On the differentiability of the value function of switched linear systems under arbitrary and controlled switching

Guillaume O. Berger[1]

***Abstract*— This paper studies the differentiability of the value function of switched linear systems under arbitrary switching and controlled switching, referred to as worst-case and optimal value functions respectively. First, we show that the value functions are Lipschitz continuous, when the cost function is Lipschitz continuous. Then, as the central contribution of this work, we show with examples that each of these functions can be non-differentiable on dense subsets of the state space, even if the cost function is smooth and Lipschitz continuous. This has implications for optimal control and reinforcement learning since it implies that the exact computation of these value functions requires templates involving functions that are non-differentiable on dense subsets.**

I. Introduction

Switched linear systems are multi-modal dynamical systems wherein each mode is a linear system. These systems appear naturally in a wide range of applications—e.g., mechanical systems with impact, digital power converters, cyber-physical systems, etc.—and as abstractions of more complex dynamical or hybrid systems [1]–[3]. The stability and stabilizability theory of switched linear systems has been extensively studied: despite intrinsic challenges (undecidability, NP-hardness, non-algebraicity, etc. of computing stabilizing controllers or stability certificates) [2], [4], several powerful tools (e.g., based on piecewise, multiple or sum-of-squares Lyapunov functions) have been provided for the stability analysis and stabilization of switched linear systems; see, e.g., [3], [5], [6] for recent surveys.

In this paper, we first focus on the problem of optimal control of switched linear systems under controlled switching, as well as the worst-case cost analysis of switched linear systems under arbitrary switching. This means that we consider a cost function mapping states to cost values, and we aim to optimize the *cost-to-go* of any trajectory of the system by choosing the switching law appropriately, or estimating the worst-case *cost-to-go* of any trajectory of the system under all possible switching sequences. Here, the *cost-to-go* refers the sum of all state costs of the trajectory. This problem has received a lot of attention in the literature [7]–[15], due its importance in applications and its challenging nature. In particular, [7], [9], [10], [12], [15] propose efficient techniques to approximate the *optimal value function* (the function mapping states of the state space to the smallest cost-to-go among all trajectories starting from them), and [14] proposes a polynomial-time algorithm for computing this function under additional assumptions on the system. Nevertheless, the question of the complexity of the optimal value function has not been properly addressed in the literature yet (see "related work" below). This question is however crucial if one aims to leave the realm of (efficient or not) approximate methods.

This paper provides a pessimistic answer to the above question. Indeed, after showing that the optimal and worst-case value functions are Lipschitz continuous if the cost function is Lipschitz continuous, we show that it may however be non-differentiable on a dense subset of the state space, even if the cost function is smooth (e.g., quadratic). This finding is deceptive for the objective of exact computation of the value functions of switched linear systems because it implies that iterative algorithms have almost zero chance of finding them in finite time (since they usually deal with piecewise smooth functions). We further show that this negative result holds also for systems whose matrices have rational entries, so that it holds with nonzero probability when working with matrices encoded with finite precision. We obtained these results by building a switched linear system for which the optimal (or worst-case) switching law leads to a closed-loop system that is shown to be an *Interval Exchange Map* [16]. This well-studied class of dynamical systems is known for their chaotic behavior, a property that we leverage to obtain the non-differentiability on dense subsets.

Related work

The paper [8] shows that the *finite-horizon* optimal value function of the linear quadratic discrete-time switched LQR problem is piecewise quadratic. It also shows that the finite-horizon optimal value function converges exponentially to the (infinite-horizon) optimal value function when the horizon tends to infinity. However, it does not discuss whether the sequence of piecewise quadratic functions converges toward a piecewise differentiable function or not.

The paper [17] discusses the continuity and smoothness of the value function for autonomous dynamical systems. In particular, conditions are given under which the value function is Lipschitz continuous, and it is shown that there are systems for which the value function is nowhere differentiable. These results differ from ours because 1) they focus on autonomous systems, 2) they assume a discount factor in the cost-to-go, and 3) they rely on a nonlinear system to prove the non-differentiability. Nevertheless, our proof of the Lipschitz continuity of the value functions of switched linear systems uses similar arguments to that in [17]. This result, which is not the main contribution of this work, is used in the proof of our main result on the possible non-

[1]G. Berger is with ICTEAM, UCLouvain, Belgium. He is an FNRS postdoctoral researcher. guillaume.berger@uclouvain.be

differentiability of the value functions of switched linear systems on dense subsets. The proof of the latter is radically different from that of the akin result in [17]. In particular, we stress that the value functions of switched linear systems are differentiable almost everywhere (consequence of being Lipschitz continuous, by Rademacher's theorem [18]), which is a striking difference with the example in [17]. Our proof relies on the existence of a switched linear system for which the optimal or worst-case switching law produces a closed-loop system called an *Interval Exchange Map*, which is known to be chaotic [16]; a fact that we leverage to obtain the non-smoothness.

Notation. $\mathbb{N}$ is the set of nonnegative integers. Given $m \in \mathbb{N}_{>0}$, $[m]$ is the set $\{1, \dots, m\}$. We denote the Euclidean vector norm by $\|\cdot\|$ (i.e., $\|x\|^2 = x^\top x$).

II. Problem statement

A. Switched linear systems

We consider a discrete-time switched linear system of the form:

$$\xi(t+1) = A_{\sigma(t)}\xi(t), \quad t \in \mathbb{N}, \tag{1}$$

where $\xi(t) \in \mathbb{R}^n$ is the *state* at time t, $\sigma(t) \in [m]$ is the *mode* at time t, and for all $i \in [m]$, $A_i \in \mathbb{R}^{n\times n}$ represents the *transition matrix* of mode i. The function $\sigma : \mathbb{N} \to [m]$ mapping times to modes is called the *switching signal*, which can be arbitrary or controlled. The system (1) is denoted by $\{A_i\}_{i=1}^m$. Given $x \in \mathbb{R}^n$, $\sigma : \mathbb{N} \to [m]$ and $t \in \mathbb{N}$, we denote by $\xi(t, x, \sigma)$ the solution at time t of (1) with signal σ and initial state $\xi(0) = x$.

In this paper, we restrict our attention to *stable* switched linear systems. This means that the *joint spectral radius* [2], which represents the worst-case linear rate of growth of the trajectories of the system, is smaller than one. Formally, the joint spectral radius of $\mathcal{A} \coloneqq \{A_i\}_{i=1}^m$ is defined by

$$\begin{aligned} \operatorname{jsr}(\mathcal{A}) = \inf\{r \geq 0 : \exists\, C \geq 0 \text{ s.t.} \\ \forall\, \xi \text{ solution of (1)}, \\ \forall\, t \in \mathbb{N},\ \|\xi(t)\| \leq C r^t \|\xi(0)\|\}, \end{aligned}$$

and we make the standing assumption that $\operatorname{jsr}(\{A_i\}_{i=1}^m) < 1$. The following result will be instrumental in our proofs:

Theorem 1 ([2]): Let $\mathcal{A} \coloneqq \{A_i\}_{i=1}^m$ be a switched linear system and $\rho > \operatorname{jsr}(\mathcal{A})$. There is a norm $\|\cdot\|_*$ such that for all solutions ξ of (1) and all $t \in \mathbb{N}$, it holds that $\|\xi(t)\|_* \leq \rho^t \|\xi(0)\|_*$.

Given a norm $\|\cdot\|'$, we denote its *eccentricity* by

$$\kappa(\|\cdot\|') = \frac{\max_{\|x\|=1} \|x\|'}{\min_{\|x\|=1} \|x\|'}.$$

B. Optimal and worst-case value functions

We let $c : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ be a cost function mapping states to cost values. Given a signal $\sigma : \mathbb{N} \to [m]$ and $x \in \mathbb{R}^n$, the *cost-to-go* of the trajectory starting from x with signal σ is defined by

$$J(x, \sigma) = \sum_{t=0}^{\infty} c(\xi(t, x, \sigma)).$$

In the context of controlled switching, the goal is to find the switching signal (which may depend on x) such that the associated cost-to-go is minimal. Hence, we define the *optimal value function* as

$$J^\star(x) = \inf_{\sigma:\mathbb{N}\to[m]} \sum_{t=0}^{\infty} c(\xi(t, x, \sigma)).$$

This can be equivalently seen as the smallest cost-to-go among all trajectories starting from x. On the other hand, in the context of arbitrary switching, we are interested in the largest cost-to-go of a trajectory starting from a given x. Hence, we define the *worst-case value function* as

$$J^\circ(x) = \sup_{\sigma:\mathbb{N}\to[m]} \sum_{t=0}^{\infty} c(\xi(t, x, \sigma)).$$

In this paper, we first show that $J^\star$ and J° are Lipschitz continuous when c is Lipschitz continuous Then, we show that they can be non-differentiable on dense subsets of the state space, even if c is differentiable everywhere.

The following well-known *dynamic programming equations* (also called *Bellman equations*; see, e.g., [19, Theorem 3.1]) will be useful in the proofs:

Proposition 1: For a switched linear system $\{A_i\}_{i=1}^m$ and a cost function c, let $J^\star$ and J° be the associated optimal and worst-case value functions respectively. It holds that

$$J^\star(x) = c(x) + \min_{i\in[m]} J^\star(A_i x), \tag{2}$$

$$J^\circ(x) = c(x) + \max_{i\in[m]} J^\circ(A_i x). \tag{3}$$

III. Lipschitz continuity of the value functions

In this section, we show that the optimal and worst-case value functions are Lipschitz continuous if the cost function is Lipschitz continuous. As a reminder, a function $f : \mathbb{R}^n \to \mathbb{R}$ is *Lipschitz continuous* on a set $X \subseteq \mathbb{R}^n$ with Lipschitz constant L if for all $x, y \in X$, it holds that

$$|f(x) - f(y)| \leq L\|x - y\|.$$

Theorem 2: Let $\mathcal{A} \coloneqq \{A_i\}_{i=1}^m \subseteq \mathbb{R}^{n\times n}$ be a switched linear system and $c : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ a cost function. Let $J^\star$ and J° be the associated optimal and worst-case value functions respectively. Assume that $\operatorname{jsr}(\mathcal{A}) < 1$ and c is Lipschitz continuous on a neighborhood $X \subseteq \mathbb{R}^n$ of the origin, with Lipschitz constant L. Let $\rho < 1$ and $\|\cdot\|_*$ be as in Theorem 1. Let $\eta > 0$ be such that $B_*(\eta) \coloneqq \{x \in \mathbb{R}^n : \|x\|_* \leq \eta\} \subseteq X$. It holds that $J^\star$ and J° are Lipschitz continuous on $B_*(\eta)$, with Lipschitz constant $M = \kappa(\|\cdot\|_*)L/(1-\rho)$.

Proof: We do the proof for $J^\star$, and it extends straightforwardly to J°. Denote $r = \kappa(\|\cdot\|_*)$ and $M = rL/(1-\rho)$. Let $x, y \in B_*(\eta)$. Let $\epsilon > 0$. We will show that $J^\star(x) \leq J^\star(y) + M\|x - y\| + \epsilon$. Since x, y and ϵ are arbitrary, this will conclude the proof. Let σ be a switching signal such that $J(y, \sigma) \leq J^\star(y) + \epsilon$. For all $t \in \mathbb{N}$, it holds that

$$\|\xi(t, x, \sigma) - \xi(t, y, \sigma)\|_* = \|\xi(t, x - y, \sigma)\|_* \leq \rho^t \|x - y\|_*.$$

Hence, it holds that for all $t \in \mathbb{N}$,

$$\|\xi(t,x,\sigma) - \xi(t,y,\sigma)\| \leq r\rho^t \|x-y\|.$$

Furthermore, for all $t \in \mathbb{N}$ and $z \in \{x, y\}$,

$$\|\xi(t,z,\sigma)\|_* \leq \rho^t \|z\|_* \leq \eta.$$

Thus, by the assumption on c, we obtain that for all $t \in \mathbb{N}$,

$$c(\xi(t,x,\sigma)) \leq c(\xi(t,y,\sigma)) + r\rho^t L \|x-y\|.$$

It follows that

$$\begin{aligned} J(x,\sigma) &= \sum_{t=0}^{\infty} c(\xi(t,x,\sigma)) \\ &\leq \sum_{t=0}^{\infty} c(\xi(t,y,\sigma)) + \frac{rL}{1-\rho}\|x-y\| \\ &= J(y,\sigma) + \frac{rL}{1-\rho}\|x-y\|. \end{aligned}$$

Hence, we get $J^\star(x) \leq J^\star(y) + \epsilon + M\|x-y\|$, concluding the proof. ■

As a corollary, we obtain that $J^\star$ and J° are differentiable almost everywhere. This is in stark contrast with the example in [17, Proposition 1] where the value function of a nonlinear dynamical system is nowhere differentiable.

Corollary 1: Let $\mathcal{A} \coloneqq \{A_i\}_{i=1}^m \subseteq \mathbb{R}^{n\times n}$ be a switched linear system and $c : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ a cost function. Let $J^\star$ and J° be the associated optimal and worst-case value functions respectively. Assume that $\mathrm{jsr}(\mathcal{A}) < 1$ and c is Lipschitz continuous on a neighborhood of the origin. It holds that $J^\star$ and J° are differentiable almost everywhere (for the Lebesgue measure) on a neighborhood of the origin.

Proof: A standard result in real analysis (see, e.g., [18, Theorem 7.20]) says that absolutely continuous functions are differentiable almost everywhere. ■

IV. Possible non-differentiability of the value functions on dense subsets

In this section, we prove the main result of this paper, stating that there exist switched linear systems for which the optimal or worst-case value function is non-differentiable on a dense subset of the state space, even if the cost function is smooth everywhere. We note that this result is *not* in contradiction with Corollary 1 since there are dense subsets of $\mathbb{R}^n$ that have zero Lebesgue measure (e.g., $\mathbb{Q}^n \subseteq \mathbb{R}^n$).

Theorem 3: Let $n \in \mathbb{N}_{\geq 2}$. There exist a switched linear system $\mathcal{A} \coloneqq \{A_i\}_{i=1}^m \subseteq \mathbb{R}^{n\times n}$ and a cost function $c : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ such that $\mathrm{jsr}(\mathcal{A}) < 1$, c is C^∞ on $\mathbb{R}^n$ and Lipschitz continuous on a neighborhood of the origin, and $J^\star$ and J° are non-differentiable on dense subsets of $\mathbb{R}^n$, where $J^\star$ and J° are the associated optimal and worst-case value functions respectively. Furthermore, the matrices in $\mathcal{A}$ can be assumed to have rational entries, i.e., $\mathcal{A} \subseteq \mathbb{Q}^{n\times n}$.

A. Proof of Theorem 3

The proof consists in building a switched linear system and a cost function that satisfy the requirements of the theorem. First, we consider the case $n = 2$, and then we generalize.

1) Case $n = 2$: The construction for $n = 2$ is as follows. Consider two angles,

$$\alpha \in (\pi/8, 3\pi/8) \ \text{ and } \ \beta = \alpha - \pi/2 \in (-3\pi/8, -\pi/8),$$

such that $\alpha/\pi \notin \mathbb{Q}$. The system consists of two matrices, which are *scaled rotations*, of angle α and β respectively, and scaling factor ρ:

$$A_1 = \rho \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}, \tag{4}$$

$$A_2 = \rho \begin{bmatrix} \cos(\beta) & -\sin(\beta) \\ \sin(\beta) & \cos(\beta) \end{bmatrix}, \tag{5}$$

where $\rho = 0.01$. We denote $\mathcal{A} = \{A_1, A_2\}$. We note that α and β can be chosen so that A_1 and A_2 have rational entries; see Lemma 14 at the end of this section. The following properties of the system will be useful.[1]

Lemma 1: The system $\mathcal{A}$ is stable with $\mathrm{jsr}(\mathcal{A}) = \rho$, and the norm $\|\cdot\|_* = \|\cdot\|$ satisfies the requirements of Theorem 1 with this ρ.

Proof: It holds that for all $x \in \mathbb{R}^2$ and $A \in \mathcal{A}$, $\|Ax\| = \rho\|(A/\rho)x\| = \rho\|x\|$ where we used that A/ρ is a rotation matrix. Hence, for all solution ξ of system $\mathcal{A}$, it holds that $\|\xi(t)\| = \rho^t\|\xi(0)\|$. ■

As for the cost function, we let

$$c(x) = x_1^2 + 2x_2^2, \quad \text{where} \quad x = [x_1, x_2]^\top.$$

The function c is C^∞ on $\mathbb{R}^2$ and Lipschitz continuous on the set $B \coloneqq \{x \in \mathbb{R}^2 : \|x\| \leq 1\}$, with Lipschitz constant $L = 4$. Moreover, c is homogeneous of degree 2, meaning that for all $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^2$, $c(\lambda x) = \lambda^2 c(x)$.

Let $J^\star$ be the optimal value function of $\mathcal{A}$ with cost c. We do the analysis of $J^\star$ below, in particular showing in several steps that it is non-differentiable on a dense subset of $\mathbb{R}^2$. (The worst-case value function will be discussed later, with similar arguments.)

First, by Theorem 2, $J^\star$ is Lipschitz continuous:

Lemma 2: The function $J^\star$ is Lipschitz continuous on B, with Lipschitz constant $M = L/(1-\rho)$.

Proof: Direct from Theorem 2 and Lemma 1. ■

Second, a consequence of c being homogeneous is that $J^\star$ is homogeneous:

Lemma 3: The function $J^\star$ is homogeneous of degree 2, i.e., for all $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^2$, $J^\star(\lambda x) = \lambda^2 J^\star(x)$.

Proof: It suffices to prove that for all $x \in \mathbb{R}^2$, $\lambda \neq 0$ and $\epsilon > 0$, $J^\star(\lambda x) \leq \lambda^2 J^\star(x) + \epsilon$. Therefore, let $x \in \mathbb{R}^2$, $\lambda \neq 0$ and $\epsilon > 0$. Let σ be a switching signal such that $J(x,\sigma) \leq J^\star(x) + \lambda^{-2}\epsilon$. By the linearity, it holds that for all $t \in \mathbb{N}$, $\xi(t,\lambda x,\sigma) = \lambda\xi(t,x,\sigma)$. Hence, $J(\lambda x,\sigma) = \lambda^2 J(x,\sigma)$, so that $J^\star(\lambda x) \leq \lambda^2 J^\star(x) + \epsilon$, concluding the proof. ■

Given $\theta \in \mathbb{R}$, we let $x(\theta) = [\cos(\theta), \sin(\theta)]^\top$. We also let

$$\tilde{c}(\theta) \coloneqq c(x(\theta)) = \cos^2(\theta) + 2\sin^2(\theta) = 1 + \sin^2(\theta).$$

See Fig. 1 for an illustration.

[1] All free variables appearing in the results (lemmas, corollaries, etc.) in the rest of this section refer to quantities defined in the body of the text.

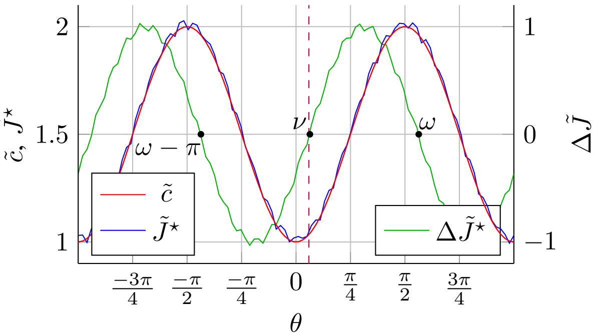

Fig. 1. The function $\tilde{c}$ and a potential plot of $\tilde{J}^\star$ (left y-axis), along with $\Delta\tilde{J}^\star$ (right y-axis) for $\alpha = 0.6$ and $\beta = \alpha - \pi/2$ (μ is the purple line).

Lemma 4: For all $x \in [\pi/8, 3\pi/8] + \pi\mathbb{Z}$, $\tilde{c}'(x) \geq \sqrt{2}/2$, and for all $x \in [-3\pi/8, -\pi/8] + \pi\mathbb{Z}$, $\tilde{c}'(x) \leq -\sqrt{2}/2$.

Proof: Straightforward from $\tilde{c}'(\theta) = \sin(2\theta)$. ■

For each $\theta \in \mathbb{R}$, we let $\Delta\tilde{c}(\theta) \coloneqq \tilde{c}(\theta+\alpha) - \tilde{c}(\theta+\beta)$. We also let $\mu \coloneqq -(\alpha+\beta)/2 = \pi/4 - \alpha = -\pi/4 - \beta$.

Lemma 5: It holds that $\Delta\tilde{c}(\theta) = \sin(2(\theta-\mu))$.

Proof: We have that

$$\begin{aligned}\tilde{c}(\theta+\alpha) &= 1 + \sin^2(\theta + \pi/4 - \mu)\\ &= 1 + (1 - \cos(2(\theta-\mu) + \pi/2))/2\\ &= 1 + (1 + \sin(2(\theta-\mu)))/2.\end{aligned}$$

Similarly,

$$\begin{aligned}\tilde{c}(\theta+\beta) &= 1 + \sin^2(\theta - \pi/4 - \mu)\\ &= 1 + (1 - \cos(2(\theta-\mu) - \pi/2))/2\\ &= 1 + (1 - \sin(2(\theta-\mu)))/2.\end{aligned}$$

Hence, $\Delta\tilde{c}(\theta) = \sin(2(\theta-\mu))$. ■

Finally, for each $\theta \in \mathbb{R}$, we let $\tilde{J}^\star(\theta) \coloneqq J^\star(x(\theta))$.

Lemma 6: For all $\theta \in \mathbb{R}$, it holds that

$$\tilde{J}^\star(\theta) = \tilde{c}(\theta) + \rho^2 \min\{\tilde{J}^\star(\theta+\alpha), \tilde{J}^\star(\theta+\beta)\}.$$

Proof: For any $\theta \in \mathbb{R}$, it holds that $A_1 x(\theta) = \rho x(\theta+\alpha)$ and $A_2 x(\theta) = \rho x(\theta+\beta)$ (because A_1 and A_2 are scaled rotations). Hence, we obtain the conclusion by using (2) and Lemma 3. ■

For each $\theta \in \mathbb{R}$, we let $\Delta\tilde{J}^\star(\theta) \coloneqq \tilde{J}^\star(\theta+\alpha) - \tilde{J}^\star(\theta+\beta)$. Denote $\delta = 0.01$. The following result is *key* (see also Fig. 1 for an illustration):

Lemma 7: There exist

- $\nu \in (\mu-\delta, \mu+\delta)$ such that $\Delta\tilde{J}^\star(\nu) = 0$;
- $\omega \in (\mu+\pi/2-\delta, \mu+\pi/2+\delta)$ such that $\Delta\tilde{J}^\star(\omega) = 0$.

Moreover, $\Delta\tilde{J}^\star$ is

- negative on $(\omega-\pi, \nu) + \pi\mathbb{Z}$;
- positive on $(\nu, \omega) + \pi\mathbb{Z}$.

Proof: First of all, let us remind that by Lemma 2, $\tilde{J}^\star$ is Lipschitz continuous on $\mathbb{R}$, with Lipschitz constant M, and for all $\theta \in \mathbb{R}$, $|\tilde{J}^\star(\theta)| \leq M$, where $M < 4.1$.[2]

[2] Indeed, recall that $L = 4$ and $\rho = 0.01$.

Negativity on the set $S_- \coloneqq [\mu - \pi/2 + \delta, \mu - \delta] + \pi\mathbb{Z}$: By Lemma 5, we have that for all $\theta \in S_-$, $\Delta\tilde{c}(\theta) < -0.019$. Hence, by Lemma 6, we obtain that for all $\theta \in S_-$,

$$\Delta\tilde{J}^\star(\theta) < -0.019 + 2\rho^2 M < -0.018 < 0.$$

Positivity on the set $S_+ \coloneqq [\mu + \delta, \mu + \pi/2 - \delta] + \pi\mathbb{Z}$: By Lemma 5, we have that for all $\theta \in S_+$, $\Delta\tilde{c}(\theta) > 0.019$. Hence, by Lemma 6, we obtain that for all $\theta \in S_+$,

$$\Delta\tilde{J}^\star(\theta) > 0.019 - 2\rho^2 M > 0.018 > 0.$$

Existence of ν: From the above, we know that $\Delta\tilde{J}^\star(\mu - \delta) < 0$ and $\Delta\tilde{J}^\star(\mu+\delta) > 0$. Hence, by continuity of $\Delta\tilde{J}^\star$, there is $\nu \in (\mu-\delta, \mu+\delta)$ such that $\Delta\tilde{J}^\star(\nu) = 0$.

Existence of ω: From the above, we know that $\Delta\tilde{J}^\star(\mu + \pi/2 - \delta) > 0$ and $\Delta\tilde{J}^\star(\mu+\pi/2+\delta) < 0$. Hence, there is $\omega \in (\mu+\pi/2-\delta, \mu+\pi/2+\delta)$ such that $\Delta\tilde{J}^\star(\omega) = 0$.

Strict increase on $S_\nearrow \coloneqq [\mu - \delta, \mu + \delta] + \pi\mathbb{Z}$: We show that $\Delta\tilde{J}^\star$ is increasing on $S_\nearrow$, which will imply that $\Delta\tilde{J}^\star$ is negative on $[\mu-\delta, \nu) + \pi\mathbb{Z}$ and positive on $(\nu, \mu+\delta] + \pi\mathbb{Z}$. Let $k \in \mathbb{Z}$ and $\mu - \delta + k\pi \leq \theta_1 < \theta_2 \leq \mu + \delta + k\pi$. By Lemma 5, we have that $\Delta\tilde{c}(\theta_2) - \Delta\tilde{c}(\theta_1) \geq \theta_2 - \theta_1$ (mean value theorem). Hence, by Lemma 6,

$$\Delta\tilde{J}^\star(\theta_2) - \Delta\tilde{J}^\star(\theta_1) \geq (1 - \rho^2 M)(\theta_2 - \theta_1) > 0.$$

Strict decrease on $S_\searrow \coloneqq [\mu+\pi/2-\delta, \mu+\pi/2+\delta] + \pi\mathbb{Z}$: We show that $\Delta\tilde{J}^\star$ is decreasing on $S_\searrow$, which will imply that $\Delta\tilde{J}^\star$ is positive on $[\mu+\pi/2-\delta, \omega) + \pi\mathbb{Z}$ and negative on $(\omega, \mu+\pi/2+\delta] + \pi\mathbb{Z}$. Let $k \in \mathbb{Z}$ and $\mu+\pi/2-\delta+k\pi \leq \theta_1 < \theta_2 \leq \mu+\pi/2+\delta+k\pi$. By Lemma 5, we have that $\Delta\tilde{c}(\theta_2) - \Delta\tilde{c}(\theta_1) \leq \theta_1 - \theta_2$ (mean value theorem). Hence, by Lemma 6,

$$\Delta\tilde{J}^\star(\theta_2) - \Delta\tilde{J}^\star(\theta_1) \geq (\rho^2 M - 1)(\theta_2 - \theta_1) < 0.$$

This concludes the proof. ■

From now on, we let ν and ω be as Lemma 7. We note that $[\nu+\beta, \nu+\alpha] \subseteq (\omega-\pi, \omega)$.[3] Similarly, $[\mu-\delta, \mu+\delta] \subseteq (\omega-\pi, \omega)$. We show that $\tilde{J}^\star$ is not differentiable at ν.

Lemma 8: The function $\tilde{J}^\star$ is not differentiable at ν.

Proof: By Lemma 6 and $\tilde{c}$ being differentiable, it suffices to show that the function

$$h : \theta \mapsto \min\{\tilde{J}^\star(\theta+\alpha), \tilde{J}^\star(\theta+\beta)\}$$

is not differentiable at ν. For that, we will show that

a) for $\mu - \delta \leq \theta < \nu$, $h(\nu) - h(\theta) \geq \frac{1}{2}(\nu - \theta)$,
b) for $\nu < \theta \leq \mu + \delta$, $h(\theta) - h(\nu) \leq \frac{-1}{2}(\theta - \nu)$.

This will imply that h is not differentiable at ν since

$$\limsup_{\theta\to\nu} \frac{h(\nu) - h(\theta)}{\nu - \theta} \geq \frac{1}{2} > \frac{-1}{2} \geq \liminf_{\theta\to\nu} \frac{h(\nu) - h(\theta)}{\nu - \theta}.$$

Proof of a): Let $\theta \in [\mu-\delta, \nu)$. By Lemma 7, it holds that $\Delta\tilde{J}^\star(\theta) < 0$, so that $\tilde{J}^\star(\theta+\alpha) < \tilde{J}^\star(\theta+\beta)$, and thus $h(\theta) = \tilde{J}^\star(\theta+\alpha)$. Similarly, $h(\nu) = \tilde{J}^\star(\nu+\alpha)$. We note that

$$\pi/8 \leq \pi/4 - \delta \leq \theta + \alpha < \nu + \alpha \leq \pi/4 + \delta \leq 3\pi/8.$$

[3] Since $\nu + \alpha < \mu + 3\pi/8 + \delta < \mu + \pi/2 - \delta$ and $\beta > \mu - 3\pi/8 - \delta > \mu - \pi/2 + \delta$ (recall that $\delta = 0.01$).

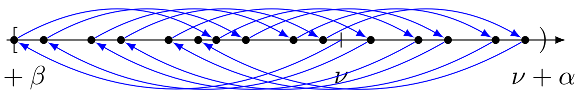

Fig. 2. The truncated backward trajectory $\{T^{-1}\nu,\ldots,T^{-15}\nu\}$ (black dots) of the map T for $\alpha = 0.6$, $\beta = \alpha - \pi/2$ and $\nu = 0.1$. The blue arrows represent transitions by T^{-1}.

(recall that $\mu = \pi/4 - \alpha$). Hence, by Lemma 4, we get that $\tilde{c}(\nu+\alpha) - \tilde{c}(\theta+\alpha) \geq \sqrt{2}/2(\nu-\theta)$ (mean value theorem). Thus, by Lemmas 6 and 2,

$$\tilde{J}^\star(\nu+\alpha) - \tilde{J}^\star(\theta+\alpha) \geq (\sqrt{2}/2 - \rho^2 M)(\nu-\theta) \geq \frac{1}{2}(\nu-\theta),$$

which gives $h(\nu) - h(\theta) \geq \frac{1}{2}(\nu-\theta)$.

Proof of b): Similar to the above: if $\theta \in (\nu, \mu+\delta]$, then we have that $h(\theta) = \tilde{J}^\star(\theta+\beta)$, $h(\nu) = \tilde{J}^\star(\nu+\beta)$ and

$$-3\pi/8 \leq -\pi/4 - \delta \leq \nu+\beta < \theta+\beta \leq -\pi/4+\delta \leq -\pi/8.$$

(recall that $\mu = -\pi/4 - \beta$). Hence, by Lemmas 4, 6 and 2, we obtain $h(\theta) - h(\nu) \leq -\frac{1}{2}(\theta-\nu)$.

This concludes the proof. ∎

Let us now consider the dynamical system T on the set $I \coloneqq [\nu+\beta, \nu+\alpha)$ defined by[4]

$$T : I \to I, \quad T(\theta) = \begin{cases} \theta+\alpha & \text{if } \theta < \nu, \\ \theta+\beta & \text{if } \theta \geq \nu. \end{cases}$$

This system is an *Interval Exchange Map (IEM)* [16], because it translates the interval $I_1 \coloneqq [\nu+\beta, \nu)$ to the interval $I_1' \coloneqq [\nu+\beta+\alpha, \nu+\alpha)$, and the interval $I_2 \coloneqq [\nu, \nu+\alpha)$ to $I_2' \coloneqq [\nu+\beta, \nu+\alpha+\beta)$. These systems are known to be chaotic when the length of I_1 (or I_2) is an irrational multiple of the length of I [16]. See Fig. 2 for an illustration.

Lemma 9: For any $\theta \in I$, the backward orbit of T from θ, i.e., the set $\{T^{-1}\theta, T^{-2}\theta, T^{-3}\theta, \ldots\}$, is dense in I.

Proof: Consider the map

$$S : [0, \pi/2) \to [0, \pi/2), \quad S(\theta) = \theta + \alpha \mod \pi/2.$$

The map S is an irrational rotation (recall that $\alpha/\pi \notin \mathbb{Q}$). Hence, its forward and backward orbits are dense in $[0, \pi/2)$. Now, observe that

$$T(\theta) = S(\theta - \phi) + \phi, \quad \text{where} \quad \phi = \nu + \beta. \tag{6}$$

Proof of (6)*:* If $\theta \in [\nu+\beta, \nu)$, then $\theta - \phi \in [0, -\beta)$ and $\theta - \phi + \alpha \in [0, \pi/2)$. Thus, $S(\theta-\phi) = \theta - \phi + \alpha$ and $T(\theta) = \theta + \alpha = S(\theta-\phi) + \phi$. Similarly, if $\theta \in [\nu, \nu+\alpha)$, then $\theta - \phi \in [-\beta, \pi/2)$ and $\theta - \phi + \alpha \in [\pi/2, \pi/2+\alpha)$, so that $S(\theta-\phi) = \theta - \phi + \alpha - \pi/2 = \theta - \phi + \beta$ and $T(\theta) = \theta + \beta = S(\theta-\phi) + \phi$.

The relation (6) between T and S shows that all backward orbits of T are dense in I. ∎

Lemma 10: The backward orbit of T from ν, i.e., the set $\{T^{-1}\nu, T^{-2}\nu, T^{-3}\nu, \ldots\}$, is contained in $I \setminus \{\nu\}$.

Proof: By contradiction, if $T^{-k}\nu = \nu$ for some $k \in \mathbb{N}_{>0}$, then the backward orbit is periodic (thus not dense), contradicting Lemma 9. ∎

[4]We sometimes note $T(\theta) = T\theta$ for convenience.

Now, we prove our second *key* result, namely that $\tilde{J}^\star$ is not differentiable at any θ in the backward orbit of ν:

Lemma 11: For every $k \in \mathbb{N}$, $\tilde{J}^\star$ is not differentiable at $T^{-k}\nu$.

Proof: The proof is by induction on k. For $k = 0$, the result follows from Lemma 8. Now, we show it for $k \in \mathbb{N}_{>0}$ assuming it holds for $k-1$. Let $\eta = T^{-k}\nu$. By Lemma 10, it holds that $\eta \in [\nu+\beta, \nu) \cup (\nu, \nu+\alpha)$. We consider two cases.

i) Assume that $\eta \in [\nu+\beta, \nu)$. By Lemmas 6 and 7, there is a neighborhood $\mathcal{V}_\eta$ of η such that for all $\theta \in \mathcal{V}_\eta$,

$$\tilde{J}^\star(\theta) = \tilde{c}(\theta) + \rho^2 \tilde{J}^\star(\theta+\alpha).$$

Also, it holds that $T(\eta) = \eta + \alpha$ (since $\eta < \nu$). Hence, for all $\theta \in \mathcal{V}_\eta$,

$$\tilde{J}^\star(\theta) = \tilde{c}(\theta) + \rho^2 \tilde{J}^\star(\theta - \eta + T^{1-k}\nu).$$

Since $\tilde{c}$ is differentiable, this implies that $\tilde{J}^\star$ is not differentiable at η, by using the induction hypothesis (since $\tilde{J}^\star$ is not differentiable at $T^{1-k}\nu$).

ii) Assume that $\eta \in (\nu, \nu+\alpha)$. Then, the same argument as above shows that $\tilde{J}^\star$ is not differentiable at η.

This concludes the proof. ∎

Corollary 2: The function $\tilde{J}^\star$ is non-differentiable on the backward orbit of T from ν, which is dense in I.

Proof: From Lemmas 9 and 11. ∎

Finally, we show that $\tilde{J}^\star$ is non-differentiable on a dense subset of $\mathbb{R}$, and then we deduce that $J^\star$ is non-differentiable on a dense subset of $\mathbb{R}^2$.

Lemma 12: The function $\tilde{J}^\star$ is non-differentiable on a dense subset of $\mathbb{R}$.

Proof: We show that $\tilde{J}^\star$ is non-differentiable on a dense subset of $(\omega-\pi, \omega)$. Since for all $\theta \in \mathbb{R}$ and $k \in \mathbb{Z}$,

$$\tilde{J}^\star(\theta + k\pi) = J^\star(\pm x(\theta)) = J^\star(x(\theta)) = \tilde{J}^\star(\theta)$$

(where we used Lemma 3 for the second equality), this will conclude the proof.

Let $\mathcal{V} \subseteq (\omega-\pi, \omega)$ be an open and nonempty interval. We show that there is $\eta \in \mathcal{V}$ at which $\tilde{J}^\star$ is not differentiable. We distinguish three cases:

i) Assume that $\mathcal{V} \cap I \neq \emptyset$. Then, the result follows from Corollary 2.

ii) Assume that $\mathcal{V} \subseteq (\omega-\pi, \nu+\beta)$. Let $\ell \in \mathbb{N}$ be the smallest positive integer such that $\mathcal{V} + \ell\alpha \cap I \neq \emptyset$. Then, note that, by Lemmas 6 and 7, for all $\theta \in \mathcal{V}$,

$$\tilde{J}^\star(\theta) = \left\{ \sum_{k=0}^{\ell-1} \rho^{2k} \tilde{c}(\theta + k\alpha) \right\} + \rho^{2\ell} \tilde{J}^\star(\theta + \ell\alpha).$$

By Corollary 2 and the assumption on ℓ, there is $\eta \in \mathcal{V}$ such that $\tilde{J}^\star$ is not differentiable at $\eta + \ell\alpha$. Using the above, this shows that $\tilde{J}^\star$ is not differentiable is not differentiable at η.

iii) Assume that $\mathcal{V} \subseteq (\nu+\alpha, \omega-)$. Let $\ell \in \mathbb{N}$ be the smallest positive integer such that $\mathcal{V} + \ell\beta \cap I \neq \emptyset$. Then, note that, by Lemmas 6 and 7, for all $\theta \in \mathcal{V}$,

$$\tilde{J}^\star(\theta) = \left\{ \sum_{k=0}^{\ell-1} \rho^{2k} \tilde{c}(\theta + k\beta) \right\} + \rho^{2\ell} \tilde{J}^\star(\theta + \ell\beta).$$

By Corollary 2 and the assumption on ℓ, there is $\eta \in \mathcal{V}$ such that $\tilde{J}^\star$ is not differentiable at $\eta + \ell\beta$. Using the above, this shows that $\tilde{J}^\star$ is not differentiable at η.

This concludes the proof. ∎

Corollary 3: The function $J^\star$ is non-differentiable on a dense subset of $\mathbb{R}^2$.

Proof: By Lemma 12, we have that $J^\star$ is not differentiable on a dense subset U of $\{x \in \mathbb{R}^2 : \|x\| = 1\}$. Hence, by Lemma 3, we deduce that $J^\star$ is not differentiable on $\mathbb{R}U$ which is dense in $\mathbb{R}^2$. ∎

This concludes the proof of the non-differentiability of $J^\star$ on a dense subset of $\mathbb{R}^2$. Now, let J° be the worst-case value function of $\mathcal{A}$ with cost c. By leveraging the result for $J^\star$, we show easily[5] that J° is non-differentiable on a dense subset of $\mathbb{R}^2$.

Corollary 4: The function J° is non-differentiable on a dense subset of $\mathbb{R}^2$.

Proof: Note that for all $\theta \in \mathbb{R}$, $\tilde{c}(\theta) = 2 - \cos^2(\theta)$. Hence, for all $\theta \in \mathbb{R}$, $\tilde{c}(\theta + \pi/2) = 3 - \tilde{c}(\theta)$. For all $\theta \in \mathbb{R}$, let $\tilde{J}^\circ(\theta) = -\tilde{J}^\star(\theta + \pi/2) + 3/(1 - \rho^2)$. Then,

$$\tilde{J}^\circ(\theta) = -\tilde{J}^\star(\theta + \pi/2) + 3/(1 - \rho^2)$$

(by Lemma 6:)

$$\begin{aligned}
&= -\tilde{c}(\theta + \pi/2) - \rho^2 \min\{\tilde{J}^\star(\theta + \pi/2 + \alpha),\\
&\qquad\qquad \tilde{J}^\star(\theta + \pi/2 + \beta)\} + 3/(1 - \rho^2)\\
&= \tilde{c}(\theta) - 3 + \rho^2 \max\{-\tilde{J}^\star(\theta + \pi/2 + \alpha),\\
&\qquad\qquad - \tilde{J}^\star(\theta + \pi/2 + \beta)\} + 3/(1 - \rho^2)\\
&= \tilde{c}(\theta) - 3 + \rho^2 \max\{\tilde{J}^\circ(\theta + \alpha), \tilde{J}^\circ(\theta + \beta)\}\\
&\qquad\qquad + 3/(1 - \rho^2) - \rho^2 3/(1 - \rho^2)\\
&= \tilde{c}(\theta) + \rho^2 \min\{\tilde{J}^\circ(\theta + \alpha), \tilde{J}^\circ(\theta + \beta)\}.
\end{aligned}$$

Hence, J° defined by $J^\circ(x) = r^2\tilde{J}^\circ(\theta)$ where $x = rx(\theta)$ with $r \geq 0$ and $\theta \in \mathbb{R}$, satisfies (3). This implies that J° is the worst-case value function, by classical arguments in dynamic programming (see, e.g., [19, Theorem 3.1]). ∎

This concludes the proof of Theorem 3 for the case $n = 2$.

2) Case $n > 2$: Now, we exploit the above to obtain a similar system for $n > 2$. Therefore, consider the matrices

$$\hat{A}_1 = \rho \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0_{1\times(n-2)} \\ \sin(\alpha) & \cos(\alpha) & 0_{1\times(n-2)} \\ 0_{(n-2)\times 1} & 0_{(n-2)\times 1} & 0_{(n-2)\times(n-2)} \end{bmatrix},$$

$$\hat{A}_2 = \rho \begin{bmatrix} \cos(\beta) & -\sin(\beta) & 0_{1\times(n-2)} \\ \sin(\beta) & \cos(\beta) & 0_{1\times(n-2)} \\ 0_{(n-2)\times 1} & 0_{(n-2)\times 1} & 0_{(n-2)\times(n-2)} \end{bmatrix},$$

where α, β and ρ are as above, and the cost function

$$c(x) = \cos^2(x_1) + 2\sin^2(x_2),$$

where $x = [x_1, \ldots, x_n]^\top$. Let $\hat{J}^\star$ and $\hat{J}^\circ$ be the associated optimal and worst-case value functions respectively.

Lemma 13: It holds that $\hat{J}^\star(x) = J^\star([x_1, x_2]^\top)$ and $\hat{J}^\circ(x) = J^\circ([x_1, x_2]^\top)$, where $x = [x_1, \ldots, x_n]^\top$ and $J^\star$ and J° are as in the above subsubsection (case $n = 2$).

[5]Note that a derivation as for $J^\star$ would also have been possible, but it is easier to build upon the work that has already been done.

Proof: Straightforward (omitted). ∎

Lemma 13 implies that $J^\star$ and J° are non-differentiable on dense subsets of $\mathbb{R}^n$.

Corollary 5: The functions $\hat{J}^\star$ and $\hat{J}^\circ$ are non-differentiable on dense subsets of $\mathbb{R}^n$.

This concludes the proof of Theorem 3 for the general case.

Rational matrices: Finally, it remains to show that there are rational matrices of the form (4)–(5).

Lemma 14: The matrices

$$A_1 = \rho \begin{bmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{bmatrix} \text{ and } A_2 = \rho \begin{bmatrix} 0.6 & 0.8 \\ -0.8 & 0.6 \end{bmatrix}$$

have the form (4)–(5) with $\alpha = \arctan(3/4) \in (\pi/8, 3\pi/8)$ and $\beta = \arctan(-4/3) \in (-3\pi/8, -\pi/8)$. Furthermore, it holds $\beta = \alpha - \pi/2$ and $\alpha/\pi \notin \mathbb{Q}$.

Proof: The form of A_1 and A_2 and the fact that $\beta = \alpha - \pi/2$ are straightforward to check. The fact that $\alpha/\pi \notin \mathbb{Q}$ follows from Niven's theorem [20, Corollary 3.12]. ∎

V. Conclusions

We showed that the optimal and worst-case value functions of switched linear systems can be highly unsmooth. Existing results in the literature (e.g., [8]) suggested it, but no formal statement or proof were available. This work addresses this gap, by providing a constructive proof that these functions can be non-differentiable on dense subsets on the state space. The implication for their computation is that it would be very hard for generic algorithms to compute them exactly, because those generally search in spaces of piecewise smooth functions. Also, the problem of finding the point x for which the optimal or worst-case value functions is minimal appears challenging, because one cannot use derivatives of these functions to find their extrema. In future work, we plan to identify sufficient conditions under which the value functions are piecewise differentiable.

References

[1] D. Liberzon, *Switching in systems and control*. Boston, MA: Birkhäuser, 2003.

[2] R. M. Jungers, *The joint spectral radius: theory and applications*. Berlin: Springer, 2009.

[3] Z. Sun and S. S. Ge, *Stability theory of switched dynamical systems*. London: Springer, 2011.

[4] V. D. Blondel and J. N. Tsitsiklis, "Complexity of stability and controllability of elementary hybrid systems," *Automatica*, vol. 35, no. 3, pp. 479–489, 1999.

[5] A. A. Ahmadi, R. M. Jungers, P. A. Parrilo, and M. Roozbehani, "Joint spectral radius and path-complete graph Lyapunov functions," *SIAM Journal on Control and Optimization*, vol. 52, no. 1, pp. 687–717, 2014.

[6] H. Lin and P. J. Antsaklis, "Stability and stabilizability of switched linear systems: a survey of recent results," *IEEE Transactions on Automatic control*, vol. 54, no. 2, pp. 308–322, 2009.

[7] A. Rantzer, "On approximate dynamic programming in switching systems," in *Proceedings of the 44th IEEE Conference on Decision and Control*. IEEE, 2005, pp. 1391–1396.

[8] W. Zhang, J. Hu, and A. Abate, "On the value functions of the discrete-time switched LQR problem," *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2669–2674, 2009.

[9] D. Görges, M. Izák, and S. Liu, "Optimal control and scheduling of switched systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 1, pp. 135–140, 2011.

[10] W. Zhang, J. Hu, and A. Abate, "Infinite-horizon switched LQR problems in discrete time: a suboptimal algorithm with performance analysis," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1815–1821, 2012.

[11] F. Zhu and P. J. Antsaklis, "Optimal control of hybrid switched systems: a brief survey," *Discrete Event Dynamic Systems*, vol. 25, no. 3, pp. 345–364, 2015.

[12] D. Antunes and W. M. Heemels, "Linear quadratic regulation of switched systems using informed policies," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2675–2688, 2017.

[13] G. Wu, L. Xiong, G. Wang, and J. Sun, "Linear quadratic regulator of discrete-time switched linear systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 12, pp. 3113–3117, 2020.

[14] Z. Wu and Q. He, "Optimal switching sequence for switched linear systems," *SIAM Journal on Control and Optimization*, vol. 58, no. 2, pp. 1183–1206, 2020.

[15] T. Hou, Y. Li, and Z. Lin, "An improved method for approximating the infinite-horizon value function of the discrete-time switched LQR problem," *Journal of Systems Science and Complexity*, vol. 37, no. 1, pp. 22–39, 2024.

[16] M. Keane, "Interval exchange transformations," *Mathematische Zeitschrift*, vol. 141, no. 1, pp. 25–31, 1975.

[17] H. Harder and S. Peitz, "On the continuity and smoothness of the value function in reinforcement learning and optimal control," in *2024 IEEE 63rd Conference on Decision and Control (CDC)*. IEEE, 2024, pp. 1935–1940.

[18] W. Rudin, *Real and complex analysis*, 3rd ed. New York, NY: McGraw-Hill, 1987.

[19] S. Meyn, *Control systems and reinforcement learning*. Padstow: Cambridge University Press, 2022.

[20] I. Niven, *Irrational numbers*. Rahway, NJ: Mathematical Association of America, 1956.