ETH zürich

On the synthesis of Bellman inequalities for data-driven optimal control

Andrea Martinelli • Matilde Gargiani • John Lygeros
Automatic Control Laboratory • ETH Zurich

60th IEEE Conference on Decision and Control • Dec. 13-17, 2021

IFA

Problem setup: optimal control

Ingredients:

- A discrete-time system $x^+ = f(x, u, \xi)$ with possibly infinite state & action spaces
- A stage-cost function $\ell : \mathbb{X} \times \mathbb{U} \to \mathbb{R}_+$

Problem setup: optimal control

Ingredients:

- A discrete-time system $x^+ = f(x, u, \xi)$ with possibly infinite state & action spaces
- A stage-cost function $\ell : \mathbb{X} \times \mathbb{U} \to \mathbb{R}_+$

The discounted ∞-horizon cost associated to a stationary feedback policy $\pi : \mathbb{X} \to \mathbb{U}$ is

$$v_\pi(x) = \mathbb{E}_\xi \left[\sum_{k=0}^{\infty} \gamma^k \ell(x_k, \pi(x_k)) \;\middle|\; x_0 = x \right]$$

Problem setup: optimal control

Ingredients:

- A discrete-time system $x^+ = f(x, u, \xi)$ with possibly infinite state & action spaces
- A stage-cost function $\ell : \mathbb{X} \times \mathbb{U} \to \mathbb{R}_+$

The discounted ∞-horizon cost associated to a stationary feedback policy $\pi : \mathbb{X} \to \mathbb{U}$ is

$$v_\pi(x) = \mathbb{E}_\xi \left[\sum_{k=0}^{\infty} \gamma^k \ell(x_k, \pi(x_k)) \;\middle|\; x_0 = x \right]$$

Objective: find an optimal policy π^* such that $v_{\pi^*}(x) = \inf_\pi v_\pi(x) = v^*(x)$

Problem setup: optimal control

Ingredients:

- A discrete-time system $x^+ = f(x, u, \xi)$ with possibly infinite state & action spaces
- A stage-cost function $\ell : \mathbb{X} \times \mathbb{U} \to \mathbb{R}_+$

The discounted ∞-horizon cost associated to a stationary feedback policy $\pi : \mathbb{X} \to \mathbb{U}$ is

$$v_\pi(x) = \mathbb{E}_\xi \left[\sum_{k=0}^{\infty} \gamma^k \ell(x_k, \pi(x_k)) \;\middle|\; x_0 = x \right]$$

Objective: find an optimal policy π^* such that $v_{\pi^*}(x) = \inf_\pi v_\pi(x) = v^*(x)$

Classical ADP methods include value iteration, policy iteration and **linear programming**

Problem setup: the linear programming (LP) approach

The optimal value function admits a recursive definition (the Bellman equation)

$$v^*(x) = \underbrace{\inf_{u \in \mathbb{U}} \left\{ \ell(x,u) + \gamma \mathbb{E}_\xi \left[v^*(f(x,u,\xi)) \right] \right\}}_{(\mathcal{T}v^*)(x)}$$

Problem setup: the linear programming (LP) approach

The optimal value function admits a recursive definition (the Bellman equation)

$$v^*(x) = \underbrace{\inf_{u \in \mathbb{U}} \left\{ \ell(x,u) + \gamma \mathbb{E}_\xi \left[v^*(f(x,u,\xi)) \right] \right\}}_{(\mathcal{T}v^*)(x)}$$

$\mathcal{T}$ is monotone and contractive, hence $v \leq \mathcal{T}v \Rightarrow v \leq v^*$ and therefore

$$\begin{aligned} &\sup_{v \in \mathbb{V}} \int_{\mathbb{X}} v(x) c(dx) \\ &\text{s.t. } \; v(x) \leq (\mathcal{T}v)(x) \quad \forall x, \end{aligned}$$

Problem setup: the linear programming (LP) approach

The optimal value function admits a recursive definition (the Bellman equation)

$$v^*(x) = \underbrace{\inf_{u \in \mathbb{U}} \left\{ \ell(x,u) + \gamma \mathbb{E}_\xi \left[v^*(f(x,u,\xi)) \right] \right\}}_{(\mathcal{T}v^*)(x)}$$

$\mathcal{T}$ is monotone and contractive, hence $v \leq \mathcal{T}v \Rightarrow v \leq v^*$ and therefore

$$\begin{aligned} &\sup_{v \in \mathbb{V}} \int_{\mathbb{X}} v(x) c(dx) \\ &\text{s.t. } \; v(x) \leq (\mathcal{T}v)(x) \quad \forall x, \end{aligned}$$

We can relax the constraints by substituting $\ell(x,u) + \gamma \mathbb{E}_\xi \left[v(f(x,u,\xi)) \right] \quad \forall (x,u)$

The LP approach in practice

Efficient methods to solve LPs exist, but several sources of intractability arise

- v is an optimization variable in the ∞-dimensional space $\mathbb{V}$
- ∞ number of constraints

$$\begin{aligned} &\sup_{v \in \mathbb{V}} \int_{\mathbb{X}} v(x) c(dx) \\ &\text{s.t.} \quad v(x) \leq \ell(x,u) + \gamma \mathbb{E}_{\xi}\left[v(f(x,u,\xi))\right] \quad \forall (x,u) \end{aligned}$$

The LP approach in practice

Efficient methods to solve LPs exist, but several sources of intractability arise

- v is an optimization variable in the ∞-dimensional space $\mathbb{V}$
- ∞ number of constraints

$$\sup_{v \in \mathbb{V}} \int_{\mathbb{X}} v(x) c(dx)$$
$$\text{s.t.} \quad v(x) \leq \ell(x,u) + \gamma \mathbb{E}_{\xi}\left[v(f(x,u,\xi))\right] \quad \forall (x,u)$$

We can substitute $\sum_i \theta_i \phi_i(x)$ and sample a finite subset of constraints

$$v(x_i) \leq \ell(x_i,u_i) + \gamma \mathbb{E}_{\xi}\left[v(f(x_i,u_i,\xi_i))\right] \quad \forall (x_i,u_i) \in \mathcal{D}$$

The LP approach in practice

Efficient methods to solve LPs exist, but several sources of intractability arise

- v is an optimization variable in the ∞-dimensional space $\mathbb{V}$
- ∞ number of constraints

$$\sup_{v \in \mathbb{V}} \int_{\mathbb{X}} v(x) c(dx)$$

$$\text{s.t. } v(x) \leq \ell(x,u) + \gamma \mathbb{E}_\xi \left[v(f(x,u,\xi)) \right] \quad \forall (x,u)$$

We can substitute $\sum_i \theta_i \phi_i(x)$ and sample a finite subset of constraints

$$v(x_i) \leq \ell(x_i,u_i) + \gamma \mathbb{E}_\xi \left[v(f(x_i,u_i,\xi_i)) \right] \quad \forall (x_i,u_i) \in \mathcal{D}$$

The LP approach in practice

Efficient methods to solve LPs exist, but several sources of intractability arise

- v is an optimization variable in the ∞-dimensional space $\mathbb{V}$
- ∞ number of constraints

$$\sup_{v \in \mathbb{V}} \int_{\mathbb{X}} v(x) c(dx)$$

$$\text{s.t. } v(x) \leq \ell(x,u) + \gamma \mathbb{E}_\xi \left[v(f(x,u,\xi)) \right] \quad \forall (x,u)$$

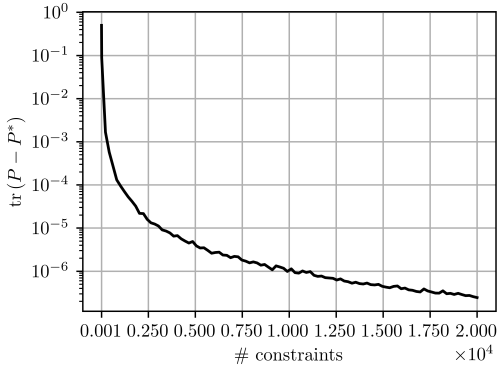
We can substitute $\sum_i \theta_i \phi_i(x)$ and sample a finite subset of constraints

$$v(x_i) \leq \ell(x_i,u_i) + \gamma \mathbb{E}_\xi \left[v(f(x_i,u_i,\xi_i)) \right] \quad \forall (x_i,u_i) \in \mathcal{D}$$

Model-free/RL framework: construct one constraint for each tuple $\{x_i, u_i, \ell_i, x_i^+\}_{i=1}^T$

Synthesis of Bellman inequalities from data

- Motivation: significant sampling/exploration is required to meet prescribed performance
- How to reconstruct Bellman inequalities from data? When a dataset is sufficiently rich?



Bellman inequalities under LQ assumptions

Assume for now deterministic linear dynamics and quadratic stage-cost

$$f(x,u,\xi) = Ax + Bu, \quad \ell(x,u) = \begin{bmatrix} x \\ u \end{bmatrix}^{\mathsf{T}} L \begin{bmatrix} x \\ u \end{bmatrix}, \quad v \in \mathbb{V}_q = \{ v : \mathbb{X} \to \mathbb{R} \mid v(x) = x^{\mathsf{T}} P x \}$$

The Bellman inequalities are

$$v(x) \leq \ell(x,u) + \gamma \mathbb{E}_{\xi}\big[v(f(x,u,\xi)) \big]$$

Bellman inequalities under LQ assumptions

Assume for now deterministic linear dynamics and quadratic stage-cost

$$f(x,u,\xi) = Ax + Bu, \quad \ell(x,u) = \begin{bmatrix} x \\ u \end{bmatrix}^\mathsf{T} L \begin{bmatrix} x \\ u \end{bmatrix}, \quad v \in \mathbb{V}_q = \{ v : \mathbb{X} \to \mathbb{R} \mid v(x) = x^\mathsf{T} P x \}$$

The Bellman inequalities are

$$\begin{gathered} v(x) \leq \ell(x,u) + \gamma \mathbb{E}_\xi \big[v(f(x,u,\xi)) \big] \\ \Downarrow \\ x^\mathsf{T} P x \leq \ell(x,u) + \gamma (Ax+Bu)^\mathsf{T} P (Ax+Bu) \end{gathered}$$

Bellman inequalities under LQ assumptions

Assume for now deterministic linear dynamics and quadratic stage-cost

$$f(x,u,\xi) = Ax + Bu, \quad \ell(x,u) = \begin{bmatrix} x \\ u \end{bmatrix}^{\mathsf{T}} L \begin{bmatrix} x \\ u \end{bmatrix}, \quad v \in \mathbb{V}_q = \{ v : \mathbb{X} \to \mathbb{R} \mid v(x) = x^{\mathsf{T}} P x \}$$

The Bellman inequalities are

$$\begin{gathered} v(x) \leq \ell(x,u) + \gamma \mathbb{E}_{\xi}\big[v(f(x,u,\xi)) \big] \\ \Downarrow \\ x^{\mathsf{T}} P x \leq \ell(x,u) + \gamma (Ax+Bu)^{\mathsf{T}} P (Ax+Bu) \\ \Downarrow \\ \operatorname{vec}\big(\underbrace{xx^{\mathsf{T}} - \gamma (Ax+Bu)(Ax+Bu)^{\mathsf{T}}}_{H(x,u)} \big)^{\mathsf{T}} \operatorname{vec}(P) \leq \ell(x,u) \end{gathered}$$

Bellman inequalities under LQ assumptions

Assume for now deterministic linear dynamics and quadratic stage-cost

$$f(x,u,\xi) = Ax + Bu, \quad \ell(x,u) = \begin{bmatrix} x \\ u \end{bmatrix}^\mathsf{T} L \begin{bmatrix} x \\ u \end{bmatrix}, \quad v \in \mathbb{V}_q = \{ v : \mathbb{X} \to \mathbb{R} \mid v(x) = x^\mathsf{T} P x \}$$

The Bellman inequalities are

$$\begin{gathered} v(x) \leq \ell(x,u) + \gamma \mathbb{E}_\xi \big[v(f(x,u,\xi)) \big] \\ \Downarrow \\ x^\mathsf{T} P x \leq \ell(x,u) + \gamma (Ax+Bu)^\mathsf{T} P (Ax+Bu) \\ \Downarrow \\ \operatorname{vec}\big(\underbrace{xx^\mathsf{T} - \gamma (Ax+Bu)(Ax+Bu)^\mathsf{T}}_{H(x,u)} \big)^\mathsf{T} \operatorname{vec}(P) \leq \ell(x,u) \end{gathered}$$

Unknown dynamics: reconstructing $H(x, u)$

We say that (X, U, X^+) is a dataset of length T when

$$X = \begin{bmatrix} x^1 & \cdots & x^T \end{bmatrix}, \quad U = \begin{bmatrix} u^1 & \cdots & u^T \end{bmatrix} \quad \text{and} \quad X^+ = AX + BU$$

Unknown dynamics: reconstructing $H(x, u)$

We say that (X, U, X^+) is a dataset of length T when

$$X = \begin{bmatrix} x^1 & \cdots & x^T \end{bmatrix}, \quad U = \begin{bmatrix} u^1 & \cdots & u^T \end{bmatrix} \quad \text{and} \quad X^+ = AX + BU$$

Inspired by behavioural theory arguments, Willem's *fundamental lemma* and many related works, we introduce the following result:

Lemma

Consider a dataset (X, U, X^+) *of length* T *with* $\begin{bmatrix} X \\ U \end{bmatrix}$ *full row-rank.*

Then, for each $(x, u) \in \mathbb{X} \times \mathbb{U}$ *there exists an* $\alpha \in \mathbb{R}^T$ *such that*

$$H(x, u) = (X\alpha)(X\alpha)^\intercal - \gamma(X^+\alpha)(X^+\alpha)^\intercal.$$

Unknown stage-cost: reconstructing $\ell(x, u)$

Proposition

Consider a dataset (X, U, X^+) *with* $\begin{bmatrix} X \\ U \end{bmatrix}$ *full row-rank and square. Then, for each* $(x, u) \in \mathbb{X} \times \mathbb{U}$ *there exists an* $\alpha \in \mathbb{R}^{\dim \mathbb{X} \times \dim \mathbb{U}}$ *such that*

$$\ell(x, u) = \alpha^\intercal L_{X,U} \alpha, \quad [L_{X,U}]_{ij} = \beta\left(\begin{bmatrix} x_i \\ u_i \end{bmatrix}, \begin{bmatrix} x_j \\ u_j \end{bmatrix}\right),$$

and β *is the bilinear form associated to* ℓ*.*

Unknown stage-cost: reconstructing $\ell(x, u)$

Proposition

Consider a dataset (X, U, X^+) *with* $\begin{bmatrix} X \\ U \end{bmatrix}$ *full row-rank and square. Then, for each* $(x, u) \in \mathbb{X} \times \mathbb{U}$ *there exists an* $\alpha \in \mathbb{R}^{\dim \mathbb{X} \times \dim \mathbb{U}}$ *such that*

$$\ell(x, u) = \alpha^\intercal L_{X,U} \alpha, \quad [L_{X,U}]_{ij} = \beta\left(\begin{bmatrix} x_i \\ u_i \end{bmatrix}, \begin{bmatrix} x_j \\ u_j \end{bmatrix}\right),$$

and β *is the bilinear form associated to* ℓ*.*

- Since $\begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} X \\ U \end{bmatrix} \alpha$, we can write $\ell(x, u) = \begin{bmatrix} x \\ u \end{bmatrix}^\intercal L \begin{bmatrix} x \\ u \end{bmatrix} = \alpha^\intercal \underbrace{\begin{bmatrix} X \\ U \end{bmatrix}^\intercal L \begin{bmatrix} X \\ U \end{bmatrix}}_{L_{X,U}} \alpha$

Unknown stage-cost: reconstructing $\ell(x, u)$

Proposition

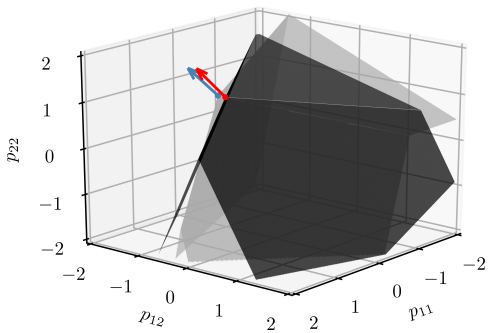
Consider a dataset (X, U, X^+) *with* $\begin{bmatrix} X \\ U \end{bmatrix}$ *full row-rank and square. Then, for each* $(x, u) \in \mathbb{X} \times \mathbb{U}$ *there exists an* $\alpha \in \mathbb{R}^{\dim \mathbb{X} \times \dim \mathbb{U}}$ *such that*

$$\ell(x, u) = \alpha^\intercal L_{X,U} \alpha, \quad [L_{X,U}]_{ij} = \beta\left(\begin{bmatrix} x_i \\ u_i \end{bmatrix}, \begin{bmatrix} x_j \\ u_j \end{bmatrix}\right),$$

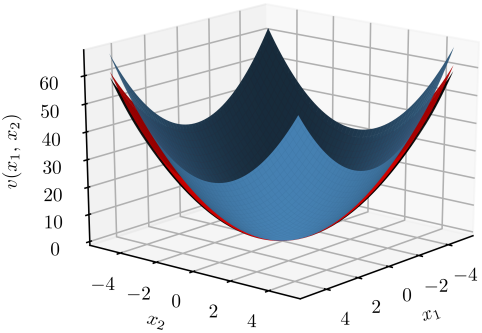
and β *is the bilinear form associated to* ℓ*.*

- Since $\begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} X \\ U \end{bmatrix} \alpha$, we can write $\ell(x, u) = \begin{bmatrix} x \\ u \end{bmatrix}^\intercal L \begin{bmatrix} x \\ u \end{bmatrix} = \alpha^\intercal \underbrace{\begin{bmatrix} X \\ U \end{bmatrix}^\intercal L \begin{bmatrix} X \\ U \end{bmatrix}}_{L_{X,U}} \alpha$
- L and $L_{X,U}$ are *congruent* and $\begin{bmatrix} X \\ U \end{bmatrix}$ takes the role of transformation matrix

2D example



(a) Bellman inequalities in the variables space



(b) Corresponding value functions

Estimation of Bellman inequalities for stochastic systems

Consider now stochastic linear systems $f(x, u, \xi) = Ax + Bu + \xi$. Estimation with iterative re-initialization can be performed as

$$v(x) \leq \ell(x, u) + \gamma \underbrace{\mathbb{E}_{\xi}\left[v(f(x, u, \xi))\right]}_{\approx \frac{1}{N}\sum_{k=1}^{N} v(f(x,u,\xi^k))}$$

Estimation of Bellman inequalities for stochastic systems

Consider now stochastic linear systems $f(x,u,\xi) = Ax + Bu + \xi$. Estimation with iterative re-initialization can be performed as

$$v(x) \leq \ell(x,u) + \gamma \underbrace{\mathbb{E}_\xi\left[v(f(x,u,\xi))\right]}_{\approx \frac{1}{N}\sum_{k=1}^{N} v(f(x,u,\xi^k))}$$

Instead, let's write again

$$\mathrm{vec}\Big(\underbrace{\mathbb{E}_\xi\left[xx^\intercal - \gamma(Ax+Bu+\xi)(Ax+Bu+\xi)^\intercal\right]}_{\mathbb{E}_\xi G(x,u,\xi)}\Big)^\intercal \mathrm{vec}(P) \leq \ell(x,u)$$

Estimation of Bellman inequalities for stochastic systems

Consider now stochastic linear systems $f(x,u,\xi) = Ax + Bu + \xi$. Estimation with iterative re-initialization can be performed as

$$v(x) \leq \ell(x,u) + \gamma \underbrace{\mathbb{E}_\xi\big[v(f(x,u,\xi))\big]}_{\approx \frac{1}{N}\sum_{k=1}^{N} v(f(x,u,\xi^k))}$$

Instead, let's write again

$$\mathrm{vec}\Big(\underbrace{\mathbb{E}_\xi\big[xx^\mathsf{T} - \gamma(Ax+Bu+\xi)(Ax+Bu+\xi)^\mathsf{T}\big]}_{\substack{\mathbb{E}_\xi G(x,u,\xi) \\ = \\ H(x,u) - \gamma\Sigma}}\Big)^\mathsf{T}\mathrm{vec}(P) \leq \ell(x,u)$$

Estimation of Bellman inequalities for stochastic systems

Proposition

Consider a linear stochastic system and $v \in \mathbb{V}_q$. *Then, given a dataset* (X, U, X^+),

$$G\left(\bar{x}, \bar{u}, \bar{\xi}\right) = \bar{x}\bar{x}^\intercal - \gamma \bar{x}^+ \bar{x}^{+\intercal},$$

where $\bar{x} = \frac{1}{N} X\mathbf{1}$, $\bar{u} = \frac{1}{N} U\mathbf{1}$, $\bar{x}^+ = \frac{1}{N} X^+\mathbf{1}$ *and* $\bar{\xi} = \frac{1}{N}\sum_{k=1}^{N} \xi^k$.

Estimation of Bellman inequalities for stochastic systems

Proposition

Consider a linear stochastic system and $v \in \mathbb{V}_q$*. Then, given a dataset* (X, U, X^+)*,*

$$G\left(\bar{x}, \bar{u}, \bar{\xi}\right) = \bar{x}\bar{x}^\intercal - \gamma \bar{x}^+ \bar{x}^{+\intercal},$$

where $\bar{x} = \frac{1}{N} X\mathbf{1}$*,* $\bar{u} = \frac{1}{N} U\mathbf{1}$*,* $\bar{x}^+ = \frac{1}{N} X^+\mathbf{1}$ *and* $\bar{\xi} = \frac{1}{N}\sum_{k=1}^{N} \xi^k$ *.*

- G can potentially be computed for any (x, u) pair depending on the available data

Estimation of Bellman inequalities for stochastic systems

Proposition

Consider a linear stochastic system and $v \in \mathbb{V}_q$*. Then, given a dataset* (X, U, X^+)*,*

$$G\left(\bar{x}, \bar{u}, \bar{\xi}\right) = \bar{x}\bar{x}^\intercal - \gamma \bar{x}^+ \bar{x}^{+\intercal},$$

where $\bar{x} = \frac{1}{N} X\mathbf{1}$*,* $\bar{u} = \frac{1}{N} U\mathbf{1}$*,* $\bar{x}^+ = \frac{1}{N} X^+\mathbf{1}$ *and* $\bar{\xi} = \frac{1}{N}\sum_{k=1}^{N} \xi^k$ *.*

- G can potentially be computed for any (x, u) pair depending on the available data
- $G(x, u, \bar{\xi}) \approx G(x, u, 0) = H(x, u)$

Estimation of Bellman inequalities for stochastic systems

Proposition

Consider a linear stochastic system and $v \in \mathbb{V}_q$*. Then, given a dataset* (X, U, X^+)*,*

$$G\left(\bar{x}, \bar{u}, \bar{\xi}\right) = \bar{x}\bar{x}^\intercal - \gamma \bar{x}^+ \bar{x}^{+\intercal},$$

where $\bar{x} = \frac{1}{N} X\mathbf{1}$*,* $\bar{u} = \frac{1}{N} U\mathbf{1}$*,* $\bar{x}^+ = \frac{1}{N} X^+\mathbf{1}$ *and* $\bar{\xi} = \frac{1}{N}\sum_{k=1}^{N} \xi^k$ *.*

- G can potentially be computed for any (x, u) pair depending on the available data
- $G(x, u, \bar{\xi}) \approx G(x, u, 0) = H(x, u)$
- The approximation is: $\mathrm{vec}(H(x, u) - \gamma\Sigma)^\intercal \mathrm{vec}(P) \leq \ell(x, u)$

Estimation of Bellman inequalities for stochastic systems

Proposition

Consider a linear stochastic system and $v \in \mathbb{V}_q$. Then, given a dataset (X, U, X^+),

$$G\left(\bar{x}, \bar{u}, \bar{\xi}\right) = \bar{x}\bar{x}^\intercal - \gamma \bar{x}^+ \bar{x}^{+\intercal},$$

where $\bar{x} = \frac{1}{N} X\mathbf{1}$, $\bar{u} = \frac{1}{N} U\mathbf{1}$, $\bar{x}^+ = \frac{1}{N} X^+\mathbf{1}$ and $\bar{\xi} = \frac{1}{N}\sum_{k=1}^{N} \xi^k$.

- G can potentially be computed for any (x, u) pair depending on the available data
- $G(x, u, \bar{\xi}) \approx G(x, u, 0) = H(x, u)$
- The approximation is: $\mathrm{vec}(H(x, u) - \gamma\Sigma)^\intercal \mathrm{vec}(P) \leq \ell(x, u)$
- Optimal value function is shifted but the optimal policy is preserved!

Conclusions and future work

In summary:

- Under LQ assumptions, a sufficiently rich dataset can be used to artificially generate all Bellman inequalities
- The associated stage-cost can be reconstructed thanks to a bilinear algebra framework
- In case of stochastic systems, we can provide an (intentionally) biased estimate of the Bellman inequalities that preserves the optimal policy without iterative re-initialization

Conclusions and future work

In summary:

- Under LQ assumptions, a sufficiently rich dataset can be used to artificially generate all Bellman inequalities
- The associated stage-cost can be reconstructed thanks to a bilinear algebra framework
- In case of stochastic systems, we can provide an (intentionally) biased estimate of the Bellman inequalities that preserves the optimal policy without iterative re-initialization

Further developments:

- How to synthesise (approximate) Bellman inequalities from data for stochastic systems
- Relax the LQ assumptions to polynomial

Contact: andremar@ethz.ch