Dynamical systems controlled by value iteration: stability, near-optimality and stopping criterion

Romain Postoyan

CNRS, CRAN, Université de Lorraine - Nancy

joint work with

- **Mathieu Granzotto** and **Jamal Daafouz**, CNRS, CRAN, Université Lorraine
- **Lucian Bușoniu**, TU Cluj - Romania
- **Dragan Nešić**, University of Melbourne - Australia

Relevance to this workshop

Reinforcement learning (RL): branch of AI

RL: optimal control of incompletely known dynamical systems [Sutton and Barto, MIT Press 2017]

Numerous open questions: stability? robustness?

RL is rooted in dynamic programming

Scope of this talk

Robust stability guarantees for systems controlled by dynamic programming **assuming knowledge of the dynamics** (no learning)

System and cost

Deterministic plant

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

where $x \in \mathbb{R}^n$, $u \in \mathcal{U}(x) \subseteq \mathbb{R}^m$, $\mathcal{U}(x)$ non-empty set of admissible inputs

Cost function

$$J_\gamma(x, \mathbf{u}) := \sum_{k=0}^{\infty} \gamma^k \ell(x(k), u(k))$$

- $\mathbf{u} = (u_0, u_1, \ldots)$
- $\ell(x, u) \geq 0$: stage cost
- $\gamma \in (0, 1]$: discount factor

Minimization problem: for any $x \in \mathbb{R}^n$,

$$V_\gamma^\star(x) := \min_{\mathbf{u}} J_\gamma(x, \mathbf{u})$$

Bellman equation

Recall

$$V_\gamma^\star(x) = \min_{\mathbf{u}} J_\gamma(x, \mathbf{u})$$

Bellman equation: for any $x \in \mathbb{R}^n$,

$$V_\gamma^\star(x) \quad = \quad \min_u \left[\ell(x,u) + \gamma V_\gamma^\star(f(x,u))\right]$$

thus

$$u_0^\star(x) \quad \in \quad \arg\min_u \left[\ell(x,u) + \gamma V_\gamma^\star(f(x,u))\right]$$

To compute $u_0^\star$, we need to know $V_\gamma^\star$

Very short introduction to value iteration

Recall Bellman equation: given $x \in \mathbb{R}^n$,

$$V_\gamma^\star(x) \quad = \quad \min_u \left[\ell(x,u) + \gamma V_\gamma^\star(f(x,u))\right]$$

Value iteration: given V_γ^0, for $i \in \mathbb{Z}_{\geq 0}$

$$V_\gamma^{i+1}(x) \quad = \quad \min_u \Big[\ell(x,u) + \gamma V_\gamma^i(f(x,u))\Big] \qquad \forall x \in \mathbb{R}^n$$

Convergence of V_γ^i to $V_\gamma^\star$ as $i \to \infty$ guaranteed under some mild conditions, see e.g., [Bertsekas, IEEE TNNLS 2017]

Objectives

Consider a plant controlled by value iteration

- Stability properties for the closed-loop system
- Stability needs to be robust [Kellett and Teel, SIAM JCON 2005]
- Exploit stability to analyse near-optimality, i.e., mismatch between V_γ^i and $V_\gamma^\star$.

Related works

Only for $\gamma = 1$ see e.g., [Wei et al., IEEE TC 2016; Heydari, IEEE TNNLS 2018; etc.]

Novelties

- Discounted cost
- More general assumptions
- More general cost
- Set stability (and not only of a point)
- Robust stability guarantees
- Near-optimality analysis

Overview

Plan

Value iteration vs MPC

Let $V_\gamma^0 = 0$ and $x \in \mathbb{R}^n$,

$$V_\gamma^1(x) \quad = \quad \min_u \left[\ell(x,u) + \gamma V_\gamma^0(f(x,u))\right] \quad = \quad \min_u \ell(x,u).$$

At step $i = 2$,

$$\begin{aligned} V_\gamma^2(x) &= \min_u \left[\ell(x,u) + \gamma V_\gamma^1(f(x,u))\right] \\ &= \min_u \left[\ell(x,u) + \gamma \ell(f(x,u), u_0)\right] \\ &= \min_{u_0, u_1} \sum_{k=0}^{1} \gamma^k \ell(x(k), u_k) \end{aligned}$$

At step $i = d \in \mathbb{Z}_{>0}$,

$$V_\gamma^d(x) \quad = \quad \min_{u_0, u_1, \ldots, u_{d-1}} \sum_{k=0}^{d-1} \gamma^k \ell(x(k), u_k)$$

We are solving a finite-horizon discounted problem

Finite-horizon discounted cost

Consider the system

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

where $x \in \mathbb{R}^n$, $u \in \mathcal{U}(x) \subseteq \mathbb{R}^m$, $\mathcal{U}(x)$ non-empty set of admissible inputs

At iteration d, we minimize the cost function

$$J_\gamma^d(x, \mathbf{u}) := \sum_{k=0}^{d-1} \gamma^k \ell(x(k), u(k))$$

Closed-loop system

$$x(k+1) \quad \in \quad f\left(x(k), \mathcal{U}_\gamma^d(x(k))\right)$$

where $\mathcal{U}_\gamma^d(x) := \left\{u_0 \,:\, \exists u_1, \ldots, u_d \;\; V_\gamma^d(x) = J_\gamma^d(x, \{u_0, \ldots, u_d\})\right\}$

Finite-horizon discounted cost

Available results in literature? No.

But

- When $\gamma = 1$, e.g., [Grimm et al., IEEE TAC 2005]
- When $d = \infty$, [Postoyan et al., IEEE TAC 2017]

Similar approach but non-trivial proof techniques

Assumptions

- Existence of optimal inputs for any d, γ, i.e., $V_\gamma^d(x)$ exists (and is finite)

To define stability, we use $\sigma : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ continuous
Examples: $\sigma(x) = |x|$, $\sigma(x) = |x|^2$, $\sigma(x) = |x|_{\mathcal{A}}$ *for some non-empty set* $\mathcal{A}$

- **Stabilizability**, i.e., there exists $\overline{\alpha}_V \in \mathcal{K}_\infty$ such that for any x, γ and d,

$$V_\gamma^d(x) \leq \overline{\alpha}_V(\sigma(x))$$

- **Detectability of the plant w.r.t. the stage cost:** Lyapunov-based conditions [Grimm et al, IEEE TAC 2005]

Stability guarantees

Theorem [Uniform semiglobal practical stability]

$\exists \beta \in \mathcal{KL}$ such that $\forall \delta, \Delta > 0$, $\exists \gamma^\star \in (0,1]$ and $d^* \in \mathbb{Z}_{\geq 0}$ such that $\forall \gamma \in (\gamma^\star, 1)$, $\forall d \geq d^\star$ and $\forall x(0) \in \{z \in \mathbb{R}^n : \sigma(z) \leq \Delta\}$, any solution to the system satisfies

$$\sigma(x(k)) \quad \leq \quad \max\{\beta(\sigma(x(0)), k), \delta\} \quad \forall k \in \mathbb{Z}_{\geq 0}.$$

Under additional conditions

- Uniform **semiglobal asymptotic** stability
- Uniform **global exponential** stability

New bounds on γ and d, which often beat those in [Grimm et al., IEEE TAC 2005] and [Postoyan et al., IEEE TAC 2017], respectively

Robustness

Recall: closed-loop system

$$x(k+1) \in f\left(x(k), \mathcal{U}_\gamma^d(x(k))\right)$$

Stability OK, but zero robustness still possible [Grimm et al., Automatica 2004]

To avoid this [Kellett and Teel, SIAM JCON 2005]:

- $f\left(x, \mathcal{U}_\gamma^d(x)\right)$ non-empty and compact for any x
- Continuous Lyapunov function, V_γ^d here

Conditions for the continuity of V_γ^d

- Relevant in its own right
- Essential for robustness
- Important for learning (approximation by radial basis functions on a compact)

Plan

Relationship between V_γ^d and $V_\gamma^\star$

Originally, for $x \in \mathbb{R}^n$,

$$V_\gamma^\star(x) = \min_{\mathbf{u}} \sum_{k=0}^{\infty} \gamma^k \ell(x(k), u(k))$$

But value iteration solves, for $d \in \mathbb{Z}_{>0}$,

$$V_\gamma^d(x) = \min_{\mathbf{u}} \sum_{k=0}^{d-1} \gamma^k \ell(x(k), u(k))$$

Fundamental question in dynamic programming [Bertsekas, 2012]

Theorem

For any $\gamma \in (0, 1]$, $d \in \mathbb{Z}_{>0}$ and $x \in \mathbb{R}^n$,

$$\left| V_\gamma^d(x) - V_\gamma^\star(x) \right| \leq \gamma^d v_{\gamma,d}(x)$$

where $v_{\gamma,d}(x) \to 0$ as $(\gamma, d) \to (1, \infty)$.

Stabilizability and detectability help

Recall:

$$\left|V_\gamma^d(x) - V_\gamma^\star(x)\right| \leq \gamma^d v_{\gamma,d}(x)$$

In the literature, e.g., [Munos and C. Szepesvári, J. of Machine Learning Research 2008]

$$\left|V_\gamma^d(x) - V_\gamma^\star(x)\right| \leq \frac{\gamma^d}{1-\gamma}\epsilon_1 + \epsilon_2$$

Pros:

- No divergence as $\gamma \to 1$
- Small error when x is close to the attractor, i.e., when $\sigma(x)$ is small

Important message

Classical control requirements do help for near-optimality analysis

Plan

When should we stop iterating?

Bound on the number of iterations available but

- Not easy to compute in general
- May be conservative

In the literature, often given $\varepsilon > 0$ stop iterating when

$$\left|V_\gamma^{d+1}(x) - V_\gamma^d(x)\right| \leq \varepsilon \qquad \forall x \text{ in a compact set}$$

Open questions:

- Stability?
- Near-optimality?
- How to select ε?

General stopping criteria

Recall

$$\left|V_\gamma^{d+1}(x) - V_\gamma^d(x)\right| \leq \varepsilon \qquad \forall x \text{ in a compact set}$$

We propose

$$\left|V_\gamma^{d+1}(x) - V_\gamma^d(x)\right| \leq c_{\text{stop}}(\varepsilon, x) \qquad \forall x \in \mathbb{R}^n,$$

where ε is a vector of tunable parameters

Examples: $c_{\text{stop}}(\varepsilon, x) = \varepsilon$, $c_{\text{stop}}(\varepsilon, x) = \varepsilon\sigma(x)$, $c_{\text{stop}}(\varepsilon, x) = \min\{\varepsilon_1, \varepsilon_2\sigma(x)\}$ etc.

Contributions: under similar assumptions as before + conditions on c_{stop}

- VI stops in a finite number of iterations
- Stability guarantees
- Clear understanding of c_{stop} on near-optimality:

$$\left|V_\gamma^d(x) - V_\gamma^\star(x)\right| \leq \alpha(c_{\text{stop}}(\varepsilon, x))$$

for some $\alpha \in \mathcal{K}_\infty$

Plan

Discussions

Recall: given V_γ^0, for $i \in \mathbb{Z}_{\geq 0}$

$$V_\gamma^{d+1}(x) \quad = \quad \min_u \left[\ell(x,u) + \gamma V_\gamma^d(f(x,u))\right] \qquad \forall x \in \mathbb{R}^n$$

Hard to compute V_γ^{d+1} exactly

In practice, typically

$$V_\gamma^{d+1}(x) \quad = \quad \min_u \left[\ell(x,u) + \gamma V_\gamma^d(f(x,u))\right] + \text{errors} \qquad \forall x \in \text{a compact set}$$

Discussions

Recall

$$V_\gamma^{d+1}(x) \quad = \quad \min_u \left[\ell(x,u) + \gamma V_\gamma^d(f(x,u))\right] + \text{errors} \qquad \forall x \in \text{a compact set}$$

Exploiting homogeneity, inspired by the notion in [Sanchez et al., IJRNC 2019]
Only solve VI equation on a given compact set $\rightarrow$ scaling of the function elsewhere

- M. Granzotto et al., *Exploiting homogeneity for the optimal control of discrete-time systems: application to value iteration*, submitted to CDC 2021

Taking explicitly into account approximation errors

- M. Granzotto et al., *Finite-horizon discounted optimal control: stability and performance*, IEEE TAC 2021
- R. Postoyan et al., *Stability guarantees for nonlinear discrete-time systems controlled by approximate value iteration*, CDC 2019

Plan

Summary and on-going works

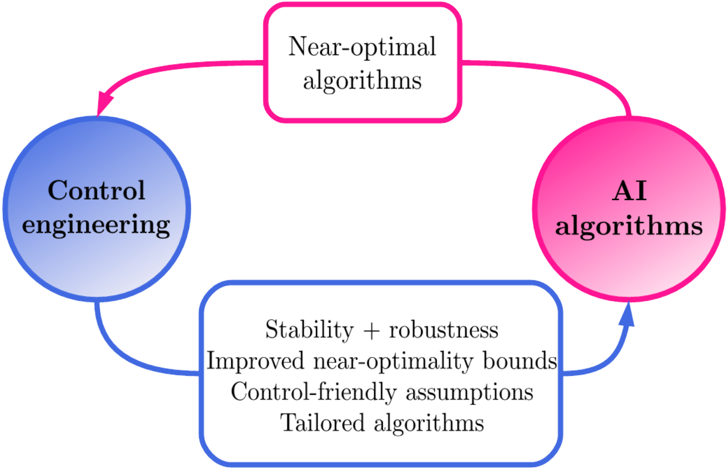
Summary

- Stability guarantees for systems controlled by value iteration
  M. Granzotto et al., *Finite-horizon discounted optimal control: stability and performance*, IEEE TAC 2021
- Stopping criterion
  M. Granzotto et al., *When to stop value iteration: stability and near-optimality versus computation*, Proceedings of Machine Learning Research (L4DC), 2021

On-going and future works

- Policy iteration
- Stochastic systems
- To take into account learning

General viewpoint



Optimistic planning [Munos, Foundations and Trends on Machine Learning 2014]

Near-optimal control inputs for systems

$$x(k+1) = f(x(k), u(k)) = f_{u(k)}(x(k))$$

where u lies **in a finite set.**

Revisit the algorithm to be applicable for the control of nonlinear switched systems

- maximization of discounted bounded cost $\rightarrow$ **minimization of undiscounted costs**
- **robust stability guarantees**
- **stronger near-optimality bounds**
- **stopping criteria to mitigate computation cost issues**

M. Granzotto et al., *Stable near-optimal control of nonlinear switched discrete-time systems: an optimistic planning-based approach*, IEEE TAC 2022

Optimistic planning [Munos, Foundations and Trends on Machine Learning 2014]

Near-optimal control inputs for systems

$$x(k+1) = f(x(k), u(k)) = f_{u(k)}(x(k))$$

where u lies **in a finite set.**

Revisit the algorithm to be applicable for the control of nonlinear switched systems

- maximization of discounted bounded cost $\rightarrow$ **minimization of undiscounted costs**
- **robust stability guarantees**
- **stronger near-optimality bounds**
- **stopping criteria to mitigate computation cost issues**

M. Granzotto et al., *Stable near-optimal control of nonlinear switched discrete-time systems: an optimistic planning-based approach*, IEEE TAC 2022

Optimistic planning [Munos, Foundations and Trends on Machine Learning 2014]

Near-optimal control inputs for systems

$$x(k+1) \quad = \quad f(x(k), u(k)) \quad = \quad f_{u(k)}(x(k))$$

where u lies **in a finite set.**

Revisit the algorithm to be applicable for the control of nonlinear switched systems

- maximization of discounted bounded cost $\rightarrow$ **minimization of undiscounted costs**
- **robust stability guarantees**
- **stronger near-optimality bounds**
- **stopping criteria to mitigate computation cost issues**

M. Granzotto et al., *Stable near-optimal control of nonlinear switched discrete-time systems: an optimistic planning-based approach*, IEEE TAC 2022

Perturbed closed-loop system

Let $\rho : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ be continuous,

$$x^+ \quad \in \quad \{v \in \mathbb{R}^n \,:\, v \in \{\eta\} + \rho(\eta)\mathcal{B},\, \eta \in F(x + \rho(x)\mathcal{B})\}$$

where

- $F(x) = f\left(x, \mathcal{U}^d_\gamma(x)\right)$ (nominal closed-loop system)
- $\mathcal{B}$ closed unit ball of $\mathbb{R}^n$
- $\rho(x) > 0$ when $\sigma(x) \neq 0$.