Policy Decomposition: Approximate Optimal Control with Sub-optimality Measure

Ashwin Khadke and Hartmut Geyer

Carnegie Mellon University

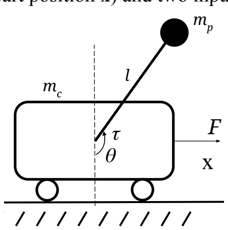
THE ROBOTICS INSTITUTE

Abstract

Dynamic programming (DP) is often applied to solve optimal control problems in robotics. However, owing to the curse of dimensionality only approximate DP methods are computationally tractable for complex systems with many degrees of freedom. Several such approximations have been proposed, relying on either local search or state-space reductions [1-3]. However, these strategies do not yield an associated measure to gauge the sub-optimality of the resulting control. We instead propose and explore policy decomposition, an alternative approximation strategy that includes such a measure. The measure indicates the similarity of the resulting closed-loop behavior to the one of the optimal control solution. We apply our ideas in computing policies for controlling the cart-pole system. Our approach identifies decompositions that require 1/10th of the compute time while minimally sacrificing on optimality.

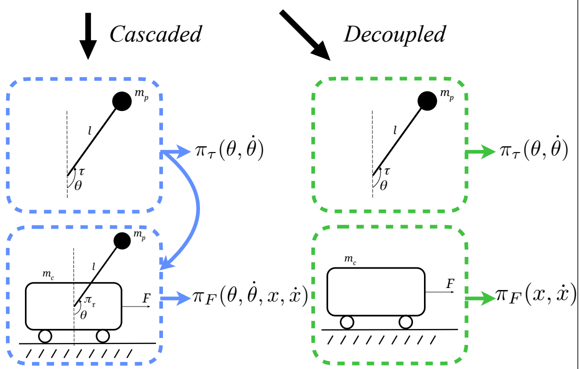
Introduction

We propose to generate control policies for complex systems based on cascaded, lower dimensional problems whose order is identified algorithmically and maximizes the similarity of the resulting control policy to the one that would be obtained if the complex system was computationally tractable. Consider, for example, designing a control policy to swing up a pole on a cart while moving the cart to a goal. This system has two degrees of freedom (pole angle θ and cart position $\mathbf{x}$) and two inputs (pole torque τ and cart force F).



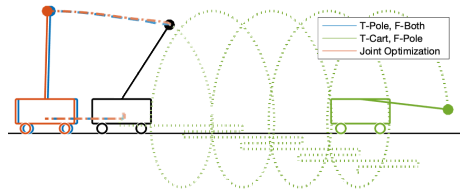
$$\frac{d}{dt}\begin{bmatrix}\theta\\ \dot\theta\end{bmatrix} = \begin{bmatrix}\dot\theta\\ G_\theta(\theta,\dot\theta,F,\tau)\end{bmatrix}$$

$$\frac{d}{dt}\begin{bmatrix}x\\ \dot x\end{bmatrix} = \begin{bmatrix}\dot x\\ G_x(\theta,\dot\theta,F,\tau)\end{bmatrix}$$



Policy Decomposition

For a dynamical system, many decompositions can be found that make control synthesis much more tractable. But the quality of resulting policies for the original system differs considerably among them.



$$\dot{\boldsymbol{x}} = f(\boldsymbol{x},\boldsymbol{u})$$
$$J = \int_0^\infty c(\boldsymbol{x}(t),\boldsymbol{u}(t))dt$$

Cascaded

$$\dot{\boldsymbol{x}}_{1:i} = f_{1:i}([\boldsymbol{x}_{1:i},\boldsymbol{x}^d_{(i+1):m}],\ [\pi_{\boldsymbol{u}_{1:(i-1)}},u_i,\mathbf{0}_{(m-i+1)}])$$
$$c_i = c([\boldsymbol{x}_{1:i},\boldsymbol{x}^d_{(i+1):m}],\ [\pi_{\boldsymbol{u}_{1:(i-1)}},u_i,\mathbf{0}])$$

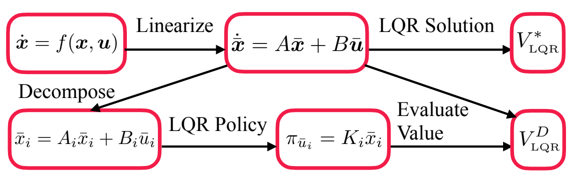
Decoupled

$$\dot{\boldsymbol{x}}_i = f_i([\boldsymbol{x}^d_{1:(i-1)},x_i,\boldsymbol{x}^d_{(i+1):m}],\ [\mathbf{0}_{(i-1)},u_i,\mathbf{0}_{(m-i+1)}])$$
$$c_i = c([\boldsymbol{x}^d_{1:(i-1)},x_i,\boldsymbol{x}^d_{(i+1):m}],\ [\mathbf{0}_{(i-1)},u_i,\mathbf{0}_{(m-i+1)}])$$

We propose two measures to predict the sub-optimality of these decompositions. These measures are:

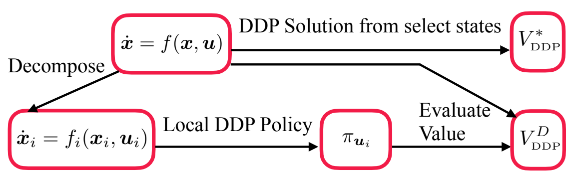
- Indicative of the closed-loop behavior of the original and decomposed problem
- Easier to compute than computing the policy

LQR Measure



$$\tilde{\mathrm{Err}}_{\mathrm{LQR}}(D) = \int_{\mathcal{S}}(V^D_{\mathrm{LQR}}(\boldsymbol{x}) - V^*_{\mathrm{LQR}}(\boldsymbol{x}))d\boldsymbol{x}$$

DDP Measure



$$\tilde{\mathrm{Err}}_{\mathrm{DDP}}(D) = \sum_{\boldsymbol{x}\sim\mathcal{S}}(V^D_{\mathrm{DDP}}(\boldsymbol{x}) - V^*_{\mathrm{DDP}}(\boldsymbol{x}))$$

LQR Measure

- ✔ Accounts for linearized system dynamics
- ✔ Accounts for objective function
- ✔ Computationally inexpensive
- ✗ Agnostic to bounds on control inputs
- ✗ Valid near the point of linearization

DDP Measure

- ✔ Accounts for full system dynamics
- ✔ Accounts for objective function
- ✔ Accounts for the bounds on control inputs
- ✗ Computationally expensive
- ✗ Provides no guarantees for values estimates

Results

decomposition	$\tilde{\mathrm{Err}}_{\mathrm{LQR}}$	$\tilde{\mathrm{Err}}_{\mathrm{DDP}}$	$\mathrm{r}_{\mathrm{DDP}}$	$\mathrm{r}_{\mathrm{LQR}}$	Err	r	time
F-Cart, τ-Both	0.21	0.36	2	2	0.34	2	7600
τ-Pole, F-Both	0.055	0.004	1	1	0.02	1	2935
τ-Cart, F-Both	2.9	27	5	5	6.75	5	3887
F-Pole, τ-Both	0.61	13	4	4	0.92	4	3758
F-Cart, τ-Pole	0.35	0.37	3	3	0.36	3	22
τ-Cart, F-Pole	∞	∞	6	6	15	6	18
Joint	0	0	-	-	0	-	85500

Table 1: $m_c = m_p$

decomposition	$\tilde{\mathrm{Err}}_{\mathrm{LQR}}$	$\tilde{\mathrm{Err}}_{\mathrm{DDP}}$	$\mathrm{r}_{\mathrm{DDP}}$	$\mathrm{r}_{\mathrm{LQR}}$	Err	r	time
F-Cart, τ-Both	0.087	0.117	2	2	0.086	3	5784
τ-Pole, F-Both	0.009	0.0005	1	1	0.006	1	3863
τ-Cart, F-Both	4.4	45	5	5	8.75	5	3978
F-Pole, τ-Both	0.92	1.62	4	4	0.5	4	6616
F-Cart, τ-Pole	0.1	0.118	3	3	0.085	3	86
τ-Cart, F-Pole	∞	251	6	6	13	6	41
Joint	0	0	-	-	0	-	99000

Table 2: $m_c \gg m_p$

Conclusions and Future Work

- ✔ Introduced strategies to decompose an optimal control problem
- ✔ Introduced closed-loop measures to gauge the quality of different decompositions
- Derive confidence bounds for the sub-optimality measures
- Extend the analysis to hybrid systems

References

[1] E. Todorov and W. Li, "A generalized iterative lqg method for locally-optimal feedback control of constrained nonlinear stochastic systems" ACC' 2005

[2] R. Tedrake, I. R. Manchester, M. Tobenkin, and J. W. Roberts, "Lqr-trees: Feedback motion planning via sums-of-squares verification" IJRR' 2010

[3] A. Gorodetsky, S. Karaman, and Y. M. Marzouk, "Efficient high-dimensional stochastic optimal motion control using tensor-train decomposition" RSS' 2015