

Stability analysis of discrete-time finite-horizon discounted optimal control

Mathieu Granzotto, Romain Postoyan, Lucian Buşoniu, Dragan Nešić, and Jamal Daafouz

Abstract—Discounted costs are considered in many fields, like reinforcement learning, for which various algorithms can be used to obtain optimal inputs for finite horizons. The related literature mostly concentrates on optimality and largely ignores stability. In this context, we study stability of general nonlinear discrete-time systems controlled by an optimal sequence of inputs that minimizes a finite-horizon discounted cost computed in a receding horizon fashion. Assumptions are made related to the stabilizability of the system and its detectability with respect to the stage cost. Then, a Lyapunov function for the closed-loop system with the receding horizon controller is constructed and a uniform semiglobal stability property is ensured, where the adjustable parameters are both the discount factor and the horizon length. Uniform global exponential stability is guaranteed by strengthening the initial assumptions, in which case explicit bounds on the discount factor and the horizon length are provided. We compare the obtained bounds in the particular cases where there is no discount or the horizon is infinite, respectively, with related results in the literature and we show our bounds improve existing ones on the examples considered.

I. INTRODUCTION

Various algorithms used in artificial intelligence generate near-optimal control inputs, which minimize infinite-horizon discounted costs for nonlinear discrete-time systems, see e.g., [2], [6], [12], [15], [17]. These works typically concentrate on optimality guarantees, ignoring stability, while the latter is of primary importance in control engineering. Analyzing stability in this context is challenging, mostly because of the discount factor $\gamma \in (0, 1)$, which makes the discounted stage cost go to zero as time grows. The recent work in [13] provides results on the stability of nonlinear discrete-time systems controlled by an *optimal* sequence of inputs, which minimizes an infinite-horizon discounted cost. When the sequence of inputs is only *near-optimal*, results are also given in [13] but these are not applicable to most algorithms because the near-optimality bound is required to be uniform with respect to γ , which is not the case in general. For instance, the usual bounds for both value iteration [12] and policy iteration [10], [11] include terms that are linear in $\frac{1}{(1-\gamma)^2}$. These algorithms sit at the core of reinforcement learning and approximate dynamic programming fields [1], [15], [18]. Therefore, there is a need for tools to analyse stability of nonlinear discrete-time systems controlled by *near-optimal* sequences of inputs, for infinite-horizon discounted costs.

M. Granzotto, R. Postoyan and J. Daafouz are with the Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France (e-mails: {name.surname}@univ-lorraine.fr).

L. Buşoniu is with the Department of Automation, Technical University of Cluj-Napoca, Romania (e-mail: lucian.busoniu@aut.utcluj.ro).

D. Nešić is with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC 3010, Australia (e-mail: dnešic@unimelb.edu.au).

A closer look at value iteration reveals an important reason for the division by $1-\gamma$: the algorithm stops after d iterations, and then accepts an error of the form $\frac{\gamma^d}{(1-\gamma)}$ to cover the remaining rewards up to the infinite horizon. This holds even when the solution is exactly represented. In fact, even though it aims for the infinite-horizon solution, when initialized with zero values and stopped after a number of iterations, the algorithm actually provides a horizon- d optimal solution. This feature is shared by other methods that work similarly, such as optimistic planning [6]. Therefore, the objective of this paper is to provide stability results for such finite-horizon discounted costs, when the system is nonlinear. We are not aware of such results in the literature. The works in [4], [5], [16] propose generic conditions when the cost is finite-horizon but undiscounted, while [13] focuses on the case where the horizon is infinite and the stage cost is discounted, as already mentioned. In [3], the authors investigate dissipativity properties of finite-horizon discounted costs, but not stability. The fact that the cost is finite horizon and involves a discount factor leads to major technical difficulties in the stability analysis. Indeed, there is an intricate interaction between the horizon and the discount factor, and stating a precise relationship under which stability holds is non-trivial. In other words, we cannot simply combine the results of [4] and [13] to obtain the desired result: a new stability analysis is needed.

We assume that the inputs are computed in a receding-horizon fashion, as in model predictive control. This would correspond to applying the state feedback obtained after d iterations of the value iteration algorithm. We start with essentially the same assumptions as in [4], [13], namely that the plant is stabilizable and the stage cost is detectable. As in [4], we use a generic measure to define stability, thus covering the stability of the origin or of more general sets in a unified way. We then present a new Lyapunov analysis, which allows us to prove that the closed-loop system satisfies a semiglobal practical stability property provided the horizon length d and the discount factor γ are sufficiently big and close to 1, respectively. Uniform global exponential stability is then provided by strengthening the assumptions, and explicit bounds on γ and d are given. The Lyapunov function used to prove stability is continuous under mild assumptions, ensuring the robustness of the property according to [9]. It appears that the Lyapunov analysis we propose differs from those in [4], [13] even when $\gamma = 1$ and $d = \infty$, respectively. This allows us to derive new bounds on d or γ , according to the considered scenario, which can be less conservative than those in [4], [13] as shown on several examples; this

constitutes an additional contribution of the paper.

The rest of the paper is organized as follows. The problem is formally stated in Section II. The main results are given in Section III. The comparison with [4], [13] is presented in Section III-D. The proof of the Lyapunov result is provided in Section IV and conclusions are given in Section V. The Appendix contains technical results needed in the main proofs. **Notation.** Let $\mathbb{R} := (-\infty, \infty)$, $\mathbb{R}_{\geq 0} := [0, \infty)$, $\mathbb{Z}_{\geq 0} := \{0, 1, 2, \dots\}$ and $\mathbb{Z}_{> 0} := \{1, 2, \dots\}$. We use (x, y) to denote $[x^T, y^T]^T$, where $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ and $n, m \in \mathbb{Z}_{> 0}$. A function $\chi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is of class \mathcal{K} if it is continuous, zero at zero and strictly increasing, and it is of class \mathcal{K}_∞ if it is of class \mathcal{K} and unbounded. A continuous function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is of class \mathcal{KL} when $\chi(\cdot, t)$ is of class \mathcal{K} for any $t \geq 0$ and $\beta(s, \cdot)$ is decreasing to 0 for any $s \geq 0$. The notation \mathbb{I} stands for the identity map from $\mathbb{R}_{\geq 0}$ to $\mathbb{R}_{\geq 0}$. For any sequence $\mathbf{u} = [u_0, u_1, \dots]$ of length $d \in \mathbb{Z}_{\geq 0} \cup \{\infty\}$ where $u_i \in \mathbb{R}^m$, $i \in \mathbb{Z}_{\geq 0}$, and any $k \in \{0, \dots, d\}$, we use $\mathbf{u}|_k$ to denote the first k elements of \mathbf{u} , i.e. $\mathbf{u}|_k = [u_0, \dots, u_{k-1}]$ and $\mathbf{u}|_0 = u_0$ by convention. Let $f : \mathbb{R} \rightarrow \mathbb{R}$, we use $f^{(k)}$ for the composition of function f to itself k times, where $k \in \mathbb{Z}_{\geq 0}$, and $f^{(0)} = \mathbb{I}$. We use $\lfloor \cdot \rfloor$ to denote the floor function. The Euclidian norm of a vector $x \in \mathbb{R}^n$ is denoted by $|x|$.

II. PROBLEM STATEMENT

Consider the system

$$x_{k+1} = f(x_k, u_k), \quad (1)$$

with state $x \in \mathbb{R}^n$, inputs $u \in \mathcal{U}(x) \subseteq \mathbb{R}^m$, where $\mathcal{U}(x)$ is the nonempty set of admissible inputs for state x , and $f : \mathcal{W} \rightarrow \mathbb{R}^n$ where $\mathcal{W} := \{(x, u) : x \in \mathbb{R}^n, u \in \mathcal{U}(x)\}$. We use $\phi(k, x, \mathbf{u}|_k)$ to denote the solution to system (1) at time $k \in \mathbb{Z}_{\geq 0}$ with initial condition x and inputs $\mathbf{u}|_k = [u_0, u_1, \dots, u_{k-1}]$, with the convention $\phi(0, x, \cdot) = x$.

We study discounted finite-horizon cost functions of the form

$$J_{\gamma, d}(x, \mathbf{u}) := \sum_{k=0}^d \gamma^k \ell(\phi(k, x, \mathbf{u}|_k), \mathbf{u}_k) \quad (2)$$

where $x \in \mathbb{R}^n$, \mathbf{u} is a sequence of $d+1$ elements of \mathbb{R}^m , $\ell : \mathcal{W} \rightarrow \mathbb{R}_{\geq 0}$, $\gamma \in (0, 1]$ is the discount factor and $d \in \mathbb{Z}_{> 0} \cup \{\infty\}$ is the horizon.

We assume that for any $x \in \mathbb{R}^n$, $\gamma \in (0, 1]$ and $d \in \mathbb{Z}_{> 0} \cup \{\infty\}$, there is a sequence \mathbf{u} that minimizes cost (2), as formalized next.

Standing Assumption (SA): For any $x \in \mathbb{R}^n$, $\gamma \in (0, 1]$ and $d \in \mathbb{Z}_{> 0} \cup \{\infty\}$, there exists a sequence of $d+1$ admissible inputs $\mathbf{u}_{\gamma, d}^*(x)$, called *optimal input sequence*, which minimizes (2), i.e.

$$J_{\gamma, d}(x, \mathbf{u}_{\gamma, d}^*(x)) = \min_{\mathbf{u}} J_{\gamma, d}(x, \mathbf{u}) =: V_{\gamma, d}(x), \quad (3)$$

where $V_{\gamma, d}$ is the *optimal cost function*. \square

Conditions to ensure the satisfaction of SA can be found in [8]. According to SA, for any $x \in \mathbb{R}^n$, $\gamma \in (0, 1]$ and

$d \in \mathbb{Z}_{> 0} \cup \{\infty\}$, the set below is non-empty

$$\mathcal{U}_{\gamma, d}^*(x) := \{u_0 : \exists u_1, \dots, u_d \in \mathbb{R}^m \text{ such that } V_{\gamma, d}(x) = J_{\gamma, d}(x, [u_0, \dots, u_d])\}. \quad (4)$$

Note that $\mathcal{U}_{\gamma, d}^*(x)$ may be a set with multiple elements because the optimal sequence may be non-unique for given x , γ and d .

We consider the scenario where system (1) is controlled in a receding horizon fashion in the sense that, at each time instant $k \in \mathbb{Z}_{\geq 0}$, the first element of the optimal input $\mathbf{u}_{\gamma, d}^*(x_k)$, which may be non-unique, is applied to system (1). This leads to the difference inclusion

$$x_{k+1} \in f(x_k, \mathcal{U}_{\gamma, d}^*(x(k))) =: F_{\gamma, d}^*(x_k), \quad (5)$$

where $f(x, \mathcal{U}_{\gamma, d}^*(x))$ is the set $\{f(x, u) : u \in \mathcal{U}_{\gamma, d}^*(x)\}$. We denote by $\phi(k, x)$, with some abuse of notation, a solution to (5) at time $k \in \mathbb{Z}_{\geq 0}$ with initial condition $x \in \mathbb{R}^n$.

Our objective is to analyse the stability of system (5) using Lyapunov-based arguments. In particular, we want to investigate the influence of the cost parameters γ and d on stability. To this end, we make the following assumptions on the stabilizability and detectability of system (1) and cost function (2).

Assumption 1: There exist $\bar{\alpha}_V, \alpha_W \in \mathcal{K}_\infty$, continuous functions $W, \sigma : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$, $\bar{\alpha}_W : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ continuous, non-decreasing and zero at zero, such that the following holds.

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

$$V_{\gamma, d}(x) \leq \bar{\alpha}_V(\sigma(x)). \quad (6)$$

(ii) For any $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$,

$$W(x) \leq \bar{\alpha}_W(\sigma(x)) \quad (7)$$

$$W(f(x, u)) - W(x) \leq -\alpha_W(\sigma(x)) + \ell(x, u). \quad (8)$$

(iii) $\mathbb{I} - \alpha_W \circ (\bar{\alpha}_V + \bar{\alpha}_W)^{-1} \in \mathcal{K}_\infty$. \square

Function σ in Assumption 1 serves as a measure of the state and will be used to define stability. When investigating the stability of the origin for instance, we typically take $\sigma(x) = |x|$, $\sigma(x) = |x|^2$, or $\sigma(x) = x^T P x$ with P symmetric, positive definite, for any $x \in \mathbb{R}^n$. When interested in stability of a set $\mathcal{A} \subseteq \mathbb{R}^n$, σ can be defined as $\sigma = |\cdot|_{\mathcal{A}}$ for instance, where $|x|_{\mathcal{A}} = \inf\{|z - x| : z \in \mathcal{A}\}$ for any $x \in \mathbb{R}^n$. Item (i) of Assumption 1 is related to the asymptotic controllability (stabilizability) of system (1) with respect to σ , see for more detail Section III in [4] and Lemma 1 in [13]. Item (ii) of Assumption 1 is a detectability property of the stage cost ℓ with respect to σ . To see it, consider the particular case where $W = 0$ so that (8) reduces to $\alpha_W(\sigma(x)) \leq \ell(x, u)$. Thus, when $\ell(x, u) = 0$, $\sigma(x) = 0$ since $\alpha_W \in \mathcal{K}_\infty$. Item (iii) of Assumption 1 is convenient for the forthcoming stability analysis. It can be assumed without loss of generality, as, if it is not verified, we can always upper-bound $\mathbb{I} - \alpha_W \circ (\bar{\alpha}_V + \bar{\alpha}_W)^{-1}$ by $\mathbb{I} - \tilde{\alpha}$, which is of class \mathcal{K}_∞ for suitable $\tilde{\alpha} \in \mathcal{K}_\infty$, according to Lemma B.1 in [7], as this is enough for the forthcoming stability analysis.

Remark 1: A more general detectability assumption is made in [4], [13], namely $W(f(x, u)) - W(x) \leq -\alpha_W(\sigma(x)) + \chi(\ell(x, u))$ instead of (8), where $\chi \in \mathcal{K}_\infty$. The extension to this case is left for future work. \square

III. STABILITY RESULTS

A. Lyapunov Properties

The satisfaction of Assumption 1 allows us to derive the next Lyapunov properties, based on which we derive our main stability result for system (5) afterwards.

Theorem 1: Suppose Assumption 1. There exist $\underline{\alpha}_Y, \bar{\alpha}_Y, \alpha_Y \in \mathcal{K}_\infty$ and, for any $\gamma \in (0, 1]$ and $d \in \mathbb{Z}_{>0} \cup \{\infty\}$, there exists $Y_{\gamma, d} : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ such that the following holds.

(i) For any $x \in \mathbb{R}^n$,

$$\underline{\alpha}_Y(\sigma(x)) \leq Y_{\gamma, d}(x) \leq \bar{\alpha}_Y(\sigma(x)). \quad (9)$$

(ii) For any $x \in \mathbb{R}^n, v \in F_{\gamma, d}^*(x)$,

$$Y_{\gamma, d}(v) - Y_{\gamma, d}(x) \leq \frac{1}{\gamma} \left(-\alpha_Y(\sigma(x)) + \Upsilon(Y_{\gamma, d}(x), \gamma, d) \right) \quad (10)$$

where $\Upsilon : \mathbb{R}_{\geq 0} \times [0, 1] \times (\mathbb{Z}_{>0} \cup \{\infty\}) \rightarrow \mathbb{R}_{\geq 0}$ is defined in Table I, and is such that, for any $s \geq 0$, $\Upsilon(s, \gamma, d) \rightarrow 0$ when $\gamma \rightarrow 1$ and $d \rightarrow \infty$. \square

Function $Y_{\gamma, d}$ plays the role of a Lyapunov function in Theorem 1, its expression as well as the expressions of $\alpha_Y, \underline{\alpha}_Y, \bar{\alpha}_Y$ are given in Table I. Item (i) states that it is positive definite and radially unbounded with respect to the set $\{x : \sigma(x) = 0\}$, uniformly in γ and d . Item (ii) of Theorem 1 shows that $Y_{\gamma, d}$ strictly decreases along the solutions to (5) up to a perturbative term Υ , which can be reduced as much as desired by selecting γ close to 1 and d big. We stress that γ has to be selected close to 1 and d has to be large in order for Υ to be small in (10), which is consistent with previous works on discounted infinite-horizon control [13] and undiscounted finite-horizon control [4] where a similar Lyapunov inequality is given. Theorem 1 is actually a generalization of Theorem 1 in [4] to discounted cost and of Theorem 1 in [13] to finite-horizon. The perturbative term Υ differs from the corresponding one in (5) in [4] when $\gamma = 1$, and from the one in item (b) of Theorem 1 in [13] when $d = \infty$, because of the way the Lyapunov analysis is carried out in the proof of Theorem 1. The new analysis we propose is motivated by the fact that it leads to different bounds on d (and γ) under which stability will be preserved. This is discussed in more detail in Section III-D.

B. Main Result

We are ready to state the main stability result, whose proof is omitted for space reasons.

Theorem 2: Consider system (5) and suppose Assumption 1 holds. There exists $\beta \in \mathcal{KL}$ such that for any $\delta, \Delta > 0$, there exist $\gamma^* \in (0, 1)$ and $d^* \in \mathbb{Z}_{>0}$ such that for any $\gamma \in (\gamma^*, 1]$, $d \in (d^*, \infty]$, $x \in \{z \in \mathbb{R}^n : \sigma(z) \leq \Delta\}$, any solution $\phi(\cdot, x)$ to system (5) satisfies, for all $k \in \mathbb{Z}_{\geq 0}$

$$\sigma(\phi(k, x)) \leq \max\{\beta(\sigma(x), k), \delta\}. \quad (11)$$

Theorem 2 ensures a semiglobal practical stability property, i.e. given any set of initial conditions $\{z \in \mathbb{R}^n : \sigma(z) \leq \Delta\}$ where $\Delta > 0$, and any (arbitrarily small) δ , we can select γ and d such that (11) holds. Conditions that allow deriving non conservative choice of γ^* and d^* for Theorem 2 are difficult to draw due to interplay between γ and d in Υ . Explicit bounds are provided in the following by strengthening the conditions of Theorem 2, which also allows us to ensure stronger stability properties.

Before that, however, we know from [9] that it is essential to work with a continuous Lyapunov function to endow the stability properties with some nominal robustness. Here, function $Y_{\gamma, d}$ in Theorem 2 serves as a Lyapunov function. To ensure it is continuous, we need to guarantee that $V_{\gamma, d}$ is, since $Y_{\gamma, d} = V_{\gamma, d} + W$ and W is continuous according to Assumption 1. Additional assumptions are needed for this purpose.

Assumption 2: The following properties hold.

(i) f and ℓ are continuous.

(ii) Either \mathcal{U} is bounded, i.e. the input domain \mathcal{U} is contained in a ball of finite radius, or for each compact set \mathcal{C} , real number η , and positive integer d , there exists $\mu > 0$ such that for $x \in \mathcal{C}$, all sequences \mathbf{u} of length $d + 1$ satisfying $J_{\gamma, d}(x, \mathbf{u}) \leq \eta$ satisfy $|u_k| \leq \mu$ for $k \in \{0, \dots, d\}$. \square

The next lemma ensures the continuity of $V_{\gamma, d}$ when d is finite, whose proof is an application of Theorem 1.17 in [14] and is omitted for space reasons.

Lemma 1: Consider system (1) and suppose Assumption 2 holds. For any $\gamma \in (0, 1]$ and $d \in \mathbb{Z}_{>0}$, function $V_{\gamma, d}$ is continuous. \square

When d is infinite, Theorem 3 in [13] can be applied to ensure that $V_{\gamma, d}$ is continuous.

C. Uniform Global Exponential Stability

We strengthen the conditions of Theorem 2 to ensure a uniform global exponential stability property in the next statement, whose proof is omitted for space reasons.

Corollary 1: Suppose that Assumption 1 is satisfied and there exist $\bar{a}_W \geq 0, a_W, \bar{a}_V > 0$ such that $\bar{\alpha}_V(s) \leq \bar{a}_V \cdot s$, $\bar{\alpha}_W(s) \leq \bar{a}_W \cdot s$, $\alpha_W(s) \geq a_W \cdot s$ for any $s \geq 0$. Let γ^*, d^* be such that

$$(1 - \gamma^*) + \frac{\bar{a}_V}{a_W} \left(1 - \frac{a_W}{\bar{a}_V + \bar{a}_W} \right)^{d^*} < \frac{a_W}{\bar{a}_V + \bar{a}_W}. \quad (12)$$

Then, there exist $K, \lambda > 0$, such that for any $\gamma \in (\gamma^*, 1]$, $d \in (d^*, \infty]$, for any $x \in \mathbb{R}^n$, the solution $\phi(\cdot, x)$ to system (5) satisfies $\sigma(\phi(k, x)) \leq K\sigma(x)e^{-\lambda k}$ for all $k \in \mathbb{Z}_{\geq 0}$. \square

Corollary 1 ensures a uniform global exponential stability property of $\{x : \sigma(x) = 0\}$. It also provides explicit conditions on the pair (γ^*, d^*) under which stability is guaranteed. Consistently with Theorem 2, to find a suitable pair (γ^*, d^*) , we have to take γ^* close to 1 and d^* large. Inequality (12) imposes conditions on the minimal allowable values of γ and of d . Indeed, the smallest value of γ ensuring (12), denoted $\bar{\gamma}$, occurs when $1 - \bar{\gamma} = \frac{a_W}{\bar{a}_V + \bar{a}_W}$,

TABLE I: Expressions of the functions used in Theorem 1

$Y_{\gamma,d} := V_{\gamma,d} + W$ $\underline{\alpha}_Y := \alpha_W$ $\bar{\alpha}_Y := \bar{\alpha}_V + \bar{\alpha}_W$ $\alpha_Y := \alpha_W$	
$\Upsilon(s, \gamma, d) := \begin{cases} (1-\gamma)s + \gamma^d \bar{\alpha}_V \circ \underline{\alpha}_Y^{-1} \circ \left(\frac{\mathbb{I} - \alpha_Y \circ \bar{\alpha}_Y^{-1}}{\gamma} \right)^{(d)}(s) & \text{when } \gamma \in (0, 1) \text{ and } d \in \mathbb{Z}_{>0} \\ \bar{\alpha}_V \circ \underline{\alpha}_Y^{-1} \circ (\mathbb{I} - \alpha_Y \circ \bar{\alpha}_Y^{-1})^{(d)}(s) & \text{when } \gamma = 1 \text{ and } d \in \mathbb{Z}_{>0} \\ (1-\gamma)s & \text{when } \gamma \in (0, 1) \text{ and } d = \infty \end{cases}$	

i.e. when $\gamma > \bar{\gamma} := 1 - \frac{a_W}{\bar{a}_V + \bar{a}_W}$. Vice-versa, the smallest value of d ensuring (12), denoted \bar{d} , is the smallest integer such that occurs when $\frac{\bar{a}_V}{a_W} \left(1 - \frac{a_W}{\bar{a}_V + \bar{a}_W}\right)^{\bar{d}} \leq \frac{a_W}{\bar{a}_V + \bar{a}_W}$, i.e. $d > \bar{d} := \lfloor \frac{\ln(\bar{a}_V(\bar{a}_V + \bar{a}_W)/a_W^2)}{-\ln(1 - \frac{a_W}{\bar{a}_V + \bar{a}_W})} \rfloor$. As a result, (12) can be used as follows to select the pair (γ, d) . We either first fix $\gamma \in (\bar{\gamma}, 1]$ and then select d such that $\frac{\bar{a}_V}{a_W} \left(1 - \frac{a_W}{\bar{a}_V + \bar{a}_W}\right)^d < \frac{a_W}{\bar{a}_V + \bar{a}_W} - (1-\gamma)$ holds, or we first fix $d > \bar{d}$ and select γ such that $(1-\gamma) < \frac{a_W}{\bar{a}_V + \bar{a}_W} - \frac{\bar{a}_V}{a_W} \left(1 - \frac{a_W}{\bar{a}_V + \bar{a}_W}\right)^d$ holds. By doing so, the pair obtained (γ, d) ensures the satisfaction of (12).

D. Comparison with existing results

It is difficult to compare the conditions on γ and d in the general case of Theorem 2 with those in [4] when $\gamma = 1$ and [13] when $d = \infty$. For this reason we perform this comparison with the bounds given by Corollary 1.

In the next lemma, whose proof is omitted for space reasons, we compare the bounds derived from Corollary 1 either when $\gamma = 1$ or $d = \infty$ with those given in Corollary 3 in [4], which we denote¹ as $d_{[4]}$, and in Corollary 2 in [13], which we denote as $\gamma_{[13]}$, respectively.

Lemma 2: Under the conditions of Corollary 1, the following holds.

- (i) When $\bar{a}_W < a_W$, $\bar{\gamma} < \gamma_{[13]} = \frac{\bar{a}_V}{\bar{a}_V + \bar{a}_W}$ where $\bar{\gamma} = 1 - \frac{a_W}{\bar{a}_V + \bar{a}_W}$ is the bound on γ given by (12) when $d = \infty$.
- (ii) $\bar{d} = -\ln d_{[4]} / \ln \left(1 - \frac{a_W}{\bar{a}_V + \bar{a}_W}\right)$ where² $d_{[4]} = \frac{\bar{a}_V(\bar{a}_V + \bar{a}_W)}{a_W^2}$ and $\bar{d} = -\ln \left(\frac{\bar{a}_V(\bar{a}_V + \bar{a}_W)}{a_W^2}\right) / \ln \left(1 - \frac{a_W}{\bar{a}_V + \bar{a}_W}\right)$ is the bound on d given by (12) when $\gamma = 1$. \square

Item (i) of Lemma 2 implies that the minimum discount factor $\bar{\gamma}$ given by Corollary 1 when $d = \infty$ is strictly smaller than the bound found in Corollary 2 from [13] when $\bar{\alpha}_W < \alpha_W$. Item (ii) of Lemma 2 provides a direct relationship between the estimate horizon $d_{[4]}$ of Corollary 3 from [4] and our minimum horizon estimate \bar{d} . It can therefore be used to infer which bound is the tightest.

We illustrate the results of Lemma 2 via examples taken from [4], [13].

¹We make a change of variable $N - 1 = d$ to align our cost function $J_{\gamma,d}$ with the one in [4].

²There is a slight abuse of notation, since \bar{d} and $d_{[4]}$ are supposed to be integers.

Example 1: Consider the following discrete cubic integrator as in [4]

$$\begin{aligned} x_1^+ &= x_1 + u \\ x_2^+ &= x_2 + u^3 \end{aligned} \quad (13)$$

where $(x_1, x_2) = x \in \mathbb{R}^2$ and $u \in \mathcal{U} = \mathbb{R}$. We take $\sigma(x) = |x_1|^3 + |x_2|$ and $\ell(x, u) = \sigma(x) + |u|^3$. From [4], Assumption 1 holds with $\bar{\alpha}_V = 14\mathbb{I}$, $\bar{\alpha}_W = 0$, $\alpha_W = \mathbb{I}$. Invoking Corollary 1 with $\gamma = 1$, we find $\bar{d} = \lfloor \frac{0 - \ln 14^2}{\ln 13 - \ln 14} \rfloor = 71$. In [4], when no terminal cost is considered, $d_{[4]} = 196$. Thus our new analysis provides a 63% improvement compared to [4]. \blacksquare

Example 2: Consider the following discrete nonholonomic integrator as in [4]

$$\begin{aligned} x_1^+ &= x_1 + u_1 \\ x_2^+ &= x_2 + u_2 \\ x_3^+ &= x_3 + x_1 u_2 - x_2 u_1 \end{aligned} \quad (14)$$

where $(x_1, x_2, x_3) = x \in \mathbb{R}^3$ and $(u_1, u_2) = u \in \mathcal{U} = \mathbb{R}^2$. We take $\sigma(x) = x_1^2 + x_2^2 + 10|x_3|$ and $\ell(x, u) = \sigma(x)$. From [4], Assumption 1 with $\bar{\alpha}_V = \frac{16}{5}\mathbb{I}$, $\bar{\alpha}_W = 0$ and $\alpha_W = \mathbb{I}$. Invoking Corollary 1 with $\gamma = 1$, we obtain $\bar{d} = \lfloor \frac{2 * (\log 5 - \log 16)}{\log 11 - \log 16} \rfloor = 6$. In [4], when no terminal cost is considered, $d_{[4]} = 11$. Thus our new analysis provides a 36% improvement compared to [4]. \blacksquare

Our analysis shows promising relaxation of the horizon d compared to [4], which is important for computation. Yet our results do not supersede [4] in the case when a terminal cost is used. For example, in [4], the minimum horizon when considering terminal cost is $d_{[4]} = 3$ for both examples, and a subsequent work [16] has found $d_{[4]} = 1$ for system (13) and $d_{[4]} = 0$ for system (14), albeit with different stage cost ℓ and measure σ . As explained in the introduction, we have a special interest in the case where no terminal cost is used, since this corresponds to the usual application of value iteration to infinite-horizon problems. Thus, extending Theorem 1 to also consider terminal costs is left for future work.

We now compare our minimum discount $\bar{\gamma}$ to the one found in [13], illustrating a case in which item (i) of Lemma 2 ensures an improvement.

Example 3: Consider system (14) with $\sigma(x) = x_1^2 + x_2^2 + 10|x_3|$ and $\ell(x, u) = \sigma(x) + |u|^2$, for $x \in \mathbb{R}^3$ and $u \in \mathbb{R}^2$. This slightly different stage cost is considered in [13] to show

that SA holds in this case. Assumption 1 is satisfied with $\bar{\alpha}_V = \frac{22}{5}\mathbb{I}$, $\bar{\alpha}_W = 0$ and $\alpha_W = \mathbb{I}$. Note that $\bar{a}_W < a_W$, so item (i) of Lemma 2 applies. Invoking Corollary 1 with $d = \infty$, we calculate $\bar{\gamma} = 1 - \frac{5}{22} = 0.773$. In³ [13], we find that $\gamma_{[13]} = \frac{22}{27} = 0.815$. As expected from item (i) of Lemma 2, our new analysis provides an improvement of about 6% compared to the bound given in [4]. ■

IV. PROOF OF THEOREM 1

We distinguish three cases depending on the value of γ and d .

Case 1: $\gamma \in (0, 1)$ and $d \in \mathbb{Z}_{>0}$.

Let $\gamma \in (0, 1)$, $d \in \mathbb{Z}_{>0}$, $x \in \mathbb{R}^n$ and $v \in F_{\gamma,d}^*(x)$. There exists $[u_0^*, u_1^*, \dots, u_d^*] = \mathbf{u}_{\gamma,d}^*(x)$ such that $v = f(x, u_0^*)$ and $\mathbf{u}_{\gamma,d}^*(x)$ is an optimal input sequence for system (1) with cost (2). Hence $V_{\gamma,d}(x) = J_{\gamma,d}(x, \mathbf{u}_{\gamma,d}^*(x))$.

The proof can be summarized as follows. We first prove item (ii) of Theorem 1. For this, we first upper-bound $V_{\gamma,d}(v)$ using Bellman-like properties. We then derive a preliminary upper-bound on $V_{\gamma,d}(v) - V_{\gamma,d}(x)$. We define $Y_{\gamma,d} = V_{\gamma,d} + W$ where W comes from Assumption 1 and we derive the desired result.

We will now prove item (ii) of Theorem 1. Since stage cost ℓ is nonnegative and in view of item (i) of Assumption 1,

$$\ell(x, u_0^*) \leq V_{\gamma,d}(x) \leq \bar{\alpha}_V(\sigma(x)). \quad (15)$$

Let $j \in \{1, \dots, d\}$. Consider the sequence $\hat{\mathbf{u}} := [u_1^*, u_2^*, \dots, u_{d-j}^*, \hat{\mathbf{u}}_j]$ where $\hat{\mathbf{u}}_j := \mathbf{u}_{\gamma,j}^*(\phi(d-j+1, x, \mathbf{u}_{\gamma,d}^*(x)|_{d-j+1}))$, $\mathbf{u}_{\gamma,d}^*(x)|_{d-j+1} = [u_0^*, \dots, u_{d-j}^*]$ and ϕ denotes the solution of system (1). The sequence $\hat{\mathbf{u}}$ consists of the first $d-j$ elements of $\mathbf{u}_{\gamma,d}^*(x)$ after u_0^* , followed by an optimal input sequence of length $j+1$ for cost $J_{\gamma,d-j}$ at state $\phi(d-j+1, x, \mathbf{u}_{\gamma,d}^*(x)|_{d-j+1})$. Note that the sequence $\hat{\mathbf{u}}_j$ exist and minimizes $J_{\gamma,j}(\phi(d-j+1, x, \mathbf{u}_{\gamma,d}^*(x)|_{d-j+1}), \hat{\mathbf{u}}_j)$ from SA. From the definition of cost $J_{\gamma,d}$ in (2) and $V_{\gamma,d}$ in (3), $V_{\gamma,d}(v) \leq J_{\gamma,d}(v, \hat{\mathbf{u}}) = J_{\gamma,d-j-1}(v, \hat{\mathbf{u}}|_{d-j}) + \gamma^{d-j} J_{\gamma,j}(\phi(d-j, v, \hat{\mathbf{u}}|_{d-j}), \hat{\mathbf{u}}_j)$. By definition of $\hat{\mathbf{u}}_j$, $V_{\gamma,d}(v) \leq \sum_{k=0}^{d-j-1} \gamma^k \ell(\phi(k, v, \hat{\mathbf{u}}|_k), \hat{\mathbf{u}}_k) + \gamma^{d-j} V_{\gamma,j}(\phi(d-j, v, \hat{\mathbf{u}}|_{d-j}))$. For any $k \in \{0, \dots, d\}$, $\phi(k, v, \hat{\mathbf{u}}|_k) = \phi(k+1, x, [u_0^*, \hat{\mathbf{u}}]_{k+1}) = \phi(k+1, x, \mathbf{u}_{\gamma,d}^*(x)|_{k+1})$. Similarly, since $j \in \{1, \dots, d\}$ implies $0 \leq d-j < d$, $\phi(d-j, v, \hat{\mathbf{u}}|_{d-j}) = \phi(d-j+1, x, \mathbf{u}_{\gamma,d}^*(x)|_{d-j+1})$. Thus $V_{\gamma,d}(v) \leq \sum_{k=0}^{d-j-1} \gamma^k \ell(\phi(k+1, x, \mathbf{u}_{\gamma,d}^*(x)|_{k+1}), [\mathbf{u}_{\gamma,d}^*(x)]_{k+1}) + \gamma^{d-j} V_{\gamma,j}(\phi(d-j+1, x, \mathbf{u}_{\gamma,d}^*(x)|_{d-j+1}))$, where $[\mathbf{u}_{\gamma,d}^*(x)]_{k+1} = u_{k+1}^*$. Using the following shorthand notation, we define the *optimal solution* $\phi_k^* := \phi(k, x, \mathbf{u}_{\gamma,d}^*(x))$ and the *optimal stage cost* $\ell_k^* := \ell(\phi_k^*, u_k^*)$ for $k \in \{0, \dots, d\}$. Hence,

$$V_{\gamma,d}(v) = V_{\gamma,d}(\phi_1^*) \leq \sum_{k=0}^{d-j-1} \gamma^k \ell_{k+1}^* + \gamma^{d-j} V_{\gamma,j}(\phi_{d-j+1}^*). \quad (16)$$

³The calculated value in [13] is misstated as $\frac{17}{22} = 0.88$.

Furthermore from the definition of $V_{\gamma,d}(x)$,

$$V_{\gamma,d}(x) = \sum_{k=0}^d \gamma^k \ell_k^* \quad (17)$$

and

$$V_{\gamma,d}(x) - \ell_0^* \geq \gamma \sum_{k=0}^{d-j-1} \gamma^k \ell_{k+1}^*. \quad (18)$$

Subtracting (17) from (16), it follows $V_{\gamma,d}(v) - V_{\gamma,d}(x) \leq -\sum_{k=0}^d \gamma^k \ell_k^* + \sum_{k=0}^{d-j-1} \gamma^k \ell_{k+1}^* + \gamma^{d-j} V_{\gamma,j}(\phi_{d-j+1}^*) = -\ell_0^* + (1-\gamma) \sum_{k=0}^{d-j-1} \gamma^k \ell_{k+1}^* + \gamma^{d-j} V_{\gamma,j}(\phi_{d-j+1}^*)$. In view of (18), we have $\sum_{k=0}^{d-j-1} \gamma^k \ell_{k+1}^* \leq \frac{V_{\gamma,d}(x) - \ell_0^*}{\gamma}$, hence

$$\begin{aligned} V_{\gamma,d}(v) - V_{\gamma,d}(x) &\leq -\ell_0^* - \frac{1-\gamma}{\gamma} \ell_0^* + \frac{1-\gamma}{\gamma} V_{\gamma,d}(x) + \gamma^{d-j} V_{\gamma,j}(\phi_{d-j+1}^*) \\ &= -\frac{\ell_0^*}{\gamma} + \frac{1-\gamma}{\gamma} V_{\gamma,d}(x) + \gamma^{d-j} V_{\gamma,j}(\phi_{d-j+1}^*). \end{aligned} \quad (19)$$

From item (i) of Assumption 1,

$$\begin{aligned} V_{\gamma,d}(v) - V_{\gamma,d}(x) &\leq -\frac{\ell_0^*}{\gamma} + \frac{1-\gamma}{\gamma} V_{\gamma,d}(x) + \gamma^{d-j} \bar{\alpha}_V(\sigma(\phi_{d-j+1}^*)) \\ &\leq \frac{1}{\gamma} \left(-\ell_0^* + (1-\gamma) V_{\gamma,d}(x) + \gamma^{d-j+1} \bar{\alpha}_V(\sigma(\phi_{d-j+1}^*)) \right). \end{aligned}$$

Adding and subtracting $\frac{1-\gamma}{\gamma} W(x)$,

$$\begin{aligned} V_{\gamma,d}(v) - V_{\gamma,d}(x) &\leq \frac{1}{\gamma} \left(-\ell_0^* - (1-\gamma)W(x) + (1-\gamma)(V_{\gamma,d}(x) + W(x)) \right. \\ &\quad \left. + \gamma^{d-j+1} \bar{\alpha}_V(\sigma(\phi_{d-j+1}^*)) \right). \end{aligned} \quad (20)$$

We define $Y_{\gamma,d} := V_{\gamma,d} + W$. In view of item (ii) of Assumption 1 and since $\gamma \leq 1$, $\gamma W(v) - W(x) \leq -\alpha_W(\sigma(x)) + \ell_0^*$. Dividing everything by γ , and since $\frac{1}{\gamma} = 1 + \frac{1-\gamma}{\gamma}$, $W(v) - W(x) - \frac{1-\gamma}{\gamma} W(x) \leq -\frac{\alpha_W(\sigma(x))}{\gamma} + \frac{\ell_0^*}{\gamma}$. Therefore,

$$W(v) - W(x) \leq \frac{1}{\gamma} \left(-\alpha_W(\sigma(x)) + \ell_0^* + (1-\gamma)W(x) \right). \quad (21)$$

In view of (20) and (21),

$$\begin{aligned} Y_{\gamma,d}(v) - Y_{\gamma,d}(x) &\leq \frac{1}{\gamma} \left(-\ell_0^* - (1-\gamma)W(x) + (1-\gamma)Y_{\gamma,d}(x) \right. \\ &\quad \left. + \gamma^{k^*} \bar{\alpha}_V(\sigma(\phi_{k^*}^*)) - \alpha_W(\sigma(x)) + \ell_0^* + (1-\gamma)W(x) \right) \\ &= \frac{1}{\gamma} \left(-\alpha_W(\sigma(x)) + (1-\gamma)Y_{\gamma,d}(x) + \gamma^{k^*} \bar{\alpha}_V(\sigma(\phi_{k^*}^*)) \right) \end{aligned} \quad (22)$$

where $k^* = d-j+1$. A major difficulty compared to [4] is how to bound $\sigma(\phi_{k^*}^*)$ in (22) because of the discount factor. For this purpose, we use Theorem 3 given in the Appendix. From item (ii) of Theorem 3, it follows for any $k \in \{0, \dots, d-1\}$ that $Y_{\gamma,d-(k+1)}(\phi_{k+1}^*) - Y_{\gamma,d-k}(\phi_k^*) \leq \frac{1}{\gamma} \left(-\alpha_Y(\sigma(\phi_k^*)) + (1-\gamma)Y_{\gamma,d-k}(\phi_k^*) \right)$. We write $Y_{\gamma,d-(k+1)}(\phi_{k+1}^*) \leq$

$\left(\frac{-\alpha_Y \circ \bar{\alpha}_Y^{-1} + \mathbb{I}}{\gamma}\right) (Y_{\gamma, d-k}(\phi_k^*))$ where $\bar{\alpha}_Y = \bar{\alpha}_V + \bar{\alpha}_W$ and $\alpha_Y = \alpha_W$, since $\alpha_Y(\sigma(\phi_k^*)) \geq \alpha_Y(\bar{\alpha}_Y^{-1}(Y_{\gamma, d-k}(\phi_k^*)))$ follows from item (i) of Theorem 3. From item (iii) of Assumption 1, $\mathbb{I} - \alpha_Y \circ \bar{\alpha}_Y^{-1} \in \mathcal{K}_\infty$. Thus, starting from $Y_{\gamma, d}(x)$ and proceeding by iteration, we have $Y_{\gamma, d-k^*}(\phi_{k^*}^*) \leq \left(\frac{\mathbb{I} - \alpha_Y \circ \bar{\alpha}_Y^{-1}}{\gamma}\right)^{(k^*)} (Y_{\gamma, d}(x))$. We can now apply item (i) of Theorem 3 and conclude that

$$\sigma(\phi_{k^*}^*) \leq \bar{\alpha}_Y^{-1} \left(\left(\frac{\mathbb{I} - \alpha_Y \circ \bar{\alpha}_Y^{-1}}{\gamma} \right)^{(k^*)} (Y_{\gamma, d}(x)) \right). \quad (23)$$

It follows from (22) and (23) that $Y_{\gamma, d}(v) - Y_{\gamma, d}(x) \leq \frac{1}{\gamma} \left(-\alpha_W(\sigma(x)) + (1 - \gamma)Y_{\gamma, d}(x) + \gamma^{k^*} \bar{\alpha}_V \circ \bar{\alpha}_Y^{-1} \circ \left(\frac{\mathbb{I} - \alpha_Y \circ \bar{\alpha}_Y^{-1}}{\gamma} \right)^{(k^*)} (Y_{\gamma, d}(x)) \right)$. Thus equation (10) of Theorem 1 is verified with $\alpha_Y = \alpha_W \in \mathcal{K}_\infty$ and $\Upsilon(s, \gamma, k^*) = (1 - \gamma)s + \gamma^{k^*} \bar{\alpha}_V \circ \bar{\alpha}_Y^{-1} \circ \left(\frac{\mathbb{I} - \alpha_Y \circ \bar{\alpha}_Y^{-1}}{\gamma} \right)^{(k^*)} (s)$. Recall that j is freely selected in $\{1, \dots, d\}$, as a result so $k^* \in \{1, \dots, d\}$. Note that $\Upsilon(s, \gamma, d) \geq 0$ for any $s \geq 0$, as $0 \leq \mathbb{I} - \alpha_Y \circ \bar{\alpha}_Y^{-1}$, which follows from $\alpha_Y = \alpha_Y \leq \bar{\alpha}_Y$. Let $s \geq 0$, consider $(1 - \gamma)s + \bar{\alpha}_V \left(\bar{\alpha}_Y^{-1} \left(\left(\frac{\mathbb{I} - \alpha_Y \circ \bar{\alpha}_Y^{-1}}{\gamma} \right)^{(d)} (s) \right) \right) \geq \Upsilon(s, \gamma, d)$. Note that $s - \alpha_Y \circ \bar{\alpha}_Y^{-1}(s) < s$ if $s \neq 0$. Indeed, suppose $s - \alpha_Y \circ \bar{\alpha}_Y^{-1}(s) = s$ and $s \neq 0$, this is only possible if $\alpha_Y \circ \bar{\alpha}_Y^{-1}(s) = 0$, we attain a contradiction. Hence, $s - \alpha_Y \circ \bar{\alpha}_Y^{-1}(s) < s$ when $s > 0$, and zero at zero. Therefore $(1 - \gamma)s + \bar{\alpha}_V \left(\bar{\alpha}_Y^{-1} \left(\left(\frac{\mathbb{I} - \alpha_Y \circ \bar{\alpha}_Y^{-1}}{\gamma} \right)^{(d)} (s) \right) \right) \rightarrow 0$ when $\gamma \rightarrow 1$ and $d \rightarrow \infty$. Finally, recall that $0 \leq \Upsilon(s, \gamma, d)$. It follows, by the sandwich rule, that $\Upsilon(s, \gamma, d) \rightarrow 0$ when $\gamma \rightarrow 1$ and $d \rightarrow \infty$. Hence, item (ii) of Theorem 1 holds.

In view of Assumption 1, $Y_{\gamma, d} \leq \bar{\alpha}_Y(\sigma(x))$ with $\bar{\alpha}_Y = \bar{\alpha}_V + \bar{\alpha}_W \in \mathcal{K}_\infty$. From item (ii) of Assumption 1, we have $W(x) \geq \alpha_Y(\sigma(x)) - \ell(x, u_0^*)$. Associated with (15), it follows $Y_{\gamma, d} \geq \alpha_W(\sigma(x)) - \ell(x, u_0^*) + \ell(x, u_0^*) = \alpha_W(\sigma(x))$. Thus $\bar{\alpha}_Y = \alpha_W \in \mathcal{K}_\infty$. Item (i) of Theorem 1 is satisfied.

Case 2: $\gamma = 1$ and $d \in \mathbb{Z}_{>0}$

By following the steps of Case 1 with $\gamma = 1$, the desired result is obtained.

Case 3: $\gamma \in (0, 1)$ and $d = \infty$

Let $\gamma \in (0, 1)$, $x \in \mathbb{R}^n$ and $v \in F_{\gamma, \infty}^*(x)$. From Bellman principle, it follows that $V_{\gamma, \infty}(x) = \ell_0^* + \gamma V_{\gamma, \infty}(v)$, thus $V_{\gamma, \infty}(v) = \frac{-\ell_0^* + V_{\gamma, \infty}(x)}{\gamma}$. By following the steps of Case 1, the desired result is obtained.

V. CONCLUSION

We have analyzed the stability of general nonlinear discrete-time systems controlled by a sequence of inputs that minimizes a discounted finite horizon cost and is computed in a receding-horizon fashion. In general, uniform semiglobal practical stability is ensured under stabilizability and detectability conditions. Under additional assumptions, uniform global exponential stability is guaranteed. The Lyapunov function used to prove stability is shown to be continuous under extra assumptions, hence endowing stability with some nominal robustness. We compare our results with previous works of the literature on undiscounted finite-horizon

and discounted infinite-horizon optimal control, respectively. A smaller minimum discount for infinite cost condition, compared to the discount from [13], is given for certain cases. A relationship between the minimum horizon from [4] and ours for undiscounted cost is provided. We show that in the examples considered by these references, less conservative bounds can be obtained.

VI. APPENDIX

The satisfaction of Assumption 1 allows to derive the next Lyapunov properties for optimal sequences, which are used in the proof of Theorem 1.

Theorem 3: Suppose Assumption 1 is verified. There exist $\bar{\alpha}_Y, \bar{\alpha}_V, \alpha_Y \in \mathcal{K}_\infty$, such that for any $\gamma \in (0, 1]$, $d \in \mathbb{Z}_{>0}$ and $k \in \{0, \dots, d\}$, there exists $Y_{\gamma, d-k} : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ such that the following holds.

- (i) For any $x \in \mathbb{R}^n$, $\alpha_Y(\sigma(x)) \leq Y_{\gamma, d-k}(x) \leq \bar{\alpha}_Y(\sigma(x))$.
- (ii) For any $x \in \mathbb{R}^n$, $Y_{\gamma, d-(k+1)}(\phi_{k+1}^*) - Y_{\gamma, d-k}(\phi_k^*) \leq \frac{1}{\gamma} \left(-\alpha_Y(\sigma(\phi_k^*)) + (1 - \gamma)Y_{\gamma, d-k}(\phi_k^*) \right)$, where $\phi_{k+1}^* \in F_{\gamma, d-k}^*(\phi_k^*)$ and $\phi_0^* = x$. \square

REFERENCES

- [1] D. P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 2. Athena Scientific, Nashua, USA, 4th edition, 2012.
- [2] D. Ernst, P. Geurts, and L. Wehenkel. Tree-based batch mode reinforcement learning. *J. of Machine Learning Research*, 6:503–556, 2005.
- [3] V. Gaitsgory, L. Grüne, M. Höger, C. M. Kellett, and S. R. Weller. Stabilization of strictly dissipative discrete time systems with discounted optimal control. *Automatica*, 93:311 – 320, 2018.
- [4] G. Grimm, M. J. Messina, S. E. Tuna, and A. R. Teel. Model predictive control: for want of a local control Lyapunov function, all is not lost. *IEEE Transactions on Automatic Control*, 50(5):546–558, 2005.
- [5] L. Grüne. Computing stability and performance bounds for unconstrained nmpc schemes. In *IEEE Conference on Decision and Control*, pages 1263–1268, New Orleans, USA, 2007.
- [6] J.-F. Hren and R. Munos. Optimistic planning of deterministic systems. In *European Workshop on Reinforcement Learning*, pages 151–164, Villeneuve d’Ascq, France, 2008. Springer.
- [7] Z.-P. Jiang and Y. Wang. Input-to-state stability for discrete-time nonlinear systems. *Automatica*, 37(6):857–869, 2001.
- [8] S. Keerthi and E. Gilbert. An existence theorem for discrete-time infinite-horizon optimal control problems. *IEEE Transactions on Automatic Control*, 30(9):907–909, 1985.
- [9] C. M. Kellett and A. R. Teel. On the robustness of \mathcal{KL} -stability for difference inclusions: Smooth discrete-time Lyapunov functions. *SIAM Journal on Control and Optimization*, 44(3):777–800, 2005.
- [10] A. Lazaric, M. Ghavamzadeh, and R. Munos. Finite-sample analysis of least-squares policy iteration. *J. of Machine Learning Research*, 13:3041–3074, 2012.
- [11] R. Munos. Error bounds for approximate policy iteration. In *Int. Conf. on Machine Learning*, pages 560–567, Washington, US, 2003.
- [12] R. Munos and C. Szepesvári. Finite time bounds for fitted value iteration. *J. of Machine Learning Research*, 9:815–857, 2008.
- [13] R. Postoyan, L. Buşoniu, D. Nešić, and J. Daafouz. Stability analysis of discrete-time infinite-horizon optimal control with discounted cost. *IEEE Transactions on Automatic Control*, 62(6):2736–2749, 2017.
- [14] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.
- [15] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, USA, 2nd edition, 2017.
- [16] S. E. Tuna, M. J. Messina, and A. R. Teel. Shorter horizons for model predictive control. In *American Control Conference*, pages 863–868, Minneapolis, USA, 2006.
- [17] C. J. C. H. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8:279–292, 1992.
- [18] M. Wiering and M. van Otterlo, editors. *Reinforcement Learning: State of the Art*, volume 12. Springer, 2012.