State Constrained Stochastic Optimal Control Using LSTMs

NYU TANDON SCHOOL OF ENGINEERING

Bolun Dai, Prashanth Krishnamurthy, Andrew Papanicolaou, Farshad Khorrami

Problem Setup

The system considered is described as a stochastic differential equation (SDE)

$$dx(t) = f(x(t),t)dt + G(x(t),t)u(t)dt + \Sigma(x(t),t)dw(t).$$

The objective is to find a control sequence that minimizes the following cost

$$J^u(x,t) = \mathbb{E}\Big[g(x(T)) + \int_t^T \Big(q(x(s)) + \frac{1}{2}u(s)^T R u(s)\Big) ds \Big| x(t) = x\Big]$$

with state constraints

$$c_{\min} \le c_s(x) \le c_{\max}$$

and control saturation

$$u \in \mathcal{U} = \{u \mid |u_i| \le U_{i,\max}\}.$$

State & Control Constraint

The control is saturated as

$$u^*(x,t) = U_{\max} * \mathrm{sig}(-R^{-1}G^T(t,x)V_x).$$

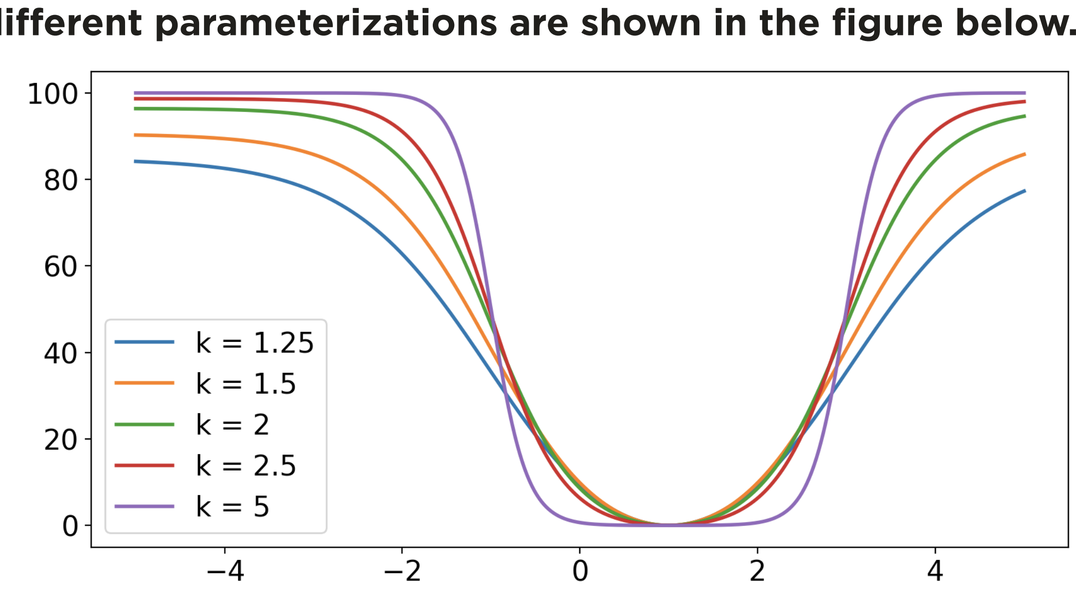
The corresponding control cost is

$$S_i(u_i) = c_i \int_0^{u_i} \mathrm{sig}^{-1}\Big(\frac{v}{U_{i,\max}}\Big) dv.$$

The state constraint is applied via a penalty function

$$p(x) = \frac{L}{1+e^{-k(c_s(x)-c_{\max})}} - \frac{L}{1+e^{-k(c_s(x)-c_{\min})}} + L - \frac{2L}{1+e^{-k(\mu - c_{\max})}}.$$

For a state constraint of [-1, 3], the penalty function with different parameterizations are shown in the figure below.



Taking both state constraints and control saturation into consideration, the overall cost function has the form

$$\mathbb{E}\Bigg[g(x(T)) + \int_t^T \Big(q(x(s)) + p(x(s)) + \sum_{i=1}^m S_i(u_i(s))\Big) ds \Big| x(t) = x\Bigg].$$

Adaptive Update Scheme

To ensure numerical stability, we use the square root of state cost variance over a fixed number of iterations as the update threshold, and gradually harden the penalty function p(x). Since the state cost variance would never decrease to zero, we also set a minimum value for the threshold.

$$k \leftarrow k + \delta$$
$$\delta \leftarrow \delta - \Delta_\delta$$
$$\beta \leftarrow \gamma\beta$$
$$\gamma \leftarrow \gamma + \Delta$$

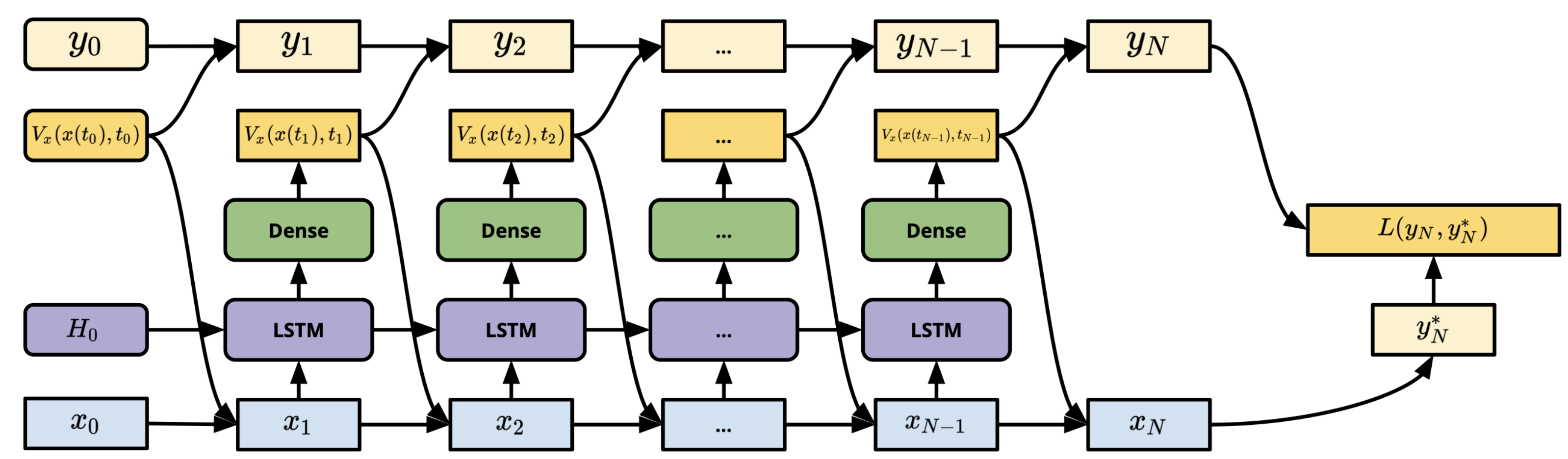
Deep FBSDE

The problem under the updated cost function can be recast as a forward-backward stochastic differential equation (FBSDE) as shown on the right, where V_x is the partial derivative of the value function w.r.t. the state, Φ represents learned weights, and the Hamiltonian is defined as

$$h(x, V_x, t, u^*) = q(x) + V_x^T G(x,t)u^*(x,t) + \sum_{i=1}^m S_i(u_i^*).$$

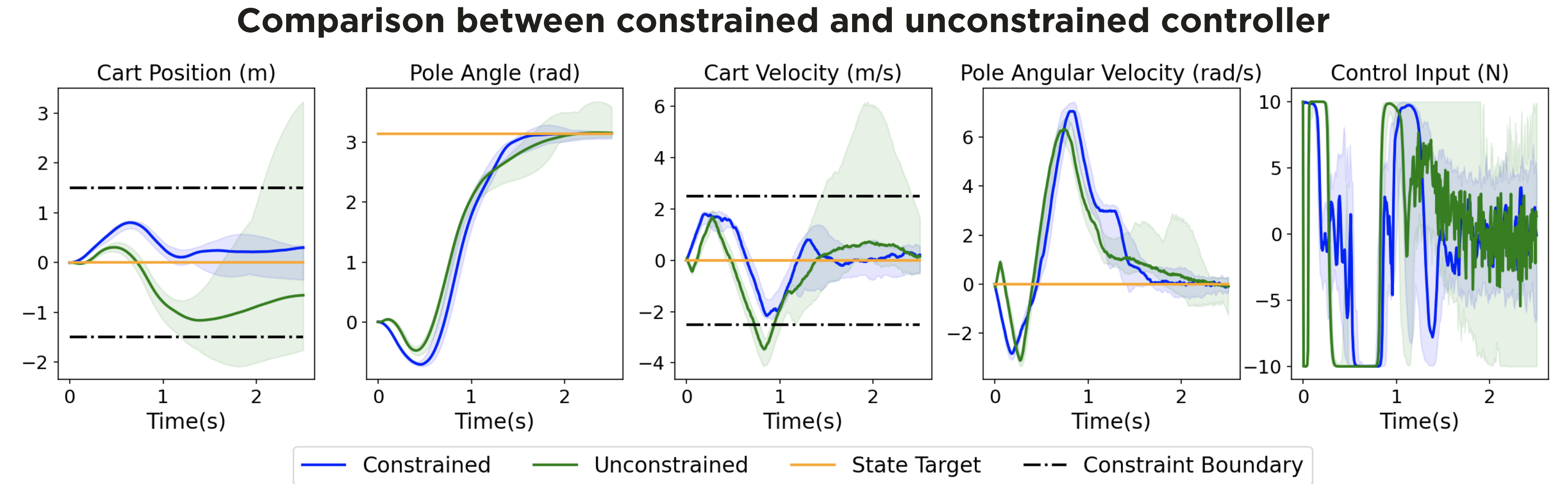
$$\begin{aligned}
dy(t) &= \Big(-h(x(t), V_x(x(t),t;\theta), t, u(t)) \\
&\quad + V_x^T(x(t),t;\theta)G(x(t),t)u(x(t),t)\Big)dt \\
&\quad + V_x^T(x(t),t;\theta)\Sigma(x(t),t)dw(t) \\
dx(t) &= \Big(f(x(t),t) + G(x(t),t)u(x(t),t)\Big)dt \\
&\quad + \Sigma(x(t),t)dw(t) \\
u(t) &= U_{\max} * \mathrm{sig}(-R^{-1}G^T(x(t),t)V_x(x(t),t;\theta) \\
y(0) &= V(\phi) \\
dy(0) &= V_x(\phi) \\
x(0) &= x_0.
\end{aligned}$$

The Deep FBSDE neural network architecture is shown below, at each time step the neural network estimates V_x. The neural network is optimized using a weight decayed L2 loss.

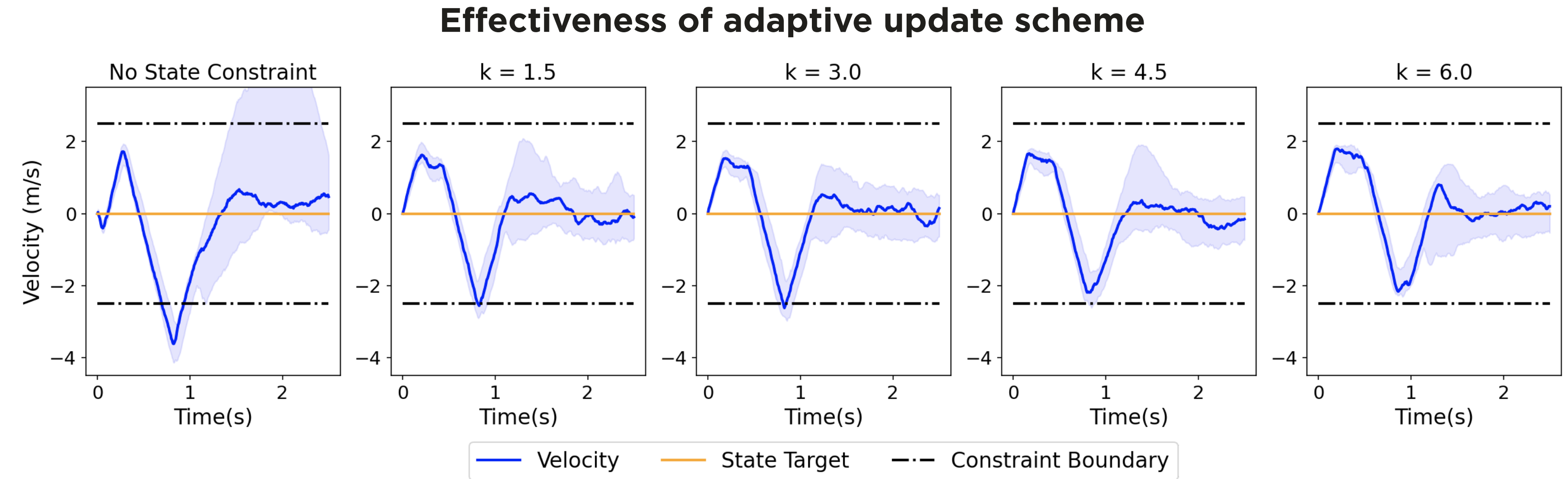


Experiments

Comparison between constrained and unconstrained controller



Effectiveness of adaptive update scheme



The presented experimental results are conducted on the cart-pole swing-up task. Two state constraint settings were tested: (i) constraining cart position and cart velocity; (ii) constraining the sum of kinetic and potential energy. We see that in both settings, the learned controller is able to respect the constraint boundaries.

Energy constraint comparison