Karlsruhe Institute of Technology

Recursive Surrogate-Modeling for Stochastic Search

Bachelor's Thesis
of

Klaus Philipp Theyssen

KIT Department of Informatics
Institute for Anthropomatics and Robotics (IAR)
Autonomous Learning Robots (ALR)

Referees: Prof. Dr. Techn. Gerhard Neumann
Prof. Dr. Ing. Tamim Asfour

Advisor: M.Sc. Maximilian Hüttenrauch

Duration: November 27th, 2020 — March 27th, 2021

Erklärung

Ich versichere hiermit, dass ich die Arbeit selbstständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht habe und die Satzung des Karlsruher Instituts für Technologie zur Sicherung guter wissenschaftlicher Praxis beachtet habe.

Karlsruhe, den 31. März 2021

Klaus Philipp Theyssen

Zusammenfassung

Abstract

The use of robots is expected to become part of our everyday life's. There are is quite a market for little house cleaning robots or toys. Most of these robots rely on being taught and programmed by a skilled human operator. This means robots still have problems to automatically acquire new skills and changing environments and requirements. Recent research is focusing on Reinforcement Learning as a framework to solve the robot control problem.

Policy search methods, a sub-field of reinforcement learning, try to optimize the robot's policy parameters for the task at hand.

In this thesis we try to improve the sample efficiency of a policy search algorithm, which is of special importance for robotics. This is done by using classical recursive estimation techniques like the Kalman filter.

We implement different versions of Filtering algorithms and compare them with previous methods and benchmark them on optimization test function and simple planar robot tasks.

Table of Contents

Chapter 1.

Introduction

1.1. Motivation

Robots are expected to transform our production methods and society as a whole. Currently mostly toy applications [cite] are widely known to the public, or highly specialized robot systems in industry [cite] which preform a special task. What lacks is a freely moving agent, able to adapt to various situations and acting autonomously. To achieve this goal recent research has focused on reinforcement learning [cite], and data driven methods.

Robotics and reinforcement learning go hand in hand, giving birth to a rich relationship similar to math and physics [cite robotics and RL survey].

- give inspiration for robotics in general terms

- write story about using a robot and solve tasks with different contexts

- introduce robots to our ever-day life (big vision), robots need to autonomously learn rich set of complex behaviors

- robotics offers platform of application for reinforcement learning (compare physics and mathematics)

- robots ability to generalize experience across similar tasks, adjust knowledge of the task to new setting or context

- give example of tasks that are solve able with policy search

- include pictures of robots solving tasks

1.2. Contribution

Trust region policy search based on Kullback-Leibler divergence has been successfully employed in some sample tasks [REPS,...]. Abdolmaleki et al. (2015) introduced MORE to alleviate issues of a lot of evaluations of objective and converging prematurely. One of the contributions of this thesis is to explore recursive estimation of the learned model of the objective function to increase sample efficiency. Besides that we aim to improve the overall runtime of the algorithm. We focus mainly on classical methods of parameter estimation and filtering like the Kalman Filter, considering more advanced methods is part of future work. We benchmark the new version of the MORE algorithm on test functions and compare them to previous results.

1.3. Structure of Thesis

The remainder of this thesis is structured as follows:

Chapter 2: In the fundamentals chapter we, we lay the foundation for the thesis, introducing Reinforcement Learning for solving Robotic tasks. Introducing basic concepts and notation. We look at policy search algorithms which leads us to stochastic search algorithms. We shortly review some optimization principles including Lagrangian multipliers. Finally we look at Bayesian parameter estimation, focusing on the Kalman Filter.

Chapter 3: Review of the related work in the field.

Chapter 4: In this chapter we develop the algorithm connecting the MORE algorithm with recursive parameter estimation, first using simple Recursive Least Squares and finally using a Kalman Filter.

Chapter 5: In the evaluation chapter we conduct experiments with the algorithms on several tests function and some toy tasks for a planar robot. We compare our algorithm with original MORE and other stochastic search algorithms.

Chapter 6: Finally, we conclude the thesis with a summary of the achieved results and an outlook on future work.

Chapter 2.

Fundamentals

This chapter introduces basic concepts used throughout this thesis. First Markov decision process and classical reinforcement learning. Then we look at the robot control problem with RL as a framework. Then discuss policy search, a sub-field od reinforcement learning, as one method to solve this problem. The Kullback-Leibler divergence (KL), as an important information theoretic distance methric between probability distributions and Dynamic Motor Primitives (DMPs) for policy representation. Then basics of stochastic optimization algorithms, lagragian multiplier and the dual function. Having introduced the underlying basics we can then discuss the MORE Algorithm as a policy search algorithm for solving the robot control problem. Finally we will look at Filtering from a Bayesian Estimation viewpoint and review the Kalman Filter for parameter estimation.

2.1. RL in Robotics

Agent explores space of possible strategies and receives feedback on the outcome of the choices he made. - [figure: clasical RL interaction loop]

2.1.1. Markov Decision Processes

- cite bellman - long term reward function - –> use notation from deisenroth talk - agent interacting with environment - policy to choose action for given state - maximize reward, episodic task - for task without end, average, expected reward

2.1.2. Robot Control as a RL Problem

- curse of dimensionality - curse of real-world samples - accumalted reward - optimal policy

2.2. Policy search

- look for optimal policy - model-free vs. model based - exploration-exploitation trade-off

2.3. Kullback-Leibler (KL) Divergence

- equation - used to limit exploitation of surrogate model

2.4. Dynamic Motor/Movement Primitives (DMPs)

- time-dependent policy representation that is commonly used in robotics - second-order dynamic system - compact representation of basic movements such as hitting and grasping - trajectory generators for real system - dynamical system: - system in which a function describes the evolution of a point in a geometrical space over time - basic example is first order linear dynamical system

- DMPs opt for statbility of second-order dynamical systems and introduce further complexity to the trajectory by adding a non-linear forcing function

- temporal scaling factor - joint positions, spring damping

–> θ are the parameters of the DMPs and will optimized by MORE algorithm to find optimal policy, and thus a robot/robotarm will perform the corresponding trajectory encoded by the DMPs

2.5. MORE Algorithm

MORE black-box optimizer, key points: - trust region method, in which optimization steps are restricted to lie within a region where the approximation of the true cost function still holds - preventing updated policies from deviating too wildly, catastrophically bad update is lessened –> monotonic improvement in policy performance - objective function is locally

approximated by a quadratic model - KL divergence for staying close to the old policy and having the policy as a proper distribution - analytic solution for the policy update derived using Lagragian multipliers (appendix) - that model used to locally optimize the objective function

2.5.1. MORE Framework

- policy search problem as constrained optimization problem - Dual problem

2.5.2. update of policy

- in gaussian case

2.5.3. Iteration step

- pseudo code for MORE

2.6. Bayesian Estimation

- Bayesian probability, and Bayesian toolbox - prior distribution - measurement model - posterior distribution

2.6.1. Least squares

- batch solution to bayes filtering problem

2.6.2. Recursive least squares

- iterative step wise solution

2.6.3. Kalman Filter

- dynamic model solution, for linear problem assumption

Chapter 3.

Related Work

3.1. Policy Search Algorithms

- episodic REPS –> use of KL-divergence

- in MORE algorithm we use surrogate model to calculate KL-divergence bound, opposed to a Taylor approximations or sample based approximation of the KL-bound

- natural gradient by NES (update direction of parameters like standard gradient but still in bound by KL-divergence)

- CMA-ES is state of the art in stochastic search optimization and the sample efficiency of it

3.2. Parameter Estimation

- look for some examples of kalman filter used for parameter estimation

Chapter 4.

Recursive Surrogate-Modeling for MORE

4.1. MORE with Recursive Least Squares for Surrogate-Modeling

- estimation of covariance of measurement and parameters with help of gradient information?

4.2. MORE with Kalman Filter for Surrogate-Modeling

- discuss problem of estimating the quadratic surrogate model

Chapter 5.

Evaluation

This chapter will introduce the setup for the experiments and the tasks.

5.1. Experiments

5.1.1. Setup

5.1.2. Test Functions for Optimization

Based on Molga and Smutnicki (2005).

Figure 5.1 shows the sphere function, one of the simplest test benchmarks.

$$f(x) = \sum_{i=1}^{n} x_i^2$$

Figure 5.2 shows the rosenbrock function, a uni-model optimization function.

$$f(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2]$$

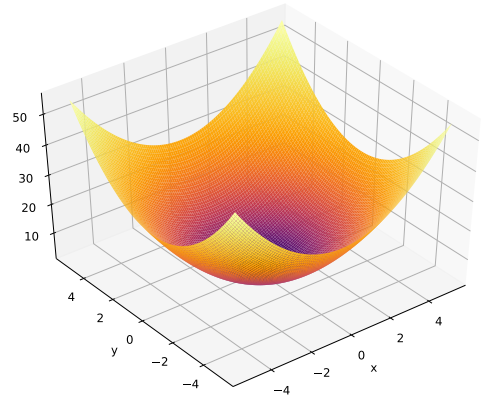

Figure 5.1.: sphere function

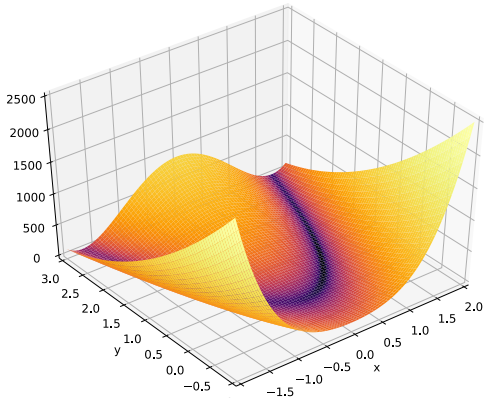

Figure 5.2.: rosenbrock function

5.1.3. Planar reaching task

- DMPs as policy representation

5.2. Evaluation

Chapter 6.

Conclusion and Future Work

6.1. Conclusion

Your conclusion.

6.2. Future Work

Your ideas about possible future works.

Bibliography

M. Molga and C. Smutnicki. Test functions for optimization needs. *Test functions for optimization needs*, 101:48, 2005.

Appendix A.

Appendix

A.1. MORE: Lagragian Dual Function