Appendix : Active Area Coverage from Equilibrium

Ian Abraham, Ahalya Prabhakar, and Todd D. Murphey

Department of Mechanical Engineering, Northwestern University,
2145 Sheridan Road, Evanston, IL 60208, USA
i-abr@u.northwestern.edu, a-prabhakar@u.northwestern.edu,
t-murphey@northwestern.edu

Here we provide a detailed description of the implementation of our algorithm including dynamic models used and weight parameters.

1 Shape Estimation Parameters

Dynamics and Equilibrium Policy

In this example, we used cart double pendulum with dynamics given in [1] with a sampling rate of 500 Hz. An approximate dynamic model is used by linearizing the nonlinear dynamics about the inverted equilibrium. The stabilizing task J_{task} is given as

$$J_{\text{task}} = \int_{t_i}^{t_t+T} x^\top \mathbf{Q} x + u^\top \mathbf{R} u dt \tag{1}$$

where

$$\mathbf{Q} = \text{diag}(0, 50, 50, 50, 700, 700) \qquad \mathbf{R} = 0.01$$

are the weights for the state and control respectively, and $T = 0.2s$ is the time horizon. A linear quadratic regulator (LQR) controller is computed using $\mathbf{Q}$ and $\mathbf{R}$ and the approximate dynamics linearized about $x_0 = \mathbf{0} \in \mathbb{R}^6$ and $u_0 = 0$.

KL-Divergence and Modeling Parameters

A Gaussian process is used with a radial basis function where we solve for the characteristic lengths by maximizing the log likelihood with respect to the data set. We fix the data set to have a memory of 100 points which we prune based on an importance measure. The Σ-approximate time-averaged statistics are calculated using $\Sigma = 0.1 \times \mathbf{I} \in \mathbb{R}^2$ where the search space is the global $x - y$ position. Here we recall the whole trajectory into the past making $t_r = 0.2$. The regularization parameter R that bounds $\mu_\star$ to $\mu(x)$ is given as $R = 20$.

2 Quadrotor State-Space Exploration

Dynamics and Equilibrium Policy

In this example we use a 22-degree of freedom quadcopter defined in [2] with sampling rate of 200 Hz. The states for the quadcopter are given by

$$x = [g, \omega, v]^\top$$

where $g \in SE(3)$ is the transformation matrix and $\omega, v \in \mathbb{R}^3$ are the angular and linear body velocities. The approximate dynamics are computed using a linearization about $x_0 = [\mathbf{I}, \mathbf{0}, \mathbf{0}]^\top$, $u_0 = \mathbf{0} \in \mathbb{R}^4$. A LQR policy is generated using the objective defined in 1 where

$$\mathbf{Q} = 10 \times \mathbf{I} \in \mathbb{R}^{22} \qquad \mathbf{R} = 0.1 \times \mathbb{R}^4$$

and the elements in $\mathbf{Q}$ corresponding to the height is set to 100. We set the time horizon as $T = 0.1s$. We specify a decaying weight on the KL-divergence measure as 100γ where $\gamma = 0.995^{i+1}$ where i is the i^{th} iteration of the algorithm.

KL-Divergence and Modeling Parameters

The Gaussian process models the body angular and linear velocity and the interaction with the control input. A radial basis function is used for the Gaussian process with parameters $\Sigma = 0.01 \times \mathbf{I} \in \mathbb{R}^{10}$ and a fixed data-set size of 80 points. The Σ-approximate time-averaged statistics are calculated using $\Sigma = 0.1 \times \mathbf{I} \in \mathbb{R}^{22}$ where the search space is the state-space. Here we recall the $t_r = 0.1s$ in the past trajectory.. The regularization parameter R that bounds $\mu_\star$ to $\mu(x)$ used is $R = 10^4 \times \mathbf{I} \in \mathbb{R}^4$.

3 Half-Cheetah Stable Exploration

Dynamics and Equilibrium Policy

In this example, we use the half-cheetah dynamical system defined in the Roboschool environment [3] with 22 dimensional state and 6 dimensional control input space. A linear policy is generated using [4] which maintains upright posture for the half-cheetah robot. We collected the state-action data during training of the equilibrium policy and created an approximate linear dynamic model using least squares. The system is sampled at $0.01s$ intervals and and horizon of $T = 0.1s$ is used. The task objective is defined as

$$J_{\text{task}} = \int_{t_i}^{t_i+T} x_{\text{height}}^2 + 0.01 x_{\text{joints}}^\top x_{\text{joints}} + x_{\text{pitch}}^2 + 0.01 u^\top u dt$$

which encourages staying upright.

KL-Divergence and Modeling Parameters

The Gaussian process model used to predict the dynamics of the half-cheetah used a radial basis function with an empirically determined variance of $\Sigma = 200 \times \mathbf{I} \in \mathbb{R}^{28}$. We fixed the data set to 40 data points which is pruned as more informative data is collected. The Σ-approximate time-averaged statistics are calculated using $\Sigma = 0.1 \times \mathbf{I} \in \mathbb{R}^{28}$. The time remembered into the past trajectory is defined as $t_r = 0.2s$. A weight of 100 is applied to the KL-divergence weight in the objective. Last, the regularization parameter used to bound $\mu_\star$ is defined as $R = 0.1 \times \mathbf{I} \in \mathbb{R}^6$.

Forward Running

During the testing phase of the learned dynamic model of the half-cheetah we used model-predictive control to maximize the forward velocity of the half-cheetah. To achieve this, we minimized the objective

$$J_{task} = \int_t^{t+T} x_{\text{height}}^2 - 2.0 x_{\text{forward vel}} + 0.01 u^\top u + x_{\text{pitch}}^2 dt \tag{2}$$

where $T = 0.2s$. The objective was minimized using [5] for both the learned model using motor babble and the learned model using our method for active exploration. A regularization parameter is set for the control input as $R = 1000 \times \mathbf{I} \in \mathbb{R}^6$.

Bibliography

[1] Wei Zhong and Helmut Rock. Energy and passivity based control of the double inverted pendulum on a cart. In *IEEE International Conference on Control Applications*, pages 896–901, 2001.

[2] Taosha Fan and Todd Murphey. Online feedback control for input-saturated robotic systems on Lie groups. In *Proceedings of Robotics: Science and Systems*, June 2016. doi: 10.15607/RSS.2016.XII.027.

[3] Oleg Klimov and John Shulman. Roboschool. `https://github.com/openai/roboschool`, 2017.

[4] Horia Mania, Aurelia Guy, and Benjamin Recht. Simple random search provides a competitive approach to reinforcement learning. *arXiv preprint arXiv:1803.07055*, 2018.

[5] A. R. Ansari and T. D. Murphey. Sequential action control: Closed-form optimal control for nonlinear and nonsmooth systems. *IEEE Transactions on Robotics*, 32 (5):1196–1214, 2016.