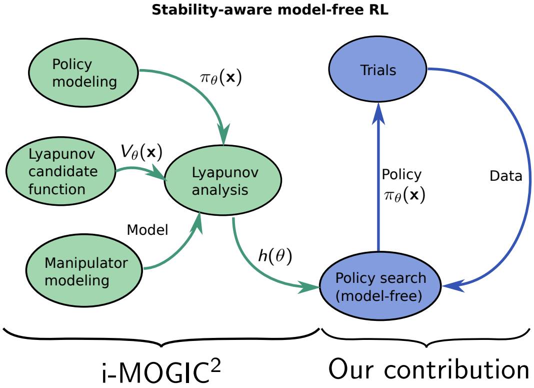
- No model is learned; only using manipulator model.
- Stability constraint $h(\theta)$ on shared parameter θ

Key enabling ideas:

- Ensure both the manipulator (controlled) and the environment to be passive.
- Lyapunov function $V_\theta(\mathbf{x})$ and policy $\pi_\theta(\mathbf{x})$ share the same parameters–updating $\pi_\theta(\mathbf{x})$ automatically updates $V_\theta(\mathbf{x})$.

Evolution Strategy (ES) based Model-Free RL[3]





Cross-Entropy Method based strategy

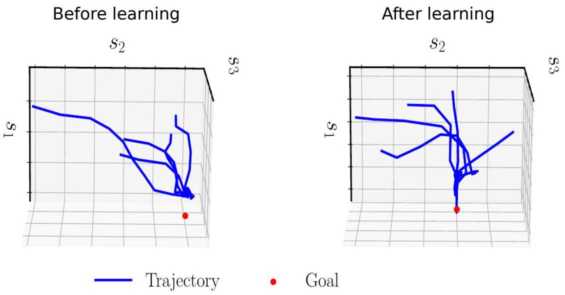
- $\theta = \{\boldsymbol{S}^0, \boldsymbol{D}^0, \boldsymbol{S}^k, \boldsymbol{D}^k, s^k, l^k\}$ for $k = 1, ..., K$, $\theta \notin \mathbb{R}^N$ (positive definite quantities)
- Novel sampling distribution $q(\theta)$ with Wishart factors (inherent constraint $h(\theta)$ satisfaction)

[2] Khansari-Zadeh, S. Mohammad, Klas Kronander, and Aude Billard. "Modeling robot discrete movements with state-varying stiffness and damping: A framework for integrated motion generation and impedance control." Proceedings of RSS 2014.

[3] Stability-Guaranteed Reinforcement Learning for Contact-rich Manipulation SA Khader, H Yin, P Falco, D Kragic - arXiv preprint arXiv:2004.10886, 2020

Experimental Results: Peg-In-Hole

- Insertion clearance of 0.5mm
- Full operational space control (translation and rotation)
- Learns in 300 trials



Conclusion and Future Work

- Reinforcement learning is important for interaction tasks.
- *All-the-time-stability* is essential for real-world deployment.
- Stable RL of peg-in-hole may be unprecedented.
- Limitations: specialized policies (i-MOGIC) may not be flexible enough and cannot incorporate (high dimensional multimodal) perception
- Future consideration: can we achieve *all-the-time-stability* using neural network policies?
- Submitted to IEEE Robotics and Automation Letters.

This work was supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM