Integrating Models and Data for Robust Manipulation with and around people

Siddhartha Srinivasa
Boeing Endowed Professor
Personal Robotics Lab
Computer Science and Engineering
University of Washington
Manipulation
Personal Robotics Lab
Carnegie Mellon University
Physics-based Manipulation
Harness the Mechanics of Manipulation to Funnel Uncertainty
Quasi-Static Pushing

How much should the robot know?

- Object mass? No.
- Object-surface friction? No.
- Object pressure distribution? Pick conservatively.
- Finger-object friction? Pick conservatively.
Analytical Capture Regions
Addressing Object Pose Uncertainty

Vision

Reported pose

Uncertainty Region

Is included in capture region of a G?

Physics-based Manipulation
Exploiting physics to manipulate objects

Autonomous control of complex dynamical systems
Global models are often only partially correct
<table>
<thead>
<tr>
<th></th>
<th>Global, analytical models</th>
<th>Model-based RL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>Inaccurate</td>
<td></td>
</tr>
<tr>
<td><strong>Policy</strong></td>
<td>No uncertainty</td>
<td></td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>No data collection</td>
<td></td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td>Fast convergence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Global, analytical models</td>
<td>Locally learned models</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td>Inaccurate</td>
<td>Locally more accurate, captures uncertainty</td>
</tr>
<tr>
<td><strong>Policy</strong></td>
<td>No uncertainty</td>
<td>Uncertainty-aware</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>No data collection</td>
<td>Requires data collection</td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td>Fast convergence</td>
<td>Slow convergence</td>
</tr>
<tr>
<td></td>
<td>Global, analytical models</td>
<td>Hybrid model</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Model</td>
<td>Inaccurate</td>
<td>Globally available, Locally accurate, captures uncertainty</td>
</tr>
<tr>
<td>Policy</td>
<td>No uncertainty</td>
<td>Uncertainty-aware</td>
</tr>
<tr>
<td>Data</td>
<td>No data collection</td>
<td>Use data to correct analytical models</td>
</tr>
<tr>
<td>Training</td>
<td>Fast convergence</td>
<td>Moderate</td>
</tr>
</tbody>
</table>
Learn the **residual** between simulation and reality

Gaussian Process Regression

Gilwoo Lee
<table>
<thead>
<tr>
<th></th>
<th>Global, analytical models</th>
<th>Hybrid model</th>
<th>Locally learned models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>Inaccurate</td>
<td>Globally available, Locally accurate, captures uncertainty</td>
<td>Locally more accurate, captures uncertainty</td>
</tr>
<tr>
<td><strong>Policy</strong></td>
<td>No uncertainty</td>
<td>Uncertainty-aware</td>
<td>Uncertainty-aware</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>No data collection</td>
<td>Use data to correct analytical models</td>
<td>Requires data collection</td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td>Fast convergence</td>
<td>Moderate</td>
<td>Slow convergence</td>
</tr>
</tbody>
</table>
Policy Search: 
Iterative Linear Quadratic Regulator

Linear dynamics,
Quadratic cost,
Iterative local improvements

Policy Search:
Iterative Linear Quadratic Regulator

Linear dynamics,
Quadratic cost,
Iterative local improvements

Policy Search: Iterative Linear Quadratic Regulator

Linear dynamics,
Quadratic cost,
Iterative local improvements
Policy Search: Iterative Linear Quadratic Gaussian Control

Linear stochastic dynamics,
Quadratic cost,
Iterative local improvements

Incorporate model uncertainty into the Iterative Linear Quadratic Regulator

Robust Iterative Linear Quadratic Regulator
Policy Search: Iterative Linear Quadratic Gaussian Control

\[ x_{t+1} = f(x_t, u_t) + \zeta(x_t, u_t), \quad \zeta \sim N(0, \Gamma) \]

Bellman update:

\[ V(x_t) = \min_{u_t} \left[ l(x_t, u_t) + E[V'(f(x_t, u_t))] \right] \]

\[ Q(\delta x, \delta u) = Q_x \delta x + Q_u \delta u + \frac{1}{2} (\delta x^T Q_{xx} \delta x + \delta u^T Q_{uu} \delta u + 2 \delta x^T Q_{ux} \delta u) \]

\[ Q_x = l_x + E [ f_x^T V' x ] = l_x + E [ (f + \zeta)_x^T V' x ] = l_x + f_x^T V' x \]
\[ Q_u = l_u + E [ f_u^T V' u ] \]
\[ Q_{xx} = l_{xx} + E [ f_{xx} V' x ] = l_{xx} + E [ (f + \zeta)_x^T V' x ] = l_{xx} + f_{xx}^T V' x + E [ \zeta_x^T V' x \zeta_x ] \]
\[ Q_{uu} = l_{uu} + E [ f_{uu}^T V' u ] \]
\[ Q_{ux} = l_{ux} + E [ f_{ux}^T V' u ] \]

Todorov, Emanuel, and Weiwei Li. “A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems.”

Policy Search: Robust ILQG

\[ x_{t+1} = f(x_t, u_t) + \zeta(x_t, u_t) \]
\[ = f_{\text{global}}(x_t, u_t) + \mu(x_t, u_t) + \xi(x_t, u_t) + \zeta(x_t, u_t) \]
\[ \zeta \sim N(0, \Gamma), \quad \xi \sim N(0, \Sigma) \]
Policy Search:
Robust ILQG

\[ x_{t+1} = f(x_t, u_t) + \zeta (x_t, u_t) \]
\[ = f_{\text{global}}(x_t, u_t) + \mu (x_t, u_t) + \xi (x_t, u_t) + \zeta (x_t, u_t) \]
\[ \zeta \sim N(0, \Gamma), \; \xi \sim N(0, \Sigma) \]

Bellman update:
\[ V(x_t) = \min_{u_t} \left[ l(x_t, u_t) + \mathbb{E}[V'(f(x_t, u_t))] \right] \]
\[ Q(\delta x, \delta u) = Q_x \delta x + Q_u \delta u + 1/2(\delta x^T Q_{xx} \delta x + \delta u^T Q_{uu} \delta u + 2 \delta x^T Q_{ux} \delta u) \]

\[ Q_x = l_x + \mathbb{E}[f_x^T V'_x] = l_x + \mathbb{E}[(f_{\text{global}}+ \mu + \xi + \zeta)_x^T V'_x] \]
\[ Q_u = l_u + \mathbb{E}[f_u^T V'_x] \]
\[ Q_{xx} = l_{xx} + \mathbb{E}[f_{xx}^T V'_{xx}] = l_{xx} + \mathbb{E}[(f_{\text{global}}+ \mu + \xi + \zeta)_x^T V'_{xx}(f_{\text{global}}+ \mu + \xi + \zeta)_x] \]
\[ Q_{uu} = l_{uu} + \mathbb{E}[f_{uu}^T V'_{xx}] \]
\[ Q_{ux} = l_{ux} + \mathbb{E}[f_{ux}^T V'_{xx}] \]
Explicitly correcting model bias and incorporating the correction as well as our uncertainty of the correction in optimal control enables lifelong learning of the system and robust control under uncertainty.

Our algorithm iterates over simulation-based optimal control, real-world data collection, and model learning, as illustrated in Figure 1. Starting from a potentially incorrect model given by the simulator, we obtain a control policy, with which we collect data in the real world. This data feeds into model learning, during which we correct model bias and estimate our uncertainty of the correction. Both the correction and its uncertainty are incorporated into computing a robust optimal control policy, which then gets used to collect more data.

Our approach improves any simulator beyond the scope of its model space to match real-world observations and produces an optimal control policy robust to model uncertainty and multiplicative noise. The improved simulator uses previous real-world observations to infer the true model when it explores previously visited space, but when it encounters a new region, it relies on the simulator’s original model. Due to this hybrid nature, our algorithm shows faster convergence to the optimal policy than a pure data-driven approach or a pure simulation-based approach. Moreover, as it permanently improves the simulator, it shows even faster convergence in new tasks in similar task domain.

Related Work

Most model-based reinforcement learning has both model learning (system identification) and policy optimization components. The data for a model comes either from real world or simulation, and is combined to construct a model via nonlinear function approximators such as Locally Weighted Regression, Gaussian Processes, or Neural Networks. Once the model is built, a typical policy gradient method computes the derivatives of the cost function with respect to control parameters.

If an analytic model is given, e.g., via equations of motion or as a simulator, one can use classical optimal control techniques such as Differential Dynamic Programming (DDP), which compute a reference trajectory as well as linear feedback control law. For robustness, Iterative Linear Quadratic Gaussian Control (ILQG) or H-1 Control can be used to incorporate multiplicative noise. Variants of have been used to generate guiding policies for data-driven RL methods.

Recently, there have been some attempts to combine DDP or ILQG with data-driven models by replacing analytical models with locally linear models learned by Gaussian Processes or Neural Networks.

We consider black-box simulators as analytical models, as derivatives can be taken by finite differencing.
## Comparison

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Global, analytical models</th>
<th>Hybrid model</th>
<th>Locally learned models</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILQG</td>
<td>Assesses perfect model</td>
<td>GP-ILQG</td>
<td>Probabilistic-DDP</td>
</tr>
<tr>
<td>Initialized with random policy</td>
<td>Initialized with ILQG policy &amp; random trajectories</td>
<td>Random policy &amp; Demonstrated trajectories</td>
<td></td>
</tr>
</tbody>
</table>

---


<table>
<thead>
<tr>
<th></th>
<th>Global, analytical models</th>
<th>GP-ILQG</th>
<th>Locally learned models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>Inaccurate</td>
<td>Globally available, Locally accurate, captures uncertainty</td>
<td>Locally more accurate, captures uncertainty</td>
</tr>
<tr>
<td><strong>Policy</strong></td>
<td>No uncertainty</td>
<td>Uncertainty-aware</td>
<td>Uncertainty-aware</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>No data collection</td>
<td>Use data to correct analytical models</td>
<td>Requires data collection</td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td>Fast convergence</td>
<td>Moderate</td>
<td>Slow convergence</td>
</tr>
</tbody>
</table>

Optimal Control

Model-based RL
Limitations

• Scaling of GPs
  Use any learner that reports uncertainty

• Robustness to model error discourages exploration
  Encourage exploration, e.g. posterior sampling
Takeaways

• Models are nice, but not perfect
  Bootstrap models with data

• Strive for minimalism
  Task-driven model learning
Integrating Models and Data for Robust Manipulation with and around people

Siddhartha Srinivasa
Boeing Endowed Professor
Personal Robotics Lab
Computer Science and Engineering
University of Washington