Learning-Based Predictive Control

Chapter 2a
Classification of Learning-based Predictive Control

Prof. Melanie Zeilinger

ETH Zurich

2023

Coauthor: Dr. Lukas Hewing

Basic MPC Problem Formulation

$$
\begin{aligned}
\min_{U} \; & \sum_{i=0}^{N-1} l(x_i, u_i) & \\
\text{s.t. } & U = \{u_0, u_1, \dots, u_{N-1}\} & \text{optimization variables} \\
& x_{i+1} = f(x_i, u_i) & \text{system model} \\
& x_i \in \mathcal{X} & \text{state constraints} \\
& u_i \in \mathcal{U} & \text{input constraints} \\
& x_0 = x(k) & \text{measurement/initialization}
\end{aligned}
$$

Challenges:
Complex tasks and objectives, complex dynamics, uncertainties,...

Outline

1. The ideal optimal control problem & approximations

Outline

1. The ideal optimal control problem & approximations

   System dynamics & uncertainties

   Cost function

   Constraints

The Real World

Simplifying assumption:

$$x(k+1) = f(x(k), u(k))$$

- System evolves in a predictable fashion

The real world:

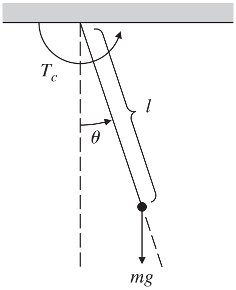
$$x(k+1) = f(x(k), u(k), w(k); \theta)$$

- Model structure is unknown (and potentially also state dimension)
- Random noise w changes the evolution of the system
- Unknown parameters θ impact the dynamics

Note: w is random, θ is unknown constant/time-varying

Parametric uncertainties

Example: Pendulum with unknown weight or length



$$m\, l^2\, \ddot{\theta} = T_c - m\, g\, l \sin(\theta)$$

$\rightarrow$ Uncertain dynamics

Source of parametric uncertainties in practice:

- Production tolerances/variations
- Parameters cannot be measured directly (e.g. tire model)
- Parameters may change over time (e.g. tires)

$\rightarrow$ Parameter uncertainties can often be learned (in principle)

- $\theta \sim \mathcal{Q}^\theta$ (stochastic treatment)
- $\theta \in \Theta$ (robust treatment)

Model Mismatch

True system model is typically not available

- accurate modeling too time/cost intensive
- complex system model not suitable for controller design
- state used in control typically reduced

Structural model uncertainty can be addressed using robust bounds

$$\|f_t(x,u) - f(x,u)\| \leq \bar{w} \quad \forall x, u$$

$\rightarrow$ Global bound difficult to obtain and can be very conservative

Disturbances

$$x(k+1) = f(x(k), u(k), \mathbf{w(k)})$$

Typical forms:

- Additive: $f(x,u) + \mathbf{w}$
- Multiplicative: $f(x,u) + \mathbf{g(x,u)w}$

Robust Approach

$$f(x,u) + \mathbf{w}, \text{ with } \mathbf{w} \in \mathcal{W}$$

Blanket term for all uncertainties (disturbances, model mismatch etc.!), but robust controller can be overly cautious/conservative

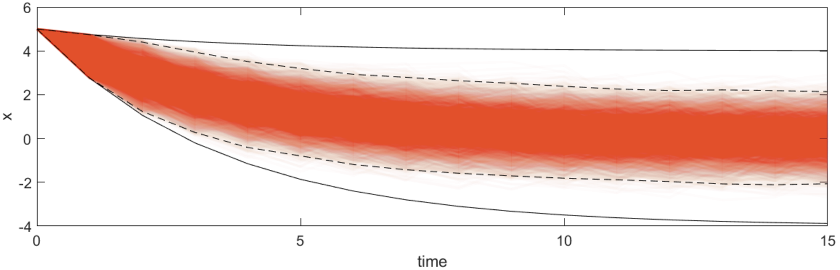
Stochastic Disturbances

$$f(x,u) + \mathbf{w}, \text{ with } \mathbf{w} \sim \mathcal{Q}^w, \text{ i.i.d.}$$

Can enable good practical solutions, **but** subject to specific assumptions

Example: Robust vs. Stochastic Treatment

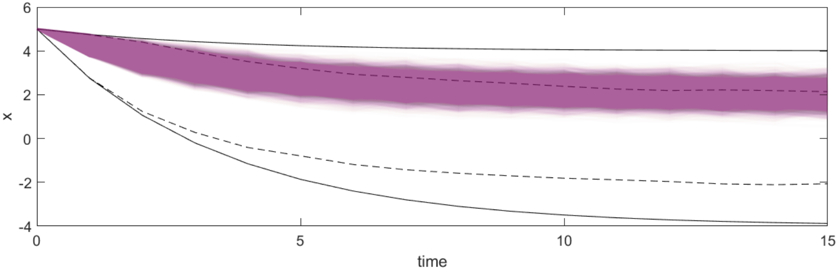
Scalar Linear system $\quad x(k+1) = 0.75x(k) + w(k)$
Additive Disturbance $\quad$ i.i.d. $w(k) \sim \mathcal{U}(-1, 1)$



- Robust bound encloses all trajectories, but 99% within dashed lines.
- Stochastic statement "more informative", but subject to specific assumptions

Example: Robust vs. Stochastic Treatment

Scalar Linear system $x(k+1) = 0.75x(k) + w(k)$
Additive Disturbance i.i.d. $\mathbf{w(k)} \sim \mathcal{U}(\mathbf{0}, \mathbf{1})$ **("$\mathcal{U}(-0.5, 0.5) + 0.5$", steady-state offset)**

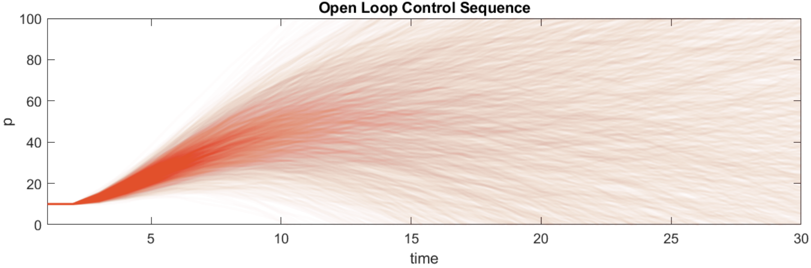


- Robust bound encloses all trajectories, but 99% within dashed lines.
- Stochastic statement "more informative", but subject to specific assumptions
- Robust bound applies whenever $x(k+1) - 0.75x(k) \in [-1, 1]$
  (model uncertainties, steady-state offset, state reduction, time-delays, ...)

Optimization over Feedback Policies

When considering model uncertainties, state evolution is not fully determined by input $u(k)$

- Simple control sequence $U = \{u(0), \ldots, u(\bar{N}-1)\}$ is suboptimal and may be insufficient



- Optimization over policies $u(k) = \pi_k(\cdot)$ (with access to all past states & inputs)

$$\Pi = \{\pi_0, \ldots, \pi_{\bar{N}-1}\}, \text{ with } \pi_k(x(0), \ldots, x(k), u(0), \ldots, u(k-1))$$

Special case: time-invariant state feedback $\pi_k = \pi(x)$

Outline

1. The ideal optimal control problem & approximations

   System dynamics & uncertainties

   Cost function

   Constraints

Cost function

Nominal cost: $$J(x(0),U) = \sum_{k=0}^{\bar{N}} l_k(x(k),u(k)), \text{ where } x(k+1) = f(x(k),u(k))$$

or $$J(x(0),U) = L(X,U) \text{ with } U = \{u(0),\dots,u(\bar{N}-1)\}, X = \{x(0),\dots,x(\bar{N})\}$$

- Optimize over complete task horizon $\bar{N}$ (possibly infinite)
- L can in principle represent complex objectives

Example:

Enforce terminal constraint $x(\bar{N}) \in \mathcal{X}_f$ via indicator function

$$L(X,U) = \mathbf{I}_{\mathcal{X}_f}(x(\bar{N}))$$

Cost function

What is a suitable cost for a “real” system, where

$$
\begin{aligned}
x(k+1) &= f(x(k), u(k), w(k), \theta) \\
u(k) &= \pi_k(x(0), \ldots, x(k), u(0), \ldots, u(k-1))
\end{aligned}
$$

and both X and U are functions of x_0, Π, $W = \{w(0), \ldots, w(\bar{N}-1)\}$ and θ?

- Minimize the expected value (requires some assumption on the distribution of w, θ)

$$J(x(0), \Pi) := \mathbb{E}\left[L(X(x(0), \Pi, W, \theta), U(x(0), \Pi, W, \theta))\right]$$

- Take the worst-case

$$J(x(0), \Pi) := \max_{W \in \mathcal{W}^{\bar{N}}, \theta \in \Theta} L(X(x(0), \Pi, W, \theta), U(x(0), \Pi, W, \theta))$$

- Take the nominal case

$$J(x(0), \Pi) := L(X(x(0), \Pi, 0, \theta_{\text{nom}}), U(x(0), \Pi, 0, \theta_{\text{nom}})) = L(X_{\text{nom}}, U_{\text{nom}})$$

Outline

1. The ideal optimal control problem & approximations

    System dynamics & uncertainties

    Cost function

    Constraints

Constraints

- Nominal control:

$$x(k) \in \mathcal{X}, u(k) \in \mathcal{U} \quad \forall k$$

- Robust constraint satisfaction for bounded uncertainties/disturbances:

$$x(k) \in \mathcal{X}, u(k) \in \mathcal{U} \quad \forall k, \forall w(k) \in \mathcal{W}, \theta \in \Theta$$

$\rightarrow$ satisfy constraints for all disturbance realizations
$\rightarrow$ generally achieved by using the concept of robust invariance

Constraints

- Nominal control:

$$x(k) \in \mathcal{X}, u(k) \in \mathcal{U} \quad \forall k$$

- Robust constraint satisfaction for bounded uncertainties/disturbances:

$$x(k) \in \mathcal{X}, u(k) \in \mathcal{U} \quad \forall k, \forall w(k) \in \mathcal{W}, \theta \in \Theta$$

$\rightarrow$ satisfy constraints for all disturbance realizations
$\rightarrow$ generally achieved by using the concept of robust invariance

- Probabilistic constraint satisfaction for stochastic uncertainties/disturbances:

$$Pr\left(x(k) \in \mathcal{X}, u(k) \in \mathcal{U}\right) \geq p \quad \forall k,\ w \sim \mathcal{Q}^w, \theta \sim \mathcal{Q}^\theta$$

$\rightarrow$ also called chance constraints

Putting things together: The "true" optimal control problem

Optimal control problem that we ideally would like to solve – the robust case

$$
\begin{aligned}
\min_{\{\pi_k\}} \quad & \max_{W,\theta} L(X,U) \\
\text{s.t.} \quad & x(k+1) = f(x(k), u(k), w(k); \theta), \\
& u(k) = \pi_k(x(0), \dots, x(k), u(0) \dots, u(k-1)), \\
& X \in \mathcal{X}^{\bar{N}}, U \in \mathcal{U}^{\bar{N}} \ \forall W \in \mathcal{W}^{\bar{N}}, \theta \in \Theta, \\
& x(0) = x_{\text{init}}
\end{aligned}
$$

- State sequence $X = [x(0)^\top, \dots, x(\bar{N}-1)^\top]^\top$
- Input sequence $U = [u(0)^\top, \dots, u(\bar{N}-1)^\top]^\top$
- Disturbance sequence $W = [w(0)^\top, \dots, w(\bar{N}-1)^\top]^\top$

But: we usually don't know f, $\mathcal{W}$, Θ, etc. exactly

Putting things together: The "true" optimal control problem

Optimal control problem that we ideally would like to solve – the stochastic case

$$
\begin{aligned}
\min_{\{\pi_k\}} \quad & \mathbb{E}\left(L(X, U)\right) \\
\text{s.t.} \quad & x(k+1) = f(x(k), u(k), w(k); \theta), \\
& u(k) = \pi_k(x(0), \ldots, x(k), u(0) \ldots, u(k-1)), \\
& W \sim \mathcal{Q}^W, \, \theta \sim \mathcal{Q}^\theta, \\
& Pr\left(x(k) \in \mathcal{X}, u(k) \in \mathcal{U}\right) \geq p, \\
& x(0) = x_{\text{init}}
\end{aligned}
$$

- State sequence $X = [x(0)^\mathsf{T}, \ldots, x(\bar{N})^\mathsf{T}]^\mathsf{T}$
- Input sequence $U = [u(0)^\mathsf{T}, \ldots, u(\bar{N})^\mathsf{T}]^\mathsf{T}$
- Disturbance sequence $W = [w(0)^\mathsf{T}, \ldots, w(\bar{N})^\mathsf{T}]^\mathsf{T}$

But: we usually don't know f, $\mathcal{Q}^W$, $\mathcal{Q}^\theta$, etc. exactly

Two common approximations

Optimal control problem that we ideally would like to solve – the stochastic case

$$
\begin{aligned}
\min_{\{\pi_k\}} \quad & \mathbb{E}\left(L(X, U)\right) \\
\text{s.t.} \quad & x(k+1) = f(x(k), u(k), w(k); \theta), \\
& u(k) = \pi_k(x(0), \ldots, x(k), u(0) \ldots, u(k-1)), \\
& W \sim \mathcal{Q}^W,\ \theta \sim \mathcal{Q}^\theta, \\
& Pr\left(x(k) \in \mathcal{X}, u(k) \in \mathcal{U}\right) \geq p, \\
& x(0) = x_{\text{init}}
\end{aligned}
$$

Model Predictive Control:

- Approximate objective, model, constraints
- Solve optimization problem numerically (in receding horizon)

Reinforcement learning:

- Evaluate objective empirically (typically in episodes)
- Iterative improvements (often model-free, no constraints)

Where Learning is used in MPC

$$
\begin{aligned}
\min_{\{\pi_k\}} \quad & \mathbb{E}\left(\sum_{k=0}^{\bar{N}} l_k(x(k), u(k))\right) \\
\text{s.t.} \quad & x(k+1) = f(x(k), u(k), w(k); \theta), \\
& u(k) = \pi_k(\cdot), \\
& w(k) \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x(k) \in \mathcal{X}) \geq p, \\
& \Pr(u(k) \in \mathcal{U}) \geq p, \\
& x(0) = x_{\text{init}}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
\min_{\{\pi_i\}} \quad & \mathbb{E}\left(l_f(x_N) + \sum_{i=0}^{N-1} l_i(x_i, u_i)\right) \\
\text{s.t.} \quad & x_{i+1} = f(x_i, u_i, w_i; \theta), \\
& u_i = \pi_i(\cdot), \\
& w_i \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x_i \in \mathcal{X}) \geq p, \\
& \Pr(u_i \in \mathcal{U}) \geq p, \\
& x_N \in \mathcal{X}_f, \\
& x_0 = x(k)
\end{aligned}
$$

- Solve over shortened horizon
- Restrict policy class (open-loop sequence, state feedback,...)

Learning-based MPC: Improve MPC approximation by learning problem components from data

Where Learning is used in MPC

$$\begin{aligned} \min_{\{\pi_k\}} \quad & \mathbb{E}\left(\sum_{k=0}^{\bar{N}} l_k(x(k), u(k))\right) \\ \text{s.t.} \quad & x(k+1) = f(x(k), u(k), w(k); \theta), \\ & u(k) = \pi_k(\cdot), \\ & w(k) \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\ & \Pr(x(k) \in \mathcal{X} \mid x(0)) \geq p, \\ & \Pr(u(k) \in \mathcal{U} \mid x(0)) \geq p, \\ & x(0) = x_{\text{init}} \end{aligned}$$

$$\Rightarrow$$

$$\begin{aligned} \min_{\{\pi_i\}} \quad & \mathbb{E}\left(l_f(x_N) + \sum_{i=0}^{N-1} l_i(x_i, u_i)\right) \\ \text{s.t.} \quad & x_{i+1} = f(x_i, u_i, w_i; \theta), \\ & u_i = \pi_i(\cdot), \\ & w_i \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\ & \Pr(x_i \in \mathcal{X} \mid x_0) \geq p, \\ & \Pr(u_i \in \mathcal{U} \mid x_0) \geq p, \\ & x_N \in \mathcal{X}_f, \\ & x_0 = x(k) \end{aligned}$$

- **Learning the system dynamics**
  Improve model using (online) measurements

$\rightarrow$ Stochastic or robust models

$\rightarrow$ Parametric or non-parametric regression

Where Learning is used in MPC

$$\begin{aligned}
\min_{\{\pi_k\}} \quad & \mathbb{E}\left(\sum_{k=0}^{\bar{N}} l_k(x(k), u(k))\right) \\
\text{s.t.} \quad & x(k+1) = f(x(k), u(k), w(k); \theta), \\
& u(k) = \pi_k(\cdot), \\
& w(k) \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x(k) \in \mathcal{X}) \geq p, \\
& \Pr(u(k) \in \mathcal{U}) \geq p, \\
& x(0) = x_{\text{init}}
\end{aligned}$$

$$\Rightarrow$$

$$\begin{aligned}
\min_{\{\pi_i\}} \quad & \mathbb{E}\left(l_f(x_N; \theta_l) + \sum_{i=0}^{N-1} l_i(x_i, u_i; \theta_l)\right) \\
\text{s.t.} \quad & x_{i+1} = f(x_i, u_i, w_i; \theta), \\
& u_i = \pi_i(\cdot), \\
& w_i \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x_i \in \mathcal{X}) \geq p, \\
& \Pr(u_i \in \mathcal{U}) \geq p, \\
& x_N \in \mathcal{X}_f, \\
& x_0 = x(k)
\end{aligned}$$

- **Learning the controller design**
  Optimize for closed-loop performance (usually episodic)

$\rightarrow$ Performance-driven learning: Bayesian / Convex optimization, Terminal components

$\rightarrow$ Inverse optimal control

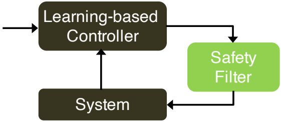
Where Learning is used in MPC

$$
\begin{aligned}
\min_{\{\pi_k\}} \quad & \mathbb{E}\left(\sum_{k=0}^{\bar{N}} l_k(x(k), u(k))\right) \\
\text{s.t.} \quad & x(k+1) = f(x(k), u(k), w(k); \theta), \\
& u(k) = \pi_k(\cdot), \\
& w(k) \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x(k) \in \mathcal{X}) \geq p, \\
& \Pr(u(k) \in \mathcal{U}) \geq p, \\
& x(0) = x_{\text{init}}
\end{aligned}
$$

$$\Rightarrow$$

$$
\begin{aligned}
\min_{\{\pi_i\}} \quad & \cancel{\mathbb{E}\left(l_f(x_N) + \sum_{i=0}^{N-1} l_i(x_i, u_i)\right)} \\
\text{s.t.} \quad & x_{i+1} = f(x_i, u_i, w_i; \theta), \\
& u_i = \pi_i(\cdot), \\
& w_i \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x_i \in \mathcal{X}) \geq p, \\
& \Pr(u_i \in \mathcal{U}) \geq p, \\
& x_N \in \mathcal{X}_f, \\
& x_0 = x(k)
\end{aligned}
$$

- **MPC for safe learning**
  MPC only for constraint satisfaction



Where Learning is used in MPC

$$
\begin{aligned}
\min_{\{\pi_k\}} \quad & \mathbb{E}\left(\sum_{k=0}^{\bar{N}} l_k(x(k), u(k))\right) \\
\text{s.t.} \quad & x(k+1) = f(x(k), u(k), w(k); \theta), \\
& u(k) = \pi_k(\cdot), \\
& w(k) \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x(k) \in \mathcal{X}) \geq p, \\
& \Pr(u(k) \in \mathcal{U}) \geq p, \\
& x(0) = x_{\text{init}}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
\color{red}{\tilde{\pi}(x(k)) \approx \arg} \min_{\{\pi_i\}} \quad & \mathbb{E}\left(l_f(x_N) + \sum_{i=0}^{N-1} l_i(x_i, u_i)\right) \\
\text{s.t.} \quad & x_{i+1} = f(x_i, u_i, w_i; \theta), \\
& u_i = \pi_i(\cdot), \\
& w_i \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x_i \in \mathcal{X}) \geq p, \\
& \Pr(u_i \in \mathcal{U}) \geq p, \\
& x_N \in \mathcal{X}_f, \\
& x_0 = x(k)
\end{aligned}
$$

- **Approximating the MPC control law**
  Learn control law offline
  (speed up evaluation)

Where Learning is used in MPC

$$
\begin{aligned}
\min_{\{\pi_k\}} \quad & \mathbb{E}\left(\sum_{k=0}^{\bar{N}} l_k(x(k),u(k))\right) \\
\text{s.t.} \quad & x(k+1) = f(x(k),u(k),w(k);\theta), \\
& u(k) = \pi_k(\cdot), \\
& w(k) \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x(k) \in \mathcal{X}) \geq p, \\
& \Pr(u(k) \in \mathcal{U}) \geq p, \\
& x(0) = x_{\text{init}}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
\min_{\{\pi_i\}} \quad & \mathbb{E}\left(l_f(x_N) + \sum_{i=0}^{N-1} l_i(x_i,u_i)\right) \\
\text{s.t.} \quad & x_{i+1} = f(x_i,u_i,w_i;\theta), \\
& u_i = \pi_i(\cdot), \\
& w_i \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x_i \in \mathcal{X}) \geq p, \\
& \Pr(u_i \in \mathcal{U}) \geq p, \\
& x_N \in \mathcal{X}_f, \\
& x_0 = x(k)
\end{aligned}
$$

- Learning the system dynamics
  Improve model using (online) measurements
- Learning the controller design
  Optimize for closed-loop performance
- MPC for safe learning
  MPC only for constraint satisfaction
- Approximating the MPC control law
  Learn control law offline (speed up evaluation)

Where Learning is used in MPC (This course)

$$
\begin{aligned}
\min_{\{\pi_k\}} \quad & \mathbb{E}\left(\sum_{k=0}^{\bar{N}} l_k(x(k), u(k))\right) \\
\text{s.t.} \quad & x(k+1) = f(x(k), u(k), w(k); \theta), \\
& u(k) = \pi_k(\cdot), \\
& w(k) \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x(k) \in \mathcal{X}) \geq p, \\
& \Pr(u(k) \in \mathcal{U}) \geq p, \\
& x(0) = x_{\text{init}}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
\min_{\{\pi_i\}} \quad & \mathbb{E}\left(l_f(x_N) + \sum_{i=0}^{N-1} l_i(x_i, u_i)\right) \\
\text{s.t.} \quad & x_{i+1} = f(x_i, u_i, w_i; \theta), \\
& u_i = \pi_i(\cdot), \\
& w_i \sim \mathcal{Q}^w,\ \theta \sim \mathcal{Q}^\theta, \\
& \Pr(x_i \in \mathcal{X}) \geq p, \\
& \Pr(u_i \in \mathcal{U}) \geq p, \\
& x_N \in \mathcal{X}_f, \\
& x_0 = x(k)
\end{aligned}
$$

- **Learning the system dynamics**
  State measurements for (online) identifcation
- Learning the controller design
  Optimize for closed-loop performance
- **MPC for safe learning**
  MPC only for constraint satisfaction
- Approximating the MPC control law
  Learn control law offline (speed up evaluation)

References and further reading

[1] Lukas Hewing, Kim P. Wabersich, Marcel Menner, and Melanie N. Zeilinger. 2020 Learning-Based Model Predictive Control: Toward Safe Learning in Control. Annual Review of Control, Robotics, and Autonomous Systems.

[2] A. Mesbah et al., "Fusion of Machine Learning and MPC under Uncertainty: What Advances Are on the Horizon?," 2022 American Control Conference (ACC).