SIAM J. CONTROL OPTIM.
Vol. 46, No. 3, pp. 999–1021

CONVEX PROGRAMS FOR TEMPORAL VERIFICATION OF NONLINEAR DYNAMICAL SYSTEMS*

STEPHEN PRAJNA† AND ANDERS RANTZER‡

Abstract. A methodology for safety verification of continuous and hybrid systems using barrier certificates has been proposed recently. Conditions that must be satisfied by a barrier certificate can be formulated as a convex program, and the feasibility of the program implies system safety in the sense that there is no trajectory starting from a given set of initial states that reaches a given unsafe region. The dual of this problem, i.e., the reachability problem, concerns proving the existence of a trajectory starting from the initial set that reaches another given set. Using insights from the linear programming duality appearing in the discrete shortest path problem, we show in this paper that reachability of continuous systems can also be verified through convex programming. Several convex programs for verifying safety and reachability, as well as other temporal properties such as eventuality, avoidance, and their combinations, are formulated. Some examples are provided to illustrate the application of the proposed methods. Finally, we exploit the convexity of our methods to derive a converse theorem for safety verification using barrier certificates.

Key words. temporal verification, safety verification, reachability analysis, barrier certificate, density function, convex programming, duality

AMS subject classifications. 93C10, 68Q60, 90C90

DOI. 10.1137/050645178

1. Introduction.

Consider a continuous-time dynamical system of the form

$$\dot{x}(t) = f(x(t)),$$

where $x(t)$ is the state of the system, taking its value in the set $\mathcal{X} \subseteq \mathbb{R}^n$. Also given are the set of possible initial states $\mathcal{X}_0 \subseteq \mathcal{X}$, the set of "bad" states $\mathcal{X}_u \subseteq \mathcal{X}$, and the set of "good" states $\mathcal{X}_r \subseteq \mathcal{X}$. In this paper, we will be concerned with methods for verifying or proving temporal properties of the system such as the following:

- *safety*: all trajectories of the system starting from $\mathcal{X}_0$ will never reach $\mathcal{X}_u$;
- *avoidance*: at least one trajectory of the system starting from $\mathcal{X}_0$ will never reach $\mathcal{X}_u$;
- *eventuality*: all trajectories of the system starting from $\mathcal{X}_0$ will reach $\mathcal{X}_r$ in finite time;
- *reachability*: at least one trajectory of the system starting from $\mathcal{X}_0$ will reach $\mathcal{X}_r$ in finite time.

They will be defined more precisely later in the paper. In addition, we will look at more complex temporal properties, which are the combinations of the above, and will also consider systems with uncertain time-varying disturbance inputs.

When the system under consideration is a discrete transition system, such as a finite automaton, the problem described above has been studied extensively in the

*Received by the editors November 15, 2005; accepted for publication (in revised form) September 26, 2006; published electronically June 29, 2007. Preliminary versions of this paper appeared in *Hybrid Systems: Computation and Control* 2005 and *Proceedings of the IFAC World Congress* 2005.

http://www.siam.org/journals/sicon/46-3/64517.html

†Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125 (prajna@cds.caltech.edu). Current address: Credit Suisse, One Cabot Square, London E14 4QJ, United Kingdom.

‡Automatic Control LTH, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden (rantzer@control.lth.se).

computer science literature, and has applications, e.g., in the verification of correctness of computer protocols, algorithms, and software. See [8, 10, 12, 18]. The methods that have been proposed fall into two mainstream approaches: *model checking* [8] and *deductive verification* (or *theorem proving*) [18]. Model checking performs an exhaustive exploration of all possible system behaviors in a fully automated way, but is applicable only to finite state systems. Deductive verification, on the other hand, verifies system properties through formal deduction based on a set of inference rules. It is applicable to infinite state systems, but has a drawback in the sense that guidance from users is often needed in the process.

Uncountable state space and continuous dynamics are introduced when we consider applications in control, since they usually involve physical plants whose dynamics is governed by differential equations. Here the need for temporal verification arises as the complexity of the system increases, especially in safety-critical applications such as air traffic management [29], automated highway systems [11], and life support systems [9]. For such systems, exact verification cannot be performed through simulation, due to the infinite number of possibilities taken by the continuous state and also the uncertainties of the system.

The success of model checking techniques in verification of discrete, finite state transition systems has prompted the development of analogous approaches for continuous and hybrid systems. These approaches (see, e.g., [1, 2, 3, 5, 7, 15, 16, 30, 31]) require explicit computation of the states reachable from the initial set, which, for example, is performed by propagating the set of states. Unfortunately, although they allow us to compute an exact or nearly exact approximation of reachable sets, it is very difficult to perform such a computation due to the uncountability of the state space, especially when the system is nonlinear and uncertain. Note also that most of the existing literature focuses on the verification of safety property, although some of their techniques can be used to verify other temporal properties stated at the beginning of this paper.

In a different vein, a deductive method for safety verification that does not require explicit computation of the reachable sets, but instead is based on functions of states termed barrier certificates, has been recently proposed in a work by the first author [21]. The idea here is to study properties of the system without the need to compute the flow explicitly. Our conditions for safety can be stated as follows. Suppose that the vector field $f(x)$ is continuous and that there exists a continuously differentiable function $B : \mathbb{R}^n \to \mathbb{R}$ such that the inequalities

$$B(x) \leq 0 \quad \forall x \in \mathcal{X}_0, \tag{1.1}$$

$$B(x) > 0 \quad \forall x \in \mathcal{X}_u, \tag{1.2}$$

$$\frac{\partial B}{\partial x}(x) f(x) \leq 0 \quad \forall x \in \mathcal{X} \tag{1.3}$$

are satisfied. Then the safety property is verified, namely, there is no trajectory $x(t)$ of the system $\dot{x} = f(x)$ 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 $t \in [0, T]$. The function $B(x)$ here is called a *barrier certificate*. When $f(x)$ is polynomial and the sets $\mathcal{X}$, $\mathcal{X}_0$, $\mathcal{X}_u$ are semialgebraic, a polynomial barrier certificate $B(x)$ can be efficiently searched using sum of squares programming [22]—a convex optimization framework based on sum of squares decompositions of multivariate polynomials [20] and semidefinite programming [6]. Because of this, our method appears to be more scalable than many other methods. The method has also

been extended to handle hybrid, uncertain, and stochastic systems [21], and successful application to a NASA life support system, which is a nonlinear hybrid system with 6 discrete modes and 10 continuous state variables, has been reported [9]. To the best of our knowledge, all other verification methods that can handle nonlinear hybrid systems are practically limited to about 5 continuous state variables.

The above method is analogous to the Lyapunov method for stability analysis [14]. Contrary to stability analysis, however, no notion of equilibrium, stability, or convergence is required in temporal verification. For example, the system does not need to have an equilibrium, and also for the eventuality and reachability properties the system is not required to stay in $\mathcal{X}_r$ once the set is reached. Our method is also related to the viability theory [4], the smallest invariant set [13], and the invariant generation [26, 27, 28] approaches to safety verification. However, one of the distinctive features of our approach is that we use convex programming to verify properties of interest, which gives benefit in terms of *computation* and in terms of its inherent *duality structure*.

In the present paper, we use insights from the linear programming duality appearing in the discrete shortest path problem [19] and the concept of density function [23, 24] to formulate a convex program for proving reachability. In fact, not only safety and reachability, but also other temporal properties such as eventuality, avoidance, and their combinations can be verified through convex programming. Several convex programs for this will be formulated. Similar to before, when the description of the system is polynomial and the sets are semialgebraic, polynomial solutions to these programs can be searched using sum of squares programming. In addition to this, we will exploit strong duality to prove a converse theorem for safety verification using barrier certificates.

The outline of the paper is as follows. In section 2, we give an intuitive illustration of some main ideas by addressing the verification of a simple discrete transition system. The convex programs for verification of continuous-time systems are presented and proven in section 3. In section 4, some examples will be presented to illustrate the applications of the proposed method. A converse theorem for barrier certificates will be stated and proven in section 5, and we offer some conclusions in section 6.

2. Discrete example. Let us consider the verification of a simple discrete transition system, shown in Figure 2.1. The system has four states, labeled 1 through 4, and three transitions between states, represented by the directed edges in the graph. We assume that node 1 is the initial state and node 4 is the bad/unsafe state.

For verifying the safety property, conditions analogous to (1.1)–(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)[1]

$$\max\ s^T B$$

$$\text{subject to } A^T B \leq 0,$$

where $B \triangleq \mathrm{col}(B_1, B_2, B_3, B_4) \in \mathbb{R}^4$ is the decision variable of the LP (i.e., the barrier

[1]Here we assume that there are only one initial state and one unsafe state. A generalization of this can be formulated by considering a bigger graph obtained by augmenting an extra "source" node and edges that connect it to all initial states, as well as an extra "sink" node and edges that connect all unsafe states to it.

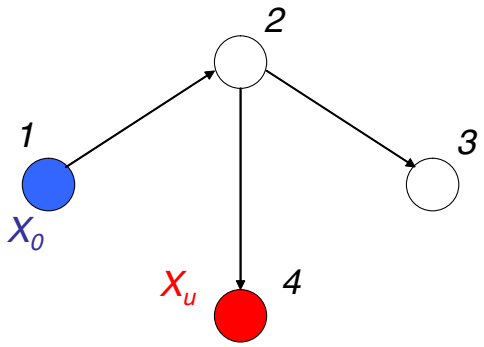


FIG. 2.1. *Verification of a simple discrete transition system. The nodes represent the states of the system, while the directed edges represent transitions between states. Node* 1 *is the initial state and node* 4 *is the unsafe state.*

certificate); A is the incidence matrix of the graph, in this case given by

$$A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}^T ;$$

and s is a 4×1 column vector whose ith entry is equal to 1 if the ith node is the unsafe state, and equal to -1 if the ith node is the initial state. This formulation is similar to the continuous case. Analogous to (1.3), we ask that $B_j \leq B_i$ if there is a directed edge from node i to node j. The objective function of the LP is just the difference between the values of B at the unsafe state and at the initial state. If there is a feasible solution to the above LP such that the objective function is strictly positive, then the safety property can be inferred; i.e., we prove that there is no path going from node 1 to node 4.

The dual of the above LP is as follows:

$$\begin{aligned} &\min 0 \\ &\text{subject to } A\rho = s, \\ &\qquad \rho \geq 0, \end{aligned}$$

where $\rho \triangleq \text{col}(\rho_{12}, \rho_{23}, \rho_{24}) \in \mathbb{R}^3$ is the dual decision variables, whose entries correspond to the edges in the graph. 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, namely, that the total flow into a node is equal to the total flow out. In addition, the first and third 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., [19] and the 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 [19].

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 [6] 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. In the continuous case, a strong duality argument will also be used to prove a converse theorem for barrier certificates later in this paper.

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. If the direction of the edge from node 2 to node 4 were reversed, for example, the optimal objective value of the corresponding primal LP would be ∞, and there would be no feasible solution to the dual LP.

Other properties of this discrete transition system such as eventuality and avoidance can also be verified by solving some appropriate LPs. We will not state them here, but instead we will now proceed to discuss the corresponding convex programs for continuous systems.

3. Continuous systems. We denote the space of m-times continuously differentiable functions mapping $X \subseteq \mathbb{R}^n$ to $\mathbb{R}^p$ by $C^m(X, \mathbb{R}^p)$. When $p = 1$, we will simply write $C^m(X)$, and for continuous functions ($m = 0$), we will omit the superscript. The solution $x(t)$ of $\dot{x} = f(x)$ starting from $x(0) = x_0$, if unique, is denoted by $\phi_t(x_0)$. For a set Z, we define $\phi_t(Z) \triangleq \{\phi_t(x) : x \in Z\}$.

The divergence of a vector field $f \in C^1(X, \mathbb{R}^n)$ is denoted by $\nabla \cdot f(x)$. Finally, let $\mathrm{cl}(X)$ denote the closure of a set X, and let ∂X denote the boundary of X.

The following version of Liouville's theorem (from [23]) will be used in the proofs of the main theorems.

LEMMA 3.1. *Let $f \in C^1(D, \mathbb{R}^n)$, where $D \subseteq \mathbb{R}^n$ is open, and let $\rho \in C^1(D, \mathbb{R})$ be integrable, i.e., $\int_D \rho(x)dx$ is finite. Consider the system $\dot{x} = f(x)$. For a measurable set Z, assume that $\phi_\tau(Z)$ is a subset of D for all τ between 0 and T. Then*

$$\int_{\phi_T(Z)} \rho(x)dx - \int_Z \rho(z)dz = \int_0^T \int_{\phi_\tau(Z)} \left[\nabla \cdot (f\rho)\right](x)dxd\tau. \tag{3.1}$$

3.1. Safety and reachability verification. At this point, we are ready to state and prove the first pair of convex programs that verify safety and reachability for continuous systems. The first convex program was proposed in [21] but will be repeated here for completeness.

THEOREM 3.2. *Consider the system $\dot{x} = f(x)$ with $f \in C(\mathbb{R}^n, \mathbb{R}^n)$. Let the sets $\mathcal{X} \subseteq \mathbb{R}^n$, $\mathcal{X}_0 \subseteq \mathcal{X}$, and $\mathcal{X}_u \subseteq \mathcal{X}$ be given. Suppose that there exists a function $B \in C^1(\mathbb{R}^n)$ satisfying*

$$B(x) \leq 0 \quad \forall x \in \mathcal{X}_0, \tag{3.2}$$

$$B(x) > 0 \quad \forall x \in \mathcal{X}_u, \tag{3.3}$$

$$\frac{\partial B}{\partial x}(x) f(x) \leq 0 \quad \forall x \in \mathcal{X}. \tag{3.4}$$

Then the safety property holds; i.e., 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 $t \in [0, T]$.

Proof. Our proof is by contradiction. Assume that there exists a barrier certificate $B(x)$ satisfying conditions (3.2)–(3.4), while at the same time the system is not safe; i.e., there exist a time instance $T \geq 0$ and an initial condition $x_0 \in \mathcal{X}_0$ such that a trajectory $x(t)$ of the system 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 (3.4) implies that the derivative of $B(x(t))$ with respect to time is nonpositive on the time interval $[0, T]$. A direct consequence of this (which, for example, can be shown using the mean value theorem) is that $B(x(T))$ must be less than or equal to $B(x(0))$, which is contradictory to (3.2)–(3.3). Thus the initial hypothesis is not correct: the safety property must hold. ∎

We will next present a convex program for verifying reachability. It can be viewed as a continuous-time version of the dual LP in section 2. The decision variable $\rho(x)$ in this convex program is termed *density function* and has an interpretation as the stationary density of a substrate that is generated and consumed in various parts of the state space, and that is transported according to the vector field of the system. It has been previously used in an almost global stability criterion in [23].

THEOREM 3.3. *Consider the system $\dot{x} = f(x)$ with $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$. Let the sets $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{X}_0 \subseteq \mathcal{X}$, and $\mathcal{X}_r \subseteq \mathcal{X}$ be given. Assume that the sets are bounded and that $\mathcal{X}_0$ has a nonempty interior. If there exists a function $\rho \in C^1(\mathbb{R}^n)$ satisfying*

$$\int_{\mathcal{X}_0} \rho(x)dx \geq 0, \tag{3.5}$$

$$\rho(x) < 0 \quad \forall x \in \mathrm{cl}(\partial\mathcal{X} \setminus \partial\mathcal{X}_r), \tag{3.6}$$

$$\nabla \cdot (\rho f)(x) > 0 \quad \forall x \in \mathrm{cl}(\mathcal{X} \setminus \mathcal{X}_r), \tag{3.7}$$

then the reachability property holds; i.e., there exists a trajectory $x(t)$ of the system such that $x(0) \in \mathcal{X}_0$, $x(T) \in \mathcal{X}_r$ for some $T \geq 0$, and $x(t) \in \mathcal{X}$ for all $t \in [0, T]$.

Proof. Let $X \subseteq \mathcal{X}_0$ be an open set on which $\rho(x) \geq 0$. We will first prove that there must be an initial condition $x_0 \in X$ whose flow $\phi_t(x_0)$ leaves $\mathcal{X} \setminus \mathcal{X}_r$ in finite time. In fact, the set of all initial conditions in X whose flows do not leave $\mathcal{X} \setminus \mathcal{X}_r$ in finite time is a set of measure zero. To show this, let Y be an open neighborhood of $\mathcal{X} \setminus \mathcal{X}_r$ such that $\nabla \cdot (\rho f)(x) > 0$ on $\mathrm{cl}(Y)$. Now define

$$Z = \bigcap_{i=1,2,\ldots} \{x_0 \in X : \phi_t(x_0) \in Y \quad \forall t \in [0, i]\}.$$

The set Z is an intersection of countable open sets and hence is measurable. It contains all initial conditions in X for which the trajectories stay in Y for all $t \geq 0$. That Z is a set of measure zero can be shown using Lemma 3.1 as follows. Since $\phi_t(Z) \subset Y$, Y is bounded, and $\rho(x)$ is continuous, the left-hand side of

$$\int_{\phi_t(Z)} \rho(x)dx - \int_Z \rho(x)dx = \int_0^t \int_{\phi_\tau(Z)} \left[\nabla \cdot (f\rho)\right](x)dxd\tau$$

is bounded for all $t \geq 0$. Therefore, for the above equation to hold, we must have $\int_{\phi_\tau(Z)} \left[\nabla \cdot (f\rho)\right](x)dx \to 0$ as $\tau \to \infty$, or, equivalently, the measure of $\phi_\tau(Z)$ converges to zero as $\tau \to \infty$. Suppose now that Z has nonzero measure. We have a contradiction since $\lim_{t\to\infty} \int_{\phi_t(Z)} \rho(x)dx = 0$, whereas $\lim_{t\to\infty} \int_0^t \int_{\phi_\tau(Z)} \left[\nabla \cdot (f\rho)\right](x)dxd\tau +$

$\int_Z \rho(x)dx$ is strictly positive, as implied by (3.5) and (3.7). Using this argument, we conclude that Z has measure zero. Since $\mathcal{X} \setminus \mathcal{X}_r \subset Y$, it follows immediately that the set of all initial conditions in X whose flows stay in $\mathcal{X} \setminus \mathcal{X}_r$ for all time is a set of measure zero.

Now take any $x_0 \in X$ whose flow leaves $\mathcal{X} \setminus \mathcal{X}_r$ in finite time; we will show that such a flow must enter $\mathcal{X}_r$ before leaving $\mathcal{X}$. Suppose to the contrary that the flow $\phi_t(x_0)$ leaves $\mathcal{X}$ without entering $\mathcal{X}_r$ first. Let $T > 0$ be the "first" time instant $\phi_t(x_0)$ leaves $\mathcal{X}$. By this we mean that either $\phi_t(x_0) \in \mathcal{X} \setminus \mathcal{X}_r$ for all $t \in [0, T)$ and $\phi_T(x_0) \notin \mathcal{X}$, or $\phi_t(x_0) \in \mathcal{X} \setminus \mathcal{X}_r$ for all $t \in [0, T]$ and $\phi_{T+\epsilon}(x_0) \notin \mathcal{X}$ for any $\epsilon > 0$. From conditions (3.6)–(3.7), it follows that for a sufficiently small neighborhood U of x_0 we have

$$\begin{aligned} \rho(x) &\geq 0 \quad \forall x \in U, \\ \rho(x) &< 0 \quad \forall x \in \phi_T(U), \\ \nabla \cdot (\rho f)(x) &> 0 \quad \forall x \in \phi_t(U), t \in [0, T]. \end{aligned}$$

Apply Lemma 3.1 again to obtain a contradiction. According to the above, the left-hand side of

$$\int_{\phi_T(U)} \rho(x)dx - \int_U \rho(x)dx = \int_0^T \int_{\phi_\tau(U)} \left[\nabla \cdot (f\rho)\right](x)dxd\tau$$

is negative, while the right-hand side is positive. Thus there is a contradiction, and we conclude that for $x(0) = x_0$ there must exist $T \geq 0$ such that $x(T) \in \mathcal{X}_r$ and $x(t) \in \mathcal{X}$ for all $t \in [0, T]$. □

Remark 3.4. Modulo the following modification on the assertion of Theorem 3.3, the conclusion will still hold even when the sets are not bounded. In particular, we need to add the condition that $\rho(x)$ is integrable on $\mathcal{X}$ (i.e., $\int_{\mathcal{X}} \rho(x)dx$ is finite) and replace (3.7) by

$$\nabla \cdot (\rho f)(x) \geq \epsilon \quad \forall x \in (\mathcal{X} \setminus \mathcal{X}_r)$$

for a positive number ϵ.

Notice that all the conditions presented in the above theorems (as well as in the theorems that will be presented later) form convex programming problems, as the sets of $B(x)$'s satisfying (3.2)–(3.4) or $\rho(x)$'s satisfying (3.5)–(3.7) are convex. This just follows from the definition of convexity. For example, if $B_1(x)$ and $B_2(x)$ satisfy (3.2)–(3.4), then for any $\alpha \in [0, 1]$, the function $\alpha B_1(x) + (1-\alpha)B_2(x)$ also satisfies the conditions. The convexity of the programs opens the possibility of computing $B(x)$ and $\rho(x)$ using convex optimization. For systems whose vector fields are polynomial and whose set descriptions are semialgebraic (i.e., described by polynomial equalities and inequalities), a computational method called sum of squares programming can be utilized if we use a polynomial parameterization for $B(x)$ or $\rho(x)$. The method is based on the sum of squares decomposition of multivariate polynomials [20] and semidefinite programming [6]. Software tools [22] are helpful for this purpose. See [21] for details.

When we set $\mathcal{X}_u = \mathcal{X}_r$, the convex programs in Theorems 3.2 and 3.3 form a pair of *weak alternatives*: at most one of them can be feasible. Nevertheless, strictly speaking it should be noted that these convex programs are not pairs of *Lagrange dual* problems [6] in the sense of convex optimization. We deliberately do not use Lagrange dual

problems to avoid computational problems when we postulate $B(x)$ or $\rho(x)$ as polynomials. For example, the Lagrange dual problem of the safety test in Theorem 3.2 will require $\nabla \cdot (\rho f)(x)$ to be zero on $\mathcal{X} \setminus (\mathcal{X}_0 \cup \mathcal{X}_u)$ (cf. section 5). Although useful for theoretical purposes, this will hinder the computation of $\rho(x)$ through polynomial parameterization and sum of squares programming. In this regard, some interesting future directions would be to see if a pair of Lagrange dual problems can be formulated so that both problems can be solved using sum of squares programming, or, more importantly, to see if the dual infeasibility certificate of one convex program can be interpreted directly as a feasible solution to the dual convex program.

3.2. Eventuality and avoidance verification. We will now present two other convex programs for verifying the eventuality and avoidance properties. Analogous to what we have in the previous subsection, when $\mathcal{X}_u = \mathcal{X}_r$ these programs form a pair of weak alternatives.

THEOREM 3.5. *Consider the system $\dot{x} = f(x)$ with $f \in C(\mathbb{R}^n, \mathbb{R}^n)$. Let $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{X}_0 \subseteq \mathcal{X}$, and $\mathcal{X}_r \subseteq \mathcal{X}$ be bounded sets. If there exists a function $B \in C^1(\mathbb{R}^n)$ satisfying*

$$B(x) \leq 0 \quad \forall x \in \mathcal{X}_0, \tag{3.8}$$

$$B(x) > 0 \quad \forall x \in \mathrm{cl}(\partial\mathcal{X} \setminus \partial\mathcal{X}_r), \tag{3.9}$$

$$\frac{\partial B}{\partial x}(x) f(x) < 0 \quad \forall x \in \mathrm{cl}(\mathcal{X} \setminus \mathcal{X}_r), \tag{3.10}$$

then the eventuality property holds; i.e., for all initial conditions $x_0 \in \mathcal{X}_0$, the trajectory $x(t)$ of the system starting at $x(0) = x_0$ satisfies $x(T) \in \mathcal{X}_r$ and $x(t) \in \mathcal{X}$ for all $t \in [0, T]$ for some $T \geq 0$.

Proof. Consider any point $x_0 \in \mathcal{X}_0$, for which $B(x_0) \leq 0$, and let $x(t)$ be a trajectory of the system starting at $x(0) = x_0$. The trajectory $x(t)$ must leave $\mathcal{X} \setminus \mathcal{X}_r$ in finite time, since the derivative inequality (3.10) holds and $B(x)$ is bounded below on $\mathcal{X}$. Now, suppose that $x(t)$ leaves $\mathcal{X}$ without entering $\mathcal{X}_r$ first, and consider the "first" time instant $t = T$ at which it happens. Similar to the proof of Theorem 3.3, by this we mean that either $x(t) \in \mathcal{X} \setminus \mathcal{X}_r$ for all $t \in [0, T)$ and $x(T) \notin \mathcal{X}$, or $x(t) \in \mathcal{X} \setminus \mathcal{X}_r$ for all $t \in [0, T]$ and $x(T + \epsilon) \notin \mathcal{X}$ for any $\epsilon > 0$. From (3.10) and $B(x_0) \leq 0$, it follows that $B(x(T))$ is nonpositive, which is contradictory to (3.9). Thus we conclude that for any trajectory $x(t)$ starting at $x(0) = x_0$ there must exist $T \geq 0$ such that $x(T) \in \mathcal{X}_r$ and $x(t) \in \mathcal{X}$ for all $t \in [0, T]$. Since x_0 is an arbitrary point in $\mathcal{X}_0$, the conclusion of the theorem follows. ∎

Remark 3.6. Similarly to before, with some modifications to the assertion of the theorem, the conclusion of Theorem 3.5 will still hold even when the sets are not bounded. In particular, we need to add the condition that $B(x)$ is bounded below on $\mathcal{X}$ and replace (3.10) by

$$\frac{\partial B}{\partial x}(x) f(x) \leq -\epsilon \quad \forall x \in (\mathcal{X} \setminus \mathcal{X}_r)$$

for a positive number ϵ.

THEOREM 3.7. *Consider the system $\dot{x} = f(x)$ with $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$. Let $\mathcal{X} \subseteq \mathbb{R}^n$, $\mathcal{X}_0 \subseteq \mathcal{X}$, and $\mathcal{X}_u \subseteq \mathcal{X}$ be some given sets, with $\mathcal{X}_0$ having a nonempty interior. If*

there exist an open set $\tilde{\mathcal{X}}$ *and a function* $\rho \in C^1(\mathbb{R}^n)$ *such that* $\mathcal{X} \subseteq \tilde{\mathcal{X}}$ *and*

$$\int_{\mathcal{X}_0} \rho(x)dx \geq 0, \tag{3.11}$$

$$\rho(x) < 0 \quad \forall x \in \mathcal{X}_u, \tag{3.12}$$

$$\nabla \cdot (\rho f)(x) \geq 0 \quad \forall x \in \tilde{\mathcal{X}}, \tag{3.13}$$

then the avoidance property holds; i.e., for some initial condition $x_0 \in \mathcal{X}_0$*, there exists no* $T \geq 0$ *such that trajectory* $x(t)$ *of the system starting at* $x(0) = x_0$ *satisfies* $x(T) \in \mathcal{X}_u$ *and* $x(t) \in \mathcal{X}$ *for all* $t \in [0, T]$.

Proof. From (3.11), it follows that there exists an open set $X \subseteq \mathcal{X}_0$ on which $\rho(x) \geq 0$. Take any x_0 in X—we will show that the trajectory starting from this point will never reach $\mathcal{X}_u$. Suppose to the contrary that there exists a $T \geq 0$ such that $\phi_T(x_0) \in \mathcal{X}_u$ and $\phi_t(x_0) \in \mathcal{X}$ for $t \in [0, T]$. Then it follows from (3.12)–(3.13) that for a sufficiently small neighborhood Z of x_0 we have

$$\rho(x) \geq 0 \quad \forall x \in Z,$$

$$\rho(x) < 0 \quad \forall x \in \phi_T(Z),$$

$$\nabla \cdot (\rho f)(x) \geq 0 \quad \forall x \in \phi_t(Z), t \in [0, T].$$

Now apply Lemma 3.1 to obtain a contradiction. Use a bounded but sufficiently large $D \subset \mathbb{R}^n$ such that $\phi_t(Z) \subset D$ for all $t \in [0, T]$; then $\rho(x)$ is integrable on D. According to the above, the left-hand side of

$$\int_{\phi_T(Z)} \rho(x)dx - \int_Z \rho(x)dx = \int_0^T \int_{\phi_\tau(Z)} \left[\nabla \cdot (f\rho)\right](x)dxd\tau$$

is negative and the right-hand side is nonnegative. Hence there is a contradiction and the proof is complete. ∎

In applications where the system has stable equilibrium points, it is often convenient to exclude a neighborhood of the equilibria from the region where the divergence inequality (3.13) must be satisfied, since the inequality is otherwise impossible to satisfy without a singularity in $\rho(x)$. This does not make the conclusion of the theorem weaker, as long as the excluded set does not intersect $\mathcal{X}_u$ and is entirely surrounded by a region of positive $\rho(x)$.

Similarly, the Lie derivative inequality (3.10) is impossible to satisfy when the system has equilibrium points in $\mathcal{X} \setminus \mathcal{X}_r$. In this case, a neighborhood of the equilibria should also be excluded from the region where the inequality is to be satisfied. The conclusion of the theorem is still valid as long as the excluded set is entirely surrounded by a region of positive $B(x)$.

3.3. Some extensions. Whereas the convex programs for safety and reachability as well as eventuality and avoidance are related since they form pairs of weak alternatives, the safety property is also related to avoidance, and eventuality to reachability, via replacing the universal quantifier with an existential quantifier. As a consequence, reachability and avoidance verification can also be performed using the

barrier certificate $B(x)$. The conditions are similar to those in Theorems 3.5 and 3.2, respectively, but with conditions (3.8) and (3.2) replaced by

$$\int_{\mathcal{X}_0} B(x)dx \leq 0,$$

where we also ask that $\mathcal{X}_0$ has a nonempty interior. The proof is similar to the proofs of Theorems 3.5 and 3.2, noting that $B(x)$ will then be less than or equal to zero in some open set contained in $\mathcal{X}_0$.

A modification of the convex program involving $\rho(x)$ in Theorem 3.7 can also be used to verify the safety property. For this, we need to replace (3.11) by

$$\rho(x) \geq 0 \quad \forall x \in \tilde{\mathcal{X}}_0,$$

where $\tilde{\mathcal{X}}_0$ is an open set containing $\mathcal{X}_0$. Note that in this case $\mathcal{X}_0$ no longer needs to have a nonempty interior.

On the other hand, an analogous modification of Theorem 3.3 can only be used to verify the eventuality property in the *weak* sense: that *almost all* trajectories of the system starting from $\mathcal{X}_0$ will reach $\mathcal{X}_r$ in finite time. This is stated in the corollary below.

COROLLARY 3.8. *Consider the system $\dot{x} = f(x)$ with $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$. Let the sets $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{X}_0 \subseteq \mathcal{X}$, and $\mathcal{X}_r \subseteq \mathcal{X}$ be given. Assume that the sets are bounded, and let $\tilde{\mathcal{X}}_0$ be an open set containing $\mathcal{X}_0$. If there exists a function $\rho \in C^1(\mathbb{R}^n)$ satisfying*

$$\rho(x) \geq 0 \quad \forall x \in \tilde{\mathcal{X}}_0, \tag{3.14}$$

$$\rho(x) < 0 \quad \forall x \in \mathrm{cl}(\partial\mathcal{X} \setminus \partial\mathcal{X}_r), \tag{3.15}$$

$$\nabla \cdot (\rho f)(x) > 0 \quad \forall x \in \mathrm{cl}(\mathcal{X} \setminus \mathcal{X}_r), \tag{3.16}$$

then the weak eventuality property holds; i.e., for almost all[2] *initial conditions $x_0 \in \mathcal{X}_0$, there exists $T \geq 0$ such that the trajectory $x(t)$ of the system starting at $x(0) = x_0$ satisfies $x(T) \in \mathcal{X}_r$ and $x(t) \in \mathcal{X}$ for all $t \in [0, T]$.*

Proof. Using an argument similar to the proof of Theorem 3.3, it can be shown that for almost all initial conditions $x_0 \in \tilde{\mathcal{X}}_0$, there exists $T \geq 0$ such that the trajectory $x(t)$ of the system starting at $x(0) = x_0$ satisfies $x(T) \in \mathcal{X}_r$ and $x(t) \in \mathcal{X}$ for all $t \in [0, T]$. Since $\mathcal{X}_0 \subseteq \tilde{\mathcal{X}}_0$, the corollary follows. ∎

Example 3.9. To show that the weak eventuality property mentioned above cannot in general be strengthened to eventuality, consider the system $\dot{x} = x$, with $\mathcal{X} = (-5, 5) \subset \mathbb{R}$, $\mathcal{X}_0 = (-1, 1)$, $\mathcal{X}_r = (-5, -4) \cup (4, 5)$. The function $\rho(x) = 1$ satisfies all the conditions that guarantee weak eventuality; hence almost all trajectories starting from $\mathcal{X}_0$ will reach $\mathcal{X}_r$ in finite time. The only exception in this case is the trajectory $x(t) = 0$.

Let us now consider the verification of a system with disturbance input $\dot{x} = f(x, d)$, where $f \in C(\mathbb{R}^{n+m}, \mathbb{R}^n)$; the disturbance signal $d(t)$ is assumed to be piecewise continuous, bounded on any finite time interval, and take its value in a set $\mathcal{D} \subseteq \mathbb{R}^m$. Then solving the convex program in Theorem 3.2 with the Lie derivative inequality (3.4) replaced by

$$\frac{\partial B}{\partial x}(x) f(x, d) \leq 0 \quad \forall (x, d) \in \mathcal{X} \times \mathcal{D}$$

[2]This is in the sense that the set of initial conditions which do not satisfy the property is a set of measure zero.

will prove safety under all possible disturbances $d(t)$. Also, solving the convex program in Theorem 3.5 with the Lie derivative inequality (3.10) replaced by

$$\frac{\partial B}{\partial x}(x)f(x,d) \le -\epsilon \quad \forall (x,d) \in (\mathcal{X} \setminus \mathcal{X}_r) \times \mathcal{D}$$

for some positive ϵ will prove eventuality under all possible disturbances $d(t)$. Similar adaptations can be applied to the convex programs that verify reachability and avoidance using $B(x)$.

At present, it is unclear how a similar worst-case analysis for systems with time-varying disturbance can be formulated using $\rho(x)$. However, as pointed out in [23], the density function $\rho(x)$ seems to have a better convexity property that is more beneficial for controller design. For a system $\dot{x} = f(x) + g(x)u(x)$, where $u(x)$ is the control input (assumed to be in a state feedback form), the inequalities (3.5)–(3.6) and

$$\nabla \cdot [\rho(f+ug)](x) > 0 \quad \forall x \in \mathrm{cl}(\mathcal{X} \setminus \mathcal{X}_r)$$

(and similarly for (3.11)–(3.13)) are certainly convex conditions on the pair $(\rho, \rho u)$. It is therefore natural to introduce $\psi = \rho u$ as a search variable and use convex optimization to find a feasible pair (ρ, ψ), then recover the control law as $u(x) = \psi(x)/\rho(x)$. Some results along this direction are available in [25].

While one may argue that the reachability and avoidance properties can be shown by running a numerical simulation of $\dot{x} = f(x)$ starting from a properly chosen $x_0 \in \mathcal{X}_0$, the merit of the convex programming tests presented before is twofold. First, a solution to the convex programs for reachability or avoidance will automatically indicate a set from which all points (or almost all points) can be chosen as the initial state. Second, the use of these convex programs allows us to also consider the verification of systems with disturbance (which obviously cannot be performed using simulation), or even the controller design problem, as we have seen above.

3.4. Other temporal properties. It is clear that the convex programs in the previous subsections can also be extended to prove combined temporal properties such as reachability–safety:

> there exists a trajectory $x(t)$ such that $x(0) \in \mathcal{X}_0$, $x(T) \in \mathcal{X}_r$ for some $T \ge 0$, and $x(t) \notin \mathcal{X}_u$, $x(t) \in \mathcal{X}$ for all $t \in [0,T]$;

and eventuality–safety[3] (or weak eventuality–safety):

> for all (or almost all) initial states $x_0 \in \mathcal{X}_0$, the trajectory $x(t)$ starting at $x(0) = x_0$ will satisfy $x(T) \in \mathcal{X}_r$ for some $T \ge 0$ and $x(t) \notin \mathcal{X}_u$, $x(t) \in \mathcal{X}$ for all $t \in [0,T]$.

Note that for the above temporal specifications, the system can reach $\mathcal{X}_u$ *after* it reaches $\mathcal{X}_r$ first.

For instance, convex programs for verifying the eventuality–safety and weak eventuality–safety properties are stated in the following corollaries.

COROLLARY 3.10. *Consider the system $\dot{x} = f(x)$ with $f \in C(\mathbb{R}^n, \mathbb{R}^n)$ and let $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{X}_0 \subseteq \mathcal{X}$, $\mathcal{X}_u \subseteq \mathcal{X}$, $\mathcal{X}_r \subseteq \mathcal{X}$ be bounded. Suppose that there exists a function*

[3] In linear temporal logic (LTL), for example, this property corresponds to the "until" operator.

$B \in C^1(\mathbb{R}^n)$ satisfying

$$B(x) \le 0 \quad \forall x \in \mathcal{X}_0, \tag{3.17}$$

$$B(x) > 0 \quad \forall x \in \mathrm{cl}(\partial\mathcal{X} \setminus \partial\mathcal{X}_r) \cup \mathcal{X}_u, \tag{3.18}$$

$$\frac{\partial B}{\partial x}(x) f(x) < 0 \quad \forall x \in \mathrm{cl}(\mathcal{X} \setminus \mathcal{X}_r). \tag{3.19}$$

Then the eventuality–safety property holds; i.e., for all initial states $x_0 \in \mathcal{X}_0$, the trajectory $x(t)$ starting at $x(0) = x_0$ will satisfy $x(T) \in \mathcal{X}_r$ for some $T \ge 0$ and $x(t) \notin \mathcal{X}_u$, $x(t) \in \mathcal{X}$ for all $t \in [0, T]$.

COROLLARY 3.11. *Consider the system $\dot{x} = f(x)$ with $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ and let $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{X}_0 \subseteq \mathcal{X}$, $\mathcal{X}_u \subseteq \mathcal{X}$, $\mathcal{X}_r \subseteq \mathcal{X}$ be bounded. If there exist an open set $\tilde{\mathcal{X}}_0$ containing $\mathcal{X}_0$ and a function $\rho \in C^1(\mathbb{R}^n)$ satisfying*

$$\rho(x) \ge 0 \quad \forall x \in \tilde{\mathcal{X}}_0, \tag{3.20}$$

$$\rho(x) < 0 \quad \forall x \in \mathrm{cl}(\partial\mathcal{X} \setminus \partial\mathcal{X}_r) \cup \mathcal{X}_u, \tag{3.21}$$

$$\nabla \cdot (\rho f)(x) > 0 \quad \forall x \in \mathrm{cl}(\mathcal{X} \setminus \mathcal{X}_r), \tag{3.22}$$

then the weak eventuality–safety property holds; i.e., for almost all initial states $x_0 \in \mathcal{X}_0$, the trajectory $x(t)$ starting at $x(0) = x_0$ will satisfy $x(T) \in \mathcal{X}_r$ for some $T \ge 0$ and $x(t) \notin \mathcal{X}_u$, $x(t) \in \mathcal{X}$ for all $t \in [0, T]$. In this case, the safety property holds also for trajectories that do not reach $\mathcal{X}_r$ in finite time.

4. Examples. We will now consider some examples to illustrate the application of the proposed methods. The MATLAB m-files for solving these examples can be found at http://www.cds.caltech.edu/∼prajna/files/PraR06.

4.1. Successive safety and reachability refinements. Consider the two-dimensional system

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = -x_1 + \frac{1}{3}x_1^3 - x_2,$$

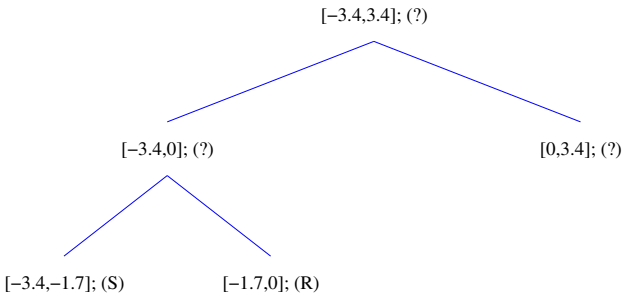
and let the set of states be $\mathcal{X} = [-3.5, 3.5] \times [-3.5, 3.5] \subset \mathbb{R}^2$. Furthermore, define

$$\mathcal{X}_0 = [-3.4, 3.4] \times [3.35, 3.45], \qquad \mathcal{X}_2 = [-3.5, 3.5] \times \{-3.5\},$$

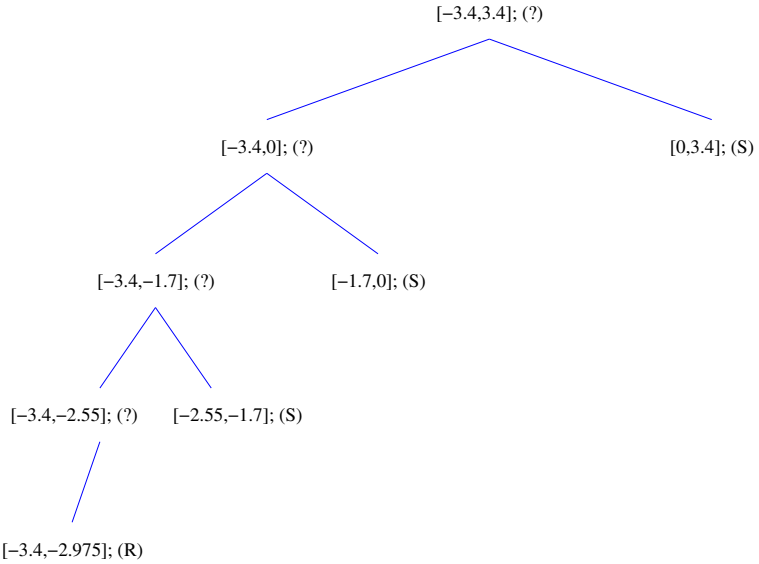
$$\mathcal{X}_1 = \{3.5\} \times [-3.5, 3.5], \qquad \mathcal{X}_3 = \{-3.5\} \times [-3.5, 3.5].$$

In this example, we will investigate the reachability of $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{X}_3$ from $\mathcal{X}_0$. Such facet-to-facet analysis is encountered when constructing a discrete abstraction of continuous or hybrid systems, or when analyzing a counterexample found during the verification of such an abstraction [1].

The convex programs in Theorems 3.2 and 3.3 will be used for our analysis. Since the vector field is polynomial and the sets are semialgebraic, we use polynomial parameterization for $B(x)$ and $\rho(x)$, and then utilize sum of squares programming to compute them. A degree bound equal to 8 is imposed on $B(x)$ and $\rho(x)$. Because of this, we might not be able to find a single $B(x)$ or $\rho(x)$ that proves safety or



(a) $\mathcal{X}_0 \to \mathcal{X}_1$



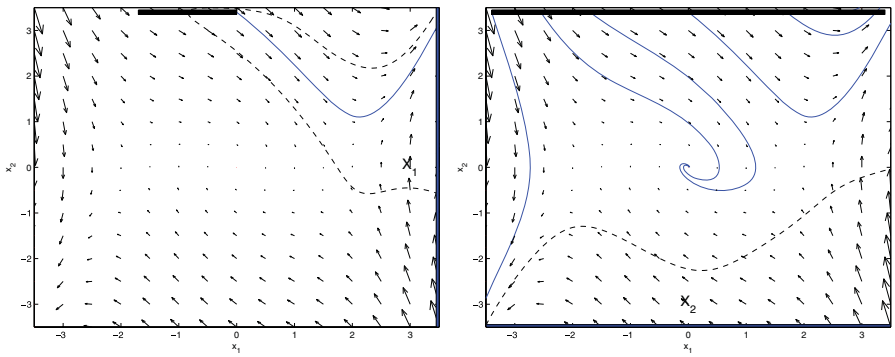
(b) $\mathcal{X}_0 \to \mathcal{X}_3$

FIG. 4.1. *Proving the reachability of $\mathcal{X}_1$ and $\mathcal{X}_3$ from $\mathcal{X}_0$ in the example of section* 4.1. *At each node we indicate the range of x_1 in $\mathcal{X}_0$ for which safety and reachability are tested. If neither is verified (denoted by ?), then the x_1-interval is divided into two and the tests are applied to the smaller sets. The annotation S (respectively, R) indicates that $B(x)$ (respectively, $\rho(x)$) is found. Breadth-first search starting from the leftmost branch is used. The verification of $\mathcal{X}_0 \nrightarrow \mathcal{X}_2$ terminates at the top node, since a barrier certificate $B(x)$ can be found directly.*

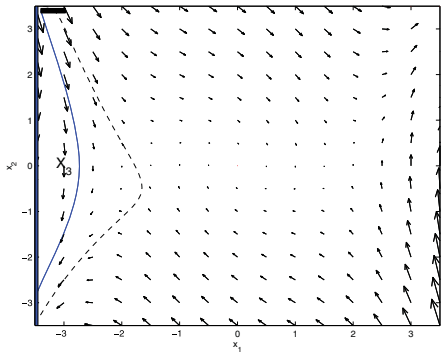
reachability for the whole $\mathcal{X}_0$. If neither $B(x)$ nor $\rho(x)$ can be found, we divide the interval of x_1 in $\mathcal{X}_0$ into two parts and apply the tests again to the smaller sets. A set is pruned if $B(x)$ is found, and this process is repeated until a $\rho(x)$ is found or the whole $\mathcal{X}_0$ is proven safe.

The result is as follows.

- We prove that the set $\mathcal{X}_1$ is reachable from $\mathcal{X}_0$. The verification progress is shown in Figure 4.1(a).



(a) $\mathcal{X}_0 \to \mathcal{X}_1$ (b) $\mathcal{X}_0 \nrightarrow \mathcal{X}_2$



(c) $\mathcal{X}_0 \to \mathcal{X}_3$

FIG. 4.2. *Possible transitions from $\mathcal{X}_0$ to $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{X}_3$ in the example of section* 4.1. *In* (a) *and* (c), *dashed curves are the zero level sets of $\rho(x)$'s that certify reachability. In* (b), *the dashed curve is the zero level set of $B(x)$ that certifies safety; trajectories starting from $\mathcal{X}_0$ cannot cross this level set to reach $\mathcal{X}_2$. Thick solid lines at the top of the figures are the initial sets for which the certificates are computed. Some trajectories of the system are depicted by solid curves.*

- It can be proven directly that $\mathcal{X}_2$ is not reachable from $\mathcal{X}_0$.
- It is proven that the set $\mathcal{X}_3$ is reachable from $\mathcal{X}_0$. See Figure 4.1(b).

For proofs of the corresponding reachability and safety, see Figure 4.2.

Obviously, the above bisection algorithm is just a simple, straightforward approach to refine and prune the initial set, and other algorithms that are more efficient can be proposed in the future.

4.2. Van der Pol oscillator. Consider the van der Pol oscillator with disturbance input:

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = x_2(1 - x_1^2) - x_1 + d,$$

where d is the disturbance input, taking its value in $\mathcal{D} = [-0.25, 0.25] \subset \mathbb{R}$. Let $\mathcal{X} = \{x \in \mathbb{R}^2 : 0.5 \leq \|x\|_2 \leq 5\}$. In addition, let

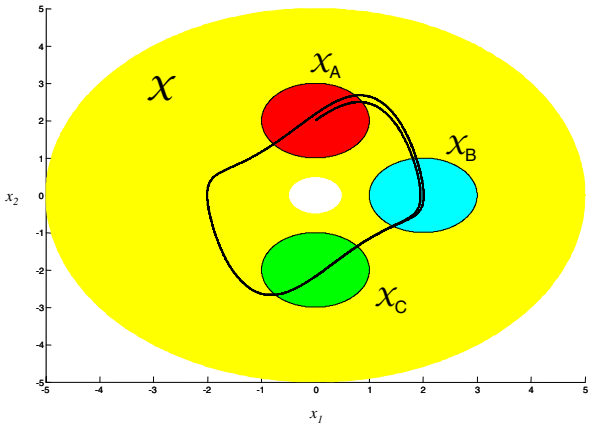


FIG. 4.3. *Verifying temporal properties of the van der Pol oscillator with disturbance. It is to be verified that under all possible disturbance input, if the system starts in $\mathcal{X}_A$, then both $\mathcal{X}_B$ and $\mathcal{X}_C$ are reached in finite time, but $\mathcal{X}_C$ will not be reached before the system reaches $\mathcal{X}_B$. The nominal trajectory of the system (i.e., for $d = 0$) starting at $x = (0, 2)$ is depicted by the solid curve.*

$$\mathcal{X}_A = \{x \in \mathbb{R}^2 : (x_1)^2 + (x_2 - 2)^2 \leq 1\},$$

$$\mathcal{X}_B = \{x \in \mathbb{R}^2 : (x_1 - 2)^2 + (x_2)^2 \leq 1\},$$

$$\mathcal{X}_C = \{x \in \mathbb{R}^2 : (x_1)^2 + (x_2 + 2)^2 \leq 1\}.$$

These sets are depicted in Figure 4.3, where a nominal trajectory of the system starting at $x = (0, 2)$ is also shown. Our objective in this example is to verify that under all possible piecewise continuous and bounded disturbances $d(t)$, if the system starts in $\mathcal{X}_A$, then both $\mathcal{X}_B$ and $\mathcal{X}_C$ are reached in finite time, but $\mathcal{X}_C$ will not be reached before the system reaches $\mathcal{X}_B$.

To verify this temporal specification, we will search for two barrier certificates $B_1(x)$ and $B_2(x)$ satisfying the following conditions:

$$\begin{cases} B_1(x) \leq 0 & \forall x \in \mathcal{X}_A, \\ B_1(x) > 0 & \forall x \in \partial\mathcal{X} \cup \mathcal{X}_C, \\ \frac{\partial B_1}{\partial x} f(x, d) \leq -\epsilon & \forall (x, d) \in (\mathcal{X} \setminus \mathcal{X}_B) \times \mathcal{D}, \end{cases}$$

$$\begin{cases} B_2(x) \leq 0 & \forall x \in \mathcal{X}_A, \\ B_2(x) > 0 & \forall x \in \partial\mathcal{X}, \\ \frac{\partial B_2}{\partial x} f(x, d) \leq -\epsilon & \forall x \in (\mathcal{X} \setminus \mathcal{X}_C) \times \mathcal{D} \end{cases}$$

for some positive ϵ. Using sum of squares programming, polynomials $B_1(x)$ and $B_2(x)$ of degree 10 can be found, and thus the temporal specification is verified.

5. A converse theorem. In this section, we will prove a converse theorem for safety verification using barrier certificates by exploiting the convexity of the problem formulation. The main result of the section can be stated as follows.

THEOREM 5.1. *Consider the system $\dot{x} = f(x)$ with $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$. Let $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{X}_0 \subseteq \mathcal{X}$, $\mathcal{X}_u \subseteq \mathcal{X}$ 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*

$$B(x) \leq 0 \quad \forall x \in \mathcal{X}_0, \tag{5.1}$$

$$B(x) > 0 \quad \forall x \in \mathcal{X}_u, \tag{5.2}$$

$$\frac{\partial B}{\partial x}(x)f(x) \leq 0 \quad \forall x \in \mathcal{X} \tag{5.3}$$

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 $\frac{\partial \tilde{B}}{\partial x}(x)f(x) < 0$ for all $x \in \mathcal{X}$. Later in the section 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 5.1 consists of two parts, given in Lemmas 5.2 and 5.4 below. In the first lemma, we use the Hahn–Banach theorem to show that the nonexistence of a $B(x)$ satisfying the conditions in Theorem 5.1 implies the existence of measures ψ_0, ψ_u, ρ satisfying some appropriate conditions. Then in Lemma 5.4 we show that the existence of such ψ_0, ψ_u, ρ actually implies that there exists an unsafe trajectory of the system.

LEMMA 5.2. *Let $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, 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 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 (5.1)–(5.3) only if there are 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. Let us 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 zero or ∞. There exists no $B \in C^1(\mathbb{R}^n)$ satisfying (5.1)–(5.3) 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:

$$\mathcal{K}_1 = \Bigg\{(z, h_0, h_u, h) \in \mathcal{K} : h_0 = B_0 - B,\ h_u = B - B_u,\ h = -\frac{\partial B}{\partial x} f \text{ on } \mathcal{X};$$
$$z = B_u - B_0;\ \text{ and } (B_0, B_u, B) \in \mathcal{B}\Bigg\},$$

$$\mathcal{K}_2 = \{(z, h_0, h_u, h) \in \mathcal{K} : z \geq 0,\ h_0 \geq 0 \text{ on } \mathcal{X}_0,\ h_u \geq 0 \text{ on } \mathcal{X}_u,\ h \geq 0 \text{ on } \mathcal{X}\}.$$

Then both $\mathcal{K}_1$ and $\mathcal{K}_2$ are convex sets, and $\mathcal{K}_2$ has a nonempty 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 $\mathcal{K}_1 \cap \operatorname{int}(\mathcal{K}_2) = \emptyset$. Now, by the Hahn–Banach theorem [17], 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

$$\sup_{k_1 \in \mathcal{K}_1} \langle k^*, k_1 \rangle \leq \inf_{k_2 \in \mathcal{K}_2} \langle k^*, k_2 \rangle, \tag{5.4}$$

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:

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

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 be equal to zero outside of $\mathcal{X}$. Then, for the left-hand side of (5.4), we have the following equality:

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

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 $(a, \psi_0, \psi_u, \rho)$, where ψ_0, ψ_u, ρ are measures of bounded variation on $\mathbb{R}^n$, such that $a \geq 0$; 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

$$\int_{\mathbb{R}^n} d\psi_0 = a,$$

$$\int_{\mathbb{R}^n} d\psi_u = a,$$

$$\nabla \cdot (\rho f) = \psi_0 - \psi_u.$$

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

$$\mathcal{L}_1 = \Bigg\{ (h_0, h_u, h) \in \mathcal{L} : h_0 = B_0 - B,\ h_u = B - B_u,\ h = -\frac{\partial B}{\partial x} f \text{ on } \mathcal{X};$$

$$\text{and } (B_0, B_u, B) \in \mathcal{B} \Bigg\},$$

$$\mathcal{L}_2 = \{(h_0, h_u, h) \in \mathcal{L} : h_0 \geq 0 \text{ on } \mathcal{X}_0,\ h_u \geq 0 \text{ on } \mathcal{X}_u,\ h \geq 0 \text{ on } \mathcal{X}\}.$$

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 (5.4). Then we have a nonzero $(\tilde{\psi}_0, \tilde{\psi}_u, \tilde{\rho}) \in (C(\mathcal{X})^*)^3$, such that

$$\sup_{\ell_1 \in \mathcal{L}_1} \langle (\tilde{\psi}_0, \tilde{\psi}_u, \tilde{\rho}), \ell_1 \rangle \leq \inf_{\ell_2 \in \mathcal{L}_2} \langle (\tilde{\psi}_0, \tilde{\psi}_u, \tilde{\rho}), \ell_2 \rangle.$$

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 5.2 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 the Liouville theorem which applies to measures.

LEMMA 5.3. *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

$$\lim_{k\to\infty}\int_X d|\rho_k-\rho|=0,$$

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

for every $X\subset D$. The lemma (cf. Lemma 3.1) was proven for the case of smooth ρ in [23], i.e.,

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

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

LEMMA 5.4. *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 $T\geq 0$ 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

$$Z_i=\bigcup_{x_0\in X_i}\{x\in\mathbb{R}^n : x=\phi_t(x_0)\text{ for some }t\geq 0\}\text{ for }i=1,2,\ldots.$$

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 5.3 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 $i=1,2,\ldots$, 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 purposes, 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

$$\phi_{\tilde{T}}(U)\subset\mathbb{R}^n\setminus(\mathcal{X}),$$

$$\phi_t(U)\cap\mathcal{X}_u=\emptyset\quad\forall t\in[0,\tilde{T}].$$

Using Lemma 5.3 again we obtain

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

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 nonpositive, 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 5.1.

($\Rightarrow$): This has been proven in Theorem 3.2.

($\Leftarrow$): This follows from Lemmas 5.2 and 5.4. □

5.1. Some remarks. The result stated in Theorem 5.1 uses the assumption that the following Slater-like condition [6] is fulfilled: 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 so does the necessity of barrier certificates) 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 5.5. *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 $\mathrm{cl}(\tilde{\mathcal{X}})$ 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 5.1 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 (3.4). 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 (5.1)–(5.3) 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 5.6. *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 $V(x)$.*[4] *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*

$$B(x) \leq 0 \quad \forall x \in \mathcal{X}_0, \tag{5.5}$$

$$B(x) > 0 \quad \forall x \in \mathcal{X}_u, \tag{5.6}$$

$$\frac{\partial B}{\partial x}(x) f(x) \leq 0 \quad \forall x \in \{x \in \mathbb{R}^n : \epsilon_1 \leq V(x) \leq \epsilon_2\} \tag{5.7}$$

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

$$x(0) \in \mathcal{X}_0, \tag{5.8}$$

$$x(T) \in \mathcal{X}_u \text{ for some } T \geq 0. \tag{5.9}$$

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 5.5, and even the Lyapunov function $V(x)$ can be used as $\tilde{B}(x)$. By Theorem 5.1, there exists a function $B \in C^1(\mathbb{R}^n)$ satisfying (5.5)–(5.7) if and only 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 $t \in [0, T]$.

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 $x(t) \in \mathcal{X}$ for all $t \in [0, T]$ is the same as the set of trajectories satisfying (5.8)–(5.9), and therefore the statement of the proposition follows. ∎

6. Conclusions. In the previous sections, we have used insights from the linear programming duality appearing in the shortest path problem and the concept of density function to formulate a convex program for reachability, which together with a convex program for safety verification using barrier certificates proposed in an earlier work form a pair of weak alternatives for safety and reachability verification. We have additionally shown that other temporal properties such as eventuality and avoidance can also be verified via convex programming and have presented convex programs to do so. This opens the possibility of performing the verification using convex optimization. In particular, sum of squares programming can be used for this purpose when the vector field of the system is polynomial and the sets are semialgebraic.

We have further commented on the use of this methodology for worst-case verification or controller synthesis. It was pointed out that the convex programs can be combined to verify properties such as reachability–safety and eventuality–safety. Some

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

examples have been presented for illustration. At the end of the paper, a converse theorem in safety verification using barrier certificates was proven.

Even though the present tests are aimed for continuous systems, they are useful for constructing discrete abstractions of hybrid systems. In addition, we expect that all of them can also be extended to handle hybrid systems directly, using an approach similar to the one presented in [21].

REFERENCES

[1] R. Alur, T. Dang, and F. Ivancic, *Progress on reachability analysis of hybrid systems using predicate abstraction*, in Hybrid Systems: Computation and Control, Lecture Notes in Comput. Sci. 2623, Springer-Verlag, Heidelberg, 2003, pp. 4–19.

[2] H. Anai and V. Weispfenning, *Reach set computations using real quantifier elimination*, in Hybrid Systems: Computation and Control, Lecture Notes in Comput. Sci. 2034, Springer-Verlag, Berlin, 2001, pp. 63–76.

[3] E. Asarin, T. Dang, and O. Maler, *The d/dt tool for verification of hybrid systems*, in Computer Aided Verification, Lecture Notes in Comput. Sci. 2404, Springer-Verlag, Berlin, 2002, pp. 365–370.

[4] J.-P. Aubin, *Viability Theory*, Birkhäuser, Boston, MA, 1991.

[5] A. Bemporad, F. D. Torrisi, and M. Morari, *Optimization-based verification and stability characterization of piecewise affine and hybrid systems*, in Hybrid Systems: Computation and Control, Lecture Notes in Comput. Sci. 1790, Springer-Verlag, Berlin, 2000, pp. 45–58.

[6] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[7] A. Chutinan and B. H. Krogh, *Computational techniques for hybrid system verification*, IEEE Trans. Automat. Control, 48 (2003), pp. 64–75.

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

[9] S. Glavaski, A. Papachristodoulou, and K. Ariyur, *Safety verification of controlled advanced life support system using barrier certificates*, in Hybrid Systems: Computation and Control, Lecture Notes in Comput. Sci. 3414, Springer-Verlag, Heidelberg, 2005, pp. 306–321.

[10] G. J. Holzmann, *Design and Validation of Computer Protocols*, Prentice–Hall, Englewood Cliffs, NJ, 1991.

[11] R. Horowitz and P. Varaiya, *Control design of an automated highway system*, Proc. IEEE, 88 (2000), pp. 913–925.

[12] M. Huth and M. Ryan, *Logic in Computer Science: Modelling and Reasoning about Systems*, Cambridge University Press, Cambridge, UK, 2000.

[13] M. Jirstrand, *Invariant sets for a class of hybrid systems*, in Proceedings of the 37th IEEE Conference on Decision and Control, 1998, pp. 3699–3704.

[14] H. K. Khalil, *Nonlinear Systems*, 2nd ed., Prentice–Hall, Upper Saddle River, NJ, 1996.

[15] A. Kurzhanski and P. Varaiya, *Ellipsoidal techniques for reachability analysis*, in Hybrid Systems: Computation and Control, Lecture Notes in Comput. Sci. 1790, Springer-Verlag, Heidelberg, 2000, pp. 203–213.

[16] G. Lafferriere, G. J. Pappas, and S. Yovine, *Symbolic reachability computations for families of linear vector fields*, J. Symbolic Comput., 32 (2001), pp. 231–253.

[17] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, New York, 1969.

[18] Z. Manna and A. Pnueli, *Temporal Verification of Reactive Systems: Safety*, Springer-Verlag, New York, 1995.

[19] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Dover, Mineola, NY, 1998.

[20] P. A. Parrilo, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 2000.

[21] S. Prajna, A. Jadbabaie, and G. J. Pappas, *A framework for worst-case and stochastic safety verification using barrier certificates*, IEEE Trans. Automat. Control, to appear (2007).

[22] S. Prajna, A. Papachristodoulou, and P. A. Parrilo, *Introducing SOSTOOLS: A general purpose sum of squares programming solver*, in Proceedings of the 41st IEEE Conference on Decision and Control, 2002, pp. 741–746; available at http://www.cds.caltech.edu/sostools

and http://www.mit.edu/∼parrilo/sostools.

[23] A. Rantzer, *A dual to Lyapunov's stability theorem*, Systems Control Lett., 42 (2001), pp. 161–168.

[24] A. Rantzer and S. Hedlund, *Duality between cost and density in optimal control*, in Proceedings of the 42nd IEEE Conference on Decision and Control, 2003, pp. 1218–1221.

[25] A. Rantzer and S. Prajna, *On analysis and synthesis of safe control laws*, in Proceedings of the 42nd Allerton Conference on Communication, Control, and Computing, 2004.

[26] S. Sankaranarayanan, H. Sipma, and Z. Manna, *Constructing invariants for hybrid systems*, in Hybrid Systems: Computation and Control, Lecture Notes in Comput. Sci. 2993, Springer-Verlag, Berlin, 2004, pp. 539–554.

[27] A. Tiwari, *Approximate reachability for linear systems*, in Hybrid Systems: Computation and Control, Lecture Notes in Comput. Sci. 2623, Springer-Verlag, Berlin, 2003, pp. 514–525.

[28] A. Tiwari and G. Khanna, *Nonlinear systems: Approximating reach sets*, in Hybrid Systems: Computation and Control, Lecture Notes in Comput. Sci. 2993, Springer-Verlag, Berlin, 2004, pp. 600–614.

[29] C. Tomlin, I. Mitchell, and R. Ghosh, *Safety verification of conflict resolution maneuvers*, IEEE Trans. Intelligent Transportation Systems, 2 (2001), pp. 110–120.

[30] C. J. Tomlin, I. Mitchell, A. M. Bayen, and M. Oishi, *Computational techniques for the verification of hybrid systems*, Proc. IEEE, 91 (2003), pp. 986–1001.

[31] H. Yazarel and G. Pappas, *Geometric programming relaxations for linear system reachability*, in Proceedings of the American Control Conference, 2004, pp. 553–559.