Policy Decomposition: Approximate Optimal Control with Sub-optimality Measure

Ashwin Khadke and Hartmut Geyer

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 \( \theta \) and cart position \( x \)) and two inputs (pole torque \( \tau \) and cart force \( F \)).

\[
\begin{array}{c}
\dot{x} = f(x, u) \\
J = \int_0^T c(x(t), u(t))dt
\end{array}
\]

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

\[
\begin{align*}
\dot{x} &= Ax + Bu & \text{LQR Solution} \\
\text{Decompose} & & \text{Evaluate Value} \\
\pi \theta &= A_t \pi + B_t u_t & V_{\text{LQR}}^t \\
\text{Err}_{\text{LQR}} &= \int_t^T (V_{\text{LQR}}^t(x) - V_{\text{LQR}}^0(x))dx
\end{align*}
\]

DDP Measure

\[
\begin{align*}
\dot{x} &= f(x, u) & \text{DDP Solution from select states} \\
\text{Decompose} & & \text{Evaluate Value} \\
\pi \theta &= f_1(x, u) & V_{\text{DDP}}^t \\
\text{Err}_{\text{DDP}} &= \sum_{x \in S} (V_{\text{DDP}}^t(x) - V_{\text{DDP}}^0(x))
\end{align*}
\]

Results

<table>
<thead>
<tr>
<th>decomposition</th>
<th>( \text{Err}_{\text{LQR}} )</th>
<th>( \text{Err}_{\text{DDP}} )</th>
<th>( r_{\text{LQR}} )</th>
<th>( r_{\text{DDP}} )</th>
<th>( \text{Err} )</th>
<th>( r )</th>
<th>( t ) time</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F )-Cart, ( \tau )-Both</td>
<td>0.21</td>
<td>0.36</td>
<td>2</td>
<td>2</td>
<td>0.34</td>
<td>2</td>
<td>7600</td>
</tr>
<tr>
<td>( \tau )-Pole, ( F )-Both</td>
<td>0.055</td>
<td>0.004</td>
<td>1</td>
<td>1</td>
<td>0.02</td>
<td>1</td>
<td>2035</td>
</tr>
<tr>
<td>( F )-Cart, ( \tau )-Both</td>
<td>2.9</td>
<td>27</td>
<td>5</td>
<td>5</td>
<td>6.75</td>
<td>5</td>
<td>3878</td>
</tr>
<tr>
<td>( F )-Pole, ( \tau )-Both</td>
<td>0.61</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>0.92</td>
<td>4</td>
<td>3758</td>
</tr>
<tr>
<td>( F )-Cart, ( \tau )-Pole</td>
<td>0.35</td>
<td>0.37</td>
<td>3</td>
<td>3</td>
<td>0.36</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>( \tau )-Cart, ( F )-Pole</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>6</td>
<td>6</td>
<td>15</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>( \tau )-Pole, ( F )-Both</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>6</td>
<td>6</td>
<td>15</td>
<td>6</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 1: \( m_c = m_p \)

Conclusions and Future Work

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

References