INVESTIGATION OF DISTURBANCES ON A NOVEL REINFORCEMENT LEARNING BASED CONTROL APPROACH

ALBERS, A[lbert]; SOMMER, H[ermann]; FRIETSCH, M[arkus] & TREMMEL, C[hristoph]

Abstract: Automatic manipulator control poses a complex challenge for control systems with various dynamic and nonlinear effects as well as high dimensionality issues. The aim of this paper is to analyze the behavior of a novel approach for a reinforcement learning based motion control of a 2-DOF robotic manipulator when it is subdues to the effect of a set of different disturbances. The implementation of the presented approach shows the performance of the manipulator when circumstances force him into unexpected situations. The experimental results are presented
Key words: machine-learning, robotics, dynamics, control

1. INTRODUCTION

1.1 Reinforcement Learning in modern Robotics

One of the biggest challenges in current research in robotics is, that robots "leave" their well structured environment and are confronted with new tasks in a more complex environment. An example for this new setting is the striving for autonomy and versatility in the fields of humanoid robotics as explained in (Peters et al., 2003). Due to this, it can only be successful resp. useful, when it is able to adapt itself and to learn from its experiences. Reinforcement Learning (RL), a branch of machine learning (Mitchell, 1997), is one possible approach to deal with this problem. However, the application of this learning process is limited due to its complexity. RL is a learning process, which uses reward and punishment signals from the interaction with the agent's environment in order to learn a distinct policy for achieving tasks. Various RL methods e.g. Q-learning (Watkins, 1989) or the SARSA algorithm have been studied in (Sutton & Barto, 1998) where it is shown that two problems must be considered (Park & Choi, 2002). At first, the high computational efforts: RL is disturbed by the "curse of dimensionality" (Sutton & Barto, 1998), which refers to the tendency of a state space to grow exponentially in its dimension.

Secondly, the information of every different task is stored separately. This results in both a storing place and computational effort issue and reduces the usefulness of RL for practical applications. Furthermore, it poses the question of how already acquired knowledge can be reused. In (Martin & De Lope, 2007), an approach is presented where a distributed architecture in RL serves as a pragmatic solution for some common robotic manipulators with different DOF. RL based approaches have been applied to various robotic systems in the past, although mostly applied to the learning of elemental tasks meant to serve as "building blocks of movement generation" as in (Peters, 2008). Nevertheless, new computations and additional storage space is required for performing new tasks.

The adaptability of these approaches regarding new, unexpected situations using already acquired knowledge has not been studied in detail. In this paper, the behavior of the novel relative approach for positioning tasks presented in (Albers et al., 2010) when exposed to a set of different disturbances is investigated and the experimental results are presented.

1.2 Model of the 2-DOF planar Robot

A 2 DOF manipulator system is chosen as a prototypical nonlinear system. Fig. 1. shows the schematic drawing of the manipulator including all relevant system parameters. The aim is to create a trajectory using as few control commands as possible. The state of the manipulator is described by:

$$s = [\theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2] \tag{1}$$

The system is detailedly described in (Denzinger & Laureyns, 2008) and adapted for the purpose of this investigation.

2. CONTROL OF SYSTEM DISTURBANCES USING THE RELATIVE APPROACH

2.1 Overview of the relative Approach

Using the relative approach developed in (Yan et al., 2009) each position can be calculated as a constant value plus a difference possible, this way every positioning task can be remodeled to a simple offset compensation, reducing every possible individual tasks to a single one. When the knowledge about this task is properly implemented as shown in (Albers et al., 2010), no new learning is required to complete new tasks. An exhaustive investigation of this approach suggests that the RL-algorithm shows a behavior similar to classical closed loop controllers. To test this behavior, a representative set of disturbances were implemented and the manipulator's response was studied.

2.2 Experimental Setup

In order to evaluate the response of the manipulator to possible disturbances it was necessary to adjust the underlying dynamic system employed for the computation of the agent's behavior.
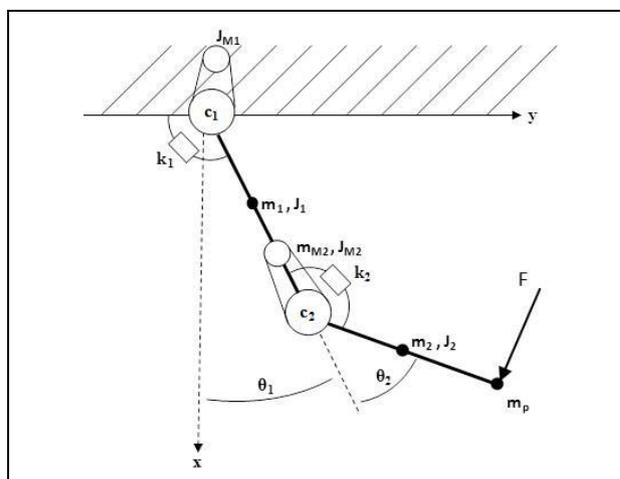

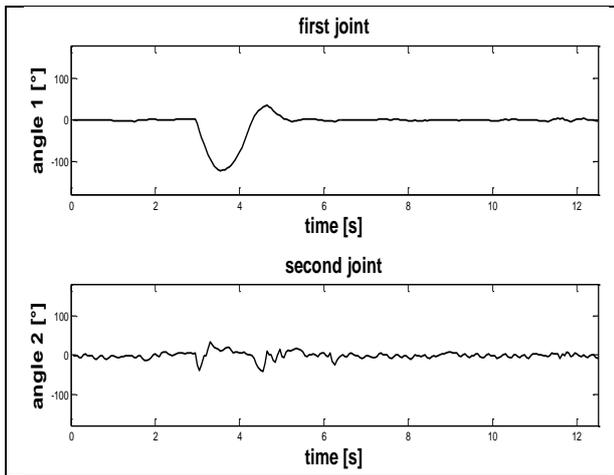Fig. 1. Dynamic model of the manipulator

Fig. 2. Force Impulse-Disturbance

Fig. 3. Force Step-Disturbance

For the purpose of our investigation we confront the manipulator control with two commonly encountered example disturbances. The first is a force impulse with a magnitude of 5 N which simulates the manipulators response to short time disturbances like a collisions or current peaks. The second is a force in form of a step signal, which represents a disturbance with a longer duration. The magnitude of the second disturbance is 0.5 N. The smaller magnitude was selected due to the lengthy application time. The point of application of the disturbing force is the Tool Center Point (TCP) of the robotic agent, located at the end of its second link. The direction of application is perpendicular to the second link at all times. The agents task during an experiment is to maintain the outstretched position $(\Delta\theta_1, \Delta\theta_2) = (0, 0)$ for 250 time steps (1 interval = 0.05 s) and to compensate the positioning offset inflicted by the disturbances if necessary.

For both cases under study the application time of the disturbances was set to 3 s. The experiments end after 12.5 s and the response curves for both links are plotted. The results for the first test case are depicted in Fig. 2, while the results for the second test case are depicted in Fig. 3. The results of conducted experiments with varied magnitudes for the applied external disturbances, which are not depicted here, showed similar behavior.

3. RESULTS

The response of the manipulator to the impulselike disturbance is depicted in Fig.2. For the first 3 seconds of the experiment the manipulator maintains the outstretched target position without noticeable difficulties. At this point the force impulse is applied to the second joint and the manipulators first joint is clearly brought out of position, while the second joint shows a smaller angular deviation as a consequence of the disturbance than the first one. After approximately one single oscillation around the target position, the first joint stabilizes. After around 5 s the manipulator has reestablished the outstretched configuration, which he then successfully maintains till the end of the experimental task as seen in Fig. 2.

In the second experimental setup the disturbing force is applied to the manipulator responding to a step function after 3 s. The first joint angle oscillates briefly before it reaches a stable position around -90° away of the manipulator, thus not being able to compensate the angular offset. The second joint remains unaffected after the disturbing force is applied and maintains its position for the rest of the experimental task. Throughout all experiments the second joint proved to be less influenced by external forces.
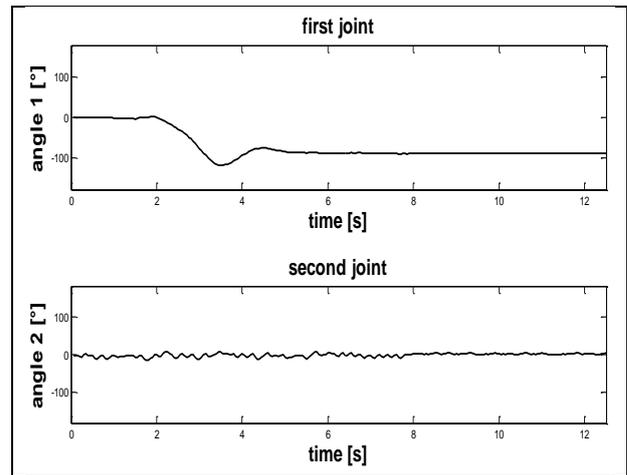
4. CONCLUSION

The manipulator is capable of reacting positively to short term disturbances and of reestablishing the target position after such a disturbance. During the time of effect of the disturbances, the underlying mapping of states and actions is no longer valid since the system dynamics are changed. For disturbances with a longer duration of effect, or even time invariant disturbances, a new learning procedure or remapping is necessary. Nevertheless, the approach does present a remarkable advantage towards a traditional open loop control. Having learnt the new mapping though, the manipulator could recognize the change in its environment and adapt or even load a previously computed new mapping of its behavior accordingly. A possibility would be to implement a model based control to recognize differing mappings and adjust the learning behavior. Future research includes this topic among others like discretization smoothness and partial recoupling of the manipulator axis. Furthermore, additional experiments will be conducted on a three DOF system.

5. REFERENCES

Albers, A.; Sommer, H. & Frietsch, M. (2010). A New Approach for Solving Positioning Tasks of Robotic Systems Based on Reinforcement Learning, Annals of DAAAM for 2010 & Proceedings of the 21st International DAAAM Symposium

Denzinger J.; Laureyns I. & et.al. (2008). A study of reward functions in reinforcement learning on a dynamic model of a two-link planar robot, The 2nd European DAAAM International Young Researchers´ and Scientists´ Conference

Mitchell, Tom M. (1997). *Machine learning*, McGraw-Hill, New York, ISBN 0-07-042807-7

Peters J., Vijayakuar S. & Schaal S. (2003) Reinforcement learning for humanoid robotics. Third IEEE-RAS International Conference on Humanoid Robots, Karlsruhe, Germany

Peters J. (2008). Machine Learning for Robotics. VDM Verlag Dr. Müller, Saarbrücken, ISBN 978-3-639-02110-3

Sutton R.S & Barto A.G. (1998). Reinforcement Learning, an introduction, *The MIT press*, MA

Yan, W & et. al. (2009). Application of reinforcement learning to a two DOF Robot arm control, Annals of DAAAM for 2009 & Proceedings of 20th DAAAM International Symposium