The Reachable Set and Safety Guarantee

23/10/2017

FLINDERS UNIVERSITY

Huda Bu Obaid

Supervisor: Murk Bottema

Submitted to the School of Computer Science, Engineering, and Mathematics in the Faculty of Science and Engineering in partial fulfilment of the requirements for the degree of master at Flinders University – Adelaide Australia.

Acknowledgment

I would first like to thank my thesis advisor the reachable set and safety guarantee, Murk Bottema of the School of Computer Science, Engineering, and Mathematics / Faculty of Science and engineering at Flinders University. The door to Prof. Bottema office was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work, but steered me in the right the direction whenever he thought I needed it.

Declaration of Originality.

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Sign: Huda Ali Bu Obaid, Date: 02/02/2018

Contents

Abstract.

This research focuses on computing the exact reachable set which can be formulated in terms of a Hamilton-Jacobi partial differential equation (PDE). Hamilton-Jacobi (HJ) reachability is a procedure which can provide an accurate analysis of the safety state of the dynamical systems. This procedure is very effective in low-dimensional dynamical systems like aircrafts and quadrotors. It guarantees the safety where there is a potential dangerous scenario. One can get the safety guarantee by computing the reachable set, particularly the backward reachable set (BRS). The system might violate the safety properties in spite of the effort from the system to stay safe. The Hamilton-Jacobi reachability cannot be used for high dimensional systems because the computation of the backward reachable sets is complex and the number of state dimensions scales exponentially. In spite of the fact that; there are many techniques which can be used for approximation and these techniques have the ability to give conservative estimates for the backward reachable set. However, they do not provide an accurate solution and they usually demand limiting assumptions about the system dynamics. Therefore, this project will use a general method to analyze dynamical systems. When the results of subsystems are connected, the high-dimensionality of backward reachable sets can be easily and quickly computed which were previously considered to be intractable or take a long time to do. Also, it gives an exact computation in lower-dimensional subspaces. The method which will be applied in this project, will project the full dimensional of the reachable sets into lower dimensional subsystems. This method is from a decomposition of the true BRS. The theoretical results will be explained through the example of aircraft collision avoidance, 3D Dubins car and a linear system example. The Hamiltonian formulation of mechanics is a stronger formulation than others such as Lagrangian formulation and Newtonian formulation. In the Hamiltonian formulation, we can use coordinates which are much wider class. It expresses the relationship between the conservation and symmetries laws.

1   Introduction.

Defining models for complex systems is difficult (I. M. Mitchell, Bayen, & Tomlin, 2005). Recently, the use of autonomous system has increased rapidly among self-driving cars, aircraft and unmanned aerial vehicles (UAV) (Hoffmann, Huang, Waslander, & Tomlin, 2007) (Kong, Pfeiffer, Schildbach, & Borrelli, 2015). These applications generate challenging questions. How can a car choose a safe path while driving? Which spacecrafts will lead to a collision if there is a target set such as an obstacle? These questions are common in the fields of optimal control and differential games which all center in the concept of reachability (Allen, Clark, Starek, & Pavone, 2014). The reachability is a procedure which can provide an accurate analysis of the safety state of the dynamical systems. So, complex models must be verified and validated and these tasks often draw the main attention in many settings and have been studied extensively in the past (C. J. Tomlin, Lygeros, & Sastry, 2000). Simulation is a very simple way to compute the validation, but the main problem with this method is that it checks only one trajectory of the system at a specific time. Therefore, for systems which contain many inputs signals and various of state values, it will be difficult to check the safety for every possible trajectory of the system by simulation only. Researchers have developed a way to catch the behavior of all possible trajectories at one time by computing reachable sets (I. M. Mitchell et al., 2005). This means that safety verification is very fundamental to define the set of states. If the system of states has pushed into unsafe region or configuration, the safety of the system can be verified to make sure that it stays outside the set. For instance, there are automatic control laws which have been proposed to ensure separation between aircrafts. The unsafe configuration or region in this case is that the distance between any couple of vehicles is less than the required minimum. Thus, we need to make sure and guarantee that the loss of separation never occurs (C. Tomlin, Mitchell, & Ghosh, 2001), (I. M. Mitchell & Tomlin, 2003). As a result, reachable set analysis is a suitable tool for this model. There are two types of reachable sets which depend on two conditions: the initial condition and the final condition. The initial condition specifies for the forward reachable set. Also, it seeks to define the set of all states which can be reached along the trajectories which start in the set. Whereas, the final condition or target set of states specifies the backward reachable set. It defines the set of states from which trajectories start can reach that given target set.
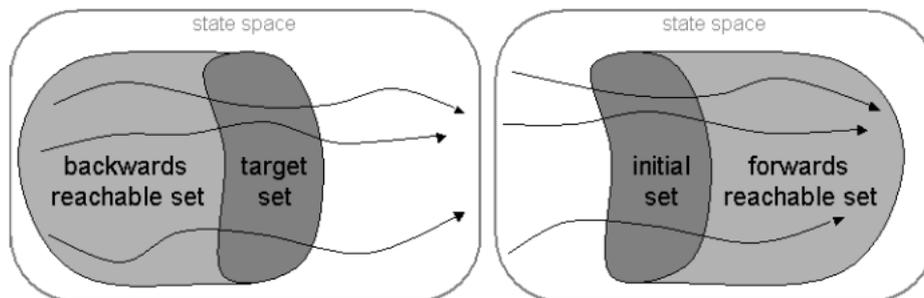
Figure 1: The difference between backwards and forwards reachability
(I. M. Mitchell & Tomlin, 2003).

Primarily, this project will focus on backwards reachable sets of a continuous system (I. M. Mitchell & Tomlin, 2003). For a given target set which contains known unsafe states, then the backwards reachable set possibly contains states which are unsafe and need to be avoided. As an example, consider how to avoid a collision between two aircrafts or vehicles. The target set consists of all states which are in a collision situation where the states are within the five mile separation

distance (A. Bayen, Santhanam, Mitchell, & Tomlin, 2003). The backwards reachable set consists of those states which will result in a collision. In this situation, the backwards reachable set is going to expand or propagate many miles ahead of the aircraft. Much more details of this example will be in the numerical example section. This thesis will show that the viscosity solution $v(x, t)$ (see equation 5 for the definition) of a specific time-dependent Hamilton-Jacobi Isaacs partial differential equation will give an implicit surface representation of the continuous backward reachable set (I. M. Mitchell et al., 2005). Also, a method to deal with the curse of dimensionality is also provided. For example, applying the method of projection via decomposition to compute the true reachable set of BRS into lower dimensional subspaces. This can address the trade off between computation complexity, reduction of small dimensionality and optimality. Also, this method can save a significant quant of time in computations (Chen, Herbert, & Tomlin, 2016a). In previous papers, researchers have studied the problem only in cylindrical coordinates. Thus, the aim of this project is to study a simple example of backward reachable sets in rectangular coordinates, and see how the decomposition method works on linear system. Also, we start computing BRS in a spherical parameterisation with three variants of the 3D Dubins models and see how far we go with this idea.

2 How to compute the reachable set.

2.1 The reachable set.

2.1.1 Defining the reachable set.

This part is going to determine the backwards reachable set of the system. Also, it is going to discuss some properties of the reachable set and determine a terminal value of the Hamilton-Jacobi Isaacs partial differential equation and its viscosity solution illustrates that reachable set (I. M. Mitchell et al., 2005). In one set, there are usually two kinds of parameters or inputs. These parameters which depend on time, describe the control input $a(.)$ and the disturbance input $b(.)$. Usually in the practical systems, the control inputs can be manipulated to force the system to achieve a goal and satisfy a property, whereas the disturbance inputs or environmental disturbances are unknown actions (I. M. Mitchell & Tomlin, 2003).

Figure 2 explains the set that this project is trying to compute. It shows many trajectories ζ_i which are starting at the same time t, but different states x. These trajectories are subjected to different input signals $a(.)$ and $b(.)$. These input signals drive the trajectories to two different outcomes. The input signal $a(.)$ is to drive them away from the target set. However, the input signal $b(.)$ is to drive the trajectories toward the target set.

Note.

In our problem, the trajectory begins at initial time $t < 0$. We want to know if the trajectory has passed into the target set by time zero. Also, sometimes we need to check or look at the length of time for the trajectory to evolve, the differential game notation $\tau = -t$ and the free variables s and r refer to the time in the domain $[t, 0]$ (I. M. Mitchell et al., 2005) and (Bokanowski & Zidani, 2011).
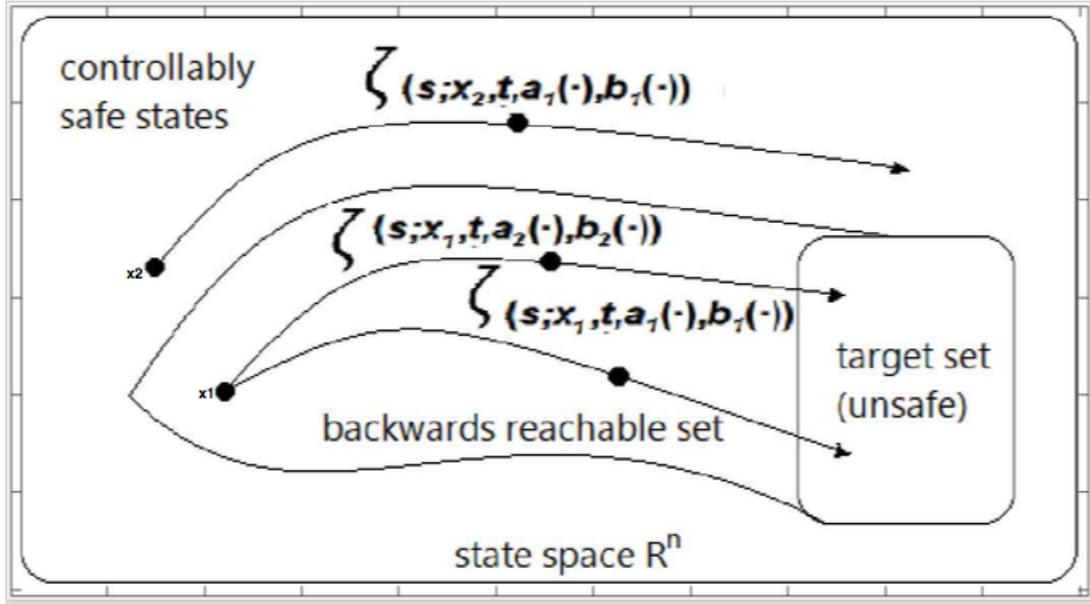
Figure 2: Target set and backwards reachable set.
Adapted from (I. M. Mitchell et al., 2005).

2.1.2 Definition of the full system dynamics.

Let x be our state variable of the system, $a(.)$ is the control input and $b(.)$ is the disturbance input. The state of our system satisfies the ordinary differential equation; therefore, our model can be formulated as

$$\frac{dx}{dt} = \dot{x} = f(x, a, b), t \in [t, 0]. \tag{1}$$

In our discussion, assume that $x \in \mathbb{R}^n$. Generally, the method which is presented can be implemented to periodic state dimensions such as angles. We consider that $\mathcal{A} \subset \mathbb{R}^{n_a}$ and $\mathcal{B} \subset \mathbb{R}^{n_b}$ are compact and $t \in [-\mathbf{T}, 0]$. Thus, the following sets can draw the inputs signals as

$$a(.) \in \mathfrak{A}(t) \triangleq \{\phi : [t, 0] \to \mathcal{A} \mid \phi(.) \text{ is measurable}\}.$$

$$b(.) \in \mathfrak{B}(t) \triangleq \{\phi : [t, 0] \to \mathcal{B} \mid \phi(.) \text{ is measurable}\}.$$

The system dynamics or flow field $f : \mathbb{R}^n \times \mathcal{A} \times \mathcal{B} \to \mathbb{R}^n$ is assumed to be bounded, uniformly continuous and Lipschitz continuous in x for fixed points a and b. Therefore, for that given fixed control input $a(.) \in \mathfrak{A}(t)$ and disturbance input $b(.) \in \mathfrak{B}(t)$ with initial point, there exists a trajectory which is considered to be unique and a solution of ODE 1 (Evans & Souganidis, 1983) and (Coddington & Levinson, 1955). Thus, denote the solutions or the trajectories of the system 1 by

$$\zeta_i(s; x, t, a(.), b(.)) : s \in [\, t, 0],$$

with the initial condition

$$\zeta_i(t; x, t, a(.), b(.)) = x.$$

7

Therefore, the trajectory of the system satisfies that initial condition and the differential equation

$$\frac{d}{ds}\zeta_i(s; x, t, a(.), b(.)) = f(\zeta_i(s; x, t, a(.), b(.)), a(s), b(s)).$$

Note.

The use of a semi-colon is to distinguish between the parameters of the trajectory x, t, $a(.)$ and $b(.)$ and the argument s of ζ_i (I. M. Mitchell et al., 2005). For the reachability problem, denote the target set as $\mathcal{T} \subset \mathbb{R}^n$. The target set is closed, and it can be formulated as the zero sub-level set of the continuous and bounded Lipschitz function $g : \mathbb{R}^n \to \mathbb{R}$

$$\mathcal{T} = \{x \in \mathbb{R}^n \mid g(x) \leqslant 0\}. \tag{2}$$

Supposing that the control input $a(.)$ is going to lead the system far away from the target set. However, the disturbance input $b(.)$ is going to push the system towards the target set. For our specification, let's determine a plan first for the disturbance function $b(.)$ as a map $\gamma : \mathfrak{A}(t) \to \mathfrak{B}(t)$. This map defines the signal of the disturbance's input as a function. This function is drawn from the nonanticipative strategies because it is easy to turn it to a HJ PDE (Bardi, 1997); so,

$$\gamma \in \Gamma(t) \triangleq \{\vartheta : \mathfrak{A}(t) \to \mathfrak{B}(t) \mid a(r) = \hat{a}(r), \forall r \in [t, s] \Rightarrow \vartheta[a](r) = \vartheta[\hat{a}](r), \forall r \in [t, s]\}.$$

The meaning of this restriction is that if the disturbance input is not able to distinguish between the input signals $a(.)$ and $\hat{a}(.)$ for the control input until time s, then the disturbance input is not going to respond in a different way to that signals input until after time s (Chen, Herbert, & Tomlin, 2016b).

2.1.3 The backwards reachable set (BRS).

This part is going to consider a general definition for the backwards reachable set (BRS) which guarantees the safety. It is going to determine the backwards reachable set, provide an explanation about how it can be used as a solution of the Hamilton-Jacobi Isaacs partial differential equation and the techniques of the computation of their approximation. This method can be used for linear and nonlinear systems. Also, the system dynamics can be modelled by ODE and it depends on control and disturbance parameters. For all the types of practical systems, the control parameters can be used as a manipulation to impose the system to reach a goal or a property. However, the disturbance parameters act as uncertainties in the system. Therefore, differential games and optimal control are employed in this project (I. M. Mitchell & Tomlin, 2003). In order to resolve the backwards reachability problem, we define the Maximal BRS and Minimal BRS respectively $\mathcal{R}(\tau)$, $\mathcal{A}(\tau)$. The BRS comprises the set of states $x \in \mathbb{R}^n$ from which the system can be lead into the target set $\mathcal{T}$ at the end of a time horizon where $\mathcal{T} \subseteq \mathbb{R}^n$. In the definition of the Maximal BRS, the system tries to enter the target set or the set of goal states. This definition comprises the set of states where the system is guaranteed to reach $\mathcal{T}$. In the second definition, the BRS represents the set of states which there exits disturbance inputs or strategies for the disturbance inputs for all possible control inputs. This definition comprises the set of states which violates the safety requirements. The mathematical form of

these definitions are as follows:

Definition 1: Maximal BRS.

$$\mathcal{R}(\tau) = \{x : \exists a(.), \forall b(.), \exists s \in [t, 0], \zeta(0; x, t, a(.), b(.)) \in \mathcal{T}\}. \tag{3}$$

Definition 2: Minimal BRS.

$$\begin{aligned}\mathcal{A}(\tau) &= \{x : \forall a(.), \exists b(.), \exists s \in [t, 0], \zeta(0; x, t, a(.), b(.)) \in \mathcal{T}\}. \\ &= \{x : \forall a(.), \exists \gamma \in \Gamma(t), \exists s \in [t, 0], \zeta(0; x, t, a(.), \gamma[a](.)) \in \mathcal{T}\}.\end{aligned} \tag{4}$$

where $b(.) = \gamma[a](.)$. The BRS refers to the system where it can be lead into the target set at the end of a time horizon (Chen, Herbert, Vashishtha, Bansal, & Tomlin, n.d.). These terms of maximal and minimal refer to the role of optimal control (I. M. Mitchell, 2007).

2.1.4 A time-dependent Hamilton-Jacobi-Isaacs equation for the reachable set.

This section is going to demonstrate one of the key theoretical results of this project, namely that the viscosity solution of a time dependent Hamilton-Jacobi-Isaacs equation can determine the backwards reachable set (Crandall, Evans, & Lions, 1984), (Margellos & Lygeros, 2011) and (I. M. Mitchell et al., 2005).

Theorem 1: Reachability Theorem.

The viscosity solution is defined as $v : \mathbb{R}^n \times [-\mathbf{T}, 0] \to \mathbb{R}$ of the terminal value HJ PDE

$$\mathbf{D}_t v(x, t) + \min[0, \mathbf{H}(x, \mathbf{D}_x v(x, t))] = 0, \quad v(x, 0) = g(x). \tag{5}$$

where the Hamiltonian is

$$\mathbf{H}(x, p) = \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} \ p^{\mathbf{T}} f(x, a, b). \tag{6}$$

Here $\mathbf{D}_x v(x, t)$ is the spatial derivative and $\mathbf{D}_t v(x, t)$ is the time derivative (I. Mitchell, 2004) and p is a vector which is called co-state of the system (A. M. Bayen, Mitchell, Oishi, & Tomlin, 2007) and (Lygeros, Tomlin, & Sastry, 1999). Therefore, we define the zero sub-level set of v which describes the backwards reachable set $\mathcal{A}(\tau)$ is defined as follows

$$\mathcal{A}(\tau) = \{x \in \mathbb{R}^n \mid v(x, t) \leqslant 0\}. \tag{7}$$

The BRS $\mathcal{A}(\tau)$ will be always used in the following explanations for a reason which will be known later through theorem 2.

The proof of the theorem is given in full in chapter 3.
Discussion of the proof.

The aim of the proof of this theorem is to show an equivalence between the reachability and a terminal cost differential game. The main problem with this game is that it can only define

the trajectory if it is in the target set at only specific time zero. If this game is employed on the original system, then the control input can avoid the target set by leading the trajectory into the target set, then the trajectory is going to be out from the other side before time zero. To avoid this case, let's provide a definition of a new system which is called augment system and it consists of some information and set of inputs of the disturbance input. In this system, the disturbance is going to freeze or stop the developments of the trajectories that the control input tries to drive them out of the target set. The trajectories of the augmented system are going to be equal to the trajectories of the original system. Also, the results of the HJI PDE of the differential game are going to be equal to

$$\mathbf{D}_t v(x,t) + \min[0, \mathbf{H}(x, \mathbf{D}_x v(x,t))] = 0, \quad v(x,0) = g(x).$$

The importance of this theorem is that we can develop numerical schemes to calculate an exact approximation of the viscosity solution. Then, taking advantage of the viscosity solution with the method of projection via decomposition to get an exact approximation of the backwards reachable set. It may take long steps to reach the final proof of this theorem (I. Mitchell, Bayen, & Tomlin, 2004). Using this formula is more efficient because it is restricted to the convexity of the Hamiltonian and target set (I. Mitchell & Tomlin, 2000).

3 Proof of the time-dependent formulation.

As stated before the proof of reachability theorem depends on the viscosity solution and the differential game which will follow in the coming sections.

3.1 Augmenting the dynamics.

This section is going to employ set of dynamics systems which are modified to augment the inputs of the disturbance with the scalar function

$$\underline{b}(.) \in \underline{\mathfrak{B}}(t) \triangleq \{\phi : [t,0] \to [0,1] \mid \phi(.) \text{ is measurable}\}.$$

The augmented input of the disturbance is

$$\tilde{b} = [b \quad \underline{b}] \in \mathcal{B} \times [0,1],$$

and in the same way $\tilde{\mathcal{B}}(t)$ and $\tilde{\mathfrak{B}}(t)$. Applying the differential game with the following dynamics

$$\tilde{f}(x, a, \tilde{b}) \triangleq \underline{b} f(x, a, b). \tag{8}$$

and its trajectory shall be indicated by $\zeta_{\tilde{f}}(s; x, t, a(.), \tilde{b}(.))$. There are three options which might be chosen by the disturbance to play the game. The three options are as follow; to select normal

dynamics by putting $\underline{b} = 1$, to pick out slowed dynamics by choosing $\underline{b} \in]0,1[$, and the last option is to choose freezing the dynamics totally by choosing $\underline{b} = 0$. We will use the third option because the disturbance input has the ability to stop the developments of the trajectory of the augmented system. Therefore, it keeps the evolution of the trajectory from leaving the target set when it enters. Obviously, it can be seen that there is a close relationship between the trajectories of the augmented system (8) and the trajectories of the original system (1). Thus, in one of the following lemmas will show the equivalence of trajectories. However, before stating that lemma, we define the pseudo-time variable $\sigma : [t, 0] \to [t, 0]$ to shape that relationship which for any $\underline{b} \in \underline{\mathfrak{B}}(t)$ is provided by

$$\sigma(s) \triangleq t + \int_t^s \underline{b}(\lambda)d\lambda. \tag{9}$$

$\underline{b}(.)$ is measurable by the definition of $\underline{b}(.)$. It is clear that σ is absolutely continuous from the definition of σ in equation 9. However, σ is not necessarily differentiable. A simple step function is an example of a measurable function $\underline{b}(.)$ for which the function σ defined in equation 9 is not differentiable. There is no genuine inverse of σ because there is no restriction of its monotonic increase. However, the proof is going to use a function σ^+, which is nearly considered to be an inverse of σ where $\sigma(\sigma^+(s)) = s$ makes sense.

The following lemma justifies the change of variables and determines σ^+ as the limit of a convergent sequence of functions

Lemma 1.

$$\int_t^{\sigma(s)} f(\zeta_f(p), a(\sigma^+(p)), b(\sigma^+(p)))dp$$
$$= \int_t^s f(\zeta_f(\sigma(\lambda)), a(\lambda), b(\lambda))\underline{b}(\lambda)d\lambda. \tag{10}$$

Proof. According to (I. M. Mitchell et al., 2005), there exits $\underline{b}_k$ a sequence of smooth functions for $k = \{1, 2, 3, \ldots\ldots\}$, such that $\underline{b}_k(s) \in [\frac{1}{k}, 1 + \frac{1}{k}]$ for all s and $\underline{b}_k$ converges pointwise to $\underline{b}$ as $k \to \infty$. The smooth functions are defined as follows

$$\sigma_k(s) \triangleq t + \int_t^s \underline{b}_k(\lambda)d\lambda.$$

When $k \to \infty$, then σ_k converges uniformly to σ. σ_k is going to be bounded and strictly monotonically increasing and there exists a smooth σ^{-1}_k because of the bounds on s and $\underline{b}_k$. Then by Helly's theorem (Ash, 2014), there exists a monotonically increasing right-continuous σ^+ and a subsequence $\sigma_{k_n}^{-1}$ such that $\sigma^+ = \lim_{k_n \to \infty} \sigma_{k_n}^{-1}$. If there are smooth $\sigma_{k_n}^{-1}$, σ_{k_n} and Lipschitz continuous f, then the change of variables which is denoted by $p = \sigma_{k_n}(\lambda)$ will be justified for every k_n:

$$\int_t^{\sigma_{k_n}(s)} f(\zeta_f(p), a(\sigma_{k_n}^{-1}(p)), b(\sigma_{k_n}^{-1}(p)))dp$$
$$= \int_t^s f(\zeta_f(\sigma_{k_n}(\lambda)), a(\lambda), b(\lambda))\underline{b}_{k_n}(\lambda)d\lambda.$$

To prove number (10), we state the following theorem.

Dominated convergence theorem (Lebesgue dominated convergence theorem).

From (Chae & Kim, 2014), let consider that (X, A, μ) a measure space and $f_n : \mathbb{R} \to [0, \infty]$ are (Lebesgue) measurable functions on X, and let g be a $[0, \infty]$-valued integrable function on X. Suppose that $f(x) = \lim_n f_n(x)$. Supposing that there is an integrable $\mid f_n(x) \mid \leqslant g(x), n = 1, 2, \dots..$ holds for every $x \in X$. Therefore; f_n is integrable and

$$\int_{\mathbb{X}} f d\mu = \lim_n \int_{\mathbb{X}} f_n d\mu.$$

Implies,

$$\int_{\mathbb{X}} \lim_n f_n(x) d\mu = \lim_n \int_{\mathbb{X}} f_n d\mu.$$

holds.

Therefore, applying the dominated convergence theorem and take the limit as $k_n \to \infty$ to be inside the integral to reach (10). $\qquad\square$

Equivalence of trajectories lemma 2.

For any $a(.) \in \mathfrak{A}(t)$ and $\tilde{b}(.) = [b(.) \quad \underline{b}(.)] \in \tilde{\mathfrak{B}}(t)$, where

$$\sigma(s) \triangleq t + \int_t^s \underline{b}(\lambda) d\lambda$$

and σ^+ as defined in lemma 1, there exists a trajectory of the augmented system

$$\tilde{f}(x, a, \tilde{b}) \triangleq \underline{b} f(x, a, b)$$

for every trajectory of the original system

$$\frac{dx}{dt} = \dot{x} = f(x, a, b).$$

Thus, these trajectories are connected through the pseudo-time variable σ

$$\zeta_f(\sigma(s); x, t, a(\sigma^+(.)), b(\sigma^+(.))) = \zeta_{\tilde{f}}(s; x, t, a(.), \tilde{b}(.))$$

for any $s \in [t, 0]$.

Proof. In (Boyce & DiPrima, 1986), we adopt the proof of this lemma from the definition of a classical ordinary differential equation uniqueness which is provided below.

Definition.

We consider that there is an ODE

$$y'(t) = f(y(t)).$$

where $t > 0, y(0) = a \in \mathbb{R}$
assuming that $f : \mathbb{R} \to \mathbb{R}$ is smooth and is Lipschitz (with $k = 1$), meaning that

$$|f(t) - f(s)| \leqslant |t - s|.$$

Having $0 < \mathbf{T} < \infty$, then $y(t)$ is a classical solution (differentiable as many times as needed) of the ODE given $y \in \mathbf{C}[0, \mathbf{T}]$. Where $\mathbf{C}[0, \mathbf{T}]$ means the set of continuous functions on the interval $[0, \mathbf{T}]$ and

$$y(t) = a + \int_0^t f(y(\tau))d\tau, \forall t \in [0, \mathbf{T}].$$

Now, introduce the shorthand

$$
\begin{aligned}
\zeta_f(s) &\triangleq \zeta_f(s; x, t, a(\sigma^+(.)), b(\sigma^+(.))), \\
\zeta_{\tilde{f}}(s) &\triangleq \zeta_{\tilde{f}}(s; x, t, a(.), \tilde{b}(.)).
\end{aligned}
\tag{11}
$$

Thus, these in (11) can be written as the classical of the ODE

$$\zeta_{\tilde{f}}(s) = \zeta_{\tilde{f}}(t) + \int_t^s \frac{d\zeta_{\tilde{f}}(\lambda)}{d\lambda}d\lambda.$$

$$= x + \int_t^s f(\zeta_{\tilde{f}}(\lambda), a(\lambda), b(\lambda))\underline{b}(\lambda)d\lambda.$$

and

$$\zeta_f(\sigma(s)) = \zeta_f(t) + \int_t^{\sigma(s)} \frac{d\zeta_f(p)}{dp}dp.$$

$$= x + \int_t^{\sigma(s)} f(\zeta_f(p), a(\sigma^+(p)), b(\sigma^+(p)))dp.
\tag{12}$$

$$= x + \int_t^s f(\zeta_f(\sigma(\lambda)), a(\lambda), b(\lambda))\underline{b}(\lambda)d\lambda.$$

since the change of variables $p = \sigma(\lambda)$ is already justified in lemma 1. Therefore, from the fact that $\underline{b}(\lambda) \in [0, 1]$ and the last two equations, then

$$\left\| \zeta_f(\sigma(s)) - \zeta_{\tilde{f}}(s) \right\|$$

$$\leqslant \int_t^s \left\| \left(f(\zeta_f(\sigma(\lambda)), a(\lambda), b(\lambda)) - f(\zeta_{\tilde{f}}(\lambda), a(\lambda), b(\lambda)) \right)\underline{b}(\lambda) \right\| d(\lambda).$$

$$\leqslant \int_t^s \left\| \left(f(\zeta_f(\sigma(\lambda)), a(\lambda), b(\lambda)) - f\left(\zeta_{\tilde{f}}(\lambda), a(\lambda), b(\lambda) \right) \right) \right\| d(\lambda).
\tag{13}$$

$$\leqslant \mathbf{K} \int_t^s \left\| \zeta_f(\sigma(\lambda)) - \zeta_{\tilde{f}}(\lambda) \right\| d(\lambda).$$

where K is the constant of the Lipschitz for the flow field f. Let

$$\psi(s) = \int_t^s \left\| \zeta_f(\sigma(\lambda)) - \zeta_{\tilde{f}}(\lambda) \right\| d\lambda,$$

It can be seen that $\psi(t) = 0$, $0 \leqslant \psi(s)$ and $\dot{\psi} = \left\| \zeta_f(\sigma(s) - \zeta_{\tilde{f}}(s) \right\|$

Rewriting (13) in terms of ψ to gain the following differential inequality

$$\dot{\psi}(s) - K\psi(s) \leqslant 0,$$

where its solution is only $\psi(s) \equiv 0$ (Boyce & DiPrima, 1986). Therefore,

$$\zeta_f(\sigma(s)) = f(\zeta_{\tilde{f}}(s)).$$

$\square$

Corollary 1.

From the definitions in (11), the trajectory $\zeta_{\tilde{f}}(.)$ of the augmented system visits only a subset of the states which is visited by the trajectory $\zeta_f(.)$ of the original system; in particular the trajectory that visited in the time interval $[t, \sigma(s)]$.

3.2 The differential game and its solution.

This section is going to deal with a finite horizon differential game over the horizon time $[-\mathbf{T}, 0]$. The dynamics of the finite horizon differential game are dominated by the following flow field

$$\tilde{f}(x, a, \tilde{b}) \triangleq \underline{b} f(x, a, b).$$

The trajectory of this game has a terminal estimate

$$\mathbf{C}(x, t, a(.), \tilde{b}(.)) = g(\zeta_{\tilde{f}}(0; x, t, a(.), \tilde{b}(.))).$$

The aim of the control input is going to maximize that cost whereas the disturbance input is going to try to minimize it. As a result, our differential game's value is going to be

$$
\begin{aligned}
v(x, t) &= \inf_{\tilde{\gamma} \in \tilde{\Gamma}(t)} \sup_{a(.) \in \mathfrak{A}(t)} \mathbf{C}(x, t, a(.), \tilde{\gamma}[a](.)). \\
&= \inf_{\tilde{\gamma} \in \tilde{\Gamma}(t)} \sup_{a(.) \in \mathfrak{A}(t)} g(\zeta_{\tilde{f}}(0; x, t, a(.), \tilde{\gamma}[a](.))).
\end{aligned}
\tag{14}
$$

Lemma 3.

In our differential game, the value function $v(x,t)$ is the viscosity solution of the Hamilton-Jacobi-Isaacs terminal value PDE

$$\mathbf{D}_t v(x,t) + \tilde{\mathbf{H}}(x, \mathbf{D}_x v(x,t)) = 0, v(x,0) = g(x). \tag{15}$$

where

$$\tilde{\mathbf{H}}(x,p) = \max_{a \in \mathcal{A}} \min_{\tilde{b} \in \tilde{\mathcal{B}}} p^{\mathbf{T}} \tilde{f}(x, a, \tilde{b}). \tag{16}$$

Proof. The proof is just a special case of the Theorem 4.1 which can be found in (Evans & Souganidis, 1983). $\qquad\square$

Lemma 4.

For $t \in [-\mathbf{T}, 0]$, the value function which is given in (14) characterises the reachable set $\mathcal{A}(\tau)$ according to (7).

Proof. Let's show that

$$x \in \mathcal{A}(\tau) \Rightarrow v(x,t) \leqslant 0 \tag{17}$$

$$v(x,t) \leqslant 0 \Rightarrow x \in \mathcal{A}(\tau) \tag{18}$$

Case one:

We consider that $x \in \mathcal{A}(\tau)$ and $v(x,t) > 0$, and derive a contradiction. Consider the implications of (14).

$$v(x,t) = \inf_{\tilde{\gamma} \in \tilde{\Gamma}(t)} \sup_{a(.) \in \mathfrak{A}(t)} \mathbf{C}(x, t, a(.), \tilde{\gamma}[a](.)) > 0,$$
$$\Rightarrow \exists \epsilon > 0, \forall \tilde{\gamma} \in \tilde{\Gamma}(t),$$
$$\sup_{a(.) \in \mathfrak{A}(t)} \mathbf{C}(x, t, a(.), \tilde{\gamma}[a](.)) > 2\epsilon > 0, \tag{19}$$
$$\Rightarrow \exists \epsilon > 0, \forall \tilde{\gamma} \in \tilde{\Gamma}(t), \exists \hat{a}(.) \in \mathfrak{A}(t),$$
$$\mathbf{C}(x, t, \hat{a}(.), \tilde{\gamma}[\hat{a}](.)) > \epsilon > 0.$$

Consider the implication of $x \in \mathcal{A}(\tau)$. From the backward reachable set in (4), we have $\gamma \in \Gamma(t)$. Thus, there exists $s \in [t, 0]$ such that $\zeta_f(s; x, t, \hat{a}(.), b(.)) \in \mathcal{T}$ for the $\hat{a}(.)$ from (19) and $b(.) = \gamma[\hat{a}](.)$. Also, from (2), $g(\zeta_f(s; x, t, \hat{a}(.), b(.))) \leqslant 0$. We select the freezing input signal

$$\underline{b}(r) = \begin{cases} 1, & \text{for } r \in [t, s[; \\ \\ 0, & \text{for } r \in [s, 0[. \end{cases}$$

15

Combining the previous $\underline{b}(.)$ with the $b(.)$ that is already chosen to get $\tilde{b}(.)$. This input is going to generate a trajectory

$$\zeta_{\tilde{f}}(r; x, t, \hat{a}(.), \tilde{b}(.)) = \begin{cases} \zeta_f(r; x, t, \hat{a}(.), b(.)), \, for \, r \in [t, s]; \\ \\ \zeta_f(s; x, t, \hat{a}(.), b(.)), \, for \, r \in [s, 0]. \end{cases}$$

Particularly,

$$\zeta_{\tilde{f}}(0; x, t, \hat{a}(.), \tilde{b}(.)) = \zeta_f(s; x, t, \hat{a}(.), b(.))$$

$$\Rightarrow g(\zeta_{\tilde{f}}(0; x, t, \hat{a}(.), \tilde{b}(.))) = \mathbf{C}(x, t, \hat{a}(.), \tilde{b}(.)) \leqslant 0.$$

It is already known that $b(.) = \gamma[\hat{a}](.)$ is nonanticipative, thus; let's design a nonanticipative strategy for $\underline{b}(.)$ with extra constraints on s. Therefore, $\tilde{b}(.)$ is nonanticipative which is a contradiction of (19). As a result, equation (17) is proved.

Case two:

Suppose that $v(x, t) \leqslant 0$ and $x \notin \mathcal{A}(\tau)$ and prove a contradiction. Consider the implications of $x \notin \mathcal{A}(\tau)$, and suppose that (4) is rejected

$$\forall \gamma \in \Gamma(t), \exists \hat{a}(.) \in \mathfrak{A}(t), \forall s \in [t, 0], \zeta_f(s; x, t, a(.), \gamma[a](.)) \notin \mathcal{T}. \tag{20}$$

Now, if any strategy $\tilde{\gamma} \in \tilde{\Gamma}(t)$ in the augmented system of the disturbance input is considered then it can be found that the response $\gamma \in \Gamma(t)$ of the disturbance input in the original system is by deleting the final elements of its output (components of $\tilde{\Gamma}(t)$ and $\Gamma(t)$ accept the same input function which is derived from $\mathfrak{A}(t)$. By selecting $\hat{a}$ which corresponds to γ in (20), and using corollary 1 to find out that the set of states which is visited by the augmented trajectory is a subset of the states which is visited by the original trajectory. Therefore, in combination with (20)

$$\zeta_f(s; x, t, \hat{a}(.), \gamma[\hat{a}](.)) \notin \mathcal{T} \, \forall s \in [t, 0] \Rightarrow \zeta_{\tilde{f}}(s; x, t, \hat{a}(.), \tilde{\gamma}[\hat{a}](.)) \notin \mathcal{T} \, \forall s \in [t, 0]. \tag{21}$$

The structure of g with $\zeta_{\tilde{f}}$ is a continuous function from the close interval $[t, 0]$ to $\mathbb{R}$, thus, it reaches its extrema. To conclude by (2) and (21), there exists $\delta > 0$ such that

$$g(\zeta_{\tilde{f}}(s; x, t, \hat{a}(.), \tilde{\gamma}[\hat{a}](.))) \geqslant \delta \, \forall s \in [t, 0]$$
$$\Rightarrow \mathbf{C}(x, t, \hat{a}(.), \tilde{\gamma}[\hat{a}](.)) = g(\zeta_{\tilde{f}}(0; x, t, \hat{a}(.), \tilde{\gamma}[\hat{a}](.))) \geqslant \delta \tag{22}$$

Now, going back to (14)

16

$$v(x,t) = \inf_{\tilde{\gamma} \in \tilde{\Gamma}(t)} \sup_{a(.) \in \mathfrak{A}(t)} \mathbf{C}(x, t, a(.), \tilde{\gamma}[a](.)) \leqslant 0,$$

$$\Rightarrow \forall \epsilon > 0, \exists \tilde{\gamma} \in \tilde{\Gamma}(t), \sup_{a(.) \in \mathfrak{A}(t)} \mathbf{C}(x, t, a(.), \tilde{\gamma}[a](.)) \leqslant \epsilon,$$

$$\Rightarrow \forall \epsilon > 0, \exists \tilde{\gamma} \in \tilde{\Gamma}(t), \forall a(.) \in \mathfrak{A}(t), \mathbf{C}(x, t, a(.), \tilde{\gamma}[a](.)) \leqslant \epsilon,$$

Selecting $\epsilon = \frac{\delta}{2}$ to infer a contradiction of (22), then proves (18). $\qquad \square$

Right now, we prove directly theorem 1.
Proof of theorem 1.

Proof. It is known that from lemma 3 the value function v is the viscosity solution to the HJI PDE (15) of the differential game (14). Also, from lemma 4 it is known that the reachable set is described by (7). Therefore, beginning with $\tilde{\mathbf{H}}$ from (16) and $\mathbf{H}$ from (6), then

$$\tilde{\mathbf{H}}(x, p) = \max_{a \in \mathcal{A}} \min_{\tilde{b} \in \tilde{\mathcal{B}}} p^{\mathbf{T}} \tilde{f}(x, a, \tilde{b})$$

,

$$= \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} \min_{\underline{b} \in [0,1]} p^{\mathbf{T}} (\underline{b} f(x, a, b)),$$

$$= \min_{\underline{b} \in [0,1]} \underline{b} \left(\max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} p^{\mathbf{T}} f(x, a, b) \right),$$

$$= \min[0, \mathbf{H}(x, p)].$$

Hence, the solutions to the two HJI PDEs (5) and (15) are equal. Then v is the solution of (5). Therefore, it is shown that the reachability is equivalent to a terminal cost differential game (I. M. Mitchell et al., 2005). $\qquad \square$

4 Numerical example.

4.1 Aircraft collision avoidance example.

This part is going to describe the method in solving (5) computationally, then employ that to a classical collision avoidance differential game. For a high air traffic control, the results of this game employed to a collision alert system due to the simplicity of the dynamics of that game. For instance, this example will use the form $f(x, a, b) = f_1(x) + f_2(x)a + f_3(x)b$ which is taken in (23) below. The aim of this example is to define the reachability for two adversarial vehicles (the evader and the pursuer) of a three dimensional kinematic model. In this game, the pursuer tries to get to the evader within a certain distance r, and the reachable set corresponds to the set where the second player can capture the first player. In recent publications, this game or problem is called the three dimensional aircraft collision avoidance example (C. J. Tomlin et

al., 2000) and (I. Mitchell, Bayen, & Tomlin, 2001). Let's model every vehicle as a simple kinematic system with a fixed linear velocity, controllable angular velocity and heading. A collision occurs, which is often called a loss of separation, if the vehicles come of one another within the defined distance. The model for each vehicle is represented as:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} v\cos x_3 \\ v\sin x_3 \\ w \end{bmatrix}$$

Where $[x_1 \; x_2]^{\mathbf{T}} \in \mathbb{R}^2$ is the location of the vehicles in the plane, $v \geqslant 0$ is the constant linear velocity, $x_3 \in [0, 2\pi]$ is its heading and w is the angular velocity of the vehicles. In (Crandall & Lions, 1984), Lax-Friedrichs (LF) approximation is used in the Hamiltonian to assure the stability. For any, not necessarily continuous function p the form of LF approximation can be defined as follows

$$\mathbf{H}(x, p^+, p^-) = \mathbf{H}\left(x, \frac{p^- + p^+}{2}\right) - \frac{1}{2}\alpha^{\mathbf{T}}(p^+ - p^-),$$

where p^- and p^+ refer to the left and right approximation of p. Also, the elements of the vector $\alpha \in \mathbb{R}^n$ depend on the partial derivatives of $\mathbf{H}$ with respect to its second argument where

$$\alpha_i = \max\left|\frac{\partial \mathbf{H}}{\partial p_i}\right|.$$

This game allows the players to choose their angular velocity. Therefore, instead of using w, let $a \in \mathcal{A} = [-\alpha, +\alpha]$ to be the angular velocity input for the evader player whereas $b \in \mathcal{B} = [-\alpha, +\alpha]$ is to be the angular velocity for the pursuer player, $v = 5$ is the linear velocity and $r = 5$ is the collision distance (Mitchell. I and Tomlin.C, 2003). Now, our aim is to define the set of states from which the second player could cause a collision to occur. Let's work on the relative coordinates $x \in \mathbb{R}^2 \times [0, 2\pi[$ because the system is simplified down into three dimensions and $\mathcal{T}$ relies on the relative positions of the vehicles. As shown in figure 3
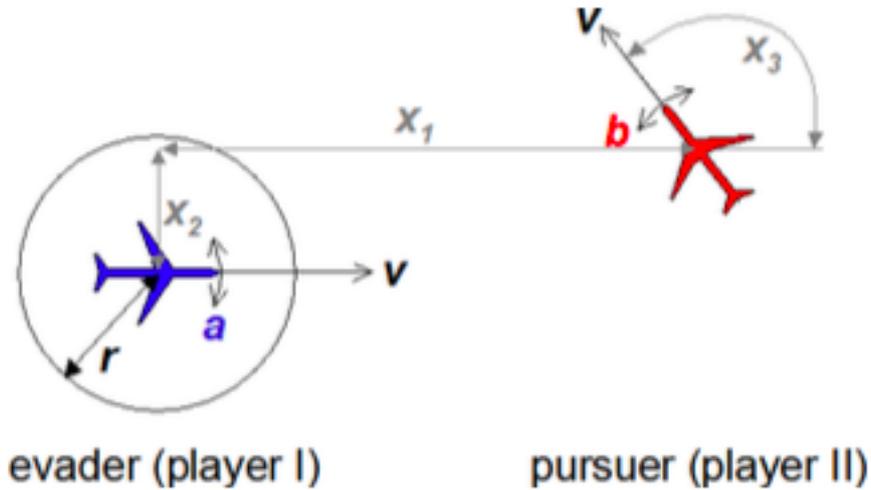
Figure 3: Relative coordinate system for collision example
(I. M. Mitchell et al., 2005).

locating the first player at the origin and facing it along the positive x_1 axis. Then, the relative

location and heading of the second player are characterized by the flow field as follow

$$\dot{x} = \frac{d}{dt} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -v_a + v_b \cos x_3 + ax_2 \\ v_b \sin x_3 - ax_1 \\ b - a \end{bmatrix} = f(x, a, b) \tag{23}$$

where v_a and v_b are the linear velocity.

In (I. M. Mitchell et al., 2005), the resulting Hamiltonian is

$$\mathbf{H}(x, p) = \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} [p^\mathbf{T} f(x, a, b)]$$
$$= \left(-p_1 v_a + p_1 v_b \cos x_3 + p_2 v_b \sin x_3 + \alpha |p_1 x_2 - p_2 x_1 - p_3| - \beta |p_3| \right). \tag{24}$$

Because the collision can happen at any relative heading, then the target set $\mathcal{T}$ relies only on x_1 and x_2, with any state within distance r of the the planar origin

$$\mathcal{T} = \{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 \leqslant r^2\}.$$

This can be converted into signed distance function for our HJI PDEs terminal conditions as follows

$$g(x) = \sqrt{x_1^2 + x_2^2} - r. \tag{25}$$

If $\sqrt{x_1^2 + x_2^2} \leqslant r$, then the collision occurs for any value $x_3 \in \mathbb{R}^3$. The set for which this holds is a cylinder of radius r centered on the x_3 axis. To solve this pursuit evasion game, we need to define the set of initial states where the second player can cause a collision even though the best efforts which can be done by the first player. Thus, the backwards reachable set needs to be defined in the limit as $\tau \to \infty$ by computing (5) backwards from $t = 0$ until finding $\mathbf{H}(x, \mathbf{D}_x v) \approx 0$. If the parameters $r = 5$, $v_a = v_b = 5$ and $\mathcal{A} = \mathcal{B} = [-1, +1]$ are considered, then the determination of the backwards reachable set occurs at $t = -2.6$ which is the convergence time. When the vehicles are identical $v_a = v_b$ and $\mathcal{A} = \mathcal{B}$, then it is possible to define the optimal inputs for both players by using the differential game theory.

Figure 4 shows the target set which is the solid cylinder and backwards reachable set (which is transparent) for the collision avoidance in 3D. However, figure 5 shows the accuracy on the top half backwards reachable set which is a solid and it analytically defines $M = 2612$ points on the boundary of the reachable set or its surface (dots). In figure 6, the target set $\mathcal{T}$ or the collision cylinder shows on the far left. However, the rest of the images shows the growth of the reachable set as τ increases from zero. Therefore, for the previous parameters the reachable set converges to a fixed point for $\tau = -2.6$. Figure 7 represents several views of that fixed point. There will be a collision if the second vehicle starts from within the reachable set. The second vehicle would choose a suitable input b no matter what input a which might be chosen by the first vehicle. On the other hand, if the pursuer begins from outside the reachable set, then there exists an input a which might be chosen by the evader leads to avoid the collision no matter what input b might be chosen by the second player. Figure 8 shows an annotated frame from an animation of the collision system. Also, figure 9 shows the series of frames from that animation. The first player begins from the left and it is surrounding by a solid collision set whereas the second player begins on the right side. The dotted shape which surrounds the evader represents the slice of the reachable set. For instance, the vehicles in the leftmost figure have relative heading $x_3 \approx \pi$. Choosing a suitable input leads the first player to not allow
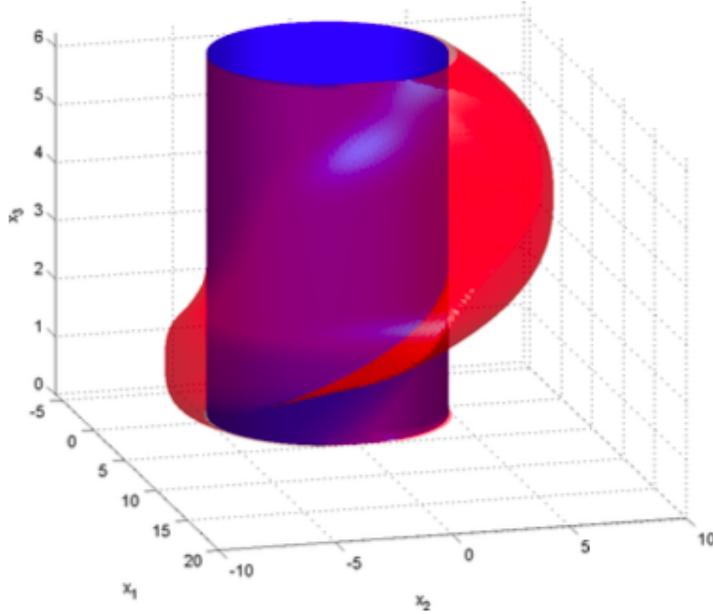
Figure 4: Target set (solid cylinder) and BRS (transparent)
(I. M. Mitchell et al., 2005).

the second player from entering the reachable set. Therefore, it keeps the pursuer away from causing a collision. Figure 10 shows the sequence from where the pursuer begins within the reachable set and causes a collision.

Note that the aircraft collision avoidance example is computed in full state space of the system. The system contains the control and disturbance inputs while other systems can be used with only the control input. The following explanation will involve the control and disturbance input. However, in the example of Dubins car will only use the control input. Thus, one goal of this thesis is to extend the Dubins car example to spherical coordinates and check what results will be obtained. However, before that example, we will work in the example of linear systems of ODEs. Now, after the determination of the reachability problem and its solution, the next section is going to characterize the reachability technique based on a projection via decomposition. It simply depends on the previous augment system where $v(x, t)$ is the viscosity solution of HJ PDE 5. Different methods can be joined together to achieve better results and reduce the problem with dimensionality (Chen et al., n.d.).

5   Reachability computation via decomposition.

The Hamilton-Jacobi-Isaacs formulation provides a method for defining the set of reachable states of a continuous dynamic game. In principle, it is known the major problem with that computation is the expense in computing the full reachable set. There is a method which reduces the cost of computation and computes the full reachable set. Thus, this method can be employed in lower dimensional subsystems. The main goal of this development is to compute exactly the true reachable set of the lower dimensional subsystems. Thus, avoiding the need of computing an expensive full dimensional BRS (I. M. Mitchell & Tomlin, 2003). Then, defining the HJ based in SCSs formulation to compute the BRS by the following algorithm:

1) Define the value function or the implicit surface function which represents the subsystem target sets $\mathcal{T}_1$ and $\mathcal{T}_2$.

2) For the self contained subsystems and for $i = 1, 2$, compute the BRS by solving (5) over the
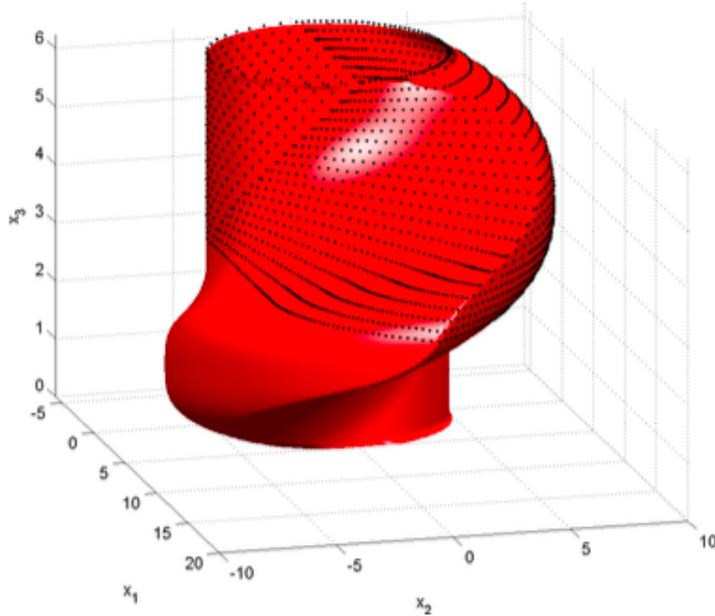
Figure 5: BRS (solid) and the determined points on the reachable set's boundary
(I. M. Mitchell et al., 2005).

Figure 6: Growth of the reachable set
(I. M. Mitchell & Tomlin, 2003).

space $\mathcal{Z}_i$.

3) Construct the full dimensional BRS which by theorem 2 gives results that the full dimensional BRS is precisely constructed (Chen et al., n.d.). The computation of the full formulation is given by the following algorithm:

4) Determine the implicit surface representation or the value function $g(x)$.

5) Solve the HJ PDE with the Hamiltonian to get the implicit surface $V(t, x)$ which represents the BRS.

5.1 Problem formulation.

This part tries to obtain the backwards reachable set by using the computations in lower dimensional subspaces. Supposing that the full system dynamics in (1) can be decomposed into self-contained subsystems (SCSs). Now, we describe some important definitions which are required to state the major results in an accurate way.
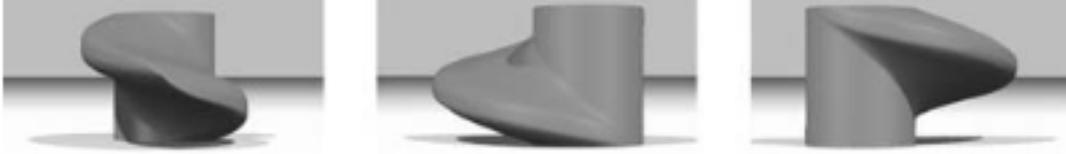
Figure 7: Other views of the reachable set
(I. M. Mitchell & Tomlin, 2003).
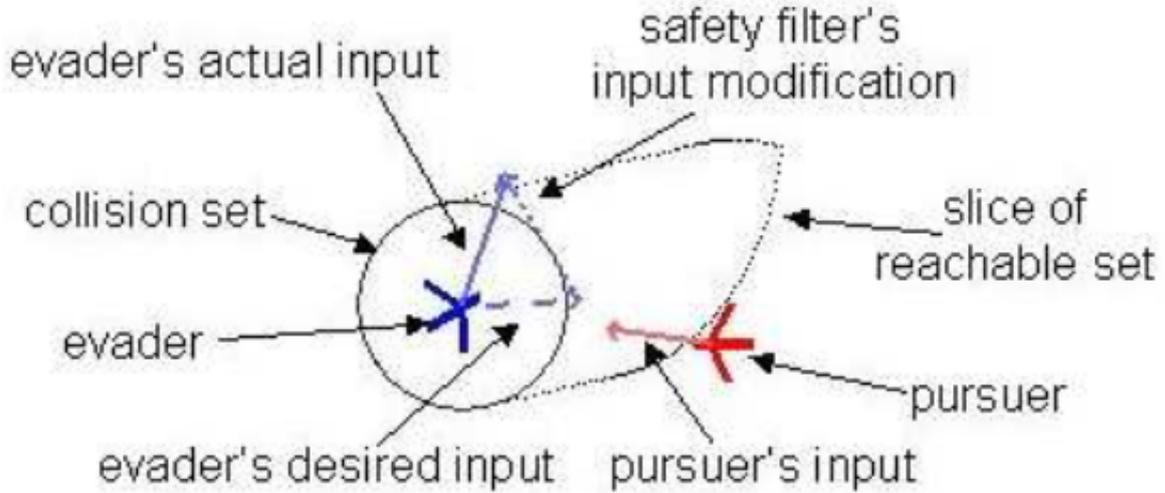
Figure 8: Annotated frame from collision avoidance example animation
(I. M. Mitchell & Tomlin, 2003).

5.2 Subsystem dynamics.

The full system $x \in \mathbb{R}^n$ is partitioned as follows

$$
\begin{aligned}
&x = (x_1, x_2, x_3), \\
&x_1 \in \mathbb{R}^{n_1}, x_2 \in \mathbb{R}^{n_2}, x_3 \in \mathbb{R}^{n_3}, \\
&n_1, n_2 > 0, n_3 \geqslant 0
\end{aligned}
\tag{26}
$$

Where $x_1 = p_x$, $x_2 = p_y$ and $x_3 = \theta$. Note that n_3 can be zero and $n_1 + n_2 + n_3 = n$. The variables x_i are called the state partitions of the full system, x_1 is the state which belongs to the subsystem one, x_2 is the state which belongs to the subsystem two and x_3 is the state which belongs to both subsystems. Therefore, the system dynamic in (1) becomes

$$
\begin{aligned}
\dot{x}_1 &= f_1(x_1, x_2, x_3, a, b). \\
\dot{x}_2 &= f_2(x_1, x_2, x_3, a, b). \\
\dot{x}_3 &= f_3(x_1, x_2, x_3, a, b).
\end{aligned}
\tag{27}
$$

Figure 9: Evader keeps pursuer from entering reachable set, and hence avoids collision
(I. M. Mitchell & Tomlin, 2003).

Figure 10: Pursuer starts within the reachable set, and thus can cause a collision despite the best efforts of the evader
(I. M. Mitchell & Tomlin, 2003).

5.3   Definition of self-contained subsystem.

Determine the SCSs states by grouping the previous states $z_1 \in \mathcal{Z}_1 = \mathbb{R}^{n_1+n_3}$, $z_2 \in \mathcal{Z}_2 = \mathbb{R}^{n_2+n_3}$ as follows:

$$z_1 = (x_1, x_3).$$

$$z_2 = (x_2, x_3).$$

Note that z_1 and z_2 have a common overlapping partition which is x_3. This thesis will only consider two subsystems but this method applies to a finite number of subsystems. The states z_i are called a self-contained subsystem where the states develop according to

$$
\begin{aligned}
\frac{dz_1}{dt} &= \dot{z}_1 = f_1(z_1, a, b) = f_1(x_1, x_3, a, b), \\
\frac{dz_2}{dt} &= \dot{z}_2 = f_2(z_2, a, b) = f_1(x_2, x_3, a, b), \\
\frac{dz_3}{dt} &= \dot{z}_3 = f_3(z_3, a, b) = f_3(x_3, a, b), t \in [t, 0],
\end{aligned}
\tag{28}
$$

where $a \in \mathfrak{A}, b \in \mathfrak{B}$

The dynamics of the two subsystems; one and two respectively are

$$\dot{z}_1 = f_1(x_1, x_3, a, b),$$
$$\dot{z}_3 = f_3(x_3, a, b),$$

23

$$\dot{z}_2 = f_2(x_2, x_3, a, b),$$
$$\dot{z}_3 = f_3(x_3, a, b),$$

The two subsystems are coupled via the partition x_3, the control a and the disturbance b.

5.4 Projection operators.

The projection of our state $x = (x_1, x_2, x_3)$ into a subsystem state space $\mathbb{R}^{n_i + n_3}$ is going

$$proj_i(x) = z_i = (x_i, x_3), i = 1, 2. \tag{29}$$

A point in the full dimensional state space is projected onto a point in the subsystem space. Also, the back-projection operator is needed because our goal is to link the backwards reachable sets of the subsystems to the backwards reachable sets of the full system as follows

$$proj^{-1}(z_i) = \{x \in \mathcal{X} : proj_i(x) = z_i = (x_i, x_3)\} \tag{30}$$

Also, applying the back-projection operator on subsystem sets $\mathcal{S}_i \subseteq \mathcal{Z}_i$. Therefore, overloading the projection operator as follow:

$$proj^{-1}(\mathcal{S}_i) = \{x \in \mathcal{X} : \exists z_i \in \mathcal{S}_i, proj_i(x) = z_i = (x_i, x_3)\} \tag{31}$$

Figure 11 describes the back-projection operator of sets $\mathcal{S}_1$ and $\mathcal{S}_2$ in the $x_1 - x_3$ and $x_2 - x_3$
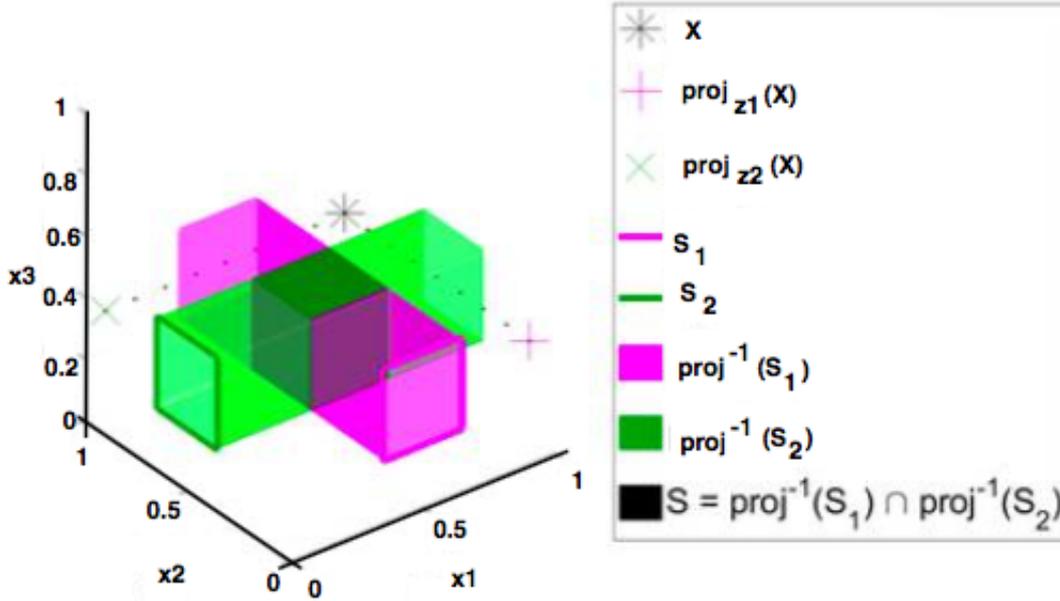
Figure 11: Back projection of sets in the $x_2 - x_3$ plane and the $x_1 - x_3$ plane into the 3D space (Chen et al., 2016a).

respectively. This is going to form the intersection of these sets which is shown as the black cube $\mathcal{S}$.

5.5 Subsystem trajectories.

It is already known that every subsystem in (28) is self-contained. The trajectories of subsystem will be denoted by $\xi_i(s; z_i, t, a(.), b(.))$. The subsystems trajectories satisfy the following subsystem dynamics and initial condition respectively.

$$\frac{d}{dt}\xi_i(s; z_i, t, a(.), b(.)) = g_i(\xi_i(s; z_i, t, a(s), b(s))$$
$$\xi_i(t; z_i, t, a(.), b(.)) = z_i, \tag{32}$$

where $g_i = (z_i, a, b) = (f_i(x_i, x_3, a, b), f_3(x_3, a, b))$. The trajectories of the full system and subsystem can be connected through the projection operator as follows

$$proj_i(\zeta(s; x, t, a(.), b(.)) = \xi_i(s; z_i, t, a(.), b(.)), \tag{33}$$

where $z_i = proj_i(x)$.

5.6 Goal of doing projection.

Suppose that the target set or the unsafe set $\mathcal{T}$ of the full system could be written or expressed as the subsystem unsafe sets $\mathcal{T}_1 \subseteq \mathcal{Z}_1$ and $\mathcal{T}_2 \subseteq \mathcal{Z}_2$ in one of the following ways:

$$\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cap proj^{-1}(\mathcal{T}_2). \tag{34}$$

This means that the intersection of the back-projections of the unsafe subsystem sets preform the full unsafe set.

$$\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cup proj^{-1}(\mathcal{T}_2). \tag{35}$$

This means that the full unsafe set or the target set is the union of the back-projections of the target sets's subsystem.

Applying the intersection operator to $\mathcal{S}_1$ and $\mathcal{S}_2$ leads to black cube as shown in figure 11, however; applying the union operator results cross-shaped set including both $proj^{-1}(\mathcal{S}_1)$ and $proj^{-1}(\mathcal{S}_2)$.

Now, let's state the subsystems BRS $\mathcal{A}_i$ in the same way as defined in (4), but with the subsystems in (28) and the subsystem of the unsafe set $\mathcal{T}_i$ where $i = 1, 2$

$$\mathcal{A}_i(\tau) = \{z_i : \forall a(.), \exists b(.), \xi_i(0; z_i, t, a(.), b(.)) \in \mathcal{T}_i\}. \tag{36}$$

Thus, the main aim for a provided system in the form of (28) and unsafe set which is presented by (34) is as follow:

5.6.1 Decomposition of BRSs.

The first thing to be done is to compute the full-dimensional of BRS by doing the computations in lower-dimensional subspaces. In particular, computing the BRS of the full system of $\mathcal{A}(\tau)$ by computing first the subsystem BRSs $\mathcal{A}_i(\tau)$. This method can reduce the complexity of computation and decompose the higher-dimensional of the full system into two lower-dimensional

subsystems. Formally, previous study investigated which one of the following two cases is true:

$$
\begin{aligned}
(34) &\Rightarrow \mathcal{A}(\tau) = proj^{-1}(\mathcal{A}_1(\tau)) \cap proj^{-1}(\mathcal{A}_2(\tau)). \\
(35) &\Rightarrow \mathcal{A}(\tau) = proj^{-1}(\mathcal{A}_1(\tau)) \cup proj^{-1}(\mathcal{A}_2(\tau)).
\end{aligned}
\tag{37}
$$

As a result it is found that the first case of the BRSs is a suitable one. The reason why this case is chosen will be explained in section 6.1.

6   Self-Contained subsystems.

It is already established background and definition of self-contained subsystem and now we present the major result in our following theorem 2. This theorem connects the lower-dimensional of BRSs to the full-dimensional of BRSs which will be computed. The outcome of this theorem is that a conservative approximation of the full-dimensional BRS will be obtained for a system of the form (28) by computing the lower-dimensional BRSs $\mathcal{A}_i(\tau)$.

Now, let's state intermediate results via lemma 5 which includes a property for the projection operator before proving theorem 2.

Lemma 5.

For some subsystem i, let $\bar{x} \in \mathcal{X}, \bar{z}_i = proj_i(\bar{x}), \mathcal{S}_i \subseteq \mathcal{Z}_i$, then

$$
\bar{z}_i \in \mathcal{S}_i \Leftrightarrow \bar{x} \in proj^{-1}(\mathcal{S}_i)
\tag{38}
$$

Proof. **Forward direction.**

Assume that $\bar{z}_i \in \mathcal{S}_i$, then $\exists z_i \in \mathcal{S}_i, proj_i(\bar{x}) = z_i$ (note that z_i is equivalent to $\bar{z}_i$ itself). Also, from the definition of the back-projection which is stated in 31, then $\bar{x} \in proj^{-1}(\mathcal{S}_i)$.

Backward direction.

Assume that $\bar{x} \in proj^{-1}(\mathcal{S}_i)$, then from the back-projection's definition in 31, $\exists z_i \in \mathcal{S}_i, proj_i(\bar{x}) = z_i$.

Let's refer to $z_i \in \mathcal{S}_i$ by $\hat{z}_i$ and assume that $\bar{z}_i \notin \mathcal{S}_i$, then getting $\hat{z}_i \neq \bar{z}_i$ which derives a contradiction because $\bar{z}_i = proj_i(\bar{x}) = \hat{z}_i$. $\qquad\square$

Corollary 2.

If $\mathcal{S} = proj^{-1}(\mathcal{S}_1) \cap proj^{-1}(\mathcal{S}_2)$, then

$$
\bar{x} \in \mathcal{S} \Leftrightarrow \bar{z}_1 \in \mathcal{S}_1 \wedge \bar{z}_2 \in \mathcal{S}_2,
\tag{39}
$$

where $\bar{z}_i = proj_i^{-1}(x)$

6.1   Theorem 2: System decomposition for computing the BRS using the intersection of the lower dimensional BRSs.

Assume that we can decompose the full system in (1) into the form of (28), then

$$
\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cap proj^{-1}(\mathcal{T}_2) \Rightarrow \mathcal{A}(\tau) = proj^{-1}(\mathcal{A}_1(\tau)) \cap proj^{-1}(\mathcal{A}_2(\tau))
\tag{40}
$$

Proof. The proof of this theorem is based on the previous corollary 2 and lemma 5. First, we prove the following equivalent statement;

$$\bar{x} \notin \mathcal{A}(\tau) \Leftrightarrow \bar{x} \notin proj^{-1}(\mathcal{A}_1(\tau)) \cap proj^{-1}(\mathcal{A}_2(\tau)) \tag{41}$$

this statement is equivalent to

$$\bar{x} \in \mathcal{A}^c(\tau) \Leftrightarrow \bar{x} \in [proj^{-1}(\mathcal{A}_1(\tau))]^c \cup [proj^{-1}(\mathcal{A}_2(\tau))]^c \tag{42}$$

Taking into consideration the relationship between the full and subsystem trajectories in (33) and define

$$\bar{z}_i = proj_i(\bar{x})$$

and

$$\xi_i(0; \bar{z}_i, t, a(.), b(.)) = proj_i(\zeta(0; \bar{x}, t, a(.), b(.))).$$

Backward direction.

From the definition of Minimal BRS which is stated in (4), we have that $\bar{x} \in \mathcal{A}^c(\tau)$ which is equivalent to $\forall a(.), \exists b(.), \zeta(0; \bar{x}, t, a(.), b(.)) \in \mathcal{T}^c$. Also,

$$\bar{x} \in [proj^{-1}(\mathcal{A}_1(\tau))]^c \cup [proj^{-1}(\mathcal{A}_2(\tau))]^c$$

Where the latest is equivalent to $\bar{z}_1 \in \mathcal{A}_1^c(\tau) \vee \bar{z}_2 \in \mathcal{A}_2^c(\tau)$. WLOG and referring to the subsystem BRS definition in (36) with supposing that $\bar{z}_1 \in \mathcal{A}^c_1(\tau)$ which is equivalent to

$$\forall a(.), \exists b(.), \xi_1(0; \bar{z}_1, t, a(.), b(.)) \in \mathcal{T}_1^c. \tag{43}$$

Therefore, by lemma 5, $\bar{x} \in [proj^{-1}(\mathcal{A}_1(\tau)]^c$ which equivalently proves the backward direction.

Forward direction.

We refer to the definition in (4), and start with $\bar{x} \in \mathcal{A}^c(\tau)$ which is equivalent to

$$\forall a(.), \exists b(.), \zeta(0; \bar{x}, t, a(.).b(.)) \in \mathcal{T}^c.$$

Then, from corollary 2, we have

$$\forall a(.), \exists b(.), \xi_1(0, \bar{z}_1, t, a(.), b(.)) \in \mathcal{T}_1^c \vee \xi_2(0, \bar{z}_2, t, a(.), b(.)) \in \mathcal{T}_2^c. \tag{44}$$

Therefore, we have:

$$\mathcal{A}(\tau) \subseteq proj^{-1}(\mathcal{A}_1(\tau)) \cap proj^{-1}(\mathcal{A}_2(\tau)). \tag{45}$$

In this situation, the reconstruction of the subsystems of the BRS is a conservative approximation of the full dimensional BRS in the right direction. Consequently, in the state x is guaranteed to be able to avoid the target or unsafe set $\mathcal{T}$ of the reconstructed BRS. □

Remark.

To sum up, the previous theorem 2 gives the following results:
In the case where $\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cup proj^{-1}(\mathcal{T}_2)$, the subsystem trajectory which reaches the corresponding subsystem unsafe set implies that the full system target set is going to be reached by the trajectory of the full system. However, in the case where $\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cap proj^{-1}(\mathcal{T}_2)$, the trajectories of both subsystems must correspond the subsystem target sets at the same time. Therefore, we are going to depend on theorem 2 to evolve the projections via decomposition in our example because our goal is to avoid the target or unsafe set.

7 Numerical examples.

In this section we are going to present and explain our method. We are going to use an example of two dimensional to illustrate HJ reachability and three dimensional Dubins car. Then we will apply our decomposition method via projections onto two dimensional subsystems. This method shows that we can reconstructed the exact full-dimensional BRS from the lower-dimensional BRSs.

7.1 A simple 2D illustrating HJ reachability.

This 2D example is to explain and summarize the Hamilton Jacobi reachability. It shows the relationship between the implicit surface function BRS, value function and target set corresponding optimal control problem. Let's consider the unsafe set $\mathcal{T} \subset \mathcal{R}$ can be represented by the implicit surface function $g(x)$. Then, the unsafe set is the zero-sublevel set of the implicit surface function: $\mathcal{T} = \{x \in \mathcal{X} : g(x) \leqslant 0\}$. There exits such a function since we can select $g(.)$ to be the signed distance function from the target set. Thus, starting with set up unsafe state and define the value function which is positive outside and negative inside the set. Therefore, by Hamilton Jacobi reachability we can propagate this value function backwards in time over a desired time horizon. We can see from this figure 12 that the boundary of the target set $\mathcal{T}$ is shown as the solid black line whereas the boundary of the BRS $\mathcal{A}(\tau)$ is shown as the dashed black line. Also, the implicit surface function $g(x)$ of the target set is represented as the green surface at time zero whereas the light gray surface represents the value function $v(t, x)$ of the BRS at $t < 0$. We can infer from that the system stays safe outside of the BRS at time $t < 0$. As a result, the system is guaranteed to not enter the target set at $t = 0$.

7.2 A 3D Dubins car example in polar coordinates.

We consider the Dubins car where the system can be decomposed into SCSs where the state is $x = (x_1, x_2, x_3)$, the angular velocity is w and the partitions are x_1, x_2 and x_3:
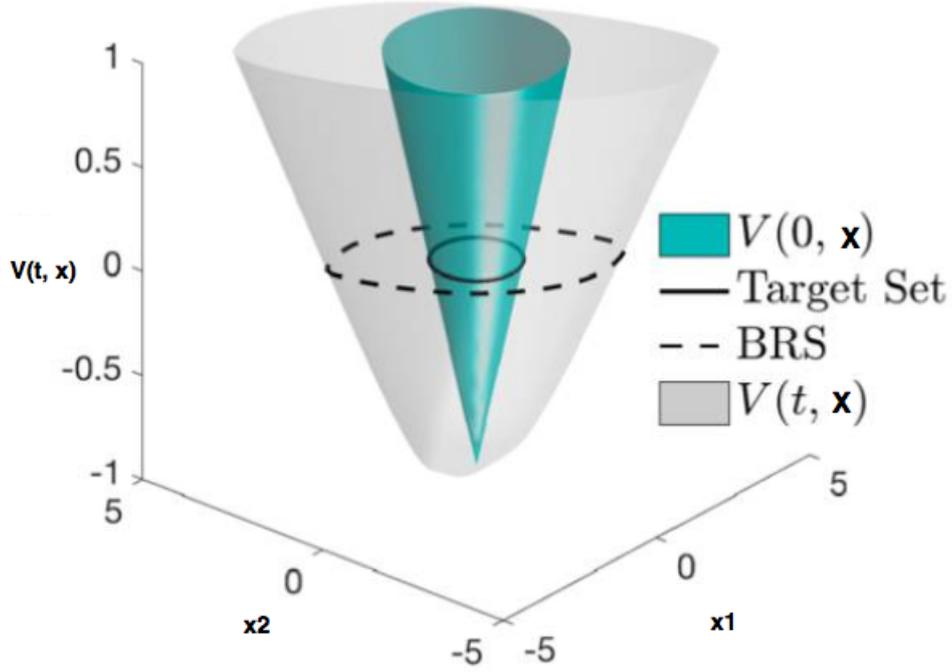
Figure 12: A 2D illustration of HJ reachability
(Chen et al., 2016a).

$$
\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} \dot{p}_x \\ \dot{p}_y \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v \cos x_3 \\ v \sin x_3 \\ w \end{bmatrix}.
\tag{46}
$$

The computation of this system is done in a 3D polar coordinate and the BRS is computed in the full-dimensional space. Therefore, we can do comparison between the full formulation with the SCS formulation. The subsystems x_i and the subsystems control and disturbance a and b are as follow:

$$
\dot{z}_1 = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} \dot{p}_x \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v \cos x_3 \\ w \end{bmatrix},
$$

$$
\dot{z}_2 = \begin{bmatrix} \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} \dot{p}_y \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v \sin x_3 \\ w \end{bmatrix},
\tag{47}
$$

where x_3 is the overlapping state. As stated in the full system, the subsystem control and disturbance function spaces $\mathbb{A}, \mathbb{B}$ and the control and disturbance signal spaces $\mathcal{A}, \mathcal{B}$ are determined suitably to the control and disturbance signal space $\mathcal{A}, \mathcal{B}$ and the function space $\mathbb{A}$ and $\mathbb{B}$ based on how they enter the dynamics of the subsystems. The evolution or development of the subsystem does not rely explicitly on each other even though if there are common overlapping states in z_1 and z_2. For example, if we disregard the subsystem z_2, then the subsystem z_1 can be considered as a full system by its self. Thus, every system can be considered as self-contained. In this example, the BRS is computed from the target set which represented as the set of positions in x_1 and x_2 dimensions. Our target set $\mathcal{T}$, which is an obstacle, is determined as $\mathcal{T} = \{(x_1, x_2, x_3) : |x_1|, |x_2| \leqslant 0.5\}$. Therefore, the vehicle must avoid that obstacle. The full

formulation of the BRS is computed from the target set $\mathcal{T}$ which is shown as the red surface in the figure below. When we need to compute the self-contained subsystem formulation of the
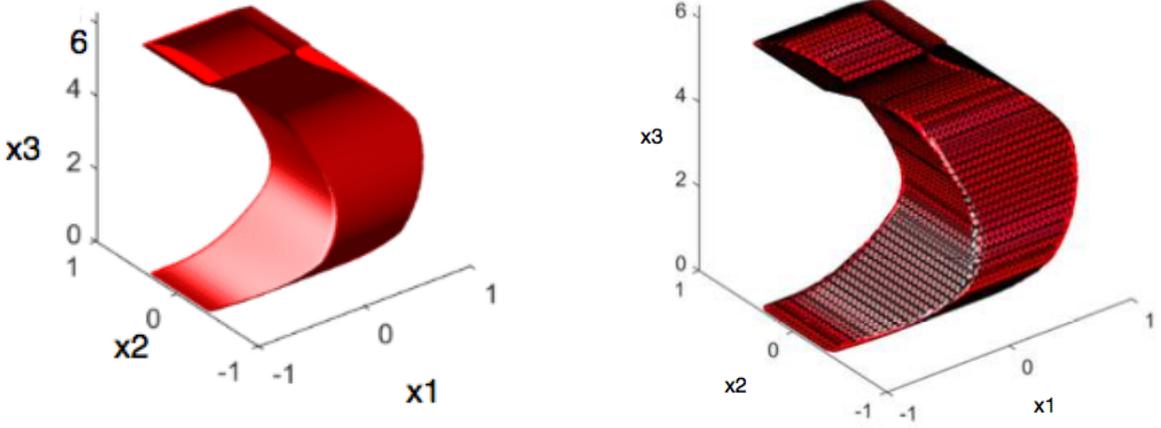
Figure 13: Computing the full formulation of the BRS when $t = -0.5$
(Chen et al., 2016a).

BRS, we need to partition the full target set $\mathcal{T}$ onto $\mathcal{T}_1$ and $\mathcal{T}_2$. Thus, we apply the theorem 2 where the system remains safe and avoid the unsafe set. Then we can write the target set as $\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cap proj^{-1}(\mathcal{T}_2)$ with

$$
\begin{aligned}
\mathcal{T}_1 &= \{(x_1, x_3) : |x_1| \leqslant 0.5\}. \\
\mathcal{T}_2 &= \{(x_2, x_3) : |x_2| \leqslant 0.5\}.
\end{aligned}
\tag{48}
$$

Now, we can compute the lower-dimensional of the BRSs $\mathcal{A}_1(\tau)$ and $\mathcal{A}_2(\tau)$ from the previous lower-dimensional target sets. Therefore, the full-dimensional of the BRS can be reconstructed from lower-dimensional BRSs $\mathcal{A}_1(\tau)$ and $\mathcal{A}_2(\tau)$ by applying theorem 2: $\mathcal{A}(\tau) = proj^{-1}\mathcal{A}_1(\tau) \cap proj^{-1}\mathcal{A}_2(\tau)$. To explain that we start to set up the unsafe set in 3D which represented as an obstacle where x_1 and x_2 positions depend on x_3. We already splitted the system into two SCSs as in (47). Then, by using the splitted subsystems, we can define the unsafe set for each subsystem. Then by back projecting these 2D subsystems into 3D. While evolving each subsystem, we can see that the full system evolves as shown in figure 14. We can see from the figure the black mesh which represents the BRS that is computed by the projection method. When we overlain the BRS of the full formulation with the BRS of the SCS, it matches exactly. The benefits of using the computation in SCS formulation that we can save time. Computing the full formulation in 3D will consume time. For instance, the computation of the the full BRS in the previous example takes 80 minutes whereas the computation in SCS via decomposition takes about 30 seconds. This computations were timed on a desktop computer with an Intel Core i7-2600K processor and 16GB of random-access memory. Also, the computation in lower-dimensional subspaces can be done very quickly. Furthermore, this method provides an accurate numerical solution.
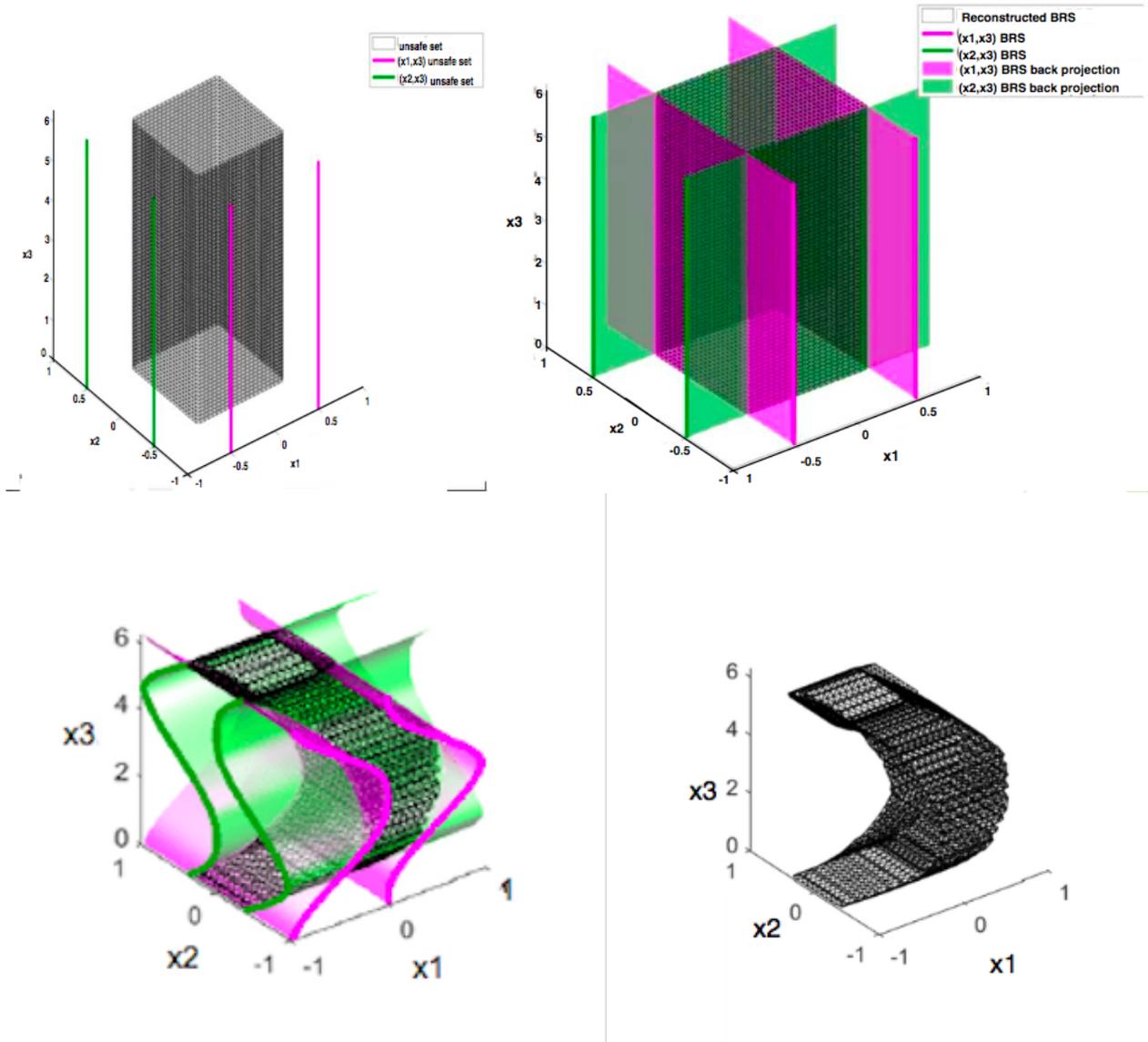
Figure 14: Set up the target set and the unsafe set for each subsystem
(Chen et al., 2016a).

8 Considering decomposition method in linear systems of ODEs.

We consider a simple first order linear problem of three ODEs in full systems as follow

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \\ \dot{x_3} \end{bmatrix} = \begin{bmatrix} ax_1 + bx_3 \\ cx_2 + dx_3 \\ ex_3 \end{bmatrix}.$$

Where a, b, c, d and e are constants, x_1, x_2 and x_3 are unknown functions $x_1(t)$, $x_2(t)$ and $x_3(t)$ of a common variable t and x_3 is a shared component. We decompose the full system into two subsystems as follow

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_3} \end{bmatrix} = \begin{bmatrix} ax_1 + bx_3 \\ ex_3 \end{bmatrix}.$$

and

$$\begin{bmatrix} \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} cx_2 + dx_3 \\ ex_3 \end{bmatrix}.$$

In matrix notation,

$$\dot{X} = AX.$$

where

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \dot{X} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix}, \qquad A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

The solutions which can solve this system look like

$$X = Ve^{\lambda t}, \qquad V = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}.$$

The eigenvalues can be found by $0 = det(A - \lambda I)$ and the corresponding eigenvectors are found by $(A - \lambda I)V = 0$ where I is the identity matrix. Then, rewrite the solution in terms of the original variables $x_1(t)$, $x_2(t)$ and $x_3(t)$

$$x_i(t) = A_1 V_1 e^{\lambda t} + A_2 V_2 e^{\lambda t} + A_3 V_3 e^{\lambda t}.$$

8.0.1 Example.

We consider the full system as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 2x_1 - 5x_3 \\ 3x_2 - 6x_3 \\ -3x_3 \end{bmatrix}.$$

The decomposed subsystems are as follows

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 2x_1 - 5x_3 \\ -3x_3 \end{bmatrix}.$$

and

$$\begin{bmatrix} \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 3x_2 - 6x_3 \\ -3x_3 \end{bmatrix}.$$

We start with the full system where

$$A = \begin{bmatrix} 2 & 0 & -5 \\ 0 & 3 & -6 \\ 0 & 0 & -3 \end{bmatrix}.$$

The eigenvalues are $\lambda_1 = 3$, $\lambda_2 = 2$ and $\lambda_3 = -3$ and the corresponding eigenvectors for λ_1, λ_2 and λ_3 are respectively

$$V_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, V_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, V_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The solution is in the form

$$X(t) = AV_1 e^{3t} + BV_2 e^{2t} + CV_3 e^{-3t}.$$

Rewrite it in terms of the original variables $x_1(t)$, $x_2(t)$ and $x_3(t)$

$$x_1(t) = Be^{2t} + Ce^{-3t}.$$
$$x_2(t) = Ae^{3t} + Ce^{-3t}.$$
$$x_3(t) = Ce^{-3t}.$$

To compute the BRS, we need to define the target set at $t = 0$, so suppose $x_1 \in [1, 2]$ and $x_2 \in [q, 3]$. The control has range $C \in [0, p]$. We introduce p and q to allow some flexibility in their final choice. Now, we need to find the targets of A and B, so that the value of $x_1(t) \in [1, 2]$ and $x_2(t) \in [q, 3]$ at $t = 0$. For the 3 dimensional system, the target set is the two dimensional rectangle $[1, 2] \times [q, 3]$. We start with $x_1(t)$ at $t = 0$, so $x_1(0) = Be^{2(0)} + Ce^{-3(0)} = B + C$. To fit the target set, B and C must satisfy $1 \leqslant B + C \leqslant 2$. Since $0 \leqslant C \leqslant p$, the restriction on B is $1 \leqslant B \leqslant 2 - p$. Thus, p must be chosen, so that $p < 1$ or there is no backward reachable set for x_1. This means that at $t = -1$,

$$x_1(-1) = Be^{2(-1)} + Ce^{-3(-1)}.$$
$$e^{-2} \leqslant x_1(-1) \leqslant (2 - p)e^{-2} + pe^3.$$

A similar calculation for $x_2(t)$ at $t = 0$ can be done so that $x_2(0) = Ae^{3(0)} + Ce^{-3(0)} = A + C$. To fit the target set, A and C must satisfy $q \leqslant A + C \leqslant 3$. Since $0 \leqslant C \leqslant p$, the restriction on A is $q \leqslant A \leqslant 3 - p$. Thus, p must be chosen, so that $3 - p > 2$. This means that at $t = -1$,

$$x_2(-1) = Ae^{3(-1)} + Ce^{-3(-1)}.$$
$$qe^{-3} \leqslant x_2(-1) \leqslant (3 - p)e^{-3} + pe^3.$$

The value of q must be less than the right side, but this is not a big restriction. It seems that $q = 2$ will be a good choice since $3 - p > 2$. Therefore, we can choose the values of $p = \frac{1}{2}$ and $q = 2$. We can infer that the backward reachable set for the 3 dimensional system (two system functions $x_1(t)$ and $x_2(t)$ and one control $x_3(t)$) is the rectangle

$$[e^{-2}, (2 - p)e^{-2} + pe^3] \times [qe^{-3}, (3 - p)e^{-3} + pe^3].$$

Therefore, the projection of this set on the first component $x_1(t)$ is the interval $[e^{-2}, (2 - p)e^{-2} + pe^3]$ and the projection of this set on the second component $x_2(t)$ is the interval $[qe^{-3}, (3 - p)e^{-3} + pe^3]$. Now, we consider the previous two subsystems for 2D, and let start with the subsystem $x_1(t)$ and $x_3(t)$ where

$$A = \begin{bmatrix} 2 & -5 \\ 0 & -3 \end{bmatrix}, \qquad X = X(t) = \begin{bmatrix} x_1(t) \\ x_3(t) \end{bmatrix}.$$

The eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = -3$ and the corresponding eigenvectors for λ_1 and λ_2 are respectively as

$$V_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, V_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Writing the solution in terms of the original variables $x_1(t)$ and $x_3(t)$

$$x_1(t) = Ae^{2t} + Be^{-3t}.$$

$$x_3(t) = Be^{-3t}.$$

Therefore,

$$\dot{x}_1 = 2x_1 - 5x_3.$$

$$\dot{x}_3 = -3x_3.$$

At $t = 0$, $x_1(0) = Ae^{2(0)} + Be^{-3(0)} = A + B$. To fit the target set, A and B must satisfy $1 \leqslant A + B \leqslant 2$. Since $0 \leqslant B \leqslant p$ is restricted by the control, the restriction on B is $1 \leqslant A \leqslant 2 - p$. Thus, p must be chosen, so that $p < 1$ or there is no backward reachable set for x_1. This means that at $t = -1$,

$$x_1(-1) = Ae^{2(-1)} + Be^{-3(-1)}.$$

$$e^{-2} \leqslant x_1(-1) \leqslant (2 - p)e^{-2} + pe^3.$$

This means that this backwards reachable set of the subsystem agrees with the one that we have previously in the 3D system. So, the projection of this component $x_1(t)$ is the interval $[e^{-2}, (2 - p)e^{-2} + pe^3]$. The same procedure can be done for the second subsystem $x_2(t)$ and $x_3(t)$ so that

$$A = \begin{bmatrix} 3 & -6 \\ 0 & -3 \end{bmatrix}, \qquad X = X(t) = \begin{bmatrix} x_2(t) \\ x_3(t) \end{bmatrix}.$$

The eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = -3$ and the corresponding eigenvectors for λ_1 and λ_2 are respectively as

$$V_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, V_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Writing the solution in terms of the original variables $x_1(t)$ and $x_3(t)$

$$x_2(t) = Ae^{3t} + Be^{-3t}.$$

$$x_3(t) = Be^{-3t}.$$

Therefore,

$$\dot{x}_2 = 3x_2 - 6x_3.$$

$$\dot{x}_3 = -3x_3.$$

At $t = 0$, $x_2(0) = Ae^{3(0)} + Be^{-3(0)} = A + B$. To fit the target set, A and B must satisfy $q \leqslant A + B \leqslant 3$. Since $0 \leqslant B \leqslant p$ is restricted by the control, the restriction on A is $q \leqslant A \leqslant 3 - p$. Thus, p must be chosen, so that $3 - p > 2$. This means that at $t = -1$,

$$x_2(-1) = Ae^{3(-1)} + Be^{-3(-1)}.$$

$$qe^{-3} \leqslant x_2(-1) \leqslant (3 - p)e^{-3} + pe^3.$$

This means that this backwards reachable set of this subsystem agrees with the one that we have previously in the 3D system. So, the projection of this component $x_2(t)$ is the interval $[qe^{-3}, (3 - p)e^{-3} + pe^3]$. As a result, computing the BRS in lower-dimensional subsystem and then combining them lead to construct the full-dimensional BRS of the full system.

9 A 3D Dubins car example in spherical coordinates.

This example will consider the following 3D variants of the Dubins model

$$
\begin{bmatrix} \dot{p}_x \\ \dot{p}_y \\ \dot{p}_z \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v \sin\theta \cos\phi \\ v \sin\theta \sin\phi \\ v \cos\theta \\ w \end{bmatrix}. \tag{49}
$$

This is a three-dimensional generalization of the Dubins model in equation (46); with a spherical parametization which describes the motion in xyz directions and a constant control w. We consider a two-dimensional Dubins car but with an additional independent bounded constraint over the speed in the z-direction, then the model becomes as follows

$$
\begin{bmatrix} \dot{p}_x \\ \dot{p}_y \\ \dot{p}_z \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v \cos\theta \\ v \sin\theta \\ cv \\ w \end{bmatrix}. \tag{50}
$$

Where $c > 0$. Also, let consider the unmanned aerial vehicle (UAV) in the Dubins Airplane with the autopilot is well tuned and the airspeed, flight-path angle and bank angle state converges with the desired response, then the kinematics is given by

$$
\begin{bmatrix} \dot{p}_x \\ \dot{p}_y \\ \dot{p}_z \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v \cos\theta \cos\phi \\ v \sin\theta \cos\phi \\ -v \sin\phi \\ \frac{g}{v} \tan\psi \end{bmatrix}. \tag{51}
$$

where g is acceleration due to gravity, ψ is the heading angle and ϕ is the bank angle.
Let's decompose each of the above system into various subsystems (self-contained subsystems) in such a way that each subsystem depends only on the subsystem states. Then, the SCSs are given as follows

$$
\dot{z}_1 = \begin{bmatrix} \dot{p}_x \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v \sin\theta \cos\phi \\ w \end{bmatrix}.
$$

$$
\dot{z}_2 = \begin{bmatrix} \dot{p}_y \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v \sin\theta \sin\phi \\ w \end{bmatrix}.
$$

$$
\dot{z}_3 = \begin{bmatrix} \dot{p}_z \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v \cos\theta \\ w \end{bmatrix}.
$$

Where $w = u$. Also, the target set $\mathcal{T}$ can be decomposed as follow:

$$
\mathcal{T}_1 = \{(p_x, \theta) : |p_x| \leqslant \alpha\}.
$$
$$
\mathcal{T}_2 = \{(p_y, \theta) : |p_y| \leqslant \alpha\}.
$$
$$
\mathcal{T}_3 = \{(p_z, \theta) : |p_z| \leqslant \alpha\}.
$$

Where α is a scalar. Then

$$
\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cap proj^{-1}(\mathcal{T}_2) \cap proj^{-1}(\mathcal{T}_3).
$$

Therefore, the SCSs for this system for the Dubins' plane is given as follow:

$$\dot{z}_1 = \begin{bmatrix} \dot{p}_x \\ \dot{p}_z \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v\cos\theta \\ cv \\ w \end{bmatrix}.$$

$$\dot{z}_2 = \begin{bmatrix} \dot{p}_y \\ \dot{p}_z \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v\sin\theta \\ cv \\ w \end{bmatrix}.$$

Thus, the target set $\mathcal{T}$ can be computed by

$$\mathcal{T}_1 = \{(p_x, p_z, \theta) : |p_x|, |p_z| < \beta\}.$$
$$\mathcal{T}_2 = \{(p_y, p_z, \theta) : |p_y|, |p_z| < \beta\}.$$

Such that

$$\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cap proj^{-1}(\mathcal{T}_2).$$

Also, the SCSs for the Dubins' Airplane can be written as

$$\dot{z}_1 = \begin{bmatrix} \dot{p}_x \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v\cos\theta\cos\phi \\ \frac{g}{v}\tan\psi \end{bmatrix}.$$

$$\dot{z}_2 = \begin{bmatrix} \dot{p}_y \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v\sin\theta\sin\phi \\ \frac{g}{v}\tan\psi \end{bmatrix}.$$

$$\dot{z}_3 = \begin{bmatrix} \dot{p}_z \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} -v\sin\phi \\ \frac{g}{v}\tan\psi \end{bmatrix}.$$

Now, the target set $\mathcal{T}$ is

$$\mathcal{T}_1 = \{(p_x, \theta) : |p_x| < \gamma\}.$$
$$\mathcal{T}_2 = \{(p_y, \theta) : |p_y| < \gamma\}.$$
$$\mathcal{T}_3 = \{(p_z, \theta) : |p_z| < \gamma\}.$$

And

$$\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cap proj^{-1}(\mathcal{T}_2) \cap proj^{-1}(\mathcal{T}_3).$$

Now, the aim is to compute the BRS for the Dubins car in 3D from the target set. It represents the positions near the origin in the p_x, p_y and p_z directions. We take the target set as

$$\mathcal{T} = \{p_x, p_y, p_z, \theta : |p_x| \leqslant 0.5, |p_y| \leqslant 0.5, |p_z| \leqslant 0.5\}.$$

Using the decomposition method, we can write the target set as

$$\mathcal{T}_1 = \{(p_x, \theta) : |p_x| < 0.5\}.$$
$$\mathcal{T}_2 = \{(p_y, \theta) : |p_y| < 0.5\}.$$
$$\mathcal{T}_3 = \{(p_z, \theta) : |p_z| < 0.5\}.$$

And

$$\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cap proj^{-1}(\mathcal{T}_2) \cap proj^{-1}(\mathcal{T}_3).$$

In the same way, we can compute the BRS for the Dubins' plane by taking $\beta = 1$, then we have

$$\mathcal{T}_1 = \{(p_x, p_z, \theta) : |p_x|, |p_z| < 1\}.$$
$$\mathcal{T}_2 = \{(p_y, p_z, \theta) : |p_y|, |p_z| < 1\}.$$

Such that

$$\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cap proj^{-1}(\mathcal{T}_2).$$

Also, for the Dubins' airplane, we take $\gamma = 1$ such that we can compute the BRS by

$$\mathcal{T}_1 = \{(p_x, \theta) : |p_x| < 1\}.$$
$$\mathcal{T}_2 = \{(p_y, \theta) : |p_y| < 1\}.$$
$$\mathcal{T}_3 = \{(p_z, \theta) : |p_z| < 1\}.$$

And

$$\mathcal{T} = proj^{-1}(\mathcal{T}_1) \cap proj^{-1}(\mathcal{T}_2) \cap proj^{-1}(\mathcal{T}_3).$$

For future work, we will continue working on this example and looking for some hypothesis to help us prove that.

10 Conclusion and future work.

In this report, we have applied two methods to solve our problem. The first method is used to compute the backwards reachable set with target set and it is based on formulating the reachability problem in terms of the viscosity solution of a time-dependent Hamilton-Jacobi-Isaacs partial differential equation. We have set a theorem which proves that the analytic solution of this equation gives an accurate reachable set as it is determined in the previous definition. The solution of time dependent HJ PDE is continuous and can be determined over the state space. As a result, the computations are less expensive. Also, the implementation is done in a three-dimensional pursuit evasion example. There is a primary weakness in the formulation of the reachable sets when the dimension of the system increases, which is the exponential growth and computational expansive. There is a way to reduce the costs of the computations which is to project the reachable set from a higher dimensional system into a group of lower dimensional subsystems. The second method is to do projections of the full system via decomposition and SCSs and solve the problem with a collection of lower dimensional. This method is combining the first method as well where combining different methods gives better results. The major goal of doing the projection is to obtain the backwards reachable sets. The self-contained subsystems formulation is used to compute that backwards reachable sets, and it reduces the burden of the computations. This means that the previous computations which are considered to be intractable would be possible to be done. Therefore, we can construct the full dimensional backwards reachable set from the lower dimensional backwards reachable set as we have seen in the example of linear systems. In this example that we have investigated we have shown the backward reachable set of the subsystems are exactly the same as the backward reachable set in the full system.

Most of this thesis has assimilated existing work on backwards reachability. The original contribution is the detailed example in Chapter 9.

In future work, looking to continue working in the example of spherical coordinates and to extend and apply the decomposition method in the backwards reachable trajectories BRTs.

References

Allen, R. E., Clark, A. A., Starek, J. A., & Pavone, M. (2014). A machine learning approach for real-time reachability analysis. In *Intelligent robots and systems (iros 2014), 2014 ieee/rsj international conference on* (pp. 2202–2208).

Ash, R. B. (2014). *Real analysis and probability: Probability and mathematical statistics: A series of monographs and textbooks.* Academic press.

Bardi, M. (1997). Some applications of viscosity solutions to optimal control and differential games. *Viscosity Solutions and Applications*, 44–97.

Bayen, A., Santhanam, S., Mitchell, I., & Tomlin, C. (2003). A differential game formulation of alert levels in etms data for high altitude traffic. In *Aiaa guidance, navigation, and control conference and exhibit* (p. 5341).

Bayen, A. M., Mitchell, I. M., Oishi, M. M., & Tomlin, C. J. (2007). Aircraft autolander safety analysis through optimal control-based reach set computation. *Journal of Guidance, Control, and Dynamics*, *30*(1), 68.

Bokanowski, O., & Zidani, H. (2011). Minimal time problems with moving targets and obstacles. *IFAC Proceedings Volumes*, *44*(1), 2589–2593.

Boyce, W. E. W. E., & DiPrima, R. C. (1986). *Elementary differential equations and boundary value problems* (Tech. Rep.).

Chae, H., & Kim, H. (2014). The validity checking on the exchange of integral and limit in the solving process of pdes. *Int. J. of Math. Anal*, *8*, 1089–1092.

Chen, M., Herbert, S., & Tomlin, C. J. (2016a). Exact and efficient hamilton-jacobi-based guaranteed safety analysis via system decomposition. *arXiv preprint arXiv:1609.05248*.

Chen, M., Herbert, S., & Tomlin, C. J. (2016b). Fast reachable set approximations via state decoupling disturbances. In *Decision and control (cdc), 2016 ieee 55th conference on* (pp. 191–196).

Chen, M., Herbert, S. L., Vashishtha, M. S., Bansal, S., & Tomlin, C. J. (n.d.). Decomposition of reachable sets and tubes for a class of nonlinear systems.

Coddington, E. A., & Levinson, N. (1955). *Theory of ordinary differential equations.* Tata McGraw-Hill Education.

Crandall, M. G., Evans, L. C., & Lions, P.-L. (1984). Some properties of viscosity solutions of hamilton-jacobi equations. *Transactions of the American Mathematical Society*, *282*(2), 487–502.

Crandall, M. G., & Lions, P.-L. (1984). Two approximations of solutions of hamilton-jacobi equations. *Mathematics of Computation*, *43*(167), 1–19.

Evans, L. C., & Souganidis, P. E. (1983). *Differential games and representation formulas for solutions of hamilton-jacobi-isaacs equations.* (Tech. Rep.). WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER.

Hoffmann, G. M., Huang, H., Waslander, S. L., & Tomlin, C. J. (2007). Quadrotor helicopter flight dynamics and control: Theory and experiment. In *Proc. of the aiaa guidance, navigation, and control conference* (Vol. 2, p. 4).

Kong, J., Pfeiffer, M., Schildbach, G., & Borrelli, F. (2015). Kinematic and dynamic vehicle models for autonomous driving control design. In *Intelligent vehicles symposium (iv), 2015 ieee* (pp. 1094–1099).

Lygeros, J., Tomlin, C., & Sastry, S. (1999). Controllers for reachability specifications for hybrid systems. *Automatica*, *35*(3), 349–370.

Margellos, K., & Lygeros, J. (2011). Hamilton–jacobi formulation for reach–avoid differential games. *IEEE Transactions on Automatic Control*, *56*(8), 1849–1861.

Mitchell, I. (2004). Demonstrating numerical convergence to the analytic solution of some backwards reachable sets with sharp features. *Dept. Comput. Sci., Univ. British Columbia,*

    Vancouver, BC, Canada, Tech. Rep. TR-2004-01.

Mitchell, I., Bayen, A., & Tomlin, C. J. (2004). Computing reachable sets for continuous dynamic games using level set methods. *Submitted January*.

Mitchell, I., Bayen, A. M., & Tomlin, C. J. (2001). Validating a hamilton-jacobi approximation to hybrid system reachable sets. In *Hscc* (pp. 418–432).

Mitchell, I., & Tomlin, C. (2000). Level set methods for computation in hybrid systems. In *Hscc* (Vol. 1790, pp. 310–323).

Mitchell, I. M. (2007). Comparing forward and backward reachability as tools for safety analysis. In *Hscc* (Vol. 4416, pp. 428–443).

Mitchell, I. M., Bayen, A. M., & Tomlin, C. J. (2005). A time-dependent hamilton-jacobi formulation of reachable sets for continuous dynamic games. *IEEE Transactions on automatic control*, *50*(7), 947–957.

Mitchell, I. M., & Tomlin, C. J. (2003). Overapproximating reachable sets by hamilton-jacobi projections. *journal of Scientific Computing*, *19*(1), 323–346.

Tomlin, C., Mitchell, I., & Ghosh, R. (2001). Safety verification of conflict resolution manoeuvres. *IEEE Transactions on Intelligent Transportation Systems*, *2*(2), 110–120.

Tomlin, C. J., Lygeros, J., & Sastry, S. S. (2000). A game theoretic approach to controller design for hybrid systems. *Proceedings of the IEEE*, *88*(7), 949–970.