Teaching the correct policy to ENMPC

Adaptive Economic NMPC and Reinforcement Learning

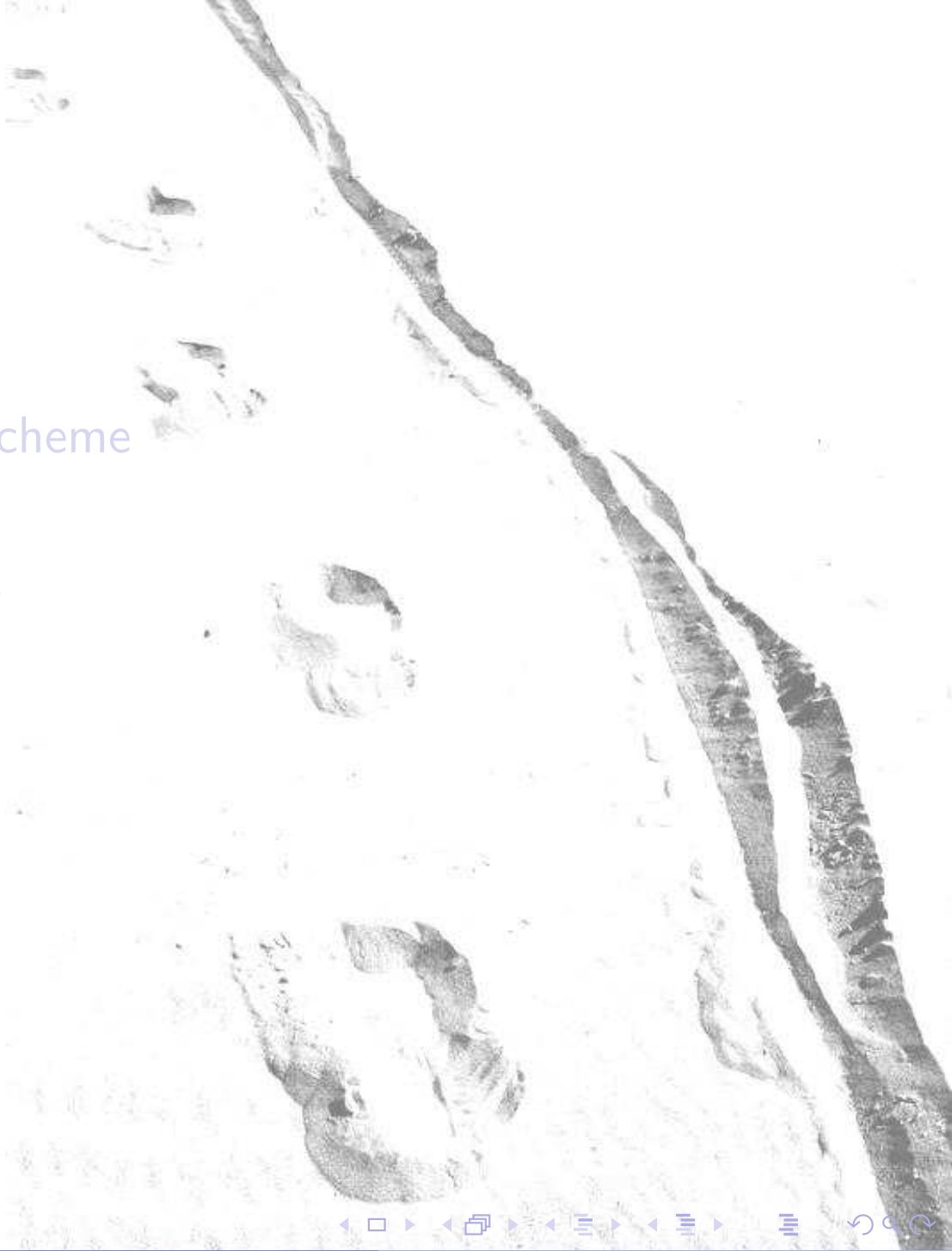
Sébastien Gros

Cybernetic, NTNU
Electrical Engineering, Chalmers

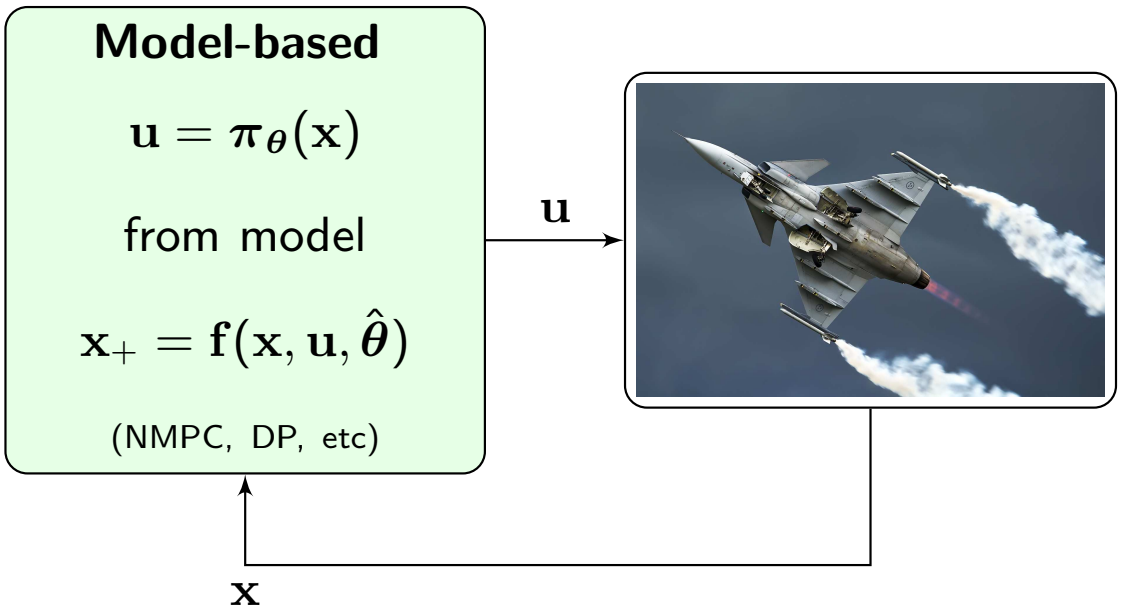
TUM

Outline

Outline

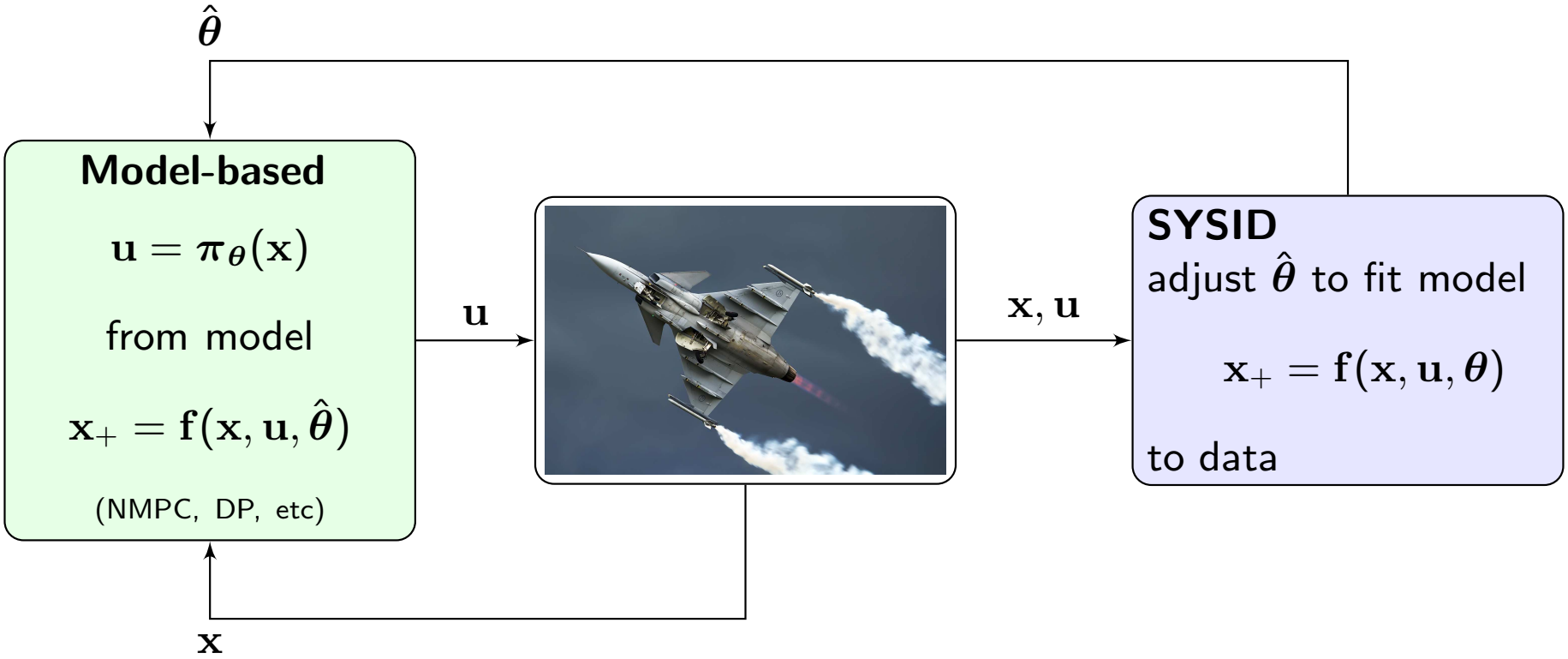


SYSID & Model-based optimal control



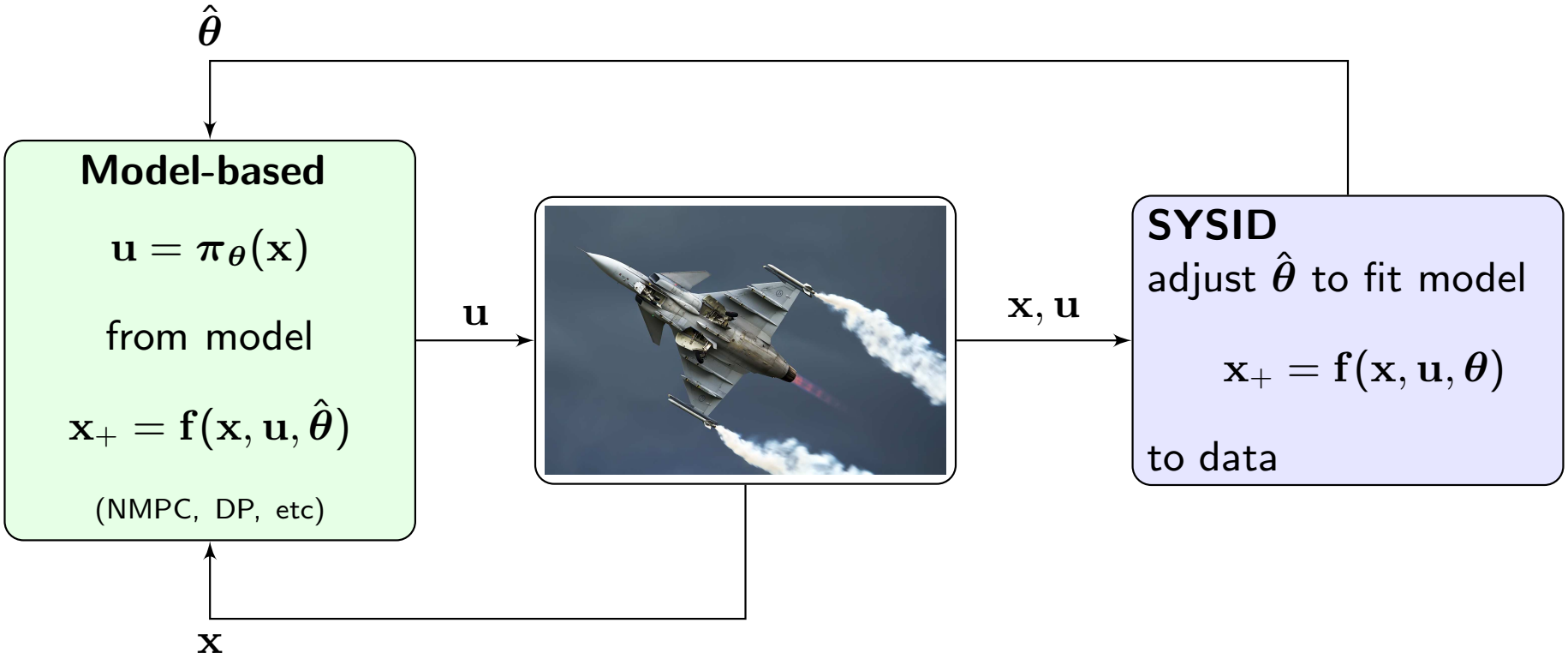
Assume we have a given performance index (i.e. a stage cost L to minimized)

SYSID & Model-based optimal control



Assume we have a given performance index (i.e. a stage cost L to minimized)

SYSID & Model-based optimal control



Assume we have a given performance index (i.e. a stage cost L to minimized)

Does this work?

- Not necessarily... problem: model may not able to capture the real system
- E.g. what is the best way of representing stochasticity using $\mathbf{f}$?
- Can degrade performance compared to keeping initial $\hat{\theta}$

Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

for given model $\mathbf{f}$.

Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$
$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$
$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

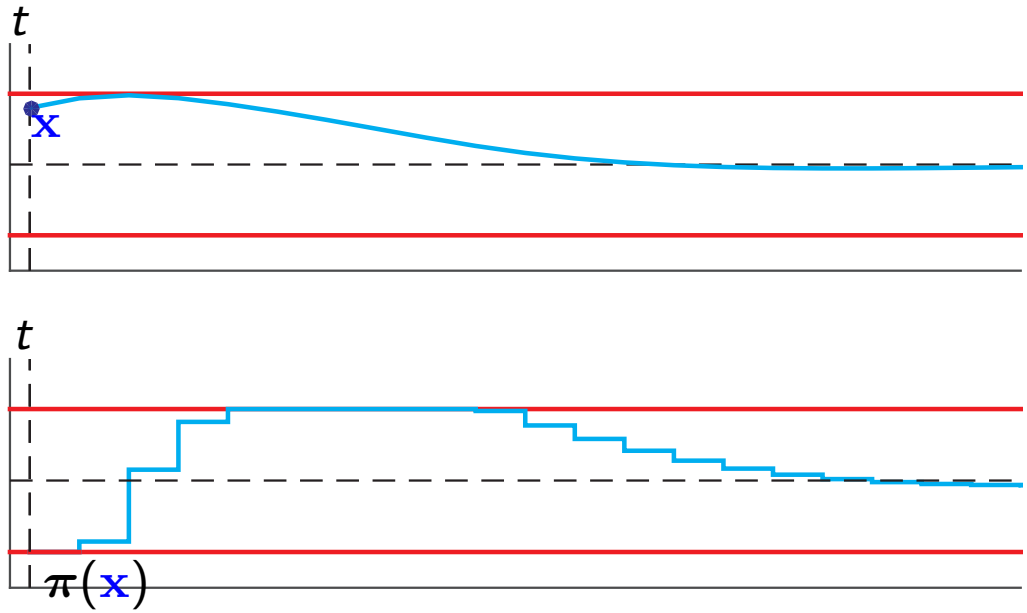
Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad \sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)$$
$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right)$$
$$\mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$
$$\mathbf{x}_0 = \mathbf{x}$$

for given model $\mathbf{f}$.



Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

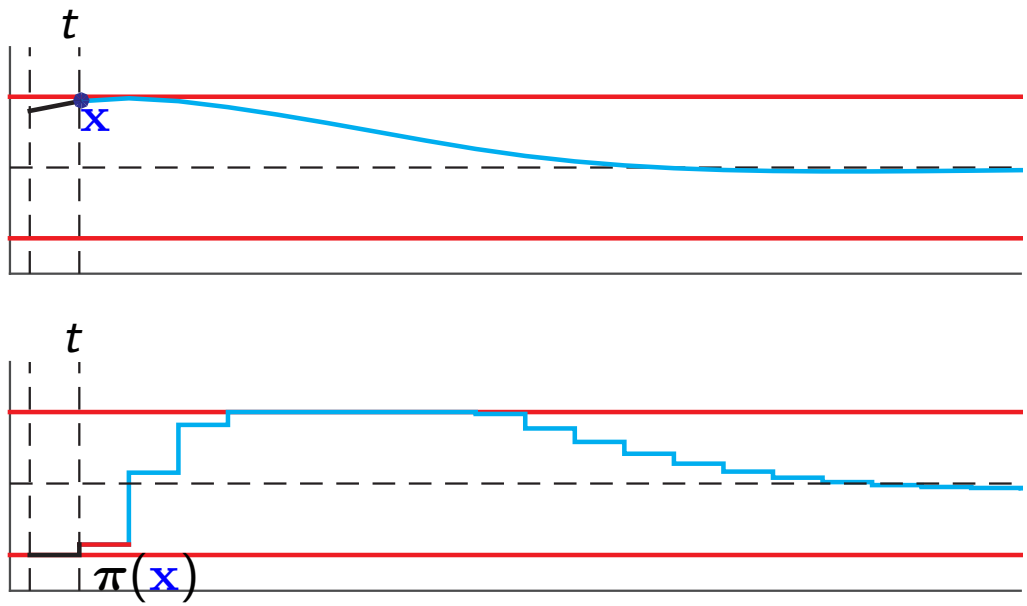
Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

for given model $\mathbf{f}$.



Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

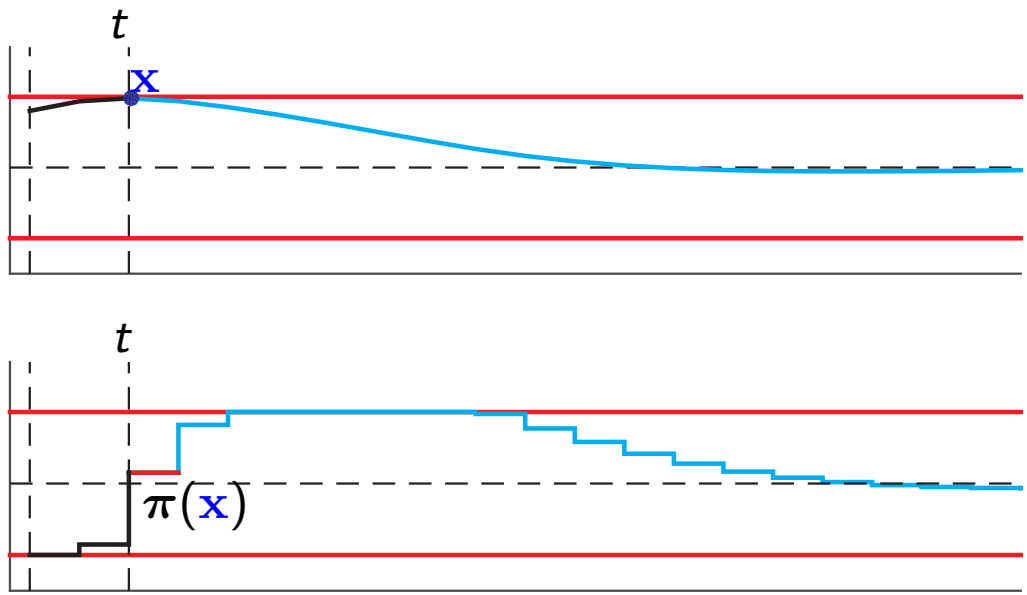
Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

for given model $\mathbf{f}$.



Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

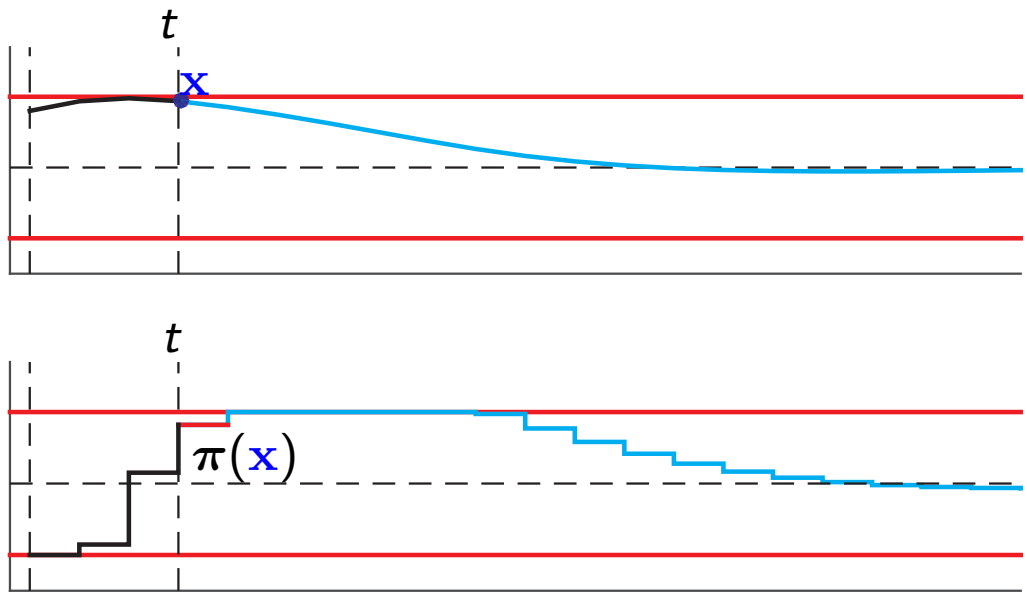
Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

for given model $\mathbf{f}$.



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$
$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$
$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

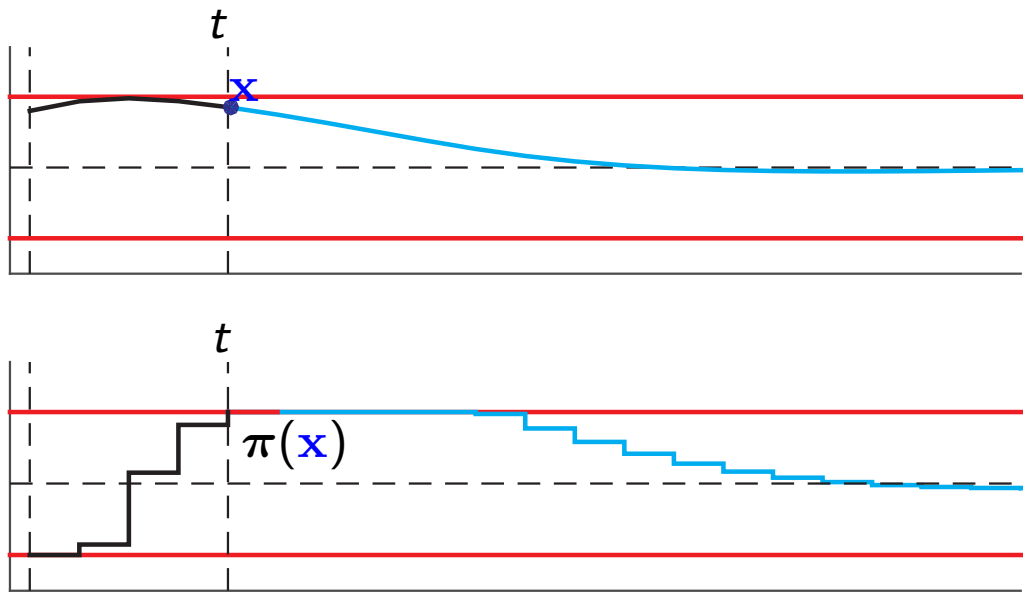
Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad \sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)$$
$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right)$$
$$\mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$
$$\mathbf{x}_0 = \mathbf{x}$$

for given model $\mathbf{f}$.



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$
$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$
$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

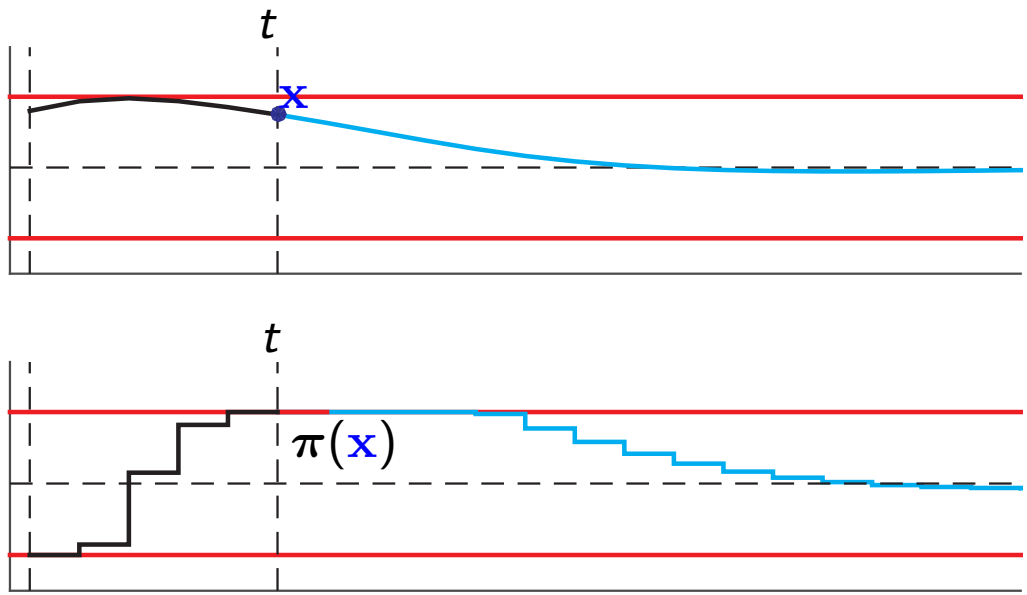
Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad \sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)$$
$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right)$$
$$\mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$
$$\mathbf{x}_0 = \mathbf{x}$$

for given model $\mathbf{f}$.



Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

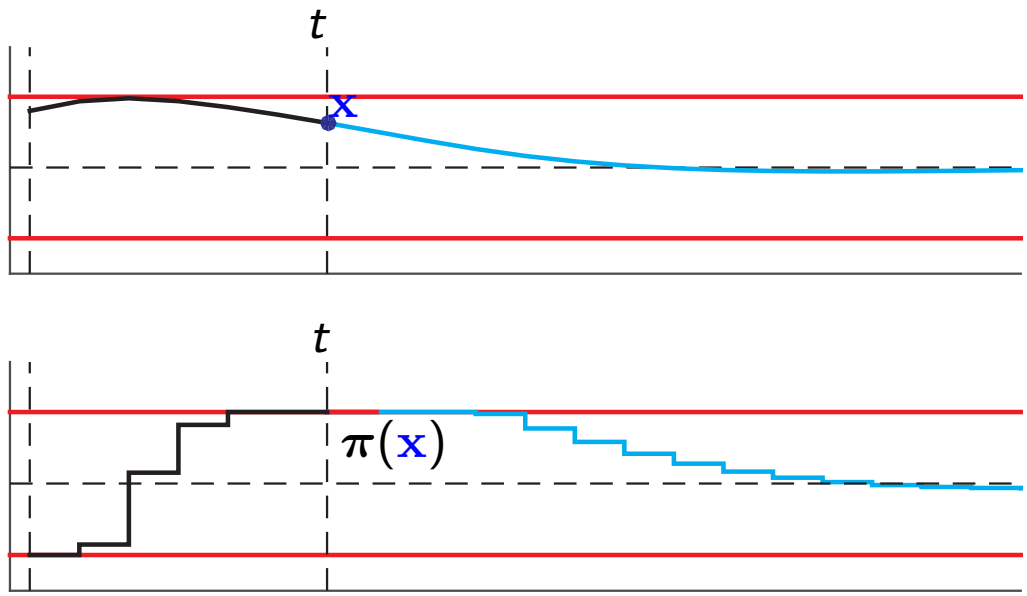
Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

for given model $\mathbf{f}$.



Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

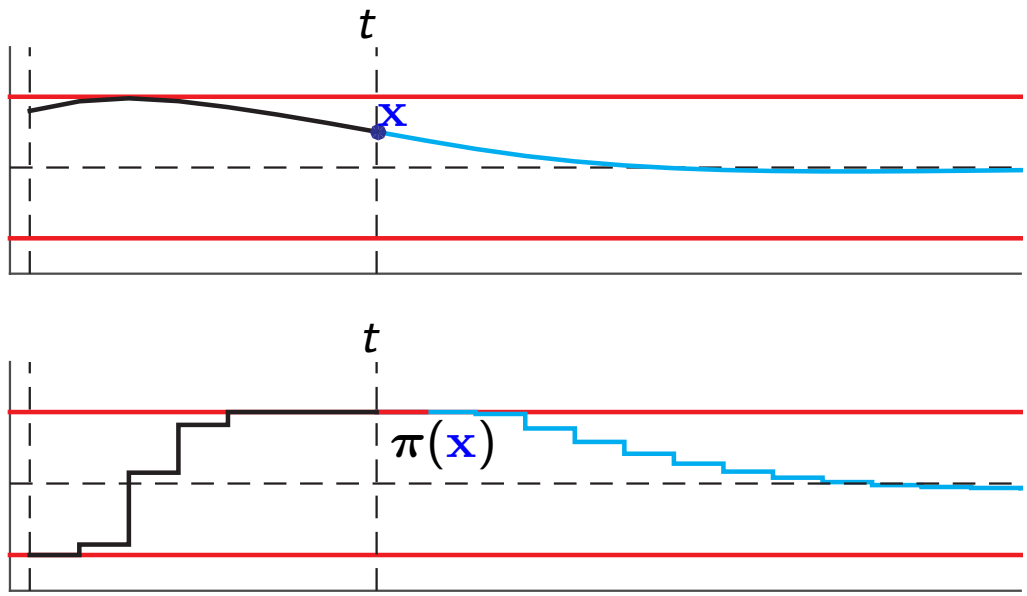
Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

for given model $\mathbf{f}$.



Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

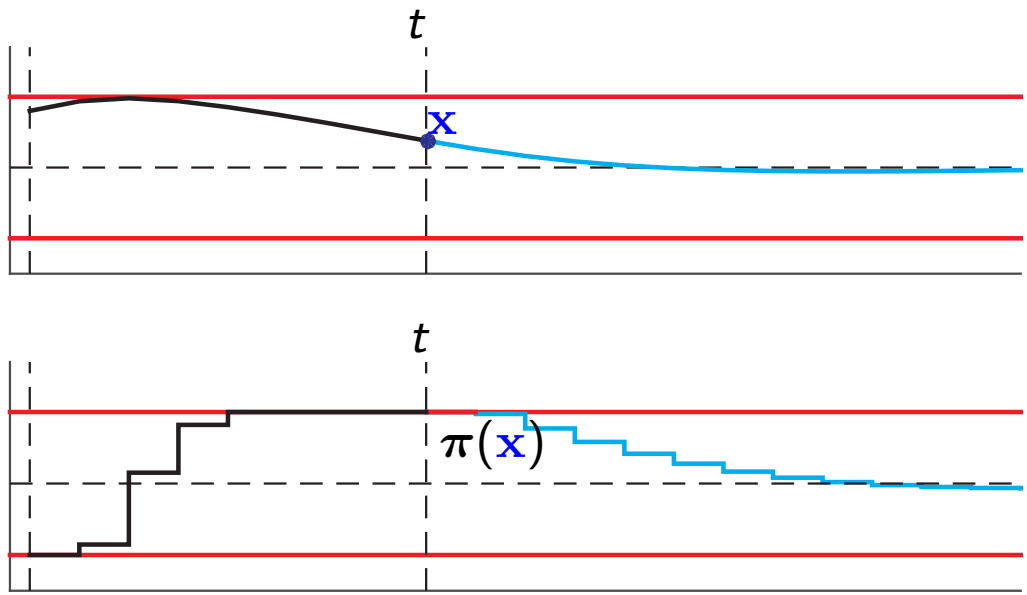
Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

for given model $\mathbf{f}$.



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$
$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$
$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

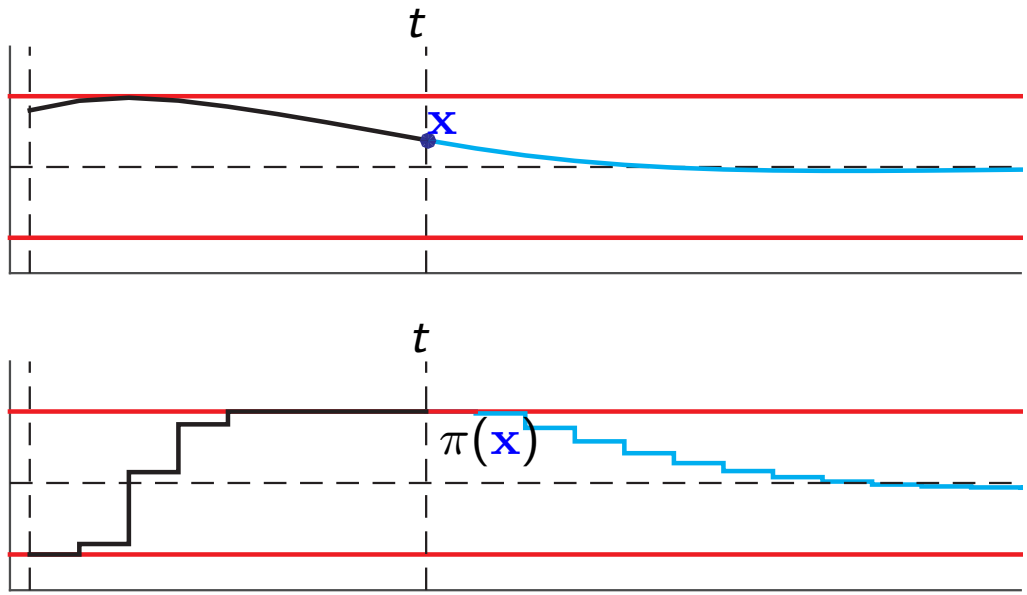
Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad \sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)$$
$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right)$$
$$\mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$
$$\mathbf{x}_0 = \mathbf{x}$$

for given model $\mathbf{f}$.



Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

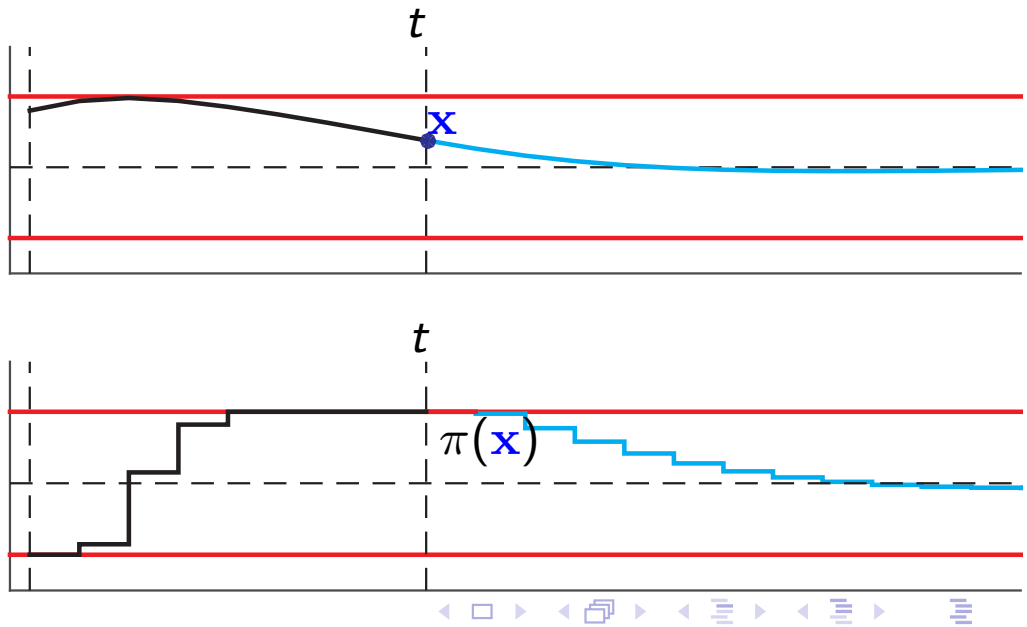
Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

for given model $\mathbf{f}$.



Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

for given model $\mathbf{f}$.

Remarks

- usually $\gamma = 1$
- Generic $L \rightarrow$ Economic NMPC

Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \pi_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

for given model $\mathbf{f}$.

Remarks

- usually $\gamma = 1$
- Generic $L \rightarrow$ Economic NMPC

User manual

- Fit model $\mathbf{f}$ to real dynamics (SYSID)
- Good choice of T

and hope that $\pi \approx \pi_\star$

Optimal control

Minimize "long-term" cost

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right] \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Model-based optimal control - Economic NMPC

Policy $\boldsymbol{\pi}(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

for given model $\mathbf{f}$.

Problem:

- model $\mathbf{f}$ may not be able to capture real dynamics
- real dynamics are often stochastic

nothing guarantees that $\boldsymbol{\pi}$ *from NMPC yields good performances*

Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

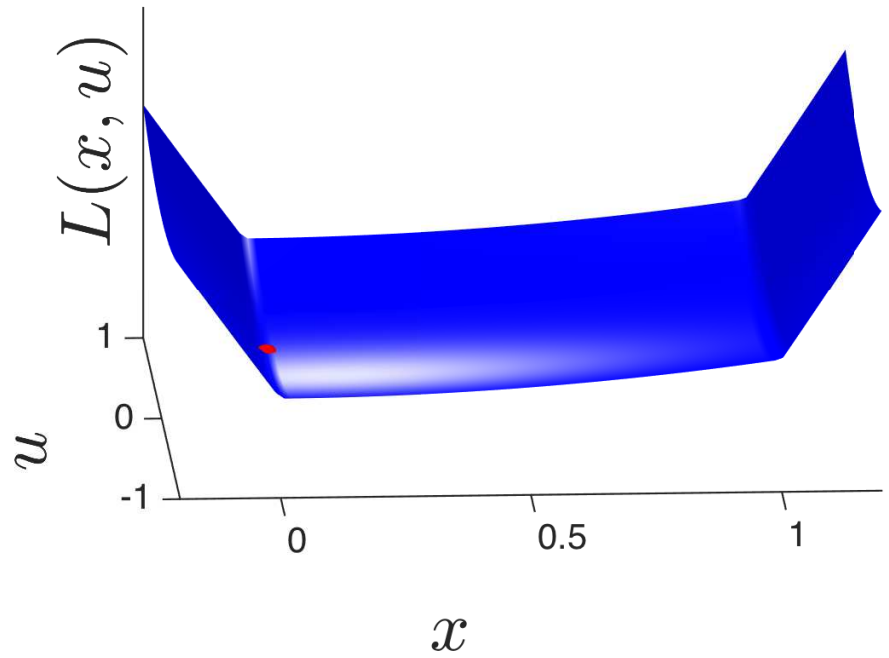
Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Our example

$$x_+ = x + 0.1u + e, \qquad e \sim \mathcal{U}([-0.1, 0])$$



- Quadratic, minimum is at $x = 0$, $u = 0$
- Strong penalty for $x \notin [0, 1]$ and $u \in [-1, 1]$
- Noise e "pushing" x "to the left"

SYSID with deterministic model:
$\hat{x}_+ = x + 0.1u - 0.05$

Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

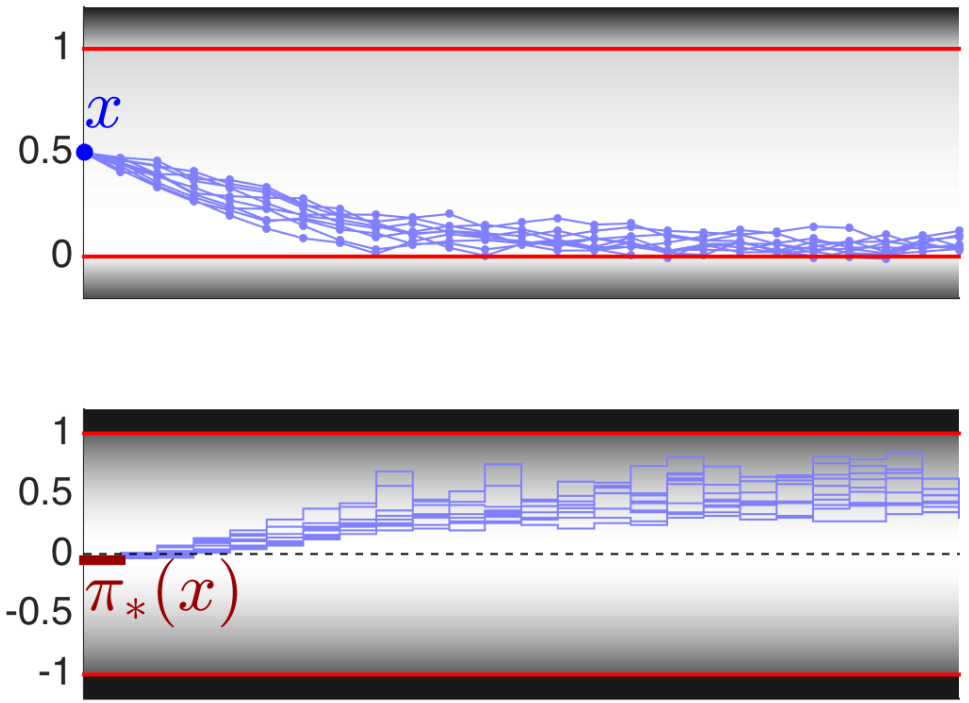
$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal trajectories (policy $\boldsymbol{\pi}_\star(\mathbf{x})$)
(dark = no good)



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

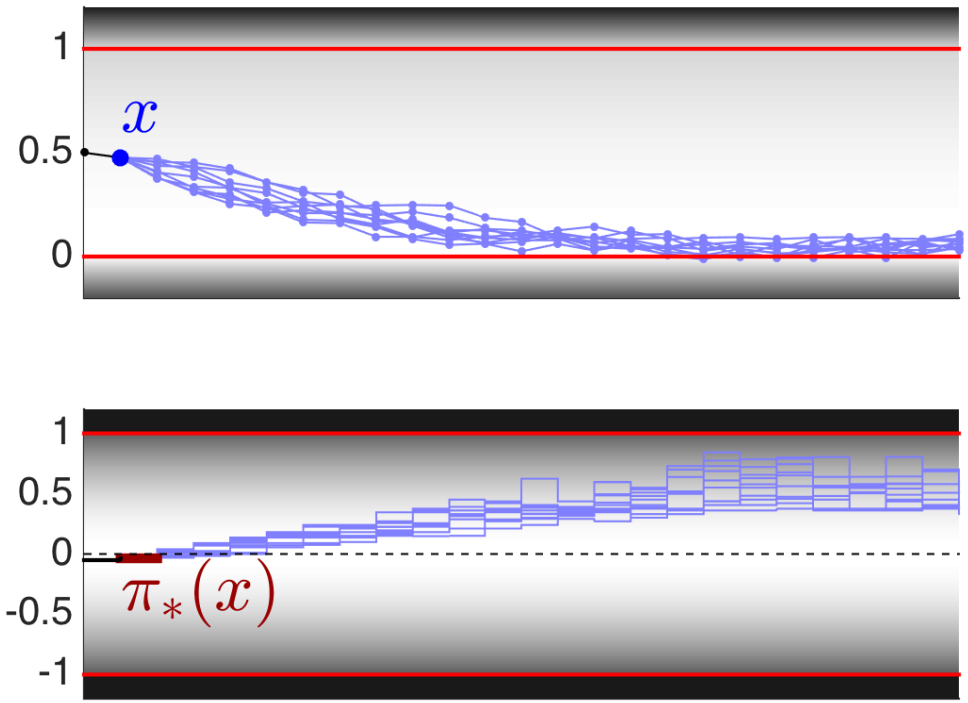
$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal trajectories (policy $\boldsymbol{\pi}_\star(\mathbf{x})$)
(dark = no good)



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

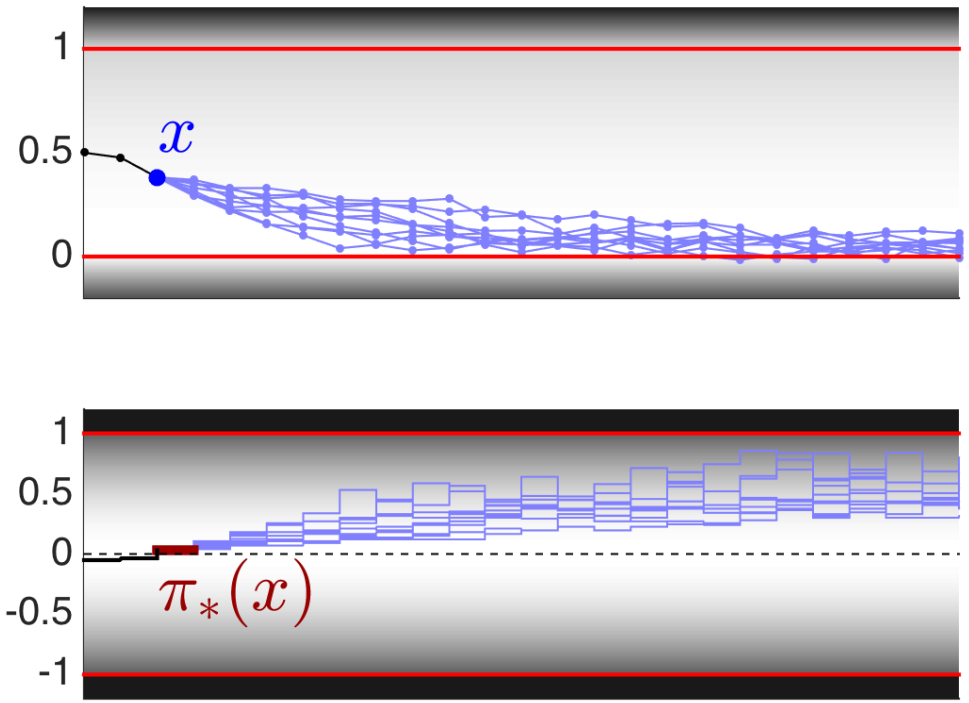
$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal trajectories (policy $\pi_\star(\mathbf{x})$)
(dark = no good)



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

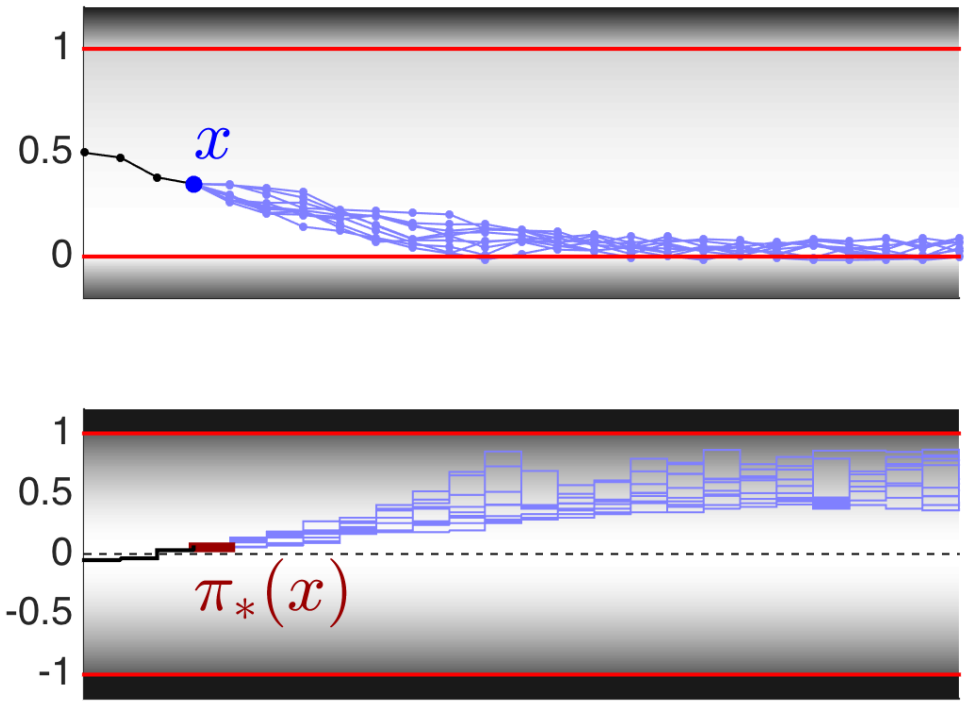
$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal trajectories (policy $\pi_\star(\mathbf{x})$)
(dark = no good)



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

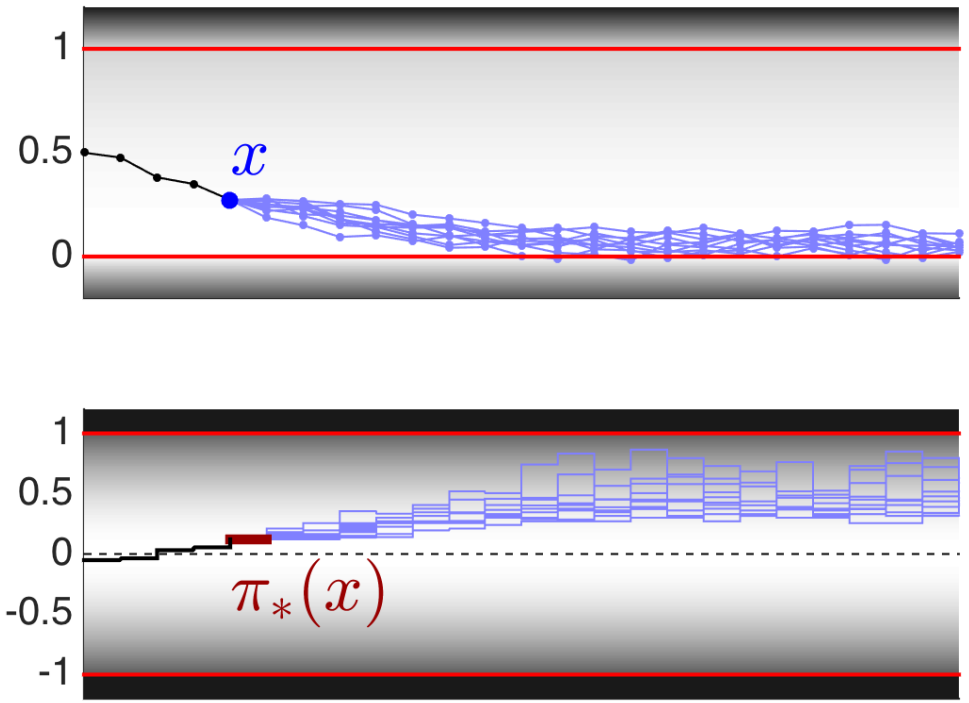
$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal trajectories (policy $\boldsymbol{\pi}_\star\left(\mathbf{x}\right)$)
(dark = no good)



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

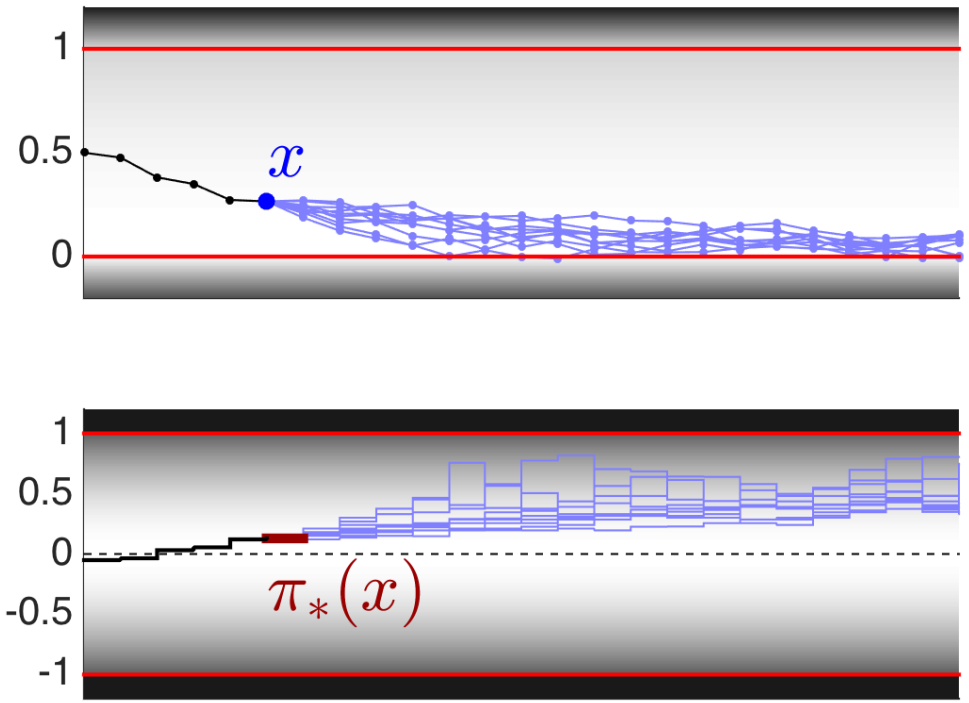
$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal trajectories (policy $\pi_\star(\mathbf{x})$)
(dark = no good)



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

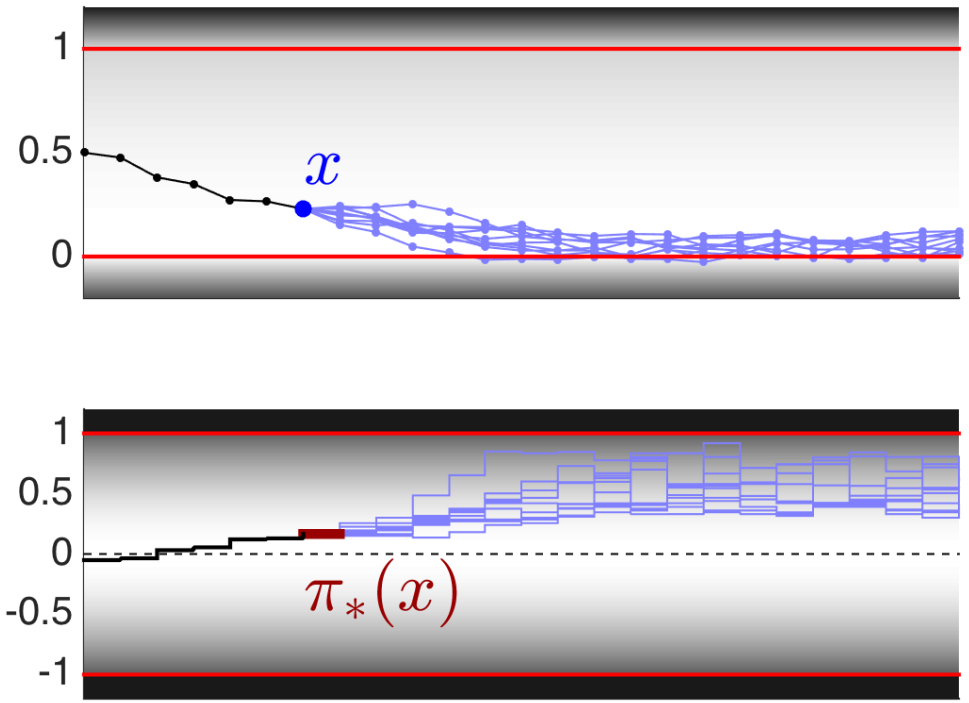
$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal trajectories (policy $\boldsymbol{\pi}_\star(\mathbf{x})$)
(dark = no good)



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

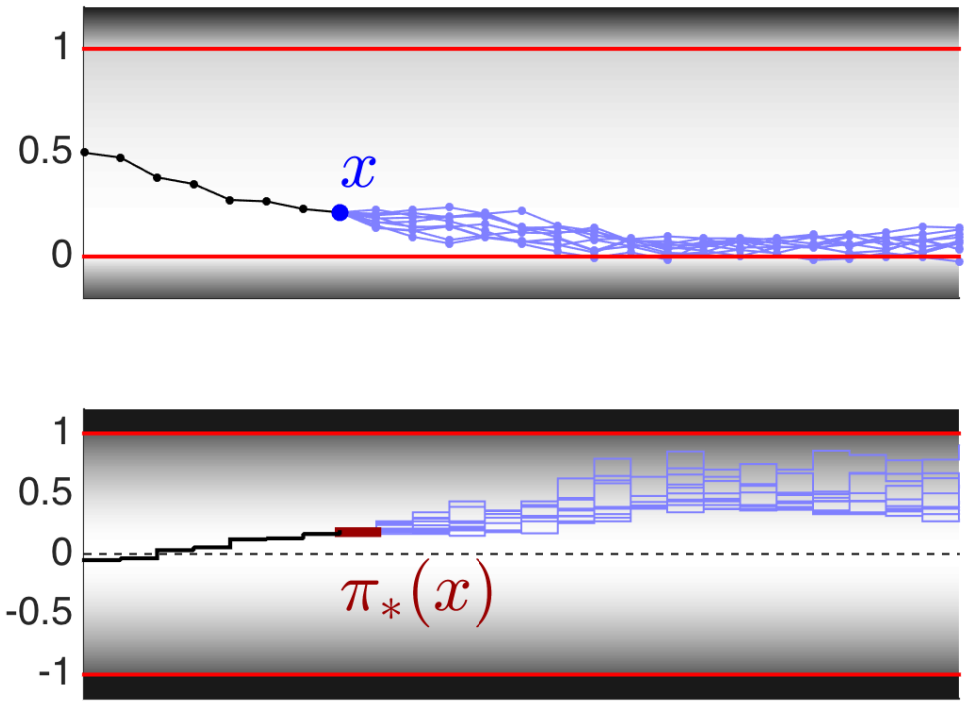
$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal trajectories (policy $\pi_\star(\mathbf{x})$)
(dark = no good)



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

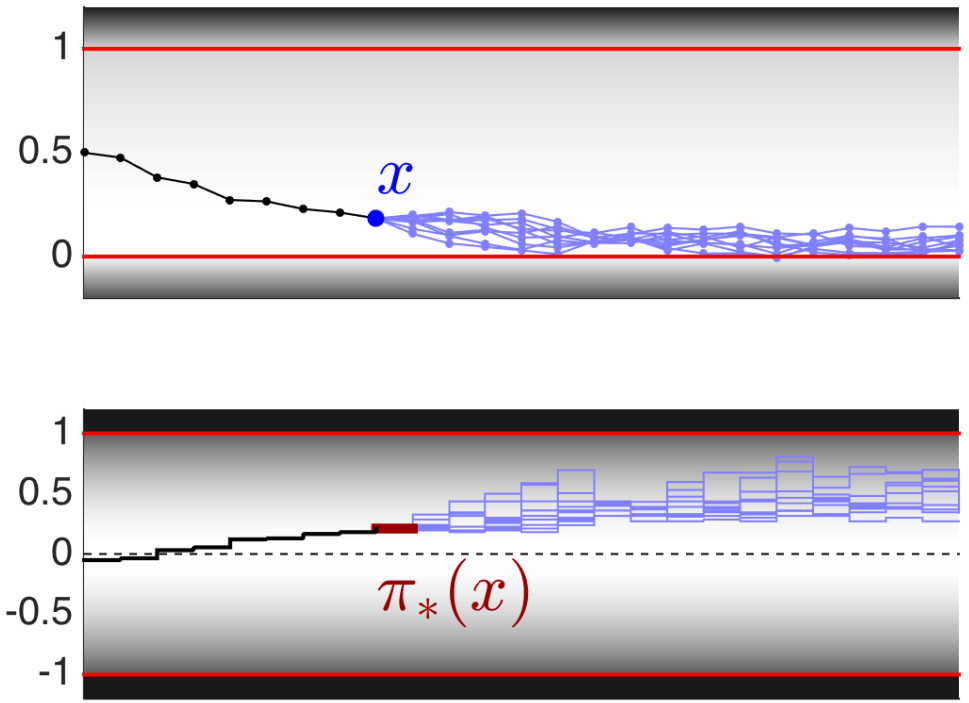
$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal trajectories (policy $\pi_\star(\mathbf{x})$)
(dark = no good)



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

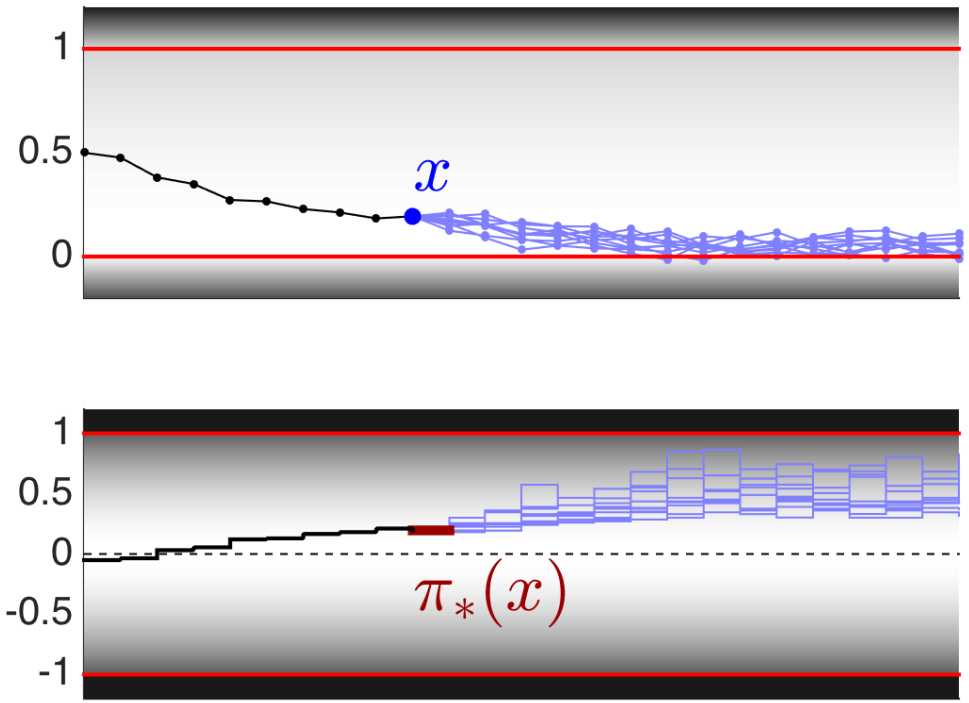
$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal trajectories (policy $\pi_\star(\mathbf{x})$)
(dark = no good)



Optimal control

Minimize "long-term" cost

$$\min_{\mathbf{x},\mathbf{u}} \quad \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right)\right]$$

$$\text{s.t.} \quad \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

given initial condition $\mathbf{x}$.

Under real dynamics:

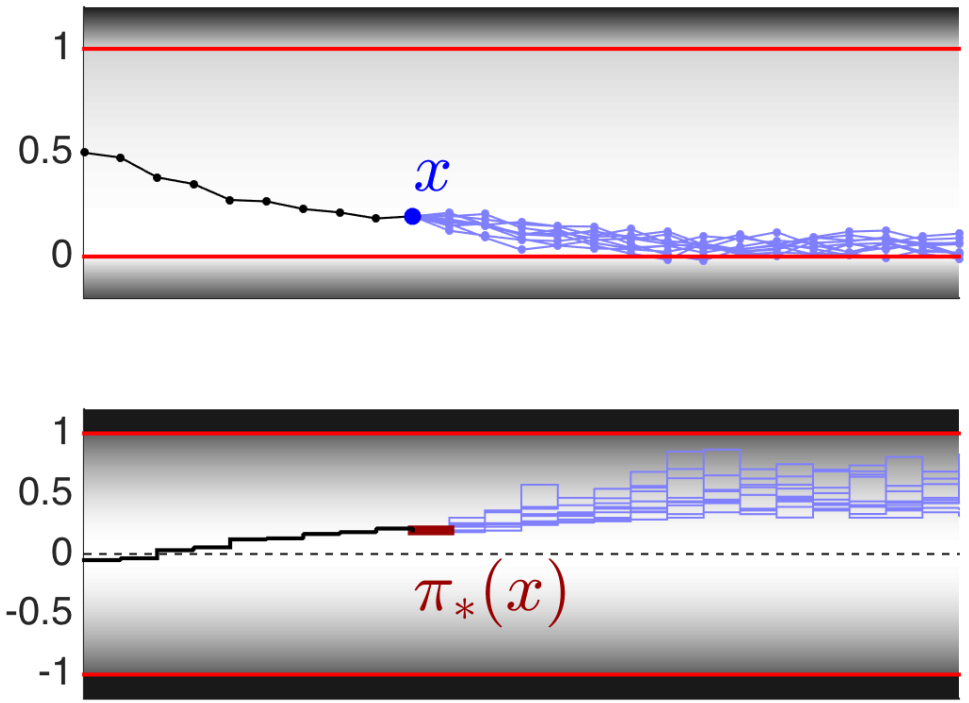
$$\mathbb{P}\left[\mathbf{x}_{k+1} \,|\, \mathbf{x}_k, \mathbf{u}_k\right]$$

Yields optimal **policy**:

$$\mathbf{u} = \boldsymbol{\pi}_\star\left(\mathbf{x}\right)$$

Always update $\mathbf{u}$ *using the latest information* $\mathbf{x}$

Optimal trajectories (policy $\pi_\star(\mathbf{x})$)
(dark = no good)



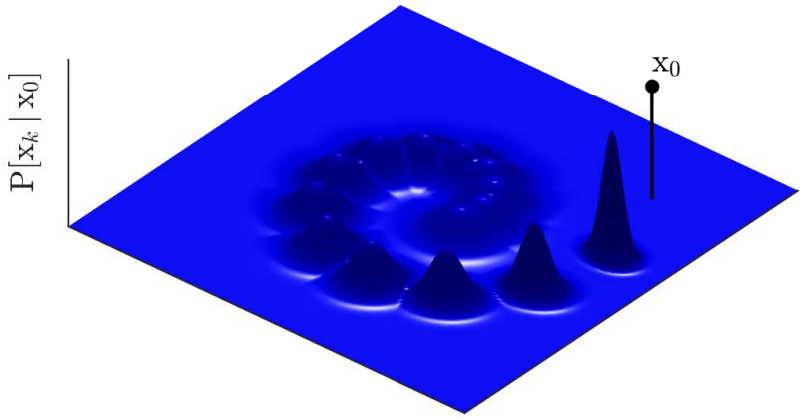
This is a Markov Decision Process (MDP):

- Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$
- Stage cost
- Discount $\gamma \in]0, 1]$

A generic framework - Markov Decision Processes

Markov Decision Process (MDP):

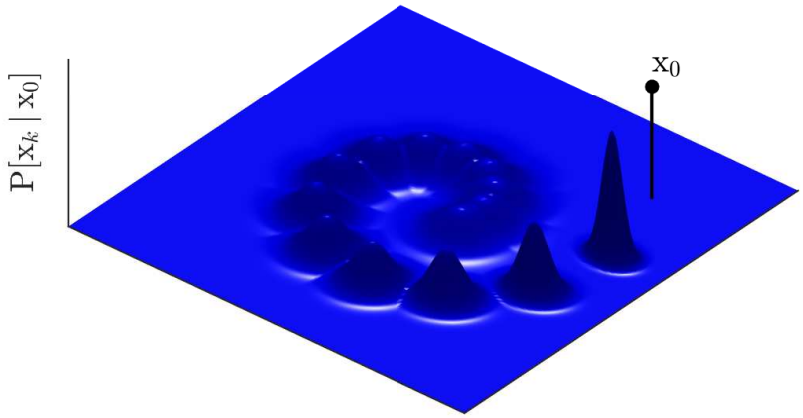
- Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$
- Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) \in \mathbb{R}$
- Discount factor: $\gamma \in [0, 1]$



A generic framework - Markov Decision Processes

Markov Decision Process (MDP):

- Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$
- Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) \in \mathbb{R}$
- Discount factor: $\gamma \in [0, 1]$



Constraints: $\mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0$, then define

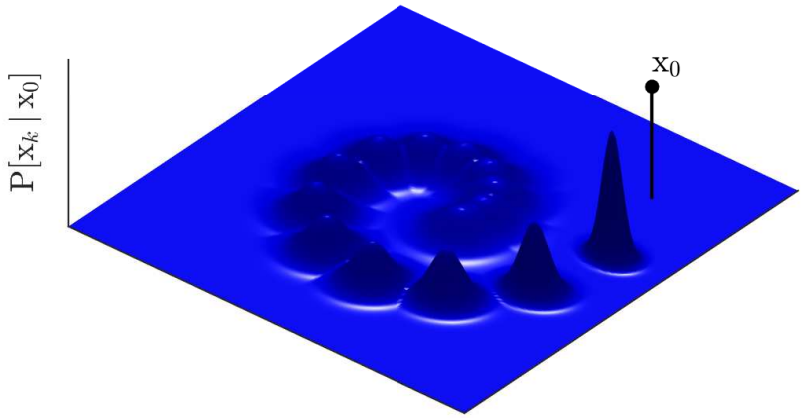
$$\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$$

where $L\left(\mathbf{x}, \mathbf{u}\right)$ is the finite stage cost

A generic framework - Markov Decision Processes

Markov Decision Process (MDP):

- Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$
- Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) \in \mathbb{R}$
- Discount factor: $\gamma \in [0, 1]$



Constraints: $\mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0$, then define

$$\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$$

where $L\left(\mathbf{x}, \mathbf{u}\right)$ is the finite stage cost

Optimal policy $\mathbf{u} = \pi_\star(\mathbf{x})$ yields value function:

$$V_\star\left(\mathbf{x}_0\right) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k \ell\left(\mathbf{x}_k, \pi_\star(\mathbf{x}_k)\right) \,|\, \mathbf{x}_0, \pi_\star\right]$$

(can take ∞ values)

Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Solving MDPs - Dynamic Programming & the Bellman equations

Real system

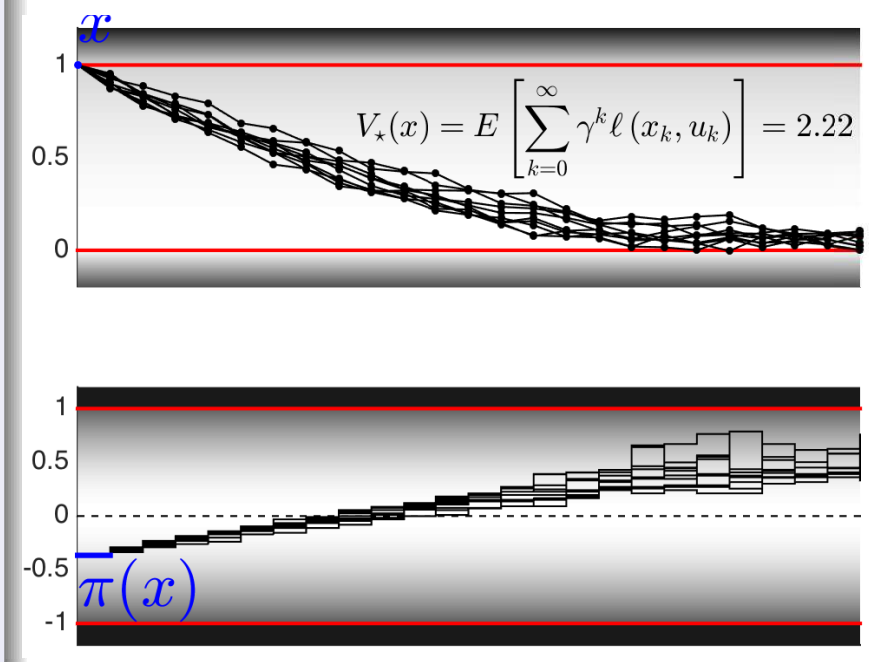
Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

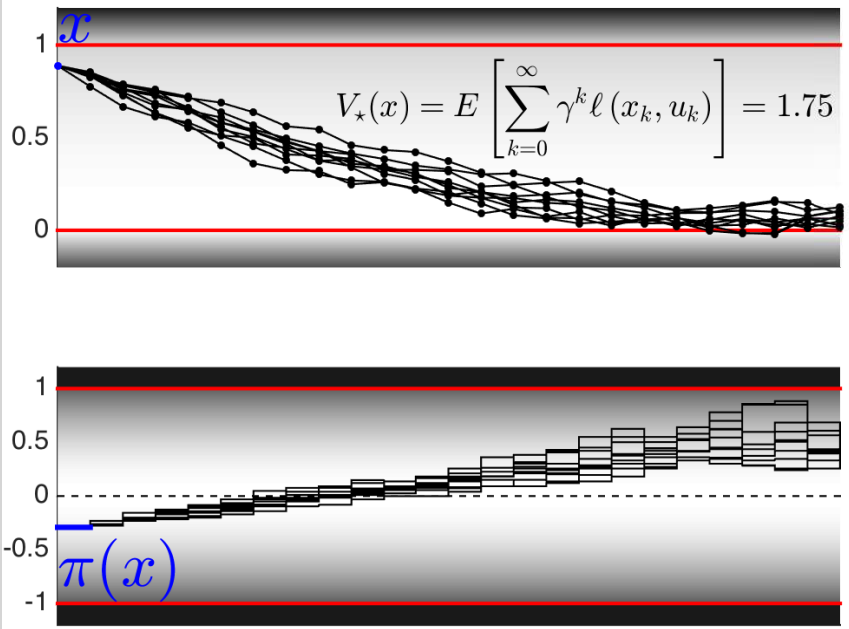
Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

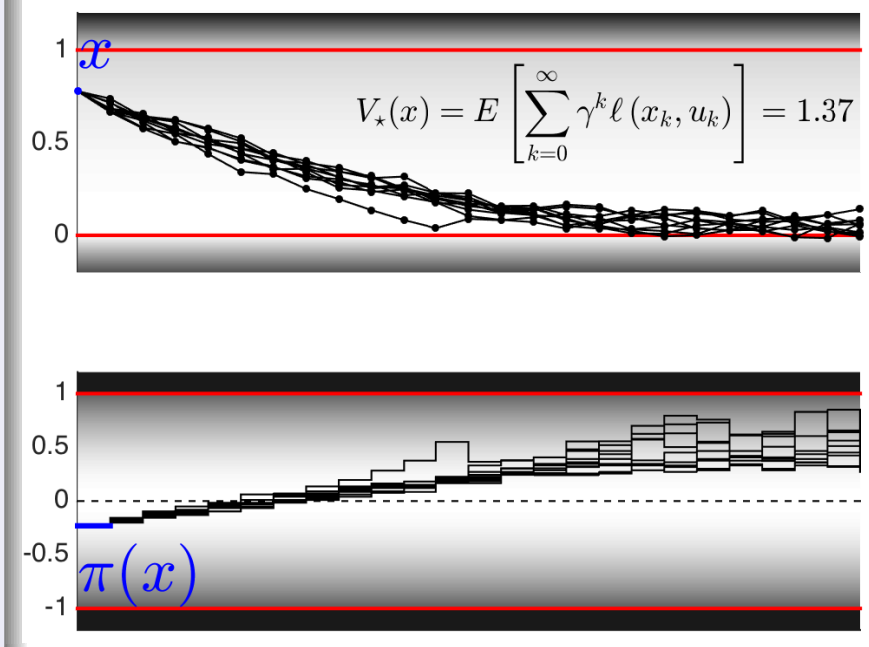
Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

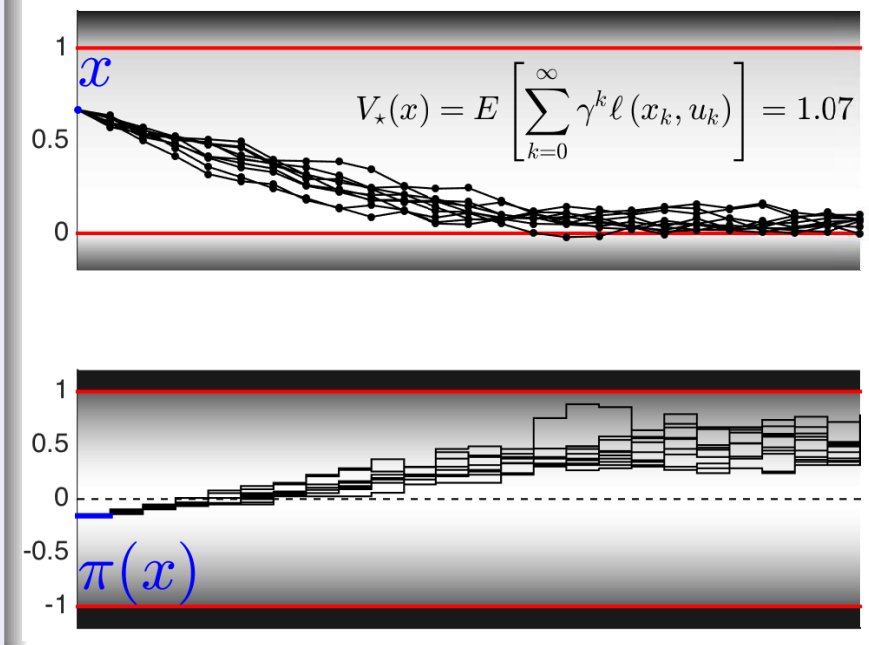
Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

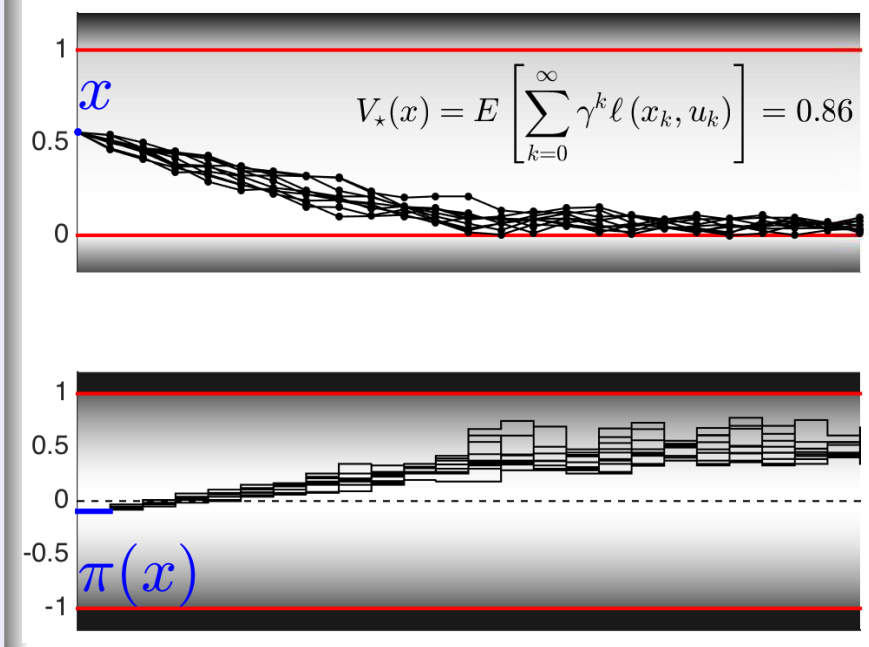
Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

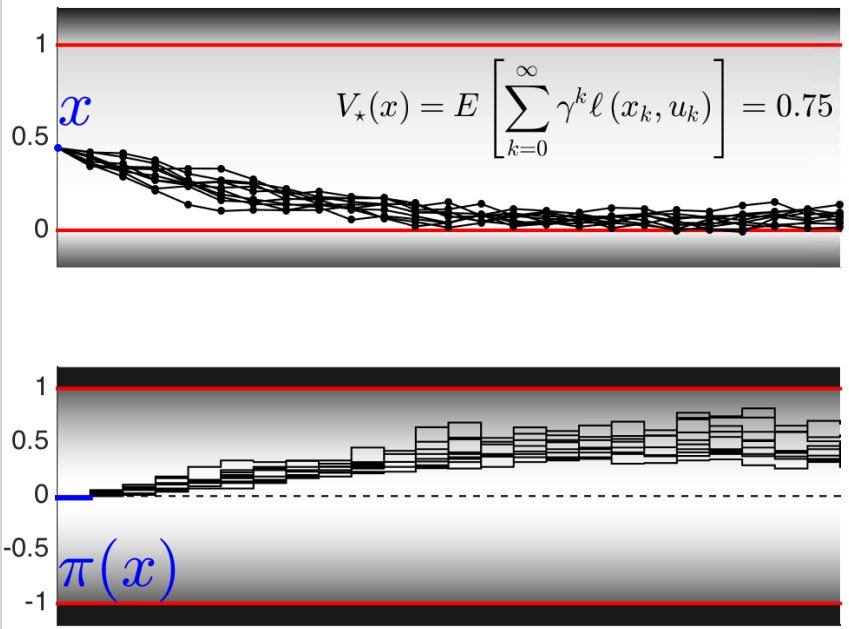
Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

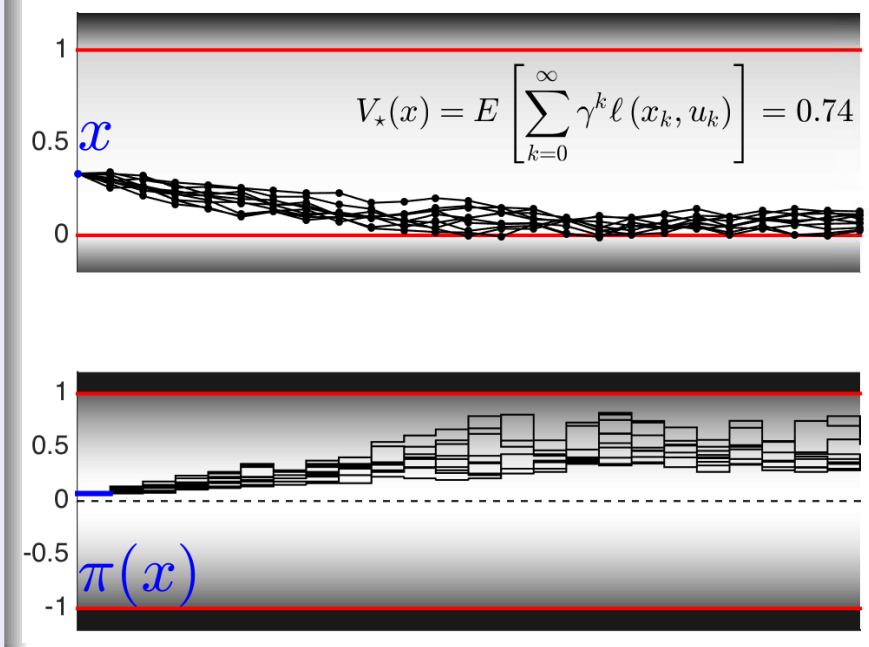
Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

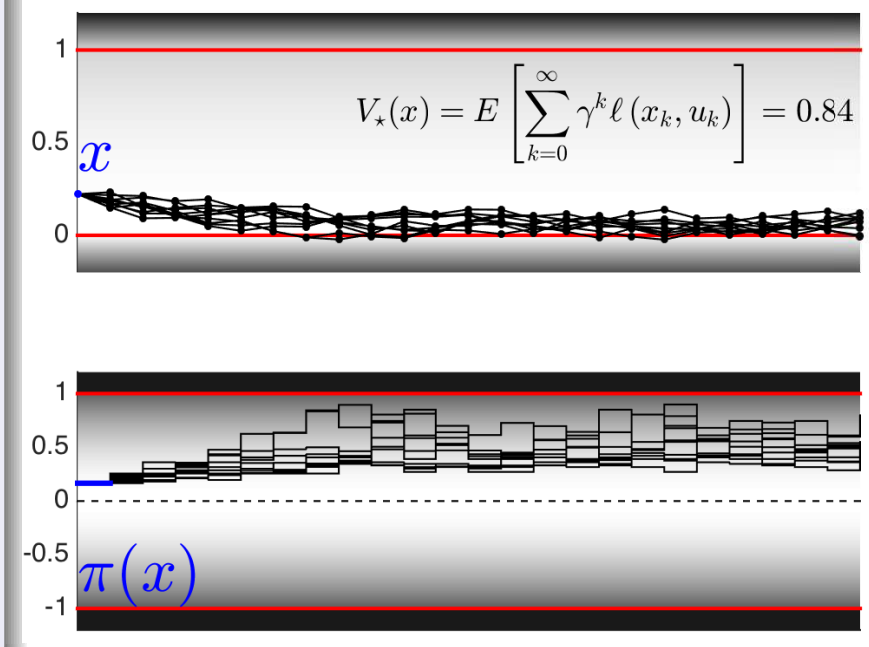
Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

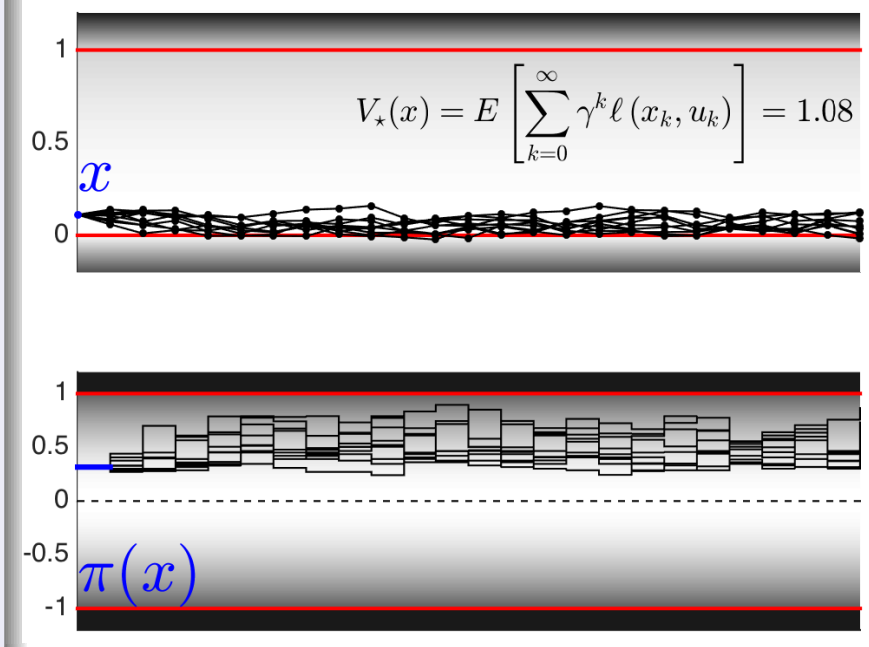
Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

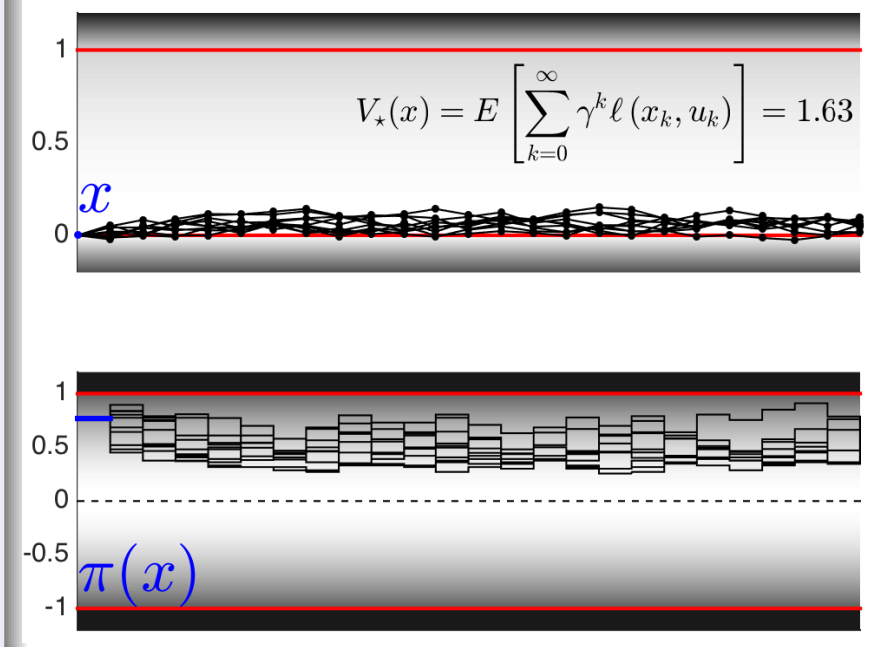
Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

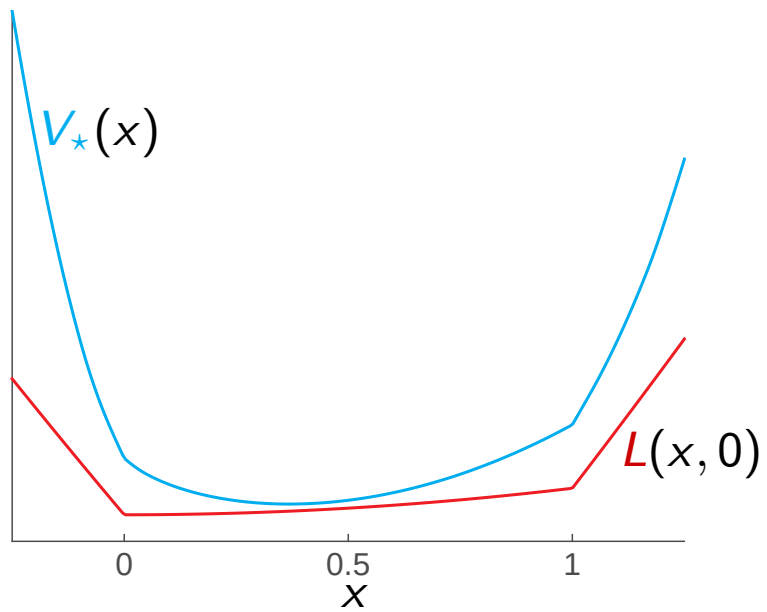
Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

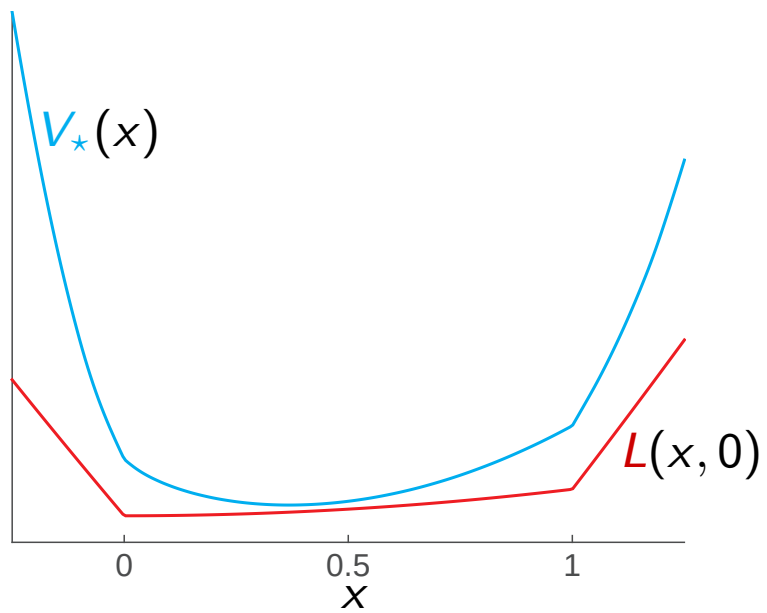
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

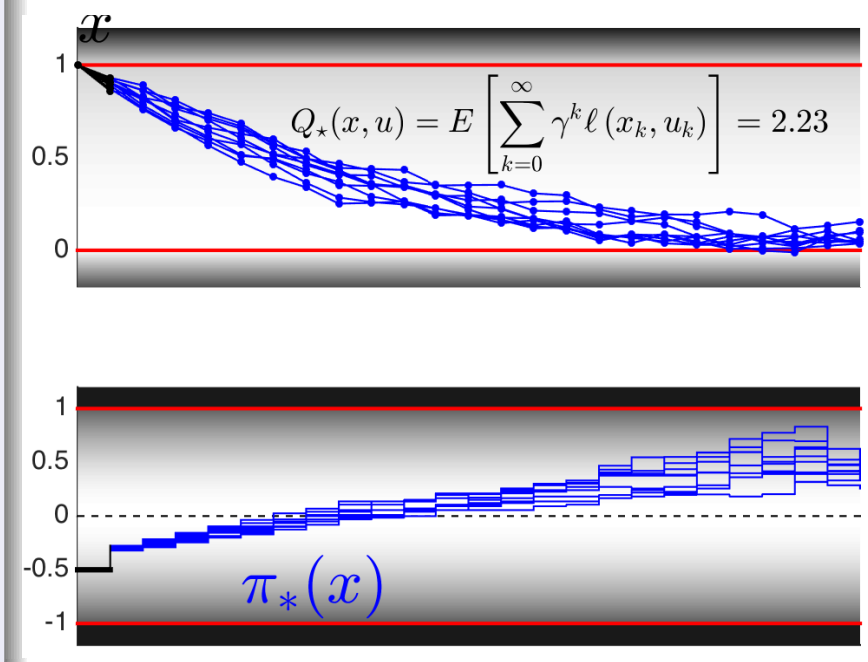
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

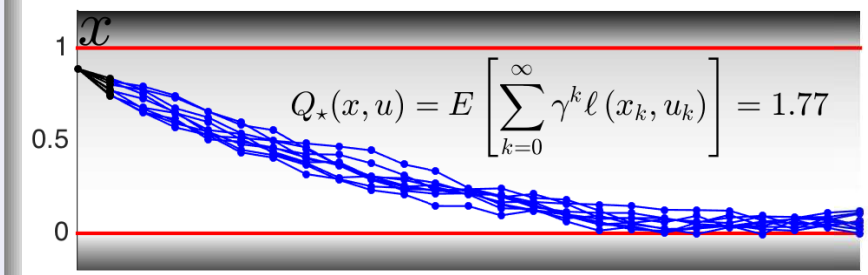
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

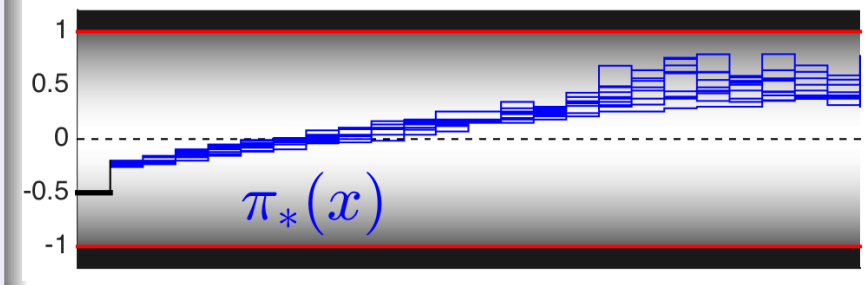
Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example





Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

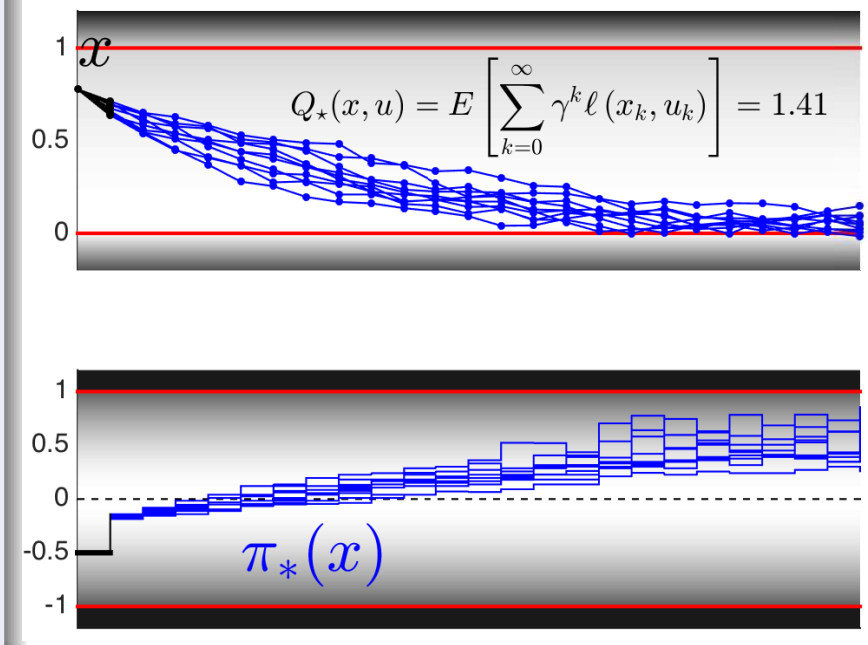
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

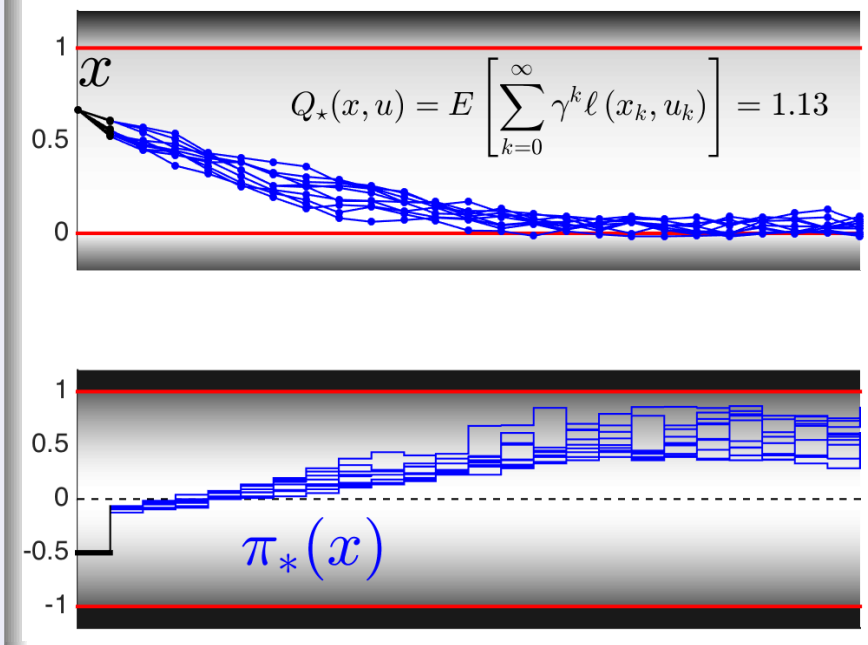
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

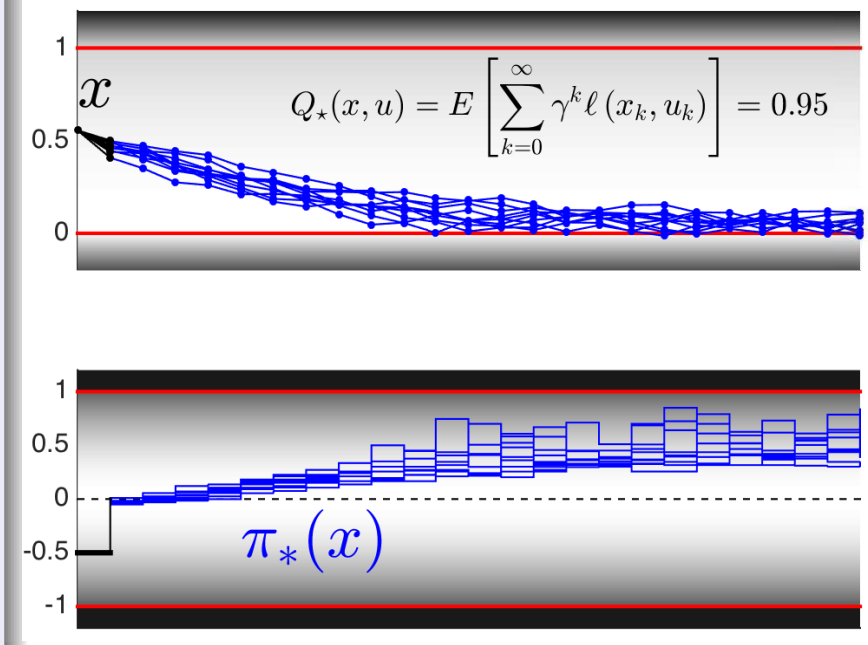
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

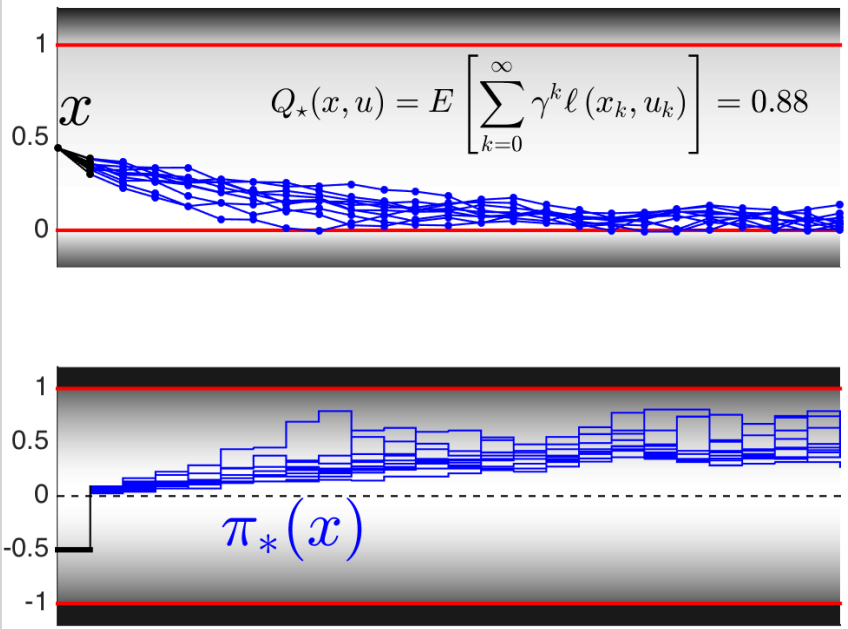
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

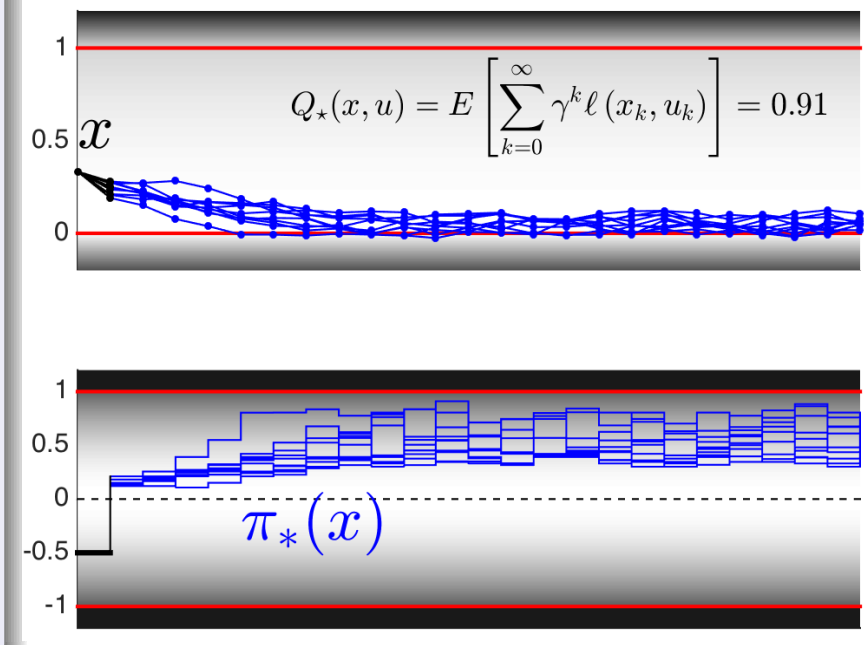
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

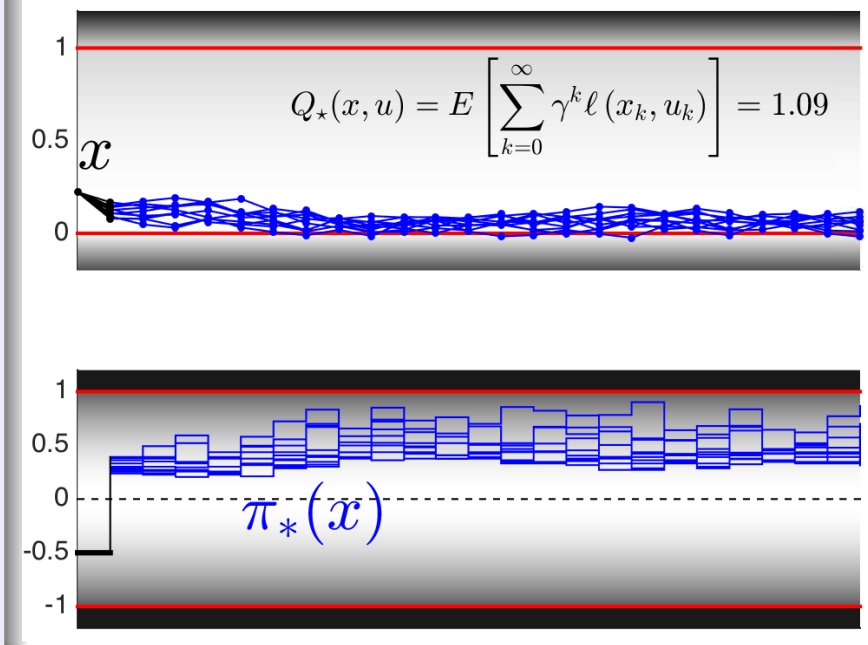
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

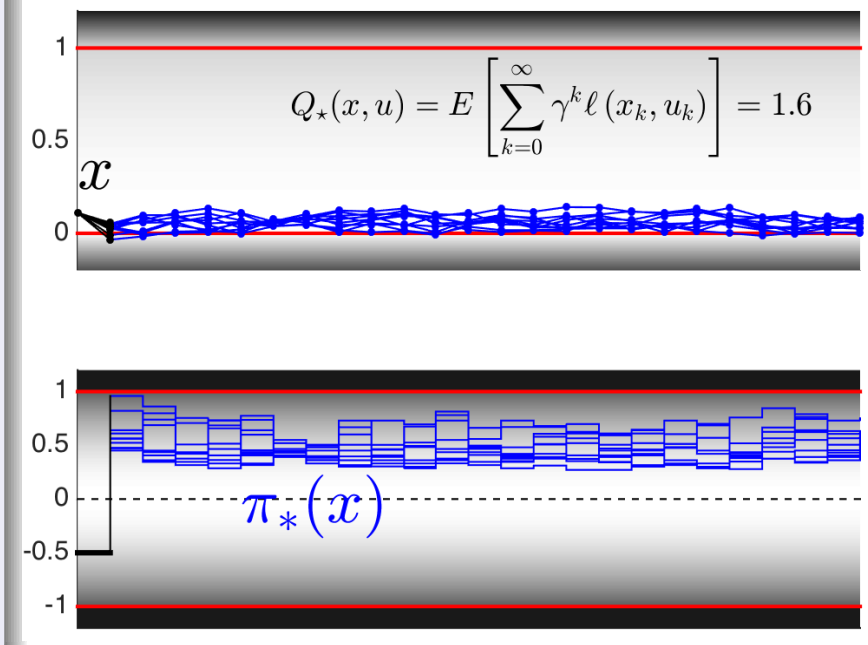
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

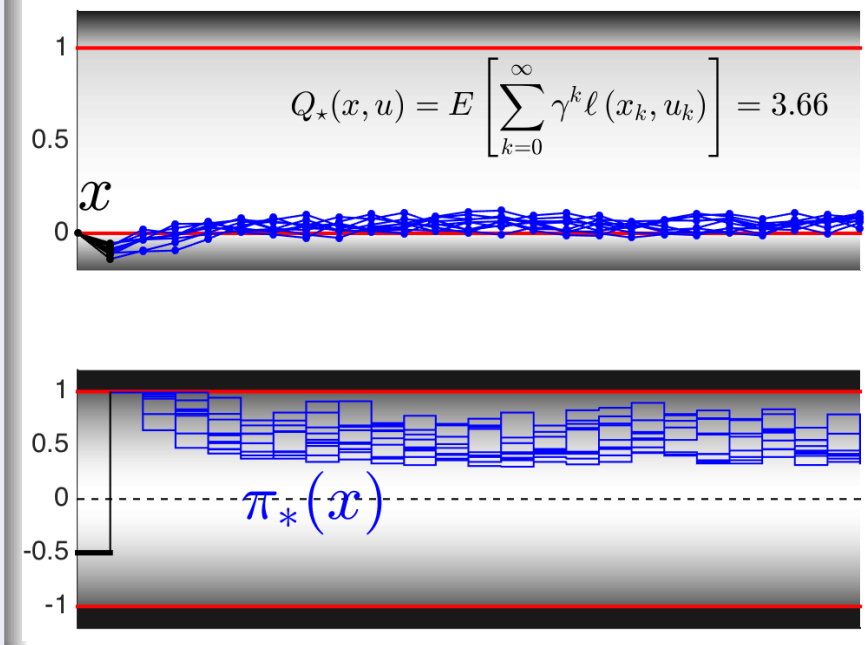
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

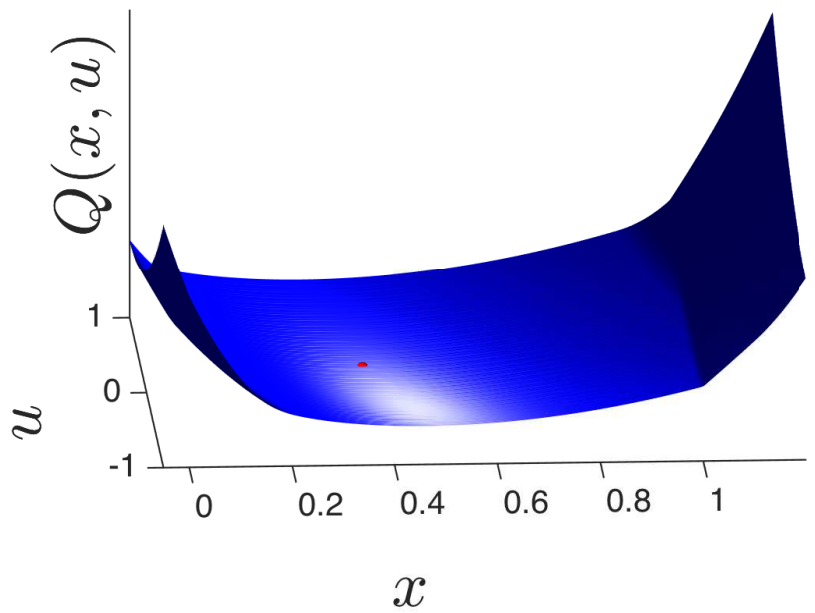
Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

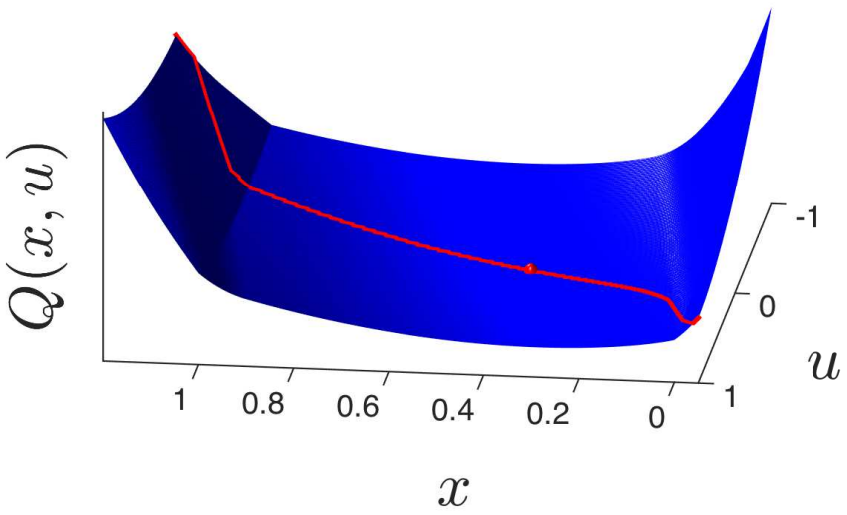
$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

Relationship $V_\star \leftrightarrow Q_\star$:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; Q_\star(\mathbf{x}, \mathbf{u}) = Q_\star(\mathbf{x}, \boldsymbol{\pi}_\star\left(\mathbf{x}\right))$$

Our trivial example



Solving MDPs - Dynamic Programming & the Bellman equations

Real system

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Bellman optimality equations:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[V_\star\left(\mathbf{x}_+\right) \,|\, \mathbf{x}, \mathbf{u}\right]$$

$$Q_\star(\mathbf{x}, \mathbf{u}) = \ell(\mathbf{x}, \mathbf{u}) + \gamma \mathbb{E}\left[\min_{\mathbf{u}'} \; Q_\star(\mathbf{x}_+, \mathbf{u}') \,|\, \mathbf{x}, \mathbf{u}\right]$$

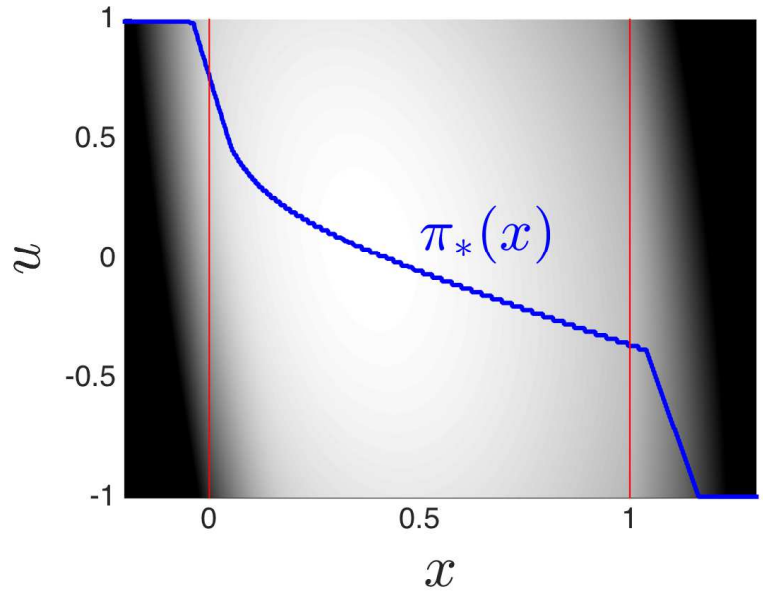
Relationship $V_\star \leftrightarrow Q_\star$:

$$V_\star\left(\mathbf{x}\right) = \min_{\mathbf{u}} \; Q_\star(\mathbf{x}, \mathbf{u}) = Q_\star(\mathbf{x}, \boldsymbol{\pi}_\star\left(\mathbf{x}\right))$$

Optimal policy:

$$\boldsymbol{\pi}_\star\left(\mathbf{x}\right) = \arg\min_{\mathbf{u}} \; Q_\star(\mathbf{x}, \mathbf{u})$$

Our trivial example



Outline

Modification of the ENMPC scheme

Real system - MDP

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Value functions: $Q_\star\left(\mathbf{x}, \mathbf{u}\right), \quad V_\star\left(\mathbf{x}\right)$

Optimal policy $\pi_\star\left(\mathbf{x}\right) = \arg\min_{\mathbf{u}} \; Q_\star(\mathbf{x}, \mathbf{u})$

Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

Modification of the ENMPC scheme

Real system - MDP

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell(\mathbf{x}, \mathbf{u}) = \begin{cases} L(\mathbf{x}, \mathbf{u}) & \text{if} \quad \mathbf{h}(\mathbf{x}, \mathbf{u}) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Value functions: $Q_\star(\mathbf{x}, \mathbf{u}), \quad V_\star(\mathbf{x})$

Optimal policy $\pi_\star(\mathbf{x}) = \arg\min_{\mathbf{u}} Q_\star(\mathbf{x}, \mathbf{u})$

Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

Modification of the ENMPC scheme

Real system - MDP

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Value functions: $Q_\star\left(\mathbf{x}, \mathbf{u}\right), \quad V_\star\left(\mathbf{x}\right)$

Optimal policy $\pi_\star\left(\mathbf{x}\right) = \arg\min_{\mathbf{u}} Q_\star(\mathbf{x}, \mathbf{u})$

Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k \hat{L}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \hat{\mathbf{h}}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

Can we gain from modifying the cost and constraints?†

† idea coming from Real-Time Optimization

Modification of the ENMPC scheme

Real system - MDP

Dynamics: $\mathbb{P}\left[\mathbf{x}_+ \,|\, \mathbf{x}, \mathbf{u}\right]$

Stage cost: $\ell\left(\mathbf{x}, \mathbf{u}\right) = \begin{cases} L\left(\mathbf{x}, \mathbf{u}\right) & \text{if} \quad \mathbf{h}\left(\mathbf{x}, \mathbf{u}\right) \leq 0 \\ \infty & \text{otherwise} \end{cases}$

Value functions: $Q_\star\left(\mathbf{x}, \mathbf{u}\right), \quad V_\star\left(\mathbf{x}\right)$

Optimal policy $\pi_\star\left(\mathbf{x}\right) = \arg\min_{\mathbf{u}} Q_\star(\mathbf{x}, \mathbf{u})$

Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k \hat{L}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \hat{\mathbf{h}}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

Can we gain from modifying the cost and constraints?†

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\pi(\mathbf{x}) = \pi_\star\left(\mathbf{x}\right)$$

even though the model $\mathbf{f}$ is wrong ?

† idea coming from Real-Time Optimization

Modification of the ENMPC scheme

Economic NMPC

Policy $\boldsymbol{\pi}(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 = \mathbf{x} \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \end{aligned}$$

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\boldsymbol{\pi}(\mathbf{x}) = \boldsymbol{\pi}_\star(\mathbf{x})$$

even though the model $\mathbf{f}$ is wrong ?

Modification of the ENMPC scheme

Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 = \mathbf{x} \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \end{aligned}$$

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\pi(\mathbf{x}) = \pi_\star(\mathbf{x})$$

even though the model $\mathbf{f}$ is wrong ?

Theorem

Under some assumptions there is a $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ such that

$$\begin{aligned} \pi(\mathbf{x}) &= \pi_\star(\mathbf{x}) \\ V(\mathbf{x}) &= V_\star(\mathbf{x}) \\ Q(\mathbf{x}, \mathbf{u}) &= Q_\star(\mathbf{x}, \mathbf{u}) \end{aligned}$$

π, V, Q delivered by ENMPC

Modification of the ENMPC scheme

Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 = \mathbf{x} \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \end{aligned}$$

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\pi(\mathbf{x}) = \pi_\star(\mathbf{x})$$

even though the model $\mathbf{f}$ is wrong ?

Theorem

Under some assumptions there is a $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ such that

$$\begin{aligned} \pi(\mathbf{x}) &= \pi_\star(\mathbf{x}) \\ V(\mathbf{x}) &= V_\star(\mathbf{x}) \\ Q(\mathbf{x}, \mathbf{u}) &= Q_\star(\mathbf{x}, \mathbf{u}) \end{aligned}$$

π, V, Q delivered by ENMPC

Bottom line: modifying the cost and constraints of the (E)NMPC may help gaining performance when the model is wrong.

Modification of the ENMPC scheme

Economic NMPC

Policy $\boldsymbol{\pi}(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 = \mathbf{x} \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \end{aligned}$$

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\boldsymbol{\pi}(\mathbf{x}) = \boldsymbol{\pi}_\star(\mathbf{x})$$

even though the model $\mathbf{f}$ is wrong ?

Modification of the ENMPC scheme

Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 = \mathbf{x} \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \end{aligned}$$

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\pi(\mathbf{x}) = \pi_\star(\mathbf{x})$$

even though the model $\mathbf{f}$ is wrong ?

Theorem

There is a $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\pi = \pi_\star,\ Q = Q_\star,\ V = V_\star$$

Modification of the ENMPC scheme

Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 = \mathbf{x} \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \end{aligned}$$

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\pi(\mathbf{x}) = \pi_\star(\mathbf{x})$$

even though the model $\mathbf{f}$ is wrong ?

Theorem

There is a $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\pi = \pi_\star, \; Q = Q_\star, \; V = V_\star$$

on $\mathbf{x} \in S$ with:

$$S = \{\mathbf{x}_0 \mid V_\star(\mathbf{x}_k) \text{ finite } \forall k\}$$

where $\mathbf{x}_{0,\ldots\infty}$ is given by

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \pi_\star(\mathbf{x}_k))$$

Modification of the ENMPC scheme

Economic NMPC

Policy $\boldsymbol{\pi}(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 = \mathbf{x} \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \end{aligned}$$

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\boldsymbol{\pi}(\mathbf{x}) = \boldsymbol{\pi}_\star(\mathbf{x})$$

even though the model $\mathbf{f}$ is wrong ?

Theorem

There is a $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

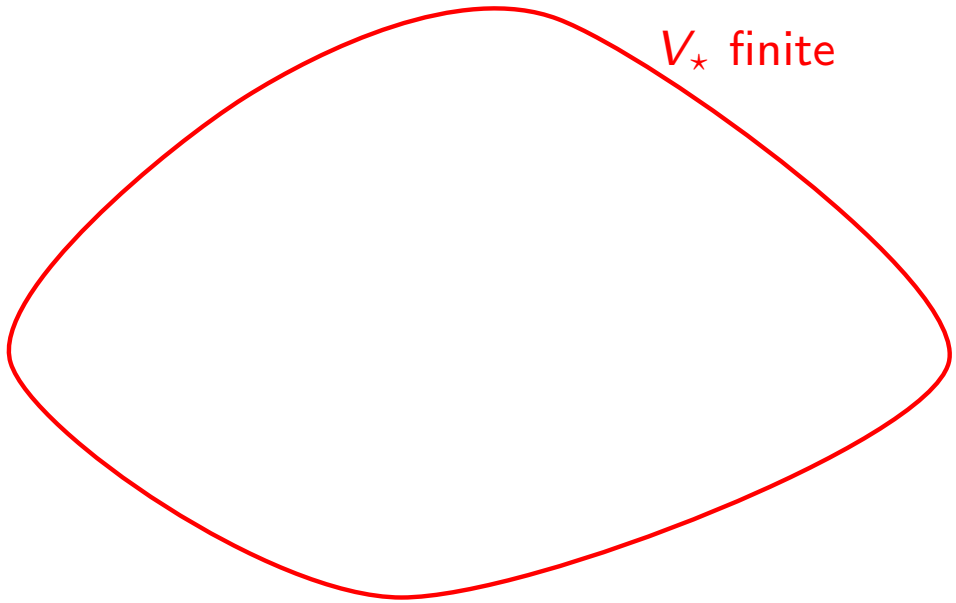
$$\boldsymbol{\pi} = \boldsymbol{\pi}_\star,\ Q = Q_\star,\ V = V_\star$$

on $\mathbf{x} \in S$ with:

$$S = \{\mathbf{x}_0 \mid V_\star(\mathbf{x}_k) \text{ finite } \forall k\}$$

where $\mathbf{x}_{0,\ldots\infty}$ is given by

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \pi_\star(\mathbf{x}_k))$$



Modification of the ENMPC scheme

Economic NMPC

Policy $\pi(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 = \mathbf{x} \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \end{aligned}$$

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\pi(\mathbf{x}) = \pi_\star(\mathbf{x})$$

even though the model $\mathbf{f}$ is wrong ?

Theorem

There is a $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

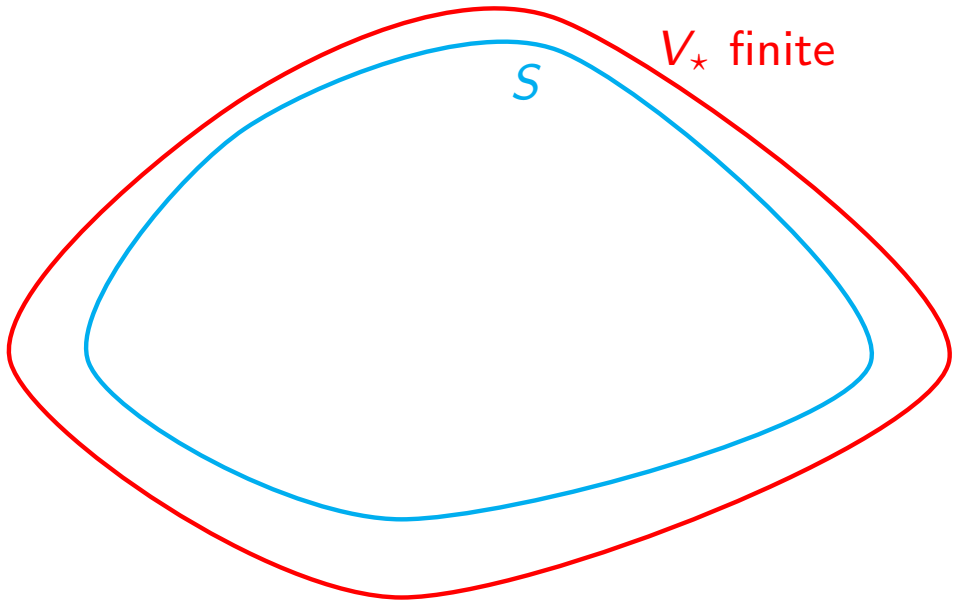
$$\pi = \pi_\star,\ Q = Q_\star,\ V = V_\star$$

on $\mathbf{x} \in S$ with:

$$S = \{\mathbf{x}_0 \mid V_\star(\mathbf{x}_k) \text{ finite } \forall k\}$$

where $\mathbf{x}_{0,\ldots\infty}$ is given by

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \pi_\star(\mathbf{x}_k))$$



Modification of the ENMPC scheme

Economic NMPC

Policy $\boldsymbol{\pi}(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 = \mathbf{x} \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \end{aligned}$$

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\boldsymbol{\pi}(\mathbf{x}) = \boldsymbol{\pi}_\star(\mathbf{x})$$

even though the model $\mathbf{f}$ is wrong ?

Theorem

There is a $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

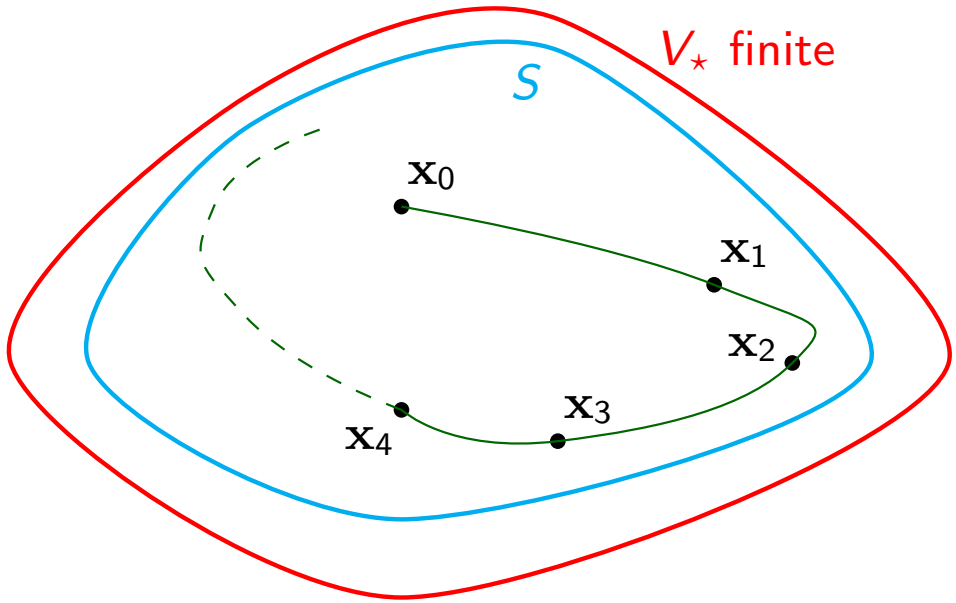
$$\boldsymbol{\pi} = \boldsymbol{\pi}_\star,\ Q = Q_\star,\ V = V_\star$$

on $\mathbf{x} \in S$ with:

$$S = \{\mathbf{x}_0 \mid V_\star(\mathbf{x}_k) \text{ finite } \forall k\}$$

where $\mathbf{x}_{0,\ldots\infty}$ is given by

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \pi_\star(\mathbf{x}_k))$$



Forward-invariant set of the model dynamics under $\boldsymbol{\pi}_\star$ within set where $V_\star$ is finite

Modification of the ENMPC scheme

Economic NMPC

Policy $\boldsymbol{\pi}(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 = \mathbf{x} \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \end{aligned}$$

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\boldsymbol{\pi}(\mathbf{x}) = \boldsymbol{\pi}_\star(\mathbf{x})$$

even though the model $\mathbf{f}$ is wrong ?

Theorem

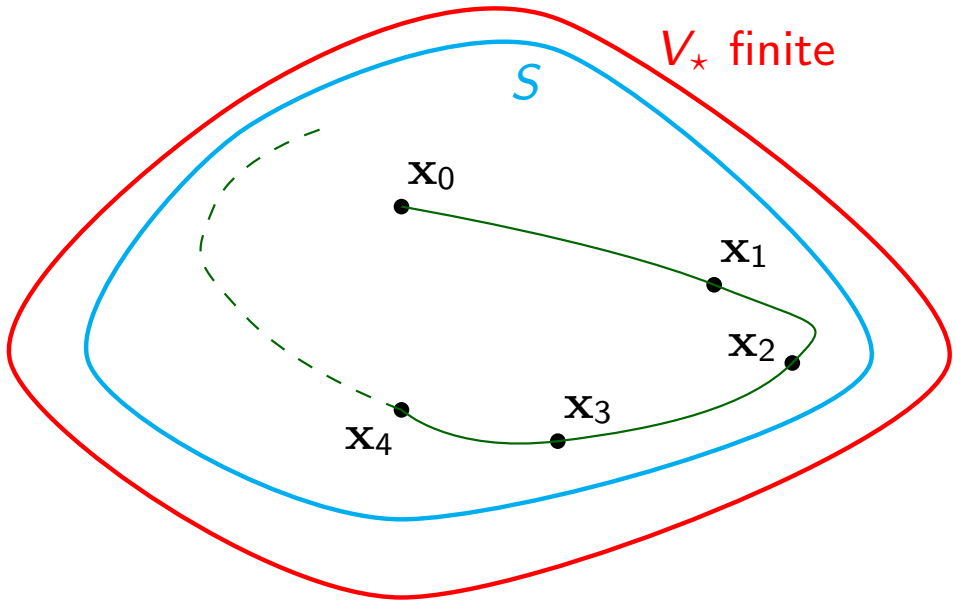
If $\hat{T} = 0$, and if for k large

$$\gamma^k V_\star(\mathbf{x}_k) \to 0$$

then there is $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\boldsymbol{\pi}(\mathbf{x}) \to \boldsymbol{\pi}_\star(\mathbf{x}) \text{ on } \mathbf{x} \in S$$

as N chosen large



Forward-invariant set of the model dynamics under $\boldsymbol{\pi}_\star$ within set where $V_\star$ is finite

Modification of the ENMPC scheme

Economic NMPC

Policy $\boldsymbol{\pi}(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned} \mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N \hat{T}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 = \mathbf{x} \\ & \hat{\mathbf{h}}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \end{aligned}$$

Is there $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\boldsymbol{\pi}(\mathbf{x}) = \boldsymbol{\pi}_\star(\mathbf{x})$$

even though the model $\mathbf{f}$ is wrong ?

Theorem

If $\hat{T} = 0$, and if for k large

$$\gamma^k V_\star(\mathbf{x}_k) \to 0$$

then there is $\hat{L}$, $\hat{\mathbf{h}}$ s.t.

$$\boldsymbol{\pi}(\mathbf{x}) \to \boldsymbol{\pi}_\star(\mathbf{x}) \text{ on } \mathbf{x} \in S$$

as N chosen large

Results also hold for

- (E)NMPC based on stochastic models (e.g. scenario tree, tube-based etc.)
- $\gamma = 1$, i.e. "classic" (E)NMPC formulations
- Deterministic real system

No free lunch: $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ can be very complex

Outline

Parametrization of the (E)NMPC scheme

NMPC policy: $\pi_{\theta}(\mathbf{x}) = \mathbf{u}_0^{\star}$ where

$$\begin{aligned}\mathbf{u}^{\star}, \mathbf{x}^{\star} = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T_{\theta}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x}\end{aligned}$$

i.e. approximate $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ and give freedom in the model

Parametrization of the (E)NMPC scheme

NMPC policy: $\pi_{\theta}(\mathbf{x}) = \mathbf{u}_0^{\star}$ where

$$\begin{aligned}\mathbf{u}^{\star}, \mathbf{x}^{\star} = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T_{\theta}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\ & \mathbf{h}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x}\end{aligned}$$

i.e. approximate $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ and give freedom in the model

- T_{θ}, L_{θ}, $\mathbf{h}_{\theta}$ are in theory enough
- Adjusting model $\mathbf{f}_{\theta}$ can only help "closing the gap" for limited parametrizations

Parametrization of the (E)NMPC scheme

NMPC policy: $\pi_\theta(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned}\mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T_\theta(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_\theta(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_\theta(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{h}_\theta(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x}\end{aligned}$$

i.e. approximate $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ and give freedom in the model

- T_θ, L_θ, $\mathbf{h}_\theta$ are in theory enough
- Adjusting model $\mathbf{f}_\theta$ can only help "closing the gap" for limited parametrizations

(E)NMPC parameters θ ought to be selected to minimize

$$J(\theta) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k \ell(\mathbf{x}_k, \pi_\theta(\mathbf{x}_k))\right]$$

where $\mathbf{x}_{1,\ldots,\infty}$ result from $\pi_\theta \rightarrow$ real system

rather than via model fitting!!

Parametrization of the (E)NMPC scheme

NMPC policy: $\pi_\theta(\mathbf{x}) = \mathbf{u}_0^\star$ where

$$\begin{aligned}\mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T_\theta(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_\theta(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_\theta(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{h}_\theta(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x}\end{aligned}$$

i.e. approximate $\hat{T}$, $\hat{L}$, $\hat{\mathbf{h}}$ and give freedom in the model

- T_θ, L_θ, $\mathbf{h}_\theta$ are in theory enough
- Adjusting model $\mathbf{f}_\theta$ can only help "closing the gap" for limited parametrizations

(E)NMPC parameters θ ought to be selected to minimize

$$J(\theta) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k \ell(\mathbf{x}_k, \pi_\theta(\mathbf{x}_k))\right]$$

where $\mathbf{x}_{1,\ldots,\infty}$ result from $\pi_\theta \rightarrow$ real system

rather than via model fitting!!

How to do that?

Reinforcement Learning - Core principles

Form function approximators:

$$Q_\theta(\mathbf{x}, \mathbf{u}), \; V_\theta(\mathbf{x}), \; \pi_\theta(\mathbf{x})$$

via ad-hoc parametrization

Reinforcement Learning - Core principles

Form function approximators:

$$Q_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{u}), \; V_{\boldsymbol{\theta}}(\mathbf{x}), \; \boldsymbol{\pi}_{\boldsymbol{\theta}}(\mathbf{x})$$

via ad-hoc parametrization

- Q-**learning methods** adjust $\boldsymbol{\theta}$ to get

$$Q_{\star}(\mathbf{x}, \mathbf{u}) \approx Q_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{u})$$

Yields policy:

$$\boldsymbol{\pi}_{\boldsymbol{\theta}}(x) = \mathrm{a}\min_{\mathbf{u}} \; Q_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{u}) \approx \mathrm{a}\min_{\mathbf{u}} \; Q_{\star}(\mathbf{x}, \mathbf{u}) = \boldsymbol{\pi}_{\star}(x)$$

Reinforcement Learning - Core principles

Form function approximators:

$$Q_\theta(\mathbf{x}, \mathbf{u}), \quad V_\theta(\mathbf{x}), \quad \pi_\theta(\mathbf{x})$$

via ad-hoc parametrization

- Q-**learning methods** adjust θ to get

$$Q_\star(\mathbf{x}, \mathbf{u}) \approx Q_\theta(\mathbf{x}, \mathbf{u})$$

Yields policy:

$$\pi_\theta(x) = \mathrm{a}\min_{\mathbf{u}} \; Q_\theta(\mathbf{x}, \mathbf{u}) \approx \mathrm{a}\min_{\mathbf{u}} \; Q_\star(\mathbf{x}, \mathbf{u}) = \pi_\star(x)$$

- **Policy gradient methods** adjust θ to get

$$\nabla_\theta J(\pi_\theta) = 0$$

yields policy $\pi_\theta(\mathbf{x}) \approx \pi_\star(\mathbf{x})$ directly

Reinforcement Learning - Core principles

Form function approximators:

$$Q_\theta(\mathbf{x}, \mathbf{u}), \quad V_\theta(\mathbf{x}), \quad \pi_\theta(\mathbf{x})$$

via ad-hoc parametrization

- Q-**learning methods** adjust θ to get

$$Q_\star(\mathbf{x}, \mathbf{u}) \approx Q_\theta(\mathbf{x}, \mathbf{u})$$

Yields policy:

$$\pi_\theta(x) = \mathrm{a}\min_{\mathbf{u}} \; Q_\theta(\mathbf{x}, \mathbf{u}) \approx \mathrm{a}\min_{\mathbf{u}} \; Q_\star(\mathbf{x}, \mathbf{u}) = \pi_\star(x)$$

- **Policy gradient methods** adjust θ to get

$$\nabla_\theta J(\pi_\theta) = 0$$

yields policy $\pi_\theta(\mathbf{x}) \approx \pi_\star(\mathbf{x})$ directly

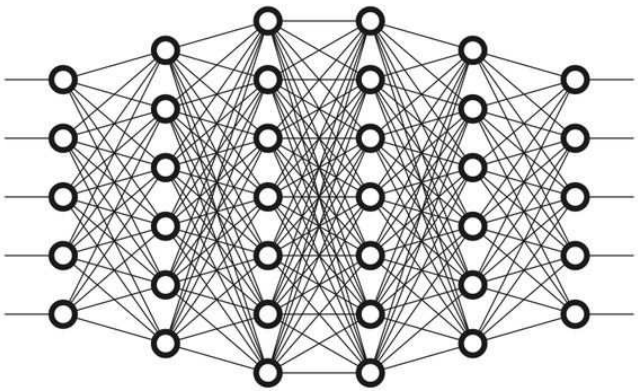
All approaches hinge on building either Q_θ or $\{\pi_\theta, V_\theta\}$

Reinforcement Learning - Core principles

Form function approximators:

$$Q_\theta(\mathbf{x}, \mathbf{u}), \quad V_\theta(\mathbf{x}), \quad \pi_\theta(\mathbf{x})$$

via ad-hoc parametrization



- Q-**learning methods** adjust θ to get

$$Q_\star(\mathbf{x}, \mathbf{u}) \approx Q_\theta(\mathbf{x}, \mathbf{u})$$

Yields policy:

$$\pi_\theta(x) = \mathrm{a}\min_{\mathbf{u}} Q_\theta(\mathbf{x}, \mathbf{u}) \approx \mathrm{a}\min_{\mathbf{u}} Q_\star(\mathbf{x}, \mathbf{u}) = \pi_\star(x)$$

- **Policy gradient methods** adjust θ to get

$$\nabla_\theta J(\pi_\theta) = 0$$

yields policy $\pi_\theta(\mathbf{x}) \approx \pi_\star(\mathbf{x})$ directly

All approaches hinge on building either Q_θ or $\{\pi_\theta, V_\theta\}$

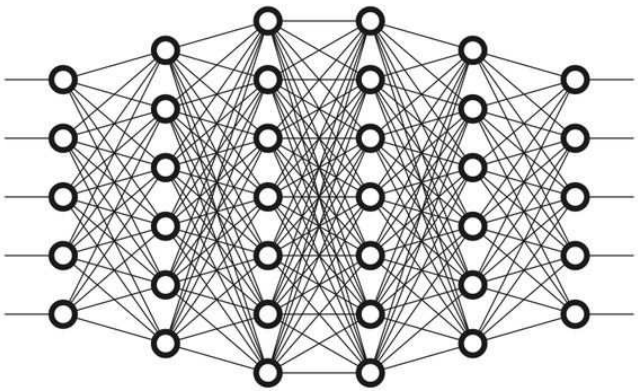
RL typically relies on DNNs as function approximators

Reinforcement Learning - Core principles

Form function approximators:

$$Q_\theta(\mathbf{x}, \mathbf{u}), \quad V_\theta(\mathbf{x}), \quad \pi_\theta(\mathbf{x})$$

via ad-hoc parametrization



- Q-**learning methods** adjust θ to get

$$Q_\star(\mathbf{x}, \mathbf{u}) \approx Q_\theta(\mathbf{x}, \mathbf{u})$$

Yields policy:

$$\pi_\theta(x) = \mathrm{a}\min_{\mathbf{u}} Q_\theta(\mathbf{x}, \mathbf{u}) \approx \mathrm{a}\min_{\mathbf{u}} Q_\star(\mathbf{x}, \mathbf{u}) = \pi_\star(x)$$

- **Policy gradient methods** adjust θ to get

$$\nabla_\theta J(\pi_\theta) = 0$$

yields policy $\pi_\theta(\mathbf{x}) \approx \pi_\star(\mathbf{x})$ directly

All approaches hinge on building either Q_θ or $\{\pi_\theta, V_\theta\}$

RL typically relies on DNNs as function approximators

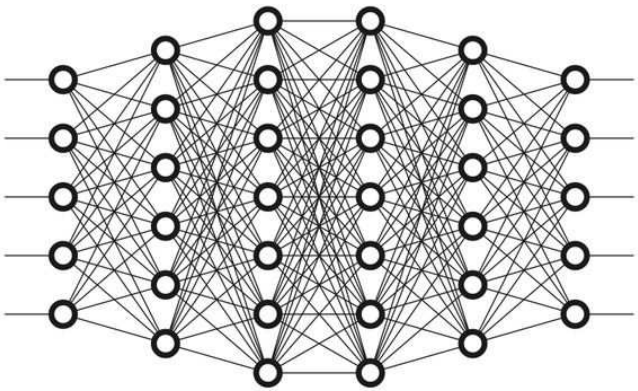
i.e. beyond core principles, the rest is alchemy...

Reinforcement Learning - Core principles

Form function approximators:

$$Q_\theta(\mathbf{x}, \mathbf{u}), \quad V_\theta(\mathbf{x}), \quad \pi_\theta(\mathbf{x})$$

via ad-hoc parametrization



- **Q-learning methods** adjust θ to get

$$Q_\star(\mathbf{x}, \mathbf{u}) \approx Q_\theta(\mathbf{x}, \mathbf{u})$$

Yields policy:

$$\pi_\theta(x) = \mathrm{a}\min_{\mathbf{u}} Q_\theta(\mathbf{x}, \mathbf{u}) \approx \mathrm{a}\min_{\mathbf{u}} Q_\star(\mathbf{x}, \mathbf{u}) = \pi_\star(x)$$

- **Policy gradient methods** adjust θ to get

$$\nabla_\theta J(\pi_\theta) = 0$$

yields policy $\pi_\theta(\mathbf{x}) \approx \pi_\star(\mathbf{x})$ directly

All approaches hinge on building either Q_θ or $\{\pi_\theta, V_\theta\}$

RL typically relies on DNNs as function approximators

i.e. beyond core principles, the rest is alchemy...

Can we be more structured and less obscure?

NMPC as a function approximator

NMPC delivers approximations of V and π

$$V_{\theta}(\mathbf{x}) = \min_{\mathbf{u},\mathbf{x}} \; \gamma^N T_{\theta}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right)$$

$$\mathbf{h}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

$$\pi_{\theta}(\mathbf{x}) = \mathbf{u}_0^{\star} \qquad \text{(ENMPC policy)}$$

NMPC as a function approximator

NMPC delivers approximations of V and π

$$V_{\theta}(\mathbf{x}) = \min_{\mathbf{u},\mathbf{x}} \ \gamma^N T_{\theta}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}(\mathbf{x}_k, \mathbf{u}_k)$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\theta}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{h}_{\theta}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x} \end{aligned}$$

$$\pi_{\theta}(\mathbf{x}) = \mathbf{u}_0^{\star} \qquad \text{(ENMPC policy)}$$

NMPC with constrained input delivers approx. of Q

$$Q_{\theta}(\mathbf{x}, \mathbf{u}) = \min_{\mathbf{u},\mathbf{x}} \ \gamma^N T_{\theta}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}(\mathbf{x}_k, \mathbf{u}_k)$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\theta}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{h}_{\theta}(\mathbf{x}_k, \mathbf{u}_k) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x}, \quad \mathbf{u}_0 = \mathbf{u} \end{aligned}$$

NMPC as a function approximator

NMPC delivers approximations of V and π

$$V_{\theta}(\mathbf{x}) = \min_{\mathbf{u},\mathbf{x}} \ \gamma^N T_{\theta}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}(\mathbf{x}_k, \mathbf{u}_k)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_{\theta}(\mathbf{x}_k, \mathbf{u}_k)$$

$$\mathbf{h}_{\theta}(\mathbf{x}_k, \mathbf{u}_k) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

$$\pi_{\theta}(\mathbf{x}) = \mathbf{u}_0^{\star} \qquad \text{(ENMPC policy)}$$

NMPC with constrained input delivers approx. of Q

$$Q_{\theta}(\mathbf{x}, \mathbf{u}) = \min_{\mathbf{u},\mathbf{x}} \ \gamma^N T_{\theta}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}(\mathbf{x}_k, \mathbf{u}_k)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_{\theta}(\mathbf{x}_k, \mathbf{u}_k)$$

$$\mathbf{h}_{\theta}(\mathbf{x}_k, \mathbf{u}_k) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}, \quad \mathbf{u}_0 = \mathbf{u}$$

DP relationships hold:

$$V_{\theta}(\mathbf{x}) = \min_{\mathbf{u}} \ Q_{\theta}(\mathbf{x}, \mathbf{u})$$

$$\pi_{\theta}(\mathbf{x}) = \arg\min_{\mathbf{u}} \ Q_{\theta}(\mathbf{x}, \mathbf{u})$$

NMPC as a function approximator

NMPC delivers approximations of V and π

$$V_{\theta}(\mathbf{x}) = \min_{\mathbf{u},\mathbf{x}} \ \gamma^N T_{\theta}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}(\mathbf{x}_k, \mathbf{u}_k)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_{\theta}(\mathbf{x}_k, \mathbf{u}_k)$$
$$\mathbf{h}_{\theta}(\mathbf{x}_k, \mathbf{u}_k) \leq 0$$
$$\mathbf{x}_0 = \mathbf{x}$$

$$\pi_{\theta}(\mathbf{x}) = \mathbf{u}_0^{\star} \qquad \text{(ENMPC policy)}$$

NMPC with constrained input delivers approx. of Q

$$Q_{\theta}(\mathbf{x}, \mathbf{u}) = \min_{\mathbf{u},\mathbf{x}} \ \gamma^N T_{\theta}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}(\mathbf{x}_k, \mathbf{u}_k)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_{\theta}(\mathbf{x}_k, \mathbf{u}_k)$$
$$\mathbf{h}_{\theta}(\mathbf{x}_k, \mathbf{u}_k) \leq 0$$
$$\mathbf{x}_0 = \mathbf{x}, \quad \mathbf{u}_0 = \mathbf{u}$$

DP relationships hold:

$$V_{\theta}(\mathbf{x}) = \min_{\mathbf{u}} \ Q_{\theta}(\mathbf{x}, \mathbf{u})$$

$$\pi_{\theta}(\mathbf{x}) = \arg\min_{\mathbf{u}} \ Q_{\theta}(\mathbf{x}, \mathbf{u})$$

Theorem says this is a valid function approximator for RL. Bonus: strong theoretical properties

NMPC as a function approximator

NMPC delivers approximations of V and π

$$V_\theta(\mathbf{x}) = \min_{\mathbf{u},\mathbf{x}} \; \gamma^N T_\theta(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_\theta(\mathbf{x}_k, \mathbf{u}_k)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_\theta(\mathbf{x}_k, \mathbf{u}_k)$$

$$\mathbf{h}_\theta(\mathbf{x}_k, \mathbf{u}_k) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}$$

$$\pi_\theta(\mathbf{x}) = \mathbf{u}_0^\star \qquad \text{(ENMPC policy)}$$

NMPC with constrained input delivers approx. of Q

$$Q_\theta(\mathbf{x}, \mathbf{u}) = \min_{\mathbf{u},\mathbf{x}} \; \gamma^N T_\theta(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_\theta(\mathbf{x}_k, \mathbf{u}_k)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_\theta(\mathbf{x}_k, \mathbf{u}_k)$$

$$\mathbf{h}_\theta(\mathbf{x}_k, \mathbf{u}_k) \leq 0$$

$$\mathbf{x}_0 = \mathbf{x}, \quad \mathbf{u}_0 = \mathbf{u}$$

DP relationships hold:

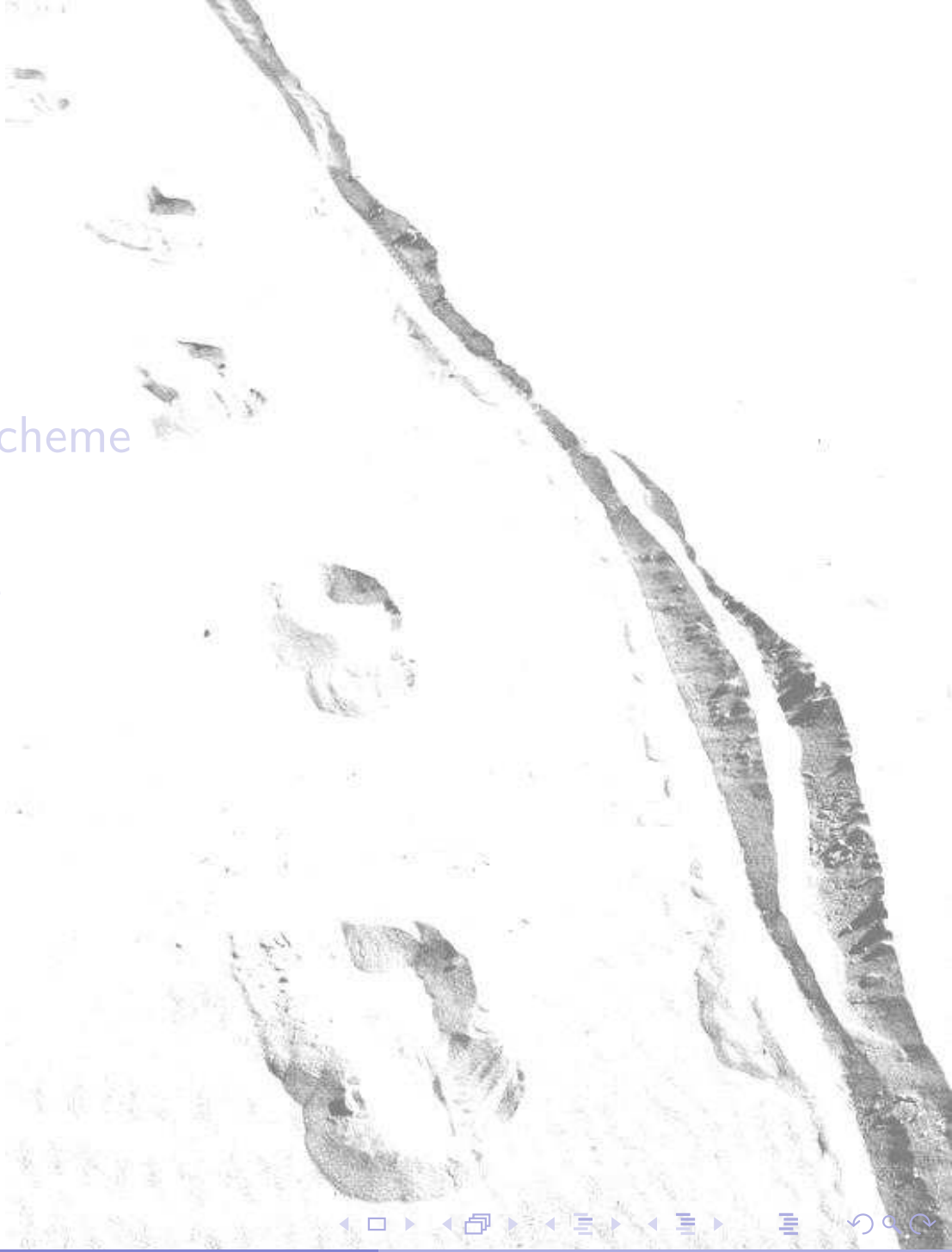
$$V_\theta(\mathbf{x}) = \min_{\mathbf{u}} \; Q_\theta(\mathbf{x}, \mathbf{u})$$

$$\pi_\theta(\mathbf{x}) = \arg\min_{\mathbf{u}} \; Q_\theta(\mathbf{x}, \mathbf{u})$$

Theorem says this is a valid function approximator for RL. Bonus: strong theoretical properties

Use RL(-like) techniques to adjust the parameters θ

Outline

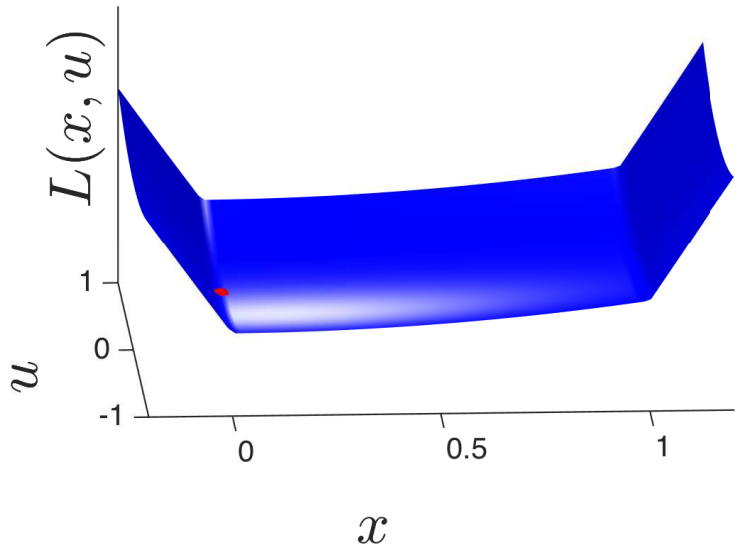


Implementation

"Learning is impossible if mistakes yield an infinite punishment"

(E)NMPC scheme with mixed-constraints relaxation

$$
\begin{aligned}
Q_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{u}) = \min_{\mathbf{x},\mathbf{u},\mathbf{s}\geq 0} \quad & \gamma^N T_{\boldsymbol{\theta}}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \left(L_{\boldsymbol{\theta}}(\mathbf{x}_k,\mathbf{u}_k) + \mathbf{w}^\top \mathbf{s}_k \right) \\
\text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\boldsymbol{\theta}}\left(\mathbf{x}_k,\mathbf{u}_k\right), \qquad \mathbf{x}_0 = \mathbf{x} \\
& \mathbf{h}^{\mathbf{u}}_{\boldsymbol{\theta}}\left(\mathbf{u}_k\right) \leq 0, \qquad \mathbf{h}_{\boldsymbol{\theta}}\left(\mathbf{x}_k,\mathbf{u}_k\right) \leq \mathbf{s}_k
\end{aligned}
$$



'Walls" = relaxation of $x \in [0,\ 1]$

Implementation

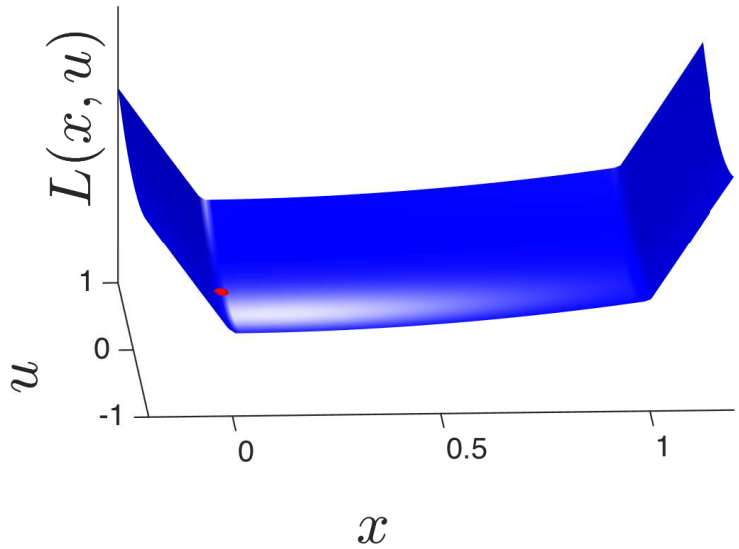
"Learning is impossible if mistakes yield an infinite punishment"

(E)NMPC scheme with mixed-constraints relaxation

$$Q_{\theta}(\mathbf{x},\mathbf{u}) = \min_{\mathbf{x},\mathbf{u},\mathbf{s}\geq 0} \overbrace{\lambda_{\theta}(\mathbf{x})}^{\rightarrow \text{dissipativity}} + \gamma^N T_{\theta}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \left(L_{\theta}(\mathbf{x}_k,\mathbf{u}_k) + \mathbf{w}^{\top}\mathbf{s}_k \right)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right), \qquad \mathbf{x}_0 = \mathbf{x}$$

$$\mathbf{h}^{\mathbf{u}}_{\theta}\left(\mathbf{u}_k\right) \leq 0, \qquad \mathbf{h}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) \leq \mathbf{s}_k$$



'Walls" = relaxation of $x \in [0,\ 1]$

Implementation

"Learning is impossible if mistakes yield an infinite punishment"

(E)NMPC scheme with mixed-constraints relaxation

$$
\begin{aligned}
Q_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{u}) = \min_{\mathbf{x},\mathbf{u},\mathbf{s}\geq 0} \quad & \overbrace{\lambda_{\boldsymbol{\theta}}(\mathbf{x})}^{\rightarrow \text{dissipativity}} + \gamma^N T_{\boldsymbol{\theta}}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \left(L_{\boldsymbol{\theta}}(\mathbf{x}_k,\mathbf{u}_k) + \mathbf{w}^\top \mathbf{s}_k \right) \\
\text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\boldsymbol{\theta}}\left(\mathbf{x}_k,\mathbf{u}_k\right), \qquad \mathbf{x}_0 = \mathbf{x} \\
& \mathbf{h}^{\mathbf{u}}_{\boldsymbol{\theta}}\left(\mathbf{u}_k\right) \leq 0, \qquad \mathbf{h}_{\boldsymbol{\theta}}\left(\mathbf{x}_k,\mathbf{u}_k\right) \leq \mathbf{s}_k
\end{aligned}
$$

TD error

$$\delta_{\boldsymbol{\theta}}\left(\mathbf{x},\mathbf{u},\mathbf{x}_+\right) = L(\mathbf{x},\mathbf{u}) + \gamma V_{\boldsymbol{\theta}}(\mathbf{x}_+) - Q_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{u})$$

E.g. use **LSTDQ**: solve

$$\mathbb{E}_{\tau_{\boldsymbol{\pi}_{\boldsymbol{\theta}}}}\left[\delta \nabla_{\boldsymbol{\theta}} Q_{\boldsymbol{\theta}}\right] = 0$$

for $\boldsymbol{\theta}$. This is on-the-fly LSPI for ENMPC

Implementation

"Learning is impossible if mistakes yield an infinite punishment"

(E)NMPC scheme with mixed-constraints relaxation

$$Q_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{u}) = \min_{\mathbf{x},\mathbf{u},\mathbf{s}\geq 0} \overbrace{\lambda_{\boldsymbol{\theta}}(\mathbf{x})}^{\rightarrow \text{dissipativity}} + \gamma^N T_{\boldsymbol{\theta}}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \left(L_{\boldsymbol{\theta}}(\mathbf{x}_k,\mathbf{u}_k) + \mathbf{w}^\top \mathbf{s}_k \right)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_k,\mathbf{u}_k), \qquad \mathbf{x}_0 = \mathbf{x}$$

$$\mathbf{h}^{\mathbf{u}}_{\boldsymbol{\theta}}(\mathbf{u}_k) \leq 0, \qquad \mathbf{h}_{\boldsymbol{\theta}}(\mathbf{x}_k,\mathbf{u}_k) \leq \mathbf{s}_k$$

TD error

$$\delta_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{u},\mathbf{x}_+) = L(\mathbf{x},\mathbf{u}) + \gamma V_{\boldsymbol{\theta}}(\mathbf{x}_+) - Q_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{u})$$

E.g. use **LSTDQ**: solve

$$\mathbb{E}_{\tau_{\boldsymbol{\pi}_{\boldsymbol{\theta}}}}[\delta \nabla_{\boldsymbol{\theta}} Q_{\boldsymbol{\theta}}] = 0$$

for $\boldsymbol{\theta}$. This is on-the-fly LSPI for ENMPC

Can learn off-policy

- Generate new (E)NMPC $\boldsymbol{\theta}'$ while using current one $\boldsymbol{\theta}$
- Update parameters $\boldsymbol{\theta}' \leftarrow \boldsymbol{\theta}$ after enough data and formal verifications

Implementation

"Learning is impossible if mistakes yield an infinite punishment"

(E)NMPC scheme with mixed-constraints relaxation

$$Q_{\theta}(\mathbf{x},\mathbf{u}) = \min_{\mathbf{x},\mathbf{u},\mathbf{s}\geq 0} \overbrace{\lambda_{\theta}(\mathbf{x})}^{\rightarrow \text{dissipativity}} + \gamma^N T_{\theta}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \left(L_{\theta}(\mathbf{x}_k,\mathbf{u}_k) + \mathbf{w}^{\top}\mathbf{s}_k \right)$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right), \qquad \mathbf{x}_0 = \mathbf{x} \\ & \mathbf{h}^{\mathbf{u}}_{\theta}\left(\mathbf{u}_k\right) \leq 0, \qquad \mathbf{h}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) \leq \mathbf{s}_k \end{aligned}$$

Deterministic policy gradient

$$\min_{\boldsymbol{\theta}} \ \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k \ell\left(\mathbf{x}_k, \boldsymbol{\pi}_{\boldsymbol{\theta}}(\mathbf{x}_k)\right)\right]$$

has necessary condition of optimality:

$$\nabla_{\boldsymbol{\theta}} J = \mathbb{E}_{\tau_{\boldsymbol{\pi}_{\boldsymbol{\theta}}}}\left[\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}}(\mathbf{x}) \nabla_{\mathbf{u}} Q_{\boldsymbol{\pi}_{\boldsymbol{\theta}}}\left(\mathbf{x}, \boldsymbol{\pi}_{\boldsymbol{\theta}}(\mathbf{x})\right)\right] = 0$$

Implementation

"Learning is impossible if mistakes yield an infinite punishment"

(E)NMPC scheme with mixed-constraints relaxation

$$Q_{\theta}(\mathbf{x},\mathbf{u}) = \min_{\mathbf{x},\mathbf{u},\mathbf{s}\geq 0} \overbrace{\lambda_{\theta}(\mathbf{x})}^{\rightarrow \text{dissipativity}} + \gamma^N T_{\theta}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \left(L_{\theta}(\mathbf{x}_k,\mathbf{u}_k) + \mathbf{w}^{\top}\mathbf{s}_k \right)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right), \qquad \mathbf{x}_0 = \mathbf{x}$$

$$\mathbf{h}^{\mathbf{u}}_{\theta}\left(\mathbf{u}_k\right) \leq 0, \qquad \mathbf{h}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) \leq \mathbf{s}_k$$

Deterministic policy gradient

$$\min_{\boldsymbol{\theta}} \ \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k \ell\left(\mathbf{x}_k, \boldsymbol{\pi}_{\boldsymbol{\theta}}(\mathbf{x}_k)\right)\right]$$

has necessary condition of optimality:

$$\nabla_{\boldsymbol{\theta}} J = \mathbb{E}_{\tau_{\boldsymbol{\pi}_{\boldsymbol{\theta}}}}\left[\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}}(\mathbf{x}) \nabla_{\mathbf{u}} Q_{\boldsymbol{\pi}_{\boldsymbol{\theta}}}\left(\mathbf{x}, \boldsymbol{\pi}_{\boldsymbol{\theta}}(\mathbf{x})\right)\right] = 0$$

- Formal condition of optimality for tuned (E)NMPC scheme
- More complex to use than Q-learning

Implementation

"Learning is impossible if mistakes yield an infinite punishment"

(E)NMPC scheme with mixed-constraints relaxation

$$Q_{\theta}(\mathbf{x},\mathbf{u}) = \min_{\mathbf{x},\mathbf{u},\mathbf{s}\geq 0} \overbrace{\lambda_{\theta}(\mathbf{x})}^{\rightarrow \text{dissipativity}} + \gamma^N T_{\theta}(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k \left(L_{\theta}(\mathbf{x}_k,\mathbf{u}_k) + \mathbf{w}^{\top}\mathbf{s}_k \right)$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right), \qquad \mathbf{x}_0 = \mathbf{x} \\ & \mathbf{h}^{\mathbf{u}}_{\theta}\left(\mathbf{u}_k\right) \leq 0, \qquad \mathbf{h}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) \leq \mathbf{s}_k \end{aligned}$$

Theorem: if real dynamics + stage cost L are dissipative, then

- can use positive-definite L_{θ} and correction λ_{θ} such that ENMPC yields $\pi_{\star}$, $V_{\star}$, $Q_{\star}$.
- λ_{θ} **is** dissipativity function
- ENMPC scheme is nominal-stable by construction

(valid for deterministic dynamics, should be extended)

Sensitivities

RL methods (may) require $\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}}$, $\nabla_{\mathbf{u}} Q_{\boldsymbol{\theta}}$, $\nabla_{\boldsymbol{\theta}} Q_{\boldsymbol{\theta}}$, $\nabla^2_{\boldsymbol{\theta}} Q_{\boldsymbol{\theta}}$, etc. how to get that?

Sensitivities

RL methods (may) require $\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}}$, $\nabla_{\mathbf{u}} Q_{\boldsymbol{\theta}}$, $\nabla_{\boldsymbol{\theta}} Q_{\boldsymbol{\theta}}$, $\nabla^2_{\boldsymbol{\theta}} Q_{\boldsymbol{\theta}}$, etc. how to get that?

NMPC delivers approx. of Q

$$
\begin{aligned}
Q_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{u}) = \min_{\mathbf{u}, \mathbf{x}} \quad & \gamma^N T_{\boldsymbol{\theta}}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\boldsymbol{\theta}}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\
\text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\boldsymbol{\theta}}\left(\mathbf{x}_k, \mathbf{u}_k\right), \quad \mathbf{x}_0 = \mathbf{x} \\
& \mathbf{h}_{\boldsymbol{\theta}}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0 \\
& \mathbf{u}_0 = \mathbf{u}
\end{aligned}
$$

Sensitivities

RL methods (may) require $\nabla_{\theta}\pi_{\theta}$, $\nabla_{\mathbf{u}}Q_{\theta}$, $\nabla_{\theta}Q_{\theta}$, $\nabla^2_{\theta}Q_{\theta}$, etc. how to get that?

NMPC delivers approx. of Q

$$
\begin{aligned}
Q_{\theta}(\mathbf{x},\mathbf{u}) = \min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T_{\theta}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) \\
\text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right), \quad \mathbf{x}_0 = \mathbf{x} \\
& \mathbf{h}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) \leq 0 \\
& \mathbf{u}_0 = \mathbf{u}
\end{aligned}
$$

Consider the Lagrange function:

$$
\begin{aligned}
\mathcal{L} = & \gamma^N T_{\theta}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) + \sum_{k=1}^{N} \boldsymbol{\lambda}_k^{\top}\left(\mathbf{x}_{k+1} - \mathbf{f}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right)\right) + \sum_{k=0}^{N-1} \boldsymbol{\mu}_k^{\top}\mathbf{h}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) \\
& + \boldsymbol{\lambda}_0^{\top}\left(\mathbf{x}_0 - \mathbf{x}\right) + \boldsymbol{\nu}^{\top}\left(\mathbf{u}_0 - \mathbf{u}\right)
\end{aligned}
$$

Sensitivities

RL methods (may) require $\nabla_{\theta}\pi_{\theta}$, $\nabla_{\mathbf{u}}Q_{\theta}$, $\nabla_{\theta}Q_{\theta}$, $\nabla^2_{\theta}Q_{\theta}$, etc. how to get that?

NMPC delivers approx. of Q

$$
\begin{aligned}
Q_{\theta}(\mathbf{x},\mathbf{u}) = \min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T_{\theta}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) \\
\text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right), \quad \mathbf{x}_0 = \mathbf{x} \\
& \mathbf{h}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) \leq 0 \\
& \mathbf{u}_0 = \mathbf{u}
\end{aligned}
$$

Consider the Lagrange function:

$$
\begin{aligned}
\mathcal{L} = & \gamma^N T_{\theta}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) + \sum_{k=1}^{N} \boldsymbol{\lambda}_k^{\top}\left(\mathbf{x}_{k+1} - \mathbf{f}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right)\right) + \sum_{k=0}^{N-1} \boldsymbol{\mu}_k^{\top}\mathbf{h}_{\theta}\left(\mathbf{x}_k,\mathbf{u}_k\right) \\
& + \boldsymbol{\lambda}_0^{\top}\left(\mathbf{x}_0 - \mathbf{x}\right) + \boldsymbol{\nu}^{\top}\left(\mathbf{u}_0 - \mathbf{u}\right)
\end{aligned}
$$

Then

$$\nabla_{\theta}Q_{\theta}(\mathbf{x},\mathbf{u}) = \nabla_{\theta}\mathcal{L}, \quad \nabla_{\mathbf{u}}Q_{\theta} = \nabla_{\mathbf{u}}\mathcal{L}$$

... evaluated on the primal-dual solution of the ENMPC scheme

Sensitivities

RL methods (may) require $\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}}$, $\nabla_{\mathbf{u}} Q_{\boldsymbol{\theta}}$, $\nabla_{\boldsymbol{\theta}} Q_{\boldsymbol{\theta}}$, $\nabla^2_{\boldsymbol{\theta}} Q_{\boldsymbol{\theta}}$, etc. how to get that?

NMPC delivers approx. of V

$$
\begin{aligned}
V_{\boldsymbol{\theta}}(\mathbf{x}) = \min_{\mathbf{u},\mathbf{x}} \quad & \gamma^N T_{\boldsymbol{\theta}}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\boldsymbol{\theta}}\left(\mathbf{x}_k, \mathbf{u}_k\right) \\
\text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}_{\boldsymbol{\theta}}\left(\mathbf{x}_k, \mathbf{u}_k\right), \quad \mathbf{x}_0 = \mathbf{x} \\
& \mathbf{h}_{\boldsymbol{\theta}}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0
\end{aligned}
$$

Sensitivities

RL methods (may) require $\nabla_{\theta}\pi_{\theta}$, $\nabla_{\mathbf{u}}Q_{\theta}$, $\nabla_{\theta}Q_{\theta}$, $\nabla^2_{\theta}Q_{\theta}$, etc. how to get that?

NMPC delivers approx. of V

$$V_{\theta}(\mathbf{x}) = \min_{\mathbf{u},\mathbf{x}} \; \gamma^N T_{\theta}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right), \quad \mathbf{x}_0 = \mathbf{x}$$

$$\mathbf{h}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

Consider the Lagrange function:

$$\mathcal{L} = \gamma^N T_{\theta}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right) + \sum_{k=1}^{N} \boldsymbol{\lambda}_k^{\top}\left(\mathbf{x}_{k+1} - \mathbf{f}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right)\right) + \sum_{k=0}^{N-1} \boldsymbol{\mu}_k^{\top} \mathbf{h}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right)$$

$$+ \boldsymbol{\lambda}_0^{\top}\left(\mathbf{x}_0 - \mathbf{x}\right)$$

Sensitivities

RL methods (may) require $\nabla_{\theta}\pi_{\theta}$, $\nabla_{\mathbf{u}}Q_{\theta}$, $\nabla_{\theta}Q_{\theta}$, $\nabla^2_{\theta}Q_{\theta}$, etc. how to get that?

NMPC delivers approx. of V

$$V_{\theta}(\mathbf{x}) = \min_{\mathbf{u},\mathbf{x}} \; \gamma^N T_{\theta}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right), \quad \mathbf{x}_0 = \mathbf{x}$$

$$\mathbf{h}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

Consider the Lagrange function:

$$\mathcal{L} = \gamma^N T_{\theta}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right) + \sum_{k=1}^{N} \boldsymbol{\lambda}_k^{\top}\left(\mathbf{x}_{k+1} - \mathbf{f}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right)\right) + \sum_{k=0}^{N-1} \boldsymbol{\mu}_k^{\top} \mathbf{h}_{\theta}\left(\mathbf{x}_k, \mathbf{u}_k\right)$$

$$+ \boldsymbol{\lambda}_0^{\top}\left(\mathbf{x}_0 - \mathbf{x}\right)$$

Then

$$\nabla_{\theta} V_{\theta}(\mathbf{x}) = \nabla_{\theta}\mathcal{L}$$

... evaluated on the primal-dual solution of the ENMPC scheme

Sensitivities

RL methods (may) require $\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}}$, $\nabla_{\mathbf{u}} Q_{\boldsymbol{\theta}}$, $\nabla_{\boldsymbol{\theta}} Q_{\boldsymbol{\theta}}$, $\nabla^2_{\boldsymbol{\theta}} Q_{\boldsymbol{\theta}}$, etc. how to get that?

NMPC delivers policy $\boldsymbol{\pi}_{\boldsymbol{\theta}}$

$$\boldsymbol{\pi}_{\boldsymbol{\theta}}(\mathbf{x}) = \mathbf{u}_0^\star$$

$$\mathbf{u}^\star, \mathbf{x}^\star = \arg\min_{\mathbf{u},\mathbf{x}} \quad \gamma^N T_{\boldsymbol{\theta}}\left(\mathbf{x}_N\right) + \sum_{k=0}^{N-1} \gamma^k L_{\boldsymbol{\theta}}\left(\mathbf{x}_k, \mathbf{u}_k\right)$$

$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{f}_{\boldsymbol{\theta}}\left(\mathbf{x}_k, \mathbf{u}_k\right), \quad \mathbf{x}_0 = \mathbf{x}$$

$$\mathbf{h}_{\boldsymbol{\theta}}\left(\mathbf{x}_k, \mathbf{u}_k\right) \leq 0$$

Sensitivity $\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}_{\boldsymbol{\theta}}(\mathbf{x}) = \nabla_{\boldsymbol{\theta}} \mathbf{u}_0^\star$, computed from:

$$\frac{\partial \mathbf{z}}{\partial \boldsymbol{\theta}} = -\frac{\partial \mathbf{R}}{\partial \mathbf{z}}^{-1} \frac{\partial \mathbf{R}}{\partial \boldsymbol{\theta}}$$

where

- $\mathbf{R}$ are the KKT conditions of ENMPC
- $\mathbf{z}$ collects the primal-dual variables
- Best treated in an primal-dual interior-point framework with "large" barrier parameter

Second-order sensitivities e.g.

$$\nabla^2_{\boldsymbol{\theta}} V_{\boldsymbol{\theta}} = \frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta}^2} + \frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta} \partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \boldsymbol{\theta}}$$

Outline

Q-learning on our trivial example

Modified NMPC scheme

$$\pi(x) \leftarrow \min_{u,x,s} \quad \gamma^N V_\star(x_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(x,u)$$

$$\text{s.t.} \quad x_{k+1} = x_k + 0.1 u_k$$

$$-1 \leq u_k \leq 1$$

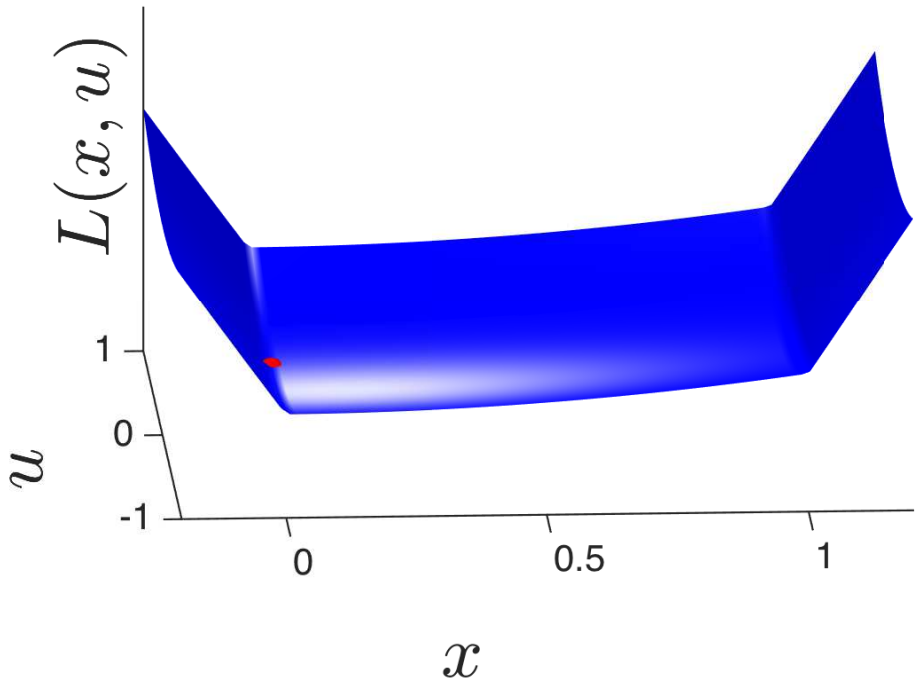
$$x_0 = x$$

Real dynamics:

$$x_+ = x + 0.1u + e$$

Noise:

$$e \sim \mathcal{U}([-0.1, 0])$$



Q-learning on our trivial example

Modified NMPC scheme

$$\pi(x) \leftarrow \min_{u,x,s} \quad \gamma^N V_\star(x_N) + \sum_{k=0}^{N-1} \gamma^k \hat{L}(x,u)$$

$$\text{s.t.} \quad x_{k+1} = x_k + 0.1 u_k$$

$$-1 \leq u_k \leq 1$$

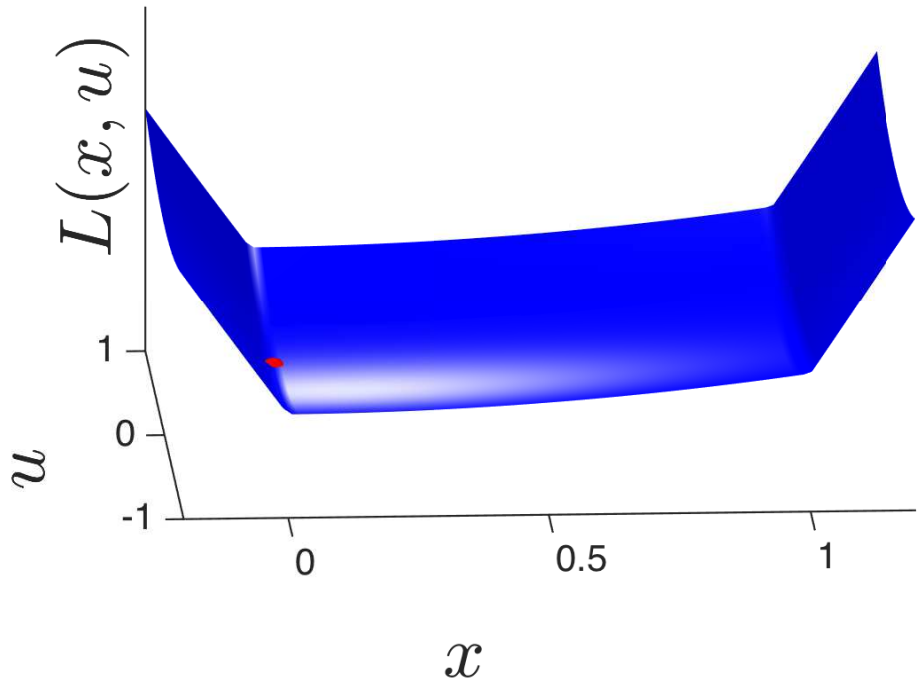
$$x_0 = x$$

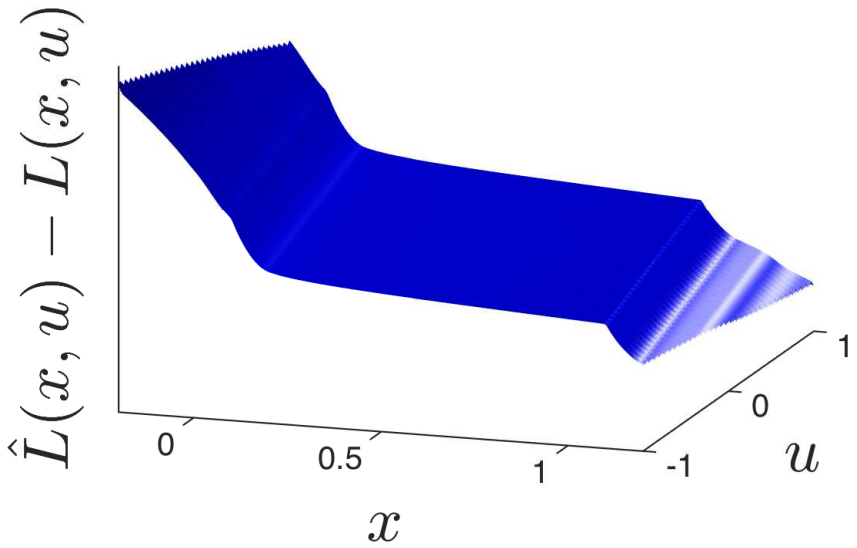
Real dynamics:

$$x_+ = x + 0.1u + e$$

Noise:

$$e \sim \mathcal{U}([-0.1, 0])$$





Q-learning on our trivial example

NMPC with relaxed constraints ($\equiv$ "walls")

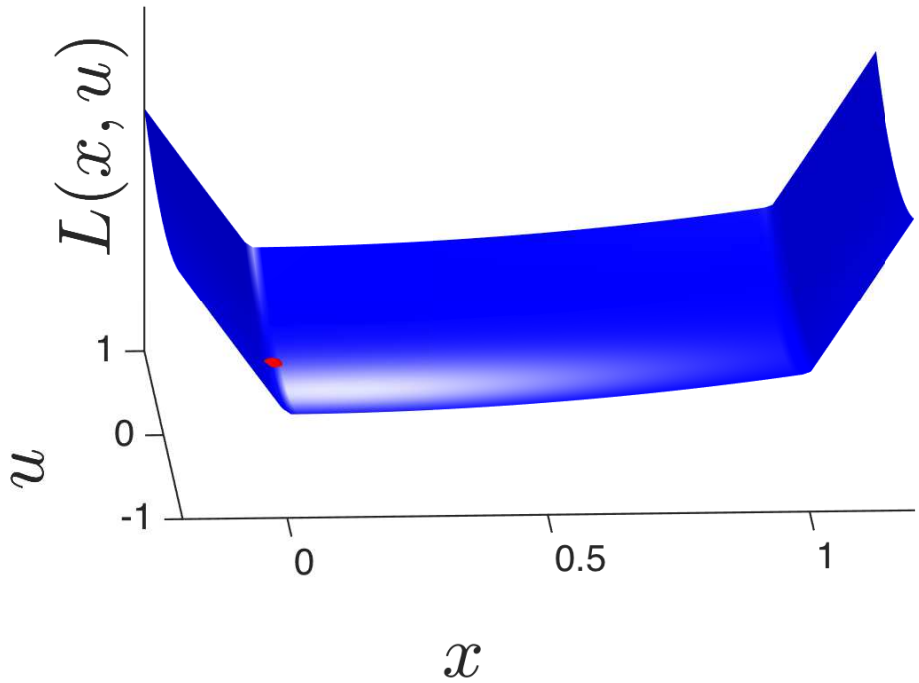
$$\begin{aligned}
\min_{u,x,s} \quad & c + \gamma^N T x_N^2 + \sum_{k=0}^{N-1} \gamma^k \left(x_k^2 + u_k^2 + w s_k + w s_k^2 + \mathbf{f}^\top \begin{bmatrix} x_k \\ u_k \end{bmatrix} \right) \\
\text{s.t.} \quad & x_{k+1} = x_k + 0.1 u_k + b \\
& -1 \leq u_k \leq 1, \quad \underline{x} - s_k \leq x_k \leq \bar{x} + s_k \\
& x_0 = x
\end{aligned}$$

Real dynamics:

$$x_+ = x + 0.1u + e$$

Noise:

$$e \sim \mathcal{U}([-0.1, 0])$$



Q-learning can tune all parameters:

- Model bias b
- Cost gradient $\mathbf{f}$
- State constraints (walls position) $\underline{x}$, $\bar{x}$
- Terminal cost T

Q-learning on our trivial example

NMPC with relaxed constraints ($\equiv$ "walls")

$$\min_{u,x,s} \quad c + \gamma^N T x_N^2 + \sum_{k=0}^{N-1} \gamma^k \left(x_k^2 + u_k^2 + w s_k + w s_k^2 + \mathbf{f}^\top \begin{bmatrix} x_k \\ u_k \end{bmatrix} \right)$$

$$\text{s.t.} \quad x_{k+1} = x_k + 0.1 u_k + b$$

$$-1 \leq u_k \leq 1, \quad \underline{x} - s_k \leq x_k \leq \bar{x} + s_k$$

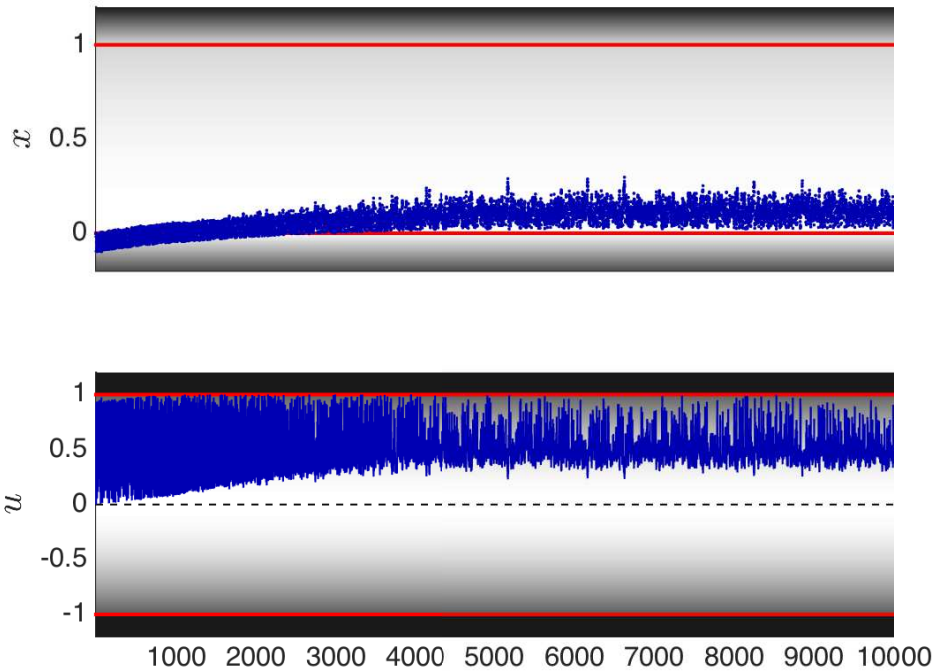
$$x_0 = x$$

Real dynamics:

$$x_+ = x + 0.1u + e$$

Noise:

$$e \sim \mathcal{U}([-0.1, 0])$$



Q-learning on our trivial example

NMPC with relaxed constraints ($\equiv$ "walls")

$$\min_{u,x,s} \quad c + \gamma^N T x_N^2 + \sum_{k=0}^{N-1} \gamma^k \left(x_k^2 + u_k^2 + w s_k + w s_k^2 + \mathbf{f}^\top \begin{bmatrix} x_k \\ u_k \end{bmatrix} \right)$$

$$\text{s.t.} \quad x_{k+1} = x_k + 0.1 u_k + b$$

$$-1 \leq u_k \leq 1, \quad \underline{x} - s_k \leq x_k \leq \bar{x} + s_k$$

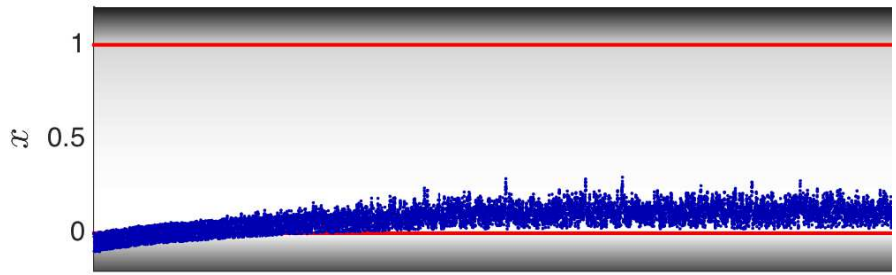
$$x_0 = x$$

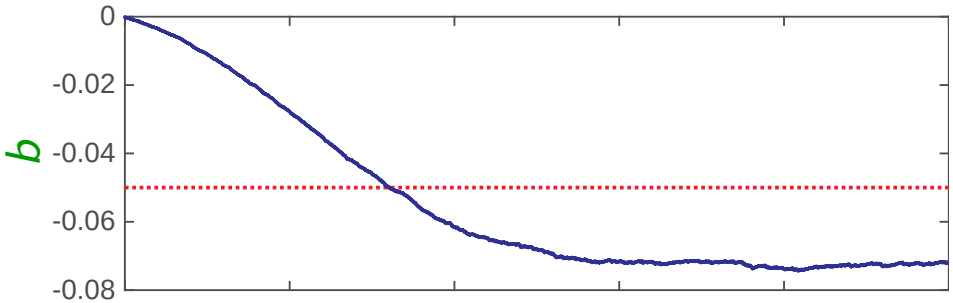
Real dynamics:

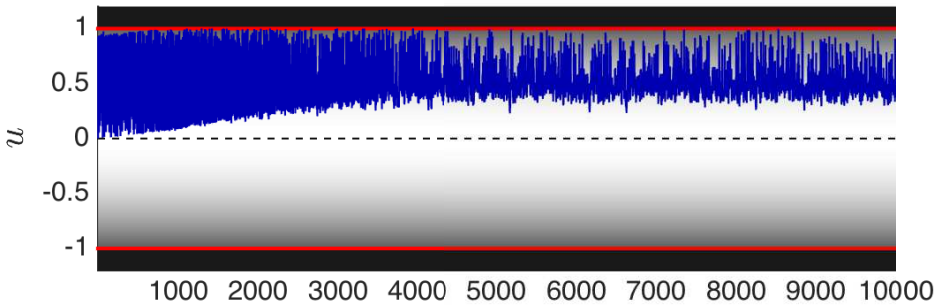
$$x_+ = x + 0.1u + e$$

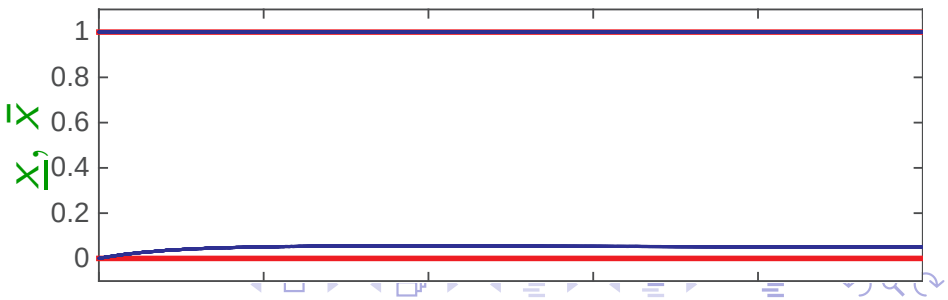
Noise:

$$e \sim \mathcal{U}([-0.1, 0])$$









Q-learning on our trivial example

NMPC with relaxed constraints ($\equiv$ "walls")

$$\min_{u,x,s} \quad c + \gamma^N T x_N^2 + \sum_{k=0}^{N-1} \gamma^k \left(x_k^2 + u_k^2 + w s_k + w s_k^2 + \mathbf{f}^\top \begin{bmatrix} x_k \\ u_k \end{bmatrix} \right)$$

$$\text{s.t.} \quad x_{k+1} = x_k + 0.1 u_k + b$$

$$-1 \leq u_k \leq 1, \quad \underline{x} - s_k \leq x_k \leq \bar{x} + s_k$$

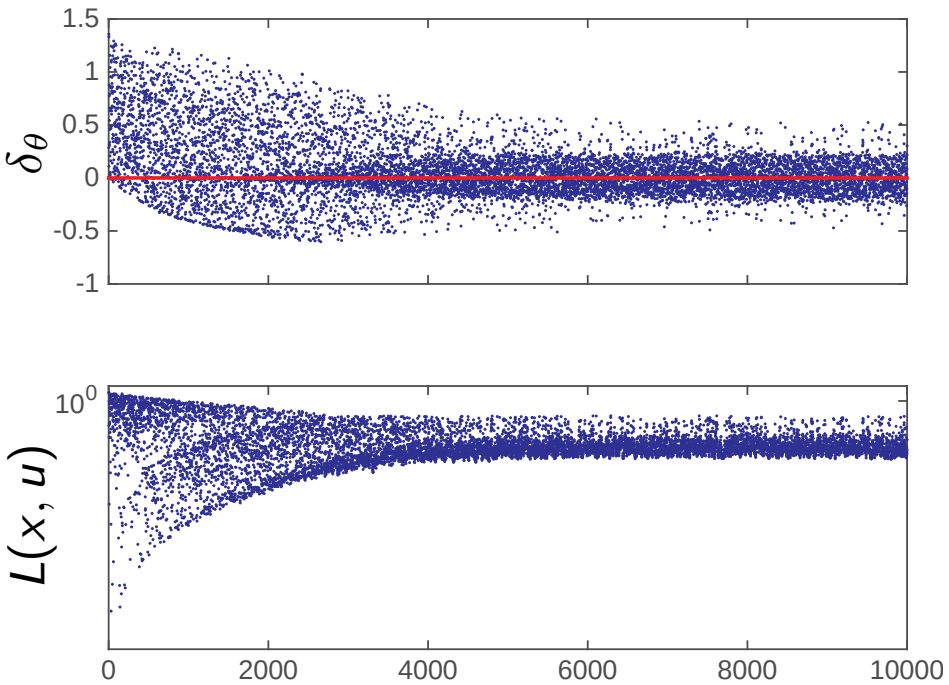
$$x_0 = x$$

Real dynamics:

$$x_+ = x + 0.1u + e$$

Noise:

$$e \sim \mathcal{U}([-0.1, 0])$$



Q-learning on our trivial example

NMPC with relaxed constraints ($\equiv$ "walls")

$$\min_{u,x,s} \quad c + \gamma^N T x_N^2 + \sum_{k=0}^{N-1} \gamma^k \left(x_k^2 + u_k^2 + w s_k + w s_k^2 + \mathbf{f}^\top \begin{bmatrix} x_k \\ u_k \end{bmatrix} \right)$$

$$\text{s.t.} \quad x_{k+1} = x_k + 0.1 u_k + b$$

$$-1 \leq u_k \leq 1, \quad \underline{x} - s_k \leq x_k \leq \bar{x} + s_k$$

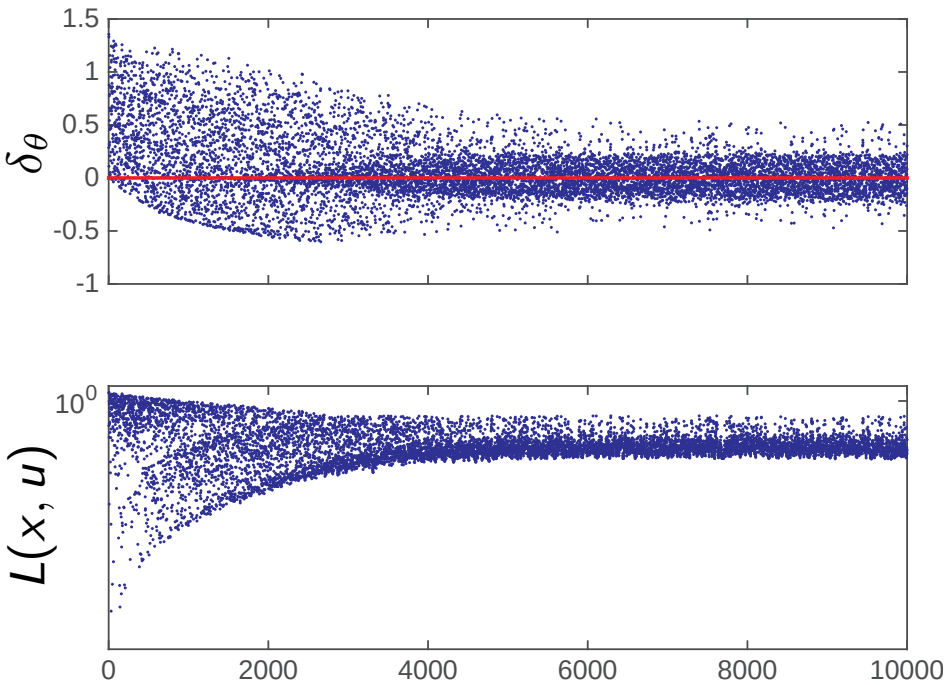
$$x_0 = x$$

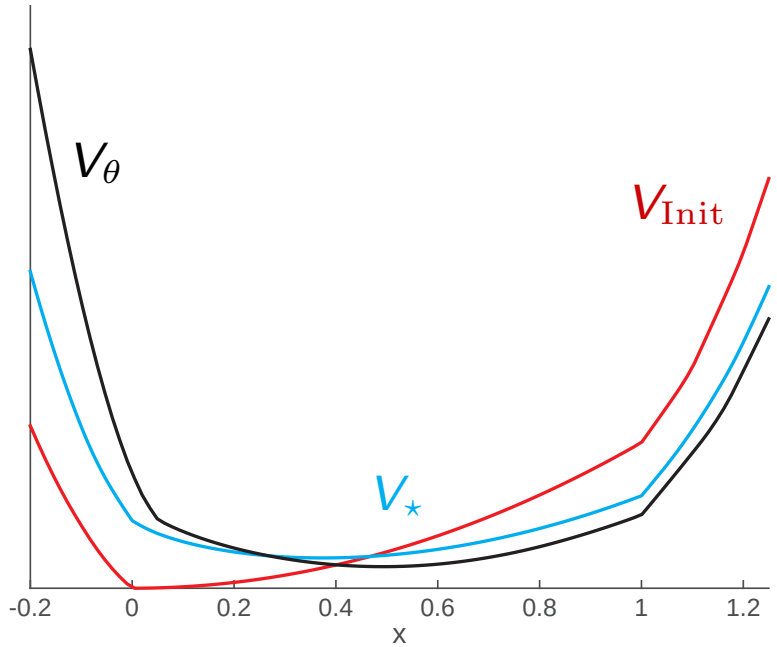
Real dynamics:

$$x_+ = x + 0.1u + e$$

Noise:

$$e \sim \mathcal{U}([-0.1, 0])$$





Q-learning on our trivial example

NMPC with relaxed constraints ($\equiv$ "walls")

$$\min_{u,x,s} \quad c + \gamma^N T x_N^2 + \sum_{k=0}^{N-1} \gamma^k \left(x_k^2 + u_k^2 + w s_k + w s_k^2 + \mathbf{f}^\top \begin{bmatrix} x_k \\ u_k \end{bmatrix} \right)$$

$$\text{s.t.} \quad x_{k+1} = x_k + 0.1 u_k + b$$

$$-1 \leq u_k \leq 1, \quad \underline{x} - s_k \leq x_k \leq \bar{x} + s_k$$

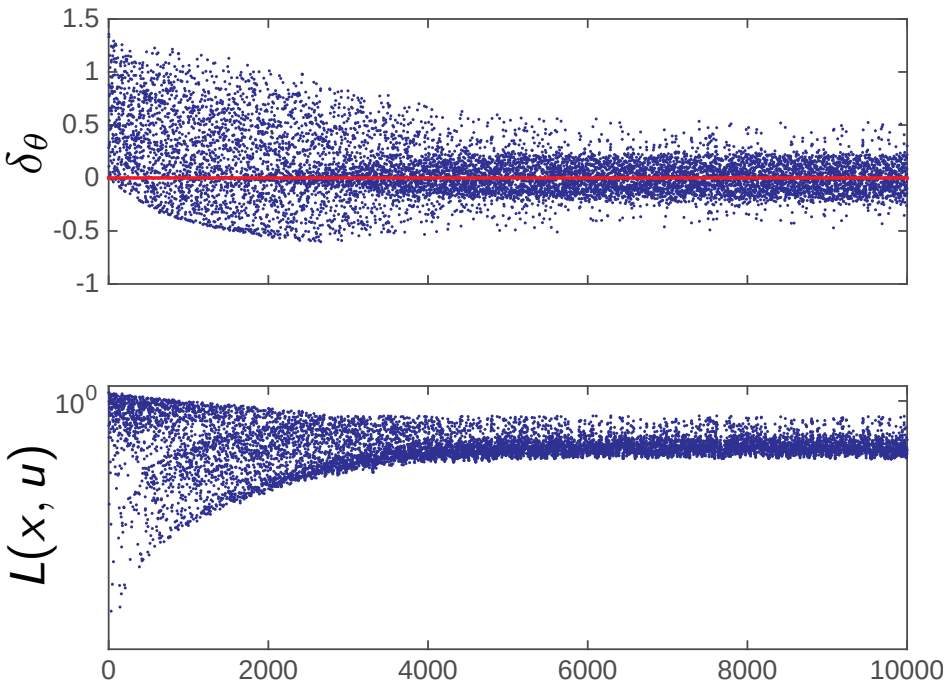
$$x_0 = x$$

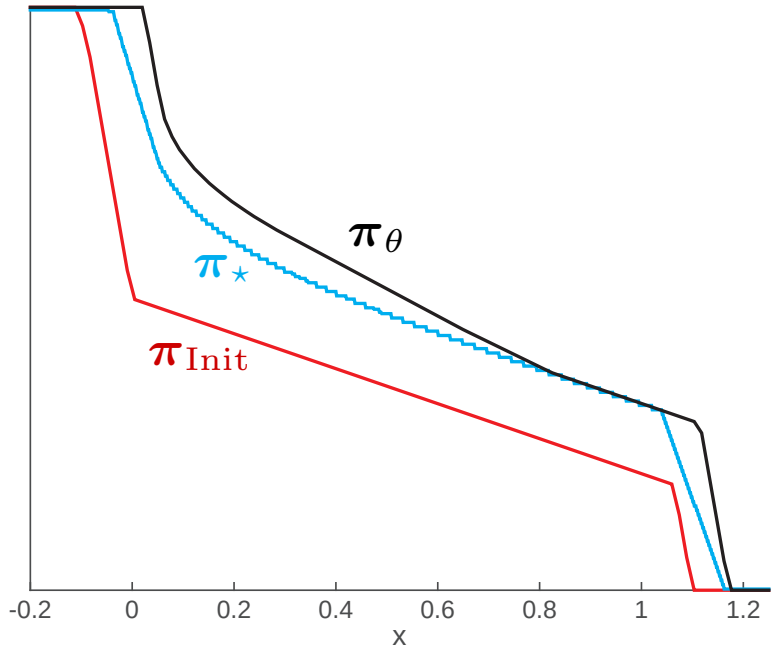
Real dynamics:

$$x_+ = x + 0.1u + e$$

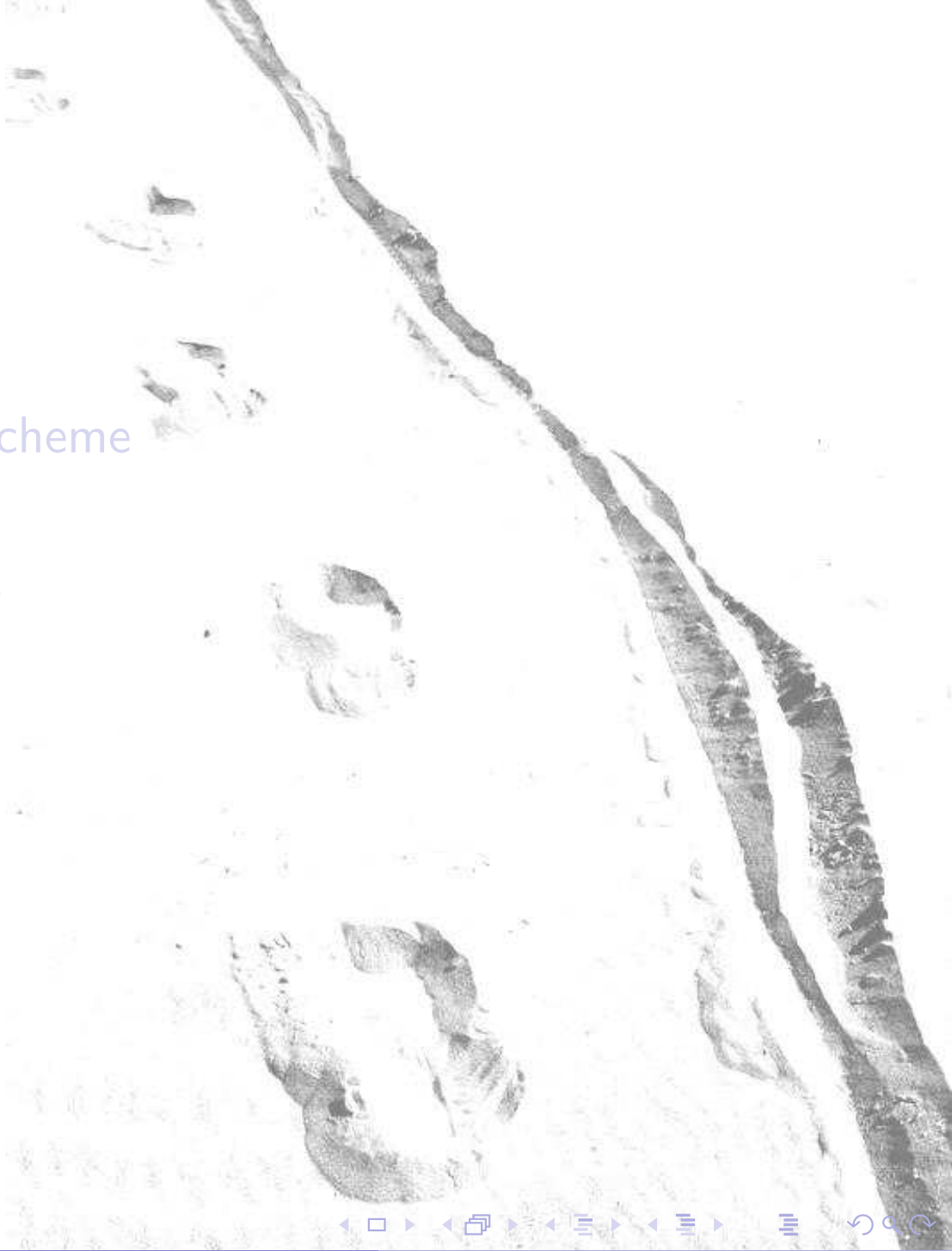
Noise:

$$e \sim \mathcal{U}([-0.1, 0])$$





Outline



Where is SYSID then?

Tune parameters such that $\nabla_\theta J(\pi_\theta) = 0$. Model fitting does not matter?

Where is SYSID then?

Tune parameters such that $\nabla_\theta J(\pi_\theta) = 0$. Model fitting does not matter?

- Model fitting the data is probably a good place to start with

Where is SYSID then?

Tune parameters such that $\nabla_\theta J(\pi_\theta) = 0$. Model fitting does not matter?

- Model fitting the data is probably a good place to start with
- $\nabla^2_\theta Q_\theta$ and $\nabla^2_\theta J(\pi_\theta)$ are (typically) rank-deficient. Some parameters are left to be adjusted $\rightarrow$ SYSID as "regularization"

Where is SYSID then?

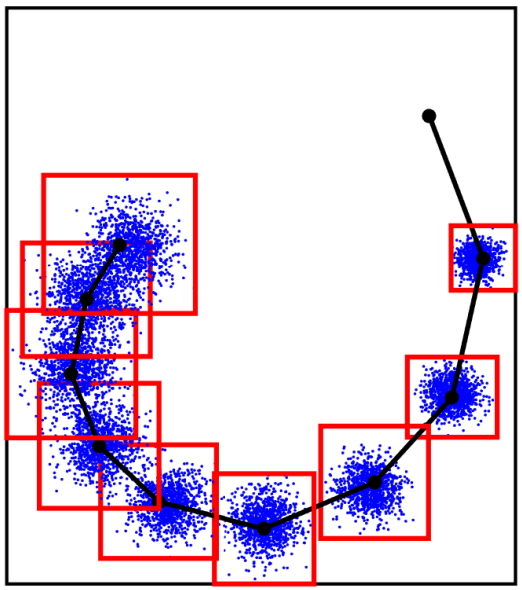
Tune parameters such that $\nabla_\theta J(\pi_\theta) = 0$. Model fitting does not matter?

- Model fitting the data is probably a good place to start with
- $\nabla^2_\theta Q_\theta$ and $\nabla^2_\theta J(\pi_\theta)$ are (typically) rank-deficient. Some parameters are left to be adjusted $\rightarrow$ SYSID as "regularization"
- Safe learning...

Where is SYSID then?

Tune parameters such that $\nabla_\theta J(\pi_\theta) = 0$. Model fitting does not matter?

- Model fitting the data is probably a good place to start with
- $\nabla^2_\theta Q_\theta$ and $\nabla^2_\theta J(\pi_\theta)$ are (typically) rank-deficient. Some parameters are left to be adjusted $\rightarrow$ SYSID as "regularization"
- Safe learning...



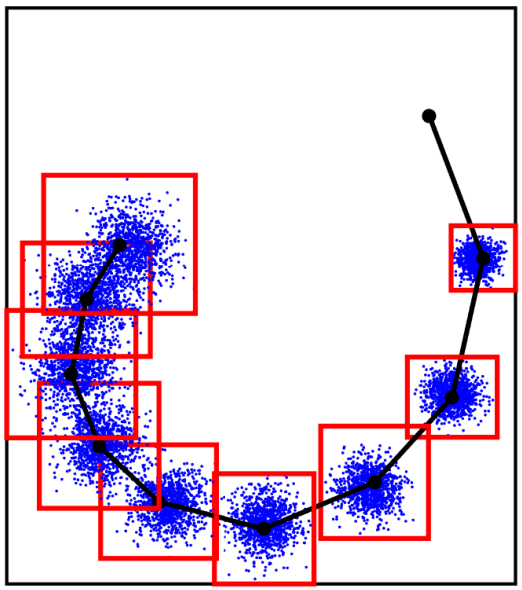
Deterministic "forward-set" model:

$$\mathbf{x}, \mathbf{u} \rightarrow \mathbf{x}_+ \in X^+_\theta(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta})$$

Where is SYSID then?

Tune parameters such that $\nabla_\theta J(\pi_\theta) = 0$. Model fitting does not matter?

- Model fitting the data is probably a good place to start with
- $\nabla^2_\theta Q_\theta$ and $\nabla^2_\theta J(\pi_\theta)$ are (typically) rank-deficient. Some parameters are left to be adjusted $\rightarrow$ SYSID as "regularization"
- Safe learning...



Deterministic "forward-set" model:

$$\mathbf{x}, \mathbf{u} \rightarrow \mathbf{x}_+ \in X^+_\theta(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta})$$

Robust ENMPC optimizes under constraint:

$$\min_{\mathbf{u},\mathbf{x}} \ \gamma^N T_\theta(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_\theta(\mathbf{x}_k, \mathbf{u}_k)$$

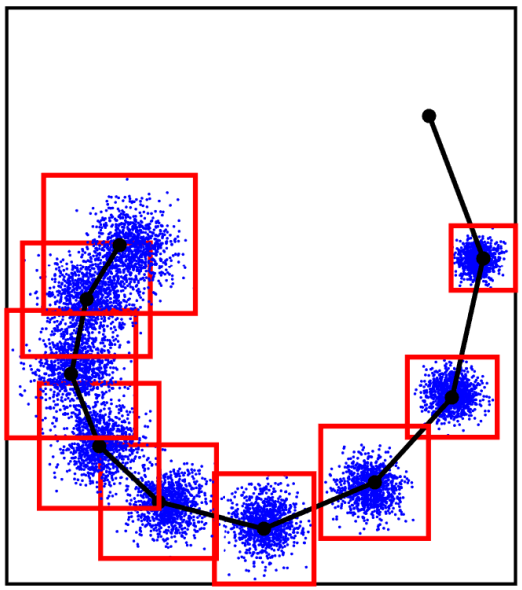
$$\text{s.t.} \ \ X^0_\theta = \mathbf{x}, \quad X^k_\theta \subseteq \mathbb{X}$$

where X^k_θ is the set-propagation of the uncertainties.

Where is SYSID then?

Tune parameters such that $\nabla_\theta J(\pi_\theta) = 0$. Model fitting does not matter?

- Model fitting the data is probably a good place to start with
- $\nabla^2_\theta Q_\theta$ and $\nabla^2_\theta J(\pi_\theta)$ are (typically) rank-deficient. Some parameters are left to be adjusted $\rightarrow$ SYSID as "regularization"
- Safe learning...



Deterministic "forward-set" model:

$$\mathbf{x}, \mathbf{u} \rightarrow \mathbf{x}_+ \in X^+_\theta(\mathbf{x}, \mathbf{u}, \theta)$$

Robust ENMPC optimizes under constraint:

$$\min_{\mathbf{u},\mathbf{x}} \ \gamma^N T_\theta(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L_\theta(\mathbf{x}_k, \mathbf{u}_k)$$

$$\text{s.t.} \quad X^0_\theta = \mathbf{x}, \quad X^k_\theta \subseteq \mathbb{X}$$

where X^k_θ is the set-propagation of the uncertainties.

Adjust θ to ensure X^+_θ "encloses the data", and minimize $J(\pi_\theta)$

Conclusions & Future work

Conclusions

- Cost function (and constraints) of an ENMPC scheme ought not necessarily match the ones describing the problem, should be adjusted

Data-driven Economic NMPC using Reinforcement Learning, S. Gros, M. Zanon, Transaction on Automatic Control
A Data-Efficient Framework for Learning NMPC controllers, M. Zanon, S. Gros, A. Bemporad, ECC 2019 Deterministic Policy Gradient for Data-Driven Economic NMPC, S. Gros, M. Zanon
System Identification and Reinforcement Learning for Data-Driven Economic NMPC, S. Gros, M. Zanon

Conclusions & Future work

Conclusions

- Cost function (and constraints) of an ENMPC scheme ought not necessarily match the ones describing the problem, should be adjusted
- Adjusting ENMPC scheme for performance $\neq$ model fitting

Data-driven Economic NMPC using Reinforcement Learning, S. Gros, M. Zanon, Transaction on Automatic Control
A Data-Efficient Framework for Learning NMPC controllers, M. Zanon, S. Gros, A. Bemporad, ECC 2019 Deterministic Policy Gradient for Data-Driven Economic NMPC, S. Gros, M. Zanon
System Identification and Reinforcement Learning for Data-Driven Economic NMPC, S. Gros, M. Zanon

Conclusions & Future work

Conclusions

- Cost function (and constraints) of an ENMPC scheme ought not necessarily match the ones describing the problem, should be adjusted
- Adjusting ENMPC scheme for performance $\neq$ model fitting
- Necessary Conditions of Optimality of an ENMPC scheme are defined by policy gradient theorems. Can be deployed to tune ENMPC schemes.

Data-driven Economic NMPC using Reinforcement Learning, S. Gros, M. Zanon, Transaction on Automatic Control
A Data-Efficient Framework for Learning NMPC controllers, M. Zanon, S. Gros, A. Bemporad, ECC 2019 Deterministic Policy Gradient for Data-Driven Economic NMPC, S. Gros, M. Zanon
System Identification and Reinforcement Learning for Data-Driven Economic NMPC, S. Gros, M. Zanon

Conclusions & Future work

Conclusions

- Cost function (and constraints) of an ENMPC scheme ought not necessarily match the ones describing the problem, should be adjusted
- Adjusting ENMPC scheme for performance $\neq$ model fitting
- Necessary Conditions of Optimality of an ENMPC scheme are defined by policy gradient theorems. Can be deployed to tune ENMPC schemes.
- ENMPC schemes can be used as a structured function approximation for Reinforcement Learning, more insights than DNN, opens new possibilities

Data-driven Economic NMPC using Reinforcement Learning, S. Gros, M. Zanon, Transaction on Automatic Control
A Data-Efficient Framework for Learning NMPC controllers, M. Zanon, S. Gros, A. Bemporad, ECC 2019 Deterministic Policy Gradient for Data-Driven Economic NMPC, S. Gros, M. Zanon
System Identification and Reinforcement Learning for Data-Driven Economic NMPC, S. Gros, M. Zanon

Conclusions & Future work

Conclusions

- Cost function (and constraints) of an ENMPC scheme ought not necessarily match the ones describing the problem, should be adjusted
- Adjusting ENMPC scheme for performance $\neq$ model fitting
- Necessary Conditions of Optimality of an ENMPC scheme are defined by policy gradient theorems. Can be deployed to tune ENMPC schemes.
- ENMPC schemes can be used as a structured function approximation for Reinforcement Learning, more insights than DNN, opens new possibilities
- Strong connections to dissipativity theory in ENMPC (not covered today)

Data-driven Economic NMPC using Reinforcement Learning, S. Gros, M. Zanon, Transaction on Automatic Control
A Data-Efficient Framework for Learning NMPC controllers, M. Zanon, S. Gros, A. Bemporad, ECC 2019 Deterministic Policy Gradient for Data-Driven Economic NMPC, S. Gros, M. Zanon
System Identification and Reinforcement Learning for Data-Driven Economic NMPC, S. Gros, M. Zanon

Conclusions & Future work

Conclusions

- Cost function (and constraints) of an ENMPC scheme ought not necessarily match the ones describing the problem, should be adjusted
- Adjusting ENMPC scheme for performance $\neq$ model fitting
- Necessary Conditions of Optimality of an ENMPC scheme are defined by policy gradient theorems. Can be deployed to tune ENMPC schemes.
- ENMPC schemes can be used as a structured function approximation for Reinforcement Learning, more insights than DNN, opens new possibilities
- Strong connections to dissipativity theory in ENMPC (not covered today)

Future work

- Computationally efficient RL methods for ENMPC (LSTD/experience replay)

Data-driven Economic NMPC using Reinforcement Learning, S. Gros, M. Zanon, Transaction on Automatic Control
A Data-Efficient Framework for Learning NMPC controllers, M. Zanon, S. Gros, A. Bemporad, ECC 2019 Deterministic Policy Gradient for Data-Driven Economic NMPC, S. Gros, M. Zanon
System Identification and Reinforcement Learning for Data-Driven Economic NMPC, S. Gros, M. Zanon

Conclusions & Future work

Conclusions

- Cost function (and constraints) of an ENMPC scheme ought not necessarily match the ones describing the problem, should be adjusted
- Adjusting ENMPC scheme for performance $\neq$ model fitting
- Necessary Conditions of Optimality of an ENMPC scheme are defined by policy gradient theorems. Can be deployed to tune ENMPC schemes.
- ENMPC schemes can be used as a structured function approximation for Reinforcement Learning, more insights than DNN, opens new possibilities
- Strong connections to dissipativity theory in ENMPC (not covered today)

Future work

- Computationally efficient RL methods for ENMPC (LSTD/experience replay)
- SYSID & RL for ENMPC context? It's a safe-learning question...

Data-driven Economic NMPC using Reinforcement Learning, S. Gros, M. Zanon, Transaction on Automatic Control
A Data-Efficient Framework for Learning NMPC controllers, M. Zanon, S. Gros, A. Bemporad, ECC 2019 Deterministic Policy Gradient for Data-Driven Economic NMPC, S. Gros, M. Zanon
System Identification and Reinforcement Learning for Data-Driven Economic NMPC, S. Gros, M. Zanon

Conclusions & Future work

Conclusions

- Cost function (and constraints) of an ENMPC scheme ought not necessarily match the ones describing the problem, should be adjusted
- Adjusting ENMPC scheme for performance $\neq$ model fitting
- Necessary Conditions of Optimality of an ENMPC scheme are defined by policy gradient theorems. Can be deployed to tune ENMPC schemes.
- ENMPC schemes can be used as a structured function approximation for Reinforcement Learning, more insights than DNN, opens new possibilities
- Strong connections to dissipativity theory in ENMPC (not covered today)

Future work

- Computationally efficient RL methods for ENMPC (LSTD/experience replay)
- SYSID & RL for ENMPC context? It's a safe-learning question...
- Big Data for ENMPC tuning?

Data-driven Economic NMPC using Reinforcement Learning, S. Gros, M. Zanon, Transaction on Automatic Control
A Data-Efficient Framework for Learning NMPC controllers, M. Zanon, S. Gros, A. Bemporad, ECC 2019 Deterministic Policy Gradient for Data-Driven Economic NMPC, S. Gros, M. Zanon
System Identification and Reinforcement Learning for Data-Driven Economic NMPC, S. Gros, M. Zanon