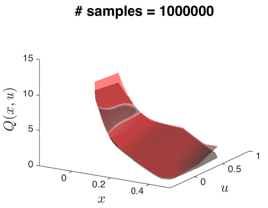
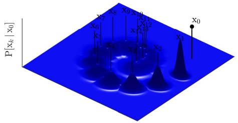
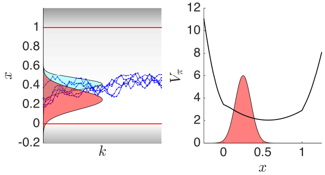
What are we going to discuss?

1. Learning for MPC - A closed-loop performance view
2. Safety & stability in Learning for MPC
3. MPC and Markov Decision Processes - When is learning beneficial?



$$Q_+(\mathbf{x},\mathbf{u}) \leftarrow L(\mathbf{x},\mathbf{u}) + \gamma \mathbb{E}\left[V\left(\mathbf{x}_+\right) \mid \mathbf{x},\mathbf{u}\right]$$





What are we going to discuss today?

MPC for MDPs

- MPC: purpose and usage
- MPC as a practical solution to tackle MDPs
- Planning vs. Policing
- Repeated planning as a policy
- MPC as an MDP model
- Optimal MPC model?

Model Predictive Control
for
Markov Decision Processes

Sébastien Gros

Cybernetic, NTNU

Freiburg PhD School

Model Predictive Control (MPC) Tuning for Performance

MPC: at current state $\mathbf{s}$ solve

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

gives policy $\boldsymbol{\pi}_{\boldsymbol{\theta}}^{\mathrm{MPC}}\left(\mathbf{s}\right) = \mathbf{u}_0^\star$

- "Classic" view
- MPC built around the model $\mathbf{f}_{\boldsymbol{\theta}}$
- $\boldsymbol{\theta}$ fits $\mathbf{f}_{\boldsymbol{\theta}}$ to data
- Model fitting→optimality is tricky

MPC: at current state $\mathbf{s}$ solve

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

gives policy $\boldsymbol{\pi}_{\boldsymbol{\theta}}^{\mathrm{MPC}}\left(\mathbf{s}\right) = \mathbf{u}_0^\star$

- "Holistic" view
- MPC is a model of $Q^\star$
- $\boldsymbol{\theta}$ fits Q^{MPC} to $Q^\star$
- RL → optimality, also BO btw!!

Model Predictive Control (MPC) Tuning for Performance

MPC: at current state $\mathbf{s}$ solve

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

gives policy $\boldsymbol{\pi}_{\boldsymbol{\theta}}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$

- "Classic" view
- MPC built around the model $\mathbf{f}_{\boldsymbol{\theta}}$
- $\boldsymbol{\theta}$ fits $\mathbf{f}_{\boldsymbol{\theta}}$ to data
- Model fitting$\rightarrow$optimality is tricky

MPC: at current state $\mathbf{s}$ solve

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

gives policy $\boldsymbol{\pi}_{\boldsymbol{\theta}}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$

- "Holistic" view
- MPC is a model of $Q^\star$
- $\boldsymbol{\theta}$ fits Q^{MPC} to $Q^\star$
- RL $\rightarrow$ optimality, also BO btw!!

Can we get $Q^{\mathrm{MPC}} = Q^\star$ from tuning the MPC model alone??
Can we get $Q^{\mathrm{MPC}} = Q^\star$ from fitting the model to data??

We are unpacking the maths

Outline

Planning vs. Policing

Infinite horizon & discounted

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

Policy π : state $\rightarrow$ action
belongs to a function space

Planning vs. Policing

Infinite horizon & discounted

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

Policy $\boldsymbol{\pi}$: state $\rightarrow$ action
belongs to a function space

Finite-horizon equivalent:

$$\boldsymbol{\pi}_{0,\ldots,N-1}^\star = \arg\min_{\boldsymbol{\pi}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)\right]$$

If $T = V_\star$, then $\boldsymbol{\pi}_{0,\ldots,N-1}^\star = \boldsymbol{\pi}_\infty^\star$

Planning vs. Policing

Infinite horizon & discounted

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

Policy $\boldsymbol{\pi}$: state $\rightarrow$ action
belongs to a function space

Finite-horizon equivalent:

$$\boldsymbol{\pi}_{0,\dots,N-1}^\star = \arg\min_{\boldsymbol{\pi}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)\right]$$

If $T = V_\star$, then $\boldsymbol{\pi}_{0,\dots,N-1}^\star = \boldsymbol{\pi}_\infty^\star$

Planning instead of policing:

$$\min_{\mathbf{u}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s}\right]$$

i.e. restrict policies to fixed $\mathbf{u}_{0,\dots,N-1}$

Planning vs. Policing

Infinite horizon & discounted

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

Policy $\boldsymbol{\pi}$: state $\rightarrow$ action
belongs to a function space

Finite-horizon equivalent:

$$\boldsymbol{\pi}_{0,\ldots,N-1}^\star = \arg\min_{\boldsymbol{\pi}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)\right]$$

If $T = V_\star$, then $\boldsymbol{\pi}_{0,\ldots,N-1}^\star = \boldsymbol{\pi}_\infty^\star$

Planning instead of policing:

$$\min_{\mathbf{u}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s}\right]$$

i.e. restrict policies to fixed $\mathbf{u}_{0,\ldots,N-1}$

Deterministic approximation:

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

i.e. adopt deterministic model

Planning vs. Policing

Infinite horizon & discounted

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

Policy $\boldsymbol{\pi}$: state $\rightarrow$ action
belongs to a function space

Finite-horizon equivalent:

$$\boldsymbol{\pi}^\star_{0,\dots,N-1} = \arg\min_{\boldsymbol{\pi}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)\right]$$

If $T = V_\star$, then $\boldsymbol{\pi}^\star_{0,\dots,N-1} = \boldsymbol{\pi}^\star_\infty$

Why attacking the problem in these ways?

Planning instead of policing:

$$\min_{\mathbf{u}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s}\right]$$

i.e. restrict policies to fixed $\mathbf{u}_{0,\dots,N-1}$

Deterministic approximation:

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

i.e. adopt deterministic model

Planning vs. Policing - Illustration

Planning instead of policing:

$$\min_{\mathbf{u}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

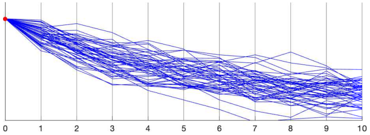
i.e. restrict policies to fixed $\mathbf{u}_{0,\dots,N-1}$

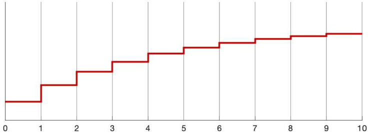
Planning vs. Policing - Illustration

Planning instead of policing:

$$\min_{\mathbf{u}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

i.e. restrict policies to fixed $\mathbf{u}_{0,\ldots,N-1}$





Planning vs. Policing - Illustration

Planning instead of policing:

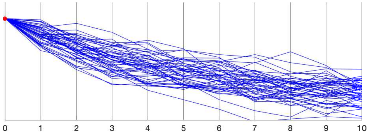
$$\min_{\mathbf{u}_{0,\ldots,N-1}} \ \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \ \middle| \ \mathbf{x}_0 = \mathbf{s} \right]$$

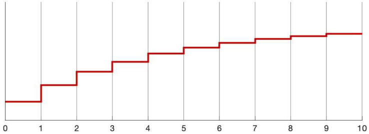
i.e. restrict policies to fixed $\mathbf{u}_{0,\ldots,N-1}$

Policing:

$$\min_{\boldsymbol{\pi}_{0,\ldots,N-1}} \ \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies





Planning vs. Policing - Illustration

Planning instead of policing:

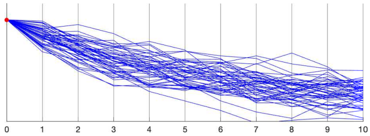
$$\min_{\mathbf{u}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

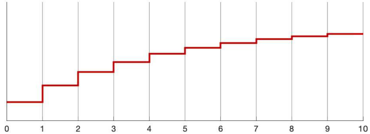
i.e. restrict policies to fixed $\mathbf{u}_{0,\ldots,N-1}$

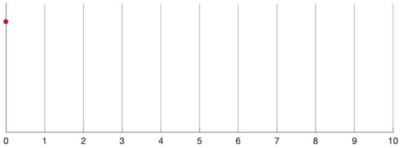
Policing:

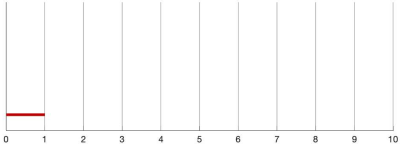
$$\min_{\boldsymbol{\pi}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies









Planning vs. Policing - Illustration

Planning instead of policing:

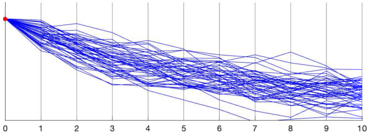
$$\min_{\mathbf{u}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

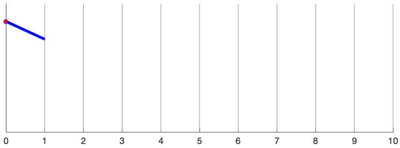
i.e. restrict policies to fixed $\mathbf{u}_{0,\dots,N-1}$

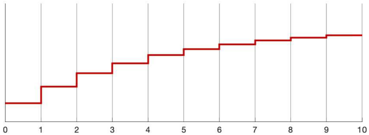
Policing:

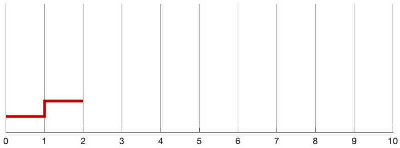
$$\min_{\boldsymbol{\pi}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies









Planning vs. Policing - Illustration

Planning instead of policing:

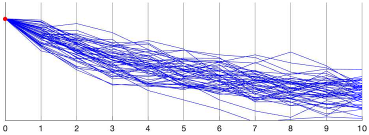
$$\min_{\mathbf{u}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

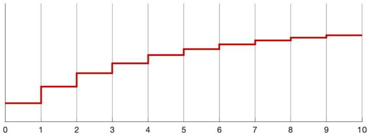
i.e. restrict policies to fixed $\mathbf{u}_{0,\dots,N-1}$

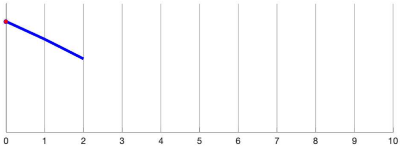
Policing:

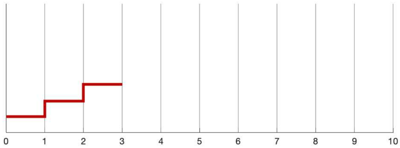
$$\min_{\boldsymbol{\pi}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies









Planning vs. Policing - Illustration

Planning instead of policing:

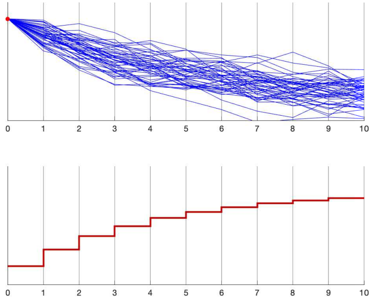
$$\min_{\mathbf{u}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

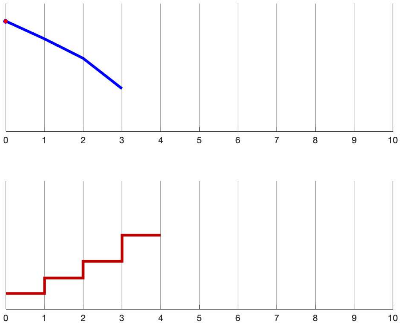
i.e. restrict policies to fixed $\mathbf{u}_{0,\ldots,N-1}$

Policing:

$$\min_{\boldsymbol{\pi}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies





Planning vs. Policing - Illustration

Planning instead of policing:

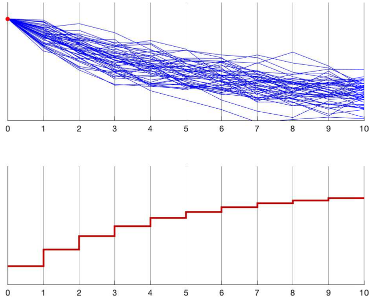
$$\min_{\mathbf{u}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

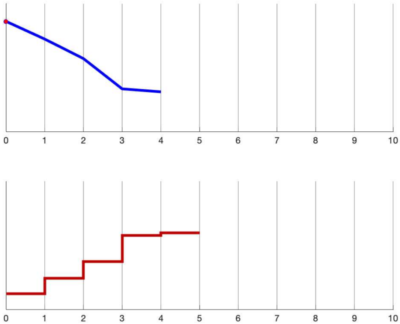
i.e. restrict policies to fixed $\mathbf{u}_{0,\ldots,N-1}$

Policing:

$$\min_{\boldsymbol{\pi}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies





Planning vs. Policing - Illustration

Planning instead of policing:

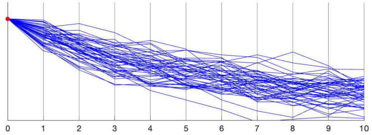
$$\min_{\mathbf{u}_{0,\ldots,N-1}} \; \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

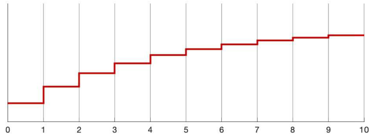
i.e. restrict policies to fixed $\mathbf{u}_{0,\ldots,N-1}$

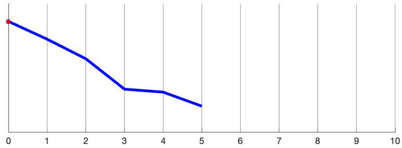
Policing:

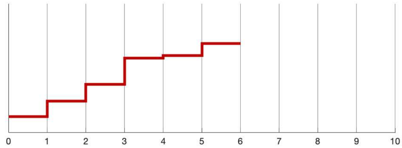
$$\min_{\boldsymbol{\pi}_{0,\ldots,N-1}} \; \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies









Planning vs. Policing - Illustration

Planning instead of policing:

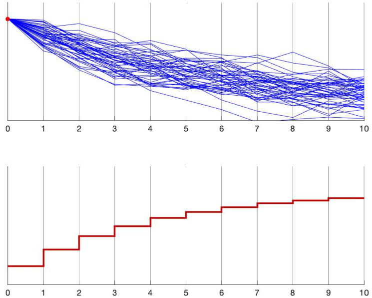
$$\min_{\mathbf{u}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

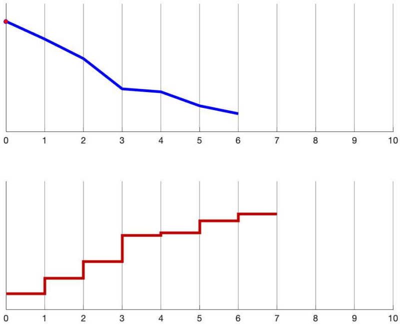
i.e. restrict policies to fixed $\mathbf{u}_{0,\ldots,N-1}$

Policing:

$$\min_{\boldsymbol{\pi}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies





Planning vs. Policing - Illustration

Planning instead of policing:

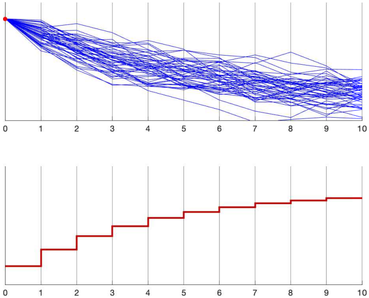
$$\min_{\mathbf{u}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

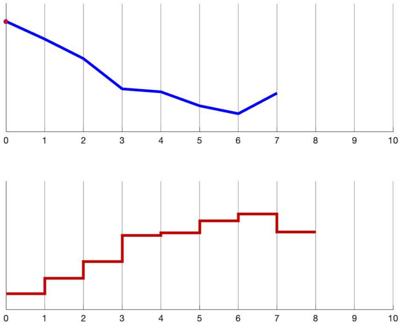
i.e. restrict policies to fixed $\mathbf{u}_{0,\dots,N-1}$

Policing:

$$\min_{\boldsymbol{\pi}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies





Planning vs. Policing - Illustration

Planning instead of policing:

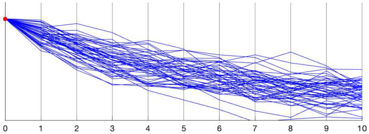
$$\min_{\mathbf{u}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

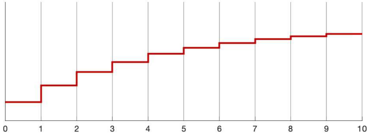
i.e. restrict policies to fixed $\mathbf{u}_{0,\dots,N-1}$

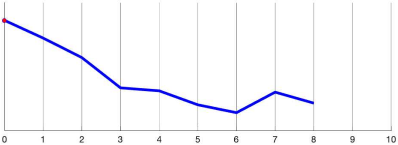
Policing:

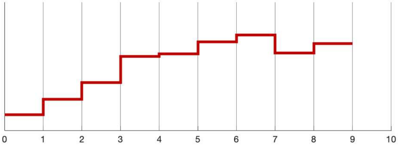
$$\min_{\boldsymbol{\pi}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies









Planning vs. Policing - Illustration

Planning instead of policing:

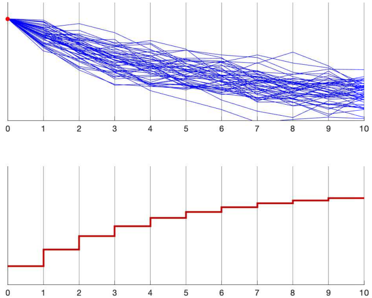
$$\min_{\mathbf{u}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

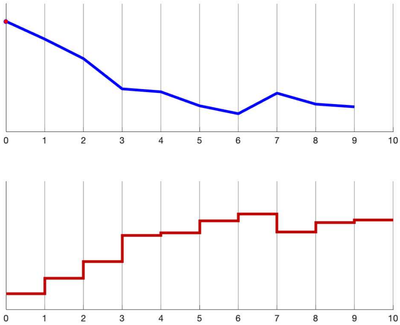
i.e. restrict policies to fixed $\mathbf{u}_{0,\ldots,N-1}$

Policing:

$$\min_{\boldsymbol{\pi}_{0,\ldots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies





Planning vs. Policing - Illustration

Planning instead of policing:

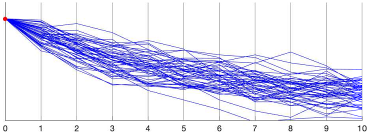
$$\min_{\mathbf{u}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

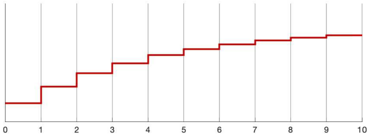
i.e. restrict policies to fixed $\mathbf{u}_{0,\dots,N-1}$

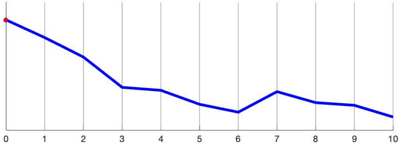
Policing:

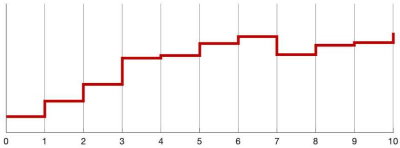
$$\min_{\boldsymbol{\pi}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies









Planning vs. Policing - Illustration

Planning instead of policing:

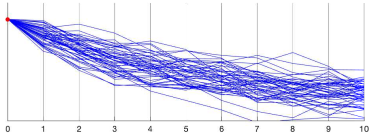
$$\min_{\mathbf{u}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \,\middle|\, \mathbf{x}_0 = \mathbf{s} \right]$$

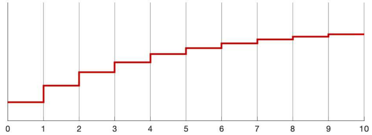
i.e. restrict policies to fixed $\mathbf{u}_{0,\dots,N-1}$

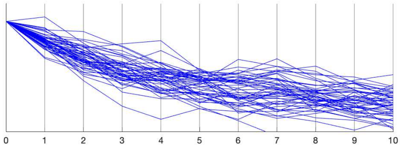
Policing:

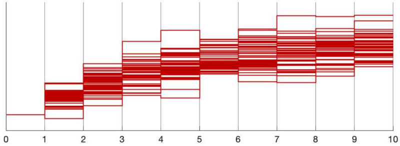
$$\min_{\boldsymbol{\pi}_{0,\dots,N-1}} \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies









Planning vs. Policing - Illustration

Deterministic approximation:

$$\min_{\mathbf{u}_{0,\dots,N-1}} \quad T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)$$

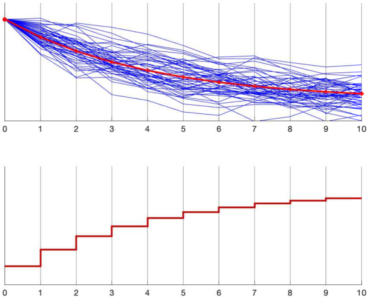
$$\text{s.t} \quad \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)$$

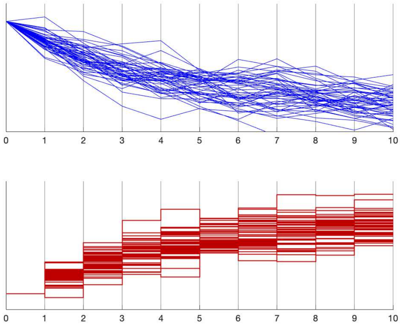
$$\mathbf{x}_0 = \mathbf{s}$$

Policing:

$$\min_{\boldsymbol{\pi}_{0,\dots,N-1}} \quad \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right]$$

i.e. optimize over policies





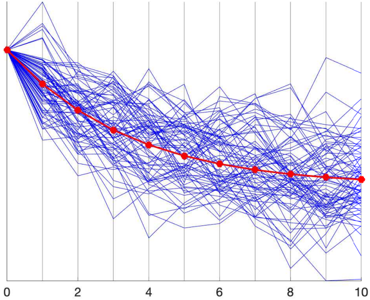
MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

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



MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

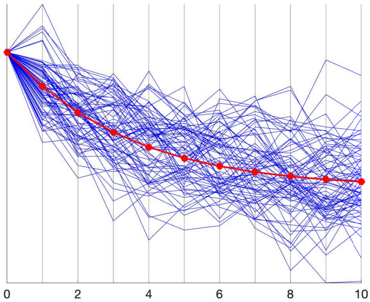
$$\begin{aligned}\min_{\mathbf{u}_{0,\dots,N-1}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s}\end{aligned}$$

Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

from repeated planning



MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

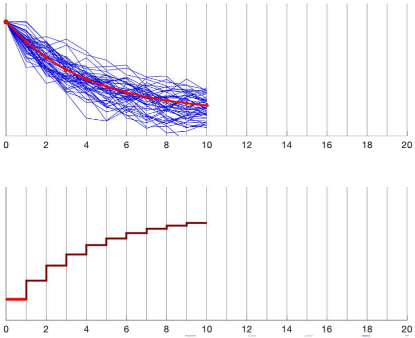
$$\min_{\mathbf{u}_{0,\dots,N-1}} \quad T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)$$
$$\text{s.t} \quad \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)$$
$$\mathbf{x}_0 = \mathbf{s}$$

Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

from repeated planning



MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

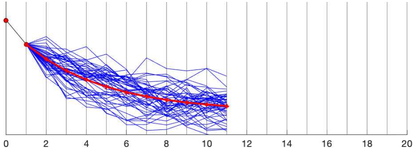
$$\begin{aligned} \min_{\mathbf{u}_{0,\ldots,N-1}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s} \end{aligned}$$

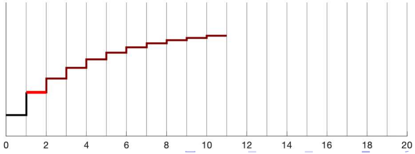
Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\boldsymbol{\pi}^{\text{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

  from repeated planning





MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

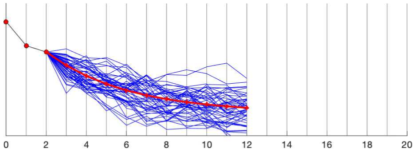
$$\begin{aligned}\min_{\mathbf{u}_{0,\ldots,N-1}} \quad & T(\mathbf{x}_N)+\sum_{k=0}^{N-1}\gamma^k L(\mathbf{x}_k,\mathbf{u}_k)\\ \text{s.t}\quad & \mathbf{x}_{k+1}=\mathbf{f}(\mathbf{x}_k,\mathbf{u}_k)\\ & \mathbf{x}_0=\mathbf{s}\end{aligned}$$

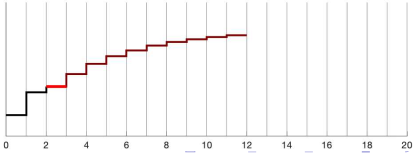
Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s})=\mathbf{u}_0^\star$$

from repeated planning





MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

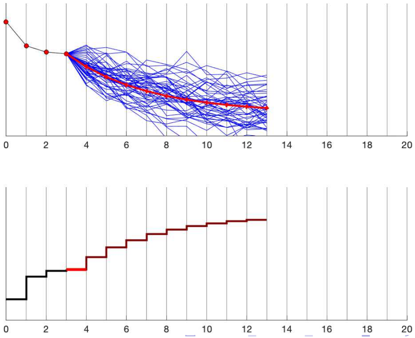
$$\begin{aligned}\min_{\mathbf{u}_{0,\ldots,N-1}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s}\end{aligned}$$

Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

from repeated planning



MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

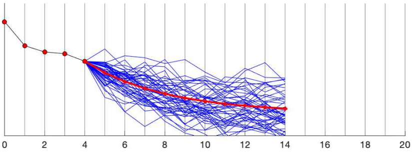
$$\begin{aligned}\min_{\mathbf{u}_{0,\dots,N-1}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s}\end{aligned}$$

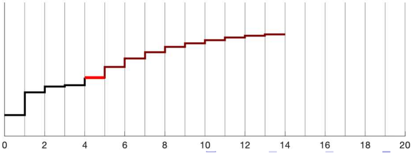
Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

from repeated planning





MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

$$\min_{\mathbf{u}_{0,\ldots,N-1}} \quad T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)$$

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

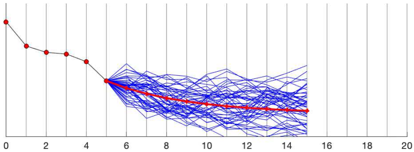
$$\mathbf{x}_0 = \mathbf{s}$$

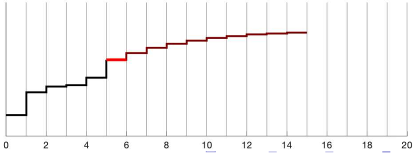
Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

from repeated planning





MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

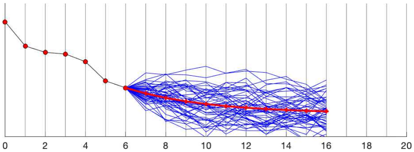
$$\begin{aligned} \min_{\mathbf{u}_{0,\dots,N-1}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s} \end{aligned}$$

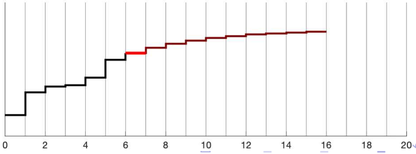
Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\pi^{\text{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

from repeated planning





MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

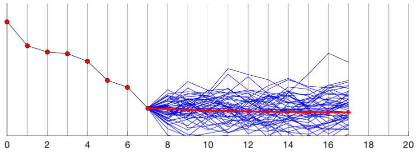
$$\begin{aligned} \min_{\mathbf{u}_{0,\ldots,N-1}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s} \end{aligned}$$

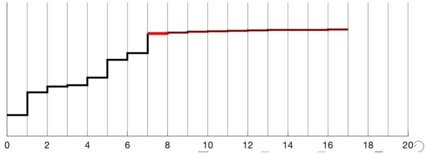
Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\boldsymbol{\pi}^{\text{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

from repeated planning





MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

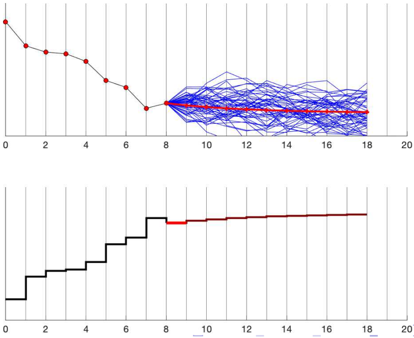
$$\begin{aligned} \min_{\mathbf{u}_{0,\dots,N-1}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s} \end{aligned}$$

Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\pi^{\text{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

from repeated planning



MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

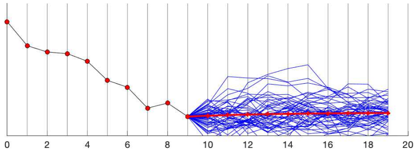
$$\begin{aligned} \min_{\mathbf{u}_{0,\ldots,N-1}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s} \end{aligned}$$

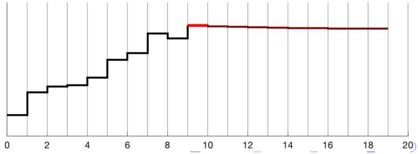
Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

from repeated planning





MPC vs. MDP?

Infinite horizon & discounted

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

Deterministic approximation:

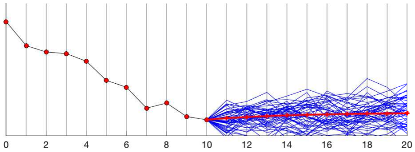
$$\begin{aligned} \min_{\mathbf{u}_{0,\ldots,N-1}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \mathrm{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s} \end{aligned}$$

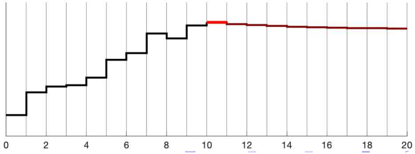
Remarks:

- MPC predictions from $\mathbf{f}$ are a simplified representation of the real dynamics
- Finer models can be built, e.g. scenario trees, more on this in a bit...
- MPC defines a policy:

$$\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

from repeated planning





MPC as a policy

Deterministic MPC:

$$\min_{\mathbf{u}_{0,\ldots,N-1}} \quad T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)$$

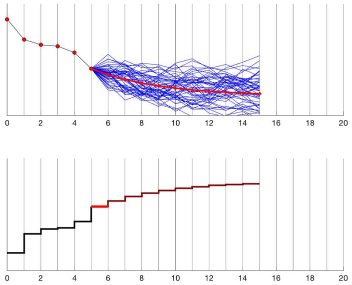
$$\text{s.t} \quad \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)$$

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

Defines policy:

$$\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$$

How does $\boldsymbol{\pi}^{\mathrm{MPC}}$ relate to $\boldsymbol{\pi}^\star$?



No reason to match:

- Planning rather than policing
- Plan ignores stochasticity

Can we clarify the relationship?

Some more context on MPC for performance

Historically MPC focuses on **constraints satisfaction & stability**. Cost is for reference tracking, not representative of the system performance.

Some more context on MPC for performance

Historically MPC focuses on **constraints satisfaction & stability**. Cost is for reference tracking, not representative of the system performance.

- "Tracking MPC"
- Classic stability theory
- Uncertainty via
  - Robust MPC
  - Stochastic MPC
- **"MPC is for constraints satisfaction"** (public statement)

Some more context on MPC for performance

Historically MPC focuses on **constraints satisfaction & stability**. Cost is for reference tracking, not representative of the system performance.

More recently, focus on **closed-loop performance**, e.g. energy, time, money. Cost is generic, representative of the system performance.

- "Tracking MPC"
- Classic stability theory
- Uncertainty via
  - Robust MPC
  - Stochastic MPC
- **"MPC is for constraints satisfaction"** (public statement)

Some more context on MPC for performance

Historically MPC focuses on **constraints satisfaction & stability**. Cost is for reference tracking, not representative of the system performance.

- "Tracking MPC"
- Classic stability theory
- Uncertainty via
  - Robust MPC
  - Stochastic MPC
- **"MPC is for constraints satisfaction"** (public statement)

More recently, focus on **closed-loop performance**, e.g. energy, time, money. Cost is generic, representative of the system performance.

- "Economic MPC"
- Dissipativity theory
- Uncertainty via
  - Robust MPC
  - Stochastic MPC

... debatable approach for performance

Some more context on MPC for performance

Historically MPC focuses on **constraints satisfaction & stability**. Cost is for reference tracking, not representative of the system performance.

- "Tracking MPC"
- Classic stability theory
- Uncertainty via
  - Robust MPC
  - Stochastic MPC
- **"MPC is for constraints satisfaction"** (public statement)

More recently, focus on **closed-loop performance**, e.g. energy, time, money. Cost is generic, representative of the system performance.

- "Economic MPC"
- Dissipativity theory
- Uncertainty via
  - Robust MPC
  - Stochastic MPC

... debatable approach for performance

MPC for closed-loop performance

- is not a very old topic
- unclear in the presence of stochasticity
- partially clarified by recent results

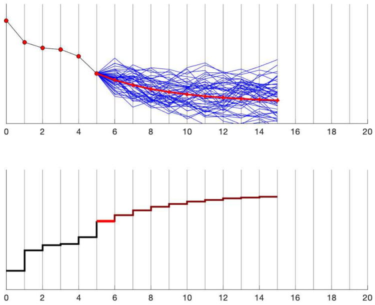
MPC as a model of the MDP

Infinite horizon & discounted

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

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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



MPC as a model of the MDP

Infinite horizon & discounted

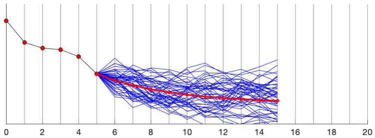
$$\boldsymbol{\pi}^\star_\infty = \arg\min_{\boldsymbol{\pi}} \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)\right]$$

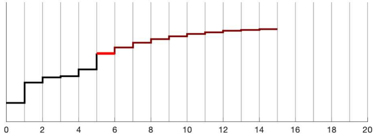
MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

MPC as a model of the MDP

$$\begin{aligned} V^{\mathrm{MPC}}(\mathbf{s}) := \min_{\mathbf{u}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \mathrm{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s} \end{aligned}$$





MPC as a model of the MDP

Infinite horizon & discounted

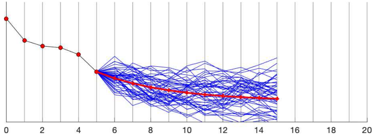
$$\pi^\star_\infty = \arg\min_{\pi} \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)\right]$$

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

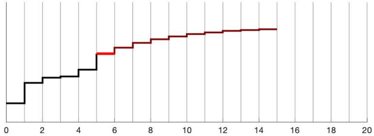
$$\begin{aligned}\min_{\mathbf{u}_{0,\ldots,N-1}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \mathrm{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s}\end{aligned}$$

MPC as a model of the MDP

$$\begin{aligned}V^{\mathrm{MPC}}(\mathbf{s}) := \min_{\mathbf{u}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \mathrm{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s}\end{aligned}$$



$$\begin{aligned}Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a}) := \min_{\mathbf{u}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \mathrm{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s}, \quad \mathbf{u}_0 = \mathbf{a}\end{aligned}$$



MPC as a model of the MDP

Infinite horizon & discounted

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

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

MPC as a model of the MDP

$$\begin{aligned}V^{\mathrm{MPC}}(\mathbf{s}) := \min_{\mathbf{u}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)\\ \mathrm{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)\\ & \mathbf{x}_0 = \mathbf{s}\end{aligned}$$

$$\begin{aligned}Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a}) := \min_{\mathbf{u}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)\\ \mathrm{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)\\ & \mathbf{x}_0 = \mathbf{s}, \quad \mathbf{u}_0 = \mathbf{a}\end{aligned}$$

MPC is consistent:

$$V^{\mathrm{MPC}}(\mathbf{s}) = \min_{\mathbf{a}} \; Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a})$$

$$\pi^{\mathrm{MPC}}(\mathbf{s}) = \arg\min_{\mathbf{a}} \; Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a})$$

MPC as a model of the MDP

Infinite horizon & discounted

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

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

MPC as a model of the MDP

$$\begin{aligned} V^{\mathrm{MPC}}(\mathbf{s}) := \min_{\mathbf{u}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \mathrm{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s} \end{aligned}$$

$$\begin{aligned} Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a}) := \min_{\mathbf{u}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \\ \mathrm{s.t} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s}, \quad \mathbf{u}_0 = \mathbf{a} \end{aligned}$$

MPC is consistent:

$$V^{\mathrm{MPC}}(\mathbf{s}) = \min_{\mathbf{a}} Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a})$$

$$\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \arg\min_{\mathbf{a}} Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a})$$

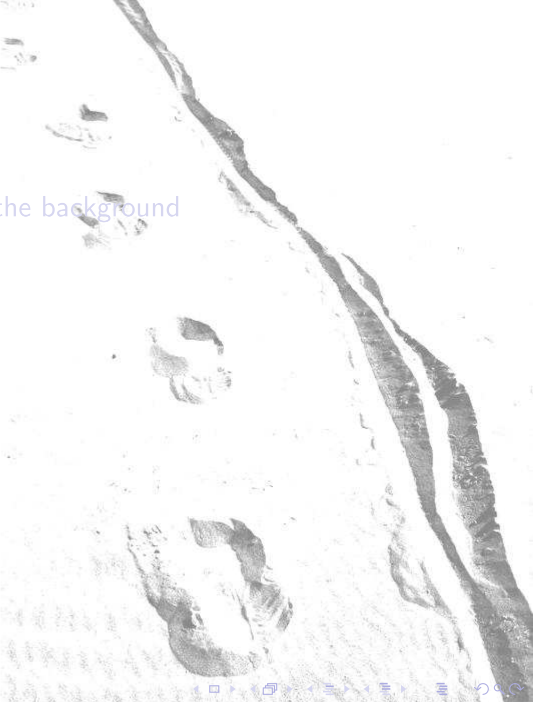
MPC is a complete model of MDP if:

$$Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a}) = Q^\star(\mathbf{s}, \mathbf{a})$$

for all $\mathbf{s}, \mathbf{a}$. Then **optimality** holds:

$$\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \boldsymbol{\pi}^\star(\mathbf{s})$$

Outline



MPC as a model of the MDP

Infinite horizon & discounted

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

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

When does

$$Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a}) = Q^\star(\mathbf{s}, \mathbf{a})$$

hold?

MPC as a model of the MDP

Infinite horizon & discounted

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

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

When does

$$Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a}) = Q^\star(\mathbf{s}, \mathbf{a})$$

hold?

Bellman equation for all $\mathbf{s}, \mathbf{a}$

$$Q^\star(\mathbf{s}, \mathbf{a}) = L(\mathbf{s}, \mathbf{a}) + \gamma \mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}\right]$$

MPC as a model of the MDP

Infinite horizon & discounted

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

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

When does

$$Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a}) = Q^\star(\mathbf{s}, \mathbf{a})$$

hold?

Bellman equation for all $\mathbf{s}, \mathbf{a}$

$$Q^\star(\mathbf{s}, \mathbf{a}) = L(\mathbf{s}, \mathbf{a}) + \gamma \mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}\right]$$

Theorem: if $T = \gamma^N V^\star$, MPC is a **complete** MDP solution† if for some c

$$Q^\star(\mathbf{s}, \mathbf{a}) = L(\mathbf{s}, \mathbf{a}) + \gamma V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

holds for all $\mathbf{s}, \mathbf{a}$ (with technical assumption)

† up to a constant

MPC as a model of the MDP

Infinite horizon & discounted

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

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

When does

$$Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a}) = Q^\star(\mathbf{s}, \mathbf{a})$$

hold?

Bellman equation for all $\mathbf{s}, \mathbf{a}$

$$Q^\star(\mathbf{s}, \mathbf{a}) = L(\mathbf{s}, \mathbf{a}) + \gamma \mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}\right]$$

Theorem: if $T = \gamma^N V^\star$, MPC is a **complete** MDP solution† if for some c

$$Q^\star(\mathbf{s}, \mathbf{a}) = L(\mathbf{s}, \mathbf{a}) + \gamma V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

holds for all $\mathbf{s}, \mathbf{a}$ (with technical assumption)

†up to a constant

Proof: telescopic sums, Bellman identities, properties of the advantage function, some measure theory, devil is in the details

MPC as a model of the MDP

Infinite horizon & discounted

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

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

When does

$$Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a}) = Q^\star(\mathbf{s}, \mathbf{a})$$

hold?

Bellman equation for all $\mathbf{s}, \mathbf{a}$

$$Q^\star(\mathbf{s}, \mathbf{a}) = L(\mathbf{s}, \mathbf{a}) + \gamma \mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}\right]$$

What does it mean?

Theorem: if $T = \gamma^N V^\star$, MPC is a **complete** MDP solution† if for some c

$$Q^\star(\mathbf{s}, \mathbf{a}) = L(\mathbf{s}, \mathbf{a}) + \gamma V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

holds for all $\mathbf{s}, \mathbf{a}$ (with technical assumption)

† up to a constant

Proof: telescopic sums, Bellman identities, properties of the advantage function, some measure theory, devil is in the details

What does that mean?

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

Theorem: if $T = \gamma^N V^\star$, MPC is complete

$$Q^\star(\mathbf{s}, \mathbf{a}) = L(\mathbf{s}, \mathbf{a}) + \gamma V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

What does that mean?

MPC policy $\pi^{\mathrm{MPC}}(\mathrm{s}) = \mathbf{u}_0^\star$ from

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

Theorem: if $T = \gamma^N V^\star$, MPC is complete

$$Q^\star(\mathrm{s}, \mathbf{a}) = L(\mathrm{s}, \mathbf{a}) + \gamma V^\star(\mathbf{f}(\mathrm{s}, \mathbf{a})) + c$$

Equivalent statement

$$\mathbb{E}\left[V^\star(\mathrm{s}_+) \mid \mathrm{s}, \mathbf{a}\right] = V^\star(\mathbf{f}(\mathrm{s}, \mathbf{a})) + c$$

requirement on model $\mathbf{f}$!!!

What does that mean?

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

Theorem: if $T = \gamma^N V^\star$, MPC is complete

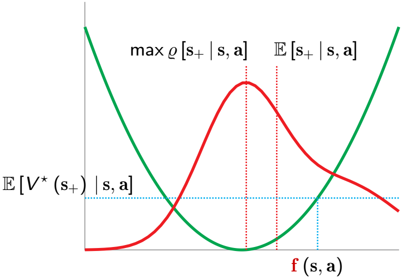
$$Q^\star(\mathbf{s}, \mathbf{a}) = L(\mathbf{s}, \mathbf{a}) + \gamma V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

Equivalent statement

$$\mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}\right] = V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

requirement on model $\mathbf{f}$!!!

E.g. $c = 0$, $V^\star$ quadratic



What does that mean?

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

Theorem: if $T = \gamma^N V^\star$, MPC is complete

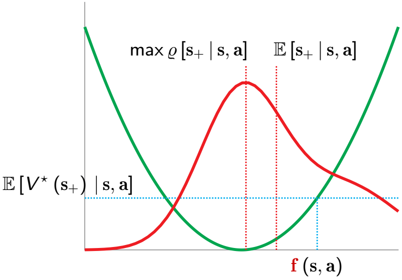
$$Q^\star(\mathbf{s}, \mathbf{a}) = L(\mathbf{s}, \mathbf{a}) + \gamma V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

Equivalent statement

$$\mathbb{E}\left[V^\star(\mathbf{s}_+) \,|\, \mathbf{s}, \mathbf{a}\right] = V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

requirement on model $\mathbf{f}$!!!

E.g. $c = 0$, $V^\star$ quadratic



Remark

Even for a simple $V^\star$ **neither**

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \mathbb{E}[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}]$$

nor

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \max \varrho[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}]$$

make the MPC complete / optimal

Why does it matter?

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

MPC is **complete** if model $\mathbf{f}$ satisfies

$$\mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}\right] = V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

Why does it matter?

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

MPC is **complete** if model $\mathbf{f}$ satisfies

$$\mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}\right] = V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

Model $\mathbf{f}$ built via **Least-Squares** fit

$$\min_{\boldsymbol{\theta}} \sum_{k=0}^{N} \left\| \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{s}_k, \mathbf{a}_k) - \mathbf{s}_{k+1} \right\|^2$$

Classic approach in SYSID & ML

Why does it matter?

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

$$\min_{\mathbf{u}_{0,\ldots,N-1}} \quad T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)$$
$$\text{s.t} \quad \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)$$
$$\mathbf{x}_0 = \mathbf{s}$$

MPC is **complete** if model $\mathbf{f}$ satisfies

$$\mathbb{E}\left[V^\star(\mathbf{s}_+) \,|\, \mathbf{s}, \mathbf{a}\right] = V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

Model $\mathbf{f}$ built via **Least-Squares** fit

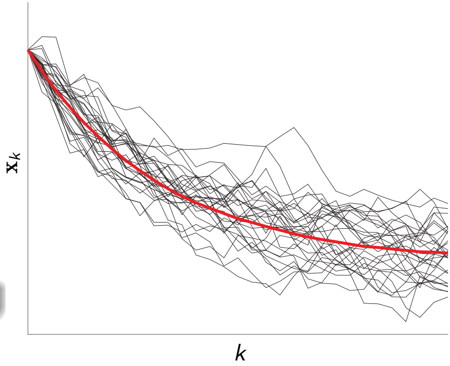
$$\min_{\boldsymbol{\theta}} \sum_{k=0}^{N} \left\| \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{s}_k, \mathbf{a}_k) - \mathbf{s}_{k+1} \right\|^2$$

Classic approach in SYSID & ML

In ideal conditions:

$$\mathbf{f}_{\boldsymbol{\theta}^\star}(\mathbf{s}, \mathbf{a}) \rightarrow \mathbb{E}\left[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\right]$$

i.e. one-step ahead expected transition



Trajectory from simulating $\mathbf{f}_{\boldsymbol{\theta}^\star}$

Why does it matter?

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

$$\min_{\mathbf{u}_{0,\ldots,N-1}} \quad T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)$$
$$\text{s.t} \quad \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)$$
$$\mathbf{x}_0 = \mathbf{s}$$

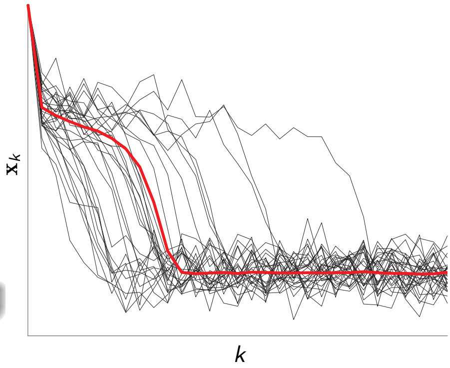
MPC is **complete** if model $\mathbf{f}$ satisfies

$$\mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}\right] = V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

Model $\mathbf{f}$ built via **Least-Squares** fit

$$\min_{\boldsymbol{\theta}} \sum_{k=0}^{N} \left\|\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{s}_k, \mathbf{a}_k) - \mathbf{s}_{k+1}\right\|^2$$

Classic approach in SYSID & ML



Trajectory from simulating $\mathbf{f}_{\boldsymbol{\theta}^\star}$

In ideal conditions:

$$\mathbf{f}_{\boldsymbol{\theta}^\star}(\mathbf{s}, \mathbf{a}) \rightarrow \mathbb{E}\left[\mathbf{s}_+ \mid \mathbf{s}, \mathbf{a}\right]$$

i.e. one-step ahead expected transition

Why does it matter?

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

$$\min_{\mathbf{u}_{0,\ldots,N-1}} \quad T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)$$
$$\text{s.t} \quad \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)$$
$$\mathbf{x}_0 = \mathbf{s}$$

Model $\mathbf{f}$ built via **Least-Squares** fit

$$\min_{\boldsymbol{\theta}} \sum_{k=0}^{N} \left\| \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{s}_k, \mathbf{a}_k) - \mathbf{s}_{k+1} \right\|^2$$

Classic approach in SYSID & ML

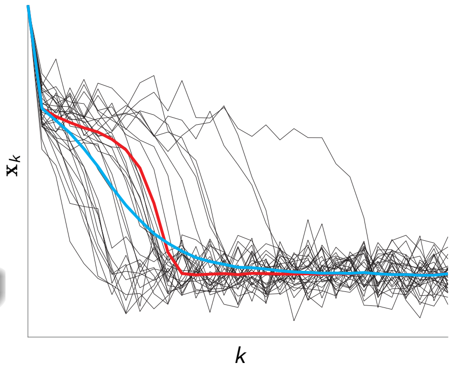
In ideal conditions:

$$\mathbf{f}_{\boldsymbol{\theta}^\star}(\mathbf{s}, \mathbf{a}) \to \mathbb{E}\left[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\right]$$

i.e. one-step ahead expected transition

MPC is **complete** if model $\mathbf{f}$ satisfies

$$\mathbb{E}\left[V^\star(\mathbf{s}_+) \,|\, \mathbf{s}, \mathbf{a}\right] = V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$



Trajectory from simulating $\mathbf{f}_{\boldsymbol{\theta}^\star}$
Expected trajectory

Why does it matter?

MPC policy $\boldsymbol{\pi}^{\text{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

MPC is **complete** if model $\mathbf{f}$ satisfies

$$\mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}\right] = V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

The **gap** in

$$Q^{\text{MPC}}(\mathbf{s}, \mathbf{a}) \approx Q^\star(\mathbf{s}, \mathbf{a}), \qquad \boldsymbol{\pi}^{\text{MPC}}(\mathbf{s}) \approx \boldsymbol{\pi}^\star(\mathbf{s})$$

when using MPC models based on one-step ahead Least-Squares fitting (PEM) comes from the **lack of commutativity (up to a constant) between** $V^\star$ **and** $\mathbb{E}[.]$, i.e.

$$\mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}\right] - V^\star\left(\mathbb{E}\left[\mathbf{s}_+ \mid \mathbf{s}, \mathbf{a}\right]\right) \neq c$$

To our best knowledge other methods (sim error, max likelihood) do not fix that

An Important Exception

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

An Important Exception - The LQR Case

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

Consider

- L is **quadratic**, no constraints
- **Real dynamics**: for some density φ

$$\varrho\left[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\right] = \varphi\left(\mathbf{s}_+ - \boldsymbol{\mu}\left(\mathbf{s}, \mathbf{a}\right)\right)$$

  where $\boldsymbol{\mu}$ is **affine**

- MPC model selected as:

$$\mathbf{f}\left(\mathbf{s}, \mathbf{a}\right) = \mathbb{E}\left[\, \mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a} \,\right]$$

An Important Exception - The LQR Case

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

Then...

- $V^\star$ is **quadratic**
- $\mathbf{f}$ is **affine**
- There is c such that

$$\mathbb{E}\left[V^\star(\mathbf{s}_+) \,|\, \mathbf{s}, \mathbf{a}\right] = V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

- MPC is **complete** for some T, i.e.

$$Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a}) = Q^\star(\mathbf{s}, \mathbf{a}),$$
$$\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \boldsymbol{\pi}^\star(\mathbf{s})$$

Consider

- L is **quadratic**, no constraints
- **Real dynamics**: for some density φ

$$\varrho\left[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\right] = \varphi\left(\mathbf{s}_+ - \boldsymbol{\mu}(\mathbf{s}, \mathbf{a})\right)$$

where $\boldsymbol{\mu}$ is **affine**

- MPC model selected as:

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \mathbb{E}\left[\,\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\,\right]$$

An Important Exception - The LQR Case

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

Then...

- $V^\star$ is **quadratic**
- $\mathbf{f}$ is **affine**
- There is c such that

$$\mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}\right] = V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + c$$

- MPC is **complete** for some T, i.e.

$$Q^{\mathrm{MPC}}(\mathbf{s}, \mathbf{a}) = Q^\star(\mathbf{s}, \mathbf{a}),$$
$$\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \boldsymbol{\pi}^\star(\mathbf{s})$$

Consider

- L is **quadratic**, no constraints
- **Real dynamics**: for some density φ

$$\varrho\left[\mathbf{s}_+ \mid \mathbf{s}, \mathbf{a}\right] = \varphi\left(\mathbf{s}_+ - \boldsymbol{\mu}(\mathbf{s}, \mathbf{a})\right)$$

where $\boldsymbol{\mu}$ is **affine**

- MPC model selected as:

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \mathbb{E}\left[\,\mathbf{s}_+ \mid \mathbf{s}, \mathbf{a}\,\right]$$

This is LQR + i.i.d state noise!

Why is this relevant for MPC?

Local Optimality of Classic MPC

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

Assume:

- $\varrho\left[\mathbf{s}_+ \mid \mathbf{s}, \mathbf{a}\right]$ is smooth in $\mathbf{s}, \mathbf{a}$, for all $\mathbf{s}_+$
- L is smooth
- $\boldsymbol{\pi}^\star$ is such that system dynamics converge to steady state density $\rho_\star(.)$ (dissipative) which is "off-constraints"

Local Optimality of Classic MPC

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

Assume:

- $\varrho\left[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\right]$ is smooth in $\mathbf{s}, \mathbf{a}$, for all $\mathbf{s}_+$
- L is smooth
- $\boldsymbol{\pi}^\star$ is such that system dynamics converge to steady state density $\rho_\star(.)$ (dissipative) which is "off-constraints"

Then MPC based on "expected-value" model

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \mathbb{E}\left[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\right]$$

yields a **locally optimal policy**, optimality loss in the order of the moments of $\rho_\star(.)$

Local Optimality of Classic MPC

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

Assume:

- $\varrho\left[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\right]$ is smooth in $\mathbf{s}, \mathbf{a}$, for all $\mathbf{s}_+$
- L is smooth
- $\boldsymbol{\pi}^\star$ is such that system dynamics converge to steady state density $\rho_\star(.)$ (dissipative) which is "off-constraints"

Then MPC based on "expected-value" model

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \mathbb{E}\left[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\right]$$

yields a **locally optimal policy**, optimality loss in the order of the moments of $\rho_\star(.)$

If problem is smooth & optimal policy drives and keeps the system "tightly" to its optimal steady state, then one can expect the MPC based on an "expected-value" model to perform well

Local Optimality of Classic MPC

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

Assume:

- $\varrho\left[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\right]$ is smooth in $\mathbf{s}, \mathbf{a}$, for all $\mathbf{s}_+$
- L is smooth
- $\boldsymbol{\pi}^\star$ is such that system dynamics converge to steady state density $\rho_\star(.)$ (dissipative) which is "off-constraints"

Then MPC based on "expected-value" model

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \mathbb{E}\left[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\right]$$

yields a **locally optimal policy**, optimality loss in the order of the moments of $\rho_\star(.)$

If problem is smooth & optimal policy drives and keeps the system "tightly" to its optimal steady state, then one can expect the MPC based on an "expected-value" model to perform well

Classic MPC paradigm works well under these conditions

Local Optimality of Classic MPC

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

Assume:

- $\varrho[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}]$ is smooth in $\mathbf{s}, \mathbf{a}$, for all $\mathbf{s}_+$
- L is smooth
- $\boldsymbol{\pi}^\star$ is such that system dynamics converge to steady state density $\rho_\star(.)$ (dissipative) which is "off-constraints"

Then MPC based on "expected-value" model

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \mathbb{E}[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}]$$

yields a **locally optimal policy**, optimality loss in the order of the moments of $\rho_\star(.)$

If problem is smooth & optimal policy drives and keeps the system "tightly" to its optimal steady state, then one can expect the MPC based on an "expected-value" model to perform well

Classic MPC paradigm works well under these conditions

Not "classic"?

- Economic / non-smooth cost
- No dissipativity / "disturbances"
- Non-smooth problem

Illustrations - Constrained LQR problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ i.i.d. *on a restricted interval.*

Cost:

$$L(\mathbf{s}, \mathbf{a}) = \mathbf{a}^2 + (\mathbf{s} - 0.5)^2$$

Constraints:

$$\mathbf{a} \in [-0.25, 0.25]$$

Illustrations - Constrained LQR problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ i.i.d. *on a restricted interval*.

Cost:

$$L(\mathbf{s}, \mathbf{a}) = \mathbf{a}^2 + (\mathbf{s} - 0.5)^2$$

Constraints:

$$\mathbf{a} \in [-0.25, 0.25]$$

MPC model:

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \mathbb{E}[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}] = \mathbf{s} + \mathbf{a}$$

Illustrations - Constrained LQR problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ i.i.d. *on a restricted interval.*

Cost:

$$L(\mathbf{s}, \mathbf{a}) = \mathbf{a}^2 + (\mathbf{s} - 0.5)^2$$

Constraints:

$$\mathbf{a} \in [-0.25, 0.25]$$

MPC model:

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \mathbb{E}[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}] = \mathbf{s} + \mathbf{a}$$

If $V^\star$ is **quadratic** i.e. $V^\star(\mathbf{s}) = \mathbf{s}^\top W \mathbf{s} + \mathbf{d}^\top \mathbf{s} + V_0$ then

Illustrations - Constrained LQR problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ i.i.d. *on a restricted interval.*

Cost:

$$L(\mathbf{s}, \mathbf{a}) = \mathbf{a}^2 + (\mathbf{s} - 0.5)^2$$

Constraints:

$$\mathbf{a} \in [-0.25, 0.25]$$

MPC model:

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \mathbb{E}[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}] = \mathbf{s} + \mathbf{a}$$

If $V^\star$ is **quadratic** i.e. $V^\star(\mathbf{s}) = \mathbf{s}^\top W \mathbf{s} + \mathbf{d}^\top \mathbf{s} + V_0$ then

$$\begin{aligned}
\mathbb{E}[V^\star(\mathbf{s}_+) \,|\, \mathbf{s}, \mathbf{a}] &= \mathbb{E}\left[\mathbf{s}_+^\top W \mathbf{s}_+ + \mathbf{d}^\top \mathbf{s}_+ + V_0 \,\middle|\, \mathbf{s}, \mathbf{a}\right] \\
&= \mathbf{f}(\mathbf{s}, \mathbf{a})^\top W \mathbf{f}(\mathbf{s}, \mathbf{a}) + \mathbf{d}^\top \mathbf{f}(\mathbf{s}, \mathbf{a}) + V_0 + \mathbb{E}\left[\mathbf{w}^\top W \mathbf{w}\right] \\
&= V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + \underbrace{\mathbb{E}\left[\mathbf{w}^\top W \mathbf{w}\right]}_{=c}
\end{aligned}$$

Illustrations - Constrained LQR problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ i.i.d.
on a restricted interval.

Cost:

$$L(\mathbf{s}, \mathbf{a}) = \mathbf{a}^2 + (\mathbf{s} - 0.5)^2$$

Constraints:

$$\mathbf{a} \in [-0.25, 0.25]$$

MPC model:

$$\mathbf{f}(\mathbf{s}, \mathbf{a}) = \mathbb{E}[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}] = \mathbf{s} + \mathbf{a}$$

If $V^\star$ is **quadratic** i.e. $V^\star(\mathbf{s}) = \mathbf{s}^\top W \mathbf{s} + \mathbf{d}^\top \mathbf{s} + V_0$ then

$$\begin{aligned}
\mathbb{E}[V^\star(\mathbf{s}_+) \,|\, \mathbf{s}, \mathbf{a}] &= \mathbb{E}\left[\mathbf{s}_+^\top W \mathbf{s}_+ + \mathbf{d}^\top \mathbf{s}_+ + V_0 \,\middle|\, \mathbf{s}, \mathbf{a}\right] \\
&= \mathbf{f}(\mathbf{s}, \mathbf{a})^\top W \mathbf{f}(\mathbf{s}, \mathbf{a}) + \mathbf{d}^\top \mathbf{f}(\mathbf{s}, \mathbf{a}) + V_0 + \mathbb{E}\left[\mathbf{w}^\top W \mathbf{w}\right] \\
&= V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) + \underbrace{\mathbb{E}\left[\mathbf{w}^\top W \mathbf{w}\right]}_{=c}
\end{aligned}$$

Hence theory predicts MPC produces optimal policy!

Illustrations - Constrained LQR problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

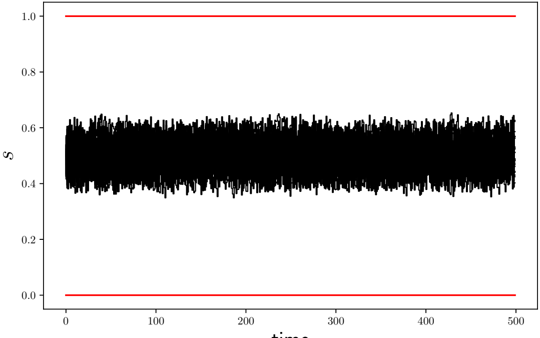
with $\mathbf{s}, \mathbf{a}$, $\mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ i.i.d. *on a restricted interval*.

Cost:

$$L(\mathbf{s}, \mathbf{a}) = \mathbf{a}^2 + (\mathbf{s} - 0.5)^2$$

Constraints:

$$\mathbf{a} \in [-0.25, 0.25]$$



Illustrations - Constrained LQR problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

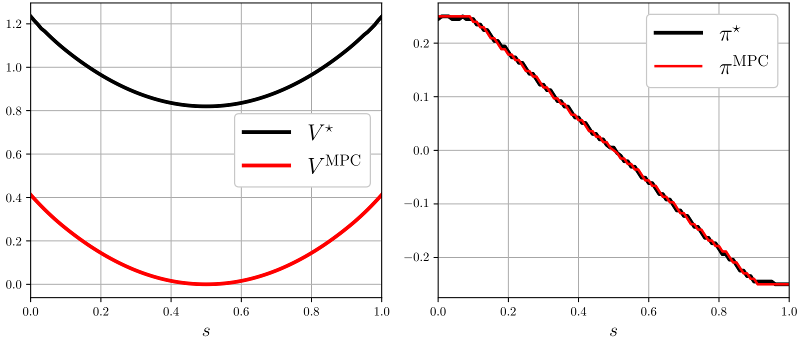
with $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ i.i.d. *on a restricted interval*.

Cost:

$$L(\mathbf{s}, \mathbf{a}) = \mathbf{a}^2 + (\mathbf{s} - 0.5)^2$$

Constraints:

$$\mathbf{a} \in [-0.25, 0.25]$$



Illustrations - Dissipative / Non-Smooth Problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with $\mathbf{s}, \mathbf{a}$, $\mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval*.

Cost:

$$L(\mathbf{s}, \mathbf{a}) = |\mathbf{a}| + |\mathbf{s} - 0.5|$$

Constraints:

$$\mathbf{a} \in [-0.25, 0.25]$$

Illustrations - Dissipative / Non-Smooth Problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

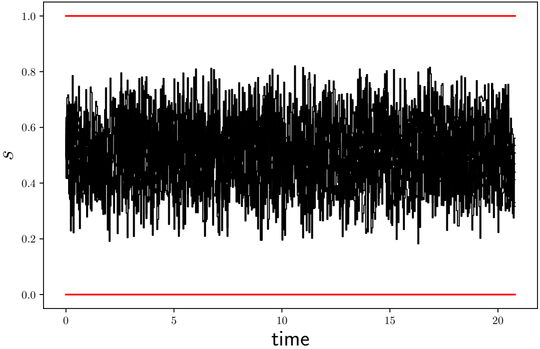
with $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval*.

Cost:

$$L(\mathbf{s}, \mathbf{a}) = |\mathbf{a}| + |\mathbf{s} - 0.5|$$

Constraints:

$$\mathbf{a} \in [-0.25, 0.25]$$



Illustrations - Dissipative / Non-Smooth Problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

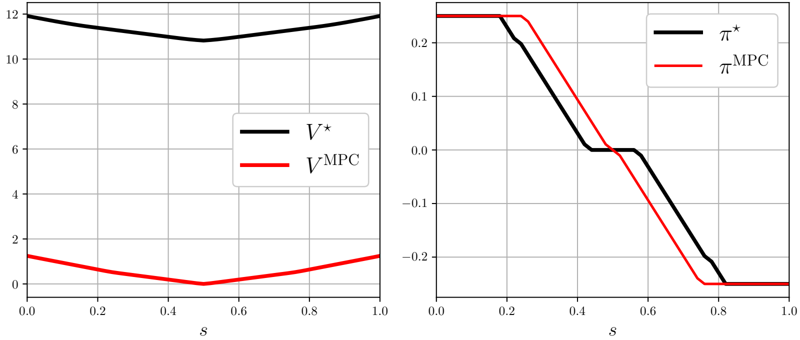
with $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval*.

Cost:

$$L(\mathbf{s}, \mathbf{a}) = |\mathbf{a}| + |\mathbf{s} - 0.5|$$

Constraints:

$$\mathbf{a} \in [-0.25, 0.25]$$



Illustrations - Non-Dissipative / Non-smooth Problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval*.

What is this problem?

Cost:

$$L(\mathbf{s}, \mathbf{a}) = \begin{cases} \mathbf{a} & \text{if} \quad \mathbf{a} \leq 0 \\ 2\mathbf{a} & \text{if} \quad \mathbf{a} > 0 \end{cases}$$

Constraints:

$$\mathbf{s} \in [0, 1], \quad \mathbf{a} \in [-0.25, 0.25]$$

Illustrations - Non-Dissipative / Non-smooth Problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval*.

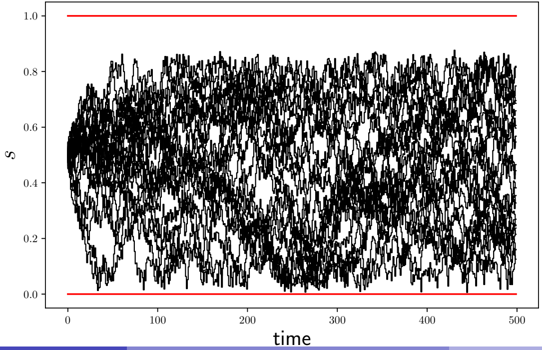
What is this problem?

Cost:

$$L(\mathbf{s}, \mathbf{a}) = \begin{cases} \mathbf{a} & \text{if} \quad \mathbf{a} \leq 0 \\ 2\mathbf{a} & \text{if} \quad \mathbf{a} > 0 \end{cases}$$

Constraints:

$$\mathbf{s} \in [0, 1], \quad \mathbf{a} \in [-0.25, 0.25]$$



Illustrations - Non-Dissipative / Non-smooth Problem

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$, and $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval.*

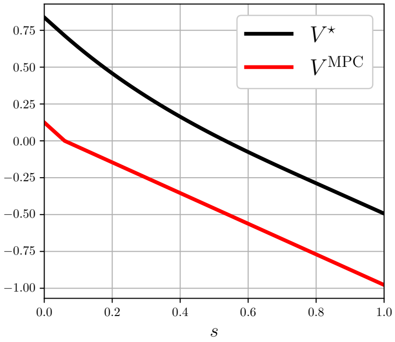
What is this problem?

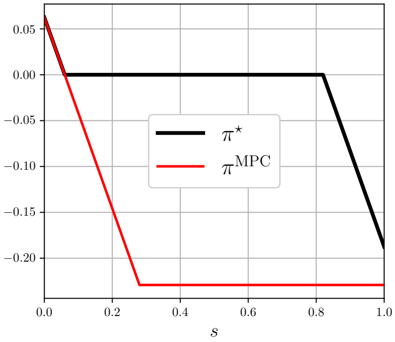
Cost:

$$L(\mathbf{s}, \mathbf{a}) = \begin{cases} \mathbf{a} & \text{if} \quad \mathbf{a} \leq 0 \\ 2\mathbf{a} & \text{if} \quad \mathbf{a} > 0 \end{cases}$$

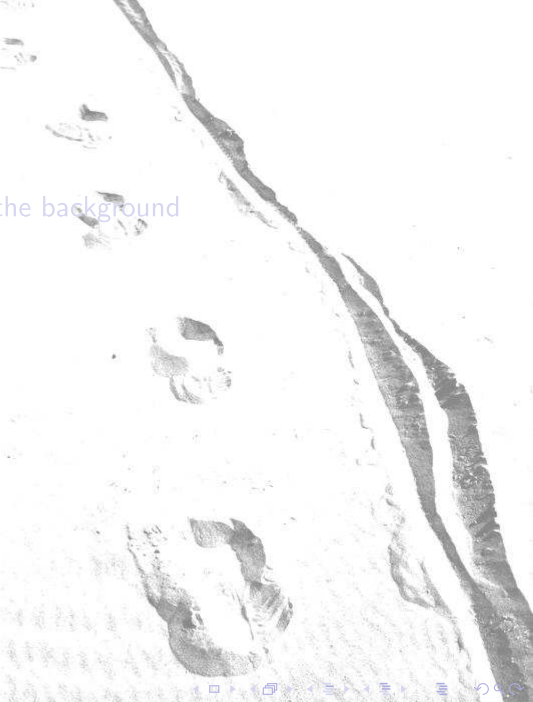
Constraints:

$$\mathbf{s} \in [0, 1], \quad \mathbf{a} \in [-0.25, 0.25]$$





Outline



MPC model & Optimality

Model $\mathrm{f}(\mathrm{s}, \mathrm{a})$ such that

$$V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) = \mathbb{E}[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}] - c \quad (1)$$

holds for some c ???

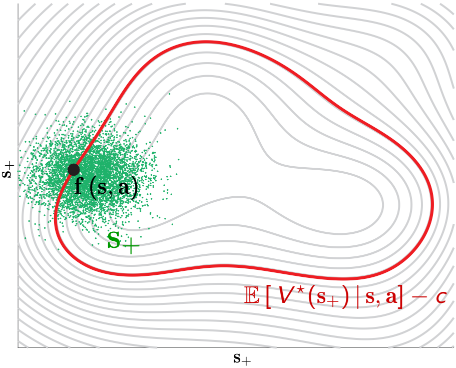
Stochastic state transition

$$\varrho[\mathbf{s}_+ \mid \mathbf{s}, \mathbf{a}]$$

Model likelihood:

$$\varrho[\mathbf{f}(\mathbf{s}, \mathbf{a}) \mid \mathbf{s}, \mathbf{a}] > 0$$

is desired, ideally maximal

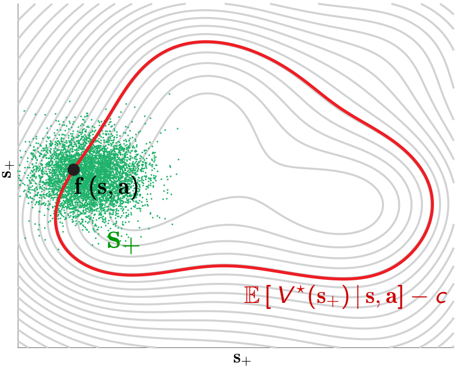


MPC model & Optimality

Model $\mathrm{f}(\mathrm{s}, \mathrm{a})$ such that

$$V^\star\left(\mathbf{f}\left(\mathbf{s}, \mathbf{a}\right)\right) = \mathbb{E}\left[V^\star\left(\mathbf{s}_+\right) \,|\, \mathbf{s}, \mathbf{a}\right] - c \tag{1}$$

holds for some c ???



Stochastic state transition

$$\varrho\left[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}\right]$$

Model likelihood:

$$\varrho\left[\mathbf{f}\left(\mathbf{s}, \mathbf{a}\right) \,|\, \mathbf{s}, \mathbf{a}\right] > 0$$

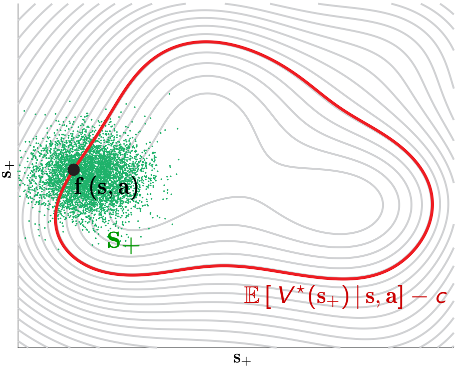
is desired, ideally maximal

MPC model & Optimality

Model $\mathbf{f}(\mathbf{s}, \mathbf{a})$ **such that**

$$V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) = \mathbb{E}[V^\star(\mathbf{s}_+) \,|\, \mathbf{s}, \mathbf{a}] - c \qquad (1)$$

holds for some c ???



Stochastic state transition

$$\varrho[\mathbf{s}_+ \,|\, \mathbf{s}, \mathbf{a}]$$

Model likelihood:

$$\varrho[\mathbf{f}(\mathbf{s}, \mathbf{a}) \,|\, \mathbf{s}, \mathbf{a}] > 0$$

is desired, ideally maximal

Existence yes... but not fully clear yet

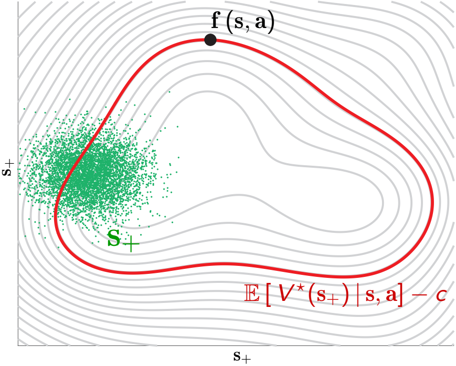
For $c = 0$ and $V^\star$ continuous and ϱ of convex support, there is a $\mathbf{f}(\mathbf{s}, \mathbf{a})$ with $\varrho[\mathbf{f}(\mathbf{s}, \mathbf{a}) \,|\, \mathbf{s}, \mathbf{a}] > 0$ and (1) for all $\mathbf{s}, \mathbf{a}$.

MPC model & Optimality

Model $\mathrm{f}(\mathrm{s}, \mathrm{a})$ such that

$$V^\star\left(\mathbf{f}\left(\mathrm{s}, \mathbf{a}\right)\right) = \mathbb{E}\left[V^\star\left(\mathrm{s}_+\right) \,|\, \mathrm{s}, \mathbf{a}\right] - c \qquad (1)$$

holds for some c ???



Stochastic state transition

$$\varrho\left[\mathrm{s}_+ \,|\, \mathrm{s}, \mathbf{a}\right]$$

Model likelihood:

$$\varrho\left[\mathbf{f}\left(\mathrm{s}, \mathbf{a}\right) \,|\, \mathrm{s}, \mathbf{a}\right] > 0$$

is desired, ideally maximal

Uniqueness no... but max likelihood

$$\mathbf{f}\left(\mathrm{s}, \mathbf{a}\right) = \arg\max_{\hat{\mathrm{s}}_+} \quad \varrho\left[\hat{\mathrm{s}}_+ \,|\, \mathrm{s}, \mathbf{a}\right]$$
$$\text{s.t.} \quad (1)$$

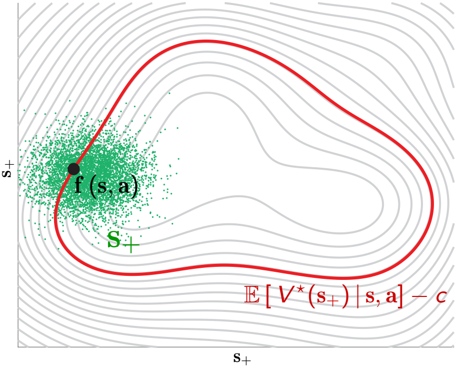
would typically fix that

MPC model & Optimality

Model $\mathbf{f}(\mathbf{s}, \mathbf{a})$ such that

$$V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) = \mathbb{E}[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}] - c \quad (1)$$

holds for some c ???



Stochastic state transition

$$\varrho[\mathbf{s}_+ \mid \mathbf{s}, \mathbf{a}]$$

Model likelihood:

$$\varrho[\mathbf{f}(\mathbf{s}, \mathbf{a}) \mid \mathbf{s}, \mathbf{a}] > 0$$

is desired, ideally maximal

Continuity not necessarily... must be imposed

Illustrations - Optimal MPC model

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with

- $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$
- $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval*

Cost:

$$L(\mathbf{s}, \mathbf{a}) = |\mathbf{a}| + |\mathbf{s} - 0.5|$$

Constraints:

$$\mathbf{s} \in [0, 1], \quad \mathbf{a} \in [-0.25, 0.25]$$

Illustrations - Optimal MPC model

Consider the dynamics:

$$\mathbf{s}_{+} = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with

- $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$
- $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval*

Cost:

$$L(\mathbf{s}, \mathbf{a}) = |\mathbf{a}| + |\mathbf{s} - 0.5|$$

Constraints:

$$\mathbf{s} \in [0, 1], \quad \mathbf{a} \in [-0.25, 0.25]$$

Expected value model:

$$\mathbb{E}[\mathbf{s}_{+}] = \mathbf{s} + \mathbf{a}$$

Illustrations - Optimal MPC model

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with

- $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$
- $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval*

Cost:

$$L(\mathbf{s}, \mathbf{a}) = |\mathbf{a}| + |\mathbf{s} - 0.5|$$

Constraints:

$$\mathbf{s} \in [0, 1], \quad \mathbf{a} \in [-0.25, 0.25]$$

Expected value model:

$$\mathbb{E}[\mathbf{s}_+] = \mathbf{s} + \mathbf{a}$$

Max Likelihood optimal model(s)

$$\mathbf{f}_c(\mathbf{s}, \mathbf{a}) =$$

$$\begin{aligned} \arg\max_{\hat{\mathbf{s}}_+} \quad & \varrho\left[\hat{\mathbf{s}}_+ \,|\, \mathbf{s}, \mathbf{a}\right] \\ \text{s.t.} \quad & V^\star(\hat{\mathbf{s}}_+) = \mathbb{E}\left[V^\star(\mathbf{s}_+) \,|\, \mathbf{s}, \mathbf{a}\right] - c \end{aligned}$$

Illustrations - Optimal MPC model

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with

- $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$
- $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval*

Cost:

$$L(\mathbf{s}, \mathbf{a}) = |\mathbf{a}| + |\mathbf{s} - 0.5|$$

Constraints:

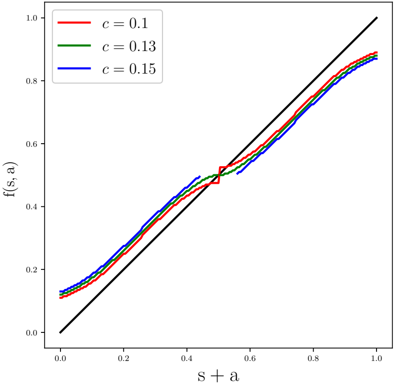
$$\mathbf{s} \in [0, 1], \quad \mathbf{a} \in [-0.25, 0.25]$$

Expected value model:

$$\mathbb{E}[\mathbf{s}_+] = \mathbf{s} + \mathbf{a}$$

Max Likelihood optimal model(s)

$$\mathbf{f}_c(\mathbf{s}, \mathbf{a}) = \arg\max_{\hat{\mathbf{s}}_+} \quad \varrho\left[\hat{\mathbf{s}}_+ \,|\, \mathbf{s}, \mathbf{a}\right] \quad \text{s.t.} \quad V^\star(\hat{\mathbf{s}}_+) = \mathbb{E}\left[V^\star(\mathbf{s}_+) \,|\, \mathbf{s}, \mathbf{a}\right] - c$$



Illustrations - Optimal MPC model

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with

- $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$
- $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval*

Cost:

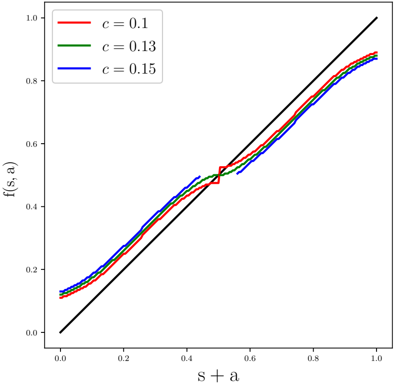
$$L(\mathbf{s}, \mathbf{a}) = |\mathbf{a}| + |\mathbf{s} - 0.5|$$

Constraints:

$$\mathbf{s} \in [0, 1], \quad \mathbf{a} \in [-0.25, 0.25]$$

Observations

- Existence: not for all c
- Continuity: not for all c
- Both: specific c(?)
- Linearity is lost



Illustrations - Optimal MPC model

Consider the dynamics:

$$\mathbf{s}_+ = \mathbf{s} + \mathbf{a} + \mathbf{w}$$

with

- $\mathbf{s}, \mathbf{a}, \mathbf{w} \in \mathbb{R}$
- $\mathbf{w} \sim \mathcal{N}(0, \sigma)$ *on a restricted interval*

Cost:

$$L(\mathbf{s}, \mathbf{a}) = |\mathbf{a}| + |\mathbf{s} - 0.5|$$

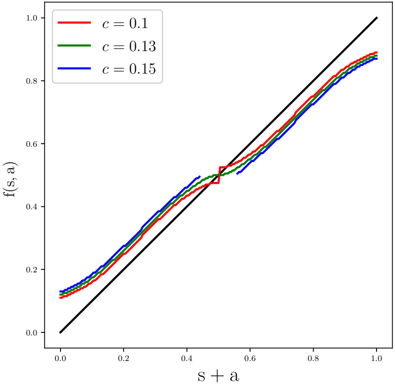
Constraints:

$$\mathbf{s} \in [0, 1], \quad \mathbf{a} \in [-0.25, 0.25]$$

Observations

- Existence: not for all c
- Continuity: not for all c
- Both: specific $c(?)$
- Linearity is lost

This is to be further investigated

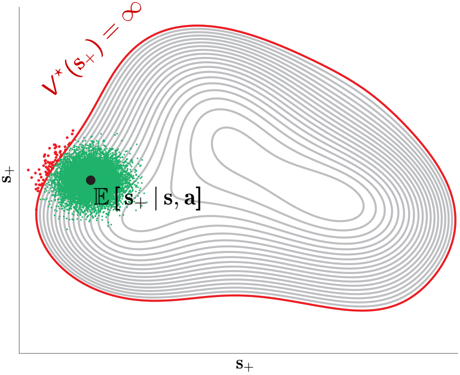


Optimal MPC model & Safety

Model $\mathbf{f}(\mathbf{s},\mathbf{a})$ such that

$$V^\star(\mathbf{f}(\mathbf{s},\mathbf{a})) = \mathbb{E}[V^\star(\mathbf{s}_+) \mid \mathbf{s},\mathbf{a}] - c$$

holds for some c ???



What if $\mathbf{s},\mathbf{a} \rightarrow$ possible infeasibility? I.e.

$$\mathbb{E}[V^\star(\mathbf{s}_+) \mid \mathbf{s},\mathbf{a}] = \infty \qquad (1)$$

holds.

What should the model do then?

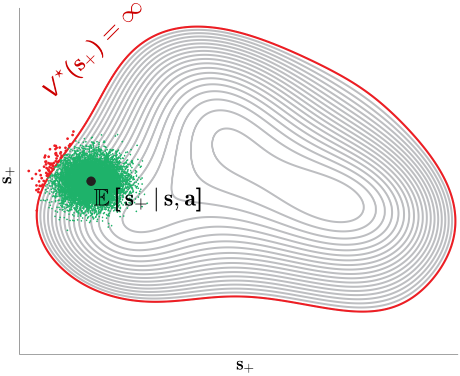
Â

Optimal MPC model & Safety

Model $\mathbf{f}(\mathbf{s},\mathbf{a})$ such that

$$V^\star(\mathbf{f}(\mathbf{s},\mathbf{a})) = \mathbb{E}\left[V^\star(\mathbf{s}_+)\,|\,\mathbf{s},\mathbf{a}\right] - c$$

holds for some c ???



What if $\mathbf{s},\mathbf{a} \rightarrow$ possible infeasibility? I.e.

$$\mathbb{E}\left[V^\star(\mathbf{s}_+)\,|\,\mathbf{s},\mathbf{a}\right] = \infty \qquad (1)$$

holds.

What should the model do then?

- (1) implies that:

$$\mathbb{P}\left[\,V^\star(\mathbf{s}_+) = \infty\,|\,\mathbf{s},\mathbf{a}\,\right] > 0$$

i.e. $\mathbf{s}_+$ may land where $V^\star$ is ∞

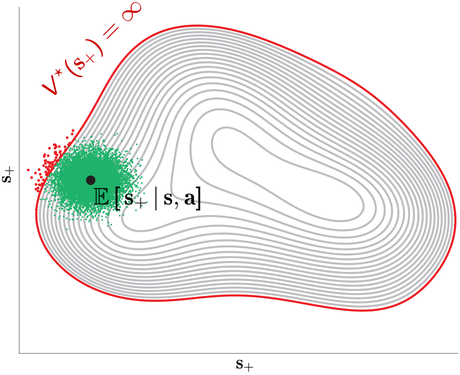
Â

Optimal MPC model & Safety

Model $\mathbf{f}(\mathbf{s},\mathbf{a})$ **such that**

$$V^\star(\mathbf{f}(\mathbf{s},\mathbf{a})) = \mathbb{E}[V^\star(\mathbf{s}_+) \,|\, \mathbf{s},\mathbf{a}] - c$$

holds for some c ???



What if $\mathbf{s},\mathbf{a} \rightarrow$ **possible infeasibility?** I.e.

$$\mathbb{E}[V^\star(\mathbf{s}_+) \,|\, \mathbf{s},\mathbf{a}] = \infty \qquad (1)$$

holds.

What should the model do then?

- (1) implies that:

  $$\mathbb{P}[\,V^\star(\mathbf{s}_+) = \infty \,|\, \mathbf{s},\mathbf{a}\,] > 0$$

  i.e. $\mathbf{s}_+$ may land where $V^\star$ is ∞

- Model must reproduce that, i.e.

  $$V^\star(\mathbf{f}(\mathbf{s},\mathbf{a})) = \infty$$

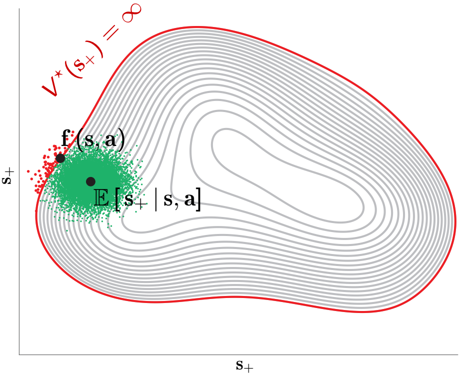
  i.e. $\mathbf{f}(\mathbf{s},\mathbf{a})$ picks a point among the infeasible onesÂ

Optimal MPC model & Safety

Model $\mathbf{f}(\mathbf{s}, \mathbf{a})$ such that

$$V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) = \mathbb{E}[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}] - c$$

holds for some c ???



What if $\mathbf{s}, \mathbf{a} \to$ possible infeasibility? I.e.

$$\mathbb{E}[V^\star(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}] = \infty \qquad (1)$$

holds.

What should the model do then?

- (1) implies that:

$$\mathbb{P}[V^\star(\mathbf{s}_+) = \infty \mid \mathbf{s}, \mathbf{a}] > 0$$

  i.e. $\mathbf{s}_+$ may land where $V^\star$ is ∞

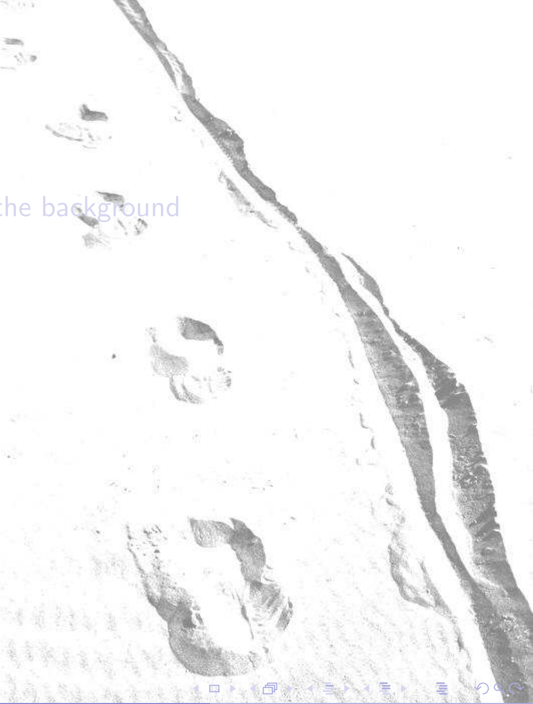
- Model must reproduce that, i.e.

$$V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) = \infty$$

  i.e. $\mathbf{f}(\mathbf{s}, \mathbf{a})$ picks a point among the infeasible onesÂ

Theory requires conservative model for constraints violations

Outline



MPC with stochastic models?

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

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

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

$$\begin{aligned} \min_{\mathbf{u}_0, \boldsymbol{\pi}_{1,\dots,N-1}} \quad & \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right] \\ \text{s.t} \quad & \mathbf{x}_{k+1} \sim \hat{\varrho}[\,\cdot\,|\,\mathbf{x}_k, \mathbf{u}_k\,], \\ & \mathbf{x}_0 = \mathbf{s}, \quad \mathbf{u}_k = \boldsymbol{\pi}_k(\mathbf{x}_k), \; k > 0 \end{aligned}$$

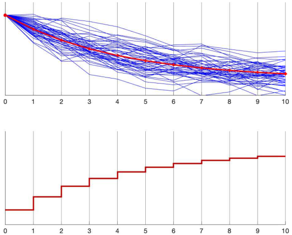
MPC with stochastic models?

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

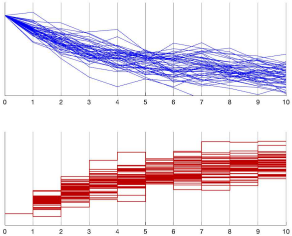
$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & T(\mathbf{x}_N) + \sum_{k=0}^{N-1} L(\mathbf{x}_k, \mathbf{u}_k) \\ \text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \\ & \mathbf{x}_0 = \mathbf{s} \end{aligned}$$

MPC policy $\pi^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

$$\begin{aligned} \min_{\mathbf{u}_0, \boldsymbol{\pi}_{1,\ldots,N-1}} \quad & \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right] \\ \text{s.t} \quad & \mathbf{x}_{k+1} \sim \hat{\varrho}[\,.\,|\,\mathbf{x}_k, \mathbf{u}_k\,], \\ & \mathbf{x}_0 = \mathbf{s}, \quad \mathbf{u}_k = \boldsymbol{\pi}_k(\mathbf{x}_k), \; k > 0 \end{aligned}$$



Policy from planning

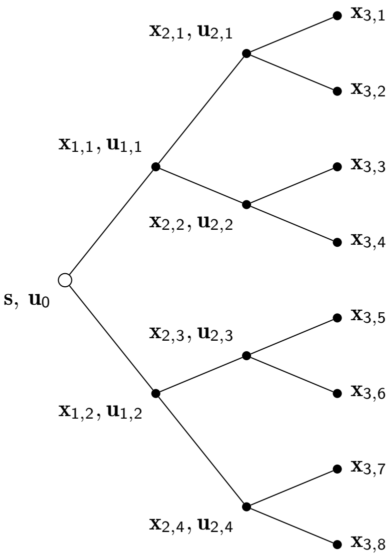


Proper policing, but difficult

MPC with stochastic models - Scenario trees

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ **from**

$$\begin{aligned} \min_{\mathbf{u}_0, \boldsymbol{\pi}_{1,\ldots,N-1}} \quad & \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k) \right] \\ \text{s.t} \quad & \mathbf{x}_{k+1} \sim \hat{\varrho}[\,.\,|\,\mathbf{x}_k, \mathbf{u}_k\,], \\ & \mathbf{x}_0 = \mathbf{s}, \quad \mathbf{u}_k = \boldsymbol{\pi}_k(\mathbf{x}_k),\ k > 0 \end{aligned}$$



Scenario tree MPC

- $\hat{\varrho}$ is a discrete probability distribution
- Tree of scenarios
- Implicitly produces decision policies
- Exploding complexity over horizon

MPC with stochastic models - Other methods

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

$$\begin{aligned}\min_{\mathbf{u}_0, \boldsymbol{\pi}_{1,\ldots,N-1}} \quad & \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1} \gamma^k L(\mathbf{x}_k, \mathbf{u}_k)\right] \\ \text{s.t} \quad & \mathbf{x}_{k+1} \sim \hat{\varrho}[\,.\,|\,\mathbf{x}_k, \mathbf{u}_k\,], \\ & \mathbf{x}_0 = \mathbf{s}, \quad \mathbf{u}_k = \boldsymbol{\pi}_k(\mathbf{x}_k), \; k > 0\end{aligned}$$

"Spectral" representations of $\hat{\varrho}$

- Gaussian Processes
- Polynomial Chaos Expansion
- RKHS

Representations of $\boldsymbol{\pi}_{1,\ldots,N-1}$

- Linear feedback

$$\boldsymbol{\pi}_k(\mathbf{s}_k) = \bar{\mathbf{u}}_k - K_k(\mathbf{s}_k - \bar{\mathbf{s}}_k)$$

- More advance forms...

Optimal MPC from stochastic models?

Model $\mathbf{f}(\mathbf{s}, \mathbf{a})$ **such that**

$$V^\star(\mathbf{f}(\mathbf{s}, \mathbf{a})) = \mathbb{E}\left[V^\star(\mathbf{s}_+) \,|\, \mathbf{s}, \mathbf{a}\right] - c$$

Optimal MPC from stochastic models?

Model $\mathrm{f}(\mathrm{s}, \mathrm{a})$ such that

$$V^{\star}(\mathbf{f}(\mathbf{s}, \mathbf{a})) = \mathbb{E}\left[V^{\star}(\mathbf{s}_{+}) \mid \mathbf{s}, \mathbf{a}\right] - c$$

Generalization?

Model $\hat{\mathrm{s}}_{+} \sim \hat{\varrho}[\,.\,|\, \mathrm{s}, \mathrm{a}\,]$ such that

$$\mathbb{E}\left[V^{\star}(\hat{\mathbf{s}}_{+}) \mid \mathbf{s}, \mathbf{a}\right] = \mathbb{E}\left[V^{\star}(\mathbf{s}_{+}) \mid \mathbf{s}, \mathbf{a}\right] - c$$

Optimal MPC from stochastic models?

Model $\mathrm{f}(\mathrm{s},\mathrm{a})$ **such that**

$$V^\star(\mathbf{f}(\mathbf{s},\mathbf{a})) = \mathbb{E}\left[V^\star(\mathbf{s}_+)\,|\,\mathbf{s},\mathbf{a}\right] - c$$

Generalization?

Model $\hat{\mathrm{s}}_+ \sim \hat{\varrho}[\,.\,|\,\mathrm{s},\mathrm{a}\,]$ **such that**

$$\mathbb{E}\left[V^\star(\hat{\mathbf{s}}_+)\,|\,\mathbf{s},\mathbf{a}\right] = \mathbb{E}\left[V^\star(\mathbf{s}_+)\,|\,\mathbf{s},\mathbf{a}\right] - c$$

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathrm{s}) = \mathbf{u}_0^\star$ from

$$\begin{aligned}
\min_{\mathbf{u}_0,\boldsymbol{\pi}_{1,\ldots,N-1}} \quad & \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1}\gamma^k L(\mathbf{x}_k,\mathbf{u}_k)\right] \\
\mathrm{s.t} \quad & \mathbf{x}_{k+1} \sim \hat{\varrho}[\,.\,|\,\mathbf{x}_k,\mathbf{u}_k\,], \\
& \mathbf{x}_0 = \mathbf{s}, \quad \mathbf{u}_k = \boldsymbol{\pi}_k(\mathbf{x}_k)
\end{aligned}$$

yields $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathrm{s}) = \boldsymbol{\pi}^\star(\mathrm{s})$ (with correct T)

Optimal MPC from stochastic models?

Model $\mathrm{f}(\mathrm{s},\mathrm{a})$ **such that**

$$V^\star\left(\mathbf{f}\left(\mathbf{s},\mathbf{a}\right)\right)=\mathbb{E}\left[V^\star\left(\mathbf{s}_+\right)\,|\,\mathbf{s},\mathbf{a}\right]-c$$

Generalization?

Model $\hat{\mathrm{s}}_+\sim\hat{\varrho}[\,.\,|\,\mathrm{s},\mathrm{a}\,]$ **such that**

$$\mathbb{E}\left[V^\star\left(\hat{\mathbf{s}}_+\right)\,|\,\mathbf{s},\mathbf{a}\right]=\mathbb{E}\left[V^\star\left(\mathbf{s}_+\right)\,|\,\mathbf{s},\mathbf{a}\right]-c$$

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathrm{s})=\mathbf{u}_0^\star$ from

$$\begin{aligned}\min_{\mathbf{u}_0,\boldsymbol{\pi}_{1,\ldots,N-1}}\quad & \mathbb{E}\left[T(\mathbf{x}_N)+\sum_{k=0}^{N-1}\gamma^k L(\mathbf{x}_k,\mathbf{u}_k)\right]\\ \mathrm{s.t}\quad & \mathbf{x}_{k+1}\sim\hat{\varrho}[\,.\,|\,\mathbf{x}_k,\mathbf{u}_k\,],\\ & \mathbf{x}_0=\mathbf{s},\quad \mathbf{u}_k=\boldsymbol{\pi}_k\left(\mathbf{x}_k\right)\end{aligned}$$

yields $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathrm{s})=\boldsymbol{\pi}^\star(\mathrm{s})$ (with correct T)

To be further explored!

Optimal MPC from stochastic models?

Model $\mathbf{f}(\mathbf{s},\mathbf{a})$ such that

$$V^\star(\mathbf{f}(\mathbf{s},\mathbf{a})) = \mathbb{E}[V^\star(\mathbf{s}_+)\,|\,\mathbf{s},\mathbf{a}] - c$$

Generalization?

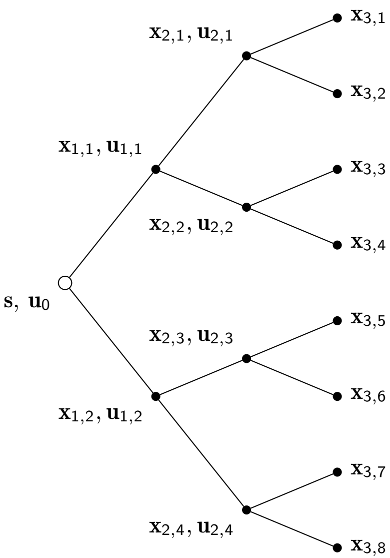
Model $\hat{\mathbf{s}}_+ \sim \hat{\varrho}[\,.\,|\,\mathbf{s},\mathbf{a}\,]$ such that

$$\mathbb{E}[V^\star(\hat{\mathbf{s}}_+)\,|\,\mathbf{s},\mathbf{a}] = \mathbb{E}[V^\star(\mathbf{s}_+)\,|\,\mathbf{s},\mathbf{a}] - c$$

MPC policy $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \mathbf{u}_0^\star$ from

$$\begin{aligned}\min_{\mathbf{u}_0,\boldsymbol{\pi}_{1,\ldots,N-1}} \quad & \mathbb{E}\left[T(\mathbf{x}_N) + \sum_{k=0}^{N-1}\gamma^k L(\mathbf{x}_k,\mathbf{u}_k)\right] \\ \mathrm{s.t} \quad & \mathbf{x}_{k+1} \sim \hat{\varrho}[\,.\,|\,\mathbf{x}_k,\mathbf{u}_k\,], \\ & \mathbf{x}_0 = \mathbf{s}, \quad \mathbf{u}_k = \boldsymbol{\pi}_k(\mathbf{x}_k)\end{aligned}$$

yields $\boldsymbol{\pi}^{\mathrm{MPC}}(\mathbf{s}) = \boldsymbol{\pi}^\star(\mathbf{s})$ (with correct T)



To be further explored!

Optimal MPC from stochastic models?

Model $\mathbf{f}(\mathbf{s},\mathbf{a})$ such that

$$V^\star(\mathbf{f}(\mathbf{s},\mathbf{a})) = \mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s},\mathbf{a}\right] - c$$

Discrete Stochastic Model

Models $\mathbf{f}_{1,\ldots,m}(\mathbf{s},\mathbf{a})$ such that

$$\sum_{i=1}^{m} \omega^i V^\star(\mathbf{f}_i(\mathbf{s},\mathbf{a})) = \mathbb{E}\left[V^\star(\mathbf{s}_+) \mid \mathbf{s},\mathbf{a}\right] - c$$

this describes a scenario tree !!