

What we have seen:

- RL MPC, why does it work & different flavors
- Fast overview of existing results, knowledge, tricks and flavors

What we have seen:

- RLMPC, why does it work & different flavors
- Fast overview of existing results, knowledge, tricks and flavors

What we will do now:

- Some practical “for-grab” open questions (high-level)
- Two hard open questions, for RL-MPC and MPC in general (more detailed)

RLMPC: Open Questions & Future Directions

Prof. Sebastien Gros

Department of Cybernetic
Faculty of Information Technology
NTNU

Freiburg PhD School

Outline

- 1 Some Practical Open Questions & Future Directions
- 2 Fundamental Open Questions (in RL MPC & general decision making)

Offline Learning & Verification Pathways



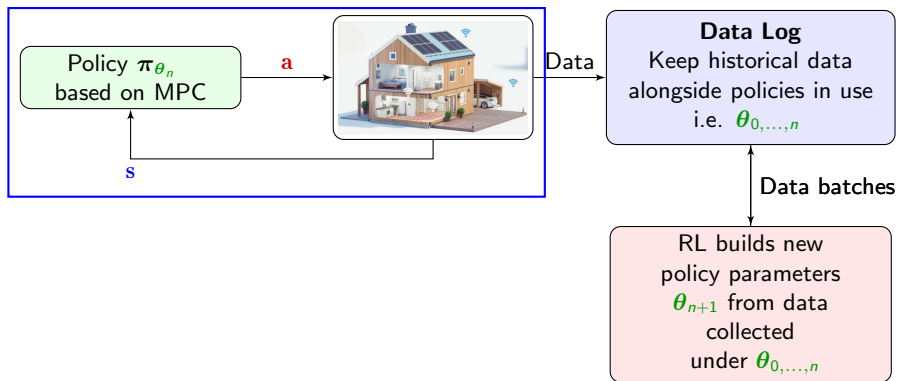
Offline Learning & Verification Pathways



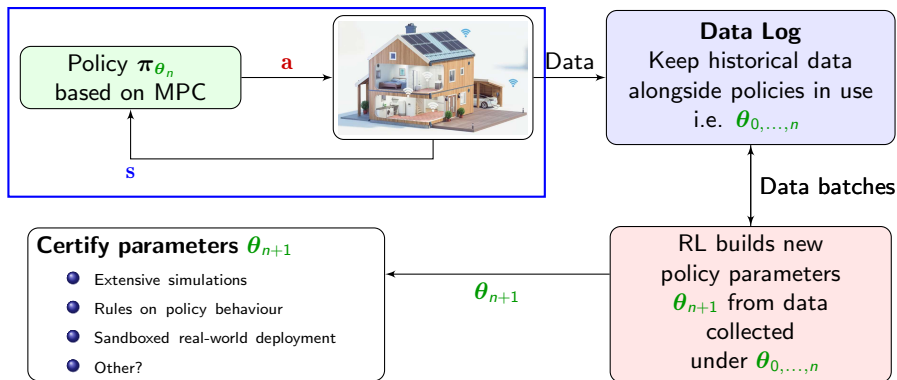
Offline Learning & Verification Pathways



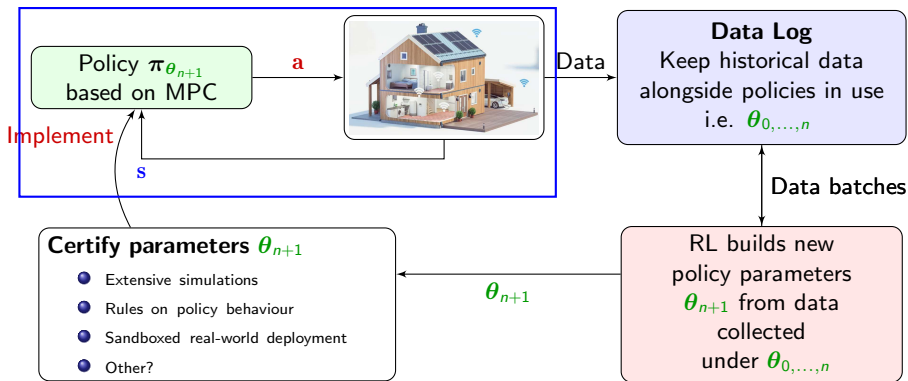
Offline Learning & Verification Pathways



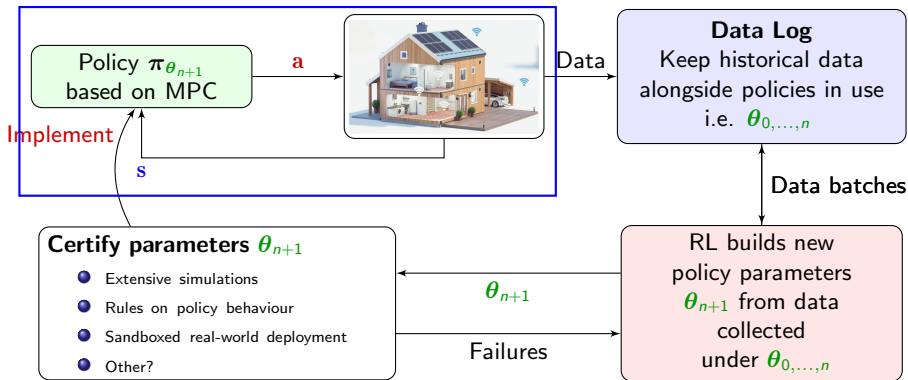
Offline Learning & Verification Pathways



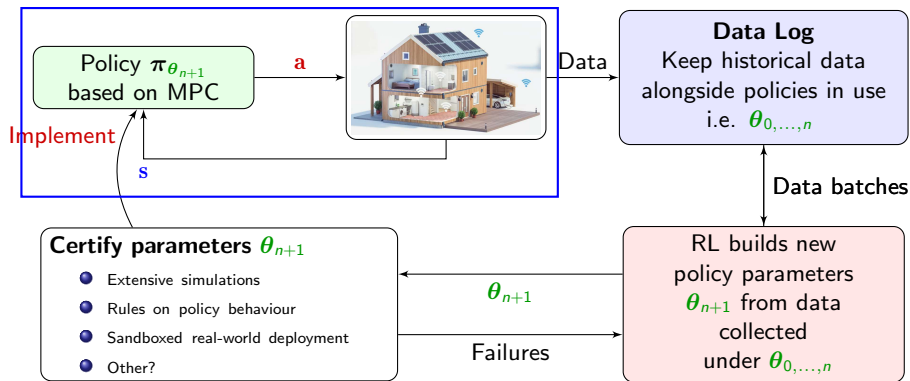
Offline Learning & Verification Pathways



Offline Learning & Verification Pathways



Offline Learning & Verification Pathways



Research Questions:

- efficient off-policy RL over MPC
- efficient embedding & correction of verification failures into the policy design
- when does the data log stop being informative (formal analysis + viable methods)?
- proof of convergence & safety

Optimal Exploration

Stochastic policy $\pi_{\theta} [\cdot | s] \equiv \mathbf{u}_0^* (\theta, s, \mathbf{d})$:

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

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

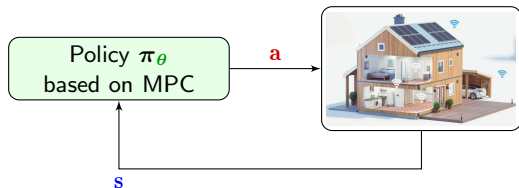
$$\mathbf{h}_{\theta} (\mathbf{x}_k, \mathbf{u}_k, \mathbf{d}) \leq 0, \quad \mathbf{x}_0 = \mathbf{s}$$

where $\mathbf{d} \sim \rho[\cdot]$

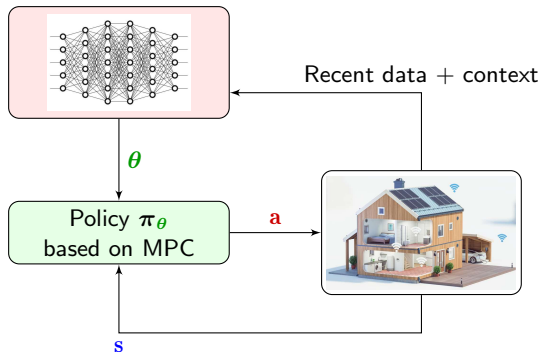
Research questions

- Can we optimize the exploration? Minimize regret?
- Adopt $\mathbf{d} \sim \rho[\cdot | s]$, dependent on our current knowledge of θ ?
- Leverage on the knowledge embedded in MPC?
- Build epistemic uncertainty on θ ?
- Should the exploration even be stochastic?

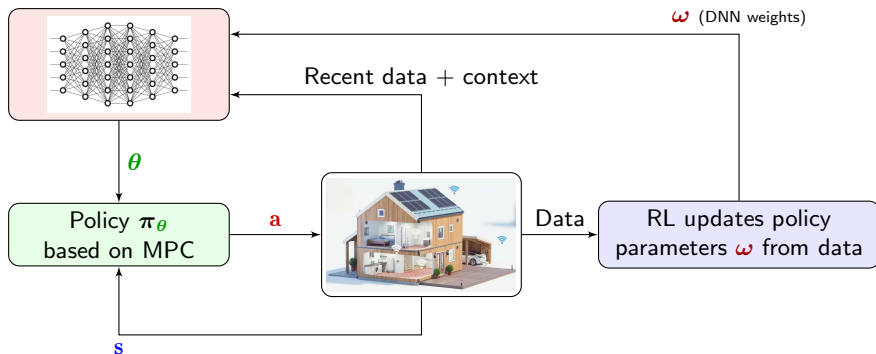
Contextual Learning



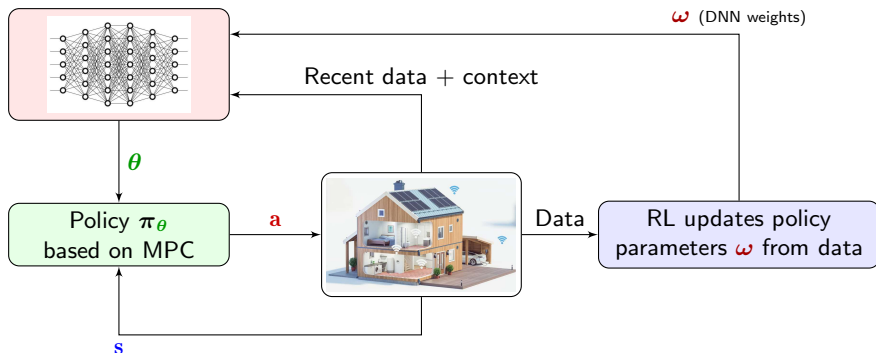
Contextual Learning



Contextual Learning



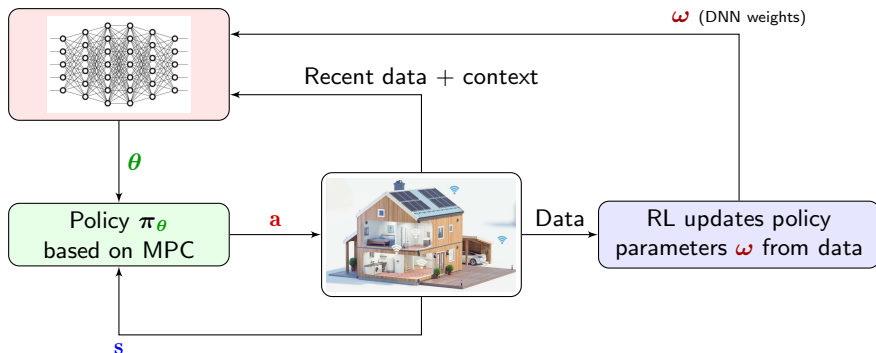
Contextual Learning



Remarks

- Architecture proposed from applications
- Good empirical results documented for e.g. racing, energy in buildings(?)
- Weak theoretical foundation and methodology, no guarantees

Contextual Learning



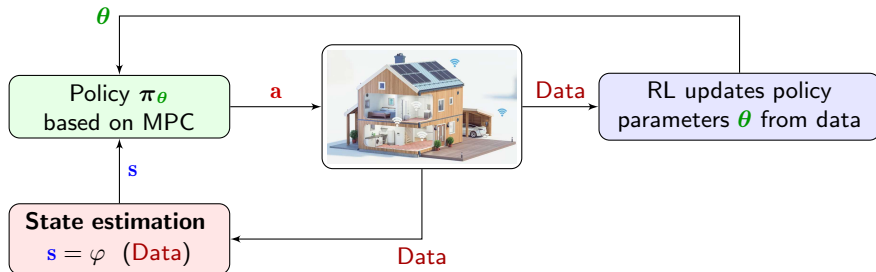
Research Questions

- How to harmonize the interactions DNN-Optimization, no DNN “interference”?
- How to retain explainability and guarantees provided by optimization?
- Can we establish formal proofs of the overall convergence and behaviour?

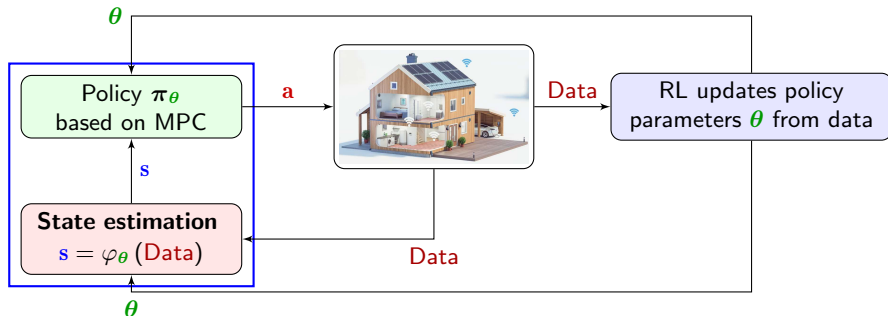
Latent state representation



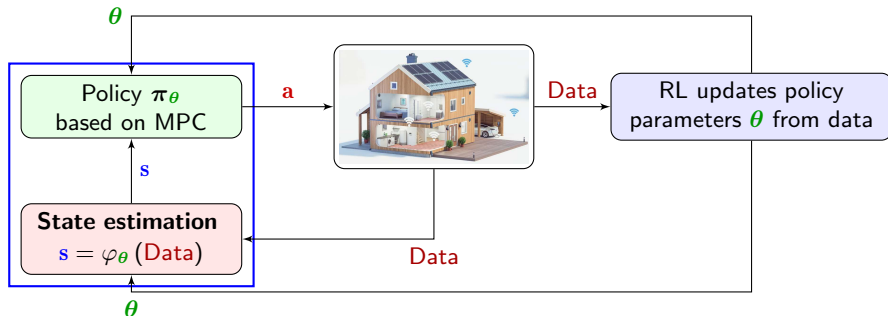
Latent state representation



Latent state representation



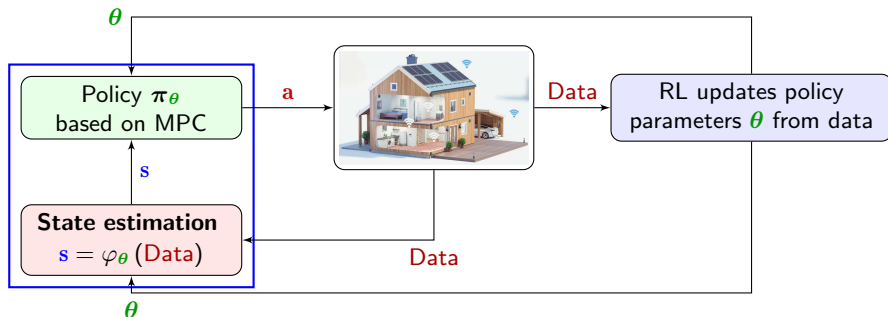
Latent state representation



Remarks

- SOTA (2023) if φ_θ is a “classical model-based” state observer (Kalman, MHE and co.)
- Extension to abstract, data-driven, latent state representation would be of high value for problems lacking simple physics-based models. State estimation becomes rather a “Learned data compression into a useful object for decision making”

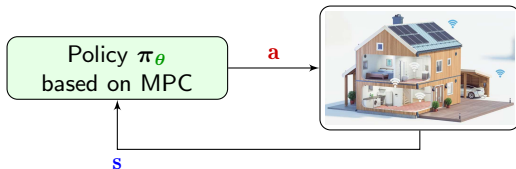
Latent state representation



Research Questions

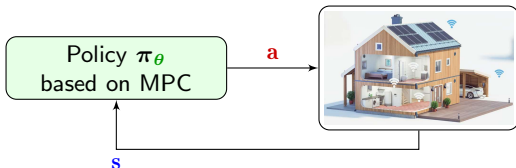
- Align latent space representation with an ontological meaning?
- Build latent space conducive of retaining explainability, safety, stability?
- Build optimization-friendly prediction models in the latent space representation?
- Embed knowledge in that context? Hybridize with system ontology?
- Harmonious combination of (semi)-supervised & reinforcement learning?
- Prove convergence (to something meaningful)?

Epistemic Uncertainty & Belief states



*We understand very well how to deal
with the “know unknown”
How to deal with the “unknown
unknown”?*

Epistemic Uncertainty & Belief states

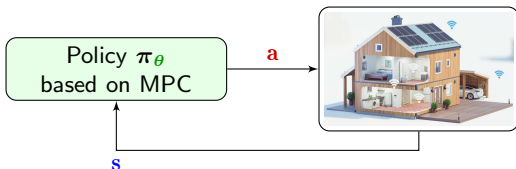


*We understand very well how to deal
with the “know unknown”
How to deal with the “unknown
unknown”?*

Remarks

- Policy carries a belief state, i.e. imperfect knowledge of the world and future behaviour

Epistemic Uncertainty & Belief states

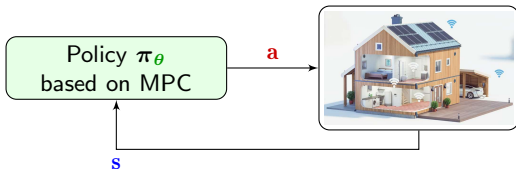


*We understand very well how to deal
with the “know unknown”
How to deal with the “unknown
unknown”?*

Remarks

- Policy carries a belief state, i.e. imperfect knowledge of the world and future behaviour
- Our belief is corrected all the time based on real-world observations

Epistemic Uncertainty & Belief states

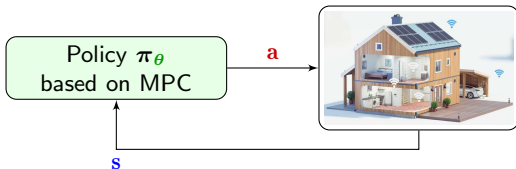


*We understand very well how to deal
with the “know unknown”
How to deal with the “unknown
unknown”?*

Remarks

- Policy carries a belief state, i.e. imperfect knowledge of the world and future behaviour
- Our belief is corrected all the time based on real-world observations
- The performance we can achieve depends on the “sharpness” of our belief

Epistemic Uncertainty & Belief states

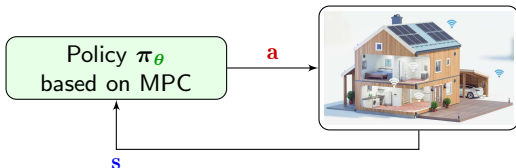


*We understand very well how to deal
with the “know unknown”
How to deal with the “unknown
unknown”?*

Remarks

- Policy carries a belief state, i.e. imperfect knowledge of the world and future behaviour
- Our belief is corrected all the time based on real-world observations
- The performance we can achieve depends on the “sharpness” of our belief
- Actions influence real world performance, and the evolution of our belief

Epistemic Uncertainty & Belief states

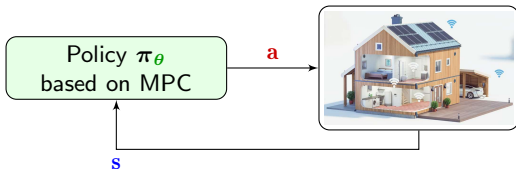


*We understand very well how to deal
with the “know unknown”
How to deal with the “unknown
unknown”?*

Remarks

- Policy carries a belief state, i.e. imperfect knowledge of the world and future behaviour
- Our belief is corrected all the time based on real-world observations
- The performance we can achieve depends on the “sharpness” of our belief
- Actions influence real world performance, and the evolution of our belief
- E.g. lose performance now to gain sharper belief and higher performance later

Epistemic Uncertainty & Belief states

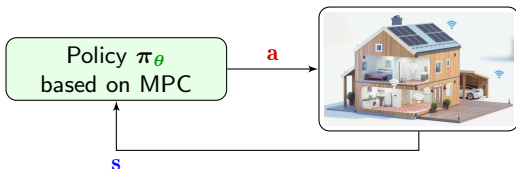


*We understand very well how to deal
with the “know unknown”
How to deal with the “unknown
unknown”?*

Remarks

- Policy carries a belief state, i.e. imperfect knowledge of the world and future behaviour
- Our belief is corrected all the time based on real-world observations
- The performance we can achieve depends on the “sharpness” of our belief
- Actions influence real world performance, and the evolution of our belief
- E.g. lose performance now to gain sharper belief and higher performance later
- In (one-shot) decision making, the need to distinguish between aleatoric & epistemic is debated

Epistemic Uncertainty & Belief states

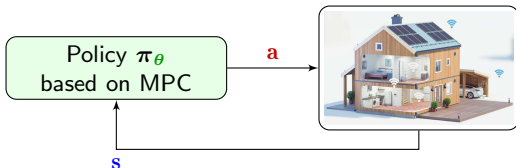


*We understand very well how to deal
with the “know unknown”
How to deal with the “unknown
unknown”?*

Remarks

- Policy carries a belief state, i.e. imperfect knowledge of the world and future behaviour
- Our belief is corrected all the time based on real-world observations
- The performance we can achieve depends on the “sharpness” of our belief
- Actions influence real world performance, and the evolution of our belief
- E.g. lose performance now to gain sharper belief and higher performance later
- In (one-shot) decision making, the need to distinguish between aleatoric & epistemic is debated
- In operational decision making, the difference is undeniable as they enter differently in the equations

Epistemic Uncertainty & Belief states



*We understand very well how to deal
with the “know unknown”
How to deal with the “unknown
unknown”?*

Research Questions

- Representation in MPC seems possible, maths are sketched, now we need to understand their implications
- Existing heuristics & approximations in SOTA (e.g. OOD management) but how do they relate to / approximate the core maths? Can they be improved accordingly?
- Efficient representations & approximations in MPC? Optimization algorithms?
- Efficient learning algorithms?

Outline

- 1 Some Practical Open Questions & Future Directions
- 2 Fundamental Open Questions (in RL MPC & general decision making)

Structure Underlying RLMPC

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

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

MDP policy $\pi^*(\mathbf{s})$ from

$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k \ell^{\text{MDP}}(\mathbf{s}_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi^*(\mathbf{s}_k)$

Structure Underlying RLMPC

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

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

MDP policy $\pi^*(\mathbf{s})$ from

$$\min_{\pi} \quad \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k \ell^{\text{MDP}}(\mathbf{s}_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi^*(\mathbf{s}_k)$

Relation MPC-MDP

$$\ell^{\text{MPC}}(\mathbf{s}, \mathbf{a}) = \ell^{\text{MDP}}(\mathbf{s}, \mathbf{a}) + \mathbb{E}[V^*(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}] - V^*(\mathbf{f}(\mathbf{s}, \mathbf{a}))$$

Structure Underlying RLMP

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

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

MDP policy $\pi^*(s)$ from

$$\min_{\pi} \quad \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k \ell^{\text{MDP}}(\mathbf{s}_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi^*(\mathbf{s}_k)$

Relation MPC-MDP

$$\ell^{\text{MPC}}(\mathbf{s}, \mathbf{a}) = \ell^{\text{MDP}}(\mathbf{s}, \mathbf{a}) + \mathbb{E}[V^*(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}] - V^*(\mathbf{f}(\mathbf{s}, \mathbf{a}))$$

- No modification if model is perfect
- MPC cost correction is intimately tied to the optimal MDP value function V^*
- Principle is tightly connected to the Bellman equations
- MPC terminal cost T is also tied to V^*
- In practice, RL sorts out the modification without really computing V^*

Structure Underlying RLMP

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

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

MDP policy $\pi^*(s)$ from

$$\min_{\pi} \quad \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k \ell^{\text{MDP}}(\mathbf{s}_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi^*(\mathbf{s}_k)$

Relation MPC-MDP

$$\ell^{\text{MPC}}(\mathbf{s}, \mathbf{a}) = \ell^{\text{MDP}}(\mathbf{s}, \mathbf{a}) + \mathbb{E}[V^*(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}] - V^*(\mathbf{f}(\mathbf{s}, \mathbf{a}))$$

- No modification if model is perfect
- MPC cost correction is intimately tied to the optimal MDP value function V^*
- Principle is tightly connected to the Bellman equations
- MPC terminal cost T is also tied to V^*
- In practice, RL sorts out the modification without really computing V^*

Fundamental open questions: what can we do when these features become a challenge?

Structure Underlying RLMP

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

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

MDP policy $\pi^*(\mathbf{s})$ from

$$\min_{\pi} \quad \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k \ell^{\text{MDP}}(\mathbf{s}_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi^*(\mathbf{s}_k)$

Relation MPC-MDP

$$\ell^{\text{MPC}}(\mathbf{s}, \mathbf{a}) = \ell^{\text{MDP}}(\mathbf{s}, \mathbf{a}) + \mathbb{E}[V^*(\mathbf{s}_+) \mid \mathbf{s}, \mathbf{a}] - V^*(\mathbf{f}(\mathbf{s}, \mathbf{a}))$$

- No modification if model is perfect
- MPC cost correction is intimately tied to the optimal MDP value function V^*
- Principle is tightly connected to the Bellman equations
- MPC terminal cost T is also tied to V^*
- In practice, RL sorts out the modification without really computing V^*

Fundamental open questions: what can we do when these features become a challenge? *Note: this happens faster than we tend to think*

Multi-agent MDPs

Example (simple yet rich & important)

$$\min_{\pi} \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right]$$

Multi-agent MDPs

Example (simple yet rich & important)

$$\min_{\pi} \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right]$$

with $\mathbf{s}_+^i \sim \rho^i(\cdot | \mathbf{s}^i, \mathbf{a}^i)$

Multi-agent MDPs

Example (simple yet rich & important)

$$\min_{\pi} \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right]$$

with $\mathbf{s}_{+}^i \sim \rho^i \left(\cdot \mid \mathbf{s}^i, \mathbf{a}^i \right)$ and

$$\sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k$$

Multi-agent MDPs

Example (simple yet rich & important)

$$\min_{\pi} \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right]$$

with $\mathbf{s}_{+}^i \sim \rho^i \left(\cdot \mid \mathbf{s}^i, \mathbf{a}^i \right)$ and

$$\sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k$$

Optimal policy

$$\underbrace{\begin{bmatrix} \mathbf{a}_k^1 \\ \vdots \\ \mathbf{a}_k^m \end{bmatrix}}_{:= \mathbf{a}_k} = \pi^* \left(\underbrace{\begin{bmatrix} \mathbf{s}_k^1 \\ \vdots \\ \mathbf{s}_k^m \end{bmatrix}}_{:= \pi^*(\mathbf{s}_k)} \right)$$

Multi-agent MDPs

Example (simple yet rich & important)

$$\min_{\pi} \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right]$$

with $\mathbf{s}_{+}^i \sim \rho^i(\cdot | \mathbf{s}^i, \mathbf{a}^i)$ and

$$\sum_{i=1}^m \mathbf{a}_k^i = 0, \quad \forall k$$

Optimal policy

$$\underbrace{\begin{bmatrix} \mathbf{a}_k^1 \\ \vdots \\ \mathbf{a}_k^m \end{bmatrix}}_{:= \mathbf{a}_k} = \underbrace{\pi^* \left(\begin{bmatrix} \mathbf{s}_k^1 \\ \vdots \\ \mathbf{s}_k^m \end{bmatrix} \right)}_{:= \pi^*(\mathbf{s}_k)}$$

This is our energy system btw.

Multi-agent MDPs

Example (simple yet rich & important)

$$\min_{\pi} \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right]$$

with $\mathbf{s}_+^i \sim \rho^i(\cdot | \mathbf{s}^i, \mathbf{a}^i)$ and

$$\sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k$$

Optimal policy

$$\underbrace{\begin{bmatrix} \mathbf{a}_k^1 \\ \vdots \\ \mathbf{a}_k^m \end{bmatrix}}_{:= \mathbf{a}_k} = \pi^* \left(\underbrace{\begin{bmatrix} \mathbf{s}_k^1 \\ \vdots \\ \mathbf{s}_k^m \end{bmatrix}}_{:= \pi^*(\mathbf{s}_k)} \right)$$

This is our energy system btw.

Optimal value functions

$$\begin{aligned} Q^*(\mathbf{s}_k, \mathbf{a}_k) &= \sum_{i=1}^m \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \\ &\quad + \gamma \mathbb{E} [V^*(\mathbf{s}_+) | \mathbf{s}_k, \mathbf{a}_k] \\ V^*(\mathbf{s}) &= \min_{\mathbf{a}_k \in \mathbb{S}} Q^*(\mathbf{s}_k, \mathbf{a}_k) \end{aligned}$$

where $\mathbb{S} := \{ \mathbf{a}_k \mid \text{s.t.} \quad \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0} \}$

Multi-agent MDPs

Example (simple yet rich & important)

$$\min_{\pi} \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right]$$

with $\mathbf{s}_+^i \sim \rho^i(\cdot | \mathbf{s}^i, \mathbf{a}^i)$ and

$$\sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k$$

Optimal policy

$$\underbrace{\begin{bmatrix} \mathbf{a}_k^1 \\ \vdots \\ \mathbf{a}_k^m \end{bmatrix}}_{:= \mathbf{a}_k} = \pi^* \left(\underbrace{\begin{bmatrix} \mathbf{s}_k^1 \\ \vdots \\ \mathbf{s}_k^m \end{bmatrix}}_{:= \pi^*(\mathbf{s}_k)} \right)$$

This is our energy system btw.

Optimal value functions

$$\begin{aligned} Q^*(\mathbf{s}_k, \mathbf{a}_k) &= \sum_{i=1}^m \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \\ &\quad + \gamma \mathbb{E} [V^*(\mathbf{s}_+) | \mathbf{s}_k, \mathbf{a}_k] \\ V^*(\mathbf{s}) &= \min_{\mathbf{a}_k \in \mathbb{S}} Q^*(\mathbf{s}_k, \mathbf{a}_k) \end{aligned}$$

where $\mathbb{S} := \{ \mathbf{a}_k \mid \text{s.t.} \quad \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0} \}$

Q^*, V^* are not separable because of $\mathbb{S}$ i.e.

$$\begin{aligned} V^*(\mathbf{s}) &\neq \sum_{i=1}^m V_i^*(\mathbf{s}^i) \\ Q^*(\mathbf{s}, \mathbf{a}) &\neq \sum_{i=1}^m Q_i^*(\mathbf{s}^i, \mathbf{a}^i) \end{aligned}$$

hence they *entangle* the states and actions

Multi-agent MDPs

Example (simple yet rich & important)

$$\min_{\pi} \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right]$$

with $\mathbf{s}_+^i \sim \rho^i(\cdot | \mathbf{s}^i, \mathbf{a}^i)$ and

$$\sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k$$

Optimal policy

$$\underbrace{\begin{bmatrix} \mathbf{a}_k^1 \\ \vdots \\ \mathbf{a}_k^m \end{bmatrix}}_{:= \mathbf{a}_k} = \pi^* \left(\underbrace{\begin{bmatrix} \mathbf{s}_k^1 \\ \vdots \\ \mathbf{s}_k^m \end{bmatrix}}_{:= \pi^*(\mathbf{s}_k)} \right)$$

This is our energy system btw.

Optimal value functions

$$\begin{aligned} Q^*(\mathbf{s}_k, \mathbf{a}_k) &= \sum_{i=1}^m \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \\ &\quad + \gamma \mathbb{E} [V^*(\mathbf{s}_+) | \mathbf{s}_k, \mathbf{a}_k] \\ V^*(\mathbf{s}) &= \min_{\mathbf{a}_k \in \mathbb{S}} Q^*(\mathbf{s}_k, \mathbf{a}_k) \end{aligned}$$

where $\mathbb{S} := \{ \mathbf{a}_k \mid \text{s.t.} \quad \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0} \}$

Q^*, V^* are not separable because of $\mathbb{S}$ i.e.

$$\begin{aligned} V^*(\mathbf{s}) &\neq \sum_{i=1}^m V_i^*(\mathbf{s}^i) \\ Q^*(\mathbf{s}, \mathbf{a}) &\neq \sum_{i=1}^m Q_i^*(\mathbf{s}^i, \mathbf{a}^i) \end{aligned}$$

hence they *entangle* the states and actions

Why does that matter?

Multi-agent MPC

Multi-agent MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right] \\ \text{s.t.} \quad & \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k \end{aligned}$$

Multi-agent MPC

Multi-agent MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right] \\ \text{s.t.} \quad & \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k \end{aligned}$$

Multi-agent MPC

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

Multi-agent MPC

Multi-agent MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right] \\ \text{s.t.} \quad & \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k \end{aligned}$$

Multi-agent MPC

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

Remarks

- There are strong & fast methods to solve this multi-agent MPC in a semi or fully distributed fashion with very limited information exchange

Multi-agent MPC

Multi-agent MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right] \\ \text{s.t.} \quad & \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k \end{aligned}$$

Multi-agent MPC

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

Remarks

- There are strong & fast methods to solve this multi-agent MPC in a semi or fully distributed fashion with very limited information exchange
- Apply RL over multi-agent MPC?

Multi-agent MPC

Multi-agent MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right] \\ \text{s.t.} \quad & \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k \end{aligned}$$

Multi-agent MPC

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

Remarks

- There are strong & fast methods to solve this multi-agent MPC in a semi or fully distributed fashion with very limited information exchange
- Apply RL over multi-agent MPC?

Multi-agent MPC

Multi-agent MPC

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

Multi-agent MPC

Multi-agent MPC

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

Distributed computation coupling constraints dualization

$$\begin{aligned} \max_{\lambda} \quad & d(\lambda) \quad \text{with} \\ d(\lambda) = \min_{\mathbf{x}, \mathbf{u}} \quad & \sum_{i=1}^m \left(T_{\theta}^i(\mathbf{x}_N^i) + \sum_{k=0}^{N-1} \gamma^k \ell_{\theta}^i(\mathbf{x}_k^i, \mathbf{u}_k^i) \right) + \sum_{k=0}^{N-1} \lambda_k^{\top} \left(\sum_{i=1}^m \mathbf{u}_k^i \right) \\ \text{s.t.} \quad & \mathbf{x}_{k+1}^i = \mathbf{f}_{\theta}^i(\mathbf{x}_k^i, \mathbf{u}_k^i), \quad \mathbf{x}_0^i = \mathbf{s}^i \end{aligned}$$

λ is the system-wide value of the shared resource at each time

Multi-agent MPC

Multi-agent MPC

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

Distributed computation distributed problem

$$\begin{aligned} \max_{\lambda} \quad & \sum_{i=1}^m d_i(\lambda) \quad \text{with} \\ d_i(\lambda) = \min_{\mathbf{x}^i, \mathbf{u}^i} \quad & T_{\theta}^i(\mathbf{x}_N^i) + \sum_{k=0}^{N-1} \gamma^k \ell_{\theta}^i(\mathbf{x}_k^i, \mathbf{u}_k^i) + \lambda_k^{\top} \mathbf{u}_k^i \\ \text{s.t.} \quad & \mathbf{x}_{k+1}^i = \mathbf{f}_{\theta}^i(\mathbf{x}_k^i, \mathbf{u}_k^i), \quad \mathbf{x}_0^i = \mathbf{s}^i \end{aligned}$$

Each agent can compute its own strategy for a given λ , the optimal λ can be computed in different ways, by maximizing $\sum_{i=1}^m d_i(\lambda)$

Multi-agent MPC

Multi-agent MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right] \\ \text{s.t.} \quad & \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k \end{aligned}$$

Remarks

- There are strong & fast methods to solve this multi-agent MPC in a semi or fully distributed fashion with very limited information exchange
- Apply RL over multi-agent MPC?

Multi-agent MPC

Multi-agent MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right] \\ \text{s.t.} \quad & \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k \end{aligned}$$

Multi-agent MPC

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

Remarks

- There are strong & fast methods to solve this multi-agent MPC in a semi or fully distributed fashion with very limited information exchange
- Apply RL over multi-agent MPC?
- But theory stating that $Q^{\text{MPC}} = Q^{\star}$ is possible does not apply. **Why?!?!**

Multi-agent MPC

Multi-agent MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right] \\ \text{s.t.} \quad & \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k \end{aligned}$$

Multi-agent MPC

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

Remarks

- There are strong & fast methods to solve this multi-agent MPC in a semi or fully distributed fashion with very limited information exchange
- Apply RL over multi-agent MPC?
- But theory stating that $Q^{\text{MPC}} = Q^{\star}$ is possible does not apply. **Why?!?!**

Theory requires

$$\begin{aligned} \sum_{i=1}^m \ell_{\theta}^i \left(\mathbf{s}^i, \mathbf{a}^i \right) &= \sum_{i=1}^m \ell^i \left(\mathbf{s}^i, \mathbf{a}^i \right) \\ &+ \mathbb{E} \left[V^{\star} \left(\mathbf{s}_{+} \right) \mid \mathbf{s}, \mathbf{a} \right] - V^{\star} \left(\mathbf{f} \left(\mathbf{s}, \mathbf{a} \right) \right) \end{aligned}$$

i.e. the **cost modification** entangles the agents in a non-separable way (incl. the terminal cost)

Multi-agent MPC

Multi-agent MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{i=1}^m \sum_{k=0}^{\infty} \gamma^k \ell^i \left(\mathbf{s}_k^i, \mathbf{a}_k^i \right) \right] \\ \text{s.t.} \quad & \sum_{i=1}^m \mathbf{a}_k^i = \mathbf{0}, \quad \forall k \end{aligned}$$

Multi-agent MPC

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

Remarks

- There are strong & fast methods to solve this multi-agent MPC in a semi or fully distributed fashion with very limited information exchange
- Apply RL over multi-agent MPC?
- But theory stating that $Q^{\text{MPC}} = Q^*$ is possible does not apply. **Why?!?!**

Theory requires

$$\begin{aligned} \sum_{i=1}^m \ell_{\theta}^i \left(\mathbf{s}^i, \mathbf{a}^i \right) &= \sum_{i=1}^m \ell^i \left(\mathbf{s}^i, \mathbf{a}^i \right) \\ &+ \mathbb{E} \left[V^* \left(\mathbf{s}_+ \right) \mid \mathbf{s}, \mathbf{a} \right] - V^* \left(\mathbf{f} \left(\mathbf{s}, \mathbf{a} \right) \right) \end{aligned}$$

i.e. the **cost modification** entangles the agents in a non-separable way (incl. the terminal cost)

Open question: this non-separable modification “ruins” efficient distributed MPC algorithms. **How to fix that?**

Problems Beyond Expectation

MDP

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

where $\mathbf{a}_k = \pi(\mathbf{s}_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(\mathbf{s}_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(\mathbf{s}_k)$

Problems Beyond Expectation

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

A lot can be done with $\mathbb{E}[\cdot]$. E.g.

- Robust problems: $\ell(s_k, \mathbf{a}_k) = \dots + \begin{cases} \infty & \text{if } \mathbf{h}(s_k, \mathbf{a}_k) > 0 \\ 0 & \text{if } \mathbf{h}(s_k, \mathbf{a}_k) \leq 0 \end{cases}$

Problems Beyond Expectation

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

A lot can be done with $\mathbb{E}[\cdot]$. E.g.

- Robust problems: $\ell(s_k, \mathbf{a}_k) = \dots + \begin{cases} \infty & \text{if } \mathbf{h}(s_k, \mathbf{a}_k) > 0 \\ 0 & \text{if } \mathbf{h}(s_k, \mathbf{a}_k) \leq 0 \end{cases}$
- Relaxed constraints: $\ell(s_k, \mathbf{a}_k) = \dots + \omega \cdot \max(0, \mathbf{h}(s_k, \mathbf{a}_k))$ (*Lagrange RL*)

Problems Beyond Expectation

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

A lot can be done with $\mathbb{E}[\cdot]$. E.g.

- Robust problems: $\ell(s_k, \mathbf{a}_k) = \dots + \begin{cases} \infty & \text{if } \mathbf{h}(s_k, \mathbf{a}_k) > 0 \\ 0 & \text{if } \mathbf{h}(s_k, \mathbf{a}_k) \leq 0 \end{cases}$
- Relaxed constraints: $\ell(s_k, \mathbf{a}_k) = \dots + \omega \cdot \max(0, \mathbf{h}(s_k, \mathbf{a}_k))$ (*Lagrange RL*)
- Penalize probability of loss: $\mathbb{E}[\ell(s_k, \mathbf{a}_k)] = \mathbb{P}[L(s_k, \mathbf{a}_k) > 0]$ can be done with

$$\ell(s_k, \mathbf{a}_k) = \begin{cases} 1 & \text{if } L(s_k, \mathbf{a}_k) > 0 \\ 0 & \text{if } L(s_k, \mathbf{a}_k) \leq 0 \end{cases}$$

Problems Beyond Expectation

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

What about e.g.

- Penalize variance of cost: $\ell(s_k, \mathbf{a}_k) = L(s_k, \mathbf{a}_k) + \omega \text{Var}(L(s_k, \mathbf{a}_k))$

Problems Beyond Expectation

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

What about e.g.

- Penalize variance of cost: $\ell(s_k, \mathbf{a}_k) = L(s_k, \mathbf{a}_k) + \omega \text{Var}(L(s_k, \mathbf{a}_k))$
- Penalize risk of loss: $\ell(s_k, \mathbf{a}_k) = \text{CVaR}_{\alpha}(L(s_k, \mathbf{a}_k))$

Problems Beyond Expectation

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

What about e.g.

- Penalize variance of cost: $\ell(s_k, \mathbf{a}_k) = L(s_k, \mathbf{a}_k) + \omega \text{Var}(L(s_k, \mathbf{a}_k))$
- Penalize risk of loss: $\ell(s_k, \mathbf{a}_k) = \text{CVaR}_{\alpha}(L(s_k, \mathbf{a}_k))$
- Constrained probability of loss: $\ell(s_k, \mathbf{a}_k) = \dots + \begin{cases} \infty & \text{if } \mathbb{P}[L(s_k, \mathbf{a}_k) > 0] > r \\ 0 & \text{if } \mathbb{P}[L(s_k, \mathbf{a}_k) > 0] \leq r \end{cases}$

Problems Beyond Expectation

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

What about e.g.

- Penalize variance of cost: $\ell(s_k, \mathbf{a}_k) = L(s_k, \mathbf{a}_k) + \omega \text{Var}(L(s_k, \mathbf{a}_k))$
- Penalize risk of loss: $\ell(s_k, \mathbf{a}_k) = \text{CVaR}_{\alpha}(L(s_k, \mathbf{a}_k))$
- Constrained probability of loss: $\ell(s_k, \mathbf{a}_k) = \dots + \begin{cases} \infty & \text{if } \mathbb{P}[L(s_k, \mathbf{a}_k) > 0] > r \\ 0 & \text{if } \mathbb{P}[L(s_k, \mathbf{a}_k) > 0] \leq r \end{cases}$

- Not all meaningful problems can be put in the “expected cost” format
- But it is worth trying (cost reformulation, state augmentation, etc)!
- “Generic MDP” $\rightarrow$ Stochastic Optimal Control
- What are the consequences, in general and for us (RLMPC)?

MDP & the Bellman Equation

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

E.g. Policy Value function

$$V^{\pi}(s) = \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \mid \begin{array}{l} s_0 = s \\ \mathbf{a}_k = \pi(s_k) \end{array} \right]$$

MDP & the Bellman Equation

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

E.g. Policy Value function

$$\begin{aligned} V^{\pi}(s) &= \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \mid \begin{array}{l} s_0 = s \\ \mathbf{a}_k = \pi(s_k) \end{array} \right] \\ &= \mathbb{E} \left[\ell(s, \mathbf{a})|_{\mathbf{a}=\pi(s)} + \gamma \ell(s_1, \mathbf{a}_1) + \gamma^2 \ell(s_2, \mathbf{a}_2) + \dots \mid \begin{array}{l} s_0 = s \\ \mathbf{a}_k = \pi(s_k) \end{array} \right] \end{aligned}$$

MDP & the Bellman Equation

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

E.g. Policy Value function

$$\begin{aligned} V^{\pi}(s) &= \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \mid \begin{array}{l} s_0 = s \\ \mathbf{a}_k = \pi(s_k) \end{array} \right] \\ &= \mathbb{E} \left[\ell(s, \mathbf{a})|_{\mathbf{a}=\pi(s)} + \gamma \ell(s_1, \mathbf{a}_1) + \gamma^2 \ell(s_2, \mathbf{a}_2) + \dots \mid \begin{array}{l} s_0 = s \\ \mathbf{a}_k = \pi(s_k) \end{array} \right] \\ &= \ell(s, \mathbf{a})|_{\mathbf{a}=\pi(s)} + \gamma \mathbb{E} \left[\ell(s_1, \mathbf{a}_1) + \gamma \ell(s_2, \mathbf{a}_2) + \dots \mid \begin{array}{l} s_0 = s \\ \mathbf{a}_k = \pi(s_k) \end{array} \right] \end{aligned}$$

MDP & the Bellman Equation

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

E.g. Policy Value function

$$\begin{aligned} V^{\pi}(s) &= \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \mid \begin{array}{l} s_0 = s \\ \mathbf{a}_k = \pi(s_k) \end{array} \right] \\ &= \mathbb{E} \left[\ell(s, \mathbf{a})|_{\mathbf{a}=\pi(s)} + \gamma \ell(s_1, \mathbf{a}_1) + \gamma^2 \ell(s_2, \mathbf{a}_2) + \dots \mid \begin{array}{l} s_0 = s \\ \mathbf{a}_k = \pi(s_k) \end{array} \right] \\ &= \ell(s, \mathbf{a})|_{\mathbf{a}=\pi(s)} + \gamma \mathbb{E} \left[\ell(s_1, \mathbf{a}_1) + \gamma \ell(s_2, \mathbf{a}_2) + \dots \mid \begin{array}{l} s_0 = s \\ \mathbf{a}_k = \pi(s_k) \end{array} \right] \\ &= \ell(s, \mathbf{a})|_{\mathbf{a}=\pi(s)} + \gamma \mathbb{E} [V^{\pi}(s_1) \mid s, \mathbf{a} = \pi(s)] \end{aligned}$$

Value function and Pathwise Bellman equations

Policy value functions

$$Q^{\pi}(s, a) = \ell(s, a) + \gamma \mathbb{E}[V^{\pi}(s_+) \mid s, a]$$

$$V^{\pi}(s) = Q^{\pi}(s, a)|_{a=\pi(s)}$$

Optimal value functions

$$Q^{\star}(s, a) = \ell(s, a) + \gamma \mathbb{E}[V^{\star}(s_+) \mid s, a]$$

$$V^{\star}(s) = Q^{\star}(s, a)|_{a=\pi^{\star}(s)} = \min_a Q^{\star}(s, a)$$

Value function and Pathwise Bellman equations

Policy value functions

$$Q^\pi(s, a) = \ell(s, a) + \gamma \mathbb{E}[V^\pi(s_+) | s, a]$$
$$V^\pi(s) = Q^\pi(s, a)|_{a=\pi(s)}$$

Optimal value functions

$$Q^*(s, a) = \ell(s, a) + \gamma \mathbb{E}[V^*(s_+) | s, a]$$
$$V^*(s) = Q^*(s, a)|_{a=\pi^*(s)} = \min_a Q^*(s, a)$$

Example: variance-averse problem

$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k \text{Var}(L(s_k, a_k)) \right] =$$
$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k (L(s_k, a_k) - \mathbb{E}[L(s_k, a_k)])^2 \right]$$

Value function and Pathwise Bellman equations

Policy value functions

$$Q^\pi(s, a) = \ell(s, a) + \gamma \mathbb{E}[V^\pi(s_+) \mid s, a]$$
$$V^\pi(s) = Q^\pi(s, a)|_{a=\pi(s)}$$

Optimal value functions

$$Q^*(s, a) = \ell(s, a) + \gamma \mathbb{E}[V^*(s_+) \mid s, a]$$
$$V^*(s) = Q^*(s, a)|_{a=\pi^*(s)} = \min_a Q^*(s, a)$$

Example: variance-averse problem

$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k \text{Var}(L(s_k, a_k)) \right] =$$
$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k (L(s_k, a_k) - \mathbb{E}[L(s_k, a_k)])^2 \right]$$

Value function

$$V^\pi(s) = \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k \text{Var}(L(s_k, a_k)) \mid a_k = \pi(s_k), s_0 = s \right]$$

Value function and Pathwise Bellman equations

Policy value functions

$$Q^\pi(s, a) = \ell(s, a) + \gamma \mathbb{E}[V^\pi(s_+) \mid s, a]$$
$$V^\pi(s) = Q^\pi(s, a)|_{a=\pi(s)}$$

Optimal value functions

$$Q^*(s, a) = \ell(s, a) + \gamma \mathbb{E}[V^*(s_+) \mid s, a]$$
$$V^*(s) = Q^*(s, a)|_{a=\pi^*(s)} = \min_a Q^*(s, a)$$

Example: variance-averse problem

$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k \text{Var}(L(s_k, a_k)) \right] =$$
$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k (L(s_k, a_k) - \mathbb{E}[L(s_k, a_k)])^2 \right]$$

Value function

$$V^\pi(s) = \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k \text{Var}(L(s_k, a_k)) \mid a_k = \pi(s_k), s_0 = s \right]$$

Naive use of Bellman principle

$$V^\pi(s) = L(s, \pi(s)) + \underbrace{\text{Var}(L(s, \pi(s)))}_{=0 \text{ (s is known)}} + \gamma \mathbb{E}[V^\pi(s_+) \mid a_k = \pi(s_k), s_0 = s]$$

Value function and Pathwise Bellman equations

Policy value functions

$$Q^\pi(s, a) = \ell(s, a) + \gamma \mathbb{E}[V^\pi(s_+) \mid s, a]$$
$$V^\pi(s) = Q^\pi(s, a)|_{a=\pi(s)}$$

Optimal value functions

$$Q^*(s, a) = \ell(s, a) + \gamma \mathbb{E}[V^*(s_+) \mid s, a]$$
$$V^*(s) = Q^*(s, a)|_{a=\pi^*(s)} = \min_a Q^*(s, a)$$

Example: variance-averse problem

$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k \text{Var}(L(s_k, a_k)) \right] =$$
$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k (L(s_k, a_k) - \mathbb{E}[L(s_k, a_k)])^2 \right]$$

Value function

$$V^\pi(s) = \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k \text{Var}(L(s_k, a_k)) \mid a_k = \pi(s_k), s_0 = s \right]$$

Naive use of Bellman principle

$$V^\pi(s) = L(s, \pi(s)) + \underbrace{\text{Var}(L(s, \pi(s)))}_{=0 \text{ (s is known)}} + \gamma \mathbb{E}[V^\pi(s_+) \mid a_k = \pi(s_k), s_0 = s]$$

Same value function as without variance

What went wrong and why?

Limitation of the classical Bellman equations

Naive application of Bellman to

$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, \mathbf{a}_k) + \gamma^k \text{Var}(L(s_k, \mathbf{a}_k)) \right]$$

yields

$$V^{\pi}(s) = L(s, \pi(s)) + \underbrace{\text{Var}(L(s, \pi(s)))}_{=0 \text{ (s is known)}} + \gamma \mathbb{E}[V^{\pi}(s_+) \mid \mathbf{a}_k = \pi(s_k), s_0 = s]$$

Limitation of the classical Bellman equations

Naive application of Bellman to

$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, \mathbf{a}_k) + \gamma^k \text{Var}(L(s_k, \mathbf{a}_k)) \right]$$

yields

$$V^{\pi}(s) = L(s, \pi(s)) + \underbrace{\text{Var}(L(s, \pi(s)))}_{=0 \text{ (s is known)}} + \gamma \mathbb{E}[V^{\pi}(s_+) \mid \mathbf{a}_k = \pi(s_k), s_0 = s]$$

if true then the following should apply:

$$V^{\pi}(s_1) = L(s_1, \pi(s_1)) + \underbrace{\text{Var}(L(s_1, \pi(s_1)))}_{=0 \text{ (s}_1 \text{ is known)}} + \gamma \mathbb{E}[V^{\pi}(s_2) \mid \mathbf{a}_1 = \pi(s_1), s_1]$$

Limitation of the classical Bellman equations

Naive application of Bellman to

$$\min_{\pi} \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k L(s_k, a_k) + \gamma^k \text{Var}(L(s_k, a_k)) \right]$$

yields

$$V^{\pi}(s) = L(s, \pi(s)) + \underbrace{\text{Var}(L(s, \pi(s)))}_{=0 \text{ (s is known)}} + \gamma \mathbb{E}[V^{\pi}(s_+) \mid a_k = \pi(s_k), s_0 = s]$$

if true then the following should apply:

$$V^{\pi}(s_1) = L(s_1, \pi(s_1)) + \underbrace{\text{Var}(L(s_1, \pi(s_1)))}_{=0 \text{ (s}_1 \text{ is known)}} + \gamma \mathbb{E}[V^{\pi}(s_2) \mid a_1 = \pi(s_1), s_1]$$

but it is wrong because $\text{Var}(L(s_1, \pi(s_1)))$ requires accounting for the alternatives that *could have happened instead of* s_1 . This intuition holds in general when this type of issue occurs.

Limitation of the classical Bellman equations

MDP

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

where $\mathbf{a}_k = \pi(\mathbf{s}_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(\mathbf{s}_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(\mathbf{s}_k)$

Limitation of the classical Bellman equations

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

Remarks

- $\ell(s_k, \mathbf{a}_k)$ with $\mathbf{a}_k = \pi(s_k)$ is a **sequence of random variable** with specific, causal dependence (filtration process)

Limitation of the classical Bellman equations

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

Remarks

- $\ell(s_k, \mathbf{a}_k)$ with $\mathbf{a}_k = \pi(s_k)$ is a **sequence of random variable** with specific, causal dependence (filtration process)
- Operator $\mathbb{E}[\cdot] : \text{random variables} \mapsto \mathbb{R}$, yields a well-posed optimization problem

Limitation of the classical Bellman equations

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

Remarks

- $\ell(s_k, \mathbf{a}_k)$ with $\mathbf{a}_k = \pi(s_k)$ is a **sequence of random variable** with specific, causal dependence (filtration process)
- Operator $\mathbb{E}[\cdot]$: random variables $\mapsto \mathbb{R}$, yields a well-posed optimization problem
- What about generic operator $\mathbb{O}[\cdot]$: random variables $\mapsto \mathbb{R}$

Limitation of the classical Bellman equations

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

Remarks

- $\ell(s_k, \mathbf{a}_k)$ with $\mathbf{a}_k = \pi(s_k)$ is a **sequence of random variable** with specific, causal dependence (filtration process)
- Operator $\mathbb{E}[\cdot]$: random variables $\mapsto \mathbb{R}$, yields a well-posed optimization problem
- What about generic operator $\mathbb{O}[\cdot]$: random variables $\mapsto \mathbb{R}$

SOC

$$\min_{\pi} \mathbb{O} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi(s_k)$

SOC (not necessarily equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{O} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

Limitation of the classical Bellman equations

MDP

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

where $\mathbf{a}_k = \pi(s_k)$

MDP (equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

Remarks

- $\ell(s_k, \mathbf{a}_k)$ with $\mathbf{a}_k = \pi(s_k)$ is a **sequence of random variable** with specific, causal dependence (filtration process)
- Operator $\mathbb{E}[\cdot]$: random variables $\mapsto \mathbb{R}$, yields a well-posed optimization problem
- What about generic operator $\mathbb{O}[\cdot]$: random variables $\mapsto \mathbb{R}$

SOC

$$\min_{\pi} \mathbb{O} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi(s_k)$

SOC (not necessarily equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{O} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

What can happen and why?

Limitation of the classical Bellman equations

SOC

$$\min_{\pi} \mathbb{O} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi(s_k)$

SOC (not necessarily equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{O}[\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

When can we do?

$$Q^*(s, \mathbf{a}) = \ell(s, \mathbf{a}) + \gamma \mathbb{O}[V^*(s_+) | s, \mathbf{a}]$$

$$V^*(s) = Q^*(s, \mathbf{a})|_{\mathbf{a}=\pi^*(s)} = \min_{\mathbf{a}} Q^*(s, \mathbf{a})$$

Limitation of the classical Bellman equations

SOC

$$\min_{\pi} \mathbb{O} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi(s_k)$

SOC (not necessarily equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{O}[\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

When can we do?

$$Q^*(s, \mathbf{a}) = \ell(s, \mathbf{a}) + \gamma \mathbb{O}[V^*(s_+) | s, \mathbf{a}]$$

$$V^*(s) = Q^*(s, \mathbf{a})|_{\mathbf{a}=\pi^*(s)} = \min_{\mathbf{a}} Q^*(s, \mathbf{a})$$

Often not...

Limitation of the classical Bellman equations

SOC

$$\min_{\pi} \mathbb{O} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi(s_k)$

SOC (not necessarily equivalent)

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{O} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

When can we do?

$$Q^*(s, \mathbf{a}) = \ell(s, \mathbf{a}) + \gamma \mathbb{O} [V^*(s_+) | s, \mathbf{a}]$$
$$V^*(s) = Q^*(s, \mathbf{a})|_{\mathbf{a}=\pi^*(s)} = \min_{\mathbf{a}} Q^*(s, \mathbf{a})$$

Often not...

Classical Bellman principle applies
if $\mathbb{O}$ is a linear operator. I.e. akin
to an expected value.

Limitation of the classical Bellman equations

SOC

$$\min_{\pi} \quad \mathbb{O} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi(s_k)$

SOC (not necessarily equivalent)

$$\min_{\pi} \quad \sum_{k=0}^{\infty} \gamma^k \mathbb{O} [\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

When can we do?

$$Q^*(s, \mathbf{a}) = \ell(s, \mathbf{a}) + \gamma \mathbb{O} [V^*(s_+) | s, \mathbf{a}]$$
$$V^*(s) = Q^*(s, \mathbf{a})|_{\mathbf{a}=\pi^*(s)} = \min_{\mathbf{a}} Q^*(s, \mathbf{a})$$

Often not...

Consequences: issues are e.g. with

- Risk-type formulation
- Probabilistic constraints
- Constraints on distributions in general

Classical Bellman principle applies if $\mathbb{O}$ is a linear operator. I.e. akin to an expected value.

Limitation of the classical Bellman equations

SOC

$$\min_{\pi} \quad \mathbb{O} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi(s_k)$

SOC (not necessarily equivalent)

$$\min_{\pi} \quad \sum_{k=0}^{\infty} \gamma^k \mathbb{O}[\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

When can we do?

$$Q^*(s, \mathbf{a}) = \ell(s, \mathbf{a}) + \gamma \mathbb{O}[V^*(s_+) | s, \mathbf{a}]$$
$$V^*(s) = Q^*(s, \mathbf{a})|_{\mathbf{a}=\pi^*(s)} = \min_{\mathbf{a}} Q^*(s, \mathbf{a})$$

Often not...

Consequences: issues are e.g. with

- Risk-type formulation
- Probabilistic constraints
- Constraints on distributions in general

Cannot be solved via “classical” methods

Classical Bellman principle applies **if** $\mathbb{O}$ is a linear operator. I.e. akin to an expected value.

Limitation of the classical Bellman equations

SOC

$$\min_{\pi} \quad \mathbb{O} \left[\sum_{k=0}^{\infty} \gamma^k \ell(s_k, \mathbf{a}_k) \right]$$

where $\mathbf{a}_k = \pi(s_k)$

SOC (not necessarily equivalent)

$$\min_{\pi} \quad \sum_{k=0}^{\infty} \gamma^k \mathbb{O}[\ell(s_k, \mathbf{a}_k)]$$

where $\mathbf{a}_k = \pi(s_k)$

When can we do?

$$Q^*(s, \mathbf{a}) = \ell(s, \mathbf{a}) + \gamma \mathbb{O}[V^*(s_+) | s, \mathbf{a}]$$
$$V^*(s) = Q^*(s, \mathbf{a})|_{\mathbf{a}=\pi^*(s)} = \min_{\mathbf{a}} Q^*(s, \mathbf{a})$$

Often not...

Consequences: issues are e.g. with

- Risk-type formulation
- Probabilistic constraints
- Constraints on distributions in general

Cannot be solved via “classical” methods

Classical Bellman principle applies **if** $\mathbb{O}$ is a linear operator. I.e. akin to an expected value.

Note: MPC with risk in the stage cost. Can do that, but it is not consistent with DP principles. Can be useful in practice, but lack of foundations.

Generalized Bellman Equations

SOC

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{L}[S_k, A_k]$$

Generalized Bellman Equations

SOC

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E}[S_k, A_k]$$

with

- State: S_k is a random variable, with $s_k \sim S_k$

Generalized Bellman Equations

SOC

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{E}[S_k, A_k]$$

with

- State: S_k is a random variable, with $s_k \sim S_k$
- Policy: $A_k = \pi(S_k)$

Generalized Bellman Equations

SOC

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{L}[S_k, A_k]$$

with

- State: S_k is a random variable, with $s_k \sim S_k$
- Policy: $A_k = \pi(S_k)$
- Cost: $\mathbb{L}[S_k, A_k] \in \mathbb{R}$ maps random variables into real numbers

Generalized Bellman Equations

SOC

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{L}[S_k, A_k]$$

with

- State: S_k is a random variable, with $s_k \sim S_k$
- Policy: $A_k = \pi(S_k)$
- Cost: $\mathbb{L}[S_k, A_k] \in \mathbb{R}$ maps random variables into real numbers
- Stochastic process: $S_{k+1} = \phi(S_k, U_k)$ with $A_k = \pi(S_k)$ yields a *filtration*

Generalized Bellman Equations

SOC

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{L}[S_k, A_k]$$

with

- State: S_k is a random variable, with $s_k \sim S_k$
- Policy: $A_k = \pi(S_k)$
- Cost: $\mathbb{L}[S_k, A_k] \in \mathbb{R}$ maps random variables into real numbers
- Stochastic process: $S_{k+1} = \phi(S_k, U_k)$ with $A_k = \pi(S_k)$ yields a *filtration*

“Generalized” Bellman Equations

$$Q^*[S, A] = \mathbb{L}[S, A] + \gamma V^*[S_+]$$

$$V^*[S] = \min_A Q^*[S, A]$$

$$\pi^*[S] = \arg \min_A Q^*[S, A]$$

Generalized Bellman Equations

SOC

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{L}[S_k, A_k]$$

with

- State: S_k is a random variable, with $s_k \sim S_k$
- Policy: $A_k = \pi(S_k)$
- Cost: $\mathbb{L}[S_k, A_k] \in \mathbb{R}$ maps random variables into real numbers
- Stochastic process: $S_{k+1} = \phi(S_k, U_k)$ with $A_k = \pi(S_k)$ yields a *filtration*

“Generalized” Bellman Equations

$$Q^*[S, A] = \mathbb{L}[S, A] + \gamma V^*[S_+]$$

$$V^*[S] = \min_A Q^*[S, A]$$

$$\pi^*[S] = \arg \min_A Q^*[S, A]$$

Bottom line:

- Recover Bellman on stochastic process

Generalized Bellman Equations

SOC

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{L}[S_k, A_k]$$

with

- State: S_k is a random variable, with $s_k \sim S_k$
- Policy: $A_k = \pi(S_k)$
- Cost: $\mathbb{L}[S_k, A_k] \in \mathbb{R}$ maps random variables into real numbers
- Stochastic process: $S_{k+1} = \phi(S_k, U_k)$ with $A_k = \pi(S_k)$ yields a *filtration*

“Generalized” Bellman Equations

$$Q^*[S, A] = \mathbb{L}[S, A] + \gamma V^*[S_+]$$

$$V^*[S] = \min_A Q^*[S, A]$$

$$\pi^*[S] = \arg \min_A Q^*[S, A]$$

Bottom line:

- Recover Bellman on stochastic process
- The “path-wise” view, with e.g. $a_k = \pi(s_k)$, is in general lost, policy needs the “entire” distribution at each time step.

Generalized Bellman Equations

SOC

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{L}[S_k, A_k]$$

with

- State: S_k is a random variable, with $s_k \sim S_k$
- Policy: $A_k = \pi(S_k)$
- Cost: $\mathbb{L}[S_k, A_k] \in \mathbb{R}$ maps random variables into real numbers
- Stochastic process: $S_{k+1} = \phi(S_k, U_k)$ with $A_k = \pi(S_k)$ yields a *filtration*

“Generalized” Bellman Equations

$$Q^*[S, A] = \mathbb{L}[S, A] + \gamma V^*[S_+]$$

$$V^*[S] = \min_A Q^*[S, A]$$

$$\pi^*[S] = \arg \min_A Q^*[S, A]$$

Bottom line:

- Recover Bellman on stochastic process
- The “path-wise” view, with e.g. $a_k = \pi(s_k)$, is in general lost, policy needs the “entire” distribution at each time step.
- The “path-wise” view is instrumental in implementable algorithms, though

Generalized Bellman Equations

SOC

$$\min_{\pi} \sum_{k=0}^{\infty} \gamma^k \mathbb{L}[S_k, A_k]$$

with

- State: S_k is a random variable, with $s_k \sim S_k$
- Policy: $A_k = \pi(S_k)$
- Cost: $\mathbb{L}[S_k, A_k] \in \mathbb{R}$ maps random variables into real numbers
- Stochastic process: $S_{k+1} = \phi(S_k, U_k)$ with $A_k = \pi(S_k)$ yields a *filtration*

“Generalized” Bellman Equations

$$Q^*[S, A] = \mathbb{L}[S, A] + \gamma V^*[S_+]$$

$$V^*[S] = \min_A Q^*[S, A]$$

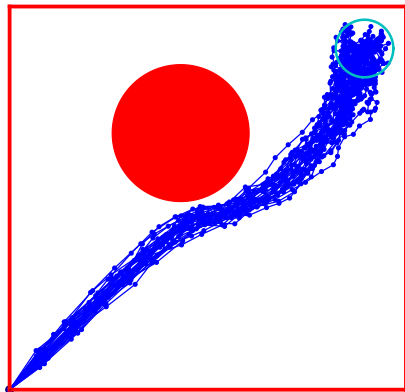
$$\pi^*[S] = \arg \min_A Q^*[S, A]$$

RLMPC (and MPC!!) on, e.g., “standard” risk measure. Operational decisions with “standard” risk measure is also problematic. These lack coherence with the lack of solution in pathwise Dynamic Programming.

But not everything is hopeless...

E.g.:

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(\mathbf{s}_k, \mathbf{a}_k) \right] \\ \text{s.t.} \quad & \mathbb{P}[\mathbf{s}_{0,\dots,\infty} \in \mathbb{S}] \geq \alpha \end{aligned}$$



But not everything is hopeless...

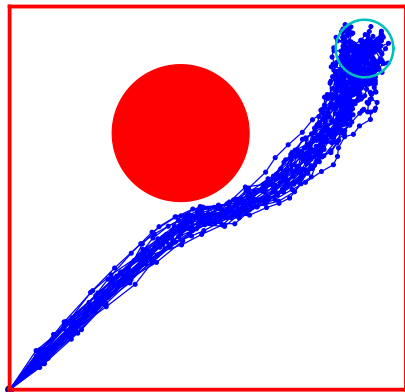
E.g.:

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \right] \\ \text{s.t.} \quad & \mathbb{P}[s_0, \dots, \infty \in \mathbb{S}] \geq \alpha \end{aligned}$$

Note: we need an absorbing set $\mathbb{T} \subset \mathbb{S}$:

$$\mathbb{P}[s_+ \in \mathbb{T} \mid s \in \mathbb{T}, a] = 1, \quad \forall a$$

and $\ell(s, a) = 0$ on $\mathbb{T}$



But not everything is hopeless...

E.g.:

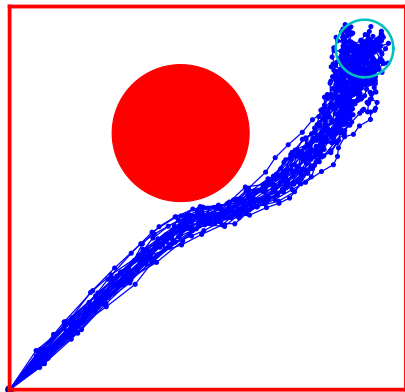
$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \right] \\ \text{s.t.} \quad & \mathbb{P}[s_0, \dots, \infty \in \mathbb{S}] \geq \alpha \end{aligned}$$

Note: we need an absorbing set $\mathbb{T} \subset \mathbb{S}$:

$$\mathbb{P}[s_+ \in \mathbb{T} \mid s \in \mathbb{T}, a] = 1, \quad \forall a$$

and $\ell(s, a) = 0$ on $\mathbb{T}$. E.g.:

- Car is parked safely at destination
- Plane is landed safely at destination



But not everything is hopeless...

E.g.:

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \right] \\ \text{s.t.} \quad & \mathbb{P}[s_0, \dots, \infty \in \mathbb{S}] \geq \alpha \end{aligned}$$

Note: we need an absorbing set $\mathbb{T} \subset \mathbb{S}$:

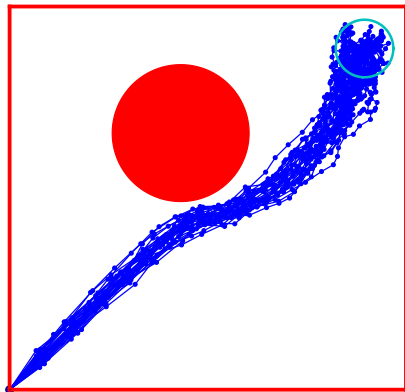
$$\mathbb{P}[s_+ \in \mathbb{T} | s \in \mathbb{T}, a] = 1, \quad \forall a$$

and $\ell(s, a) = 0$ on $\mathbb{T}$. E.g.:

- Car is parked safely at destination
- Plane is landed safely at destination

Reformulation

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \right] \\ \text{s.t.} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{s}_k \right] \leq 1 - \alpha \end{aligned}$$



But not everything is hopeless...

E.g.:

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \right] \\ \text{s.t.} \quad & \mathbb{P}[s_0, \dots, \infty \in \mathbb{S}] \geq \alpha \end{aligned}$$

Note: we need an absorbing set $\mathbb{T} \subset \mathbb{S}$:

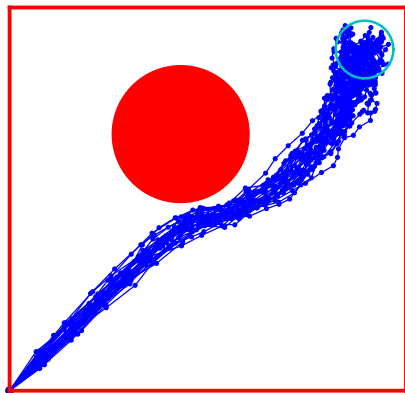
$$\mathbb{P}[s_+ \in \mathbb{T} | s \in \mathbb{T}, a] = 1, \quad \forall a$$

and $\ell(s, a) = 0$ on $\mathbb{T}$. E.g.:

- Car is parked safely at destination
- Plane is landed safely at destination

Reformulation

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \right] \\ \text{s.t.} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{s}_k \right] \leq 1 - \alpha \end{aligned}$$



With **augmented state** $\tilde{s}_k$ such that

$$\sum_{k=0}^{\infty} \tilde{s}_k = 1 \quad \text{if} \quad s_k \notin \mathbb{S} \text{ for some } k$$

such that

$$\mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{s}_k \right] = \mathbb{P}[s_0, \dots, \infty \notin \mathbb{S}]$$

Constrained MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(\mathbf{s}_k, \mathbf{a}_k) \right] \\ \text{s.t.} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{\mathbf{s}}_k \right] \leq 1 - \alpha \end{aligned}$$

has **policy improvement** solutions

Constrained MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, \mathbf{a}_k) \right] \\ \text{s.t.} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{s}_k \right] \leq 1 - \alpha \end{aligned}$$

has **policy improvement** solutions

Define

$$\begin{aligned} J(\pi_{\theta}) &= \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, \mathbf{a}_k) \mid \mathbf{a}_k = \pi_{\theta}(s_k) \right] \\ \tilde{J}(\pi_{\theta}) &= \mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{s}_k \mid \mathbf{a}_k = \pi_{\theta}(s_k) \right] \end{aligned}$$

Constrained MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \right] \\ \text{s.t.} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{s}_k \right] \leq 1 - \alpha \end{aligned}$$

has **policy improvement** solutions

Define

$$\begin{aligned} J(\pi_{\theta}) &= \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \mid a_k = \pi_{\theta}(s_k) \right] \\ \tilde{J}(\pi_{\theta}) &= \mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{s}_k \mid a_k = \pi_{\theta}(s_k) \right] \end{aligned}$$

Constrained policy gradient

$$\begin{aligned} \min_{\Delta \theta} \quad & \frac{1}{2\alpha} \|\Delta \theta\|^2 + \nabla_{\theta} J(\pi_{\theta})^{\top} \Delta \theta \\ \text{s.t.} \quad & \tilde{J}(\pi_{\theta}) + \nabla_{\theta} \tilde{J}(\pi_{\theta})^{\top} \Delta \theta \leq 1 - \alpha \end{aligned}$$

Constrained MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \right] \\ \text{s.t.} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{s}_k \right] \leq 1 - \alpha \end{aligned}$$

has **policy improvement** solutions

Define

$$\begin{aligned} J(\pi_{\theta}) &= \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \mid a_k = \pi_{\theta}(s_k) \right] \\ \tilde{J}(\pi_{\theta}) &= \mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{s}_k \mid a_k = \pi_{\theta}(s_k) \right] \end{aligned}$$

Constrained policy gradient

$$\begin{aligned} \min_{\Delta \theta} \quad & \frac{1}{2\alpha} \|\Delta \theta\|^2 + \nabla_{\theta} J(\pi_{\theta})^{\top} \Delta \theta \\ \text{s.t.} \quad & \tilde{J}(\pi_{\theta}) + \nabla_{\theta} \tilde{J}(\pi_{\theta})^{\top} \Delta \theta \leq 1 - \alpha \end{aligned}$$

Could be applied to MPC to manage probabilistic constraints in the sense of $\mathbb{P}[s_{0,\dots,\infty} \in \mathbb{S}] \geq \alpha$. Open question: variance of estimation of $\tilde{J}$ and $\nabla_{\theta} \tilde{J}$ will be high for $\alpha \approx 1$ and naive approaches.

Constrained MDP

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \right] \\ \text{s.t.} \quad & \mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{s}_k \right] \leq 1 - \alpha \end{aligned}$$

has **policy improvement** solutions

Define

$$\begin{aligned} J(\pi_{\theta}) &= \mathbb{E} \left[\sum_{k=0}^{\infty} \ell(s_k, a_k) \mid a_k = \pi_{\theta}(s_k) \right] \\ \tilde{J}(\pi_{\theta}) &= \mathbb{E} \left[\sum_{k=0}^{\infty} \tilde{s}_k \mid a_k = \pi_{\theta}(s_k) \right] \end{aligned}$$

Constrained policy gradient

$$\begin{aligned} \min_{\Delta \theta} \quad & \frac{1}{2\alpha} \|\Delta \theta\|^2 + \nabla_{\theta} J(\pi_{\theta})^{\top} \Delta \theta \\ \text{s.t.} \quad & \tilde{J}(\pi_{\theta}) + \nabla_{\theta} \tilde{J}(\pi_{\theta})^{\top} \Delta \theta \leq 1 - \alpha \end{aligned}$$

Could be applied to MPC to manage probabilistic constraints in the sense of $\mathbb{P}[s_{0,\dots,\infty} \in \mathbb{S}] \geq \alpha$. Open question: variance of estimation of $\tilde{J}$ and $\nabla_{\theta} \tilde{J}$ will be high for $\alpha \approx 1$ and naive approaches.

Bottom line: certain things may be possible if we define operational risks in a way that is compatible with pathwise Bellman equations.

Some conclusions

- Many interesting and important new developments and problems to look at
- Low-hanging fruits and hard problems
- Some questions may require reconsidering certain naive approach (e.g. risk-based MPC using classical risk measures)