On Analysis and Synthesis of Safe Control Laws

Rantzer, Anders; Prajna, Stephen

Published in:
Proceedings of the Allerton Conference on Communication, Control, and Computing

2004

On Analysis and Synthesis of Safe Control Laws

Anders Rantzer*
Department of Automatic Control
LTH, Lund University
Box 118, SE-224 74 Lund, SWEDEN
rantzer@control.lth.se

Stephen Prajna
Control and Dynamical Systems
California Institute of Technology
Pasadena, CA 91125, USA
prajna@cds.caltech.edu

Abstract

Controller synthesis for nonlinear systems is considered with the following objective: no trajectory starting from a given set of initial states is allowed to enter into a given set of forbidden (unsafe) states. A methodology for safety verification using barrier certificates has recently been proposed. Here it is shown how a safe control law together with a corresponding certificate can be computed by means of convex optimization. A basic tool is the theory for density functions in analysis of nonlinear systems. Computational examples are considered.

1 Introduction

Safety verification or reachability analysis addresses the question whether an unsafe region in the state space is reachable by some system trajectories starting from a set of initial states. The need for safety verification is crucial in many engineering disciplines, especially in the presence of nonlinear dynamics.

Various methods have been proposed for safety verification. For verification of discrete (finite state) systems, model checking techniques [3] have been quite successful and have garnered a popularity that prompts the development of analogous approaches for verification of continuous systems, mostly requiring computational propagation of initial states (see e.g. [1, 6]). Unfortunately, while these techniques 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, a method for safety verification was recently proposed based on barrier certificates [7, 8], closely related to the notion of Lyapunov function for stability analysis. Computation of barrier certificates was done using polynomial parameterizations and convex optimization.

In this paper, we point out that the same problem can be addressed using density functions, as introduced in [12]. This makes it possible to also include synthesis in the convex optimization problem, to find a controller that satisfies the safety specification and at the same time a certificate that verifies the safety.

The paper is structured as follows. The technique with barrier certificates is reviewed in Section 2. After that, the corresponding technique based on density functions is introduced in Section 3 and the synthesis procedure is described in Section 4.

*This paper was written while the first author spent a sabbatical year 2004/05 at Caltech supported by the Swedish Foundation for Strategic Research.

2 Safety Verification Using Barrier Certificates

Our conditions for safety can be stated as follows. Given a system $\dot{x} = f(x)$ with the state x taking values in $\mathcal{X}$, 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 $B : \mathcal{X} \to \mathbb{R}$ such that

$$
\begin{aligned}
B(x) &\leq 0 && \forall x \in \mathcal{X}_0, && (1)\\
B(x) &> 0 && \forall x \in \mathcal{X}_u, && (2)\\
\frac{\partial B}{\partial x} f(x) &\leq 0 && \forall x \in \mathcal{X}. && (3)
\end{aligned}
$$

Then the system is safe, namely, 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 [5], and is also closely related to the use of viability theory [2] and invariant sets [4] 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 [10]. The method can also be extended to handle hybrid, uncertain, and stochastic systems [8, 9].

Example 1 Consider the two-dimensional system (taken from [5, page 180])

$$
\begin{aligned}
\dot{x}_1 &= x_2,\\
\dot{x}_2 &= -x_1 + \frac{1}{3}x_1^3 - x_2,
\end{aligned}
$$

with $\mathcal{X} = \mathbb{R}^2$. We want to verify that no trajectory of the system starting at

$$\mathcal{X}_0 = \{x \in \mathbb{R}^2 : (x_1 - 1.5)^2 + x_2^2 < 0.25\}$$

will ever reach the unsafe set

$$\mathcal{X}_u = \{x \in \mathbb{R}^2 : (x_1 + 1)^2 + (x_2 + 1)^2 < 0.16\}.$$

Using sum of squares programming, we are able to find a polynomial barrier certificate $B(x)$ that satisfies (1)–(3), e.g.,

$$B(x) = -13 + 7x_1^2 + 16x_2^2 - 6x_1^2x_2^2 - \frac{7}{6}x_1^4 - 3x_1x_2^3 + 12x_1x_2 - \frac{12}{3}x_1^3x_2.$$

For example, non-positivity of the Lie derivative $\frac{\partial B}{\partial x} f(x)$ can be shown by considering the quadratic form

$$-\frac{\partial B}{\partial x} f(x) = \begin{bmatrix} x_2 \\ x_2^2 \\ x_1 \\ x_1x_2 \\ x_1^2x_2 \\ x_1^3 \end{bmatrix}^T \begin{bmatrix} 20 & 0 & 15 & 0 & -15/2 & -5 \\ 0 & 3 & 0 & 3/2 & 0 & 0 \\ 15 & 0 & 12 & 0 & -6 & -4 \\ 0 & 3/2 & 0 & 6 & 0 & 0 \\ -15/2 & 0 & -6 & 0 & 3 & 2 \\ -5 & 0 & -4 & 0 & 2 & 4/3 \end{bmatrix} \begin{bmatrix} x_2 \\ x_2^2 \\ x_1 \\ x_1x_2 \\ x_1^2x_2 \\ x_1^3 \end{bmatrix}.$$

In this quadratic form, the coefficient matrix is positive semi-definite, which implies the existence of a sum of squares decomposition for $-\frac{\partial B}{\partial x} f(x)$ (and hence its nonnegativity). That (1)–(2) are satisfied is depicted in Figure 1, and in fact can also be shown by sum of squares arguments. Hence the safety of the system is verified.

□

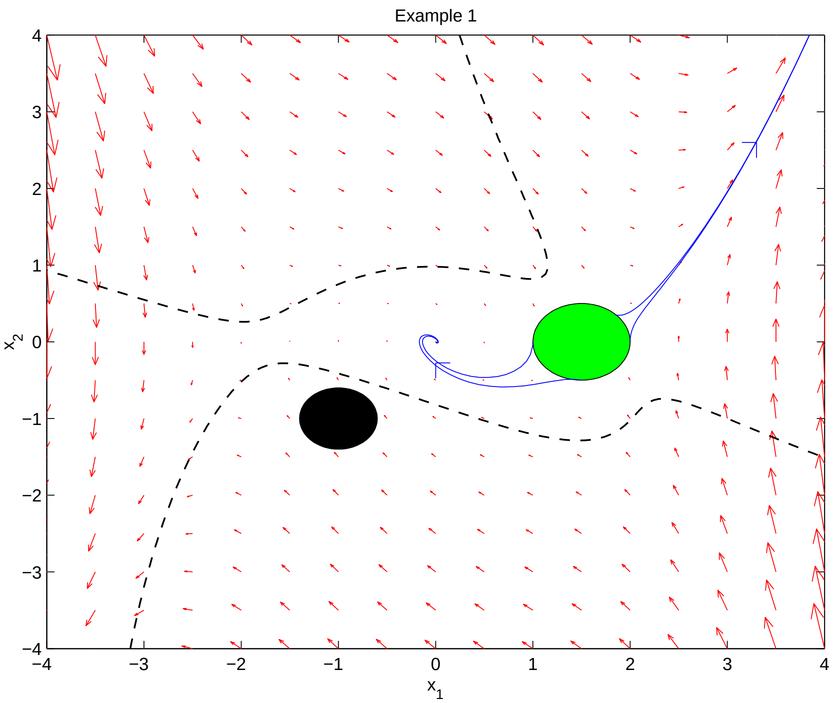

Figure 1: Phase portrait of the system in Example 1. Solid patches are (from the left) $\mathcal{X}_u$ and $\mathcal{X}_0$, respectively. Dashed curves are the zero level set of $B(x)$, whereas solid curves are some trajectories of the system. The function $B(x)$ is greater than zero for all $x \in \mathcal{X}_u$ and less than zero for all $x \in \mathcal{X}_0$.

3 Safety Verification Using Density Functions

An alternative criterion for safety of the system $\dot{x}(t) = f(x(t))$ is stated using a density function ρ in the following way.

Theorem 1 *Consider $f \in \mathbf{C}^1(\mathbf{R}^n, \mathbf{R}^n)$ and let $\mathcal{X}_0$ and $\mathcal{X}_u$ be open subsets of the open set $\mathcal{X} \subset \mathbf{R}^n$. Suppose there exists a function $\rho \in \mathbf{C}^1(\mathbf{R}^n, \mathbf{R})$ such that*

$$[\nabla \cdot (\rho f)](x) \geq 0 \qquad \forall x \in \mathcal{X}, \tag{4}$$

$$\rho(x) > 0 \qquad \forall x \in \mathcal{X}_0, \tag{5}$$

$$\rho(x) \leq 0 \qquad \forall x \in \mathcal{X}_u. \tag{6}$$

Then there exists no solution to the equation $\dot{x}(t) = f(x(t))$ with $x(0) \in \mathcal{X}_0$, $x(T) \in \mathcal{X}_u$ for some $T > 0$ and $x(t) \in \mathcal{X}$ for all $t \in [0, T]$.

Remark 1 It should be noted that in points where $\rho(x) = 0$ the divergence inequality reduces to a gradient constraint:

$$0 \leq [\nabla \cdot (\rho f)](x) = \rho(x) \nabla \cdot f(x) + \nabla \rho \cdot f(x) = \nabla \rho \cdot f(x).$$

Hence, on the zero level set, the density function ρ plays a very similar role to the barrier functions of the previous section. □

Remark 2 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 must be satisfied, since the inequality is otherwise impossible to satisfy without a singularity in ρ. 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 ρ. □

Theorem 1 will be proved using the following version of Liouville's theorem (from [12]).

Lemma 2 *Let $f \in \mathbf{C}^1(D, \mathbf{R}^n)$ where $D \subset \mathbf{R}^n$ is open and let $\rho \in \mathbf{C}^1(D, \mathbf{R})$ be integrable. For $x_0 \in \mathbf{R}^n$, let $\phi_t(x_0)$ be the solution $x(t)$ of $\dot{x} = f(x)$, $x(0) = x_0$. For a measurable set Z, assume that $\phi_\tau(Z) = \{\phi_\tau(x) \mid x \in Z\}$ is a subset of D for all τ between 0 and t. Then*

$$\int_{\phi_t(Z)} \rho(x)dx - \int_Z \rho(z)dz = \int_0^t \int_{\phi_\tau(Z)} \left[\nabla \cdot (f\rho)\right](x)dxd\tau. \tag{7}$$

Proof of Theorem 1 Assume that the theorem is false. Then there exists an $x_0 \in \mathcal{X}_0$ such that $\phi_T(x_0) \in \mathcal{X}_u$ for some $T > 0$ and $\phi_t(x_0) \in \mathcal{X}$ for $t \in [0, T]$. Let $Z \subset \mathcal{X}_0$ be a ball surrounding x_0 such that also $\phi_T(Z) \subset \mathcal{X}_u$ and $\phi_t(Z) \subset \mathcal{X}$ for $t \in [0, T]$. Let D be a bounded open set containing $\phi_t(x)$ for $x \in Z$, $t \in [0, T]$, and apply Lemma 2 to obtain a contradiction. According to the assumptions of Theorem 1, the left hand side of (7) is negative and the right hand side is non-negative. Hence, there is a contradiction and the proof is complete. □

Remark 3 If the system satisfies the additional property that $\nabla \cdot f(x) \leq 0$, then conditions (5)–(6) can be replaced by a less stringent condition

$$\rho(x_0) - \rho(x_u) > 0 \qquad\qquad \forall (x_0, x_u) \in \mathcal{X}_0 \times \mathcal{X}_u.$$

In this case, the volume of $\phi_t(Z)$ is non-increasing along time, and thus when we apply Lemma 2 the left hand side of (7) is still negative. □

Example 2 Consider again the verification problem of Example 1. Note that the system has a stable equilibrium at the origin and saddle points at $(\pm\sqrt{3}, 0)$. In this example, we will exclude a neighborhood of $(0, 0)$ and $(\sqrt{3}, 0)$ from the region where the divergence inequality needs to be satisfied. More specifically, we ask that

$$\begin{aligned} \rho(x) &\geq 0.1 & \qquad & \forall x \in \mathcal{X}_0, \\ \rho(x) &\leq -0.1 & & \forall x \in \mathcal{X}_u, \\ \nabla \cdot (\rho f)(x) &\geq 0 & & \forall x \in \mathcal{X} \setminus \mathcal{X}_{\text{excl}}, \end{aligned}$$

where $\mathcal{X}_{\text{excl}} = \{x : x_1^2 + x_2^2 \leq 0.25\} \cup \{x : (x_1 - 1.5)^2 + x_2^2 \leq 0.16\}$. To ensure that the safety statement is valid in terms of the original set $\mathcal{X}$, we also ask that $\rho(x) \geq 0.1$ at the boundary of $\mathcal{X}_{\text{excl}}$ (cf. Remark 2). A polynomial $\rho(x)$ satisfying these conditions, computed using sum of squares optimization, is

$$\begin{aligned} \rho(x) = \ & 7.6152 - 12.1597x_1 - 3.85628x_2 - 18.5818x_1^2 - 3.92213x_1x_2 - 17.2032x_2^2 \\ & + 51.0454x_1^3 + 15.8062x_1^2x_2 + 45.3684x_1x_2^2 - 32.2778x_1^4 - 9.27854x_1^3x_2 \\ & - 28.5075x_1^2x_2^2 + 1.03161x_1^2x_2^3 - 3.64267x_2^5 + 5.75452x_1^6 - .498478x_1^4x_2^2 \\ & - 1.60486x_1^2x_2^4 - 6.96206x_2^6 - 1.23579x_1^7 + 2.33577x_1^5x_2^2 - 1.86292x_1^4x_2^3 \\ & + .503928x_1^7x_2 + 2.19945x_1^3x_2^5, \end{aligned}$$

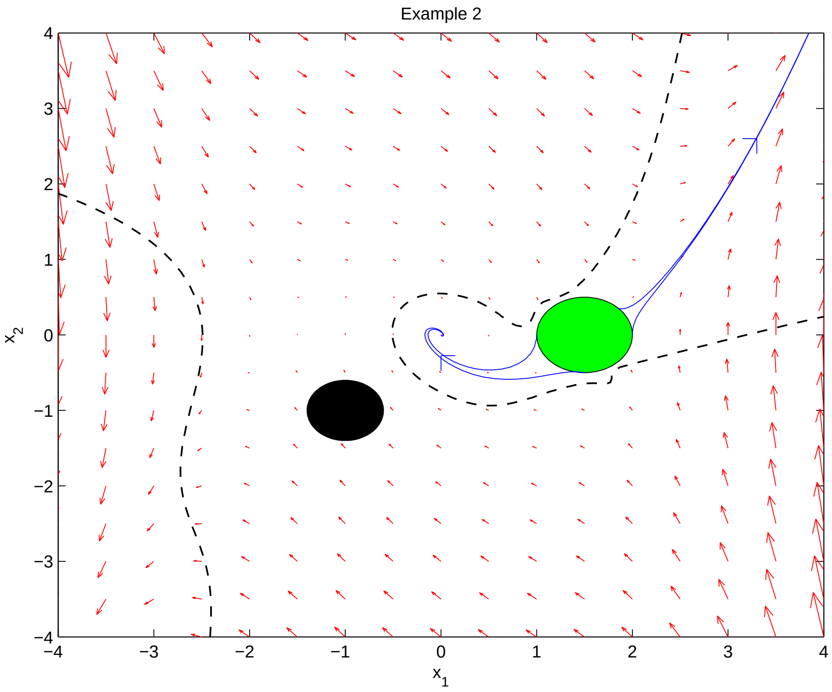

Figure 2: The level set $\rho(x) = 0$ is shown by the dashed curve, for a density function satisfying the conditions of Theorem 1 in Example 2. The existence of such a function proves that the system is safe.

and thus the system is safe. The phase portrait of the system and the zero level set of $\rho(x)$ are shown in Figure 2.

In this example we have $\nabla \cdot f \leq 0$. This makes it possible to relax the conditions on ρ (cf. Remark 3) and only ask that

$$\begin{aligned} \rho(x_0) - \rho(x_u) &\geq 0.1 & \qquad & \forall (x_0, x_u) \in \mathcal{X}_0 \times \mathcal{X}_u, \\ \nabla \cdot (\rho f)(x) &\geq 0 & & \forall x \in \mathcal{X}. \end{aligned}$$

The resulting $\rho(x)$ is simpler:

$$\rho(x) = -1.617 + 1.001x_1 + .2512x_2 - .4257x_1^2 - .7985x_1x_2 - .5973x_2^2 + .4203x_1^3x_2.$$

A level set of $\rho(x)$ which separates $\mathcal{X}_u$ and $\mathcal{X}_0$ is shown in Figure 3.

□

4 Safe Control Synthesis Using Density Functions

It has been widely recognized that density functions are useful to formulate nonlinear control synthesis in terms of convex optimization [12, 11]. However, a straightforward attempt to use the conditions of Theorem 1 yields some unexpected difficulties. The inequalities

$$\begin{aligned} \nabla \cdot [\rho(f + ug)](x) &\geq 0 & \qquad & \text{for } x \in \mathbf{R}^n, \\ \rho(x) &> 0 & & \text{for } x \in \mathcal{X}_0, \\ \rho(x) &\leq 0 & & \text{for } x \in \mathcal{X}_u \end{aligned}$$

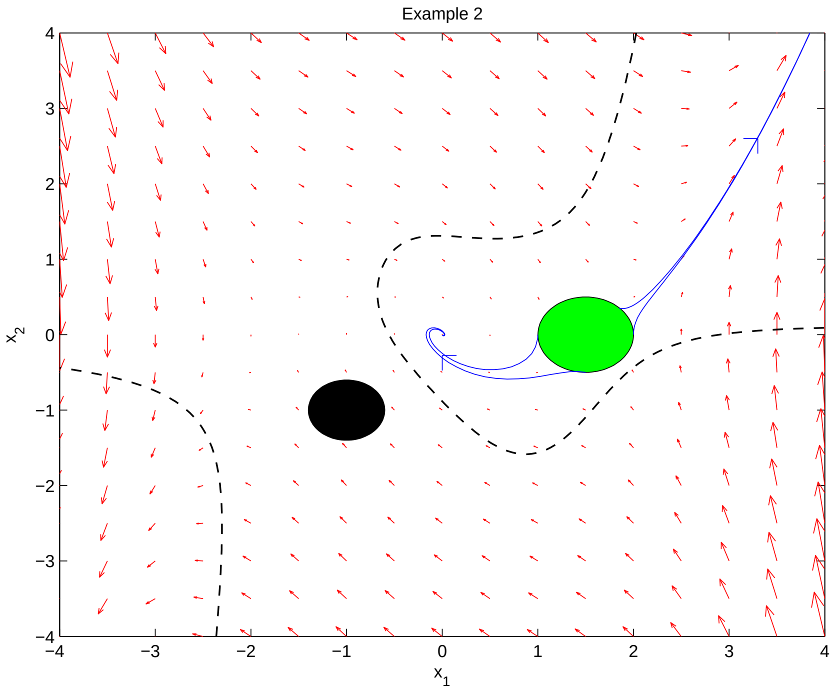

Figure 3: A level set separating $\mathcal{X}_0$ and $\mathcal{X}_u$ for the density function corresponding to the simplified criterion in Example 2 is depicted by the dashed curves.

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 $u(x) = \psi(x)/\rho(x)$. The difficulty arises because $u(x)$ becomes singular at points where $\rho(x) = 0$, unless $\psi(x)$ also vanishes there.

To overcome the difficulty, we will synthesize controllers under the magnitude constraint $|u(x)| \leq c$, where $c > 0$, using the conditions

$$\begin{aligned}
\nabla \cdot [\rho f + \psi g](x) &\geq 0 && \text{for } x \in \mathcal{X}, \\
|\psi(x)| &\leq c\rho(x) && \text{for } x \in \mathcal{X}, \\
\rho(x) &> 0 && \text{for } x \in \mathcal{X}_0, \\
\rho(x) &= 0 && \text{for } x \in \mathcal{X}_u.
\end{aligned}$$

Given a solution (ρ, ψ) to the inequalities, the control law $u(x) = \psi(x)/\rho(x)$ makes the region $\rho(x) > 0$ invariant.

The new conditions maintain convexity in the pair (ρ, ψ) and do not generate singular controllers. However, they do not allow for polynomial parameterizations of ρ and ψ since any polynomial vanishing identically in the open region $\mathcal{X}_u$ must be zero everywhere. Instead, we have in the following example implemented the search by gridding the state space.

Example 3 We consider the problem of navigating a boat on a river. The kinematic model of the system is given by

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = f(x, y) + g(x, y)u,$$

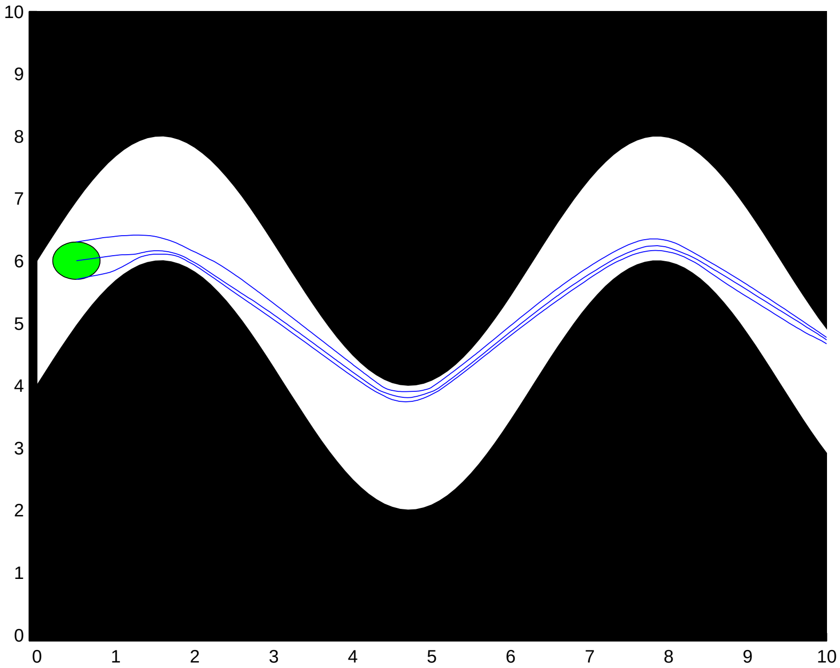

Figure 4: Dark region is the unsafe set $\mathcal{X}_u$ in Example 3, whereas the small circle near the left edge is the initial set $\mathcal{X}_0$. Solid curves are some trajectories of the closed loop system.

where the drift current profile $f(x, y)$ is given by

$$\begin{bmatrix} 1 + 0.125\cos(0.5x) - 0.125\sin(0.5y) \\ 0 \end{bmatrix},$$

and $g(x, y) = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$. We assume that magnitude bound $|u| \leq 1$ is imposed on the control input.

The set of states we consider is

$$\mathcal{X} = \{(x, y) \in \mathbb{R}^2 : -0.1 \leq x \leq 10, 0 \leq y \leq 10\},$$

while the initial and the unsafe sets are

$$\begin{aligned} \mathcal{X}_0 &= \{(x, y) \in \mathcal{X} : (x - 0.5)^2 + (y - 6)^2 \leq 0.09\}, \\ \mathcal{X}_u &= \{(x, y) \in \mathcal{X} : (2\sin x + 6 - y)(y - 2\sin x - 4) \leq 0 \text{ or } x \leq 0\}. \end{aligned}$$

See Figure 4. We use gridding with step size equal to 0.1 to solve the synthesis problem, where centered difference approximation is used to approximate the derivatives in the divergence inequality. The resulting linear program is solved for ρ and ψ, and then the controller is given by $u(x) = \psi(x)/\rho(x)$, with linear interpolation between the grid points. Some trajectories of the closed loop system are depicted in Figure 4, and the corresponding $\rho(x)$ is shown in Figure 5.

□

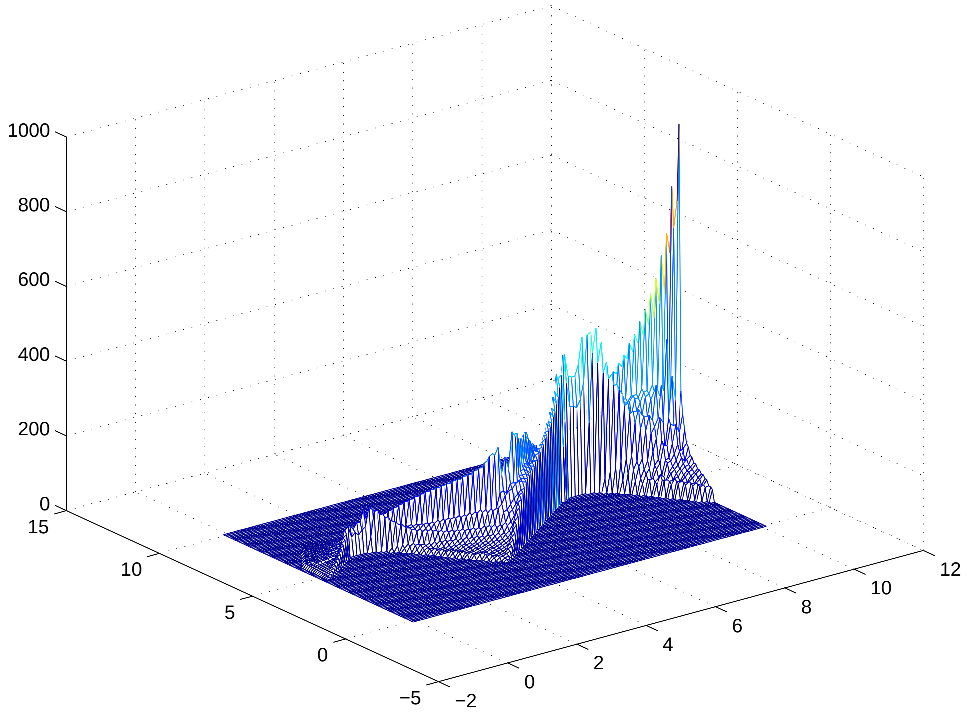

Figure 5: The density function obtained by gridding and convex optimization in Example 3.

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, LNCS 2623*, pages 4–19. Springer-Verlag, Heidelberg, 2003.

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

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

[4] M. Jirstrand. Invariant sets for a class of hybrid systems. In *Proceedings IEEE Conference on Decision and Control*, 1998.

[5] H. K. Khalil. *Nonlinear Systems*. Prentice-Hall, Inc., Upper Saddle River, NJ, second edition, 1996.

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

[7] S. Prajna. Barrier certificates for nonlinear model validation. In *Proceedings IEEE Conference on Decision and Control*, 2003.

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

[9] S. Prajna, A. Jadbabaie, and G. J. Pappas. Stochastic safety verification using barrier certificates. In *Proceedings IEEE Conference on Decision and Control*, 2004.

[10] S. Prajna, A. Papachristodoulou, and P. A. Parrilo. Introducing SOSTOOLS: A general purpose sum of squares programming solver. In *Proceedings IEEE Conference on Decision and Control*, 2002. Available at http://www.cds.caltech.edu/sostools and http://www.aut.ee.ethz.ch/~parrilo/sostools.

[11] S. Prajna, P. A. Parrilo, and A. Rantzer. Nonlinear control synthesis by convex optimization. *IEEE Transactions on Automatic Control*, 49(2):310–314, 2004.

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