Department of Mechanical Engineering

Cybersecurity analysis and robust controller synthesis against false data injections in cooperative driving

Graduation Project (45 EC)

Student:	*ID-number:*	
T.J.A. van Oorschot	1352725	
TU/e supervisors:	*Department*	*Research group:*
prof. dr. ir. E. Lefeber	Mechanical Engineering	Dynamics & Control
ir. M. Huisman	Mechanical Engineering	Dynamics & Control
Graduation committee:	*Department*	*Research group:*
prof. dr. ir. E. Lefeber	Mechanical Engineering	Dynamics & Control
prof. dr. ir. M.S.T. Chong	Mechanical Engineering	Dynamics & Control
prof. dr. ir. M.R.U. Salazar	Mechanical Engineering	Control Systems Technology
prof. dr. ir. C.G. Murguia	Mechanical Engineering	Dynamics & Control

This report was made in accordance with the TU/e Code of Scientific Conduct for the Master thesis

Eindhoven, June 19, 2025

Abstract

This research addresses the cybersecurity of a cooperative adaptive cruise control (CACC) algorithm in vehicle platooning. We consider a specific implementation of CACC in a platoon that is attacked by disturbing its measured and communicated signals with false data injections, and focus on the algorithm's resilience to these attacks, rather than preventing the attacks. Previous work has established that there exist infinitely many realizations of the CACC. Each realization corresponds to using a different combination of sensors, such that each realization exhibits the same platooning behaviour in the absence of attacks. However, in the presence of attacks, the robustness of these realizations varies significantly. In the context of cybersecurity, robustness is measured by the *reachable set*, which is the set of states attackers can reach by injecting bounded false data into the CACC. Since computation of the reachable set is often intractable, previous work has considered the ellipsoidal outer approximation of the reachable set instead. Though this approximation was suitable for two-vehicle configurations, the propagation of attacks through larger platoons was not evaluated.

In this study, we therefore extend the analysis to multi-vehicle platoons. We highlight the limitations of using an ellipsoid to outer approximate the reachable set, and we propose to use the *hyperrectangular outer approximation* instead. We develop a dynamical model that describes the motion of the complete vehicle platoon. Using this model, we reveal that false data injections cannot drive the state of the platoon anywhere, but that this state is constrained to the so-called *attackable subspace*. We then formulate safety guarantees of the propagation of the reachable set of vehicles along the platoon, by evaluating the hyperrectangular outer approximation of the platoon's reachable set. We formalize these guarantees by introducing a new notion of string stability, which we call $\mathcal{L}_\infty$-q *string stability*, and show that the platoon satisfies this notion. Finally, we propose a *controller synthesis* framework aimed at selecting the optimal controller realization of the CACC scheme that minimizes the platoon's vulnerability to false data injections, by minimizing its reachable set.

Contents

1 Introduction

In recent years, traffic congestion has become a growing concern. Both the frequency and duration of traffic jams are steadily rising, causing an increase in travel times [1]. Moreover, highway congestion near urban areas significantly contributes to daily pollution, increases fuel consumption, and leads to substantial time loss for commuters. In the United States, commuters experience an average annual delay of 52 hours, resulting in an estimated 121 billion in combined costs due to lost time and fuel consumption [2]. A promising way to improve highway capacity that does not require physical expansion of existing infrastructure is to reduce the distance between vehicles without sacrificing speed. However, maintaining such small gaps is unsafe with human drivers alone [3], and thus requires automation in longitudinal control.

Cooperative Adaptive Cruise Control (CACC) enables automated longitudinal vehicle-following by regulating spacing between subsequent vehicles to a desired value [4], creating a so-called vehicle platoon. This vehicle-following is achieved through a combination of onboard measurements (e.g., radar, lidar) and Vehicle-to-Vehicle (V2V) wireless communication [5, 6]. CACC is known to allow for time gaps between vehicles of less than 1 second, which is the standardized minimum value for currently available adaptive cruise control (ACC) systems [7]. Benefits of CACC compared to human drivers include improved traffic flow [8], reduced fuel consumption—especially for heavy-duty vehicles—due to lower aerodynamic drag [6, 9], and increased road capacity. The design of CACC controllers has been widely studied; multiple schemes have been proposed and evaluated, see, e.g., [10–16]. The work in this report builds on the scheme introduced by Ploeg et al. [17].

Due to their reliance on wireless communication, CACC systems are vulnerable to cyberattacks that may degrade safety or performance [18–20]. Attackers can exploit vulnerabilities in the wireless network to inject false data into sensor signals to disrupt coordination and compromise safety by ,e.g., causing collisions. Although various detection and mitigation strategies have been developed [21, 22], these strategies often face limitations due to process and measurement disturbances [23], constrained resources (e.g., limited computing power) [24], or attacker knowledge of the system [24, 25]. This motivates the need for inherently robust control approaches.

Rather than focusing on prevention of cyberattacks, this work aims to mitigate the impact of attacks and ensure system safety under adversarial conditions. In literature, two general approaches exist: *active* methods and *passive methods*. Active methods switch control modes upon detecting an attack (e.g., from CACC to ACC [26]). Passive methods, on the other hand, are designed to tolerate disturbances without switching. While active methods rely on timely and reliable detection, passive methods typically sacrifice nominal (i.e., attack-free) performance for robustness.

A recent passive approach by Huisman et al. [27, 28] enhances robustness without compromising nominal behavior. By reformulating the dynamics of the CACC controller, this controller admits multiple equivalent realizations. Each controller realization corresponds to using a different combination of sensors. The terms *controller realization* and *sensor configuration* (that is, the combination of sensors that is used in the CACC scheme) are thus equivalent in this context, and therefore used interchangeably in the remainder of this thesis. Under normal (i.e., attack-free) conditions, the vehicle platoon behaves identically under each controller realization. However, each controller realization yields different behaviour when the platoon is attacked. The equivalent attack-free behaviour and varying sensitivity to attacks of the different controller realizations allow for selecting the most robust sensor configuration, without affecting nominal performance.

In order to quantify robustness to adversarial attacks, prior work employs reachability analysis. By imposing bounds on the magnitude of the false data injections, robustness is measured by the set of states that the system can reach by injecting false data. This set of states is called the *reachable set*, and is used to provide insight into the potential damage that attacks can induce, e.g., cause a collision. Since exact computation of the reachable set is often intractable, ellipsoidal outer approximations are used. Since this ellipsoid is an *outer* approximation, any safety guarantees of the approximation also apply for the true reachable set. This enables the formulation of a synthesis framework that finds the controller realization (i.e., sensor configuration) that minimizes the approximated reachable set. However, this prior work focuses only on two-vehicle configurations and does not analyze how the effects of an attack propagate through larger platoons.

This thesis addresses these limitations by extending the analysis to multi-vehicle platoons. It considers the case where only the sensor signals of the first follower vehicle are attacked, and investigates how these attacks propagate through the platoon. First, the conventional *ellipsoidal* outer approximation is elaborated and its limitations are highlighted. Consequently, a *hyperrectangular* (box-shaped) outer approximation of the reachable set is proposed as an alternative for the ellipsoidal outer approximation. This hyperrectangular outer approximation offers greater interpretability and generalizability. A method is formulated that finds this hyperrectangular outer approximation of a system's reachable set.

Next, the dynamics that describe the motion of the complete platoon are formulated. This formulation is then evaluated to identify in which *directions* each vehicle's state can be steered by injecting false data into the sensor signals. Analysis reveals that the platoon cannot be driven in the complete state space, but that attack-induced trajectories are constrained to the so-called *attackable subspace*, regardless of the attack magnitude.

Then, the hyperrectangular outer approximation of the multi-vehicle platoon (i.e., more than two vehicles) is computed. This hyperrectangular outer approximation is used to formulate safety guarantees for the complete platoon, by introducing the notion of $\mathcal{L}_\infty$- q *string stability.* The platoon is said to be $\mathcal{L}_\infty$- q string stable when the outer approximated reachable sets of all subsequent vehicles downstream along the platoon do not grow, starting from at most the q^{th} follower vehicle. Furthermore, the reachable sets of all vehicles downstream from vehicle q are contained by the union of the reachable set of the first q vehicles. Evaluation of the hyperrectangular approximation reveals that the platoon under consideration is guaranteed to be $\mathcal{L}_\infty$- q string stable with $q \leq 3$. This string stability is important since it implies that; if safety is guaranteed for the first three follower vehicles, this safety is also guaranteed for all vehicles downstream.

Finally, a *controller synthesis framework* is formulated to find a sensor configuration that minimizes the platoon's susceptibility to attacks. Since the platoon is guaranteed to be $\mathcal{L}_\infty$- q string stable with $q \leq 3$, only the reachable sets of the first three follower vehicles are considered. The optimal sensor configuration is defined as the configuration that minimizes the union of the reachable sets of these first three vehicles, which is found by solving a convex optimization problem.

In summary, this thesis addresses the limitations of current understanding of cybersecurity in cooperative driving by extending the analysis from two-vehicle platoons to *multi-vehicle* configurations. The main difference in approach from prior work lies in using a *hyperrectangular approximation* technique, instead of the conventional ellipsoidal method, which enables the analysis of these multi-vehicle platoons. Although the research is performed in the context of vehicle platooning, some results of this work are extended to more general systems where possible.

The remainder of this report is structured as follows: Chapter 2 introduces theoretical preliminaries and notation. Chapter 3 describes the problem setup and research objectives. Chapter 4 reviews related literature. Chapter 5 describes the method to outer approximate the reachable set with a hyperrectangle. Chapter 6 models the dynamics of the full vehicle platoon. Chapter 7 introduces and analyzes the attackable subspace. Chapter 8 defines the notion of $\mathcal{L}_\infty$-q string stability and evaluates this notion for the platoon. Chapter 9 presents the controller synthesis method. Finally, Chapter 10 summarizes the conclusions and outlines directions for future research.

2 Preliminaries and notation

This chapter introduces the theoretical concepts and notation used throughout the report. First, some signal and system norms are elaborated, followed by relevant elements from set theory and convex analysis. Some theoretical concepts are adapted to the specific context of this problem to avoid unnecessary complexity.

Signal norms

Let $u(t)$ be a time-dependent vector signal $u(t)^\top = [u_1(t),\, u_2(t),\, \dots,\, u_n(t)]$. A signal norm is a real scalar number which gives an overall measure of the signal $u(t)$. The signal p-norm (or $\mathcal{L}_p$-norm) of signal $u(t)$ is defined as

$$\|u(t)\|_{\mathcal{L}_p} := \left(\int_{-\infty}^{\infty} \sum_{i=1}^{n} |u_i(t)|^p \, dt \right)^{\frac{1}{p}} . \tag{2.1}$$

Signals with bounded $\mathcal{L}_p$-norm belong to the Lebesgue space $\mathcal{L}_p$. An important signal norm in the context of safety guarantees is the $\mathcal{L}_\infty$-norm

$$\|u(t)\|_{\mathcal{L}_\infty} := \lim_{p\to\infty} \left(\int_{-\infty}^{\infty} \sum_{i=1}^{n} |u_i(t)|^p \, dt \right)^{\frac{1}{p}} = \max_i \sup_t |u_i(t)| \tag{2.2}$$

where sup (from supremum) denotes the least upper bound, which is equal to the maximum value. The $\mathcal{L}_\infty$-norm thus comes down to taking the "channel" of the signal $u(t)$ with the highest peak value over time such that $||u(t)||_{\mathcal{L}_\infty} = \max_i ||u_i(t)||_{\mathcal{L}_\infty}$.

System norms

Now let $G(s)$ be the transfer function matrix of a linear time-invariant continuous-time system with input signal $u(t)$ and output signal $y(t)$ of appropriate dimensions, and $g(t)$ the corresponding impulse response matrix of $G(s)$. The $\mathcal{L}_1$ norm of this impulse response $g(t)$ can be regarded as being induced by the $\mathcal{L}_\infty$-norms of input $u(t)$ and output $y(t)$ such that

$$||g(t)||_{\mathcal{L}_1} = \max_{u\neq 0} \frac{||y(t)||_{\mathcal{L}_\infty}}{||u(t)||_{\mathcal{L}_\infty}} . \tag{2.3}$$

From this, the output signal $y(t)$ can be bounded for any input signal $u(t)$ such that

$$||y(t)||_{\mathcal{L}_\infty} \leq ||g(t)||_{\mathcal{L}_1} \, ||u(t)||_{\mathcal{L}_\infty} . \tag{2.4}$$

The notation in (2.3) implies that this is a tight inequality, in the sense that there always exists an input signal $u(t)$ for which equality holds. The $\mathcal{L}_1$-norm of the convolution of two signals $u_1(t)$ and $u_2(t)$ can be bounded according to Young's inequality for convolutions

$$\|u_1(t) * u_2(t)\|_{\mathcal{L}_1} \leq \|u_1(t)\|_{\mathcal{L}_1} \, \|u_2(t)\|_{\mathcal{L}_1} . \tag{2.5}$$

Set theory

Let the set $\mathcal{S}$ be a collection of n-dimensional position vectors s_i. Let these vectors s_i be expressed in terms of the cartesian coordinates $x^\top = [x_1,\, \dots,\, x_n]$ with $x_i \in \mathbb{R}$. The set $\mathcal{S}$ is *point symmetric* if $s \in \mathcal{S} \Rightarrow -s \in \mathcal{S}$. A *hyperrectangular* set is a rectangular (box-shaped) set generalized to any dimension n. The Minkowski sum of two sets of position vectors $\mathcal{S}_1$ and $\mathcal{S}_2$, denoted by $\oplus$, is formed by adding each vector in $\mathcal{S}_1$ to each vector in $\mathcal{S}_2$ such that

$$\mathcal{S}_1 \oplus \mathcal{S}_2 = \left\{ \, s_1 + s_2 \, \middle| \, s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2 \, \right\} . \tag{2.6}$$

An outer approximation $\mathcal{A}(\mathcal{S})$ of the set $\mathcal{S}$ is any set that contains $\mathcal{S}$ such that $\mathcal{S} \subseteq \mathcal{A}$. The projection of set $\mathcal{S}$ onto the coordinates $\tilde{x} = [x_k,\, \dots,\, x_l]$ (or more precisely, onto the subspace that is spanned by these coordinate axes) is denoted by

$$\Pi_{\tilde{x}}[\mathcal{S}]. \tag{2.7}$$

We say that an approximation $\mathcal{A}(\mathcal{S})$ is *tight* around set $\mathcal{S}$ if the projections onto *each individual coordinate* $x_i \in \mathbb{R}$ of the original set $\mathcal{S}$ and its approximation $\mathcal{A}(\mathcal{S})$ are equivalent. More formally,

$$\mathcal{A}(\mathcal{S}) \text{ is tight around } \mathcal{S} \quad \Leftrightarrow \Pi_{x_i}[\mathcal{A}(\mathcal{S})] = \Pi_{x_i}[\mathcal{S}] \text{ for all } i = 1, 2, \dots, n. \tag{2.8}$$

This means that *whether or not an approximation is tight depends on the choice of coordinate frame* in which the vectors are expressed. If an approximation $\mathcal{A}(\mathcal{S})$ is hyperrectangular, centered in the origin, aligned with the coordinate frame axes, and tight around the set $\mathcal{S}$, then the set $\mathcal{S}$ touches each boundary (i.e., face) of the approximation $\mathcal{A}(\mathcal{S})$.

To clarify the concept of tightness of an approximation, consider the example presented below. The figure depicts a set $\mathcal{S}$ along with its (hyperrectangular) outer approximation $\mathcal{A}(\mathcal{S})$, relative to two coordinate frames: the original frame $x = (x_1, x_2)$ and a rotated frame $x' = (x'_1, x'_2)$. In the coordinate frame x, the projection of $\mathcal{S}$ onto each axis is equivalent to that of $\mathcal{A}(\mathcal{S})$. That is, the projections of $\mathcal{S}$ and its approximation $\mathcal{A}(\mathcal{S})$ are identical along both x_1 and x_2. This indicates that $\mathcal{A}(\mathcal{S})$ is a tight outer approximation of $\mathcal{S}$ *in terms of the x-coordinate frame*. However, *in the rotated frame x'*, the projections no longer align: the projection of $\mathcal{A}(\mathcal{S})$ onto x'_1 is wider than that of $\mathcal{S}$, and the same holds for x'_2. This means that $\mathcal{A}(\mathcal{S})$ does not tightly enclose $\mathcal{S}$ when viewed in the x' frame. This example illustrates that the tightness of an outer approximation is not an absolute property—it depends on the choice of coordinate frame.

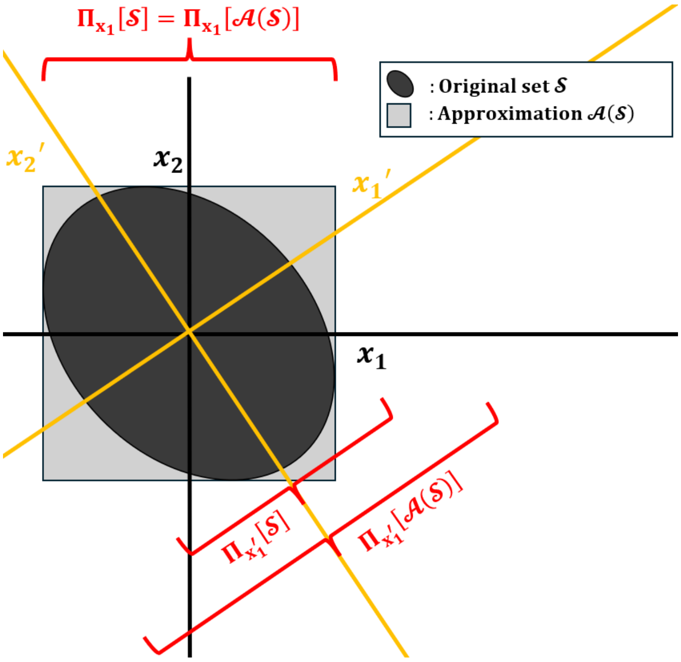
$\Pi_{x_1}[\mathcal{S}] = \Pi_{x_1}[\mathcal{A}(\mathcal{S})]$
● : Original set $\mathcal{S}$
□ : Approximation $\mathcal{A}(\mathcal{S})$
x_2'
x_2
x_1'
x_1
$\Pi_{x_1'}[\mathcal{S}]$
$\Pi_{x_1'}[\mathcal{A}(\mathcal{S})]$

Figure 2.1: Illustration of the tightness of a hyperrectangular outer approximation. In the coordinate frame $x = (x_1, x_2)$, the hyperrectangle $\mathcal{A}(\mathcal{S})$ tightly encloses the set $\mathcal{S}$. In contrast, in the rotated frame $x' = (x'_1, x'_2)$, the approximation is no longer tight. This highlights that tightness is frame-dependent.

Convexity

The set $\mathcal{S}$ is said to be convex if for any two points $s_1, s_2 \in \mathcal{S}$ the line segment connecting s_1 and s_2 is entirely contained in $\mathcal{S}$. More formally

$$\mathcal{S} \text{ is convex} \quad \Leftrightarrow \quad \forall s_1, s_2 \in \mathcal{S},\, \forall \lambda \in [0,1],\, \lambda s_1 + (1-\lambda) s_2 \in \mathcal{S}. \tag{2.9}$$

The convex hull of a set $\mathcal{S}$ is the smallest convex set that contains it, denoted by $\operatorname{conv}(\mathcal{S})$. If set $\mathcal{S}$ is already convex, then its convex hull is the set itself such that $\operatorname{conv}(\mathcal{S}) = \mathcal{S}$. The function $f(x) : \mathbb{R}^n \to \mathbb{R}$ is convex if for any two points $x_1, x_2 \in \mathbb{R}^n$ the function lies below the straight line segment connecting the points $(x_1, f(x_1))$ and $(x_2, f(x_2))$. More formally,

$$f(x) \text{ is convex} \quad \Leftrightarrow \quad \forall x_1, x_2 \in \mathcal{S},\, \forall \lambda \in [0,1],\, f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2). \tag{2.10}$$

Taking the maximum of multiple convex functions, as well as taking the conic combination of convex functions, also results in a convex function [29]. More formally,

$$\begin{aligned} f_1(x), f_2(x) \text{ are convex} \quad &\Leftrightarrow \quad \max\{f_1(x), f_2(x)\} \text{ is convex.} \\ f_1(x), f_2(x) \text{ are convex} \quad &\Leftrightarrow \quad p_1\, f_1(x) + p_2\, f_2(x) \text{ with } p_1, p_2 > 0 \text{ is convex.} \end{aligned} \tag{2.11}$$

Finally, the vector $\mathbf{e}_i$ denotes the standard column vector with a 1 in the i^{th} entry and zeros elsewhere. In summary, this chapter has introduced the required theoretical foundations and notations used in this thesis. These concepts included signals and system norms, followed by relevant elements from set theory and convex analysis. In the next chapter, the problem statement and research objectives of this thesis are formulated.

3 Problem statement

This chapter formulates the problem statement of this research. First, the dynamics used to model each vehicle in the platoon are introduced, as well as the dynamics of the CACC scheme employed to synchronize the vehicles' relative states. The available sensor signals are elaborated, together with the false data injections into these sensor signals. It is demonstrated that there exist multiple sensor configurations to realize the CACC controller. Each realization corresponds to using a different combination (or configuration) of sensors. All sensor configurations lead to identical platooning behaviour in attack-free conditions, but differ in their sensitivity to attacks. Then, the reachable set—induced by bounded false data injections—is introduced as a measure for robustness, and the formal definition of this reachable set is presented. These concepts form the framework of the problem statement. Based on this framework, the main research objectives of this thesis are formulated, which all serve the main goal; extending the understanding of cybersecurity from two-vehicle platoons to multi-vehicle configurations.

3.1 Longitudinal vehicle model

Consider a platoon of m vehicles. The vehicles are enumerated with index $i = 1, \ldots, m$ with $i = 1$ indicating the lead vehicle, see Figure 3.1. To model the platoon, the longitudinal vehicle model from [17] is adopted. Here, the longitudinal dynamics of vehicle i are described by the differential equations

$$\begin{aligned} \dot{q}_i &= v_i \\ \dot{v}_i &= a_i \\ \dot{a}_i &= -\frac{1}{\tau}a_i + \frac{1}{\tau}u_i \end{aligned} \qquad \text{for all } i = 1, \ldots, m. \tag{3.1}$$

The scalars q_i, v_i and a_i denote the absolute position of the rear bumper, the velocity and the acceleration of vehicle i respectively. The control input is denoted by u_i and can be interpreted as the desired acceleration of vehicle i, since the acceleration a_i asymptotically converges to input u_i for constant u_i. The scalar $\tau > 0$ is a time constant that captures the responsiveness of vehicles to longitudinal acceleration commands; vehicles with quicker responses (e.g., racing cars) are modeled with a smaller time constant, while slower-responding vehicles (e.g., trucks) have a larger time constant. A homogeneous platoon is considered, such that each vehicle i is assumed to have the same time constant τ.

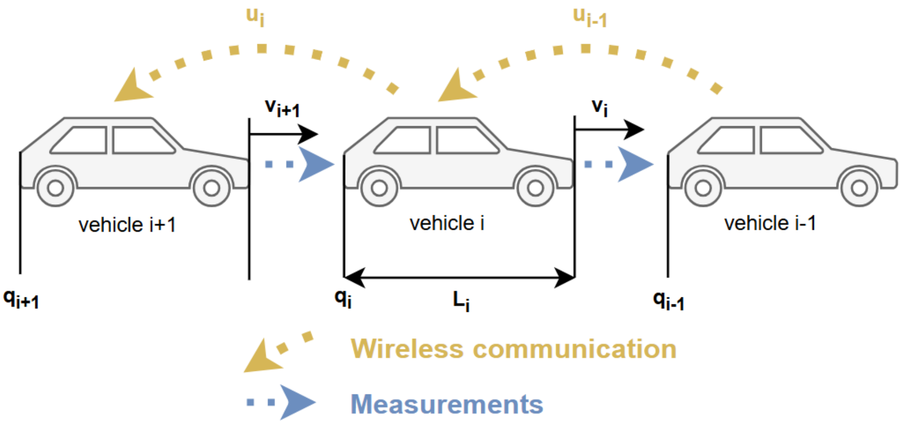

Figure 3.1: Visualization of CACC-equipped vehicle platoon. Each vehicle has onboard sensors (e.g. radars, camera and velocity/acceleration sensors) and can receive wireless signals from its predecessor.

3.2 CACC scheme

The objective of the CACC scheme is to achieve synchronization in the relative states of the vehicles in the platoon. Consider the inter-vehicular distance $d_i := q_{i-1} - q_i - L_i$, where the scalar L_i denotes the length of vehicle i. This inter-vehicular distance d_i represents the distance from the front bumper of vehicle i to the rear bumper of its preceding vehicle $i - 1$, and is therefore defined for all vehicles except the lead vehicle.

The so-called constant time-gap policy is employed. Here, the objective of each follower is to keep a desired reference distance $d_{r,i}$ to its preceding vehicle. This reference distance is defined as $d_{r,i} = r + hv_i$ for all $i = 2, \ldots, m$. The scalars $r > 0$ [m] and $h > 0$ [s] represent the standstill distance and time gap respectively. Consequently, the spacing error $e_i := d_r - d_{r,i}$ is the deviation of the inter-vehicular distance with respect to its reference distance.

The inter-vehicular distance of each vehicle is driven to its reference distance by a CACC algorithm. In this thesis, the algorithm adopted from [17] is considered, whose dynamics are described by

$$\begin{aligned} \dot{\xi}_i &= -\frac{1}{h}\xi_i + \frac{1}{h}\left(k_p e_i + k_d \dot{e}_i\right) + \frac{1}{h}u_{i-1}, \\ u_i &= \xi_i, \end{aligned} \tag{3.2}$$

with the internal controller state ξ_i. This controller ensures that the spacing error e_i asymptotically converges to zero with the control gains $k_p > 0$ and $k_d > 0$. It is important to note that the CACC algorithm of vehicle i thus uses information of both vehicle i and its preceding vehicle $i-1$. The CACC scheme in (3.2) is employed by all vehicles except the leader of the platoon, which controls its longitudinal velocity independently. Typically, this lead vehicle maintains a constant reference velocity v_r using a standard cruise control algorithm.

3.3 Controller realization and attack modeling

To show that there exist multiple sensor configurations that realize the same CACC scheme, the available sensor signals are introduced. The sensor signals of vehicle i are denoted by $y_{i,j}$, where j represents the sensor number. The available sensor signals of vehicle i are

$$\begin{aligned} &y_{i,1}(t) := d_i(t) + \delta_{i,1}(t), && y_{i,2}(t) := v_i(t) + \delta_{i,2}(t), && y_{i,3}(t) := a_i(t) + \delta_{i,3}(t), \\ &y_{i,4}(t) := v_{i-1}(t) - v_i(t) + \delta_{i,4}(t), && y_{i,5}(t) := a_{i-1}(t) + \delta_{i,4}(t), && y_{i,6}(t) := u_{i-1}(t) + \delta_{i,6}(t). \end{aligned} \tag{3.3}$$

The signals $\delta_{i,j}(t)$ with $j = 1, 2, 3, 4, 5, 6$ represent false data injections from attackers (i.e., hackers) into the sensor signals of vehicle i. All sensor signals are stacked into the column vector $y_i^\top = [y_{i,1},\, y_{i,2},\, y_{i,3},\, y_{i,4},\, y_{i,5},\, y_{i,6}]$. By means of the linear coordinate transformation on the internal controller state $\bar{\xi}_i = \xi_i + \beta y_i$ with $\beta_i = [\beta_{i,1},\, \beta_{i,2},\, \beta_{i,3},\, \beta_{i,4},\, \beta_{i,5},\, 0]$ where each $\beta_{i,j} \in \mathbb{R}$, there exist infinitely many controller realizations. Note that in the proposed change of coordinates $\beta_{i,6} = 0$ (effectively excluding $y_{i,6}$), since this would require information about $\dot{u}_{i-1}$ and thus also knowledge about the control structure of vehicle $i-1$ (and possibly sensor data from vehicle $i-2$), which is shown in [28]. As a result of the linear coordinate transformation $\bar{\xi}_i = \xi_i + \beta y_i$, the CACC dynamics in (3.2) can be alternatively represented by

$$\begin{aligned} \dot{\bar{\xi}}_i &= f_\xi(\beta_i)\bar{\xi}_i + f_y(\beta_i)y_i, \\ u_i &= \bar{\xi}_i - \beta_i y_i. \end{aligned} \tag{3.4}$$

This notation is adopted from [28]. Analytical expressions for the functions $f_\xi(\beta_i)$ and $f_y(\beta_i)$ are presented in Appendix A.1. The vector β_i determines how much each sensor is used in the CACC scheme, and therefore parametrizes the sensor configuration. All sensor configurations yield equivalent platooning behaviour in the absence of attacks, but differ in their sensitivity to the false data injections $\delta_i(t)$. In other words, the controller realization determines how much the CACC scheme depends on each sensor signal, and therefore also determines how false data injections enter the system.

In this research, the case is considered where only the measurements and communication between the lead vehicle and the first follower vehicle are compromised, such that the sensor signals of vehicle 2 are injected with false data, see Figure 3.2. Though any vehicle in the platoon could be attacked, this is equivalent to attacking the lead vehicle by choosing the vehicle in front of the attacked vehicle as the leader. Consequently, this research only focuses on the sensor configuration of the first follower vehicle, denoted by β_2.

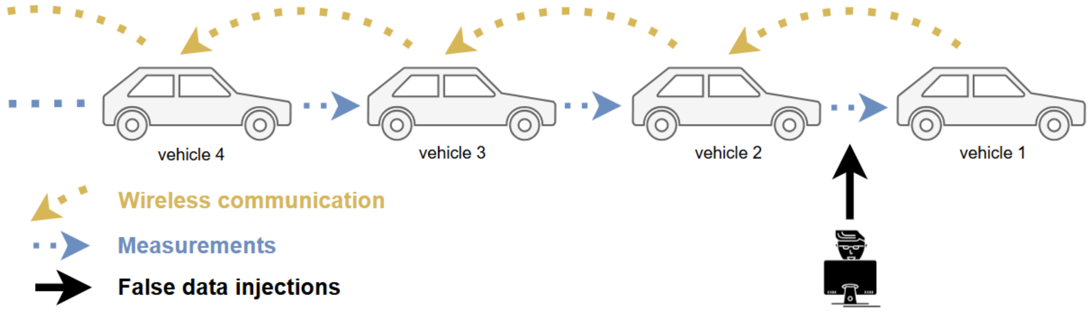

Figure 3.2: Visualization of the problem setup, where the measurements and communication between the lead vehicle and its first follower is compromised by hackers with false data injections.

It is assumed that false data injections $\delta_{i,j}(t)$ are undetectable as long as their $\mathcal{L}_\infty$-norm remains below a predefined threshold $\bar{\delta}_j \geq 0$, i.e., as long as $||\delta_{i,j}||_{\mathcal{L}_\infty} \leq \bar{\delta}_j$. In other words, undetectable false data injections are contained within the hyperrectangular set

$$\mathcal{U} := \left\{ \ \delta \in \mathbb{R}^p \ \middle| \ |\delta_j| \leq \bar{\delta}_j \text{ with } \bar{\delta}_j \geq 0 \text{ for all } j = 1, 2, \ldots, p \ \right\} \tag{3.5}$$

such that $\delta_{i,j}(t) \in \mathcal{U}$ for all time t. Since this thesis considers a *passive* cybersecurity method, the CACC scheme should be inherently robust (i.e., guarantee safety) against any attacks within this set. Attacks outside of $\mathcal{U}$ are detected and can be mitigated through *active* methods, such as switching control strategy.

3.4 Reachable set

To quantify the robustness of the platoon against undetected attacks in the set $\mathcal{U}$, the reachable set, denoted by $\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})}$, is introduced. The reachable set is the set of states that the system can reach, starting from any initial condition in a predefined set $\mathcal{X}_0$, induced by false data injections that lie in the set $\mathcal{U}$. This set thus measures the size of the state space portion that attacks can induce, and therefore gives insight into the potential damage to the platoon, such as collisions.

To provide a formal definition of the reachable set, let the state of the complete vehicle platoon be denoted by $x^\top = [x_1^\top, x_2^\top, \ldots, x_m^\top]$. Here, the state vector of the complete platoon x is composed of the state vectors of each vehicle i, denoted by x_i. These state vectors are defined in Chapter 6, where it is demonstrated that the dynamics of the complete platoon can be written as the linear time-invariant continuous-time system

$$\dot{x}(t) = A\,x(t) + B(\beta_2)\,\delta(t) \tag{3.6}$$

with state $x \in \mathbb{R}^n$ and input $\delta \in \mathbb{R}^p$ with $p = 6$. Note that only the input matrix $B(\beta_2)$ depends on the sensor configuration of vehicle 2, denoted by β_2. The matrix A—characterizing the closed-loop dynamics of the platoon—is invariant to the sensor configuration β_2, highlighting that the sensor configuration only affects how false data injections enter the system, and that the closed-loop dynamics remain unaffected. This is derived in Chapter 6, which also demonstrates that by means of a coordinate translation, the system is asymptotically stable to the origin. This origin corresponds to a fully synchronized platoon.

A state x lies in the reachable set $\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})}$ if there exists a time $T \geq 0$, an initial condition $x(0)$ with $x(0) \in \mathcal{X}_0$ and input signal $\delta(t)$ with $\delta(t) \in \mathcal{U}$ for all time $t \in [0, T]$ for which the system can reach state x. Vice versa, if a state x does not lie in the reachable set, there exists no such initial condition $x(0)$ and input signal $\delta(t)$. Formally, the reachable set of system (3.6) is

$$\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})} := \left\{ \ x \in \mathbb{R}^n \ \middle| \ \begin{aligned} \exists T \geq 0 \text{ such that } x(T) \text{ solution to (3.6)}, \\ \text{and } x(0) \in \mathcal{X}_0, \\ \text{and } \delta(t) \in \mathcal{U} \text{ for all } t \in [0, T]. \end{aligned} \right\}. \tag{3.7}$$

This thesis studies the cascaded system formed by a platoon of vehicles. The state vector of an individual vehicle i is denoted by x_i, and the complete state vector of the platoon is constructed by stacking the individual states such that

$$x^\top = \begin{bmatrix} x_1^\top & x_2^\top & \ldots & x_m^\top \end{bmatrix}. \tag{3.8}$$

To analyze the reachable set of an individual vehicle, the platoon's complete reachable set is projected onto the states of this vehicle. The reachable set of vehicle i is defined as the reachable set of the complete platoon, projected onto vehicle i's coordinates x_i (or more precisely, on the subspace spanned by these coordinate axes) such that;

$$\text{the reachable set of vehicle } i \text{ is } \Pi_{x_i}[\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})}]. \tag{3.9}$$

Computing the exact reachable set of a dynamical system is generally intractable. Several numerical methods have been developed that rely on temporal discretization of the system. However, their computational complexity scales poorly with the state dimension, rendering them impractical for high-dimensional systems such as vehicle platoons (see Chapter 4).

To address this issue, an *outer approximation* of the reachable set is considered instead, denoted by $\mathcal{A}(\mathcal{R})$. This outer approximation fully contains the true reachable set $\mathcal{R}$, ensuring that any safety guarantees established for $\mathcal{A}(\mathcal{R})$ also apply to $\mathcal{R}$. Note that this approach can introduce some degree of conservatism.

Previous studies have used *ellipsoidal* sets for outer approximations. However, ellipsoids have notable limitations, making it challenging to extend this method to multi-vehicle platoons, as elaborated in Chapter 4. This motivates the use of different sets to outer approximate the reachable set, such as *hyperrectangles*.

3.5 Research objectives

This chapter has introduced the theoretical concepts on which this thesis builds. The dynamics that are used to model each individual vehicle in the platoon have been presented, alongside the formulation of the CACC scheme. It has been shown that there exist multiple sensor configurations that realize this scheme. Each sensor configuration has equivalent attack-free dynamics, but a different sensitivity when the sensors are attacked. To measure the robustness of the platoon under cyberattacks, the notion of the reachable set has been introduced, defined as the set of states the system can reach under bounded false data injection. Since numerical methods to compute the reachable set are numerically inefficient, previous work has considered the ellipsoidal outer approximation of the reachable set to formulate safety guarantees. This ellipsoidal approach, however, has only been considered for the two-vehicle case, and is challenging to apply to multi-vehicle platoons. The main goal of this thesis is therefore to extend the understanding of cybersecurity in two-vehicle platoons to multi-vehicle configurations. This goal is divided into five distinct objectives.

The first objective of this thesis is to address the limitations of the ellipsoidal outer approximation of the reachable set, by **introducing a technique to compute the hyperrectangular (box-shaped) outer approximation of the reachable set**. This is done in Chapter 5, where the conservativeness of this hyperrectangular outer approximation is also elaborated.

The second objective of this thesis is to **develop a dynamical model that captures the behaviour of the complete vehicle platoon**, extending the two-vehicle case to multi-vehicle platoons. In Chapter 6, this is achieved by interconnecting the dynamics of consecutive vehicles to form a cascaded system. The resulting model is formulated in two coordinate frames: the *original coordinates*, which offer an intuitive physical interpretation, and the *error coordinates*, which make it easier to identify key system properties.

With the full platoon dynamics established, the third objective is to **characterize the directions in the state space that are vulnerable to false data injections**, i.e., the directions the system state can be driven toward, regardless of the attack magnitude. The subspace formed by these directions is referred to as the *attackable subspace*. The attackable subspace provides a qualitative measure of robustness; there are some directions where the state of the platoon cannot be steered, no matter how much false data is injected. This analysis is carried out in Chapter 7, where the attackable subspace is first derived in the error coordinates and then transformed and interpreted in the original coordinates.

Building on this, the fourth objective is to **formulate safety guarantees on the propagation of the reachable sets along the platoon**. Chapter 8 computes the hyperrectangular approximation of the complete vehicle platoon in terms of the original coordinates, and establishes bounds on how the reachable set evolves from one vehicle to the next. A new notion of string stability—*$\mathcal{L}_\infty$-q string stability*—is introduced. This notion ensures that the size of the reachable set does not grow downstream from vehicle q, providing a formal safety guarantee. It is shown that the platoon is $\mathcal{L}_\infty$-q string stable with $q \leq 3$.

Finally, the fifth objective is to **develop a synthesis framework for shaping the reachable set by choosing the sensor configuration**. Given that the reachable sets do not grow downstream from the third vehicle, Chapter 9 proposes a convex optimization-based framework to minimize the union of the reachable set approximations for the first three vehicles. This enables the identification of sensor configurations that optimize robustness in the presence of bounded attacks.

Before addressing the five research objectives outlined above, the next chapter presents a literature review on reachable sets of dynamical systems. This review highlights the computational challenges associated with numerical methods to compute the reachable set, particularly in high-dimensional systems such as vehicle platoons. It also discusses the limitations of conventional ellipsoidal outer approximations, which motivate the need for alternative approximation methods.

4 Literature review

The previous chapter has formulated the research objectives of this thesis. Before addressing these objectives, this chapter presents a literature review on research on reachable sets of dynamical systems. First, a numerical algorithm for computing the reachable set of a dynamical system via discretization of its continuous-time dynamics is examined. While this method provides precise results, it suffers from scalability issues due to its computational intensity, especially for high-dimensional systems. As such, in this thesis it is primarily used as a benchmark to evaluate approximation methods. Next, an ellipsoidal outer-approximation technique for the reachable set is reviewed, adopted from [30,31]. This method is used in [28] to formulate a controller synthesis framework. This approach too requires discretization of the dynamics, but provides a tractable alternative for higher-dimensional systems. However, this ellipsoidal approach has some limitations—such as a lack of generalizability to multi-vehicle platoons—which are elaborated, and motivate the need for alternative approximation methods.

4.1 Discrete-time reachable set

The methods to compute the reachable set that are discussed in this chapter require temporal discretization of the system. Though the specific discretization method depends on the context, this thesis considers exact discretization for simplicity. Consequently, the continuous-time system in (3.6), with state vector x whose dynamics are characterized by matrix A and B, is discretised using a fixed sampling time T_s, resulting in the discrete-time system

$$x(k+1) = A_d\, x(k) + B_d\, u(k) \text{ with } A_d = e^{AT_s} \text{ and } B_d = \left(\int_0^{T_s} e^{A\tau} d\tau \right) B. \tag{4.1}$$

Analogous to the continuous-time reachable set defined in (3.7), its discrete-time counterpart is defined, denoted by $\mathcal{R}^{\text{discrete}}_{(\mathcal{X}_0,\mathcal{U})}$. As the sampling time T_s decreases, the discrete-time reachable set approaches the continuous-time reachable set. The discrete-time reachable set is formally defined as

$$\mathcal{R}^{\text{discrete}}_{(\mathcal{X}_0,\mathcal{U})} := \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{r} \exists K \geq 0 \text{ such that } x(K) \text{ solution to (4.1)}, \\ \text{and } x(0) \in \mathcal{X}_0, \\ \text{and } u(k) \in \mathcal{U} \text{ for all } k \in [0,K]. \end{array} \right\}. \tag{4.2}$$

4.2 Numerical algorithm to compute reachable set

Now, a common numerical algorithm to compute the reachable set of the discrete-time system in (4.1) is introduced. Let $\mathcal{X}_k$ denote the discrete-time reachable set at time k, which approaches $\mathcal{X}_k \to \mathcal{R}^{\text{discrete}}_{(\mathcal{X}_0,\mathcal{U})}$ as $k \to \infty$. Both the reachable set $\mathcal{X}_k$ and the input set $\mathcal{U}$ are modeled as convex hulls of point sets $\mathcal{S}_{x_k}$ and $\mathcal{S}_u$ containing a number of n_{x_k} and n_u points respectively such that

$$\begin{aligned} \mathcal{X}_k &:= \text{conv}(\mathcal{S}_{x_k}) \text{ with } \mathcal{S}_{x_k} = \left\{ x_k^1, \ldots, x_k^{n_{x_k}} \right\}, \\ \mathcal{U} &:= \text{conv}(\mathcal{S}_u) \text{ with } \mathcal{S}_u = \left\{ u^1, \ldots, u^{n_u} \right\}. \end{aligned} \tag{4.3}$$

Algorithm 1 (Reachable set computation)**.** *Consider the discrete-time dynamical system in* (4.1) *and the discrete-time reachable set in* (4.2) *with set of allowed inputs* $\mathcal{U}$ *in* (3.5) *and set of initial conditions* $\mathcal{X}_0$ *.*

1. ***Initialization:*** *Set* $k = 0$*. The sets* $\mathcal{S}_{x_k}$ *and* $\mathcal{S}_u$ *are known. The reachable set* $\mathcal{X}_0 = conv\{\mathcal{X}_0\}$*.*
2. ***Propagation:*** *For each* $x_k \in \mathcal{S}_{x_k}$ *and* $u_k \in \mathcal{S}_u$*, compute* $x_{k+1} = A_d\, x_d + B_d\, u_k$*. Store the resulting* $\tilde{n}_{x_{k+1}} = (n_{x_k} \times n_u)$ *points in set* $\tilde{\mathcal{S}}_{x_{k+1}}$*.*
3. ***Prune:*** *Keep only the vertices of* $conv(\tilde{\mathcal{S}}_{x_{k+1}})$ *to form* $\mathcal{S}_{k+1}$*. Let set* $\mathcal{X}_{k+1} = conv(\mathcal{S}_{x_{k+1}})$*.*
4. ***Check termination:*** *If the subsequent reachable sets are sufficiently similar such that* $\mathcal{X}_{k+1} \cong \mathcal{X}_k$ *(within tolerance), proceed to Step 5. Otherwise, increment* k *and repeat from Step 2.*
5. ***Output:*** *The approximate reachable set is* $\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})} \cong \mathcal{X}_{k+1}$*.*

Note that the pruning in Step 3 is not necessary to obtain $\mathcal{X}_{k+1}$, but drastically improves efficiency. Without pruning, n_{x_k} grows exponentially as $n_{x_{k+1}} = n_{x_k} n_{u_k}$. Pruning decreases this growth rate—e.g., it ensures at most linear growth in two-dimensional systems (with state dimension $n = 2$) such that $n_{x_{k+1}} \leq n_{x_k} + n_{u_k}$, see [32].

Though the pruning improves the numerical efficiency of the algorithm, the computational load still scales poorly with the state dimension. In systems with a state dimension $n \geq 3$, there is no explicit bound on the growth rate of the amount of vertices of the set. To demonstrate this, the numerical algorithm is applied to the two- and three-dimensional systems

$$\Sigma_1 : \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \delta \text{ and } \Sigma_2 : \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -3 & 1 & 0 \\ -2 & -4 & 1 \\ 0 & -2 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \delta. \tag{4.4}$$

Both systems are discretised with sampling time $T_s = 0.1$ [s] such that Algorithm 1 can be applied. The set of initial states of both systems is the origin $\mathcal{X}_0 = \{0\}$. The set of allowed inputs is $\mathcal{U}$ in (3.5) with $\bar{\delta}_j = 1$ for all $j = 1, 2$ and $j = 1, 2, 3$ for system Σ_1 and Σ_2 respectively. The computed reachable sets are visualized in Figure 4.1 and Figure 4.2 respectively.

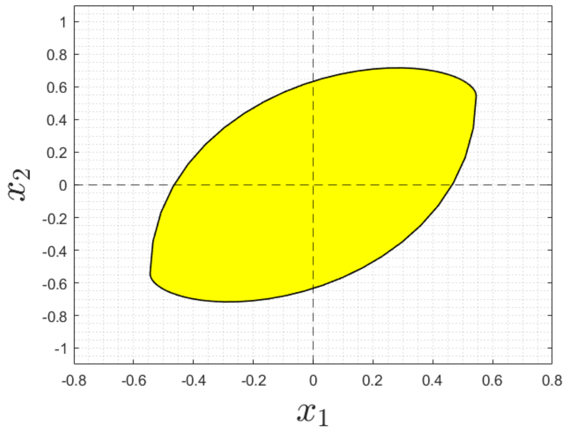

Figure 4.1: Visualization of the reachable set of the two-dimensional system Σ_1 in (4.4), discretised with sampling time $T_s = 0.1$ [s].

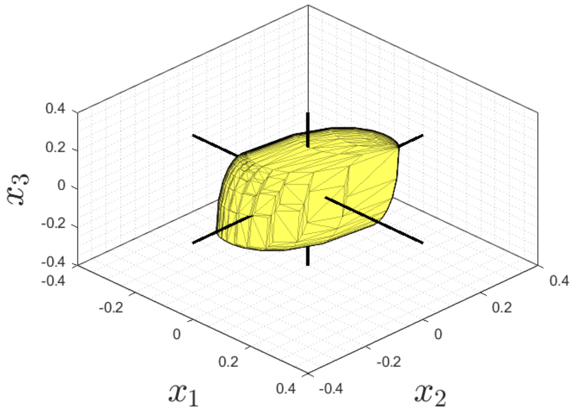

Figure 4.2: Visualization of the reachable set of the three-dimensional system Σ_2 in (4.4), discretised with sampling time $T_s = 0.1$ [s].

To highlight the poor scalability of the computational load with respect to the state dimension, Figure 4.3 shows the number of vertices over time for the first 30 iterations. As expected, the number of vertices grows linearly in the two-dimensional set but much faster in three-dimensional set.

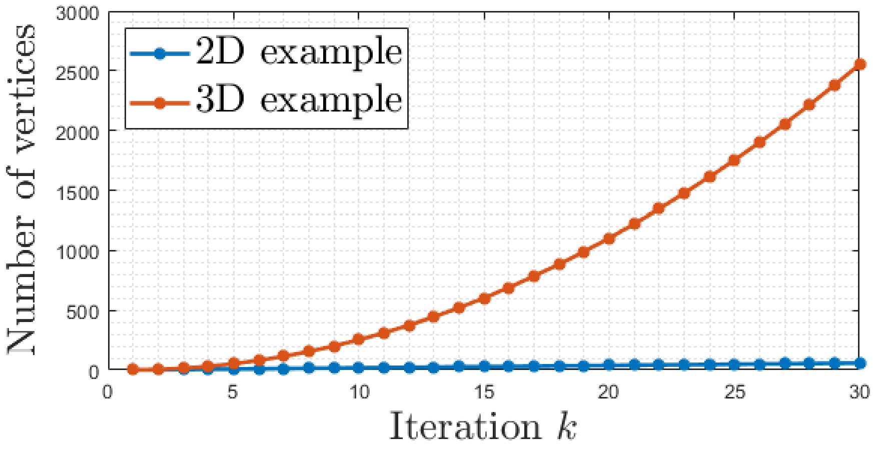

Figure 4.3: Number of points in the set $\mathcal{S}_{x_1}$ that form the vertices of the polygon, obtained by executing the numerical algorithm. The figure shows linear growth for the two-dimensional example, and must faster growth for the three-dimensional example.

Since this thesis considers a platoon consisting of many vehicles, the dynamical system typically has a high state dimension, rendering Algorithm 1 impractical. Instead of the actual reachable set, an outer approximation of the true reachable set is therefore considered. The method in Algorithm 1 is thus not used to analyze the reachable set of the vehicle platoon. Instead, it serves as a benchmark to evaluate and compare different outer approximation techniques.

4.3 Ellipsoidal approximation of reachable set

Given the poor scalability of the computational load of the method described above, this method is thus not used to analyze the reachable set of the vehicle platoon. Instead, previous work considers an outer approximation of the true reachable set. Since the outer approximation fully contains the true reachable set, any safety guarantees (such as bounds on the velocity deviation) established for the outer approximation also apply to the true reachable set. There exist multiple techniques to compute an outer approximation of the reachable set. Huisman et al. [27, 28] propose using ellipsoidal outer approximations of the form

$$\mathcal{E} := \left\{ \; x \in \mathbb{R}^n \;\middle|\; x^\top P x \leq 1 \; \right\} \quad \text{with } P \succ 0. \tag{4.5}$$

These ellipsoids are computed using the convex optimization-based approach developed in [30, 31]. Similarly to Algorithm 1, this method also requires temporal discretization of the system in (3.6). The method is summarized in the following lemma:

Lemma 1 (Ellipsoidal approximation). *Consider the discrete-time system in* (4.1) *with system matrices* (A_d, B_d) *and the discrete-time reachable set in* (4.2) *with set of allowed inputs* $\mathcal{U}$ *in* (3.5) *and set of initial conditions* $\mathcal{X}_0$. *For a given* $a \in [\max_i \left\{\lambda_i^2(A_d)\right\}, 1)$, *if there exist constants* $a_1, \ldots, a_p$ *and matrix* P *that is the solution of the convex program*

$$\begin{aligned} \min_{P, a_1, \ldots, a_6} \quad & -\log\det[P] \\ s.t. \quad & a_1, \ldots, a_p \in (0,1) \\ & a_1 + \cdots + a_p \geq a \\ & P \succ 0 \\ & \begin{bmatrix} aP & A_d^\top P & 0 \\ PA_d & P & PB_d \\ 0 & B_d^\top P & W_a \end{bmatrix} \succeq 0 \end{aligned} \tag{4.6}$$

with matrix $W_a := diag[(1-a_1)\frac{1}{\sqrt{\bar{\delta}_1}}, \ldots, (1-a_p)\frac{1}{\sqrt{\bar{\delta}_p}}]$, *then set* $\mathcal{E} = \left\{ \; x \in \mathbb{R}^n \;\middle|\; x^\top(\frac{1-a}{p-a}P)x \leq 1 \; \right\}$ *is an outer approximation of the reachable set such that* $\mathcal{R}^{discrete}_{(\{0\},\mathcal{X}_0)} \subseteq \mathcal{E}$. *This ellipsoid* $\mathcal{E}$ *has the minimum volume among all outer ellipsoidal approximations of the reachable set.*

This method thus computes the ellipsoidal set with the minimum volume that still fully contains the true reachable set. To demonstrate this, the ellipsoidal outer approximation is computed by solving the convex optimization program in Lemma 1 for system Σ_1 and Σ_2 in (4.4). The resulting two- and three-dimensional ellipsoids, overlaid on the true reachable sets, are visualized in Figure 4.4 and Figure 4.5 respectively.

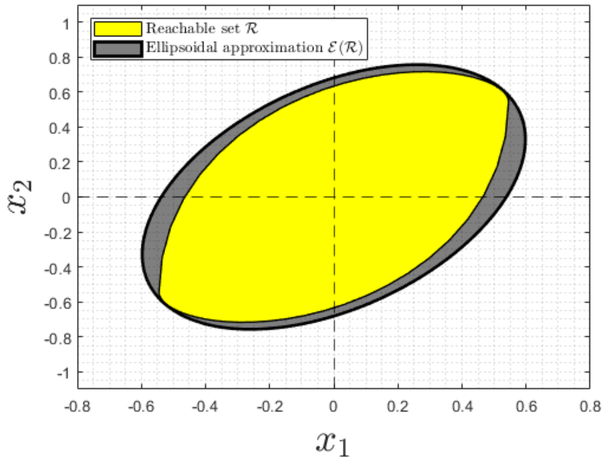

Figure 4.4: Reachable set of the two-dimensional system Σ_1 with its ellipsoidal approximation.

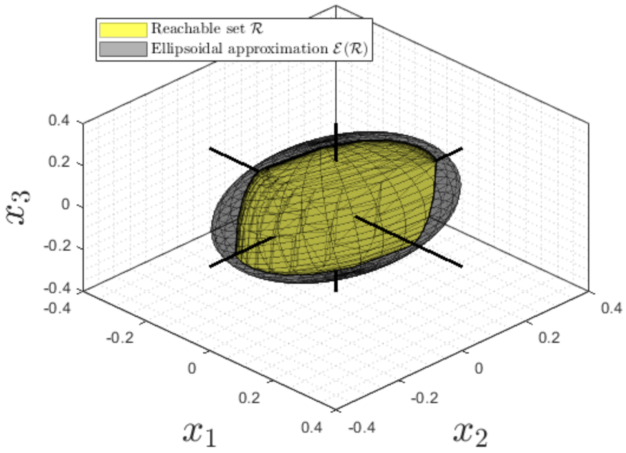

Figure 4.5: Reachable set of the three-dimensional system Σ_2 with its ellipsoidal approximation.

This method computes the ellipsoidal outer approximation of the true reachable set *for a specific sensor configuration* β_2. In subsequent work, Huisman et al. [28] formulate a controller synthesis framework that selects an optimal realization β_2 that minimizes the volume of the ellipsoidal outer approximation of the reachable set. In other words, the reachable set is shaped by selecting β such that its ellipsoidal outer approximation has the minimum volume. This is formulated as a convex optimization problem, described in the following lemma:

Lemma 2 (Controller synthesis)**.** *Consider the discrete-time system in* (4.1) *with system matrices* (A_d, B_d) *and the discrete-time reachable set in* (4.2) *with set of allowed inputs* $\mathcal{U}$ *in* (3.5) *and set of initial conditions* $\mathcal{X}_0$*. For a given* $a \in [\max_i \left\{\lambda_i^2(A_d)\right\}, 1)$*, if there exist constants* $a_1, \ldots, a_p$ *and vector* β_2 *and matrix* Y *that is the solution of the convex program*

$$
\begin{aligned}
\min_{Y,\beta_2,a_1,\ldots,a_p} \quad & tr[Y] \\
s.t. \quad & a_1, \ldots, a_p \in (0,1) \\
& a_1 + \cdots + a_p \geq a \\
& Y \succ 0 \\
& \begin{bmatrix} aY & YA_d^\top & 0 \\ A_d^\top & Y & B_d(\beta_2) \\ 0 & B_d^\top(\beta_2) & W_a \end{bmatrix} \succeq 0
\end{aligned}
\tag{4.7}
$$

with matrix $W_a := diag[(1-a_1)\frac{1}{\sqrt{\bar{\delta}_1}}, \ldots, (1-a_p)\frac{1}{\sqrt{\bar{\delta}_p}}]$*, then set* $\mathcal{E} = \left\{ \; x \in \mathbb{R}^n \;\middle|\; x^\top(\frac{1-a}{p-a}P)x \leq 1 \; \right\}$ *with* $P = Y^{(-1)}$ *is the ellipsoidal outer approximation of the reachable set such that* $\mathcal{R}^{discrete}_{(\{0\},\mathcal{X}_0)} \subseteq \mathcal{E}$*. Furthermore, this is the ellipsoidal outer approximation with the minimum volume.*

The method in Lemma 2 thus shapes the reachable set such that the ellipsoidal outer approximation has the minimum volume by choosing the optimal sensor configuration β_2. Though the method to compute the ellipsoidal outer approximation in Lemma 1 provides a tractable alternative for the technique described in Algorithm 1, the method has some limitations.

Firstly, the computation of the ellipsoidal outer approximation of the reachable set discussed in this thesis requires temporal discretization of the continuous-time system dynamics in (3.6), therefore approximating the *discrete-time* reachable set in (4.2) rather than the *continuous-time* set in (3.7). Though these two sets can be similar for small sampling times T_s, a fundamental discrepancy is introduced, especially in systems where the discretization of time is non-trivial.

Secondly, ellipsoidal outer approximations tend to overestimate the reachable set along the axes of the coordinate frame. E.g., consider the ellipsoidal outer approximation of the reachable set in Figure 4.4. The projection of the true reachable set onto coordinate x_1 is smaller than the projection of its ellipsoidal outer approximation onto this coordinate x_1, such that $\Pi_{x_1}[\mathcal{R}] \subset \Pi_{x_1}[\mathcal{E}(\mathcal{R})]$. This is also true for the projection onto coordinate x_2. In other words, the ellipsoidal outer approximation is often *not tight* around the true reachable set in the considered coordinate frame. The axes of the coordinate frame typically represent states of direct interest, such as velocity or inter-vehicular distance. Ellipsoidal outer approximations therefore tend to be introduce conservatism in the directions of interest. Alternative techniques to compute an outer approximation that are guaranteed to be tight around the true reachable set would therefore reduce conservatism.

Finally and most importantly, analyzing how the reachable set evolves along a vehicle platoon—i.e., evaluating string stability—is challenging when using the ellipsoidal outer approximation. The techniques to compute these ellipsoids in Lemmas 1 and 2 involve optimization programs. While these programs can be solved numerically for a small number of vehicles in the platoon, there is no analytical expression for the optimal solution. This makes it challenging to generalize analysis to arbitrary platoon lengths m and arbitrary system parameters. This motivates the need of an alternative outer approximation that can be expressed analytically.

In summary, this chapter has presented an overview of the relevant literature about research on reachable sets of dynamical systems. A numerical method to compute the reachable set has been elaborated, as well as a method to compute the corresponding ellipsoidal outer approximation of the reachable set. However, their limitations have also been highlighted, such as the lack of scalability to multi-vehicle platoons. This motivates the need for an alternative outer approximation method. To this end, the next chapter proposes to use a hyperrectangular outer approximation of the reachable set.

5 Hyperrectangular outer approximation of reachable set

The previous chapter has highlighted the limitations of using an ellipsoid to outer approximate the reachable set. This chapter therefore proposes an alternative: the hyperrectangular outer approximation of the reachable set. First, a formal definition of a hyperrectangular outer approximation is presented. Then the method to compute such an approximation is described, which does not include temporal discretization of the continuous-time system, allowing for an analytical expression of the hyperrectangle. Then, some claims regarding its conservatism with respect to the true reachable set are formulated. This is done by considering the autonomous and origin-reachable set individually (which are defined in this chapter). Finally, an example is presented where the hyperrectangular outer approximation of the reachable set is computed for a two- and three-dimensional system.

5.1 Formal definition

Consider the class of linear time-invariant continuous-time asymptotically stable systems in (3.6) and their reachable set $\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})}$ in (3.7). Instead of the actual reachable set, a *hyperrectangular outer approximation* is considered. A hyperrectangle is a rectangle generalized to any dimension n. The hyperrectangular approximation should fully contain the true reachable set such that it is an outer approximation. This ensures that safety guarantees (such as upper bounds on the velocity deviations) of the outer approximation also apply to the true reachable set. The hyperrectangle is chosen to be *centered in the origin* and *aligned with the axes of the coordinate frame*, which makes projections onto individual states more intuitive. The hyperrectangular outer approximation $\mathcal{H}(\mathcal{R})$ of the true reachable set $\mathcal{R}$ is defined as

$$\mathcal{H}(\mathcal{R}) := \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{l} |x_i| \leq \bar{x}_i \text{ for all } i = 1, 2, \dots, n \\ \text{such that } \mathcal{R} \subseteq \mathcal{H} \end{array} \right\}. \tag{5.1}$$

The scalars $\bar{x}_i$ represent the dimensions (i.e., the size of the sides of the box, or the *half-widths* of this box) of the hyperrectangle. This is visualized in Figure 5.1. Note that the approximation depends on the choice of coordinate frame. This means that the approximation $\mathcal{H}(\mathcal{R})$ can be tight around the reachable set $\mathcal{R}$ when viewed from one coordinate frame x, but not when viewed from another coordinate frame x'.

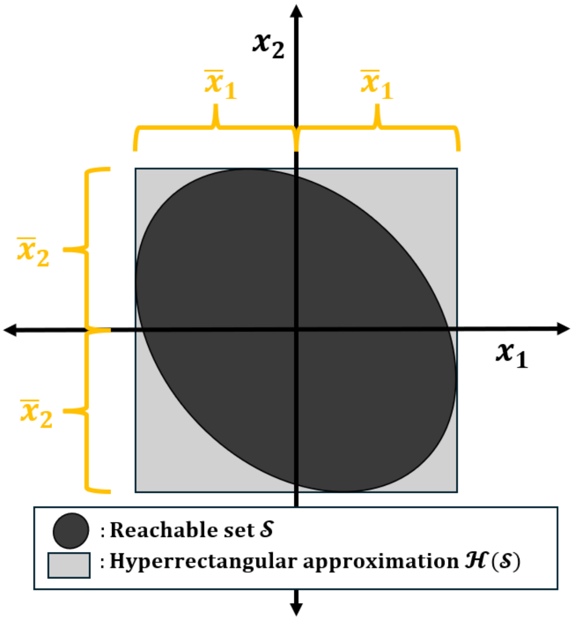
x_2
$\bar{x}_1$
$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_2$
x_1
● : Reachable set $\mathcal{S}$
▢ : Hyperrectangular approximation $\mathcal{H}(\mathcal{S})$

Figure 5.1: Visualization of the hyperrectangular outer approximation $\mathcal{H}(\mathcal{R})$ of a reachable set $\mathcal{R}$. The size of the hyperrectangle is parametrized by $\bar{x}_1$ and $\bar{x}_2$, which represent the half-widths of its sides. The outer approximation is tight around the reachable set when viewed in the coordinate frame $x = (x_1, x_2)$.

To compute the hyperrectangular outer approximation of the reachable set, this set is divided into two parts; the *origin-reachable set* and the *autonomous reachable set.* Before analyzing the complete reachable set, the hyperrectangular outer approximation of these two individual sets is elaborated.

5.2 Approximation of origin-reachable set

First, the origin-reachable set is considered. This is the reachable set where it is imposed that all trajectories start in the origin such that $\mathcal{X}_0 = \{0\}$. The origin-reachable set is denoted by $\mathcal{R}_{(\{0\},\mathcal{U})}$. To compute the hyperrectangular outer approximation of the origin-reachable set, the dynamical system in (3.6) with state vector x and input vector δ whose dynamics are described by the system matrices (A, B) are represented in the Laplace domain as

$$\hat{x}(s) = G(s)\,\hat{\delta}(s) \text{ with } G(s) = (sI - A)^{-1}B. \tag{5.2}$$

Here, $\hat{x}(s)$ and $\hat{\delta}(s)$ represent the Laplace transforms of $x(t)$ and $\delta(t)$ respectively. Let $\hat{x}_i(s)$ denote the i^{th} state in the state vector $\hat{x}(s)$, and let $\hat{\delta}_j(s)$ denote the j^{th} input in the input vector $\hat{\delta}(s)$. To compute the origin-reachable set, first the reachable set induced by each individual input $\delta_j(t)$ is considered separately. This is the reachable set where it is imposed that $\delta_k(t) = 0$ for all $k \neq j$. This set is denoted by $\mathcal{R}^j_{(\mathcal{X}_0,\mathcal{U})}$ and formally defined as

$$\mathcal{R}^j_{(\mathcal{X}_0,\mathcal{U})} := \left\{ \; x \in \mathcal{R}_{(\mathcal{X}_0,\mathcal{U})} \;\middle|\; \delta_k(t) = 0 \text{ for all } t \in [0, T] \text{ and for all } k \neq j \; \right\}. \tag{5.3}$$

The superscript j specifies that this is the reachable set resulting from input $\delta_j(t)$, setting all other inputs to zero. Since each input $\delta_j(t)$ can be chosen independently, and since the dynamical system is linear, the complete origin-reachable set $\mathcal{R}_{(0,\mathcal{U})}$ is the Minkowski sum of the individual sets $\mathcal{R}^j_{(0,\mathcal{U})}$. This is a well known result from reachability analysis of linear systems. For completeness, the corresponding proof is recalled in the following lemma:

Lemma 3 (Minkowski decomposition of origin-reachable set)**.** *Consider the linear time-invariant continuous-time asymptotically stable system in* (3.6) *with state vector x, input vector δ and system matrices (A, B) and hyperrectangular set of allowed inputs $\mathcal{U}$ in* (3.5)*. Let $\mathcal{R}^j_{(\{0\},\mathcal{U})}$ denote the origin-reachable set induced by input δ_j defined in* (5.3)*. The complete origin-reachable set in* (3.7)*, denoted by $\mathcal{R}_{(\{0\},\mathcal{U})}$, is equivalent to the Minkowski-sum of the origin-reachable sets induced by each individual input δ_j such that*

$$\mathcal{R}_{(\{0\},\mathcal{U})} = \mathcal{R}^1_{(\{0\},\mathcal{U})} \oplus \cdots \oplus \mathcal{R}^p_{(\{0\},\mathcal{U})}. \tag{5.4}$$

Proof: *Let $x \in \mathcal{R}_{(\{0\},\mathcal{U})}$. This implies that there exists an input signal $\delta(t) \in \mathcal{U}$ for all t such that the trajectory starting from the origin satisfies $x = x(t;\delta)$ for some $t \geq 0$. Since $\mathcal{U}$ is a hyperrectangle aligned with the coordinate axes, the input can be decomposed as*

$$\delta(t) = \sum_{j=1}^{p} \delta^j(t) \text{ with } \delta^j(t) := \delta_j(t)\mathbf{e}_j. \tag{5.5}$$

Each $\delta^j(t)$ has only one non-zero component and satisfies $\delta^j(t) \in \mathcal{U}$ for all t. By linearity of the system, the response to $\delta(t)$ is the sum of the responses to each $\delta^j(t)$ such that

$$x(t;\delta) = \sum_{j=1}^{p} x^j \text{ with } x^j = x(t;\delta^j), \tag{5.6}$$

where each $x^j = x(t;\delta^j) \in \mathcal{R}^j_{(\{0\},\mathcal{U})}$ by construction. Therefore, the state induced by all inputs δ_j is $x = x^1 + \cdots + x^p \in \mathcal{R}^1_{(\{0\},\mathcal{U})} \oplus \cdots \oplus \mathcal{R}^p_{(\{0\},\mathcal{U})}$, which proves the inclusion

$$\mathcal{R}_{(\{0\},\mathcal{U})} \subseteq \mathcal{R}^1_{(\{0\},\mathcal{U})} \oplus \cdots \oplus \mathcal{R}^p_{(\{0\},\mathcal{U})}. \tag{5.7}$$

Intuitively, any input signal can be decomposed into individual channels. System linearity implies that the resulting state can be expressed as the sum of responses to these individual input components. Thus, every reachable state belongs to the Minkowski sum of the individually reachable set, which is captured in the inclusion in (5.7)*.*

Now, let $x \in \mathcal{R}^1_{(\{0\},\mathcal{U})} \oplus \cdots \oplus \mathcal{R}^p_{(\{0\},\mathcal{U})}$. This implies that there exist states $x^j \in \mathcal{R}^j_{(\{0\},\mathcal{U})}$ induced by individual inputs δ_j such that the state induced by all inputs is $x = x^1 + \cdots + x^p$. By definition of each $\mathcal{R}^j_{(\{0\},\mathcal{U})}$, there exists an input $\delta^j(t) = \delta_j(t)\mathbf{e}_j$ that drives the system from the origin to x^j. Construct the input signal $\delta(t)$ similarly to the summation in (5.5)*, and note that $\delta(t) \in \mathcal{U}$ for all t. By linearity of the system,*

$$x(t;\delta) = \sum_{j=1}^{p} x(t;\delta^j) = x^1 + \cdots + x^p = x, \tag{5.8}$$

which shows that $x \in \mathcal{R}_{(\{0\},\mathcal{U})}$*. Hence, this proves the inclusion*

$$\mathcal{R}^1_{(\{0\},\mathcal{U})} \oplus \cdots \oplus \mathcal{R}_{(\{0\},\mathcal{U})} \subseteq \mathcal{R}_{(\{0\},\mathcal{U})}. \tag{5.9}$$

Intuitively, any state in the Minkowski sum of the individually reachable sets can be expressed as a sum of states, each induced by an individual input channel. By constructing an input signal that consists of the sum of these individual components, it is ensured that every state in the Minkowski sum of the individually reachable sets is indeed reachable, which is captured in the inclusion in (5.9)*. Combining both inclusions in* (5.7) *and* (5.9) *completes the proof that*

$$\mathcal{R}_{(\{0\},\mathcal{U})} = \mathcal{R}^1_{(\{0\},\mathcal{U})} \oplus \cdots \oplus \mathcal{R}^p_{(\{0\},\mathcal{U})}. \tag{5.10}$$

■

Lemma 3 demonstrates that the origin-reachable can thus be decomposed into the reachable sets induced by the individual inputs δ_j. In other words, the complete origin-reachable set is equivalent the Minkowski sum

$$\mathcal{R}_{(\{0\},\mathcal{U})} = \mathcal{R}^1_{(\{0\},\mathcal{U})} \oplus \cdots \oplus \mathcal{R}^p_{(\{0\},\mathcal{U})}. \tag{5.11}$$

To obtain a hyperrectangular outer approximation of the reachable set that is aligned with the coordinate axes, each state $\hat{x}^j_i(s)$ of the state vector $\hat{x}^j(s)$ is considered individually. The following notation is adopted:

$$\hat{x}^j_i(s) = \mathbf{e}_i^\top G(s)\mathbf{e}_j\, \hat{\delta}_j(s). \tag{5.12}$$

Again, the superscript j specifies that this is the state response induced by input $\delta_j(t)$. Let function $g(t)$ denote the impulse response function corresponding to the transfer function $G(s)$. The function $\mathbf{e}_i^\top g(t)\mathbf{e}_j$ therefore represents the impulse response function from input $\delta_j(t)$ to state $x_i(t)$. The $\mathcal{L}_\infty$-norm of the input $\delta(t)$ and the output $x(t)$ is induced by the $\mathcal{L}_1$-norm of the impulse response function $g(t)$ according to (2.4) such that

$$\left\|x^j_i(t)\right\|_{\mathcal{L}_\infty} \le \left\|\mathbf{e}_i^\top g(t)\mathbf{e}_j\right\|_{\mathcal{L}_1} \left\|\delta_j(t)\right\|_{\mathcal{L}_\infty}. \tag{5.13}$$

Since the set of allowed inputs $\mathcal{U}$ in (3.5) is a hyperrectangle with half-widths $\bar{\delta}_j$, centered in the origin and aligned with the coordinate axes, each $\|\delta_j(t)\|_{\mathcal{L}_\infty} = \bar{\delta}_j$. The dimensions (or half-widths), denoted by $\bar{x}_i$, of the hyperrectangle in (5.1) that outer approximates the reachable set in (5.3) are therefore

$$\bar{x}_i = \left\|\mathbf{e}_i^\top g(t)\mathbf{e}_j\right\|_{\mathcal{L}_1} \bar{\delta}_j \text{ , which implies that } \left\|x^j_i(t)\right\|_{\mathcal{L}_\infty} \le \bar{\delta}^j_i. \tag{5.14}$$

Recall that there always exists an input signal $\delta_j(t)$ for which the equality holds in equation (5.13). Furthermore, by the point symmetry of the input set $\mathcal{U}$ and by linearity of the system, the origin-reachable set $\mathcal{R}^j_{(\{0\},\mathcal{U})}$ is also point symmetric. This results in the conclusion that

$$\mathcal{H}(\mathcal{R}^j_{(\{0\},\mathcal{U})}) \text{ in (5.1) with } \bar{x}_i = \left\|\mathbf{e}_i^\top g(t)\mathbf{e}_j\right\|_{\mathcal{L}_1} \bar{\delta}_j \text{ is tight around } \mathcal{R}^j_{(\{0\},\mathcal{U})}. \tag{5.15}$$

This is an important result, since now an analytical expression is formulated for a *tight* hyperrectangular outer approximation around each individual reachable set $\mathcal{R}^j_{(\{0\},\mathcal{U})}$. Recall from (5.11) that the total reachable set is equivalent to the Minkowski sum of the individual reachable sets, induced by each input δ_j. Since the outer approximation around each individual reachable set $\mathcal{R}^j_{(\{0\},\mathcal{U})}$, denoted by $\mathcal{H}(\mathcal{R}^j_{(\{0\},\mathcal{U})})$, is tight, the Minkowski sum of these hyperrectangles

$$\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})}) = \mathcal{H}(\mathcal{R}^1_{(\{0\},\mathcal{U})}) \oplus \cdots \oplus \mathcal{H}(\mathcal{R}^p_{(\{0\},\mathcal{U})}) \text{ is also tight around } \mathcal{R}_{(\{0\},\mathcal{U})}. \tag{5.16}$$

In other words, the tightness of the approximations is preserved under the Minkowski addition. The corresponding proof of this claim is presented in the following lemma:

Lemma 4 (Preservation of tightness under Minkowksi addition)**.** *Let* $\mathcal{S}_1, \ldots, \mathcal{S}_p$ *be sets in* $\mathbb{R}^n$ *expressed in the coordinates* $x^\top = [x_1, \ldots, x_n]$ *with* $x_i \in \mathbb{R}$*. Let* $\mathcal{H}(\mathcal{S}_j)$ *be the hyperrectangle that is centered in the origin and aligned with the coordinate frame axes, such that this hyperrectangle is an outer approximation of set* $\mathcal{S}_j$*. Each outer approximation is tight such that the projection* $\Pi_{x_i}[\mathcal{H}(\mathcal{S}_j)] = \Pi_{x_i}[\mathcal{S}_j]$ *for all* $i = 1, \ldots, n$ *and* $j = 1, \ldots, p$*. The Minkowski sum of the sets* $\mathcal{S}_j$ *is denoted by* $\mathcal{S} = \mathcal{S}_1 \oplus \cdots \oplus \mathcal{S}_p$*. The Minkowski sum of the corresponding outer approximations* $\mathcal{H}(\mathcal{S}_j)$ *is denoted by* $\mathcal{H}(\mathcal{S}) = \mathcal{H}(\mathcal{S}_1) \oplus \cdots \oplus \mathcal{H}(\mathcal{S}_p)$*. The tightness of the outer approximation is preserved under the Minkowski sum. This means that the Minkowski sum* $\mathcal{H}(\mathcal{S})$ *is also tight around the Minkowski sum* $\mathcal{S}$ *such that*

$$\Pi_{x_i}[\mathcal{H}(\mathcal{S})] = \Pi_{x_i}[\mathcal{S}] \textit{ for all } i = 1, \ldots, n. \tag{5.17}$$

Proof: *Due to linearity of the projection operator, the projection of the Minkowski sum of the original sets $\mathcal{S}$ onto the state x_i satisfies*

$$\Pi_{x_i}[\mathcal{S}] = \sum_{j=1}^{p} \Pi_{x_i}[\mathcal{S}_j] \tag{5.18}$$

for each state x_i. Similarly, for the Minkowski sum of the corresponding hyperrectangular approximations $\mathcal{H}(\mathcal{S})$, the projection onto the state x_i satisfies

$$\Pi_{x_i}[\mathcal{H}(\mathcal{S})] = \sum_{j=1}^{p} \Pi_{x_i}[\mathcal{H}(\mathcal{S}_j)] \tag{5.19}$$

for each state x_i. Since each hyperrectangle $\mathcal{H}(\mathcal{S}_i)$ is tight around the set $\mathcal{S}_i$, their projections satisfy

$$\Pi_{x_i}[\mathcal{H}(\mathcal{S}_j)] = \Pi_{x_i}[\mathcal{S}_j] \tag{5.20}$$

for each j. Summing both sides over j gives that

$$\sum_{j=1}^{p} \Pi_{x_i}[\mathcal{H}(\mathcal{S}_j)] = \sum_{j=1}^{p} \Pi_{x_i}[\mathcal{S}_j], \text{ which implies that } \Pi_{x_i}[\mathcal{H}(\mathcal{S})] = \Pi_{x_i}[\mathcal{S}] \forall i = 1, \dots, n. \tag{5.21}$$

■

Lemma 4 thus implies an important result: an analytical expression can be formulated of a *tight* hyperrectangular outer approximation of the complete origin-reachable set $\mathcal{R}_{(\{0\},\mathcal{U})}$, by taking the sum of the dimensions of the individual origin-reachable sets $\mathcal{R}^j_{(\{0\},\mathcal{U})}$. The proposed method thus yields a hyperrectangular outer approximation that is tight around true reachable set. In other words,

$$\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})}) \text{ in (5.1) with } \bar{x}_i = \sum_{j=1}^{N} \left\| \mathbf{e}_i^\top g(t) \mathbf{e}_j \right\|_{\mathcal{L}_1} \bar{\delta}_j \text{ is tight around origin-reachable set } \mathcal{R}_{(\{0\},\mathcal{U})}. \tag{5.22}$$

This is an important result, since this approximation can be expressed analytically, which is an advantage compared to the ellipsoidal outer approximation. The ellipsoidal outer approximation could only be computed numerically by means of an optimization program. Now that the analytical expression for the hyperrectangular outer approximation of the *origin-reachable set* is formulated, the *autonomous reachable set* is analyzed.

5.3 Approximation of autonomous reachable set

Now, the autonomous reachable set is considered, which is denoted by $\mathcal{R}_{(\mathcal{X}_0,\{0\})}$. This is the reachable set where it is imposed that no inputs are applied to the system such that $\mathcal{U} = \{0\}$. The system dynamics in (3.6) therefore simplify to

$$\dot{x}(t) = A\,x(t). \tag{5.23}$$

There exist several methods to compute this reachable set, each of which have their advantages and limitations. However, this thesis does not focus on these methods. In the remainder of this report, it is therefore assumed that there is a formulation of a hyperrectangular outer approximation of the autonomous reachable set, centered in the origin and aligned with the coordinate frame axes. Furthermore, it is assumed that this approximation tight around the autonomous reachable set. The reachable set $\mathcal{R}_{(\mathcal{X}_0,\{0\})}$ is said to by tightly outer approximated by the hyperrectangle in (5.1) with half-widths

$$\bar{x}_i = \bar{x}_i^{\text{auto}}. \tag{5.24}$$

5.4 Approximation of complete reachable set

The complete reachable set is divided into the *origin-reachable set* and *autonomous reachable* set, both of which have been analyzed separately. Now, the hyperrectangular approximation of the *complete reachable set* $\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})}$ in (3.7) is elaborated. Here, the trajectory may start in any initial condition in the set $\mathcal{X}_0$ and may use any input in the set of allowed inputs $\mathcal{U}$. This reachable set can be outer bounded by the Minkowski sum of the origin- and autonomous reachable set, such that

$$\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})} \subseteq \mathcal{R}_{(\{0\},\mathcal{U})} \oplus \mathcal{R}_{(\mathcal{X}_0,\{0\})}. \tag{5.25}$$

This is also a well known result from reachability analysis in linear systems. For completeness, the corresponding proof is recalled in the following lemma:

Lemma 5 (Minkowski decomposition of reachable set)**.** *Consider the linear time-invariant continuous-time asymptotically stable system in* (3.6) *with state vector x, inputs vector u and system matrices (A, B) and set of allowed inputs $\mathcal{U}$ in* (3.5) *and set of initial conditions $\mathcal{X}_0$. Let the origin-reachable set be denoted by $\mathcal{R}_{(\{0\},\mathcal{U})}$, the autonomous reachable set be denoted by $\mathcal{R}_{(\mathcal{X}_0,\{0\})}$, and the complete reachable set be denoted by $\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})}$. The complete reachable set is included by the Minkowski sum of the origin-reachable set and the autonomous reachable set such that*

$$\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})} \subseteq \mathcal{R}_{(\{0\},\mathcal{U})} \oplus \mathcal{R}_{(\mathcal{X}_0,\{0\})}. \tag{5.26}$$

Proof: *For any state $x \in \mathcal{R}_{(\mathcal{X}_0,\mathcal{U})}$, there exists an initial condition $x_0 \in \mathcal{X}_0$ and an input $\delta(t) \in \mathcal{U}$ for all t such that the trajectory satisfies*

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}B\delta(\tau)d\tau. \tag{5.27}$$

Define the two individual components; $x_{(\mathcal{X}_0,\{0\})}(t) = e^{At}x_0$, which belongs to $\mathcal{R}_{(\mathcal{X}_0,\{0\})}$, and $x_{(\{0\},\mathcal{U})}(t) = \int_0^t e^{A(t-\tau)}B\delta(\tau)d\tau$, which belongs to $\mathcal{R}_{(\{0\},\mathcal{U})}$. Since any reachable state x satisfies

$$x(t) = x_{(\mathcal{X}_0,\{0\})}(t) + x_{(\{0\},\mathcal{U})}(t), \tag{5.28}$$

the complete reachable set is included by the Minkowski sum of the origin-reachable set and the autonomous reachable set such that

$$\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})} \subseteq \mathcal{R}_{(\{0\},\mathcal{U})} \oplus \mathcal{R}_{(\mathcal{X}_0,\{0\})}. \tag{5.29}$$

■

Lemma 5 thus proves the inclusion in (5). Note that the complete reachable set is *included by* and not *equivalent to* the Minkowski sum of the origin-reachable set and the autonomous reachable set. Intuitively, this can be understood by considering the case where the autonomous response decays very fast whereas the response due to an input increases very slowly.

The complete reachable set is thus contained by the Minkowski sum of the origin- and autonomous reachable set. Furthermore, the hyperrectangular approximations of the origin-reachable set and the autonomous reachable set fully contain these reachable sets. Combining this gives that the complete reachable set can be outer bounded such that

$$\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})} \subseteq \mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})}) \oplus \mathcal{H}(\mathcal{R}_{(\mathcal{X}_0,\{0\})}). \tag{5.30}$$

Again, these sets need not be equivalent. Though the approximations around the origin- and autonomous reachable set may be tight, their Minkowski sum need not be tight around the complete reachable set. Therefore, the inequality in (5.30) implies that the hyperrectangular set in (5.1) with half-widths

$$\bar{x}_i = \sum_{j=1}^{N} \left\|\mathbf{e}_i^\top g(t)\mathbf{e}_j\right\|_{\mathcal{L}_1} \bar{\delta}_j + \bar{x}_i^{\text{auto}} \tag{5.31}$$

is an outer approximation of the true reachable set $\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})}$, and that this hyperrectangle need not be tight around the true reachable set.

5.5 Numerical example

To demonstrate the technique to compute a hyperrectangular approximation of the reachable set that has been elaborated, this technique is applied to the two-dimensional system Σ_1 and three-dimensional system Σ_2 in (4.4). Again, the set of initial states of both systems is the origin $\mathcal{X}_0 = \{0\}$, and the set of allowed inputs is $\mathcal{U}$ in (3.5) with $\bar{\delta}_j = 1$ for all $j = 1, 2$ and $j = 1, 2, 3$ for system Σ_1 and Σ_2 respectively. Note that the method to compute the hyperrectangular approximation does not require temporal discretization of the systems.

The computed hyperrectangular approximations of the reachable sets are presented in Figure 5.2 and Figure 5.3 for system Σ_1 and Σ_2 respectively. Since it is imposed that all trajectories of both systems start in the origin (so essentially the origin-reachable set is computed), the result in (5.22) says that these approximations are tight. The sets show that the true reachable set indeed touches its hyperrectangular approximation on each boundary (or facet) as predicted.

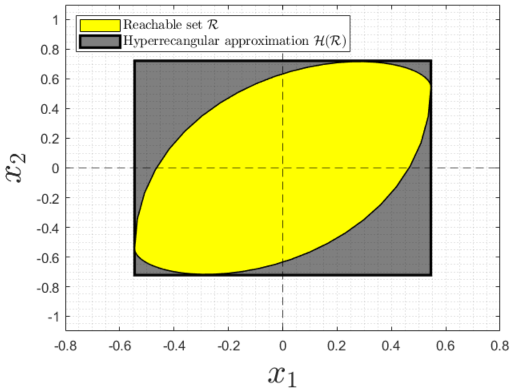

Figure 5.2: Reachable set of the two-dimensional system Σ_1 with its hyperrectangular approximation. The approximation is tight around the reachable set.

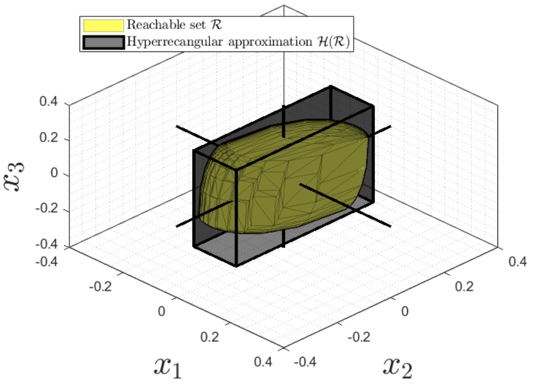

Figure 5.3: Reachable set of the two-dimensional system Σ_1 with its hyperrectangular approximation. The approximation is tight around the reachable set.

In summary, this chapter has presented an alternative to the *ellipsoidal* approximation of the reachable set, in the form of a *hyperrectangular* outer approximation. This alternative addresses the limitations of the ellipsoidal approach, since it does not require temporal discretization of the continuous-time system and can be expressed analytically, without the need of any optimization program. In the next chapter, a dynamical model that captures the behaviour of the complete platoon is developed. In subsequent chapters, this model is used to compute and analyze the hyperrectangular outer approximation of the reachable set of this platoon.

6 Platoon dynamics

The previous chapter has introduced the hyperrectangular outer approximation of the reachable set, as an alternative for the ellipsoidal outer approximation. In this chapter, a dynamical model that captures the behaviour of the complete platoon is developed, extending the two-vehicle model in previous work to multi-vehicle platoons. This is achieved by interconnecting the dynamics of consecutive vehicles to form the cascaded system that is the platoon. The resulting description of the complete vehicle platoon serves as the foundation for further analysis throughout this thesis, as it specifies the dynamics of the complete vehicle platoon in (3.6). The model is formulated in terms of two coordinate frames: the *original coordinates* and the *error coordinates*. The original coordinates offer an intuitive physical interpretation, and are therefore considered when formulating safety guarantees for the platoon. The error coordinates, on the other hand, make it easier to identify some key system properties, and are therefore considered when, e.g., identifying the attackable subspace.

6.1 Dynamics in original coordinates

Consider again the platoon of m vehicles, enumerated with index $i = 1, \dots, m$ with $i = 1$ indicating the lead vehicle (see Figure 3.1). The longitudinal dynamics of each vehicle i are described by (3.1). These dynamics in terms of the inter-vehicular distance $d_i = q_{i-1} - q_i - L_i$ for each vehicle i, except the lead vehicle, are described by

$$\left.\begin{aligned} \dot{d}_i &= v_{i-1} - v_i \\ \dot{v}_i &= a_i \\ \dot{a}_i &= -\frac{1}{\tau}a_i + \frac{1}{\tau}u_i \end{aligned}\right. \quad \text{for all } i = 2, \dots, m. \tag{6.1}$$

To simplify the derivation of the platoon model, for the moment it is assumed that there are no false data injections. Consequently, the dynamics do not depend on the controller realization, and the choice of the sensor configuration β_i is arbitrary. By choosing the realization $\beta_i = \mathbf{0}$, the CACC dynamics can be represented by the dynamical system in (3.4). Note that, since the control input u_i in this realization is equivalent to the internal controller state ξ_i, this internal controller state is redundant and can therefore be omitted. The CACC scheme with controller realization $\beta_i = \mathbf{0}$ can thus be represented as

$$\dot{u}_i = -\frac{1}{h}u_i + \frac{k_p}{h}e_i + \frac{k_d}{h}\dot{e}_i + \frac{1}{h}u_{i-1}. \tag{6.2}$$

The inter-vehicular distance $d_i := q_{i-1} - q_i - L_i$ asymptotically converges to the reference distance $d_{r,i} := r_i + hv_i$ for all vehicles $i = 2, \dots, m$. Consequently, the spacing error $e_i := d_i - d_{r,i} = d_i - r_i - hv_i$ asymptotically converges to zero. The time-derivative of the spacing error is $\dot{e}_i = v_{i-1} - v_i - ha_i$. Substituting these expressions for the errors into the CACC dynamics gives

$$\dot{u}_i = -\frac{1}{h}u_i + \frac{k_p}{h}d_i - \left(k_p + \frac{k_d}{h}\right)v_i - k_d a_i + \frac{k_d}{h}v_{i-1} + \frac{1}{h}u_{i-1} - \frac{k_p}{h}r_i. \tag{6.3}$$

First, the *original coordinates* are defined. In the original coordinate frame, the state vector of the lead vehicle $i = 1$ reads $x_{d,1}^\top = [v_1,\, a_1]$. The state vector of all following vehicles $i = 2, \dots, m$ reads $x_{d,i}^\top = [d_i,\, v_i,\, a_i,\, u_i]$. By interconnecting dynamics of subsequent vehicles, consisting of the longitudinal vehicle model in (6.1) and the CACC scheme in (6.3), a dynamical model of the complete vehicle platoon can be formulated. The dynamics of this model in terms of the original coordinates read

$$\underbrace{\begin{bmatrix} \dot{x}_{d,1} \\ \dot{x}_{d,2} \\ \dot{x}_{d,3} \\ \vdots \\ \dot{x}_{d,m} \end{bmatrix}}_{\dot{x}_d} = \underbrace{\begin{bmatrix} A_d^1 & 0 & 0 & \dots & 0 \\ \tilde{A}_d^1 & A_d^2 & 0 & \dots & 0 \\ 0 & \tilde{A}_d^2 & A_d^3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & \tilde{A}_d^{m-1} & A_d^m \end{bmatrix}}_{A_d} \underbrace{\begin{bmatrix} x_{d,1} \\ x_{d,2} \\ x_{d,3} \\ \vdots \\ x_{d,m} \end{bmatrix}}_{x_d} + \underbrace{\begin{bmatrix} B_d^1 \\ B_d^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{B_d^u} u_1 + \underbrace{\begin{bmatrix} 0 \\ r_d^2 \\ r_d^3 \\ \vdots \\ r_d^m \end{bmatrix}}_{r_d}, \tag{6.4}$$

or more compactly $\dot{x}_d = A_d\, x_d + B_d^u\, u_1 + r_d$. Here, the matrices B_d^1, A_d^1 and $\tilde{A}_d^1$—characterizing the dynamics of the lead vehicle—respectively read

$$A_d^1 = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{1}{\tau} \end{bmatrix}, \; B_d^1 = \begin{bmatrix} 0 \\ \frac{1}{\tau} \end{bmatrix}, \text{ and } \tilde{A}_d^1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ \frac{k_d}{h} & 0 \end{bmatrix}. \tag{6.5}$$

The matrices B_d^2, A_d^i and $\tilde{A}_d^i$, that characterize the dynamics of the follower vehicles, and the vector r_d^i, that induces that each vehicle converges to its reference inter-vehicular distance, respectively read

$$A_d^i = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{\tau} & \frac{1}{\tau} \\ \frac{k_p}{h} & -\left(k_p + \frac{k_d}{h}\right) & -k_d & -\frac{1}{h} \end{bmatrix}, \tilde{A}_d^i = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{k_d}{h} & 0 & \frac{1}{h} \end{bmatrix}, B_d^2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{h} \end{bmatrix} \text{ and } r_d^i = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -\frac{k_p}{h} r \end{bmatrix}. \tag{6.6}$$

Now, it is assumed that there are false data injections into the sensors of vehicle 2, which are denoted by δ_2. Consequently, the model describing the dynamics of the complete vehicle platoon in (6.4) is extended such that it takes these false data injections into account. This results in the model

$$\dot{x}_d = A_d\, x_d + B_d^u\, u_1 + r_d + B_d^\delta(\beta_2)\, \delta_2. \tag{6.7}$$

Here, matrix $B_d^\delta(\beta_2)$ describes how false data injection enter the dynamics of the platoon. The analytical expression of matrix $B_d^\delta(\beta_2)$ is derived in [28]. This matrix is affine in the controller realization β_2 and reads

$$B_d^\delta(\beta_2) = \begin{bmatrix} 0 \\ B_{d,\delta}^2(\beta_2) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ with } B_{d,\delta}^2(\beta_2) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\beta_{2,1}}{\tau} & -\frac{\beta_{2,2}}{\tau} & -\frac{\beta_{2,3}}{\tau} & -\frac{\beta_{2,4}}{\tau} & -\frac{\beta_{2,5}}{\tau} & 0 \\ \frac{\beta_{2,1}+k_p}{h} & \frac{\beta_{2,2}}{h} - k_p & \beta_{2,2} - \beta_{2,4} - k_d + \beta_{2,3}\left(\frac{1}{h} - \frac{1}{\tau}\right) & \beta_{2,1} + \frac{\beta_{2,4}+k_d}{h} & \beta_{2,4} + \beta_{2,5}\left(\frac{1}{h} - \frac{1}{\tau}\right) & \frac{\beta_{2,5}}{\tau} + \frac{1}{h} \end{bmatrix}. \tag{6.8}$$

Note that the control input of the lead vehicle, denoted by u_1, is still an external input in the dynamical model. Typically, the control input of the lead vehicle u_1 is computed by a cruise controller that asymptotically drives its velocity v_1 to some reference velocity v_r. Consequently, the stacked state x_d vector does not converge to the origin, but to some non-zero equilibrium state denoted by x_{eq}. This equilibrium state reads

$$x_{\text{eq}}^\top = \left[\begin{array}{cc|cccc|cccc|c|cccc} v_r & 0 & r + hv_r & v_r & 0 & 0 & r + hv_r & v_r & 0 & 0 & \dots & r + hv_r & v_r & 0 & 0 \end{array}\right]. \tag{6.9}$$

See Appendix A.2 for the derivation of this equilibrium state. This means that the synchronized state of the platoon—which is the state that the system asymptotically converges to—is not the origin but the non-zero equilibrium x_{eq}. The techniques that are used in this thesis to compute the outer approximation of the reachable set, however, assume that the system asymptotically converges to the origin. To this end, the new state vector $\tilde{x}_d$ is defined by means of a coordinate translation $\tilde{x}_d = x_d - x_{\text{eq}}$. The new coordinates are called the *shifted* original coordinates. In the shifted original coordinates, the state vector of each vehicle is

$$\begin{aligned} \tilde{x}_{d,1}^\top &= [\tilde{v}_1,\, a_1] \quad \text{for the lead vehicle } i = 1 \text{ and} \\ \tilde{x}_{d,i}^\top &= [\tilde{d}_i,\, \tilde{v}_i,\, a_i,\, u_i] \quad \text{for all following vehicles } i = 2, \dots, m. \end{aligned} \tag{6.10}$$

The translated inter-vehicular distance $\tilde{d}_i = d_i - (r_i + hv_r)$ is the deviation of the inter-vehicular distance d_i with respect to its equilibrium intervehicle distance $r_i + hv_r$. Similarly, the translated velocity $\tilde{v}_i = v_i - v_r$ is simply the deviation of the velocity v_i with respect to the equilibrium velocity v_r. The acceleration a_i and control input u_i remain equivalent in both coordinate frames. By means of the coordinate translation $\tilde{x}_d = x_d - x_{\text{eq}}$, the dynamical system in (6.7) is written in terms of the shifted original coordinates $\tilde{x}_d$. This system thus converges asymptotically to the origin and its dynamics are described by

$$\underbrace{\begin{bmatrix} \dot{\tilde{x}}_{d,1} \\ \dot{\tilde{x}}_{d,2} \\ \dot{\tilde{x}}_{d,3} \\ \vdots \\ \dot{\tilde{x}}_{d,m} \end{bmatrix}}_{\dot{\tilde{x}}_d} = \underbrace{\begin{bmatrix} A_d^1 & 0 & 0 & \dots & 0 \\ \tilde{A}_d^1 & A_d^2 & 0 & \dots & 0 \\ 0 & \tilde{A}_d^2 & A_d^3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & \tilde{A}_d^{m-1} & A_d^m \end{bmatrix}}_{A_d} \underbrace{\begin{bmatrix} \tilde{x}_{d,1} \\ \tilde{x}_{d,2} \\ \tilde{x}_{d,3} \\ \vdots \\ \tilde{x}_{d,m} \end{bmatrix}}_{\tilde{x}_d} + \underbrace{\begin{bmatrix} B_d^1 \\ B_d^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{B_d^u} u_1 + \underbrace{\begin{bmatrix} 0 \\ B_{d,\delta}^2(\beta_2) \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{B_d^\delta(\beta_2)} \delta_2. \tag{6.11}$$

The chain-like structure in matrix A_d characterizes the cascaded nature of the system. Note that the control input of vehicle 1, denoted by u_1, enters only the state vector of vehicle 1 and 2 directly through the input matrix B_d^u. Similarly, the false data injections δ_2 into the sensor signals of vehicle 2 only enter the state vector of vehicle 2 through input matrix $B_d^\delta(\beta_2)$.

The original coordinates offer an intuitive physical interpretation, which helps with the formulation of safety guarantees. However, when identifying system properties of the platoon, alternative coordinate frames are more suitable.

6.2 Dynamics in error coordinates

To facilitate the identification of some key system properties of the platoon, the error coordinates are introduced. In the error coordinate frame, denoted by x_e, the state vector for each vehicle in the platoon reads

$$\begin{aligned} \tilde{x}_{e,1}^{\top} &= [\tilde{v}_1,\, a_1] \quad \text{for the lead vehicle } i = 1 \text{ and} \\ \tilde{x}_{e,i}^{\top} &= [e_i,\, \dot{e}_i,\, \ddot{e}_i,\, u_i] \quad \text{for all following vehicles } i = 2, \dots, m. \end{aligned} \tag{6.12}$$

In the error coordinates, the state vector of all following vehicles thus includes the spacing error e_i, its time-derivative $\dot{e}_i$ and its double time-derivative $\ddot{e}_i = a_{i-1} + \left(\frac{h}{\tau} - 1\right) a_i - \frac{h}{\tau} u_i$. The relation between the complete state vector in error coordinates x_e and in original coordinates x_d is thus described by

$$\underbrace{\begin{bmatrix} x_{e,1} \\ x_{e,2} \\ x_{e,3} \\ \vdots \\ x_{e,m} \end{bmatrix}}_{x_e} = \underbrace{\begin{bmatrix} T^1 & 0 & 0 & \dots & 0 \\ \tilde{T}^1 & T^2 & 0 & \dots & 0 \\ 0 & \tilde{T}^2 & T^3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & T^m \end{bmatrix}}_{T} \underbrace{\begin{bmatrix} x_{d,1} \\ x_{d,2} \\ x_{d,3} \\ \vdots \\ x_{d,m} \end{bmatrix}}_{x_d} + \underbrace{\begin{bmatrix} 0 \\ r_t^2 \\ r_t^3 \\ \vdots \\ r_t^m \end{bmatrix}}_{r_t}, \tag{6.13}$$

or more compactly $x_e = T\, x_d + r_t$. The matrix T consists of the individual components T^1, $\tilde{T}^1$, T^i and $\tilde{T}^i$, and the column vector r_t consists of the individual components r_t^i. These components respectively read

$$T^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ \tilde{T}_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},\ T^i = \begin{bmatrix} 1 & -h & 0 & 0 \\ 0 & -1 & -h & 0 \\ 0 & 0 & \left(\frac{h}{\tau} - 1\right) & -\frac{h}{\tau} \\ 0 & 0 & 0 & 1 \end{bmatrix},\ \tilde{T}^i = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } r_t^i = \begin{bmatrix} -r \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{6.14}$$

The coordinate transformation in (6.13) is employed to transform the dynamics of the complete platoon from the original coordinates x_d in (6.7) to the error coordinates x_e. The resulting model—in terms of the error coordinates x_e—asymptotically converges to the origin, and its corresponding dynamics are described by

$$\underbrace{\begin{bmatrix} \dot{x}_{e,1} \\ \dot{x}_{e,2} \\ \dot{x}_{e,3} \\ \vdots \\ \dot{x}_{e,m} \end{bmatrix}}_{\dot{x}_e} = \underbrace{\begin{bmatrix} A_e^1 & 0 & 0 & \dots & 0 \\ \tilde{A}_e^1 & A_e^2 & 0 & \dots & 0 \\ 0 & \tilde{A}_e^2 & A_e^3 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \tilde{A}_e^{m-1} & A_e^m \end{bmatrix}}_{A_e} \underbrace{\begin{bmatrix} x_{e,1} \\ x_{e,2} \\ x_{e,3} \\ \vdots \\ x_{e,m} \end{bmatrix}}_{x_e} + \underbrace{\begin{bmatrix} B_e^1 \\ B_e^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{A_e^u} u_1 + \underbrace{\begin{bmatrix} 0 \\ B_e^{\delta,2}(\beta_2) \\ B_e^{\delta,3}(\beta_2) \\ \vdots \\ 0 \end{bmatrix}}_{B_e^{\delta}(\beta_2)} \delta_2, \tag{6.15}$$

or more compactly by $\dot{x}_e = A_e\, x_e + B_e^u\, u_1 + B_e^{\delta}(\beta_2)\, \delta_2$. See Appendix A.3 for the derivation of this model, which is done using the coordinate transformation $x_e = T x_d + r_t$. The matrices B_e^1, A_e^1 and $\tilde{A}_e^1$—characterizing the dynamics of the lead vehicle—respectively read

$$A_e^1 = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{1}{\tau} \end{bmatrix},\ B_e^1 = \begin{bmatrix} 0 \\ \frac{1}{\tau} \end{bmatrix},\ \text{and } \tilde{A}_e^1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{6.16}$$

The matrices B_e^2, A_e^i and $\tilde{A}_e^i$—characterizing the dynamics of the follower vehicles— respectively read

$$A_e^i = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\frac{k_p}{\tau} & -\frac{k_d}{\tau} & -\frac{1}{\tau} & 0 \\ \frac{k_p}{h} & \frac{k_d}{h} & 0 & -\frac{1}{h} \end{bmatrix},\ \tilde{A}_e^i = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{h} \end{bmatrix} \text{ and } B_e^2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{h} \end{bmatrix}. \tag{6.17}$$

The matrices $B_{e,\delta}^2$ and $B_{e,\delta}^3$ in input matrix B_e^{δ} describe how the false data injections enter the dynamics of vehicle 2 and 3 respectively. These matrices are also affine in the controller realization β_2 and read

$$\begin{aligned} B_{e,\delta}^2(\beta_2) &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\beta_{2,1} h}{\tau} & \frac{\beta_{2,2} h}{\tau} & \frac{\beta_{2,3} h}{\tau} & \frac{\beta_{2,4} h}{\tau} & \frac{\beta_{2,5} h}{\tau} & 0 \\ -\frac{\beta_{2,1} h + k_p \tau}{\tau^2} & -\frac{h(\beta_{2,2} - k_p \tau)}{\tau^2} & \frac{h(\beta_{2,4} - \beta_{2,2} + k_d)}{\tau} & -\frac{\beta_{2,4} h + \tau(k_d + \beta_{2,1} h)}{\tau^2} & -\frac{\beta_{2,4} h}{\tau} & -\frac{\tau + \beta_{2,5} h}{\tau^2} \\ \frac{\beta_{2,1} + k_p}{h} & \frac{\beta_{2,2}}{h} - k_p & \beta_{2,2} - \beta_{2,4} - k_d + \beta_{2,3}\left(\frac{1}{h} - \frac{1}{\tau}\right) & \beta_{2,1} + \frac{\beta_{2,4} + k_d}{h} & \beta_{2,4} + \beta_{2,5}\left(\frac{1}{h} - \frac{1}{\tau}\right) & \frac{\beta_{2,5}}{\tau} + \frac{1}{h} \end{bmatrix} \\ \text{and } B_{e,\delta}^3(\beta_2) &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\beta_{2,1}}{\tau} & -\frac{\beta_{2,2}}{\tau} & -\frac{\beta_{2,3}}{\tau} & -\frac{\beta_{2,4}}{\tau} & -\frac{\beta_{2,5}}{\tau} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned} \tag{6.18}$$

Again, the chain-like structure in the system matrix A_e in (6.15) characterizes the cascaded nature of the system. Note that in the error coordinates x_e, the false data injections into the sensor signals of vehicle 2 not only directly affect the dynamics of vehicle 2, but also of vehicle 3 through input matrix $B_e^{\delta}(\beta_2)$.

In summary, this chapter has developed a dynamical model that captures the behaviour of the complete vehicle platoon. The model is formulated in terms of the original coordinates and the error coordinates. The original coordinates offer an intuitive physical interpretation. The error coordinates, however, make it easier to identify some key system parameters. This is demonstrated in the next chapter, where the platoon is analyzed in terms of the error coordinates to evaluate controllability. This helps to identify the attackable subspace. This attackable subspace is then considered in terms of the original coordinates to facilitate physical interpretation of the subspace.

7 Attackable subspace

In the previous chapter, the dynamics of the complete vehicle platoon have been formulated in terms of both the original coordinates and the error coordinates. In this chapter, the vulnerability of the platoon to attacks is analyzed by identifying the so-called attackable subspace. This is the subspace spanned by the directions where the state can be driven toward induced by attacks, *regardless of the attack magnitude*. This is therefore a qualitative measure of robustness. First, a formal definition of the attackable subspace is presented. Subsequently, the attackable subspace is identified for the platoon in terms of the error coordinates. This attackable subspace is then transformed to the original coordinates for physical interpretation. The results are demonstrated in both coordinate frames through simulation.

7.1 Formal definition

Consider the linear time-invariant continuous-time system asymptotically stable in (3.6), with state vector x and input vector δ, whose dynamics are described by the system matrices (A, B). The *attackable subspace* is the subspace of states that can be affected by its input (i.e., false data injections). This concept is closely related to the *controllable* or *reachable* subspace in linear system theory. However, as our objective is not to control the system but to characterize where attackers can steer the system, the term attackable subspace is adopted instead. Formally, the attackable subspace of the asymptotically stable system is defined as

$$\mathcal{X}^{\text{att}} := \left\{ \, x \in \mathbb{R}^l \;\middle|\; \exists\, \delta(.) \text{ such that } x(t) = \int_0^t e^{A(t-\tau)} B\delta(\tau)\, d\tau \text{ for some } t \geq 0 \, \right\} \tag{7.1}$$

with $l \leq n$ such that $\mathcal{X}^{att} \subseteq \mathbb{R}^n$. The attackable subspace is thus a qualitative measure of robustness of the platoon; there are some directions where the state of the platoon cannot be steered, regardless of the attack magnitude. These directions are inherent to the system, and do not depend on the magnitude of the attacks. Since the system is asymptotically stable, all trajectories converge to this attackable subspace, regardless of the applied input $\delta(t)$. Naturally, when the system is fully controllable, the attackable subspace is simply the entire state space. However, if it is not controllable, the attackable subspace is a lower dimensional subspace of $\mathbb{R}^n$.

7.2 Attackable subspace in error coordinates

The attackable subspace is identified by evaluating the platoon in terms of the error coordinates x_e. Evaluating the platoon in these coordinates allows for an intuitive evaluation. Since the attackable subspace only considers the influence of the false data injections δ_2 on the follower vehicles, the control input of the lead vehicle u_1 is set to zero. This input u_1 and the state vector of the lead vehicle $x_{e,1}$—which remains unaffected by the attacks—are therefore removed from the system in (6.15) to obtain

$$\underbrace{\begin{bmatrix} \dot{x}_{e,2} \\ \dot{x}_{e,3} \\ \dot{x}_{e,4} \\ \vdots \\ \dot{x}_{e,m} \end{bmatrix}}_{\dot{x}_e} = \underbrace{\begin{bmatrix} A_e^2 & 0 & 0 & \dots & 0 \\ \tilde{A}_e^2 & A_e^3 & 0 & \dots & 0 \\ 0 & \tilde{A}_e^3 & A_e^4 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & \tilde{A}_e^{m-1} & A_e^m \end{bmatrix}}_{A_e} \underbrace{\begin{bmatrix} x_{e,2} \\ x_{e,3} \\ x_{e,4} \\ \vdots \\ x_{e,m} \end{bmatrix}}_{x_e} + \underbrace{\begin{bmatrix} B_e^{\delta,2}(\beta_2) \\ B_e^{\delta,3}(\beta_2) \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{B_e^{\delta}(\beta_2)} \delta_2. \tag{7.2}$$

First, it is checked whether this system is controllable, since full controllability would imply that the complete state space $\mathbb{R}^n$ is attackable. Typically, some test is used, e.g., checking the controllability matrix or the Popov-Belevitch-Hautus (PBH) lemma. In this particular case, however, closer inspection of the dynamics reveals that certain states are clearly uncontrollable. This is evident from the structure of the system matrix A_e, which consists of the matrices $\tilde{A}_e^i$ and A_e^i. For each vehicle $i = 2, \dots, m$, the system matrices corresponding to this vehicle are

$$\tilde{A}_e^{i-1} = \left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \frac{1}{h} \end{array}\right], \; A_e^i = \left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\frac{k_p}{\tau} & -\frac{k_d}{\tau} & -\frac{1}{\tau} & 0 \\ \hline \frac{k_p}{h} & \frac{k_d}{h} & 0 & -\frac{1}{h} \end{array}\right] \text{ with state } x_{e,i} = \left[\begin{array}{c} e_i \\ \dot{e}_i \\ \ddot{e}_i \\ \hline u_i \end{array}\right]. \tag{7.3}$$

Matrix $\tilde{A}_e^{i-1}$ describes how the state vector of the preceding vehicle $i-1$ enters the dynamics of vehicle i. This matrix clearly shows that only the control input u_{i-1} affects vehicle i, or more specifically, the control input of vehicle i, denoted by u_i. Moreover, matrix A_e^i shows that the dynamics of the error coordinates of vehicle $[x_{e,i}, \dot{x}_{e,i}, \ddot{x}_{e,i}]$ remain unaffected by the control input u_{i-1} and u_i. Note that this is only the case for vehicles $i = 4, \ldots, m$, since vehicle 2 and 3 can still be influenced by attacks δ_2 through input matrix B_e^δ. In other words, the error dynamics of the states $[x_{e,i}, \dot{x}_{e,i}, \ddot{x}_{e,i}]$ for each vehicle $i \geq 4$ are autonomous, and can thus not be affected by false data injections δ_2.

Determining the controllability of the other states in the complete state vector x_e is more challenging. Their controllability must be verified numerically using, e.g., the PBH lemma. This is not very insightful, especially since this controllability can change depending on the vehicle parameters, platoon length m and controller realization β_2. Since this thesis aims to formulate safety guarantees, the "worst case" is considered by assuming that all other states of the state vector x_e are controllable with respect to the false data injections. In other words, all states except $[x_{e,i}, \dot{x}_{e,i}, \ddot{x}_{e,i}]$ for vehicles $i = 4, \ldots, m$ are assumed to be controllable with respect to the false data injections δ_2. This is a conservative assumption, since some states could become uncontrollable depending on the system parameters. Since the state vector without the errors $[x_{e,i}, \dot{x}_{e,i}, \ddot{x}_{e,i}]$ for vehicles $i = 4, \ldots, m$ is $(m + 5)$-dimensional, the attackable subspace is thus $(m + 5)$-dimensional.

Consider now again the model of the platoon in terms of the error coordinates x_e *with the lead vehicle included*, as introduced in (6.15). By saying that all states of this state vector x_e, except the errors $[x_{e,i}, \dot{x}_{e,i}, \ddot{x}_{e,i}]$ with $i \geq 4$, can be affected by attacks δ_2, the attackable subspace in terms of the error coordinates can thus be formulated as

$$\mathcal{X}_e^{\text{att}} = \text{Im}[V^e] \text{ with } V^e = \left[\begin{array}{cccc|cccc|c|c|c|c} \mathbf{e}_3 & \mathbf{e}_4 & \mathbf{e}_5 & \mathbf{e}_6 & \mathbf{e}_7 & \mathbf{e}_8 & \mathbf{e}_9 & \mathbf{e}_{10} & \mathbf{e}_{14} & \mathbf{e}_{18} & \ldots & \mathbf{e}_{4m-2} \end{array} \right]. \tag{7.4}$$

Simulation in error coordinates:

To demonstrate the attackable subspace in terms of the error coordinates, the platooning dynamics in (6.15) are simulated. The lead vehicle uses a dynamic cruise controller of the form $\dot{u}_1 = -k_v\,(v_1 - v_r) - k_a\,a_1 - k_u\,u_1$ with feedback gains $k_v, k_a, k_u > 0$. A platoon consisting of $m = 4$ vehicles is considered. Each vehicle has a time constant $\tau = 0.1$ [s]. The CACC of each vehicle uses a time gap $h = 0.5$ [s] with feedback gains $k_p = 0.2$ and $k_d = 0.7$ and a controller realization $\beta_2^\top = [0.5, 0.5, 0.5, 0.5, 0.5, 0]$. The platoon starts in synchronization, except for vehicle 4, which has an initial spacing error $e_4(0) = 3$ [m]. The sensor signals of vehicle 2 are disturbed with the false data injections

$$\begin{aligned} &\delta_{2,1}(t) := 0.5\sin(0.5t), && \delta_{2,2}(t) := 0.5\cos(t), && \delta_{2,3}(t) := 0.2e^{-0.1t}, \\ &\delta_{2,4}(t) := 0.2\,\text{sgn}[\cos(0.3t)], && \delta_{2,5}(t) := 0.2, && \delta_{2,6}(t) := 0.1w(t), \end{aligned} \tag{7.5}$$

where $w(t)$ is zero-mean white noise with a standard deviation of 1. The resulting spacing error e_i, its time-derivative $\dot{e}_i$ and double time-derivative $\ddot{e}_i$ of vehicle $i = 2, 3, 4$ are visualized in Figure 7.1. The errors do not converge to zero for vehicle 2 and 3, indicating that attackers can indeed influence the dynamics of these vehicles. However, the error coordinates of vehicle 3 converge to zero and remain unaffected by the attacks. This supports the claim that the error states of the following vehicles have autonomous dynamics, as established from the notation in (7.3).

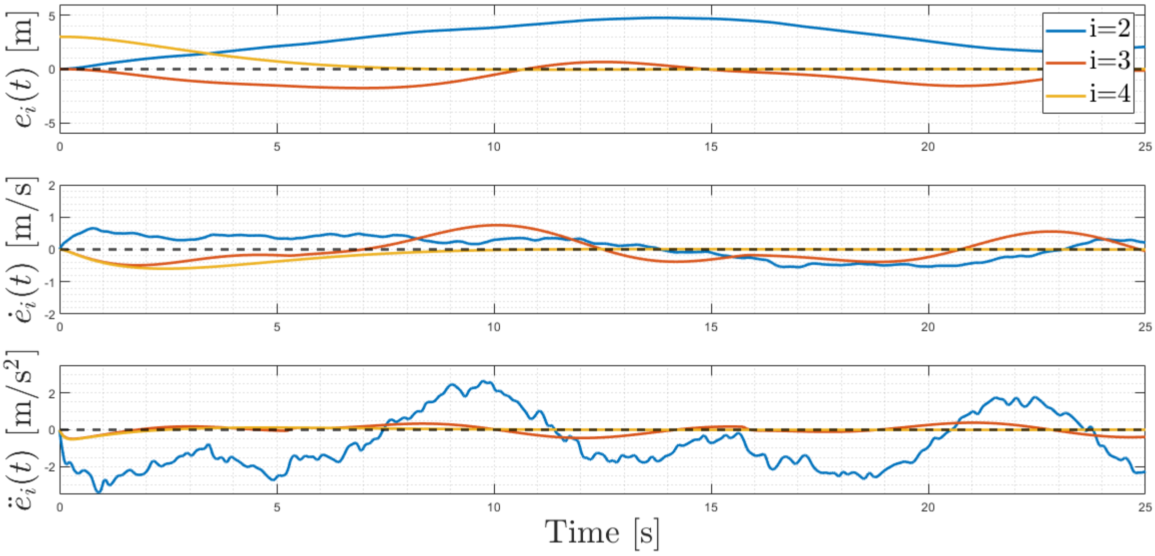

Figure 7.1: Visualization of the trajectory of the platoon in terms of the error coordinates when disturbed by the false data injections in (7.5), demonstrating that the error coordinates of vehicle 4 remain unaffected by the attacks.

In conclusion, writing the platooning in terms of the error coordinates defined in (6.12) decouples its dynamics. The error states e_i, $\dot{e}_i$ and $\ddot{e}_i$ behave autonomously and are not influenced by the false data injections. Though this does help with identifying the attackable subspace, the physical interpretation of these coordinates is not intuitive. To this end, the attackable subspace is transformed to and evaluated in terms of the original coordinates.

7.3 Attackable subspace in original coordinates

The attackable subspace has been identified in terms of the error coordinates x_e. Since these coordinates lack intuitive physical interpretation, the attackable subspace is transformed to the original coordinates x_d. To this end, the coordinate transformation in (6.13) is used, yielding

$$\mathcal{X}_d^{\text{att}} = \text{Im}[T^{-1}\, V^e] + r_t = \text{Im}[V^d] + r_t. \tag{7.6}$$

To illustrate matrix V^d, its symbolic expression for a platoon of $m = 4$ vehicles is provided in Appendix A.4. Due to the high dimensionality of the space $\mathcal{X}_d^{\text{att}}$, it is projected onto the individual vehicle states (d_i, v_i, a_i) of each vehicle $i = 2, 3, \ldots, m$. This projected space is denoted by $\mathcal{X}_i^{\text{att}}$. The attackable space projected onto the states of vehicle 2 and 3 are

$$\mathcal{X}_2^{\text{att}} = \mathcal{X}_3^{\text{att}} = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}. \tag{7.7}$$

These subspaces are three-dimensional. In other words, false data injections can steer the states of vehicle 2 ans 3 anywhere in their three-dimensional space. The attackable space projected onto the states of all following vehicles $i = 4, 5, \ldots, m$ is given by

$$\mathcal{X}_i^{\text{att}} = \text{span}\left\{ \begin{bmatrix} -h \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} + \begin{bmatrix} r_i + h\, v_r \\ v_r \\ 0 \end{bmatrix}. \tag{7.8}$$

In contrast to $\mathcal{X}_2^{\text{att}}$ and $\mathcal{X}_3^{\text{att}}$, the subspace $\mathcal{X}_i^{\text{att}}$ for all following vehicles is thus two-dimensional. Therefore, the state of these vehicles cannot be steered anywhere by false data injections. Instead, the state is constrained to the attackable subspace $\mathcal{X}_i^{\text{att}}$.

Simulation in original coordinates:

To demonstrate the attackable subspace in terms of the original coordinates, the platoon is simulated using the same system parameters and false data injections as before, only now the platoon is initially completely synchronized. The resulting shifted inter-vehicular distance $\tilde{d}_i$, shifted velocity $\tilde{v}_i$ and acceleration a_i are visualized in Figure 7.2.

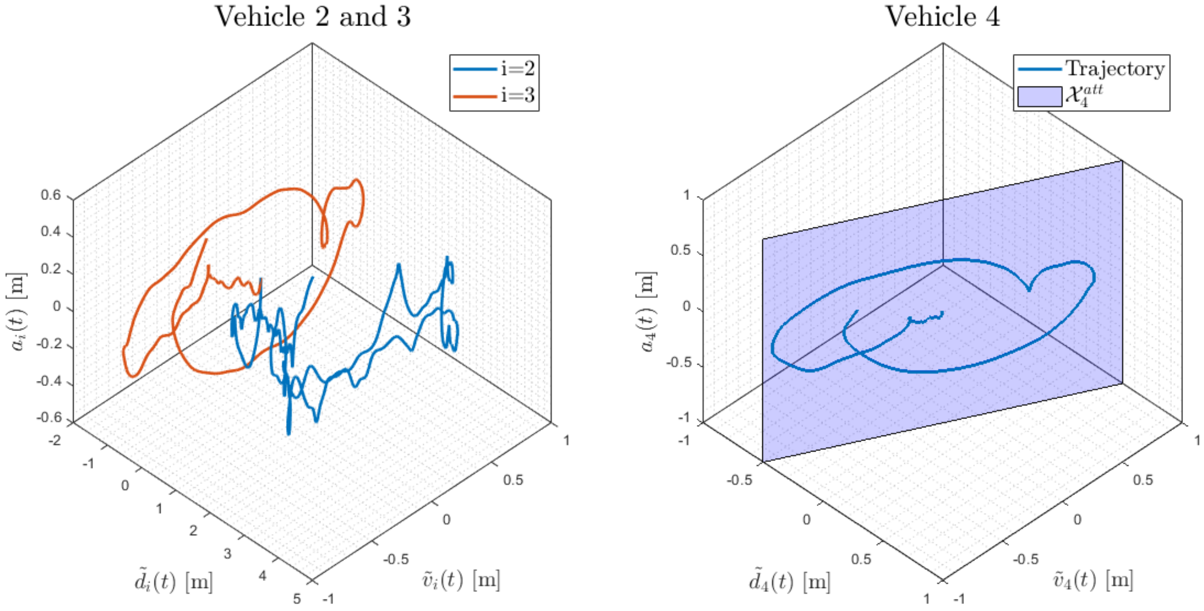

Figure 7.2: Visualization of the trajectory of the platoon in terms of the shifted original coordinates in (6.10) when disturbed by the false data injections in (7.5). This demonstrates that the trajectory of vehicle 2 and 3 moves freely in the three-dimensional space (left), whereas the trajectory of vehicle 4 remain in the attackable subspace (right).

It is observed that the trajectory of vehicle 2 and 3 indeed moves in the complete three-dimensional space $(\tilde{d}_i, \tilde{v}_i, a_i)$. However, the trajectory of vehicle 4 remains in the two-dimensional projected subspace $\mathcal{X}_4^{\text{att}}$ as predicted. This supports the claim that the state of all vehicles $i = 4, \ldots, m$ are constrained to the attackable subspace.

In summary, this chapter has defined the attackable subspace as a qualitative measure of robustness, and identified this attackable subspace for the complete vehicle platoon. The next chapter evaluates reachable set of the vehicle platoon, by evaluation this set's hyperrectangular outer approximation. Based on this approximation, some claims regarding the propagation of the reachable set along vehicles in the platoon are formulated. These claims are formalized by introducing a new notion of string stability; the so-called $\mathcal{L}_\infty$- q string stability.

8 String stability under attacks

The previous chapter has introduced and evaluated the attackable subspace of the platoon. This chapter considers the reachable set of the vehicle platoon, induced by bounded false data injections, which is analyzed by computing the hyperrectangular outer approximation of this reachable set. This reachable set is formulated in terms of the original coordinates, since these coordinates represent intuitive physical states. Some claims are formulated regarding the propagation of the reachable set through the vehicles in the platoon. These claims are formalized by introducing a new notion of string stability that regards the reachable set, called $\mathcal{L}_\infty$-q *string stability.* Intuitively, this notion says that the platoon is $\mathcal{L}_\infty$-q string stable when the reachable set of vehicles stops growing along the platoon, starting at the q^{th} follower vehicle. It is then demonstrated that the origin-reachable set of the platoon under consideration is guaranteed to be $\mathcal{L}_\infty$-q string stable with $q \leq 3$, regardless of the sensor configuration β_2. Finally, the string stability of the platoon's reachable set is demonstrated through a numerical example, and the influence of some key system parameters is evaluated. The results in this chapter form the foundation for the controller synthesis framework developed in the next chapter.

8.1 Formal definition

Consider the reachable set of a vehicle platoon, induced by bounded false data injection attacks. Though there exist many notions of string stability, none of these notions addresses how the reachable set evolves along vehicles in the platoon. To this end, the following notion of $\mathcal{L}_\infty$-q string stability is introduced:

Definition 1 ($\mathcal{L}_\infty$-q string stability)**.** *Consider the cascaded dynamical system in the form* (3.6) *with stacked state vector* $x^\top = \begin{bmatrix} x_1^\top \; x_2^\top \; \dots \; x_m^\top \end{bmatrix} \in \mathbb{R}^n$*, where each* $x_i \in \mathbb{R}^{n_i}$ *denotes the individual state of subsystem* i*. Let* $\mathcal{R}$ *denote the system's reachable set. The reachable set* $\mathcal{R}$ *is said to be* $\mathcal{L}_\infty$-q *string stable if it has a finite volume and if, for any number of subsystems* m*, there exists a finite index* q *such that the projections satisfy the inclusion*

$$\bigcup_{i=1}^{q} \Pi_{x_i}[\mathcal{R}] \supseteq \Pi_{x_{q+1}}[\mathcal{R}] \supseteq \cdots \supseteq \Pi_{x_m}[\mathcal{R}]. \tag{8.1}$$

Similarly, a hyperrectangular approximation of this reachable set, denoted by $\mathcal{H}(\mathcal{R})$*, is* $\mathcal{L}_\infty$-q *string stable if it has a finite volume and if, for any number of vehicles* m*, there exists a finite index* q *such that the projections satisfy the inclusion*

$$\bigcup_{i=1}^{q} \Pi_{x_i}[\mathcal{H}(\mathcal{R})] \supseteq \Pi_{x_{q+1}}[\mathcal{H}(\mathcal{R})] \supseteq \cdots \supseteq \Pi_{x_m}[\mathcal{H}(\mathcal{R})]. \tag{8.2}$$

Intuitively, $\mathcal{L}_\infty$-q string stability ensures that the reachable set of each subsystem downstream from index q is contained within the union of the reachable sets of the first q subsystems. That is, from subsystem $q+1$ onward, the reachable set no longer expands, and can even start shrinking. This behaviour suggests that disturbances become increasingly attenuated throughout the cascade.

Definition 1 implies that if the reachable set is $\mathcal{L}_\infty$-q stable, then it is also $\mathcal{L}_\infty$-q' stable for all $q' > q$. In practice, the smallest such q is pursued, since this index q indicates the earliest subsystem at from which the reachable set of downstream subsystems no longer expands. A special and desirable case is therefore $\mathcal{L}_\infty$-1 string stability where the reachable set of subsystems monotonically decreases along the entire platoon such that

$$\Pi_{x_1}[\mathcal{H}(\mathcal{R})] \supseteq \Pi_{x_2}[\mathcal{H}(\mathcal{R})] \supseteq \cdots \supseteq \Pi_{x_m}[\mathcal{H}(\mathcal{R})]. \tag{8.3}$$

The notion of $\mathcal{L}_\infty$-q string stability is related to the notion of $\mathcal{L}_p$ string stability stability introduced in [33], since both approaches consider how the $\mathcal{L}_p$-norm of the state x_i evolves through the platoon. However, in $\mathcal{L}_p$ string stability, the $\mathcal{L}_p$-norm of the state vector x_i of subsystem i is considered. This is in contrast to the definition in this thesis, which considers the $\mathcal{L}_\infty$-norm of each state in the state vector x_i individually, enabling the connection to the reachable set.

8.2 Structure of platoon dynamics

To investigate whether the hyperrectangular approximation of the platoon's reachable set is $\mathcal{L}_\infty$- q string stable, the platoon's dynamics are analyzed in terms of the shifted original coordinates $\tilde{x}_d$ as defined in (6.10). The model describing the dynamics of the platoon in terms of these shifted original coordinates is $\dot{\tilde{x}}_d = A_d\tilde{x}_d + B_d^u u_1 + B_d^\delta(\beta_2)\delta_2$, which is defined in (6.11). These coordinates ensure asymptotic convergence to the origin, corresponding to a completely synchronized vehicle platoon. Since the reachable set induced by the attacks δ_2 is considered, the influence of the control input of the lead vehicle u_1 is not relevant. Therefore, this input u_1 and the state vector of the lead vehicle $\tilde{x}_{d,1}$—which remains unaffected by the attacks—are removed from the platoon's state vector to obtain

$$\underbrace{\begin{bmatrix} \dot{\tilde{x}}_{d,2} \\ \dot{\tilde{x}}_{d,3} \\ \dot{\tilde{x}}_{d,4} \\ \vdots \\ \dot{\tilde{x}}_{d,m} \end{bmatrix}}_{\dot{\tilde{x}}_d} = \underbrace{\begin{bmatrix} A_d^2 & 0 & 0 & \dots & 0 \\ \tilde{A}_d^2 & A_d^3 & 0 & \dots & 0 \\ 0 & \tilde{A}_d^2 & A_d^3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & \tilde{A}_d^{m-1} & A_d^m \end{bmatrix}}_{A_d} \underbrace{\begin{bmatrix} \tilde{x}_{d,2} \\ \tilde{x}_{d,3} \\ \tilde{x}_{d,4} \\ \vdots \\ \tilde{x}_{d,m} \end{bmatrix}}_{\tilde{x}_d} + \underbrace{\begin{bmatrix} B_{d,\delta}^2(\beta) \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{B_d^\delta(\beta)} \delta_2. \tag{8.4}$$

To guarantee $\mathcal{L}_\infty$- q string stability of the vehicle platoon's reachable set, it is aimed to find a pattern in the dynamics of the platoon. To this end, the dynamics of the platoon are represented in the Laplace domain as

$$\hat{\tilde{x}}(s) = G(\beta_2, s)\,\hat{\delta}_2(s) \text{ with } G(\beta_2, s) = (sI - A_d)^{-1}B_d^\delta(\beta_2). \tag{8.5}$$

The transfer function $G(\beta_2, s)$ depends affinely on the controller realization β_2. Recall that the state vector of each vehicle i is $\tilde{x}_i^\top = [\tilde{d}_i,\, \tilde{v}_i,\, a_i,\, u_i]$. First, consider the transfer function from the false data injections δ_2 to the shifted inter-vehicular distance $\tilde{d}_i = d_i - (r_i + hv_r)$—which is the deviation of the inter-vehicular distance with respect to the equilibrium inter-vehicular distance—of all vehicles $i = 2, 3, \dots, m$. This transfer function reads

$$\hat{\tilde{d}}_i(s) = \begin{cases} G_{d,i}(\beta_2, s)\,\hat{\delta}_2(s) = \sum_{j=1}^6 G_{d,i}(\beta_2, s)\mathbf{e}_j\,\hat{\delta}_{2,j}(s) & \text{if } i = 2, 3, 4. \\ H_{\text{lp}}^{i-4}(s)G_{d,4}(\beta_2, s)\,\hat{\delta}_2(s) = \sum_{j=1}^6 H_{\text{lp}}^{i-4}(s)G_{d,4}(\beta_2, s)\mathbf{e}_j\,\hat{\delta}_{2,j}(s) & \text{if } i = 5, 6, \dots, m. \end{cases} \tag{8.6}$$

Here, $H_{\text{lp}} = \frac{1}{hs+1}$ is a low-pass filter and the transfer function $G_{d,i}(\beta_2, s)$ is affine in the sensor configuration β_2. A repeating structure emerges; for large enough i, the transfer function to vehicle i is the transfer function to its predecessor $i-1$, multiplied by the low-pass filter $H_{\text{lp}}(s)$. For the shifted inter-vehicular distance, *this starts at vehicle* 5. To see whether this is also the case for the other states, consider the transfer function from the false data injections δ_2 to the shifted velocity $\tilde{v}_i = v_i - v_r$—which is the velocity deviation with respect to the reference velocity—of all vehicles $i = 2, 3, \dots, m$. This transfer function reads

$$\hat{\tilde{v}}_i(s) = \begin{cases} G_{v,i}(\beta_2, s)\,\hat{\delta}_2(s) = \sum_{j=1}^6 G_{v,i}(\beta_2, s)\mathbf{e}_j\,\hat{\delta}_{2,j}(s) & \text{if } i = 2, 3. \\ H_{\text{lp}}^{i-3}(s)G_{v,3}(\beta_2, s)\,\hat{\delta}_2(s) = \sum_{j=1}^6 H_{\text{lp}}^{i-3}(s)G_{v,3}(\beta_2, s)\mathbf{e}_j\,\hat{\delta}_{2,j}(s) & \text{if } i = 4, 5, \dots, m. \end{cases} \tag{8.7}$$

The transfer function $G_{v,i}(\beta_2, s)$ is affine in the controller realization β_2. The same repeating structure emerges; for large enough i, the transfer function to vehicle i is the transfer function to its predecessor $i-1$, multiplied by the low-pass filter $H_{\text{lp}}(s)$. For the shifted velocity, however, *this already starts at vehicle* 4. Finally, consider the transfer function from the false data injections δ_2 to the (shifted) acceleration a_i of all vehicles $i = 2, 3, \dots, m$. This transfer function reads

$$\hat{a}_i(s) = \begin{cases} G_{a,i}(\beta_2, s)\,\hat{\delta}_2(s) = \sum_{j=1}^6 G_{a,i}(\beta_2, s)\mathbf{e}_j\,\hat{\delta}_{2,j}(s) & \text{if } i = 2, 3. \\ H_{\text{lp}}^{i-3}(s)G_{a,3}(\beta_2, s)\,\hat{\delta}_2(s) = \sum_{j=1}^6 H_{\text{lp}}^{i-3}(s)G_{a,3}(\beta_2, s)\mathbf{e}_j\,\hat{\delta}_{2,j}(s) & \text{if } i = 4, 5, \dots, m. \end{cases} \tag{8.8}$$

The transfer function $G_{a,i}(\beta_2, s)$ is affine in the controller realization β_2. See Appendix A.5 for analytical expressions of the transfer functions $G_{d,i}(\beta_2, s)$, $G_{v,i}(\beta_2, s)$ and $G_{a,i}(\beta_2, s)$. Again, the repeating structure is observed. Similarly to the shifted velocity $\tilde{v}_i$, the repeating structure for the acceleration a_i already *starts at vehicle* 4. The transfer function to the control input u_i is not considered, since this is not a physical state of interest, and it depends on the controller realization, which is ignored in the dynamical system description. Therefore, the control input u_i is not taken into account in the analysis of $\mathcal{L}_\infty$- q string stability in the remainder of this research.

In summary, a repeating structure emerges in the platooning dynamics: for large enough i, the transfer function to vehicle i is the transfer function of its predecessor $i-1$ multiplied by the low-pass filter $H_{\text{lp}}(s)$. For the inter-vehicular distance $\tilde{d}_i$, this starts at vehicle 5. For the velocity $\tilde{v}_i$ and acceleration a_i, this starts at vehicle 4. This pattern enables analysis of the system's $\mathcal{L}_\infty$- q string stability.

8.3 String stability of origin-reachable set approximation

A repeating structure in the platooning dynamics has been established. It is now demonstrated that the *hyperrectangular outer approximation of the platoon's origin-reachable* set satisfies $\mathcal{L}_\infty$-q string stability. In the origin-reachable set, it is imposed that all trajectories start in the origin $\mathcal{X}_0 = \{0\}$. The set of allowed inputs is the hyperrectangle $\mathcal{U}$ in (3.5) with half-widths $\bar{\delta}_{2,j}$. The results on the hyperrectangular outer aproximation of the origin-reachable set in Chapter 5 say that

$$\begin{aligned}
\left\|\tilde{d}_i(\beta_2,t)\right\|_{\mathcal{L}_\infty} &\leq \sum_{j=1}^{6} \|g_{d,i}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1}\ \bar{\delta}_{2,j} = \bar{\bar{d}}_i(\beta_2),\\
\|\tilde{v}_i(\beta_2,t)\|_{\mathcal{L}_\infty} &\leq \sum_{j=1}^{6} \|g_{v,i}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1}\ \bar{\delta}_{2,j} = \bar{\tilde{v}}_i(\beta_2),\\
\|a_i(\beta_2,t)\|_{\mathcal{L}_\infty} &\leq \sum_{j=1}^{6} \|g_{a,i}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1}\ \bar{\delta}_{2,j} = \bar{a}_i(\beta_2).
\end{aligned} \tag{8.9}$$

The result in (5.22) says that the hyperrectangular set $\mathcal{H}$ in (5.1) with half-widths $\bar{\bar{d}}_i(\beta_2)$, $\bar{\tilde{v}}_i(\beta_2)$ and $\bar{a}_i(\beta_2)$ is tight around the origin-reachable set (ignoring the control input u_i). Since a multiplication in the Laplace domain corresponds to a convolution in the time-domain,

$$\begin{aligned}
&G_{d,i}(s)\mathbf{e}_j = H_{\text{lp}}(s)\,G_{d,i-1}(s)\mathbf{e}_j \text{ corresponds to } g_{d,i}(t)\mathbf{e}_j = h_{\text{lp}}(t) * g_{d,i-1}(t)\mathbf{e}_j \text{ for all } i = 5,6,\dots,m,\\
&G_{v,i}(s)\mathbf{e}_j = H_{\text{lp}}(s)\,G_{v,i-1}(s)\mathbf{e}_j \text{ corresponds to } g_{v,i}(t)\mathbf{e}_j = h_{\text{lp}}(t) * g_{v,i-1}(t)\mathbf{e}_j \text{ for all } i = 4,5,\dots,m,\\
&G_{a,i}(s)\mathbf{e}_j = H_{\text{lp}}(s)\,G_{a,i-1}(s)\mathbf{e}_j \text{ corresponds to } g_{a,i}(t)\mathbf{e}_j = h_{\text{lp}}(t) * g_{a,i-1}(t)\mathbf{e}_j \text{ for all } i = 4,5,\dots,m,
\end{aligned} \tag{8.10}$$

(omitting the dependence on β_2 in the notation for clarity). In Appendix A.6, it is shown that the $\mathcal{L}_1$-norm of the impulse response of the low-pass filter $\|h_{\text{lp}}(t)\|_{\mathcal{L}_1} = 1$. Combining this with with Young's inequality for convolutions gives that

$$\begin{aligned}
\|g_{d,i}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1} &\leq \|h_{\text{lp}}(\beta_2,t)\|_{\mathcal{L}_1}\ \|g_{d,i-1}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1} = \|g_{d,i-1}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1} \text{ for all } i = 5,6,\dots,m,\\
\|g_{v,i}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1} &\leq \|h_{\text{lp}}(\beta_2,t)\|_{\mathcal{L}_1}\ \|g_{v,i-1}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1} = \|g_{v,i-1}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1} \text{ for all } i = 4,5,\dots,m,\\
\|g_{a,i}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1} &\leq \|h_{\text{lp}}(\beta_2,t)\|_{\mathcal{L}_1}\ \|g_{a,i-1}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1} = \|g_{a,i-1}(\beta_2,t)\mathbf{e}_j\|_{\mathcal{L}_1} \text{ for all } i = 4,5,\dots,m.
\end{aligned} \tag{8.11}$$

Applying this result to the inequalities in (8.9) gives bounds on the size (i.e., half-widths) of the hyperrectangular set that outer approximates the reachable set such that

$$\begin{aligned}
\bar{\bar{d}}_i(\beta_2) &\leq \bar{\bar{d}}_{i-1}(\beta_2) \text{ for all } i = 5,6,\dots,m,\\
\bar{\tilde{v}}_i(\beta_2) &\leq \bar{\tilde{v}}_{i-1}(\beta_2) \text{ for all } i = 4,5,\dots,m,\\
\bar{a}_i(\beta_2) &\leq \bar{a}_{i-1}(\beta_2) \text{ for all } i = 4,5,\dots,m.
\end{aligned} \tag{8.12}$$

Since no assumptions are made for the controller realization, this result holds true for any controller realization β_2. Consequently, it is guaranteed that for any platoon length m, there exists a hyperrectangular approximation of the origin-reachable set, denoted by $\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})})$, such that

$$\bigcup_{i=2}^{4} \Pi_{x_i}[\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})})] \supseteq \Pi_{x_5}[\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})})] \supseteq \cdots \supseteq \Pi_{x_m}[\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})})]. \tag{8.13}$$

This implies that the hyperrectangular approximation of the origin-reachable set $\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})})$ is $\mathcal{L}_\infty$-3 string stable. Note that the indexing in (8.4) starts at $i = 2$, as the lead vehicle $i = 1$ is unaffected by attacks and therefore excluded from the analysis. This causes the shift in the index such that the approximated set is $\mathcal{L}_\infty$-3 string stable. Intuitively, this thus means that the reachable set of all vehicles $i = 5,6,\dots,m$ is contained by the union of the reachable sets of vehicle $i = 2,3,4$, provided that the trajectory starts in the origin. This origin corresponds to all vehicles being initially synchronized.

Importantly, the platoon's origin-reachable set is $\mathcal{L}_\infty$-3 string stable across all system parameter choices and controller realizations β_2. Young's inequality for convolutions in (8.4) introduces some conservatism. This means that the approximated reachable set could also be $\mathcal{L}_\infty$-q string stable for a smaller value of q. However, such tighter bounds cannot be guaranteed for all combinations of system parameters and controller realizations β_2, hence the conservative bound $q = 3$ is adopted.

8.4 String stability of complete reachable set approximation

It has now been established that the hyperrectangular approximation of the *origin-reachable set* is $\mathcal{L}_\infty$-3 string stable. Now, this result is generalized to include non-zero initial conditions by considering the *autonomous reachable set*. This generalization is practically relevant, as in real-world scenarios the platoon is not always synchronized at the start of an attack.

While it is possible to numerically compute the hyperrectangular approximation of the autonomous reachable set $\mathcal{H}(\mathcal{R}_{(\mathcal{X}_0,\{0\})})$, this process becomes computationally intensive for systems with high-dimensional state spaces. Therefore, a more fundamental analytical approach is pursued.

Recall from (8.13) that the hyperrectangular approximation of the origin-reachable set $\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})})$ is $\mathcal{L}_\infty$-3 string stable. By definition, this implies that any trajectory starting in the origin-reachable set $\mathcal{R}_{(\{0\},\mathcal{U})}$ remains within its approximation $\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})})$ for all time $t \geq 0$. Therefore

$$\text{if } \mathcal{X}_0 = \mathcal{R}_{(\{0\},\mathcal{U})} \text{ then } \mathcal{H}(\mathcal{R}_{(\mathcal{X}_0,\mathcal{U})}) \text{ is } \mathcal{L}_\infty\text{-3-string stable.} \tag{8.14}$$

Verifying whether a specific initial condition lies within the set $\mathcal{X}_0 = \mathcal{R}_{(\{0\},\mathcal{U})}$ can be non-trivial. Hence, the purpose of this result is not to provide a practical verification method, but to demonstrate the existence of a nontrivial set of initial conditions for which the platoon remains string stable. In practical applications, where the length of the platoon is known, numerical techniques can be used to approximate such a set of initial conditions for which $\mathcal{L}_\infty$- q string stability holds.

8.5 Numerical example

To illustrate the $\mathcal{L}_\infty$- q string stability of the hyperrectangular approximation of the platoon's origin-reachable set, the system dynamics in terms of the shifted original coordinates $\tilde{x}_d$ in (8.4) are evaluated. A platoon that consists of $m = 15$ vehicles is considered, each with time constant $\tau = 0.1$ [s]. The CACC uses a time gap $h = 0.5$ [s] with controller gains $k_p = 0.2$ and $k_d = 0.7$ and controller realization $\beta_2 = [0, 0, 0, 0, 0, 0]$. The set of allowed inputs $\mathcal{U}$ in (3.5) is bounded by $\bar{\delta}_{2,j} = 0.1$ for all $j = 1, 2, 3, 4, 5, 6$. The hyperrectangular approximation of the origin-reachable set is computed. Figure 8.1 shows the projection of this approximation onto the state vectors of the first five follower vehicles (again excluding the control input u_i).

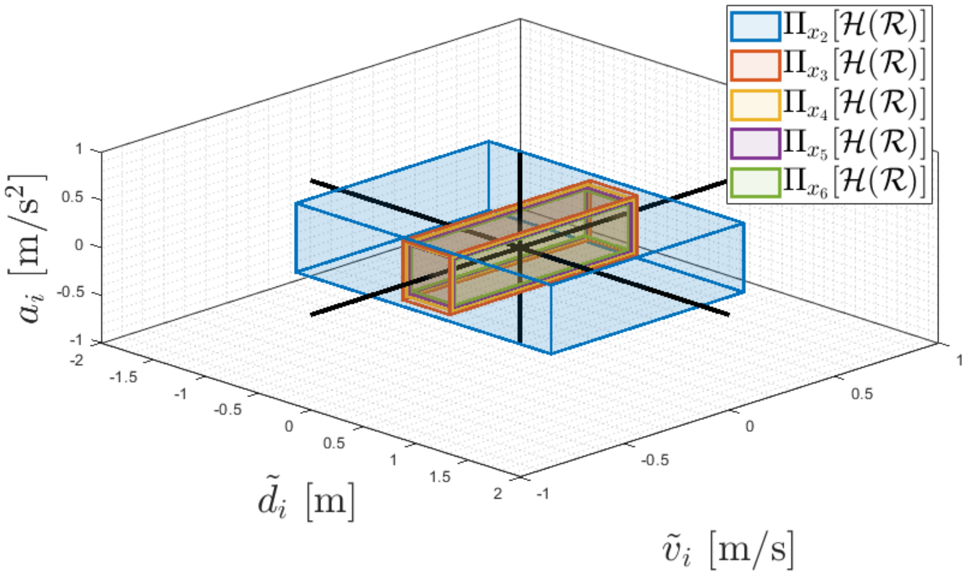

Figure 8.1: Approximated origin-reachable sets of first 5 follower vehicles in a platoon, demonstrating its $\mathcal{L}_\infty$-1-string stability.

The projections reveal the nested structure: the approximated reachable set of the first follower vehicle ($i = 2$) contains the reachable sets of all subsequent follower vehicles $i = 3, 4, 5, 6$ such that

$$\Pi_{x_2}[\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})})] \supseteq \Pi_{x_3}[\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})})] \supseteq \cdots \supseteq \Pi_{x_m}[\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})})]. \tag{8.15}$$

In other words, under the given parameters, the hyperrectangular approximation of the platoon's origin-reachable set $\mathcal{H}(\mathcal{R}_{(\{0\},\mathcal{U})})$ is actually $\mathcal{L}_\infty$-1 string stable. Furthermore, the size of the hyperrectangular outer approximation of each subsequent vehicles is shrinking, suggesting that the effect of the attacks become increasingly attenuated throughout the platoon.

To investigate how the size of each vehicle's approximated reachable set evolves along the platoon, Figure 8.2 plots the size (i.e., half-widths) of the hyperrectangular outer approximation of the origin-reachable set, projected onto the states of the first 14 follower vehicles across various time gaps $h = \{0.1, 0.5, 1, 2\}$ [s].

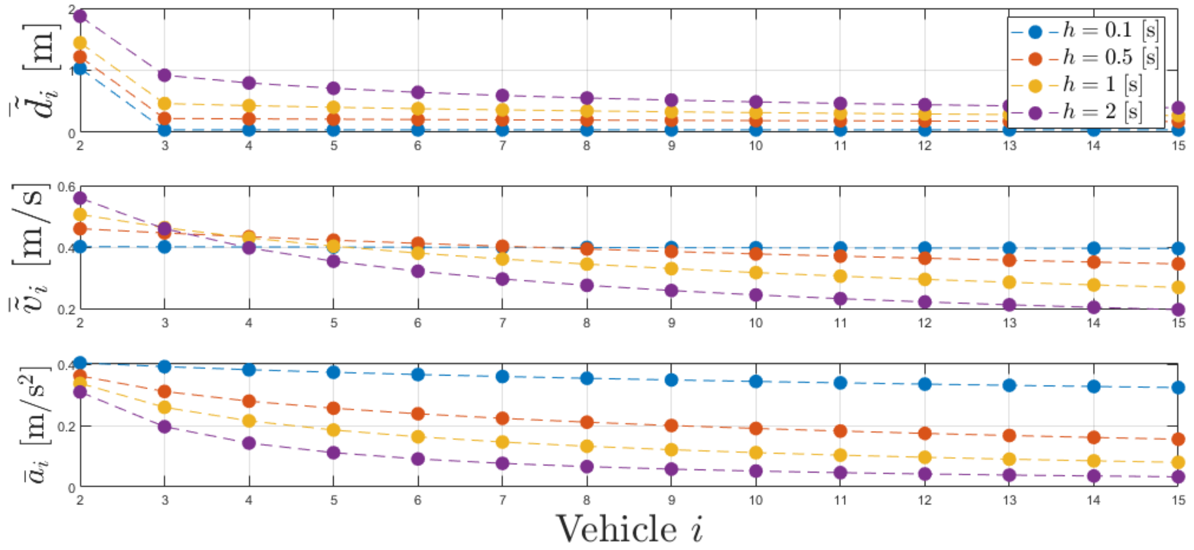

Figure 8.2: Size of hyperrectangular outer approximation of the origin-reachable set, projected onto the states of the first 14 follower vehicles for $h = \{0.1, 0.5, 1, 2\}$ and $\tau = 0.1$ [s].

Across all cases, vehicle $i = 2$ consistently has the largest reachable set, demonstrating $\mathcal{L}_\infty$-1 string stability. As the time gap h increases, the reachable set of vehicle $i = 2$ also increases in the intervehicle distance $\tilde{d}_i$ and velocity $\tilde{v}_i$. However, the reachable sets decrease more rapidly with each subsequent vehicle for larger h, particularly in the shifted velocity $\tilde{v}_i$ and acceleration a_i. This trend may be attributed to the attenuation effect of the low-pass filter $H_{\text{lp}}(s)$, which becomes more suppressive as h increases. Therefore, selecting the time gap h in practical settings involves a trade-off between the magnitude of the reachable set for the first follower and the rate of attenuation of this reachable set along the platoon.

In addition to the time gap, the influence of the vehicle time constant τ on the evolution of the reachable set along the platoon is examined. Figure 8.3 shows the size (i.e., half-widths) of the hyperrectangular outer approximation of the origin-reachable set, projected onto the states of the first 14 follower vehicles across different time constants $\tau = \{0.1, 0.5, 1, 2\}$ [s], with the time gap fixed at $h = 0.5$ [s].

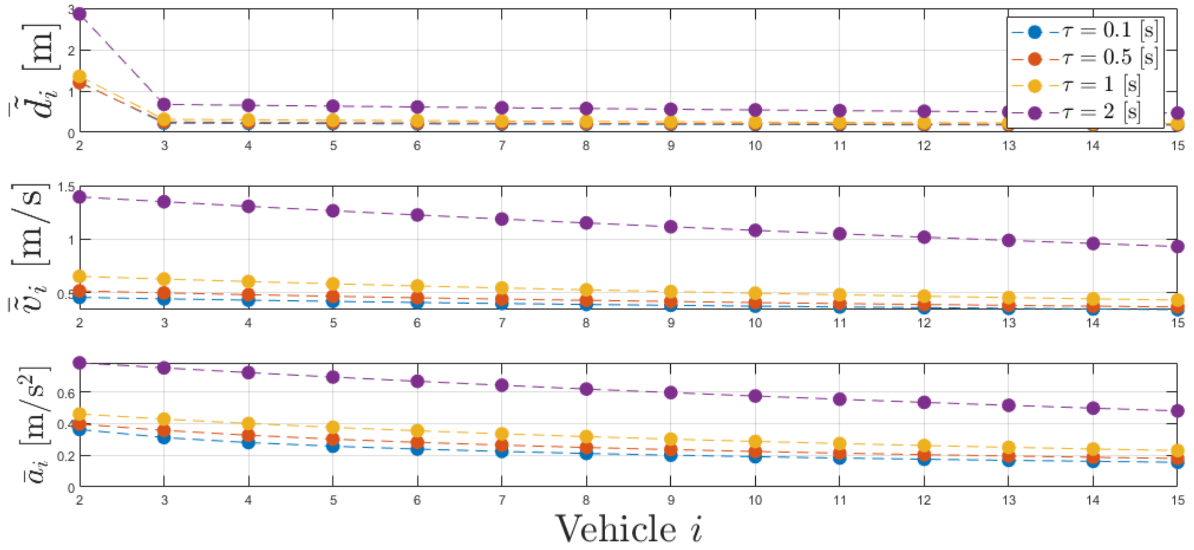

Figure 8.3: Size of hyperrectangular outer approximation of the origin-reachable set, projected onto the states of the first 14 follower vehicles for $\tau = \{0.1, 0.5, 1, 2\}$ and $h = 0.5$ [s].

The results suggests that the rate at which the reachable set shrinks along the platoon is not affected by the value of τ. However, the absolute size of the approximated reachable sets increases with larger τ values. This implies that platoons composed of vehicles with higher time constants (e.g., trucks) are more vulnerable to input perturbations than those composed of vehicles with lower time constants (e.g., passenger cars).

Note that all results in Figure 8.2 and Figure 8.3 consistently suggest that, for typical CACC parameters, the system satisfies the stronger $\mathcal{L}_\infty$-1 string stability condition. Although this thesis has not yet found a formal proof, the observed behavior across all tested cases provides strong empirical evidence that the approximated reachable set of platoon may indeed be inherently $\mathcal{L}_\infty$-1-string stable.

In summary, this chapter has formulated guarantees of the propagation of the vehicles' reachable set along the platoon. These guarantees have been formalized by introducing a new notion of string stability, called $\mathcal{L}_\infty$- q string stability. The platoon under consideration has been guaranteed to be $\mathcal{L}_\infty$- q string stable with $q \leq 3$ under all sensor configurations, which has been demonstrated numerically for typical CACC parameters. In the next chapter, this safety guarantee is exploited to develop a framework to synthesize a sensor configuration that optimizes the robustness of the platoon.

9 Robust controller synthesis

Previous chapters have presented a method to compute the hyperrectangular outer approximation of the origin-reachable set of the platoon and demonstrated that this approximation guarantees $\mathcal{L}_\infty$-q string stability. The volume of the approximation, as well as the index q, depend on the sensor configuration, which is characterized by β_2. In this chapter, an approach is developed to synthesize a controller realization that shapes the reachable set (more precisely, its outer approximation) according to specific design preferences. First, a general controller synthesis framework is formulated for the system in (3.6), cast as an optimization problem. This framework is then tailored to the vehicle platoon model. Finally, its application is demonstrated through a numerical example, and the influence of key system parameters on the optimal solution is examined.

9.1 General controller synthesis framework

Consider the general linear time-invariant continuous-time system in (3.6), where the input matrix $B(\beta_2)$ is affine in the parameter β_2. To minimize the effect of the attacks δ_2, the framework's objective is to find the optimal β_2 that minimizes the volume of the reachable set. Since the reachable set is contained by the Minkowski sum of the origin-reachable set and the autonomous reachable set (see Lemma 5)—and the latter is independent on β_2—the origin-reachable set is considered exclusively. As computing the origin-reachable set is often not tractable, instead its hyperrectangular outer approximation in (5.1) is considered. The size of this hyperrectangle is quantified using a volume measure defined as the conic combination of its half-widths

$$V(\beta_2) = \sum_{i=1}^{n} p_i \, \bar{x}_i(\beta_2) \text{ with } p_i \geq 0. \tag{9.1}$$

The scalars p_i allow scaling of the size of the hyperrectangle, enforcing that certain dimensions become more important. The optimization objective is to minimize $V(\beta_2)$ such that

$$\beta_2^* = \arg\min_{\beta_2} \quad V(\beta_2). \tag{9.2}$$

This optimization program is convex. To see this, Lemma 6 proves that any SISO transfer function that is affine in parameter β_2 has an impulse response function whose $\mathcal{L}_1$-norm is convex in β_2. Lemma 7 then proves that the volume of the hyperrectangle $V(\beta_2)$ is consequently also convex in β_2.

Lemma 6 (Convexity of $\mathcal{L}_1$-norm)**.** *Consider the single-input-single-output (SISO) dynamical system whose dynamics in Laplace domain read $\hat{x}(s) = P(\beta_2, s)\, \hat{u}(s)$ and associated impulse response $p(\beta_2, t)$. If the transfer function is affine in parameter β_2, then the $\mathcal{L}_1$-norm of its impulse response function is convex in the parameter β_2.*

Proof: *Since the transfer function $P(\beta_2, s)$ is affine in parameter β_2, it can be decomposed into $P(\beta_2, s) = P_1(s) + \beta_2 \, P_2(s)$, as well as its impulse response into $p(\beta_2, t) = p_1(\beta_2) + \beta_2 \, p_2(\beta_2, t)$. The $\mathcal{L}_1$-norm of the impulse response function is*

$$\|p(\beta_2, s)\|_{\mathcal{L}_1} = \int_0^\infty f(\beta_2, t) dt \quad \textit{with} \quad f(\beta_2, t) = |p_1(t) + \beta_2 \, p_2(t)| \,. \tag{9.3}$$

For a fixed time t, the function $f(\beta_2, t)$ is convex in parameter β_2, meaning that the inequality in (2.11) holds. Integrating both sides of this inequality over time t gives that

$$\begin{aligned} f(\lambda\beta_2 + (1-\lambda)\beta_2', t) &\leq \lambda f(\beta_2, t) + (1-\lambda) f(\beta_2', t), \\ \int_0^\infty f(\lambda\beta_2 + (1-\lambda)\beta_2', t)\, dt &\leq \int_0^\infty \lambda f(\beta_2, t) + (1-\lambda) f(\beta_2', t)\, dt, \\ \int_0^\infty f(\lambda\beta_2 + (1-\lambda)\beta_2', t)\, dt &\leq \lambda \int_0^\infty f(\beta_2, t)\, dt + (1-\lambda) \int_0^\infty f(\beta_2', t)\, dt, \end{aligned} \tag{9.4}$$

which proves convexity. Consequently, the $\mathcal{L}_1$-norm of the impulse response function $\|p(\beta_2, s)\|_{\mathcal{H}_1} = \int_0^\infty f(\beta_2, t) dt$ is convex in parameter β. ■

Lemma 7 (Convexity of hyperrectangular volume)**.** *Consider the multi-input-multi-output (MIMO) system with dynamics (3.6). If the input matrix $B(\beta_2)$ is affine in the parameter β_2, then the volume of the hyperrectangular outer approximation of the reachable set in (9.1) is convex in parameter β_2.*

Proof: *The size (i.e., half-width) of the hyperrectangular approximation of the reachable set along state x_i is given by the weighted sum*

$$\bar{x}_i(\beta_2) = \sum_{j=1}^{N} \left\| \mathbf{e}_i^\top g(\beta_2, t)\mathbf{e}_j \right\|_{\mathcal{L}_1} \bar{\delta}_j \text{ with } \bar{\delta}_j \geq 0. \tag{9.5}$$

Lemma 6 says that the norm $\left\| \mathbf{e}_i^\top g(\beta_2, t)\mathbf{e}_j \right\|_{\mathcal{L}_1}$ is convex in parameter β_2 for each $j = 1, \ldots, N$. Consequently, since $\bar{\delta}_j \geq 0$, their weighted sum $\bar{x}_i(\beta)$ is also convex in the parameter β_2. Since the volume $V(\beta_2) = \sum_{i=1}^{n} p_i \, \bar{x}_i(\beta_2)$ is the weighted sum of all dimensions $\bar{x}_i(\beta_2)$ with $p_i \geq 0$, this volume $V(\beta_2)$ is also convex in parameter β_2. ■

While alternative volume measures (e.g. based on the product of the length of the sides) may seem more natural, these are in general not convex. Thus, the (slightly less intuitive) conic combination in (9.1) is adopted. This conic combination also allows for tuning of the individual states x_i using the parameters p_i.

9.2 Controller synthesis framework for vehicle platoon

The general synthesis framework in (9.2) is now applied to the vehicle platoon model in (8.4), with β_2 characterizing the controller realization. As shown in Chapter 8, the hyperrectangular approximation of the platoon's origin-reachable set is guaranteed to be $\mathcal{L}_\infty$-3 string stable. That is, the hyperrectangular approximation of the union of the approximated reachable sets of the first 3 follower vehicles contains the approximated reachable sets of all downstream vehicles. Hence, the synthesis is restricted to vehicles 2, 3 and 4. The volume of (the hyperrectangular approximation of) the approximated reachable sets of vehicles 2, 3 and 4 is defined as

$$V(\beta_2) = p_d \max_i \left\{ \bar{\tilde{d}}_i(\beta_2) \right\} + p_v \max_i \left\{ \bar{\tilde{v}}_i(\beta_2) \right\} + p_a \max_i \left\{ \bar{\tilde{a}}_i(\beta_2) \right\} \text{ with } i = 2, 3, 4. \tag{9.6}$$

The scalars $p_d, p_v, p_a \geq 0$ are weighting factors for the inter-vehicular distance, velocity and acceleration respectively. This leads to the convex optimization problem

$$\beta_2^* = \arg\min_{\beta_2} V(\beta_2). \tag{9.7}$$

Since each half-width $\bar{\tilde{d}}_i(\beta_2)$, $\bar{\tilde{v}}_i(\beta_2)$ and $\bar{\tilde{a}}_i(\beta_2)$ is convex in the controller realization β_2 and the pointwise maximum of convex functions remains convex, the entire optimization problem remains convex. This is illustrated in Figure 9.1, which shows the normalized volume $V(\beta_2)$ as a function of individual components $\beta_{2,j}$, with all other components set to zero.

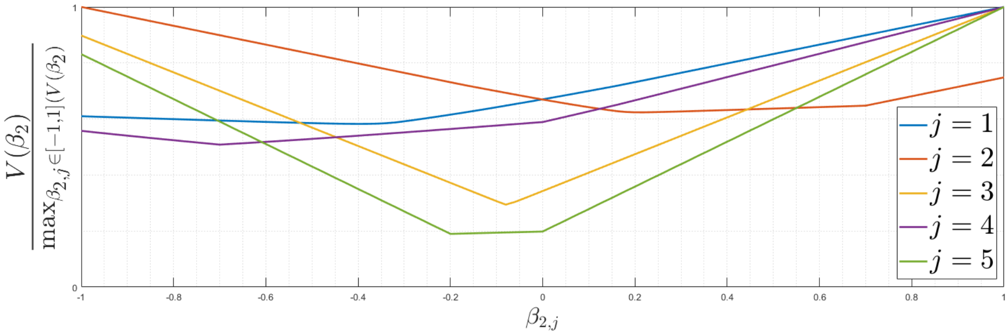

Figure 9.1: Volume of the hyperrectangular set $V(\beta_2)$ (normalized by its maximum volume in the considered domain) in (9.7) as a function of $\beta_{2,i}$, with all $\beta_{2,j\neq i} = 0$ and $p_d, p_v, p_a = 1$, highlighting its convexity in controller realization β.

To ensure safety, a minimum inter-vehicular distance $d_i^{\min}$ is typically enforced such that $d_i \geq d_i^{\min}$. In the optimization, the shifted inter-vehicular distance $\tilde{d}_i = d_i - (r_i + hv_r)$ is considered, rather than the actual inter-vehicular distance d_i. Since the dynamics are invariant under the coordinate translation between these states, as shown in (6.9), the requirement $d_i \geq d_i^{\min}$ can be satisfied by choosing r such that

$$r \geq \min_{i=2,3,4} \left\{ \bar{\tilde{d}}_i(\beta_2) \right\} + d_i^{\min} - hv_r, \tag{9.8}$$

where h is the time gap and v_r is the reference velocity of the lead vehicle. In practice, the goal is minimize $d_i^{\min}$ to reduce aerodynamic drag and improve traffic throughput, while maintaining a positive distance $d_i^{\min} > 0$ to avoid collisions.

9.3 Numerical example

To illustrate the synthesis framework in (9.7), the optimal controller realization is computed for the complete platoon whose dynamics are described by (8.4). Since the optimization problem is convex, standard numerical methods can be employed. The CACC uses gains $k_p = 0.2$ and $k_d = 0.7$, and the set of allowed inputs $\mathcal{U}$ in (3.5) is bounded by $\bar{\delta}_{2,j} = 0.1$ for all $j = 1, 2, 3, 4, 5, 6$. To analyze the influence of the time gap h, the optimal controller realization β_2^* and corresponding optimal volume $V(\beta_2^*)$ are plotted as h varies over $\{0.1, 0.3, 0.5, 0.75, 1, 1.25, 1.5, 2\}$ [s] for a fixed vehicle time constant $\tau = 0.1$ [s]. The results are shown in Figure 9.2. It is observed that the optimal volume $V(\beta_2^*)$ increases with the time gap h, implying that lower time gaps—which improve traffic throughput and reduce drag—also enhance robustness against false data injection attacks.

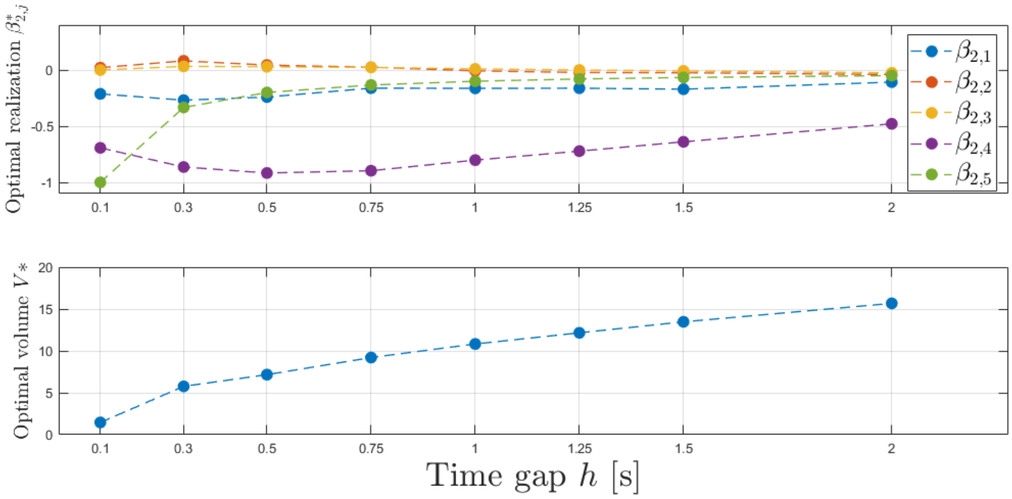

Figure 9.2: Optimal controller realization β_2^* and corresponding optimal volume $V(\beta_2^*)$ as a function of the time gap $h = \{0.1, 0.3, 0.5, 0.75, 1, 1.25, 1.5, 2\}$ [s] for a vehicle time constant $\tau = 0.1$ [s].

Next, the influence of the vehicle time constant τ is investigated. The time gap $h = 0.5$ [s] is fixed, and the time constant $\tau = \{0.1, 0.3, 0.5, 0.75, 1, 1.25, 1.5, 2\}$ [s] is varied. The results, shown in Figure 9.3, indicate that the optimal volume $V(\beta_2^*)$ increases with τ. This implies that vehicle platoons with higher time constants (e.g., trucks) are more susceptible to attacks than those with lower time constants (e.g., passenger cars). Once more, all optimal controller realizations yield a hyperrectangular approximation that satisfies $\mathcal{L}_\infty$-1 string stability, lending further support to the suspicion that the platoon is inherently $\mathcal{L}_\infty$-1 string stable.

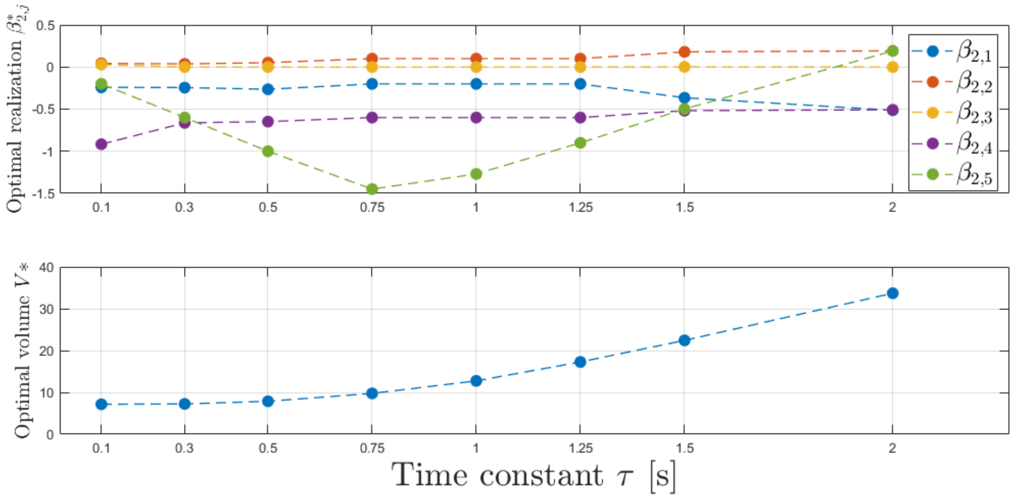

Figure 9.3: Optimal controller realization β_2^* and corresponding optimal volume $V(\beta_2^*)$ as a function of the vehicle time constant $\tau = \{0.1, 0.3, 0.5, 0.75, 1, 1.25, 1.5, 2\}$ [s] for time gap $h = 0.1$ [s].

In summary, this chapter has developed a synthesis framework to find a sensor configuration that maximizes the robustness of the vehicle platoon to attacks. In the next chapter, the main conclusions of this thesis are recalled and summarized, and suggestions for future work are proposed.

10 Conclusions and recommendations

The demand on transportation infrastructure is predicted to continue to grow, leading to increased traffic congestion. In an era of rapid technological advancement and automation, Cooperative Adaptive Cruise Control (CACC) presents a promising solution to enhance highway capacity without requiring physical expansion of existing infrastructure. However, CACC's reliance on onboard sensors and vehicle-to-vehicle wireless communication introduces vulnerabilities to cyberattacks—false data injections into the sensor signals aimed to compromise safety. This thesis has explored the cybersecurity of a specific implementation of CACC that allows multiple realizations, all of which are equivalent in the absence of attacks but significantly differ in their robustness to false data injections. In this chapter, a summary of the thesis and its primary conclusions and recommendations for future research are stated.

Conclusions

Existing literature was reviewed, which measured resilience to attacks by the reachable set induced by the false data injections. This work used an ellipsoid to outer approximate the reachable set, which was computed with a convex optimization program that required discretizing the system dynamics. It also extended this program to a synthesis framework, such that it could be used to find the optimal controller realization that yields the minimum-volume ellipsoid. Though this method was shown suitable for two-vehicle configurations, it was challenging analyze the propagation of attacks through larger platoons.

In order to describe the motion of the complete vehicle platoon, a dynamical model was derived to model and predict the states of all the vehicles. The controller realization did not affect their autonomous dynamics, but only influenced how false data injections entered the system. Based on this model, the attackable subspace—the subspace that can be influenced by attacks—was identified and shown to have reduced dimensionality. This revealed that attackers cannot steer the platoon anywhere, but are constrained to this attackable subspace.

To address the limitations of using ellipsoids to approximate the reachable set, the hyperrectangular (box-shaped) outer approximation was introduced. This approach did not require discretization, but instead relied on the $\mathcal{L}_1$-norm of the system's impulse response, allowing for an analytical expression. This method was then applied to the vehicle platoon, which enabled outer approximation of the reachable set of a platoon of arbitrary length.

Subsequently, the notion of $\mathcal{L}_\infty$- q string stability was introduced, which extended classical notions of string stability by providing formal guarantees on the evolution of the reachable set of vehicles along the platoon. The repeating structure in the platoon's dynamics was exploited to guarantee $\mathcal{L}_\infty$-3 string stability of reachable set's hyperrectangular approximation, which was verified numerically for typical system parameters. Simulations indicated that the attenuation of the reachable set along the platoon increases with the time gap h, and even suggested stronger string stability $q = 1$, though this was not yet formally proven.

Finally, a controller synthesis framework was developed, aimed at finding the realization that minimizes the reachable set's hyperrectangular outer approximation, in the form of a convex optimization program. This framework was then specified such that it could be applied to the vehicle platoon of arbitrary length, using the $\mathcal{L}_\infty$-3 string stability guarantee. The program was solved for typical system parameters, which indicated that the approximated reachable set increases with time gap h and vehicle time constant τ.

In conclusion, this thesis has advanced the understanding of the effect of cyberattacks in cooperative driving by extending existing analyses from two-vehicle platoons to multi-vehicle configurations. In particular, it has introduced safety guarantees that characterize how attacks propagate through larger platoons. Furthermore, a synthesis framework has been developed to determine the sensor configuration for the CACC scheme that maximizes robustness against such attacks. Overall, this work thus contributes to a deeper understanding of cybersecurity in cooperative driving systems and provides tools to enhance their resilience.

Recommendations

The work in this thesis opens several promising directions for further investigation. A list of the most important recommendations for future work is provided below:

Formal proof of strict string stability: While $\mathcal{L}_\infty$-q string stability has been formally guaranteed only for $q = 3$, numerical simulations indicate that stability may also hold for $q = 1$. A natural next step is to reduce the conservatism in the current derivation and provide a formal proof for this stricter case.

Reducing dimensionality of approximation: In this thesis, the concepts of the attackable subspace and the outer approximation of the reachable set were treated separately. A promising direction is to integrate these two concepts, ensuring that the approximated reachable set lies entirely within the attackable subspace—effectively collapsing the dimensionality of the approximation and decreasing conservatism.

Alternative adversarial settings: The current study focused exclusively on attacks targeting the first follower vehicle. However, coordinated attacks on multiple vehicles are both realistic and potentially more harmful. Future work should explore such scenarios, aiming to identify conditions—such as constraints on the number, placement, or correlation of attacks—under which safety guarantees are maintained.

More realistic CACC implementation: Although the vehicle platoon was modeled as a continuous-time system, practical CACC implementations operate in discrete time and often rely on wireless communication, which introduces delays. These factors can affect both performance and stability, yet were not considered in this thesis. Extending the analysis to include discrete-time implementation effects and communication delays would provide a more realistic framework.

Experimental validation: The results in this thesis were validated through numerical simulations. A logical next step is to verify the proposed CACC scheme in a hardware-in-the-loop setup or real-vehicle experiment, to assess whether the theoretical guarantees hold under practical conditions.

References

[1] Cambridge Systematics Inc. and Texas Transportation Institute, "Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation," Tech. Rep., 2005. [Online]. Available: https://ops.fhwa.dot.gov/congestion_report/executive_summary.htm

[2] D. Schrank, T. Lomax, and B. Eisele, "Annual Urban Mobility Report," *Traffic*, no. September, 2011. [Online]. Available: https://transportationops.org/research/annual-urban-mobility-report

[3] L. Xiao and F. Gao, "A comprehensive review of the development of adaptive cruise control systems," *Vehicle System Dynamics*, vol. 48, no. 10, pp. 1167–1192, 2010. [Online]. Available: https://doi.org/10.1080/00423110903365910

[4] V. Milanés and S. E. Shladover, "Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data," *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 285–300, 2014. [Online]. Available: https://doi.org/10.1016/j.trc.2014.09.001

[5] A. Vahidi and A. Eskandarian, "Research advances in intelligent collision avoidance and adaptive cruise control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 4, no. 3, pp. 132–153, 2003. [Online]. Available: https://doi.org/10.1109/TITS.2003.82129

[6] S. E. Shladover, "Automated vehicles for highway operations (automated highway systems)," *Proceedings of the Institution of Mechanical Engineers. Part I: Journal of Systems and Control Engineering*, vol. 219, no. 2, pp. 53–75, 2005. [Online]. Available: https://doi.org/10.1243/095440705X9407

[7] Adaptive Cruise Control Systems—Performance Requirements And Test Procedures, "Standard BS ISO 15 622," Tech. Rep., 2002. [Online]. Available: https://www.iso.org/standard/71515.html

[8] B. Van Arem, C. J. Van Driel, and R. Visser, "The impact of cooperative adaptive cruise control on traffic-flow characteristics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 4, pp. 429–436, 2006. [Online]. Available: https://doi.org/10.1109/TITS.2006.884615

[9] A. Al Alam, A. Gattami, and K. H. Johansson, "An experimental study on the fuel reduction potential of heavy duty vehicle platooning," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, no. October, pp. 306–311, 2010. [Online]. Available: https://doi.org/10.1016/j.ifacol.2024.10.312

[10] D. Swaroop and J. K. Hedrick, "Constant spacing strategies for platooning in automated highway systems," *ASME Journal of Dynamic Systems, Measurement, and Control*, pp. 462–470, 1999. [Online]. Available: https://doi.org/10.1115/1.2802497

[11] G. J. Naus, R. P. Vugts, J. Ploeg, M. J. Van De Molengraft, and M. Steinbuch, "String-stable CACC design and experimental validation: A frequency-domain approach," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4268–4279, 2010. [Online]. Available: https://ieeexplore.ieee.org/document/5571043

[12] S. S. Desoer and C. A., "Longitudinal control of a platoon of vehicles with no communication of lead vehicle information: A system level study," *IEEE Trans. Veh. Technol.*, vol. vol. 42, no. no. 4, pp. 546–554, 1993. [Online]. Available: https://doi.org/10.1109/25.260756

[13] R. R. Zhu and C., "Semi-autonomous adaptive cruise control systems," *IEEE Trans. Veh. Technol.*, pp. 1186–1192, 2002. [Online]. Available: https://doi.org/10.1109/TVT.2002.800617

[14] O. Gehring and H. Fritz, "Practical results of a longitudinal control concept for truck platooning with vehicle to vehicle communication," *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, pp. 117–122, 1997. [Online]. Available: https://doi.org/10.1109/ITSC.1997.660461

[15] J. Hansson and E. Tegling, "A Closed-Loop Design for Scalable High-Order Consensus," *Proceedings of the IEEE Conference on Decision and Control*, pp. 7388–7394, 2023. [Online]. Available: https://arxiv.org/abs/2304.12064

[16] T. v. Oorschot, M. Jeeninga, and E. Tegling, "Experimental verification of a scalable protocol for vehicle platooning," 2024. [Online]. Available: https://vindulamj.github.io/rss24-avas-workshop/papers/

[17] J. Ploeg, B. T. Scheepers, E. Van Nunen, N. Van De Wouw, and H. Nijmeijer, "Design and experimental evaluation of cooperative adaptive cruise control," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, no. 2011, pp. 260–265, 2011. [Online]. Available: https://doi.org/10.1109/ITSC.2011.6082981

[18] M. Amoozadeh, A. Raghuramu, C. N. Chuah, D. Ghosal, H. Michael Zhang, J. Rowe, and K. Levitt, “Security vulnerabilities of connected vehicle streams and their impact on cooperative driving,” *IEEE Communications Magazine*, vol. 53, no. 6, pp. 126–132, 2015. [Online]. Available: https://doi.org/10.1109/MCOM.2015.7120028

[19] Z. Ju, H. Zhang, X. Li, X. Chen, J. Han, and M. Yang, “A Survey on Attack Detection and Resilience for Connected and Automated Vehicles: From Vehicle Dynamics and Control Perspective,” *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 815–837, 2022. [Online]. Available: https://doi.org/10.1109/TIV.2022.3186897

[20] X. Sun, F. R. Yu, and P. Zhang, “A Survey on Cyber-Security of Connected and Autonomous Vehicles (CAVs),” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6240–6259, 2022. [Online]. Available: https://doi.org/10.1109/TITS.2021.3085297

[21] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, “A secure control framework for resource-limited adversaries,” *Automatica*, vol. 51, pp. 135–148, 2015. [Online]. Available: https://doi.org/10.1016/j.automatica.2014.10.067

[22] A. Chowdhury, G. Karmakar, J. Kamruzzaman, A. Jolfaei, and R. Das, “Attacks on self-driving cars and their countermeasures: A survey,” *IEEE Access*, vol. 8, pp. 207 308–207 342, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.3037705

[23] X. M. Zhang, Q. L. Han, X. Ge, D. Ding, L. Ding, D. Yue, and C. Peng, “Networked control systems: A survey of trends and techniques,” *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 1, pp. 1–17, 2020. [Online]. Available: https://doi.org/10.1109/JAS.2019.1911651

[24] S. C. Anand, A. M. Teixeira, and A. Ahlen, “Risk assessment and optimal allocation of security measures under stealthy false data injection attacks,” *2022 IEEE Conference on Control Technology and Applications, CCTA 2022*, no. July, pp. 1347–1353, 2022. [Online]. Available: https://doi.org/10.1109/CCTA49430.2022.9966025

[25] A. Teixeira, H. Sandberg, and K. H. Johansson, “Strategic stealthy attacks: The output-to-output 2-gain,” *Proceedings of the IEEE Conference on Decision and Control*, vol. 54rd IEEE, no. Cdc, pp. 2582–2587, 2015. [Online]. Available: https://doi.org/10.1109/CDC.2015.7402605

[26] R. Van Der Heijden, T. Lukaseder, and F. Kargl, “Analyzing attacks on cooperative adaptive cruise control (CACC),” *IEEE Vehicular Networking Conference, VNC*, vol. 2018-Janua, pp. 45–52, 2017. [Online]. Available: https://doi.org/10.1109/VNC.2017.8275598

[27] M. Huisman, C. Murguia, E. Lefeber, and N. Van De Wouw, “Impact Sensitivity Analysis of Cooperative Adaptive Cruise Control Against Resource-Limited Adversaries,” *Proceedings of the IEEE Conference on Decision and Control*, no. Cdc, pp. 5105–5110, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2304.02395

[28] M. Huisman, C. Murguia, E. Lefeber, and N. V. D. Wouw, “Optimal Controller Realizations against False Data Injections in Cooperative Driving,” no. 101069748. [Online]. Available: https://doi.org/10.48550/arXiv.2404.05361

[29] Wikipedia, “Convex function,” 2025. [Online]. Available: https://en.wikipedia.org/wiki/Convex_function

[30] C. Murguia, I. Shames, J. Ruths, and D. Nesic, “Security Metrics of Networked Control Systems under Sensor Attacks (extended preprint),” no. June, 2018. [Online]. Available: http://arxiv.org/abs/1809.01808

[31] C. Murguia, I. Shames, J. Ruths, and D. Nešić, “Security metrics and synthesis of secure control systems,” *Automatica*, vol. 115, p. 108757, 2020. [Online]. Available: https://doi.org/10.1016/j.automatica.2019.108757

[32] CGAL (Computational Geometry Algorithms Library), “On the bounded number of vertices of polygons.” [Online]. Available: https://doc.cgal.org/Manual/3.3/doc_html/cgal_manual/Minkowski_sum_2/Chapter_main.html

[33] J. Ploeg, N. Van De Wouw, and H. Nijmeijer, “Lp string stability of cascaded systems: Application to vehicle platooning,” *IEEE Transactions on Control Systems Technology*, vol. 22, no. 2, pp. 786–793, 2014. [Online]. Available: https://doi.org/10.1109/TCST.2013.2258346

A Appendix A

This appendix provides supplementary derivations and analytical expressions that support the analysis presented in the main text. While not essential for understanding the main results, they are included for completeness and to provide deeper insight.

A.1 Analytical form of CACC scheme:

In Chapter 3, it was shown that the CACC system dynamics admit an infinite set of realizations under a linear coordinate transformation. Each realization is parametrized by vector β_2, resulting in the model in (3.4). The analytical forms of the functions $f_\xi(\beta_2)$ and $f_y(\beta_2)$, omitted from the main text for clarity, are

$$f_\xi(\beta_2) = \frac{\beta_{2,3}}{\tau} - \frac{1}{h} \quad \text{and} \quad f_y(\beta_2) = \begin{bmatrix} \frac{\beta_{2,1}+\mathrm{k_p}}{h} - \frac{\beta_{2,1}\,\beta_{2,3}}{\tau} \\ \frac{\beta_{2,2}-h\,\mathrm{kp}}{h} - \frac{\beta_{2,2}\,\beta_{2,3}}{\tau} \\ \beta_2 - \beta_{2,4} - \mathrm{k_d} + \frac{\beta_{2,3}}{h} - \frac{\beta_{2,3}\,(\beta_{2,3}+1)}{\tau} \\ \frac{\beta_{2,4}+\mathrm{k_d}}{h} - \frac{\beta_{2,3}\,\beta_{2,4}-\beta_{2,1}\,\tau}{\tau} \\ \frac{\beta_{2,5}+\beta_{2,4}\,h}{h} - \frac{\beta_{2,5}\,(\beta_{2,3}+1)}{\tau} \\ \frac{\beta_{2,5}}{\tau} + \frac{1}{h} \end{bmatrix}^\top. \tag{A.1}$$

A.2 Derivation of non-zero equilibrium position:

In Chapter 6, the system in (6.7), expressed in original coordinates x_d, does not converge to the origin under nominal (attack-free) conditions. Instead, it stabilizes at a non-zero equilibrium x_{eq}. This is demonstrated below. Consider the system without attacks

$$\dot{x}_d = A_d\, x_d + B_d^u\, u_1 + r_d, \tag{A.2}$$

where u_1 is the control input of the lead vehicle. A simple cruise controller ensures convergence to a reference velocity v_r. Consider the cruise controller with dynamics $\dot{u}_1 = -k_v(v_1 - v_r) - k_a a_1 - k_u u_1$. This leads to an augmented state of the lead vehicle $x_{d,1} = [v_1,\, a_1,\, u_1]^\top$, therefore modifying the system matrices of A_d such that

$$A_d^1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -\frac{1}{\tau} & \frac{1}{\tau} \\ -k_v & -k_a & -k_u \end{bmatrix} \text{ and } \tilde{A}_d^1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{k_d}{h} & 0 & \frac{1}{h} \end{bmatrix}. \tag{A.3}$$

Since matrix A_d is Hurwitz, all trajectories of the system asymptotically converge to the equilibrium state x_{eq} such that $\dot{x}_{\mathrm{eq}} = A_d\, x_{\mathrm{eq}} + r_d = 0$. Re-arranging this equation gives that the equilibrium state is $x_{\mathrm{eq}} = -A_d^{-1}\, r_d$, which reads

$$x_{\mathrm{eq}}^\top = \left[\begin{array}{cc|cccc|cccc|c|cccc} v_r & 0 & r + hv_r & v_r & 0 & 0 & r + hv_r & v_r & 0 & 0 & \dots & r + hv_r & v_r & 0 & 0 \end{array}\right]. \tag{A.4}$$

A.3 Coordinate transformation from original to error coordinates:

In Chapter 6, the dynamics of the complete vehicle platoon are described in terms of the original coordinates x_d in (6.4). The coordinate transformation $x_e = T\, x_d + r_t$ is applied, to express the dynamics in terms of error coordinates x_e to obtain (6.15). In other words, the system

$$\dot{x}_d = A_d\, x_d + B_d^u\, u_1 + B_d^\delta(\beta_2)\, \delta_2 \text{ is transformed into } \dot{x}_e = A_e\, x_e + B_e^u\, u_1 + B_e^\delta(\beta_2)\, \delta_2. \tag{A.5}$$

The details of this coordinate transformation to obtain the platooning dynamics in terms of the error coordinates is omitted from the main body for brevity, and therefore presented here. The coordinate transformation from the original coordinates x_d to the error coordinates x_e is

$$x_e = T\, x_d + r_t \text{ or equivalently } x_d = T^{-1}\,(x_e - r_t) \text{ and } \dot{x}_d = T^{-1}\, \dot{x}_e. \tag{A.6}$$

Substituting this into the platooning dynamics in terms of the original coordinates yields

$$\begin{aligned} \dot{x}_d &= A_d\, x_d + B_d^u\, u_1 + r_d + B_d^\delta(\beta)\, \delta_2, \\ T^{-1}\, \dot{x}_e &= A_d\, T^{-1}\,(x_e - r_t) + B_d^u\, u_1 + r_d + B_d^\delta(\beta)\, \delta_2, \\ \dot{x}_e &= TA_dT^{-1}\, x_e - TA_dT^{-1}\, r_t + TB_d^u\, u_1 + TB_d^\delta(\beta)\, \delta_2 \text{ with } TA_dT^{-1}\, r_t = 0, \\ \dot{x}_e &= TA_dT^{-1}\, x_e + TB_d^u\, u_1 + TB_d^\delta(\beta)\, \delta_2, \\ \dot{x}_e &= A_e\, x_e + B_e^u\, u_1 + B_e^\delta(\beta)\, \delta_2 \text{ which gives the system in (6.15).} \end{aligned} \tag{A.7}$$

A.4 Analytical expression of matrix V_d:

In Chapter 7, the attackable subspace of the complete vehicle platoon is identified in terms of the error coordinates x_e and then transformed into the original coordinates x_d. In these original coordinates, the attackable subspace is $\mathcal{X}_d^{\text{att}} = \text{Im}[V^d] + r_t$ with $V^d = T^{-1}V^e$. In the main body, only the projection of this attackable subspace onto each vehicle is presented, and an analytical expression for matrix V^d is omitted. This analytical expression of matrix V^d—which characterizes the attackable subspace in terms of the original coordinates x_d— for a platoon of $m = 4$ vehicles reads

$$V^d = \left[\begin{array}{c|c|c|c|c|c}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
\hline
1 & -h & -\frac{h^2\,\tau}{h-\tau} & -\frac{h^3}{h-\tau} & 0 & 0 \\
0 & -1 & -\frac{h\,\tau}{h-\tau} & -\frac{h^2}{h-\tau} & 0 & 0 \\
0 & 0 & \frac{\tau}{h-\tau} & \frac{h}{h-\tau} & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
\hline
0 & -h & -\frac{h^2\,\tau\,(h-2\,\tau)}{(h-\tau)^2} & -\frac{h^3\,(h-2\,\tau)}{(h-\tau)^2} & -\frac{h^3}{h-\tau} & 0 \\
0 & -1 & -\frac{h\,\tau\,(h-2\,\tau)}{(h-\tau)^2} & -\frac{h^2\,(h-2\,\tau)}{(h-\tau)^2} & -\frac{h^2}{h-\tau} & 0 \\
0 & 0 & -\frac{\tau^2}{(h-\tau)^2} & -\frac{h\,\tau}{(h-\tau)^2} & \frac{h}{h-\tau} & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
\hline
0 & -h & -\frac{h^2\,\tau\,\left(h^2-3\,h\,\tau+3\,\tau^2\right)}{(h-\tau)^3} & -\frac{h^3\,\left(h^2-3\,h\,\tau+3\,\tau^2\right)}{(h-\tau)^3} & -\frac{h^3\,(h-2\,\tau)}{(h-\tau)^2} & -\frac{h^3}{h-\tau} \\
0 & -1 & -\frac{h\,\tau\,\left(h^2-3\,h\,\tau+3\,\tau^2\right)}{(h-\tau)^3} & -\frac{h^2\,\left(h^2-3\,h\,\tau+3\,\tau^2\right)}{(h-\tau)^3} & -\frac{h^2\,(h-2\,\tau)}{(h-\tau)^2} & -\frac{h^2}{h-\tau} \\
0 & 0 & \frac{\tau^3}{(h-\tau)^3} & \frac{h\,\tau^2}{(h-\tau)^3} & -\frac{h\,\tau}{(h-\tau)^2} & \frac{h}{h-\tau} \\
0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]. \tag{A.8}$$

A.5 Analytical expression of transfer functions:

In Chapter 8, the transfer function from the false data injections $\hat{\delta}_2$ to the shifted original coordinates of each vehicle in the platoon is considered. Starting at some vehicle in the platoon, the transfer function to vehicle i is equivalent to the transfer function to vehicle $i-1$, multiplied by a low-pass filter. For the first vehicles in the platoon, however, this is not yet the case. To this end, the analytical expressions of these transfer functions are presented in this appendix. Consider first the transfer function to the shifted inter-vehicular distance, denoted by $G_{d,i}(s)$. The first and second column of this transfer function matrix to vehicle 2, 3 and 4 are

$$\begin{bmatrix} G_{d,2}(s) \\ G_{d,3}(s) \\ G_{d,4}(s) \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}^\top = \begin{bmatrix} -\frac{k_p-\beta_{2,1}hs}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} & \frac{h(k_p+\beta_{2,2}s)}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} \\ -\frac{\beta_{2,1}+s(\beta_{2,1}h-hk_p)+\beta_{2,1}h^2s^2}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & -\frac{\beta_{2,2}+s(k_p h^2+\beta_{2,2}h)+\beta_{2,2}h^2s^2}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} \\ \frac{hs(\beta_{2,1}+k_p)}{(hs+1)^3(\tau s^3+s^2+k_d s+k_p)} & \frac{hs(\beta_{2,2}-hk_p)}{(hs+1)^3(\tau s^3+s^2+k_d s+k_p)} \end{bmatrix}. \tag{A.9}$$

The third and fourth column of this transfer function matrix to vehicle 2, 3 and 4 are

$$\begin{bmatrix} G_{d,2}(s) \\ G_{d,3}(s) \\ G_{d,4}(s) \end{bmatrix} \begin{bmatrix} \mathbf{e}_3 \\ \mathbf{e}_4 \end{bmatrix}^\top = \begin{bmatrix} \frac{\beta_{2,3}h+h\tau(\beta_{2,4}-\beta_{2,2}+k_d+\beta_{2,3}s)}{\tau(hs+1)(\tau s^3+s^2+k_d s+k_p)} & -\frac{k_d+\beta_{2,1}h-\beta_{2,4}hs}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} \\ -\frac{\tau(\beta_{2,3}+s(\beta_{2,3}h-\beta_{2,2}h^2+\beta_{2,4}h^2+h^2k_d)+\beta_{2,3}h^2s^2)+\beta_{2,3}h^2s}{\tau(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & -\frac{\beta_{2,4}-s(hk_d-\beta_{2,4}h+\beta_{2,1}h^2)+\beta_{2,4}h^2s^2}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} \\ -\frac{s(h^2(\beta_{2,3}+\tau(\beta_{2,4}-\beta_{2,2}+k_d))-\beta_{2,3}h\tau)}{\tau(hs+1)^3(\tau s^3+s^2+k_d s+k_p)} & \frac{s(\beta_{2,1}h^2+(\beta_{2,4}+k_d)h)}{(hs+1)^3(\tau s^3+s^2+k_d s+k_p)} \end{bmatrix}. \tag{A.10}$$

The fifth and sixth column of this transfer function matrix to vehicle 2, 3 and 4 are

$$\begin{bmatrix} G_{d,2}(s) \\ G_{d,3}(s) \\ G_{d,4}(s) \end{bmatrix} \begin{bmatrix} \mathbf{e}_5 \\ \mathbf{e}_6 \end{bmatrix}^\top = \begin{bmatrix} \frac{\beta_{2,5}h-h\tau(\beta_{2,4}-\beta_{2,5}s)}{\tau(hs+1)(\tau s^3+s^2+k_d s+k_p)} & -\frac{\tau+\beta_{2,5}h}{\tau(hs+1)(\tau s^3+s^2+k_d s+k_p)} \\ -\frac{\tau(\beta_{2,5}+s(\beta_{2,5}h-\beta_{2,4}h^2)+\beta_{2,5}h^2s^2)+\beta_{2,5}h^2s}{\tau(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & \frac{hs(\tau+\beta_{2,5}h)}{\tau(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} \\ -\frac{s(\beta_{2,5}h^2-h\tau(\beta_{2,5}+\beta_{2,4}h))}{\tau(hs+1)^3(\tau s^3+s^2+k_d s+k_p)} & \frac{hs(\tau+\beta_{2,5}h)}{\tau(hs+1)^3(\tau s^3+s^2+k_d s+k_p)} \end{bmatrix}. \tag{A.11}$$

Now, consider the transfer function to the shifted velocity, denoted by $G_{v,i}(s)$. The first and second column of this transfer function matrix to vehicle 2 and 3 are

$$\begin{bmatrix} G_{v,2}(s) \\ G_{v,3}(s) \\ G_{v,4}(s) \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}^\top = \begin{bmatrix} \frac{(k_p-\beta_{2,1}hs)(s+h)}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} & \frac{h(k_p+\beta_{2,2}s)}{(\tau s^3+s^2+k_d s+k_p)} \\ \frac{\beta_{2,1}(1+hs)^2-k_phs(1+hs)}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & -\frac{h(k_p(1+hs)+\beta_{2,2}s)}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} \\ \frac{hs(k_p+\beta_{2,1})}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & \frac{hs(\beta_{2,2}-hk_p)}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} \end{bmatrix}. \tag{A.12}$$

The third and fourth column of this transfer function matrix to vehicle 2 and 3 are

$$\begin{bmatrix} G_{v,2}(s) \\ G_{v,3}(s) \\ G_{v,4}(s) \end{bmatrix} \begin{bmatrix} \mathbf{e}_3 \\ \mathbf{e}_4 \end{bmatrix}^\top = \begin{bmatrix} -\frac{h(\beta_{2,3}(s+h)+\tau(\beta_{2,3}s+\beta_{2,4}-\beta_{2,2}+k_d))}{\tau(hs+1)(\tau s^3+s^2+k_d s+k_p)} & \frac{(k_d+\beta_{2,1}h-\beta_{2,4}hs)(s+h)}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} \\ \frac{(1+hs)^2(\beta_{2,3}+\tau(\beta_{2,4}-\beta_{2,2}+k_d))-\beta_{2,3}\tau(1+hs)^2}{\tau(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & \frac{(1+hs)(\beta_{2,4}+k_d-\beta_{2,1}hs)}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} \\ \frac{s(h^2(\beta_{2,3}+\tau(\beta_{2,4}-\beta_{2,2}+k_d))-\beta_{2,3}h\tau)}{\tau(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & -\frac{s(\beta_{2,1}h^2+(\beta_{2,4}+k_d)h)}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} \end{bmatrix}. \tag{A.13}$$

The fifth and sixth column of this transfer function matrix to vehicle 2 and 3 are

$$\begin{bmatrix} G_{v,2}(s) \\ G_{v,3}(s) \\ G_{v,4}(s) \end{bmatrix} \begin{bmatrix} \mathbf{e}_5 \\ \mathbf{e}_6 \end{bmatrix}^\top = \begin{bmatrix} -\frac{h(\beta_{2,5}(s+h)-\tau(\beta_{2,4}-\beta_{2,5}s))}{\tau(hs+1)(\tau s^3+s^2+k_d s+k_p)} & \frac{\tau+\beta_{2,5}h}{\tau(\tau s^3+s^2+k_d s+k_p)} \\ \frac{(1+hs)^2\beta_{2,5}+\tau(1+hs)(\beta_{2,4}-\beta_{2,5}s)}{\tau(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & -\frac{(1+hs)(\tau+\beta_{2,5}h)}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} \\ \frac{s(\beta_{2,5}h^2-h\tau(\beta_{2,5}+\beta_{2,4}h))}{\tau(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & -\frac{hs(\tau+\beta_{2,5}h)}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} \end{bmatrix}. \tag{A.14}$$

Finally, consider the transfer function to the acceleration, denoted by $G_{v,i}(s)$. The first and second column of this transfer function matrix to vehicle 2 and 3 are

$$\begin{bmatrix} G_{a,2}(s) \\ G_{a,3}(s) \\ G_{a,4}(s) \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}^\top = \begin{bmatrix} \frac{(k_p-\beta_{2,1}hs)(s+h)s}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} & \frac{hs(k_p+\beta_{2,2}s)}{(\tau s^3+s^2+k_d s+k_p)} \\ \frac{s\big(\beta_{2,1}(1+hs)^2-k_p hs(1+hs)\big)}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & -\frac{hs(k_p(1+hs)+\beta_{2,2}s)}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} \\ \frac{hs^2(k_p+\beta_{2,1})}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & \frac{hs^2(\beta_{2,2}-hk_p)}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} \end{bmatrix}. \tag{A.15}$$

The third and fourth column of this transfer function matrix to vehicle 2 and 3 are

$$\begin{bmatrix} G_{a,2}(s) \\ G_{a,3}(s) \\ G_{a,4}(s) \end{bmatrix} \begin{bmatrix} \mathbf{e}_3 \\ \mathbf{e}_4 \end{bmatrix}^\top = \begin{bmatrix} -\frac{hs(\beta_{2,3}(s+h)+\tau(\beta_{2,3}s+\beta_{2,4}-\beta_{2,2}+k_d))}{\tau(hs+1)(\tau s^3+s^2+k_d s+k_p)} & \frac{(k_d+\beta_{2,1}h-\beta_{2,4}hs)(s+h)s}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} \\ \frac{s((1+hs)^2(\beta_{2,3}+\tau(\beta_{2,4}-\beta_{2,2}+k_d))-\beta_{2,3}\tau(1+hs)^2)}{\tau(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & \frac{s(1+hs)(\beta_{2,4}+k_d-\beta_{2,1}hs)}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} \\ \frac{s^2(h^2(\beta_{2,3}+\tau(\beta_{2,4}-\beta_{2,2}+k_d))-\beta_{2,3}h\tau)}{\tau(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & -\frac{s^2(\beta_{2,1}h^2+(\beta_{2,4}+k_d)h)}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} \end{bmatrix}. \tag{A.16}$$

The fifth and sixth column of this transfer function matrix to vehicle 2 and 3 are

$$\begin{bmatrix} G_{a,2}(s) \\ G_{a,3}(s) \\ G_{a,4}(s) \end{bmatrix} \begin{bmatrix} \mathbf{e}_5 \\ \mathbf{e}_6 \end{bmatrix}^\top = \begin{bmatrix} -\frac{hs(\beta_{2,5}(s+h)-\tau(\beta_{2,4}-\beta_{2,5}s))}{\tau(hs+1)(\tau s^3+s^2+k_d s+k_p)} & \frac{s(\tau+\beta_{2,5}h)}{\tau(\tau s^3+s^2+k_d s+k_p)} \\ \frac{s((1+hs)^2\beta_{2,5}+\tau(1+hs)(\beta_{2,4}-\beta_{2,5}s))}{\tau(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & -\frac{s(1+hs)(\tau+\beta_{2,5}h)}{(hs+1)(\tau s^3+s^2+k_d s+k_p)} \\ \frac{s^2(\beta_{2,5}h^2-h\tau(\beta_{2,5}+\beta_{2,4}h))}{\tau(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} & -\frac{hs^2(\tau+\beta_{2,5}h)}{(hs+1)^2(\tau s^3+s^2+k_d s+k_p)} \end{bmatrix}. \tag{A.17}$$

A.6 Signal 1-norm of low-pass filter:

In Chapter 8, the inequality derived with Young's convolution in (8.11) is used to evaluate the string stability of the platoon. Here, it is used that the $\mathcal{L}_1$-norm of the impulse response function of the low-pass filter $H_{\text{lp}}(s)$ is equal to one, which is proven here. Consider the low-pass filter

$$H_{\text{lp}}(s) = \frac{1}{hs+1} \text{ with impulse response function } h_{\text{lp}}(t) = \mathcal{L}^{-1}\left(H_{\text{lp}}(s)\right) = \frac{1}{h}e^{-\frac{1}{h}t}u(t), \tag{A.18}$$

where $u(t)$ is the unit step function such that $u(t) = 0$ for all $t < 0$ and $u(t) = 1$ for all $t \geq 0$. The $\mathcal{L}_1$-norm of this signal is computed by integrating its absolute value, which yields that

$$\begin{aligned} \|h_{\text{lp}}\|_{\mathcal{L}_1} &= \int_{-\infty}^{\infty} \left| \frac{1}{h}e^{-\frac{1}{h}t}u(t) \right| dt \\ &= \int_{0}^{\infty} \left| \frac{1}{h}e^{-\frac{1}{h}t} \right| dt \\ &= \left[-e^{-\frac{1}{h}t} + C \right]_0^{\infty} \\ &= 0 - (-1) \\ &= 1. \end{aligned} \tag{A.19}$$

EINDHOVEN
UNIVERSITY OF
TECHNOLOGY

Declaration concerning the TU/e Code of Scientific Conduct for the Master’s thesis

I have read the TU/e Code of Scientific Conduct[i].

I hereby declare that my Master’s thesis has been carried out in accordance with the rules of the TU/e Code of Scientific Conduct

Date

19-06-2025 ………………………………………………………..…………..

Name

Thijs Jacob André van Oorschot ………..…………..

ID-number

1352725 ………………………………………………..…………..

Signature

………………………………………………………..…………..

Submit the signed declaration to the student administration of your department.

[i] See: https://www.tue.nl/en/our-university/about-the-university/organization/integrity/scientific-integrity/
The Netherlands Code of Conduct for Scientific Integrity, endorsed by 6 umbrella organizations, including the VSNU, can be found here also. More information about scientific integrity is published on the websites of TU/e and VSNU

February 21, 2020