Supervised Learning: A potential tool for feasible motion planning in State Space

Mukunda Bharatheesha[1] and Martijn Wisse[1]

***Abstract*— Planning motions that enable robots perform complex tasks has been widely studied. Motions that adhere to the kinodynamic constraints would enable robots to perform such tasks by maximizing the use of natural dynamics. Sampling-based planners such as Rapidly Exploring Random Trees are a useful tool for kinodynamic motion planning. In such planners, the choice of distance metric plays an important role in the feasibility of generated motion plans. However, computing the distance metric, which is the optimal cost-to-go between states, takes a significant amount of time. On the other hand, using a supervised learning algorithm to obtain reasonably good estimates of the distance metric considerably alleviates the problem of metric computation time. In this work, we present a generic framework that combines the domain of optimal control and supervised learning that approximates distance metrics in state space in quick time.**

I. INTRODUCTION

The most common solution space for kinodynamic motion planning problems is the state space of a robotic system as trajectories in state space implicitly consider the associated dynamical constraints. Sampling-based approaches such as Rapidly Exploring Random Trees (RRT) [1] and Probabilistic Road Map (PRM) [2] have provided important insights towards solving kinodynamic motion planning problems. In our work, we use the RRT algorithm, but our approach itself is not restricted to RRT alone.

Ever since the basic RRT algorithm proposed the Euclidean distance as the choice of distance metric in [1], significant amount of research has been conducted in order to improve the estimation of distance metric in state space. For example, the authors of [3], [4] present Linear Quadratic Regulator (LQR) based heuristics to compute state space distances and show that these approximations significantly improve the solutions from RRT algorithms over Euclidean distance. An important drawback however, is that the point-based linearization technique that is an integral part of LQR-based approaches significantly compromises the use of natural system dynamics. Also, it is well known that optimal cost determination is computationally expensive to be used in the nearest neighbor search of RRT.

II. APPROACH

The central idea of our framework is to generate and use a learning-based approximation of the optimal cost-to-go. This enables us to leverage the benefits of optimal control with the speed advantage from learning which is, in principle, a good combination for motion planning in state space using RRT. In other words, we benefit from having a reasonable approximation of the distance metric in state space with the computational cost of a learning approximation, which is much lower compared to computing an optimal control solution between pairs of multiple states [7]. The basic idea of our scheme is represented in the block diagram of Fig. 1.

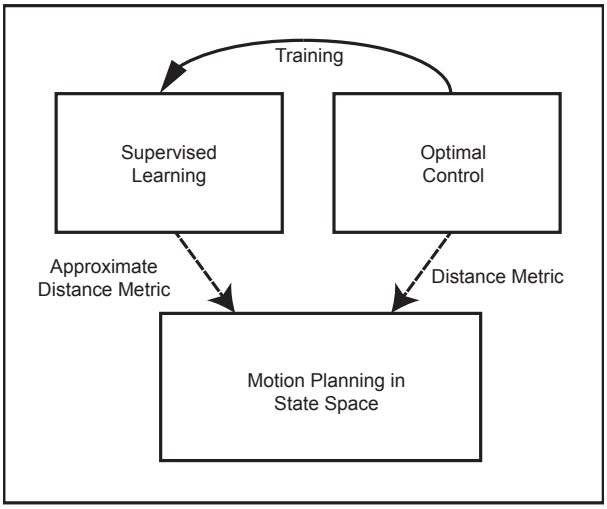

Fig. 1. Schematic of the learning-based approach to distance metric approximation in state space.

The generality of our approach stems from the fact that our scheme allows for experimentation with a variety of combinations of supervised learning and optimal control schemes. We consider one such combination. In the following section we briefly introduce the learning and the optimal control methods used and reason further regarding the choice we make.

III. COMBINATION OF OPTIMAL CONTROL AND LEARNING

We use Iterative Linear Quadratic Gaussian (iLQG) [5] and Locally Weighted Projection Regression (LWPR) [6] as the candidates for optimal control and supervised learning respectively. The motivation for the choice of the optimal controller is the trajectory-based linearization scheme that is used in iLQG which aids significantly in accounting for the natural dynamics of a given system much better than point based linearization schemes such as in [4], [3]. Also, the choice for LWPR is motivated by the ability of the learning

[1]Faculty of Mechanical, Maritime and Materials Engineering, Mekelweg 2, Delft University of Technology, 2628 CD Delft, The Netherlands {m.bharatheesha, m.wisse}@tudelft.nl

algorithm to incrementally learn nonlinear relationships between input and output data in high dimensions. However, we have not been able to completely utilize this benefit yet which we will detail further in the discussion section. Due to space restrictions, we turn our focus now to discussing the current results we have using our framework and refer to [7] and the references therein for further details. Subsequently, we conclude with a discussion on some important observations which provide directions for further research to gain maximum benefits from our framework.

IV. Current Results

To get useful approximations of the optimal cost-to-go, it is necessary to provide sufficient number of training samples to LWPR learning. This is accomplished by Algorithm 1. The training samples are generated as input-output pairs with inputs being two states between which the optimal cost-to-go is computed and the output is the optimal cost computed by iLQG controller. We have considered a quadratic cost function of the control effort.

Algorithm 1 LearnMetric (nSamples, dt)

for $j = 1,\ldots n_{\text{Samples}}$ **do**
  $x_i \leftarrow$ `Sample`
  $x_f \leftarrow$ `Sample`
  $cost_{optimal} \leftarrow$ `ILQG` (x_i, x_f, dt)
  $X_{tr}(j) \leftarrow [x_i; x_f]$
  $Y_{tr}(j) \leftarrow cost_{optimal}$
end for
return $\rho_{\text{learn}} =$ `LWPRlearnmetric` (X_{tr}, Y_{tr})

The `LearnMetric` procedure ensures unbiased learning by sampling the states uniformly at random. These random samples are equally divided into a training and a test data set by the `LWPRlearnmetric` procedure in order to generate and validate the learning model .

In our experiments, we studied the performance of the learning-based approximation on three different dynamical systems in simulation; a simple pendulum, the acrobot and a differentially driven mobile robot. The learnt models were subsequently used in the basic RRT algorithm to compute the nearest neighbors to a randomly sampled state. And the correctness of the models were verified by computing the squared difference between predicted and actual cost of traversal between the chosen nearest neighbor and the randomly sampled state. The prediction error mean and variance for the three models are indicated in Table I. It is to be noted here that, both input and output training data were normalized to a range of $[0, 1]$ and the errors were computed by normalizing the actual cost of traversal using the same range used for training. The mean and the variance of the squared prediction errors for all three models were calculated by creating an RRT of 500 nodes.

From the mean and variance values of the squared prediction errors, it is clear that the magnitudes of prediction errors are relatively low. Thus, in all three cases with different state dimensions, the results show that a successful approximation of the cost-to-go between different states is achieved. In the following section, we discuss the main aspects that we are currently studying further in order to achieve the best performance from our framework.

TABLE I
Mean and variance of squared prediction errors.

System	State dimension	Learning dimension	Mean/ Variance
Pendulum	2	4	$1.8e-3$ $5.3e-5$
Acrobot	4	8	$4.4e-4$ $1.7e-7$
Mobile Robot (Differential Drive)	5	10	$7.2e-3$ $1.3e-4$

V. Discussion and Future Work

It was briefly mentioned in the previous section that the input and output data were normalized to a $[0, 1]$ range. For this purpose, a sufficiently large number of randomly generated states were used. This enabled us to assume a reliable range for the velocities for normalization. Furthermore, we identified that there is a significant difference between the prediction accuracy of the learnt models with and without input-output normalization. The impact of input normalization on the prediction accuracy is quite understandable as that provides a uniform range along the input dimensions for generating learning models. An important aspect we are currently investigating is the influence of output normalization on the prediction accuracy.

Currently, the incremental learning benefits of LWPR are unused as large number of training samples are required. However, this issue provides us with a new direction of embedding dynamical model of the systems in the learning algorithm. This is based on the fact that state pairs which are on the trajectory of the natural evolution of system dynamics need not interfere with learning the cost models as they are all zero cost state pairs. We believe other supervised learning approaches such as deep learning [8] can also serve as interesting options to study the quality of the cost approximation in comparison to LWPR.

Finally, the main benefit of the learning-based approximations for feasible motion planning in state space can be observed in the resulting motion plans from the basic RRT algorithm. In particular, for the pendulum and acrobot swing-up problems where natural dynamics play an important role in achieving the desired goal positions, the performance difference is quite significant in comparison to approaches such as in [4]. Due to space restrictions, we are confined to only stating these benefits here. Our next step is to use this framework for kinodynamic motion planning on robot arms such as the UR5 and the Kuka LWR.

REFERENCES

[1] S. M. LaValle and J. J. Kuffner-Jr, "Randomized kinodynamic planning," *The International Journal of Robotics Research*, vol. 20, pp. 378–400, 2001.

[2] D. Hsu, R. Kindel, J.-C. Latombe, and S. Rock, "Randomized kinodynamic motion planning with moving obstacles," *The International Journal of Robotics Research*, vol. 21, pp. 233–255, 2002.

[3] E. Glassman and R. Tedrake, "A quadratic regulator-based heuristic for rapidly exploring state space," in *IEEE International Conference on Robotics and Automation*, 2010.

[4] A. Perez, R. Platt-Jr, G. Konidaris, L. Kaelbling, and T. Lozano-Perez, "LQR-RRT*: Optimal sampling-based motion planning with automatically derived extension heuristics," in *IEEE International Conference on Robotics and Automation*, 2012.

[5] Y. Tassa, T. Erez, and E. Todorov, "Synthesis and stabilization of complex behaviors through online trajectory optimization," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.

[6] S. Vijayakumar and S. Schaal, "Incremental online learning in high dimensions," *Neural Computation*, vol. 17, pp. 2602–2634, 2005.

[7] M. Bharatheesha, W. Caarls, W. Wolfslag, and M. Wisse, "Distance metric approximation for state space rrts using supervised learning," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014.

[8] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, 2009.