Learning Force Control Policies for Compliant Robotic Manipulation

Mrinal Kalakrishnan[*] KALAKRIS@USC.EDU
Ludovic Righetti[*†] LUDOVIC.RIGHETTI@A3.EPFL.CH
Peter Pastor[*] PASTORSA@USC.EDU
Stefan Schaal[*†] SSCHAAL@USC.EDU
[*]CLMC Lab, University of Southern California, Los Angeles, CA 90089, USA
[†]Max-Planck-Institute for Intelligent Systems, 72076 Tübingen, Germany

1. Introduction

Developing robots capable of fine manipulation skills is of major importance in order to build truly assistive robots. These robots need to be compliant in their actuation and control in order to operate safely in human environments. Manipulation tasks imply complex contact interactions with an unstructured environment, and involve reasoning about the forces and torques to be applied. In order for robots to co-exist in an environment with humans, safety is a prime consideration. Therefore, touching and manipulating an unstructured world requires a certain level of compliance while achieving the intended tasks accurately.

Methods for planning kinematic trajectories for manipulators are well-studied and widely used. Rigid body dynamics models even allow us to plan trajectories that take the robot dynamics into account. However, once the robot comes into contact with the environment, planning algorithms would require precise dynamics models of the resulting contact interactions. These models are usually unavailable, or so imprecise that the generated plans are unusable. This seems to suggest alternate solutions that can learn these manipulation skills through trial and error.

In this abstract, we present an approach to learning manipulation tasks on compliant robots through reinforcement learning. We demonstrate our approach on two different manipulation tasks: opening a door with a lever door handle, and picking up a pen off the table (Fig. 1). We show that our approach can learn the force control policies required to achieve both tasks successfully. The contributions of this work are two-fold: (1) we demonstrate that learning force control policies enables compliant execution of manipulation tasks with increased robustness as opposed to stiff position control, and (2) we introduce a policy parameterization that uses finely discretized trajectories coupled with a cost function that ensures smoothness during exploration and learning. A full version of this paper is available (Kalakrishnan et al., 2011).

Appearing in *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012.

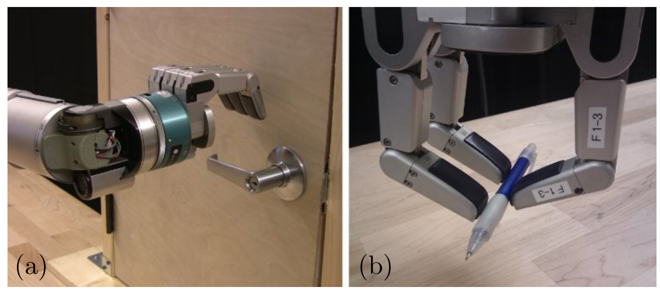


Figure 1. Force control policies for two different manipulation tasks were learnt using our method: (a) opening a door, and (b) picking up a pen from the table.

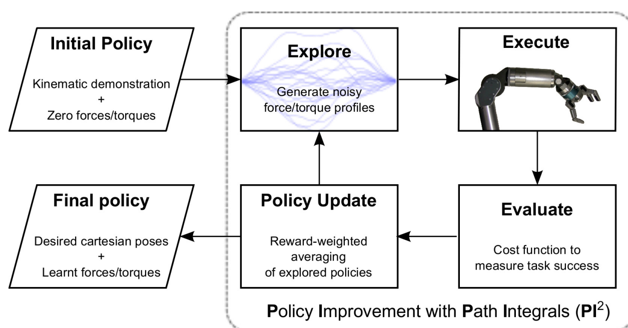


Figure 2. A high-level overview of our approach to learning force control policies for robotic manipulation.

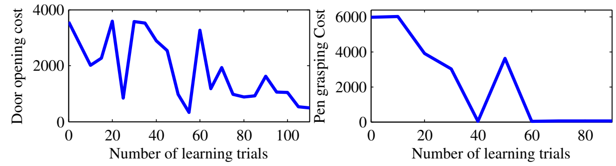
2. Algorithms

Fig. 2 shows a high-level overview of our approach to learning manipulation tasks on a compliant robot. The control policy is represented as a discretized trajectory of desired end-effector positions, orientations, forces and torques. Positions and orientations are initialized from a kinesthetic demonstration of the task. The required forces and torques cannot be observed during this process, and are initialized to zero. The policies are then optimized using the **P**olicy **I**mprovement with **P**ath **I**ntegrals (**PI**^{2}) reinforcement learning algorithm (Theodorou et al., 2010). This allows acquisition of a suitable force/torque control policy through trial and error. Policy performance is measured by a cost function that measures task success and penalizes squared accelerations of the trajectory. The latter part of the cost function is designed to be quadratic in the trajectory parameters. This specific choice of control cost in conjunction with the PI^2 algorithm allows for generation of smooth trajectories for exploration that do not deviate from the start or goal points, and ensures smoothness of the trajectory after every iteration. The algorithm samples trajectories around the current policy, measures their cost by executing them on the robot, and subsequently updates the policy as a weighted average of the samples. This process is iterated until convergence. More details may be found in the full paper (Kalakrishnan et al., 2011).

The combined position/force trajectory needs to be controlled by the robot in a suitable way during policy execution. PI^2, being a *model-free* reinforcement learning algorithm, is indeed agnostic to the type of controller used. It simply optimizes the policy parameters to improve the resulting cost function, treating the intermediate controllers and unmodeled system dynamics as a black box.

3. Experiments

Our approach was verified using two different manipulation tasks: opening a door and picking up a pen lying on a table. These tasks were chosen because each one involves significant contact with the environment, and are thus appropriate test-beds for the use of force control policies. The learning process and final executions of both tasks are shown in the cited video (vid). Both tasks were performed on the 7 degree of freedom Barrett WAM arm, equipped with a three-fingered Barrett Hand and a 6-DOF force-torque sensor at the wrist. Success in the door-opening task was measured by attaching an inertial measurement unit to the door handle to verify that the handle was turned and the door was indeed opened. For the pen-grasping task, success was quantified by measuring the amount of time the pen stayed within the fingers of the robot without slipping out. Both tasks were learnt in around 100 trials, and their final policies achieved 100% success. Fig. 3 shows the evolution of cost functions for each task during learning. The force control policy learnt for the pen-grasping task is shown in Fig. 4. This policy was also found to be robust to errors in the position and orientation of the pen.



(a) (b)

Figure 3. Evolution of cost functions during learning for the two manipulation tasks: (a) door opening, and (b) pen grasping.

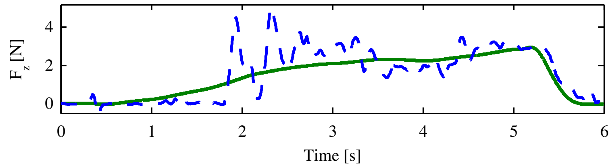


Figure 4. Learnt force/torque control policy and tracking errors for the pen grasping task. The green solid line shows the learnt desired force in the z axis. The blue dashed line indicates the corresponding measured force.

4. Conclusion

The need for compliant actuation in robotics is well understood. In this abstract, we have presented a learning approach for compliant manipulation. The approach relies on initialization of a desired position trajectory through kinesthetic demonstration, followed by learning of a desired force profile through reinforcement learning. We have successfully demonstrated the application of this approach to two different manipulation tasks. Explicit training of these policies for robustness is a promising direction for future work.

Acknowledgments

This research was supported in part by National Science Foundation grants ECS-0326095, IIS-0535282, IIS-1017134, CNS-0619937, IIS-0917318, CBET-0922784, EECS-0926052, CNS-0960061, the DARPA program on Autonomous Robotic Manipulation, the Army Research Office, the Okawa Foundation, the ATR Computational Neuroscience Laboratories, and the Max-Planck-Society.

References

Video. `http://www.youtube.com/watch?v=LkwQJ9_i6vQ`.

Kalakrishnan, M., Righetti, L., Pastor, P., and Schaal, S. Learning force control policies for compliant manipulation. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pp. 4639–4644. IEEE, 2011.

Theodorou, Evangelos, Buchli, Jonas, and Schaal, Stefan. A generalized path integral control approach to reinforcement learning. *Journal of Machine Learning Research*, pp. 3137–3181, 2010.