ON THE NECESSITY OF BARRIER CERTIFICATES

Stephen Prajna [*,1] **Anders Rantzer** [**,1]

[*] *Control and Dynamical Systems*
California Institute of Technology
Pasadena, CA 91125, USA
E-mail: prajna@cds.caltech.edu

[**] *Department of Automatic Control*
Lund Institute of Technology
SE 221 00 Lund, Sweden
E-mail: rantzer@control.lth.se

Abstract: A methodology for safety verification of nonlinear systems using barrier certificates has been proposed recently. The condition was stated in a sufficiency form: if there exists a barrier certificate, then the system is safe, in the sense that there is no trajectory starting from a given set of initial states that reaches a given unsafe region. Using the concepts of convex duality and density functions, in this paper we derive a converse statement for barrier certificates, showing that in a quite general setting the existence of a barrier certificate is also necessary for safety.

Keywords: Safety analysis, barrier certificates, density functions, nonlinear systems, convex duality

1. INTRODUCTION

Safety verification addresses the question whether an unsafe or bad region in the state space is reachable by some system trajectories starting from a set of initial states. The need for safety verification arises as the complexity of the system increases, and is also underscored by the safety critical nature of the system.

Various methods have been proposed for safety verification. For verification of finite state systems, model checking techniques (Clarke, Jr. *et al.*, 2000) have been quite successful and have garnered a popularity that prompts the development of analogous approaches for verification of continuous systems, which mostly require computing the propagation of initial states (see e.g. (Alur *et al.*, 2003; Kurzhanski and Varaiya, 2000)). Unfortunately, while these methods allow us to compute an exact or near exact approximation of reachable sets, it is difficult to perform such a computation when the system is nonlinear and uncertain.

Using a different approach, we recently proposed a method for safety verification that is based on what we term barrier certificates (Prajna and Jadbabaie, 2004). Our conditions for safety can be stated as follows. Given a system $\dot{x} = f(x)$ with the state x taking its value in $\mathcal{X} \subseteq \mathbb{R}^n$, a set of initial states $\mathcal{X}_0 \subseteq \mathcal{X}$, and an unsafe set $\mathcal{X}_u \subseteq \mathcal{X}$, suppose there exists a continuously differentiable function $B : \mathbb{R}^n \to \mathbb{R}$ such that

$$B(x) \leq 0 \quad \forall x \in \mathcal{X}_0, \tag{1}$$

[1] The collaboration between the authors was supported by an exchange grant from the Swedish Foundation for International Cooperation in Research and Higher Education.

$$B(x) > 0 \quad \forall x \in \mathcal{X}_u, \tag{2}$$

$$\frac{\partial B}{\partial x}(x) f(x) \leq 0 \quad \forall x \in \mathcal{X}. \tag{3}$$

Then the system is safe, i.e., there is no trajectory $x(t)$ of the system such that $x(0) \in \mathcal{X}_0$, $x(T) \in \mathcal{X}_u$ for some $T \geq 0$, and $x(t) \in \mathcal{X}$ for all $t \in [0, T]$.

A function $B(x)$ satisfying (1)–(3) is called a barrier certificate. The above method is analogous to the Lyapunov method for stability analysis (Khalil, 1996), and is also closely related to the use of viability theory (Aubin, 1991) and invariant sets (Jirstrand, 1998) for safety verification. When the vector field $f(x)$ is polynomial and the sets $\mathcal{X}$, $\mathcal{X}_0$, $\mathcal{X}_u$ are semialgebraic, a polynomial barrier certificate $B(x)$ can be searched using sum of squares programming (Prajna *et al.*, 2002). The method can also be extended to handle hybrid, uncertain, and stochastic systems (Prajna and Jadbabaie, 2004; Prajna *et al.*, 2004), or to verify other system properties such as reachability and eventuality (Prajna and Rantzer, 2005).

In the present paper, we derive a converse statement for barrier certificates. We use convex duality and density functions (Rantzer, 2001; Rantzer and Hedlund, 2003) to show that under some reasonable technical conditions, there exists a barrier certificate if and only if the system is safe. In Section 2, we give an intuitive illustration of the main idea by addressing the verification of a simple discrete system. The main result of the paper is presented and proven in Section 3. Some concluding remarks will be given in Section 4.

Notations: We denote the spaces of m-times continuously differentiable functions mapping $\mathcal{X} \subseteq \mathbb{R}^n$ to $\mathbb{R}^n$ by $C^m(\mathcal{X}, \mathbb{R}^n)$, and $\mathcal{X}$ to $\mathbb{R}$ by $C^m(\mathcal{X})$. The spaces of continuous functions are denoted by $C(\mathcal{X}, \mathbb{R}^n)$ and $C(\mathcal{X})$, equipped with the supremum norm if necessary. The zero subscript as in $C_0^1(\mathbb{R}^n)$ indicates compact support. For a normed vector space $\mathcal{K}$, the dual space is denoted by $\mathcal{K}^*$. Finally, the flow of $\dot{x} = f(x)$ starting at x_0 is denoted by $\phi_t(x_0)$.

2. A DISCRETE EXAMPLE

To give an intuitive flavor of the ideas used in the main theorem, let us consider the verification of a simple discrete system, shown in Figure 1. The system has four states, labelled 1 through 4, and three transitions between states, represented by the directed edges. We assume that node 1 is the initial state and node 4 is the unsafe state.

For this system, conditions analogous to (1)–(3) that must be satisfied by a barrier certificate can be formulated. One way to find a barrier certificate which proves safety is by solving the linear program (LP):

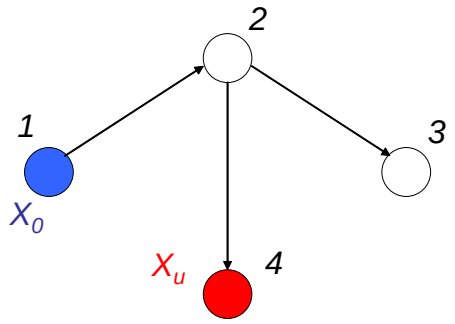

Fig. 1. A simple discrete system. The nodes represent the states of the system, while the directed edges represent transitions between states.

$$\begin{aligned} &\max\ B_4 - B_1 \\ &\text{subject to}\ B_2 - B_1 \leq 0, \\ &\qquad\qquad B_3 - B_2 \leq 0, \\ &\qquad\qquad B_4 - B_2 \leq 0, \end{aligned}$$

where the decision variables B_1, B_2, B_3, B_4 are reals, and B_i corresponds to the value of the barrier certificate at node i. If there is a feasible solution of the above problem such that the objective function is strictly positive, then there exists a barrier certificate for the system, and consequently there is no path going from node 1 to node 4.

The dual of the above LP is as follows:

$$\begin{aligned} &\min\ 0 \\ &\text{subject to}\ \rho_{12} \geq 0, \rho_{23} \geq 0, \rho_{24} \geq 0, \\ &\qquad\qquad \rho_{12} = 1,\ \rho_{24} = 1,\ \rho_{23} = 0, \\ &\qquad\qquad \rho_{24} + \rho_{23} - \rho_{12} = 0, \end{aligned}$$

The dual decision variable ρ_{ij} can be interpreted as the transportation density from node i to node j. The equality constraints basically state that conservation of flows holds at each node – the total flow into a node is equal to the total flow out. In addition, the first and second equality constraints indicate that there exist a unit source at node 1, i.e., the initial state, and a unit sink at node 4, i.e., the unsafe state. This duality interpretation has been studied extensively in the past; see, e.g., (Papadimitriou and Steiglitz, 1998) and references therein.

The existence of a feasible solution to the dual LP implies the existence of a path from the initial state to the unsafe state. This can be shown using the facts that the flows are conserved and that there are a unit source and a unit sink at the initial state and unsafe state, respectively. Hence, solving the dual LP can be used for verifying reachability. As a matter of fact, we obtain a linear programming formulation of the shortest path problem if we also add the objective function $\sum \rho_{ij}$ to the dual LP. In this case, the nonzero entries corresponding to any optimal vertex solution to the LP will indicate a shortest path from

the initial node to the unsafe node (Papadimitriou and Steiglitz, 1998).

This duality argument can also be used to prove that the existence of a barrier certificate is both sufficient and necessary for safety. For this, suppose that there exists no barrier certificate for the system, which is equivalent to the maximum objective value of the primal LP being equal to zero. This objective value is attained by, e.g., $B_i = 0$ for all i. The linear programming duality (Boyd and Vandenberghe, 2004) implies that there exists a feasible solution to the dual LP, from which we can further conclude the existence of a path from the initial state to the unsafe state, as explained in the previous paragraph. It is exactly the continuous counterpart of this argument that we will develop in the next section.

For the above example, the optimal objective value of the primal LP is equal to zero, and hence the safety property does not hold. The unique feasible solution to the dual LP is given by $\rho_{12} = 1$, $\rho_{23} = 0$, $\rho_{24} = 1$, which shows the path from node 1 to node 4. Had the direction of the edge from node 2 to node 4 been reversed, for example, the optimal objective value of the corresponding primal LP will be ∞, and there will be no feasible solution to the dual LP.

3. CONTINUOUS SYSTEMS

The main result of the paper is as follows.

Theorem 1. Consider the system $\dot{x} = f(x)$ with $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$. Let $\mathcal{X} \subset \mathbb{R}^n$, and $\mathcal{X}_0 \subseteq \mathcal{X}$, $\mathcal{X}_u \subseteq \mathcal{X}$ be compact sets, and suppose that there exists a function $\tilde{B} \in C^1(\mathbb{R}^n)$ such that $\frac{\partial \tilde{B}}{\partial x}(x)f(x) < 0$ for all $x \in \mathcal{X}$. Then there exists a function $B \in C^1(\mathbb{R}^n)$ that satisfies

$$B(x) \leq 0 \quad \forall x \in \mathcal{X}_0, \tag{4}$$
$$B(x) > 0 \quad \forall x \in \mathcal{X}_u, \tag{5}$$
$$\frac{\partial B}{\partial x}(x)f(x) \leq 0 \quad \forall x \in \mathcal{X} \tag{6}$$

if and only if the safety property holds, i.e., if there exists no trajectory $x(t)$ of the system such that $x(0) \in \mathcal{X}_0$, $x(T) \in \mathcal{X}_u$ for some $T \geq 0$, and $x(t) \in \mathcal{X}$ for all $t \in [0, T]$.

Notice that in the theorem we have used a seemingly strong assumption that there exists a function $\tilde{B} \in C^1(\mathbb{R}^n)$ such that $\frac{\partial \tilde{B}}{\partial x}(x)f(x) < 0$ $\forall x \in \mathcal{X}$. In Section 4 we will show that in many cases of interest the existence of such $\tilde{B}(x)$ is actually guaranteed.

Our proof of the converse statement in Theorem 1 consists of two parts, given in Lemmas 2 and 4 below. In the first lemma, we use the Hahn-Banach theorem to show that the non-existence of a $B(x)$ satisfying the conditions in Theorem 1 implies the existence of measures ψ_0, ψ_u, ρ satisfying some appropriate conditions. Then, in Lemma 4 we show that the existence of such ψ_0, ψ_u, ρ actually implies that there exists an unsafe trajectory of the system.

Lemma 2. Let $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, and $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{X}_0 \subseteq \mathcal{X}$, $\mathcal{X}_u \subseteq \mathcal{X}$ be compact sets. Suppose there exists a function $\tilde{B} \in C^1(\mathbb{R}^n)$ such that $\frac{\partial \tilde{B}}{\partial x}(x)f(x) < 0$ for all $x \in \mathcal{X}$. Then there exists no $B \in C^1(\mathbb{R}^n)$ satisfying (4)–(6) only if there exist measures of bounded variation ψ_0, ψ_u, ρ (each defined on $\mathbb{R}^n$) such that ψ_0, ψ_u, ρ are nonnegative on $\mathbb{R}^n$ and equal to zero outside $\mathcal{X}_0$, $\mathcal{X}_u$, and $\mathcal{X}$ respectively; and

$$\int_{\mathcal{X}_0} d\psi_0 = 1, \ \int_{\mathcal{X}_u} d\psi_u = 1,$$
$$\nabla \cdot (\rho f) = \psi_0 - \psi_u,$$

where $\nabla \cdot (\rho f)$ is interpreted as a distributional derivative.

Proof. Consider the convex optimization problem

$$\begin{aligned} &\sup\ B_u - B_0, \\ &\text{subject to } B(x) - B_0 \leq 0 \quad \forall x \in \mathcal{X}_0, \\ &\qquad\qquad B(x) - B_u \geq 0 \quad \forall x \in \mathcal{X}_u, \\ &\qquad\qquad \frac{\partial B}{\partial x}(x)f(x) \leq 0 \quad \forall x \in \mathcal{X}, \end{aligned}$$

with the supremum denoted by γ, and taken over all $B_0 \in \mathbb{R}$, $B_u \in \mathbb{R}$, and $B \in C^1(\mathbb{R}^n)$. Since $B_0 = 0$, $B_u = 0$, and $B(x) = 0$ satisfy the constraint, γ must be greater than or equal to zero. In addition, since the objective function and the constraints are all linear, the value of γ is either 0 or ∞. There exists no $B \in C^1(\mathbb{R}^n)$ satisfying (4)–(6) if and only if the value of γ is equal to zero.

Now suppose that $\gamma = 0$. Let $\mathcal{K} = \mathbb{R} \times (C(\mathcal{X}))^3$, $\mathcal{B} = \mathbb{R}^2 \times C_0^1(\mathbb{R}^n)$, and define $\mathcal{K}_1$, $\mathcal{K}_2$ as follows:

$$\begin{aligned} \mathcal{K}_1 = \{&(z, h_0, h_u, h) \in \mathcal{K} : h_0 = B_0 - B \text{ on } \mathcal{X}, \\ &h_u = B - B_u \text{ on } \mathcal{X},\ h = -\frac{\partial B}{\partial x} f \text{ on } \mathcal{X}, \\ &z = B_u - B_0, \text{ and } (B_0, B_u, B) \in \mathcal{B}\}, \\ \mathcal{K}_2 = \{&(z, h_0, h_u, h) \in \mathcal{K} : z \geq 0,\ h_0 \geq 0 \text{ on } \mathcal{X}_0, \\ &h_u \geq 0 \text{ on } \mathcal{X}_u,\ h \geq 0 \text{ on } \mathcal{X}\}. \end{aligned}$$

Then both $\mathcal{K}_1$ and $\mathcal{K}_2$ are convex sets, and $\mathcal{K}_2$ has non-empty interior in $\mathcal{K}$. Furthermore, since $\gamma = 0$, it follows that the first component in $\mathcal{K}_1$ is less than or equal to zero when the second, third, and fourth components are greater than or equal to zero, and therefore $\mathcal{K}_1 \cap \text{int}(\mathcal{K}_2) = \emptyset$. Now, by the Hahn-Banach theorem (Luenberger, 1969), there exists a nonzero $k^* = (a, \tilde{\psi}_0, \tilde{\psi}_u, \tilde{\rho}) \in \mathcal{K}^* = \mathbb{R} \times (C(\mathcal{X})^*)^3$ such that

$$\sup_{k_1\in\mathcal{K}_1}\langle k^*,k_1\rangle \le \inf_{k_2\in\mathcal{K}_2}\langle k^*,k_2\rangle, \tag{7}$$

where $C(\mathcal{X})^*$ in this case is the set of measures on $\mathcal{X}$ with bounded variation. The right-hand side of the inequality can be expanded as follows

$$\begin{aligned}
&\inf_{k_2\in\mathcal{K}_2}\langle k^*,k_2\rangle\\
&=\inf_{(z,h_0,h_u,h)\in\mathcal{K}_2} az+\langle\tilde\psi_0,h_0\rangle+\langle\tilde\psi_u,h_u\rangle+\langle\tilde\rho,h\rangle\\
&=\begin{cases}0, & \text{if } a\ge 0;\ \tilde\psi_0,\ \tilde\psi_u,\ \tilde\rho\ge 0;\ \text{and}\\ & \tilde\psi_0,\ \tilde\psi_u \text{ are zero outside } \mathcal{X}_0,\mathcal{X}_u \text{ resp.},\\ -\infty, & \text{otherwise.}\end{cases}
\end{aligned}$$

Now denote the extension of $\tilde\psi_0$, $\tilde\psi_u$, $\tilde\rho$ to the whole $\mathbb{R}^n$ by ψ_0, ψ_u, ρ, which are obtained by letting them equal to zero outside of $\mathcal{X}$. Then, for the left-hand side of (7), we have the following equality:

$$\begin{aligned}
&\sup_{k_1\in\mathcal{K}_1}\langle k^*,k_1\rangle\\
&=\sup_{(B_0,B_u,B)\in\mathcal{B}} a(B_u-B_0)+\langle\psi_0,B_0-B\rangle\\
&\qquad+\langle\psi_u,B-B_u\rangle+\langle\rho,-\frac{\partial B}{\partial x}f\rangle\\
&=\sup_{(B_0,B_u,B)\in\mathcal{B}}(-a+\int d\psi_0)B_0+(a-\int d\psi_u)B_u\\
&\qquad+\langle-\psi_0+\psi_u+\nabla\cdot(\rho f),B\rangle\\
&=\begin{cases}0, & \text{if } \int_{\mathbb{R}^n} d\psi_0=a,\int_{\mathbb{R}^n} d\psi_u=a,\ \text{and}\\ & -\psi_0+\psi_u+\nabla\cdot(\rho f)=0\\ \infty, & \text{otherwise,}\end{cases}
\end{aligned}$$

where $\nabla\cdot(\rho f)$ is interpreted as a distributional derivative. Thus, for the supremum to be less than or equal to the infimum, we must have a nonzero (a,ψ_0,ψ_u,ρ), where ψ_0, ψ_u, ρ are measures of bounded variation on $\mathbb{R}^n$, such that $a\ge 0$; ψ_0, ψ_u, ρ are nonnegative; ψ_0, ψ_u, ρ are equal to zero outside $\mathcal{X}_0$, $\mathcal{X}_u$, and $\mathcal{X}$ respectively; and

$$\begin{aligned}
&\int_{\mathbb{R}^n} d\psi_0=a,\ \int_{\mathbb{R}^n} d\psi_u=a,\\
&\nabla\cdot(\rho f)=\psi_0-\psi_u.
\end{aligned}$$

We will next show that because of the assumption that there exists a $\tilde B\in C^1(\mathbb{R}^n)$ such that $\frac{\partial\tilde B}{\partial x}(x)f(x)<0$ for all $x\in\mathcal{X}$, we must have $a>0$. For this, let $\mathcal{L}=(C(\mathcal{X}))^3$, and define

$$\begin{aligned}
\mathcal{L}_1=\{&(h_0,h_u,h)\in\mathcal{L}: h_0=B_0-B \text{ on } \mathcal{X},\\
&h_u=B-B_u \text{ on } \mathcal{X},\ h=-\frac{\partial B}{\partial x}f \text{ on } \mathcal{X},\\
&\text{and } (B_0,B_u,B)\in\mathcal{B}\},\\
\mathcal{L}_2=\{&(h_0,h_u,h)\in\mathcal{L}: h_0\ge 0 \text{ on } \mathcal{X}_0,\\
&h_u\ge 0 \text{ on } \mathcal{X}_u,\ h\ge 0 \text{ on } \mathcal{X}\}.
\end{aligned}$$

Note in particular that due to the above assumption and the compactness of $\mathcal{X}_0$, $\mathcal{X}_u$, $\mathcal{X}$, we have $\mathcal{L}_1\cap\text{int}(\mathcal{L}_2)\neq\emptyset$. Now consider $k^*=(a,\tilde\psi_0,\tilde\psi_u,\tilde\rho)$ that we have before. Suppose that $a=0$ and substitute this to (7). Then we have a nonzero $(\tilde\psi_0,\tilde\psi_u,\tilde\rho)\in(C(\mathcal{X})^*)^3$, such that

$$\sup_{\ell_1\in\mathcal{L}_1}\langle(\tilde\psi_0,\tilde\psi_u,\tilde\rho),\ell_1\rangle\le\inf_{\ell_2\in\mathcal{L}_2}\langle(\tilde\psi_0,\tilde\psi_u,\tilde\rho),\ell_2\rangle.$$

This implies that $\mathcal{L}_1\cap\text{int}(\mathcal{L}_2)=\emptyset$, which is contradictory to the above. Thus a must be strictly positive. Without loss of generality, assume that k^* is scaled such that $a=1$. This completes the proof of our lemma. ■

Next, we will show that the existence of ψ_0, ψ_u, ρ in the conclusion of Lemma 2 implies that there exists an unsafe trajectory of the system. Since in this case we have a density function ρ which is in fact a measure, we need a version of Liouville theorem which applies to measures.

Lemma 3. Let $f\in C^1(D,\mathbb{R}^n)$ where $D\subseteq\mathbb{R}^n$ is open. For a measurable set Z, assume that $\phi_t(Z)$ is a subset of D for all t between 0 and T. If ρ is a measure of bounded variation on D such that ρ has a compact support and the distributional derivative $\nabla\cdot(\rho f)$ is also a measure of bounded variation with compact support, then

$$\int_{\phi_T(Z)} d\rho-\int_Z d\rho=\int_0^T\int_{\phi_t(Z)} d(\nabla\cdot(\rho f))dt.$$

Proof. Choose $\rho_1,\rho_2,\ldots\in C_0^\infty(D)$ such that $\rho_k\to\rho$ in the (weak) topology of distributions. Then also $\nabla\cdot(\rho_k f)\to\nabla\cdot(\rho f)$ in the sense of distributions. In particular

$$\begin{aligned}
&\lim_{k\to\infty}\int_X d|\rho_k-\rho|=0,\\
&\lim_{k\to\infty}\int_X d|\nabla\cdot(\rho_k f)-\nabla\cdot(\rho f)|=0
\end{aligned}$$

for every $X\subset D$. The lemma was proven for the case of smooth ρ in (Rantzer, 2001), i.e.,

$$\begin{aligned}
&\int_{\phi_T(Z)}\rho_k(x)dx-\int_Z\rho_k(x)dx\\
&\qquad=\int_0^T\int_{\phi_t(Z)}[\nabla\cdot(\rho_k f)(x)]dxdt.
\end{aligned}$$

So the desired equality is obtained in the limit as $k\to\infty$. ■

Lemma 4. Consider the system $\dot x=f(x)$ with $f\in C^1(\mathbb{R}^n,\mathbb{R}^n)$, and let $\mathcal{X}\subset\mathbb{R}^n$, $\mathcal{X}_0\subseteq\mathcal{X}$, $\mathcal{X}_u\subseteq\mathcal{X}$ be compact sets. Suppose there exist measures of bounded variations ψ_0, ψ_u, ρ such that ψ_0,ψ_u,ρ are nonnegative on $\mathbb{R}^n$ and equal to zero outside $\mathcal{X}_0$, $\mathcal{X}_u$, and $\mathcal{X}$ respectively; and $\int_{\mathcal{X}_0} d\psi_0=1$, $\int_{\mathcal{X}_u} d\psi_u=1$, $\nabla\cdot(\rho f)=\psi_0-\psi_u$. Then there exists a $T\ge 0$ and a trajectory $x(t)$ of the system such that $x(0)\in\mathcal{X}_0$, $x(T)\in\mathcal{X}_u$, and $x(t)\in\mathcal{X}$ for all $t\in[0,T]$.

Proof. Let $X_1,X_2,\ldots\subseteq\mathbb{R}^n$ be a sequence of open sets such that $\mathcal{X}_0\subseteq X_i$ for all i and $\lim_{i\to\infty}X_i=\mathcal{X}_0$. In addition, define the measurable sets

$$Z_i = \bigcup_{x_0 \in X_i} \{x \in \mathbb{R}^n : x = \phi_t(x_0) \text{ for some } t \geq 0\},$$

for $i = 1, 2, 3$, and so on. By the assertions of the lemma, both ρ and $\nabla \cdot (\rho f)$ are measures with bounded variation and compact support, so we can use Lemma 3 and $\nabla \cdot (\rho f) = \psi_0 - \psi_u$ to obtain the relation

$$\int_{\phi_t(Z_i)} d\rho - \int_{Z_i} d\rho = \int_0^t \int_{\phi_\tau(Z_i)} d(\psi_0 - \psi_u) d\tau$$

for all $t \geq 0$. Since $\rho \geq 0$ and $\phi_t(Z_i) \subseteq Z_i$ for all $t \geq 0$, the left-hand side of the above expression is less than or equal to zero. It follows from $\int_{\mathcal{X}_0} d\psi_0 = 1$ and $\psi_0 \geq 0$ that $\mathcal{X}_u \cap Z_i \neq \emptyset$ for all $i = 1, 2, ...$, for otherwise the right-hand side of the expression can be made strictly greater than zero by taking some $t > 0$ and we obtain a contradiction. Since the sets $\mathcal{X}_0$ and $\mathcal{X}_u$ are closed, we conclude that $\phi_T(x_0) \in \mathcal{X}_u$ for some $T \geq 0$ and $x_0 \in \mathcal{X}_0$. For our purpose, let T be the first time instance such that $\phi_T(x_0) \in \mathcal{X}_u$.

The case in which $T = 0$ is trivial since $\mathcal{X}_0 \subseteq \mathcal{X}$. Consider now the case in which $T > 0$. We will show that $\phi_t(x_0) \in \mathcal{X}$ for all $t \in [0, T]$ by a contradiction. Suppose to the contrary that there exists $\tilde{T} \in (0, T)$ such that $\phi_{\tilde{T}}(x_0) \notin \mathcal{X}$. Then, for a sufficiently small open neighborhood U of x_0, we have

$$\phi_{\tilde{T}}(U) \subset \mathbb{R}^n \setminus (\mathcal{X}),$$
$$\phi_t(U) \cap \mathcal{X}_u = \emptyset \quad \forall t \in [0, \tilde{T}].$$

Use again Lemma 3:

$$\int_{\phi_{\tilde{T}}(U)} d\rho - \int_U d\rho = \int_0^{\tilde{T}} \int_{\phi_\tau(U)} d(\psi_0 - \psi_u) d\tau.$$

Since $\rho = 0$ on $\mathbb{R}^n \setminus (\mathcal{X})$, the first term on the left is equal to zero, and therefore the left-hand side is non-positive, which leads to a contradiction since the right-hand side is strictly greater than zero. This lets us conclude that $\phi_t(x_0) \in \mathcal{X}$ for all $t \in [0, T]$, thus finishing the proof of the lemma. ■

We are now ready to present the proof of the main theorem.

Proof of Theorem 1.

($\Rightarrow$): Assume that there exists a $B(x)$ satisfying (1)–(3), while at the same time the system is not safe, i.e., there is an initial condition $x_0 \in \mathcal{X}_0$ such that the flow $x(t)$ of the model $\dot{x} = f(x)$ starting at $x(0) = x_0$ satisfies $x(t) \in \mathcal{X}$ for all $t \in [0, T]$ and $x(T) \in \mathcal{X}_u$. Condition (3) states that the Lie derivative of $B(x)$ along this flow is non-positive. A direct consequence of this is that $B(x(T))$ must less than or equal to $B(x(0))$, which is contradictory to (1)–(2). Thus the initial assumption is not correct: the system is safe.

($\Leftarrow$): Follows from Lemmas 2 and 4. ■

4. CONCLUDING REMARKS

The result stated in Theorem 1 uses the assumption that the following Slater-like condition (Boyd and Vandenberghe, 2004) is fulfilled: that there exists a function $\tilde{B} \in C^1(\mathbb{R}^n)$ such that $\frac{\partial \tilde{B}}{\partial x}(x) f(x) < 0$ for all $x \in \mathcal{X}$. While in the discrete case strong duality holds (and hence the necessity of barrier certificates too) without such an assumption, its proof depends on a special property of polyhedral convex sets, which does not carry over to the continuous case. Eliminating this condition in the continuous case will presumably require a different proof technique than the one presented in this paper. Nevertheless, there are cases in which the condition is automatically fulfilled, for instance when the trajectories of the system starting from any $x_0 \in \mathcal{X}$ leave a neighborhood of $\mathcal{X}$ at least once, as shown in the following proposition.

Proposition 5. Consider the system $\dot{x} = f(x)$ with $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ and let $\mathcal{X} \subset \mathbb{R}^n$ be a compact set. Suppose there exist an open neighborhood $\tilde{\mathcal{X}}$ of $\mathcal{X}$ and a time instant $T > 0$ such that for all initial conditions $x_0 \in \mathcal{X}$, we have the flow $\phi_t(x_0)$ outside of $\text{cl}(\tilde{\mathcal{X}})$ for some $t \in [0, T]$. Then there exists a function $\tilde{B} \in C^1(\mathbb{R}^n)$ such that $\frac{\partial \tilde{B}}{\partial x}(x) f(x) < 0$ for all $x \in \mathcal{X}$.

Proof. Let $\mathcal{Y}$ be an open neighborhood of $\mathcal{X}$ such that its closure is contained in $\tilde{\mathcal{X}}$. In addition, let $\xi \in C^1(\mathbb{R}^n)$ be a nonnegative function such that $\xi(x) = 1$ for all $x \in \mathcal{Y}$ and $\xi(x) = 0$ for all $x \notin \tilde{\mathcal{X}}$; also let $\psi \in C^1(\mathbb{R}^n)$ be a function such that $\psi(x) > 0$ for all $x \in \mathcal{X}$ and $\psi(x) = 0$ for all $x \notin \mathcal{Y}$. Now consider the differential equation $\dot{x} = \xi(x) f(x)$. Denote the flow of $\dot{x} = \xi(x) f(x)$ starting at x_0 by $\tilde{\phi}_t(x_0)$. Modulo a time reparameterization, the flows $\tilde{\phi}_t(x_0)$ and $\phi_t(x_0)$ are identical up to some finite time. Next define

$$\tilde{B}(x_0) = \int_0^\infty \psi(\tilde{\phi}_t(x_0)) dt.$$

For all x_0 in a neighborhood of $\mathcal{X}$, the flow $\tilde{\phi}_t(x_0)$ is outside of $\mathcal{Y}$ for large t, and thus by its construction $\psi(\tilde{\phi}_t(x_0))$ is equal to zero for large t and for all such x_0. It follows that $\tilde{B}(x)$ is well defined on a neighborhood of $\mathcal{X}$. The function $\tilde{B}(x)$ is continuously differentiable on $\mathcal{X}$ since both $\psi(x)$ and $\tilde{\phi}_t(x)$ are also continuously differentiable. Taking the total derivative of $\tilde{B}(x)$ with respect to time, we obtain

$$\frac{\partial \tilde{B}}{\partial x}(x) \xi(x) f(x) = -\psi(x),$$

which is strictly less than zero, on $\mathcal{X}$. Finally, recall that on $\mathcal{X}$ we have $\xi(x) = 1$. This completes the proof of the proposition. ■

While the above Slater-like condition excludes the possibility of applying Theorem 1 when there is, e.g., an equilibrium point in $\mathcal{X}$, analysis can still be performed by excluding a neighborhood of the equilibrium point from $\mathcal{X}$ in the condition (3). If the excluded region is either backward or forward invariant, and does not intersect $\mathcal{X}_0$ and $\mathcal{X}_u$, then the safety criterion (4)–(6) will still apply in terms of the original sets.

Finally, note also that when *all* the connected components of $\mathbb{R}^n \setminus \mathcal{X}$ are either forward or backward invariant, an even stronger safety criterion can be obtained, as in the following proposition.

Proposition 6. Let the system $\dot{x} = f(x)$ with $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ and the compact sets $\mathcal{X}_0 \subset \mathbb{R}^n$, $\mathcal{X}_u \subset \mathbb{R}^n$ be given, with $0 \notin \mathcal{X}_0 \cup \mathcal{X}_u$. Suppose that the origin is a globally asymptotically stable equilibrium of the system with a global strict Lyapunov function $V(x)$ [2]. Let $\epsilon_1 = \min_{x \in \mathcal{X}_0 \cup \mathcal{X}_u} V(x)$ and $\epsilon_2 = \max_{x \in \mathcal{X}_0 \cup \mathcal{X}_u} V(x)$. Then there exists a function $B \in C^1(\mathbb{R}^n)$ satisfying

$$B(x) \leq 0 \quad \forall x \in \mathcal{X}_0, \tag{8}$$

$$B(x) > 0 \quad \forall x \in \mathcal{X}_u, \tag{9}$$

$$\frac{\partial B}{\partial x}(x) f(x) \leq 0 \quad \forall x \in \{x \in \mathbb{R}^n : \epsilon_1 \leq V(x) \leq \epsilon_2\}, \tag{10}$$

if and only if there exists no trajectory $x(t)$ of the system such that

$$x(0) \in \mathcal{X}_0, \tag{11}$$

$$x(T) \in \mathcal{X}_u \text{ for some } T \geq 0. \tag{12}$$

Proof. Define $\mathcal{X} = \{x \in \mathbb{R}^n : \epsilon_1 \leq V(x) \leq \epsilon_2\}$. In this case, the existence of a function $\tilde{B} \in C^1(\mathbb{R}^n)$ such that $\frac{\partial \tilde{B}}{\partial x}(x) f(x) < 0$ for all $x \in \mathcal{X}$ is guaranteed by Proposition 5, and even the Lyapunov function $V(x)$ can be used as $\tilde{B}(x)$. By Theorem 1, there exists a function $B \in C^1(\mathbb{R}^n)$ satisfying (8)–(10) iff there exists no trajectory $x(t)$ of the system such that $x(0) \in \mathcal{X}_0$, $x(T) \in \mathcal{X}_u$ for some $T \geq 0$, and $x(t) \in \mathcal{X}\ \forall t \in [0, T]$.

Since the connected components of $\mathbb{R}^n \setminus \mathcal{X}$ are either forward or backward invariant, however, there can be no trajectory $x(t)$ of the system and time instants T_1, T_2, T_3 such that $T_1 < T_2 < T_3$, $x(T_1) \in \mathcal{X}$, $x(T_2) \in \mathbb{R}^n \setminus \mathcal{X}$, and $x(T_3) \in \mathcal{X}$. This combined with the fact that $\mathcal{X}_0, \mathcal{X}_u \subseteq \mathcal{X}$ implies that the set of trajectories satisfying $x(0) \in \mathcal{X}_0$, $x(T) \in \mathcal{X}_u$ for some $T \geq 0$, and $x(t) \in \mathcal{X}\ \forall t \in [0, T]$ is the same as the set of trajectories satisfying (11)–(12), and therefore the statement of the proposition follows. ■

[2] That is, $V \in C^1(\mathbb{R}^n)$ is radially unbounded, $V(x) > 0\ \forall x \neq 0$, and $\frac{\partial V}{\partial x}(x) f(x) < 0\ \forall x \neq 0$.

REFERENCES

Alur, R., T. Dang and F. Ivancic (2003). Progress on reachability analysis of hybrid systems using predicate abstraction. In: *Hybrid Systems: Computation and Control, LNCS 2623*. pp. 4–19. Springer-Verlag. Heidelberg.

Aubin, J.-P (1991). *Viability Theory*. Birkhäuser. Boston, MA.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press. Cambridge.

Clarke, Jr., E. M., O. Grumberg and D. A. Peled (2000). *Model Checking*. MIT Press. Cambridge, MA.

Jirstrand, M. (1998). Invariant sets for a class of hybrid systems. In: *Proceedings of the IEEE Conference on Decision and Control*.

Khalil, H. K. (1996). *Nonlinear Systems*. second ed.. Prentice-Hall, Inc.. Upper Saddle River, NJ.

Kurzhanski, A. and P. Varaiya (2000). Ellipsoidal techniques for reachability analysis. In: *Hybrid Systems: Computation and Control, LNCS 1790*. pp. 203–213. Springer-Verlag. Heidelberg.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. John Wiley & Sons. New York, NY.

Papadimitriou, C. H. and K. Steiglitz (1998). *Combinatorial Optimization: Algorithms and Complexity*. Dover Publications Inc.. Mineola, NY.

Prajna, S., A. Jadbabaie and G. J. Pappas (2004). Stochastic safety verification using barrier certificates. In: *Proceedings of the IEEE Conference on Decision and Control*.

Prajna, S., A. Papachristodoulou and P. A. Parrilo (2002). Introducing SOSTOOLS: A general purpose sum of squares programming solver. In: *Proceedings of the IEEE Conference on Decision and Control*. Available at http://www.cds.caltech.edu/sostools and http://www.mit.edu/~parrilo/sostools.

Prajna, S. and A. Jadbabaie (2004). Safety verification of hybrid systems using barrier certificates. In: *Hybrid Systems: Computation and Control, LNCS 2993*. pp. 477–492. Springer-Verlag. Heidelberg.

Prajna, S. and A. Rantzer (2005). Primal–dual tests for safety and reachability. In: *Hybrid Systems: Computation and Control, LNCS 3414*. pp. 542–556. Springer-Verlag. Heidelberg.

Rantzer, A. (2001). A dual to Lyapunov's stability theorem. *Systems and Control Letters* **42**(3), 161–168.

Rantzer, A. and S. Hedlund (2003). Duality between cost and density in optimal control. In: *Proceedings of the IEEE Conference on Decision and Control*.