Solving Mission-Wide Chance-Constrained Optimal Control Using Dynamic Programming

Kai Wang, Sébastien Gros

Department of Engineering Cybernetics
Norwegian University of Science and Technology (NTNU)

Overview

Background

- Markov Control Models
- Chance Constraints
- Mission-Wide Chance-Constrained Optimal Control

Contributions

- How Does Standard Dynamic Programming Work?
- Functional State Augmentation

Markov Control Model

State transition: $\underbrace{\mathbf{s}, \mathbf{a}}_{\text{current state\&input}} \xrightarrow[\text{or, state space model } \mathbf{f}]{\text{s.t. transition law } \rho} \underbrace{\mathbf{s}^{+}}_{\text{successor state}}$

- (Valid) transition probability function

$$\rho\left(\mathbf{s}^{+} \mid \mathbf{s}, \mathbf{a}\right) : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \to [0, \infty)$$

  such that

$$\mathbb{P}\left[\mathbf{s}_{k+1} \in \mathbb{B} \mid \mathbf{s}_k, \mathbf{a}_k\right] = \int_{\mathbb{B}} \rho\left(\mathbf{s}' \,\middle|\, \mathbf{s}_k, \mathbf{a}_k\right) \mathrm{d}\mathbf{s}', \quad \text{for all } \mathbb{B} \subseteq \mathcal{S}.$$

- In classic control, it is specified by the equation of the form

$$\mathbf{s}_{k+1} = \mathbf{f}(\mathbf{s}_k, \mathbf{a}_k, \mathbf{w}_k),$$

  where $\{\mathbf{w}_k\}$ is i.i.d with values in $\mathcal{W}$ and distribution μ, independent of $\mathbf{s}_0$.

- Model ρ is more "general"; Model $\mathbf{f}$ is more "intuitive". Relation:

$$\int_{\mathbb{B}} \rho\left(\mathbf{s}' \mid \mathbf{s}_k, \mathbf{a}_k\right) \mathrm{d}\mathbf{s}' = \int_{\mathcal{W}} \mathbf{1}_{\mathbb{B}}\big(\mathbf{f}(\mathbf{s}_k, \mathbf{a}_k, \mathbf{w})\big)\mu(\mathbf{w})\mathrm{d}\mathbf{w}, \quad \text{given } \mathbf{s}_k,\, \mathbf{a}_k\,.$$

Markov Control Model

State transition: $\underbrace{\mathbf{s},\mathbf{a}}_{\text{current state\&input}} \xrightarrow[\text{or, state space model } \mathbf{f}]{\text{s.t. transition law } \rho} \underbrace{\mathbf{s}^+}_{\text{successor state}}$

- (Valid) transition probability function

$$\rho\left(\mathbf{s}^+ \mid \mathbf{s},\mathbf{a}\right) : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \to [0,\infty)$$

  such that

$$\mathbb{P}\left[\mathbf{s}_{k+1} \in \mathbb{B} \mid \mathbf{s}_k, \mathbf{a}_k\right] = \int_{\mathbb{B}} \rho\left(\mathbf{s}' \,\middle|\, \mathbf{s}_k, \mathbf{a}_k\right) \mathrm{d}\mathbf{s}', \quad \text{for all } \mathbb{B} \subseteq \mathcal{S}.$$

- In classic control, it is specified by the equation of the form

$$\mathbf{s}_{k+1} = \mathbf{f}(\mathbf{s}_k, \mathbf{a}_k, \mathbf{w}_k),$$

  where $\{\mathbf{w}_k\}$ is i.i.d with values in $\mathcal{W}$ and distribution μ, independent of $\mathbf{s}_0$.

- Model ρ is more "general"; Model $\mathbf{f}$ is more "intuitive". Relation:

$$\int_{\mathbb{B}} \rho\left(\mathbf{s}' \mid \mathbf{s}_k, \mathbf{a}_k\right) \mathrm{d}\mathbf{s}' = \int_{\mathcal{W}} \mathbf{1}_{\mathbb{B}}\big(\mathbf{f}(\mathbf{s}_k, \mathbf{a}_k, \mathbf{w})\big)\mu(\mathbf{w})\mathrm{d}\mathbf{w}, \quad \text{given } \mathbf{s}_k,\ \mathbf{a}_k\,.$$

Markov Control Model

State transition: $\underbrace{\mathbf{s}, \mathbf{a}}_{\text{current state\&input}} \xrightarrow[\text{or, state space model } \mathbf{f}]{\text{s.t. transition law } \rho} \underbrace{\mathbf{s}^+}_{\text{successor state}}$

- (Valid) transition probability function

$$\rho\left(\mathbf{s}^+ \mid \mathbf{s}, \mathbf{a}\right) : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \to [0, \infty)$$

  such that

$$\mathbb{P}\left[\mathbf{s}_{k+1} \in \mathbb{B} \mid \mathbf{s}_k, \mathbf{a}_k\right] = \int_{\mathbb{B}} \rho\left(\mathbf{s}' \,\middle|\, \mathbf{s}_k, \mathbf{a}_k\right) \mathrm{d}\mathbf{s}', \quad \text{for all } \mathbb{B} \subseteq \mathcal{S}.$$

- In classic control, it is specified by the equation of the form

$$\mathbf{s}_{k+1} = \mathbf{f}(\mathbf{s}_k, \mathbf{a}_k, \mathbf{w}_k),$$

  where $\{\mathbf{w}_k\}$ is i.i.d with values in $\mathcal{W}$ and distribution μ, independent of $\mathbf{s}_0$.

- Model ρ is more "general"; Model $\mathbf{f}$ is more "intuitive". Relation:

$$\int_{\mathbb{B}} \rho\left(\mathbf{s}' \mid \mathbf{s}_k, \mathbf{a}_k\right) \mathrm{d}\mathbf{s}' = \int_{\mathcal{W}} \mathbf{1}_{\mathbb{B}}\big(\mathbf{f}(\mathbf{s}_k, \mathbf{a}_k, \mathbf{w})\big)\mu(\mathbf{w})\mathrm{d}\mathbf{w}, \quad \text{given } \mathbf{s}_k,\ \mathbf{a}_k\,.$$

Markov Control Model

State transition: $\underbrace{\mathbf{s}, \mathbf{a}}_{\text{current state\&input}} \xrightarrow[\text{or, state space model } \mathbf{f}]{\text{s.t. transition law } \rho} \underbrace{\mathbf{s}^{+}}_{\text{successor state}}$

- (Valid) transition probability function

$$\rho\left(\mathbf{s}^{+} \mid \mathbf{s}, \mathbf{a}\right) : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \to [0, \infty)$$

  such that

$$\mathbb{P}\left[\mathbf{s}_{k+1} \in \mathbb{B} \mid \mathbf{s}_k, \mathbf{a}_k\right] = \int_{\mathbb{B}} \rho\left(\mathbf{s}' \,\middle|\, \mathbf{s}_k, \mathbf{a}_k\right) \mathrm{d}\mathbf{s}', \quad \text{for all } \mathbb{B} \subseteq \mathcal{S}.$$

- In classic control, it is specified by the equation of the form

$$\mathbf{s}_{k+1} = \mathbf{f}(\mathbf{s}_k, \mathbf{a}_k, \mathbf{w}_k),$$

  where $\{\mathbf{w}_k\}$ is i.i.d with values in $\mathcal{W}$ and distribution μ, independent of $\mathbf{s}_0$.

- Model ρ is more "general"; Model $\mathbf{f}$ is more "intuitive". Relation:

$$\int_{\mathbb{B}} \rho\left(\mathbf{s}' \mid \mathbf{s}_k, \mathbf{a}_k\right) \mathrm{d}\mathbf{s}' = \int_{\mathcal{W}} \mathbf{1}_{\mathbb{B}}\big(\mathbf{f}(\mathbf{s}_k, \mathbf{a}_k, \mathbf{w})\big)\mu(\mathbf{w})\mathrm{d}\mathbf{w}, \quad \text{given } \mathbf{s}_k,\, \mathbf{a}_k\,.$$

Markov Control Model

Finite-horizon performance criterion:

$$J(\boldsymbol{\pi}, \mathbf{s}_0) = \mathbb{E}_{\mathbf{s}_0}^{\boldsymbol{\pi}} \left[\sum_{k=0}^{N-1} \ell(\mathbf{s}_k, \mathbf{a}_k) + \ell_N(\mathbf{s}_N) \right],$$

- Mission horizon:N
- stage cost ℓ, and terminal cost ℓ_N
- Deterministic Markov policy: $\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{N-1}) \in \Pi$

Chance Constraints

Mission-wide chance constraint:

$$\mathbb{P}\left[\mathbf{s}_{1,\ldots,N} \in \mathbb{S} \mid \mathbf{s}_0, \boldsymbol{\pi}\right] \geq 1 - \varepsilon$$

- $\mathbb{S}$ is regarded as a safe region
- $\varepsilon \in [0,1]$ is a predefined risk bound
- describe a reasonable reliability concerning the whole mission

Stage-wise chance constraints:

$$\mathbb{P}[\mathbf{s}_k \in \mathbb{S}] \geq 1 - \epsilon_k, \ \forall k \in \{0, \ldots, N\}$$

- describe reliability only at individual stages

Cumulative chance constraints:

$$\sum_{k=0}^{N} \mathbb{P}[x_k \in \mathbb{S}] \geq r$$

- safe approximations of mission-wide chance constraint if one sets $r = 1 - \varepsilon$, i.e., a feasible policy here must be feasible for the mission-wide chance constraint too.

Chance Constraints

Mission-wide chance constraint:

$$\mathbb{P}\left[\,\mathbf{s}_{1,\ldots,N} \in \mathbb{S} \mid \mathbf{s}_0, \boldsymbol{\pi}\right] \geq 1 - \varepsilon$$

- $\mathbb{S}$ is regarded as a safe region
- $\varepsilon \in [0, 1]$ is a predefined risk bound
- describe a reasonable reliability concerning the whole mission

Stage-wise chance constraints:

$$\mathbb{P}\left[\,\mathbf{s}_k \in \mathbb{S}\,\right] \geq 1 - \epsilon_k, \;\; \forall k \in \{0, \ldots, N\}$$

- describe reliability only at individual stages

Cumulative chance constraints:

$$\sum_{k=0}^{N} \mathbb{P}[\, x_k \in \mathbb{S}\,] \geq r$$

- safe approximations of mission-wide chance constraint if one sets $r = 1 - \varepsilon$, i.e., a feasible policy here must be feasible for the mission-wide chance constraint too.

Chance Constraints

Mission-wide chance constraint:

$$\mathbb{P}\left[\, \mathbf{s}_{1,\ldots,N} \in \mathbb{S} \mid \mathbf{s}_0, \boldsymbol{\pi} \right] \geq 1 - \varepsilon$$

- $\mathbb{S}$ is regarded as a safe region
- $\varepsilon \in [0, 1]$ is a predefined risk bound
- describe a reasonable reliability concerning the whole mission

Stage-wise chance constraints:

$$\mathbb{P}\left[\, \mathbf{s}_k \in \mathbb{S} \,\right] \geq 1 - \epsilon_k, \;\; \forall k \in \{0, \ldots, N\}$$

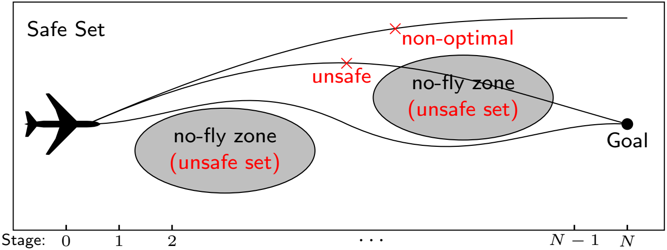
- describe reliability only at individual stages

Cumulative chance constraints:

$$\sum_{k=0}^{N} \mathbb{P}[\, x_k \in \mathbb{S}\,] \geq r$$

- safe approximations of mission-wide chance constraint if one sets $r = 1 - \varepsilon$, i.e., a feasible policy here must be feasible for the mission-wide chance constraint too.

Mission-Wide Chance-Constrained Optimal Control



Mission-wide chance constrained optimal control problem:

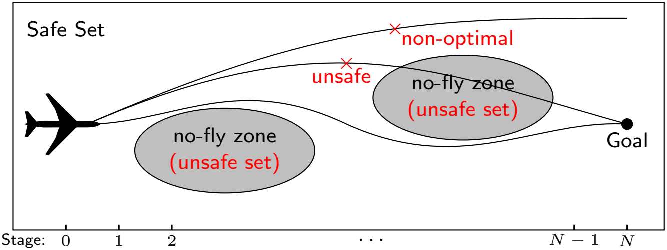
$$J^\star(\mathbf{s}_0) \stackrel{\text{def}}{=} \min_{\boldsymbol{\pi}\in\Pi} \ \mathbb{E}^{\boldsymbol{\pi}}_{\mathbf{s}_0}\left[\sum_{k=0}^{N-1} \ell(\mathbf{s}_k, \mathbf{a}_k) + \ell_N(\mathbf{s}_N)\right]$$

$$\text{s.t.} \ \ \mathbb{P}\left[\mathbf{s}_{1,\dots,N} \in \mathbb{S} \mid \mathbf{s}_0, \boldsymbol{\pi}\right] \geq 1 - \varepsilon\,.$$

- assume that this problem is well-defined and non-trivial, and attains its minimum value
- The goal is to find an optimal policy $\boldsymbol{\pi}^\star \subseteq \Pi$ for all feasible $\mathrm{s}_0 \in \mathcal{S}$.

In this paper, we provides a Dynamic Programming (DP) algorithm, finding both the value function $J^\star$ and an optimal policy $\pi^\star$.

Mission-Wide Chance-Constrained Optimal Control



Mission-wide chance constrained optimal control problem:

$$J^\star(\mathbf{s}_0) \stackrel{\text{def}}{=} \min_{\boldsymbol{\pi}\in\Pi} \ \mathbb{E}^{\boldsymbol{\pi}}_{\mathbf{s}_0}\left[\sum_{k=0}^{N-1} \ell(\mathbf{s}_k, \mathbf{a}_k) + \ell_N(\mathbf{s}_N)\right]$$

$$\text{s.t.} \ \ \mathbb{P}\left[\mathbf{s}_{1,\ldots,N} \in \mathbb{S} \mid \mathbf{s}_0, \boldsymbol{\pi}\right] \geq 1 - \varepsilon \, .$$

- assume that this problem is well-defined and non-trivial, and attains its minimum value
- The goal is to find an optimal policy $\boldsymbol{\pi}^\star \subseteq \Pi$ for all feasible $\mathrm{s}_0 \in \mathcal{S}$.

In this paper, we provides a Dynamic Programming (DP) algorithm, finding both the value function $J^\star$ and an optimal policy $\boldsymbol{\pi}^\star$.

Overview

Background

- Markov Control Models
- Chance Constraints
- Mission-Wide Chance-Constrained Optimal Control

Contributions

- How Does Standard Dynamic Programming Work?
- Functional State Augmentation

Overview

Background

- Markov Control Models
- Chance Constraints
- Mission-Wide Chance-Constrained Optimal Control

Contributions

- How Does Standard Dynamic Programming Work?
- Functional State Augmentation

How does Standard DP work?

The classic way: transform the constraint into objective via penalty function

- First, define functions:

$$V_k^{\boldsymbol{\pi}}(\mathbf{s}) = \mathbb{P}\left[\mathbf{s}_{k+1,\ldots,N} \in \mathbb{S} \,\middle|\, \mathbf{s}_k = \mathbf{s}\right] \text{ and } V_0^{\boldsymbol{\pi}}(\mathbf{s}_0) = \mathbb{P}\left[\mathbf{s}_{1,\ldots,N} \in \mathbb{S} \mid \mathbf{s}_0, \boldsymbol{\pi}\right]$$

- By Lemma 1 in the paper

$$V_k^{\boldsymbol{\pi}}(\mathbf{s}_k) = \mathbb{E}_{\mathbf{s}_{k+1}}\left[V_{k+1}^{\boldsymbol{\pi}}(\mathbf{s}_{k+1}) \mid \mathbf{s}_k\right]$$

- Transform the constraint into objective function via penalty function

$$\tilde{J}^\star(\mathbf{s}_0) = \min_{\boldsymbol{\pi}} \mathbb{E}_{\mathbf{s}_0}^{\boldsymbol{\pi}}\left[\sum_{k=0}^{N-1} \ell(\mathbf{s}_k, \mathbf{a}_k) + \ell_N(\mathbf{s}_N)\right] + \zeta\left(V_0^{\boldsymbol{\pi}}(\mathbf{s}_0)\right)$$

where $\zeta : [0,1] \to [-\infty, +\infty]$ is some penalty functions.

What would function ζ be like to ensure that DP algorithms exist?

How does Standard DP work?

The classic way: transform the constraint into objective via penalty function

- First, define functions:

$$V_k^{\boldsymbol{\pi}}(\mathbf{s}) = \mathbb{P}\left[\mathbf{s}_{k+1,\ldots,N} \in \mathbb{S} \,\middle|\, \mathbf{s}_k = \mathbf{s}\right] \text{ and } V_0^{\boldsymbol{\pi}}(\mathbf{s}_0) = \mathbb{P}\left[\mathbf{s}_{1,\ldots,N} \in \mathbb{S} \mid \mathbf{s}_0, \boldsymbol{\pi}\right]$$

- By Lemma 1 in the paper

$$V_k^{\boldsymbol{\pi}}(\mathbf{s}_k) = \mathbb{E}_{\mathbf{s}_{k+1}}\left[V_{k+1}^{\boldsymbol{\pi}}(\mathbf{s}_{k+1}) \mid \mathbf{s}_k\right]$$

- Transform the constraint into objective function via penalty function

$$\tilde{J}^\star(\mathbf{s}_0) = \min_{\boldsymbol{\pi}} \ \mathbb{E}_{\mathbf{s}_0}^{\boldsymbol{\pi}}\left[\sum_{k=0}^{N-1} \ell(\mathbf{s}_k, \mathbf{a}_k) + \ell_N(\mathbf{s}_N)\right] + \zeta\big(V_0^{\boldsymbol{\pi}}(\mathbf{s}_0)\big)$$

where $\zeta : [0,1] \to [-\infty, +\infty]$ is some penalty functions.

What would function ζ be like to ensure that DP algorithms exist?

How does Standard DP work?

The classic way: transform the constraint into objective via penalty function

- First, define functions:

$$V_k^{\boldsymbol{\pi}}(\mathbf{s}) = \mathbb{P}\left[\mathbf{s}_{k+1,\ldots,N} \in \mathbb{S} \,\middle|\, \mathbf{s}_k = \mathbf{s}\right] \text{ and } V_0^{\boldsymbol{\pi}}(\mathbf{s}_0) = \mathbb{P}\left[\mathbf{s}_{1,\ldots,N} \in \mathbb{S} \mid \mathbf{s}_0, \boldsymbol{\pi}\right]$$

- By Lemma 1 in the paper

$$V_k^{\boldsymbol{\pi}}(\mathbf{s}_k) = \mathbb{E}_{\mathbf{s}_{k+1}}\left[V_{k+1}^{\boldsymbol{\pi}}(\mathbf{s}_{k+1}) \mid \mathbf{s}_k\right]$$

- Transform the constraint into objective function via penalty function

$$\tilde{J}^{\star}(\mathbf{s}_0) = \min_{\boldsymbol{\pi}} \; \mathbb{E}_{\mathbf{s}_0}^{\boldsymbol{\pi}}\left[\sum_{k=0}^{N-1} \ell(\mathbf{s}_k, \mathbf{a}_k) + \ell_N(\mathbf{s}_N)\right] + \zeta\big(V_0^{\boldsymbol{\pi}}(\mathbf{s}_0)\big)$$

where $\zeta : [0,1] \to [-\infty, +\infty]$ is some penalty functions.

What would function ζ be like to ensure that DP algorithms exist?

How does DP work?

Solving: $\tilde{J}^\star(\mathbf{s}_0) = \min_{\boldsymbol{\pi}} \ \mathbb{E}^{\boldsymbol{\pi}}_{\mathbf{s}_0}\left[\sum_{k=0}^{N-1} \ell(\mathbf{s}_k, \mathbf{a}_k) + \ell_N(\mathbf{s}_N)\right] + \zeta\left(V_0^{\boldsymbol{\pi}}(\mathbf{s}_0)\right)$ **via DP**.

Proposition 1: If the penalty function ζ commutes with the expectation operator $\mathbb{E}_{\mathbf{s}_{k+1}}$, i.e.,

$$\zeta\left(V_k^{\boldsymbol{\pi}}(\mathbf{s}_k)\right) = \zeta\left(\mathbb{E}_{\mathbf{s}_{k+1}}\left[V_{k+1}^{\boldsymbol{\pi}}(\mathbf{s}_{k+1}) \mid \mathbf{s}_k\right]\right) = \mathbb{E}_{\mathbf{s}_{k+1}}\left[\zeta\left(V_{k+1}^{\boldsymbol{\pi}}(\mathbf{s}_{k+1})\right) \mid \mathbf{s}_k\right].$$

Then the DP solutions are given by the following recursion:

$$\tilde{J}_N(\mathbf{s}_N) = \ell_N(\mathbf{s}_N) + \zeta\left(V_N(\mathbf{s}_N)\right)$$

$$\tilde{J}_k(\mathbf{s}_k) = \min_{\mathbf{a}_k \in \mathcal{A}} \ell(\mathbf{s}_k, \mathbf{a}_k) + \int_{\mathcal{S}} \tilde{J}_{k+1}(\mathbf{s}_{k+1}) \rho(\mathbf{s}_{k+1} | \mathbf{s}_k, \mathbf{a}_k) \mathrm{d}\mathbf{s}_{k+1}$$

Remark:

- If ζ is an affine function, then **Proposition 1** holds.
- However, to enforce exact penalty of mission-wide chance constraint, ζ is of the form:

$$\zeta(x) = \begin{cases} 0, & \text{if } x \geq 1 - \varepsilon \\ +\infty, & \text{otherwise}, \end{cases}$$

which does not commute with the expectation operator. Proposition 1 fails to apply.

How does DP work?

Solving: $\tilde{J}^\star(\mathbf{s}_0) = \min_{\boldsymbol{\pi}} \; \mathbb{E}^{\boldsymbol{\pi}}_{\mathbf{s}_0}\left[\sum_{k=0}^{N-1} \ell(\mathbf{s}_k, \mathbf{a}_k) + \ell_N(\mathbf{s}_N)\right] + \zeta\left(V^{\boldsymbol{\pi}}_0(\mathbf{s}_0)\right)$ **via DP**.

Proposition 1: If the penalty function ζ commutes with the expectation operator $\mathbb{E}_{\mathbf{s}_{k+1}}$, i.e.,

$$\zeta\left(V^{\boldsymbol{\pi}}_k(\mathbf{s}_k)\right) = \zeta\left(\mathbb{E}_{\mathbf{s}_{k+1}}\left[V^{\boldsymbol{\pi}}_{k+1}(\mathbf{s}_{k+1}) \mid \mathbf{s}_k\right]\right) = \mathbb{E}_{\mathbf{s}_{k+1}}\left[\zeta\left(V^{\boldsymbol{\pi}}_{k+1}(\mathbf{s}_{k+1})\right) \mid \mathbf{s}_k\right].$$

Then the DP solutions are given by the following recursion:

$$\tilde{J}_N(\mathbf{s}_N) = \ell_N(\mathbf{s}_N) + \zeta\left(V_N(\mathbf{s}_N)\right)$$

$$\tilde{J}_k(\mathbf{s}_k) = \min_{\mathbf{a}_k \in \mathcal{A}} \ell(\mathbf{s}_k, \mathbf{a}_k) + \int_{\mathcal{S}} \tilde{J}_{k+1}(\mathbf{s}_{k+1})\rho(\mathbf{s}_{k+1}|\mathbf{s}_k, \mathbf{a}_k)\mathrm{d}\mathbf{s}_{k+1}$$

Remark:

- If ζ is an affine function, then **Proposition 1** holds.
- However, to enforce exact penalty of mission-wide chance constraint, ζ is of the form:

$$\zeta(x) = \begin{cases} 0, & \text{if } x \geq 1 - \varepsilon \\ +\infty, & \text{otherwise}, \end{cases}$$

which does not commute with the expectation operator. Proposition 1 fails to apply.

Functional State Augmentation

State Augmentation: enlarge the state information, based on which one makes decisions

- First, define forward dynamics:

$$F_0(\mathbf{s}) = \mathbf{1}_{\mathbb{S}}(\mathbf{s}), \text{ and } F_{k+1}(\mathbf{s}) = \int_{\mathbb{S}} F_k(\mathbf{s}')\rho\left[\mathbf{s} \mid \mathbf{s}', \mathbf{a}_k\right] \mathrm{d}\mathbf{s}'.$$

such that

$$\int_{\mathbb{S}} F_k(\mathbf{s})\mathrm{d}\mathbf{s} = \mathbb{P}\left[\mathbf{s}_{1,\ldots,k} \in \mathbb{S} \,\middle|\, \mathbf{s}_0, \boldsymbol{\pi}\right] \text{ and } \int_{\mathbb{S}} F_N(\mathbf{s})\mathrm{d}\mathbf{s} = \mathbb{P}\left[\mathbf{s}_{1,\ldots,N} \in \mathbb{S} \,\middle|\, \mathbf{s}_0, \boldsymbol{\pi}\right]$$

- Transform the constraint into objective function via penalty function:

$$J^\star(\mathbf{s}_0) = \min_{\boldsymbol{\pi}} \; \mathbb{E}^{\boldsymbol{\pi}}_{\mathbf{s}_0}\left[\sum_{k=0}^{N-1} \ell(\mathbf{s}_k, \mathbf{a}_k) + \ell_N(\mathbf{s}_N)\right] + \zeta\left(\int_{\mathbb{S}} F_N(\mathbf{s})\mathrm{d}\mathbf{s}\right)$$

with ζ being an EXACT penalty function, i.e.,

$$\zeta(x) = \begin{cases} 0, & \text{if } x \geq 1-\varepsilon \\ \infty, & \text{otherwise}, \end{cases}$$

Functional State Augmentation

State Augmentation: enlarge the state information, based on which one makes decisions

- First, define forward dynamics:

$$F_0(\mathbf{s}) = \mathbf{1}_{\mathbb{S}}(\mathbf{s}), \text{ and } F_{k+1}(\mathbf{s}) = \int_{\mathbb{S}} F_k(\mathbf{s}')\rho\left[\mathbf{s} \mid \mathbf{s}', \mathbf{a}_k\right] \mathrm{d}\mathbf{s}'.$$

  such that

$$\int_{\mathbb{S}} F_k(\mathbf{s})\mathrm{d}\mathbf{s} = \mathbb{P}\left[\mathbf{s}_{1,\dots,k} \in \mathbb{S} \,\middle|\, \mathbf{s}_0, \boldsymbol{\pi}\right] \text{ and } \int_{\mathbb{S}} F_N(\mathbf{s})\mathrm{d}\mathbf{s} = \mathbb{P}\left[\mathbf{s}_{1,\dots,N} \in \mathbb{S} \,\middle|\, \mathbf{s}_0, \boldsymbol{\pi}\right]$$

- Transform the constraint into objective function via penalty function:

$$J^\star(\mathbf{s}_0) = \min_{\boldsymbol{\pi}} \ \mathbb{E}_{\mathbf{s}_0}^{\boldsymbol{\pi}}\left[\sum_{k=0}^{N-1} \ell(\mathbf{s}_k, \mathbf{a}_k) + \ell_N(\mathbf{s}_N)\right] + \zeta\left(\int_{\mathbb{S}} F_N(\mathbf{s})\mathrm{d}\mathbf{s}\right)$$

  with ζ being an EXACT penalty function, i.e.,

$$\zeta(x) = \begin{cases} 0, & \text{if } x \geq 1 - \varepsilon \\ \infty, & \text{otherwise}, \end{cases}$$

Functional State Augmentation

State Augmentation: adding F_k to the state components

$$\begin{pmatrix} \mathbf{s}_{k+1} \\ F_{k+1}(\mathbf{s}) \end{pmatrix} = \begin{pmatrix} \mathbf{f}(\mathbf{s}_k, \mathbf{a}_k, \mathbf{w}_k) \\ \int_{\mathbb{S}} F_k(\mathbf{s}')\rho[\mathbf{s} \mid \mathbf{s}', \mathbf{a}_k]\mathrm{d}\mathbf{s}' \end{pmatrix}$$

Proposition 2: By defining $\delta_k = (\mathbf{s}_k, F_k)$ as the new state, then the DP solutions are given by the following recursion:

$$J_N(\delta_N) = \ell_N(\mathbf{s}_N) + \zeta\left(\int_{\mathbb{S}} F_N(\mathbf{s})\mathrm{d}\mathbf{s}\right)$$

$$J_k(\delta_k) = \min_{\mathbf{a}_k} \ell(\mathbf{s}_k, \mathbf{a}_k) + \int_{\mathcal{S}} J_{k+1}(\delta_{k+1})\rho[\mathbf{s}_{k+1}|\mathbf{s}_k, \mathbf{a}_k]\mathrm{d}\mathbf{s}_{k+1}$$

Remark:

- $\mathbf{s}_k$ is stochastic, but the functional state F_k is deterministic.
- Penalty only applies to the final state, specifically, F_N.
- Functional state F_k are infinite-dimensional.

Functional State Augmentation

State Augmentation: adding F_k to the state components

$$\begin{pmatrix} \mathbf{s}_{k+1} \\ F_{k+1}(\mathbf{s}) \end{pmatrix} = \begin{pmatrix} \mathbf{f}(\mathbf{s}_k, \mathbf{a}_k, \mathbf{w}_k) \\ \int_{\mathbb{S}} F_k(\mathbf{s}')\rho[\mathbf{s} \mid \mathbf{s}', \mathbf{a}_k]\mathrm{d}\mathbf{s}' \end{pmatrix}$$

Proposition 2: By defining $\delta_k = (\mathbf{s}_k, F_k)$ as the new state, then the DP solutions are given by the following recursion:

$$J_N(\delta_N) = \ell_N(\mathbf{s}_N) + \zeta \left(\int_{\mathbb{S}} F_N(\mathbf{s})\mathrm{d}\mathbf{s} \right)$$

$$J_k(\delta_k) = \min_{\mathbf{a}_k} \ell(\mathbf{s}_k, \mathbf{a}_k) + \int_{\mathcal{S}} J_{k+1}(\delta_{k+1})\rho[\mathbf{s}_{k+1}|\mathbf{s}_k, \mathbf{a}_k]\mathrm{d}\mathbf{s}_{k+1}$$

Remark:

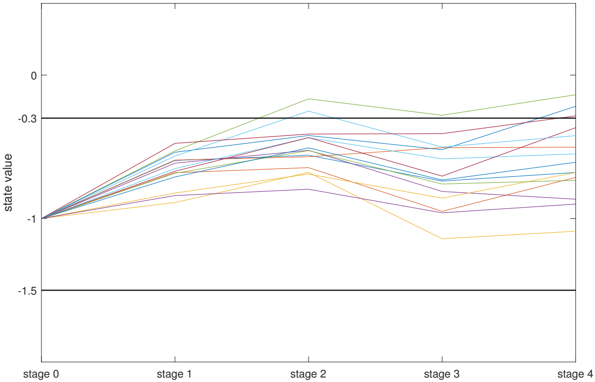
- $\mathbf{s}_k$ is stochastic, but the functional state F_k is deterministic.
- Penalty only applies to the final state, specifically, F_N.
- Functional state F_k are infinite-dimensional.

Case Study

Linear systems (one dimensional) with additive Gaussian disturbance

$$\mathbf{s}_{k+1} = 0.95\mathbf{s}_k + 1.05\mathbf{a}_k + \mathbf{w}_k, \quad \mathbf{w}_k \sim \mathcal{N}(0,\, 0.11^2)$$

$\mathbb{S} = [-1.5,\, -0.3]$; $\mathcal{A} = \{-0.2,\, -0.1,\, 0.1,\, 0.2,\, 0.3\}$; $N = 4$; $\ell(\mathbf{s}, \mathbf{a}) = 5\mathbf{s}^2 + \mathbf{a}^2$;
$\ell_N(\mathbf{s}) = 5.7\mathbf{s}^2$; $\mathbf{s}_0 = -1$; safety bound 0.8.



- Running 10^5 missions, we get empirical safety is 0.84, satisfying the safety guarantee.
- we get empirical cost 12.98, close to the theoretical minimum $J^\star(-1) = 12.99$.

Summary

Conclusions

- Mission-wide chance constraints express a meaningful reliability of the whole mission
- Standard DP only works on the cases where the penalty function commutes with expectation operator.
- We derive a DP algorithm by adding an additional functional state

Future work:

- investigate some approximation methods based on the proposed DP scheme to make numerical simulations tractable
- make use of data-driven method e.g., reinforcement learning to solve this mission-wide chance-constrained OCPs.

Contact: `kai.wang@ntnu.no`

Thank you! Questions?

Summary

Conclusions

- Mission-wide chance constraints express a meaningful reliability of the whole mission
- Standard DP only works on the cases where the penalty function commutes with expectation operator.
- We derive a DP algorithm by adding an additional functional state

Future work:

- investigate some approximation methods based on the proposed DP scheme to make numerical simulations tractable
- make use of data-driven method e.g., reinforcement learning to solve this mission-wide chance-constrained OCPs.

Contact: `kai.wang@ntnu.no`

Thank you! Questions?